

# Molecular mechanisms underpinning host *O*-glycan metabolism by human gut symbionts

Wouter van Bakel

A thesis submitted for the degree of Doctor of Philosophy (PhD)  
To the University of East Anglia

Quadram Institute  
Gut Microbes and Health  
Norwich Research Park  
Norwich  
NR4 7UQ

June 2023

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

## Abstract

*Ruminococcus gnavus* is considered as a prevalent member of the 'normal' gut microbiota found in 90% of people and its ability to degrade mucin glycans is strain dependent. *R. gnavus* shows a disproportionate representation in several diseases. Previously, the capacity of *R. gnavus* ATCC 29149 glycoside hydrolases (GHs) to cleave fucose and sialic acid epitopes from mucin glycans has been demonstrated as an important colonisation factor. Here, we showed that *R. gnavus* ATCC 29149 RgGH98 releases blood group A trisaccharide (BgAtri) from mucin glycans, enabling *R. gnavus* E1 strain, which is not able to grow on mucin, to grow on RgGH98-treated mucin. Both *R. gnavus* ATCC 29149 and E1 strain could benefit from the released BgAtri and from access to the underlying mucin glycans. To analyse the full complement of GHs involved in mucin glycan degradation, we performed RNA sequencing (RNAseq) analysis of *R. gnavus* ATCC 29149 grown on mucin glycans. 12 GHs were chosen for further analysis, the activity of three fucosidases on mucin glycans was confirmed. In addition, two GH13 enzymes were analysed; RgGH13-1 and RgGH13-2 were found to be able to hydrolyse  $\alpha$ 1-4 and  $\alpha$ 1-6 links in pullulan, respectively. These enzymes may be implicated in glycogen metabolism. Using stable isotope probing (SIP), we investigated the capacity of gut microbes (from human faecal microbiota) to utilise Neu5Ac and 2,7-anhydro-Neu5Ac. *Collinsella aerofaciens* and *Dorea longicatena* were found to be highly induced in the presence of Neu5Ac while no large difference in the faecal microbiota was observed upon 2,7-anhydro-Neu5Ac supplementation. Together these data demonstrate that the mucin glycan foraging strategy of *R. gnavus* ATCC 29149 is based on the removal and utilisation of mucin glycan epitopes like fucose, Neu5Ac and BgA, and identified new enzymes and pathways that may be involved in *R. gnavus* adaptation to the mucus niche.

## **Access Condition and Agreement**

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

## **Acknowledgements**

This PhD was funded by the European Commission and is part of the Marie Skłodowska-Curie Actions Innovative Training Network: Sweet Crosstalk.

I would like to thank Prof Maher Abou Hachem and Dr Fred Warren for taking their time to read and discuss this PhD thesis.

I want to specifically thank Prof Nathalie Juge for her guidance with my scientific work and my writing process. I want to further thank my supervisory team: Dr Mark Webber, Dr Emmanuelle Crost, Dr Dimitris Latousakis, and Dr Andrew Bell for lots of suggestions and training.

I would like to acknowledge all the support I received from all QIB members who have trained me, performed analyses, or otherwise helped me. I want to specifically thank all members of the Juge group. Thank you Dr Emmanuelle Crost, Dr Dimitris Latousakis, Dr Andrew Bell, and Dr Ryan Griffiths for always helping me with my many questions.

During my project I have had the chance to work with some excellent collaborations: Dr Tom Wennekes (Utrecht Universiteit, The Netherlands), Jun Yang Ong (Utrecht Universiteit), Dr Gwénaëlle Le Gall (University of East Anglia, UK), Dr Serena Monaco (University of East Anglia), Dr Marcela Hernández Garcia (University of East Anglia), Prof J. Colin Murrell (University of East Anglia), and Prof Štefan Janeček (Slovak Academy of Sciences, Slovakia). These people have helped me with their expertise, and many aspects of my project would not have been possible without them.

I want to thank my whole ITN, with which I have had many joyful moments, but also had the chance to discuss our scientific work and even collaborate: Jun Yang Ong, Victor Laplanche, Valentina Borlandelli, Molly Pither, Ferran Nieto, Schahzad Saqib, Erianna Alvarado, Dollwin Matharu, Paola Paone, Romane Breysse, Harshitha Jois, Krishna Desai, Pedro J. Hernando.

I want to thank everyone who has helped me get through these last years. I want to thank my Heavenly Father, for helping me get through both tough and happy times. I want to thank Ana Martínez Gascueña for being the best emotional support I could wish for. Thank you for getting me through my most difficult moments and thank you for sharing my most

precious moments. I want to thank all the friends I have made, who have shared in every complaint, breakthrough, difficulty, and joyful moment: Ana Martínez Gascueña, Victor Laplanche, Anne Jordan, Hélène Yvonne, and Jack Hales. Thank you all for all the lunches, long boardgame evenings, and pub visits.

## Table of contents

<b>Abstract</b> .....	<b>2</b>
<b>Acknowledgements</b> .....	<b>3</b>
<b>Table of contents</b> .....	<b>5</b>
<b>List of abbreviations</b> .....	<b>9</b>
<b>List of figures</b> .....	<b>11</b>
<b>List of tables</b> .....	<b>12</b>
<b>List of supplemental figures</b> .....	<b>13</b>
<b>1. Introduction</b> .....	<b>15</b>
1.1 Architecture and function of the gastrointestinal (GI) tract.....	15
1.1.1 The GI epithelium.....	17
1.1.2 The intestinal mucus layer .....	18
1.1.3 Mucin glycosylation .....	20
1.1.3.1 Chemical structure .....	20
1.1.3.2 Role of intestinal mucin glycosylation in health and disease .....	22
1.2 The gut microbiota.....	24
1.2.1 Structure and function.....	24
1.2.2 Role of the gut microbiota in intestinal diseases.....	30
1.2.3 The mucus-associated microbiota .....	31
1.2.3.1 Composition in health and disease.....	31
1.2.3.2. Factors of colonisation.....	32
1.2.4 Metabolism of host glycans by gut bacteria .....	32
1.3 <i>Ruminococcus gnavus</i> .....	35
1.3.1. Occurrence and role in health and disease.....	35
1.3.2 <i>R. gnavus</i> glycan utilisation.....	36
<b>2. Materials and Methods</b> .....	<b>41</b>
2.1 Materials, strains, and media.....	41
2.1.1 Materials .....	41
2.1.2. Strains .....	41
2.1.3. Media .....	41
2.2 Bacterial growth assays.....	42
2.2.1 <i>R. gnavus</i> growth assays on carbohydrate sources .....	42
2.2.2 Faecal slurry growth assays .....	44
2.3 Molecular biology .....	45

2.3.1 DNA extraction, quantification and quality control.....	45
2.3.2. RNA extraction .....	45
2.3.3 qPCR analysis .....	46
2.3.4 RT-qPCR.....	47
2.3.5 Analysis of RNA sequencing data.....	47
2.4 Heterologous expression .....	48
2.4.1. Cloning of target genes .....	48
2.4.2. Transformation .....	49
2.4.3 Production and purification of recombinant proteins.....	49
2.5 Enzymatic assays.....	50
2.5.1 Enzymatic reactions .....	50
2.5.2. Kinetics .....	50
2.5.3 Enzymatic synthesis of 2,7-anhydro-Neu5Ac.....	51
2.5.4 Thin-layer chromatography (TLC) analysis.....	53
2.5.5 High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) analysis.....	54
2.5.6 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI- ToF-MS) analysis .....	54
2.5.7 NMR analysis.....	54
2.5.8 Electrospray ionization mass spectrometry (ESI-MS) analysis.....	55
2.6 Stable isotope probing (SIP).....	55
2.6.1. Preparation of samples .....	55
2.6.2. Gradient generation, and fractionation by ultracentrifugation.....	55
2.6.3 Analysis of 16S amplicon sequencing data .....	56
2.7 Bioinformatics .....	56
2.7.1 Prediction of protein function and subcellular localisation.....	56
2.7.2 Sequence alignment and AlphaFold modelling .....	57
2.8 Statistical analysis .....	57
<b>3. Role of RgGH98 in growth of <i>R. gnavus</i> strains on mucin glycans .....</b>	<b>59</b>
3.1 Introduction .....	59
3.2 Results.....	61
3.2.1 RgGH98 is a galactosidase active on blood group A antigen.....	61
.....	63
3.2.2 RgGH98 mediates the growth of <i>R. gnavus</i> strains on mucins .....	63
3.2.2.1 Validation of a suitable minimum media for <i>R. gnavus</i> growth .....	63
3.2.2.2 <i>R. gnavus</i> growth on mucin and BgA is strain-specific .....	64

3.2.3 The <i>RgGH98</i> cluster is upregulated during growth of <i>R. gnavus</i> ATCC 29149 on BgAtri and BgAtetra .....	71
3.3 Discussion.....	76
<b>4. Identification and characterisation of <i>R. gnavus</i> GHs involved in mucin glycan degradation.....</b>	<b>80</b>
4.1 Introduction .....	80
4.2 Results.....	81
4.2.1 Growth of <i>R. gnavus</i> ATCC 29149 on mucin glycans .....	81
4.2.2 Transcriptomics analysis of <i>R. gnavus</i> ATCC 29149 .....	84
4.2.2.1 <i>R. gnavus</i> ATCC 29149 gene expression during growth on mucin glycans.....	84
4.2.2.2 Dedicated mucin glycan degrading operons are upregulated during <i>R. gnavus</i> growth on mucin glycans.....	86
4.2.2.3 Selection of GH targets for functional characterisation .....	90
4.2.3 Functional characterisation of selected <i>R. gnavus</i> ATCC 29149 GHs .....	92
4.2.3.1 Heterologous expression and purification of recombinant <i>RgGHs</i> .....	92
4.2.3.2 Enzymatic activity of <i>R. gnavus</i> ATCC 29149 GH29 and GH95 fucosidases.....	93
4.2.3.3 Enzymatic activity of <i>R. gnavus</i> ATCC 29149 GH13 enzymes .....	94
4.2.3.4 <i>R. gnavus</i> ATCC 29149 GHs with no demonstrated functional activity .....	96
4.3 Discussion.....	97
<b>5. Structure and function of <i>R. gnavus</i> GH13 pullulanases .....</b>	<b>101</b>
5.1 Introduction .....	101
5.2 Results.....	103
5.2.1 Structural basis for <i>RgGH13</i> substrate specificity.....	103
5.2.2 <i>RgGH13</i> activity on pullulan.....	106
5.2.3 Functional characterisation of <i>RgGH13-1</i> and <i>RgGH13-2</i> .....	112
5.2.3.1 <i>RgGH13</i> substrate specificity on glucose oligo- and polysaccharides .....	112
5.2.3.2 <i>RgGH13-2</i> enzyme kinetics .....	116
5.2.4 Growth of <i>R. gnavus</i> ATCC 29149 on glucose oligo-and polysaccharides .....	117
5.3 Discussion.....	119
<b>6. 2,7-anhydro-Neu5Ac and Neu5Ac utilisers in the human faecal microbiome .....</b>	<b>123</b>
6.1 Introduction .....	123
6.2 Results.....	126
6.2.1 Enzymatic synthesis of 2,7-anhydro-Neu5Ac from ManNAc and pyruvate.....	126
6.2.2 Stable isotope probing of 2,7-anhydro-Neu5Ac metabolism in human faecal microbiota .....	128
6.2.2.1 Rate of 2,7-anhydro-Neu5Ac utilisation and selection of SIP conditions .....	128
6.2.2.2 Stable isotope probing using <sup>13</sup> C <sub>7</sub> labelled or unlabelled 2,7-anhydro-Neu5Ac and Neu5Ac.....	130



6.2.2.3 Analysis of the gut microbiota composition .....	132
6.3 Discussion.....	135
<b>7. General discussion .....</b>	<b>139</b>
<b>Appendix:.....</b>	<b>143</b>
<b>Appendix 1. Supplemental material.....</b>	<b>143</b>
<b>Appendix 2. LAB stock solutions composition .....</b>	<b>162</b>
<b>References .....</b>	<b>163</b>

## List of abbreviations

2'FL	2'-Fucosyllactose
3'SL	3'-sialyllactose
3FL	3-Fucosyllactose
aa	Amino acid
ADHD	Attention-deficit/hyperactivity disorder
Asp	Aspartic acid
BgA	Blood group A
BgAtetra	BgA tetrasaccharide
BgAtri	BgA trisaccharide
BgB	Blood group B
BgH	Blood group H
BHI-YH	Brain Heart Infusion - Yeast extract - Haemin
BLI	Bio-Layer Interferometry
CAZymes	Carbohydrate-Active Enzymes
Cd	Catalytic domain
CD	Crohn's disease
CRC	Colorectal cancer
CsTFA	Caesium trifluoroacetate
CTP	Cytidine 5'-triphosphate
DCM	Dichloromethane
DMSO	Dimethyl sulfoxide
EHEC	Enterohemorrhagic <i>E. coli</i>
ESI-MS	Electrospray ionisation-mass spectrometry
EtOAc	Ethyl acetate
EtOH	Ethanol
FPKM	Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced
Fuc	Fucose
Gal	Galactose
GalNAc	<i>N</i> -acetylgalactosamine
GF	Germ-free
GHs	Glycoside hydrolases
GI	Gastrointestinal
Glc	Glucose
GlcNAc	<i>N</i> -acetylglucosamine
GlcNAcase	<i>N</i> -acetylglucosaminidase
Glu	Glutamic acid
GNB	Galacto- <i>N</i> -biose
HMOs	Human milk oligosaccharides
HPAEC-PAD	High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection
HRA	Health Research Authority
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome

IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Lac	Lactose
Le <sup>a</sup>	Lewis a
Le <sup>b</sup>	Lewis b
Le <sup>x</sup>	Lewis x
Le <sup>y</sup>	Lewis y
LNnT	Lacto-N-neotetraose
LNT	Lacto-N-tetraose
MeOH	Methanol
MurNAc	<i>N</i> -acetylmuramic acid
MWCO	Molecular Weight Cut-Off
NEB	New England Biolabs
Neu5Ac	<i>N</i> -acetylneuraminic acid
N-term GBLD	N-terminal galactose-binding-like domain
OD	Optical density
PCoA	Principle component analysis
PGM	Pig gastric mucin
PLs	Polysaccharide lyases
<i>p</i> NP	<i>Para</i> -nitrophenol
pPGM	Purified pig gastric mucin
PULs	Polysaccharide utilization loci
RFOs	Raffinose family oligosaccharides
RNAseq	RNA sequencing
RT	Reverse Transcription
SCFAs	Short-chain fatty acids
SEA	Sea urchin-enterokinase-agrin
SIP	Stable isotope probing
TIM	triosephosphate isomerase
TLC	Thin-layer chromatography
UC	Ulcerative colitis
YCFA	Yeast extract - Casitone - Fatty Acids

## List of figures

Figure 1.1. Overview of the layers of the GI tract.....	16
Figure 1.2. Overview of the gut epithelium in the small intestine and colon .....	17
Figure 1.3. Thickness of the mucus layer along the GI tract.....	19
Figure 1.4. Overview of mucus thickness and type throughout the GI tract.....	20
Figure 1.5. Overview of mucin glycosylation .....	22
Figure 1.6. Example of HMO structures. A. building blocks of HMOs .....	26
Figure 1.7. Overview of the adult microbiome in the GI tract.....	27
Figure 1.8. Occurrence of GH families across <i>R. gnavus</i> strains .....	37
Figure 1.9. Schematic representation of <i>R. gnavus</i> ATCC 29149 nan operon .....	38
Figure 3.1. Schematic representation of <i>R. gnavus</i> ATCC 29149 gene-cluster including the RgGH98 gene .....	61
Figure 3.2. RgGH98 enzymatic characterisation .....	63
Figure 3.3. Growth curves of <i>R. gnavus</i> ATCC 29149 and E1 in LAB or YCFA medium supplemented with 0.5% glucose .....	64
Figure 3.4. Growth of <i>R. gnavus</i> ATCC 29149 and E1 in LAB supplemented with untreated or RgGH98-treated pPGM .....	65
Figure 3.5. Analysis of BgAtri utilisation during growth of <i>R. gnavus</i> strains on mucin .....	66
Figure 3.6. qPCR analysis of <i>R. gnavus</i> growth in LAB supplemented with RgGH98-treated or untreated mucin .....	67
Figure 3.7. Growth curves of <i>R. gnavus</i> ATCC 29149 and E1 on BgAtri and BgAtetra.....	69
Figure 3.8. Composition of mucin glycans following <i>R. gnavus</i> growth on untreated and RgGH98-treated mucin.....	70
Figure 3.9. Growth of <i>R. gnavus</i> ATCC 29149 and E1 on dialysed RgGH98-treated pPGM.....	71
Figure 3.10. Growth of <i>R. gnavus</i> ATCC 29149 on BgAtri and BgAtetra .....	72
Figure 3.11. Copy numbers per ml during <i>R. gnavus</i> ATCC 29149 growth on BgAtri and BgAtetra.....	73
Figure 3.12. Transcriptomics analysis of RgGH98 cluster during <i>R. gnavus</i> ATCC 29149 growth on BgAtri or BgAtetra compared to growth on glucose .....	75
Figure 3.13 Model of <i>R. gnavus</i> mucin glycan degradation strategy.....	78
Figure 4.1. MALDI-TOF MS analysis of $\beta$ -eliminated mucin glycans from PGM .....	82
Figure 4.2. Growth of <i>R. gnavus</i> ATCC 29149 and E1 on mucin and mucin-released glycans.....	83
Figure 4.3. Growth of <i>R. gnavus</i> ATCC 29149 in LAB supplemented with 0.5% glucose or 1% mucin glycans.....	84
Figure 4.4. Volcano plot of significantly up- and downregulated genes during <i>R. gnavus</i> ATCC 29149 growth on mucin glycans versus Glc.....	85
Figure 4.5. Expression of RgGH98 gene cluster during growth on mucin glycans.....	87
Figure 4.6. Expression of nan cluster during growth of <i>R. gnavus</i> ATCC 29149 on mucin glycans ..	88
Figure 4.7. Expression of prophage cluster during growth of <i>R. gnavus</i> ATCC 29149 on mucin glycans.....	89
Figure 4.8. HPAEC-PAD analysis of RgGH13 enzymatic reaction on mucin glycans .....	95
Figure 4.9. NMR analysis of RgGH13-2 enzymatic reaction with mucin glycans.....	96
Figure 5.1. Schematic overview of pullulan degradation products .....	103
Figure 5.2. TLC analysis of the enzymatic reaction of RgGH13-1 and RgGH13-2 on pullulan .....	108
Figure 5.3. AlphaFold model of RgGH13-1 superimposed on <i>Geobacillus stearothermophilus</i> neopullulanase TRS40 in complex with panose.....	110
Figure 5.4. AlphaFold model of RgGH13-2 superimposed on <i>Paenibacillus barengoltzii</i> type I pullulanase .....	111

Figure 5.5. TLC analysis of the enzymatic reaction of RgGH13-1 D357N and RgGH13-2 D336A catalytic mutants, and RgGH13-2 CBM48 domain with pullulan .....	111
Figure 5.6. Analysis of RgGH13-1 substrate specificity .....	113
Figure 5.7. Analysis of RgGH13-2 substrate specificity .....	114
Figure 5.8. TLC analysis of the enzymatic reaction of RgGH13-1 and RgGH13-2 on glycogen .....	115
Figure 5.9. RgGH13-2 activity on pullulan turnover rate at a range of pullulan concentrations ...	116
Figure 5.10. Growth curves of <i>R. gnavus</i> ATCC 29149 on glucose oligo- and polysaccharides .....	118
Figure 6.1. Chemical overview of the enzymatic synthesis of 2,7-anhydro-Neu5Ac from Neu5Ac .....	127
Figure 6.2. TLC and ESI-MS monitoring the synthesis of 3'SL .....	128
Figure 6.3. Utilisation of 2,7-anhydro-Neu5Ac and Neu5Ac by human faecal microbiota in batch model .....	129
Figure 6.4. Relative 16S cDNA copy numbers per gradient fraction of faecal slurry grown on Neu5Ac .....	129
Figure 6.5. Overview of the average refractive index of SIP fractions .....	130
Figure 6.6. Relative bacterial 16S cDNA copy numbers per SIP fraction .....	131
Figure 6.7. PCoA plots of the heavy and light fractions of bacteria grown on <sup>12</sup> C and <sup>13</sup> C Neu5Ac and 2,7-anhydro-Neu5Ac .....	133
Figure 6.8. Relative abundance of bacterial species in heavy and light SIP fractions per major bacterial phyla of faecal microbiota grown on sialic acid derivatives .....	134

## List of tables

Table 2.1. Conditions of activity assays. ....	50
Table 2.2. Reaction overview of ManNAc + Pyruvate → Neu5Ac .....	53
Table 2.3. Reaction overview of Neu5Ac + Lactose-Ncbz + CTP → 3'SL-Ncbz. ....	53
Table 3.1. Description of the genes in <i>R. gnavus</i> ATCC 29149 RgGH98 cluster .....	61
Table 3.2. Concentrations of BgAtri released from pPGM by RgGH98 .....	68
Table 3.3. Concentration of BgAtri and BgAtetra in medium during growth of <i>R. gnavus</i> ATCC 29149. ....	74
Table 4.1. Gene expression analysis of GH genes during <i>R. gnavus</i> ATCC 29149 growth on mucin glycan vs Glc .....	85
Table 4.2. Predicted gene functions of <i>R. gnavus</i> ATCC 29149 prophage genes RGna_RS06995 to RGna_RS07120 .....	89
Table 4.3. Overview of selected <i>R. gnavus</i> ATCC 29149 GHs. ....	91
Table 4.4. Fold change increase of RGna_RS04290 (RgGH29-2) gene during growth on mucin glycans vs Glc. ....	92
Table 4.5. Enzymatic activity of <i>R. gnavus</i> ATCC 29149 fucosidases .....	94
Table 5.1. Overview of genes flanking RgGH13-1 (RGNA_RS06060) .....	105
Table 5.2. Overview of genes flanking RgGH13-2 (RGna_RS08510) .....	106
Table 5.3. Enzyme kinetics of RgGH13-2 with pullulan. ....	116
Table 6.1. Overview of published SIP experiments in microbial communities .....	125

## List of supplemental figures

Supplemental Table S1. Overview of characterised GHs from gut bacteria active on mucin glycans. ....	143
Supplemental Table S2. Primers used in this study .....	145
Supplemental Table S3. RNA concentration and quality extracted from <i>R. gnavus</i> ATCC 29149 growth on glucose, BgAtri, or BgAtetra .....	146
Supplemental Table S4. RNA concentration and quality extracted from growth on glucose and mucin glycans.....	146
Supplemental Table S5. List of selected <i>R. gnavus</i> ATCC 29149 GHs.....	147
Supplemental Table S6. Overview of all glycans used as substrate for <i>R. gnavus</i> growth.....	148
Supplemental Figure S1. Agarose gel electrophoresis of amplified <i>R. gnavus</i> GHs by PCR .....	149
Supplemental Figure S2. Agarose gel electrophoresis of recombinant plasmids encoding selected <i>R. gnavus</i> GHs .....	150
Supplemental Figure S3. SDS-PAGE of selected <i>R. gnavus</i> recombinant GHs.....	151
Supplemental Figure S4. MALDI-TOF MS analysis of RgGH13-2 activity on pPGM mucin glycans	152
Supplemental Figure S5. TLC analysis of the enzymatic reaction of RgGH73 and RgGH23 with peptidoglycan.....	153
Supplemental Figure S6. HPAEC-PAD analysis of enzymatic reaction of RgGHs on mucin glycans. ....	154
Supplemental Figure S7. HPAEC-PAD analysis of enzymatic reaction of RgGH88 and RgGH2-1 on predicted substrates .....	155
Supplemental Figure S8. Amino acids sequence alignment of RgGH13-1 and RgGH13-2 with fully annotated pullulanase sequences .....	156
Supplemental Figure S9. Relative abundance of bacterial phyla in heavy and light SIP fractions .	157
Supplemental Figure S10. Relative abundance of Actinobacteriota species in heavy and light SIP fractions per major bacterial phyla.....	158
Supplemental Figure S11. Relative abundance of Bacteroidota species in heavy and light SIP fractions per major bacterial phyla.....	159
Supplemental Figure S12. Relative abundance of Firmicutes species in heavy and light SIP fractions per major bacterial phyla.....	160
Supplemental Figure S13. Relative abundance of Proteobacteria species in heavy and light SIP fractions per major bacterial phyla.....	161

# Chapter 1

## Introduction

## **1. Introduction**

### **1.1 Architecture and function of the gastrointestinal (GI) tract**

The human gastrointestinal (GI) tract is responsible for the digestion of food and uptake of nutrients. The upper GI tract consists of the mouth, oesophagus, stomach, duodenum, jejunum, and ileum. The lower GI tract is made up of the colon, rectum, and the anus (Greenwood-Van Meerveld et al., 2017). In the jejunum and ileum, nutrients are further broken down and absorbed (Kiela & Ghishan, 2016), while in the colon, the undigested food is exposed to the gut microbiota which degrades complex carbohydrates (Den Besten et al., 2013). Furthermore, many of the remaining salts and most of the water are extracted in the colon (Debongnie & Phillips, 1978; Sandle, 1998). The GI tract contains four layers: the innermost layer is the mucosa, underneath is the submucosa, followed by the muscularis propria and finally, the outermost layer - the adventitia (Figure 1.1). The structure of these layers varies, in different regions of the digestive system, depending on their function. The adventitia, or in certain parts called the serosa, consists of supporting tissue. The muscularis propria consists of muscle cells responsible for rhythmic peristalsis. The submucosa contains blood vessels, lymphatic vessels, submucosal glands, and nerves. The mucosa consists of the epithelium and the mucus layer (Reed & Wickham, 2009).



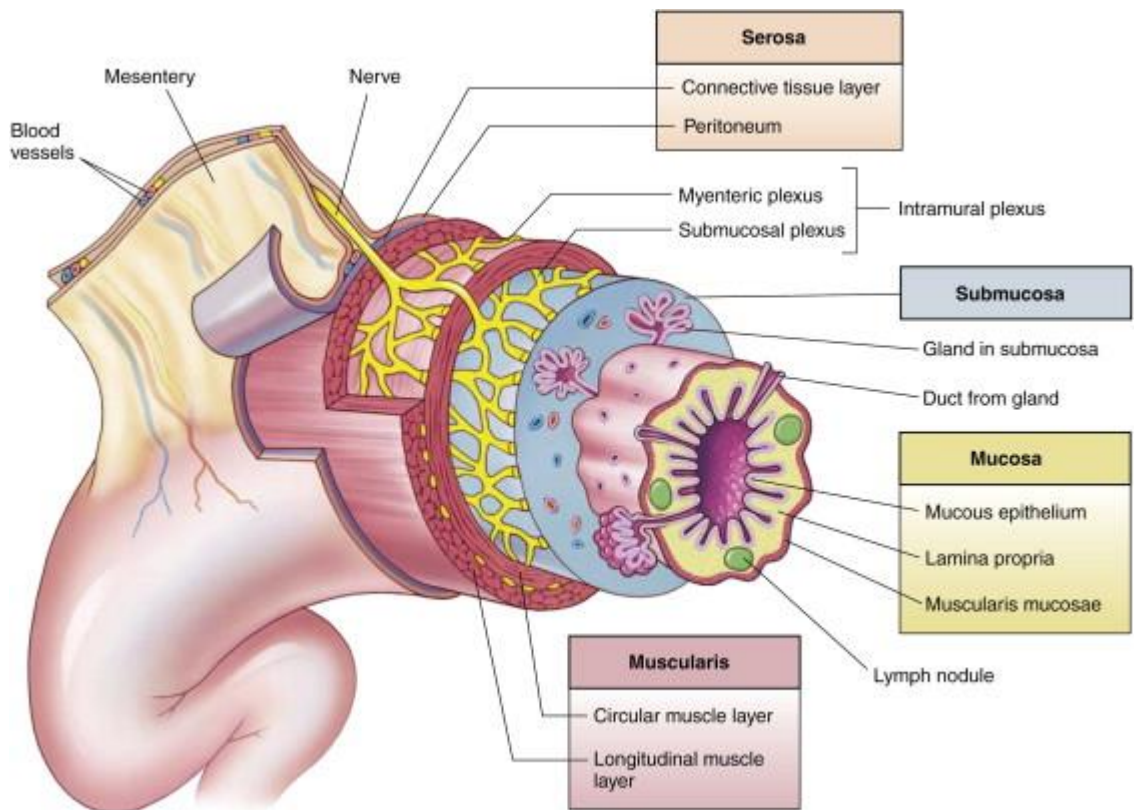


Figure 1.1. Overview of the layers of the GI tract. Reprinted from (Reed & Wickham, 2009), with permission from Elsevier.

### 1.1.1 The GI epithelium

The epithelial layer covering the GI tract is made of a monolayer of cells which separates the body from the intestinal lumen. The epithelium is involved in nutrient uptake and protection of the tissue against the microbes in the GI tract. The epithelial layer is renewed through differentiation of stem cells residing in the crypts of Lieberkühn. These crypts form pits in the intestinal layer; at the bottom, the stem cells divide and their daughter cells migrate upwards while differentiating into different cell types (van der Flier & Clevers, 2009) (Figure 1.2). Many cells in this layer are absorptive enterocytes which are responsible for the uptake of nutrients and other metabolic tasks (Peterson & Artis, 2014). Other specialised cells in the epithelium include goblet cells, which produce mucins and other mucus components (Y. S. Kim & Ho, 2010), Paneth cells, which are only located in the small intestine, and produce a range of anti-microbial peptides (Bevins, 2006; Hooper et al., 2003), and enteroendocrine cells which produce various hormones (Engelstoft et al., 2013). The epithelial cells have multiple pathogen-recognition receptors, which are able to recognise the presence of microorganisms close to the epithelium, the epithelial cells can

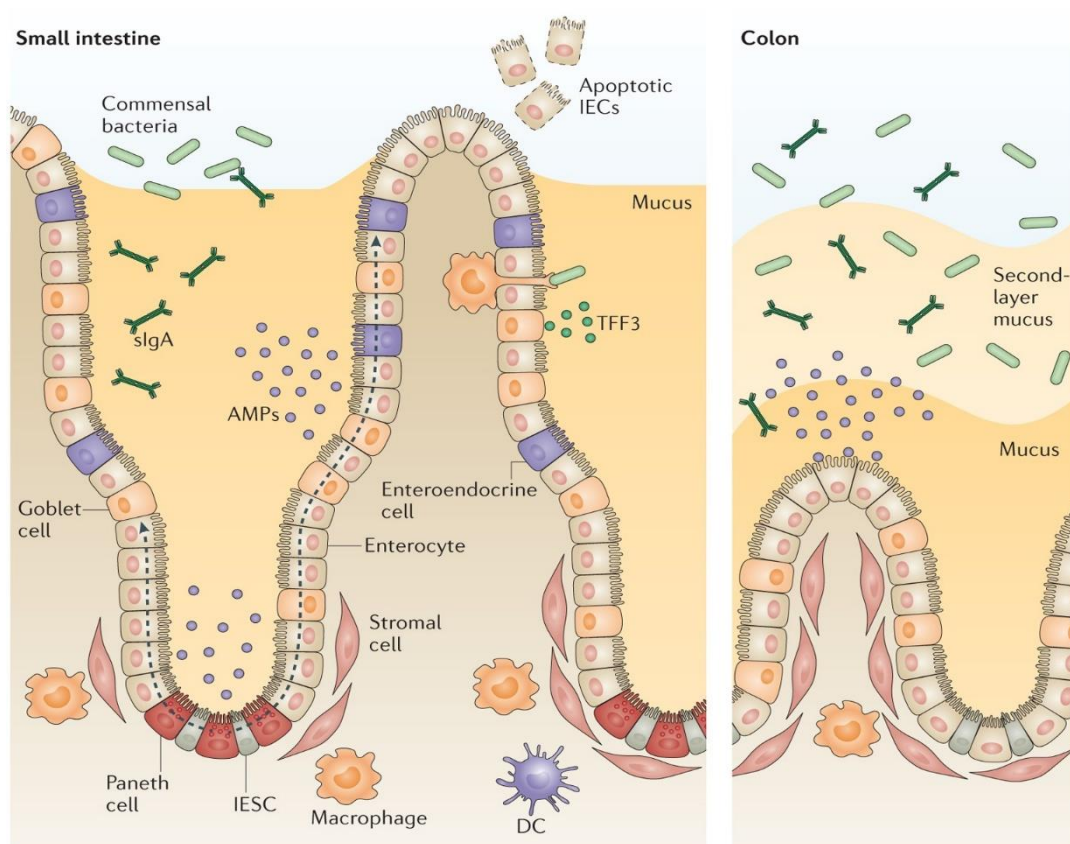


Figure 1.2. Overview of the gut epithelium in the small intestine and colon. Image adapted from (Peterson & Artis, 2014), with permission from Springer Nature.

work together with immune cells to mount an immune reaction (Santaolalla & Abreu, 2012). In the small intestine the surface area of the gut is greatly increased by the villi, projections of the epithelium into the lumen.

### **1.1.2 The intestinal mucus layer**

The mucus layer together with a gradient of anti-microbial peptides protects the epithelium from the large numbers of microorganisms in the gut. The organisation and composition of mucus varies along the GI tract (Figure 1.3). In the stomach and the colon, the epithelium is lined with a double layered mucus, whereas a single mucus layer covers the small intestine (Etienne-Mesmin et al., 2019; Johansson & Hansson, 2016; Juge, 2012). In the colon, the inner layer consists of highly structured mucin networks which are almost impenetrable to bacteria (Johansson et al., 2008; Johansson, Holmén Larsson, et al., 2011) and attached to the epithelium (Atuma et al., 2001; Johansson, Ambort, et al., 2011). In contrast, the outer mucus layer facing the lumen is loose and heavily colonised with bacteria. The composition of these two layers is highly similar (Johansson et al., 2008), and the outer layer derives from the inner layer (Ermund et al., 2013). MUC2 is the main component of the colonic mucus layers (Figure 4), MUC2 is a gel-forming mucin which forms large polymers (> 100 MDa) (Johansson, Ambort, et al., 2011). Inside the cells, MUC2 is tightly packed and kept in this form due to a low pH and high  $\text{Ca}^{2+}$  concentration, during secretion the pH is increased and the  $\text{Ca}^{2+}$  concentration drops, facilitating a 1000-fold expansion of the mucins (G. C. Hansson, 2019). Once secreted, mucins form layers of tight net-like sheets (Ambort et al., 2012; Johansson et al., 2008), which physically block bacteria from entering the inner layer (Ermund et al., 2013; Johansson et al., 2008) and can encapsulate bacteria in mucus to immobilise them (Bergstrom et al., 2020). The concentration of MUC2 is four times higher in the firmer inner layer compared to the outer layer (Johansson et al., 2008), leading to an increase in size of four times (Johansson, Holmén Larsson, et al., 2011). A mucus associated lectin-like protein, ZG16, was found to further prevent gram-positive bacteria from entering the inner mucus layer (Bergström et al., 2016).

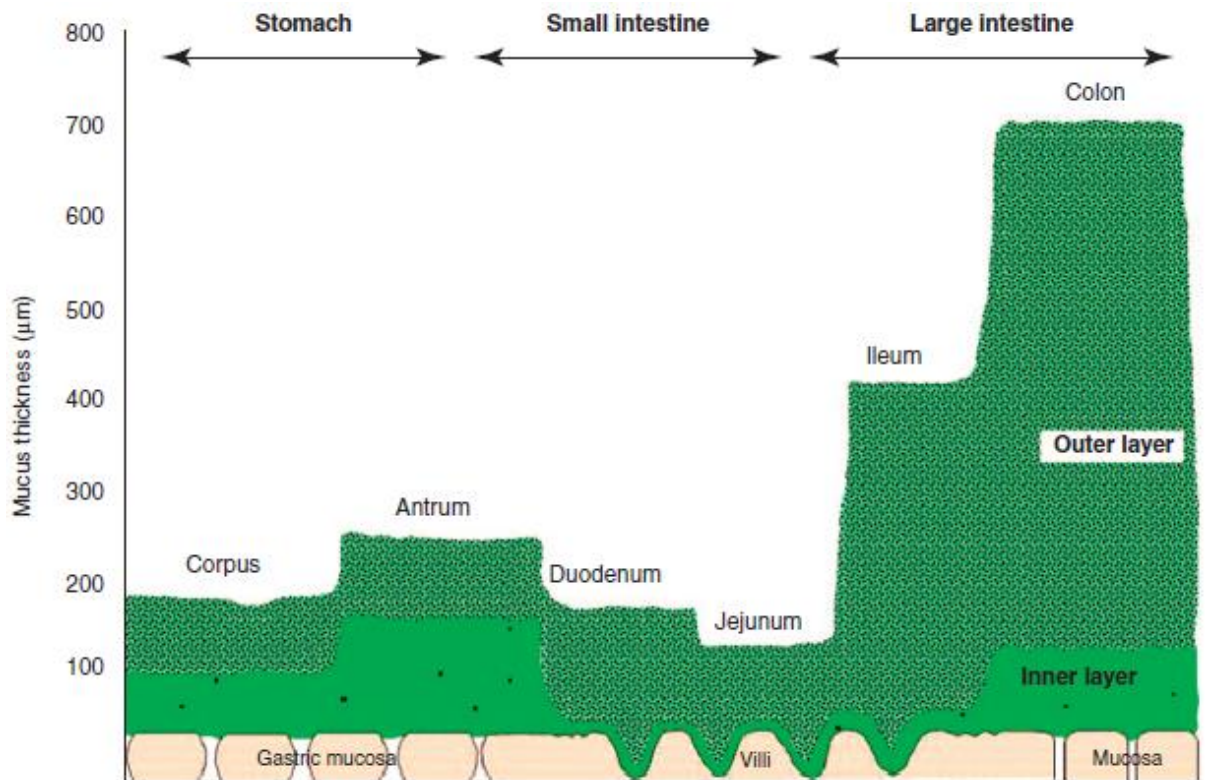


Figure 1.3. Thickness of the mucus layer along the GI tract. Reprinted from (Juge, 2012), with permission from Elsevier.

In addition to secreted mucins, the glycocalyx of the GI tract comprises membrane-bound or transmembrane mucins (Figure 1.4). Transmembrane mucins are, similarly to secreted mucins, heavily glycosylated with complex glycan structures (Pelaseyed & Hansson, 2020). Most transmembrane mucins contain one or more SEA (sea urchin-enterokinase-agrin) domains, these domains are located between the heavily glycosylated extracellular domain and the transmembrane domain. The SEA domains are cleaved after the protein is formed, but still held in place by a strong non-covalent forces. This is to protect the epithelium from tearing if forces are shearing the mucins (Macao et al., 2006; Pelaseyed et al., 2013). Transmembrane mucins cover epithelial cells in a thick glycan cover, providing protection by physical means, but the release of mucins helps keeping pathogens at bay that have bound to the mucins (Lindén et al., 2009; Pelaseyed & Hansson, 2020). Transmembrane mucins have also been shown to be involved in forming the shape of cells (Shurer et al., 2019).

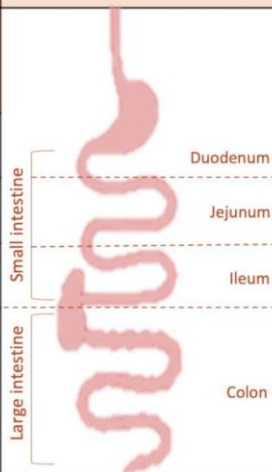
MICE	GASTROINTESTINAL TRACT	HUMANS				
MUCUS THICKNESS		GOBLET CELLS/ ENTEROCYTE RATIO	MUCINS			
			Secreted	Transmembrane		
~ 500 $\mu\text{m}$	Duodenum	4 %	MUC2, MUC6	<i>Functions:</i> <ul style="list-style-type: none"> <li>Source of nutrients and attachment sites for the gut microbiota</li> </ul>	MUC1, 12, 15, 17	<i>Functions:</i> <ul style="list-style-type: none"> <li>Attachment sites for the gut microbiota</li> <li>Involved in host-microbiota interactions</li> </ul>
~ 250 $\mu\text{m}$	Jejunum	6 %	MUC2	<ul style="list-style-type: none"> <li>Gel forming-mucin properties</li> <li>Diffusion barrier</li> </ul>	MUC1, 12, 15, 17	<ul style="list-style-type: none"> <li>Glycocalyx formation</li> <li>Cell protection</li> <li>Intracellular signal transduction</li> <li>Sensors for the luminal milieu</li> </ul>
~ 200 $\mu\text{m}$	Ileum	12 %	MUC2		MUC1, 12, 15, 17	
~ 110-150 $\mu\text{m}$	Colon	16 %	MUC2, MUC5B		MUC3, 4, 12, 15, 17, 20, 21	

Figure 1.4. Overview of mucus thickness and type throughout the GI tract. Reprinted from (Paone & Cani, 2020), under the Creative Commons CC-BY-NC license.

### 1.1.3 Mucin glycosylation

#### 1.1.3.1 Chemical structure

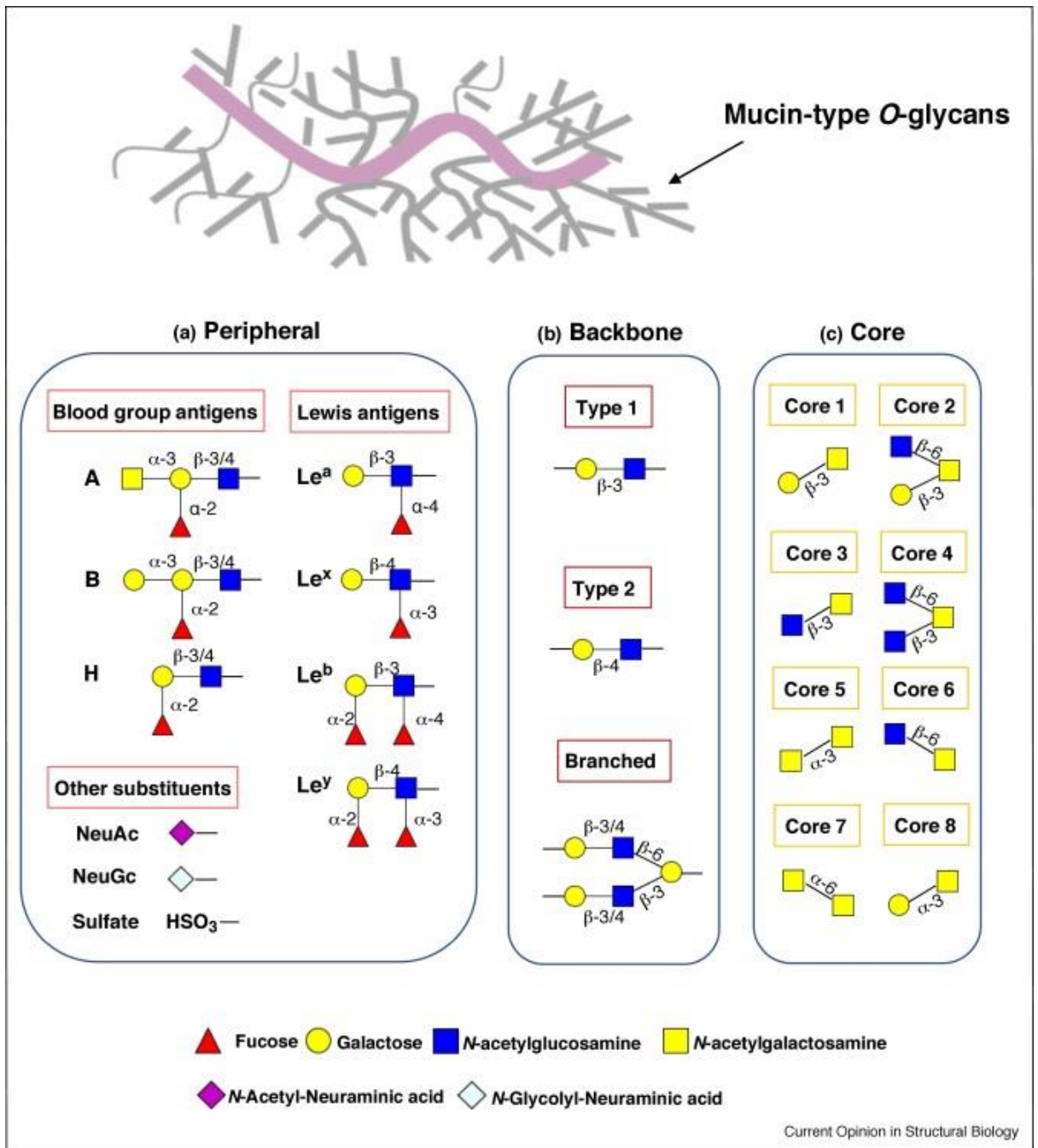
Mucins are glycoproteins which consists for more than 80% of their weight of *O*-glycans (Gunnar C. Hansson, 2020) and the sidechains are complex glycans of up to 12 monomers attached by glycosidic bonds of different linkages (Holmén Larsson et al., 2009) (Figure 1.5). Mucin glycosylation starts with the addition of *O*-linked *N*-acetylgalactosamine (GalNAc) to serine or threonine in the mucin protein via the action of a GalNAc-transferase (ppGalNAcT) (Brockhausen et al., 2022; Juge, 2012; Peter-Katalinić, 2005). From this first GalNAc, eight core structures can be synthesised by glycosyltransferases, of which the first four are most common in the GI tract (Figure 1.5) (Guzman-Aranguez & Argüeso, 2010). Core 1 consists of a galactose (Gal)  $\beta$ 1-3 attached to GalNAc by the Gal-transferase C1GALT1, to produce Gal $\beta$ 1-3GalNAc, while core 3 consists of GlcNAc $\beta$ 1-3GalNAc, due to the activity of GlcNAc-transferase  $\beta$ 3GNT6. Core 2 is an elongation of core 1 while core 4 is an elongation of core 3 with  $\beta$ 1-6 GlcNAc, resulting in GlcNAc $\beta$ 1-6(Gal $\beta$ 1-3)GalNAc and GlcNAc $\beta$ 1-6(GlcNAc $\beta$ 1-3)GalNAc, respectively. The GlcNAc is added to these cores by GlcNAc-transferases GCNT1, GCNT3 and GCNT4. These core structures are further elongated in several ways. In type 1 chains, a Gal residue is added in a  $\beta$ 1-3 linkage to a terminal *N*-acetylglucosamine (GlcNAc). In Type 2 chains, Gal is added in a  $\beta$ 1-4 linkage to a terminal GlcNAc, resulting in a lactosamine unit (Gal $\beta$ 1-4GlcNAc). Type 2 chains can be further elongated from both branches but with a preference for the branch originating from the core starting with the  $\beta$ 1-6 GlcNAc (found in Core 2 and 4). The formation of Type 1 and 2 is performed by several

GlcNAc-transferases and Gal-transferases, like B3GNT3 and B3GALT5. Type 2 branches are more likely to polymerise than Type 1 branches, however both can gain poly-*N*-acetyllactosamines units. Type 2 chains can be capped by a GlcNAc  $\alpha$ 1-4 attached to the terminal Gal, an example Type 2 glycan would be:  $\text{GlcNAc}\alpha$ 1-4 $\text{Gal}\beta$ 1-4 $\text{GlcNAc}\beta$ 1-6( $\text{GlcNAc}\alpha$ 1-4 $\text{Gal}\beta$ 1-3)GalNAc. Gal in these chains can potentially form branching points by adding a GlcNAc with a  $\beta$ 1-6 linkage. (Brockhausen et al., 2022). Glycan structures with core 3 are the most prevalent in the human colonic mucus (Capon et al., 2001; Holmén Larsson et al., 2009), while core 1 and 2 are more common in stomach mucin (Slomiany et al., 1984).

Following further elongation, these structures are usually terminated with Fucose (Fuc), *N*-acetylneuraminic acid (Neu5Ac) or sulphate (Figure 1.5) (Juge et al., 2016; Z. Li & Chai, 2019). Fucose can be attached  $\alpha$ 1-2 to Gal and  $\alpha$ 1-3 or  $\alpha$ 1-4 to GlcNAc (Brockhausen et al., 2022). Fucosyltransferases FUT1 and FUT2 can add fucose to synthesise blood group H (BgH) epitopes. With GalNAc-transferases and Gal-transferases a GalNAc or Gal is added to BgH in order to produce blood group A (BgA) and blood group B (BgB) epitopes, respectively (Brockhausen et al., 2022). Neu5Ac is often attached either  $\alpha$ 2-3 or  $\alpha$ 2-6 to mucin glycans by several sialyltransferases. In addition to ABO blood group epitopes, Lewis epitopes are commonly found in mucin glycans as well. Lewis structures are either of type 1 ( $\text{Gal}\beta$ 1-3GlcNAc) or type 2 ( $\text{Gal}\beta$ 1-4GlcNAc). Lewis a ( $\text{Le}^a$ ) is produced by the action of FUT3, which adds a  $\alpha$ 1-4 fucose to the GlcNAc of a type 1 chain. Lewis b ( $\text{Le}^b$ ) is formed by the simultaneous expression of FUT2 and FUT3 and is similar to  $\text{Le}^a$  with an additional fucose  $\alpha$ 1-2 attached to the Gal. Lewis x ( $\text{Le}^x$ ) and Lewis y ( $\text{Le}^y$ ) are similar to  $\text{Le}^a$  and  $\text{Le}^b$ , but on a type 2 chain. Because the type 2 chain consists of  $\text{Gal}\beta$ 1-4GlcNAc the fucose is added  $\alpha$ 1-3.  $\text{Le}^y$  is similar to  $\text{Le}^b$  but on a type 2 chain. A combination of FUT4, 5, 6, and 7 are involved in  $\text{Le}^x$  and  $\text{Le}^y$  synthesis. Neu5Ac can be added  $\alpha$ 2-3 to the terminal gal instead of fucose, this results in sialyl-lewis a and sialyl-lewis x (Stanley et al., 2022).

In humans, there is an increasing gradient of sialic acid in mucin from the small intestine throughout the colon. This gradient is reversed for fucosylated glycans (Robbe et al., 2003, 2004). Interestingly, a reverse pattern is observed in mice (Holmén Larsson et al., 2013).





**Figure 1.5.** Overview of mucin glycosylation. **(A)** common mucin epitopes. **(B)** mucin backbone. **(C)** Mucin core structures. The way glycans are linked is described by first showing the configuration of the linkage in either  $\alpha$  or  $\beta$ . The following numbers indicate which type of linkage it is. This starts with the anomeric carbon (reducing end), which is generally carbon number 1. The second number indicates the carbon to which the linkage has formed. This originally was any unmodified hydroxyl group attached to a carbon, generally 2, 3, 4, or 6. So an example of Gal $\beta$ 1-3 GalNAc would be a linkage of the 1st carbon (C1) of Gal via an ether bond to the third carbon (C3) of GalNAc. This example is in a  $\beta$  configuration. Reprinted from (Z. Li & Chai, 2019), with permission from Elsevier.

### 1.1.3.2 Role of intestinal mucin glycosylation in health and disease

The human gut is colonised by trillions of microbes, including potential pathogenic bacteria. The GI tract has evolved strategies to combat microbes or keep them at bay by the production of antimicrobials, antibodies, and by reducing colonisation. The GI tract also

promotes colonisation by commensal bacteria, which in turn not only take space and nutrients away from potential pathogens, but also are able to themselves combat pathogens. As described above, the mucus layer is one of the major protective measures against potential microbial threats, not only by physically blocking threats, but also by modulating the microbial community by the assortment of mucin glycan epitopes (Bergstrom & Xia, 2022). In addition, accumulating evidence from both clinical and animal studies indicates that intestinal mucins and mucin-type *O*-glycans play an important role in inflammatory bowel disease (IBD), colorectal cancer (CRC), and pathogen infection, as reviewed recently (Grondin et al., 2020; Pothuraju et al., 2021; Y. Zhang et al., 2021).

Mucin-type *O*-glycans, primarily core 1- and core 3-derived *O*-glycans, are the major mucus barrier components throughout the GI tract (see section 1.1.3.1). Using transgenic mouse models, it was shown that mice lacking core 1- and/or core 3-derived intestinal *O*-glycan showed more severe chronic colitis and spontaneous CRC tumours (An et al., 2007; Bergstrom et al., 2016; Fu et al., 2011). Mucin glycan chains are shorter in ulcerative colitis (UC) (Holmén Larsson et al., 2011), in line with studies in humans showing that in UC patients the mucus layer is thinner (Strugala et al., 2008) (for a review see, (Fang et al., 2021)). The FUT2 gene is responsible for the secretor status, which determines the ABO blood group antigens in secreted mucins (Kelly et al., 1995). Non-secretors, which carry a homozygous inactivating mutation of the FUT2 gene, have been shown to be more susceptible for Crohn's disease (CD) (McGovern et al., 2010). It is likely that this is due to the effect the ABO blood group antigens have on shaping the gut microbiota (Y. Qin et al., 2022; Tong et al., 2014).

In several cancer types, including colon cancer, a mutation in a  $\beta$ 1–3 galactosyltransferase pathway results in the tumour associated Tn antigen (Yu et al., 2015), a truncated mucin glycan consisting of only the initial GalNAc. Upregulated gene expression in ST6GalNAc-1, a sialyltransferase, results in increased immunosuppressive tumour associated STn antigen (Neu5Ac $\alpha$ 2-6GalNAc) frequency (Beatson et al., 2015). Both unsialylated and sialylated lewis epitopes have been shown to be upregulated in several cancer types, including mucins in colon cancer (Pearce, 2018). For an extensive review of cancer related glycan epitopes, see (Pearce, 2018).



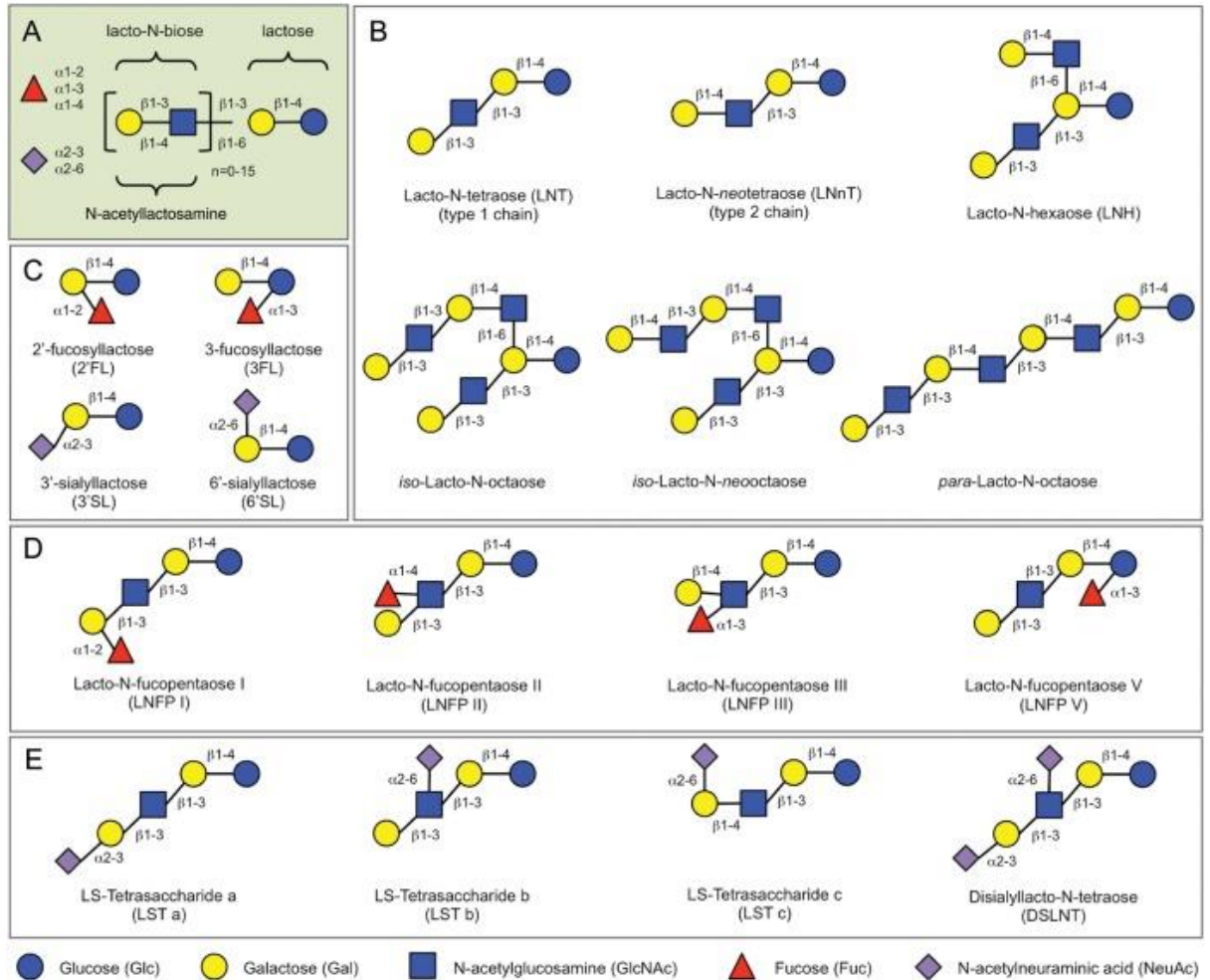
Fucose released from mucin glycans can influence the virulence of enterohaemorrhagic *E. coli* (Pacheco et al., 2012). Similarly, the release of fucose from mucin glycans by *Bacteroides fragilis* has been reported to promote the growth of *Campylobacter jejuni* in the presence of mucins (Luijckx et al., 2020). Several pathogens like *E. coli*, *C. difficile*, and *Salmonella typhimurium* species do not encode any sialidases, and are unable to degrade mucin glycans which are capped by sialic acid. When uncapped by sialidases of other species, growth of these pathogens can be promoted, either due to direct growth on sialic acid or the access to underlying mucin glycans, indicating a protective function of sialic acid residues (Y. L. Huang et al., 2015; Ng et al., 2013).

## **1.2 The gut microbiota**

### **1.2.1 Structure and function**

At birth, infants are virtually sterile, but they are immediately exposed to many microorganisms from different sources. Already the mode of birth gives rise to different microorganisms; bacterial diversity is lower in children born by caesarean section than by vaginal birth, furthermore, the composition of microorganisms is different (Rutayisire et al., 2016). The introduction of bacteria continues with close contact of the child with the mother. Then, with breastfeeding more bacteria are introduced (Hunt et al., 2011; Solís et al., 2010). The gut microbiota of breast-fed infants is dominated by *Bifidobacterium* and other milk-associated bacteria (Ehrlich et al., 2020; Turroni et al., 2012). In infants, human milk oligosaccharides (HMOs) play a key role in shaping the gut microbiota. HMOs are complex sugars found in breast milk in concentrations ranging from 10-15 g/l (Azad et al., 2018; Bode, 2015; Kunz et al., 2000). Until now, more than two hundred HMOs have been identified (Kobata, 2010; Petschacher & Nidetzky, 2016). All HMOs are made from the same five monosaccharides: glucose (Glc), Gal, Fuc, GlcNAc, and Neu5Ac (Bode, 2015; Kunz et al., 2000). HMO synthesis starts with a lactose residue, which is a combination of  $\beta$ 1-4 linked Gal and Glc. This can be expanded with either  $\beta$ 1-3 or  $\beta$ 1-4 linked Gal-GlcNAc. These Gal-GlcNAc disaccharides can be added to the backbone in two ways, either with a  $\beta$ 1-3 link, which elongates the chain, or with a  $\beta$ 1-6 link, which creates a branch. This can form complex chains which can be modified with Fuc or Neu5Ac (Figure 1.6) (Bode, 2012, 2015; Kunz et al., 2000).

Humans are unable to break down HMOs and these compounds therefore reach the colon mostly intact (Engfer et al., 2000) where they are degraded and metabolised by bacteria (LoCascio et al., 2007; Marcobal et al., 2010; Ward et al., 2006; B. Zhang et al., 2022) and help shape the gut microbiota in the first days of an infant. Bacteria that degrade HMOs like *Bifidobacterium longum* subsp. *infantis*, *Bifidobacterium bifidum*, and *Bifidobacterium breve* dominate the infant gut microbiota during breastfeeding (for a recent review see Ioannou et al., 2021)). *Bifidobacterium* species are specialised in HMO metabolism and gain a growth advantage when grown in the presence of HMOs (Bode, 2015; LoCascio et al., 2007; Ward et al., 2006). However, other gut symbionts have been shown to benefit from HMO either directly or through cross-feeding including *Akkermansia muciniphila*, which is specialised in mucus degradation and associated with human health (Kostopoulos et al., 2020), or Clostridiales species, which utilises similar pathways in HMO and mucin glycan metabolism (Pichler et al., 2020). This expansion in beneficial bacteria can help suppress the growth of pathogenic bacteria due to competition for nutrients (for review see Walsh et al., 2020). Because HMOs are built from similar building blocks as epithelial surface or mucin glycans, they can also serve to combat pathogens which use these glycans as nutrient or binding sites (Bode, 2015; Kunz et al., 2000; Walsh et al., 2020) or be used to promote gut barrier function (Natividad et al., 2022; Šuligoj et al., 2020).



**Figure 1.6.** Example of HMO structures. A. building blocks of HMOs. B. different examples of  $\beta$ 1-3 linked Gal-GlcNAc (type 1 chain) and  $\beta$ 1-4 linked Gal-GlcNAc (type 2 chain). Branching occurs when Gal-GlcNAc is linked  $\beta$ 1-6 to the backbone, while elongation occurs with a  $\beta$ 1-3 link. C. different tri-saccharide HMOs with either Fuc or Neu5Ac. D-E. Addition of either Fuc or Neu5Ac generates new isomers. Reprinted from: Bode, L. (2012), with permission from Oxford University Press.

Towards the age of 3, the gut microbiota stabilises and becomes more adult-like (Koenig et al., 2011; Yatsunenکو et al., 2012). The most abundant organisms in the adult gut are bacteria, but archaea (C. Hoffmann et al., 2013; Mafra et al., 2022), viruses (Minot et al., 2011; Nayfach et al., 2021), fungi (Ghannoum & Mukherjee, 2013; C. Hoffmann et al., 2013; Pérez, 2021) and other eukaryotes (Hooks & O'Malley, 2020; Parfrey et al., 2011) are also present, albeit in lower abundancies and many species are only transient in the gut.

The gut microbiota helps mature the immune system, secretory IgA modulates the gut microbiome by specifically targeting certain invasive or highly abundant species, while improving a rich diversity (Sutherland et al., 2016). Microbial molecular patterns, like LPS and flagellin, further engage the immune system by activating several immune receptors like Toll-like receptors and cytoplasmic NOD-like receptors (van den Elsen et al., 2019). This

can activate inflammatory pathways, but can also teach the adaptive immune system to recognise these molecular patterns (K. D. McCoy et al., 2019).

The abundance of bacteria in the small intestine is much lower than in the colon (Sender et al., 2016; Zoetendal et al., 2012) due to a higher oxygen concentration, higher acidity, high amounts of antimicrobial compounds, and a much shorter transit time, all contributing to inhibit extensive colonisation (Donaldson et al., 2015) (Figure 1.7). The jejunum is dominated by Firmicutes and Proteobacteria (Sundin et al., 2017), while Bacteroidetes and Clostridia are more abundant in the ileum (Vasapolli et al., 2019). In the colon the total bacterial abundance is reaching  $10^{11}$  per ml (Sender et al., 2016). Species from the phyla Bacteroidetes and Firmicutes dominate (Vasapolli et al., 2019); other major phyla in the colon of adults are the Actinobacteria, Proteobacteria, and Fusobacteria (Andersson et al., 2008).

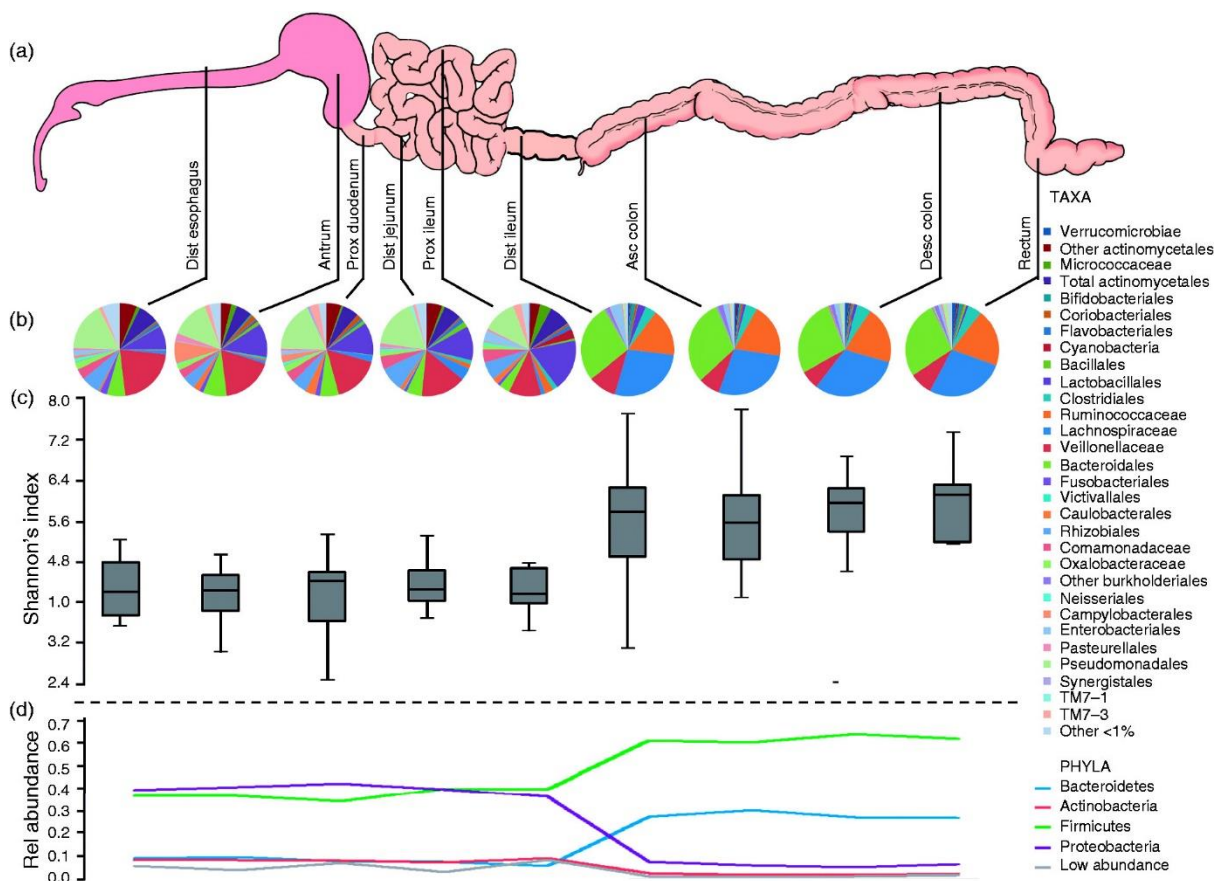


Figure 1.7. Overview of the adult microbiome in the GI tract. **A.** Schematic overview of the GI tract. **B.** Bacterial taxa present in the different sections of the GI tract. **C.** Diversity of bacterial taxa present per GI tract section represented by Shannon's index, which is a measure for both diversity and abundance. **D.** Relative abundance of major bacterial phyla in the GI tract. Reprinted from (Vuik et al., 2019), under the Creative Commons CC-BY-NC license.

Commensal bacteria in the gut help degrade complex carbohydrates found in the diet such as HMOs (in infants) or plant polysaccharides in adults due to the extensive repertoire of Carbohydrate-Active Enzymes (CAZymes) encoded by their genomes to break down these complex glycans. There are two types of CAZymes that cleave glycosidic bonds between carbohydrates or between a carbohydrate and a non-carbohydrate moiety: glycoside hydrolases (GHs), which cleave bonds by the insertion of a water molecule (hydrolysis), and polysaccharide lyases (PLs), which cleave complex carbohydrates using an elimination mechanism ([www.cazy.org](http://www.cazy.org)). Specialised bacteria can break down dietary (plant-based) carbohydrates such as mannan (Cuskin et al., 2015), fructans (Sonnenburg et al., 2010), xyloglucan (Larsbrink et al., 2014), xylan (Rogowski et al., 2015), raffinose (Anggraeni, 2022), or resistant starch (Ze et al., 2012).

Plant cell walls consist of highly complex carbohydrates including, cellulose, hemicellulose, and pectin, which reach the colon undigested and there they become substrates for the gut microbiota. Cellulose consists of  $\beta$ 1-4 linked glucose chains and is effectively broken down by gut microbiomes specialised in plant fibre digestion like in ruminants and termites (Brune & Friedrich, 2000; Oyeleke & Okusanmi, 2008). Cellulose only differs from starch by the linkage configuration between the glucose units, starch has  $\alpha$ 1-4 linked units. The ability to degrade cellulose has been found in the human gut (Chassard et al., 2011), but only at low levels. Hemicelluloses and pectins have more complex branched structures with variation between plants, consisting of different monosaccharide units (Koropatkin et al., 2012; Mohnen, 2008)(Koropatkin et al., 2012; Mohnen, 2008).

As a consequence, consuming different diets (plant-based or animal-based) can alter the gut microbiota composition, even in the short-term (David et al., 2014). The gut microbiota of humanised mice can rapidly change following a switch from a plant polysaccharide-rich and low-fat diet to a high-fat and high-sugar diet (Turnbaugh et al., 2009). This is reflected in metagenomics studies in human populations showing dietary changes have a profound effect on the human microbiota (Wilson et al., 2020), although intrinsic (e.g., age, sex, ancestry) and extrinsic (e.g., lifestyle, geography) factors can often confound the results (J. Li et al., 2022; Troci et al., 2022; Wilson et al., 2020).

The major end products of fermentation are the short-chain fatty acids (SCFAs), butyrate, acetate, and propionate, which provide approximately 10% of the calories we absorb (the

value varies depending on our diets) and are involved in numerous physiological processes (for a recent review see (Blaak et al., 2020)). For example, butyrate is a main energy source for colonocytes (Bourassa et al., 2016; Roediger, 1980) and has been associated with a reduced risk of CRC (Bourassa et al., 2016; McNabney & Henagan, 2017). In germ-free (GF) mice, the lack of butyrate production leads to autophagy in the colonocytes (Donohoe et al., 2011). Acetate is the most abundantly produced SCFA, and is used in lipid production (Den Besten et al., 2013). On the other hand, propionate has been shown to inhibit lipid production (Morrison & Preston, 2016). SCFAs produced by gut bacteria play an important role in the development of the immune system. Butyrate induces regulatory T-cell generation and differentiation (Arpaia et al., 2013; Furusawa et al., 2013). In addition to being the main energy source for colonocytes, SCFAs enhance barrier function (Mathewson et al., 2016; Suzuki et al., 2008). SCFAs from gut bacteria increase mucus production (Burger-van Paassen et al., 2009; Gaudier et al., 2004)

In addition to SCFAs, the gut microbiome helps in the production of vitamins. These vitamins are not only used by the host but are also essential for many gut microbes. Many B vitamins are produced by the gut microbiota, with the estimated maximum of the daily reference intake of B vitamins produced by bacteria being between 2.3% (B1) and 86% (B6) (Magnúsdóttir et al., 2015), however this did not take into account that some vitamins are only able to be absorbed at the ileum level, and thus any of those vitamins produced will not be taken up by the host (Barone et al., 2022). In addition to B vitamins, vitamin K is also produced by gut bacteria (Dai et al., 2023). The diet supplies the gut microbiota with essential vitamins, for example *A. muciniphila* and *bifidobacterium* species were promoted by vitamins C, B2, D and E *in vitro* (Pham et al., 2021).

The gut microbiota also plays an important role in the defence against pathogenic bacteria (Ducarmon et al., 2019; Endt et al., 2010). There are several ways gut commensal bacteria can reduce infection by pathogens, including competition for nutrients or binding sites (for recent reviews, see (H. Y. Cheng et al., 2019; Ducarmon et al., 2019)). The beneficial role of commensal bacteria is also shown by the fact that after an extensive antibiotic treatment, which kills the majority of commensals, the rate of intestinal infections increases (Buffie et al., 2012; Sekirov et al., 2008). Furthermore, faecal microbiota transplantation from a healthy donor is effective in alleviating recurring *Clostridium difficile* infections (Van Nood

et al., 2013). Some commensal bacteria have the ability to secrete antimicrobial compounds themselves which target specific pathogenic bacteria (Rea et al., 2010). For example, *Bacteroides thetaiotaomicron* can induce the secretion of microbicidal compounds secreted by Paneth cells, which are effective against certain pathogenic bacteria, while being less effective against *B. thetaiotaomicron* (Hooper et al., 2003).

### **1.2.2 Role of the gut microbiota in intestinal diseases**

Many GI diseases are associated with an alteration in the gut microbiota or 'dysbiosis', with examples provided below.

UC and CD are both chronic intestinal diseases that are part of IBD. Microbial communities in IBD patients are generally characterised by a decreased diversity and reduced abundance in several bacterial phyla (Andoh & Nishida, 2022; Lepage et al., 2011; Martinez et al., 2008; Sokol et al., 2006). For example, in CD patients Proteobacteria were found to be significantly increased in abundance, while the abundance of Firmicutes and Bacteroidetes were significantly reduced (Nishino et al., 2018). It does remain unclear whether the changes in microbial diversity and abundance are a consequence or the cause of the inflammation (Andoh & Nishida, 2022). A study analysing the results of 22 articles showed the families Enterobacteriaceae and Lactobacillaceae, and the genus *Bacteroides* were increased, while Clostridiales, and the genera *Faecalibacterium* and *Bifidobacterium* were decreased in irritable bowel syndrome (IBS) patients (Pittayanon et al., 2019).

