Complete Genome Sequence of the Polysaccharide-Degrading Rumen Bacterium *Pseudobutyrivibrio xylanivorans* MA3014.

Nikola Palevich[1,2,*], Paul H. Maclean[1], William J. Kelly[3], Sinead C. Leahy[1], Jasna Rakonjac[2] and Graeme T. Attwood[1].

[1]AgResearch Limited, Grasslands Research Centre, Palmerston North, New Zealand.

[2]Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand.

[3]Donvis Limited, Palmerston North, New Zealand.

[*]Corresponding author: E-mail: nik.palevich@agresearch.co.nz.

Keywords

Pseudobutyrivibrio xylanivorans, genome, rumen, bacteria, polysaccharide, xylan, starch.

Abstract

Ruminants are essential for maintaining the global population and managing greenhouse gas emissions. In the rumen, bacterial species belonging to the genera rumen *Butyrivibrio* and *Pseudobutyrivibrio* constitute the core bacterial rumen microbiome and are important degraders of plant-derived complex polysaccharides. *Pseudobutyrivibrio xylanivorans* MA3014 was selected for genome sequencing in order to examine its ability to breakdown and utilize plant polysaccharides. The complete genome sequence of MA3014 is 3.58 Mb, consists of three replicons (a chromosome, chromid and plasmid), has an overall G+C content of 39.6% and encodes 3,265 putative protein-coding genes (PCGs). Comparative pan-genomics of all cultivated and currently available *P. xylanivorans* genomes has revealed highly open genomes and a strong correlation of orthologous genes within this species of rumen bacteria. MA3014 is

metabolically versatile and capable of utilizing a range of simple mono- or oligosaccharides to complex plant polysaccharides such as pectins, mannans, starch and hemicelluloses for growth, with lactate, butyrate and formate as the principal fermentation end-products. The genes encoding these metabolic pathways have been identified and MA3014 is predicted to encode an extensive repertoire of Carbohydrate-Active enZYmes (CAZymes) with 80 Glycoside Hydrolases (GHs), 28 Carbohydrate Esterases (CEs) and 51 Glycosyl Transferases (GTs), that suggest its role as an initiator of primary solubilization of plant matter in the rumen.

Introduction

Butyrivibrio and *Pseudobutyrivibrio* represent the most commonly isolated butyrate-producing anaerobic rumen bacteria (Henderson, et al. 2015), and are among the small number of rumen genera capable of utilizing the complex plant structural polysaccharide xylan (Bryant and Small 1956; Hungate 1966). *Pseudobutyrivibrio* [family Lachnospiraceae, order Clostridiales] are anaerobic, monotrichous, butyrate-producing, curved rods and have been isolated from the gastrointestinal tracts of various ruminants, monogastric animals and humans (Kopecný, et al. 2003; Willems and Collins 2009). The *Butyrivibrio* and *Pseudobutyrivibrio* genera originally consisted of only one species, *Butyrivibrio fibrisolvens* (Bryant and Small 1956). In addition to phenotypic characterisations (Hazlewood, et al. 1986; Shane, et al. 1969), studies have utilized DNA-DNA hybridization (Mannarelli 1988; Mannarelli, et al. 1990), 16*S* rRNA gene sequencing (Forster, et al. 1996; Willems, et al. 1996) and 16*S* rRNA-based hybridization probes (Forster, et al. 1997), to differentiate these organisms. To accommodate the observed diversity amongst the newly discovered bacterial strains, a new genus, *Pseudobutyrivibrio*, was described in which only *P. ruminis* and *P. xylanivorans* species are currently recognized (Kopecný, et al. 2003; Van Gylswyk, et al. 1996). *P. xylanivorans* are common anaerobic rumen bacteria found in domestic

and wild ruminants and the type strain is Mz 5^T (DSM 14809) (Henderson, et al. 2015; Kopecný, et al. 2003). *P. xylanivorans* Mz 5^T is non-proteolytic but is able to utilize xylan or hemicellulose and various oligo- and monosaccharides as substrates for growth (Zorec, et al. 2000). Gaining an insight into the role of these microbial primary plant polysaccharide fermenters is important for understanding rumen function. Here we present the complete genome sequence of *P. xylanivorans* MA3014, a strain isolated from a New Zealand pasture-grazed dairy cow (Seshadri, et al. 2018), and describe its comparison with other representative *P. xylanivorans* genomes.

