



# Genomic insights from *Monoglobus pectinilyticus*: a pectin-degrading specialist bacterium in the human colon

Caroline C. Kim<sup>1,2</sup> · Genelle R. Lunken<sup>1,3</sup> · William J. Kelly<sup>4</sup> · Mark L. Patchett<sup>2</sup> · Zoe Jordens<sup>2</sup> · Gerald W. Tannock<sup>5</sup> · Ian M. Sims<sup>6</sup> · Tracey J. Bell<sup>6</sup> · Duncan Hedderley<sup>1</sup> · Bernard Henrissat<sup>7,8,9</sup> · Douglas I. Rosendale<sup>1</sup>

Received: 19 April 2018 / Revised: 7 January 2019 / Accepted: 19 January 2019 / Published online: 6 February 2019  
© International Society for Microbial Ecology 2019, corrected publication 2023

## Abstract

Pectin is abundant in modern day diets, as it comprises the middle lamellae and one-third of the dry carbohydrate weight of fruit and vegetable cell walls. Currently there is no specialized model organism for studying pectin fermentation in the human colon, as our collective understanding is informed by versatile glycan-degrading bacteria rather than by specialist pectin degraders. Here we show that the genome of *Monoglobus pectinilyticus* possesses a highly specialized glycobiome for pectin degradation, unique amongst *Firmicutes* known to be in the human gut. Its genome encodes a simple set of metabolic pathways relevant to pectin sugar utilization, and its predicted glycobiome comprises an unusual distribution of carbohydrate-active enzymes (CAZymes) with numerous extracellular methyl/acetyl esterases and pectate lyases. We predict the *M. pectinilyticus* degradative process is facilitated by cell-surface S-layer homology (SLH) domain-containing proteins, which proteomics analysis shows are differentially expressed in response to pectin. Some of these abundant cell surface proteins of *M. pectinilyticus* share unique modular organizations rarely observed in human gut bacteria, featuring pectin-specific CAZyme domains and the cell wall-anchoring SLH motifs. We observed *M. pectinilyticus* degrades various pectins, RG-I, and galactan to produce polysaccharide degradation products (PDPs) which are presumably shared with other inhabitants of the human gut microbiome (HGM). This strain occupies a new ecological niche for a primary degrader specialized in foraging a habitually consumed plant glycan, thereby enriching our understanding of the diverse community profile of the HGM.

## Introduction

The human diet includes plant cell wall (PCW) polysaccharides that serve as fermentable nutrients for the

complex microbial community found in the lower gastrointestinal tract. Along with cellulose and hemicellulose, pectin is a major polysaccharide constituting the PCW. Thick layers of pectin also cover the surface of the PCW, forming middle lamellae between the shared cell wall interfaces [1]. Pectin is the most complex polysaccharide found in PCW, consisting of structurally heterogeneous components, such as homogalacturonan (HG), rhamnogalacturonan-I (RG-I), and

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-019-0363-6>) contains supplementary material, which is available to authorized users.

✉ Caroline C. Kim  
caroline.kim@plantandfood.co.nz

✉ Douglas I. Rosendale  
douglas.rosendale@plantandfood.co.nz

<sup>1</sup> The New Zealand Institute for Plant and Food Research, Palmerston North 4474, New Zealand

<sup>2</sup> Institute of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand

<sup>3</sup> Massey Institute of Food Science and Technology, School of Food and Nutrition, Massey University, Palmerston North, New Zealand

<sup>4</sup> Donvis Limited, Ashurst 4810, New Zealand

<sup>5</sup> Department of Microbiology and Immunology, Microbiome Otago, University of Otago, Dunedin 9016, New Zealand

<sup>6</sup> Ferrier Research Institute, Victoria University of Wellington, Gracefield Research Centre, Lower Hutt 5040, New Zealand

<sup>7</sup> Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille University, Marseille F-13288, France

<sup>8</sup> Institut National de la Recherche Agronomique, USC1408 Architecture et Fonction des Macromolécules Biologiques, Marseille F-13288, France

<sup>9</sup> Department of Biological Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia

rhamnogalacturonan-II (RG-II) [2]. HG is a homogenous polymer of  $\alpha$ -1,4-linked-D-galacturonic acid (D-GalpA) which constitutes the majority of uronic acid contents of pectin, and 65–70% of the total pectin mass [2, 3]. Approximately half of D-GalpA residues present in HG are either methyl-esterified at C-6 or acetyl-esterified at O-2 and/or O-3 [3]. Non-esterified D-GalpA residues carry a negative charge, enabling the formation of a gel-like texture by chelating  $\text{Ca}^{2+}$  ions [4]. The RG-I and RG-II regions of pectin are compositionally heterogeneous, containing diverse neutral sugars. Depending on the plant species, up to 20–80% of L-rhamnose (L-Rhap) residues in RG-I are branched by arabinan (polymers of  $\alpha$ -L-1,5-arabinofuranose (L-Araf) units branched at O-2 and O-3 with  $\alpha$ -L-Araf residues), galactan (unbranched polymers of  $\beta$ -D-1,4-galactopyranose (D-Galp) residues), and arabinogalactan (a linear  $\beta$ -1,4-galactan substituted with  $\alpha$ -L-1,5-Araf oligosaccharides) [2]. Some arabinan and galactan are substituted with ferulic acid side chains, which can dimerize to strengthen the pectin network [5]. The backbone of RG-I consists of alternating diglycosyl units of  $\alpha$ -D-GalpA and  $\alpha$ -L-Rhap [ $\rightarrow$ 2)- $\alpha$ -L-Rhap-(1 $\rightarrow$ 4)- $\alpha$ -D-GalpA-(1 $\rightarrow$ ] [6]; whereas, the backbone of RG-II is made of a linear  $\alpha$ -1,4-L-GalpA residues, and does not contain L-Rhap units as a part of the basal structure [7]. The RG-II is less abundant than RG-I, but shows a higher degree of structural complexity as it contains at least 13 glycosyl residues covalently linked together by more than 21 different types of glycosidic linkages [8, 9]. In the primary cell wall, RG-II predominantly occurs as a dimer crosslinked by a borate diester [7].

The structural complexity of pectin poses a considerable challenge to the human digestive system, and humans must rely on the concerted actions of CAZymes produced by symbiotic gut bacteria for pectin degradation. Related research with environmental and animal gut bacteria suggest that the erosion of middle lamellae by pectin-degrading bacteria is a necessary prerequisite for initiating PCW degradation, exposing other cell wall polymers and allowing the establishment of a more heterogeneous bacterial colonization along the PCW [10, 11]. So far, such association has been difficult to examine in humans partly due to the technical difficulty of cultivating anaerobic gut bacteria. Our current knowledge is best exemplified by *Bacteroides* spp., Gram-negative generalist carbohydrate degraders well-studied for their versatile glycan-foraging strategy using gene clusters called polysaccharide utilisation loci (PULs) [8, 12–14]. Upon detecting a signalling carbohydrate, a PUL is activated to express surface glycan-binding proteins, outer membrane oligosaccharide transporters, surface/periplasmic CAZymes, and SusC and SusD homologues [15]. This synchronous production of components comprising a complete glycan-foraging unit underpins the sequestration model of *Bacteroides* which efficiently binds, degrades, sequesters,

and transports PDPs into the intracellular space while reducing losses to other members of the HGM [12, 14, 16], although cases of glycan sharing were also reported [17, 18]. Although the important role of Gram-positive *Firmicutes* as primary degraders that break down dietary glycans to release products to the HGM in a cross-feeding relationship has been proposed [12], the mechanisms by which *Firmicutes* degrade pectin in the human gut are not well understood. Currently known pectin-degrading *Firmicutes* species are few in number, namely *Eubacterium eligens* [19, 20] and *Faecalibacterium prausnitzii* [21] that possess relatively small repertoires of CAZyme-encoding genes involved in pectin degradation. Previously, we reported the isolation of a novel *Firmicutes* bacterium from human faeces (*M. pectinilyticus* 14 T), whose selective growth on pectin was mediated through a tight cell-substrate interaction [22] which is considered as a hallmark trait of primary PCW-degrading bacteria [23]. This raised the possibility that within the human gut there is an insufficiently characterized ecological niche for pectin-degrading specialists, such as *M. pectinilyticus*, which initiate the cascade of PCW degradation by dissolving the obstructive pectin layers to expose attachment sites for heterogeneous bacterial species, and release oligosaccharides for utilization by secondary feeders of the microbial community. Building on this earlier discovery and recognizing the scarcity of bacterial model systems for studying the colonic pectin degradation, we sought to provide a first insight into the primary PCW degradation by examining the pectin degradation by *M. pectinilyticus* through combined proteogenomic and biochemical approaches. We generated a high-quality genome of *M. pectinilyticus* which revealed genetic, metabolic, and glycomicrobiological specializations for pectin degradation and utilization, as well as genes encoding unique protein features indicative of an extracellular glycan degradation strategy previously unseen in the HGM. To examine the host-bacterial relationships, we assessed the prevalence and abundance of *M. pectinilyticus* among healthy human subjects, and revealed a novel phylogenetic lineage of *Ruminococcaceae* currently represented by *M. pectinilyticus* and its uncultured relatives from gut systems of various terrestrial organisms.

## Results

### Taxonomic affiliation of *M. pectinilyticus* and its presence in human studies

We first sought to determine whether *M. pectinilyticus* or any related uncultured bacteria are true human gut commensals. An unfiltered-BLAST search of the GenBank database identified 77 near full-length 16S rRNA gene sequences (59 from human faeces and 18 from other human/

animal sources) with  $\geq 92\%$  sequence identities (query cover  $\geq 80\%$ ) to *M. pectinilyticus* (Supplementary Information S1). These sequences phylogenetically diverged from *Clostridium* clusters III and IV with a high bootstrap support, forming a novel polyphyletic clade within the family *Ruminococcaceae* (Fig. 1). The sequences from this clade formed five clusters (arbitrarily named MP-I to MP-V) in the phylogenetic tree, of which *M. pectinilyticus* and its uncultured relatives from human faeces formed a dominant cluster (MP-I). To further establish *M. pectinilyticus* as a human gut commensal, we performed quantitative PCR using faecal DNA samples collected from 44 healthy New Zealand volunteers, and detected *M. pectinilyticus* from 10 donors at a mean relative abundance of  $\sim 0.3\%$  (excluding an outlying donor 24) (Supplementary Information S2). To relate the presence of *M. pectinilyticus* to dietary consumption in donors, subjects were asked to provide four sets of 3-day diet records which were used to calculate the average daily consumption of different food groups (Supplementary Information S3). Among the 10 individuals that tested positive for *M. pectinilyticus*, 8 reported to consume more than the recommended amount of dietary fibre ( $>25$  g/day for females and  $>30$  g/day for males), and 5 individuals consumed  $>5$  g of pectin per day. Overall, members of the *M. pectinilyticus*-positive group showed higher median values for fibre and pectin intakes compared to those in the negative group (Supplementary Fig. S1a, b). We further calculated the daily servings of vegetable, fruit, grain, and protein for each donor, but statistically significant correlation to *M. pectinilyticus* was not observed under these categories (Supplementary Fig. S1c-f).

