Pectin degradation and metabolism in *Monoglobus pectinilyticus* 14^T from human faeces

By Caroline Chae-hyun Kim

A thesis presented in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy
in
Microbiology

at Massey University, Manawatu, New Zealand

2017

Supervisory Committee

Dr Douglas I. Rosendale (The New Zealand Institute for Plant and Food Research)

Supervisor

Dr Mark L. Patchett (Institute of Fundamental Sciences, Massey University)

Supervisor

Dr Zoe Jordens (Institute of Fundamental Sciences, Massey University)

Supervisor

Dr William J. Kelly (Donvis Limited)

Supervisor

Professor Gerald W. Tannock (Department of Microbiology and Immunology, University of Otago)

Supervisor

Abstract

Pectin is a conspicuous plant polysaccharide, comprising one third of the dry weight of dietary fibre in common vegetables and fruit. Although pectin is almost completely digested by the human gut microbiota, few bacterial species are known to possess a comprehensive glycobiome to challenge the structurally complex pectin. The current understanding of the colonic degradation of pectin is incomplete, as the knowledge has almost exclusively derived from studying the sequestration system of *Bacteroides* spp. Here I report the isolation and characterization of *Monoglobus pectinilyticus*, and the sequencing of its genome which so far encodes the most pectin-specialized repertoire of carbohydrate active enzymes (CAZymes) found from the human gut. *M. pectinilyticus* also possesses an extracellular pectin degradation system consisting of novel protein constituents which did not find significant sequence homology and functional matches using the most up-to-date nucleotide and protein sequence databases. Proteome analysis of *M. pectinilyticus* using iTRAQ quantification revealed that pectin-degrading CAZymes and the potential constituents of the novel pectin degradation system were differentially up-regulated in response to the availability of pectin. Finally, using quantitative PCR, a positive correlation was observed between the prevalence of *M. pectinilyticus* and the consumption of fibre, vegetables, and pectin in individuals living in NZ. The discovery of *M. pectinilyticus* may add a new layer of complexity onto our interpretation of the colonic pectin degradation by presenting a system highly relevant to the pectin-rich diet of humans, and by suggesting a possibility outside the established paradigms of microbial polysaccharide degradation. The presence of *M. pectinilyticus* and the related uncultured bacteria in the gastrointestinal systems of humans and animals indicated that the organisms of this lineage are frequent terrestrial gut commensals, prompting an investigation into the genomic and molecular properties underlying their carbohydrate degradation potentials.

Acknowledgement

The support and help of many individuals are gratefully acknowledged. I would like to thank all my supervisors for providing me with the opportunity to do this research. I am deeply grateful for their mentorship, support, guidance, and optimism which were unreservedly given to me throughout the duration of my studies. I would also like to extend my sincere appreciations to all members of Food, Nutrition & Health group at Plant & Food Research, for their friendship and contributions to my works. A special thanks to Dr Ian Sims and Dr Tracey Bell for their help with respect to carbohydrate analysis. I am also grateful for the administrative assistance provided by many staffs at Plant & Food Research and Massey University. I am endlessly thankful to my parents and sisters for their encouragement, patience, and understanding throughout my studies. Finally, this work would not have been possible without the financial support of Ministry of Business, Innovation and Employment of New Zealand ('Foods for Health at Different Life Stages' C11X1312).

Table of Contents

List of Figures

List of Tables

List of Appendices

List of copyrighted material for which permission was obtained

Material	Source	Note
Figure 1.1	Flint *et al*., 2012. Gut Microbes. 3, 289-306.	No permission required
Figure 1.2	Mohnen, 2008. Current Opinion in Plant Biology. 11, 266-277.	Permission granted
Figure 1.3	Simoneit *et al*., 1999. Atmospheric Environment. 33, 173-182.	Permission granted
Figure 1.4	Scheller and Ulvskov, 2010. Annual Review of Plant Biology. 61, 263-89	No permission required
Figure 1.5	Rincon *et al*., 2007. Journal of Bacteriology. 189, 4774-4783.	No permission required
Figure 1.6	Martens and Koropatkin, 2009. Journal of Biological Chemistry. 284, 24673-24677.	No permission required
Figure 1.7	Ndeh *et al*., 2017. Nature. 544, 65-70.	Permission granted
Figure 1.8	Condemine and Robert-Baudouy, 1987. Journal of Bacteriology. 169, 1972-1978.	No permission required
Appendix 21	National Health and Nutrition Examination Survey (NHANES), 2003-2006.	No permission required

Abbreviations

2-*O*-Me-Fuc	2-*O*-methyl L-fucose
2-*O*-Me-Xyl	2-*O*-methyl-xylose
AceA	L-aceric acid
Api	D-apiose
Ara*f*	arabinofuranose
Ara*p*	arabinopyranose
BN	basal nutrient medium
CAZyme	carbohydrate active enzyme
CBM	carbohydrate-binding module
CDS	coding sequence
CE	carbohydrate esterase
COG	Cluster of orthologous groups
CRISPR	Clustered regularly interspaced short palindromic repeats
DDH	DNA-DNA hybridization
DE	degree of esterification
DhaA	3-deoxy-D-*lyxo*-heptulosaric acid
DSMZ	German Collection of Microorganisms and Cell Cultures
Fru	fructose
Fuc*p*	fucose pyranose
Gal*p*	galactopyranose
Gal*p*A	galacturonic acid
GH	glycoside hydrolase
GI	gastrointestinal tract
Glc*p*A	glucuronic acid
Glu	glucose
GT	glycosyltransferase
HG	homogalacturonan
HGT	horizontal gene transfer
HMM	Hidden Markov model
HMP	Human Microbiome Project
HPAEC-PAD	High-Performance Anion-Exchange Chromatography Coupled with Pulsed Amperometric Detection
HSP	high-scoring segment pairs
IJSEM	International Journal of Systematic and Evolutionary Microbiology
JCM	Japan Culture Collection
kdoA	2-keto-3-deoxy-D-*manno*-octulosonic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani medium
LPS	lipoprotein signal peptides
Man	mannose
ML	middle lamella
NCBI	National Center for Biotechnology Information
ORF	open reading frame
Ori	origin of replication
PBS	phosphate buffered saline

PCW	plant cell wall
PL	polysaccharide lyase
POCP	the percentage of conserved proteins
PUL	polysaccharide utilization locus
RC	reinforced clostridial medium
RG	rhamnogalacturonan
Rha*p*	rhamnopyranose
SLH	S-layer homology
Sus	starch utilization system or sequestration system
TEM	transmission electron microscope
XGA	xylogalacturonan
Xyl*p*	xylopyranose

Chapter 1

Literature review

1.1 Plant cell wall degradation by colonic bacteria

Fibre is an important part of our diet. Dietary fibres provide many health benefits including reducing the risks of heart diseases, stroke, diabetes, obesity, and various gastrointestinal disorders (Anderson *et al*., 2009). Fibres are traditionally defined as dietary plant materials that resist the degradative actions of human digestive enzymes. Plant cell walls of common vegetables, fruits, whole grain cereals, and legumes make up a large portion of dietary fibres, constituting 20 – 30 % of daily consumption of dietary fibres (Harris *et al*., 2006). Plant cell walls consist of a complex polysaccharide matrix formed by intricate interactions between cellulose, hemicelluloses, and pectin (Figure 1.1). The surface of the plant cell wall is covered with pectin layers called the middle lamellae which cement two adjacent plant cells together.

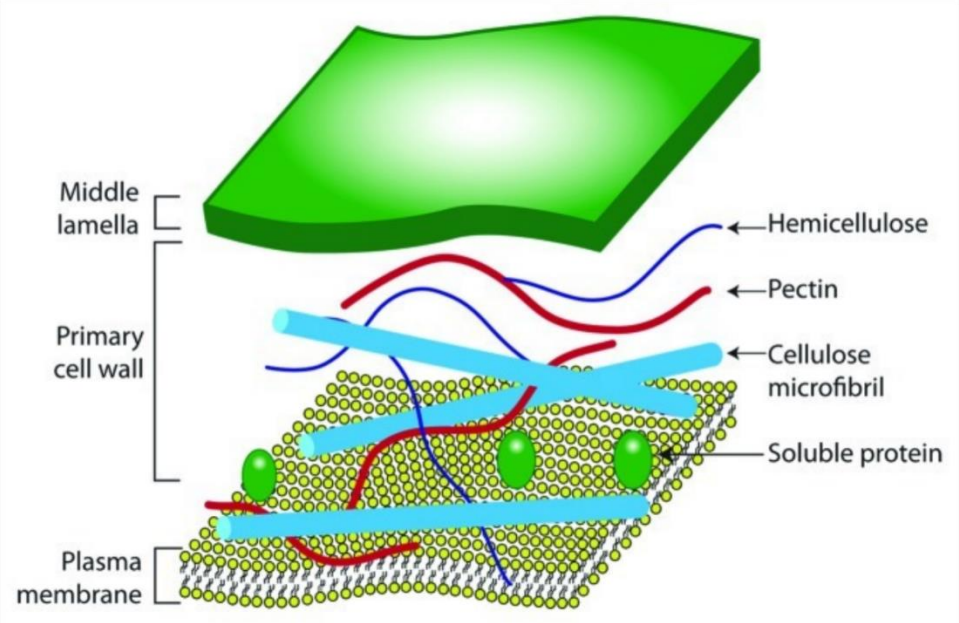


Figure 1.1 Plant cell wall structure. Schematic representation of major structural polysaccharides of a "typical" primary plant cell wall (Flint *et al*., 2012)**.** No copyright permission is required for the use of the figure.

Common flowering vegetables (e.g. cabbages, broccoli, cauliflower, carrots, and spinach) possess lignified secondary cell walls that are rich in cellulose and heteroxylan, whereas fruits usually do not have secondary cell walls and are low in lignin (Sarkar *et al*., 2009). The presence of thick and lignified secondary cell walls is responsible for the tough and stringy texture of some vegetables (e.g. broccoli and asparagus) and gritty texture of some fruits (pears and feijoas), causing undesirable effects on palatability and texture (Harris and Smith, 2006, Muller *et al*., 2003). Stem and vascular regions, compared to the edible parts of vegetables, tend to have heavily lignified and thick secondary cell walls with higher proportions of xylan which forms cross-links with cellulose microfibrils (Muller *et al*., 2003, Carpita and Gibeaut, 1993, Fujino *et al*., 2000).

Due to the complexity and variability of covalent and non-covalent cross-linkages between the polysaccharide components of the plant cell wall, it is difficult to extricate free-end of carbohydrate polymer for degradation without disrupting the overall chemical structure of the surrounding region. As a result, plant cell walls pose a considerable challenge to the human digestive systems, as the insoluble polysaccharides efficiently resist degradation by endogenous enzymes. Gastric digestions within the stomach are known to cause a limited degree of erosion on the surface of ingested plant materials (Kong and Singh, 2009). A study showed that the gastric juice failed to penetrate into the central region of food particles in a simulated human stomach model, suggesting the intercellular integrity of plant tissues remained undamaged in the undigested central regions of plant materials (Kong and Singh, 2009). The colonic degradation of plant polysaccharides is achieved by the action of a diverse array of gut bacteria enriched with genes encoding fibrolytic enzymes that cleave various glycosidic linkages. The human gastrointestinal (GI) tract is densely populated with an estimate of 100 trillion prokaryotic and eukaryotic microorganisms, achieving the highest densities of microbial cells and extraneous genomes existing in the human body (Backhed *et al*., 2005). It is estimated that there are approximately 9,000,000 unique genes in the human gut bacterial community; a staggering number considering the number of genes estimated for human genome is about 23,000 (Yang *et al*., 2009). The blending of genomes of the human host and the gut microbiome partially relieve humans from the burden of being solely responsible for evolving metabolic traits that may confer evolutionary

advantages to human digestive functions. The concerted efforts of gut microorganisms often result in metabolic versatility far greater than any humans can achieve. Such symbiotic co-evolution of the human host and gut microorganisms may have played important roles during the long struggle for human survival by expanding the metabolic capabilities and providing room for genetic variations (Turnbaugh *et al*., 2007). Despite the great abundance and diversity of bacterial communities colonizing the human gut, the ability to utilize the insoluble cell wall polysaccharides trapped within the recalcitrant cell wall matrix has been found in only a few bacterial species. Termed "keystone species" or "primary degraders", these bacteria specialize in the degradation of insoluble plant polymers and essentially initiate the plant cell wall degradation to release more accessible-forms of polymers which can be utilized by other gut microbes (Flint *et al*., 2012).

This review addresses the current research of keystone gut bacteria and microbial enzymes involved in the human colonic degradation of plant cell wall polysaccharides, with a focus on pectin.

1.2 Plant cell wall polysaccharides

1.2.1 Pectin

Pectin is a major non-cellulosic component of plant cell wall polysaccharides, comprising approximately one third of the cell wall dry weight of dicotyledons and non-Poales monocotyledon plants, and 0.5 – 4 % of the fresh weight of edible plant materials (Caffall and Mohnen, 2009; Jayani et al., 2005). Some Poales plants, including grass, contain less than 10% pectin by the dry weight of their cell walls (Caffall and Mohnen, 2009). The relative molecular weights of naturally occurring pectin greatly vary depending on the plant species, and are estimated to be within the range of 25 – 360 kDa (Jayani et al., 2005). The pectic polysaccharides have a linear backbone structure containing galacturonic acid (Gal*p*A) and rhamnose (Rha*p*) units joined in alternating chains via α-1, 4-glycosidic linkages, accompanied by side chains containing neutral sugars such as arabinose, galactose and xylose (Figure 1.2). The major structural components of the pectin backbone includes

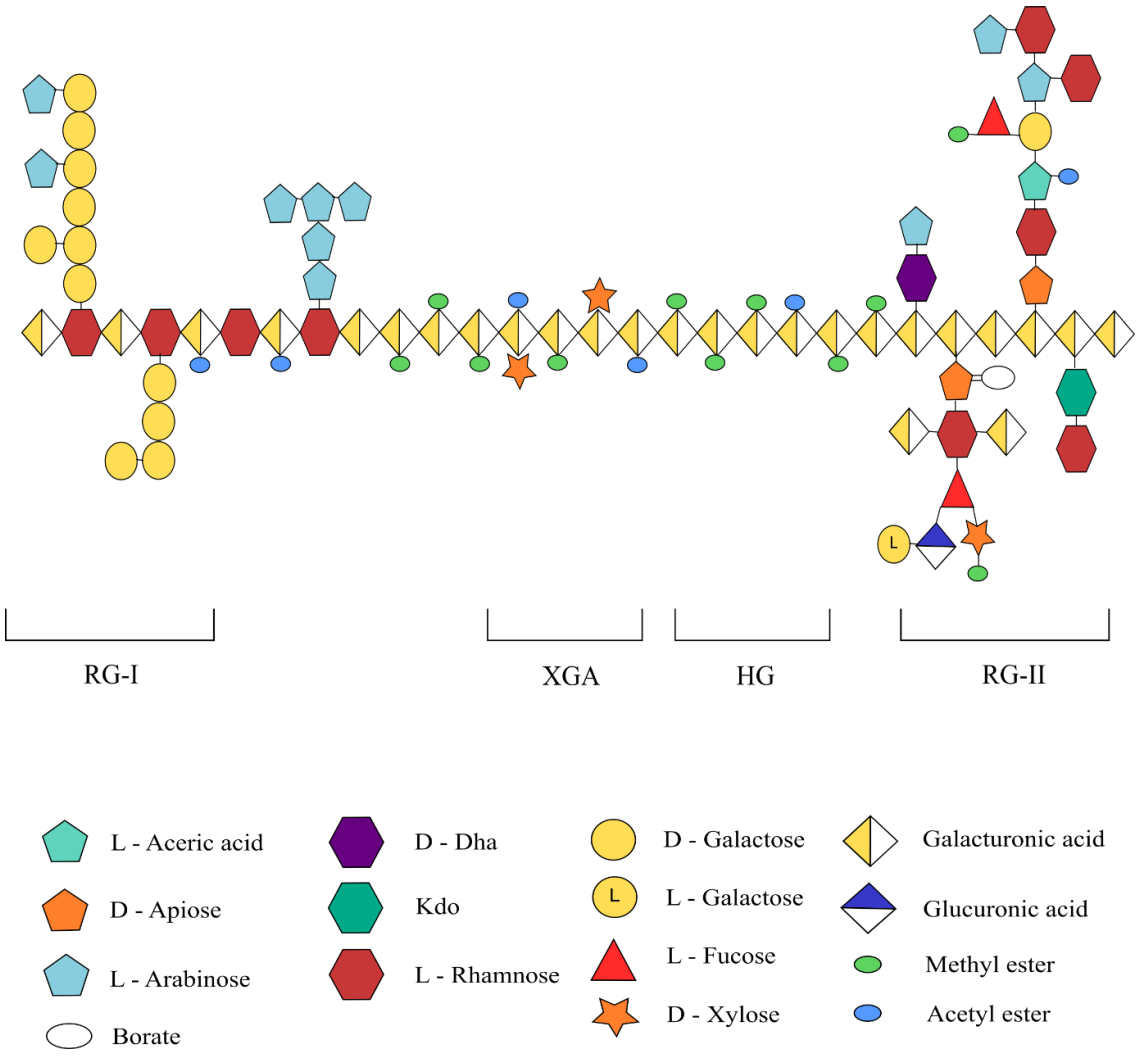


Figure 1.2 Pectin structure depicted with the four pectic polysaccharides homogalacturonan (HG), rhamnogalacturonan-I (RG-I), rhamnogalacturonan (RG-II), and xylogalacturonan (XGA) linked to each other. This diagram was adopted from Mohnen (2008). Copyright permission has been granted for the use of the figure. Monosaccharide symbols are taken from Symbol and Text Nomenclature for representation of Glycan Structure from Consortium for Functional Glycomics.

homogalacturonan (HG), rhamnogalacturonan-I (RG-I), rhamnogalacturonan-II (RG-II), and xylogalacturonan (XGA). HG is a homogenous polymer of α-1,4-linked-D-galacturonic acid which constitutes the majority of uronic acid contents of pectin (Bush *et al*., 2001). HG is the most predominant component of pectin, comprising 65 – 70 % of the total pectin mass (Mohenan, 2008). Approximately half of the galacturonic acid residues present in HG are either methyl-esterified at C-6 or acetylated at O-2 and/or O-3 (Ridley *et al*., 2001). The fluctuating degrees of methylation and acetylation during different developmental stages play significant roles in regulating the plant cell growth and ripening of fruits (Mohenan, 2008). Pectin may be categorized as high-ester (DE $>$ 50 %) or low-ester (DE $<$ 50 %) pectins depending on the degree of esterification (DE) at galacturonic acid residues in HG (Harris and Smith, 2006). Citrus (64 %), apple (~70 %), kiwifruit (~90 %) pectins are typically extensively methylated, whereas pectin from sunflower, pear, potato, and peas are known to have low DEs (Harris and Smith, 2006; Yuliarti *et al*., 2015; Sato *et al*., 2011; Voragen *et al*., 1986). Plant cell walls contain up to 7 % of acetyl groups esterified to O-2 and/or O-3 groups of pectin and xylan (Shevchik and Hugouvieux-Cotte-Pattat, 2003). Sugarbeet pectin has an unusual sugar composition in that up to ~ 20 % of its galacturonic acid residues were found to be acetylated (Williams *et al*., 2005). Gal*p*A residues that are neither methyl-esterified nor acetylated carry a negative charge, enabling the formation of a characteristic gel-like texture by chelating Ca^{2+} ions (Caffall and Mohnen, 2009). HG forms successive connections to RG-I and RG-II that carry most of the side chains found on pectin. In XGA, a short polymer of β-D-xylose was also found to substitute at O-3 of HG backbone in growing leaves and peas (Le Goff *et al.*, 2001).

The RG-I and RG-II regions of pectin are compositionally heterogeneous, containing diverse neutral sugars. Rhamnose residues in RG-I are branched by homoarabinans, homogalactans, and arabinogalactan side chains which can contain up to 100 – 200 sugars (Mohenan, 2008). The types and compositions of RG-I side chains may show a high degree of heterogeneity depending on the source of pectin (Caffall and Mohnen, 2009). Approximately 20 – 80 % of rhamnose residues contain side chains depending on the plant species (Willats *et al*., 2001). The backbone of arabinan side

chains are made of α-L-1,5-Ara*f* units which in turn are branched at O2 and O3 with α-L-Ara*f* residues (Mohenan, 2008). Galactan side chains are usually made of unbranched polymers of 43 – 47 β-D-1,4-Gal residues (Mohenan, 2008). The type I arabinogalactan side chains of pectin are typically made of a linear chain of β-1,4-galactan substituted with α-1,5 linked arabino-oligosaccharides (Caffall and Mohnen, 2009). The type II arabinogalactans are made of a β-1,3-D-galactan backbone substituted with β-1,6-D-galactan side chains decorated with a heterogeneous sugar mixture including arabinose and glucuronic acid residues (Sakamoto *et al*., 2013). Type II arabinogalactans are seldom found as a constituent of pectin as they are generally linked to polypeptides, forming arabinogalactan-protein (AGP) (Caffall and Mohnen, 2009). Some arabinan and galactan side chains can be substituted with ferulic acid side chains, which can dimerize to strengthen the pectin network (Zykwinska *et al*., 2005).

The RG-II class is less abundant than RG-I, but has a higher degree of structural complexity. While RG-I comprises 20 – 35 % of the total pectin mass, RG-II typically constitutes less than 10 % of the total pectic contents (Silva *et al*., 2016; Voragen *et al*., 2009). Unlike RG-I, the backbone of RG-II is made of a linear α-1,4-L-Gal*p*A residues, and does not contain Rha*p* units as a part of the basal structure (Yapo, 2011). RG-II carries four groups of side chains, designated A to D, that contain at least 13 glycosyl residues covalently linked together by more than 21 different types of glycosidic linkages (O'Neill *et al*., 2004; Ndeh *et al*., 2017). RG-II contains a heterogeneous mixture of sugar residues such as galacturonic acid, rhamnose, arabinose, galactose, D-apiose (Api), L-aceric acid (AceA), 2-keto-3-deoxy-D-*manno*-octulosonic acid (kdoA), 2-*O*-methyl L-fucose (2-*O*-Me-Fuc), glucuronic acid (GlcA), 2-*O*-methyl-xylose (2-*O*-Me-Xyl) and 3-deoxy-D-*lyxo*-heptulosaric acid (DhaA) (Yapo, 2011). The exact locations and the orientations of the side chains of RG-II relative to the backbone structure have not yet been elucidated (Yapo, 2011). In the primary cell wall, RG-II predominantly occurs as a dimer crosslinked by a borate diester (Yapo, 2011). Changes in the degree of dimerization in RG-II play a major role in structuring the pectic network by during various plant developmental stages, enabling the dynamic structural modifications that facilitate the plant cell growth (Yapo, 2011).

1.2.2 Cellulose

Cellulose is a tightly packed aggregate of linear polymers of 1,4-linked β-D-glycopyranose. The polymers aggregate to form long microfibrils (of 7,000 -12,000 glucose molecules/units) varying in length, width, and in the degree of order, depending on the plant sources (Figure 1.3) (Simoneit *et al*., 1999). For the last two decades, plant cell wall polysaccharides have been envisaged as a series of cellulose microfibrils consisting of a highly organized crystalline domain of cellulose linked to less organized, and more easily accessible amorphous regions of cellulose that are embedded into a highly cross-linked matrix of hemicelluloses and pectin (Wei *et al*., 2009). Crystalline regions of the cell wall cellulose form when adjacent cellulose microfibrils stack in a parallel manner, linked to each other through the formation of hydrogen bonds and van der Waals interactions (Wei *et al*., 2009). However, recent studies of cellulose chemical structure revealed that the "crystalline" and "amorphous" regions of cellulose microfibrils may actually be structural distortions caused by the use of harsh chemical means to extract cellulose component from the plant cell wall (Wei *et al*., 2009). Electron microscopic structural analysis of plant cell wall cellulose indicated that cellulose is cross-linked by the constituents of hemicelluloses, such as xyloglucans (Somerville *et al*., 2004). Due to the steric restraint caused by its side chain positions, only part of the xyloglucan backbone of hemicelluloses can form hydrogen bonds along the surface of cellulose microfibrils (Hayashi *et al*., 1987; Vincken *et al*., 1994). Although no direct linkage between cellulose and pectin was initially believed to exist, growing evidence suggests that the neutral sugars of pectin side chains such as

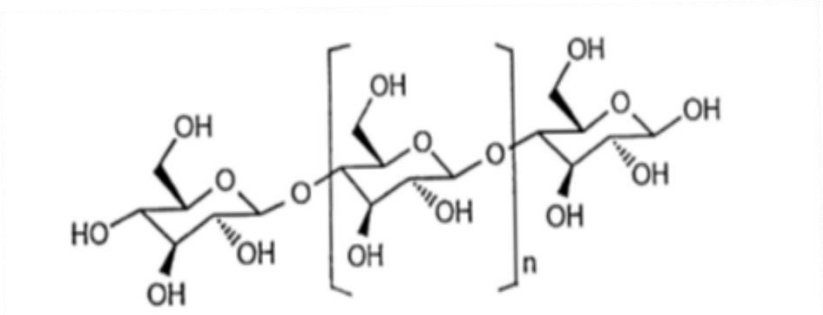


Figure 1.3 Cellulose structure (Simoneit *et al*., 1999). Copyright permission has been granted for the use of the figure.

arabinan, galactan, and arabinogalactan, can bind to cellulose, as shown in an *in vitro* binding assay using pectin and cellulose isolated from the cell walls of potato and apple (Zykwinska *et al*., 2005). The exact mechanism of cross-linking between cellulose and pectin is not yet elucidated, although hydrogen bond formation between the arabinan and galactan side chains of pectin and cellulose microfibrils is suspected (Zykwinska *et al*., 2005; Lin *et al*., 2016). The strength of pectin-cellulose interaction was found to be weaker than that of hemicelluloses-cellulose interaction, despite the abundance of arabinan and galactan side chains on pectin (Zykwinska *et al*., 2005). This may suggest that not all pectin side chains are capable of bonding to cellulose, possibly because the steric interference and bulkiness of pectin side chains discourage all side chains from engaging in hydrogen bond formation with cellulose (Zykwinska *et al*., 2005). Although yet to be discovered, the presence of covalent linkages between cellulose microfibrils and the complex side chains of pectic polysaccharides may add considerable degrees of mechanical strength and structural integrity to the weight-sustaining framework of the plant cell wall.

1.2.3 Hemicellulose

A collective term "hemicellulose" is used to describe a group of plant cell wall polysaccharides that are characterised by having β-1,4-linked backbones of glucose, mannose, and xylose (Scheller and Ulvskov, 2010). Unlike cellulose, hemicellulose is branched, and forms polymers that are shorter in length (about 200 sugar molecules/unit) (Figure 1.4) (Simoneit *et al*., 1999). The backbone of hemicellulose has a very heterogeneous chemical composition which can be categorized into various structural components including xylans, xyloglucans, mannans, galactomannans, glucuronoxylan, glucoronoarabinoxylan, β-1, 3- and β-1,4-linked glucans, depending on the type of side chains it possesses (Scheller and Ulvskov, 2010). Terminal ferulic acid and *p*-coumaric acid esterified to arabinose residues facilitate the cross-linking between hemicellulosic structures (Van Laere *et al*., 1997, Mirande *et al*., 2010). Xyloglucan is the most abundantly present form of carbohydrate in hemicellulose found in vegetables and cereal brans, consisting of a backbone of 1,4-linked β-D-glucose residues with short side chains containing xylose, arabinose, galactose, and terminal fucose (Scheller and Ulvskov, 2010). Xyloglucan in vegetables has less xylose content than cereal bran, and

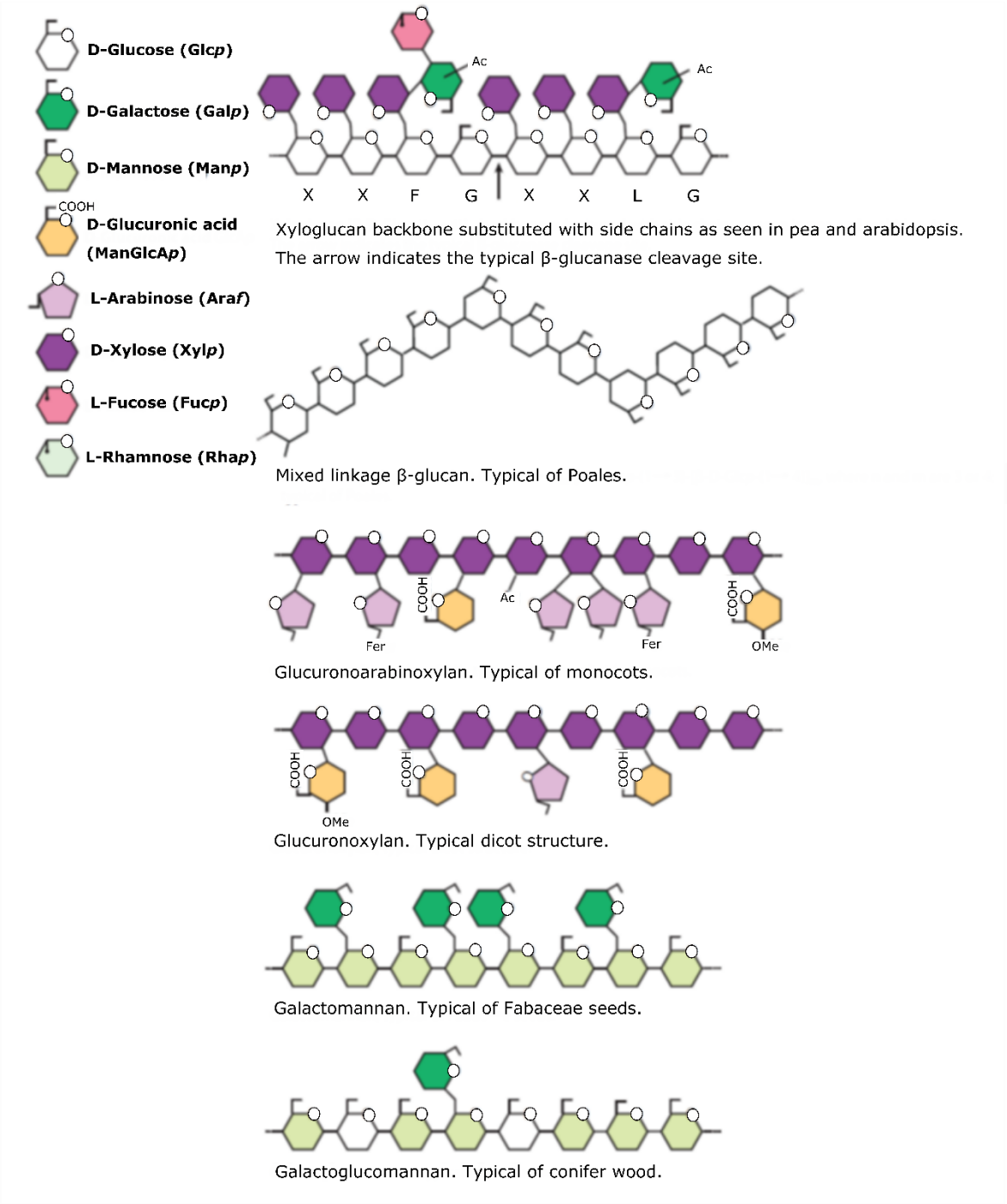


Figure 1.4 Hemicelluloses found in the plant cell walls (Scheller and Ulvskov, 2010). No copyright permission is required for the use of the figure. Symbols are used to indicate common side chains of the xyloglucan molecule. X: α-1,6-D-xylose; F: α-L-Fucose; G: glucose; L: L-β-galactose; Ac: acetylated; OMe: O-methylated; Fer: ferulic acid.

shows a smaller degree of arabinoxylan substitution (Dodd *et al*., 2011, Hopkins *et al*., 2003). Examination by rapid-freezing and deep-etching techniques (RFDE) after treatment with endoglucanase and subsequent immunogold-labelling with anti-xyloglucan antibody showed a significant decrease in cross-linking between xyloglucans and cellulose in the cell wall (Fujino *et al*., 2000). An apparent collapse of cell wall network and the removal of cross-links between cellulose microfibrils were also observed in another study which examined the onion cell wall structure using RFDE after the sequential extractions of pectic and hemicellulosic polymers (McCann *et al*., 1990). The exact nature of interaction between xyloglucan and cellulose microfibrils at an ultrastructural level is still unclear. However, the needs for harsh chemical means to disrupt the crystalline region of cellulose to dissociate xyloglucan from cellulose microfibrils suggested that parts of xyloglucan backbone of hemicelluloses may extend along the surface of crystalline cellulose through the formation of hydrogen bonds (Hayashi *et al*., 1987; Vincken *et al*., 1994). Some studies showed an alkaline-stable connection existed between RG-I side chains of pectin and up to 30 – 50 % of xyloglucan of hemicellulose, although the exact nature of the linkage is still unclear (Thompson and Fry, 2000; Popper and Fry, 2008).

1.3 Fermentability of plant cell wall polysaccharides in the human large intestine

Upon the ingestion of dietary plants, the fibrolytic bacterial populations inhabiting the lower digestive tract initially encounter the middle lamella overlaying the intercellular spaces. The erosion of pectin layers coating the outer surface of the plant cells by pectin-degrading bacteria would expose the intact plant cell walls for bacterial attachment, allowing the access of degradative enzymes at the zone of penetration (Cheng *et al.*, 1979). A more or less complete digestion of pectin by gut microflora was previously reported, indicating pectin becomes extensively fermented in the human large intestine (Cummings and Englyst, 1987; Cummings *et al*., 1979; Gibson *et al*., 1990). Compared to cellulose, pectin is more efficiently digested in the human gut, and its contribution to faecal bulking was found to be minimal (Spiller *et al*., 1980). The total faecal weight from individuals subjected to a diet supplemented with high-methoxyl pectin did not significantly differ from those subjected to a low fibre diet (Spiller *et al*., 1980; Cummings *et al*., 1979). The degree of methyl esterification appeared

to affect the rate of *in vitro* pectin fermentation by the colonic bacterial cultures in that the oligomeric breakdown products of high-methoxyl pectin accumulated at a slower rate compared to low methoxyl-pectin (Dongowski *et al*., 2000). Using *in vivo* animal models, the faecal recovery of uronic acids from low-methoxyl and high-methoxyl pectins were 19 % and 25 %, respectively (Nyman and Asp, 1982). These results indicated that the removal of methyl groups from the pectin backbone is potentially a rate-limiting step for the overall plant cell wall degradation process, suggesting the presence of pectin de-esterifying bacteria may promote a rapid colonic digestion of plant materials.

An extensive disintegration of cellulose is desirable to expose the interior parts of the plant cell wall matrix to the enzymatic attack by various gut microorganisms. However, despite its structural simplicity, cellulose is highly recalcitrant and insoluble, thus efficiently resists the degradation and fermentation by gut microorganisms in the lower intestine. Due to the low digestibility and high water-absorbing capacity, cellulose quickly transits to the lower intestine with a substantial amount of its polymers still intact or only partially degraded, thus significantly contributing to the faecal bulking (Spiller *et al*., 1980). The recovery of isotopically labelled cellulose excreted in human faeces showed that only 38% of ingested cellulose became digested and absorbed, suggesting that a low nutritional value was generated for the human host from the consumption of cellulose (Kelleher *et al*., 1984). Another study based on a diet supplemented with a mixed fibre preparation showed a mean faecal excretion of ingested cellulose was 59%, although a considerable degree of individual variations between subjects were observed (Prynne and Southgate, 1979).

Xylan, a major component of hemicellulose, is the primary polysaccharide constituent of cereal grains consume by humans (Dodd *et al*., 2011). The structural complexity of hemicellulose, as well as its close association with cellulose and pectin makes hemicellulose a difficult substrate to access, requiring a large repertoire of enzymatic activities to break the various chemical linkages (Dodd *et al*., 2011). Similar to pectin, an almost complete fermentation of hemicellulose occurred in the human large intestine, indicating that very efficient microbial degradation of hemicellulose takes place (Holloway *et al*., 1978; Slavin *et al*., 1981). From these studies, the digestibility of hemicellulose was reported to range between 72 – 96 % (Holloway *et al*., 1978; Slavin *et al*., 1981).

1.4 Microbial strategies of complex polysaccharide degradation

Rapid and efficient degradation of plant-derived polysaccharides occurs in the human large intestine. The process of plant cell wall degradation is initiated by particle-associated saccharolytic bacteria closely adhered to the substrate surfaces (Flint *et al*., 2012). These "primary degraders" or "keystone species" often produce highly specialized enzymatic machineries to form a physical adherence to the plant cell wall surfaces prior to penetrating and disrupting the polysaccharide layers (Flint *et al*., 2012). Currently, two contrasting paradigms of microbial polysaccharide utilization have been established, although detailed functional and molecular study of these systems has been undertaken for only a few bacterial species (Flint and Bayer, 2008). In this review, the cellulosome system of Gram positive cellulolytic bacteria and the starch degradation system of Gram negative *Bacteroides thetaiotaomicron* are discussed briefly. The generalistic *Bacteroides* polysaccharide utilization system contrasts with the single substrate-targeting strategy adopted by the cellulose-degrading specialist *Ruminococcus flavefaciens*.

1.4.1 The cellulosome system

Since an external secretion of cellulases is not an ideal enzymatic system in anaerobic environments such as the human gut or the animal rumen in that hydrolysed products may quickly diffuse into the aqueous surroundings before their uptake, some cellulolytic members of the phylum *Firmicutes* developed the cellulosome system to enable a tight adhesion to the cellulose substrates (Ransom-Jones *et al*., 2012). (Figure 1.5). The cellulosome is a bacterial cell surface-associated multi-protein complex, composed of catalytic enzymes and non-catalytic structural proteins (Alber *et al*., 2009). The cellulosome system was first described in *Clostridium thermocellum*, a thermophilic, anaerobic bacterium whose great efficiency in cellulose degradation has been exploited to develop a next-generation, environmental-friendly technology for processing agricultural lignocellulosic biomass into useful byproducts such as ethanol (Bayer *et al*., 1983; Akinosho *et al*., 2014). In gut/rumen environments, the cellulosome system has been exclusively found in Gram positive *Ruminococcus* spp. of the family *Ruminococcaceae*. In recent years, *R. flavefaciens*, an isolate from the animal rumen, has been used as a model organism for modelling the microbial system for gastrointestinal

cellulose degradation (Flint *et al*., 2008b; Fontes and Gilbert, 2010). The cellulosome system of *R. flavefaciens* is arguably the most elaborate and specialized microbial cellulose degradation mechanism described to date (Fontes and Gilbert, 2010). *R. flavefaciens* produces a wide range of catalytic enzymes that catalyse the cleavage of glycosidic bonds linking various types of neutral sugars found in cellulose, hemicellulose, and pectin. Some enzymes possess bi- or tri-catalytic domains, enabling the simultaneous hydrolysis and de-esterification of more than one type of carbohydrate substrates (Flint *et al*., 2008b). A large number of catalytic enzymes found in *R. flavefaciens* have a multi-domain organization in which catalytic modules are appended to non-catalytic structural and substrate recognition domains (Flint *et al*., 2008). Catabolic enzymes such as

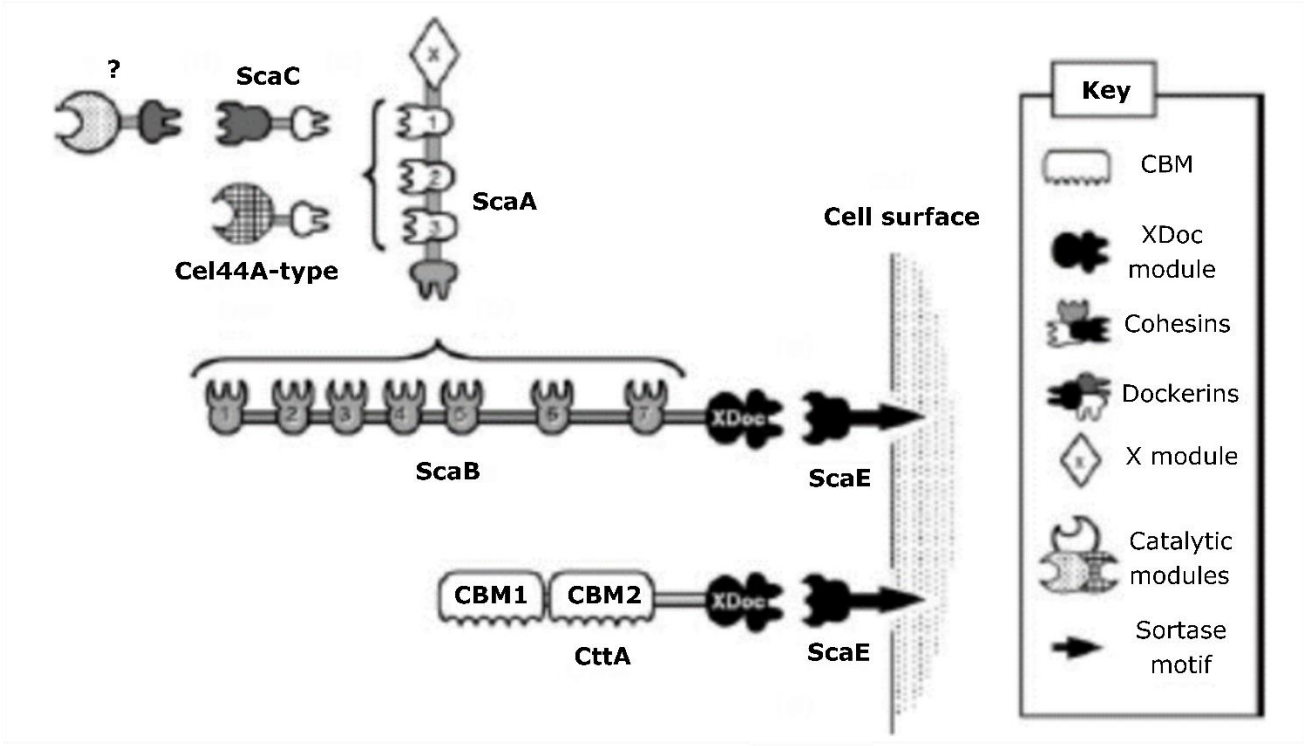


Figure 1.5 Schematic diagram of cellulosome system in Ruminococcus flavefaciens strain 17 (Rincon *et al*., 2007). No copyright permission is required for the use of this figure.

cellulases, xylanases, and esterases often contain C-terminal dockerin sequences which interact with cohesins modules present on large structural proteins called scaffoldins, resulting in the formation of multi-enzyme complexes into which various degradative enzymes become incorporated (Flint *et al*., 2008b). Scaffoldins are responsible for bringing the cohort of enzymatic fractions close to the surface of cellulose. Complex cellulosomes are often made of 3 major types of scaffoldins, primary, anchoring, and adaptor scaffoldins (Artzi *et al*., 2017). Primary scaffoldins contain multiple cohesin units to facilitate the interactions with dockerins, whereas anchoring scaffoldins serve to dock the cellulosome complexes onto the bacterial cell surface. Adaptor scaffoldins usually connect two scaffoldins together to expand the catalytic diversity of cellulosomes by recruiting more enzymes into the protein complex (Artzi *et al*., 2017). The close proximity of cellulosome complexes to the substrate is maintained by the dockerin-associated carbohydrate binding modules (CBMs) which are non-catalytic protein domains with binding specificities for various carbohydrate substrates. CBMs will be discussed in more detail in section 1.5 of this review. Note that the presence of dockerins and cohesins is not a unique feature of the cellulosome system, as these protein domains were found in all domains of life (Peer *et al*., 2009). Approximately 14 % of available bacterial genomes contained either cohesin and/or dockerin modules, although these modules are usually present in small numbers, and mostly not in dockerin-cohesin pairs (Peer *et al*., 2009). The majority of dockerins and cohesins found in nature are components of enzymes serving currently unknown functions, with very few involved in the cellulosome assembly (Fontes and Gilbert, 2010).

The scaffoldin components of the cellulosomes of *R. flavefaciens* strains 17 and FD-1 are encoded by the 17 kb *sca* gene cluster (*sca*C-*sca*A-*sca*B-*ctt*A-*sca*E) (Jindou *et al*., 2008). ScaA and ScaB proteins are both primary scaffoldins containing cohesin modules. Depending on the strains, ScaA carries 2 – 3 cohesin repeats, whereas ScaB may contain up to 9 contiguous cohesin units (Jindou *et al*., 2006). Discovered on the upstream of *sca*A and *sca*B genes, *sca*C encodes a 31.8 kDa adaptor scaffoldin which expands the repertoire of degrading proteins to participate in cellulosome functions (Rincon *et al*., 2004). Cohesins of ScaB are recognized by the dockerins of ScaA, which in turn has a binding affinity for ScaC (Flint *et al*., 2008b). The cellulosome anchorage to the bacterial cell wall is mediated

by ScaE which interacts with ScaB via cohesion-dockerin interactions (Rincon *et al*., 2005). ScaE gene sequence possesses a cohesion domain at N-terminus, and LPXTG-like motif (a sorting signal) at C terminus (Rincon *et al*., 2007). LPXTG motif is commonly found in Gram positive bacteria, and it is recognized prior to proteolytic cleavage as a signal by sortase which mediates transpeptidation mechanism to dock their cellulosome onto the bacterial cell wall (Fontes and Gilbert, 2010). Among the members of the *Clostridium* cluster III such as *C. thermocellum*, *Acetivibrio cellulolyticus*, and *Clostridium clariflavum*, the peptidoglycan-binding S-layer homology domain repeats are used instead of LPXTG motif to attach the cellulosome assembly onto the bacterial cell surface (Artzi *et al*., 2017). Using a mutant strain of *R. flavefaciens* that had lost the ability to degrade cotton cellulose, researchers have identified *ctt*A gene upstream of *scaE* gene (Rincon *et al*., 2007). Sequence analysis of *ctt*A region revealed that it is a non-catalytic protein domain containing two novel carbohydrate binding modules (CBMs) which bind to microcrystalline cellulose (Rincon *et al*., 2007). CttA is required for degrading highly recalcitrant forms of cellulose, such as those found in the plant cell walls (Rincon *et al*., 2007). Despite the large repertoire of hemicellulases and pectinases produced by *R. flavefaciens*, the species was not isolated using hemicellulose and pectin, and did not show a significant growth using these substrates (Hespell *et al*., 1987). Therefore, the principle role of hemicellulases and pectinases integrated in the cellulosome appeared to be a systematic deterioration of the polysaccharides surrounding the cellulose microfibrils, thus making cellulose fibres more accessible to the enzymatic hydrolysis by endoglucanases.

Despite the efficiency of the system, the cellulosome is considered an unconventional system for polysaccharide degradation in gut environments since the presence of cellulosome has been confirmed in a very few *Ruminococcus* spp. (Flint *et al*., 2012). Although the frequent occurrence of *Ruminococcus* spp. in faecal particulates enriched with plant cellulose suggested that *Ruminococcus* spp. may be the principle cellulolytic organisms in the human gut, the first and the only evidence to support the existence of true cellulosomes in the human gut was the discovery of *Ruminococcus champanellensis* (Walker *et al*., 2008; David *et al*., 2015). The cellulosome system of *R. champanellensis* closely resembled that of *R. flavefaciens* in the overall architecture, fundamental

structural elements, and genetic organizations (David *et al*., 2015). *Ruminococcus bromii*, a keystone bacterial species with an exceptional ability to degrade resistant starch and starch particles in the human gut, uses the dockerin-cohesin interactions to mediate the anchorage of amylases onto the ScaE-like proteins present on the bacterial cell walls (Ze *et al*., 2015). This 'amylosome' of *R. bromii* showed for the first time that non-cellulolytic ruminococci can also rely on the dockerin-cohesin interactions to augment their fibrolytic capacity, suggesting variations in cellulosome organization may result in the formation of simple cellulosome structures conferred with an unconventional substrate specificity (Ze *et al*., 2015).

1.4.2 The polysaccharide utilization locus (PULs)

B. thetaiotaomicron is a versatile glycan foraging bacterium of the human gut microbiota. *B. thetaiotaomicron* can grow on a wide range of plant polysaccharides, and shows a high degree of nutritional flexibility to adapt to the substrates available in the gut (Koropatkin *et al*., 2008). The genome analysis of *B. thetaiotaomicron* revealed a comprehensive carbohydrate utilization system called the 'Sus' (starch utilization system or sequestration system) which is used to degrade and utilize starch. The genes encoding the proteins of the Sus system are organized into gene clusters called polysaccharide utilization loci (PULs). A minimum single unit PUL may possess genes encoding a set of enzymes that degrade a specific form of glycan substrate or closely related group of glycans, an outer membrane-associated substrate binding protein, and a substrate-specific receptor (Martens *et al*., 2011). The Sus locus typically consists of a cluster of 8 genes, *sus*RABCDEFG, each encoding the components of a multi-protein complex that forms a cell-associated apparatus which binds, cleaves, and transports oligomeric starch into the intracellular regions where the imported substrates are enzymatically degraded (Figure 1.6) (Martens *et al*., 2009). SusC is an outer membrane-spanning β-barrel TonB-dependent receptor family that carries out a high affinity binding and the intake of substrate into the periplasmic region using the proton motive force (Martens *et al*., 2009). In general, active transport system is required especially when the substrates are poorly permeable through porins (e.g. large oligosaccharides). SusDEFG are lipoproteins containing bacterial signal peptidase II recognition motif (Marten *et al*., 2009). The concerted efforts of SusCDEF mediate

substrate binding close to the cell surface (Marten *et al*., 2009). SusC and SusD directly interact with each other, and are considered as a minimum unit of starch binding complex; disruptions in SusC and SusD completely abolished starch binding (Cho and Salyers, 2001). SusC and SusD are responsible for approximately 60% of starch binding observed in wildtype *B. thetaiotaomicron* (Martens *et al* 2009). SusD mediates starch and malto-oligosaccharide binding via a single starch-binding site (Koropatkin *et al*., 2008). The analysis of SusD crystal structure revealed that the protein has a unique α-helical fold which recognizes the 3D helical structure of starch molecules instead of the relative spatial arrangement of individual glucose residues (Koropatkin *et al*., 2008). SusE and SusF are surface-exposed lipoproteins containing multiple starch-binding domains which facilitate a high-affinity substrate binding by *B. thetaiotaomicron* (Cameron *et al*., 2012). Although not essential for growth, the inclusion of SusE in SusCDE complex increased the starch binding affinity to more than 80%, while the inclusion of SusF in SusCEDF complex completely restored the starch binding affinity

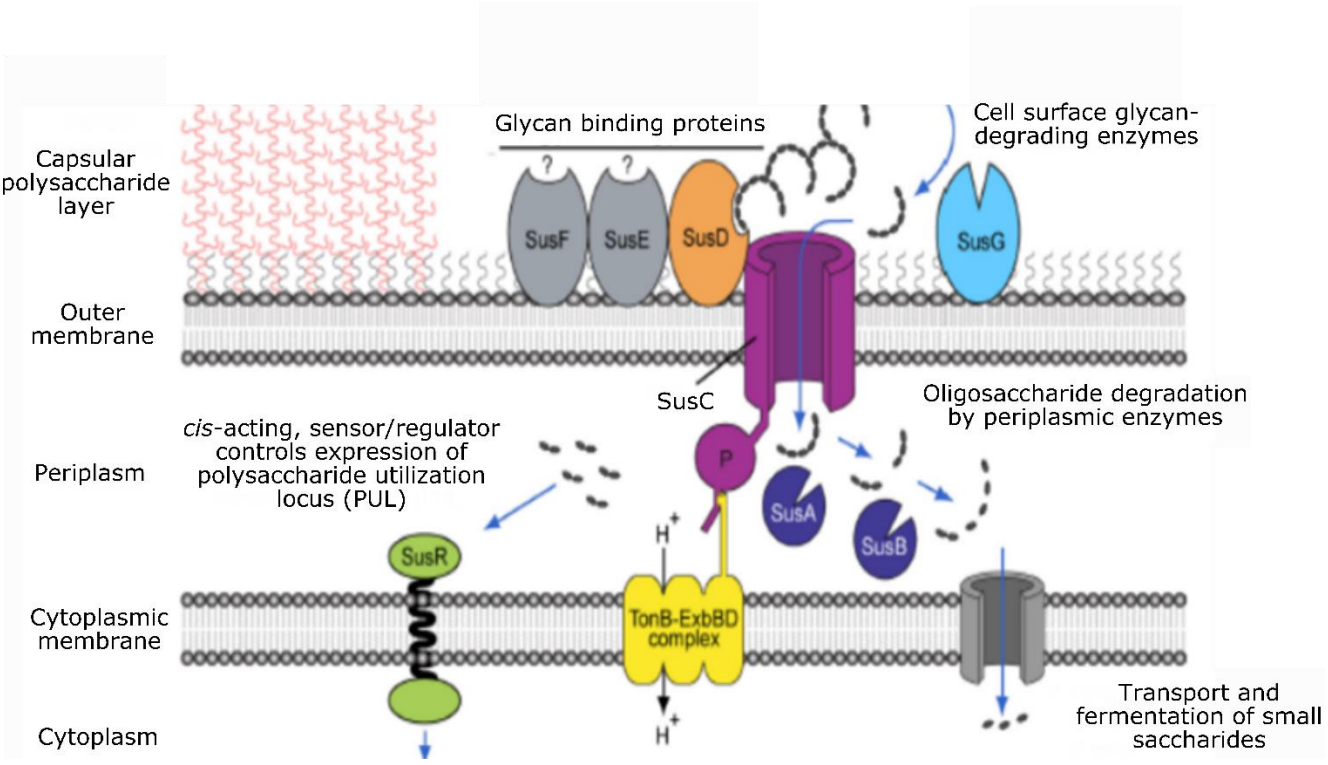


Figure 1.6 Schematic diagram of starch utilization system (Sus) of B. thetaiotaomicron (Martens *et al*., 2009). No copyright permission is required for the use of this figure.

to 100% (Martens *et al*., 2009, Shipman *et al*., 2000). SusG is an endo-acting starch-degrading enzyme tethered to the outer membrane protein (Shipman *et al*., 1999). Sharing a high sequence similarity to α-amylase, the starch-degrading ability of SusG is essential for the organism's growth on long chain starch molecules (Shipman *et al*., 1999). It is speculated that the endo-acting SusG may introduce internal cuts within the bound starch, generating shorter oligosaccharides that can be transported through SusC porins (Martens *et al*., 2009). Once in the periplasm, glycoside hydrolases (GHs) SusA and SusB degrade the oligosaccharides into simpler sugars before their transport into the cytosol. SusA is a neopullulanase homologue which is a typical starch-hydrolysing protein (Kitamura *et al*., 2008). SusB is an α-glycosidase belonging to GH family 97 (Kitamura *et al*., 2008). SusR is a transcriptional level controller for SusABCDEFG expression (Martens *et al*., 2009). While SusR is expressed constitutively in wildtype, increasing the expression of SusR increased the expression of SusA, B, and C as well as the overall growth rate of *B. thetaiotaomicron* on starch (DElla *et al*., 1996).

B. thetaiotaomicron and other *Bacteroides* spp. use the Sus-like systems to degrade a wide range of plant cell wall polysaccharides and the host-derived mucin *O*-glycans. The Sus paradigm is a typical example of a highly conserved carbohydrate utilization system showing a diversity in saccharolytic ability due to the variations in the catalytic components of the multi-protein complex. For example, the transcriptomic profiling of *B. thetaiotaomicron* grown on pectic glycans revealed that at least 10 PULs showed a significantly induced expression, indicating the enzymes encoded by pectin-specific PULs were activated in response to the availability of pectin substrates (Martens *et al*., 2011). Unlike *B. thetaiotaomicron* that cannot degrade hemicellulose, the PULs of *B. ovatus* were enriched with genes enabling the metabolism of all types of hemicelluloses (Martens *et al*., 2011). *B. ovatus* PULs contained at least a pair of genes encoding homologs of *B. thetaiotaomicron sus*C/D and various enzyme families implicated in degrading xylan, xyloglucan and galactomannan, which showed increased expressions in the presence of hemicellulose (Martens *et al*., 2011). In *Bacteroides cellulosilyticus*, the expression of *sus* gene homologues was significantly upregulated by the consumption of high plant polysaccharide diet (McNulty *et al*., 2013). Although some strains of *B.*

cellulolyticus were capable of degrading cellulose, their preferred source of carbohydrate was xylan, a major component of hemicelluloses. Among the most highly expressed carbohydrate-degrading enzymes of *B. cellulolyticus* in response to the plant polysaccharide-rich diet, many belonged to the enzyme families required for degrading xylan and pectin (McNulty *et al*., 2013). *B. xylanisolvens* showed a higher xylanolytic activity than *B. ovatus*, and was able to degrade xylans from diverse sources (Mirande *et al*., 2010). *B. xylanisolvens* genome also contained over 100 genes coding for enzymes potentially involved in the pectin degradation (Despres *et al*., 2016). The genes coding for pectate lyases and polygalacturonases targeting the pectin backbone structure were found in 5 out of 6 PULs which showed a significant overexpression when *B. xylanisolvens* was grown using citrus and apple pectin (Despres *et al*., 2016).

The homologues of Sus system genes were also found in the rumen/gut starch- and hemicelluloses-degrading species of *Prevotella* (previously classified as *Bacteroides*), indicating a deep evolutionary root and the distribution of the system among the members of *Bacteroidetes* occurring through an extensive gene transfer and gene duplications (Dodd *et al*., 2010).

1.5 CAZy database and CAZymes

The Carbohydrate-Active Enzyme database (CAZy; http://www.cazy.org) is a web-based and continuously curated collection of enzymes involved in the assembly and breakdown of complex carbohydrates and glycoconjugates (Lombard *et al*., 2014; Cantarel *et al*., 2009). Collectively called Carbohydrate-Active enZymes (CAZymes), these enzymes are classified into families and subfamilies based on amino acid sequence similarity. The CAZy database provides a comprehensive resource for establishing the sequence-to-specificity relationships in CAZymes, presently covering $>$ 370 families of CAZyme categories including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), glycosyltransferases (GTs), and carbohydrate-binding modules (CBMs).

GHs are present in all domains of life, constituting the largest proportion of currently described CAZyme families (Cantarel *et al*., 2009; Naumoff, 2011). GHs catalyse the hydrolytic and/or transglycolytic cleavage of *O*-, *N*-, and *S*-linked glycosidic bonds (Naumoff, 2011). GHs can be either

exo- or endo-acting depending on the topology of active sites and where the cleavage occurs. Exo-acting enzymes use a pocket topology to recognize the non-reducing end of a saccharide and to cleave the terminal sugar, while endo-acting enzymes use open cleft or tunnel structures to feed the polysaccharide through the active site to introduce random cleavages along the polysaccharides to generate oligomeric products with varying sizes (Davies and Henrissat, 1995). Currently, 145 families have been assigned into GH, accounting for > 40 % of all CAZymes listed on the CAZy database.

PLs specifically target polysaccharides containing uronic acids (galacturonate, glucoronate, iduronate, mannuronate, or guluronate) for cleavage via β-elimination mechanism to generate oligomers with an unsaturation between C4 and C5 positions between adjacent uronic acid residues (Hugouvieux-Cotte-Pattat *et al*., 2014). The structural constituents of pectin, alginate, chondroitin, heparin, and hyaluronate are often targeted by PLs (Lombard *et al*., 2010). Similar to GHs, PLs typically act using either endo- or exo-type mechanisms. Some families of PLs, especially several groups of pectate lyases, have a mandatory requirement of divalent cations (Ca^{2+} or Mg^{2+}) for activity (Hugouvieux-Cotte-Pattat *et al*., 2014). Currently there are 26 families of PLs available on the CAZy database.

CEs catalyse the *O*- or *N*-deacylation of substituted esters or amides in mono-, oligo-, and polysaccharides, enabling the subsequent cleavage of de-esterified polysaccharides by GHs and PLs (Biely, 2012; Cantarel *et al*., 2009). CEs have a broad range of substrate specificities including acetylxylan esterases, acetyl esterases, pectin acetyl esterases, pectin methyl esterases, chitin deacetylases, peptidoglycan deacetylases, feruloyl esterases, glucuronoyl esterases, and deacetylases acting on amino sugar derivatives (Biely, 2012). Currently there are 16 different CE families listed in the CAZy database.

The non-catalytic CBMs augment the catalytic ability of CAZyme domains by specifically targeting the substrate and retaining the enzyme-substrate interaction for a prolonged period of time to increase the protein concentration on the polysaccharide surface (Boraston *et al*., 2004). Catalytic enzymes are often found appended to CBMs which display substantial variations in ligand specificity. Although the majority of CBMs described to date bind to plant structural polysaccharides, some have affinity for storage polysaccharides and the cell-surface glycans (Boraston *et al*., 2004). The exact

mechanisms which facilitate the CBM-carbohydrate interactions are currently not well understood. However, the positional arrangement of aromatic amino acid residues (Trp, Tyr, and Phe) in the binding sites to form hydrophobic surfaces, as well as the chelation of calcium ions to form polar surfaces to directly bind targeted ligands were proposed to play significant roles during the CBM-carbohydrate interaction (Boraston *et al*., 2004; Gilbert *et al*., 2013). The nature of CBMs is described as 'promiscuous' and 'multivalent' to indicate that some CBMs recognize multiple types of ligands, and the weak strength of CBM-carbohydrate interaction can be compensated by the use of multiple CBM domains in a single protein or multiple clustered binding sites in individual CBMs to increase the overall affinity for the substrate (Boraston *et al*., 2004). Currently 81 families of CBMs are listed in the CAZy database.

1.6 Pectin-degrading enzymes

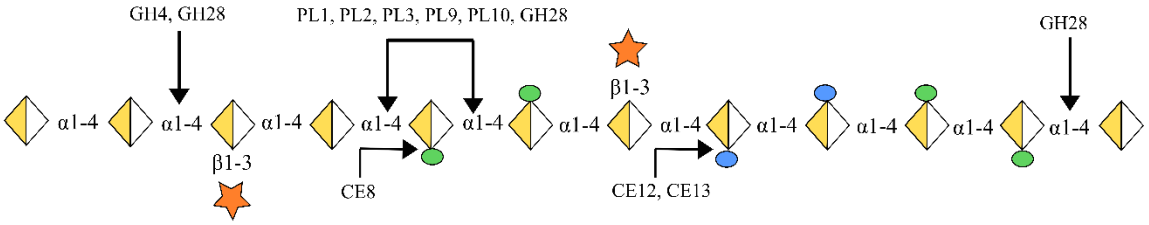
Pectinolytic enzymes are abundantly produced in higher plants and microorganisms. These enzymes are of prime importance for controlling the cell wall expansion during plant growth and fruit development, and for catalysing the depolymerisation of the plant cell wall polysaccharides for sugar utilization by pathogenic and colonic microorganisms. Purified fungal pectinases, most commonly from *Aspergillus niger*, are commercially valuable for their use in food industry to speed up the juice extraction process from fruit and vegetable pulps (Kashyap *et al*., 2001). Due to the structural complexity of pectin, a large arsenal of degradative enzymes is required to selectively target 17 different sugar types and at least 21 distinct glycosidic linkages present in pectic polysaccharides (Ndeh *et al*., 2017). This review will focus on assessing the enzymes involved in the degradation of HG, RG-I, and RG-II, of which together constitute up to 90 % of the total pectin mass. The list of pectin-degrading enzymes targeting glycosidic linkages of HG, RG-I, and RG-II regions of pectin is summarized in Appendix 1 and Figure 1.7. The following sections (1.6.1 – 1.6.4) summarize relevant literature on bacterial pectin-degrading CAZymes which may help the readers to understand the findings from this study.

1.6.1 Homogalacturonan (HG)-degrading enzymes

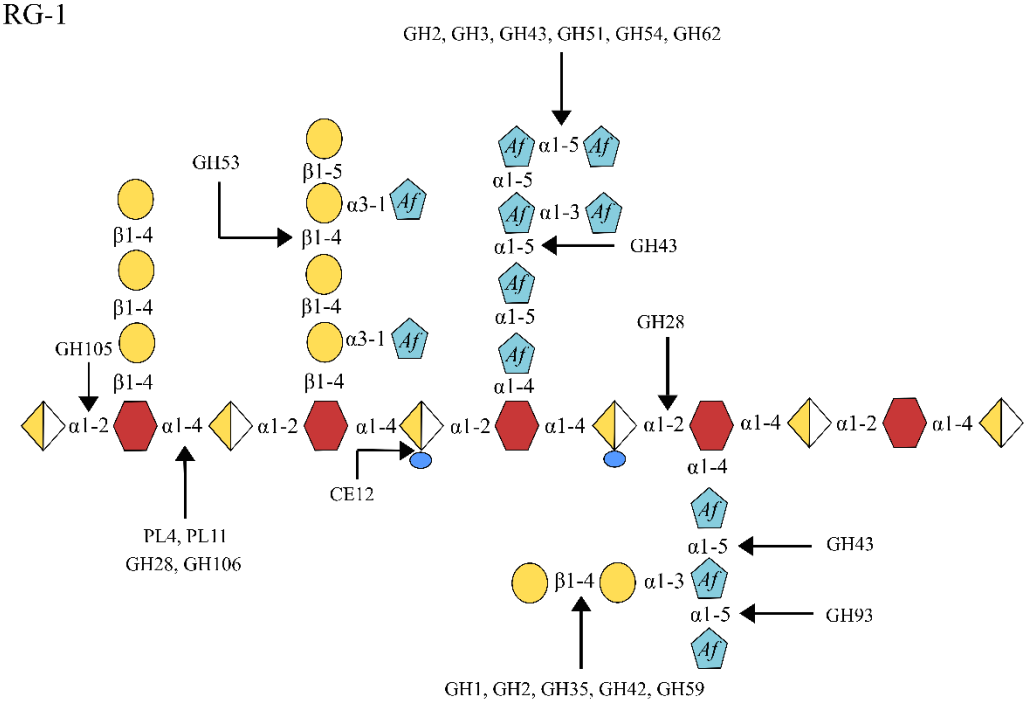
1.6.1.1 Pectate lyases

HG-degrading enzymes are mostly active on the HG backbone of pectin, and catalyse the specific cleavages of D-Gal*p*A-α-1,4-D-Gal*p*A linkages. Currently, the majority of HG-degrading enzymes belong to various families and sub-families of pectate lyases which have been extensively studied from bacterial (e.g. *Dickeya* and *Pectobacterium* spp.) and fungal pathogens that are known to have devastating effects on crop plants (Hugouvieux-Cotte-Pattat *et al*., 2014). As a result, despite the small number of families, PLs constitute the highest proportion of biochemically characterized enzymes in the CAZyme database (Cantarel *et al*., 2008). Depending on the cleavage pattern and substrate specificities, pectate lyases can be grouped into exo- or endo-acting pectate lyases, exo- or endo-acting pectin lyase, and exo- or endo-acting rhamnogalacturonan lyases. Endo-acting lyases cleave randomly inside polygalacturonates or RG-I chain to create products varying in sizes, whereas exo-acting lyases cleave the terminal end of polygalacturonates or RG-I chain to generate 4,5-unsaturated di- or trigalacturonates (Hugouvieux-Cotte-Pattat *et al*., 2014). Pectin lyases (also known as polymethyl galacturonate lyases) prefer to attack highly methylated (~98 %) regions of polygalacturonate. Although pectin lyases are capable of splitting pectin polymers independent of the degree of methylation, such lyases are mostly found in fungi (Dongowski *et al*., 2000). Only few bacterial species such as *Bacillus* spp. and *Pseudomonas marginalis* are currently known to produce pectin lyases (Sakiyama *et al*., 2001; Kashyap 2000; Hayashi 1997). According to the CAZy database, HG-degrading pectate lyases are divided into five families, PL1, PL2, PL3, PL9, and PL10, with the addition of PL22 which exclusively acts on short galacturonate oligomers. Families of PL fall into different protein-folding categories, such as parallel β-helix (PL1, PL3, and PL9), $(\alpha/\alpha)_7$ barrel (PL2), $(\alpha/\alpha)_3$ barrel (PL10), or β-propeller (PL22) (Hugouvieux-Cotte-Pattat *et al*., 2014). Although pectate lyases generally prefer unmethylated or partially methylated polygalacturonates, some families or subfamilies of pectate lyases recognize and act on polygalacturonate with varying degrees of methylation (Hugouvieux-Cotte-Pattat *et al*., 2014). Aside from being the cause of steric hindrance to pectin-degrading enzymes, methyl groups on HG also alter the overall surface charge of pectin by neutralizing the negative charge on individual galacturonate monomers (Herron *et al*., 2000). Plant

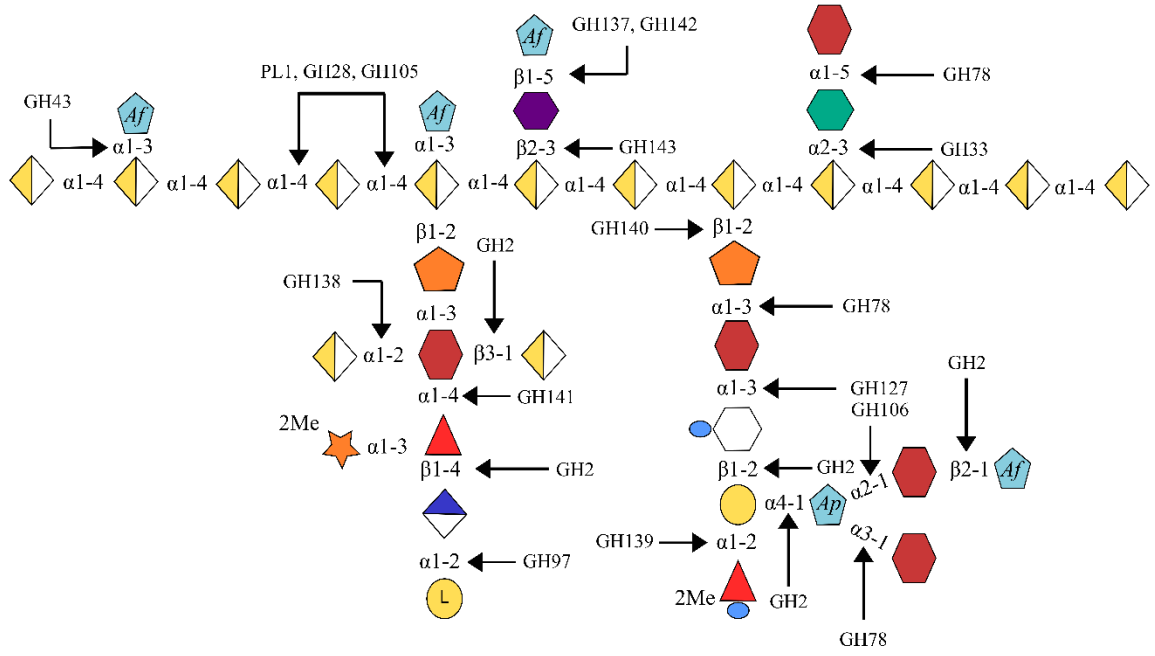
HG


GH4, GH28
PL1, PL2, PL3, PL9, PL10, GH28
GH28
β1-3
α1-4
β1-3
CE8
CE12, CE13

RG-1


GH2, GH3, GH43, GH51, GH54, GH62
GH53
β1-5
α3-1
Af
α1-5
α1-3
GH43
β1-4
GH105
GH28
α1-2
α1-4
CE12
PL4, PL11
GH28, GH106
GH93
GH1, GH2, GH35, GH42, GH59

RG-I1


GH43
PL1, GH28, GH105
GH137, GH142
GH78
GH143
GH133
α1-3
β1-5
β2-3
α2-3
α1-4
GH140
β1-2
GH2
GH138
α1-2
β3-1
GH141
2Me
β1-4
GH97
GH127
GH106
β2-1
α2-1
α4-1
Ap
GH139
α3-1
GH78

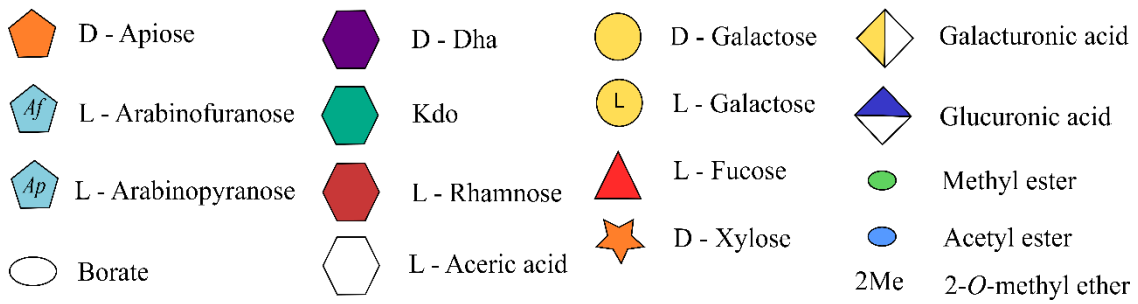


Figure 1.7 Activities of pectin-degrading enzymes on HG, RG-I, and RG-II. The diagram of RG-II structure was adopted from Ndeh *et al*., 2017. Copyright permission has been granted for the use of this figure.

pathogens produce a heterogeneous population of pectate lyases with varying substrate specificities to orchestrate the degradation of HG which significantly changes its degree of esterification over the course of plant developmental stages. Unlike most enzyme families that display a high degree of structural conservation in the active sites and the regions surrounding the active sites, the variations in the length, amino acid sequences, and protein-folding topology found within the immediate region around the catalytic sites of pectate lyases were proposed as structural determinants for distinguishing methylated and non-methylated substrates (Herron *et al*., 2000). The comparison of the substrate-binding sites in pectin and pectate lyases using 3D structural modelling showed that pectin lyases possessed a hydrophobic pocket primarily consisting of an unusual cluster of tryptophan and tyrosine residues to accommodate a methyl group added to an uronic acid moiety (Herron *et al*., 2000). With the exception of PL2 which uses Co^{2+}, Mn^{2+}, and Ni^{2+}, most families of pectate lyases require Ca^{2+} as an essential cofactor to mediate the interaction between protein and sugar substrate (Creze *et al*., 2008). Ca^{2+} is readily obtainable from the middle lamella and the plant cell wall in which the localized calcium concentration may reach up to 1.0 mM (Herron *et al*., 2000). Ca^{2+} ions confer a positive charge to the surface of the groove surrounding the Ca^{2+}-binding sites, possibly contributing to the stabilization of a negative intermediate, or to the transfer of a proton to the target glycosidic bond to promote the acidification at C-5 position (Herron *et al*., 2000; Hugouvieux-Cotte-Pattat *et al*., 2014). In contrast to pectate lyases, pectin lyases that target fully methylated pectate do not usually require cations for activity as the positive charge is conferred by Arg-176 residue conserved within the pectin lyase subfamilies (Herron *et al*., 2000; Hugouvieux-Cotte-Pattat *et al*., 2014).

PL1

PL1 represents the largest pectate lyase family which is divided into 11 subfamilies. PL1 preferably targets low-methylated HG (Hugouvieux-Cotte-Pattat *et al*., 2014; Shevchik *et al*., 1997). PL1 pectate lyases from *Dickeya dadantii* have been extensively characterized. *D. dadantii* produces 6 secreted PL1s encoded by genes organized in two gene clusters, *pel*A-*pel*E-*pel*D and *pel*B-*pel*C-*pel*Z, and an outer membrane-associated PL1 encoded by *pnl*H (Hugouvieux-Cotte-Pattat *et al*., 2014; Ferrandez and Condemine, 2008). The 5 major PL1s, PelA, PelB, PelC, PelD, and PelE have endo-acting activities with an absolute requirement for Ca^{2+} ions (Tardy *et al*., 1997). PelB and PelC which share 84 % sequence identity are most active on partially methylated pectin with up to 22 % DE (Tardy *et al*., 1997). On the other hand, PelA, PelD, and PelE that share 60 – 80 % sequence homology retain 80 % of enzymatic activities towards low-methoxyl pectin with up to 7 % DE (Tardy *et al*., 1997). *D. dadantii* also produces secondary endo-pectate lyases PelL and PelZ which belong to PL9 and PL1 families, respectively. Similar to PelB and PelC, PelZ is an endo-acting extracellular pectate lyase mostly targeting low-methoxyl pectin, although PelZ shows a very low catalytic activity compared to the major PL1s (Pissavin *et al*., 1998). Each *pel* gene is regulated under an individual transcriptional control, indicating PL1s with varying substrate specificities may be differentially expressed to adapt to the changes in pectin structures (Tardy *et al*., 1997). PL1 family of pectate lyases are abundantly present in a wide range of ecological niches including plant pathogens such as *D. dadantii* (8 PL1s) and *Pectobacterium atrosepticum* (6 PL1s), a plant symbiont *Paenibacillus mucilaginosus* (12 PL1s), a soil bacterium *Sorangium cellulosum* (6 PL1s), and the gut symbionts *B. thetaiotaomicron* (5 PL1s) and *B. ovatus* (9 PL1s) (Hugouvieux-Cotte-Pattat *et al*., 2014).

PL9

PL9 family includes both endo- and exo-acting pectate lyases found in *D. dadantii*, *Bacillus* sp. and few fungi (Hassan *et al*., 2013; Jenkins *et al*., 2004; Hla *et al*., 2015). Some environmental isolates of cellulolytic *Clostridium* spp. such as *Clostridium stercorarium, C. thermocellum,* and *Clostridium cellulovorans* are also known to produce small numbers of PL1 and PL9 pectate lyases possibly to facilitate the extraction of cellulose polymers by disrupting the cellulose-pectin interactions in the

plant cell wall (Hugouvieux-Cotte-Pattat *et al*., 2014). Currently, 10 bacterial PL9 have been biochemically characterized. *D. dadantii* is known to produce three PL9 pectate lyases, including endo-acting PelL, PelN, and exo-acting PelX (Hassan *et al*., 2013). The secreted endopectate lyase PelL is a low active, Ca^{2+}-dependent enzyme which converts polygalacturonate with a degree of polymerization of $>$ 5 GalA residues to produce di- and trigalacturonates (Roy *et al*., 1999). PelN prefers to act on polygalacturonate, although it retains $>$ 50 % of catalytic activity against pectin with 91 % methylation (Hassan *et al*., 2013). As opposed to most pectate lyases with an absolute requirement for Ca^{2+}, PelN activity becomes stimulated in the presence of an unusual cation cofactor Fe^{2+} (Hassan *et al*., 2013). As suggested by its periplasmic localization, PelX showed maximum activity on GalA oligomers with 4 – 6 residues, liberating unsaturated digalacturonate as the end-product (Shevchik *et al*., 1999b). PelX activities against polygalacturonate and pectin with up to 22 % methylation were observed, although reaction rates were slower compared to oligomers (Shevchik *et al*., 1999b). While PelX activity was weakly stimulated using Ca^{2+}, the inhibitory bivalent cations such as Mn^{2+}, Co^{2+}, Ni^{2+}, and Cu^{2+} also weakly activated PelX, indicating the enzyme had unusual cofactor requirements (Shevchik *et al*., 1999b).

1.6.1.2 Glycoside hydrolases

GH28

GH28 is found in all domains of life and is one of the largest family of glycoside hydrolases described to date (Abbott and Boraston, 2007b). GH28 belongs to the GH-N clan characterized by the right-handed $(\beta)_3$ solenoid fold composed of four parallel sheets connected by ten complete turns (Abbott and Boraston, 2007b; Naumoff 2011). In *Enterobacteriaceae*, GH28s are divided into either exo-acting or endo-acting classes depending on the mode of action. The active site of endo-acting enzymes forms an open-ended channel structure in which the reducing ends of polygalacturonates become randomly cleaved to release n $>$ 3 oligogalacturonates (Abbott and Boraston, 2007b). A GH28 endo-polygalacturonase PehA from *Pectobacterium caratovorum* adopted an open-ended active site to attack the internal galacturonic acid residues within the polygalacturonate chain protruding out from the catalytic site at either end (Abbott and Boraston, 2007b; Pickersgill *et al*.,

1998). In contrast, the periplasmic exo-acting GH28 uses a pocket-like active site to hydrolyse the terminal non-reducing ends of oligogalacturonates to liberate mono-, di-, or trigalacturonate residues that are suitable for intracellular transport (Abbott and Boraston, 2007b). *D. dadantii* produces three GH28 exo-polygalacturonases encoded by *peh*V, *peh*W, and *peh*X organized in a single gene cluster (Nasser *et al*., 1999). The high sequence similarities between *peh*V, *peh*W, and *peh*X indicated that these genes arose from gene duplication events (Nasser *et al*., 1999). While the exo-acting PL9 PelX cleaves 6-residue polygalacturonate to produce a tetragalacturonate and an unsaturated digalacturonate, PehX uses the same substrate to produce a tetragalacturonate and a saturated digalacturonate as the sole end-product (Shevchik *et al*., 1999b). Although both PelX and PehX cleaved GalA oligomers with 4 – 6 residues, the presence of unsaturation at the non-reducing end of polygalacturonate adversely affected the activity of PehX, indicating PehX preferred to act on the saturated oligomeric products from endo-polygalacturonases (Shevchik *et al*., 1999b). Using a modified hexogalacturonate with an artificially reduced end, PehX produced a digalacturonate and reduced tetragalacturonate, indicating PehX attacks at the non-reducing end of the polygalacturonate (Shevchik *et al*., 1999b). In contrast, the modification of the reducing end inhibited PelX activity on hexogalacturonate, indicating PelX attacked from the reducing end of polygalacturonate (Shevchik *et al*., 1999b).

Rhamnogalacturonase and xylogalacturonan hydrolases are also pectin-associated hydrolases of GH28 family. However, these enzymes have been exclusively found and characterized from fungi, therefore will not be discussed in this review.

1.6.1.3 Carbohydrate esterases

Pectin methylesterases (CE8)

Pectin methylesterases facilitate the action of pectate lyases by catalysing the removal of methyl esters from the pectin backbone to release methanol, leading to the formation of polygalacturonate (Hugouvieux-Cotte-Pattat *et al*., 1996). An extensive de-methylation of pectin causes the cross-linking of HG regions via the chelation of Ca^{2+} ions as the removal of methyl esters creates negatively charged uronic acid residues (Øbro *et al*., 2009). The majority of pectin methylesterases have been

studied from plant enzymes that play significant roles during the regulation of plant development and fruit ripening processes. Pectin methylesterases can be described with three general patterns of polysaccharide modification: (1) a single-chain mechanism in which pectin methylesterases move in a processive fashion to remove all methyl groups from a single pectin chain before dissociating into the solvent; (2) a multi-chain or distributive mechanism in which the enzyme randomly removes a single methyl ester before dissociating; and (3) a multi-attack mechanism where pectin methylesterases carry out a number of reaction cycles before dissociation (Øbro *et al*., 2009; Fries *et al*., 2007; Duvetter *et al*., 2006). Plant pectin methylesterases usually operate in a processive manner by removing methyl esters in a block-wise arrangement while fungal enzymes generate randomly distributed de-esterified galacturonic acids using multi-chain mechanism (Duvetter *et al*., 2006). Bacterial pectin methylesterases use all three mechanisms, although environmental factors such as pH and cation concentrations may affect the enzyme activities (Øbro *et al*., 2009). Few CE8 pectin methylesterases from bacterial sources have been characterized, including PemA and PemB from *D. dadantii*. Despite the presence of PemA and PemB, *D. dadantii* cannot grow on high methyl pectin (98 %) (Shevchik *et al*., 1996). PemA is an extracellular pectin methylesterase encoded by *pem*A gene located at the downstream of the *pel*A-*pel*E-*pel*D locus (Hugouvieux-Cotte-Pattat *et al*., 1996). PemB is an outer membrane-associated lipoprotein that carries out the periplasmic de-esterification of methylated oligomeric products (Shevchik *et al*., 1996). While PemB was able to remove ester groups from 98 % DE pectin, the activity of PemB was 100-fold higher on methylated oligogalacturonate (Shevchik *et al*., 1996). PemA was also more active on methylated oligogalacturonate than natural pectin, but only by 5-fold (Shevchik *et al*., 1996).

Pectin acetylesterases (CE12)

In general, pectin acetylesterases cleave acetyl groups esterified to the O-2 and/or O-3 positions of galacturonic acid residues present in HG to liberate acetate (Shevchik *et al*., 1997). The biochemical properties of pectin acetylesterases from *D. dadantii* (PaeY and PaeX), and *Bacillus licheniformis* (*Bli*PAE) have been characterized (Shevchik *et al*., 1997; Shevchik *et al*., 2003; Remoroza *et al*., 2014). The *pae*Y gene is present in the *pel*A-*pel*E-*pel*D-*pae*Y-*pem*A locus which encodes various

types of pectinases acting in a synergetic manner for an efficient breakdown of pectin in plant tissues (Shevchik *et al.*, 1996; Shevchik *et al*., 1997). The *pae*X gene is present in the *pel*W-*tog*MNAB-*kdg*M-*pae*X operon which contains a cluster of genes involved in the catabolism and transport of pectic oligomers (Condemine and Ghazi, 2007). PaeY possesses a secretory signal peptide sequence, whereas PaeX becomes localized at the periplasm (Shevchik *et al.*, 1997; Shevchik *et al*., 2003). It was proposed that PaeY and PaeX have different substrate preferences for acetyl esters linked to pectin either at O-2 or O-3 configurations, although the exact mode of action is currently not well understood (Shevchik *et al*., 2003). Despite these differences, both PaeY and PaeX showed a hierarchical deacetylation of sugar beet pectin in which the highest deacetylation activities were observed with pectin pre-treated with PemA and pectate lyases, followed by de-methylated pectin (Shevchik *et al.*, 1997; Shevchik *et al*., 2003). PaeY and PaeX removed only 15 % and 3 % of acetyl groups from untreated natural pectin, respectively, indicating non-methylated oligogalacturonates were the best substrates for pectin acetylesterases (Shevchik *et al.*, 1997; Shevchik *et al*., 2003). *Bli*PAE from *B. licheniformis* showed a similar preference for low methoxyl pectin in that a 2-fold increase in the *Bli*PAE activity was observed when sugar beet pectin was pre-treated with pectin methylesterases (Romoroza *et al*., 2014). Pre-treating pectin with PemA and PaeY significantly improved the pectate lyase activities, suggesting the synergetic action of pectin methylesterases, pectin acetylesterases, and pectate lyases facilitated the efficient plant cell wall degradation by *D. dadantii* (Shevchik *et al*., 1997).

1.6.2 Rhamnogalacturonan-I (RG-I)-degrading enzymes

1.6.2.1 RG-I lyases

RG-I lyases catalyse the exo- and endo-cleavage of L-α-Rha*p*-1,4-α-D-Gal*p*A bond in RG-I region to release oligosaccharides with α-Δ4,5-unsaturated Gal*p*A at non-reducing ends. PL4 and PL11 families are currently known to possess RG-I lyases according to the CAZy database. PL11 family of RG-I lyases was first described in *Pseudomonas cellulosa* (McKie *et al*., 2001). Since then, five PL11 family of RG-I lyases were characterized from bacterial species including *B. licheniformis*, *Bacillus subtilis*, *Cellvibrio japonicas*, and *Clostridium cellulolyticum*. The acetyl groups and the side chains

made of galactan, arabinan, and arabinogalactan oligosaccharides present on the RG-I backbone affect the catalytic efficiency of RG-I lyases (Silva *et al*., 2016). Rgl11A from *P. cellulosa* and Rgl11Y from *C. cellulolyticum* both showed a preferential activity for cleaving RG-I containing galactose substitutions, indicating galactan side chains are probably required for their enzyme activity (McKie *et al*., 2001; Pages *et al*., 2003). The presence of acetyl group appeared to hinder the enzyme efficiency of Rgl11A in that the enzyme did not attack RG-I containing esterified galacturonic acid residues (McKie *et al*., 2001). These two enzymes were also active on pure galactans from potato pectin, although due to the natural origin of these substrates, the galactans may still have contained a trace amount of RG-I backbone (Silva *et al*., 2016). PL11 family of YesW and YesX from *B. subtilis* remained inactive on pure galactans, but showed activity towards potato RG-I substituted with side chains containing 12 % galactose residues (Ochiai *et al*., 2007). In *B. subtilis*, the endo-acting YesW was more active towards RG-I, leading to the production of oligosaccharides with differing degrees of polymerization which in turn were cleaved into the unsaturated RG disaccharides by the exo-acting function of YesX (Ochiai *et al*., 2007). The steric hindrance caused by the presence of acetyl groups and galactose moieties on RG-I backbone may be abated by the co-production of RG-I lyases and auxiliary enzymes such as β-D-galactosidase and rhamnogalacturonan acetylesterases (CE12). The YesW-encoding gene was found adjacent to a gene coding for β-D-galactosidase in *B. subtilis*, indicating β-D-galactosidase may catalyse the removal of galactan side chains prior to the activity of YesW on RG-I backbone (Ochiai *et al*., 2007; Silva *et al*., 2016). Along with Pel4A from *C. cellulovorans*, and PL1 and PL9 from *R. flavefaciens*, Rgl11Y is one of few lyases found incorporated in the cellulosome complexes, suggesting their involvement in the plant cell wall degradation (Pages *et al*., 2003; Venditto *et al*., 2016). Unlike PL4 which does not require divalent cations for activity, PL11 RG-I lyases either have an essential requirement for calcium, or show enhanced enzymatic activities in the presence of calcium (Silva *et al*., 2016).

1.6.2.2 RG-I hydrolases

RG-I hydrolases catalyse exo- and-endo hydrolysis of α-D-Gal*p*A-1,2-α-L-Rha*p* bond in RG-I backbone. Currently, seven endo-acting and two exo-acting RG-I hydrolases of GH28 family are

registered in the CAZy database, all of them from a fungal origin of the genus *Aspergillus* (Silva *et al*., 2016). No bacterial RG-I hydrolases of GH28 family have been reported so far, suggesting bacteria may prefer to use the β-elimination mechanism of RG-I lyases to cleave the RG-I backbone. The RG-I hydrolases from GH28 and GH105 families are structurally different in that GH28 possesses a β-helix structure consisting of 13 turns of parallel β-strands arranged in a helical pattern, while GH105 has an (α/α) double toroid structure with six α-hairpins forming a double helical barrel (Silva *et al*., 2016). Unsaturated rhamnogalacturonyl hydrolases (URGH) of GH105 has shown to cleave the unsaturated uronic acid at the non-reducing end of two- and three-chain unsaturated polysaccharides from plant and algal origins, as well as the non-reducing β-linked end of rhamnose residue to from fully saturated oligomers (Germane *et al*., 2015, Collen *et al*., 2014). So far, two URGHs, YteR and YesR, which belong to the GH105 family have been biochemically characterized from *B. subtilis* (Itoh *et al*., 2006). Similar to the PL22 family of oligogalacturonide lyases, YteR and YesR acted specifically on unsaturated RG disaccharides with α-Δ4,5-unsaturated-Gal*p*A at the non-reducing end generated by the activity of pectate lyases (Itoh *et al*., 2006). Recently, a structural analysis of URGH of GH105 family was carried out in *Clostridium acetobutylicum* (Germane *et al*., 2015). In *C. acetobutylicum*, GH105 was found to possess an active binding pocket for a trisaccharide containing an unsaturated galacturonate (Germane *et al*., 2015). In contrast to PL22 which produces two monomeric end-products galacturonate and 5-keto-4-deoxyuronate (DKI), the resultant products of GH105 are either a galacturonate monomer and a rhamnose residue, or a galacturonate or rhamnose monomer and saturated dimers of Rha*p*-Gal*p*A, Rha*p*-Rha*p* or Gal*p*A-Gal*p*A (Itoh *et al*., 2006). The dimeric products can be further cleaved into monomers by the activities of GH78 and GH106 families of α-rhamnosidases and rhamnogalacturonan-rhamnohydrolase, and GH28 family of polygalacturonases.