Materials and Methods

Growth Conditions and Fermentation End Product Analysis

P. xylanivorans MA3014 was isolated from the rumen contents of fistulated Friesian dairy cattle and sequenced (Noel 2013; Seshadri, et al. 2018). MA3014 was grown in RM02 medium (Kenters, et al. 2011) with 10 mM glucose and 0.1% yeast extract but without rumen fluid and culture purity was confirmed by Gram stain. The morphological features of MA3014 cells were determined by both scanning (SEM) and transmission (TEM) electron microscopy of cells grown on RM02 medium alone or with the addition of neutral detergent fraction (NDF) of plant material as previously described (Palevich, et al. 2017; Palevich, et al. 2018).

Growth on soluble substrates was assessed as an increase in culture density OD_{600nm} compared to cultures without carbon source added (all tested at 0.5% w/v final concentration), whereas total VFA production was used as an indicator of substrate utilization and growth for insoluble polymers (Supplementary Table S3). VFA production was determined from triplicate broth cultures grown overnight in RM02 medium with cellobiose as substrate and analysed for formate, acetate, propionate, n-butyrate, iso-valerate and lactate on a HP 6890 series GC

(Hewlett Packard) with 2-ethylbutyric acid (Sigma-Aldrich, St. Louis, MO, USA) as the internal standard. To derivatize formic, lactic and succinic acids, samples were mixed with HCl ACS reagent (Sigma-Aldrich, St. Louis, MO, USA) and diethyl ether, with the addition of N-methyl-N-t-butyldimethylsilyltri-fluoroacetamide (MTBSTFA) (Sigma-Aldrich, St. Louis, MO, USA) (Richardson, et al. 1989).

Preparation of Genomic DNA for Whole-Genome Sequencing

Genomic DNA was extracted from freshly grown cells by a modification of the standard cell lysis method previously described (Palevich, et al. 2018; Seshadri, et al. 2018), followed by phenol-chloroform extraction, and purification using the Qiagen Genomic-Tip 500 Maxi kit (Qiagen, Hilden, Germany). Specificity of genomic DNA was verified by automated Sanger sequencing of the 16*S* rRNA gene following PCR amplification from genomic DNA. Total DNA amounts were determined using a NanoDrop® ND-1000 (Thermo Scientific Inc.) and a Qubit Fluorometer dsDNA BR Kit (Invitrogen, USA), in accordance with the manufacturer's instructions. Genomic DNA integrity was verified by agarose gel electrophoresis and using a 2000 BioAnalyzer (Agilent, USA).

Genome Sequencing, Assembly and Comparison

Pseudobutyrivibrio xylanivorans MA3014 was selected for genome sequencing as a NZ strain and only representative member of *P. xylanivorans* from the Hungate1000 collection ((Seshadri, et al. 2018): Supplementary Table S1). The complete genome sequence of MA3014 was determined by pyrosequencing 3 kb mate paired-end sequence libraries using the 454 GS FLX platform with Titanium chemistry (Macrogen, Korea). Pyrosequencing reads provided 55× coverage of the genome and were assembled using the Newbler assembler (version 2.7, Roche

