Starch Utilisation by the Beneficial Microbiota Genus *Bifidobacterium* for Improved Health Outcomes

Molly Eve Millar

A thesis submitted for the degree of Doctor of Philosophy to the University of East Anglia, for research conducted at the Quadram Institute Bioscience.

Quadram Institute Bioscience

March 2023

Abstract

The gut microbiome is an important axis of health and disease. A key function of gut bacteria is fermentation of dietary fibre producing short chain fatty acids (SCFA) which directly benefit host health. Fibre includes forms of starch which are resistant to digestion. Degraders of resistant starch (RS) include *Bifidobacterium*, a genus of beneficial, Gram-positive, anaerobic commensals. They first colonise the infant gut as degraders of human milk sugars and reduce in relative abundance with age as available substrates decline. Adequate fibre and RS intake is associated with higher bifidobacterial abundance and is generally regarded as a marker of a 'healthy gut microbiome'.

This thesis explored the relationship between starch hydrolysis capabilities of bifidobacteria and their associated niche e.g. infant or adult microbiome. *Bifidobacterium* isolates were assessed for RS degrading phenotype using a range of RS structures. The ecological impacts of strain supplementation combined with a high-amylose maize starch ('synbiotic') using an in vitro fermentation model of the infant and adult microbiome were tested. The biomolecular interface between strong degrader *Bifidobacterium pseudolongum* and RS was researched using genomic, transcriptomic, proteomic, and metabolomic investigation.

Bifidobacterium strains presented niche-specific starch degradation phenotypes: pre-weaning species were non-degraders; transitional (weaning) species were moderate utilisers; and strong degraders were associated with the adult or

ruminant microbiota. A large gene cluster with novel protein functions (amylase flanked by starch binding modules, including RS-binding module CBM74) was discovered in primary degrader *B. pseudolongum* which was found to be upregulated in the presence of starch. The structure-function relationship between RS and *Bifidobacterium* with regards to SCFA production was starch structure and strain specific. New information in this field could aid in precision microbiome modulation, development of next generation synbiotics, or influence our understanding of how RS in foods could enhance *Bifidobacterium* levels in the human gut microbiota.

Access Condition & Agreement

Preface

Acknowledgements

I am extremely grateful to the supervisors of my project Dr Fred Warren, Professor Lindsay Hall, and Dr Hannah Harris. They have each provided the best guidance and encouragement I could have hoped for, not only as scientists but as my mentors. I have learned a tremendous amount about the scientific method and different disciplines of science which will be eternally valuable to me. I am very thankful for the amount of time and patience each of them has dedicated towards my project and my own development as a scientist. Thank you again Fred, Hannah, and Lindsay.

I am indebted to several scientific collaborators for their contributions not only for completion of this project but for infinitely expanding my skills and knowledge as a scientist. Nucleic acid extraction and sequencing was facilitated by the following individuals who contributed training or performance of laboratory work: Ms Rhiannon Evans and Mr David Baker, Ms Nancy Teng, Mr Thomas Atkinson, and Ms Cara Moss. Many thanks to Dr Raymond Kiu who kindly (and quickly) performed a whole genome assembly of a bacterial strain we sequenced. I acknowledge the significant bioinformatics training, support, and work by Dr Perla Rey. I acknowledge the support of Dr Andrea Telatin and Dr Sumeet Tiwari for running pipelines to process RNA and DNA data. Dr Trey Koev and Dr Sergey Nepogodiev provided training and assisted with NMR metabolomics experiments which I am very grateful for. I acknowledge and thank Dr Christina Ludwig and Ms Miriam Abele who contributed technical support and carried out proteomics from their lab in Germany. Thanks to David Lawson at JIC who ran a protein through AlphaFold at the last minute. I would like to also extend my thanks to and acknowledge the PEARL study team Professor Lindsay Hall, Ms Sarah Phillips, Ms Rachel Watt, Mr Thomas Atkinson, Ms Hayley Summerfield, Ms Laura Harris for facilitating the donation of baby stool samples as well as Ms Jennifer Ketskemety and Dr Melissa Lawson who isolated the BAMBI strains used in this thesis.

I would like to acknowledge and thank lab managers Dr Mike Ridout and Dr Andrea Hinkova and the technical team who run the labs I used for circa 4 years. Thank you to Dr Magdalena Kujawska who has donated her time and expertise from the beginning of my project. I acknowledge the contributions by means of additional training, advice, or lab support also kindly provided by Dr Emma Waters, Dr Manu Crost, Ms Barbora Nemeckova, Dr James Lazenby, Dr Ryan Griffiths, Dr George Savva, Dr Sally Dreger, Dr Andrew Page, and Dr Lee Kellingray. Fred and Hannah aided lab work and training for which I am very grateful. I would also like to thank Professor Nathalie Juge for assessing my work at the beginning of my

List of Contents

List of Contents

List of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Project outline

Starch exists as one of the most abundant biomolecules and a major carbohydrate reserve among plants. During human evolution, starch has become a cornerstone of the human diet thanks to cooking. Due to physicochemical variability in starch structure, small amounts of indigestible starches, known as resistant starch (RS), bypass the human digestive process in the small intestine, reaching the large intestine as fibre. Minimally processed forms of various plant foods contribute to daily intake of RS which are less prevalent in westernised dietary patterns (Miketinas, Shankar et al. 2020). Different types of fibre such as RS are a carbon source for certain bacteria which reside in our colon, collectively known as the gut microbiota. Fermentative processes influence health and disease states through metabolism of fibrous dietary substrates, usually by many taxa, which produces beneficial metabolites. The health effects of these metabolites include acting as an energy source for colonocytes and stimulating the production of satiety hormones, inhibiting cancer proliferation, and suppressing microbial pathogens in the gut (Makki, Deehan et al. 2018).

How starch structure influences degradation by bacterial taxa, both in terms of mechanism and scope of use by specific taxa, is not fully understood. Including chemical modification of starch, genetic alterations or breeding of plant crops, and

processing innovations, a wide variation in RS structure exists (Dobranowski and Stintzi 2021). Among several prominent degraders of RS is the *Bifidobacterium* genus, of which several species degrade RS, its components, and various structures. The roles of bifidobacteria as part of the gut ecosystem begin during microbial colonisation of the infant gut: the ecosystem becomes monopolised by *Bifidobacterium* which can efficiently and cooperatively degrade human breastmilk oligosaccharides (Masi and Stewart 2022). This thesis aims to explore how different starches interact with bifidobacteria. The mechanisms involved and the ecological impact of the interaction will be addressed in the context of how bifidobacteria co-evolved with humans as microbial commensals key to human health from early life and into adulthood. This work aims to contribute to current understandings of how bifidobacteria and RS contribute to health. New knowledge in this area may inform future microbiota modulation therapies or products which initiate health promoting processes as a result of the bacterial fermentation of RS.

1.2 Starch

1.2.1 Starch composition and biosynthesis

Starch is one of the most abundant biomolecules alongside cellulose (plants), chitin (insects, fungi), and glycogen (animals) (Robyt 2008, Gilbert 2011). As well as use in human and animal nutrition, industrial applications of starch are economically significant (Bertolini 2009). The importance of starch as an energy storage molecule for plants mirrors its ubiquity across plant species, varieties, and organs (Smith 2001). Starch comprises glucose polymers arranged in an organised structure into a compact starch granule. These polymers are glucose

sugar polysaccharides called α-glucans, namely amylose and amylopectin (Martin and Smith 1995). In principle, starch polymers are formed from glucose via a condensation reaction (Figure 1-1A); the breakdown of starch is therefore primarily by enzymatic hydrolysis. Amylose is an α(1→4)-linked, relatively unbranched (1 branch per 1000 residues) α-glucan which makes up around 15-30% of starch (Martin and Smith 1995, Tatge, Marshall et al. 1999). Amylopectin is a highly branched glucan also linked by α(1→4)-linked glucose residues, with α(1→6) linkages to form branch points (Smith 2001).

While there are multiple enzymes which participate in the complex process of starch production, the synthesis process in higher plants involves three main enzymes to facilitate the sequential starch biosynthetic pathway, exclusively taking place in the plastid or amyloplast (Figure 1-1B). Firstly, ADP glucose pyrophosphorylase (AGPase) removes phosphate from glucose-1-phosphate to form the sugar nucleotide ADP-glucose. Secondly, starch synthase (SS) catalyses α(1→4) glycosidic bond formation between the non-reducing end of a growing glucan chain and ADP-glucose. Finally for the branched-chain polymer amylopectin, starch branching enzyme (SBE) cuts the nascent chain to form an α(1→6) linkage between the reducing end of an α-glucan and the sixth carbon (C6) of another glucose residue on a different chain (Smith 2001). The involvement of enzymes such as amylase, debranching enzymes, and phosphorylation enzymes called α-glucan water dikinases are also important to starch granule formation in spite of being generally considered as degradative (Smith 2001, Merida and Fettke 2021).
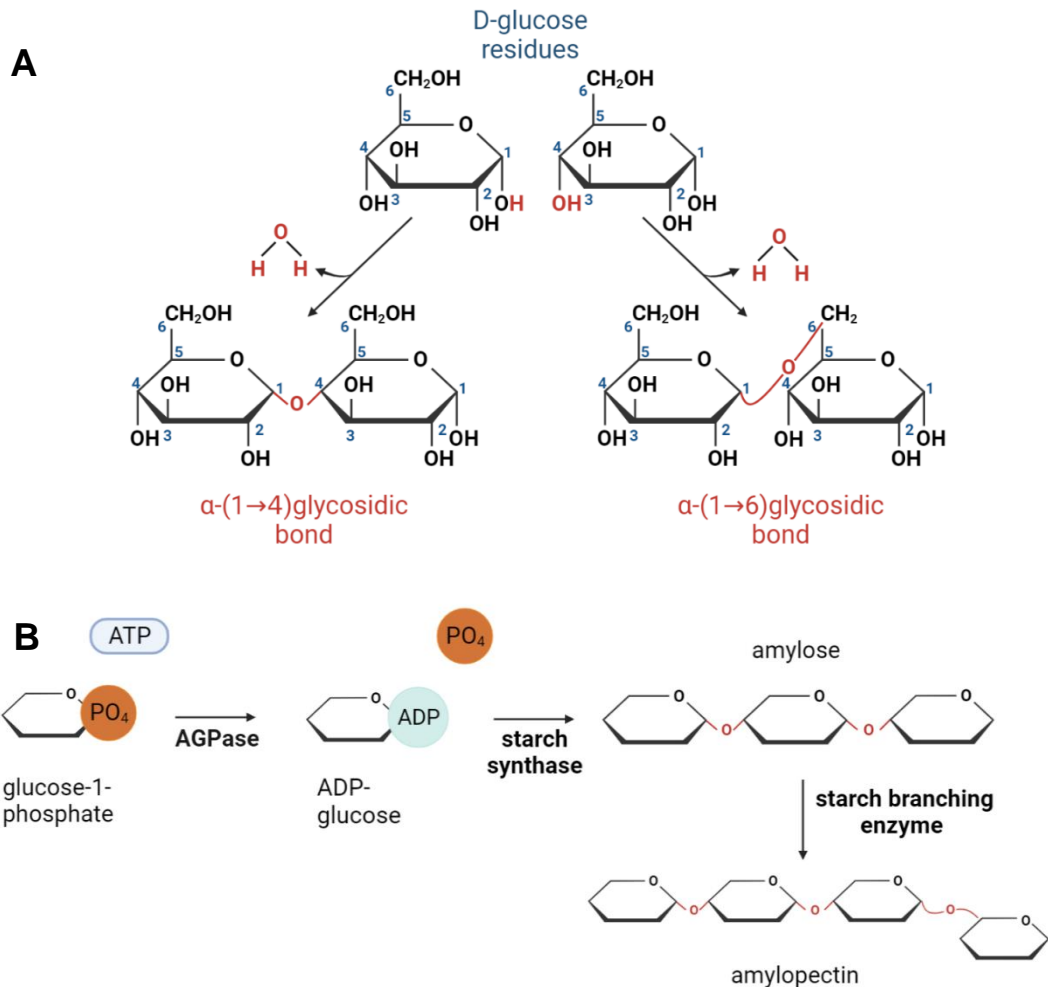
Figure 1-1 (A) The formation of glycosidic bonds involves a condensation reaction of hydroxyl groups from two carbons on adjacent monosaccharides [top], resulting in one carbon losing its OH group and another losing a hydrogen, producing one water molecule. The oxygen molecule bridges the two carbon rings to form a covalent bond. The structure of starch is built on glycosidic bonds between the first and fourth carbon of adjacent glucose residues, and the first and sixth carbon in a branched chain, respectively called α-(1→4) [left] and α-(1→6) linkages [right]. (B) In vivo, the starch biosynthetic pathway in plant chloroplasts involves a series of steps and multiple enzymes to form glycosidic bonds, branch points, and granule structure. Graphic created using BioRender.com.

Amylopectin contains multiple branch points where glucose chains attach. The various chains are categorised into A, B, and C chains: A chains are outer chains attached to B (inner) chains at branch points. The C chain carries other branches but contains the only reducing residue per molecule of starch (Martin and Smith

1995, Buléon, Colonna et al. 1998). The polymers arrange into alternating crystalline and amorphous lamellae, forming growth rings which create compact particles called starch granules (Figure 1-2).
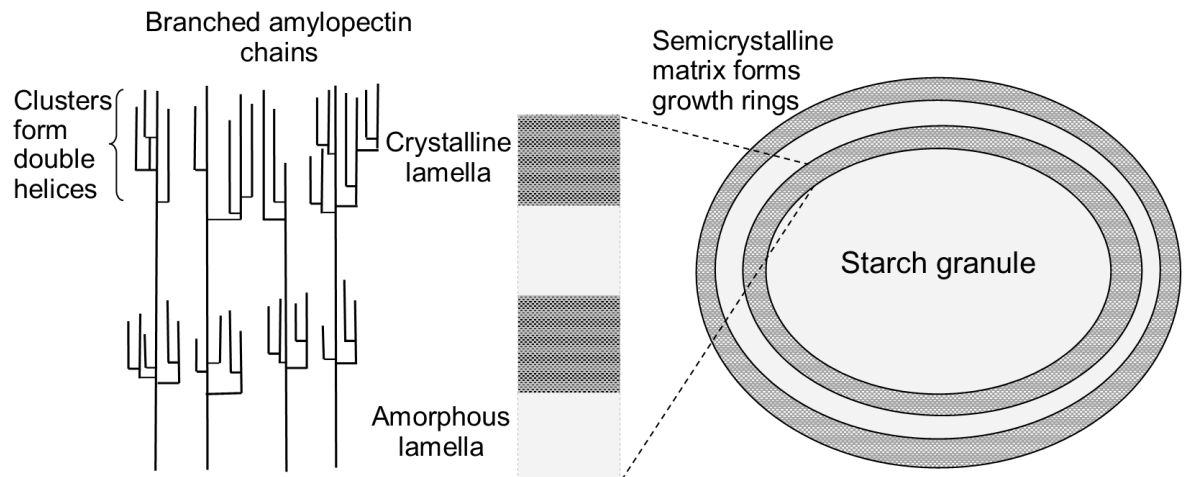
Figure 1-2 Starch granule formation showing the semi-crystalline nature of the final granule, comprised of alternating crystalline double-helical branches with amorphous regions of amylose and between branch points. Figure adapted with permission from Smith, A. M. (2001). "The biosynthesis of starch granules." Biomacromolecules 2(2): 335-341. Copyright (2001) American Chemical Society.

The initiation of starch granule formation has been studied extensively (Ball, Guan et al. 1996, Buléon, Colonna et al. 1998, Seung, Schreier et al. 2018). Although a basic understanding of starch biosynthesis and its chemical structure have been well-established for decades, new discoveries in the field are still being made. For example, Seung et al. revealed three key proteins in the chloroplast which interact with each other to initiate starch granule synthesis (Seung, Schreier et al. 2018).

The assembly of the starch granule (and thus its overall structure) is controlled by the branching of radially arranged amylopectin chains. At heavily branched regions of the amylopectin molecule, the branched chains form clusters where neighbouring chains associate with each other through hydrogen bonding (Smith

2001). Clusters allow double-helix formation between adjacent strands leading to crystallites during the formation of a starch granule.

1.2.2 Variables affecting starch structure

The type of crystallinity present in a starch is dependent on botanical origin. Plants produce different types of crystalline structure (polymorphs) described as A-type, B-type, and C-type (Gidley and Bociek 1985, Wang, Bogracheva et al. 1998, Gilbert, Witt et al. 2013). A-type crystallinity is present mostly in dry, above-ground plant organs, like grains. The geometry of A-type single-cell units allow tighter packing of amylopectin double-helices in the crystalline lamellar regions of the granule. The tighter the packing of the double-helices, the less water is trapped within the crystal structure. B-type starch is found in tubers and C-type exists in legumes as a mixture of A- and B-type polymorphs. Underground storage organs such as tubers tend to have a higher ambient moisture content and these conditions contribute to B-type crystallinity due to less compact helical packing which can allow increased water-binding (Englyst, Kingman et al. 1992, Roder, Gerard et al. 2009).

Plant genetics (species, variety) largely pre-determines amylose content (Deatherage, MacMasters et al. 1955), although environmental factors also contribute (Martin and Smith 1995). α-glucans in starch can vary depending on a variety of factors related to isoforms of enzymes in the biosynthetic pathway, which ultimately affects the number of glucose residues in a chain (degree of polymerisation) or extent of branching. The amylopectin chain lengths and their distribution relative to total chains are important for the polymers' ability to

associate helically and pack tightly together. Differences in molecular structure of starch can be determined by measuring the chain length distribution (CLD) of amylopectin, defined as the number of chains as a function of its length (Yu, Li et al. 2019). Different chain lengths and extent of branching leads to variations in shape and extent of crystallites (Jenkins and Donald 1995, Noda, Takahata et al. 1998) and amylopectin assembly in granules (Smith, Denyer et al. 1997, Gilbert, Witt et al. 2013).

The CLD of amylopectin has been shown to be determined by SS and its various isoforms, and its direct or indirect interaction with SBE (Maddelein, Libessart et al. 1994), as well as different expression levels of AGPase enzyme subunits, and post-translational modifications (Tetlow, Morell et al. 2004). Since the nature of SBE involves cutting (shortening) the amylopectin chain in order to form an $\alpha(1 \rightarrow 6)$ linkage, the extent of branching and chain length are linked through the action of the enzymes in this pathway. While it is unclear why plants produce amylopectin with specific CLD, the subsequent differences in amylopectin CLD and amylose content affects the assembly of starch within the granule, leading to predictable physical and functional changes relevant to human uses of starch (Vandeputte and Delcour 2004, Gous, Warren et al. 2017).

Amylose content normally lies between 15-30% of total starch content, unless the plant variety has been specifically selected or mutated to have an abnormally high or low amount of amylose. The maize crop has been extensively selected to produce 0%, 50%, and 70% amylose varieties, amongst others (Jenkins and Donald 1995). While amylose is not necessary for granule formation, it plays an

important role in disrupting the amylopectin packing by impeding double-helix formation in the crystalline regions by co-crystallising, counterintuitively increasing the crystalline region size (Jenkins and Donald 1995). Granule-bound starch synthase (GBSS) gene knockouts produce waxy (low amylose) varieties of starch as GBSS is the key enzyme involved in amylose synthesis (Tatge, Marshall et al. 1999, Zeeman, Tiessen et al. 2002). High amylose varieties are produced by knocking out branching enzyme SBEII (Liu, Ahmed et al. 2012) leading to longer B chains of amylopectin and an increase in amylose which affects crystallinity and increases gelatinisation temperature (You and Izydorczyk 2002). This phenomenon could be related to the fact that GBSSs are also known to elongate amylopectin (Maddelein, Libessart et al. 1994). In summary, the basic building blocks of starch, amylose and amylopectin, are a fundamental aspect which govern starch structure and functionality.

1.3 Starch in the diet

1.3.1 Dietary adaptation to starch

A specimen of 280-million-year-old fossilised starch granules was found on the exterior surface of a seed, indicating that the starch was not for germination (Liu, Bomfleur et al. 2018). This demonstrates that plants evolved to entice land animals with starch as an edible reward in exchange for carrying out seed dispersal, an example of evolutionary mutualism. The evolution of apes led to the departure of hominins as distinct from ancestors of modern chimpanzees over 6 million years ago. During the emergence the *Homo* genus, several key events in human evolution occurred (Figure 1-3). Expanding human populations brought

intense competition for resources, resulting in dramatic migration around 1-2 million years ago (Hardy, Brand-Miller et al. 2015). Occupation of regions which were less botanically diverse and scarcer in food resulted in rapid evolution: unfamiliar niches, growing brains, and increasing populations. Evolutionary pressures facilitated adaptation to increase dietary diversity and innovation to utilise new food sources.
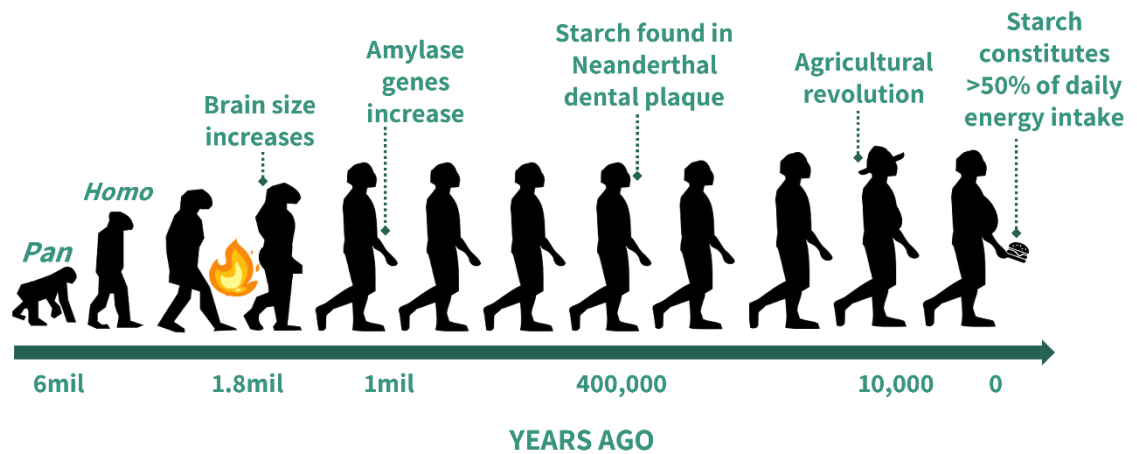
Figure 1-3 The long-standing relationship between humans and starch over our evolutionary history, culminating in 'the starch diet' where most populations derive the majority of calories from starch.

Around 1.8 million years ago, human ancestors developed the use and control of fire (Wrangham, Jones et al. 1999, Wrangham and Conklin-Brittain 2003). Wielding fire to cook, we increase caloric yield from starch-rich foods in particular, prominently available in arid, food scarce regions (Wrangham and Carmody 2010, Lordkipanidze, Ponce de León et al. 2013). Starch granules were found in the dental calculus of *Homo neanderthalensis* (Power, Salazar-García et al. 2018), who lived alongside *Homo sapiens* until around 400,000 years ago.

Just as cooking facilitated a remarkable diversification of human diets, so too did agriculture. Starch-rich foods such as tubers, seeds, and cereals are available in a

variety of forms, climates, regions, and time of year (Carmody and Wrangham 2009). Cultivation of starchy staples were therefore a reliable source of easily digested, energy dense, year-round food (Hardy, Blakeney et al. 2009, Hardy, Brand-Miller et al. 2015). Decreases in jaw force potential and expansion of salivary amylase gene AMY1 is evidence of a complex biological reliance on starch (Lucas, Ang et al. 2006). The energy contributed by cooked starches is also particularly relevant in early life where starch has a role in promoting health during infancy (Rodríguez, León et al. 2022). Human infants have large energy and nutrient requirements for adequate development of the body and brain, meaning the critical stage of weaning in infants from energy dense breastmilk to easily digestible solid foods also rich in energy was likely a driving factor in infant survival (Lin and Nichols 2017). Our reliance on starch has remained a strong feature of our species over evolutionary timescales, with agriculture, processing technology, and cooking revolutionising the bioavailability of nutrients (Power, Salazar-García et al. 2018).

Technological advances in cultivating plants not only increased yield of a variety of crops but helped to optimise caloric density of seeds or grains (Slavin, Jacobs et al. 2000). Crushing, grinding, and milling of grains is a process which achieves several objectives: increasing surface area, facilitating release of starch and removal of fibrous aspects such as the husk, decreasing grain volume, and increasing caloric density. These new characteristics improve digestibility, cooking time, and reduces the cost of transporting foods. While separating the bran from grains increases the calorie density which is beneficial in the short-term, it can have some drawbacks for long-term health. Highly processed grain products

increase postprandial glycaemia more rapidly and are generally lower in micronutrient density (Ross, Brand et al. 1987, Liyana-Pathirana and Shahidi 2007).

1.3.2 Human starch digestion

The importance of starch in the human diet today is reflected in most agricultural regions of the world where the majority of calories are gained from starch (Perry, Dominy et al. 2007, Gilbert, Witt et al. 2013). Additionally, the α-glucans present in starch make up the majority (<80%) of all polysaccharides in the human diet (Englyst and Hudson 1996). The most notable physiological adaptation related to starch digestion is the duplication of amylase genes (Carpenter, Dhar et al. 2015, Atkinson, Hancock et al. 2018), thought to have occurred around 1 million years ago through a retroviral insertion event (Samuelson, Wiebauer et al. 1990). This suggests our ancestors gained an evolutionary advantage from eating starch allowing it to persist universally in human physiology, and importantly, chronologically correlated to human use of fire for cooking (Hardy, Brand-Miller et al. 2015).

Cooked starches are a revolutionary introduction to the human diet since uncooked starch is largely indigestible. Alongside development of cooking as a technological trait of our species, adaptations in the digestive process evolved to break down the cooked starches being introduced. Digestion begins in the mouth, where mastication of food by teeth mixed with saliva forms a cohesive food bolus before swallowing (Lucas, Ang et al. 2006). Stimulated by food, α-amylase present in saliva is secreted by serous parotid glands and hydrolyses starch present in

food (Pedersen, Bardow et al. 2002). Although, only about 5% of starch is broken down in our mouths (Lee, Bello-Pérez et al. 2013). Muscular contractions allow food to be transported from mouth to stomach, where it continues to be mixed with enzymes and acid by peristalsis. Reaching the duodenum in the small intestine, the chyme contains several pancreatic enzymes such as pancreatic α-amylase and proteases. The pancreatic α-amylase in duodenal secretions converts starch into primarily oligosaccharides or disaccharides (e.g. maltose, α-limit dextrin). The *AMY1* and *AMY2* genes encode α-amylase secreted primarily in the salivary glands and pancreas, respectively. The two genes have 97% sequence similarity (Butterworth, Warren et al. 2011). Brush-border disaccharidases complete the final step by turning the small glucose chains into absorbable glucose, which is assimilated in the jejunum by the action of microscopic, folded protrusions called villi (Lee, Bello-Pérez et al. 2013).

Postprandial glycaemia and insulinaemia denote the increase of glucose in the blood following a meal, and the subsequent increase of insulin in the blood triggering glucose to enter somatic cells. Glycaemia is affected by a multitude of factors related to the origin, biochemical structure, processing, cooking, and context of a food as part of a meal. As defined by Jenkins et al., glycaemic index (GI) is a quantitative *in vivo* measurement of carbohydrate quality in terms of how quickly after ingestion carbohydrates are absorbed into the blood in the form of glucose. It is calculated as the area under the curve of postprandial blood glucose for two hours after the ingestion of a carbohydrate compared to a positive control food such as glucose or white bread (Jenkins, Wolever et al. 1981).

1.3.3 Resistant starch

To compliment the GI method, Englyst and colleagues pioneered the enzymatic classification of starch (Englyst and Hudson 1996). They categorised starch into three groups: rapidly digestible starch (RDS), slowly digestible starch (SDS) and RS (Englyst, Kingman et al. 1992). Various starches were measured using controlled enzymatic hydrolysis with pancreatin and amyloglucosidase and released glucose was measured by colorimetry using a glucose oxidase kit. These measurements reflect the rate of starch digestion *in vivo*. RDS and SDS are both fully hydrolysed and absorbed in the small intestine as glucose, just at different rates. RS refers to the fraction or sum of starch and starch degradation products that pass into the large intestine, meaning that the starch is either fully intact or only partially hydrolysed (Englyst, Kingman et al. 1992).

The physical structure of food is a major determinant of the rate of digestion. By the Englyst definition, RS is unable to be hydrolysed fully by endogenous enzymes and will enter the large intestine where it will be fermented by the gut microbiota. Minimally processed (whole) grains and legumes even when fully cooked have higher amounts of fibre and larger particle size compared to fine-milled equivalents which impact both glycaemia and bacterial fermentation rates (Yao, Flanagan et al. 2023). Different subcategories of RS arise due to various physical and biochemical obstacles which prevent hydrolysis in the upper GI tract, and thus mechanisms by which RS evades digestion in the small intestine is explained by four main categories Figure 1-4.

Encapsulated starch (RS Type 1) is within whole or partly disrupted plant structures such as whole grains or seeds (including legumes) and intact plant cells. This starch is inaccessible to endogenous amylolytic enzymes. In addition, rigid cell walls can inhibit swelling and dispersion of starch, processes which affects the ability of enzymes to hydrolyse starch fully. Similarly, densely packed starch in foods such as pasta is encapsulated in a protein matrix (Fuentes-Zaragoza, Riquelme-Navarrete et al. 2010).

Type 1	Type 2	Type 3	Type 4
Physical inaccessibility due to intact structures impeding enzyme	Presence or retention of native starch which is highly ordered	Gelatinised starch allowed to partially re-crystallise, increasing order	Chemical alteration of the chemical structure inhibits enzyme access
Encapsulation	Crystallinity	Retrogradation	Modification
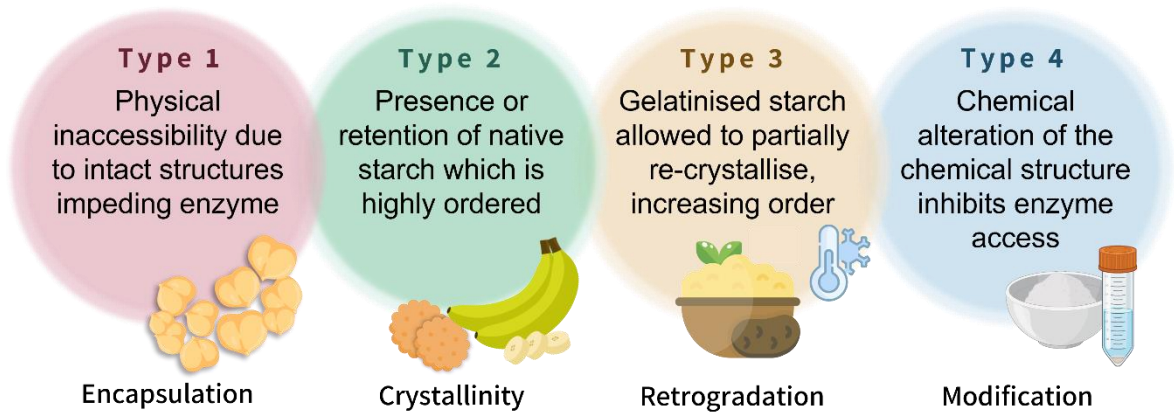
Figure 1-4 Descriptions of each type of resistant starch including the four main types: (1) encapsulated, (2) crystalline, (3) retrograded, and (4) chemically modified starch. Graphic created using BioRender.com

Native starch (RS Type 2) present or retained in food in its granular form is present in uncooked foods such as green bananas and dry heat processed foods such as biscuits or crackers. Highly ordered granular starch inhibits enzyme activity by steric hindrance due to inability of enzymes to penetrate granular starch.

Retrogradation of starch (RS Type 3) involves partial recrystallisation of amylose during a cooking and cooling process made possible due to hydrogen bonding between chains of relatively unbranched amylose. B- and C-type crystallinity generally have increased resistance to enzymatic breakdown compared to A-type

(Planchot, Colonna et al. 1997). Re-crystallisation and the fact that retrograded starch forms B-type crystallinity makes this type of starch particularly resistant. RS Type 4 is chemically modified starch. While these classifications explain the mechanisms of digestive evasion in the upper GI tract, the physiological effects once in the intestine do not necessarily correspond with the current classification system. Thus, some researchers suggest that since the RS primary effects occur in the gut, the RS types may be better classified as to indicate their physiological consequences rather than evasion properties (Warren, Fukuma et al. 2018).

While starch intake varies between 40-75% total energy depending on region (Englyst and Hudson 1996), RS varies more significantly as a result of regional food culture and traditional ingredients. For example, in Northern Europe or North America, RS intake is likely to be as low as 3 grams per day (Fuentes-Zaragoza, Riquelme-Navarrete et al. 2010, Miketinas, Shankar et al. 2020), compared to Italy which is around 9 grams (Brighenti, Casiraghi et al. 1998). Regions with higher intakes are in South and East Asia (10-15 grams) (Chen, Liu et al. 2010) and in areas of rural Africa, around 38 grams per day (O'Keefe, Li et al. 2015). These studies show that high dietary intake of RS is associated with grain products (porridge, rice, pasta, or noodles), tubers and legumes, especially if cooked and then cooled.

Recognised limitations during estimations of RS intake include differences between and variation within approaches (*in vitro* or *in vivo*) e.g. evaluating RS content in typical foods (storage, moisture content, cooking method) versus *in vivo* measurements of remaining RS in stool or ileostomy effluent samples (Miketinas,

Shankar et al. 2020). Therefore, commentary on RS intake based on factors such as geography or time period based on these estimations make it difficult to draw comparisons.

1.4 Starch fermentation

1.4.1 The gut microbiota

The *in utero* foetal microbiome has long-believed to be sterile as reiterated by recent investigations (Kennedy, de Goffau et al. 2023). After birth, the colon goes on to become host to an ecosystem of multiple kingdoms of life – bacteria, archaea, viruses, and fungi – which act as an interface between dietary components and host cells (Thursby and Juge 2017, Ogunrinola, Oyewale et al. 2020). The human body is host to approximately 3.8×10^{13} bacterial cells which is on the same order of magnitude to human cells in the body (Sender, Fuchs et al. 2016). The greatest number of bacterial cells concentrated to one organ of the human body is the large intestine containing 10^{11} cells per mL which dwarfs the upper gastrointestinal tract by several orders of magnitude (Sender, Fuchs et al. 2016). Beneficial microbes' functions include fermentation of dietary components, resisting colonisation of pathogens, controlling inflammation processes, and maintenance of microbial functionality (Lawley and Walker 2013).

The physiological contribution of the gut microbiota to general homeostasis and prevention of disease cannot be understated (Makki, Deehan et al. 2018). In terms of prevention of pathogenic bacterial colonisation, the mucin of the intestinal lining acts as a physical barrier for pathogens, and beneficial microbes participate in the maintenance of this barrier and promote tight junctions which enhance the

defences of the gut against pathogenic bacteria and fungi (Johansson, Ambort et

al. 2011, Camilleri, Madsen et al. 2012). *Bifidobacterium longum* has been shown

to produce extracellular vesicles containing mucin-binding proteins which may aid

in long-term colonisation of the intestine, showing complex methods of adapting to

and inhabiting the microbial ecosystem in the human gut (Nishiyama, Takaki et al.

2020). Production of bacteriocins (antimicrobial peptides) and extracellular

vesicles are important mechanisms of interaction between opportunistic pathogens

and regulation of host immunity (Ibrahim 2019, Hosseini-Giv, Basas et al. 2022).

Perturbations of intestinal microbial ecosystems can occur (sometimes termed 'gut

dysbiosis') where the symbiotic balance shifts towards negative physiological

processes, whether that be involving blooms of inflammatory or pathogenic

bacteria which produce toxins or fluctuations in short chain fatty acids (SCFA)

produced by general bacterial metabolism (Walker and Lawley 2013). In particular,

a common result of intestinal perturbations is the reduction in intestinal barrier

integrity which makes the epithelia vulnerable to infiltration of pathogens or their

toxins which increases inflammation (Camilleri, Madsen et al. 2012). Common

pathologies associated with these processes include the umbrella term

inflammatory bowel disease (IBD) which encompasses a spectrum of intestinal

pathologies involving chronic inflammation. Diet, especially dietary fibre, can be a

trigger for IBD patients (Adolph and Zhang 2022).

Dietary fibre fermentation by gut bacteria also plays an important role in metabolic

health and hunger hormone signalling. Gastrointestinal L cells secrete hunger

reduction hormones, namely peptide tyrosine tyrosine (PYY) and glucagon-like

peptide 1 (GLP-1), which stimulate satiety signals (De Silva and Bloom 2012,

Burcelin, Gourdy et al. 2014). The SCFA acetate has been shown in humans to reduce appetite via stimulation of L cells (Frost, Sleeth et al. 2014). However, some evidence shows the process is not stimulated in acute cases of SCFA released from fibre fermentation, but over the course of a long-term diet rich in fermentable fibres (Rahat-Rozenbloom, Fernandes et al. 2017). The physiological mechanisms of GLP-1 on appetite are well understood such that administration of a GLP-1 analogue has been successful in humans to improve fat mass reduction by decreasing hunger (Mehta, Marso et al. 2017).

Human evolution was predicated on dietary flexibility and variation, presenting a range of nutrients throughout the digestive tract. Development of microbial ecological adaptation strategies, in addition to symbiotic co-evolution with the human host, resulted in diversity of function in the gut microbiota. Thus, the gut microbiota is responsive to and modifiable by diet (Walker, Ince et al. 2011, David, Maurice et al. 2014, Carmody, Gerber et al. 2015, Aguirre, Eck et al. 2016, Chung, Walker et al. 2016). Accordingly, the higher the amount of resistant carbohydrates ingested, the higher the number and type of carbohydrate-degrading, or saccharolytic, bacteria present with enzymatic capabilities to utilise them (Walker, Ince et al. 2011, Lee, Bello-Pérez et al. 2013, Warren, Fukuma et al. 2018, Wardman, Bains et al. 2022). As a result of the variety and quantity of fermentable carbohydrates available, bacteria have adapted a wide range of highly effective carbohydrate degrading enzymes (Kaoutari, Armougom et al. 2013). Human co-evolution with gut bacteria adapted to fermentation of new dietary carbohydrates improved fitness for both host and bacterium (Brooks, Kohl et al. 2016, Koskella,

Hall et al. 2017). Thus, gut bacteria are an important aspect of human survival which may have allowed us to adapt to new niches (Moeller, Li et al. 2014).

The intake of fibre as a dominant food component over evolutionary timescales up to the present day varies depending on geography, food culture, availability; the gut microbiome composition of individuals with different nationalities reflects this (Yadav, Ghosh et al. 2016). Fermentation is a main physiological outcome for fibre, however the positive effects of fibre passing through the intestinal lumen are not limited to just fermentation processes. They include the ability of soluble fibres to bind water which softens the stool and binding to bile salts which lowers serum cholesterol (Li and Komarek 2017); insoluble fibre's bulking effect adds volume, decreasing transit time and impacting energy extraction (Boekhorst, Venlet et al. 2022). Fibre can serve one or many of these properties (not all types are fermentable) and not all bacteria are capable of degrading fibre.

Fermentation is an enzymatic process which commonly refers to microbial substrate degradation of an organic substance. Chemical energy from a carbon source is converted to ATP using microbial metabolic pathways (Nicholson, Holmes et al. 2012). The major end product metabolites are organic acids including butyrate, acetate, and propionate, contributing up to 10% of host calorie intake, in addition to well-established health-promoting mechanisms for the host, and cross-feeding other bacterial taxa (Makki, Deehan et al. 2018). There is significant evidence to support the effects of SCFAs, particularly butyrate, on health which has been previously well-reviewed (Makki, Deehan et al. 2018). Specifically, SCFAs contribute to colonocyte health through acting as an energy

source and they exert anti-neoplastic effects (Roediger 1982, Fattahi, Heidari et al. 2020). The production of acids lowers luminal pH of the large intestine, a known stress inducer of bacteria and thus pathogen suppressor (Ricci, Mackie et al. 2022). Other mechanisms of pathogen suppression includes SCFAs influencing bacterial virulence gene expression (Fukuda, Toh et al. 2011, Mirzaei, Dehkhodaie et al. 2022). SCFA production also encourages crosstalk between bacteria and the host in addition to metabolites such as quorum sensing molecules mediated by dietary input (Oliveira, Cabral et al. 2021, Falà, Álvarez-Ordóñez et al. 2022).

1.4.2 Bacterial amylolytic enzymes

Starch is a macromolecule with many potential recalcitrant structures. Multiple enzymes are required to associate with the surface, degrade the supramolecular structure, and break down the polymer into oligosaccharides and sugars. Although some bacteria are competitive, degradation of substrates as a collective to crowdsource stages of the process using enzymes maximises the energy gain for least input on individuals (Cerqueira, Photenhauer et al. 2020). Several bacteria have been identified as key players in starch fermentation and their microbe interactions can be described as hierarchical in structure (Dobranowski and Stintzi 2021). Non-exhaustively, there are known primary and secondary degraders of RS, the members of which belong to multiple genera of common gut bacteria described as 'keystone degraders' (Ze, Duncan et al. 2012, Centanni, Lawley et al. 2018). Examples of proficient, prominent degraders of starch are species such as *Ruminococcus bromii, Eubacterium rectale, Bacteroides thetaiotaomicron*, and *Bifidobacterium adolescentis.* The ability of bacteria to utilise RS is frequently strain-specific and there are differences in how efficiently bacteria hydrolysis

different starch structures depending on factors such as botanical origin (Dobranowski and Stintzi 2021). This emphasises the need for study into both strain-level phenotype and mechanism of utilisation.

Bacterial biomolecular equipment to ferment fibre includes a variety of Carbohydrate-Active enZYmes (CAZymes) and carbohydrate binding modules (CBM). There are several families of CAZymes indicating the diversity of carbohydrates occurring in nature (Kaoutari, Armougom et al. 2013, Warren, Fukuma et al. 2018). CAZymes have a sequence-based classification system: families have conserved features such as the catalytic residues, molecular mechanism of degradation, and stereochemical outcome (Cantarel, Coutinho et al. 2009, Wardman, Bains et al. 2022). Members of the same family can have varying substrate specificity; enzymes themselves can also have dual activity e.g. binding and/or catalytic. Enzymes which cleave glycosidic bonds in carbohydrates (or between a carbohydrate and a non-carbohydrate moiety) include glycoside hydrolases (GH) by insertion of a water molecule i.e. hydrolysis. According to CAZypedia, there are around 128 families of GHs (Available at www.cazypedia.org [Accessed 28/11/2022]). Among the bacteria able to ferment fibre in the gut microbiota, their capabilities to utilise dietary carbohydrates far outweighs that of the host. The human genome possesses around 17 GH CAZymes for the breakdown of food glycans, whereas a single bacterium could have hundreds (Flint, Scott et al. 2012, Bhattacharya, Ghosh et al. 2015). Important to starch hydrolysis is the GH13 family of CAZymes, which includes α-amylases involved in the cleavage of α1→4 linkages and multiple other enzymes which can cleave α1→6 linkages, a summary of these is presented in Table 1-1.

Table 1-1 A compact list of some GH families associated with starch utilisation in bacteria, including examples of enzymes and their activities.

CAZymes family or subfamily	Enzyme	Bond breakage	Substrate examples	Hydrolysis end products
GH13_1 GH13_5 GH13_7 GH13_15 GH13_24 GH13_27 GH13_28 GH13_32 GH57	α-amylase	α1→4	starch, amylose, amylopectin	Maltose, maltotriose, α- and β-limit dextrins, glucose
GH13_17 GH13_21 GH13_30	α-glucosidase (maltase)	α1→4	terminal, non-reducing end of maltose, maltotriose, starch	glucose
GH13_31	oligo-1,6-glucosidase	α1→6	starch, isomaltose, isomaltotriose	glucose
GH13_11	Isoamylase (debranching enzyme)	α1→6	Starch, amylopectin, β-limit dextrins	maltose
GH13_12 GH13_13 GH13_14	Pullulanase (type 1) (debranching enzyme)	α1→6	pullulan, amylopectin, α- and β-limit dextrins	maltotriose, maltose, glucose
GH13_20	Neopullulanase (type 2)	α1→4	amylose, amylopectin, pullulan	panose, maltose, glucose

Commonly associated with amylolytic enzymes are CBMs called 'starch binding domains' some of which are starch structure specific (Table 1-2) (Cockburn, Suh et al. 2018, Cockburn, Kibler et al. 2021, Cerqueira, Photenhauer et al. 2022). These modules facilitate the physical positioning of the active site and improve catalytic activity of enzymatic domains. Starch associated CBMs are 20, 21, 25, 26, 34, 41, 45, 48, 53, 58, 68, 69, 74, 82, and 83 (Cerqueira, Photenhauer et al. 2020).

Table 1-2 A compact list of starch related CBM families and their respective binding affinities (http://www.cazy.org [Accessed 19/1/23]) Koropatkin et al. (2020).

CBM	Binding activity	Co-occurrence
20	Starch granules, amylose, and amylopectin; maltooligosaccharides: maltose, maltoheptaose, maltodecaose	CBM25 CBM34 CBM48
21	Starch granules, amylose oligosaccharides derived from starch including maltose through maltoheptaose, β- and γ-cyclodextrins, isomaltotriose and isomaltotetraose	
25	Unknown	CBM74, CBM20
26	Unknown	CBM74
41	Amylopectin, amylose, glycogen, and pullulan; oligosaccharides including maltose, maltotriose, longer maltooligosaccharides up to DP7, glucosyl-maltotriose and glucosyl-maltotriosyl-maltotriose. CBM41 modules are specific for α(1→4)-linked glucose chains and may accommodate a linear α(1→6)-linked glucose moiety.	
48	Linear and cyclic α-glucans related to and derived from starch containing both α(1→4)- and α(1→6)-linkages	
58	Binds exclusively to α(1→4)-linked glucan structures in starch; founding member from the neopullulanase SusG of the human gut symbiont *Bacteroides thetaiotaomicron*	
74	Soluble potato starch, boiled granular potato, wheat, and waxy corn starch (type 3 RS), amylose and amylopectin, raw granular starches (type 2 RS) from potato, wheat, and waxy corn; higher affinity for potato derived starch over that from wheat and maize suggesting that the CBM74 domain has a higher affinity for starches with B-type crystallinity over A-type.	GH13_28 GH13_19 GH13_32
82	Amylopectin and pullulan, beta-cyclodextrin and glycogen with similar affinity, with slightly weaker affinity for maltoheptaose; binds to corn starch granules, both from a wild-type source and from a high amylose source (HiMaize 260) with approximately equal affinity but did not demonstrate binding to potato starch or a chemically crosslinked starch	

Another enzyme group, lytic polysaccharide monooxygenases (LMPO), act on

relatively crystalline structures (Lo Leggio, Simmons et al. 2015). LPMOs have the

ability to oxidise glycans at α(1→4) linkages, therefore causing chain breakage. A

relatively new discovery, they are classed as auxiliary activity enzymes in the CAZy database, as their activity allows difficult polysaccharide structures to be more susceptible to hydrolysis by glycoside hydrolases.

1.4.3 Starch hydrolysis systems in bacteria

In the presence of starch, bacteria pivot their cellular machinery to starch breakdown which involves a genetic expression shift to produce enzymes able to bind to and attack the substrate and import the sugars intracellularly (Foley, Martens et al. 2018). Under strong competition, some bacteria have evolved specialised mechanisms to optimise the degradative process, summarised in Table 1-3.

Table 1-3 Some bacteria evolved highly specialised cellular machinery to hydrolyse RS starch in the competitive environment of the gut microbiota.

Phylum	Example	Mechanism	Reference
Bacteroidetes	*Bacteroides thetaiotaomicron*	Specialised starch uptake system (sus)	(Thursby and Juge 2017)
Firmicutes	*Ruminococcus bromii*	Amylosome: amylase complex on their exterior which is enabled by cohesin-dockerin interactions, and starch adherence system (sas)	(Ze, Ben David et al. 2015, Ndeh and Gilbert 2018, Cerqueira, Photenhauer et al. 2022)
Actinobacteria	*Bifidobacterium adolescentis*	Amylase production and adhesion to starch, particularly RS	(Crittenden, Laitila et al. 2001, Kim, Shin et al. 2018)

R. bromii possesses a unique starch utilisation mechanism which underpins its status as a keystone degrader of RS. The amylosome is a protein complex on the cell exterior which arranges α-amylase proteins facilitated by cohesin-dockerin interactions (Ze, David et al. 2015). Elements of this system work in concert: signal

peptides target the proteins for secretion, CBMs enable capture of substrates, enzymes contain catalytic modules for the degradation starch, and cohesins and dockerins arrange the proteins together in a complex.

The *R. bromii* system to utilise starch is in contrast to *B. thetaiotaomicron,* another highly proficient starch degrader, which relies on a sequestration-type system called a starch uptake system (sus). The proteins expressed from the starch utilisation locus of *B. thetaiotaomicron* are transcriptionally induced by the presence of maltose (Foley, Martens et al. 2018). Genes encoding glycan cleaving enzymes in the *sus* of *B. thetaiotaomicron* have a signal peptide which acts as a secretion signal and enables the proteins to be extracellularly expressed (Kaoutari, Armougom et al. 2013). Sus encodes 8 genes including 5 outer membrane proteins (SusCDEFG), the latter four of which have starch-binding sites (Arnal, Cockburn et al. 2018, Foley, Martens et al. 2018). At the cell surface, SusDEFG capture extracellular starch. SusG is a membrane-bound α-amylase which can degrade starch at the cell surface into shorter chains such as malto-oligosaccharides which are transported across the outer membrane by SusC. SusA and SusB, α-glucosidase and neopullulanase, complete hydrolysis to glucose which is transported to the cytoplasm (Foley, Martens et al. 2018).

Clusters of CAZyme genes which are activated in the presence of their substrates are defined as polysaccharide utilisation loci (PUL). *B. thetaiotaomicron* possesses 88 different PULs, the first discovered being *sus* (Foley, Martens et al. 2018). Gene clustering is present across several phyla but is most well described in *Bacteroides.* Generally, PULs encode glycan degrading systems are characterised

by adjacent gene pairs for an outer membrane transporter (SusCh) and an associated glycan-binding protein (SusDh) (Foley, Cockburn et al. 2016). Examples of well characterised starch-degrading gene clusters are *E. rectale* Amy13K and Amy13B and *B. thetaiotaomicron sus* gene cluster (Cerqueira, Photenhauer et al. 2020). In contrast, *R. bromii* has non-clustered amylase, dockerin, scaffoldin, and cohesion genes which assemble the amylosome extracellularly, precluding the need for the enzymes to be translated together (Ze, David et al. 2015).

1.4.4 Bacterial metabolism of carbohydrates

A primary reason for SCFA production in a bacterial ecosystem is as a waste product required to maintain redox balance in the anoxic environment of the gut (van Hoek and Merks 2012). The SCFA themselves then act on bacterial fitness and host health by stimulating physiological adaptations and promoting health. In Figure 1-5, oxidation and reduction during multiple steps balance electrons and maintain anoxia in the ecosystem by production of hydrogen and carbon dioxide. Bacterial hexose and pentose sugar metabolism involves a complex fatty acid synthesis pathway using the phosphotransferase system (PTS) through the action of the phosphoenolpyruvate (PEP) enzyme shown in Figure 1-5. Deoxy sugars (which have hydrogen at carbon 6 instead of a hydroxyl group) such as fucose and rhamnose can inter the PTS but also its own degradative pathway, called the propanediol pathway, to produce propionate.

The functionality of the PTS depends on multiple factors (Macfarlane and Macfarlane 2003). For example, the availability of a carbon substrate can affect

the end products of fermentation in bifidobacteria which favour acetate and formate production in carbon excess, and acetate and lactate in carbon scarcity. Acetate is the primary SCFA produced because it yields more ATP from pyruvate with fewer enzyme catalysed steps (van Hoek and Merks 2012). Some bacteria preferentially produce butyrate as a primary metabolic product (Figure 1-5). The butyrate production pathway utilises intermediates and end-products of the acetate pathway, one mechanism of interaction between microbes in the gut (Louis and Flint 2017). The cross-feeding of metabolites between acetate and butyrate producers benefits the host and is especially prevalent during starch degradation. For example, *R. bromii* is implicated in cross-feeding other important health-promoting species of bacteria such as *E. rectale* because a proportion of the outputs of the amylosome are released (Walker, Ince et al. 2011). *E. rectale* has been shown to exhibit a variety of CBMs appended to its cell surface expressed GHs which target specific starch structures (Cockburn, Suh et al. 2018). *E. rectale* is implicated in a cross-feeding-enabled production of butyrate thanks to capturing these fermentation products of primary degraders such as *R. bromii* and *Bifidobacterium* spp. (Cockburn, Suh et al. 2018, Baxter, Schmidt et al. 2019).

Figure 1-5 Conversion of metabolites balances electrons to form the SCFA biosynthetic pathway producing end-products acetate, propionate, and butyrate. Intermediates can become involved in multiple additional pathways meaning one or more end product can be produced, e.g., Acetyl-CoA entering the Ethanol pathway. PEP, phosphoenolpyruvate; DHAP, dihydroxy-acetonephosphate. Figure is adapted from Koh et al. under CC BY licence (Koh, De Vadder et al. 2016).
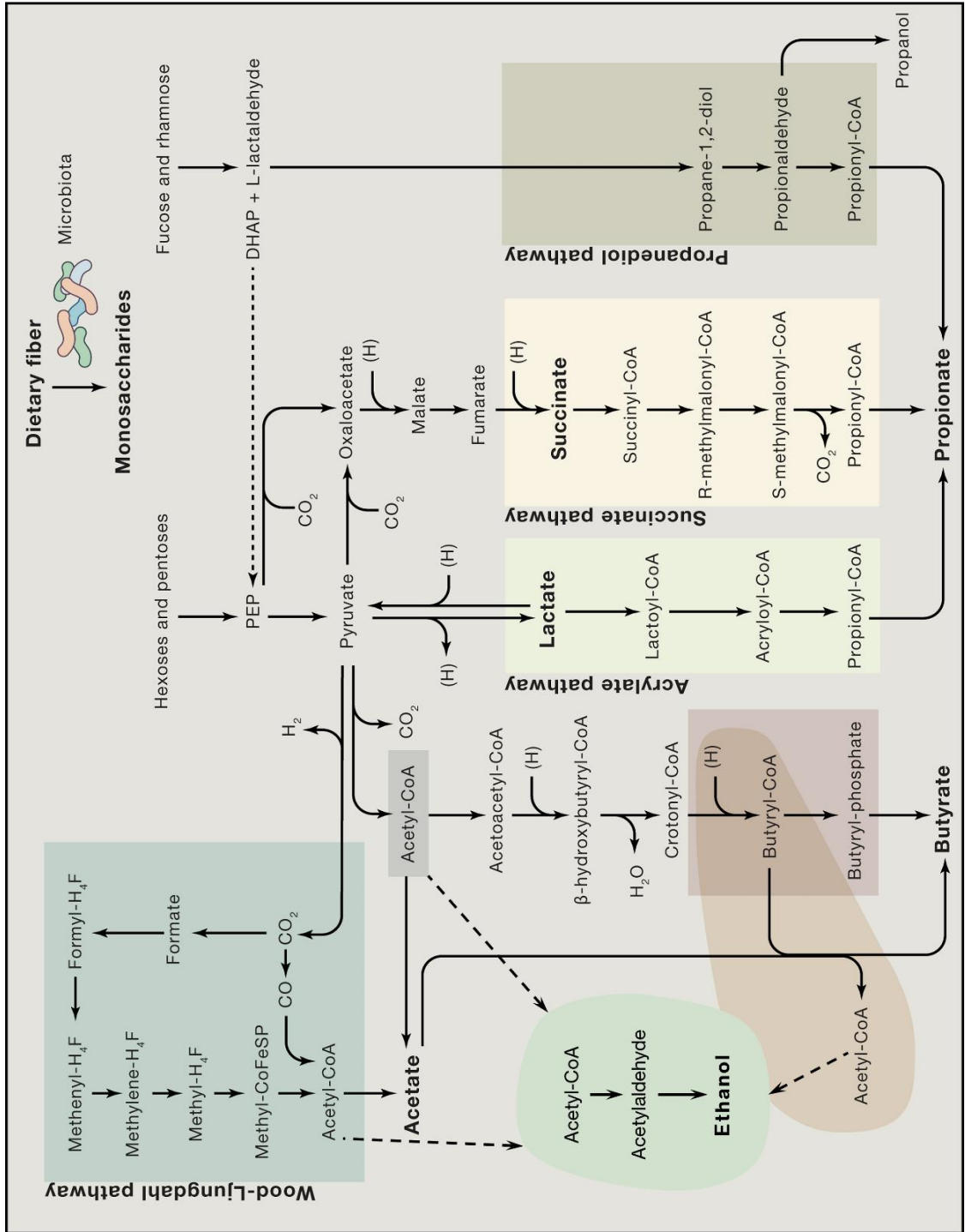
1.5 *Bifidobacterium* and starch

1.5.1 *Bifidobacterium* genus overview

Bifidobacterium is a Gram-positive genus of bacteria in the Actinomycetota (previously Actinobacteria) phylum, that primarily reside in the gastrointestinal tract. Bifidobacteria were discovered in 1899 by Tisser describing an organism called *Bacillus bifidus* derived from the Latin 'bifidus', from bi- 'doubly' + fidus (from findere 'to split') shown in Figure 1-6.
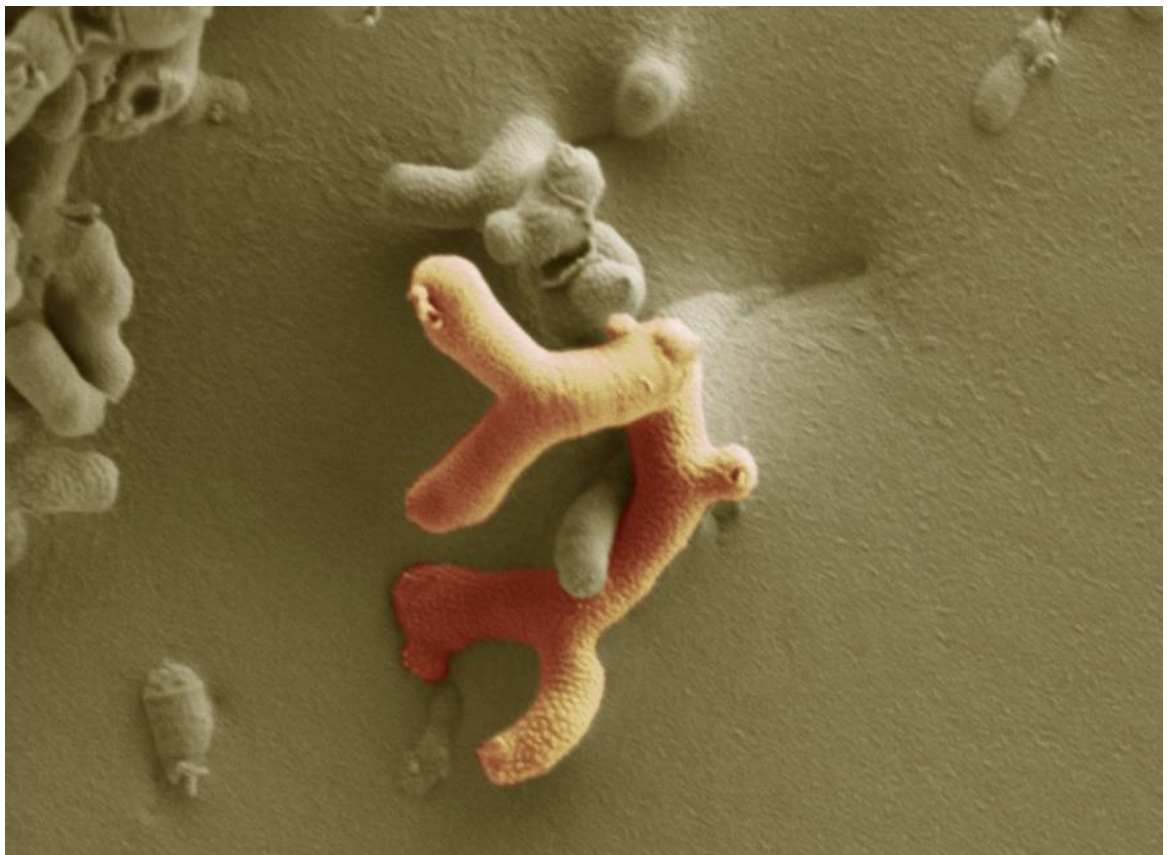
Figure 1-6 Scanning electron microscopy image of Bifidobacterium breve *UCC2003, courtesy of Kathryn Cross, Quadram Institute, Norwich, UK.*

Until the 1960s, *Bifidobacterium* species were collectively referred to as *Lactobacillus bifidum*. The following seven species of human origin were described as the *Bifidobacterium* genus by Reuter in 1963: *Bifidobacterium*

infantis, "*B. liberorum*," (now known as *B. infantis*) "*B. lactensis*," (now known as

B. longum), *B. breve*, "*B. parvulorum*," (now known as *B. adolescentis*), and

B. bifidum (Reuter 1971). Its taxonomy is described as follows: Bacterial

Terrabacteria group; Actinomycetota*;* Actinomycetia; Bifidobacterialies,

Bifidobacteriaceae. Currently there are over 100 recorded validly named species

with a validly published source belonging to the *Bifidobacterium* genus according

to the International Code of Nomenclature of Prokaryotes (ICNP) (Source: <u>Genus:</u>

<u>Bifidobacterium (dsmz.de)</u> [Accessed 21/9/2022]). These 100 species and

subspecies comprise over 400 publicly documented *Bifidobacterium* strains

(Rodriguez and Martiny 2020).

1.5.2 *Bifidobacterium* in infancy

While the adult human microbiota is purported to be relatively stable and

functionally consistent across populations (Lozupone, Stombaugh et al. 2012), the

rapid taxonomic and functional microbiota changes that occurs during early life is

an important research area. Shortly after birth, the GI tract of infants becomes

home to the early life microbiome, which itself is moulded by a variety of factors

such as gestational age at birth, birth mode, probiotic or antibiotic use, and infant

diet (Stewart, Ajami et al. 2018). Important in infant nutrition is the consumption of

breastmilk (where possible) due to the presence of human milk oligosaccharides

(HMOs) which act as health promoting molecules including as prebiotics (Masi and

Stewart 2022). Gut bacteria proliferate by breaking down these HMOs, which are

not digested by infants and instead fermented in the large intestine.

Bifidobacteria are well-known as early colonisers of the human gut in infancy

(Underwood, German et al. 2015, Lawson, O'Neill et al. 2020). In infants

consuming mainly breastmilk, HMOs are almost exclusively and productively broken down by *Bifidobacterium* (Garrido, Dallas et al. 2013, Stewart, Ajami et al. 2018). As discussed, fermentation causes production of beneficial SCFAs including acetate, and this process serves to benefit general gut ecology and has immunomodulatory effects (Gomez de Agüero, Ganal-Vonarburg et al. 2016) which continues into adulthood (Nomura 2022). Supplementation of HMOs to formula milk (which ordinary lacks them) improves the taxonomic composition of infant gut microbiota and increases the relative abundance of bifidobacteria (Vandenplas, Berger et al. 2018). Additionally, in one of the largest early life microbiome studies performed using stool from over 900 children, the intake of breast milk (exclusively or partially) was the most significant driver of microbiome structure. The taxa most correlated to breastmilk intake up to 14 months belonged to *Bifidobacterium* (Stewart, Ajami et al. 2018).

Although exact distributions of bacterial taxa are subject to high inter-individual variability, bifidobacteria dominate the microbiome in infancy (60-70% relative abundance) and decline to less than 10% in adulthood, although this can be much higher in some individuals (Arboleya, Watkins et al. 2016). A 2020 study confirms these figures reporting that bifidobacteria represented between 16.5% and 100% of the microbiota in infants (Kujawska, La Rosa et al. 2020). The complement of species and strains present in the gut microbiome changes with age, largely due to drastic dietary changes from 0-3 years of life (Tsukuda, Yahagi et al. 2021). *B. longum* is considered to be the most common species (although not the most abundant) found in the infant, adult, and elderly gut. *B. longum* subsp*. infantis* detected in infants and *B. longum* subsp*. longum* is widely distributed in both

infants and adults (Table 1-4) (Odamaki, Bottacini et al. 2018). The most common *Bifidobacterium* species during infancy are *B. bifidum*, *B. longum* (including subsp. *infantis*), and *B. breve* (*Table 1-4*) (Arboleya, Watkins et al. 2016, Rodriguez and Martiny 2020). The nuances of these generalisations are demonstrated in one study where *B. breve* accounted for the highest number of taxa in breastfed infants from before 6 months, which was then substituted by *B. bifidum* after 6 months of life (Stewart, Ajami et al. 2018). In this study, the authors considered the gut microbiome up to 12 months as developmental, the second year of life as transitional, and after 3 years of age as stable.

A large-scale bioinformatic investigation of *Bifidobacterium* natural hosts and/or environments used 400 publicly available *Bifidobacterium* strains where 287 strains were of human origin (Rodriguez and Martiny 2020). Additionally, 264 strains (covering 15 species) were isolated specifically from human stool donated by people of varying ages (Table 1-4). *B. breve* was found to be the most commonly isolated species of human-associated *Bifidobacterium*, likely due to its tolerance to lab-culture, but also because it is generally present across all stages of life. Although these data may not comprehensively represent species distributions across ages due to lack of metadata, sampling biases, or sequencing/genomic assembly variances, the findings largely confirm the consensus about age of colonisation of *Bifidobacterium* taxa. While *Bifidobacterium animalis* including subsp*. animalis* was not commonly isolated, they are frequently used as probiotics and in dairy products such as yogurt. Notably, *B. thermophilum* and *B. pseudolongum* are more commonly isolated from animals such as ruminants but are infrequently also isolated from infant faeces

(von Ah, Mozzetti et al. 2007, Vazquez-Gutierrez, Lacroix et al. 2015). These two

species are degraders of highly resistant granular starch which is likely related to

animal diets lacking (hot) cooked starches (Jung and Park 2023).

Table 1-4 An analysis was performed using the 287 publicly reported human faecal Bifidobacterium *strains (Rodriguez and Martiny 2020). The number of strains and their species classification was collated here alongside the associated metadata (age group, milk isolate). The colour-filled boxes on the left-hand column are considered to be typical species of that age group (infant (0-1 years) or adult). Child age was defined as 2-6 years.*

Species	Human faecal isolates	Infant	Child	Adult	Elderly	Human milk isolates
Bifidobacterium breve	76	yes		yes	yes	yes
Bifidobacterium bifidum	34	yes	yes	yes		yes
Bifidobacterium longum subsp. *infantis*	12	yes			yes	
Bifidobacterium animalis subsp. *lactis*	11	yes		yes		yes
Bifidobacterium pseudocatenulatum	13	yes				yes
Bifidobacterium longum	49			yes		yes
Bifidobacterium longum subsp. *longum*	32		yes	yes	yes	
Bifidobacterium adolescentis	23			yes		yes
Bifidobacterium catenulatum	4			yes		
Bifidobacterium gallicum	2			yes		
Bifidobacterium animalis	3				yes	
Bifidobacterium animalis subsp. *animalis*	1		yes			
Bifidobacterium kashiwanohense	2	yes				
Bifidobacterium thermophilum	1	yes				
Bifidobacterium pseudolongum	1	yes				

1.5.3 *Bifidobacterium* during weaning

Weaning is the process of complementing a milk diet with solid foods, which in most cultures includes introducing cereals and starchy vegetables after 6 months of human milk or formula nutrition (Lin and Nichols 2017). As discussed previously, the role of cooked starches in providing energy for weaning infants links to the evolutionary changes under which *Bifidobacterium* adapted to dietary changes of the host. HMOs followed by starch necessitate a flexibility of gut bacteria to utilise different carbohydrate sources. In fact, the underdeveloped upper gastrointestinal tract in humans means that starches are digested less completely, which is suggested to facilitate the maturation of the gut microbiome in weaning infants (Lin and Nichols 2017). The transition of bifidobacterial taxa during the weaning process represents a shift of available carbohydrates of which strains can use at least one of the following: one or many types of HMO; solid food-derived glycans such as starch; host-derived substrates such as mucin; degradation by-products of either HMO or plant glycans. Certain taxa can persist through multiple life stages due to adaptation to changing nutritional landscape of the intestinal milieu through changes in host diet.

By monitoring and strain-resolution sequencing of 75 isolated *B. longum* strains in the first 18 months of infants' lives, overall changes in *Bifidobacterium* species and subspecies present corresponded to the changing infant diet from milk to solid foods (Kujawska, La Rosa et al. 2020). Initial Bifidobacterium relative abundance decreased from 16.5% and 100% of the microbiota in infants to post-weaning between 4.6% and 12.1% of bacterial abundance in the gut. The complement of *Bifidobacterium* present in the gut shifts in response to changes in nutritional

environment, however many taxa span age groups, meaning they must be able to adapt to these changes. *B. longum* including subsp. *infantis* and subsp. *longum* are particularly key to this transitional period (O'Callaghan and van Sinderen 2016, Kujawska, La Rosa et al. 2020, Vatanen, Ang et al. 2022). In fact, *B. longum* persists in the gut microbiota throughout life thanks to its substrate promiscuity (Kujawska, La Rosa et al. 2020).

Kujawska *et al.* (2020) tied together phenotype, genotype, and enzyme production using a wide variety of strains and substrates. The carbohydrate utilisation of *B. longum* (and subsp. *infantis*) corresponded to the changing nutritional intake that occurs when weaning infants from breastmilk or formula to solid foods. These data also show evidence of intra-species variation in the ability to degrade substrates as well as sub-species variation and nutrient-specific adaptation of bifidobacteria during weaning (Kujawska, La Rosa et al. 2020). This was recently corroborated in a larger study of 222 weaning infants where distinct clades of *B. longum* arose during the introduction of solid foods whereby the transitional clade possessed enzymes to utilise both breastmilk and food substrates (Vatanen, Ang et al. 2022).

Shifting enzymatic gears appears to be common during weaning, as *B. longum* subsp. *infantis* was shown to have enzymes to utilise by-products of HMO broken down by other taxa (Lawson, O'Neill et al. 2020), and can additionally shift to other commonly available substrates such as host-derived mucin in the intestinal barrier (Alcon-Giner, Dalby et al. 2020, Duranti, Longhi et al. 2020), the authors noting that these capabilities are strain-specific. A screening was carried out for animal

associated strains of *B. castoris* isolated from wild mice being able to utilise starch and wide variety of different substrates allowing them to switch hosts, indicating a trend in adaptation strategies of the genus (Kujawska, Raulo et al. 2022).

1.5.4 Bifidobacteria as part of the microbial community

The gut microbial community response to RS is interesting and varied due to both vast differences in inter-host microbiota composition and RS structures. In some instances, the presence of bifidobacteria and starch alters microbial ecology, often to the benefit of the host and community (Turroni, Milani et al. 2016, Warren, Fukuma et al. 2018, Baxter, Schmidt et al. 2019).

A dietary intervention investigating the response of bacteria to starch intake showed associations between primary degraders and butyrate producers (Baxter, Schmidt et al. 2019), adding to evidence of bifidobacteria being primary degraders of starch and cooperating with other community members to complete the process to produce SCFAs like butyrate (Belenguer, Duncan et al. 2006, Rios-Covian, Gueimonde et al. 2015, Centanni, Lawley et al. 2018). More recently, an *in vitro* fermentation study using infant faecal inocula (pre- and post-weaning) showed that abundance of *Bifidobacterium* genus members significantly increased in the presence of a high amylose maize starch (Gopalsamy, Mortimer et al. 2019).

It is suggested that the co-existence of many genus members within a single ecosystem is facilitated by their cooperative behaviour (Kujawska, La Rosa et al. 2020, Lawson, O'Neill et al. 2020). Lawson et al. (2020) showed that bifidobacterial species share resources in order to optimise nutrient consumption (mucin, HMO) by carrying out a series of cross-feeding experiments. The spent media from *B. longum* strains grown in 2'-Fucosyllactose (2'-FL) (the most

prevalent HMO) was shown to support growth of the non-HMO user

B. pseudocatenulatum from within the same infant microbial ecosystem (Lawson,

O'Neill et al. 2020). Evidence of cooperation and formation of a pangenome

means the species and strains present collectively perform fermentation of specific

substrates. Strain-specific carbohydrate utilisation genes supports a pangenome

effect, whereby the genus works cooperatively to utilise available dietary

substrates.

1.5.5 *Bifidobacterium* as probiotics

Probiotics were defined by the Food and Agriculture Organization of the United

Nations (FAO) and the WHO as "live microorganisms that, when administered in

adequate amounts, confer a health benefit on the host" (Hill, Guarner et al. 2014).

Although, as the term is highly regulated in some European countries, the term

'live cultures' is used as an alternative. The most common probiotics used belong

to *Bifidobacterium* and *Lactobacillus* genera (O'Callaghan and van Sinderen

2016). *B. animalis* is commonly used in food microbiology applications for

acidifying and antimicrobial properties as well as in probiotics. Several

Bifidobacterium strains belonging to different species are frequently utilised for

treatment of acute diarrhoea facilitated by strain level certification as Generally

Recognised As Safe (GRAS) (Picard, Fioramonti et al. 2005, Di Gioia, Aloisio et al.

2014). These species included but are not limited to *B. bifidum*, *B. breve*,

B. animalis (including subsp. *animalis* and subsp. *lactis*), *B. longum* (including

subsp. *infantis* (Fijan 2014, Odamaki, Horigome et al. 2015). Whilst some benefits

of probiotics are well-documented, inconsistencies in magnitude of the positive

effects are noted as well as factors such as survivability of strains through the

gastrointestinal tract (Ritchie and Romanuk 2012, Sharma, Wasan et al. 2021).

Furthermore, transfer of resistance genes from *Bifidobacterium* probiotics to other

bacteria in the gut are among the safety concerns of live bacterial therapeutics

(Sharma, Wasan et al. 2021).

Increasingly, the strain specificity required to precisely modulate the microbiome is

becoming more apparent (Odamaki, Horigome et al. 2015). Probiotic strains of

Bifidobacterium and *Lactobacillus* have been used in preterm infants as a

preventative treatment for gastrointestinal conditions such as necrotising

enterocolitis (Alcon-Giner, Dalby et al. 2020). This is particularly important since

the preterm infant gut does not naturally resemble the term infant microbiome,

meaning abnormal colonisation patterns can lead to developmental issues or

infection. Studies consistently showed that probiotic administration in pre-term

infants significantly impacted the infant microbiome in a strain-dependent manner

(Beck, Masi et al. 2022), was positively correlated with faecal acetate

concentrations (Alcon-Giner, Dalby et al. 2020), and results in lower abundance of

potential pathobionts (AlFaleh and Anabrees 2014). Additionally, probiotics as

adjuvant therapy has been shown to beneficially influence vaccine response in

some cohorts of infants (Vitetta, Saltzman et al. 2017). The use of probiotics in

early life therefore aids in shaping of a health microbiome and development of life-

long health (Oliphant and Claud 2022).

1.5.6 *Bifidobacterium* starch utilisation

After 3 years of age, the microbiome is relatively stable in its composition and

function. Compared to a mono diet of breast milk and a handful of infant suitable

soft foods, a more diverse diet leads to higher microbial diversity (Stewart, Ajami

et al. 2018, Vatanen, Ang et al. 2022). As the relative abundance of

Bifidobacterium declines to less than 10% of the mature gut microbiota, species

present are commonly specialised to host food sources such as fermentable fibre.

Common *Bifidobacterium* species found in adults include *B. longum* subsp*.*

longum, B. adolescentis, and *B. pseudocatenulatum* (Derrien, Turroni et al. 2022).

Starch utilisation by the *Bifidobacterium* genus was first thoroughly investigated

with a screening approach which investigated 16 species and 42 strains for starch,

amylopectin, and pullulan utilisation. Pullulan, commonly a product of starch

hydrolysis, is a linear polymer consisting of maltotriose (three glucose residues

linked by $\alpha(1 \rightarrow 4)$ glycosidic bonding) repeating units each connected by $\alpha(1 \rightarrow 6)$

linkages (Cheng, Demirci et al. 2011). There were 19 strains belonging to

B. adolescentis, *B. breve, B. dentium, B. globosum, B. longum* subsp. *infantis, B.*

longum, B. magnum, B. pseudolongum, B. pseudolongum subsp*. globosum, and*

B. thermophilum which were able to degrade starch (Ryan, Fitzgerald et al. 2006).

Six *B. breve* strains and one of each *B. pseudolongum, B. pseudolongum* subsp*.*

globosum, B. thermophilum, and *B. longum* subsp. *infantis* could utilise all three

substrates. Similarly, growth and binding assays of *B. adolescentis* IVS-1 showed

significant utilisation of three different Type 4 RS of different botanical origin

(maize, potato, tapioca) (Deehan, Yang et al. 2020).

Some *B. adolescentis* strains are considered as primary degraders proposed as

cross-feeders of butyrate producers (Belenguer, Duncan et al. 2006, Baxter,

Schmidt et al. 2019). Genomic analysis of *B. adolescentis* 22L revealed a nutrient

acquisition strategy that targets plant-derived glycans, in particular starch and

starch-like carbohydrates (Duranti, Turroni et al. 2014). Starch degradation by the

Bifidobacterium genus is well reviewed, but findings are non-exhaustive as many

more strains are yet to be tested, especially as not all species are tolerant to lab

culture and testing of starch utilisation is non-standardised especially in regard to

preparation of starch. Some key findings are summarised in Table 1-5.

Table 1-5 A summary of key findings regarding Bifidobacterium *starch utilisation*

Bifidobacterium species or strains	**Main findings**	**Reference**
B. adolescentis spp. *B. angulatum* *B. bifidum* *B. breve* spp. *B. longum* spp. *B. pseudocatenulatum*	- 2 *B. adolescentis* strains were able to utilise potato starch - *B. bifidum* and *B. pseudocantenulatum* found to utilise starch, but to a lesser degree	(Belenguer, Duncan et al. 2006)
B. breve UCC2003 *B. breve* NCFB2258 *B. pseudolongum* DSM20095 *B. thermophilum* JCM7027	- Strains contain conserved gene regions predicted to be attributed to amylase and pullulanase	(Ryan, Fitzgerald et al. 2006)
B. breve UCC 2003	- Identified putative ApuB amylopullulanase	(Ryan, Fitzgerald et al. 2006)
B. breve UCC 2003	- the enzymatic activity of the purified product of apuB was classified as an amylopullulanase, the first to be identified in the genus. - confirmed by knock-out in *B. breve* UCC2003	(Connell Motherway, Fitzgerald et al. 2008)
B. adolescentis 22L	- Expression of a set of chromosomal loci responsible for starch metabolism as well as for pilus production - extracellular structures include sortase-dependent type IVb pili	(Duranti, Turroni et al. 2014)
B. adolescentis pan-genome (18 strains)	- genetic variation between strains combined with a larger repertoire of genes (relative to *B. breve* or *B. bifidum*) capable of degrading a wide variety diet-derived glycans indicates this species has evolved to be more adaptable to varied adult diet	(Duranti, Milani et al. 2016)

B. adolescentis P2P3 co-culture with *B. thetaiotaomicron* ATCC 29148	- *B. adolescentis* P2P3 showed very strong RS granule utilisation activity - can attach to RS granules and form them into clusters, including high-amylose corn starch granules - Co-culture with *B. thetaiotaomicron* ATCC 29148, was able to grow using carbon sources generated from RS granules by *B. adolescentis* P2P3	(Jung, Kim et al. 2019)
204 strains of bifidobacteria	- soluble wheat starch was utilised by 114 strains belonging to 37 species and subspecies	(Modrackova, Makovska et al. 2019)
Bifidobacteriaceae	- Metaproteomic responses of in vitro gut microbiomes to resistant starches - *In vitro* culture with RS showed *Bifidobacteriaceae* was significantly - Metaproteomics indicated *Bifidobacteriaceae* mainly contributed to starch and sucrose metabolism and pentose phosphate metabolism pathways	(Li, Loponen et al. 2020)
B. pseudocatenulatum *B. longum* *B. breve* *B. longum* subsp. *infantis*	- total of 39 different GH families were found in all *Bifidobacterium* strains isolated - 62 GH genes in *B. pseudocatenulatum* strains - 48 GH genes per *B. longum* genome - 46 GH genes per *B. breve* genome - 42 GH genes per *B. longum* subsp. *infantis* genome	(Lawson, O'Neill et al. 2020)
B. catenulatum KCTC 3221 *B. bifidum* BGN4 *B. longum* RD72, *B. animalis* ATCC 25557 *B. animalis* subsp. *lactis* RD68 *B. animalis* subsp. *lactis* SH5 *B. animalis* subsp. *lactis* ad011	- Co-culture with *Bifidobacterium catenulatum* improves the growth and colonisation and butyrate production of *Faecalibacterium prausnitzii*	(Kim, Jeong et al. 2020)

1.5.7 *Bifidobacterium* amylolytic enzymes

B. pseudocatenulatum is commonly found in infants and adults. Analysis of infant

gut derived strains showed that the predominant GH family in all 19 strains of

B. pseudocatenulatum was GH13, more specifically subfamilies GH13_4, GH13_13, GH13_15, GH13_28, all of which have α-amylase noted as their primary enzyme activity (Table 1-1). Further comparative genomic analyses of 49 strains of *B. longum, B. breve,* and *B. animalis* showed that *B. longum* had a much wider range of GH family representation, consistent with their ability to occupy the microbiota of multiple age groups (Odamaki, Horigome et al. 2015). CAZyme families present show that they are able to adapt to different carbohydrate sources including starch (GH13), xylan and arabinose (GH43 and GH51), suggesting that they are more 'generalist' than other species. In contrast to 'specialist' bacteria which prefer a small number of substrates, generalist bacteria use a broader array of substrates for growth and energy (Bell and Bell 2021). The authors noted significant differences between strains of the same species in membrane carbohydrate transporters sequences, suggesting intra-species specialisation (Odamaki, Horigome et al. 2015).