Some studies have shown a link between the gut microbiome and CRC, especially in inflamed tissue (Arthur et al., 2012). *Fusobacterium* species have been shown in multiple studies to be correlated with colorectal tumours (Castellarin et al., 2012; Kostic et al., 2012; A. N. McCoy et al., 2013; T. C. Yu et al., 2017) whereas Bacteroidetes and Firmicutes seem to be lowered (Kostic et al., 2012). *Fusobacterium nucleatum* directly promotes colorectal tumour growth and metastasis (Rubinstein et al., 2013; Shang & Liu, 2018; Yin et al., 2022). Recently pks+ *E. coli*, which can produce the genotoxin colibactin, was shown to be directly linked to CRC development (Pleguezuelos-Manzano et al., 2020).

Several microbial species have been linked with obesity and other metabolic diseases, like *Bacteroides vulgatus* (Aron-Wisnewsky et al., 2019). A large metatranscriptome study showed an increased abundance of Firmicutes and a decreased abundance of

Bacteroidetes in obesity, while the secretome (all secreted proteins by the microbiome) revealed a significant increase in secreted CAZymes (Gallardo-Becerra et al., 2020). *A. muciniphila* was found in reduced abundance in obese patients (Yun et al., 2017).

### **1.2.3 The mucus-associated microbiota**

#### **1.2.3.1 Composition in health and disease**

In humans, the mucosa-associated microbiota appears to be different from the lumen microbiota and uniform throughout the colon (see for example Jiao et al., 2022; Zoetendal et al., 2002). However, due to the invasiveness of the procedure, studies profiling the mucus-associated microbiota in humans are limited and suffer from small sample size, confounding effects of colonic lavage, and variability in the anatomical site being sampled, the nature of the sample, and sampling method (Juge, 2022). As a result, knowledge on mucus-associated microbiota mainly comes from *in vitro* and mouse studies or from analysis of human samples from CRC or IBD (Juge, 2022). In the mucosa the most abundant phyla are Firmicutes, Bacteroidetes, and Proteobacteria (Jiao et al., 2022), and compared to the lumen there is an increase of Lachnospiraceae, Ruminococcaceae, and the mucin degrader *A. muciniphila* (Ouwerkerk et al., 2013). Mucus-associated bacteria such as *A. muciniphila* are able to increase mucus production, providing them with an energy source (Cani et al., 2022; Everard et al., 2013; Van Der Lugt et al., 2019), while protecting the host against over-fermentation and some pathogenic bacteria which are able to break down the mucus layer (Donaldson et al., 2015; Y. S. Kim & Ho, 2010). Some mucus degrading enzymes from pathogens actually trigger an increase in mucus production to 'wash out' these pathogens (Navarro-Garcia et al., 2010). In IBD, the mucosal microbiota diversity is decreased, with a reduction in Firmicutes and an increase in Bacteroidetes (Walker et al., 2011). Significant differences were also observed between the mucosa-associated microbiota in inflamed and non-inflamed colonic sites (Walker et al., 2011). Several mucosa-associated species such as *F. nucleatum* and *E. coli* are associated with CRC (W. Chen et al., 2012; Peng et al., 2020) while mucosa-associated *Faecalibacterium* species with anti-inflammatory properties are reduced in CRC patients (W. Chen et al., 2012; Sokol et al., 2008).



### 1.2.3.2. Factors of colonisation

The adaptation of gut bacteria to the mucus layer relies on several mechanisms (for a recent review, see (B. X. Wang et al., 2021)). As discussed above, the mucus layer acts as a physical barrier to prevent microbes from coming into contact with the gut epithelium, while at the same time mucins provide bacteria with epitopes which can be used for bacterial adhesion (J. Y. Huang et al., 2011; Tasteyre et al., 2001; B. X. Wang et al., 2021). Some commensal bacteria, like *B. fragilis*, have been shown to bind to immunoglobulin A (IgA) for mucosal colonisation (Donaldson et al., 2018). IgA secreted by the human host prevent pathogens from colonising the epithelial layer by forming bacterial clumps which can be removed more easily from the epithelial layer (Moor et al., 2017), but can also participate in biofilm formation of gut commensal bacteria (Huus et al., 2021). MyD88, which regulates IgA signalling, knock-out mice show an increase in IgA binding of several bacteria species, including *Ruminococcus gnavus* (Kubinak et al., 2015). IgA targets and regulates mucus-associated bacteria more than lumen associated bacteria (Kubinak et al., 2015). Interestingly, IgA has been shown to induce mucin associated genes pathways in *B. thetaiotaomicron*, aiding in colonisation and symbiosis with other commensal bacteria (Nakajima et al., 2018). *Lactobacilli* species use mucus-binding adhesins to facilitate colonising of the mucus layer (Boekhorst et al., 2006; Muscariello et al., 2020; Roos & Jonsson, 2002; von Ossowski et al., 2011). Recently, an adhesin in *R. gnavus* E1 was found to bind mucins and immunoglobulins (Maresca et al., 2021). One of the main factors promoting the association of gut bacteria with the mucus layer is their capacity to forage on mucin glycan chains as described below (for a review see, (Bell & Juge, 2021)).

### 1.2.4 Metabolism of host glycans by gut bacteria

The capacity of gut commensal bacteria to degrade HMOs or mucin glycans relies on the complement of GHs encoded in their genomes, as recently reviewed (Bell & Juge, 2021; Coker et al., 2021; Crost et al., 2016; Tailford, Crost, et al., 2015) (see Supplemental Table S1). Several mucin glycan foraging bacterial species have been characterised to date (Berkhout et al., 2022; Tailford, Crost, et al., 2015). These include for example: *R. gnavus* (Crost et al., 2013), *A. muciniphila* (Derrien et al., 2004), *B. thetaiotaomicron* (Martens et al., 2008), *Bacteroides caccae* (Desai et al., 2016), *B. fragilis* (Pudlo et al., 2015), *B. bifidum* (Turroni et al., 2010), and *Clostridium perfringens* (Low et al., 2021).

To cleave off terminal mucin glycans, bacteria rely on sialidases (GH33) and fucosidases (GH29 and GH95), releasing terminal Neu5Ac and Fuc, respectively. These enzymes have been functionally characterised in *R. gnavus* (Crost et al., 2013; H. Wu, Rebello, et al., 2021), *A. muciniphila* (Shuoker et al., 2023), *B. bifidum* strains (Katayama et al., 2005; Kiyohara et al., 2011), *C. perfringens* (Fan et al., 2016), and *Bacteroides* species such as *B. thetaiotaomicron* (Sakurama et al., 2012; J. Xu et al., 2003) or *B. fragilis* (Luijckx et al., 2020). *B. thetaiotaomicron* also encodes multiple *O*-glycans active sulfatases (Luis, Baslé, et al., 2022), as recently reviewed (Luis, Yates, et al., 2022). Following removal of the glycan epitopes, other GH families can break down the mucin glycan chains further as described below.

Galactosidases from the GH2, GH35, GH42, GH98, GH110 families have been shown to be active on mucin glycans. GH2 and GH35  $\beta$ -galactosidases in *A. muciniphila* ATCC BAA-835 were found to be active on defucosylated, desialylated, and desulphurylated mucin core structures, releasing terminal Gal (Kosciow & Deppenmeier, 2020). GH110 from *B. bifidum* JCM 1254 can release Gal from BgB epitopes, to produce BgH antigens (Wakinaka et al., 2013). GH98 from *C. perfringens* ATCC 10543 is able to release both BgA and BgB trisaccharide from glycans containing these terminal structures, with a preference for glycans with a core 2 structure over a core 1 structure (Anderson et al., 2005). Surprisingly, a recent study showed that GH16 enzymes from *B. thetaiotaomicron*, *B. fragilis*, *B. caccae*, and *A. muciniphila* were active on Gal $\beta$ 1-4GlcNAc linkages in poly-LacNAc chains (type 2) in mucins while the activity of GH16s had previously only been reported on plant polysaccharides (Crouch et al., 2020).

Other GHs such as those belonging to GH20, GH84, GH85, and GH89 families catalyse the release of terminal GlcNAc epitopes. A GH89 from *C. perfringens* strain 13 was shown to release GlcNAc from terminal GlcNAc $\alpha$ 1-4Gal in pig gastric mucin (PGM) (M. Fujita et al., 2011), and a similar activity was found in a GH89 from *B. bifidum* JCM 1254 (Shimada et al., 2015). A GH20 enzyme from *A. muciniphila* was shown to be active on desialylated *para*-nitrophenol (*p*NP) linked core 2 mucin structures, removing the beta-linked GlcNAc (M. Wang et al., 2018). A GH20 from *Prevotella* strain RS2 was found to cleave terminal 6-SO<sub>3</sub>-GlcNAc from sulphated mucin glycans (Rho et al., 2005).

GHs catalysing the release of GalNAc residues in mucins have been demonstrated in GH27, GH36, GH101, GH109, and GH129 families. A GH129 from *B. bifidum* JCM 1254 was shown to be active on the Tn antigen (starting GalNAc attached to serine or threonine), releasing GalNAc from the protein backbone (Kiyohara et al., 2012). GH101 from *B. longum* was found to be able to cleave Gal $\beta$ 1,3GalNAc (core 1) from serine or threonine (K. Fujita et al., 2005) and a similar GH101 from *C. perfringens* strain 13 was shown to release core 1 and core 2 from mucin pre-treated with other GHs (Ashida et al., 2008).

Additional GHs have been found to be active on structures present in mucin glycans but were not tested on mucin itself, like GH109 (Teze et al., 2020) and GH35 from *A. muciniphila* ATCC BAA-835 (Kosciow & Deppenmeier, 2019); GH2 and GH42 from *B. longum* subsp. *infantis* ATCC15697 (Yoshida et al., 2012). Screening of metagenomic DNA from the human gut microbiome also identified GH109 is active on blood group type A-antigen and a new subfamily of GH31 that specifically cleaves the initial  $\alpha$ -GalNAc from mucin-type O-glycans (Rahfeld et al., 2019).

No bacteria possess the full complement of GHs required to fully degrade the mucin glycan chains, and as a consequence mucin-derived oligosaccharides or monosaccharides are used through cross-feeding by other members of the gut microbiota or pathogens in the mucosal environment. For example, *A. muciniphila* enables cross-feeding of both *Roseburia hominis* and *Roseburia inulinivorans* (Pichler et al., 2020). *R. inulinivorans* sialic acid metabolism was highly upregulated during cross-feeding on mucin glycans, even though it does not encode a sialidase, suggesting that it benefits from *A. muciniphila*-released sialic acid. Additionally, *R. inulinivorans* encodes a GH98 enzyme able to release BgA and BgB epitopes from mucin glycans (Pichler et al., 2020). *B. breve* is unable to grow on mucins, however it is able to grow when in co-culture with *B. bifidum* (Bunesova et al., 2018) due to the production of sialidases from *B. bifidum*, while *B. breve* is not able to release sialic acid from mucin glycans (Nishiyama et al., 2018). In addition to the utilisation of released oligo- or monosaccharides, cross-feeding on mucin glycans of *A. muciniphila* and other species, like *Anaerostipes caccae*, shows further use of other metabolites as SCFAs and vitamin precursors (Chia et al., 2018; Shuoker et al., 2023).

Because mucin glycans share structural similarities with HMOs many bacteria that can (partially) degrade mucins can also degrade several HMOs. This is shown by the ability of

mucin degrading specialist *A. muciniphila* to grow on HMOs (Kostopoulos et al., 2020) while not able to grow on most dietary glycans (Derrien et al., 2004; Ottman et al., 2017). Some bacteria only have specific fucosidases or sialidases to cleave off fucose or Neu5Ac groups, while others might only be able to degrade the backbone of an HMO, therefore promoting a diversity of bacteria (Bode, 2015; Marcobal et al., 2010).

### **1.3 *Ruminococcus gnavus***

#### **1.3.1. Occurrence and role in health and disease**

*R. gnavus* is considered as a prevalent member of the 'normal' gut microbiota found in 90% of people (J. Qin et al., 2010). It is part of the Firmicutes phylum, class Clostridia and the family Lachnospiraceae. *R. gnavus* is an early coloniser of the infant gut (Nilsen et al., 2020; Sagheddu et al., 2016) but persists in adults where it belongs to 57 frequently detected species (J. Qin et al., 2010). Furthermore, *R. gnavus* shows a disproportionate representation in diseases such as IBD, malnutrition, brain disorders or infection with a positive or negative association depending on the condition (Blanton et al., 2016; Crost et al., 2023; Grabinger et al., 2019; Hall et al., 2017; Henke et al., 2019).

For example, *R. gnavus* was shown to be one of the species significantly reduced in children with attention-deficit/hyperactivity disorder (ADHD) in a case-control study (Wan et al., 2020). This alteration was associated with differences in the metabolic pathways of neurotransmitters (e.g. serotonin and dopamine) and may contribute to ADHD symptoms. Mice mono-colonised with *R. gnavus* ATCC 29149 showed an alteration in *O*-acetylated sialic acids and other metabolites associated with changes in brain function and behaviour (Coletto et al., 2022).

*R. gnavus* has been shown to be significantly increased in infants with allergies, and mice intragastrically infected with *R. gnavus* developed asthma (Chua et al., 2018).

Pangenomic studies identified an enrichment of specific *R. gnavus* clades in IBD patients (Hall et al., 2017; Lloyd-Price et al., 2019). However, no causal effect has been established. A recent study showed that some isolates of *R. gnavus* produce a capsular polysaccharide that promotes a tolerogenic immune response, whereas strains without a capsule elicit a strong proinflammatory response *in vitro* (Henke et al., 2021), likely through inflammatory

glucorhamnan (Henke et al., 2019). In addition, mouse studies showed that *R. gnavus* supplementation led to enhanced Reg3 $\gamma$  expression, an important antimicrobial peptide that helps maintain spatial segregation between the epithelium and the microbiota and promote gut homeostasis (Surana & Kasper, 2017) and that increasing *R. gnavus* led to a decreased severity of inflammation in mouse models of colitis (Grabinger et al., 2019). In contrast, studies have shown an association of *R. gnavus* in CD (Henke et al., 2019).

In germ-free mice monocolonised with *R. gnavus*, it was shown that *R. gnavus* was correlated with more severe IBS symptoms in mice (Zhai et al., 2023). In monocolonised mice with *R. gnavus*, increased IBS symptoms were observed, including diarrhoea (Zhai et al., 2023). Oral gavage of *R. gnavus* induced IBS symptoms in mice with normal microbiota (Zhai et al., 2023). *R. gnavus* was found to be correlated with Serotonin, a neurotransmitter, which can induce diarrhoea (Zhai et al., 2023).

Using gnotobiotic mouse models, it was shown that, *R. gnavus* could ameliorate growth and metabolic abnormalities when added to the microbiota from undernourished donors in recipient animals by modulating acylcarnitine metabolites, revealing *R. gnavus* as a potential therapeutic target (Blanton et al., 2016). The importance of *R. gnavus* on host metabolism was further demonstrated by metatranscriptomics in GF mice mono-colonised by different bacterial species (T. W. Hoffmann et al., 2016). A recent metagenomics study directly correlated *R. gnavus* with body fat percentage and associated *R. gnavus* with obesity (Grahnmemo et al., 2022).

Given the potential implication of *R. gnavus* in influencing human health and disease, there is therefore great interest in developing strategies to target and modulate this bacterial species in humans.

### **1.3.2 *R. gnavus* glycan utilisation**

*R. gnavus* strains encode a different set of GHs (Figure 1.8) (Croft et al., 2013, 2016, 2023) which have been shown to be involved in *R. gnavus* capacity to degrade and utilise both dietary and host -derived glycans such as mucins and HMOs.

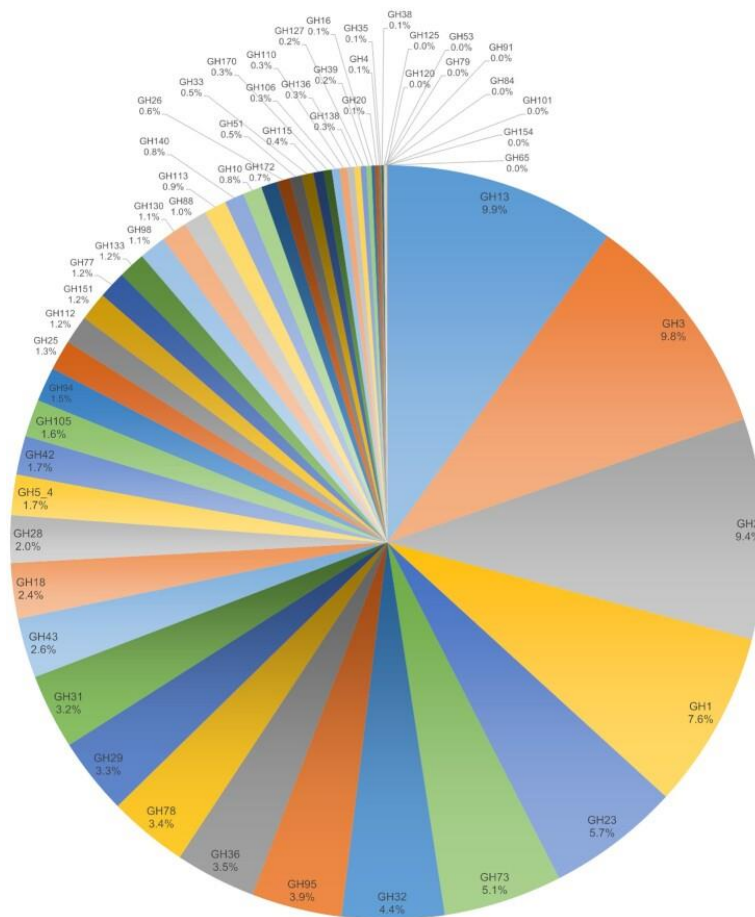


Figure 1.8. Occurrence of GH families across *R. gnavus* strains. Reprinted from (Crost et al., 2023) under CC BY 4.0 licence.

Plant  $\alpha$ -galactosides belonging to the raffinose family oligosaccharides (RFOs) and considered as prebiotics, are commonly degraded by  $\alpha$ -galactosidases produced by the human gut microbiome. *R. gnavus* E1 encodes GH36 enzymes, *RgAga1*, *RgAga2* and *RgAgaSK*, able to hydrolyse melibiose, raffinose, stachyose, and verbascose (Aguilera et al., 2012; Bruel et al., 2011; Cervera-Tison et al., 2012; Lafond et al., 2020). Interestingly, *AgaSK* also exhibits kinase activity on sucrose (Lafond et al., 2020).

Starch is another source of complex carbohydrate available to the gut microbiota (Tian & Sun, 2020). Although *R. gnavus* ATCC 29149 cannot grow on starch as sole carbon source *in vitro*, it benefits by cross-feeding from starch-degrading organisms such as *Ruminococcus bromii*. *R. bromii* L2-63 grown in co-culture with *R. gnavus* ATCC 29149 promoted *R. gnavus* growth by providing resistant starch digestion products, like glucose and maltose, which *R. gnavus* is able to utilise (Crost et al., 2018). When in co-culture on starch, acetate was the main SCFA produced and there was a significant increase in the induction of Trp biosynthesis genes and a decrease of vitamin B12-dependent methionine biosynthesis as

compared to the *R. bromii* mono-culture (Crost et al., 2018). These results also showed the importance of cross-feeding activities in modulating the microbial metabolite profile (and therefore of their nutritional niches *in vivo*).

Some *R. gnavus* strains are inhabiting the mucus layer overlying the GI tract (Bell et al., 2019). The ability of *R. gnavus* to grow on host glycans, such as mucin glycans and HMOs is strain-dependent (Crost et al., 2013, 2016). Both *R. gnavus* ATCC 29149 and E1 are able to grow on 2'-Fucosyllactose (2'FL) and 3-Fucosyllactose (3FL), only *R. gnavus* E1 is able to grow on lactose and LacNAc, while only *R. gnavus* ATCC 29149 is able to grow on 3'-sialyllactose (3'SL) (Crost et al., 2013). *R. gnavus* ATCC 29149 and E1 strains produce different fucosidases able to hydrolyse 2'FL, 3FL, Le<sup>a</sup>, and Le<sup>x</sup> structures. *R. gnavus* E1 fucosidase E1\_10125 could hydrolyse fucose from sialylated sLe<sup>a</sup> and sLe<sup>x</sup> (H. Wu, Rebello, et al., 2021). The ability of *R. gnavus* ATCC 29149 to utilise fucose from 3FL or 2'FL led to an increased propanol and propionate production (most likely via the propanediol pathway) (Crost et al., 2013).

*R. gnavus* strains ATCC 35913 and ATCC 29149 can grow on mucin as a sole carbon source (purified pig gastric mucin, pPGM). Transcriptomics analyses of *R. gnavus* ATCC 29149 and ATCC 35913 strains grown on pPGM showed upregulation of specific GH33 sialidase, and GH29 and GH95 fucosidases in the mucin utilisation pathway (Crost et al., 2013, 2016), suggesting that the strategy utilised by *R. gnavus* strains to grow on mucins is focused on the utilisation of mucin glycan epitopes.



Figure 1.9. Schematic representation of *R. gnavus* ATCC 29149 *nan* operon, *nanH* encodes an IT *trans*-sialidase (RGna\_RS08335 in new genome annotation). Taken from (Bell et al., 2019), with permission from Springer Nature.

When present in *R. gnavus* genomes (such as ATCC 29149 and ATCC 35913), the gene encoding GH33 is found part of a *nan* cluster which is dedicated to sialic acid metabolism (Figure 1.9) (Bell et al., 2019; Crost et al., 2013, 2016). The *R. gnavus* strain E1, which is not

able to grow on mucin, lacks the *nan* cluster (Croft et al., 2013). Biochemical and enzymatic characterisation of RgGH33 enzyme from *R. gnavus* ATCC 29149 revealed that the enzyme was able to cleave  $\alpha$ 2-3 linked sialic acid residues releasing 2,7-anhydro-Neu5Ac instead of Neu5Ac, classifying this enzyme as an intramolecular *trans*-sialidase (Tailford, Owen, et al., 2015). Interestingly *R. gnavus* ATCC 29149 and ATCC 35913 are not able to grow on Neu5Ac as sole carbon source but can grow on 2,7-anhydro-Neu5Ac (Croft et al., 2013, 2016). Further studies unravelled the metabolic pathway of 2,7-anhydro-Neu5Ac utilisation in *R. gnavus* ATCC 29149. The *nan* operon contains an ABC transporter (RUMGNA\_02698-RUMGNA\_02696/RGna\_RS08355-RGna\_RS08345). RgSBP (RUMGNA\_02698/RGna\_RS08355) transporter is specific for 2,7-anhydro-Neu5Ac and is not able to transport Neu5Ac. RgNanOx (RUMGNA\_02695/RGna\_RS08340) was found to be an oxidoreductase which is able to reversibly convert 2,7-anhydro-Neu5Ac into Neu5Ac. RgNanA (RUMGNA\_02692/RGna\_RS08325), a sialic acid aldolase, then produces ManNAc from Neu5Ac. NanK (RUMGNA\_02691/RGna\_RS08320) and NanE (RUMGNA\_02693/RGna\_RS08330) are then involved in further metabolism of ManNAc (Bell et al., 2019). YjhC, an *E. coli* oxidoreductase, was shown to have a similar activity to RgNanOx, converting 2,7-anhydro-Neu5Ac into Neu5Ac using the same mechanism (Bell et al., 2020).

The presence of this unique sialic acid (2,7-anhydro-Neu5Ac) metabolism pathway in mucus-colonising strains confers *R. gnavus* strains with a competitive advantage in the gut as shown in gnotobiotic mice where the *nan* mutant was outcompeted by the wild-type strain and less effective in colonising the mucus layer (Bell et al., 2019). Further *in silico* analysis of sialic acid gene clusters across bacterial genomes supports the specialisation of *R. gnavus* for 2,7-anhydro-Neu5Ac metabolism (Severi et al., 2021), suggesting that 2,7-anhydro-Neu5Ac metabolism is more widespread than just *R. gnavus*. This ability does help *R. gnavus* colonise its niche in the mucus layer and helps shape the human gut microbiota.



## Chapter 2

### Materials and Methods

## 2. Materials and Methods

This Materials and Methods section includes expert methodology carried out and written by others: bioinformatics of RNA sequencing (2.3.5) and 16S amplicon sequencing (2.6.3) by Dr Andrea Telatin; NMR analysis (2.5.7) by Dr Gwénaëlle Le Gall, and statistical analyses of qPCR growth data (2.8) by Dr George Savva.

### 2.1 Materials, strains, and media

#### 2.1.1 Materials

All chemicals were obtained from Merck (Darmstadt, Germany) unless otherwise stated. Purified PGM (pPGM) was prepared as described previously (Gunning et al., 2013). BgA trisaccharide (BgAtri) and BgA tetrasaccharide (BgAtetra) were purchased from Elicityl (Crolles, France). ManNAc was purchased from Carbosynth (Compton, United Kingdom). Pyruvate was purchased from Fluka (Munich, Germany). Cytidine 5'-triphosphate (CTP) was purchased from Chemily (Peachtree Corners, GA, USA).  $^{13}\text{C}_6$  ManNAc and Lactose-NCbz were kindly provided by Jun Jang Ong.  $\alpha$ -Amylase from *Bacillus subtilis* subsp. *subtilis* str. 168 was purchased from Prozomix (Northumberland, UK). Recombinant RgGH98 and RgGH33 from *R. gnavus* ATCC 29149 were produced in-house (Tailford, Owen, et al., 2015; H. Wu, Crost, et al., 2021). Peptidoglycan from *Streptomyces venezuelae* cell wall was a kind gift from Dr Susan Schlimpert (John Innes Centre, Norwich UK). Recombinant plasmids expressing RGna\_RS05890, RGna\_RS04290, RGna\_RS14395, and RGna\_RS14690 were a kind gift from Dr Haiyang Wu (Quadram Institute Bioscience, Norwich UK).

#### 2.1.2. Strains

*R. gnavus* strains used were ATCC 29149 (Moore & Holdeman, 1974) and E1 (Ramare et al., 1993), both isolated from healthy adults. *E. coli* strains used were DH5 $\alpha$  (Invitrogen) for cloning and site-directed mutagenesis and TunerDE3pLacI for heterologous expression.

#### 2.1.3. Media

*R. gnavus* strains were either grown in rich medium BHI-YH (Brain Heart Infusion - Yeast extract - Haemin; as described in Crost et al., 2013) or in minimum medium supplemented with carbohydrates. The two minimum media used were the semi defined medium YCFA (Yeast extract - Casitone - Fatty Acids) (Duncan et al., 2002) and the synthetic medium LAB (Tramontano et al., 2018).

BHI-YH (1 l) contained 37 g brain heart infusion, 5 g yeast extract and 10 ml of haemin 500 mg/l stock solution, in water.

YCFA (1 l) contained 10 g casitone, 2.5 g yeast extract, 4 g NaHCO<sub>3</sub>, 1 g L-cysteine hydrochloride, 0.45 g K<sub>2</sub>HPO<sub>4</sub>, 0.45 g KH<sub>2</sub>PO<sub>4</sub>, 0.9 g NaCl, 0.09 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.09 g CaCl<sub>2</sub>, 1 mg resazurine, 10 mg haemin, 10 µg biotin, 10 µg cobalamine, 30 µg p-4-aminobenzoic acid, 50 µg folic acid, 150 µg pyridoxamine dihydrochloride, 50 mg thiamine, 50 mg riboflavin, 1.889 ml acetic acid, 0.6714 ml propionic acid, 0.0927 ml isobutyric acid, 0.1104 ml isovaleric acid, and 0.1088 ml valeric acid, in water.

LAB medium (1 l) contained 3.1 g KH<sub>2</sub>PO<sub>4</sub>, 6.48 g K<sub>2</sub>HPO<sub>4</sub>, and 1.5 g tricine added to 600 ml water. 10 ml ascorbic acid 100x stock solution, 1 ml vitamins and antioxidant 1000x stock solution, 10 ml nucleotides 100x stock solution, 100 ml amino acid 10x stock solution, 1 ml salts and minerals 1 1000x solution, and 100 ml salts and minerals 2 10x solution were added separately (see Appendix 2 for the composition of stock solutions). The pH was checked and adjusted to 6.4 if necessary, and the solution was completed to 1 l with water. The solution was then filter-sterilised into a sterile bottle and stored in the dark at 4 °C.

*E. coli* strains were grown in LB media supplemented with 50 µg/ml kanamycin (final concentration) (see 2.4.2 and 2.4.3). LB (1l) contained 10 g tryptone, 5 g yeast extract and 10 g NaCl in 1 l water. For LB agar plates 1.5% (w/v) agar was added and subsequently autoclaved.

The growth of faecal samples was carried out in batch medium. One litre of batch medium contained 2 g peptone, 2 g yeast extract, 0.1 g NaCl, 0.04 g K<sub>2</sub>HPO<sub>4</sub>, 0.04 g KH<sub>2</sub>PO<sub>4</sub>, 0.01 g MgSO<sub>4</sub>·6H<sub>2</sub>O, 0.01 g CaCl<sub>2</sub>·2H<sub>2</sub>O, 2 g NaHCO<sub>3</sub>, 0.5 g cysteine HCl, 0.5 g bile salts, 2 ml Tween 80, 0.02 g haemin and 10 µl 0.5% (v/v) vitamin K, in water.

## **2.2 Bacterial growth assays**

### **2.2.1 *R. gnavus* growth assays on carbohydrate sources**

BHI-YH medium was pre-reduced by exposure to anaerobic conditions (85% N<sub>2</sub>, 10% H<sub>2</sub> and 5% CO<sub>2</sub>) in an anaerobic cabinet (Don Whitley, Shipley, UK) for 48 hours. Starter cultures were made by adding 2-4% (v/v) of *R. gnavus* glycerol stock maintained at -80 °C to 5 ml pre-reduced BHI-YH. The culture was grown in the anaerobic cabinet overnight at 37 °C and an aliquot of the starter culture (1 ml) was centrifuged at 3300 g for 5 min and the pellet resuspended in a volume of LAB medium to have an OD of 1.33. The inoculation of fresh

LAB medium was made with 1-2% of this cell suspension. The cultures were grown anaerobically at 37 °C in 14 ml tubes or in a 96-wells plate (200 µl/well). Optical density (OD) was measured in tubes with an Ultrospec 10 cell density meter from GE Healthcare (Chicago, IL, USA) at 600 nm or in plate using a plate reader (Infinite F50 plate reader from Tecan (Männedorf, Zürich, Switzerland) at 595 nm or Absorbance 96 from Byonoy at 600 nm (Hamburg, Germany)).

For growth assays on LAB medium supplemented with carbohydrates as sole carbon source, LAB was supplemented with Glc or Fuc at a final concentration of 0.5% and 0.46%, respectively, or with BgAtri and BgAtetra diluted at 1.5 mM or 1 mM in LAB. For growth assays with pPGM, a final concentration of 10 mg/ml pPGM was added to LAB medium without vitamins and autoclaved, then filter-sterilised Vitamins & Antioxidants 1000x solution and Ascorbic acid 100x solution (see Appendix 2) were added. For growth assays in the presence of *RgGH98*-treated pPGM, *RgGH98* enzyme was diluted with a small volume of LAB medium (1:1) and filter sterilised, then it was added to the LAB medium containing pPGM at a final concentration of 1.3 µM and incubated for 48 hours at 37 °C before conducting the growth assay. For growth assays on LAB medium supplemented with maltooligosaccharides or polysaccharides as sole carbon source, LAB was supplemented with 27.7 mM maltooligosaccharides or 0.5% polysaccharides.

For growth assays on *RgGH98*-treated pPGM mucin or oligosaccharide fractions, pPGM (10 g/l) was first treated with *RgGH98* (at 1.3 µM final concentration) in 0.1 M citrate buffer pH 6.0 for 72 hours at 37 °C. The sample was then heated at 50 °C for 10 min to inactivate the enzyme and dialysed using a 7KDa MWCO cut-off membrane in 2 l water for 3 hours followed by a change of water and a further 3-hour dialysis. The retentate, corresponding to the enzymatically treated pPGM, was freeze-dried. The dialysate was concentrated using a Büchi rotavapor R-210 (Flawil, Sankt Gallen, Switzerland), washed with 3x 1 ml MQ water and run through a graphitised carbon column for clean-up. First, the column was washed with 4 ml acetonitrile with 0.1% 2,2,2-trifluoroacetic acid (TFA), and equilibrated with 4x 1 ml H<sub>2</sub>O supplemented with 0.1% TFA, before the sample (3x 1ml) was loaded onto the column and successively eluted using 4 ml 0.1% TFA, 4 ml 10% acetonitrile with 0.1% TFA, 4 ml 25% acetonitrile with 0.1% TFA, and 4 ml 50% acetonitrile with 0.1% TFA. Fractions for each step were collected separately. The fractions were then analysed by HPAEC-PAD as

described in 2.5.5 and the fraction corresponding to the oligosaccharides was freeze-dried. The growth culture was supplemented with 1% *RgGH98*-treated pPGM, or the equivalent amount of the mucin fraction of BgAtri fraction.

For growth assays on mucin glycans,  $\beta$ -elimination of PGM or pPGM was performed as previously described (Pruss et al., 2021). Briefly, NaBH<sub>4</sub> was dissolved in 150 mM aqueous NaOH at 750 mM final concentration, and pPGM or PGM was added to this solution at a final concentration of 1%. Following incubation overnight at 48 °C (for at least 20 hours), the reaction was neutralised with HCl (10 M) until effervescence stopped. The sample was then centrifuged at 14,000 × g for 30 min at 4 °C and the supernatant filtered using 0.22  $\mu$ m filters (Thermo Fisher) and dialysed against dH<sub>2</sub>O with 1 kDa MWCO (Molecular Weight Cut-Off) membranes (Spectra/Por 7 Dialysis Tubing 1 kDa MWCO, Repligen, Waltham, MA, USA). Each 100 ml sample was dialysed in 5 l dH<sub>2</sub>O, which was refreshed 6 times with at least 7 hours in between each dialysis. The solution was then lyophilised. The mucin glycans were added to the media at a final concentration of 1%.

### **2.2.2 Faecal slurry growth assays**

Faecal samples were obtained from one human donor and resuspended in PBS with 10% glycerol. The donor was a participant recruited onto the QIB Colon Model study. The criteria for participants of the QIB Colon Model study state that they have to be healthy adults over 18 years old, living within 10 miles of the Norwich Research Park. Eligible participants displayed a normal bowel habit, regular defecation between three times a week and three times a day, with an average stool type of 3–5 on the Bristol Stool Chart, and no diagnosed chronic gastrointestinal health problems, such as irritable bowel syndrome, inflammatory bowel disease, or coeliac disease. Participants were also asked that they had not taken antibiotics or probiotics within the last four weeks, had not experienced gastrointestinal complaints, such as vomiting or diarrhoea, within the last 72 h, and had not recently returned from travel. Access to human faecal samples was granted by the NHS Health Research Authority (HRA) approval of The QIB Colon Model study (Full title: Stool Sample Collection For Research Using the QIB Colon Model; REC reference: 15/LO/2169. IRAS project ID: 171172).

For growth of faecal slurry on 2,7-anhydro-Neu5Ac or Neu5Ac, 10% faecal slurry was added to batch media with 2 mg/ml <sup>13</sup>C<sub>7</sub>-Neu5Ac or <sup>13</sup>C<sub>7</sub>-2,7-anhydro-Neu5Ac, and grown in

anaerobic conditions. After 8 hours RNA was extracted via the phenol-chloroform extraction protocol as described below (see 2.3.2).

## **2.3 Molecular biology**

### **2.3.1 DNA extraction, quantification and quality control**

*R. gnavus* genomic DNA extraction was done with the GeneJET Genomic DNA Purification Kit (Thermo Fischer Scientific, Waltham, MA, USA) using the Gram-Positive Bacteria Genomic DNA Purification Protocol or the New England Biolabs (NEB, Ipswich, MA, USA) Monarch DNA extraction kit following the manufacturer instructions with the following modifications. First bacteria were centrifuged at 5,000 g for 10 min and resuspended in 80 µl cold PBS by vortexing. Then 20 µl lysozyme solution (25 mg/ml) was added and briefly vortexed. Subsequently, 100 µl NEB Tissue Lysis Buffer was added, vortexed, and incubated at 37 °C for 10 min. Proteinase K (10 µl) (Thermo Fisher) was added, briefly vortexed and incubated at 56 °C for 30 min under agitation at 1400 rpm. RNase A (3 µl) (Thermo Fisher) was added, briefly vortexed and incubated at 56 °C for 5 min under agitation at 1400 rpm.

DNA concentration was measured Qubit 2.0 fluorometer (Thermo Fischer Scientific) and DNA quality assessed with NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fischer Scientific).

### **2.3.2. RNA extraction**

RNA from *R. gnavus* or faecal growth assays was extracted following the phenol-chloroform RNA extraction protocol: 0.4 ml phenol-ethanol solution (1:9 phenol pH 4.3-ethanol) was added to 2 ml bacterial culture and mixed by inverting. The tubes were placed on ice for 30 min and then centrifuged at 10,000 g for 5 min at 4 °C and supernatant removed. Cell pellets were then frozen at -80 °C, until all samples were collected and extraction was performed. Then the pellets were thawed on ice for 10 min and subsequently resuspended in 500 µl ice cold TES buffer (50 mM Tris, 5 mM EDTA, 50 mM NaCl, pH 7.5). This suspension was added to 2 ml tubes containing 1 ml acid washed silica beads, 600 µl acid phenol, and 100 µl chloroform, and then vortexed. The tubes were run for 3 x 40 s at speed setting 6 m/s in a MP Biomedical FastPrep-24™ Classic bead-beater, while chilling on ice for 5 min between runs. The tubes were then centrifuged at 10,000 g for 10 min at 4 °C. The top

phase (500 µl) was then added to a new tube containing 600 µl phenol and 100 µl chloroform, and then vortexed and centrifuged at 17,000 g for 5 min at 4 °C. The top phase (400 µl) was then added to a new tube containing 600 µl phenol and 100 µl chloroform, and then vortexed and centrifuged at 17,000 g for 5 min at 4 °C. The top phase (350 µl) was then added to a new tube containing 600 µl chloroform and centrifuged at 17,000 g for 5 min at 4 °C. The top phase (200 µl) was then added to new tubes containing 20 µl 3 M sodium acetate (pH 5.2) and 550 µl ice cold 95% ethanol. Following incubation at -80 °C for 20 min to precipitate the RNA, the tube was centrifuged at 17,000 g for 20 min at 4 °C and the supernatant discarded. The RNA pellet was washed with 600 µl 70% ethanol and centrifuged at 17,000 g for 10 min at 4 °C. Then the supernatant was discarded and the RNA pellet air dried. When dried, the pellet was resuspended in 200 µl RNase free water.

RNA was treated by using the TURBO DNA-free kit (Thermo Fischer Scientific). The concentration of RNA was determined before and after DNase treatment using NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fischer Scientific). Quantity and quality were then determined using Qubit 2.0 fluorometer (Thermo Fischer Scientific) and High-Sensitivity RNA ScreenTape on Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA, USA), respectively.

For RNA sequencing, 50 µl RNA of at least 50 ng/µl was sent to Novogene (Hong Kong, China) and the data analysed as described under section 2.3.5.

For RT-PCR analysis, up to 1 µg RNA was used following the method described under section 2.3.4.

### **2.3.3 qPCR analysis**

qPCR was run in StepOnePlus Real-Time PCR System (Thermo Fischer Scientific) with the following programme: 95 °C for 10 min, followed by 40 cycles of 95 °C for 10 s, 60 °C for 35 s, after which a denature step at 95 °C for 15 s was performed followed by complete annealing at 60 °C for 60 s, followed by a disassociation curve gradually increasing to 95 °C and holding there for 15 s, then followed by a final 60 °C for 15 s.

For qPCR to quantify bacterial growth, standard curves were prepared using 10-fold dilutions of DNA amplified from the targeted region ranging from  $10^2$  to  $10^7$  copies per 2 µl diluted in 5 ng/µl single stranded salmon sperm DNA and showed a linear relationship of

log copy number and the Ct value. Each well contained 2 µl DNA (1 ng/µl), 2.6 µl water, 0.2 µl of both forward and reverse primer (Supplemental Table S2), and 5 µl QuantiFast SYBR Green (Qiagen, Venlo, The Netherlands). If DNA concentration was below 1 ng/µl, more template DNA and less water was added to the well to achieve the same final DNA concentration. For the no-template control, 5 ng/µl single stranded salmon sperm DNA was used instead of DNA. Log copies per well (equivalent to Log copies per 2ng of DNA) were calculated using the standard curve. Copy number per ml of culture was calculated using Log copies per 2ng of DNA and the total DNA amount extracted from the cells from 1 ml of culture.

#### **2.3.4 RT-qPCR**

For RT-qPCR to determine gene transcription, standard curves were prepared using 2-fold dilutions of *R. gnavus* ATCC 29149 gDNA showing a linear relationship of log gDNA concentration and the Ct value. Reverse transcription was performed using QuantiTect Reverse Transcription (RT) kit (Qiagen) using up to 1 µg RNA extracted as described under 2.3.2. To check for gDNA contamination of RNA samples, RT negative controls were prepared by replacing reverse transcriptase with water. The qPCR reactions were performed using StepOnePlus Real-Time PCR System (Applied Biosystems Japan, Tokyo, Japan) using the same programme as described in 2.3.3. Each well contained 2 µl cDNA (1 ng/µl), 2.6 µl water, 0.2 µl of both forward and reverse primer (Supplemental Table 1), and 5 µl QuantiFast SYBR Green (Qiagen). The no-template controls used water instead of cDNA. Relative gene expression was calculated using the  $2^{-\Delta\Delta Ct}$  method, using the housekeeping gene *gyrB* for normalisation and growth on glucose as the control condition.

#### **2.3.5 Analysis of RNA sequencing data**

RNAseq data analysis was performed by Dr Andrea Telatin. Briefly, FASTQ files containing base calls and quality information for all reads that passed quality filtering were generated. Reads were mapped to the *R. gnavus* reference genome using TopHat2 (Pasakarnis et al., 2006). The mismatch parameter was set to two, and other parameters were set to default. Appropriate parameters were also set, such as the longest intron length. Filtered reads were used to analyse the mapping status of RNA-seq data to the reference genome. The HTSeq software was used to analyse the gene expression levels, using the union mode



(Anders & Huber, 2010). In order for the gene expression levels estimated from different genes and experiments to be comparable, the FPKM (Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced) was used to take into account the effects of both sequencing depth and gene length. The reads-to-mapping workflow was performed using nf-core/rnaseq (version 3.1) (P. A. Ewels et al., 2020), using the '`--aligner hisat`' mode, using the following tools: Fastp (S. Chen et al., 2018), SortMeRNA (Kopylova et al., 2012), Salmon (Patro et al., 2017), HiSat2 (D. Kim et al., 2015), Samtools (Danecek et al., 2021), Picard, and MultiQC (P. Ewels et al., 2016). The differential gene expression analysis was carried out using the DESeq2 package (Anders & Huber, 2010; Love et al., 2014) and the read counts from gene expression level analysis as input data. A list of differently expressed genes was generated using a Log2-fold change of 1 and -1 as threshold for the change in gene expression, and a p-adjusted value threshold of 0.05 for the significance.

## **2.4 Heterologous expression**

### **2.4.1. Cloning of target genes**

RGna\_RS14450, RGna\_RS07225, RGna\_RS14905, RGna\_RS06060, RGna\_RS08510, RGna\_RS10330, RGna\_RS07930, and RGna\_RS02745 were amplified from *R. gnavus* ATCC 29149 genomic DNA (obtained as described in 2.3.1) using Phusion high fidelity polymerase (New England Biolabs) and gene-specific primers (Supplemental Table 1). 10 µl 5x Phusion HF buffer, 1 µl dNTPs (200 µM final concentration) (Invitrogen, Waltham, MA, USA), 2.5 µl forward and reverse primers (0.5 µM final concentration), and 0.5 µl Phusion DNA polymerase, were added to 100 ng gDNA. The PCR reaction was carried out under the following conditions: 98 °C for 3 min, 34 cycles at 98 °C for 10 s, 58 °C for 30 s, and 72 °C for 2 min, followed by 72 °C for 10 min. The amplified DNA product was purified using the NEB PCR/DNA clean-up kit following the manufacturer's instructions.

Target genes were cloned into pET-28a plasmids, in frame with an N-terminal HIS-tag using NdeI/NheI-HF and XhoI (New England Biolabs) restriction enzymes and subsequent ligation using T4 DNA Ligase (New England Biolabs). Briefly, 1 µg amplified DNA was incubated with 5 µl 10X rCutSmart Buffer (New England Biolabs), 1 µl of each restriction enzyme was added and filled with nuclease free water to 50 µl for 60 min at 37 °C. After digestion the DNA was purified again using the NEB PCR/DNA clean-up kit and quantified by NanoDrop 2000 UV-

Vis Spectrophotometer. Ligation was performed by incubating digested plasmid DNA and digested inserts in a 1:3 mass ratio with 2  $\mu$ l T4 DNA Ligase Buffer (10X), 1  $\mu$ l T4 DNA Ligase and filled with nuclease free water to 20  $\mu$ l.

#### **2.4.2. Transformation**

For transformation of ligation reactions into *E. coli*, 100 ng DNA was added to 50  $\mu$ l competent DH5 $\alpha$  cells (Invitrogen), incubated on ice for 30 min and heat shocked at 42  $^{\circ}$ C for 45 s. Following incubation on ice for 2-3 min, 300  $\mu$ l SOC medium was added, and the suspension was incubated for 1 hour at 37  $^{\circ}$ C while shaking at 180 rpm. One litre of SOC medium contained 20 g tryptone, 5 g yeast extract, 0.58 g NaCl, 0.19 g KCl, 2.03 g MgCl<sub>2</sub>, 2.46 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 3.6 g glucose, adjusted at pH 7. Cells were then plated on LB supplemented with 30  $\mu$ g/ml kanamycin and incubated overnight at 37  $^{\circ}$ C. Plasmids were extracted and purified using the NEB plasmid miniprep kit. Successful transformations were checked by analysing uncut plasmid and plasmid digested with NdeI/NheI-HF and XhoI in order to remove the insert and analysed by electrophoresis on 1% agarose gel. The integrity of the recombinant plasmids was further determined by DNA sequencing performed in-house by David Baker and Rhiannon Evans on an Illumina Nextseq500 instrument.

#### **2.4.3 Production and purification of recombinant proteins**

Recombinant plasmids were transformed into *E. coli* TunerDE3pLacI cells as described above. TunerDE3pLacI cells were grown overnight in 10 ml LB supplemented with 30  $\mu$ g/ml kanamycin at 37  $^{\circ}$ C shaking at 180 rpm. This culture was added to 1 l LB supplemented with 30  $\mu$ g/ml kanamycin and incubated at 37  $^{\circ}$ C shaking at 180 rpm for 2-3 hours. It was then cooled to room temperature and 200  $\mu$ l 1 M isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) was added, and subsequently incubated at 16  $^{\circ}$ C shaking at 180 rpm overnight. The culture was centrifuged at 1500 g for 35 min at 4  $^{\circ}$ C. The supernatant was discarded and the pellet was resuspended in 15 ml binding buffer (20 mM Tris-HCl, 500 mM NaCl, pH 7.9). Cells were put on ice and lysed using sonification for 3 min at 30% amplitude in 30 s bursts and 59.9 s cooling and subsequently centrifuged at 17,000 g for 20 min at 4  $^{\circ}$ C.

The purification of the recombinant proteins was carried out using ÄKTA Pure (GE Healthcare, Little Chalfont, UK). The supernatant was first loaded onto a 5 ml HiTrap Fast Flow Crude column (Cytiva, Marlborough, MA, USA), equilibrated with binding buffer for immobilised metal affinity chromatography (IMAC). Target proteins were eluted using

elution buffer (20 mM TRIS-HCl, 500 mM NaCl, 150 mM imidazole, pH 7.9). Subsequently proteins were separated into fractions using a gel filtration column (Hi Load 16/600 Superdex 200 pg (Cytiva)) with gel filtration buffer (20 mM TRIS-HCl, 150 mM NaCl, pH 7.9). Protein purity in ÄKTA fractions were analysed using a NuPAGE Novex 4-12 % Bis-Tris gel (Invitrogen). Protein fractions were pooled and concentrated using 10-kDa MWCO Vivaspinn column (Sartorius, Göttingen, Germany).

## 2.5 Enzymatic assays

### 2.5.1 Enzymatic reactions

The recombinant enzymes were tested against putative substrates or mucin glycans using the reaction parameters in Table 2.1 for 24 hours at 37 °C. Reactions were performed in 50 mM citrate buffer at pH 6. The products of these reactions were analysed by HPAEC-PAD (see 2.5.5), if mono- or oligosaccharides were used as substrate or thin-layer chromatography (TLC) if the substrate was a polysaccharide (see 2.5.4). Quantification of released fucose was performed by HPAEC-PAD or k-fucose kit (Megazyme, Bray, Ireland) according to manufacturer's instructions.

Table 2.1. Conditions of activity assays.

	Substrate-enzyme combination (µl)	Substrate only control (µl)	Enzyme only control (µl)
500 mM Citrate buffer pH 6 (10x)	15	15	15
Substrate 1 mM or 10mg/ml	15	15	
Enzyme 100 µM	15		15
mQ water	105	120	120
Total:	150	150	150

0.1 mM substrate (final concentration) was used, except for large molar mass substrates (starch, pullulan, mucin glycans) where 1 mg/ml (final concentration) was used.

### 2.5.2. Kinetics

*RgGH13-2* enzyme kinetics were determined using the copper-bicinchoninate reducing sugar assay as previously described (Doner & Irwin, 1992). Enzyme reactions were performed as described in 2.5.1. Pullulan was used as the substrate with concentrations ranging from 0.05 mg/ml to 1 mg/ml and was incubated with 0.1 µM recombinant *RgGH13-2* taking 40 µl samples at 0, 2.5, 5, 10, 20, 30, and 45 min. These samples were instantly quenched in 160 µl stopping solution. When all timepoints were taken all samples were heated at 80 °C for 30 min and then cooled to 4 °C for 15 min and subsequently spun down

at 4.000 g for 5 min at 4 °C. Then 150 µl per sample was transferred to a flatbottom microtiter plate and OD<sub>530</sub> was measured. Stop solution: 1:1:1.2 mix of Solution A (5.427 g anhydrous Na<sub>2</sub>CO<sub>3</sub> and 2.42 g NaHCO<sub>3</sub> in 90 ml water was made, to this, 194.2 mg disodium-2,2'-bicinchoninate hydrate was added, and the final volume was adjusted to 100 ml with water), Solution B (79.8 mg CuSO<sub>4</sub>·H<sub>2</sub>O and 126.1 mg L-serine were dissolved in 90 ml water and then adjusted to a final volume of 100 ml with water). A standard curve of maltotriose using concentrations ranging from 4 mM to 0.02 mM in the same conditions as the reaction was used to calculate the concentration of released maltotriose per sample. By plotting the maltotriose concentrations over time the rate of maltotriose release was calculated. Michaelis-Menten kinetic constants were derived using non-linear regression analysis of *RgGH13-2* enzymatic rate on pullulan with a substrate-inhibition model  $v = V_{max}/(1 + K_m/[S] + [S]/K_i)$  for best fit using excel solver

### 2.5.3 Enzymatic synthesis of 2,7-anhydro-Neu5Ac

Neu5Ac was first synthesised from ManNAc and pyruvate using 0.25 mmol ManNAc and 0.05 mmol pyruvate in 300 µl 0.05 M potassium phosphate buffer supplemented with 6 µl 1% NaN<sub>3</sub>. The pH was adjusted to 7.2 and 0.66 mg (from a 10 mg/ml solution) sialic acid aldolase (Carbosynth) was added. This reaction was incubated at 37 °C for 72 hours while shaking at 160 rpm. The reaction was followed using TLC. At the end of the reaction, an equal portion of ice cold 96% EtOH was added to the suspension. After 15 min incubation at -20 °C, the suspension was centrifuged at 1900 g for 5 min. The supernatant was collected and the pellet was quenched once more with a 1:1 mixture of water and ice-cold ethanol, spun down and both supernatants were combined, evaporated and freeze dried. An anion-exchange column (Dowex 1x8 (Formate)) was then used to separate the ManNAc from Neu5Ac. The column was first flushed with 2 M formic acid twice and then with ddH<sub>2</sub>O until the outflow was ~pH 6. The sample was loaded onto the column and water was added until all ManNAc was eluted, as confirmed by TLC. Neu5Ac was then eluted using 2 M formic acid. This process was exactly repeated for the synthesis of <sup>13</sup>C<sub>7</sub> Neu5Ac from <sup>13</sup>C<sub>6</sub> ManNAc and 2-<sup>13</sup>C pyruvate.

To produce 3'SL from Neu5Ac, 0.049 mmol Neu5Ac, 0.053 mmol lactose benzyl chloroformate (Lac-NHCbz), and 0.058 mmol cytidine triphosphate (CTP) were added in

1.25 ml 100 mM Tris-HCl buffer containing 20 mM MgCl<sub>2</sub> at 37 °C. The pH was adjusted to 8.5. 70 µl of *Neisseria meningitidis* Cytidine monophosphate sialic acid synthetase (*NmCSS*) (Martin et al., 1998; H. Yu et al., 2004), 20 µl 2mg/ml *Pasteurella multocida* sialyltransferase 3 (*PmST3*) (H. Yu et al., 2005) were added. This reaction was incubated at 37 °C for 72 hours while shaking at 160 rpm. The reaction was monitored using TLC. After completion, the reaction was quenched with an equal portion of ice cold 96% EtOH. Following 15 min incubation at -20 °C, the suspension was centrifuged at 1900 g for 5 min. The supernatant was collected, and the pellet was quenched once more with a 1:1 mixture of water and ice-cold ethanol, spun down and both supernatants were combined, evaporated and freeze dried. The sample was loaded onto a P2 size exclusion column containing Bio-Gel P2 resin (Bio-Rad, Hercules, CA, USA). Fractions were confirmed by TLC. If needed, the sample was further purified using a silica column made of silica gel G60 (Silicycle 60 – 200 µm, 60 Å). Fractions and purity were also confirmed by electrospray ionisation-mass spectrometry (ESI-MS). This process was repeated for the synthesis of <sup>13</sup>C<sub>7</sub> 3'SL from <sup>13</sup>C<sub>7</sub> labelled Neu5Ac.

To produce 2,7-anhydro-Neu5Ac from 3'SL, 2.2 µl 0.3 mg/ml *RgNanH* in 20 mM Tris-HCl with 150 mM NaCl of pH 7.9 was added to 0.25 mg 3'SL in 97.8 µl H<sub>2</sub>O. The reaction was incubated at 37 °C at 160 rpm for 5 min. The reaction was monitored by TLC stained with a p-anisaldehyde sugar stain. Fractions and purity were also confirmed by ESI-MS. After completion, the reaction was quenched with an equal portion of ice-cold 96% EtOH. Following 15 min incubation at -20 °C, the suspension was centrifuged at 1900 g for 5 min. The supernatant was collected, and the pellet was quenched once more with 1:1 mixture of water and ice-cold ethanol, spun down and both supernatants were combined, evaporated and freeze dried. The sample was loaded onto a P2 size exclusion column containing Bio-Gel P2 resin (Bio-Rad). Fractions were confirmed by TLC. If needed, the sample was further purified using a silica column made of silica gel G60 (Silicycle 60 – 200 µm, 60 Å). Fractions and purity were also confirmed by ESI-MS (negative mode). This process was repeated for the synthesis of <sup>13</sup>C<sub>7</sub> 2,7-anhydro-Neu5Ac from <sup>13</sup>C<sub>7</sub> labelled 3'SL.

Table 2.2. Reaction overview of ManNAc + Pyruvate → Neu5Ac.

Chemical	MW	Mass (mg)	mmol	equivalence
<sup>13</sup> C <sub>6</sub> ManNAc	221.21 (227.14)	55 (57)	0.25	5
<sup>13</sup> C Pyruvate	110 (111.04)	5.5 (5.5)	0.05	1
<sup>13</sup> C <sub>7</sub> Neu5Ac	309.27 (316.18)	15.5 (15.5)	0.05	1
Recovered <sup>13</sup> C <sub>6</sub> ManNAc	221.21 (227.14)	45.5 (44)	0.2	4

The molar weight of <sup>13</sup>C compounds is given in brackets. Mass used in <sup>13</sup>C reaction is given in brackets.

Table 2.3. Reaction overview of Neu5Ac + Lactose-Ncbz + CTP → 3'SL-Ncbz.

Chemical	MW	Mass (mg)	mmol	equivalence
<sup>13</sup> C <sub>7</sub> Neu5Ac	309.27 (316.18)	15.2 (15.5)	0.049	0.95
LacNcbz	475.45	25.2	0.053	1
CTP	527.12	30.6	0.058	1.1
<sup>13</sup> C <sub>7</sub> 3'SLNcbz	765.3 (773.62)	37.5 (37.9)	0.049	0.95

The molar weight of <sup>13</sup>C compounds is given in brackets. Mass used in <sup>13</sup>C reaction is given in brackets.

#### 2.5.4 Thin-layer chromatography (TLC) analysis

TLC was carried out by spotting 4 µl reaction onto a TLC silica gel 60 F254 plate (Merck). The plates were developed in a solvent solution of isopropanol, 25% ammonium hydroxide solution, and water in a ratio of 6:3:1 (by volume). Carbohydrates were stained using an orcinol solution (150 ml EtOH, 10 ml H<sub>2</sub>O and 360 mg orcinol, the solution was cooled on ice and then slowly 20 ml concentrated sulphuric acid was added), and coloured spots were developed by heating of the TLC plate with a heat gun.

TLC was also used to monitor the enzymatic synthesis reactions to produce Neu5Ac, 3'SL and 2,7-anhydro-Neu5Ac. For this purpose, the TLC plate to monitor the reaction to produce Neu5Ac was developed in a solvent solution of 96% ethanol (EtOH) containing dichloromethane (DCM), 1 M ammonium bicarbonate ((NH<sub>4</sub>)HCO<sub>3</sub>) and H<sub>2</sub>O in a ratio of 6:3:0.5:0.5 (by volume). The spots were revealed with a molybdenum stain (made of 235 ml H<sub>2</sub>O, 12 g ammonium molybdate, 0.5 g ceric ammonium molybdate, and 15 ml concentrated sulphuric acid). The TLC to monitor the reaction producing 3'SL or 2,7-anhydro-Neu5Ac was developed using a solvent solution of ethyl acetate (EtOAc), methanol (MeOH), H<sub>2</sub>O at a 5:2:1 ratio (by volume). The spots were revealed with a *p*-anisaldehyde stain (made of 135 ml absolute EtOH, 5 ml concentrated sulphuric acid, 1.5 ml glacial acetic acid, and 3.7 ml *p*-Anisaldehyde, the solution was then stirred vigorously to ensure homogeneity) The TLC was developed by heating of the TLC plate with a heat gun.

### **2.5.5 High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD) analysis**

To prepare samples for HPAEC-PAD analysis, samples were centrifuged at 17,000 g for 10 min at room temperature, and supernatant was boiled for 10 min at 95 °C. The samples were then analysed on a Dionex ICS-5000 (Thermo Scientific, Hemel Hempstead, UK) using a Carbo-Pac PA1 column (Thermo Scientific). Samples were run according to the following gradient: 0 min, 18 mM NaOH; 20-35 min, 100 mM NaOH; 35.1-50 min, 18 mM NaOH. Data were analysed using the Chromeleon 6.80SR13 software.

### **2.5.6 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-ToF-MS) analysis**

For mucin glycan analysis, glycans were first released from pPGM by reductive  $\beta$ -elimination. Briefly, 1 g pPGM was dissolved in 100 ml 0.5 M NaBH<sub>4</sub> and 50 mM NaOH and then incubated at 45 °C for 16 hours and subsequently neutralised using drops of 10% acetic acid until effervescence stopped. The sample was then desalted by cation exchange chromatography using DOWEX 50W x8 H<sup>+</sup> resin. Samples were then freeze dried, and borates were removed by co-evaporation with methanol under a nitrogen stream. When dried, the glycans were permethylated by adding 200  $\mu$ l Dimethyl sulfoxide (DMSO), 300  $\mu$ l NaOH in DMSO (prepared as described in Shajahan et al., 2019), and 150  $\mu$ l iodomethane while vigorously shaken for 30 min at room temperature. The reactions were quenched by addition of 1 ml H<sub>2</sub>O. Excess of iodomethane was removed under nitrogen and the samples were loaded onto Supel Swift HLB SPE cartridges. Permethylated glycans were eluted in methanol, following washing of the cartridge with H<sub>2</sub>O. Samples were then dried and analysed using MALDI-ToF MS on a Bruker Autoflex (Bruker Daltonics, Billerica, MA, United States). 2,5-Dihydroxybenzoic acid was used as the matrix at 10 mg/ml final concentration.

### **2.5.7 NMR analysis**

NMR analysis was performed by Dr Gwénaëlle Le Gall. An aliquot of the enzymatic reaction was evaporated to dryness and reconstituted in 600  $\mu$ l of NMR buffer (100 ml D<sub>2</sub>O containing 0.26 g NaH<sub>2</sub>PO<sub>4</sub>, 1.41 g K<sub>2</sub>HPO<sub>4</sub>, and 1 mM deuterated trimethyl silylpropionate (TSP) as a reference compound) before <sup>1</sup>H NMR spectroscopic analysis. <sup>1</sup>H NMR spectra were recorded using a 600-MHz Bruker Avance spectrometer fitted with a 5-mm TCI proton-optimised triple resonance NMR inverse cryoprobe and autosampler (Bruker).

Sample temperature was controlled at 300 K. Spectra were acquired with 64 scans, a spectral width of 12500 Hz and an acquisition time of 2.6s. The “noesypr1d” presaturation sequence was used to suppress the residual water signal with a low-power selective irradiation at the water frequency during the recycle delay. Spectra were then transformed with a 0.3-Hz line broadening and zero filling, manually phased, baseline corrected, and referenced by setting the TSP-d4 signal to 0 ppm. Metabolites (amino acids and organic acids) were identified using the software Chenomx (V8.6).

### **2.5.8 Electrospray ionization mass spectrometry (ESI-MS) analysis**

Reactions to produce 2,7-anhydro-Neu5Ac and Neu5Ac were analysed by ESI-MS by diluting 1 µl samples in 50 µl or 250 µl of acetonitrile in H<sub>2</sub>O (1:1), then between 0.1 to 1.0 µl was injected into a Shimadzu (Kyoto, Japan) ESI LC-MS QP8000, depending on the intensity of ions. Sialic acid derivative peaks were detected using ESI negative mode.

## **2.6 Stable isotope probing (SIP)**

### **2.6.1. Preparation of samples**

Growth of faecal slurry and subsequent RNA extraction was performed as described above (2.2.2 and 2.3.2).

### **2.6.2. Gradient generation, and fractionation by ultracentrifugation**

The CsTFA gradient solution was generated by adding 4.8 ml caesium trifluoroacetate (CsTFA) was added to 1 ml gradient buffer (100 mM Tris, 100 mM KCl, 1 mM EDTA, pH 8.0). 20 µl was taken to measure the refractive index using a Reichert AR200 Full-Range Digital Refractometer (Buffalo, New York). 1257.9 µl (3.5% final volume) formamide was added to the CsTFA gradient solution. Gradient buffer was added to the CsTFA gradient solution with formamide until a refractive index of  $1.3725 \pm 0.0002$  was achieved. The mixture was added to centrifuge cartridges and for each sample exactly 500 ng RNA was added to a cartridge. The cartridges were centrifuged at 177,000 g (Beckman Vti 65.2 rotor) for 64 hours without braking. The gradients were subsequently fractionated by pumping the exact volume of the cartridge over 12 min into the top of each cartridge using an automatic peristaltic pump. A needle was used to pierce the bottom of each cartridge and fractions were collected for 1



min each. The refractive index of each fraction was measured. In each fraction RNA was then precipitated by adding 40 µl sodium acetate (3M, pH 5.2), 1 µl glycogen, and 1 ml 100% EtOH, and the solution stored at -20 °C for at least 2 hours. These samples were then centrifuged at max speed ( $\geq 20,000$  g) for 30 min at 4 °C. The precipitated RNA was washed with 500 µl 70% EtOH and centrifuged at 13,000 g for 10 min at 4 °C twice.

The RNA was subsequently airdried and then resuspended in 5 µl RNase free water and frozen at -80 °C. RNA was converted into cDNA using QuantiTect Reverse Transcription kit. Total bacterial 16S copies were analysed by qPCR (see 2.3.3) and sent for Illumina 16S amplicon sequencing to MRDNA labs (Shallowater, TX, USA).

### **2.6.3 Analysis of 16S amplicon sequencing data**

This analysis was performed by Dr Andrea Telatin. The quality profile of the raw reads (in FASTQ format) was assessed using Fastp 0.20.0 (S. Chen et al., 2018), which was also used to remove reads containing ambiguous bases, while the qualified region for DADA2 was determined using seqfu qual from SeqFu 1.17 (Telatin et al., 2021) with default parameters. The nf-core/ampliseq 2.5 pipeline (P. A. Ewels et al., 2020; Straub et al., 2020) was executed setting the locus specific primers to (forward: CCTACGGGNGGCWGCAG, reverse: GGACTACHVGGGTATCTAATCC for 16S rDNA). The workflow uses DADA2 (Callahan et al., 2016) to identify the Amplicon Sequence Variants (ASVs) using its Qiime2 2022.11.1 (Bolyen et al., 2019) wrapper. The taxonomic assignment was performed using against the SILVA database release 138 (Quast et al., 2013) using the DECIPHER R package (Wright, 2016). Data normalization and diversity were produced using the Rhea scripts (Lagkouvardos et al., 2017), and the final tables exported to be further analysed and plotted using MicrobiomeAnalyst (Dhariwal et al., 2017), and the built-in plotting provided by Dadaist2 (Ansorge et al., 2021).

## **2.7 Bioinformatics**

### **2.7.1 Prediction of protein function and subcellular localisation**

Protein function was predicted by BLAST (Altschul et al., 1990) analysis, amino acid sequences were blasted against the fully annotated Swiss-Prot protein database, using the Uniprot standard settings, and results used to predict GH function. Subcellular localisation

was determined using SignalP 5.0 (Almagro Armenteros et al., 2019) and Psort 3.0 (Yu et al., 2010). Transmembrane domains were predicted using TMHMM 2.0 (Krogh et al., 2001).

### **2.7.2 Sequence alignment and AlphaFold modelling**

All sequence alignments were carried out using Clustal Omega (Madeira et al., 2022). AlphaFold 2 modelling was performed using ColabFold v1.5.2 (Mirdita et al., 2022) using the standard settings: `msa_mode: mmseqs2_uniref_env`, `pair_mode: unpaired_paired`, `model_type: auto`, `num_recycles: auto`, `recycle_early_stop_tolerance: auto`.

## **2.8 Statistical analysis**

Statistical analyses were performed by Dr George Savva. For the analysis of qPCR data of *R. gnavus* ATCC 29149 and E1 grown on with 1% pPGM pre-treated with RgGH98, growth was analysed using a linear mixed model. Three technical replicates (measurements) were available at each time point. Estimated concentrations were transformed onto a logarithmic scale, then the concentration was modelled using the interaction of time, strain, and enzyme and all of their 2- and 3-way interactions as predictors, with nested random effects corresponding to biological replicate and time within biological replicate. Models were estimated using lmerTest/lme4 for R as above, and the effects of enzyme on growth between 0 and 9 and between 0 and 48 h for each strain was calculated from each model using emmeans (Bates et al., 2015; Kuznetsova et al., 2017; Lenth et al., 2022).

## Chapter 3

### Role of RgGH98 in growth of *R. gnavus* strains on mucin glycans

### 3. Role of *RgGH98* in growth of *R. gnavus* strains on mucin glycans

Some of the work described in this chapter has been published elsewhere (H. Wu, Crost, et al., 2021) and appears with permission of the authors. All work described in Chapter 3 has been performed by Wouter van Bakel, with the exception of 3.2.1, which was performed by Dr. Haiyang Wu (Quadram Institute Bioscience, Norwich, UK) and Ana Martínez Gascueña (Quadram Institute Bioscience)

#### 3.1 Introduction

The ability of *R. gnavus* to utilise mucin glycans is strain-dependent (Crost et al., 2013, 2016). Despite *R. gnavus* proficiency at using host-derived oligosaccharides, galactose (Gal), fucose (Fuc), *N*-acetylglucosamine (GlcNAc), 2'-fucosyllactose, (2'FL), 3-fucosyllactose (3FL) as carbon sources, only some *R. gnavus* strains (e.g. ATCC 29149 and ATCC 35913) but not E1, can utilise mucin glycans as metabolic substrates (Crost et al., 2013, 2016).

Both *R. gnavus* ATCC 29149 and E1 strains contain several fucosidase genes encoding GH29 and GH95 enzymes, five and eight fucosidases, respectively. Interestingly, even though *R. gnavus* E1 cannot grow on mucin as a sole carbon source, it produces fucosidases which are active on Lewis A and Lewis X epitopes (H. Wu, Rebello, et al., 2021), common mucin epitopes (De Bolós et al., 1995). In addition to mucins, blood group epitopes are also commonly found in human milk oligosaccharides (HMOs) (Garcher et al., 1994; Ravn & Dabelsteen, 2000; Thurl et al., 1997; Yan et al., 2019). An undefined galactosidase from *R. gnavus* ATCC 35913 was found to be active on blood group B, cleaving of the terminal galactose (Hata & Smith, 2004). The genes encoding GH33 and GH98 enzymes are exclusively present in *R. gnavus* ATCC 29149 and ATCC 35913 and not in E1 (Crost et al., 2013, 2016). Previous transcriptomic analyses of *R. gnavus* ATCC 29149 grown on purified pig gastric mucin (pPGM) showed upregulation of the *nan* cluster, a cluster dedicated to sialic acid metabolism (Crost et al., 2013). The *nan* cluster contains the gene coding for a GH33 intramolecular *trans*-sialidase, *RgNanH*, which is needed for *R. gnavus* ATCC 29149 growth on mucin and sialoglycans (Bell et al., 2019; Tailford, Owen, et al., 2015).

Previous, bioinformatics analysis of the *R. gnavus* ATCC 29149 genome revealed that the gene encoding *RgGH98* is part of a cluster containing thirteen genes with *RGna\_RS103115* to *RGna\_RS10360* being part of the same operonic structure (H. Wu, Crost, et al., 2021) (for the full list of genes identified in the cluster, see Table 3.1). Interestingly, in addition to the

*RgGH98*-encoding gene, two other genes are predicted to code for GH enzymes (GH95 and GH73) (Figure 3.1), suggesting that this cluster may also be dedicated to glycan utilisation.

Here we tested the hypothesis that *RgGH98* contributes to mucin glycan metabolism of *R. gnavus* strains and further explored the importance of *RgGH98* in the ability of *R. gnavus* strains to utilise mucin glycans.

Table 3.1. Description of the genes in *R. gnavus* ATCC 29149 RgGH98 cluster.

Locus tag	protein name	Old locus tag*	Putative function	Size (aa)
RGna_RS10300	WP_082923165.1	RUMGNA_01260	PTS ascorbate transporter subunit IIC	51
RGna_RS10305	WP_004841949.1	RUMGNA_01259	SPFH domain-containing protein; putative membrane protein	333
RGna_RS10310	WP_004841948.1	RUMGNA_01258	<b>GH31 <math>\alpha</math>-xylosidase (or <math>\alpha</math>-glucosidase)</b>	796
RGna_RS10315	WP_105084834.1	RUMGNA_03121	<b>GH95 fucosidase</b>	2236
RGna_RS10320	WP_004844095.1	RUMGNA_03120	leucine-rich repeat domain-containing protein	679
RGna_RS10325	WP_039959992.1	RUMGNA_03119	<b>GH98 blood-group endo-<math>\beta</math>-1,4-galactosidase</b>	1357
RGna_RS10330	WP_004844093.1	RUMGNA_03118	<b>GH73 peptidoglycan hydrolase with endo-b-N-acetylglucosaminidase specificity</b>	420
RGna_RS10335	WP_039959990.1	RUMGNA_03117	phage holin family protein	146
RGna_RS10340	WP_004844091.1	RUMGNA_03116	hypothetical protein	163
RGna_RS10345	WP_009244114.1	RUMGNA_03115	hypothetical protein	83
RGna_RS10350	WP_009244115.1	RUMGNA_03113	hypothetical protein	138
RGna_RS10355	WP_004844087.1	RUMGNA_03112	hypothetical protein	164
RGna_RS10360	WP_070101560.1	RUMGNA_03111	hypothetical protein	210

RGna\_RS10300- RGna\_RS10310 are not part of the RgGH98 operon. \*Locus tag from former version of *R. gnavus* genome annotation. The putative function is the automatically predicted function from the NCBI database.