## Genome organization and general features

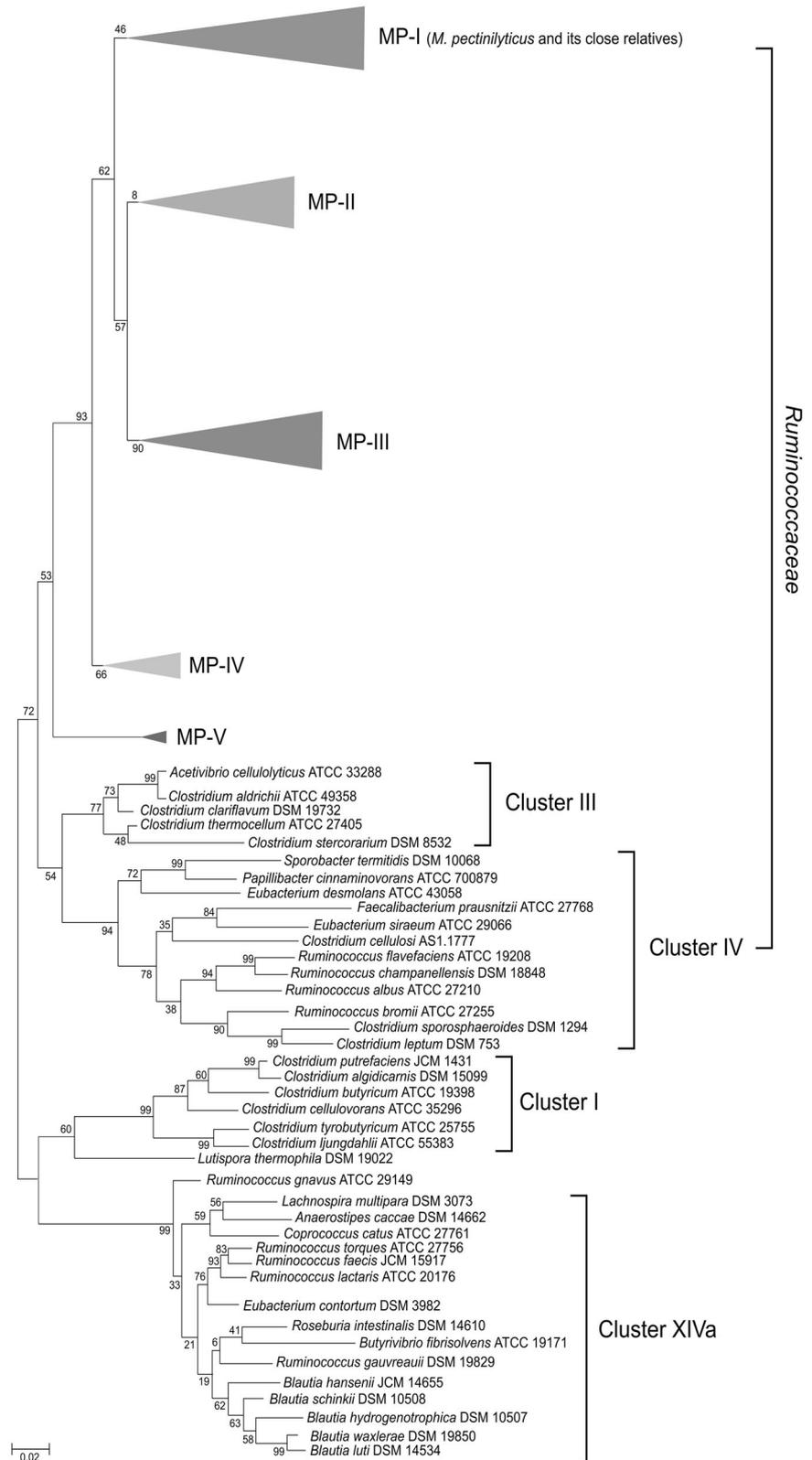
The draft genome of *M. pectinilyticus* (GenBank accession CP020991) was sequenced and assembled using the Illumina HiSeq 2500 system and SPAdes genome assembler [24]. The final genome assembly consists of a single contig of 2,757,678 bp, only lacking closure from a  $\sim 2$  kb sequencing gap within a predicted prophage region. Using a combination of Prokka [25] and other protein function prediction tools, a total of 2263 protein-coding sequences (CDS) were annotated, and HMMER-based pipelines were used to assign Clusters of Orthologous Groups (COG) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway genes (Supplementary Fig. S2). The annotated genome features a complete set of amino acid biosynthesis pathways, along with 46 transfer RNAs for all 20 natural L-amino acids. The genome lacks biosynthesis pathways for biotin, riboflavin, ascorbate, and folate, whereas complete pathways for nicotinate, vitamin B5, vitamin B6, and thiamine are present. No genes involved in flagella assembly or exopolysaccharide biosynthesis are present, in accordance

with our previous observations [22]. Genetic elements of four complete or partially degraded prophage genomes are present, as well as two putative integrated, horizontally transferred plasmids ( $\sim 51$  kb and  $\sim 73$  kb), and CRISPR type I-C and II-A systems.

## Glycobiome analysis

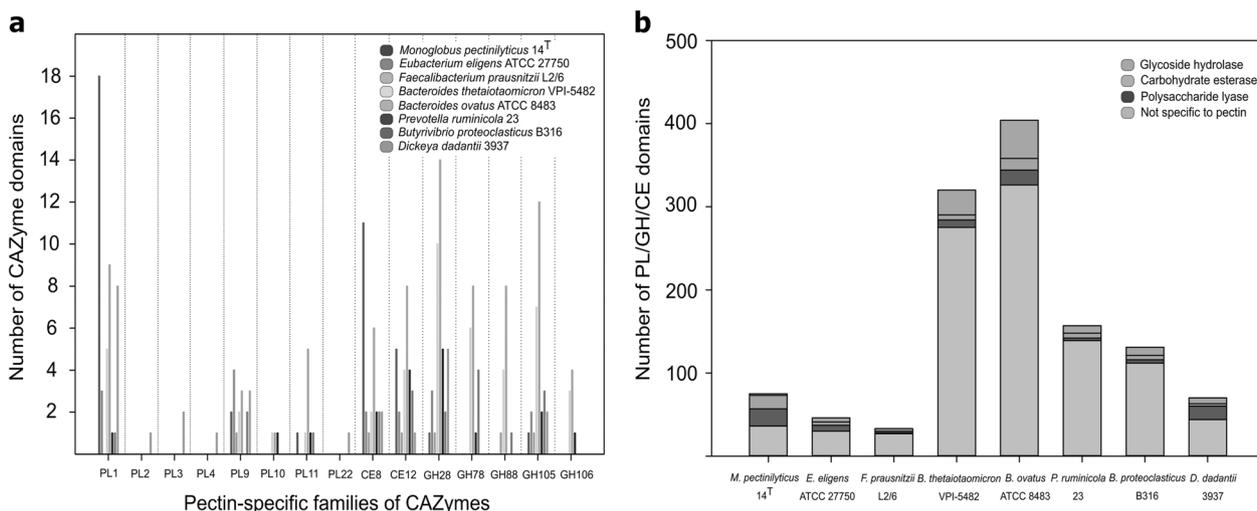
We identified glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), non-catabolic glycosyl transferases (GTs), and carbohydrate-binding modules (CBMs) using the CAZy database, finding 91 genes annotated to encode 108 putative CAZyme domains (Supplementary Information S4). The genome encodes 48 putative pectin-degrading CAZyme domains including 20 pectate/pectin lyases and rhamnogalacturonan lyases (PL1 and PL9), one rhamnogalacturonan lyase (PL11), eight pectin methyl-esterases (CE8), four pectin acetyl esterases (CE12), one polygalacturonase (GH28), one 2-keto-3-deoxynononic acid hydrolase/sialidase (GH33), three  $\beta$ -D-xylosidase/ $\alpha$ -L-arabinofuranosidases (GH43), one  $\alpha$ -L-arabinofuranosidase/ $\beta$ -D-xylosidase/ $\beta$ -1,4-D-xylanase/endoglucanase (GH51), one  $\alpha$ -D-galactosidase/ $\alpha$ -D-glucosidase (GH97), one unsaturated homogalacturonyl/rhamnogalacturonyl hydrolase (GH105), one apiosidase (GH140), and one multi-domain CE12/CE8. In addition, a novel CAZyme combination consisting of GH95/CE8 is encoded by B9O19\_1299 and B9O19\_1681, whose exact role in the pectin degradation is currently unclear. *M. pectinilyticus* also produces degradative CAZymes related to other PCW polymers, including five poly-specific GH3 family enzymes, three  $\alpha$ -L-fucosidases/ $\alpha$ -L-galactosidase (GH95), one  $\beta$ -D-glucosidase/ $\beta$ -D-xylosidase (GH116), and one poly-specific GH5 family enzyme. We acknowledge that predicting CAZyme functions and specificities based on informatics analysis has limitations as these enzyme families are often poly-specific, and lack functional characterization [26]. Considering the moderate glycobiome size of *M. pectinilyticus* compared to other pectinolytic bacteria available in the CAZy database, this bacterium possesses disproportionately large numbers of genes for CEs and PLs predicted to be involved in the initiation of pectin degradation (Fig. 2a). While pectate lyases are abundantly produced by fungal or bacterial plant pathogens (e.g. *Dickeya dadantii*), it is unusual amongst gut bacteria to derive a larger share of pectinolytic activity from PLs than GHs (Fig. 2b). Abundant production of CEs presumably provides a competitive advantage by a hierarchical removal of methyl- and acetyl-groups to facilitate rapid access by PLs, which in turn cleave HG and RG backbones to generate unsaturated oligomeric end-products of  $\beta$ -elimination [27]. Consistent with this, all PLs and most CE8 and CE12 contain putative signal peptide sequences, suggesting pectin degradation mostly occurs in the extracellular environment. We confirmed this by observing