Five strains of *B. breve* investigated bioinformatically also contained conserved genes consistent with amylopullulanase activity which would suggest an ability to break both types of glycosidic bond contained in starch molecules (Table 1-1) (Ryan, Fitzgerald et al. 2006). The *B. breve* UCC2003 genome was found to contain 26 GH families (55 total CAZymes), 6 CBM families (10 total modules). Of the 26 GH families represented, the GH13 family covers 14 out of 55 total enzymes. Enzyme activity assays also determined that certain enzymes were inducible by the presence of starch and produced extracellularly. The amylopullulanase gene was the first discovered gene operon in *Bifidobacterium* specific to starch degradation. *B. breve* UCC2003 was found to contain a

bifunctional class II amylopullulanase (called ApuB) that is necessary to metabolise starch. Interestingly, this strain also contains several genes needed for proficiency in utilised HMOs explaining why it persists through infancy and adulthood (Connell Motherway, Fitzgerald et al. 2008, James, O'Connell Motherway et al. 2018).

ApuB is a multi-modular protein containing amylase and pullulan catalytic domains co-localised with starch binding CBM domains, resulting in an enzyme which binds starch and can break both α(1→4) and α(1→6) glycosidic bonds (Ryan, Fitzgerald et al. 2006). Functional amylase enzymes have 3 main domains: alpha/beta barrel active site, Ca binding site (conserved as essential for stability), and a C terminal domain variable region. Within the ApuB operon, 2 active sites, Ca binding domain, and Aamy_C domain (alpha-amylase C-terminal domain) were detected (Figure 1-7).
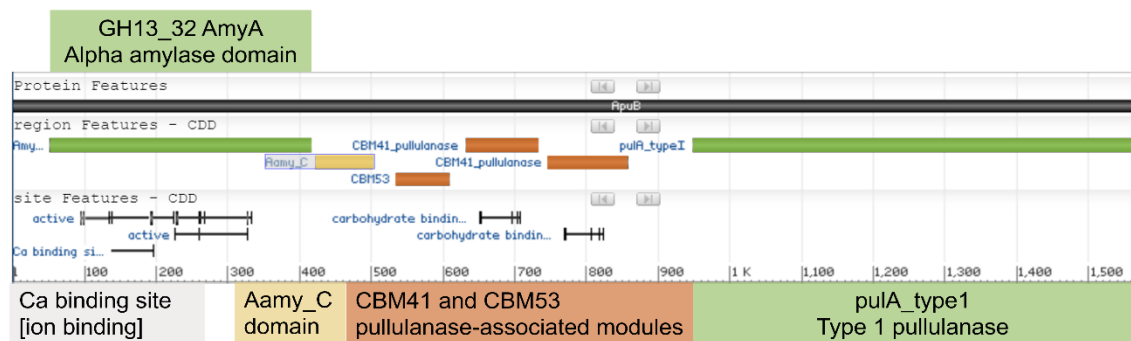
Figure 1-7 The ApuB gene cluster in B. breve *UCC2003 contains a bifunctional amylopullulanase enzyme which is capable of breaking both types of glycosidic bond contained in starch.*

The presence of alpha-amylase and type I pullulanase enzymes which cleave α(1→4) and α(1→6) glycosidic bonds respectively means it can hydrolyse linear chains and debranch starch i.e. a bifunctional enzyme. In addition, CBM53 and

CBM41 from pullulanase-like enzymes at the same gene locus implies substrate-binding capabilities of the final gene product. The ApuB operon was confirmed by knock-out to be essential for starch degradation (Connell Motherway, Fitzgerald et al. 2008). A similar enzyme gene cluster is found in *B. bifidum, B. dentium, B. adolescentis, B. pseudocatenulatum, B. animalis* subsp*. lactis, and B. thermophilum* (www.ncbi.nlm.nih.gov/protein/?term=ApuB [Accessed 19/01/22]). Multidomain amylolytic polypeptides are unique to RS degrading organisms and are viewed as essential for amylolytic phenotypes (Cerqueira 2022).

1.5.8 Starch binding modules in *Bifidobacterium*

Bacterial adhesion mechanisms in *Bifidobacterium* species are sometimes employed but are not necessary for degradation of substrates (Crittenden, Laitila et al. 2001). Co-localisation of the bacteria to the surface of the substrate followed by formation of pores allowing bacteria to permeate inwards has been demonstrated in *R. bromii,* but not *B. adolescentis,* indicating differences in starch degradation patterns or behaviour depending on starch structure (Koev, Harris et al. 2022). More prominently studied in other amylolytic species of gut bacteria are the specific CBMs which commonly associate with amylase enzymes produced on the outside of the cell. Of particular interest is CBM74 which specifically binds to RSII, RSIII, soluble potato starch, amylose, and amylopectin. The CBM74 family is found almost exclusively in *Bifidobacterium*: 71 gut microbes encode CBM74 family members, 50 are found in bifidobacteria (Valk, Lammerts van Bueren et al. 2016, Cerqueira, Photenhauer et al. 2020). There are few reports of bifidobacteria which contain CBM74 domain containing amylases investigating their ability to

enhance RS binding or degradation, including one recent study utilising a recombinantly expressed CBM74-amylase protein from *B. adolescentis* P2P3 showing degradative activity on high-amylose corn starch granules (Jung, Seo et al. 2020). Hybrid metagenome assemblies confirm the upregulation of CBM74 sequences in response to RS being present in an *in vitro* fermentation model (Ravi, Troncoso-Rey et al. 2022).

CBM74 is a relatively novel RS binding module which commonly associates with multi-domain alpha amylases which also contain flanking CMB25 or CBM26 modules and are suggested to work synergistically to dock to starch granules (Valk, Lammerts van Bueren et al. 2016, Cerqueira, Photenhauer et al. 2020). The protein structure of a CBM74 has only recently been solved revealing a unique topological arrangement which has been hypothesised to allow efficacious binding. In *R. bromii*, CBM26 was shown to bind short malto-oligosaccharides while the CBM74 module bound to single and double helical starch (Photenhauer, Cerqueira et al. 2022).

1.5.9 *Bifidobacterium* import and cell metabolism of carbohydrates

After bacteria have bound to and begun hydrolysing a complex substrate such as starch, there are specialised systems which import the lower molecular weight products, e.g. mono-, di- and oligo-saccharides, where they are met by a new set of enzymes for further processing inside the cell. Import mechanisms in bacteria are facilitated by two main systems: ABC-type (ATP-binding cassette) transporters or major facilitator superfamily (MFS) transport systems PEP-PTS systems (Turroni, Strati et al. 2012, O'Callaghan and van Sinderen 2016).

Bifidobacteria have a few characteristic mechanisms of importing and metabolising carbohydrates. Firstly, unlike other bacteria which utilise both systems, bifidobacteria almost exclusively use ABC family transporters (Pokusaeva, Fitzgerald et al. 2011, Turroni, Strati et al. 2012). Thus, approximately half of the average *Bifidobacterium* genome is dedicated to carbohydrate uptake through ABC systems (Pokusaeva, Fitzgerald et al. 2011). ABC transporters capture and import oligo- and mono-saccharide sugars, after which they enter metabolic pathways to produce ATP (Foley, Cockburn et al. 2016). Sugars can enter a pathway unique to *Bifidobacterium* called the F6PK pathway, which contains primary enzyme fructose-6-phosphoketolase (Figure 1-8) (De Vries and Stouthamer 1967).

The arrangement of genes involved is named the 'bifid shunt' and is considered to be a key characteristic of the genus as well as implicated in its success in the complex environment of the human gut microbiota. This is due to its high efficiency of energy turnover of sugars which are fermented into lactate and acetate, and less efficiently into ethanol and formate (Figure 1-8). Hexose sugars yield 2.5 mol ATP from 1 mol of sugar, as well as 1.5 mol acetate and 1 mol of lactate (Palframan, Gibson et al. 2003, Kelly, Munoz-Munoz et al. 2021). Hexose sugars enter the F6PK pathway as fructose-6-phosphate and pentose as ribulose-5-phosphate or xylulose-5-phosphate (Egan and Van Sinderen 2018). Pentose (5 carbon) sugars are pre-processed into ribulose-5-phosphate, which was described in *B. breve* UCC2003 using a gene cluster (rbsACBDK) induced by the presence of ribose, which is a commonly found pentose in the gut (Pokusaeva, Neves et al. 2010).

Pentoses can enter the F6PK pathway yielding 1 mol acetate and 1 mol of lactate. Bifidobacteria also only product acetate under certain conditions, meaning the ratios of lactate and acetate produced can vary (Devika and Raman 2019, Kelly, Munoz-Munoz et al. 2021). The bifid shunt genetic signature was used in metaproteomic methodology to show an enrichment of *B. pseudolongum* in the presence of a high fibre diet in mice, including fibre degradation enzymes and the bifid shunt pathway (Killinger, Whidbey et al. 2022).
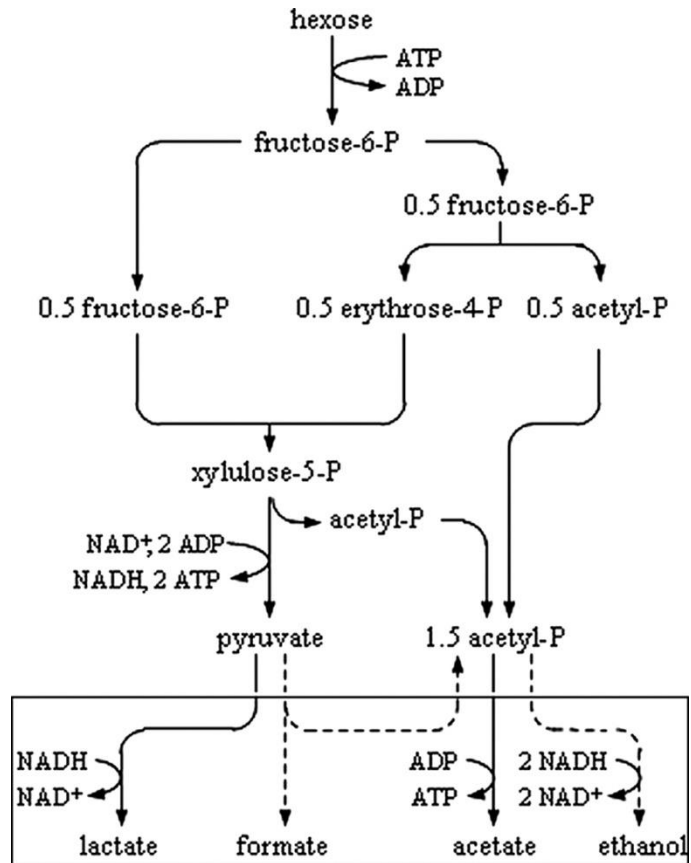
Figure 1-8 The bifid shunt represents a sugar fermentation metabolic pathway called the F6PK pathway present only in Bifidobacterium. *Figure used with permission of American Society for Microbiology, from "Kinetics and metabolism of* Bifidobacterium adolescentis *MB 239 growing on glucose, galactose, lactose, and galactooligosaccharides", Amaretti et al., Vol.73, (2007); permission conveyed through Copyright Clearance Center, Inc.*

1.6 Outlook

Bifidobacterium is a genus able to pivot their collective function by moderating the abundance of members able to ferment HMO, starch, or both. To fully explore the relationship between *Bifidobacterium* and starch, more discussion is needed about

the evolutionary pressures as humans transition their diet from infancy into adulthood. The *Bifidobacterium* genus utilises a cooperative strategy to dominate the infant microbiome which is particularly necessary since there are over 200 structurally different oligosaccharides in breastmilk (Masi and Stewart 2022). It could be hypothesised this cooperative process is only possible due to the relatively low levels of competition experienced in infancy when *Bifidobacterium* is most prevalent and microbial diversity is low. In order to continue competing for an ecological niche, *Bifidobacterium* pivot their metabolism to degrade starch structures during weaning onto solid foods. However, the starch degradation mechanisms of *Bifidobacterium* are understudied.

While the introduction of starch during weaning stages is advantageous for infant health, *Bifidobacterium* likely coordinate the transition of weaning infants' microbiota (Wang, Mortimer et al. 2019). Probiotic supplementation of bifidobacteria is effective in certain clinical presentations, although the engraftment or long-term colonisation of supplemented *Bifidobacterium* remains a challenge (Neumann, Mahnert et al. 2023). Some proof-of-concept synbiotic studies show that precision combinations of strain and a substrate can improve engraftment of a probiotic without the need for antibiotic use (Button, Autran et al. 2022). This has been shown in infants where the risks of synbiotics are low compared to antibiotic use, the main problem being the costs associated with human milk oligosaccharides (Button, Autran et al. 2022). There is growing interest in RS as an easy-to-implement 'missing ingredient' when formulating live biotherapeutic products, especially as it cross-feeds many other beneficial taxa (Gopalsamy, Mortimer et al. 2019, Liang, Xie et al. 2023)

The impacts of starch composition and structure on microbial response in adults are still being clarified. Progress in precision microbiome modulation is aided by the functional pliability of starch. For example, the amylose content, type of crystallinity of starch, or chemical modifications can define specific ecosystem responses (Warren, Fukuma et al. 2018, Deehan, Yang et al. 2020, Dobranowski and Stintzi 2021). The advantages of precision microbiome modulation include potentially triggering physiologically beneficial responses during processes such as appetite regulation, insulin control, or improvement of IBD symptoms. However, the exact interactions between dietary substrate and fermentative bacteria need to be better studied in order to mechanistically understand the outcomes.

1.7 Conclusions

The ability of *Bifidobacterium* spp. to degrade a variety of substrates provides avenues to boost *Bifidobacterium* populations in the human gut microbiota. The structural variability of starch offers opportunities to deliver discrete fibre types to the colon which can induce health-promoting outcomes (Baxter, Schmidt et al. 2019, Deehan, Yang et al. 2020). This is particularly the case for adult nutrition where starch is present ubiquitously across many food types and cultures (Cassidy, Bingham et al. 1994, Jones, Hunt et al. 2011).

Bifidobacterium have a broad preference for starch as a substrate. Supplementation of complementary RS and specific strains could promote engraftment in either the infant or adult microbiome. For these reasons, more knowledge of the molecular underpinnings of *Bifidobacterium* starch fermentation

is needed as well as the ecological impacts of starch fermentation by bifidobacteria.

1.8 Thesis Aims and Objectives

This thesis aims to investigate the interaction between *Bifidobacterium* and starch. Of particular focus is the development of new knowledge of the biomolecular underpinnings and the potential impacts of this knowledge on combinations of the two on microbial ecology and beneficial metabolite production.

This thesis aims to expand on knowledge of interaction between starch and selected *Bifidobacterium* strains known to utilise starch available from culture collections. Additional unique isolates derived from infant stool donations were selected to assess the starch-degrading capabilities of infant bifidobacteria. The following work aims to understand the genetic basis of utilisation as well as the combinations of genes or clusters of genes relevant to enzymatic breakdown. It is postulated that different starch active CAZymes and CBMs work synergistically to dock to and degrade starch. While targeted gene knockouts of these modules would confirm the essential roles these proteins have during RS degradation, this thesis explores the selection and investigation of two strains of *B. pseudolongum* where one had a diminished capacity to degrade high-amylose starch.

RS reaching the human colon causes amylolytic bacteria to ferment starch into SCFAs benefitting the microbial community and host health through mechanisms previously discussed such as pH lowering, immune modulation, and hormone signalling. Therefore, this thesis explores the spectrum of SCFAs produced by bifidobacteria metabolising different types of starch. Quantitative NMR

metabolomics in bacterial isolate cultures and *in vitro* batch fermentation systems attempt to understand the potential health impacts of this interaction. A combination of supplementary probiotic bifidobacteria and starch introduced to the microbiome could have a series of effects: enhance proliferation of the *Bifidobacterium* strain introduced; cross-feed other beneficial taxa and enhance their abundance; positively impact bacterial diversity; lead to an increase in production of SCFA. The final stage of work investigates the ecology of starch and *Bifidobacterium* combinations.

This thesis seeks to investigate the effects of the interaction between *Bifidobacterium* and starch with the following objectives. Utilise a variety of different starches (potato, maize, high amylose varieties etc), some of which in an autoclaved or granular form, as a substrate for reference and unique strains of bifidobacteria and assess their growth and metabolite production (Chapter 3). Bioinformatically investigate the genomes of strains for genes annotated as starch-degrading or starch-binding (Chapter 3). Transcriptomic analysis of two strains of *B. pseudolongum* which exhibit starch structure specific variation in starch utilisation, as part of a multi-omics experiment (transcriptome, proteome, metabolome, genome) to interrogate the biomolecular mechanisms of the interaction (Chapter 4). And finally, assess the impacts of tailored *Bifidobacterium* and starch combinations using *in vitro* batch fermentation to simulate a human gut microbial ecosystem (Chapter 5).

Chapter 2

General Materials and Methods

2.1 General materials

The water used throughout (unless otherwise stated) was Avidity Science Triple

Red deionised, ultrapure water. Ethanol absolute (≥99.8%) used was purchased

from VWR International, Poole, UK (Cat. No. 20821.330).

2.1.1 Media and substrates

General media and reagents used to prepare nutrient broth or agar for bacterial

culturing are detailed in Table 2-1. Additionally, substrates such as individual

sugars and starches are also listed.

Table 2-1 Media and substrate details used for general bacterial culturing.

General Media	Source	Cat. No.
Difco™ Lactobacilli deMan, Rogosa and Sharpe (MRS) Broth	Scientific Laboratory Supplies	288130
Reinforced Clostridial Medium (RCM)	Thermo Fisher	CM0149B
Brain heart infusion broth	Thermo Fisher	CM1135
Additional media ingredients	**Source**	**Cat. No.**
L-Cysteine hydrochloride anhydrous, ≥98%	Sigma-Aldrich	C1276
Sodium Hydroxide (NaOH) 98-100.5%, pellets	Sigma-Aldrich	06203
Glycerol, ≥99%	Thermo Fisher	G/0650/08
Oxoid™ Agar	Thermo Fisher	LP0011B
Modified MRS	**Source**	**Cat. No.**
Difco™ Tryptone Peptone	Thermo Fisher	211921
Oxoid™ Yeast extract	Sigma-Aldrich	LP0021
Bacto™ Tryptose	Thermo Fisher	211713
Dipotassium Phosphate, ≥98% (K_2HPO_4)	Sigma-Aldrich	P3786

Monopotassium Phosphate (KH_2PO_4), ≥99.0%	Sigma-Aldrich	P5655
Tri-ammonium citrate, ≥97%	Sigma-Aldrich	A1332
Pyruvic acid, 98%	Sigma-Aldrich	107360
L-Cysteine hydrochloride anhydrous, ≥98%	Sigma-Aldrich	C1276
Tween® 80	Sigma-Aldrich	P1754
Magnesium sulfate heptahydrate, ReagentPlus® ≥99.0% ($MgSO_4 \cdot 7H_2O$)	Sigma-Aldrich	M1880
Manganese sulfate tetrahydrate, 99% ($MnSO_4 \cdot 4H_2O$)	Thermo Fisher	B22081
Iron (II) sulfate heptahydrate ($FeSO_4 \cdot 7H_2O$), ACS reagent ≥99.0%	Sigma-Aldrich	215422
Carbohydrate substrates	**Source**	**Cat. No.**
Starch from corn (Unmodified regular corn starch containing approx. 73% amylopectin and 27% amylose)	Sigma-Aldrich	S4126
Starch from corn (practical grade high-amylose corn starch)	Sigma-Aldrich	S4180
Hylon VII® maize	Gifted by Ingredion	03451B00
Starch from potato	Sigma-Aldrich	S4251
D-(+)-Glucose, ≥99.5%	Sigma-Aldrich	G7021
Lactose Monohydrate, Extra Pure	Thermo Fisher	L/0200/60
D-(+)-Maltose monohydrate from potato, ≥99%	Sigma-Aldrich	M5885
Scientific Laboratory Supplies, Nottingham, UK; Thermo Fisher Scientific™, Loughborough, UK; Sigma-Aldrich, Dorset, UK; Ingredion Inc, Manchester, UK.		

2.1.2 Bacterial strains

Reference strains are publicly available from recognised culture collections for laboratory testing (Kyrpides, Hugenholtz et al. 2014). Eight reference strains were purchased from the National Collection of Industrial, Food and Marine Bacteria (NCIMB) as lyophilised ampules, 4 purchased directly from the supplier and 4 were kindly provided by Hall Lab as detailed in Section 2.1.2. The complete genome assemblies for reference strains were accessed and downloaded from the

NCBI Assembly public database and from existing WGS from Hall Lab as detailed in Table 2-2.

A total of 13 bacterial strains kindly provided by Hall Lab were acquired as unique isolates from infant stool from the Baby-Associated MicroBiota of the Intestine (BAMBI) study, detailed in Table 2-3. For 10 of these isolates, associated metadata and research was published by Lawson et al. (2020), for which faeces were collected from healthy, full-term breast-fed infants in accordance with protocols laid out by the National Research Ethics Service approved UEA/QIB Biorepository (Licence no: 11208) and Quadram Institute Bioscience Ethics Committee. Three infants (V1, V2, V3) donated at a similar age (mean = 145 ± 38d) enabling isolation of 19 unique isolates. Mother V3 took antibiotics (0-14 days) and probiotics (3 months before birth and 6 months after); mothers V1 and V2 did not take any antibiotics or probiotics during infant sampling. None of the infants were administered antibiotics or probiotics during faecal sampling. Strains isolated from the same infant at the same faecal donation time point are considered clonal.

Table 2-2 *A full list of all bacterial strains used in this thesis, including the reference strains purchased from culture collections and those which are unique isolates. Where possible, the original publication and accession number are detailed. *Species classification is unresolved.*

Genus	Species	Strain ID / Cat. No.	Description	Origin	Reference	GenBank accession number
Bifidobacterium	*breve*	UCC2003 / NCIMB 8807	Reference strain	Breast-fed infant faeces	(Connell Motherway, Fitzgerald et al. 2008)	GCA_000220135.1
Lacticaseibacillus	*casei* subsp. *casei*	NCIMB 4114	Reference strain	Fermented ketchup	(Toh, Oshima et al. 2013)	GCA_000829055.1
Bifidobacterium	*breve*	NCIMB 702258 / NCFB 2258	Reference / type strain	Infant faeces	(O'Riordan and Fitzgerald 1999)	GCA_000569035.1
Bifidobacterium	*pseudolongum* subsp. *pseudolongum*	NCIMB 702244 / DSM 20099	Reference / type strain	Pig faeces	This thesis	GCA_027945455.1
Bifidobacterium	*pseudolongum* subsp. *globosum* / *pseudolongum**	NCIMB 702245 / DSM 20092	Reference strain	Bovine rumen	Park et al. (2016) [Unpublished]	GCA_002706665.1
Bifidobacterium	*longum* subsp. *longum*	NCIMB 8809	Reference strain	Breast-fed infant faeces	(O'Callaghan, Bottacini et al. 2015)	GCA_001446255.1
Bifidobacterium	*adolescentis*	NCIMB 702204 / ATCC15703 / DSM20083	Reference strain	Human adult intestine	Suzuki et al. (2006) [Unpublished]	GCA_000010425.1
Bifidobacterium	*longum* subsp. *infantis*	ATCC 15697 / DSM 20088	Reference strain	Infant intestine	(Sela, Chapman et al. 2008)	GCA_000020425.1

Bifidobacterium	*breve*	LH21	Unique isolate	Breast-fed infant faeces	(Lawson, O'Neill et al. 2020)	N/A
Bifidobacterium	*breve*	LH24	Unique isolate	Breast-fed infant faeces	(Lawson, O'Neill et al. 2020)	N/A
Bifidobacterium	*pseudocatenulatum*	LH662 LH663 LH656 LH657 LH658 LH659	Unique isolates; clonal	Breast-fed infant faeces	(Lawson, O'Neill et al. 2020)	GCA_902505405.1 GCA_902505395.1 GCA_913066385.1 GCA_902505415.1 GCA_913248015.1 GCA_902505425.1
Bifidobacterium	*pseudocatenulatum*	LH9	Unique isolate	Breast-fed infant faeces	(Lawson, O'Neill et al. 2020)	N/A
Bifidobacterium	*longum* subsp. *infantis*	LH277	Unique isolate	Breast-fed infant faeces	(Lawson, O'Neill et al. 2020)	N/A
Bifidobacterium	*bifidum*	LH80	Unique isolate	Premature infant faeces	This thesis	N/A
Bifidobacterium	*breve*	LH36 LH37	Unique isolates; clonal	Breast-fed infant faeces	This thesis	N/A

Infant ID	Gender	Birth mode	Diet	Age at sampling (days)	Faecal isolates	Strain
V1	Male	Vaginal	Breast milk	102	LH9	*B. pseudocatenulatum*
					LH11	*B. pseudocatenulatum*
					LH12	*B. longum* subsp. *longum*
					LH13	*B. pseudocatenulatum*
					LH14	*B. pseudocatenulatum*
V2	Female	Vaginal	Breast milk	174	LH21	*B. breve*
					LH23	*B. longum* subsp. *infantis*
					LH24	*B. breve*
V3	Female	Vaginal	Breast milk	37	LH277	*B. longum* subsp. *infantis*
				159	LH656	*B. pseudocatenulatum*
					LH657	*B. pseudocatenulatum*
					LH658	*B. pseudocatenulatum*
					LH659	*B. pseudocatenulatum*
					LH662	*B. pseudocatenulatum*
					LH663	*B. pseudocatenulatum*

Table 2-3 Metadata of infants from which unique faecal strains (LH strains) were isolated. (Lawson, O'Neill et al. 2020)

2.1.3 Sequencing materials

Various materials were used to perform the extraction, quality assessment,

quantification, and sequencing of nucleic acids, the details of which are detailed in

Table 2-4.

Table 2-4 A comprehensive list of each kit or reagent used for the purposes of preparing nucleic acid for sequencing.

Materials, kits, and reagents	Purpose	Source	
General		**Supplier**	**Cat. No.**
Invitrogen™ RT-PCR Grade Water	Dilutions	Thermo Fisher	AM9935
Ethanol absolute (≥99.8%, AnalaR NORMAPUR® ACS	Decontamination; solvent	VWR International	20821.330
Invitrogen™ RNase AWAY™ Decontamination Reagent	Decontamination	Thermo Fisher	10328011
Oxoid™ Phosphate Buffered Saline (PBS) Tablets	Cell wash	Thermo Fisher	BR0014G
Nucleic acid extraction		**Supplier**	**Cat. No.**
FastDNA™ SPIN Kit for Soil	Bacterial isolate DNA extraction	Thermo Fisher	116560-2000
Maxwell RSC Purefood GMO and Authentication Kit	Metagenomic DNA extraction	Promega	AS1600
RNeasy® Mini extraction kit	Bacterial isolate RNA extraction	Qiagen	74104
MP Biomedicals™ Lysing Matrix E tubes	Cell lysis	Thermo Fisher	MP116914050
β -mercaptoethanol, molecular biology grade	Cell lysis	Sigma-Aldrich	444203
Invitrogen™ RNA*later*™ Stabilisation Solution	Cell RNA preservation	Thermo Fisher	AM7020
Nucleic acid extract quantification		**Supplier**	**Cat. No.**
Qubit™ dsDNA HS and BR Assay Kit	DNA quantification	Thermo Fisher	Q32851
Qubit™ RNA High Sensitivity (HS)	RNA quantification	Thermo Fisher	Q32852
Sample quality control (QC)		**Supplier**	**Cat. No.**
RNA extract QC			

High Sensitivity RNA ScreenTape	RNA integrity check	Agilent	5067- 5579
High Sensitivity RNA ScreenTape Sample Buffer	RNA integrity check	Agilent	5067- 5580
RNA Kit (15NT)	RNA integrity check	Agilent	DNF-471-0500
Library QC			
HS NGS Fragment Kit	RNA library validation	Agilent	DNF-474-0500
Quantifluor ds DNA system	Illumina library quantification	Promega	E2670
D5000 ScreenTape	Illumina library QC	Agilent	5067-5588
D5000 Reagents		Agilent	5067-5589
D5000 Ladder		Agilent	5067-5590
Library preparation		**Supplier**	**Cat. No.**
NEBNext Ultra II RNA Library Prep Kit	RNA transcriptomics	New England Biolabs	
Illumina® DNA Prep, Tagmentation kit (formerly Nextera DNA Flex)	Kit used for Illumina and Nanopore DNA sequencing prep	Illumina	20018704
Illumina DNA sequencing prep (additional reagents)			
Nextera XT Index Kit v2 index primers	Molecular sequencing labels	Illumina	FC-131-2001 to 2004
KAPA 2G Fast Hot Start Ready Mix	DNA polymerase	Merck	KK5601
Oxford Nanopore DNA sequencing prep (additional reagents)			
LongAmp® Taq 2X Master Mix	DNA polymerase	New England Biolabs	M0287L
rRNA depletion		**Supplier**	**Cat. No.**
NEBNext rRNA Depletion Kit (Bacteria)	Ribosomal RNA removal	New England Biolabs	E7850S
Sequencing		**Supplier**	**Cat. No.**
Ligation Sequencing Kit	Genomic ligation protocol (Nanopore sequencing)	ONT	SQK-LSK109

NSQ® 500 Mid Output Flowcell (KT v2(300 CYS))	Illumina Nextseq500 sequencing	Illumina	FC-404-2003
% PhiX spike in (PhiX Control v3)			FC-110-3001
NovaSeq 6000 Reagent Kits v1.5	Illumina NovaSeq 6000 sequencing	Illumina	20028312
Agilent Technologies Ltd, Cheadle, UK; New England Biolabs, Ipswich, USA; Oxford Nanopore Technologies (ONT), Oxford, UK; Merck, Poole, UK; Illumina, Cambridge, UK; Promega, Southampton, UK; Thermo Fisher Scientific™, Loughborough, UK, Sigma-Aldrich, Dorset, UK; Qiagen, Manchester, UK			

2.1.4 Proteomics materials

The materials used to perform the extraction and quantification of bacterial

proteins are detailed in Table 2-5.

Table 2-5 Proteomics materials and reagents

Reagent	Source	Cat. No.
Tris buffer	Sigma-Aldrich	GE17-1321-01
Trifluoroacetic acid (TFA) 100%	Sigma-Aldrich	299537
Oxoid™ Phosphate Buffered Saline (PBS) Tablets	Thermo Fisher Scientific	BR0014G
Bradford Solution (Coomassie brilliant blue)	Sigma-Aldrich	B6916
Bovine serum albumin (BSA), lyophilised powder, ≥96%	Sigma-Aldrich	A2153

2.2 Methods

2.2.1 Bacterial Isolate Studies: Starch Growth Assays

2.2.1.1 Bacterial Isolate Studies media preparation

Difco™ MRS or Thermo Scientific™ RCM were used to culture bacteria from frozen 30% glycerol stocks and to maintain live cultures. 500 mL RCM liquid broth was prepared by adding 19 g RCM powder to a 1 L Duran bottle and thoroughly mixing with 500 mL demineralised water. MRS medium was prepared by adding 27.5 g MRS powder to 500 mL demineralised water. To prepare a solid medium, 15 g Oxoid™ Agar was added per L of media. The bottle was sterilised to 126 °C for 15 min in a Prestige 2100 Classic Autoclave. Per 500 mL media used, 0.5 mL of 50% w/v cysteine solution was added and mixed well by shaking gently. Liquid media were stored at room temperature for a maximum of 2 weeks. Molten agar media were poured onto 90 mm triple vent petri dishes at a volume of 25 mL per plate, or 140 mm petri dishes at a volume of 62 mL per plate, and stored for a maximum of two weeks at 4 °C.

A modified MRS (mMRS) was prepared since standard media typically contain bacterial growth substrates such as starch or glucose. mMRS was used in the absence of any substrate as a negative control. To test the bacterial growth on specific substrates, the medium was supplemented with substrates. mMRS was prepared by combining the following base ingredients in demineralised water: Trypticase peptone (1% w/v), Yeast extract (0.25% w/v), Tryptose (0.3% w/v), Dipotassium phosphate (0.3% w/v), Monopotassium Phosphate (0.3% w/v), Tri-ammonium citrate (0.2% w/v), Pyruvic acid (0.2 mL/L), Tween 80 (1 mL/L).

Magnesium Sulphate heptahydrate (0.0575% w/v) and Manganese Sulphate tetrahydrate (0.012% w/v) solutions were prepared as 1000X stock solutions and sterilised under aseptic conditions within a Class II Walker microbiological safety cabinet; the solution was transferred to a sterile syringe and filter-sterilised through a 0.22 mm filter attached to the tip into a clean, sterile 15 mL plastic centrifuge tube. 1 mL of each was added per 1 L media and subsequently stored at - 20 °C between uses. Iron sulphate heptahydrate (final concentration 0.004% w/v) and Cysteine HCl (final concentration 0.03% w/v) were both prepared freshly prior to use by making a 1000x stock solution and adding 1 mL per L of media.

The final medium was then topped with demineralised water to the total volume required and pH adjusted using a pH meter to 6.8 by adding concentrated NaOH drop-wise with a plastic Pasteur pipette. The medium was sterilised to 126 °C for 15 min in a Prestige 2100 Classic Autoclave.

To maintain an anaerobic environment for bacteria, all media were supplemented with 0.05% L-cysteine after autoclaving as a reducing agent. A 50% w/v stock solution was prepared: to a 15 mL sterile centrifuge tube containing 10 mL demineralised water, 5 g L-cysteine HCl powder was dissolved. In a microbiological safety cabinet, the solution was filter sterilised as previously described. The stock solution was added by mixing 1 mL per L of media after autoclaving and stored at - 20 °C in 1 mL aliquots to avoid repeated freeze-thawing.

2.2.1.2 Substrate preparation

To prepare a 0.5% w/v glucose mMRS medium, a 50% w/v glucose stock solution

was prepared by adding 5 g D-glucose to 50 °C demineralised water and dissolved, the solution was made up to a final volume of 10 mL. The solution was filter sterilised. The concentrated glucose solution was added to sterile mMRS medium: 1 mL per 100 mL of sterile mMRS media. Similarly, to prepare a 0.1% mMRS medium, a 10% w/v glucose solution was made by adding 1 g D-glucose to 10 mL 50 °C demineralised water and filter-sterilised as in Section 2.2.1.1; 1 mL concentrated sterile glucose solution was added per 100 mL sterile mMRS medium. Other sugar (0.1% or 0.5% w/v maltose or lactose) solutions were prepared using the same method.

mMRS supplemented with starch was prepared by various methods. A range of concentrations were tested (0.1% - 2% w/v) e.g., for 1% w/v starch 2.5 g was added to 250 mL mMRS. To integrate cooked starch into media and to ensure accessibility for use as a bacterial growth substrate, "cooked" starch was prepared initially by adding starch powder and mixing on a stirring block with room temperature mMRS medium (prepared as per protocol in Section 2.2.1.1) and sterilising by autoclave to 126 °C for 15 min in a Prestige 2100 Classic Autoclave. However, adding the starch directly to media at room temperature encouraged the formation of a starch gel at the bottom of the Duran bottle, meaning the starch was less integrated into the media either as a solute or as a suspension. This effect was more prominent for higher amylose starches such as Hylon VII which more readily formed gels.

To improve the protocol for cooked starches after mixing with the media (specifically, not for plate reader assays but for CFU assays), it was then brought

to a boil (100 °C) on a hot plate with stirring, or in a microwave with mixing at regular intervals, for 1-3 min or until starch was "cooked" and then sterilised at 126 °C for 15 min in a Prestige 2100 Classic Autoclave.

Ungelatinised (granular) starch which was not heated prior to inoculation was prepared by development of a method involving UV treatment of starch powders. A 5 g soluble potato starch sample was weighed and was evenly distributed in a thin even layer using a fine mesh sieve on a large sterile petri dish in a Class II Walker microbiological safety cabinet. The lid was replaced and transported to a PCR cabinet with a UV radiation lamp. The dish was raised to within ~8 cm of the UV lamp, lid removed and irradiated for 90 min total.

To establish presence of contaminants, the starch was tested by mixing starch powder in 10 mL sterile MRS and RCM using an ethanol sterilised metal spatula. Non-UV treated starch was prepared in the same manner as a positive control. 100 µL of each sample was aseptically spread on 2 BHI agar plates. One plate of each was incubated anaerobically in a Don Whitely Scientific miniMACS anaerobic incubator and one aerobically in a Binder GmbH Model BD 23 incubator at 37 °C for 24 – 48 h. All anaerobic incubators used were set to a temperature of 37 °C maintaining a constantly monitored anoxic environment using a mixture of three gases: 5% CO_2, 10% hydrogen, 85% nitrogen gas.

To further reduce UV sterilisation time, after placing the starch in the petri dish around 15 mL of ethanol was poured to submerge the powder, the lid was replaced and left on the benchtop to soak overnight. The following day, the dish was placed under UV lamp and the time was reduced to 45 min. The method was

proven to be an effective sterilisation protocol for introducing native starch into bacterial growth assays as a substrate. For use in bacterial growth assays, the UV-treated starch was mixed with mMRS medium at 1% w/v before inoculation.

2.2.1.3 Culturing procedure

Frozen 30% glycerol stocks of bacterial strains stored at - 80 °C were aseptically scraped and suspended in either 10 mL MRS/RCM or streaked onto an MRS/RCM agar plate (90 mm triple vent sterile petri dish). Thereafter, the culture was incubated anaerobically for 48 – 72 h at 37 °C in a Don Whitely Scientific minMACS anaerobic incubator. Strains were sub-cultured using 800 µL in 40 mL or 500 µL in 10 mL in MRS (depending on the experimental media volume) for 48 h in the same anaerobic incubator. The day before needed, the strains were pre-cultured anaerobically in MRS to ensure optimal bacterial growth for inoculating the experimental media. For each experiment, bacteria were introduced into the media in triplicate with a 1:100 dilution for inoculation.

2.2.1.4 OD plate reader measurement

Bifidobacterial isolate studies using different substrates were performed using 21 strains. Isolate study methods were the basis for further experiments used to assess bacterial phenotype e.g. RNA transcription (Section 2.2.2) or protein expression (Section 2.2.5)

Starches of different origin and preparation were used to assess growth of each strain in a defined liquid medium containing one available carbon source. In the presence of starch, the changes in bacterial concentration over time ("growth") were monitored and compared relative to the controls, the same concentration w/v

of a sugar or no substrate. Cooked starches included as part of the growth media were always prepared within 2-18h of inoculation of the experiment and stored at room temperature to minimise retrogradation of starches, which occurs more readily at refrigeration temperatures. This included high-amylose content starches, for which retrogradation occurs more readily due to the propensity of amylose to retrograde more quickly than other starch components such as amylopectin. Retrogradation rate of starch is a function of amylose content, storage temperature, and duration of time (Eerlingen, Crombez et al. 1993).

Two different methods to measure bacterial growth on starch were tested. Optical density (OD) is a measure of the light absorbance of a material relative to the light transmitted to the material. Instruments which measure OD transmit a specified wavelength of light into a test liquid. Measuring the solvent as a 'blank' can be subtracted by the instrument or post hoc during analysis to mitigate against the absorption of the solvent. In microbiology, the blank is the sterile medium.

For assessing bacterial growth, although not without its limitations, using an OD spectrophotometer plate reader to measure cultures in a 96-well plate format to measure bacterial growth is a high-throughput method, allowing several strains of bacteria and conditions to be assessed. The sterile media was measured as a blank to mitigate against background absorption. Additionally, OD was measured at a light wavelength of 595 nm to counter the absorbance due to colour, and MRS is best defined as yellow. The amount of light scattered by bacterial cells suspended in a turbid liquid indicates concentration of bacteria by using turbidity as a proxy for the cell population. Measurements taken multiple times over a

growth period can indicate the increase or decrease of bacterial population, thus, replication rate and bacterial phases can be ascertained.

For assessing bacterial growth, an OD spectrophotometer plate reader was used to measure bacterial growth in a 96-well plate format. This is a high-throughput method, allowing several strains of bacteria and conditions to be assessed. Modified MRS (mMRS) was supplemented with specified carbohydrates, prepared according to previously described protocols (Section 2.2.1.2). For each assay, mMRS with no added substrate was used as a negative control, and mMRS supplemented with glucose or lactose of the same concentration as starch was used as a positive control. In aseptic conditions, 190 µL of each media type (up to 4 per assay including positive and negative controls) was transferred using a multi-channel pipette and a sterile reservoir to fill each row of a sterile clear polystyrene 96-well microplate (compatible with plate reader used) with lid.

Bacterial strains were cultured (up to 6 strains per assay, row A and H were media blanks) as described in Section 2.2.1.3. For the 24 h preculture, strains were subcultured into the 96-well plate by adding 10 µL of inoculum to 190 µL fresh MRS media per well. The plate was covered with adhesive gas-permeable seal and incubated in a Ruskinn Concept Plus anaerobic cabinet at 37 °C for 24 h. Prior to inoculation, cultures were mixed well using sterile, filtered p200 pipette tips attached to a 12-tip multichannel pipette. A 1:100 inoculation was added to experimental media in the prepared 96-well plate: 2 µL inoculum culture was added in triplicate to 200 µL mMRS experimental media e.g., 1% starch, 2% starch, 2% lactose, and no substrate. The plate was sealed with adhesive gas-

permeable seal and incubated in a Ruskinn anaerobic cabinet at 37 °C for 48 h. OD measurements were taken by a Tecan Infinite F50 plate reader within the Ruskinn anaerobic cabinet, which were taken at a wavelength of 595 nm every 15 minutes, with shaking for 60s and rest for 60s prior to each read, for the total duration of 48 h.

2.2.1.5 Colony forming unit measurement

The second method used for assessing the bacterial concentration was enumeration of cells plated on solid agar media (Dave and Shah 1996). Pure bacterial samples are ordinarily too concentrated to plate individual cells: the sample was subject to a series of 10-fold dilutions in sterile media. The original bacterial culture being defined as x, the first dilution in the series is $x \times 10^{-1}$, followed by each further dilution decreasing by a factor of 10 i.e., $x \times 10^{-2}, x \times 10^{-3}, x \times 10^{-4}$. Each dilution was plated on a solid agar medium. For dilute concentrations, each single bacterium present on the plate forms a single colony over the incubation period, which is visible by eye. One colony represents one bacterial unit (a colony forming unit, or CFU). By counting each colony, it is possible to calculate the concentration of bacteria in the original culture from which the diluted sample came and express this value as a number: bacterial cells per mL of media (CFU/mL).

To perform serial dilutions of the sample culture, a 96-well microplate was prepared using a multi-channel pipette to fill rows B-G, columns 1-9 with 90 μL sterile MRS broth. Aseptically, 10 μL samples from each experimental tube were removed and deposited to wells in row A, resulting in the first 10^{-1} dilution. To

perform the subsequent dilutions, sequentially 10 µl were removed from rows to serially dilute the culture e.g., from A and dispensed in row B etc. before transferring to the next row, each well was mixed using 100 µL tips and a multi-channel pipette to ensure that the bacterial were well-distributed. Once up to 6 dilutions were complete, 20 µL was removed from the wells and dropped onto large 140 mm MRS agar plates, using 3 dilutions from each triplicate on each plate (9 drops total on each plate). Plates were incubated for 2-3 days, and colonies counted. Serial dilutions and plating were performed at pre-determined timepoints (including 0 h) to obtain CFU/mL over the course of the experiment.

Data from each CFU count was input into a spreadsheet and colony forming units (CFU/mL) were calculated by multiplying the number of colonies by the dilution factor, giving the CFU of the 20 µL dilution dropped onto the plate. To calculate the CFU per mL, the number was multiplied by 50. The CFU/mL triplicate values were averaged to show bacterial concentrations at each timepoint in a line graph.

To generate growth curves of bacterial growth in each media type, the culture volume was typically 10 mL – 40 mL (depending on the volume of sample being removed) per experimental variable at various pre-determined timepoints over the course of 24 or 48 h.

2.2.2 Next Generation Sequencing Methodology

In order to identify bacteria in complex starch fermentation ecosystems or gene expression of bacteria in the presence of starch, the nucleic acid material of bacteria was sequenced. All sequencing performed for data production in this thesis belong to the Next Generation Sequencing (NGS) class of sequencing

technologies. Sanger sequencing sequences one DNA fragment at a time; contrary to this, NGS sequences multiple (millions) of fragments in parallel per run (Levy and Myers 2016). The advantages and reasons for emergence of NGS are high sensitivity, quick sequencing runs, combined with a greater capacity to sequence multiplexed pooled samples in one run, meaning sequencing jobs can be extremely high throughput.

2.2.2.1 Library preparation

A feature of NGS is sequencing of a number of samples at once through the generation of DNA libraries. Each sample is equipped with a molecular label ("barcode") which means all samples are combined into a single pool, whereby each sample can be recognised and sequenced individually by the flow-cell. This process includes a series of steps using molecular biology techniques, the first of which is called tagmentation. This method relies on a naturally occurring enzyme called transposase, which is colloquially known as a "cut and paste" enzyme (Adey 2021). A stable enzyme complex known as a transposome can be linked to beads at a known concentration for a more efficient workflow, known as Bead-Linked Transposomes or BLT (Bruinsma, Burgess et al. 2018).

During tagmentation, DNA is cut into fragments and simultaneously adapters ("barcodes") several base pairs (bp) in length are attached. These labels comprise a mosaic end sequence of double-stranded DNA, as well as a single-stranded overhang sequence, which act as primers for subsequent PCR, where further indexes can be added. When sequencing is performed, the mosaic end and

adapter sequence are identified by primers which allow paired reads of the DNA and their index sequences (Adey 2021).

2.2.2.2 Sequencing by synthesis

There are some distinct differences between detection and chemistry used between sequencing platforms chosen for data production in this thesis. Sequencing by synthesis technology is a commonly employed chemistry employed by Illumina sequencing platforms. First, the DNA fragments are copied by PCR and loaded into a flow cell. A flow cell is a microfluidic hollow channel, or lane, containing complementary oligonucleotides to the adapter sequences ligated to sample DNA libraries (Ambardar, Gupta et al. 2016). The flow cell is lined with oligonucleotides which bind to the amplified sample DNA due to natural complementarity (Fuller, Middendorf et al. 2009). The generation of clonal sequence clusters occurs by the DNA polymerase first creating a 'reverse' strand, which then bends to attach to an adjacent oligo which is complementary to the adapter sequence on the opposite end of the strand. A copy of this strand is then made (the original sequence), to create two strands, a forward and reverse strand which are clonal copies of the original sequence. This is known as bridge amplification by clonal clustering.

By synthesising copies of the sample DNA with fluorescently tagged nucleotides, each base has a unique tag able to be optically detected by the instrument sensors, allowing for each base to be identified in the (reverse) order of the original sample DNA (Ambardar, Gupta et al. 2016). Illumina sequencers perform reads of both the forward and reverse of the fragments, called paired-end

sequencing. For RNA sequencing, a conventional strategy is to synthesise and sequence complementary DNA (cDNA).

2.2.2.3 Nanopore sequencing

Oxford Nanopore Technology sequencing involves a more direct measurement of the translocation of DNA through a nanopore sensor in the flow cell. The fragment adapters and strand capture mechanism of the flow cell is similar to Illumina, however a processive enzyme is also co-located with the DNA fragment (Jain, Olsen et al. 2016). Instead of a synthesis chemistry approach, the DNA adapters are concentrated at the nanopore membrane surface, where the DNA is 'captured' and contiguously fed through the nanopore one nucleotide at a time by procession of the enzyme. A sensor in the nanopore detects the variation in disturbance to an ionic current based on which nucleotide is passing through. In this way, the nanopore sensor can sequence the DNA by identification of each nucleotide in the sequence.

2.2.2.4 Sequencing data outputs

The main data output differences between Illumina and Nanopore are length and number of reads produced per run. Illumina prioritises quantity of data output per run, while producing relatively short reads (35-300 bases per read) (Levy and Myers 2016). As already mentioned, Illumina does however facilitate paired-end sequencing: reading each end of the DNA fragment and its full length improves the quality of the read. Oxford Nanopore sequencing can commonly produce read lengths of tens of thousands of bases per read (Levy and Myers 2016). Both technologies have excellent utility for different applications. For example, in this

thesis, Illumina technology has been used to sequence complex metagenomic DNA derived from thousands of microbial taxa in parallel, known as shotgun metagenomics. Meanwhile, the long reads produced by Oxford Nanopore's minION facilitated a *de novo* genome assembly of a bacterial strain (*B. pseudolongum* subsp*. pseudolongum* NCIMB 702244).

2.2.3 Whole Genome Sequencing of Bacterial Isolates

Due to lack of a publicly available complete genome for *B. pseudolongum* subsp. *pseudolongum* NCIMB 702244, the bacterium was sequenced employing both long- and short-read sequencing (Table 2-6).

Table 2-6 Summary of sequencing performed for whole genome sequencing		
Experiment	Whole Genome Sequencing strategy for *B. pseudolongum* NCIMB 702244	
Description	Short-read sequencing	Long-read sequencing
Library prep by	QIB Sequencing	QIB Sequencing
Sequencing by	QIB Sequencing	QIB Sequencing
Sequencing Technology	Illumina NextSeq™ 550 Sequencing System	Oxford Nanopore Technologies MinION sequencer

2.2.3.1 Bacterial Isolate DNA extraction

To achieve successful cell lysis, the FastDNA™ SPIN Kit for Soil was used, following manufacturer instructions, with an extended 3 min bead-beating. For bifidobacteria, the process was specifically optimised to use 3 min of homogenisation due to the bacterial cell walls being difficult to lyse (Alcon-Giner, Dalby et al. 2020).

For genome sequencing of a reference type strain purchased, *Bifidobacterium pseudolongum subsp. pseudolongum* DSM 20099 / NCIMB 702244, the strain was

cultured from 30% glycerol stock using the methods described in Section 2.2.1.3. The strain was sub-cultured using 500 μL in 10 mL in MRS and incubated anaerobically for 48 h at 37 °C in a Don Whitely Scientific miniMACS anaerobic incubator for each incubation. Oxoid™ Phosphate Buffered Saline (PBS) Tablets were used to prepare PBS buffer as per product recommendations: one tablet per 100 mL demineralised water (1X PBS solution: 137 mM NaCl, 9.5 mM phosphate, 3 mM KCl, pH 7.3). The PBS was sterilised to 126 °C for 15 min in a Prestige 2100 Classic Autoclave. The culture was removed from the incubator and centrifuged for 10 min at 4000 g. The supernatant was discarded, and 5 mL PBS was added to resuspend the bacterial pellet. 200 uL of this bacterial solution was added to an MP Biomedicals™ Lysing Matrix E tube and the protocol was followed using the method published by Alcon-Giner et al. (2020) for extracting DNA from infant faecal samples. The method utilises FastDNA™ SPIN Kit for Soil, following manufacturer instructions, with extended 3 min bead-beating.

2.2.3.2 Nucleic acid quantification

Qubit is a well-established, commonly used quantification technology for DNA, RNA, and protein samples using extremely low volumes and/or concentrations of sample. The platform relies on the principle of fluorometry: a sample molecule is subject to an excitation (radiation) which is absorbed and causes the sample to emit a different wavelength of light, also known as fluorescence (Hawker, Genzen et al. 2018). According to the manufacturer, the Qubit kit includes fluorescent reagents (dyes) which bind to the desired molecule (DNA) and only exhibits fluorescent properties when bound to the target molecules. Using fluorometry to excite the reagents, the fluorometer reads the light emitted as proportional to the

concentration of analyte molecules. Other platforms such as the Promega QuantiFluor® dsDNA System also employ the fluorometric properties of analyte-bound dyes.

DNA extract concentration was quantified using a Qubit® 2.0 fluorometer as described in the Qubit™ dsDNA HS and BR Assay Kit instructions. In brief, a Qubit™ working solution was made by diluting the Qubit™ dsDNA BR Reagent 1:200 in Qubit™ dsDNA BR Buffer in a clean, plastic, 5 mL snap-cap tube. To prepare the standards, two 0.5 mL, clear PCR tubes were labelled Standard 1 and Standard 2, and the standard was mixed with the working solution to a final volume of 200 µL with a ratio of 10 µL standard to 190 µL working solution. The samples to be quantified were also mixed (in separate, clean tubes) at a ratio of 1 µL to 199 µL working solution. The samples and the standards were vortexed for 2-3 seconds and then incubated, covered with foil, for 2 minutes. The standards were measured first, followed by measuring the samples.

2.2.3.3 Short-read Sequencing

The subsequent steps (short- and long-read sequencing) were carried out by Dave Baker and Rhiannon Evans of the Quadram Institute Core Sequencing department.

The DNA extracts were stored at -20 °C. After thawing, genomic DNA was normalised to 5 ng/µL with Elution Buffer (10mM Tris-HCl). To prepare the genome libraries, which facilitates flow-cell identification and adherence, a method was employed based on a modified Illumina® DNA prep using 20-fold less tagmentation reagent and specific step omission to save on resources (Baker,

Aydin et al. 2021). A Tagmentation Mix was prepared by combining three reagents: TB1 Tagment DNA Buffer (0.5 μL) was mixed with 0.5 μL BLT, Tagment DNA Enzyme (Illumina), and 4 μL PCR grade water. The total volume prepared (5 uL) was added to a chilled 96-well plate. A volume of 2 μL of normalised DNA (10 ng total) was pipette mixed with the 5 μL of the Tagmentation Mix and heated to 55 °C for 15 minutes in a PCR block.

A PCR master mix was prepared using a ratio of 10 μL KAPA 2G Fast Hot Start Ready Mix (Merck) to 2 μL PCR grade water; 12 μL of PCR master mix was required per sample and so scaled up appropriately to the number of samples. PCR master mix was added to each well in a clean 96-well plate. P7 and P5 of Nextera XT Index Kit v2 index primers (Illumina) were added to each well at a quantity of 1 μL per primer. Finally, the 7 μL of Tagmentation Mix and DNA was added to each well and pipette mixed.

The PCR parameters were programmed: 72 °C for 3 minutes, 95 °C for 1 minute, 14 cycles of 95 °C for 10s, 55 °C for 20 seconds and 72 °C for 3 minutes. The libraries were quantified according to the manufacturer's instructions using fluorometry-based quantification assay, QuantiFluor® dsDNA System (Promega). Library concentrations were measured using a GloMax® Discover Microplate Reader (Cat. No. GM3000, Promega, Southampton, UK).

Following quantification, libraries were pooled in equal quantities. The final pool was double-SPRI size selected between 0.5 - 0.7X bead volumes and quantified using a Qubit 3.0 instrument. A D5000 ScreenTape (Agilent) was used to calculate

the final library pool molarity using the Agilent Tapestation 4200 (Cat. No. G2991BA, Agilent Technologies Ltd, Cheadle, UK)

The pool was sequenced at a final concentration of 1.5 pM on an Illumina Nextseq500 instrument using a Mid Output Flowcell (NSQ® 500 Mid Output KT v2(300 CYS), Illumina) following the Illumina recommended denaturation and loading recommendations which included a 1% PhiX spike in (Illumina). The data was uploaded to Basespace (www. basespace.illumina.com) where the raw data was converted to 8 FASTQ files for each sample.

2.2.3.4 Long-read Sequencing

A different library preparation protocol was followed for downstream long-read sequencing using the Nanopore MinION sequencing platform. Generally, short- and long- read sequencing require different library prep methods, as with the latter, preservation of longer DNA fragments is desirable. While incorporating the modified tagmentation steps of the Illumina® DNA Prep kit method, the Nanopore library prep is a different PCR-based method for long-read Nanopore sequencing which is used for smaller genomic DNA (bacteria rather than animal, for example). Primers were designed with a 3' end compatible with the Nextera transposon insert and a 24bp barcode at the 5' end including a pad and spacer. The same barcode was used at each end using 96 combinations. Primer pairs were mixed at a concentration of 10 µM.

The tagmentation reaction was reduced 20-fold. A Tagmentation Mix was prepared by combining three reagents: TB1 Tagment DNA Buffer (0.5 µL) was mixed with 0.5 µL bead-linked transposomes (BLT) (Illumina) and 4 µL PCR grade

water. The total volume prepared (5 uL) was added to a chilled 96-well plate. DNA was normalised to between 50-100 ng/µL. Thereafter, 2 µL of normalised DNA was pipette mixed with the 5 µL of the Tagmentation Mix and heated to 55 °C for 15 minutes in a PCR thermoblock.

A PCR master mix was made using a ratio of 10 µL NEB LongAmp Taq 2X Master Mix to 2 µL PCR grade water (12 µL required per sample). To each well of a new 96-well plate, 12 µL of the master mix was added. Additionally, 1 µL of the appropriate primer pair at 10 µM was added to each well. Finally, the 7 µL of Tagmentation mix was added and mixed by pipette. The PCR was run using the following parameters: 1 cycle at 94 °C for 2 min; 14 cycles at 94°C for 30 sec; 55 °C for 1 min; 65 °C for 15 min; 1 cycle at 65 °C for 15 min; and finally hold set at 10°C.

Following the PCR reaction, the libraries were quantified according to the manufacturer's instructions using QuantiFluor® dsDNA System (Promega) and run on a GloMax® Discover Microplate Reader (Promega). Libraries were pooled together in equal nanograms and SPRI size selected at 0.6X bead volume using Illumina Sample purification beads with final elution in 32 µL. A 0.75% Agarose Cassette was used to run 30 µL of sample for processing using a sageELF (Cat. No. ELF0001, Sage Science, Beverly, USA). sageELF fractionates the DNA using electrophoresis to separate the fragments based on size. Using a running time of 3 hours and 40 minutes on the 1-18kb run mode, 10kb fragments were in the middle of the 12 elution wells. The eluted fractions were then run on an Agilent 4200 TapeStation using D5000 Genomic ScreenTape (Agilent); fragments of 3kb and above were pooled together. Pooled fractions underwent a 1X SPRI clean-up and

were eluted in 50 µL. The eluate then underwent genomic ligation protocol using a Ligation Sequencing Kit (Oxford Nanopore Technologies) and was loaded on a MinION device according to the manufacturer's instructions for sequencing.

2.2.3.5 Whole Genome Assembly

A whole genome assembly was carried out by Raymond Kiu as described below. The long reads acquired by Nanopore MinION sequencing were used by to assemble the reads into one complete sequence, representing a *de novo* assembly of a previously unknown whole genome sequence. To ensure the sequence obtained from many reads is assembled in the correct order, software needs to identify the regions of the reads which overlap with other reads, which is called a consensus region. The consensus region can be identified if two reads manage to overlap in certain segments of the DNA, meaning they are adjacent or contiguous. These consensus regions are known as contigs. Using long reads from MinION sequencing data, a draft genome of a single contig containing 1,916,804 bp was generated as follows.

Firstly, Filtlong (v0.2.1, Anaconda, Inc., Austin, USA) was used to filter long reads by preferentially selecting higher quality, longer reads from the read set, in which the worst 10% reads and reads <1kbp were discarded (Wick and Menzel 2017). Next, de novo assembler Flye v2.9 (--nano-hq mode) was used to assemble the Nanopore long reads into a draft genome with 5 polishing iterations (with option -i 5) (Kolmogorov et al. 2019). Subsequently, the genome assembly was further corrected and polished with Medaka v1.4.3 (Oxford Nanopore Technologies, Oxford, UK). A final tool, Circlator (v1.5.5, (Hunt, Silva et al. 2015)) was used

attempting to circularise the genome assembly. However, the final genome of one contig was not fully circularised.

2.2.4 Transcriptomic Profiling of *B. pseudolongum*

Transcription is an essential process for life. RNA is made through the process of DNA being copied into RNA transcripts, thereafter, they are translated into protein by a ribosome. Intermediate RNA is known as messenger RNA (mRNA). When RNA is extracted and sequenced from an organism (RNA-seq), the transcripts represent the gene expression at time of extraction (Lowe, Shirley et al. 2017).

RNA-seq forms the basis of a transcriptomics experiment, where the gene expression can provide information about the response of organisms, especially in relation to specific genes which may be differentially regulated under different conditions. The sequencing strategy is outlined in Table 2-7 for an experimental design involving two strains of the same species cultured in the presence of starch.

Table 2-7 Transcriptomics sequencing strategy.

Experiment summary	Transcriptomic profiling of *B. pseudolongum* in the presence or absence of starch
Strains	*B. pseudolongum* subsp. *globosum* NCIMB 702245 and *B. pseudolongum* subsp. *pseudolongum* NCIMB 702244
Library prep by	Azenta (formerly Genewiz), Cambridge, UK
Sequencing by	Azenta (formerly Genewiz), Cambridge, UK
Sequencing Technology	Illumina NovaSeq™ 6000 Sequencing System
Sequencing depth	20 million paired end reads
Other data obtained	NMR metabolomics + CFU

Summarised in Figure 2-1, media used was prepared according to previously described protocols in Section 2.2.1.1, including modified MRS (mMRS) which

was supplemented with specified carbohydrates the following substrates prepared as described in Section 2.2.1.2: as a positive control, mMRS supplemented with 0.5% D-glucose (Sigma-Aldrich); the cooked starch substrates were prepared as mMRS with 1% high amylose maize starch Hylon VII (Ingredion) or 1% normal maize starch (Sigma-Aldrich).
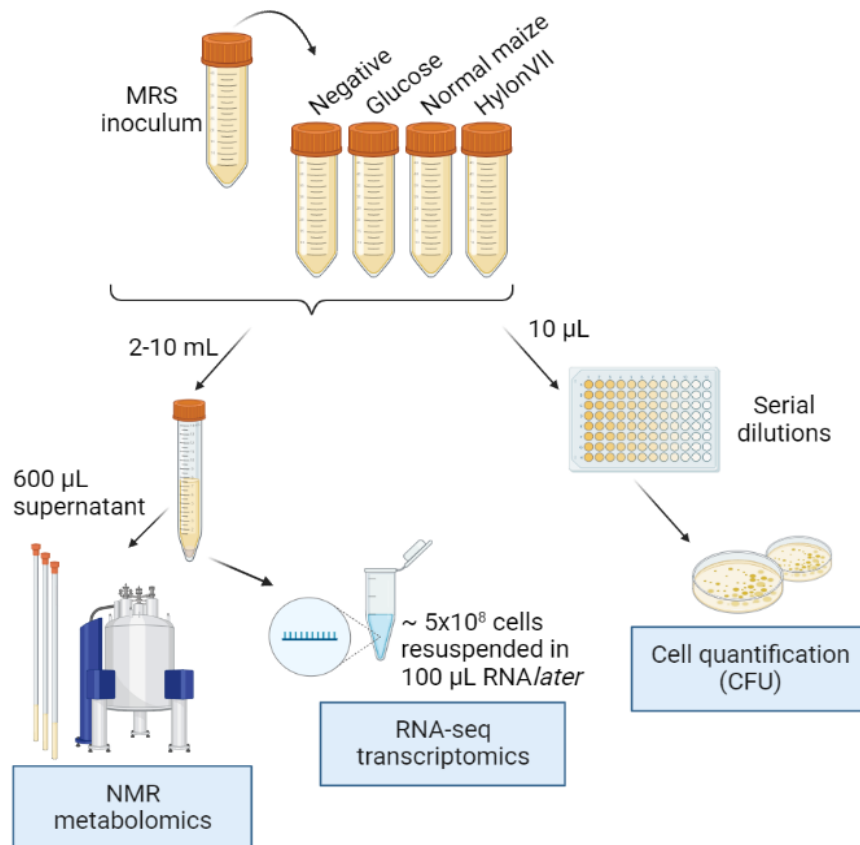
Figure 2-1 A visual summary of transcriptomics experimental design and methods. All conditions were performed in biological triplicate; inoculum sampling occurred at 0h. For sampling at 12h and 24h, parallel cultures of each condition and replicate were implemented. Figure created using BioRender.com.

Cultures of *B. pseudolongum* NCIMB 702244 and *B. pseudolongum* NCIMB 202245 were carried out as per the culturing procedure detailed in Section 2.2.1.3. For each culture, 40 mL fresh MRS were inoculated using 800 µL in triplicate. The

final 24 h pre-culture (the inoculum for the experiment) was prepared in triplicate.

Two sets of parallel experimental tubes for each bacterial strain were inoculated

by adding 400 µL of inoculum per 40 mL media and anaerobically incubated at

37 °C in a Ruskinn Concept Plus anaerobic incubator. At 12 h, one set of

experimental tubes were removed for sampling. At 24 h, the second set of

experimental tubes containing the same media and strain conditions were

removed for sampling and processing in an identical manner.

2.2.4.1 Experimental sampling and processing

To generate bacterial growth curves of each strain in each media type, at

timepoints 0, 12, and 24 h, 10 µL samples were removed for serial dilution, plating

and CFU quantification as described in Section 2.2.1.5. For sampling the cultures

to assess metabolite production, 1 mL from each timepoint (0, 12, 24 h) was

removed and stored in 2 mL snap-cap plastic microtubes (including each parallel

culture). The samples were stored at – 80 °C until processed for NMR

metabolomics according to protocols in Section 2.2.7.

For sampling the cultures for RNA extraction, the inoculum used was prepared

using 40 mL MRS. A 2 mL sample was reserved for RNA extraction and

transcriptome profiling of the bacterial cells in their initial transcriptional state (at 0

h) at point of inoculation. For each timepoint, 12 and 24 h, the corresponding

parallel culture tubes were sacrificed for sampling 2 mL of the starch and glucose

cultures, and 10 mL of the mMRS (negative control) tubes. The samples were

either deposited in 10 mL sterile plastic tubes (2 mL samples) or remained in their

original tubes (negative control) and centrifuged in an Eppendorf 5810 R at 4000 g

for 10 minutes. The supernatant was discarded and 100 µl of RNA*later*™ Stabilisation Solution (Invitrogen™) was used to resuspend the cell pellets. The suspension was transferred to a 2 mL snap-cap plastic microtube and stored at 4 °C for a maximum of 14 h before RNA extraction.

2.2.4.2 RNA extraction and sequencing

RNeasy® Mini kits by Qiagen were used to perform the RNA extraction. In a fume cupboard, 10 µL β-mercaptoethanol (Sigma-Aldrich) was added to 1 mL Buffer RLT Plus before use. Bacterial cells containing RNAlater™ (Thermo Fisher) were mixed with 600 µL RLT lysis buffer and transferred to lysing matrix E-tube and placed in an MP Biomedicals™ Fastprep 24 – speed setting 6 for 60 seconds, repeated 3 times (3 mins total) keeping samples on ice between the bead beating steps. Tubes were centrifuged for 10 min at 14000 g. The supernatant was transferred to a 2 mL Low-bind Eppendorf tube before adding 600 µL of 70% ethanol and mixed well by pipetting gently. A volume of 700 µL was then transferred to an RNeasy spin column and centrifuged at 8000 g for 15 sec and the flow through was discarded, repeating twice until all lysate has been filtered with the same column. One volume (700 µL) of buffer RW1 was added and centrifuged at 8000 g for 30 sec. The flow-through was discarded and the spin column was placed in a fresh collection tube. RPE was added at a volume of 500 µL and centrifuged at 8000 g for 30 sec. The flow-through was discarded. A further 500 µL RPE was added and centrifuged at 8000 g for 2 min. The spin column was placed in a fresh collection tube and centrifuged at 8000 g for 2 min. The spin column was placed in a fresh 1.5 mL collection tube and 35 µL RNase free water was directly added to the spin column membrane. Samples were incubated for 1

min at room temp. before being centrifuged at 8000g for 1 min – the spin column was discarded after eluate was obtained.

TURBO DNA-free™ kit for DNA removal was followed using the routine protocol. TURBO DNase™ Enzyme was added to the RNA extract samples, 1 µL per sample, in addition to 3.5 µL 10X TURBO DNase™ buffer was added per sample and mixed gently. All samples were incubated at 37 °C for 20–30 minutes in a Binder GmbH Model BD 23 aerobic incubator. The DNase Inactivation Reagent was added at a volume of 3.5 µL per sample, mixing well, and incubated for 5 minutes at 24 °C. To ensure distribution of deactivation reagent, each tube was flicked 2 times during the incubation period. The samples were centrifuged for 1.5 min at 10,000 g and subsequently the supernatant containing DNA-free RNA extract was carefully transferred to a fresh tube without disturbing the inactivation reagent pellet. The samples were aliquoted – one plate containing 5 µL for QC and another containing 30 µL for sequencing – into chilled 96 well plates and were stored at -80 °C.

2.2.4.3 RNA QC

With the lab assistance of Rhiannon Evans, RNA was quantified using Qubit™ RNA High Sensitivity (HS) quantification kit and Qubit™ 2.0 spectrometer, which was performed according to previously described method in Section 2.2.3.2 with the following amendments. RNA quantification was performed using Qubit™ working solution by diluting the Qubit™ RNA BR Reagent 1:200 in Qubit™ HS RNA buffer and preparing standards using supplied RNA standards. The protocols

for preparing standards and quantifying samples are the same as previously described.

Agilent 4200 TapeStation and materials listed in Table 2-4 were used to assess RNA quality, performed by myself with the assistance of Rhiannon Evans. The 4200 TapeStation instrument is an electrophoresis tool specifically designed for nucleic acids. Using proprietary gel tapes which allow high throughput and automated gel electrophoresis of samples, consistent assessment of concentration, integrity, and size (bp) occurs. RNA integrity number equivalent, or RIN^e, expressed as a value between 1-10, is a function of the RNA degradation of the sample. The software compares expected ribosomal peaks (for prokaryotes at 16S and 23S) to any other peaks detected which are deemed degradation products.

Reagents and ScreenTapes stored at 4 °C were equilibrated to room temperature for 30 min prior to use. Firstly, the RNA samples were diluted to 1 ng/µL using PCR-grade nuclease-free water. RNA sample buffer was added to clean, labelled microtubes (5 µL). A volume of 1 µL RNA ladder was added to the first tube and 1 µL RNA sample to the remainder. Using a Programmable Thermal Cycler (MiniAmp Plus, Applied Biosystems) the RNA was denatured at 72 °C for 3 mins before cooling on ice for 2 mins and finally pulse centrifuging the tubes. High Sensitivity RNA ScreenTape® was inserted into the Agilent 4200 TapeStation and the tubes were inserted before starting the program.

2.2.4.4 RNA sequencing protocol

The RNA extracts were shipped on dry ice to Azenta in Cambridge, UK, where library preparation and sequencing was carried out. A detailed list of the materials used (Section 2.1.3). RNA samples were first quantified using Qubit 4.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked with RNA Kit on Agilent 5300 Fragment Analyzer (Agilent Technologies, Palo Alto, CA, USA). rRNA depletion was performed using NEBNext rRNA Depletion Kit (Bacteria). RNA sequencing library preparation was performed using NEBNext Ultra II RNA Library Prep Kit for Illumina by following the manufacturer's recommendations (NEB, Ipswich, MA, USA). Briefly, enriched RNAs were fragmented according to manufacturer's instruction. First strand and second strand cDNA were subsequently synthesized. cDNA fragments were end repaired and adenylated at 3'ends, and universal adapter was ligated to cDNA fragments, followed by index addition and library enrichment with limited cycle PCR. Sequencing libraries were validated using NGS Kit on the Agilent 5300 Fragment Analyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by using Qubit 4.0 Fluorometer (Invitrogen, Carlsbad, CA). The sequencing libraries were multiplexed and loaded on the flow cell on the Illumina NovaSeq 6000 instrument according to manufacturer's instructions. The sequencing parameters selected were depth of ~20 million paired end, 150bp reads per sample. The samples were sequenced using a 2x150 Pair-End (PE) configuration v1.5. Image analysis and base calling were conducted by the NovaSeq Control Software v1.7 on the NovaSeq instrument. Raw sequence data (.bcl files) generated from Illumina NovaSeq was converted into fastq files and de-multiplexed using Illumina

bcl2fastq program version 2.20. One mismatch was allowed for index sequence

identification. Raw data was delivered in FASTQ format and downloaded by

Andrea Telatin.

Table 2-8 Description of B. pseudolongum 45 reference genome used for mapping of RNA transcripts.			
Genus	**Species**	**Subspecies**	**Strain / Culture collection**
Bifidobacterium	*pseudolongum*	*globosum*	DSM 20092 NCIMB 702245
Genome details	**Length**	**Sequencing**	**Description**
Annotated reference genome	1,915,040 bp	whole genome shotgun (WGS) Illumina Hiseq	63 contigs ranging from 204-210,742 basepairs in length
BioProject		**Accession number**	**Reference**
SAMN02442013 PRJNA237576		JDTV01000001.1	(Sun, Zhang et al. 2015)

2.2.4.5 Transcriptomics Data Analysis

Andrea Telatin perform the following steps to process the transcriptomic data. The

reads were mapped to genes from the reference genome (Table 2-8). The

percentage of mapped reads for each sample was displayed using a bar plot in

Microsoft Excel. The data was analysed using iDEP.96 (Ge, Son et al. 2018) to

create plots and perform pairwise comparisons of all variables to identify

Differential Expressed Genes (DEGs) using DeSeq2 (Love, Huber et al. 2014).

The Benjamini–Hochberg algorithm was used to produce an FDR value, the cut-off

set to q = 0.05 and minimum fold change of 2. Substrate (no substrate, glucose,

NMS, HylonVII) was selected as the main factor.

The total number of differentially expressed genes were downloaded as tables and

displayed as bar charts in Microsoft Excel. The specific gene tables containing up

or downregulated were selected (Table 2-9). The top differentially expressed

genes were selected and cross-referenced with genes also upregulated in the other conditions to assess patterns in upregulation. Genes upregulated in both glucose and starch were assumed to be involved in sugar transport and carbon metabolism. Genes which were (a) uniquely upregulated in the starch conditions versus no substrate but not the glucose and (b) also upregulated in the starch conditions vs glucose were assigned as starch degradation related genes. Genes uniquely upregulated in Normal maize vs HylonVII and vice versa were also assessed.

Table 2-9 Statistical comparisons performed for RNA-seq data

Statistical comparisons using DESeq2	
Normal maize	No substrate
HylonVII	No substrate
Glucose	No substrate
Normal maize	Glucose
HylonVII	Glucose
Normal maize	HylonVII

2.2.5 Proteomics: LCMS

To complement the hypotheses for transcriptomic profiling experiment (Section 2.2.4), a small proteomics experiment was performed using *B. pseudolongum* subsp. *globosum* NCIMB 702245 in the presence of starch or glucose after 24 h (Table 2-10). Proteomics is defined as an experimental assessment of all proteins being produced either within cells or externally, or both. Liquid chromatography mass spectrometry (LCMS) was used as an analytical chemistry technique which can identify and quantify proteins by leveraging the ability of liquid chromatography to separate compounds and mass spectrometry's detection of mass-to-charge ratio. Addressing the limitations of RNA-seq in that transcripts may not be translated into functional proteins, measuring the protein products of upregulated genes provides more evidence of the microbial response to a treatment.

Table 2-10 Proteomics experimental summary

Experiment summary	Proteomics of *B. pseudolongum* in the presence or absence of starch
Strains	*B. pseudolongum* subsp. *globosum* 702245
Conditions	Glucose Normal maize
Timepoint(s)	24h
LCMS performed at	BayBioMS Technical University of Munich

2.2.5.1 Bacterial Lysis with Trifluoroacetic Acid (TFA)

Culturing of *B. pseudolongum* NCIMB 702245 (Section 2.2.1.3) was performed in 10 mL in 0.5% NMS starch and 0.5% D-glucose (Section 2.2.1.2) modified MRS per sample, in triplicate, for a total experimental culturing time of 24 h. Samples were stored at -80 °C. Neutralization Buffer (NB) was prepared by dissolving 24.22 g Tris (Sigma-Aldrich) in a final volume of 100 mL demineralised water. Oxoid™ Phosphate Buffered Saline (PBS) Tablets were used to prepare PBS buffer as described in Section 2.2.3.1. Bacterial cell cultures were thawed and centrifuged for 10 min at 4,000 g. The supernatant was discarded, 5 mL PBS was added, and cells were washed by vortexing. The samples were centrifuged for 10 min at 4,000 g and supernatant was discarded. The cells were resuspended in 100 µL 100 % TFA and transferred to clean plastic microtubes which were incubated in a thermocycler at 55 °C. A volume of 900 µL NB was added and tubes were vortexed. Using pH paper, samples were adjusted to pH 8.2 if necessary. The lysed samples were quantified by Bradford assay and stored at -80 °C.

2.2.5.2 Protein Quantification by Bradford Assay

Determination of total protein concentration for preparing samples for LCMS can be carried out using a Bradford Assay. Utilising a dye called Coomassie brilliant

blue which binds protein carboxyl and amino groups, the dye changes colour from brown to blue proportionally to concentration. Using a colorimeter and a standard protein series, the total protein concentration can be calculated.

To generate a calibration curve, known concentrations of Bovine Serum Albumin (BSA) were tested. A BSA stock solution was prepared by adding 100 mg to 50 mL pure water and dissolving. Defined volumes of this stock solution were added to microtubes containing pure water to create a concentration series. To create calibration standards, each stock solution (Stock A – Stock I) was diluted in equal volumes of NB containing 10 % TFA to mimic the sample conditions e.g. 20 µL stock and 20 µL NB. Each protein sample was diluted in equal volumes of pure water.

The standards and protein samples were pipetted in triplicate (5 µL each) into a flat bottomed 96 well plate. To each well, 250 µL Bradford Solution was added. The samples were incubated for 10 min at room temperature, replacing lid to prevent evaporation. The plate (without lid) was inserted into a CLARIOstar Plus plate reader (BMG Labtech, Ortenberg, Germany). The absorbance was measured at $\lambda = 595$ nm and $\lambda = 690$ nm. The BSA standard calibration curve was fitted using a polynomial regression model and the sample protein concentration was calculated using the resultant equation.

2.2.5.3 Liquid Chromatography Mass Spectrometry

Samples were shipped to Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS) at the Technical University of Munich where they were processed and

analysed by Christina Ludwig and lab members using LCMS to assess the identity and relative abundance of proteins in each sample.

2.2.5.4 Data Analysis

Omicsviewer 1.1.5 (Meng 2022) was used to create volcano plots and perform t-tests of differential protein abundances between *B. pseudolongum* subsp. *globosum* NCIMB 702245 in the presence of normal maize starch or glucose which was downloaded as a table. The top upregulated genes in normal maize starch (NMS) compared to glucose were identified and displayed by table in Microsoft Excel.

2.2.6 Model Colon Fermentation Studies

A series of *in vitro* fermentation experiments were carried out to investigate the effects of starch and bifidobacterial supplementation on adult and infant microbiota ecology, which will be described in detail in Chapter 5.

2.2.6.1 Materials

Adult faecal samples used in the colon model experiments were obtained from participants recruited into the QIB Colon Model study. The collection of stool donations from health adult donors was approved by the Quadram Institute Bioscience (formally Institute of Food Research) Human Research Governance committee (IFR01/2015), and London - Westminster Research Ethics Committee (15/LO/2169). The criteria for these samples were that the donors are adult and healthy men and women aged 18 years or older with a normal bowel habit, no self-reported diagnosed chronic gastrointestinal health problems. Immediately prior to donation, donors were asked to confirm that they had not taken antibiotics or

probiotics within the last four weeks, had not experienced a gastrointestinal complaint within the last 72 h, were not currently pregnant or breast-feeding, had not recently had an operation requiring general anaesthetic, and were not taking iron or multivitamin supplements. The informed consent of all participating subjects was obtained, and the trial is registered at http://www.clinicaltrials.gov (NCT02653001). Adult stool donor samples were collected, aliquoted and frozen at -20 °C in 50 mL sterile, screw-cap plastic tubes.

Infant stool samples were transferred following ethics approval and distribution procedures for the PEARL study (IRAS project ID 241880) (Phillips, Watt et al. 2021). Pregnant women were recruited into the PEARL study with informed consent of the expectant mother to collect samples of their infant after birth. The women's health criteria to be eligible were BMI between 18 - 35 and non-smoker; not taken antimicrobials in the last 3 months or steroids in the last 6 months; no gastrointestinal (incl. long-term constipation), immune, or liver abnormalities; history of cancer (except non-metastatic, treated skin cancer); major dietary changes in last 30 days; history of substance abuse (Phillips, Watt et al. 2021).

Samples from infants were collected, aliquoted and stored at -20 °C. The inclusion criteria for infant stool donors used in this thesis were that infant donors selected were during the weaning stage of life (consuming breastmilk and/or formula in addition to solid foods including starchy foods) and were otherwise healthy; the infants ranged from 8-12 months in age at point of donation (Table 2-11).

	Age at donation (months)	Breastmilk (feedings per week)	Formula (feedings per week)	Starchy foods* (feedings per week)
Infant1	8	14	21	9
Infant2	12	0	28	6
Infant3	8	21	0	7

Table 2-11 Infant stool donor information and dietary metadata

*Baby cereal or other cereals incl. breakfast cereals, teething biscuits, crackers, breads, rice, pasta

2.2.6.2 Model colon media preparation

Using the reagents and ingredients detailed in Table 2-12, trace mineral, fatty acids, resazurin, and haemin solutions were individually prepared and combined with basal solution as follows. Trace mineral solution was prepared in a clean glass beaker containing 800 mL 0.02 M HCl , each of the trace minerals (listed in media ingredients Table 2-12) were added in order on a stirring block, ensuring to allow sufficient time for each to dissolve before the next was added. The solution was made up to 1 L with 0.02 M HCl in a volumetric flask and stored at 4 °C for up to 6 months in a sealed bottle. Fatty acids solution was prepared by in a fume hood by making 100 mL of 0.2M NaOH and transferring to a beaker. The stated acids were added by pipette (Table 2-12) and stored at 4 °C in a sealed Duran. Resazurin solution was made by adding 0.05 g Resazurin to 40 mL of deionised water and mixed well to dissolve. The solution was made up to 50 mL volume in a clean volumetric flask and stored in a sealed bottle covered with foil at 4 °C. Haemin solution was made by adding 0.05 g of haemin to 25 mL of 0.05M NaOH (0.05 g NaOH dissolved in 25 mL with deionised water) and dissolving. This solution was made up to 500 mL with boiled deionised water in a volumetric flask, allowed to cool, and stored in a foil-wrapped bottle at 4°C. Finally, a basal solution

was prepared by adding the ingredients (detailed in Table 2-12) to a large conical flask and stirring with boiled deionised water until fully incorporated. It was carefully transferred to a volumetric flask and topped up to the required volume using boiled deionised water.