1.6.2.3 RG-I auxiliary enzymes

In this review, auxiliary enzymes refer to the enzymes which act on cleaving the glycosidic linkages present in the side chains of pectin. Auxiliary enzymes such as β-D-galactosidases are often co-produced with the enzymes catalysing the hydrolysis of pectin backbone to facilitate a coordinated

process of pectin depolymerisation (Ochiai *et al*., 2007; Silva *et al*., 2016). GH43 family is made of a heterogeneous population of enzymes including endo-acting α-L-arabinanases and exo-acting α-L-arabinofuranosidases. The members of GH43 family are abundantly present in bacteria that can degrade the major components of the plant cell wall, especially arabinose- and xylose-containing polysaccharides such as arabinoxylan, arabinan, and the side chains of RG-I (Cartmell *et al*., 2011). Endo-acting arabinanases of GH43 family introduced internal cleavages at α-1,5-L-Ara*f* linkages in the linear-α-1,5-L-arabinan side chains of RG-I to release arabino-oligosaccharides and L-arabinose monomers (Inacio and Sa-Nogueira 2008). The structure and function of GH43 arabinanases have been characterized in bacterial species such as *B. subtilis*, *B. licheniformis*, *B. thetaiotaomicron*, *Cellvibrio japonicus*, and *R. champanellensis*. The activity of arabinanases can be affected by the presence of additional branching within the arabinan side chains, as shown in the two endo-arabinanases Bt0360 and Bt0367 from *B. thetaiotaomicron* displaying differing preferences for branched and linear arabinan substrates, respectively (Cartmell *et al*., 2011). The activity of exo-arabinanases that release a terminal arabinose or arabino-oligosaccharides from α-1,5-L-arabinan has been reported in GH93 family, although so far, all exo-acting arabinases have been described in fungi (Carapito *et al*., 2009; Sakamoto and Thibault, 2001). The exo-acting α-L-arabinofuranosidases catalyse the hydrolysis of α-1,2, α-1,3, and α-1,5 terminal arabinofuranose residues from arabinan (Carapito *et al*., 2009; Miyanaga *et al*., 2004; Paes *et al*., 2008; Taylor *et al*., 2006; Shi *et al*., 2014). Currently, α-L-arabinofuranosidases are found in GH2, GH3, GH43, GH51, GH54, and GH62 families. Despite their abundance in bacterial communities, relatively few α-L-arabinofuranosidases have been biochemically characterized. Furthermore, enzymes of the same family often display varying substrate specificities, making it difficult to assign a defined function for each family (Taylor *et al*., 2006). For example, α-L-arabinofuranosidases of GH3 family usually occur as a bifunctional β-D-xylosidase/α-L-arabinosidase that can cleave both β-1,4,-xylose-xylose and β-1,3-xylose-arabinose linkages (Mai *et al*., 2000). GH43 family of α-L-arabinofuranosidases typically target the α-1,5-Ara*f* linkages present in arabino-oligosaccharides (Carapito *et al*., 2009; Nurizzo *et al*., 2002; Sakamoto *et al*., 2001). GH43 also includes bifunctional enzymes displaying β-D-xylosidase and α-L-arabinosidase activities, as well as arabinoxylan α-arabinofuranohydrolases that act on α-1,2 and α-

1,3-Ara*f* substituted with a single β-D-Xyl*p* residue (Wang *et al*., 2014; Viborg *et al*., 2014; Bourgois *et al*., 2007). GH51 and GH54 share a similar broad range of substrate specificities in that both enzymes act on α-1,2-, α-1,3-, and α-1,5-L-Ara*f* linkages, although GH51 shows preferential activities on short and unsubstituted arabino-oligosaccharides, whereas GH54 can act on both short substrates and arabinose-containing polysaccharides such as arabinan and arabinoxylan (Miyanaga *et al*., 2004; Ioannes *et al*., 2000; Koseki *et al*., 2003; Beldman *et al*., 1993; Taylor *et al*., 2006; Hoffmam *et al*., 2013). GH62 showed an exceptionally high activity on arabinoxylan. GH62 was able to cleave Ara*f* substituents that are α-1,2- and α-1,3-linked to xylopyranosyl (Xyl*p*) backbone in arabinoxylan, and also showed a low activity against α-1,5-L-Ara*f* linkages in debranched arabinans (Wang *et al*., 2014; Maehara *et al*., 2014; Wilkens *et al*., 2016; Sakamoto *et al*., 2011; Kaur *et al*., 2014).

Endo-galactanases catalyse the hydrolysis of β-1,4-Gal*p* linkages present in type I arabinogalactan and galactan side chains of RG-I. Currently, endo-acting β-galactanases with pectin-associated activities are found only in GH53 family. Endo-β-1,4-galactanses of family GH53 are predominantly found in bacteria, and the enzymes have been characterized in fibrolytic bacterial species such as *B. subtilis*, *B. licheniformis*, *B. thetaiotaomicron*, *D. dadantii*, *C. japonicus*, and *R. thermocellum*. GH53 showed a limited activity on galactan oligomers made of less than three Gal*p* residues, indicating exo-acting β-galactosidases were required to process D-Gal*p*-β-1,4-D-Gal*p* dimers into galactose monomers (Sakamoto *et al*., 2013; van Bueren *et al*., 2016). Exo-acting β-galactosidases catalyse the hydrolysis of terminal β-D-galactosyl residues from non-reducing ends of galacto-oligosaccharides. Exo-acting β-galactosidases have been divided into five different families: GH1, GH2, GH35, GH42, and GH59. The majority of microbial β-galactosidases has been classified into GH1 and GH2 families, whereas some enzymes active in extreme environments such as thermophilic hot springs, hypersaline brines, and psychrophilic soils are found in family GH42 (Gupta *et al*., 2012; Shipkowski *et al*., 2006; Karan *et al*., 2013). GH35 family has been found in all domains of life, mostly from plants and fungi (Gupta *et al*., 2012; Lee *et al*., 2017). The β-galactosidases of GH1 and GH2 are best represented by lactose-hydrolysing enzymes, most notably the *lacZ* gene of *Escherichia coli* (Talens-Perales *et al*., 2016; Lee *et al*., 2017). On the other hand, β-galactosidases belonging GH35 and GH42

families are often associated with galactose-containing polysaccharides such as galactan side chains of RG-I (Gamauf *et al*., 2007; Talens-Perales *et al*., 2016). GH35 family of β-galactosidases have demonstrated substrate specificities towards β-1,3-, β-1,6- or β-1,4-galactosidic linkages, of which β-1,4-D-Gal*p* bonds are predominantly found in pectic galactans (Kim *et al*., 2006; Gamauf *et al*., 2007). In family GH42, lactase functions are not always found in all members, as several GH42 enzymes are highly active on galacto-oligosaccharides and galactans, suggesting a function in the degradation of pectic polysaccharides (Shipkowski *et al*., 2006; Kosugi *et al*., 2002). In these studies, GH42 β-galactosidases from *B. subtilis* and *C. cellulovorans* were found to catalyse the hydrolysis of β-1,4-D-Gal*p* linkages in galacto-oligosaccharides from the galactan backbone of type I arabinogalactan.

Along with pectin acetylesterases, RG acetylesterases belong to CE12 family. RG acetylesterases facilitate the action of RG-I lyases and hydrolases by removing the acetyl groups at C-2 and C-3 positions of galacturonic acid residues in RG-I (Navarro-Fernandez *et al*., 2008). Bacterial RG acetylesterases have been studied little, and only two enzymes of this classification have been characterized from *Bacillus halodurans* (BhRgae) and *R. thermocellum* (Rgae12A) (Navarro-Fernandez *et al*., 2008). The substrate specificities of BhRgae and Rgae12A on pectic polysaccharides are currently unknown; BhRgae has shown activity towards acetylated xylan (Navarro-Fernandez *et al*., 2008).

1.6.3 Rhamnogalacturonan-II (RG-II)-degrading enzymes

The mechanism of RG-II degradation and the bacterial utilization of RG-II components are not clearly understood (Ndeh *et al*., 2017). Very recently, 12 *Bacteroides* spp. from the human gut were found to show positive growths on RG-II purified from apple juice (Ndeh *et al*., 2017). In this study, *B. thetaiotaomicron* was used as a model organism to uncover the metabolic mechanism of RG-II degradation for the first time. *B. thetaiotaomicron* completely degraded RG-II by producing enzymes with highly stringent target specificities to cleave 20 out of 21 glycosidic linkages present in the glycan, possibly compensating for the high energy costs incurred by maintaining the complicated catabolic machinery (Ndeh *et al*., 2017). Novel enzymatic functions for degrading RG-II components

were assigned to existing CAZymes families GH2 and GH127. The GH2 β-D-galacturonidase and α-L-arabinopyrosidase, and the GH127 3-*C*-carboxy-5-deoxy-L-xylose hydrolase were novel functions assigned to these CAZymes families (Ndeh *et al*., 2017). Seven novel GH families (GH137 – GH143) containing β-L-arabinofuranosidase, α-galacturonidase, α-2-methyl-fucosidase, apiosidase, α-L-fucosidase, and DHA hydrolase, and a novel family of pectin methylesterase were discovered and newly added to the RG-II degradome (Ndeh *et al*., 2017). These new enzymatic specificities suggested the current model of RG-II structure required revision (Ndeh *et al*., 2017). In the RG-II side chains A and B, it was suggested that Rha*p*-Api*f* linkage should be in α-configuration as opposed to the previously proposed β-configuration (Ndeh *et al*., 2017). Also, a novel side chain F consisting of a backbone of Gal*p*A substituted with α-Ara*f* at O3 was found to be linked to chain A (Ndeh *et al*., 2017). The location of methyl-esterified GalpA residues were pinpointed to occur closely to chains A and F (Ndeh *et al*., 2017).

1.6.4 Pectin-associated carbohydrate binding modules (CBMs)

Currently, there are 9 CBM families characterized with the binding affinity for pectin or monomeric components of pectin as listed in Appendix 2. Despite the importance of pectin in maintaining the cell wall structure in the majority of terrestrial plants, remarkably few pectin-targeting CBMs have been described to date, possibly because a large number of CBMs were studied using the cellulose- and glycan-degrading bacteria as model organisms (Cid *et al*., 2010). The carbohydrate binding domains of CAZymes often reflect the target substrate specificity of the enzymes, providing clues for the enzyme functions (Fujimoto *et al*., 2002). For example, arabinose-binding CBM42 family was in association with GH43, GH51, and GH54 families of α-L-arabinofuranosidases (Fujimoto *et al*., 2008; Miyanaga *et al*., 2004). Similarly, the CBM13 domain appended to the endo-β-1,4-D xylanase (GH10) of *Streptomyces olivaceoviridis* showed a preferential binding affinity for xylose or xylo-oligosaccharides from plant hemicellulose (Fujimoto *et al*., 2002). CBM13 of *S. olivaceoviridis* was also shown to be capable of binding single units of galactose and lactose in its xylan-binding site, as it shared the conserved polar residues Asp, Gln, Asn, and Tyr with the galactose binding sites of *Ricinus communis* (castor-oil plant) (Fujimoto *et al*., 2002). Similar capacity to bind galactose and the

galactose moiety of lactose was demonstrated using the CBM32 appended to the sialidase of *Micromonospora viridifaciens* (Newstead *et al*., 2005). CBM32s from *Clostridium perfringens* and *Cladobotryum dendroides* have also been assigned with similar galactose-binding functions (Abbott *et al*., 2007). On the other hand, the periplasmic CBM32 of *Yersinia enterocolitica* (YeCBM32) shared only ~20 % amino acid sequence similarities with other CBM32s, and possessed a structurally unique binding site architecture in which only a single active site residue (Trp) was conserved (Abbott *et al*., 2007). The unique binding site topology of YeCBM32 led to the alkalinisation of the surface to accommodate the internal binding of acidic oligosaccharides, resulting in the non-typical binding specificity to polymerized galacturonic acid residues, potato RG-I, and potato pectin galactan (Abbott *et al*., 2007c). YeCBM32 distinguishes itself from other characterized CBM32s in that it stands as an independent protein, and not as an accessory module of a larger catalytic protein (Abbott *et al*., 2007c). Apart from the cellulose-binding CBM of CttA protein in the cellulosome of *R. flavefaciens,* and the Sus proteins of *B. thetaiotaomicron,* CBMs appended within non-catalytic polypeptides have been rarely reported (Flint *et al*., 2008; Abbott *et al.,* 2007c; Cameron *et al*., 2012). Ligand binding prior to intracellular transport and the prolonged retention of the substrate close to the bacterial cell surface have been proposed as potential functions of independent CBM domains (Abbott *et al.,* 2007c). Substrate binding affinity for unsaturated Δ4,5-Gal*p*A was reported from CBM35 domains appended to PL10 (Pel-CBM35), CE12 (Rhe-CBM35), and GH10 xylanase (Xyl-CBM35) from an environmental isolate, *R. thermocellum*, and *C. japonicus*, respectively (Montanier *et al*., 2009). All CBM35s were tested positive for binding the resultant unsaturated pectin product of β-elimination, but showed no affinity for pectin pre-digested with glycoside hydrolases, indicating CBM35 recognized Δ4,5-Gal*p*A (Montanier *et al*., 2009). While Xyl-CBM35 may target the GlcA side chains present in some xylans, its binding to degraded pectic structure may play a role in recruiting xylanases at the sites of active plant degradation (Montanier *et al*., 2009). CBM62 appended to the GH5 endo-β-1,4-xylanase shows a similar CBM-dependent targeting of catalytic modules into close proximity with the degrading complex polysaccharides (Montanier *et al*., 2011). The CBM62 from the cellulosomal GH5 endo-β-1,4-xylanase of *C. thermocellum* (*Ct*CBM62) is currently the only constituting member of family CBM62 (Montanier *et al*., 2011). Although *Ct*CBM62 showed binding affinity with

complex polysaccharides such as galactomannan, arabinogalactan, and xyloglucan, the binding is mediated by single ligand-binding sites which recognize single units of D-galactose and L-arabinopyranose decorating the target glycans (Montanier *et al*., 2011). *Ct*CBM62 showed a significantly augmented binding affinity with complex polysaccharides in the presence of calcium ions which play important roles during the oligomerization of CtCBM62, resulting in the formation of multivalent ligand recognition sites (Montanier *et al*.,2011). CBM60 and CBM61 both displayed binding capacity for β-1,4-galactans (Montanier *et al*., 2011; Cid *et al*., 2010). CBM60 from GH11 xylanase binds tightly to 4 – 6 sugar residues of both β-1,4-galactans and xylan, although CBM60 was almost always associated with a catalytic module of xylanases, and xylan was usually selected as a preferred substrate for binding in the primary cell walls of tobacco stems (Montanier *et al*., 2010). CBM61 was first described in the functionally obscure ~160 amino acid region of GH53 endo-β-1,4-galactanase from *Thermotoga maritima* (Cid *et al*., 2010). CBM61 showed the greatest binding affinity to pectin and β-galactan motifs present in complex polysaccharides (Cid *et al*., 2010). Using glycan microarray and immunofluorescence microscopy, similar substrate binding profile was observed between CBM61 and anti-β-1,4-galactan monoclonal antibody LM5, although CBM61 showed a greater efficiency at accessing a wide range of cell wall sections of *Arabidopsis thaliana* stem, indicating CBM61 was able to tolerate glycan decorations present on the target polysaccharide (Cid *et al*., 2010). In *Streptomyces avermitilis*, CBM67 appended to GH78 α-L-rhamnosidase showed binding to L-rhamnose, and a reduced affinity for L-mannose (Fujimoto *et al*., 2013). CBM67 activity is calcium-dependent, and the EDTA-induced chelation of calcium ions led to a significantly reduced enzyme activity of GH78 against gum arabic, a rhamnose-containing complex polysaccharide (Fujimoto *et al*., 2013). CBM67 domains are also found in PL1 of *Sorangium cellulosum* and α-L-rhamnosidase of *Paenibacillus* sp., and the calcium binding motif within the ligand recognition sites was conserved in both CBM67s (Fujimoto *et al*., 2013). CBM77 was recently discovered in a cellulosomal enzyme which possessed bifunctional pectate lyase families 1 and 9 from *R. flavefaciens* (Venditto *et al*., 2016). $CBM77_{PL1/9}$ exclusively targeted low methylated homogalacturonan with a degree of polymerization $>$ 7 or 8 Gal*p*A residues (Venditto *et al*., 2016). Despite the mandatory requirements for divalent cations in pectate lyases, CBM77 activity was not affected by the use of

EDTA (Venditto *et al*., 2016). Currently, CtCBM62 and CBM77$_{PL1/9}$ are the only cellulosomal CBM domains from which the target specificities for pectic polysaccharides have been demonstrated.

1.7 Pectin metabolism

While the catabolic action of polygalacturonases produces saturated digalacturonate which eventually becomes degraded to form D-galacturonate monomers, pectate lyases form unsaturated digalacturonate which becomes cleaved by oligogalacturonate-lyase to form two different types of monomeric end-products; D-galacturonate and 5-keto-4-deoxyuronate (DKI) (Figure 1.8). In isomerase pathway, D-galacturonate becomes processed by uronate isomerase (UxaC; EC5.3.1.12), NADH-utilizing D-tagaturonate reductase (UxaB; EC1.1.1.58), altronate dehydratase (UxaA; EC4.2.1.7), 2-keto-3-deoxy-D-gluconate kinase (KdgK; EC2.7.1.45), and 2-keto-3-deoxy-6-phosphogluconate aldolase (KdgA; EC4.1.2.14) to form intermediates D-tagaturonate, D-

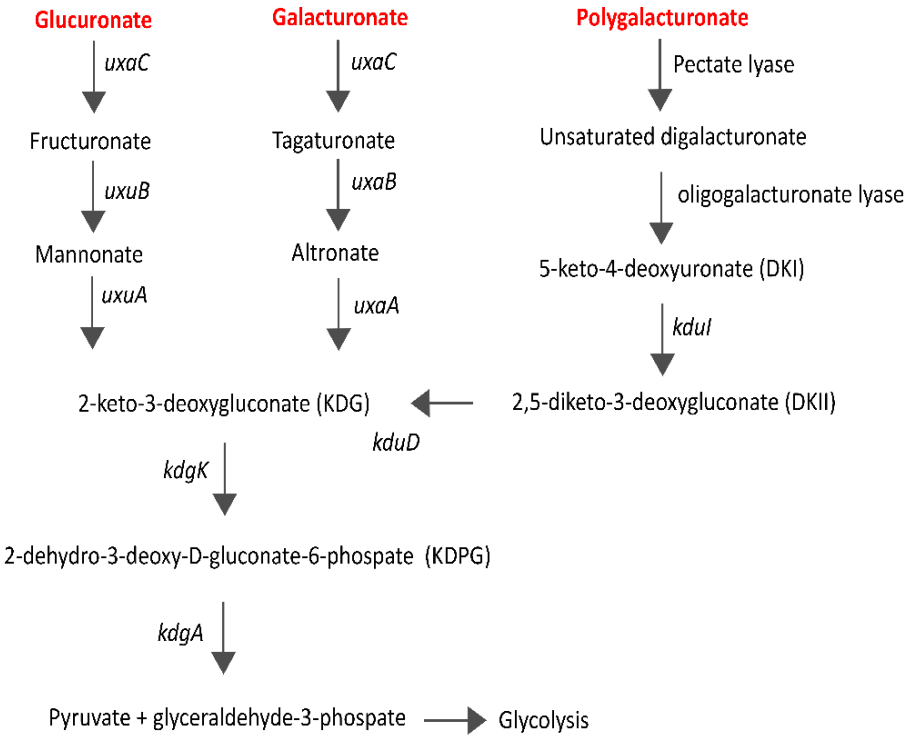


Figure 1.8 Pectin degradation pathway characterized in *Dickeya dadantii*. This diagram was adopted from Condemine and Robert-Baudouy (1987). No copyright permission is required for the use of this figure.

altronate, 2-keto-3-deoxygluconate (KDG), 2-dehydro-3-deoxy-D-gluconate-6-phosphate (KDPG), eventually converting KDPG to pyruvate and glyceraldehyde-3-phosphate (Richard and Hilditch 2009). Glucuronate metabolism follows a pathway similar to the isomerase pathway, but instead uses fructuronate reductase (UxuB; EC1.1.1.57) and mannonate dehydratase (UxuA; EC4.2.1.8) to convert D-glucuronate to D-fructuronate and D-mannonate which eventually converge into KDG of the isomerase pathway. DKI, which is reported to constitute 20 % of the end-product from the lyase-mediated pectin catabolism (Richard and Hilditch 2009), is converted to 2,5-diketo-3-deoxygluconate (DKII) and then to KDG by the consecutive actions of 5-dehydro-4-deoxy-D-glucuronate isomerase (KduI; EC 5.3.1.17) and 2-deoxy-D-gluconate 3-dehydrogenase (KduD; EC1.1.1.125).

1.8 Pectin degradation in the human large intestine

Currently, the majority of pectinolytic bacteria isolated from human gut are of *Bacteroides* spp., with *Faecalibacterium prausnitzii*, *Eubacterium eligens*, and an uncharacterized *Clostridium* sp. being the few pectinolytic organisms from the phylum *Firmicutes* (Appendix 3). As a result, our current understanding of the metabolic process of pectin degradation has been largely derived from studying Gram negative models. For *Bacteroides* spp. which lack an extracellular machinery for efficient substrate binding to maintain a proximal cell-to-substrate distance, the extracellular secretion of enzymes may not be an ideal strategy as the enzymes will diffuse before encountering the substrates. As a result, *Bacteroides* spp. do not usually release a large number of CAZymes into the extracellular environment, and therefore do not share their degradative ability with neighbouring bacterial species less adapted to utilizing complex polysaccharide substrates (Comstock, 2009). In *B. thetaiotaomicron*, up to 80 % of genes encoding CAZymes contained predicted signal peptide sequences, indicating these enzymes were targeted to either the periplasm or the outer membrane (El Kaoutari *et al*. 2013). In one study, no polygalacturonase or pectate lyase activities were detected in the extracellular supernatant of *B. thetaiotaomicron*, as these enzyme activities were mostly found to be cell-associated, indicating B. *thetaiotaomicron* mostly produced outer membrane-associated and periplasmic CAZymes to carry out pectin degradation using a sequestration mechanism (McCarthy 1985; Flint *et al*., 2008). *Bacteroides* spp. are known to produce high numbers of glycoside hydrolases

and polysaccharide lyases, although the number of glycoside hydrolases often greatly outnumbers that of polysaccharide lyases (El Kaoutari *et al*., 2013). In fact, polysaccharide lyases made up less than 2 % of the glycobiome of a simulated human gut microbiome, making it easy to assume that the pectin decomposition would mostly rely on the enzymatic activities of glycoside hydrolases (El Kaoutari *et al*., 2013). However, the by-product of the pectin degradation by human faecal microflora and the pure culture of *B. thetaiotaomicron* suggested that the resultant pectin oligomers followed the β-elimination pattern of pectate lyases, suggesting pectate lyases were likely to be the key enzymes during the pectin degradation in the large intestine (Dongowski *et al*., 2000). Furthermore, the activity of pectin methylesterases appeared to make a significant contribution to the pectin degradation process in that the degree of methylation on the pectin backbone significantly affected the digestibility of pectin, as high methoxyl pectin became depolymerized at a much slower rate than low methoxyl pectin (Dongowski *et al*., 2000). Interestingly, although the simulated human microbiome contained 5 times more *Firmicutes* genomes than *Bacteroidetes* genomes, almost 3 times more polysaccharide lyase genes were found to be associated with *Bacteroidetes* (El Kaoutari *et al*., 2013). The highest number of polysaccharide lyases in the human gut microbiome are produced by the phylum *Bacteroidetes* may explain why the majority of pectinolytic bacteria isolated from the humans so far are *Bacteroides* spp. Therefore, pectin-specific PULs may be highly conserved across many *Bacteroides* spp. as demonstrated for the Sus-like pectin utilization system of *B. xylanisolvens* which shared the similar pectin-PUL architectures with other *Bacteroides* spp. (Despres *et al*., 2016). Among *Firmicutes*, *E. eligens* and *F. prausnitzii* are known to break down pectin in the human gut (Salyers *et al*., 1977a; Lopez-Siles *et al.*, 2012). Both organisms possess moderate sized glycobiomes; *E. eligens* ATCC 27750 produces 5 GHs, 7 PLs, and 4 CEs specifically targeting pectin backbone structure compared to *F. prausnitzii* L2-6 that produces 2 GHs, 1 PL, and 2 CEs devoted to pectin degradation (CAZy database). Salyers (1977a) reported that *E. eligens* was the only organism tested positive for pectinolytic ability out of 153 strains of genera *Bifidobacterium, Coprococcus, Eubacterium, Faecalibacterium, Lactobacillus, Peptostreptococcus,* and *Ruminococcus*. A recent study showed a dramatic increase in the abundance of *E. eligens* in a simulated gut microbial community using *in vitro* fermenter vessels fed with apple pectin (Chung *et al.,* 2016). Using inulin as a negative control,

E. eligens remained undetectable throughout the fermentation, indicating this organism thrives in the presence of pectin, and may play a key role in degrading pectin in the human gut. Although some strains of *F. prausnitzii* were reported to ferment pectin, their pectinolytic activity was limited to apple pectin as no growth occurred using polygalacturonate or pectin from citrus peels (Lopez-Siles *et al.,* 2012). Another study reported that five strains of *F. prausnitzii* were tested negative for the utilization of pectin and polygalacturonate (Salyers *et al.,* 1977). Interestingly, most strains of *F. prausnitzii* were able to grow using galacturonate monomer as a sole carbon source, as later confirmed by the presence of uronic acid utilization pathway in the genome of the organism (Lopez-Siles *et al.,* 2012; Heinken *et al.,* 2014). Due to its narrow range of substrate specificity, the extent to which *F. prausnitzii* contributes towards the pectin degradation in the human gut remains unclear. Chung (2016) reported a pH-dependent increase of *F. prausnitzii* in apple pectin-fed fermenter vessels, but suggested that the increase in the abundance was likely due to an opportunistic growth of *F. prausnitzii* in the event of the low pH inhibition of *Bacteroides* spp. At neutral pH, the abundance of *F. prausnitzii* remained low and changed little compared to the inulin control. A further assessment may be required to determine if *F. prausnitzii* should be considered as pectinolytic or a polygalacturonate-utilizing bacterium.

Members of *Enterobacteriaceae*, such as *Y. enterocolitica* and *Escherichia coli* that colonize the human gut, was found to possess pectinases and possibly the abridged pectin degradation pathways (Abbott and Boraston, 2008). *Yersinia enterocolitica* were found to produce both endo- and exo-types of pectate lyases (PelY and PelW) that are periplasmic and cytoplasmic, respectively (Hugouvieux-Cotte-Pattat 2014). It also possesses YeCE8, an intracellular pectin methylesterase (Boraston 2012)

1.9 Rationale of the research

Humans rely on the complex gut microbiota to maintain their omnivorous digestive functions. The diverse fibrolytic enzymes produced by gut bacteria provide metabolic versatility which enables an efficient degradation of the ingested plant materials. Our current understanding of the microbial degradation of the plant cell wall was mostly derived from studying ruminant animals. However, as the health benefits of plant-rich diets are being widely recognized, there are growing interests to

understand the underlying mechanisms of plant fibre degradation in the human colon. Pectin is a major non-cellulosic component which comprises approximately one third of the cell wall dry weight of food plant materials (Caffall and Mohnen, 2009). The intercellular space between adjacent plant cells is covered with pectin layers called the middle lamella. The microbial degradation of ingested plant materials typically begins with the solubilisation of the middle lamella by pectin-degrading bacteria which subsequently expose adhesion sites along the plant cell wall surfaces for further microbial colonization (Cheng *et al.*, 1979). Despite the abundant pectin consumption in our modern-day diet, few pectin-degrading bacteria have been found in the human gut, and none with the genomic and nutritional specialization for utilizing pectin. Even though successful cultivation of microorganisms remains crucial for producing good quality genomes and investigating microbial functions at molecular levels, it is often challenged by the difficulty of replicating the natural physiology of the colonic environment and the fastidious anaerobic conditions in laboratories. As a result, significance of the microbial pectin degradation with respect to improving the colonic digestibility of insoluble fibres has not been properly addressed. Also, knowledge of the ecological niche for primary pectin degraders has remained lacking for a long time. This study sought to isolate pectin-degrading keystone bacterial species from the human gut, and to provide a genomic insight into the 'missing link' that currently exists in our understanding of the hierarchical degradation of the plant cell wall.

1.10 Aims

Due to the scarcity of cultured pectinolytic species of the phylum *Firmicutes*, Gram positive model of particulate-associated pectin degradation has not yet been discovered from the human large intestine. Fibrolytic members of the phylum *Firmicutes* are highly sought after for cultivation and genome sequencing as their elaborate enzymatic machineries attract a considerable amount of research attention to elucidate the metabolic processes of plant digestion in the human gut. Given the conspicuous presence of pectin in the cell walls of dietary plants, a Gram positive organism which specializes in the pectin degradation would be expected to play important roles in breaching the insoluble barrier of plant cells, thus facilitating a rapid and efficient fibre digestion. Therefore, the

main aim of this research was to isolate and characterise a novel human gut bacterium showing strong pectinolytic activities. This was achieved by investigating the following questions:

1. Can a pectin-degrading specialist gut bacterium be isolated from human faecal material and, if so,
2. What are the general biochemical and phenotypic characteristics of such an organism?
3. What are the features of its glycobiome and the pectin degradation system that can be deduced from the genome?
4. What does proteomics suggest about the pectin-induced differential expression of candidate proteins potentially involved in the pectin degradation system?
5. How abundant is this organism in a sample of the New Zealand population?

Chapter 2 Materials and Methods

2.1 General methods

2.1.1 Media and additives

Unless specified otherwise, all anaerobic media and additives in this study were prepared using strict anaerobic techniques described in McSweeney *et al.,* 2005. Final media preparations were maintained under anaerobic condition in either 16 × 125 mm Hungate tubes (Bellco) sealed with open top screw caps and rubber stoppers (Bellco) or 50 ml glass serum vials (Supelco) plugged with butyl rubber stoppers and aluminium seals (Supelco). Liquid medium was sterilized by autoclaving at 121℃ for 20 min. Any additions of components into sterilized media were carried out by inserting N_2-flushed syringes and needles through the stoppers or inside a UV-sterilized anaerobic chamber (COY Laboratory Products Inc. Michigan).

2.1.2 Basal nutrient (BN) medium

Basal nutrient medium was used as an enrichment medium for pectin-utilizing bacteria from human faeces. Per litre of basal nutrient medium contained 2 g Difco™ peptone water; 2 g Bacto™ yeast extract; 0.5 g bile salts (Oxoid); 10 ml Mineral A stock solution (0.4 g K_2HPO_4 dissolved in 100 ml distilled water); 10 ml Mineral B stock solution (0.4 g KH_2PO_4; 0.1 g $MgSO_4.7H_2O$; 0.1 g $CaCl_2.6H_2O$; and 1.0 g NaCl dissolved in 100 ml distilled water); 2 ml Tween 80 (Fischer Scientific); 0.05 g hemin (Sigma); 10 µl vitamin K (Sigma); 0.5 g L-cysteine-HCl; 0.02 g resazurin; and 4.2 g $NaHCO_3$. All components except hemin, vitamin K, L-cysteine-HCl, and $NaHCO_3$ were dissolved in distilled water and boiled, and then were transferred on ice while hot. The medium was bubbled with N_2 while being cooled on ice. Hemin, vitamin K, L-cysteine-HCl, and $NaHCO_3$ were added into the cooled medium. Required volumes of the medium were dispensed into Hungate tubes or serum vials under N_2, resulting in the headspace of the tubes to be filled with 100 % N_2. Vessels sealed with caps and stoppers were sterilized by autoclaving. Pectin substrates and 5 % (v/v) clarified rumen fluid were added immediately prior to inoculation.

2.1.3 Basal nutrient (BN) agar medium prepared in roll-tubes

9 ml of basal nutrient medium was dispensed into Hungate tubes containing 150 mg of agar. Medium was sterilized by autoclaving and cooled to 50 ℃ in a water bath. Additives were injected through rubber stoppers to make up the final volume of the medium to 10 ml. Hot medium was spread over the inner surface of stoppered tubes under cold tap water as described in Hungate (1969).

2.1.4 Mineral medium

Routine cultivation and characterization of strain 14^T was carried out using mineral medium. Per litre of the medium contained 1.4 g KH_2PO_4; 0.6 g $(NH_4)_2SO_4$; 1.5 g KCl; 1 g yeast extract; 0.02 g resazurin; 4.2 g $NaHCO_3$; and 0.5 g L-cysteine-HCl. $NaHCO_3$ and L-cysteine-HCl were added after boiling and cooling the medium on ice under 100 % CO_2. Required volumes of the medium were dispensed into Hungate tubes or serum vials under CO_2. Sealed vessels were sterilized by autoclaving. Substrates and 5 % (v/v) clarified rumen fluid were added to the medium prior to inoculation.

2.1.5 Reinforced clostridial (RC) medium

Genomic DNA was extracted from strain 14^T grown on RC medium (CM0149) purchased from Oxoid. RCM contains 13 g/L yeast extract; 10 g/L peptone; 5 g/L glucose; 1 g/L soluble starch; 5 g/L NaCl; 3 g/L sodium acetate; 0.5 g/L; cysteine hydrochloride; and 0.5 g/L agar. Following the manufacturer's instruction, 38 g of RCM was dissolved in 950 ml of distilled water in 1 litre Duran® glass bottles. The pH of the medium was adjusted to 8.0 using 2M NaOH. Medium was boiled and cooled on ice while bubbling with N_2. Medium was autoclaved and kept inside a UV-sterilized anaerobic chamber overnight to equilibrate with the anaerobic ambient. Additives (2 % (v/v) clarified rumen fluid and 50 ml of 10 % (w/v) D-fructose solution, filter-sterilized) were aseptically added to the medium prior to inoculation.

2.1.6 Luria-Bertani (LB) broth and agar medium

20 g of LB broth powder (Sigma) was dissolved in 950 ml of distilled water. Medium was sterilized by autoclaving. Autoclaved medium was allowed to cool to 50℃ in a water bath before adding antibiotics if required. LB agar medium was prepared by adding 15 g of agar per litre of the medium before autoclaving. Autoclaved agar medium containing antibiotics was poured into Petri dishes.

2.1.7 Clarified rumen fluid

Rumen fluid collected from hay-fed fistulated cows was provided by AgResearch Ltd (Grasslands Research Centre, Palmerston North). Rumen fluids were bubbled with N_2 for 20 min and dispensed into 50 ml serum vials sealed with rubber stoppers and aluminium caps. Rumen fluid was autoclaved. Autoclaved rumen fluid was poured into a beaker and mixed thoroughly with $MgCl_2.6H_2O$ (1.63 g per 100 ml) and $CaCl_2.2H_2O$ (1.18 g per 100 ml) by stirring for 30 min. Precipitates were removed by centrifuging at 6,000 g for 20 min at 4℃. Supernatant was collected in a clean beaker and bubbled with N_2 for 20 min. While under N_2, supernatant was filter-sterilized through 0.22 µm filter (Merck Millipore) into sterile and N_2-washed serum vials sealed with rubber stoppers and aluminium caps. 2 % (v/v) anaerobically prepared vitamin solution was added by using N_2-washed syringes and needles. The resulting mixture was termed 'Clarified rumen fluid' in this study.

2.1.8 Vitamin solution

RPMI-1640 Vitamin Solution (R 7256) was purchased from Sigma-Aldrich. RPMI-1640 Vitamin Solution contains 0.02 g/L D-biotin; 0.3 g/L Choline Chloride; 0.1 g/L Folic Acid; 3.5 g/L myo-Inositol; 0.1 g/L Niacinamide; 0.1 g/L p-Amino Benzoic Acid; 0.025 g/L D-Pantothenic Acid·$^1/_2$Ca; 0.1 g/L Pyridoxine·HCl; 0.02 g/L riboflavin; 0.1 g/L Thiamine·HCl; 0.0005 g/L Vitamin B12; 0.2 g/L KCl; 0.2 g/L KH_2PO_4 (anhyd); 8 g/L NaCl; and 1.15 g/L Na_2HPO_4 (anhyd). The solution was poured into a beaker and bubbled with N_2 for 10 min. While bubbling with N_2, the solution was filter-sterilized (0.22 µm) into sterile and N_2-washed serum vials. Vitamin Solution was stored frozen at -20℃ until use.

2.1.9 Anaerobic water

1 litre of distilled water was boiled and cooled on ice under a continuous flow of N_2. 0.5 g of L-cysteine HCl was added. Required volumes of water were dispensed into Hungate tubes or serum vials and sealed using caps and rubber stoppers under N_2. Anaerobically prepared water was sterilized by autoclaving.

2.1.10 Anaerobic phosphate buffered saline (PBS)

1 litre of 10 mM PBS (MP Biomedicals) was boiled and cooled on ice under a continuous flow of N_2. 0.5 g of L-cysteine HCl was added to the cooled solution. Required volumes of PBS were dispensed into anaerobic vessels and sealed under N_2. PBS was sterilized by autoclaving.

2.1.11 Anaerobic glycerol solution for a long-term preservation of bacteria

50 % (v/v) glycerol was mixed with distilled water and boiled and cooled on ice under N_2. 0.5 g of L-cysteine HCl was added per litre of glycerol solution. Solution was dispensed into anaerobic glass vessels under N_2 and sealed using caps and rubber stoppers. Solution was sterilized by autoclaving. 50 % glycerol solution was mixed vigorously with exponentially growing bacterial culture at 1:1 ratio. The mixture was stored at -80℃.

2.2 Materials and methods for isolation and characterization of strain 14^T

2.2.1 Preparation of pectin from kiwifruit

2.2.1.1 *in vitro* gastric/ileal digestion of kiwifruit

Pepsin solution

1 g of pepsin from porcine gastric mucosa (250 units/mg; Sigma) was dissolved in 10 ml of 50 mM HCl.

Pancreatin solution

5 % (w/v) of pancreatin from porcine pancreas (protease 250 FIP-U/g; lipase 6,000 FIP-U/g; amylase 7,500 FIP-U/g; Merck) was dissolved in 20 ml of saline solution. Saline solution was prepared by dissolving 10 g of NaCl and 2 g of $CaCl_2$ in 1 L of distilled water.

***in vitro* digestion of kiwifruit**

Seasonal green kiwifruit was harvested at maturity from Plant & Food Research orchard (Te Puke, New Zealand). Each fruit was weighed and peeled before blending into pulps using a low speed blender (15 rpm/sec). Pulps were centrifuged at a low speed (1,000 g, 5 min) to remove seeds. Kiwifruit pulps were subjected to *in vitro* gastric and ileal digestions as described by Monro *et al*., 2010 and modified by Carnachan *et al*., 2012. Kiwifruit pulps were poured into a beaker and mixed

on a heated magnetic stirrer until the temperature reached 37℃. Gastric digestion was initiated by adjusting pH of pulps to 2.5 by slowly adding 2M HCl while stirring constantly. For every 200 g of pulps, 10 ml of pepsin solution was added, and the gastric digestion was carried out for 30 min at 37℃. The transition to ileal digestion was carried out by increasing the pH of pulps to 6.5 by adding 2M NaOH. For every 200 g of pulps, 20 ml of pancreatin solution and 0.5 ml of amyloglucosidase (soluble starch 3,260 U/ml; Megazyme) were added. Ileal digestion was carried out for 120 min at 37℃ with a constant stirring.

2.2.1.2 Chemical extraction of pectin from digested kiwifruit

CDTA solution

4.55 g of *trans*-1,2-diaminocyclohexane-N,N,N′,N′-tetraacetic acid monohydrate (CDTA) was dissolved in 200 ml water. 2M KOH was added until pH reached 6.0 and CDTA completely dissolved. 1.23 g of potassium acetate was added and pH was further adjusted to 6.5 with 2 M KOH. The final volume of the solution was made up to 250 ml with distilled water.

Na_2CO_3 solution

Na_2CO_3 solution was freshly prepared immediately prior to use. 1.32 g of Na_2CO_3 was dissolved in 200 ml of water. 0.19 g of $NaBH_4$ was added, and distilled water was added to make up the final volume to 250 ml.

Ammonium acetate buffer

7.708 g of ammonium acetate was dissolved in 950 ml of distilled water. Glacial acetic acid was added to adjust pH to 6.5. The final volume was made up to 1 L with distilled water. Ammonium acetate buffer was freshly prepared before use.

Fractionation of water-soluble pectin using ethanol

Pectic polysaccharides were sequentially extracted from the cell walls of digested kiwifruit as described by Melton and Smith (2001). Digested kiwifruit materials were collected by centrifugation at 5,000 g for 15 min at 4℃. Supernatants were pooled and thoroughly mixed with four volumes of 80% ethanol overnight. Precipitated materials were spun down at 5,000 g for 15 min at 4℃, and

supernatants were discarded. Precipitates were washed with a small volume of 100 % acetone twice to remove residual ethanol, and then were mixed with a small amount of distilled water until a fine paste was formed. Paste was kept frozen at -80℃ overnight, and freeze-dried. This initial soluble pectin extract will be called "Ethanol fraction" in this study.

Fractionation of cell wall-associated pectin using CDTA and Na_2CO_3

Soluble cell wall polysaccharides in kiwifruit digesta were further fractionated by successive treatments in CDTA solution and Na_2CO_3 solution, followed by dialysis. For every 50 g of kiwifruit digesta, 20 ml of CDTA solution was added. After vigorous mixing, the mixture was centrifuged at 1,000 g for 5 min at room temperature. The supernatant was decanted and retained at 4℃. The pellet was resuspended in 100 ml of CDTA solution, and stirred at room temperature for 6 h. The mixture was centrifuged at 1,000 g for 5 min at room temperature. Supernatant was pooled and combined with previous supernatant. This process was repeated and all three supernatants were combined and retained at 4℃. The resulting mixture of supernatants was termed 'CDTA fraction' in this study. Pellets were vigorously mixed in 20 ml of Na_2CO_3 solution and centrifuged at 1,000 g for 5 min at room temperature. The supernatant was decanted and retained at 4℃. The pellets were resuspended in 100 ml of Na_2CO_3 solution and stirred for 16 h at 4℃, and then for 2 h at room temperature. Centrifugation was carried out at 1,000 g for 10 min at room temperature. The resulting supernatant was combined with previous supernatant and kept at 4℃. Finally, the pellet was resuspended in 100 ml of Na_2CO_3 solution, and stirred for 2 h at room temperature. The mixture was centrifuged at 1,000 g for 10 min at room temperature. This process was repeated. The resulting supernatants were pooled and combined with previous supernatants. The combined supernatants were termed 'Na_2CO_3 fraction' in this study.

Dialysis

12,000 to 14,000 MWCO dialysis tubing (Visking size 6, 37 × 32 mm diameter, Medicell International) were cut and rinsed in distilled water. Dialysis was carried out at 4℃. 250 ml of CDTA fraction was poured into knotted dialysis tubing and dialysed against 10 L of ammonium acetate buffer for 1 day, with 3 changes of buffer. Further dialysis was carried out in 10 L of distilled water

over 3 days with 3 changes of water per day. Using 1M acetic acid, pH of Na_2CO_3 fraction was adjusted to 6.5. 250 ml of Na_2CO_3 fraction in dialysis tubing was dialysed against 10 L of distilled water for 3 days with 3 changes of water per day.

Freeze-drying

Dialysed fractions were kept frozen at -80℃ overnight. Samples were freeze-dried at -50℃ at 513 mTorr (Flexi-Dry™ freeze-drier).

2.2.1.3 Sterilization by gamma irradiation

Freeze-dried samples of Ethanol, CDTA, and Na_2CO_3 fractions were sterilized by gamma-irradiation at 17 kGy (MSD Animal Health, Wellington).

2.2.1.4 Separation of insoluble and soluble polysaccharides from digested kiwifruit

In order to compare the constituent sugar compositions of soluble and insoluble kiwifruit materials before and after *in vitro* digestion, kiwifruit pulp was retained at each stage of *in vitro* digestion (pre-digestion, after gastric digestion, and after ileal digestion). Kiwifruit materials were collected by centrifugation at 5,000 g for 15 min at 4℃. Pellets were retained and termed "insoluble kiwifruit fraction" in this study. Supernatant was thoroughly mixed with four volumes of 80% ethanol overnight. Precipitated materials were centrifuged at 5,000 g for 15 min at 4℃, and supernatants were discarded. Precipitated materials were termed "soluble kiwifruit fraction" in this study. Insoluble and soluble kiwifruit fractions from pre-digestion, gastric digestion, and ileal digestion were kept frozen at -80℃ overnight. Samples were freeze-dried at -50℃ at 513 mTorr (Flexi-Dry™ freeze-drier).

2.2.1.5 Chemical constituent analysis of pectin fractions from kiwifruit

Standard sugar mix

Stock solutions containing 5 mg/ml of L-rhamnose, D-fucose, L-arabinose, D-xylose, D-mannose, D-galactose, D-glucose, D-ribose, D-glucuronic acid, and D-galacturonic acid were prepared by dissolving 50 mg of each sugar in 10 ml of distilled water in dry volumetric flasks. 1 ml aliquots of each sugar stock solution were pipetted into the same volumetric flask. 15 ml of distilled water was

added to make up the final volume to 25 ml. The final concentration of each sugar in the stock solution was 0.2 mg/ml. The total sugar concentration in the stock solution was 2 mg/ml.

Trifluoroacetic acid (TFA) containing allose internal standard

10 mg of allose was carefully weighed and dissolved in 25 ml of 2 M TFA. The final concentration of the allose internal standard in TFA was 0.4 mg/ml.

HPAEC-PAD chromatography

Constituent monosaccharide compositions of pectin extracts were analysed using HPAEC-PAD (High-Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection) chromatography at Ferrier Research Institute (Victoria University of Wellington). The protocol was developed and optimized by Tracey Bell and Ian Sims at Ferrier Research Institute. 1 mg of freeze-dried samples of Ethanol, CDTA, Na_2CO_3, insoluble and soluble kiwifruit fractions, and pectin from citrus peels (Sigma) were accurately weighed, and dispensed in duplicate into Kimax screw-cap tubes that were kept in a drying oven overnight.1 ml of standard sugar mix was dispensed in triplicate into Kimax tubes and freeze-dried overnight. 500 µl of 3 N methanolic HCl were into each sample and standard. The tubes were filled with argon before tightly sealing the caps. Mixtures were incubated on a dry block heater at 80℃ overnight. Methanolic HCl was evaporated under a stream of air. 500 µl of 2 M aqueous TFA containing allose internal standard was added into each tube which was then topped with argon. Caps were tightly sealed and mixtures were incubated on a dry heat block at 120℃ for 1 h. TFA was evaporated under a stream of air while heating at 40℃ on a heat block. Hydrolysates were resuspended in 1 ml of distilled water and mixed vigorously. Incubation at 30℃ was carried out for 30 min to solubilize the hydrolysates. 100 µl of hydrolysates were transferred into HPLC vials containing 900 µl of distilled water. 20 µl of aliquots were separated at 30℃ on a 4 × 250 mm CarboPac PA-1 (Thermo Fisher Scientific) column equilibrated in 25 mM NaOH. Elution was carried out at a flow rate of 1 ml/min using simultaneous gradients of NaOH (25-10 mM from 0-10 min, and 10-100 mM from 10-30 min) for separating mono-, di-, tri-, and tetrasaccharides, and sodium acetate (0-500 mM from 30-55 min) for separating larger carbohydrates. Underivatized mono- and

oligosaccharide analytes were monitored by pulsed amperometric detection (PAD), using Dionex (Thermo Fisher Scientific) standard carbohydrate waveform.

2.2.2 Isolation of pectin degrading bacteria from human faeces

2.2.2.1 Acquiring fresh faecal samples

Fresh faecal samples were obtained from a healthy female donor who had not been prescribed with antibiotics for 6 months prior to the sample collection. Faecal sample was transferred into an anaerobic chamber within 5 min of defecation.

2.2.2.2 Pectin Solutions

For the use of irradiated kiwifruit pectin extracts as growth substrates, freeze-dried powders of Ethanol, CDTA, and Na_2CO_3 fractions were mixed in 4:1:1 ratio and were dispensed into the sterile medium at a final concentration of 0.2 % (w/v) inside a UV-sterilized anaerobic chamber. Pectin from citrus peel (esterification ≥6.4 %) and pectin from apple pomace (esterification ≥75 %) were weighed and dispensed into anaerobically prepared water to make up 5 % (w/v) pectin solutions. After dissolving citrus and apple pectins at 65℃ for 25 min, solutions were filtered (0.45 μm, Merck Millipore) into sterile and N_2-washed Hungate tubes. Pectin solutions were added into sterile medium at a final concentration of 0.5 % (w/v).

2.2.2.3 Isolation of pectin-enriched bacteria

Inside an anaerobic chamber, stool was weighed and 5 g of stool was mixed with 50 ml of anaerobic PBS solution to make up 10 % faecal slurry. 10 % faecal slurry was serially diluted (10-folds) in anaerobic PBS solution until 10^9-fold dilution was achieved. 0.5 ml of 10^{-7}, 10^{-8}, and 10^{-9} dilutions were inoculated in 8.5 ml BN medium containing 0.2 % (w/v) kiwifruit pectin mixture. After 15 days of incubation at 37℃ under a constant shaking, 0.3 ml of culture was spread over the roll-tube surface of 9 ml BN agar medium containing 0.2 % (w/v) kiwifruit pectin mixture. After further incubation at 37℃ for 15 days, colonies were picked and inoculated into 9 ml of BN medium containing 0.5 % (w/v) citrus pectin. 0.5 ml of 1-week-old cultures were subcultured three successive times in 9 ml BN medium with or without 0.5 % (w/v) citrus pectin. Cells from 1-2 weeks old cultures were Gram

stained and observed under a light microscope to differentiate the strains which maintained a consistent growth over multiple transfers only in the presence of pectin. Strain 14^T was selected for further characterization.

2.2.3 Characterization of strain 14^T

2.2.3.1 Gram staining

Gram staining was carried out using the method described by Bartholomew and Mittwer (1952).

Crystal violet solution

Solution A was prepared by adding 2 g of crystal violet (dye content ≥ 90 %) into 20 ml of 95 % (v/v) ethanol. Solution B was prepared by dissolving 0.8 g of ammonium oxalate in 80 ml of distilled water. Solution A and B were mixed and stored in dark for 24 h before use.

Gram iodine solution

1 g of iodine and 2 g of potassium iodide were finely grinded using a mortal and a pestle. 300 ml of distilled water was slowly added until iodine completely dissolved. Solution was kept in an amber bottle until use.

Decolourizing agent

50 ml of 95 % (v/v) ethanol and 50 ml of 100 % (v/v) acetone were mixed.

Safranin solution

A stock solution of safranin was prepared by dissolving 2.5 g of safranin O in 100 ml of 95 % (v/v) ethanol. A working solution was prepared by diluting 10 ml of the stock solution in 90 ml of distilled water.

Gram staining method

A small volume of bacterial suspension was spread on a glass microscope slide and air-dried. Air-dried culture was fixed by passing the slide briefly over flame. The slide was flooded with crystal violet solution for 1 min and washed with a gentle stream of tap water. The slide was then covered with Gram iodine solution for 1 min and washed with tap water. Cells were briefly washed with Decolourizing agent and drained well. Safranin solution was used to counterstain for 30 sec. The slide

was washed with tap water and drained using a paper towel. The cells were observed at 1,000× power magnification using immersion oil mount using Olympus CX21 microscope.

2.2.3.2 Endospore staining

Malachite green stain solution

0.5 g of Malachite green was dissolved in 100 ml of distilled water. The solution was kept in the dark overnight before use.

Endospore staining method

Cells of strain 14^T were grown in 9 ml mineral medium containing 0.5 % (w/v) citrus pectin. Cells were incubated at 37℃ for 31 days with a constant shaking. Endospore formation was examined by Schaeffer-Fulton staining method (Schaeffer and Fulton, 1933). A small volume of bacterial suspension was spread on a glass microscope slide. The suspension was air-dried and heat fixed over Bunsen flame. A piece of blotting paper was used to cover the slide which was then saturated with Malachite green stain solution. The slide was steamed for 5 min over boiling water while gentling dripping Malachite green solution to keep the paper moist. The slide was washed with tap water, followed by counterstaining with Safranin solution (from Gram staining method) for 30 sec. The slide was washed with tap water and drained using a paper towel. Cells were observed under the oil immersion lens (1,000×) using Olympus CX21 microscope for the presence of endospores.

2.2.3.3 Cell motility by phase contrast microscopy

A small volume of 1-week-old bacterial culture was added onto a glass microscope slide. A glass coverslip was placed on top of the cell suspension. Excess liquid was dried using a paper towel. Cell motility was observed under phase contrast with a DM2500 microscope at x400 magnification (Leica Microsystems, Germany). Digital images of cells were captured using Leica Application suite software.

2.2.3.4 Transmission electron microscopy (TEM)

Imaging free cells

Cells of strain 14^T were grown at 37℃ for 96 h in 9 ml mineral medium containing 0. 5 % (w/v) D-fructose. Cells were pelleted by centrifugation at 10,000 g for 5 min at room temperature. Cells were fixed, stained, prepared as ultra-thin sections for TEM, and imaged at Manawatu Microscopy & Imaging Centre (Palmerston North, New Zealand). Pelleted cells were fixed in modified Karnovsky's fixative (3 % glutaraldehyde (v/v) 2 % formaldehyde (w/v) in 0.1 M phosphate buffer (pH 7.2)) for 2 h at room temperature. Fixed cells were pelleted by centrifuging at 4,000 g for 4 min, and the fixatives were drained and discarded. 3 drops of 20 % (w/v) bovine serum albumin (BSA) were added to the pellet. Cells were centrifuged again at 4,000 g for 4 min. BSA was coagulated by adding 1 drop of 25 % (w/v) glutaraldehyde and leaving at room temperature for 5 min. Pellets were cut into thin slices and rinsed three times in 0.1 M phosphate buffer (pH 7.2) for 10 min. Cells were post fixed in 1 % (w/v) osmium tetroxide dissolved in 0.1 M phosphate buffer (pH 7.2) for 30 min at room temperature. Cells were rinsed three times in 0.1 M phosphate buffer (pH 7.2) before being dehydrated in increasing concentrations of acetone (25 % (v/v), 50 % (v/v), 75 % (v/v), 95 % (v/v), and 100 % (v/v)) for 10 min each. Cells were further dehydrated by incubating in 100 % (v/v) acetone twice for 1 h each. Cells were stirred overnight in 50:50 (v/v) resin:acetone mix (ProSciTech, Australia). Once the acetone evaporated off, fresh 100 % (v/v) resin (Procure812, ProSciTech, Australia) was added and the cells were stirred for 8 h. This step was repeated twice. Samples were embedded with fresh 100 % (v/v) resin, and cured in a 60 ℃ oven for 48 h. The embedded block of cells was sliced into ultra-thin sections (100 nm) using a diamond knife (Diatome, Austria) on Ultramicrotome (Leica Microsystems, Germany). Sections were stretched over a copper grid and stained with saturated uranyl acetate in 50 % (v/v) ethanol for 4 min. Grids were washed with 50 % (v/v) ethanol and MilliQ water, followed by a further staining in lead citrate for 4 min. Samples were viewed using FEI Tecnai G2 Spirit BioTWIN (FEI Company, Czech Republic).

Bacterial adhesion to the cell walls of orange peel and kiwifruit

Orange and kiwifruit were purchased from a local supermarket. The surface of fruit was wiped using 80 % (v/v) ethanol before placing them in a UV-sterilized Infinity Class II biosafety cabinet (ESCO). Using a sterile razor blade, orange peel and kiwifruit were cut into 0.5-1 mm slices. Fruit sections were transferred to a UV-sterilized anaerobic chamber inside sterile plastic containers. Using sterile forceps, 5 – 10 pieces of fruit sections were dispensed into N_2-washed sterile Hungate tubes that were then sealed with screw caps and rubber stoppers. The tubes were topped with 9 ml of mineral medium which was then inoculated with 1-week-old bacterial cultures. Fruit sections were collected after 72 h of incubation at 37℃ with a constant shaking. Sections were washed briefly with anaerobic PBS and immediately submerged in modified Karnovsky's fixative. TEM imaging was carried out as above without the centrifugation procedures.

2.2.3.5 Genomic DNA extraction

TES buffer (pH 8.0)

0.254 g of Tris-HCl; 0.048 g of Tris; 0.116 g of NaCl; and 0.068 g of EDTA were dissolved in 400 ml of distilled water. The solution was sterilized by autoclaving.

Lysozyme solution

200 mg of lysozyme from chicken egg white (≥ 40,000 U/mg; Sigma) were dissolved in 10 ml of autoclaved water. The solution was dispensed into 1 ml microcentrifuge tubes in 250 µl aliquots, and stored at -20℃ until use.

Proteinase K solution

200 mg of proteinase K from *Tritirachium album* (41 U/mg; Sigma) were dissolved in 10 ml of autoclaved water. The solution was dispensed into 1 ml microcentrifuge tubes in 250 µl aliquots, and stored at -20℃ until use.

RNase solution

200 mg of ribonuclease A from bovine pancreas (≥ 50 Kunitz U/mg; Sigma) were dissolved in 10 ml of autoclaved water. 250 µl aliquots of the solution were kept at -20℃ until use.

TNS solution (pH 8.0)

500 mM Tris-HCl; 100 mM NaCl; and 10 % (w/v) SDS were dissolved in distilled water by autoclaving.

PEG solution

1.6 M NaCl was dissolved in distilled water. 30 % (w/v) polyethylene glycol 6000 was added and dissolved by autoclaving.

EB buffer

10 mM Tris-HCl was dissolved in distilled water. 2M NaOH was used to adjust the pH to 8.5. The solution was sterilized by autoclaving.

Sodium acetate solution

3 M sodium acetate was dissolved in distilled water. Using 2M HCl, the pH of the solution was adjusted to 5.2. Solution was sterilized by autoclaving.

Phenol-chloroform extraction of genomic DNA

Bacterial cell pellets were obtained from 1-week-old cultures grown on RC medium containing 1 % (w/v) D-fructose. Cells were suspended in 0.5 ml of TES buffer and mixed vigorously. 10 µl of lysozyme was added to the cell suspension followed by vigorous mixing. Suspension was incubated in a heat block at 37℃ for 2 h. Samples were immediately transferred on ice. 10 µl of proteinase K and 10 µl of RNase were added to the suspension. The mixture was vortexed and incubated in a water bath at 65℃ for 1 h. 100 µl of TNS solution was added and samples were further incubated at 65℃ for 30 min. Samples were transferred to 2 ml clean Eppendorf tubes. 700 µl of phenol/chloroform/isoamylalcohol (25:24:1, pH8; Sigma) was added and mixed with the sample by gentle inversion. Samples were centrifuged at 20,000 g for 20 min at 4℃. 500 µl of upper layer was carefully pipetted and transferred into a fresh 2 ml Eppendorf tube. 500 µl of chloroform/isoamylalcohol (24:1; Sigma) was added and mixed with the sample by gentle inversion, followed by centrifugation at 20,000 g for 5 min at 4℃. 400 µl of upper layer was carefully pipetted and transferred into a clean 2 ml tube. 800 µl of PEG solution was added to the sample and mixed by

gentling inverting the vial 5 times. Precipitated DNA was centrifuged at 20,000 g for 1 h at 4℃. Liquids were carefully poured out and 500 µl of ice cold 70 % (v/v) ethanol was added to wash the pellet. Precipitates were centrifuged at 20,000 g for 10 min at 4℃. The supernatant was carefully poured out and ethanol-wash step was repeated. Ethanol was carefully poured out and the tubes were kept inverted on a paper towel to air-dry the pellet. DNA precipitates were eluted in 50 µl EB buffer, and left to dissolve in the fridge (4℃) overnight. DNA concentration and purity were checked using Nanodrop®-1000. DNA samples were further purified and concentrated using ethanol precipitation method.

Ethanol precipitation of DNA

Extracted DNA samples were combined and the volumes of DNA samples were carefully measured and recorded. 1/10 volume of sodium acetate solution was added to the DNA samples and mixed by gentle inversion. 2.5 volume of ice cold 100 % (v/v) ethanol was added and mixed by gentle inversion. The mixture was stored at -80℃ overnight. The mixture was centrifuged at 20,000 g for 15 min at 4℃, and the supernatant was carefully poured out. The pellet was washed with 1 ml of ice cold 70 % (v/v) ethanol. Precipitated DNA was centrifuged at 20,000 g for 10 min at 4℃, and ethanol was carefully decanted. Ethanol-wash step was repeated and supernatant was carefully poured out. Tubes were kept upside down on a paper towel to air-dry the pellets. Pellets were eluted in 20 µl of EB buffer and left to dissolve overnight at 4℃.

Quantification of DNA

DNA quantification was carried out using Qubit™ dsDNA BR Assay Kit (Invitrogen) according to the manufacturer's instructions.

2.2.3.6 Agarose gel electrophoresis

TAE buffer (1×)

50× TAE stock solution was prepared by mixing 2 M Tris; 50 mM EDTA; and 60 ml of glacial acetic acid in 940 ml of distilled water. 2M NaOH was used to adjust pH to 8.0 of necessary. Stock solution was diluted 50× in distilled water to make 1× TAE buffer.

SYBR® Green solution

SYBR® Green I Nucleic Acid Gel Stain was purchased from Invitrogen. 100 µl of dye was mixed with 900 µl of 1× TAE buffer (2.6.7.1). Aliquots were stored at -20℃ until use. 100 µl of dye-buffer mixture was added per 100 ml of agarose gel.

Orange G loading dye (10×)

100 mg of orange G and 15 ml glycerol were dissolved in 35 ml of distilled water. The solution was sterilized by autoclaving and stored at -20℃ until use. 1 µl of loading dye was used per 5 µl of DNA sample.

Gel electrophoresis

1 % (w/v) agarose gel was prepared with 1× TAE buffer. DNA was stained using SYBR® Green I Nucleic Acid Gel Stain. DNA samples were mixed with 10 × Orange G loading dye before being loaded onto the gel along with 1 Kb Plus DNA Ladder standard (Invitrogen). 1 L of 1× TAE buffer was used as a running buffer. Electrophoresis was carried out at 80 V for 50 min at ambient temperature using Electrophoresis Subsystem 150 apparatus (LabNet). Bands were visualized by UV Transilluminator at 254 nm, and imaged using GLEDOC software.

2.2.3.7 Phylogenetic analysis

TE buffer

10 mM Tris-HCl and 1mM EDTA were mixed in distilled water. 2M NaOH was used to adjust the pH of the solution to 8.0. Solution was sterilized by autoclaving.

PCR amplification

PCR amplification of 16S rRNA gene of strain 14^T was performed using universal primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1510R (5'-GGTTACCTTGTTACGACTT-3'). 25 mM of lyophilised primers were purchased from Life Technologies NZ Ltd. Primers were reconstituted in TE buffer to make up 100 µM primer stock solutions. HotStarTaq® Master Mix Kit was purchased from QIAGEN. A 50 µl PCR reaction contained 25 µl of HotStarTaq® Master Mix (contains 2.5 units of HotStarTaq® DNA polymerase; 1× $MgCl_2$ (1.5 mM) buffer; and 200 µM of each dNTP); 10 µl of 2

µM forward primer; 10 µl of 2 µM reverse primer; 4 µl RNase-free water; and 1 µl of template DNA. Optimized PCR cycling conditions recommended by the manufacturer were used: initial heat polymerase activation at 95℃ for 15 min, 35 cycles of 95℃ for 30 sec (denaturation), 56℃ for 30 sec (annealing), 72℃ for 90 sec (extension), final extension at 72℃ for 10 min, and storage at 4℃ until further processing. PCR was performed using Applied Biosystems GeneAmp® PCR System 9700. PCR products were cleaned using QIAquick® PCR Purification Kit (QIAGEN) according to the manufacturer's instruction. The purity and quality of DNA samples were checked by Nanodrop®-1000 measurement and gel electrophoresis. DNA quantification was carried out using Qubit™ dsDNA BR assay.

Vector ligation and transformation of competent cells

Purified PCR products were ligated into pGEM®-T Easy Vector (Promega) using 1:5 vector: insert ratio. A standard ligation reaction contained 10 µl 2× Rapid Ligation Buffer; 1 µl of pGEM®-T Easy Vector (50 ng); 5 µl of PCR product (~250 ng); 2 µl of T4 DNA ligase (3 Weiss units/µl); and 2 µl of sterile water. After mixing by gentle pipetting, the reactions were incubated at 4℃ overnight. 3 µl of ligation reaction was added into a vial of One Shot® TOP10 Chemically Competent *E.coli* (Invitrogen) and mixed by gentle tapping. The vials were incubated on ice for 30 min and then heat-shocked for 30 sec at 42℃ in a water bath. Vials were immediately transferred on ice and incubated for further 2 min. 250 µl of pre-warmed S.O.C medium (Invitrogen) was added into each vial of competent cells. Vials were shaken horizontally at 37℃ for 1 h at 225 rpm. 50 µl of transformation mixture was spread on pre-warmed LB agar plates containing 100 µg/ml ampicillin and 1.6 mg X-gal per plate. Plates were incubated at 37℃ overnight. White colonies were selected and grown in 5 ml LB broth medium containing 100 µg/ml ampicillin overnight. Cells were collected for plasmid extraction by centrifuging at 10,000 g for 5 min at room temperature.

Isolation of plasmid DNA

Plasmid DNA was extracted from transformed cells using PureLink® Quick Plasmid Miniprep Kits (K2100-10, Invitrogen) using the manufacturer's instruction.

DNA sequencing

Sequencing of recombinant plasmids was carried out at Genome Service Centre (Massey University, New Zealand) using BigDye® Terminator v3.1 or v3.0 Cycle Sequencing Kit on ABI3730 capillary instrumentation (Applied Biosystems). Each sequencing reaction contained 100-300 ng of DNA; 3.2 pmole primer; 1× BigDye® Terminator chemistry; 5× sequencing reaction buffer; and sterile water to make up the final volume to 20 µl. M13 forward (5'-CCCAGTCACGACGTTGTAAAACG-3') and M13 reverse (5'-AGCGGATAACAATTTCACACAGG-3') primers were supplied by the sequencing service provider. Sequencing results were provided in ABI format.

Phylogenetic analysis

Multiple sequence alignments were performed on ClustalW using default parameters using Mega7 software (Kumar *et al*., 2016). Phylogenetic tree was constructed using neighbor-joining method (Saitou and Nei, 1987). Bootstrap values were calculated using 2,000 re-sampling to evaluate the support of tree topology. Reference 16S rRNA gene sequences from type strains and cloned 16S rRNA gene sequences from uncultured bacteria were obtained from GenBank database. 16S rRNA gene sequence from *Lutispora thermophila* strain EBR46T (GenBank accession number NR_041236) was used as outgroup.

2.2.3.8 G+C mol % analysis of DNA

The G+C content of DNA was determined as a part of whole-genome sequencing process using Illumina HiSeq 2500 (Macrogen, Korea).

2.2.3.9 Cellular fatty acid analysis by GC

Cellular fatty acid contents were determined by gas chromatography by the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, Braunscheweig, Germany).

2.2.3.10 Oxygen sensitivity

Strain 14^T was subcultured 5 times under aerobic and microaerophilic (2 % or 4 % O_2 inside Thermo™ Scientific CO_2 incubator) conditions. Cell viability after 1 week of each transfer was observed using Olympus CX21 microscope.

2.2.3.11 Optimum temperature

Strain 14^T was inoculated in mineral medium containing 0.5 % (w/v) citrus pectin. Cultures were incubated at a range of temperatures (25-45℃ at 5℃ intervals). Optical density at 595 nm was measured daily on Spectronic 20 spectrophotometer (Bausch & Lomb) until stationary phase was reached.

2.2.3.12 Optimum pH

Growth of strain 14^T over a pH range of 5.0-9.0 (0.5 intervals) at 36℃ was monitored. Using 2M NaOH and 2M HCl, pH of RC medium was adjusted before autoclaving. RC medium was used instead of mineral medium as it was difficult to manipulate pH in a CO_2-gased medium. Prior to inoculation, 0.5 % (w/v) citrus pectin was added to 9 ml sterile medium. Changes in optical density at 595 nm were measured daily on Spectronic 20 spectrophotometer (Bausch & Lomb) until stationary phase was reached.

2.2.3.13 Substrate utilization

Carbohydrate Solutions

Starch from wheat, arabinan from sugar beet (Megazyme), xylan from oat spelt, galactan from ex gum arabic (Sigma-Aldrich), β-glucan from barley, arabinoxylan from wheat (Megazyme), and oligofructose-enriched inulin from chicory (Orafti® Synergy1) were weighed and dispensed into anaerobic water to make up 5 % (w/v) solutions. Substrates were dissolved by heating in a water bath at 65℃ for 20 min. Solubilized substrates were filtered (0.45 µm; Merck Millipore) into N_2-washed sterile Hungate tubes. Substrate solutions were added into mineral medium at a final concentration of 0.5 % (w/v) prior to inoculation.

Sugar Solution

D-glucose, D-mannitol, D-lactose, D-sucrose, D-maltose, salicin, D-xylose, L-arabinose, cellobiose, D-mannose, D-sorbitol, L-rhamnose, D-galacturonic acid, D-raffinose, D-galactose, D-fructose, D-galactitol, and D-fucose were weighed and dissolved in anaerobic water to make up 5 % (w/v) sugar solutions. Solutions were filter-sterilized using 0.22 µm filters (Merck Millipore) into N_2-washed sterile Hungate tubes. Prepared solutions were injected into mineral medium at a final concentration

of 0.5 % (w/v) prior to inoculation. 5 % (v/v) glycerol solution was boiled and cooled on ice while bubbling with N_2. 0.05 % (w/v) L-cysteine-HCl was added before dispensing the solution into Hungate tubes. Solution was sterilized by autoclaving. Glycerol solution was added to mineral medium at a final concentration of 0.5 % (v/v).

Insoluble Substrates

Whatman filter papers (number 1) were cut into 5 × 5 mm squares. 10 pieces were added into an empty Hungate tube which was then topped with 9 ml of freshly prepared mineral medium dispensed under CO_2. 50 mg of Avicel® PH-101 (~50 µm particle size; Fluka Analytical) and Sigmacell cellulose (Type 101) were weighed and dispensed into empty Hungate tubes. 9 ml of freshly prepared mineral medium was dispensed into the tubes under CO_2. Sealed medium containing filter papers and cellulose substrates was sterilized by autoclaving.

Growth experiments

The growth of strain 14^T on substrates mentioned above was examined by monitoring the culture viability over five successive transfers. The changes in optical density at 595 nm was measured daily on FLUOstar Optima Microplate Reader (BMG Labtech). Gram stained cells were observed under a light microscope to observe the growth of cells.

2.2.3.14 Size exclusion chromatography (SEC)

Sample preparation

100 µl of clarified rumen fluid, 100 µl of 5 % (w/v) apple or citrus pectin solutions, and 100 µl of 1-week-old inoculum were added in triplicate to 1.7 ml mineral medium. Cultures were grown at 37℃ for 3 days with a constant shaking. 900 µl of samples were taken out at 48 h and 72 h of incubation. These samples were centrifuged at 12,000 g for 10 min at room temperature. Supernatants were transferred to clean Eppendorf tubes and sent to Ferrier Research Institute (Victoria University of Wellington) for SEC analysis.

SEC analysis

Polysaccharide substrates were dissolved in 0.1 M $NaNO_3$ (2 mg/mL), allowed to hydrate fully by standing at room temperature overnight and centrifuged (14,000 g, 10 min) to clarify. The soluble material and samples of spent culture media (100 µL) were injected and eluted with 0.1 M $NaNO_3$ (0.5 mL/min, 60°C) from three columns (TSK-Gel G5000PWXL, G4000PWXL and G3000PWXL, 300x 7.8 mm, Tosoh Corp., Tokyo, Japan) connected in series. The eluted material was detected using a refractive index monitor. The system was also calibrated with a series of pullulan molecular weight standards (6–850 kDa; Shodex, Showa Denko K.K. Tokyo, Japan).

Spent culture media (100 µL) were also injected and eluted with 0.1 M $NaNO_3$ (0.5 mL/min, 60°C) from two Superdex Peptide (GE Healthcare) columns in series. The eluted material was detected using a refractive index monitor. The system was also calibrated with a series of pullulan molecular weight standards (6–24 kDa and the trisaccharide raffinose).

2.2.3.15 Headspace gas production

Production of gas (CO_2 and H_2) was examined by gas chromatography (GC). Using syringes and needles, headspace gas (0.5 ml) was taken from 5 day-old cultures grown on 0.5 % (w/v) citrus pectin. Headspace gas was injected into Aerograph 660 (Varian Associates, Palo Alto, CA, USA) fitted with a Porapak Q80/100 mesh column and a thermal conductivity detector operated at 100℃. The column was operated at room temperature using N_2 as a carrier gas at 12 cm^3/min.

2.2.3.16 Measurement of organic acids using GC

Standard mix solution

Indicated volumes of formate (370 µl), acetate (572 µl), propionate (739 µl), isobutyrate (918 µl), butyrate (909 µl), lactate (655 µl), and succinate (1,175 mg) were added into separate 50 ml volumetric flasks. Distilled water was added the make up the final volume to 50 ml in order to prepare 200 mM stock solutions of each acid. 5 ml of each acid stock solution was transferred and combined in 50 ml volumetric flask. 15 ml of 10 mM PBS was added to make up the volume to 50 ml. The final concentration of each acid in the standard mix solution is 20 mM.

Gas chromatography

Strain 14^T was grown in 50 ml mineral medium containing 0.5 % (w/v) pectin from citrus peels, kiwifruit pectin mixture, D-galacturonic acid, D-xylose, L-arabinose, or D-fructose. Cultures were prepared in triplicate for each substrate. 1 ml of samples were collected daily from growing cultures over a period of 10-14 days. Collected samples were centrifuged at 12,000 g for 10 min at room temperature. Supernatants were transferred into a clean 1.5 ml Eppendorf tube and stored at -80℃ until subsequent analysis. 250 µl of supernatant was mixed with 250 µl of 10 mM PBS containing a final concentration of 5 mM internal standard (2-ethyl butyrate; Sigma). After vigorous mixing, 500 µl of samples was transferred into clean 2 ml Eppendorf tubes. 500 µl of standard mix solutions and a blank (diethyl ether) were also transferred into clean 2 ml Eppendorf tubes. 250 µl of concentrated hydrochloric acid and 1000 µl of diethyl ether were carefully pipetted into samples, standard solutions, and blank. Organic acids were transferred into the diethyl ether phase by vigorous vortexing. Samples were centrifuged at 10,000 g for 5 min at 4℃. The upper layer was transferred into clean Eppendorf tubes for storage (-80℃) or immediate derivatisation. 20 µl of MTBSTFA and 100 µl of diethyl ether extracts were added to a 2 ml GC vial (Grace) which was then tightly capped. Derivatization was carried out in a water bath at 80℃ for 20 min. Vials were further incubated at room temperature for 48 h to allow complete derivatization. Concentrations of organic acids were quantified on a Shimadzu capillary gas chromatograph system (GC-17A, Kyoto, Japan) equipped with a flame ionization detector and fitted with a HP-1 column (10 m × 0.53 mm ID × 2.65 µm) (Agilent Technologies). Helium was used as a carrier gas at a flow rate of 21.2 ml/min at 131.2 kPa. Make-up gas was 100 % nitrogen. The temperature program was initiated at 70℃ increasing to 115℃ at the increment of 6℃ per min, with the final increase to 300℃ at the increment of 60℃ per min. The temperature was held at 300℃ for 3 min. Flow control mode was set to a linear velocity at 37.5 cm/sec. Injector temperature was 260℃ and detector temperature was 310℃. 1 µl of samples were injected with a split injection (10:1 split ratio). Shimadzu GC Work Station LabSolutions software (version 5.3) was used to operate the GC and process the data. The results were expressed as µmol organic acid present in 1 ml of the culture samples.

2.2.3.17 Biochemical tests

Catalase test

A small amount of 1-week-old culture was taken out and smeared on a glass microscope slide. Using a Pasteur pipette, 1 drop of 3 % hydrogen peroxide solution (Sigma) was placed on the smeared cells. Positive results are indicated by the formation of bubbles of oxygen.

Oxidase test

A cotton swab was soaked with the solution of 1 % (w/v) tetra-methyl-p-phenylenediamine dihydrochloride dissolved in distilled water. The swab was dipped into 1-week-old culture and incubated at room temperature for 10-20 sec. Positive results are indicated by violet/purple colour.

Indole test

Strain 14^T was grown in mineral medium which had yeast extract component replaced with 1 % (w/v) tryptophan. The medium was supplemented with 0.5 % (w/v) citrus pectin. 0.5 ml of Kovac's reagent was directly added into triplicates of 1-week-old culture. Kovac's reagent was prepared by dissolving 10 g of *p*-dimethylaminobenzaldehyde 150 ml of isoamyl alcohol by gentle heating. 50 ml of concentrated hydrochloric acid was slowly added into the aldehyde-alcohol mixture. The solution was stored in dark overnight before use. A positive result is shown by red/violet colour in the surface of broth.

Casein hydrolysis test

Skim milk agar contains (per L of the medium) 28 g of skim milk powder; 5 g of casein from bovine milk (Sigma); 2.5 g of Bacto™ yeast extract; 1 g of D-glucose; 0.5 g of L-cysteine HCl; and 15 g of agar. The medium was boiled and cooled on ice under N_2. After adding L-cysteine HCl, 10 ml of cooled medium was dispensed into Hungate tubes under N_2. Sealed medium was sterilized by autoclaving. The medium was cooled to 50℃ in a water bath, and 0.5 % (w/v) citrus pectin was added by injection. Medium was put in a slanted position overnight. The formation of clearing zones on agar slant indicates positive results.

Esculin hydrolysis test

0.445 g of BBL™ Bile Esculin Agar (BD) was pre-dispensed into individual Hungate tubes. 1 L of distilled water was boiled and cooled on ice under N_2. 0.5 g of L-cysteine HCl was added to the cooled water, and 10 ml was dispensed into Hungate tubes under N_2. Sealed medium was sterilized by autoclaving. The medium was cooled to 50℃ in a water bath, and 0.5 % (w/v) citrus pectin was added by injection. Medium was put in a slanted position overnight. Positive results for esculin hydrolysis are indicated by the formation of deep brown or black coloring.

Urease test

29 g of BBL™ Urea Agar Base (BD) was dissolved in 100 ml of distilled water. The solution was boiled with N_2 for 10 min and then filter-sterilized (0.22 µm; Merck Millipore) into N_2-washed sterile Hungate tubes. Urea Agar Base solution was kept at 4℃ until use. 0.15 g of agar powder was pre-dispensed into clean Hungate tubes and then topped with 9 ml of anaerobic water under N_2. Tubes sealed with caps and rubber stoppers were sterilized by autoclaving. The medium was cooled to 50℃ in a water bath, and 0.5 % (w/v) citrus pectin and 1 ml of Urea Agar Base solution was added by injection. The medium was mixed well and then placed in a slanted position at room temperature until the medium solidified. Positive reaction is indicated by color change from orange to pink within ~12 h of inoculation.

Lipase/lecithinase test

0.63 g of BBL™ Baird Parker Agar Base (BD) was added separately into clean Hungate tubes. 10 ml of anaerobically prepared distilled water (index) was added under N_2. After sterilization by autoclaving, medium was cooled to 50℃ in a water bath. 0.5 % (w/v) citrus pectin and 0.5 ml of Egg Yolk Emulsion (Oxoid) were added by injection. Medium was mixed well and placed in slanted position and kept at room temperature overnight. Zones of clearance are expected to form for lipase/lecithinase positive organisms.

Gelatinase test

1 L of nutrient gelatin contains 3 g of Lab-Lemco powder (Oxoid); 5 g of Difco™ peptone; 120 g of gelatin (Fisher Scientific); and 0.5 g of L-cysteine HCl. 1.2 g of gelatin was pre-dispensed into

individual Hungate tubes. Lab-Lemco powder and peptone were dissolved in distilled water, boiled and cooled on ice under N_2. L-cysteine HCl was added to the cooled medium. 10 ml of medium was dispensed under N_2 into the tubes containing pre-dispensed gelatin. Sealed medium was sterilized by autoclaving. The medium was cooled to 50℃ in a water bath, and 0.5 % (w/v) citrus pectin was added by injection. Medium was placed in slanted position and kept at 4℃ overnight. The presence of gelatinase activity is indicated by liquefaction of nutrient gelatin.

2.3 Materials and methods for genome analysis

2.3.1 List of bioinformatics tools

Table 2.1 lists the bioinformatics tools used in this study. Computational tools were selected based on the recommendations from bioinformatics experts, the user-friendliness of the programs, and the research impact inferred from the number of citations. These tools are routinely used to assist the bacterial genome annotation procedures.

2.3.2 Illumina HiSeq 2500 sequencing

The bacterial genome was sequenced at Macrogen (South Korea) using Illumina HiSeq 2500. A paired-end TruSeq DNA PCR-Free (350 bp insert) library and 3 kb and 8 kb Nextera mate-pair (gel plus) libraries were generated for this genome. Sequencing data were digitally delivered in FASTQ format. A total of 11 GB of clean sequencing data were obtained, resulting in approximately 4,000-fold genome coverage. The quality of raw sequencing data was assessed using FastQC (Andrews, 2010).

2.3.3 Pre-processing raw sequencing data

Adaptor sequences present in paired-end reads were trimmed and quality-filtered using Trimmomatic at default settings (Bolger *et al*., 2014). Out of 33,501,036 total reads, 99.74 % of paired-end reads survived the quality-filtering and were included in *de novo* genome assembly. NxTrim was used to trim adaptor sequences from 3 kb and 8 kb mate-pair reads and to select for sets of true mate pairs (O'Connell *et al*., 2015). 100 % of 38,303,662 total reads in 3 kb mate-pair library passed the purity filter set at default parameters and were included in *de novo* genome assembly. 100 % of

Table 2.1 List of bioinformatics tools used in this study.

Tool name	Description	References
ABCdb	ABC transporter database	Fichant *et al*., 2006
BaSys	Web-based genome annotation	Van Domselaar *et al*., 2005
CATH-Gene3D	Protein domain/family prediction	Pearl *et al*., 2005
CDD	Identify conserved protein domains	Marchler-Bauer *et al*., 2011
CGViewer	Graphical visualization of genome properties	Grant and Stothard 2008
CheckM	Assessment of genome quality	Parks *et al*., 2015
ClustalW	Multiple protein sequence alignment	Larkin *et al*., 2007
CRISPR Finder	Database of CRISPR repeats and spacer sequences	Grissa *et al*., 2007
dbCAN	CAZyme identification	Yin *et al*., 2012
eggNOG	Gene orthology inference	Jensen *et al*, 2008
FastQC	Quality control raw sequencing data	Andrews, 2010
GapFiller	*In silico* genome gap closure	Boetzer and Pirovano, 2012
GGDC	Genome-to-genome distance calculator	Meier-Kolthoff *et al*., 2013
HAMAP	Identify protein families	Lima *et al*., 2009
HMMER	Searching sequence homologs	Eddy and Wheeler, 2013
InterProScan 5	A consortium of protein function prediction tools	Jones *et al*., 2014
Islandviewer 3	Genomic island prediction	Dhillon *et al*. 2015
KEGG	Molecular pathway database	Kanehisa *et al*., 2016
Mega7	Sequence alignment and phylogenetic tree construction	Kumar *et al*., 2016
MobiDB	Annotate protein disorder	Domenico *et al*., 2012
NxTrim	Mate-pair sequence trimming and quality control	O'Connell *et al*., 2015
PANTHER	Protein function/domain identification	Mi *et al*., 2013
Pfam	Curated protein family database	Finn *et al*., 2014
PHAST	Prophage search tool	Zhou *et al*., 2011
PIRSF	Protein family classification system	Wu *et al*., 2004
PRED-LIPO	Predict lipoprotein and secretory signal peptide sequences in Gram positive bacteria	Bagos *et al*., 2008
PRINTS	A collection of conserved protein motifs	Attwood *et al*., 2002
Prodigal	Open reading frame prediction	Hyatt *et al*., 2010
ProDom	Protein domain database	Corpet *et al*., 1998
Prokka	Genome annotation	Seemann, 2014
PROSITE	Database of protein families and domains	Hulo *et al*., 2006
RAST	Web-based genome annotation	Aziz *et al*., 2008
RNAmmer	Identify ribosomal operon regions	Lagesen *et al*. 2007
SFLD	Structure-function linkage database	Akiva *et al*., 2014
SignalP	Identify secretory signal peptide sequences	Peterson *et al*., 2011
SMART	Analysis of domain architectures and mobile domains	Schultz *et al*., 2000
SPAdes	*De Novo* genome assembly	Bankevich *et al*., 2012
SPIDER2	Protein secondary structure prediction	Heffernan et al., 2016
SSPACE	Genome scaffold extension	Boetzer *et al*., 2011
SUPERFAMILY	Library of HMM profiles of all proteins of known structure	Gough, 2000
TIGRFAM	Curated collection of protein families	Haft *et al*., 2003
TMHMM	Predict transmembrane helix domains	Krogh *et al*., 2001
Trimmomatic	Paired-end sequence trimming and quality control	Bolger *et al*., 2014
WebLogo	Sequence logo generation tool	Crooks *et al*., 2004

36,608,300total reads in 8 kb mate-pair library were classified as true mate-pairs and were used for additional contig extension.

2.3.4 *de novo* genome assembly

de novo genome assembly was performed on SPAdes (St. Petersburg Genome Assembler) assembler using default parameters (Bankevich *et al*., 2012). Scaffolds generated using an overlapping *K-mer* length of 77 were selected for further processing. Additional scaffold extension was carried out on SSPACE (SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension) using the output data from SPAdes *de novo* assembly and trimmed sequencing data from 8 kb mate-pair library (Boetzer *et al*., 2011). An *in silico* gap closure approach was attempted using GapFiller (Boetzer and Pirovano, 2012). A draft genome consisting of three major scaffolds (782,032 bp, 959,713, and 1,007,898 bp in size) was constructed as a result of *de novo* assembly, indicating the circular genome was fragmented at three sites.

2.3.5 Genome gap closure by primer walking

3 sets of primers were designed to hybridize at 200-300 bp upstream of the truncation sites on each scaffold (Table 2.2). Long-range PCR was carried out to amplify the genome gap sequences using Phusion Green High-Fidelity DNA Polymerase (Thermo® Scientific). 50 µl of PCR reaction contained 10 µl of 5× Phusion Green GC Buffer; 1 µl of 10 mM dNTPs; 5 µl of 5 µM forward primer; 5 µl of 5 µM reverse primer; 1 µl of high quality genomic DNA; 27.5 µl of PCR-grade water; and 0.5 µl of Phusion DNA Polymerase. Optimized PCR cycling conditions recommended by manufacturer was used: initial denaturation at 98℃ for 30 sec; 35 cycles of denaturation at 98℃ for 10 sec and annealing/extension at 72℃ for 10 min; final extension at 72℃ for 10 min; hold at 4℃.

Table 2.2 List of long-range PCR primers used to amplify genome gaps.

Gap #	Orientation	Sequence (5'→3')
1	Forward	GCATGCTGCGGTGAATACGTTCCCGGG
	Reverse	CCACCTCACCACTGAATAAATGGCTTAGGC
2	Forward	CGTTGGGTATGTCACTCGGAGCTGGGAG
	Reverse	AGGGGGAGGCGTTGTTGTGCGTGAGC
3	Forward	CAGAGACCGATTTGTGGAACTGGGATTAATGATTTC
	Reverse	CGCAGTCGAAGGTGGGATTGATAATTAGGGTG

PCR products were cleaned, quality-checked on agarose gel, and quantified as described before. Blunt-end PCR products were ligated into pCR™-BluntII-TOPO® vector using Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen). 6 µl of ligation reaction contained 4 µl of PCR product; 1 µl of Salt Solution; and 1 µl of vector plasmid. Incubation was carried out at room temperature for 30 min before proceeding to transformation of competent cells. Extracted plasmids were sequenced using M13 forward and M13 reverse primers as described before. Using newly designed primers, primer walk sequencing of extracted plasmids was continued until forward-walking sequences overlapped with reverse-walking sequences. Sequences were aligned using Geneious alignment (version 10.0.3) with 65 % similarity cost matrix and 12 and 3 gap open and extension penalties, respectively.

2.3.6 Genome quality assessment

The completeness and quality of the assembled genome was assessed using CheckM (Parks *et al*., 2015).

2.3.7 Genome analysis and annotation

A preliminary annotation of the genome was carried out using Prokka (Seemann, 2014). The draft genome was exported in FASTA format. Open reading frames were predicted using Prodigal, a built-in software in Prokka annotation pipeline (Hyatt *et al*., 2010). Sequences were queried in a hierarchical manner against the default protein database (UniProt) and a series of user-specified databases using BLAST + blastp. By default, a best significant match below an *e*-value threshold of 10^{-6} was used to annotate the putative gene products. Three user-specified databases were constructed by concatenating selected microbial genomes obtained from GenBank (Table 2.3). The "*Ruminococcaceae*" database was built from partial/complete annotated genome sequences of 9 cellulolytic strains of the family *Ruminococcaceae* (*Clostridium* cluster III and IV). The "*Bacteroides*" and "*Dickeya*" databases were built by combining partial/complete annotated genomes of pectinolytic strains of the genus *Bacteroides* and *Dickeya* (formerly known as *Erwinia*). Web-based annotation was carried out on RAST server by digitally submitting the draft genome sequence of *M. pectinilyticus* for automatic ORF prediction and functional assignment of putative genes (Aziz *et al*., 2008). Protein function prediction was carried out using InterProScan 5 which combines a consortium

Table 2.3 List of microbial genomes used to construct user-specified Prokka databases for genome annotation

Databases	Species	GenBank accession	Size (bp)	Status
Ruminococcaceae	*Acetivibrio cellulolyticus* CD2	JH556653	2,192,750	GS[1]
	Clostridium cellulolyticum H10	NC011898	4,068,724	CG[2]
	Clostridium thermocellum ATCC 27405	CP000568	3,843,301	CG
	Clostridium thermocellum DSM 1313	CP002416	3,561,619	CG
	Ruminiclostridium thermocellum AD2	CP013828	3,554,854	CG
	Ruminococcus albus 7	CP002403	3,685,408	CG
	Ruminococcus flavefaciens ATCC19208	KI912489	711,801	GS
	Ruminococcus flavefaciens FD-1	ACOK01000119	201,816	WGS[3]
	Ruminococcus flavefaciens ND2009	KK211346	783,253	GS
Bacteroides	*Bacteroides cellulosilyticus* WH2	CP012801	7,084,828	CG
	Bacteroides finegoldii DSM 17565	GG688317	539,253	GS
	Bacteroides fragilis YCH46	AP006841	5,277,274	CG
	Bacteroides thetaiotaomicron 7330	CP012937	6,487,685	CG
	Bacteroides thetaiotaomicron VPI-5482	AE015928	6,260,361	CG
	Bacteroides xylanisolvens XB1A	FP929033	5,976,145	DG[4]
Dickeya	*Dickeya chrysanthemi* NCPPB 402	CM001974	4,669,628	WGS
	Dickeya chrysanthemi NCPPB 516	CM001904	4,549,457	WGS
	Dickeya chrysanthemi NCPPB 3533	CM001981	4,681,779	WGS
	Dickeya chrysanthemi L11	JSYH01000019	382,940	WGS
	Dickeya chrysanthemi M074	JRWY01000013	253,194	WGS
	Dickeya zeae Ech1591	CP001655	4,813,854	CG
	Pectobacterium carotovorum subsp. *carotovorum* PC1	CP001657	4,862,913	CG

[1]Genome scaffold; [2] complete genome; [3] whole genome shotgun sequencing; [4] draft genome

of protein search tools including CATH-Gene3D (Pearl *et al*., 2005), CDD (Marchler-Bauer *et al*., 2011), MobiDB (Domenico *et al*., 2012), HAMAP (Lima *et al*., 2009), PANTHER (Mi *et al*., 2013), Pfam (Finn *et al*., 2014), PIRSF (Wu *et al*., 2004), PRINTS (Attwood, 2002), ProDom (Corpet *et al*., 1998), PROSITE (Hulo *et al*., 2006), SFLD (Akiva *et al*., 2014), SMART (Schultz *et al*., 2000),SUPERFAMILY (Gough, 2000), and TIGRFAM (Haft *et al*., 2003) in a single software package (Jones *et al*., 2014). For the construction of KEGG metabolic pathways, nucleotide sequences of the target enzyme from 10-15 species of the family *Ruminococcacaea* and family *Clostridiaceae* were manually downloaded from GenBank. The sequences were aligned using Geneious alignment (version 10.0.3) with 65 % similarity cost matrix and 12 and 3 gap open and extension penalties, respectively. Aligned sequences were exported into Linux environment and converted into a HMMER database using *hmmbuild* function (Eddy and Wheeler, 2013). Using *hmmsearch* function, the

database was queried against the draft genome of strain 14^T to find the most significant match to the full target sequence below an *e*-value threshold of 10^{-5}. Genes assigned with enzyme functions were manually mapped into reference metabolic pathways stored in KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa *et al*., 2016). In order to assign each protein-coding gene into an orthologous group, all available Hidden Markov Models (HMMs) for the target taxonomic group (Bacteria) were manually downloaded from eggNOG database (Jensen *et al*, 2008). Obtained sequences were concatenated into a HMMER database using *hmmpress* function. Using *hmmscan* function, all bacterial orthologous groups available in eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database were queried against the genome database of strain 14^T. The most significant matches with the lowest *e*-values were used to predict the orthologous groups of each gene. The presence of signal peptides in each gene was identified using SignalP version 4.1 (Petersen *et al*., 2011). TMHMM (Transmembrane Helix Hidden Markov Models) server was used to predict transmembrane helix domains in proteins (Krogh *et al*., 2001). tRNA sequences were identified using a combination of Prokka and tRNAscan-SE (Lowe and Eddy, 1997). Genes coding for CAZymes were identified by the CAZy database, at our request to Dr Bernard Henrissat (Centre National de la Recherche Scientifique & Aix-Marseille Université). The most recent release of dbCAN (version 5.0) allows a genome-wide *hmmscan* search against the signature domain features of a total of 360 CAZyme families and subfamilies, including GHs, PLs, CEs, GTs, CBMs, dockerins, cohesins, S-layer homology domains (SLH) and AA (auxiliary activities). SLH domains were identified by querying the genome of strain 14^T against dbCAN database and then manually verifying the results by BLAST searching against microbial genomes available in GenBank (Yin *et al*., 2012). Default parameters (alignment length > 80 amino acids, use *e*-value < 10^{-5}, otherwise use *e*-value < 10^{-3}) were used. Each gene entry in the genome of strain 14^T was manually annotated by combining and comparing the annotation results across all databases above. Genes with no apparent matches to any known protein functions or domains were annotated as hypothetical proteins.

2.3.8 Genome curation and depository

The raw sequencing reads were deposited Sequence Read Archive (SRA) under the BioProject accession PRJNA383867, and BioSample accession SAMN06817956. The GenBank accession number for the draft genome of strain 14^T is CP020991 with the name *M. pectinilyticus* CP020991.