**Fig. 1** Phylogenetic analysis reveals *M. pectinilyticus* is a human gut commensal from a previously unexplored taxonomic branch of *Ruminococcaceae*. This novel branch of *Ruminococcaceae* consists of polyphyletic sequence clusters (MP-I to MP-V) in which *M. pectinilyticus* and closely related uncultured bacterial 16S rRNA gene sequences are phylogenetically placed in the MP-I cluster. The identity of 16S rRNA gene sequences from each MP cluster is given in Supplementary Information S1. Bootstrap values were calculated using 2000 re-samplings



the gradual degradation of citrus and apple pectins in the culture supernatant of *M. pectinilyticus* using size exclusion chromatography (SEC) (Supplementary Fig. S3). The SEC

analysis results showed that *M. pectinilyticus* carries out an extracellular degradation of high molecular weight pectic carbohydrates while simultaneously releasing degraded



**Fig. 2** The glyco biome of *M. pectinilyticus* shows an unusual PL- and CE-enriched profile of extracellular CAZymes. **a** Comparison of PLs, GHs and CEs with pectin-specific activities between pectin-degrading strains isolated from the human gut (*M. pectinilyticus*, *E. eligens*, *F. prausnitzii*, *Bacteroides thetaiotaomicron*, and *Bacteroides ovatus*),

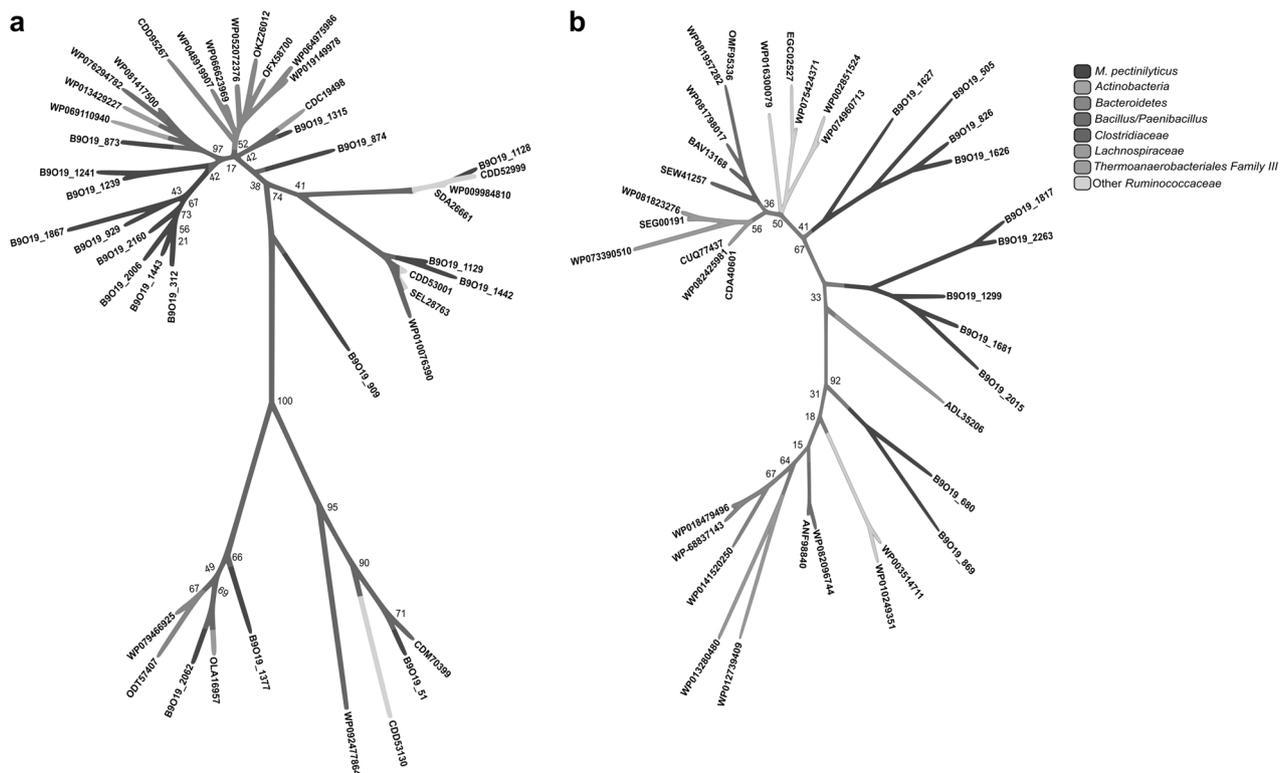
oligomers into the culture supernatant. Most pectin-related GH family enzymes, including GH28 and GH105, lack signal peptide sequences, indicating their roles in the intracellular processing of oligomers. Based on the informatics data, we initially predicted the primary target for degradation to be the pectin backbone rather than the side chains, as *M. pectinilyticus* lacks identifiable  $\beta$ -galactosidases/galactanases and terminal rhamnosidases required for the degradation of RG-I. Furthermore, the comprehensive RG-II degradome recently described from *B. thetaiotaomicron* [8] is only partially present in *M. pectinilyticus* (GH33, GH43, and GH140). To test our genome-based predictions, the oligosaccharide and monosaccharide degradation remnants in the 0 h, 48 h, and 72 h culture supernatants of *M. pectinilyticus* grown on arabinan (sugar beet), arabinogalactan (larch wood), RG-I (potato), and galactan (potato) were examined using high-performance anion-exchange chromatography (HPAEC) with pulsed amperometric detection (PAD). Unexpectedly, *M. pectinilyticus* degraded RG-I and galactan to generate smaller weight degradation products, whereas arabinan and arabinogalactan did not appear to be degraded (Supplementary Fig. S4 a-d). While *M. pectinilyticus* grows on RG-I (data not shown), no growth was observed on galactan, arabinan, and arabinogalactan over three successive transfers (data not shown), consistent with our previous data [22]. The degradation profiles from RG-I show the accumulation of oligosaccharides at 48 and 72 h (Supplementary Fig. S4 e), but not of L-Rhap monomers (Supplementary Fig. S4 g). This was consistent with our genome-based prediction that *M. pectinilyticus* does not produce rhamnosidases which enable the cleavage of terminal L-Rhap. Although *M. pectinilyticus* degrades potato galactan, the resulting D-Galp monomers

the animal rumen (*Prevotella ruminicola* and *Butyrivibrio proteoclasticus*), and plants infested with soft rot disease (*D. dadantii*). **b** Comparison of proportion of pectin-specific CAZymes relative to the total number of degradative CAZymes (PLs, GHs, and CEs) for each strain

accumulate in the culture supernatant without being utilized (Supplementary Fig. S4 f, h), consistent with our previous observation [22]. We attempted to predict carbohydrate recognition sites in silico, but few CBM families could be identified from the CAZymes of *M. pectinilyticus*, perhaps due to the small number of pectin-binding CBMs in the current CAZy database [28].

### PL1 and CE8 CAZyme phylogeny

The unusual PL1 and CE8-dominated CAZyme profile of *M. pectinilyticus* prompted us to attempt to infer the evolutionary history of these enzymes. We used a BlastP search of the NCBI protein database to find the closest sequence relatives to PL1 and CE8 of *M. pectinilyticus*. We then extracted catalytic domains from 193 PL1 and 85 CE8 sequences to construct phylogenetic trees using the maximum-likelihood method. 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 (Fig. 3 with extended phylogenetic trees in Supplementary Information S5), suggesting that these enzymes have independently evolved, and may fulfil functions specific to *M. pectinilyticus*. 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*. 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



**Fig. 3** 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

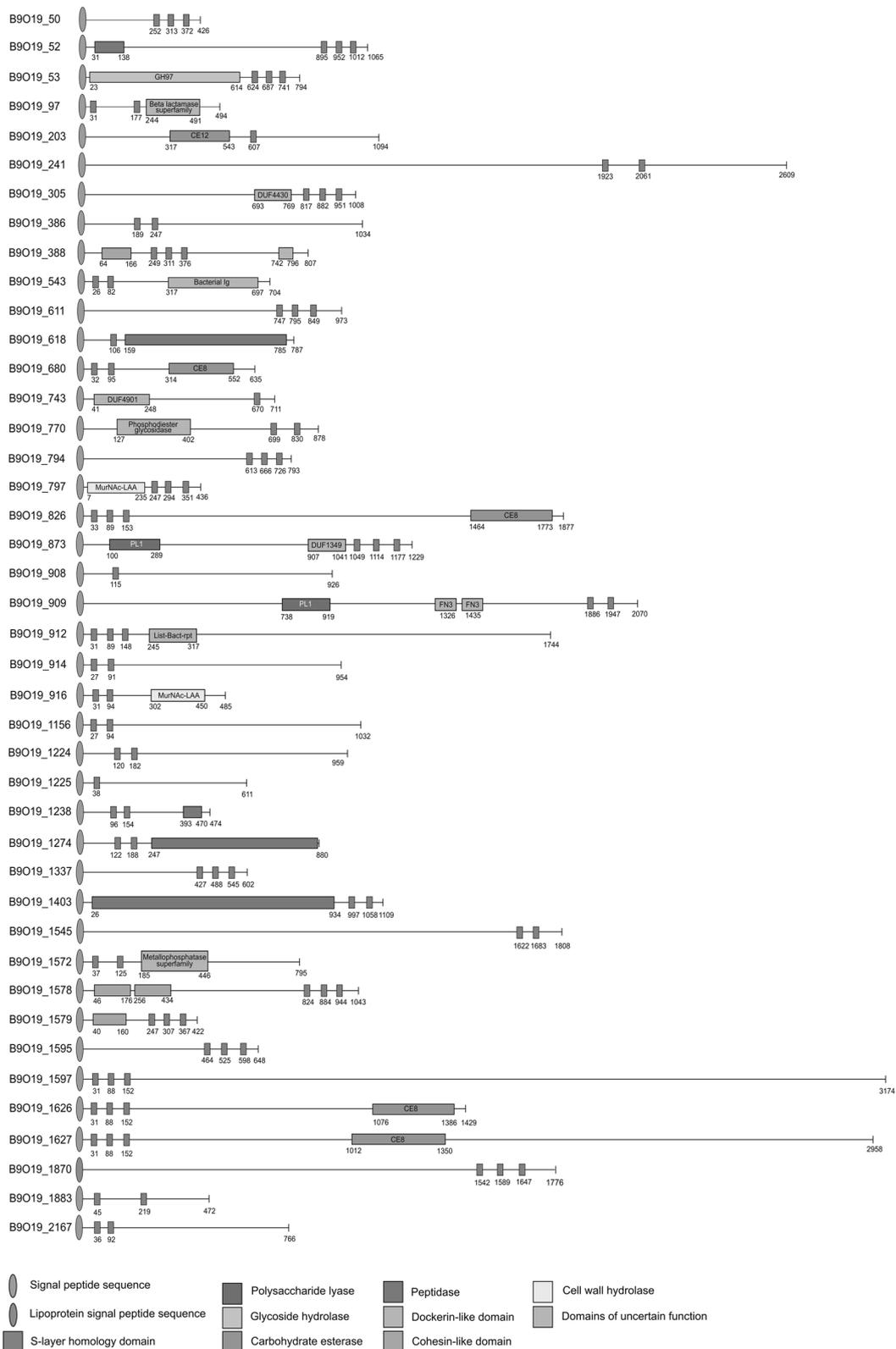
*M. pectinilyticus* PL1 and CE8 sequences. GenBank accession numbers and bootstrap values from 1000 re-samplings are indicated. Bacterial source of sequences are indicated with colours. Extended phylogenetic trees are given in Supplementary Information S5

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 PL1 catalytic domain showed  $< 40\%$  peptide sequence similarities ( $< 100$  bit-score; and  $e$ -value  $> e^{-30}$ ) to other PL1 domain sequences used for comparison 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 bormimense* (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. Although non-cellulosomal dockerin/cohesin modules are found in 14% of the known bacterial genomes, their functions remain obscure as most of these sequences occurs integral to hypothetical proteins with unknown functions [29]. All dockerin- and cohesin-containing proteins found in the *M. pectinilyticus* genome contain secretory signal peptide sequences and/or SLH domains, suggesting these proteins may function at the bacterial cell surface. Further examination of these proteins may reveal potentially novel catalytic and/or carbohydrate-binding functions in that dockerin and/or cohesin modules were previously found appended to CAZyme domains in a single protein among few members of the HGM [29].