Table 2-12 Contents of the Model Colon experiment fermentation media

Basal Media preparation

Trace mineral solution	Quantity per 1 L	Source	Cat. No.
Solvent: HCl (0.02M)	1 L	Thermo Fisher Scientific	15676840
Manganese chloride ($MnCl_2 \cdot 4H_2O$)	25.0 mg	Sigma-Aldrich	M3634
Ferrous sulphate ($FeSO_4 \cdot 7H_2O$)	20.0	Sigma-Aldrich	F7002
Zinc chloride ($ZnCl_2$)	25.0	Sigma-Aldrich	793523
Copper chloride ($CuCl \cdot 2H_2O$)	25.0	Sigma-Aldrich	307483
Cobalt chloride ($CoCl_2 \cdot 6H_2O$)	50.0	Sigma-Aldrich	255599
Selenium dioxide (SeO_2)	50.0	Sigma-Aldrich	213365
Nickel chloride ($NiCl_2 \cdot 6H_2O$)	250.0	Sigma-Aldrich	223387
Sodium molybdate ($Na_2MoO_4 \cdot 2H_2O$)	250.0	Sigma-Aldrich	331058
Sodium metavanadate ($NaVO_3$)	31.4	Sigma-Aldrich	590088
Boric acid (H_3BO_3)	250.0	Sigma-Aldrich	31146
Fatty acid solution	**Quantity per 1 L**	**Source**	**Cat. No.**
Solvent: NaOH (0.2M)	1 L	Sigma-Aldrich	06203
Acetic acid	6.85 mL	Sigma-Aldrich	A6283
Propionic acid	3.00 mL	Sigma-Aldrich	402907
Butyric acid	1.84 mL	Sigma-Aldrich	B103500
Isobutyric acid	0.47 mL	Sigma-Aldrich	58360
2-Methylbutyric acid	0.55 mL	Sigma-Aldrich	W269514
Valeric acid	0.55 mL	Sigma-Aldrich	240370
Isovaleric acid	0.55 mL	Sigma-Aldrich	129542

Resazurin solution	Quantity per 1 L	Source	Cat. No.
Solvent: Boiling deionised water	1 L		
Resazurin	1.0 g	Sigma-Aldrich	R7017
Haemin solution	**Quantity per 1 L**		
Solvent: NaOH (0.05M)	1 L	Sigma-Aldrich	06203
Haemin	100.0 mg	Sigma-Aldrich	51280
Potassium hydroxide solution (6M)	**Quantity per 1 L**	**Source**	**Cat. No.**
Solvent: deionised water	1 L		
Potassium hydroxide (KOH)	336 g	Sigma-Aldrich	
Basal solution	**Quantity per 1 L**	**Source**	**Cat. No.**
Solvent: Boiling deionised water	963.16 L		
Potassium chloride (KCl)	0.7134 g	Sigma-Aldrich	P3911
Sodium chloride (NaCl)	0.7134 g	Sigma-Aldrich	S7653
Calcium chloride dihydrate (CaCl$_2$·2H$_2$O)	0.2378 g	Sigma-Aldrich	223506
Magnesium sulphate heptahydrate (MgSO$_4$·7H$_2$O)	0.5945 g	Sigma-Aldrich	M1880
PIPES buffer	1.567 g	Sigma-Aldrich	P6757
Ammonium chloride (NH$_4$Cl)	0.642 g	Sigma-Aldrich	A9434
Bacto™ Trypticase Peptone	1.189 g	VWR International	1.07213.1000
Reagents:			
Trace Mineral solution	11.89 mL		
Fatty Acid solution	11.89 mL		
Haemin solution	11.89 mL		
Resazurin solution	1.17 mL		

KOH	pH adjustment to 6.8		
Fermentation Vessel preparation			
Vitamin Phosphate solution	**Quantity per 1 L**	**Source**	**Cat. No.**
Solvent: Potassium phosphate solution			
Deionised water	1 L		
Potassium phosphate monobasic (KH_2PO_4)	54.7 g	Sigma-Aldrich	P9791
Reagents:			
Biotin	20.4 mg	Sigma-Aldrich	B4639
Folic acid	20.6 mg	Sigma-Aldrich	F8758
Calcium dpantothenate	164.0 mg	Sigma-Aldrich	P5155
Nicotinamide	164.0 mg	Sigma-Aldrich	72340
Riboflavin	164.0 mg	Sigma-Aldrich	R9504
Thiamine HCL	164.0 mg	Sigma-Aldrich	T4625
Pyridoxine HCl	164.0 mg	Sigma-Aldrich	181986
P*ara*-amino benzoic acid	20.4 mg	Sigma-Aldrich	A9878
Cyanocobalamin (vitamin B12)	20.6 mg	Sigma-Aldrich	C3607
Sodium Carbonate solution	**Quantity per 1 L**	**Source**	**Cat. No.**
Solvent: Boiling deionised water	1 L		
Sodium carbonate (Na_2CO_3)	82 g	Sigma-Aldrich	223484
Sodium Carbonate + Vitamin Phosphate solution	**Quantity per 75 mL**	**Source**	**Cat. No.**
Sodium carbonate solution	60 mL		
Vitamin phosphate solution	15 mL		
Reducing agent	**Quantity per 1 L**	**Source**	**Cat. No.**
Solvent: Boiling deionised water	1 L		
L-Cysteine HCl	20.0 g	VWR International	ACRO434850010
$Na_2S \cdot 9H_2O$	20.0 g	Sigma-Aldrich	431648

Final vessel constituents	Quantity per vessel		
Basal solution	76 mL		
Reducing agent	1 mL		
Sodium Carbonate + Vitamin Phosphate solution	5 mL		
Thermo Fisher Scientific™, Loughborough, UK Sigma-Aldrich, Dorset, UK VWR International, Poole, UK			

The basal solution was transferred to the large conical flask again, and the following solutions were added (mL per 1L of media): Resazurin solution (1.174 mL), Trace Mineral solution (11.891 mL), Haemin solution (11.891 mL), Fatty Acid solution (11.891 mL). A concentrated solution of 6M KOH was made by dissolving 33.6 g KOH in 100 mL deionised water and stored in a sealed Duran bottle. Using a pH meter, the media was adjusted to 6.8 with the concentrated KOH, turning the media from pink to deep purple. To reduce the media, CO_2 was bubbled through the solution until cooled and pink, 0.5 - 1 hour(s) per L. Media was then divided in volume into conical flasks (starch and no substrate control), 0.5% w/v HylonVII starch was added to one flask and boiled to 100 °C for 5 min on a stirring plate.

2.2.6.3 Vessel preparation

Once media was reduced with CO_2 and cooled, 76 mL basal media was dispensed into clean and autoclaved serum bottles. The septa and aluminium caps were crimped and autoclaved at 121 °C for 15 min. For 0.5% starch media, the bottles were used the same day for experiments (to prevent starch retrogradation) while the no substrate control media was stored at 4 °C until subsequent use.

To prepare the vessels for inoculation, two final solutions – vitamin buffer solution and reducing solution – were prepared. First, a sodium carbonate solution was made by adding 41 g Na_2CO_3 (sodium carbonate anhydrous) to 500 mL boiled deionised water. The solution was bubbled with CO_2 for around 30 min. 60 mL of the solution was dispensed into serum bottles, capped, crimped, and autoclaved at 121 °C for 15 min. Bottles were stored at room temperature. Vitamin phosphate solution was prepared in a beaker by dissolving 27.35 g KH_2PO_4 in 500 mL deionised water, followed by adding each chemical in order (Table 2-12), only adding the next when the previous was fully dissolved. This was filter-sterilised into sterilised serum bottles, septa were inserted, and bottles were capped and crimped; they were stored at 4 °C until use. Finally, to bottles containing 60 mL autoclaved sodium carbonate solution, 15 mL vitamin/phosphate solution was added by injecting into the septa with a sterile needle and syringe. A reducing solution was made by combining 1 g L-Cysteine HCl and 1 g $Na_2S·9H_2O$ was weighed and dissolved in 50 mL boiled deionised water. The pH was adjusted using pH paper to pH 10. This solution was used on the day of vessel prep only but was stored at 4 °C until use. Vitamin phosphate solution was added to autoclaved media by injecting 15 mL into the septa of serum bottles with a sterile syringe and 23G needle. The same was performed to add 1 mL of the reducing solution to each serum bottle using a sterile.

One medium bottle (from each medium type) was sacrificed (cap removed) to ensure pH was ~ 6.5 – 7 using a pH meter. The pH of the media was adjusted accordingly using KOH or HCl, noting down the adjustment volume required and using a needle and syringe, adding the same volume of acid/base to each bottle.

Finally, the serum bottles containing the pH corrected experimental media were purged with CO_2 for 2-3 mins using plastic tubing with sterile needles attached and inserted into the septa of bottles and ensuring there was an outlet needle to prevent gas pressure build-up. The bottles were incubated at 37 °C overnight in an aerobic incubator.

2.2.6.4 Experiment procedure

The model colon media and vessels were prepared to accommodate duplicate experimental conditions outlined in Table 2-13.

Table 2-13 In vitro *batch fermentation studies: treatments and variables summarised. All conditions were performed using two biological replicates.*

Experiment summary	Fermentation of starch using simulated infant and adult *in vitro* model colon
Substrate	Negative; HylonVII
Donor age	Infant (3 donors); Adult (3 donors)
Bacterial supplement	*B. pseudolongum* NCIMB 702245 *B. longum* subsp. *infantis* LH277
Sequencing timepoints	0h, 24h
Library prep by	Quadram Institute Bioscience Sequencing
Sequencing by	Novogene, Cambridge, UK
Sequencing Technology	Illumina NovaSeq™ 6000 Sequencing System
NMR timepoints	0h, 8h, 24h, 48h

For preparing faecal inocula, sterile PBS was pre-reduced for 24 h in an anaerobic cabinet. The donor stool was defrosted overnight in a 4 °C fridge the day before inoculation. Once defrosted, the stool was weighed using a tongue depressor into a stomacher bag in a Class II microbiological safety cabinet (at least 5 g) and a 1:10 slurry was made by adding pre-reduced, pre-warmed (37 °C) PBS e.g., for 5 g of stool, 50 mL PBS was added. The stomacher bag was placed into the stomacher and mixed for 30 seconds. Inoculation of serum bottles was carried out

using a 19/18G needle and syringe by injected 3 mL of slurry into the serum bottle containing medium, buffers, and substrate.

For the bacterial strain supplementation, cultures of *Bifidobacterium pseudolongum* NCIMB 702245 and *B. longum* subsp. *infantis* LH277 were cultured. Frozen 30% glycerol stocks of bacterial strains were aseptically scraped and suspended in 10 mL MRS broth and incubated anaerobically for 48 h at 37 °C in a Don Whitely Scientific minMACS anaerobic incubator. The day before (24 h) inoculation, the cultures were pre-cultured under the same conditions (temperature, incubator) for 24 h in MRS broth autoclaved with 1% w/v normal maize starch. Bacteria were injected into the media (to maintain anaerobic conditions) in duplicate with a 1:100 dilution for inoculation (e.g., 850 μL of inoculum per 85 mL medium), changing the needle each time to avoid cross-contamination.

2.2.6.5 Metabolomics sampling and processing

1 mL of model colon medium was removed at 0, 8, 24, 48h, centrifuged and the supernatant removed for NMR metabolomics protocol and data processing as described in Section 2.2.7. Remaining supernatant and pellet were retained and frozen at -80 °C. The samples were processed for NMR quantification and data was processed and analysed as described in Section 2.2.7.2.

2.2.6.6 Metagenomics sampling procedure and processing

At timepoints 0 h and 24 h, 1 mL of fermentation liquid was removed using a sterile 18G hypodermic needle and sterile syringe and transferred to lysing matrix E tubes. Tubes were frozen at -80 °C until needed for bacterial DNA extractions,

sample QC, library preparation sequencing, and data processing, all carried out as described in Section 2.2.6.7.

The Maxwell® RSC 48 Instrument was utilised to expedite DNA extraction with the assistance of Fred Warren, Cara Moss, and Hannah Harris. This platform uses technology which automates DNA or RNA extraction and has been validated previously (Phillips, Watt et al. 2021). The Maxwell instrument utilises prefilled cartridges containing reagents including paramagnetic particles which associate to the nucleic acids. By moving magnetic particles instead of liquid, the risk of contamination is lower, and the process yields purer, consistent extracts. Extractions were performed using Maxwell RSC Purefood GMO and Authentication Kit according to the kit instructions as follows. Samples were defrosted at room temperature for 45 minutes and centrifuged for 5 min at 13,000 rcf. For a negative control sample, 200 uL molecular grade H_2O was added to a Lysing Matrix E tube. The supernatant was carefully removed by pipette and discarded. Using the kit reagents, 1 mL CTAB was added to each sample and vortexed for 30 seconds, before heating in a hot block at 95 °C for 5 min. Samples were vortexed again for 1 minute. Samples were homogenized in the FastPrep Instrument for 45 seconds at a speed setting of 6.0 m/s and pulse centrifuged for 3 seconds to reduce bubbles in the samples. From the kit, 40 µL of proteinase K and 20µl of RNase A was added to samples and vortexed to mix. Samples were heated in a hot block at 70 °C for 10 minutes. The samples were then centrifuged at 14,000 x g for 5 minutes to pellet debris. The Maxwell cartridges were loaded into the instrument and 300µl of lysis buffer was added to the first well of the cartridge. Elution buffer (100 µL) was added into supplied elution tubes and placed

inside the instrument rack. The supernatant (300 µL) from each sample was added to well 1 of the cartridges. Maxwell RSC program was selected corresponding to the kit used (Maxwell RSC Purefood Protocol).

2.2.6.7 Metagenomics Library Preparation and Sequencing

DNA extracts were prepped exclusively by Rhiannon Evans according to short-read sequencing protocols described in Section 2.2.3.3, except the final sequencing steps were performed by Novogene. Secondary library QC was performed by Novogene using Qubit fluorometry for quantification and real-time PCR combined with a bioanalyzer (unspecified) for size distribution detection. Paired end 150 (PE150) partial lane sequencing was performed using a NovaSeq 6000 and NovaSeq 6000 Reagent Kits v1.5 using sequencing cycles 151+10+10+151. The average number of raw reads per (non-blank) sample was 11,025,497.

2.2.6.8 Metagenomics Data Processing

Data QC carried out by Novogene included discarding the adapter containing reads and the paired reads when uncertain nucleotides constitute more than 10% of either read. In addition, paired reads containing more than 50% low quality nucleotides for either read were discarded.

Bioinformaticians Andrea Telatin and Sumeet Tiwari performed data processing protocols. Raw reads clean-up involves aligning ("mapping") every read produced to a host genome (in this case human) using Kraken and removing them from the dataset (Salzberg and Wood 2021). These host-filtered microbial reads undergo classification using Kraken 2, whereby a database containing sequences

belonging to bacterial, viral, fungal, archaeal, and animal genomes, exact-matching sequencing reads in the dataset to a species level classification. Instead of alignment of reads based on sequence similarity to taxa in the database, Kraken identifies reads by exactly matching reads to short sequences of length k, called k-mers (Salzberg and Wood 2021). The number of occurrences of sequences of k length which match to a lowest common ancestor are counted (Marcais and Kingsford, 2011). The taxon information is assigned from the NCBI taxonomy database to produce a Kraken 2 report containing the taxon identity to which the k-mer is exactly matched, and the number of times the read is matched to that taxon (Salzberg and Wood 2021).

Bracken is the final read processing tool used to estimate the species abundance, essentially by adjusting the reads identified by Kraken 2 (Lu, Breitwieser et al. 2017). The efficacy of the program to detect species is reliant on the completeness of the genome database used, however targeted alignments can be performed to search for specific genome alignments, if the whole-genome sequence is available.

Upon delivery of the final Bracken read table containing how many reads from each taxon and their relative abundance value, I performed the remaining steps: The data was transferred to R where it was further processed using packages phyloseq and phylosmith to visualise the data (McMurdie and Holmes 2013, Smith 2019). R package Maaslin2 was used to perform statistical associations between experimental variables and microbial metagenomic and metabolomic features (Mallick, Rahnavard et al. 2021).

2.2.7 Metabolomics: Nuclear Magnetic Resonance Spectroscopy

2.2.7.1 NMR Methodology

Nuclear Magnetic Resonance (NMR) Spectroscopy is a quantitative method used to characterise organic compounds. It utilises the predictable physical behaviours of individual nuclei within organic substances when a strong magnetic force is applied. An NMR spectrometer uses a powerful, superconducting magnet to create a magnetic field. The sample is placed in the epicentre of the field inside an NMR probe where it is excited, and the frequency of the signals returned is recorded by the probe. The frequencies and the intensity recorded by the instrument are representative and proportional to specific molecules and their relative concentration.

The principles of NMR rely on magnetism of atomic nuclei which is possible due to their spin and charge properties. Nuclei of atoms can be excited by a magnetic field settling at equilibrium in the direction of the applied field, *z* (Figure 2-2) (Keeler 2010). For this first step, the superconducting magnet provides a constant field to align every magnetised nucleus in the sample in the same direction.

The second step involves tilting nuclei off their alignment using radiofrequency pulses. A coil present in the NMR probe can deliver radiofrequency pulses along the *x* axis at a resonant frequency (Figure 2-2). When the weak oscillating pulse occurs, the magnetised nuclei tilt towards the *x* and *y* axes in a spinning-top motion. A *pulse sequence* can repeatedly, temporarily rotate the magnetisation away from the *z* axis, causing nuclear spin known as precession. Several *scans* are necessary for pulse sequences to generate multiple acquisitions and combine

them in order to remove unwanted noise in

the data. A feature of the spectrometer is to

execute the pulses and period between

scans, defined as the *relaxation delay*

(Keeler 2010).
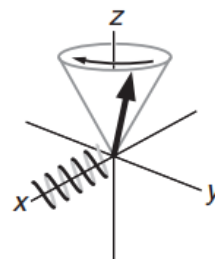
Precession is detected by the NMR probe.

The motion of the magnetised nuclei creates

a current through a coil within the probe,

known as the free induction signal. The

tendency of the nucleus to return to

equilibrium on the *z* axis is called relaxation.

Relaxation, a process which occurs over

time, is known as the free induction decay.

Figure 2-2 Nuclei spin can be magnetised to equilibrium along a z axis, the direction of the applied field. Radiofrequency pulses can tilt the nucleus towards the x or y axis, generating a current through a coil. This current is detected by the NMR probe. Figure used with permission from John Wiley and Sons, "Understanding NMR Spectroscopy", Keeler (2002); permission conveyed through Copyright Clearance Center, Inc.

As the nuclei equilibriate, the amplitude of the current generated in the probe coil

proportionally declines. The evolution of the magnetisation – the recorded

frequencies of nuclei transitioning from precession to free induction decay – is

known as *acquisition time*. The oscillation frequency generated is amplified and

recorded by the probe (Keeler 2010).

The NMR probe can detect specific chemical signatures since the angle at which

the nuclei are tilted is specific to chemical arrangements within molecules. A

molecule's peak intensity is proportional to concentration. The NMR buffer is

prepared to include a reference molecule, 3-(trimethylsilyl)-propionate-d_4 (TMSP),

whereby all molecules can be quantified. For ^{1}H NMR, TMSP has a chemical shift

which is zeroed at 0 ppm such that the other peaks are measured relative to this reference. The buffer employs deuterated solvents such as D_2O which minimises the protons from the water solvent peak in ^1H NMR spectra obscuring peaks of interest.

NMR was used to reliably assess metabolite production by bacteria during starch degradation. Identification and quantification of SCFAs and sugars provides information about bacterial phenotype by indicating metabolic processes occurring and rate of starch breakdown by bacteria over the course of starch fermentation.

2.2.7.2 NMR sample preparation

An NMR Buffer was prepared according to previous work (Crost, Le Gall et al. 2018) using 100 mL D_2O and concentrations of reagents detailed in Table 2-14. For preparation of 100 mL of buffer, the following was added to a clean glass, twist-seal capped bottle: 0.26 g NaH_2PO_4, 1.44 g K_2HPO_4, 17.0 mg TMSP-d_4, 56.1 mg NaN3. A final volume of 100 mL D_2O was added and vortexed until solubilised. The bottle was sealed with parafilm until needed and between uses.

Culture medium aliquots of 600 mL were stored in 1.5 mL snap-cap plastic microtubes at -20 °C (if not immediately processed). If frozen, the aliquots were thawed at room temperature or in 4 °C overnight. Each sample was centrifuged at 13,000 rpm for 3 min. 400 µL of the supernatant was pipetted into a labelled NMR tube (Norell® Standard Series™ (5 mm), Cat. No. NOR509UP7-5EA, Sigma-Aldrich, Dorset, UK), followed by 200 µL NMR Buffer.

2.2.7.3 Instrument parameters

Protocols used in this thesis were designed by Trey Koev and Sergey Nepogodiev. Metabolomics protocols were executed by me with some assistance from the two NMR scientists and Fred Warren. For identification and quantification of metabolites present in fermentation media were recorded on a Bruker Avance II NMR spectrometer, operating at a ^{1}H frequency of 500.11 MHz equipped with an inverse triple resonance z-gradient probe. Spectra recorded were obtained using two different pulse sequences, used interchangeably due to data output being identical, Bruker's 'noesygppr1d' and '*zggppewg*'. All other parameters were: 128 scans, a spectral width of 7500 Hz and an acquisition time of 4.4 s, and relaxation delay of 10 s.

Table 2-14 NMR buffer ingredients and concentrations, including supplier information.

NMR buffer	Volume	Source	Cat. No.
Solvent: Deuterium oxide, 99.9% atom % D	100 mL	Sigma-Aldrich	151882
	Concentration		
Monopotassium phosphate, KH$_2$PO$_4$, ≥99.0%	82.7 mM	Sigma-Aldrich	P5655)
Sodium phosphate monobasic NaH$_2$PO$_4$	21.7 mM	Sigma-Aldrich	71507
Sodium azide (NaN$_3$) BioXtra	8.6 mM	Sigma-Aldrich	S8032
3-(trimethylsilyl)-propionate-d4 (TMSP)	1.0 mM	Sigma-Aldrich	269913

Spectra were recorded for RNA-seq experiments and model colon fermentation assays (Chapters 4 and 5) using a Bruker Avance NEO 600 MHz instrument with cryoprobe operating at a ^{1}H frequency of 600.17 MHz. Spectra were recorded using the same pulse sequence as before Bruker's 'noesygppr1d' however using 64 scans, spectral width of 12500 Hz, an acquisition time of 2.6 s, and relaxation

delay of 4 s. These parameters were an improvement on previous NMR experiments due to using a more powerful magnet, meaning a reduction of scans, acquisitions, and delays has minimal impact on the data output but being able to run more experiments in the same amount of time.

2.2.7.4 Metabolomics Data Analysis

Using NMR Suite Processor v8.41 (Chenomx®, Edmonton, Canada), the software converts the signal recorded to intensity as a function of frequency, taking into consideration the time it takes for free induction decay, using the Fourier transformation. This converts the raw frequency spectra into chemical shift peaks. If necessary, the TMSP reference peak was baseline corrected to 0.0 ppm which allows quantification of other molecules recorded on the spectrum.

As discussed in Section 2.2.7.1, a molecule's peak intensity is proportional to concentration. Once the files were processed into spectra containing peaks, the software can automatically fit the spectra for each molecular peak pattern from a database. Metabolites detected on at least one occasion were acetate, alanine, butyrate, cysteine, ethanol, formate, glucose, glycine, isobutyrate, isoleucine, isovalerate, lactate, leucine, maltose, methionine, phenylalanine, propionate, pyroglutamate, succinate, threonine, tryptophan, tyrosine, valerate, and valine. Following this, each detected metabolite peak was manually adjusted in Chenomx Profiler.

The data was exported to Microsoft Excel where the concentrations were adjusted based on the dilution factor due to addition of NMR buffer. The adjusted concentrations were copied to GraphPad Prism 9 for statistical analysis.

2.2.8 Whole Genome Bioinformatics

Investigation into the genomic and genetic basis of bacterial carbohydrate utilisation is a broad and well investigated research area. Whole genome sequences (WGS) for BAMBI study (LH isolates) and reference strains were acquired as described in Section 2.1.2. A final strain, *Bifidobacterium pseudolongum* subsp. *pseudolongum* DSM 20099 / NCIMB 702244 was sequenced and assembled in-house as detailed in Section 2.2.3.4; a single circular contig (genome) was assembled using methods described in Section 2.2.3.5.

2.2.8.1 Strain similarity rating

Several strains of LH isolates belonging to *B. pseudocatenulatum* were evaluated for similar using ANI values with pyani v0.2.7 (Yoon, Ha et al. 2017). The same tool was used to assess the similarity of *B. pseudolongum* subsp. *pseudolongum* 202244 = DSM20099 and *B. pseudolongum* subsp. *globosum* 202245 = DSM20092.

2.2.8.2 CAZyme profiles

The presence of relevant enzyme families tested for starch utilisation was carried out computationally using WGS of the strains. dbCAN2 is web server designed to identify annotated genes that are functionally and categorically organised on the carbohydrate-active enzyme (CAZyme) database Cazy (Zhang, Yohe et al. 2018, Wardman, Bains et al. 2022). dbCAN2 generated hits for presence of relevant CAZymes (families and subfamilies) utilising three tools: Hotpep, HMMER, and DIAMOND. HMMER and DIAMOND perform sequence alignments against the Cazy database, whereas Hotpep uses a peptide database to predict function of

the annotated genes more accurately. The hits obtained from these tools indicates the number of families and subfamilies detected in the WGS which was processed to array the number of hits for each CAZyme family, subfamily, and CBM identified and displayed by heatmap using GraphPad Prism v9.5.0.

Chapter 3

Bifidobacterium genus-wide variation in starch metabolism

3.1 Abstract

Members of the *Bifidobacterium* genus are prominent commensals of the human gut from birth. While there is significant knowledge of its niche occupancy in the infant gut, and starch degrading capabilities of *B. adolescentis* in adulthood, there is little known about the evolutionary impacts of the transition from milk to solid foods containing starch (i.e. weaning). The variables which impact starch hydrolysis and enzyme kinetics are also relatively understudied in bifidobacteria.

Detailed growth assays of 7 starch degrading strains of *Bifidobacterium* belonging to *B. breve, B. pseudocatenulatum,* and *B. pseudolongum* show that increasing amylose content significantly impacted the rate of degradation of gelatinised maize starch suggesting it is resistant to enzymatic hydrolysis. Complementary metabolomic investigation mirrored the growth kinetics showing significant increase in the health promoting SCFA acetate. There was also significant intraspecies variation in degradation of resistant high amylose maize starch underlining the importance of strain-specific investigation. Bioinformatic analyses also highlighted the differences between strains in the presence of starch CAZymes from groups GH13, CBM48 and CBM26. Four strains with starch utilisation capabilities contained RS binding CBM74 modules.

These detailed growth kinetics and metabolite production of different types of starch provide new insights in *Bifidobacterium* genus bacteria and could be informative for the design of functional pro- and pre-biotics which modulate the microbiome.

3.2 Introduction

Bifidobacterial substrate utilisation or preference is a result of their niche occupancy: strains which occupy the infant gut are suited to HMO utilisation (Lawson, O'Neill et al. 2020, Walsh, Lane et al. 2022). However, less discussed are the ecological impacts of starch in the gut on bifidobacterial populations, particularly throughout weaning where starch is a common food component (Vatanen, Ang et al. 2022). The present work tests the same bacterial isolates from a previous (HMO-focussed) study (Lawson, O'Neill et al. 2020) for starch utilisation phenotypes, to build a more complete picture about niche-shifting behaviour of *Bifidobacterium* strains in a changing dietary environment. Starch utilisation phenotypes of *Bifidobacterium* are relevant during early development, as well as in adults, where the starch present allows proliferation of *Bifidobacterium* and other important health-promoting bacteria (Belenguer, Duncan et al. 2006). A greater understanding about the interactions of *Bifidobacterium* and starch, especially different structures, will be key to developing biotherapeutics leveraging the health benefits of *Bifidobacterium* and starch in the human gut.

This chapter seeks to address starch utilisation phenotypes among a range of reference strains using a variety of different starch structures. In typical screening studies of this nature, the structure of starch is less well-considered during study

design. Starch physical structure and therefore accessibility to degradative enzymes can vary depending on composition (the proportions of amylose to amylopectin), the starch preparation (and/or sterilisation) method, or botanical origin (impacting starch crystallinity) (Wang, Hasjim et al. 2014, Soler, Velazquez et al. 2020, Zhong, Liu et al. 2020). These variables will be explored to assess their impact on growth of various species and strains of *Bifidobacterium*.

In particular, the impact of starch accessibility to bacterial enzymes and associated growth kinetics has not previously been explored in bifidobacteria. This is particularly pertinent because there has not been standardisation or consistencies for methods such as media type, starch preparation, measurement technology, or the timepoints of measurement in the current literature. The data presented in this chapter addresses how starch structure impacts growth kinetics and aims to shed light on several pitfalls and advantages of bacteria screening assays.

Complementary to the growth phenotyping, an analysis of the metabolic output of several strains was employed to strengthen evidence related to utilisation and production of metabolic end products such as acetate and ethanol. Additionally, the production of reducing sugars during the degradation process is a proposed mechanism for how primary degraders of starch facilitate cross-feeding of other bacterial taxa (Belenguer, Duncan et al. 2006, Crost, Le Gall et al. 2018). Both production of acetate and sugars enable cross-feeding which can promote production of butyrate (Rios-Covian, Gueimonde et al. 2015, LaBouyer, Holtrop et al. 2022). Butyrate and acetate are SCFAs well known to promote health (Makki, Deehan et al. 2018). This chapter therefore combines the growth kinetics and

metabolomic studies to discuss how starch structure impacts metabolism of bacteria.

The hypotheses proposed for this chapter are firstly that there will be a range of abilities to utilise starch depending on the ecological niche the species or strain naturally occupies. For example, early gut colonisers such as *B. bifidum* would be expected to utilise starch least well whilst bacteria which persist in later life would have starch degrading capabilities e.g. *B. breve*. Secondly, strains which utilise starch would have variation in capability to degrade more resistant forms of starch and that their capabilities would be connected to the presence of GH13 and CBMs in their genomes. Lastly, the metabolites produced may indicate how the bacteria are able to metabolise starch and confirm the health-promoting aspects of starch fermentation by *Bifidobacterium*.

3.3 Methods

Two methods were used to assess starch utilisation phenotypes of bifidobacterial strains by measuring an increase in growth in the presence of different starches detailed in Table 3-1.

Table 3-1 A summary of starches used in bacterial growth assays throughout this thesis and their properties.

Substrate	Abbr.	Features
Soluble potato starch	SPS	Soluble in water once gelatinised. Accessible when gelatinised and resistant when granular.
Normal maize starch maize	NMS	Soluble in water once gelatinised.
High amylose maize starches	HAM HylonVII	Higher amylose affects starch structure after gelatinisation and increases gelation propensity.

The media preparation, culturing, and quantification methods (OD plate reader and Colony forming unit measurement) are described in Section 2.2.1. The strains tested are detailed in Section 2.1.2. Details of each set of assays is shown in Table 3-2. Metabolomics studies were performed as described in Section 2.2.7. Whole genome sequencing and bioinformatics were carried out according to methods described in Section 2.2.8.

Table 3-2 Experimental overview of phenotyping studies using reference strains and unique isolates derived from (Lawson, O'Neill et al. 2020) (LH strains). All % are w/v. Strains tested are detailed in Section 2.1.2.

Strains	Assay	Substrates	Controls	Replicates
Reference	plate reader 1	1% + 2% autoclaved SPS	2% lactose, no substrate	Triplicate
Reference	plate reader 2	0.25%, 0.5%, 1%, + 2% autoclaved SPS	1% lactose, no substrate, MRS	Duplicate
Reference	plate reader 3	2% autoclaved SPS, 2% maltose, 2% glucose, 2% lactose	No substrate, MRS	Duplicate
Reference	plate reader 4	1% autoclaved SPS, NMS, and HylonVII	No substrate	Triplicate
LH strains	plate reader 5	2% autoclaved SPS, 2% autoclaved NMS	2% glucose, no substrate	Triplicate
Reference + LH strains	CFU-1	1% autoclaved NMS and HAM	1% glucose, no substrate	Triplicate
Reference + LH strains	CFU-2	1% autoclaved NMS and HylonVII	1% glucose, no substrate	Triplicate
Reference + LH strains	CFU-3	1% autoclaved SPS and 1% granular (raw) SPS	No substrate, +1% glucose	Triplicate

As previously mentioned in Section 2.1.2 and discussed in this chapter, some LH strains isolated from the same infant at the same timepoint are very closely related and are probably clonal strains, with minor single nucleotide polymorphism differences in their genomes. Thus, where several clonal strains were tested, the

data were combined into one graph as the results were replicates of the same strain and statistical tests were performed including these additional replicates. Where this has been done, the strain identifiers are listed separated by a forward slash e.g. *B. pseudocatenulatum* LH657/659.

3.4 Results

3.4.1 Plate reader assays of reference strains

Initial assays of *Bifidobacterium* reference strains were carried out to optimise the method of high-throughput measurement of bacterial growth. Plate reader assays 1 and 2 test the phenotype of strains for starch utilisation of different concentrations (Table 3-2). Across two experimental runs using the plate reader, the strains were tested using 2% w/v, 1% w/v, 0.5% w/v and 0.25% w/v soluble potato starch (SPS) (Figure 3-1). Four strains were selected from the literature identified as having starch utilisation capabilities (two each from *B. breve* and *B. pseudolongum)* as well as two negative control strains from *B. longum* and *Lacticaseibacillus casei* which have previously been shown not to utilise starch (Ryan, Fitzgerald et al. 2006). For one of the plate reader experiments, a third control strain was included (*B. longum* subsp*. infantis* 20088*)*.

For both concentrations, *B. pseudolongum* subsp*. pseudolongum* NCIMB 702244 (*B. pseudolongum* 44) consistently achieved stationary phase earlier than any other strain, whereas *B. breve* UCC2003 achieved the highest OD readings. After 24 h, all strains of *B. breve* and *B. pseudolongum* had significantly higher growth on both 1% and 2% SPS compared to the negative control, with no significant difference in growth between the two concentrations (Figure 3-2 and

Supplementary Table S1). Meanwhile, neither of the negative strains (*L. casei* 4114 or *B. longum* 8809) had significantly higher growth on starch relative to the negative control (Figure 3-2 and Supplementary Table S1).
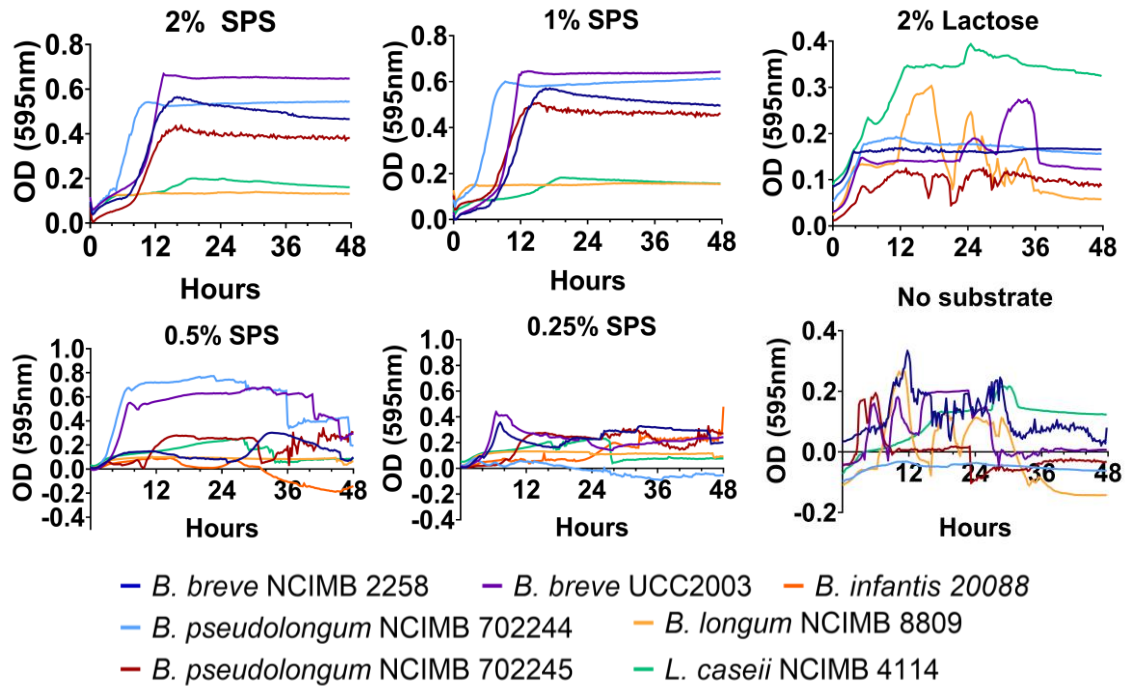
Figure 3-1 Growth of reference strains on varying concentrations of soluble potato starch (SPS) was measured by reading OD at 595nm. All % concentrations shown are w/v. The growth curves represent the mean of two duplicate OD values taken every 15 minutes over the course of 48h.

From these results, there are observed intra-species variations in starch utilisation: *B. pseudolongum* 44 was a better degrader of autoclaved soluble potato starch than *B. pseudolongum* subsp. *globosum* NCIMB 702245 (*B. pseudolongum* 45), and the same trend was observed between *B. breve* UCC2003 and *B. breve* 2258. These strains were all competent utilisers of SPS with varying degrees of capability. The lactose control, which was intended as a positive control, did not promote growth of any strains aside from *L. casei* (Figure 3-2).
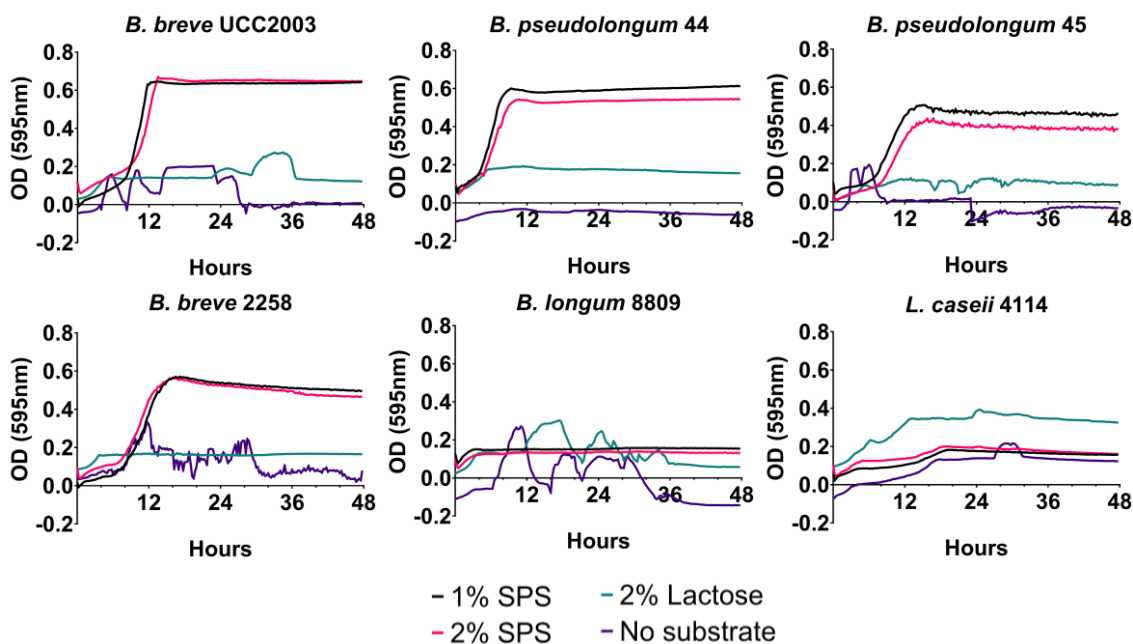
Figure 3-2 Growth curves of individual strains on 1 % or 2% soluble potato starch (SPS) utilisation by strains of Bifidobacterium *and* Lacticaseibacillus *as a negative control. Measurements were performed every 15 minutes and the mean of three replicates is displayed for each time point.*

Below 1% starch, the OD readings became less reproducible making the trends less clear (Figure 3-1). To address this, these strains were assayed (plate reader assay 3,) for utilisation of the sugars maltose, glucose and lactose each at a concentration of 2% w/v as well as 2% autoclaved SPS (Figure 3-3) using the same method and plate reader. Compared to the negative control (no substrate), there were no statistically significant differences for any of the sugars or starch at the 24 h timepoint (Table 3-3), likely due to the inconsistent reads produced. The results of the first plate reader assay were not reproduced and was deemed invalid. Additionally, the OD readings from different concentrations of starch and sugars showed that the OD reading of blank media with no bacteria increases as starch concentration increased (Figure 3-3). Since these readings remained consistent over 48 h, issues with erratic readings may have been due to bacteria and not solely starch turbidity or particle settling.

Table 3-3 Plater reader assay statistical summary. The associated p-values from statistical testing of strain OD at 24h (54h for plate reader 4) relative to the negative control with a two-way ANOVA Tukey's multiple comparisons test.

	Plate reader assay				
	1	3		4	
Strain	1% SPS	2% SPS	2% Maltose	1% NMS	1% SPS
B. breve 2003	<0.0001	0.9796	0.3758	0.0039	<0.0001
B. breve 2258	0.0002	0.1237	0.4778	<0.0001	<0.0001
B. pseudolongum 44	<0.0001	0.2379	0.9013	0.0029	0.0079
B. pseudolongum 45	<0.0001	0.9452	0.2164	<0.0001	0.8227
B. longum 8809	0.8319	0.5691	0.5087	0.9984	0.5809
B. infantis 20088		0.2392	0.9997	0.0001	0.7327
L. casei 4114	0.9999	0.9423	0.9993	0.0298	0.0665

A final assay (plate reader assay 4, Table 3-2) using the reference strains compared the utilisation of 1% SPS, normal maize starch (NMS), and a high amylose maize starch (HylonVII) after 54 h (due to equipment failure) (Figure 3-4). The data reproduced the findings of the first starch assay that *B. breve* and *B. pseudolongum* had significantly higher growth on SPS and NMS, except for *B. pseudolongum* 45 which only showed significantly higher growth on NMS (Table 3-3). *B. pseudolongum* 45 exhibited significantly higher growth on normal maize compared to potato starch under the same conditions (p=0.0006) (at 54h) (Figure 3-4). There was no significant difference between soluble potato or maize starch for *B. pseudolongum* 44 (p=0.986). *B. infantis* 20088 and *L. casei* 4114 exhibited significantly lower growth on NMS than the no substrate control (Table 3-3). HylonVII results were discarded: due to extensive coagulation of starch in the media when the blank media reading was subtracted from the 54 h reading all values were negative (Figure 3-4).
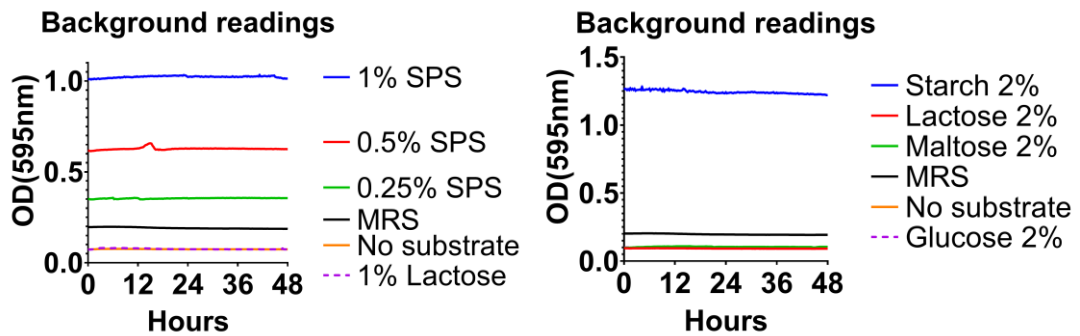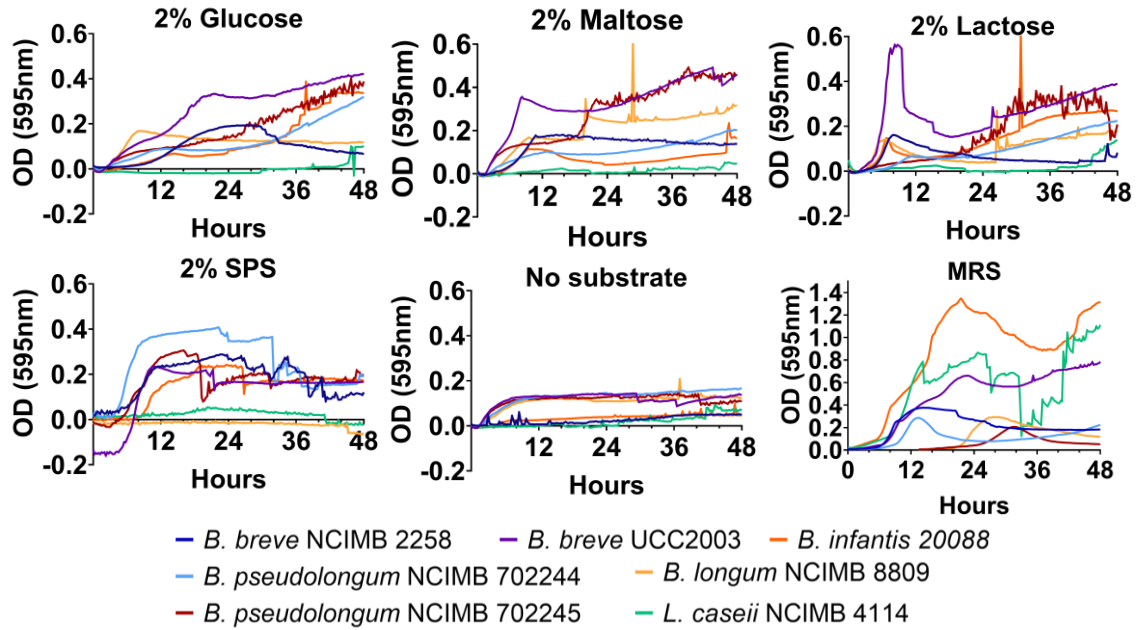
Figure 3-3 [Top panel] Growth curves showing utilisation of various substrates including soluble potato starch (SPS) and sugars; each measurement every 15 minutes is displayed as the mean of biological duplicates. [Bottom panel] Background readings of the media containing varying concentrations of substrate were measured in duplicate over the course of 48h in the absence of bacterial inoculum.

3.4.2 Plate reader assay of LH strains

Assaying of isolates from a previous study of *B. pseudocatenulatum* (6 total) and

B. breve (1) using 1% soluble potato starch and normal maize starch was

performed (plate reader assay 5, Table 3-2) (Lawson, O'Neill et al. 2020). This

showed that one strain (LH656) achieved higher overall growth but growth on SPS

or NMS was not significantly different from the control media containing no

substrate (p=0.1926, p=0.9999, respectively) (Figure 3-5).
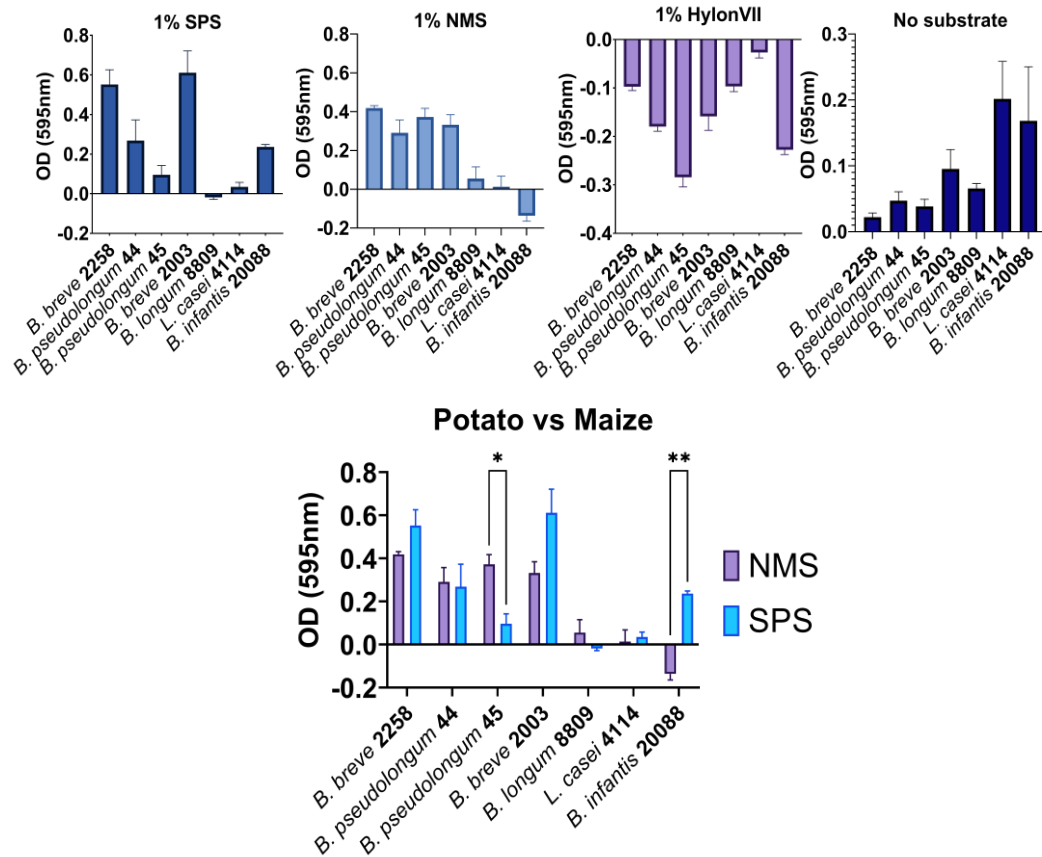
Figure 3-4 Normal maize starch (NMS) and high amylose maize starch HylonVII utilisation by bacteria is displayed as the mean of duplicate OD values taken after 54h of culture. **p<0.01 *p<0.05.

Strains isolated from the same infant at the same time point (age) were proposed to be clonal (Lawson, O'Neill et al. 2020) which was confirmed post hoc by bacterial genomic analysis by authors indicating that *B. pseudocatenulatum* LH662, LH663, LH656, LH657, LH658, and LH659 have only very small differences in Average Nucleotide Identity (ANI) values (>99.99%) (Table 3-4). These starch utilisation results are consistent with the fact the strains are clonal, although LH656 mean growth at 24 h for any substrate condition was significantly higher than every other strain (p<0.0001), suggesting it was more tolerant to lab culture. However, the individual growth curves for each strain were unusual,

further calling into question the validity of this technique for measuring starch

utilisation of bacteria (Figure 3-5).
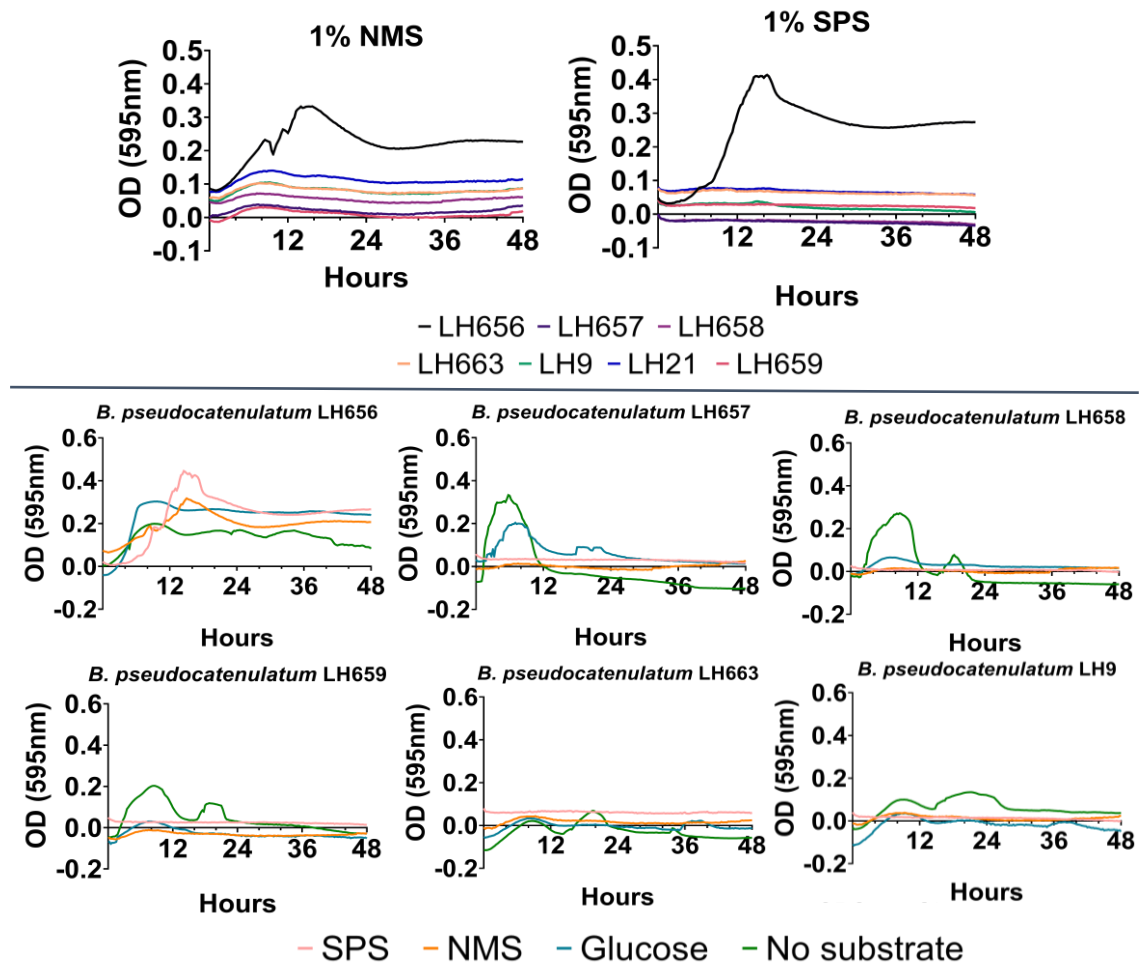
Figure 3-5 Unique isolates (LH strains) measured using plate reader for SPS and NMS utilisation, including glucose as a control. All substrates were supplemented into the media at a concentration of 1% w/v). Values displayed are the mean of duplicate OD (595nm) readings taken every 15 minutes.

Table 3-4 Whole-genome similarity of LH strains of B. pseudocatenulatum *as Average Nucleotide Identity (ANI) scores. A score of 1.00 denotes 100% sequence similarity (green); a similarity score > 1.00 denotes less similarity (yellow). The least similar strains are LH656 and LH657.*

Strain	LH656	LH657	LH658	LH659	LH662	LH663
LH656	1.00000000					
LH657	0.99994515	1.00000000				
LH658	0.99998963	0.99996040	1.00000000			
LH659	0.99998824	0.99999334	0.99999290	1.00000000		
LH662	0.99998558	0.99995832	0.99999184	0.99999091	1.00000000	
LH663	0.99998926	0.99996107	0.99999290	0.99999286	0.99999176	1.00000000

3.4.3 Screening bacterial growth using colony counting

Following reproducibility issues and assessment of high background readings of starch media, phenotyping of strains was continued using a colony counting approach. A combination of reference strains and LH strains were tested using 1% autoclaved NMS and 1% autoclaved high amylose maize (HAM), including 1% glucose and no substrate for controls (Figure 3-6). Out of nine strains tested (by combining clonal strains LH657-662 and LH36-37), 6 were found to have statistically significantly higher growth using either NMS or HAM, confirming the capability to degrade starch to provide energy for growth found in plate reader assays (Figure 3-6). The significance ranged from moderately significant (p = 0.0405) in *B. breve* LH24 to highly significant (p <0.0001) in *B. pseudolongum* 45, *B. pseudocatenulatum* LH strains, and *B. breve* LH36/37.

While *B. breve* 2258 had increases in growth on both NMS and HAM, these results were not statistically significant (Figure 3-6); it is possible *B. breve* 2258 prefers degrading potato starch as opposed to maize. Opposite to this, there were highly significant increases in the growth of *B. pseudocatenulatum* LH strains on
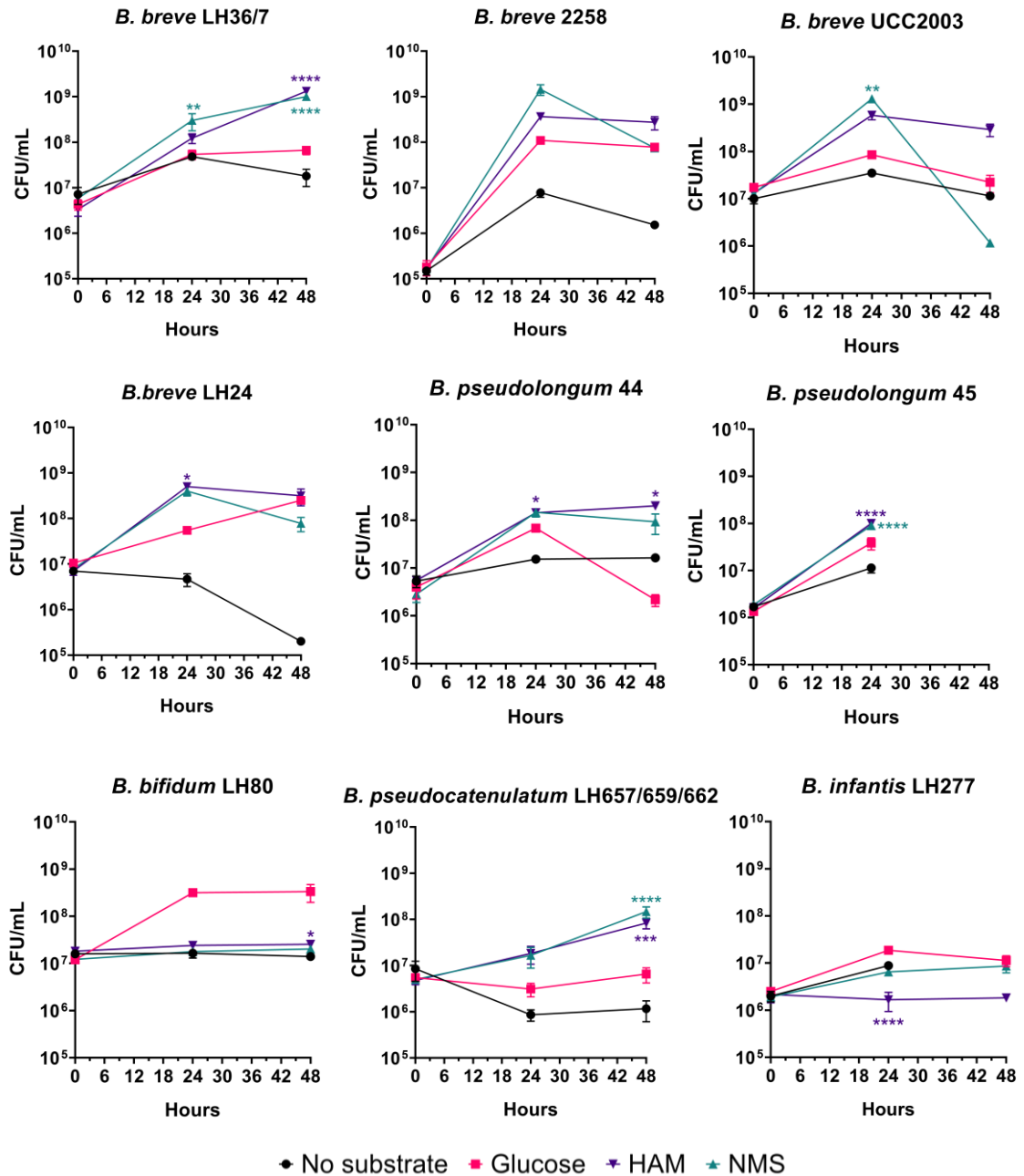
Figure 3-6 Utilisation of various strains of glucose, normal maize starch (NMS) and high amylose maize (HAM) by measurement of CFU. All substrate concentrations were 1% w/v. Values at each timepoint are displayed as mean of triplicate CFU measurements. Asterisks indicate a significant difference compared to the No substrate control. **** $p<0.0001$ ***$p<0.001$ **$p<0.01$ *$p<0.05$

NMS and HAM, but only at the 48 h timepoint, suggesting these strains could be

utilising the substrate at a slower rate than other strains. In fact, the bacterial cell

counts of B. pseudocatenulatum LH strains and B. breve LH36/7 did not plateau

after 24h; bacterial counts utilising either NMS and HAM increase between 24 and 48 h, suggesting the substrate was not depleted (Figure 3-6). The *B. longum* subsp. *infantis* LH277 strain was found to have significantly lower growth on HAM relative to the negative control and no significant differences for NMS suggesting it did not utilise starch. Similarly, *B. bifidum* LH80 was found not to have significantly higher growth on either starch.

Generally, a trend occurred where the bacterial counts in the NMS media declined between 24 and 48 h suggesting the substrate was used up quickly (Figure 3-6). Either the cells suffered an accumulation of waste products or simply senesced due to rapid replication, or both. However, in most cases, the cell counts in the presence of HAM remained stationary at the measured 48h timepoint.

Higher amylose content could slow enzymatic hydrolysis of the starch, allowing a slower release of sugars from the starch molecule, sustaining growth of bacteria for longer periods than a more accessible starch. For example, *B. breve* 2003 appears to be a rapid degrader of NMS and after 48 h the cell counts were lower than the negative control, while the cell counts remained stationary in the presence of HAM (Figure 3-6). However, from these two timepoints it was unclear where the true growth peak was due to having limited sampling timepoints. This is an important finding as the substrates were observed to be metabolised and used up at different timepoints depending on the strain and starch used meaning that in studies which measure growth at limited or single timepoints could be missing important subtleties in bacterial phenotype. Thus, these data demonstrate the growth kinetics of starch utilisation is strain and starch structure specific.

Further to this, several experiments were carried out to establish optimal timepoints to select to capture important phases of bacterial growth on starch of *B. breve* UCC2003 such as the exponential, peak growth, and stationary phase: 0, 4, 8, 14, 15, 16, 18, 20, 22, 24, 28, 30, 36, 48, 52, 72, and 76 h. In these assays, autoclaved SPS and granular sterilised SPS were tested for utilisation by *B. breve* 2003. Conclusively, there was significantly higher growth on autoclaved SPS compared to the negative control at earlier timepoints (prior to 24 h), and no significant differences in growth on granular SPS at any timepoint (Figure 3-7).
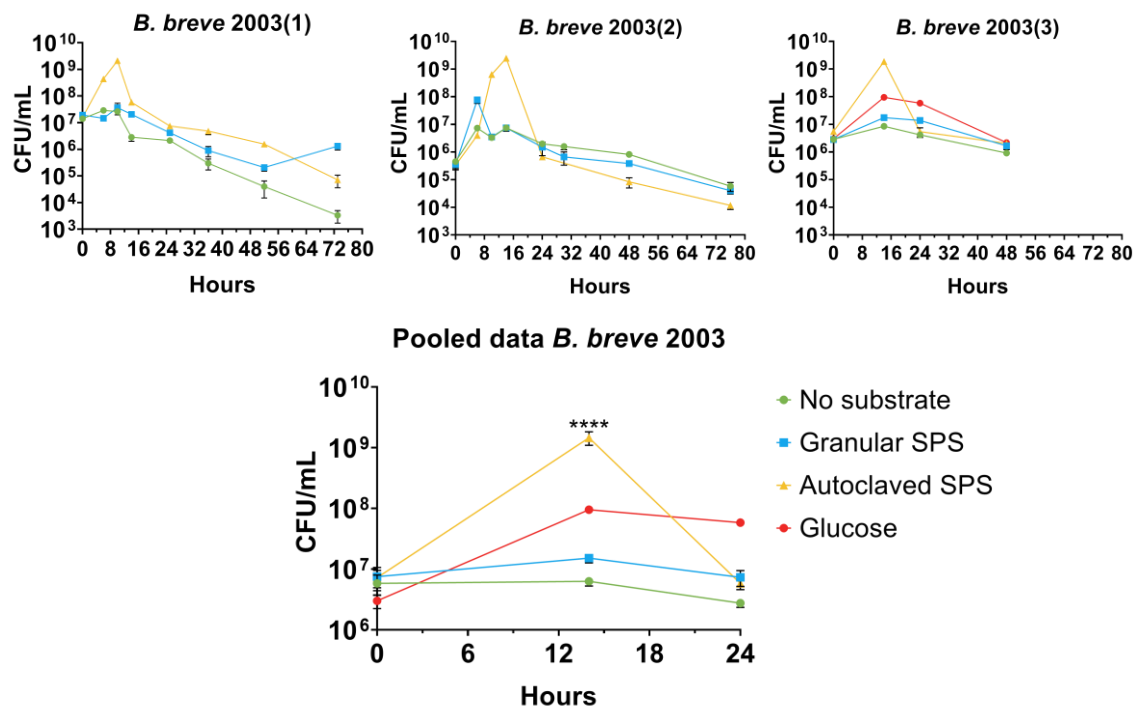
Figure 3-7 B. breve *UCC2003 growth curves utilising granular or autoclaved SPS. All substrate concentrations were 1% w/v. Values at each timepoint are displayed as mean of triplicate CFU measurements. [Top] The experiment was repeated 3 times in triplicate (*B. breve 2003(1)*, B. breve *2003(2)*, B. breve *2003(3)*). [Bottom] All values from each experiment at timepoints 0, 14, and 24 hours were averaged. The asterisks displayed represent a t-test performed comparing Autoclaved SPS CFU value compared to No Substrate at 14h. **** p<0.0001.*

These data confirm that growth kinetics on accessible starch were occurring at around the 12 h timepoint for this strain. Starch resistance or accessibility to

hydrolysis should be considered when selecting timepoints to assess utilisation by strains. Granular potato starch was not further investigated as a useful substrate to compare strains as the structure is extremely resistant to breakdown due to having a larger granule size (Raigond, Ezekiel et al. 2015).

3.4.4 Amylose content significantly impacts growth and metabolite output

As a result of HAM allowing sustained growth over time (implying a higher resistance to hydrolysis) and granular forms of starch showing no significant growth in *B. breve* 2003, a higher amylose variety of maize (HylonVII, 70% amylose (Koev, Muñoz-García et al. 2020)) was used to test the reference and LH strains for differences in utilisation of more resistant starch as amylose increases starch crystallinity after hydrothermal treatment (Soler, Velazquez et al. 2020). Timepoints 0, 12, 18, 24, and 48 h were selected to capture crucial kinetics for both types of starch.

Assessing the relationship between bacterial starch hydrolysis efficiency and starch accessibility can be evaluated using the growth curve kinetics on two metrics: the extent of bacterial growth achieved (peak cell count) and how quickly bacterial counts declined. If bacterial growth transcended a lag, exponential, and stationary phase followed by a death phase it can be assumed that the bacteria utilised the substrate to depletion. Bacteria still in an exponential phase of growth by 48 h indicates their growth is sustained because of slower degradation. The utilisation of NMS and HylonVII was confirmed again in *B. breve* 2003, *B. breve* LH24, *B. pseudocatenulatum* LH662, and both *B. pseudolongum* strains 44 and 45 (Figure 3-8). *B. breve* 2003, *B. breve* LH24, and *B. pseudolongum* 45 had significantly higher cell counts using NMS count relative to the control at 24 h but

not at 48 h suggesting they are efficient at degradation (Figure 3-8). Each strain

increased its cell counts by three orders of magnitude which also suggests they

are 'moderate degraders' of NMS and HylonVII.
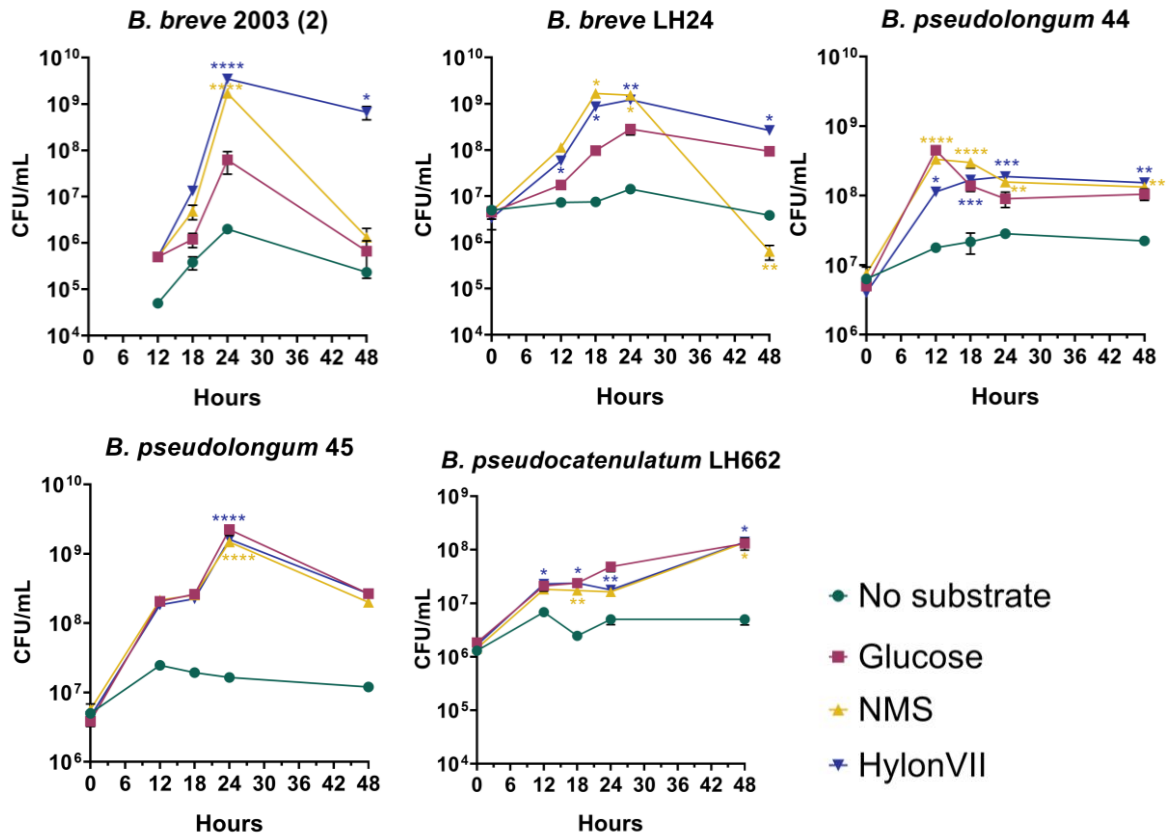
Figure 3-8 Resistant starch (HylonVII) utilisation curves demonstrated a starch structure specific effect on bacterial growth. Strains were assessed using normal maize starch (NMS), high amylose maize (HylonVII) using CFU at 5 timepoints between 0h 48 hours. Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.

For each strain except *B. pseudolongum* 45, cell counts at 48 h were significantly

higher than the control because it had not entered a significant death phase and

were still utilising the substrate. *B. pseudolongum* 45 could be suggested as being

a 'strong degrader' of this starch being equally efficient at degrading NMS and

HylonVII through all the growth phases prior to 48 h. *B. breve* 2003 and LH24 had

a decline in cell counts at 48h which suggests they were beginning to deplete the substrate (Figure 3-8).

B. pseudolongum 44 degrading the accessible maize starch (NMS) achieved significantly higher growth at 12 h and remained significantly higher for the duration of the assay (48 h), which was similarly observed in *B. pseudocatenulatum* LH662. They achieved a cell count increase of two orders of magnitude, implying that the growth of these strains is being sustained in the NMS, so they could be considered 'slow degraders' of this starch. Alternatively, they had slow replication under lab conditions compared to other strains.

The experiments testing growth kinetics of bifidobacterial strains on NMS and HylonVII were accompanied by simultaneous quantitative metabolomic analysis which provides a fuller picture of the utilisation and subsequent metabolism of the substrate intracellularly. Metabolites detected in the supernatant of the cultures were primarily hydrolysis by-products (maltose, glucose) and bacterial metabolites (acetate, ethanol) produced over the course of the fermentation. Around 20 molecules were identified at least once in the bacterial starch fermentation media. The complement of metabolites produced are limited in isolate cultures contrary to complex bacterial communities which can together ferment substrates into a wider variety of SCFAs and other metabolic by-products (Louis and Flint 2017, LaBouyer, Holtrop et al. 2022). Acetate, propionate, and butyrate are the primary end-product SCFA produced by gut microbes and butyrate increases are associated with starch fermentation; however, acetate is the main SCFA produced by *Bifidobacterium* (Flint, Scott et al. 2012).

As previously established, the utilisation of NMS and HylonVII by

B. pseudolongum 44 was significantly higher compared to the negative control at

every timepoint measured after 0 h; its growth was sustained for the duration of

the assay (Figure 3-9). However, there were significant differences in cell counts

between substrates that show this strain achieved significantly lower growth on the

more resistant HylonVII compared to NMS at 12 and 18 h timepoints (p<0.0001,

p=0.0015 respectively). Statistically significantly higher levels of ethanol were

produced relative to the negative control in the presence of both starches (Figure

3-9).
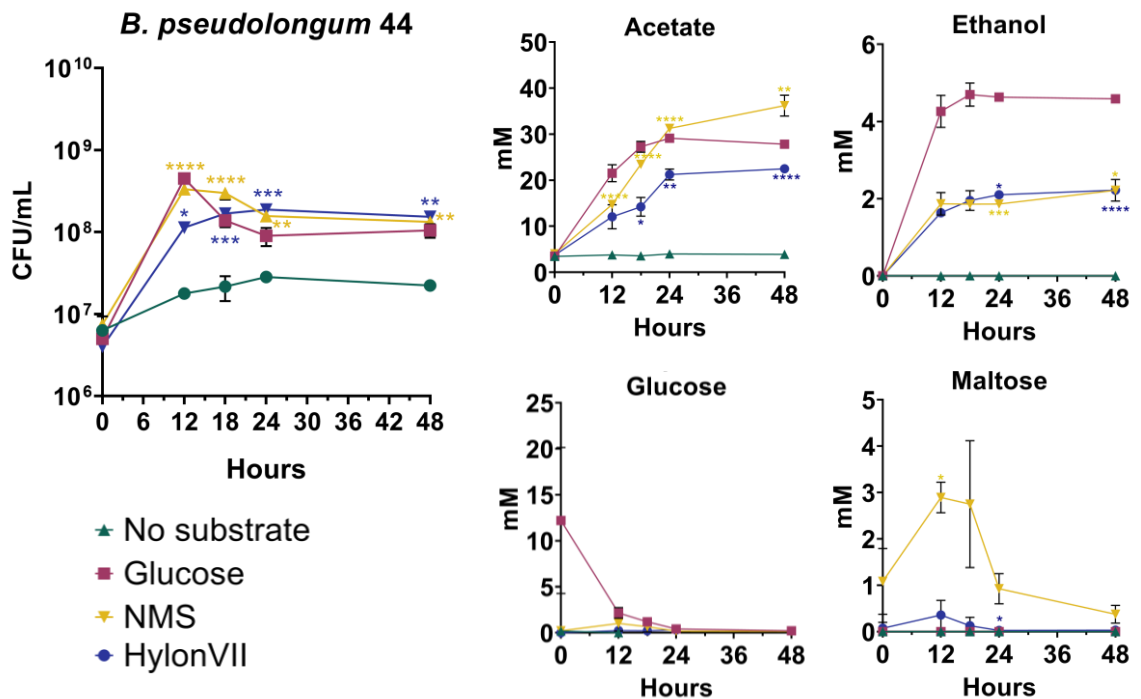
Figure 3-9 B. pseudolongum *44 growth curve compared to metabolite production (acetate, ethanol, glucose, maltose) in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

The pattern of acetate production over time of *B. pseudolongum* 44 for NMS and

HylonVII resembles the growth pattern implying that as the bacteria metabolise the

starch for growth, they produce acetate in the process (Figure 3-9). At each time point after 0 h, there were significantly higher acetate concentrations for both starch substrates compared to the negative control (Figure 3-9). There were also significantly lower concentrations of acetate using HylonVII at 18, 24, and 48 h timepoints compared to NMS (p=0.0324, p=0.0014, p=0.0128 respectively) suggesting the bacterial growth was stunted by resistance of HylonVII to hydrolysis leading to slower degradation, growth, and lower acetate production. The production of maltose from NMS was significantly higher at 12h compared to the negative control, but not for HylonVII which confirms that NMS was much more readily hydrolysed into sugars than HylonVII by *B. pseudolongum* 44. The concentration of maltose produced from NMS at 12 h was significantly higher than from HylonVII (p=0.0023) when fermented by *B. pseudolongum* 44.

There were no significant differences observed in cell counts of *B. pseudolongum* 45 between NMS and HylonVII at any time point (p ≥ 0.7562) suggesting it is equally capable of using accessible and resistant forms of maize starch (Figure 3-10). Interestingly, *B. pseudolongum* 45 exhibited a lag phase in growth at earlier timepoints 12 and 18h before then achieving higher growth rates on both starch types (Figure 3-10). This could be due to a delay in production of enzymes which bind to and hydrolyse starch. *B. pseudolongum* 45 had statistically higher levels of acetate and ethanol produced in the presence of NMS and HylonVII relative to the negative control, and a significant difference in acetate concentrations between NMS and HylonVII was only observed at 12 h (p = 0.0055) and not significantly different at any other timepoints. There were significantly higher levels of maltose produced in the presence of NMS at all timepoints after 0h, and in the presence of

HylonVII at 0 and 12 h (Figure 3-10). There were significantly lower concentrations of maltose between HylonVII and NMS at 12 h (p=0.0019) and 18 h (p= 0.0159), implying that while growth rates were the same, HylonVII is still more resistant to enzymatic hydrolysis than NMS when it is degraded by *B. pseudolongum* 45.
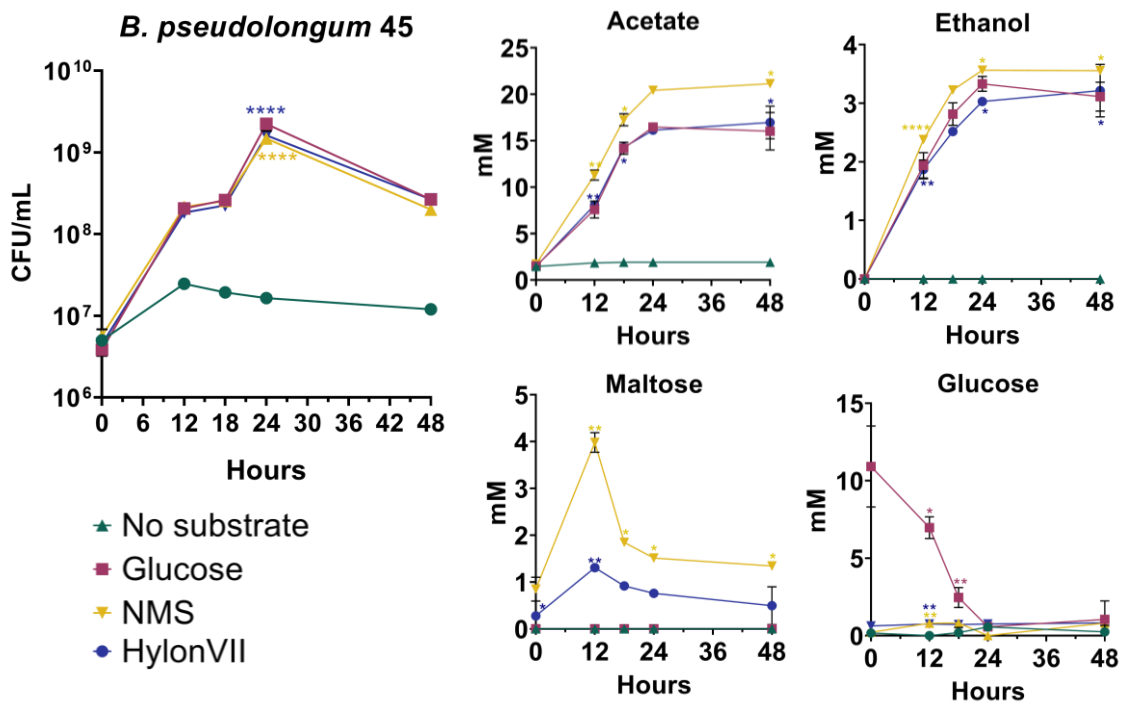
Figure 3-10 B. pseudolongum *45 growth curve compared to metabolite production (acetate, ethanol, glucose, maltose) in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

A comparison of metabolites produced between different strains is physiologically relevant to the gut ecosystem and to host health, since SCFAs and sugars exert effects on other microbial taxa. By comparing *B. pseudolongum* 44 and *B. pseudolongum* 45 with statistical modelling, the differences in metabolite outputs (acetate, ethanol, glucose) of the two strains were assessed in the presence of both NMS and HylonVII (Figure 3-11). Maltose production was not significantly different between strains when utilising NMS, however

B. pseudolongum 45 produced significantly higher concentrations of maltose from

HylonVII (Figure 3-11). The overall acetate production was found to be higher

when *B. pseudolongum* 44 utilised both NMS and HylonVII, in spite of having

significantly lower growth and lower ability to utilise HylonVII (Figure 3-11).
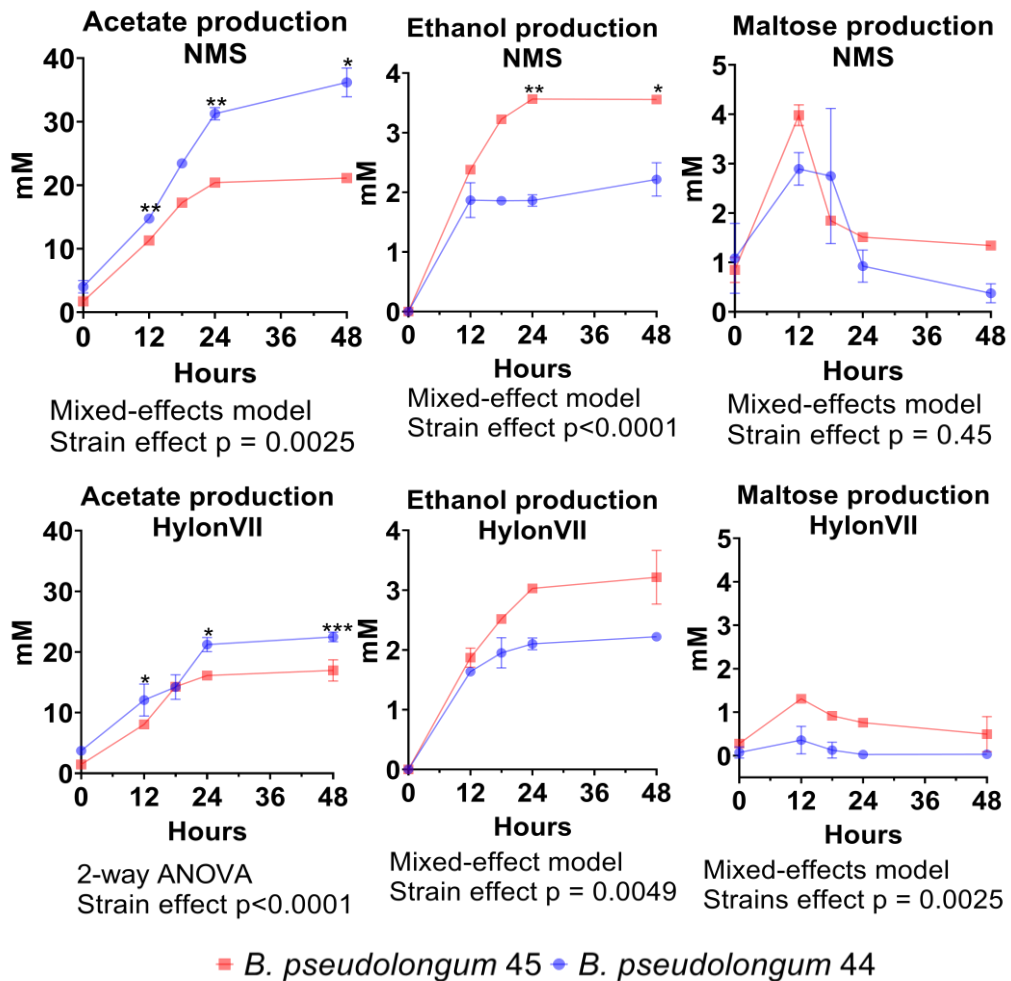
Figure 3-11 Comparison of metabolites (acetate, ethanol, maltose) produced by B. pseudolongum *44 and* B. pseudolongum *45 in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII). Statistical models were used to ascertain if strain had a significant effect on the production of relevant metabolites. Values displayed for each timepoint are the mean of three biological replicates. Asterisks represent a significant difference (using repeated measures ANOVA) between strains at each individual timepoint. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

This finding suggests that degradative abilities of individual strains may contradict metabolite output. This underlines the importance of detailed investigation of phenotype not limited to growth in the presence of a substrate, as these differences could be biologically relevant. These results may be explained by the differences in ethanol produced by the two strains: *B. pseudolongum* 45 appeared to produce significantly higher levels of ethanol in the presence of NMS or HylonVII (Figure 3-11). Therefore, *B. pseudolongum* 44 more efficiently converts starch into acetate, a beneficial metabolite which can participate in cross-feeding interactions with other microbes and contribute to host health. To conclude these analyses, growth kinetics may reflect differences in ability of bacterial strains to use starch however the physiologically relevant outcomes such as acetate production can contradict these.