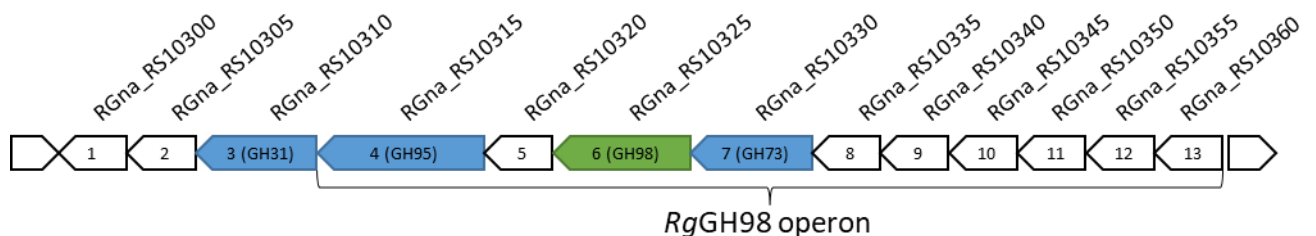


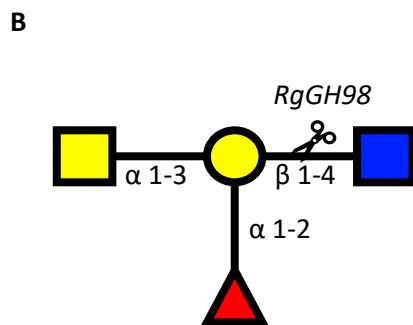
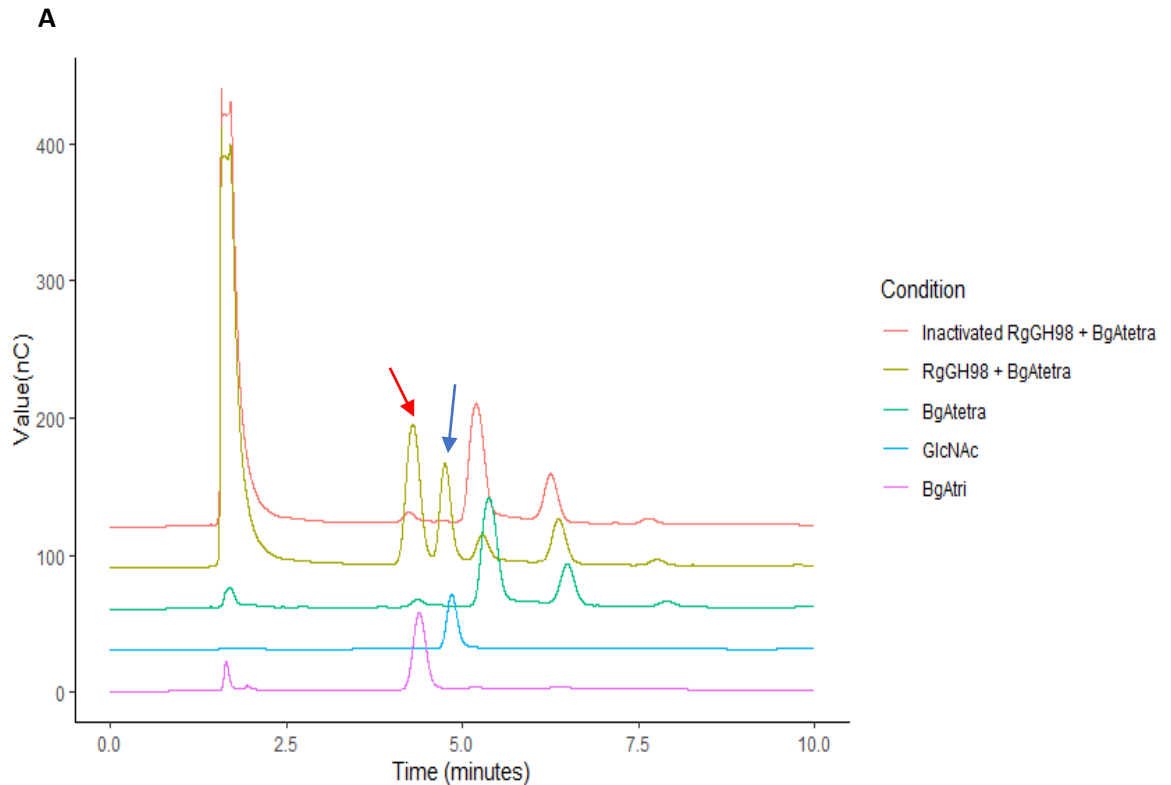
Figure 3.1. Schematic representation of *R. gnavus* ATCC 29149 gene-cluster including the RgGH98 gene. The RgGH98 encoding gene is indicated in green and other predicted GH-encoding genes are indicated in blue. RGna\_RS10300- RGna\_RS10310 are not part of the RgGH98 operon.

## 3.2 Results

### 3.2.1 RgGH98 is a galactosidase active on blood group A antigen

In order to experimentally validate the predicted enzymatic function of RgGH98, the RGna\_RS10325 gene encompassing the N-terminal galactose-binding-like domain (N-term GBLD), the catalytic domain (Cd), and the C-terminal domain was heterologously expressed in *Escherichia coli* (H. Wu, Crost, et al., 2021). Structural alignment to homologous GH98 enzymes identified the general acid catalytic residue as Glu 411 (H. Wu, Crost, et al., 2021).

The recombinant *RgGH98* was tested against blood group A tetrasaccharide type II (BgAtetra, GalNAc $\alpha$ 1-3[Fuca $\alpha$ 1-2]Gal $\beta$ 1-4GlcNAc) as a substrate, a common mucin epitope (Garcher et al., 1994; Ravn & Dabelsteen, 2000). The products of this enzymatic reaction were analysed by High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). The data showed the appearance of peaks corresponding to BgA trisaccharide (BgAtri, GalNAc $\alpha$ 1-3(Fuca $\alpha$ 1-2)Gal) and GlcNAc and the reduction of the peak corresponding to BgAtetra (Figure 3.2A). The peaks were absent when the catalytic mutant (E411A) was used as control. Further work by Dr. Haiyang Wu showed that *RgGH98* was specific for BgA antigen and not active on BgB or BgH antigens (H. Wu, Crost, et al., 2021).



**Figure 3.2.** *RgGH98* enzymatic characterisation. **(A)** HPAEC-PAD analysis of the enzymatic reaction using *RgGH98* or *RgGH98* E411A catalytic inactive mutant on BgAtetra. BgAtetra is cleaved into BgAtri (red arrow) and GlcNAc (blue arrow). Data kindly provided by Ana Martínez Gascueña. **(B)** Schematic of the enzymatic reaction. *RgGH98* cleaves BgAtetra  $\beta$ 1-4 glycosidic linkage to produce free GlcNAc and BgAtri.

### 3.2.2 *RgGH98* mediates the growth of *R. gnavus* strains on mucins

#### 3.2.2.1 Validation of a suitable minimum media for *R. gnavus* growth

The newly developed LAB medium (Tramontano et al., 2018) was tested as a minimal medium for *R. gnavus* growth. The advantage of a defined medium is that the composition is known, which is essential to assay the role of specific compounds in a metabolic system. *R. gnavus* E1 and ATCC 29149 strains were anaerobically grown in LAB and YCFA media in the presence or absence of glucose. No growth was observed in LAB in the absence of a



carbon source while the bacteria could grow to high levels in LAB supplemented with 0.5% glucose (Figure 3.3). The growth of both strains was significantly higher in the LAB medium supplemented with glucose, with an OD<sub>595</sub> reaching around 1.6 whereas the strains reached an OD of around 1.1 in YCFA supplemented with glucose (Figure 3.3). In both media, *R. gnavus* E1 strain showed a longer lag phase, which was more pronounced in YCFA supplemented with glucose.

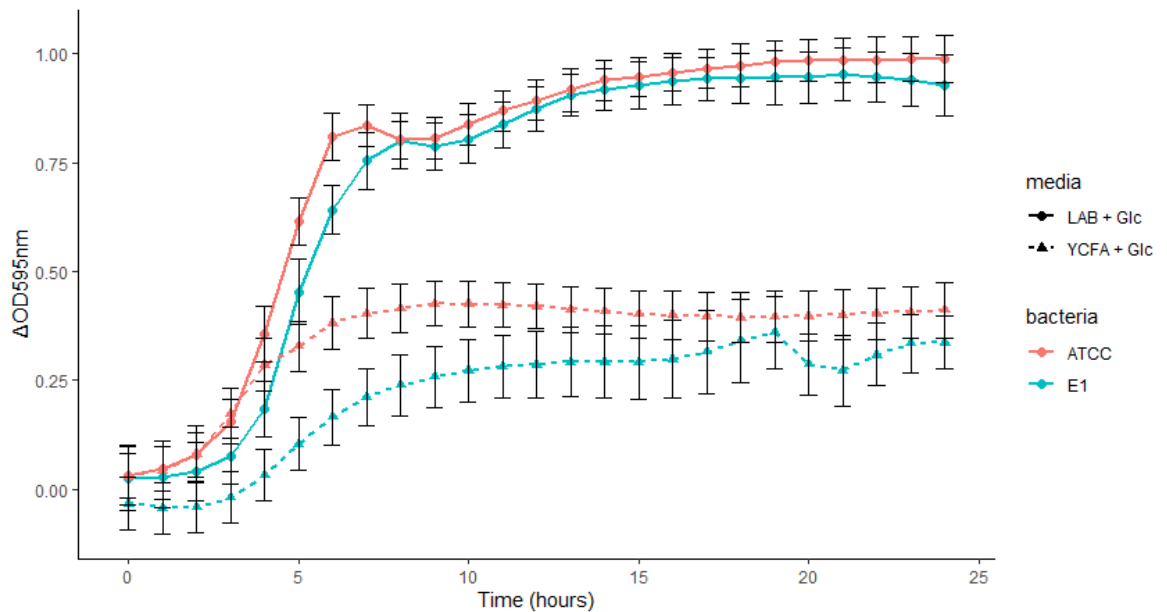


Figure 3.3. Growth curves of *R. gnavus* ATCC 29149 and E1 in LAB or YCFA medium supplemented with 0.5% glucose. The points show the average OD values of four replicates, error bars indicate standard deviation.  $\Delta$ OD was calculated by subtracting the negative control (LAB + bacteria) from the experimental samples (LAB + Glc + bacteria).

### 3.2.2.2 *R. gnavus* growth on mucin and BgA is strain-specific

To further investigate the role of *RgGH98* in the ability of *R. gnavus* strains to grow on mucins, we tested the ability of *R. gnavus* E1 strain to grow in LAB medium supplemented with 1% mucin (pPGM) pre-treated with *RgGH98*. *R. gnavus* E1 strain showed a significant increase in growth in *RgGH98*-treated mucin ( $p = 0.023$  at 9 hours and  $p = 0.033$  at 48 hours) as compared to the control (untreated mucin). The *RgGH98* treatment had less of an effect on the growth of *R. gnavus* ATCC 29149, which is in line with the ability of this strain to express *RgGH98* (Figure 3.4A). The data were confirmed by qPCR analysis of *R. gnavus* copy numbers at T= 0-, 9-, and 24-hours using *R. gnavus* 16S, *R. gnavus* ATCC 29149 specific primers, and *R. gnavus* E1 specific primers (Figure 3.5), with significant increases in *R. gnavus* ATCC 29149 copy numbers after 9 hours of growth on both untreated pPGM

( $p < 0.000$ ) and *RgGH98*-treated pPGM ( $p < 0.000$ ). *R. gnavus* E1 copy numbers after 9 hours of growth on untreated pPGM was not significant ( $P = 0.53$ ) but was significant on *RgGH98*-treated pPGM ( $p = 0.02$ ).

Since both *R. gnavus* E1 and ATCC 29149 strains could grow on *RgGH98*-treated mucin, we then hypothesised that the bacteria could metabolise BgAtri, the product of the *RgGH98* enzymatic reaction. Using HPAEC-PAD, we showed that when *R. gnavus* E1 was grown on *RgGH98*-treated pPGM, the BgAtri was consumed and no longer detected in the spent medium after 9 hours of growth, while it was detected in the no bacteria control (Figure 3.4B-D).

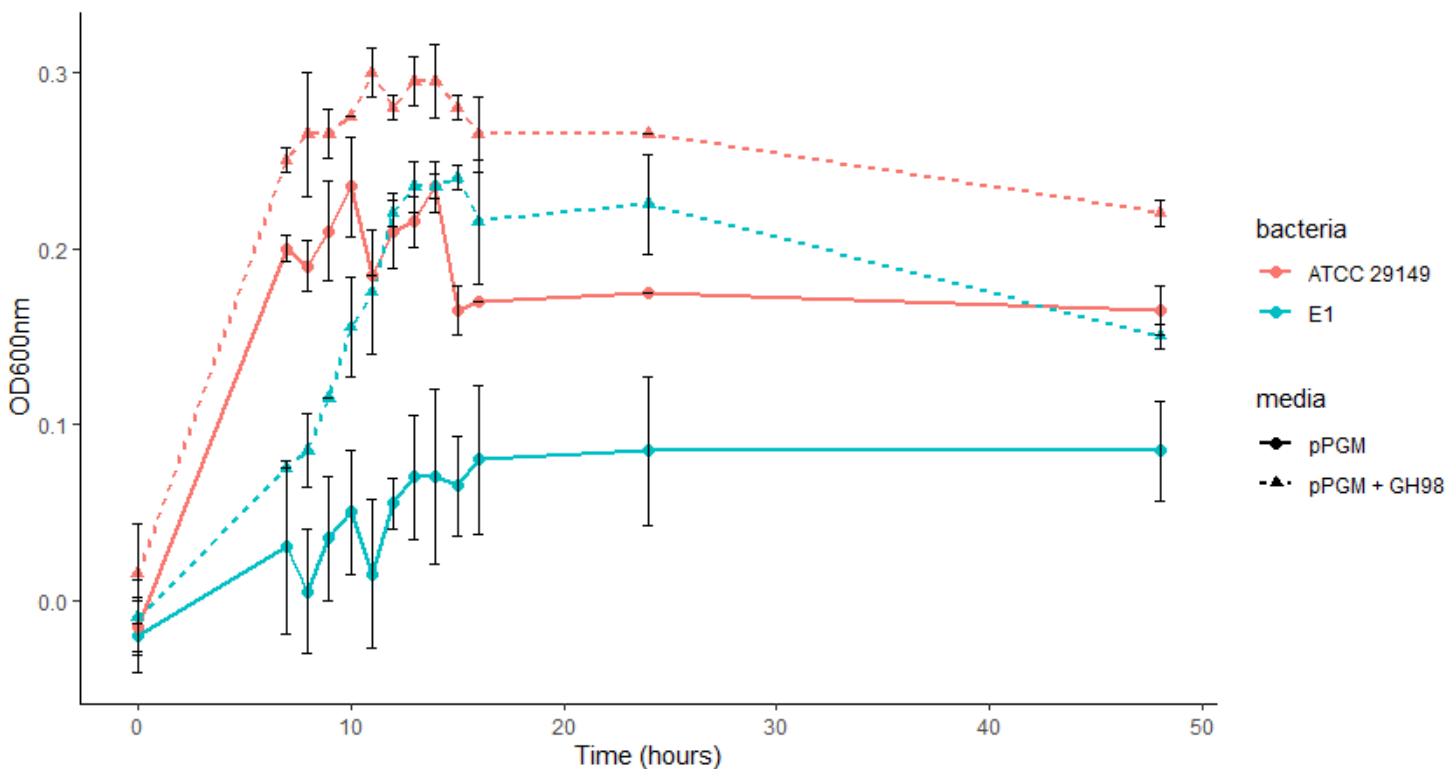


Figure 3.4. Growth of *R. gnavus* ATCC 29149 and E1 in LAB supplemented with untreated or *RgGH98*-treated pPGM over 48 hours.  $\Delta OD$  was calculated by subtracting the negative control (LAB + mucin) from the experimental samples (LAB + mucin + bacteria). Averages are calculated from the average of two biological repeats, error bars indicate standard deviation.

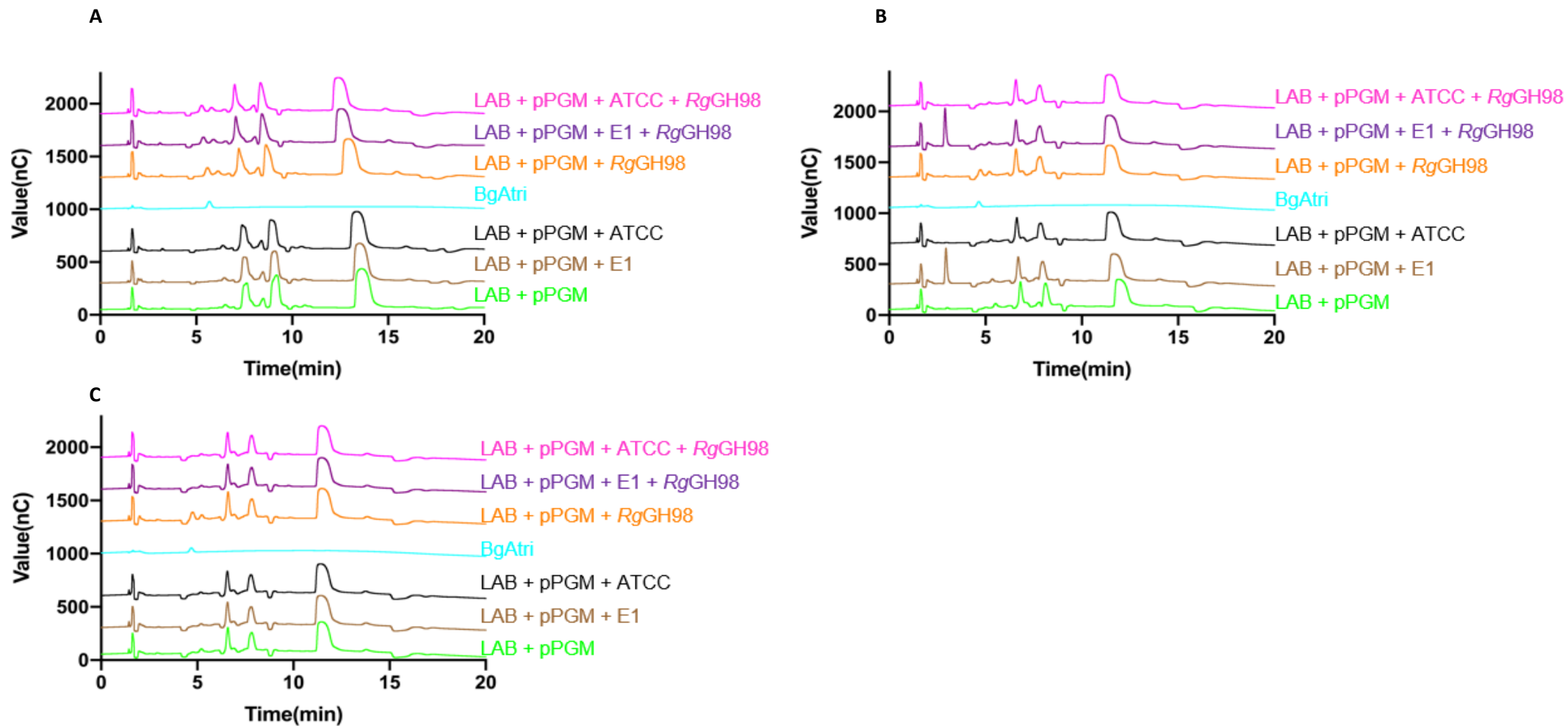


Figure 3.5. Analysis of BgAtri utilisation during growth of *R. gnavus* strains on mucin. HPAEC-PAD analysis of the supernatant of *R. gnavus* ATCC 29149 and E1 cultures in LAB supplemented with untreated or *RgGH98*-treated pPGM at (A) T= 0 hours, (B) T= 9 hours and (C) T= 48 hours. BgAtri, LAB supplemented with pPGM, and LAB supplemented with *RgGH98*-treated pPGM were used as controls. This figure has also been published in (Wu, Crost, et al., 2021) and appears here with permission.

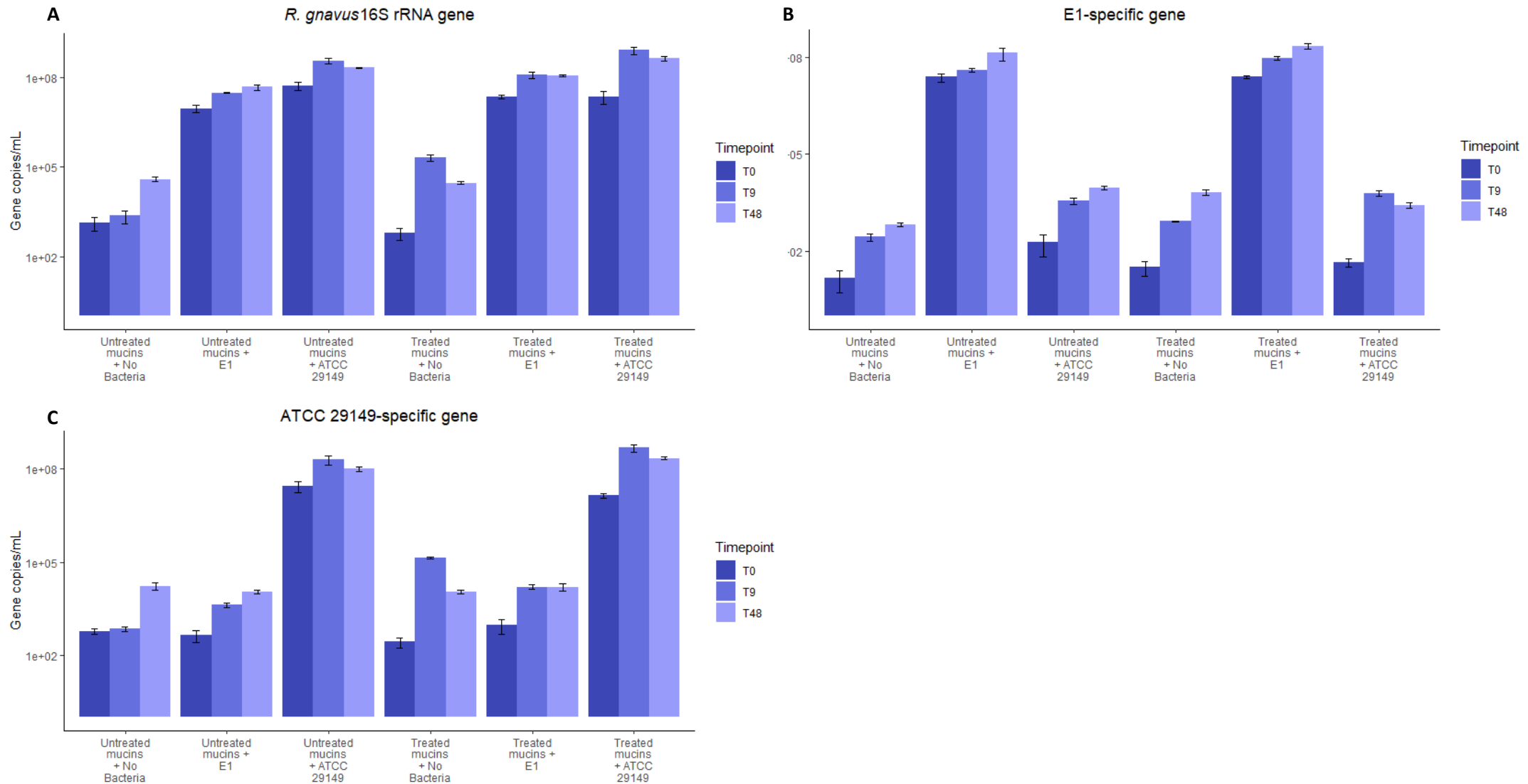


Figure 3.6. qPCR analysis of *R. gnavus* growth in LAB supplemented with RgGH98-treated or untreated mucin. DNA was extracted at T=0, T=9, and T=48 hours of *R. gnavus* E1 or ATCC 29149 grown in LAB supplemented with 1% pPGM or 1% RgGH98-treated pPGM and analysed by qPCR using (A) *R. gnavus* 16S universal primer (Rg5 16S), (B) *R. gnavus* E1 specific primers (RUMGNEv3\_60291), and (C) *R. gnavus* ATCC 29149 specific primers (RUMGNA\_02237). The absolute number of bacterial quantity through qPCR was determined using a calibration curve generated using the amplified regions for each primer pair. Data are expressed as gene copy numbers per ml. Two biological replicates with three technical replicates each are shown. Error bars correspond to standard error.

HPAEC-PAD was used to quantify the released BgAtri in the growth medium after 48 hours of pre-treatment. When pPGM was treated with 1.3  $\mu\text{M}$  *RgGH98*, which is the concentration used across all experiments, an average concentration of 66.5  $\mu\text{M}$  BgAtri was detected in the absence of bacteria (Table 3.2).

Table 3.2. Concentrations of BgAtri released from pPGM by *RgGH98*.

	BgA concentration in $\mu\text{M}$		
	T = 48	T = 57	T = 96
pPGM + <i>RgGH98</i> (1)	67.4	69.6	81.6
pPGM + <i>RgGH98</i> (2)	65.7	66.1	80.4
Average	66.5	67.85	81

Samples were taken after 48, 57, and 96 hours of treatment as quantified by HPAEC. (1) and (2) correspond to biological repeats.

To validate these findings, we next evaluated the ability of these strains to grow on commercially available BgAtri and BgAtetra (Figure 3.6). The bacteria were grown on LAB supplemented with 1.5 mM, 0.5 mM, and 0.05 mM concentrations of commercial BgAtri or BgAtetra. *R. gnavus* ATCC 29149 grew to high density on all three concentrations of BgAtri as well as on 1.5 mM and 0.5 mM BgAtetra, but not on 0.05 mM BgAtetra. *R. gnavus* E1 however was not able to grow on BgAtetra at the concentrations tested but grew at the two highest concentrations of BgAtri (Figure 3.6). At the lowest concentration, which is similar to the concentration of BgAtri released after 48 hours of pre-treatment of *RgGH98* with pPGM, only *R. gnavus* ATCC 29149 (but not E1) showed growth on commercial BgAtri (Figure 3.6). Together these results suggest that the released BgAtri was utilised but not sufficient to fully sustain *R. gnavus* E1 growth.

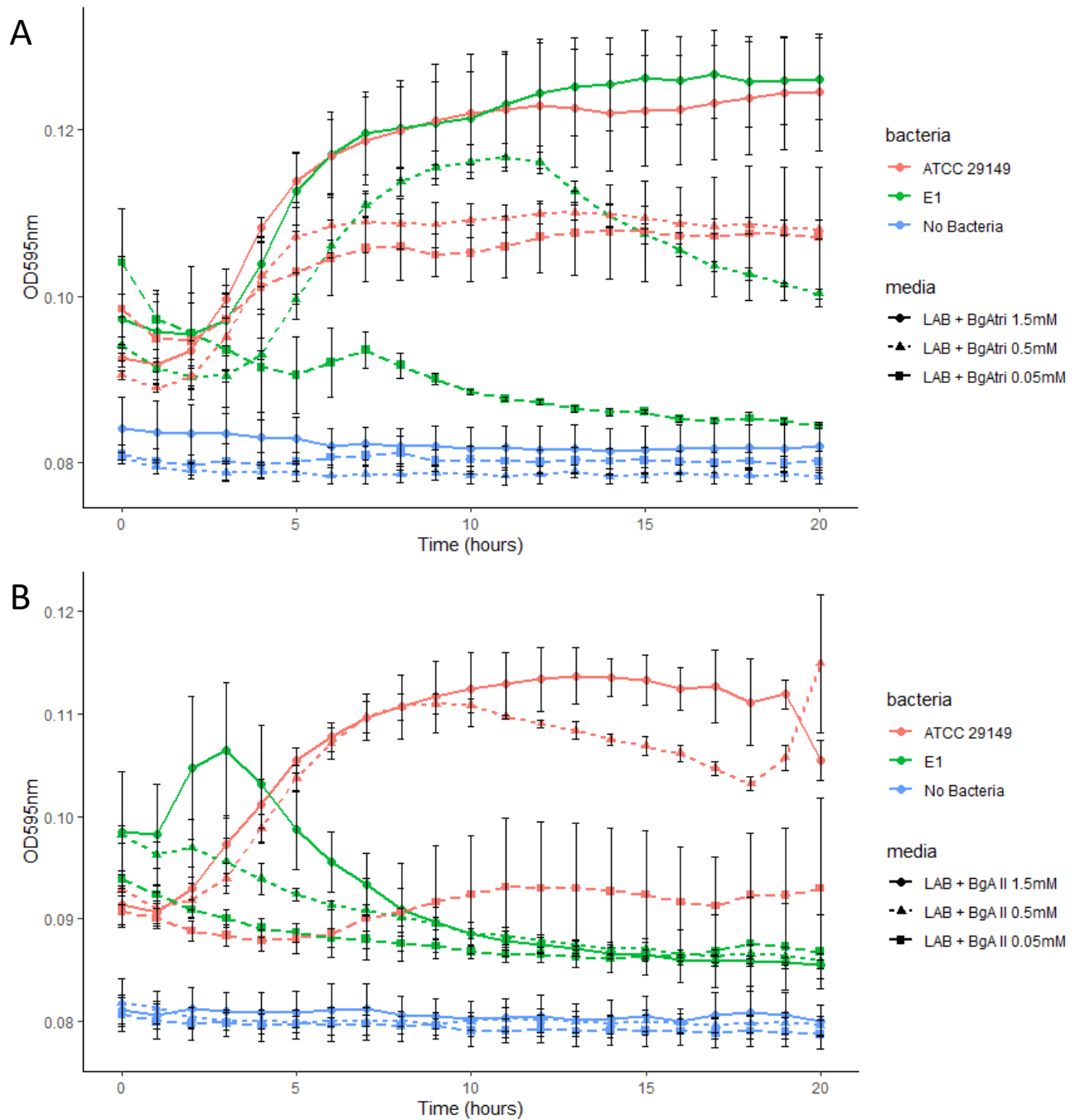


Figure 3.7. Growth curves of *R. gnavus* ATCC 29149 and E1 on BgAtri and BgAtetra. *R. gnavus* growth in LAB supplemented with 1.5 mM, 0.5 mM, and 0.05 mM commercial BgAtri (A) or BgAtetra (B). Averages of three replicates are shown, error bars indicate standard deviation. This figure has also been published in (Wu, Crost, et al., 2021) and appears here with permission.

In addition, since *R. gnavus* ATCC 29149 expresses *RgGH98*, the observed growth on BgAtetra was consistent with *RgGH98* substrate specificity to hydrolyse BgAtetra into BgAtri and GlcNAc. Further, the effect of *RgGH98*-treatment on mucin was confirmed by MALDI-TOF/TOF mass spectrometry. After *RgGH98*-treatment, increased levels of BgA epitopes were detected in the growth medium due to the release of BgAtri from the mucin. Following growth with *R. gnavus* E1 or ATCC 29149, BgA abundance was significantly decreased after 24 hours (Figure 3.7), indicating consumption of BgA by both strains, whereas the relative abundance of the other fucosylated peaks remained largely the same. These results are in line with the ability of the strains to utilise BgAtri as a metabolic substrate.

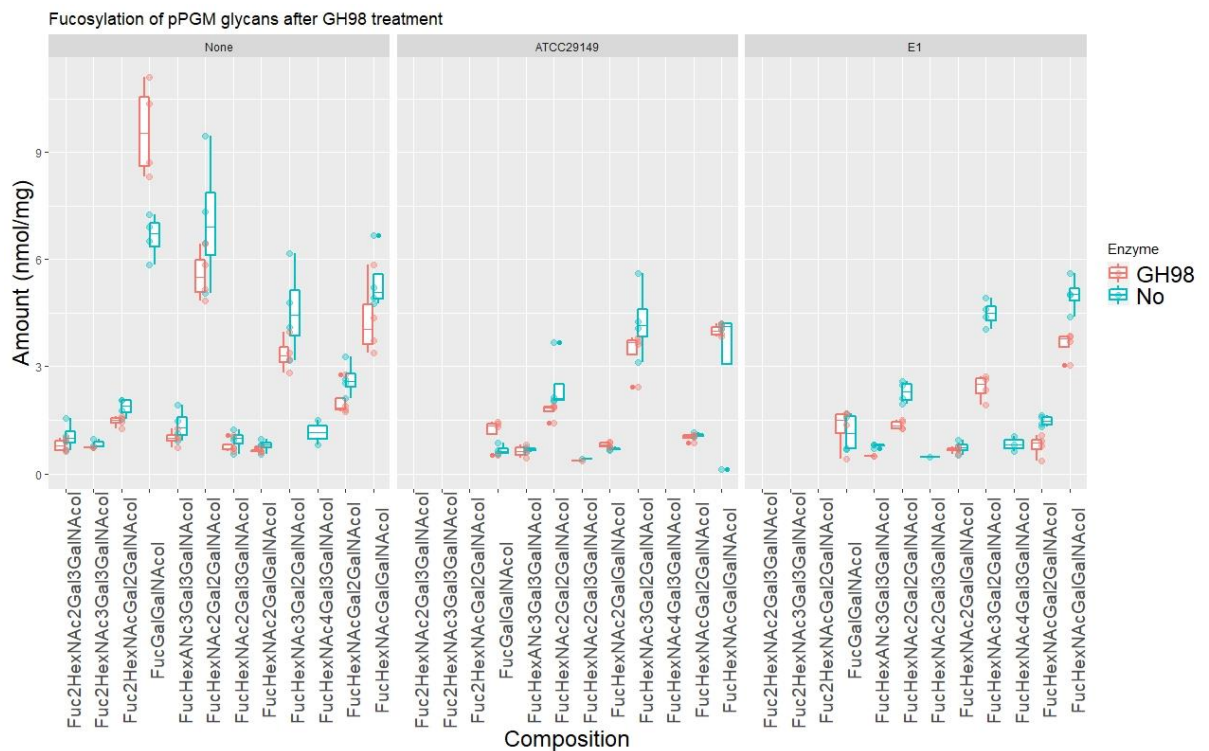


Figure 3.8. Composition of mucin glycans following *R. gnavus* growth on untreated and *RgGH98*-treated mucin. Relative abundance in nmol/mg and composition of mucin glycans of untreated and *RgGH98*-treated pPGM after 24 hours of *R. gnavus* E1 and ATCC 29149 growth, determined by MS analysis. Red boxes are samples from *RgGH98*-treated pPGM, blue boxes are from untreated pPGM. FucGalGalNAcol indicates BgAtri residue or Fuc-Gal-GalNAc-ol. This figure has also been published in (H. Wu, Crost, et al., 2021) and appears here with permission.

We next tested whether, when growing on *RgGH98*-treated pPGM, the bacteria benefited from the release of BgAtri and/or from access to the rest of the mucin glycan chain. Following treatment of pPGM with *RgGH98*, the suspension was dialysed, to separate the

enzymatically treated mucin from the released oligosaccharides (see section 2.2.1). Both strains were grown in LAB supplemented with these two separate fractions. *R. gnavus* ATCC 29149 showed growth on the ‘leftover’ pPGM fraction, while *R. gnavus* E1 showed a delayed growth in this fraction compared to *R. gnavus* ATCC 29149. No growth was observed on the released oligosaccharide fractions (Figure 3.8), which contrasts with the results obtained on commercial BgAtri (Figure 3.6). The lowest concentration BgAtri that *R. gnavus* was grown on was slightly less than the concentration *RgGH98* releases from pPGM (50  $\mu$ M compared to 60  $\mu$ M), which might explain this difference.

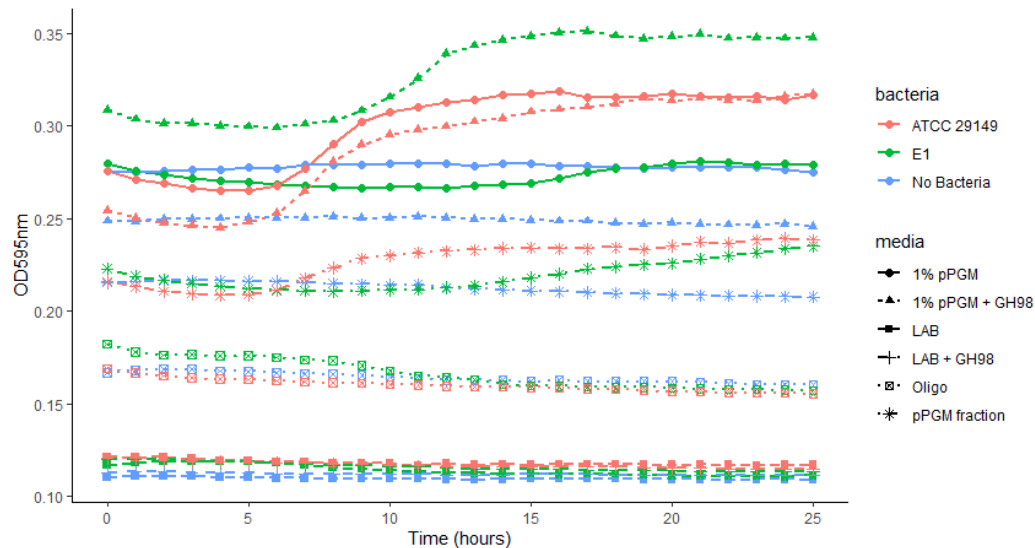


Figure 3.9. Growth of *R. gnavus* ATCC 29149 and E1 on dialysed *RgGH98*-treated pPGM. Growth in LAB supplemented with dialysed *RgGH98*-treated pPGM fraction, dialysed released oligosaccharide fraction, or LAB supplemented with untreated or *RgGH98*-treated pPGM over 25 hours. Averages of three replicates are shown.

### 3.2.3 The *RgGH98* cluster is upregulated during growth of *R. gnavus* ATCC 29149 on BgAtri and BgAtetra

In order to gain more insights into the metabolic pathway involved in BgA utilisation, we analysed the transcriptomic response of the *RgGH98* cluster (Figure 3.1) by RT-qPCR when *R. gnavus* ATCC 29149 was grown on LAB supplemented with 0.5% glucose, 1 mM BgAtri, or 1 mM BgAtetra (Figure 3.9). The lag phase during growth on both BgAtri and BgAtetra was longer than observed previously (Figure 3.6), but a high bacterial density could be observed after 9 hours of growth on BgAtetra, while growth on BgAtri reached a plateau



between 11 and 24 hours of growth. Growth was verified by qPCR analysis using *R. gnavus* 16S and *R. gnavus* ATCC 29149 specific primers (Figure 3.10).

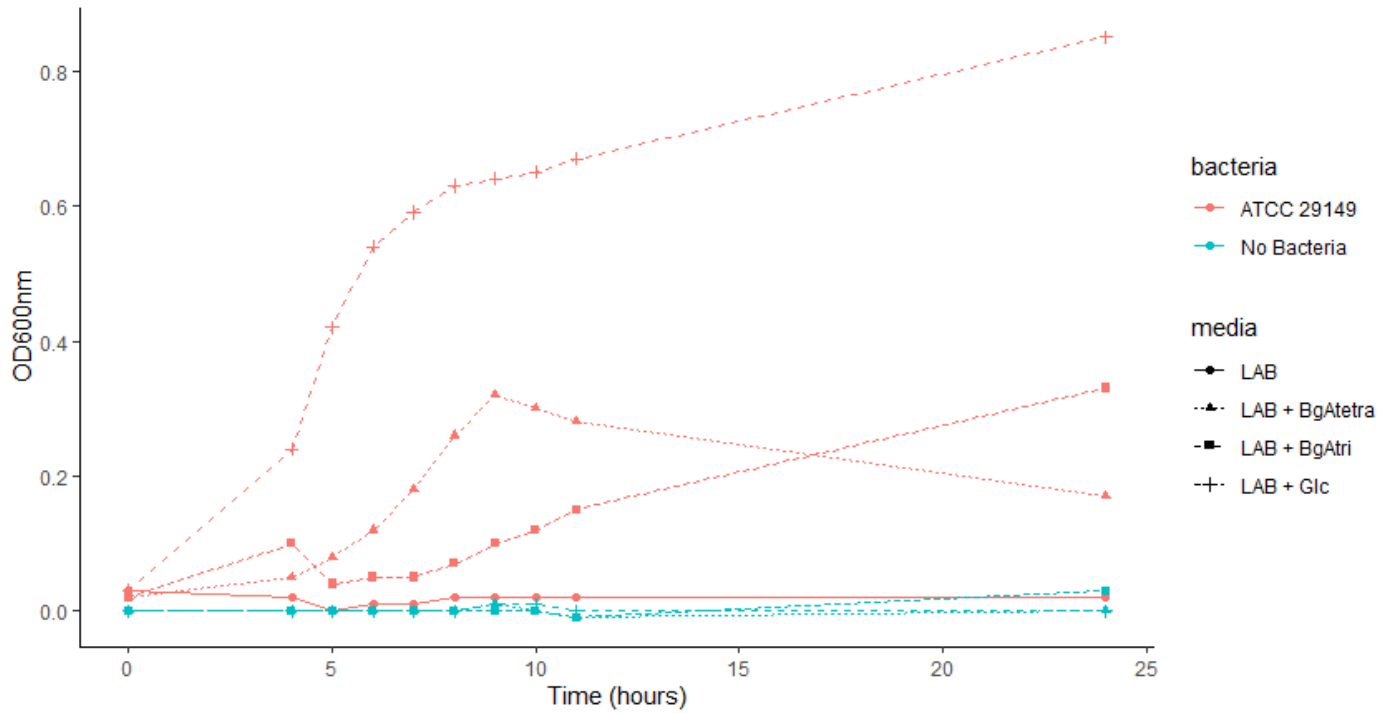


Figure 3.10. Growth of *R. gnavus* ATCC 29149 on BgAtri and BgAtetra. *R. gnavus* ATCC 29149 was grown in LAB supplemented with 0.5% glucose (27,7 mM), 1 mM BgAtri or 1 mM BgAtetra. A single repeat is shown.

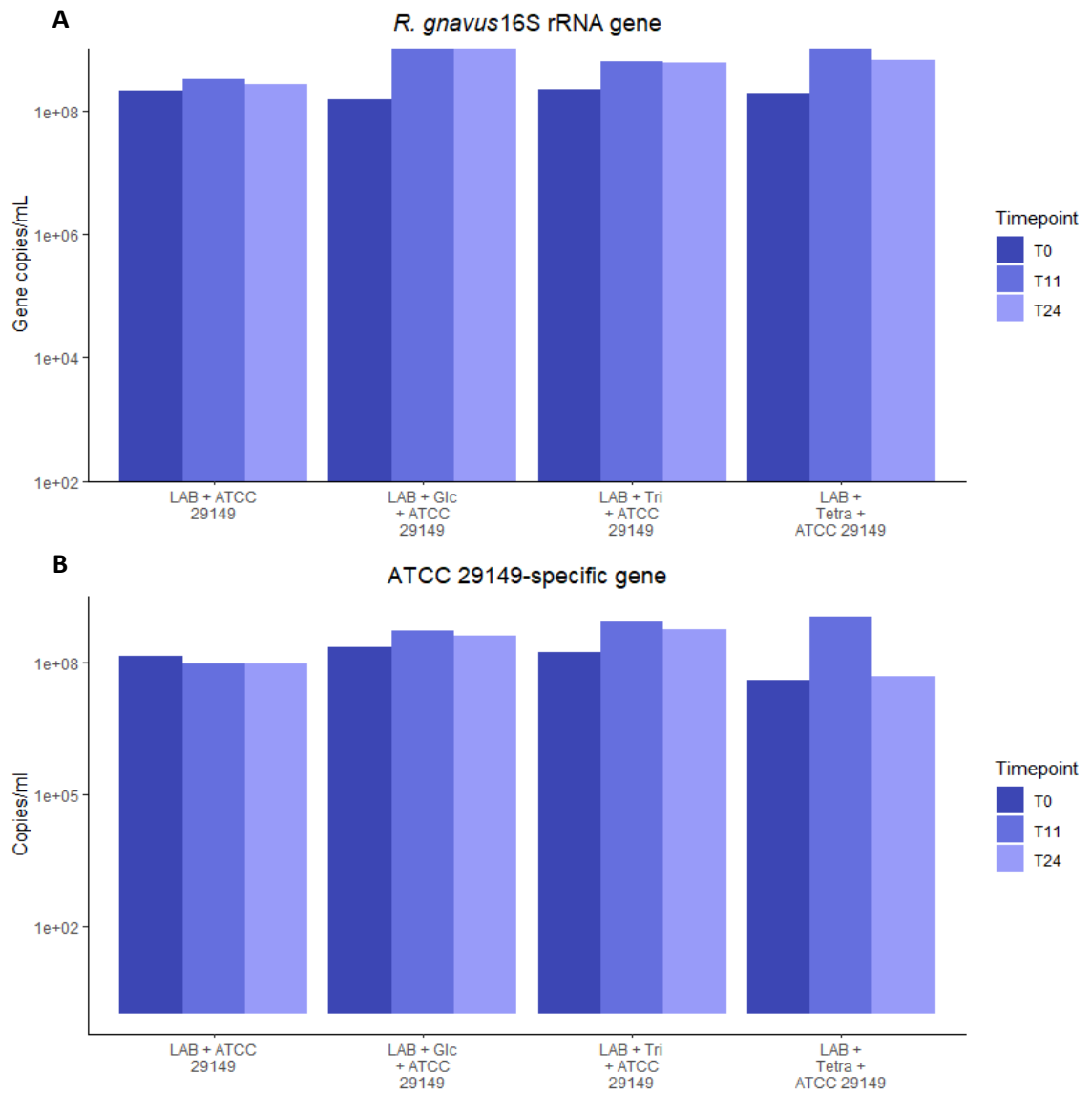


Figure 3.11. Copy numbers per ml during *R. gnavus* ATCC 29149 growth on BgAtri and BgAtetra. Samples of *R. gnavus* ATCC 29149 grown on LAB, LAB + 0.5% glucose (27,7 mM), LAB + 1mM BgAtri or 1mM BgAtetra were taken at T=0, T=11, and T=24 hours. **(A)** *R. gnavus* 16S rRNA gene copies/ml. **(B)** *R. gnavus* ATCC 29149 RUMGNA\_02237 gene copies/ml.

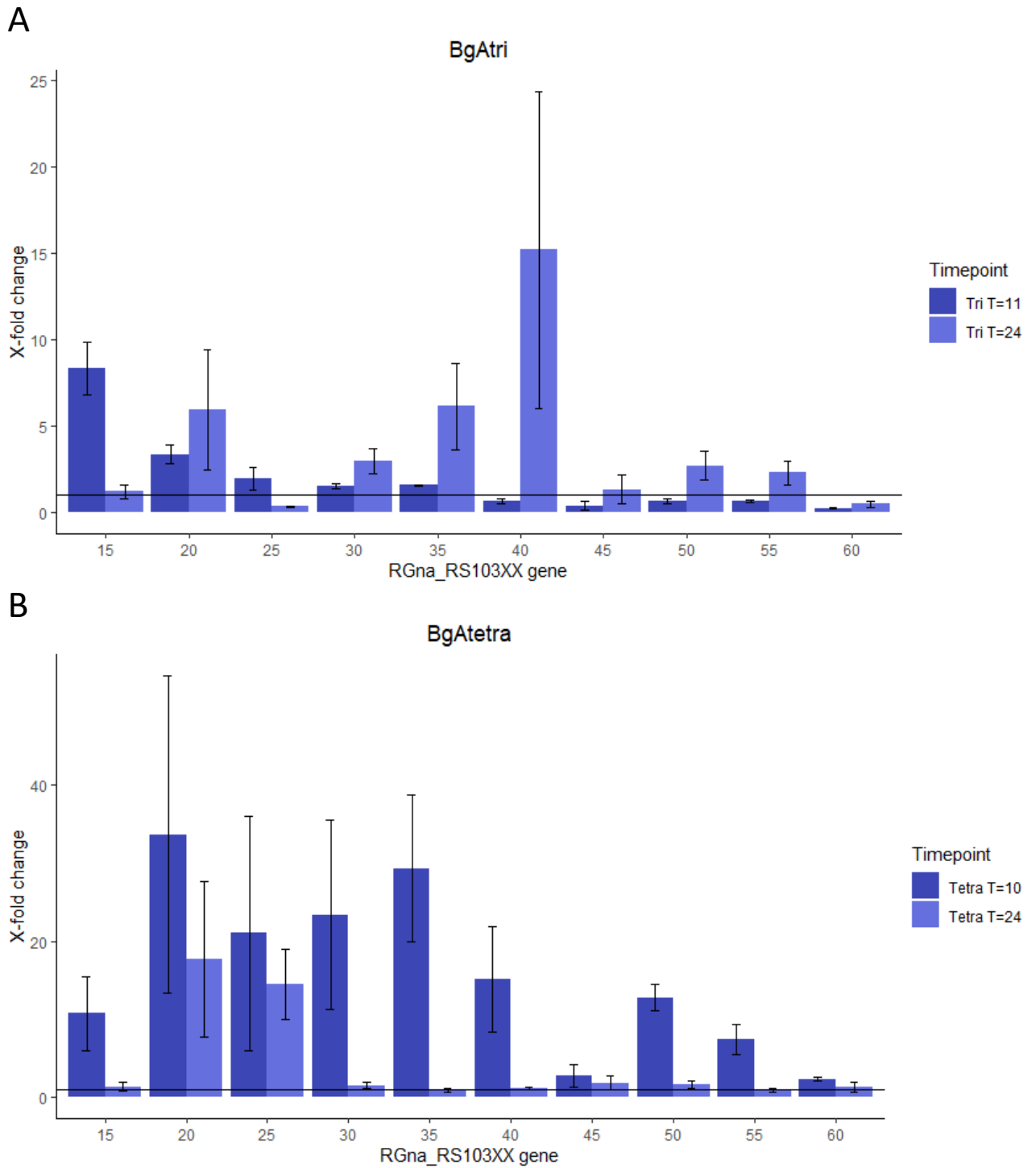
LC-MS analysis confirmed that *R. gnavus* ATCC 29149 could utilise BgAtri, with concentrations of BgAtri in the spent medium decreasing significantly from 11 hours to 24 hours (Table 3.3). The concentration of BgAtetra did not decrease when *R. gnavus* ATCC 29149 was grown on BgAtetra, even though growth was observed.

Table 3.3. Concentration of BgAtri and BgAtetra in medium during growth of *R. gnavus* ATCC 29149.

	BgA concentration in $\mu\text{M}$		
	T0	T10/11	T24
LAB + BgAtri + ATCC 29149	514 $\mu\text{M}$	388 $\mu\text{M}$	0 $\mu\text{M}$
LAB + BgAtetra + ATCC 29149	844 $\mu\text{M}$	906 $\mu\text{M}$	988 $\mu\text{M}$

Consumption of BgAtri and BgAtetra after 10 or 11 and 24 hours of *R. gnavus* ATCC 29149 growth as quantified by LC-MS. BgAtri was measured after 11 and 24 hours; BgAtetra was measured after 10 and 24 hours.

RNA was extracted from samples collected at late exponential phase (T=7-11 hours depending on growth) and after 24 hours (for concentration and quality of RNA see Supplemental Table S3) and converted into cDNA. RT-qPCR was performed for genes within the *RgGH98* cluster using gene specific primers and the Gyrase B housekeeping gene for normalisation and growth on glucose as the control condition. The log-fold change in gene expression showed differences between the first three genes in the cluster (RGna\_RS10315-25), including the *RgGH98* encoding gene (RGna\_RS10325), and genes RGna\_RS10330-55. During early growth on BgAtri, only the first genes (RGna\_RS10315-35) were upregulated, while after 24 hours, RGna\_RS10330- RGna\_RS10355 and RS10320 were upregulated. During growth on BgAtetra the whole cluster was upregulated. However, after 24 hours, gene expression returned to basal expression, with the exception of RGna\_RS10320 and RGna\_RS10325 (the *RgGH98* encoding gene). RGna\_RS10325, the *RgGH98* encoding gene, was highly upregulated during growth on BgAtetra, while only slightly during early BgAtri growth, in line with its substrate specificity. The expression of RGna\_RS10315 (coding for a predicted GH95 fucosidase) was found to be highly upregulated during growth on both BgAtri and BgAtetra while RGna\_RS10330 (encoding for a predicted GH73 acetylglucosaminidase) was upregulated during growth on both BgAtri and BgAtetra (Figure 3.11).



**Figure 3.12.** Transcriptomics analysis of *RgGH98* cluster during *R. gnavus* ATCC 29149 growth on BgAtri or BgAtetra compared to growth on glucose. Bars indicate fold changes of RGna\_RS10315 to RGna\_RS10360 expression during **(A)** growth on BgAtri at T11 and T24 hours and **(B)** growth on BgAtetra at T10 and T24 hours based on three repeats. Horizontal lines indicate a fold change of 1, indicating an unchanged expression level compared to growth on glucose. X-fold changes are shown based on differences in expression calculated from changes in  $2^{-\Delta\Delta Ct}$  after normalising against the *GyrB* housekeeping gene grown in LAB supplemented with 0.5% glucose as control condition. Error bars correspond to the standard-error. 15-60 correspond to RGna\_RS10315 to RGna\_RS10360 genes. See Table 3.1 for full list of genes and predicted function.

### 3.3 Discussion

We showed that *RgGH98* plays a critical role in the ability of *R. gnavus* strains to grow on mucin. The specificity of *RgGH98* for BgA might give a competitive advantage in mucosal regions where blood group epitopes are highly expressed such as in ileum, cecum, and proximal colon (Robbe et al., 2003, 2004). This work complements previous work from the group showing the importance of GH29 and GH95 fucosidases and GH33 sialidases in the ability of *R. gnavus* to grow on mucin (Bell et al., 2019; Crost et al., 2013, 2016; Owen et al., 2017; H. Wu, Rebello, et al., 2021). Together these data point towards the role of *R. gnavus* exo-GHs in targeting the main terminal glycan epitopes (sialic acid, fucose, BgA) in intestinal mucins, which can then be used as metabolic substrate for *R. gnavus* strains (Figure 3.12B). This was recently demonstrated for 2,7-anhydro-Neu5Ac, the product of *R. gnavus* sialidase reaction, which has been shown to be transported and metabolised in *R. gnavus* strains due to the exquisite specificity of the transporter for 2,7-anhydro-Neu5Ac, and the presence of an oxidoreductase converting 2,7-anhydro-Neu5Ac back to Neu5Ac (Bell et al., 2019, 2020).

The *RgGH98* enzyme is part of a gene cluster. We showed that during exponential growth on BgAtri and BgAtetra, the whole cluster was upregulated, indicating that it is dedicated to BgA metabolism. The genes encoding for a GH73 *N*-acetylglucosaminidase (GlcNAcase) and a GH95 fucosidase were upregulated when *R. gnavus* was grown on both BgAtri and BgAtetra. The GH73 overexpression is surprising as no GlcNAc residue is present in BgAtri and annotated proteins with high sequence similarity to this GH73 are all peptidoglycan hydrolases. Previously, the whole *RgGH98* cluster was shown to be also upregulated during growth on pPGM (H. Wu, Crost, et al., 2021), suggesting a broader role of this operon in mucin glycan metabolism. Further, we showed that, in addition to consuming BgA, the *R. gnavus* strains were able to grow on the *RgGH98*-enzymatically treated mucin glycan chains, suggesting that the uncapping allows other GHs to access the underlying glycan linkages, releasing new metabolic substrates for the bacteria (Figure 3.12).

*RgGH98* was found to be specific for blood group A by HPAEC-PAD analysis. This was further confirmed by STD-NMR and isothermal titration calorimetry (ITC) analysis (H. Wu, Crost, et al., 2021). In contrast to *RgGH98*, other functionally characterised GH98s have been shown to release both BgA and BgB trisaccharide from glycans, like *Streptococcus pneumoniae*

SP3-BS71 Sp3GH98 (Higgins et al., 2009), and GH98 E-ABase from *Clostridium perfringens* ATCC 10543, which has a preference for glycans with a core 2 structure over a core 1 structure (Anderson et al., 2005). *S. pneumoniae* TIGR4 Sp4GH98, however, showed specificity to Lewis Y epitopes (Higgins et al., 2009). The specificity of RgGH98 for BgA epitopes can have potential applications as tools to discriminate between different blood group epitopes.

Blood group epitopes have been shown to be involved in pathogen infection, as they can be used as recognition sites (Cooling, 2015; Ruvoën-Clouet et al., 2013). For example, SARS-CoV-2 receptor-binding domain has been shown to preferentially recognise blood group A (S. C. Wu et al., 2021), which may underpin the higher risk of SARS-COV-2 infection reported in patients with blood group A (Khder Mustafa et al., 2023). Advanced understanding of the structure and function of these enzymes involved in the modification of blood group epitopes can help future development of drugs targeting these pathogens.

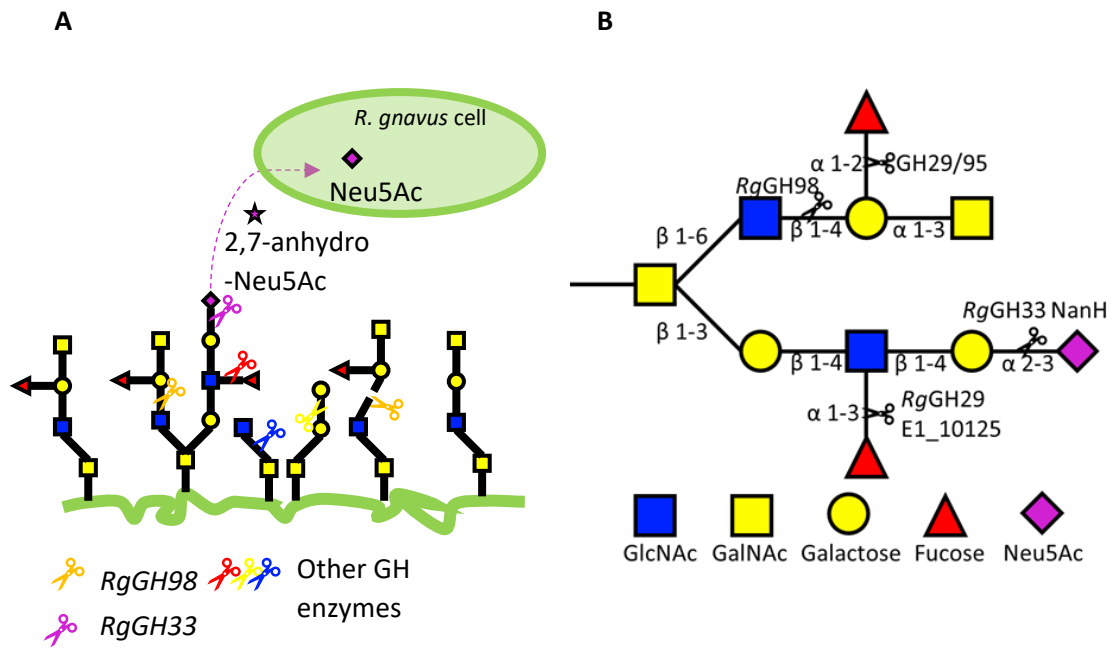


Figure 3.13 Model of *R. gnavus* mucin glycan degradation strategy. **(A)** *RgGH98* cleaves BgAtri epitopes, which opens space and allows other GH enzymes to cleave more glycans. *RgGH33* cleaves Neu5Ac epitopes releasing 2,7-anhydro-Neu5Ac, once transported into the cell this is converted back to Neu5Ac. **(B)** Functionally characterised GHs involved in *R. gnavus* degradation of mucin glycan chains. *RgGH98* is active on  $\beta$ 1-4 linked BgA epitopes GH29 and GH95 enzymes are active on fucose epitopes. *RgGH29* E1\_10125 is active on  $\alpha$ 1-3 linked fucose on sLe<sup>x</sup> epitopes. *RgGH33* NanH is active on  $\alpha$ 2-3 linked Neu5Ac.

## Chapter 4

### Identification and characterisation of *R. gnavus* GHs involved in mucin glycan degradation



## 4. Identification and characterisation of *R. gnavus* GHs involved in mucin glycan degradation

### 4.1 Introduction

Mucin glycans provide a constant source of nutrients to the bacteria inhabiting the intestinal mucus niche. There, bacteria produce sulfatases (Luis et al., 2021), fucosidases (GH29 and GH95) and sialidases (GH33) to remove glycan epitopes. The mucin glycan chains are then further broken down by GHs hydrolysing glycosidic linkages to galactose (Gal) (GH2, GH35, GH42, GH98, GH110), *N*-acetylgalactosamine (GalNAc) (GH27, GH36, GH101, GH109, GH129), or *N*-acetylglucosamine (GlcNAc) (GH20, GH84, GH85, GH89) (Bell & Juge, 2021; Raba & Luis, 2023; Tailford, Crost, et al., 2015). Human milk oligosaccharides have similar structures to mucin glycans, and thus require a similar set of GH families to be broken down (Marcobal et al., 2011).

*R. gnavus*, *A. muciniphila* (Derrien et al., 2004), *B. thetaiotaomicron* (Martens et al., 2008), *Bacteroides caccae* (Desai et al., 2016), *B. fragilis* (Pudlo et al., 2015), *B. bifidum* (Kiyohara et al., 2011), and *C. perfringens* (Low et al., 2021) have all been shown to degrade mucin glycans and HMOs (see section 1.2.4). In *Bacteroides* species, these mucin glycan degrading GHs are organised into polysaccharide utilization loci (PULs) (Bjursell et al., 2006; Feng et al., 2022) whereas in Firmicutes and other bacteria genes involved in the metabolism of mucin glycans are clustered in operons (Sela et al., 2008; Tailford, Crost, et al., 2015).

The capacity of *R. gnavus* ATCC 29149 to utilise mucin glycans is based on the complement of GHs encoded by its genome. *RgGH98* was shown in the previous chapter to be involved in the mucin degradation strategy of this strain. Furthermore, several fucosidases and a sialidase were shown to be active on glycan structures which are present in mucins (Tailford, Owen, et al., 2015; H. Wu, Rebello, et al., 2021). However, the full complement of GHs involved in mucin glycan degradation of *R. gnavus* is not yet known. Here we have used RNA sequencing (RNAseq) to identify other GHs potentially involved in mucin glycan degradation. The activity of recombinant enzymes was then tested experimentally.

## 4.2 Results

### 4.2.1 Growth of *R. gnavus* ATCC 29149 on mucin glycans

*R. gnavus* E1 is able to grow on RgGH98-treated pPGM and on the underlying mucin fraction (see 3.2.2.2). To further explore the role of other GHs in *R. gnavus*' ability to grow on mucin, the growth of *R. gnavus* strains was tested on mucin glycans released from the protein backbone by reductive  $\beta$ -elimination from PGM using NaOH and NaBH<sub>4</sub>, as previously described (Pruss et al., 2021). Following dialysis to remove the borates from  $\beta$ -elimination, the mucin glycan content was analysed by MALDI-TOF MS. Glycans of a size up to pentasaccharides were significantly reduced after dialysis, while larger glycan structures were at similar levels before and after dialysis (Figure 4.1). The growth of both *R. gnavus* ATCC 29149 and E1 strains was tested on pPGM, PGM, pPGM mucin glycans, and PGM mucin glycans (Figure 4.2). During growth on released PGM glycans, *R. gnavus* ATCC 29149 displayed a shorter lag phase compared to growth on pPGM and PGM, suggesting that mucin glycans are easier to metabolise once released from the protein backbone. The lag phase of *R. gnavus* ATCC 29149 growth on pPGM was 4 hours, while the lag phase on PGM was 10 hours. It should be noted that while *R. gnavus* ATCC 29149 grew well on pPGM, PGM, and PGM-released mucin glycans, no growth was observed on pPGM mucin glycans which needs further investigation, this condition was not pursued in the rest of the experiments. *R. gnavus* E1 did not grow on either pPGM and PGM, as expected from previous reports (Crost et al., 2013). However, it did show some growth on mucin glycans released from PGM, albeit to low levels (Figure 4.2).

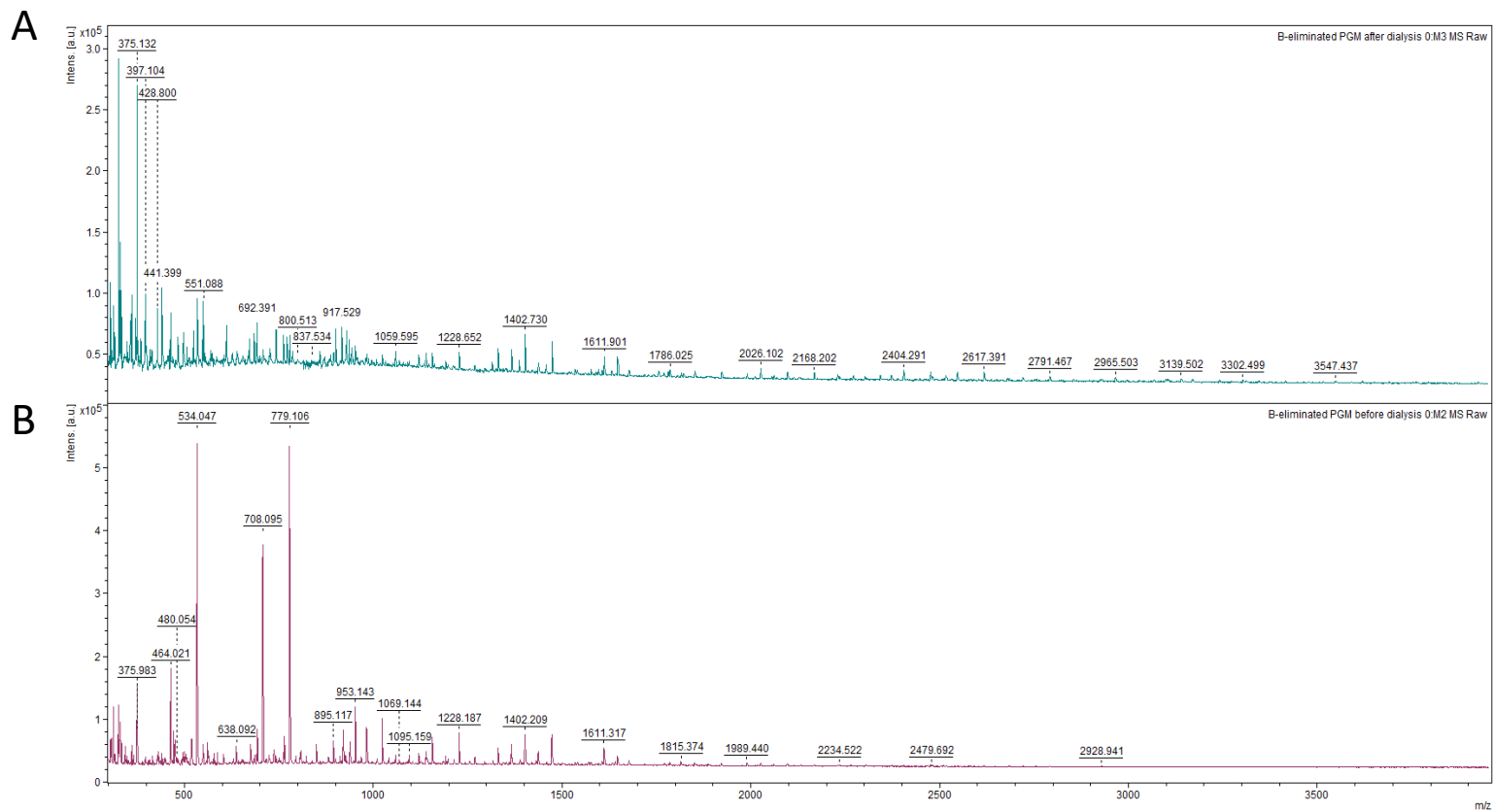


Figure 4.1. MALDI-TOF MS analysis of  $\beta$ -eliminated mucin glycans from PGM. Mucin glycans after (A) and before (B) dialysis by 1 kDa MWCO membrane. Glycans with masses lower than 1 kDa were significantly reduced.

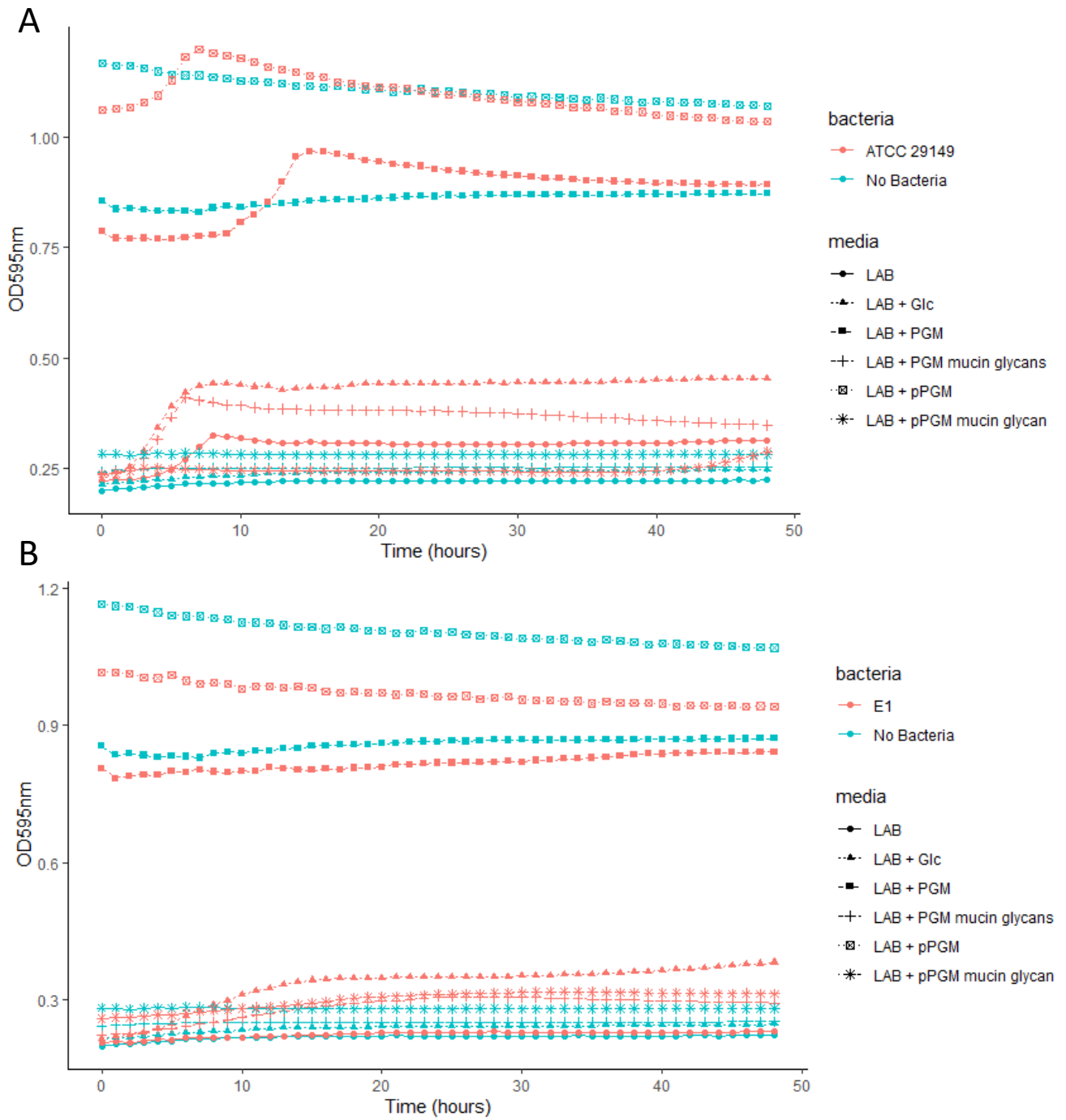


Figure 4.2. Growth of *R. gnavus* ATCC 29149 and E1 on mucin and mucin-released glycans. The culture in LAB supplemented with PGM, PGM mucin glycans, pPGM, pPGM mucin glycans was monitored by spectrophotometry over 48 hours. Averages of three replicates are shown.

For transcriptomics analysis, *R. gnavus* growth was carried out on 1% mucin glycans or 0.5% glucose (Figure 4.3). RNA was then extracted in the late exponential phase, which corresponded to 6 hours for growth on glucose and 7 hours for growth on mucin glycans (Figure 4.3). RNAseq requires high quality (RNA Integrity Number  $\geq 6$ ) and high quantity ( $\geq 50$  ng/ $\mu$ l) of RNA, quality and quantity were assessed and found to be of sufficient quality and quantity (Supplemental Table S4) and the RNA was sequenced using an Illumina HiSeq platform by Novogene.