### Modular organization of SLH domain-containing proteins

The *M. pectinilyticus* genome encodes 42 extracellular ‘SLH proteins’ predicted to attach to the extracellular surface via one to three SLH modules (Fig. 4). To our



**Fig. 4** The 42 SLH module-containing proteins of *M. pectinilyticus* contain few identifiable domains, including pectin-degrading CAZymes, putative dockerins and cohesins, and FN3- and Big-like domains with presumed functions in carbohydrate-binding and degradation. Protein lengths are drawn to scale. Amino acid positions

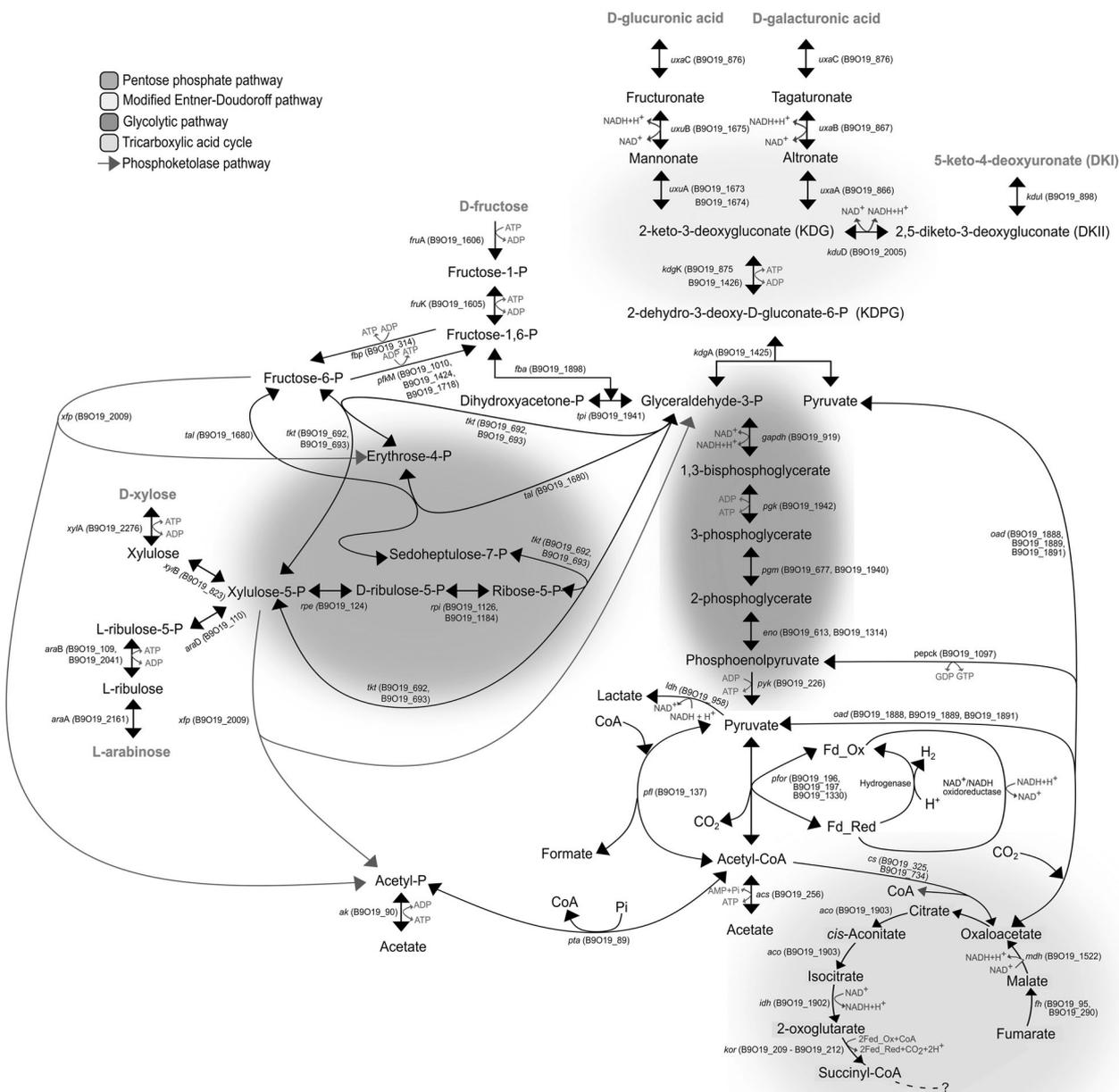
are indicated with numbers. Predicted protein domains are shown as rectangles, and coloured according to the domain category shown in the key. Where possible, additional domain details are labelled within domain rectangles. Remaining regions, with no known homology to existing protein domains, are represented by lines

knowledge, this is one of the largest number of SLH proteins reported from any obligate anaerobic bacterium (Supplementary Information S6). 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 42–318 kDa). All but one of the SLH proteins possess an N-terminal type I signal peptide; whereas, the sole exception (B9O19\_1870) contains a lipoprotein type II signal peptide. CAZymes attached to the bacterial surface via SLH modules have been previously implicated in the attachment and deconstruction of lignocellulose biomass in numerous *Firmicutes* organisms isolated from environmental samples [30–40], although to our knowledge, this mechanism is very rarely observed in the human gut, and *M. pectinilyticus* is the first isolate to use it in the strict context of pectin degradation. While some of the *M. pectinilyticus* SLH proteins have similar domain architectures to the structural protein components of the S-layer in *Bacillus anthracis* [41], SLH proteins containing CAZyme domains, cohesin- and dockerin-like domains, Fn3, and bacterial immunoglobulin (Big)-like domains are presumably involved in carbohydrate degradation [32, 42–44]. Using BlastP, at least 2 SLH protein sequences of *M. pectinilyticus* have been found in the NIH human stool metagenome databases generated from 12 subjects (out of 85) living in the US (~100% continuous sequence matches along >100 amino acids;  $e$ -value <1e–5), affirming that these SLH proteins are abundant within the HGM (Supplementary Information S7).

### Pectin metabolism and organization of pectin-gene clusters

We mapped the *M. pectinilyticus* genes to the KEGG metabolic pathways. The presence of the KEGG metabolic pathways for utilizing D-GalpA, 5-keto-4-deoxyuronate (DKI), D-glucuronic acid, D-Xylp, and L-Araf supports that monomeric derivatives of pectin and hemicellulose are the primary carbon sources used by *M. pectinilyticus* (Fig. 5). Fructose is the only non-PCW sugar which *M. pectinilyticus* utilizes, and for which the strain possesses its only phosphotransferase system (PTS) for the uptake across the cell membrane. Next we looked at the location of some of these genes within the *M. pectinilyticus* genome. While others are present in the genome as isolated loci, some genes encoding putative pectin-degrading CAZymes are located adjacent to the genes coding for SLH proteins (Fig. 6a, c, d) or as part of the two notable ‘pectin clusters’ consisting of distinct genomic regions of pectin-utilizing functions (Fig. 6b, e).

The two most notable pectin clusters were observed between B9O19\_864 and B9O19\_877, and between B9O19\_2005 and B9O19\_2015. The former spans a ~22 kb genomic segment and contains genes coding for CAZymes (GH28, GH43, PL1, and CE8); GalpA-metabolising enzymes (*uxaA*, *uxaB*, and *uxaC*); KDG kinase (*kdgK*); 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 (*kduD*); a phosphoketolase (*xfp*); a GntR type of transcription regulator; a biotin transporter; and proteins of ambiguous functions. Phosphoketolases are thiamine diphosphate-dependent enzymes responsible for two phosphoroclastic reactions: (1) conversion of xylulose-5-P (X5P) into glyceraldehyde-3-P (G3P) and acetyl-P, and (2) conversion of fructose-6-phosphate (F6P) to erythrose-4-P and acetyl-P. The position of the *xfp* gene within a pectin cluster is intriguing since X5P is a metabolic derivative of xylose and arabinose – the latter a predominant pentose sugar of pectin that is utilized by *M. pectinilyticus*. Notably present in *Bifidobacteria* [45, 46], *Lactobacillales* [47, 48], and *Clostridium acetobutylicum* [49, 50], mono- or bifunctional phosphoketolases have been found to shunt X5P and/or F6P through a catabolic alternative of the pentose-phosphate pathway (PPP) by bypassing carbon flux through the lower portion of glycolysis, thus avoiding a complete oxidation of carbon to CO<sub>2</sub>. As a result, the phosphoketolase pathway (PKP) yields fewer ATP than PPP [49], but also has lesser needs for NAD<sup>+</sup> regeneration as it adopts an enzymatic ‘shortcut’ through the conversion of X5P to acetate [49, 51]. Consistent with the traditional view of the pentose sugar metabolism in *Clostridia* [51], *M. pectinilyticus* also possesses enzymes of the PPP, including transketolase, transaldolase, and epimerase. Theoretically, more than one route of pentose sugar metabolism may exist in *M. pectinilyticus*, allowing greater metabolic flexibility. The existence of pectin clusters suggests that gene responses may be regulated at a transcriptional level to synchronize the catabolic and metabolic processes of pectin degradation. Further studies are necessary to conclude if *M. pectinilyticus* uses general regulatory mechanisms such as catabolite repression, the two-component systems, or alternative  $\sigma$ -factors [52] we know to exist within the genome (data not shown) to sense and react to the presence pectic polysaccharides in its extracellular surroundings. Using the ABC transporter database [53] we identified three ABC-related transporters (B9O19\_98–B9O19\_100, B9O19\_870–B9O19\_872, and B9O19\_1089) presumably involved in carbohydrate-binding and transport. An additional non-ABC type



**Fig. 5** 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 Supplementary Information S8. Question mark indicates missing genes (*scs* and *sdh*) in an otherwise complete TCA cycle

extracellular transporter (B9019\_1628) is present adjacent to genes coding for CE8 domain-containing SLH proteins, possibly suggesting its role in transporting pectic PDPs across the cell wall.