Metabolomic analysis of *B. breve* LH24 and *B. breve* 2003 culture media showed that acetate and maltose was produced in higher concentrations in the presence of NMS than in HylonVII (Figure 3-12). Metabolite concentrations indicated the rate of starch hydrolysate release which was found to correlate to starch resistance. Meanwhile, clonal strains of *B. pseudocatenulatum* LH659 and LH662 were also assayed multiple times in the presence of HylonVII, showing that growth did not diminish over the course of 48 h again demonstrating that these strains utilise starch more slowly (Figure 3-13). Acetate and ethanol produced by LH662 did not plateau before 48 h like in several other strains further restating that the LH662 strain starch utilisation was slower than other stronger degraders (Figure 3-13).
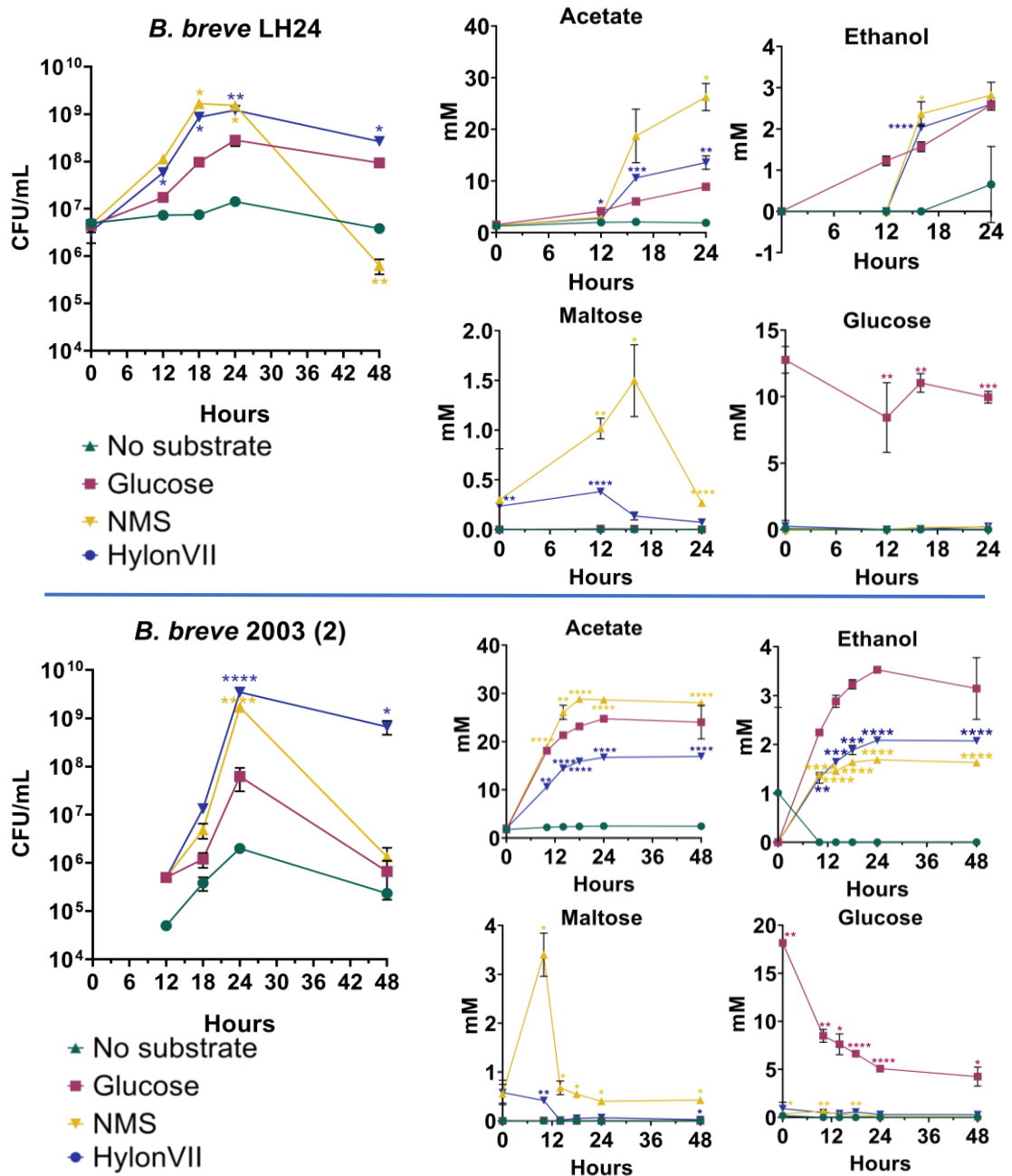
Figure 3-12 B. breve *growth curves compared to metabolite production (acetate, ethanol, glucose, maltose) in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

A similar trend was observed for *B. breve* 2258 where the production of acetate

trended upwards instead of declining (although metabolomics was only carried out

for 24 h for this strain) (Figure 3-14). There was growth observed in *B. breve* 2258

but there were no statistically significant results for either starch (Figure 3-14).

Although, there were statistically significantly higher levels of acetate, ethanol, and

maltose produced for both starches (Figure 3-14). *B. breve* 2258 exhibited a

starch utilisation phenotype but did so at a slower rate, meaning that they could

also be considered 'slow degraders' of gelatinised NMS and HylonVII.
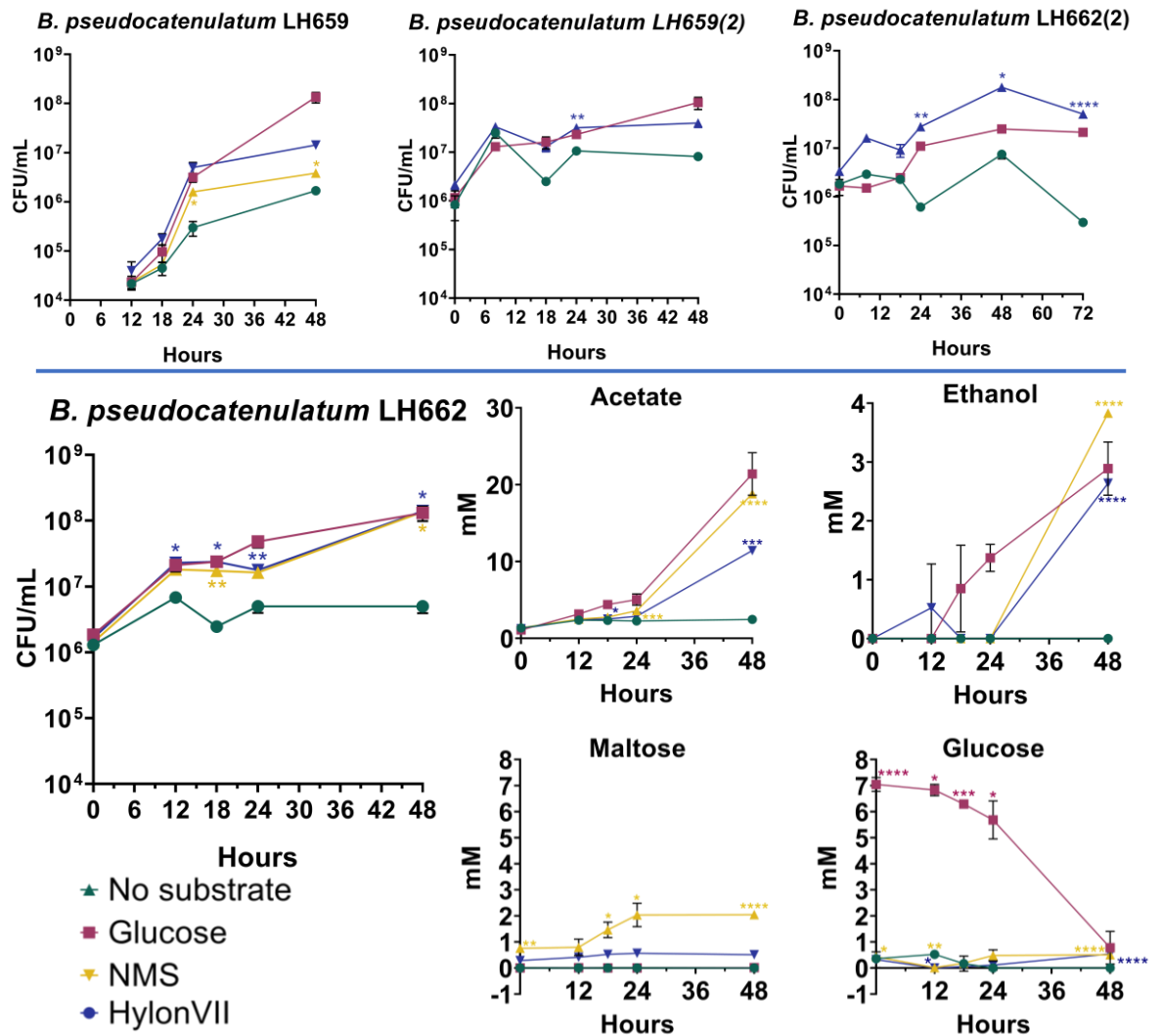
Figure 3-13 B. pseudocatenulatum *growth curves compared to metabolite production (acetate, ethanol, glucose, maltose) in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

B. adolescentis 702204 (NCIMB 702204 / ATCC15703 / DSM20083) was

assessed for utilisation of gelatinised NMS and HylonVII given that previous

studies yielded ambiguous results about its starch degrading capabilities. In 2018,

the bacterial growth of this strain *B. adolescentis* 702204 was assessed for

utilisation of a retrograded high amylose maize starch by decrease in pH as a

proxy for utilisation (Kim, Shin et al. 2018). The authors found a significant

decrease in pH and declared confirmation of its ability to utilise the starch and

stated that time course monitoring of CFU confirmed utilisation, but the data was

not shown. In 2019, it was found not to degrade granular, high amylose corn

starch (Jung, Kim et al. 2019). To date, it does not appear to have been tested for

gelatinised starch utilisation.
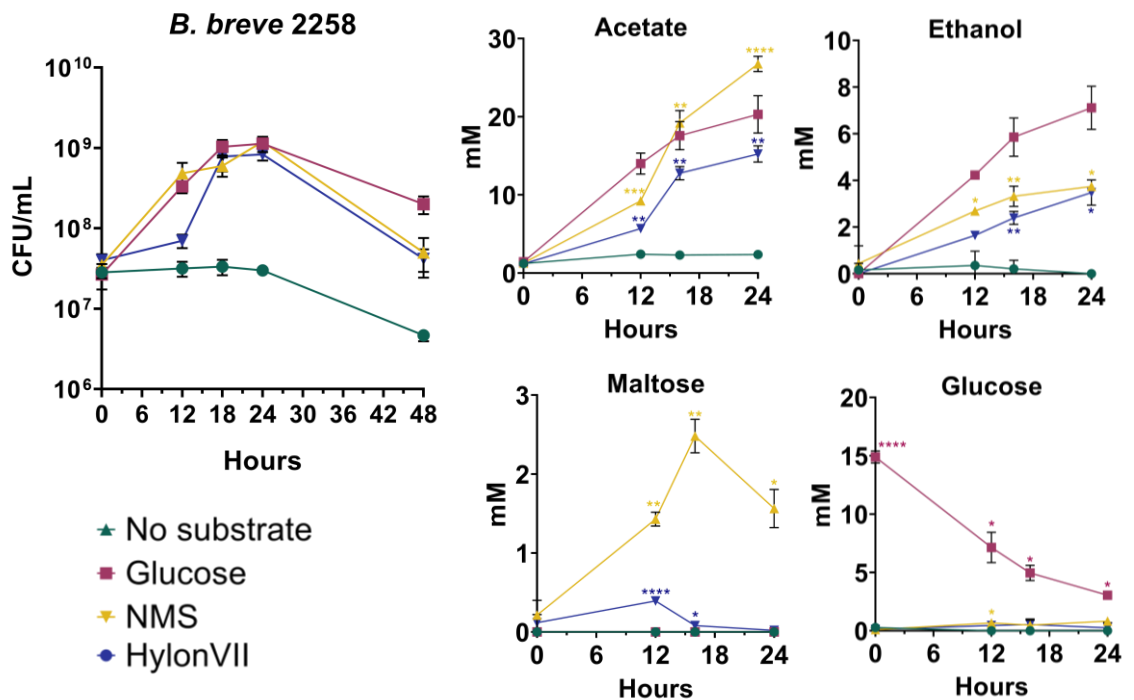
*Figure 3-14 B. breve 2258 growth curve compared to metabolite production (acetate, ethanol, glucose, maltose) in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

157

The current growth assays assessing gelatinised NMS and HylonVII utilisation by *B. adolescentis* 702204 show a minimal but statistically significant increase in utilisation of both substrates relative to the negative control (Figure 3-15). Additionally, there were statistically significantly higher levels of maltose produced in the presence of HylonVII, suggesting that there was a small degree of starch hydrolysis (Figure 3-15). Maltose was also produced from NMS however intra-sample variation diminished any statistical significance. The acetate production decreased minorly which may be indicative of acetate utilisation by the bacteria.

Kim et al. (2018) assumed a decrease in pH (presumably due to acetate production) was indicative of an ability to utilise RS as a carbon source ((Kim, Shin et al. 2018), and in the absence of a negative/positive control or the CFU data, it is not easy to interpret the significance of their results. The current work demonstrates that *B. adolescentis* 702204 produces significantly lower acetate in the presence of glucose compared to no substrate or starch (Supplementary Table S2). The fact that acetate production is significantly higher in the no substrate control compared to glucose (p=0.0042) is interesting since the acetate production has an inverse relationship with bacterial growth. It is possible that this strain favoured lactate production instead in these experimental conditions.

However, it is not clear why *B. adolescentis* 702204 produces acetate in the absence of bacterial growth. Earlier studies show it possesses genes for two β-galactosidases which collectively allow highly efficient use of maltooligo- and mono-saccharides isomaltose, isomaltotriose, and maltose (Hinz, van den Broek et al. 2004). It is possible that *B. adolescentis* 702204 has no ability to use

extremely resistant granular or retrograded starches, while the current work shows
it has limited ability to utilise gelatinised NMS or HylonVII, likely due to having
some ability to partially hydrolyse starch and metabolise the starch hydrolysates.
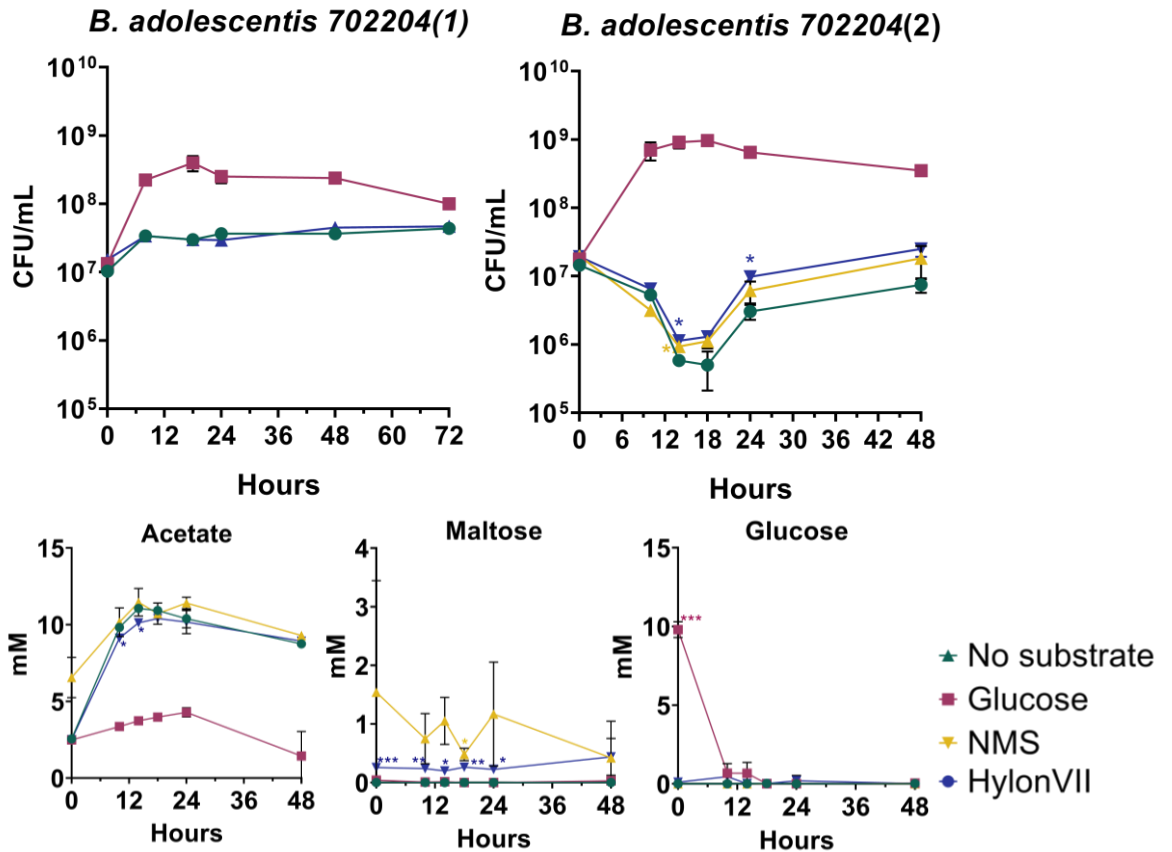
Figure 3-15 B. adolescentis *growth curve compared to metabolite production (acetate, ethanol, glucose, maltose) in the presence of different types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

Finally, the various concentrations of maltose produced by strains over time
presented some notable patterns (Figure 3-16). Moderate degraders *B. breve*
2258 and *B. pseudocatenulatum* LH662 had proficiency in producing maltose
when utilising NMS however the concentrations did not rapidly decline as was
observed in prominent starch users *B. breve* 2003 for example (Figure 3-16). This

could be interpreted as proficiency in hydrolysis but not sugar import, differences

in lab culture tolerance, or generally utilising starch more slowly. Cooperative

processes can be performed by primary degraders of substrates in order to share
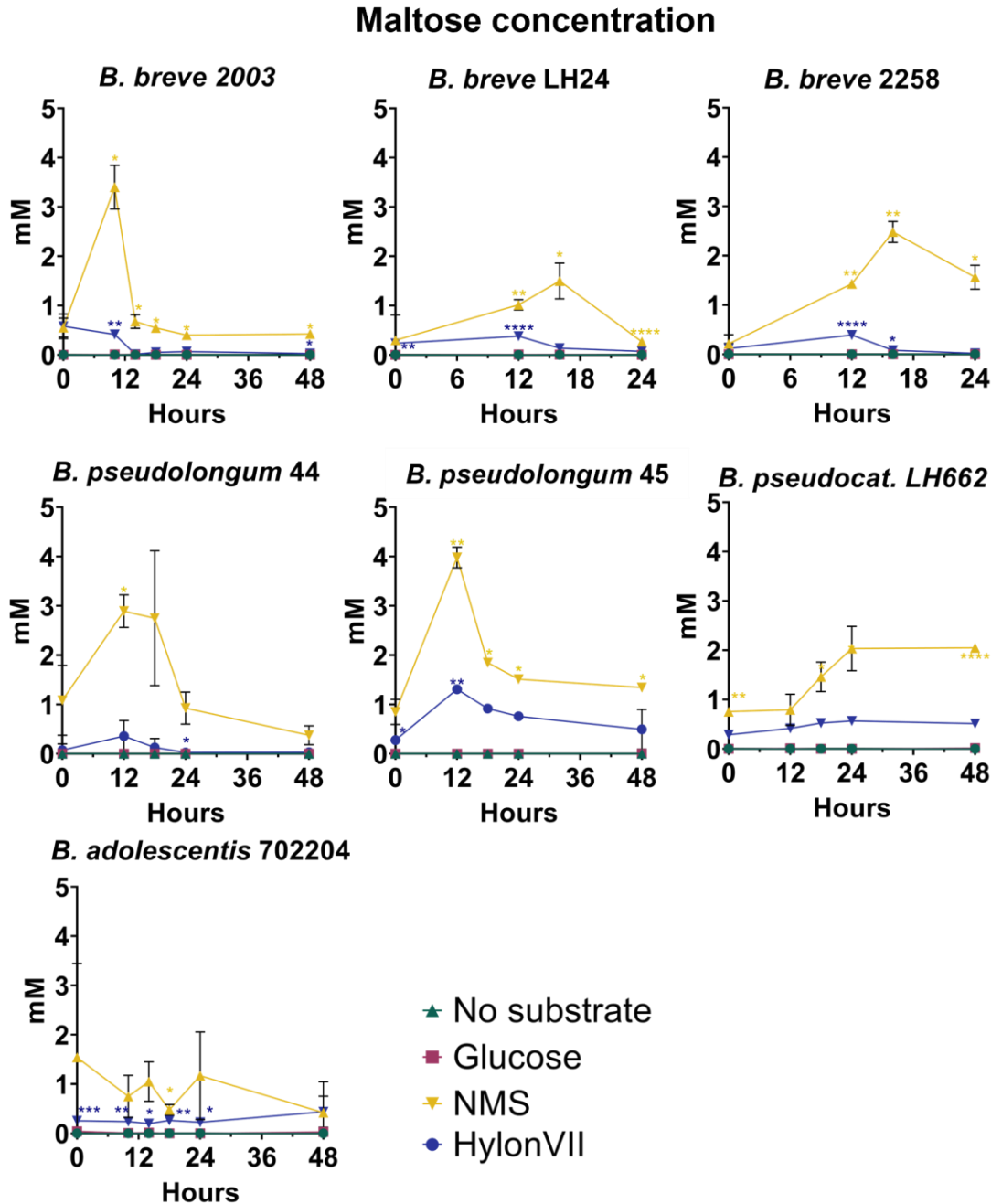
Figure 3-16 Maltose production of strains over time in the presence of two types of maize starch, normal maize starch (NMS), high amylose maize (HylonVII) . Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** $p<0.0001$ ***$p<0.001$ **$p<0.01$ *$p<0.05$.

degradation products, as noted during *Bifidobacterium* HMO degradation (Lawson, O'Neill et al. 2020). Cooperative degradation of substrates is characteristic of *Bifidobacterium*, and if confirmed as true during starch degradation would confer health benefits to other microbial taxa and to the host especially if downstream products include butyrate (Milani, Lugli et al. 2015, Rios-Covian, Gueimonde et al. 2015, Kim, Jeong et al. 2020).

3.4.5 Bioinformatic analysis of CAZymes present in WGS of strains tested

Bioinformatic analysis revealed genetic traits contributing to the differences in starch utilisation between the strains tested and additional strains from the LH strain bank. The total number of GH family counts detected across 21 strains varied between 36 (*B. bifidum* LH34) and 86 (*B. pseudocatenulatum* LH656) (Table 3-5). In previous similar analysis of 12 Actinobacteria genomes, the mean GH count was 48 (Kaoutari, Armougom et al. 2013).

The most prevalent CAZyme family detected was GH13 (Figure 3-17) which was the only GH family found in these strains with known starch hydrolysis enzyme activity (see Table 1-2 for enzyme activity of each family). The GH13 family is the largest of the GH class, with several enzyme activities designated to 46 GH13 subfamilies (Table 1-2). There were overall 14 GH13 subfamilies detected in this set of strains (Figure 3-18). *B. pseudolongum* 45 had the highest overall counts of GH13, in particular the GH13_28 subfamily whose enzyme activity is designated as α-amylase. This finding suggests a plausible mechanism for their superior ability to degrade normal and high amylose varieties of maize starch.

	B. breve NCIMB 2258	48
	B. breve LH21	45
	B. breve UCC2003	62
	B. breve LH24	48
	B. pseudolongum subsp. *pseudolongum* NCIMB 702244	66
	B. pseudolongum subsp. *globosum* NCIMB 702245	46
	B. longum subsp. *infantis* NCIMB 20088	42
	B. breve LH36	61
	B. breve LH37	58
This thesis	*B. pseudocatenulatum* LH662	88
	B. pseudocatenulatum LH664	64
	B. pseudocatenulatum LH656	89
	B. pseudocatenulatum LH657	83
	B. pseudocatenulatum LH659	88
	B. pseudocatenulatum LH11	78
	B. longum subsp. *longum* NCIMB 8809	65
	B. longum subsp. infantis LH277	67
	B. adolescentis NCIMB 702204 / ATC15703	66
	B. longum LH12	72
	B. bifidum LH80	37
	B. bifidum LH34	36
	Median	**64**
(Kaoutari, Armougom et al. 2013)	*Bifidobacterium adolescentis* ATCC 15703	55
	Bifidobacterium animalis subsp. *lactis* AD011	36
	Bifidobacterium bifidum NCIMB 41171	42
	Bifidobacterium breve UCC2003	53
	Bifidobacterium longum DJO10A	37
	Bifidobacterium pseudocatenulatum D2CA	60

Table 3-5 [Top] GH counts in 21 bifidobacterial strains examined in this thesis and [Bottom] GH counts in 6 bifidobacterial strains in a previous similar analysis show similarities in the inter-species variation that exists.

Other prominent families were GH43 among *B. pseudocatenulatum* LH strains

which is involved in arabinoxylan degradation and glycosyltransferase family 2

(GT2) which are commonly associated with synthesis of glycosidic bonds.

B. pseudocatenulatum LH strains had the highest number of total CAZymes from a

variety of these groups (GH13, GH43, GT2 and GT4) and is likely able to degrade

a range of substrates. The reasons for the correlation between GT2 with GH13

prevalence is perplexing due to the fact their enzyme actions are generally

oppositional i.e. synthetic versus hydrolytic. Cyclodextrin glycosyltransferases form

glycosidic bonds to transform the starch into a cyclic form, which may be easier to

hydrolyse (Bautista, Esclapez et al. 2012).
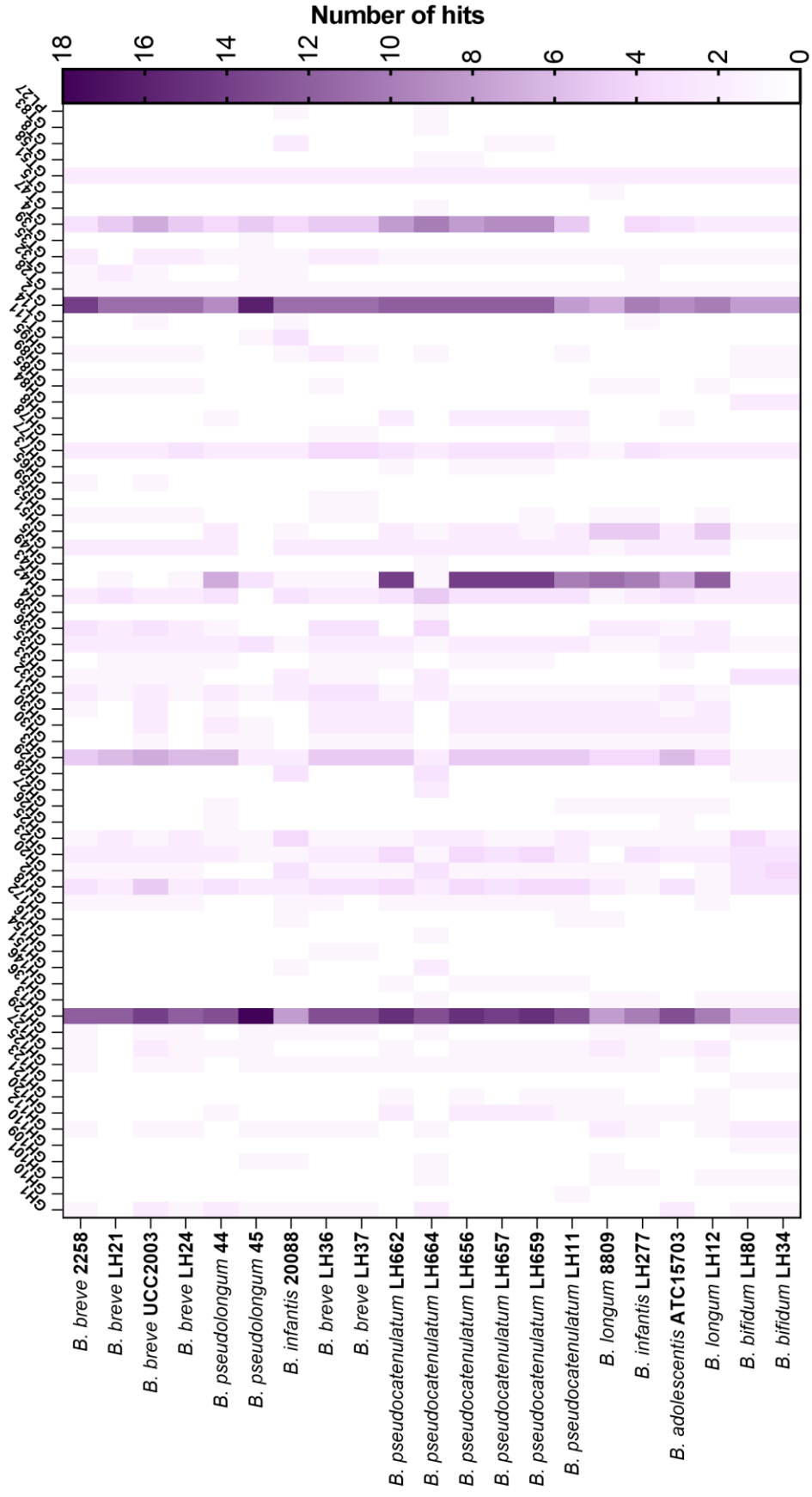
Figure 3-17 The total number of CAZyme families identified in strains' whole genome sequences from the results of dbCAN where the hit was detected by 2 or more tools (Ausland, Zheng et al. 2021).
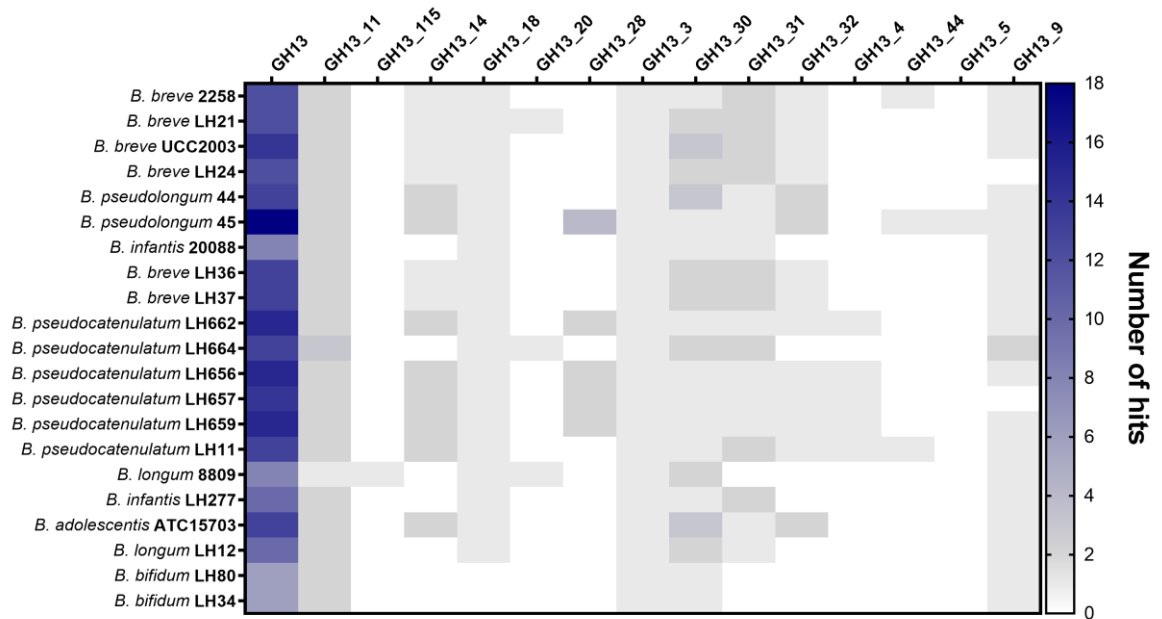
Figure 3-18 The total number of GH13 family and subfamilies identified in strains' whole genome sequences from the results of dbCAN where the hit was detected by 2 or more tools (Ausland, Zheng et al. 2021).

In addition, while glycosyl hydrolases are generally degradative, they can also catalyse transglycosylation, which causes formation of glycan chains or elongate existing oligosaccharides (Van den Broek, Hinz et al. 2005). β-Galactosidases extracted from *B. longum* subsp. *infantis* 20088 and *B. pseudolongum* 44 and used to elongate oligosaccharides produced from lactose, caused the highest growth of *Bifidobacterium* strains using β-galactooligosaccharides produced by its 'own' β-galactosidases (Rabiu, Jay et al. 2001).

The enzyme catalogue that some *Bifidobacterium* have can define its phenotype as only preferring longer glycans rather simple sugars. For example, beta-galactosidase enzymes from *B. adolescentis* 702204 prefers beta(1,4)-galactosides over simple sugar lactose (Hinz, 2004).This suggests that synthesis enzymes could play a role in promoting growth of bacteria as they have proven

mechanisms of glycan synthesis, providing an explanation for the high prevalence of GT enzymes present.

The bioinformatic investigation identified 27 different groups of CBM across all strains (Figure 3-19). CBM48 was the most commonly found from this analysis. In both *B. bifidum* strains (which commonly occupy the infant niche involving milk intake) contained relatively low numbers of all other CBM but high numbers of CBM32 which are lactose and galactose binding (Figure 3-19) (Arboleya, Watkins et al. 2016) (Fushinobu 2010). Among starch-degrading strains *B. pseudolongum* 45, *B. pseudocatenulatum* LH662/565/657/659, higher counts of CBM25, CBM26, CBM42 were observed.

Figure 3-19 The total number of carbohydrate binding modules (CBM) identified in strains' whole genome sequences from the results of dbCAN where the hit was detected by 2 or more tools (Ausland, Zheng et al. 2021).

B. pseudocatenulatum strains had relatively higher counts of L-rhamnose binding module CBM67 compared to all other strains, which are potentially involved in pathogen suppression as this rare sugar is commonly produced in pathogenic

bacterial cell walls (Fujimoto, Jackson et al. 2013, Mistou, Sutcliffe et al. 2016). The LH isolates *B. breve* LH36/37 and *B. pseudocatenulatum* LH662, LH656, LH657, and LH659 uniquely possess starch binding domain CBM68. Similar recent analyses show this is a relatively rare starch-binding CBM for *Bifidobacterium* (Dobranowski and Stintzi 2021).

Overall, 14 GH13 subfamilies and 7 starch binding domains were identified (Figure 3-20). The strains which did not utilise starch had the fewest GH13 family counts and binding modules: *B. longum* subsp. *infantis* 20088, *B. longum* subsp. *infantis* LH277, *B. longum* 8809, and *B. bifidum* LH80. The detailed growth curves generated coupled with bioinformatic analysis confirms that these strains are not suited to starch degradation as these species are more commonly found in infant microbiota where starch is not present.

The presence of CBM74 (a strong RS binding domain) was uniquely present in small numbers (1-2 counts per strain) in *B. pseudolongum* 45 and *B. pseudocatenulatum* LH662/665/657/659 (Valk, Lammerts van Bueren et al. 2016, Photenhauer, Cerqueira et al. 2022). This CBM commonly associates with GH13_28, CBM25 and CBM26 which was found to be true for these data (Valk, Lammerts van Bueren et al. 2016) (Figure 3-20). In fact, these strains were the only ones tested which contained GH13_28 family enzymes suggesting it may confer an advantage for starch utilisation.

Interestingly, the clonal strains *B. breve* LH36/LH37 and *B. pseudocatenulatum* LH662-LH659 had slight differences in total CAZyme counts (Table 3-5). *B. pseudocatenulatum* LH strains had subtle differences in the counts of GH13

family enzymes and CBM25 (Figure 3-20). This suggests that while these strains

are extremely similar (Table 3-4), SNP differences in the genomes may be

detected as different enzyme sequences by the Cazy database.
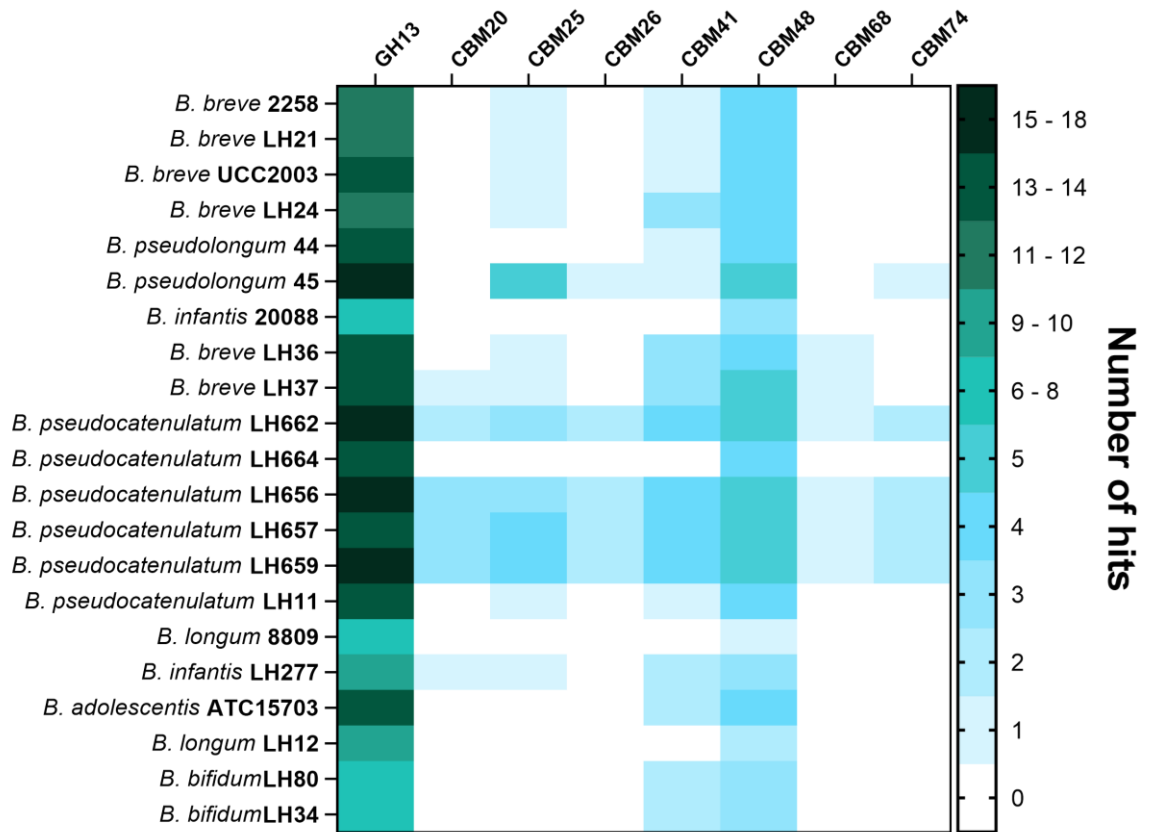
Figure 3-20 The total number of CAZymes families and CBMs detected in strains related to starch hydrolysis or binding identified by dbCAN (where the hit was detected by 2 or more tools) (Ausland, Zheng et al. 2021).

3.5 Discussion

3.5.1 Limitations in methodology to measure microbial starch utilisation phenotype

The advantage of 96-well microplate growth assays using an OD

spectrophotometer include high throughput assaying of multiple strains and/or

conditions at frequent intervals, convenient for generating detailed growth curves.

For measuring growth on bacterial substrates like starch, there were some

limitations in this methodology. Bacteria in culture scatter light delivered by the spectrophotometer proportional to their population density. Suspended starch increased the overall background reading which may have affected the reliability of the OD reading. The background readings of each starch increased stepwise with each increase in w/v concentration, although in other studies authors would manually dilute cultures in fresh media before taking a reading to dilute the starch (Ryan, Fitzgerald et al. 2006). Although not experimentally explored in this thesis, turbid starch suspensions affect light transmission (at 620 nm) in a time-dependant manner (Wang, Li et al. 2015). This may interfere with cell density measurements since starch can (a) increase in turbidity over the course of the experiment (Shujun, Wenyuan et al. 2006) or (b) potentially reduce in turbidity as bacterial starch degradation progresses. It was found that using a plate reader to measure OD of bacterial starch utilisation yielded inconsistent results between experiments.

While soluble starches used (NMS or SPS) were easily integrated into media to provide a manageable substrate for initial screenings, careful method development had to be carried out to solubilise the starch through boiling before autoclaving to reduce gelling. Moreover, high amylose starches (HAM, HylonVII) more easily gel which prevents accurate readings which was markedly observed in HylonVII (Wang, Li et al. 2015). These techniques are not widely published or discussed as a methodological hurdle for starch utilisation assays involving OD measurements. Furthermore, raw starch granules are insoluble meaning the particles could be unevenly distributed between readings. Yet, OD measurements are used for measurement of bacteria growth in the presence of granular starch (Jung, Seo et al. 2020). The problem of insoluble starch generating turbidity has been previously

identified by other researchers, who have used proxies for bacterial metabolism such as measuring decreases in pH (Kim, Shin et al. 2018), plating of cultures on agar containing starch, or depletion of the substrate over time (Centanni, Lawley et al. 2018). These alternatives were considered but colony enumeration was found to be the best alternative due to it being a direct measure of bacterial replication over time, from which bacterial growth kinetics can be generated.

Aside from being a direct measure of bacterial concentration rather than relying on instrumentation, the CFU method can be subject to human error such as differences in mixing dilutions or diligence in pipetting and plating (Brugger, Baumberger et al. 2012). Careful statistical comparisons were made only between each strain and no substrate control in biologically independent triplicates to ensure that conditions and data produced were as consistent and reproducible as possible. Moreover, where possible, the assays were replicated to ensure reproducibility.

3.5.2 Starch utilisation by *Bifidobacterium* genus members

In the majority of publications discussing the ability of microbial taxa in the human gut to degrade RS, the most commonly mentioned is *B. adolescentis* which is indeed a primary degrader of RS and the most well characterised (Rodriguez and Martiny 2020). To contribute new knowledge to this area, *B. adolescentis* strains were not preferred for investigation in the present work aside from a single strain of *B. adolescentis* 702204 which had ambiguous previous results regarding starch utilisation (Kim, Shin et al. 2018, Jung, Kim et al. 2019). Crucially, *B. adolescentis* is not a highly abundant inhabitant of the infant gut microbiota (Duranti, Milani et al. 2016). Commonly accepted primary degraders of highly inaccessible, granular

starch also include *B. pseudolongum* and *B. choerinum* strains which are animal-associated species (Centanni, Lawley et al. 2018, Jung and Park 2023). Seldom discussed in detail are other species and strains belonging to the *Bifidobacterium* genus with starch utilisation capabilities or their potential impact on health. As highlighted by Centanni et al. (2018), primary degraders of starch do not need to be present in high relative abundance to be important for a healthy microbiota as they may serve important cross-feeding roles, and this could be leveraged for development of synbiotic biotherapeutics (Centanni, Lawley et al. 2018). The current results show starch of varying structures were utilised by 5 strains belonging to *B. breve*, a presumed clade of *B. pseudocatenulatum*, 2 strains of *B. pseudolongum*, and additionally proved negligible utilisation by strains of *B. bifidum, B. longum, B. longum* subsp. *infantis*, and *B. adolescentis*. These findings largely replicate the earliest *Bifidobacterium* starch utilisation screening (Ryan, Fitzgerald et al. 2006). The current work testing of various starch types and assessment of growth kinetics provides new insights into the factors which impact both growth and metabolism of starch by *Bifidobacterium* members not commonly investigated.

3.5.3 Starch structure impacts bacterial isolate growth kinetics

As previously discussed, there are multiple factors which give rise to a wide variation in starch and there is growing understanding for the need to include multiple starch structures in bacterial phenotyping or microbiome analysis (DeMartino and Cockburn 2020, Teichmann and Cockburn 2021). Considering the variety of starch present across plants and between preparations of different foods, there is a broad spectrum of RS structures potentially ingested as part of a

human diet. RS is therefore not confined to highly inaccessible starch which reaches the colon largely intact, but research on bacterial strain utilisation moving forward should accept that a variety of starch structures likely reach the microbiota, many of which are previously cooked or may be partially hydrolysed. Therefore, exploring the variations in starch utilisation kinetics of various *Bifidobacterium* strains and gelatinised starches reflects this reality to understand the spectrum of utilisation across different species.

The phenotypic studies presented in this chapter indicate there is strong strain-level variation among *Bifidobacterium* starch usage depending on starch type. *B. pseudolongum* 45 utilised gelatinised NMS significantly more readily than gelatinised SPS suggesting a preference of starch based on botanical origin. This effect was not observed in *B. pseudolongum* 44 which could be explained by the fact they were isolated from animals (*B. pseudolongum* 44 from pig faeces and *B. pseudolongum* 45 from cow rumen) that have different diets. The two *B. pseudolongum* strains also had different starch growth kinetics using NMS or HylonVII. This shows intra-species variation in starch utilisation is common depending on both botanical origin and amylose content, two variables which impact starch structure.

The growth kinetics of various strains utilising HAM and HylonVII relative to NMS shows that amylose content of starch is a significant factor impacting accessibility of cooked starches to degradative enzymes. Higher amylose varieties of maize starch were utilised more slowly by most strains, leading to a plateau of bacterial growth by the 48 h timepoint. This process could have functional benefits *in vivo*

as a high amylose starch could provide a more sustaining energy source due to a slower release of sugars from the starch molecule, rather than being used up quickly. Growth kinetics of different starch types should be considered when testing strains using limited timepoints.

3.5.4 Metabolomic analysis bolsters growth kinetics data

The combining of growth kinetics with metabolomics datasets are relatively underutilised when investigating phenotype of strains. Quantitative measurement of metabolites provides more details about the rates of starch hydrolysis and the end-products of bacteria metabolism. The SCFAs produced provide a clearer picture of what biological impacts starch fermentation could have *in vivo*. The pH altering effect of fibre fermentation is shown to be a large contributor to variation in SCFA ratios and bacterial populations in the human gut (Walker, Duncan et al. 2005). As observed in the differences between acetate production in the two *B. pseudolongum* strains tested, there are strain-level variations to consider. Whether these differences would translate to significant impacts *in situ* may be considered but on a large scale over time it is possible. For precision modulation of the microbiome, there are two considerations to note from the current findings: degraders of starch which release more maltose than they import, releasing the sugars into the community, and those which internalise and rapidly breakdown starch to produce SCFAs which then benefit the host and can enter a secondary conversion to butyrate (Belenguer, Duncan et al. 2006, Baxter, Schmidt et al. 2019, Kim, Jeong et al. 2020, Teichmann and Cockburn 2021). In fact, increased SCFA production such as acetate facilitates butyrate conversion by lower pH

affecting the stoichiometry of the butyrogenic pathways (LaBouyer, Holtrop et al. 2022).

B. pseudolongum 45 was observed to be equally adept at degrading the two different starch structures, NMS and HylonVII, while *B. pseudolongum* 44 had less capacity to degrade HylonVII. Acetate and maltose levels in both strains were lower when using HylonVII compared to NMS. *B. pseudolongum* 44 had a significantly reduced ability to produce maltose in the presence of HylonVII but produced significantly more acetate than *B. pseudolongum* 45 in both conditions. It is clear that *B. pseudolongum* 45 instead produced significantly more ethanol suggesting it is a less efficient metaboliser of starch. This finding demonstrates that an increase in bacterial replication may not correlate with the beneficial metabolites produced. *B. longum* subsp. *infantis* has been demonstrated to shift to ethanol production during the inefficient degradation of HMOs (Özcan and Sela 2018) and detailed studies in *B. adolescentis* confirm that ethanol is produced at the expense of acetate, yielding fewer ATP molecules (Amaretti, Bernardi et al. 2007). This is the first description of two *Bifidobacterium* strains exhibiting a different distribution of carbon fluxes through the bifid shunt pathway in response to starch. The differences in how the two *B. pseudolongum* strains hydrolyse and metabolise starch can be further investigated by researching the mechanisms they employ. As far as is known, this is the first detailed demonstration of how different starch structures impact the growth of individual bacterial strains and their metabolite outputs.

In *B. pseudocatenulatum* LH662, the growth on NMS was more gradual over 48h, leading to the highest growth recorded at 48 h, indicating it was slower at utilising NMS. The slow degrader phenotype by *B. pseudocatenulatum* has been previously demonstrated (Belenguer, Duncan et al. 2006). The bioinformatics analysis showing they have a large number of GH enzymes (100+) from different families may implicate them as being able to use different carbohydrate sources and occupy multiple niches in the gut. Following from this, the strain and starch type reveal interesting variations which could indicate the role they play in the ecosystem. Metabolomics analyses revealed that they could produce acetate and maltose from NMS and HylonVII but that the concentrations continued to increase or plateau over the time course, respectively. This is concordant with *R. bromii*, another primary degrader of starch which cross feeds other taxa with its starch hydrolysates ((Ze, Duncan et al. 2012, Crost, Le Gall et al. 2018). Additionally, LH662 was isolated alongside many clonal strains which are postulated to cooperate with each other in a single ecosystem (the gut microbiome of the healthy infant) (Lawson, O'Neill et al. 2020). If these *B. pseudocatenulatum* strains are derived from a common ancestor, the clade could have a pangenome effect where the collective utilises starch, which would preclude an ability to effectively utilise starch isolation. Maltose is also likely to be a regulatory molecule for the transcriptional upregulation of starch degradation genes as with *B. thetaiotaomicron* (Foley, Martens et al. 2018). If the strains release extra maltose than they can use, they may be doing so to trigger a starch utilisation phenotype in community members. The *B. pseudocatenulatum* LH662 or *B. breve* 2258 strains could be investigated for cross-feeding propensity or reduced ability

to uptake maltose to strengthen this hypothesis. A goal of precision modulation of the gut microbiome could be to cross-feed other microbial taxa through initial degradation of resistant starch. It is characteristic of bifidobacteria to possess an altruistic strategy as part of their previously demonstrated cooperative behaviours (Lawson, O'Neill et al. 2020) and so it is feasible these interactions could be leveraged to improve the gut microbiota.

3.5.5 Unique isolates of *B. pseudocatenulatum* may straddle infant dietary niches

For isolation of the *B. pseudocatenulatum* LH strains, faeces was collected from healthy, full-term, exclusively breast-fed infants who had not received any antibiotics/probiotics prior to sampling (Lawson, O'Neill et al. 2020). The isolates tested were obtained at different ages of the same infant: *B. longum* subsp. *infantis* LH277 at 37 days; *B. pseudocatenulatum* LH656-662 at day 159 (all from the same stool sample). *B. breve* LH24 was isolated from a different infant at 174 days after birth. It was hypothesised by the authors that that the breastfed infant microbiome harbours HMO degrading strains of *Bifidobacterium* (Lawson, O'Neill et al. 2020). The day 37 isolate of *B. longum* subsp. *infantis* LH277 possessed 3 complete HMO cluster homologues and could utilise two HMOs for growth, 2′FL and LNnT. *B. breve* LH24 possessed 5 clusters, however, not the *lac* cluster, and did not grow on 2′FL but on LNnT. Meanwhile, the *B. pseudocatenulatum* strains all lacked any complete or partial HMO clusters, phenotypically did not demonstrate utilisation of either HMO tested, but 2′FL conditioned media enhanced the growth of all tested strains, indicating the strains in this ecosystem had a propensity to share resources (Lawson, O'Neill et al. 2020). It is possible that the variation in ANI scores is caused by sequencing error but as there were

some phenotypic differences and slight variation in CAZymes detected, there appear to be genuine SNPs that might have caused these variations.

Not explicitly discussed in the original work but elsewhere is the effect of time (age of the infant) on the phenotype/genotype of strains in the infant gut microbiota, which has been recently discussed the species *B. longum* (Kujawska, La Rosa et al. 2020, Vatanen, Ang et al. 2022, Zeng, Patangia et al. 2022). The observations of LH strains isolated from an infant's stool show the carbohydrate degradative abilities are variable between donation timepoints (Lawson, O'Neill et al. 2020). The ability of an early isolate (37 days) *B. longum* subsp. *infantis* LH277 to degrade two HMOs used in growth assays was confirmed by establishing growth kinetics and presence of complete or partial HMO gene clusters. The current work demonstrates a negative phenotype for starch utilisation, which is concordant with its role as an HMO degrader. Subsequent assays in the original study demonstrated cross-feeding between HMO-degraders and non-degraders, which were isolated between 5-6 months (*B. breve* LH24 and *B. pseudocatenulatum* LH662) (Lawson, O'Neill et al. 2020). This stage of life also experiences a maturation of the infant GI tract during which endogenous enzyme production is still in development meaning that digestion of substrates prior to the large intestine can be diminished and subject to individual variation (Bourlieu, Ménard et al. 2014). The current work confirms that these HMO non-degraders contain enzyme family profiles consistent with an ability to degrade dietary substrates including starch rather than breastmilk.

Evidence of *B. pseudocatenulatum* LH strains being able to cross-feed HMO products and degrade starch *in vitro* indicates these strains could be "straddling" infant dietary niches, addressing the dilemma of being a non-HMO degrader in an HMO dominated environment (Figure 3-21). Niche shifting between cross-feeding HMO degraders whilst being prepared for arrival of solid dietary substrates ensures the presence of taxa able to thrive in multiple environments, agility that other *Bifidobacterium* species have demonstrated (Kujawska, Raulo et al. 2022). *B. pseudocatenulatum* LH strains had the highest number of CAZymes from a variety of families such as the starch related GH13, CBM48, CBM25, and CBM74, as well as arabinoxylan enzyme family GH43, otherwise known as bacterial generalism (Bell and Bell 2021). This suggests that the *B. pseudocatenulatum* LH strains have a large variety of enzymes which would make them well suited for solid food substrates normally introduced during weaning. This supports the current hypothesis of a temporal development of infant bacterial isolates to transition towards weaning prior to solid foods being introduced. Meanwhile, the *B. breve* LH24 strain (isolated from a different infant at age 6 months) was adept at degrading NMS and HylonVII, which adds weight to the hypothesis.

Bifidobacteria are a commonly used probiotic therapeutic administered to infants (Alcon-Giner, Dalby et al. 2020, Beck, Masi et al. 2022). Understanding the ecological niches bifidobacteria occupy in the infant microbiota and their ability to adapt to dietary shifts will be important for the development of precision modulation of the human microbiota through this critical stage of life. The fact that the current work demonstrates a transitional *B. pseudocatenulatum* clade before 6 months indicates the bacteria appear to shift to taxa able to utilise solid food

substrates in a similar manner to that observed in *B. longum* (Kujawska, La Rosa et al. 2020, Vatanen, Ang et al. 2022). Thus, the transitional bifidobacteria facilitating the maturation process may vary as long as they are functionally comparable. The temporal shift of taxa prior to the weaning process could be important for the timing of pre- or pro-biotic administration. Provisionally, at least in the infant the strains were isolated from, *B. pseudocatenulatum* could be considered a transitional clade to aid in the introduction of solid food substrates to the microbiome.
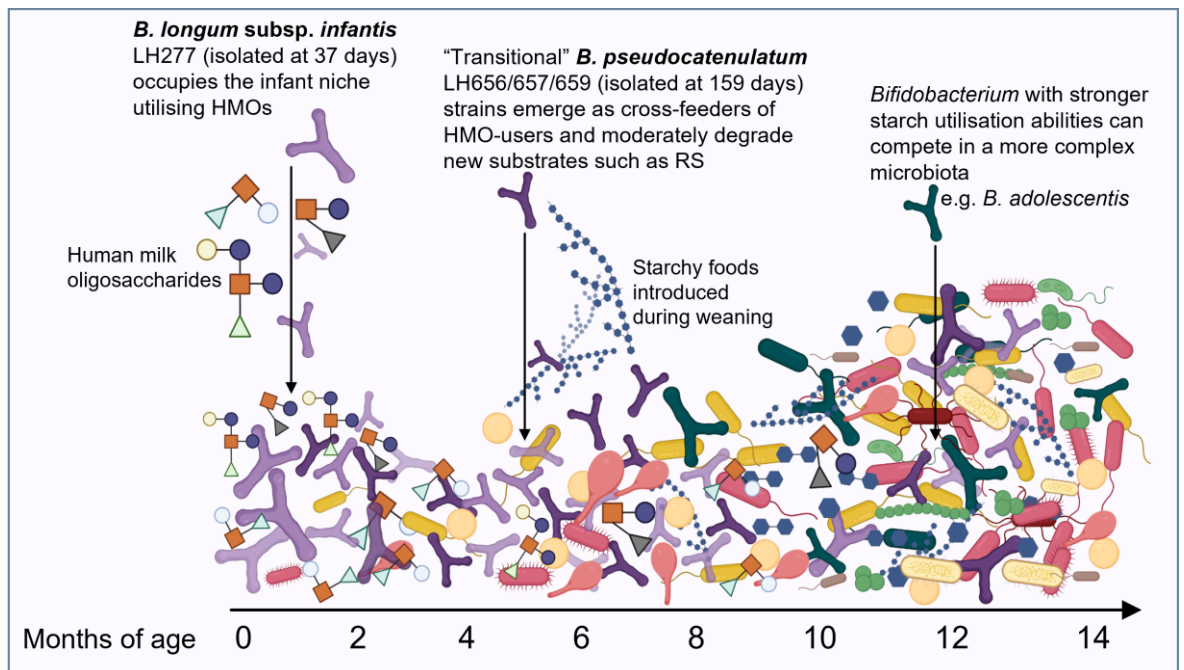
Figure 3-21 Visual representation of the time of isolation of LH strains from an infant at 37 days vs 159 after birth and the variations observed in phenotype of these isolates. Figure created using BioRender.com.

3.6 Conclusions

In this chapter, different starch structures were investigated for breakdown by a variety of different *Bifidobacterium* genus members. Raw or gelatinised potato starch and maize starches with various amylose contents were used to assess the

kinetics of degradation by various strains. There was a significant effect of amylose content on rate of degradation of gelatinised maize starch suggesting it is resistant to enzymatic hydrolysis by bacteria. The rates of bifidobacterial starch hydrolysis are relatively understudied. Detailed growth assays of 7 starch degrading strains of *Bifidobacterium* belonging to *B. breve, B. pseudocatenulatum,* and *B. pseudolongum* demonstrate a significant intraspecies variation in degradation of high amylose maize starch. Bioinformatic analysis also contributed potential mechanistic insights of utilisation differences between strains.

Metabolomics studies matched to growth curves provided valuable perspectives into the biological impacts of starch metabolism in bacteria; in the majority of cases the metabolite outputs reflected the growth kinetics on different starches. Metabolites produced (maltose and acetate) are known to positively impact the gut microbiota and host health. Understanding the variation in degradation of different types of starch and their metabolic impacts will be informative for the design of functional pro- and pre-biotics.

Chapter 4

Mechanistic insights into *Bifidobacterium pseudolongum* RS utilisation

4.1 Abstract

The mechanisms of interaction between *Bifidobacterium* spp. and resistant starch are relatively understudied. Aside from commonly accepted utilisation methods including expression of amylase genes with or without adhesion or starch-binding properties, there have been relatively few new findings. The current chapter sought to investigate the genome, transcriptome, proteome, and metabolome of two *B. pseudolongum* isolates in the presence of starch to interrogate the biomolecular tools the cells use to degrade the substrate. The results show that there is a starch-specific gene locus containing a cluster of amylases flanked by complementary binding modules which may be co-expressed at the cell surface. This gene cluster was also found to be present in multiple species of *Bifidobacterium* including *B. adolescentis* and *B. pseudocatenulatum*. These genes were more strongly upregulated in the presence of a type 2 resistant starch (high amylose maize, HylonVII) compared to an accessible maize starch. The approach used to investigate the phenotype of starch-degrading strains has provided rationale for the use of multiomics experimental approaches and highlighted the presence of a potentially novel polysaccharide utilisation locus (PUL) in *Bifidobacterium* specific for RS.

4.2 Introduction

Resistant forms of starch interact with a variety of *Bifidobacterium* species as important members of the gut microbiome (O'Callaghan and van Sinderen 2016). In both humans and several animal species, they can act as important primary degraders of starch to facilitate cross-feeding (Duranti, Turroni et al. 2014, Rios-Covian, Gueimonde et al. 2015, Duranti, Longhi et al. 2020). The fermentation of resistant starch requires the upregulation of genes which encode a combination of carbohydrate modules which specifically bind starch and enzymes which contain moieties capable of hydrolysing $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ glycosidic linkages present in starch molecules: amylose and amylopectin (Kaoutari, Armougom et al. 2013, Duranti, Turroni et al. 2014, Turroni, Milani et al. 2016, Cerqueira, Photenhauer et al. 2020). Several strains of bifidobacteria have been identified previously to utilise starch as an energy source, some of which are in possession of a gene operon called *ApuB* which produces a bifunctional amylopullulanase enzyme (Connell Motherway, Fitzgerald et al. 2008). Other mechanisms that *Bifidobacterium* use to degrade starch are currently under-explored and the extent of starch fermentation in other species is under-studied. Identifying further proteins or operons (especially if conserved in the genus) or mechanisms of utilisation could have implications for understanding of the role of *Bifidobacterium* in the gut microbiome. New findings could be important for improving human or animal health especially during targeted bifidogenic interventions or synbiotic therapies.

Common approaches to assessing the molecular and physiological changes that drive substrate utilisation in bacteria involve measurement of its gene expression or protein production (Duranti, Turroni et al. 2014). As Gram-positive bacteria,

bifidobacteria contain a cellular envelope which is a multi-layered protective structure comprising a thick peptidoglycan cell wall (Silhavy, Kahne et al. 2010, Pasquina-Lemonche, Burns et al. 2020). This feature results in difficulty during cell lysis for extracting intracellular components. It has been suggested that the origin and recalcitrant nature of the bifidobacterial cell wall could be two-fold: thick cell walls assist bacteria in resisting low pH environments such as the human stomach, in food products, and in the acidification process during normal metabolism (Cotter and Hill 2003). This would be an evolutionary driver for the bacteria to produce a thicker, more recalcitrant cell wall. Bifidobacteria produce a range of SCFAs, the main metabolite being acetate and the most abundant SCFA present in the gut. The acidification process is of benefit to the host by inhibiting non-commensal bacterial colonisation, as has been previously demonstrated by *Bifidobacterium* acetate production (Fukuda, Toh et al. 2011). Secondly, the bifidobacterial cell wall may contribute to immunological interactions with the host (Sekine, Watanabe-Sekine et al. 1994). Due to this morphology, processes such as bifidobacterial nucleic acid extraction and genetic transformation are not without challenges. Thus, several variations were tested to control for both Bifidobacterium cell wall and interference of starch, which has been noted previously (Centanni, Lawley et al. 2018).

Two subspecies of *B. pseudolongum* isolated from domesticated animals (cow rumen and pig intestine) were identified in previous screening assays (Chapter 3) to be strong utilisers of high amylose starch and were in possession of starch-specific enzymes and binding modules. *B. pseudolongum* 45 was more capable than *B. pseudolongum* 44 at utilising high-amylose resistant starch. In this chapter,

the two strains were examined by means of genomic, transcriptomic, and proteomic analysis to ascertain the molecular mechanisms causing their phenotypic differences. The hypotheses of this chapter include that *B. pseudolongum* will have specificity to ferment resistant starch because they possess gene sequences able to produce functional proteins which bind to and break down starch. Secondly, that the production of proteins specific to starch hydrolysis would be induced in the presence of starch via a transcriptional change which upregulates starch-specific gene activity. Given some proteins have specificity to RS, the variation in starch molecular fine structure between a normal maize starch and a high-amylose maize starch may illicit a differential transcriptomic response.

4.3 Methods

Whole genome bioinformatics was carried out using genomes and tools in Table 4-1. RNA-seq and proteomics experiments were performed as described in Section 2.2.4 and Section 2.2.5, respectively.

*Table 4-1 Details of strains' WGS and bioinformatics tools used to analyse transcriptomics data. *Strain designation unresolved (Lugli, Duranti et al. 2019).*

Genus	Species / subspecies	Strain	Abbr.
Bifidobacterium	*pseudolongum* subsp. *pseudolongum*	NCIMB 702244 / DSM 20099	*B. pseudolongum* 44
Bifidobacterium	*pseudolongum* subsp. *globosum* / *pseudolongum**	NCIMB 702245 / DSM 20092	*B. pseudolongum* 45
Strain	**Description**	**Origin**	**Accession**
B. pseudolongum 44	Reference strain	Pig faeces	GCA_027945455.1
B. pseudolongum 45	Type strain	Bovine rumen	GCA_002706665.1

Bioinformatic tools	Reference	Description
pyani v0.2.7.	(Yoon, Ha et al. 2017)	Strain similarity score
BLAST	(Altschul, Gish et al. 1990)	Basic local alignment search tool
Clustal Omega	(Madeira, Pearce et al. 2022)	Search and sequence analysis tools services from EMBL-EBI
InterPro	(Paysan-Lafosse, Blum et al. 2023)	Classification of protein families using a consortium of functional annotation tools
Range Extractor DNA	(Stothard 2000)	Obtain sequences from a genome using base position information
dbCAN	(Huang, Zhang et al. 2018)	Prediction of glycan substrates for CAZymes by searching against dbCAN-sub, and for CAZyme gene clusters (CGCs)
HMMER	(Finn, Clements et al. 2011)	HMMER search for CAZyme family annotation vs. dbCAN CAZyme domain HMM database
DIAMOND	(Buchfink, Xie et al. 2015)	Search for BLAST hits in the CAZy database
dbCAN-PUL	(Ausland, Zheng et al. 2021)	Experimentally validated CAZyme containing loci known to act on a carbohydrate substrate i.e. not only the *Bacteroides sus* architype
PULDB	(Terrapon, Lombard et al. 2018)	PULs experimentally characterized in the literature

4.3.1 RNA extraction method development protocols

4.3.1.1 Cell input optimisation

The cellular inputs for Qiagen RNeasy® Mini kit are recommended between 1×10^7 and a maximum of 4×10^9 cells. The cell growth stage should be captured at the exponential phase of growth to ensure maximal transcriptional activity. To assess optimal volumes of culture to extract RNA from, culture concentration (CFU/mL) was evaluated at each timepoint (0, 12, 24 h) according to Section 2.2.1.5.

4.3.1.2 Use of RNAlater™ Stabilisation Solution

The use of RNA*later*™ Stabilisation Solution (Invitrogen™) and cryogenic storage conditions were also assessed. Two duplicate sets of bacterial cultures were used to test two variables in extracting RNA: one was stored at -80 °C in 100 µL RNAlater™ and the other set was extracted in real-time. RNA quality and quantity was compared from bacterial isolate cultures extracted in real-time and after being stored in RNAlater™ As described in Section 2.2.4.3, RNA extracted was quantified using Qubit 2.0 fluorometry and quality using the Agilent 4200 TapeStation with High Sensitivity RNA ScreenTape for ribosomal RNA prokaryotic peaks at 23S, 16S and 5S.

4.3.1.3 Extraction protocol optimisation

Three protocols were tested (Table 4-2).

Table 4-2 Overview of various protocols trialled using the Qiagen RNeasy Mini kit.

RNeasy Mini kit	Differences	Purpose for amendment
Protocol A	None	
Protocol B	Protocol A + β-mercaptoethanol	Enhance cell lysis
	+ Bead beating	Enhance cell lysis
Protocol C	Protocol B + additional centrifugation step	Removal of precipitated starch
	DNase treatment	Removal of DNA contamination

Protocol A

Protocol A is a method using the Qiagen RNeasy® Mini kit, utilised successfully for *Salmonella typhi* RNA extractions by (Waters, Tucker et al. 2022) used according to manufacturer's instructions with no amendments. Protocol A uses RNeasy® Mini kits (Qiagen) following the kit procedures as follows. A 10 mL

volume of cell culture was prepared according to the culturing methods described in Section 2.2.1.3. Bifidobacterial cells were pre-cultured for 12h and 24h and each 10 mL culture was pelleted at 4000 g for 10 mins using an Eppendorf 5810 R centrifuge. The supernatant was removed, and cell pellet was resuspended in 100 µL RNA*later*™ RNA stabilization reagent, storing for up to 14 h at 4 °C. Buffer RLT Plus was added (600 µL) and pipette mixed and transferred to the kit-supplied RNeasy spin column. Samples were centrifuged at 8000g for 30 s and flow-through discarded. Buffer RW1 (700 µL) was added to the spin column and centrifuged at 8000g for 30 s. Flow-through was discarded again, and 500 µL buffer RPE was added to the spin column. The samples were centrifuged at 8000g for 30 s and flow-through discarded. Buffer RPE (500 µL) was added to spin column and centrifuged at 8000g for 30 s. The flow-through was discarded and finally centrifuged again at 16,000g for 1 min to remove residual wash buffer. The spin column was transferred to a new snap-cap plastic microtube. RNAse-free water (35 µL) was added to spin column, and RNA was eluted at 8000g for 1 min. The RNA was stored at -80 °C.

Protocol B

Protocol B uses the Qiagen RNeasy® mini kit to perform RNA extraction from bacterial culture but with the addition of bead-beating using Lysing Matrix E tubes and homogenisation step with an MP Biomedicals™ Fastprep 24 and β-mercaptoethanol both of which enhance cell lysis. 10 mL pure culture was used as the initial starting cellular input regardless of bacterial concentration to test the method. Since cells were to be stored at 4 °C as a safe stopping point for up to 14 h during the experiment, as a precaution to preserve the RNA integrity the

protocol continued to include addition of RNA*later*™ RNA stabilization reagent to the cell pellets by resuspending the cell pellet in RNA*later*™ 100 µL and transferring to DNA LoBind Eppendorfs and thereafter kept at 4°C until ready to continue.

The same kit (RNeasy® Mini kits (Qiagen) was used alongside homogenisation with Lysing Matrix E tubes (MP Biomedicals™). In a fume cupboard, 10 µL of β-mercaptoethanol (β-ME) was added per 1 mL Buffer RLT Plus before use. The tube containing resuspend bacterial cell pellet in RNA*later*™ was mixed with 600 µL RLT lysis buffer and transferred to Lysing Matrix E tubes. Using an MPbio Fastprep 24 on speed setting 6, samples were homogenised for 60 seconds, repeated 3 times (3 mins total). Samples were kept on ice in between homogenisation. Samples were then centrifuged for 10 min at 14000 g and the supernatant/lysate was pipetted to a fresh microtube containing 700 µL 70% ethanol. This was transferred (450 µL at a time) to the kit-supplied RNeasy spin column and centrifuged at 8000 g for 15 sec. Flow through was discarded. The protocol was carried out as described prior in Protocol A.

Protocol C

In order to address starch interfering with the spin column, Protocol C was performed as per Protocol B however including an additional centrifugation at 8000 g for 2 min after 700 µL ethanol was added and before transfer to RNA spin column, as starch present precipitated in the presence of ethanol.

4.4 Results

4.4.1 RNA-seq method development

Protocol A

The RNA yield of Protocol A was below the requirement of 50 ng/µL to perform

library preparation and sequencing: glucose samples had a maximum RNA

quantity of ~13 ng/µL (Figure 4-1). Due to the low yield of RNA, the RNA*later*™

A

RNA concentration comparing real-time extraction to RNAlater™ treated

B

Bacterial concentration of B. pseudolongum NCIMB 702244 with no substrate

Bacterial concentration of B. pseudolongum NCIMB 702244 with glucose
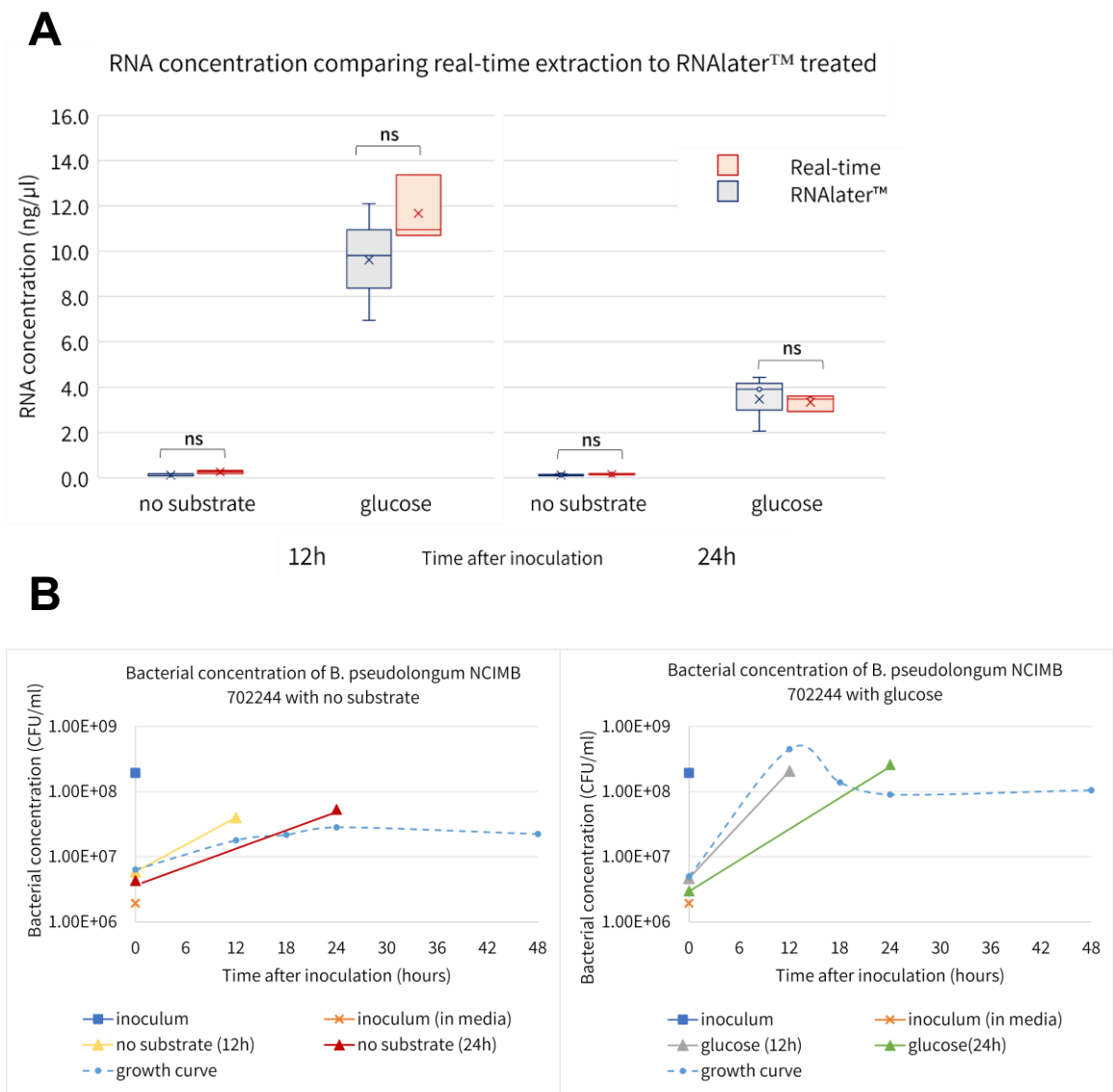
Figure 4-1 RNA extraction optimisation involving [A] RNA yield under different culturing conditions such as the presence or absence of glucose and [B] the cell count of bacteria over time to establish the volume of culture to achieving adequate RNA yield.

189

RNA stabilization results were not reliable. There were no significant differences in the RNA quantities that were obtained, but for the no substrate samples the RNA yield was below the level quantified by the assay (Figure 4-1A).

After assessing the growth of *B. pseudolongum* cells, it was determined that 2 mL from glucose, inoculum, or starch samples would be sufficient to provide a minimum of 5×10^8 cells and 10 mL would be required for no substrate samples (Figure 4-1B). Low RNA Integrity (RIN^e) values were obtained across all samples (median = 4.0; a minimum of 6.0 is required) suggesting poor RNA integrity and indicating that the process of releasing the cellular RNA needed optimisation (Figure 4-2).

Protocol B

Protocol B had improved RIN^e values compared to protocol A. The samples containing starch, however, had lower RIN^e values when compared to the other extractions (Figure 4-2).