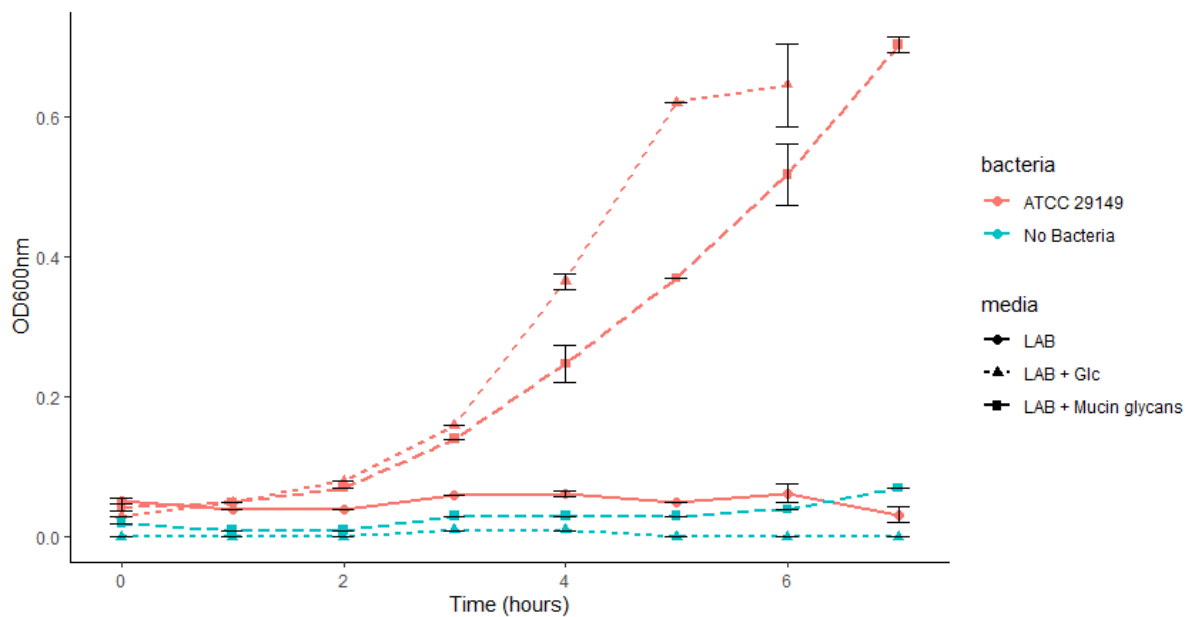


Figure 4.3. Growth of *R. gnavus* ATCC 29149 in LAB supplemented with 0.5% glucose or 1% mucin glycans. OD was measured hourly of a single repeat. OD was measured for four repeats at T=0, 4, 6, and 7 hours. Error bars indicate standard deviation.

## 4.2.2 Transcriptomics analysis of *R. gnavus* ATCC 29149

### 4.2.2.1 *R. gnavus* ATCC 29149 gene expression during growth on mucin glycans

RNAseq data was analysed using DESeq2, which generated a list of significantly differentially expressed genes during *R. gnavus* growth on mucin glycan compared to growth on glucose. This list was generated using a Log2 fold change of 1 and -1 as threshold for the change in gene expression, and a p-adjusted value threshold of 0.05 for the significance. Using these thresholds, 717 genes were found to be significantly upregulated, and 652 genes were significantly downregulated in growth on mucin glycans versus growth on glucose (Figure 4.4). The genome of *R. gnavus* ATCC 29149 contains 60 GH encoding genes; during mucin glycan growth 22 were significantly upregulated while 3 were

significantly downregulated (Table 4.1). Among GHs involved in mucin glycan degradation, genes encoding GH2s, GH29s, GH33, GH36, GH95, and GH98 enzymes were significantly upregulated. *R. gnavus* ATCC 29149 GH33, and GH29 and GH95 fucosidases have been shown to be upregulated during growth on mucin (pPGM) and to be active on mucin glycan epitopes, but have not been tested on mucin (Crosth et al., 2016; Owen et al., 2017; H. Wu, Rebello, et al., 2021). GH2 galactosidases have been shown to have mucin glycan degrading activity in *A. muciniphila* ATCC BAA-835 (Kosciow & Deppenmeier, 2020). *RgGH98* was shown to be active on mucin glycans in chapter 3 of this thesis (H. Wu, Crosth, et al., 2021). Other significantly upregulated GHs were GH3, 4 GH13s, GH23, GH28, 2 GH31s, 2 GH32s, GH73, GH77, GH88, and GH112 (Table 4.1).

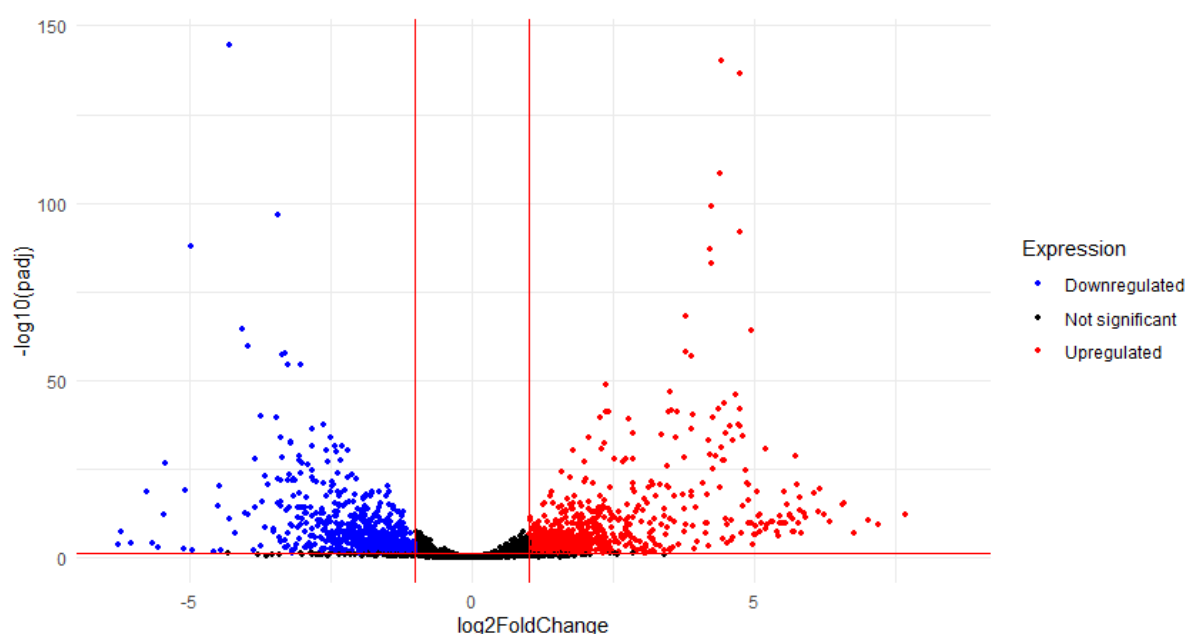


Figure 4.4. Volcano plot of significantly up- and downregulated genes during *R. gnavus* ATCC 29149 growth on mucin glycans versus Glc. Red dots indicate a significantly upregulated gene. Blue dots indicate significantly downregulated genes. Horizontal line is the threshold denoting significance (p-adjusted value < 0.05). Vertical lines denote Log2 fold change cut-offs (-1 and 1).

Table 4.1. Gene expression analysis of GH genes during *R. gnavus* ATCC 29149 growth on mucin glycan vs Glc.

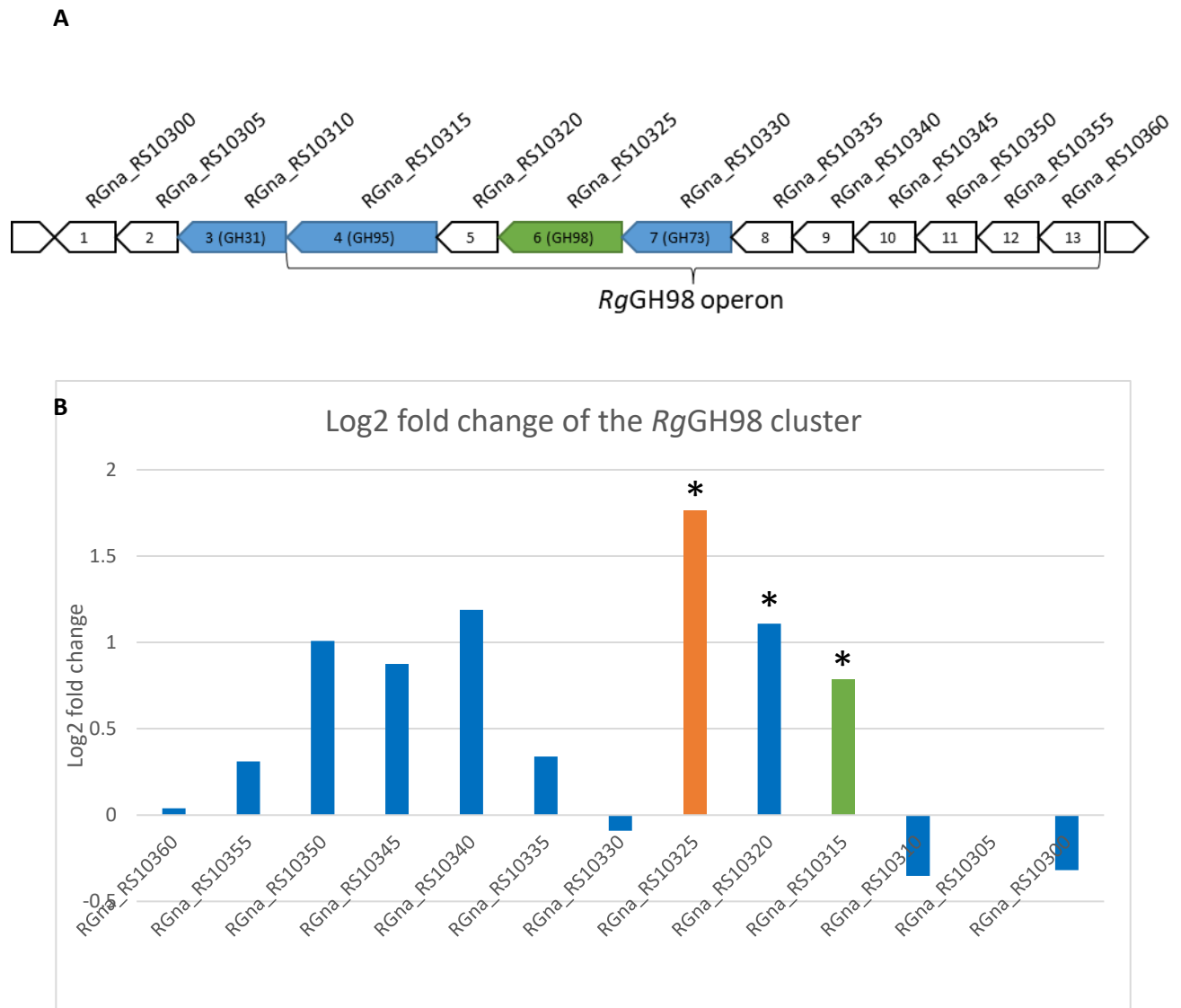
GH family	Significantly upregulated	Non-significant/below Log2 fold threshold	Significantly downregulated	Missing data
GH1		3		2
GH2	2	2	1	1
GH3	1	3		
GH5				2
GH13	4	4		1
GH18		1	1	

GH23	1	2		1
GH25		1		1
GH28	1			
GH29	1			1*
GH31	2	1		
GH32	2	1		
GH33	1			
GH36	1	1		
GH42		1		
GH73	1	1		
GH77	1			
GH78		1		
GH88	1			
GH95	1	2		
GH98	1			
GH105		1	1	
GH112	1			
GH113		1		
Total	22	26	3	9

Thresholds for significance: Log2 fold change >1 or <-1; p-adj < 0.05. \*This GH29 gene has a high log2 fold change, but a p-adj of NA due to an outlier in the data.

#### 4.2.2.2 Dedicated mucin glycan degrading operons are upregulated during *R. gnavus* growth on mucin glycans

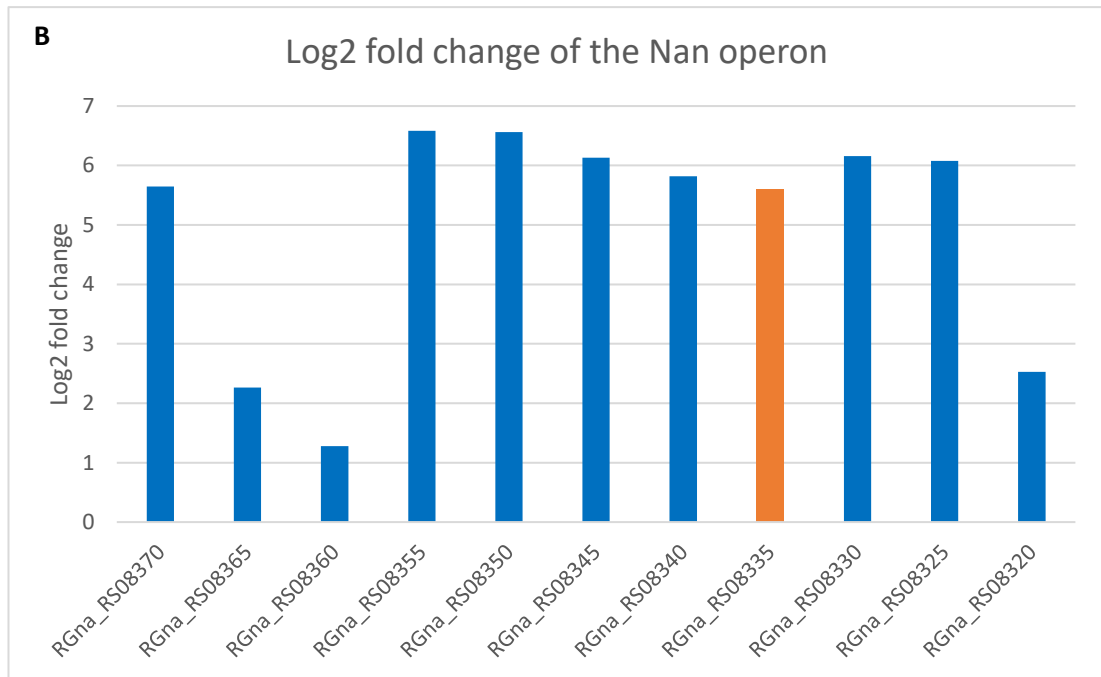
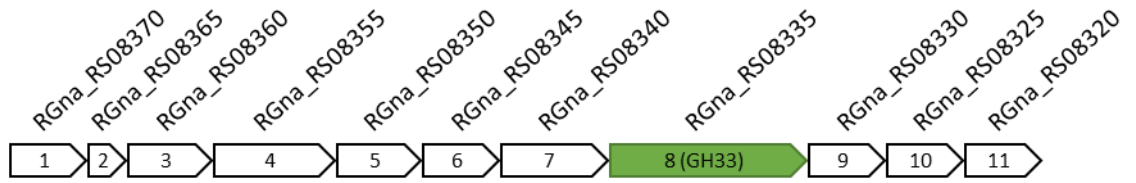
From the *RgGH98* operon, only *RGna\_RS10315* (GH95), *RGna\_RS10320*, and *RGna\_RS10325* (*RgGH98*) were shown to be significantly upregulated, but *RGna\_RS10315* (GH95) transcription was not above the Log2 threshold (Figure 4.5). The normalised level of transcription data showed that this cluster was also expressed during growth of *R. gnavus* ATCC 29149 on glucose. In contrast, the Nan cluster (*RGna\_RS08320* to *RGna\_RS08370*), which is dedicated to 2,7-anhydro-Neu5Ac metabolism (Bell et al., 2019; Tailford, Owen, et al., 2015), was found to be highly upregulated across the entire operon (Figure 4.6). Some genes in this cluster were among the most highly upregulated of all genes. Furthermore, a GH29 (*RGna\_RS05890*) and a GH95 (*RGna\_RS14395*) gene, both part of separate expected operons including a sugar ABC transporter, were found to be upregulated. Interestingly, we found that an cluster of 29 genes belonging to a prophage was highly induced on *R. gnavus* grown in mucin as compared to glucose growth. BLAST analysis indicates that these genes are part of the Myoviridae bacteriophage family. This is an unexpected development which will be worth pursuing in the future (Figure 4.7 and Table 4.2).



**Figure 4.5.** Expression of *RgGH98* gene cluster during growth on mucin glycans. **(A)** Schematic of *RgGH98* gene cluster. **(B)** log<sub>2</sub> fold change of *RgGH98* gene cluster during growth of *R. gnavus* ATCC 29149 on mucin glycans compared to growth on Glc. *RgGH98* operon is RGna\_RS10360- RGna\_RS10315. RGna\_RS10310- RGna\_RS10300 are not part of the operon. *RgGH98* is shown in orange. The GH95 in the operon is shown in green. \* = p-adj < 0.05.



**A**



**Figure 4.6.** Expression of *nan* cluster during growth of *R. gnavus* ATCC 29149 on mucin glycans. **(A)** Schematic of *RgGH33* gene cluster. **(B)** log<sub>2</sub> fold change of *RgGH33* gene cluster during growth of *R. gnavus* ATCC 29149 on mucin glycans compared to growth on Glc. *RgGH33* is shown in orange. All genes in the operon were significantly upregulated: Log<sub>2</sub> fold change >1 and p-adj < 0.05.

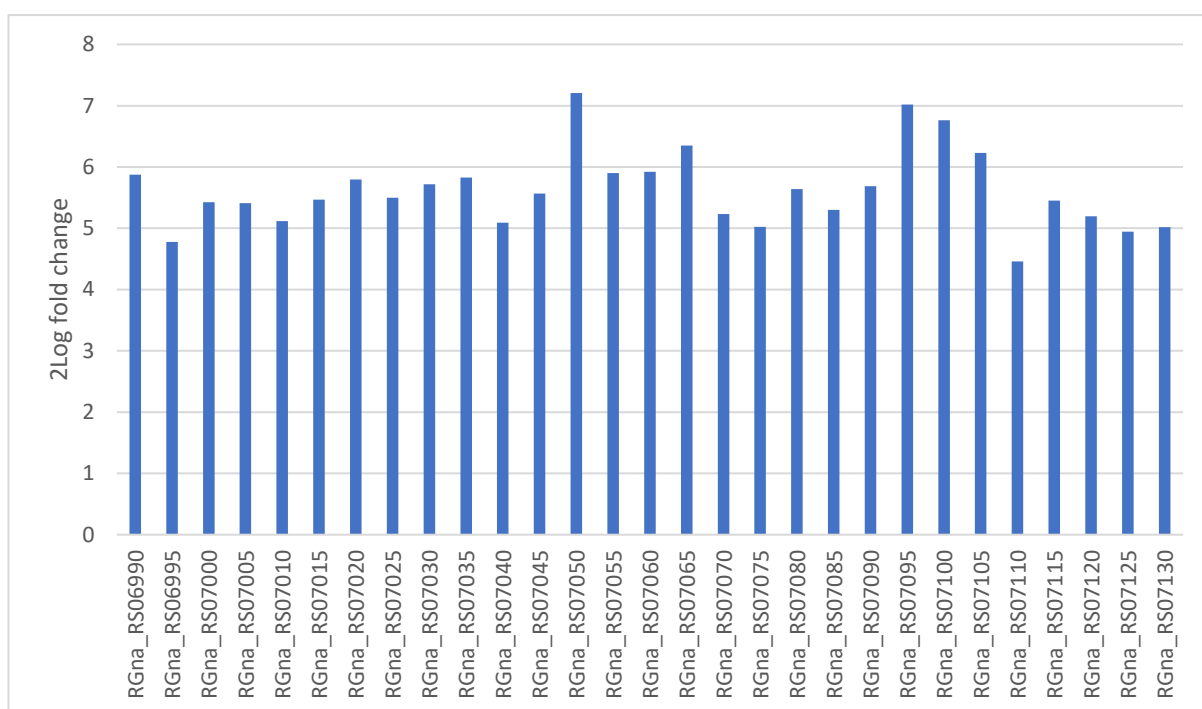


Figure 4.7. Expression of prophage cluster during growth of *R. gnnavus* ATCC 29149 on mucin glycans. Log2 fold change of prophage gene cluster during growth of *R. gnnavus* ATCC 29149 on mucin glycans compared to growth on Glc. All 29 genes in the cluster were significantly upregulated: Log2 fold change >1 and p-adj < 0.05.

Table 4.2. Predicted gene functions of *R. gnnavus* ATCC 29149 prophage genes RGna\_RS06995 to RGna\_RS07120.

RGna_RS06995	NCBI annotation
RGna_RS07000	MULTISPECIES: XkdX family protein
RGna_RS07005	MULTISPECIES: hypothetical protein
RGna_RS07010	RNA-directed DNA polymerase
RGna_RS07015	phage tail protein
RGna_RS07020	phage tail protein
RGna_RS07025	baseplate J/gp47 family protein
RGna_RS07030	MULTISPECIES: hypothetical protein
RGna_RS07035	phage tail protein
RGna_RS07040	hypothetical protein
RGna_RS07045	hypothetical protein
RGna_RS07050	tail protein X
RGna_RS07055	phage tail tape measure protein
RGna_RS07060	hypothetical protein
RGna_RS07065	phage major tail tube protein
RGna_RS07070	phage tail sheath family protein
RGna_RS07075	hypothetical protein
RGna_RS07080	hypothetical protein

RGna_RS07085	phage tail protein
RGna_RS07090	MULTISPECIES: hypothetical protein
RGna_RS07095	MULTISPECIES: hypothetical protein
RGna_RS07100	MULTISPECIES: DUF2190 family protein
RGna_RS07105	hypothetical protein
RGna_RS07110	HK97 family phage prohead protease
RGna_RS07115	phage portal protein
RGna_RS07120	hypothetical protein
RGna_RS07125	phage terminase large subunit family protein
RGna_RS07130	MULTISPECIES: hypothetical protein
RGna_RS07135	hypothetical protein

NCBI annotation is the automatically predicted function from the NCBI database.

#### 4.2.2.3 Selection of GH targets for functional characterisation

In total 11 GH gene targets were chosen for functional characterisation of the predicted enzymes (Supplemental Table S5). Each amino acid sequence was BLASTed against the Swiss-prot database to determine closely related fully annotated proteins (Table 4.3). RGna\_RS14450, RGna\_RS07225, RGna\_RS14905, RGna\_RS06060, RGna\_RS08510, RGna\_RS14690, RGna\_RS07930, and RGna\_RS02745 were chosen as highly significantly induced when *R. gnavus* ATCC 29149 was grown on mucin glycans. These are predicted to code for GH32, GH23, GH31, GH13 and GH2 enzymes (see Table 4.3 for assignments). RGna\_RS14450, RGna\_RS14905, RGna\_RS06060, RGna\_RS08510, RGna\_RS14690, and RGna\_RS07930 are part of possible operons including other GHs and sugar transporter genes. We also selected RGna\_RS05890, RGna\_RS04290, and RGna\_RS14395 encoding fucosidases herein named *RgGH29-1*, *RgGH29-2* and *RgGH95* previously characterised on fucosylated oligosaccharides but not on mucin (H. Wu, Rebello, et al., 2021). Lastly RGna\_RS10330 coding for a predicted GH73 was selected as this gene is part of the *RgGH98* operon (see 3.1) and was found to be significantly upregulated during *R. gnavus* ATCC 29149 growth on pPGM (Crost et al., 2016), although not during growth on mucin glycans.

Table 4.3. Overview of selected *R. gnavus* ATCC 29149 GHs.

NZ_CP027002.1 locus tag	GH family	NCBI annotation	Swiss-Prot predicted function	Log2 fold change	P-adjusted value	Predicted localisation
RGna_RS05890	GH29	$\alpha$ -L-fucosidase	$\alpha$ -L-fucosidase	3.76	6.07E-29	Cell wall
RGna_RS04290	GH29	$\alpha$ -L-fucosidase	$\alpha$ -L-fucosidase	4.36	NA	Cytoplasmic
RGna_RS14395	GH95	Glycoside hydrolase N-terminal domain-containing protein	Fibronectin type III domain protein	3.77	1.98E-14	Cell wall
RGna_RS14450	GH32	Sucrose-6-phosphate hydrolase	Sucrose-6-phosphate hydrolase	3.41	2.49E-14	Cytoplasmic
RGna_RS07225	GH23	Transglycosylase SLT domain-containing protein	Membrane-bound lytic murein transglycosylase F	3.24	0.024088	<b>Extracellular</b>
RGna_RS14905	GH31	DUF4968 domain-containing protein	$\alpha$ -glucosidase	3.25	5.14E-18	Cytoplasmic
RGna_RS06060	GH13	Glycoside hydrolase family 13 protein	Cyclomaltodextrinase/Neopullulanase	3.21	2.35E-21	Cytoplasmic
RGna_RS08510	GH13	Type I pullulanase	Pullulanase	2.03	6.07E-08	Cytoplasmic
RGna_RS14690	GH88	Glycoside hydrolase family 88 protein	Unsaturated chondroitin disaccharide hydrolase	1.24	0.000493	Cytoplasmic
RGna_RS10330	GH73	Glucosaminidase domain-containing protein	Peptidoglycan hydrolase FlgJ	-0.09	0.851899013	<b>Extracellular</b>

RGna_RS07930	GH2	Hypothetical protein	$\beta$ -galactosidase or $\beta$ -glucuronidase	1.23	1.89E-05	Cytoplasmic
RGna_RS02745	GH2	DUF4981 domain- containing protein	$\beta$ -galactosidase	1.60	0.0000658	Cytoplasmic

Log2 fold change and p-adjusted value are taken from RNAseq results. NCBI annotation is the automatically predicted function from the NCBI database. Swiss-Prot is the function predicted by Swiss-Prot BLAST analysis, using the Uniprot standard settings and bacterial results with high identity were used.

It should be noted that while RGna\_RS04290 (coding for *RgGH29-2*) showed high upregulation on mucin glycans versus glucose, the DESeq2 analysis assigned a p-adjusted value (p-adj) of NA. This was due to variations between triplicates, with one of them being so high that the DESeq2 analysis assigned it as outlier and gave it a p-adj of NA. To verify that this gene was indeed highly upregulated during growth on mucin glycans, a targeted RT-qPCR analysis was performed using the same RNA sample used for the RNAseq and primers specific for RGna\_RS04290. This confirmed that all replicates were highly upregulated compared to *R. gnavus* grown in Glc (Table 4.4) and that the first replicate was much higher upregulated, mirroring the results obtained by the RNAseq.

Table 4.4. Fold change increase of RGna\_RS04290 (*RgGH29-2*) gene during growth on mucin glycans vs Glc.

	X-fold-change	
	Replica 1	Replica 2
<b>Sample 1</b>	539.5	689.1
<b>Sample 2</b>	4.2	5.3
<b>Sample 3</b>	4.3	5.2
<b>Sample 4</b>	6.6	3.6

X-fold change was determined of four samples; samples 1-3 were used for RNAseq. RT-qPCR was performed in duplicate. X-fold changes are shown based on differences in expression calculated from changes in  $2^{-\Delta\Delta Ct}$  after normalising against the *GyrB* housekeeping gene.

### 4.2.3 Functional characterisation of selected *R. gnavus* ATCC 29149 GHs

#### 4.2.3.1 Heterologous expression and purification of recombinant *RgGHs*

Selected GH-encoding genes were amplified by PCR and cloned into the pET-28a plasmid containing an N terminal His-tag following transformation in *E. coli* DH5 $\alpha$  cells, with the exception of the GH29 and GH95 fucosidases, of which previously produced plasmids were used (H. Wu, Rebello, et al., 2021). Using specific primers for selected GH-encoding genes (see 2.4.1), PCR products were amplified from *R. gnavus* ATCC 29149 genomic DNA at the

correct size (Supplemental Figure S1). No recombinant plasmid could be obtained for RGna\_RS14450 (GH32), RGna\_RS14905 (GH31), and RGna\_RS02745 (GH2-2). All other genes were successfully cloned as shown on agarose gel where bands corresponding to the insert were observed following digestion of the plasmid with restriction enzymes targeting flanking regions of the inserts (Supplemental Figure S2). The recombinant plasmids were then used to transform *E. coli* TunerDE3pLacI cells for expression, and recombinant His-tag proteins *RgGH2-1*, *RgGH13-1*, *RgGH13-2*, *RgGH23*, *RgGH73*, *RgGH88*, *RgGH29-1*, *RgGH29-2* and *RgGH95* were purified to homogeneity using immobilised metal affinity chromatography (IMAC), and further purified by gel filtration. Protein production and purification was monitored by SDS-PAGE (Supplemental Figure S3). *RgGH13-1*, *RgGH13-2*, *RgGH73*, *RgGH23*, *RgGH2-1*, *RgGH95*, *RgGH29-1*, *RgGH29-2*, and *RgGH88* run at the calculated size of 69.85 kDa, 73.26 kDa, 48.07 kDa, 27.83 kDa, 84.15 kDa, 103.62 kDa, 84.37 kDa, 49.72 kDa, 46.53 kDa, respectively. In summary, nine of the selected GHs were successfully cloned, as confirmed by sequencing of the recombinant plasmids, and expressed as shown by SDS-PAGE of the purified protein.

#### **4.2.3.2 Enzymatic activity of *R. gnavus* ATCC 29149 GH29 and GH95 fucosidases**

The enzymatic activity of purified recombinant *RgGH29-1*, *RgGH29-2*, and *RgGH95* was first tested on 3-fucosyllactose (3FL). *RgGH29-2* and *RgGH95* were previously reported to be active on 3FL as was E1\_10125 from *R. gnavus* E1 strain and this protein shares 94.5% sequence similarity with *RgGH29-1* (H. Wu, Rebello, et al., 2021). HPAEC-PAD analysis showed 100% release of fucose from 3FL after 24 hours for all three enzymes (Table 4.5). The recombinant enzymes were then tested using mucin glycans (released from PGM or pPGM) as substrate. Following 24 hours incubation with *RgGH29-1* and *RgGH95*, ~100  $\mu$ M fucose was released from PGM mucin glycans, while 55  $\mu$ M was released with *RgGH29-2*. *RgGH29-2* activity was half that of the other fucosidases on mucin glycans from both PGM and pPGM. The amount of fucose released from pPGM mucin glycans was about half compared to PGM mucin glycans. In summary, the three fucosidases tested were able to release fucose from mucin glycans, indicating a role of these fucosidases in mucin glycan metabolism.

Table 4.5. Enzymatic activity of *R. gnavus* ATCC 29149 fucosidases.

Enzyme	Released fucose ( $\mu\text{M}$ ) on		
	3FL	PGM mucin glycans	pPGM mucin glycan
<b><i>RgGH29-1</i></b>	101.3 $\pm$ 0.7	97.7 $\pm$ 2.3	45.1 $\pm$ 2.3
<b><i>RgGH29-2</i></b>	103.0 $\pm$ 1.5	55.0 $\pm$ 1.4	27.5 $\pm$ 4.6
<b><i>RgGH95</i></b>	99.7 $\pm$ 0.4	99.1 $\pm$ 5.0	44.8 $\pm$ 1.0

Concentration of released fucose shown in  $\mu\text{M}$  following 24 hours incubation with *R. gnavus* ATCC 29149 fucosidases. Released concentration is average of three replicates. Substrate concentration of 3FL was 100  $\mu\text{M}$ ; substrate concentration of PGM and pPGM mucin glycans was 0.1 mg/ml.  $\pm$  = standard deviation.

#### 4.2.3.3 Enzymatic activity of *R. gnavus* ATCC 29149 GH13 enzymes

The GH13 family is a large family consisting of 45 subfamilies with broad substrate specificity towards a range of substrates including starch and pullulan (Janeček et al., 2014; Stam et al., 2006; www.cazy.org). Based on their sequence identity, *RgGH13-1* and *RgGH13-2* are predicted to be a neopullulanase and type I pullulanase, respectively.

When the *RgGH13* enzymes, *RgGH13-1* or *RgGH13-2*, were tested against mucin glycans released from PGM or pPGM, HPAEC results showed the appearance of a peak at 5 min retention time which was not present in the control reactions with the *RgGH13* enzymes in the absence of substrate or with the substrate in the absence of enzymes (Figure 4.8). This indicates that the additional peak was not due to spontaneous degradation of the substrate over time or to the presence of a contaminant in the enzyme or substrate preparation. In addition, this peak was not observed in the experiment using  $\alpha$ -amylase from *Bacillus subtilis* subsp. *subtilis* 168 as a control performed in the same conditions, ruling out a possible contamination of the substrate with glucose polymers (Figure 4.8). These reactions were further analysed using MALDI-TOF MS and NMR. However, these analyses were unable to identify a potential released glycan from the reaction of GH13 with mucin glycans (Figure 4.9 and Supplemental Figure S4).

*RgGH13-1* and *RgGH13-2* were then further characterised and shown to be active on their predicted substrates (see chapter 5).

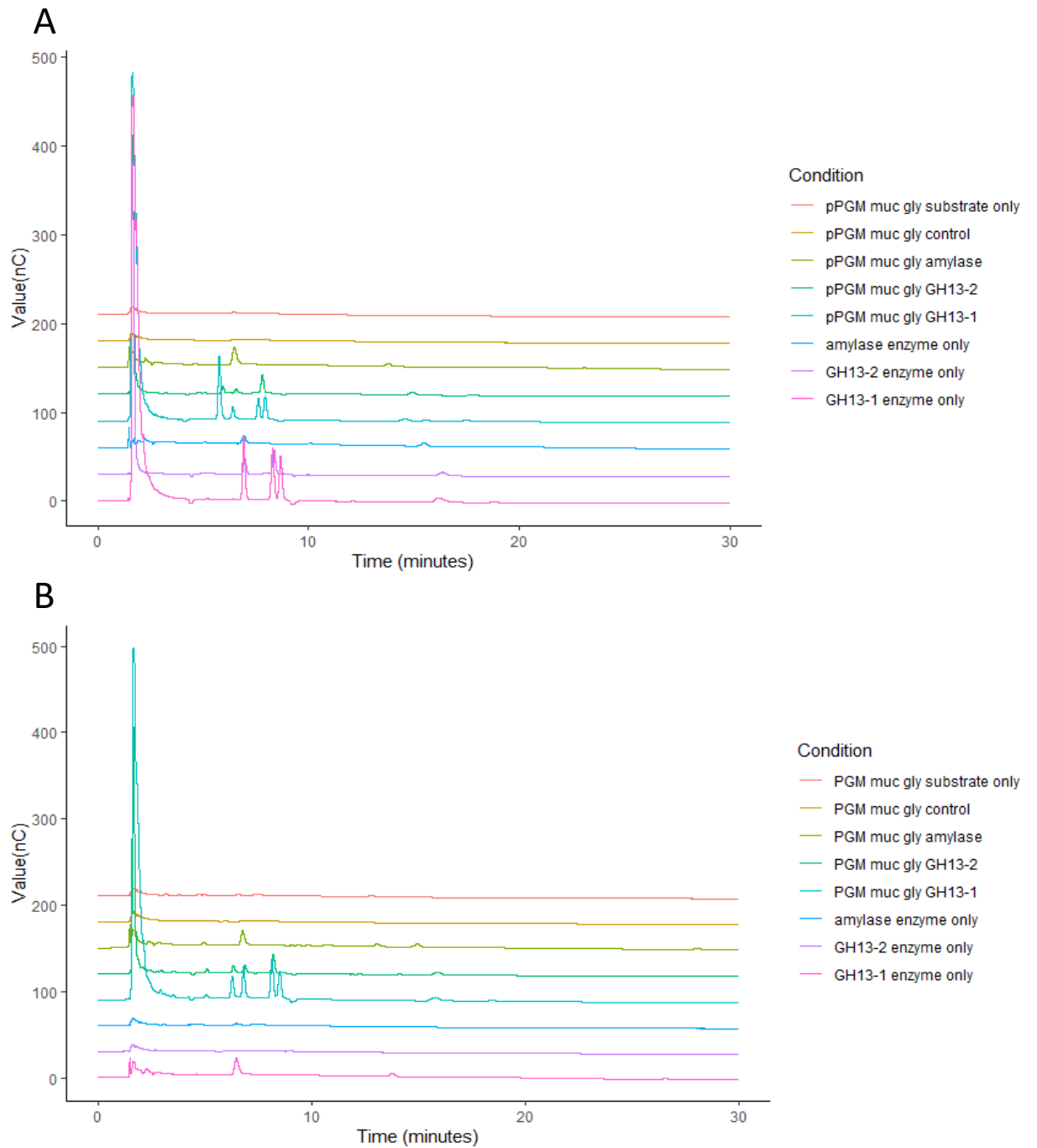


Figure 4.8. HPAEC-PAD analysis of *RgGH13* enzymatic reaction on mucin glycans. **(A)** Analysis of the products of the reaction with pPGM-released mucin glycans **(B)** Analysis of the products of the reaction with PGM-released mucin glycans. *RgGH13-1* and *RgGH13-2* were used at 10  $\mu\text{M}$  concentration and the reaction performed for 24 hours.  $\alpha$ -Amylase from *Bacillus subtilis* subsp. *subtilis* 168 (10  $\mu\text{M}$   $\mu\text{M}$ ) was used as a control.



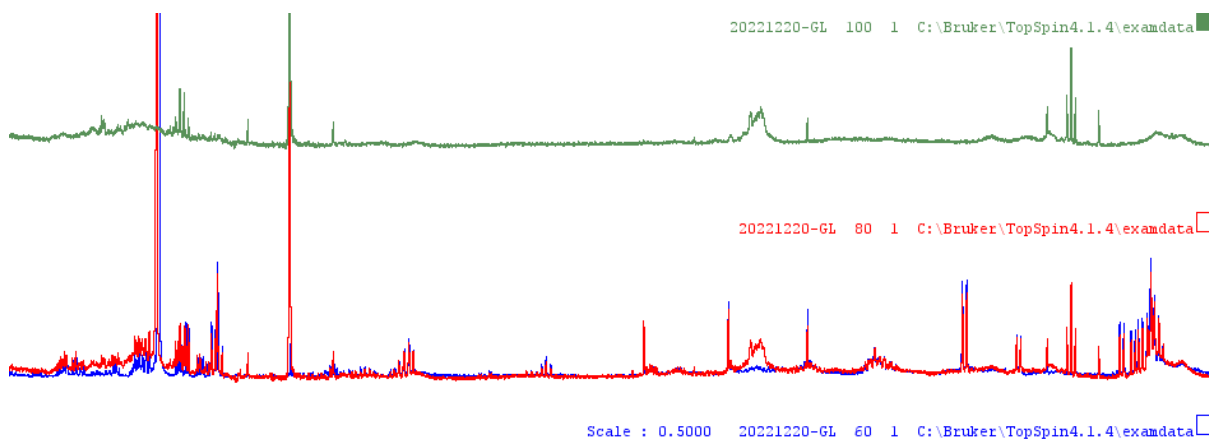


Figure 4.9. NMR analysis of *RgGH13-2* enzymatic reaction with mucin glycans. Comparison of NMR spectra of *RgGH13-2* (blue), mucin glycans (green), and mucin glycans (1 mg/ml) incubated with 10  $\mu$ M *RgGH13-2* (red).

#### 4.2.3.4 *R. gnavus* ATCC 29149 GHs with no demonstrated functional activity

Based on sequence identity, *RgGH23* (RGna\_RS07225) and *RgGH73* (RGna\_RS10330) are predicted to be peptidoglycan hydrolysing enzymes (Table 4.3). Peptidoglycan is a polysaccharide consisting of GlcNAc $\beta$ 1-4*N*-acetylmuramic acid (MurNAc) disaccharide units with a short amino acid chain attached to each disaccharide unit. Using TLC, no release of GlcNAc was observed in the enzymatic reaction of *RgGH23* or *RgGH73* with peptidoglycan from *Streptomyces venezuelae* cell wall (Supplemental Figure S5). No released products were detected by HPAEC using mucin glycans as substrate (Supplemental Figure S6 A and B).

*RgGH88* (RGna\_RS14690) is predicted to be a chondroitin disaccharide hydrolase (Table 4.3). Here the recombinant enzyme was tested against non-sulphated chondroitin, a disaccharide consisting of glucuronic acid and GlcNAc. No activity was detected using HPAEC on chondroitin disaccharide or on mucin glycans as substrate (Supplemental Figures S6C and S7A).

*RgGH2-1* (RGna\_RS07930) is predicted to be a  $\beta$ -galactosidase (Table 4.3). The enzymatic activity of recombinant *RgGH2-1* was tested against lactose (Lac; Gal $\beta$ 1-4Glc), lacto-*N*-tetraose (LNT; Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc), lacto-*N*-neotetraose (LNnT; Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc), and lacto-*N*-triose II (GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc). LNT and LNnT contain terminal Gal in a  $\beta$ 1-3 and  $\beta$ 1-4 conformation, attached to a GlcNAc, a glycosidic linkage which is also found in mucin glycans. A galactosidase active on these substrates would result in the release of Gal and lacto-*N*-triose II. However, no peak corresponding to

Gal or lacto-N-triose II was detected by HPAEC in the conditions tested (Supplemental Figure S7B). *RgGH2-1* was tested against mucin glycans and no released compound was detected by HPAEC after an incubation of 24 hours (Supplemental Figure S6D).

The absence of functional activity of these recombinant enzymes may be due to issues with protein folding of the recombinant enzymes, to the enzymatic conditions tested (e.g. pH, temperature, enzyme/substrate concentrations), or that the tested compound was not a substrate for the enzyme.

### 4.3 Discussion

RNAseq has been used successfully in the past to identify genes involved in both dietary glycan (Briliūtė et al., 2019) and mucin glycan degradation by *B. thetaiotaomicron* ATCC 29148 (Pudlo et al., 2015). Here, we used this approach to identify the complement of GHs involved in *R. gnavus* ATCC 29149 mucin glycan degradation. *R. gnavus* ATCC 29149 showed faster growth and to a higher density on mucin glycans compared to pPGM, suggesting that the chemical release of the mucin glycan chains from the peptidic mucin backbone made the mucin glycans more easily accessible to the bacteria. Full transcriptome analysis comparing growth of *A. muciniphila* ATCC BAA-835, a beneficial mucin degrader (Cani et al., 2022), on mucin and glucose has been previously performed, revealing a metabolic model for mucin degradation for this strain (Ottman et al., 2017). On further analysis of the data, half of *A. muciniphila* fucosidases and half of its sialidases were found to be significantly upregulated during mucin glycan growth (not shown), suggesting that not all fucosidases and sialidases are involved in mucin glycan metabolism. This was recently validated experimentally (Shuoker et al., 2023).

Here, we showed that during *R. gnavus* growth on mucin glycan, 22 GHs were significantly upregulated and three significantly downregulated as compared to growth on glucose. The genes encoding predicted galactosidases (2 GH2, 1 GH98), GalNAcase (GH36), fucosidases (2 GH29, 1 GH95) and sialidase (GH33) were among the significantly upregulated GHs when *R. gnavus* ATCC 29149 was grown on mucin glycans as compared to glucose. Previously, transcriptomic analysis by RNAseq of *R. gnavus* ATCC 29149 grown on pPGM (rather than released mucin glycans) showed 22 GHs to be significantly upregulated compared to

growth on glucose (Crost et al., 2016). Among these, 13 were found to be significantly upregulated in both RNAseq datasets, including *RgGH13-2*, *RgGH33*, *RgGH29-2*, *RgGH32*, *RgGH88*, *RgGH2-1*, and *RgGH2-2*. Surprisingly, *RgGH13-2* was found to be the highest upregulated GH enzyme on pPGM, in line with our results showing that *RgGH13-1* and *RgGH13-2* encoding genes were found highly expressed during *R. gnavus* ATCC 29149 growth on mucin glycans. Both *RgGH13-1* and *RgGH13-2* were found to be active on pullulan (see chapter 5). HPAEC-PAD suggested that the enzymes may be active towards mucin glycans (from pPGM and PGM) and mucin (both pPGM and PGM). However, it was not possible to identify the product of the reaction using MALDI-TOF and NMR analysis. GH13 enzymes are characterised by their activity on glucose polysaccharides like starch and pullulan and therefore GH13s are not expected to be active on mucin glycans. Further research is needed to identify the released compound from mucin glycans by *RgGH13-1* and *RgGH13-2*.

Interestingly, RGna\_RS10330 (encoding for a predicted GH73), which is part of the *RgGH98* operon was upregulated when *R. gnavus* ATCC 29149 was grown on pPGM (Crost et al., 2016), but not on mucin glycans (this work), suggesting variation in the expression of this operon during mucin glycan metabolism, although care should be taken in comparing these datasets as the gene expression level on glucose used as control in both transcriptomics analyses was not always similar. Unfortunately, it has not been possible to characterise the enzymatic activity of the recombinant enzyme.

We showed that fucosidases *RgGH29-1*, *RgGH29-2*, and *RgGH95* were active on mucin, releasing fucose. *RgGH29-2* and *RgGH95* were previously characterised, showing preference for  $\alpha$ 1-2 linkages (H. Wu, Rebello, et al., 2021). *RgGH29-1* has 94.5% sequence identity to E1\_10125, a fucosidase from *R. gnavus* E1 strain, which was found to be highly specific for  $\alpha$ 1-3 and  $\alpha$ 1-4 linked fucose and able to release fucose from sialylated lewis A and sialylated lewis X structures (H. Wu, Rebello, et al., 2021). It is therefore likely that *RgGH29-1* has similar substrate specificity. *RgGH29-1* was 2-fold more active on mucin glycans compared to *RgGH29-2*, which may be due to its ability to accommodate sialylated epitopes although this would need to be confirmed. It is of note that while *RgGH29-2* was also found to be significantly upregulated when the bacteria were grown on pPGM (Crost et al., 2016), *RgGH29-1* and *RgGH95* were significantly downregulated and two other GH95

fucosidases were significantly upregulated (Croft et al., 2016). Taken together these results may suggest that *RgGH29s* and *RgGH95s* share some level of redundancies but with different substrate specificities depending on accessibility of the mucin glycan chains.

Interestingly, RNAseq analysis of *R. gnavus* ATCC 29149 grown on mucin glycans also showed the induction of prophage genes in *R. gnavus* ATCC 29149 grown on mucin glycans. Prophage induction in gut bacteria has been reported as a response to stress conditions like antibiotics, temperature, and oxidative stress (Hu et al., 2021). Moreover, phage production was promoted in *Lactobacillus reuteri* 6475 by dietary fructose, galactose, and xylose (Oh et al., 2019). Prophages in *L. reuteri* 6475 were also induced by short-chain fatty acids (Oh et al., 2019), a product of bacterial glycan metabolism, including mucin glycans (Blaak et al., 2020). Several bacteriophages able to infect *R. gnavus* have been described, interestingly these phages had no negative effect on the growth on *R. gnavus* (Buttimer et al., 2023), suggesting a potential beneficial role of these phages. It would be interesting for future research to investigate the role of monosaccharides released from mucin glycans and mucin glycan metabolic products on prophage induction. The presence of genes in sialic acid utilisation encoded by prophages has been reported to confer a growth advantage to enterohemorrhagic *E. coli* (EHEC) by outcompeting other *E. coli* strains in Neu5,9Ac2 rich environments, such as mucus in animal and human gut (Feuerbaum et al., 2018; Saile et al., 2018). Another strategy by which prophages may contribute to the gut homeostasis and pathogenesis is through their capacity to promote mucin production (Nishio et al., 2021). The induction of prophage genes when *R. gnavus* is grown on mucin glycans may provide a competitive advantage to *R. gnavus* strains inhabiting the mucus niche.

## Chapter 5

### Structure and function of *R. gnavus* GH13 pullulanases

## 5. Structure and function of *R. gnavus* GH13 pullulanases

### 5.1 Introduction

In addition to host glycans, a wide variety of dietary glycans are available to the gut microbiome. While humans only encode eight digestive GH families, the human gut microbiome encodes extensive numbers of GHs (Kaoutari et al., 2013), enabling the degradation of complex carbohydrates reaching the colon undigested. The released oligosaccharides can be used for consumption and metabolism by gut microbes (Hamaker & Tuncil, 2014; Ndeh et al., 2017; Ndeh & Gilbert, 2018).

A major component of dietary glycans is starch, which is a main dietary source of carbohydrates, consisting of  $\alpha$ 1-4 linked glucose units with  $\alpha$ 1-6 linked sidechains (Tian & Sun, 2020). Non-resistant starch can be broken down by humans, while resistant starch reaches the colon mostly intact (Svihus & Hervik, 2016). Only specialised gut bacteria are able to break down resistant starch including *Ruminococcus bromii* and *Bifidobacterium adolescentis* (Ze et al., 2012). Another glucose-based polysaccharide is pullulan, which consists of  $\alpha$ 1-6 linked maltotriose (Glc $\alpha$ 1-4Glc $\alpha$ 1-4Glc) units. Pullulan is produced by certain fungi, most importantly *Aureobasidium pullulans* (Prasongsuk et al., 2018), which is widespread in soil and plant material and has been used as an effective biocontrol agent combatting post-harvest pathogens in fruits (Prasongsuk et al., 2018). Moreover, pullulan is increasingly being used as food additive and as a medicine carrier (K. C. Cheng et al., 2011; Prasongsuk et al., 2018). Both starch and pullulan are substrates for some of the enzymes found in the GH13 family (Stam et al., 2006). This is a large GH family with 45 subfamilies (Janeček et al., 2014) ([www.cazy.org](http://www.cazy.org)); subfamilies are subgroups found within a family that share a more recent ancestor and that are usually more uniform in molecular function ([www.cazy.org](http://www.cazy.org)). The GH13 family is the most abundant GH family in gut bacteria (Kaoutari et al., 2013). GH13 family enzymes are characterised by several features, they are active on  $\alpha$ -glucosidic linkages, they contain a triosephosphate isomerase (TIM) barrel catalytic domain, the catalytic domain contains four to seven specific conserved regions, and the catalytic residues are two aspartic acids and one glutamic acid (Janeček et al., 2014; Toor et al., 2020). GH13 subfamilies 12, 13, and 14 have reported pullulanase enzyme activity, while subfamily 20 enzymes have reported neopullulanase activity, while also

exhibiting further cyclomaltoextrinase, and maltogenic  $\alpha$ -amylase activities (Janeček et al., 2014; Stam et al., 2006).

Pullulanases are divided into several types. Type I and type II pullulanases catalyse the hydrolysis of  $\alpha$ 1-6 linkages in pullulan, therefore releasing only maltotriose units. For type I pullulanases, this activity is exclusive while type II pullulanases can also hydrolyse  $\alpha$ 1-4 linkages in other glucose polysaccharides like starch, but not pullulan. In contrast to pullulanases, pullulan hydrolases type I (neopullulanase) and type II (Isopullulanase) cleave  $\alpha$ 1-4 linkages in pullulan, but not  $\alpha$ 1-6 linkages, resulting in panose (Glc $\alpha$ 1-6Glc $\alpha$ 1-4Glc) and isopanose (Glc $\alpha$ 1-4Glc $\alpha$ 1-6Glc), respectively, as end products. Generally, neopullulanases are not able to cleave smaller oligosaccharides, but some have been reported to cleave oligosaccharides as small as maltose (Glc $\alpha$ 1-4Glc) (Cheong et al., 2002). Pullulan hydrolase type III cleaves both  $\alpha$ 1-4 and  $\alpha$ 1-6 linkages in pullulan, resulting in a mix of maltotriose, panose, maltose, and glucose (Hii et al., 2012; Niehaus et al., 2000). To date, only 2 Type III pullulan hydrolases have been described (Toor et al., 2020), both in archaea: *Thermococcus aggregans* (Niehaus et al., 2000) and *Thermococcus kodakarensis* KOD1 (Ahmad et al., 2014).

*R. gnavus* ATCC 29149 encodes 9 predicted GH13 enzymes. In the previous chapter, we successfully expressed and purified *RgGH13-1* and *RgGH13-2* and assayed their activity against mucin glycans. Here, we characterised the substrate specificity of *RgGH13-1* and *RgGH13-2* and gained insights into their structural features and classification into GH13 subfamilies. Further, we tested the ability of *R. gnavus* ATCC 29149 to grow on glucose oligo- and polysaccharides.

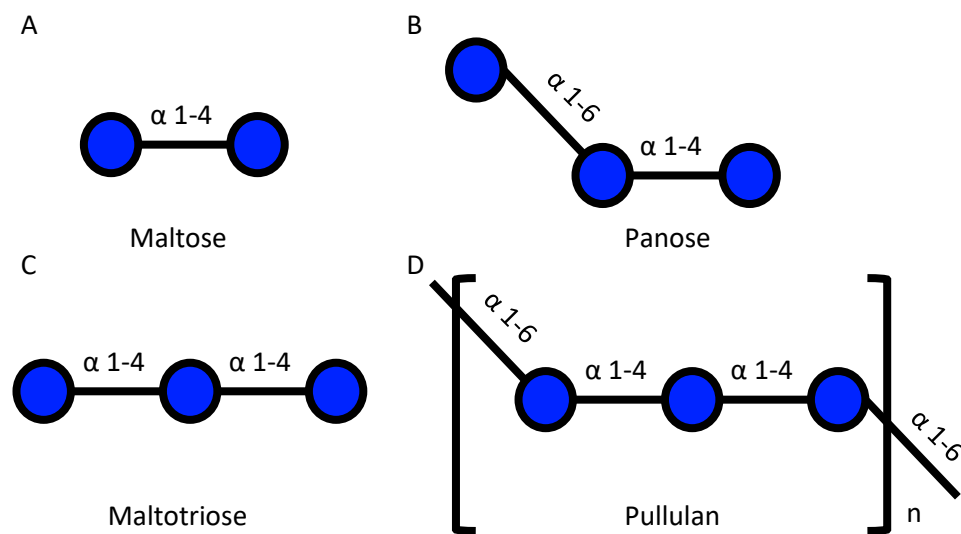


Figure 5.1. Schematic overview of pullulan degradation products. (A) Maltose consists of two  $\alpha$ 1-4 linked glucose units (Glc $\alpha$ 1-4Glc). (B) Panose consists of an  $\alpha$ 1-6 linked glucose unit and two  $\alpha$ 1-4 linked glucose units (Glc $\alpha$ 1-6Glc $\alpha$ 1-4Glc). (C) Maltotriose consists of three  $\alpha$ 1-4 linked glucose units ((Glc $\alpha$ 1-4Glc $\alpha$ 1-4Glc). (D) Pullulan is a polysaccharide consisting of maltotriose subunits, which are linked with  $\alpha$ 1-6 links.

## 5.2 Results

### 5.2.1 Structural basis for *RgGH13* substrate specificity

*RgGH13-1* and *RgGH13-2* show a modular organisation characterised by the presence of a catalytic domain and a carbohydrate binding module belonging to CBM34 family for *RgGH13-1* and CBM48 for *RgGH13-2* ([www.cazy.org](http://www.cazy.org)). These CBM types are widespread in GH13 enzymes, often characterised by binding affinity to starch and pullulan (Janeček et al., 2019). *RgGH13-1* and *RgGH13-2* do not have any other predicted domains, including signal peptides, suggesting that these enzymes are intracellular.

Alignment analysis of conserved sequence regions across functionally characterised GH13 enzymes indicated that *RgGH13-1* belongs to subfamily GH13\_20, known as a neopullulanase subfamily, and *RgGH13-2* to subfamily GH13\_14 (Prof. Štefan Janeček, Bratislava, Slovakia; personal communication). CBM34 is common in GH13 subfamilies 20 and 21, while CBM48 is, among others, highly represented in subfamilies 12, 13, and 14 (Janeček et al., 2019), which is in line with the classification of *RgGH13-1* and *RgGH13-2* into their GH13 subfamilies.



Further sequence analysis was carried out to identify functionally characterised enzymes sharing the highest amino acid (aa) identity to *RgGH13-1* and *RgGH13-2*. *RgGH13-1* was found to have 37.5% and 42.5% aa sequence identity with two neopullulanases, Uniprot ID: NEPU2\_THEVU from *Thermoactinomyces vulgaris* and Uniprot ID: NEPU\_GEOSE from *Geobacillus stearothermophilus*, respectively, 41.4% aa identity with an intracellular maltogenic amylase from *Bacillus subtilis* (Uniprot ID: BBMA\_BACIU), and 43.6% aa sequence identity with a cyclomaltodextrinase from *Geobacillus thermopakistaniensis* (Uniprot ID: CDAS\_GEOTM), consistent with *RgGH13-1* substrate specificity.

*RgGH13-2* displayed 64.9% aa sequence identity with a type I pullulanase from *Faecalicatena orotica* (Uniprot ID: A0A2Y9B8P7\_9FIRM) and 38.23% aa identity with a type I pullulanase from *Paenibacillus barengoltzii* (P. Huang et al., 2020), also consistent with *RgGH13-2* substrate specificity.

Genes surrounding the *RgGH13-1* and *RgGH13-2* loci were upregulated with a similar Log<sub>2</sub> fold change, indicating these GHs might be part of an operon. *RgGH13-1* has three maltodextrin ABC transporter genes with similar expression flanking the gene (Table 5.1). *RgGH13-2* is flanked by two maltodextrin ABC transporter genes with similar expression (Table 5.2). In addition, *RgGH13-2* is flanked by two predicted GH13 enzymes, one of which is predicted to be a neopullulanase, a predicted glycogen debranching enzyme, and a predicted Maltodextrin phosphorylase.

Table 5.1. Overview of genes flanking *RgGH13-1* (RGNA\_RS06060).

NZ_CP027002.1 locus tag	NCBI annotation	Swiss-Prot predicted function	Log2 fold change	P-adjusted value
RGNA_RS06045	DUF1667 domain-containing protein	N-acetyl-gamma-glutamyl-phosphate reductase	4.46	6.82E-06
RGNA_RS06050	FAD-dependent oxidoreductase	Thioredoxin reductase	5.07	2.01E-12
RGNA_RS06055	NAD(P)/FAD-dependent oxidoreductase	Glycerol 3-phosphate dehydrogenase	5.82	1.38E-17
RGNA_RS06060	glycoside hydrolase family 13 protein	Neopullulanase	3.21	2.35E-21
RGNA_RS06065	maltose ABC transporter substrate-binding protein	Maltooligosaccharide ABC transporter solute-binding lipoprotein	2.85	5.14E-12
RGNA_RS06070	sugar ABC transporter permease	Maltodextrin transport system permease protein MalD	1.87	6.94E-10
RGNA_RS06075	sugar ABC transporter permease	Maltodextrin transport system permease protein MalC/F	1.71	1.77E-06

Log2 fold change and p-adjusted value are taken from RNAseq results. NCBI annotation is the predicted function from the NCBI database. Swiss-Prot is the function predicted by Swiss-Prot BLAST analysis, using the Uniprot standard settings and bacterial results with high identity were used.

Table 5.2. Overview of genes flanking *RgGH13-2* (RGna\_RS08510).

NZ_CP027002.1 locus tag	NCBI annotation	Swiss-Prot predicted function	Log2 fold change	P-adjusted value
RGna_RS08485	alpha-glucosidase	Oligo-1,6-glucosidase	0.78	0.0003
RGna_RS08490	glycoside hydrolase family 13 protein	Neopullulanase / Intracellular maltogenic amylase / Cyclomaltodextrinase	1.03	0.0046
RGna_RS08495	4-alpha-glucanotransferase	4-alpha-glucanotransferase	1.76	0.0005
RGna_RS08500	glycogen/starch/alpha-glucan family phosphorylase	Maltodextrin phosphorylase	0.53	0.0198
RGna_RS08505	glycogen debranching enzyme family protein	Glycogen debranching enzyme	0.81	0.0467
RGna_RS08510	type I pullulanase	type I pullulanase	2.03	6.07E-08
RGna_RS08515	sugar ABC transporter permease	Maltose/maltodextrin transport system permease protein MalD/G	2.09	1.73E-06
RGna_RS08520	sugar ABC transporter permease	Maltose/maltodextrin transport system permease protein MalC/F	2.03	6.84E-06

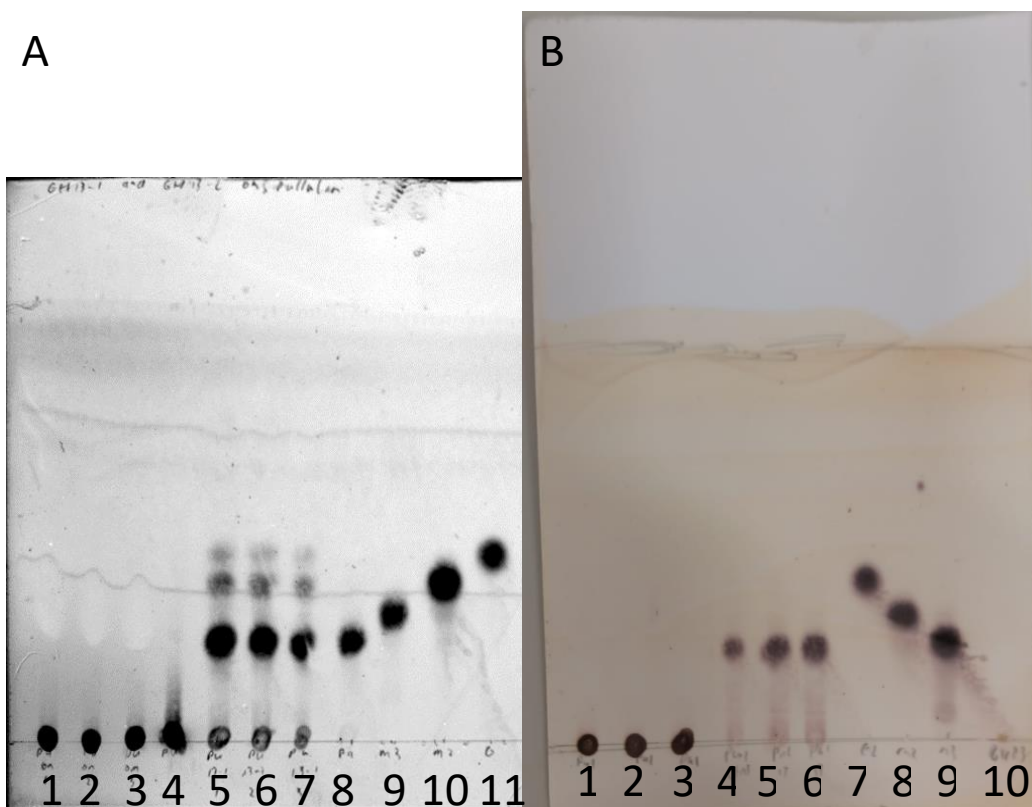
Log2 fold change and p-adjusted value are taken from RNAseq results. NCBI annotation is the predicted function from the NCBI database. Swiss-Prot is the function predicted by Swiss-Prot BLAST analysis, using the Uniprot standard settings and bacterial results with high identity were used.

### 5.2.2 *RgGH13* activity on pullulan

Because both GH13s are predicted to be pullulanases, the enzymatic activity of *R. gnavus* recombinant enzymes *RgGH13-1* and *RgGH13-2* was first tested towards pullulan. The enzymatic reaction products of *RgGH13-1* with pullulan were first analysed by thin-layer chromatography (TLC). A spot corresponding to panose was observed but no spot corresponding to maltotriose could be identified (Figure 5.2), indicating that the enzyme was able to cleave  $\alpha$ 1-4 linkages, but not  $\alpha$ 1-6 linkages. As shown by TLC, panose was further degraded into a product which may correspond to isomaltose (Glc $\alpha$ 1-6Glc) and glucose following further  $\alpha$ 1-4 linkage hydrolysis while the  $\alpha$ 1-6 linkage between two glucose units remained intact. However, it was not possible to distinguish between maltose and isomaltose standards by TLC.

*RgGH13-2* was also found to be active on pullulan. The TLC showed the formation of maltotriose, indicating that this enzyme could hydrolyse only the  $\alpha$ 1-6 linkages. Together

these results suggest that *RgGH13-1* is a neopullulanase (pullulan hydrolase type I) and *GH13-2* a pullulan hydrolase type I or type II.



*Figure 5.2.* TLC analysis of the enzymatic reaction of *RgGH13-1* and *RgGH13-2* on pullulan. **(A)** 1-3 = Pullulan only (assay conditions), 4 = Pullulan (standard), 5-7 = *RgGH13-1* on pullulan, 8 = panose (standard), 9 = maltotriose (standard), 10 = maltose (standard), 11 = glucose (standard). **(B)** 1-3 = Pullulan only (assay conditions), 4-6 = *RgGH13-2* on pullulan, 7 = glucose (standard), 8 = maltose (standard), 9 = maltotriose (standard), 10 = *RgGH13-2* only (assay conditions).

Sequence alignments with sequences with high aa identity (see 5.2.1) showed conservation of the catalytic residues including an aspartic acid (Asp) acting as the catalytic nucleophile, a glutamic acid (Glu) as catalytic proton donor, and another aspartic acid acting as transition-state stabiliser (Supplemental Figure S8). These correspond to Asp357, Glu386 and Asp458 in *RgGH13-1*, and to Asp336, Glu367 and Asp458 in *RgGH13-2*. Two histidine and one arginine implicated in substrate binding were also found to be conserved in all aligned sequences and could also be identified as His262, Arg355, and His457 in *RgGH13-1* and His271, Arg334, and His457 in *RgGH13-2* (Supplemental Figure S8). However, Asp468 and Arg472, shown to be involved in binding to panose, maltotetraose, and isopanose in NEPU\_GEOSE from *G. stearothermophilus* (Hondoh et al., 2003), were either missing from both *RgGH13-1* and *RgGH13-2*, as it is the case for Asp468 or from *RgGH13-2* as is the case for Arg472 (Supplemental Figure S8).

A 3D AlphaFold 2 model of full length *RgGH13-1* was generated and superimposed on *G. stearothersophilus* neopullulanase TRS40 (Hondoh et al., 2003) in complex with panose (Figure 5.3). Full-length *RgGH13-2* AlphaFold 2 3D structure was superimposed with *P. barengoltzii* type I pullulanase (P. Huang et al., 2020) (Figure 5.4). This analysis confirmed the nature and location of the catalytic residues within the enzyme catalytic groove. To validate this experimentally, two catalytic mutants were generated based on sequence alignments with functionally characterised enzymes (Cockburn et al., 2021; Kamitori et al., 2002). *RgGH13-1* D357N and *RgGH13-2* D336A catalytic mutants were heterologously expressed and the purified mutants tested against pullulan. *RgGH13-2* D336A showed total loss of catalytic activity on pullulan (Figure 5.5B) while traces of activity could be detected with *RgGH13-1* D357N (Figure 5.5A), albeit much reduced compared to the wild-type enzyme.

In addition, the CBM48 and the catalytic domain of *RgGH13-2* were recombinantly expressed separately, however we were unable to obtain a recombinant *RgGH13-2* catalytic domain. As expected, the CBM48 domain was inactive on pullulan as shown by TLC (Figure 5.5C).

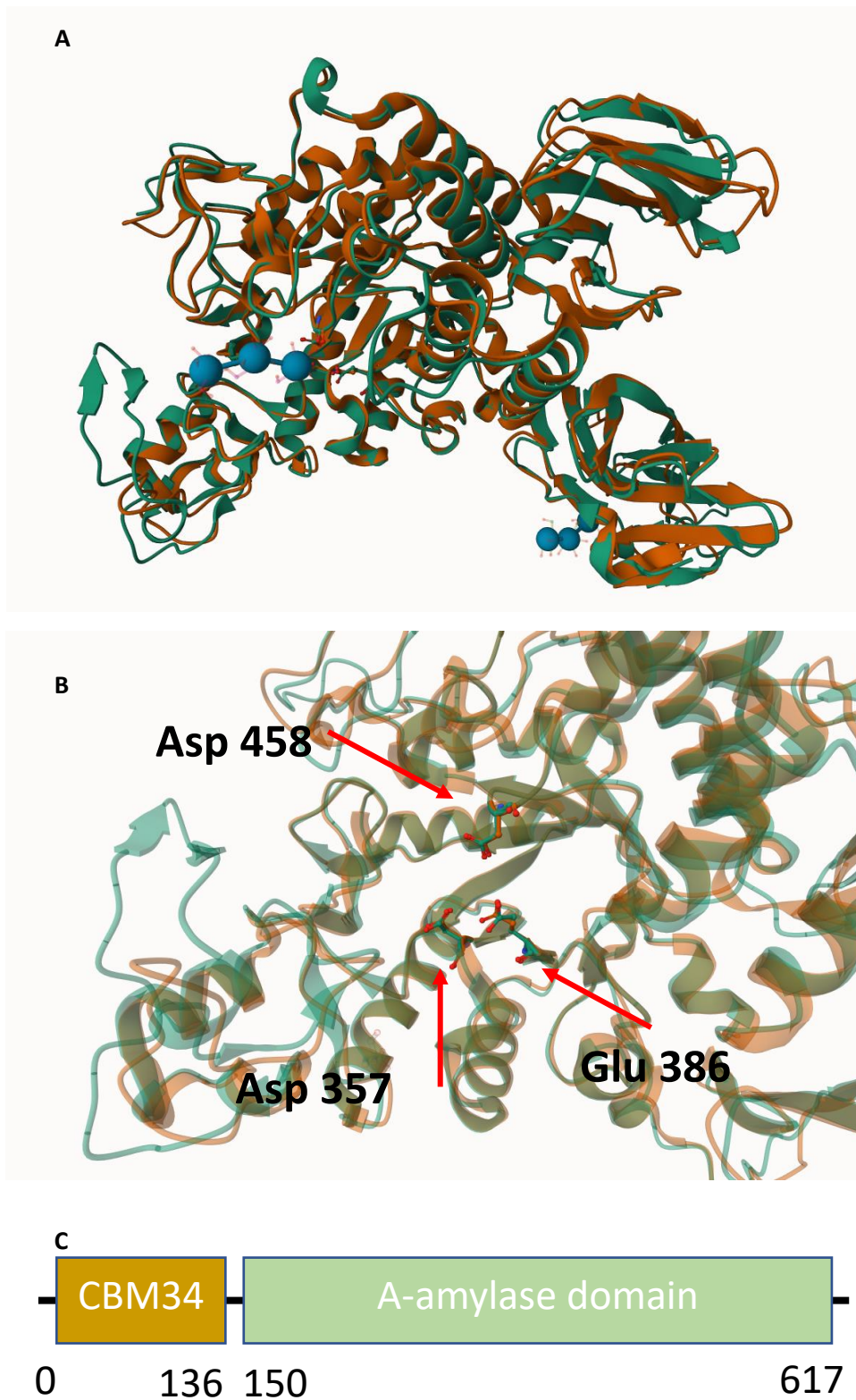


Figure 5.3. AlphaFold model of *RgGH13-1* superimposed on *Geobacillus stearothermophilus* neopullulanase TRS40 in complex with panose. **(A)** Overview of *RgGH13-1* model superimposed on *G. stearothermophilus* neopullulanase TRS40 **(B)** Detail of catalytic residues of both neopullulanases. *RgGH13-1* catalytic residues are highlighted. **(C)** Modular organisation of *RgGH13-1*.

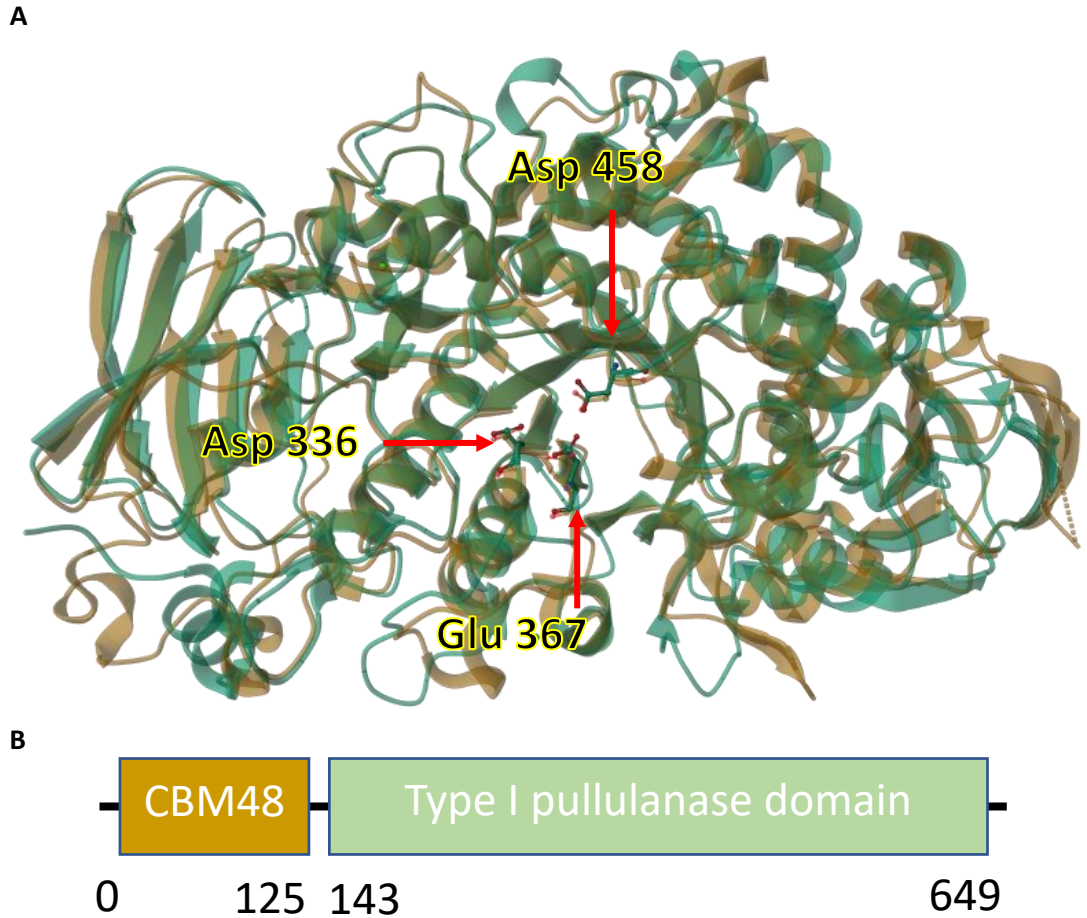


Figure 5.4. AlphaFold model of RgGH13-2 superimposed on *Paenibacillus barengoltzii* type I pullulanase. Overview of RgGH13-2 model superimposed on *P. barengoltzii* type I pullulanase. RgGH13-2 catalytic residues are highlighted. **(B)** Modular organisation of RgGH13-2.