**Comparative proteomic profiling**

To examine the possible role of SLH proteins in pectin degradation, we used isobaric tags for relative and absolute

quantification (iTRAQ) technology to preliminarily measure the differential protein production in *M. pectinilyticus* in response to pectins from apple and kiwifruit, compared to the cells grown using the simple sugar fructose. Using three biological replicates, we identified 159 proteins that were commonly detected during the growth of *M. pectinilyticus* on both types of pectins (Fig. 7, extended data in Supplementary Information S9). Several non-catalytic (B9019\_52, B9019\_611, and B9019\_1156) and CE8 domain-



B9O19\_912, B9O19\_914, B9O19\_1225, B9O19\_1337, B9O19\_1578, B9O19\_1595, and B9O19\_1626), and the pattern of expression was not always consistent with both types of pectins. While the exact functions of SLH proteins are currently not understood, it is possible that *M. pectinilyticus* has varying anchoring requirements for SLH proteins to facilitate the binding and degradation of different pectin structures. 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 downregulated in pectin-grown cells. Despite the abundant PL sequences identified in the *M. pectinilyticus* genome, our proteome data did not show upregulation of PLs in response to pectins. It is possible that either these PLs are unrelated to pectin degradation, or CAZymes that are non-covalently attached to the cell wall or lack the cell wall-retention domains such as SLH modules became lost from the whole-cell proteome. The SEC analysis results showed that the degradation of high molecular weight pectic materials occurs at an early stage of growth (48 h), suggesting PLs may be highly expressed during the log/exponential phase. However, due to the difficulties of harvesting a sufficient mass of slowly growing cells (~0.3 OD<sub>595</sub> reached after 1-week growth), only cultures that had reached stationary phase were used in this study.

## Discussion

Until recently, most research on colonic pectin degradation was carried out using non-specialist bacteria. To our knowledge, *M. pectinilyticus* is the first human gut bacterium to show the functional specialization for pectin degradation and utilization. Notably, it is a member of the *Ruminococcaceae* as are other specialist human gut bacteria capable of initiating the degradation of resistant polysaccharides [54]. *M. pectinilyticus* takes a 'quality-not-quantity' approach by directing a significant proportion of its fibrolytic potential (48 CAZymes out of 108) to focus on pectin degradation. The *M. pectinilyticus* glycobiome is enriched with pectate/pectin lyases and pectin esterases, contrasting with the GH-rich CAZyme profiles more conventionally observed in human gut bacteria. Pectin lyases are capable of splitting pectin polymers independent of the degree of methylation. However, most microbial pectin lyases have been characterized from fungi [27, 55], with rare exceptions found in endophytic environmental bacteria such as *Bacillus/Paenibacillus* spp., *Pseudomonas* spp., and *Erwinia* spp [56]. The striking over-representation of PLs in the glycobiome of *M. pectinilyticus*, and their phylogenetic

similarity to enzymes from environmental bacteria suggest that *M. pectinilyticus* should be further examined for potential pectin lyase activities. Pectate lyases, on the other hand, are commonly produced by human gut bacteria for the enteric digestion of pectin [27]. PL family enzymes make up less than 2% of the total glycobiome in the HGM [57], but pectate lyases play key roles in mediating the extracellular degradation of complex pectin in the human colon. In previous studies conducted elsewhere [55, 58], the microbial degradation products of pectin were examined from the human faecal fluid, the colonic contents from rats, and the culture supernatant of *B. thetaiotaomicron*. In all cases, the resulting D-GalpA oligosaccharides followed the  $\beta$ -elimination pattern of pectate lyases by retaining the unsaturated double bonds at the non-reducing ends of molecules. Extensive methyl- and acetyl-esterification on the pectin backbone is a sterically limiting factor for pectate lyase activities. Pre-processing pectin with methyl- and acetyl-esterases significantly improves substrate accessibility by pectate lyases, and accelerates the enzyme reaction rates [59, 60]. Complementary actions of pectate lyases and pectin esterases would provide *M. pectinilyticus* the enzymatic flexibility to deal with the varying degrees of esterification observed in different types of the PCW pectins. Partial trimming of RG-I side chains is expected to occur by an unknown extracellular  $\beta$ -D-1,4-galactanase/galactosidase. A relatively small number of GH family enzymes may carry out the downstream processing of intracellular pectin oligomers. Only GH140 is specific to the RG-II degradation, suggesting that it is unlikely *M. pectinilyticus* can cope with the structural complexity of RG-II.

Inferring from previous studies with environmental bacteria, we predict that the pectin-binding and degradation in *M. pectinilyticus* are mediated through the SLH protein machinery, which to date has been mostly found in lignocellulolytic systems of various environmental isolates. While organisms with surface-associated glycan-binding proteins benefit from bringing glycan substrates in close contact with degradative enzymes [12], the exact mechanisms by which these SLH proteins may confer *M. pectinilyticus* with the ability to bind pectin are not yet understood. The SLH proteins are non-covalently tethered to the bacterial surface via the SLH modules which typically contain 50–60 amino acid residues that have binding affinity for the pyruvylated bacterial secondary cell wall [61]. Amongst *Firmicutes*, the SLH-mediated protein-anchoring motifs are used to tether a modular CAZyme structure consisting of various combinations of CAZymes, and non-catalytic CBMs, Fn3, and Big-like domains to the bacterial cell surface [30–40]. These non-catalytic regions of CAZymes presumably play supporting roles by functioning as CBM or a linker to swing out and orient the catalytic domains with the biomass substrates [32]. Little is known about

functionally ambiguous modules such as Fn3 and Big. Fn3 domains have previously demonstrated binding affinity to insoluble lignocellulose biomass [43] while significantly enhancing the activity of the appended CAZyme domains [44]. Big-like domains were found in bacterial proteins involved in cell–cell adhesion and extracellular carbohydrate hydrolysis [44]. The CAZyme-free SLH proteins may serve as scaffolding proteins for assembling multi-protein complex, or independent carbohydrate-binding protein units. For instance, the closest cultured relatives of *M. pectinilyticus*, such as *A. cellulolyticus* [62], *C. thermocellum*, and *Clostridium clariflavum* [63] use the SLH domain-containing scaffoldin proteins to attach the cellulosome assembly on the bacterial cell surface. Recently, novel types of lignocellulolytic degradation systems consisting of large uncharacterized SLH proteins were described in *Paenibacillus curdlanolyticus* and *Caldicellulosiruptor* spp. *P. curdlanolyticus* produces a 1450 kDa extracellular multi-enzyme complex consisting of 11 xylanase subunits that are assembled onto a non-catalytic SLH S1 protein [64]. In *Caldicellulosiruptor saccharolyticus*, the largest ORF (Csac\_2722) of the genome encodes an independent non-catalytic SLH protein showing binding affinity for Avicel cellulose [42]. The gene expressions of non-catalytic SLH proteins in *Caldicellulosiruptor kronotskyensis* were upregulated during growth on crystalline cellulose and switchgrass [32]. *Lactobacillus acidophilus*, which are human commensal *Firmicutes*, expresses an extracellular cell-attached GH13 pullulanase tethered through an S-layer associated protein (SLAP) domain [65]. The Pfam database contains several cell wall-anchoring motifs associated with S-layers, including SLH domains (PF00395), SLAP (PF03217), cell wall binding repeat 2 (CWB2; PF04122), S-layer like family, C-terminal region (S\_layer\_C; PF05125), and S-layer like family, N-terminal region (S\_layer\_N; PF05123). We showed that *M. pectinilyticus* produces numerous CAZyme-containing and non-catalytic SLH proteins, of which some constitute the largest ORFs of the genome and are upregulated in response to pectins, suggesting that these proteins may play roles in the mechanical and/or catalytic degradation of pectin. S-layer protein-mediated glycan degradation strategies are rarely observed in a human gut bacterium, differentiating *M. pectinilyticus* and *L. acidophilus* from the PUL system of *Bacteroides*, the Gram-positive PUL (gpPUL) system of the *Roseburia/Eubacterium rectale* group [66], or the cellulosome/amylosome organizations in *Ruminococcus champanellensis* [67, 68] and *Ruminococcus bromii* [69].

Consistent with the glycobiome specialized for pectin degradation, *M. pectinilyticus* possesses a narrow spectrum of metabolic pathways for utilizing uronic acids and pentose sugars of pectin/hemicellulose, and fructose. The *M. pectinilyticus* genome encodes all key enzymes of the variant

Entner-Doudoroff (ED) pathway to catabolize the uronic acid end-products of pectin-degrading lyases, D-GalpA and DKI. Uronic acids are low ATP-yielding substrates as their key metabolic intermediate KDPG is cleaved to form one pyruvate and one G3P through the upper portion of the ED pathway, with only the latter contributing towards ATP production through substrate-level phosphorylation [70–72]. In comparison, hexose sugars are split to form G3P and dihydroxyacetone-P (DHAP), from which twice as many net ATPs are yielded through the Embden–Meyerhof–Parnas (EMP) pathway. The ED pathway-dependent ATP production is therefore energy limiting. Anaerobic gut bacteria that must rely on glycolysis to generate ATP in the absence of aerobic respiration often lack the ED pathway [72, 73], and rarely use uronic acids as the main energy source [20]. However, although metabolic alternatives such as the ED pathway and PKP produce less ATPs than their canonical counterparts (EMP and PPP), 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 [71]. Furthermore, as uronic acids are already in a highly oxidized state, cells encounter less need to consume ATP to produce electron sinks (e.g. lactate, ethanol, and butyrate) to regenerate reducing equivalents, sparing more pyruvate for ATP production via acetate synthesis to compensate for the energy loss occurring in the ED pathway [73, 74]. In the case of PKP, there are fewer enzymatic steps (8 versus 13) involved in pentose metabolism than PPP [49], possibly reducing the amount of energy and biomass required for an equivalent 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.

The highly specialized glycobiome comprising extracellular secretion of pectate/pectin lyases, RG-lyases, and pectin esterases, together with the adhesion-based colonization of PCW suggests a niche specialization of *M. pectinilyticus* in proximity to plant particulate matter in the bowel, particularly the HG-dense region of the middle lamellae. However, *M. pectinilyticus* may not necessarily be an efficient utilizer of the resulting PDPs. As implied by the scarcity of intracellular pectin-degrading CAZymes and carbohydrate-associated transporters; the narrow substrate utilization spectrum; and the presence of PDP remnants in the culture supernatant, *M. pectinilyticus* likely shares a significant proportion of its PDPs with other members of HGM that are less adapted for pectin degradation.