2.4 Materials and methods for proteomics analysis

2.4.1 Protein sample preparation

M. pectinilyticus was grown using four different types of substrates: 0.5 % (w/v) D-fructose, 0.2 % (w/v) apple pectin, 0.2 % (w/v) citrus pectin, and 0.2 % (w/v) kiwifruit pectin. For each substrate, 5 × 50 ml of mineral medium containing 5 % (v/v) clarified rumen fluid was used as the base medium. Cultures were incubated at 37℃ with a constant shaking until stationary phase was reached. In order to enable the detection of the highest possible number of proteins potentially related to pectin metabolism, proteins were extracted from *M. pectinilyticus* cultures growing in a stationary phase in which the highest level of protein secretion is believed to take place (Antelmann *et al*., 2001; Voigt *et al*., 2006; Evans *et al*., 2007). Proteins containing an N-terminal signal peptide have also been reported to become expressed and secreted into the growth medium during the stationary phase since these proteins lack cell wall-retention domains such as SLH modules which show strong affinity for peptidoglycan cell wall (Antelmann *et al*., 2001; Rincon *et al*., 2005). Fully grown cultures were combined (250 ml each) and centrifuged at 12,000 g for 10 min. Cell pellets were stored frozen at -80℃. The supernatant was filter-sterilized through 0.22 μm filters (Merck Millipore) and then concentrated by ultrafiltration using 10 kDa Amicon Ultra filters by centrifuging at 4,000 rpm for 20 min (Millipore Analytical). Supernatant proteins were further concentrated using 50 kDa Amicon Ultra filters by centrifuging at 4,000 rpm for 20 min (Millipore Analytical). Pectinase assay and Qubit™ protein assay was carried out for both > 50 kDa and < 50 kDa supernatant fractions to ensure a negligible amount of enzyme loss occurred during ultrafiltration process. Frozen cell pellets and > 50 kDa supernatant samples were sent to Auckland Science Analytical Services (University of Auckland) for quantitative proteomics analysis.

2.4.2 Pectinase assay

2.4.2.1 Imidazole-HCl buffer

Imidazole-HCl buffer (pH 7.2) contained 50 mM imidazole, 20 mM HCl, 1.5 g/L sodium azide, 10 mM $CaCl_2$, 10 mM DDT, and 0.5% (w/v) citrus pectin.

2.4.2.2 DNSA reagent

DNSA solution contained 1% (w/v) 3,5-dinitrosalicylic acid, 0.05 % (w/v) sodium sulfite, and 1 % (w/v) sodium hydroxide. Solution was stored in dark overnight before use.

2.4.2.3 Pectinase assay method

Colorimetric assay was performed using DNSA assay method described by Kashyap *et al*., (2000), developed by Miller (1959), and modified by Aguillar and Huitron (1990) using galacturonic acid standard. 100 µl of concentrated supernatant proteins (> 50 kDa) and the filtrates (< 50 kDa) were added to 100 µl of imidazole-HCl assay buffer. Mixture was incubated at 36℃ for 3 hr under constant shaking. 400 µl of DNSA reagent were added and the mixture was boiled for 15 min to develop the color. Mixture was diluted with 4.4 ml of deionized water. Absorbance was measured at 530 nm using Jenway 6100 Spectrophotometer. Enzyme activity was calculated by defining one unit of enzyme as the amount of enzyme which catalyzes the formation of 1 µmol of reducing sugar per min at a fixed pH. Enzyme activity (U) per µg of protein was calculated.

2.4.3 iTRAQ quantitative proteomics

2.4.3.1 iTRAQ sample preparation

Three independently produced biological replicates of 4-plex iTRAQ were used to analyse the proteome profiles of whole-cell lysates. Due to the limited amount of resources allocated to this part of the study, supernatant proteins were analysed without biological or technical replicates. Cellular proteins were extracted by sonication of cell pellets in lysis buffer (50 mM ammonium bicarbonate containing 7M urea, 2M thiourea and 10 mM DTT). The total protein content of each sample was assayed using the EZQ Protein Quantitation Kit (Life Technologies). Aliquots containing 25µg of protein were taken for each sample and volumes adjusted to 100ul. Samples were reduced by

incubation at 56°C for 30 minutes in the presence of 10 mM DTT. Samples were cooled and alkylated with 50mM iodoacetamide (GE Healthcare) in the dark at RT for 30 minutes, and digested with 1µg of sequencing grade trypsin (Promega). Digests were acidified with formic acid, desalted on Oasis HLB SPE cartridges (Waters), and dried down in a vacuum centrifuge. Samples were reconstituted with 30ul of 0.5M TEAB (Sigma) and labelled with 4-plex iTRAQ labels (Sciex) according to the manufacturer's supplied protocol. Labelling reactions were concentrated in a vacuum centrifuge and resuspended in 0.1% formic acid. Pools were prepared using equal amounts of each sample, and desalted on Oasis HLB cartridges as above.

2.4.3.2 LC-MS/MS and database search

Samples were injected onto a 0.3 × 10 mm trap column packed with Reprosil C18 media and desalted for 5 min at 2 µl/min before being separated on a 0.075 × 200 mm picofrit column (New Objective) packed in-house with Reprosil C18 media. The following gradient was formed at 300 nl/min using a NanoLC 400 UPLC system (Eksigent): 0 min 10 %B; 50 min, 35 %B; 52 min, 90 %B; 55 min, 90 %B; 56 min, 10 %B; 60 min, 10 %B, where A was 0.1 % formic acid in water and B was 0.1% formic acid in acetonitrile. The picofrit spray was directed into a TripleTOF 6600 Quadrupole-Time-of-Flight mass spectrometer (Sciex) scanning from 350-1200 m/z for 250 ms, followed by 40 ms MS/MS scans on the 40 most abundant multiply-charged peptides (m/z 60-1200) for a total cycle time of ~2 seconds. The mass spectrometer and HPLC system were under the control of the Analyst TF 1.7 software package (Sciex). The resulting data from each pool were searched against a database containing the UniProt sequences for pentapetalae from August 2015 (1,895,602 entries) appended to a custom database of *M. pectinilyticus* entries including common contaminant sequences (2418 entries) using ProteinPilot version 5.0 (Sciex). Raw iTRAQ peak area data were processed through a combination of Paragon™ and Pro Group™ algorithms to reduce the protein inference ambiguities coming from protein modifications and to bundle peptides into winner protein groups. False Discovery Rate analysis was enabled. Search parameters were as follows: Sample Type, iTRAQ 4-plex (Peptide Labelled); Search Effort, Thorough; Cys Alkylation, Iodoacetamide; Digestion, Trypsin: FDR analysis, Yes. Using the peak height of reporter ions as a proxy for marker mass abundance,

peptide ratios across 114 (apple), 115 (citrus), 116 (fructose), and 117 (kiwifruit) samples were determined in log space using 116 values as the baseline denominators. After performing bias correction by applying a correction factor of < 20 % across the samples, an average ratio was calculated for each protein. P-values were calculated to assess the possibility of random distribution of peptide ratios contributing to protein inference. Peptide and Protein summary files were exported in Excel format for further statistical analyses.

2.4.3.3 Statistics

While the direction of changes in differential expression are not usually affected, iTRAQ experiments rarely reported fold-change ratios exceeding 2 (Pascovici *et al*., 2015). Therefore, the typically used 2- or 0.5-cutoffs to estimate up- or down-regulation of protein production are often considered too strict, and were readjusted with less stringent thresholds of 1.2- and 0.83-folds in this study (Pascovici *et al*., 2015). The large proportion of identified proteins represented by a single peptide is another issue causing a significant amount of data exclusion from the total proteome coverage in iTRAQ datasets (Evans *et al*., 2012). From the acquired dataset, it was noted that proteins represented by one or two unique peptides often suffer from a high statistical variability. Because reliability of fold-changes ratios are often directly proportional to the number of peptides contributing towards the ratio calculation, only those proteins represented by ≥ 3 unique peptides with $>$ 95 % confidence intervals were included for further analysis in this study (Pascovici *et al*., 2015). To ensure the statistical significance of the dataset, unused protein score of > 2, and *p*-value of < 0.05 inferred from using ProteinPilot software (version 5.0) were coupled with the quantification data to remove unreliable peptide identification results. In ProteinPilot, unused scores reflect the amount of total and unique peptide evidence which identifies with the protein hit. While a protein with an unused score of 0 is considered unreliable as it suggests that no unique peptide was identified for this particular protein, an unused score of > 2 gives a solid metric for deciding if a protein has been confidently identified. In order to inspect the presence of potential peptide outliers, Q-Q plots were constructed for each protein by plotting the logarithm of peptide ratios against the expected normal distribution. The qqmaths function in R package lattice was used to produce the graphs.

2.5 Materials and methods for human carriage study

2.5.1 Participant selection

For this part of study, ethical approval was obtained from the Massey University Human Ethics Committee (Southern A, Application 15/34). 44 participants were recruited for faecal sample donation and dietary intake assessment by Genelle Healey (Institute of Food Science and Technology, School of Food and Nutrition, Massey University) as a part of her PhD project "The influence of habitual dietary fibre intake on the responsiveness of the gut microbiota to a prebiotic". 19 - 65 years old male/female participants were selected based on the following criteria: 1) has not taken antibiotics within the last 6 months of study; 2) has not taken laxatives and prebiotic/probiotic supplements within the last month; 3) no past history of food intolerance or clinically significant diseases including cancer, gastrointestinal disorders, autoimmune disorders, diabetes, heart disease, or renal failure; 4) body mass index (BMI) within the range of 18.5 – 30 kg/m^2; 5) no significant changes in weight or dietary habits within past year; 6) not pregnant or breastfeeding; and 7) non-smokers and standard alcohol consumers (< 15 drinks per week for males and < 10 drinks per week for females; more than two alcohol-free days per week). Each participant completed 4 sets of 3-day diet records over a period of 10 weeks. Dietary fibre and food composition analysis was conducted using Foodworks version 8 software (Xyris Software Pty Ltd). Participants were divided into high, moderate, and low dietary fibre intake groups. The high dietary fibre intake cut-offs were chosen to reflect the New Zealand recommended dietary fibre intake which is > 25 g/day for females and > 30 g/day for males. The average dietary fibre intake in New Zealand (17.5 g/day for females and 22.1 g/day for males) was chosen as the low dietary fibre intake cut-offs, which is below recommended amounts. The amount of pectin consumed per day by each participant was calculated by using an established food composition database (Marlett and Cheung, 1997).

2.5.2 DNA extraction from faecal samples

Bacterial DNA was extracted from 0.25g of faecal samples using MoBio PowerLyzer® Powersoil DNA® isolation kit as per the manufacturer's instructions. A FastPrep-24™ 5G (MP Biomedicals) was

used to homogenise the samples. The DNA was eluted in EB buffer. NanoDrop®-1000 spectrophotometry was used to quantify the DNA concentration.

2.5.3 Primer design and specificity testing

The primer sets used in this study are listed in Table 2.4. The *M. pectinilyticus*-specific primers sets were designed by using NCBI Primer Blast (Ye *et al*., 2012). Using Primer Blast, the primer specificity was virtually tested against 16S rRNA gene sequences from cultured and uncultured bacterial species available in the NCBI database. Parameters were set to meet the primer requirements recommended by Lightcycler® which suggested the use annealing temperatures around 60℃, and an optimal amplicon size of ~500 bp. The optimum annealing temperatures were calculated using OligoCalc, an online-based calculator for oligonucleotide properties (Kibbe, 2007). The specificity of *M. pectinilyticus*-specific primers was further tested by performing colony PCR using frozen stock cultures of the following strains; *Eubacterium rectale* ATCC 35183, *Faecalibacterium prausnitzii* DSM 17677, *Roseburia intestinalis* DSM 14610, *Lactobacillus plantarum* ATCC 14917, *Lactobacillus acidophilus* ATCC 11975, *Staphylococcus aureus* ATCC 25923, *Ruminococcus gnavus* ATCC 29149, *Lachnospira multipara* ATCC 19207, and *M. pectinilyticus*. With the exception of *L. multipara* which was isolated from the animal rumen, all strains were human enteric bacterial species of the phylum *Firmicutes*, available from the in-house bacterial culture collection. PCR-grade water was used as a negative control. Each 20 µl of PCR reaction contained 4.7 µl of PCR-grade water; 4 µl of 5× Phusion Green HF Buffer (Thermo® Scientific); 0.4 µl of 10 mM dNTPs; 5 µl of 5 µM forward primer; 5 µl of 5 µM reverse primer; 0.2 µl of Phusion DNA Polymerase (Thermo® Scientific); and 1 µl of frozen stock cultures or negative control. Optimized PCR cycling conditions

Table 2.4 Sequences of universal and *M. pectinilyticus*-specific PCR primers used in this study.

Primer set	Target	Sequence (5'→3')	Annealing Tm (℃)	Product size (bp)
Uni331F	Bacteria	TCCTACGGGAGGCAGCAGT	58	466
Uni797R		GGACTACCAGGGTATCTAATCCTGTT	55	
MP1087F	*M. pectinilyticus*	GAGCGCAACCCTTACTGTCA	54	495
MP1581R		CTCTTACTTCCGCTCTCCGC	56	

recommended by the manufacturer was used: initial denaturation step of 98 ℃ for 30 sec; 30 cycles of 98 ℃ for 10 sec (denaturation), 65 ℃ for 20 sec (annealing), and 72 ℃ for 15 sec (extension); and a final elongation step at 72 ℃ for 5 min. Purified PCR products were analysed on 1 % (w/v) agarose gel, and the bands were visualized by E-Gel® Safe Imager™ (Invitrogen™).

2.5.4 Preparation of qPCR standard

M. pectinilyticus was grown in RC medium containing 0.5 % (w/v) D-fructose. *E.coli* strain Nissle was grown on standard LB broth medium. Cell densities were determined by manual cell counting using a Neubauer haemocytometer. Cultures were diluted using a sterile medium to achieve 1.0 x 10^9cells/ml. Diluted cultures were centrifuged at 12,000 g for 10 min at room temperature. Cell pellets were homogenized using FastPrep-24™ 5G (MP Biomedicals). Genomic DNA was extracted using MoBio PowerLyzer® Powersoil DNA® isolation kit as per the manufacturer's instructions. Extracted DNA was eluted in EB buffer and quantified using Qubit™ dsDNA BS assay. Genomic DNA samples were serially diluted (10-fold) to construct standard curves ranging between 1.0×10^3 cells/ml and 1.0×10^9 cells/ml.

2.5.5 Quantitative PCR

Samples and standards were run in triplicate by absolute quantification on Roche Lightcycler® 480 real-time PCR instrument. Lightcycler® 480 SYBR Green I Master Mix was used for specific detection of double-stranded PCR-amplified products. A 20 µl reaction contained 10 µl SYBR® Green I Master Mix; 4 µl of 2.5 µM forward primer; 4 µl of 2.5 µM reverse primer; and 2 µl of template DNA from samples and standards. Negative controls were prepared with PCR-grade sterile water in place of DNA samples. Quantitative PCR was performed using the following conditions: one activation cycle at 95℃ for 10 min; 45 run cycles of denaturation (95℃ for 10 sec), annealing (60℃ for 20 sec), and extension (72℃ for 20 sec); one cycle of melting (95℃ for 30 sec, 65 ℃ for 1 min, followed by 65℃ to 95℃ at 0.1℃ increment per second with continuous fluorescence acquisition); and a cooling cycle at 40℃. Results were analysed and visualized using Lightcycler® 480 software package (version 1.5).

2.5.6 Statistics and calculations

Log_{10} concentration of *M. pectinilyticus* and the % abundance of *M. pectinilyticus* relative to the total bacterial concentration were plotted to determine the split point at which the participants were separated into *M. pectinilyticus*-positive and *M. pectinilyticus*-negative groups. A chi-square test was performed to calculate *p*-values to assess the distribution relationships between high fibre consumption and the number of participants showing a positive presence of *M. pectinilyticus*. The median amounts (g) or servings of fibre, pectin, vegetable, fruit, grain, and protein intake per day were calculated for *M. pectinilyticus*-positive and –negative groups, and the statistical significance of the data was assessed based on *p*-values obtained through performing a non-parametric Mann-Whitney-U test. The p-value < 0.05 was considered statistically significant.

2.5.7 Metagenomic analysis

Full-length protein sequences of 23 S-layer homology (SLH) domain-containing proteins from the *M. pectinilyticus* were manually extracted from the genome database. As the SLH proteins of *M. pectinilyticus* did not find sequence matches with a significant homology in the currently available protein databases (Chapter 4), this suggested that the possession of these proteins may be a unique characteristic of *M. pectinilyticus* and possibly its closely related uncultured bacterial species. Therefore, these SLH protein sequences were used as query sequences to search for the presence of *M. pectinilyticus* in the metagenome databases constructed from the stool samples of 85 donors living in the US, as part of the Human Microbiome Project (HMP) initiated by National Institutes of Health (NIH). Although the faecal samples were collected from each donor over two separate visits, only the data from the first visit was used in this study. The metagenome databases were accessed through the Integrated Microbial Genomes with Microbiome Samples (IMG/M) system of Joint Genome Institute (JGI) funded by the United States Department of Energy (Chen *et al*., 2017). A built-in Blast Genome function of IMG/M system was used to carry out metagenome mining in which the FASTA sequences of 23 SLH proteins were simultaneously searched against the selected databases using BlastP with an *e*-value threshold of 1e-5. A stringent cut-off was used to identify true positives which showed 80 –

100 % identical amino acid residue matches along the same positions over a relatively long (> 100 amino acids) length of aligned protein sequences (Nelson *et al*., 2010).

Chapter 3 Isolation and characterization of *Monoglobus pectinilyticus* 14^T

3.1 Introduction

This chapter describes the isolation and characterization of *Monoglobus pectinilyticus* 14^T, a pectin-degrading bacterium from human faeces. This work was derived from the necessity to fill in the currently vacant niche for specialist pectin degraders in the human gut. It was hypothesized that a hierarchical cross-feeding relationship between pectin specialists and generalist bacteria (e.g. *Bacteroides* spp.) may lead to a more efficient pectin digestion in the human gut through a hand-down transfer of more readily digestible form of pectic oligomers for microbial utilization. The use of high molecular weight pectin retaining a complex structure was considered ideal for isolating bacteria that can degrade the pectin backbone while comprehensively removing the side chains and the methyl/acetyl groups. Therefore, kiwifruit pectin retaining a heterogeneous composition of side chains was prepared as the enrichment substrate for isolating pectinolytic bacteria. Since a naturally occurring pectin > 800 kDa in size with 60~90 % degree of methylation is unlikely to maintain its intact structure during the transit through the gastrointestinal tract, kiwifruit was initially treated with *in vitro* digestion prior to extraction to expose the fruit to simulated gastric and ileal conditions. *In vitro* digestion treatment of green kiwifruit had been previously shown to increase the release of soluble pectin materials from the plant cell wall at the expense of moderate decreases in the molecular weight of pectin and degree of methyl esterification (Carnachan *et al.*, 2012). Alkaline fractionation method was chosen to extract pectic materials from the digested kiwifruit, as this method was reported to achieve up to 82 % of polygalacturonic acid extraction from the fruit cell wall in a previous study (Redgwell *et al.*, 1997). The use of increasingly alkaline conditions (ethanol, CDTA, and Na_2CO_3) further fractionated the branched components of pectin (RG-I and RG-II) from the plant cell wall, as reflected by the changes in the ratio of uronic acid and rhamnose in these fractions (Parkar *et al.*, 2010). After the extraction, kiwifruit pectin was sterilized using gamma-irradiation to avoid the extensive pectin depolymerisation (up to 90 % reduction in molecular weight) observed during harsh thermal sterilization procedures (Munarin *et al.*, 2013; Dongowski *et al.*, 2002). Finally,

the constituent sugar analysis of kiwifruit pectin extracts was carried out using 'High-Performance Anion-Exchange Chromatography Coupled with Pulsed Electrochemical Detection' (HPAEC-PAD) (Wee *et al*., 2014).

Using kiwifruit pectin as the sole added source of carbohydrate, a novel pectin-degrading bacterium (strain 14^T) was isolated from human faeces. Strain 14^T was particularly interesting due to its inability to grow without pectin substrates in the growth medium. TEM microscopic examination showed a strong localization of bacterial cells of strain 14^T in proximity to the plant cell wall and within the middle lamella regions. Strain 14^T was selected for further characterization due to the strong pectinolytic trait and its phylogenetic novelty within the family *Ruminococcaceae*. As a prelude to genome sequencing, phenotypic and genotypic characterizations of strain 14^T were performed in accordance with the guidelines provided by International Journal of Systematic and Evolutionary Microbiology (IJSEM). Strain 14^T was given a taxonomic name *Monoglobus pectinilyticus* gen. nov., sp. nov., type strain 14^T (=JCM 31914^T= DSM 104782^T).

3.2 Constituent sugar analysis of pectin extracts from kiwifruit

3.2.1 Soluble and insoluble kiwifruit fractions

Table 3.1 summarizes the constituent sugar composition of insoluble and soluble polysaccharides from kiwifruit following the simulated gastric and ileal digestions, in comparison with the undigested control. Glucose was the most abundantly present sugar in all insoluble fractions, followed by galactose and xylose, reflecting the presence of cellulosic and hemicellulosic materials. The amount of neutral sugars in insoluble kiwifruit fraction changed little after gastric digestion, but slightly

Table 3.1 Constituent sugar composition of insoluble and soluble fractions from pre-digested, gastric digested, and ileal digested green kiwifruit materials

		Sugar (μg/mg of dry weight)								
		Fuc	Rha	Ara	Gal	Glu	Man	Xyl	Neutral	UA
Insoluble	Pre-digested	3.2	3.1	7.8	37.2	128.7	13.9	37.1	231.0	26.6
	Gastric	3.2	3.1	7.5	41.0	127.5	14.6	38.9	235.8	26.6
	Ileal	4.2	3.5	8.5	47.6	138.9	18.5	51.6	272.8	24.6
Soluble	Pre-digested	2.0	9.8	9.9	34.0	68.8	2.7	2.0	129.2	318.5
	Gastric	2.0	9.4	9.4	34.5	91.1	2.2	1.9	150.5	334.5
	Ileal	1.8	9.6	9.5	33.8	59.0	3.4	0.0	117.1	308.3

increased after the ileal digestion treatment. Regardless of the digestion status of kiwifruit, the amount of uronic acid in insoluble fractions made up approximately 10 % of the total carbohydrate, which slightly decreased after the ileal digestion. In soluble fractions of pre-digested, gastric digested, and ileal digested kiwifruit, approximately 70 % of the total polysaccharide was made of uronic acid, suggesting that most pectin-rich materials were separated from the insoluble kiwifruit materials by using simple ethanol precipitation. The sugar compositions of soluble and insoluble kiwifruit polysaccharides changed little after the *in vitro* digestion treatment, consistent with the previously reported results from literature (Carnachan *et al.,* 2012). The increase in the amount of uronic acid in soluble fractions after *in vitro* gastric digestion was also consistent with the previous finding which reported that the low pH used during gastric digestion increased the yields of soluble sugars and uronic acid (Carnachan *et al.,* 2012). The amount of uronic acid in soluble fraction was the lowest after ileal digestion treatment for this study and the reference literature (Carnachan *et al.*, 2012). The effect of *in vitro* ileal digestion on the pectin structure may be attributed to the acidic pH used during the gastric digestion, and the activity of endogenous pectinases released from the disrupted plant cells. This study did not determine the molecular weight and the degree of pectin methylesterification of the digested kiwifruit pectin. In the literature, the molecular weights changes of ileal-digested kiwifruit pectin were marked by reductions in high molecular weight materials (~ 500 kDa) and increases in lower molecular weight materials (~30 kDa) (Carnachan *et al.*, 2012). The literature also reported that there was ~14 % reduction in the degree of methylesterification of kiwifruit pectin after ileal digestion (Carnachan *et al.*, 2012). The structure and the constituent sugar composition of ileal-digested kiwifruit pectin should reflect the post-ingestion status of kiwifruit pectin available for bacterial fermentation in the large intestine.

3.2.2 Fractionation of branched pectin from insoluble kiwifruit materials

In order to extract the remaining pectic materials cross-linked to cellulose and hemicellulose from the ileal-digested insoluble kiwifruit fraction, additional extractions were carried out using CDTA, and Na_2CO_3 (2.2.1). The size of *in vitro* digestion and fractionation experiments were scaled up to prepare a sufficient amount of substrate for the isolation experiment. Table 3.2 summarizes the

Table 3.2 Constituent sugar compositions of pectic oligosaccharides extracted from green kiwifruit digested *in vitro*.

	Sugar (µg/mg of dry weight)								
	Fuc	Rha	Ara	Gal	Glu	Man	Xyl	Neutral	UA
Ethanol	3.8	9.4	15.3	36.0	82.3	4.4	6.3	157.4	428.3
CDTA	10.4	13.7	17.6	43.9	6.9	9.2	5.1	106.8	614.1
Na_2CO_3	1.6	15.7	12.3	34.0	11.9	1.6	6.9	84.0	345.0
Citrus pectin	1.4	16.0	21.3	29.1	2.9	0.0	0.0	70.7	754.0

constituent sugar analysis results for ileal-digested kiwifruit pectin sequentially extracted in Ethanol, CDTA, and Na_2CO_3, in comparison to the commercial pectin from citrus peels (Sigma). Uronic acid, which is a major constituent sugar of the HG and RG backbones, was predominantly found in all fractions, constituting approximately 73 %, 85 %, and 80 % of the total carbohydrate, respectively. These values were slightly lower compared to the uronic acid content of citrus pectin (92 % of the total carbohydrate). However, kiwifruit pectin fractions showed greater heterogeneity in the neutral sugar composition, and higher neutral sugar-uronic acid ratios compared to citrus pectin. The absence of mannose and xylose in citrus pectin indicated that the non-covalent linkages between RG-I and hemicellulose were probably extensively trimmed during the chemical processing. Starch contaminant released from the disrupted plant cell walls may have caused the large amount of glucose present in Ethanol fraction. The rhamnose and uronic acid ratio, which reflects the proportion of RG-I region in the pectin, was highest for Na_2CO_3 fraction, and was similar between Ethanol and CDTA fractions. Arabinose, galactose, and fucose were found in the largest amount in CDTA fraction, indicating the structural components of RG-I and RG-II were likely extracted in CDTA and Na_2CO_3. In order to isolate pectinolytic bacteria from human faeces, Ethanol, CDTA, and Na_2CO_3 fractions were mixed in a 4:1:1 ratio to reflect the general proportions of HG (~70 %), RG-I (20 – 35 %), and RG-II (~10 %) in pectin.

3.3 Isolation and selection of strain 14^T

Strain 14^T was isolated from 10^{-8}-dilution of fresh faeces of a healthy female donor living in NZ. Strain 14^T grew poorly on the roll-tubes and on agar surfaces, forming a small number of colonies which were too small in size to allow an accurate morphological description. The BN medium used

for isolation was not optimal to grow strain 14^T, as demonstrated by the negligible changes in optical density (595 nm) and the small number of cells observed under the light microscope. The use of mineral medium and RC medium significantly improved the growth of strain 14^T. While mineral medium did not contain background carbohydrates, RC medium contained 0.5 % (w/v) D-glucose and 0.1 % (w/v) soluble starch as integral ingredients. Strain 14^T did not grow on the glucose- and starch-containing RC medium, unless pectin was also present. The inability of strain 14^T to grow on these readily usable types of carbohydrates indicated that the strain had an unusual nutrient requirement for pectin for growth. In addition to the availability of pectin, strain 14^T also required clarified rumen fluid and O_2-free strict anaerobic conditions to sustain growth over continuous multiple transfers. The requirement for rumen fluid to be included in the growth medium usually indicates the needs for trace metals, vitamins, or ions of microbial/fungal origins which are present within the anaerobic gut environment as a result of the fermentation of food materials and the extensive microbial cross-feeding interactions (Saleem *et al.,* 2013). Strain 14^T was selected for further characterization based on the preliminary observations which indicated a strong preference for pectin as a growth substrate and the signs of adaptation in the anaerobic gut environments.

3.4 Morphological descriptions of strain 14^T

3.4.1 TEM examination

Cells of strain 14^T were cocci (approximately 0.6 μm in diameter) most often found as single cells or in pairs. A relatively thick Gram positive type bacterial cell wall structure (50 – 60 nm) was present (Figure 3.1a). Strain 14^T was not encapsulated and did not form extracellular coat materials on the cell surface. Cells lacked flagella and were confirmed to be non-motile by phase-contrast microscopic examination. Although the penetration of structurally intact plant cell walls did not occur after 72 h of incubation, strain 14^T formed zones of degradation in the pectin-rich regions (middle lamella) overlaying the plant cell walls (Figure 3.1b). Some bacterial cells appeared to degrade fibrous material in proximity to undamaged plant cell walls (Fig. 3.1c, d). Most cells were positioned at a slight distance (~50 nm) from the sites of attachment, possibly indicating diffusion of degradative

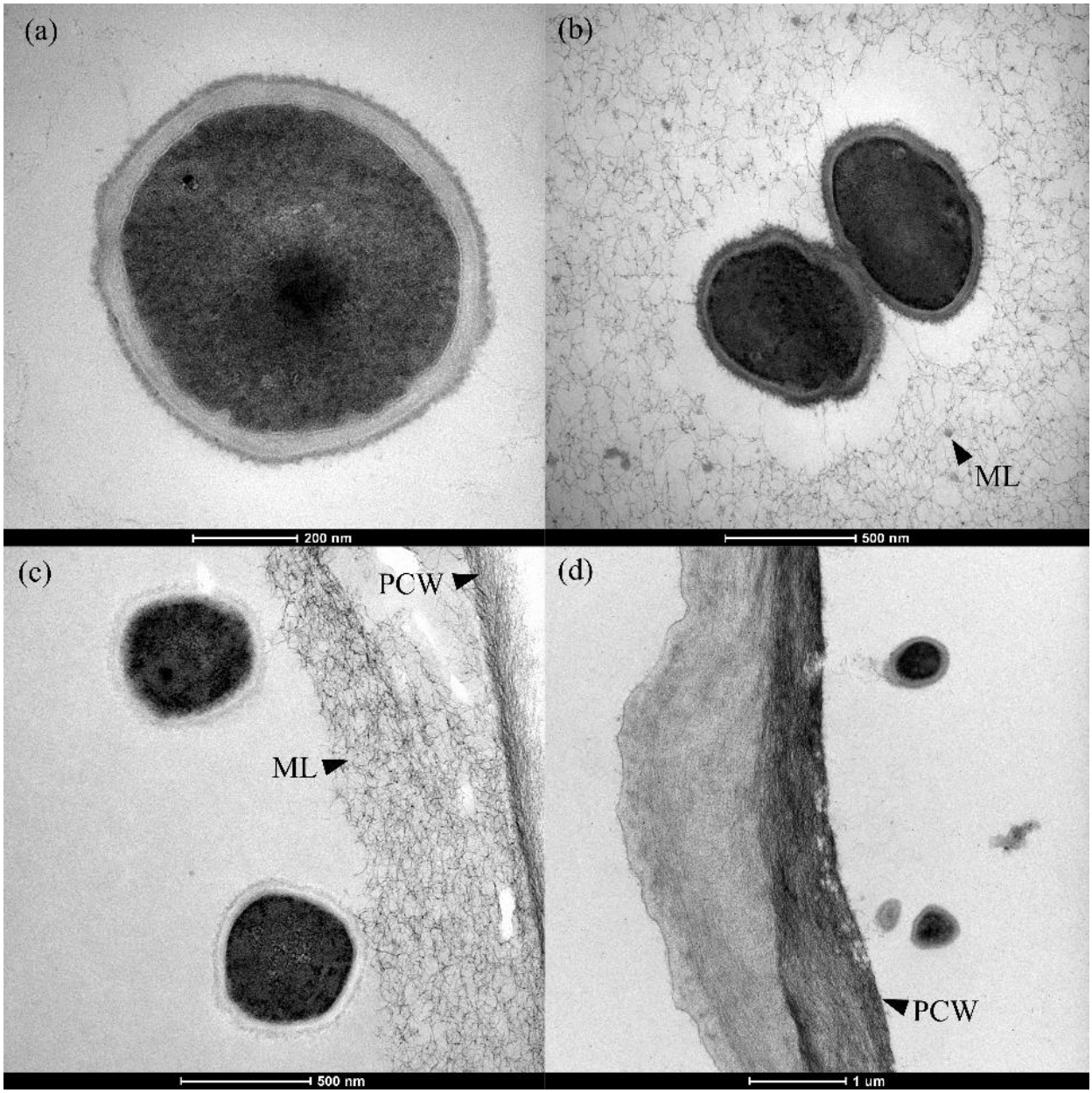


Figure 3.1 Micrographs of strain 14^T. (a) TEM micrograph of strain 14^T grown in mineral medium plus 0.5 % (w/v) fructose. (b, c) TEM micrographs of strain 14^T showing attachment to the middle lamella of kiwifruit (b) and orange peel (c). Liquid culture-grown cells bound to thin slices of kiwifruit or orange peel were collected after 72 h of incubation at 37℃. (d) Cells of strain 14^T in proximity to the plant cell wall of orange peel. ML, middle lamella; PCW, plant cell wall.

enzymes from bacterial cells. Extremely thin fibres from the middle lamella and the plant cell wall appeared to be anchored to the surface of the bacterial cell wall. Adhesion of bacterial cells at damaged or cut surfaces of the plant cell walls was occasionally observed, indicating strain 14^T may also be capable of degrading pectin materials present within the plant cell wall.

3.4.2 Endospore formation

Endospore formation was not detected from nutrient-deprived cells after 4 weeks of incubation at 37℃.

3.5 Phylogenetic analysis

Note that only near full-length 16S rRNA sequences (1,350 – 1,500 bp) were used for conducting phylogenetic analysis in this study (2.2.3.7). The neighbour-joining tree of 16S rRNA gene sequences grouped strain 14^T within the family *Ruminococcaceae* (Figure 3.2). Strain 14^T shared the highest 16S rRNA gene sequence similarities with the members of *Clostridium* cluster III, *Acetivibrio cellulolyticus* NRC 2248^T (89.7 %), *Clostridium clariflavum* DSM 19732^T (89.3 %), *Clostridium thermocellum* ATCC 27405^T (88.8 %), and *Clostridium aldrichii* DSM 6159^T (89.1 %). These members of *Clostridium* cluster III are cellulolytic Gram positive rods isolated from non-gut anaerobic environments, such as anaerobic sewage and agricultural waste fermenters (Patel *et al*., 1980; Madden, 1983; Kato *et al*., 2004; Shiratori *et al*., 2009; Lv and Yu; 2012). Of the strains isolated from the human gut, *Ruminococcus bromii* ATCC 27255^T (85.4 %) and *Ruminococcus champanellensis* $18P13^T$ (84.5 %) were the closest relatives of strain 14^T. Overall, these sequence similarity values to cultured bacteria were considerably below the threshold values used for species (≥ 97 %) and genus (≥ 95 %) delineations (Qin *et al*., 2014). It was noted that the phylogenetic divergence of strain 14^T occurred outside *Clostridium* cluster III, forming a distinct lineage of descent within the family *Ruminococcaceae*. The closest GenBank relatives of strain 14^T were 16S rRNA gene sequences of uncultured bacterial species derived from human gut communities, with 16S rRNA gene sequence similarity of ~ 99 %. An unfiltered-BLAST search against GenBank database identified 63 near full-length cloned 16S rRNA gene sequences from uncultured bacteria of various gut origins that had ≥ 92 % sequence similarities to strain 14^T (Appendix 4). These sequences were obtained from 19 independent studies with no apparent geographical or periodic correlations. Of the 63 sequences, 50 sequences were isolated from the human gut, and the rest of the sequences were isolated from animal guts including cattle, mice, rabbit, pig, gazelle, and dhole. The 16 rRNA gene

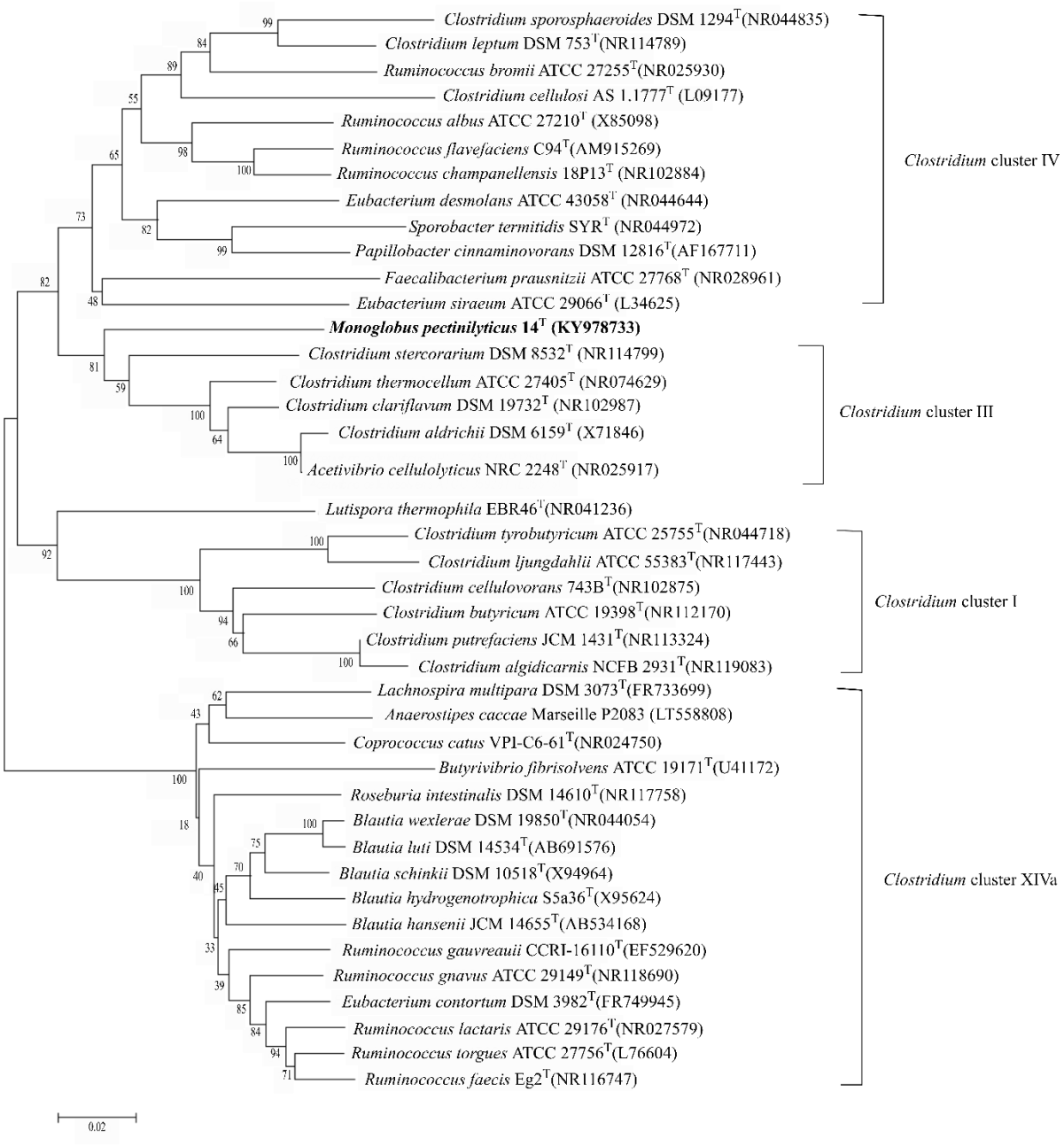
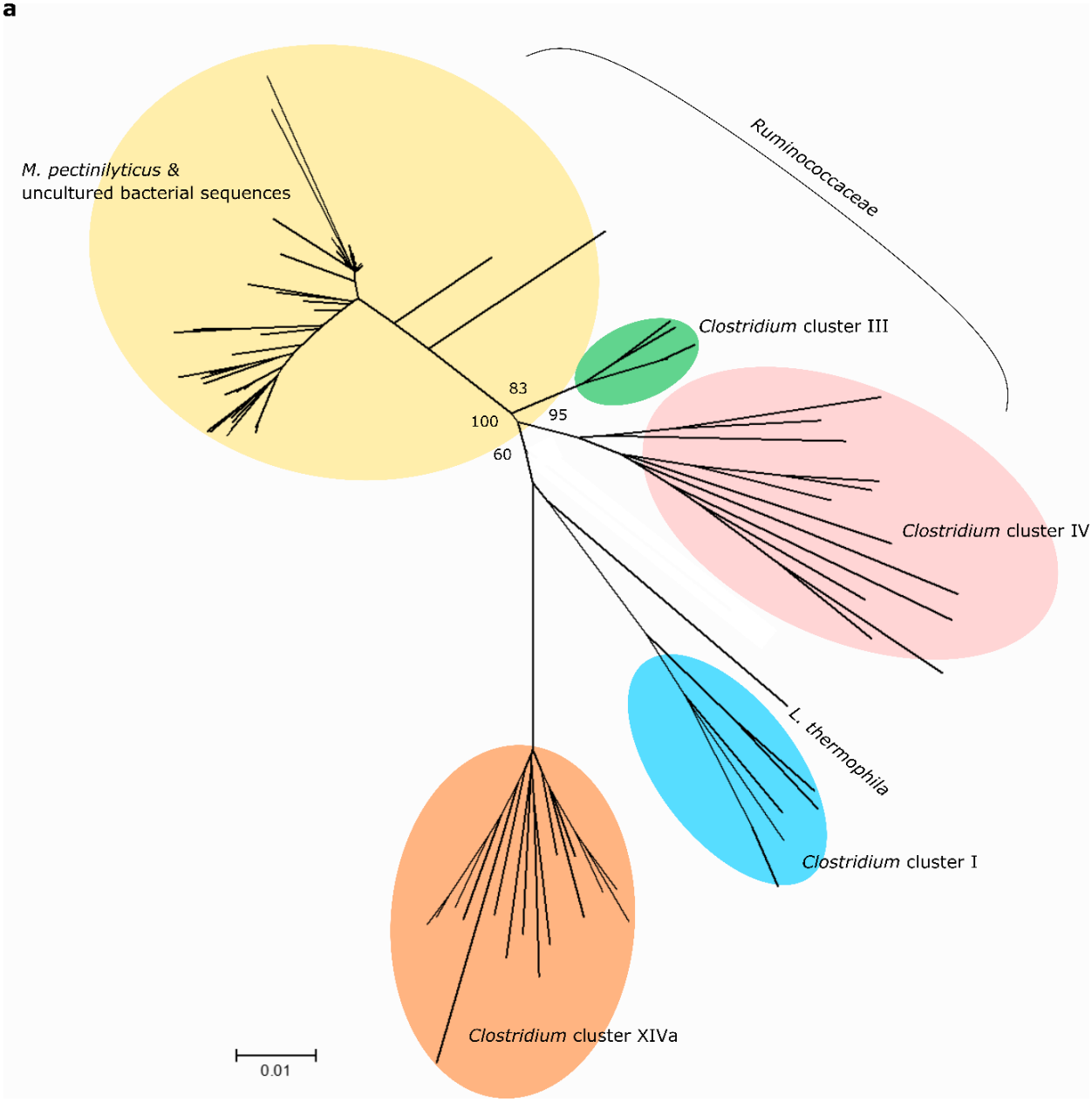


Figure 3.2 Neighbor-joining phylogenetic tree constructed based on 16S rRNA gene sequences to infer the phylogenetic relationship of strain 14^T to members of *Clostridium* clusters I, III, IV and XIVa. Bootstrap test (2,000 replicates) results are shown as percentage values next to the branches.

sequences of *M. pectinilyticus* and its genus relatives from human faeces formed a dominant cluster, with 29 sequences passing the ≥ 95% genus threshold, and 27 sequences passing the ≥ 97 % species threshold. Several non-human sequences (GQ175428, GQ175455, EU458641, and AB821761) were identified as *Monoglobus* spp., suggesting while *M. pectinilyticus* is strongly conserved in the human colon, closely related organisms are present in other terrestrial gut systems. Regardless of the isolation sources, strain 14^T and all sequences from uncultured bacteria formed a distinct polyphylectic cluster separated from *Clostridium* clusters III and IV (Figure 3.3a). The phylogenetic divergence of strain 14^T from the rest of the family *Ruminococcaceae* was supported by 100 % bootstrap value with 2,000 re-sampling. These results suggested that a previously uncharacterized cluster exists within the family *Ruminococcaceae*, and strain 14^T represented the first cultured member of a novel phylogenetic lineage of this taxon. The members of this lineage appeared to be present in various terrestrial gut systems of humans and animals (Figure 3.3b).


a
M. pectinilyticus &
uncultured bacterial sequences
Ruminococcaceae
Clostridium cluster III
83
100
95
60
Clostridium cluster IV
L. thermophila
Clostridium cluster I
Clostridium cluster XIVa
0.01

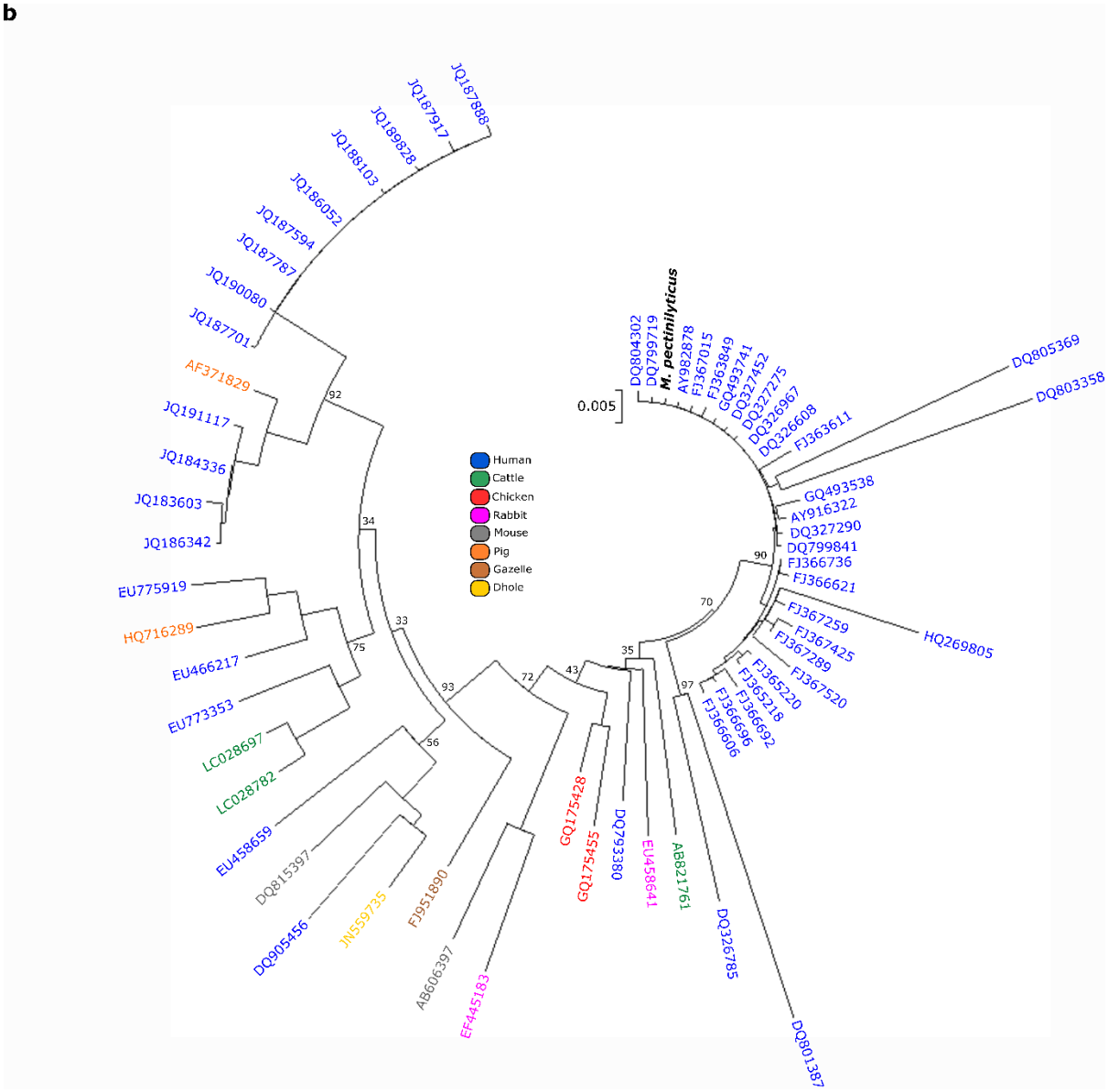


Figure 3.3 (a) The evolutionary history of 105 sequences was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016). (b) A neighbour-joining phylogenetic tree of near-full length 16S rRNA gene sequences from *M. pectinilyticus* and its uncultured GenBank relatives. The sources of sequences are indicated according to the colour table. Bootstrap values were calculated with 2,000 re-sampling.

3.6 Optimum growth condition

Strain 14^T was strictly anaerobic and exposure to aerobic or microaerophilic conditions prevented its growth. A temperature range of 30℃ - 40℃ and the acidity range of pH 6.5 - 8.0 provided optimal growth conditions (Figures 3.4 & 3.5). Even under optimal conditions, strain 14^T had a doubling time of 40 h and a log phase of 72 h until reaching stationary phase. Maximum optical density (OD595) ranged from 0.2 – 0.3. The addition of rumen fluid and vitamin solution was required to maintain growth during multiple successive transfers.

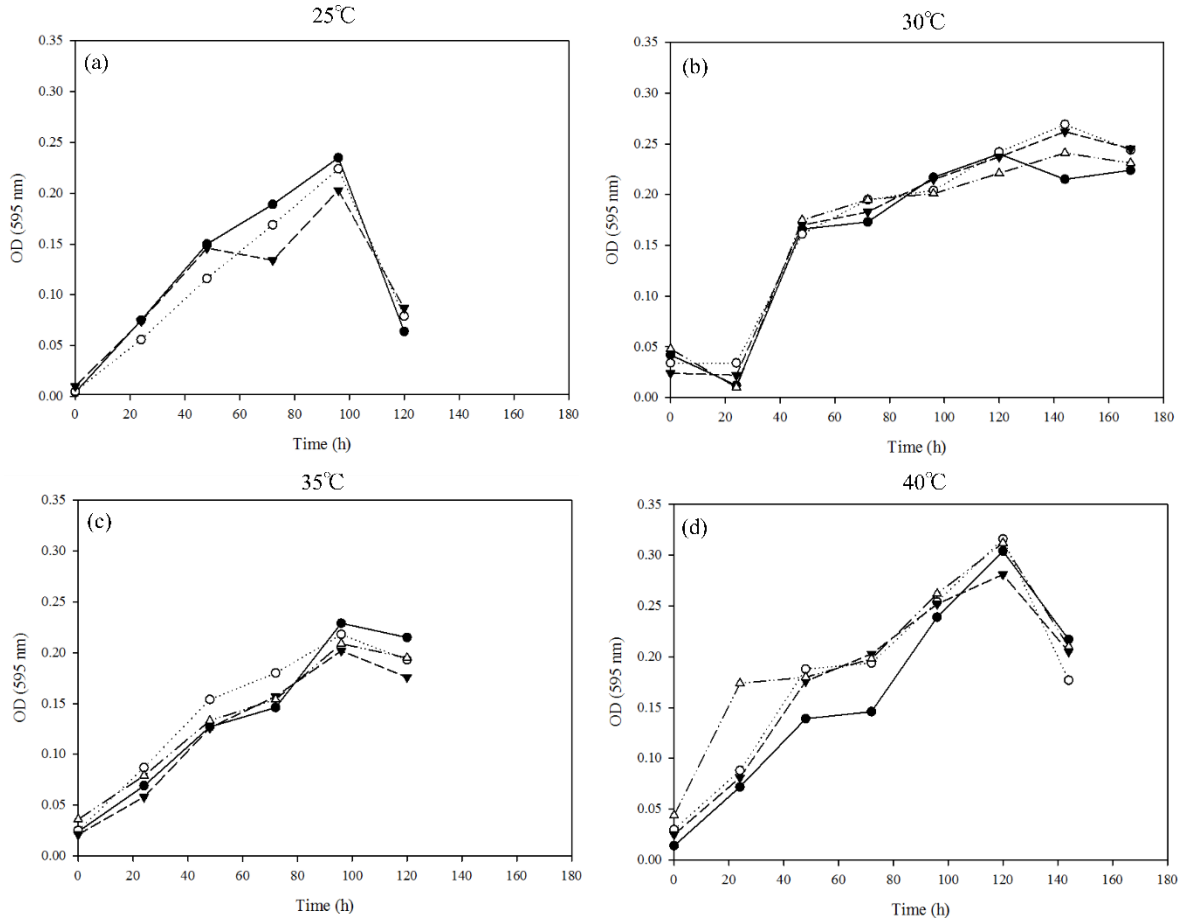


Figure 3.4 Growth curves of strain 14^T growing at 25℃ (a), 30℃ (b), 35℃ (c), and 40℃ (d) when the acidity of the medium was adjusted to pH 6.8. Optical density was measured at 595 nm. Each assay was performed using triplicate of cultures.

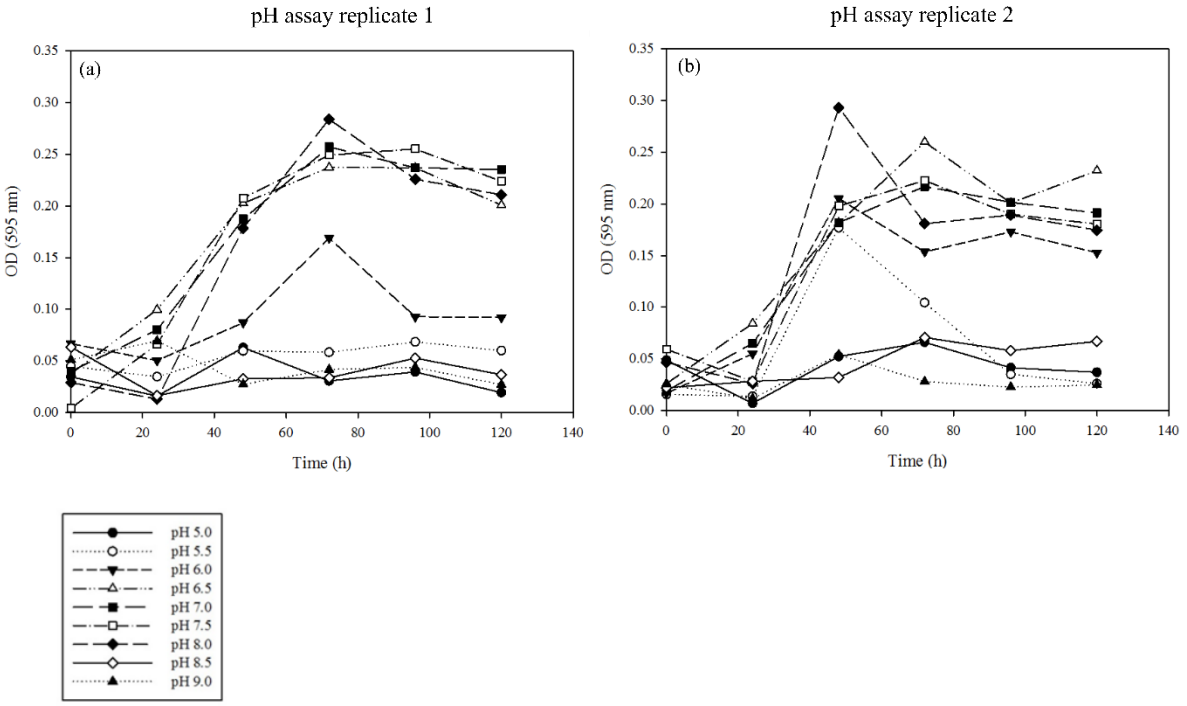


Figure 3.5 Growth curves of strain 14^T over an acidity range of pH 5.0 – pH 9.0. Incubation was carried out at 37℃. Assays were performed in duplicate (a) and (b) using triplicate of cultures. Optical density was measured at 595 nm.

3.7 Substrate utilization and organic acid production

Growth was monitored visually using a microscope. The cultures were considered viable if sufficient number of cells were present under the microscopic vision. Strain 14^T was able to grow on D-galacturonic acid, D-xylose, L-arabinose, D-fructose, and pectin from citrus peels, apple, and kiwifruit. With the exception of D-fructose, strain 14^T grew on a narrow spectrum of substrates which were either pectin or monosaccharides of pectin and hemicellulose (arabinose, xylose and galacturonic acid). Pectin-associated polysaccharides such as galactan and arabinan were unable to support the growth of strain 14^T. The organic acid production from galacturonic acid, arabinose, xylose, fructose, glucose, and pectin from citrus peels, apple, and kiwifruit was measured using GC (Figure 3.6 – 3.7). Acetate and formate formed as the major fermentation end-products of pectin, galacturonic acid, xylose, arabinose, and fructose. Strain 14^T was also able to produce acetate and gas from D-glucose, but very low cell number was observed under the microscope, and a negligible change in OD600 occurred throughout the growth. The amount of acetate and formate productions

was higher in pectin-grown cultures compared to the cells grown using monomeric carbohydrates. Acetate production from pectins reached maximum concentration ranges of 1.5 – 1.9 µmol/ml, compared to 0.3 – 0.6 µmol/ml in the presence of monomeric sugars. Formate production peaked at a range of 0.4 – 0.7 µmol/ml for pectin-grown cultures, but remained below 0.1 µmol/ml for the cultures grown on galacturonic acid, xylose, and arabinose. Fructose was the only monomeric substrate to achieve a formate concentration above 0.3 µmol/ml. Lactate production remained below 0.2 µmol/ml for all substrates tested including pectin, and declined below the threshold of detection limit over time. The phenotypic characteristics between strain 14^T and the related strains were compared in Table 3.3.

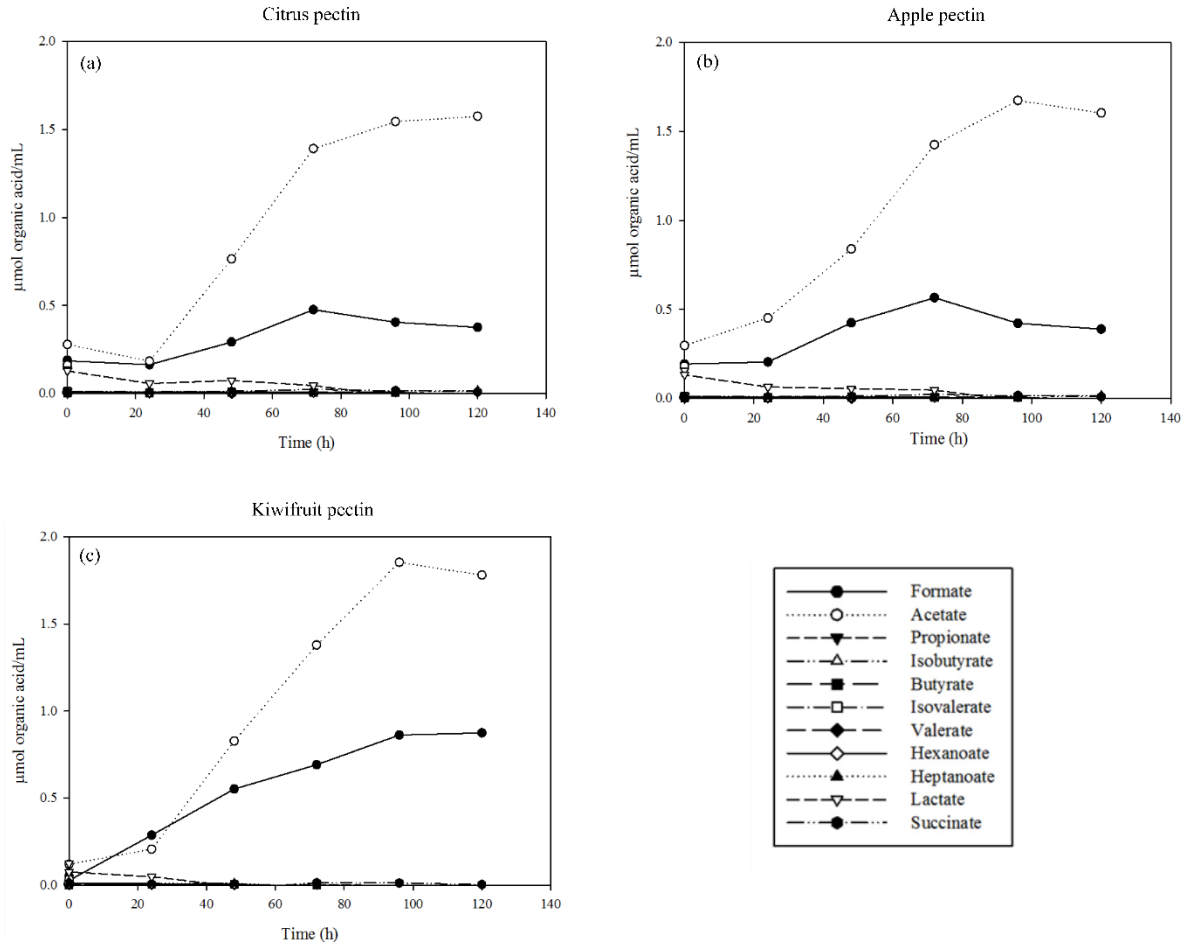


Figure 3.6 Organic acid production from pectin from citrus peels (a), apple (b), and kiwifruit (c). Each line shows an average value of triplicate samples.

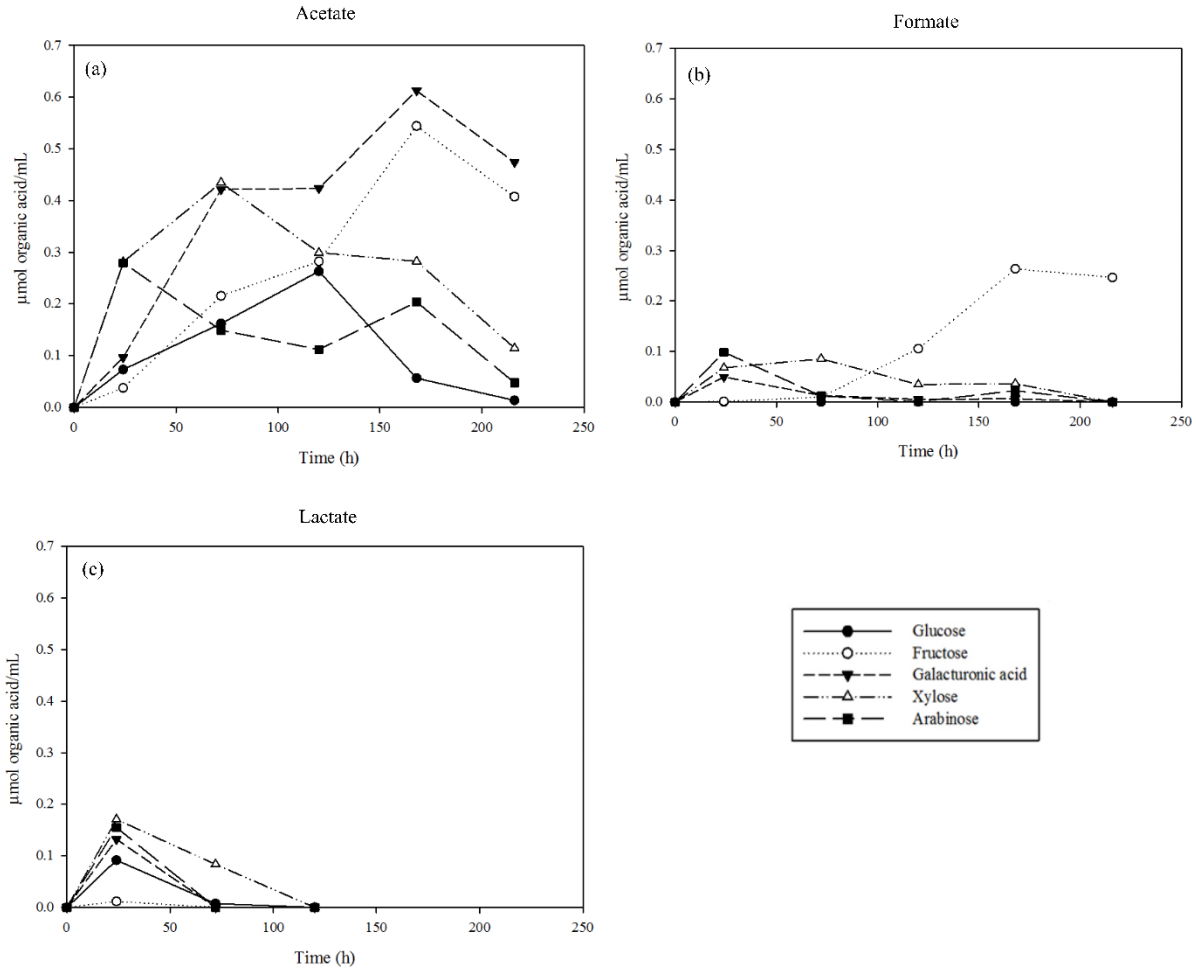


Figure 3.7 Acetate (a), formate (b), and lactate (c) production from D-glucose, D-fructose, D-galacturonic acid, D-xylose, and L-arabinose. Each line shows an average value of triplicate samples.

3.8 Gas production

CO_2 and H_2 were detected from the headspace of fully grown cultures of strain 14^T.

3.9 Depolymerization of pectin measured by SEC

SEC of 0 h, 48 h, and 72 h fermented citrus and apple pectin showed three distinct peaks eluting at 19 ml, 22 ml, and 27 ml, respectively (Figure 3.8 (a), (b)). Comparison with the pullulan molecular weight standards showed that these peaks are equivalent to molecular weights of 380 kDa, 48 kDa, and 6 kDa, respectively. These results suggested that the molecular weight of pectin decreased over the time of incubation. Similar degradation profiles were observed for both apple and citrus pectin. SEC using Superdex columns was carried out to monitor the accumulation of small oligosaccharides

Table 3.3 Differential phenotypic characteristics between strain 14^T and related species of *Clostridium* clusters III and IV. Strains: 1, Strain 14^T; 2, *Clostridium clariflavum* DSM 19732^T (data from Shiratori *et al*., 2009); 3, *Acetivibrio cellulolyticus* NRC 2248^T (Patel *et al*., 1980); 4, *Clostridium thermocellum* ATCC 27405^T (Shiratori *et al*., 2009); 5, *Clostridium stercorarium* subsp. *stercorarium* (Madden, 1983; Shiratori *et al*., 2009); 6, *Clostridium aldrichii* DSM 6159^T (Yang *et al*., 1990); 7, *Ruminococcus bromii* ATCC 27255^T (Moore *et al*., 1972); 8, *Ruminococcus champanellensis* 18P13^T (Chassard *et al*., 2012). Characteristics are scored as: +, positive; -, negative.

Characteristics	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Isolation source	Human faeces	Sludge	Sludge	Widespread	Compost	Wood fermenter	Human faeces	Human faeces
Morphology	Cocci	Rod	Rod	Rod	Rod	Rod	Cocci	Cocci
Gram strain	+	Variable	+	+	–	+	+	+
Cell size (μm)								
Length/Diameter	0.6 – 0.7	2.0 – 5.0	4.0 – 10	2.5 – 5.0	2.0 – 4.0	3.0 – 5.0	0.7 – 1.1	0.9 – 1.3
Width	-	0.4 – 0.5	0.5 – 0.8	0.5 – 0.7	0.3 – 0.4	0.5 – 1	-	-
Optimum temperature (℃)	30 – 40	55 – 60	35	60 – 64	65	35	37	33 – 39
Growth using								
Arabinose	+	-	-	-	+	-	-	-
Cellobiose	-	+	+	+	+	+	-	+
Cellulose	-	+	+	+	+	+	-	+
Fructose	+	-	-	-	+	-	+	-
Galactose	-	-	-	-	+	-	+	-
Galacturonic acid	+	-	-	-	-	-	-	-
Glucose	-	-	-	-	+	-	-	-
Glycogen	-	-	-	-	+	-	-	-
Lactose	-	-	-	-	+	-	-	-
Maltose	-	-	-	-	+	-	+	-
Mannitol	-	-	-	-	-	-	-	-
Mannose	-	-	-	-	+	-	-	-
Pectin	+	-	-	-	-	-	-	-
Salicin	-	-	+	-	-	-	-	-
Starch	-	-	-	-	+	-	+	-
Sucrose	-	-	-	-	+	-	-	-
Xylan	-	-	-	-	-	+	-	+
Xylose	+	-	-	-	+	-	-	-
Fermentation products†	A,F,L	A,F,L,E	A,P,B	A,L,E	A,L,E	A,L,P,B,S	A,F,L,P	A,S
DNA G+C content (mol %)	37.2	36.9	35.5	39	39	40	41.4	53.1

†A, acetate; F, formate; L, lactate; E, ethanol; P, propionate; B, butyrate; S, succinate.

over time as the result of pectin degradation (Figure 3.8 (c), (d)). For 0 h samples, a single peak was eluted at 30 min. The peak height decreased steadily over 48 h and 72 h incubation period, followed by the appearance of smaller peaks eluted at 54, 56, 60, and 64 min. These peaks gradually increased in heights after 48 h and 72 h incubation, indicating a gradual accumulation of oligosaccharides over time. The SEC analysis results showed that strain 14^T degrades high molecular weight pectin materials, leading to the accumulation of low molecular weight degraded products.

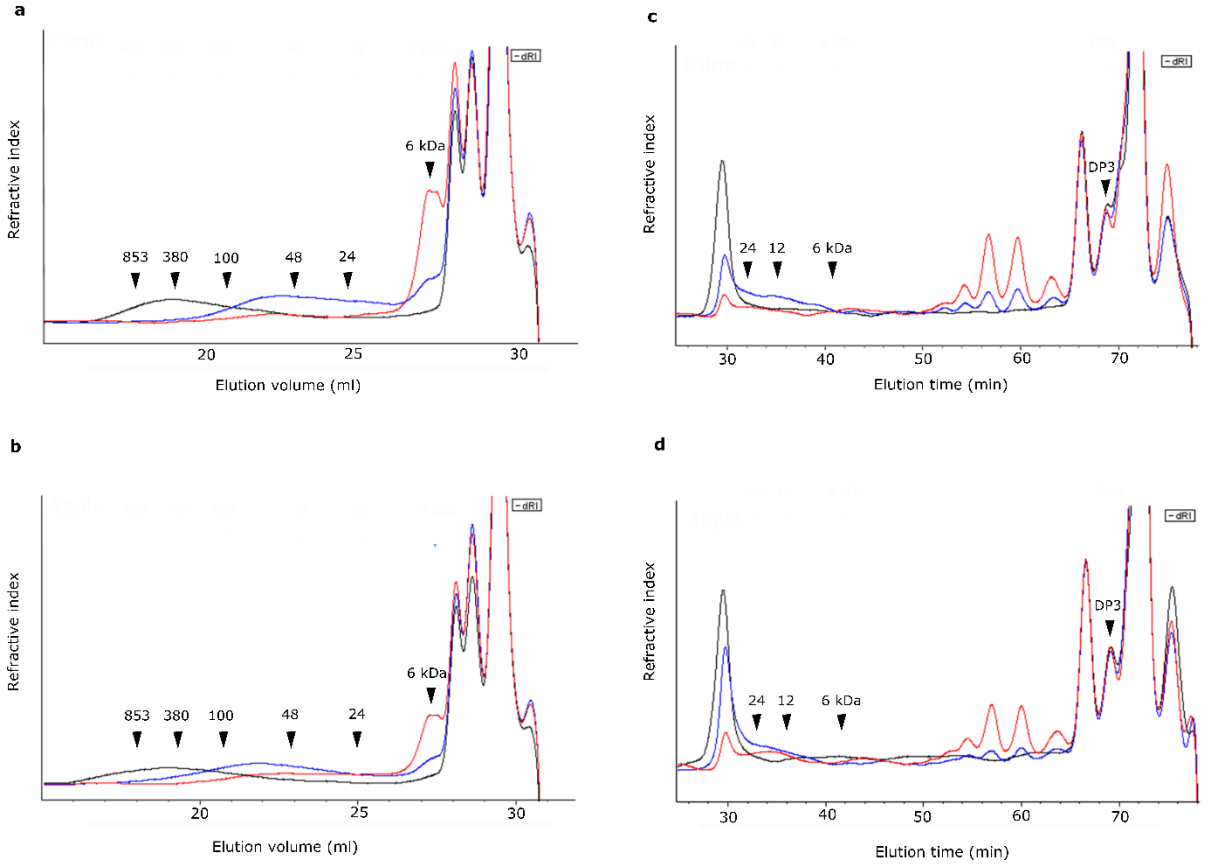


Figure 3.8 (a) & (b) SEC on TSK 5000, 4000 and 3000 columns. Citrus (top) and apple (bottom) pectins were fermented for 0 (black), 48 (blue) and 72 (red) hours. Arrows indicate the elution volumes of pullulan molecular weight standards. (c) & (d) SEC on two Superdex Peptide columns. Citrus (top) and apple (bottom) pectins were fermented for 0 (black), 48 (blue) and 72 (red) hours. Arrows indicate the elution volumes of pullulan molecular weight standards and trisaccharide (DP3, raffinose).

3.10 G+C content and cellular fatty acid

The DNA G+C content of strain 14^T was 37.21%, as determined from a sequenced genome. The major fatty acids of strain 14^T were $C_{18:1}\omega 9c$ (21.72 %), $C_{16:0}$ (19.16 %), ante-$C_{18:0}$ (12.02 %), iso-$C_{15:0}$ (11.80 %), anteiso-$C_{15:0}$ (11.80 %), anteiso-$C_{17:0}$ (8.48 %), and $C_{18:0}$ (5.26 %).

3.11 Biochemical tests

Strain 14^T tested positive for catalase, and negative for casein hydrolysis, esculin hydrolysis, gelatinase, indole production, lipase/lecithinase, cytochrome oxidase, and urease.

3.12 Curation and publication of strain 14^T

Strain 14^T was given a taxonomic name *Monoglobus pectinilyticus* gen. nov., sp. nov., type strain 14^T (Mo.no.glo'bus. Gr. adj. *monos*, single, solitary; L. masc. n. *globus*, a ball, sphere; N.L. masc. n. *Monoglobus* single sphere; pec.ti.ni.ly'ti.cus. N.L. neut. n. *pectinum* pectin; N.L. adj. *lyticus* (from Gr. adj. *lytikos*) able to loosen, able to dissolve; N.L. masc. adj. *pectinilyticus*, pectin-dissolving). A type strain (designated by superscripted 'T') is defined in the "International Code of Nomenclature of Bacteria" as the "nomenclatural type of the species", and is the "reference point" to which all other strains are compared to know whether they belong to that species. JCM and DSM accession numbers (strain 14^T=JCM 31914^T= DSM 104782^T) were issued from the respective public culture collections as evidence of culture deposition. The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of *M. pectinilyticus* 14^T is KY978733. The results from this chapter were published in International Journal of Systematic and Evolutionary Microbiology, under the title of "*Monoglobus pectinilyticus* gen. nov., sp. nov., a pectinolytic bacterium isolated from human faeces" (DOI: 10.1099/ijsem.0.002395).

Chapter 4 Genomic overview of *Monoglobus pectinilyticus* strain 14^{T}

4.1 Introduction

This chapter describes the *de novo* assembly and annotation of a near-complete genome of *Monoglobus pectinilyticus* 14^{T}. *M. pectinilyticus* was selected for further genomic characterizations based on its phylogenetic novelty, the strong pectin-degrading traits, and the potential presence of extracellular pectin-adhesion machineries. The assembled genome was analysed to provide a genomic overview of the bacterium with a strong emphasis on its glycobiome and metabolic capabilities. Due to the taxonomic closeness of *M. pectinilyticus* to the cellulosome-producing organisms, it was initially speculated that the pectinolytic ability of *M. pectinilyticus* may have resulted from the presence of a cellulosome-like system ('pectinosome'). However, a more detailed analysis of its genome revealed that *M. pectinilyticus* not only produces a highly pectin-specialized repertoire of CAZymes, but also possesses a carbohydrate degradation system consisting of novel protein constituents which failed to find significant sequence homology and functional matches using the most up-to-date nucleotide and protein sequence databases. These novel features may represent a possibility outside the two major paradigms of the microbial polysaccharide degradation (e.g. the cellulosome system and the Sus system) described from the human gut, and a system highly relevant to the pectin-rich diet of humans. The information obtained here will improve our understanding of the process of microbial pectin degradation in the human gut, and provide useful reference data for future genomics and metagenomics work.

4.2 *de novo* genome assembly and primer walking

A total of 11 Gb sequencing reads were generated using the Illumina HiSeq2500 platform. Raw sequences were quality-filtered and assembled into a draft genome. Sequencing reads were assembled to produce a draft genome consisting of three large contigs (782,032 bp, 959,713 bp, and 1,007,898 bp in sizes). Because the *in silico* method of closing the genome gaps using GapFiller (Boetzer and Pirovano, 2012) was not successful, a series of primer waking procedures were carried out to manually sequence the genome gap regions. Using the combination of long-range PCR and cloning

techniques, the gap sequences were amplified and the sequence sizes were estimated on agarose gel (Figure 4.1). The approximate sizes of Gap 1, Gap 2, and Gap 3 were estimated to be 4 kb, 2 kb, and 5 kb, respectively. A series of primer walking procedures was carried out to manually close the sequencing gaps (Figure 4.2). Subsequent genome annotation revealed that Gap 1 and 3 had occurred within the two copies of the ribosomal operon regions. Gap 1 became fully closed with 3,489 bp added between contig 1 and contig 2. Gap 3 was also fully sequenced by extending 4,546 bp from

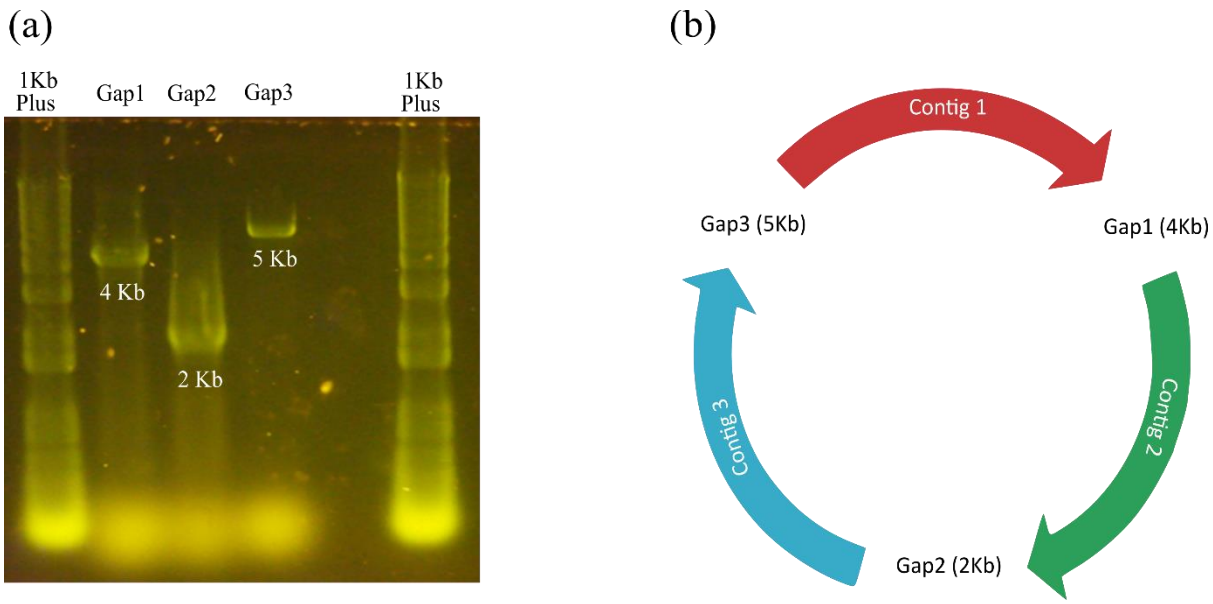
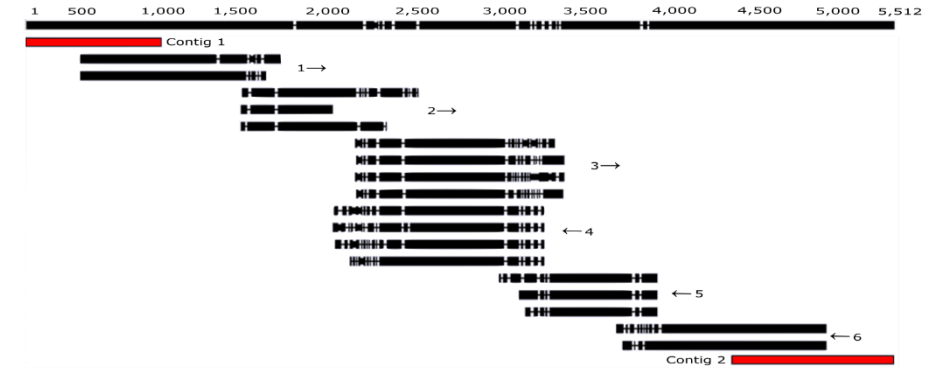
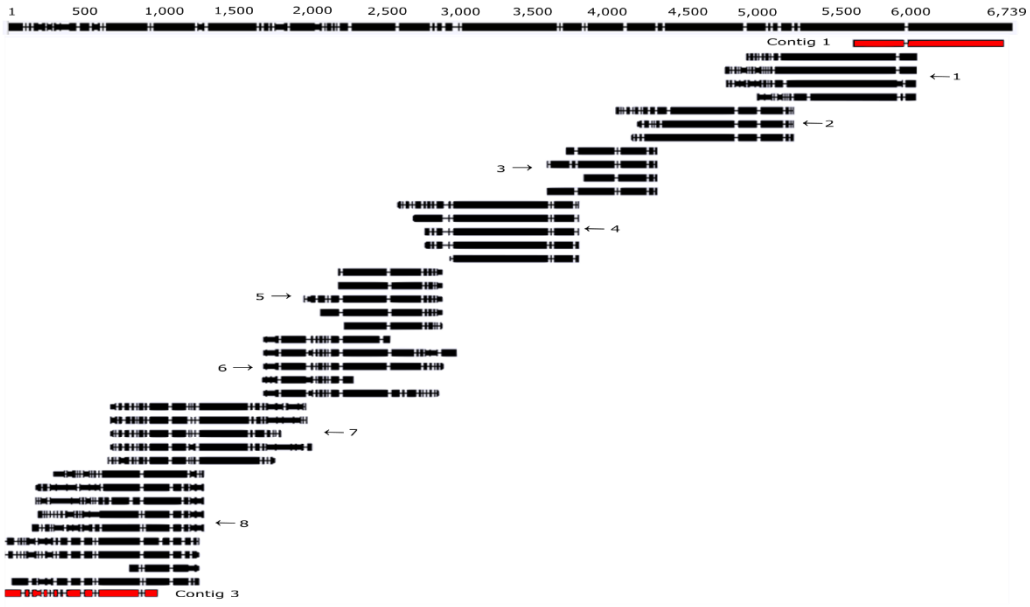


Figure 4.1 (a) High-fidelity PCR-amplified sequences from genome gaps electrophoresed on 1 % agarose gel stained with SYBR Green I Nucleic Acid Gel Stain; (b) A graphical visualization of genome contigs and sequencing gaps.

(a)



(b)



	Primer #	Sequence (5'→3')
Gap 1	1	GCATGCTGCGGTGAATACGTTCCCGGG
	2	GCGAAGAAGAGCACATGGAGAATGCCTTGG
	3	CGACCCGAAACCGAGTGACCTACCC
	4	CTTACCTGACAAGGAATTTCGC
	5	CCCTTCAAATTTCCTGCGCCCGCGAC
	6	CCACCTCACCACTGAATAAATGGCTTAGGC
Gap 3	1	CAGAGACCGATTTGTGGAACTGGGATTAATGATTTC
	2	GACGTAGTATATGATACAGGTC
	3	CCTAACGCGGCAAACTGCG
	4	CCGGGAGGGGAGTGAAAGAGAACC
	5	GGTCCCAAAATAGGTTTAAGTGG
	6	GTTCTGACCCGCACGAATGGCG
	7	GCGCCCTACCGATGTTCGC
	8	GCATAAATGCGCCCTACCGATGTTCGC

Figure 4.2 Primer walking procedures for closing the sequencing Gap 1 (a) and Gap 3 (b). Arrows indicate forward and reverse sequencing orientations. The contig ends from the assembled genome are indicated in red. The sequence information for each numbered primer is listed in table.

the 5'-end of contig 1, until the newly sequenced region overlapped with the 3'-end of contig 3. For reasons unclear, the PCR-amplified region of Gap 2 failed to become sequenced using either the Illumina HiSeq2500 technology or the traditional Sanger sequencing approach using the ABI3730 capillary instrumentation (Applied Biosystems). Despite the numerous sequencing attempts and method optimizations by the sequencing service provider (Massey Genome Service), a premature truncation of the sequencing reads, weak signalling intensity, and overlapping peaks rendered an unbiased base calling difficult. Considering that the 3'-end of contig 2 and the 5'-end of contig 3 both harboured phage-related genes, it was speculated that Gap 2 may consist of a 2 – 3 kb DNA segment from a genetically unstable region containing the elements of a prophage. The newly attained gap sequences were aligned with the existing contigs to produce a single continuous contig of 2,757,678 bp, suggesting the full-size genome of *M. pectinilyticus* was estimated to be ~ 2.76 Mb.

4.3 Assessing the genome completeness

Because the unfilled gap region was estimated to be small (2 kb), and transfer RNAs for all 20 amino acids were identified, the assembled genome of *M. pectinilyticus* appeared to have reached an almost-complete draft genome stage. To be sure, the completeness and cleanness of the draft genome was assessed using CheckM software (Parks *et al*., 2015). CheckM assesses the completeness of a bacterial genome by the presence and absence of universally present marker genes. Marker genes account for ~10 % of all genes, and they are seldom distributed in an even manner within the genome (Sharon and Banfield, 2013). Using the family *Ruminococcaceae* as the marker lineage, the draft genome of *M. pectinilyticus* was diagnosed as 95.24 % complete, with 1.37 % contamination, and 0 % strain heterogeneity. These values fell within the HMP-recommended levels of finishing for gold standard genome quality, as reported from the self-assessment of the CheckM software which found that 96 % of the bacterial genomes in public database were estimated to be $\geq$ 95 % complete, with $\leq$ 5 % contamination, and were considered as superior quality genomes suitable for genomic and metagenomic analysis (Parks *et al*., 2015).

4.4 Genome-based phylogenetic analysis

In prokaryotic taxonomic classification, 16S rRNA gene sequence similarity and the level of DNA-DNA hybridization (DDH) play crucial roles in determining whether a novel strain should be placed within an existing genus or be identified as the first species representing a novel genus (Qin *et al*., 2014). With the increasing availability of high quality genome sequences, computational techniques to integrate the whole-genome data into *in silico* prokaryotic taxonomic analysis are becoming common (Meier-Kolthoff *et al*., 2013). Although the need for experimental validation of computational results is unlikely to become completely eliminated, the volume and depth of sequencing data provide highly reproducible means to predict the prokaryote phylogeny through genome-to-genome comparison (Meier-Kolthoff *et al*., 2013). The draft genome of *M. pectinilyticus* was aligned against the complete genomes of *C. thermocellum* ATCC 27405 (GenBank accession CP000568.1), *C. clariflavum* DSM 19732 (GenBank accession CP003065.1), *R. albus* DSM 20455 (GenBank accession CP002403.1), and *R. champanellensis* JCM 17042 (GenBank accession FP929052.1) using Genome-to-Genome Distance Calculator which performs an *in silico* genome-wide DNA-DNA hybridization by producing a set of high-scoring segment pairs (HSPs) also known as intergenomic matches (Meier-Kolthoff *et al*., 2013). The reference organisms scored significantly below the recommended > 70 % hybridization threshold for reliable species/sub-species delineation (*C. thermocellum* 26.8 %; *C. clariflavum* 27.5 %; *R. albus* 25.5 %; and *R. champanellensis* 24.1 %). Qin *et al*. (2014) has recently proposed the use of the percentage of conserved proteins (POCP) as a novel criterion for estimating the phylogenetic distance between two strains. Using POCP of 50 % as a parameter for defining a genus boundary, Qin detected a high degree of genus-level diversity within the organisms classified as the genus *Clostridium*, thus echoing the concern expressed in *Bergey's manual of systematics of archaea and bacteria* that the clostridial lineage of *Firmicutes* requires an extensive restructuring and renaming (Rainey *et al*., 2015). The amino acid sequences of all CDSs were extracted from the genome of *M. pectinilyticus*, and aligned in a pair-wise manner against the complete genome sequences of *C. thermocellum* ATCC 27405, *C. clariflavum* DSM 19732, *C. cellulolyticum* H10 (GenBank accession CP001348.1), *C. stercorarium* DSM 8532 (GenBank accession CP004044.1), *C. cellulosi* DG5 (GenBank accession LM995447.1), and *R. albus* DSM

20455 using a locally installed BLASTP program to find the best match per query protein as described in Qin *et al*. (2014). This process was continued until all organisms were used as queries at least once against another to create a POCP matrix (Figure 4.3). Proteins from the query genome were considered as conserved if the predicted bit-score (or max score) was > 200. In accordance with Qin's results, *C. thermocellum* and *C. clariflavum,* which shared 96 % 16S rRNA sequence similarity with each other, showed > 50 % POCP, suggesting the two strains should be classified into a single genus. *C. thermocellum, C. clariflavum, C. cellulolyticum* and *C. stercorarium* belong to the *Clostridium* cluster III which has been reported to be a multi-genus group of organisms with intracluster 16S rRNA sequence similarities ranging 87 – 99 % (Collins *et al*., 1994). While all four species are known

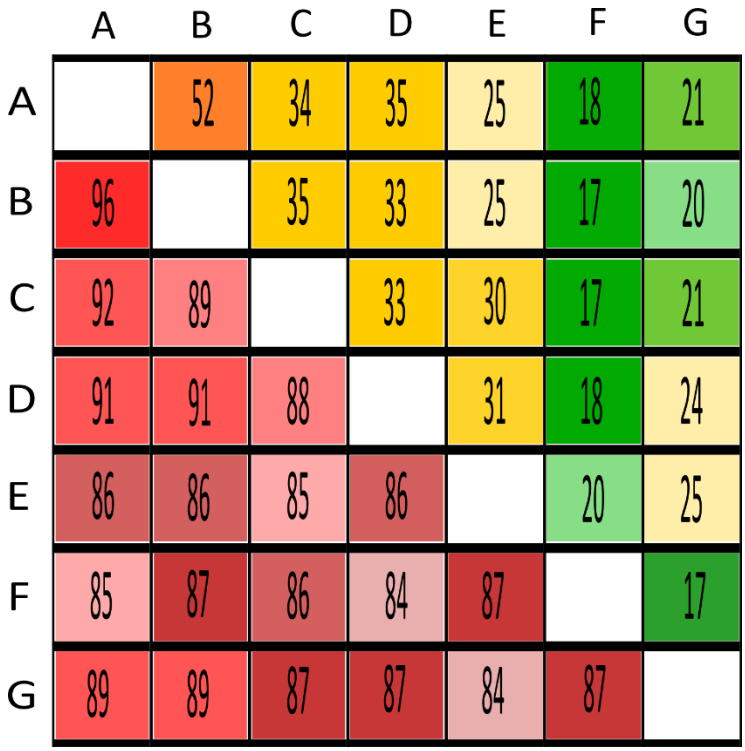


Figure 4.3 Matrix diagram showing the POCP of bacterial genomes and the 16S rRNA gene sequence similarities. On the upper right corner of the matrix, a pair-wise comparison of POCPs between (A) *C. thermocellum*; (B) *C. clariflavum*; (C) *C. cellulolyticum*; (D) *C. stercorarium*; (E) *C. cellulosi*; (F) *R. albus*; and (G) *M. pectinilyticus* is shown. The bottom left corner of the matrix displays the relative 16S rRNA gene sequence similarities between the strains mentioned above. All numbers are expressed in percentage. The colour gradient is an indication of the relative percentage when 100 % was used as the highest level of intensity.

to be cellulolytic, only *C. thermocellum*, *C. clariflavum,* and *C. cellulolyticum* possess the multi-enzyme cellulosome complex, whereas no conclusive genetic evidence of cellulosome was found in *C. stercorarium* (Desvaux *et al*., 2005). Nevertheless, a substantial amount of core proteins (33 – 35 %) appeared conserved outside cellulosomal elements, and the presence/absence of cellulosome seemed to have little influence over dictating the phylogeny of these strains. Another cellulosome-producing strain, *R. albus* significantly deviated from the *Clostridium* cluster III and *M. pectinilyticus*, sharing only 17 – 18 % of its core proteins with other strains. Interestingly, *C. cellulosi* shared 30 – 31 % of its core proteins with *C. cellulolyticum* and *C. stercorarium* compared to 20% with its phylogenetic neighbour *R. albus* with which it constitutes the *Clostridium* cluster IV, suggesting the core proteins of *Ruminococcus* spp. may be strongly conserved within its genus. Despite the phenotypic likeness to *Ruminococcus* spp. which were previously the only known Gram positive coccus-type bacteria of the family *Ruminococcaceae*, *M. pectinilyticus* shared only 17 % of its protein repertoire with *R. albus*, thus suggesting that *M. pectinilyticus* has evolved independently from *Ruminococcus* spp. Overall, the results of POCP analysis were consistent with the phylogenetic relationships inferred based on the 16S rRNA gene sequence similarities. *M. pectinilyticus* showed a higher POCP value (24 %) with *C. stercorarium* compared to *C. thermocellum* and *C. clariflavum* (21 % and 20 %, respectively), reflecting the divergence of *M. pectinilyticus* from the *Clostridium* cluster III which likely occurred before the event of *C. stercorarium* diverging from *C. thermocellum* and *C. clariflavum.* Based on POCP analysis, the number of conserved proteins between *M. pectinilyticus* and the reference species of the *Clostridium* cluster III and IV were significantly below the proposed 50 % genus boundary for prokaryotic lineage, consolidating the proposed classification of *M. pectinilyticus* into a novel genus/species.