Protocol C

This protocol failed to improve on starch interference problems as it reduced overall RNA yields (Figure 4-2). As the RIN^e values for all samples were acceptable using Protocol B, I therefore proceeded using this method. Previous studies have utilised a solid culture media containing starch to harvest cells for RNA extraction and while more convenient practically, the present data show that liquid culture can also be used (Centanni, Lawley et al. 2018).
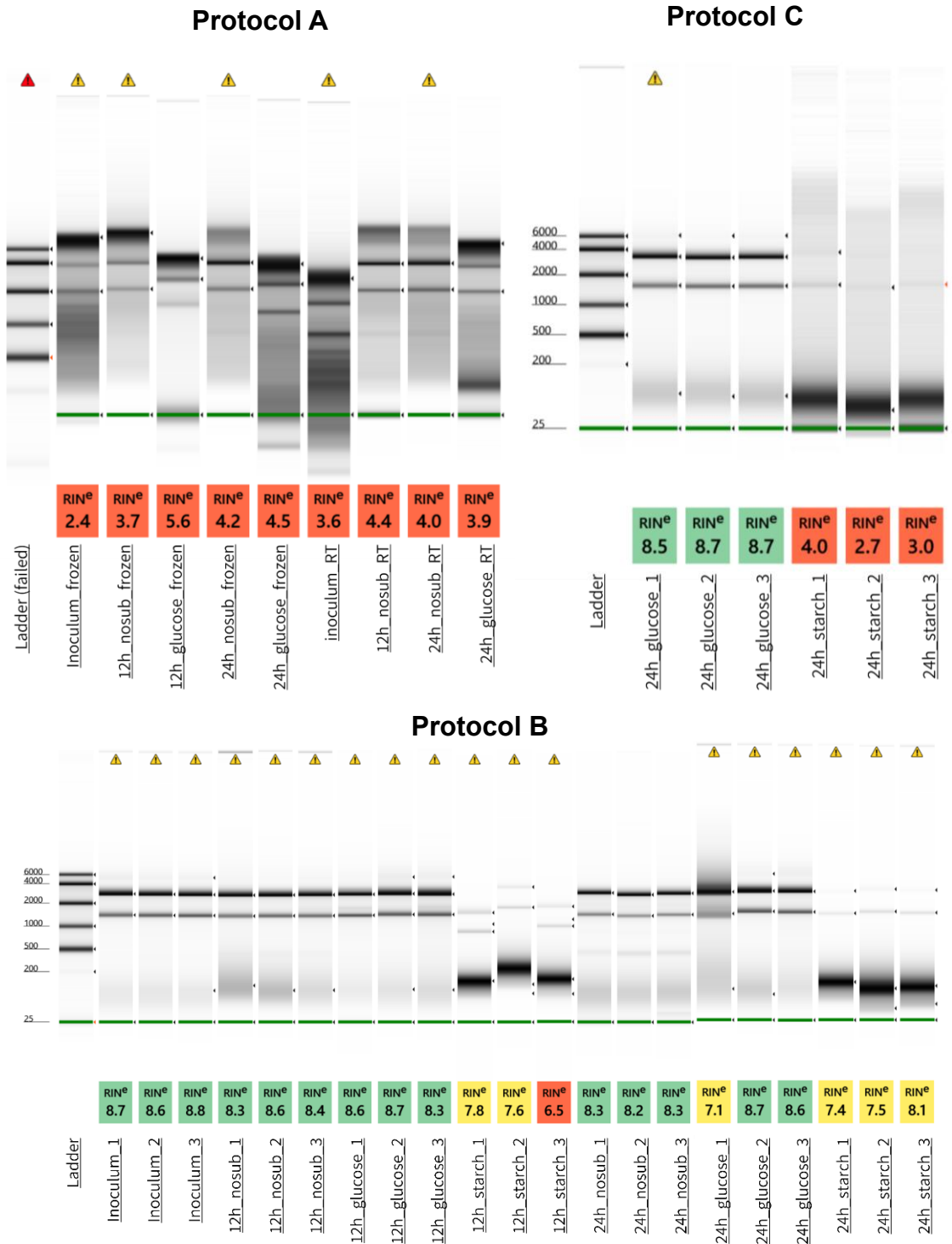
Figure 4-2 RNA integrity (RIN) values and gel images showing the varying efficacies of each protocol tested to extract RNA from Bifidobacterium. *Values and images were generated by Agilent Tapestation.*

4.4.2 Genomic analysis of *Bifidobacterium*

4.4.2.1 Strain selection and taxonomy

Due to inconsistencies in taxonomy between two strains *B. pseudolongum* subsp.

pseudolongum NCIMB 702244 and *B. pseudolongum* subsp. *globosum /*

pseudolongum NCIMB 702245, the WGS were assessed for strain similarity score

(ANI) using pyani v0.2.7 for which a score of >95% indicates that they belong to

the same species and 98% to described subspecies (Yoon, Ha et al. 2017). The

results of this analysis show that the strains are closely related with and ANI value

of 93.4% and thus not members of the same species (Table 4-3). The two strains

have a sequence identity of 91.78% as characterised by BLAST alignment of the

two WGS. These results would

warrant a recommendation to

reclassify one of the strains to a

different species, however, their

current taxonomy will be used for the

remainder of this thesis.

Table 4-3 ANI values comparing the two subsp. of B. pseudolongum.

Metric	Value
OrthoANIu value (%)	93.40
Genome B.p44 length (bp)	1,915,560
Genome B.p45 length (bp)	1,977,780
Average aligned length (bp)	1,192,078
Genome A coverage (%)	62.23
Genome B coverage (%)	60.27

4.4.2.2 Bioinformatic investigation of *Bifidobacterium* starch utilisation mechanisms

Polysaccharide utilisation gene clusters experimentally validated for degrading

starch were searched using PULDB. Defined primarily for the *Bacteroides* genus,

a PUL is a set of physically linked genes organised around a susCD gene pair

(Terrapon, Lombard et al. 2018). A total of 13 strains contained experimentally

characterised starch degradation PUL, none of which belonged to *Bifidobacterium.*

One PUL for maltose was experimentally validated in *Bifidobacterium animalis*

subsp. *animalis.* Thus, there are currently no known experimentally validated

PULs specific for starch in any *Bifidobacterium* species in the PULDB database.

The starch-utilisation gene present in *B. breve* UCC2003, *ApuB*, produces a bifunctional amylopullanase enzyme and is presumably not characterised as a PUL due to the gene product being only 1 protein, homologs of which are found in *B. adolescentis* P2P3 (Kim, Kim et al. 2021).

As described in Chapter 3, dbCAN was used to characterise the starch related CAZymes present in WGS of strains tested, which is summarised in Table 4-4. The well-characterised type strain *B. breve* UCC2003 and non-starch degrading *B. longum* subsp. *infantis* 20088 were included for comparison (as covered in Chapter 3). The results from this generally show that *B. pseudolongum* possesses more GH13 family enzymes and CBMs related to starch hydrolysis than non-starch degrader *B. longum* subsp. *infantis* 20088 (Table 4-4). However, criticisms of CAZyme count analysis include that the number of enzymes present in the genome does not always confer substrate specificity or imply rate of utilisation, especially regarding discrete molecular structures of starch (Louis, Solvang et al. 2021). This is due to the enzyme families possessing multiple substrate binding and/or catalytic functions, especially in GH13 which contains multiple subfamilies each with their own catalytic and functional designations. However, amylases (GH13 family) which contain repeated homogenous (two CBM25 domains) or heterogeneous CBMs (CBM25 followed by CBM26) have been shown to enhance adsorption to corn starch granules compared to amylases flanked by single CBMs (Dobranowski and Stintzi 2021).

Table 4-4 dbCAN results showing the number of counts for each starch-relevant family including GH13 and CBMs.

CAZyme	B. breve UCC2003	B. pseudolongum subsp. pseudolongum NCIMB 702244	B. pseudolongum subsp. globosum NCIMB 702245	B. longum subsp. infantis 20088
GH13	14	13	18	8
CBM20	0	0	0	0
CBM25	1	0	5	0
CBM26	0	0	1	0
CBM41	1	1	1	0
CBM48	4	4	5	3
CBM68	0	0	0	0
CBM74	0	0	1	0

4.4.2.3 *B. pseudolongum* subsp. *pseudolongum* NCIMB 702244

Generally, the two *B. pseudolongum* subspecies contain a number of starch-related genes with functions which could theoretically complete the steps required to degrade starch, including extracellularly expressed starch-binding amylases, intracellular amylases to degrade glycosidic bonds after import of oligosaccharides, examples of which are shown from the *B. pseudolongum* 44 genome (Table 4-5). Interestingly, there was a protein identified which has alpha amylase catalytic domains but is not experimentally characterised as such instead being defined as an Ig-like domain containing protein, which has previously been reported in other important gut species such as *Roseburia inulinovorans* and *E. rectale* (Cerqueira, Photenhauer et al. 2020). They produce a cell-wall anchored GH13 amylase that includes at least one CBM26 domain and Ig-like domains, which authors propose may be unidentified starch-specific CBMs or structurally important for degradation. In the prolific carbohydrate-degrader *B. thetaiotaomicron*, elements of its large starch uptake system contain several

starch-binding domains which contain a canonical Ig-like/b-sandwich fold attached

to starch-active enzymes (Cerqueira, Photenhauer et al. 2020).

Table 4-5 dbCAN results of B. pseudolongum *44 which are recognised in the NCBI databases as related to starch degradation.*

Protein ID	Signal peptide	GH13	Protein description	Description
WP_051918316.1	Yes(1-53)	13_28 + CBM25	Ig-like domain-containing protein	Alpha amylase catalytic domain
WP_152593139.1	Yes(1-20)	13_32	alpha-amylase family protein	signal peptide present
WP_033489309.1	No	13_31	alpha-glucosidase	oligo-1,6-glucosidase (also called isomaltase) possibly acting on imported oligos
WP_033489699.1	No	13_14	type I pullulanase	Possibly acting on imported oligos
WP_033489712.1	No	13_30	alpha-amylase family glycosyl hydrolase	oligo-1,6-glucosidase (also called isomaltase) possibly acting on imported oligos
WP_152593138.1	No	13_5	alpha-amylase	possibly acting on imported oligos
WP_162175189.1	No	13_3	DUF3416 domain-containing protein	possibly acting on imported maltose (starch synthase (maltosyl-transferring) enzyme)

Using dbCAN, the WGS of *B. pseudolongum* 44 contains one Ig-like domain

containing protein which appears to be multi-modular protein containing 1396

amino acids and is comprised of GH13_32 flanked by several binding modules

(CBM25+CBM41+CBM41+CBM48) (Huang, Zhang et al. 2018)). Table 4-6 shows

the InterPro tool region functional attributions which confirms many of the catalytic

domains, however the tool does not identify the additional CBMs CBM41 and

CMB48. Therefore, it is possible the Ig-like domains are involved in starch binding

but may not have been experimentally validated or accepted by protein

homology/functional annotation tools used by InterPro.

Table 4-6 Functional annotation of each domain in a protein (WP_051918316.1) discovered by dbCAN in the B. pseudolongum *45 genome using InterPro. The current description for this protein in NCBI database is 'Ig-like domain-containing protein' which are putatively starch-binding domains.*

Region name	Functional attribution
Ig-like domain-containing protein	AmyAc_bac1_AmyA: Alpha amylase catalytic domain found in bacterial Alpha-amylases
N/A	Ca binding site [ion binding]
N/A	catalytic site [active]
CBM26	Starch-binding module 26; pfam16738
Big_2	Bacterial Ig-like domain (group 2); pfam02368
CBM_25	Carbohydrate binding domain; smart01066
YjdB	Uncharacterized conserved protein YjdB, contains Ig-like domain
SLH	S-layer homology domain; pfam00395
SLH	S-layer homology domain; pfam00395
SLH	S-layer homology domain; pfam00395

4.4.2.4 *B. pseudolongum* subsp. *globosum* NCIMB 702245

The *B. pseudolongum* 45 genome contains more GH13 enzymes and CBMs

compared to *B. pseudolongum* 44 which could explain its increased tenacity to

degrade HylonVII, a high-amylose resistant starch (Table 4-4). This result was

reproduced where colony counts were performed (in parallel to RNA sequencing

experiment to be discussed in Section 4.4.3) showing that *B. pseudolongum* 45

could degrade HylonVII at a higher rate than *B. pseudolongum* 44 (Figure 4-3).

B. pseudolongum 45 also contains CBM25, CBM26, and CBM74 whereas

B. pseudolongum 44 contains none of these modules (Table 4-4). The relatively

recently characterised CBM74 of *R. bromii* is capable of tightly binding to crystalline starches meaning a protein containing this domain is a potential candidate for understanding the RS utilisation phenotype (Valk, Lammerts van Bueren et al. 2016, Photenhauer, Cerqueira et al. 2022). In the WGS of *B. pseudolongum* 45 this CBM74-containing protein is assigned the function of Ig-like domain-containing protein (WP_099309720.1) meaning its protein function is not yet fully characterised. dbCAN confirms it contains GH13_28, CBM74, CBM25 implying that it is likely to be a starch-binding amylase enzyme (Figure 4-4).
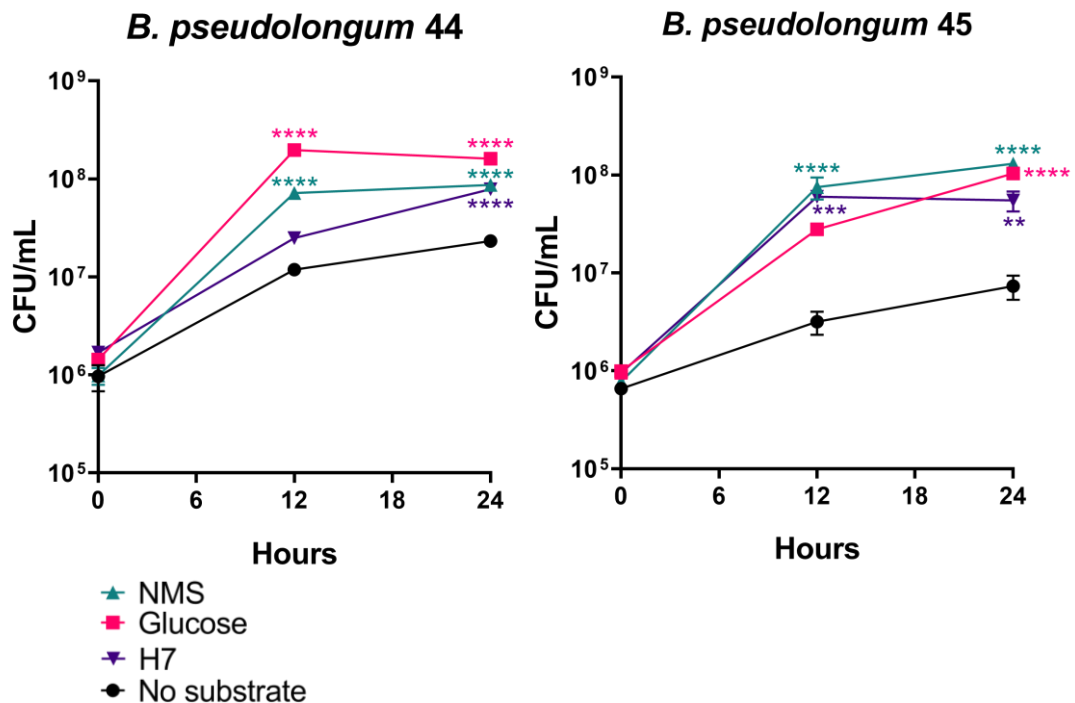
Figure 4-3 *Resistant starch (HylonVII) utilisation curves demonstrated a starch structure specific effect on bacterial growth depending on strain of* B. pseudolongum. *Strains were assessed using normal maize starch (NMS), high amylose maize (H7) using CFU at 3 timepoints between 0h 48 hours. Values displayed for each timepoint are the mean of three biological replicates. Asterisks indicate a significant difference compared to the No substrate control. **** p<0.0001 ***p<0.001 **p<0.01 *p<0.05.*

This protein is particularly large due to the bigger size of CBM74 (relative to other CBMs) and also contains a signal peptide indicating it is extracellularly expressed

(Janeček, Mareček et al. 2019). By BLAST alignment of this gene sequence, there is a % identity of 73.51% in *B. pseudolongum* 44 indicating significant protein homology between the two subspecies. The key difference between the two proteins from each strain is the absence of CBM74 module in *B. pseudolongum* 44. CAZyme gene clusters (CGCs) were identified by dbCAN revealing that this protein is a part of a cluster of genes which contains the CBM74 Ig-like domain protein (Figure 4-4).

Ig-like domain containing protein gene (WP_099309720.1)
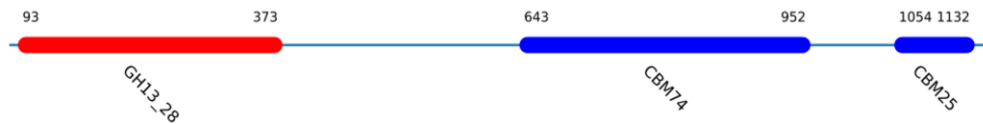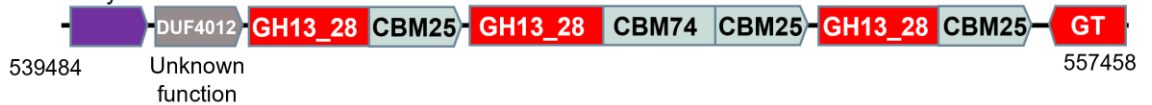
Start	Stop	Strand	Protein product	Length	Protein Name
538366	539352	+	WP_026643827.1	328	homoserine kinase
539556	540968	+	WP_236822972.1	470	polysaccharide biosynthesis tyrosine autokinase
541039	542865	+	WP_051592063.1	608	DUF4012 domain-containing protein
543286	547479	+	WP_051894162.1	1397	Ig-like domain-containing protein
547883	552475	+	WP_099309720.1	1530	Ig-like domain-containing protein
552693	556346	+	WP_162288110.1	1217	Ig-like domain-containing protein
556442	557458	-	WP_051592152.1	338	glycosyltransferase
557585	558655	+	WP_080699309.1	356	polysaccharide pyruvyl transferase family protein
558702	560207	-	WP_026644253.1	501	oligosaccharide flippase family protein
560357	561733	+	WP_026644252.1	458	Coenzyme F420 hydrogenase/dehydrogenase, beta subunit C-terminal domain
561807	562820	-	WP_080699307.1	337	glycosyltransferase
563297	564427	+	WP_026644250.1	376	EpsG family protein
564533	565120	+	WP_026644249.1	195	IS607 family transposase
565113	566345	+	WP_225531910.1	410	IS607 family element RNA-guided endonuclease TnpB
566425	567558	+	WP_033502307.1	377	glycosyltransferase family 1 protein
567567	568691	+	WP_051592015.1	374	glycosyltransferase
568761	569018	-	WP_034878930.1	85	S-layer homology domain-containing protein
569104	569991	-	WP_026643696.1	295	Ig-like domain-containing protein

Figure 4-4 Gene cluster discovered in B. pseudolongum *45 which contains hypothetical amylase domains (Ig-like domain-containing proteins) flanked by starch binding modules (CBM74, CBM25).*

The proteins are downstream of polysaccharide biosynthesis tyrosine autokinase proteins which is a transporter protein involved in anchoring extracellular enzymes to the cell wall (Mesnage, Fontaine et al. 2000). InterPro was used to further characterise each protein domain showing three open reading frames, which is evidence they are three separate proteins that are at the same gene locus (Supplementary Tables S3-S5). Proximal glycosyl transferases and another Ig-like domain containing protein outside this cluster imply it could be an important locus related to carbohydrate degradation (Figure 4-4).

Additionally, by BLAST search using the NCBI database, this particular Ig-like domain containing protein is almost exclusive to the *Bifidobacterium* genus, and most common in *B. pseudolongum* meaning it is a conserved protein for this species (Table 4-7). There were 23 total strains containing protein homology of <64.5% sequence identity, including 4 strains of *B. pseudocatenulatum* and 7 *B. adolescentis* strains. Additionally, the original protein in which the CBM74 was first discovered in *Microbacterium aurum* (MaAmyA) has significant sequence homology to this protein (35.96% identity, E value = 5e-45) (Pearson 2013). MaAmyA is a large (1417 amino acid) and multi-modular protein that contains a C terminus-located CBM74. In conclusion, proving biological function of proteins in *Bifidobacterium* in the gene cluster as involved in starch hydrolysis may be important for animal health (which are a natural host to *B. pseudolongum*) and by proposing new mechanisms of starch utilisation by human-health relevant bifidobacteria such as *B. adolescentis* and *B. pseudocatenulatum*.

Table 4-7 Homology of CBM74-Ig-like domain protein in B. pseudolongum *subsp.* globosum *700245 (WP_099309720.1) in other* Bifidobacterium *species and strains.*

Description	Query Cover	E value	% Identity	Accession
B. pseudolongum 46-E-1	99%	0	77.73	NZ_CP084313.1
B. pseudolongum 46-D-2	99%	0	77.73	NZ_CP084312.1
B. pseudolongum subsp. *globosum* DSM 20092	99%	0	77.73	NZ_CP017695.1
B. pseudolongum 46-D-6	99%	0	77.88	NZ_CP085326.1
B. pseudolongum 49-D-6	99%	0	77.88	NZ_CP084314.1
B. pseudolongum PV8-2	75%	0	78.23	NZ_CP007457.1
B. pseudolongum UMB-MBP-01	77%	0	78.28	NZ_CP022544.1
B. choerinum strain FMB-1	97%	0	78.37	NZ_CP018044.1
B. pseudolongum NjM1	40%	9.00E-142	68.41	NZ_CP071805.1
B. pseudocatenulatum YIT11953	47%	3.00E-115	65.28	NZ_CP079233.1
B. pseudocatenulatum DSM 20438 = JCM 1200	47%	3.00E-115	65.28	NZ_AP012330.1
B. pseudocatenulatum JCLA3	47%	4.00E-114	65.23	NZ_CP090598.1
B. pseudocatenulatum YIT11027	47%	4.00E-114	65.23	NZ_CP079236.1
B. adolescentis P2P3	27%	9.00E-72	68.94	NZ_CP024959.1
B. adolescentis 6	27%	9.00E-72	68.94	NZ_CP023005.1
B. adolescentis 22L	40%	3.00E-71	68.87	NZ_CP007443.1
B. adolescentis BB23	37%	3.00E-71	68.87	NZ_CP010437.1
B. adolescentis isolate MGYG-HGUT-02395	37%	3.00E-71	68.87	NZ_LR698990.1
B. adolescentis PRL2019	27%	1.00E-69	68.78	NZ_CP053072.1
B. adolescentis ZJ2	29%	1.00E-69	68.78	NZ_CP047129.1
B. angulatum DSM 20098 = JCM 7096	26%	2.00E-61	66.36	NZ_AP012322.1
B. subtile KCTC 3272	30%	2.00E-61	64.48	NZ_CP062939.1
B. angulatum GT102	32%	1.00E-56	85	NZ_CP014241.1

4.4.3 RNA-seq

4.4.3.1 Sequencing read statistics

Final read statistics produced of the mapped reads indicate that there was bias

detected due to the reference strain being used belonging to *B. pseudolongum* 45

meaning that due to lack or presence of genes in one subspecies and not the

other, total mapped read counts were significantly different among sample groups (p= 8.99e-10) based on ANOVA test. The total read counts bias was correlated with Strain (p = 6.58e-12). The number of reads mapped to the RNA sequences detected was lower in *B. pseudolongum* 44 (~80%) compared to the *B. pseudolongum* 45 (~95%) to which the reference sequence belongs (Supplementary Figure S1). The practical implication of this is the fact that some genes present in *B. pseudolongum* 44 were not mapped and thus were not analysed for differential abundance.

The read statistics for *B. pseudolongum* 44 show that there were relatively low total read counts in no substrate (m1/m2/m3) samples which was likely caused by lower cell inputs, however, the data was normalised using variance-stabilising transformation tool as part of the iDEP data processing pipeline (Figure 4-5). However, this was not observed with for *B. pseudolongum* 45 where the read counts were both higher and more even between sample types. This is likely due to the reference genome used to map the reads belonging to *B. pseudolongum* 45.

4.4.3.2 Factors impacting RNA expression

The main differences between the strains are exemplified in the PCA plot showing that the largest sources of variation in between samples are due to Strain in addition to the presence/absence of starch (Figure 4-6). Thus, for further analyses, the two datasets were analysed separately by strain. The PCA plot of the samples for *B. pseudolongum* 44 shows that the sample distance was greatest between glucose (including baseline), no substrate, and starch indicating that gene expression was different in these culture conditions (Figure 4-6). There was an

observed effect of time (between 12 and 24h) for the Glucose and HylonVII

samples. In *B. pseudolongum* 45, there was a similar pattern observed except it

appears that the Normal maize and HylonVII had a clearer separation (Figure 4-6).

B. pseudolongum 44   *B. pseudolongum 45*

Strain	Timepoint	Substrate	Replicate
44	0/ino = 0h (inoculum)	m = no substrate	1/2/3
45	12 = 12h	g = glucose	
	24 = 24h	N = normal maize	
		H = HylonVII	
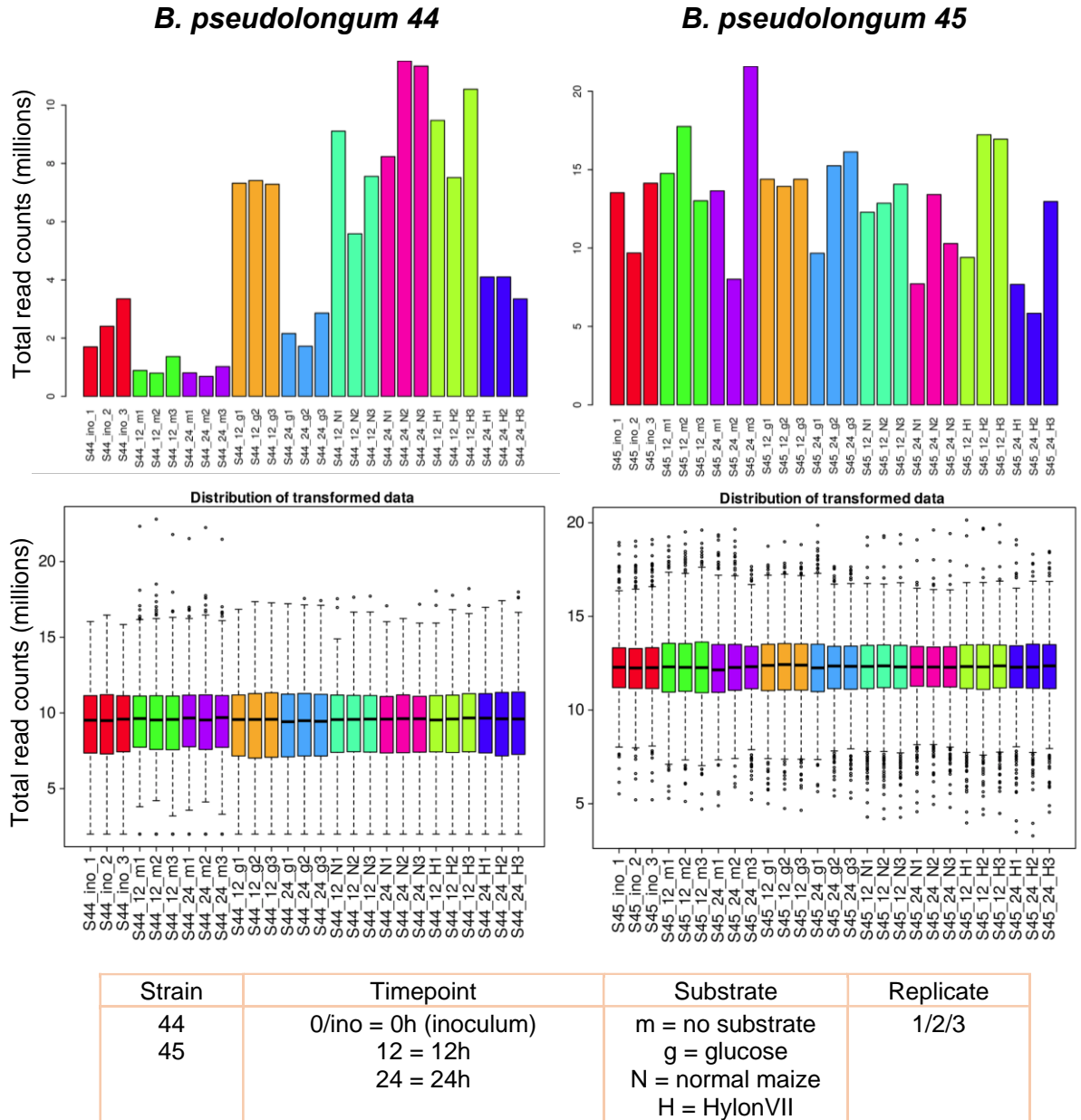
Figure 4-5 Total read counts (millions) of each strain for each sample type and replicate including [Top] raw counts and [Bottom] transformed data. Data was normalised using variance-stabilising transformation tool as part of the iDEP data processing pipeline (Ge, Son et al. 2018).
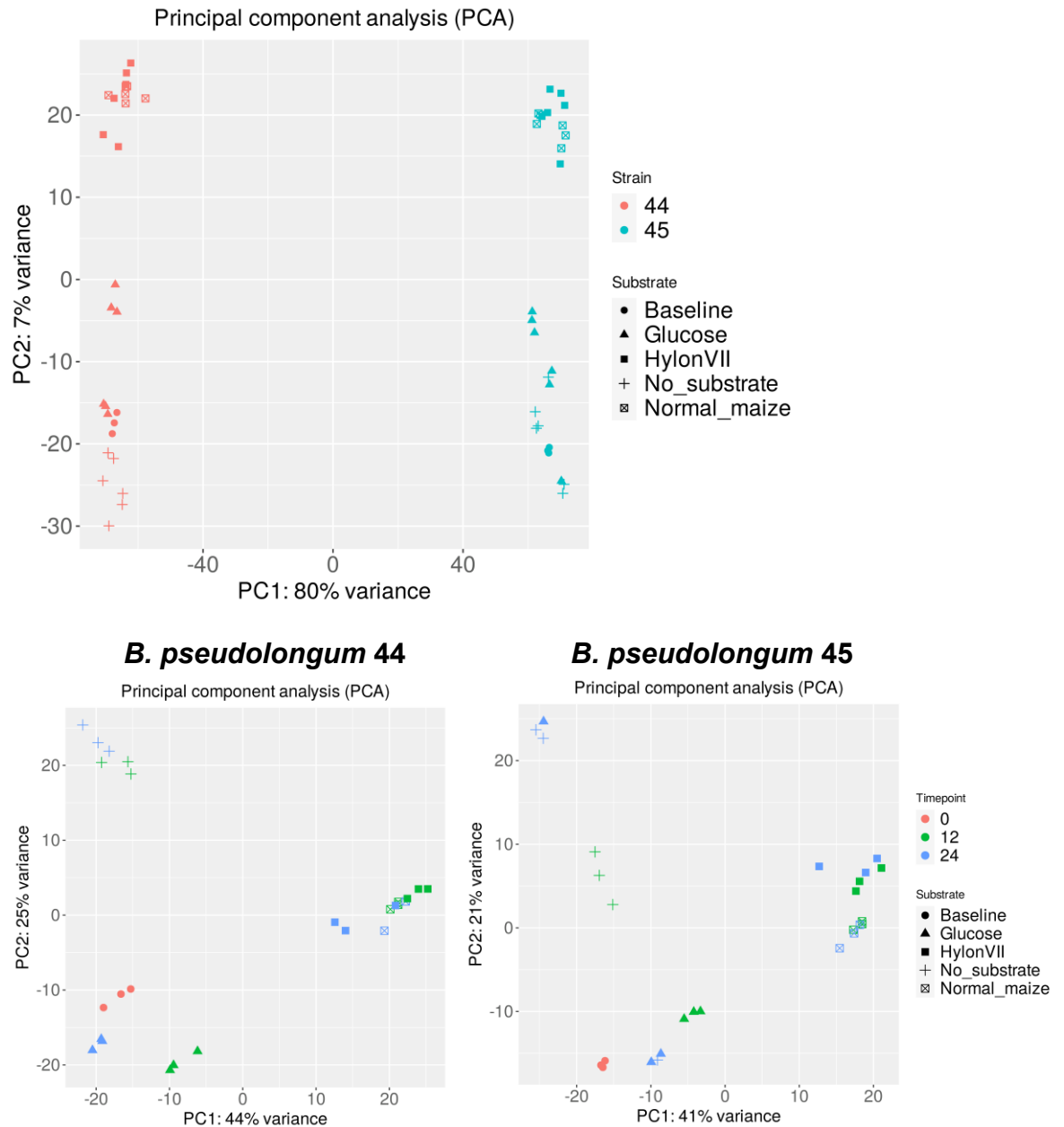
Figure 4-6 Principal component analysis of gene transcript abundance detected by RNA-seq. [Top] Displayed is the effect of strain (B. pseudolongum *44 and* B. pseudolongum *45) on transcriptional response, as well as other experimental variables such as Substrate. [Bottom] Data are stratified by strain to display the effects of timepoint and substrate.*

Of the top 100 differentially expressed genes, in HylonVII and NMS the same

complement of genes was upregulated *B. pseudolongum* 44 and 45, which

included an Ig-like domain containing protein (Figure 4-7 and Figure 4-8). The

protein was upregulated in NMS and HylonVII for *B. pseudolongum* 44 but only in

HylonVII for *B. pseudolongum* 45. In the starch condition (NMS and HylonVII) the same genes were co-upregulated compared to non-starch conditions (Figure 4-9).

The total number of up and down regulated genes in each condition for each strain shows that there are more genes upregulated in no substrate versus all other conditions which could be due to nutrient depletion triggering a starvation response (Figure 4-10) (Fung, Chan et al. 2010). NMS and HylonVII had more upregulated genes than downregulated compared to the glucose group which is a more appropriate condition to compare as the cells are also performing carbohydrate metabolism. A small number of genes were found to be more upregulated in HylonVII compared to NMS suggesting that starch structure impacts bacterial gene expression and/or enzyme response but to a lesser extent.
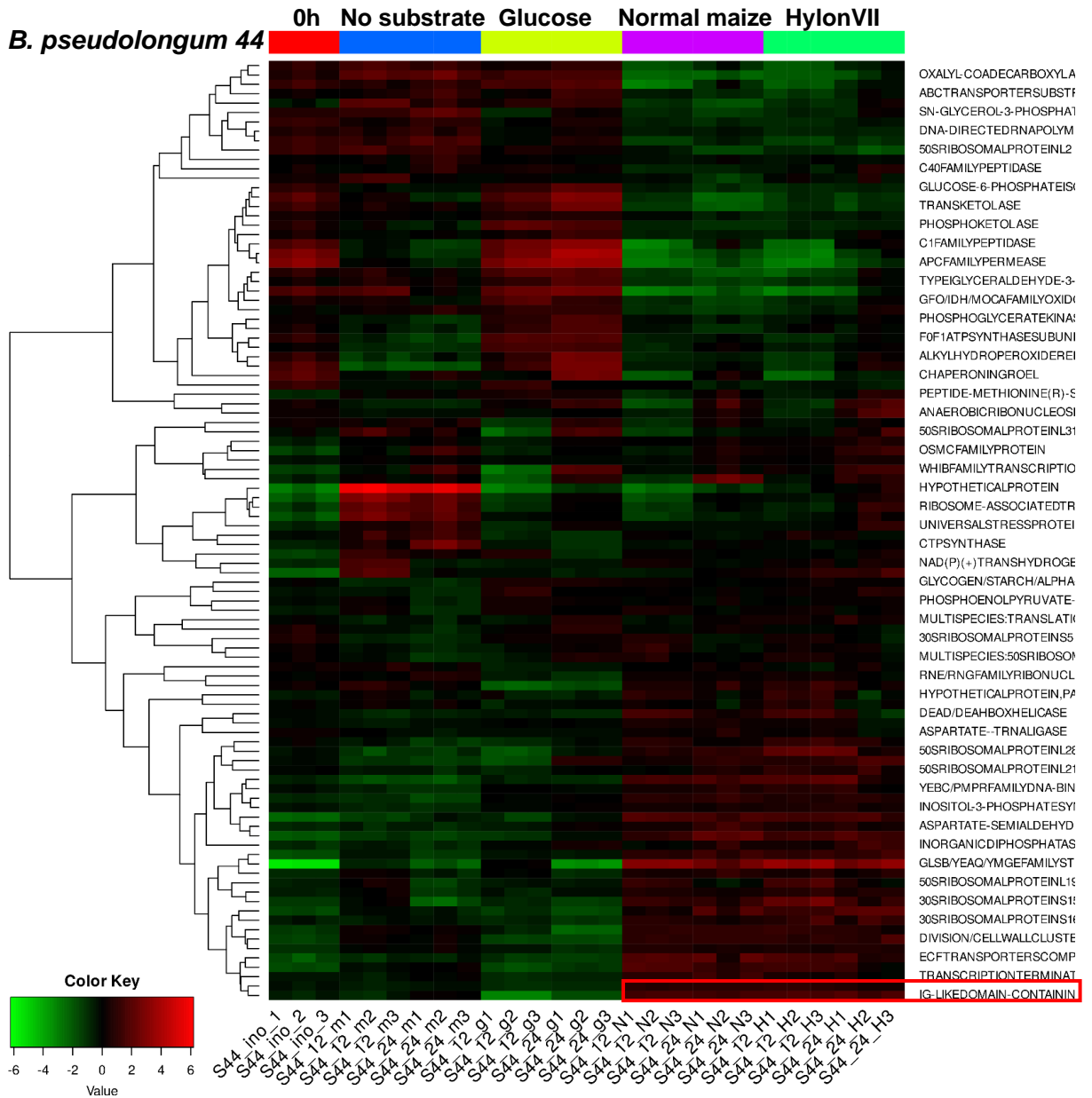
Figure 4-7 Differential gene expression of B. pseudolongum *44 clustered by substrate compared to 'baseline' transcription at 0h. A positive value (red) indicates an upregulation of genes compared to negative values (green) which indicate downregulation of gene expression.*
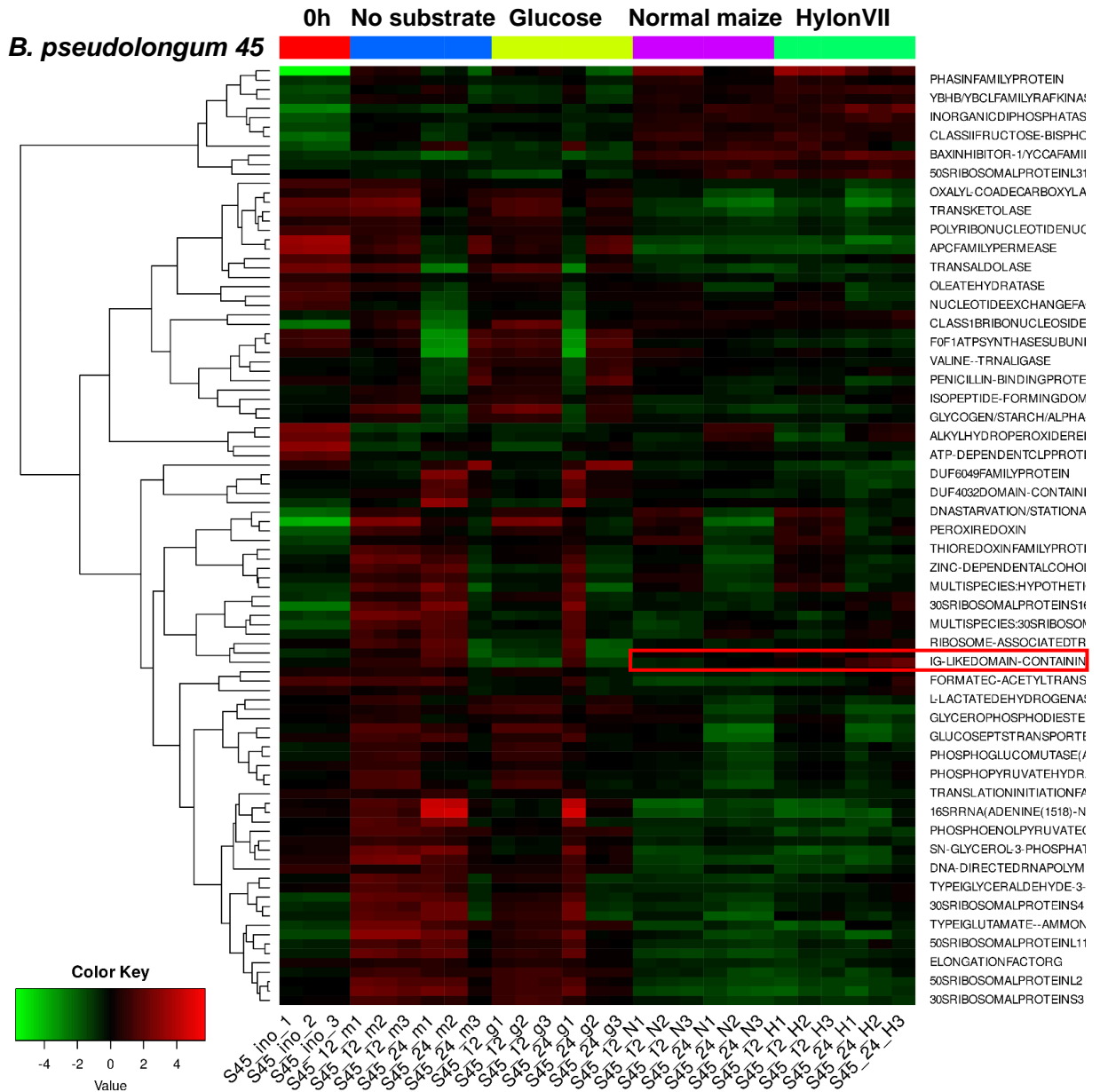
Figure 4-8 Differential gene expression of B. pseudolongum *45 clustered by substrate compared to 'baseline' transcription at 0h. A positive value (red) indicates an upregulation of genes compared to negative values (green) which indicate downregulation of gene expression.*
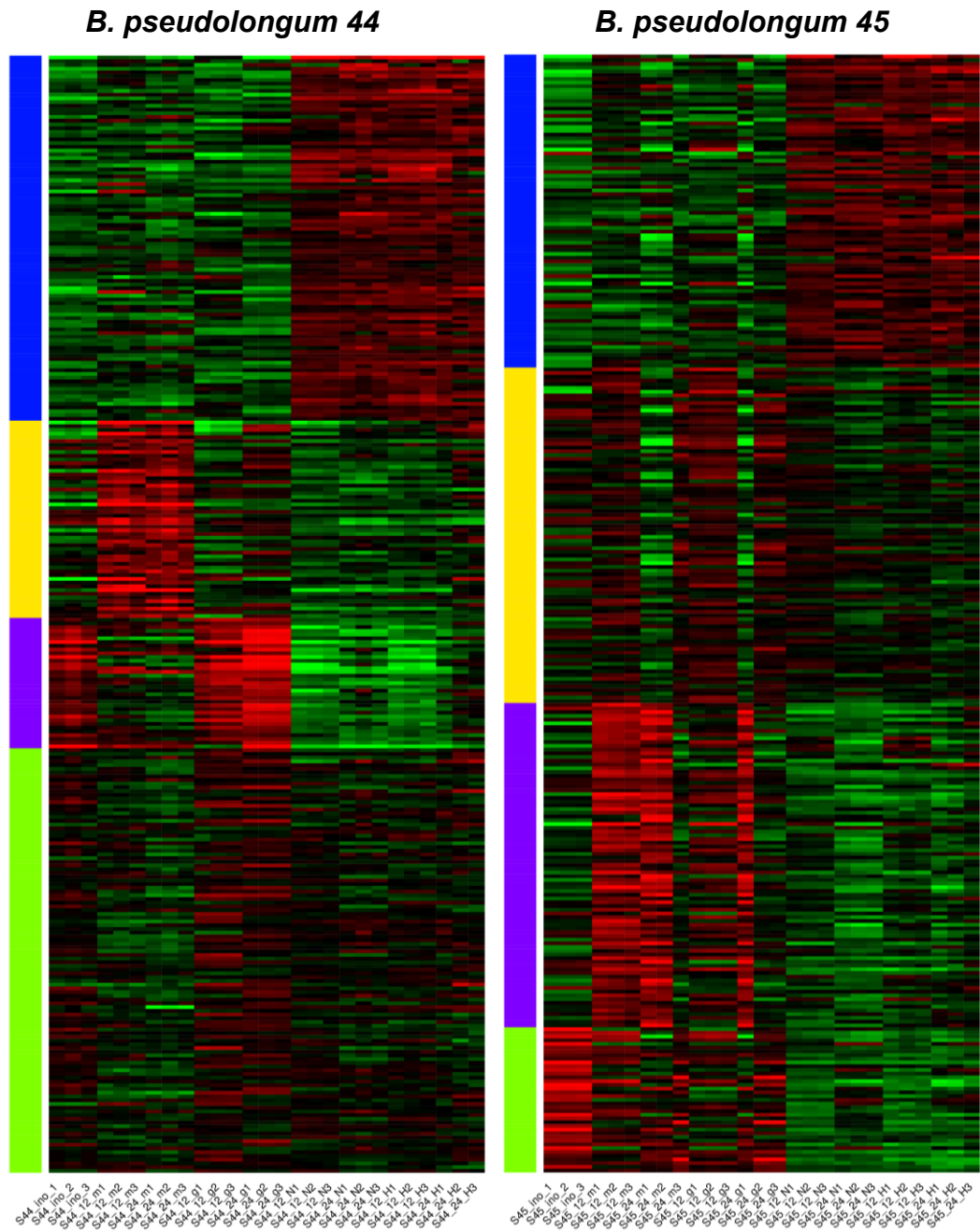
Figure 4-9 Differential gene expression of both strains of B. pseudolongum *clustered by genes up or downregulated together per sample.*
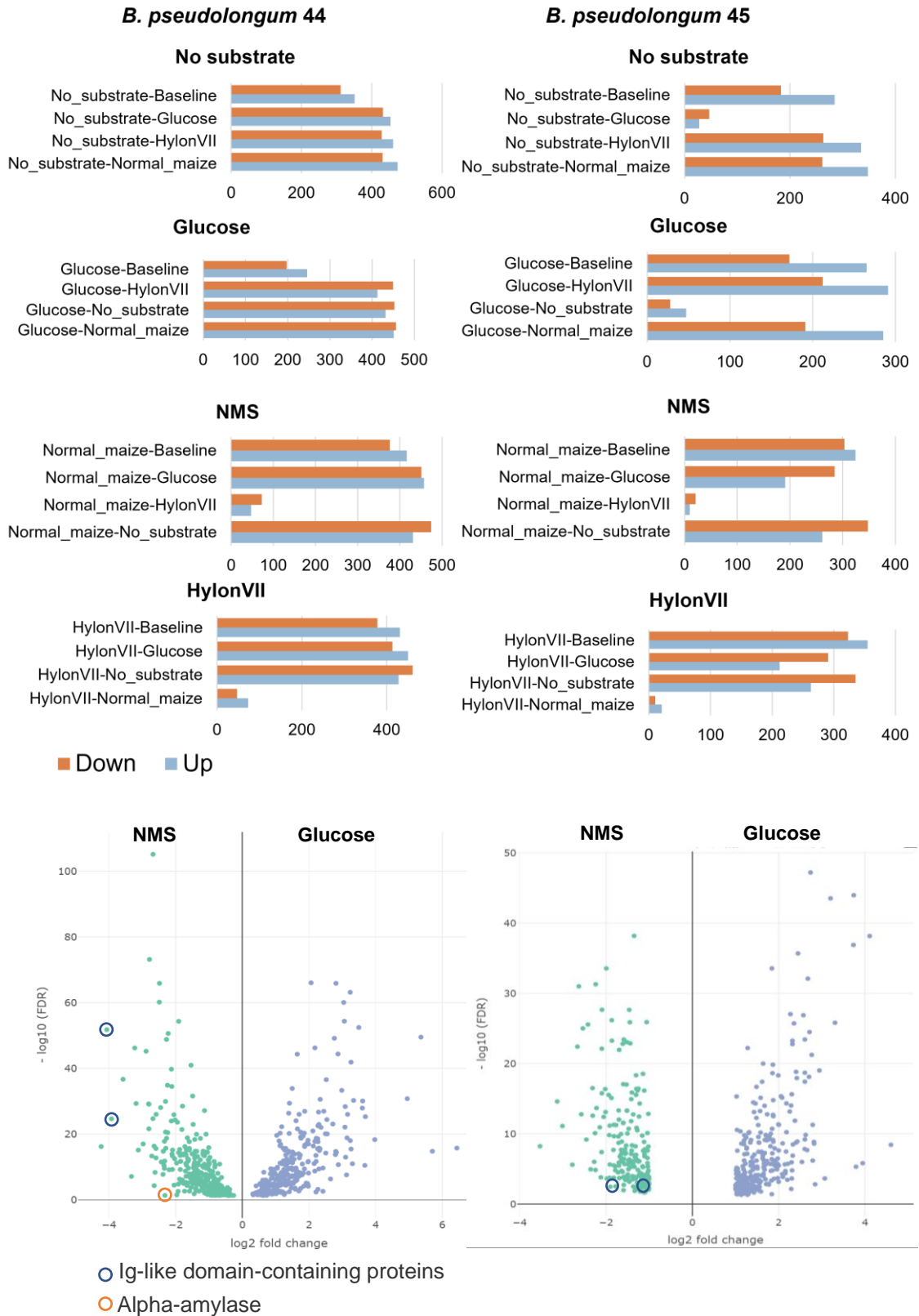
Figure 4-10 [Top] Total number of significantly up or down regulated genes (pairwise analysis of two substrate conditions) (q>0.05) performed by DESeq2. [Bottom] Volcano plots comparing the differential gene expression of NMS and Glucose.

4.4.3.3 Differential gene expression of specific genes

Specific genes significantly upregulated (q <0.05) in glucose and both starch conditions compared to no substrate is suggestive of involvement in sugar recognition, internalisation, and intracellular metabolism (Table 4-8 and Table 4-9). This hypothesis is supported by the fact that several genes upregulated are characterised as ATP-binding cassette proteins (ABC) commonly involved in sugar transportation across membranes (Rees, Johnson et al. 2009).

Table 4-8 B. pseudolongum *44 upregulated gene transcripts with a role in sugar utilisation. Asterisk denotes upregulated genes in both strains tested.*

B. pseudolongum 44

Gene product fold increase vs no substrate	Glucose	NMS	HylonVII
FAD-dependent oxidoreductase	4.30	1.62	1.99
DUF4191 domain-containing protein	3.42	1.24	1.84
Fe-S cluster assembly ATPase SufC	2.99	2.55	2.88
F0F1 ATP synthase subunit gamma	2.97	2.49	1.63
DNA-directed RNA polymerase subunit omega	2.56		1.30
class 1b ribonucleoside-diphosphate reductase subunit beta	2.55		2.44
PspA/IM30 family protein	2.39		2.26
UMP kinase	2.32	2.65	2.99
*ABC transporter permease	2.22	1.17	
aldo/keto reductase	2.15	2.46	2.10
SufS family cysteine desulfurase	2.08	2.93	3.49
*ABC transporter ATP-binding protein	1.87	2.41	1.35
*DUF2207 domain-containing protein	1.76	3.35	2.68
DUF4190 domain-containing protein	1.72	2.50	2.77
glycogen/starch/alpha-glucan phosphorylase	1.63	1.12	1.20
ROK family protein	1.62	2.20	2.70
hypoxanthine phosphoribosyltransferase	1.45	2.40	2.90
PFL family protein	1.36	2.42	2.19
Ppx/GppA family phosphatase	1.33	3.14	3.41
Fe-S cluster assembly protein SufB	1.17	2.32	2.79
phosphate ABC transporter ATP-binding protein PstB	1.11		1.41
*succinate--CoA ligase subunit alpha	1.02	2.61	2.33
GTP-binding DUF697 domain-containing protein	1.02	2.20	2.69
tRNA dihydrouridine synthase DusB	1.01	2.67	2.46

Several genes were significantly upregulated in both strains (marked by *). Of particular note among these genes was the pattern of several being more highly upregulated in the presence of starch compared to glucose, implying that starch caused a stronger transcriptional response. For example, the fold increase of succinate--CoA ligase subunit alpha was almost double in the starch conditions for both subspecies compared to glucose; this enzyme is directly involved in the succinate production pathway (Joyce, Fraser et al. 1999).

Table 4-9 B. pseudolongum *45 upregulated gene transcripts in the presence of starch or glucose implying their role in sugar utilisation. Asterisk denotes upregulated sequences in both strains under the same conditions.*

B. pseudolongum 45

Gene product fold increase vs no substrate	Glucose	NMS	HylonVII
*ABC transporter ATP-binding protein	2.38	2.65	3.71
*ABC transporter permease	2.22	2.30	3.52
DEAD/DEAH box helicase	1.84	2.23	2.44
DUF6466 family protein	1.53	1.67	1.09
VWA domain-containing protein	1.48	1.93	1.39
dihydroorotate dehydrogenase electron transfer subunit	1.47	2.22	1.45
*succinate--CoA ligase subunit alpha	1.46	2.66	2.64
*DUF2207 domain-containing protein	1.44	1.10	1.17
VWA domain-containing protein	1.43	1.51	
orotidine-5\\'-phosphate decarboxylase	1.34	1.06	
UPF0182 family protein	1.33		1.06
Stk1 family PASTA domain-containing Ser/Thr kinase	1.32	1.25	
phosphoribosylformylglycinamidine cyclo-ligase	1.26	2.30	2.15
orotate phosphoribosyltransferase	1.24	2.76	2.39
DUF6350 family protein	1.21	1.68	1.63
23S rRNA (adenine(2503)-C(2))-methyltransferase RlmN	1.20	1.45	1.63
WHG domain-containing protein	1.19	2.47	2.76
hypothetical protein	1.18	1.40	1.12
energy-coupling factor transporter ATPase	1.17	2.57	2.28
CDP-alcohol phosphatidyltransferase family protein	1.02	2.69	2.56
DUF6350 family protein	1.21	1.68	1.63

In *B. pseudolongum* 44 and 45, comparing glucose to NMS, proteins containing the Ig-like domain and alpha-amylase were significantly upregulated in the NMS starch condition more so in *B. pseudolongum* 44 than in 45 (Figure 4-10) indicating that these enzymes are directly related to starch utilisation. Since the reference genome used was a different assembly to that used in previous bioinformatic analysis (Section 4.4.2) the Ig-like domain protein sequence was subsequently assessed using InterProScan for characterisation of the protein domains. The protein contains glycoside hydrolase domains, namely alpha amylase, CBM25 domains, as well as a twin-arginine translocation (Tat) pathway (involved in transporting folded proteins across cell membrane), S-layer homology domain (anchoring of cell-surface proteins), Intimin (involved in adhesion to a target) and Ig-like folds (molecular recognition and binding) (Kelly, Prasannan et al. 1999, Goosens and van Dijl 2017, Alves, Fontes et al. 2021). These results are strongly suggestive that *B. pseudolongum* upregulated genes which may result in extracellularly expressed, cell-wall anchored starch-specific proteins.

Genes significantly upregulated in both HylonVII and NMS were assumed to be specifically related to starch utilisation especially if they were significantly upregulated proteins in all four categories: upregulated in the presence of both NMS and starch compared to glucose and no substrate (Table 4-10 and Table 4-11). In *B. pseudolongum* 44 and 45, the most highly upregulated proteins were related to stress response, transcriptional regulation, sugar uptake, and Ig-like domain containing proteins which have alpha-amylase catalytic domains.

Table 4-10 B. pseudolongum *44 genes upregulated in the presence of starch statistically compared to glucose or no substrate.*

Fold increase vs glucose		Gene products – *B. pseudolongum* 44	Fold increase vs no substrate	
HylonVII	NMS		NMS	HylonVII
4.22	3.71	GlsB/YeaQ/YmgE family stress response protein*	4.35	4.86
3.91	4.35	Ig-like domain-containing protein*	4.15	3.72
2.75	3.25	hypothetical protein DUF6591*	3.56	3.06
2.57	2.48	ribonuclease P protein component	3.47	3.56
3.65	3.63	hypothetical protein	3.25	3.27
3.13	-3.07	hypothetical protein		
2.23	2.43	Bax inhibitor-1/YccA family protein*	3.18	2.98
2.71	4.43	extracellular solute-binding protein	3.02	1.30
2.43	2.85	orotate phosphoribosyltransferase	2.90	2.48
2.32	2.21	primosomal protein N\\'	2.85	2.96
2.28	2.09	GuaB3 family IMP dehydrogenase-related protein	2.83	3.02
2.91	2.84	ATP-binding cassette domain-containing protein	2.82	2.90
2.29	2.46	helix-turn-helix domain-containing protein*	2.81	2.66
1.85	1.79	phosphoribosylformylglycinamidine cyclo-ligase	2.73	2.80
2.79	3.14	ECF transporter S component	2.72	2.37
3.71	4.33	helix-turn-helix transcriptional regulator*	2.70	2.08
1.49	1.38	Holliday junction resolvase RuvX*	2.67	2.78
3.57	3.31	CarD family transcriptional regulator	2.58	2.84
	1.85	energy-coupling factor transporter ATPase	2.56	1.67
1.37	1.29	YigZ family protein	2.52	2.61
2.02	1.72	galactokinase*	2.51	2.81
	1.68	alpha-glucosidase	2.42	1.27
1.22	1.58	serine--tRNA ligase	2.47	2.13
1.83	1.21	adenosine deaminase	2.03	2.66
	1.93	50S ribosomal protein L28	2.24	3.29
	2.88	ABC transporter ATP-binding protein	2.19	1.44
	3.59	hypothetical protein	2.04	1.25
	3.64	hypothetical protein	1.91	1.14
4.06	3.84	Ig-like domain-containing protein, partial*	1.91	2.13
	2.85	ABC-2 transporter permease	1.90	2.27
	2.24	30S ribosomal protein S15	1.89	2.76
2.32	1.90	class II fructose-bisphosphate aldolase	1.85	2.28
2.69		hypothetical protein	1.84	2.42
2.67	3.09	FtsX-like permease family protein	1.75	1.33
3.17	2.57	30S ribosomal protein S20	1.65	2.26
2.87	2.29	AMP-binding protein	1.40	1.99
2.47	2.99	ABC transporter ATP-binding protein	1.38	
2.61	3.43	MULTISPECIES: 50S ribosomal protein L33	1.34	
3.20	3.86	MerR family transcriptional regulator	1.34	
1.96	2.89	cystathionine gamma-synthase	1.01	
2.77	2.72	division/cell wall transcriptional repressor MraZ	1.01	1.06
2.71	1.65	FGGY-family carbohydrate kinase		
2.34	1.16	flp pilus-assembly TadE/G-like family protein		

Some stress response proteins are upregulated due to the high requirement for protein folding (unfolded proteins indicates stress as usually due to heat or chemical stress), but it could also be in response to the decreasing pH of the media as the bacteria produce SCFAs into the milieu (Duguay and Silhavy 2004, Sun and O'Riordan 2013). Bax inhibitor-1 is involved in preventing apoptosis during endoplasmic reticulum (ER) stress and acts to retain Ca^{2+} homeostasis presumably due to a high number of hydrolytic proteins being transcribed and required to be folded in the ER (Bultynck, Kiviluoto et al. 2014).

B. pseudolongum 44 had significantly higher upregulation compared to glucose or no substrate of 2 Ig-like domain containing proteins containing alpha-amylase domains which were over 4-fold upregulated. In contrast, *B. pseudolongum* 45 did not have the specific proteins related to starch hydrolysis upregulated in both the NMS and HylonVII group, although several of the upregulated genes were hypothetical or classified as domains of unknown function (DUF). Alternatively, if *B. pseudolongum* 45 is highly adapted to starch degradation it is possible the proteins are constitutively expressed and would not be differentially expressed.

One of the hypothetical proteins was determined by InterPro to have a membrane embedded region but the majority of the protein is an extracellular DUF6591 domain (InterPro accession: IPR046526). While highly conserved among several other organisms (namely, expressed in the skin of *Danio rerio* (zebrafish)), DUF6591 is commonly found in *Bifidobacterium,* established by BLAST search (UniProt Entry: B0UXS4 [Accessed 3/3/23]). It is associated with zinc-ribbon domains and in *Bifidobacterium castoris* a homolog is characterised as a

transmembrane chemotaxis sensory transducer (UniProt Entry: A0A430F739 [Accessed 3/3/23]). If DUF6591 performs a similar function in *B. pseudolongum*, it may have been upregulated for nutrient sensing (Keegstra, Carrara et al. 2022, Norris, Alcolombri et al. 2022).

Table 4-11 B. pseudolongum *45 genes upregulated in the presence of starch statistically compared (pairwise) to glucose or no substrate.*

Fold increase vs glucose		Gene products – *B. pseudolongum* 45	Fold increase vs no substrate	
HylonVII	**NMS**		**NMS**	**HylonVII**
2.66	2.45	Bax inhibitor-1/YccA family protein*	2.92	3.14
1.70	2.53	hypothetical protein	2.71	1.89
	1.78	SufS family cysteine desulfurase	2.68	2.25
2.22	2.59	hypothetical protein DUF6591*	2.65	2.43
1.65	1.69	tRNA dihydrouridine synthase DusB	2.57	2.53
	1.99	guanylate kinase	2.56	2.13
	1.64	Holliday junction resolvase RuvX*	2.55	2.35
	1.62	Fe-S cluster assembly protein SufD	2.54	2.01
2.54	2.15	hypothetical protein	2.45	2.84
1.82	2.08	galactokinase*	2.40	2.14
1.87	2.05	PFL family protein	2.37	2.19
2.14	1.89	peptide deformylase	2.32	2.57
2.63	2.43	GNAT family N-acetyltransferase	2.30	2.49
1.64	2.25	sigma-70 family RNA polymerase sigma factor	2.29	1.68
	1.34	aspartate-semialdehyde dehydrogenase	2.25	1.71
	1.34	Ppx/GppA family phosphatase	2.17	2.31
1.79	1.50	DUF2469 domain-containing protein	2.09	2.38
1.94	1.49	polysaccharide pyruvyl transferase family protein	1.92	2.37
2.09	1.68	inositol monophosphatase family protein	1.92	2.33
2.24	1.70	MerR family transcriptional regulator	1.89	2.43
2.42	2.07	protein-(glutamine-N5) methyltransferase	1.84	2.19
2.02	1.76	hypothetical protein	1.77	2.02
2.78	2.03	GlsB/YeaQ/YmgE family stress response protein*	1.72	2.47
3.53	2.69	hypothetical protein	1.71	2.55
2.10	2.10	DUF3039 domain-containing protein	1.67	1.66
3.01	2.38	helix-turn-helix domain-containing protein*	1.46	2.09
2.36	2.13	WhiB family transcriptional regulator	1.44	1.66
2.10	1.85	hypothetical protein	1.40	1.66
2.46	1.58	hypothetical protein	1.39	2.26
2.57	1.89	helix-turn-helix transcriptional regulator*	1.04	1.72

There were genes significantly upregulated in only NMS or HylonVII compared to

glucose and/or differentially expressed comparing NMS to HylonVII (Table 4-12).

Table 4-12 B. pseudolongum *44 upregulated genes only in the presence of one of the two starch substrates*

B. pseudolongum subsp. *pseudolongum* NCIMB 702244

Upregulated in NMS vs	HylonVII	
extracellular solute-binding protein	1.72	
ABC transporter permease	1.24	
sugar ABC transporter permease	1.22	
alpha-glucosidase	1.15	
carbohydrate ABC transporter permease	1.11	
NUDIX hydrolase	1.05	
MFS transporter	1.01	
Upregulated in HylonVII vs	**Glucose**	**NMS**
helix-turn-helix transcriptional regulator	3.70	
ISL3 family transposase	3.31	
hypothetical protein	3.08	
Ig-like domain-containing protein, partial	2.96	
ABC transporter permease	2.85	
glycoside hydrolase 43 family protein	2.80	
transcriptional regulator NrdR	2.80	
hypothetical protein	2.74	
FGGY-family carbohydrate kinase	2.71	1.06
DUF6508 domain-containing protein	2.67	
LytR C-terminal domain-containing protein	2.64	
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	2.63	
type I pullulanase	2.56	
ABC transporter substrate-binding protein	2.35	
flp pilus-assembly TadE/G-like family protein	2.34	1.18
GntR family transcriptional regulator	2.02	
hypothetical protein	1.75	1.83
hypothetical protein	2.15	1.73
MFS transporter		1.69
hypothetical protein		1.53
L-ribulose-5-phosphate 4-epimerase		1.52
DUF2974 domain-containing protein		1.48
PspA/IM30 family protein		1.39
hypothetical protein		1.34
HPr family phosphocarrier protein		1.28
aldo/keto reductase		1.23
phosphate ABC transporter ATP-binding protein PstB		1.20
nucleoside hydrolase		1.16
G5 domain-containing protein		1.11
formate C-acetyltransferase		1.09

In *B. pseudolongum* 44, there were few genes which were significantly highly upregulated in the NMS condition, the top result was an extracellular solute-binding protein was upregulated by 1.72-fold, putatively involved in nutrient sensing (Tam and Saier Jr 1993). Genes uniquely upregulated in HylonVII only compared to glucose (i.e. not found to be significantly upregulated in the NMS condition) included two ABC transporter genes, Ig—like domain-containing protein, type I pullulanase, and flp pilus assembly protein (associated with carbohydrate binding) (Iyer, Leipe et al. 2004). A recent investigation into *B. adolescentis* 22L RS utilisation also uncovered extracellular pilus structure which is potentially involved in substrate adhesion (Duranti, Turroni et al. 2014). These results suggest that the transcriptional response of this subspecies is partially mediated by starch structure, which could be due to the slower rate of degradation meaning that starch-related proteins are needed in higher amounts and/or for longer periods, meaning that there are more transcripts at the time of extraction.

In *B. pseudolongum* 45, there were fewer genes upregulated in NMS compared to HylonVII which is suggestive that starch structure impacts gene expression of starch-related genes (Table 4-13). Genes upregulated in NMS by ~2-fold compared to glucose/no substrate were only marginally upregulated by (>1.4-fold) in NMS compared to HylonVII which implies that the effect is relatively small. The genes upregulated in HylonVII compared to No substrate, Glucose, and NMS were representative of the effect of high-amylose starch on transcriptional response of *B. pseudolongum* 45. There were 4 Ig-like domain-containing proteins uniquely expressed only in HylonVII, as well an ABC transporter. The most highly upregulated protein upregulated uniquely in HylonVII however was an S-layer

homology domain-containing protein. S-layer homology domain is important for cell-wall anchoring of proteins, essential to this process is polysaccharide pyruvyl transferase which was also found to be upregulated in HylonVII. In Gram-positive bacteria, the protein display on the surface of the cell is important for nutrient acquisition and/or hydrolysis and this S-layer homology proteins are a conserved and important mechanism in bacteria as part of carbohydrate-binding proteins (Mesnage, Fontaine et al. 2000).

Additionally, another gene of interest more upregulated in HylonVII was VWA-domain containing protein, which are commonly known to associate or interact with other domains during formation of multiprotein complexes (Whittaker and Hynes 2002). In the whole genome sequence of *B. pseudolongum* 45, several of these proteins are proximally located indicating that they could share a transcriptional regulator or be translated as part of the same operon. The protein IDs of the cluster identified by dbCAN in Section 4.4.2 (which contains 3 proteins each containing alpha-amylase catalytic domains) match the 3 Ig-like domains found to be upregulated (Figure 4-5).

The functional analysis of the three proteins using InterPro showed that each of these proteins contains alpha-amylase, Ig-like folds, and S-layer homology domains and are proximally located to genes encoding polysaccharide pyruvyl transferase and several GTs (Supplementary Tables S3-S5 and Figure 4-4). This evidence could implicate these proteins as cell-wall anchored, starch-active, co-transcribed and transported proteins which are possibly formed into a complex.

Table 4-13 B. pseudolongum *45 upregulated genes only in the presence of one of the two starch substrates.*

B. pseudolongum subsp. *globosum* NCIMB 702245

Upregulated in NMS vs	No substrate	Glucose	HylonVII
thioredoxin domain-containing protein	2.08	1.99	1.19
hypothetical protein	1.92	1.97	1.34
dTMP kinase	1.79	1.96	1.02
dipeptide/oligopeptide/nickel ABC transporter ATP-binding protein	1.32	1.93	1.18
methylated-DNA-	1.10		1.30
extracellular solute-binding protein			1.30
ABC transporter ATP-binding protein			1.23
YraN family protein			1.21
ABC transporter permease			1.14
zinc ribbon domain-containing protein, partial			1.10

Upregulated in HylonVII vs	No substrate	Glucose	NMS
S-layer homology domain-containing protein, partial	2.30	1.71	2.06
hypoxanthine phosphoribosyltransferase	1.87		1.16
hypothetical protein	1.87	1.82	1.20
hypothetical protein	1.72	1.93	1.09
GntR family transcriptional regulator	1.69		2.09
VWA domain-containing protein	1.57	2.01	
SPFH domain-containing protein	1.55		1.02
DNA-directed RNA polymerase subunit omega	1.47		1.07
ECF transporter S component	1.46		1.18
polysaccharide pyruvyl transferase family protein	1.43		1.08
ABC transporter ATP-binding protein	1.42	1.78	1.68
DUF3046 domain-containing protein	1.36	1.59	1.15
Ig-like domain-containing protein, partial	1.31	1.83	
Ig-like domain-containing protein, partial	1.29		
Ig-like domain-containing protein, partial	1.17	1.78	1.86
hypothetical protein	1.15	1.85	1.02
CarD family transcriptional regulator	1.15		
HAD hydrolase family protein	1.09		1.20
Ig-like domain-containing protein, partial			1.54
phage holin family protein			1.28
amidophosphoribosyltransferase			1.18
MULTISPECIES: 50S ribosomal protein L33			1.14

4.4.4 Proteomics

4.4.4.1 Protein QC statistics

The protein extracted from *B. pseudolongum* 45 cells grown in glucose or NMS was adequate to perform LCMS; the yield in the starch samples was slightly diminished due to the interference of starch in the extraction process (Table 4-14). The protein input was normalised by Tina Ludwig by using different volumes of each sample to input the same mass (µg) of protein for analysis, however, there were more differentially detected proteins in the glucose samples which may be related to more efficient extraction (starch interferes with extraction) (Figure 5-16).

Table 4-14 Protein extraction was performed on cultures of B. pseudolongum *45 in the presence of 1% glucose or in 1% normal maize starch (NMS). Each condition was carried out in triplicate. Using Bradford assay to quantify protein extracts, the quantities of protein obtained are affected by starch being present.*

Culture condition	Replicate	Protein concentration (µg/µl)
Glucose	1	0.667
Glucose	2	0.509
Glucose	3	0.469
NMS	1	0.267
NMS	2	0.278
NMS	3	0.200

4.4.4.2 Differential protein production in response to starch

The top upregulated proteins in the NMS condition confirmed the findings of the RNA-seq experiment (Figure 4-11). Namely, the three Ig-like domain-containing proteins contained in the CBM74-domain gene cluster, and additional GH13 protein and additional adjunct proteins which orchestrate the process such as transcriptional regulator WhiB, substrate sensing and recognition by extracellular solute-binding protein, ABC transporter substrate-binding protein, and G5 domain containing proteins (Bateman, Holden et al. 2005, Maqbool, Horler et al. 2015,

Alves, Fontes et al. 2021). The three Ig-like domain containing proteins discussed earlier were the top upregulated proteins and the upregulation of VWA containing protein adds weight to the possibility these proteins are a multi-protein complex anchored extracellularly to the cell wall (Whittaker and Hynes 2002). PFL family protein is involved in the acetyl-CoA and formate production pathway (Leppänen, Merckel et al. 1999).

However, the Ig-like proteins upregulated in NMS detected by proteomics contradicts the finding that these proteins are transcribed in the presence of HylonVII but not NMS. A possible explanation for this is that these proteins were transcribed and translated exceptionally quickly. The transcripts may not have been detectable at 12h (when RNA was extracted) because they were already translated, which is plausible due to growth plateauing after 12h in the presence of NMS (Figure 4-3).

In the presence of HylonVII, the transcripts may have been (a) more strongly upregulated and detectable by RNA-seq or (b) transcribed for a longer period due to HylonVII being more difficult to hydrolyse. In either case, these data show that HylonVII initiated a stronger transcriptional response of these novel amylase genes. The effects of time on the production of transcripts vs proteins would be important to investigate more thoroughly, especially regarding the earlier stages of growth kinetics (between 0-12h).

Ig-like domain-containing protein

G5 domain containing protein

VWA domain-containing protein

Peptide counts	Fasta headers	# proteins	-LOG10(p)	Mean diff.	Enzyme activity	Reference
	Maxmium p value threshold = 0.05		1.30			
27;27;16;14 72;60;57;10;6	WP_162288110.1| Ig-like domain-containing protein	4	3.84	1.51	alpha-amylase +binding	This study
;6;6;2	WP_051894162.1	Ig-like domain-containing protein	8	4.62	1.51	alpha-amylase +binding	This study
41;41;32;28	WP_099309720.1| Ig-like domain-containing protein	4	4.41	1.28	alpha-amylase +binding	This study
7	WP_026643470.1| VWA domain-containing protein	1	3.35	1.27	multi-protein complexes	Whittaker, 2002
11	WP_026643911.1| G5 domain-containing protein	1	5.32	0.92	substrate recognition	Bateman, 2005
12	WP_026643667.1| G5 domain-containing protein	1	4.50	0.65	substrate recognition	Bateman, 2005
25;25;2;2	WP_160268232.1| S-layer homology domain-containing protein	4	2.93	0.55	Cell-wall anchoring	Alves, 2021
9	ATO39409.1| nucleoside hydrolase	1	3.84	0.55	DNA replication	Palatnik-de-Sousa, 2019
10	ATO39702.1| ABC transporter substrate-binding protein	1	2.27	0.54	sugar transport	Maqbool, 2015
24	ATO39615.1| PFL family protein	1	3.97	0.43	SCFA anabolism	Leppänen, 1999
7	ATO39894.1| WhiB family transcriptional regulator	1	3.30	0.39	gene upregulation	Khoroshkin, 2016
15	WP_026644244.1| glycoside hydrolase family 13 protein	1	2.37	0.24	isomaltase	Kumar, 2011
25	WP_026644126.1| extracellular solute-binding protein	1	1.71	0.22	chemotaxis	Roland, 1993

Figure 4-11 [Top] Volcano plot to show differential protein expression of B. pseudolongum 45 in the presence of glucose or NMS. The purple box indicates significant results of t-test p<0.05 visualised on a log(p value) scale. [Bottom]

4.5 Discussion

The health contribution of RS fermentation in the gut is well-understood whilst the mechanisms which members of *Bifidobacterium* use to do so are limited. Previous studies investigating differential RNA expression in *B. pseudolongum* strains highlight the upregulation of known starch-active enzymes in response to high-amylose maize starch (Centanni, Lawley et al. 2018). Authors of the study highlighting a novel RS-binding domain as part of a large and multi-domain α-amylase from *Microbacterium aurum* (1417 aa) highlight that the starch-binding module CBM74 is ubiquitous among *Bifidobacterium* α-amylases (Valk, Lammerts van Bueren et al. 2016), which was reinforced in a later study showing an increase in CBM74 sequences in assembled metagenomes in response to highly resistant starch (Ravi, Troncoso-Rey et al. 2022).

The present study provides strong evidence of the essential role of the CBM74 domain in degradation of RS by intentional selection of two *B. pseudolongum* subspecies which utilised hydrothermally processed high-amylose maize starch at different rates. The better RS-degrader contained a CBM74-flanked alpha-amylase protein in its genome which was significantly transcriptionally upregulated in the presence of HylonVII after 12 h but not a normal maize starch. This protein is also part of a cluster of genes which appeared to be co-upregulated in both transcriptomic and proteomic experiments. This gene cluster could be the first polysaccharide utilisation loci identified in *Bifidobacterium* which is more transcriptionally upregulated in the presence of a resistant high-amylose maize starch compared to a normal maize starch, and there are indications that the proteins may complexed on the cell surface (Figure 4-12).
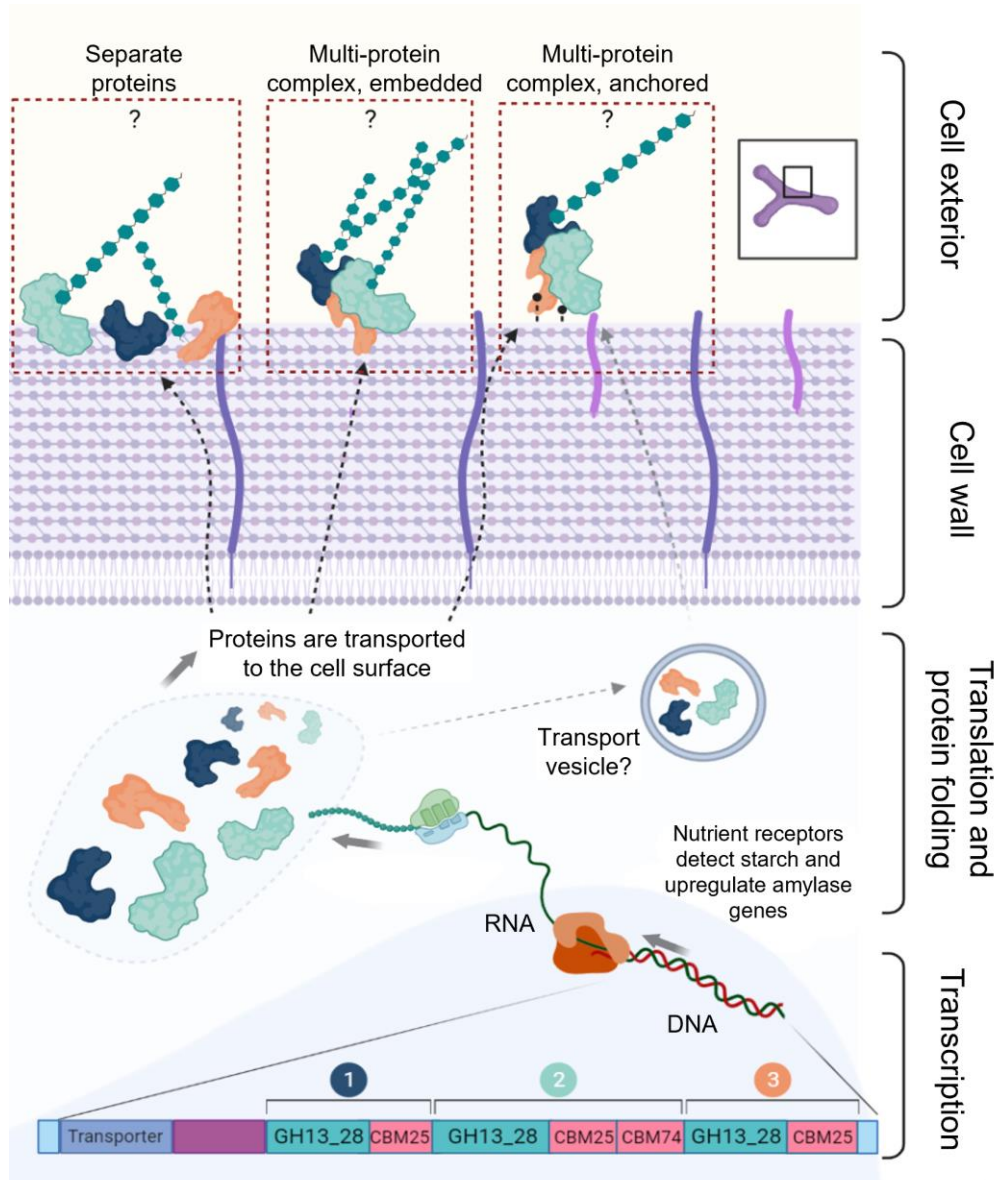
Figure 4-12 Proposed mechanism for the gene cluster discovered in B. pseudolongum *45 and potential arrangement of these proteins at the cell surface. The data proposes a novel function of the proteins which are upregulated in the presence of starch and enhance the ability of this bacterium to degrade RS. Graphic created using BioRender.com.*

It was unexpected that characterised amylolytic enzymes did not appear to be significantly upregulated such as the type 1 pullulanase or alpha amylase present in both subspecies' genomes. It is possible that these enzymes are constitutively expressed, and the Ig-like proteins are upregulated due to more efficient binding and degradation of starch when it is detected by the bacteria. The resemblance of

each of the proximally located Ig-like domain proteins in the starch utilisation cluster in *B. pseudolongum* 45 suggests they may possibly be a result of a gene duplication via the action of transposases which were also found in this genome to be upregulated, such as ISL3 family transposase in Table 4-12 (Sakanaka, Nakakawaji et al. 2018).

The presence and upregulation of sequences and proteins involved in nutrient sensing, protein transmembrane transport, cell-wall protein anchoring, substrate recognition, multi-protein complex facilitators, sugar uptake, and SCFA production provide a deeper level of insight into the molecular biomechanisms that underpin starch utilisation by *B. pseudolongum*. Commonly found during assembly of bacterial extracellular enzyme complexes are cohesion and dockerin molecules which were not investigated in the present study and may only be prevalent in other taxa such as *R. bromii* (Ze, David et al. 2015). Whilst the VWA-domain can plausibly facilitate multi-protein complexes, further evidence of complex assembly is needed to confirm if these proteins are physically associated (Whittaker and Hynes 2002, Ze, Ben David et al. 2015, Louis, Solvang et al. 2021). The upregulation of these proteins in response to a high-amylose starch implies that there are characteristics of the substrate which initiate transcriptional changes and protein production, which has been shown in *Bifidobacterium* previously (Kelly, Munoz-Munoz et al. 2021). A slower hydrolytic rate of breakdown may allow the transcriptional changes to continue whilst the substrate is still being hydrolysed. Alternatively, hydrolytic products (sugars, oligosaccharides) of high-amylose starch may act as a molecular signal. Although in *B. pseudolongum* 45, there was not a significant difference in the hydrolysis rate between NMS and HylonVII or lag

phase difference, suggesting there may be aspects of the molecular starch structure which controls the regulation of RS-degrading enzymes.

Finally, the starch-degrading strains of *B. pseudocatenulatum* evaluated for starch-utilisation phenotype were found to contain CBM74 and could be important taxa for the transition of the infant to the mature microbiome. The presence of these RS binding domains isolated from infants could be indicative that the domains are more essential for general RS degradation in the human gut than was previously recognised. Thus, these new findings of starch-utilisation mechanisms in *B. pseudolongum* could be more widely applicable to other beneficial taxa.

4.6 Conclusion

Proteins containing the CBM74 domain and other Ig-like domain containing proteins are upregulated in response to starch utilisation and the genes are differentially expressed in response to a more resistant starch structure. These Ig-like proteins therefore may be proposed as novel starch-degrading enzymes which are found to be present in other *Bifidobacterium* spp. such as *B. adolescentis*. There is evidence of a cluster of enzymes in *B. pseudolongum* 45 which may be proposed as a novel starch utilisation gene cluster, or a PUL, which would be the first of its kind in *Bifidobacterium*. This new knowledge could have an impact on how the efficiency of RS digestion in the mammalian gastrointestinal tract is understood with regard to specific enzymes and their associated substrate binding modules.