Despite certain phenotypic similarities to environmental cellulolytic bacteria, *M. pectinilyticus* focuses on pectin,

potentially reflecting adaptation to the human gut environment. There are other indications that this gut bacterium may have evolved from an environmental ancestor, with a notable example being the enzymes of *M. pectinilyticus* often finding closest sequence homology with enzymes from environmental bacteria. The species-specific clustering of PL1 and CE8 family enzymes suggests that *M. pectinilyticus* may have evolved as a part of HGM for a sufficiently long time to allow a phylogenetic divergence from its environmental relatives. In *M. pectinilyticus*, the glycobiome evolution directed towards pectin utilization may have been a logical adaptation strategy to forage a readily available plant glycan from the human colon. *M. pectinilyticus* is currently the only representative of the genus *Monoglobus* and the only cultured species of a novel phylogenetic lineage of the *Ruminococcaceae*. The observation that uncultured bacteria found in other gut environments have similar 16S rRNA gene sequences suggests that additional cultures from this lineage exist and could be obtained by prioritizing microbial cultivation. The availability of more isolates would help 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 protein-based carbohydrate degradation strategy to target a broader range of substrates.

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 novelty. *M. pectinilyticus* likely occupies a spatial niche associated with the middle lamellae and PCW, and a functional niche as a primary pectin degrader, illustrating the functional compartmentalization and ecological diversity of the HGM community. The study begun here increases our understanding of microbial pectin degradation in the human colon by presenting a possibility outside of the currently existing paradigms of gut microbial polysaccharide degradation, and also showcases a specialist bacterium which has potentially co-evolved with its host to accommodate the pectin-rich diet of humans.

## Materials and methods

### Phylogenetic analysis

Multiple sequence alignments were performed on CLUSTALW using default parameters using MEGA7 software [75]. Phylogenetic tree was constructed using neighbour-joining method [76]. Bootstrap values were calculated using 2000 re-samplings 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 the GenBank database.

### Participant selection

Forty-four healthy subjects living in New Zealand were recruited for faecal sample donation and dietary intake assessment as part of a previous human intervention study [77]. Each participant completed four sets of 3-day diet records over a period of 10 weeks. Dietary analysis was conducted using FoodWorks version 8 software (Xyris Software Pty Ltd). The Australian database in FoodWorks was used (AusBrands and AusFoods 2015 data sources) to conduct nutrient intake and food group analysis. Participants were divided into low, moderate, and high habitual dietary fibre intake groups. The high dietary fibre intake cutoffs were chosen to reflect the New Zealand recommended dietary fibre intake which is >25 g/day for females and >30 g/day for males [78]. The average dietary fibre intake in New Zealand (17.5 g/day for females and 22.1 g/day for males) [79] was chosen as the low dietary fibre intake cutoffs. The amount of pectin consumed per day by each participant was calculated using an established food composition database [80].

### Quantitative PCR

Bacterial DNA was extracted from faecal samples using MoBio PowerLyzer<sup>®</sup> Powersoil DNA<sup>®</sup> isolation kit as per the manufacturer's instructions with minor amendments. Faecal samples (0.25 ± 0.025 g) were weighed into PowerLyzer<sup>®</sup> glass bead tubes. A FastPrep-24<sup>™</sup> 5G (MP Bio-medicals) was used to homogenise the samples at a speed of 5.5 m/s for four 90 s cycles with a 60 s break between each cycle. The DNA was eluted in 10 mM Tris (pH 8.0). NanoDrop 1000 spectrophotometry was used to quantify the DNA concentration. Primers MP1087F (5'-GAGCGC AACCCTTACTGTCA-3') and MP1581R (5'-CTCTTA CTCCGCTCTCCGC-3') were designed using NCBI Primer Blast [81]. Samples and standards were run in triplicate by absolute quantification on Roche Lightcycler<sup>®</sup> 480 real-time PCR instrument. Lightcycler<sup>®</sup> 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<sup>®</sup> 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 °C for 10 min; 45 run cycles of denaturation (95 °C for 10 s), annealing (60 °C for 20 s), and extension (72 °C for 20 s); one cycle of melting (95 °C for 30 s, 65 °C for 1 min,

followed by 65 °C to 95 °C at 0.1 °C increment per second with continuous fluorescence acquisition); and a cooling cycle at 40 °C. Results were analysed and visualized using Lightcycler® 480 software package (version 1.5). Log<sub>10</sub> concentration of *M. pectinilyticus* and the % abundance of *M. pectinilyticus* relative to the total bacterial concentration were plotted and participants were separated into *M. pectinilyticus*-positive and *M. pectinilyticus*-negative groups at a clear gap in the distribution at 0.01% abundance, 3.5 log<sub>10</sub> concentration. A chi-square test was performed to check whether the proportion of high fibre consumers was similar in the *M. pectinilyticus*-positive and *M. pectinilyticus*-negative groups (8/10 compared to 14/34). Intakes of dietary fibre and pectin were compared between the *M. pectinilyticus*-positive and *M. pectinilyticus*-negative groups using a non-parametric Mann–Whitney *U*-test. The *p*-value <0.05 was considered statistically significant.

### Genomic sequencing

Routine cultivation of *M. pectinilyticus* and genomic DNA extraction were carried out as described previously [22]. The genome of *M. pectinilyticus* 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 4000-fold genome coverage. The quality of raw sequencing data was assessed using FastQC [82]. Adaptor sequences present in paired-end reads were trimmed and quality-filtered using Trimmomatic at default settings [83]. 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 to 8 kb mate-pair reads and to select for sets of true mate pairs [84]. Hundred percent 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. Hundred percent of 36,608,300 total reads in 8 kb mate-pair library were classified as true mate-pairs and were used for additional contig extension. The de novo genome assembly was performed on SPAdes assembler using default parameters [24]. Scaffolds generated using an overlapping *K*-mer length of 77 were selected for further processing. Additional scaffold extension was carried out on SSPACE using the output data from SPAdes de novo assembly and trimmed sequencing data from 8 kb mate-pair library [85]. 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 the initial assembly, indicating the circular genome was fragmented at three sites.

### Primer walking

Primers were designed to hybridize at 200–300 bp upstream of the truncation sites on each scaffold. Long-range PCR was carried out to amplify the genome gap sequences using Phusion Green High-Fidelity DNA Polymerase (Thermo® Scientific). Fifty microliters of PCR reaction contained 10 µl of 5X 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 the manufacture were used: initial denaturation at 98 °C for 30 s; 35 cycles of denaturation at 98 °C for 10 s and annealing/extension at 72 °C for 10 min; final extension at 72 °C for 10 min; hold at 4 °C. Blunt-end PCR products were ligated into pCR™-BluntII-TOPO® vector using Zero Blunt® TOPO® PCR Cloning Kit (invitrogen). Six microliters 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. 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 software (version 10.0.3) with 65% similarity cost matrix, 12 gap open, and 3 extension penalties.

### Genome annotation

A preliminary annotation of the genome was carried out using Prokka [25]. The draft genome was exported in FASTA format. Open reading frames were predicted using Prodigal, a built-in software in Prokka annotation pipeline [86]. 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<sup>-6</sup> was used to annotate the putative gene products. Protein function prediction was carried out using InterProScan 5 [87]. For the construction of KEGG metabolic pathways, nucleotide sequences of the target enzyme from 10 to 15 species of the family *Ruminococcaceae* and family *Clostridiaceae* were manually downloaded from GenBank. The sequences were aligned using Geneious alignment (version 10.0.3). Aligned sequences were exported into Linux environment and converted into a HMMER database using *hmmbuild* function [88]. Using *hmmsearch* function, the database was queried against the draft genome of *M. pectinilyticus* to find the most significant match to the full target sequence below an *e*-value threshold of 10<sup>-5</sup>. Genes

assigned with enzyme functions were manually mapped into reference metabolic pathways stored in KEGG database [89]. In order to assign each protein-coding gene into an orthologous group, all available HMM models for the target taxonomic group (Bacteria) were manually downloaded from the eggNOG database [90]. Obtained sequences were concatenated into a HMMER database using *hmmcompress* function. Using *hmmsearch* function, all bacterial orthologous groups available in the eggNOG database were queried against the genome database of *M. pectinilyticus*. 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 [91]. TMHMM server was used to predict transmembrane helix domains in proteins [92]. tRNA sequences were identified using a combination of Prokka and tRNAscan-SE [93]. Genes coding for CAZymes were identified using the CAZy database. SLH domains were identified by querying the genome of *M. pectinilyticus* against dbCAN database [94] and then manually verifying the results by BLAST searching against GenBank. For this, default dbCAN parameters (alignment length >80 amino acids, use *e*-value <10<sup>-5</sup>, otherwise use *e*-value <10<sup>-3</sup>) were used. Each gene entry in the genome of *M. pectinilyticus* 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.

### 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 *M. pectinilyticus* is CP020991.

### Carbohydrate analysis

Pectins from apple (93854) and citrus peels (P9135) were purchased from Sigma. Arabinan from sugar beet (P-ARAB), arabinogalactan from larch wood (P-ARGAL), galactan from potato (P-GALPOT), and RG-I from potato (P-RHAM1) were purchased from Megazyme. Hundred microliters of clarified rumen fluid, 100 µl of 5% (w/v) apple or citrus pectins, and 100 µl of 1-week-old inoculum were added in triplicate to 1.7 ml mineral medium. Cultures were grown at 37 °C for 3 days with a constant shaking. Nine hundred microliters of samples was taken out at 48 h and 72 h of incubation. These samples were centrifuged at 12,000 × *g* for 10 min at room temperature. Polysaccharide substrates were dissolved in 0.1 M NaNO<sub>3</sub> (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<sub>3</sub> (0.5 mL/min, 60 °C) from three columns (TSK-Gel G5000PWXL, G4000PWXL and G3000PWXL, 300 × 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<sub>3</sub> (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). For detection of oligosaccharide, spent culture media (2 µL) were injected and eluted with a simultaneous gradient of NaOH and sodium acetate (1 mL min<sup>-1</sup>) from a CarboPac PA-100 (4 × 250 mm) column. For detection of monosaccharides, spent culture media (2 µL) were injected and eluted with a simultaneous gradient of NaOH and sodium acetate (1 mL min<sup>-1</sup>) from a CarboPac PA-1 (4 × 250 mm) column. The sugars were identified from their elution times relative to standard sugar mixes.