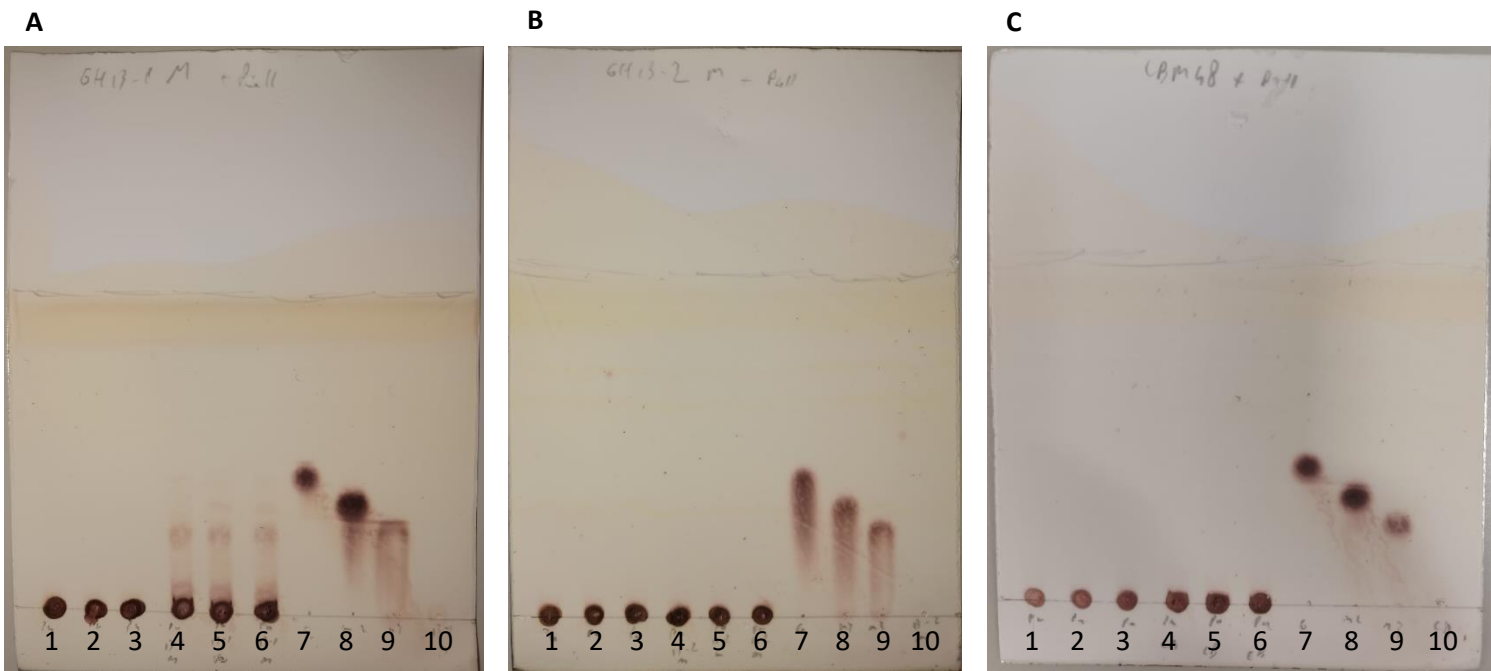


Figure 5.5. TLC analysis of the enzymatic reaction of RgGH13-1 D357N and RgGH13-2 D336A catalytic mutants, and RgGH13-2 CBM48 domain with pullulan. **(A)** RgGH13-1 D357N catalytic mutant with pullulan. **(B)** RgGH13-2 D336A catalytic mutant with pullulan. **(C)** RgGH13-2 CBM48 domain with pullulan. 1-3 = pullulan only (assay conditions), 4-6 = RgGH13-1/2 mutant or CBM48 domain with pullulan, 7 = glucose (standard), 8 = maltose (standard), 9 = maltotriose (standard), 10 = RgGH13-1/2 mutant or CBM48 domain only.



### 5.2.3 Functional characterisation of *RgGH13-1* and *RgGH13-2*

#### 5.2.3.1 *RgGH13* substrate specificity on glucose oligo- and polysaccharides

In order to further determine the substrate specificity of *RgGH13-1* and *RgGH13-2*, these enzymes were tested on a range of glucose-containing oligo- and polysaccharides. *RgGH13-1* enzyme was found to be active on starch, releasing maltotriose, maltose and glucose, as shown by HPAEC (Figure 5.6A). Further substrate specificity was determined by testing the two enzymes on maltose, isomaltose, maltotriose, panose,  $\alpha$ -cyclodextrin, and  $\beta$ -cyclodextrin at 1 mM concentrations for 24 hours. HPAEC analysis of the enzymatic reactions revealed that recombinant *RgGH13-1* released (iso)maltose and glucose from maltotriose or panose (Figure 5.6B), while glucose was the product of the reaction with maltose (Figure 5.3B). *RgGH13-1* was less active on panose and maltose, only releasing small amounts of maltose and glucose after 24 hours. *RgGH13-1* was found to be active on both  $\alpha$ -cyclodextrin and  $\beta$ -cyclodextrin, releasing maltose and glucose (Figure 5.6C). Because *RgGH13-1* is predicted to be a cyclomaltoamylase in addition to a neopullulanase, this activity was expected. No activity was found on isomaltose (Figure 5.6C).

*RgGH13-2* showed no activity on starch, maltotriose, isomaltose, and maltose (Figure 5.7 A and B). It released very small amounts of glucose and maltose from panose and  $\alpha$ -cyclodextrin and  $\beta$ -cyclodextrin (Figure 5.7 A and B), but the activity after 24 hours was very low compared to the  $\alpha$ 1-6 activity seen on pullulan. Even though starch does have  $\alpha$ 1-6 linked branches, no activity was found on starch.

The enzymes were then tested on glycogen, which contains  $\alpha$ 1-4 linked glucose chains with  $\alpha$ 1-6 linked sidechains, and the enzymatic reaction products analysed by TLC. *RgGH13-1* released glucose and maltose from glycogen while a small smear was seen on TLC for *RgGH13-2*, which may be indicative of de-branching activity of *RgGH13-2* (Figure 5.8).

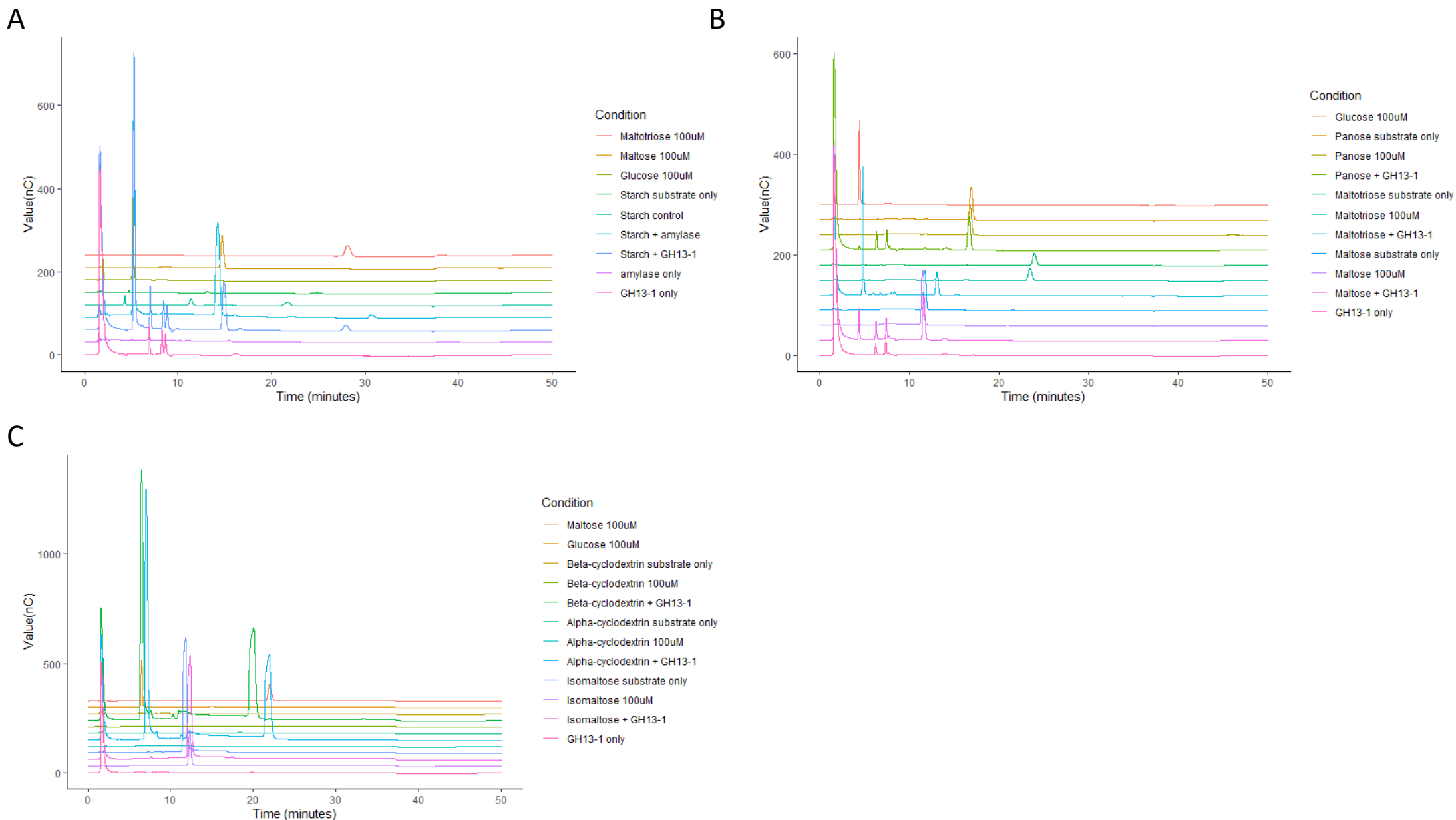
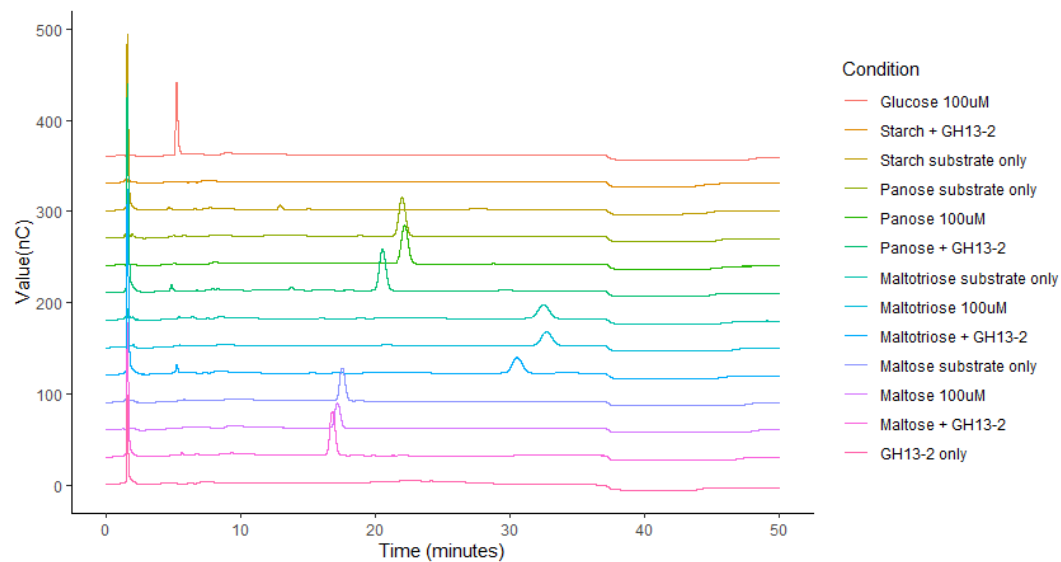


Figure 5.6. Analysis of RgGH13-1 substrate specificity. HPAEC-PAD chromatograms of enzymatic reactions of RgGH13-1 with (A) starch, (B) panose, maltotriose, maltose, (C)  $\alpha$ -cyclodextrin,  $\beta$ -cyclodextrin, and isomaltose. All substrate concentrations were 0.1 mM, except starch at 1 mg/ml concentrations. Control reactions were performed under the same conditions.

A



B

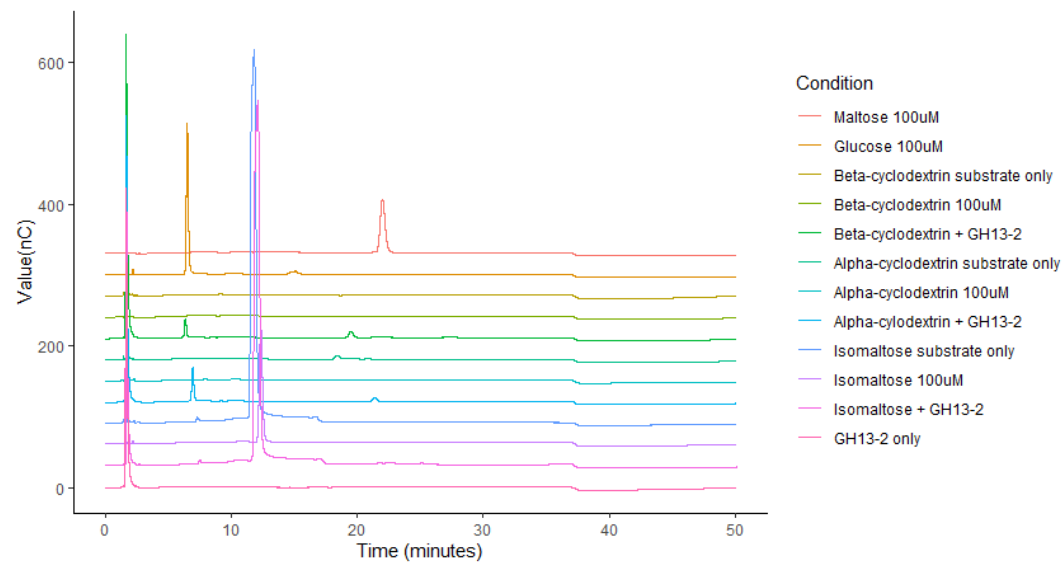


Figure 5.7. Analysis of RgGH13-2 substrate specificity. HPAEC-PAD chromatograms of enzymatic reactions of RgGH13-2 with (A) starch, panose, maltotriose, and maltose, and (B)  $\alpha$ -cyclodextrin,  $\beta$ -cyclodextrin, and isomaltose. All substrate concentrations were 0.1 mM, except starch at 1 mg/ml concentrations. Control reactions were performed under the same conditions.



Figure 5.8. TLC analysis of the enzymatic reaction of *RgGH13-1* and *RgGH13-2* on glycogen. 1-3 = glycogen only (assay conditions), 4 = glycogen control. 5-7 = *RgGH13-1* on glycogen, 8 = maltotetraose (standard), 9 = maltotriose (standard), 10 = maltose (standard), 11 = glucose (standard) 12-14 = *RgGH13-2* on glycogen.

### 5.2.3.2 RgGH13-2 enzyme kinetics

To gain further insight into the enzymatic activity of RgGH13-2, the kinetic parameters of this enzyme were determined using pullulan as substrate with concentrations ranging from 0.05 mg/ml to 1 mg/ml. A standard curve of maltotriose was used to calculate the release of maltotriose over time. The enzymatic reaction reached a maximum turnover rate at a pullulan concentration of 0.4 mg/ml (Figure 5.9). The curve shows a reduction at higher substrate concentrations, which could indicate substrate inhibition. The  $V_{\max}$  was calculated to be  $0.6 \mu\text{moles L}^{-1} \text{s}^{-1}$ , and an apparent  $K_m$  of 0.26 mg/ml (Table 5.3).

Table 5.3. Enzyme kinetics of RgGH13-2 with pullulan.

	$K_m$ (mg/ml)	$V_{\max}$ ( $\mu\text{moles/s}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )
Pullulan	0.26	0.6	5.8

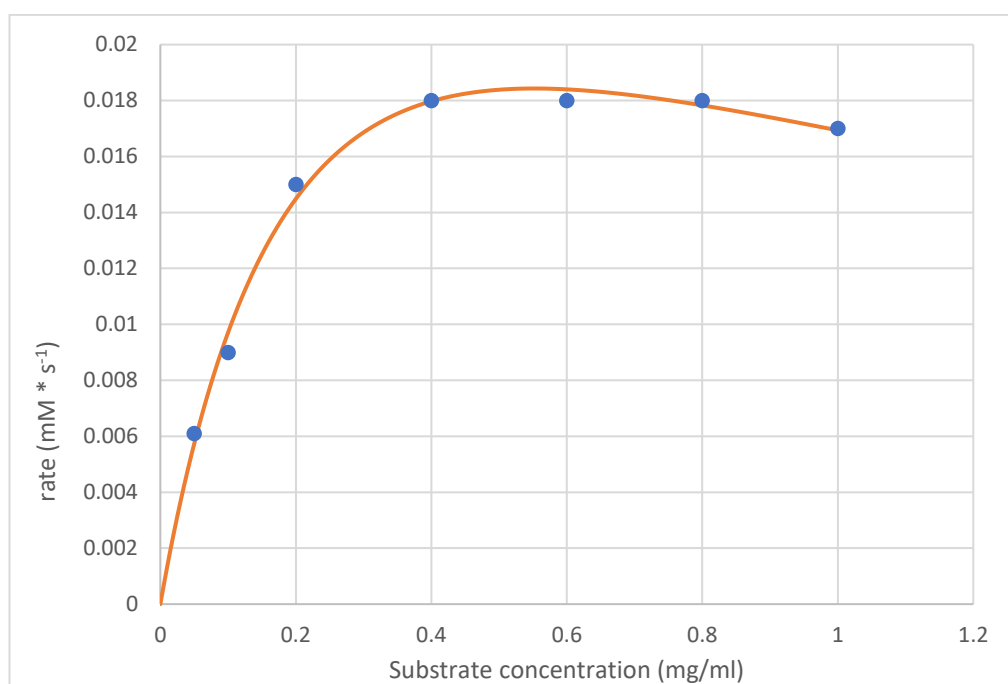


Figure 5.9. RgGH13-2 activity on pullulan turnover rate at a range of pullulan concentrations. The orange line reflects the non-linear substrate inhibition model.

#### 5.2.4 Growth of *R. gnavus* ATCC 29149 on glucose oligo- and polysaccharides

As shown above, *RgGH13-1* and *RgGH13-2* from *R. gnavus* ATCC 29149 are active on starch and pullulan. However, *R. gnavus* ATCC 29149 has previously been reported not to be able to grow on starch as sole carbon source (Croft et al., 2018). Here, we tested the growth of *R. gnavus* ATCC 29149 on several glucose oligo- and polysaccharides including 0.5% pullulan, starch, and glycogen, and 27.7 mM  $\alpha$ -cyclodextrin,  $\beta$ -cyclodextrin, panose, maltotriose, isomaltose, maltose, and glucose.

*R. gnavus* ATCC 29149 grew on glucose, maltose, maltotriose, as reported before (Croft et al., 2018), and panose (Figure 5.10). The bacteria grew to highest density in the presence of glucose. Growth on panose and maltose was similar during the lag phase and beginning of the exponential phase, but after 16 hours of growth, the bacteria grown on panose reached a plateau (Figure 5.10). Because the sugars were used in equimolar concentrations, more glucose units were available to the bacteria in the presence of panose, a trisaccharide compared to maltose, a disaccharide. These results are consistent with the capacity of *RgGH13-1* and *RgGH13-2* to cleave  $\alpha$ 1-4 linkage in panose, releasing glucose, while they are unable to cleave the  $\alpha$ 1-6 linkage confirmed by the inability of *R. gnavus* ATCC 29149 to grow on isomaltose (Glc $\alpha$ 1-6Glc). During growth on glucose, *R. gnavus* ATCC 29149 showed a shorter lag phase and a higher final OD compared to growth on maltose, even though maltose has twice the amount of glucose units present in the growth media.

Both *RgGH13-1* and *RgGH13-2* are active on pullulan, releasing a mix of panose, maltose, and glucose, or maltotriose, respectively. Additionally, *RgGH13-1* showed activity on starch and  $\alpha$ -cyclodextrin and  $\beta$ -cyclodextrin, releasing maltose and glucose. *R. gnavus* ATCC 29149 was able to grow on all four of the released sugars, panose, maltose, glucose, and maltotriose (Figure 5.10). However, no growth was observed on pullulan, starch, glycogen,  $\alpha$ -cyclodextrin, or  $\beta$ -cyclodextrin (Figure 5.10). The inability of *R. gnavus* ATCC 29149 to grow on pullulan, starch,  $\alpha$ -cyclodextrin, or  $\beta$ -cyclodextrin despite *RgGH13-1* and *RgGH13-2* substrate specificity towards these substrates, is in line with the predicted intracellular localisation of these enzymes in the bacteria (see 4.2.2.3).

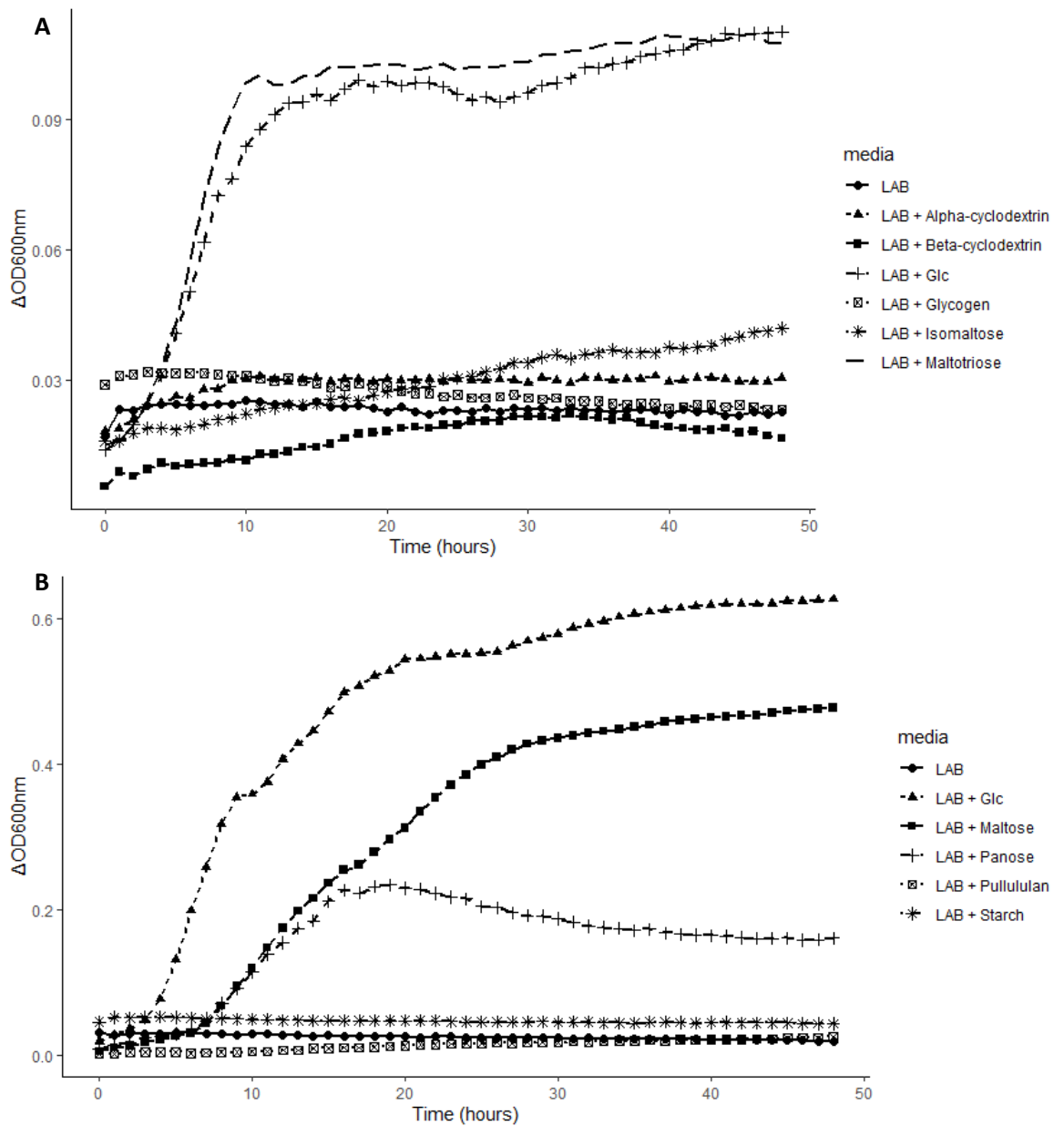


Figure 5.10. Growth curves of *R. gnavus* ATCC 29149 on glucose oligo- and polysaccharides. **(A)** *R. gnavus* growth in LAB supplemented with 27.7 mM glucose, maltose, panose or 0.5% pullulan or starch. **(B)** *R. gnavus* growth in LAB supplemented with 27.7 mM glucose,  $\alpha$ -cyclodextrin,  $\beta$ -cyclodextrin, isomaltose, maltotriose or 0.5% glycogen.  $\Delta OD$  was calculated by subtracting the negative control (LAB + carbohydrate) from the experimental samples (LAB + carbohydrate + bacteria).

### 5.3 Discussion

*RgGH13-1* is active on pullulan releasing panose, isomaltose and glucose, and on  $\alpha$ -cyclodextrin and  $\beta$ -cyclodextrin and shows a strict specificity for  $\alpha$ 1-4 linkages, indicating that *RgGH13-1* is a neopullulanase. Based on homology modelling and sequence alignments (Prof. Štefan Janeček, Bratislava, Slovakia; personal communication), this enzyme was classified in GH13 subfamily 20 also characterised by the presence of a CBM34 domain. Among gut bacteria, *RgGH13-1* shares similar substrate specificity to a neopullulanase characterised from the human gut symbiont *B. thetaiotaomicron* 95-1, which was shown to release panose from pullulan (Smith & Salyers, 1991).

*RgGH13-2* is active on pullulan, releasing maltotriose and shows a strict specificity for  $\alpha$ 1-6 linkages, indicating that *RgGH13-2* is a type I pullulanase. *RgGH13-2* was found to be part of GH13 subfamily 14 with a CBM48 domain. Among gut bacteria, *RgGH13-2* shares similar substrate specificity to functionally characterised Amy10 from *Ruminococcus bromii* L2-63 (Cockburn et al., 2021).

*RgGH13-2* showed a  $V_{\max}$  of 0.6  $\mu\text{moles L}^{-1} \text{s}^{-1}$  and an apparent  $K_m$  of 0.26 mg/ml. This compares to a  $K_m$  of 10.5 mg/ml of *Klebsiella pneumoniae* type I pullulanase (Ali et al., 2013, 2015), a  $K_m$  of 1.33 mg/ml of *Clostridium thermosulfurogenes* EM1 type II pullulanase (Spreinat & Antranikian, 1990), and a  $K_m$  of 2.0 and 2.38 mg/ml from *T. kodakarensis* KOD1 and *T. aggregans* type III pullulanases (Ahmad et al., 2014), respectively. Therefore *RgGH13-2* shows highest affinity for pullulan compared to other pullulanases functionally characterised to date.

*RgGH13-2* is part of a potential cluster of genes containing two more GH13 encoding genes (RGNA\_RS08485 and RGNA\_RS08490), a starch/glycogen phosphorylase (RGNA\_RS08500 *glgP*, involved in starch or glycogen metabolism), a glycogen debranching enzyme (RGNA\_RS08505), and an ABC sugar transporter (RGNA\_RS08515 and RGNA\_RS08520). It is therefore likely that this cluster is involved in *R. gnavus* glycogen and/or starch metabolism.

*R. gnavus* ATCC 29149 contains predicted glycogen synthesis genes (*glgA*, *glgB*, *glgC/D*) and glycogen metabolism genes (*glgP*). *GlgP* releases glucose-1-phosphate from glycogen sidechains, reducing the chain length until 4 glycosyl units is reached. A debranching



enzyme then removes the maltotetraose chain which can then be cleaved into maltose units by a maltogenic amylase (L. Wang & Wise, 2011). *RgGH13-1* and *RgGH13-2* may play a role in glycogen metabolism, with *RgGH13-1* acting as an exo-acting enzyme to release maltose from glycogen sidechains, and *RgGH13-2* as a debranching enzyme. *R. gnavus* ATCC 29149 does not have a gene annotated as a *glgX*, the glycogen debranching enzyme in this pathway but *RGNA\_RS08505*, which is predicted to be a glycogen debranching enzyme, or *RgGH13-2* could act as putative *glgX*. Pullulanases have been shown before to have glycogen debranching activity (Shim et al., 2009). Testing *RgGH13-2* activity on branched  $\beta$ -cyclodextrin and on glycogen would help to confirm its role as a potential debranching enzyme.

Although we demonstrated that *R. gnavus* ATCC 29149 produces two *RgGH13* enzymes functionally active on pullulan and starch (and encodes another 7 GH13 in its genome), our results showed that this strain is unable to grow on pullulan or starch while it can utilise degradation products maltotriose, maltose, and glucose. Both *RgGH13* enzymes have a predicted cytoplasmic localisation which would explain the inability of *R. gnavus* ATCC 29149 to utilise pullulan and starch.

However, a role of *RgGH13-1* and *RgGH13-2* enzymes in starch metabolism cannot be ruled out. *R. gnavus* ATCC 29149 cross-feeding on starch with *R. bromii* has been shown before (Crost et al., 2018). During starch metabolism smaller maltooligosaccharides are released and taken up by transporters to be further broken down inside the cell (Cerqueira et al., 2020; Cockburn et al., 2015; Crost et al., 2018). Further research on the specificity of maltooligosaccharide transporters in *R. gnavus* ATCC 29149 in combination with the testing of *R. gnavus* ATCC 29149 growth and *RgGH13-1* and *RgGH13-2* activity on a wider range of maltooligosaccharides would be necessary in order to determine the role of these GH13 enzymes during starch cross-feeding.

For *RgGH13-1*, this is consistent with the fact that neopullulanases often have cyclomaltodextrinase activity (Hii et al., 2012). Neopullulanases, cyclomaltodextrinases, and maltogenic amylases have been reported to have high sequence identity (40-86%) and are often indistinguishable from each other (Lee et al., 2002). Cyclomaltodextrinases are intracellular enzymes involved in the metabolism of cyclodextrins which are produced extracellularly during starch metabolism and subsequently transported into the cell (Aroob

et al., 2021). However, *R. gnavus* ATCC 29149 was not able to grow on  $\alpha$ -cyclodextrin and  $\beta$ -cyclodextrin, suggesting that this strain does not express a functional cyclodextrin transporter or that the enzymes were not expressed in the conditions tested. Further research is needed to understand the metabolic role of *RgGH13-1* inside the bacteria. In addition, none of the *R. gnavus* ATCC 29149 GH13 enzymes have predicted signal peptides and are thus likely to be intracellular. Together with the absence of growth of *R. gnavus* ATCC 29149 on starch indicates that the GH13 encoded in its genome fulfil activities other than starch degradation.

Pullulan is not a common dietary glycan and as discussed above many pullulanases have other primary enzymatic functions, like starch debranching (Hii et al., 2012). The advantage of pullulanases as starch debranching enzymes is that they are so-called 'direct' debranching enzymes, which do not require any modification of the substrate to be active (Hii et al., 2012). In addition, pullulanases require only two  $\alpha$ 1-4 linked glycosyl units on each side of the  $\alpha$ 1-6 linkage, while debranching enzymes like isoamylases require larger chains attached to the  $\alpha$ 1-6 linked glucose (Hii et al., 2012). These are advantageous properties in industry where pullulanases are utilised in starch saccharification (Hii et al., 2012; Toor et al., 2020). Further characterisation of *RgGH13-1* and *RgGH13-2* on pullulan are warranted to further explore the suitability of these enzymes for potential biotechnological applications and establish their biological roles in bacteria.

## Chapter 6

# 2,7-anhydro-Neu5Ac and Neu5Ac utilisers in the human faecal microbiome

## 6. 2,7-anhydro-Neu5Ac and Neu5Ac utilisers in the human faecal microbiome

### 6.1 Introduction

Sialic acid is a common terminal epitope in mucins and other glycans. Gut bacteria encoding sialidases can release Neu5Ac from mucins to utilise as a carbon source directly or to access the underlying glycan chains. Sialic acid released into the gut environment is an important source of cross-feeding as it can be scavenged by other microbes including strains from the same species, as recently reviewed (Bell et al., 2023; Bell & Juge, 2021; Berkhout et al., 2022). *Bifidobacterium breve* is unable to grow on mucins, however it is able to grow when in co-culture with *Bifidobacterium bifidum* (Bunesova et al., 2018), due to the release of sialic acid from *B. bifidum* sialidases (Nishiyama et al., 2018). Interestingly, *B. thetaiotaomicron* ATCC 29148 and *A. muciniphila* DSM 22959 encode sialidases able to release sialic acid from mucin glycans in order to gain access to underlying glycans, but are unable to utilise sialic acid (Brigham et al., 2009; K. Huang et al., 2015; Shuoker et al., 2023; Tailford, Owen, et al., 2015). Sialic acid cross-feeding was also observed in co-cultures of *Roseburia inulinivorans* or *Faecalibacterium prausnitzii* and *A. muciniphila* during growth on mucin glycans (Shuoker et al., 2023). Sialic acid was found to be a main driver behind *R. inulinivorans* and *F. prausnitzii* butyrate production in co-culture (Shuoker et al., 2023). Several pathogens like *E. coli*, *Clostridium difficile*, and *Salmonella typhimurium* species do not encode any sialidases, and are unable to degrade mucin glycans which are capped by sialic acid. When uncapped by sialidases of other species, growth of these pathogens can be promoted, either due to direct growth on sialic acid or the access to underlying mucin glycans (K. Huang et al., 2015; Ng et al., 2013).

As reported in sections 1.3.2 and 4.2.2.2, *R. gnavus* ATCC 29149 encodes a GH33 intramolecular *trans*-sialidase (*RgNanH*) which cleaves off  $\alpha$ 2,3-linked terminal sialic acid residues in mucin, releasing 2,7-anhydro-Neu5Ac (Tailford, Owen, et al., 2015). This enzyme is part of the *nan* operon, which is highly upregulated during growth on mucin glycans (see chapter 4.2.2.2) and 2,7-anhydro-Neu5Ac (Croft et al., 2013, 2016). *R. gnavus* is able to metabolise 2,7-anhydro-Neu5Ac due to the selectivity of the solute binding protein (*RgSBP*) part of the ABC transporter and the presence of an oxido-reductase (*RgNanOx*) converting 2,7-anhydro-Neu5Ac into Neu5Ac, which then becomes substrate

for the Neu5Ac aldolase (*RgNanA*) (Bell et al., 2019, 2020). So far, *R. gnavus* strains are the only bacterial species shown to be able to release 2,7-anhydro-Neu5Ac from Neu5Ac bound to glycans in the gut, prompting questions about the possibility for other gut bacteria to benefit from this released sialic acid form.

Our group recently showed that *E. coli* BW25113 was shown to be able to grow on 2,7-anhydro-Neu5Ac due to the presence of a sialic acid transporter (YjhB) able to transport 2,7-anhydro-Neu5Ac and an oxidoreductase (YjhC) able to transform 2,7-anhydro-Neu5Ac into Neu5Ac (Bell et al., 2020). *E. coli* BW25113 does not encode an IT-sialidase able to release 2,7-anhydro-Neu5Ac from glycan structures and thus relies on cross-feeding to utilise 2,7-anhydro-Neu5Ac (Bell et al., 2020). Interestingly, *E. coli* can metabolise both Neu5Ac and 2,7-anhydro-Neu5Ac (Bell et al., 2020) while *R. gnavus* is only able to grow on 2,7-anhydro-Neu5Ac (Crost et al., 2013, 2016). Further *in silico* genomic and phylogenetic analysis of bacterial sialic acid transporters suggest that the capacity of transporting 2,7-anhydro-Neu5Ac may occur across bacterial pathogens and commensals, although limited in the gut environment (Severi et al., 2021). *Streptococcus pneumoniae* strains express a IT-sialidase able to release 2,7-anhydro-Neu5Ac (Gut et al., 2008; G. Xu et al., 2008). This IT-sialidase has been used in the past to enzymatically produce 2,7-anhydro-Neu5Ac *in vitro* (W. Li et al., 2019; Xiao et al., 2018). *S. pneumoniae* D39 further contains a *nan* operon encoding an ABC transporter which has been suggested to be specific for 2,7-anhydro-Neu5Ac and an oxidoreductase gene, suggesting that *S. pneumoniae* D39 might also be able to release, transport and metabolise 2,7-anhydro-Neu5Ac in the respiratory tract (Bell et al., 2020).

Metabolic dependencies through cross-feeding in microbial communities influences the gut microbiota composition (Mataigne et al., 2021). This could be through direct sharing of released sugars such as sialic acid mentioned above or from sharing metabolites resulting from the fermentation of the sugars (Culp & Goodman, 2023). To address the complexity and experimental limitations caused by the large number of possible interactions, new methodologies such as stable-isotope probing (SIP) used in environmental ecology have been applied to the study of gut microbial communities. SIP is dependent on the use of substrates labelled with stable isotopes like  $^2\text{H}$  hydrogen,  $^{15}\text{N}$  nitrogen, or  $^{13}\text{C}$  carbon (Dumont & Hernández García, 2019) in order to identify bacterial species able to

metabolise specific compounds (Radajewski et al., 2000). When species in a bacterial community metabolise these compounds, the heavy isotope is incorporated into their DNA and RNA, this heavy DNA can be distinguished from the DNA of species that were not able to metabolise the compound, which will only have unlabelled DNA, this is followed 16S amplicon sequencing (Dumont & Hernández García, 2019). RNA-SIP can be used as an alternative method due to the higher amount of RNA in the cells and a high turnover rate (Weis et al., 2019). After fractionation RNA is converted to cDNA and sequenced. Both DNA- and RNA-SIP have been used successfully in the past to determine metabolic activity in several bacterial communities using a variety of labelled compounds (Table 6.1).

Table 6.1. Overview of published SIP experiments in microbial communities

Bacterial community	Labelled compound	DNA/RNA SIP	Reference
Soil	Methanol	DNA	Radajewski et al., 2000
	CO <sub>2</sub>	DNA	Jia & Conrad, 2009
		RNA	Ostle et al., 2003
	Methane	DNA	Jia et al., 2019
	phenanthrene	DNA and RNA	Schwarz et al., 2018
	Acetate	DNA and RNA	Liu et al., 2018
	Propionate	RNA	Lueders et al., 2004
	Benzo[a]pyrene	DNA	Zhao et al., 2022
Industrial waste bioreactor	Phenol	RNA	Manefield et al., 2002
Activated sludge (wastewater treatment)	Phenol	RNA	Manefield et al., 2005
Rat caecal	Inulin	RNA	Tannock et al., 2014
Mouse caecum	Potato starch	RNA	Herrmann et al., 2018
Mouse faeces	Potato starch	RNA	Herrmann, et al., 2017a
	Glucose	RNA	Herrmann, et al., 2017b
Pig caecum	Neu5Ac	RNA	Young et al., 2015
<i>In vitro</i> small intestine fermentation model	Glucose	RNA	Egert et al., 2007
<i>In vitro</i> colon fermentation model	Starch	RNA	Kovatcheva-Datchary et al., 2009
	Galacto-oligosaccharide	RNA	Maathuis et al., 2012

Here we used RNA-SIP with enzymatically synthesised <sup>13</sup>C<sub>7</sub>-labelled 2,7-anhydro-Neu5Ac and <sup>13</sup>C<sub>7</sub>-labelled Neu5Ac to investigate the ability of members of the human gut microbiota to metabolise these sialic acids, using faecal slurry from human donors.

## 6.2 Results

### 6.2.1 Enzymatic synthesis of 2,7-anhydro-Neu5Ac from ManNAc and pyruvate

The enzymatic synthesis of sialic acid derivatives has been thoroughly investigated by the group of Xi Chen (W. Li et al., 2019; Xiao et al., 2018; H. Yu et al., 2004, 2006). Here,  $^{13}\text{C}_7$ -labelled Neu5Ac was first synthesised from  $^{13}\text{C}$ -labelled pyruvate and  $^{13}\text{C}_6$  ManNAc using a commercial sialic acid aldolase (W. Li et al., 2019). The follow-up reactions from Neu5Ac to 2,7-anhydro-Neu5Ac were performed in two steps based on *RgNanH* enzymatic hydrolysis of 3'-sialyllactose (3'SL), as described in Figure 6.1.

In the first step **(A)**, Neu5Ac was first converted into CMP-Neu5Ac using *Neisseria meningitidis* CMP-sialic acid synthetase (*NmCSS*) (Martin et al., 1998) and cytidine triphosphate (CTP), as previously described (W. Li et al., 2019). CMP-Neu5Ac together with lactose were then used as substrates of *Pasteurella multocida* sialyltransferase 3 (*PmST3*) (H. Yu et al., 2005), leading to the formation of 3'SL (Figure 6.1A). Lactose tagged with the nonpolar NHCbz (benzyl chloroformate) tag, lactose-NHCbz, was used to aid with subsequent purification. Following optimisation, this reaction was repeated with  $^{13}\text{C}$ -labelled Neu5Ac. The progress of the reaction was monitored by thin-layer chromatography (TLC) (Figure 6.2A). The TLC of the reaction showed spots corresponding to the starting materials, lactose-NHCbz and Neu5Ac, as well as the reaction product 3'SL-NHCbz (Figure 6.2A). The reaction was continued until completion or until no further product was obtained. A P2 size exclusion column, in which large molecules elute first, followed by a silica column, which separates molecules based on their polarity, was used to separate the reaction products. The nonpolar NHCbz tag both increases the size of the lactose molecule and makes it less polar, 3'SL-NHCbz is larger and less polar compared to Neu5Ac. Purification was confirmed by TLC and by electrospray ionisation-mass spectrometry (ESI-MS) (Figure 6.2B). A peak at was observed at 772.3 m/z, which is the expected mass corresponding to 3'SL-NHCbz plus 7 for the  $^{13}\text{C}$  carbons in Neu5Ac (Figure 6.2B). No other peaks were observed, indicating that the purification process led to a pure compound. However, the final yield of 3'SL was low (~5%) and the reaction therefore required further optimisation.

In the second step **(B)**, purified 3'SL-NHCbz was used as a substrate for *RgNanH* to produce 2,7-anhydro-Neu5Ac (Figure 6.1B). Following the reaction, a P2 size exclusion column

followed by a silica column, then another P2 size exclusion column was used to separate the reaction products, here the large and relatively non-polar lactose-NHCbz can be separated from the smaller polar 2,7-anhydro-Neu5Ac. Purification was confirmed by TLC, ESI-MS, and NMR (not shown). The purification process was further optimised to obtain a pure product with a yield of with a yield of 30% by Jun Yang Ong (Utrecht University).

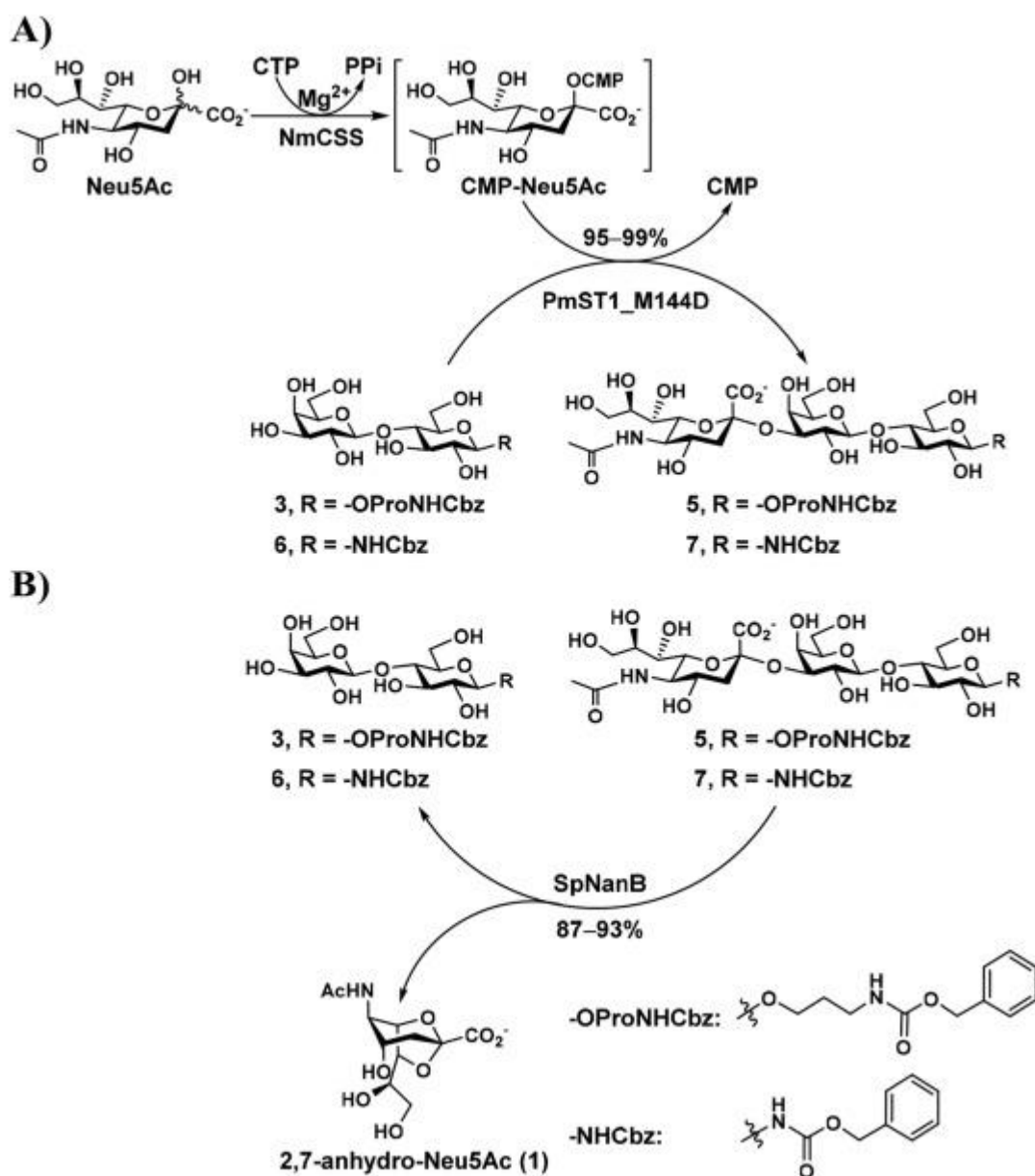
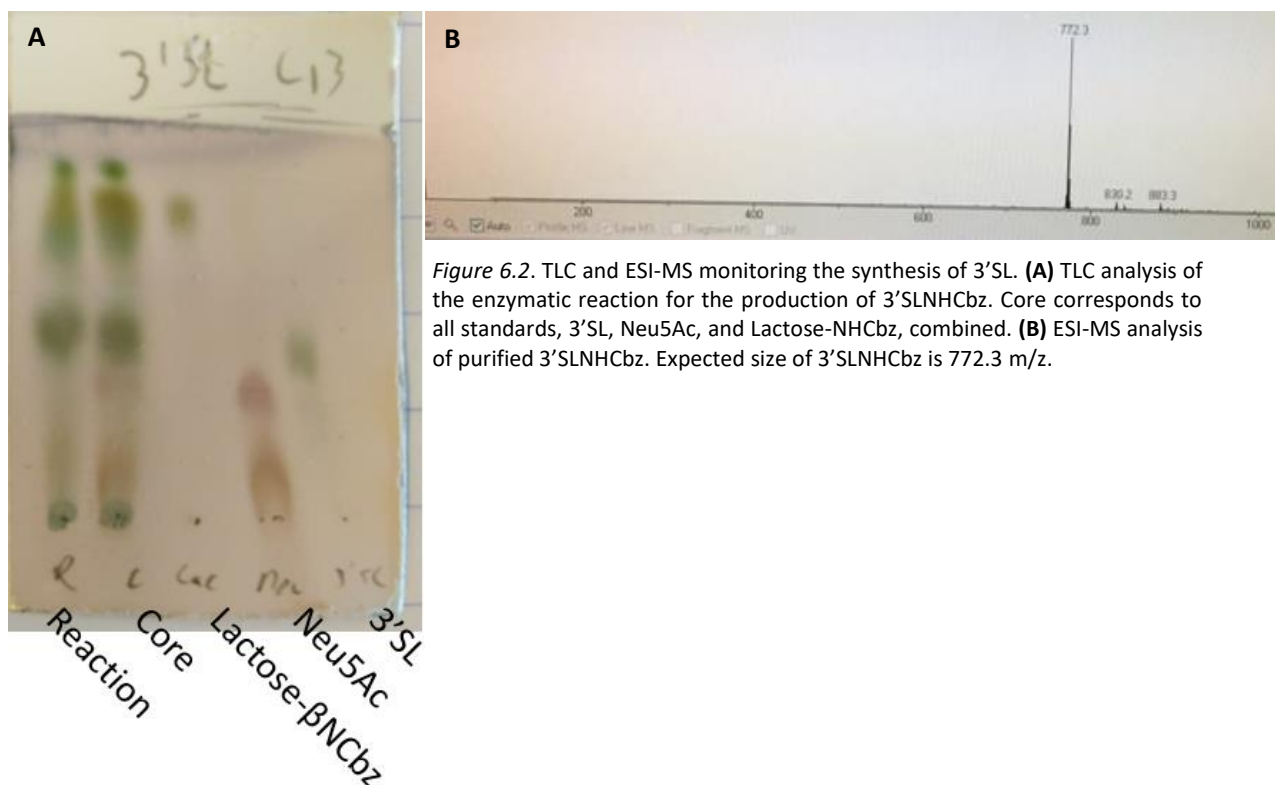


Figure 6.1. Chemical overview of the enzymatic synthesis of 2,7-anhydro-Neu5Ac from Neu5Ac. **(A)** Synthesis of 3'SL-NHCbz from Neu5Ac. **(B)** Synthesis of 2,7-anhydro-Neu5Ac from 3'SL-NHCbz. Figure reprinted from (W. Li et al., 2019), with permission from Elsevier. Our strategy differs from the one depicted in the figure in the following: *PmST3* was used instead of *PmST1\_M144D*, and *RgNanH* was used instead of *SpNanB*. Of the two nonpolar tags showed only Lactose-NHCbz was used.





## 6.2.2 Stable isotope probing of 2,7-anhydro-Neu5Ac metabolism in human faecal microbiota

### 6.2.2.1 Rate of 2,7-anhydro-Neu5Ac utilisation and selection of SIP conditions

During an RNA-SIP experiment, the labelled compound is metabolised over time, resulting in increased RNA labelling with longer incubation times. Sampling too early may not provide sufficient labelled RNA. However, due to cross-feeding, longer incubation times may result in labelled RNA in species which did not directly metabolise the labelled compound. It is therefore important to first determine the rate of 2,7-anhydro-Neu5Ac consumption by the bacteria in the faecal slurry. Here, faecal slurry from human donors was grown in batch media supplemented with either 11.1 mM 2,7-anhydro-Neu5Ac or 11.1 mM Neu5Ac, as described in section 2.2.2. After 24 hours growth under anaerobic conditions, 40% of 2,7-anhydro-Neu5Ac and 60% of Neu5Ac was found to be consumed (Figure 6.3).

In a preliminary SIP experiment, the faecal slurry was grown with 2 mg/ml  $^{13}\text{C}_7$  Neu5Ac and sampled at 4, 8, 12, and 24 hours. After RNA extraction, 'heavy' labelled RNA was separated from 'light' unlabelled RNA by ultracentrifugation using a caesium trifluoroacetate (CsTFA) gradient, as described in section 2.6. This gradient was fractionated and relative RNA concentration in each gradient was determined by qPCR using universal bacterial 16S

primers. An increase in RNA in heavier fractions was observed after 8 hours with an additional peak above a density of 1.81, while after 12 hours a single peak was detected, indicating cross-feeding (Figure 6.4). Based on these data, a sampling timepoint at 8 hours was chosen for RNA extractions in subsequent SIP experiments.

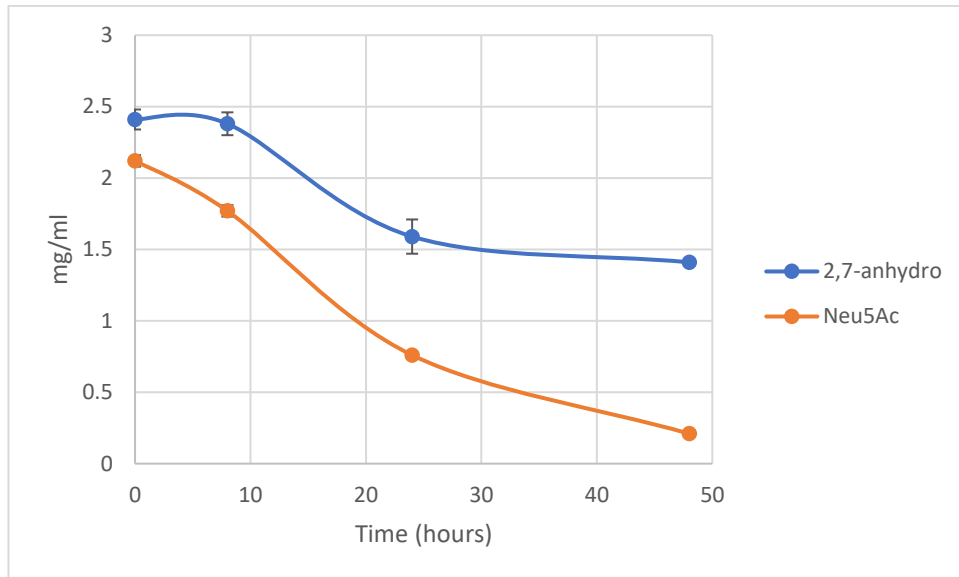


Figure 6.3. Utilisation of 2,7-anhydro-Neu5Ac and Neu5Ac by human faecal microbiota in batch model. Concentration is given in mg/ml. Average of three replicates is shown. Error bars indicate standard deviation.

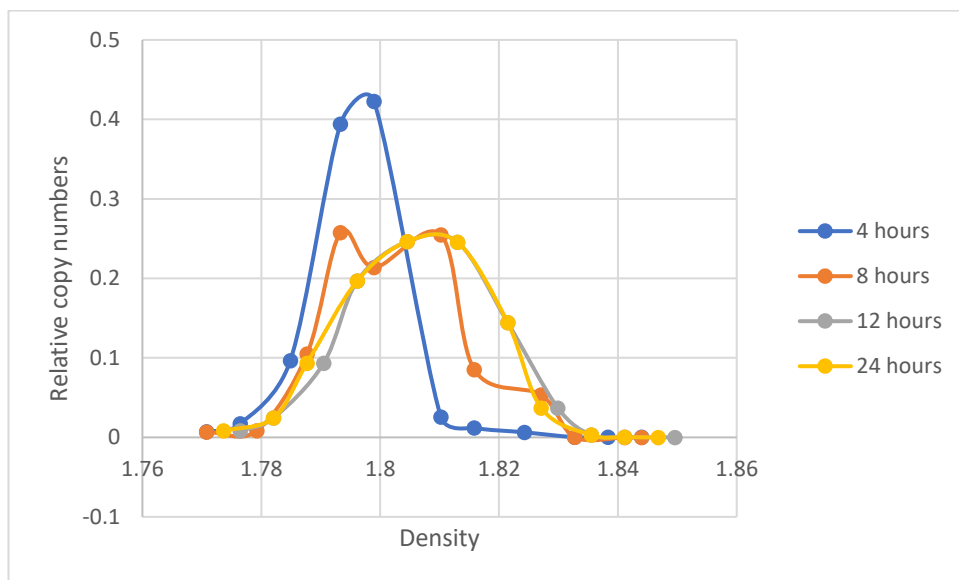


Figure 6.4. Relative 16S cDNA copy numbers per gradient fraction of faecal slurry grown on Neu5Ac. cDNA per fraction is shown for four different timepoints.

### 6.2.2.2 Stable isotope probing using $^{13}\text{C}_7$ labelled or unlabelled 2,7-anhydro-Neu5Ac and Neu5Ac

Human faecal slurry was grown in batch medium supplemented with both unlabelled ( $^{12}\text{C}$ ) or  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac and Neu5Ac. RNA was extracted after 8 hours of growth and separated as described above. The refractive index of the fractions showed that corresponding fractions from different gradients, for example the fifth fraction of each gradient, were consistent, (Figure 6.5) and subsequent fractions to form a consistent gradient ( $R^2=0.998$ ; Figure 6.5). 16S qPCR analysis showed two distinct peaks for the  $^{13}\text{C}$  samples, a peak in the fractions with lower density comprised of unlabelled RNA and a peak in fractions of high density comprised of partially  $^{13}\text{C}$  labelled RNA (Figure 6.6). Only a single peak for the  $^{12}\text{C}$  samples could be seen as expected since the RNA is not  $^{13}\text{C}$  labelled (Figure 6.6). All fractions with a density higher than 1.81 g/ml were pooled as “heavy” fractions, and fractions below a density of 1.81 g/ml were pooled as “light” fractions. All cDNA acquired of both heavy and light samples were sequenced using 16S amplicon sequencing.

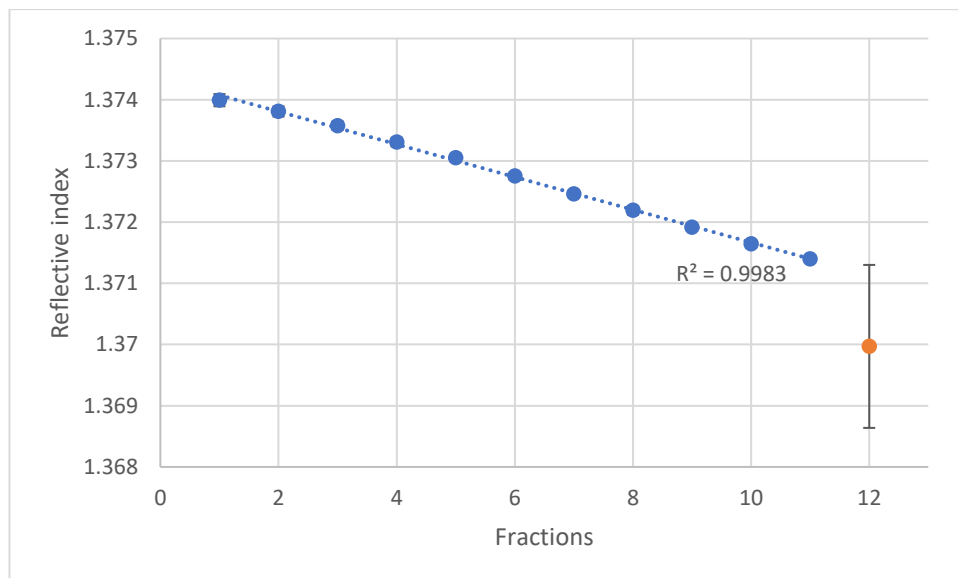
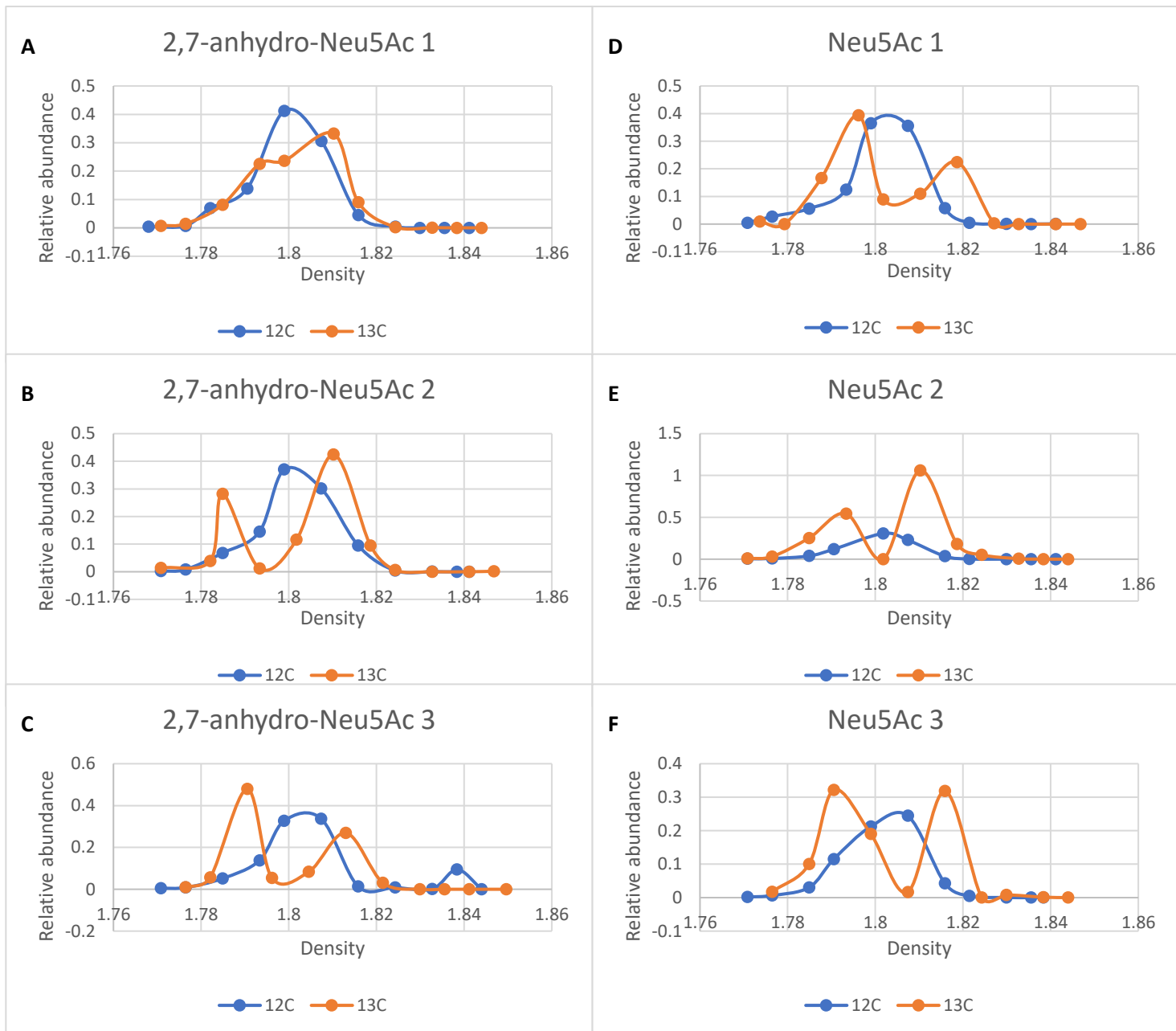


Figure 6.5. Overview of the average refractive index of SIP fractions. Average is taken over all  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac and Neu5Ac and unlabelled 2,7-anhydro-Neu5Ac and Neu5Ac fractions. Error bars indicate the standard deviation.  $R^2$  was calculated using only the first 11 fractions, which contain all RNA (blue dots).



**Figure 6.6.** Relative bacterial 16S cDNA copy numbers per SIP fraction. For each fraction the density and relative abundance of bacterial 16S genes are shown. **(A-C)** Relative bacterial 16S cDNA copy numbers of bacteria grown on  $^{13}\text{C}_7$  2,7-anhydro-Neu5Ac. **(D-F)** Relative bacterial 16S cDNA copy numbers of bacteria grown on  $^{13}\text{C}_7$  Neu5Ac.

### 6.2.2.3 Analysis of the gut microbiota composition

Principle component analysis (PCoA) showed a separation between the light fraction and heavy fraction of bacteria grown on  $^{13}\text{C}$  Neu5Ac, while both  $^{12}\text{C}$  fractions clustered together (Figure 6.7). For the samples grown on 2,7-anhydro-Neu5Ac this separation was less evident (Figure 6.7). Samples of bacteria grown on batch medium with no supplemented sugar for 0 or 8 hours showed a high degree of separation from bacteria grown on sialic acid derivatives (Figure 6.7).

At the phylum level, an increase of Actinobacteriota could be seen in the heavy fraction of  $^{13}\text{C}_7$  2,7-anhydro-Neu5Ac compared to the light fraction, while in the heavy fraction of  $^{13}\text{C}_7$  Neu5Ac both Firmicutes and Actinobacteriota were more abundant compared to the light fraction (Supplemental Figure S9).

Several Firmicutes species showed a large change in abundance in the heavy fraction of Neu5Ac growth and the light fraction of  $^{13}\text{C}_7$  Neu5Ac growth. An unknown *Holdemanella* species was highly increased in the heavy fraction of Neu5Ac (46-48% relative abundance of Firmicutes species) compared to the light fraction (16-25% relative abundance of Firmicutes species) (Figure 6.8C). In addition, *Dorea longicatena* relative abundance was 25-33% of Firmicutes species in the heavy Neu5Ac fraction compared to 7-12% in the light fraction. Changes in relative abundance were less pronounced in Bacteroidota and Proteobacteria phyla. An unknown *Dorea* species was also increased (9-12% of Firmicutes species in heavy fraction to 2-4% in light fraction) (Figure 6.8C). This was accompanied by a large increase of an unknown *Romboutsia* species in the light fraction (4-11% of all Firmicutes in heavy fraction; 41-56% in light fraction) and an unknown *Streptococcus* species (0.5-1.5% in heavy fraction; 5-8% in light fraction) (Figure 6.8C).

Some differences were also observed at the species level between heavy and light fractions after growth on  $^{13}\text{C}_7$  Neu5Ac in other species. For example, *Collinsella aerofaciens* showed a relative abundance of total actinobacteria of 36-53% in the heavy fraction and 9-14% in the light fraction (Figure 6.8A). This coincided with a decrease in an unknown *Bifidobacterium* species from 74-79% in the light fraction to 39-54% in the heavy fraction, however, this change was much less pronounced in the read numbers, indicating the decrease in relative abundance was mostly due to the increase of *C. aerofaciens*. This

indicates *C. aerofaciens* is able to metabolise Neu5Ac better than most other bacteria in the samples.

Unlike Neu5Ac, only minor differences were observed between heavy and light fractions after growth on  $^{13}\text{C}_7$  2,7-anhydro-Neu5Ac (Figure 6.8). Only an unknown *Holdemanella* species and *C. aerofaciens* showed small differences in relative abundance (Figure 6.8 A and C). The *Holdemanella* species showed a small increase in relative abundance of Firmicutes species (12-13% in the light fraction to 14-17% in the heavy fraction) (Figure 6.8C) whereas *C. aerofaciens* showed a doubling in relative abundance of Actinobacteria species, 7-9% in the light fraction compared to 14-15% in the heavy fraction (Figure 6.8A). The lack of significant changes following supplementation with  $^{13}\text{C}$ -labelled or unlabelled 2,7-anhydro-Neu5Ac cannot be due to labelling issues as shown by qPCR analysis and read numbers from 16S sequencing which demonstrated differences between heavy and light fractions.

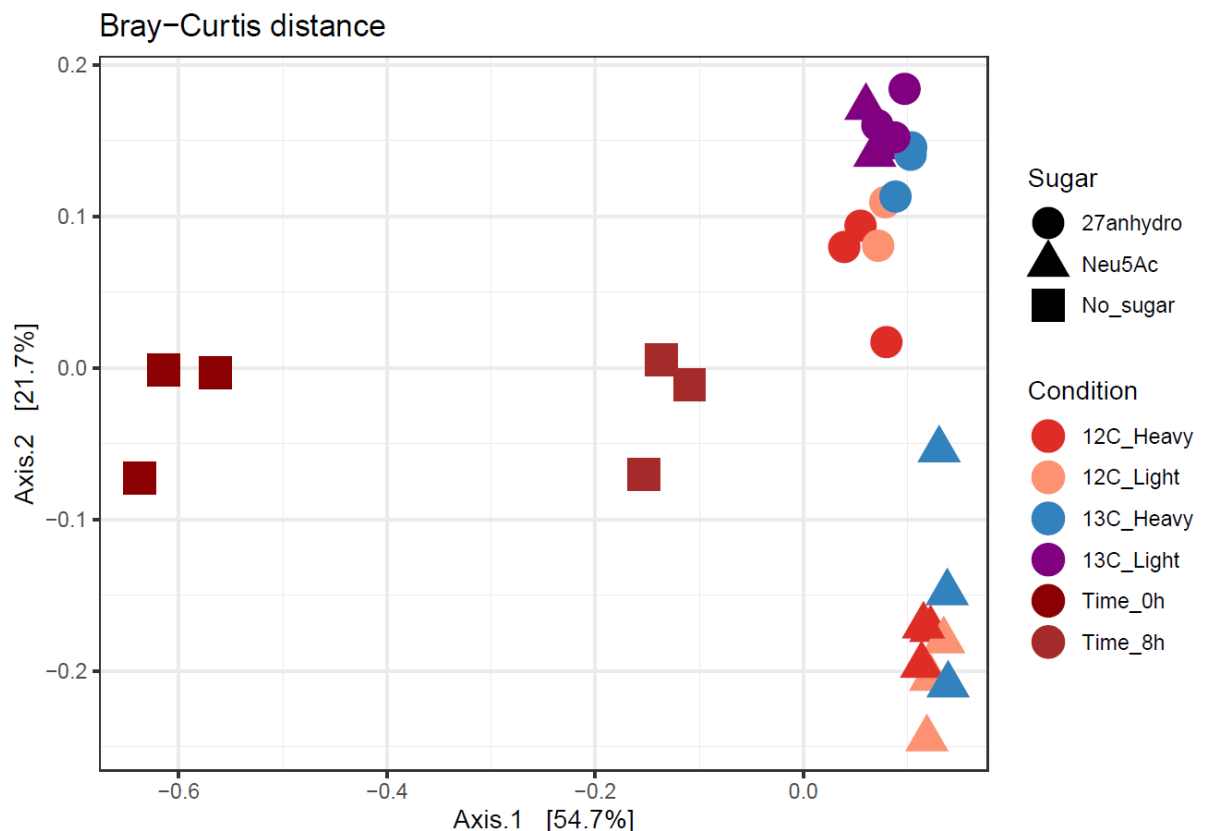
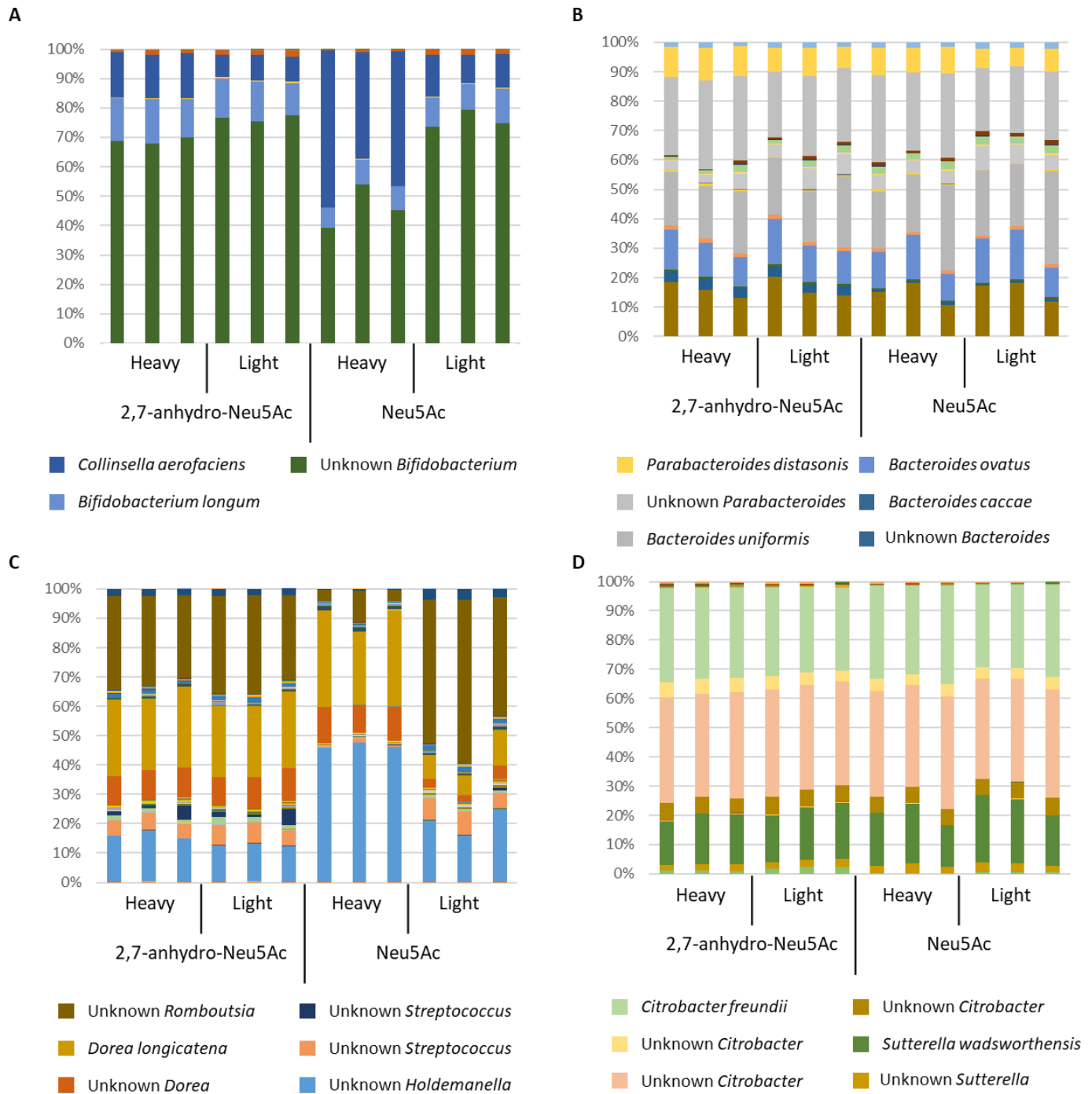


Figure 6.7. PCoA plots of the heavy and light fractions of bacteria grown on  $^{12}\text{C}$  and  $^{13}\text{C}$  Neu5Ac and 2,7-anhydro-Neu5Ac. No sugar 0 hour and 8 hour controls are indicated. Data analysed and figure made by Helena Pardo Casales.