4.5 Genome organization and general features

The genome of *M. pectinilyticus* consists of a single circular chromosome of an estimated size of 2.76 Mb. The general features of *M. pectinilyticus* genome are summarized in Table 4.1. Genome assemblies, raw sequencing data, and annotation were deposited at GenBank under the accession number CP020991, and BioProject number PRJNA383867. Locus tag prefix ‘B9O19’ was assigned.

Table 4.1 Key genomic properties of *M. pectinilyticus*

General genome properties	
GenBank accession number	CP020991
Locus tag prefix	B9O19
Genome size (bp)	2,757,678
Genome state	Draft
Number of contigs	1
G+C content (%)	37.21
Number of genes	2,310
Number of CDS	2,263
Coding percentage	98.01
Genes assigned with COGs	1,515
Genes with signal peptide sequence	227
Ribosomal RNAs	2
Transfer RNAs	46
Transfer-messenger RNAs	1
CRISPR regions	2
Number of prophages	4

The high percentage of the protein coding density (98.0 %) suggested that the genome of *M. pectinilyticus* was highly evolved, and followed the tendency of bacterial genomes to delete non-essential DNA to maintain compact genetic information (Ussery *et al*., 2009). Two regions of ribosomal RNA were located at positions 1,792,742 → 1,798,042 and 1,008,566 → 1,012,661 using RNAmmer 1.2 server (Lagesen *et al*. 2007). *M. pectinilyticus* possessed the highly conserved 16S-23S-5S arrangement of ribosomal genes within the rRNA operon. A total of 46 transfer RNA-encoding genes were identified, and all 20 essential amino acids have been accounted for. A visual representation of *M. pectinilyticus* genome which shows the guanine abundance (GC skew), G+C content, the frequency of CDS occurring in leading and lagging DNA strands, and COG (Cluster of orthologous groups) functional categories was constructed using a combination of CGViewer and BaSys annotation system (Figure 4.4a; Grant and Stothard 2008; Van Domselaar, 2005). GC skew refers to the position where the guanine-rich region splits from the guanine-poor region which usually indicates the origin of replication (Ori). Accordingly, chromosomal replication initiation protein (DnaA) was identified at the position 1,344,987 of the genome. *M. pectinilyticus* followed the general rule of bacteria to position essential protein-coding genes on the guanine-rich leading strand of DNA.

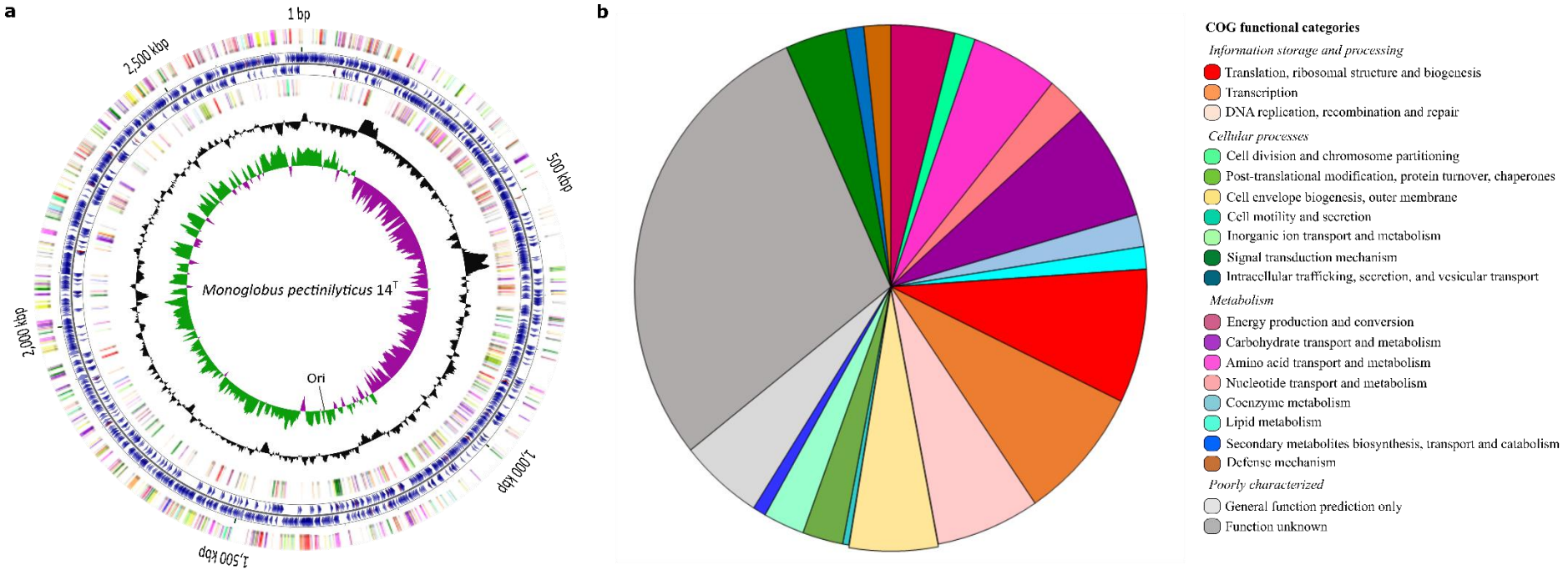


Figure 4.4 A genome map of *M. pectinilyticus* chromosome. (a) Genomic positions are indicated in kbp. From the outer most ring, ring 1 and 4 show the COG assignment of CDSs; ring 2 and 3 show the predicted CDSs on leading and lagging DNA strands, respectively; ring 5 shows the G+C content throughout the genome; and GC skew is shown in the innermost ring 6. Origin of replication ('Ori') is indicated. (b) Distribution of COG functional categories analysed using the eggNOG database. Colour codes for COG functional categories are shown on the right.

Approximately ~67 % of the genome showed a preference for guanine to be positioned on the leading strand. Also, ~72 % of CDSs were present on the leading strand of replication. In general, a strong selective pressure favours the majority of highly expressed protein-coding genes to be located on the leading strand of a chromosome, as the transcription from the lagging strand tends to be less efficient due to the possible steric collision between RNA and DNA polymerases moving in opposite directions (Mao *et al*., 2012). CDS frequency of *M. pectinilyticus* fell within the reported ranges for the phylum *Firmicutes* (70 – 80 %) (Wegmann *et al*., 2014; Chou *et al*., 2015; Suen *et al*., 2011; Mao *et al*., 2012). To ensure the accuracy of the COG functional prediction, eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database was used to construct an additional eggNOG profile for *M. pectinilyticus* using an HMMER-based pipeline (Jensen *et al*., 2008). The eggNOG database was built based on the original COGs, but extended with additional non-supervised orthologous groups (NOGs) of proteins which could not be assembled into conventional COG categories (Jensen *et al*., 2008). NOGs are not manually curated, but it uses a powerful computational analysis of lineage inferences to provide a high-accuracy function annotation covering a broader range of genes and genome databases outside COG (Jensen *et al*., 2008). Each result was manually inspected. Overall, the eggNOG profile of *M. pectinilyticus* showed the distribution of genes into appropriate functional categories (Figure 4.4b). Approximately 66 % of the total genes were assigned into the following eggNOG categories; energy production and conversion (4 %); cell cycle control, cell division, and chromosome partitioning (1 %); amino acid transport and metabolism (6 %); nucleotide transport and metabolism (2 %); carbohydrate transport and metabolism (7 %); coenzyme transport and metabolism (2 %); lipid transport and metabolism (1 %); translation, ribosomal structure, and biogenesis (8 %); transcription (8 %); replication, recombination, and repair (7 %); cell wall/membrane/envelope biogenesis (6 %); cell motility (0.2 %); post-translation modification, protein turnover, and chaperones (3 %); inorganic ion transport and metabolism (3 %); secondary metabolites biosynthesis, transport, and catabolism (1 %); signal transduction mechanisms (4 %); intracellular trafficking, secretion, and vesicular transport (1 %); and defence mechanisms (2 %). 34 % of genes were assigned into ‘function unknown’ or ‘general function prediction only’ categories which represent proteins with poorly characterized functions. A small degree of erroneous

function prediction was observed from the genes assigned into the cell motility category, as *M. pectinilyticus* is non-motile, and lacks the genes associated with flagella assembly. It appeared that eggNOG program mistakenly inferred a common ancestry relationship between the GH95 family of α-L-fucosidase and the eukaryotic fucosidases which are known for their involvement in cell-cell adhesion and cell migration of leukocytes (Ali *et al*., 2008). The GH95 family enzymes are often associated with cleaving the terminal L-fucose residues during the hemicellulose degradation in prokaryotes. The presence of N-terminal signal peptides was predicted using SignalP (Peterson *et al*., 2011). Approximately 10 % of CDSs were shown to contain type I signal peptides, suggesting that these proteins may be secreted.

4.6 Extraneous DNA

4.6.1 Prophage genomes

M. pectinilyticus harbours 4 prophage regions (I – IV) which appear to be either complete or remnants of degraded phage genomes. These included B9O19_454 – 500 (~ 35 kb), B9O19_634 – 650 (~16.6 kb), B9O19_1735 – 1795 (~43 kb), and B9O19_2287 – 2310 /1 – 35 (~42 kb). Figure 4.5 shows a schematic diagram of the gene organizations of prophages. The sizes of prophage I, III, and IV were above the $>$ 30 kb genome sizes observed from the majority of the dsDNA self-transmissible temperate phage genomes available in the current GenBank database (Touchon *et al*., 2016). The sequencing of *M. pectinilyticus* genome was disrupted within the prophage IV region. Despite the fragmented genome status, prophage IV appeared to maintain a relatively complete repertoire of bacteriophage-related enzymes including lysins, recombinase, terminase, resolvase, and structural phage proteins such as head, coat, tail, portal, and holin components. Therefore, prophage I, III, and IV may be *bona fide* prophages which can be induced into entering the lytic cycle under unfavourable conditions. On the other hand, prophage II appeared to be a partially degraded phage remnant which probably no longer retains the ability to trigger the lytic cycle. Prophage genome elements accounted for ~3.2 % of the total genomic content of *M. pectinilyticus*. This value was close to the reported average (3.1 %) of prophage elements constituting the lysogenic bacterial genomes, but the number of

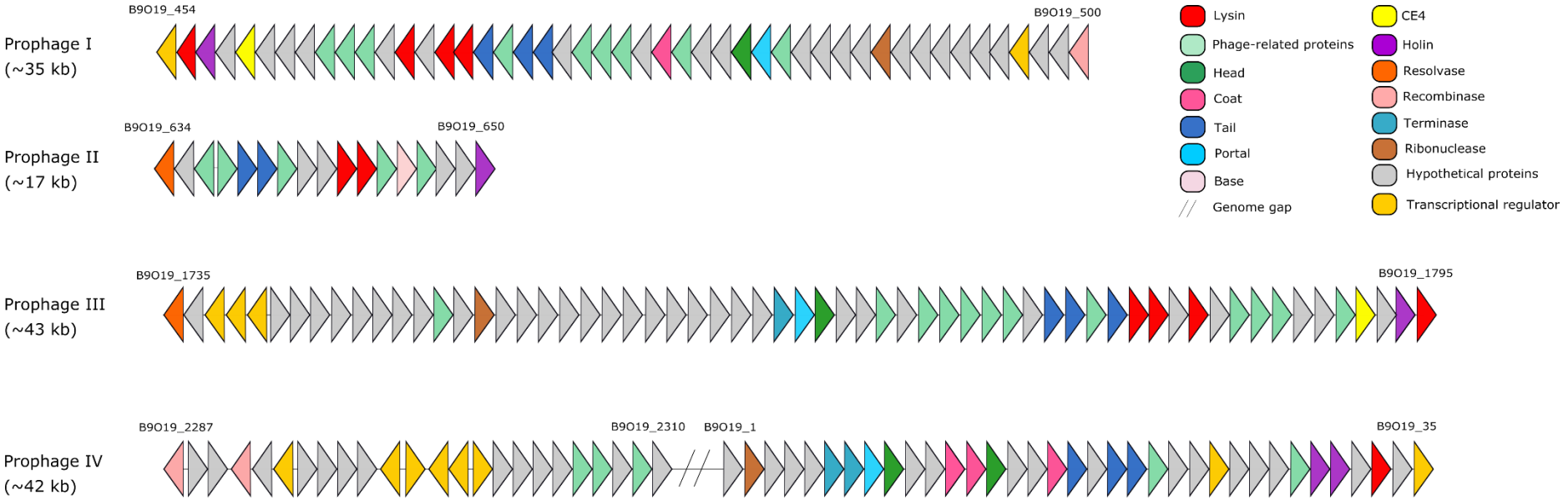


Figure 4.5 Composition of prophage-associated gene clusters present within the *M. pectinilyticus* genome. CDSs are coloured according to the product category. The direction of transcription is indicated by arrows.

prophages was greater than the average of ~1 prophage reported for bacteria with 2 – 3 Mb genome sizes (Touchon *et al*., 2016).

4.6.2 Horizontal gene transfer (HGT)

The presence of erratic oscillation of the GC skew and the regions of non-typical G+C content in a genome usually indicates the evidence of horizontal gene transfer (HGT), also known as lateral transfer, of foreign genetic materials associated with bacteria, phage, transposons, and plasmids. While *M. pectinilyticus* showed a low degree of GC fluctuations throughout its genome, two notably deviated GC peaks were observed at approximate genomic positions 155,765 – 206,581 (~51 kb) and 573,907 – 646,837 (~73 kb). Using Islandviewer 3 (Dhillon *et al*. 2015), the two large aberrant G+C peaks were predicted to overlap with two major HGT events, suggesting the non-typical G+C contents observed in these regions were likely to have resulted from the acquisition of foreign DNA. While some of the laterally transferred genes in the *M. pectinilyticus* genome were predicted to be involved in DNA integration and mobilization, the majority of genes were annotated as 'hypothetical', as there were no currently known functions for these genes. Within the ~51 kb HGT region, genes encoding plasmid partitioning proteins ParA and ParB (B9O19_141 – 142), transposable elements TnpY (B9O19_153) and TnpV (B9O19_173), plasmid conjugative transfer protein TraG (B9O19_160), DNA/RNA integrases (B9O19_174 and B9O19_188), putative plasmid-related excisionase family proteins (B9O19_189, B9O19_190, B9O19_192, and B9O19_193), and mutagenesis-inducing DNA polymerase IV (B9O19_182) were found, indicating this genome region may contain remains of an inserted plasmid. The ~73 kb HGT occurs immediately after the prophage I, suggesting the repeated acquisition of transmissible genetic elements may have rendered this genomic region unstable and susceptible to evolution. Genes encoding DNA mobilization proteins such as relaxases (B9O19_530, B9O19_531, B9O19_553, B9O19_572, B9O19_586), TraG (B9O19_589), and TrsE (B9O19_564), TnpV (B9O19_532 and B9O19_579), integrases (B9O19_521 and B9O19_536), and excisionase (B9O19_537) were present, indicating these genetic elements may also have a plasmid origin. The predicted HGT regions showed unusual G+C contents, suggesting that these genes may have been acquired recently, and undergoing the process of amalgamation into the

host genome. HGT elements accounted for approximately 4.6 % of the total genomic content of *M. pectinilyticus*.

4.7 CRISPR

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated genes are evidence of an adaptive microbial immune system to protect against foreign genetic elements such as bacteriophages, plasmids, and transposons (Bhaya *et al*., 2011). The 21 – 48 bp partially palindromic repeat sequences are flanked by spacer sequences which derive from the memory of previous exposures to non-self nucleic acids. Upon the detection of spacer sequences from a foreign object, spacers are transcribed to produce small non-coding RNA (crRNA) which interact with various Cas proteins such as nuclease, helicase, and RNA-binding proteins to act as a surveillance system to destroy or silence the invading DNA (Bhaya *et al*., 2011). Almost 47 % of bacterial genomes are known to possess at least one type of Cas system, and the occurrence of CRISPR was more frequently observed among bacterial genomes carrying prophages (Touchon *et al*., 2016). *M. pectinilyticus* possesses CRISPR type I-C and II-A systems in its genome (Figure 4.6). Cas3 and Cas9 are the signature proteins of Type I and Type II systems, respectively. Signature Cas proteins carry out the nucleotide-specific recognition of the target DNA and initiate the cleavage. Although there are exceptions, type-dependent Cas genes (*cas* 1, 3, 4, 5, 6) are usually involved in the expression of the Cas locus. Type II-A system is characterized by the presence of Csn2 involved in spacer integration; type II-B system lacks Csn2 but has Cas4 family protein instead (Chylinski *et al*., 2014). A CRISPR-

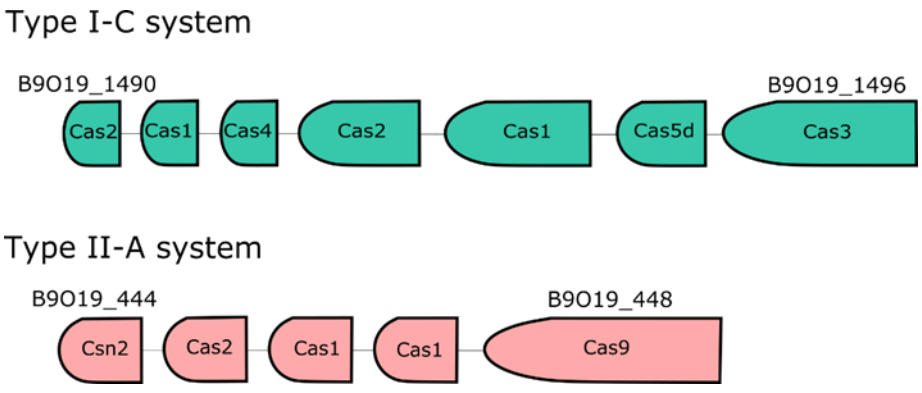


Figure 4.6 CRISPR-Cas systems within the *M. pectinilyticus* genome. The rounded ends indicate the direction of transcription.

associated repeat region (438 bp) of unknown function was located at 875,017 – 875,454 region of the genome. The genes coding for Cas3 and Cas9 appeared to be intact without truncation as their lengths were close to the reported values, suggesting these signature nucleases were likely to be active for initiating the DNA cleavage interference processes. Using CRISPR Finder (Grissa *et al*., 2007), a web-based database of CRISPR repeats and spacer sequences, the spacer-acquiring Cas1 and Cas2 proteins universally present in CRISPR-Cas system, and the type-dependent Csn2 were BLAST searched against the CRISPR database. For type I-C system, a single 19-bp spacer sequence (5'-AGCATTTGATGGAAATACT-3') was identified in the Cas1 protein encoded by B9O19_1494. For type II-A system, a 23-bp spacer sequence (5'-ATAATATACCTATTGAAATACAT-3') and a 19 bp-direct repeat (5'-AATCAGTCTATTGCAGGAT-3') were found in Csn2- and Cas2-encoding genes B9O19_444 and B9O19_445, respectively. The spacers were not recognized from any prophage- and plasmid-associated DNA present in *M. pectinilyticus* genome. Further BLAST searches of the spacer sequences using the PHAST prophage genome database (Zhou *et al*., 2011) yielded no results, but such is not uncommon as bacteriophages infecting clostridial species are known to be significantly underrepresented in the currently available phage genome database (Tomazetto *et al*., 2016).

4.8 The glycobiome of *M. pectinilyticus*

M. pectinilyticus possessed a highly specialized repertoire of CAZymes predicted to be involved in the degradation of pectic polysaccharides. CAZyme domains including glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), glycosyl transferases (GTs), and carbohydrate binding modules (CBMs) were identified using the CAZy database, finding 91 genes annotated to encode 108 putative CAZyme domains (Appendix 5). The 47 putative pectin-associated CAZyme domains include 20 pectate lyases (PL1 and PL9), 1 rhamnogalacturonan lyase (PL11), 8 pectin methylesterases (CE8), 4 pectin acetylesterases (CE12), 1 polygalacturonase (GH28), 1 2-keto-3-deoxynononic acid hydrolase (GH33), 3 β-xylosidase/α-arabinofuranosidases (GH43), 1 α-galactosidase (GH97), 1 unsaturated rhamnogalacturonyl hydrolase (GH105), 1 apiosidase (GH140) and 3 multi-domain enzymes consisting of GH95/CE8 or CE12/CE8. *M. pectinilyticus* also produces 9 putative hemicellulolytic enzymes including 5 β-glucosidase/β-xylosidase/α-arabinofuranosidase

(GH3), 3 α-L-fucosidases (GH95), and 1 β-glucosidase/β-xylosidase (GH116). Among the cellulolytic enzymes, 2 endoglucanases were assigned to GH5, and 1 to GH51. The total number of individual CAZyme domains and the number of different CAZyme families represented in *M. pectinilyticus* genome (3 families of PLs; 17 families of GHs; 3 families of CEs; 8 families of GTs; and 5 CBMs) were smaller compared to the highly developed glycobiome pools reported from cellulosome-producing bacteria such as *C. thermocellum* ATCC 27405 (220), *C. cellulolyticum* H10 (189), *Clostridium cellulovorans* 743B (250), *Ruminococcus flavefaciens* FD-1 (> 200), and *R. albus 8* ($>$ 200) (Munir *et al*., 2014, Dassa *et al*., 2014). The glycobiome size of *M. pectinilyticus* was slightly larger than those of *Eubacterium eligens* ATCC 27750 (83) and *Faecalibacterium prausnitzii* L2/6 (56), both Gram positive pectin-degrading bacteria which are not considered as conspicuous fibre degraders due to the small number of pectin-associated CAZymes present in their genomes. However, *M. pectinilyticus* differentiated itself from other fibrolytic strains by producing pectin-specific CAZymes in a disproportionally high number in relative to the size of its glycobiome. The distribution of pectin-specific CAZymes among various pectin degraders isolated from the human gut (*M. pectinilyticus*, *E. eligens*, *F. prausnitzii*, *B. thetaiotaomicron*), the animal rumen (*Prevotella ruminicola* and *Butyrivibrio proteoclasticus*), and the environment (*Dickeya dadantii*) is compared in Figure 4.7. *M. pectinilyticus* displayed an unusual PL- and CE-dominated CAZyme profile which was characterized by having significantly higher numbers of PL1 and CE8 than other pectin degraders with larger genome sizes. On the other hand, only two genes (B9O19_868 and B9O19_2275) were identified to encode the pectin-specific GH28 and GH105, contrasting with the GH-rich CAZyme compositions usually observed in generalist bacteria such as *B. thetaiotaomicron* and *B. ovatus* (Figure 4.8). Interestingly, *D. dadantii*, an aggressive plant pathogen, showed a PL-dominated glycobiome profile similar to *M. pectinilyticus*, although the pathogen produced a much broader range of PL families. While pectate lyases are abundantly produced by plant fungal/bacterial pathogens to inflict extensive plant tissue maceration, it is unusual for a gut bacterium to derive the majority of its pectinolytic activity from PL families. The unique PL- and CE-dominant CAZyme profile of *M. pectinilyticus*, and its nutritional specialization for pectin/hemicellulose-related monomeric

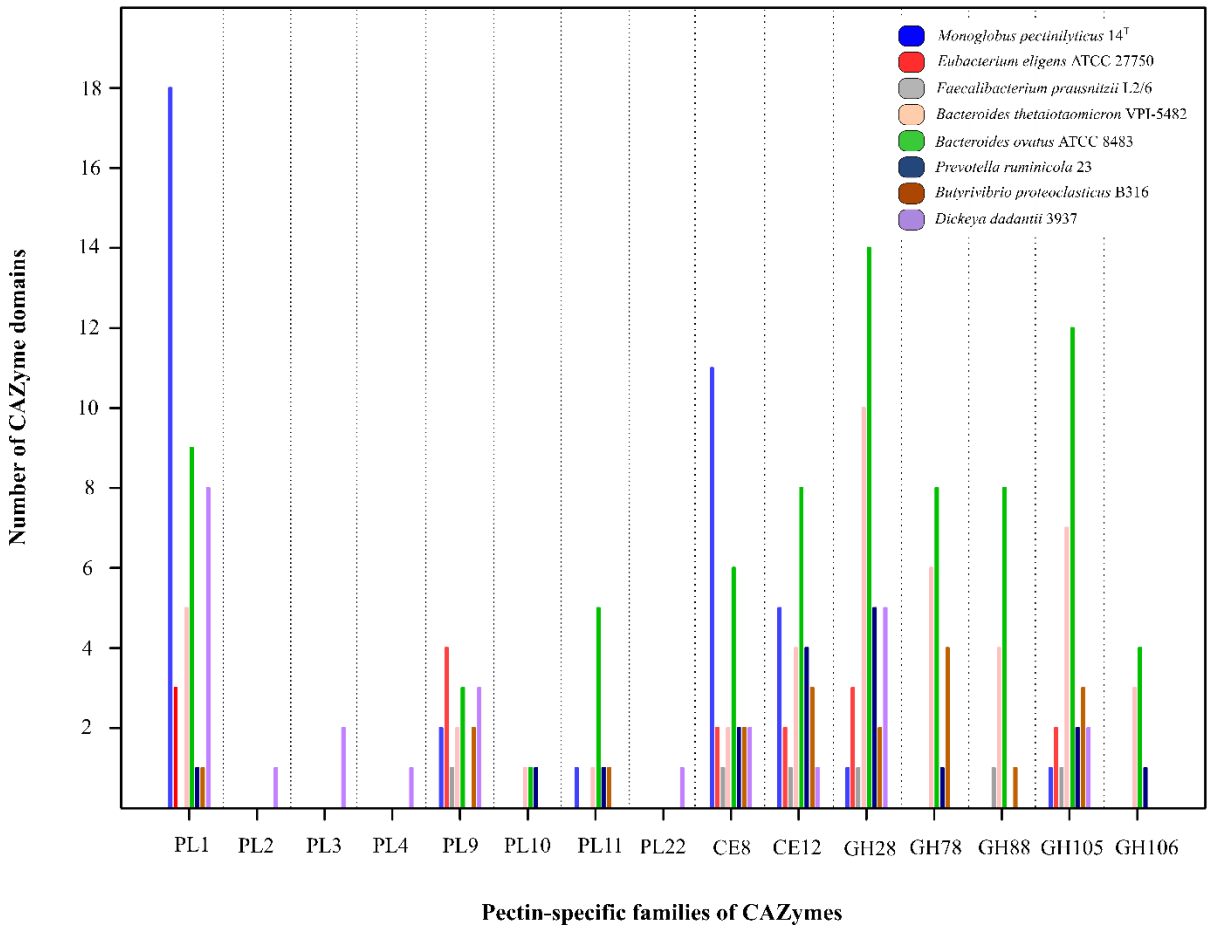


Figure 4.7 Comparison of the selected families of PLs, GHs, and CEs with pectin-specific activities between *M. pectinilyticus* and pectin-degrading strains isolated from the human gut (*E. eligens*, *F. prausnitzii*, *B. thetaiotaomicron*, and *B. ovatus*), the animal rumen (*P. ruminicola* and *B. proteoclasticus*), and plants infested with soft rot disease (*D. dadantii*). The reported genome sizes of these strains are as follow: *M. pectinilyticus* (2.76 Mb), *E. eligens* (2.14 Mb), *F. prausnitzii* (3.32 Mb), *B. thetaiotaomicron* (6.26 Mb), *B. ovatus* (6.48 Mb) *P. ruminicola* (3.62 Mb), *B. proteoclasticus* (3.55 Mb), and *D. dadantii* (4.92 Mb).

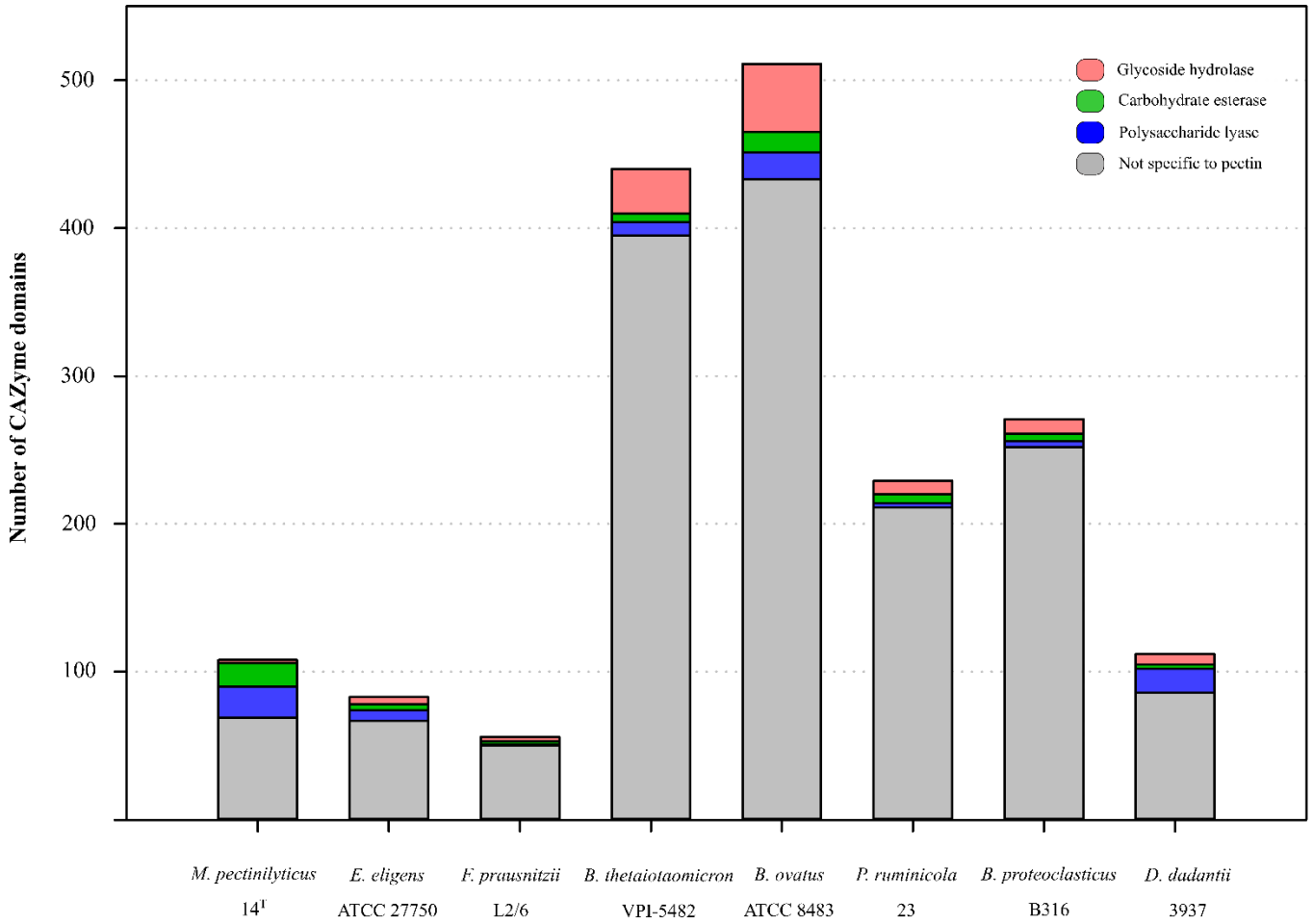


Figure 4.8 Comparison of the number of pectin-specific CAZymes in relative to the total number of CAZymes. PL families of 1, 2, 3, 4, 9, 10, 11, 12; CE families of 8, 12; and GH families of 28, 78, 88, 105, 106 were included among the pectin-specific CAZymes. The number of CAZymes for each reference strain was obtained through the CAZy database.

carbohydrates (most notably galacturonic acid) suggested that the strain has evolved a unique glycobiome for degrading and utilizing the structural components of pectin as the primary carbon source. All PLs and the majority of CE8 and CE12 were detected with secretory signal peptides, suggesting that the solubilisation of HG and RG backbones, and the de-esterification of methyl- and acetyl-groups would mostly take place in the extracellular space. In contrast, GH28 and GH105 were predicted with intracellular activities, indicating the downstream degradation of pectic oligomers would occur after their translocation into the cell. All GH43 proteins lacked identifiable signal peptide sequences, indicating the putative β-xylosidase/α-arabinofuranosidase functions have intracellular activities. Also, *M. pectinilyticus* lacked putative β-galactosidases/galactanases usually required for the degradation of RG-I galactan side chains. Furthermore, while the putative GH33, GH43, GH97, and GH140 may be involved in the partial degradation of RG-II side chains, most of these enzymes were designated with intracellular activities, and *M. pectinilyticus* lacked the comprehensive RG-II degradome recently described from *B. thetaiotaomicron* (Ndeh *et al*., 2017). The glycobiome of *M. pectinilyticus* encoded 3 extracellular GH95 domains which are known to target the fucosyl substitutes from xyloglucan rather than the 2-*O*-Me-L-Fuc*p*-α-1,2-D-Gal*p* linkages present in RG-II. However, the 2 putative GH95/CE8 multi-domain enzymes displayed a novel CAZyme combination which may be useful for cleaving the 2-*O*-methyl-α-L-Fuc*p* in RG-II (Glushka *et al*., 2003). Signal peptide sequences were identified from genes encoding three of five putative GH3 (poly-specific hemicellulases), two of three GH5 endoglucanase, and GH116 β-glucosidase/β-xylosidase, suggesting a partial degradation of hemicellulose and cellulose may occur outside the cell. The cellulolytic and hemicellulolytic CAZyme contents of *M. pectinilyticus* may play roles in the strategic disintegration of the surrounding polysaccharides (e.g. xyloglucan) to release pectin from the plant cell wall structures.

Very few CBM families were identified in *M. pectinilyticus*, possibly due to the underrepresentation of pectin-targeting CBMs characterized in the current version of the CAZy database. Putative CBM13s with a reported binding affinity for xylose, xylan, and galactose were found within PL1 (B9O19_51), and an uncharacterized extracellular proteins (B9O19_1469). Putative cellulose-binding

CBM6 domains were found in association with GH3 (B9O19_398) and GH116 (B9O19_2087). CBM34 and CBM48, which were reported with glycogen- or starch-binding functions, was identified in GH13 family of α-glucosidases (B9O19_1043, B9O19_1110, and B9O19_1731). Additionally, chitin- or peptidoglycan-binding CBM50 domains were found in GH18 chitinase (B9O19_232), lysin motif (LysM)-containing proteins (B9O19_1285), and a spore assembly protein (B9O19_76). Overall, few catalytic enzymes predicted to be involved in plant carbohydrate degradation were associated with identifiable CBMs. It was noted that secreted enzymes were frequently larger in the predicted molecular weight sizes, and often contained more extended regions of unidentifiable domains than intracellular enzymes (Appendix 6).

The unusual PL1 and CE8-dominated CAZyme profile of *M. pectinilyticus* prompted for an attempt to infer the evolutionary history of these enzymes. Using BlastP search of the NCBI protein database, the closest sequence relatives to PL1 and CE8 of *M. pectinilyticus* were found. Catalytic domains were manually extracted from 193 PL1 and 85 CE8 sequences to construct phylogenetic trees using the maximum-likelihood method (Figure 4.9, Appendix 7). The majority of PL1 and CE8 domains of *M. pectinilyticus* forms species-specific clusters that are separate from other clusters consisting of NCBI sequence relatives, suggesting that these enzymes have independently evolved, and may fulfil functions specific to *M. pectinilyticus*. A group of six PL1 (B9O19_312, B9O19_929, B9O19_1443, B9O19_1867, B9O19_2006, and B9O19_2160), and another group of two PL1 (B9O19_1239 and B9O19_1241) show significant protein sequence homology (> 200 bit-score; *e*-value < e-50; and query cover >50 %) only within the members of the groups, suggesting that these subsets of PL1s may have arisen from gene duplication events. A ~207 kDa protein (B9O19_909) containing PL1, fibronectin type III (Fn3), and SLH repeat domains has no close orthologues in NCBI database, and its catalytic domain is unrelated to other PL1 sequences used in this study. B9O19_874 and B9O19_1315 are distantly related to CDC19498 of an uncultured *Eubacterium* species. B9O19_51 is the only PL1 of *M. pectinilyticus* that contain identifiable CBM domains (CBM13-PL1-CBM13) and shows 72 % protein sequence homology with a CBM13-containing PL1 of *Clostridium bornimense* (CDM70399). B9O19_2062 contains extensions of unknown function at both ends of its catalytic PL1

domain (UNK-PL1-UNK) and is related to OLA16957 of an uncultured *Eubacterium* sp. B9O19_873 and B9O19_1377 are distantly related PL1 clusters consisting of heterogeneous groups of bacteria. PL1s encoded by B9O19_1128, B9O19_1129, and B9O19_1442 are distantly related to PL1s bearing putative dockerin domains that are often associated with the assembly of clostridial and ruminococcal cellulosomes. However, no dockerin sequences have been identified from any of the CAZymes in *M. pectinilyticus*, indicating that the few dockerin- and cohesin-like domains present in hypothetical proteins of *M. pectinilyticus* likely occurred outside of a cellulosomal context. All CE8 domain sequences of *M. pectinilyticus* form discrete species-specific clusters, with B9O19_680 and B9O19_869 distantly related to CE8 of *Clostridium thermocellum*, *Acetivibrio cellulolyticus*, and *Paenibacillus mucilaginosus*.

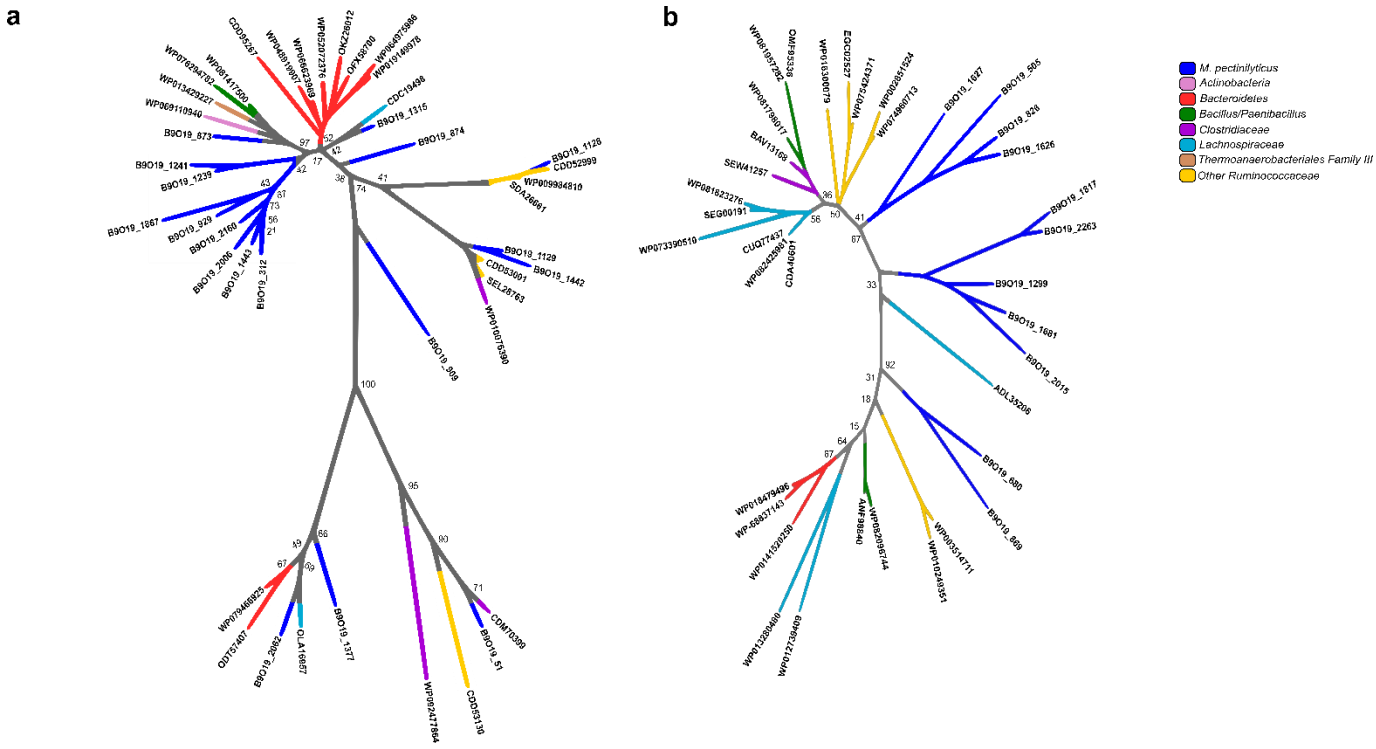


Figure 4.9 PL1 and CE8 domain sequences of *M. pectinilyticus* form discrete species-specific clusters. The PL1 (a) and CE8 (b) catalytic domains were extracted from enzyme sequences, aligned, and used to construct a maximum-likelihood phylogenetic trees. Reference sequences were chosen based on the highest BlastP identity matches to *M. pectinilyticus* PL1 and CE8 sequences. GenBank accession numbers and bootstrap values from 1,000 re-samplings are indicated. Colour indicates the bacterial source of each sequence. Extended phylogenetic trees are given in Appendix 7.

M. pectinilyticus produced 21 putative GTs, with the majority of GTs classified into GT2, GT4, and GT28 families. Due to the large diversity of functions within the same GT families, the exact functions of these poly-specific enzymes are difficult to predict based on the computational data. *M. pectinilyticus* possessed 7 genes encoding CE4 (putative peptidoglycan N-acetylmuramic acid deacetylase), higher than the numbers observed from *R. thermocellum* ATCC27405 (3), *R. albus* 7 (4), *B. thetaiotaomicron* 7330 (3), and *B. ovatus* ATCC 8483 (6) (the CAZy database). Although the biological function of CE4 is little understood except for its presumed contribution to the lysozyme resistance, recent studies suggested that this enzyme class may also be involved in determining the cell shape and separation (Moynihan *et al*., 2014). The presence of GT5 (B9O19_1046) indicated that the synthesis of bacterial glycogen is catalysed by glycogen synthase through the sugar activation to form ADP-glucose, followed by the chain elongation to form α-1,4 glucosidic bond in the glycogen molecule (Preiss, 2006). The enzymes involved in the de-branching of the α-glucosidic bonds present in storage glycogen includes GT5 (B9O19_1596), GH77 (B9O19_60), and GH13 (B9O19_229, B9O19_1043, B9O19_1110, B9O19_1731, and B9O19_2192).

4.9 S-layer homology domain-containing proteins

The genome of *M. pectinilyticus* was searched against currently available protein databases and the CAZyme database to identify dockerins, cohesins, and scaffoldins which constitute the assembly of cellulosome. Only 2 dockerins were found - in hypothetical proteins encoded by B9O19_388 and B9O19_1577 - and 3 cohesin-like domains - in B9O19_1578 and B9O19_1579 - indicating these domains likely occurred outside the cellulosome context. Instead of the genes coding for cellulosomal proteins, *M. pectinilyticus* genome possessed 42 genes encoding extracellular proteins that were predicted to attach to the bacterial cell surface via S-layer homology (SLH) modules. Proteins were considered to contain SLH modules if SLH domains were identified by using both the dbCAN database and BlastP against the GenBank database. Among the SLH proteins of *M. pectinilyticus*, eight are associated with pectin-degrading CAZymes including PL1, CE8, CE12, and GH97; five contain peptidases; two contain N-acetylmuramoyl-L-alanine amidase; four contain domains with poorly characterized functions; and 23 have few identifiable domains other than SLH modules (M_w of

42 kDa – 318 kDa) (Figure 4.10). Searching the 23 uncharacterized SLH proteins and B9O19_241 against various protein databases (UniProt/SwissProt, Pfam, TIGRfam, and InterPro-entailed databases) based on the protein sequence homology revealed that some SLH proteins contained little or no verifiable domains from which functional prediction could be made. SLH proteins were also searched against the GenBank database using the BlastP function. The GenBank protein database operated by NCBI contains a vast amount of peptide sequence data ranging from thoroughly characterized proteins to hypothetical proteins with no known functions, protein sequences from complete and draft quality bacterial genomes, as well as shotgun metagenome sequencing data of uncultured organisms. Proteins were considered to have found homologous matches in the GenBank protein database if the aligned proteins passed the minimum BlastP thresholds of 1) $>$ 200 bit-score; or 2) $<$ 200 bit-score, but $>$ 20 % sequence coverage and $>$ 40 % sequence identity. Among the 23 uncharacterized SLH proteins, only B9O19_1224 and B9O19_2167 found homologous matches which met the minimum requirements. B9O19_1224 showed ~ 500 bit-score, 88 % query cover, and 41 % identity to 'S-layer domain-containing protein' and 'hexagonal wall protein' from a *Firmicutes* bacterium CAG:41 (accession number CDB95905.1) and an uncultured *Clostridium* sp. (accession number SCH67055.1), respectively. B9O19_2167 showed ~ 500 bit-score, 98 % query cover, and 41 % sequence identity to a hypothetical protein from *Sedimentibacter* sp. B4 (accession number WP_019228207.1). While these matches should be considered as proteins of significant sequence similarity, no functional information could be deduced from the results. Therefore, it was concluded that all available search options were exhausted, and that the SLH proteins of *M. pectinilyticus* belong to previously undiscovered groups of proteins with potentially novel characteristics, structures, and functions. All but one (B9O19_1870) SLH proteins possessed N-terminal type I signal peptide, suggesting these proteins were secreted to the bacterial cell surface. B9O19_1870 contained a lipoprotein type II signal peptide. An *in silico* analysis of protein secondary structure using SPIDER2 predicted that the SLH proteins mostly consisted of a β-sheet profile spanning the majority of the protein regions, while the α-helix-rich regions of SLH proteins usually overlapping with CAZyme domains, dockerin-like domains, or SLH modules (Heffernan *et al*., 2016). The 23 uncharacterized

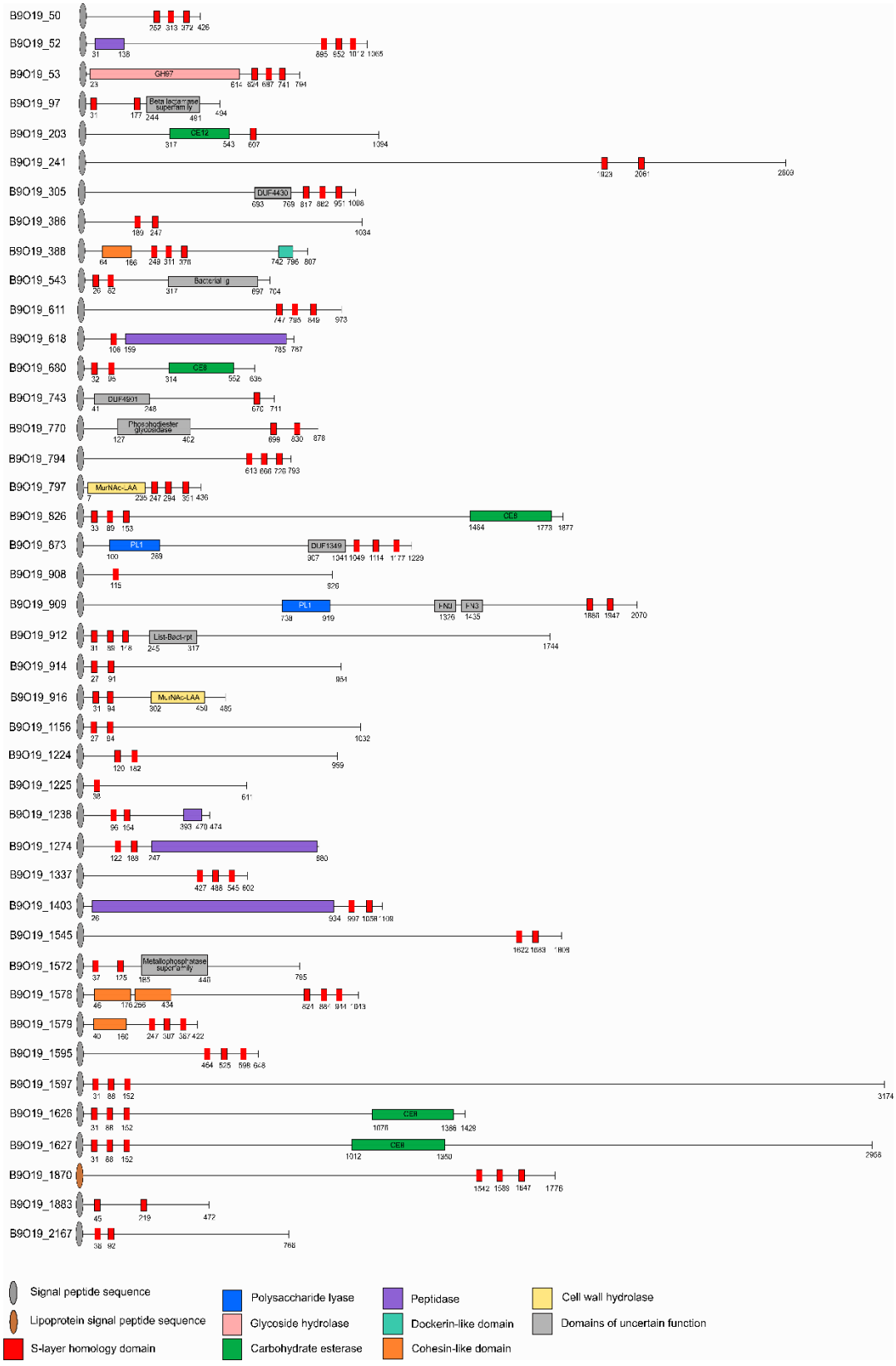


Figure 4.10 The 42 SLH module-containing proteins of *M. pectinilyticus*. Protein lengths are drawn to scale. Amino acid positions are indicated with numbers. Predicted protein domains are shown as rectangles, and coloured according to the domain category shown in the key. Where possible, additional domain details are labelled within domain rectangles. Remaining regions, with no known homology to existing protein domains, are represented by lines.

SLH proteins were aligned using ClustalW and BlastP to compare the protein sequence similarity (Figure 4.11). The most significant sequence similarities were observed between B9O19_50 and B9O19_1579, and between B9O19_1578 and B9O19_1579. While B9O19_1578 and B9O19_1579 had similar protein domain compositions made of SLH modules and putative type I and type II cohesion-like domains, respectively, B9O19_50 contained no known protein domains apart from the SLH repeats present at the C-terminal. The duplicated segments of type I cohesin-like domains on B9O19_1578 showed a high degree of protein sequence homology (74 %) to each other. B9O19_611 shared 35 % and 33 % sequence identities with B9O19_1545 and B9O19_1870, respectively. Overall, with few exceptions, SLH proteins lacked significant sequence homology to each other, suggesting

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1. B9O19_1238	100	21	24	18	15	11	17	16	15	14	18	14	16	23	20	22	23	22	21	16	26	21	23
2. B9O19_912	21	100	28	13	17	14	20	14	17	16	12	22	20	24	16	29	33	31	31	19	31	27	33
3. B9O19_1597	24	28	100	17	16	14	20	15	15	16	14	18	19	20	16	26	27	26	27	15	27	24	30
4. B9O19_908	18	13	17	100	23	19	15	14	16	19	25	16	17	14	23	17	17	14	16	16	16	15	18
5. B9O19_914	15	17	16	23	100	22	18	14	20	21	22	23	26	22	22	25	23	21	27	18	25	24	24
6. B9O19_1883	11	14	14	19	22	100	14	11	13	16	15	16	17	18	16	19	14	15	18	12	19	20	18
7. B9O19_1225	17	20	20	15	18	14	100	18	16	21	19	15	16	19	17	21	32	22	23	19	23	23	25
8. B9O19_241	16	14	15	14	14	11	18	100	18	17	15	17	16	18	18	18	15	16	16	15	15	17	19
9. B9O19_2167	15	17	15	16	20	13	16	18	100	23	20	24	22	17	20	22	23	25	23	19	22	25	27
10. B9O19_1156	14	16	16	19	21	16	21	17	23	100	26	23	22	21	24	23	19	20	24	15	28	22	27
11. B9O19_1224	18	12	14	25	22	15	19	15	20	26	100	19	20	20	25	21	22	17	25	17	22	23	28
12. B9O19_1870	14	22	18	16	23	16	15	17	24	23	19	100	33	24	18	18	18	21	19	20	19	20	19
13. B9O19_0611	16	20	19	17	26	17	16	16	22	22	20	33	100	35	17	16	16	18	19	16	18	17	19
14. B9O19_1545	23	24	20	14	22	18	19	18	17	21	20	24	35	100	14	17	17	24	23	22	25	22	24
15. B9O19_386	20	16	16	23	22	16	17	18	20	24	25	18	17	14	100	19	19	19	20	21	22	25	22
16. B9O19_305	22	29	26	17	25	19	21	18	22	23	21	18	16	17	19	100	21	21	22	24	24	22	26
17. B9O19_52	23	33	27	17	23	14	32	15	23	19	22	18	16	17	19	21	100	25	22	26	24	24	26
18. B9O19_794	22	31	26	14	21	15	22	16	25	20	17	21	18	24	19	21	25	100	28	26	27	24	28
19. B9O19_1337	21	31	27	16	27	18	23	16	23	24	25	19	19	23	20	22	22	28	100	30	30	28	32
20. B9O19_388	16	19	15	16	18	12	19	15	19	15	17	20	16	22	21	24	26	26	30	100	30	28	32
21. B9O19_50	26	31	27	16	25	19	23	15	22	28	22	19	18	25	22	24	24	27	30	30	100	37	52
22. B9O19_1578	21	27	24	15	24	20	23	17	25	22	23	20	17	22	25	22	24	24	28	28	37	100	49
23. B9O19_1579	23	33	30	18	24	18	25	19	27	27	28	19	19	24	22	26	26	28	32	32	52	49	100

Figure 4.11 Percentage identity matrix of SLH domain-containing proteins of *M. pectinilyticus*. Protein sequences were aligned using ClustalW (Larkin *et al*., 2007) which expressed the protein identity in percentage format. Colour intensity corresponds to the sequence similarities relative to 100 % sequence identity which suggests perfect amino acid matches are found along the entire length of alignments.

that either these proteins were either not related in structures or functions, or the primary amino acid sequences of these proteins may not be the determining factor for shaping their 3-dimensional architecture.

4.10 Metabolic capacity of *M. pectinilyticus*

In section 3.7, I have reported *M. pectinilyticus* utilizes few pectin- and hemicellulose-associated monomeric sugars, including L-arabinose, D-xylose, and D-galacturonic acid. D-fructose was the only non-pectin sugar which the strain could utilize. In accordance with this observation, the KEGG analysis of *M. pectinilyticus* genome revealed that the strain possessed metabolic pathways for utilizing uronic acids, D-xylose, L-arabinose, and D-fructose (Figure 4.12; Appendix 8). *M. pectinilyticus* possessed the two best-known ways for the bacterial utilization of uronic acid; the hydrolase-mediated isomerase pathway for metabolising galacturonic/glucuronic acids, and the lyase-mediated pathway for metabolising DKI (Richard and Hilditch 2009). The intracellular activities of GH28 (B9O19_868) and GH105 (B9O19_2275) were predicted to cleave the pectic oligomers to produce D-galacturonate and DKI which then enter the uronic acid utilization pathways. The scarcity of intracellular enzymes may reflect the specialist nature of *M. pectinilyticus* which actively carries out the extracellular degradation of pectin, but has a limited metabolic capacity to utilize the end-products of degradation. D-xylose and L-arabinose are both pentose sugars which become processed through the intermediates of the pentose phosphate pathway to produce glyceraldehyde-3-phosphate and pyruvate. D-xylose monomers are converted to D-xylulose and then to xylulose-5-phosphate by xylose isomerase (*xyl*A; EC5.3.1.5) and xylulokinase (*xyl*B; EC2.7.1.17). L-arabinose becomes converted to L-ribulose and then to L-ribulose-5-phosphate by the consecutive actions of L-arabinose isomerase (*ara*A; EC5.3.1.4) and L-ribulokinase (*ara*B; EC2.7.1.16), eventually converging to form the intermediary xylulose-5-phosphate by the action of L- ribulose-5-phosphate 4-epimerase (*ara*D; EC5.1.3.4). Phosphoketolases are thiamine diphosphate-dependent enzymes responsible for two phosphoroclastic reactions: (1) conversion of xylulose-5-phosphate into glyceraldehyde-3-phosphate and acetyl phosphate, and (2) conversion of fructose-6-phosphate to erythrose-4-phosphate and acetyl phosphate.

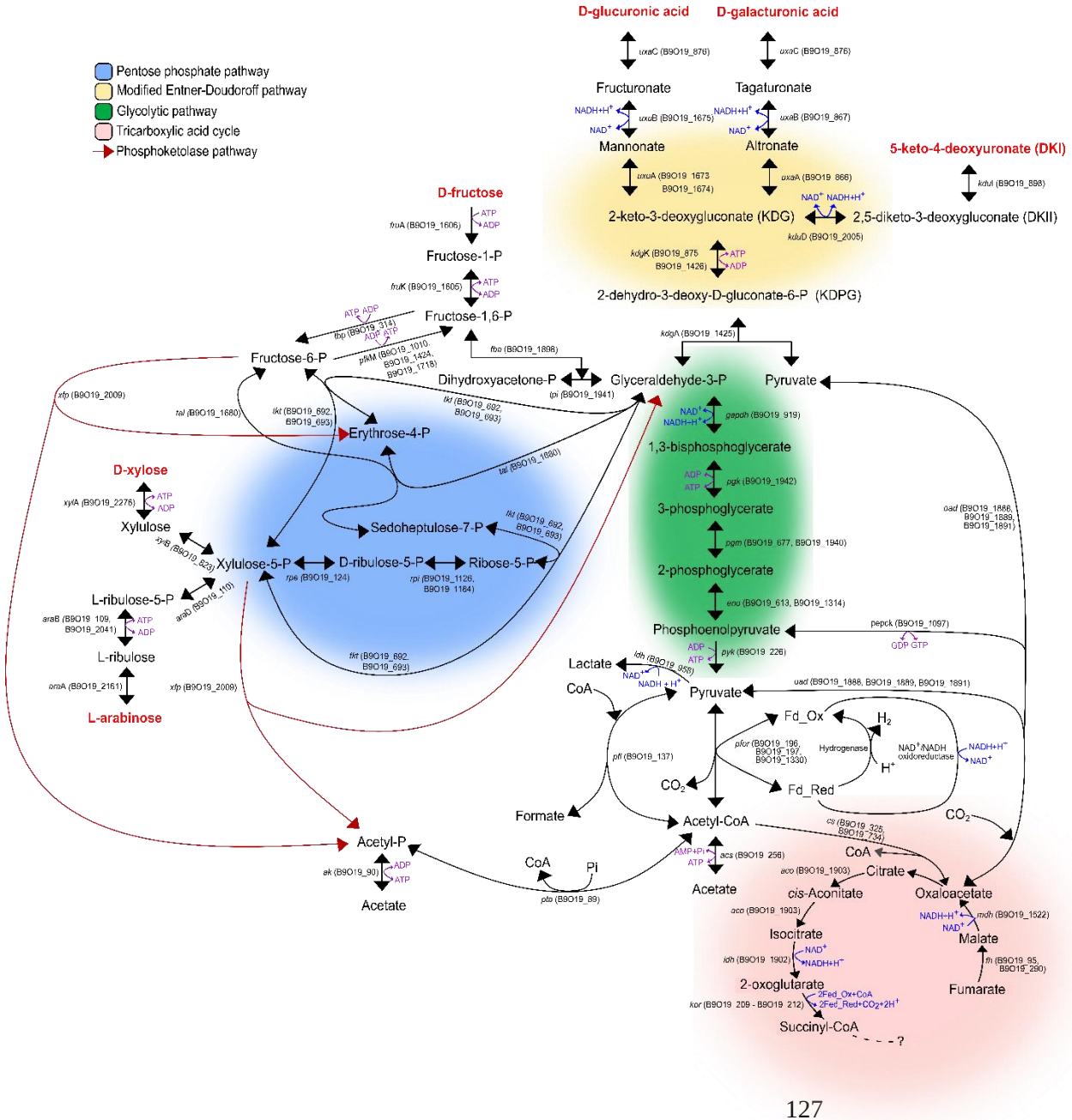


Figure 4.12 Reconstruction of sugar metabolism by *M. pectinilyticus*. Colour shadings distinguish major metabolic pathways. Arrows indicate the direction of enzymatic reactions. Locus tag numbers are given. KEGG numbers and enzyme identities of each gene are given in Appendix 8. Question mark indicates missing genes (*scs* and *sdh*) in an otherwise complete TCA cycle.

Xylulose-5-phosphate is a metabolic derivative of xylose and arabinose - predominant pentose sugars of pectin that are utilized by *M. pectinilyticus*. Notably present in *Bifidobacteria* (Meile *et al*., 2002; Yin *et al*., 2005), *Lactobacillale*s (Tanaka *et al*., 2002; Okano *et al*., 2009), and *Clostridium acetobutylicum* (Liu *et al*., 2012; Sund *et al*., 2015), mono- or bi-functional phosphoketolases have been found to shunt xylulose-5-phosphate and/or fructose-6-phosphate through a catabolic alternative of the pentose-phosphate pathway by bypassing carbon flux through the lower portion of glycolysis, thus avoiding a complete oxidation of carbon to CO_2. As a result, the phosphoketolase pathway yields fewer ATP than pentose phosphate pathway (Liu *et al*., 2012), but also has lesser needs for NAD^+ regeneration as it adopts an enzymatic 'shortcut' through the conversion of xylulose-5-phosphate to acetate (Liu *et al*., 2012; Servinsky *et al*., 2012). Theoretically, more than one route of pentose sugar metabolism may exist in *M. pectinilyticus*, allowing greater metabolic flexibility. The cellular transport of D-fructose is carried out using the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS^{Fru}) which requires the initial phosphorylation of D-fructose into D-fructose-1-phosphate before the transport of sugar across the cell membrane. *M. pectinilyticus* possessed a single PTS system designated for the uptake of D-fructose, and lacked cellular transport systems for sugars such as D-mannose, D-glucose, D-sucrose, and D-mannitol. Genes encoding the proteins of PTS^{Fru} (cytoplasmic EI and HPr, membrane-bound permease complex made of EIIA, EIIB, and EIIC, sugar-phosphorylating enzyme phosphotransferase, and phosphofructokinase (EC2.7.1.11) which carries out the conversion of fuctose-1-phosphate to fructose-1,6-bisphosphate) were organized in a *lac* operon-like cluster (B9O19_1604 – B9O19_1609), beginning with a gene coding for a transcription repressor of DeoR family, and ending with a gene coding for histidine kinase-type of sensory regulator (Voigt *et al*., 2014). A full set of glycolytic enzymes in the Embden-Meyerhof-Parnass (EMP) pathway were present in the *M. pectinilyticus* genome. Glucose utilization begins with the glucokinase (EC2.7.1.2)-mediated conversion of α- or β-D-glucose to α- or β-D-glucose-6-phosphate. Both α- and β-forms of D-glucose-6-phosphate become converted to fructose-6-phosphate and then to fructose-1,6-phosphate by the sequential actions of glucose-6-phosphate isomerase (EC5.3.1.9) and 6-phosphofructokinase (EC2.7.1.11) before progressing to the glycolytic conversion to pyruvate. The lack of growth on D-glucose, as well as the absence of glucose-type

transporters may suggest that *M. pectinilyticus* has a limited use of the EMP pathway for glucose utilization. KEGG analysis of *M. pectinilyticus* genome revealed that acetate is produced from acetyl CoA through the action of acetate synthase (B9O19_256; EC6.2.1.1) and alternatively, through a combination of phosphate acetyltransferase (B9O19_89; EC2.3.1.8) and acetate kinase (B9O19_90; EC2.7.2.1). Furthermore, formate and lactate are predicted to be produced via formate C-acetyltransferase (B9O19_137; EC2.3.1.54) and L-lactate dehydrogenase (B9O19_958; EC1.1.1.27).

4.11 Organization of pectin gene clusters

The organization of carbohydrate utilization genes in *M. pectinilyticus* was characterized by the formation of gene clusters encoding various combinations of carbohydrate-degrading enzymes, metabolic enzymes, transcription regulators, and ABC-type transporter system, forming distinct genomic regions of pectin-utilizing functions. The two most notable 'pectin clusters' were observed between B9O19_864 and B9O19_877, and between B9O19_2005 and B9O19_2015 (Figure 4.13b & e). The former spans a ~22 kb genomic segment and contains genes coding for CAZymes (GH28, GH43, PL1, and CE8); Gal*p*A-metabolising enzymes *(uxa*A, *uxa*B, and *uxa*C); KDG kinase (*kdg*K); a sugar-binding ABC type transporter unit; and an AraC type transcription regulator. Another cluster of genes between B9O19_2005 and B9O19_2015 forms a ~21 kb genomic segment encoding CAZymes (PL1, GH33, GH51, GH95, and CE8); 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (*kdu*D); a phosphoketolase (*xfp*); a GntR type of transcription regulator; a biotin transporter; and proteins of ambiguous functions. Some genes encoding SLH proteins were found adjacent to the genes coding for putative CAZymes involved in the pectin degradation (Figure 4.13a, c & d). B9O19_50 – B9O19_53 contained genes coding for two SLH proteins, PL1, and GH97 associated with SLH modules, while B9O19_908 and B9O19_1238 that encode SLH proteins were found adjacent to the genes coding for PL1s. In addition, a number of SLH proteins, including B9O19_1156, B9O19_1224, B9O19_1225, and B9O19_1337 were also found adjacent to or in vicinity of the genes encoding CE4, GT2, and GT26. Whether these SLH proteins participate in assisting the catalytic functions of CAZymes remains unknown. The organization of gene clusters of mixed groups of CAZymes, metabolic enzymes, transcription regulators, and transporters suggests that the gene regulation may occur at a

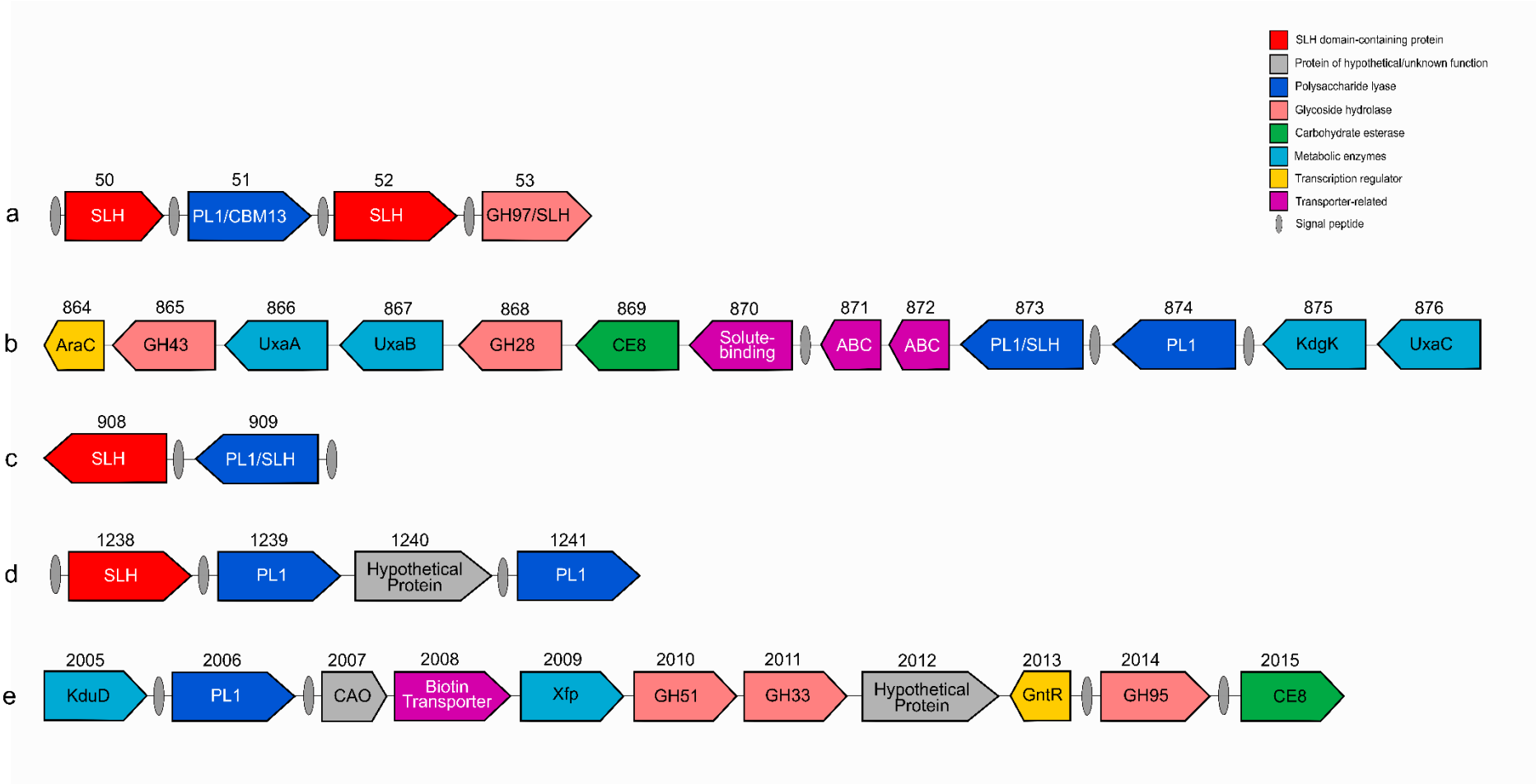


Figure 4.13 Schematic representation of gene clusters concerning the pectin degradation/utilization. Gene orientation is indicated by arrows. Gene numbers (locus tag B9O19) are indicated above the arrows. Colours corresponding to gene functions are marked in the product category.

transcriptional level to synchronize the catabolic and metabolic processes to enhance the efficiency of the pectin degradation. Using the ABC transporter database (Fichant *et al*., 2006), we identified three ABC-related transporters (B9O19_98 – B9O19_100, B9O19_870 – B9O19_872, and B9O19_1089) presumably involved in carbohydrate-binding and transport. Additionally, B9O19_1628 was a non-ABC type extracellular transporter located next to two CE8 proteins containing SLH proteins (B9O19_1626 and B9O19_1627). Although the exact function of this transporter is currently unknown, its extracellular nature and the proximity to pectin-degrading CAZymes may suggest its involvement in transporting carbohydrates across the cell wall. Some genes with pectin-related functions appeared to be under the control of AraC- or GntR-type transcription regulators. GntR type of DNA-binding activators were found adjacent to glucuronate metabolism genes such as *uxu*A (B9O19_1673 and B9O19_1674) and *uxu*B (B9O19_1675), and also GH95 (B9O19_2014), while an AraC type regulator was present at the end of the large pectin cluster found at B9O19_864 – B9O19_877.

4.12 Endospore formation, cofactor synthesis and motility

Very few Gram positive coccus-type bacteria are known to form endospores. Some examples that do are the aerobic soil bacterium, *Sporosarcina ureae* (Robinson and Spotts 1983) and *R. bromii* (Mukhopadhya *et al*., 2017). In section 3.4.2, it was shown that nutrient-starved conditions did not induce the endospore formation in *M. pectinilyticus*. However, *M. pectinilyticus* possessed genes encoding sporulation-specific sigma factors (spo0A, σ^E, σ^F, σ^G, σ^H, and σ^K), a cluster of sporulation stage II proteins for asymmetric cell division, sporulation stage III proteins involved in the cell engulfment process, spore coat structural proteins, and spore proteases and phosphatases. A recent study of spore-forming population in the human gut microbiota revealed that 60 % of bacterial genera, most notably found within the family *Ruminococcaceae* and *Lachnospiraceae*, contained 'signature' sporulation genes (Browne *et al*., 2016). Brown *et al*. (2016) reported that potential spore formers isolated using ethanol-treated plates were shown to have a signature sporulation gene coverage of > 50 %, although the spore-forming ability of these isolates could not always be verified. In *M. pectinilyticus*, 34 genes found matches in the list of 66 highly conserved sporulation genes that were

used to make the computational prediction of the spore-forming ability of bacteria (Appendix 9). The predicted signature score of *M. pectinilyticus* (~52 %) was lower than the values obtained from distantly related spore-formers (53 – 63 %) such as *Lysinibacillus sphaericus*, *Sulfobacillus acidophilus*, and *Caldicellulosiruptor kristjanssonii*, or the 70 – 85 % signature range reported from *Clostridium* and *Bacillus* spp. (Browne *et al*., 2016). As starvation-induced endospore formation does not often occur in a gut environment, the endospore-forming mechanism may play a role in increasing the resilience of the strain to trace oxygen exposure. *M. pectinilyticus* also possessed MreD (B9O19_715), MreB (B9O19_717 and B9O19_1004), FtsW (B9O19_1358 and B9O19_1582), and RodA (B9O19_1231) genes involved in rod-shape forming and non-spherical cell maintenance, suggesting the strain may have been a rod-shaped cell once, but somehow reverted to the default coccus cell type possibly as an adaptation strategy to the gut environment or to the laboratory growth conditions. No genes involved in flagella assembly or exopolysaccharide biosynthesis were found in *M. pectinilyticus*, in accordance with our previous observations (section 3.4). Metabolic pathways for the synthesis of all 20 essential amino acids were present. The essential growth requirement for the addition of vitamin cofactors could be explained by the lack of biosynthesis pathways for biotin, riboflavin, ascorbate, and folate (section 3.3). On the other hand, nicotinate (vitamin B3), vitamin B5, vitamin B6, and thiamine (water-soluble vitamin B) appeared to be cellularly synthesized.

Chapter 5 Proteomic profiling of *Monoglobus pectinilyticus* using iTRAQ

5.1 Introduction

As described in the previous chapter, *Monoglobus pectinilyticus* is the only human gut bacterium known so far to possess a glycobiome specifically shaped to target pectin for degradation and utilization. The genome of *M. pectinilyticus* encodes sequences for 42 proteins containing S-layer homology (SLH) modules, and also 108 putative CAZyme domains spread across 34 CAZy families, of which 47 CAZyme domains are predicted with pectinolytic functions. A number of pectin-degrading CAZymes are putatively anchored to the cell surface via SLH domains, while others seem to be secreted to the environment. The exact roles of SLH proteins remain unclear, and their involvement in conferring pectinolytic ability to *M. pectinilyticus* has not been explored previously. However, large uncharacterized SLH proteins were recently identified as constituents of novel lignocellulolytic degradation systems in *Paenibacillus curdlanolyticus* B-6 and *Caldicellulosiruptor* spp., indicating that SLH proteins of *M. pectinilyticus* may represent constituents of a novel pectin degradation system (Conway *et al*., 2016; Ratanakhanokchai *et al*., 2013). This chapter presents the findings from the investigation into the protein production of *M. pectinilyticus* in response to the availability of pectin. Based on the assumption that the extracellular proteins bearing SLH domains are involved in pectin degradation, it was hypothesized that the presence of complex pectin in the growth medium may induce *M. pectinilyticus* to produce these proteins in a greater abundance on the bacterial cell surface compared to the cells growing on simple sugars such as fructose as a sole carbon source. Proteins produced from *M. pectinilyticus* cells grown on 0.2 % (w/v) pectins from citrus peel, apple pomace, and kiwifruit were quantified relative to the cells grown on 0.5 % (w/v) D-fructose using isobaric tags for relative and absolute quantification (iTRAQ). iTRAQ is a mass spectrometric technology which measures the intensities of reporter ions released from iTRAQ-tagged peptides upon their fragmentation (Karp *et al*., 2010). Figure 5.1 provides the workflow of sample processing and iTRAQ data analysis. The stable isotope tags used in iTRAQ consist of a peptide reactive group

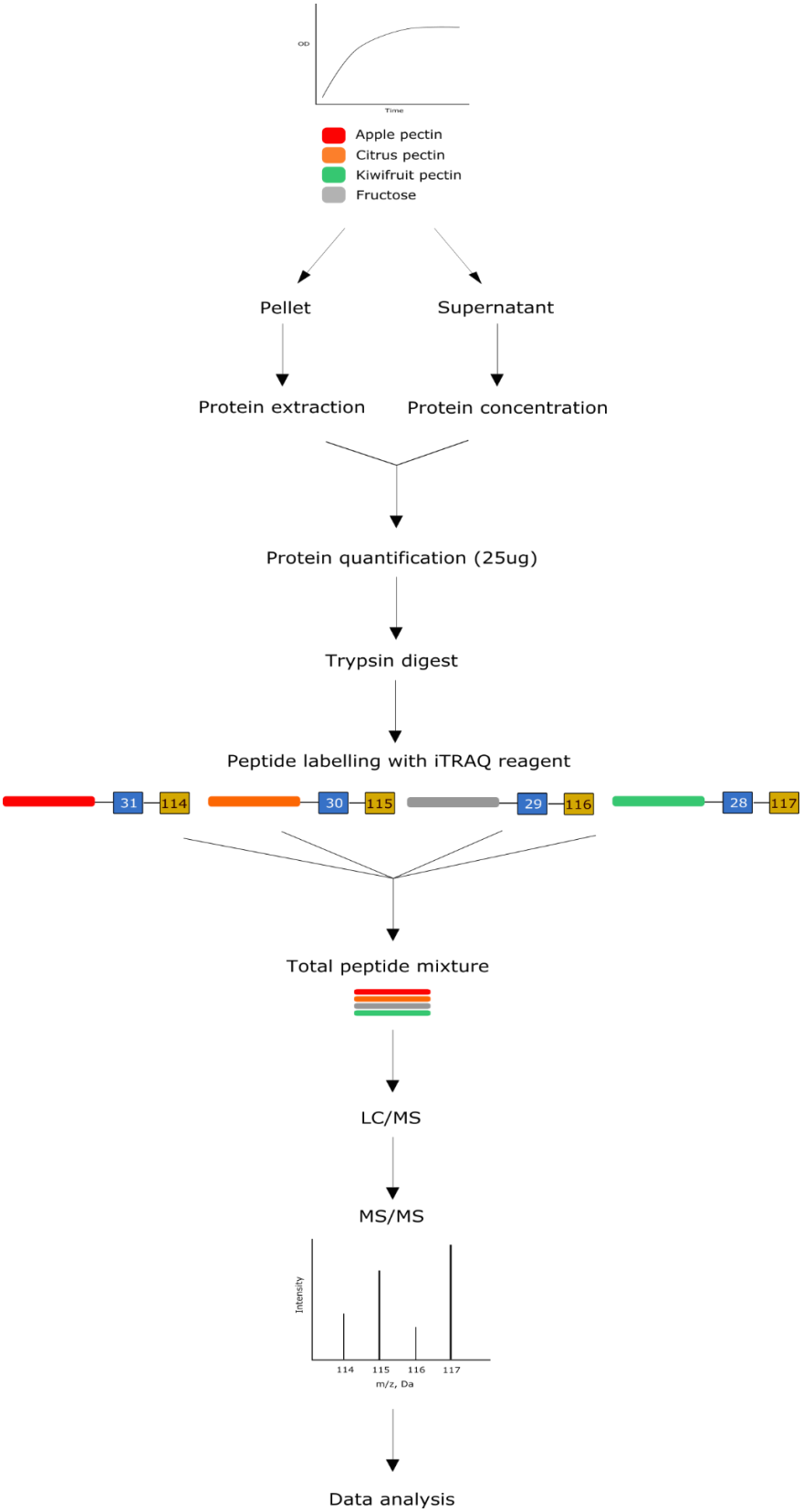


Figure 5.1 Workflow diagram of quantitative proteomics using iTRAQ labelling and mass spectrometry.

which reacts with primary amines of peptides, and a uniquely charged reporter group (28 – 31 Da), and a balance group (114 – 117 Da) to maintain the overall mass of 145 Da (Fuller and Morris, 2012). iTRAQ technology allows multiplexing of 4 and 8 separately labelled samples within a single experiment, which are then separated by LC/MS based on the peptide sizes, subsequently followed by peptide sequencing using tandem mass spectrometry (MS/MS). During the MS/MS run, the balance group is released, thus exposing the charged reporter group to provide relative quantification information of peptides across multiple samples by measuring the relative intensities of iTRAQ reporter ions. The iTRAQ protein quantification experiment was designed to investigate the differential expression of CAZymes and SLH proteins in the presence of pectin in the growth medium. The iTRAQ quantification of whole-cell proteome and the exoproteome of *M. pectinilyticus* revealed a set of SLH domain-containing proteins showing more elevated levels of protein expression in the presence of apple, citrus, and kiwifruit pectin compared to fructose-grown cells. Furthermore, the CAZymes secreted without SLH modules were mostly identified from the exoproteome of culture supernatant, suggesting *M. pectinilyticus* may take an advantage of a concerted effort of freely secreted enzymes and SLH proteins tethered to the cell surface to enhance the efficiency of pectin degradation.

5.2 iTRAQ protein identification

Three biological replicates were used across all pellet samples. Together, a total of 1,676 proteins were identified from 11,398 peptides for pellet samples at 5 % critical false discovery rates (FDR). In this study, only the proteins from *M. pectinilyticus* which were represented by at least three peptides with unused scores of > 2 were included in the total protein abundances. Using these criteria, 920 proteins were represented by ≥ 3 peptides from *M. pectinilyticus*, and with unused scores > 2. Overall, approximately 17 % of the total CDSs in *M. pectinilyticus* was identified in the pellet proteome. For supernatant proteins, a total of 133 proteins were identified with ≥ 3 peptides with unused score > 2 from 3,004 peptides at 5 % FDR in a single sample without replicate. Contaminant protein entries such as human keratin introduced through sample handling, and the traces of endogenous plant proteins from apple, citrus, and kiwifruit pectin were manually removed. Proteins

represented by at least three peptides with 95 % confidence, FDR-corrected *p*-value < 0.05, and fold-changes of > 1.20 or < 0.83 relative to the fructose control were regarded as differentially produced with statistical significance. The z-score for each replicate was calculated from the *p*-value using the NORMINV function in Excel. The mean z-score was calculated as the arithmetic average of the z-scores for the three replicates. A mean z-score value of <1.65 indicates that the average protein production ratios for the three replicates lie outside the normal distribution (outermost 10 % of the population). The mean relative protein expression is the geometric mean of the relative expressions in the three replicates (Appendix 10).

5.3 Differential production of pellet proteins identified using iTRAQ

Approximately 159 proteins were found to be commonly shared in all three pellet proteomes. An overview of the relative changes in protein production between pectins and fructose is given in Figure 5.2 and the raw data is given in Appendices 11 – 13. Overall, extracellular proteins containing SLH modules were often pronouncedly up-regulated in all three pellet samples, suggesting components of pectin may serve as a regulatory trigger to increase the production of these proteins in *M. pectinilyticus*. Among the SLH proteins, B9O19_52 and B9O19_611 showed the tendency of up-regulation in the presence of all three types of pectins. The gene coding for B9O19_52 is found within a cluster of genes that encode PL1 and GH97-containing SLH protein, tentatively suggesting that B9O19_52 may serve a function related to the pectin degradation (Figure 4.13). B9O19_1156 (SLH protein) and B9O19_1627 (CE8/SLH protein) were commonly up-regulated in the presence of apple and kiwifruit pectins, but down-regulated with citrus pectin. B9O19_1597 (SLH protein) was up-regulated in the presence of apple pectin, and not significantly changed in the presence of citrus and kiwifruit pectin. B9O19_1156 and B9O19_1597 were represented by the highest numbers of peptides in all three replicates. CE8/SLH protein (B9O19_1626) was consistently down-regulated with all three pectins, while B9O19_826 (CE8/SLH protein) showed an increased production with apple pectin. SLH proteins encoded by B9O19_908, B9O19_912, B9O19_1337, B9O19_1578, and B9O19_1595 were either not significantly changed or down-regulated in the presence of any pectins. B9O19_682 (hypothetical protein) was commonly up-regulated in the presence of apple and

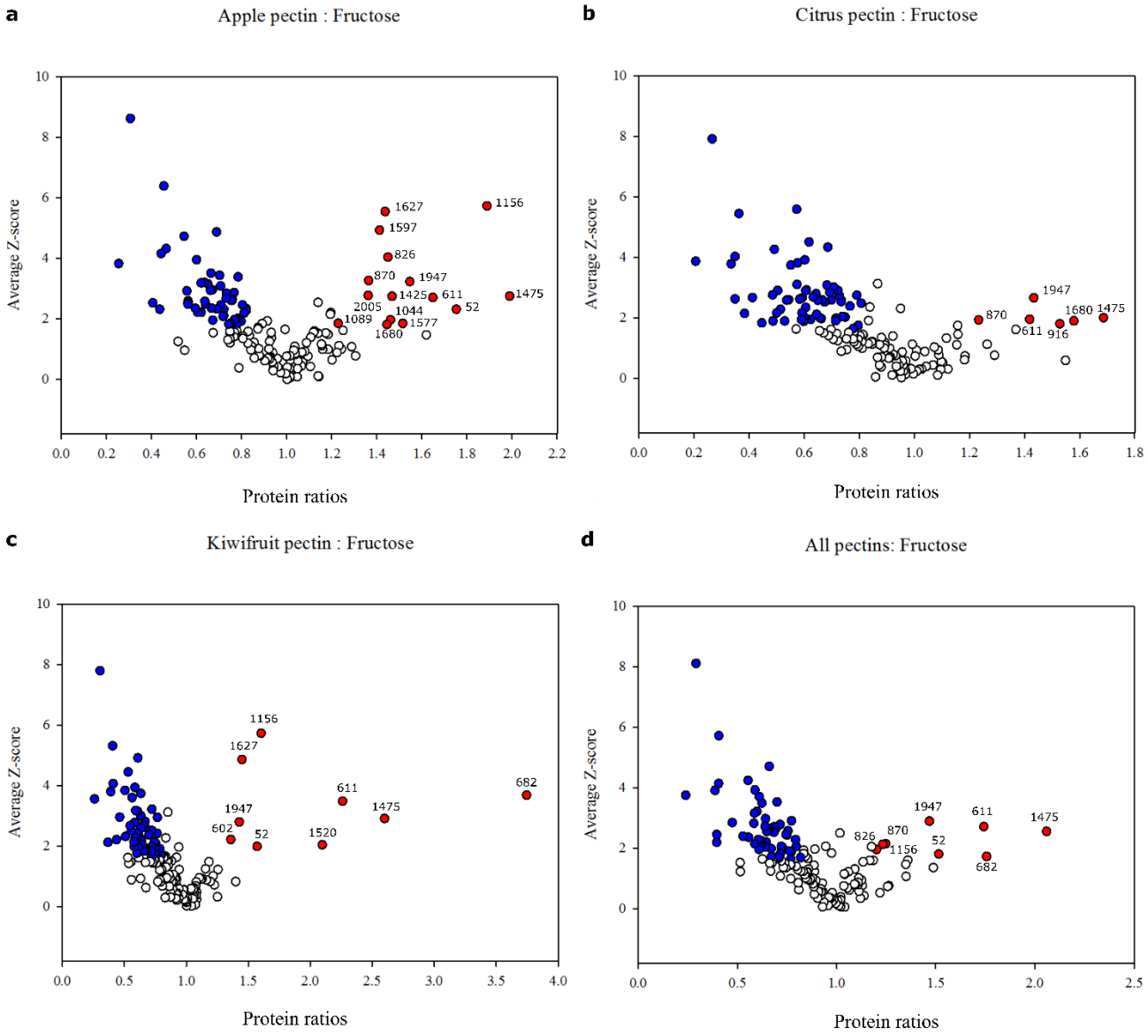


Figure 5.2 Protein ratios between *M. pectinilyticus* grown on fructose control and pectins (a, apple pectin; b, citrus pectin; c, kiwifruit pectin; and d, all nine replicates). Red, up-regulated proteins; blue, down-regulated proteins; white, not significantly changed. Locus tag (B9O19) numbers are indicated for up-regulated proteins. Protein ratios of $>$ 1.20 or $<$ 0.83 relative to the fructose control, and z-score value of $\geq$ 1.65 were regarded as differential with statistical significance.

kiwifruit pectins, although protein ratio was much greater for cells grown using kiwifruit pectin. B9O19_1475 (C1A family cysteine peptidase) and B9O19_1947 (elongation factor G) were up-regulated with all three pectins, but the reason for this is currently unclear. A transcription regulator (B9O19_1520) was up-regulated in the presence of apple and kiwifruit pectin. B9O19_1520 is found adjacent to two genes coding for cell surface-associated hypothetical proteins (B9O19_1517 and B9O19_1518; B9O19_1519 is a XRE family transcription regulator). An extracellular solute-binding protein (B9O19_870) was up-regulated with apple and citrus pectin. As mentioned in section 4.10, B9O19_870 is found within one of the major pectin gene clusters (Figure 4.13). A sugar-binding ABC transporter (B9O19_1089) and another sugar-binding protein (B9O19_2227) were up-regulated in the presence of apple and kiwifruit pectins, respectively. Metabolic enzymes involved in the uronate utilization (B9O19_1425 and B9O19_2005) and pentose phosphate pathway (B9O19_1680) were in pectin samples were often up-regulated compared to the fructose control. When the pectin:fructose protein ratios were combined and compared across all nine replicates, B9O19_52, B9O19_611, B9O19_826, B9O19_682, B9O19_870, B9O19_1475, and B9O19_1947 were consistently up-regulated with statistical significance, suggesting these proteins may be highly produced in response to the availability of pectin (Appendix 14). Proteins involved in fructose uptake (B9O19_1606 - B9O19_1608), glycolysis (B9O19_1010, B9O19_1941, and B9O19_1942), the PKP (B9O19_2009), TCA cycle (B9O19_1097, B9O19_1902, and B9O19_1903) and acetate/formate production (B9O19_89, B9O19_90, and B9O19_137) were mostly down-regulated in pectin-grown cells.

5.4 Protein ratios of supernatant proteins identified using iTRAQ

Appendix 15 shows proteins with ≥ 3 peptides identified from the secretome fractions. Because the supernatant proteome data was generated using a single sample without replicate, these results should be treated as preliminary until further confirmed with repeats. In the proteome profiles of supernatant samples, a total of 20 SLH proteins were identified from 1,648 peptides, making up 9.0 % of the total number of proteins and 59.5 % of the total peptides. Of these, 5 SLH proteins contained putative CAZyme domains (PL1 and CE8). SLH proteins were the most abundant proteins in the secretome fractions of *M. pectinilyticus*. B9O19_1156 (SLH protein) was the most abundantly represented

protein in both the pellet and the secretome samples. SLH proteins containing PL1 and CE8 domains (B9O19_909 and B9O19_1627) were also abundantly present in the secretome samples. In contrast to the pellet proteome which had no proteins identified as freely secreted CAZymes, the supernatant proteome had 341 peptides assigned to 17 CAZymes which contained secretory signal peptide sequences, but not SLH modules. Together, peptides from SLH proteins and CAZymes made up 22.2 % of the total proteins and 71.9 % of the total peptides, suggesting the peptides identified from the supernatant proteome were predominantly made of extracellular proteins. Most of SLH proteins were down-regulated in the supernatant samples from apple, citrus, and kiwifruit-pectin grown cultures, compared to the fructose control. The abundance of SLH proteins in fructose-grown cultures may have resulted from the release of these proteins into extracellular environment due to the absence of any adherable solid substrates (e.g. pectin). One notable difference between pellet and secretome proteomes was the presence of copper-amine oxidase-like domain-containing (CAO) proteins in the secretome samples. Also, although it was initially hypothesized that pectin-degrading CAZymes that lack the anchorage to the cell surface via SLH modules are freely secreted and detected in the growth medium at a greater abundance than the fructose control, the protein production of these enzymes mostly remained neutral or only slightly greater compared to the fructose control. However, the fold-change ratios for secreted CAZymes often suffered from *p*-values above the acceptable threshold (*p*-value ≤ 0.05). Compared to the pellet samples, fewer proteins were found to be differentially up-regulated in the supernatant proteomes, and the fold-change ratios mostly occurred below ~1.5. None of the up-regulated proteins were associated with SLH domains. CAZymes up-regulated in the supernatant of apple pectin-grown cultures included GH3, CE8, GH95/CE8, although these proteins were not significantly differentiated in the supernatant proteomes of citrus and kiwifruit pectin-grown cultures. Two CE4 domains encoded by B9O19_973 and B9O19_1528 were up-regulated in more than one supernatant protein preparations, in this case the secretomes from citrus and kiwifruit pectin-grown cultures. The presence of cytoplasmic proteins lacking signal peptides such as aspartate semialdehyde dehydrogenase (B9O19_775), glutamine synthetase (B9O19_1105), phosphoketolase (B9O19_2009), L-arabinose isomerase (B9O19_2161), and EF-Tu (B9O19_1271) in secretomes

suggested a small degree of cell lysis may have occurred during the sample preparation process which may have caused the leakage of some intracellular proteins into the growth medium.