454 Life Sciences, USA) which resulted in 116 contigs across 13 scaffolds. Gap closure was managed using the Staden package (Staden, et al. 1999) and gaps were closed using additional Sanger sequencing by standard and inverse PCR techniques. In addition, MA3014 genomic DNA was sequenced using shotgun sequencing of 2 kb paired-end sequence libraries using the Illumina MiSeq platform (Macrogen, Korea) which provided 677-fold sequencing coverage. A *de novo* assembly was performed using the assemblers Velvet version 3.0 (Zerbino and Birney 2008), and EDENA version 3.120926 (Hernandez, et al. 2008). The resulting sequences were combined with the Newbler assembly using the Staden package and Geneious, version 8.1 (Kearse, et al. 2012). Genome assembly was confirmed by pulsed-field gel electrophoresis (Palevich 2011; Palevich N, et al. 2019b) and genome annotation was performed as described previously (Kelly, et al. 2010). Genome comparisons of orthologous gene clusters within *Pseudobutyrivibrio* genomes were performed using OrthoVenn version 2 (Wang, et al. 2015).

Results and Discussion

Genome Assembly, Properties and Annotation

To sequence the genome of *P. xylanivorans* MA3014, short-read 454 GS FLX Titanium and Illumina technologies based on 9.9 million paired-end (PE) reads was applied (Supplementary Tables S1 and S2). The genome of *P. xylanivorans* MA3014 consists of three replicons (Palevich 2011, 2016); a single chromosome (3,412,851 bp, %G+C 39.7), a chromid or secondary chromosome (PxyII, 88,942 bp, %G+C 36.9) and a plasmid (pNP95, 82,698 bp, %G+C 37.4) (Figure 1A). The total size of the closed genome is 3,584,491 bp with an overall %G+C content of 39.6%. The MA3014 assembly with high coverage of 677× was achieved using insert sizes that ranged between 238 bp (Illumina MiSeq) and 2.5kb (454 GS-FLX Titanium). In total, 2.6

Gb of trimmed and filtered sequence data was retained for the reported assembly (Supplementary Table S2). The overall genome assembly statistics of MA3014 are similar to the Mz 5^T (DSM 14809) and NCFB 2399 (DSM 10317) ((Kopečný, et al. 2003): Supplementary Table S4).

Ab initio gene prediction resulted in a total of 3,365 genes annotated in MA3014, of which 3,265 (97.03%) were PCGs (Supplementary Table S4). Among these, a putative function was assigned to 2,364 (70.25%), while 601 PCGs were annotated as hypothetical proteins or proteins of unknown function. In total, 840 (24.96%) genes have clear homology to proteins in the KEGG database, 2,506 (74.47%) and 2,593 (77.06%) of annotated genes have well-defined PFAM and InterPro protein domains, respectively. In contrast, 153 (4.55%) of annotated genes have identified signal peptide protein domain hits and are predicted have extracellular functions. The MA3014 chromosome encodes 3,098 PCGs while the PxyII and pNP95 encode 96 and 71 genes, respectively. Overall, the coding region comprises 89.77% of the genome, typical of rumen bacterial genomes (Seshadri, et al. 2018; Palevich, et al. 2019b). However, in order to elucidate the actual genetic divergence within the rumen *Butyrivibrio* and *Pseudobutyrivibrio,* future efforts should focus on the generation of complete genomes, starting with the Hungate1000 collection.

Genome Comparison

A comparison of the *P. xylanivorans* MA3014 genome with the draft genomes of *P. xylanivorans* Mz 5^T (DSM 14809) and NCFB 2399 (DSM 10317) (Kopečný, et al. 2003) is shown in Supplementary Table S4. MA3014 is the largest *P. xylanivorans* genome to date, where it is 163,567 bp and 370,547 bp larger, also contains 187 and 375 more PCGs than Mz 5^T and NCFB 2399, respectively. A novel feature of MA3014 and other well-characterized *Butyrivibrio* genomes is the presence of chromids or secondary chromosomes (Kelly, et al. 2010;