**Figure 6.8.** Relative abundance of bacterial species in heavy and light SIP fractions per major bacterial phyla of faecal microbiota grown on sialic acid derivatives. **(A)** Relative abundance of Actinobacteria species per SIP fraction **(B)** Relative abundance of Bacteroidota species per SIP fraction **(C)** Relative abundance of Firmicutes species per SIP fraction **(D)** Relative abundance of Proteobacteria species per SIP fraction. Heavy fractions contain RNA from all gradient fractions with a density of more than 1.81 g/ml, light fractions contain RNA from all gradient fractions with a density of less than 1.81 g/ml. 2,7-anhydro-Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac. Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled Neu5Ac. Figures with complete legends are shown in Supplemental Figures S10-S13.

### 6.3 Discussion

The SIP results showed an increase in *C. aerofaciens*, *D. longicatena* relative abundance as well as of an unknown *Holdemanella* species following supplementation of human gut microbiota with  $^{13}\text{C}_7$ -Neu5Ac. *C. aerofaciens* is a Gram-positive obligate anaerobe that is the most abundant actinobacterium in a healthy human gut (Bag et al., 2017; Tourlousse et al., 2020) but also associated with a number of diseases including IBS (Malinen et al., 2010), IBD (Joossens et al., 2011), and rheumatoid arthritis (J. Chen et al., 2016). However, little information is available on the role of sialic acid metabolism by this species in the human gut. Previous studies in piglets identified *C. aerofaciens* as one of the species being significantly increased in the gut of 6'-SL fed piglets (Jacobi et al., 2016). In addition, metatranscriptomics analyses of caecal contents showed that *C. aerofaciens* MC1 and *C. aerofaciens* MC2 strains were among the top 5 species with most significant differently expressed genes in piglets harbouring a consortium of gut bacteria from Malawian babies, fed with a diet supplemented with sialylated bovine milk oligosaccharides (Charbonneau et al., 2016). Another species which we found to be highly induced during growth on  $^{13}\text{C}_7$  Neu5Ac is *D. longicatena*, another Gram-positive obligately anaerobe. Not much information is available on the role of this species in the human gut but it encodes a *nan* cluster dedicated to sialic acid metabolism (Almagro-Moreno & Boyd, 2009). Therefore, the expansion of *C. aerofaciens* and *D. longicatena* in Neu5Ac SIP is in line with these previous reports. It is also of note that both *D. longicatena* and *C. aerofaciens* have been shown to be increased in obese individuals (Companys et al., 2021; Gallardo-Becerra et al., 2020).

However, our analysis did not reveal induction of other known Neu5Ac utilisers of the human gut. For example, *R. inulinivorans* and *F. prausnitzii*, which have both been shown to benefit from Neu5Ac released by *A. muciniphila* when in co-culture *in vitro* (Shuoker et al., 2023), were detected in the starting material and after 8 hours in the batch fermentation without added sugar, these bacteria species could not be detected in the conditions supplemented with sialic acid derivatives. *R. gnavus* was only found at low levels in samples containing  $^{12}\text{C}$  Neu5Ac and  $^{12}\text{C}$  2,7-anhydro-Neu5Ac. The genus *Bifidobacterium* was highly abundant in all conditions, however its abundance was reduced in the heavy fraction of the bacteria grown on Neu5Ac, indicating this bacteria was outcompeted by other species. However, the analysis performed did not provide



information at the strain level, or in cases even species level, and this level of resolution would be needed to better capture strains implicated in sialic acid metabolism. For example, some *Bifidobacterium* species such as *B. bifidum* are able to release Neu5Ac, but not metabolise it, while others such as *B. breve* are not able release sialic acid but are able to utilise it through cross-feeding (Bunesova et al., 2018; Egan et al., 2014; Nishiyama et al., 2018).

The faecal sample used for the SIP experiment was exposed to oxygen during processing. This might result in some strict anaerobes not being able to recover and grow during the anaerobic growth conditions. Many Proteobacteria are obligatory anaerobes, which explains the increase in abundance of Proteobacteria between 0 and 8 hours of growth in the media without added sugars (Supplemental Figure S9). Future experiments should take care to keep the faecal sample anaerobic.

Previous SIP aimed at identifying Neu5Ac utilisation in pig caecal content found only the genus *Prevotella* to be in a higher abundance in the heavy fraction of caecal content grown on  $^{13}\text{C}_3$  Neu5Ac (Young et al., 2015). However, it should be noted that this study used  $^{13}\text{C}_3$  Neu5Ac rather than  $^{13}\text{C}_7$  Neu5Ac (our work), which would result in a lower percentage of labelled DNA or RNA, and an incubation time of 24 hours rather than 8 hours (our work), potentially leading to further cross-feeding.

The nature of a SIP experiment requires early sample times to avoid cross-feeding, but this might favour fast-growing Neu5Ac metabolisers over slow-growing Neu5Ac metabolisers. In an experiment investigating faecal bacteria growth rates, both *C. aerofaciens* and *D. longicatena* were found to be dominating in conditions favouring fast growing species (Adamberg & Adamberg, 2018). Here, it is likely that the SIP conditions led to slow-growing Neu5Ac metabolisers to be outcompeted by *C. aerofaciens* and *D. longicatena*. In the future, it would be interesting to carry out SIP experiments using conditions promoting different Neu5Ac utilisers.

In contrast to Neu5Ac, no large difference was observed in the heavy and light fractions of faecal slurry grown on  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac. Small increases in relative abundance in both an unknown *Holdemanella* species and *C. aerofaciens* were found in the heavy fraction of 2,7-anhydro-Neu5Ac, indicating that these species might be able to

metabolise 2,7-anhydro-Neu5Ac faster compared to other species. In addition, it may be informative to carry out RNAseq analysis on the RNA-SIP fractions to help determine common or unique Neu5Ac pathways utilised by the gut microbiota.

In the future, other approaches may be applied to complement this work. Recently, D<sub>2</sub>O was used to determine metabolism of mucin *O*-glycan monosaccharides including Neu5Ac by mouse colonic bacteria, by monitoring the incorporation of deuterium in bacterial peptides (Pereira et al., 2020). SIP could also be applied to address the metabolism of complex mucin glycans by the gut microbiota using <sup>13</sup>C labelled mucin which has been shown to be produced by a lung adenocarcinoma cell line grown on <sup>13</sup>C<sub>6</sub> glucose (Evert et al., 2019).

## Chapter 7

### General discussion

## 7. General discussion

The capacity of *R. gnavus* strains to grow on mucins is strain specific (Crost et al., 2013) and associated with GHs targeting mucin glycan epitopes such as fucose and sialic acids (Tailford, Owen, et al., 2015; H. Wu, Rebello, et al., 2021). For a full overview of the substrates *R. gnavus* strains can utilise, see Supplemental Table S6.

In the first part of this thesis, we focused on *RgGH98*, an enzyme which cleaves the Gal $\beta$ 1-4GlcNAc linkages in BgA epitopes, releasing BgA trisaccharide (BgAtri). BgA epitopes are found in terminal location of mucin glycan chains (Brockhausen et al., 2022), and are used by pathogens as recognition sites (Cooling, 2015; Ruvoën-Clouet et al., 2013). The gene encoding *RgGH98* is only present in *R. gnavus* strains growing on mucin such as *R. gnavus* ATCC 29149 and ATCC 35913, and absent in the E1 strain (Crost et al., 2013, 2016). We showed that while only *R. gnavus* ATCC 29149 could grow on BgA tetrasaccharide (BgAtetra), both *R. gnavus* E1 and ATCC 29149 strains could grow on BgAtri, indicating that these strains could metabolise the product of the *RgGH98* enzymatic reaction. In line with this result, we showed that *R. gnavus* E1 could grow on *RgGH98*-pretreated mucin, benefiting from both the released BgAtri and from access to the underlying mucin chain. In *R. gnavus* ATCC 29149, qPCR analysis confirmed the transcription of the *RgGH98* cluster during growth on BgAtri and BgAtetra. These data further support the model that *R. gnavus*' strategy to adapt to mucus is through foraging mucin glycan epitopes. Interestingly, this cluster also contains a predicted GH95 fucosidase, and a predicted GH73 peptidoglycan hydrolase. In the future, it will be interesting to characterise the function of these GHs in order to confirm the metabolic importance of this cluster, especially during mucin glycan metabolism. In addition, *RgGH98* may have potential application in therapeutics to limit infection by pathogens targeting BgA epitopes such as Sars-Cov-2 (S. Wu et al., 2022).

This work also underscores the importance of characterising gut commensal bacteria at the strain level. The importance of strain level resolution in analyses of the metabolic capabilities has been demonstrated in other gut bacteria such as *A. muciniphila* where the H2 strain has a reduced complement of GHs involved in mucin glycan metabolism compared to other strains which may lead to different health effects (Ouwerkerk et al., 2022) or *Bifidobacterium* strains, which genomic analyses showed significant variation in

GH diversity depending on their host and environment (Rodriguez & Martiny, 2020; Turrone et al., 2019).

To identify the complement of GHs involved in the process of mucin glycan degradation, RNAseq was carried out on *R. gnavus* ATCC 29149 grown on mucin glycans. Here, we found 717 genes to be significantly upregulated, and 652 genes significantly downregulated, as compared to growth on glucose. Among the significantly upregulated genes, 22 were GH-encoding genes. Of these, GH33 was the most expressed together with the *nan* cluster dedicated to 2,7-anhydro-Neu5Ac metabolism, in line with previous data showing the importance of the GH33 intramolecular *trans*-sialidase (IT-sialidase) and *nan* cluster in *R. gnavus* adaptation to the mucus niche (Bell et al., 2019, 2020). Sialic acid plays an important role in the gut microbiota for food and adhesion (Bell & Juge, 2021; Sokolovskaya et al., 2022). *R. gnavus* ATCC 29149 IT-sialidase releases 2,7-anhydro-Neu5Ac from sialylated glycoconjugates including mucins (Tailford, Owen, et al., 2015). The ability to utilise Neu5Ac is widespread in mucosa-associated bacteria (Bell & Juge, 2021) and pathogens (K. Huang et al., 2015; Ng et al., 2013). In contrast, the ability to release, transport, and metabolise 2,7-anhydro-Neu5Ac is unique to *R. gnavus* in the gut environment (Bell et al., 2020), conferring *R. gnavus* ATCC 29149 a nutritional advantage in scavenging the available Neu5Ac in mucin epitopes and helping *R. gnavus* to colonise the mucus niche (Bell et al., 2019). Here, SIP experiments based on human faecal microbiota provided new insights into the capacity of microbes to utilise Neu5Ac or 2,7-anhydro-Neu5Ac. The results suggested that that *C. aerofaciens* and *D. longicatena* could metabolise Neu5Ac at a higher rate compared to other gut microbes whereas no significant changes in gut microbiota was observed upon 2,7-anhydro-Neu5Ac supplementation. Future shotgun analysis of RNAseq analysis of the SIP samples could be done to identify the metabolic pathways expressed in the microbial community upon supplementation with these sialic acid derivatives.

In addition, we showed that *R. gnavus* ATCC 29149 GH29 and GH95 fucosidases, previously shown to hydrolyse 2'FL, 3FL, LeA, and BgAtri (H. Wu, Crost, et al., 2021; H. Wu, Rebello, et al., 2021), were also active on mucin. In humans, mucin glycan fucosylation decreases from the ileum to the rectum (Robbe et al., 2003). An opposite gradient is found for sialic acid epitopes in mucin (Robbe et al., 2003). In the ileum and caecum the highest quantity of BgA epitopes can be found (Robbe et al., 2004). The capacity of *R. gnavus* ATCC 29149 to utilise

fucose, sialic acid and BgA epitopes is likely to underpin *R. gnavus*' fitness to colonise the GI tract and its prevalence in adults (J. Qin et al., 2010). In addition, *R. gnavus* has been shown to be consistently elevated in IBD (Croft et al., 2023). Ulcerative colitis is characterised by a decrease in complex mucin glycan chains (Holmén Larsson et al., 2011) which may favour *R. gnavus*' mucin glycan degradation strategy as compared to *A. muciniphila* which is decreased in UC (Earley et al., 2019; Zakerska-Banaszak et al., 2021) and more adapted to the degradation of complex mucin glycan chains (Cani et al., 2022).

Interestingly, the RNAseq analysis of *R. gnavus* ATCC 29149 grown on mucin glycans, also identified a highly induced prophage, which may indicate a role for this prophage in mucin glycan foraging. Interestingly, recent work in mice showed that phages co-exist with *R. gnavus* population in the gut, and that patients with IBD have an over-representation of *R. gnavus* virome (Buttimer et al., 2023). In the future, it will be interesting to study prophage induction during growth on mucin glycan and the role of these phages in the ability of gut bacteria to colonise the mucus niche in health and disease.

In addition to their ability to utilise mucin glycans, *R. gnavus* strains have also been tested on range of HMOs (Croft et al., 2013), and dietary oligosaccharides such as raffinose oligosaccharides (Bruel et al., 2011) and starch degradation products (Croft et al., 2018) (see Supplemental Table S6).

Here, we investigated the substrate specificity of two *R. gnavus* ATCC 29149 GH13 enzymes. We showed that *RgGH13-1* could hydrolyse  $\alpha$ 1-4 links in pullulan, starch, glycogen, maltose, maltotriose, panose,  $\alpha$ -cyclodextrin and  $\beta$ -cyclodextrin, while *RgGH13-2* was found to hydrolyse  $\alpha$ 1-6 links in pullulan. *RgGH13-2* was also found to be active on glycogen, but the specific activity on this substrate was not determined. *R. gnavus* ATCC 29149 is not able to grow on pullulan, starch, or glycogen. Since these GH13 enzymes are predicted to be intracellular, they may fulfil a different role, for example in glycogen metabolism or in the catabolism of oligosaccharides released by other bacteria during cross-feeding on starch. In the competitive gut environment, the ability to manage energy during both nutrient rich and nutrient poor conditions is essential. Glycogen is one of the compounds used in bacteria for energy storage. For example, *Lactobacillus acidophilus* was shown to increase glycogen content during growth on raffinose, trehalose, and lactose, while glycogen biosynthesis was repressed by glucose (Goh & Klaenhammer, 2013).

However, in other bacteria it was found that growth on glucose led to an increase of glycogen (Bonafonte et al., 2000). Depending on the metabolic niche and substrate availability, conditions may favour bacterial expansion or storage of energy in preparation for energy sparse conditions. This was demonstrated in *E. coli* BW25113 where glycogen reserves were found to be beneficial during changing nutrient sources *in vitro* (Sekar et al., 2020). Here, we showed that during growth of *R. gnavus* on mucin glycans, glycogen synthesis genes were upregulated. This was paralleled with an upregulation of several GH13 enzymes, as also shown in a previous study using mucin as source of carbohydrate (Crost et al., 2018). In the future, it would be interesting to monitor glycogen contents in *R. gnavus* ATCC 29149 during growth on mucin glycans. Knowledge on *R. gnavus* ATCC 29149 glycogen biosynthesis and metabolism during growth on mucin glycans, combined with further functional characterisation of *R. gnavus* GH13 enzymes on glycogen will help further advance our understanding of the adaptation and persistence of *R. gnavus* ATCC 29149 to the gut as a way to regulate nutrient storage and mobilisation upon dietary changes in the gut environment. The ability of *R. gnavus* ATCC 29149 to grow on a wide range of maltooligosaccharides should also be further investigated, specifically oligosaccharides released during cross-feeding on starch. Further analysis of the specificity of transporters and *RgGH13-1* and *RgGH13-2* will further determine their role in starch cross-feeding.

Together these results showed the importance of GHs in *R. gnavus*' capacity to colonise different nutritional niches in health and disease and identified new enzymes and pathways that may be involved in *R. gnavus* adaptation to the mucus niche, opening the way for novel avenues of research.

## Appendix:

### Appendix 1. Supplemental material

Supplemental Table S1. Overview of characterised GHs from gut bacteria active on mucin glycans.

Name	GH family and function	Species	Mucin substrates tested	reference
<i>AmGH29C</i>	GH29 Fucosidase	<i>A. muciniphila</i> BAA-835	Mix of PGM, PCM, and BF*	(Shuoker et al., 2023)
<i>AmGH29D</i>	GH 29 Fucosidase	<i>A. muciniphila</i> BAA-835	Mix of PGM, PCM, and BF*	(Shuoker et al., 2023)
<i>RgGH29-1</i>	GH29 Fucosidase	<i>R. gnavus</i> ATCC 29149	pPGM mucin glycans and PGM mucin glycans	This work
<i>RgGH29-2</i>	GH29 Fucosidase	<i>R. gnavus</i> ATCC 29149	pPGM mucin glycans and PGM mucin glycans	This work
<i>Afc2</i>	GH29 Fucosidase	<i>C. perfringens</i> ATCC 13124	PGM	(Fan et al., 2016)
<i>AmGH95A</i>	GH95 Fucosidase	<i>A. muciniphila</i> BAA-835	Mix of PGM, PCM, and BF*	(Shuoker et al., 2023)
<i>AmGH95B</i>	GH95 Fucosidase	<i>A. muciniphila</i> BAA-835	Mix of PGM, PCM, and BF*	(Shuoker et al., 2023)
<i>RgGH95</i>	GH95 Fucosidase	<i>R. gnavus</i> ATCC 29149	pPGM mucin glycans and PGM mucin glycans	This work
<i>Afc3</i>	GH95 Fucosidase	<i>C. perfringens</i> ATCC 13124	PGM	(Fan et al., 2016)
<i>SiaBb2</i>	GH33 Sialidase	<i>B. bifidum</i> JCM1254	pPGM	(Kiyohara et al., 2011)
<i>AmGH33A</i>	GH33 Sialidase	<i>A. muciniphila</i> BAA-835	PCM and Mouse MUC2	(Shuoker et al., 2023)
<i>AmGH33B</i>	GH33 Sialidase	<i>A. muciniphila</i> BAA-835	PCM and Mouse MUC2	(Shuoker et al., 2023)
<i>AmGH181</i>	GH181 Sialidase	<i>A. muciniphila</i> BAA-835	PCM and Mouse MUC2	(Shuoker et al., 2023)
<i>Amuc_0824</i>	GH2 Galactosidase	<i>A. muciniphila</i> BAA-835	Defucosylated, desialylated, and desulphurylated PGM	(Kosciow & Deppenmeier, 2020)
<i>Amuc_0771</i>	GH35 Galactosidase	<i>A. muciniphila</i> BAA-835	Defucosylated, desialylated, and desulphurylated PGM	(Kosciow & Deppenmeier, 2020)



BT2824	GH16 Galactosidase	<i>B. thetaiotaomicron</i> VPI- 5482	porcine small intestinal (SI) and PGM	(Crouch et al., 2020)
BF4058	GH16 Galactosidase	<i>B. fragilis</i> NCTC9343	SI and PGM	(Crouch et al., 2020)
BF4060	GH16 Galactosidase	<i>B. fragilis</i> NCTC9343	SI and PGM	(Crouch et al., 2020)
BACCAC_02679	GH16 Galactosidase	<i>B. caccae</i> ATCC 43185	SI and PGM	(Crouch et al., 2020)
BACCAC_02680	GH16 Galactosidase	<i>B. caccae</i> ATCC 43185	SI and PGM	(Crouch et al., 2020)
BACCAC_03717	GH16 Galactosidase	<i>B. caccae</i> ATCC 43185	SI and PGM	(Crouch et al., 2020)
Amuc_0724	GH16 Galactosidase	<i>A. muciniphila</i> BAA-835	SI and PGM	(Crouch et al., 2020)
Amuc_875	GH16 Galactosidase	<i>A. muciniphila</i> BAA-835	SI and PGM	(Crouch et al., 2020)
Amuc_2108	GH16 Galactosidase	<i>A. muciniphila</i> BAA-835	SI and PGM	(Crouch et al., 2020)
RgGH98	GH98 Galactosidase	<i>R. gnavus</i> ATCC 29149	pPGM	(H. Wu, Crost, et al., 2021); This work
E-ABase	GH98 Galactosidase	<i>C. perfringens</i> ATCC 10543	PGM	(Anderson et al., 2005)
AgaBb	GH110 Galactosidase	<i>B. bifidum</i> JCM 1254	Salivary mucin	(Wakinaka et al., 2013)
SGL	GH20 <i>N</i> - acetylglucosaminidase	<i>Prevotella</i> strain RS2	PGM	(Rho et al., 2005)
AgnC	GH89 <i>N</i> - acetylglucosaminidase	<i>C. perfringens</i> strain 13	PGM	(M. Fujita et al., 2011)
GngC	GH89 <i>N</i> - acetylglucosaminidase	<i>B. bifidum</i> JCM 1254	PGM	(Shimada et al., 2015)
EngCP	GH129 <i>N</i> - acetylgalactosaminidase	<i>B. bifidum</i> JCM 1254	GH pre-treated PGM	(Ashida et al., 2008)

\*BF = bovine fetuin

Supplemental Table S2. Primers used in this study.

<b>qPCR confirming growth of <i>R. gnavus</i> on pPGM, RgGH98-treated pPGM, BgAtri, and BgAtetra (3.2.2.2; 3.2.3):</b>		
16SRg5	<b>F:</b> TGGCGGCGTGCTTAACA	(Joossens et al., 2011)
Universal <i>R. gnavus</i> 16S	<b>R:</b> TCCGAAGAAATCCGTCAAGGT	
RUMGNA_02237	<b>F:</b> GCTTTATGGTCTTGCACT	Kindly provided by
<i>R. gnavus</i> ATCC 29149 specific primer	<b>R:</b> AGAAAGCGTGGGTGGTATTG	Dr Emmanuelle Crost
RUGNEv3_60291	<b>F:</b> CCGAGTCCACATCAAATAGC	Kindly provided by
<i>R. gnavus</i> E1 specific primer	<b>R:</b> CAAGTCTCTGTTTTCTCCACTGTA	Dr Emmanuelle Crost
<b>RgGH98 operon transcriptomics RT-qPCR (3.2.3):</b>		
Gyrase B ATCC-00867-c	<b>F:</b> GGAGCAGACCAGATCCAAAT <b>R:</b> CCAATATACATTCCCGGTCTTT	(Crost et al., 2013)
RS10315-2	<b>F:</b> TTCCCGGGGACATTGATCAA <b>R:</b> GCTTTTGACCAGCCAGTTGA	Kindly provided by Dr Emmanuelle Crost
RS10320-2	<b>F:</b> CTGCGCCGAAAGTGTTTAT <b>R:</b> AGCCTTTCCTGACTCTGTCC	Kindly provided by Dr Emmanuelle Crost
RS10325-1	<b>F:</b> AGAAGGGAGCAAAGCCAGAT <b>R:</b> TCCTCTGCCAGATTGACCAG	<b>F:</b> Kindly provided by Dr Emmanuelle Crost <b>R:</b> (H. Wu, Crost, et al., 2021)
RS10330-2	<b>F:</b> AAGTTGCCGGGGAAGTGATA <b>R:</b> TGAACAGCATCCGGATCGTA	<b>F:</b> (H. Wu, Crost, et al., 2021) <b>R:</b> Kindly provided by Dr Emmanuelle Crost
RS10335-1	<b>F:</b> ATCACAGGCGGGATTCTTCT <b>R:</b> TCATTCTTTTCGGCACAGC	<b>F:</b> (H. Wu, Crost, et al., 2021) <b>R:</b> (H. Wu, Crost, et al., 2021)
RS10340-1	<b>F:</b> GGAGGTACGAAGGGCATGTA <b>R:</b> CCGTCCATTGCCGTATTTT	<b>F:</b> (H. Wu, Crost, et al., 2021) <b>R:</b> (H. Wu, Crost, et al., 2021)
RS10345-2	<b>F:</b> TGCGTATAGTCTTTGCGTTTGT <b>R:</b> TCATCATAAACTGCTTTCATGCC	This study
RS10350-1	<b>F:</b> TGCGAGGTACAGAAAGATGGT <b>R:</b> CCACTTCGCCATCATTACTTCC	<b>F:</b> (H. Wu, Crost, et al., 2021) <b>R:</b> (H. Wu, Crost, et al., 2021)
RS10355-1	<b>F:</b> ACCTGGGCTCCGAAAGAATT <b>R:</b> ACACCCGTTGCTAATCCCAA	<b>F:</b> (H. Wu, Crost, et al., 2021) <b>R:</b> Kindly provided by Dr Emmanuelle Crost
RS10360-2	<b>F:</b> AGGGATGCCAGATTCAGTGT <b>R:</b> TCGCCGCTATCCATGTCATA	This study
<b>RgGH29-2 expression confirmation RT-qPCR (4.2.2.3):</b>		
RGna_RS04290_RG6_2	<b>F:</b> TGGCATCATCCGGACTATCC <b>R:</b> CCGGCTGATGTTTTCTGACC	This study
<b>Cloning primers (4.2.3):</b>		

RGna_RS14450_NdeI_F	<b>F:</b> GCATCGCATATGATGAGACCCAAGTATCAT	This study
RGna_RS14450_XhoI_R	<b>R:</b> GCATCGCTCGAGCTATGCCTCCATCCTCCATATTTT	
RGna_RS07225_NheI_F	<b>F:</b>	This study
RGna_RS07225_XhoI_R	GCATCGGCTAGCGGAAGAGAGAAAGAGATCAATGC <b>R:</b> GCATCGCTCGAGTCACTGCTTTTTCTCCTTTGC	
RGna_RS14905_NdeI_F	<b>F:</b>	This study
RGna_RS14905_XhoI_R	GCATCGCATATGATGAAGATATGTTTACAGGCAGGAG <b>R:</b> GCATCGCTCGAGTTACATTCCGATCATGTC	
RGna_RS06060_NdeI_F	<b>F:</b>	This study
RGna_RS06060_XhoI_R	GCATCGCATATGATGAATTTTACAACAGTTTTGCATC <b>R:</b> GCATCGCTCGAGTTATTCAAGTTTCTTTAAAAATCAG	
RGna_RS08510_NheI_F	<b>F:</b>	This study
RGna_RS08510_XhoI_R	GCATCGGCTAGCATGGGAAAGACAGCAAAGCAATGG <b>R:</b> GCATCGCTCGAGTTACCTGCTCTCTCTTTTTTCC	
RGna_RS14690_NdeI_F	<b>F:</b>	This study
RGna_RS14690_XhoI_R	GCATCGCATATGATGATTAAGAAGATAAATATTGAAG <b>R:</b> GCATCGCTCGAGTTACCACCATACCTTCCAGTCTTTC	
RGna_RS07930_NheI_F	<b>F:</b> GCATCGGCTAGCATGAGACGAAAAATAGAA	This study
RGna_RS07930_XhoI_R	<b>R:</b> GCATCGCTCGAGTTAAGAATTTTTTACATGCC	
RGna_RS02745_NheI_F	<b>F:</b> GCATCGGCTAGCATGATTATTCAGAGACAT	This study
RGna_RS02745_XhoI_R	<b>R:</b> GCATCGCTCGAGCTGTAAGGAAGACAGGGTATAA	
<b>Cloning primers for <i>RgGH13-2</i> CBM48 domain (5.2.2)</b>		
RGna_RS14690_NdeI_F	<b>F:</b>	This study
GH13_2_XhoI_CBM48_R	GCATCGGCTAGCATGGGAAAGACAGCAAAGCAATGG <b>R:</b> GCATCGCTCGAGCGTTTTTCTCAAATCCACCGCC	
<b>Stable Isotope Probing fraction qPCR (6.2.2)</b>		
16S Universal	<b>F:</b> AGAGTTTGATCMTGGCTCAG <b>R:</b> ACGGCTACCTGTTCAGACTT	<b>F:</b> (Heuer et al., 1997) <b>R:</b> (Weisburg et al., 1991)

Numbers in brackets correspond to the thesis chapter where their use is described.

*Supplemental Table S3.* RNA concentration and quality extracted from *R. gnavus* ATCC 29149 growth on 0.5% (27.7 mM) glucose, 1 mM BgAtri, or 1 mM BgAtetra.

Sample	RNA Concentration (ng/ $\mu$ l)	Tapestation RINe
Glucose T=7	16.7	8.1
BgAtri T=11	1.06	7.4
BgAtetra T=10	2.05	8.7
Glucose T=24	2.22	7.9
BgAtri T=24	6.24	8.5
BgAtetra T=24	<2	7.5

Extraction volume was 3 ml. RNA was extracted at late exponential phase (T=7-11 hours) and after 24 hours. RNA concentration was measured by Qubit after DNase treatment. RINe was determined by Tapestation analysis.

*Supplemental Table S4.* RNA concentration and quality extracted from *R. gnavus* ATCC 29149 growth on 0.5% glucose and 1 mM mucin glycans.

Sample	RNA Concentration (ng/ $\mu$ l)	Tapestation RINe
Glc 1	172	8.6
Glc 2	142	8.6
Glc 3	142	8.5
Glc 4	116	8.5

Mucin glycans 1	89	8.8
Mucin glycans 2	144	8.7
Mucin glycans 3	156	7.9
Mucin glycans 4	136	8.7

Extraction volume was 5 ml. RNA was extracted at late exponential phase (T=6 for glucose samples and T=7 for mucin glycan samples. RNA concentration was measured by Qubit after DNase treatment. RINe was determined by TapeStation analysis.

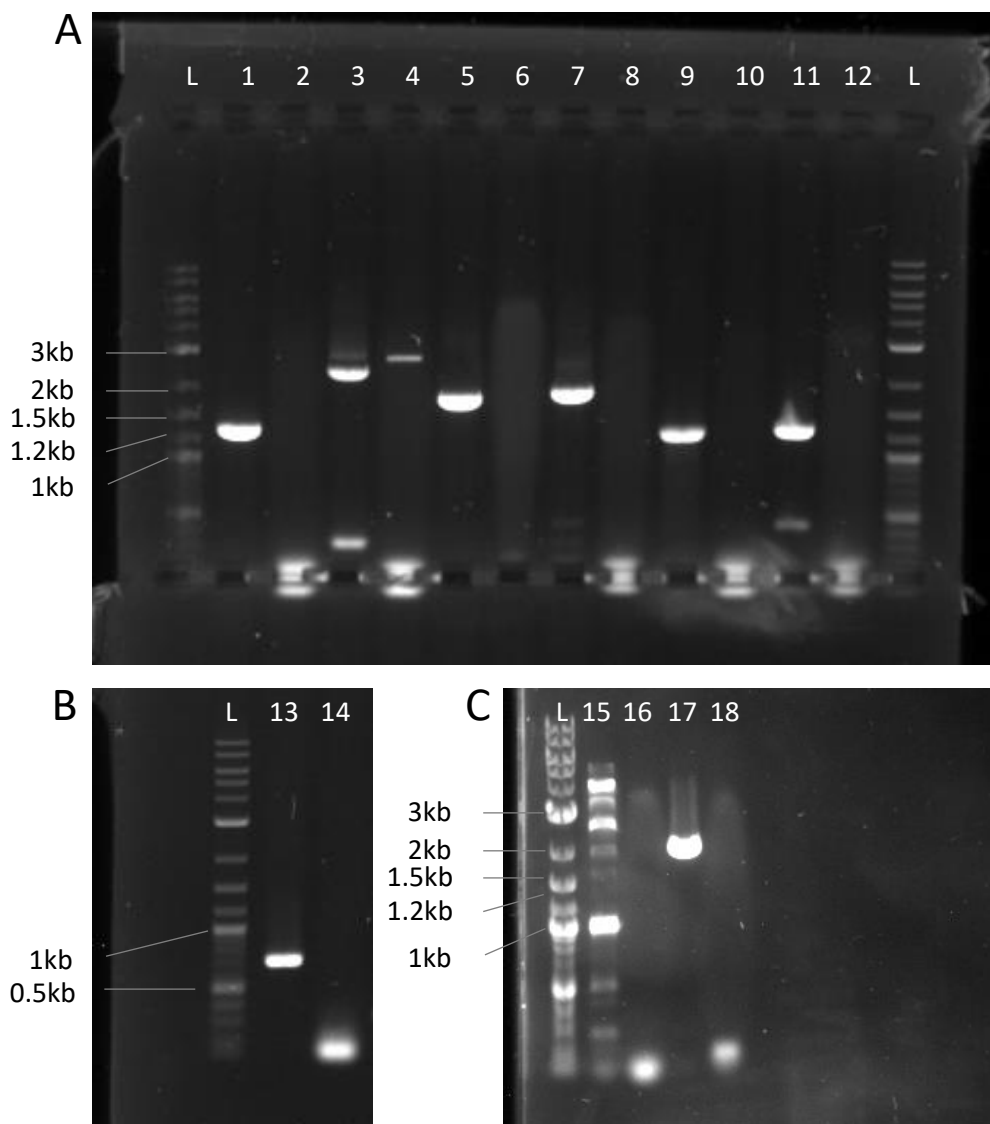
*Supplemental Table S5.* List of selected *R. gnavus* ATCC 29149 GHs.

New locus tag	Old Locus tag	GH family	Protein product
RGna_RS05890	RUMGNA_03411	GH29	WP_233447390.1
RGna_RS04290	RUMGNA_03833	GH29	WP_004844769.1
RGna_RS14395	RUMGNA_00842	GH95	WP_004841212.1
RGna_RS14450	RUMGNA_00830	GH32	WP_004841190.1
RGna_RS07225	RUMGNA_00204	GH23	WP_233447397.1
RGna_RS14905	RUMGNA_01009	GH31	WP_004841524.1
RGna_RS06060	RUMGNA_03447	GH13	WP_004844422.1
RGna_RS08510	RUMGNA_01668	GH13	WP_004842563.1
RGna_RS14690	RUMGNA_00961	GH88	WP_004841431.1
RGna_RS10330	RUMGNA_03118	GH73	WP_004844093.1
RGna_RS07930	RUMGNA_02609	GH2	WP_004843532.1
RGna_RS02745	RUMGNA_00001	GH2	WP_039959057.1

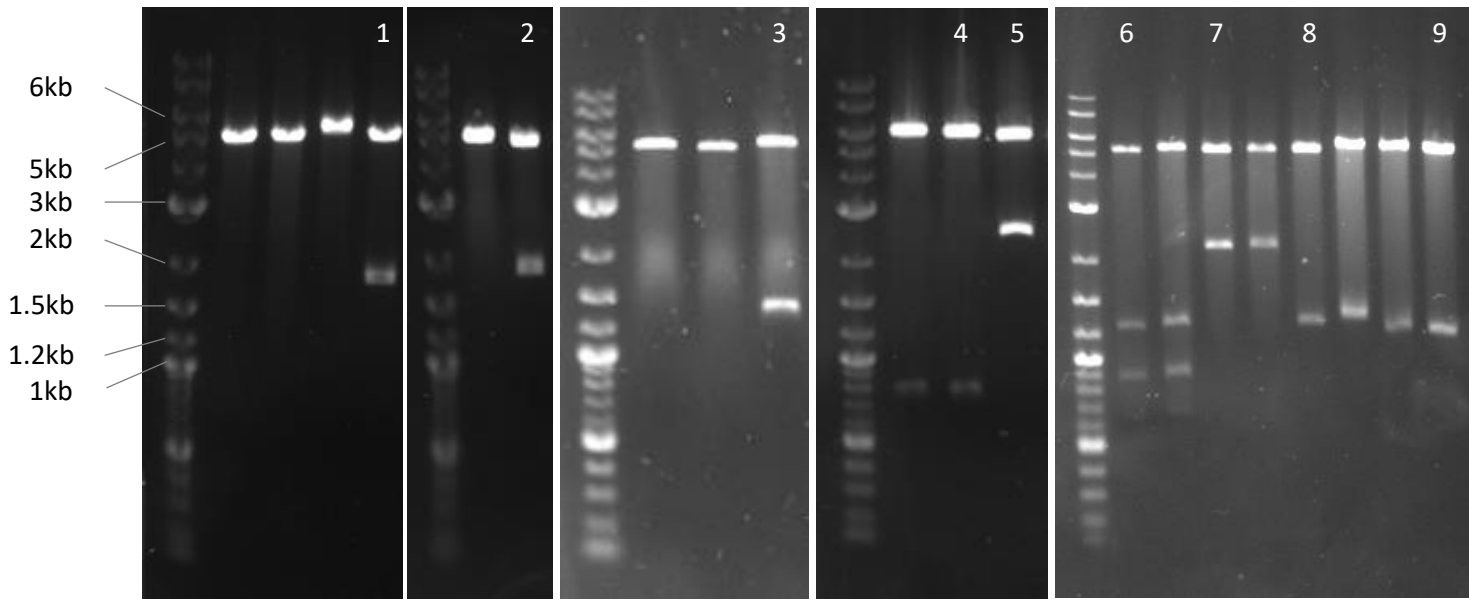
Supplemental Table S6. Overview of all glycans used as substrate for *R. gnavus* growth.

	ATCC 29149	E1	References
<b>Mucin and mucin glycan components</b>			
Glucose	+	+	(Crosth et al., 2013, 2016, 2018; Grabinger et al., 2019); This work
Gal	+	+	(Crosth et al., 2013)
GalNAc	-	-	(Crosth et al., 2013)
GlcNAc	+	+	(Crosth et al., 2013)
Fucose	+	+	(Crosth et al., 2013; Grabinger et al., 2019)
Neu5Ac	-	-	(Crosth et al., 2013, 2016)
2,7-anhydro-Neu5Ac	+	-	(Crosth et al., 2016)
BgA trisaccharide	+	+	(H. Wu, Crosth, et al., 2021); This work
BgA tetrasaccharide	+	-	(H. Wu, Crosth, et al., 2021); This work
pPGM	+	-	(Crosth et al., 2013; H. Wu, Crosth, et al., 2021); This work
BgAtri decapped pPGM	+	+	This work
Mucin glycans	+	-	This work
<b>HMOs</b>			
Lactose	-*	+	(Crosth et al., 2013)
2'FL	+	+	(Crosth et al., 2013)
3FL	+	+	(Crosth et al., 2013; Grabinger et al., 2019)
3'SL	+	-	(Crosth et al., 2013, 2016)
6'SL	-	-	(Crosth et al., 2013, 2016)
LNT	-	-	(Crosth et al., 2013)
LNnT	-	-	(Crosth et al., 2013)
LacNAc	-	+	(Crosth et al., 2013)
<b>Glucose oligo- and polysaccharides</b>			
Soluble starch	-	Nt	(Crosth et al., 2018); This work
Resistant starch	-	Nt	(Crosth et al., 2018)
Pullulan	-	Nt	This work
Glycogen	-	Nt	This work
Maltose	+	Nt	(Crosth et al., 2018); This work
Isomaltose	-	Nt	This work
Maltotriose	+	Nt	(Crosth et al., 2018); This work
Panose	+	Nt	This work
Maltotetraose	+	Nt	(Crosth et al., 2018)
$\alpha$ -Cyclodextrin	-	Nt	This work
$\beta$ -Cyclodextrin	-	Nt	This work
<b>Other dietary glycans</b>			
Melibiose	Nt	+	(Bruehl et al., 2011)
Raffinose	Nt	+	(Bruehl et al., 2011)

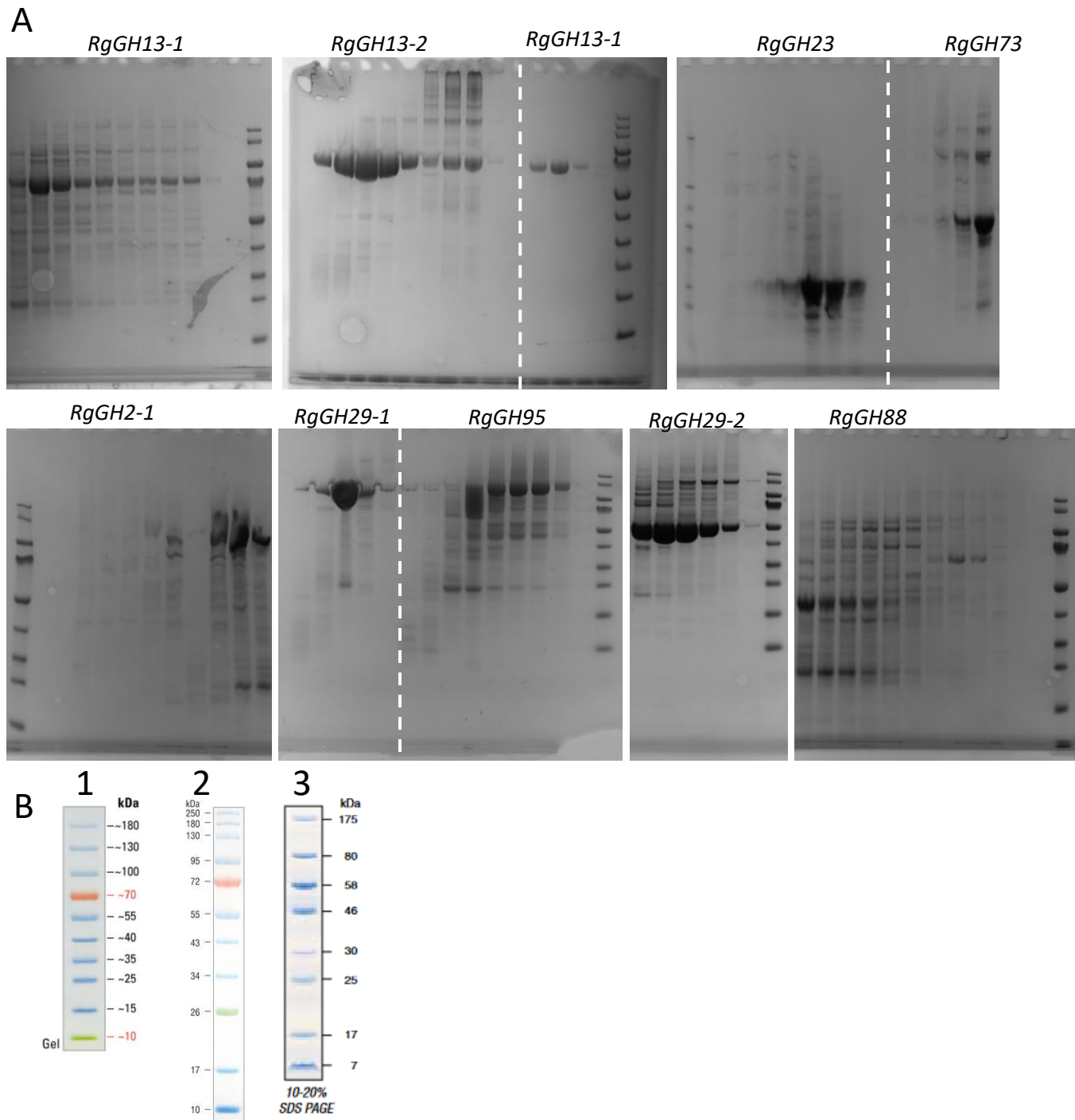
Nt = Not tested. \* *R. gnavus* was shown to grow on lactose after 60 hours in (Grabinger et al., 2019).



*Supplemental Figure S1.* Agarose gel electrophoresis of amplified *R. gnavus* GHs by PCR. GHs amplified from *R. gnavus* ATCC 29149 gDNA (primers listed in 2.4.1). Lanes **1**: RGna\_RS14450 (insert size: 1.30 Kb); **2**: RGna\_RS14450 negative; **3**: RGna\_RS14905 (gene size: 2.50 Kb); **4**: RGna\_RS14905 negative; **5**: RGna\_RS06060 (insert size: 1.85 Kb); **6**: RGna\_RS06060 negative; **7**: RGna\_RS08510 (insert size: 1.95 Kb); **8**: RGna\_RS08510 negative; **9**: RGna\_RS14690 (insert size: 1.22 Kb); **10**: RGna\_RS14690 negative; **11**: RGna\_RS10330 (insert size: 1.26 Kb); **12**: RGna\_RS10330 negative; **13**: RGna\_RS07225 (insert size: 0.66 Kb); **14**: RGna\_RS07225 negative; **15**: RGna\_RS02745 (insert size: 3.11 Kb); **16**: RGna\_RS02745 negative; **17**: RGna\_RS07930 (insert size: 2.24 Kb); **18**: RGna\_RS07930 negative. Negatives are PCR runs without gDNA.

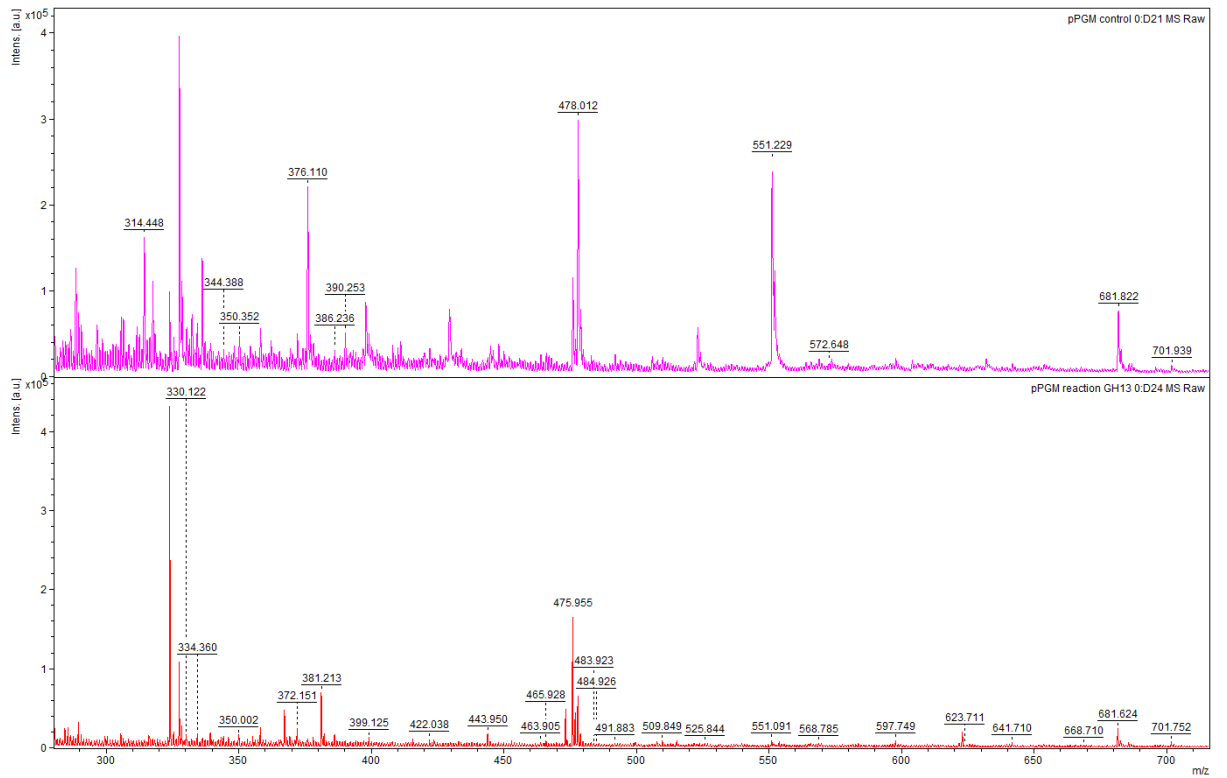


*Supplemental Figure S2.* Agarose gel electrophoresis of recombinant plasmids encoding selected *R. gnavus* GHs. Agarose gel of pET-28a recombinant plasmids digested with NdeI/Nhe-HF and XhoI. Lanes **1**: RGna\_RS06060 (insert size: 1.85 Kb); **2**: RGna\_RS08510 (insert size: 1.95 Kb); **3**: RGna\_RS10330 (gene size: 1.26 Kb); **4**: RGna\_RS07225 (insert size: 0.66 Kb); **5**: RGna\_RS07930 (insert size: 2.24 Kb); **6**: RGna\_RS14395 (insert size: 2.78 Kb (cut 3 times due to NdeI restriction site in gene)); **7**: RGna\_RS05890 (insert size: 2.25 Kb); **8**: RGna\_RS04290 (insert size: 1.31 Kb); **9**: RGna\_RS14690 (insert size: 1.22 Kb). pET-28a plasmid size = 5.369 Kb.

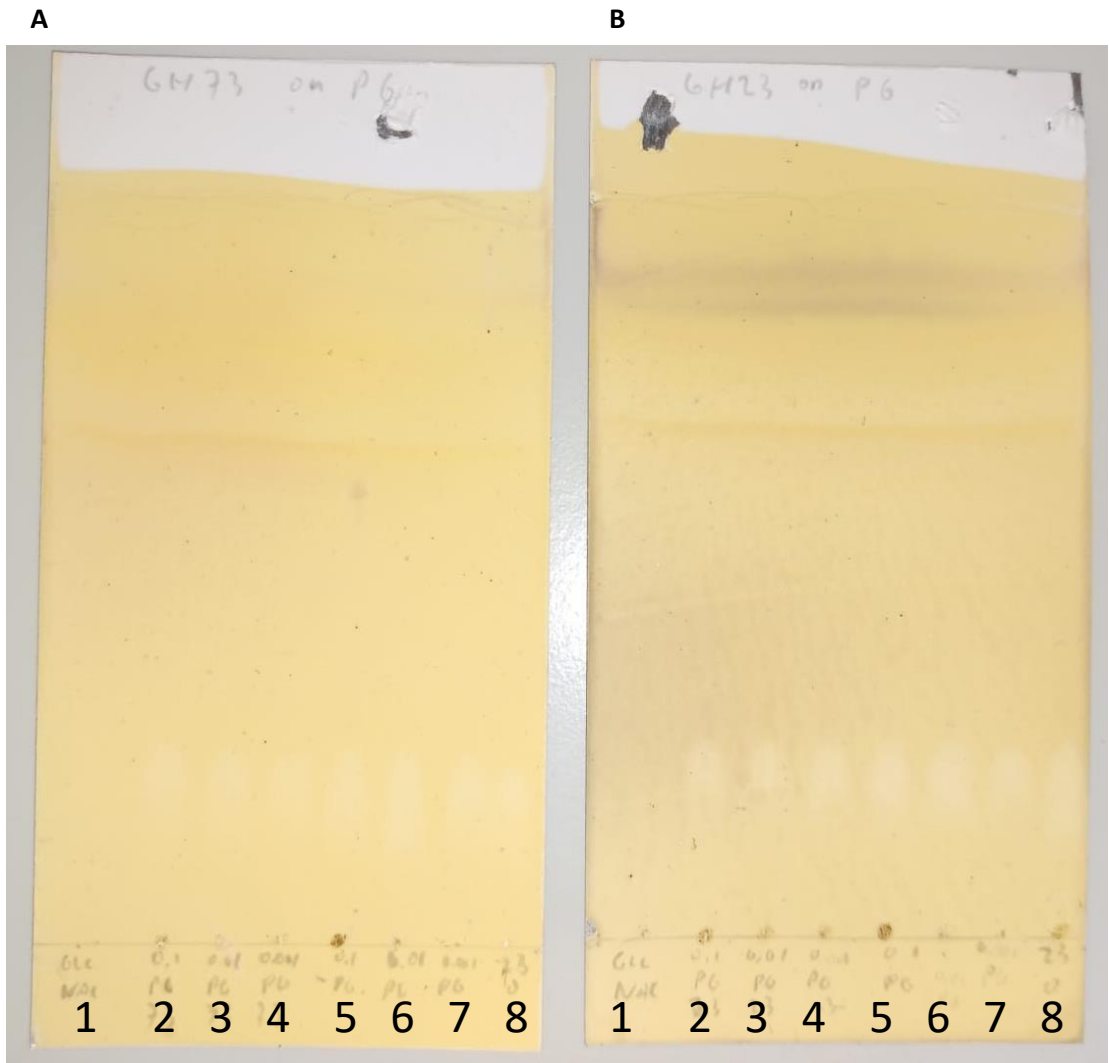


**Supplemental Figure S3.** SDS-PAGE of selected *R. gnavus* recombinant GHs. **(A)** *RgGH13-1*, *RgGH2-1*, *RgGH29-1*, *RgGH95*, *RgGH29-2*, and *RgGH88* were run with ladder 1. The gel containing *RgGH13-2* and more *RgGH13-1* was run with ladder 2. The gel containing *RgGH23* and *RgGH73* was run with ladder 3. Lanes correspond to AKTA fractions chosen based on the analysis of the AKTA peak. Protein names in the gels are given above the gels, for gels with two proteins the white dashed line demarcates the different protein fractions. **(B)** Protein ladders used in SDS-PAGE, 1: PageRuler™ Prestained Protein Ladder, 10 to 180 kDa. 2: Color Prestained Protein Standard, Broad Range (10-250 kDa), New England Biolabs. 3: Prestained Protein Marker Broad Range (7-175 kDa), New England Biolabs.

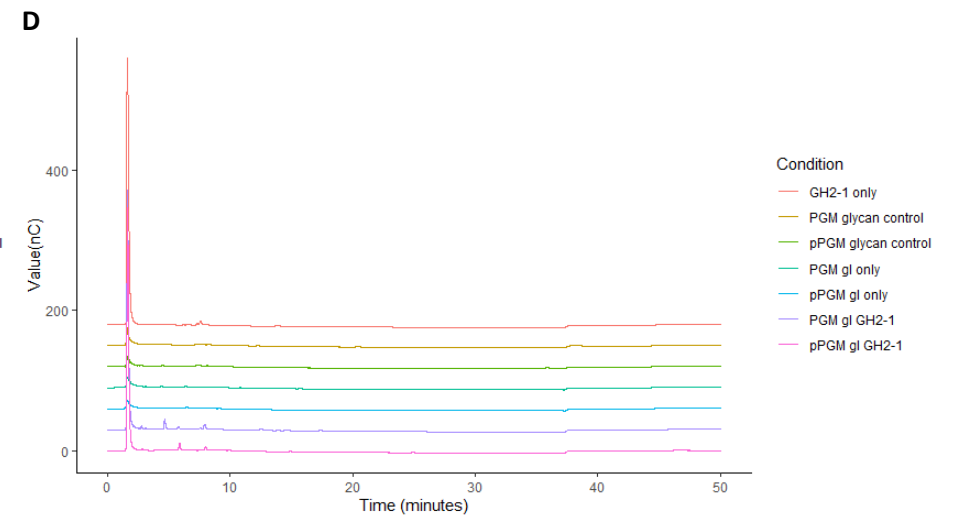
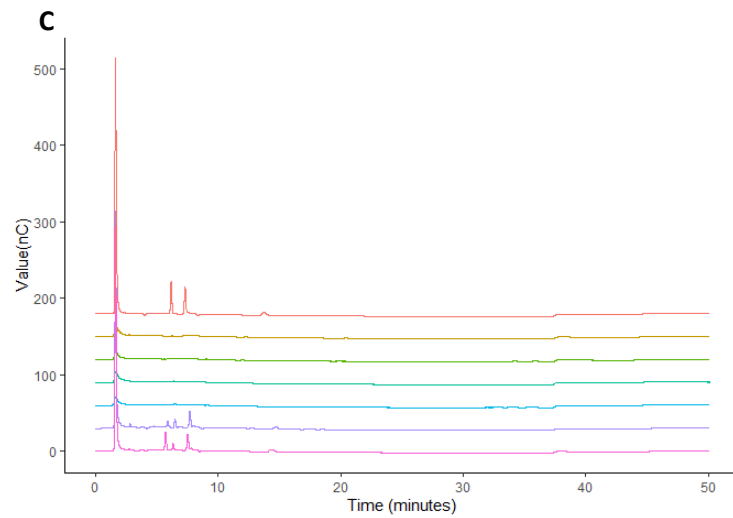
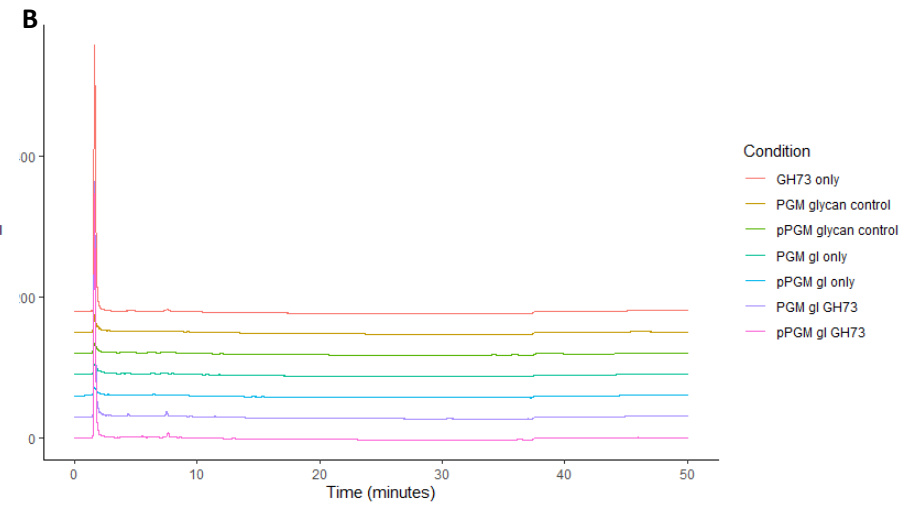
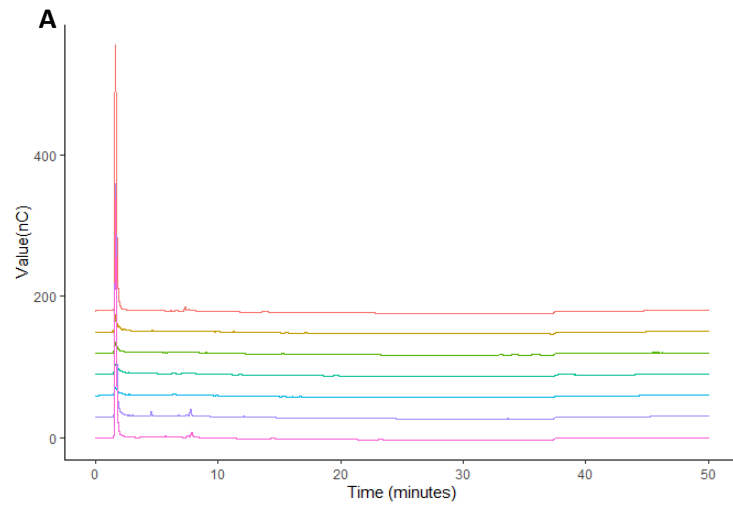




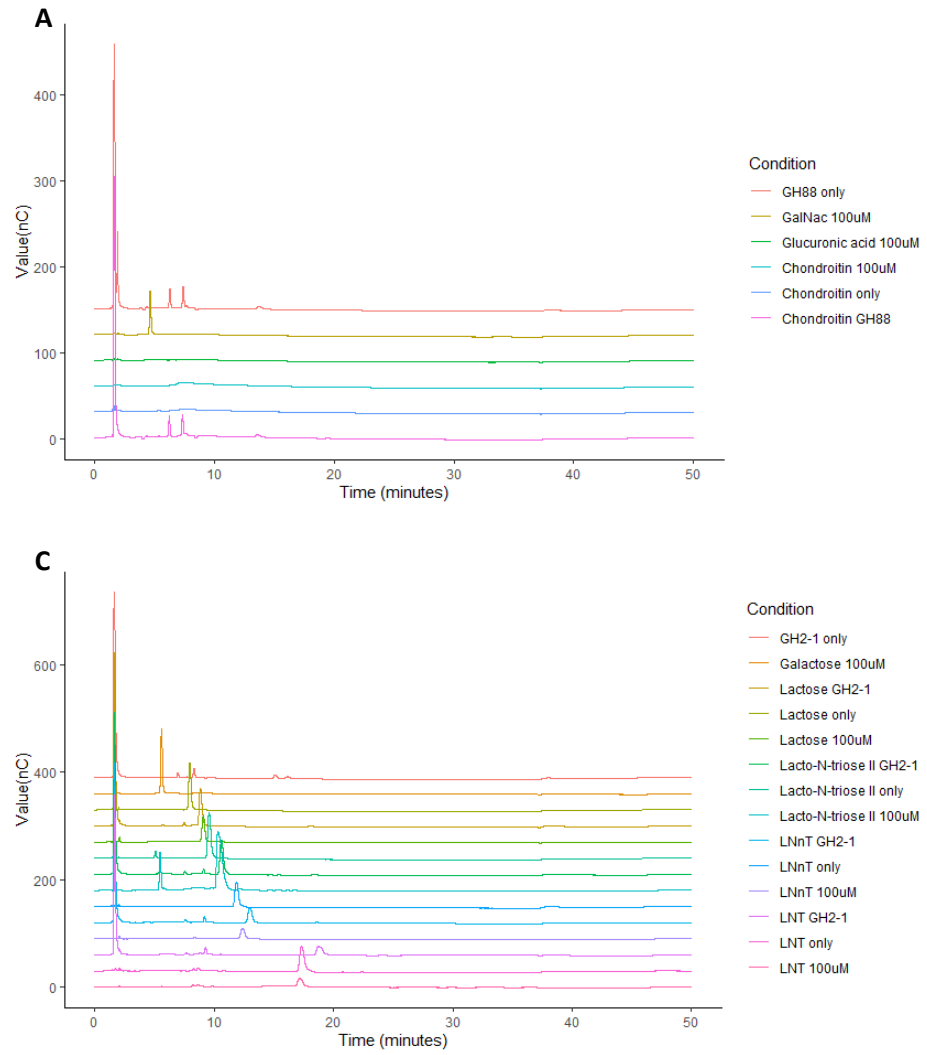
Supplemental Figure S4. MALDI-TOF MS analysis of RgGH13-2 activity on pPGM mucin glycans. Comparison of pPGM mucin glycans (pink) and pPGM mucin glycans treated with RgGH13-2 (red).



Supplemental Figure S5. TLC analysis of the enzymatic reaction of *RgGH73* and *RgGH23* with peptidoglycan. 1 = GlcNAc (standard), 2 = *RgGH73/23* + Peptidoglycan 0.1x, 3 = *RgGH73/23* + Peptidoglycan 0.01x, 4 = *RgGH73/23* + Peptidoglycan 0.001x, 5 = Peptidoglycan 0.1x (assay conditions), 6 = Peptidoglycan 0.01x (assay conditions), 7 = Peptidoglycan 0.001x (assay conditions), 8 = *RgGH73/23* only.



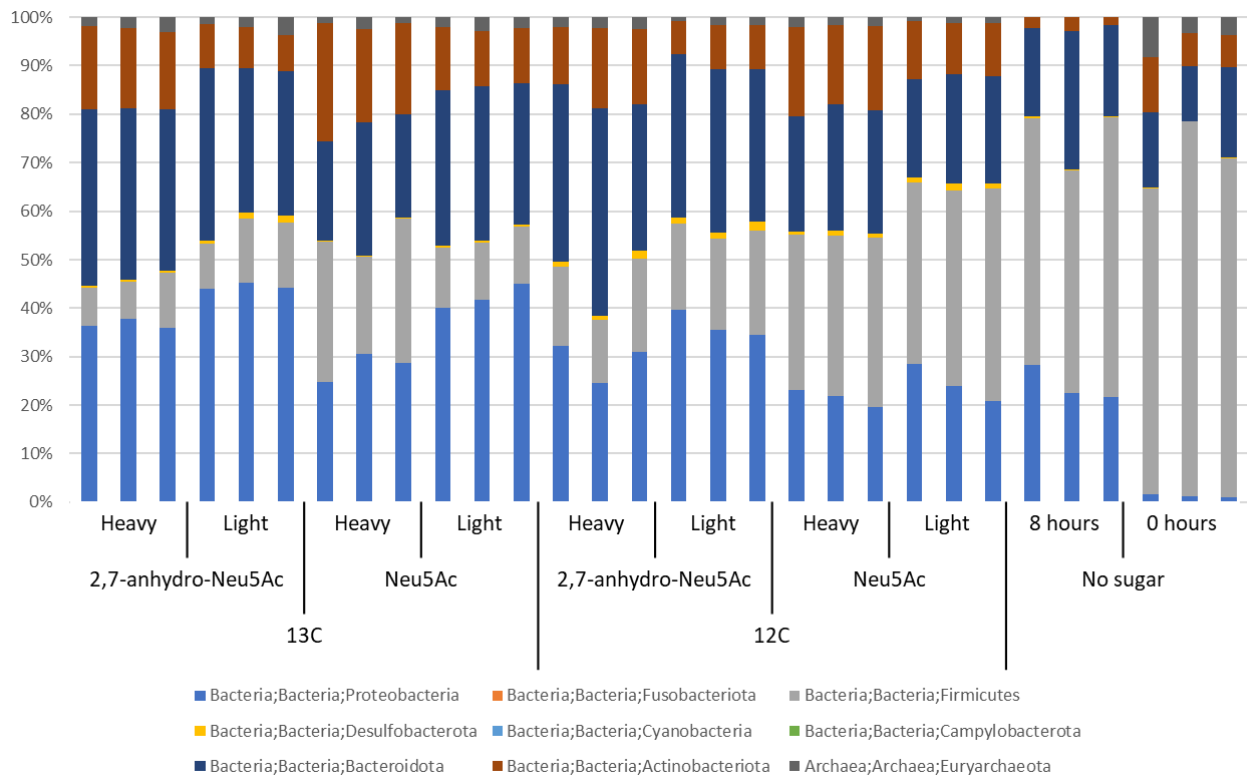
Supplemental Figure S6. HPAEC-PAD analysis of enzymatic reaction of *RgGHs* on mucin glycans. **(A)** Enzymatic reaction of *RgGH23* with mucin glycans. **(B)** Enzymatic reaction of *RgGH73* with mucin glycans. **(C)** Enzymatic reaction of *RgGH88* with mucin glycans. **(D)** Enzymatic reaction of *RgGH2-1* with mucin glycans.



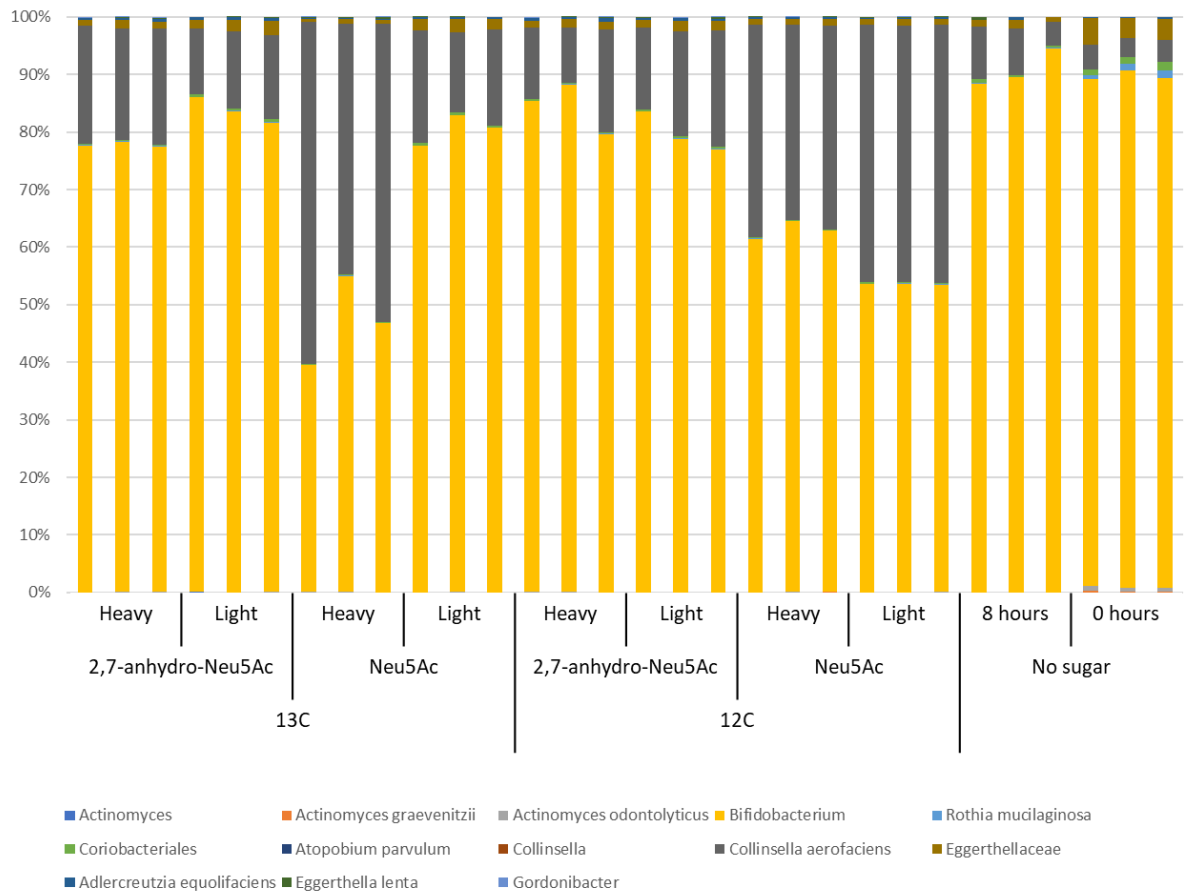
**Supplemental Figure S7.** HPAEC-PAD analysis of enzymatic reaction of *RgGH88* and *RgGH2-1* on predicted substrates. **(A)** Enzymatic reaction of *RgGH88* with chondroitin. **(B)** Enzymatic reaction of *RgGH2-1* with HMOs.