Chapter 5

In vitro fermentation of HylonVII starch

5.1 Abstract

Starchy foods are nutritionally important for health especially during weaning. The weaning infant microbiota is a functionally transitional ecosystem important for the maturation of the microbiome and immunodevelopment. The responsiveness and impacts of high-amylose maize (HylonVII) on the microbiome of weaning infants in comparison to the mature, adult microbiome has not been assessed. In this chapter, the impacts of HylonVII are assessed in weaning Infant (n=3) and Adult (n=3) using an *in vitro* batch fermentation model by monitoring metagenomic changes and SCFA production. In addition, supplements of *B. pseudolongum* subsp. *globosum* NCIMB 702245 and *B. longum* subsp. *infantis* LH277 were introduced to assess synbiotic efficacy.

Infant taxonomic changes after HylonVII fermentation included increases in *Clostridium butyricum* and *Enterococcus faecium,* and in one donor significant increases in *Clostridium perfringens*. Adult responses included increases in taxa such as *Bifidobacterium, Phocaeicola*, and *Erysipelatoclostridium*. Acetate was significantly higher in both the Infant and Adult group (p = 0.0026, p = 0.0013, respectively). Infant donor models produced significantly more butyrate compared to the no substrate control (p = 0.0182) and to Adult group (p = 0.0378). The effect of HylonVII plus *B. longum* subsp. *infantis* LH277 to enhance acetate production in

two Adult microbiomes was highly significant (p<0.0001), indicating its potential use as a synbiotic in spite of the strain itself not having starch degradative capabilities. Adult and Infant donor microbiome composition and ratios of acetate-butyrate varied during HylonVII fermentation. The results suggest HylonVII may benefit the weaning infant microbiome and that the role of starch during weaning could be potentially pivotal to the microbiome maturation process.

5.2 Introduction

Colon models such as *in vitro* batch fermentation systems are static, monocompartmental glass vessels containing a rich microbial medium; they are inoculated with faecal inoculum from a human or an animal as described in early studies using this method (McBurney and Thompson 1989, Adiotomre, Eastwood et al. 1990) and since reviewed widely (Guerra, Etienne-Mesmin et al. 2012, Williams, Walton et al. 2015). This system is designed to facilitate testing the effects of certain treatments on a closed microbial ecosystem derived from the inoculum not involving testing live subjects (Payne, Zihler et al. 2012). The effects of a treatment can be evaluated by testing experimental outputs such as gas production, pH measurement, metabolite concentrations, microbial composition etc. Outputs tested are often sampled multiple times over the course of the fermentation to express the changes in the output over time.

Batch fermentation model colon methodology can effectively simulate the colonic microbial milieu (Gunn, Murthy et al. 2020); in one specific example, when gas production was tested both *in vitro* and *in vivo* during inulin fermentation, the two methods correlated (Gunn, Abbas et al. 2022). Previous work has investigated the

impacts of RS in batch fermentations on aspects such as microbial community shifts and metabolite output (Fässler, Arrigoni et al. 2006, Fässler, Arrigoni et al. 2006, Warren, Fukuma et al. 2018, Teichmann and Cockburn 2021, Ravi, Troncoso-Rey et al. 2022).

The limitations of the batch fermentation model colon system lie in its lack of dynamism typical of living systems, such as continual flow of nutrients and waste products, interaction with host immune cells or organs (for example, the human gastrointestinal mucosal system), and uptake of nutrients by the gastrointestinal epithelia, including SCFAs (Herath, Hosie et al. 2020). Subsequent to these limitations is the fact that some microbes may not survive an *in vitro* environment (Guerra, Etienne-Mesmin et al. 2012). Applications of model colon experiments include utilising the batch starch fermentation supernatants in cancer cell viability assays (Fässler, Gill et al. 2007) or the potential for investigating global microbial gene expression with meta-transcriptomics experiments (Payne, Zihler et al. 2012, Bashiardes, Zilberman-Schapira et al. 2016).

The response of bifidobacteria to starch in isolate cultures has been previously assessed using growth measurement, quantitative metabolomics, transcriptomics, and proteomics to understand the interaction between strains and different starch structures previously in Chapter 3 and 4. However, the potential impact of this interaction on health is yet to be explored. Fermentable carbohydrates confer health benefits through multiple mechanisms, one of which is via the action of bacterial fermentation releasing SCFAs into the intestinal milieu which are ultimately absorbed by the large intestine and promote health (Makki, Deehan et

al. 2018). In the gut microbiome, there is a broad spectrum of different bacteria which can vary depending on factors such as age or diet. This chapter aims to investigate the ecological impacts of donor age and fermentable starch HylonVII using *in vitro* colon model methodology using 3 donors each of infant and adult age. Additionally, the effects of *Bifidobacterium* supplementation in the presence of HylonVII was assessed using a small pilot study design (Figure 5-1).

Bifidobacteria are considered important but not highly abundant taxa in the adult gut microbiome, but highly prevalent early in life (O'Callaghan and van Sinderen 2016, Derrien, Turroni et al. 2022). Strain-specific *Bifidobacterium* supplementation has benefits in human studies of both infants and adults (Fijan 2014, Alcon-Giner, Dalby et al. 2020) and proof-of-concept *in vitro* studies have recently been published (Teichmann and Cockburn 2021). Two bifidobacterial supplements were introduced into the model: *B. pseudolongum* 702245 (*B. pseudolongum* 45) and *B. longum* subsp. *infantis* LH277 (*B. longum* subsp. *infantis* 277). Being commonly isolated from pig and cow GI tracts, *B. pseudolongum* 45 is a RS degrading bacterium with enzymatic machinery which binds to and efficiently degrades HylonVII, as discovered in previous chapters, and produces beneficial metabolites of fermentation such as acetate. Since it is not commonly present in human stool, there was opportunity to test the ecological effects of introducing a "foreign" microbe able to ferment RS and if it persists in the model over time. Conversely, *B. longum* subsp. *infantis* LH277 was isolated from a young infant and does not have capabilities to degrade starch. The two strains were introduced individually or in combination.
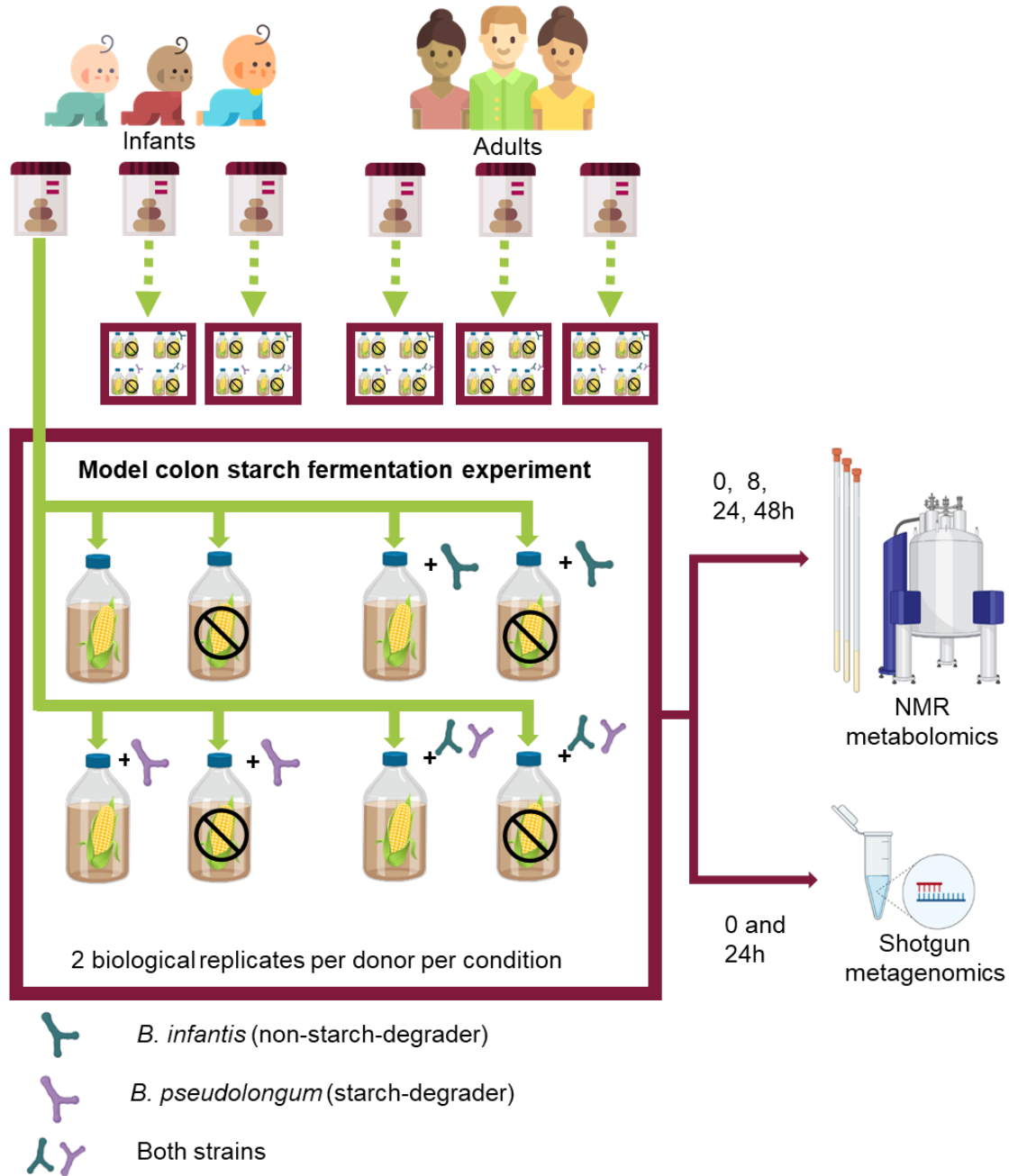
Figure 5-1 Study design for investigation of the effects of donor age and Bifidobacterium *supplementation on starch fermentation in a model colon system. Graphic created using BioRender.com.*

Several hypotheses were proposed for this chapter including that the presence of

HylonVII will illicit microbial taxa abundance increases in bacteria which are

involved in starch degradation. Second, HylonVII will cause an increase in SCFA

produced due to fermentation processes of amylolysis bacteria and cross-feeding

interactions. The effects of donor age (≤12 months or adult) on microbiome response will vary due to the differences in bacterial taxa present and their ability to degrade resistant starch. The supplementation of bacterial supplements will illicit different microbial responses depending on donor age due to large ecosystem differences. Lastly, the supplementation of *B. longum* subsp. *infantis* LH277 may illicit a different ecosystem response to *B. pseudolongum* 45 because the latter can utilise resistant starch.

5.2.1 Methodology

5.3 Methods

The colon model methodology (Section 2.2.6) was used in order to perform Illumina platform short-read shotgun sequencing of the microbial ecosystem at timepoints 0 h and 24 h under different experimental conditions (Figure 5-1). Untargeted metabolomics of the model supernatant was carried out at 0, 8, 24, and 48 h timepoints using methodology detailed in Section 2.2.7.

5.3.1 Data analysis

A number of approaches were taken when analysing microbiome datasets due to large number of variables: experimental time course, age group, individual donor (within age group), HylonVII or no substrate, and the presence or absence of bacterial supplement(s). Differences in Shannon diversity index was established using Phyloseq R package and assessed statistically by t-test in GraphPad Prism 9 by comparing simple variables e.g. Adult and Infant, 0 and 24h. Metagenomic and metabolomic datasets were used to perform linear models using Maaslin2 R package which compared two variables for each of the 28 statistical tests performed (McMurdie and Holmes 2013, Mallick, Rahnavard et al. 2021). Fixed

effects were designated as Donor number (n = 6) while variable effects were tested e.g. time, substrate, age (Supplementary Table S6). Maaslin2 parameters used to filter the data were a minimum taxa prevalence = 0.9, and minimum abundance = 0.001. Results considered significant were assessed by q value (FDR) with cut off of <0.05. The FDR was selected because while a larger number of results were statistically significant going by p value, due to the number of statistical tests performed the FDR controls for the proportion of discoveries that are false (Benjamini and Yekutieli 2001).

Unless explicitly stated, phylogeny plots and heatmaps of taxa abundance were filtered to only include taxa which had a prevalence in 20% across samples (frequency = 0.20), and an abundance threshold of 0.01 in order to detect less common taxa which may have changed which was important when comparing taxa at a species level. Metabolomic statistics were performed using Repeated Measures ANOVA in GraphPad Prism 9, in some datasets including Multiple Comparisons of the mean cell values of each variable at each timepoint.

5.4 Results

5.4.1 Metagenomics sequencing statistics and results

For the sequencing parameters selected (NovaSeq6000, 2 × 150 bp) the percentage of bases with a Q30 score of ≥ 75% is considered excellent read quality (Ravi, Walton et al. 2018). Across all 191 samples sequenced, samples returned very high-quality sequences with a mean 88.5% of bases with a Q value ≥ 30. Mean raw reads generated was 11,033,038 with a mean GC content of 46.8%. Adapter sequences were trimmed and removed by Fastp. Eight samples

were removed resulting in 183 samples with high-quality metagenomic sequence reads.

5.4.2 Alpha diversity

Shannon diversity is expressed as an index of the number of species (richness) and their relative abundance (evenness). Shannon diversity was significantly higher in Adult than in Infant samples regardless of all other variables (Figure 5-2A). Generally, diversity significantly decreased between 0h and 24h in (Figure 5-2A). However, diversity was not significantly different between no substrate and HylonVII (Figure 5-2A).

Three donors from each age group were chosen in efforts to (a) assess generalities in response between three individually unique microbiomes and (b) assess how inter-donor variation impacts microbiome response. Alpha diversity values (Shannon) were plotted grouped by donor, substrate, and supplement (Figure 5-2C). At 0h, there was close clustering of all sample types within donors meaning the sample diversity was similar at the start of the experiment (Figure 5-2B). At 0h, diversity varied between all donors (p < 0.0001) except for Infant 2 and 3 whose diversity values did not differ significantly from each other (p = 0.9983). Thus, the baseline microbiome diversity generally varied between donors, as well as the response to experimental variables. For example, donors had a significant drop in diversity between 0 and 24h except for Infant 2 and Infant 3 where diversity did not significantly decrease after 24h (Figure 5-2B). At 24h, Adult 2 and Infant 1 had a significant decline in diversity in the presence of HylonVII (Figure 5-2C) (p = 0.0147 and p < 0.0001, respectively). There were no significant differences with No substrate implying that there was a negative effect

of HylonVII on microbial diversity (p = 0.1007 in Adult 2, p = 0.2046 in Infant 1)

(Figure 5-2C).
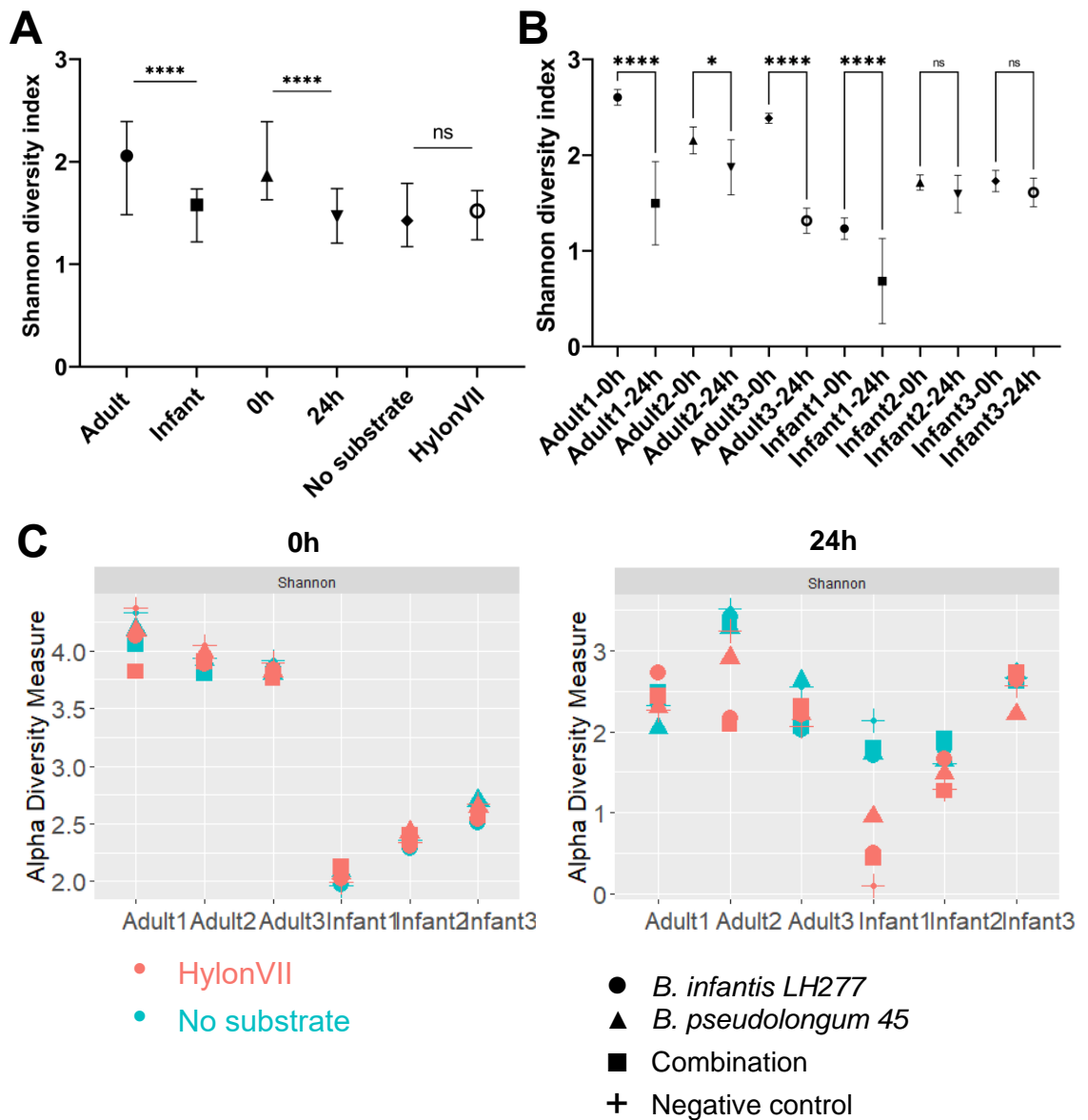
Figure 5-2 Shannon diversity plots of metagenomic reads. [A] T-test statistical comparison between Shannon values of different sample variables such as age group, timepoint, and substrate. [B] Statistical comparisons between each sample type over time (0h-24h) in the HylonVII condition. [C] Each sample plotted individually for each donor and timepoint. **** p<0.0001, *** p<0.001, ** p<0.01, * p<0.05, ns p>0.05.

Since diversity declined over time between 0 and 24h (Figure 5-2A), the effects of

time on specific taxa were statistically assessed (Supplementary Table S6, results

summarised in Table 5-1). There were 6,521 total bacterial taxa detected across all 183 samples. Among the Adult samples, 537 species were significantly more relatively abundant at 0h, whilst only 49 were more abundant at 24h regardless of substrate (Figure 5-3). In Infant samples, 260 species were more abundant at 0h and 77 more abundant at 24h (Figure 5-4). These results confirm a loss of bacterial diversity may be caused by a decrease in specific species between 0 and 24h. Taxa which are not suited to lab culture conditions or suffer lack of specific substrates for growth may contribute to a loss of bacterial taxa over time in the *in vitro* model (Williams, Walton et al. 2015).

Table 5-1 Summary of the statistical tests performed (pairwise comparisons) using Maaslin2. The number of significant effects of timepoint, age, and substrate on differences in microbial taxa are shown. Parameters using Maaslin2 were a minimum taxa prevalence = 90%, minimum abundance = 0.001, q value (FDR) cut off <0.05.

Significant differences in abundance	Results	Total significant results
Over time (0-24h) (all substrates)	Adult: 49 more abundant in 24h, 537 in 0h	586
	Infant: 77 more abundant in 24h, 260 in 0h	337
Between no substrate and HylonVII (0h)	Adult: ns	0
	Infant: ns	0
Between no substrate and HylonVII (24h)	Adult: 19 more abundant in HylonVII, 370 more in No Substrate	389
	Infant: 4 taxa more abundant in HylonVII, 106 more abundant in No substrate	110
Between age groups at 24h	2 taxa more abundant in infant; 5 more abundant in adults	7
Between age groups at 24h in HylonVII	No taxa significantly different in infant; 2 more abundant in adult samples	2

Differential microbial abundance of bacteria between 0 and 24 hours

Adult

Adult samples n = 90
Top 49/573 results shown

Adult samples n = 90
Top 49/49 results shown
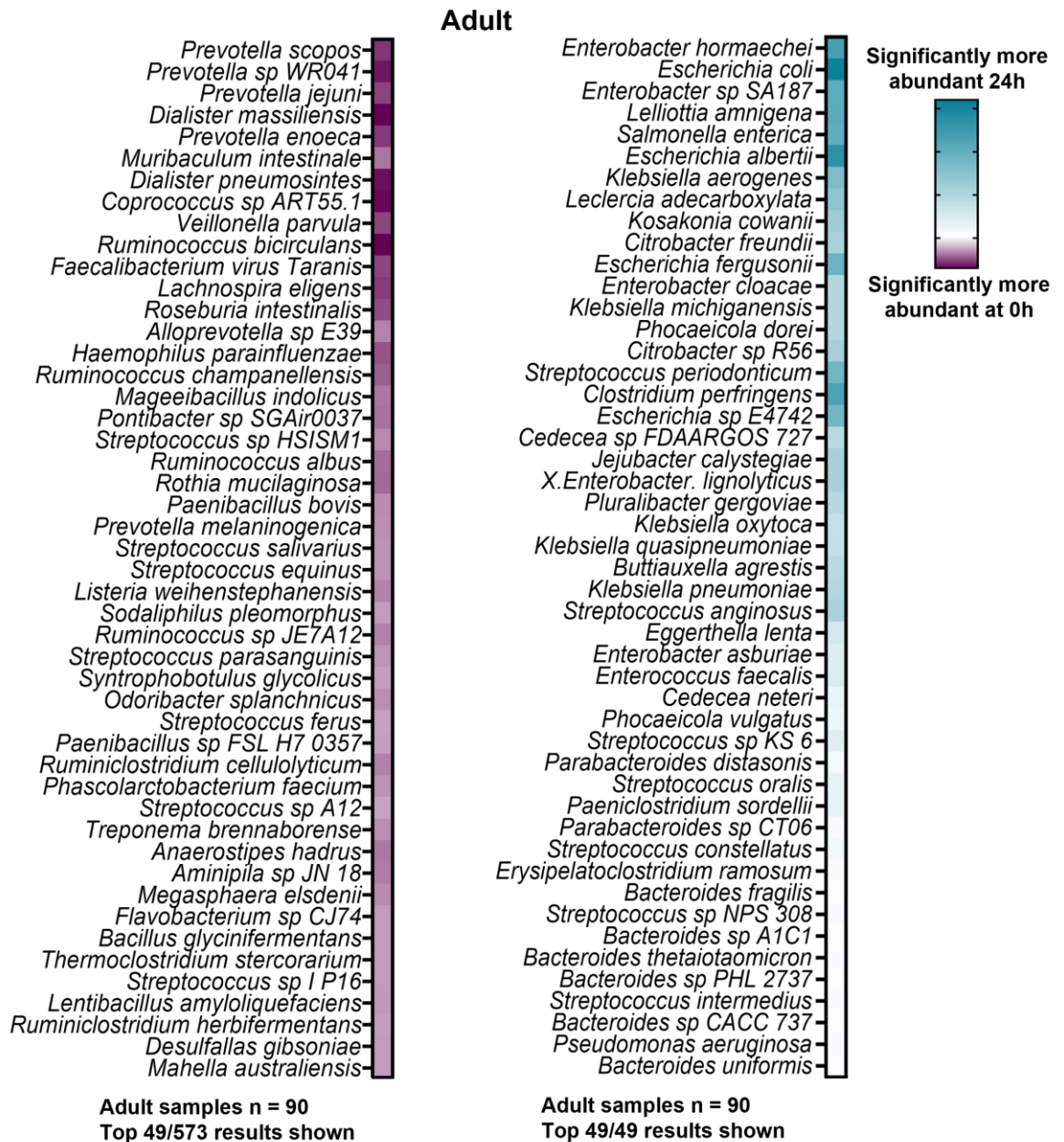
Figure 5-3 Specific taxa significantly more abundant in the Adult metagenome at the start of the experiment (0h) compared to after 24h after in the in vitro *fermentation model (q<0.05). Colour saturation represents the coefficient value (magnitude of the differential response) multiplied by the negative log of q value. Statistics performed using Maaslin2 R package.*

Differential microbial abundance of bacteria between 0 and 24 hours
Infant

Veillonella parvula	Clostridium perfringens
X.Ruminococcus. gnavus	Clostridium butyricum
Blautia obeum	Clostridium botulinum
Anaerobutyricum hallii	Clostridium acetobutylicum
Blautia sp SC05B48	Clostridium sp CT4
Coprococcus catus	Clostridium saccharobutylicum
Streptococcus infantarius	Clostridium intestinale
X.Ruminococcus. torques	Clostridium chauvoei
Longibaculum sp KGMB06250	Clostridium saccharoperbutylacetonicum
Streptococcus gwangjuense	Clostridium isatidis
Lacticaseibacillus rhamnosus	Clostridium sp JN 1
Faecalitalea cylindroides	Clostridium cochlearium
Rothia mucilaginosa	Clostridium gasigenes
Escherichia coli	Clostridium fermenticellae
Lacticaseibacillus paracasei	Clostridium taeniosporum
Coprococcus comes	Clostridium septicum
Salmonella enterica	Clostridium pasteurianum
Streptococcus mitis	Clostridium tetani
Streptococcus pseudopneumoniae	Clostridium kluyveri
Roseburia intestinalis	Clostridium sp DL VIII
Gemella haemolysans	Clostridium baratii
Eubacterium sp NSJ 61	Clostridium drakei
Christensenella minuta	Clostridium carboxidivorans
Streptococcus pneumoniae	Clostridium thermarum
Streptococcus australis	Clostridium sp JN 9
Streptococcus sp A12	Clostridium sp AWRP
Anaerostipes hadrus	Clostridium cellulovorans
Blautia argi	Clostridium argentinense
Streptococcus salivarius	Clostridium novyi
X.Eubacterium. cellulosolvens	Clostridium formicaceticum
Citrobacter freundii	Enterococcus faecalis
Streptococcus sp oral taxon 061	Enterococcus hirae
Bifidobacterium thermophilum	Enterococcus mundtii
Streptococcus oralis	Enterococcus durans
Klebsiella pneumoniae	Enterococcus thailandicus
Gemella sanguinis	Enterococcus sp M190262
Faecalibacillus intestinalis	Hathewaya histolytica
Bifidobacterium pseudocatenulatum	Enterococcus gallinarum
Blautia sp LZLJ 3	Enterococcus faecium
Streptococcus thermophilus	Tetragenococcus halophilus
Mogibacterium diversum	Enterococcus rotai
Bifidobacterium angulatum	Buchnera aphidicola
Bifidobacterium adolescentis	Fusobacterium varium
Bifidobacterium bifidum	Gottschalkia acidurici
Bifidobacterium dentium	Paeniclostridium sordellii
Lancefieldella parvula	Caloramator sp E03
Coprococcus sp ART55.1	Vagococcus hydrophili
X.Clostridium. scindens	Sarcina sp JB2
Lachnospira eligens	Listeria monocytogenes
Massilistercora timonensis	Enterococcus avium

Significantly more abundant 24h

Significantly more abundant at 0h

**Infant samples n = 93
Top 50/260 results shown**

**Infant samples n = 93
Top 50/70 results shown**

Figure 5-4 Specific taxa significantly more abundant in the Infant metagenome at the start of the experiment (0h) compared to after 24h after in the in vitro fermentation model (q<0.05). Colour saturation represents the coefficient value (magnitude of the differential response) multiplied by the negative log of q value. Statistics performed using Maaslin2 R package.

5.4.3 Beta diversity

Beta diversity indices calculate the quantitative difference in the overall taxonomic composition between samples, two of which were used: Bray-Curtis index and Jaccard index (which ignores exact abundances and considers only presence/absence of taxa). The data was transformed to adjust for sequencing depth. At 0h, close clustering of sample types confirmed high similarity of samples type within respective donor (Figure 5-5). The Adult and Infant samples were not closely grouped, which confirmed both differential composition (Bray-Curtis) as well as a presence of taxa in the Adults not found in the Infants and vice versa (Jaccard). All three Adult donors clustered closely together in both Bray-Curtis and Jaccard plots, whilst the Infants were not closely clustered together in either which implies that they had more variation in microbiome composition compared to Adults. Statistical modelling of taxa more abundance in Adult versus Infant revealed specific taxa significantly more abundant in each group (Figure 5-5).

After 24h, there is a stratification of samples due to HylonVII across all donors in Bray-Curtis but they are not as closely grouped as in Jaccard (Figure 5-5). In Jaccard, there was more distance between Adult and Infant, suggesting that time has enhanced the differences in which taxa were present/absent in each age group. In addition, there was a division of samples in each donor between HylonVII and No substrate in Jaccard, implying the presence of starch led to taxa shifting in response to starch; this was particularly prominent in Infant 1, Adult 2, and Adult 3 (Figure 5-5). Subsequently, the two age groups were analysed separately as this was the largest source of variation.
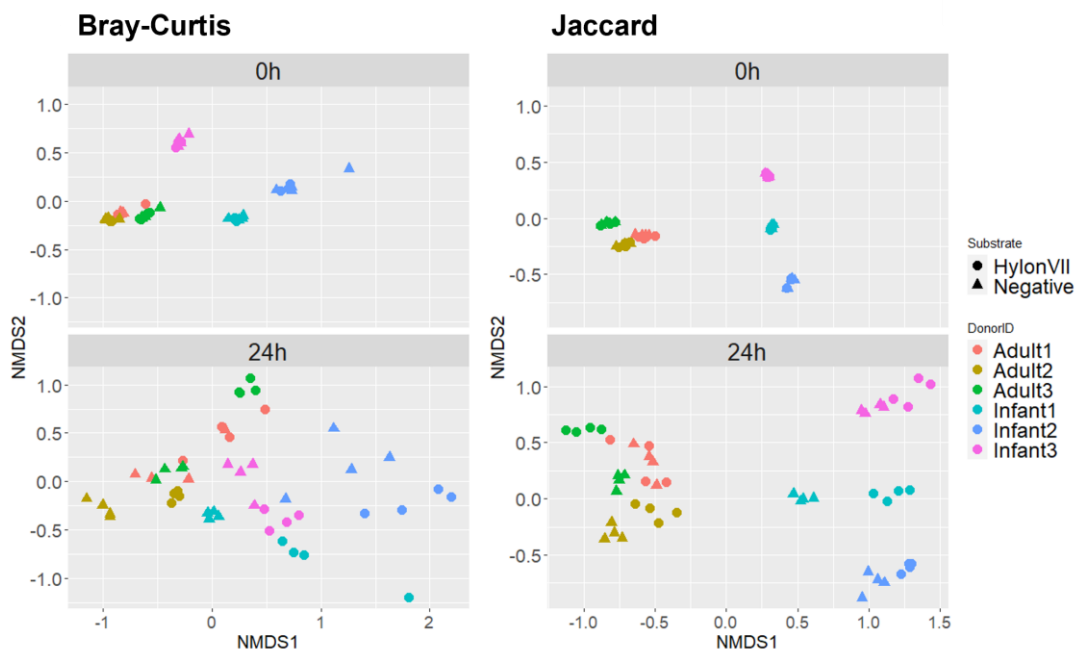
More abundant at 24h in adult vs infant (all substrates)
Eubacterium limosum
Anaerobutyricum hallii
Longibaculum sp. KGMB06250
Bacteroides fragilis
Faecalibacillus intestinalis
More abundant at 24h in infant vs adult (all substrates)
Enterococcus wangshanyuanii
Lactococcus lactis
More abundant at 24h-HylonVII in adult vs infant
Anaerobutyricum hallii
Eubacterium limosum
More abundant at 24h-HylonVII in infant vs adult
None

Figure 5-5 [Top] Bray-Curtis and Jaccard Beta diversity plots displaying the (dis)similarity of samples using multidimensional plots. [Bottom] The presence or absence of taxa was demonstrated statistically using Maaslin2 showing the statistically higher abundance of specific species comparing infant and adult.

5.4.4 Metabolomic response to HylonVII

In the presence of HylonVII, total SCFA production values for Adult and Infant donors were not significantly different (Figure 5-6A). However, the ratios of SCFAs produced differed in the presence of HylonVII which suggests that the microbiomes of the two age groups metabolised HylonVII differently (Figure 5-6B).

Infant total butyrate production was significantly higher than adult in the presence

of HylonVII, whilst Adult had significantly higher propionate and succinate

(Figure5-6C).
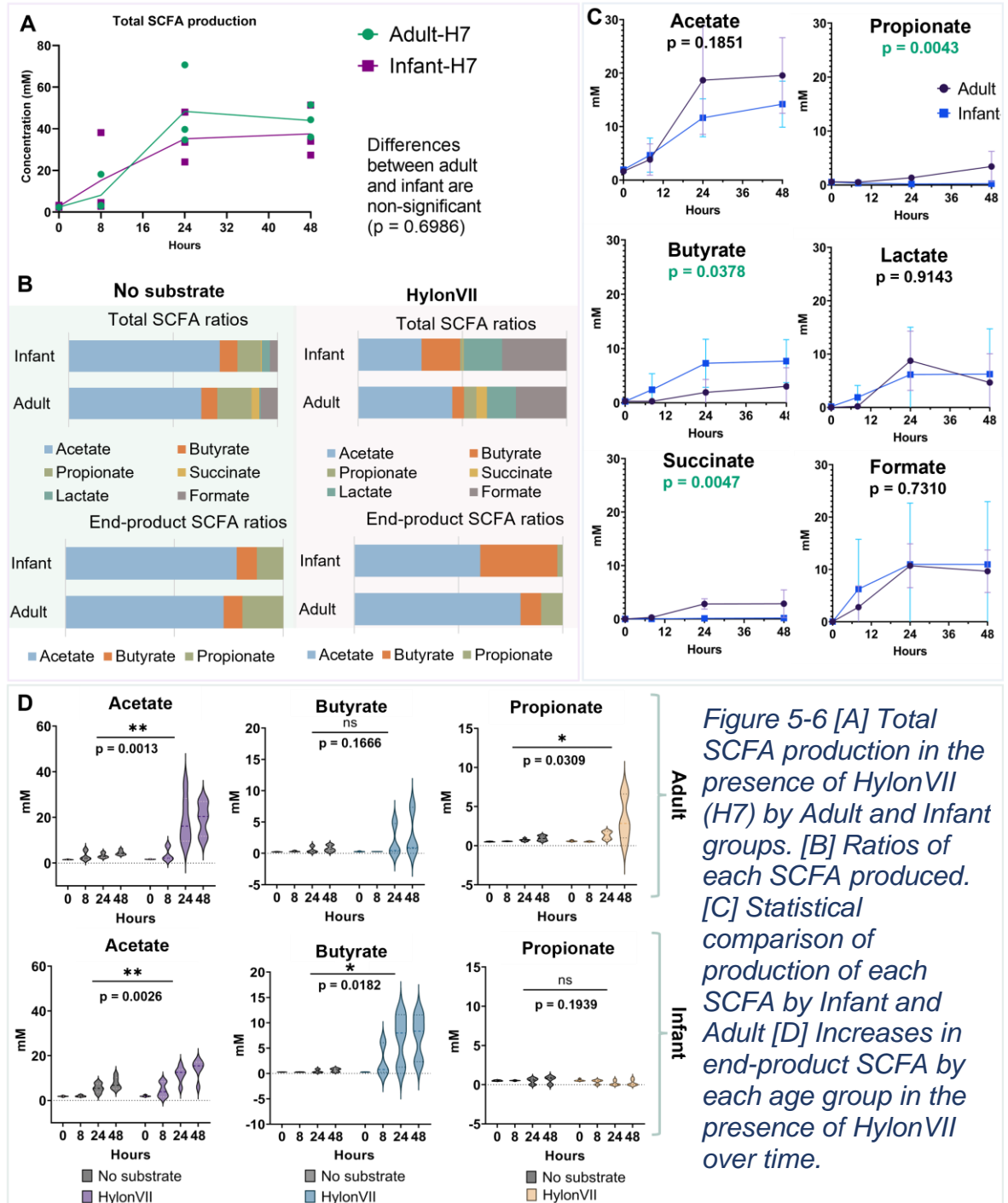
Figure 5-6 [A] Total SCFA production in the presence of HylonVII (H7) by Adult and Infant groups. [B] Ratios of each SCFA produced. [C] Statistical comparison of production of each SCFA by Infant and Adult [D] Increases in end-product SCFA by each age group in the presence of HylonVII over time.

Acetate, lactate, and formate production were not significantly different between age groups (Figure 5-6C). Confirming these findings, there was significantly higher acetate production in the presence of HylonVII compared to No substrate for both age groups (p = 0.0026 for Infant, p = 0.0013 for Adult); only the Infant group had significantly higher butyrate production in the presence of HylonVII (p = 0.0182); and adults had significantly higher propionate production (p= 0.0309) (Figure5-6D).

5.4.5 Infant microbiome response to HylonVII

To assess the specific taxa driving the differences observed between experimental variables, the relative abundance of genera and species of each donor was plotted to compare the changes in taxa between 0 and 24 hours in the presence or absence of HylonVII (Figure 5-6, Figure 5-7A+B). There were no significant differences in bacterial taxa at 0h for all Infant samples pooled together between no substrate and HylonVII (Figure 5-7A), which was confirmed by statistical testing (Table 5-1). At 0h, there are observed similarities between individual infant donors as all three had a relative abundance of *Bifidobacterium* at >70% of taxa (Figure 5-7A). Infant 1 had >5% of other taxa such as *Ruminococcus, Lacticaseibacillus,* and *Escherichia* (Figure 5-7A). *Blautia* SC05B48 was particularly present in Infant 2 compared to the other donors, as well as <5% relative abundance of *Enterococcus* and *Clostridium*. Infant 3 had a relative abundance of ~20% *Escherichia* genus taxa present at 0h, which was presumably *E. coli,* as well as ~5% *Lacticaseibaccillus* (Figure 5-7B). All donors had <3% *Erysipelatoclostridium* (now called *Thomasclavelia (*spp. *ramosa)*, (Lawson, Saavedra Perez et al. 2023) (Figure 5-7A) and very low relative abundance of

C. perfringens at 0h (Figure 5-7B). *Faecalibacterium prausnitzii* was found to be

present in all three infant donors.

Figure 5-7 [A] Phylogeny plot of changes in microbiome composition (genus level) for each donor, presence or absence of the substrate HylonVII, over time. [B] Species level heatmap displaying changes occurring across the same variables. [C] Prevalence value reduced to minimum (0.0) to show the changes in specific species only some of which are prevalent in fewer than 20% of samples.

After 24h, the response of Infant 1 microbiome to HylonVII was defined by a *Clostridium* bloom which accounted for almost the entire microbiome in the presence of HylonVII. Whilst in the No substrate control, it contained ~30% *Clostridium*, 50% *Bifidobacterium* and 5% *Erysipelatoclostridium* (Figure 5-7B). This suggests that is it not only starch driving increases in *Clostridium*-like bacteria. These results confirm the affinity of *Clostridium*-like bacteria to multiply in the presence of starch as well as their proclivity for rapid proliferation (Wang, Conway et al. 1999, Kiu and Hall 2018). *C. butyricum* and *C. perfringens* were statistically higher at 24h in HylonVII Infant samples (all three donors pooled for one statistical test) compared to No substrate (Figure 5-8).

To examine the inter-donor responses to the presence of HylonVII, the prevalence filter was reduced to 0.0 to assess the presence of less common taxa which increased in relative abundance (Figure 5-8C). Infant 1 was found to be dominated by *C. perfringens* whilst Infant 2 had significant increases of *C. butyricum* (Figure 5-7C). Clinical metadata reports that Infant 1's mother noted at age of 4 months (donation used for this thesis was at 8 months of age) the infant's stool over the previous 7 days was "semi-watery and soft" whereas on all other timepoints she reported the stools to be "formed". This could be indicative of a mild infection or microbial perturbance in the months prior to donation.

Infant 2 and 3 had similar responses to HylonVII. Even though Infant 2 had higher levels of *C. butyricum* relative to the control where *Clostridioides difficile* was more abundant (Figure 5-7C), the dominant genera were *Enterococcus, Blautia,* and *Bifidobacterium* (Figure 5-7A). Infant 3 had some increases in relative abundance

of *Erysipelatoclostridium, Paeniclostridium,* and *Clostridium* in HylonVII (accounting for around 30% of relative abundance) reproducing that starch caused increases in *Clostridium* and related taxa. Infant 2 and 3 both had increases in *Enterococcus faecium* in the presence of HylonVII (Figure 5-7C). Comparing HylonVII to the No substrate control, *Enterococcus faecium* could be competing with *Bifidobacterium breve* and *B. longum* for the substrate.

The relative abundance of *Clostridium* and related taxa at 0h was low across all three infant donors (Figure 5-7C). The increases in abundance in all three infant donors are observed as partially a function of time (regardless of substrate) and the presence of HylonVII (Figure 5-4, Figure 5-7C). The resistance of Infant 2 to blooms of *Clostridium* could be related to its later donation age (12 months versus 8 months) which could imply this infant has a more mature, resilient microbiome.

There were statistically significant increases of *Ligilactobacillus salivarius* and *Lactobacillus crispatus* detected when all 3 Infant data were pooled (Figure 5-8). These related species require simple sugars (including maltose) for growth, implying the presence of HylonVII consistently promoted growth of these secondary cross-feeders of starch across all 3 Infant donor microbiomes *in vitro* (Li, Loponen et al. 2020). Whilst found to be positively correlated to acetate production in this Infant dataset, the taxa were present in low relative abundance due to having low read counts (Figure 5-9).
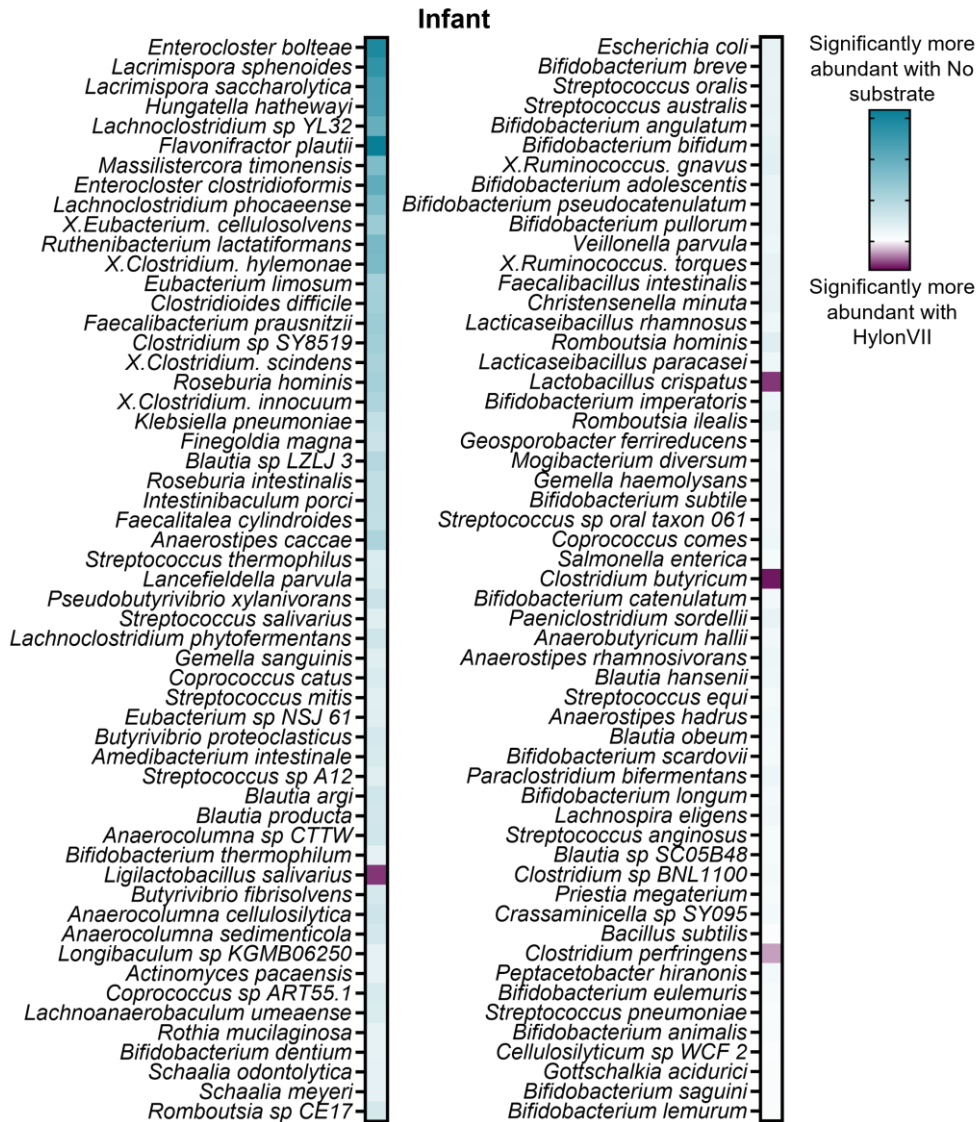
Figure 5-8 Specific taxa which were significantly more abundant in the No substrate or the HylonVII condition in all Infant samples (q<0.05). Colour saturation represents the coefficient value (magnitude of the differential response) multiplied by the negative log of q value. Statistics performed using Maaslin2 R package.

Comparing the relative abundance of only *Bifidobacterium* taxa, at 0h there are similar relative abundances of *Bifidobacterium* between the HylonVII and No substrate in all donors which indicates sequencing was reproducible (Figure 5-10). After 24h, the relative abundance of *Bifidobacterium* declined drastically in Infant 1, especially in the HylonVII group. This is likely due to the *Clostridium*

bloom which may have outcompeted all other taxa for utilisation of the substrate. Infant 2 maintained its relative abundance of *Bifidobacterium* after 24h in both HylonVII and no substrate, Infant 3 had slight decline in the *Bifidobacterium* group, also likely due to competition with other taxa such as *Clostridium* and *Paeniclostridium*.

The most prevalent species after 24h with HylonVII were *B. breve, B. longum, B. catenulatum, B. bifidum*, and *B. adolescentis,* some of which contain known amylolytic strains (Ryan, Fitzgerald et al. 2006). *B. catenulatum* was mainly present in Infant 2, and *B. adolescentis* in Infant 3. The presence of HylonVII had a potentially negative effect on relative abundance of *Bifidobacterium* taxa (especially on a species level) due to competition with other more taxa such as *Clostridium* and relatives or *Enterococcus faecium.*
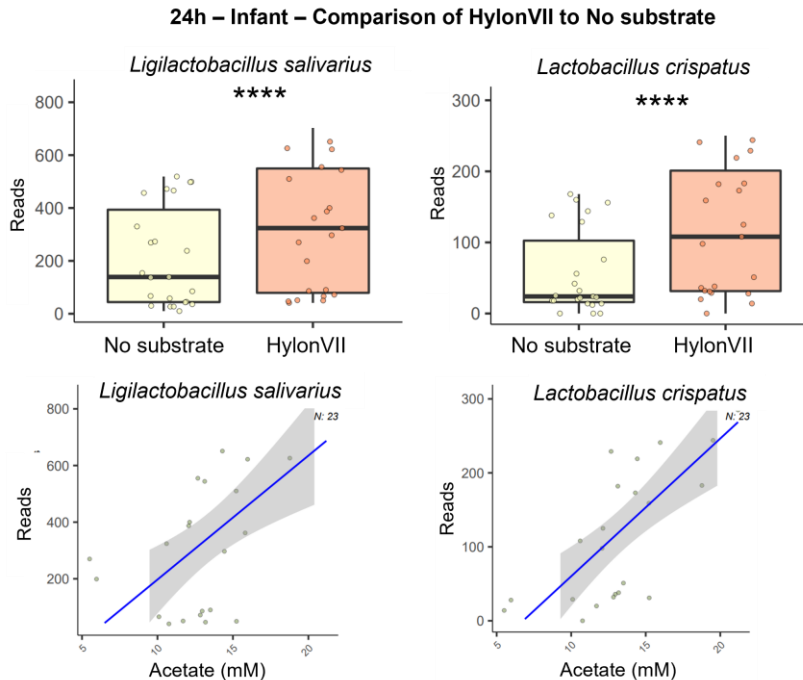
Figure 5-9 Two bacterial species found to be significantly higher in abundance in all Infant samples in the HylonVII condition were found to be positively correlated with acetate production.

Figure 5-10 Plots showing the changes in relative abundance of Bifidobacterium *taxa [A] overall and [B] specific species in Infant donors.*
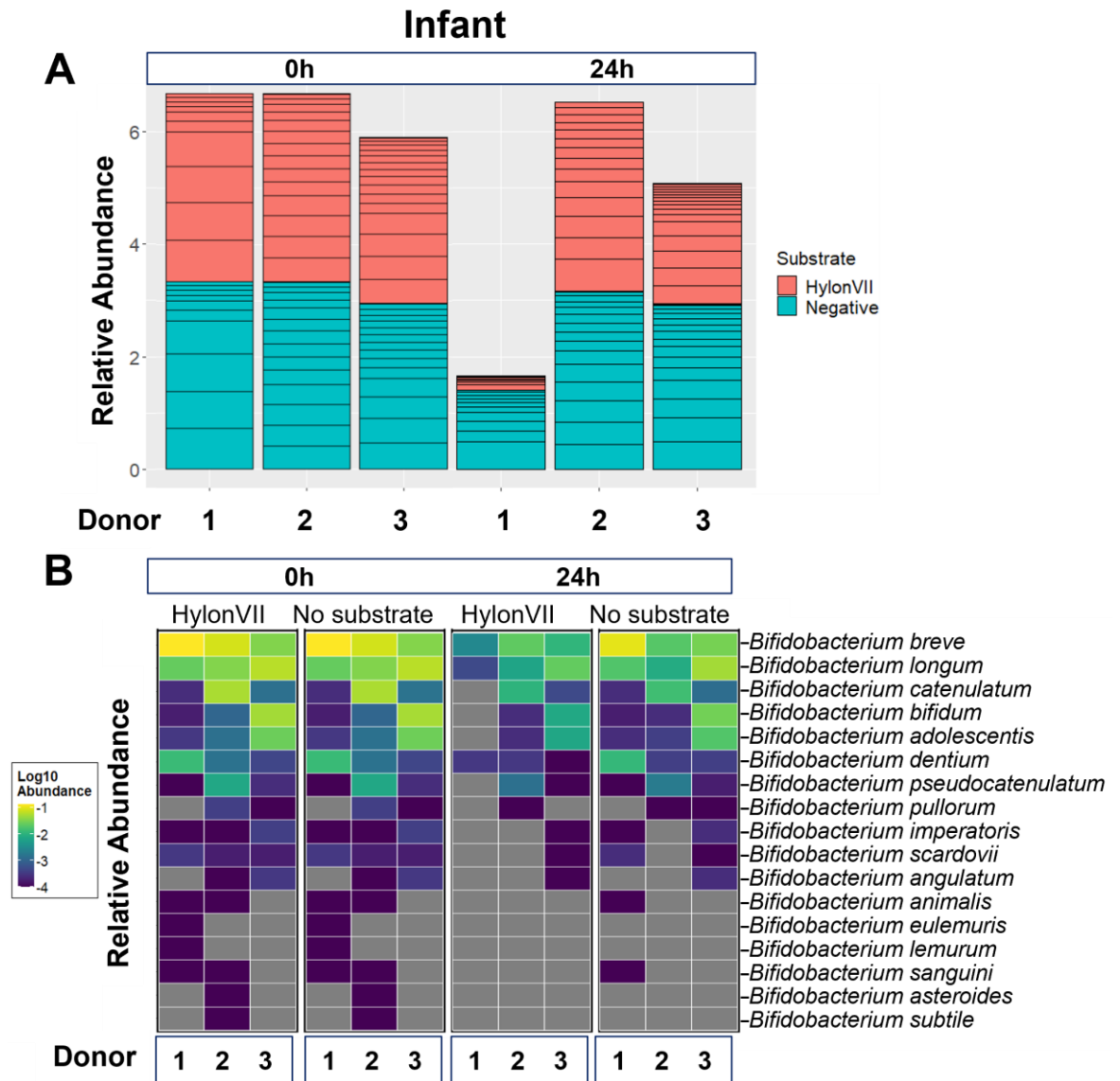
5.4.6 Adult microbiome response to HylonVII

Adult relative abundances were not statistically different between HylonVII and

No substrate at 0h (Table 5-1). The baseline adult microbiome of Adult 1 and 2

was characterised by beneficial and common dominant taxa from the Bacillota

(formerly Firmicutes) phylum, *Blautia*, *Roseburia*, Actinomycetota/Actinobacteria

members *Collinsella* and *Bifidobacterium*, *Faecalibacterium*, *Coprococcus*,

Anaerobutyricum, and *Ruminococcus* (Figure 5-11A). Adult 2 had high relative

abundance of *Faecalibacterium praunitzii* specifically (Figure 5-11B). Adult 3 was Bacteroidota (*Bacteroides* and *Phocaeicola*) dominant, lacking *Colinsella* and *Bifidobacterium* but still comprising moderate relative abundance (5-10%) of Baccilota phylum members *Blautia*, *Anaerobutyricum*, *Faecalibacterium*, *Coprococcus,* and *Roseburia*. *Bacteroides* species present at 0h in Adult 3 were *Bacteroides ovatus*, *B. uniformis, B. intestinalis*, and *Phoaeicola (Bacteroides) dorei* (Figure 5-11B).

After 24h, in Adult 1 and 3 there were increases in *Escherichia* in both No substrate and HylonVII (Figure 5-11A). There were significant increases in *E. coli* in the Adult group between 0 and 24h regardless of substrate which suggests that characteristics of the *in vitro* fermentation system encourage *E. coli* blooms (Figure 5-4). *E. coli* was less abundant in the presence of HylonVII suggesting that the bacteria which metabolise it may have suppressed *E. coli* in Adult 1 and 3 (Figure 5-11A).

Aside from *E. coli* blooms, the Adult 1 response to HylonVII compared to the no substrate control demonstrated higher relative abundances of *Phocaeicola dorei*, *Bacteroides uniformis*, *Bifidobacterium longum* and *B. adolescentis*, and *Collinsella aerofaciens* compared to the no substrate. In both HylonVII and No susbtrate at 24h, Adult 1 had decreases of *Roseburia*, *Ruminococcus*, *Anaerostipes*, *Coprococcus*, and *Blautia* (Figure 5-11A). The microbiome of Adult 1 in the presence of HylonVII after 24h was *Bacteroidota* and *Pseudomonadota* phyla dominant.
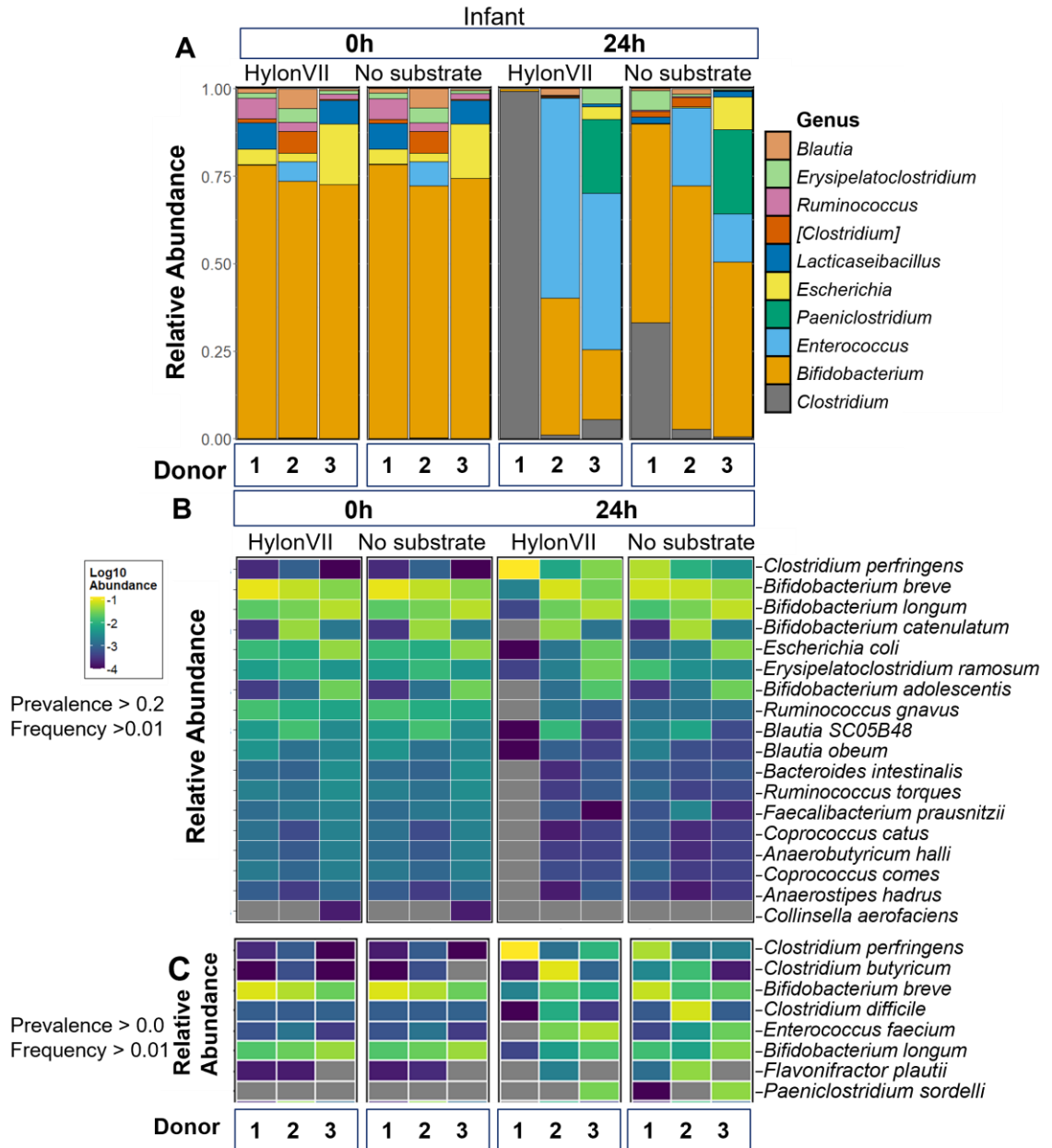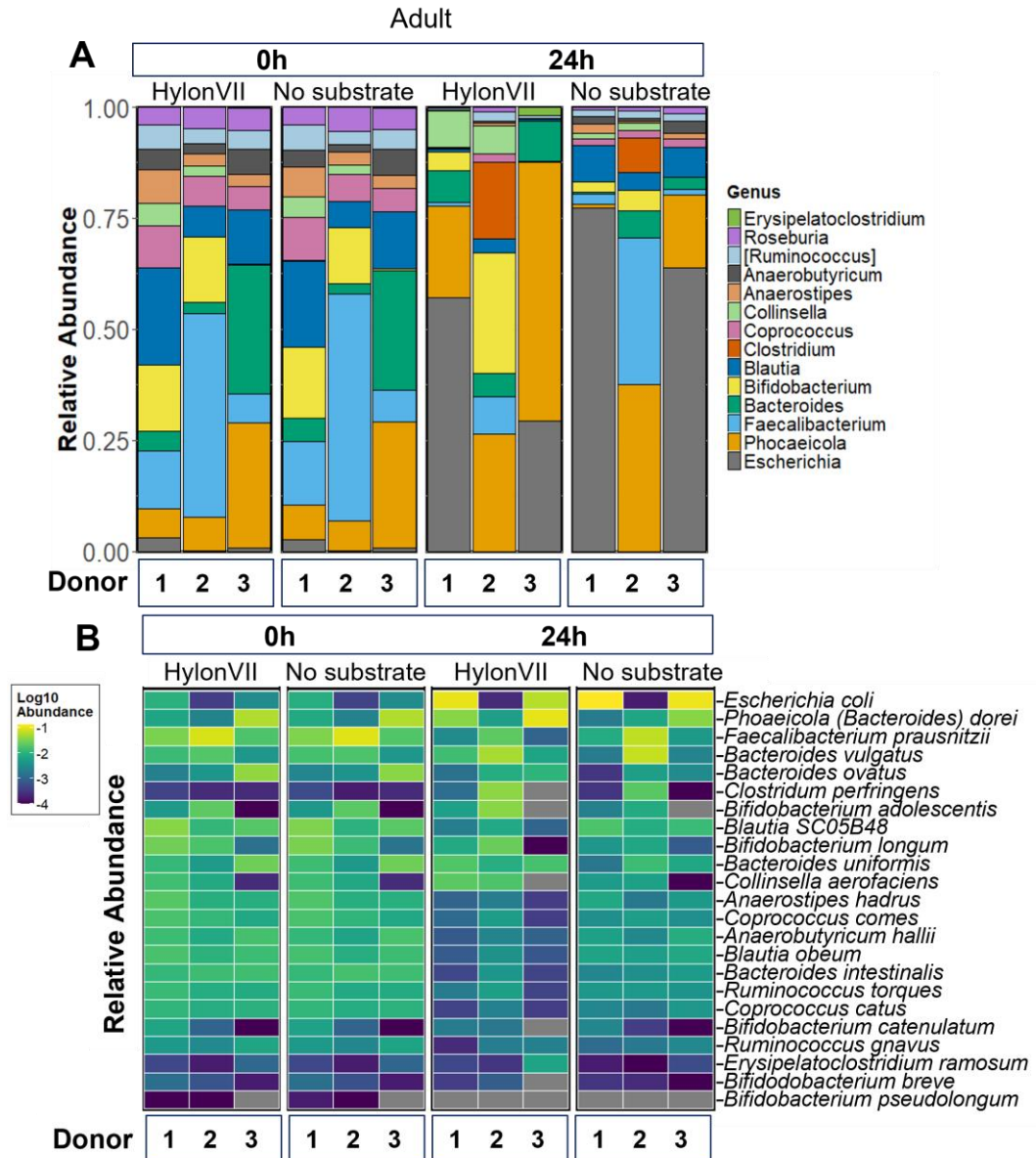
Figure 5-11 [A] Phylogeny plot of changes in Adult microbiome composition (genus level) for each donor, presence or absence of the substrate HylonVII, over time. [B] Species level heatmap displaying changes occurring across the same variables.

In contrast, Adult 2 response to HylonVII after 24h was defined by increases in

starch degrading Bacillota members *Clostridium, Phocaeicola,* and

Actinobacteria *Collinsella* and *Bifidobacterium* (Figure 5-11A). In the absence of

a substrate, *Faecalibacterium*, *Phocaeicola,* and *Clostridium* were dominant in

Adult 2. Thus, HylonVII encouraged *Clostridium*, *Bifidobacterium,* and *Collinsella* growth.

Adult 3 response to HylonVII compared to the No substrate control was characterised by increases in *Phocaeicola*, *Bacteroides* and *Erysipelaclostridium* shifting the microbiome to a Bacteroidota and Pseudomonadota dominant 'enterotype' (Figure 5-11A). Whilst defining gut microbiota 'enterotypes' remains a point of contention as the microbiome is both rapidly modifiable and is more broadly accepted as existing as a gradient rather than categorical, it is worth noting Adult 3's microbiome composition differed from the other two Adult donors (Jeffery, Claesson et al. 2012, David, Maurice et al. 2014, Costea, Hildebrand et al. 2018).

Statistical testing of Adult samples at 24h confirmed a significant increase in taxa such as *Bifidobacterium animalis* and *E. ramosum* in HylonVII (Figure 5-12A). There were significant increases of *Prevotella* spp. (6 total) even though they were not relatively highly abundant (Figure 5-12B). *Phocaeicola* spp. also had significantly higher abundance and were positively correlated with acetate production.
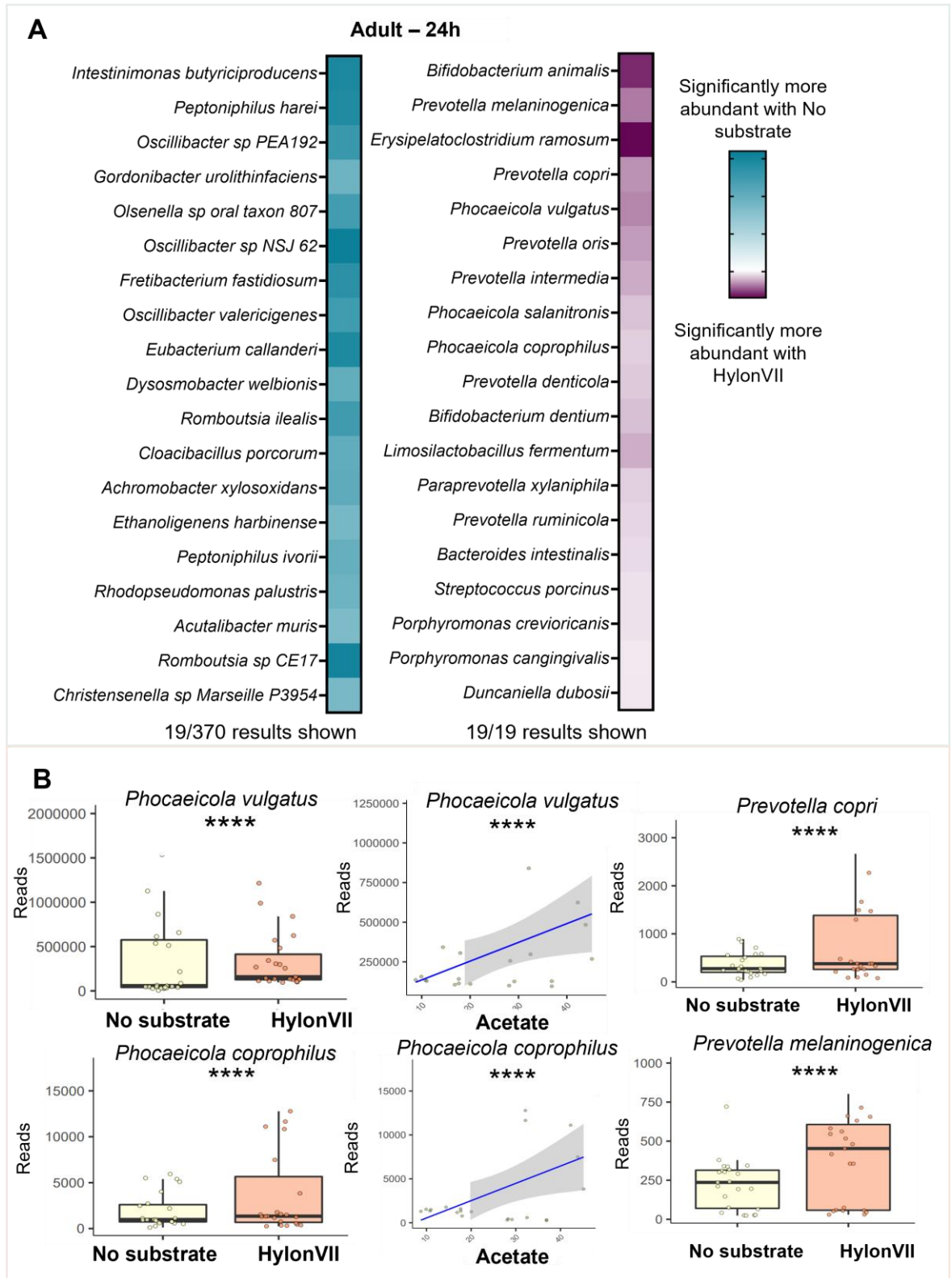
Figure 5-12 [A] Adult samples significantly more abundant taxa in the presence of absence of HylonVII [B] Prevotella *spp. had significantly higher abundance in HylonVII whilst* Phoecaeicola *spp. were more highly abundant in HylonVII and were positively correlated to acetate production.*

Adult 3 had the lowest relative abundance of *Bifidobacterium* for both time points and substrate conditions (Figure 5-13A). In the adults, *Bifidobacterium* declined to very low relative abundance in the no substrate control, while in the HylonVII group Infant 1 and 2 maintained or increased (respectively) relative abundance of *Bifidobacterium*, which suggests HylonVII promoted the relative population abundances of *Bifidobacterium* only where (starch-degrading) *Bifidobacterium* were present initially.
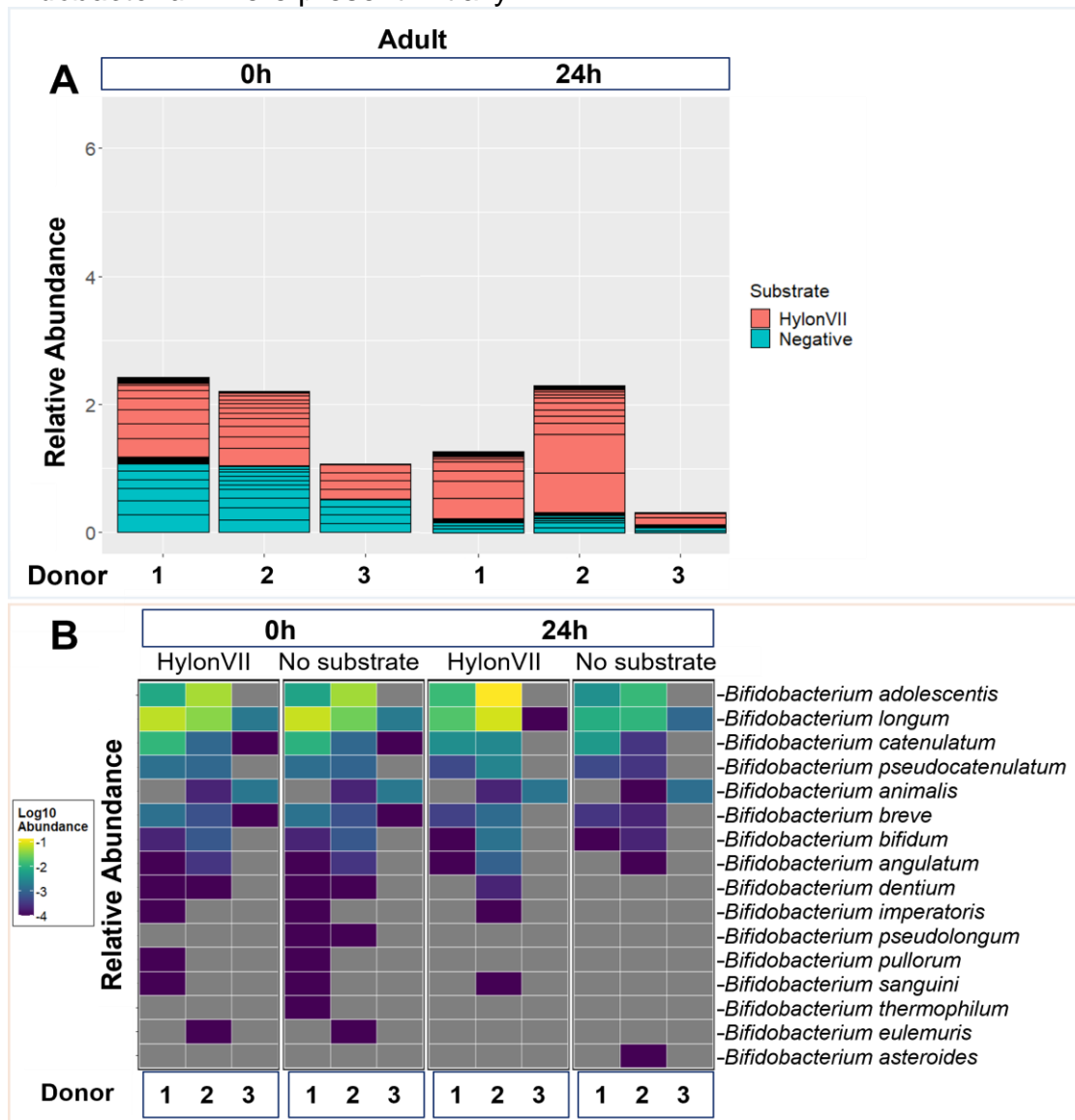
Figure 5-13 Plots showing the changes in relative abundance of Bifidobacterium *taxa [A] overall and [B] specific species in Adult donors.*

The highest increases in *Bifidobacterium* in the presence of HylonVII were in Adult 2 (Figure 5-13A) in *B. adolescentis* and *B. longum* which generally have starch-degrading capabilities (Dobranowski and Stintzi 2021) (Figure 5-13B*)*. Pooling the donors for statistical analysis resulted in a non-significant increase of *B. adolescentis* in the presence of HylonVII in adult donors (q = 0.09669). Thus, the increase of specific taxa which can degrade specific substrates is dependent on the original microbiota composition of individual donors.

5.4.7 Effect of bifidobacterial supplement on *in vitro* HylonVII fermentation

The datasets were divided into each sample type in the presence of HylonVII and plotted using NMDS to display Jaccard diversity (Figure 5-14). Generally, at 24h there are no observed trends or differences in beta diversity between supplement groups.
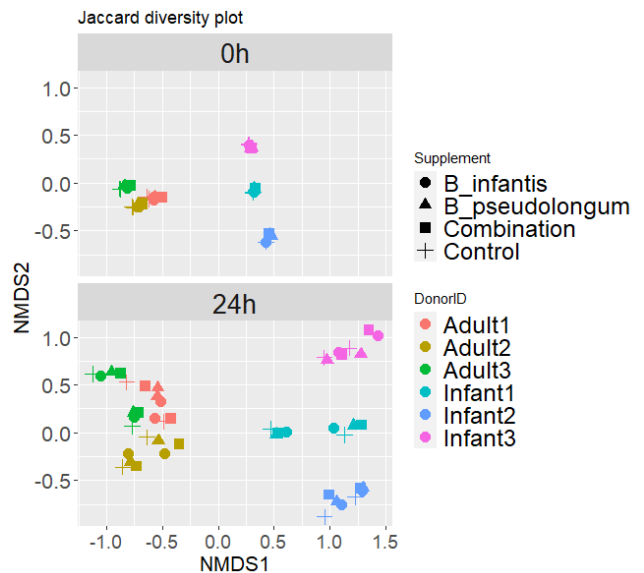
Figure 5-14 Multidimensional plot displaying supplement group beta diversity.

5.4.7.1 Infant microbiome response to bifidobacterial supplementation

Infant 1 was dominated by *Clostridium* in the negative control (no supplement) but maintained some levels of *Bifidobacterium* in the *B. pseudolongum* 45, Combination and *B. longum* subsp. *infantis* 277 groups (Figure 5-15A). The highest levels of *Bifidobacterium* maintained after 24h was in the *B. pseudolongum* 45 group and relative abundance of *B. pseudolongum* itself was higher than other *Bifidobacterium* species detected (Figure 5-15B). Infant 2 and

3 had similar responses to HylonVII after 24h: dominant genera were

Enterococcus and *Bifidobacterium* with low levels of *Clostridium*, and Infant 3

additional had blooms of *Paeniclostridium,* and *Erysipelatoclostridium* across all

supplement groups including the negative control.
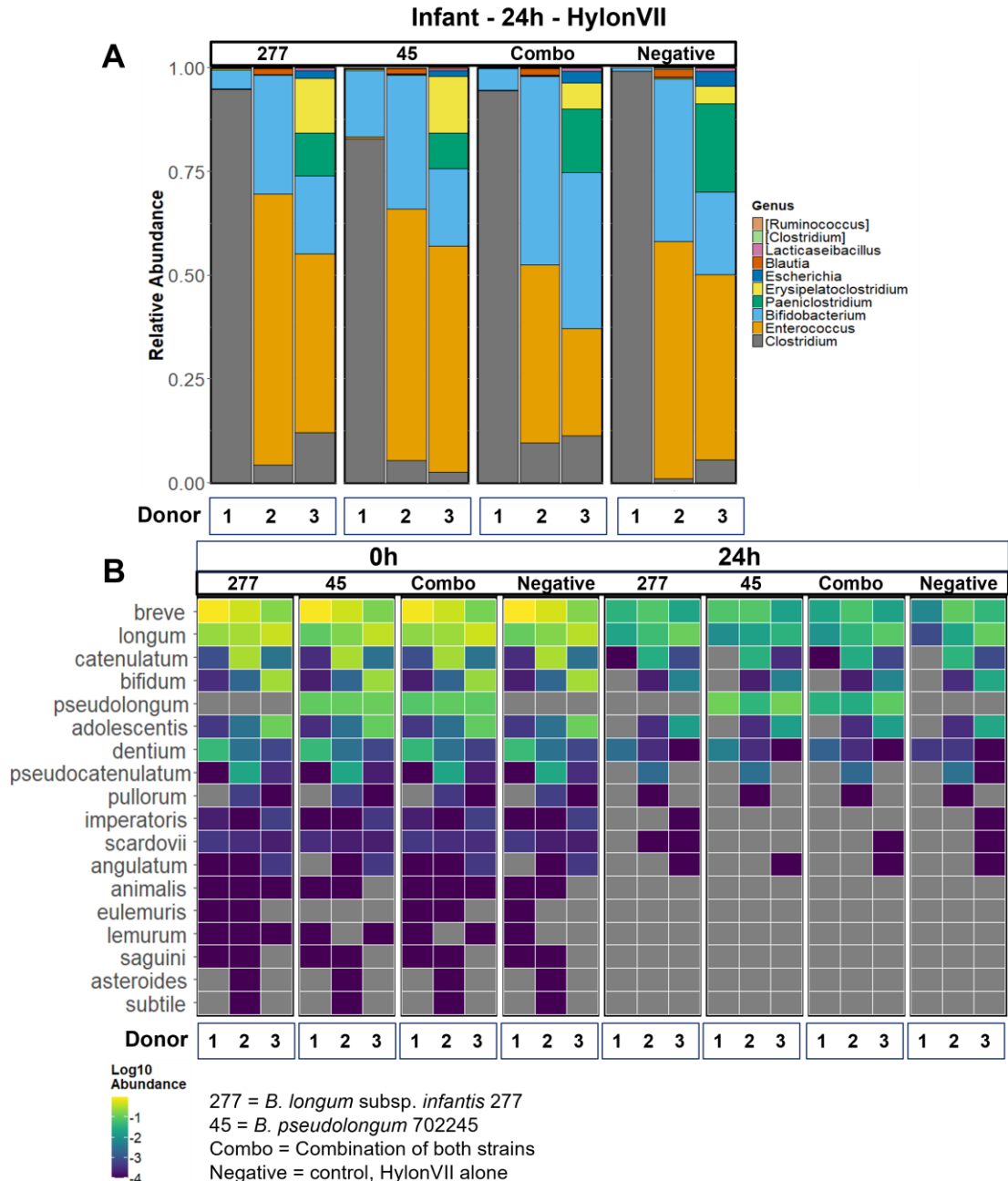
Figure 5-15 [A] Phylogeny plot of bacterial genera in the supplemented Infant samples compared to a negative control in the presence of HylonVII and [B] the relative abundance of Bifidobacterium *species.*

There were no apparent differences in response when comparing *B. longum* subsp. *infantis* 277 and *B. pseudolongum* 45 supplementations in Infant 2 and 3 (Figure 5-15). The presence of *B. pseudolongum* 45 at both 0h and 24h timepoints confirmed that (a) the supplemented strain was undetectable in all three donors when not supplemented and (b) persisted in the model over the course of 24h (Figure 5-15B).