Chapter 6 Ecological abundance of *Monoglobus pectinilyticus* in the human gut

6.1 Introduction

In this part of the study, the prevalence and abundance of *M. pectinilyticus* were investigated using quantitative PCR (qPCR) and metagenomic data mining approaches. It was hypothesized that people who consumed the recommended amount (>5 g/day) of pectic fibres would likely harbour *M. pectinilyticus* as part of their commensal gut microbiota. To investigate the potential correlation between the frequency of *M. pectinilyticus* and the consumption of dietary pectin as part of natural diets, the presence of *M. pectinilyticus* was examined in correlation to normal diets high/low in natural fibres without putting subjects through any dietary intervention, for example, by adding pectin supplementation to artificially modify their gut microbiota. Quantitative PCR was carried out using faecal DNA extracted from NZ participants whose dietary habits were recorded in four sets of 3-day dietary records over a period of 10 weeks, from which the daily nutritional contents were calculated. *M. pectinilyticus* was more frequently detected from NZ participants consuming diets high in fibre, vegetables, and pectin, indicating a positive correlation may exist between the presence of *M. pectinilyticus* and greater fibre consumption. The presence of *M. pectinilyticus* among the US population was also confirmed through Blast-searching the SLH protein sequences of *M. pectinilyticus* in the NIH metagenome datasets, suggesting *M. pectinilyticus* may be a frequently present gut bacterium within Western populations.

6.2 Quantification of *M. pectinilyticus* by qPCR

6.2.1 Primer design and specificity testing

M. pectinilyticus-specific MP1087F and MP1581R primers were designed and tested for potential cross-amplifications of 16S rRNA gene sequences from other species of the phylum *Firmicutes* (Appendix 16). *M. pectinilyticus* is very distantly related to most of the known human gut bacteria, thus the commonly encountered problem of primer cross-activity between phylogenetic relatives was not observed in this study. Of the 9 strains, a ~ 500 bp band was amplified only from *M.*

pectinilyticus, suggesting MP1087F and MP1581R specifically amplified PCR products with a desired amplicon size from the target organism. Virtual NCBI primer specificity test showed that MP1087F was predicted to anneal to the 16S rRNA gene sequences of marine bacterial strains of *Thalassobaculum salexigens* (NR_116122.1), *Thalassobaculum litoreum* (NR_044165.1), and *Thalassocola ureilytica* (NR_145648.1) with 100 % identity, while MP1581R was estimated to contain > 6 single-base mismatches to its closest target sequence from *Heliobacterium chlorum* (NR_025932.1), a photosynthetic soil bacterium. These results suggested that MP1087F and MP1581R showed no cross-activity towards non-target organisms, and therefore they were suitable qPCR primers for detecting the presence of *M. pectinilyticus* in a mixed bacterial population from human stool samples. In this study, Uni331F and Uni797R were used to enumerate the total bacterial number from each stool DNA sample and to construct the *E.coli* DNA standard. Uni331F and Uni797R are widely used as a broad-range primer set to determine the total bacterial load in quantitative PCR (Hopkins *et al*., 2005, Nadkarni *et al*., 2002).

6.2.2 Construction of qPCR standard curves

It was previously reported that the qPCR enumeration of slow-growing obligate anaerobes using DNA standards constructed from fast-growing bacteria led to a significant underestimation of the abundance of slow-growing strains due to their low cell viability and the fewer number of 16S rRNA operons present in their genomes (Nadkarni *et al*., 2002). Therefore, it was predicted that the use of *E. coli* DNA standard to enumerate *M. pectinilyticus* would result in a significant underestimation of the abundance of the strain in the stool DNA samples. Two separate standard curves were constructed using serially diluted genomic DNA samples prepared from known number of cells of *E. coli* and *M. pectinilyticus* collected during the stationary phase of growth (Appendix 17). Standard curves constructed by plotting $\log_{10}$ cell numbers against the crossing point (CP) values generated by Lightcycler® 480 software are shown in Appendix 18. CP values correspond to the round of PCR amplification at which the intensity of fluorescent signal released by the 5'-exonuclease activity of *Taq* polymerase reaches a threshold to exceed the background fluorescence (Nadkarni *et al*., 2002). As a low amount of DNA material typically requires a large number of PCR amplification cycles to

occur before achieving the fluorescence detection limit, CP values are inversely proportional to the amount of nucleic acids present in the samples. Overall, good quality standard curves were generated as shown by values conforming closely to form straight lines in both *M. pectinilyticus* and *E. coli* standard graphs. The small amount of bacterial DNA amplified using universal primers from the no-template-control suggested that the enzyme preparation or chemical reagents used in PCR may be the source of contamination, and 45 run cycles of amplification may have exceeded the threshold at which the PCR reaction becomes overly sensitive to the trace amount of contaminant DNA presents in reagents. As in general, *E. coli* DNA standards were used to determine the total bacterial number in the stool DNA samples, while *M. pectinilyticus* standards were used to detect the relative abundance of *M. pectinilyticus* compared to the total bacterial number.

6.2.3 Participant recruitment and dietary composition analysis

The participant information including the age, body mass index (BMI), ethnicity, gender, the amounts of fibre and pectin intakes, and the daily consumption of selected food groups (vegetables, fruit, grains, and protein) are summarized in Appendix 19.

6.2.4 The relative abundance of *M. pectinilyticus* in stool DNA samples

Participants were considered as *M. pectinilyticus*-positive when the strain was detected at a relative abundance of $>$ 0.01 % and $>$ 6.0 $\log_{10}$ cells per gram of faeces. These cut-offs were determined by plotting the % abundance of the strain against its $\log_{10}$ concentration, which formed two clearly separated groups of high- and low-*M. pectinilyticus* individuals (Appendix 20). *M. pectinilyticus* was present at detectable levels in 10 donors at a mean of 8.056 $\log_{10}$ cells per gram of faeces (Table 6.1). The relative abundance of *M. pectinilyticus* in positive individuals ranged between 0.036 – 8.399 %, with a median value of ~1.1 %. With the exceptions of donors 3, 16, 24, and 33, the total numbers of bacteria were within the range of 10.0 – 11.5 $\log_{10}$ cells per gram of faeces reported in the literature (Hopkins *et al*., 2005; Bartosch *et al*., 2004; Matsuki *et al*., 2004). In donors 3, 16, 24, and 33, the number of total bacteria was almost one order of magnitude lower compared to other participants. This may have been caused as a result of the variations in 16 rRNA gene copy numbers in different

Table 6.1 Comparison of *M. pectinilyticus* and the total bacterial populations in 1 gram of stool samples collected from 44 healthy donors determined by qPCR. Donors showing positive presence of *M. pectinilyticus* are highlighted.

Donor ID	*M. pectinilyticus* log10 ± std dev (gram faeces)	Total bacteria log10 ± std dev (gram faeces)	% abundance of *M. pectinilyticus*
1	3.863 ± 0.15	11.176 ± 0.03	0
2	3.422 ± 0.16	11.217 ± 0.01	0
3	3.696 ± 0.13	9.953 ± 0.01	0
4	3.859 ± 1.12	10.983 ± 0.00	0
5	2.837 ± 0.07	10.830 ± 0.03	0
6	3.362 ± 0.14	10.895 ± 0.03	0
7	7.791 ± 0.04	11.042 ± 0.02	0.056
8	3.930 ± 0.50	10.937 ± 0.00	0
9	7.302 ± 0.05	10.744 ± 0.01	0.036
10	4.611 ± 0.24	10.429 ± 0.03	0
11	4.749 ± 0.08	10.498 ± 0.02	0
12	5.337 ± 0.10	10.733 ± 0.01	0
13	9.303 ± 0.01	11.447 ± 0.00	0.717
14	8.772 ± 0.02	11.284 ± 0.03	0.307
15	4.283 ± 0.04	11.447 ± 0.00	0
16	6.398 ± 0.08	8.912 ± 0.04	0.305
17	4.210 ± 0.07	11.447 ± 0.00	0
18	3.927 ± 0.99	10.322 ± 0.01	0
19	3.826 ± 0.84	11.198 ± 0.01	0
20	7.814 ± 0.03	10.332 ± 0.01	0.303
21	4.127 ± 0.17	10.739 ± 0.02	0
22	3.047 ± 0.13	10.726 ± 0.02	0
23	5.329 ± 0.03	10.284 ± 0.00	0.001
24	8.517 ± 0.02	9.593 ± 0.01	8.399
25	3.462 ± 0.22	11.191 ± 0.02	0
26	3.613 ± 0.26	10.588 ± 0.03	0
27	3.423 ± 0.05	11.085 ± 0.03	0
28	4.037 ± 0.13	10.874 ± 0.07	0
29	3.849 ± 0.16	11.453 ± 0.01	0
30	3.631 ± 0.10	11.279 ± 0.02	0
31	8.878 ± 0.01	11.172 ± 0.01	0.507
32	4.702 ± 0.08	11.213 ± 0.04	0
33	3.542 ± 0.19	8.170 ± 0.04	0.002
34	3.426 ± 0.13	10.386 ± 0.01	0
35	8.494 ± 0.01	10.964 ± 0.01	0.339
36	4.643 ± 0.09	10.963 ± 0.01	0
37	2.973 ± 0.33	10.521 ± 0.15	0
38	3.603 ± 0.39	10.204 ± 0.01	0
39	4.008 ± 0.19	10.326 ± 0.01	0
40	7.288 ± 0.00	10.485 ± 0.01	0.064
41	4.785 ± 0.05	11.111 ± 0.08	0
42	2.899 ± 0.29	10.086 ± 0.02	0
43	4.379 ± 0.13	11.115 ± 0.01	0
44	4.159 ± 0.02	11.246 ± 0.04	0

bacterial species, as well as the use of a fast-growing aerobic bacterial standard to enumerate slow-growing obligate anaerobes (Nadkarni *et al*., 2002). The triplicate samples yielded results with very little statistical discrepancies, suggesting the variable ranges of the total bacteria and the cell number of *M. pectinilyticus* observed in each donor have not resulted from technical errors. Of the 10 *M. pectinilyticus*-positive donors, 8 individuals reported to consume more than the recommended 25 g per day, while donor 35 and 40 were classified into low and moderate fibre groups, respectively. Because this study used more female participants (29 females) than males (15 males), a gender-biased presence of *M. pectinilyticus* was not considered. Ethnicity was also not considered as an influencing factor due to the small sampling size. However, it was noted that all *M. pectinilyticus*-positive individuals identified themselves as European or NZ Maori, while all 12 Asian participants tested negative for the presence of the strain. Figure 6.1 compares the differences in the median intakes for fibre, vegetables, fruit, protein, grain, and pectin consumptions between the *M. pectinilyticus*-positive and -negative groups. The amount of food intake and the presence of *M. pectinilyticus* in each donor was considered to correlate with a statistical significance if the calculated *p*-values were < 0.05. In this study, a positive correlation was observed between the presence of *M. pectinilyticus* and the consumption of fibre and pectin, as the *M. pectinilyticus*-positive group showed higher median values for fibre and pectin intakes compared to the negative group. While these results may suggest *M. pectinilyticus* is more frequently present in donors consuming diets rich in fibre- and pectin, the data should be interpreted with caution due to the small sampling size of the *M. pectinilyticus*-positive group, as well as the relatively marginal *p*-values. *M. pectinilyticus*-positive individuals were also often among the highest vegetable consumers from this study. The mean value for daily vegetable intake was 1.3 servings higher for the *M. pectinilyticus*-positive group than the negative group, and the *p*-value was only slightly higher than the 0.05 cut-off. Interestingly, donor 24 who showed an unusually high % abundance of *M. pectinilyticus* consumed the largest amounts of fibre, pectin, vegetable, and grains among all participants. *M. pectinilyticus* was present in donors 14, 20, 24, and 31, who were among the highest consumers of dietary fibre, vegetable and pectin. The consumption of fruit, grain, and protein foods did not show statistically significant correlations to the presence of *M. pectinilyticus* in that the Mann-Whitney test scores significantly exceeded the 0.05 *p*-value cut-off.

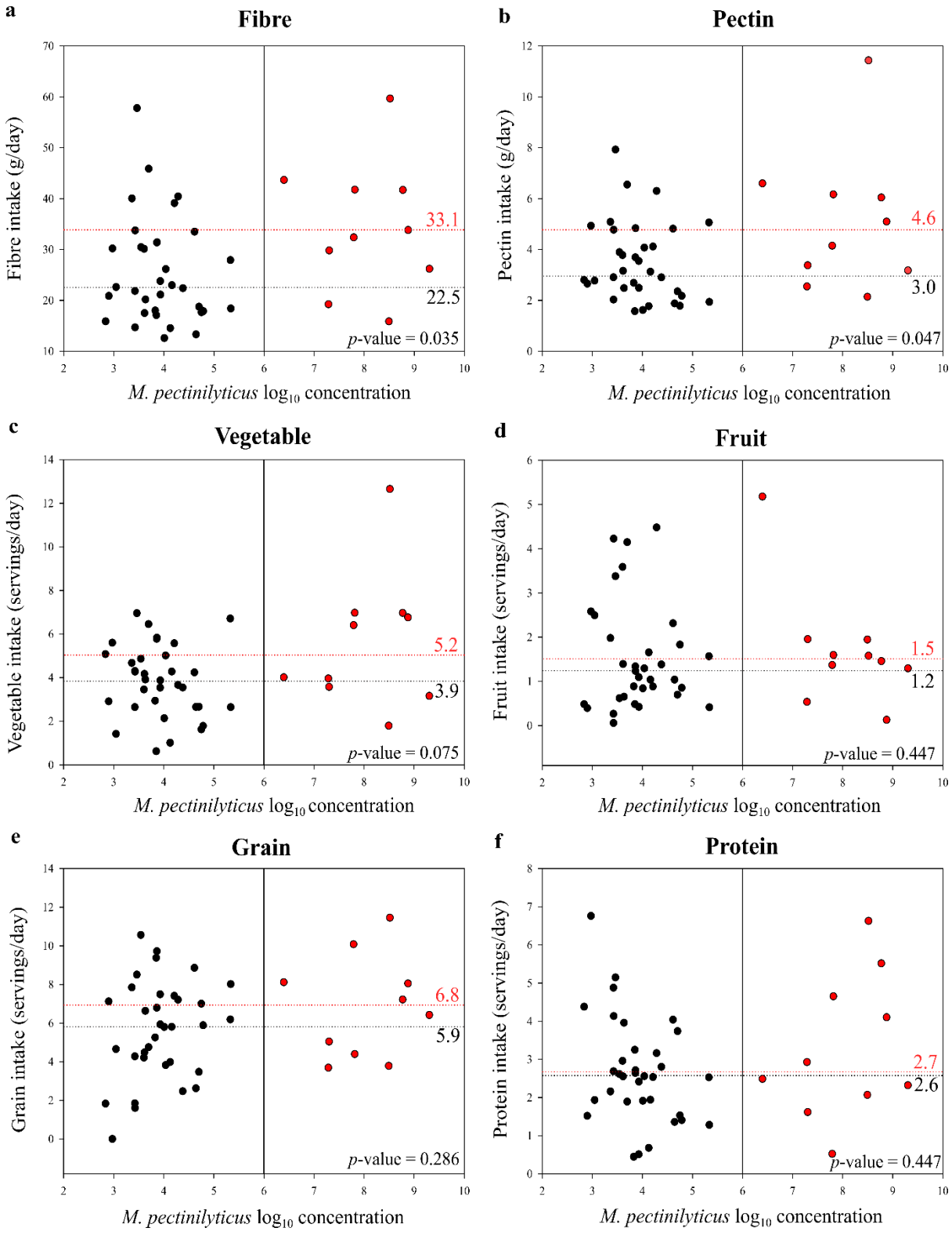


Figure 6.1 Comparison of $\log_{10}$ concentration of *M. pectinilyticus* and the consumption of different food categories (a. fibre; b. pectin; c. vegetable; d. fruit; e. grain; and f. protein) determined based on four sets of 3-day dietary record over a period of 10 weeks. Participants were considered to possess *M. pectinilyticus* if the strain's relative abundance against the total bacteria was > 0.01 % and the $\log_{10}$ concentration was > 6 (indicated by the vertical split line). *M. pectinilyticus*-positive and -negative groups were indicated in red and black dots, respectively. The median values for the intake of each food category were calculated for positive (red dotted line) and negative groups (black dotted line). The amount of food intakes and the abundance of *M. pectinilyticus* were considered significantly correlated if *p*-values calculated using a non-parametric Mann-Whitney test were < 0.05.

Despite the abundance of pectin in most fruit, *M. pectinilyticus* was not present in the majority of donors who reported to consume the recommended > 2 servings of fruit per day. The median intake values for fruit and protein showed little difference between the *M. pectinilyticus*-positive and – negative groups. The daily grain consumption for positive group was only 0.9 servings higher than the negative group. Overall, *M. pectinilyticus* was more frequently present in individuals who consumed higher-than-average amounts of fibre and pectin, suggesting that fibre-rich diet which is also naturally high in pectin may be an important nutritional factor for establishing *M. pectinilyticus* as a commensal gut microbe. *M. pectinilyticus* was detected from 36.4 % of the high fibre-consuming participants, compared to 9.1 % of low-moderate fibre group, suggesting the organism may be commonly present in the healthy-eating population in NZ.

6.3 The presence of *M. pectinilyticus* in the metagenomic datasets

SLH domain-containing protein sequences of *M. pectinilyticus* were identified from the metagenome databases of 12 out of 85 individuals living in the US, suggesting approximately 14 % of the sampled population likely possessed *M. pectinilyticus* or closely related strains (Appendix 21). The majority of SLH protein sequences retrieved from the metagenome databases were significantly shorter than the length of the original query sequences, and some individuals possessed only 2 – 3 fragmented peptides of SLH proteins. While cost-effective and fast, the nature of whole-genome shotgun sequencing (WGS) requires a random fragmentation of genomes, and the current next-generation sequencing technology can only be used to sequence fairly short DNA strands of 50 – 300 bp. Furthermore, without a reference genome, assembling short shotgun sequences together can be a difficult process which requires vast amounts of computing power and is often error-prone (Sharpton, 2014). Also, the diversity and complexity of metagenome data make it difficult to determine the genomic origins of the sequences, and most often genomes are not completely represented by the sequence reads (Sharpton, 2014). The majority of SLH proteins identified from the metagenome databases showed very high sequence similarities (98 – 100 %) to the query sequences. Some peptides contained mismatched amino acids and gaps within the alignments which may have resulted from errors during the sequence assembly, but the sequence similarity was still within the range of 76 – 95

%, which was significantly higher than the 40 – 41 % homologous sequence matches found from the GenBank protein database. Protein functions of SLH proteins could not be deduced from the metagenome datasets, as all proteins were identified as 'unnamed protein product'. Because the dietary information of the HMP participants are not available, the US nationwide dietary intake data collected by National Health and Nutrition Examination Survey (NHANES) for the years 2003 – 2006 was used to indirectly assess the dietary fibre intake in the general US population (Appendix 22). Using the fibre category cut-offs used in this study (≥ 25 g/day for females and ≥ 30 g/day for males), only the top 95th percentile of US males in the age group of 51 – 70, and pregnant females were found to consume sufficient amounts of dietary fibre to classify into high fibre group. However, due to the different methodologies used to collect the dietary information and to calculate of the amount of fibres from the food intakes, the results from this study and the NHANES survey should not be considered to be comparable to each other. Nevertheless, the US dataset showed that the majority of the NHANES survey participants were consuming low amounts of fibre, in agreement with the mean dietary fibre intake of adults living in the US which averaged 16 grams per day, approximately half the recommended amount of the daily fibre consumption (Hoy *et al*., 2014).

Chapter 7 Discussions

7.1 Isolation of a novel pectin-degrading bacterium

The first objective of this research was to isolate a novel pectin-degrading bacterium from human faeces. I hypothesized that the use of high molecular weight pectin (from kiwifruit) with a heterogeneous side chain composition would selectively enrich for bacteria that have the fibrolytic capability to systematically deconstruct the complex pectin structures, and metabolize the downstream products of the pectin degradation. Using kiwifruit pectin as an isolation substrate, a novel genus/species *Monoglobus pectinilyticus* 14^T was isolated, as described in chapter 3. Chapter 3 provides biochemical and physiological descriptions of *M. pectinilyticus*, which were performed as part of the routine procedures to formally introduce a novel bacterial species to the public. Together with the uncultured GenBank relatives, *M. pectinilyticus* formed a distinct phylogenetic lineage of descent clearly separated from the neighbouring clostridial groups within the family *Ruminococcaceae*. *M. pectinilyticus* differed from its phylogenetic relatives of the *Clostridium* clusters III and IV in terms of spore formation, motility, optimum temperature range for growth, cellular fatty acid composition, G+C content, and substrate utilization. With the exception of D-fructose, *M. pectinilyticus* was able to grow on a narrow spectrum of substrates which were either pectin or monosaccharides of pectin and hemicellulose (arabinose, xylose and galacturonic acid). Although D-glucose appeared to be fermented by *M. pectinilyticus*, almost negligible growth was supported. Based on TEM images, *M. pectinilyticus* directly adhered to the plant pectin materials, and formed fibrous aggregations on bacterial cell surfaces which suggested the presence of an extracellular substrate adhesion machinery. On the basis of the phenotypic and phylogenetic data obtained in this work, a novel genus of the family *Ruminococcaceae*, represented by the type species *M. pectinilyticus* gen. nov., sp. nov. was proposed. The widespread presence of *M. pectinilyticus* and the uncultured GenBank relatives from humans, ruminant, and non-ruminant animals suggested that the members of this lineage were probably frequent terrestrial gut commensals. *M. pectinilyticus* and its relatives extended a novel phylogenetic branch within the family *Ruminococcaceae*. Discovering a

novel phylogenetic lineage within the order *Clostridiales* is not uncommon due to the immensity of the taxon. A recent example is the isolation of *Lutispora thermophila* which is considered as a sole cultured representative of a novel *Clostridium* cluster interposing between *Clostridium* cluster II and III (Shiratori *et al.,* 2008). The taxonomic disarray within the order *Clostridiales* was caused by the traditional classification of Gram positive, anaerobic, and endospore-forming organisms into a single genus, *Clostridium*, primarily based on phenotypic characteristics. To address this issue, *Bergey's manual of systematics of archaea and bacteria* has suggested extensive restructuring and renaming of the taxon may take place in the near future (De Vos *et al*., 2009). The family *Ruminococcaceae* is best known for its fibrolytic roles to break down complex plant polysaccharides in anaerobic environments. The majority of cellulose-digesting bacteria isolated from the animal rumen (*Ruminococcus flavefaciens* and *Ruminococcus albus*), the human large intestine (*Ruminococcus champanellensis*), and anaerobic agricultural waste fermenters (*Clostridium aldrichii*, *Clostridium cellulolyticum, Clostridium stercorarium, Clostridium straminisolvens, Clostridium thermocellum, Clostridium clariflavum*, and *Acetivibrio cellulolyticus*) belong to this taxon. The few known species of *Ruminococcaceae* inhabiting the human gut play unique and important functions in assisting the digestion of plant materials and maintaining the gut health. For example, *R. champanellensis* is so far the only human colonic bacterium known to be capable of degrading a crystalline form of cellulose using the cellulosome system (David *et al.,* 2015; Chassard *et al.,* 2012). *R. bromii* plays a key role in breaking down and creating energy from resistant starch which reaches the gut undigested (Ze *et al.,* 2012). *F. prausnitzii* is one of the most abundantly present butyrate-producing bacterium in the human gut, and it is an indicator of healthy gut as the reduction of this species is often associated with inflammatory bowel disease and Crohn's disease (Lopez-Siles *et al.,* 2012). For this reason, a large number of *Ruminococcaceae* was included in the 'most wanted' priority list for cultivation and whole genome sequencing from the human microbiome (Fodor *et al.,* 2012). Despite the abundance of the family *Ruminococcaceae* within the human gut microbiota, relatively few bacterial species of this taxon have been isolated and cultivated. The lack of cultured representatives of the family *Ruminococcaceae* from the human gut may be attributed to the difficulty of growing fastidious strict anaerobes, as well as the recent neglect of microbial culture techniques in favour of next-generation

sequencing approaches (Lagier *et al.*, 2016). Even with the advent of single cell genomics, isolating microorganisms in pure cultures remains crucial for producing good quality genomes and investigating microbial functions.

M. pectinilyticus has a unique metabolic capacity in that the use of uronic acid was traditionally regarded as an unusual trait for human colonic bacteria, with the exception of *Bacteroides* spp. (Salyers *et al.*, 1977b). Salyers (1977b) showed that 47 % of 188 isolates of *Bacteroides* spp. surveyed tested positive for pectin degradation, and 72 % tested positive for galacturonate utilization. *Eubacterium eligens*, one of the few Gram positive organisms known to utilize pectin in the human colon, failed to grow using galacturonate monomer (Salyers *et al.*, 1977a). The author suggested the metabolic pathway for uronic acid utilization may be largely absent outside the *Bacteroidetes*. So far, *M. pectinilyticus* and *Faecalibacterium prausnitzii* are the only members of the family *Ruminococcaceae* known to degrade large molecular weight pectin and to utilize uronic acid as a sole carbon source, although the pectinolytic activity of *F. prausnitzii* was reported to be limited to apple pectin, as no growth occurred using polygalacturonate or pectin from citrus peels (Salyers *et al.*, 1977; Lopez-Siles *et al.*, 2012). If the search was expanded to the realm of the rumen microbiome, *Lachnospira multipara* and *Butyrivibrio fibrisolvens* of the family *Lachnospiraceae* have shown to degrade pectin, although only the latter was able to utilize galacturonic acid monomers for growth (Hespell, 1992; Marounek *et al.*, 1999; Paggi *et al.*, 2005). Therefore, the ability to utilize uronic acid was previously seen as a rare trait among the *Firmicutes*.

The presence of extracellular adhesion machinery on the surface of Gram positive organisms is considered as a necessary prerequisite to facilitate the adhesion-based colonization of the plant cell wall against the fluid nature of intestinal/ruminal contents (Cheng *et al.*, 1979). In mixed microbial communities from the animal rumen, the sequential degradation of the plant cell wall was initiated by recruiting Gram positive bacteria such as *Ruminococcus* spp. and *L. multipara* to the sites of the degradation (Cheng *et al.*, 1980; Cheng *et al.*, 1979; Cheng *et al.*, 1984). *M. pectinilyticus* shared some phenotypic and phylogenetic characteristics with *Ruminococcus* spp. in that both were non-motile Gram positive cocci of the family *Ruminococcaceae*. However, *M. pectinilyticus* possessed

thicker cell wall structures, and lacked the extracellular glycoprotein coat which serves as a continuous biofilm to assist the attachment of *Ruminococcus* spp. to the surface of the plant cell wall (Patterson *et al.,* 1975; Latham *et al.,* 1978). The ability to utilize pectin as a primary growth substrate is also an unusual metabolic trait among its mostly cellulolytic neighbours of the *Ruminococcaceae*. The differences between *M. pectinilyticus* and *Ruminococcus* spp. were further highlighted in the G+C content, and the types of major cellular fatty acids. The G+C content of *M. pectinilyticus* was close to the reported % G+C ranges for *A. cellulolyticus* NRC 2248^T (38 %), *C. thermocellum* YS (39 %), *C. clariflavum* DSM 19732^T (36.9 %), and *C. stercorarium* DSM 8532^T (39 %) (Patel *et al.,* 1980; Brown *et al.,* 2012; Shiratori *et al.,* 2009; Madden, 1983). DNA G+C content of *M. pectinilyticus* was lower compared to *Ruminococcus flavefaciens* C94^T (39 – 44 %), *Ruminococcus albus* ATCC 27210^T (42.6 – 45.8 %), *R. champanellensis* 18P13^T (53.05 %), and *R. bromii* ATCC 27255^T (41.05 %) (De Vos *et al.,* 2009; Chassard *et al.,* 2012). The cellular fatty acid profile of *M. pectinilyticus* was unusual in that $C_{18:1}$ ω9c, an 18C-straight monounsaturated fatty acid, seldom occurs as a major cellular fatty acid in its close relatives. Previous findings showed predominantly 15C- or 16C-rich fatty acid profiles for *R. thermocellum*, *C. clariflavum*, *C. stercorarium*, *R. flavefaciens*, *R. albus*, and *R. champanellensis* (Ifkovits and Reghab, 1968; Shiratori *et al.,* 2009; Timmons *et al.,* 2009; Chassard *et al.,* 2012). 18C-fatty acids were either absent or produced as minor fatty acids in these organisms. Although 18C-fatty acids made up 39 % of the total fatty acid in *M. pectinilyticus*, 15C-, 16C-, and 17C-fatty acids were also present in significant amounts. Despite these differences, the tendency of *M. pectinilyticus* to associate in proximity to the plant cell wall, and to form fibrous aggregations onto the bacterial cell surfaces resembled the fibrolytic actions observed from the cellulosome-producing strains of *Ruminococcus* spp. and the pectin-degrading *L. multipara*. Our results from TEM and SEC suggested that an extracellular adhesion machinery may be present on the cell surface of *M. pectinilyticus*, which may enable the strain to directly engage in the degradation of large molecular weight pectin materials. To our knowledge, *M. pectinilyticus* is the first human gut bacterium to display the capacity to carry out the extracellular degradation of pectin in a potentially systematic manner, as the potential constituents of extracellular adhesion machineries were not reported from the genomes of *F. prausnitzii*, *E. eligens*, and *Bacteroides* spp.

It is unclear if the use of manually prepared kiwifruit pectin has contributed to the isolation of a novel bacterium which has successfully evaded laboratory cultivation for a long time. Regardless of the source of pectin, preserving the structural integrity and complexity of pectin should be an important factor for consideration when preparing enrichment substrates for isolating bacteria with specialized nutritional requirements for pectin. The use of harsh chemicals and thermal sterilization should be avoided to minimize the destruction of the pectin structures.

From these results, it was hypothesized that *M. pectinilyticus* may have adopted a distinctive microbial carbohydrate degradation strategy in order to accommodate the pectin-rich diet of humans. *M. pectinilyticus* was selected for further genomic characterizations based on its phylogenetic novelty, pectinolytic ability, and the potential presence of extracellular adhesion machineries.

7.2 Genomic characterization suggests *M. pectinilyticus* is a pectin-degrading specialist

The second objective of this thesis was to provide a general overview of the genome of *M. pectinilyticus*. Chapter 4 describes the *de novo* assembly and annotation of *M. pectinilyticus* genome which was then analysed to elucidate the general genomic features, and properties underlying the pectinolytic capability of the strain. Based on the phenotypic and phylogenetic similarities shared between *M. pectinilyticus* and the cellulosome-producing organisms, it was initially hypothesized that the pectinolytic ability of *M. pectinilyticus* may have resulted from the presence of pectin-degrading CAZymes incorporated into a cellulosome-like system (‘pectinosome’). From this study, a ~2.76 Mb draft genome of *M. pectinilyticus* was produced. Using the CAZy database, a moderately sized glycobiome of *M. pectinilyticus* was identified, which encodes a disproportionally large number of putative pectinolytic enzymes accounting for ~42 % (47 out of 108 CAZymes) of the total glycobiome. The strain appeared to take a ‘quality-not-quantity’ approach by directing a significant proportion of its fibrolytic potential to focus solely on the pectin degradation, thereby compensating for its less elaborate glycobiome in comparison with cellulosome-producing bacteria (>200 CAZymes); or versatile PUL-possessing *Bacteroides* spp. (>400 CAZymes), which comprise the two most well studied carbohydrate degrading paradigms. To our knowledge, *M. pectinilyticus* is the only bacterium known so far to possess glycobiome and metabolic capacity shaped for serving a specific

purpose of pectin degradation and utilization. A high degree of genomic and nutritional specialization for the utilization of a single type of polysaccharide is an unusual trait among gut bacteria, and considered as a hallmark for a keystone bacterial species (Flint *et al*., 2012). *M. pectinilyticus* differs from other pectin-utilizing gut bacteria in that its glycobiome encodes PL- and CE-rich CAZyme profiles which contrasts with the GH-rich compositions usually observed in most gut bacteria, suggesting the strain has probably adopted unique carbohydrate utilization strategies and environmental niches to harness pectin as a primary growth substrate. *M. pectinilyticus* encoded PL1s and CE8s in significantly greater numbers than most other pectin degraders, but produced few CAZyme families targeting RG-I and RG-II, suggesting its primary target was the structural components of the pectin backbone rather than the side chains. The majority of PLs and CEs contained secretory signal peptide sequences, indicating these enzymes were mostly associated with extracellular activities. Abundant production of pectate lyases may confer several competitive advantages to the strain. For example, the unsaturated oligomeric by-products formed as a result of β-elimination cleavage are further processed to monomers by few enzyme families such as PL22 and GH105 that are often absent in secondary degraders (e.g. *Bifidobacterium* spp.). Furthermore, uronic acids are utilized through the intermediates of the modified ED pathway which occurs uncommonly in most anaerobes (Richard and Hilditch 2009; Flamholz *et al*., 2013). As a result, *M. pectinilyticus* would not need to share the oligomeric/monomeric products from degraded pectin with most of the neighbouring gut bacteria that are less adapted for pectin utilization. The niche specialization of *M. pectinilyticus* in proximity to the plant cell wall and the middle lamella may be ideal for providing divalent cations (Ca^{2+} or Mg^{2+}) required for most pectate lyases activities, as these ions are localized at high concentrations in plant tissues (Herron *et al*., 2000). The abundance of CE8 and CE12 would facilitate a rapid de-esterification of methyl- and acetyl-moieties from pectin to allow an efficient access of PLs to the HG and RG backbone structures. *M. pectinilyticus* possessed few putative CAZyme families (GH33, GH43, GH51, GH97, and GH140) targeting the side chain components of RG-I and RG-II, and most of these enzymes lacked secretory signal peptide sequences. Therefore, a partial degradation of RG-I and RG-II side chains may occur, but mostly as part of an intracellular downstream processing of substituted pectin oligomers. Although the majority of PLs and CEs were

predicted with extracellular pectin-degrading activities, only 1 PL1 (B9O19_51) contained identifiable CBM13 domains. The few families of CBMs identified in *M. pectinilyticus* mostly had predicted binding functions to cellulose, glycogen, and the bacterial cell wall polysaccharides. CBMs are usually found within the non-catalytic regions adhered to CAZyme domains, and are considered to play a pivotal role in mediating the interaction between the catalytic domains and the substrates. The lack of identifiable CBMs associated with CAZymes produced by *M. pectinilyticus* may reflect the scarcity of pectin-targeting CBMs which have been functionally characterized to date (Cid *et al*., 2010). Due to the difficulty of experimental validation, currently available CAZy catalogue of CBMs is far from complete, and active research is being carried out to discover and characterize novel types of CBMs. Therefore, the large uncharacterized regions of CAZymes merit further investigation as they may constitute novel types of CBMs or substrate-binding domains specifically targeting the components of pectin. Several PL1s produced in *M. pectinilyticus* showed the signs of recent duplication events, and the majority of PL1s were only distantly related to PL1s from soil bacteria. It is currently unclear if some of the genes coding for these PL1s are acquired through the lateral gene transfer, and will require further investigation in future. Horizontal transfer of β-porphyranases from a marine bacterium (*Zobellia galactanivorans*) to seaweed glycan-degrading *Bacteroides plebeius* from Japanese gut microbiota was previously demonstrated, suggesting the contact with environmental bacteria is an important driver of the CAZyme evolution of the human gut microbe (Hehemann *et al*., 2010; Hehemann *et al*., 2012).

M. pectinilyticus also produced putative hemicellulases and cellulases, presumably involved in the disintegration of the polysaccharides surrounding the pectin in the plant cell wall matrix. *M. pectinilyticus* lacked dockerin domains associated with CAZymes, thus ruling out the initial hypothesis that the strain relied on the dockerin-cohesin interactions to integrate its enzymes into a cohesive multi-protein complex. Interestingly, *M. pectinilyticus* has several genetic remnants potentially associated with a clostridial ancestor, but some of these phenotypes appear to have been either lost (e.g. endospore formation and rod-shaped cell morphology) or changed in order to adapt to the human gut environment. The most notable suggestion of adaptation is the presumed pectin-

binding mediated through the SLH protein machinery, which to date has been exclusively found in lignocellulolytic systems of various environmental isolates (Table 7.1). SLH modules are typically 50 – 60 amino acids long, and the SLH motifs from Gram positive organisms have demonstrated binding affinity for pyruvylated bacterial cell wall (Sara, 2001). SLH proteins have been implicated in the attachment and deconstruction of the lignocellulose biomass in several *Firmicutes* organisms. For example, the SLH-mediated protein anchoring motif is a highly conserved characteristic among the members of the *Clostridium* cluster III such as *C. thermocellum, A. cellulolyticus* (Xu *et al*., 2003), and *C. clariflavum* which use SLH modules to attach the cellulosome assembly onto the bacterial cell surface (Artzi *et al*., 2015). In *C. thermocellum* cellulosomes, the anchoring scaffoldins SdbA, OlpB, and Orf2p are tethered to the cell surface via three segments of SLH modules (Fontes and Gilbert, 2010). Recently, novel types of lignocellulolytic degradation systems consisting of large uncharacterized SLH proteins were described in *Paenibacillus curdlanolyticus* B-6 and *Caldicellulosiruptor* spp. (Conway *et al*., 2016; Ratanakhanokchai *et al*., 2013). *P. curdlanolyticus* is a facultative anaerobic bacterium isolated from a pineapple waste digester. *P. curdlanolyticus* is known to produce a 1,450 kDa extracellular multi-enzyme complex consisting of 11 protein subunits with xylanase activities (Ratanakhanokchai *et al*., 2013). The genus *Caldicellulosiruptor* consists of extremely thermophilic anaerobic bacteria that can attach to, and degrade lignocellulolytic polysaccharides (Conway *et al*., 2016). *Caldicellulosiruptor kronotskyensis* produces up to 19 catalytic and non-catalytic SLH proteins, although it is currently unknown if these proteins form subunits of multi-protein complexes (Conway *et al*., 2016). The genes coding for non-catalytic SLH proteins of *C. kronotskyensis* were reported to show a moderate-high level of transcription when the strain was grown on crystalline cellulose and switchgrass (Conway *et al*., 2016). In *Caldicellulosiruptor saccharolyticus*, the largest ORF (2,593 amino acids) of the genome was encoded by Csac_2722, a non-catalytic SLH protein that showed binding affinity to Avicel cellulose (Ozdemir *et al*., 2012). *M. pectinilyticus* encodes ~42 extracellular proteins with SLH modules. Many of these SLH proteins had similar protein domain architectures to the structural protein components of the S-layer in *Bacillus anthracis* in that the tandemly arranged SLH repeats at either N- or C-

Table 7.1 List of bacterial species that produce CAZyme domain-containing proteins with SLH modules for presumed peptidoglycan anchorage.

Species	Strain	Source	O_2 tolerance	Temperature requirement	# of SLH proteins
Caldanaerobius polysaccharolyticus	DSM 13641	Organic waste	Anaerobic	Thermophile	NA
Caldicellulosiruptor bescii	DSM 6725	Thermal spring	Anaerobic	Thermophile	12
Caldicellulosiruptor kronotskyensis	DSM 18902	Thermal spring	Anaerobic	Thermophile	19
Clostridium cellulolyticum	H10	Compost	Anaerobic	Mesophile	26
Clostridium clariflavum	DSM19732	Bioreactor	Anaerobic	Thermophile	31
Clostridium stercorarium	DSM 3532	Compost	Anaerobic	Thermophile	15
Clostridium thermocellum	ATCC27405	Compost	Anaerobic	Thermophile	25
Monoglobus pectinilyticus	DSM 104782	Human faeces	Anaerobic	Mesophile	42
Paenibacillus sp.	JDR-2	Soil	Facultative anaerobic	Mesophile	78
Thermoanaerobacterium saccharolyticum	DSM 8691	Thermal spring	Anaerobic	Thermophile	12

terminal of the proteins were followed by domains encoding β-lactamase, peptidoglycan amidases, peptidases, and domains of unknown functions (DUFs) (Fagan and Fairweather, 2014). However, some SLH proteins contained CAZyme domains, cohesin- and dockerin-like domains, and fibronectin type III (Fn3), indicating these proteins may play supporting roles in carbohydrate-binding and modulating the catalytic activities of CAZymes. *M. pectinilyticus* produces 8 pectin-degrading CAZymes directly adhered to the bacterial cell surface via SLH modules. Many lignocellulose-degrading clostridial bacteria are known to produce similar multi-domain proteins consisting of various combinations of CAZymes, SLH modules, CBMs, Fn3, and bacterial immunoglobulin (Big)-like domains. The CAZymes containing SLH modules have been characterized from *Paenibacillus* spp., *Caldicellulosiruptor* spp., *Clostridium* spp., *Bacillus* spp., *Caldanaerobius polysaccharilyticus*, and *Thermoanaerobacterium* spp. (Adelsberger *et al*., 2004; Ali *et al*., 1999; Ali *et al*., 2001; Ali *et al*., 2005; Brechtel *et al*., 1999; Cheng *et al*., 2009; Conway *et al*., 2016; Feng *et al*., 2000; Fuchs *et al*., 2003; Han *et al*., 2010; Han *et al*., 2012; Hung *et al*., 2011; Huang *et al*., 2015; Ito *et al*., 2003; Itoh *et al*., 2014; Kim *et al*., 2000; Lee *et al*., 1994; Liu *et al*., 1996; Matuschek *et al*., 1994; Morris *et al*., 1999; Shao *et al*., 1995; St John *et al*., 2006; St John *et al*., 2012; Waeonukul *et al*., 2009). In *Caldicellulosiruptor* spp., the uncharacterized extensions in catalytic SLH proteins were presumed to function as a linker to tether the CAZyme domains to the cell surface, while swinging out catalytic domains to reach, and attach to the biomass substrates (Conway *et al*., 2016). Although the non-catalytic regions of CAZymes are thought to play supporting roles in modulating the catalytic activities, very little is known about the functionally ambiguous modules such as Fn3, Big, and repeat domains. Fn3 domains have previously demonstrated binding affinity to insoluble lignocellulose biomass (Ozdemir *et al*., 2012). Fn3 found in the cellulobiohydrolase CbhA (GH9) of *R. thermocellum* showed weak binding to acid-swollen cellulose, and significantly enhanced the hydrolytic activities of GH9 by loosening up the surface of cellulose fibres (Kataeva *et al*., 2002). Big-like domains are commonly found in bacterial proteins involved in cell-cell adhesion and extracellular carbohydrate hydrolysis (Fraser *et al*., 2006). So far, *M. pectinilyticus* is the only human gut bacterium known to anchor fibrolytic enzymes to the bacterial cell surface via S-layer domains, suggesting SLH proteins potentially represented elements of a novel carbohydrate degradation

strategy which was previously unknown in the human gut. While these data suggest that *M. pectinilyticus* may long ago have an environmental origin, the history of pectin consumption by humans may have exerted nutritional pressure to drive *M. pectinilyticus* evolution towards utilizing pectin - one of the most readily available plant glycans - from the human colon.

In accordance with the previous observations from chapter 3, the KEGG analysis of *M. pectinilyticus* genome revealed metabolic pathways for utilizing uronic acids, D-xylose, L-arabinose, and D-fructose. Taken together, these results suggested that *M. pectinilyticus* is a true pectinolytic primary degrader with a high degree of genomic and metabolic specialization for utilizing naturally occurring pectin in plant. While ED and EMP pathways share the lower glycolysis pathway, the upper portion of reactions differ in that a single phosphorylation event occurs in ED pathway to form one pyruvate and one glyceraldehyde-3-phosphate, whereas the monomeric substrates are phosphorylated twice in EMP pathway, forming glyceraldehyde-3-phosphate and dihydroxyacetone phosphate (Figure 4.12 & Figure 7.1). Both triose-phosphates of EMP pathway produce ATP through substrate-level phosphorylation (Flamholz *et al*., 2013). As a result, EMP pathway yields twice more ATPs than ED pathway, which has only one glyceraldehyde-3-phsophate to rely on for ATP production. The ED pathway-dependent ATP production is therefore thermodynamically unfavourable, and rarely occurs in anaerobes which must rely on glycolysis to generate ATP in the absence of aerobic respiration (Flamholz *et al*., 2013). The use of glycolysis and pentose phosphate pathways requires NAD^+ to serve as the reducing equivalents to take away H^+ from the highly reduced forms of hexose and pentose sugars (Macfarlane and Macfarlane 2003). Eventually, cells face the need to recycle NADH to NAD^+ to sustain the metabolic activities, and this is achieved by placing NADH to partially oxidized metabolic intermediates such as lactate, succinate, and ethanol, which serve as electron sinks and eventually become excreted into the environment as fermentation waste products (Wolfe *et al*., 2005). A balanced production of acetate and ethanol is an important regulator of the redox reactions in the anaerobic environment in that acetyl-CoA, which is produced by the decarboxylation of pyruvate, can either become converted into energy-yielding ATP via acetate synthesis or become reduced to an electron sink ethanol (Wolfe *et al*., 2005).

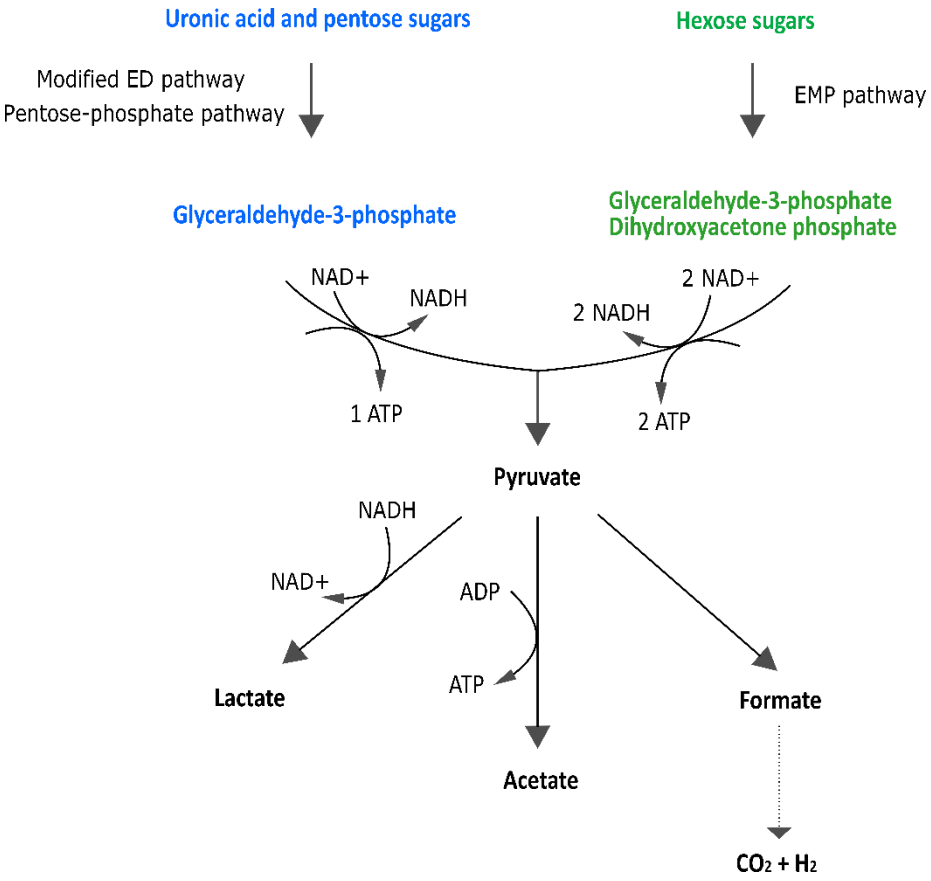


Figure 7.1 Simplified diagram of sugar utilization and the production of acetate, formate, lactate, CO_2 and H_2 in *M. pectinilyticus*.

For *M. pectinilyticus*, recycling reducing equivalents is not a significant process as uronic acids are highly oxidized form of sugars, and therefore the cells do not need to produce ethanol as uronic acid is already redox-balanced. Instead of producing electron sinks, *M. pectinilyticus* may direct pyruvate for producing large amount of acetate, thus producing ATP to compensate for the energy loss occurring in the ED pathway-mediated glycolysis. In chapter 3, pectin- or galacturonic acid-grown *M. pectinilyticus* has shown to produce the largest amount of acetate and almost negligible amounts of formate and lactate, reflecting its metabolic needs for additional ATP production via acetate production. Formate production was only noticeable when the strain was grown using D-fructose as a sole carbon source, suggesting the full glycolytic pathway is in use to oxidize D-fructose to pyruvate, shunting additional pyruvate into formate production. Although metabolic alternatives such as the ED pathway and the phosphoketolase pathway produce less ATPs than their canonical counterparts (EMP and pentose phosphate pathway), their metabolic pathways tend to be more streamlined, hence likely

incur less protein costs. For example, in *Escherichia coli*, 3.5-fold less enzyme mass is needed to catabolize glucose through the ED pathway than the EMP pathway to achieve the same carbon flux (Flamholz *et al*., 2013). In the case of PKP, there are fewer enzymatic steps (8 versus 13) involved in pentose metabolism than PPP (Liu *et al*., 2012), possibly reducing the amount of energy and biomass required for the same carbon flux. The energy saving strategies of *M. pectinilyticus* may illustrate how this bacterium matches its metabolic capacity to its specialized glycobiome to meet the thermodynamic demands of surviving on low energy-yielding substrates.

Due to the lack of isolates to study, a 'missing link' remained in our understanding of the systematic deconstruction of the plant cell wall, and the contribution of pectin degradation to the colonic digestive functions has not been fully appreciated. For the first time, I have obtained a high quality genome of a human gut bacterium which has evolved a glycobiome with a single-substrate specificity for pectin. The initial glimpse into the glycobiome and metabolic capability of *M. pectinilyticus* implied a unique niches specialization by *M. pectinilyticus*, which is mediated by producing cell surface proteins anchored to the S-layer to form a physical adherence to the plant cell wall and the middle lamella. From these results, it was hypothesized that these proteins may constitute elements of a novel pectin degradation system, and the expression of pectin-degrading CAZymes and SLH proteins may be regulated depending on the availability of pectin.

7.3 Proteomics supports putative extracellular pectin degradation

In this part of study, a labelling-dependent iTRAQ protein quantification method was used to measure the production of whole-cell and secreted proteins of *M. pectinilyticus* in response to the availability of pectin in the growth medium. iTRAQ is a widely used shotgun proteomic methodology which relies on peptide labelling for quantitative protein comparison across multiple samples using a combination of liquid and mass spectrometry. The robust isobaric tags recognize the N-terminus and the ε side chains of lysines, thus ensuring every fragmented peptide generates signal intensity which can be used to quantify the relative abundance of proteins that are differentially expressed under specific conditions (Evans *et al*., 2012). The ability of iTRAQ to simultaneously prepare and analyse multiple samples in a single experiment to reduce the potential variations occurring during sample

processing and data acquisition was considered as an attractive approach to compare the proteome profiles of *M. pectinilyticus* cells grown using pectin from different sources (apple pomace, orange peels, and kiwifruit pulps) against that of fructose control. However, similar to other tools of quantitative proteomics such as ICAT, SILAC, and label-free methods, iTRAQ technology is not without its limits (Ow *et al*., 2009; Karp *et al*., 2010). Underestimation of fold-change ratios arising from factors such as isotopic impurities, ion suppression, and co-fragmentation of peptides with similar mass-to-charge ratio values (*m/z*) which ends up contributing to background signals is a well-recognized issue affecting the reliability of iTRAQ quantification (Evans *et al*., 2012; Pascovici *et al*., 2015). Furthermore, the number of proteins identified in the iTRAQ proteome tends to be significantly less than the genome-based prediction of the total number of proteins (Berghoff *et al*., 2013). Notwithstanding the limitations, the total number of proteins identified, and the number of differentially produced proteins observed in this study were within comparable ranges reported from similar studies in which iTRAQ labelling was used to assess proteomic profiles of cellulosome-producing *Clostridium termidis* growing on various substrates, biofilm-producing *Staphylococcus epidermidis,* and *Escherichia coli* incubated with antimicrobial peptides (Munir *et al*., 2016; Carvalhais *et al*., 2015; Zhou and Chen 2011).

The SLH protein encoded by B9O19_1156 was represented by the highest number of peptides from both whole-cell and supernatant proteomes. B9O19_1156 was located in proximity to the two genes (B9O19_1158 and B9O19_1163) annotated as GT2 which has been characterized with poly-specific functions including cellulose synthase, chitin synthase, β-1,3-glucan synthase, chondroitin polymerase, N-acetylglucosaminyltransferase, N-acetylgalactosaminyltransferase, and hyaluronan synthase in various bacteria (the CAZy database). Since N-linked glycan moieties are often found to decorate the S-layer glycoproteins, the gene product of B9O19_1156 is presumably involved in the formation of S-layer structure to cover the cell surface of *M. pectinilyticus* with a protein coat. S-layer proteins have been reported to be the most abundantly expressed proteins in S-layer-producing bacteria such as *Caldicellulosiruptor* spp. *Clostridium phytofermentans,* and *Clostridium termidis* (Blumer-Schuette *et al*., 2010; Muddiman *et al*., 2010; Tolonen *et al*., 2011; Munir, 2014). S-layer

proteins are known to self-assemble, providing attachment scaffolds to extracellular enzymes, and also maintaining close proximity with the substrates (Sara and Sleytr, 1987; Sleytr and Messner, 1988; Engelhardt, 2007). The combination of a genome-based functional prediction for proteins and the proteomic profiling of cells grown in the presence and absence of pectin suggested that extracellular proteins which were implicated with structural homology to cellulosomal components, such as the SLH domain-mediated protein anchorage and cohesin- and dockerin-like domains, may play important roles in the process of pectin degradation. It was therefore interesting to observe that the proteins showing elevated production in the presence of apple and kiwifruit pectin, often displayed a different pattern of expression in the presence of citrus pectin. The main differences between citrus pectin and the two other types of pectin (apple and kiwifruit) used in this study would be the differing neutral sugar compositions, the degree of esterification, and the amount of starch contaminations. The sugar compositional analysis of in-house prepared kiwifruit pectin and commercial citrus pectin showed that kiwifruit pectin possessed a greater amount of total neutral sugar and a higher degree of heterogeneity in the sugar composition compared to citrus pectin (section 3.2). Similarly, apple pectin has been reported to possess longer RG-I regions and higher neutral sugar contents than citrus pectin (Despres *et al*., 2016). Likewise, the high glucose content detected in the in-house kiwifruit pectin preparation and the previously reported co-extraction of starch with apple pectin suggested that kiwifruit and apple pectin were more heavily affected by starch contamination compared to citrus pectin which is traditionally prepared using orange peels rather than fruit pulps (Despres *et al*., 2016). Although apple, citrus, and kiwifruit pectins are all known to be highly methylated, citrus pectin is the least methylated of three (section 1.2.1). Therefore, in order to accommodate a highly complex structure of apple and kiwifruit pectin, an increased anchoring requirement for a greater number of SLH domain-containing proteins may be required. While it is possible that the highly refined structure of citrus pectin may have caused *M. pectinilyticus* to develop a simpler proteome profile than apple and kiwifruit-grown cells, it remained unclear why the expression levels of some proteins potentially involved in citrus pectin degradation were lower compared to the fructose control. Future investigation should address the variability in pectin structures, as well as the reproducibility and

stability of proteome data by including a larger number of replicates, and possibly by performing a parallel analysis with transcriptomic data.

During the stationary phase, fructose-grown cells appeared to maintain a higher level of production of basic metabolic proteins than pectin-grown cells, possibly due to the need for an increased amount of protein production for producing glycolytic intermediate enzymes for processing a hexose sugar. As a result, protein intermediates of the energy production and redox reactions (e.g. glycolysis, TCA cycle, pyruvate metabolism, and SCFA production) were generally down-regulated when cells were grown on pectins. The proteins involved in the biosynthesis of amino acids, cofactors, and nucleic acid, sugar transport, and antibiotic resistance also showed a reduced level of production, while the intermediate protein for storage starch production increased its production, suggesting the cells grown using pectin had depleted the available nutrient, and entered the stationary phase at the time of sample collection. During the stationary phase of growth, a reduction in growth rate, protein synthesis, nucleic acid production, and the transport and metabolism of carbohydrate has been observed from nutrient-starved Gram negative cells (Llorens *et al*., 2010). Once the nutrient becomes exhausted, the carbon flow is known to reverse from glycolysis to activate the production of storage compounds through gluconeogenesis before shutting down unused metabolic pathways such as the pentose-phosphate pathway (Shimizu *et al*., 2014). At the same time, cells entering the stationary growth phase increase the production of cell envelope proteins, stress-related chaperons, and antibiotic resistance-conferring proteins to effectively protect the dormant cells (Llorens *et al*., 2010).

One interesting observation was that cytoplasmic proteins lacking a signal peptide were often detected in the supernatant proteome. Called ‘moon-lighting proteins’, these cytoplasmic proteins found on the bacterial cell surface are known to be secreted using an unconventional protein secretion mechanism, and are predicted to have functions in cell invasion or adhesion (Yu *et al*., 2012; Esaka *et al*., 2015). Glycolytic enzymes, chaperones, and translational factors are known to be frequently found in association with the bacterial cell wall, suggesting the proteins present in the supernatant proteome were not likely to have resulted from extensive cell lysis (Yu *et al*., 2012).

Overall, the supernatant data provided an exoproteome profile of *M. pectinilyticus* which identified several bacterial cell wall-coating SLH and non-SLH proteins that were not detected from the pellet proteome. One such example was the secreted proteins containing an N-terminal homology domain of copper amine oxidase (CAO). Proteins containing CAO N-terminal domains have been previously reported from clostridial genomes including *Clostridium thermocellum, Clostridium straminisolvens,* and *Acetivibrio cellulolyticus*. CAOs are widely present in both prokaryotes and eukaryotes, and are known to carry out the oxidation of biological amines, enabling the organisms to utilize amines as the source of carbon or nitrogen. CAOs are homodimers which require one copper ion and an organic cofactor such as the redox cofactor, topaquinone (TPQ), per subunit for activity (Parsons *et al*., 1995). The mushroom-shaped dimers are hoisted up by a 85-amino acid N-terminal stalk domain consisting of five-stranded anti-parallel β-sheets twisting around an α-helix (Parsons *et al*., 1995). Therefore, the CAO N-terminal-like domains in *M. pectinilyticus* may function as a structural support to hoist extracellular proteins on the bacterial cell wall. Copper-containing oxidases are widely found in all life forms, although the exact functions of these proteins are largely unclear. Recently, lytic polysaccharide monooxygenases (LPMOs), a class of powerful copper-dependent oxidative enzymes with the ability to break glycosidic bonds in polysaccharides such as cellulose, xyloglucan, glucomannan, xylan, starch, and chitin have been discovered in saccharolytic bacteria and fungi (Vaaje-Kolstad *et al*., 2010; Forsberg *et al*., 2011; Phillips *et al*., 2011; Quinlan *et al*., 2011; Agger *et al*., 2014; Beeson *et al*., 2012; Frommhagen *et al*., 2015; Lo Leggio *et al*., 2015). LPMOs were previously classified as CBM33 in bacteria and GH61 in fungi, but these enzymes have been recategorized as Auxiliary Activity (AA) families AA9, AA10, AA11, and AA13 in the CAZy database (Lombard *et al*., 2014). LPMOs disrupt the cellulose crystalline architecture by weakening the hydrogen bonds and van der Waals interactions, making the cellulose packing to become less-ordered and more susceptible to the action of cellulases (Villares *et al*., 2017). The copper-containing active sites of LPMOs cycle between Cu (I) and Cu (II) to activate molecular oxygen into a super oxide in a reducing agent-dependent manner by abstracting a hydrogen (Westereng *et al*., 2015; Villares *et al*., 2017). Electron donors that can induce LPMO action include small molecules (e.g. ascorbic acid, gallic acid, and phenolic compounds), macromolecules (e.g. lignin), and enzymes (e.g.

cellobiose dehydrogenases) (Frommhagen *et al*., 2017). The CAO proteins are distant relative to LPMOs, and currently there is no evidence that CAO proteins support the growth of *M. pectinilyticus* on plant carbohydrates, prompting a future investigation.

Another group of secreted proteins found in abundance in the supernatant samples was CAZymes without identifiable SLH anchorage domains. All 17 non-SLH CAZymes found in the supernatant proteome contained a signal peptide-encoding sequence, and the majority of enzymes were predicted to be involved in cleaving pectin-associated glycosidic linkages with the exceptions of GH3, GH5, and CE4. It remains to be determined how these enzymes are presented to the substrates as they lacked cell-wall retention domains such as SLH modules which have affinity for the peptidoglycan structure or dockerin-like modules to become incorporated into protein complexes. One proposed possibility was that the CAZymes secreted without SLH domains may become freely released into the surrounding, and function as independent saccharolytic enzymes to enhance the pectin degradation process. *Clostridium cellulovorans*, a cellulosome-producing anaerobic mesophilic bacterium, has been studied for its ability to maximize the efficiency of cellulose degradation using cooperative activities of cellulosome complexes and freely secreted non-cellulosomal CAZymes (Esaka *et al*., 2015). As a result, the labelling-based LC/MS/MS exoproteome profiling of *C. cellulovorans* in the presence of various cellulose substrates revealed a mixture of cellulosomal and non-cellulosomal proteins becoming co-secreted into the supernatant of the stationary growth phase cultures (Esaka *et al*., 2015). *M. pectinilyticus* may possess a similar system in which freely secreted enzymes enhance the process of pectin degradation by cooperating with the cell wall-bound SLH proteins that are potentially involved in the substrate recognition and the enzymatic degradation of pectin biomass. For an extensive secretion of free enzymes to occur, *M. pectinilyticus* would likely possess an efficient substrate-binding machinery to maintain an intimate cell-to-substrate distance to minimize the enzyme diffusion into the fluid environment of the large intestine. Freely secreted enzymes diffused into the site of cell-substrate interaction may work in a cooperative manner with the proteins tethered to the cell surface to enhance the efficiency of pectin degradation.

7.4 Ecological presence of *M. pectinilyticus* in human colonic microbiota

In chapter 6, the ecological abundance and frequency of *M. pectinilyticus* in the sampled populations of NZ and the US were assessed using qPCR method and metagenome mining. In the NZ study, qPCR results were combined with the nutrient composition analysis to examine the potential link between the amount of different food intakes and the presence of *M. pectinilyticus*. The concentration of *M. pectinilyticus* per gram of faeces obtained from qPCR was within comparable ranges to the reported median abundance for *Prevotella*, *Roseburia*, and *Methanobrevibacter smithii* from healthy and lean individuals (Larsen *et al*., 2010; Armougom *et al*., 2009). In these studies, *Prevotella*, *Roseburia*, and *M. smithii* showed mean concentrations of 8.1, 8.5, and 8.0 $\log_{10}$ bacteria per gram of faeces, respectively. While *Prevotella* can constitute up to 10 – 40 % of the total microbiota within the rural populations that generally consume a high fibre and low-protein diet, its relative abundance was reported to be less than 5 % within the general European population consuming a Western style diet (Gorvitovskaia *et al*., 2016). The mean relative abundance of *Roseburia* was reported to be about 2 %, while *M. smithii* was also present at a low abundance in the human gut, since only 0.8 % of genes from the faecal metagenome was assigned to Archaea (Gorvitovskaia *et al*., 2016; Qin *et al*., 2010). While the family *Ruminococcaceae* can represent 10 – 24 % of the gut microbiota of healthy individuals, the relative abundance of fibre-digesting keystone bacterial species was reported to be moderate (Lay *et al*., 2005). Despite their important roles in the degradation of insoluble plant polysaccharides, *Ruminococcus bromii* and *Ruminococcus champanellensis*, the two insoluble fibre-digesting keystone species reported from the human intestine, each represented less than 5 % of the total gut bacteria (Abell *et al*., 2008; Reau *et al*., 2016; Venkataraman *et al*., 2016). The mean relative abundance of *Ruminococcus* spp. associated with the human host was reported to be less than 1 % (Reau *et al*., 2016). Our results showed that *M. pectinilyticus* constituted approximately 1.1 % of the total faecal bacteria, indicating the strain is present in the human gut at a relatively low abundance. The relative abundance of *M. pectinilyticus* showed a potential positive correlation with the consumption of fibre and pectin. Of the 10 donors who reported to consume the daily recommendation of $>$ 5 grams of pectin, 5 donors were tested positive for the presence of *M. pectinilyticus* (Willats *et al*., 2006). In donor 24 who reported a daily pectin consumption of $>$ 11

grams, the relative abundance of *M. pectinilyticus* increased to 8.4 %, suggesting a high pectin diet may enrich for *M. pectinilyticus* to a significant level of abundance in the human gut. Previous studies demonstrated a diet-induced alteration of the relative abundance of a human enteric keystone bacterial species using *R. bromii,* which significantly increased in abundance *in vivo* in response to a diet high in resistant starch (RS) (Walker *et al*., 2011; Abell *et al*., 2008; Martinez *et al*., 2010; Venkataraman *et al*., 2016). Interestingly, while amylolytic strains of *Bifidobacterium adolescentis*, *Eubacterium rectale,* and the *Roseburia* group were also shown to surge in response to RS-rich diet, up to 69 % of RS survived the digestion process and was recovered from faeces of individuals characterized with low *R. bromii* counts (Walker *et al*., 2011; Martinez *et al*., 2010; Venkataraman *et al*., 2016). Therefore, while the functional redundancy that comes with a diverse set of gut microbes ensures that the basic metabolic and digestive functions are maintained, the efficiency of insoluble fibre degradation becomes significantly enhanced by the presence of keystone bacterial species with fibrolytic specialty. While *M. pectinilyticus* showed phenotypic potential for being described as a key pectinolytic species, it is currently unclear how the presence or absence of *M. pectinilyticus* may affect the overall pectin digestibility in the human large intestine. It is interesting that *M. pectinilyticus* was not detected from previous studies which investigated the *in vivo* and *in vitro* pectin-induced changes in the gut microbial composition using cultivation-independent methods (Licht *et al*., 2010; Chung *et al*., 2016). Using qPCR and taxonomic sequencing approaches, both studies observed significant increases in the relative abundance of *Clostridium* cluster XIVa including the *Roseburia*/*Eubacterium* group and *Eubacterium eligens* in apple pectin-fed rats and the *in vitro* human gut models enriched with apple pectin. *Bacteroides* spp. showed increases in their relative abundance in the *in vitro* gut model, but showed a significant reduction in animals fed with a diet supplemented with pectin. These results may suggest *Bacteroides* spp. efficiently compete in a relatively stable condition of fermenters in which the absence of fluctuation after periodic meals and the use of a single type of carbohydrate substrate may have resulted in the formation of a relatively simple and artificial microbiota (Chung *et al*., 2016). The absence of *M. pectinilyticus* in these studies may suggest that the strain is not a well-established member of the rat gut commensals, and its slow growth rate and low abundance may have prevented it from efficiently competing with other bacterial

groups within a closed fermenter system. In addition to qPCR detection, the SLH protein sequences from *M. pectinilyticus* were also found to be present in 14 % of the US metagenome datasets, indicating *M. pectinilyticus* is a member of the colon microbiota of humans, and not a phenomenon exclusively observed in NZ.

7.5 A proposed model for pectin degradation by *M. pectinilyticus*

Based on the genomic and proteomic interpretations of the glycobiome capability of *M. pectinilyticus*, we propose a putative model for its pectin degradation mechanism (Figure 7.2). The proposed model attempts to address the sequential process of pectin degradation by the cooperative action of enzymes annotated with CAZymes functions, and the putative mechanisms for carbohydrate recognition and binding. However, the proposed model mostly relies on the computational prediction of CAZyme functions which would remain hypothetical until they are experimentally validated. Furthermore, the novelty of the strain has left us with few biological evidences from which the mechanistic detail of its carbohydrate degradation system could be deduced, especially with regards to the exact roles of SLH proteins. Although the association of SLH proteins with pectin-degrading CAZymes, and the genomic positions of SLH protein-coding genes within the pectin-related gene clusters suggested that SLH proteins presumably play important roles in pectin degradation, elucidating the molecular and structural nature of these proteins and their roles in carbohydrate interactions will require an extensive amount of future research.

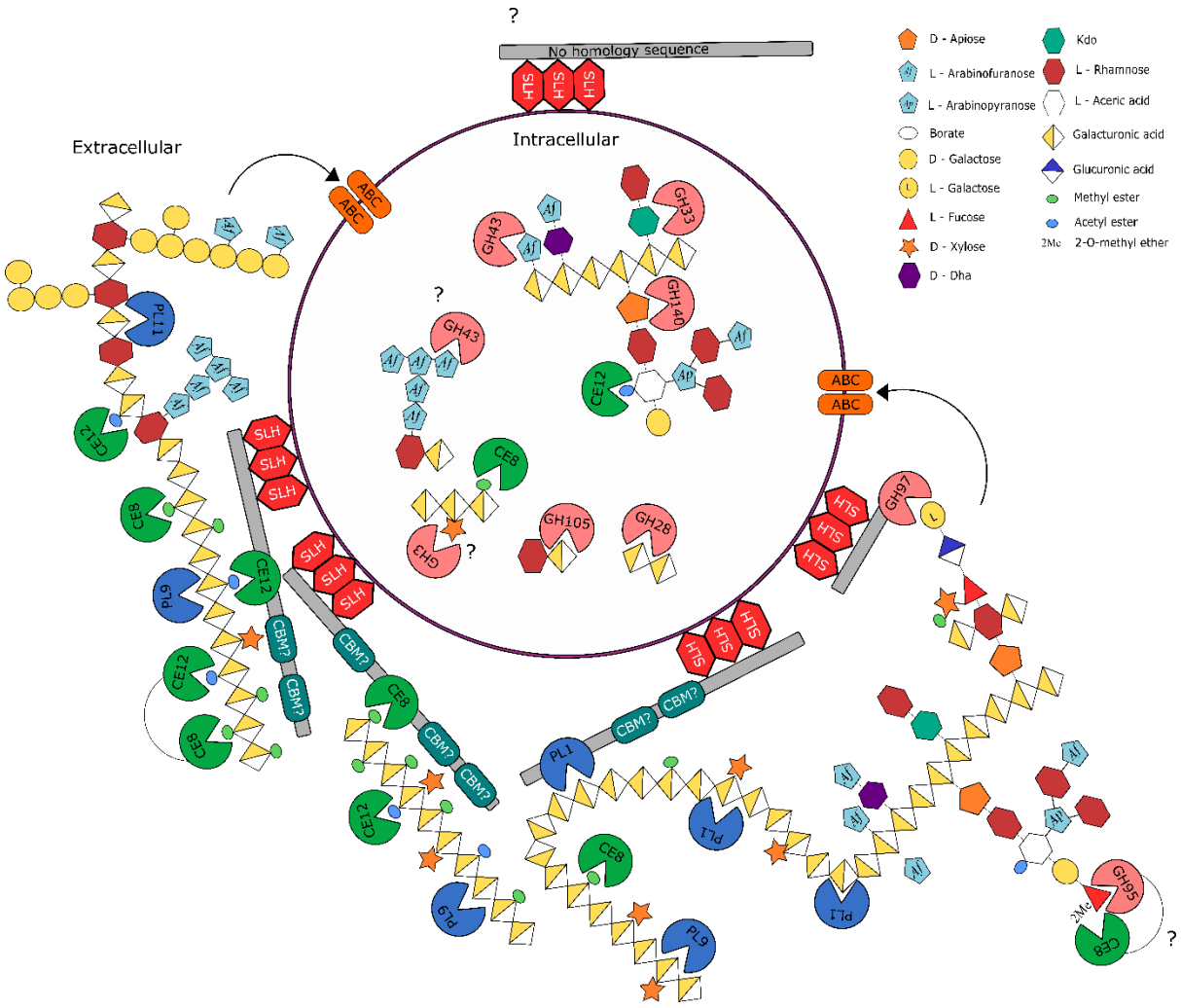


Figure 7.2 A proposed model for pectin degradation by *M. pectinilyticus*. (1) Cells are brought close to the plant cell wall by putative CBMs appended to SLH proteins with CAZyme functions. The substrate specificities of the CBMs remain to be determined. Large uncharacterized SLH proteins presumably play roles in carbohydrate-binding and assisting the pectin degradation process, but the exact mechanism is unknown. (2) De-esterification of methyl/acetyl groups begins with extracellular CE8 and CE12, subsequently allowing the access of PL1, PL9, and PL11 for lyase-mediated degradation of the pectin backbone. While some CAZymes remain attached to the cell surface via SLH modules, others are freely secreted into the surroundings, and function as independent saccharolytic enzymes to enhance the pectin degradation process. GH95/CE8 multi-domain enzymes exhibit a novel CAZyme combination which may be useful for cleaving the 2-*O*-methyl-α-L-Fuc*p* in RG-II. (3) Degraded pectins are transported into the intracellular space via ABC-related transporters. (4) Oligomeric uronic acid sugars are further broken down into monomers by intracellular GH28 and GH105. GH3, GH33, GH43, and GH140 carry out partial degradation of RG-I and RG-II side chains, although the exact functions of poly-specific enzymes (e.g. GH3 and GH43) are difficult to predict without experimental validation. (5) Galacturonic acid, L-arabinose, and D-xylose monomers enter the metabolic pathways for subsequent utilization.

7.6 Conclusions

In conclusion, the findings from this study shed new light on the existence of a novel phylogenetic lineage currently typified by a rare pectin-degrading specialist bacterium which possesses putative carbohydrate-associated extracellular proteins of striking novelty. The genome analysis of *M. pectinilyticus* revealed that the organism is the first human gut bacterium to possess the keystone potential to exclusively degrade pectin. In *M. pectinilyticus*, a high efficiency pectin degradation is achieved through a novel type of pectin-degrading strategy which is distinct from other microbial carbohydrate degradation systems studied to date. The study began here increases our understanding of the microbial pectin degradation in the human colon by presenting a possibility outside the current paradigms of microbial polysaccharide degradation, and by providing a microbial model which has potentially co-evolved with its host to accommodate the pectin-rich diet of humans.

7.7 Future directions

This thesis reports the discovery of *M. pectinilyticus*, the first pectin specialist bacterium isolated from human faeces. *M. pectinilyticus* is an ideal model organism for studying various aspects of colonic pectin degradation, as it has evolved a sophisticated glycobiome for serving a specific purpose of the pectin degradation and utilization. The genomic and phylogenetic novelty of *M. pectinilyticus* leaves ample scope for future research. The presence of putative pectin-binding machinery on the cell surface suggests an evolutionary adaptation which may enable *M. pectinilyticus* to grow on pectin while resisting the fluid movement in the bowel by forming physical adhesion to the plant materials. Although the protein multi-modularity of CAZyme domains and CBMs is known to be crucial features of most anaerobic polysaccharide degradation systems, *M. pectinilyticus* had few identifiable CBMs that found homologous matches in the current CAZy database, most of them unrelated to pectin. Therefore, the exact nature of the pectin-binding mechanism remains to be determined, and it was speculated that *M. pectinilyticus* may contain numerous uncharacterized pectin-binding CBMs which require experimental validation. Characterizing putative CBMs from bacteria typically involves cloning and expressing recombinant proteins, and testing the glycan binding ability of purified proteins through affinity gel electrophoresis and titration assays. In a recent study, antibody-mediated

glycan microarray was used to develop a high throughput detection methodology to simultaneously identify novel CBM families appended to GHs and PLs, and their binding specificities to various plant glycans (Venditto *et al*., 2016). Due to the complex nature of pectin structure, CBMs targeting a variety of different pectic glycans are expected to be present in *M. pectinilyticus*. In addition to CBMs, the molecular mechanisms underlying the presumed carbohydrate-binding mediated by SLH proteins will need to be elucidated. It remains unknown whether each SLH protein functions as an independent protein unit to modulate the cell attachment to pectin, or several SLH proteins use unknown protein interaction mechanisms to incorporate CAZymes and other proteins to form cellulosome-like protein complexes. The presence of multi-protein complexes in the culture supernatant or lysed cells can be investigated by using native-polyacrylamide gel electrophoresis (native-PAGE) to detect high molecular weight bands. These protein bands may be further dissociated into individual protein components under denaturing conditions (e.g. sodium dodecyl sulfate), and each of the solubilized protein units may be tested for catalytic and carbohydrate-binding activities. Advanced molecular biology techniques based on the traditional homologous recombination system or a more recently developed CRISPR/Cas9 platform may be used to design gene knockout or deletion mutant strains expressing non-functional SLH proteins to compare the phenotypic differences to the wildtype strain. However, it is important to note that developing genetic engineering platforms for a novel bacterial species may not be a trivial task, especially for an obligate anaerobic Gram positive bacterium which tends to suffer from issues such as low transformation efficiency and technical difficulties in mutant strain selection. The catalytic functions and substrate specificities of putative CAZymes will need to be experimentally characterized, and their long protein extensions merit further search for potentially unique functional properties such as novel CBM families or protein-protein interaction domains which may enable the assemply of a multi-protein pectin degradation system. In addition to enzyme activity assays and degradation end-products analysis, crystallographic studies are typically carried out to characterize the overall 3D architecture of sugar-binding domains and catalytic sites of CAZymes. Various biochemical techniques are available to examine the potential protein-protein interactions between recombinant proteins, including cross-linkage analysis, pull-down assays, co-immunoprecipitation methods, and Western-blot analysis.

Mutation analysis of protein interacting domains may further reveal the essential amino acid residue compositions contributing to the protein complex formation. The ecological roles of *M. pectinilyticus* during the hierarchical degradation of the plant cell wall should be assessed *in vitro* and *in vivo*. Whether *M. pectinilyticus* serves as a keystone bacterial species in the human colon remains to be determined by assessing the changes in the overall digestibility of plant carbohydrates with the inclusion of *M. pectinilyticus* either in co-culture with other keystone bacterial species or *in vivo* microbiome. To further support the positive relationships between the presence of *M. pectinilyticus* and the amount of fibre, pectin, and vegetable intakes, the frequency and abundance of *M. pectinilyticus* will need to be addressed within non-Western populations, such as the agricultural African communities, whose lifestyles are often characterized by diets high in domestically produced fruit, vegetables, and grains. *M. pectinilyticus* is currently the only cultured species representing the genus *Monoglobus* and a novel phylogenetic lineage of the *Ruminococcaceae*. However, the uncultured GenBank relatives of *M. pectinilyticus* appear to be reasonably common terrestrial gut commensals, indicating the significant underrepresentation of cultured members of this lineage can be addressed by prioritizing microbial cultivation practices. Such prioritization would provide more isolates to answer questions such as whether the pectin-degrading capacity is a common trait among the members of the lineage, or whether these organisms use a similar SLH-based carbohydrate degradation strategy to target a broader range of substrates. Certainly, additional microbiological and biochemical characterization of isolates is required to aid genomic studies aiming to provide better understanding of the physiological and ecological properties of this taxon.

References

Abbott, D. W. and A. B. Boraston (2007). "A family 2 pectate lyase displays a rare fold and transition metal assisted β-elimination." Journal of Biological Chemistry 282(48): 35328-35336.

Abbott, D. W. and A. B. Boraston (2007b). "The structural basis for exopolygalacturonase activity in a family 28 glycoside hydrolase." Journal of Molecular Biology 368(5): 1215-1222.

Abbott, D. W. and A. B. Boraston (2008). "Structural biology of pectin degradation by *Enterobacteriaceae*." Microbiology and Molecular Biology Reviews 72(2): 301-316.

Abbott, D. W., *et al*. (2007). "Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*." Journal of Molecular Biology 367(4): 1023-1033.

Abbott, D. W., *et al*. (2010). "The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate β-elimination." Journal of Biological Chemistry 285(**50**): 39029-39038.

Abell, G. C. J., *et al*. (2008). "Phylotypes related to *Ruminococcus bromii* are abundant in the large bowel of humans and increase in response to a diet high in resistant starch." Fems Microbiology Ecology 66(3): 505-515.

Adelsberger, H., *et al*. (2004). "Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the *in vivo* system from recombinant enzymes." Microbiology-Sgm 150: 2257-2266.

Agger, J., *et al*. (2014). "Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell degradation." Proceedings of the National Academy of Sciences 111: 6287-6292.

Aguillar, G. and Huitron, C. (1990) "Constitutive exopectinase produced by *Aspergillus* sp. CH-Y-1043 on different carbon sources." Biotechnology Letters 12: 655-660.

Akinosho, H., *et al*. (2014). "The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications." Frontiers in Chemistry 2.

Akiva, E., *et al*. (2014). "The structure-function linkage database." Nucleic Acids Research 42(D1): D521-D530.

Alber, O., *et al*. (2009). "Cohesin diversity revealed by the crystal structure of the anchoring cohesin from *Ruminococcus flavefaciens*." Proteins-Structure Function and Bioinformatics 77(3): 699-709.

Al-Hinai, M. A., *et al*. (2015). "The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation." Microbiology and Molecular Biology Reviews 79(1): 19-37.

Ali, E., *et al*. (2005). "Functions of family 22 carbohydrate-binding modules in *Clostridium josui* Xyn10A." Bioscience Biotechnology and Biochemistry 69(12): 2389-2394.

Ali, M. K., *et al*. (1999). "Cloning, sequencing, and expression of the gene encoding the *Clostridium stercorarium* xylanase C in *Escherichia coli*." Bioscience Biotechnology and Biochemistry 63(9): 1596-1604.

Ali, M. K., *et al*. (2001). "The multidomain xylanase Xyn10B as a cellulose-binding protein in *Clostridium stercorarium*." Fems Microbiology Letters 198(1): 79-83.

Ali, S., *et al*. (2008). "Leukocyte extravasation: An immunoregulatory role for alpha-L-Fucosidase?" Journal of Immunology 181(4): 2407-2413.

Anderson, J. W., *et al*. (2009). "Health benefits of dietary fiber." Nutrition Reviews 67(4): 188-205.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Armougom, F., *et al*. (2009). "Monitoring bacterial community of human gut microbiota reveals an increase in *Lactobacillus* in obese patients and methanogens in anorexic patients." Plos One 4(9).

Artzi, L., *et al*. (2015). "*Clostridium clariflavum*: Key cellulosome players are revealed by proteomic analysis." Mbio 6(3).

Artzi, L., *et al*. (2017). "Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides." Nature Reviews Microbiology 15(2): 83-95.

Attwood, K. (2002). "The PRINTS database: A resource for identification of protein families." Brief Bioinformatics 3(3): 252-63.
Available online at: http://eddylab.org/software/hmmer3/3.1b2/Userguide.pdf

Aziz, R. K., *et al*. (2008). "The RAST server: Rapid annotations using subsystems technology." BMC Genomics 9: 75

Backhed, F., *et al*. (2005). "Host-bacterial mutualism in the human intestine." Science 307(5717): 1915-1920.

Bagos, P. G., *et al*. (2008). "Prediction of lipoprotein signal peptides in Gram-positive bacteria with a hidden Markov model." Journal of Proteome Research **7**(12): 5082-5093.

Bankevich, A., *et al*. (2012). "SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing." Journal of Computational Biology 19(5): 455-477.

Bartholomew, J. W. and Mittwer, T. (1952). "The Gram stain." Bacteriology Review 16:1-29.

Bartosch, S., *et al*. (2004). "Characterization of bacterial communities in feces from healthy elderly volunteers and hospitalized elderly patients by using real-time PCR and effects of antibiotic treatment on the fecal microbiota." Applied and Environmental Microbiology 70(6): 3575-3581.

Bayer, E. A., *et al*. (1983). "Adherence of *Clostridium thermocellum* to cellulose." Journal of Bacteriology 156(2): 818-827.

Bayliss, C. E. and A. P. Houston (1984). "Characterization of plant polysaccharide and mucin fermenting anaerobic bacteria from human feces." Applied and Environmental Microbiology 48(3): 626-632.

Beeson, W. T., *et al*. (2012). "Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases." Journal of the American Chemical Society 134: 890-892.

Bekri, M. A., *et al*. (1999). "*Azospirillum irakense* produces a novel type of pectate lyase." Journal of Bacteriology 181(8): 2440-2447.

Beldman, G., *et al*. (1993). "Degradation of arabinans by arabinanases from *Aspergillus aculeatus* and *Aspergillus niger*." Carbohydrate Polymers 20(3): 159-168.

Ben David, Y., *et al*. (2015). "Ruminococcal cellulosome systems from rumen to human." Environmental Microbiology 17(9): 3407-3426.

Betian, H. G., *et al*. (1977). "Isolation of a cellulotytic *Bacteroides* sp. from human feces." Appl Environ Microbiol 33(4): 1009-1010.

Bhaya, D., *et al*. (2011). "CRISPR-Cas systems in Bacteria and Archaea: versatile small RNAs for adaptive defense and regulation." Annual Review of Genetics 45: 273-297.

Biely, P. (2012). "Microbial carbohydrate esterases deacetylating plant polysaccharides." Biotechnology Advances 30(6): 1575-1588.

Boetzer, M. and W. Pirovano (2012). "Toward almost closed genomes with GapFiller." Genome Biology 13(6).

Boetzer, M., *et al*. (2011). "Scaffolding pre-assembled contigs using SSPACE." Bioinformatics 27(4): 578-579.

Bolger, A. M., *et al*. (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics 30(15): 2114-2120.

Book, A. J., *et al*. (2014). "Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases." Biotechnology for Biofuels 7:109.

Boraston, A. B. and D. W. Abbott (2012). "Structure of a pectin methylesterase from *Yersinia enterocolitica*." Acta Crystallographica Section F-Structural Biology and Crystallization Communications 68: 129-133.

Boraston, A. B., *et al*. (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochemical Journal 382: 769-781.

Bourgois, T. M., *et al*. (2007). "Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp *subtilis* ATCC 6051: a GH 43 arabinoxylan arabinofuranohydrolase." Applied Microbiology and Biotechnology 75(6): 1309-1317.

Brechtel, E., *et al*. (1999). "Cell wall of *Thermoanaerobacterium thermosulfurigenes* EM1: isolation of its components and attachment of the xylanase XynA." Archives of Microbiology 171(3): 159-165.

Breton, C., *et al*. (2012). "Recent structures, evolution and mechanisms of glycosyltransferases." Current Opinion in Structural Biology 22(5): 540-549.

Brown, I. E., *et al*. (2001). "Pectate lyase 10A from *Pseudomonas cellulosa* is a modular enzyme containing a family 2a carbohydrate-binding module." Biochemical Journal 355: 155-165.

Brown, S. D., *et al*. (2012). **"**Draft genome sequences for *Clostridium thermocellum* wild type strain YS and derived cellulose adhesion-defective mutant strain AD2." Journal of Bacteriology 194:3290-3291.

Browne, H. P., *et al*. (2016). "Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation." Nature 533(7604): 543.

Bush, M. S., *et al*. (2001). "Developmental regulation of pectic epitopes during potato tuberisation." Planta 213(6): 869-880.

Caffall, K. H. and D. Mohnen (2009). "The structure, function, and biosynthesis of plant cell wall pectic polysaccharides." Carbohydrate Research 344(14): 1879-1900.

Cameron, E. A., *et al*. (2012). "Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism." Journal of Biological Chemistry 287(41): 34614-34625.

Cantarel, B. L., *et al*. (2009). "The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics." Nucleic Acids Research 37: D233-D238.

Carapito, R., *et al*. (2009). "Molecular Basis of arabinobio-hydrolase activity in phytopathogenic fungi crystal structure and catalytic mechanism of *Fusarium graminearum* GH93 exo-α-L-arabinanase." Journal of Biological Chemistry 284(18): 12285-12296.

Carnachan, S. M., *et al*. (2012). "Effects of simulated digestion *in vitro* on cell wall polysaccharides from kiwifruit (*Actinidia* spp.)." Food Chemistry 133(1): 132-139.

Carpita, N. C. and D. M. Gibeaut (1993). "Structural models of primary cell walls in flowering plants - consistency of molecular structure with the physical properties of the walls during growth." Plant Journal 3(1): 1-30.

Cartmell, A., *et al*. (2011). "The structure and function of an arabinan-specific α-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases." Journal of Biological Chemistry 286(**17**): 15483-15495.

Chassard, C., *et al*. (2012). "*Ruminococcus champanellensis* sp nov., a cellulose-degrading bacterium from human gut microbiota." International Journal of Systematic and Evolutionary Microbiology 62: 138-143.

Chen, I. M. A., *et al*. (2017). "IMG/M: integrated genome and metagenome comparative data analysis system." Nucleic Acids Research 45(D1): D507-D516.

Chen, L., *et al*. (2016). "First report on the bacterial diversity in the distal gut of dholes (*Cuon alpinus*) by using 16S rRNA gene sequences analysis." Journal of Applied Genetics 57(2): 275-283.

Cheng, K. J., *et al*. (1979). "Maceration of clover and grass leaves by *Lachnospira multiparus*." Applied and Environmental Microbiology 38(4): 723-729.

Cheng, K. J., *et al*. (1980). "Sequence of events in the digestion of fresh legume leaves by tumen bacteria." Applied and Environmental Microbiology 40(3): 613-625.

Cheng, K. J., *et al*. (1984). "Electron-microscopy of bacteria involved in the digestion of plant cell walls." Animal Feed Science and Technology 10: 93-120.

Cheng, Y. M., *et al*. (2009). "Cloning and functional characterization of a complex endo-β-1,3-glucanase from *Paenibacillus* sp." Applied Microbiology and Biotechnology 81(6): 1051-1061.

Cho, K. H. and A. A. Salyers (2001). "Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*." Journal of Bacteriology 183(24): 7224-7230.

Chou, W. C., *et al*. (2015). "Analysis of strand-specific RNA-seq data using machine learning reveals the structures of transcription units in *Clostridium thermocellum*." Nucleic Acids Research 43(10).

Chung, W. S. F., *et al*. (2016). "Modulation of the human gut microbiota by dietary fibres occurs at the species level." Bmc Biology 14.

Chylinski, K., *et al*. (2014). "Classification and evolution of type II CRISPR-Cas systems." Nucleic Acids Research 42(10): 6091-6105.

Cid, M., *et al*. (2010). "Recognition of the helical structure of β-1,4-galactan by a new family of carbohydrate-binding modules." Journal of Biological Chemistry 285(46): 35999-36009.

Collen, P. N., *et al*. (2014). "A novel unsaturated beta-glucuronyl hydrolase involved in ulvan degradation unveils the versatility of stereochemistry requirements in family GH105." Journal of Biological Chemistry 289(9): 6199-6211.

Collins, M. D., *et al*. (1994). "The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations." International Journal of Systematics and Evolutionary Microbiology 44(4) 812-826.

Comstock, L. E. (2009). "Importance of glycans to the host-*Bacteroides* mutualism in the mammalian intestine." Cell Host & Microbe 5(6): 522-526.

Condemine, G. and A. Ghazi (2007). "Differential regulation of two oligogalacturonate outer membrane channels, KdgN and KdgM, of *Dickeya dadantii* (*Erwinia chrysanthemi*)." Journal of Bacteriology 189(16): 5955-5962.

Condemine, G. and J. Robertbaudouy (1987). "2-keto-3-deoxygluconate transport system in *Erwinia chrysanthemi*." Journal of Bacteriology 169(5): 1972-1978.

Conway, J. M., *et al*. (2016). "Multi-domains, surface layer associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species." Journal of Biological Chemistry Available online at doi: 10.1074/jbc.M115.707810.

Corpet, F., *et al*. (1998). "The ProDom database of protein domain families." Nucleic Acids Research 26(1): 323-326.

Creze, C., *et al*. (2008). "The crystal structure of pectate lyase PelI from soft rot pathogen *Erwinia chrysanthemi* in complex with its substrate." Journal of Biological Chemistry 283(26): 18260-18268.

Cummings, J. H. and H. N. Englyst (1987). "Fermentation in the human large-intestine and the available substrates." American Journal of Clinical Nutrition 45(5): 1243-1255.

Cummings, J. H., *et al*. (1979). "Digestion of pectin in the human gut and its effect on calcium-absorption and large bowel function." British Journal of Nutrition 41(3): 477-485.

Dassa, B., *et al*. (2014). "Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six Ruminococcal strains." Plos One 9(7).

David *et al*., (2015). "Ruminococcal cellulosome systems from rumen to human." Environmental Microbiology. 17(9): 3407-3426.

Davies, G. and B. Henrissat (1995). "Structures and mechanisms of glycosyl hydrolases." structure 3(9): 853-859.

De Ioannes, P., *et al*. (2000). "An α-L-arabinofuranosidase from *Penicillium purpurogenum*: production, purification and properties." Journal of Biotechnology 76(2-3): 253-258.

De Vos, P., *et al*. (2009). "Bergey's Manual of Systematic Bacteriology". 2nd ed. Vol. 3. New York. Springer.

DElla, J. N. and A. A. Salyers (1996). "Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron on starch*." Journal of Bacteriology 178(24): 7173-7179.

Despres, J., *et al*. (2016). "Unraveling the pectinolytic function of *Bacteroides xylanisolvens* using a RNA-seq approach and mutagenesis." *BMC Genomics* 17, DOI 10.1186/s12864-016-2472-1.

Desvaux, M. (2005). "*Clostridium cellulolyticum*: Model organism of mesophilic cellulolytic clostridia." Fems Microbiology Reviews 29(4): 741-764.

Dhillon, B. K., *et al*. (2015). "IslandViewer 3: More flexible, interactive genomic island discovery, visualization and analysis." Nucleic Acids Research 43(W1): W104-W108.

Di Domenico, T., *et al*. (2012). "MobiDB: a comprehensive database of intrinsic protein disorder annotations." Bioinformatics 28(15): 2080-2081

Din, N., *et al*. (1994). "C-1-C-X revisited - intramolecular synergism in a cellulase." Proceedings of the National Academy of Sciences of the United States of America 91(24): 11383-11387.