RGNA_RS08510 (GH13-2)	QALHRQGFVIMDVVYNTYSL----DSWLQ----KTPMWFYFVRVWEDG----SVSNG-	299
tr A0A3E3K1S3 A0A3E3K1S3_9FIRM	KALHKNQFRVIMDVVYNTYFL----DSWFQ----RTVPWYFVRVHENG----ETANG-	299
tr Q9P9A0 Q9P9A0_THEAG	EEAHRKRGIRIFDFVFNHSGI GHWAFLDVASRGKSPYWNWYFVQRPFKLGDGKAYLG-	427
tr Q5JID9 Q5JID9_THEKO	DEAHRGRMRVIFDFVFNHCGI GNPAFLDWEKGNESPYWDWPFVKKWPFKLGDSAYVG-	464
RGNA_RS06060 (GH13-1)	KEAHQGRIRIMLDVAVFNHCGYQHPFQDVLHMGKESKYDYFYILDADKPIFDGQVLDGV	304
sp Q08751 NEPU2_THEVU	DEAHRRGIRIILDAVFNHAGDQFFAFRDVLQKGEQSRKDWFFIEDFPVVS-----	276
sp Q9R9H8 BBMA_BACIU	SQLHQGRMRIMLDVAVFNHIGSASPQWQDQVVKNGDQSRKDWPHIHSFPVVT-----	281
sp P38940 NEPU_GEOSE	DRCHEKGRVIMLDVAVFNHCGYEFAPFQDVKWNGESSKYKDWPHIHEFPFLQ-----	279
sp A0A7U9P668 CDAS_GEOTM	KRCHEKGRVIMLDVAVFNHCGYEFQDVLKNGAASRYKDWPHIREFPFLQ-----	279
	. * . * . * . * . * . * . * . * . * . * . * . * . * . * . * . * . * . * .	
RGNA_RS08510 (GH13-2)	-----SACGNVASEQAMCAKYILESVLYWAAEYHMDGFRFLMGLL-----	341
tr A0A3E3K1S3 A0A3E3K1S3_9FIRM	-----SACGNVASERIMCANYIVDSVLYWAGEYHIDGFRFLMGLL-----	341
tr Q9P9A0 Q9P9A0_THEAG	-----W-WLGSLSLPLKNTANPEVKEYLIGAALHWLD-FGFDGRIIDAPQELI	472
tr Q5JID9 Q5JID9_THEKO	-----W-WGFGSLPLKNTANQEVREYLIGAALHWIE-FGFDGRIIDVNEVL	509
RGNA_RS06060 (GH13-1)	PQEIIPREELNRTFAY-TPTMFKWNTGNPEVREYLLAACAQFWTEKXIHIDGWRLLVSNV-	362
sp Q08751 NEPU2_THEVU	-----KTSRTNRYETFAVQVPMFKLRTENPEVKEYLFDVARFVME-QGIDGWRLLVANEV-	330
sp Q9R9H8 BBMA_BACIU	-----D---DNVDRFAF-TADMPKLTANPEVQRYLLDIALYWIREFDIDGWRLLVANEV-	332
sp P38940 NEPU_GEOSE	-----TEPRPNYDTFRF-VFQMPKLTANPEVKRYLLDVATYWIREFDIDGWRLLVANEI-	333
sp A0A7U9P668 CDAS_GEOTM	-----TEPRPNYDTFAP-VFQMPKLTANPEVKRYLLDVATYWIREFDIDGWRLLVANEI-	333
	. : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : .	
RGNA_RS08510 (GH13-2)	-DVELMNRIRKGLDARYGTGEKLVFGEIPWRADETAVEGNALLADKAHILLDEQVGMFSD	400
tr A0A3E3K1S3 A0A3E3K1S3_9FIRM	-DTGLMNRIMELSSRYGKGEKLVFGEIPWASSETPMEQALPALKKNIIRLNPQVGFCD	400
tr Q9P9A0 Q9P9A0_THEAG	NAEEFFSFLRKAIEK--HPDAYIVGIEWELSPRWVQGNMFDSIMN-YALGRDILLAYAR	529
tr Q5JID9 Q5JID9_THEKO	DPGTFFPELRRKAVKEK--KPDAYLVGIEIWTLSPEWVKGRDSDSIMN-YALGRDILLNAYK	566
RGNA_RS06060 (GH13-1)	-PHDFWREFRKRVKDR--NPEYLILGENWNSLPLWLMGDPDAVMN-YEFAMPWKYFRK	418
sp Q08751 NEPU2_THEVU	-DHAFWREFRRLVKSL--NPDALIVGIEIWHDSAGWLMGDPDSDVMN-YLFRESVIRFFAT	386
sp Q9R9H8 BBMA_BACIU	-DHVFWKTFRQAVSTE--KPDVYILGIEIWHSAEPWLRGDEFFHAAMN-YPFTEPMIEYFAD	388
sp P38940 NEPU_GEOSE	-DHEFWREFRQEVKAL--KPDVYILGIEIWHDSAGWLMGDPDAVMN-YPTDGVLRFFAK	389
sp A0A7U9P668 CDAS_GEOTM	-DHQFWREFRQAVKAL--KPDVYILGIEIWHDSAGWLMGDPDAVMN-YPLADAALRFFAK	389
	. : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : .	
RGNA_RS08510 (GH13-2)	DTRDAIKGSVFEEAEPGFANGKDLLEEQILASVKAWCGQKPEPKVKAPSQVITYVSAHDNH	460
tr A0A3E3K1S3 A0A3E3K1S3_9FIRM	DTRDAIKGHTFEGGTPGFVNGGKQEGAILSSVKAWCGPG-SKVQAPSQVITYVSAHDNW	459
tr Q9P9A0 Q9P9A0_THEAG	GDW-----NGERTL-----ELLGRIYASY-----GENVIAMGFNLVSSHDT	566
tr Q5JID9 Q5JID9_THEKO	GLL-----SGESAM-----KMMGRYASY-----GENVVAMGFNLVSDHDT	603
RGNA_RS06060 (GH13-1)	EKE-----GERIYSEEQFR-----YAI GRLLTSY-----PKNVTANFNLESHTD	460
sp Q08751 NEPU2_THEVU	GEI-----HAERFD-----AELTRARMLY-----PEQAAQLWNLDSHDT	423
sp Q9R9H8 BBMA_BACIU	QTI-----SASRMA-----HRVNAHLMNG-----MKQANVMFNLLDSHDTK	425
sp P38940 NEPU_GEOSE	E EI-----SARQFA-----NQMMHVLHSY-----PNNVNEAAFNLGSHDTS	426
sp A0A7U9P668 CDAS_GEOTM	EDM-----SASEFA-----DRLMHVLSY-----PKQVNEAAFNLGSHDTP	426
	. : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : .	
RGNA_RS08510 (GH13-2)	TLWDKLCET---TASEELRRKQYRLAAGMYLTCQGI PFLLSGEEFARTKGRDNTYNAPI	517
tr A0A3E3K1S3 A0A3E3K1S3_9FIRM	TLWDKLSIT---VEDEAERMRANRMAAAVYMT CQGRIFFLSGEEASRTKGLNSYNAPI	516
tr Q9P9A0 Q9P9A0_THEAG	RVLTDLGGGNLGDTPKPEAI QRLKLLSTLLYTLPGMPVTFQDGERGLLDGKE--HFD--S	622
tr Q5JID9 Q5JID9_THEKO	RVLTDLGGGKLGDTSPNESIQRLKLLSTLLYALPGTPTVTFQDGERGLLDGDKG--HYD--E	659
RGNA_RS06060 (GH13-1)	RILNRAGQD-----VSIVKLAFLYLFMFI FPGTPCVYVYGGIEGMGGGHE-----	503
sp Q08751 NEPU2_THEVU	RFLTSCGN-----EAKFRILAVLFQMTYLGTPPLIYYGDEIGMAGATD--P----D	467
sp Q9R9H8 BBMA_BACIU	RLLTRCRND-----EKKARALLAEMFAQTGSPCIYYGTEIGLDGEND--P----L	469
sp P38940 NEPU_GEOSE	RILTVCGGD-----IRKVKLLFLFQLTFTGSPCIYYGDEIGMTGGND--P----E	470
sp A0A7U9P668 CDAS_GEOTM	RLLTVCGGD-----VRKVKLLFLFQLTFTGSPCIYYGDEIGMTGGND--P----E	470
	. : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : .	
RGNA_RS08510 (GH13-2)	TINRLDWEQAYKEQELVEYYRGLITLRKMLPGLWDKSAQAARILKTKWTDGCVGLYVDN	577
tr A0A3E3K1S3 A0A3E3K1S3_9FIRM	EINRLDWNRMYEQYELVEYYRGLIALRRLPGLCDKSKDAWKRIRDEWRTDGVVGFVLDN	576
tr Q9P9A0 Q9P9A0_THEAG	HRYPICQWDT---VNEDVLNHYKSLADLRKSVPAL-TSSKIKFY-----TSKE--GVLAF	671
tr Q5JID9 Q5JID9_THEKO	CRYPICQWDT---VNEDVLNHYRALAELRKRVPAL-RSSAMRFY-----TAKG--GYMAFF	708
RGNA_RS06060 (GH13-1)	NRQCMWEEKEKQRELFEFISFMIELRKYDSF-CSTELQWI-----NSE--NSLALK	553
sp Q08751 NEPU2_THEVU	CRRPMIWEEKEKQRELFEFIFYKELIRLRHLASL-TRGNVRSW-----HADKQANLYAFV	520
sp Q9R9H8 BBMA_BACIU	CRKCMVWEEKEKQNDMLQFMKRLIALRQENTLLTEGHLEWN-----LLDDKNDPISFS	523
sp P38940 NEPU_GEOSE	CRKCMVWDPMQNKELHQHVQLIALRQYRSL-RRGEISFL-----HADDEMNYLIYK	523
sp A0A7U9P668 CDAS_GEOTM	CRKCMVWDPKQNKELYEHVKQLIALRQYRAL-RRGDVAFL-----AADDEVNHLVYA	523
	. : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : .	

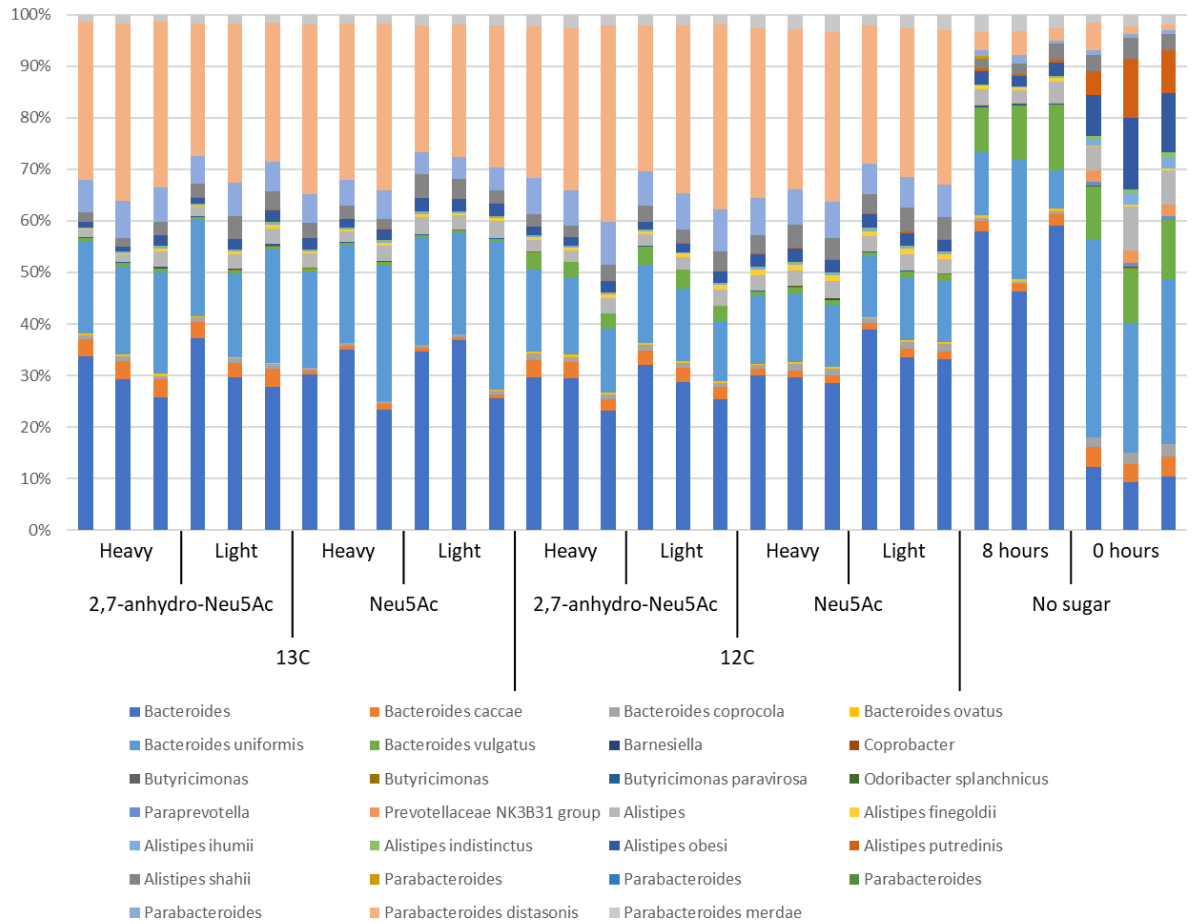
Supplemental Figure S8. Amino acids sequence alignment of RgGH13-1 and RgGH13-2 with fully annotated pullulanase sequences. tr|A0A3E3K1S3|A0A3E3K1S3\_9FIRM is a type I pullulanase from *Faecalicatena orotica*. tr|Q9P9A0|Q9P9A0\_THEAG is a type III pullulanase from *Thermococcus aggregans*. tr|Q5JID9|Q5JID9\_THEKO is a type III pullulanase from *Thermococcus kodakarensis*. sp|Q08751|NEPU2\_THEVU is a neopullulanase from *Thermoactinomyces vulgaris*. sp|Q9R9H8|BBMA\_BACIU is an intracellular maltogenic amylase from *Bacillus subtilis*. sp|P38940|NEPU\_GEOSE is a neopullulanase from *Geobacillus stearothermophilus*. sp|A0A7U9P668|CDAS\_GEOTM is a cyclomaltodextrinase from *Geobacillus thermopakistaniensis*. Highlighted in green are catalytic residues. Highlighted in blue are binding residues. \* correspond to fully conserved residues, : to residues with highly similar properties. . to residues with weakly similar properties.



*Supplemental Figure S9.* Relative abundance of bacterial phyla in heavy and light SIP fractions. Heavy fractions contain RNA from all gradient fractions with a density of more than 1.81 g/ml, light fractions contain RNA from all gradient fractions with a density of less than 1.81 g/ml. 2,7-anhydro-Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac. Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled Neu5Ac.

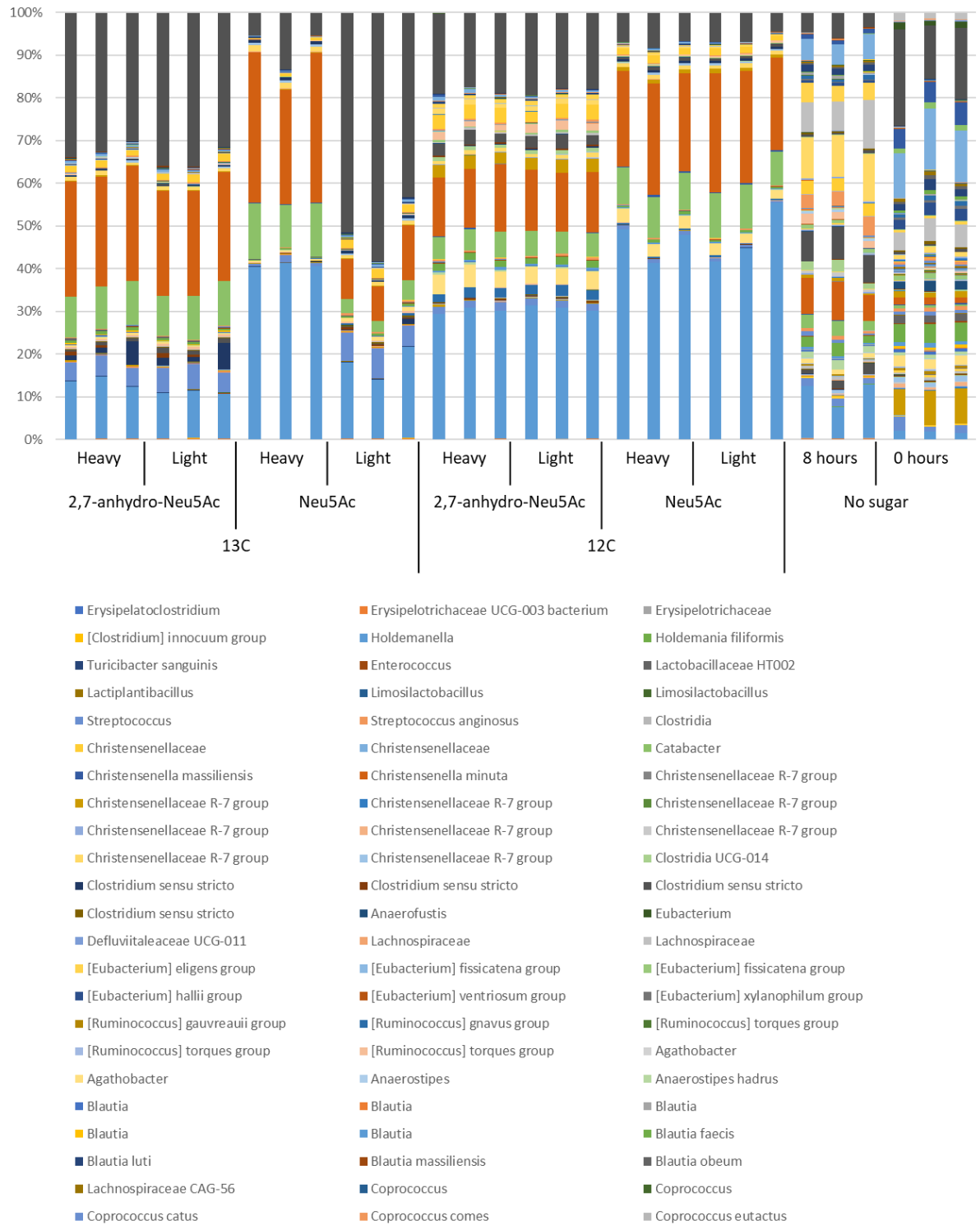


*Supplemental Figure S10.* Relative abundance of Actinobacteriota species in heavy and light SIP fractions per major bacterial phyla. Heavy fractions contain RNA from all gradient fractions with a density of more than 1.81 g/ml, light fractions contain RNA from all gradient fractions with a density of less than 1.81 g/ml. 2,7-anhydro-Neu5Ac fractions contain RNA from bacteria grown on <sup>13</sup>C<sub>7</sub> labelled 2,7-anhydro-Neu5Ac. Neu5Ac fractions contain RNA from bacteria grown on <sup>13</sup>C<sub>7</sub> labelled Neu5Ac.

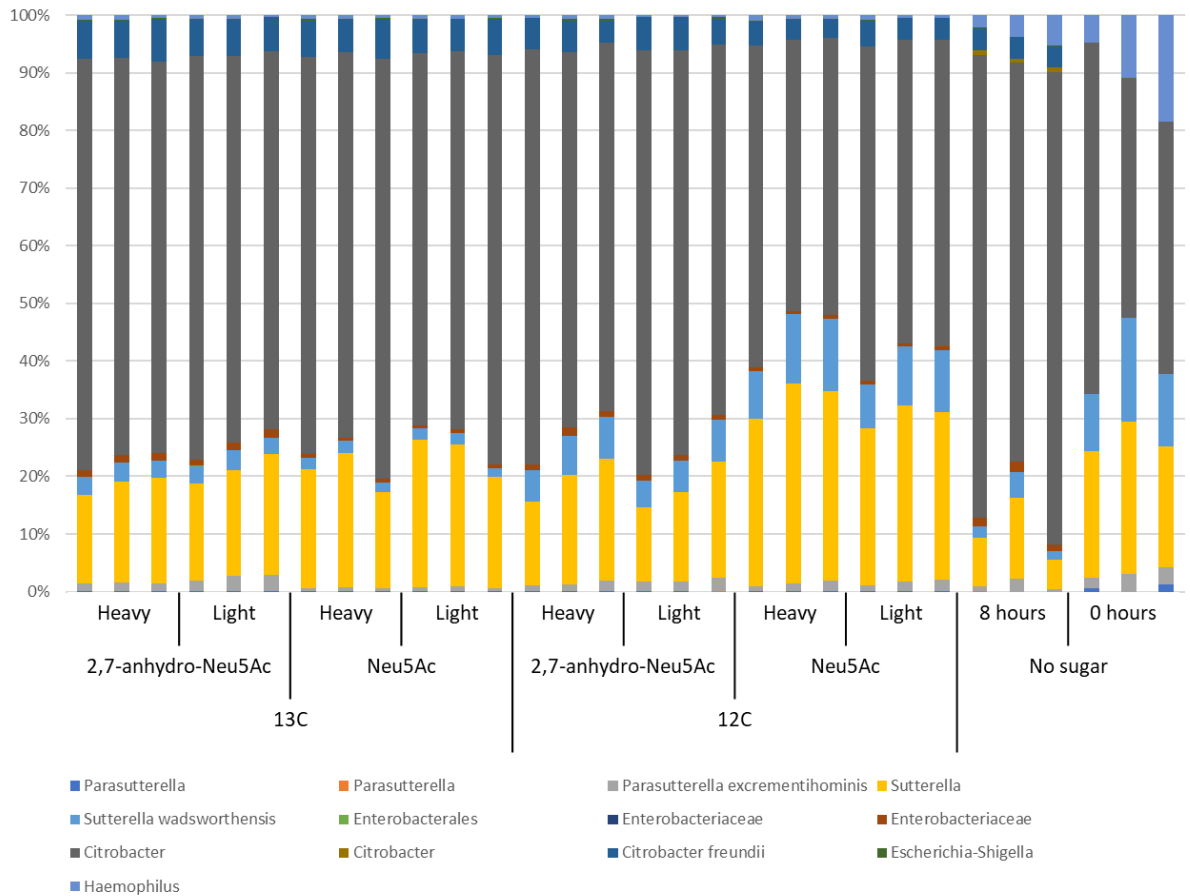


**Supplemental Figure S11.** Relative abundance of Bacteroidota species in heavy and light SIP fractions per major bacterial phyla. Heavy fractions contain RNA from all gradient fractions with a density of more than 1.81 g/ml, light fractions contain RNA from all gradient fractions with a density of less than 1.81 g/ml. 2,7-anhydro-Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac. Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled Neu5Ac.





**Supplemental Figure S12.** Relative abundance of Firmicutes species in heavy and light SIP fractions per major bacterial phyla. Heavy fractions contain RNA from all gradient fractions with a density of more than 1.81 g/ml, light fractions contain RNA from all gradient fractions with a density of less than 1.81 g/ml. 2,7-anhydro-Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac. Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled Neu5Ac.



*Supplemental Figure S13.* Relative abundance of Proteobacteria species in heavy and light SIP fractions per major bacterial phyla. Heavy fractions contain RNA from all gradient fractions with a density of more than 1.81 g/ml, light fractions contain RNA from all gradient fractions with a density of less than 1.81 g/ml. 2,7-anhydro-Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled 2,7-anhydro-Neu5Ac. Neu5Ac fractions contain RNA from bacteria grown on  $^{13}\text{C}_7$  labelled Neu5Ac.

## Appendix 2. LAB stock solutions composition

### Amino acid 10x

Ingredients:	g/L	MnSO <sub>4</sub> ·H <sub>2</sub> O	0.11
I-Histidine	1.7	(NH <sub>4</sub> ) <sub>6</sub> Mo <sub>7</sub> O <sub>24</sub> ·4H <sub>2</sub> O	0.19
I-Isoleucine	2.4	FeCl <sub>3</sub>	3
I-Leucine	10	FeSO <sub>4</sub> ·7H <sub>2</sub> O	4
I-Methionine	1.25	EDTA	7.34
I-Valine	7	Nitilotriacetic acid	7.34
I-Arginine	7.2		
I-Cysteine	2		
I-Glutamic acid	6		
I-Phenylalanine	4		
I-Proline	7		
I-Asparagine	5		
I-Aspartic acid	0.5		
I-Glutamine	6		
I-Serine	5		
I-Threonine	5		
I-Alanine	4		
Glycine	3		
I-Lysine	5		
I-Tryptophan	2		
I-Tyrosine	3		

### Salt & minerals 2 10x

Ingredients:	g/L
KCH <sub>3</sub> CO <sub>2</sub> (Potassium acetate)	9
Ammonium citrate dibasic	17
MgCl <sub>2</sub>	3.86
NaCl	30
CaCl <sub>2</sub> (anhydrous)	0.302
K <sub>2</sub> SO <sub>4</sub>	0.23

### Vitamins & Antioxidants 1000x

Ingredients:	g/L
myo-Inositol	2
L-Glutathione reduced	15
Biotin	6
Thiamine HCl	0.56
Riboflavin	0.9
Pyridoxamine·2HCl	5
Niacin	0.9
Pyridoxine HCl	4.8
Calcium Pantothenate	1.2
Folic acid	0.56
p-Aminobenzoic acid	0.056
Lipoic acid	1

### Nucleotides 100x

Ingredients:	g/L
Adenine	1.1
Guanine	0.56
Uracil	2.3
Xanthine	0.38

### Salt & minerals 1 1000x

Ingredients:	g/L
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	5
CoCl <sub>2</sub> ·6H <sub>2</sub> O	0.19
CuSO <sub>4</sub> (anhydrous)	0.12
H <sub>3</sub> BO <sub>3</sub>	0.75
KI	0.11

### Ascorbic acid 100x

Ingredients:	g/L
Ascorbic acid	50

## References

- Adamberg, K., & Adamberg, S. (2018). Selection of fast and slow growing bacteria from fecal microbiota using continuous culture with changing dilution rate. *Microbial Ecology in Health and Disease*, *29*(1), 1549922. <https://doi.org/10.1080/16512235.2018.1549922>
- Aguilera, M., Rakotoarivonina, H., Brutus, A., Giardina, T., Simon, G., & Fons, M. (2012). Aga1, the first alpha-Galactosidase from the human bacteria *Ruminococcus gnavus* E1, efficiently transcribed in gut conditions. *Research in Microbiology*, *163*(1), 14–21. <https://doi.org/10.1016/j.resmic.2011.10.005>
- Ahmad, N., Rashid, N., Haider, M. S., Akram, M., & Akhtar, M. (2014). Novel maltotriose-hydrolyzing thermoacidophilic type III pullulan hydrolase from *Thermococcus kodakarensis*. *Applied and Environmental Microbiology*, *80*(3), 1108–1115. <https://doi.org/10.1128/AEM.03139-13>
- Ali, G., Dulong, V., Gasmi, S. N., Rihouey, C., Picton, L., & Le Cerf, D. (2015). Covalent immobilization of pullulanase on alginate and study of its hydrolysis of pullulan. *Biotechnology Progress*, *31*(4), 883–889. <https://doi.org/10.1002/btpr.2093>
- Ali, G., Rihouey, C., Le Cerf, D., & Picton, L. (2013). Effect of carboxymethyl groups on degradation of modified pullulan by pullulanase from *Klebsiella pneumoniae*. *Carbohydrate Polymers*, *93*(1), 109–115. <https://doi.org/10.1016/j.carbpol.2012.07.039>
- Almagro-Moreno, S., & Boyd, E. F. (2009). Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evolutionary Biology*, *9*(1), 1–16. <https://doi.org/10.1186/1471-2148-9-118>
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Ambort, D., Johansson, M. E. V., Gustafsson, J. K., Nilsson, H. E., Ermund, A., Johansson, B. R., Koeck, P. J. B., Hebert, H., & Hansson, G. C. (2012). Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(15), 5645–5650. <https://doi.org/10.1073/pnas.1120269109>
- An, G., Wei, B., Xia, B., McDaniel, J. M., Ju, T., Cummings, R. D., Braun, J., & Xia, L. (2007). Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *Journal of Experimental Medicine*, *204*(6), 1417–1429. <https://doi.org/10.1084/jem.20061929>
- Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10). <https://doi.org/10.1186/gb-2010-11-10-r106>
- Anderson, K. M., Ashida, H., Maskos, K., Dell, A., Li, S. C., & Li, Y. T. (2005). A clostridial endo- $\beta$ -galactosidase that cleaves both blood group A and B glycotopes: The first member of a new glycoside hydrolase family, GH98. *Journal of Biological Chemistry*, *280*(9), 7720–7728. <https://doi.org/10.1074/jbc.M414099200>
- Andersson, A. F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P., & Engstrand, L. (2008).

- Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE*, 3(7). <https://doi.org/10.1371/journal.pone.0002836>
- Andoh, A., & Nishida, A. (2022). Alteration of the Gut Microbiome in Inflammatory Bowel Disease. *Digestion*, 1–8. <https://doi.org/10.1159/000525925>
- Anggraeni, A. A. (2022). Mini-Review: The potential of raffinose as a prebiotic. *IOP Conference Series: Earth and Environmental Science*, 980(1). <https://doi.org/10.1088/1755-1315/980/1/012033>
- Ansorge, R., Birolo, G., James, S. A., & Telatin, A. (2021). Dadaist2: A toolkit to automate and simplify statistical analysis and plotting of metabarcoding experiments. *International Journal of Molecular Sciences*, 22(10). <https://doi.org/10.3390/ijms22105309>
- Aron-Wisnewsky, J., Prifti, E., Belda, E., Ichou, F., Kayser, B. D., Dao, M. C., Verger, E. O., Hedjazi, L., Bouilliot, J. L., Chevallier, J. M., Pons, N., Le Chatelier, E., Levenez, F., Ehrlich, S. D., Dore, J., Zucker, J. D., & Clément, K. (2019). Major microbiota dysbiosis in severe obesity: Fate after bariatric surgery. *Gut*, 68(1), 70–82. <https://doi.org/10.1136/gutjnl-2018-316103>
- Aroob, I., Ahmad, N., & Rashid, N. (2021). Cyclodextrin-preferring glycoside hydrolases: properties and applications. *Amylase*, 5(1), 23–37. <https://doi.org/10.1515/amylase-2021-0003>
- Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., DeRoos, P., Liu, H., Cross, J. R., Pfeffer, K., Coffey, P. J., & Rudensky, A. Y. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature*, 504(7480), 451–455. <https://doi.org/10.1038/nature12726>
- Arthur, J. C., Perez-Chanona, E., Mühlbauer, M., Tomkovich, S., Uronis, J. M., Fan, T., Campbell, B. J., Abujamel, T., Dogan, B., Rogers, A. B., Rhodes, J. M., Stintzi, A., Simpson, K. W., Hansen, J. J., Keku, T. O., Fodor, A. A., & Jobin, C. (2012). Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science*, 338(6103), 120–123. <https://doi.org/10.1126/science.1224820>
- Ashida, H., Maki, R., Ozawa, H., Tani, Y., Kiyohara, M., Fujita, M., Imamura, A., Ishida, H., Kiso, M., & Yamamoto, K. (2008). Characterization of two different endo- $\alpha$ -N-acetylgalactosaminidases from probiotic and pathogenic enterobacteria, *Bifidobacterium longum* and *Clostridium perfringens*. In *Glycobiology* (Vol. 18, Issue 9, pp. 727–734). <https://doi.org/10.1093/glycob/cwn053>
- Atuma, C., Strugala, V., Allen, A., & Holm, L. (2001). The adherent gastrointestinal mucus gel layer: Thickness and physical state in vivo. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 280(5 43-5), 922–929. <https://doi.org/10.1152/ajpgi.2001.280.5.g922>
- Azad, M. B., Robertson, B., Atakora, F., Becker, A. B., Subbarao, P., Moraes, T. J., Mandhane, P. J., Turvey, S. E., Lefebvre, D. L., Sears, M. R., & Bode, L. (2018). Human Milk Oligosaccharide Concentrations Are Associated with Multiple Fixed and Modifiable Maternal Characteristics, Environmental Factors, and Feeding Practices. *Journal of Nutrition*, 148(11), 1733–1742. <https://doi.org/10.1093/jn/nxy175>
- Bag, S., Ghosh, T. S., & Das, B. (2017). Complete genome sequence of *Collinsella aerofaciens* isolated from the gut of a healthy Indian subject. *Genome Announcements*, 5(47). <https://doi.org/10.1128/genomeA.01361-17>
- Barone, M., D'Amico, F., Brigidi, P., & Turroni, S. (2022). Gut microbiome–micronutrient interaction: The key to controlling the bioavailability of minerals and vitamins? *BioFactors*, 48(2), 307–314. <https://doi.org/10.1002/biof.1835>

- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Beatson, R., Maurstad, G., Picco, G., Arulappu, A., Coleman, J., Wandell, H. H., Clausen, H., Mandel, U., Taylor-Papadimitriou, J., Sletmoen, M., & Burchell, J. M. (2015). The breast cancer-associated glycoforms of MUC1, MUC1-Tn and sialyl-Tn, are expressed in COSMC wild-type cells and bind the C-type lectin MGL. *PLoS ONE*, 10(5), 1–21. <https://doi.org/10.1371/journal.pone.0125994>
- Bell, A., Brunt, J., Crost, E., Vaux, L., Nepravishta, R., Owen, C. D., Latousakis, D., Xiao, A., Li, W., Chen, X., Walsh, M. A., Claesen, J., Angulo, J., Thomas, G. H., & Juge, N. (2019). Elucidation of a sialic acid metabolism pathway in mucus-foraging *Ruminococcus gnavus* unravels mechanisms of bacterial adaptation to the gut. *Nature Microbiology*, 4(December). <https://doi.org/10.1038/s41564-019-0590-7>
- Bell, A., & Juge, N. (2021). Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology*, 31(6), 691–696. <https://doi.org/10.1093/glycob/cwaa097>
- Bell, A., Severi, E., Lee, M., Monaco, S., Latousakis, D., Angulo, J., Thomas, G. H., Naismith, J. H., & Juge, N. (2020). Uncovering a novel molecular mechanism for scavenging sialic acids in bacteria. *Journal of Biological Chemistry*, 295(40), 13724–13736. <https://doi.org/10.1074/jbc.RA120.014454>
- Bell, A., Severi, E., Owen, C. D., Latousakis, D., & Juge, N. (2023). Biochemical and structural basis of sialic acid utilization by gut microbes. *Journal of Biological Chemistry*, 299(3), 102989. <https://doi.org/10.1016/j.jbc.2023.102989>
- Bergström, J. H., Birchenough, G. M. H., Katona, G., Schroeder, B. O., Schütte, A., Ermund, A., Johansson, M. E. V., & Hansson, G. C. (2016). Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. *Proceedings of the National Academy of Sciences of the United States of America*, 113(48), 13833–13838. <https://doi.org/10.1073/pnas.1611400113>
- Bergstrom, K., Liu, X., Zhao, Y., Gao, N., Wu, Q., Song, K., Cui, Y., Li, Y., McDaniel, J. M., McGee, S., Chen, W., Huycke, M. M., Houchen, C. W., Zenewicz, L. A., West, C. M., Chen, H., Braun, J., Fu, J., & Xia, L. (2016). Defective Intestinal Mucin-Type O-Glycosylation Causes Spontaneous Colitis-Associated Cancer in Mice. *Gastroenterology*, 151(1), 152-164.e11. <https://doi.org/10.1053/j.gastro.2016.03.039>
- Bergstrom, K., Shan, X., Casero, D., Batushansky, A., Lagishetty, V., Jacobs, J. P., Hoover, C., Kondo, Y., Shao, B., Gao, L., Zandberg, W., Noyovitz, B., Michael McDaniel, J., Gibson, D. L., Pakpour, S., Kazemian, N., McGee, S., Houchen, C. W., Rao, C. V., ... Xia, L. (2020). Proximal colon-derived O-glycosylated mucus encapsulates and modulates the microbiota. *Science*, 370(6515), 467–472. <https://doi.org/10.1126/science.aay7367>
- Bergstrom, K., & Xia, L. (2022). The barrier and beyond: Roles of intestinal mucus and mucin-type O-glycosylation in resistance and tolerance defense strategies guiding host-microbe symbiosis. *Gut Microbes*, 14(1), 1–29. <https://doi.org/10.1080/19490976.2022.2052699>
- Berkhout, M. D., Plugge, C. M., & Belzer, C. (2022). How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding. *Glycobiology*, 32(3), 182–200. <https://doi.org/10.1093/glycob/cwab105>
- Bevins, C. L. (2006). Paneth cell defensins: Key effector molecules of innate immunity. *Biochemical Society Transactions*, 34(2), 263–266. <https://doi.org/10.1042/BST20060263>

- Bjursell, M. K., Martens, E. C., & Gordon, J. I. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *Journal of Biological Chemistry*, *281*(47), 36269–36279. <https://doi.org/10.1074/jbc.M606509200>
- Blaak, E. E., Canfora, E. E., Theis, S., Frost, G., Groen, A. K., Mithieux, G., Nauta, A., Scott, K., Stahl, B., van Harselaar, J., van Tol, R., Vaughan, E. E., & Verbeke, K. (2020). Short chain fatty acids in human gut and metabolic health. *Beneficial Microbes*, *11*(5), 411–455. <https://doi.org/10.3920/BM2020.0057>
- Blanton, L. V., Charbonneau, M. R., Salih, T., Barratt, M. J., Ilkaveya, O., Subramanian, S., Manary, M. J., Trehan, I., Jorgensen, J. M., Fan, Y., Henrissat, B., Leyn, S. A., Dmitry, A., Osterman, A. L., Maleta, K. M., Newgard, C. B., Ashorn, P., Dewey, K. G., Gordon, J., ... Nutrition, C. (2016). Gut bacteria that rescue growth impairments transmitted by immature microbiota from undernourished children. *Science*, *351*(6275), 1–18. <https://doi.org/10.1126/science.aad3311>.Gut
- Bode, L. (2012). Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology*, *22*(9), 1147–1162. <https://doi.org/10.1093/glycob/cws074>
- Bode, L. (2015). The functional biology of human milk oligosaccharides. *Early Human Development*, *91*(11), 619–622. <https://doi.org/10.1016/j.earlhumdev.2015.09.001>
- Boekhorst, J., Helmer, Q., Kleerebezem, M., & Siezen, R. J. (2006). Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria. *Microbiology*, *152*(1), 273–280. <https://doi.org/10.1099/mic.0.28415-0>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- Bonafonte, M. A., Solano, C., Sesma, B., Alvarez, M., Montuenga, L., García-Ros, D., & Gamazo, C. (2000). The relationship between glycogen synthesis, biofilm formation and virulence in *Salmonella enteritidis*. *FEMS Microbiology Letters*, *191*(1), 31–36. [https://doi.org/10.1016/S0378-1097\(00\)00366-9](https://doi.org/10.1016/S0378-1097(00)00366-9)
- Bourassa, M. W., Alim, I., Bultman, S. J., & Ratan, R. R. (2016). Butyrate, neuroepigenetics and the gut microbiome: Can a high fiber diet improve brain health? *Neuroscience Letters*, *625*, 56–63. <https://doi.org/10.1016/j.neulet.2016.02.009>
- Brigham, C., Caughlan, R., Gallegos, R., Dallas, M. B., Godoy, V. G., & Malamy, M. H. (2009). Sialic acid (N-acetyl neuraminic acid) utilization by *Bacteroides fragilis* requires a novel N-acetyl mannosamine epimerase. *Journal of Bacteriology*, *191*(11), 3629–3638. <https://doi.org/10.1128/JB.00811-08>
- Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N., & Crouch, L. I. (2019). Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*, *4*(9), 1571–1581. <https://doi.org/10.1038/s41564-019-0466-x>
- Brockhausen, I., Wandall, H. H., Ten Hagen, K. G., & Stanley, P. (2022). O-GalNAc Glycans. In A. Varki, J. D. E. Richard D. Cummings, & and P. H. S. Pamela Stanley, Gerald W. Hart, Markus Aebi, Debra Mohnen, Taroh Kinoshita, Nicolle H. Packer, James H. Prestegard, Ronald L.

- Schnaar (Eds.), *Essentials of Glycobiology. 4th edition.* (4th Editio). Laboratory Press; 2022. <https://www.ncbi.nlm.nih.gov/books/NBK579918/> doi: 10.1101/9781621824213
- Bruel, L., Sulzenbacher, G., Tison, M. C., Pujol, A., Nicoletti, C., Perrier, J., Galinier, A., Ropartz, D., Fons, M., Pompeo, F., & Giardina, T. (2011).  $\alpha$ -galactosidase/sucrose kinase (AgaSK), a novel bifunctional enzyme from the human microbiome coupling galactosidase and kinase activities. *Journal of Biological Chemistry*, *286*(47), 40814–40823. <https://doi.org/10.1074/jbc.M111.286039>
- Brune, A., & Friedrich, M. (2000). Microecology of the termite gut: Structure and function on a microscale. *Current Opinion in Microbiology*, *3*(3), 263–269. [https://doi.org/10.1016/S1369-5274\(00\)00087-4](https://doi.org/10.1016/S1369-5274(00)00087-4)
- Buffie, C. G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A., Viale, A., Ubeda, C., Xavier, J., & Pamer, E. G. (2012). Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to *Clostridium difficile*-induced colitis. *Infection and Immunity*, *80*(1), 62–73. <https://doi.org/10.1128/IAI.05496-11>
- Bunesova, V., Lacroix, C., & Schwab, C. (2018). Mucin Cross-Feeding of Infant Bifidobacteria and *Eubacterium hallii*. *Microbial Ecology*, *75*(1), 228–238. <https://doi.org/10.1007/s00248-017-1037-4>
- Burger-van Paassen, N., Vincent, A., Puiman, P. J., van der Sluis, M., Bouma, J., Boehm, G., van Goudoever, J. B., Van Seuningen, I., & Renes, I. B. (2009). The regulation of intestinal mucin MUC2 expression by short-chain fatty acids: Implications for epithelial protection. *Biochemical Journal*, *420*(2), 211–219. <https://doi.org/10.1042/BJ20082222>
- Buttimer, C., Khokhlova, E. V., Stein, L., Hueston, C. M., Govi, B., Draper, L. A., Ross, R. P., Shkorporov, A. N., & Hill, C. (2023). Temperate bacteriophages infecting the mucin-degrading bacterium *Ruminococcus gnavus* from the human gut. *Gut Microbes*, *15*(1). <https://doi.org/10.1080/19490976.2023.2194794>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cani, P. D., Depommier, C., Derrien, M., Everard, A., & de Vos, W. M. (2022). Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. *Nature Reviews Gastroenterology and Hepatology*, *19*(10), 625–637. <https://doi.org/10.1038/s41575-022-00631-9>
- Capon, C., Maes, E., Michalski, J. C., Leffler, H., & Kim, Y. S. (2001). Sda-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochemical Journal*, *358*(3), 657–664. <https://doi.org/10.1042/bj3580657>
- Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R. A., & Holt, R. A. (2012). *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Research*, *22*(2), 299–306. <https://doi.org/10.1101/gr.126516.111>
- Cerqueira, F. M., Photenhauer, A. L., Pollet, R. M., Brown, H. A., & Koropatkin, N. M. (2020). Starch Digestion by Gut Bacteria: Crowdsourcing for Carbs. *Trends in Microbiology*, *28*(2), 95–108. <https://doi.org/10.1016/j.tim.2019.09.004>
- Cervera-Tison, M., Tailford, L. E., Fuell, C., Bruel, L., Sulzenbacher, G., Henrissat, B., Berrin, J. G., Fons, M., Giardina, T., & Juge, N. (2012). Functional analysis of family GH36  $\alpha$ -galactosidases



from *Ruminococcus gnavus* E1: Insights into the metabolism of a plant oligosaccharide by a human gut symbiont. In *Applied and Environmental Microbiology* (Vol. 78, Issue 21, pp. 7720–7732). <https://doi.org/10.1128/AEM.01350-12>

- Charbonneau, M. R., O'Donnell, D., Blanton, L. V., Totten, S. M., Davis, J. C. C., Barratt, M. J., Cheng, J., Guruge, J., Talcott, M., Bain, J. R., Muehlbauer, M. J., Ilkayeva, O., Wu, C., Struckmeyer, T., Barile, D., Mangani, C., Jorgensen, J., Fan, Y. M., Maleta, K., ... Gordon, J. I. (2016). Sialylated Milk Oligosaccharides Promote Microbiota-Dependent Growth in Models of Infant Undernutrition. *Cell*, *164*(5), 859–871. <https://doi.org/10.1016/j.cell.2016.01.024>
- Chassard, C., Delmas, E., Robert, C., Lawson, P. A., & Bernalier-Donadille, A. (2011). *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from human gut microbiota. *International Journal of Systematic and Evolutionary Microbiology*, *62*(1), 138–143. <https://doi.org/10.1099/ijs.0.027375-0>
- Chen, J., Wright, K., Davis, J. M., Jeraldo, P., Marietta, E. V., Murray, J., Nelson, H., Matteson, E. L., & Taneja, V. (2016). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Medicine*, *8*(1), 1–14. <https://doi.org/10.1186/s13073-016-0299-7>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen, W., Liu, F., Ling, Z., Tong, X., & Xiang, C. (2012). Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS ONE*, *7*(6). <https://doi.org/10.1371/journal.pone.0039743>
- Cheng, H. Y., Ning, M. X., Chen, D. K., & Ma, W. T. (2019). Interactions between the gut microbiota and the host innate immune response against pathogens. *Frontiers in Immunology*, *10*(MAR), 1–11. <https://doi.org/10.3389/fimmu.2019.00607>
- Cheng, K. C., Demirci, A., & Catchmark, J. M. (2011). Pullulan: Biosynthesis, production, and applications. *Applied Microbiology and Biotechnology*, *92*(1), 29–44. <https://doi.org/10.1007/s00253-011-3477-y>
- Cheong, K.-A., Kim, T.-J., Yoon, J.-W., Park, C.-S., Lee, T.-S., Kim, Y.-B., Park, K.-H., & Kim, J.-W. (2002). Catalytic activities of intracellular dimeric neopullulanase on cyclodextrin, acarbose and maltose. *Biotechnology and Applied Biochemistry*, *35*(1), 27. <https://doi.org/10.1042/ba20010052>
- Chia, L. W., Hornung, B. V. H., Aalvink, S., Schaap, P. J., de Vos, W. M., Knol, J., & Belzer, C. (2018). Deciphering the trophic interaction between *Akkermansia muciniphila* and the butyrogenic gut commensal *Anaerostipes caccae* using a metatranscriptomic approach. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *111*(6), 859–873. <https://doi.org/10.1007/s10482-018-1040-x>
- Chua, H. H., Chou, H. C., Tung, Y. L., Chiang, B. L., Liao, C. C., Liu, H. H., & Ni, Y. H. (2018). Intestinal Dysbiosis Featuring Abundance of *Ruminococcus gnavus* Associates With Allergic Diseases in Infants. *Gastroenterology*, *154*(1), 154–167. <https://doi.org/10.1053/j.gastro.2017.09.006>
- Cockburn, D. W., Kibler, R., Brown, H. A., Duvall, R., Moraš, S., Bayer, E., & Koropatkin, N. M. (2021). Structure and substrate recognition by the *Ruminococcus bromii* amylosome pullulanases. *Journal of Structural Biology*, *213*(3), 107765. <https://doi.org/10.1016/j.jsb.2021.107765>
- Cockburn, D. W., Orlovsky, N. I., Foley, M. H., Kwiatkowski, K. J., Bahr, C. M., Maynard, M.,

- Demeler, B., & Koropatkin, N. M. (2015). Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*. *Molecular Microbiology*, *95*(2), 209–230. <https://doi.org/10.1111/mmi.12859>
- Coker, J. K., Moyne, O., Rodionov, D. A., & Zengler, K. (2021). Carbohydrates great and small, from dietary fiber to sialic acids: How glycans influence the gut microbiome and affect human health. *Gut Microbes*, *13*(1), 1–18. <https://doi.org/10.1080/19490976.2020.1869502>
- Coletto, E., Latousakis, D., Pontifex, M. G., Crost, E. H., Vaux, L., Perez Santamarina, E., Goldson, A., Brion, A., Hajihosseini, M. K., Vauzour, D., Savva, G. M., & Juge, N. (2022). The role of the mucin-glycan foraging *Ruminococcus gnavus* in the communication between the gut and the brain. *Gut Microbes*, *14*(1). <https://doi.org/10.1080/19490976.2022.2073784>
- Companys, J., Gosalbes, M. J., Pla-Pagà, L., Calderón-Pérez, L., Llauradó, E., Pedret, A., Valls, R. M., Jiménez-Hernández, N., Sandoval-Ramirez, B. A., Del Bas, J. M., Caimari, A., Rubió, L., & Solà, R. (2021). Gut microbiota profile and its association with clinical variables and dietary intake in overweight/obese and lean subjects: A cross-sectional study. *Nutrients*, *13*(6). <https://doi.org/10.3390/nu13062032>
- Cooling, L. (2015). Blood groups in infection and host susceptibility. *Clinical Microbiology Reviews*, *28*(3), 801–870. <https://doi.org/10.1128/CMR.00109-14>
- Crost, E. H., Coletto, E., Bell, A., & Juge, N. (2023). *Ruminococcus gnavus*: friend or foe for human health. *FEMS Microbiology Reviews*, *47*(2), 1–23. <https://doi.org/10.1093/femsre/fuad014>
- Crost, E. H., Le Gall, G., Laverde-Gomez, J. A., Mukhopadhyaya, I., Flint, H. J., & Juge, N. (2018). Mechanistic insights into the cross-feeding of *Ruminococcus gnavus* and *Ruminococcus bromii* on host and dietary carbohydrates. *Frontiers in Microbiology*, *9*(NOV), 1–13. <https://doi.org/10.3389/fmicb.2018.02558>
- Crost, E. H., Tailford, L. E., Le Gall, G., Fons, M., Henrissat, B., & Juge, N. (2013). Utilisation of Mucin Glycans by the Human Gut Symbiont *Ruminococcus gnavus* Is Strain-Dependent. *PLoS ONE*, *8*(10). <https://doi.org/10.1371/journal.pone.0076341>
- Crost, E. H., Tailford, L. E., Monestier, M., Swarbreck, D., Henrissat, B., Crossman, L. C., & Juge, N. (2016). The mucin-degradation strategy of *Ruminococcus gnavus*: The importance of intramolecular trans-sialidases. *Gut Microbes*, *7*(4), 302–312. <https://doi.org/10.1080/19490976.2016.1186334>
- Crouch, L. I., Liberato, M. V., Urbanowicz, P. A., Baslé, A., Lamb, C. A., Stewart, C. J., Cooke, K., Doona, M., Needham, S., Brady, R. R., Berrington, J. E., Madunic, K., Wuhrer, M., Chater, P., Pearson, J. P., Glowacki, R., Martens, E. C., Zhang, F., Linhardt, R. J., ... Bolam, D. N. (2020). Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-17847-5>
- Culp, E. J., & Goodman, A. L. (2023). Cross-feeding in the gut microbiome: Ecology and mechanisms. *Cell Host and Microbe*, *31*(4), 485–499. <https://doi.org/10.1016/j.chom.2023.03.016>
- Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Verweken, W., Hakki, Z., Speciale, G., ... Gilbert, H. J. (2015). Corrigendum: Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature*, *520*(7547), 388. <https://doi.org/10.1038/nature14334>

- Dai, L., Mafra, D., Shiels, P. G., Hackeng, T. M., Stenvinkel, P., & Schurgers, L. J. (2023). *Vitamin K and Hallmarks of Ageing : Focus on Diet and Gut Microbiome*. 1–15.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. <https://doi.org/10.1093/gigascience/giab008>
- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet Rapidly Alters the Human Gut Microbiota. *Nature*, *505*(7484), 559–563. <https://doi.org/10.1038/nature12820>.Diet
- De Bolós, C., Garrido, M., & Real, F. X. (1995). MUC6 apomucin shows a distinct normal tissue distribution that correlates with Lewis antigen expression in the human stomach. *Gastroenterology*, *109*(3), 723–734. [https://doi.org/10.1016/0016-5085\(95\)90379-8](https://doi.org/10.1016/0016-5085(95)90379-8)
- Debonnie, J. C., & Phillips, S. F. (1978). Capacity of the human colon to absorb fluid. *Gastroenterology*, *74*(4), 698–703. [https://doi.org/10.1016/0016-5085\(78\)90246-9](https://doi.org/10.1016/0016-5085(78)90246-9)
- Den Besten, G., Van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D. J., & Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of Lipid Research*, *54*(9), 2325–2340. <https://doi.org/10.1194/jlr.R036012>
- Derrien, M., Vaughan, E. E., Plugge, C. M., & de Vos, W. M. (2004). Akkermansia municipiphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *54*(5), 1469–1476. <https://doi.org/10.1099/ijs.0.02873-0>
- Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Núñez, G., & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, *167*(5), 1339–1353.e21. <https://doi.org/10.1016/j.cell.2016.10.043>
- Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research*, *45*(W1), W180–W188. <https://doi.org/10.1093/nar/gkx295>
- Donaldson, G. P., Lee, S. M., & Mazmanian, S. K. (2015). Gut biogeography of the bacterial microbiota. *Nature Reviews Microbiology*, *14*(1), 20–32. <https://doi.org/10.1038/nrmicro3552>
- Donaldson, Ladinsky, M. S., Yu, K. B., Sanders, J. G., Yoo, B. B., Chou, W. C., Conner, M. E., Earl, A. M., Knight, R., Bjorkman, P. J., & Mazmanian, S. K. (2018). Gut microbiota utilize immunoglobulin a for mucosal colonization. *Science*, *360*(6390), 795–800. <https://doi.org/10.1126/science.aaq0926>
- Doner, L. W., & Irwin, P. L. (1992). Assay of reducing end-groups in oligosaccharide homologues with 2,2'-bicinchoninate. In *Analytical Biochemistry* (Vol. 202, Issue 1, pp. 50–53). [https://doi.org/10.1016/0003-2697\(92\)90204-K](https://doi.org/10.1016/0003-2697(92)90204-K)
- Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K., & Bultman, S. J. (2011). The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the Mammalian Colon. *Cell Metabolism*, *13*(5), 517–526. <https://doi.org/10.1016/j.cmet.2011.02.018>

- Ducarmon, Q. R., Zwiittink, R. D., Hornung, B. V. H., van Schaik, W., Young, V. B., & Kuijper, E. J. (2019). Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. *Microbiology and Molecular Biology Reviews*, 83(3), 1–29. <https://doi.org/10.1128/mubr.00007-19>
- Dumont, M. G., & Hernández García, M. (2019). *Stable Isotope Probing* (M. G. Dumont & M. Hernández García (eds.); Vol. 2046, Issue August). Springer New York. <https://doi.org/10.1007/978-1-4939-9721-3>
- Duncan, S. H., Hold, G. L., Harmsen, H. J. M., Stewart, C. S., & Flint, H. J. (2002). Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 52(6), 2141–2146. <https://doi.org/10.1099/ijs.0.02241-0>
- Earley, H., Lennon, G., Balfe, Á., Coffey, J. C., Winter, D. C., & O’Connell, P. R. (2019). The abundance of *Akkermansia muciniphila* and its relationship with sulphated colonic mucins in health and ulcerative colitis. *Scientific Reports*, 9(1), 1–9. <https://doi.org/10.1038/s41598-019-51878-3>
- Egan, M., Motherway, M. O., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M., & van Sinderen, D. (2014). Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium (ene bifido profiteert van glucosidase activiteit van de andere). *BMC Microbiology*, 14, 1–14.
- Egert, M., De Graaf, A. A., Maathuis, A., De Waard, P., Plugge, C. M., Smidt, H., Deutz, N. E. P., Dijkema, C., De Vos, W. M., & Venema, K. (2007). Identification of glucose-fermenting bacteria present in an in vitro model of the human intestine by RNA-stable isotope probing. *FEMS Microbiology Ecology*, 60(1), 126–135. <https://doi.org/10.1111/j.1574-6941.2007.00281.x>
- Ehrlich, A. M., Pacheco, A. R., Henrick, B. M., Taft, D., Xu, G., Huda, M. N., Mishchuk, D., Goodson, M. L., Slupsky, C., Barile, D., Lebrilla, C. B., Stephensen, C. B., Mills, D. A., & Raybould, H. E. (2020). Indole-3-lactic acid associated with *Bifidobacterium*-dominated microbiota significantly decreases inflammation in intestinal epithelial cells. *BMC Microbiology*, 20(1), 1–13. <https://doi.org/10.1186/s12866-020-02023-y>
- Endt, K., Stecher, B., Chaffron, S., Slack, E., Tchitchek, N., Benecke, A., van Maele, L., Sirard, J. C., Mueller, A. J., Heikenwalder, M., Macpherson, A. J., Strugnell, R., von Mering, C., & Hardt, W. D. (2010). The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal salmonella diarrhea. *PLoS Pathogens*, 6(9). <https://doi.org/10.1371/journal.ppat.1001097>
- Engelstoft, M. S., Egerod, K. L., Lund, M. L., & Schwartz, T. W. (2013). Enteroendocrine cell types revisited. *Current Opinion in Pharmacology*, 13(6), 912–921. <https://doi.org/10.1016/j.coph.2013.09.018>
- Engfer, M. B., Stahl, B., Finke, B., Sawatzki, G., & Daniel, H. (2000). Human milk oligosaccharides are resistant to enzymatic hydrolysis in the upper gastrointestinal tract. *American Journal of Clinical Nutrition*, 71(6), 1589–1596. <https://doi.org/10.1093/ajcn/71.6.1589>
- Ermund, A., Schütte, A., Johansson, M. E. V., Gustafsson, J. K., & Hansson, G. C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer’s patches. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 305(5), 341–347. <https://doi.org/10.1152/ajpgi.00046.2013>

- Etienne-Mesmin, L., Chassaing, B., Desvaux, M., De Paepe, K., Gresse, R., Sauvaitre, T., Forano, E., De Wiele, T. Van, Schüller, S., Juge, N., & Blanquet-Diot, S. (2019). Experimental models to study intestinal microbes–mucus interactions in health and disease. *FEMS Microbiology Reviews*, 43(5), 457–489. <https://doi.org/10.1093/femsre/fuz013>
- Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J. P., Druart, C., Bindels, L. B., Guiot, Y., Derrien, M., Muccioli, G. G., Delzenne, N. M., De Vos, W. M., & Cani, P. D. (2013). Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America*, 110(22), 9066–9071. <https://doi.org/10.1073/pnas.1219451110>
- Evert, C., Loesekann, T., Bhat, G., Shajahan, A., Sonon, R., Azadi, P., & Hunter, R. C. (2019). Generation of <sup>13</sup>C-Labeled MUC5AC Mucin Oligosaccharides for Stable Isotope Probing of Host-Associated Microbial Communities. *ACS Infectious Diseases*, 5(3), 385–393. <https://doi.org/10.1021/acsinfecdis.8b00296>
- Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, 38(3), 276–278. <https://doi.org/10.1038/s41587-020-0439-x>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Fan, S., Zhang, H., Chen, X., Lu, L., Xu, L., & Xiao, M. (2016). Cloning, characterization, and production of three  $\alpha$ -l-fucosidases from *Clostridium perfringens* ATCC 13124. In *Journal of Basic Microbiology* (Vol. 56, Issue 4, pp. 347–357). <https://doi.org/10.1002/jobm.201500582>
- Fang, J., Wang, H., Zhou, Y., Zhang, H., Zhou, H., & Zhang, X. (2021). Slimy partners: the mucus barrier and gut microbiome in ulcerative colitis. *Experimental and Molecular Medicine*, 53(5), 772–787. <https://doi.org/10.1038/s12276-021-00617-8>
- Feng, J., Qian, Y., Zhou, Z., Ertmer, S., Vivas, E. I., Lan, F., Hamilton, J. J., Rey, F. E., Anantharaman, K., & Venturelli, O. S. (2022). Polysaccharide utilization loci in *Bacteroides* determine population fitness and community-level interactions. *Cell Host and Microbe*, 30(2), 200–215.e12. <https://doi.org/10.1016/j.chom.2021.12.006>
- Feuerbaum, S., Saile, N., Pohlentz, G., Müthing, J., & Schmidt, H. (2018). De-O-Acetylation of mucin-derived sialic acids by recombinant NanS-p esterases of *Escherichia coli* O157:H7 strain EDL933. *International Journal of Medical Microbiology*, 308(8), 1113–1120. <https://doi.org/10.1016/j.ijmm.2018.10.001>
- Fu, J., Wei, B., Wen, T., Johansson, M. E. V., Liu, X., Bradford, E., Thomsson, K. a, McGee, S., Mansour, L., Tong, M., McDaniel, J. M., Sferra, T. J., Turner, J. R., Chen, H., Hansson, G. C., Braun, J., & Xia, L. (2011). Loss of intestinal core 1–derived O-glycans causes spontaneous colitis in mice. *Journal of Clinical Investigation*, 121(4), 1657–1666. <https://doi.org/10.1172/JCI45538>
- Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H., & Yamamoto, K. (2005). Identification and molecular cloning of a novel glycoside hydrolase family of core 1 type O-glycan-specific endo- $\alpha$ -N-acetylgalactosaminidase from *Bifidobacterium longum*. In *Journal of Biological Chemistry* (Vol. 280, Issue 45, pp. 37415–37422). <https://doi.org/10.1074/jbc.M506874200>
- Fujita, M., Tsuchida, A., Hirata, A., Kobayashi, N., Goto, K., Osumi, K., Hirose, Y., Nakayama, J.,

- Yamanoi, T., Ashida, H., & Mizuno, M. (2011). Glycoside hydrolase family 89  $\alpha$ -N-acetylglucosaminidase from clostridium perfringens specifically acts on GlcNAc $\alpha$ 1,4Gal $\beta$ 1R at the non-reducing terminus of O-glycans in fastric mucin. In *Journal of Biological Chemistry* (Vol. 286, Issue 8, pp. 6479–6489). <https://doi.org/10.1074/jbc.M110.206722>
- Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., Takahashi, M., Fukuda, N. N., Murakami, S., Miyauchi, E., Hino, S., Atarashi, K., Onawa, S., Fujimura, Y., Lockett, T., ... Ohno, H. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature*, *504*(7480), 446–450. <https://doi.org/10.1038/nature12721>
- Gallardo-Becerra, L., Cornejo-Granados, F., García-López, R., Valdez-Lara, A., Bikel, S., Canizales-Quinteros, S., López-Contreras, B. E., Mendoza-Vargas, A., Nielsen, H., & Ochoa-Leyva, A. (2020). Metatranscriptomic analysis to define the Secrebiome, and 16S rRNA profiling of the gut microbiome in obesity and metabolic syndrome of Mexican children. *Microbial Cell Factories*, *19*(1), 1–18. <https://doi.org/10.1186/s12934-020-01319-y>
- Garcher, C., Bara, J., Bron, A., & Oriol, R. (1994). Expression of mucin peptide and blood group ABH- and Lewis-related carbohydrate antigens in normal human conjunctiva. *Investigative Ophthalmology and Visual Science*, *35*(3), 1184–1191.
- Gaudier, E., Jarry, A., Blottière, H. M., De Coppet, P., Buisine, M. P., Aubert, J. P., Labois, C., Cherbut, C., & Hoebler, C. (2004). Butyrate specifically modulates MUC gene expression in intestinal epithelial goblet cells deprived of glucose. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *287*(6 50-6), 1168–1174. <https://doi.org/10.1152/ajpgi.00219.2004>
- Ghannoum, M. A., & Mukherjee, P. K. (2013). The human mycobiome and its impact on health and disease. *Current Fungal Infection Reports*, *7*(4), 345–350. <https://doi.org/10.1007/s12281-013-0162-x>
- Goh, Y. J., & Klaenhammer, T. R. (2013). A functional glycogen biosynthesis pathway in *Lactobacillus acidophilus*: Expression and analysis of the glg operon. *Molecular Microbiology*, *89*(6), 1187–1200. <https://doi.org/10.1111/mmi.12338>
- Grabinger, T., Garzon, J. F. G., Hausmann, M., Geirnaert, A., Lacroix, C., & Henet, T. (2019). Alleviation of intestinal inflammation by oral supplementation with 2-fucosyllactose in mice. *Frontiers in Microbiology*, *10*(JUN), 1–14. <https://doi.org/10.3389/fmicb.2019.01385>
- Grahnmø, L., Nethander, M., Coward, E., Gabrielsen, M. E., Sree, S., Billod, J. M., Engstrand, L., Abrahamsson, S., Langhammer, A., Hveem, K., & Ohlsson, C. (2022). Cross-sectional associations between the gut microbe *Ruminococcus gnavus* and features of the metabolic syndrome. *The Lancet Diabetes and Endocrinology*, *10*(7), 481–483. [https://doi.org/10.1016/S2213-8587\(22\)00113-9](https://doi.org/10.1016/S2213-8587(22)00113-9)
- Greenwood-Van Meerveld, B., Johnson, A. C., & Grundy, D. (2017). Gastrointestinal Physiology and Function. In B. Greenwood-Van Meerveld (Ed.), *Gastrointestinal Pharmacology* (pp. 1–16). Springer International Publishing. [https://doi.org/10.1007/164\\_2016\\_118](https://doi.org/10.1007/164_2016_118)
- Grondin, J. A., Kwon, Y. H., Far, P. M., Haq, S., & Khan, W. I. (2020). Mucins in Intestinal Mucosal Defense and Inflammation: Learning From Clinical and Experimental Studies. *Frontiers in Immunology*, *11*(September), 1–19. <https://doi.org/10.3389/fimmu.2020.02054>
- Gunning, A. P., Kirby, A. R., Fuell, C., Pin, C., Tailford, L. E., & Juge, N. (2013). Mining the “glycocode” - Exploring the spatial distribution of glycans in gastrointestinal mucin using force spectroscopy. *FASEB Journal*, *27*(6), 2342–2354. <https://doi.org/10.1096/fj.12-221416>

- Gut, H., King, S. J., & Walsh, M. A. (2008). Structural and functional studies of *Streptococcus pneumoniae* neuraminidase B: An intramolecular trans-sialidase. *FEBS Letters*, *582*(23–24), 3348–3352. <https://doi.org/10.1016/j.febslet.2008.08.026>
- Guzman-Aranguéz, A., & Argüeso, P. (2010). Structure and biological roles of mucin-type O-glycans at the ocular surface. *The Ocular Surface*, *8*(1), 8–17. [https://doi.org/10.1016/s1542-0124\(12\)70213-6](https://doi.org/10.1016/s1542-0124(12)70213-6)
- Hall, A. B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G. K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D., Ananthakrishnan, A. N., Kugathasan, S., Lander, E. S., Blainey, P., ... Huttenhower, C. (2017). A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Medicine*, *9*(1), 1–12. <https://doi.org/10.1186/s13073-017-0490-5>
- Hamaker, B. R., & Tuncil, Y. E. (2014). A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. *Journal of Molecular Biology*, *426*(23), 3838–3850. <https://doi.org/10.1016/j.jmb.2014.07.028>
- Hansson, G. C. (2019). Mucus and mucins in diseases of the intestinal and respiratory tracts. *Journal of Internal Medicine*, *285*(5), 479–490. <https://doi.org/10.1111/joim.12910>
- Hansson, Gunnar C. (2020). Mucins and the Microbiome. *Annual Review of Biochemistry*, *89*(66), 769–793. <https://doi.org/10.1146/annurev-biochem-011520-105053>
- Hata, D. J., & Smith, D. S. (2004). Blood Group B Degrading Activity of *Ruminococcus gnavus*  $\alpha$ -Galactosidase. *Artificial Cells, Blood Substitutes, and Biotechnology*, *32*(2), 263–274. <https://doi.org/10.1081/BIO-120037831>
- Henke, M. T., Brown, E. M., Cassilly, C. D., Vlamakis, H., Xavier, R. J., & Clardy, J. (2021). Capsular polysaccharide correlates with immune response to the human gut microbe *Ruminococcus gnavus*. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(20), 1–7. <https://doi.org/10.1073/pnas.2007595118>
- Henke, M. T., Kenny, D. J., Cassilly, C. D., Vlamakis, H., Xavier, R. J., & Clardy, J. (2019). *Ruminococcus gnavus*, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(26), 12672–12677. <https://doi.org/10.1073/pnas.1904099116>
- Herrmann, E., Young, W., Reichert-Grimm, V., Weis, S., Riedel, C. U., Rosendale, D., Stoklosinski, H., Hunt, M., & Egert, M. (2018). In vivo assessment of resistant starch degradation by the caecal microbiota of mice using rna-based stable isotope probing—a proof-of-principle study. *Nutrients*, *10*(2). <https://doi.org/10.3390/nu10020179>
- Herrmann, E., Young, W., Rosendale, D., Conrad, R., Riedel, C. U., & Egert, M. (2017). Determination of resistant starch assimilating bacteria in fecal samples of mice by in vitro RNA-based stable isotope probing. *Frontiers in Microbiology*, *8*(JUL), 1–13. <https://doi.org/10.3389/fmicb.2017.01331>
- Herrmann, E., Young, W., Rosendale, D., Reichert-Grimm, V., Riedel, C. U., Conrad, R., & Egert, M. (2017). RNA-Based Stable Isotope Probing Suggests *Allobaculum* spp. as Particularly Active Glucose Assimilators in a Complex Murine Microbiota Cultured in Vitro. *BioMed Research International*, *2017*. <https://doi.org/10.1155/2017/1829685>
- Heuer, H., Krsek, M., Baker, P., Smalla, K., & Wellington, E. M. H. (1997). Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic

- separation in denaturing gradients. *Applied and Environmental Microbiology*, 63(8), 3233–3241. <https://doi.org/10.1128/aem.63.8.3233-3241.1997>
- Higgins, M. A., Whitworth, G. E., El Warry, N., Randriantsoa, M., Samain, E., Burke, R. D., Vocadlo, D. J., & Boraston, A. B. (2009). Differential recognition and hydrolysis of host carbohydrate antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases. *Journal of Biological Chemistry*, 284(38), 26161–26173. <https://doi.org/10.1074/jbc.M109.024067>
- Hii, S. L., Tan, J. S., Ling, T. C., & Ariff, A. Bin. (2012). Pullulanase: Role in starch hydrolysis and potential industrial applications. *Enzyme Research*, 2012. <https://doi.org/10.1155/2012/921362>
- Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., Lewis, J. D., & Bushman, F. D. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0066019>
- Hoffmann, T. W., Pham, H. P., Bridonneau, C., Aubry, C., Lamas, B., Martin-Gallausiaux, C., Moroldo, M., Rainteau, D., Lapaque, N., Six, A., Richard, M. L., Fargier, E., Le Guern, M. E., Langella, P., & Sokol, H. (2016). Microorganisms linked to inflammatory bowel disease-associated dysbiosis differentially impact host physiology in gnotobiotic mice. *ISME Journal*, 10(2), 460–477. <https://doi.org/10.1038/ismej.2015.127>
- Holmén Larsson, J. M., Karlsson, H., Crespo, J. G., Johansson, M. E. V., Eklund, L., Sjövall, H., & Hansson, G. C. (2011). Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflammatory Bowel Diseases*, 17(11), 2299–2307. <https://doi.org/10.1002/ibd.21625>
- Holmén Larsson, J. M., Karlsson, H., Sjövall, H., & Hansson, G. C. (2009). A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology*, 19(7), 756–766. <https://doi.org/10.1093/glycob/cwp048>
- Holmén Larsson, J. M., Thomsson, K. A., Rodríguez-Piñeiro, A. M., Karlsson, H., & Hansson, G. C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 305(5), 357–363. <https://doi.org/10.1152/ajpgi.00048.2013>
- Hondoh, H., Kuriki, T., & Matsuura, Y. (2003). Three-dimensional structure and substrate binding of *Bacillus stearothermophilus* neopullulanase. *Journal of Molecular Biology*, 326(1), 177–188. [https://doi.org/10.1016/S0022-2836\(02\)01402-X](https://doi.org/10.1016/S0022-2836(02)01402-X)
- Hooks, K. B., & O'Malley, M. A. (2020). Contrasting Strategies: Human Eukaryotic Versus Bacterial Microbiome Research. *Journal of Eukaryotic Microbiology*, 67(2), 279–295. <https://doi.org/10.1111/jeu.12766>
- Hooper, L. V., Stappenbeck, T. S., Hong, C. V., & Gordon, J. I. (2003). Angiogenins: A new class of microbicidal proteins involved in innate immunity. *Nature Immunology*, 4(3), 269–273. <https://doi.org/10.1038/ni888>
- Hu, J., Ye, H., Wang, S., Wang, J., & Han, D. (2021). Prophage Activation in the Intestine: Insights Into Functions and Possible Applications. *Frontiers in Microbiology*, 12(December), 1–13. <https://doi.org/10.3389/fmicb.2021.785634>
- Huang, J. Y., Lee, S. M., & Mazmanian, S. K. (2011). The human commensal *Bacteroides fragilis* binds intestinal mucin. *Anaerobe*, 17(4), 137–141. <https://doi.org/10.1016/j.anaerobe.2011.05.017>



- Huang, K., Wang, M. M., Kulinich, A., Yao, H. L., Ma, H. Y., Martínez, J. E. R., Duan, X. C., Chen, H., Cai, Z. P., Flitsch, S. L., Liu, L., & Voglmeir, J. (2015). Biochemical characterisation of the neuraminidase pool of the human gut symbiont *Akkermansia muciniphila*. *Carbohydrate Research*, *415*, 60–65. <https://doi.org/10.1016/j.carres.2015.08.001>
- Huang, P., Wu, S., Yang, S., Yan, Q., & Jiang, Z. (2020). Structural basis of carbohydrate binding in domain C of a type i pullulanase from *Paenibacillus barengoltzii*. *Acta Crystallographica Section D: Structural Biology*, *76*(2006), 447–457. <https://doi.org/10.1107/S205979832000409X>
- Huang, Y. L., Chassard, C., Hausmann, M., Von Itzstein, M., & Hennet, T. (2015). Sialic acid catabolism drives intestinal inflammation and microbial dysbiosis in mice. *Nature Communications*, *6*. <https://doi.org/10.1038/ncomms9141>
- Hunt, K. M., Foster, J. A., Forney, L. J., Schütte, U. M. E., Beck, D. L., Abdo, Z., Fox, L. K., Williams, J. E., McGuire, M. K., & McGuire, M. A. (2011). Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS ONE*, *6*(6), 1–8. <https://doi.org/10.1371/journal.pone.0021313>
- Huus, K. E., Petersen, C., & Finlay, B. B. (2021). Diversity and dynamism of IgA–microbiota interactions. In *Nature Reviews Immunology* (Vol. 21, Issue 8, pp. 514–525). <https://doi.org/10.1038/s41577-021-00506-1>
- Ioannou, A., Knol, J., & Belzer, C. (2021). Microbial Glycoside Hydrolases in the First Year of Life: An Analysis Review on Their Presence and Importance in Infant Gut. *Frontiers in Microbiology*, *12*(May), 1–13. <https://doi.org/10.3389/fmicb.2021.631282>
- Jacobi, S. K., Yatsunenkov, T., Li, D., Dasgupta, S., Yu, R. K., Berg, B. M., Chichlowski, M., & Odle, J. (2016). Dietary isomers of sialyllactose increase ganglioside sialic acid concentrations in the corpus callosum and cerebellum and modulate the colonic microbiota of formula-fed piglets. *Journal of Nutrition*, *146*(2), 200–208. <https://doi.org/10.3945/jn.115.220152>
- Janeček, Š., Mareček, F., MacGregor, E. A., & Svensson, B. (2019). Starch-binding domains as CBM families—history, occurrence, structure, function and evolution. *Biotechnology Advances*, *37*(8). <https://doi.org/10.1016/j.biotechadv.2019.107451>
- Janeček, Š., Svensson, B., & MacGregor, E. A. (2014).  $\alpha$ -Amylase: An enzyme specificity found in various families of glycoside hydrolases. In *Cellular and Molecular Life Sciences* (Vol. 71, Issue 7, pp. 1149–1170). <https://doi.org/10.1007/s00018-013-1388-z>
- Jia, Z., Cao, W., & Hernández García, M. (2019). DNA-based stable isotope probing. In M. G. Dumont & M. Hernández García (Eds.), *Stable Isotope Probing* (New York,). Springer New York.
- Jia, Z., & Conrad, R. (2009). Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. *Environmental Microbiology*, *11*(7), 1658–1671. <https://doi.org/10.1111/j.1462-2920.2009.01891.x>
- Jiao, L., Kourkoumpetis, T., Hutchinson, D., Ajami, N. J., Hoffman, K., White, D. L., Graham, D. Y., Hair, C., Shah, R., Kanwal, F., Jarbrink-Sehgal, M., Husain, N., Hernaez, R., Hou, J., Cole, R., Velez, M., Ketwaroo, G., Kramer, J., El-Serag, H. B., & Petrosino, J. F. (2022). Spatial Characteristics of Colonic Mucosa-Associated Gut Microbiota in Humans. *Microbial Ecology*, *83*(3), 811–821. <https://doi.org/10.1007/s00248-021-01789-6>
- Johansson, M. E. V., Ambort, D., Pelaseyed, T., Schütte, A., Gustafsson, J. K., Ermund, A., Subramani, D. B., Holmén Larsson, J. M., Thomsson, K. A., Bergström, J. H., Van Der Post, S.,