There were 175 taxa significantly higher in abundance with a *Bifidobacterium* supplement compared to without after 24h (Figure 5-16). *B. pseudolongum* 45 supplement had a minimal but significant effect on *B. breve*, whilst the main response was to *B. longum* subsp. *infantis* 277 (alone or in Combination) (Figure 5-16). The responsive taxa were *Bifidobacterium*, *Streptococcus, Roseburia, Eurbaceterium, Lactiplantibacillus plantarum,* and *Enterococcus* spp. and butyrate producing *F. prausnitzii.*

Just as the taxonomic shifts due to HylonVII were variable depending on donor, the metabolic responses of individual donor microbiota to the *in vitro* fermentation model were variable. In the presence of HylonVII, the concentrations of acetate increased in all infant donors to peak levels of around 20mM by the 48 h timepoint (Figure 5-17). Butyrate levels were the highest in Infant 1 and 2 likely due to prevalence of *C. perfringens* and *C. butyricum* respectively. Propionate levels remained unchanged (Figure 5-17). In Infant donors, there appeared to be no effect of supplement on acetate, butyrate, or propionate concentrations in the presence of HylonVII (Figure 5-17), which was confirmed statistically for butyrate when all Infant donors were pooled (p = 0.9877). To conclude, whilst

there were several key taxa which increased in abundance in response to the
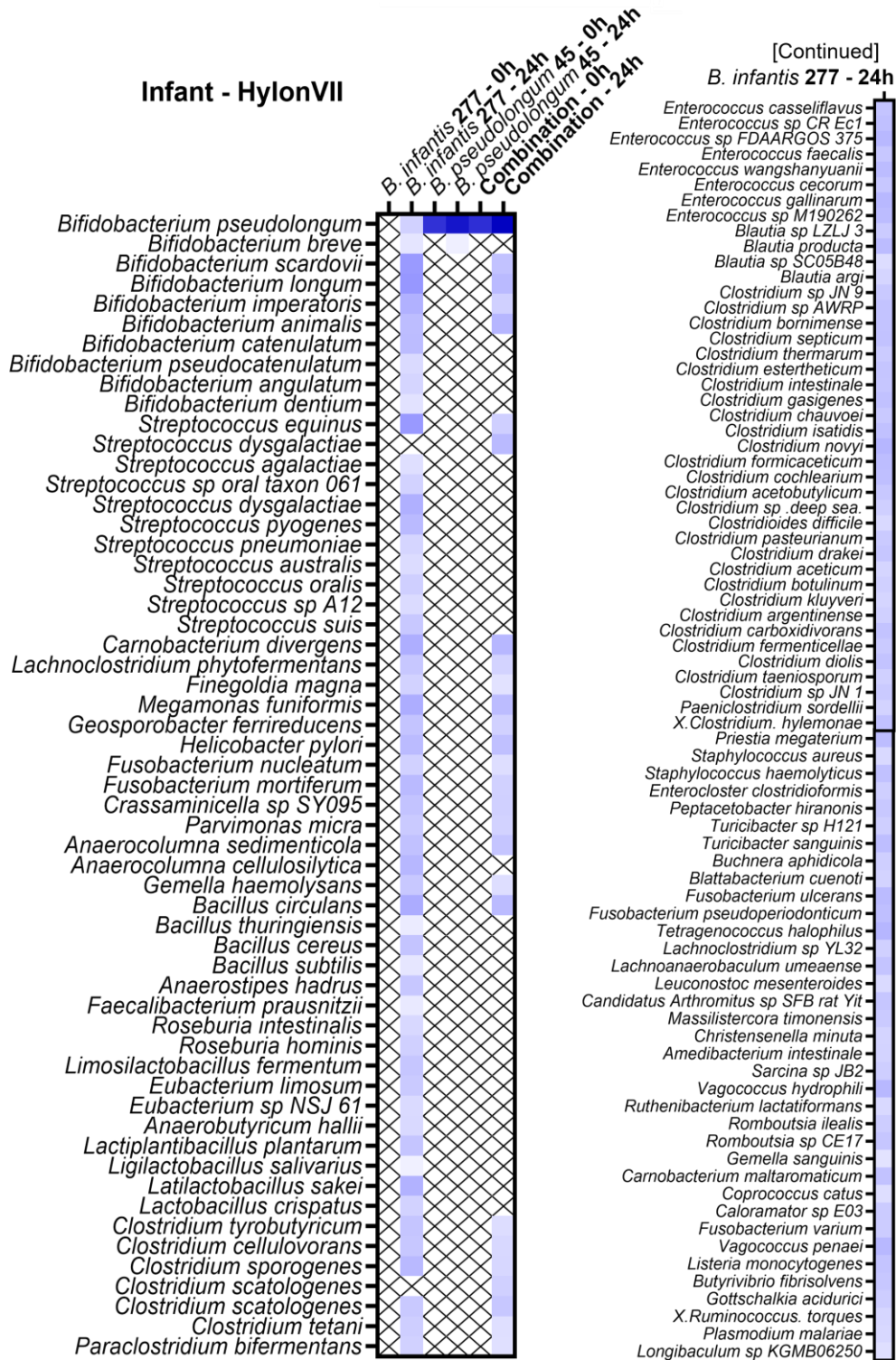
Figure 5-16 Species with significantly higher (q<0.05) relative abundance in the presence of bacterial supplement compared to no supplement, performed using Maaslin2 R package. Colour saturation demonstrates the coefficient multiplied by the negative log of the q value.

presence of a *B. longum* subsp. *infantis* 277 supplement, there were few

discernible positive effects of the treatment on SCFA production.

SCFA production with HylonVII
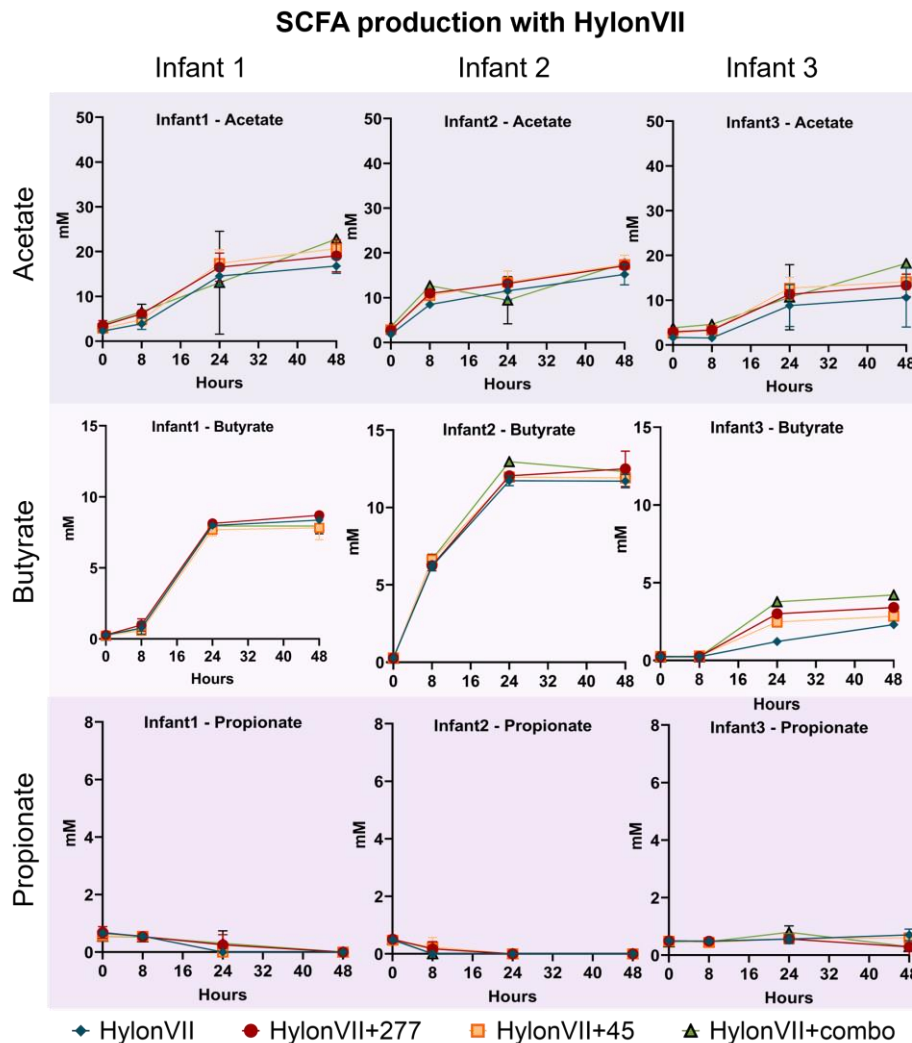
Figure 5-17 Acetate, butyrate, and propionate production over the course of 48h by each Infant donor stratified by bacterial supplement in the presence of HylonVII.

5.4.7.2 Adult response to bifidobacterial supplementation

As described previously, the negative control (HylonVII, no supplement) resulted in

Adult 1 and 3 having Bacteroidota-dominant enterotypes, whilst Adult 2 had

dominant Bacillota phylum taxa present (Figure 5-18). Adult 1 response to

B. longum subsp. *infantis* 277 was defined by higher levels of *Bifidobacterium,*

Collinsella, *Faecalibacterium* and lower levels of *Phocaeicola*, *Escherichia*, and

Bacteroides, essentially shifting the microbiome to a Bacillota dominant enterotype (Figure 5-18A). This suggests that activities related to *B. longum* subsp. *infantis* 277 exert positive effects on Bacillota members in the presence of starch. Species
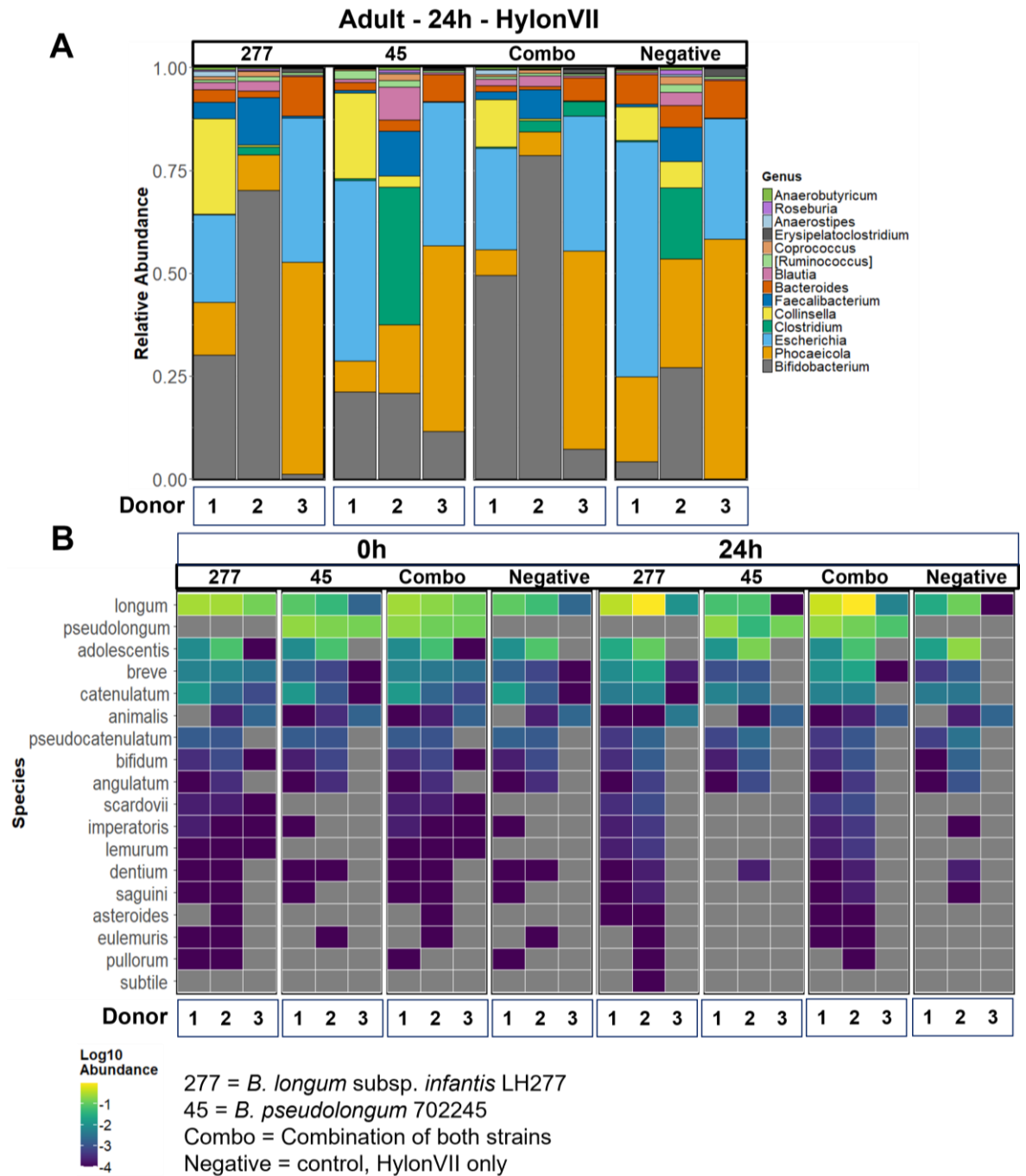
Figure 5-18 [A] Phylogeny plot of bacterial genera in the supplemented Adult samples compared to a negative control in the presence of HylonVII and [B] the relative abundance of Bifidobacterium *species.*

more abundant in the absence of *B. longum* subsp. *infantis* 277 were *E. coli*,

Phocaeicola dorei, Bacteroidoes uniformis (Figure 5-19).
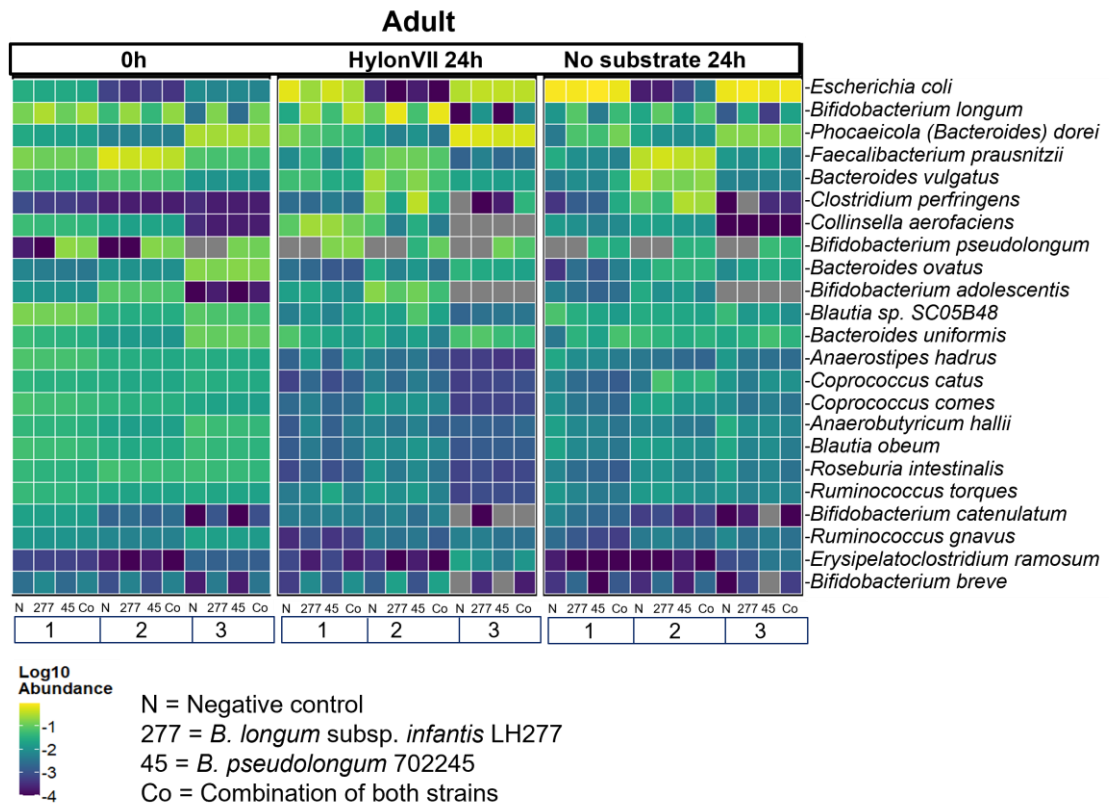
Figure 5-19 Species relative abundance heatmap showing the effects of Bifidobacterium *supplementation on Adult samples.*

The response of Adult 1 to *B. pseudolongum* 45 supplementation was similar in

terms of increases in *Bifidobacterium*, *Collinsella*, and reduction in *Phoaeicola*,

however there was higher abundance of *Escherichia* compared to *B. longum*

subsp. *infantis* 277 group (Figure 5-18A). The combination of the two strains

appeared to confound the effect of administration of each strain individually as the

dominant taxon was *Bifidobacterium* (50% relative abundance). Thus, the

supplementation of bifidobacterial strains individually and in combination increased

beneficial taxa such as *Bifidobacterium*, *Collinsella* and *Faecalibacterium*.

Adult 2 response to *B. longum* subsp. *infantis* 277 and Combination of strains was characterised by significant increases in *Bifidobacterium* (Figure 5-18A) which was designated as *B. longum* (Figure 5-18B). The relative abundance of *B. longum* increased between 0 and 24h only in HylonVII (Figure 5-19), suggesting the populations of this species thrived in the presence of starch in this microbiome. This implies that the levels of *B. longum* were dramatically impacted by supplementing the model with HylonVII and *B. longum* subsp. *infantis* 277 in tandem in spite of it not itself utilising HylonVII in isolated cultures. It has not yet been confirmed if the increases of *B. longum* are attributable to the subspecies/strain supplemented (subsp. *infantis* 277).

In Adult 2, taxa which decreased in the presence of *B. longum* subsp. *infantis 277* (including the Combination) were *Clostridium, Bacteroides, Phoaeicola, Collinsella* (Figure 5-18A). In contrast, in the *B. pseudolongum* 45 supplement group, Adult 2 had a decrease of *Bifidobacterium* and *Collinsella*, *Bacteroides*, and *Phocaelicola*, and an increase in butyrogenic taxa *Clostridium*, *Blautia*, and *Faecalibacterium*. The species which increased were *C. perfringens*, *Blautia sp. SC05B48* and *F. prausnitzii* (Figure 5-19). Therefore, the bacterial strains elicited two distinct changes to the microbiome composition of Adult 2 in the presence of HylonVII. Adult 3 had a minimal response to bifidobacterial supplementation, however there was an increase in relative abundance of *Bifidobacterium* in all supplement groups, especially the *B. pseudolongum* 45 group (Figure 5-18).

Statistical testing of the adult microbiome response confirmed significant increases in 20 taxa in the *B. longum* subsp. *infantis 277* or the Combination but not the *B.*

pseudolongum 45 group, suggesting that any taxonomic changes were solely due to *B. longum* subsp. *infantis* (Figure 5-20). Statistical correlations between SCFA concentrations and the taxa abundance values showed that several acetate producing species were upregulated in the *B. longum* subsp. *infantis* 277/Combination groups relative to the no supplement group: *Bifidobacterium breve, Bifidobacterium lemurum, Bifidobacterium longum,* and *Eggerthella* spp. *YY7918* (Figure 5-20). Butyrate producing *F. prausnitzii, Adlercreutzia equolifaciens,* and *Schaalia odontolytica* were also upregulated. These results suggest that *B. longum* subsp. *infantis* 277 caused microbiome shifts which favour acetate and butyrate producing bacteria and that this effect is statistically robust even across 3 adults with different microbiome structure (before and after HylonVII).

Taxa significantly more abundant in the absence of a supplement were *Bacteroides* spp., *E. ramosum, Ruthenibacterium lactatiformans* which were found in this dataset to be negatively correlated with acetate production (Figure 5-20). It is possible the combination of the two *Bifidobacterium* strains specifically work synergistically to decrease levels of these bacteria possibly in a competitive process for the substrate either directly in the downstream utilisation of hydrolysates or metabolic products. The results suggest that that while *B. pseudolongum* 45 persists in the *in vitro* model it is not necessarily impacting microbial composition by promoting the increase of taxa, but it potentially causes a decrease of *Bacteroides* spp. when combined with *B. longum* subsp. *infantis* 277 (Figure 5-21).
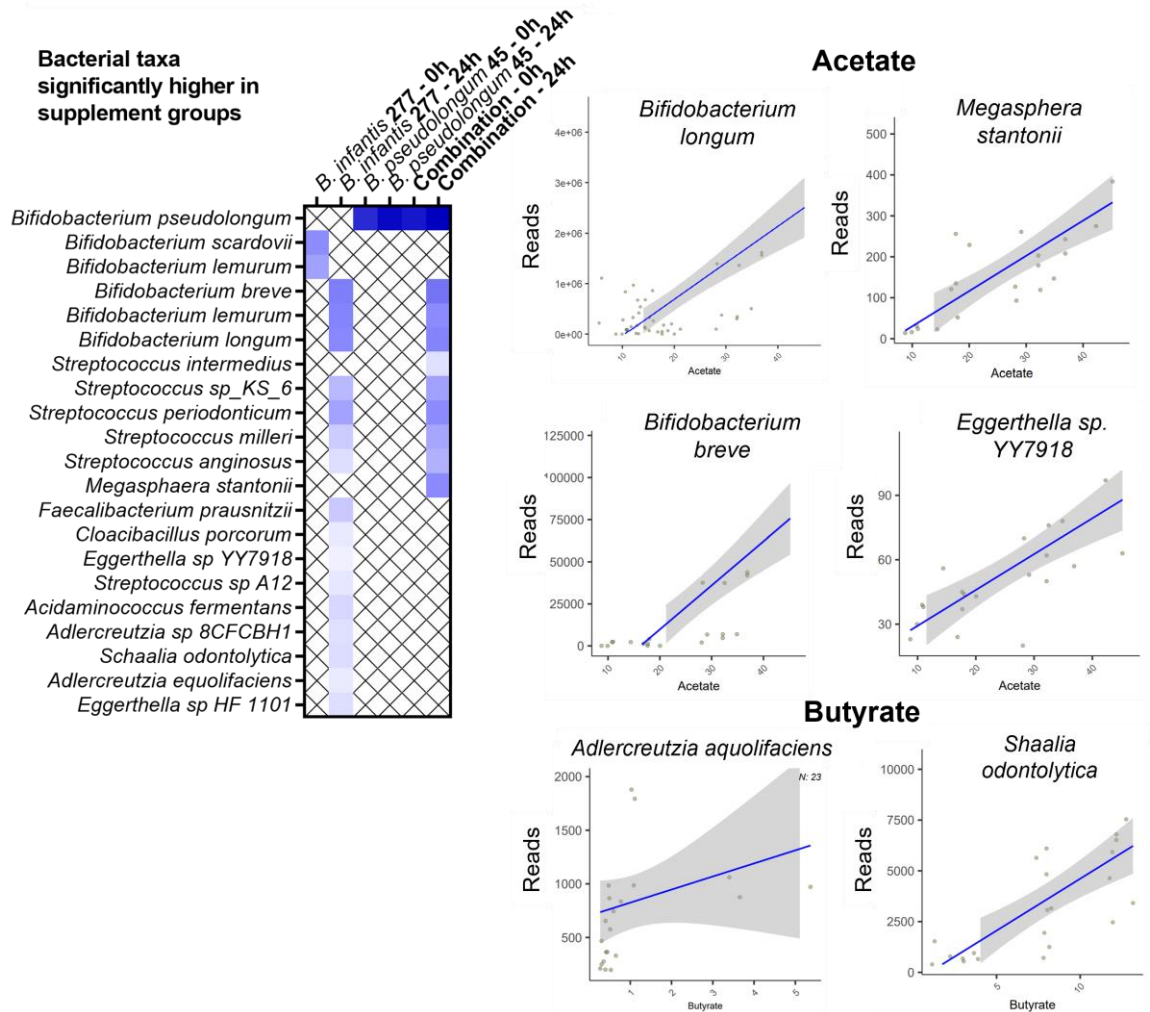
Figure 5-20 Statistically more abundant taxa in Bifidobacterium *supplemented Adult samples in HylonVII and correlation plots of species and acetate/butyrate concentrations performed by Maaslin2.*

In the presence of HylonVII, Adult 1 had the highest production of acetate in the supplemented groups, suggesting there was a positive effect of supplementation on metabolism of the starch in the presence of supplemented species (Figure 5-22). The highest production of acetate resulted from the Combination supplement, which may have been driven by the significant increases in acetate producing bacteria including ones supplemented: *B. pseudolongum*, *B. longum*, and *B. breve*. Adult 2 had the highest increases in acetate across all Adult donors and moderate increases in butyrate (Figure 5-22). In Adult 1, there was a lag phase

between 0h and 24h in butyrate concentrations which implies that butyrate was produced downstream in the metabolic pathway rather than being directly produced. Acetate production peaked and plateaued, whilst butyrate and propionate continued to increase at 48h especially in the supplement groups (Figure 5-22).
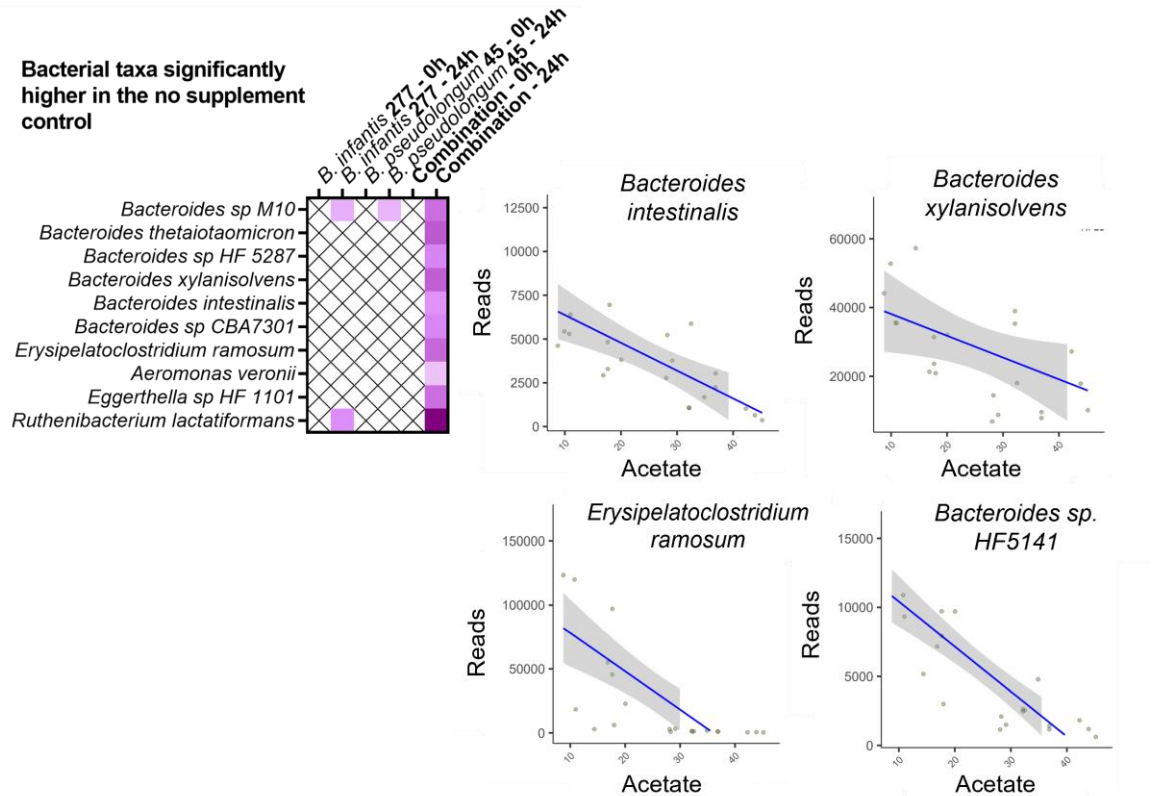
Figure 5-21 Significantly higher abundance species presence in the absence of a Bifidobacterium supplement, some of which are negatively correlated with acetate production.

Interestingly for Adult 2, acetate levels were highest in the *B. longum* subsp. *infantis* 277 supplemented groups, whilst the butyrate was highest in the no supplement and *B. pseudolongum* 45 supplement groups. The differential concentrations are likely due to the high levels of *Bifidobacterium* which produce acetate in the *B. longum* subsp. *infantis* 277 supplement and the increase in butyrate is likely caused predominantly by *Clostridium* or *F. prausnitzii* in the *B.*

pseudolongum 45 supplemented group which produce butyrate. However, the butyrate levels with the *B. pseudolongum* 45 supplement were similar to the non-supplemented group suggesting there may not be any practical advantages to supplementing with this strain.
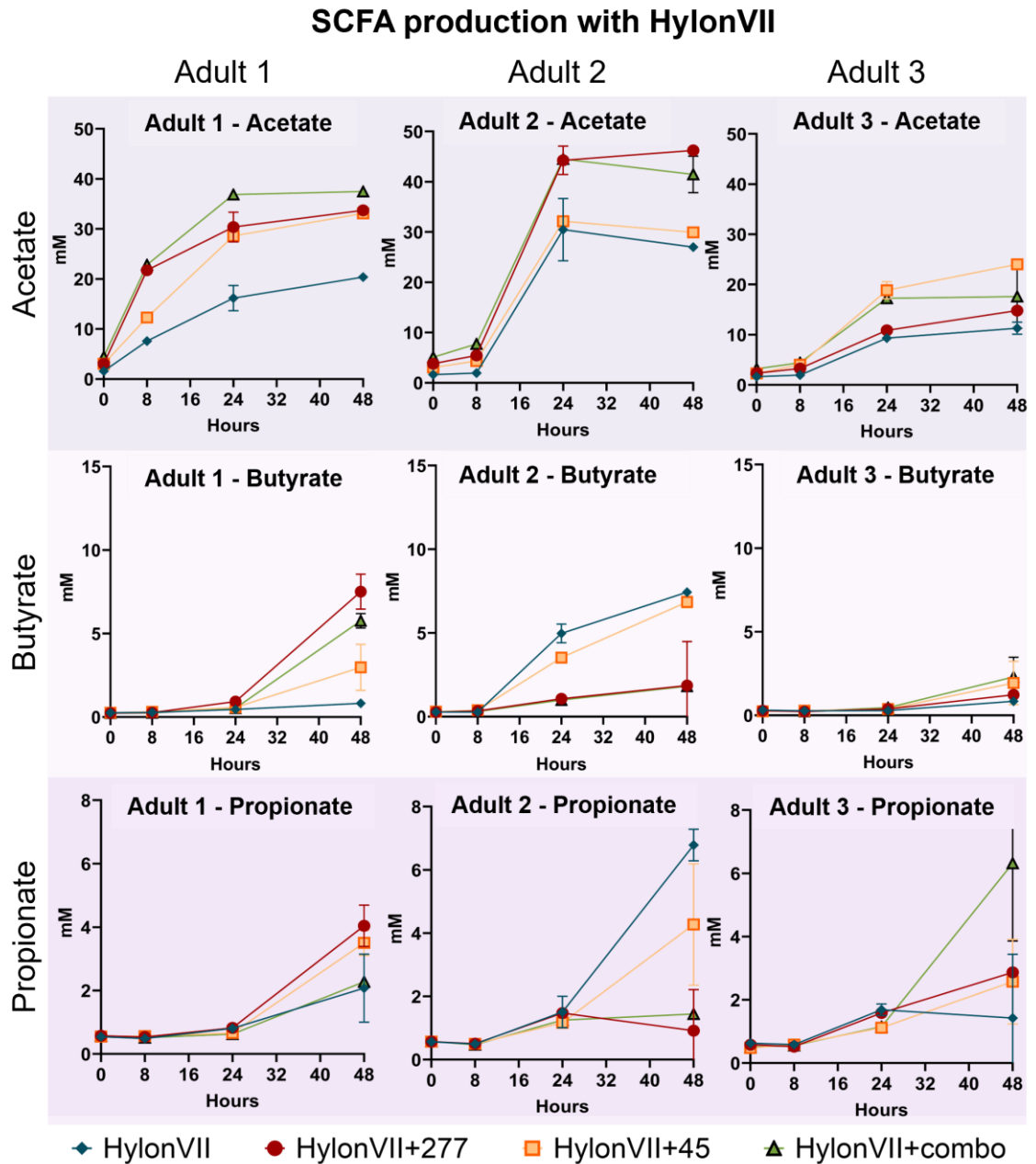
Figure 5-22 Acetate, butyrate, and propionate production over the course of 48h by each Adult donor stratified by bacterial supplement in the presence of HylonVII.

There were no significant effects of supplementation of acetate production when all donor data was pooled (Figure 5-23A), nor when only adults are pooled (Figure 5-23B). Combining Adult 1 and 2, supplementation of *B. longum* subsp. *infantis* 277 or Combination significantly increased acetate compared to no supplement in HylonVII ($p < 0.0001$) (Figure 5-23C).
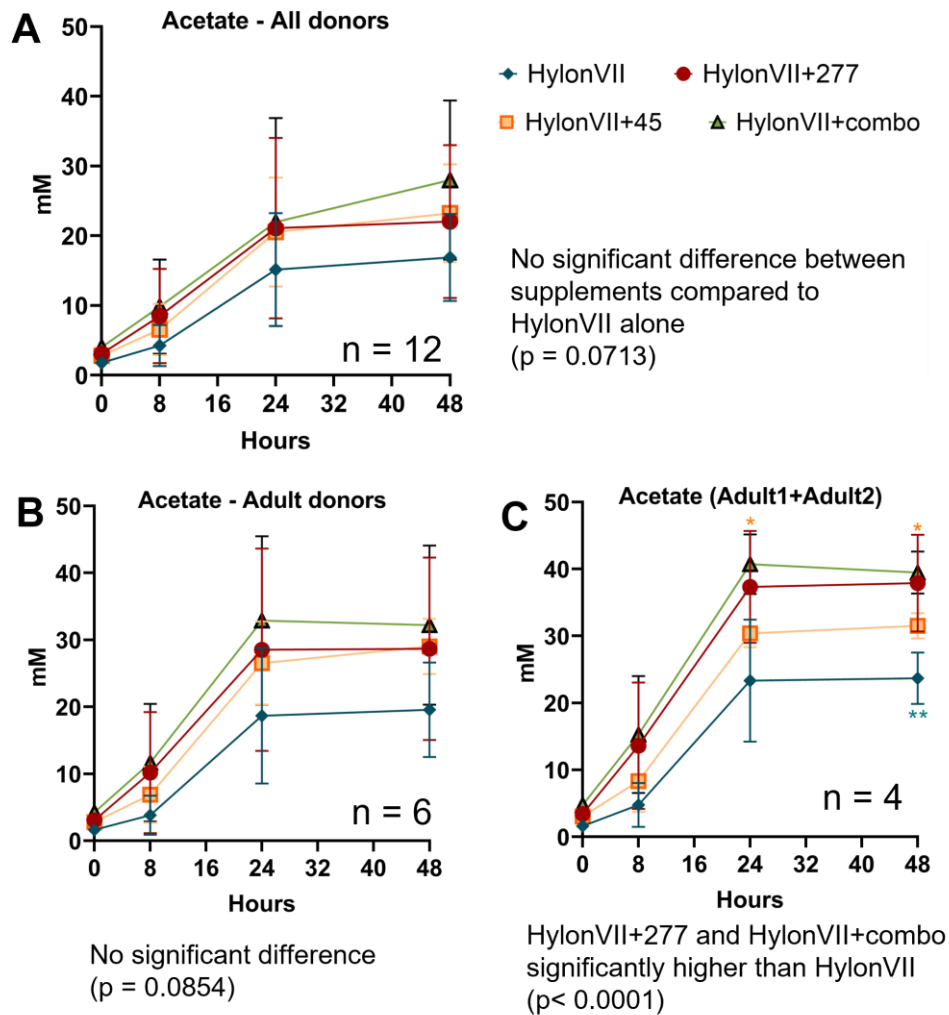
Figure 5-23 Acetate production by Adult donors showing the significant positive effects of 277 supplementation on acetate production but only in Adult 1 and Adult 2.

Adult 3 SCFAs concentrations to HylonVII were generally much lower compared to Adult 1 and 2 (Figure 5-22). *B. pseudolongum* 45 supplemented samples had

slightly higher acetate production which may be related to higher *Bifidobacterium* relative abundance observed. The propionate levels were highest in the Combination group but given the error between the replicates it is unlikely this effect was due to the supplement.

5.5 Discussion

Microbial diversity in infants was found to be significantly lower than adults which is generally accepted due to factors such as limited diet and continuing development of digestive system (Favier, Vaughan et al. 2002, Dobbler, Procianoy et al. 2017). *Bifidobacterium* dominate the niche having evolved to be the most adept bacterial genus at being able to break down fermentable oligosaccharides in human breastmilk (Masi and Stewart 2022). During weaning, diversity is higher with other taxa such as *Clostridium*, *Bacteroides*, and *Lactocaseibacillus* commonly found to become more abundant in human infant cohort studies which other analyses of weaning infant faecal stool composition (Penders, Thijs et al. 2006, Bergström, Skov et al. 2014, Wang, Mortimer et al. 2019). The pre-fermentation bacterial composition of the three infant donors tested in the present study reflect this. The role of starch in maturation of the infant microbiome is still being elucidated, however weaning is generally associated with increasing microbial diversity and decreasing *Bifidobacterium* due to competition with other bacterial taxa being able to utilise starch (McKeen, Young et al. 2019, Wang, Mortimer et al. 2019).

The infants selected for donations of faecal inocula for the *in vitro* fermentation were all weaning infants 8-12 months of age meaning the microbiota would be

adapting to the introduction of new fermentable substrates. The high amylose starch used in this small study, HylonVII, improved microbial diversity in 2 out of 3 infant microbiota tested, which confirms recent findings in a larger study which also tested the capacity of high amylose maize starch to modify the weaning infant microbiome (n=17) (Gopalsamy, Mortimer et al. 2019).

The *Clostridium* bloom in Infant 1 may have been caused by a previous gastrointestinal perturbance especially given the high growth rate of *Clostridium* (Kiu and Hall 2018). Infant 2 had the most positive response to HylonVII in the low abundance of potentially pathogenic taxa as well as achieving the highest levels of acetate and butyrate, which may be related to the donation being provided at 12 months of age rather than 8 months. The establishment of an enterotype-like structure has been estimated to occur between the age of 9 and 36 months in humans (Bergström, Skov et al. 2014). During the critical months following introduction of solid foods at 6 months of age, it is possible that Infant 2 had a more mature microbiome which could resist the rapid growth of *Clostridium* and had relatively high levels of *Enterococcus faecium*, some species of which can utilise starch and have probiotic characteristics depending on strain (Velikova, Stoyanov et al. 2016, Holzapfel, Arini et al. 2018, Krawczyk, Wityk et al. 2021, Krishna, Koujalagi et al. 2022).

In the current study, there were increases in species such as *Enterococcus faecium, C. butyricum, C. perfringens*, and *E. ramosum* which are known to utilise starch and/or starch hydrolysates such as maltose or maltodextrin (Wang, Conway et al. 1999). Bacterial isolate studies show that *C. butyricum*, *C. sordellii* (found in

Infant 3), and *C. perfringens* produce large amounts of butyrate in pure culture (Sharp and Macfarlane 2000). These results confirm the findings of another similar study of weaning infants (n = 23) *in vitro* high-amylose maize starch response (Wang, Mortimer et al. 2019). *C. butyricum* can thrive on starch equally well to glucose and produce butyrate directly through a butyrate kinase pathway rather than conversion from acetate (Wang, Conway et al. 1999, Tsukuda, Yahagi et al. 2021). Recent work has statistically associated the weaning process with an increase in Clostridiales order taxa and butyrate production due to use of acetyl-CoA enzyme pathway to produce butyrate as metabolic end-products (often converting acetate produced by other taxa) (Tsukuda, Yahagi et al. 2021). This would explain why there was a significant difference in butyrate between infant and adults in the current study. Recent studies show *Clostridium* colonisation in the infant gut are involved in butyrate production by colonising the intestinal mucus layer, meaning they are well adapted to intestinal colonisation of the gut (Van den Abbeele, Belzer et al. 2013). While the roles of *C. butyricum* in the gut can vary from commensal or pathogenic, it is considered to be an important emergent probiotic (Saarela 2019, Guo, Zhang et al. 2020). Thus, the butyrogenic interaction between *Clostridium* spp. and starch has the potential to be harnessed for formulation of precision synbiotics.

Some *Clostridium* members play a role in promotion of infant health not only during SCFA production but also by inducing beneficial immune responses which suggests that they may have co-evolved with humans as a gut commensal (Bhattacharjee, Flores et al. 2022). It is probable that as early commensals of the infant gut they evolved to degrade starch as it is a common food component

introduced in infancy, similar to *Bifidobacterium* spp. However, *C. perfringens* is commonly found to have pathogenic properties depending on the strain, although, there are hundreds of commensal strains (Kiu, Caim et al. 2017). *C. perfringens* and other similar taxa can cause common gastrointestinal infections in infants due to production of specific toxins. Some studies suggest the action of higher levels of butyrate may exacerbate intestinal necrotic lesions although and recent evidence shows butyrate production attenuated strains of *Clostridium* lead to reduced enteropathogenicity (Ferraris, Balvay et al. 2023). There were increases in butyrate-producing *Clostridium* in all Infant donors, which confirms the proliferation of *Clostridium* in the presence of starch, as well as significantly impacting butyrate production. The increase in potential pathogens in the presence of starch suggests that starch may modulate the niche-occupation of *Clostridium* spp. and could have clinical implications.

Clostridia are a diverse group of organisms, some members of which commensally occupy the human gut and opportunistically take advantage of vulnerabilities, such as in pre-term infants (Kiu, Shaw et al. 2021). The benefits received by commensal occupation of a wide variety of *Clostridium* members potentially outweigh the negative effects of the opportunistic pathogenic members. The weaning process may repeatedly perturb the microbiota hence orchestrating a maturation process to increase competition between microbes in the ecosystem. Some *Clostridia* members could conductors of that maturation and immune priming process. The ecological relationship between starch and growth of pathogenic strains in infants which harbour them would be important to assess further. It is worth reiterating that *Clostridium* increased in the non-starch

Infant samples in this study and while the presence of starch exacerbated proliferation, there could be other factors associated with the *in vitro* methodology which caused proliferation. The ability of highly competitive species to outcompete in the batch fermentation model has been noted previously as a draw-back of the method (Pham and Mohajeri 2018).

E. ramosum is known as an opportunistic pathogen involved in the development of life-threating conditions such as bacteraemia and septicaemia. The species used to be known as *Bacillus ramosum*, *Clostridium ramosum* and unofficially renamed in 2013 to *Erysipelatoclostridium ramosum* where it is known in the databases used for taxonomic designation in the present study. Recently and after analysis of these data, it was re-classified again as *Thomasclavelia ramosa* but for continuity in this thesis it will be referred to as *E. ramosum* (Lawson, Saavedra Perez et al. 2023). *E. ramosum* are found to be more abundant in starch-rich diets and positively correlated to adiposity, the proposed mechanism being by modulating host lipid metabolism as shown in one study in chickens whose primary feed ingredient was cornmeal (Liu, Wang et al. 2022). This was also observed in a resistant starch feeding study in pigs (Haenen, Zhang et al. 2013). In another study in pigs, the amylopectin and starch contents affected the abundance of *E. ramosum*: cereals high in amylose caused proliferation of *C. butyricum*, whilst cereals high in amylopectin promoted *E. ramosum* and *Bacteroides* (Tiwari, Singh et al. 2019). In the absence of testing a high amylopectin starch in this small study, it is unclear if the opposite of this effect occurred or if generally the *E. ramosum* proliferated due to the presence of starch not necessarily hinging on the amylose/amylopectin ratios. In any case, *E. ramosum* was promoted in the

presence of a high amylose maize starch. The role of *E. ramosum* in the human gut microbiota is less well understood, thus the reasons for consistent increases in the presence of starch could be the focus of future study (Bhattacharjee, Flores et al. 2022).

The Adult donors included in this small study provide insight into the response of a mature, stable microbiome to the introduction of HylonVII and bifidobacterial supplementation with regards to various downstream effects on microbial composition and metabolism. The Bacillota dominant enterotype of Adult 1 and 2 differed compared to Adult 3 which had a Bacteroidota-dominant microbiome pre-fermentation, which has been a previously identified variable impacting response to fermentation substrates (Bendiks, Knudsen et al. 2020) (Costea, Hildebrand et al. 2018). The Adult response to HylonVII confirmed previous studies: high amylose maize starch reduced bacterial alpha-diversity but altered beta diversity and ultimately changed the global structure of the microbiome (Bendiks, Knudsen et al. 2020, Deehan, Yang et al. 2020, Teichmann and Cockburn 2021, Bendiks, Guice et al. 2022). The purported mechanism of a decline of diversity is related to the selection of specific taxa (often to the species level) of discrete fibre structures which supports the hypothesis that a variety of fermentation substrates is required to maintain diversity; studies exploring the combination of complex combinations of substrates corroborate this (Chung, Walker et al. 2019).

In the present study, there were increases in genera such as *Bacteroides* and *Prevotella* in response to resistant and/or high-amylose maize starch concurrent to previous research (Wang, Conway et al. 1999, Warren, Fukuma et al. 2018,

271

Bendiks, Guice et al. 2022). Some report Actinobacteria (*Bifidobacterium* and *Collinsella*) as the most responsive phylum to high amylose maize (RS Type 2) supplementation (Ordiz, May et al. 2015). These genera were both found to be statistically higher in the presence of RS across all Adults in the current study especially *B. animalis* confirming previous work that show it is correlated to human dietary patterns including resistant starch (Rosés, Cuevas-Sierra et al. 2021). Increases in other *Bifidobacterium* spp. were observed particularly in Adult 1 and 2, as baseline levels of *Bifidobacterium* and *Collinsella* in Adult 3 were very low. Adult 3 had very few detectable taxa assigned to *Bifidobacterium* meaning they could not increase if they were not occupying a niche in that individual's microbiome; this was the reasoning behind another *in vitro* fermentation study also exploring supplementation of starch-degrading strains (Teichmann and Cockburn 2021). Adult 3 had a different microbial composition before and after the introduction of starch and had the lowest production of SCFAs. Though, Adult 3 mildly responded to introduction of amylolytic *B. pseudolongum* strain, during which it maintained levels of *Bifidobacterium* in the presence of starch and caused a higher production of both acetate and propionate compared to the control group with no bacterial supplement. *B. pseudolongum* was identified as a pivotal, keystone degrader of RS in rats meaning that even in low abundances its presence can promote growth of other taxa to benefit the microbiome (Centanni, Lawley et al. 2018). In the Combination supplement, several *Bacteroides* spp. decreased in abundance. It is possible that *B. pseudolongum* 45 supplementation could enhance populations of *Bifidobacterium* and other beneficial taxa in Bacteroidota dominant microbiomes to boost SCFA production, however future

studies should assess this with more donors with this enterotype as it is not conclusive in the present study due to limited sample size.

In Adult 2, increases in *B. adolescentis* and *B. longum* were observed both of which have been generally shown to utilise starch in isolation and *in vitro* fermentations (Sharp and Macfarlane 2000, Duranti, Turroni et al. 2014, Jung, Seo et al. 2020, Tarracchini, Milani et al. 2021). As a primary degrader of starch, *B. adolescentis* can also cross-feed other taxa detected in this study such as *Anaerobutyricum hallii* and *F. prausnitzii* which can complete metabolic conversion of bifidobacterial metabolites (acetate) to butyrate (Belenguer, Duncan et al. 2006, Rios-Covian, Gueimonde et al. 2015). *A. hallii* (formerly *Eubacterium hallii* (Shetty, Zuffa et al. 2018)) and *Eubacterium limosum* were significantly more highly abundant in Adult donors compared to Infant donors in the presence of HylonVII, implying the presence of these taxa are (a) more common in Adult than Infant donors and (b) are upregulated in the presence of HylonVII. Growth has been reported for *E. limosum* in the presence of high amylose maize starch which implies it has the capability to utilise it in isolation culture, although growth rates were low (Wang, Conway et al. 1999). Therefore, as well as being a producer of butyrate in pure culture, it could be participating in degradation alongside other stronger degraders and/or utilising any hydrolysates e.g. dextrins produced by stronger degraders (Wang, Conway et al. 1999, Sharp and Macfarlane 2000, Dobranowski and Stintzi 2021).

However, there was not a significant increase in butyrate in response to HylonVII as the majority of SCFA present was acetate, suggesting production and/or

conversion to butyrate was low in these three Adult microbiomes. One of the purported benefits of starch as a fermentable fibre is the accessibility and wide capability of microbial gut taxa to convert starch into beneficial SCFAs including butyrate especially as a targeted treatment of conditions such as irritable bowel syndrome (Kaoutari, Armougom et al. 2013, Walker and Lawley 2013, Cerqueira, Photenhauer et al. 2020). Although as previously addressed, inter-individual variation is continually the primary determinant of responsiveness to modulation of the microbiome with dietary components (Walker and Lawley 2013). In efforts to address this, some studies researching microbiota response to a treatment delineate human samples by microbiome enterotypes as a rudimentary stratification measure (Chen, Long et al. 2017, Xie, Ding et al. 2021, Yu, Arıoğlu-Tuncil et al. 2021). Although not without its limitations, advances in approaches to measuring and analysis of microbiome composition means that it is easier to define enterotypes dependant on Bacteroidota-Bacillota ratio (Costea, Hildebrand et al. 2018).

There were several taxa which were enriched and increased in relative abundance in the presence of HylonVII such as *Bacteroides vulgatus*, *Collinsella aerofaciens*, *B. longum*, *B. adolescentis* but were not found to be statistically significant due to the prevalence cut off being set to prevalent among 90% of samples. Due to the differences in enterotype especially in Adult 3, there were several taxa missing and therefore the statistical analyses conducted here preclude any significant results for these taxa which are commonly attributed to RS intake, as shown *in vivo* (Walker, Ince et al. 2011).

Adult 1 and 2 responded with increased levels of acetate and butyrate in the presence of starch which was significantly enhanced both in metabolic output and taxonomic composition to the introduction of *B. longum* subsp. *infantis* 277. It appears that this supplement shifted the microbiome structure of Adult 1 from *Escherichia, Bacteroides*, and *Phocaeicola* dominant composition towards an Actinobacteria-dominant one. This supplement caused an increase in several other taxa such as butyrate producing *F. prausnitzii*, two species of *Bifidobacterium*, and other butyrate and acetate producing bacterial taxa. The mechanisms for this happening are unclear due to the fact that this strain is not amylolytic as shown in Chapter 3. There was a significant increase in *B. longum* abundance in the presence of starch which was higher than the 0h timepoint, which means that the presence of *B. longum* subsp. *infantis* 277 itself increased in abundance or other spp. of *B. longum* increased. It is also possible that there was cross-feeding of starch hydrolysates from primary degraders present (*B. adolescentis, B. breve*) to *B. longum* subsp. *infantis* 277 or to other *B. longum* species. More rigorous research is required to address the effects of this supplement on the human microbiome *in vitro*, especially by developing a higher-powered study design.

5.6 Limitations

In the present *in vitro* study, I sought to investigate the ecological impacts of high amylose maize starch combined with *Bifidobacterium* strains. There were several limitations which were due to the experimental design, methodology selected, and bioinformatics support which prevented conclusive analysis and impact.

As discussed in the introduction of this chapter, there are known differences between batch *in vitro* fermentations of substrates using stool samples compared to chemostat (dynamic) systems which can more closely simulate the gut environment. Secondly, the timely procurement of eligible human infant and adult stool samples limited the sample size (three from each age category), and all donations were frozen prior to experimentation. The dominance of facultative anaerobes after 24h indicates that the stool was in contact with oxygen at some stage between donation and the end of the experiment in a manner which may have impacted the bacteria detected.

Additionally, the three individuals (using two biological replicates) had highly varied responses to each experimental condition, meaning that with n = 3, there were few conclusions which could be made about the effect of each condition on either the bacterial composition or metabolites produced. A higher sample size from one age group and/or more biological replicates would be essential to ameliorating the statistical power to assess the efficacy of each treatment.

Lastly, analysis of the metagenomic reads with regards to the classification of each read to a respective taxonomic species was limited by the reference database used. During the use of Kraken 2, the taxon identity is assigned from the NCBI taxonomy database, it is apparent that a bespoke classification is required for mis-classified taxa which did not biologically align with what was expected from an adult or infant gut microbiome.

5.7 Conclusion

There were observed and statistically confirmed positive effects of HylonVII in both donor age groups selected, Adult and Infant. The presence of HylonVII enhanced the diversity of weaning Infant donors inoculated *in vitro* fermentations and caused significant increase of several taxa which are primary and secondary degraders of starch resulting in significantly increased acetate and butyrate production. The response may provide insight into the role of the impacts of weaning and starch intake during infancy, including potentially negative effects such as *Clostridium* blooms.

Adult donors yielded interesting results with regard to the effectiveness of *Bifidobacterium* strains supplemented into the *in vitro* fermentation of HylonVII, showing that the two strains elicited different compositional and metabolic shifts which may have applications in endeavours in precision modulation of the microbiome. The efficacy of both HylonVII and bacterial supplementation was affected by varying enterotypes of the donors which may also be important to address when personalising synbiotic or dietary approaches.

Chapter 6

6.1 Discussion

Bifidobacteria are highly abundant in the infant microbiome and are prominent species in the adult gut as they are capable of fermenting resistant starch (Arboleya, Watkins et al. 2016, O'Callaghan and van Sinderen 2016, Duranti, Longhi et al. 2020). This thesis explored the interaction between resistant starch and *Bifidobacterium*. The varying physical structure of starch and its resistant fractions provides unique opportunity to exploit its versatility as a dietary fibre (DeMartino and Cockburn 2020, Dobranowski and Stintzi 2021, Louis, Solvang et al. 2021). The impacts of starch structure on fermentation by *Bifidobacterium* strains was assessed by measuring isolate growth, gene expression, metabolite output using starches with different processing, botanical origins, and amylose content. There was a focus on expanding knowledge of species and strains of *Bifidobacterium* not including *B. adolescentis* which is already well studied for its RS utilisation. However, there have been few recent advances in understanding the mechanisms of degradation by *Bifidobacterium* in general.

Proving starch utilisation under lab culturing conditions presented various challenges, especially during measurement of cell replication, extraction of nucleic acids and protein: light scattering of starch in spectrophotometry, pipetting, interfering with nucleic acid flow columns and the impacts of heat (sterilisation) on starch structure. For each of these, considerable development was required

including the development of a non-heat sterilisation method involving ethanol and UV treatment of starch powders.

In the infant microbiome, the high abundance of *Bifidobacterium* is due to the general ability of certain species and strains to utilise HMOs, often cooperatively, which provides health benefits to the host (Lawson, O'Neill et al. 2020, Walsh, Lane et al. 2022). The *Bifidobacterium* pangenome contains a high number of genes specialised in HMO degradation which results in *Bifidobacterium* dominating the infant microbiome (Derrien, Turroni et al. 2022). These evolutionary circumstances saw *Bifidobacterium* become key stone degraders of HMOs as the majority of the genus gained (and/or horizontally transferred) the genes to convert breastmilk sugars (Milani, Turroni et al. 2016). There is therefore not a requirement for a single species or strain of *Bifidobacterium* to independently degrade a substrate as they grow cooperatively (Milani, Lugli et al. 2015).

The introduction of solid food including starch-rich foods during the weaning phase triggers maturation of the gut microbiome (Lin and Nichols 2017). In this thesis, evidence is presented of a clade of *B. pseudocatenulatum* which was isolated from an infant at 6 months of age which was a cross-feeder of HMO by-products (Lawson, O'Neill et al. 2020) and in the present study had moderate starch degrading capabilities. The isolates obtained contained an average of around 100 GH family enzymes specific to a variety of substrates indicating it may be a transitional species in the infant gut (Table 3-5). Interestingly, the clade also possessed CBM74 modules which may have been involved in its starch utilisation phenotype as observed in other organisms (Valk, Lammerts van Bueren et al.

2016, Photenhauer, Cerqueira et al. 2022). Recent studies show there is a shift in the complement of *Bifidobacterium* species and strains present pre- and post-weaning, namely in *B. longum* (Kujawska, La Rosa et al. 2020, Vatanen, Ang et al. 2022).

Disproportionately in low- and middle-income countries, weaning is associated with enteric infections providing ample reason to further understand how the health-promoting bacteria such as *Bifidobacterium* and starch impact gut ecology (Wang, Mortimer et al. 2019). During weaning, *Bifidobacterium* were present in relatively high abundance which may have conferred a selective advantage for adapting to the starch in the diet. The abundance and permanence of starch in the human diet may have led to the emergence of highly prolific starch degraders (e.g. *R. bromii* and *B. thetaiotaomicron*) which carved a niche by evolving more efficacious enzyme strategies (Carmody, Bisanz et al. 2019, Cerqueira, Photenhauer et al. 2020). Additionally, there is likely a negative impact of cooking and processing on dietary intake of RS. In animals whose dietary components are not often cooked, there are *Bifidobacterium* species much more adept at degrading resistant starch (Centanni, Lawley et al. 2018, Jung and Park 2023). While cooking increased the overall quantity of starch in the human diet, it is possible that the advent of cooked starchy food low in resistant starch portions reduced the evolutionary pressures for prevalent, health-promoting commensals such as *Bifidobacterium* to evolve strong RS binding and degradative strategies (Carmody, Bisanz et al. 2019). Milani et al. performed *in silico* gene loss or decay analysis of *Bifidobacterium* and only detected a significant gene decay process for GH13 enzymes in bifidobacteria isolated from insects (Milani, Turroni et al. 2016).

The mechanisms of starch utilisation by *Bifidobacterium* are limited to a few examples of the production of amylase enzymes with relevant CBM domains in *B. adolescentis* and *B. breve* (Connell Motherway, Fitzgerald et al. 2008, Jung, Seo et al. 2020). The isolate studies performed in this thesis of individual strains used preparations of starch with varying resistance to fermentation which was developed to reproducibly highlight variation in utilisation between strains. This thesis presented data that shows novel functions for three Ig-like domain containing proteins with presumed amylolytic function and cell-surface anchoring which form a cluster in the *B. pseudolongum* subsp. *globosum* NCIMB 702245 genome. These genes were found to be more highly expressed in the presence of a resistant starch and their protein products were the most significantly upregulated proteins in the presence of a normal maize starch. The gene cluster contains a protein transporter upstream which is suggestive that they may be transcribed and transported extracellularly as part of one operon. This is evidence of potentially the first starch related multi-protein gene cluster, or PUL, in *Bifidobacterium*. The presence of flanking CBM74, CBM25, and CBM26 domains are likely to be fundamental to efficiency of the amylase enzymes produced.

The data produced from the *in vitro* colon models inoculated with infant or adult stool provided insight into the age-related response of the microbiome to high amylose starch. Both age groups responded with significantly higher levels of SCFAs in the presence of high amylose maize starch, whilst the microbiome compositional response varied by individual in each age group which is typical of RS microbiome studies (Venkataraman, Sieber et al. 2016, Bendiks, Knudsen et al. 2020, DeMartino and Cockburn 2020). The infant microbiome response to RS

was significantly more butyrogenic which may be attributed to the early colonisation of *Clostridium* taxa common in early life especially associated with the weaning process and butyrate production generally (Tsukuda, Yahagi et al. 2021). Butyrate is commonly found to increase in response to RS in adults, but inter-individual variation remains a major challenge in microbiome research and the reasoning behind the interest in "personalised" microbiome interventions (Costea, Hildebrand et al. 2018, Klimenko, Odintsova et al. 2022). Combined with advances in understanding how molecular fine structure of starch can uniquely impact microbiome or "precision modulation", there may be some imminent material advances in efforts to improve human health using RS (Deehan, Yang et al. 2020, Dobranowski and Stintzi 2021, Ravi, Troncoso-Rey et al. 2022).

The supplemented strains of *Bifidobacterium* (*B. pseudolongum* subsp. *globosum* NCIMB 702245 and *B. longum* subsp. *infantis* LH277) when included in the fermentation of starch demonstrated the ability of strains to persist over time *in vitro* as well as the probiotic potential of *B. longum* subsp. *infantis* LH277 which significantly impacted the microbiome composition of both adult and infant donors. Moreover, *B. longum* was higher in abundance at 24h of starch fermentation implying that the bacterium not only persisted in the model but may have (a) proliferated to achieve higher abundance or (b) caused the proliferation of other *B. longum* species. *B. longum* is known to be a beneficial strain to human health and is associated adapting to the changing infant microbiome (Kujawska, La Rosa et al. 2020). In the presence of starch, the acetate production was significantly increased by *B. longum* subsp. *infantis* LH277 in adult donors only in spite of the strain not being amylolytic. Positive effects of synbiotics are typically due to the

strain and substrate combination being compatible, as has been demonstrated in *Bifidobacterium lactis* (Le Leu, Hu et al. 2010).

6.2 Future Perspectives

It would be important to assess the protein characterisations of the novel, cell-wall anchored, amylolytic Ig-like proteins discovered in *B. pseudolongum* subsp. *globosum* NCIMB 702245, the structure of which has been predicted using AlphaFold (Figure 6-1). The 3-dimensional structure using crystallography and potentially the necessity of each domain for cell-wall anchoring (S-layer homology) and RS binding (CBM74) such as that carried out in a similar protein, Amy13K, in *E. rectale* could be investigated (Cockburn, Suh et al. 2018, Cerqueira, Photenhauer et al. 2022). Knocking out the CBMs from Amy13K reduced enzyme activity on corn starches by ~40-fold. It would also be interesting to assess if the proteins are complexed in any way, as the three proteins identified were on the same genetic locus and it would be logical for them to be co-located on the outside of the cell.

Genetic modification of strains to produce gene knockouts for candidate genes (such as those identified in the bioinformatic and proteomic analyses) would help to confirm the protein functions. While it is challenging to carry out in this genus of bacteria, gene knockout and gene cloning of the amylopullanase in *B. breve* UCC2003 has been successfully carried out previously (Connell Motherway, Fitzgerald et al. 2008). As an imperfect proxy for gene knockouts, the two subspecies of *B. pseudolongum* were selected such that one contained the CBM74 containing amylase gene cluster which was absent in the other.
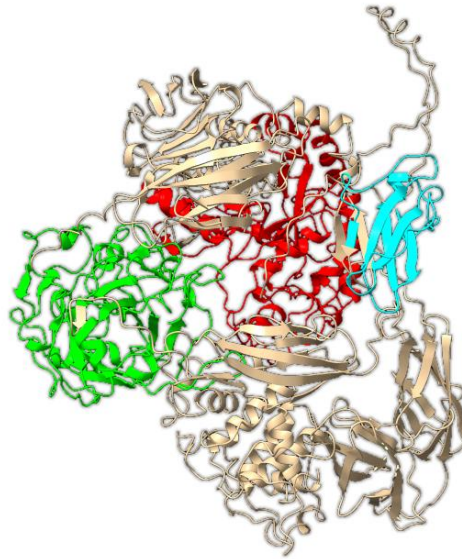
Figure 6-1 Protein structure of Ig-like domain-containing protein (WP_099309720.1) containing domains: amylase (red), CBM74 (green), CBM25 (cyan) in B. pseudolongum *45 predicted by AlphaFold, courtesy of David Lawson, John Innes Centre, Norwich, UK.*

There was significant sequence homology of at least one of the novel amylolytic proteins in *B. pseudocatenulatum* and *B. adolescentis* spp. It would be pertinent to assess the importance of these proteins in other bifidobacteria. Moreover, there were CBM74 modules detected in *B. pseudocatenulatum* isolates tested in this thesis. Further research could be directed towards the presence and function of this module in other *Bifidobacterium* spp. especially as the *B. pseudocatenulatum* clade may be important transitional taxa for the microbiome development during weaning, as discussed in Chapter 3. As well as phenotypically assessing taxa such as *B. adolescentis* strains in possession of CBM74 versus without for RS binding and hydrolysis, further phenotypic testing using a wider range of starch structures should be considered. RS utilisation was assessed using a pre-gelatinised high amylose maize starch for the majority of strains; however, future work could evaluate granular starch (potato, maize) or retrograded starches

(normal or high amylose variety). In particular the variation in localisation of bacterial cells or biofilm-like organisation on starch gel surfaces may be interesting to research via microscopy (Koev, Harris et al. 2022). For example, strains *B. pseudolongum* subsp. *globosum* 45 and *B. breve* LH24 utilised gelatinised high-amylose maize but were not tested on granular forms of starch. Combining bacteria with module CBM74 proteins and RS may improve the next generation of synbiotics.

B. longum subsp. *infantis* LH277 and the *B. pseudocatenulatum* LH strains were isolated from the same infant. Cross-feeding co-culture experiments with strong vs weak starch utilising *Bifidobacterium* would highlight any cooperative behaviours especially in the context of weaning. The LH strains isolated from the same infant at different timepoints may be interesting candidates to understand the interaction between isolates within the same ecosystem (Lawson, O'Neill et al. 2020). The potential health effects of *B. longum* subsp. *infantis* LH277 may also be investigated as a new probiotic in gut organoid models for example (Lauder, Kim et al. 2020, Meng, Sommella et al. 2020, Zhou, Fang et al. 2020). Especially, the driving mechanisms behind modulatory effects on microbiome composition and metabolite output would an interesting research question to address.

Finally, the transcriptomic response of *B. pseudolongum* to starch gave rise to new, intriguing questions about the upregulation of enzymes involved in starch degradation. There is little knowledge about transcriptional regulators of genes regarding the biochemical triggers, nutrient sensing mechanisms, physicochemical signatures specific to RS, whilst there are many of these details for substrates

such as maltose (Khoroshkin, Leyn et al. 2016). Do various structures of starch initiate a differential transcriptomic response? Given more time, there would have been an opportunity to analyse the temporal upregulation of enzymes after the introduction of starch to the bacterial cells as was intended by the experimental design (which included 12h and 24h RNA sequencing). Additionally, it would be important to map the RNA sequence reads from each subsp. of *B. pseudolongum* to the corresponding reference genomes.

Supplementary Materials

Table S1 Two-way ANOVA Tukey's multiple comparison test of reference strains in the presence or absence of gelatinised 1 or 2 % SPS.		
B. breve NCIMB 2258	**Significance**	**Adjusted P Value**
Starch 1% vs. Starch 2%	ns	0.9947
Starch 1% vs. No substrate	***	0.0002
Starch 2% vs. No substrate	***	0.0005
B. breve UCC2003		
Starch 1% vs. Starch 2%	ns	0.9999
Starch 1% vs. No substrate	****	<0.0001
Starch 2% vs. No substrate	****	<0.0001
B. pseudolongum NCIMB 702244		
Starch 1% vs. Starch 2%	ns	0.8793
Starch 1% vs. No substrate	****	<0.0001
Starch 2% vs. No substrate	****	<0.0001
B. pseudolongum NCIMB 702245		
Starch 1% vs. Starch 2%	ns	0.9893
Starch 1% vs. No substrate	****	<0.0001
Starch 2% vs. No substrate	****	<0.0001
B. longum NCIMB 8809		
Starch 1% vs. Starch 2%	ns	0.991
Starch 1% vs. No substrate	ns	0.8319
Starch 2% vs. No substrate	ns	0.6664
L. caseii NCIMB 4114		
Starch 1% vs. Starch 2%	ns	0.9424
Starch 1% vs. No substrate	ns	>0.9999
Starch 2% vs. No substrate	ns	0.9283
B. longum subsp. infantis 20088		
Starch 1% vs. No substrate	ns	0.5229

Table S2 Two-way ANOVA (Repeated measures) comparing Acetate concentrations by B. adolescentis *702204.*

0h	Summary	Adjusted P Value
Glucose vs. HylonVII	ns	0.9969
Glucose vs. No substrate	ns	0.7754
Glucose vs. Normal maize	ns	0.0805
HylonVII vs. No substrate	ns	0.9906
HylonVII vs. Normal maize	ns	0.0764
No substrate vs. Normal maize	ns	0.0825
10 h		
Glucose vs. HylonVII	****	<0.0001
Glucose vs. No substrate	****	<0.0001
Glucose vs. Normal maize	*	0.0154
HylonVII vs. No substrate	*	0.0252
HylonVII vs. Normal maize	ns	0.4269
No substrate vs. Normal maize	ns	0.9164
14 h		
Glucose vs. HylonVII	****	<0.0001
Glucose vs. No substrate	****	<0.0001
Glucose vs. Normal maize	*	0.0103
HylonVII vs. No substrate	*	0.0293
HylonVII vs. Normal maize	ns	0.2758
No substrate vs. Normal maize	ns	0.8582
18 h		
Glucose vs. HylonVII	***	0.0003
Glucose vs. No substrate	****	<0.0001
Glucose vs. Normal maize	**	0.0086
HylonVII vs. No substrate	ns	0.2567
HylonVII vs. Normal maize	ns	0.8757
No substrate vs. Normal maize	ns	0.9606
24 h		
Glucose vs. HylonVII	**	0.0061
Glucose vs. No substrate	**	0.0019
Glucose vs. Normal maize	****	<0.0001
HylonVII vs. No substrate	ns	0.9759
HylonVII vs. Normal maize	ns	0.2307
No substrate vs. Normal maize	ns	0.2180
48 h		
Glucose vs. HylonVII	*	0.0362
Glucose vs. No substrate	*	0.0364
Glucose vs. Normal maize	*	0.0303
HylonVII vs. No substrate	ns	0.5269
HylonVII vs. Normal maize	ns	0.2881

| No substrate vs. Normal maize | ns | 0.1301 |

Table S3 Ig-like domain containing protein (WP_051894162.1) Ig-like domain-containing protein (GH13_28 + CBM25)

ORF	Size	Database	Start	Stop	E-value	ID	Protein domain
orf40	1397	PRINTS	253	264	1.10E-06	IPR006046	Alpha amylase
orf40	1397	Pfam	841	930	3.70E-22	IPR005085	CBM25
orf40	1397	SUPERFAMILY	1035	1107	3.37E-13	IPR008964	Invasin/intimin cell-adhesion fragments
orf40	1397	Pfam	148	270	1.50E-18	IPR006047	GH13, catalytic domain
orf40	1397	Phobius	35	52	-	-	C-terminal region of a signal peptide.
orf40	1397	CDD	84	446	0	-	AmyAc_bac1_AmyA
orf40	1397	Phobius	1	25	-	-	N-terminal region of a signal peptide.
orf40	1397	SUPERFAMILY	444	525	1.40E-09	-	GH domain
orf40	1397	Pfam	939	1015	2.30E-16	IPR003343	Bacterial Ig-like domain, group 2
orf40	1397	Pfam	1344	1386	1.10E-05	IPR001119	S-layer homology domain
orf40	1397	Pfam	1205	1248	3.10E-09	IPR001119	S-layer homology domain
orf40	1397	SUPERFAMILY	1125	1196	3.10E-14	IPR008964	Invasin/intimin cell-adhesion fragments
orf40	1397	PANTHER	97	408	6.30E-22	-	ALPHA-AMYLASE FAMILY MEMBER
orf40	1397	SMART	840	918	6.40E-16	IPR005085	CBM25
orf40	1397	SUPERFAMILY	936	1016	4.32E-18	IPR008964	Invasin/intimin cell-adhesion fragments
orf40	1397	Phobius	26	34	-	-	Hydrophobic region of a signal peptide.
orf40	1397	Gene3D	617	721	4.50E-11	IPR013783	Ig-like fold
orf40	1397	SUPERFAMILY	85	427	3.48E-50	IPR017853	(Trans)glycosidases
orf40	1397	Gene3D	835	933	2.20E-29	IPR013783	Ig-like fold
orf40	1397	SMART	85	436	5.40E-11	IPR006047	GH13, catalytic domain
orf40	1397	Phobius	1	52	-	-	Signal peptide region
orf40	1397	Pfam	624	700	6.60E-10	IPR031965	Starch-binding CBM26
orf40	1397	SMART	938	1015	1.10E-18	IPR003343	Bacterial Ig-like domain, group 2
orf40	1397	SMART	1113	1195	2.00E-08	IPR003343	Bacterial Ig-like domain, group 2
orf40	1397	SMART	751	830	6.60E-12	IPR003343	Bacterial Ig-like domain, group 2
orf40	1397	Gene3D	80	441	#######	-	Glycosidases
orf40	1397	Phobius	53	1397	-	-	Region of a membrane-bound protein predicted as extracellular
orf40	1397	TMHMM	21	43	-	-	Region predicted to be membrane-embedded

| orf40 | 1397 | SUPERFAMILY | 753 | 832 | 1.26E-11 | IPR008964 | Invasin/intimin cell-adhesion fragments |
| orf40 | 1397 | Gene3D | 442 | 527 | 9.70E-22 | IPR013780 | Glycosyl hydrolase, all-beta |

Table S4 Ig-like domain protein (WP_099309720.1) Ig-like domain-containing protein (GH13_28 + CBM74 +CBM25)							
ORF	Total size	Database/Tool	Start	Stop	E-value	ID	Protein domain
orf88	1537	SMART	92	457	6.70E-12	IPR006047	GH13, catalytic domain
orf88	1537	SMART	467	549	0.0075	IPR031319	Alpha-amylase, C-terminal domain
orf88	1537	PRINTS	364	376	1.80E-08	IPR006046	Alpha amylase
orf88	1537	SMART	1059	1148	3.90E-10	IPR005085	CBM25
orf88	1537	PANTHER	82	530	9.40E-33	-	ALPHA-AMYLASE
orf88	1537	SUPERFAMILY	90	448	6.21E-50	IPR017853	(Trans)glycosidases
orf88	1537	SUPERFAMILY	991	1053	3.18E-16	IPR008964	Invasin/intimin cell-adhesion fragments
orf88	1537	Pfam	1062	1160	1.00E-19	IPR005085	CBM25
orf88	1537	CDD	91	467	1.05E-19	-	AmyAc_bac1_AmyA
orf88	1537	SUPERFAMILY	467	546	1.34E-09	-	GH domain
orf88	1537	Gene3D	1056	1158	4.20E-28	IPR013783	Immunoglobulin-like fold
orf88	1537	Gene3D	463	548	1.10E-24	IPR013780	Glycosyl hydrolase, all-beta
orf88	1537	TMHMM	35	57	-	-	Region predicted to be membrane-embedded
orf88	1537	Pfam	105	291	3.40E-21	IPR006047	GH13, catalytic domain
orf88	1537	SUPERFAMILY	1264	1337	8.63E-14	IPR008964	Invasin/intimin cell-adhesion fragments
orf88	1537	Pfam	1345	1388	3.40E-09	IPR001119	S-layer homology domain
orf88	1537	Pfam	989	1052	2.50E-16	IPR003343	Bacterial Ig-like domain, group 2

Table S5 Ig-like domain protein (WP_162288110.1) Ig-like domain-containing protein (GH13_28 + CBM25) and polysaccharide pyruvyl transferase protein which is involved in cell-wall anchoring of extracellular proteins.

ORF	Total size	Database/Tool	Start	Stop	E-value	ID	Protein domain
orf134	1251	Gene3D	425	509	6.10E-18	IPR013780	Glycosyl hydrolase, all-beta
orf134	1251	SMART	884	960	5.00E-10	IPR003343	Bacterial Ig-like domain, group 2
orf134	1251	SMART	795	872	2.60E-20	IPR003343	Bacterial Ig-like domain, group 2
orf134	1251	SMART	967	1049	7.20E-08	IPR003343	Bacterial Ig-like domain, group 2
orf134	1251	SMART	602	686	8.70E-10	IPR003343	Bacterial Ig-like domain, group 2
orf134	1251	Gene3D	691	788	1.00E-28	IPR013783	Immunoglobulin-like fold
orf134	1251	SUPERFAMILY	432	509	4.63E-08	-	Glycosyl hydrolase domain
orf134	1251	PANTHER	69	490	1.30E-24	-	ALPHA-AMYLASE
orf134	1251	SUPERFAMILY	625	688	3.26E-15	IPR008964	Invasin/intimin cell-adhesion fragments
orf134	1251	Gene3D	74	424	#######	-	Glycosidases
orf134	1251	SUPERFAMILY	887	961	3.69E-10	IPR008964	Invasin/intimin cell-adhesion fragments
orf134	1251	Pfam	145	348	1.80E-13	IPR006047	GH13, catalytic domain
orf134	1251	SMART	696	774	2.60E-10	IPR005085	Carbohydrate binding module family 25
orf134	1251	SMART	79	419	5.00E-08	IPR006047	GH13, catalytic domain
orf134	1251	CDD	78	429	#######	-	AmyAc_bac1_AmyA
orf134	1251	SUPERFAMILY	793	873	1.96E-17	IPR008964	Invasin/intimin cell-adhesion fragments
orf134	1251	SUPERFAMILY	74	410	1.31E-45	IPR017853	Glycoside hydrolase superfamily
orf134	1251	Pfam	795	873	8.00E-18	IPR003343	Bacterial Ig-like domain, group 2
orf134	1251	PRINTS	241	252	1.90E-06	IPR006046	Alpha amylase
orf134	1251	Pfam	1129	1171	9.40E-05	IPR001119	S-layer homology domain
orf134	1251	Pfam	697	792	2.00E-20	IPR005085	Carbohydrate binding module family 25
orf134	1251	TMHMM	21	43	-	-	Region predicted to be membrane-embedded
orf524	723	Pfam	373	662	1.40E-27	IPR007345	Polysaccharide pyruvyl transferase

Table S6 Statistical comparisons made for metagenomic datasets				
Age group	**Substrate**	**Timepoint (h)**	**Statistical variables**	
Testing the effect of time on microbial changes all infant or adult samples				
Infant	Both	Both	0h	24h
Adult	Both	Both	0h	24h
Testing the effect of substrate at each timepoint in infant or adult samples				
Infant	Both	0	HylonVII	No substrate
Adult	HylonVII	0	HylonVII	No substrate
Infant	HylonVII	24	HylonVII	No substrate
Adult	HylonVII	24	HylonVII	No substrate
Testing the effect of age at 24h (No substrate + HylonVII, or HylonVII)				
Both	Both	24	Adult	Infant
Both	HylonVII	24	Adult	Infant
Testing the associations between taxa prevalence and SCFA production				
Adult	HylonVII	24	HylonVII	Acetate
Adult	HylonVII	24	HylonVII	Butyrate
Adult	HylonVII	24	HylonVII	Propionate
Infant	HylonVII	24	HylonVII	Acetate
Infant	HylonVII	24	HylonVII	Butyrate
Infant	HylonVII	24	HylonVII	Propionate
Testing the effect of bacterial supplements against a no supplement control				
Adult	HylonVII	0	*B. infantis* LH277	No supplement
Adult	HylonVII	24	*B. infantis* LH277	No supplement
Adult	HylonVII	0	*B. pseudolongum* 45	No supplement
Adult	HylonVII	24	*B. pseudolongum* 45	No supplement
Adult	HylonVII	0	Combination	No supplement
Adult	HylonVII	24	Combination	No supplement
Infant	HylonVII	0	*B. infantis* LH277	No supplement
Infant	HylonVII	24	*B. infantis* LH277	No supplement
Infant	HylonVII	0	*B. pseudolongum* 45	No supplement
Infant	HylonVII	24	*B. pseudolongum* 45	No supplement

Infant	HylonVII	0	Combination	No supplement
Infant	HylonVII	24	Combination	No supplement
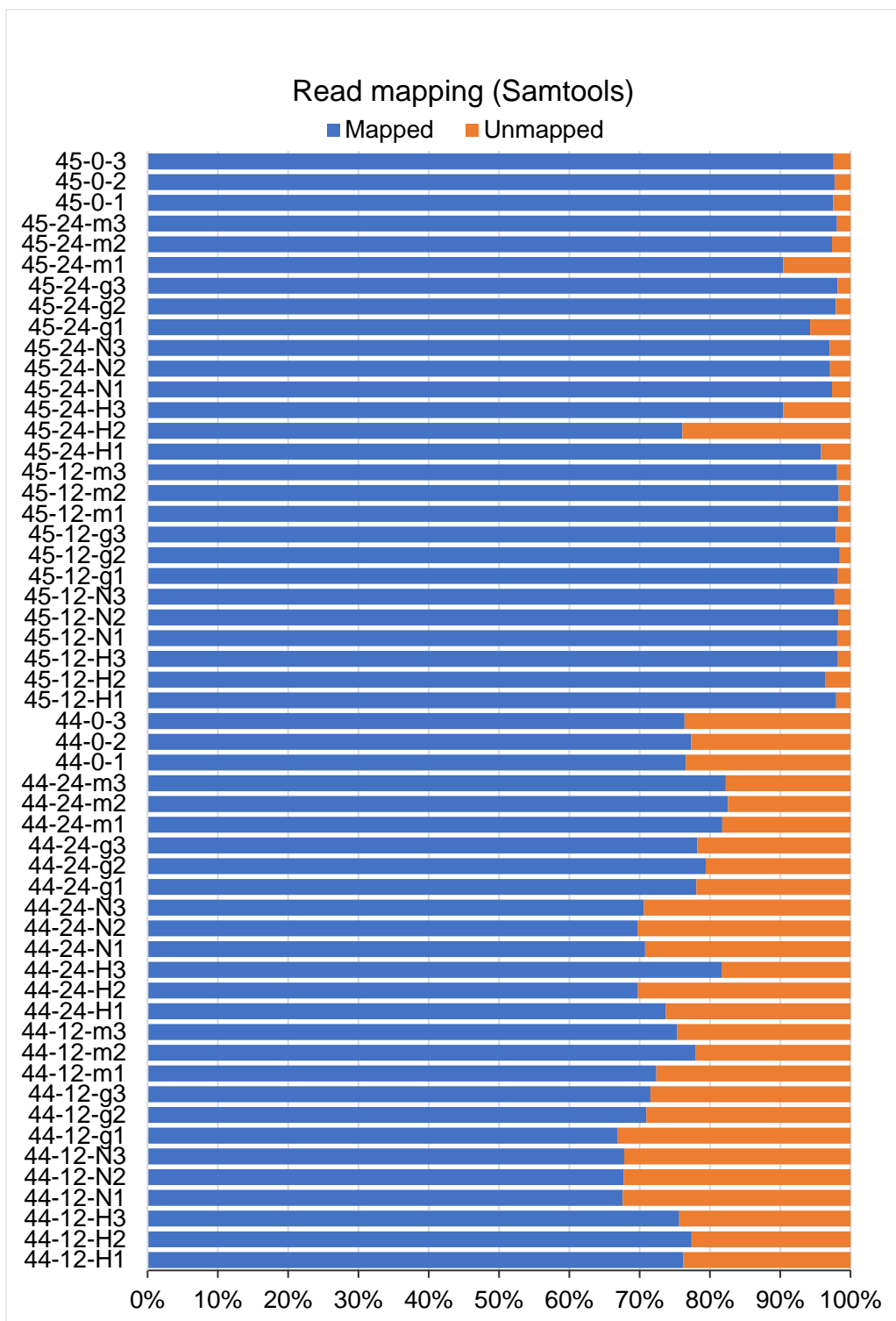
Figure S1 The % of reads from RNA-seq which were successfully mapped to the reference genome

References

Adey, A. C. (2021). "Tagmentation-based single-cell genomics." <u>Genome Research</u> **31**(10): 1693-1705.

Adiotomre, J., M. A. Eastwood, C. Edwards and W. G. Brydon (1990). "Dietary fiber: in vitro methods that anticipate nutrition and metabolic activity in humans." <u>The American journal of clinical nutrition</u> **52**(1): 128-134.

Adolph, T. E. and J. Zhang (2022). "Diet fuelling inflammatory bowel diseases: Preclinical and clinical concepts." <u>Gut</u> **71**(12): 2574-2586.

Aguirre, M., A. Eck, M. E. Koenen, P. H. Savelkoul, A. E. Budding and K. Venema (2016). "Diet drives quick changes in the metabolic activity and composition of human gut microbiota in a validated in vitro gut model." <u>Research in microbiology</u> **167**(2): 114-125.