Dodd, D., *et al*. (2011). "Xylan degradation, a metabolic property shared by rumen and human colonic *Bacteroidetes*." Molecular Microbiology 79(2): 292-304.

Dongowski, G., *et al*. (2000). "Degradation of pectins with different degrees of esterification by *Bacteroides thetaiotaomicron* isolated from human gut flora." Appl Environ Microbiol 66(4): 1321-1327.

Dongowski, G., *et al*. (2002). "The degree of methylation influences the degradation of pectin in the intestinal tract of rats and *in vitro*." Journal of Nutrition 132(7): 1935-1944.

Duvetter, T., *et al*. (2006). "Mode of de-esterification of alkaline and acidic pectin methyl esterases at different pH conditions." Journal of Agricultural and Food Chemistry 54(20): 7825-7831.

Eckburg, P. B., *et al*. (2005). "Diversity of the human intestinal microbial flora." Science 308(5728): 1635-1638.

Eddy, S. and Wheeler, T. (2013) HMMER user's guide version 3.1b1. Available online at: http://eddylab.org/software/hmmer3/3.1b2/Userguide.pdf

El Kaoutari, A., *et al*. (2013). "The abundance and variety of carbohydrate-active enzymes in the human gut microbiota." Nature Reviews Microbiology 11(**7**): 497-504.

Fagan, R. P. and N. F. Fairweather (2014). "Biogenesis and functions of bacterial S-layers." Nature Reviews Microbiology 12(3): 211-222.

Feng, J. X., *et al*. (2000). "Cloning, sequencing, and expression of the gene encoding a cell-bound multi-domain xylanase from *Clostridium josui*, and characterization of the translated product." Bioscience Biotechnology and Biochemistry 64(12): 2614-2624.

Ferrandez, Y. and G. Condemine (2008). "Novel mechanism of outer membrane targeting of proteins in Gram-negative bacteria." Molecular Microbiology 69(6): 1349-1357.

Fichant, G., *et al*. (2006). "ABCdb: an online resource for ABC transporter repertories from sequenced archaeal and bacterial genomes." Fems Microbiology Letters 256(2): 333-339.

Finn, R. D., *et al*. (2014). "Pfam: the protein families database." Nucleic Acids Research 42(D1): D222-D230.

Flamholz, A., *et al*. (2013). "Glycolytic strategy as a trade-off between energy yield and protein cost." Proceedings of the National Academy of Sciences of the United States of America 110(24): 10039-10044.

Flint, H. J. and E. A. Bayer (2008). "Plant cell wall breakdown by anaerobic microorganisms from the mammalian digestive tract." Incredible Anaerobes: From Physiology to Genomics to Fuels 1125: 280-288.

Flint, H. J., *et al*. (2008b). "Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis." Nature Reviews Microbiology 6(**2**): 121-131.

Flint, H. J., *et al*. (2012). "Microbial degradation of complex carbohydrates in the gut." Gut Microbes 3(4): 289-306.

Fodor, A. A., *et al*. (2012). "The "most wanted" taxa from the human microbiome for whole genome sequencing." Plos One **7**(7).

Fontes, C. M. G. A. and H. J. Gilbert (2010). "Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates." Annual Review of Biochemistry, 79: 655-681.

Forsberg, Z., *et al*. (2011). “Cleavage of cellulose by a CBM33 protein.” Protein Science 20(9): 1479-1483.

Fraser, J. S., *et al*. (2006). "Ig-like domains on bacteriophages: A tale of promiscuity and deceit." Journal of Molecular Biology 359(2): 496-507.

Fries, M., *et al*. (2007). "Molecular basis of the activity of the phytopathogen pectin methylesterase." Embo Journal 26(17): 3879-3887.

Frommhagen, M., *et al*. (2015). “Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase.” Biotechnology for Biofuels 8:101.

Frommhagen, M., *et al*. (2015). “Boosting LPMO-driven lignocellulose degradation by polyphenol oxidase-activated lignin building blocks.” Biotechnology for Biofuels 10: 121.

Fuchs, K. P., *et al*. (2003). "Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo-β-1, 3-glucanase bound to the outer cell surface." Microbiology-Sgm 149: 1021-1031.

Fujimoto, Z., *et al*. (2002). "Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: Sugar binding structure of the family 13 carbohydrate binding module." Journal of Molecular Biology 316(1): 65-78.

Fujimoto, Z., *et al*. (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.

Fujino, T., *et al*. (2000). "Characterization of cross-links between cellulose microfibrils, and their occurrence during elongation growth in pea epicotyl." Plant and Cell Physiology 41(**4**): 486-494.

Gamauf, C., *et al*. (2007). "Characterization of the bga1-encoded glycoside hydrolase family 35 β-galactosidase of *Hypocrea jecorina* with galacto-β-D-galactanase activity." Febs Journal 274(**7**): 1691-1700.

Germane, K. L., *et al*. (2015). "Structural analysis of *Clostridium acetobutylicum* ATCC 824 glycoside hydrolase from CAZy family GH105." Acta Crystallographica Section F-Structural Biology Communications 71: 1100-1108.

Gibson, G. R., *et al*. (1990). “The fermentability of polysaccharides by mixed faecal bacteria in relation to their suitability as bulk-forming laxatives.” Letters in Applied Microbiology 11:251–254.

Gilbert, H. J., *et al*. (2013). "Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules." Current Opinion in Structural Biology 23(5): 669-677.

Gill, S. R., *et al*. (2006). "Metagenomic analysis of the human distal gut microbiome." Science 312(5778): 1355-1359.

Glushka, J. N., *et al*. (2003). "Primary structure of the 2-O-methyl-α-L-fucose-containing side chain of the pectic polysaccharide, rhamnogalacturonan II." Carbohydrate Research 338(4): 341-352.

Gorvitovskaia, A., *et al*. (2016). "Interpreting *Prevotella* and *Bacteroides* as biomarkers of diet and lifestyle." Microbiome 4.

Gough, J. (2002). "The SUPERFAMILY database in structural genomics." Acta Crystallographica Section D-Biological Crystallography 58: 1897-1900.

Grant, J. R. and P. Stothard (2008). "The CGView Server: a comparative genomics tool for circular genomes." Nucleic Acids Research 36: W181-W184.

Grissa, I., *et al*. (2007). "CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats." Nucleic Acids Research 35: W52-W57.

Gupta, R., *et al*. (2012). "Characterization of a glycoside hydrolase family 1 beta-galactosidase from hot spring metagenome with transglycosylation activity." Applied Biochemistry and Biotechnology 168(6): 1681-1693.

Haft, D. H., *et al*. (2003). "The TIGRFAMs database of protein families." Nucleic Acids Research 31(1): 371-373.

Han, Y. J., *et al*. (2010). "Comparative analyses of two thermophilic enzymes exhibiting both β-1,4 mannosidic and β-1,4 glucosidic cleavage activities from *Caldanaerobius polysaccharolyticus*." Journal of Bacteriology 192(16): 4111-4121.

Han, Y. J., *et al*. (2012). "Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*." Journal of Biological Chemistry 287(42): 34946-34960.

Harris, P. J. and B. G. Smith (2006). "Plant cell walls and cell-wall polysaccharides: structures, properties and uses in food products." International Journal of Food Science and Technology 41: 129-143.

Hassan, S., *et al*. (2013). "PelN is a new pectate lyase of *Dickeya dadantii* with unusual characteristics." Journal of Bacteriology 195(10): 2197-2206.

Hatada, Y., *et al*. (2000). "Deduced amino-acid sequence and possible catalytic residues of a novel pectate lyase from an alkaliphilic strain of *Bacillus*." European Journal of Biochemistry 267(8): 2268-2275.

Hayashi, K., *et al*. (1997). "Pectinolytic enzymes from *Pseudomonas marginalis* MAFF 03-01173." Phytochemistry 45(**7**): 1359-1363.

Hayashi, T., *et al*. (1987). "Pea xyloglucan and cellulose." Plant Physiology 83(**2**): 384-389.

Heffernan, R., *et al*. (2015). "Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning." Scientific Reports 5.

Hehemann, J., *et al*. (2010). "Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota." Nature 464(8): 908-912.

Hehemann, J., *et al*. (2012). "Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes." Proceedings of the National Academy of Sciences of the United States of America 109(48): 19786-19791.

Heinken, A., *et al*. (2014). "Functional metabolic map of *Faecalibacterium prausnitzii*, a beneficial human gut microbe." Journal of Bacteriology 196(18): 3289-3302.

Herron, S. R., *et al*. (2000). "Structure and function of pectic enzymes: virulence factors of plant pathogens." Proceedings of the National Academy of Sciences of the United States of America 97(16): 8762-8769.

Hespell, R. B. (1992). "Fermentation of xylans by *Butyrivibrio fibrisolvens* and *Thennoanaerobacter* strain B6A: utilization of uonic acids and xylanolytic activities." Current Microbiology 25: 189 – 195.

Hespell, R. B., *et al*. (1987). "Fermentation of xylans by *Butyrivibrio fibrisolvens* and other ruminal Bacteria." Applied and Environmental Microbiology 53(12): 2849-2853.

Hla, S. S., *et al*. (2005). "A novel thermophilic pectate lyase containing two catalytic modules of *Clostridium stercorarium*." Biosc. Biotechnol. Biochem 69(11): 2138-2145.

Hoffmam, Z. B., *et al*. (2013). "Characterization of a hexameric exo-acting GH51 α-L-arabinofuranosidase from the mesophilic *Bacillus subtilis*." Molecular Biotechnology 55(3): 260-267.

Holloway, W. D., *et al*. (1978). "Digestion of certain fractions of dietary fiber in humans." American Journal of Clinical Nutrition 31(6): 927-930.

Hopkins, M. J., *et al*. (2003). "Degradation of cross-linked and non-cross-linked arabinoxylans by the intestinal microbiota in children." Applied and Environmental Microbiology 69(11): 6354-6360.

Hopkins, M. J., *et al*. (2005). "Characterisation of intestinal bacteria in infant stools using real-time PCR and northern hybridisation analyses." Fems Microbiology Ecology 54(1): 77-85.

Hoy, M. K. and Goldman, J. D. (2014). "Fiber intake of the U.S. population." Food Surveys Research Group Dietary Data Brief 12.

Huang, X. L., *et al*. (2015). "Improved expression and characterization of a multidomain xylanase from *Thermoanaerobacterium aotearoense* SCUT27 in *Bacillus subtills*." Journal of Agricultural and Food Chemistry 63(28): 6430-6439.

HugouvieuxCottePattat, N., *et al*. (1996). "Regulation of pectinolysis in *Erwinia chrysanthemi*." Annual Review of Microbiology 50: 213-257.

Hugouvieux-Cotte-Pattat, N., *et al*. (2014). "Bacterial pectate lyases, structural and functional diversity." Environmental Microbiology Reports 6(5): 427-440.

Hulo, N., *et al*. (2006). "The PROSITE database." Nucleic Acids Research 34: D227-D230.

Hung, K. S., *et al*. (2011). "Characterization of a salt-tolerant xylanase from *Thermoanaerobacterium saccharolyticum* NTOU1." Biotechnology Letters 33(7): 1441-1447.

Hungate, R. (1969). A roll tube method for cultivation of strict anaerobes. Methods Microbiolgy 3B: 117–132.

Hyatt, D., *et al*. (2010). "Prodigal: prokaryotic gene recognition and translation initiation site identification." BMC Bioinformatics 11: 119

Ifkovits, R. W. and H. S. Ragheb (1968). "Cellular fatty acid composition and identification of rumen bacteria." Applied Microbiology 16(9): 1406-&.

Inacio, J. M. and I. de Sa-Nogueira (2008). "Characterization of abn2 (yxiA), encoding a *bacillus subtilis* GH43 arabinanase, Abn2, and its role in arabino-polysaccharide degradation." Journal of Bacteriology 190(12): 4272-4280.

Ito, Y., *et al*. (2003). "Cloning, expression, and cell surface localization of *Paenibacillus* sp strain W-61 xylanase 5, a multidomain xylanase." Applied and Environmental Microbiology 69(12): 6969-6978.

Itoh, T., *et al*. (2006). "A novel glycoside hydrolase family 105: The structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide." Journal of Molecular Biology 360(**3**): 573-585.

Itoh, T., *et al*. (2014). "Overexpression, purification, and characterization of *Paenibacillus* cell surface-expressed chitinase ChiW with two catalytic domains." Bioscience Biotechnology and Biochemistry 78(4): 624-634.

Jayani, R. S., *et al*. (2005). "Microbial pectinolytic enzymes: A review." Process Biochemistry 40(9): 2931-2944.

Jenkins, J., *et al*. (2004). "The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*." Journal of Biological Chemistry 279(10): 9139-9145.

Jensen, L. J., *et al*. (2008). "eggNOG: automated construction and annotation of orthologous groups of genes." Nucleic Acids Research 36: D250-D254.

Jensen, N. S. and E. Canale-Parola (1986). "*Bacteroides pectinophilus* sp. nov. and *Bacteroides galacturonicus* sp. nov.: two pectinolytic bacteria from the human intestinal tract." Appl Environ Microbiol 52(4): 880-887.

Jindou, S., *et al*. (2006). "Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens*." Journal of Bacteriology 188(**22**): 7971-7976.

Jindou, S., *et al*. (2008). "Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*." Fems Microbiology Letters 285(2): 188-194.

Jones, P., *et al*. (2014). "InterProScan 5: genome-scale protein function classification." Bioinformatics 30(9): 1236-1240.

Kalmokoff, M., *et al*. (2011). "Continuous feeding of antimicrobial growth promoters to commercial swine during the growing/finishing phase does not modify faecal community erythromycin resistance or community structure." Journal of Applied Microbiology 110(6): 1414-1425.

Kanehisa, M., *et al*. (2016). "KEGG as a reference resource for gene and protein annotation." Nucleic Acids Research 44(D1): D457-D462

Karan, R., *et al*. (2013). "Cloning, overexpression, purification, and characterization of a polyextremophilic β-galactosidase from the Antarctic *haloarchaeon Halorubrum lacusprofundi*." Bmc Biotechnology 13.

Kashyap, D. R., *et al*. (2000). "Production, purification and characterization of pectinase from a *Bacillus* sp DT7." World Journal of Microbiology & Biotechnology 16(3): 277-282.

Kashyap, D. R., *et al*. (2001). "Applications of pectinases in the commercial sector: a review." Bioresource Technology 77(3): 215-227.

Kataeva, I. A., *et al*. (2002). "The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface." Applied and Environmental Microbiology 68(9): 4292-4300.

Kato, S., *et al*. (2004). "*Clostridium straminisolvens* sp nov., a moderately thermophilic, aerotolerant and cellulolytic bacterium isolated from a cellulose-degrading bacterial community." International Journal of Systematic and Evolutionary Microbiology 54: 2043-2047.

Kaur, A. P., *et al*. (2015). "Functional and structural diversity in GH62 α-L-arabinofuranosidases from the thermophilic fungus *Scytalidium thermophilum*." Microbial Biotechnology 8(3): 419-433.

Kelleher, J., *et al*. (1984). "Degradation of cellulose within the gastrointestinal tract in man." Gut 25(**8**): 811-815.

Kibbe, W. A. (2007). "OligoCalc: an online oligonucleotide properties calculator." Nucleic Acids Research 35: W43-W46.

Kim, H., *et al*. (2000). "Molecular characterization of xynX, a gene encoding a multidomain xylanase with a thermostabilizing domain from *Clostridium thermocellum*." Applied Microbiology and Biotechnology 54:521-527.

Kim, Y. W., *et al*. (2006). "Catalytic properties of a mutant β-galactosidase from *Xanthomonas manihotis* engineered to synthesize galactosyl-thio-β-1,3 and -β-1,4-glycosides." Febs Letters 580(18): 4377-4381.

Kitamura, M., *et al*. (2008). "Structural and functional analysis of a glycoside hydrolase family 97 enzyme from *Bacteroides thetaiotaomicron*." Journal of Biological Chemistry 283(52): 36328-36337.

Kong, F. B. and R. P. Singh (2009). "Modes of disintegration of solid foods in simulated gastric environment." Food Biophysics 4(3): 180-190.

Koropatkin, N. M., *et al*. (2008). "Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices." Structure 16(**7**): 1105-1115.

Koseki, T., *et al*. (2003). "Role of two α-L-arabinofuranosidases in arabinoxylan degradation and characteristics of the encoding genes from shochu koji molds, *Aspergillus kawachii* and *Aspergillus awamori*." Journal of Bioscience and Bioengineering 96(3): 232-241.

Kosugi, A., *et al*. (2002). "Characterization of two noncellulosomal subunits, ArfA and BgaA, from *Clostridium cellulovorans* that cooperate with the cellulosome in plant cell wall degradation." Journal of Bacteriology 184(24): 6859-6865.

Krogh, A., *et al*. (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." Journal of Molecular Biology 305(3): 567-580.

Kumar, S., *et al*. (2016). "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." Molecular Biology and Evolution 33(7): 1870-1874.

Laatu, M. and G. Condemine (2003). "Rhamnogalacturonate lyase RhiE is secreted by the out system in *Erwinia chrysanthemi*." Journal of Bacteriology 185(5): 1642-1649.

Lagesen, K., *et al*. (2007). "RNAmmer: consistent and rapid annotation of ribosomal RNA genes." Nucleic Acids Research 35(9): 3100-3108.

Lagier, J. C., *et al*. (2016). "Culture of previously uncultured members of the human gut microbiota by culturomics." Nature Microbiology 1(12).

Lamed, R., *et al*. (1983). "Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*." *Journal of Bacteriology* 156:828-836.

Larkin, M. A., *et al*. (2007). "Clustal W and clustal X version 2.0." Bioinformatics 23(21): 2947-2948.

Larsen, N., *et al*. (2010). "Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults." Plos One 5(2).

Latham, M. J., *et al*. (1978). "*Ruminococcus flavefaciens* cell coat and adhesion to cotton cellulose and to cell walls in leaves of perennial ryegrass (*Lolium-Perenne*)." Applied and Environmental Microbiology 35(1): 156-165.

Lay, C., *et al*. (2005). "Colonic microbiota signatures across five northern European countries." Applied and Environmental Microbiology 71(7): 4153-4155.

Lee, J. Y., *et al*. (2017). "Microbial β-galactosidase of *Pediococcus pentosaceus* ID-7: isolation, cloning, and molecular characterization." Journal of Microbiology and Biotechnology 27(3): 598-609.

Lee, S. P., *et al*. (1994). "Cloning of the aapT gene and characterization of its product, α-amylase-pullulanase (aapT), from Thermophilic and alkaliphilic *Bacillus* sp strain Xal601." Applied and Environmental Microbiology 60(10): 3764-3773.

Leser, T. D., *et al*. (2002). "Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited." Applied and Environmental Microbiology 68(2): 673-690.

Ley, R. E., *et al*. (2006). "Microbial ecology - Human gut microbes associated with obesity." Nature 444(7122): 1022-1023.

Ley, R. E., *et al*. (2008). "Evolution of mammals and their gut microbes." Science 320(5883): 1647-1651.

Li, E., *et al*. (2012). "Inflammatory bowel diseases phenotype, *C. difficile* and NOD2 genotype are associated with shifts in human ileum associated microbial composition." Plos One 7(6).

Licht, T. R., *et al*. (2010). "Effects of apples and specific apple components on the cecal environment of conventional rats: role of apple pectin." Bmc Microbiology 10.

Lima, T., *et al*. (2009). "HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot." Nucleic Acids Research 37: D471-D478.

Lin, A., *et al*. (2013). "Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States." Plos One 8(1).

Lin, D. H., *et al*. (2016). "Interactions of pectins with cellulose during its synthesis in the absence of calcium." Food Hydrocolloids 52: 57-68.

Liu, L., *et al*. (2012). "Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by ^{13}C metabolic flux analysis." Journal of Bacteriology. 194: 5413 – 5422.

Liu, S. Y., *et al*. (1996). "Cloning, sequencing, and expression of the gene encoding a large S-layer-associated endoxylanase from *Thermoanaerobacterium* sp strain JW/SL-YS 485 in Escherichia coli." Journal of Bacteriology 178(6): 1539-1547.

Lo Leggio, L., *et al*. (2015). "Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. Nature Communications 6:5961.

Lombard, V., *et al*. (2010). "A hierarchical classification of polysaccharide lyases for glycogenomics." Biochemical Journal 432: 437-444.

Lombard, V., *et al*. (2014). "The carbohydrate-active enzymes database (CAZy) in 2013." Nucleic Acids Research 42(D1): D490-D495.

Lopez-Siles, M., *et al*. (2012). "Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth." Applied and Environmental Microbiology 78(2): 420-428.

Lowe, T. M. and S. R. Eddy (1997). "tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence." Nucleic Acids Research 25(5): 955-964.

Lv, W. and Z. Yu (2013). "Isolation and characterization of two thermophilic cellulolytic strains of *Clostridium thermocellum* from a compost sample." Journal of Applied Microbiology 114(4): 1001-1007.

Macfarlane, S. and G. T. Macfarlane (2003). "Regulation of short-chain fatty acid production." Proceedings of the Nutrition Society 62(1): 67-72.

Madden, R. H. (1983). "Isolation and characterization of *Clostridium stercorarium* sp nov, cellulolytic thermophile." International Journal of Systematic Bacteriology 33(4): 837-840.

Maehara, T., *et al*. (2014). "Crystal structure and characterization of the glycoside hydrolase family 62 alpha-L-arabinofuranosidase from *Streptomyces coelicolor*." Journal of Biological Chemistry 289(11): 7962-7972.

Mai, V., *et al*. (2000). "Cloning, sequencing, and characterization of the bifunctional xylosidase-arabinosidase from the anaerobic thermophile *Thermoanaerobacter ethanolicus*." Gene 247(1-2): 137-143.

Mai, V., *et al*. (2006). "Effect of bowel preparation and colonoscopy on post-procedure intestinal microbiota composition." Gut 55(12): 1822-1823.

Mao, X. Z., *et al*. (2012). "The percentage of bacterial genes on leading versus lagging strands is influenced by multiple balancing forces." Nucleic Acids Research 40(17): 8210-8218.

Marchler-Bauer, A., *et al*. (2011). "CDD: a Conserved Domain Database for the functional annotation of proteins." Nucleic Acids Research 39: D225-D229.

Marlett, J. A. and T. F. Cheung (1997). "Database and quick methods of assessing typical dietary fiber intakes using data for 228 commonly consumed foods." Journal of the American Dietetic Association 97(10): 1139-&.

Marounek, M. and D. Duskova (1999). "Metabolism of pectin in rumen bacteria *Butyrivibrio fibrisolvens* and *Prevotella ruminicola*." Letters in Applied Microbiology 29(6): 429-433.

Martens, E. C., *et al*. (2009). "Complex glycan catabolism by the human gut microbiota: The *Bacteroidetes* Sus-like paradigm." Journal of Biological Chemistry 284(37): 24673-24677.

Martens, E. C., *et al*. (2011). "Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts." Plos Biology 9(12).

Martinez, I., *et al*. (2010). "Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects." Plos One 5(11).

Matsuki, T., *et al*. (2004). "Use of 16S rRNA gene-targeted group-specific primers for real-time PCR analysis of predominant bacteria in human feces." Applied and Environmental Microbiology 70(12): 7220-7228.

Matsuura, Y. (1991). "Pectic acid degrading enzymes from human feces." Agricultural and Biological Chemistry 55(3): 885-886.

Matuschek, M., *et al*. (1994). "Pullulanase of *Thermoanaerobacterium thermosulfurigenes* Em1 (*Clostridium thermosulfurogenes*) - Molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface." Journal of Bacteriology 176(11): 3295-3302.

McCann, M. C., *et al*. (1990). "Direct visualization of cross-links in the primary plant cell wall." Journal of Cell Science 96: 323-334.

McCarthy, R. E., *et al*. (1985). "Location and characteristics of enzymes involved in the breakdown of polygalacturonic acid by *Bacteroides thetaiotaomicron*." Journal of Bacteriology 161(2): 493-499.

McKie, V. A., *et al*. (2001). "A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose." Biochemical Journal 355: 167-177.

McNulty, N. P., *et al*. (2013). "Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome." Plos Biology 11(8).

McSweeney, C. S., *et al*., (2005). “Rumen bacteria.” In P. M. Harinder, H. P. Makkar & C. S. McSweeney (Eds.), Methods in gut microbial ecology for ruminants (Vol 201, pp. 23-37): Springer

Meier-Kolthoff, J. P., *et al*. (2013). "Genome sequence-based species delimitation with confidence intervals and improved distance functions." Bmc Bioinformatics 14.

Meile, L., *et al*. (2002). “. Characterization of the D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase gene (*xfp*) from *Bifidobacterium lactis*.” Journal of Bacteriology. 183: 2929–2936.

Melton, L. D. and Smith, B. G. (2001). “Isolation of plant cell walls and fractionation of cell wall polysaccharides.” Current Protocols in Food Analytical Chemistry E3.1.1-E3.1.23

Mi, H. Y., *et al*. (2013). "Large-scale gene function analysis with the PANTHER classification system." Nature Protocols 8(8): 1551-1566.

Miller, L. G. (1959). “Use of dinitrosalicyclic acid reagent for the determination of reducing sugars.” Analytical Chemistry 31: 426-428

Mirande, C., *et al*. (2010). "Dietary fibre degradation and fermentation by two xylanolytic bacteria *Bacteroides xylanisolvens* XB1AT and *Roseburia intestinalis* XB6B4 from the human intestine." Journal of Applied Microbiology 109(2): 451-460.

Miyanaga, A., *et al*. (2004). "Crystal structure of a family 54 α-L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose." Journal of Biological Chemistry 279(43): 44907-44914.

Mohnen, D. (2008). "Pectin structure and biosynthesis." Current Opinion in Plant Biology 11(3): 266-277.

Monro, J. A., *et al*. (2010). "Baselines representing blood glucose clearance improve in vitro prediction of the glycaemic impact of customarily consumed food quantities." British Journal of Nutrition 103(2): 295-305.

Montanier, C. Y., *et al*. (2011). "A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization." Journal of Biological Chemistry 286(25): 22499-22509.

Montanier, C., *et al*. (2009). "Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function." Proceedings of the National Academy of Sciences of the United States of America 106(9): 3065-3070.

Montanier, C., *et al*. (2010). "Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules." Journal of Biological Chemistry 285(41): 31742-31754.

Monteils, V., *et al*. (2008). "Potential core species and satellite species in the bacterial community within the rabbit caecum." Fems Microbiology Ecology 66(3): 620-629.

Moore, W. E. C., *et al*. (1972). “*Ruminococcus bromii* sp. n. and emendation of the description of *Ruminococcus sijpestein*.” *International Journal of Systematic Bacteriology* 1972; 22:78-80.

Morris, D. D., *et al*. (1999). "Family 10 and 11 xylanase genes from *Caldicellulosiruptor* sp strain Rt69B.1." Extremophiles 3(2): 103-111.

Mukhopadhya, L., *et al*. (2017). "Sporulation capability and amylosome conserved among diverse human colonic and rumen isolates of the keystone starch-degrader *Ruminococcus bromii*." Environmental Microbiology. DOI: 10.1111/1462-2920.14000

Muller, S., *et al*. (2003). "Cell wall composition of vascular and parenchyma tissues in broccoli stems." Journal of the Science of Food and Agriculture 83(13): 1289-1292.

Munarin, F., *et al*. (2013). "Sterilization treatments on polysaccharides: effects and side effects on pectin." Food Hydrocolloids 31(1): 74-84.

Munir, R. I., *et al*. (2014). "Comparative analysis of carbohydrate active enzymes in *Clostridium termitidis* CT1112 reveals complex carbohydrate degradation ability." Plos One 9(8).

Nadkarni, M. A., *et al*. (2002). "Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set." Microbiology-Sgm 148: 257-266.

Nakajima, N., *et al*. (1999). "Degradation of pectic substances by two pectate lyases from a human intestinal bacterium, *Clostridium butyricum-beijerinckii* group." Journal of Bioscience and Bioengineering 88(3): 331-333.

Nasser, W., *et al*. (1999). "Analysis of three clustered polygalacturonase genes in *Erwinia chrysanthemi* 3937 revealed an anti-repressor function for the PecS regulator." Molecular Microbiology 34(4): 641-650.

Naumoff, D. G. (2014). "Hierarchical classification of glycoside hydrolases." Febs Journal 281: 569-569.

Navarro-Fernandez, J., *et al*. (2008). "Characterization of a new rhamnogalacturonan acetyl esterase from *Bacillus halodurans* C-125 with a new putative carbohydrate binding domain." Journal of Bacteriology 190(4): 1375-1382.

Ndeh, D., *et al*. (2017). "Complex pectin metabolism by gut bacteria reveals novel catalytic functions." Nature 544(7648): 65-+.

Nelson, K. E., *et al*. (2010). "A catalog of reference genomes from the human microbiome." Science 328(5981): 994-999.

Newstead, S. L., *et al*. (2005). "Galactose recognition by the carbohydrate-binding module of a bacterial sialidase." Acta Crystallographica Section D-Biological Crystallography 61: 1483-1491.

Nurizzo, D., *et al*. (2002). "*Cellvibrio japonicas* α-L-arabinanase 43A has a novel five-blade β-propeller fold." Nature Structural Biology 9(9):665-668.

Nyman, M. and N. G. Asp (1982). "Fermentation of dietary fibre components in the rat intestinal tract." Br J Nutr 47(3): 357-366.

Obro, J., *et al*. (2009). "High-throughput screening of *Erwinia chrysanthemi* pectin methylesterase variants using carbohydrate microarrays." Proteomics 9(7): 1861-1868.

Ochiai, A., *et al*. (2007). "Plant cell wall degradation by saprophytic *Bacillus subtilis* strains: Gene clusters responsible for rhamnogalacturonan depolymerization." Applied and Environmental Microbiology 73(12): 3803-3813.

O'Connell, J., *et al*. (2015). "NxTrim: optimized trimming of Illumina mate pair reads." Bioinformatics 31(12): 2035-2037.

Okano, K., *et al*. (2009). "Improved production of homo-D-lactic acid via xylose fermentation by introduction of xylose assimilation genes and redirection of the phosphoketolase pathway to the pentose phosphate pathway in L-lactate dehydrogenase gene-deficient *Lactobacillus plantarum*." Applied Environmental Microbiology. 75: 7858–7861.

O'Neill, M. A., *et al*. (2004). "Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide." Annual Review of Plant Biology, Vol 61 55: 109-139.

Ozdemir, I., *et al*. (2012). "S-Layer homology domain proteins Csac_0678 and Csac_2722 are implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*." Applied and Environmental Microbiology 78(3): 768-777.

Paes, G., *et al*. (2008). "The structure of the complex between a branched pentasaccharide and *Thermobacillus xylanilyticus* GH-51 arabinofuranosidase reveals xylan-binding determinants and induced fit." Biochemistry 47(28): 7441-7451.

Pages, S., *et al*. (2003). "A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome." Journal of Bacteriology 185(16): 4727-4733.

Paggi, R. A., *et al*. (2005). "Growth and pectate-lyase activity of the ruminal bacterium *Lachnospira multiparus* in the presence of short-chain organic acids." Letters in Applied Microbiology 41(5): 434-439.

Parkar, S. G., *et al*. (2010). "Gut health benefits of kiwifruit pectins: Comparison with commercial functional polysaccharides." Journal of Functional Foods 2(3): 210-218.

Parks, D. H., *et al*. (2015). "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." Genome Research 25(7): 1043-1055.

Patel, G. B., *et al*. (1980). "Isolation and characterization of an anaerobic, cellulolytic microorganism, *Acetivibrio cellulolyticus* gen. nov., sp nov" International Journal of Systematic Bacteriology 30(1): 179-185.

Patterson, H., *et al*. (1975). "Ultrastructure and adhesion properties of *Ruminococcus albus*." Journal of Bacteriology 122(1): 278-287.

Pearl, F., *et al*. (2005). "The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis." Nucleic Acids Research 33: D247-D251.

Petersen, T. N., *et al*. (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nature Methods 8(10): 785-786.

Phillips C, M., *et al*. (2011). "Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. ACS Chemical Biology 6(12): 1399-1406.

Pickersgill, R., *et al*. (1998). "Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. *carotovora*." Journal of Biological Chemistry 273(38): 24660-24664.

Pissavin, C., *et al*. (1998). "Biochemical characterization of the pectate lyase PelZ of *Erwinia chrysanthemi* 3937." Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology 1383(2): 188-196.

Preiss, J. (2006). "Bacterial glycogen inclusions: enzymology and regulation of synthesis." Inclusion in Prokaryotes. DOI 10.1007/7171_004. Springer.

Prynne, C. J. and D. A. Southgate (1979). "The effects of a supplement of dietary fibre on faecal excretion by human subjects." British Journal of Nutrition 41(3): 495-503.

Qin, J. J., *et al*. (2010). "A human gut microbial gene catalogue established by metagenomic sequencing." Nature 464(7285): 59-U70.

Qin, Q. L., *et al*. (2014). "A proposed genus boundary for the prokaryotes based on genomic insights." Journal of Bacteriology 196(12): 2210-2215.

Quinlan, R. J., *et al*. (2011). "Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proceedings of the National Academy of Sciences 108(37): 15079-15084.

Rainey, F. A., *et al*. (2015). "Bergey's Manual of Systematics of Archaea and Bacteria". 1-122.

Rajan, S. S., *et al*. (2004). "Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD^+/$Mn2^+$-dependent phospho-α-glucosidase from *Bacillus subtilis*." Structure 12:1619-1629.

Ransom-Jones, E., *et al*. (2012). "The *Fibrobacteres*: an important phylum of cellulose-degrading Bacteria." Microb Ecol 63(2): 267-281.

Ratanakhanokchai K., *et al*. (2013). "*Paenibacillus curdlanolyticus* strain B-6 multienzyme complex: A novel system for biomass utilization." In: Matovic M. D. (ed) Biomass Now-Cultivation and Utilization. pp 369-394 doi:10.5772/51820

Rawls, J. F., *et al*. (2006). "Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection." Cell 127(2): 423-433.

Reau, A. G., *et al*. (2016). "Sequence-based analysis of the genus *Ruminococcus* resolves its phylogeny and reveals strong host association." Microb Genomics 2 (000099).

Redgwell, R. J., *et al*. (1997). "Galactose loss and fruit ripening: high-molecular-weight arabinogalactans in the pectic polysaccharides of fruit cell walls." Planta 203(2): 174-181.

Remoroza, C., *et al*. (2014). "A *Bacillus licheniformis* pectin acetylesterase is specific for homogalacturonans acetylated at O-3." Carbohydrate Polymers 107: 85-93.

Richard, P. and S. Hilditch (2009). "D-galacturonic acid catabolism in microorganisms and its biotechnological relevance." Applied Microbiology and Biotechnology 82(4): 597-604.

Richardson, A. J., *et al*. (1989). "Simultaneous determination of volatile and non-volatile acidic fermentation products of anaerobes by capillary gas-chromatography." Letters in Applied Microbiology 9:5-8.

Ridley, B. L., *et al*. (2001). "Pectins: structure, biosynthesis, and oligogalacturonide-related signaling." Phytochemistry 57(6): 929-967.

Rincon, M. T., *et al*. (2004). "ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the *Ruminococcus flavefaciens* 17 cellulosome." Journal of Bacteriology 186(9): 2576-2585.

Rincon, M. T., *et al*. (2005). "Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface." Journal of Bacteriology 187(22): 7569-7578.

Rincon, M. T., *et al*. (2007). "A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of *Ruminococcus flavefaciens*." Journal of Bacteriology 189(13): 4774-4783.

Robert, C. and A. Bernalier-Donadille (2003). "The cellulolytic microflora of the human colon: evidence of microcrystalline cellulose-degrading bacteria in methane-excreting subjects." Fems Microbiology Ecology 46(1): 81-89.

Robinson, R. W. and C. R. Spotts (1983). "The ultrastructure of sporulation in *Sporosarcina-Ureae*." Canadian Journal of Microbiology 29(7): 807-814.

Roy, C., *et al*. (1999). "Modes of action of five different endopectate lyases from *Erwinia chrysanthemi* 3937 (vol 181, pg 3705, 1999)." Journal of Bacteriology 181(18): 5889-5889.

Rye, C. S. and S. G. Withers (2000). "Glycosidase mechanisms." Current Opinion in Chemical Biology 4(5): 573-580.

Saitou, N. and M. Nei (1987). "The neighbor-joining method - a new method for reconstructing phylogenetic trees." Molecular Biology and Evolution 4(4): 406-425.

Sakamoto, T. and J. F. Thibault (2001). "Exo-arabinanase of *Penicillium chrysogenum* able to release arabinobiose from α-1,5-L-arabinan." Applied and Environmental Microbiology 67(7): 3319-3321.

Sakamoto, T., *et al*. (2011). "Identification of a GH62 α-L-arabinofuranosidase specific for arabinoxylan produced by *Penicillium chrysogenum*." Applied Microbiology and Biotechnology 90(1): 137-146.

Sakamoto, T., *et al*. (2013). "Biochemical characterization of a GH53 endo-β-1,4-galactanase and a GH35 exo-β-1,4-galactanase from *Penicillium chrysogenum*." Applied Microbiology and Biotechnology 97(7): 2895-2906.

Sakiyama, C. C. H., *et al*. (2001). "Characterization of pectin lyase produced by an endophytic strain isolated from coffee cherries." Letters in Applied Microbiology 33(2): 117

Saleem, F., *et al*. (2013). "The bovine ruminal fluid metabolome." Metabolomics 9(2): 360-378.

Salyers, A. A., *et al*. (1977a). "Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon." Applied Environmental Microbiology. 34:529 –533.

Salyers, A. A., *et al*. (1977b). "Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. Applied Environmental Microbiology." 33:319 –322.

Sambrook, J. *et al*. (1989). Molecular cloning: A laboratory manual 2nd edition. Cold Spring Harbor Laboratory Press 3, E3 - E4.

Sara, M. (2001). "Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria?" Trends in Microbiology 9(2): 47-49.

Sarkar, P., *et al*. (2009). "Plant cell walls throughout evolution: towards a molecular understanding of their design principles." Journal of Experimental Botany 60(13): 3615-3635.

Sato, M. D., *et al*. (2011). "Chemical and instrumental characterization of pectin from dried pomace of eleven apple cultivars." Acta Scientiarum-Agronomy 33(3): 383-389.

Schaeffer, A. B. and Fulton, M. D. (1933). "Simplified method of staining endospores." *Science* 77:194.

Scheller, H. V. and P. Ulvskov (2010). "Hemicelluloses." Annual Review of Plant Biology, Vol 61 61: 263-289.

Schultz, J., *et al*. (2000). "SMART: a web-based tool for the study of genetically mobile domains." Nucleic Acids Research 28(1): 231-234.

Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." Bioinformatics 30(14): 2068-2069.

Servinsky, M. D., *et al*. (2012). "Arabinose is metabolized via a phosphoketolase pathway in *Clostridium acetobutylicum* ATCC 824." Journal of Industrial Microbiolology and Biotechnology. 39: 1859–1867.

Shao, W. L., *et al*. (1995). "A high molecular weight, cell-associated xylanase isolated from exponentially growing *Thermoanaerobacterium* Sp strain Jw/Sl-Ys485." Applied and Environmental Microbiology 61(3): 937-940.

Sharon, I. and J. F. Banfield (2013). "Genomes from metagenomics." Science 342(6162): 1057-1058.

Sharpton, T. J. (2014). "An introduction to the analysis of shotgun metagenomic data." Frontiers in Plant Science 5.

Sheridan, P. O., *et al*. (2016). "Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*." Microbial Genomics, Published online. doi: 10.1099/mgen.0.000043.

Shevchik, V. E. and N. Hugouvieux-Cotte-Pattat (2003). "PaeX, a second pectin acetylesterase of *Erwinia chrysanthemi* 3937." Journal of Bacteriology 185(10): 3091-3100.

Shevchik, V. E., *et al*. (1996). "Characterization of pectin methylesterase B, an outer membrane lipoprotein of *Erwinia chrysanthemi* 3937." Molecular Microbiology 19(3): 455-466.

Shevchik, V. E., *et al*. (1997). "Pectate lyase PelI of *Erwinia chrysanthemi* 3937 belongs to a new family." Journal of Bacteriology 179(23): 7321-7330.

Shevchik, V. E., *et al*. (1999). "The exopolygalacturonate lyase PelW and the oligogalacturonate lyase Ogl, two cytoplasmic enzymes of pectin catabolism in *Erwinia chrysanthemi* 3937." Journal of Bacteriology 181(13): 3912-3919.

Shevchik, V. E., *et al*. (1999b). "Characterization of the exopolygalacturonate lyase PelX of *Erwinia chrysanthemi* 3937." Journal of Bacteriology 181(5): 1652-1663.

Shipkowski, S. and J. E. Brenchley (2006). "Bioinformatic, genetic, and biochemical evidence that some glycoside hydrolase family 42 beta-galactosidases are arabinogalactan type I oligomer hydrolases." Applied and Environmental Microbiology 72(12): 7730-7738.

Shipman, J. A., *et al*. (1999). "Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*." Journal of Bacteriology 181(23): 7206-7211.

Shipman, J. A., *et al*. (2000). "Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*." Journal of Bacteriology 182(19): 5365-5372.

Shiratori, H., *et al*. (2008). "*Lutispora thermophila* gen. nov., sp nov., a thermophilic, spore-forming bacterium isolated from a thermophilic methanogenic bioreactor digesting municipal solid wastes." International Journal of Systematic and Evolutionary Microbiology 58: 964-969.

Shiratori, H., *et al*. (2009). "*Clostridium clariflavum* sp nov and *Clostridium caenicola* sp nov., moderately thermophilic, cellulose-/cellobiose-digesting bacteria isolated from methanogenic sludge." International Journal of Systematic and Evolutionary Microbiology 59: 1764-1770.

Silva, I. R., *et al*. (2016). "Rhamnogalacturonan I modifying enzymes: an update." New Biotechnology 33(1): 41-54.

Simoneit, B. R. T., *et al*. (1999). "Levoglucosan, a tracer for cellulose in biomass burning and atmospheric particles." Atmospheric Environment 33(2): 173-182.

Slavin, J. L., *et al*. (1981). "Neutral detergent fiber, hemicellulose and cellulose digestibility in human subjects." J Nutr 111(2): 287-297.

Somerville, C., *et al*. (2004). "Toward a systems approach to understanding plant-cell walls." Science 306(5705): 2206-2211.

Soriano, M., *et al*. (2000). "An unusual pectate lyase from a *Bacillus* sp with high activity on pectin: cloning and characterization." Microbiology-Sgm 146: 89-95.

Spiller, G. A., *et al*. (1980). "Effect of purified cellulose, pectin, and a low-residue diet on fecal volatile fatty acids, transit time, and fecal weight in humans." Am J Clin Nutr 33(4): 754-759.

St. John, F. J., *et al*. (2006). "*Paenibacillus* sp. strain JDR-2 and XynA1: a novel system for methylglucuronoxylan utilization." Applied and Environmental Microbiology 72(2): 1496-1506.

St. John, F. J., *et al*. (2012). "Novel structural features of xylanase A1 from *Paenibacillus* sp JDR-2." Journal of Structural Biology 180(2): 303-311.

Suen, G., *et al*. (2011). "Complete genome of the cellulolytic ruminal bacterium *Ruminococcus albus* 7." Journal of Bacteriology 193(19): 5574-5575.

Sund, C. J., *et al*. (2015). "Phosphoketolase flux in *Clostridium acetobutylicum* during growth on L-arabinose." Microbiology 161: 430–440.

Talens-Perales, D., *et al*. (2016). "Analysis of domain architecture and phylogenetics of family 2 glycoside hydrolases (GH2)." Plos One 11(12).

Tamura, K. and Nei, M. (1993). “Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees.” Molecular Biology and Evolution 10:512-526.

Tamura, K., *et al*. (2004). "Prospects for inferring very large phylogenies by using the neighbor-joining method." Proceedings of the National Academy of Sciences of the United States of America 101(30): 11030-11035.

Tanaka, K., *et al*. (2002). “Two different pathways for D-xylose metabolism and the effect of xylose concentration on the yield coefficient of L-lactate in mixed-acid fermentation by the lactic acid bacterium *Lactococcus lactis* IO-1.” Applied Microbiology and Biotechnology. 60: 160–167.

Tardy, F., *et al*. (1997). "Comparative analysis of the five major *Erwinia chrysanthemi* pectate lyases: Enzyme characteristics and potential inhibitors." Journal of Bacteriology 179(8): 2503-2511.

Thompson, J., *et al*. (2013). "α-galacturonidase(s): A new class of Family 4 glycoside hydrolases with strict specificity and a unique CHEV active site motif." Febs Letters 587(6): 799-803.

Tierny, Y., *et al*. (1994). "Molecular cloning and expression in *Escherichia coli* of genes encoding pectate lyase and pectin methylesterase activities from *Bacteroides thetaiotaomicron*." Journal of Applied Bacteriology 76(6): 592-602.

Timmons, M. D., *et al*. (2009). "Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains." Applied Microbiology and Biotechnology 82(5): 929-939.

Tomazetto, G., *et al*. (2016). "Complete genome analysis of *Clostridium bornimense* strain M2/40(T): A new acidogenic *Clostridium* species isolated from a mesophilic two-phase laboratory-scale biogas reactor." Journal of Biotechnology 232: 38-49.

Touchon, M., *et al*. (2016). "Genetic and life-history traits associated with the distribution of prophages in bacteria." Isme Journal 10(11): 2744-2754.

Turnbaugh, P. J., *et al*. (2007). "The human microbiome Project." Nature 449(7164): 804-810.

Turnbaugh, P. J., *et al*. (2009a). "A core gut microbiome in obese and lean twins." Nature 457(7228): 480-U487.

Turnbaugh, P. J., *et al*. (2009b). "The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice." Science Translational Medicine 1(6).

Tylor, E. J., *et al*. (2006). “Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum*.” Biochem J 395(1):31-37.

Ussery, D. W., *et al*. (2009). “Computing for comparative microbial genomics: Bioinformatics for microbiologists”. London: Springer.

Vaaje-Kolstad, G., *et al*. (2010). “An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharide.” Science 330: 219-222.

Van Bueren, A. L., *et al*. (2017). "Prebiotic galactooligosaccharides activate mucin and pectic galactan utilization pathways in the human gut symbiont *Bacteroides thetaiotaomicron*." Scientific Reports 7.

Van Domselaar, G. H., *et al*. (2005). "BASys: a web server for automated bacterial genome annotation." Nucleic Acids Research 33: W455-W459.

Van Laere, K. M. J., *et al*. (1997). "A new arabinofuranohydrolase from *Bifidobacterium adolescentis* able to remove arabinosyl residues from double-substituted xylose units in arabinoxylan." Applied Microbiology and Biotechnology 47(3): 231-235.

Venditto, I., *et al*. (2016). "The complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition." Glycobiology 26(12): 1491-1492.

Venkataraman, A., *et al*. (2016). "Variable responses of human microbiomes to dietary supplementation with resistant starch." Microbiome 4.

Viborg, A. H., *et al*. (2014). "A β 1-6/β 1-3 galactosidase from *Bifidobacterium animalis* subsp *lactis* Bl-04 gives insight into sub-specificities of β-galactoside catabolism within *Bifidobacterium*." Molecular Microbiology 94(5): 1024-1040.

Villares, A., *et al*. (2017). "Lytic polysaccharide monooxygenases disrupt the cellulose fibres structure." Scientific Reports. 7: 40262.

Vincken, J. P., *et al*. (1994). "The effect of xyloglucans on the degradation of cell wall-embedded cellulose by the combined action of cellobiohydrolase and endoglucanases from *Trichoderma viride*." Plant Physiology 104(1): 99-107.

Voigt, C., *et al*. (2014). "Identification of PTSFru as the major fructose uptake system of *Clostridium acetobutylicum*." Applied Microbiology and Biotechnology 98(16): 7161-7172.

Voragen, A. G. J., *et al*. (1986). "Determination of the degree of methylation and acetylation of pectins by HPLC." Food Hydrocolloids 1(1): 65-70.

Voragen, A. G. J., *et al*. (2009). "Pectin, a versatile polysaccharide present in plant cell walls." Structural Chemistry 20(2): 263-275.

Waeonukul, R., *et al*. (2009). "Cloning, sequencing, and expression of the gene encoding a multidomain endo-β-1,4-xylanase from *Paenibacillus curdlanolyticus* B-6, and characterization of the recombinant enzyme." Journal of Microbiology and Biotechnology 19(3): 277-285.

Walker, A. W., *et al*. (2011). "Dominant and diet-responsive groups of bacteria within the human colonic microbiota." Isme Journal 5(2): 220-230.

Wang, W., *et al*. (2014). "Elucidation of the molecular basis for arabinoxylan-debranching activity of a thermostable family GH62 α-L-arabinofuranosidase from *Streptomyces thermoviolaceus*." Applied and Environmental Microbiology 80(17): 5317-5329.

Wee, M. S. M., *et al*. (2014). Structure of a shear-thickening polysaccharide extracted from the New Zealand black tree fern, *Cyathea medullaris*. International Journal of Biological Macromolecules 70: 86–91.

Wegmann, U., *et al*. (2014). "Complete genome of a new *Firmicutes* species belonging to the dominant human colonic microbiota ('*Ruminococcus bicirculans*') reveals two chromosomes and a selective capacity to utilize plant glucans." Environmental Microbiology 16(9): 2879-2890.

Wei, H., *et al*. (2009). "Natural paradigms of plant cell wall degradation." Curr Opin Biotechnol 20(3): 330-338.

Westereng, B., *et al*. (2015). “Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer.” Scientific Reports 5: 18561.

White, B. A., *et al*. (2014). “Biomass utilization by gut microbiomes.” Annual Review of Microbiology 68:279-296.

Wilkens, C., *et al*. (2016). "An efficient arabinoxylan-debranching α-L-arabinofuranosidase of family GH62 from *Aspergillus nidulans* contains a secondary carbohydrate binding site." Applied Microbiology and Biotechnology 100(14): 6265-6277.

Willats, W. G. T., *et al*. (2001). "Modulation of the degree and pattern of methyl-esterification of pectic homogalacturonan in plant cell walls - Implications for pectin methyl esterase action, matrix properties, and cell adhesion." Journal of Biological Chemistry 276(22): 19404-19413.

Willats, W. G. T., *et al*. (2006). "Pectin: new insights into an old polymer are starting to gel." Trends in Food Science & Technology 17(3): 97-104.

Williams, P. A., *et al*. (2005). "Elucidation of the emulsification properties of sugar beet pectin." Journal of Agricultural and Food Chemistry 53(9): 3592-3597.

Wolfe, A. J. (2005). "The acetate switch." Microbiology and Molecular Biology Reviews 69(1): 12.

Wu, C. H., *et al*. (2004). "PIRSF: family classification system at the Protein Information Resource." Nucleic Acids Research 32: D112-D114.

Xu, Q., *et al*. (2003). “The Cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein.” Journal of Bacteriology 185(15): 4548-4557.

Yang, J. C., *et al*. (1990). “*Clostridium aldrichii* sp. nov., a cellulolytic mesophile inhabiting a wood-fermenting anaerobic digester. International Journal of Systematic Bacteriology 40:268-272.

Yang, X., *et al*. (2009). "More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body." Plos One 4(6).

Yapo, B. M. (2011). “Pectin rhamnogalacturonan II: on the “small stem with four branches” in the primary cell walls of plants.” International Journal of Carbohydrate Chemistry. 2011:1–11.

Ye, J., *et al*. (2012). "Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction." Bmc Bioinformatics 13.

Yin, X., *et al*. (2005). “. The gene encoding xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (*xfp*) is conserved among *Bifidobacterium* species within a more variable region of the genome and both are useful for strain identification.” FEMS Microbiology Letter. 246: 251 – 257.

Yin, Y. B., *et al*. (2012). "dbCAN: a web resource for automated carbohydrate-active enzyme annotation." Nucleic Acids Research 40(W1): W445-W451.

Yuliarti, O., *et al*. (2015). "Characterization of gold kiwifruit pectin from fruit of different maturities and extraction methods." Food Chemistry 166: 479-485.

Ze, X. L., *et al*. (2012). "*Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon." ISME Journal 6(8): 1535-1543.

Ze, X. L., *et al*. (2015). "Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic *Firmicutes* bacterium *Ruminococcus bromii*." Mbio 6(5).

Zhou, Y., *et al*. (2011). "PHAST: A Fast Phage Search Tool." Nucleic Acids Research 39: W347-W352.

Zykwinska, A. W., *et al*. (2005). "Evidence for *in vitro* binding of pectin side chains to cellulose." Plant Physiology 139(1): 397-407.

Appendices

Appendix 1 CAZymes involved in the degradation of HG, RG-I, and RG-II.

Enzyme	Action	Target	Function	EC	CAZy category
		Homogalacturonan (HG)			
Pectate lyase	Endo	α-1,4-D-Gal*p*A	Cleave α-1,4-D-galacturonan to release Δ4,5-unsaturated galacturonic acid residues	4.2.2.2	PL1, 2, 3, 9, 10
Pectate lyase	Exo	α-1,4-D-Gal*p*A	Cleave α-1,4-D-galacturonan to release Δ4,5- unsaturated digalacturonic acid residues	4.2.2.9	PL1, 2, 9
Pectate lyase	Exo	α-1,4-D-Gal*p*A	Cleave the reducing ends of PGA to release trigalacturonates as a major product	4.2.2.22	PL1, 2, 9
Polymethyl galacturonate lyase	Endo	Unsaturated polymethyl-D-digalacturonates	Endo-cleave α-1,4-D-galacturonan methyl ester to release unsaturated monogalacturonates	4.2.2.10	PL1
polymethyl galacturonate lyase	Exo	Unsaturated poly methyl-D-digalacturonates	Exo-cleavage of α-1,4-D-galacturonan methyl ester to release unsaturated methyl-monogalacturonates	4.2.2.10	PL1
Oligogalacturonide lyase	Exo	Digalacturonate Δ4,5- unsaturated digalacturonate	Exo-cleavage of DKI from short chain oligogalacturonides	4.2.2.6	PL22
Polygalacturonase	Exo	Low-methyl α-1,4-D-galacturonic acid	Hydrolyse the chain ends of PGA to release monogalacturonates	3.2.1.67	GH4
Polygalacturonase	Endo	Low-methyl α-1,4-D-galacturonic acid	Internally hydrolyse PGA into oligogalacturonates	3.2.1.15	GH28
Oligogalacturonate hydrolase	Exo	Trigalacturonate	Hydrolyse the chain ends of trigalacturonate to release monogalacturonates	3.2.1.67	GH28
Digalacturono hydrolase	Exo	Low-methyl α-1,4-D- Gal*p*A	Hydrolyse PGA from non-reducing ends to release digalacturonates	3.2.1.82	GH28

Methyl galacturonase	Exo	High-methyl α-1,4-D- Gal*p*A	Hydrolyse the chain ends of PGA to release oligogalacturonates	3.2.1.82	GH28
Methyl galacturonase	Endo	High-methyl α-1,4-D- Gal*p*A	Hydrolyse the internal chain of PGA into methyl-esterified oligogalacturonates	3.2.1.82	GH28
Xylogalacturonan hydrolase	Endo	Xylogalacturonan	Cleave glycosidic linkages between two Gal*p*A substituted with D-Xyl*p* at O3 to release dimers of α-D-Gal*p*A-β-1,4-D-Xyl*p*	3.2.1.-	GH28
Pectin methyl esterase	Exo	Methoxy groups esterified to HG	De-esterify pectin by removing methoxy esters to release pectin and methanol	3.1.1.11	CE8
Pectin acetyl esterase	Exo	Acetyl groups esterified to HG	De-esterify pectin by removing esterified acetyl groups	3.1.1.-	CE12
Pectin acetyl esterase	Exo	Acetyl groups at esterified to HG	De-esterify pectin by removing esterified acetyl groups	3.1.1.-	CE13
			Rhamnogalacturonan-I (RG-I)		
Rhamnogalacturonan lyase	Endo	L-α-Rha*p*-1,4-α-D-Gal*p*A	Endo-cleavage of L-α-Rha*p*-1,4-α-D-Gal*p*A to release oligosaccharides containing L-Rha*p* at reducing ends and unsaturated D-Gal*p*A acid at non-reducing ends	4.2.2.23	PL4, 11
Rhamnogalacturonan lyase	Exo	L-α-Rha*p*-1,4-α-D-Gal*p*A	Exo-cleavage of the break down products of RG endo-lyase. The products are α-L-disaccharide rhamnose and RG oligosaccharides containing unsaturated D-Gal*p*A at non-reducing ends.	4.2.2.24	PL4, 11
Rhamnogalacturonase	Endo	α-D-Gal*p*A-1,2-α-L-Rha*p*	Endo-hydrolysis of α-D-Gal*p*A-1,2-α-L-Rha*p* to release oligogalacturonates with β-D-GalA at reducing end	3.2.1.171	GH28
Rhamnogalacturonan-galacturonohydrolase	Exo	α-D-Gal*p*A-1,2-α-L-Rha*p*	Exo-hydrolysis of α-D-Gal*p*A-1,2-α-L-Rha*p* to release D-Gal*p*A from non-reducing end	3.2.1.173	GH28

Rhamnogalacturonan-rhamnohydrolase	Exo	α-L-Rha*p*-1,4-α-D-Gal*p*A	Exo-hydrolysis of α-L-Rha*p*-1,4-α-D-Gal*p*A to release β-L-Rha*p* from non-reducing end	3.2.1.174	GH28, 106
Unsaturated rhamnogalacturonyl hydrolase	Exo	Unsaturated rhamnogalacturonan	Hydrolysis of unsaturated RG to release a galacturonate monomer and L-Rha*p*	3.2.1.172	GH105
β-galactosidase	Exo	Terminal β-D-Gal*p*	Hydrolysis of terminal non-reducing β-D-Gal*p* residues in β-D-galactosides	3.2.1.23	GH1, 2, 35, 42, 59
α-L-arabinofuranosidase	Exo	Terminal α-L-Ara*f*	Hydrolysis of terminal non-reducing α-L-Ara*f* residues in α-L-arabinosides	3.2.1.55	GH2, 3, 43, 51, 54, 62
Arabinanase	Endo	Linear 1,5-α-L-arabinan	Endo-hydrolysis of α-1,5-L-Ara*f* linkages in (1→5)-arabinans	3.2.1.99	GH43
β-1,4-galactanase	Endo	Arabinogalactan I	hydrolyses β-1,4-D-galactosidic linkages in type I arabinogalactans to release a tetrasaccharide	3.2.1.89	GH53
α-L-1,5-arabinanase	Exo	Linear α-1,5-L- Ara*f*	Hydrolysis of terminal Ara*f* residues in α-1,5-linked linear L-arabinosides	3.2.1.-	GH93
Rhamnogalacturonan acetylesterase	Exo	Acetyl groups at esterified to RG-I	De-esterify pectin by removing esterified acetyl groups	3.1.1.-	CE12

Rhamnogalacturonan-II (RG-II)

Pectate lyase	Endo	α-1,4-D-Gal*p*A	Cleave long α-1,4-D-Gal*p*A oligosaccharides	4.2.2.2	PL1
Polygalacturonase	Endo	α-1,4-D-Gal*p*A	Release D-Gal*p*A	3.2.1.15	GH28
D-4,5-unsaturated β-glucuronyl hydrolase	Endo	α-1,4-D-Gal*p*A	Release unsaturated 4,5-anhydro-GalA	3.2.1.-	GH105
β-glucuronidase	Exo	L-Fuc*p*-β-1,4-D-Glc*p*A	Release D-Glc*p*A	3.2.1.31	GH2
β-L-arabinofuranosidase	Exo	L-Rha*p*-β-2,1-L-Ara*f*	Release L-Ara*f*	3.2.1.55	GH2
α-L-arabinopyranosidase	Exo	D-Gal*p*-α-4,1-L-Ara*p*	Release L-Ara*p*	ND	GH2
β-D-galacturonidase	Exo	L-Rha*p*-β-3,1-D-Gal*p*A	Release D-Gal*p*A	ND	GH2
β-galactosidase	Exo	L-aceric acid-β-1,2-D-Gal*p*	Release D-Gal*p*	3.2.1.23	GH2
2-keto-3-deoxynononic acid hydrolase	Exo	D-Gal*p*A-α-2,3-D-Kdo	Release D-Kdo	3.2.1.-	GH33

α-L-arabinofuranosidase	Exo	D-Gal*p*A-α-1,3-L-Ara*f*	Release L-Ara*f*	3.2.1.55	GH43
α-L-rhamnosidase	Endo	D-Api*f*-α-1,3-L-Rha*p*	Release L-Rha*p*	3.2.1.40	GH78
α-L-rhamnosidase	Exo	L-Ara*p*-α-3,1-L-Rha*p*	Release L-Rha*p*	3.2.1.40	GH78
α-L-rhamnosidase	Exo	D-Kdo-α-1,5-L-Rha*p*	Release L-Rha*p*	3.2.1.40	GH78
α-L-galactosidase	Exo	D-Glc*p*A-α1,2-L-Gal*p*	Release L-Gal*p*	3.2.1.-	GH97
D-4,5-unsaturated β-glucuronyl hydrolase	Exo	Terminal D-Glc*p*A	Cleave the unsaturated oligomers at a glucuronyl residue at the non-reducing end	3.2.1.-	GH88
α-L-rhamnosidase	Exo	L-Ara*p*-α-2,1-L-Rha*p*	Release L-Rha*p*	3.2.1.40	GH106
3-*C*-carboxy-5-deoxy-L-xylose hydrolase	Exo	L-Rha*p*-α-1,3-L-aceric acid	Release L-aceric acid	ND	GH127
β-L-arabinofuranosidase	Exo	L-Ara*f*-β-1,5-D-DHA	Release L-Ara*f*	3.2.1.185	GH137, 142
α-galacturonidase	Exo	L-Rha*p*-α-1,2-D-Gal*p*A	Release D-Gal*p*A	3.2.1.-	GH138
α-2-methyl-fucosidase	Exo	2-*O*-Me-L-Fuc*p*-α-1,2-D-Gal*p*	Release 2-*O*-Me-L-Fuc*p*	3.2.1.-	GH139
Apiosidase	Endo	D-Gal*p*A-β-1,2-D-Api*f*	Release L-Rha*p*-α-1,3-D-Api*f*	3.2.1.-	GH140
α-L-fucosidase	Endo	L-Rha*p*-α-1,4-L-Fuc*p*	Release D-MeXyl*p*-L-Fuc*p*	3.2.1.51	GH141
DHA hydrolase	Exo	D-DHA-β-2,3-D-Gal*p*A	Release D-DHA	3.2.1.-	GH143

Appendix 2 CBM domains associated with binding the pectic polysaccharides or the monomeric components of pectin

CBM family	Type	Fold	Ligand	Associated enzyme	CAZy category	Organism	References
CBM13	C	β-trefoil	Xylose, xylan, and galactose	Endo-β-1,4,-D-xylanase (EC 3.2.1.8)	GH10	*Streptomyces olivaceoviridis*	Fujimoto *et al*., 2002
CBM32	C	β-sandwich	Galactose	Sialidase (EC 3.2.1.18)	GH33, GH34	*Micromonospora viridifaciens,*	Newstead *et al*., 2005
			Pectin	None	NA	*Yersinia enterolitica*	Abbott *et al*., 2007
CBM35	ND	β-sandwich	Δ4,5-anhydrogalacturonic acid	Pectate lyase (EC 4.2.2.2)	PL10	Environmental isolate	
				Rhamnogalacturonan acetyl esterase (EC 3.1.1.-)	CE12	*Ruminiclostridium thermocellum*	Montanier *et al*., 2009
				Endo-β-1,4-xylanase (EC 3.2.1.8)	GH10	*Cellvibrio japonicus*	
CBM42	C	β-trefoil	Arabinofuranose	α-L-arabinofuranosidase (EC 3.2.1.55)	GH43	*Streptomyces avermitilis*	Fujimoto *et al*., 2008
				α-L-arabinofuranosidase (EC 3.2.1.55)	GH51, 54	*Aspergillus kawachii*	Miyanaga *et al*., 2004
CBM60	ND	β-trefoil	Galactan	Xylanase	GH11	*Cellvibrio japonicus*	Montanier *et al*., 2011
CBM61	ND	β-sandwich	Galactan	Endo-β-1,4-galactanase (EC 3.2.1.89)	GH53	*Thermotoga maritima*	Cid *et al*., 2010

CBM62	ND	β-sandwich	Galactose and arabinose from arabinogalactan	Endo-β-1,4-xylanase (EC 3.2.1.8)	GH5	*Ruminiclostridium thermocellum*	Montanier *et al*., 2011
CBM67	ND	β-sandwich	L-rhamnose	α-L-rhamnosidases (EC 3.2.1.40)	GH78	*Streptomyces avermitilis*	Fujimoto *et al*., 2013
CBM77	B	β-sandwich	Homogalacturonan	Pectate lyase (EC 4.2.2.2)	PL1, PL9	*Ruminococcus flavefaciens*	Venditto *et al*., 2016

Appendix 3 Pectin-degrading bacteria reported from the human gut

Phylum	Genus/species	Source	Isolation substrate	Reference
Firmicutes	*Clostridium butyricum-Clostridium beijerinckii* group	Human faeces	Pectic acid (0~75% methylated galacturonate units)	Matsuura, 1991 Nakajima *et al*., 1999
	Faecalibacterium prausnitzii	Human faeces	Apple pectin	Lopez-Siles *et al*., 2012
	Eubacterium eligens	Human faeces	Commercial pectin	Salyers *et al*., 1977
Bacteroidetes	*Bacteroides thetaiotaomicron*	Human faeces		Tierny *et al*., 1994
	Bacteroides cellulolyticus	Human faeces	Citrus pectin	Robert *et al*., 2007
	Bacteroides sp.	Human faeces	Polygalacturonate (Sigma)	Jensen and Canale-Parola, 1986; MacCarthy *et al*., 1985
	Bacteroides sp.		Commercial pectin	Betian *et al*., 1977
	Bacteroides distasonis	Human faeces	Citrus pectin	Bayliss and Houston, 1984
	Bacteroides ovatus	Human faeces	Citrus pectin	Bayliss and Houston, 1984; Salyers *et al*., 1977b
	Bacteroides caccae	Human faeces	Citrus pectin	Salyers *et al*., 1977b
	Bacteroides fragilis	Human faeces	Citrus pectin	Salyers *et al*., 1977b
	Bacteroides vulgatus	Human faeces	Citrus pectin	Salyers *et al*., 1977b
	Bacteroides pectinophilus	Human faeces	Polygalacturonate (Sigma)	Jensen and Canale-Parola, 1986
	Bacteroides galacturonicus	Human faeces	Polygalacturonate (Sigma)	Jensen and Canale-Parola, 1986

Appendix 4 Uncultured bacterial 16S rRNA gene sequences from GenBank with ≥ 92 % sequence similarity to strain 14^T.

GenBank Accession number	Source	Sequence similarity (%)	references
AB606397	mice gut	93.9	Muramatsu and Matsumoto (unpublished)
AB821761	cattle rumen	96.0	Koike (unpublished)
AF371829	swine gut	94.5	Leser *et al.*, 2002
AY916322	human gut	99.6	Eckburg *et al.*, 2005
AY982878	human gut	98.7	Gill *et al.*, 2006
DQ326608	human gut	99.7	
DQ326785	human gut	93.7	
DQ326967	human gut	99.4	
DQ327275	human gut	99.6	
DQ327290	human gut	99.6	
DQ327452	human gut	99.6	
DQ793380	human gut	95.5	Ley *et al.*, 2006
DQ799719	human gut	97.6	
DQ799841	human gut	99.5	
DQ801387	human gut	92.5	
DQ803358	human gut	95.3	
DQ804302	human gut	99.6	
DQ805369	human gut	95.4	
DQ815397	mouse caecum	92.6	Rawls *et al.*, 2006
DQ905456	human gut	93.3	Mai *et al.*, 2006
EF445183	rabbit caecum	94.0	Monteils *et al.*, 2008
EU458641	rabbit caecum	96.0	Ley *et al.*, 2008
EU458659	human gut	92.9	
EU466217	human gut	93.5	
EU773353	human gut	93.9	
EU775919	human gut	93.6	
FJ363611	human gut	99.0	Turnbaugh *et al.*, 2009a
FJ363849	human gut	99.4	
FJ365218	human gut	99.0	
FJ365220	human gut	99.0	
FJ366606	human gut	99.1	
FJ366621	human gut	99.1	
FJ366692	human gut	98.6	
FJ366696	human gut	98.9	

FJ366736	human gut	99.1	
FJ367015	human gut	99.3	
FJ367259	human gut	98.8	
FJ367289	human gut	98.5	
FJ367425	human gut	98.7	
FJ367520	human gut	98.5	
FJ951890	gazelle rumen	93.5	Cai *et al.*, (unpublished)
GQ175428	chicken gut	95.3	Massias and Urdaci (unpublished)
GQ175455	chicken gut	94.9	
GQ493538	human gut	99.1	Turnbaugh *et al.*, 2009b
GQ493741	human gut	99.6	
HQ716289	swine gut	93.5	Kalmokoff *et al.*, 2011
HQ769805	human gut	97.0	Li *et al.*, 2012
JN559735	dhole gut	92.8	Chen *et al.*, 2016
JQ183603	human gut	93.2	Lin *et al.*, 2013
JQ184336	human gut	93.5	
JQ186052	human gut	93.5	
JQ186342	human gut	93.4	
JQ187594	human gut	93.4	
JQ187701	human gut	93.1	
JQ187787	human gut	93.4	
JQ187888	human gut	93.4	
JQ187917	human gut	93.4	
JQ188103	human gut	93.4	
JQ189828	human gut	93.4	
JQ190080	human gut	93.3	
JQ191117	human gut	93.4	
LC028697	cattle rumen	92.3	Kobayashi (unpublished)
LC028782	cattle rumen	92.0	

Appendix 5 *M. pectinilyticus* 14^T CDSs encoding putative enzymes and CBMs involved in polysaccharide degradation/modification.

	Locus tag #	CAZyme category	Predicted CAZyme function	PCW component	Length (aa)	Signal peptide	SLH	CBM
PL	B9O19_51	PL1	Pectate lyase	HG/RG-II	709	Yes	No	CBM13/CBM13
	B9O19_312	PL1	Pectate lyase	HG/RG-II	906	Yes	No	ND
	B9O19_428	PL9	Pectate lyase	HG/RG-II	759	Yes	No	ND
	B9O19_873	PL1	Pectate lyase	HG/RG-II	1230	Yes	Yes	ND
	B9O19_874	PL1	Pectate lyase	HG/RG-II	1525	Yes	No	ND
	B9O19_909	PL1	Pectate lyase	HG/RG-II	2071	Yes	Yes	ND
	B9O19_929	PL1	Pectate lyase	HG/RG-II	1022	Yes	No	ND
	B9O19_1128	PL1	Pectate lyase	HG/RG-II	1261	Yes	No	ND
	B9O19_1129	PL1	Pectate lyase	HG/RG-II	980	Yes	No	ND
	B9O19_1239	PL1	Pectate lyase	HG/RG-II	1000	Yes	No	ND
	B9O19_1241	PL1	Pectate lyase	HG/RG-II	604	Yes	No	ND
	B9O19_1315	PL1	Pectate lyase	HG/RG-II	771	Yes	No	ND
	B9O19_1377	PL1	Pectate lyase	HG/RG-II	841	Yes	No	ND
	B9O19_1442	PL1	Pectate lyase	HG/RG-II	1110	Yes	No	ND
	B9O19_1443	PL1	Pectate lyase	HG/RG-II	1201	Yes	No	ND
	B9O19_1615	PL9	Pectate lyase	HG/RG-II	581	Yes	No	ND
	B9O19_1653	PL11	Rhamnogalacturonan lyase	RG-I	1333	Yes	No	ND
	B9O19_1867	PL1	Pectate lyase	HG/RG-II	897	Yes	No	ND
	B9O19_2006	PL1	Pectate lyase	HG/RG-II	1080	Yes	No	ND
	B9O19_2062	PL1	Pectate lyase	HG/RG-II	1392	Yes	No	ND
	B9O19_2160	PL1	Pectate lyase	HG/RG-II	1146	Yes	No	ND
GH	B9O19_53	GH97	α-galactosidase	RG-II	795	Yes	Yes	ND
	B9O19_60	GH77	α-glucosidase		494	No	No	ND
	B9O19_229	GH13	α-glucosidase		437	No	No	ND

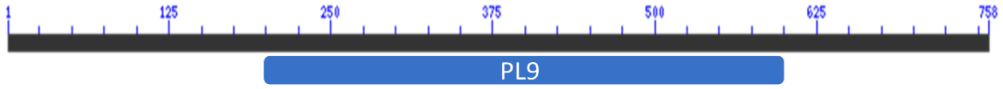
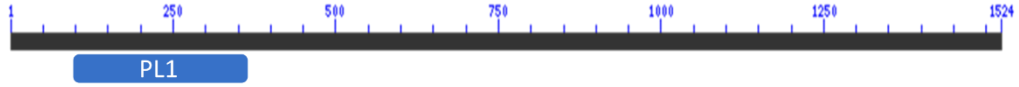
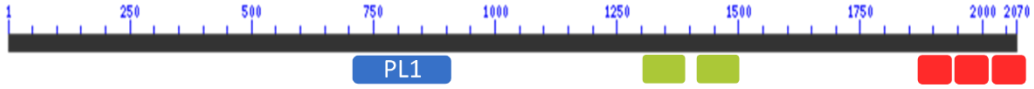
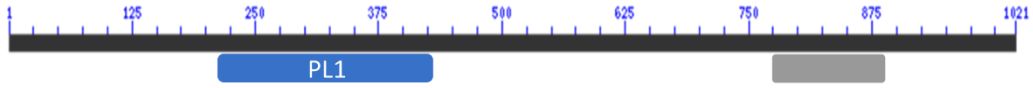
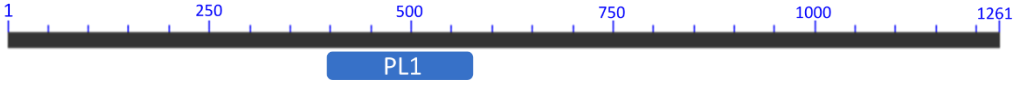
	B9O19_232	GH18	Chitinase		429	No	No	CBM50/CBM50
	B9O19_390	GH23	Lysozyme		215	LPS	No	ND
	B9O19_398	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	Hemicellulose	2019	Yes	No	CBM6
	B9O19_466	GH73	Lysozyme		693	No	No	ND
	B9O19_609	GH5	Endoglucanase	Cellulose	1141	Yes	No	ND
	B9O19_681	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	Hemicellulose	815	No	No	ND
	B9O19_804	GH43	β-xylosidase; α-arabinofuranosidase	RG-I/RG-II	638	No	No	ND
	B9O19_865	GH43	β-xylosidase; α-arabinofuranosidase	RG-I/RG-II	513	No	No	ND
	B9O19_868	GH28	Polygalacturonase	HG/RG-I	493	No	No	ND
	B9O19_956	GH140	endo-apiosidase	RG-II	434	No	No	ND
	B9O19_989	GH95	α-fucosidase	Hemicellulose	1305	Yes	No	ND
	B9O19_1043	GH13	α-glucosidase		687	No	No	CBM48
	B9O19_1110	GH13	α-glucosidase		664	No	No	CBM34
	B9O19_1304	GH43	β-xylosidase; α-arabinofuranosidase	RG-I/RG-II	512	No	No	ND
	B9O19_1438	GH95	α-fucosidase	Hemicellulose	833	Yes	No	ND
	B9O19_1543	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	Hemicellulose	1781	No	No	ND
	B9O19_1703	GH5	Endoglucanase	Cellulose	1125	Yes	No	ND
	B9O19_1731	GH13	α-glucosidase		611	No	No	CBM48
	B9O19_1828	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	Hemicellulose	2013	Yes	No	ND
	B9O19_2010	GH51	Endoglucanase	Cellulose	657	No	No	ND
	B9O19_2011	GH33	2-keto-3-deoxynononic acid hydrolase	RG-II	348	No	No	ND
	B9O19_2014	GH95	α-fucosidase	Hemicellulose	1386	Yes	No	ND
	B9O19_2023	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	Hemicellulose	494	Yes	No	ND
	B9O19_2087	GH116	β-glucosidase; β-xylosidase	Hemicellulose	1368	Yes	No	CBM6
	B9O19_2192	GH13	α-glucosidase		483	No	No	ND
	B9O19_2275	GH105	Unsaturated rhamnogalacturonyl hydrolase	RG-I	367	No	No	ND
CE	B9O19_203	CE12	Pectin acetylesterase	HG/RG-I	1095	Yes	Yes	ND
	B9O19_505	CE8	Pectin methylesterase	HG	1168	Yes	No	ND
	B9O19_680	CE8	Pectin methylesterase	HG	636	Yes	No	ND

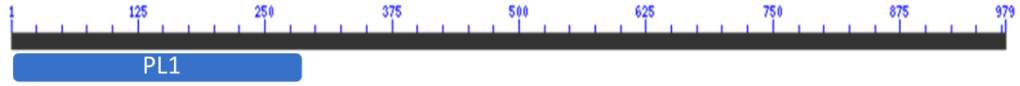
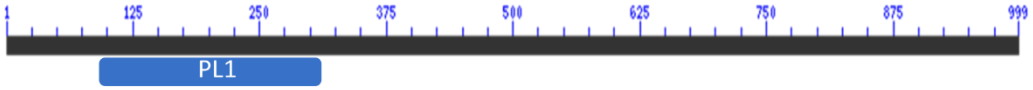
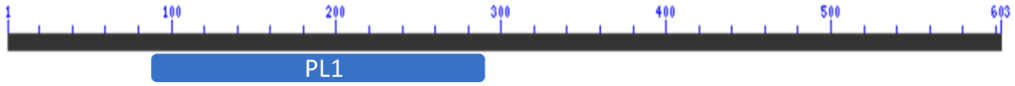
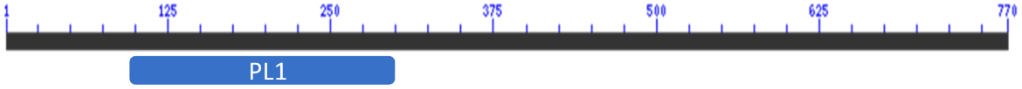
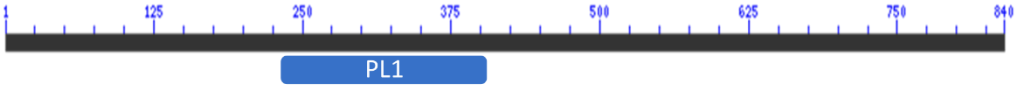
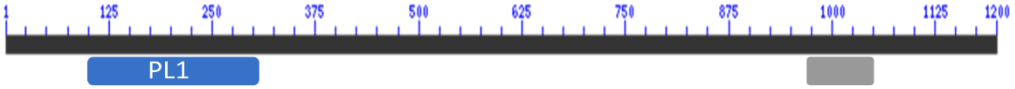
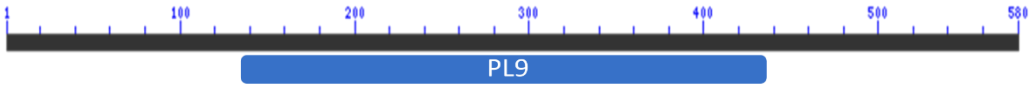
	B9O19_826	CE8	Pectin methylesterase	HG	1808	Yes	Yes	ND
	B9O19_869	CE8	Pectin methylesterase	HG	303	No	No	ND
	B9O19_884	CE12	Pectin acetylesterase	HG/RG-I	329	No	No	ND
	B9O19_973	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		367	Yes	No	ND
	B9O19_1338	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		318	Yes	No	ND
	B9O19_1397	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		272	LPS	No	ND
	B9O19_1528	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		399	Yes	No	ND
	B9O19_1626	CE8	Pectin methylesterase	HG	1430	Yes	Yes	ND
	B9O19_1627	CE8	Pectin methylesterase	HG	2959	Yes	Yes	ND
	B9O19_1817	CE8	Pectin methylesterase	HG	1153	Yes	No	ND
	B9O19_1829	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		422	Yes	No	ND
	B9O19_2015	CE8	Pectin methylesterase	HG	645	No	No	ND
	B9O19_2077	CE12	Pectin acetylesterase	HG/RG-I	363	No	No	ND
	B9O19_2119	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		302	LPS	No	ND
	B9O19_2231	CE12	Pectin acetylesterase	HG/RG-I	868	Yes	No	ND
	B9O19_2279	CE4	Peptidoglycan N-acetylmuramic acid deacetylase		270	LPS	No	ND
Multi-domain enzymes	B9O19_1299	GH95/CE8	α-fucosidase/pectin methylesterase	RG-II/HG	2430	Yes	No	ND
	B9O19_1681	GH95/CE8	α-fucosidase/pectin methylesterase	RG-II/HG	2000	Yes	No	ND
	B9O19_2263	CE12/CE8	Pectin acetylesterase/pectin methylesterase	HG/RG-I	2132	Yes	No	ND
GT	B9O19_261	GT51	Murein polymerase		683	LPS	No	ND
	B9O19_279	GT2	Poly-specific transferases		418	LPS	No	ND
	B9O19_280	GT2	Poly-specific transferases		420	LPS	No	ND
	B9O19_291	GT39	Protein O-mannosyltransferase		1163	LPS	No	ND
	B9O19_294	GT4	Poly-specific transferases		761	No	No	ND
	B9O19_594	GT2	Poly-specific transferases		406	LPS	No	ND
	B9O19_1041	GT2	Poly-specific transferases		354	No	No	ND
	B9O19_1042	GT2	Poly-specific transferases		327	No	No	ND
	B9O19_1046	GT5	Glycogen synthase		478	No	No	ND
	B9O19_1158	GT2	Poly-specific transferases		315	No	No	ND

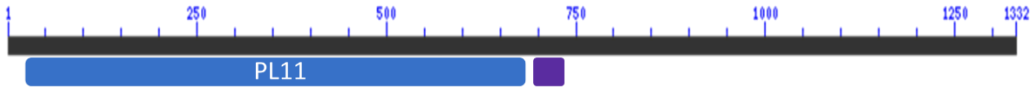
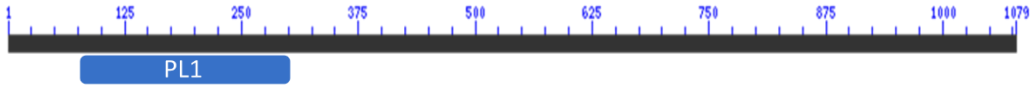
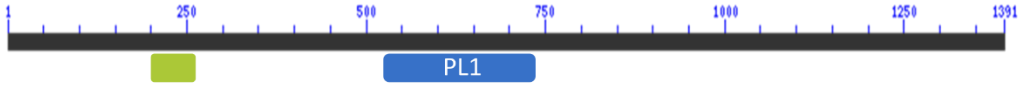
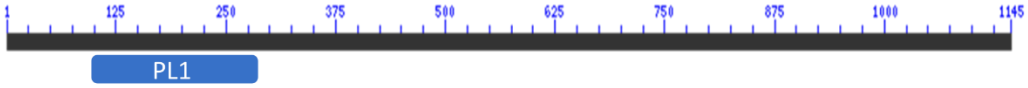
	B9O19_1163	GT2	Poly-specific transferases		325	No	No	ND
	B9O19_1226	GT26	UDP-ManNAcA: β-N-acetyl mannosaminuronyltransferase		246	No	No	ND
	B9O19_1237	GT2	Poly-specific transferases		320	No	No	ND
	B9O19_1359	GT28	Poly-specific transferases		371	No	No	ND
	B9O19_1363	GT28	Poly-specific transferases		371	No	No	ND
	B9O19_1596	GT35	Glycogen or starch phosphorylase		843	No	No	ND
	B9O19_1637	GT4	Poly-specific transferases		378	No	No	ND
	B9O19_1638	GT4	Poly-specific transferases		406	No	No	ND
	B9O19_1645	GT2	Poly-specific transferases		342	No	No	ND
	B9O19_1646	GT2	Poly-specific transferases		341	No	No	ND
	B9O19_1809	GT4	Poly-specific transferases		368	No	No	ND
Non-catalytic CBMs	B9O19_76	CBM50	Peptidoglycan-binding		203	Yes	No	-
	B9O19_1285	CBM50	Peptidoglycan-binding		565	No	No	-
	B9O19_1469	CBM13	Xylose/xylan/galactose-binding	Hemicellulose	732	Yes	No	-

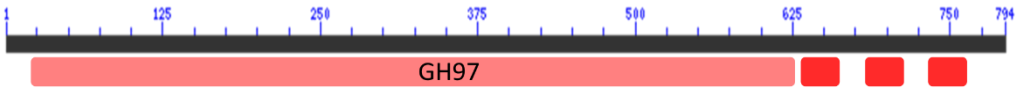
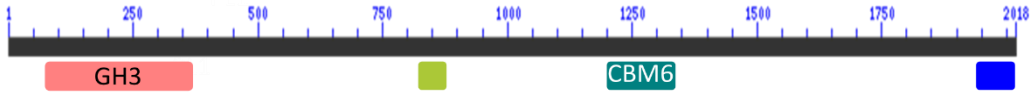
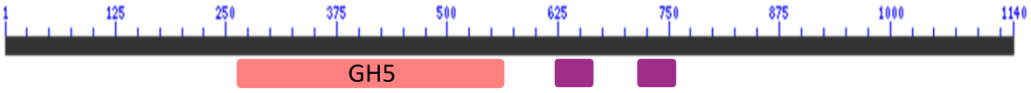
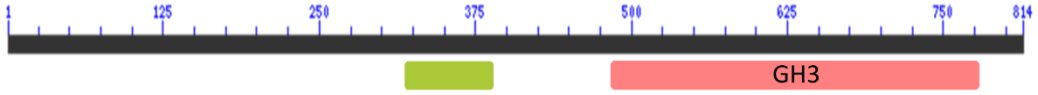
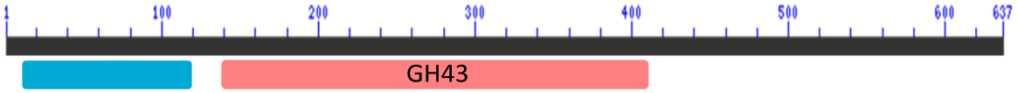
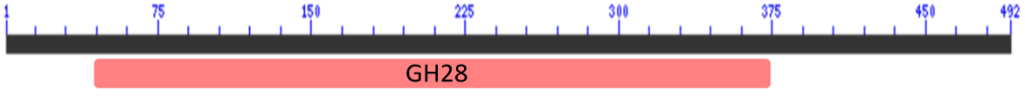
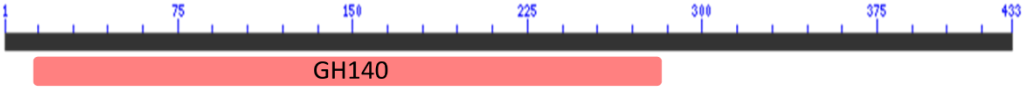
Abbreviations: SLH, S-layer homology; PCW, plant cell wall; ND, not detected; LPS, lipoprotein signal peptide

Appendix 6 Identifiable protein domain architectures in plant carbohydrate-associated PLs, GHs, and CEs of *M. pectinilyticus*. The numbers on the scale indicate the amino acid positions. Proteins are not drawn in scale. The CAZy family classification, predicted protein functions, and the presence/absence of secretory signal peptides are indicated.