- Rodriguez-Piñeiro, A. M., Sjövall, H., Bäckström, M., & Hansson, G. C. (2011). Composition and functional role of the mucus layers in the intestine. *Cellular and Molecular Life Sciences*, *68*(22), 3635–3641. <https://doi.org/10.1007/s00018-011-0822-3>
- Johansson, M. E. V., & Hansson, G. C. (2016). Immunological aspects of intestinal mucus and mucins. *Nature Reviews Immunology*, *16*(10), 639–649. <https://doi.org/10.1038/nri.2016.88>
- Johansson, M. E. V., Holmén Larsson, J. M., & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4659–4665. <https://doi.org/10.1073/pnas.1006451107>
- Johansson, M. E. V., Phillipson, M., Petersson, J., Velcich, A., Holm, L., & Hansson, G. C. (2008). The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(39), 15064–15069. <https://doi.org/10.1073/pnas.0803124105>
- Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P., & Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut*, *60*(5), 631–637. <https://doi.org/10.1136/gut.2010.223263>
- Juge, N. (2012). Microbial adhesins to gastrointestinal mucus. *Trends in Microbiology*, *20*(1), 30–39. <https://doi.org/10.1016/j.tim.2011.10.001>
- Juge, N. (2022). Relationship between mucosa-associated gut microbiota and human diseases. *Biochemical Society Transactions*, *50*(5), 1225–1236. <https://doi.org/10.1042/BST20201201>
- Juge, N., Tailford, L., & Owen, C. D. (2016). Sialidases from gut bacteria: A mini-review. *Biochemical Society Transactions*, *44*, 166–175. <https://doi.org/10.1042/BST20150226>
- Kamitori, S., Abe, A., Ohtaki, A., Kaji, A., Tonzuka, T., & Sakano, Y. (2002). Crystal structures and structural comparison of *Thermoactinomyces vulgaris* R-47  $\alpha$ -amylase 1 (TVAI) at 1.6 Å resolution and  $\alpha$ -amylase 2 (TVAII) at 2.3 Å resolution. In *Journal of Molecular Biology* (Vol. 318, Issue 2, pp. 443–453). [https://doi.org/10.1016/S0022-2836\(02\)00111-0](https://doi.org/10.1016/S0022-2836(02)00111-0)
- Kaoutari, A. El, Armougom, F., Gordon, J. I., Raoult, D., & Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Reviews Microbiology*, *11*(7), 497–504. <https://doi.org/10.1038/nrmicro3050>
- Katayama, T., Fujita, K., & Yamamoto, K. (2005). Novel bifidobacterial glycosidases acting on sugar chains of mucin glycoproteins. *Journal of Bioscience and Bioengineering*, *99*(5), 457–465. <https://doi.org/10.1263/jbb.99.457>
- Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G., & Lowe, J. B. (1995). Sequence and expression of a candidate for the human Secretor blood group  $\alpha$ (1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *Journal of Biological Chemistry*, *270*(9), 4640–4649. <https://doi.org/10.1074/jbc.270.9.4640>
- Khder Mustafa, S., Zrar Omar, S., Kamal Ahmad, K., & Basil Khudhur, L. (2023). The association of ABO blood group distribution and clinical characteristics in patients with SARS-CoV-2. *Journal of Infection in Developing Countries*, *17*(1), 18–22. <https://doi.org/10.3855/jidc.17430>
- Kiela, P. R., & Ghishan, F. K. (2016). Physiology of intestinal absorption and secretion. *Best Practice and Research: Clinical Gastroenterology*, *30*(2), 145–159. <https://doi.org/10.1016/j.bpg.2016.02.007>

- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kim, Y. S., & Ho, S. B. (2010). Intestinal goblet cells and mucins in health and disease: Recent insights and progress. *Current Gastroenterology Reports*, *12*(5), 319–330. <https://doi.org/10.1007/s11894-010-0131-2>
- Kiyohara, M., Nakatomi, T., Kurihara, S., Fushinobu, S., Suzuki, H., Tanaka, T., Shoda, S., Kitaoka, M., Katayama, T., Yamamoto, K., & Ashida, H. (2012).  $\alpha$ -N-Acetylgalactosaminidase from Infant-associated Bifidobacteria Belonging to Novel Glycoside Hydrolase Family 129 Is Implicated in Alternative Mucin Degradation Pathway. *Journal of Biological Chemistry*, *287*(1), 693–700. <https://doi.org/10.1074/jbc.M111.277384>
- Kiyohara, M., Tanigawa, K., Chaiwangsri, T., Katayama, T., Ashida, H., & Yamamoto, K. (2011). An exo- $\alpha$ -sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates. *Glycobiology*, *21*(4), 437–447. <https://doi.org/10.1093/glycob/cwq175>
- Kobata, A. (2010). Structures and application of oligosaccharides in human milk. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, *86*(7), 731–747. <https://doi.org/10.2183/pjab.86.731>
- Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T., & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(SUPPL. 1), 4578–4585. <https://doi.org/10.1073/pnas.1000081107>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>
- Koropatkin, N. M., Cameron, E. a, & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology*, *10*(5), 323–335. <https://doi.org/10.1038/nrmicro2746>
- Kosciow, K., & Deppenmeier, U. (2019). Characterization of a phospholipid-regulated  $\beta$ -galactosidase from Akkermansia muciniphila involved in mucin degradation. In *MicrobiologyOpen* (Vol. 8, Issue 8). <https://doi.org/10.1002/mbo3.796>
- Kosciow, K., & Deppenmeier, U. (2020). Characterization of three novel  $\beta$ -galactosidases from Akkermansia muciniphila involved in mucin degradation. In *International Journal of Biological Macromolecules* (Vol. 149, pp. 331–340). <https://doi.org/10.1016/j.ijbiomac.2020.01.246>
- Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Taberero, J., Baselga, J., Liu, C., Shivdasani, R. A., Ogino, S., Birren, B. W., Huttenhower, C., Garrett, W. S., & Meyerson, M. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Research*, *22*(2), 292–298. <https://doi.org/10.1101/gr.126573.111>
- Kostopoulos, I., Elzinga, J., Ottman, N., Klievink, J. T., Blijenberg, B., Aalvink, S., Boeren, S., Mank, M., Knol, J., de Vos, W. M., & Belzer, C. (2020). Akkermansia muciniphila uses human milk oligosaccharides to thrive in the early life conditions in vitro. *Scientific Reports*, *10*(1), 1–17. <https://doi.org/10.1038/s41598-020-71113-8>
- Kovatcheva-Datchary, P., Egert, M., Maathuis, A., Rajilić-Stojanović, M., De Graaf, A. A., Smidt, H.,

- De Vos, W. M., & Venema, K. (2009). Linking phylogenetic identities of bacteria to starch fermentation in an in vitro model of the large intestine by RNA-based stable isotope probing. *Environmental Microbiology*, *11*(4), 914–926. <https://doi.org/10.1111/j.1462-2920.2008.01815.x>
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Kubinak, J. L., Petersen, C., Stephens, W. Z., Soto, R., Bake, E., O’Connell, R. M., & Round, J. L. (2015). MyD88 signaling in T cells directs IgA-mediated control of the microbiota to promote health. *Cell Host and Microbe*, *17*(2), 153–163. <https://doi.org/10.1016/j.chom.2014.12.009>
- Kunz, C., Rudloff, S., Baier, W., Klein, N., & Strobel, S. (2000). Oligosaccharides in Human Milk: Structural, Functional, and Metabolic Aspects. *Annual Review of Nutrition*, *20*(1), 699–722. <https://doi.org/10.1146/annurev.nutr.20.1.699>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). <https://doi.org/10.18637/jss.v082.i13>
- Lafond, M., Tauzin, A. S., Bruel, L., Laville, E., Lombard, V., Esque, J., André, I., Vidal, N., Pompeo, F., Quinson, N., Perrier, J., Fons, M., Potocki-Veronese, G., & Giardina, T. (2020).  $\alpha$ -Galactosidase and Sucrose-Kinase Relationships in a Bi-functional AgaSK Enzyme Produced by the Human Gut Symbiont *Ruminococcus gnavus* E1. In *Frontiers in Microbiology* (Vol. 11). <https://doi.org/10.3389/fmicb.2020.579521>
- Lagkouvardos, I., Fischer, S., Kumar, N., & Clavel, T. (2017). Rhea: A transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ*, *2017*(1). <https://doi.org/10.7717/peerj.2836>
- Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O., Klintner, S., Pudlo, N. A., Urs, K., Koropatkin, N. M., Creagh, A. L., Haynes, C. A., Kelly, A. G., Cederholm, S. N., Davies, G. J., Martens, E. C., & Brumer, H. (2014). A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature*, *506*(7489), 498–502. <https://doi.org/10.1038/nature12907>
- Lee, H. S., Kim, M. S., Cho, H. S., Kim, J. I., Kim, T. J., Choi, J. H., Park, C., Lee, H. S., Oh, B. H., & Park, K. H. (2002). Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other. *Journal of Biological Chemistry*, *277*(24), 21891–21897. <https://doi.org/10.1074/jbc.M201623200>
- Lenth, R. V., Buerkner, P., Herve, M., Love, J., & Singmann, H. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.3. 2021.* <https://doi.org/10.1080/00031305.1980.10483031>>.License
- Lepage, P., Höslner, R., Spehlmann, M. E., Rehman, A., Zvirbliene, A., Begun, A., Ott, S., Kupcinkas, L., Doré, J., Raedler, A., & Schreiber, S. (2011). Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology*, *141*(1), 227–236. <https://doi.org/10.1053/j.gastro.2011.04.011>
- Li, J., Markowitz, R. H. G., Brooks, A. W., Mallott, E. K., Leigh, B. A., Olszewski, T., Zare, H., Bagheri, M., Smith, H. M., Friese, K. A., Habibi, I., Lawrence, W. M., Rost, C. L., Lédeczi, Á., Eeds, A. M., Ferguson, J. F., Silver, H. J., & Bordenstein, S. R. (2022). Individuality and ethnicity eclipse a short-term dietary intervention in shaping microbiomes and viromes. In *PLoS Biology* (Vol. 20, Issue 8). <https://doi.org/10.1371/journal.pbio.3001758>

- Li, W., Ghosh, T., Bai, Y., Santra, A., Xiao, A., & Chen, X. (2019). A substrate tagging and two-step enzymatic reaction strategy for large-scale synthesis of 2,7-anhydro-sialic acid. *Carbohydrate Research*, 479, 41–47. <https://doi.org/10.1016/j.carres.2019.05.002>
- Li, Z., & Chai, W. (2019). Mucin O-glycan microarrays. *Current Opinion in Structural Biology*, 56, 187–197. <https://doi.org/10.1016/j.sbi.2019.03.032>
- Lindén, S. K., Sheng, Y. H., Every, A. L., Miles, K. M., Skoog, E. C., Florin, T. H. J., Sutton, P., & McGuckin, M. A. (2009). MUC1 limits *Helicobacter pylori* infection both by steric hindrance and by acting as a releasable decoy. *PLoS Pathogens*, 5(10). <https://doi.org/10.1371/journal.ppat.1000617>
- Liu, P., Pommerenke, B., & Conrad, R. (2018). Identification of Syntrophobacteraceae as major acetate-degrading sulfate reducing bacteria in Italian paddy soil. *Environmental Microbiology*, 20(1), 337–354. <https://doi.org/10.1111/1462-2920.14001>
- Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., ... Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, 569(7758), 655–662. <https://doi.org/10.1038/s41586-019-1237-9>
- LoCascio, R. G., Ninonuevo, M. R., Freeman, S. L., Sela, D. A., Grimm, R., Lebrilla, C. B., Mills, D. A., & German, J. B. (2007). Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. *Journal of Agricultural and Food Chemistry*, 55(22), 8914–8919. <https://doi.org/10.1021/jf0710480>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Low, K. E., Smith, S. P., Abbott, D. W., & Boraston, A. B. (2021). The glycoconjugate-degrading enzymes of *Clostridium perfringens*: Tailored catalysts for breaching the intestinal mucus barrier. *Glycobiology*, 31(6), 681–690. <https://doi.org/10.1093/glycob/cwaa050>
- Lueders, T., Pommerenke, B., & Friedrich, M. W. (2004). Stable-isotope probing of microorganisms thriving at thermodynamic limits: Syntrophic propionate oxidation in flooded soil. *Applied and Environmental Microbiology*, 70(10), 5778–5786. <https://doi.org/10.1128/AEM.70.10.5778-5786.2004>
- Luijckx, Y. M. C. A., Bleumink, N. M. C., Jiang, J., Overkleeft, H. S., Wösten, M. M. S. M., Strijbis, K., & Wennekes, T. (2020). *Bacteroides fragilis* fucosidases facilitate growth and invasion of *Campylobacter jejuni* in the presence of mucins. *Cellular Microbiology*, 22(12), 1–12. <https://doi.org/10.1111/cmi.13252>
- Luis, A. S., Baslé, A., Byrne, D. P., Wright, G. S. A., London, J. A., Jin, C., Karlsson, N. G., Hansson, G. C., Eysers, P. A., Czjzek, M., Barbeyron, T., Yates, E. A., Martens, E. C., & Cartmell, A. (2022). Sulfated glycan recognition by carbohydrate sulfatases of the human gut microbiota. *Nature Chemical Biology*, 18(8), 841–849. <https://doi.org/10.1038/s41589-022-01039-x>
- Luis, A. S., Jin, C., Pereira, G. V., Glowacki, R. W. P., Gugel, S. R., Singh, S., Byrne, D. P., Pudlo, N. A., London, J. A., Baslé, A., Reihill, M., Oscarson, S., Eysers, P. A., Czjzek, M., Michel, G., Barbeyron, T., Yates, E. A., Hansson, G. C., Karlsson, N. G., ... Martens, E. C. (2021). A single sulfatase is required to access colonic mucin by a gut bacterium. *Nature*, 598(7880), 332–337. <https://doi.org/10.1038/s41586-021-03967-5>

- Luis, A. S., Yates, E. A., & Cartmell, A. (2022). Functions and specificity of bacterial carbohydrate sulfatases targeting host glycans. *Essays in Biochemistry*, 0(November 2022), 429–442. <https://doi.org/10.1042/ebc20220120>
- Maathuis, A. J. H., van den Heuvel, E. G., Schoterman, M. H. C., & Venema, K. (2012). Galacto-oligosaccharides have prebiotic activity in a dynamic in vitro colon model using a <sup>13</sup>C-labeling technique. *Journal of Nutrition*, 142(7), 1205–1212. <https://doi.org/10.3945/jn.111.157420>
- Macao, B., Johansson, D. G. A., Hansson, G. C., & Härd, T. (2006). Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nature Structural and Molecular Biology*, 13(1), 71–76. <https://doi.org/10.1038/nsmb1035>
- Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, 50(W1), W276–W279. <https://doi.org/10.1093/nar/gkac240>
- Mafra, D., Ribeiro, M., Fonseca, L., Regis, B., Cardozo, L. F. M. F., Fragoso dos Santos, H., Emiliano de Jesus, H., Schultz, J., Shiels, P. G., Stenvinkel, P., & Rosado, A. (2022). Archaea from the gut microbiota of humans: Could be linked to chronic diseases? *Anaerobe*, 77, 102629. <https://doi.org/10.1016/j.anaerobe.2022.102629>
- Magnúsdóttir, S., Ravcheev, D., De Crécy-Lagard, V., & Thiele, I. (2015). Systematic genome assessment of B-vitamin biosynthesis suggests cooperation among gut microbes. *Frontiers in Genetics*, 6(MAR). <https://doi.org/10.3389/fgene.2015.00148>
- Malinen, E., Krogius-Kurikka, L., Lyra, A., Nikkilä, J., Jääskeläinen, A., Rinttilä, T., Vilpponen-Salmela, T., von Wright, A. J., & Palva, A. (2010). Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome. *World Journal of Gastroenterology*, 16(36), 4532–4540. <https://doi.org/10.3748/wjg.v16.i36.4532>
- Manefield, M., Griffiths, R. I., Leigh, M. B., Fisher, R., & Whiteley, A. S. (2005). Functional and compositional comparison of two activated sludge communities remediating coking effluent. *Environmental Microbiology*, 7(5), 715–722. <https://doi.org/10.1111/j.1462-2920.2004.00746.x>
- Manefield, M., Whiteley, A. S., Griffiths, R. I., & Bailey, M. J. (2002). RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Applied and Environmental Microbiology*, 68(11), 5367–5373. <https://doi.org/10.1128/AEM.68.11.5367-5373.2002>
- Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., German, J. B., Lebrilla, C. B., & Mills, D. A. (2010). Consumption of Human Milk Oligosaccharides by Gut-Related Microbes. *Journal of Agricultural and Food Chemistry*, 58(9), 5334–5340. <https://doi.org/10.1021/jf9044205>
- Marcobal, A., Barboza, M., Sonnenburg, E. D., Pudlo, N., Martens, E. C., Desai, P., Lebrilla, C. B., Weimer, B. C., Mills, D. A., German, J. B., & Sonnenburg, J. L. (2011). Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways. *Cell Host & Microbe*, 10(5), 507–514. <https://doi.org/10.1016/j.chom.2011.10.007>
- Maresca, M., Alatou, R., Pujol, A., Nicoletti, C., Perrier, J., Giardina, T., Simon, G., Méjean, V., & Fons, M. (2021). Rada, a mscramm adhesin of the dominant symbiote ruminococcus gnavus e1, binds human immunoglobulins and intestinal mucins. *Biomolecules*, 11(11). <https://doi.org/10.3390/biom11111613>

- Martens, E. C., Chiang, H. C., & Gordon, J. I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host and Microbe*, 4(5), 447–457. <https://doi.org/10.1016/j.chom.2008.09.007>
- Martin, R., Witte, K. L., & Wong, C. H. (1998). The synthesis and enzymatic incorporation of sialic acid derivatives for use as tools to study the structure, activity, and inhibition of glycoproteins and other glycoconjugates. *Bioorganic and Medicinal Chemistry*, 6(8), 1283–1292. [https://doi.org/10.1016/S0968-0896\(98\)00121-7](https://doi.org/10.1016/S0968-0896(98)00121-7)
- Martinez, C., Antolin, M., Santos, J., Torrejon, A., Casellas, F., Borruel, N., Guarner, F., & Malagelada, J. R. (2008). Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *American Journal of Gastroenterology*, 103(3), 643–648. <https://doi.org/10.1111/j.1572-0241.2007.01592.x>
- Mataigne, V., Vannier, N., Vandenkoornhuysen, P., & Hacquard, S. (2021). Microbial Systems Ecology to Understand Cross-Feeding in Microbiomes. *Frontiers in Microbiology*, 12(December). <https://doi.org/10.3389/fmicb.2021.780469>
- Mathewson, N. D., Jenq, R., Mathew, A. V., Koenigsnecht, M., Hanash, A., Toubai, T., Oravec-Wilson, K., Wu, S. R., Sun, Y., Rossi, C., Fujiwara, H., Byun, J., Shono, Y., Lindemans, C., Calafiore, M., Schmidt, T. C., Honda, K., Young, V. B., Pennathur, S., ... Reddy, P. (2016). Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. *Nature Immunology*, 17(5), 505–513. <https://doi.org/10.1038/ni.3400>
- McCoy, A. N., Araújo-Pérez, F., Azcárate-Peril, A., Yeh, J. J., Sandler, R. S., & Keku, T. O. (2013). *Fusobacterium* Is Associated with Colorectal Adenomas. *PLoS ONE*, 8(1). <https://doi.org/10.1371/journal.pone.0053653>
- McCoy, K. D., Burkhard, R., & Geuking, M. B. (2019). The microbiome and immune memory formation. *Immunology and Cell Biology*, 97(7), 625–635. <https://doi.org/10.1111/imcb.12273>
- McGovern, D. P. B., Jones, M. R., Taylor, K. D., Marcianti, K., Yan, X., Dubinsky, M., Ippoliti, A., Vasiliauskas, E., Berel, D., Derkowski, C., Dutridge, D., Fleshner, P., Shih, D. Q., Melmed, G., Mengesha, E., King, L., Pressman, S., Haritunians, T., Guo, X., ... Rotter, J. I. (2010). Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Human Molecular Genetics*, 19(17), 3468–3476. <https://doi.org/10.1093/hmg/ddq248>
- McNabney, S. M., & Henagan, T. M. (2017). Short chain fatty acids in the colon and peripheral tissues: A focus on butyrate, colon cancer, obesity and insulin resistance. *Nutrients*, 9(12), 1–28. <https://doi.org/10.3390/nu9121348>
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., Lewis, J. D., & Bushman, F. D. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research*, 21(10), 1616–1625. <https://doi.org/10.1101/gr.122705.111>
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>
- Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology*, 11(3), 266–277. <https://doi.org/10.1016/j.pbi.2008.03.006>
- Moor, K., Diard, M., Sellin, M. E., Felmy, B., Wotzka, S. Y., Toska, A., Bakkeren, E., Arnoldini, M., Bansept, F., Co, A. D., Völler, T., Minola, A., Fernandez-Rodriguez, B., Agatic, G., Barbieri, S.,

- Piccoli, L., Casiraghi, C., Corti, D., Lanzavecchia, A., ... Slack, E. (2017). High-avidity IgA protects the intestine by enchainning growing bacteria. *Nature*, *544*(7651), 498–502. <https://doi.org/10.1038/nature22058>
- Moore, W. E. C., & Holdeman, L. V. (1974). Human Fecal Flora: The Normal Flora of 20 Japanese-Hawaiians. *Applied Microbiology*, *27*(5), 961–979. <https://doi.org/10.1128/am.27.5.961-979.1974>
- Morrison, D. J., & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes*, *7*(3), 189–200. <https://doi.org/10.1080/19490976.2015.1134082>
- Muscariello, L., De Siena, B., & Marasco, R. (2020). Lactobacillus Cell Surface Proteins Involved in Interaction with Mucus and Extracellular Matrix Components. *Current Microbiology*, *77*(12), 3831–3841. <https://doi.org/10.1007/s00284-020-02243-5>
- Nakajima, A., Vogelzang, A., Maruya, M., Miyajima, M., Murata, M., Son, A., Kuwahara, T., Tsuruyama, T., Yamada, S., Matsuura, M., Nakase, H., Peterson, D. A., Fagarasan, S., & Suzuki, K. (2018). IgA regulates the composition and metabolic function of gut microbiota by promoting symbiosis between bacteria. *Journal of Experimental Medicine*, *215*(8), 2019–2034. <https://doi.org/10.1084/jem.20180427>
- Natividad, J. M., Marsaux, B., Rodenas, C. L. G., Rytz, A., Vandevijver, G., Marzorati, M., Van den Abbeele, P., Calatayud, M., & Rochat, F. (2022). Human Milk Oligosaccharides and Lactose Differentially Affect Infant Gut Microbiota and Intestinal Barrier In Vitro. *Nutrients*, *14*(12). <https://doi.org/10.3390/nu14122546>
- Navarro-Garcia, F., Gutierrez-Jimenez, J., Garcia-Tovar, C., Castro, L. A., Salazar-Gonzalez, H., & Cordova, V. (2010). Pic, an autotransporter protein secreted by different pathogens in the Enterobacteriaceae family, is a potent mucus secretagogue. *Infection and Immunity*, *78*(10), 4101–4109. <https://doi.org/10.1128/IAI.00523-10>
- Nayfach, S., Páez-Espino, D., Call, L., Low, S. J., Sberro, H., Ivanova, N. N., Proal, A. D., Fischbach, M. A., Bhatt, A. S., Hugenholtz, P., & Kyrpides, N. C. (2021). Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nature Microbiology*, *6*(7), 960–970. <https://doi.org/10.1038/s41564-021-00928-6>
- Ndeh, D., & Gilbert, H. J. (2018). Biochemistry of complex glycan depolymerisation by the human gut microbiota. *FEMS Microbiology Reviews*, *42*(2), 146–164. <https://doi.org/10.1093/femsre/fuy002>
- Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y., O'Neill, M. A., Urbanowicz, B. R., York, W. S., ... Gilbert, H. J. (2017). Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature*, *544*(7648), 65–70. <https://doi.org/10.1038/nature21725>
- Ng, K. M., Ferreyra, J. A., Higginbottom, S. K., Lynch, J. B., Kashyap, P. C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B. C., Monack, D. M., & Sonnenburg, J. L. (2013). Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature*, *502*(7469), 96–99. <https://doi.org/10.1038/nature12503>
- Niehaus, F., Peters, A., Groudieva, T., & Antranikian, G. (2000). Cloning, expression and biochemical characterisation of a unique thermostable pullulan-hydrolysing enzyme from the hyperthermophilic archaeon *Thermococcus aggregans*. *FEMS Microbiology Letters*, *190*(2), 223–229. [https://doi.org/10.1016/S0378-1097\(00\)00339-6](https://doi.org/10.1016/S0378-1097(00)00339-6)



- Nilsen, M., Saunders, C. M., Angell, I. L., Arntzen, M., Lødrup Carlsen, K. C., Carlsen, K. H., Haugen, G., Hagen, L. H., Carlsen, M. H., Hedlin, G., Jonassen, C. M., Nordlund, B., Rehbinder, E. M., Skjerven, H. O., Snipen, L., Staff, A. C., Vettukattil, R., & Rudi, K. (2020). Butyrate levels in the transition from an infant-to an adult-like gut microbiota correlate with bacterial networks associated with eubacterium rectale and ruminococcus gnavus. *Genes*, *11*(11), 1–15. <https://doi.org/10.3390/genes11111245>
- Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., Inatomi, O., Bamba, S., Sugimoto, M., Kawahara, M., Naito, Y., & Andoh, A. (2018). Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *Journal of Gastroenterology*, *53*(1), 95–106. <https://doi.org/10.1007/s00535-017-1384-4>
- Nishio, J., Negishi, H., Yasui-Kato, M., Miki, S., Miyanaga, K., Aoki, K., Mizusawa, T., Ueno, M., Aina, A., Muratani, M., Hangai, S., Yanai, H., Hasegawa, H., Ishii, Y., Tanji, Y., & Taniguchi, T. (2021). Identification and characterization of a novel Enterococcus bacteriophage with potential to ameliorate murine colitis. *Scientific Reports*, *11*(1), 1–13. <https://doi.org/10.1038/s41598-021-99602-4>
- Nishiyama, K., Nagai, A., Uribayashi, K., Yamamoto, Y., Mukai, T., & Okada, N. (2018). Two extracellular sialidases from Bifidobacterium bifidum promote the degradation of sialyl-oligosaccharides and support the growth of Bifidobacterium breve. *Anaerobe*, *52*, 22–28. <https://doi.org/10.1016/j.anaerobe.2018.05.007>
- Oh, J. H., Alexander, L. M., Pan, M., Schueler, K. L., Keller, M. P., Attie, A. D., Walter, J., & van Pijkeren, J. P. (2019). Dietary Fructose and Microbiota-Derived Short-Chain Fatty Acids Promote Bacteriophage Production in the Gut Symbiont Lactobacillus reuteri. *Cell Host and Microbe*, *25*(2), 273–284.e6. <https://doi.org/10.1016/j.chom.2018.11.016>
- Ostle, N., Whiteley, A. S., Bailey, M. J., Sleep, D., Ineson, P., & Manefield, M. (2003). Active microbial RNA turnover in a grassland soil estimated using a <sup>13</sup>C<sub>2</sub> spike. *Soil Biology and Biochemistry*, *35*(7), 877–885. [https://doi.org/10.1016/S0038-0717\(03\)00117-2](https://doi.org/10.1016/S0038-0717(03)00117-2)
- Ottman, N., Davids, M., Suarez-Diez, M., Boeren, S., Schaap, P. J., dos Santos, V. A. P. M., Smidt, H., Belzer, C., & de Vos, W. M. (2017). Genomescale model and omics analysis of metabolic capacities of Akkermansia muciniphila reveal a preferential mucin-degrading lifestyle. *Applied and Environmental Microbiology*, *83*(18). <https://doi.org/10.1128/AEM.01014-17>
- Ouwkerk, J. P., De Vos, W. M., & Belzer, C. (2013). Glycobiome: Bacteria and mucus at the epithelial interface. *Best Practice and Research: Clinical Gastroenterology*, *27*(1), 25–38. <https://doi.org/10.1016/j.bpg.2013.03.001>
- Ouwkerk, J. P., Tytgat, H. L. P., Elzinga, J., Koehorst, J., Van den Abbeele, P., Henrissat, B., Gueimonde, M., Cani, P. D., Van de Wiele, T., Belzer, C., & de Vos, W. M. (2022). Comparative Genomics and Physiology of Akkermansia muciniphila Isolates from Human Intestine Reveal Specialized Mucosal Adaptation. *Microorganisms*, *10*(8), 1–17. <https://doi.org/10.3390/microorganisms10081605>
- Owen, C. D., Tailford, L. E., Monaco, S., Šuligoj, T., Vaux, L., Lallement, R., Khedri, Z., Yu, H., Lecointe, K., Walshaw, J., Tribolo, S., Horrex, M., Bell, A., Chen, X., Taylor, G. L., Varki, A., Angulo, J., & Juge, N. (2017). Unravelling the specificity and mechanism of sialic acid recognition by the gut symbiont Ruminococcus gnavus. *Nature Communications*, *8*(1). <https://doi.org/10.1038/s41467-017-02109-8>
- Oyeleke, S. B., & Okusanmi, T. A. (2008). Isolation and characterization of cellulose hydrolysing microorganism from the rumen of ruminants. *African Journal of Biotechnology*, *7*(10), 1503–1504.

- Pacheco, A. R., Munera, D., Waldor, M. K., Sperandio, V., & Ritchie, J. M. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature*, *492*(7427), 113–117. <https://doi.org/10.1038/nature11623>
- Paone, P., & Cani, P. D. (2020). Mucus barrier, mucins and gut microbiota: The expected slimy partners? *Gut*, *69*(12), 2232–2243. <https://doi.org/10.1136/gutjnl-2020-322260>
- Parfrey, L. W., Walters, W. A., & Knight, R. (2011). Microbial eukaryotes in the human microbiome: Ecology, evolution, and future directions. *Frontiers in Microbiology*, *2*(JULY), 1–6. <https://doi.org/10.3389/fmicb.2011.00153>
- Pasakarnis, T. S., Gorski, C. A., O’Loughlin, E., Parkin, G. F., & Scherer, M. M. (2006). Abiotic reduction of chlorinated ethenes in the presence of anaerobic bacteria. *ACS, Division of Environmental Chemistry - Preprints of Extended Abstracts*, *46*(1), 957–961.
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- Pearce, O. M. T. (2018). Cancer glycan epitopes: Biosynthesis, structure and function. *Glycobiology*, *28*(9), 670–696. <https://doi.org/10.1093/glycob/cwy023>
- Pelaseyed, T., & Hansson, G. C. (2020). Membrane mucins of the intestine at a glance. *Journal of Cell Science*, *133*(5). <https://doi.org/10.1242/JCS.240929>
- Pelaseyed, T., Zäch, M., Petersson, Å. C., Svensson, F., Johansson, D. G. A., & Hansson, G. C. (2013). Unfolding dynamics of the mucin SEA domain probed by force spectroscopy suggest that it acts as a cell-protective device. *FEBS Journal*, *280*(6), 1491–1501. <https://doi.org/10.1111/febs.12144>
- Peng, C., Ouyang, Y., Lu, N., & Li, N. (2020). The NF-κB Signaling Pathway, the Microbiota, and Gastrointestinal Tumorigenesis: Recent Advances. In *Frontiers in Immunology* (Vol. 11). <https://doi.org/10.3389/fimmu.2020.01387>
- Pereira, F. C., Wasmund, K., Cobankovic, I., Jehmlich, N., Herbold, C. W., Lee, K. S., Sziranyi, B., Vesely, C., Decker, T., Stocker, R., Warth, B., von Bergen, M., Wagner, M., & Berry, D. (2020). Rational design of a microbial consortium of mucosal sugar utilizers reduces *Clostridiodes difficile* colonization. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-18928-1>
- Pérez, J. C. (2021). Fungi of the human gut microbiota: Roles and significance. *International Journal of Medical Microbiology*, *311*(3). <https://doi.org/10.1016/j.ijmm.2021.151490>
- Peter-Katalinić, J. (2005). *Methods in Enzymology: O-Glycosylation of Proteins* (pp. 139–171). [https://doi.org/10.1016/S0076-6879\(05\)05007-X](https://doi.org/10.1016/S0076-6879(05)05007-X)
- Peterson, L. W., & Artis, D. (2014). Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. *Nature Reviews Immunology*, *14*(3), 141–153. <https://doi.org/10.1038/nri3608>
- Petschacher, B., & Nidetzky, B. (2016). Biotechnological production of fucosylated human milk oligosaccharides: Prokaryotic fucosyltransferases and their use in biocatalytic cascades or whole cell conversion systems. *Journal of Biotechnology*, *235*, 61–83. <https://doi.org/10.1016/j.jbiotec.2016.03.052>
- Pham, V. T., Fehlbaum, S., Seifert, N., Richard, N., Bruins, M. J., Sybesma, W., Rehman, A., & Steinert, R. E. (2021). Effects of colon-targeted vitamins on the composition and metabolic

- activity of the human gut microbiome— a pilot study. *Gut Microbes*, 13(1), 1–20. <https://doi.org/10.1080/19490976.2021.1875774>
- Pichler, M. J., Yamada, C., Shuoker, B., Alvarez-Silva, C., Gotoh, A., Leth, M. L., Schoof, E., Katoh, T., Sakanaka, M., Katayama, T., Jin, C., Karlsson, N. G., Arumugam, M., Fushinobu, S., & Abou Hachem, M. (2020). Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-17075-x>
- Pittayanon, R., Lau, J. T., Yuan, Y., Leontiadis, G. I., Tse, F., Surette, M., & Moayyedi, P. (2019). Gut Microbiota in Patients With Irritable Bowel Syndrome—A Systematic Review. *Gastroenterology*, 157(1), 97–108. <https://doi.org/10.1053/j.gastro.2019.03.049>
- Pleguezuelos-Manzano, C., Puschhof, J., Huber, A. R., van Hoeck, A., Wood, H. M., Nomburg, J., Gurjao, C., Manders, F., Dalmaso, G., Stege, P. B., Paganelli, F. L., Geurts, M. H., Beumer, J., Mizutani, T., van der Linden, R., van Elst, S., Top, J., Willems, R. J. L., Giannakis, M., ... Clevers, H. (2020). Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. *Nature*. <https://doi.org/10.1038/s41586-020-2080-8>
- Pothuraju, R., Chaudhary, S., Rachagani, S., Kaur, S., Roy, H. K., Bouvet, M., & Batra, S. K. (2021). Mucins, gut microbiota, and postbiotics role in colorectal cancer. *Gut Microbes*, 13(1). <https://doi.org/10.1080/19490976.2021.1974795>
- Prasongsuk, S., Lotrakul, P., Ali, I., Bankeeree, W., & Punnapayak, H. (2018). The current status of *Aureobasidium pullulans* in biotechnology. *Folia Microbiologica*, 63(2), 129–140. <https://doi.org/10.1007/s12223-017-0561-4>
- Pruss, K. M., Marcobal, A., Southwick, A. M., Dahan, D., Smits, S. A., Ferreyra, J. A., Higginbottom, S. K., Sonnenburg, E. D., Kashyap, P. C., Choudhury, B., Bode, L., & Sonnenburg, J. L. (2021). Mucin-derived O-glycans supplemented to diet mitigate diverse microbiota perturbations. *ISME Journal*, 15(2), 577–591. <https://doi.org/10.1038/s41396-020-00798-6>
- Pudlo, N. A., Urs, K., Kumar, S. S., German, J. B., Mills, D. A., & Martens, E. C. (2015). Symbiotic human gut bacteria with variable metabolic priorities for host mucosal glycans. *MBio*, 6(6). <https://doi.org/10.1128/mBio.01282-15>
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., ... Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59–65. <https://doi.org/10.1038/nature08821>
- Qin, Y., Havulinna, A. S., Liu, Y., Jousilahti, P., Ritchie, S. C., Tokolyi, A., Sanders, J. G., Valsta, L., Brożyńska, M., Zhu, Q., Tripathi, A., Vázquez-Baeza, Y., Loomba, R., Cheng, S., Jain, M., Niiranen, T., Lahti, L., Knight, R., Salomaa, V., ... Méric, G. (2022). Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nature Genetics*, 54(2), 134–142. <https://doi.org/10.1038/s41588-021-00991-z>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), 590–596. <https://doi.org/10.1093/nar/gks1219>
- Raba, G., & Luis, A. S. (2023). Mucin utilization by gut microbiota: recent advances on characterization of key enzymes. *Essays in Biochemistry*, 0(April), 345–353. <https://doi.org/10.1042/ebc20220121>
- Radajewski, S., Ineson, P., Parekh, N. R., & Murrell, J. C. (2000). Stable-isotope probing as a tool in

- microbial ecology. *Nature*, 403(6770), 646–649. <https://doi.org/10.1038/35001054>
- Rahfeld, P., Wardman, J. F., Mehr, K., Huff, D., Morgan-Lang, C., Chen, H. M., Hallam, S. J., & Withers, S. G. (2019). Prospecting for microbial  $\alpha$ -N-acetylgalactosaminidases yields a new class of GH31 O-glycanase. In *Journal of Biological Chemistry* (Vol. 294, Issue 44, pp. 16400–16415). <https://doi.org/10.1074/jbc.RA119.010628>
- Ramare, F., Nicoli, J., Dabard, J., Corring, T., Ladire, M., Gueugneau, A. M., & Raibaud, P. (1993). Trypsin-dependent production of an antibacterial substance by a human Peptostreptococcus strain in gnotobiotic rats and in vitro. *Applied and Environmental Microbiology*, 59(9), 2876–2883. <https://doi.org/10.1128/aem.59.9.2876-2883.1993>
- Ravn, V., & Dabelsteen, E. (2000). Tissue distribution of histo-blood group antigens. Review article. *APMIS*, 108(1), 1–28. <https://doi.org/10.1034/j.1600-0463.2000.d01-1.x>
- Rea, M. C., Sit, C. S., Clayton, E., O'Connor, P. M., Whittal, R. M., Zheng, J., Vederas, J. C., Ross, R. P., & Hill, C. (2010). Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. *Proceedings of the National Academy of Sciences of the United States of America*, 107(20), 9352–9357. <https://doi.org/10.1073/pnas.0913554107>
- Reed, K. K., & Wickham, R. (2009). Review of the Gastrointestinal Tract: From Macro to Micro. *Seminars in Oncology Nursing*, 25(1), 3–14. <https://doi.org/10.1016/j.soncn.2008.10.002>
- Rho, J. H., Wright, D. P., Christie, D. L., Clinch, K., Furneaux, R. H., & Robertson, A. M. (2005). A novel mechanism for desulfation of mucin: Identification and cloning of a mucin-desulfating glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. *Journal of Bacteriology*, 187(5), 1543–1551. <https://doi.org/10.1128/JB.187.5.1543-1551.2005>
- Robbe, C., Capon, C., Coddeville, B., & Michalski, J. C. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochemical Journal*, 384(2), 307–316. <https://doi.org/10.1042/BJ20040605>
- Robbe, C., Capon, C., Maes, E., Rousset, M., Zweibaum, A., Zanetta, J. P., & Michalski, J. C. (2003). Evidence of regio-specific glycosylation in human intestinal mucins: Presence of an acidic gradient along the intestinal tract. *Journal of Biological Chemistry*, 278(47), 46337–46348. <https://doi.org/10.1074/jbc.M302529200>
- Rodriguez, C. I., & Martiny, J. B. H. (2020). Evolutionary relationships among bifidobacteria and their hosts and environments. *BMC Genomics*, 21(1), 1–12. <https://doi.org/10.1186/s12864-019-6435-1>
- Roediger, W. E. W. (1980). Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. *Gut*, 21(9), 793–798. <https://doi.org/10.1136/gut.21.9.793>
- Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., Baslé, A., Morland, C., Day, A. M., Zheng, H., Rogers, T. E., Thompson, P., Hawkins, A. R., Yadav, M. P., Henrissat, B., Martens, E. C., Dupree, P., Gilbert, H. J., & Bolam, D. N. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. *Nature Communications*, 6(May). <https://doi.org/10.1038/ncomms8481>
- Roos, S., & Jonsson, H. (2002). A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. *Microbiology*, 148(2), 433–442. <https://doi.org/10.1099/00221287-148-2-433>
- Rubinstein, M. R., Wang, X., Liu, W., Hao, Y., Cai, G., & Han, Y. W. (2013). *Fusobacterium nucleatum* Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/ $\beta$ -Catenin

- Signaling via its FadA Adhesin. In *Cell Host and Microbe* (Vol. 14, Issue 2, pp. 195–206). <https://doi.org/10.1016/j.chom.2013.07.012>
- Rutayisire, E., Huang, K., Liu, Y., & Tao, F. (2016). The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: A systematic review. *BMC Gastroenterology*, *16*(1), 1–12. <https://doi.org/10.1186/s12876-016-0498-0>
- Ruvoën-Clouet, N., Belliot, G., & Le Pendu, J. (2013). Noroviruses and histo-blood groups: The impact of common host genetic polymorphisms on virus transmission and evolution. *Reviews in Medical Virology*, *23*(6), 355–366. <https://doi.org/10.1002/rmv.1757>
- Saggheddu, V., Patrone, V., Miragoli, F., Puglisi, E., & Morelli, L. (2016). Infant early gut colonization by Lachnospiraceae: High frequency of Ruminococcus gnavus. *Frontiers in Pediatrics*, *4*(JUN), 7–12. <https://doi.org/10.3389/fped.2016.00057>
- Saile, N., Schwarz, L., Eißenger, K., Klumpp, J., Fricke, F. W., & Schmidt, H. (2018). Growth advantage of Escherichia coli O104:H4 strains on 5-N-acetyl-9-O-acetyl neuraminic acid as a carbon source is dependent on heterogeneous phage-Borne nanS-p esterases. *International Journal of Medical Microbiology*, *308*(4), 459–468. <https://doi.org/10.1016/j.ijmm.2018.03.006>
- Sakurama, H., Tsutsumi, E., Ashida, H., Katayama, T., Yamamoto, K., & Kumagai, H. (2012). Differences in the substrate specificities and active-site structures of two  $\alpha$ -L-fucosidases (glycoside hydrolase family 29) from Bacteroides thetaiotaomicron. *Bioscience, Biotechnology and Biochemistry*, *76*(5), 1022–1024. <https://doi.org/10.1271/bbb.111004>
- Sandle, G. I. (1998). Salt and water absorption in the human colon: A modern appraisal. *Gut*, *43*(2), 294–299. <https://doi.org/10.1136/gut.43.2.294>
- Santaolalla, R., & Abreu, M. T. (2012). Innate immunity in the small intestine. *Current Opinion in Gastroenterology*, *28*(2), 124–129. <https://doi.org/10.1097/MOG.0b013e3283506559>
- Schwarz, A., Adetutu, E. M., Juhasz, A. L., Aburto-Medina, A., Ball, A. S., & Shahsavari, E. (2018). Microbial Degradation of Phenanthrene in Pristine and Contaminated Sandy Soils. *Microbial Ecology*, *75*(4), 888–902. <https://doi.org/10.1007/s00248-017-1094-8>
- Sekar, K., Linker, S. M., Nguyen, J., Grünhagen, A., Stocker, R., & Sauer, U. (2020). Bacterial Glycogen Provides Short-Term Benefits in Changing Environments. *Applied and Environmental Microbiology*, *86*(9), 1–12. <https://doi.org/10.1128/AEM.00049-20>
- Sekirov, I., Tam, N. M., Jogova, M., Robertson, M. L., Li, Y., Lupp, C., & Finlay, B. B. (2008). Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infection and Immunity*, *76*(10), 4726–4736. <https://doi.org/10.1128/IAI.00319-08>
- Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B., German, J. B., Price, N. P., Richardson, P. M., & Mills, D. A. (2008). The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(48), 18964–18969. <https://doi.org/10.1073/pnas.0809584105>
- Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), 1–14. <https://doi.org/10.1371/journal.pbio.1002533>
- Severi, E., Rudden, M., Bell, A., Palmer, T., Juge, N., & Thomas, G. H. (2021). Multiple evolutionary origins reflect the importance of sialic acid transporters in the colonization potential of

- bacterial pathogens and commensals. *Microbial Genomics*, 7(6).  
<https://doi.org/10.1099/MGEN.0.000614>
- Shajahan, A., Supekar, N., Heiss, C., & Azadi, P. (2019). High-Throughput Automated Micro-permethylation for Glycan Structure Analysis. *Analytical Chemistry*, 91(2), 1237–1240.  
<https://doi.org/10.1021/acs.analchem.8b05146>
- Shang, F. M., & Liu, H. L. (2018). *Fusobacterium nucleatum* and colorectal cancer: A review. *World Journal of Gastrointestinal Oncology*, 10(3), 71–81. <https://doi.org/10.4251/wjgo.v10.i3.71>
- Shim, J. H., Park, J. T., Hong, J. S., Kim, K. W., Kim, M. J., Auh, J. H., Kim, Y. W., Park, C. S., Boos, W., Kim, J. W., & Park, K. H. (2009). Role of maltogenic amylase and pullulanase in maltodextrin and glycogen metabolism of *Bacillus subtilis* 168. *Journal of Bacteriology*, 191(15), 4835–4844. <https://doi.org/10.1128/JB.00176-09>
- Shimada, Y., Watanabe, Y., Wakinaka, T., Funeno, Y., Kubota, M., Chaiwangsri, T., Kurihara, S., Yamamoto, K., Katayama, T., & Ashida, H. (2015).  $\alpha$ -N-Acetylglucosaminidase from *Bifidobacterium bifidum* specifically hydrolyzes  $\alpha$ -linked N-acetylglucosamine at nonreducing terminus of O-glycan on gastric mucin. *Applied Microbiology and Biotechnology*, 99(9), 3941–3948. <https://doi.org/10.1007/s00253-014-6201-x>
- Shuoker, B., Pichler, M. J., Jin, C., Sakanaka, H., Wu, H., Gascueña, A. M., Liu, J., Nielsen, T. S., Holgersson, J., Nordberg Karlsson, E., Juge, N., Meier, S., Morth, J. P., Karlsson, N. G., & Abou Hachem, M. (2023). Sialidases and fucosidases of *Akkermansia muciniphila* are crucial for growth on mucin and nutrient sharing with mucus-associated gut bacteria. *Nature Communications*, 14(1), 1833. <https://doi.org/10.1038/s41467-023-37533-6>
- Shurer, C. R., Kuo, J. C. H., Roberts, L. D. M., Gandhi, J. G., Colville, M. J., Enoki, T. A., Pan, H., Su, J., Noble, J. M., Hollander, M. J., O'Donnell, J. P., Yin, R., Pedram, K., Möckl, L., Kourkoutis, L. F., Moerner, W. E., Bertozzi, C. R., Feigenson, G. W., Reesink, H. L., & Paszek, M. J. (2019). Physical Principles of Membrane Shape Regulation by the Glycocalyx. *Cell*, 177(7), 1757–1770.e21. <https://doi.org/10.1016/j.cell.2019.04.017>
- Slomiany, A., Zdebska, E., & Slomiany, B. L. (1984). Structures of the neutral oligosaccharides isolated from A-active human gastric mucin. *Journal of Biological Chemistry*, 259(23), 14743–14749. [https://doi.org/10.1016/s0021-9258\(17\)42665-2](https://doi.org/10.1016/s0021-9258(17)42665-2)
- Smith, K. A., & Salyers, A. A. (1991). Characterization of a neopullulanase and an  $\alpha$ -glucosidase from *Bacteroides thetaiotaomicron* 95-1. *Journal of Bacteriology*, 173(9), 2962–2968. <https://doi.org/10.1128/jb.173.9.2962-2968.1991>
- Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J. J., Blugeon, S., Bridonneau, C., Furet, J. P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottière, H. M., Doré, J., Marteau, P., Seksik, P., & Langella, P. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43), 16731–16736. <https://doi.org/10.1073/pnas.0804812105>
- Sokol, H., Seksik, P., Rigottier-Gois, L., Lay, C., Lepage, P., Podglajen, I., Marteau, P., & Doré, J. (2006). Specificities of the fecal microbiota in inflammatory bowel disease. *Inflammatory Bowel Diseases*, 12(2), 106–111. <https://doi.org/10.1097/01.MIB.0000200323.38139.c6>
- Sokolovskaya, O. M., Tan, M. W., & Wolan, D. W. (2022). Sialic acid diversity in the human gut: Molecular impacts and tools for future discovery. *Current Opinion in Structural Biology*, 75, 102397. <https://doi.org/10.1016/j.sbi.2022.102397>

- Solís, G., de los Reyes-Gavilan, C. G., Fernández, N., Margolles, A., & Gueimonde, M. (2010). Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe*, *16*(3), 307–310. <https://doi.org/10.1016/j.anaerobe.2010.02.004>
- Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N., & Sonnenburg, J. L. (2010). Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations. *Cell*, *141*(7), 1241–1252. <https://doi.org/10.1016/j.cell.2010.05.005>
- Spreinat, A., & Antranikian, G. (1990). Applied Microbiology Biotechnology Purification and properties of a thermostable pullulanase from Clostridium thermosulfurogenes EM1 which hydrolyses both  $\alpha$ -1,6 and  $\alpha$ -1,4-glycosidic linkages. *Appl Microbiol Biotechnol*, *33*, 511–518.
- Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., & Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of  $\alpha$ -amylase-related proteins. In *Protein Engineering, Design and Selection* (Vol. 19, Issue 12, pp. 555–562). <https://doi.org/10.1093/protein/gzl044>
- Stanley, P., Wuhrer, M., Lauc, G., Stowell, S. R., & Cummings, R. D. (2022). Structures Common to Different Glycans. In A. Varki, J. D. E. Richard D. Cummings, & P. H. S. Pamela Stanley, Gerald W. Hart, Markus Aebi, Debra Mohnen, Taroh Kinoshita, Nicolle H. Packer, James H. Prestegard, Ronald L. Schnaar (Eds.), *Essentials of Glycobiology. 4th edition*. (4th Editio). Laboratory Press; 2022. <https://www.ncbi.nlm.nih.gov/books/NBK579929/>
- Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., & Kleindienst, S. (2020). Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline. *Frontiers in Microbiology*, *11*(October), 1–18. <https://doi.org/10.3389/fmicb.2020.550420>
- Strugala, V., Dettmar, P. W., & Pearson, J. P. (2008). Thickness and continuity of the adherent colonic mucus barrier in active and quiescent ulcerative colitis and Crohn's disease. *International Journal of Clinical Practice*, *62*(5), 762–769. <https://doi.org/10.1111/j.1742-1241.2007.01665.x>
- Šuligoj, T., Vignæs, L. K., Van den Abbeele, P., Apostolou, A., Karalis, K., Savva, G. M., McConnell, B., & Juge, N. (2020). Effects of human milk oligosaccharides on the adult gut microbiota and barrier function. *Nutrients*, *12*(9), 1–21. <https://doi.org/10.3390/nu12092808>
- Sundin, O. H., Mendoza-Ladd, A., Zeng, M., Diaz-Arévalo, D., Morales, E., Fagan, B. M., Ordoñez, J., Velez, P., Antony, N., & McCallum, R. W. (2017). The human jejunum has an endogenous microbiota that differs from those in the oral cavity and colon. *BMC Microbiology*, *17*(1), 1–17. <https://doi.org/10.1186/s12866-017-1059-6>
- Surana, N. K., & Kasper, D. L. (2017). Moving beyond microbiome-wide associations to causal microbe identification. *Nature*, *552*(7684), 244–247. <https://doi.org/10.1038/nature25019>
- Sutherland, D. B., Suzuki, K., & Fagarasan, S. (2016). Fostering of advanced mutualism with gut microbiota by Immunoglobulin A. *Immunological Reviews*, *270*(1), 20–31. <https://doi.org/10.1111/imr.12384>
- Suzuki, T., Yoshida, S., & Hara, H. (2008). Physiological concentrations of short-chain fatty acids immediately suppress colonic epithelial permeability. *British Journal of Nutrition*, *100*(2), 297–305. <https://doi.org/10.1017/S0007114508888733>
- Svihus, B., & Hervik, A. K. (2016). Digestion and metabolic fates of starch, and its relation to major

- nutrition-related health problems: A review. *Starch/Staerke*, 68(3–4), 302–313. <https://doi.org/10.1002/star.201500295>
- Tailford, L. E., Crost, E. H., Kavanaugh, D., & Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. *Frontiers in Genetics*, 5(FEB). <https://doi.org/10.3389/fgene.2015.00081>
- Tailford, L. E., Owen, C. D., Walshaw, J., Crost, E. H., Hardy-Goddard, J., Le Gall, G., De Vos, W. M., Taylor, G. L., & Juge, N. (2015). Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nature Communications*, 6(May), 1–12. <https://doi.org/10.1038/ncomms8624>
- Tannock, G. W., Lawley, B., Munro, K., Sims, I. M., Lee, J., Butts, C. A., & Roy, N. (2014). RNA-stable-isotope probing shows utilization of carbon from inulin by specific bacterial populations in the rat large bowel. *Applied and Environmental Microbiology*, 80(7), 2240–2247. <https://doi.org/10.1128/AEM.03799-13>
- Tasteyre, A., Barc, M., Collignon, A., Boureau, H., & Karjalainen, T. (2001). Role of FliC and FliD Flagellar Proteins of *Clostridium difficile* in Adherence and Gut Colonization. *Infection and Immunity*, 69(12), 7937–7940. <https://doi.org/10.1128/IAI.69.12.7937-7940.2001>
- Telatin, A., Fariselli, P., & Birolo, G. (2021). Seqfu: A suite of utilities for the robust and reproducible manipulation of sequence files. *Bioengineering*, 8(5), 1–7. <https://doi.org/10.3390/bioengineering8050059>
- Teze, D., Shuoker, B., Chaberski, E. K., Kunstmann, S., Fredslund, F., Nielsen, T. S., Stender, E. G. P., Peters, G. H. J., Karlsson, E. N., Welner, D. H., & Hachem, M. A. (2020). The Catalytic Acid–Base in GH109 Resides in a Conserved GGHGG Loop and Allows for Comparable  $\alpha$ -Retaining and  $\beta$ -Inverting Activity in an N -Acetylgalactosaminidase from *Akkermansia muciniphila*. *ACS Catalysis*, 10(6), 3809–3819. <https://doi.org/10.1021/acscatal.9b04474>
- Thurl, S., Henker, J., Siegel, M., Tovar, K., & Sawatzki, G. (1997). Detection of four human milk groups with respect to Lewis blood group dependent oligosaccharides. *Glycoconjugate Journal*, 14(7), 795–799. <https://doi.org/10.1023/a:1018529703106>
- Tian, S., & Sun, Y. (2020). Influencing factor of resistant starch formation and application in cereal products: A review. *International Journal of Biological Macromolecules*, 149, 424–431. <https://doi.org/10.1016/j.ijbiomac.2020.01.264>
- Tong, M., McHardy, I., Ruegger, P., Goudarzi, M., Kashyap, P. C., Haritunians, T., Li, X., Graeber, T. G., Schwager, E., Huttenhower, C., Fornace, A. J., Sonnenburg, J. L., McGovern, D. P. B., Borneman, J., & Braun, J. (2014). Reprograming of gut microbiome energy metabolism by the FUT2 Crohn’s disease risk polymorphism. *ISME Journal*, 8(11), 2193–2206. <https://doi.org/10.1038/ismej.2014.64>
- Toor, K. J., Ahmad, N., Muhammad, M. A., & Rashid, N. (2020). TK-PUL, a pullulan hydrolase type III from *Thermococcus kodakarensis*, a potential candidate for simultaneous liquefaction and saccharification of starch. *Amylase*, 4(1), 45–55. <https://doi.org/10.1515/amylase-2020-0004>
- Tourlousse, D. M., Sakamoto, M., Miura, T., Narita, K., Ohashi, A., Uchino, Y., Yamazoe, A., Kameyama, K., Terauchi, J., Ohkuma, M., Kawasaki, H., & Sekiguchi, Y. (2020). Complete Genome Sequence of *Collinsella aerofaciens* JCM 10188 T. *Microbiology Resource Announcements*, 9(16), 10–11. <https://doi.org/10.1128/mra.00134-20>
- Tramontano, M., Andrejev, S., Pruteanu, M., Klünemann, M., Kuhn, M., Galardini, M., Jouhten, P., Zelezniak, A., Zeller, G., Bork, P., Typas, A., & Patil, K. R. (2018). Nutritional preferences of



- human gut bacteria reveal their metabolic idiosyncrasies. *Nature Microbiology*, 3(4), 514–522. <https://doi.org/10.1038/s41564-018-0123-9>
- Troci, A., Rausch, P., Waschina, S., Lieb, W., Franke, A., & Bang, C. (2022). Long-Term Dietary Effects on Human Gut Microbiota Composition Employing Shotgun Metagenomics Data Analysis. *Molecular Nutrition and Food Research*, 2101098, 1–12. <https://doi.org/10.1002/mnfr.202101098>
- Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine*, 1(6), 6ra14-6ra14. <https://doi.org/10.1126/scitranslmed.3000322>
- Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J. H., Zomer, A., Sánchez, B., Bidossi, A., Ferrarini, A., Giubellini, V., Delledonne, M., Henrissat, B., Coutinho, P., Oggioni, M., Fitzgerald, G. F., Mills, D., Margolles, A., Kelly, D., Van Sinderen, D., & Ventura, M. (2010). Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proceedings of the National Academy of Sciences of the United States of America*, 107(45), 19514–19519. <https://doi.org/10.1073/pnas.1011100107>
- Turroni, F., Duranti, S., Milani, C., Lugli, G. A., van Sinderen, D., & Ventura, M. (2019). *Bifidobacterium bifidum*: A key member of the early human gut microbiota. *Microorganisms*, 7(11), 1–13. <https://doi.org/10.3390/microorganisms7110544>
- Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D., Fuligni, F., Gueimonde, M., Margolles, A., de Bellis, G., O'Toole, P. W., van Sinderen, D., Marchesi, J. R., & Ventura, M. (2012). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE*, 7(5), 20–24. <https://doi.org/10.1371/journal.pone.0036957>
- van den Elsen, L. W. J., Garssen, J., Burcelin, R., & Verhasselt, V. (2019). Shaping the gut microbiota by breastfeeding: The gateway to allergy prevention? *Frontiers in Pediatrics*, 7(FEB), 1–10. <https://doi.org/10.3389/fped.2019.00047>
- van der Flier, L. G., & Clevers, H. (2009). Stem Cells, Self-Renewal, and Differentiation in the Intestinal Epithelium. *Annual Review of Physiology*, 71(1), 241–260. <https://doi.org/10.1146/annurev.physiol.010908.163145>
- Van Der Lugt, B., Van Beek, A. A., Aalvink, S., Meijer, B., Sovran, B., Vermeij, W. P., Brandt, R. M. C., De Vos, W. M., Savelkoul, H. F. J., Steegenga, W. T., & Belzer, C. (2019). *Akkermansia muciniphila* ameliorates the age-related decline in colonic mucus thickness and attenuates immune activation in accelerated aging Ercc1 - $\Delta$ 7 mice. *Immunity and Ageing*, 16(1), 1–17. <https://doi.org/10.1186/s12979-019-0145-z>
- Van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E. G., De Vos, W. M., Visser, C. E., Kuijper, E. J., Bartelsman, J. F. W. M., Tijssen, J. G. P., Speelman, P., Dijkgraaf, M. G. W., & Keller, J. J. (2013). Duodenal infusion of donor feces for recurrent *clostridium difficile*. *New England Journal of Medicine*, 368(5), 407–415. <https://doi.org/10.1056/NEJMoa1205037>
- Vasapolli, R., Schütte, K., Schulz, C., Vital, M., Schomburg, D., Pieper, D. H., Vilchez-Vargas, R., & Malforteiner, P. (2019). Analysis of Transcriptionally Active Bacteria Throughout the Gastrointestinal Tract of Healthy Individuals. *Gastroenterology*, 157(4), 1081-1092.e3. <https://doi.org/10.1053/j.gastro.2019.05.068>
- von Ossowski, I., Satokari, R., Reunanen, J., Lebeer, S., De Keersmaecker, S. C. J., Vanderleyden, J., de Vos, W. M., & Palva, A. (2011). Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG. *Applied and Environmental*

*Microbiology*, 77(13), 4465–4472. <https://doi.org/10.1128/AEM.02497-10>

- Vuik, F. E. R., Dicksved, J., Lam, S. Y., Fuhler, G. M., van der Laan, L. J. W., van de Winkel, A., Konstantinov, S. R., Spaander, M. C. W., Peppelenbosch, M. P., Engstrand, L., & Kuipers, E. J. (2019). Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. *United European Gastroenterology Journal*, 7(7), 897–907. <https://doi.org/10.1177/2050640619852255>
- Wakinaka, T., Kiyohara, M., Kurihara, S., Hirata, A., Chaiwangsri, T., Ohnuma, T., Fukamizo, T., Katayama, T., Ashida, H., & Yamamoto, K. (2013). Bifidobacterial  $\alpha$ -galactosidase with unique carbohydrate-binding module specifically acts on blood group B antigen. *Glycobiology*, 23(2), 232–240. <https://doi.org/10.1093/glycob/cws142>
- Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G., & Petrovska, L. (2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiology*, 11. <https://doi.org/10.1186/1471-2180-11-7>
- Walsh, C., Lane, J. A., van Sinderen, D., & Hickey, R. M. (2020). Human milk oligosaccharides: Shaping the infant gut microbiota and supporting health. *Journal of Functional Foods*, 72(June), 104074. <https://doi.org/10.1016/j.jff.2020.104074>
- Wan, L., Ge, W. R., Zhang, S., Sun, Y. L., Wang, B., & Yang, G. (2020). Case-Control Study of the Effects of Gut Microbiota Composition on Neurotransmitter Metabolic Pathways in Children With Attention Deficit Hyperactivity Disorder. *Frontiers in Neuroscience*, 14(February), 1–9. <https://doi.org/10.3389/fnins.2020.00127>
- Wang, B. X., Wu, C. M., & Ribbeck, K. (2021). Home, sweet home: how mucus accommodates our microbiota. *FEBS Journal*, 288(6), 1789–1799. <https://doi.org/10.1111/febs.15504>
- Wang, L., & Wise, M. J. (2011). Glycogen with short average chain length enhances bacterial durability. *Naturwissenschaften*, 98(9), 719–729. <https://doi.org/10.1007/s00114-011-0832-x>
- Wang, M., Zhang, X.-Y., Guo, R.-R., Cai, Z.-P., Hu, X.-C., Chen, H., Wei, S., Voglmeir, J., & Liu, L. (2018). Cloning, purification and biochemical characterization of two  $\beta$ -N-acetylhexosaminidases from the mucin-degrading gut bacterium *Akkermansia muciniphila*. *Carbohydrate Research*, 457, 1–7. <https://doi.org/10.1016/j.carres.2017.12.007>
- Ward, R. E., Niñonuevo, M., Mills, D. A., Lebrilla, C. B., & German, J. B. (2006). In vitro fermentation of breast milk oligosaccharides by *Bifidobacterium infantis* and *Lactobacillus gasseri*. *Applied and Environmental Microbiology*, 72(6), 4497–4499. <https://doi.org/10.1128/AEM.02515-05>
- Weis, S., Schnell, S., & Egert, M. (2019). RNA-Based Stable Isotope Probing (RNA-SIP) in the Gut Environment. In *Stable Isotope Probing* (pp. 221–231).
- Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, 173(2), 697–703. <https://doi.org/10.1128/jb.173.2.697-703.1991>
- Wilson, A. S., Koller, K. R., Ramaboli, M. C., Nesengani, L. T., Ocvirk, S., Chen, C., Flanagan, C. A., Sapp, F. R., Merritt, Z. T., Bhatti, F., Thomas, T. K., & O’Keefe, S. J. D. (2020). Diet and the Human Gut Microbiome: An International Review. *Digestive Diseases and Sciences*, 65(3), 723–740. <https://doi.org/10.1007/s10620-020-06112-w>

- Wright, E. S. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. *R Journal*, 8(1), 352–359. <https://doi.org/10.32614/rj-2016-025>
- Wu, H., Crost, E. H., Owen, C. D., van Bakel, W., Martínez Gascueña, A., Latousakis, D., Hicks, T., Walpole, S., Urbanowicz, P. A., Ndeh, D., Monaco, S., Sánchez Salom, L., Griffiths, R., Reynolds, R. S., Colvile, A., Spencer, D. I. R., Walsh, M., Angulo, J., & Juge, N. (2021). The human gut symbiont *Ruminococcus gnavus* shows specificity to blood group A antigen during mucin glycan foraging: Implication for niche colonisation in the gastrointestinal tract. *PLOS Biology*, 19(12), e3001498. <https://doi.org/10.1371/journal.pbio.3001498>
- Wu, H., Rebello, O., Crost, E. H., Owen, C. D., Walpole, S., Bennati-Granier, C., Ndeh, D., Monaco, S., Hicks, T., Colvile, A., Urbanowicz, P. A., Walsh, M. A., Angulo, J., Spencer, D. I. R., & Juge, N. (2021). Fucosidases from the human gut symbiont *Ruminococcus gnavus*. *Cellular and Molecular Life Sciences*, 78(2), 675–693. <https://doi.org/10.1007/s00018-020-03514-x>
- Wu, S., Arthur, C. M., Jan, H., Garcia-Beltran, W. F., Patel, K. R., Rathgeber, M. F., Verkerke, H., Cheedarla, N., Jajosky, R. P., Paul, A., Neish, A. S., Roback, J. D., Josephson, C. D., Wesemann, D. R., Kalman, D., Rakoff-Nahoum, S., Cummings, R. D., & Stowell, S. R. (2022). Blood Group A Enhances SARS-CoV-2 Infection. *Blood*, 140(Supplement 1), 79–80. <https://doi.org/10.1182/blood-2022-170083>
- Wu, S. C., Arthur, C. M., Wang, J., Verkerke, H., Josephson, C. D., Kalman, D., Roback, J. D., Cummings, R. D., & Stowell, S. R. (2021). The SARS-CoV-2 receptor-binding domain preferentially recognizes blood group A. *Blood Advances*, 5(5), 1305–1309. <https://doi.org/10.1182/bloodadvances.2020003259>
- Xiao, A., Slack, T. J., Li, Y., Shi, D., Yu, H., Li, W., Liu, Y., & Chen, X. (2018). Streptococcus pneumoniae Sialidase SpNanB-Catalyzed One-Pot Multienzyme (OPME) Synthesis of 2,7-Anhydro-Sialic Acids as Selective Sialidase Inhibitors. *Journal of Organic Chemistry*, 83(18), 10798–10804. <https://doi.org/10.1021/acs.joc.8b01519>
- Xu, G., Potter, J. A., Russell, R. J. M., Oggioni, M. R., Andrew, P. W., & Taylor, G. L. (2008). Crystal Structure of the NanB Sialidase from *Streptococcus pneumoniae*. *Journal of Molecular Biology*, 384(2), 436–449. <https://doi.org/10.1016/j.jmb.2008.09.032>
- Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., & Gordon, J. I. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science*, 299(5615), 2074–2076. <https://doi.org/10.1126/science.1080029>
- Yan, J., Ding, J., Jin, G., Duan, Z., Yang, F., Li, D., Zhou, H., Li, M., Guo, Z., Chai, W., & Liang, X. (2019). Profiling of human milk oligosaccharides for lewis epitopes and secretor status by electrostatic repulsion hydrophilic interaction chromatography coupled with negative-ion electrospray tandem mass spectrometry. *Analytical Chemistry*, 91(13), 8199–8206. <https://doi.org/10.1021/acs.analchem.9b00687>
- Yatsunenkov, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., ... Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, 486(7402), 222–227. <https://doi.org/10.1038/nature11053>
- Yin, H., Miao, Z., Wang, L., Su, B., Liu, C., Jin, Y., Wu, B., Han, H., & Yuan, X. (2022). *Fusobacterium nucleatum* Promotes Liver Metastasis in Colorectal Cancer by Regulating the Hepatic Immune Niche and Altering Gut Microbiota. *Aging*, 14(4), 1941–1958. <https://doi.org/10.18632/aging.203914>

- Yoshida, E., Sakurama, H., Kiyohara, M., Nakajima, M., Kitaoka, M., Ashida, H., Hirose, J., Katayama, T., Yamamoto, K., & Kumagai, H. (2012). Bifidobacterium longum subsp. infantis uses two different  $\beta$ -galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology*, *22*(3), 361–368. <https://doi.org/10.1093/glycob/cwr116>
- Young, W., Egert, M., Bassett, S. A., & Bibiloni, R. (2015). Detection of sialic acid-utilising bacteria in a caecal community batch culture using RNA-based stable isotope probing. *Nutrients*, *7*(4), 2109–2124. <https://doi.org/10.3390/nu7042109>
- Yu, Du, Z., Sun, X., Shi, C., Zhang, H., & Hu, T. (2015). Aberrant Cosmc genes result in Tn antigen expression in human colorectal carcinoma cell line HT-29. *International Journal of Clinical and Experimental Pathology*, *8*(3), 2590–2602.
- Yu, H., Chokhawala, H. A., Huang, S., & Chen, X. (2006). One-pot three-enzyme chemoenzymatic approach to the synthesis of sialosides containing natural and non-natural functionalities. *Nature Protocols*, *1*(5), 2485–2492. <https://doi.org/10.1038/nprot.2006.401>
- Yu, H., Chokhawala, H., Karpel, R., Yu, H., Wu, B., Zhang, J., Zhang, Y., Jia, Q., & Chen, X. (2005). A multifunctional Pasteurella multocida sialyltransferase: A powerful tool for the synthesis of sialoside libraries. *Journal of the American Chemical Society*, *127*(50), 17618–17619. <https://doi.org/10.1021/ja0561690>
- Yu, H., Yu, H., Karpel, R., & Chen, X. (2004). Chemoenzymatic synthesis of CMP-sialic acid derivatives by a one-pot two-enzyme system: Comparison of substrate flexibility of three microbial CMP-sialic acid synthetases. *Bioorganic and Medicinal Chemistry*, *12*(24), 6427–6435. <https://doi.org/10.1016/j.bmc.2004.09.030>
- Yu, T. C., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., Qian, Y., Kryczek, I., Sun, D., Nagarsheth, N., Chen, Y., Chen, H., Hong, J., Zou, W., & Fang, J. Y. (2017). Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell*, *170*(3), 548-563.e16. <https://doi.org/10.1016/j.cell.2017.07.008>
- Yu, Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), 1608–1615. <https://doi.org/10.1093/bioinformatics/btq249>
- Yun, Y., Kim, H. N., Kim, S. E., Heo, S. G., Chang, Y., Ryu, S., Shin, H., & Kim, H. L. (2017). Comparative analysis of gut microbiota associated with body mass index in a large Korean cohort. *BMC Microbiology*, *17*(1), 1–9. <https://doi.org/10.1186/s12866-017-1052-0>
- Zakerska-Banaszak, O., Tomczak, H., Gabryel, M., Baturo, A., Wolko, L., Michalak, M., Malinska, N., Mankowska-Wierzbicka, D., Eder, P., Dobrowolska, A., Slomski, R., & Skrzypczak-Zielinska, M. (2021). Dysbiosis of gut microbiota in Polish patients with ulcerative colitis: a pilot study. *Scientific Reports*, *11*(1), 1–13. <https://doi.org/10.1038/s41598-021-81628-3>
- Ze, X., Duncan, S. H., Louis, P., & Flint, H. J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME Journal*, *6*(8), 1535–1543. <https://doi.org/10.1038/ismej.2012.4>
- Zhai, L., Huang, C., Ning, Z., Zhang, Y., Zhuang, M., Yang, W., Wang, X., Wang, J., Zhang, L., Xiao, H., Zhao, L., Asthana, P., Lam, Y. Y., Chow, C. F. W., Huang, J. dong, Yuan, S., Chan, K. M., Yuan, C. S., Lau, J. Y. N., ... Bian, Z. xiang. (2023). Ruminococcus gnavus plays a pathogenic role in diarrhea-predominant irritable bowel syndrome by increasing serotonin biosynthesis. *Cell Host and Microbe*, *31*(1), 33-44.e5. <https://doi.org/10.1016/j.chom.2022.11.006>

- Zhang, B., Li, L. Q., Liu, F., & Wu, J. Y. (2022). Human milk oligosaccharides and infant gut microbiota: Molecular structures, utilization strategies and immune function. *Carbohydrate Polymers*, 276(September 2021), 118738. <https://doi.org/10.1016/j.carbpol.2021.118738>
- Zhang, Y., Wang, L., Wiredu Ocansey, D. K., Wang, B., Wang, L., & Xu, Z. (2021). Mucin-type O-glycans: Barrier, microbiota, and immune anchors in inflammatory bowel disease. *Journal of Inflammation Research*, 14, 5939–5953. <https://doi.org/10.2147/JIR.S327609>
- Zhao, X., Li, J., Zhang, D., Huang, Z., Luo, C., Jiang, L., Huang, D., & Zhang, G. (2022). Mechanism of salicylic acid in promoting the rhizosphere benzo[a]pyrene biodegradation as revealed by DNA-stable isotope probing. *Science of the Total Environment*, 810, 152202. <https://doi.org/10.1016/j.scitotenv.2021.152202>
- Zoetendal, E. G., Raes, J., Van Den Bogert, B., Arumugam, M., Booijink, C. C., Troost, F. J., Bork, P., Wels, M., De Vos, W. M., & Kleerebezem, M. (2012). The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME Journal*, 6(7), 1415–1426. <https://doi.org/10.1038/ismej.2011.212>
- Zoetendal, E. G., Von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. L., & De Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Applied and Environmental Microbiology*, 68(7), 3401–3407. <https://doi.org/10.1128/AEM.68.7.3401-3407.2002>