Palevich, et al. 2019a). Chromids are replicons with %G+C content similar to that of their main chromosome, but have plasmid-type maintenance and replication systems, are usually smaller than the chromosome (but larger than plasmids) and contain genes essential for growth along with several core genus-specific genes (Harrison, et al. 2010). The PxyII replicon has been designated as a chromid of MA3014 as it possesses all of these characteristics and contains genes encoding enzymes that have a role in carbohydrate metabolism and transport. Since the PxyII chromid of MA3014 is 2,834 bp smaller than the Bhu II chromid of MB2003, it is now the smallest chromid reported for bacteria. Although several plasmid replication genes have been identified in the Mz 5^T but not in NCFB 2399 draft genomes, the presence of extrachromosomal elements requires experimental validation.

Comparison of MA3014, Mz 5^T and NCFB 2399 genomes based on COG category (Supplementary Table S5) and synteny analysis (Figure 1B-C), show that these *Pseudobutyrivibrio* strains are genetically similar. Despite the differences in genome sizes of MA3014 and Mz 5^T, the basic metabolism of these two rumen bacteria are comparable. Comparative pan-genomics of these rumen bacterial strains revealed highly open genomes and a strong correlation of orthologous genes among these species (Figure 1D). Most of the predicted MA3014 genes were found to have homologs (BLASTP e-value cut-off 10^{-5}) in the other two strains (2,356; 73%), with the *P. xylanivorans* represented by 768 orthologous clusters and 1,996 single-copy genes. In total, 2,036 core genes were found to be orthologous among the three *P. xylanivorans* genomes compared, with only 58 genes found to be unique to MA3014 (Figure 1D). Genomic comparisons with other species within the genera *Butyrivibrio* and *Pseudobutyrivibrio* have revealed strong collinearities (Palevich, et al. 2019b), that will facilitate our understanding of genome evolution of rumen bacteria.

Polysaccharide Degradation

The Carbohydrate-Active enZYmes database was used to identify glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and carbohydrate-binding protein module (CBM) families within the MA3014 genome. Overall, CAZyme profile of MA3014 is similar to other *Pseudobutyrivibrio* that are in general not as extensive as those of *Butyrivibrio* (Palevich 2016; Palevich, et al. 2019b). Analysis of the functional domains of enzymes involved in the breakdown or synthesis of complex carbohydrates, has revealed the polysaccharide-degrading potential of this rumen bacterium (Supplementary Table S6). Approximately 2% of the MA3014 genome (69 CDSs) is predicted to encode either 22 secreted (21 GHs and one CE) and 47 intracellular (41 GHs, 4 CEs and two GTs) proteins dedicated to polysaccharide degradation. The enzymatic profiles of MA3014 and Mz 5^T are almost identical, as both possess the same genes encoding predicted secreted and intracellular CAZymes in their genomes. Out of the 22 genes predicted to encode secreted polysaccharide degrading enzymes, only β-glucosidase *bgl3K* (FXF36_15770) is encoded by the MA3014 chromid (PxyII). The majority (40) of MA3014 genes encoding intracellular proteins involved in polysaccharide breakdown (excluding GTs), had corresponding homologues in Mz 5^T. The most abundant Pfam domains included GH families (GH3, GH13 and GH43) and CE1, most of which did not contain signal sequences and predicted to be located intracellularly. Similarly, CAZymes with predicted roles in xylan (GH8, GH51, GH115), dextrin and starch (GH13 and GH77) degradation families were also predicted to be mostly intracellular.