### Enzyme phylogenetic analysis

Reference PL1 and CE8 sequences were obtained as the results of BlastP queries against NCBI protein database to find the closest matches to each of PL1 and CE8 sequence from *M. pectinilyticus*. The catalytic domains within sequences were identified using the dbCAN database [94], and PL1 and CE8 domain sequences were manually extracted using Geneious software (version 10.0.3). Extracted sequences were aligned by ClustalW [95] and used to construct a maximum-likelihood phylogenetic tree using MEGA7 software [75]. Bootstrap values were calculated using 1000 re-samplings to evaluate the support of tree topology.

### Metagenome mining

Full-length protein sequences of all SLH proteins from the *M. pectinilyticus* were manually extracted from the genome database. These SLH protein sequences were used as query sequences to search the metagenome databases constructed from the stool samples of 85 donors living in the US, collected as part of the Human Microbiome Project (HMP) initiated by National Institutes of Health (NIH). 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 [96]. A built-in Blast Genome function of IMG/M system was used to BlastP sequences with an *e*-value threshold of  $1e^{-5}$ . A stringent cutoff was used to identify true positives which showed ~100% identical amino acid residue matches along the same positions over a relatively long (>100 amino acids) length of aligned protein sequences.

### Preparing samples for iTRAQ quantitative protein analysis

In three-independent experiments, *M. pectinilyticus* was grown using three different types of substrates: 0.5% (w/v) D-fructose, 0.2% (w/v) apple pectin, and 0.2% (w/v) kiwifruit pectin. Cultures were incubated at 37 °C with a constant rotary shaking until stationary phase was reached. Fully grown cultures were centrifuged at  $12,000 \times g$  for 10 min to collect cell pellets. The total protein content of each pellet sample was assayed using the EZQ Protein Quantitation Kit (Life Technologies). Aliquots containing 25 µg of protein were taken for each sample and diluted to 100 µl with 50 mM ammonium bicarbonate. Samples were reduced by addition of DTT to 10 mM final concentration and incubated at 56 °C for 10 min. Samples were cooled and alkylated with 50 mM iodoacetamide (GE Healthcare) in the dark at room temperature for 30 min, and digested with 1 µg of sequencing grade trypsin (Promega) in a chilled microwave (CEM) at 40 °C for 2 h at 15 W power, followed by overnight incubation at 37 °C. Digests were acidified with formic acid and desalted on Oasis HLB SPE cartridges (Waters) and dried down in a vacuum centrifuge. Samples were reconstituted with 30 µl of 0.5 M TEAB (Sigma) and labelled with 4-plex iTRAQ labels (Sciex) as per the manufacturer's instructions. Pools were prepared using equal amounts of each sample, and were then concentrated in a vacuum centrifuge to reduce solvent content, then desalted and concentrated to 40 µl as above.

### LC-MS/MS and database search

Samples were injected onto a  $0.3 \times 10$  mm trap column packed with Reprosil C18 media and desalted for 5 min at 2 µl/min before being separated on a  $0.075 \times 200$  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 to 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 s. 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), 116 (fructose), and 117 (kiwifruit) samples were determined in log space using 116 (fructose) 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. The *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 analysis.

### Proteomics statistics

Reliability of protein fold-changes ratios were ensured by using those proteins represented by  $\geq 3$  unique peptides with >95% confidence intervals. 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. Using these criteria, 920 proteins across three biological replicates were selected. Contaminant protein entries such as human keratin introduced through sample handling, and the traces of endogenous plant proteins from apple and kiwifruit pectin were further removed manually. Proteins represented by at least three peptides with 95% confidence, FDR-corrected *p*-value <0.05, and fold-changes of  $\geq 1.20$  or  $\leq 0.83$  relative to the fructose control were regarded as differentially expressed 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  $\geq 1.65$  indicates that the average protein expression 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.

**Acknowledgements** We thank staff at Macrogen (South Korea) and Massey Genome Service (Palmerston North, New Zealand) for sequencing services; and Martin Middleditch and Leo Payne (University of Auckland, New Zealand) for iTRAQ proteomics analysis. We are grateful to Louise Brough (Massey Institute of Food Science and Technology, School of Food and Nutrition, Massey University, Palmerston North, New Zealand), Chrissie Butts (Department of Food, Nutrition and Health, Plant and Food Research Limited, Palmerston North, New Zealand), Rinki Murphy (Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand), and Jane Coad (Massey Institute of Food Science and Technology, School of Food and Nutrition, Massey University, Palmerston North, New Zealand) for assistance and supervision of the clinical study conducted by GRL during her PhD studies. We thank Peter Janssen at AgResearch Ltd (Grasslands Research Centre, Palmerston North) for providing rumen fluid. We are grateful to David Brummell (Plant and Food Research Limited, Palmerston North) for his valuable advice on pectin structure. This work has been carried out with the financial support of Ministry of Business, Innovation and Employment of New Zealand ('Foods for Health at Different Life Stages' C11X1312).

**Author contributions** CCK performed all experiments and data analysis related to genome sequencing of *M. pectinilyticus*. CCK conducted 16S rRNA phylogenetic analysis and quantitative PCR detection of *M. pectinilyticus* in human faecal samples. GRL designed the clinical study, obtained ethics approval, recruited donors, conducted the study, collected dietary records, and faecal samples, and prepared DNA samples. GL performed the dietary intake analysis. CCK prepared fermented pectin samples, and IMS and TJB assisted with SEC and data analysis. CCK prepared samples for iTRAQ protein quantification, and analysed the data. CCK performed all enzyme phylogeny and metagenome analysis. DH assisted with all statistical analysis used in this study. CAZyme domains were identified by BH using the CAZy database. This study was conceived and supervised by IMS, GWT, WJK, DIR, ZJ, and MLP, and directed by CCK and DIR. All authors contributed to research designing and planning. CCK and DIR wrote the article, with contributions from all other authors.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Daher FB, Braybrook SA. How to let go: pectin and plant cell adhesion. *Front Plant Sci.* 2015; <https://doi.org/10.3389/fpls.2015.00523>
- Mohnen D. Pectin structure and biosynthesis. *Curr Opin Plant Biol.* 2008;11:266–77.
- Ridley BL, O'Neill MA, Mohnen D. Pectins: structure, biosynthesis, and oligogalacturonide-related signalling. *Phytochemistry.* 2001;57:929–67.
- Caffall KH, Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res.* 2009;344:1879–1900.
- Zykwinska AW, Ralet MJ, Garnier CD, Thibault JJ. Evidence for in vitro binding of pectin side chains to cellulose. *Plant Physiol.* 2005;139:397–407.
- Yapo BM. Rhamnogalacturonan-I: a structurally puzzling and functionally versatile polysaccharide from plant cell walls and mucilages. *Polym Rev.* 2011;51:391–413.
- Yapo BM. Pectin rhamnogalacturonan II: on the "small stem with four branches" in the primary cell walls of plants. *Int J Carbohydr Chem.* 2011;2011:1–11.
- Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature.* 2017;544:65–70.
- O'Neill MA, Ishii T, Albersheim P, Darvill AG. Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol.* 2004; 55:109–39.
- Cheng KJ, Dinsdale D, Stewart CS. Maceration of clover and grass leaves by *Lachnospira multiparus*. *Appl Environ Microbiol.* 1979;38:723–9.
- Chung D, Pattathil S, Biswal AK, Hahn MG, Mohnen D, Westpheling J. Deletion of a gene cluster encoding pectin degrading enzymes in *Caldicellulosiruptor bescii* reveals an important role for pectin in plant biomass recalcitrance. *Biotechnol Biofuels.* 2014;7:147.
- Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. *Nat Microbiol.* 2017; <https://doi.org/10.1038/s41564-017-0079-1>
- Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. *J Biol Chem.* 2009;284:24673–7.
- Comstock LE. Importance of glycans to the host-*Bacteroides* mutualism in the mammalian intestine. *Cell Host Microbes.* 2009;5:522–6.
- Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol.* 2012;10:323–35.
- Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Nicholas A, et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature.* 2015;517:165–9.
- Rogowski A, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Lowe, EC, et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun.* 2015; <https://doi.org/10.1038/ncomms8481>
- Leth ML, Ejby M, Workman C, Ewald DA, Pedersen SS, Sternberg C, et al. Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat Microbiol.* 2018;3:570–80.
- Chung WSF, Meijerink M, Zeuner B, Holck J, Louis P, Meyer A, et al. Prebiotic potential of pectin and pectic oligosaccharides to promote anti-inflammatory commensal bacteria in the human colon. *FEMS Microbiol Ecol.* 2017; <https://doi.org/10.1093/femsec/fix127>
- Salyers AA, West SE, Vercellotti JR, Wilkins TD. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl Environ Microbiol.* 1977;34:529–33.
- Lopez-Siles M, Khan TM, Duncan SH, Harmsen HJM, Garcia-Gil LJ, Flint HJ. Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl Environ Microbiol.* 2012;78:420–8.
- Kim CC, Kelly WJ, Patchett ML, Tannock GW, Jordens Z, Stoklosinski HM, et al. *Monoglobus pectinilyticus* gen. nov., sp. nov., a pectinolytic bacterium isolated from human faeces. *Int J Syst Evol Microbiol.* 2017; <https://doi.org/10.1099/ijsem.0.002395>

23. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol*. 2008;6:121–31.
24. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
25. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
26. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2013;42:D490–D495.
27. Hugouvieux-Cotte-Pattat N, Condemine G, Shevchik VE. Bacterial pectate lyases, structural and functional diversity. *Environ Microbiol Rep*. 2014; <https://doi.org/10.1111/1758-2229.12166>
28. Cid M, Pedersen HL, Kaneko S, Coutinho PM, Henrissat B, Willats WGT, et al. Recognition of the helical structure of  $\beta$ -1,4-galactan by a new family of carbohydrate-binding modules. *J Biol Chem*. 2010;285:25999–6009.
29. Peer A, Smith SP, Bayer EA, Lamed R, Borovok I. Non-cellulosomal cohesin- and dockerin-like modules in the three domains of life. *FEMS Microbiol Lett*. 2009;291:1–16.
30. Adelsberger H, Hertel C, Glawischnig E, Zverlov VV, Schwarz WH. Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the in vivo system from recombinant enzymes. *Microbiology*. 2004;150:2257–66.
31. Ali E, Araki R, Zhao G, Sakka M, Karita S, Kimura T, et al. Functions of family 22 carbohydrate-binding modules in *Clostridium josui* Xyn10A. *Biosci Biotechnol Biochem*. 2005;69:2389–94.
32. Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, et al. Multi-domain, surface layer associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. *J Biol Chem*. 2016; <https://doi.org/10.1074/jbc.M115.707810>
33. Fuchs KP, Zverlov VV, Velikodvorskaya GA, Lottspeich F, Schwarz WH. Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo- $\beta$ -1,3-glucanase bound to the outer cell surface. *Microbiology*. 2003;149:1021–31.
34. Han YJ, Agarwal V, Dodd D, Kim J, Bae B, Mackie RI, et al. Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *J Biol Chem*. 2012;287:34946–60.
35. Hung KS, Liu SM, Fang TY, Tzou WS, Lin FP, Sun KH, et al. Characterization of a salt-tolerant xylanase from *Thermoanaerobacterium saccharolyticum* NTOU1. *Biotechnol Lett*. 2011;33:1441–7.
36. Huang X, Li Z, Du C, Wang J, Li S. Improved expression and characterization of a multidomain xylanase from *Thermoanaerobacterium aotearoense* SCUT27 in *Bacillus subtilis*. *J Agric Food Chem*. 2015;63:6430–9.
37. Itoh T, Sugimoto I, Hibi T, Suzuki F, Matsuo K, Fujii Y, et al. Overexpression, purification, and characterization of *Paenibacillus* cell surface-expressed chitinase ChiW with two catalytic domains. *Biosci Biotechnol Biochem*. 2014;78:624–34.
38. Lee SP, Morikawa M, Takagi M, Imanaka T. Cloning of the *aapT* gene and characterization of its product,  $\alpha$ -amylase-pullulanase (AapT), from Thermophilic and alkaliphilic *Bacillus* sp strain Xal601. *Appl Environ Microbiol*. 1994;60:3764–73.
39. St John FJ, Preston JF, Pozharski E. Novel structural features of xylanase A1 from *Paenibacillus* sp JDR-2. 2012. *J Struct Biol*. 2012;180:303–11.
40. Waeonukul R, Pason P, Kyu KL, Sakka K, Kosugi A, Mori Y, et al. Cloning, sequencing, and expression of the gene encoding a multidomain endo- $\beta$ -1,4-xylanase from *Paenibacillus curdlanolyticus* B-6, and characterization of the recombinant enzyme. *J Microbiol Biotechnol*. 2009;19:277–85.
41. Fagan RP, Fairweather NF. Biogenesis and functions of bacterial S-layers. *Nat Rev Microbiol*. 2014;12:211–22.
42. Ozdemir I, Blumer-Schuette SE, Kelly RM. S-layer homology domain proteins Csac\_0678 and Csac\_2722 are implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl Environ Microbiol*. 2012;78:768–77.
43. Fraser JS, Yu Z, Maxwell KL, Davidson AR. Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol*. 2006;359:496–507.
44. Kataeva IA, Seidel RD, Shah A, West LT, Li XL, Ljungdahl LG. The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CdhA promotes hydrolysis of cellulose by modifying its surface. *Appl Environ Microbiol*. 2002;68:4294–4300.
45. Meile L, Rohr LM, Geissmann TA, Herensperger M, Teuber M. Characterization of the D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase gene (*xfp*) from *Bifidobacterium lactis*. *J Bacteriol*. 2002;183:2929–36.
46. Yin X, Chambers JR, Barlow K, Park AS, Wheatcroft R. 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 Microbiol Lett*. 2005;246:251–7.
47. Tanaka K, Komiyama A, Sonomoto K, Ishizaki A, Hall SJ, Stanbury PF. Two different pathways for D-xylulose 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. *Appl Microbiol Biotechnol*. 2002;60:160–7.
48. Okano K, Yoshida S, Yamada R, Tanaka T, Ogino C, Fukuda H, et al. 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*. *Appl Environ Microbiol*. 2009;75:7858–61.
49. Liu L, Zhang L, Tang W, Gu Y, Hua Q, Yang S, et al. Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by  $^{13}\text{C}$  metabolic flux analysis. *J Bacteriol*. 2012;194:5413–22.
50. Sund CJ, Liu S, Germane KL, Servinsky MD, Gerlach ES, Hurley MM. Phosphoketolase flux in *Clostridium acetobutylicum* during growth on L-arabinose. *Microbiology*. 2015;161:430–40.
51. Servinsky MD, Germane KL, Liu S, Kiel JT, Clark AM, Shankar J, et al. Arabinose is metabolized via a phosphoketolase pathway in *Clostridium acetobutylicum* ATCC 824. *Ind Microbiol Biotechnol*. 2012;39:1859–67.
52. Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, Bayer EA, et al. *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc Natl Acad Sci USA*. 2010;107:18646–51.
53. Fichant G, Basse MJ, Quentin Y. ABCdb: an online resource for ABC transporter repertoires from sequenced archaeal and bacterial genomes. *FEMS Microbiol Lett*. 2006;256:333–9.
54. Mukhopadhyaya I, Morais S, Laverde-Gomez J, Sheridan PO, Walker AW, Kelly W, et al. Sporulation capability and amylose conserved among diverse human colonic and rumen isolates of the keystone starch-degrader *Ruminococcus bromii*. *Environ Microbiol*. 2017; <https://doi.org/10.1111/1462-2920.14000>
55. Dongowski G, Lorenz A, Anger H. Degradation of pectins with different degrees of esterification by Bacteroides thetaiotaomicron isolated from human gut flora. *Appl Environ Microbiol*. 2000;66:1321–7.

56. Yadav S, Yadav PK, Yadav D, Yadav KDS. Pectin lyase: a review. *Process Biochem.* 2009;44:1–10.
57. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. 2013. *Nat Rev Microbiol.* 2013; 11:497–504.
58. Dongowski G, Lorenz A, Proll J. The degree of methylation influences the degradation of pectin in the intestinal tract of rats and in vitro. *J Nutr.* 2002;132:1935–44.
59. Shevchik VE, Hugouvieux-Cotte-Pattat N. Identification of a bacterial pectin acetyl esterase in *Erwinia chrysanthemi* 3937. *Mol Microbiol.* 1997;24:1285–301.
60. Shevchik VE, Hugouvieux-Cotte-Pattat N. PaeX, a second pectin acetyl esterase of *Erwinia chrysanthemi* 3937. *J Bacteriol.* 2003;185:3091–3100.
61. Sara M. Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in Gram-positive bacteria? *Trends Microbiol.* 2001;9:47–49.
62. Xu Q, Gao W, Ding SY, Kenig R, Shoham Y, Bayer EA, et al. The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. *J Bacteriol.* 2003;185:4548–57.
63. Artzi L, Morag E, Barak Y, Lamed R, Bayer E. *Clostridium clariflavum*: key cellulosome players are revealed by proteomic analysis. *mBio.* 2015;6:e00411–15.
64. Ratanakhanokchai K, Waeonukul R, Pason P, Tachaapaikoon C, Kyu KL, Sakka K, et al. *Paenibacillus curdolanolyticus* strain B-6 multienzyme complex: a novel system for biomass utilization. In: Matovic MD editor. *Biomass Now-Cultivation and Utilization*; 2013. <https://doi.org/10.5772/51820>
65. Møller MS, Goh YJ, Rasmussen KB, Cypryk W, Celebioglu HU, Klaenhammer TR, et al. An extracellular cell-attached pullulanase confers branched  $\alpha$ -glucan utilization in human gut *Lactobacillus acidophilus*. *Appl Environ Microbiol.* 2017;83:e00402–17.
66. Sheridan PO, Martin JC, Lawley TD, Browne H, Harris HM, Bernalier-Donadille A, et al. Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microbial Genom.* 2016; <https://doi.org/10.1099/mgen.0.000043>
67. Morais S, David YB, Bensoussan L, Duncan SH, Koropatkin NM, Martens EC, et al. Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, *Ruminococcus champanellensis*, reveals a fine-tuned system for cohesin-dockerin recognition. *Environ Microbiol.* 2015; <https://doi.org/10.1111/1462-2920.13047>
68. Cann I, Bernardi RC, Mackie RI. Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm. *Environ Microbiol.* 2016; <https://doi.org/10.1111/1462-2920.13152>
69. Ze X, David YB, Laverde-Gomez JA, Dassa B, Sheridan PO, Duncan SH, et al. Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic *Firmicutes* bacterium *Ruminococcus bromii*. *mBio.* 2015;6:e01058–15.
70. Condemine G, Robertbaudouy J. 2-keto-3-deoxygluconate transport system in *Erwinia chrysanthemi*. *J Bacteriol.* 1987;169:1972–8.
71. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a trade-off between energy yield and protein cost. *Proc Natl Acad Sci USA.* 2013;110:10039–44.
72. Richard P, Hilditch S. D-galacturonic acid catabolism in microorganisms and its biotechnological relevance. *Appl Microbiol Biotechnol.* 2009;82:597–604.
73. Macfarlane S, Macfarlane GT. Regulation of short-chain fatty acid production. *Proc Nutr Soc.* 2003;62:67–72.
74. Wolfe AJ. The acetate switch. *Microbiol Mol Biol Rev.* 2005;69:12–50.
75. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
76. Saitou N, Nei M. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4:406–25.
77. Healey G, Murphy R, Butts C, Brough L, Whelan K, Coad J. Habitual dietary fibre intake influences gut microbiota response to an inulin-type fructan prebiotic: a randomised, double-blind, placebo-controlled, cross-over, human intervention study. *Br J Nutr.* 2018; <https://doi.org/10.1017/S0007114517003440>
78. Australian Government, National Health and Medical Research Council. Nutrient reference values for Australia and New Zealand. 2006; Available online at: <https://www.nhmrc.gov.au/guidelines-publications/n35-n36-n37>
79. New Zealand Government, Ministry of Health. A focus on nutrition: key findings from the 2008/09 NZ adult nutrition survey. 2011; Available online at: <https://www.health.govt.nz/publication/focus-nutrition-key-findings-2008-09-nz-adult-nutrition-survey>
80. Marlett JA, Cheung TF. Database and quick methods of assessing typical dietary fiber intakes using data for 228 commonly consumed foods. *J Am Diet Assoc.* 1997;97:1139–51.
81. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinforma.* 2012;13:134.
82. Andrews S FastQC: a quality control tool for high throughput sequence data. 2010; Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
83. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
84. O'Connell J, Schulz-Trieglaff O, Carlson E, Hims MM, Gormley NA, Cox AJ. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics.* 2015;31:2035–7.
85. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 2011;27:578–9.
86. Hyatt D, Chen G, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 2011;11:119.
87. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–40.
88. Eddy S, Wheeler T. HMMER user's guide version 3.1b1. 2013; Available online at: <http://eddylab.org/software/hmmer3/3.1b2/Userguide.pdf>
89. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44:D457–D462.
90. Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, et al. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* 2008;36:D250–D254.
91. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8:785–6.
92. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
93. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25:955–64.
94. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y, et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2012;40:W445–W451.

95. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
96. Chen IA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, et al. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res*. 2017;45:D507–D516.
97. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res*. 2008;36:W181–W184.
98. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, et al. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res*. 2005;33:W455–W459.