Alcon-Giner, C., M. J. Dalby, S. Caim, J. Ketskemety, A. Shaw, K. Sim, M. A. Lawson, R. Kiu, C. Leclaire and L. Chalklen (2020). "Microbiota supplementation with Bifidobacterium and Lactobacillus modifies the preterm infant gut microbiota and metabolome: an observational study." <u>Cell Reports Medicine</u> **1**(5): 100077.

AlFaleh, K. and J. Anabrees (2014). "Probiotics for prevention of necrotizing enterocolitis in preterm infants." <u>Evidence-Based Child Health: A Cochrane Review Journal</u> **9**(3): 584-671.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." <u>Journal of molecular biology</u> **215**(3): 403-410.

Alves, V. D., C. M. Fontes and P. Bule (2021). "Cellulosomes: highly efficient cellulolytic complexes." <u>Macromolecular Protein Complexes III: Structure and Function</u>: 323-354.

Amaretti, A., T. Bernardi, E. Tamburini, S. Zanoni, M. Lomma, D. Matteuzzi and M. Rossi (2007). "Kinetics and metabolism of Bifidobacterium adolescentis MB 239 growing on glucose, galactose, lactose, and galactooligosaccharides." <u>Applied and Environmental Microbiology</u> **73**(11): 3637-3644.

Ambardar, S., R. Gupta, D. Trakroo, R. Lal and J. Vakhlu (2016). "High Throughput Sequencing: An Overview of Sequencing Chemistry." <u>Indian Journal of Microbiology</u> **56**(4): 394-404.

Arboleya, S., C. Watkins, C. Stanton and R. P. Ross (2016). "Gut bifidobacteria populations in human health and aging." <u>Frontiers in microbiology</u> **7**: 1204.

Arnal, G., D. W. Cockburn, H. Brumer and N. M. Koropatkin (2018). "Structural basis for the flexible recognition of α-glucan substrates by Bacteroides thetaiotaomicron SusG." Protein Science **27**(6): 1093-1101.

Atkinson, F. S., D. Hancock, P. Petocz and J. C. Brand-Miller (2018). "The physiologic and phenotypic significance of variation in human amylase gene copy number." The American journal of clinical nutrition **108**(4): 737-748.

Ausland, C., J. Zheng, H. Yi, B. Yang, T. Li, X. Feng, B. Zheng and Y. Yin (2021). "dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates." Nucleic Acids Research **49**(D1): D523-D528.

Baker, D. J., A. Aydin, T. Le-Viet, G. L. Kay, S. Rudder, L. de Oliveira Martins, A. P. Tedim, A. Kolyva, M. Diaz and N.-F. Alikhan (2021). "CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes." Genome medicine **13**(1): 1-11.

Ball, S., H.-P. Guan, M. James, A. Myers, P. Keeling, G. Mouille, A. Buléon, P. Colonna and J. Preiss (1996). "From Glycogen to Amylopectin: A Model for the Biogenesis of the Plant Starch Granule." Cell **86**(3): 349-352.

Bashiardes, S., G. Zilberman-Schapira and E. Elinav (2016). "Use of metatranscriptomics in microbiome research." Bioinformatics and biology insights **10**: BBI. S34610.

Bateman, A., M. T. Holden and C. Yeats (2005). "The G5 domain: a potential N-acetylglucosamine recognition domain involved in biofilm formation." Bioinformatics **21**(8): 1301-1303.

Bautista, V., J. Esclapez, F. Pérez-Pomares, R. M. Martínez-Espinosa, M. Camacho and M. J. Bonete (2012). "Cyclodextrin glycosyltransferase: a key enzyme in the assimilation of starch by the halophilic archaeon Haloferax mediterranei." Extremophiles **16**: 147-159.

Baxter, N. T., A. W. Schmidt, A. Venkataraman, K. S. Kim, C. Waldron and T. M. Schmidt (2019). "Dynamics of human gut microbiota and short-chain fatty acids in response to dietary interventions with three fermentable fibers." MBio **10**(1): e02566-02518.

Beck, L. C., A. C. Masi, G. R. Young, T. Vatanen, C. A. Lamb, R. Smith, J. Coxhead, A. Butler, B. J. Marsland and N. D. Embleton (2022). "Strain-specific impacts of probiotics are a significant driver of gut microbiome development in very preterm infants." Nature Microbiology **7**(10): 1525-1535.

Belenguer, A., S. H. Duncan, A. G. Calder, G. Holtrop, P. Louis, G. E. Lobley and H. J. Flint (2006). "Two Routes of Metabolic Cross-Feeding between

Bifidobacterium adolescentis and Butyrate-Producing Anaerobes from the Human Gut." Applied and Environmental Microbiology **72**(5): 3593.

Belenguer, A., S. H. Duncan, A. G. Calder, G. Holtrop, P. Louis, G. E. Lobley and H. J. Flint (2006). "Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut." Applied and environmental microbiology **72**(5): 3593-3599.

Bell, T. H. and T. Bell (2021). "Many roads to bacterial generalism." FEMS microbiology ecology **97**(1): fiaa240.

Bendiks, Z. A., J. Guice, D. Coulon, A. M. Raggio, R. C. Page, D. G. Carvajal-Aldaz, M. Luo, D. A. Welsh, B. D. Marx and C. M. Taylor (2022). "Resistant starch type 2 and whole grain maize flours enrich different intestinal bacteria and metatranscriptomes." Journal of Functional Foods **90**: 104982.

Bendiks, Z. A., K. E. Knudsen, M. J. Keenan and M. L. Marco (2020). "Conserved and variable responses of the gut microbiome to resistant starch type 2." Nutrition Research **77**: 12-28.

Benjamini, Y. and D. Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency." Annals of statistics: 1165-1188.

Bergström, A., T. H. Skov, M. I. Bahl, H. M. Roager, L. B. Christensen, K. T. Ejlerskov, C. Mølgaard, K. F. Michaelsen and T. R. Licht (2014). "Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants." Applied and environmental microbiology **80**(9): 2889-2900.

Bertolini, A. (2009). Starches: characterization, properties, and applications, CRC Press.

Bhattacharjee, D., C. Flores, C. Woelfel-Monsivais and A. M. Seekatz (2022). "Diversity and prevalence of Clostridium innocuum in the human gut microbiota." Msphere: e00569-00522.

Bhattacharya, T., T. S. Ghosh and S. S. Mande (2015). "Global profiling of carbohydrate active enzymes in human gut microbiome." PloS one **10**(11): e0142038.

Boekhorst, J., N. Venlet, N. Procházková, M. L. Hansen, C. B. Lieberoth, M. I. Bahl, L. Lauritzen, O. Pedersen, T. R. Licht and M. Kleerebezem (2022). "Stool energy density is positively correlated to intestinal transit time and related to microbial enterotypes." Microbiome **10**(1): 1-10.

Bourlieu, C., O. Ménard, K. Bouzerzour, G. Mandalari, A. Macierzanka, A. R. Mackie and D. Dupont (2014). "Specificity of infant digestive conditions: some clues for developing relevant in vitro models." Critical reviews in food science and nutrition **54**(11): 1427-1457.

Brighenti, F., M. C. Casiraghi and C. Baggio (1998). "Resistant starch in the Italian diet." British journal of nutrition **80**(4): 333-341.

Brooks, A. W., K. D. Kohl, R. M. Brucker, E. J. van Opstal and S. R. Bordenstein (2016). "Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History." PLOS Biology **14**(11): e2000225.

Brugger, S. D., C. Baumberger, M. Jost, W. Jenni, U. Brugger and K. Mühlemann (2012). "Automated counting of bacterial colony forming units on agar plates." PloS one **7**(3): e33695.

Bruinsma, S., J. Burgess, D. Schlingman, A. Czyz, N. Morrell, C. Ballenger, H. Meinholz, L. Brady, A. Khanna and L. Freeberg (2018). "Bead-linked transposomes enable a normalization-free workflow for NGS library preparation." BMC genomics **19**(1): 1-16.

Buchfink, B., C. Xie and D. H. Huson (2015). "Fast and sensitive protein alignment using DIAMOND." Nature methods **12**(1): 59-60.

Buléon, A., P. Colonna, V. Planchot and S. Ball (1998). "Starch granules: structure and biosynthesis." International Journal of Biological Macromolecules **23**(2): 85-112.

Bultynck, G., S. Kiviluoto and A. Methner (2014). "Bax inhibitor-1 is likely a pH-sensitive calcium leak channel, not a H+/Ca2+ exchanger." Science signaling **7**(343): pe22-pe22.

Burcelin, R., P. Gourdy and S. Dalle (2014). "GLP-1-based strategies: a physiological analysis of differential mode of action." Physiology **29**(2): 108-121.

Butterworth, P. J., F. J. Warren and P. R. Ellis (2011). "Human α-amylase and starch digestion: An interesting marriage." Starch-Stärke **63**(7): 395-405.

Button, J. E., C. A. Autran, A. L. Reens, C. M. Cosetta, S. Smriga, M. Ericson, J. V. Pierce, D. N. Cook, M. L. Lee and A. K. Sun (2022). "Dosing a synbiotic of human milk oligosaccharides and B. infantis leads to reversible engraftment in healthy adult microbiomes without antibiotics." Cell Host & Microbe **30**(5): 712-725. e717.

Camilleri, á., K. Madsen, R. Spiller, B. Van Meerveld and G. Verne (2012). "Intestinal barrier function in health and gastrointestinal disease." Neurogastroenterology & Motility **24**(6): 503-512.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics." Nucleic acids research **37**(suppl_1): D233-D238.

Carmody, R. N., J. E. Bisanz, B. P. Bowen, C. F. Maurice, S. Lyalina, K. B. Louie, D. Treen, K. S. Chadaideh, V. Maini Rekdal and E. N. Bess (2019). "Cooking shapes the structure and function of the gut microbiome." Nature microbiology **4**(12): 2052-2063.

Carmody, R. N., G. K. Gerber, J. M. Luevano Jr, D. M. Gatti, L. Somes, K. L. Svenson and P. J. Turnbaugh (2015). "Diet dominates host genotype in shaping the murine gut microbiota." Cell host & microbe **17**(1): 72-84.

Carmody, R. N. and R. W. Wrangham (2009). "The energetic significance of cooking." Journal of human evolution **57**(4): 379-391.

Carpenter, D., S. Dhar, L. M. Mitchell, B. Fu, J. Tyson, N. A. Shwan, F. Yang, M. G. Thomas and J. A. Armour (2015). "Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes." Human molecular genetics **24**(12): 3472-3480.

Cassidy, A., S. Bingham and J. Cummings (1994). "Starch intake and colorectal cancer risk: an international comparison." British journal of cancer **69**(5): 937-942.

Centanni, M., B. Lawley, C. A. Butts, N. C. Roy, J. Lee, W. J. Kelly and G. W. Tannock (2018). "Bifidobacterium pseudolongum in the ceca of rats fed Hi-Maize starch has characteristics of a keystone species in bifidobacterial blooms." Applied and environmental microbiology **84**(15): e00547-00518.

Cerqueira, F. (2022). Biochemical Features of Resistant Starch Degradation by Ruminococcus bromii.

Cerqueira, F. M., A. L. Photenhauer, H. L. Doden, A. N. Brown, A. M. Abdel-Hamid, S. Moraïs, E. A. Bayer, Z. Wawrzak, I. Cann and J. M. Ridlon (2022). "Sas20 is a highly flexible starch-binding protein in the Ruminococcus bromii cell-surface amylosome." Journal of Biological Chemistry **298**(5).

Cerqueira, F. M., A. L. Photenhauer, R. M. Pollet, H. A. Brown and N. M. Koropatkin (2020). "Starch Digestion by Gut Bacteria: Crowdsourcing for Carbs." Trends Microbiol **28**(2): 95-108.

Chen, L., R. Liu, C. Qin, Y. Meng, J. Zhang, Y. Wang and G. Xu (2010). "Sources and intake of resistant starch in the Chinese diet." Asia Pacific journal of clinical nutrition **19**(2): 274-282.

Chen, T., W. Long, C. Zhang, S. Liu, L. Zhao and B. R. Hamaker (2017). "Fiber-utilizing capacity varies in Prevotella-versus Bacteroides-dominated gut microbiota." Scientific reports **7**(1): 2594.

Cheng, K.-C., A. Demirci and J. M. Catchmark (2011). "Pullulan: biosynthesis, production, and applications." Applied microbiology and biotechnology **92**: 29-44.

Chung, W. S. F., A. W. Walker, P. Louis, J. Parkhill, J. Vermeiren, D. Bosscher, S. H. Duncan and H. J. Flint (2016). "Modulation of the human gut microbiota by dietary fibres occurs at the species level." BMC biology **14**(1): 1-13.

Chung, W. S. F., A. W. Walker, J. Vermeiren, P. O. Sheridan, D. Bosscher, V. Garcia-Campayo, J. Parkhill, H. J. Flint and S. H. Duncan (2019). "Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota." FEMS microbiology ecology **95**(1): fiy201.

Cockburn, D. W., R. Kibler, H. A. Brown, R. Duvall, S. Moraïs, E. Bayer and N. M. Koropatkin (2021). "Structure and substrate recognition by the Ruminococcus bromii amylosome pullulanases." Journal of Structural Biology **213**(3): 107765.

Cockburn, D. W., C. Suh, K. P. Medina, R. M. Duvall, Z. Wawrzak, B. Henrissat and N. M. Koropatkin (2018). "Novel carbohydrate binding modules in the surface anchored alpha-amylase of Eubacterium rectale provide a molecular rationale for the range of starches used by this organism in the human gut." Mol Microbiol **107**(2): 249-264.

Cockburn, D. W., C. Suh, K. P. Medina, R. M. Duvall, Z. Wawrzak, B. Henrissat and N. M. Koropatkin (2018). "Novel carbohydrate binding modules in the surface anchored α-amylase of Eubacterium rectale provide a molecular rationale for the range of starches used by this organism in the human gut." Molecular Microbiology **107**(2): 249-264.

Connell Motherway, M., G. F. Fitzgerald, S. Neirynck, S. Ryan, L. Steidler and D. van Sinderen (2008). "Characterization of ApuB, an Extracellular Type II Amylopullulanase from Bifidobacterium breve UCC2003." Applied and Environmental Microbiology **74**(20): 6271.

Costea, P. I., F. Hildebrand, M. Arumugam, F. Bäckhed, M. J. Blaser, F. D. Bushman, W. M. De Vos, S. D. Ehrlich, C. M. Fraser and M. Hattori (2018). "Enterotypes in the landscape of gut microbial community composition." Nature microbiology **3**(1): 8-16.

Cotter, P. D. and C. Hill (2003). "Surviving the acid test: responses of gram-positive bacteria to low pH." Microbiology and molecular biology reviews **67**(3): 429-453.

Crittenden, R., A. Laitila, P. Forssell, J. Matto, M. Saarela, T. Mattila-Sandholm and P. Myllarinen (2001). "Adhesion of bifidobacteria to granular starch and its implications in probiotic technologies." Applied and Environmental Microbiology **67**(8): 3469-3475.

Crost, E. H., G. Le Gall, J. A. Laverde-Gomez, I. Mukhopadhya, H. J. Flint and N. Juge (2018). "Mechanistic insights into the cross-feeding of Ruminococcus gnavus and Ruminococcus bromii on host and dietary carbohydrates." Frontiers in microbiology **9**: 2558.

Dave, R. and N. Shah (1996). "Evaluation of media for selective enumeration of Streptococcus thermophilus, Lactobacillus delbrueckii ssp. bulgaricus, Lactobacillus acidophilus, and bifidobacteria." Journal of dairy science **79**(9): 1529-1536.

David, L. A., C. F. Maurice, R. N. Carmody, D. B. Gootenberg, J. E. Button, B. E. Wolfe, A. V. Ling, A. S. Devlin, Y. Varma and M. A. Fischbach (2014). "Diet rapidly and reproducibly alters the human gut microbiome." Nature **505**(7484): 559-563.

De Silva, A. and S. R. Bloom (2012). "Gut hormones and appetite control: a focus on PYY and GLP-1 as therapeutic targets in obesity." Gut and liver **6**(1): 10.

De Vries, W. and A. Stouthamer (1967). "Pathway of glucose fermentation in relation to the taxonomy of bifidobacteria." Journal of bacteriology **93**(2): 574-576.

Deatherage, W., M. M. MacMasters and C. Rist (1955). "A partial survey of amylose content in starch from domestic and foreign varieties of corn, wheat and sorghum and from some other starch-bearing plants." Trans. Am. Assoc. Cereal Chem **13**(3): 13-42.

Deehan, E. C., C. Yang, M. E. Perez-Muñoz, N. K. Nguyen, C. C. Cheng, L. Triador, Z. Zhang, J. A. Bakal and J. Walter (2020). "Precision microbiome modulation with discrete dietary fiber structures directs short-chain fatty acid production." Cell Host & Microbe **27**(3): 389-404. e386.

DeMartino, P. and D. W. Cockburn (2020). "Resistant starch: impact on the gut microbiome and health." Current opinion in biotechnology **61**: 66-71.

Derrien, M., F. Turroni, M. Ventura and D. van Sinderen (2022). "Insights into endogenous Bifidobacterium species in the human gut microbiota during adulthood." Trends in Microbiology.

Devika, N. and K. Raman (2019). "Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models." Scientific Reports **9**(1): 1-9.

Di Gioia, D., I. Aloisio, G. Mazzola and B. Biavati (2014). "Bifidobacteria: their impact on gut microbiota composition and their applications as probiotics in infants." Applied microbiology and biotechnology **98**: 563-577.

Dobbler, P. T., R. S. Procianoy, V. Mai, R. C. Silveira, A. L. Corso, B. S. Rojas and L. F. Roesch (2017). "Low microbial diversity and abnormal microbial succession is associated with necrotizing enterocolitis in preterm infants." Frontiers in microbiology **8**: 2243.

Dobranowski, P. A. and A. Stintzi (2021). "Resistant starch, microbiome, and precision modulation." Gut Microbes **13**(1): 1926842.

Duguay, A. R. and T. J. Silhavy (2004). "Quality control in the bacterial periplasm." Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1694**(1-3): 121-134.

Duranti, S., G. Longhi, M. Ventura, D. van Sinderen and F. Turroni (2020). "Exploring the Ecology of Bifidobacteria and Their Genetic Adaptation to the Mammalian Gut." Microorganisms **9**(1).

Duranti, S., C. Milani, G. A. Lugli, L. Mancabelli, F. Turroni, C. Ferrario, M. Mangifesta, A. Viappiani, B. Sánchez, A. Margolles, D. van Sinderen and M. Ventura (2016). "Evaluation of genetic diversity among strains of the human gut commensal Bifidobacterium adolescentis." Scientific Reports **6**(1): 23971.

Duranti, S., F. Turroni, G. A. Lugli, C. Milani, A. Viappiani, M. Mangifesta, L. Gioiosa, P. Palanza, D. van Sinderen and M. Ventura (2014). "Genomic characterization and transcriptional studies of the starch-utilizing strain Bifidobacterium adolescentis 22L." Applied and Environmental Microbiology **80**(19): 6080-6090.

Eerlingen, R., M. Crombez and J. Delcour (1993). "Enzyme-resistant starch. 1. Quantitative and qualitative influence of incubation-time and temperature of autoclaved starch on resistant starch formation." Cereal chemistry **70**(3): 339-344.

Egan, M. and D. Van Sinderen (2018). Carbohydrate metabolism in Bifidobacteria. The Bifidobacteria and Related Organisms, Elsevier**:** 145-164.

Englyst, H. N. and G. J. Hudson (1996). "The classification and measurement of dietary carbohydrates." Food chemistry **57**(1): 15-21.

Englyst, H. N., S. M. Kingman and J. Cummings (1992). "Classification and measurement of nutritionally important starch fractions." European journal of clinical nutrition **46**: S33-50.

Falà, A. K., A. Álvarez-Ordóñez, A. Filloux, C. Gahan and P. D. Cotter (2022). "Quorum sensing in human gut and food microbiomes: Significance and potential for therapeutic targeting." Frontiers in Microbiology: 4389.

Fässler, C., E. Arrigoni, K. Venema, F. Brouns and R. Amadò (2006). "In vitro fermentability of differently digested resistant starch preparations." Molecular nutrition & food research **50**(12): 1220-1228.

Fässler, C., E. Arrigoni, K. Venema, V. Hafner, F. Brouns and R. Amadò (2006). "Digestibility of resistant starch containing preparations using two in vitro models." European journal of nutrition **45**(8): 445-453.

Fässler, C., C. I. Gill, E. Arrigoni, I. Rowland and R. Amado (2007). "Fermentation of resistant starches: influence of in vitro models on colon carcinogenesis." Nutrition and cancer **58**(1): 85-92.

Fattahi, Y., H. R. Heidari and A. Y. Khosroushahi (2020). "Review of short-chain fatty acids effects on the immune system and cancer." Food Bioscience **38**: 100793.

Favier, C. F., E. E. Vaughan, W. M. De Vos and A. D. Akkermans (2002). "Molecular monitoring of succession of bacterial communities in human neonates." Applied and environmental microbiology **68**(1): 219-226.

Ferraris, L., A. Balvay, D. Bellet, J. Delannoy, C. Maudet, T. Larcher, J.-C. Rozé, C. Philippe, T. Meylheuc and M.-J. Butel (2023). "Neonatal necrotizing enterocolitis: Clostridium butyricum and Clostridium neonatale fermentation metabolism and enteropathogenicity." Gut Microbes **15**(1): 2172666.

Fijan, S. (2014). "Microorganisms with claimed probiotic properties: an overview of recent literature." International journal of environmental research and public health **11**(5): 4745-4767.

Finn, R. D., J. Clements and S. R. Eddy (2011). "HMMER web server: interactive sequence similarity searching." Nucleic acids research **39**(suppl_2): W29-W37.

Flint, H. J., K. P. Scott, S. H. Duncan, P. Louis and E. Forano (2012). "Microbial degradation of complex carbohydrates in the gut." Gut microbes **3**(4): 289-306.

Foley, M. H., D. W. Cockburn and N. M. Koropatkin (2016). "The Sus operon: a model system for starch uptake by the human gut Bacteroidetes." Cellular and Molecular Life Sciences **73**(14): 2603-2617.

Foley, M. H., E. C. Martens and N. M. Koropatkin (2018). "SusE facilitates starch uptake independent of starch binding in B. thetaiotaomicron." Molecular microbiology **108**(5): 551-566.

Frost, G., M. L. Sleeth, M. Sahuri-Arisoylu, B. Lizarbe, S. Cerdan, L. Brody, J. Anastasovska, S. Ghourab, M. Hankir, S. Zhang, D. Carling, J. R. Swann, G. Gibson, A. Viardot, D. Morrison, E. Louise Thomas and J. D. Bell (2014). "The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism." Nature Communications **5**(1): 3611.

Fuentes-Zaragoza, E., M. Riquelme-Navarrete, E. Sánchez-Zapata and J. Pérez-Álvarez (2010). "Resistant starch as functional ingredient: A review." Food Research International **43**(4): 931-942.

Fujimoto, Z., A. Jackson, M. Michikawa, T. Maehara, M. Momma, B. Henrissat, H. J. Gilbert and S. Kaneko (2013). "The structure of a Streptomyces avermitilis α-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement." Journal of Biological Chemistry **288**(17): 12376-12385.

Fukuda, S., H. Toh, K. Hase, K. Oshima, Y. Nakanishi, K. Yoshimura, T. Tobe, J. M. Clarke, D. L. Topping and T. Suzuki (2011). "Bifidobacteria can protect from enteropathogenic infection through production of acetate." Nature **469**(7331): 543-547.

Fuller, C. W., L. R. Middendorf, S. A. Benner, G. M. Church, T. Harris, X. Huang, S. B. Jovanovich, J. R. Nelson, J. A. Schloss, D. C. Schwartz and D. V. Vezenov (2009). "The challenges of sequencing by synthesis." Nature Biotechnology **27**(11): 1013-1023.

Fung, D. K., E. W. Chan, M. L. Chin and R. C. Chan (2010). "Delineation of a bacterial starvation stress response network which can mediate antibiotic tolerance development." Antimicrobial agents and chemotherapy **54**(3): 1082-1093.

Fushinobu, S. (2010). "Unique sugar metabolic pathways of bifidobacteria." Bioscience, biotechnology, and biochemistry **74**(12): 2374-2384.

Garrido, D., D. C. Dallas and D. A. Mills (2013). "Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications." Microbiology **159**(Pt 4): 649.

Ge, S. X., E. W. Son and R. Yao (2018). "iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data." BMC bioinformatics **19**(1): 1-24.

Gidley, M. J. and S. M. Bociek (1985). "Molecular organization in starches: A carbon 13 CP/MAS NMR study." Journal of the American Chemical Society **107**(24): 7040-7044.

Gilbert, R. G. (2011). "Size-separation characterization of starch and glycogen for biosynthesis–structure–property relationships." Analytical and bioanalytical chemistry **399**(4): 1425-1438.

Gilbert, R. G., T. Witt and J. Hasjim (2013). "What is being learned about starch properties from multiple-level characterization." Cereal Chemistry **90**(4): 312-325.

Gomez de Agüero, M., S. C. Ganal-Vonarburg, T. Fuhrer, S. Rupp, Y. Uchimura, H. Li, A. Steinert, M. Heikenwalder, S. Hapfelmeier and U. Sauer (2016). "The maternal microbiota drives early postnatal innate immune development." Science **351**(6279): 1296-1302.

Goosens, V. J. and J. M. van Dijl (2017). "Twin-arginine protein translocation." Protein and Sugar Export and Assembly in Gram-positive Bacteria: 69-94.

Gopalsamy, G., E. Mortimer, P. Greenfield, A. R. Bird, G. P. Young and C. T. Christophersen (2019). "Resistant starch is actively fermented by infant faecal microbiota and increases microbial diversity." Nutrients **11**(6): 1345.

Gous, P. W., F. Warren, R. Gilbert and G. P. Fox (2017). "Drought-Proofing Barley (Hordeum vulgare): The Effects of Stay Green on Starch and Amylose Structure." Cereal Chemistry **94**(5): 873-880.

Guerra, A., L. Etienne-Mesmin, V. Livrelli, S. Denis, S. Blanquet-Diot and M. Alric (2012). "Relevance and challenges in modeling human gastric and small intestinal digestion." Trends in biotechnology **30**(11): 591-600.

Gunn, D., Z. Abbas, H. C. Harris, G. Major, C. Hoad, P. Gowland, L. Marciani, S. K. Gill, F. J. Warren and M. Rossi (2022). "Psyllium reduces inulin-induced colonic gas production in IBS: MRI and in vitro fermentation studies." Gut **71**(5): 919-927.

Gunn, D., R. Murthy, G. Major, V. Wilkinson-Smith, C. Hoad, L. Marciani, J. Remes-Troche, S. Gill, M. Rossi, H. Harris, J. Ahn-Jarvis, F. Warren, K. Whelan and R. Spiller (2020). "Contrasting effects of viscous and particulate fibers on colonic fermentation in vitro and in vivo, and their impact on intestinal water studied by MRI in a randomized trial." The American Journal of Clinical Nutrition **112**(3): 595-602.

Guo, P., K. Zhang, X. Ma and P. He (2020). "Clostridium species as probiotics: potentials and challenges." Journal of animal science and biotechnology **11**(1): 1-10.

Haenen, D., J. Zhang, C. Souza da Silva, G. Bosch, I. M. van der Meer, J. van Arkel, J. J. van den Borne, O. Pérez Gutiérrez, H. Smidt and B. Kemp (2013). "A diet high in resistant starch modulates microbiota composition, SCFA concentrations, and gene expression in pig intestine." The Journal of nutrition **143**(3): 274-283.

Hardy, K., T. Blakeney, L. Copeland, J. Kirkham, R. Wrangham and M. Collins (2009). "Starch granules, dental calculus and new perspectives on ancient diet." Journal of Archaeological Science **36**(2): 248-255.

Hardy, K., J. Brand-Miller, K. D. Brown, M. G. Thomas and L. Copeland (2015). "The importance of dietary carbohydrate in human evolution." The Quarterly review of biology **90**(3): 251-268.

Hawker, C. D., J. R. Genzen and C. T. Wittwer (2018). "Automation in the clinical laboratory." Tietz Textbook of Clinical Chemistry and Molecular Diagnostics: 370. e371-370. e324.

Herath, M., S. Hosie, J. C. Bornstein, A. E. Franks and E. L. Hill-Yardin (2020). "The role of the gastrointestinal mucus system in intestinal homeostasis: implications for neurological disorders." Frontiers in Cellular and Infection Microbiology: 248.

Hill, C., F. Guarner, G. Reid, G. R. Gibson, D. J. Merenstein, B. Pot, L. Morelli, R. B. Canani, H. J. Flint and S. Salminen (2014). "Expert consensus document: The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic." Nature reviews Gastroenterology & hepatology.

Hinz, S. W., L. A. van den Broek, G. Beldman, J.-P. Vincken and A. G. Voragen (2004). "β-Galactosidase from Bifidobacterium adolescentis DSM20083 prefers β (1, 4)-galactosides over lactose." Applied microbiology and biotechnology **66**: 276-284.

Holzapfel, W., A. Arini, M. Aeschbacher, R. Coppolecchia and B. Pot (2018). "Enterococcus faecium SF68 as a model for efficacy and safety evaluation of pharmaceutical probiotics." Beneficial microbes **9**(3): 375-388.

Hosseini-Giv, N., A. Basas, C. Hicks, E. El-Omar, F. El-Assaad and E. Hosseini-Beheshti (2022). "Bacterial extracellular vesicles and their novel therapeutic applications in health and cancer." Frontiers in Cellular and Infection Microbiology: 1661.

Huang, L., H. Zhang, P. Wu, S. Entwistle, X. Li, T. Yohe, H. Yi, Z. Yang and Y. Yin (2018). "dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation." Nucleic Acids Research **46**(D1): D516-D521.

Hunt, M., N. D. Silva, T. D. Otto, J. Parkhill, J. A. Keane and S. R. Harris (2015). "Circlator: automated circularization of genome assemblies using long sequencing reads." Genome biology **16**(1): 1-10.

Ibrahim, O. O. (2019). "Classification of antimicrobial peptides bacteriocins, and the nature of some bacteriocins with potential applications in food safety and bio-pharmaceuticals." EC Microbiol **15**(7): 591-608.

Iyer, L. M., D. D. Leipe, E. V. Koonin and L. Aravind (2004). "Evolutionary history and higher order classification of AAA+ ATPases." Journal of structural biology **146**(1-2): 11-31.

Jain, M., H. E. Olsen, B. Paten and M. Akeson (2016). "The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community." Genome Biology **17**(1): 239.

James, K., M. O'Connell Motherway, C. Penno, R. L. O'Brien and D. van Sinderen (2018). "Bifidobacterium breve UCC2003 Employs Multiple Transcriptional Regulators To Control Metabolism of Particular Human Milk Oligosaccharides." Appl Environ Microbiol **84**(9).

Janeček, Š., F. Mareček, E. A. MacGregor and B. Svensson (2019). "Starch-binding domains as CBM families–history, occurrence, structure, function and evolution." Biotechnology Advances **37**(8): 107451.

Jeffery, I. B., M. J. Claesson, P. W. O'toole and F. Shanahan (2012). "Categorization of the gut microbiota: enterotypes or gradients?" Nature Reviews Microbiology **10**(9): 591-592.

Jenkins, D. J., T. Wolever, R. H. Taylor, H. Barker, H. Fielden, J. M. Baldwin, A. C. Bowling, H. C. Newman, A. L. Jenkins and D. V. Goff (1981). "Glycemic index of foods: a physiological basis for carbohydrate exchange." The American journal of clinical nutrition **34**(3): 362-366.

Jenkins, P. and A. Donald (1995). "The influence of amylose on starch granule structure." International journal of biological macromolecules **17**(6): 315-321.

Johansson, M. E., D. Ambort, T. Pelaseyed, A. Schütte, J. K. Gustafsson, A. Ermund, D. B. Subramani, J. M. Holmén-Larsson, K. A. Thomsson and J. H. Bergström (2011). "Composition and functional role of the mucus layers in the intestine." Cellular and molecular life sciences **68**(22): 3635-3641.

Jones, M., H. Hunt, E. Lightfoot, D. Lister, X. Liu and G. Motuzaite-Matuzeviciute (2011). "Food globalization in prehistory." World archaeology **43**(4): 665-675.

Joyce, M. A., M. E. Fraser, E. R. Brownie, M. N. James, W. A. Bridger and W. T. Wolodko (1999). "Probing the nucleotide-binding site of Escherichia coli succinyl-CoA synthetase." Biochemistry **38**(22): 7273-7283.

Jung, D.-H., G.-Y. Kim, I.-Y. Kim, D.-H. Seo, Y.-D. Nam, H. Kang, Y. Song and C.-S. Park (2019). "Bifidobacterium adolescentis P2P3, a human gut bacterium having strong non-gelatinized resistant starch-degrading activity."

Jung, D.-H. and C.-S. Park (2023). "Resistant starch utilization by Bifidobacterium, the beneficial human gut bacteria." Food Science and Biotechnology: 1-12.

Jung, D.-H., D.-H. Seo, Y.-J. Kim, W.-H. Chung, Y.-D. Nam and C.-S. Park (2020). "The presence of resistant starch-degrading amylases in Bifidobacterium adolescentis of the human gut." International Journal of Biological Macromolecules **161**: 389-397.

Kaoutari, A. E., F. Armougom, J. I. Gordon, D. Raoult and B. Henrissat (2013). "The abundance and variety of carbohydrate-active enzymes in the human gut microbiota." Nature Reviews Microbiology **11**(7): 497-504.

Keegstra, J. M., F. Carrara and R. Stocker (2022). "The ecological roles of bacterial chemotaxis." Nature Reviews Microbiology **20**(8): 491-504.

Keeler, J. (2010). Understanding NMR spectroscopy, John Wiley & Sons.

Kelly, G., S. Prasannan, S. Daniell, K. Fleming, G. Frankel, G. Dougan, I. Connerton and S. Matthews (1999). "Structure of the cell-adhesion fragment of intimin from enteropathogenic Escherichia coli." Nature structural biology **6**(4): 313-318.

Kelly, S. M., J. Munoz-Munoz and D. Van Sinderen (2021). "Plant glycan metabolism by bifidobacteria." Frontiers in Microbiology **12**: 609418.

Kennedy, K. M., M. C. de Goffau, M. E. Perez-Muñoz, M.-C. Arrieta, F. Bäckhed, P. Bork, T. Braun, F. D. Bushman, J. Dore and W. M. de Vos (2023). "Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies." Nature **613**(7945): 639-649.

Khoroshkin, M. S., S. A. Leyn, D. Van Sinderen and D. A. Rodionov (2016). "Transcriptional regulation of carbohydrate utilization pathways in the Bifidobacterium genus." Frontiers in microbiology **7**: 120.

Killinger, B. J., C. Whidbey, N. C. Sadler, A. J. DeLeon, N. Munoz, Y.-M. Kim and A. T. Wright (2022). "Activity-based protein profiling identifies alternating activation of enzymes involved in the bifidobacterium shunt pathway or mucin degradation in

the gut microbiome response to soluble dietary fiber." npj Biofilms and Microbiomes **8**(1): 1-10.

Kim, H., Y. Jeong, S. Kang, H. J. You and G. E. Ji (2020). "Co-culture with Bifidobacterium catenulatum improves the growth, gut colonization, and butyrate production of Faecalibacterium prausnitzii: in vitro and in vivo studies." Microorganisms **8**(5): 788.

Kim, H. J., S. I. Shin, S. J. Lee, T. W. Moon and C. J. Lee (2018). "Screening and selection of Bifidobacterium strains isolated from human feces capable of utilizing resistant starch." Journal of the Science of Food and Agriculture **98**(15): 5901-5907.

Kim, S.-Y., H. Kim, Y.-J. Kim, D.-H. Jung, D.-H. Seo, J.-H. Jung and C.-S. Park (2021). "Enzymatic analysis of truncation mutants of a type II pullulanase from Bifidobacterium adolescentis P2P3, a resistant starch-degrading gut bacterium." International Journal of Biological Macromolecules **193**: 1340-1349.

Kiu, R., S. Caim, S. Alexander, P. Pachori and L. J. Hall (2017). "Probing genomic aspects of the multi-host pathogen Clostridium perfringens reveals significant pangenome diversity, and a diverse array of virulence factors." Frontiers in microbiology **8**: 2485.

Kiu, R. and L. J. Hall (2018). "An update on the human and animal enteric pathogen Clostridium perfringens." Emerging microbes & infections **7**(1): 1-15.

Kiu, R., A. Shaw, K. Sim, H. Bedwell, E. Cornwell, D. Pickard, G. Belteki, J. Malsom, S. Philips and G. R. Young (2021). "Dissemination and pathogenesis of toxigenic Clostridium perfringens strains linked to neonatal intensive care units and Necrotising Enterocolitis." bioRxiv: 2021.2008. 2003.454877.

Klimenko, N. S., V. E. Odintsova, A. Revel-Muroz and A. V. Tyakht (2022). "The hallmarks of dietary intervention-resilient gut microbiome." npj Biofilms and Microbiomes **8**(1): 77.

Koev, T. T., H. C. Harris, S. Kiamehr, Y. Z. Khimyak and F. J. Warren (2022). "Starch hydrogels as targeted colonic drug delivery vehicles." Carbohydrate Polymers **289**: 119413.

Koev, T. T., J. C. Muñoz-García, D. Iuga, Y. Z. Khimyak and F. J. Warren (2020). "Structural heterogeneities in starch hydrogels." Carbohydrate Polymers **249**: 116834.

Koh, A., F. De Vadder, P. Kovatcheva-Datchary and F. Bäckhed (2016). "From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites." Cell **165**(6): 1332-1345.

Koskella, B., L. J. Hall and C. J. E. Metcalf (2017). "The microbiome beyond the horizon of ecological and evolutionary theory." <u>Nature Ecology & Evolution</u> **1**(11): 1606-1615.

Krawczyk, B., P. Wityk, M. Gałęcka and M. Michalik (2021). "The many faces of Enterococcus spp.—commensal, probiotic and opportunistic pathogen." <u>Microorganisms</u> **9**(9): 1900.

Krishna, K. V., K. Koujalagi, R. U. Surya, M. Namratha and A. Malaviya (2022). "Enterococcus species and their probiotic potential: Current status and future prospects." <u>Journal of Applied Biology and Biotechnology</u> **11**(1): 36-44.

Kujawska, M., S. L. La Rosa, L. C. Roger, P. B. Pope, L. Hoyles, A. L. McCartney and L. J. Hall (2020). "Succession of Bifidobacterium longum strains in response to a changing early life nutritional environment reveals dietary substrate adaptations." <u>Iscience</u> **23**(8): 101368.

Kujawska, M., A. Raulo, M. Millar, F. Warren, L. Baltrūnaitė, S. C. Knowles and L. J. Hall (2022). "Bifidobacterium castoris strains isolated from wild mice show evidence of frequent host switching and diverse carbohydrate metabolism potential." <u>ISME Communications</u> **2**(1): 1-14.

Kyrpides, N. C., P. Hugenholtz, J. A. Eisen, T. Woyke, M. Göker, C. T. Parker, R. Amann, B. J. Beck, P. S. Chain and J. Chun (2014). "Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains." <u>PLoS biology</u> **12**(8): e1001920.

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

Lauder, E., K. Kim, T. M. Schmidt and J. L. Golob (2020). "Organoid-derived adult human colonic epithelium responds to co-culture with a probiotic strain of Bifidobacterium longum." <u>BioRxiv</u>: 2020.2007. 2016.207852.

Lawley, T. D. and A. W. Walker (2013). "Intestinal colonization resistance." <u>Immunology</u> **138**(1): 1-11.

Lawson, M. A. E., I. J. O'Neill, M. Kujawska, S. Gowrinadh Javvadi, A. Wijeyesekera, Z. Flegg, L. Chalklen and L. J. Hall (2020). "Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem." <u>The ISME Journal</u> **14**(2): 635-648.

Lawson, P. A., L. Saavedra Perez and K. Sankaranarayanan (2023). "Reclassification of Clostridium cocleatum, Clostridium ramosum, Clostridium spiroforme and Clostridium saccharogumia as Thomasclavelia cocleata gen. nov., comb. nov., Thomasclavelia ramosa comb. nov., gen. nov., Thomasclavelia spiroformis comb. nov. and Thomasclavelia saccharogumia comb. nov." International Journal of Systematic and Evolutionary Microbiology **73**(1): 005694.

Le Leu, R. K., Y. Hu, I. L. Brown, R. J. Woodman and G. P. Young (2010). "Synbiotic intervention of Bifidobacterium lactis and resistant starch protects against colorectal cancer development in rats." Carcinogenesis **31**(2): 246-251.

Lee, B. H., L. A. Bello-Pérez, A. H. M. Lin, C. Y. Kim and B. R. Hamaker (2013). "Importance of location of digestion and colonic fermentation of starch related to its quality." Cereal chemistry **90**(4): 335-343.

Leppänen, V.-M., M. C. Merckel, D. L. Ollis, K. K. Wong, J. W. Kozarich and A. Goldman (1999). "Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase." Structure **7**(7): 733-744.

Levy, S. E. and R. M. Myers (2016). "Advancements in next-generation sequencing." Annu Rev Genomics Hum Genet **17**(1): 95-115.

Li, Q., J. Loponen and M. G. Ganzle (2020). "Characterization of the extracellular fructanase FruA in Lactobacillus crispatus and its contribution to fructan hydrolysis in breadmaking." Journal of Agricultural and Food Chemistry **68**(32): 8637-8647.

Li, Y. O. and A. R. Komarek (2017). "Dietary fibre basics: Health, nutrition, analysis, and applications." Food quality and safety **1**(1): 47-59.

Liang, T., X. Xie, L. Wu, L. Li, L. Yang, T. Jiang, M. Du, M. Chen, L. Xue and J. Zhang (2023). "Metabolism of resistant starch RS3 administered in combination with Lactiplantibacillus plantarum strain 84-3 by human gut microbiota in simulated fermentation experiments in vitro and in a rat model." Food Chemistry: 135412.

Lin, A. H. M. and B. L. Nichols (2017). "The digestion of complementary feeding starches in the young child." Starch-Stärke **69**(7-8): 1700012.

Liu, F., Z. Ahmed, E. A. Lee, E. Donner, Q. Liu, R. Ahmed, M. K. Morell, M. J. Emes and I. J. Tetlow (2012). "Allelic variants of the amylose extender mutation of maize demonstrate phenotypic variation in starch structure resulting from modified protein–protein interactions." Journal of Experimental Botany **63**(3): 1167-1183.

Liu, F., B. Bomfleur, H. Peng, Q. Li, H. Kerp and H. Zhu (2018). "280-my-old fossil starch reveals early plant–animal mutualism." Geology **46**(5): 423-426.

Liu, J., J. Wang, Y. Zhou, H. Han, W. Liu, D. Li, F. Li, D. Cao and Q. Lei (2022). "Integrated omics analysis reveals differences in gut microbiota and gut-host metabolite profiles between obese and lean chickens." Poultry Science **101**(11): 102165.

Liyana-Pathirana, C. M. and F. Shahidi (2007). "The antioxidant potential of milling fractions from breadwheat and durum." Journal of Cereal Science **45**(3): 238-247.

Lo Leggio, L., T. J. Simmons, J.-C. N. Poulsen, K. E. Frandsen, G. R. Hemsworth, M. A. Stringer, P. Von Freiesleben, M. Tovborg, K. S. Johansen and L. De Maria (2015). "Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase." Nature communications **6**(1): 1-9.

Lordkipanidze, D., M. S. Ponce de León, A. Margvelashvili, Y. Rak, G. P. Rightmire, A. Vekua and C. P. Zollikofer (2013). "A complete skull from Dmanisi, Georgia, and the evolutionary biology of early Homo." Science **342**(6156): 326-331.

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

Louis, P., M. Solvang, S. H. Duncan, A. W. Walker and I. Mukhopadhya (2021). "Dietary fibre complexity and its influence on functional groups of the human gut microbiota." Proceedings of the Nutrition Society **80**(4): 386-397.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome biology **15**(12): 1-21.

Lowe, R., N. Shirley, M. Bleackley, S. Dolan and T. Shafee (2017). "Transcriptomics technologies." PLoS computational biology **13**(5): e1005457.

Lozupone, C. A., J. I. Stombaugh, J. I. Gordon, J. K. Jansson and R. Knight (2012). "Diversity, stability and resilience of the human gut microbiota." Nature **489**(7415): 220-230.

Lu, J., F. P. Breitwieser, P. Thielen and S. L. Salzberg (2017). "Bracken: estimating species abundance in metagenomics data." PeerJ Computer Science **3**: e104.

Lucas, P. W., K. Y. Ang, Z. Sui, K. R. Agrawal, J. F. Prinz and N. J. Dominy (2006). "A brief review of the recent evolution of the human mouth in physiological and nutritional contexts." Physiology & behavior **89**(1): 36-38.

Lugli, G. A., S. Duranti, K. Albert, L. Mancabelli, S. Napoli, A. Viappiani, R. Anzalone, G. Longhi, C. Milani and F. Turroni (2019). "Unveiling genomic diversity among members of the species Bifidobacterium pseudolongum, a widely

distributed gut commensal of the animal kingdom." Applied and Environmental Microbiology **85**(8): e03065-03018.

Macfarlane, S. and G. T. Macfarlane (2003). "Regulation of short-chain fatty acid production." Proceedings of the Nutrition Society **62**(1): 67-72.

Maddelein, M. L., N. Libessart, F. Bellanger, B. Delrue, C. Dhulst, N. Vandenkoornhuyse, T. Fontaine, J. M. Wieruszeski, A. Decq and S. Ball (1994). "TOWARD AN UNDERSTANDING OF THE BIOGENESIS OF THE STARCH GRANULE - DETERMINATION OF GRANULE-BOUND AND SOLUBLE STARCH SYNTHASE FUNCTIONS IN AMYLOPECTIN SYNTHESIS." Journal of Biological Chemistry **269**(40): 25150-25157.

Madeira, F., M. Pearce, A. R. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov and R. Lopez (2022). "Search and sequence analysis tools services from EMBL-EBI in 2022." Nucleic acids research **50**(W1): W276-W279.

Makki, K., E. C. Deehan, J. Walter and F. Bäckhed (2018). "The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease." Cell Host & Microbe **23**(6): 705-715.

Mallick, H., A. Rahnavard, L. J. McIver, S. Ma, Y. Zhang, L. H. Nguyen, T. L. Tickle, G. Weingart, B. Ren and E. H. Schwager (2021). "Multivariable association discovery in population-scale meta-omics studies." PLoS computational biology **17**(11): e1009442.

Maqbool, A., R. S. Horler, A. Muller, A. J. Wilkinson, K. S. Wilson and G. H. Thomas (2015). "The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity." Biochemical Society Transactions **43**(5): 1011-1017.

Martin, C. and A. M. Smith (1995). "Starch biosynthesis." The plant cell **7**(7): 971.

Masi, A. C. and C. J. Stewart (2022). "Untangling human milk oligosaccharides and infant gut microbiome." Iscience **25**(1): 103542.

McBurney, M. and L. Thompson (1989). "Effect of human faecal donor on in vitro fermentation variables." Scandinavian journal of gastroenterology **24**(3): 359-367.

McKeen, S., W. Young, J. Mullaney, K. Fraser, W. C. McNabb and N. C. Roy (2019). "Infant complementary feeding of prebiotics for the microbiome and immunity." Nutrients **11**(2): 364.

McMurdie, P. J. and S. Holmes (2013). "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PloS one **8**(4): e61217.

Mehta, A., S. P. Marso and I. Neeland (2017). "Liraglutide for weight management: a critical review of the evidence." Obesity science & practice **3**(1): 3-14.

Meng, C. (2022). "Fast analyzing, exploring and sharing quantitative omics data using omicsViewer." bioRxiv: 2022.2003. 2010.483845.

Meng, D., E. Sommella, E. Salviati, P. Campiglia, K. Ganguli, K. Djebali, W. Zhu and W. A. Walker (2020). "Indole-3-lactic acid, a metabolite of tryptophan, secreted by Bifidobacterium longum subspecies infantis is anti-inflammatory in the immature intestine." Pediatric research **88**(2): 209-217.

Merida, A. and J. Fettke (2021). "Starch granule initiation in Arabidopsis thaliana chloroplasts." The Plant Journal **107**(3): 688-697.

Mesnage, S., T. Fontaine, T. Mignot, M. Delepierre, M. Mock and A. Fouet (2000). "Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation." The EMBO journal **19**(17): 4473-4484.

Miketinas, D. C., K. Shankar, M. Maiya and M. A. Patterson (2020). "Usual dietary intake of resistant starch in US adults from NHANES 2015–2016." The Journal of Nutrition **150**(10): 2738-2747.

Milani, C., G. A. Lugli, S. Duranti, F. Turroni, L. Mancabelli, C. Ferrario, M. Mangifesta, A. Hevia, A. Viappiani and M. Scholz (2015). "Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut." Scientific reports **5**(1): 1-14.

Milani, C., F. Turroni, S. Duranti, G. A. Lugli, L. Mancabelli, C. Ferrario, D. van Sinderen and M. Ventura (2016). "Genomics of the genus Bifidobacterium reveals species-specific adaptation to the glycan-rich gut environment." Applied and Environmental Microbiology **82**(4): 980-991.

Mirzaei, R., E. Dehkhodaie, B. Bouzari, M. Rahimi, A. Gholestani, S. R. Hosseini-Fard, H. Keyvani, A. Teimoori and S. Karampoor (2022). "Dual role of microbiota-derived short-chain fatty acids on host and pathogen." Biomedicine & Pharmacotherapy **145**: 112352.

Mistou, M.-Y., I. C. Sutcliffe and N. M. Van Sorge (2016). "Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria." FEMS microbiology reviews **40**(4): 464-479.

Modrackova, N., M. Makovska, C. Mekadim, E. Vlkova, V. Tejnecky, P. Bolechova and V. Bunesova (2019). "Prebiotic potential of natural gums and starch for bifidobacteria of variable origins." Bioactive Carbohydrates and Dietary Fibre **20**: 100199.

Moeller, A. H., Y. Li, E. Mpoudi Ngole, S. Ahuka-Mundeke, E. V. Lonsdorf, A. E. Pusey, M. Peeters, B. H. Hahn and H. Ochman (2014). "Rapid changes in the gut microbiome during human evolution." Proceedings of the National Academy of Sciences **111**(46): 16431-16435.

Ndeh, D. and H. J. Gilbert (2018). "Biochemistry of complex glycan depolymerisation by the human gut microbiota." FEMS microbiology reviews **42**(2): 146-164.

Neumann, C. J., A. Mahnert, C. Kumpitsch, R. Kiu, M. J. Dalby, M. Kujawska, T. Madl, S. Kurath-Koller, B. Urlesberger and B. Resch (2023). "Clinical NEC prevention practices drive different microbiome profiles and functional responses in the preterm intestine." Nature Communications **14**(1): 1349.

Nicholson, J. K., E. Holmes, J. Kinross, R. Burcelin, G. Gibson, W. Jia and S. Pettersson (2012). "Host-Gut Microbiota Metabolic Interactions." Science **336**(6086): 1262-1267.

Nishiyama, K., T. Takaki, M. Sugiyama, I. Fukuda, M. Aiso, T. Mukai, T. Odamaki, J.-z. Xiao, R. Osawa and N. Okada (2020). "Extracellular vesicles produced by Bifidobacterium longum export mucin-binding proteins." Applied and environmental microbiology **86**(19): e01464-01420.

Noda, T., Y. Takahata, T. Sato, I. Suda, T. Morishita, K. Ishiguro and O. Yamakawa (1998). "Relationships between chain length distribution of amylopectin and gelatinization properties within the same botanical origin for sweet potato and buckwheat." Carbohydrate Polymers **37**(2): 153-158.

Nomura, M. (2022). "Association of the gut microbiome with cancer immunotherapy." International Journal of Clinical Oncology: 1-7.

Norris, N., U. Alcolombri, J. M. Keegstra, Y. Yawata, F. Menolascina, E. Frazzoli, N. M. Levine, V. I. Fernandez and R. Stocker (2022). "Bacterial chemotaxis to saccharides is governed by a trade-off between sensing and uptake." Biophysical Journal **121**(11): 2046-2059.

O'Callaghan, A., F. Bottacini, M. O'Connell Motherway and D. van Sinderen (2015). "Pangenome analysis of Bifidobacterium longum and site-directed mutagenesis through by-pass of restriction-modification systems." BMC Genomics **16**: 832.

O'Callaghan, A. and D. van Sinderen (2016). "Bifidobacteria and Their Role as Members of the Human Gut Microbiota." Front Microbiol **7**: 925.

O'Keefe, S. J., J. V. Li, L. Lahti, J. Ou, F. Carbonero, K. Mohammed, J. M. Posma, J. Kinross, E. Wahl, E. Ruder, K. Vipperla, V. Naidoo, L. Mtshali, S. Tims, P. G. Puylaert, J. DeLany, A. Krasinskas, A. C. Benefiel, H. O. Kaseb, K. Newton, J. K. Nicholson, W. M. de Vos, H. R. Gaskins and E. G. Zoetendal (2015). "Fat, fibre and cancer risk in African Americans and rural Africans." Nat Commun **6**: 6342.

O'Riordan, K. and G. F. Fitzgerald (1999). "Molecular characterisation of a 5.75-kb cryptic plasmid from Bifidobacterium breve NCFB 2258 and determination of mode of replication." FEMS microbiology letters **174**(2): 285-294.

Odamaki, T., F. Bottacini, K. Kato, E. Mitsuyama, K. Yoshida, A. Horigome, J.-z. Xiao and D. van Sinderen (2018). "Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan." Scientific reports **8**(1): 1-12.

Odamaki, T., A. Horigome, H. Sugahara, N. Hashikura, J. Minami, J.-z. Xiao and F. Abe (2015). "Comparative genomics revealed genetic diversity and species/strain-level differences in carbohydrate metabolism of three probiotic bifidobacterial species." International journal of genomics **2015**.

Ogunrinola, G. A., J. O. Oyewale, O. O. Oshamika and G. I. Olasehinde (2020). "The human microbiome and its impacts on health." International Journal of Microbiology **2020**.

Oliphant, K. and E. C. Claud (2022). "Early probiotics shape microbiota." Nature Microbiology **7**(10): 1506-1507.

Oliveira, R. A., V. Cabral and K. B. Xavier (2021). "Microbiome–diet interactions drive antibiotic efficacy." Nature Microbiology **6**(7): 824-825.

Ordiz, M. I., T. D. May, K. Mihindukulasuriya, J. Martin, J. Crowley, P. I. Tarr, K. Ryan, E. Mortimer, G. Gopalsamy and K. Maleta (2015). "The effect of dietary resistant starch type 2 on the microbiota and markers of gut inflammation in rural Malawi children." Microbiome **3**(1): 1-9.

Özcan, E. and D. A. Sela (2018). "Inefficient metabolism of the human milk oligosaccharides lacto-N-tetraose and lacto-N-neotetraose shifts Bifidobacterium longum subsp. infantis physiology." Frontiers in Nutrition **5**: 46.

Palframan, R. J., G. R. Gibson and R. A. Rastall (2003). "Carbohydrate preferences of Bifidobacterium species isolated from the human gut." Current issues in intestinal microbiology **4**(2): 71-75.

Pasquina-Lemonche, L., J. Burns, R. Turner, S. Kumar, R. Tank, N. Mullin, J. Wilson, B. Chakrabarti, P. Bullough and S. Foster (2020). "The architecture of the Gram-positive bacterial cell wall." Nature 582(7811): 294-297.

Payne, A. N., A. Zihler, C. Chassard and C. Lacroix (2012). "Advances and perspectives in in vitro human gut fermentation modeling." Trends in biotechnology 30(1): 17-25.

Paysan-Lafosse, T., M. Blum, S. Chuguransky, T. Grego, B. L. Pinto, G. A. Salazar, M. L. Bileschi, P. Bork, A. Bridge and L. Colwell (2023). "InterPro in 2022." Nucleic Acids Research 51(D1): D418-D427.

Pearson, W. R. (2013). "An introduction to sequence similarity ("homology") searching." Current protocols in bioinformatics 42(1): 3.1. 1-3.1. 8.

Pedersen, A., A. Bardow, S. B. Jensen and B. Nauntofte (2002). "Saliva and gastrointestinal functions of taste, mastication, swallowing and digestion." Oral diseases 8(3): 117-129.

Penders, J., C. Thijs, C. Vink, F. F. Stelma, B. Snijders, I. Kummeling, P. A. van den Brandt and E. E. Stobberingh (2006). "Factors influencing the composition of the intestinal microbiota in early infancy." Pediatrics 118(2): 511-521.

Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain and R. Misra (2007). "Diet and the evolution of human amylase gene copy number variation." Nature genetics 39(10): 1256-1260.

Pham, V. and M. Mohajeri (2018). "The application of in vitro human intestinal models on the screening and development of pre-and probiotics." Beneficial Microbes 9(5): 725-742.

Phillips, S., R. Watt, T. Atkinson, G. M. Savva, A. Hayhoe and L. J. Hall (2021). "The Pregnancy and EARly Life study (PEARL)-a longitudinal study to understand how gut microbes contribute to maintaining health during pregnancy and early life." BMC pediatrics 21(1): 1-11.

Photenhauer, A. L., F. M. Cerqueira, R. Villafuerte-Vega, K. M. Armbruster, F. Mareček, T. Chen, Z. Wawrzak, J. B. Hopkins, C. W. Vander Kooi and Š. Janeček (2022). "The Ruminococcus bromii amylosome protein Sas6 binds single and double helical α-glucan structures in starch." bioRxiv.

Picard, C., J. Fioramonti, A. Francois, T. Robinson, F. Neant and C. Matuchansky (2005). "bifidobacteria as probiotic agents–physiological effects and clinical benefits." Alimentary pharmacology & therapeutics 22(6): 495-512.

Planchot, V., P. Colonna and A. Buleon (1997). "Enzymatic hydrolysis of α-glucan crystallites." Carbohydrate Research **298**(4): 319-326.

Pokusaeva, K., G. F. Fitzgerald and D. van Sinderen (2011). "Carbohydrate metabolism in Bifidobacteria." Genes & nutrition **6**: 285-306.

Pokusaeva, K., A. R. Neves, A. Zomer, M. O'Connell-Motherway, J. MacSharry, P. Curley, G. F. Fitzgerald and D. Van Sinderen (2010). "Ribose utilization by the human commensal Bifidobacterium breve UCC2003." Microbial biotechnology **3**(3): 311-323.

Power, R. C., D. C. Salazar-García, M. Rubini, A. Darlas, K. Harvati, M. Walker, J.-J. Hublin and A. G. Henry (2018). "Dental calculus indicates widespread plant use within the stable Neanderthal dietary niche." Journal of Human Evolution **119**: 27-41.

Rabiu, B. A., A. J. Jay, G. R. Gibson and R. A. Rastall (2001). "Synthesis and fermentation properties of novel galacto-oligosaccharides by β-galactosidases from Bifidobacterium species." Applied and Environmental Microbiology **67**(6): 2526-2530.

Rahat-Rozenbloom, S., J. Fernandes, J. Cheng and T. M. Wolever (2017). "Acute increases in serum colonic short-chain fatty acids elicited by inulin do not increase GLP-1 or PYY responses but may reduce ghrelin in lean and overweight humans." European journal of clinical nutrition **71**(8): 953-958.

Raigond, P., R. Ezekiel and B. Raigond (2015). "Resistant starch in food: a review." Journal of the Science of Food and Agriculture **95**(10): 1968-1978.

Ravi, A., P. Troncoso-Rey, J. Ahn-Jarvis, K. R. Corbin, S. Harris, H. Harris, A. Aydin, G. L. Kay, T. Le Viet and R. Gilroy (2022). "Hybrid metagenome assemblies link carbohydrate structure with function in the human gut microbiome." Communications biology **5**(1): 1-13.

Ravi, R. K., K. Walton and M. Khosroheidari (2018). "MiSeq: a next generation sequencing platform for genomic analysis." Disease gene identification: methods and protocols: 223-232.

Rees, D. C., E. Johnson and O. Lewinson (2009). "ABC transporters: the power to change." Nature reviews Molecular cell biology **10**(3): 218-227.

Reuter, G. (1971). "Designation of type strains for Bifidobacterium species." International Journal of Systematic and Evolutionary Microbiology **21**(4): 273-275.

Ricci, L., J. Mackie, G. E. Donachie, A. Chapuis, K. Mezerová, M. D. Lenardon, A. J. Brown, S. H. Duncan and A. W. Walker (2022). "Human gut bifidobacteria inhibit

the growth of the opportunistic fungal pathogen Candida albicans." FEMS Microbiology Ecology **98**(10): fiac095.

Rios-Covian, D., M. Gueimonde, S. H. Duncan, H. J. Flint and C. G. de Los Reyes-Gavilan (2015). "Enhanced butyrate formation by cross-feeding between Faecalibacterium prausnitzii and Bifidobacterium adolescentis." FEMS microbiology letters **362**(21): fnv176.

Ritchie, M. L. and T. N. Romanuk (2012). "A meta-analysis of probiotic efficacy for gastrointestinal diseases." PloS one **7**(4): e34938.

Robyt, J. F. (2008). Starch: structure, properties, chemistry, and enzymology. Glycoscience, Springer**:** 1437-1472.

Roder, N., C. Gerard, A. Verel, T. Y. Bogracheva, C. L. Hedley, P. R. Ellis and P. J. Butterworth (2009). "Factors affecting the action of α-amylase on wheat starch: Effects of water availability. An enzymic and structural study." Food Chemistry **113**(2): 471-478.

Rodriguez, C. I. and J. B. Martiny (2020). "Evolutionary relationships among bifidobacteria and their hosts and environments." BMC genomics **21**(1): 1-12.

Rodríguez, M. D., A. E. León and M. C. Bustos (2022). "Starch Digestion in Infants: An Update of Available In Vitro Methods—A Mini Review." Plant Foods for Human Nutrition: 1-8.

Roediger, W. (1982). "Utilization of nutrients by isolated epithelial cells of the rat colon." Gastroenterology **83**(2): 424-429.

Rosés, C., A. Cuevas-Sierra, S. Quintana, J. I. Riezu-Boj, J. A. Martínez, F. I. Milagro and A. Barceló (2021). "Gut microbiota bacterial species associated with Mediterranean diet-related food groups in a northern Spanish population." Nutrients **13**(2): 636.

Ross, S., J. Brand, A. Thorburn and A. Truswell (1987). "Glycemic index of processed wheat products." The American journal of clinical nutrition **46**(4): 631-635.

Ryan, S. M., G. F. Fitzgerald and D. van Sinderen (2006). "Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains." Applied and environmental microbiology **72**(8): 5289-5296.

Saarela, M. H. (2019). "Safety aspects of next generation probiotics." Current Opinion in Food Science **30**: 8-13.

Sakanaka, M., S. Nakakawaji, S. Nakajima, S. Fukiya, A. Abe, W. Saburi, H. Mori and A. Yokota (2018). "A Transposon Mutagenesis System for Bifidobacterium

longum subsp. longum Based on an IS 3 Family Insertion Sequence, IS Blo11." Applied and Environmental Microbiology **84**(17): e00824-00818.

Salzberg, S. L. and D. E. Wood (2021). "Releasing the Kraken." Frontiers in Bioinformatics **1**: 75.

Samuelson, L. C., K. Wiebauer, C. Snow and M. H. Meisler (1990). "Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution." Molecular and cellular biology **10**(6): 2513-2520.

Sekine, K., E. Watanabe-Sekine, T. Toida, T. Kasashima, T. Kataoka and Y. Hashimoto (1994). "Adjuvant activity of the cell wall of Bifidobacterium infantis for in vivo immune responses in mice." Immunopharmacology and immunotoxicology **16**(4): 589-609.

Sela, D. A., J. Chapman, A. Adeuya, J. H. Kim, F. Chen, T. R. Whitehead, A. Lapidus, D. Rokhsar, C. Lebrilla and J. German (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 **105**(48): 18964-18969.

Sender, R., S. Fuchs and R. Milo (2016). "Revised estimates for the number of human and bacteria cells in the body." PLoS biology **14**(8): e1002533.

Seung, D., T. B. Schreier, L. Bürgy, S. Eicke and S. C. Zeeman (2018). "Two plastidial coiled-coil proteins are essential for normal starch granule initiation in Arabidopsis." The Plant Cell **30**(7): 1523-1542.

Sharma, M., A. Wasan and R. K. Sharma (2021). "Recent developments in probiotics: An emphasis on Bifidobacterium." Food Bioscience **41**: 100993.

Sharp, R. and G. T. Macfarlane (2000). "Chemostat enrichments of human feces with resistant starch are selective for adherent butyrate-producing clostridia at high dilution rates." Applied and Environmental Microbiology **66**(10): 4212-4221.

Shetty, S. A., S. Zuffa, T. P. N. Bui, S. Aalvink, H. Smidt and W. M. De Vos (2018). "Reclassification of Eubacterium hallii as Anaerobutyricum hallii gen. nov., comb. nov., and description of Anaerobutyricum soehngenii sp. nov., a butyrate and propionate-producing bacterium from infant faeces." International Journal of Systematic and Evolutionary Microbiology **68**(12): 3741-3746.

Shujun, W., G. Wenyuan, J. Wei and X. Peigen (2006). "Crystallography, morphology and thermal properties of starches from four different medicinal plants of Fritillaria species." Food Chemistry **96**(4): 591-596.

Silhavy, T. J., D. Kahne and S. Walker (2010). "The bacterial cell envelope." <u>Cold Spring Harbor perspectives in biology</u> **2**(5): a000414.

Slavin, J. L., D. Jacobs and L. Marquart (2000). "Grain processing and nutrition." <u>Critical Reviews in Food Science and Nutrition.</u> **40**(4): 309-326.

Smith, A. M. (2001). "The biosynthesis of starch granules." <u>Biomacromolecules</u> **2**(2): 335-341.

Smith, A. M., K. Denyer and C. Martin (1997). "The synthesis of the starch granule." <u>Annual review of plant biology</u> **48**(1): 67-87.

Smith, S. (2019). phylosmith: an R-package for reproducible and efficient microbiome analysis with phyloseq-objects. J Open Source Softw 4: 1442.

Soler, A., G. Velazquez, R. Velazquez-Castillo, E. Morales-Sanchez, P. Osorio-Diaz and G. Mendez-Montealvo (2020). "Retrogradation of autoclaved corn starches: Effect of water content on the resistant starch formation and structure." <u>Carbohydrate Research</u> **497**: 108137.

Stewart, C. J., N. J. Ajami, J. L. O'Brien, D. S. Hutchinson, D. P. Smith, M. C. Wong, M. C. Ross, R. E. Lloyd, H. Doddapaneni and G. A. Metcalf (2018). "Temporal development of the gut microbiome in early childhood from the TEDDY study." <u>Nature</u> **562**(7728): 583-588.

Stothard, P. (2000). "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences." <u>Biotechniques</u> **28**(6): 1102-1104.

Sun, Y. and M. X. O'Riordan (2013). "Regulation of bacterial pathogenesis by intestinal short-chain fatty acids." <u>Advances in applied microbiology</u> **85**: 93-118.

Sun, Z., W. Zhang, C. Guo, X. Yang, W. Liu, Y. Wu, Y. Song, L. Y. Kwok, Y. Cui and B. Menghe (2015). "Comparative genomic analysis of 45 type strains of the genus Bifidobacterium: a snapshot of its genetic diversity and evolution." <u>PLoS One</u> **10**(2): e0117912.

Tam, R. and M. H. Saier Jr (1993). "Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria." <u>Microbiological reviews</u> **57**(2): 320-346.

Tarracchini, C., C. Milani, G. A. Lugli, L. Mancabelli, F. Fontana, G. Alessandri, G. Longhi, R. Anzalone, A. Viappiani and F. Turroni (2021). "Phylogenomic disentangling of the Bifidobacterium longum subsp. infantis taxon." <u>Microbial Genomics</u> **7**(7).

Tatge, H., J. Marshall, C. Martin, E. Edwards and A. Smith (1999). "Evidence that amylose synthesis occurs within the matrix of the starch granule in potato tubers." Plant, Cell & Environment **22**(5): 543-550.

Teichmann, J. and D. W. Cockburn (2021). "In vitro fermentation reveals changes in butyrate production dependent on resistant starch source and microbiome composition." Frontiers in microbiology **12**: 640253.

Terrapon, N., V. Lombard, É. Drula, P. Lapébie, S. Al-Masaudi, H. J. Gilbert and B. Henrissat (2018). "PULDB: the expanded database of Polysaccharide Utilization Loci." Nucleic Acids Res **46**(D1): D677-d683.

Tetlow, I. J., M. K. Morell and M. J. Emes (2004). "Recent developments in understanding the regulation of starch metabolism in higher plants." Journal of experimental botany **55**(406): 2131-2145.

Thursby, E. and N. Juge (2017). "Introduction to the human gut microbiota." Biochemical journal **474**(11): 1823-1836.

Tiwari, U. P., A. K. Singh and R. Jha (2019). "Fermentation characteristics of resistant starch, arabinoxylan, and β-glucan and their effects on the gut microbial ecology of pigs: a review." Animal Nutrition **5**(3): 217-226.

Toh, H., K. Oshima, A. Nakano, M. Takahata, M. Murakami, T. Takaki, H. Nishiyama, S. Igimi, M. Hattori and H. Morita (2013). "Genomic adaptation of the Lactobacillus casei group." PloS one **8**(10): e75073.

Tsukuda, N., K. Yahagi, T. Hara, Y. Watanabe, H. Matsumoto, H. Mori, K. Higashi, H. Tsuji, S. Matsumoto, K. Kurokawa and T. Matsuki (2021). "Key bacterial taxa and metabolic pathways affecting gut short-chain fatty acid profiles in early life." The ISME Journal **15**(9): 2574-2590.

Turroni, F., C. Milani, S. Duranti, L. Mancabelli, M. Mangifesta, A. Viappiani, G. A. Lugli, C. Ferrario, L. Gioiosa and A. Ferrarini (2016). "Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach." The ISME journal **10**(7): 1656-1668.

Turroni, F., F. Strati, E. Foroni, F. Serafini, S. Duranti, D. van Sinderen and M. Ventura (2012). "Analysis of predicted carbohydrate transport systems encoded by Bifidobacterium bifidum PRL2010." Applied and environmental microbiology **78**(14): 5002-5012.

Underwood, M. A., J. B. German, C. B. Lebrilla and D. A. Mills (2015). "Bifidobacterium longum subspecies infantis: champion colonizer of the infant gut." Pediatric research **77**(1): 229-235.

Valk, V., A. Lammerts van Bueren, R. M. van der Kaaij and L. Dijkhuizen (2016). "Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain α-amylase enzymes." The FEBS journal **283**(12): 2354-2368.

Van den Abbeele, P., C. Belzer, M. Goossens, M. Kleerebezem, W. M. De Vos, O. Thas, R. De Weirdt, F.-M. Kerckhof and T. Van de Wiele (2013). "Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model." The ISME journal **7**(5): 949-961.

Van den Broek, L. A., S. W. Hinz, G. Beldman, C. H. Doeswijk-Voragen, J.-P. Vincken and A. G. Voragen (2005). "Glycosyl hydrolases from Bifidobacterium adolescentis DSM20083. An overview." Le Lait **85**(1-2): 125-133.

van Hoek, M. J. and R. M. Merks (2012). "Redox balance is key to explaining full vs. partial switching to low-yield metabolism." BMC systems biology **6**(1): 1-10.

Vandenplas, Y., B. Berger, V. P. Carnielli, J. Ksiazyk, H. Lagström, M. Sanchez Luna, N. Migacheva, J.-M. Mosselmans, J.-C. Picaud and M. Possner (2018). "Human milk oligosaccharides: 2′-fucosyllactose (2′-FL) and lacto-N-neotetraose (LNnT) in infant formula." Nutrients **10**(9): 1161.

Vandeputte, G. and J. A. Delcour (2004). "From sucrose to starch granule to starch physical behaviour: a focus on rice starch." Carbohydrate polymers **58**(3): 245-266.

Vatanen, T., Q. Y. Ang, L. Siegwald, S. A. Sarker, C. I. Le Roy, S. Duboux, O. Delannoy-Bruno, C. Ngom-Bru, C. L. Boulangé and M. Stražar (2022). "A distinct clade of Bifidobacterium longum in the gut of Bangladeshi children thrives during weaning." Cell **185**(23): 4280-4297. e4212.

Vazquez-Gutierrez, P., C. Lacroix, C. Chassard, J. Klumpp, M. J. Stevens and C. Jans (2015). "Bifidobacterium pseudolongum strain PV8-2, isolated from a stool sample of an anemic Kenyan infant." Genome announcements **3**(1): e01469-01414.

Velikova, P., A. Stoyanov, G. Blagoeva, L. Popova, K. Petrov, V. Gotcheva, A. Angelov and P. Petrova (2016). "Starch utilization routes in lactic acid bacteria: New insight by gene expression assay." Starch-Stärke **68**(9-10): 953-960.

Venkataraman, A., J. Sieber, A. Schmidt, C. Waldron, K. Theis and T. Schmidt (2016). "Variable responses of human microbiomes to dietary supplementation with resistant starch." Microbiome **4**: 1-9.

Vitetta, L., E. T. Saltzman, M. Thomsen, T. Nikov and S. Hall (2017). "Adjuvant probiotics and the intestinal microbiome: enhancing vaccines and immunotherapy outcomes." Vaccines **5**(4): 50.

von Ah, U., V. Mozzetti, C. Lacroix, E. E. Kheadr, I. Fliss and L. Meile (2007). "Classification of a moderately oxygen-tolerant isolate from baby faeces as Bifidobacterium thermophilum." BMC microbiology **7**(1): 1-11.

Walker, A. W., S. H. Duncan, E. C. McWilliam Leitch, M. W. Child and H. J. Flint (2005). "pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon." Applied and environmental microbiology **71**(7): 3692-3700.

Walker, A. W., J. Ince, S. H. Duncan, L. M. Webster, G. Holtrop, X. Ze, D. Brown, M. D. Stares, P. Scott and A. Bergerat (2011). "Dominant and diet-responsive groups of bacteria within the human colonic microbiota." The ISME journal **5**(2): 220-230.

Walker, A. W., J. Ince, S. H. Duncan, L. M. Webster, G. Holtrop, X. Ze, D. Brown, M. D. Stares, P. Scott, A. Bergerat, P. Louis, F. McIntosh, A. M. Johnstone, G. E. Lobley, J. Parkhill and H. J. Flint (2011). "Dominant and diet-responsive groups of bacteria within the human colonic microbiota." The ISME journal **5**(2): 220-230.

Walker, A. W. and T. D. Lawley (2013). "Therapeutic modulation of intestinal dysbiosis." Pharmacological research **69**(1): 75-86.

Walsh, C., J. A. Lane, D. van Sinderen and R. M. Hickey (2022). "Human milk oligosaccharide-sharing by a consortium of infant derived Bifidobacterium species." Scientific Reports **12**(1): 4143.

Wang, K., J. Hasjim, A. C. Wu, R. J. Henry and R. G. Gilbert (2014). "Variation in amylose fine structure of starches from different botanical sources." Journal of agricultural and food chemistry **62**(19): 4443-4453.

Wang, S., C. Li, L. Copeland, Q. Niu and S. Wang (2015). "Starch Retrogradation: A Comprehensive Review." Comprehensive Reviews in Food Science and Food Safety **14**(5): 568-585.

Wang, T. L., T. Y. Bogracheva and C. L. Hedley (1998). "Starch: as simple as A, B, C?" Journal of experimental botany **49**(320): 481-502.

Wang, X., P. L. Conway, I. L. Brown and A. J. Evans (1999). "In vitro utilization of amylopectin and high-amylose maize (amylomaize) starch granules by human colonic bacteria." Applied and environmental microbiology **65**(11): 4848-4854.

Wang, Y., E. K. Mortimer, K. G. Katundu, N. Kalanga, L. E. Leong, G. L. Gopalsamy, C. T. Christophersen, A. C. Richard, A. Shivasami and G. C. Abell (2019). "The capacity of the fecal microbiota from Malawian infants to ferment resistant starch." Frontiers in microbiology **10**: 1459.

Wardman, J. F., R. K. Bains, P. Rahfeld and S. G. Withers (2022). "Carbohydrate-active enzymes (CAZymes) in the gut microbiome." Nature Reviews Microbiology **20**(9): 542-556.

Warren, F. J., N. M. Fukuma, D. Mikkelsen, B. M. Flanagan, B. A. Williams, A. T. Lisle, P. O. Cuiv, M. Morrison and M. J. Gidley (2018). "Food Starch Structure Impacts Gut Microbiome Composition." Msphere **3**(3).

Warren, F. J., N. M. Fukuma, D. Mikkelsen, B. M. Flanagan, B. A. Williams, A. T. Lisle, P. Ó Cuív, M. Morrison and M. J. Gidley (2018). "Food starch structure impacts gut microbiome composition." Msphere **3**(3): e00086-00018.

Waters, E. V., L. A. Tucker, J. K. Ahmed, J. Wain and G. C. Langridge (2022). "Impact of Salmonella genome rearrangement on gene expression." bioRxiv.

Whittaker, C. A. and R. O. Hynes (2002). "Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere." Molecular biology of the cell **13**(10): 3369-3387.

Wick, R. R. and P. Menzel (2017). "Filtlong." Github, https://github.com/rrwick/Filtlong, Editor.

Williams, C., G. Walton, L. Jiang, S. Plummer, I. Garaiova and G. R. Gibson (2015). "Comparative analysis of intestinal tract models." Annu Rev Food Sci Technol **6**(1): 329-350.

Wrangham, R. and N. Conklin-Brittain (2003). "Cooking as a biological trait." Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology **136**(1): 35-46.

Wrangham, R. W. and R. N. Carmody (2010). "Human adaptation to the control of fire." Evolutionary Anthropology.

Wrangham, R. W., J. H. Jones, G. Laden, D. Pilbeam and N. Conklin-Brittain (1999). "The raw and the stolen: cooking and the ecology of human origins." Current anthropology **40**(5): 567-594.

Xie, Z., L. Ding, Q. Huang, X. Fu, F. Liu, S. Dhital and B. Zhang (2021). "In vitro colonic fermentation profiles and microbial responses of propionylated high-amylose maize starch by individual Bacteroides-dominated enterotype inocula." Food Research International **144**: 110317.

Yadav, D., T. S. Ghosh and S. S. Mande (2016). "Global investigation of composition and interaction networks in gut microbiomes of individuals belonging to diverse geographies and age-groups." Gut pathogens **8**: 1-21.

Yao, H., B. M. Flanagan, B. A. Williams, D. Mikkelsen and M. J. Gidley (2023). "Particle size of dietary fibre has diverse effects on in vitro gut fermentation rate and end-products depending on food source." Food Hydrocolloids **134**: 108096.

Yoon, S.-H., S.-M. Ha, J. Lim, S. Kwon and J. Chun (2017). "A large-scale evaluation of algorithms to calculate average nucleotide identity." Antonie Van Leeuwenhoek **110**(10): 1281-1286.

You, S. and M. S. Izydorczyk (2002). "Molecular characteristics of barley starches with variable amylose content." Carbohydrate polymers **49**(1): 33-42.

Yu, M., S. Arıoğlu-Tuncil, Z. Xie, X. Fu, Q. Huang, T. Chen and B. Zhang (2021). "In vitro fecal fermentation profiles and microbiota responses of pulse cell wall polysaccharides: enterotype effect." Food & Function **12**(18): 8376-8385.

Yu, W., H. Li, W. Zou, K. Tao, J. Zhu and R. G. Gilbert (2019). "Using starch molecular fine structure to understand biosynthesis-structure-property relations." Trends in Food Science & Technology **86**: 530-536.

Ze, X., Y. Ben David, J. A. Laverde-Gomez, B. Dassa, P. O. Sheridan, S. H. Duncan, P. Louis, B. Henrissat, N. Juge and N. M. Koropatkin (2015). "Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium Ruminococcus bromii." MBio **6**(5): e01058-01015.

Ze, X., Y. B. David, J. A. Laverde-Gomez, B. Dassa, P. O. Sheridan, S. H. Duncan, P. Louis, B. Henrissat, N. Juge and N. M. Koropatkin (2015). "Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic Firmicutes bacterium Ruminococcus bromii." MBio **6**(5): e01058-01015.

Ze, X., S. H. Duncan, P. Louis and H. J. Flint (2012). "Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon." The ISME journal **6**(8): 1535-1543.

Zeeman, S. C., A. Tiessen, E. Pilling, K. L. Kato, A. M. Donald and A. M. Smith (2002). "Starch synthesis in Arabidopsis. Granule synthesis, composition, and structure." Plant physiology **129**(2): 516-529.

Zeng, S., D. Patangia, A. Almeida, Z. Zhou, D. Mu, R. Paul Ross, C. Stanton and S. Wang (2022). "A compendium of 32,277 metagenome-assembled genomes

and over 80 million genes from the early-life human gut microbiome." <u>Nature communications</u> **13**(1): 1-15.

Zhang, H., T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, P. K. Busk, Y. Xu and Y. Yin (2018). "dbCAN2: a meta server for automated carbohydrate-active enzyme annotation." <u>Nucleic Acids Research</u> **46**(W1): W95-W101.

Zhong, Y., L. Liu, J. Qu, A. Blennow, A. R. Hansen, Y. Wu, D. Guo and X. Liu (2020). "Amylose content and specific fine structures affect lamellar structure and digestibility of maize starches." <u>Food Hydrocolloids</u> **108**: 105994.

Zhou, C., X. Fang, J. Xu, J. Gao, L. Zhang, J. Zhao, Y. Meng, W. Zhou, X. Han and Y. Bai (2020). "Bifidobacterium longum alleviates irritable bowel syndrome-related visceral hypersensitivity and microbiota dysbiosis via Paneth cell regulation." <u>Gut Microbes</u> **12**(1): 1782156.

Fin