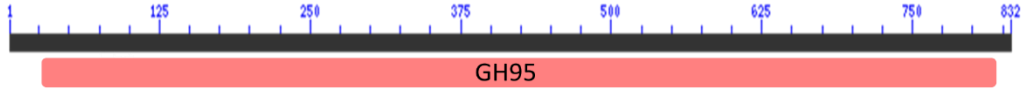
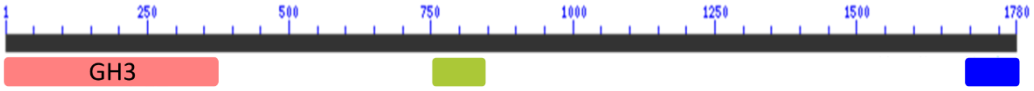
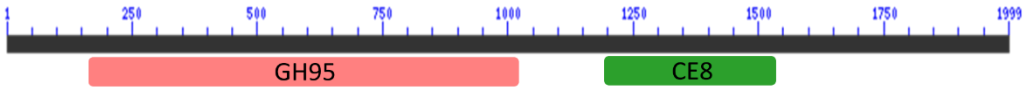
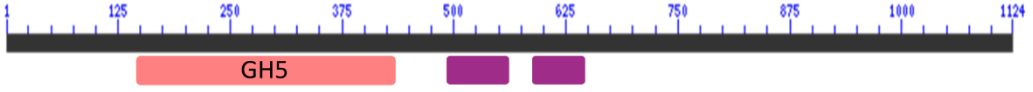
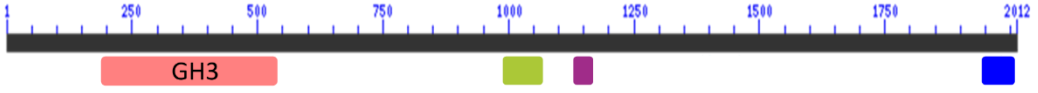
B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
51	PL1	Pectate lyase	709	Yes
				
312	PL1	Pectate lyase	906	Yes
				
428	PL9	Pectate lyase	759	Yes
				
873	PL1 / S-layer domain-containing protein	Pectate lyase	1,230	Yes
				
874	PL1	Pectate lyase	1,525	Yes
				
909	PL1 / S-layer domain-containing protein	Pectate lyase	2,071	Yes
				
929	PL1	Pectate lyase	1,022	Yes
				
1128	PL1	Pectate lyase	1,261	Yes
				

B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
1129	PL1	Pectate lyase	980	Yes
				
1239	PL1	Pectate lyase	1,000	Yes
				
1241	PL1	Pectate lyase	604	Yes
				
1315	PL1	Pectate lyase	771	Yes
				
1377	PL1	Pectate lyase	841	Yes
				
1442	PL1	Pectate lyase	1,110	Yes
				
1443	PL1	Pectate lyase	1,201	Yes
				
1615	PL9	Pectate lyase	581	Yes
				

B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
1653	PL11	Rhamnogalacturonan lyase	1,333	Yes
				
1867	PL1	Pectate lyase	897	Yes
				
2006	PL1	Pectate lyase	1,080	Yes
				
2062	PL1	Pectate lyase	1,392	Yes
				
2160	PL1	Pectate lyase	1,146	Yes
				

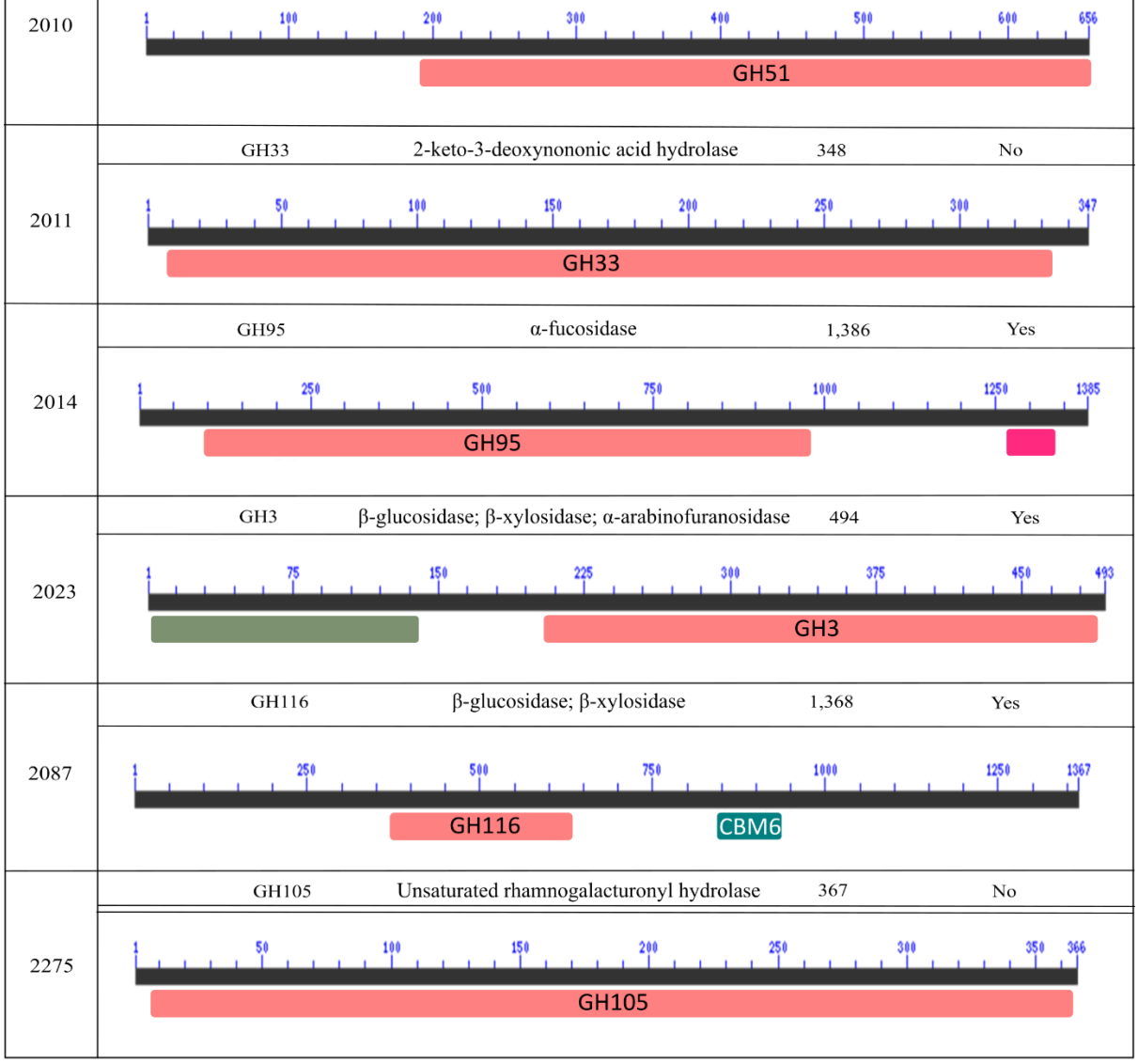
B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
53	GH97 / S-layer domain-containing protein	α-galactosidase	795	Yes
				
398	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	2,019	Yes
				
609	GH5	Endoglucanase	1,141	Yes
				
681	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	815	No
				
804	GH43	β-xylosidase; α-arabinofuranosidase	638	No
				
865	GH43	β-xylosidase; α-arabinofuranosidase	513	No
				
868	GH28	Polygalacturonase	493	No
				
956	GH140	Apiosidase	434	No
				

S-layer homology domain | Glycoside hydrolase | Pectinesterase | CBM | Fibronectin type III-like domain | Proline-rich protein
Rhodanese homology domain | Bacterial Ig-like domain | Ankyrin repeat | Copper amine oxidase N-terminal domain | Discoidin domain family protein

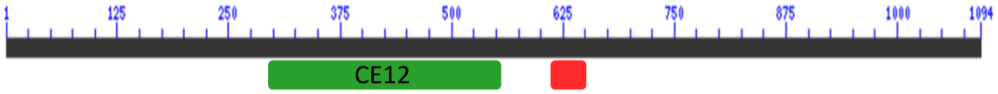
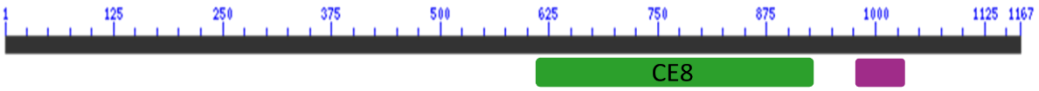
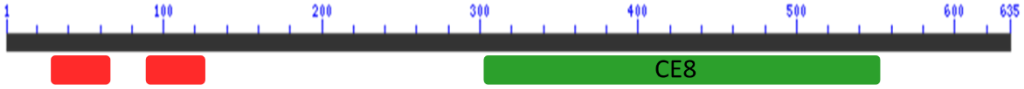
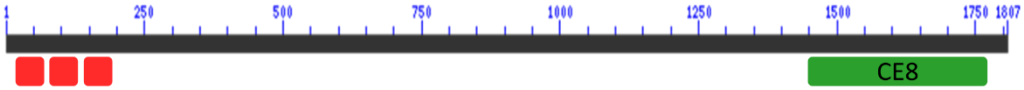
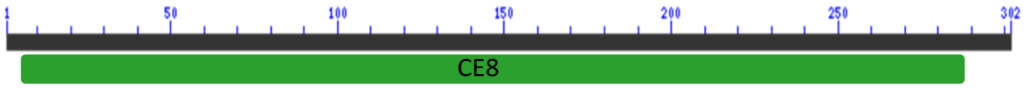
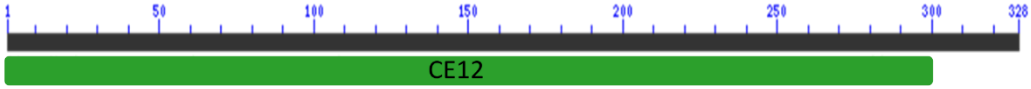
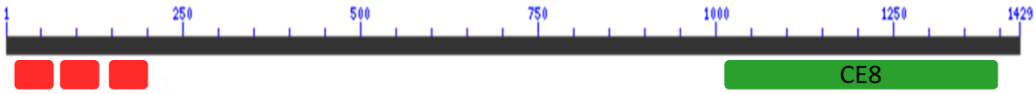
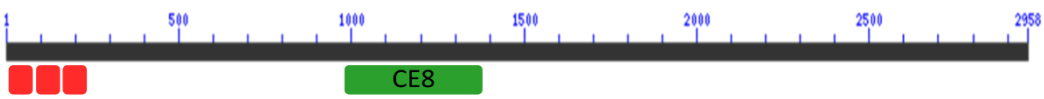
B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
989	GH95	α-fucosidase	1,305	Yes
				
1299	GH95 / CE8	α-fucosidase / pectin methylesterase	2,430	Yes
				
1304	GH43	β-xylosidase; α-arabinofuranosidase	512	No
				
1438	GH95	α-fucosidase	833	Yes
				
1543	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	1,781	No
				
1681	GH95 / CE8	α-fucosidase / pectin methylesterase	2,000	Yes
				
1703	GH5	Endoglucanase	1,125	Yes
				
1828	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	2,013	Yes
				

S-layer homology domain | Glycoside hydrolase | Pectinesterase | CBM | Fibronectin type III-like domain | Proline-rich protein
Rhodanese homology domain | Bacterial Ig-like domain | Ankyrin repeat | Copper amine oxidase N-terminal domain | Discoidin domain family protein

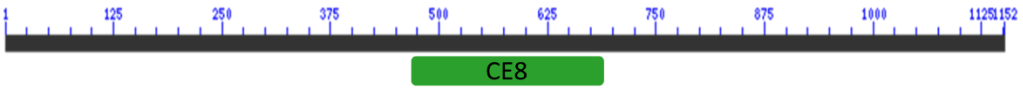
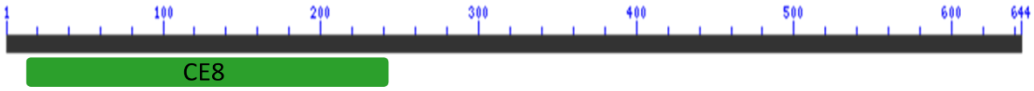
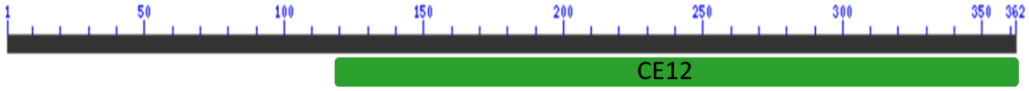
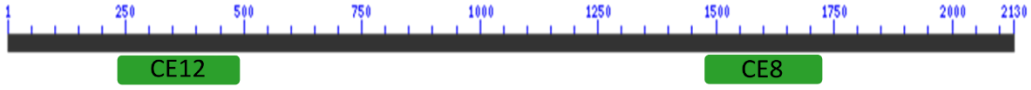
B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
2010	GH51	Endoglucanase	657	No
2011	GH33	2-keto-3-deoxynononic acid hydrolase	348	No
2014	GH95	α-fucosidase	1,386	Yes
2023	GH3	β-glucosidase; β-xylosidase; α-arabinofuranosidase	494	Yes
2087	GH116	β-glucosidase; β-xylosidase	1,368	Yes
2275	GH105	Unsaturated rhamnogalacturonyl hydrolase	367	No



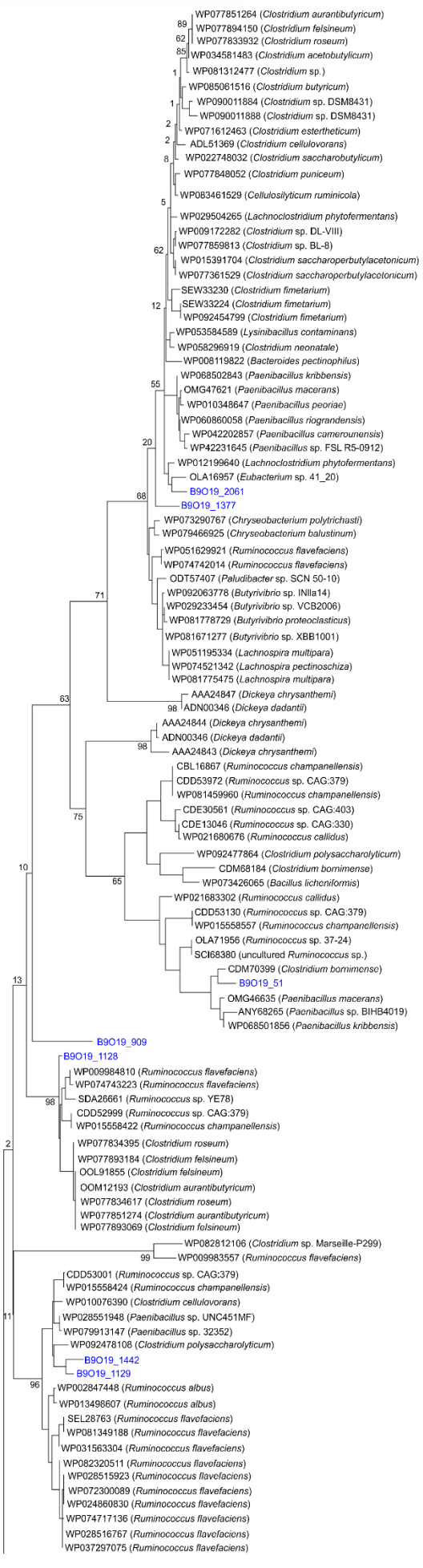
S-layer homology domain
Glycoside hydrolase
Pectinesterase
CBM
Fibronectin type III-like domain
Proline-rich protein
Rhodanese homology domain
Bacterial Ig-like domain
Ankyrin repeat
Copper amine oxidase N-terminal domain
Discoidin domain family protein

B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
203	CE12 / S-layer domain-containing protein	pectin acetylesterase	1,095	Yes
				
505	CE8	pectin methylesterase	1,168	Yes
				
680	CE8 / S-layer domain-containing protein	pectin methylesterase	636	Yes
				
826	CE8 / S-layer domain-containing protein	pectin methylesterase	1,808	Yes
				
869	CE8	pectin methylesterase	303	No
				
884	CE12	pectin acetylesterase	329	No
				
1626	CE8 / S-layer domain-containing protein	pectin methylesterase	1,430	Yes
				
1627	CE8 / S-layer domain-containing protein	pectin methylesterase	2,959	Yes
				

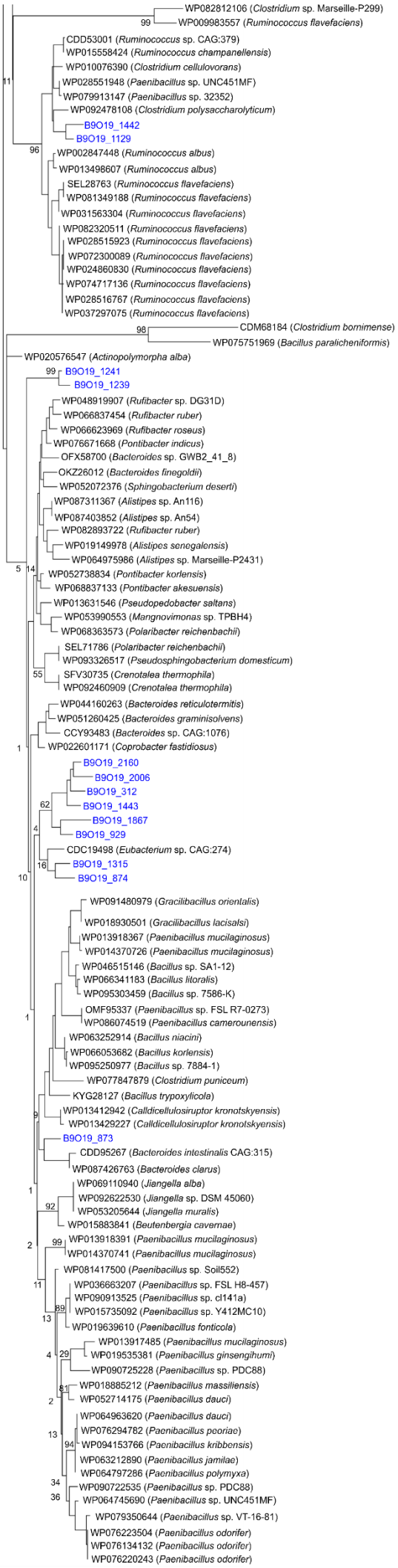
S-layer homology domain Pectinesterase CBM Fibronectin type III-like domain Bacterial Ig-like domain

B9O19	Annotation	Predicted function	Length (aa)	Signal peptide
1817	CE8	pectin methylesterase	1,153	Yes
				
2015	CE8	pectin methylesterase	645	No
				
2077	CE12	pectin acetylesterase	363	No
				
2231	CE12	pectin acetylesterase	868	Yes
				
2263	CE8 / CE12	pectin methylesterase / acetylesterase	2,132	Yes
				

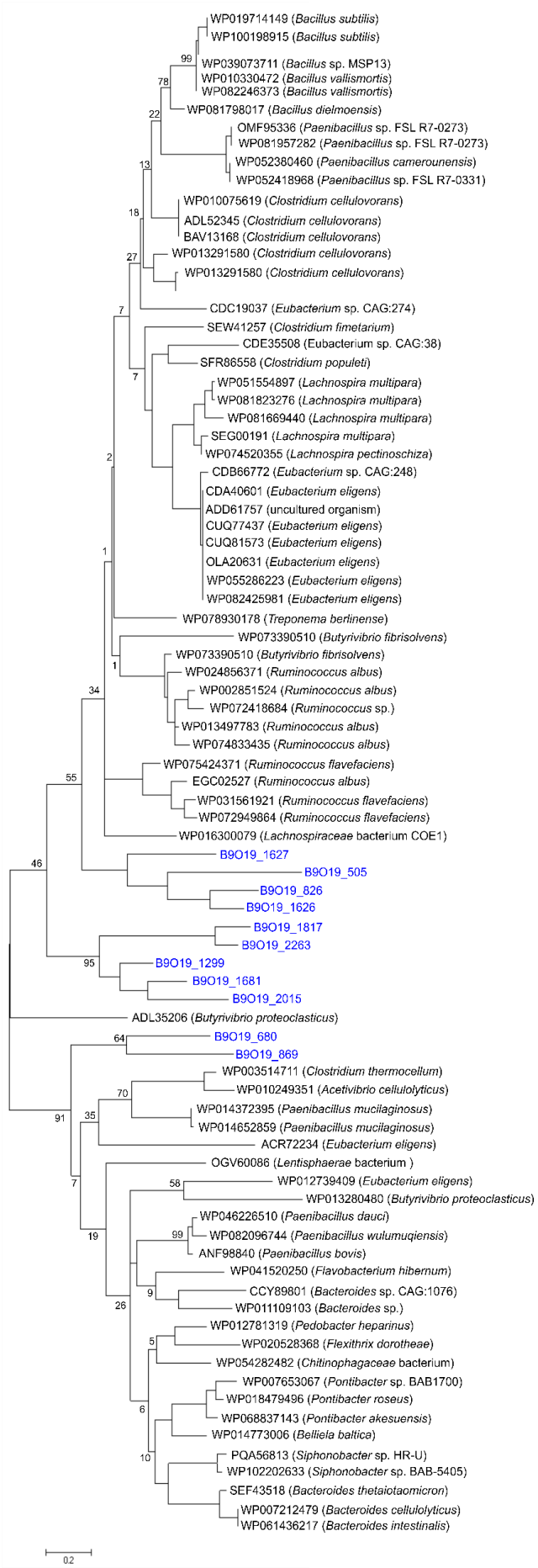
a



a (continued)



b



Appendix 7 Extended phylogenetic trees of PL1 (a) and CE8 (b) catalytic domain sequences. GenBank accession numbers, and the bacterial source of sequences are indicated. Sequences from *M. pectinilyticus* are indicated with blue. Reference sequences were obtained from the GenBank protein database based on the highest BlastP identity matches to the *M. pectinilyticus* PL1 and CE8 sequences. Catalytic domain sequences were manually extracted, aligned with ClustalW, and used to construct a maximum-likelihood phylogenetic tree using MEGA7 software. Bootstrap values calculated using 1,000 re-samplings are indicated.

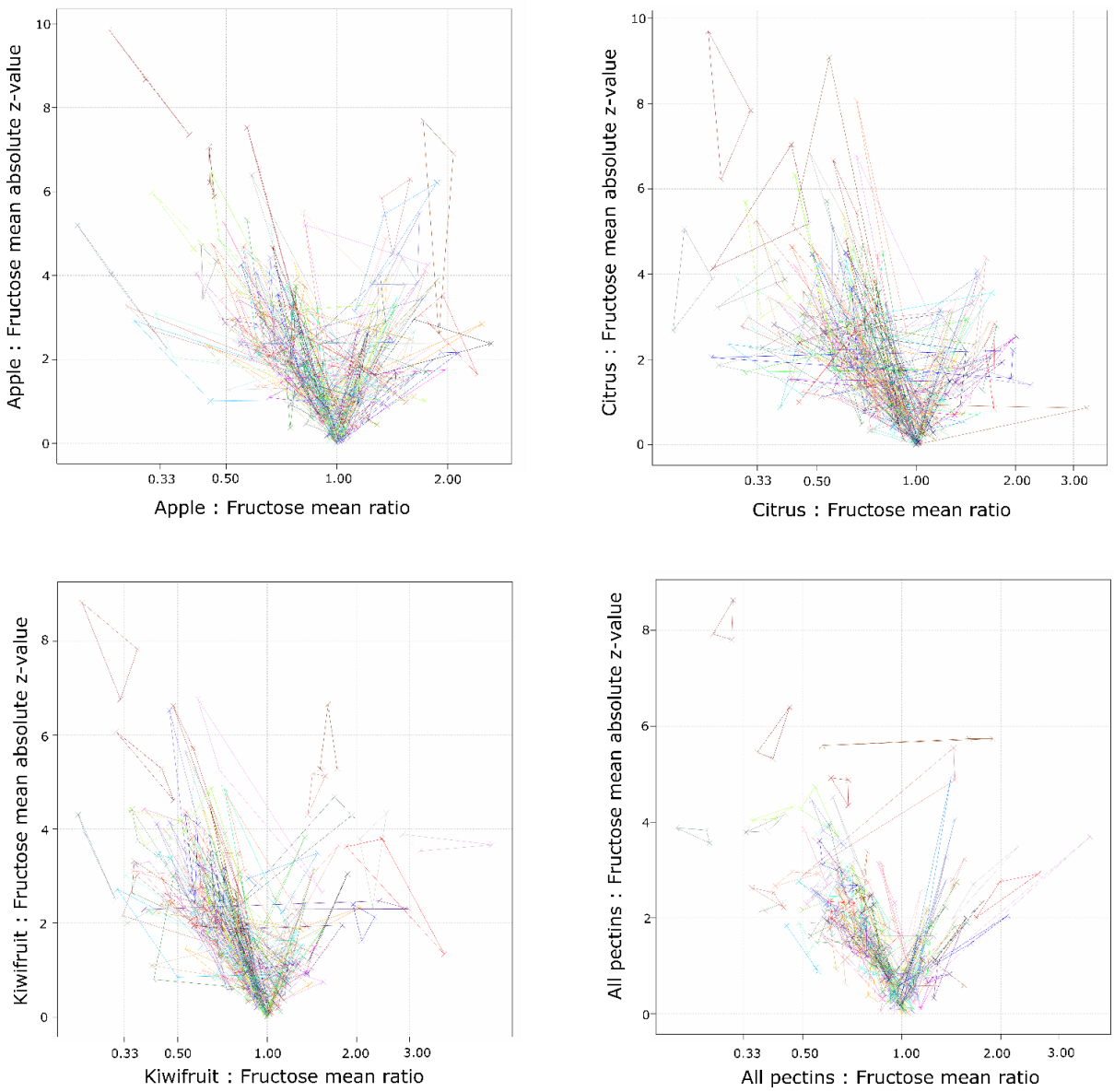
Appendix 8 EC numbers and enzyme names associated with gene names given in Figure 4.12.

Gene name	EC number	Enzyme name
aco	EC 4.2.1.3	Aconitate hydratase
acs	EC 6.2.1.1	Acetyl-CoA synthetase
ak	EC 2.7.2.1	Acetate kinase
araA	EC 5.3.1.4	L-arabinose isomerase
araB	EC 2.7.1.16	Ribulokinase
araD	EC 5.1.3.4	L-ribulose-5-phosphate 4-epimerase
cs	EC 2.3.3.1	Citrate synthase
eno	EC 4.2.1.11	Enolase
fba	EC 4.1.2.13	Fructose-bisphosphate aldolase
fbp	EC 3.1.3.11	Fructose bisphosphatase
fh	EC 4.2.1.2	Fumarate hydratase
fruA	EC 2.7.1.4	Fructokinase
fruK	EC 2.7.1.56	1-phosphofructokinase
gapdh	EC 1.2.1.12	Glyceraldehyde 3-phosphate dehydrogenase
idh	EC 1.1.1.42	Isocitrate dehydrogenase
kdgA	EC 4.1.2.14	2-dehydro-3-deoxy-phosphogluconate aldolase
kdgK	EC 2.7.1.45	2-dehydro-3-deoxygluconokinase
kduD	EC 1.1.1.125	2-deoxy-D-gluconate 3-dehydrogenase
kduI	EC 5.3.1.17	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
kor	EC 1.2.7.3	2-oxoglutarate synthase
ldh	EC 1.1.1.27	Lactate dehydrogenase
mdh	EC 1.1.1.37	Malate dehydrogenase
oad	EC 4.1.1.3	Oxaloacetate decarboxylase
pepck	EC 4.1.1.32	Phosphoenolpyruvate carboxykinase
pfkM	EC 2.7.1.11	Phosphofructokinase-1
pfl	EC 2.3.1.54	Pyruvate formate-lyase
pfor	EC 1.2.7.1	Pyruvate:ferredoxin oxidoreductase
pgk	EC 2.7.2.3	Phosphoglycerate kinase
pgm	EC 5.4.2.12	Phosphoglycerate mutase
pta	EC 2.3.1.8	Phosphate acetyltransferase
rpe	EC 5.1.3.1	Ribulose-phosphate 3-epimerase
rpi	EC 5.3.1.6	Ribose 5-phosphate isomerase
scs	EC 6.2.1.4	Succinate-CoA synthetase
sdh	EC 1.3.5.1	Succinic dehydrogenase
tal	EC 2.2.1.2	Transaldolase
tkt	EC 2.2.1.1	Transketolase
tpi	EC 5.3.1.1	Triose-phosphate isomerase
uxaA	EC 4.2.1.7	Altronate dehydratase
uxaB	EC 1.1.1.58	Tagaturonate reductase
uxaC	EC 5.3.1.12	Uronate isomerase
uxuA	EC 4.2.1.8	Mannonate dehydratase
uxuB	EC 1.1.1.57	D-mannonate oxidoreductase
xfp	EC 4.1.2.9	Phosphoketolase
xylA	EC 5.3.1.5	Xylose isomerase
xylB	EC 2.7.1.17	Xylulokinase

Appendix 9 Sporulation-related genes of *M. pectinilyticus*. 'Signature' sporulation genes (Browne *et al*., 2016) are indicated.

Gene name	Locus number	Annotation	Signature
*sig*E	B9O19_58	RNA polymerase sigma factor SigE	Yes
*sig*H	B9O19_91	RNA polymerase sigma-H factor	Yes
*yun*B	B9O19_228	sporulation protein YunB	No
*ger*A	B9O19_251	GerA spore germination protein	No
*ger*C	B9O19_252	spore germination protein GerC	No
*abr*B	B9O19_253	transcription regulator AbrB	Yes
*spo0*A	B9O19_265	stage 0 sporulation protein A	Yes
*spoIII*D	B9O19_324	sporulation transcriptional regulator SpoIIID	No
*ytx*C	B9O19_371	sporulation protein YtxC	Yes
spl	B9O19_378	spore photoproduct lyase	No
*trx*B	B9O19_379	thioredoxin-disulfide reductase	Yes
*spoII*M	B9O19_395	stage II sporulation protein M	No
*spoIII*AH	B9O19_436	stage III sporulation protein AH	Yes
*spoIII*AG	B9O19_437	stage III sporulation protein AG	Yes
*spoIII*AF	B9O19_438	stage III sporulation protein AF	No
*spoIII*AE	B9O19_439	stage III sporulation protein AE	Yes
*spoIII*AD	B9O19_440	stage III sporulation protein AD	Yes
*spoIII*AC	B9O19_441	stage III sporulation protein AC	Yes
*spoIII*AB	B9O19_442	stage III sporulation protein AB	Yes
*spoIII*AA	B9O19_443	stage III sporulation protein AA	Yes
*sig*K	B9O19_656	RNA polymerase sporulation specific sigma factor SigK	Yes
*spoI*VA	B9O19_657	stage IV sporulation protein A	Yes
*whi*A	B9O19_659	sporulation transcription regulator WhiA	No
*spoII*D	B9O19_774	stage II sporulation protein SpoIID	No
*dap*G	B9O19_778	aspartokinase	Yes

*ytf*J	B9O19_798	sporulation protein YtfJ	No
*yqf*D	B9O19_813	sporulation protein YqfD	No
*yqf*C	B9O19_814	sporulation protein YqfC	No
*ylb*J	B9O19_822	sporulation integral membrane protein YlbJ	Yes
*spo*VT	B9O19_851	stage V sporulation protein T	Yes
gpr	B9O19_854	germination protease	No
*ylm*C/*ymx*H	B9O19_900	sporulation protein,YlmC/YmxH family	No
*rpl*J	B9O19_1050	50S ribosomal protein L10	Yes
*par*B	B9O19_1072	chromosome(plasmid) partitioning protein ParB	Yes
*spoII*E	B9O19_1116	stage II sporulation protein E	No
*dna*A	B9O19_1170	chromosomal replication initiation protein DnaA	Yes
*sig*G	B9O19_1311	sporulation sigma factor SigG	Yes
*spm*B	B9O19_1349	spore maturation protein B	Yes
*spm*A	B9O19_1350	spore maturation protein A	Yes
*fts*Q	B9O19_1361	cell division protein FtsQ	Yes
*spoII*P	B9O19_1417	stage II sporulation protein P	Yes
*wal*R	B9O19_1498	transcriptional regulatory protein WalR	Yes
*sod*A	B9O19_1713	superoxide dismutase [Cu-Zn] precursor	Yes
*spoII*R	B9O19_1840	stage II sporulation protein R	Yes
*sig*F	B9O19_1914	RNA polymerase sigma-F factor	Yes
*spo*VAC	B9O19_1915	stage V sporulation protein AC	No
*spo*VAD	B9O19_1916	stage V sporulation protein AD	Yes
*spo*VAE	B9O19_1917	stage V sporulation protein AE	No
*yab*P	B9O19_1923	sporulation protein YabP	Yes
*yab*Q	B9O19_1924	spore cortex biosynthesis protein YabQ	No
*sep*F	B9O19_1961	cell division protein SepF	Yes
*spoI*VB	B9O19_1994	spoIVB peptidase	No
*spo0*A	B9O19_1995	sporulation transcription factor Spo0A	Yes
*cot*JB	B9O19_2285	spore coat protein CotJB	No



Appendix 10 Geometric distribution of z-scores for individual protein expression ratios. The statistical significance of the mean relative protein expression was determined based on the geometric mean of the z-scores in the three replicates.

Appendix 11 iTRAQ proteomics expression ratios between apple pectin-grown *M. pectinilyticus* cells and fructose-grown cells. E, extracellular; I; intracellular, L, lipoprotein signal; R1, replicate 1; R2, replicate 2; R3, replicate 3; R mean, average R values; P1 – P3, *p*-values for each replicate; Z1, z-score for P1; Z2, z-score for P2; Z3, z-score for P3; abs Z, average z-score. Red, up-regulated (R mean > 1.2); yellow, not differentiated; green; down-regulated (R mean < 0.83). R mean values are statistically significant if abs Z is > 1.65.

Locus tag (B9O19)	Gene product description		R1	P1	# of peptides (95%)	R2	P2	# of peptides (95%)	R3	P3	# of peptides (95%)	R mean	z1	z2	z3	abs z
52	S-layer domain-containing protein	E	2.63	0.02	4	1.26	0.11	11	1.62	0.00	11	1.75	2.38	1.62	2.96	2.32
1475	peptidase C1A papain	E	2.41	0.10	3	1.94	0.00	13	1.69	0.00	17	1.99	1.66	3.53	3.07	2.75
1156	S-layer domain-containing protein	E	2.07	0.00	121	1.72	0.00	336	1.89	0.01	354	1.89	6.90	7.69	2.63	5.74
611	S-layer domain-containing protein	E	1.98	0.02	4	1.37	0.02	12	1.65	0.00	9	1.65	2.38	2.38	3.35	2.70
1520	transcriptional regulator	I	1.03	0.95	4	1.93	0.03	11	2.14	0.03	7	1.62	0.06	2.18	2.16	1.47
1947	elongation factor G	I	1.74	0.00	5	1.31	0.01	17	1.61	0.00	14	1.55	4.11	2.47	3.11	3.23
1577	Dockerin-like domain	E	2.48	0.00	6	1.08	0.36	9	1.30	0.07	9	1.52	2.84	0.92	1.78	1.85
1425	2-dehydro-3-deoxy phosphogluconate aldolase	I	1.21	0.10	18	1.40	0.01	27	1.87	0.00	24	1.47	1.67	2.69	3.89	2.75
1044	glucose-1-phosphate adenylyltransferase	I	1.58	0.01	4	1.51	0.04	9	1.31	0.17	6	1.46	2.49	2.04	1.38	1.97
826	Pectinesterase (CE8) / S-layer domain-containing protein	E	1.67	0.00	20	1.25	0.00	47	1.47	0.00	45	1.45	3.80	3.79	4.54	4.04
1680	Transketolase	I	1.75	0.09	4	1.39	0.13	10	1.24	0.03	11	1.45	1.70	1.53	2.20	1.81
1627	Pectinesterase (CE8) / S-layer domain-containing protein	E	1.58	0.00	47	1.32	0.00	26	1.42	0.00	38	1.44	6.30	5.84	4.51	5.55
1597	S-layer domain-containing protein	E	1.87	0.00	92	1.12	0.00	120	1.35	0.00	147	1.41	6.21	3.14	5.46	4.94
870	extracellular solute-binding protein family 1	L	1.28	0.03	5	1.30	0.00	21	1.53	0.00	15	1.36	2.23	3.80	3.77	3.27
2005	2-deoxy-D-gluconate 3-dehydrogenase	I	0.91	0.33	19	1.65	0.00	33	1.68	0.00	40	1.36	-0.98	3.83	5.48	2.78
972	Thioredoxin	I	1.55	0.28	6	0.96	0.66	10	1.50	0.09	9	1.31	1.09	-0.44	1.70	0.78
916	N-acetylmuramoyl-L-alanine amidase	E	1.45	0.09	6	1.03	0.92	5	1.44	0.15	5	1.29	1.69	0.10	1.45	1.08
608	NIF system FeS cluster assembly, NifU, N-terminal	I	1.15	0.23	9	1.30	0.09	12	1.31	0.05	14	1.25	1.20	1.71	1.94	1.62
1225	S-layer domain-containing protein	E	1.65	0.09	5	0.95	0.37	15	1.23	0.03	21	1.24	1.71	-0.89	2.19	1.00
1045	glucose-1-phosphate adenylyltransferase	I	1.16	0.25	3	1.22	0.57	7	1.32	0.10	9	1.23	1.15	0.56	1.63	1.11
1089	Multiple sugar ABC transporter, ATP-binding protein	I	0.96	0.93	4	1.35	0.02	12	1.44	0.00	10	1.23	-0.08	2.38	3.26	1.85
682	hypothetical protein	E	1.78	0.02	6	0.93	0.49	12	1.10	0.28	14	1.22	2.27	-0.69	1.08	0.89
2227	Periplasmic sugar-binding proteins	L	0.94	0.86	5	1.11	0.70	7	1.66	0.08	7	1.20	-0.17	0.38	1.73	0.65

2203	IMP cyclohydrolase	I	0.93	0.61	4	1.35	0.00	14	1.37	0.00	14	1.20	-0.52	3.21	3.92	2.21
1928	band 7 protein	I	0.87	0.45	16	1.29	0.01	21	1.51	0.00	18	1.19	-0.75	2.80	4.42	2.15
2204	IMP cyclohydrolase	I	1.48	0.07	6	1.00	0.99	17	1.12	0.24	14	1.18	1.82	-0.01	1.17	0.99
751	transcriptional regulator, CdaR	I	0.95	0.59	7	1.19	0.13	17	1.46	0.00	20	1.18	-0.54	1.53	3.47	1.49
1570	dihydroxy-acid dehydratase	I	1.04	0.79	5	1.16	0.14	16	1.34	0.00	15	1.17	0.26	1.48	2.87	1.53
1046	starch synthase	I	1.05	0.94	5	1.15	0.09	8	1.30	0.23	10	1.16	0.07	1.69	1.21	0.99
876	Uronate isomerase	I	1.10	0.70	6	1.06	0.54	10	1.33	0.02	11	1.16	0.39	0.61	2.34	1.11
1005	transcriptional regulator	I	0.63	0.02	11	1.43	0.02	23	1.72	0.00	20	1.16	-2.37	2.42	3.45	1.17
914	S-layer domain-containing protein	E	1.77	0.00	19	0.82	0.00	67	1.03	0.49	64	1.14	4.25	-5.21	0.68	0.09
1554	hypothetical protein	I	1.97	0.08	3	0.73	0.12	6	1.03	0.89	9	1.14	1.75	-1.56	0.14	0.11
2157	Periplasmic [Fe] hydrogenase large subunit	I	1.07	0.79	6	1.11	0.22	13	1.25	0.00	11	1.14	0.26	1.22	3.09	1.52
2161	L-arabinose isomerase	I	0.91	0.67	10	1.20	0.00	26	1.36	0.00	19	1.14	-0.43	3.20	4.86	2.55
827	carbamoyl-phosphate synthase large chain	I	1.10	0.27	4	1.08	0.24	19	1.19	0.01	23	1.12	1.11	1.17	2.52	1.60
602	Heat shock protein 60 family chaperone GroEL	I	0.79	0.03	17	1.31	0.00	72	1.36	0.00	71	1.12	-2.19	3.98	2.89	1.56
1888	oxaloacetate decarboxylase alpha chain	I	0.96	0.83	4	1.21	0.01	17	1.17	0.08	13	1.11	-0.22	2.61	1.76	1.39
1372	ribosome recycling factor	I	1.32	0.20	6	0.95	0.48	13	1.07	0.48	10	1.10	1.29	-0.71	0.70	0.43
273	phosphoglucosamine mutase	I	1.01	0.96	4	1.12	0.19	8	1.16	0.35	8	1.09	0.05	1.31	0.93	0.76
1678	argininosuccinate synthase	I	0.89	0.54	3	1.12	0.14	9	1.29	0.01	12	1.09	-0.61	1.46	2.79	1.21
2084	ATP synthase beta chain	I	0.51	0.20	4	1.62	0.01	5	1.54	0.17	6	1.08	-1.28	2.67	1.39	0.92
1446	DNA-directed RNA polymerase, beta subunit	I	1.14	0.34	7	0.93	0.25	20	1.16	0.06	28	1.07	0.96	-1.14	1.88	0.57
2196	phosphoribosylformylglycinamidine synthase	I	1.09	0.30	14	0.96	0.66	23	1.17	0.04	19	1.07	1.05	-0.44	2.08	0.90
908	S-layer domain-containing protein	E	1.19	0.24	13	0.90	0.05	46	1.12	0.04	50	1.07	1.18	-1.95	2.11	0.45
1426	2-dehydro-3-deoxy gluconokinase KdgK	I	1.04	0.78	13	1.02	0.77	14	1.09	0.21	17	1.05	0.27	0.30	1.25	0.61
692	Transketolase, C-terminal domain	I	0.92	0.52	8	1.17	0.22	13	1.07	0.63	13	1.05	-0.65	1.22	0.48	0.35
607	hypothetical protein	I	1.06	0.79	8	0.89	0.13	12	1.22	0.13	12	1.05	0.27	-1.53	1.52	0.09
1898	fructose-bisphosphate aldolase	I	0.78	0.24	5	1.13	0.21	11	1.27	0.02	12	1.04	-1.17	1.27	2.38	0.83
1303	Cell division protein FtsZ	I	0.99	0.93	18	1.00	0.99	19	1.13	0.01	20	1.04	-0.09	-0.01	2.67	0.86
909	Pectate lyase (PL1) / S-layer domain-containing protein	E	1.43	0.00	20	0.81	0.00	64	0.96	0.26	63	1.04	3.34	-5.50	-1.12	1.09
898	4-deoxy-L-threo-5-hexosulose-uronateketol-isomerase	I	0.82	0.64	4	1.02	0.88	7	1.31	0.00	8	1.03	-0.47	0.15	3.21	0.96
743	Propeptide PepSY and peptidase M4	E	1.03	0.89	8	0.88	0.17	13	1.16	0.28	13	1.02	0.14	-1.37	1.07	0.05

2018	translation elongation factor P	I	1.15	0.15	3	0.93	0.46	6	0.98	0.89	8	1.01	1.45	-0.74	-0.13	0.19
1578	S-layer domain-containing protein	E	1.27	0.00	28	0.82	0.00	53	1.01	0.88	58	1.01	3.31	-3.19	0.15	0.09
724	tetrahydrofolate synthase	I	1.73	0.31	3	0.75	0.07	7	0.81	0.27	9	1.01	1.02	-1.82	-1.10	0.63
1271	Translation elongation factor Tu	I	0.87	0.23	36	1.03	0.64	58	1.17	0.00	60	1.01	-1.19	0.46	3.07	0.78
1445	DNA-directed RNA polymerase, beta subunit	I	0.93	0.56	10	0.98	0.75	26	1.14	0.05	29	1.01	-0.58	-0.32	1.99	0.36
1522	malate dehydrogenase	I	0.83	0.02	10	0.96	0.76	12	1.27	0.13	14	1.00	-2.40	-0.30	1.51	0.40
901	LL-diaminopimelate aminotransferase	I	0.98	0.86	11	0.93	0.44	21	1.10	0.34	28	1.00	-0.17	-0.77	0.96	0.00
1287	Rubrerythrin	I	0.65	0.27	5	1.10	0.39	8	1.40	0.01	8	1.00	-1.09	0.86	2.52	0.76
366	amino acid-binding ACT domain protein	I	0.85	0.36	3	1.03	0.80	9	1.12	0.22	8	1.00	-0.92	0.25	1.24	0.19
1178	ornithine carbamoyltransferase	I	1.54	0.32	5	0.75	0.00	11	0.84	0.01	12	0.99	0.99	-3.27	-2.67	1.65
919	glyceraldehyde-3-phosphate dehydrogenase	I	0.86	0.12	19	0.99	0.87	27	1.08	0.28	26	0.97	-1.57	-0.17	1.09	0.22
949	homoserine dehydrogenase	I				0.80	0.06	11	1.15	0.30	7	0.96		-1.86	1.03	0.41
2045	glycine hydroxymethyltransferase	I	0.92	0.41	8	0.92	0.05	17	1.04	0.47	16	0.96	-0.83	-1.95	0.72	0.69
1972	orotate phosphoribosyltransferase	I	0.86	0.16	4	0.91	0.40	10	1.07	0.32	8	0.94	-1.40	-0.84	0.99	0.42
1346	aspartate transaminase	I	0.77	0.23	3	1.05	0.83	8	1.03	0.90	5	0.94	-1.21	0.21	0.12	0.29
1337	S-layer domain-containing protein	E	1.28	0.65	5	0.78	0.08	9	0.83	0.18	6	0.94	0.45	-1.75	-1.34	0.88
198	phosphoribosylamine--glycine ligase	I	0.95	0.83	6	0.81	0.06	5	1.06	0.71	8	0.94	-0.21	-1.87	0.37	0.57
617	30S ribosomal protein S2	I	1.53	0.14	5	0.69	0.00	12	0.76	0.03	15	0.93	1.48	-3.40	-2.17	1.36
912	S-layer domain-containing protein	E	1.13	0.44	10	0.76	0.00	31	0.94	0.27	31	0.93	0.77	-4.32	-1.09	1.55
1090	N-acetyl-gamma-glutamyl-phosphatereductase	I				0.83	0.08	7	1.04	0.79	6	0.93		-1.76	0.26	0.75
1219	DNA-directed RNA polymerase subunit alpha	I	0.99	0.96	3	0.87	0.17	9	0.92	0.47	9	0.93	-0.06	-1.37	-0.73	0.72
2111	D-3-phosphoglycerate dehydrogenase	I	0.91	0.34	12	0.91	0.10	23	0.95	0.25	22	0.92	-0.94	-1.66	-1.15	1.25
414	SCP-2 sterol transfer family	I				0.91	0.49	6	0.93	0.51	6	0.92		-0.69	-0.66	0.67
403	Transcriptional regulator, GntR family	I	0.88	0.28	5	0.84	0.01	17	1.05	0.60	15	0.92	-1.08	-2.45	0.52	1.01
947	aspartatekinase	I	1.39	0.02	12	0.73	0.01	10	0.75	0.07	10	0.92	2.26	-2.56	-1.80	0.70
1330	pyruvate synthase	I	0.91	0.18	32	0.88	0.00	45	0.95	0.26	50	0.91	-1.34	-3.32	-1.13	1.93
1969	orotidine 5'-phosphate decarboxylase	I	0.80	0.25	3	0.81	0.11	6	1.15	0.36	4	0.90	-1.16	-1.59	0.91	0.61
1892	acetyl-CoA carboxylase	I	0.91	0.60	10	0.89	0.02	20	0.88	0.01	19	0.89	-0.52	-2.24	-2.76	1.84
287	3-dehydroquinate synthase	I				0.91	0.62	7	0.85	0.30	9	0.88		-0.49	-1.05	0.77
2082	ATP synthase alpha chain	I	1.07	0.52	4	0.70	0.05	5	0.91	0.27	5	0.88	0.65	-1.95	-1.11	0.80

2112	phosphoserine transaminase	I	0.74	0.05	12	0.92	0.36	18	0.99	0.87	18	0.88	-1.96	-0.91	-0.17	1.01
89	phosphate acetyltransferase	I	0.98	0.85	6	0.81	0.00	10	0.84	0.06	11	0.87	-0.18	-3.03	-1.92	1.71
1269	30Sribosomal protein S7	I	1.62	0.26	3	0.61	0.00	15	0.67	0.00	11	0.87	1.12	-3.16	-2.95	1.66
1243	DNA gyrase subunit A	I				0.97	0.84	7	0.75	0.23	7	0.86		-0.21	-1.20	0.70
1858	malonyl CoA-acyl carrier protein transacylase	I	0.85	0.55	8	0.81	0.05	15	0.90	0.26	12	0.86	-0.60	-1.92	-1.14	1.22
368	2-oxoacid ferredoxin oxidoreductase	I	0.75	0.35	3	0.83	0.05	10	0.97	0.69	16	0.84	-0.93	-1.97	-0.40	1.10
1066	Arginine-binding extracellular protein ArtP	L	0.89	0.45	5	0.73	0.10	8	0.93	0.62	10	0.84	-0.75	-1.66	-0.50	0.97
1195	30S ribosomal protein S3	I	1.37	0.38	5	0.61	0.00	11	0.70	0.10	11	0.83	0.88	-3.57	-1.65	1.45
1513	Chaperone protein DnaK	I	0.73	0.27	8	0.77	0.00	20	1.01	0.92	20	0.83	-1.10	-3.74	0.10	1.58
506	O-acetylhomoserine sulfhydrylase	I	0.89	0.47	4	0.80	0.13	7	0.80	0.17	5	0.83	-0.72	-1.51	-1.37	1.20
775	aspartate-semialdehyde dehydrogenase	I	0.91	0.68	13	0.74	0.00	16	0.82	0.01	16	0.82	-0.41	-3.92	-2.68	2.34
1010	6-phosphofructokinase	I	1.01	0.95	7	0.68	0.00	16	0.80	0.01	15	0.82	0.07	-4.23	-2.66	2.27
727	NADP-specific glutamate dehydrogenase	I	0.82	0.09	9	0.77	0.00	15	0.85	0.20	12	0.81	-1.68	-3.61	-1.27	2.19
950	hypothetical protein	I				0.77	0.07	4	0.84	0.31	6	0.81		-1.83	-1.02	1.43
1826	3-isopropylmalate dehydratase small subunit	I	0.68	0.06	5	0.80	0.06	4	0.96	0.78	5	0.81	-1.86	-1.90	-0.28	1.35
616	translation elongation factor Ts	I	1.06	0.79	8	0.67	0.00	16	0.73	0.00	18	0.80	0.27	-4.66	-2.99	2.46
389	peptidase M18 aminopeptidase I	I				0.77	0.09	8	0.83	0.03	6	0.80		-1.72	-2.13	1.92
1273	2-isopropylmalate synthase	I	0.44	0.02	4	1.04	0.65	19	1.07	0.46	16	0.79	-2.35	0.45	0.74	0.38
1265	pyruvate phosphate dikinase	I	0.76	0.01	18	0.78	0.00	33	0.82	0.00	36	0.78	-2.51	-3.91	-3.74	3.39
277	proline--tRNA ligase	I	0.84	0.13	6	0.67	0.01	5	0.86	0.07	7	0.78	-1.53	-2.68	-1.83	2.01
1821	acetolactate synthase	I	0.76	0.26	4	0.75	0.00	13	0.83	0.15	8	0.78	-1.13	-3.32	-1.45	1.97
1218	30Sribosomal protein S4	I	1.38	0.19	4	0.56	0.00	8	0.61	0.04	7	0.78	1.30	-4.64	-2.11	1.81
1051	50S ribosomal protein L1	I	1.29	0.07	6	0.56	0.00	12	0.66	0.06	9	0.78	1.80	-4.69	-1.90	1.60
1206	30Sribosomal protein S5	I	1.59	0.01	9	0.54	0.00	12	0.53	0.00	9	0.77	2.48	-3.60	-3.52	1.55
542	phosphoribosylaminoimidazolesuccinocarboxamide synthase	I	1.03	0.81	9	0.60	0.00	10	0.73	0.04	8	0.77	0.24	-4.08	-2.08	1.97
1626	Pectinesterase (CE8) / S-layer domain-containing protein	E	1.09	0.07	26	0.59	0.00	40	0.71	0.00	34	0.77	1.84	-6.39	-4.06	2.87
678	adenylosuccinate synthase	I	0.61	0.18	6	0.86	0.21	13	0.86	0.10	15	0.77	-1.34	-1.25	-1.63	1.41
1903	aconitate hydratase	I	0.75	0.06	20	0.76	0.00	28	0.77	0.01	27	0.76	-1.87	-3.56	-2.50	2.64
1595	S-layer domain-containing protein	E	0.91	0.43	7	0.66	0.00	10	0.73	0.00	11	0.76	-0.80	-3.29	-3.74	2.61
1823	Ketol-acid reductoisomerase	I	0.69	0.03	9	0.76	0.00	27	0.82	0.01	20	0.75	-2.11	-2.97	-2.76	2.61

326	Rubrerythrin	I	0.61	0.03	4	0.71	0.00	5	0.99	0.96	5	0.75	-2.20	-3.32	-0.05	1.86
128	aspartate--tRNA ligase	I	1.04	0.84	4	0.60	0.02	10	0.68	0.14	11	0.75	0.20	-2.35	-1.48	1.21
1091	acetylglutamate kinase	I	0.75	0.71	3	0.72	0.03	11	0.78	0.05	9	0.75	-0.38	-2.23	-2.00	1.53
1189	50S ribosomal protein L3	I	1.28	0.10	3	0.54	0.00	6	0.61	0.03	4	0.75	1.63	-3.43	-2.23	1.34
1827	3-isopropylmalate dehydrogenase	I	0.91	0.51	3	0.61	0.03	6	0.76	0.05	6	0.75	-0.65	-2.19	-1.92	1.59
1192	50S ribosomal protein L2	I	1.08	0.89	4	0.60	0.00	9	0.63	0.01	11	0.74	0.13	-2.84	-2.78	1.83
1607	Phosphotransferase system, phosphocarrier protein HPr	I	0.78	0.24	4				0.70	0.15	4	0.74	-1.17		-1.44	1.30
541	amidophosphoribosyltransferase	I	0.75	0.01	7	0.66	0.00	17	0.79	0.01	15	0.73	-2.61	-3.40	-2.53	2.85
90	acetate kinase	I	0.80	0.11	6	0.69	0.00	14	0.71	0.00	17	0.73	-1.60	-3.56	-2.88	2.68
59	branched-chain-amino-acid transaminase	I	0.65	0.04	3	0.76	0.01	7	0.77	0.02	9	0.72	-2.06	-2.53	-2.35	2.31
1203	30Sribosomal protein S8	I	0.99	0.96	4	0.60	0.00	8	0.63	0.00	10	0.72	-0.05	-3.30	-2.96	2.10
1155	CTPsynthase	I	0.89	0.47	5	0.66	0.02	9	0.63	0.00	10	0.72	-0.72	-2.42	-3.09	2.08
92	histidinol dehydrogenase	I				0.70	0.01	10	0.72	0.02	7	0.71		-2.61	-2.31	2.46
540	GMP synthase [glutamine-hydrolyzing]	I	0.77	0.08	8	0.61	0.00	14	0.75	0.00	12	0.71	-1.77	-4.29	-3.21	3.09
1097	Phosphoenolpyruvate carboxykinase [GTP]	I	0.75	0.04	25	0.66	0.00	28	0.70	0.00	29	0.70	-2.01	-5.19	-3.11	3.44
289	3-deoxy-7-phosphoheptulonate synthase	I	0.48	0.00	6	0.78	0.01	11	0.87	0.19	10	0.69	-2.98	-2.70	-1.30	2.33
1270	Translation elongation factor G	I	0.93	0.56	19	0.57	0.00	37	0.61	0.00	40	0.69	-0.58	-7.52	-6.52	4.88
1386	30S ribosomal protein S16	I	0.97	0.93	3	0.52	0.05	3	0.60	0.01	4	0.67	-0.09	-1.97	-2.57	1.54
1953	ribose-phosphate diphosphokinase	I	0.71	0.02	5	0.52	0.01	5	0.82	0.25	7	0.67	-2.27	-2.45	-1.15	1.96
1050	50S ribosomal protein L10	I	0.90	0.39	5	0.49	0.00	7	0.68	0.01	14	0.67	-0.86	-5.24	-2.80	2.97
1190	50S ribosomal protein L4	I	1.14	0.73	6	0.46	0.00	11	0.56	0.01	11	0.67	0.35	-4.71	-2.69	2.35
1902	isocitrate dehydrogenase	I	0.68	0.02	7	0.65	0.00	15	0.66	0.00	24	0.66	-2.28	-3.85	-4.41	3.51
1105	Glutamine synthetase	I	0.87	0.13	10	0.55	0.00	12	0.61	0.00	10	0.66	-1.51	-4.26	-3.06	2.94
1201	50S ribosomal protein L5	I	0.97	0.60	3	0.59	0.00	12	0.48	0.00	11	0.65	-0.53	-4.12	-4.82	3.16
2138	polyribonucleotide nucleotidyltransferase	I	0.53	0.00	5	0.67	0.01	12	0.75	0.06	11	0.64	-2.82	-2.72	-1.91	2.48
129	histidine--tRNA ligase	I	0.78	0.08	5	0.52	0.00	4	0.64	0.00	6	0.64	-1.74	-2.94	-3.07	2.58
627	aspartate--tRNA ligase	I	0.76	0.11	7	0.56	0.00	9	0.59	0.00	11	0.63	-1.60	-4.49	-3.52	3.21
1608	phosphoenolpyruvate-protein phosphotransferase	I	0.55	0.02	4	0.64	0.02	6	0.69	0.05	5	0.62	-2.36	-2.29	-2.00	2.21
1941	triose-phosphate isomerase	I	0.66	0.00	8	0.52	0.00	11	0.70	0.02	11	0.62	-3.10	-4.18	-2.28	3.19
1379	50S ribosomal protein L27	I				0.54	0.01	5	0.68	0.05	5	0.61		-2.43	-2.00	2.22

357	Cell division trigger factor	I				0.57	0.00	14	0.63	0.01	6	0.60		-5.32	-2.59	3.96
127	tyrosine--tRNA ligase	I				0.65	0.02	6	0.55	0.01	5	0.59		-2.25	-2.44	2.35
1052	50S ribosomal protein L11	I	0.86	0.38	4	0.41	0.00	7	0.50	0.00	8	0.56	-0.87	-4.04	-2.88	2.60
356	ATP-dependent Clp endopeptidase, proteolytic subunit ClpP	I				0.52	0.01	5	0.61	0.02	6	0.56		-2.71	-2.35	2.53
1606	PTS system fructose-specific EIIABC component	I				0.54	0.00	6	0.59	0.04	8	0.56		-2.95	-2.02	2.48
1049	50S ribosomal protein L7/L12	I	0.67	0.08	3	0.48	0.00	10	0.53	0.00	15	0.56	-1.74	-3.42	-3.61	2.92
1151	tryptophan synthase alpha chain	I	1.28	0.30	4	0.28	0.00	4	0.45	0.31	4	0.55	1.05	-2.92	-1.01	0.96
137	pyruvate formate-lyase	I	0.74	0.00	17	0.46	0.00	20	0.48	0.00	22	0.54	-2.90	-6.45	-4.84	4.73
1150	tryptophan synthase beta chain	I	1.21	0.16	6	0.32	0.00	4	0.36	0.04	5	0.52	1.40	-3.10	-2.06	1.25
2009	Xylulose-5-phosphate phosphoketolase	I	0.79	0.02	17	0.31	0.00	18	0.41	0.00	22	0.47	-2.35	-5.99	-4.64	4.33
1942	phosphoglycerate kinase	I	0.45	0.00	28	0.45	0.00	42	0.46	0.00	24	0.46	-6.22	-7.07	-5.89	6.39
1940	phosphoglycerate mutase	I	0.43	0.00	9	0.43	0.00	13	0.47	0.00	13	0.44	-3.45	-4.71	-4.34	4.16
1205	50S ribosomal protein L18	I				0.40	0.01	3	0.48	0.05	5	0.44		-2.66	-1.98	2.32
349	translation initiation factor IF-3	I	0.78	0.10	4	0.27	0.00	6	0.32	0.01	5	0.41	-1.65	-3.25	-2.68	2.53
2145	acetaldehyde dehydrogenase	I	0.40	0.00	45	0.24	0.00	74	0.30	0.00	65	0.31	-7.35	-9.86	-8.68	8.63
2146	Aminoglycoside phosphotransferase	I	0.34	0.02	5	0.20	0.00	21	0.24	0.00	14	0.25	-2.27	-5.20	-4.04	3.83

Appendix 12 iTRAQ proteomics expression ratios between citrus pectin-grown *M. pectinilyticus* cells and fructose-grown cells. E, extracellular; I; intracellular, L, lipoprotein signal; R1, replicate 1; R2, replicate 2; R3, replicate 3; R mean, average R values; P1 – P3, *p*-values for each replicate; Z1, z-score for P1; Z2, z-score for P2; Z3, z-score for P3; abs Z, average z-score. Red, up-regulated (R mean > 1.2); yellow, not differentiated; green; down-regulated (R mean < 0.83). R mean values are statistically significant if absZ is > 1.65.

Locus tag (B9O19)	Gene product description		R1	P1	# of peptides (95%)	R2	P2	# of peptides (95%)	R3	P3	# of peptides (95%)	R mean	z1	z2	z3	abs z
1475	peptidase C1A papain	E	1.72	0.40	3	1.73	0.00	13	1.61	0.02	17	1.69	0.85	2.85	2.33	2.01
1680	Transketolase	I	1.59	0.21	4	1.41	0.10	10	1.75	0.01	11	1.58	1.27	1.66	2.78	1.90
1045	glucose-1-phosphate adenylyltransferase	I	3.29	0.39	3	1.00	1.00	7	1.13	0.35	9	1.55	0.86	0.00	0.93	0.60
916	N-acetylmuramoyl-L-alanine amidase	E	1.31	0.14	6	1.35	0.16	5	2.01	0.01	5	1.53	1.47	1.41	2.54	1.81
1947	elongation factor G	I	1.48	0.07	5	1.22	0.07	17	1.63	0.00	14	1.43	1.81	1.80	4.38	2.66
611	S-layer domain-containing protein	E	1.46	0.32	4	1.22	0.12	12	1.61	0.00	9	1.42	0.99	1.58	3.30	1.95
1044	glucose-1-phosphate adenylyltransferase	I	1.11	0.60	4	1.37	0.15	9	1.67	0.00	6	1.37	0.53	1.43	2.89	1.62
1554	hypothetical protein	I	1.94	0.01	3	0.85	0.39	6	1.31	0.48	9	1.29	2.45	-0.85	0.70	0.77
52	S-layer domain-containing protein	E	1.36	0.35	4	1.05	0.79	11	1.42	0.03	11	1.26	0.94	0.26	2.18	1.13
870	extracellular solute-binding protein family 1	L	0.99	0.98	5	1.19	0.03	21	1.59	0.00	15	1.23	-0.02	2.14	3.69	1.94
2227	Periplasmic sugar-binding proteins	L	0.69	0.43	5	1.32	0.27	7	1.82	0.05	7	1.18	-0.80	1.10	1.93	0.74
682	hypothetical protein	E	1.58	0.10	6	0.95	0.45	12	1.10	0.34	14	1.18	1.63	-0.76	0.96	0.61
826	Pectinesterase (CE8) / S-layer domain-containing protein	E	0.99	0.94	20	1.03	0.71	47	1.53	0.00	45	1.16	-0.08	0.38	4.07	1.46
2005	2-deoxy-D-gluconate 3-dehydrogenase	I	0.69	0.04	19	1.47	0.00	33	1.53	0.00	40	1.16	-2.03	3.32	3.95	1.75
1570	dihydroxy-acid dehydratase	I	0.99	0.97	5	1.15	0.22	16	1.35	0.03	15	1.15	-0.04	1.23	2.13	1.11
273	phosphoglucosamine mutase	I	1.4	0.60	4	1.05	0.49	8	0.96	0.78	8	1.12	0.53	0.68	-0.28	0.31
1678	argininosuccinate synthase	I	1.04	0.94	3	1.04	0.43	9	1.29	0.00	12	1.12	0.08	0.78	3.13	1.33
2084	ATP synthase beta chain	I	0.79	0.40	4	1.21	0.27	5	1.40	0.19	6	1.10	-0.85	1.11	1.30	0.52
898	4-deoxy-L-threo-5-hexosulose-uronateketol-isomerase	I	0.91	0.76	4	0.99	0.96	7	1.45	0.04	8	1.10	-0.30	-0.05	2.03	0.56
2204	IMP cyclohydrolase	I	1.28	0.50	6	0.98	0.85	17	1.04	0.70	14	1.10	0.68	-0.19	0.38	0.29
876	Uronate isomerase	I	0.93	0.81	6	1.14	0.33	10	1.21	0.11	11	1.09	-0.25	0.96	1.62	0.78
751	transcriptional regulator, CdaR	I	0.59	0.00	7	1.29	0.03	17	1.70	0.00	20	1.09	-2.89	2.15	3.57	0.94
198	phosphoribosylamine--glycine ligase	I	1.54	0.48	6	0.85	0.35	5	0.98	0.92	8	1.08	0.70	-0.93	-0.10	0.11

1888	oxaloacetate decarboxylase alpha chain	I	0.85	0.68	4	1.15	0.04	17	1.27	0.14	13	1.07	-0.41	2.02	1.48	1.03
1046	starch synthase	I	1.08	0.79	5	1.02	0.78	8	1.11	0.43	10	1.07	0.27	0.27	0.79	0.44
1225	S-layer domain-containing protein	E	1.12	0.17	5	0.79	0.01	15	1.29	0.02	21	1.04	1.38	-2.47	2.27	0.39
1005	transcriptional regulator	I	0.52	0.26	11	1.30	0.08	23	1.67	0.04	20	1.04	-1.12	1.74	2.04	0.89
2157	Periplasmic [Fe] hydrogenase large subunit	I	0.91	0.76	6	0.99	0.91	13	1.20	0.17	11	1.03	-0.31	-0.12	1.37	0.31
1597	S-layer domain-containing protein	E	0.95	0.62	92	0.93	0.08	120	1.19	0.00	147	1.02	-0.50	-1.75	3.15	0.30
2196	phosphoribosylformylglycinamidine synthase	I	1.05	0.72	14	0.90	0.34	23	1.09	0.19	19	1.01	0.35	-0.95	1.32	0.24
2203	IMP cyclohydrolase	I	0.81	0.41	4	1.18	0.03	14	1.08	0.15	14	1.01	-0.82	2.17	1.44	0.93
1898	fructose-bisphosphate aldolase	I	0.61	0.04	5	1.14	0.15	11	1.46	0.00	12	1.00	-2.04	1.43	2.82	0.74
1446	DNA-directed RNA polymerase, beta subunit	I	1.01	0.95	7	0.88	0.08	20	1.11	0.22	28	1.00	0.06	-1.75	1.23	0.15
2161	L-arabinose isomerase	I	0.65	0.06	10	1.13	0.06	26	1.33	0.00	19	0.99	-1.90	1.85	3.44	1.13
607	hypothetical protein	I	0.88	0.27	8	0.89	0.27	12	1.23	0.21	12	0.99	-1.11	-1.11	1.25	0.32
1243	DNA gyrase subunit A	I				1.07	0.39	7	0.88	0.68	7	0.97		0.87	-0.42	0.22
1520	transcriptional regulator	I	0.24	0.04	4	1.97	0.03	11	1.94	0.12	7	0.97	-2.05	2.21	1.56	0.58
827	carbamoyl-phosphate synthase large chain	I	0.74	0.08	4	1.00	1.00	19	1.22	0.03	23	0.97	-1.75	0.01	2.17	0.14
1346	aspartate transaminase	I	0.72	0.75	3	1.00	1.00	8	1.25	0.16	5	0.97	-0.32	0.00	1.40	0.36
1426	2-dehydro-3-deoxy gluconokinase KdgK	I	0.9	0.58	13	0.98	0.81	14	1.01	0.94	17	0.96	-0.55	-0.24	0.08	0.24
1969	orotidine 5'-phosphate decarboxylase	I	1.08	0.84	3	0.78	0.08	6	1.05	0.81	4	0.96	0.21	-1.75	0.24	0.43
602	Heat shock protein 60 family chaperone GroEL	I	0.52	0.01	17	1.18	0.00	72	1.41	0.03	71	0.96	-2.67	3.07	2.22	0.87
1425	2-dehydro-3-deoxy phosphogluconate aldolase	I	0.45	0.00	18	1.29	0.01	27	1.50	0.02	24	0.96	-3.05	2.44	2.35	0.58
1972	orotate phosphoribosyltransferase	I	0.81	0.33	4	0.94	0.72	10	1.13	0.15	8	0.95	-0.98	-0.36	1.44	0.03
914	S-layer domain-containing protein	E	1.43	0.25	19	0.66	0.00	67	0.91	0.19	64	0.95	1.15	-6.78	-1.32	2.32
901	LL-diaminopimelate aminotransferase	I	0.89	0.73	11	0.86	0.11	21	1.12	0.23	28	0.95	-0.35	-1.58	1.21	0.24
743	Propeptide PepSY amd peptidase M4	E	0.99	0.96	8	0.83	0.14	13	1.02	0.86	13	0.94	-0.05	-1.46	0.17	0.45
1577	Dockerin-like domain	E	0.73	0.15	6	0.86	0.03	9	1.31	0.05	9	0.94	-1.45	-2.23	1.97	0.57
1522	malate dehydrogenase	I	0.86	0.68	10	0.79	0.06	12	1.20	0.29	14	0.93	-0.41	-1.85	1.07	0.40
972	Thioredoxin	I	1.25	0.82	6	0.74	0.11	10	0.87	0.37	9	0.93	0.23	-1.60	-0.90	0.76
287	3-dehydroquinate synthase	I				0.89	0.18	7	0.96	0.81	9	0.92		-1.34	-0.24	0.79
2111	D-3-phosphoglycerate dehydrogenase	I	1.05	0.86	12	0.83	0.00	23	0.89	0.18	22	0.92	0.18	-3.06	-1.33	1.40
692	Transketolase, C-terminal domain	I	0.7	0.29	8	0.99	0.96	13	1.12	0.43	13	0.92	-1.06	-0.05	0.78	0.11

506	O-acetylhomoserine sulfhydrylase	I	0.88	0.70	4	0.85	0.07	7	1.01	0.94	5	0.91	-0.38	-1.82	0.08	0.71
1090	N-acetyl-gamma-glutamyl-phosphatereductase	I				0.84	0.07	7	0.98	0.87	6	0.91		-1.84	-0.16	1.00
608	NIF system FeS cluster assembly, NifU, N-terminal	I	0.77	0.14	9	0.95	0.39	12	1.03	0.77	14	0.91	-1.47	-0.86	0.29	0.68
1928	band 7 protein	I	0.48	0.01	16	1.11	0.12	21	1.38	0.00	18	0.91	-2.49	1.57	2.93	0.67
1192	50S ribosomal protein L2	I	2.23	0.16	4	0.46	0.00	9	0.69	0.05	11	0.89	1.40	-3.03	-1.94	1.19
949	homoserine dehydrogenase	I				0.74	0.01	11	1.08	0.63	7	0.89		-2.68	0.48	1.10
414	SCP-2 sterol transfer family	I				0.90	0.51	6	0.88	0.18	6	0.89		-0.66	-1.35	1.01
1303	Cell division protein FtsZ	I	0.74	0.49	18	0.90	0.08	19	1.03	0.73	20	0.88	-0.69	-1.78	0.34	0.71
2045	glycine hydroxymethyltransferase	I	0.87	0.59	8	0.81	0.00	17	0.97	0.84	16	0.88	-0.55	-3.31	-0.20	1.35
1287	Rubrerythrin	I	0.41	0.13	5	1.24	0.24	8	1.33	0.14	8	0.88	-1.53	1.18	1.48	0.38
368	2-oxoacid ferredoxin oxidoreductase	I	0.85	0.54	3	0.82	0.02	10	0.96	0.67	16	0.87	-0.61	-2.31	-0.43	1.12
366	amino acid-binding ACT domain protein	I	0.74	0.71	3	0.90	0.36	9	0.99	0.97	8	0.87	-0.37	-0.92	-0.04	0.44
909	Pectate lyase (PL1) / S-layer domain-containing protein	E	1.08	0.67	20	0.66	0.00	64	0.91	0.08	63	0.87	0.43	-8.08	-1.75	3.13
129	histidine--tRNA ligase	I	1.38	0.47	5	0.50	0.04	4	0.95	0.45	6	0.87	0.72	-2.05	-0.75	0.70
1271	Translation elongation factor Tu	I	0.52	0.00	36	0.97	0.64	58	1.26	0.00	60	0.86	-2.83	-0.47	3.14	0.05
1219	DNA-directed RNA polymerase subunit alpha	I	0.8	0.39	3	0.85	0.12	9	0.90	0.21	9	0.85	-0.86	-1.54	-1.26	1.22
1445	DNA-directed RNA polymerase, beta subunit	I	0.63	0.03	10	0.90	0.06	26	1.06	0.54	29	0.85	-2.15	-1.86	0.61	1.13
1178	ornithine carbamoyltransferase	I	1.02	0.98	5	0.67	0.01	11	0.89	0.30	12	0.84	0.03	-2.62	-1.03	1.21
1330	pyruvate synthase	I	0.74	0.01	32	0.79	0.00	45	1.00	0.95	50	0.84	-2.53	-4.49	-0.06	2.36
727	NADP-specific glutamate dehydrogenase	I	0.83	0.37	9	0.75	0.00	15	0.93	0.26	12	0.83	-0.89	-3.73	-1.12	1.92
1337	S-layer domain-containing protein	E	1.14	0.91	5	0.66	0.01	9	0.77	0.36	6	0.83	0.12	-2.63	-0.92	1.14
389	peptidase M18 aminopeptidase I	I				0.78	0.08	8	0.86	0.26	6	0.81		-1.74	-1.13	1.43
1595	S-layer domain-containing protein	E	1.2	0.31	7	0.64	0.02	10	0.70	0.00	11	0.81	1.03	-2.43	-2.96	1.46
1089	Multiple sugar ABC transporter, ATP-binding protein	I	0.3	0.09	4	1.25	0.08	12	1.42	0.00	10	0.81	-1.71	1.77	2.93	1.00
1273	2-isopropylmalate synthase	I	0.65	0.18	4	0.89	0.14	19	0.92	0.39	16	0.81	-1.33	-1.47	-0.85	1.22
1010	6-phosphofructokinase	I	1.12	0.78	7	0.61	0.00	16	0.76	0.00	15	0.81	0.28	-4.50	-3.26	2.49
947	aspartatekinase	I	1.01	0.98	12	0.69	0.02	10	0.74	0.05	10	0.80	0.03	-2.27	-1.98	1.41
1372	ribosome recycling factor	I	0.69	0.46	6	0.84	0.11	13	0.87	0.40	10	0.80	-0.73	-1.61	-0.84	1.06
908	S-layer domain-containing protein	E	0.67	0.16	13	0.71	0.00	46	1.06	0.47	50	0.80	-1.39	-4.60	0.73	1.75
724	tetrahydrofolate synthase	I	0.89	0.45	3	0.72	0.12	7	0.79	0.11	9	0.80	-0.76	-1.55	-1.59	1.30

1892	acetyl-CoA carboxylase	I	0.75	0.09	10	0.76	0.00	20	0.87	0.03	19	0.79	-1.68	-4.46	-2.11	2.75
1953	ribose-phosphate diphosphokinase	I	0.98	0.92	5	0.53	0.00	5	0.95	0.67	7	0.79	-0.10	-2.87	-0.42	1.13
1858	malonyl CoA-acyl carrier protein transacylase	I	0.67	0.15	8	0.72	0.01	15	0.98	0.90	12	0.78	-1.44	-2.77	-0.12	1.45
2018	translation elongation factor P	I	0.72	0.36	3	0.78	0.02	6	0.85	0.16	8	0.78	-0.92	-2.28	-1.42	1.54
277	proline--tRNA ligase	I	1.01	0.96	6	0.66	0.01	5	0.72	0.24	7	0.78	0.04	-2.50	-1.17	1.21
403	Transcriptional regulator, GntR family	I	0.57	0.01	5	0.79	0.01	17	1.05	0.71	15	0.78	-2.54	-2.79	0.37	1.65
1195	30S ribosomal protein S3	I	1.12	0.69	5	0.53	0.00	11	0.80	0.24	11	0.78	0.40	-3.16	-1.17	1.31
1265	pyruvate phosphate dikinase	I	0.71	0.17	18	0.74	0.00	33	0.86	0.05	36	0.77	-1.36	-3.84	-1.97	2.39
617	30S ribosomal protein S2	I	0.73	0.72	5	0.65	0.00	12	0.94	0.61	15	0.76	-0.35	-3.69	-0.52	1.52
2082	ATP synthase alpha chain	I	0.86	0.81	4	0.56	0.07	5	0.89	0.41	5	0.75	-0.24	-1.80	-0.82	0.95
1155	CTPsynthase	I	0.88	0.45	5	0.60	0.05	9	0.80	0.23	10	0.75	-0.76	-1.98	-1.19	1.31
1821	acetolactate synthase	I	0.69	0.07	4	0.61	0.00	13	0.99	0.93	8	0.75	-1.82	-4.18	-0.08	2.03
1513	Chaperone protein DnaK	I	0.57	0.07	8	0.69	0.00	20	1.03	0.72	20	0.74	-1.81	-4.68	0.36	2.04
89	phosphate acetyltransferase	I	0.68	0.08	6	0.71	0.00	10	0.83	0.05	11	0.73	-1.75	-4.03	-1.97	2.59
678	adenylosuccinate synthase	I	0.45	0.22	6	0.94	0.44	13	0.94	0.49	15	0.73	-1.24	-0.77	-0.69	0.90
616	translation elongation factor Ts	I	0.92	0.90	8	0.61	0.00	16	0.71	0.01	18	0.73	-0.13	-4.79	-2.70	2.54
912	S-layer domain-containing protein	E	0.71	0.06	10	0.60	0.00	31	0.90	0.34	31	0.73	-1.86	-5.43	-0.95	2.75
1903	aconitate hydratase	I	0.7	0.16	20	0.70	0.00	28	0.77	0.00	27	0.72	-1.42	-3.70	-3.58	2.90
1826	3-isopropylmalate dehydratase small subunit	I	0.59	0.30	5	0.66	0.01	4	0.96	0.86	5	0.72	-1.04	-2.53	-0.18	1.25
919	glyceraldehyde-3-phosphate dehydrogenase	I	0.43	0.01	19	0.89	0.10	27	0.96	0.77	26	0.72	-2.49	-1.66	-0.29	1.48
950	hypothetical protein	I				0.70	0.04	4	0.73	0.05	6	0.72		-2.02	-1.96	1.99
540	GMP synthase [glutamine-hydrolyzing]	I	0.77	0.46	8	0.65	0.00	14	0.73	0.06	12	0.71	-0.74	-3.06	-1.90	1.90
1578	S-layer domain-containing protein	E	0.62	0.01	28	0.64	0.00	53	0.89	0.08	58	0.71	-2.72	-4.58	-1.76	3.02
1066	Arginine-binding extracellular protein ArtP	L	0.56	0.21	5	0.64	0.01	8	0.99	0.93	10	0.71	-1.25	-2.56	-0.09	1.30
90	acetate kinase	I	0.67	0.10	6	0.64	0.00	14	0.80	0.01	17	0.70	-1.64	-3.67	-2.46	2.59
1827	3-isopropylmalate dehydrogenase	I	1.21	0.70	3	0.44	0.02	6	0.63	0.00	6	0.70	0.39	-2.40	-2.83	1.62
775	aspartate-semialdehyde dehydrogenase	I	0.7	0.30	13	0.64	0.00	16	0.74	0.01	16	0.69	-1.05	-4.80	-2.68	2.84
2112	phosphoserine transaminase	I	0.37	0.00	12	0.88	0.18	18	0.99	0.91	18	0.69	-2.82	-1.33	-0.12	1.42
1270	Translation elongation factor G	I	0.88	0.35	19	0.56	0.00	37	0.66	0.00	40	0.69	-0.93	-6.66	-5.44	4.34
92	histidinol dehydrogenase	I				0.61	0.00	10	0.76	0.06	7	0.68		-4.31	-1.86	3.09

1823	Ketol-acid reductoisomerase	I	0.54	0.07	9	0.70	0.00	27	0.82	0.03	20	0.67	-1.79	-3.58	-2.21	2.53
1608	phosphoenolpyruvate-protein phosphotransferase	I	0.7	0.40	4	0.65	0.03	6	0.67	0.14	5	0.67	-0.85	-2.21	-1.46	1.51
1051	50S ribosomal protein L1	I	1.01	0.96	6	0.48	0.00	12	0.60	0.00	9	0.67	0.04	-4.68	-3.14	2.59
128	aspartate--tRNA ligase	I	1.04	0.77	4	0.50	0.00	10	0.56	0.01	11	0.66	0.30	-3.66	-2.58	1.98
1190	50S ribosomal protein L4	I	1.17	0.55	6	0.42	0.00	11	0.58	0.05	11	0.66	0.60	-4.63	-1.97	2.00
1386	30S ribosomal protein S16	I	1.17	0.82	3	0.38	0.01	3	0.64	0.02	4	0.66	0.23	-2.63	-2.32	1.58
1269	30Sribosomal protein S7	I	0.77	0.52	3	0.55	0.00	15	0.65	0.00	11	0.65	-0.65	-3.82	-3.29	2.59
2138	polyribonucleotide nucleotidyltransferase	I	0.48	0.01	5	0.66	0.01	12	0.85	0.32	11	0.65	-2.74	-2.63	-0.99	2.12
627	aspartate--tRNA ligase	I	0.85	0.44	7	0.51	0.00	9	0.61	0.00	11	0.64	-0.78	-4.31	-3.32	2.80
1201	50S ribosomal protein L5	I	0.87	0.45	3	0.54	0.00	12	0.54	0.00	11	0.63	-0.76	-3.95	-3.37	2.69
1189	50S ribosomal protein L3	I	1.03	0.97	3	0.45	0.00	6	0.52	0.01	4	0.62	0.04	-3.25	-2.63	1.95
1626	Pectinesterase (CE8) / S-layer domain-containing protein	E	0.73	0.03	26	0.48	0.00	40	0.68	0.00	34	0.62	-2.12	-6.84	-4.59	4.52
541	amidophosphoribosyltransferase	I	0.56	0.06	7	0.61	0.00	17	0.66	0.00	15	0.61	-1.89	-3.85	-3.04	2.92
59	branched-chain-amino-acid transaminase	I	0.44	0.32	3	0.69	0.00	7	0.73	0.00	9	0.61	-1.00	-2.87	-3.17	2.35
542	phosphoribosylaminoimidazolesuccinocarboxamide synthase	I	0.52	0.02	9	0.59	0.00	10	0.73	0.03	8	0.61	-2.24	-4.26	-2.23	2.91
289	3-deoxy-7-phosphoheptulonate synthase	I	0.34	0.02	6	0.66	0.00	11	0.98	0.86	10	0.60	-2.27	-3.54	-0.17	2.00
357	Cell division trigger factor	I				0.54	0.00	14	0.68	0.03	6	0.60		-5.71	-2.14	3.92
1206	30Sribosomal protein S5	I	0.9	0.70	9	0.47	0.00	12	0.51	0.02	9	0.60	-0.39	-3.36	-2.36	2.04
127	tyrosine--tRNA ligase	I				0.63	0.18	6	0.57	0.02	5	0.60		-1.35	-2.43	1.89
1050	50S ribosomal protein L10	I	0.76	0.27	5	0.42	0.00	7	0.65	0.01	14	0.59	-1.11	-4.25	-2.65	2.67
1203	30Sribosomal protein S8	I	1	0.99	4	0.46	0.00	8	0.45	0.00	10	0.59	0.01	-3.60	-2.90	2.17
1091	acetylglutamate kinase	I	0.68	0.63	3	0.52	0.01	11	0.57	0.00	9	0.59	-0.49	-2.51	-2.83	1.94
1941	triose-phosphate isomerase	I	0.68	0.17	8	0.48	0.00	11	0.61	0.00	11	0.58	-1.38	-3.54	-2.95	2.62
1902	isocitrate dehydrogenase	I	0.56	0.05	7	0.55	0.00	15	0.61	0.00	24	0.58	-1.92	-5.14	-4.40	3.82
1105	Glutamine synthetase	I	0.61	0.00	10	0.48	0.00	12	0.64	0.07	10	0.57	-3.02	-4.47	-1.82	3.10
1156	S-layer domain-containing protein	E	0.42	0.00	121	0.55	0.00	336	0.81	0.01	354	0.57	-5.14	-9.08	-2.56	5.60
1607	Phosphotransferase system, phosphocarrier protein HPr	I	0.5	0.21	4				0.66	0.04	4	0.57	-1.24		-2.03	1.63
1097	Phosphoenolpyruvate carboxykinase [GTP]	I	0.41	0.03	25	0.61	0.00	28	0.67	0.00	29	0.55	-2.23	-5.15	-3.87	3.75
1379	50S ribosomal protein L27	I				0.48	0.01	5	0.62	0.01	5	0.54		-2.73	-2.47	2.60
1606	PTS system fructose-specific EIIABC component	I				0.50	0.00	6	0.58	0.15	8	0.54		-3.75	-1.43	2.59

326	Rubrerythrin	I	0.25	0.06	4	0.71	0.00	5	0.82	0.32	5	0.53	-1.86	-2.83	-1.00	1.90
1052	50S ribosomal protein L11	I	0.77	0.34	4	0.38	0.00	7	0.46	0.01	8	0.51	-0.96	-3.35	-2.54	2.28
1627	Pectinesterase (CE8) / S-layer domain-containing protein	E	1.03	0.77	47	0.33	0.00	26	0.38	0.00	38	0.50	0.30	-5.26	-3.76	2.91
356	ATP-dependent Clp endopeptidase, proteolytic subunit ClpP	I				0.44	0.01	5	0.57	0.07	6	0.50		-2.55	-1.79	2.17
137	pyruvate formate-lyase	I	0.57	0.04	17	0.43	0.00	20	0.49	0.00	22	0.49	-2.01	-6.32	-4.50	4.27
1150	tryptophan synthase beta chain	I	1.26	0.36	6	0.29	0.00	4	0.32	0.01	5	0.49	0.92	-3.91	-2.71	1.90
1218	30Sribosomal protein S4	I	0.39	0.38	4	0.51	0.00	8	0.58	0.00	7	0.49	-0.88	-4.25	-3.15	2.76
1151	tryptophan synthase alpha chain	I	0.73	0.05	4	0.27	0.02	4	0.45	0.25	4	0.45	-1.98	-2.36	-1.16	1.83
1049	50S ribosomal protein L7/L12	I	0.42	0.13	3	0.36	0.00	10	0.46	0.00	15	0.41	-1.52	-3.46	-3.03	2.67
1205	50S ribosomal protein L18	I				0.31	0.02	3	0.47	0.04	5	0.38		-2.28	-2.01	2.15
1942	phosphoglycerate kinase	I	0.24	0.00	28	0.42	0.00	42	0.47	0.00	24	0.36	-4.14	-7.04	-5.19	5.46
349	translation initiation factor IF-3	I	0.47	0.03	4	0.29	0.01	6	0.31	0.00	5	0.35	-2.11	-2.78	-3.00	2.63
2009	Xylulose-5-phosphate phosphoketolase	I	0.34	0.00	17	0.30	0.00	18	0.41	0.00	22	0.35	-2.99	-5.68	-3.45	4.04
1940	phosphoglycerate mutase	I	0.25	0.00	9	0.38	0.00	13	0.40	0.00	13	0.33	-3.22	-4.27	-3.88	3.79
2145	acetaldehyde dehydrogenase	I	0.26	0.00	45	0.23	0.00	74	0.31	0.00	65	0.27	-6.23	-9.70	-7.84	7.92
2146	Aminoglycoside phosphotransferase	I	0.18	0.01	5	0.20	0.00	21	0.24	0.00	14	0.21	-2.68	-5.07	-3.89	3.88

Appendix 13 iTRAQ proteomics expression ratios between kiwifruit pectin-grown *M. pectinilyticus* cells and fructose-grown cells. E, extracellular; I; intracellular, L, lipoprotein signal; R1, replicate 1; R2, replicate 2; R3, replicate 3; R mean, average R values; P1 – P3, *p*-values for each replicate; Z1, z-score for P1; Z2, z-score for P2; Z3, z-score for P3; abs Z, average z-score. Red, up-regulated (R mean > 1.2); yellow, not differentiated; green; down-regulated (R mean < 0.83). R mean values are statistically significant if absZ is > 1.65.

Locus tag (B9O19)	Gene product description		R1	P1	# of peptides (95%)	R2	P2	# of peptides (95%)	R3	P3	# of peptides (95%)	R mean	z1	z2	z3	abs z
682	hypothetical protein	E	5.68	0.00	6	2.85	0.00	12	3.25	0.00	14	3.75	3.67	3.89	3.53	3.69
1475	peptidase C1A papain	E	3.92	0.18	3	2.43	0.00	13	1.84	0.00	17	2.60	1.34	3.79	3.62	2.92
611	S-layer domain-containing protein	E	2.26	0.02	4	2.04	0.00	12	2.51	0.00	9	2.26	2.34	3.79	4.34	3.49
1520	transcriptional regulator	I	2.29	0.03	4	1.94	0.02	11	2.07	0.10	7	2.10	2.13	2.39	1.62	2.05
1156	S-layer domain-containing protein	E	1.50	0.00	121	1.60	0.00	336	1.72	0.00	354	1.60	5.28	6.65	5.29	5.74
52	S-layer domain-containing protein	E	1.38	0.22	4	1.51	0.09	11	1.86	0.00	11	1.57	1.22	1.72	3.04	1.99
1627	Pectinesterase (CE8) / S-layer domain-containing protein	E	1.57	0.00	47	1.37	0.00	26	1.42	0.00	38	1.45	5.13	4.31	5.18	4.87
1947	elongation factor G	I	1.74	0.00	5	1.19	0.08	17	1.41	0.00	14	1.43	3.66	1.74	3.01	2.80
2227	Periplasmic sugar-binding proteins	L	0.39	0.02	5	2.37	0.01	7	2.98	0.02	7	1.40	-2.29	2.47	2.30	0.83
602	Heat shock protein 60 family chaperone GroEL	I	0.78	0.02	17	1.68	0.00	72	1.92	0.00	71	1.36	-2.30	4.69	4.29	2.23
1044	glucose-1-phosphate adenylyltransferase	I	1.59	0.14	4	1.06	0.50	9	1.18	0.12	6	1.26	1.48	0.67	1.54	1.23
916	N-acetylmuramoyl-L-alanine amidase	E	0.81	0.06	6	1.36	0.37	5	1.78	0.05	5	1.25	-1.88	0.91	1.96	0.33
1554	hypothetical protein	I	1.44	0.14	3	0.95	0.77	6	1.35	0.33	9	1.23	1.48	-0.30	0.97	0.72
2157	Periplasmic [Fe] hydrogenase large subunit	I	1.13	0.51	6	1.12	0.17	13	1.42	0.31	11	1.22	0.66	1.38	1.01	1.02
1089	Multiple sugar ABC transporter, ATP-binding protein	I	0.98	0.89	4	1.11	0.48	12	1.56	0.05	10	1.19	-0.14	0.70	1.94	0.83
1425	2-dehydro-3-deoxy phosphogluconate aldolase	I	1.03	0.81	18	1.13	0.28	27	1.44	0.00	24	1.18	0.24	1.07	3.15	1.49
1046	starch synthase	I	1.14	0.73	5	1.16	0.33	8	1.19	0.19	10	1.16	0.35	0.98	1.32	0.88
870	extracellular solute-binding protein family 1	L	1.23	0.39	5	1.07	0.44	21	1.17	0.03	15	1.15	0.86	0.77	2.12	1.25
1597	S-layer domain-containing protein	E	1.48	0.00	92	0.86	0.00	120	1.13	0.03	147	1.13	3.49	-2.95	2.24	0.93
826	Pectinesterase (CE8) / S-layer domain-containing protein	E	1.33	0.02	20	0.92	0.28	47	1.14	0.16	45	1.12	2.39	-1.08	1.40	0.90
2084	ATP synthase beta chain	I	0.66	0.41	4	1.40	0.06	5	1.46	0.07	6	1.10	-0.82	1.90	1.79	0.95
876	Uronate isomerase	I	1.09	0.84	6	1.01	0.95	10	1.16	0.23	11	1.08	0.20	0.06	1.21	0.49

2204	IMP cyclohydrolase	I	1.54	0.45	6	0.84	0.19	17	0.98	0.86	14	1.08	0.75	-1.30	-0.18	0.24
1005	transcriptional regulator	I	1.32	0.20	11	0.86	0.52	23	1.11	0.61	20	1.08	1.29	-0.65	0.51	0.38
2203	IMP cyclohydrolase	I	1.02	0.95	4	1.12	0.26	14	1.10	0.21	14	1.08	0.07	1.13	1.25	0.82
1680	Transketolase	I	0.79	0.25	4	1.17	0.44	10	1.34	0.01	11	1.08	-1.16	0.77	2.46	0.69
692	Transketolase, C-terminal domain	I	1.36	0.35	8	0.92	0.54	13	0.99	0.91	13	1.07	0.93	-0.62	-0.11	0.07
751	transcriptional regulator, CdaR	I	0.91	0.50	7	1.07	0.50	17	1.25	0.07	20	1.07	-0.67	0.67	1.84	0.61
1225	S-layer domain-containing protein	E	1.12	0.56	5	0.86	0.22	15	1.24	0.09	21	1.06	0.58	-1.24	1.71	0.35
1287	Rubrerythrin	I	0.77	0.35	5	1.27	0.26	8	1.23	0.25	8	1.06	-0.94	1.13	1.15	0.45
898	4-deoxy-L-threo-5-hexosulose-uronateketol-isomerase	I	0.84	0.33	4	1.08	0.63	7	1.31	0.14	8	1.06	-0.97	0.49	1.47	0.33
1045	glucose-1-phosphate adenylyltransferase	I	1.26	0.11	3	0.86	0.74	7	1.08	0.62	9	1.06	1.58	-0.34	0.49	0.58
1522	malate dehydrogenase	I	0.93	0.69	10	0.96	0.74	12	1.30	0.13	14	1.05	-0.40	-0.33	1.52	0.26
1577	Dockerin-like domain	E	2.06	0.02	6	0.69	0.19	9	0.79	0.34	9	1.04	2.33	-1.31	-0.96	0.02
1928	band 7 protein	I	0.95	0.73	16	0.96	0.68	21	1.22	0.01	18	1.04	-0.34	-0.41	2.56	0.60
2005	2-deoxy-D-gluconate 3-dehydrogenase	I	0.80	0.04	19	1.14	0.20	33	1.14	0.11	40	1.01	-2.05	1.27	1.61	0.28
1570	dihydroxy-acid dehydratase	I	0.96	0.72	5	1.00	0.98	16	1.05	0.76	15	1.00	-0.37	-0.02	0.30	0.03
198	phosphoribosylamine--glycine ligase	I	1.18	0.45	6	0.94	0.73	5	0.91	0.40	8	1.00	0.76	-0.35	-0.84	0.14
1446	DNA-directed RNA polymerase, beta subunit	I	1.06	0.76	7	0.97	0.74	20	0.96	0.55	28	1.00	0.30	-0.33	-0.59	0.21
2161	L-arabinose isomerase	I	1.04	0.84	10	0.89	0.17	26	1.05	0.55	19	0.99	0.20	-1.37	0.59	0.19
1888	oxaloacetate decarboxylase alpha chain	I	1.02	0.87	4	0.93	0.21	17	1.01	0.94	13	0.99	0.17	-1.26	0.08	0.34
1513	Chaperone protein DnaK	I	0.77	0.51	8	0.92	0.14	20	1.33	0.00	20	0.98	-0.65	-1.47	3.82	0.56
2196	phosphoribosylformylglycinamidine synthase	I	1.12	0.49	14	0.88	0.36	23	0.96	0.62	19	0.98	0.69	-0.91	-0.49	0.24
827	carbamoyl-phosphate synthase large chain	I	1.18	0.31	4	0.85	0.05	19	0.93	0.47	23	0.98	1.01	-1.98	-0.72	0.56
1346	aspartate transaminase	I	1.11	0.91	3	0.92	0.60	8	0.89	0.42	5	0.97	0.12	-0.52	-0.81	0.40
908	S-layer domain-containing protein	E	1.29	0.29	13	0.70	0.00	46	0.92	0.31	50	0.94	1.06	-3.73	-1.02	1.23
389	peptidase M18 aminopeptidase I	I				0.78	0.18	8	1.13	0.42	6	0.94		-1.36	0.80	0.28
129	histidine--tRNA ligase	I	0.99	0.88	5	0.61	0.02	4	1.39	0.49	6	0.94	-0.15	-2.33	0.69	0.60
909	Pectate lyase (PL1) / S-layer domain-containing protein	E	1.35	0.05	20	0.73	0.00	64	0.84	0.00	63	0.94	1.94	-4.74	-3.10	1.97
743	Propeptide PepSY amd peptidase M4	E	0.91	0.65	8	0.97	0.82	13	0.91	0.51	13	0.93	-0.45	-0.22	-0.66	0.44
1090	N-acetyl-gamma-glutamyl-phosphatereductase	I				0.81	0.41	7	1.05	0.70	6	0.92		-0.82	0.39	0.21
1330	pyruvate synthase	I	1.23	0.10	32	0.72	0.00	45	0.89	0.10	50	0.92	1.62	-4.87	-1.64	1.63

1271	Translation elongation factor Tu	I	0.97	0.76	36	0.78	0.00	58	1.04	0.63	60	0.92	-0.31	-2.97	0.48	0.93
273	phosphoglucosamine mutase	I	1.09	0.58	4	0.91	0.46	8	0.78	0.01	8	0.92	0.55	-0.73	-2.51	0.89
1898	fructose-bisphosphate aldolase	I	0.86	0.70	5	0.79	0.03	11	1.14	0.19	12	0.92	-0.39	-2.19	1.30	0.43
368	2-oxoacid ferredoxin oxidoreductase	I	1.02	0.92	3	0.81	0.03	10	0.94	0.65	16	0.92	0.10	-2.16	-0.45	0.84
1678	argininosuccinate synthase	I	0.74	0.23	3	1.03	0.76	9	1.01	0.94	12	0.91	-1.19	0.30	0.08	0.27
1445	DNA-directed RNA polymerase, beta subunit	I	0.93	0.71	10	0.84	0.05	26	0.97	0.69	29	0.91	-0.37	-1.95	-0.40	0.91
1303	Cell division protein FtsZ	I	0.94	0.71	18	0.83	0.14	19	0.93	0.54	20	0.90	-0.37	-1.48	-0.61	0.82
617	30S ribosomal protein S2	I	1.48	0.19	5	0.69	0.02	12	0.70	0.02	15	0.89	1.31	-2.24	-2.27	1.07
1243	DNA gyrase subunit A	I				1.05	0.75	7	0.76	0.14	7	0.89		0.32	-1.49	0.58
607	hypothetical protein	I	0.88	0.55	8	0.73	0.03	12	1.07	0.76	12	0.88	-0.60	-2.19	0.31	0.83
1426	2-dehydro-3-deoxy gluconokinase KdgK	I	0.91	0.51	13	0.81	0.00	14	0.92	0.43	17	0.88	-0.66	-3.09	-0.79	1.51
1337	S-layer domain-containing protein	E	0.98	0.94	5	0.82	0.52	9	0.83	0.50	6	0.87	-0.08	-0.64	-0.67	0.46
1607	Phosphotransferase system, phosphocarrier protein HPr	I	1.08	0.81	4				0.71	0.38	4	0.87	0.24		-0.88	0.32
1972	orotate phosphoribosyltransferase	I	0.89	0.47	4	0.80	0.16	10	0.92	0.46	8	0.87	-0.72	-1.41	-0.74	0.96
1969	orotidine 5'-phosphate decarboxylase	I	0.76	0.42	3	0.79	0.11	6	1.08	0.68	4	0.86	-0.81	-1.61	0.41	0.67
1858	malonyl CoA-acyl carrier protein transacylase	I	1.00	0.98	8	0.69	0.01	15	0.92	0.52	12	0.86	0.02	-2.52	-0.64	1.05
506	O-acetylhomoserine sulfhydrylase	I	0.86	0.25	4	0.91	0.66	7	0.81	0.11	5	0.86	-1.15	-0.44	-1.59	1.06
912	S-layer domain-containing protein	E	1.21	0.19	10	0.61	0.00	31	0.86	0.14	31	0.86	1.32	-4.41	-1.48	1.52
727	NADP-specific glutamate dehydrogenase	I	0.96	0.81	9	0.78	0.01	15	0.84	0.23	12	0.86	-0.24	-2.69	-1.19	1.37
914	S-layer domain-containing protein	E	1.53	0.01	19	0.58	0.00	67	0.69	0.00	64	0.85	2.65	-6.77	-5.26	3.12
901	LL-diaminopimelate aminotransferase	I	0.79	0.24	11	0.84	0.02	21	0.90	0.30	28	0.84	-1.19	-2.27	-1.04	1.50
1155	CTPsynthase	I	1.05	0.86	5	0.69	0.04	9	0.83	0.32	10	0.84	0.17	-2.04	-1.00	0.96
949	homoserine dehydrogenase	I				0.71	0.01	11	0.99	0.94	7	0.84		-2.54	-0.07	1.30
1578	S-layer domain-containing protein	E	1.06	0.56	28	0.65	0.00	53	0.85	0.01	58	0.83	0.58	-4.86	-2.54	2.27
2045	glycine hydroxymethyltransferase	I	0.97	0.81	8	0.67	0.00	17	0.88	0.20	16	0.83	-0.24	-3.48	-1.28	1.66
1273	2-isopropylmalate synthase	I	0.53	0.14	4	1.00	0.98	19	1.09	0.58	16	0.83	-1.47	0.02	0.55	0.30
2111	D-3-phosphoglycerate dehydrogenase	I	0.97	0.80	12	0.71	0.01	23	0.78	0.02	22	0.82	-0.25	-2.55	-2.24	1.68
608	NIF system FeS cluster assembly, NifU, N-terminal	I	0.94	0.57	9	0.72	0.01	12	0.80	0.11	14	0.82	-0.56	-2.70	-1.59	1.62
678	adenylosuccinate synthase	I	0.61	0.32	6	0.90	0.51	13	0.96	0.70	15	0.81	-1.00	-0.66	-0.39	0.68
366	amino acid-binding ACT domain protein	I	0.69	0.17	3	0.83	0.31	9	0.90	0.24	8	0.80	-1.38	-1.02	-1.18	1.19

287	3-dehydroquinate synthase	I				0.87	0.62	7	0.74	0.05	9	0.80		-0.50	-1.95	1.23
1219	DNA-directed RNA polymerase subunit alpha	I	0.79	0.31	3	0.74	0.01	9	0.88	0.13	9	0.80	-1.02	-2.58	-1.50	1.70
2082	ATP synthase alpha chain	I	0.75	0.14	4	0.67	0.38	5	0.99	0.93	5	0.79	-1.49	-0.89	-0.09	0.82
1372	ribosome recycling factor	I	0.71	0.10	6	0.80	0.06	13	0.86	0.27	10	0.79	-1.64	-1.86	-1.10	1.53
89	phosphate acetyltransferase	I	0.86	0.33	6	0.66	0.00	10	0.86	0.29	11	0.79	-0.98	-3.69	-1.06	1.91
1606	PTS system fructose-specific EIIABC component	I				0.67	0.03	6	0.92	0.84	8	0.78		-2.21	-0.21	1.21
972	Thioredoxin	I	0.94	0.90	6	0.67	0.41	10	0.76	0.08	9	0.78	-0.12	-0.83	-1.73	0.89
2018	translation elongation factor P	I	1.12	0.41	3	0.63	0.00	6	0.68	0.01	8	0.78	0.83	-2.83	-2.64	1.55
326	Rubrerythrin	I	0.44	0.01	4	0.85	0.64	5	1.21	0.62	5	0.77	-2.51	-0.46	0.50	0.82
1265	pyruvate phosphate dikinase	I	0.80	0.06	18	0.71	0.00	33	0.80	0.00	36	0.77	-1.88	-3.49	-3.48	2.95
403	Transcriptional regulator, GntR family	I	0.79	0.08	5	0.64	0.00	17	0.88	0.32	15	0.76	-1.78	-4.51	-0.99	2.43
1821	acetolactate synthase	I	0.91	0.57	4	0.59	0.00	13	0.83	0.31	8	0.76	-0.57	-3.63	-1.02	1.74
1953	ribose-phosphate diphosphokinase	I	0.82	0.27	5	0.71	0.10	5	0.75	0.12	7	0.76	-1.11	-1.65	-1.56	1.44
1010	6-phosphofructokinase	I	0.84	0.35	7	0.71	0.03	16	0.72	0.00	15	0.75	-0.93	-2.20	-3.16	2.10
2112	phosphoserine transaminase	I	0.68	0.02	12	0.75	0.08	18	0.83	0.07	18	0.75	-2.27	-1.77	-1.83	1.96
1595	S-layer domain-containing protein	E	0.92	0.66	7	0.59	0.01	10	0.77	0.05	11	0.75	-0.44	-2.49	-1.96	1.63
947	aspartatekinase	I	1.04	0.88	12	0.62	0.00	10	0.65	0.03	10	0.75	0.14	-2.81	-2.24	1.63
1178	ornithine carbamoyltransferase	I	0.89	0.27	5	0.60	0.01	11	0.78	0.01	12	0.75	-1.11	-2.77	-2.50	2.13
277	proline--tRNA ligase	I	0.93	0.76	6	0.65	0.00	5	0.68	0.09	7	0.74	-0.30	-3.23	-1.69	1.74
414	SCP-2 sterol transfer family	I				0.70	0.18	6	0.79	0.08	6	0.74		-1.33	-1.75	1.54
775	aspartate-semialdehyde dehydrogenase	I	0.75	0.20	13	0.71	0.00	16	0.77	0.05	16	0.74	-1.29	-3.20	-1.93	2.14
724	tetrahydrofolate synthase	I	0.98	0.97	3	0.63	0.08	7	0.64	0.01	9	0.73	-0.04	-1.75	-2.55	1.45
1608	phosphoenolpyruvate-protein phosphotransferase	I	0.73	0.56	4	0.60	0.00	6	0.87	0.54	5	0.73	-0.59	-3.18	-0.61	1.46
289	3-deoxy-7-phosphoheptulonate synthase	I	0.69	0.19	6	0.66	0.01	11	0.83	0.22	10	0.72	-1.30	-2.44	-1.22	1.65
1626	Pectinesterase (CE8) / S-layer domain-containing protein	E	0.99	0.92	26	0.53	0.00	40	0.72	0.00	34	0.72	-0.11	-5.66	-3.89	3.22
919	glyceraldehyde-3-phosphate dehydrogenase	I	0.63	0.00	19	0.77	0.04	27	0.78	0.01	26	0.72	-3.01	-2.07	-2.48	2.52
1051	50S ribosomal protein L1	I	1.14	0.38	6	0.53	0.00	12	0.62	0.01	9	0.72	0.89	-4.37	-2.49	1.99
128	aspartate--tRNA ligase	I	0.93	0.73	4	0.64	0.02	10	0.62	0.01	11	0.72	-0.34	-2.43	-2.74	1.84
1826	3-isopropylmalate dehydratase small subunit	I	0.61	0.06	5	0.66	0.03	4	0.88	0.65	5	0.71	-1.87	-2.15	-0.45	1.49
616	translation elongation factor Ts	I	1.11	0.68	8	0.54	0.00	16	0.57	0.00	18	0.70	0.42	-3.71	-3.85	2.38

1201	50S ribosomal protein L5	I	1.38	0.24	3	0.47	0.00	12	0.48	0.00	11	0.68	1.18	-3.83	-3.17	1.94
1827	3-isopropylmalate dehydrogenase	I	0.70	0.06	3	0.57	0.05	6	0.79	0.06	6	0.68	-1.90	-1.95	-1.85	1.90
1269	30Sribosomal protein S7	I	1.06	0.69	3	0.52	0.00	15	0.56	0.01	11	0.68	0.39	-3.18	-2.74	1.84
1386	30S ribosomal protein S16	I	1.13	0.52	3	0.41	0.27	3	0.66	0.15	4	0.67	0.64	-1.09	-1.43	0.63
357	Cell division trigger factor	I				0.64	0.00	14	0.70	0.06	6	0.67		-3.36	-1.86	2.61
1189	50S ribosomal protein L3	I	1.18	0.49	3	0.45	0.01	6	0.56	0.01	4	0.67	0.69	-2.63	-2.44	1.46
1903	aconitate hydratase	I	0.67	0.05	20	0.64	0.00	28	0.69	0.01	27	0.67	-1.98	-3.86	-2.62	2.82
1066	Arginine-binding extracellular protein ArtP	L	0.61	0.01	5	0.60	0.08	8	0.78	0.23	10	0.66	-2.52	-1.78	-1.21	1.84
90	acetate kinase	I	0.68	0.00	6	0.58	0.00	14	0.71	0.07	17	0.65	-2.99	-3.25	-1.82	2.69
127	tyrosine--tRNA ligase	I				0.69	0.04	6	0.59	0.02	5	0.64		-2.04	-2.41	2.22
627	aspartate--tRNA ligase	I	0.86	0.47	7	0.50	0.00	9	0.61	0.02	11	0.64	-0.73	-3.46	-2.31	2.17
950	hypothetical protein	I				0.64	0.07	4	0.63	0.02	6	0.64		-1.81	-2.29	2.05
540	GMP synthase [glutamine-hydrolyzing]	I	0.72	0.09	8	0.59	0.00	14	0.60	0.01	12	0.64	-1.70	-3.44	-2.67	2.61
137	pyruvate formate-lyase	I	0.53	0.00	17	0.63	0.00	20	0.77	0.01	22	0.63	-3.96	-4.80	-2.48	3.75
1823	Ketol-acid reductoisomerase	I	0.55	0.01	9	0.64	0.00	27	0.72	0.00	20	0.63	-2.66	-3.47	-2.88	3.00
1206	30Sribosomal protein S5	I	1.30	0.06	9	0.42	0.03	12	0.46	0.01	9	0.63	1.86	-2.22	-2.44	0.93
1379	50S ribosomal protein L27	I				0.60	0.14	5	0.64	0.08	5	0.62		-1.47	-1.76	1.62
59	branched-chain-amino-acid transaminase	I	0.67	0.11	3	0.55	0.04	7	0.63	0.02	9	0.61	-1.58	-2.05	-2.27	1.97
356	ATP-dependent Clp endopeptidase, proteolytic subunit ClpP	I				0.55	0.01	5	0.68	0.02	6	0.61		-2.44	-2.34	2.39
1270	Translation elongation factor G	I	0.83	0.02	19	0.48	0.00	37	0.56	0.00	40	0.61	-2.43	-6.62	-5.72	4.92
1892	acetyl-CoA carboxylase	I	0.58	0.06	10	0.64	0.00	20	0.61	0.00	19	0.61	-1.90	-3.56	-4.00	3.15
1190	50S ribosomal protein L4	I	1.03	0.92	6	0.44	0.00	11	0.48	0.01	11	0.60	0.10	-2.96	-2.43	1.77
92	histidinol dehydrogenase	I				0.64	0.01	10	0.56	0.03	7	0.60		-2.70	-2.20	2.45
541	amidophosphoribosyltransferase	I	0.66	0.03	7	0.55	0.00	17	0.57	0.00	15	0.59	-2.18	-3.60	-3.75	3.18
1218	30Sribosomal protein S4	I	0.99	0.98	4	0.43	0.00	8	0.47	0.00	7	0.59	-0.02	-3.46	-3.38	2.29
1097	Phosphoenolpyruvate carboxykinase [GTP]	I	0.53	0.00	25	0.57	0.00	28	0.66	0.00	29	0.59	-3.20	-5.29	-3.36	3.95
542	phosphoribosylaminoimidazolesuccinocarboxamide synthase	I	0.65	0.02	9	0.47	0.00	10	0.65	0.06	8	0.58	-2.37	-4.13	-1.85	2.78
1195	30S ribosomal protein S3	I	0.76	0.57	5	0.49	0.00	11	0.52	0.02	11	0.58	-0.57	-3.10	-2.29	1.99
2138	polyribonucleotide nucleotidyltransferase	I	0.44	0.02	5	0.63	0.05	12	0.70	0.04	11	0.58	-2.40	-1.93	-2.10	2.14
1105	Glutamine synthetase	I	0.59	0.00	10	0.53	0.00	12	0.57	0.01	10	0.56	-4.09	-4.30	-2.44	3.61

1151	tryptophan synthase alpha chain	I	1.10	0.37	4	0.31	0.01	4	0.50	0.40	4	0.56	0.90	-2.72	-0.85	0.89
1192	50S ribosomal protein L2	I	0.72	0.49	4	0.43	0.00	9	0.56	0.01	11	0.56	-0.69	-4.10	-2.72	2.50
1091	acetylglutamate kinase	I	0.42	0.42	3	0.48	0.00	11	0.84	0.54	9	0.55	-0.80	-3.01	-0.62	1.47
1941	triose-phosphate isomerase	I	0.58	0.02	8	0.50	0.00	11	0.57	0.01	11	0.55	-2.25	-3.04	-2.74	2.68
1150	tryptophan synthase beta chain	I	1.14	0.34	6	0.41	0.03	4	0.33	0.00	5	0.54	0.96	-2.23	-3.02	1.43
1902	isocitrate dehydrogenase	I	0.55	0.01	7	0.47	0.00	15	0.58	0.00	24	0.53	-2.65	-6.53	-4.21	4.46
1203	30Sribosomal protein S8	I	1.17	0.27	4	0.36	0.00	8	0.35	0.01	10	0.53	1.10	-3.33	-2.65	1.63
1052	50S ribosomal protein L11	I	0.91	0.68	4	0.36	0.00	7	0.40	0.00	8	0.51	-0.41	-3.24	-3.31	2.32
1050	50S ribosomal protein L10	I	0.64	0.00	5	0.39	0.00	7	0.52	0.00	14	0.51	-2.99	-4.43	-4.13	3.85
1049	50S ribosomal protein L7/L12	I	0.59	0.01	3	0.35	0.00	10	0.47	0.00	15	0.46	-2.72	-3.10	-3.06	2.96
349	translation initiation factor IF-3	I	0.70	0.13	4	0.33	0.03	6	0.36	0.00	5	0.44	-1.50	-2.16	-3.01	2.22
2009	Xylulose-5-phosphate phosphoketolase	I	0.51	0.00	17	0.34	0.00	18	0.39	0.00	22	0.41	-4.18	-4.40	-3.64	4.07
1942	phosphoglycerate kinase	I	0.31	0.00	28	0.44	0.00	42	0.48	0.00	24	0.41	-6.07	-5.28	-4.62	5.32
1940	phosphoglycerate mutase	I	0.46	0.00	9	0.35	0.00	13	0.37	0.00	13	0.39	-3.25	-4.42	-3.75	3.81
1205	50S ribosomal protein L18	I				0.34	0.05	3	0.40	0.02	5	0.37		-2.00	-2.26	2.13
2145	acetaldehyde dehydrogenase	I	0.37	0.00	45	0.24	0.00	74	0.32	0.00	65	0.30	-7.83	-8.82	-6.76	7.80
2146	Aminoglycoside phosphotransferase	I	0.31	0.01	5	0.23	0.00	21	0.24	0.00	14	0.26	-2.49	-4.30	-3.90	3.56

Appendix 14 iTRAQ proteomics expression ratios across nine replicates from all pectins combined. R mean, average R values; Z1, z-score for P1; Z2, z-score for P2; Z3, z-score for P3; abs Z, average z-score. Positive and negative z-scores indicate the direction of protein expressions (positive, up-regulation; negative, down-regulation). Red, up-regulated (R mean > 1.2); yellow, not differentiated; green; down-regulated (R mean < 0.83). R mean values are statistically significant if absZ is > 1.65.

			Apple			Citrus			Kiwifruit			
Locus tag (B9O19)	Gene product description	R mean	z1	z2	z3	z1	z2	z3	z1	z2	z3	abs z
1475	peptidase C1A papain	2.06	1.66	3.53	3.07	0.85	2.85	2.33	1.34	3.79	3.62	2.56
682	hypothetical protein	1.76	2.27	-0.69	1.08	1.63	-0.76	0.96	3.67	3.89	3.53	1.73
611	S-layer domain-containing protein	1.74	2.38	2.38	3.35	0.99	1.58	3.30	2.34	3.79	4.34	2.72
52	S-layer domain-containing protein	1.52	2.38	1.62	2.96	0.94	0.26	2.18	1.22	1.72	3.04	1.81
1520	transcriptional regulator	1.49	0.06	2.18	2.16	-2.05	2.21	1.56	2.13	2.39	1.62	1.36
1947	elongation factor G	1.47	4.11	2.47	3.11	1.81	1.80	4.38	3.66	1.74	3.01	2.90
1044	glucose-1-phosphate adenylyltransferase	1.36	2.49	2.04	1.38	0.53	1.43	2.89	1.48	0.67	1.54	1.61
916	N-acetylmuramoyl-L-alanine amidase	1.35	1.69	0.10	1.45	1.47	1.41	2.54	-1.88	0.91	1.96	1.07
1680	Transketolase	1.35	1.70	1.53	2.20	1.27	1.66	2.78	-1.16	0.77	2.46	1.47
1045	glucose-1-phosphate adenylyltransferase	1.26	1.15	0.56	1.63	0.86	0.00	0.93	1.58	-0.34	0.49	0.76
2227	Periplasmic sugar-binding proteins	1.26	-0.17	0.38	1.73	-0.80	1.10	1.93	-2.29	2.47	2.30	0.74
870	extracellular solute-binding protein family 1	1.25	2.23	3.80	3.77	-0.02	2.14	3.69	0.86	0.77	2.12	2.15
826	Pectinesterase (CE8) / S-layer domain-containing protein	1.23	3.80	3.79	4.54	-0.08	0.38	4.07	2.39	-1.08	1.40	2.14
1554	hypothetical protein	1.22	1.75	-1.56	0.14	2.45	-0.85	0.70	1.48	-0.30	0.97	0.53
1156	S-layer domain-containing protein	1.20	6.90	7.69	2.63	-5.14	-9.08	-2.56	5.28	6.65	5.29	1.96
1425	2-dehydro-3-deoxy phosphogluconate aldolase	1.18	1.67	2.69	3.89	-3.05	2.44	2.35	0.24	1.07	3.15	1.60
1597	S-layer domain-containing protein	1.18	6.21	3.14	5.46	-0.50	-1.75	3.15	3.49	-2.95	2.24	2.06
2005	2-deoxy-D-gluconate 3-dehydrogenase	1.17	-0.98	3.83	5.48	-2.03	3.32	3.95	-2.05	1.27	1.61	1.60
1577	Dockerin-like domain	1.14	2.84	0.92	1.78	-1.45	-2.23	1.97	2.33	-1.31	-0.96	0.43
602	Heat shock protein 60 family chaperone GroEL	1.13	-2.19	3.98	2.89	-2.67	3.07	2.22	-2.30	4.69	4.29	1.55
1046	starch synthase	1.13	0.07	1.69	1.21	0.27	0.27	0.79	0.35	0.98	1.32	0.77
2157	Periplasmic [Fe] hydrogenase large subunit	1.13	0.26	1.22	3.09	-0.31	-0.12	1.37	0.66	1.38	1.01	0.95

2204	IMP cyclohydrolase	1.12	1.82	-0.01	1.17	0.68	-0.19	0.38	0.75	-1.30	-0.18	0.35
1225	S-layer domain-containing protein	1.11	1.71	-0.89	2.19	1.38	-2.47	2.27	0.58	-1.24	1.71	0.58
751	transcriptional regulator, CdaR	1.11	-0.54	1.53	3.47	-2.89	2.15	3.57	-0.67	0.67	1.84	1.01
876	Uronate isomerase	1.11	0.39	0.61	2.34	-0.25	0.96	1.62	0.20	0.06	1.21	0.79
1570	dihydroxy-acid dehydratase	1.11	0.26	1.48	2.87	-0.04	1.23	2.13	-0.37	-0.02	0.30	0.87
2084	ATP synthase beta chain	1.10	-1.28	2.67	1.39	-0.85	1.11	1.30	-0.82	1.90	1.79	0.80
2203	IMP cyclohydrolase	1.09	-0.52	3.21	3.92	-0.82	2.17	1.44	0.07	1.13	1.25	1.32
1005	transcriptional regulator	1.09	-2.37	2.42	3.45	-1.12	1.74	2.04	1.29	-0.65	0.51	0.81
898	4-deoxy-L-threo-5-hexosulose-uronateketol-isomerase	1.06	-0.47	0.15	3.21	-0.30	-0.05	2.03	-0.97	0.49	1.47	0.62
1089	Multiple sugar ABC transporter, ATP-binding protein	1.06	-0.08	2.38	3.26	-1.71	1.77	2.93	-0.14	0.70	1.94	1.23
1888	oxaloacetate decarboxylase alpha chain	1.05	-0.22	2.61	1.76	-0.41	2.02	1.48	0.17	-1.26	0.08	0.69
273	phosphoglucosamine mutase	1.04	0.05	1.31	0.93	0.53	0.68	-0.28	0.55	-0.73	-2.51	0.06
1928	band 7 protein	1.04	-0.75	2.80	4.42	-2.49	1.57	2.93	-0.34	-0.41	2.56	1.14
2161	L-arabinose isomerase	1.04	-0.43	3.20	4.86	-1.90	1.85	3.44	0.20	-1.37	0.59	1.16
1678	argininosuccinate synthase	1.04	-0.61	1.46	2.79	0.08	0.78	3.13	-1.19	0.30	0.08	0.76
827	carbamoyl-phosphate synthase large chain	1.02	1.11	1.17	2.52	-1.75	0.01	2.17	1.01	-1.98	-0.72	0.39
1446	DNA-directed RNA polymerase, beta subunit	1.02	0.96	-1.14	1.88	0.06	-1.75	1.23	0.30	-0.33	-0.59	0.07
2196	phosphoribosylformylglycinamidine synthase	1.02	1.05	-0.44	2.08	0.35	-0.95	1.32	0.69	-0.91	-0.49	0.30
1627	Pectinesterase (CE8) / S-layer domain-containing protein	1.02	6.30	5.84	4.51	0.30	-5.26	-3.76	5.13	4.31	5.18	2.50
692	Transketolase, C-terminal domain	1.01	-0.65	1.22	0.48	-1.06	-0.05	0.78	0.93	-0.62	-0.11	0.10
198	phosphoribosylamine--glycine ligase	1.01	-0.21	-1.87	0.37	0.70	-0.93	-0.10	0.76	-0.35	-0.84	0.27
1522	malate dehydrogenase	1.00	-2.40	-0.30	1.51	-0.41	-1.85	1.07	-0.40	-0.33	1.52	0.18
1898	fructose-bisphosphate aldolase	0.99	-1.17	1.27	2.38	-2.04	1.43	2.82	-0.39	-2.19	1.30	0.38
972	Thioredoxin	0.98	1.09	-0.44	1.70	0.23	-1.60	-0.90	-0.12	-0.83	-1.73	0.29
1287	Rubrerythrin	0.98	-1.09	0.86	2.52	-1.53	1.18	1.48	-0.94	1.13	1.15	0.53
608	NIF system FeS cluster assembly, NifU, N-terminal	0.98	1.20	1.71	1.94	-1.47	-0.86	0.29	-0.56	-2.70	-1.59	0.23
914	S-layer domain-containing protein	0.97	4.25	-5.21	0.68	1.15	-6.78	-1.32	2.65	-6.77	-5.26	1.84
607	hypothetical protein	0.97	0.27	-1.53	1.52	-1.11	-1.11	1.25	-0.60	-2.19	0.31	0.35
743	Propeptide PepSY amd peptidase M4	0.96	0.14	-1.37	1.07	-0.05	-1.46	0.17	-0.45	-0.22	-0.66	0.31
1426	2-dehydro-3-deoxy gluconokinase KdgK	0.96	0.27	0.30	1.25	-0.55	-0.24	0.08	-0.66	-3.09	-0.79	0.38

1346	aspartate transaminase	0.96	-1.21	0.21	0.12	-0.32	0.00	1.40	0.12	-0.52	-0.81	0.11
909	Pectate lyase (PL1) / S-layer domain-containing protein	0.94	3.34	-5.50	-1.12	0.43	-8.08	-1.75	1.94	-4.74	-3.10	2.06
1303	Cell division protein FtsZ	0.94	-0.09	-0.01	2.67	-0.69	-1.78	0.34	-0.37	-1.48	-0.61	0.22
901	LL-diaminopimelate aminotransferase	0.93	-0.17	-0.77	0.96	-0.35	-1.58	1.21	-1.19	-2.27	-1.04	0.58
1271	Translation elongation factor Tu	0.93	-1.19	0.46	3.07	-2.83	-0.47	3.14	-0.31	-2.97	0.48	0.07
908	S-layer domain-containing protein	0.93	1.18	-1.95	2.11	-1.39	-4.60	0.73	1.06	-3.73	-1.02	0.85
1972	orotate phosphoribosyltransferase	0.92	-1.40	-0.84	0.99	-0.98	-0.36	1.44	-0.72	-1.41	-0.74	0.45
1090	N-acetyl-gamma-glutamyl-phosphatereductase	0.92		-1.76	0.26		-1.84	-0.16		-0.82	0.39	0.66
1445	DNA-directed RNA polymerase, beta subunit	0.92	-0.58	-0.32	1.99	-2.15	-1.86	0.61	-0.37	-1.95	-0.40	0.56
1969	orotidine 5'-phosphate decarboxylase	0.91	-1.16	-1.59	0.91	0.21	-1.75	0.24	-0.81	-1.61	0.41	0.57
1243	DNA gyrase subunit A	0.90		-0.21	-1.20		0.87	-0.42		0.32	-1.49	0.35
949	homoserine dehydrogenase	0.90		-1.86	1.03		-2.68	0.48		-2.54	-0.07	0.94
1330	pyruvate synthase	0.89	-1.34	-3.32	-1.13	-2.53	-4.49	-0.06	1.62	-4.87	-1.64	1.97
2045	glycine hydroxymethyltransferase	0.89	-0.83	-1.95	0.72	-0.55	-3.31	-0.20	-0.24	-3.48	-1.28	1.23
366	amino acid-binding ACT domain protein	0.89	-0.92	0.25	1.24	-0.37	-0.92	-0.04	-1.38	-1.02	-1.18	0.48
2111	D-3-phosphoglycerate dehydrogenase	0.89	-0.94	-1.66	-1.15	0.18	-3.06	-1.33	-0.25	-2.55	-2.24	1.45
1372	ribosome recycling factor	0.89	1.29	-0.71	0.70	-0.73	-1.61	-0.84	-1.64	-1.86	-1.10	0.72
1337	S-layer domain-containing protein	0.88	0.45	-1.75	-1.34	0.12	-2.63	-0.92	-0.08	-0.64	-0.67	0.83
368	2-oxoacid ferredoxin oxidoreductase	0.88	-0.93	-1.97	-0.40	-0.61	-2.31	-0.43	0.10	-2.16	-0.45	1.02
287	3-dehydroquinate synthase	0.87		-0.49	-1.05		-1.34	-0.24		-0.50	-1.95	0.93
506	O-acetylhomoserine sulfhydrylase	0.87	-0.72	-1.51	-1.37	-0.38	-1.82	0.08	-1.15	-0.44	-1.59	0.99
617	30S ribosomal protein S2	0.86	1.48	-3.40	-2.17	-0.35	-3.69	-0.52	1.31	-2.24	-2.27	1.32
1178	ornithine carbamoyltransferase	0.85	0.99	-3.27	-2.67	0.03	-2.62	-1.03	-1.11	-2.77	-2.50	1.66
1219	DNA-directed RNA polymerase subunit alpha	0.85	-0.06	-1.37	-0.73	-0.86	-1.54	-1.26	-1.02	-2.58	-1.50	1.21
2018	translation elongation factor P	0.85	1.45	-0.74	-0.13	-0.92	-2.28	-1.42	0.83	-2.83	-2.64	0.97
389	peptidase M18 aminopeptidase I	0.85		-1.72	-2.13		-1.74	-1.13		-1.36	0.80	1.21
414	SCP-2 sterol transfer family	0.85		-0.69	-0.66		-0.66	-1.35		-1.33	-1.75	1.07
1513	Chaperone protein DnaK	0.84	-1.10	-3.74	0.10	-1.81	-4.68	0.36	-0.65	-1.47	3.82	1.02
1578	S-layer domain-containing protein	0.84	3.31	-3.19	0.15	-2.72	-4.58	-1.76	0.58	-4.86	-2.54	1.74
724	tetrahydrofolate synthase	0.84	1.02	-1.82	-1.10	-0.76	-1.55	-1.59	-0.04	-1.75	-2.55	1.12

912	S-layer domain-containing protein	0.83	0.77	-4.32	-1.09	-1.86	-5.43	-0.95	1.32	-4.41	-1.48	1.94
727	NADP-specific glutamate dehydrogenase	0.83	-1.68	-3.61	-1.27	-0.89	-3.73	-1.12	-0.24	-2.69	-1.19	1.83
1858	malonyl CoA-acyl carrier protein transacylase	0.83	-0.60	-1.92	-1.14	-1.44	-2.77	-0.12	0.02	-2.52	-0.64	1.24
947	aspartatekinase	0.82	2.26	-2.56	-1.80	0.03	-2.27	-1.98	0.14	-2.81	-2.24	1.25
403	Transcriptional regulator, GntR family	0.82	-1.08	-2.45	0.52	-2.54	-2.79	0.37	-1.78	-4.51	-0.99	1.69
1273	2-isopropylmalate synthase	0.81	-2.35	0.45	0.74	-1.33	-1.47	-0.85	-1.47	0.02	0.55	0.63
2082	ATP synthase alpha chain	0.81	0.65	-1.95	-1.11	-0.24	-1.80	-0.82	-1.49	-0.89	-0.09	0.86
129	histidine--tRNA ligase	0.80	-1.74	-2.94	-3.07	0.72	-2.05	-0.75	-0.15	-2.33	0.69	1.29
89	phosphate acetyltransferase	0.80	-0.18	-3.03	-1.92	-1.75	-4.03	-1.97	-0.98	-3.69	-1.06	2.07
919	glyceraldehyde-3-phosphate dehydrogenase	0.80	-1.57	-0.17	1.09	-2.49	-1.66	-0.29	-3.01	-2.07	-2.48	1.41
1010	6-phosphofructokinase	0.79	0.07	-4.23	-2.66	0.28	-4.50	-3.26	-0.93	-2.20	-3.16	2.29
1595	S-layer domain-containing protein	0.77	-0.80	-3.29	-3.74	1.03	-2.43	-2.96	-0.44	-2.49	-1.96	1.90
1265	pyruvate phosphate dikinase	0.77	-2.51	-3.91	-3.74	-1.36	-3.84	-1.97	-1.88	-3.49	-3.48	2.91
1155	CTPsynthase	0.77	-0.72	-2.42	-3.09	-0.76	-1.98	-1.19	0.17	-2.04	-1.00	1.45
277	proline--tRNA ligase	0.77	-1.53	-2.68	-1.83	0.04	-2.50	-1.17	-0.30	-3.23	-1.69	1.65
678	adenylosuccinate synthase	0.77	-1.34	-1.25	-1.63	-1.24	-0.77	-0.69	-1.00	-0.66	-0.39	1.00
2112	phosphoserine transaminase	0.77	-1.96	-0.91	-0.17	-2.82	-1.33	-0.12	-2.27	-1.77	-1.83	1.46
1821	acetolactate synthase	0.76	-1.13	-3.32	-1.45	-1.82	-4.18	-0.08	-0.57	-3.63	-1.02	1.91
1892	acetyl-CoA carboxylase	0.75	-0.52	-2.24	-2.76	-1.68	-4.46	-2.11	-1.90	-3.56	-4.00	2.58
775	aspartate-semialdehyde dehydrogenase	0.75	-0.41	-3.92	-2.68	-1.05	-4.80	-2.68	-1.29	-3.20	-1.93	2.44
616	translation elongation factor Ts	0.74	0.27	-4.66	-2.99	-0.13	-4.79	-2.70	0.42	-3.71	-3.85	2.46
1826	3-isopropylmalate dehydratase small subunit	0.74	-1.86	-1.90	-0.28	-1.04	-2.53	-0.18	-1.87	-2.15	-0.45	1.36
1953	ribose-phosphate diphosphokinase	0.74	-2.27	-2.45	-1.15	-0.10	-2.87	-0.42	-1.11	-1.65	-1.56	1.51
1066	Arginine-binding extracellular protein ArtP	0.73	-0.75	-1.66	-0.50	-1.25	-2.56	-0.09	-2.52	-1.78	-1.21	1.37
1269	30Sribosomal protein S7	0.73	1.12	-3.16	-2.95	-0.65	-3.82	-3.29	0.39	-3.18	-2.74	2.03
1195	30S ribosomal protein S3	0.72	0.88	-3.57	-1.65	0.40	-3.16	-1.17	-0.57	-3.10	-2.29	1.58
1051	50S ribosomal protein L1	0.72	1.80	-4.69	-1.90	0.04	-4.68	-3.14	0.89	-4.37	-2.49	2.06
1192	50S ribosomal protein L2	0.72	0.13	-2.84	-2.78	1.40	-3.03	-1.94	-0.69	-4.10	-2.72	1.84
1607	Phosphotransferase system, phosphocarrier protein HPr	0.72	-1.17		-1.44	-1.24		-2.03	0.24		-0.88	1.08
950	hypothetical protein	0.72		-1.83	-1.02		-2.02	-1.96		-1.81	-2.29	1.82

1903	aconitate hydratase	0.72	-1.87	-3.56	-2.50	-1.42	-3.70	-3.58	-1.98	-3.86	-2.62	2.79
128	aspartate--tRNA ligase	0.71	0.20	-2.35	-1.48	0.30	-3.66	-2.58	-0.34	-2.43	-2.74	1.68
1827	3-isopropylmalate dehydrogenase	0.71	-0.65	-2.19	-1.92	0.39	-2.40	-2.83	-1.90	-1.95	-1.85	1.70
1626	Pectinesterase (CE8) / S-layer domain-containing protein	0.70	1.84	-6.39	-4.06	-2.12	-6.84	-4.59	-0.11	-5.66	-3.89	3.53
90	acetate kinase	0.69	-1.60	-3.56	-2.88	-1.64	-3.67	-2.46	-2.99	-3.25	-1.82	2.65
1823	Ketol-acid reductoisomerase	0.69	-2.11	-2.97	-2.76	-1.79	-3.58	-2.21	-2.66	-3.47	-2.88	2.71
540	GMP synthase [glutamine-hydrolyzing]	0.68	-1.77	-4.29	-3.21	-0.74	-3.06	-1.90	-1.70	-3.44	-2.67	2.53
1189	50S ribosomal protein L3	0.68	1.63	-3.43	-2.23	0.04	-3.25	-2.63	0.69	-2.63	-2.44	1.58
326	Rubrerythrin	0.67	-2.20	-3.32	-0.05	-1.86	-2.83	-1.00	-2.51	-0.46	0.50	1.53
1608	phosphoenolpyruvate-protein phosphotransferase	0.67	-2.36	-2.29	-2.00	-0.85	-2.21	-1.46	-0.59	-3.18	-0.61	1.73
289	3-deoxy-7-phosphoheptulonate synthase	0.67	-2.98	-2.70	-1.30	-2.27	-3.54	-0.17	-1.30	-2.44	-1.22	1.99
1386	30S ribosomal protein S16	0.67	-0.09	-1.97	-2.57	0.23	-2.63	-2.32	0.64	-1.09	-1.43	1.25
1206	30Sribosomal protein S5	0.66	2.48	-3.60	-3.52	-0.39	-3.36	-2.36	1.86	-2.22	-2.44	1.51
92	histidinol dehydrogenase	0.66		-2.61	-2.31		-4.31	-1.86		-2.70	-2.20	2.66
1270	Translation elongation factor G	0.66	-0.58	-7.52	-6.52	-0.93	-6.66	-5.44	-2.43	-6.62	-5.72	4.71
1201	50S ribosomal protein L5	0.65	-0.53	-4.12	-4.82	-0.76	-3.95	-3.37	1.18	-3.83	-3.17	2.60
542	phosphoribosylaminoimidazolesuccinocarboxamide synthase	0.65	0.24	-4.08	-2.08	-2.24	-4.26	-2.23	-2.37	-4.13	-1.85	2.56
59	branched-chain-amino-acid transaminase	0.64	-2.06	-2.53	-2.35	-1.00	-2.87	-3.17	-1.58	-2.05	-2.27	2.21
1190	50S ribosomal protein L4	0.64	0.35	-4.71	-2.69	0.60	-4.63	-1.97	0.10	-2.96	-2.43	2.04
541	amidophosphoribosyltransferase	0.64	-2.61	-3.40	-2.53	-1.89	-3.85	-3.04	-2.18	-3.60	-3.75	2.98
627	aspartate--tRNA ligase	0.64	-1.60	-4.49	-3.52	-0.78	-4.31	-3.32	-0.73	-3.46	-2.31	2.72
1091	acetylglutamate kinase	0.63	-0.38	-2.23	-2.00	-0.49	-2.51	-2.83	-0.80	-3.01	-0.62	1.65
357	Cell division trigger factor	0.62		-5.32	-2.59		-5.71	-2.14		-3.36	-1.86	3.50
2138	polyribonucleotide nucleotidyltransferase	0.62	-2.82	-2.72	-1.91	-2.74	-2.63	-0.99	-2.40	-1.93	-2.10	2.25
1606	PTS system fructose-specific EIIABC component	0.62		-2.95	-2.02		-3.75	-1.43		-2.21	-0.21	2.09
127	tyrosine--tRNA ligase	0.61		-2.25	-2.44		-1.35	-2.43		-2.04	-2.41	2.15
1097	Phosphoenolpyruvate carboxykinase [GTP]	0.61	-2.01	-5.19	-3.11	-2.23	-5.15	-3.87	-3.20	-5.29	-3.36	3.71
1203	30Sribosomal protein S8	0.61	-0.05	-3.30	-2.96	0.01	-3.60	-2.90	1.10	-3.33	-2.65	1.96
1218	30Sribosomal protein S4	0.61	1.30	-4.64	-2.11	-0.88	-4.25	-3.15	-0.02	-3.46	-3.38	2.29
1105	Glutamine synthetase	0.60	-1.51	-4.26	-3.06	-3.02	-4.47	-1.82	-4.09	-4.30	-2.44	3.22

1902	isocitrate dehydrogenase	0.59	-2.28	-3.85	-4.41	-1.92	-5.14	-4.40	-2.65	-6.53	-4.21	3.93
1379	50S ribosomal protein L27	0.59		-2.43	-2.00		-2.73	-2.47		-1.47	-1.76	2.14
1050	50S ribosomal protein L10	0.59	-0.86	-5.24	-2.80	-1.11	-4.25	-2.65	-2.99	-4.43	-4.13	3.16
1941	triose-phosphate isomerase	0.58	-3.10	-4.18	-2.28	-1.38	-3.54	-2.95	-2.25	-3.04	-2.74	2.83
356	ATP-dependent Clp endopeptidase, proteolytic subunit ClpP	0.56		-2.71	-2.35		-2.55	-1.79		-2.44	-2.34	2.36
137	pyruvate formate-lyase	0.55	-2.90	-6.45	-4.84	-2.01	-6.32	-4.50	-3.96	-4.80	-2.48	4.25
1052	50S ribosomal protein L11	0.53	-0.87	-4.04	-2.88	-0.96	-3.35	-2.54	-0.41	-3.24	-3.31	2.40
1151	tryptophan synthase alpha chain	0.51	1.05	-2.92	-1.01	-1.98	-2.36	-1.16	0.90	-2.72	-0.85	1.23
1150	tryptophan synthase beta chain	0.51	1.40	-3.10	-2.06	0.92	-3.91	-2.71	0.96	-2.23	-3.02	1.53
1049	50S ribosomal protein L7/L12	0.47	-1.74	-3.42	-3.61	-1.52	-3.46	-3.03	-2.72	-3.10	-3.06	2.85
1942	phosphoglycerate kinase	0.41	-6.22	-7.07	-5.89	-4.14	-7.04	-5.19	-6.07	-5.28	-4.62	5.72
2009	Xylulose-5-phosphate phosphoketolase	0.41	-2.35	-5.99	-4.64	-2.99	-5.68	-3.45	-4.18	-4.40	-3.64	4.15
349	translation initiation factor IF-3	0.40	-1.65	-3.25	-2.68	-2.11	-2.78	-3.00	-1.50	-2.16	-3.01	2.46
1205	50S ribosomal protein L18	0.39		-2.66	-1.98		-2.28	-2.01		-2.00	-2.26	2.20
1940	phosphoglycerate mutase	0.39	-3.45	-4.71	-4.34	-3.22	-4.27	-3.88	-3.25	-4.42	-3.75	3.92
2145	acetaldehyde dehydrogenase	0.29	-7.35	-9.86	-8.68	-6.23	-9.70	-7.84	-7.83	-8.82	-6.76	8.12
2146	Aminoglycoside phosphotransferase	0.24	-2.27	-5.20	-4.04	-2.68	-5.07	-3.89	-2.49	-4.30	-3.90	3.76

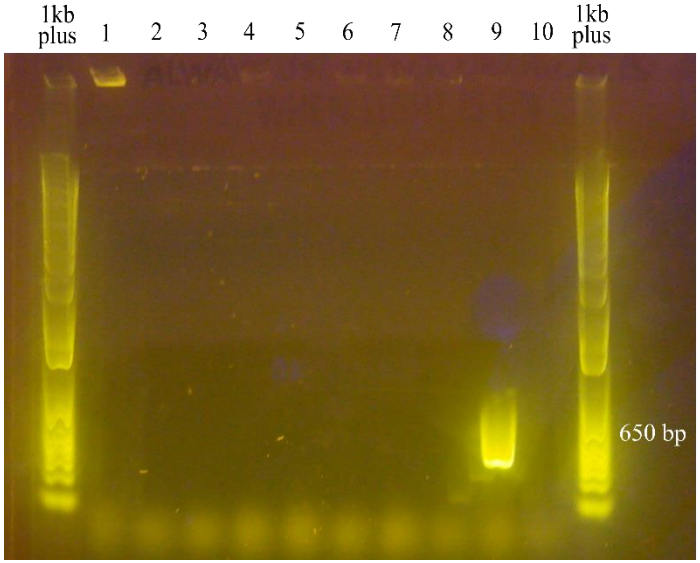
Appendix 15 iTRAQ expression ratios in *M. pectinilyticus* extraproteomes. A, apple; C, citrus; K, kiwifruit; F, fructose. Red, up-regulated; yellow, not differentiated; green; down-regulated. These results have not been replicated, hence should be used as a preliminary data only.

Locus tag (B9O19)	Gene product name	# of peptides (95%)	A:F	*p*-value (A:F)	C:F	*p*-value (C:F)	K:F	*p*-value (K:F)
52	S-layer domain-containing protein	41	0.98	0.803	0.88	0.085	0.58	0.000
87	hypothetical protein	17	0.66	0.000	0.62	0.000	0.50	0.000
90	acetate kinase	4	0.74	0.257	0.67	0.112	0.80	0.062
97	beta-lactamase domain protein	8	0.82	0.322	0.71	0.084	0.65	0.020
137	pyruvate formate-lyase	13	0.67	0.082	0.70	0.072	0.74	0.005
254	copper amine oxidase-like domain-containing protein	13	0.70	0.007	0.67	0.002	0.58	0.001
292	copper amine oxidase-like domain-containing protein	8	0.52	0.004	0.50	0.003	0.55	0.006
293	copper amine oxidase-like domain-containing protein	5	0.71	0.188	0.59	0.177	0.65	0.009
305	S-layer domain-containing protein	6	0.72	0.150	0.69	0.046	0.94	0.813
312	PL1	25	0.60	0.000	0.61	0.000	0.58	0.000
380	copper amine oxidase-like domain-containing protein	7	0.56	0.001	0.64	0.102	0.46	0.008
398	GH3	50	1.08	0.216	0.95	0.399	0.86	0.020
505	CE8	4	1.29	0.024	1.15	0.223	1.10	0.282
543	S-layer domain-containing protein	12	1.29	0.053	1.19	0.024	1.09	0.474
602	Heat shock protein 60 family chaperone GroEL	16	0.74	0.075	0.74	0.063	0.78	0.058
608	NIF system FeS cluster assembly, NifU, N-terminal	9	0.55	0.007	0.60	0.042	0.69	0.058
609	GH5	42	0.97	0.591	0.96	0.684	0.78	0.000
611	S-layer domain-containing protein	38	0.82	0.020	0.79	0.003	0.76	0.007
682	hypothetical protein	25	0.90	0.107	0.73	0.002	0.27	0.000
683	Cell surface protein	3	0.54	0.010	0.64	0.010	1.12	0.574
743	S-layer domain-containing protein	55	0.75	0.003	0.67	0.000	0.51	0.000
758	copper amine oxidase domain protein	8	1.00	0.990	0.87	0.014	0.79	0.102
775	aspartate-semialdehyde dehydrogenase	14	0.56	0.010	0.58	0.010	0.57	0.000
826	CE8 / S-layer domain-containing protein	29	0.61	0.000	0.67	0.000	0.59	0.000

870	extracellular solute-binding protein family 1	10	0.82	0.167	0.85	0.282	1.43	0.046
873	PL1 / S-layer domain-containing protein	23	0.72	0.006	0.75	0.014	0.69	0.000
874	PL1	17	0.98	0.893	1.00	0.985	0.79	0.013
876	Uronate isomerase	7	0.59	0.021	0.72	0.103	0.83	0.385
908	S-layer domain-containing protein	87	0.58	0.000	0.56	0.000	0.41	0.000
909	PL1 / S-layer domain-containing protein	127	0.62	0.000	0.58	0.000	0.38	0.000
912	S-layer domain-containing protein	20	0.49	0.000	0.57	0.000	0.45	0.000
914	S-layer domain-containing protein	168	0.59	0.000	0.56	0.000	0.39	0.000
916	N-acetylmuramoyl-L-alanine amidase	12	0.68	0.116	0.64	0.065	0.67	0.104
973	CE4	5	1.16	0.336	1.38	0.031	1.44	0.014
1097	Phosphoenolpyruvate carboxykinase [GTP]	12	0.77	0.017	1.01	0.965	0.90	0.240
1105	Glutamine synthetase	12	0.66	0.002	0.80	0.067	0.63	0.000
1156	S-layer domain-containing protein	719	0.47	0.000	0.37	0.000	0.28	0.000
1218	30Sribosomal protein S4	3	0.73	0.253	0.43	0.009	1.73	0.101
1225	S-layer domain-containing protein	35	0.72	0.029	0.63	0.000	0.56	0.000
1238	S-layer domain-containing protein	7	0.65	0.068	0.63	0.036	0.46	0.003
1241	PL1	8	0.93	0.685	0.86	0.343	0.60	0.001
1271	Translation elongation factor Tu	6	1.41	0.268	1.05	0.788	3.60	0.016
1299	GH95 / CE8	33	1.13	0.075	1.15	0.058	1.06	0.503
1315	PL1	11	0.96	0.770	0.87	0.400	0.68	0.031
1344	Cell wall-binding protein	23	0.89	0.411	1.16	0.375	0.73	0.043
1415	copper amine oxidase-like domain-containing protein	5	0.71	0.097	0.70	0.256	0.58	0.014
1425	2-dehydro-3-deoxy phosphogluconate aldolase	12	0.61	0.054	0.56	0.004	0.64	0.069
1426	2-dehydro-3-deoxy gluconokinase KdgK	7	0.57	0.069	0.54	0.099	0.80	0.226
1528	CE4	17	1.20	0.261	1.35	0.003	1.34	0.005
1545	S-layer domain-containing protein	9	0.89	0.465	0.81	0.139	0.72	0.064
1554	hypothetical protein	9	1.24	0.101	1.38	0.058	2.04	0.000
1568	hypothetical protein	12	0.57	0.004	0.52	0.002	0.50	0.000
1578	S-layer domain-containing protein	51	0.51	0.000	0.47	0.000	0.39	0.000
1595	S-layer domain-containing protein	17	0.56	0.000	0.55	0.001	0.66	0.001

1597	S-layer domain-containing protein	70	0.64	0.000	0.68	0.000	0.53	0.000
1615	PL9	4	1.09	0.709	1.48	0.255	0.63	0.025
1621	copper amine oxidase-like domain-containing protein	11	0.62	0.003	0.63	0.069	0.48	0.006
1626	CE8 / S-layer domain-containing protein	35	0.42	0.000	0.42	0.000	0.34	0.000
1627	CE8 / S-layer domain-containing protein	92	0.69	0.000	0.70	0.000	0.56	0.000
1653	PL11	3	1.21	0.641	1.54	0.393	1.46	0.057
1654	hypothetical protein	34	1.19	0.214	0.97	0.830	0.49	0.000
1681	GH95 / CE8	23	1.25	0.022	1.12	0.397	1.08	0.430
1828	GH3	29	2.45	0.000	0.97	0.732	1.03	0.693
1942	phosphoglycerate kinase	27	0.32	0.000	0.28	0.000	0.40	0.000
1989	copper amine oxidase	19	0.63	0.021	0.56	0.000	0.41	0.000
2002	hypothetical protein	3	0.98	0.915	0.88	0.474	0.66	0.047
2009	Xylulose-5-phosphate phosphoketolase	19	0.63	0.004	0.57	0.003	0.68	0.024
2112	phosphoserine transaminase	4	0.60	0.113	0.54	0.063	0.91	0.675
2145	acetaldehyde dehydrogenase	21	0.78	0.037	0.79	0.031	0.99	0.956
2160	PL1	33	0.86	0.048	0.85	0.024	0.67	0.000
2161	L-arabinose isomerase	15	0.71	0.003	0.71	0.003	0.81	0.030
2167	S-layer domain-containing protein	7	0.67	0.047	0.78	0.200	0.74	0.098
2190	NLP/P60 protein	5	1.64	0.046	1.24	0.069	1.46	0.094
2231	CE12	16	0.93	0.526	1.00	0.970	0.71	0.000
2263	CE8 / CE12	21	1.09	0.370	1.09	0.353	0.84	0.019

Appendix 16 Testing the target specificity of *M. pectinilyticus*-specific qPCR primers. Electrophoresis was carried out at 80V for 50 min at ambient temperature using 1 % agarose gel. Bands were visualized using E-Gel® Safe Imager™ (Invitrogen™).



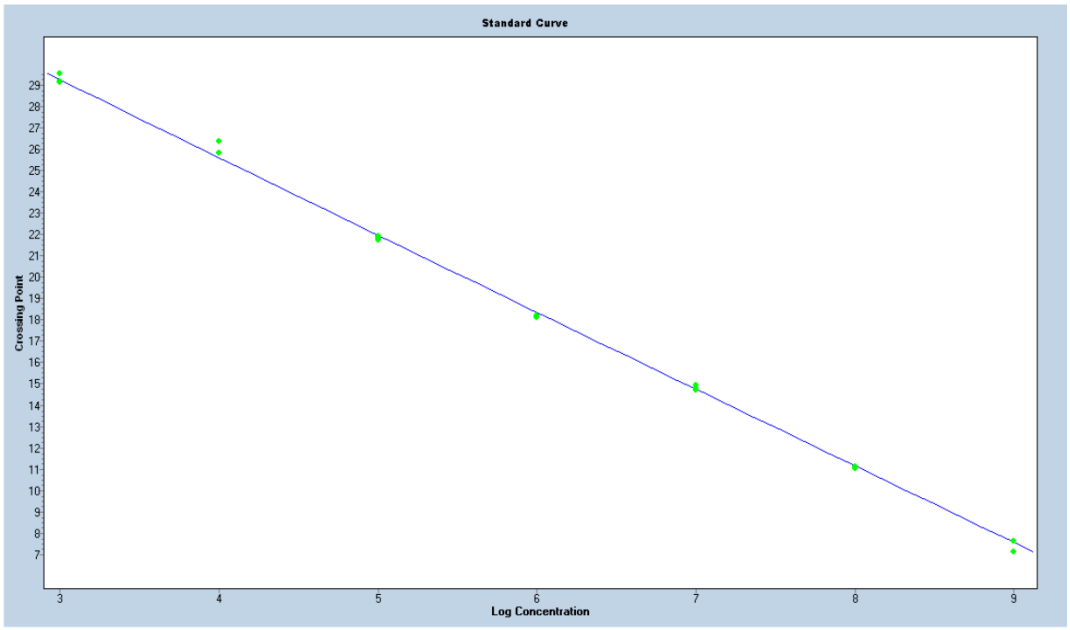
1. *Eubacterium rectale* ATCC 35183
2. *Faecalibacterium prausnitzii* DSM 17677
3. *Roseburia intestinalis* DSM 14610
4. *Lactobacillus plantarum* ATCC 14917
5. *Lactobacillus acidophilus* ATCC 11975
6. *Staphylococcus aureus* ATCC 25923
7. *Ruminococcus gnavus* ATCC 29149
8. *Lachnospira multipara* ATCC 19207
9. *Monoglobus pectinilyticus* DSM 104782
10. Negative control

Appendix 17 Enumeration of cell numbers of *M. pectinilyticus* (V6-V8) and *E. coli* (V3-V4) by viable cell count and qPCR. Genomic DNA from known number of cells was amplified using M. pectinilyticus-specific or universal bacterial primer sets to construct two separate qPCR standard curves (Appendix 17).

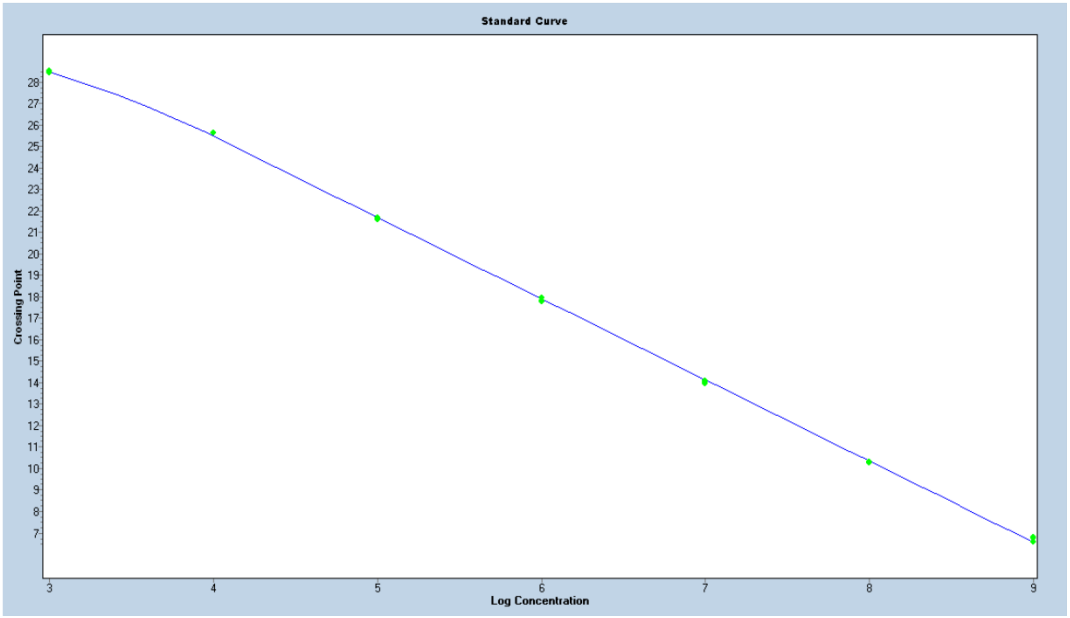
	M. pectinilyticus			***E. coli***		
Number of cells	**Standard DNA concentration (ng/ul)**	**Concentration of amplified 16S rRNA**	**log_{10} ± std dev**	**Standard DNA concentration (ng/ul)**	**Concentration of amplified 16S rRNA**	**log_{10} ± std dev**
1.00E+09	1.73E+01	1.33E+09 9.68E+08 9.76E+08	9.033 ± 0.08	3.50E+01	8.98E+08 9.88E+08 8.69E+08	8.962 ± 0.03
1.00E+08	1.73E+00	1.08E+08 1.01E+08 1.05E+08	8.020 ± 0.01	3.50E+00	1.06E+08 1.03E+08 1.05E+08	8.020 ± 0.01
1.00E+07	1.73E-01	9.77E+06 1.04E+07 8.79E+06	6.984 ± 0.04	3.50E-01	1.06E+07 1.03E+07 1.10E+07	7.021 ± 0.01
1.00E+06	1.73E-02	1.15E+06 1.17E+06 1.09E+06	6.055 ± 0.02	3.50E-02	9.64E+05 1.05E+06 1.08E+06	6.013 ± 0.03
1.00E+05	1.73E-03	1.16E+05 1.10E+05 1.03E+05	5.040 ± 0.03	3.50E-03	9.98E+04 1.06E+05 1.05E+05	5.015 ± 0.01
1.00E+04	1.73E-04	6.22E+03 8.69E+03 8.73E+03	3.891 ± 0.08	3.50E-04	9.14E+03 8.97E+03 9.06E+03	3.957 ± 0.00
1.00E+03	1.73E-05	8.34E+02 1.05E+03 1.08E+03	2.992 ± 0.06	3.50E-05	NA 1.56E+03 1.64E+03	3.204 ± 0.02
Negative control	Water	0.00E+00 0.00E+00 0.00E+00	0.000 ± 0.00	Water	2.48E+01 3.21E+01 2.30E+01	1.421 ± 0.08

Appendix 18 Quantitative PCR standard curves constructed by plotting $\log_{10}$ cell numbers of *M. pectinilyticus* (a) and *E. coli* (b) against the crossing points (CP). CP values correspond to the cycle at which the fluorescence initially achieves a defined threshold with a statistical significance. CP values are inversely proportional to the amount of nucleic acids present in the samples.

a



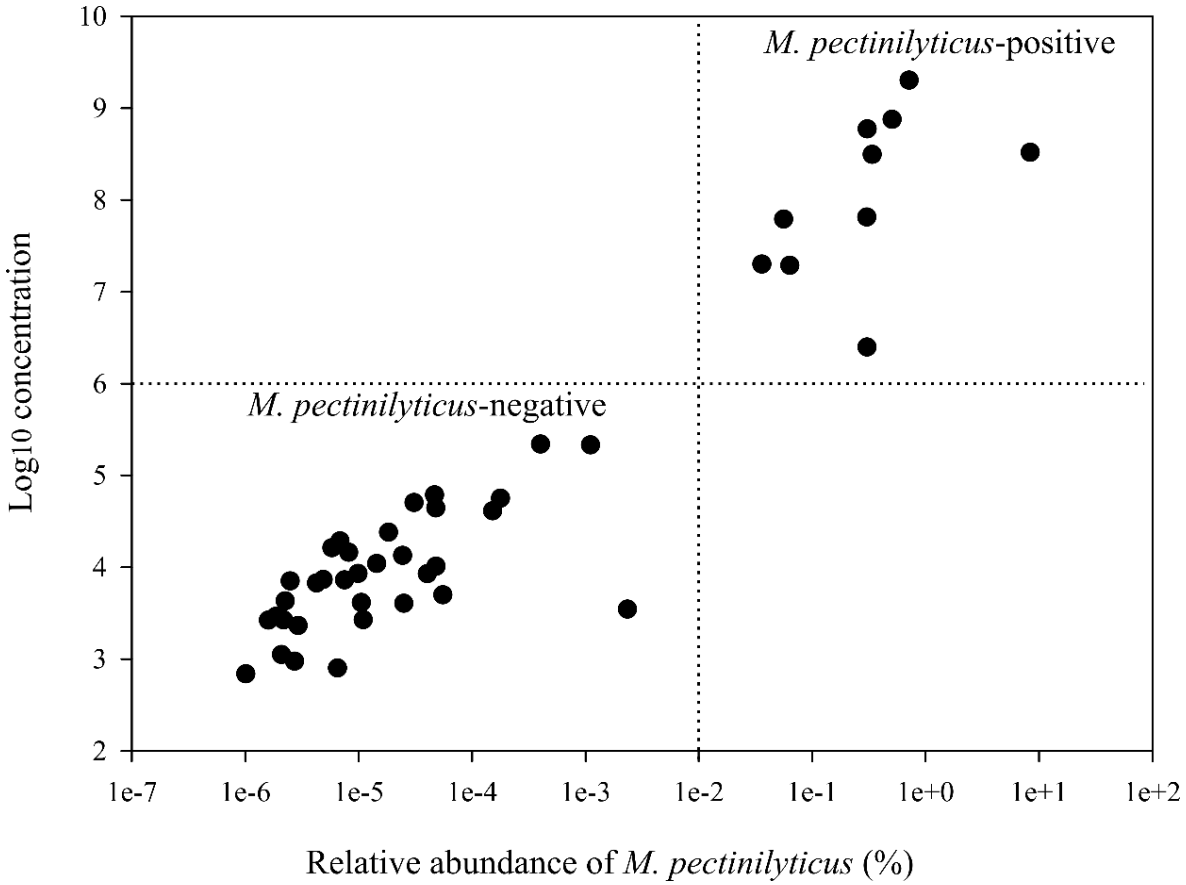
b



Appendix 19 Participant information and the daily food group intake calculated from four sets of 3-day dietary records. Participants were divided into high, moderate, and low fibre groups based on the amount of daily fibre intake recommended by Ministry of Health NZ. Pectin intake (gram) per day was calculated based on the food composition table. Vegetable, fruit, grain, and protein intakes are expressed in servings per day.

Donor ID	Age	BMI	Ethnicity	Gender	Fibre intake (g/day)	Fibre group	Pectin intake (g/day)	Vegetables (serves/day)	Fruit (serves/day)	Grains (serves/day)	Protein (serves/day)
1	26	23.1	NZ European	M	31.43	High	4.83	5.83	1.24	9.72	2.71
2	40	27.2	Indian	M	21.83	Low	2.90	2.64	0.27	4.28	4.88
3	52	20.6	NZ European	F	45.90	High	6.55	6.46	4.15	4.75	1.89
4	52	22.3	Chinese	F	31.10	High	3.69	5.77	1.34	6.80	2.64
5	28	24	NZ European	F	15.88	Low	2.80	5.07	0.49	1.83	4.38
6	59	22.5	NZ European	M	40.05	High	5.08	4.67	1.98	7.85	2.16
7	49	21.4	Caucasian	F	32.40	High	4.15	6.41	1.37	10.08	0.52
8	42	20.7	NZ European	M	21.13	Low	2.49	3.87	0.42	5.93	2.41
9	20	20.2	NZ European	F	29.80	High	3.38	3.58	1.95	5.04	1.62
10	22	28	Indian	M	33.55	High	4.81	4.24	2.32	8.86	4.04
11	43	24.1	Middle Eastern	F	17.65	Low	1.78	1.63	1.83	7.01	1.53
12	59	28.4	NZ European	M	18.38	Low	1.94	2.64	0.41	8.02	1.28
13	37	22	Irish	F	26.20	High	3.18	3.16	1.29	6.42	2.32
14	22	22.6	NZ European	M	41.73	High	6.04	6.97	1.46	7.22	5.51
15	34	20.5	Indian	M	40.43	High	6.30	3.66	4.48	7.21	3.16
16	32	19	European	F	43.68	High	6.60	4.01	5.18	8.12	2.49
17	37	24.6	NZ European	M	39.13	High	4.11	5.57	0.88	7.41	2.54
18	30	21.6	Indian	M	23.80	Moderate	3.55	3.54	1.09	7.49	0.51
19	26	22.5	Indian	F	17.98	Low	2.69	2.94	0.89	5.25	0.45
20	25	23.6	Romanian	F	41.78	High	6.16	6.98	1.60	4.39	4.65
21	33	22.3	Indian	F	14.53	Low	1.77	1.01	1.65	3.98	0.68
22	20	26.3	NZ European	F	22.60	Moderate	2.78	1.42	2.50	4.65	1.93
23	35	22.3	NZ European, Maori	F	27.93	High	5.06	6.71	1.57	6.19	2.53
24	19	18.4	NZ European	F	59.70	High	11.43	12.65	1.58	11.46	6.63
25	20	19.8	NZ European	F	57.80	High	7.93	6.96	3.38	8.51	5.15
26	33	26.5	NZ European	F	17.53	Low	3.16	4.17	1.39	4.48	2.55
27	36	22.5	NZ European	F	14.70	Low	2.03	4.30	0.06	1.84	2.69
28	64	23.6	Pakeha	F	26.15	High	4.07	5.01	1.30	3.82	2.56
29	21	28.2	Sri Lankan	M	17.08	Low	1.57	0.62	0.49	9.38	3.25
30	34	27.3	Maori	M	20.20	Low	2.48	3.91	0.65	6.63	3.96
31	35	27.5	NZ European, Maori	F	33.83	High	5.10	6.76	0.13	8.05	4.10
32	52	24.2	NZ European	M	18.73	Low	2.35	2.66	0.70	3.48	3.74
33	48	27.7	NZ European	M	30.45	High	3.89	4.86	0.62	10.56	2.61
34	53	20.1	Chinese	F	33.73	High	4.78	4.27	4.23	1.60	4.14
35	36	25.1	Maori	F	15.90	Low	2.14	1.79	1.95	3.78	2.07
36	50	22	NZ European	F	13.30	Low	1.88	2.66	1.04	2.62	1.36
37	25	23	NZ European	M	29.03	High	4.93	5.60	2.58	0.00	6.76
38	52	21.3	Chinese	F	29.93	High	3.78	3.46	3.59	4.21	2.96
39	33	21.1	Filipino	F	12.58	Low	1.62	2.14	0.84	5.80	1.91
40	32	22.5	NZ European	F	19.28	Moderate	2.55	3.96	0.54	3.69	2.93
41	34	22.2	Sri Lankan	F	17.88	Low	2.17	1.78	0.85	5.89	1.41
42	45	20.5	British	F	20.88	Moderate	2.65	2.91	0.40	7.12	1.52
43	28	20.2	NZ European	F	22.38	Moderate	2.90	3.55	1.38	2.47	2.80
44	40	22.7	Other European	F	22.98	Moderate	3.13	4.28	1.04	5.81	1.94

Appendix 20 Determining the split point for separating *M. pectinilyticus*-positive and *M. pectinilyticus*-negative groups by plotting the $\log_{10}$ concentration against the % abundance of *M. pectinilyticus* relative to the total bacteria.



Appendix 21 SLH protein sequences of *M. pectinilyticus* identified in the NIH human stool metagenome databases. The metagenome databases are accessible through the JGI genome portal using the JGI genome number as a reference. JGI gene number and protein identity were assigned by the metagenome databases. Only sequences with > 200 bit-score were considered to have found significant matches. Lower *e*-values indicated more significant alignments. Identities are expressed in percentage, and indicate the extent to which two amino acid sequences have the same residue at the same position in an alignment. Positive scores are expressed in percentage, and indicate the extent to which two protein sequences are related. Gaps are indicatives of insertion and deletion of amino acid residues, and scores are penalized as the result.

JGI genome number/database name	B9O19 number	Protein length	Gene number	Protein identity	Bit-score	*e*-value	Identities	Positives	Gaps
7000000038	241	2601	159371	unnamed protein product	1740	0.00E+00	845/845 (100%)	845/845 (100%)	0/845 (0%)
Subject 508703490	1578	1044	159968	unnamed protein product	981	0.00E+00	638/638 (100%)	638/638 (100%)	0/638 (0%)
7000000111	241	2601	173967	unnamed protein product	453	1.00E-144	219/219 (100%)	219/219 (100%)	0/219 (0%)
Subject 765620695	908	927	130244	unnamed protein product	237	2.00E-72	116/116 (100%)	116/116 (100%)	0/116 (0%)
	912	1745	124378	unnamed protein product	218	2.00E-64	110/110 (100%)	110/110 (100%)	0/110 (0%)
	914	955	154706	unnamed protein product	280	4.00E-87	156/156 (100%)	156/156 (100%)	0/156 (0%)
	1224	960	155234	unnamed protein product	254	7.00E-78	157/157 (100%)	157/157 (100%)	0/157 (0%)
	1545	1809	164178	unnamed protein product	300	4.00E-92	151/151 (100%)	151/151 (100%)	0/151 (0%)
	1578	1044	155937	unnamed protein product	251	1.00E-76	121/121 (100%)	121/121 (100%)	0/121 (0%)
	1597	3175	173519	unnamed protein product	391	9.00E-123	186/205 (91%)	197/205 (96%)	0/205 (0%)
7000000158	50	426	213757	unnamed protein product	384	4.00E-132	187/188 (99%)	187/188 (99%)	0/188 (0%)
Subject 764062976	241	2601	233210	unnamed protein product	1172	0.00E+00	572/587 (97%)	580/587 (99%)	0/587 (0%)
	305	1009	73739	unnamed protein product	889	0.00E+00	513/603 (85%)	561/603 (93%)	4/603 (1%)
	386	1035	208090	unnamed protein product	566	0.00E+00	302/304 (99%)	302/304 (99%)	0/304 (0%)
	388	808	219727	unnamed protein product	595	0.00E+00	343/350 (98%)	345/350 (99%)	0/350 (0%)
	794	781	235253	unnamed protein product	1144	0.00E+00	582/594 (98%)	588/594 (99%)	0/594 (0%)
	908	927	238031	unnamed protein product	1103	0.00E+00	553/724 (76%)	628/724 (87%)	19/724 (3%)
	912	1745	253810	unnamed protein product	1979	0.00E+00	1117/1404 (80%)	1216/1404 (87%)	30/1404 (2%)
	914	955	253812	unnamed protein product	1148	0.00E+00	586/587 (99%)	586/587 (99%)	0/587 (0%)
	1156	1033	252914	unnamed protein product	1919	0.00E+00	995/1032 (96%)	1018/1032 (99%)	0/1032 (0%)

	1224	960	250532	unnamed protein product	1425	0.00E+00	801/803 (99%)	803/803 (100%)	0/803 (0%)
	1225	612	250531	unnamed protein product	1189	0.00E+00	610/611 (99%)	610/611 (99%)	0/611 (0%)
	1238	475	80749	unnamed protein product	805	0.00E+00	460/461 (99%)	461/461 (100%)	0/461 (0%)
	1337	603	237142	unnamed protein product	707	0.00E+00	369/370 (99%)	370/370 (100%)	0/370 (0%)
	1545	1809	209036	unnamed protein product	624	0.00E+00	309/309 (100%)	309/309 (100%)	0/309 (0%)
	1578	1044	245395	unnamed protein product	884	0.00E+00	501/558 (90%)	514/558 (92%)	20/558 (4%)
	1579	423	231664	unnamed protein product	360	4.00E-123	180/181 (99%)	180/181 (99%)	0/181 (0%)
	1597	3175	236788	unnamed protein product	737	0.00E+00	431/433 (99%)	431/433 (99%)	0/433 (0%)
	1870	1777	225748	unnamed protein product	814	0.00E+00	448/449 (99%)	448/449 (99%)	0/449 (0%)
	1883	473	247935	unnamed protein product	911	0.00E+00	471/472 (99%)	471/472 (99%)	0/472 (0%)
	2167	767	231690	unnamed protein product	1055	0.00E+00	550/557 (99%)	557/557 (100%)	0/557 (0%)
7000000239	912	1745	226505	unnamed protein product	380	3.00E-120	211/212 (99%)	211/212 (99%)	0/212 (0%)
Subject 763496533	914	955	199844	unnamed protein product	283	2.00E-88	157/157 (100%)	157/157 (100%)	0/157 (0%)
	1156	1033	148060	unnamed protein product	223	3.00E-67	108/108 (100%)	108/108 (100%)	0/108 (0%)
	1224	960	198169	unnamed protein product	307	2.00E-97	155/155 (100%)	155/155 (100%)	0/155 (0%)
	1597	3175	207291	unnamed protein product	335	7.00E-104	169/169 (100%)	169/169 (100%)	0/169 (0%)
7000000271	50	426	147551	unnamed protein product	349	1.00E-118	188/189 (99%)	188/189 (99%)	0/189 (0%)
Subject 508703490	241	2601	159371	unnamed protein product	1740	0.00E+00	845/845 (100%)	845/845 (100%)	0/845 (0%)
	305	1009	153722	unnamed protein product	801	0.00E+00	498/498 (100%)	498/498 (100%)	0/498 (0%)
	386	1035	158128	unnamed protein product	912	0.00E+00	472/472 (100%)	472/472 (100%)	0/472 (0%)
	388	808	161940	unnamed protein product	1440	0.00E+00	786/786 (100%)	786/786 (100%)	0/786 (0%)
	794	781	148326	unnamed protein product	702	0.00E+00	353/358 (99%)	355/358 (99%)	0/358 (0%)
	908	927	78566	unnamed protein product	1855	0.00E+00	926/926 (100%)	926/926 (100%)	0/926 (0%)
	912	1745	160217	unnamed protein product	1769	0.00E+00	930/937 (99%)	932/937 (99%)	0/937 (0%)
	914	955	22515	unnamed protein product	1876	0.00E+00	952/954 (99%)	953/954 (99%)	0/954 (0%)
	1156	1033	164933	unnamed protein product	1978	0.00E+00	1032/1032 (100%)	1032/1032 (100%)	0/1032 (0%)
	1224	960	157677	unnamed protein product	1257	0.00E+00	694/694 (100%)	694/694 (100%)	0/694 (0%)
	1225	612	155688	unnamed protein product	1098	0.00E+00	565/565 (100%)	565/565 (100%)	0/565 (0%)
	1238	475	60984	unnamed protein product	499	1.00E-175	295/295 (100%)	295/295 (100%)	0/295 (0%)

	1337	603	156522	unnamed protein product	775	0.00E+00	422/422 (100%)	422/422 (100%)	0/422 (0%)
	1545	1809	156331	unnamed protein product	1224	0.00E+00	619/619 (100%)	619/619 (100%)	0/619 (0%)
	1578	1044	159968	unnamed protein product	981	0.00E+00	638/638 (100%)	638/638 (100%)	0/638 (0%)
	1579	423	159967	unnamed protein product	349	3.00E-119	173/173 (100%)	173/173 (100%)	0/173 (0%)
	1597	3175	75644	unnamed protein product	1429	0.00E+00	799/799 (100%)	799/799 (100%)	0/799 (0%)
	1870	1777	161002	unnamed protein product	1233	0.00E+00	631/631 (100%)	631/631 (100%)	0/631 (0%)
	1883	473	162738	unnamed protein product	913	0.00E+00	472/472 (100%)	472/472 (100%)	0/472 (0%)
	2167	767	158509	unnamed protein product	562	0.00E+00	291/291 (100%)	291/291 (100%)	0/291 (0%)
7000000317	794	781	204308	unnamed protein product	241	3.00E-74	119/122 (98%)	120/122 (98%)	0/122 (0%)
Subject 763577454	908	927	202595	unnamed protein product	246	3.00E-75	120/120 (100%)	120/120 (100%)	0/120 (0%)
	912	1745	180923	unnamed protein product	209	5.00E-61	100/100 (100%)	100/100 (100%)	0/100 (0%)
	1224	960	194674	unnamed protein product	214	5.00E-64	102/102 (100%)	102/102 (100%)	0/102 (0%)
	1545	1809	66505	unnamed protein product	368	8.00E-116	209/209 (100%)	209/209 (100%)	0/209 (0%)
	1597	3175	214208	unnamed protein product	245	7.00E-73	117/134 (87%)	123/134 (92%)	0/134 (0%)
7000000332	50	426	173355	unnamed protein product	259	1.00E-84	126/126 (100%)	126/126 (100%)	0/126 (0%)
Subject 763901136	241	2601	200578	unnamed protein product	688	0.00E+00	342/342 (100%)	342/342 (100%)	0/342 (0%)
	305	1009	137059	unnamed protein product	268	7.00E-83	135/135 (100%)	135/135 (100%)	0/135 (0%)
	386	1035	116031	unnamed protein product	240	3.00E-73	114/114 (100%)	114/114 (100%)	0/114 (0%)
	794	781	201240	unnamed protein product	563	0.00E+00	343/347 (99%)	345/347 (99%)	0/347 (0%)
	908	927	218372	unnamed protein product	1137	0.00E+00	576/576 (100%)	576/576 (100%)	0/576 (0%)
	912	1745	203470	unnamed protein product	713	0.00E+00	365/367 (99%)	366/367 (99%)	0/367 (0%)
	914	955	203145	unnamed protein product	707	0.00E+00	364/364 (100%)	364/364 (100%)	0/364 (0%)
	1156	1033	204434	unnamed protein product	698	0.00E+00	377/377 (100%)	377/377 (100%)	0/377 (0%)
	1224	960	203241	unnamed protein product	689	0.00E+00	365/365 (100%)	365/365 (100%)	0/365 (0%)
	1225	612	186828	unnamed protein product	531	0.00E+00	255/255 (100%)	255/255 (100%)	0/255 (0%)
	1238	475	186459	unnamed protein product	416	2.00E-143	254/254 (100%)	254/254 (100%)	0/254 (0%)
	1337	603	190190	unnamed protein product	456	1.00E-156	272/272 (100%)	272/272 (100%)	0/272 (0%)
	1545	1809	192862	unnamed protein product	573	0.00E+00	287/287 (100%)	287/287 (100%)	0/287 (0%)
	1578	1044	191004	unnamed protein product	425	2.00E-139	245/245 (100%)	245/245 (100%)	0/245 (0%)

	1579	423	205482	unnamed protein product	512	0.00E+00	305/305 (100%)	305/305 (100%)	0/305 (0%)
	1597	3175	211196	unnamed protein product	709	0.00E+00	443/443 (100%)	443/443 (100%)	0/443 (0%)
	1870	1777	134799	unnamed protein product	266	9.00E-81	132/132 (100%)	132/132 (100%)	0/132 (0%)
	1883	473	189272	unnamed protein product	546	0.00E+00	268/268 (100%)	268/268 (100%)	0/268 (0%)
	2167	767	195818	unnamed protein product	586	0.00E+00	306/306 (100%)	306/306 (100%)	0/306 (0%)
7000000334	50	426	89320	unnamed protein product	729	0.00E+00	413/413 (100%)	413/413 (100%)	0/413 (0%)
Subject 764143897	241	2601	242058	unnamed protein product	5211	0.00E+00	2581/2581 (100%)	2581/2581 (100%)	0/2581 (0%)
	305	1009	243910	unnamed protein product	1826	0.00E+00	1008/1008 (100%)	1008/1008 (100%)	0/1008 (0%)
	386	1035	228997	unnamed protein product	1582	0.00E+00	807/807 (100%)	807/807 (100%)	0/807 (0%)
	388	808	237428	unnamed protein product	1487	0.00E+00	807/807 (100%)	807/807 (100%)	0/807 (0%)
	794	781	235583	unnamed protein product	1394	0.00E+00	769/780 (99%)	774/780 (99%)	0/780 (0%)
	908	927	12054	unnamed protein product	1855	0.00E+00	926/926 (100%)	926/926 (100%)	0/926 (0%)
	912	1745	12058	unnamed protein product	3142	0.00E+00	1717/1725 (99%)	1720/1725 (99%)	0/1725 (0%)
	914	955	12060	unnamed protein product	1843	0.00E+00	937/939 (99%)	938/939 (99%)	0/939 (0%)
	1156	1033	98222	unnamed protein product	1978	0.00E+00	1032/1032 (100%)	1032/1032 (100%)	0/1032 (0%)
	1224	960	237639	unnamed protein product	1711	0.00E+00	959/959 (100%)	959/959 (100%)	0/959 (0%)
	1225	612	237641	unnamed protein product	1192	0.00E+00	611/611 (100%)	611/611 (100%)	0/611 (0%)
	1238	475	100477	unnamed protein product	805	0.00E+00	461/461 (100%)	461/461 (100%)	0/461 (0%)
	1337	603	224886	unnamed protein product	987	0.00E+00	602/602 (100%)	602/602 (100%)	0/602 (0%)
	1545	1809	231439	unnamed protein product	1878	0.00E+00	943/943 (100%)	943/943 (100%)	0/943 (0%)
	1578	1044	127895	unnamed protein product	976	0.00E+00	620/620 (100%)	620/620 (100%)	0/620 (0%)
	1579	423	87985	unnamed protein product	701	0.00E+00	419/419 (100%)	419/419 (100%)	0/419 (0%)
	1597	3175	242975	unnamed protein product	2239	0.00E+00	1316/1316 (100%)	1316/1316 (100%)	0/1316 (0%)
	1870	1777	99797	unnamed protein product	1969	0.00E+00	999/1003 (99%)	1001/1003 (99%)	0/1003 (0%)
	1883	473	65836	unnamed protein product	913	0.00E+00	472/472 (100%)	472/472 (100%)	0/472 (0%)
	2167	767	229477	unnamed protein product	1377	0.00E+00	714/714 (100%)	714/714 (100%)	0/714 (0%)
7000000352	50	426	112170	unnamed protein product	333	3.00E-112	180/180 (100%)	180/180 (100%)	0/180 (0%)
Subject 763860675	241	2601	300133	unnamed protein product	3116	0.00E+00	1522/1523 (99%)	1522/1523 (99%)	0/1523 (0%)

	305	1009	123556	unnamed protein product	1060	0.00E+00	540/540 (100%)	540/540 (100%)	0/540 (0%)
	386	1035	98967	unnamed protein product	2005	0.00E+00	1034/1034 (100%)	1034/1034 (100%)	0/1034 (0%)
	388	808	98969	unnamed protein product	1487	0.00E+00	807/807 (100%)	807/807 (100%)	0/807 (0%)
	794	781	7109	unnamed protein product	813	0.00E+00	415/422 (98%)	418/422 (99%)	0/422 (0%)
	908	927	15857	unnamed protein product	1855	0.00E+00	926/926 (100%)	926/926 (100%)	0/926 (0%)
	912	1745	15861	unnamed protein product	3142	0.00E+00	1717/1725 (99%)	1720/1725 (99%)	0/1725 (0%)
	914	955	11610	unnamed protein product	1843	0.00E+00	937/939 (99%)	938/939 (99%)	0/939 (0%)
	1156	1033	127944	unnamed protein product	1978	0.00E+00	1032/1032 (100%)	1032/1032 (100%)	0/1032 (0%)
	1224	960	110454	unnamed protein product	1711	0.00E+00	959/959 (100%)	959/959 (100%)	0/959 (0%)
	1225	612	110456	unnamed protein product	1192	0.00E+00	611/611 (100%)	611/611 (100%)	0/611 (0%)
	1238	475	115852	unnamed protein product	805	0.00E+00	461/461 (100%)	461/461 (100%)	0/461 (0%)
	1337	603	118447	unnamed protein product	987	0.00E+00	602/602 (100%)	602/602 (100%)	0/602 (0%)
	1545	1809	304778	unnamed protein product	1878	0.00E+00	943/943 (100%)	943/943 (100%)	0/943 (0%)
	1578	1044	127895	unnamed protein product	976	0.00E+00	620/620 (100%)	620/620 (100%)	0/620 (0%)
	1579	423	127894	unnamed protein product	707	0.00E+00	422/422 (100%)	422/422 (100%)	0/422 (0%)
	1597	3175	6319	unnamed protein product	4084	0.00E+00	2525/2547 (99%)	2527/2547 (99%)	20/2547 (1%)
	1870	1777	301253	unnamed protein product	1405	0.00E+00	716/716 (100%)	716/716 (100%)	0/716 (0%)
	1883	473	304831	unnamed protein product	913	0.00E+00	472/472 (100%)	472/472 (100%)	0/472 (0%)
	2167	767	298012	unnamed protein product	1484	0.00E+00	766/766 (100%)	766/766 (100%)	0/766 (0%)
7000000417	912	1745	136645	unnamed protein product	319	3.00E-99	154/154 (100%)	154/154 (100%)	0/154 (0%)
Subject 158742018	1883	473	91428	unnamed protein product	223	2.00E-70	106/106 (100%)	106/106 (100%)	0/106 (0%)
7000000475	241	2601	136676	unnamed protein product	245	3.00E-73	120/120 (100%)	120/120 (100%)	0/120 (0%)
Subject 160704339	305	1009	115404	unnamed protein product	143	1.00E-38	105/105 (100%)	105/105 (100%)	0/105 (0%)
	908	927	122829	unnamed protein product	232	2.00E-70	109/109 (100%)	109/109 (100%)	0/109 (0%)
	912	1745	149022	unnamed protein product	269	7.00E-82	129/131 (98%)	130/131 (99%)	0/131 (0%)
	914	955	111206	unnamed protein product	202	5.00E-60	102/102 (100%)	102/102 (100%)	0/102 (0%)
	1156	1033	170170	unnamed protein product	270	5.00E-83	155/155 (100%)	155/155 (100%)	0/155 (0%)
	1238	475	136596	unnamed protein product	211	2.00E-65	119/119 (100%)	119/119 (100%)	0/119 (0%)
	1337	603	141522	unnamed protein product	208	3.00E-63	124/124 (100%)	124/124 (100%)	0/124 (0%)

	1545	1809	188580	unnamed protein product	370	1.00E-116	185/185 (100%)	185/185 (100%)	0/185 (0%)
	1597	3175	144945	unnamed protein product	226	2.00E-66	112/112 (100%)	112/112 (100%)	0/112 (0%)
	1870	1777	126604	unnamed protein product	238	2.00E-71	113/113 (100%)	113/113 (100%)	0/113 (0%)
7000000736	241	2601	279128	unnamed protein product	729	0.00E+00	352/372 (95%)	363/372 (98%)	1/372 (0%)
Subject 763678604	305	1009	238079	unnamed protein product	281	7.00E-87	162/185 (88%)	172/185 (93%)	0/185 (0%)
	386	1035	265316	unnamed protein product	552	0.00E+00	272/273 (99%)	272/273 (99%)	0/273 (0%)
	388	808	75239	unnamed protein product	348	2.00E-113	171/172 (99%)	171/172 (99%)	0/172 (0%)
	794	781	165479	unnamed protein product	209	8.00E-63	106/106 (100%)	106/106 (100%)	0/106 (0%)
	908	927	205074	unnamed protein product	283	1.00E-88	137/137 (100%)	137/137 (100%)	0/137 (0%)
	912	1745	292233	unnamed protein product	527	3.00E-172	279/280 (99%)	279/280 (99%)	0/280 (0%)
	914	955	276541	unnamed protein product	704	0.00E+00	346/346 (100%)	346/346 (100%)	0/346 (0%)
	1156	1033	281524	unnamed protein product	684	0.00E+00	364/389 (94%)	379/389 (97%)	0/389 (0%)
	1224	960	275162	unnamed protein product	603	0.00E+00	334/335 (99%)	335/335 (100%)	0/335 (0%)
	1225	612	296396	unnamed protein product	806	0.00E+00	390/391 (99%)	390/391 (99%)	0/391 (0%)
	1238	475	274566	unnamed protein product	417	2.00E-143	252/253 (99%)	253/253 (100%)	0/253 (0%)
	1337	603	260182	unnamed protein product	363	9.00E-121	202/202 (100%)	202/202 (100%)	0/202 (0%)
	1545	1809	265485	unnamed protein product	488	9.00E-158	273/275 (99%)	274/275 (99%)	0/275 (0%)
	1578	1044	276870	unnamed protein product	526	5.00E-177	345/348 (99%)	345/348 (99%)	0/348 (0%)
	1579	423	278811	unnamed protein product	513	0.00E+00	304/308 (99%)	306/308 (99%)	0/308 (0%)
	1597	3175	258667	unnamed protein product	460	3.00E-146	244/245 (99%)	245/245 (100%)	0/245 (0%)
	1870	1777	262008	unnamed protein product	468	6.00E-151	251/251 (100%)	251/251 (100%)	0/251 (0%)
	1883	473	283950	unnamed protein product	755	0.00E+00	369/370 (99%)	369/370 (99%)	0/370 (0%)
	2167	767	267331	unnamed protein product	552	0.00E+00	279/284 (98%)	281/284 (99%)	0/284 (0%)

Appendix 22 US nationwide dietary intake data for the years 2003 - 2006. The data was collected in 'What We Eat in America', a program in the National Health and Nutrition Examination Survey (NHANES) which is part of the Centers for Disease Control and Prevention (CDC). Two days of 24-hour dietary records were collected over a period of 3 - 10 days. No copyright permission is required for the use of this data.

Dietary Fiber (g): Usual Intakes[#] from Food and Water, 2003-2006, Percentiles and Standard Errors

Age in years	Percentiles (SE)													
	5th		10th		25th		50th		75th		90th		95th	
Males and females:														
1-3	4.7	(0.21)	5.6	(0.20)	7.1	(0.18)	9.1	(0.18)	11.4	(0.24)	13.7	(0.35)	15.2	(0.43)
4-8	7.4	(0.38)	8.2	(0.33)	9.7	(0.23)	11.7	(0.16)	13.9	(0.29)	16.2	(0.56)	17.6	(0.75)
Males:														
9-13	8.7	(0.68)	9.7	(0.63)	11.7	(0.56)	14.2	(0.53)	17.1	(0.66)	20.0	(0.95)	22.0	(1.19)
14-18	7.8	(0.66)	9.1	(0.61)	11.5	(0.52)	14.7	(0.39)	18.5	(0.40)	22.4	(0.67)	25.0	(0.93)
19-30	8.6	(0.62)	10.0	(0.59)	12.8	(0.52)	16.6	(0.47)	21.2	(0.60)	26.0	(0.92)	29.2	(1.18)
31-50	8.9	(0.51)	10.5	(0.47)	13.6	(0.37)	17.5	(0.32)	22.1	(0.43)	26.8	(0.68)	29.9	(0.85)
19-50	8.7	(0.43)	10.2	(0.40)	13.3	(0.34)	17.2	(0.29)	21.7	(0.36)	26.5	(0.56)	29.6	(0.72)
51-70	7.8	(0.25)	9.3	(0.24)	12.4	(0.24)	16.5	(0.33)	21.4	(0.52)	26.7	(0.78)	30.2	(0.97)
71 and over	7.7	(0.35)	9.2	(0.35)	12.0	(0.35)	15.8	(0.37)	20.2	(0.46)	24.9	(0.65)	27.9	(0.77)
51 and over	7.8	(0.22)	9.3	(0.20)	12.3	(0.19)	16.3	(0.26)	21.1	(0.42)	26.2	(0.65)	29.6	(0.81)
19 and over	8.4	(0.27)	9.9	(0.25)	12.9	(0.22)	16.9	(0.22)	21.5	(0.28)	26.4	(0.41)	29.7	(0.51)
Females:														
9-13	8.3	(0.77)	9.2	(0.69)	10.7	(0.53)	12.6	(0.39)	14.7	(0.49)	16.8	(0.81)	18.2	(1.06)
14-18	6.2	(0.50)	7.2	(0.47)	9.0	(0.39)	11.4	(0.32)	14.2	(0.38)	17.2	(0.65)	19.1	(0.86)
19-30	6.4	(0.42)	7.4	(0.41)	9.4	(0.40)	12.1	(0.43)	15.3	(0.55)	18.6	(0.77)	20.8	(0.95)
31-50	6.4	(0.36)	7.6	(0.33)	10.1	(0.29)	13.2	(0.29)	16.9	(0.39)	20.8	(0.60)	23.3	(0.77)
19-50	6.3	(0.26)	7.5	(0.26)	9.9	(0.25)	12.9	(0.27)	16.4	(0.37)	20.1	(0.53)	22.6	(0.66)
51-70	6.9	(0.35)	8.1	(0.36)	10.5	(0.36)	13.7	(0.36)	17.4	(0.40)	21.3	(0.50)	23.8	(0.61)
71 and over	6.3	(0.28)	7.5	(0.27)	9.9	(0.27)	13.1	(0.32)	16.7	(0.49)	20.5	(0.77)	23.1	(1.00)
51 and over	6.8	(0.26)	8.0	(0.26)	10.4	(0.26)	13.5	(0.26)	17.1	(0.29)	21.0	(0.36)	23.5	(0.44)
19 and over	6.5	(0.21)	7.7	(0.21)	10.1	(0.21)	13.2	(0.23)	16.8	(0.28)	20.6	(0.39)	23.1	(0.47)
Pregnant 19-50	8.3	(0.75)	9.8	(0.69)	12.6	(0.57)	16.6	(0.55)	21.1	(0.79)	25.9	(1.22)	29.2	(1.65)
Males and females:														
1 and over	6.8	(0.16)	8.1	(0.16)	10.7	(0.17)	14.1	(0.17)	18.2	(0.20)	22.4	(0.27)	25.3	(0.34)