Growth experiments showed MB2003 to be a metabolically versatile bacterium able to grow on a wide variety of monosaccharides and disaccharides (Supplementary Table S3). However, unlike Mz 5^T (Kopecný, et al. 2003), MA3014 was unable to utilize the insoluble substrate pectin for

growth. This difference is due to Mz 5^T possession of 4 pectate lyases (1 PL1 and 3 PL3) predicted to be involved in pectin degradation and utilization, of which MA3014 has none. The ability of MA3014 to breakdown starch and xylan is predicted to be based on four large (>1,000 aa) cell-associated proteins shown to be significantly up-regulated in related *B. hungatei* MB2003 and *B. proteoclasticus* $B316^T$ cells grown on xylan (Kelly, et al. 2010; Palevich, et al. 2019a). These are: α-amylase *amy13E* (FXF36_11320), arabinogalactan endo-1,4-β-galactosidase *agn53A* (FXF36_02635), xylosidase/arabinofuranosidase *xsa43D* (FXF36_08285), endo-1,4-β-xylanase xyn10A (FXF36_14365). These proteins contain multiple cell wall binding repeat domains (CW-binding domain, Pfam01473) at their C-termini that are predicted to anchor the protein to the peptidoglycan cell membrane (Dunne, et al. 2011). In addition, the secreted α-amylase *amy13E* (FXF36_11320) contains a CBM26 (Pfam16738) domain with predicted starch-binding functions (Gilbert, et al. 2013; McCartney, et al. 2004).

Electron microscopy of MA3014 cells grown in liquid media supplemented with plant material has revealed the copious production of exopolysaccharides (EPS) (Figure 2A-B). EPS is a characteristic of *Butyrivibrio* strains and is composed of the neutral sugars rhamnose, fucose, mannose, galactose and glucose (Stack 1988), made available by recycling the plant polysaccharides breakdown products. Our findings also show the presence of cytoplasmic inclusions (Figure 2C), similar to those seen in $B316^T$ and other *Butyrivibrio* strains containing glycogen-like material (Hespell, et al. 1993). The MA3014 genome encodes a complete repertoire of genes for glycogen synthesis and degradation, suggesting that a variety of complex oligosaccharides resulting from extracellular hydrolysis are metabolized within the cell and that glycogen has a role in the storage of excess carbohydrate.

Enolase Loss and Metabolic Flexibility

Pseudobutyrivibrio and *Butyrivibrio* are key members of the degradative microbiota found in a highly carbohydrate-rich environment and appear to have evolved beyond using glycolysis as the central pathway. The pathways for butyrate production for these rumen bacteria presume the possession of a complete Embden-Meyerhof-Parnas (EMP) glycolytic pathway. The lack of an enolase (*eno*, EC4.2.1.11), that converts 2-phospho-D-glycerate to phosphoenolpyruvate in the second to last step of the EMP pathway, is extremely unusual. As part of the Hungate1000 project in which the genomes of 410 rumen microbes were sequenced (Seshadri, et al. 2018), we discovered that many specialized polysaccharide fermenters lacked an enolase gene. Of all 21 *Pseudobutyrivibrio* genomes sequenced, only *P. xylanivorans* MA3014 and *P. ruminis* AD2017 reported to date lack a detectable enolase. *Eno*- *P. xylanivorans* MA3014 and several *Butyrivibrio* strains were recently confirmed using PCR screens with *eno*-specific primers (Kelly, et al. 2010; Palevich, et al. 2018). Given the observed polysaccharide-degrading abilities and lactate production, the Methylglyoxal Shunt (MS) and uronic acid metabolic pathways (Figure 2D), have been suggested as alternatives to the EMP pathway (Cooper 1984). Previous work has also reported similar findings in other *Butyrivibrio* strains (Kelly, et al. 2010; Palevich, et al. 2019a). In this pathway the dihydroxyacetone phosphate is transformed to pyruvate via methylglyoxal and D-lactate dehydrogenase encoded by *ldhA*. The MA3014 genome possesses methylglyoxal synthase, *mgsA* (FXF36_12340), glyoxalases *gloA/B* (FXF36_00730, FXF36_01130 and FXF36_09530) and both D- and L-lactate dehydrogenases *ldh* (FXF36_04170 and FXF36_11135) genes. In addition, MA3014 has the same set of genes as the previously reported and well-characterized *B. hungatei* MB2003 and *B. proteoclasticus* B316^{T} for the

production of butyrate, formate, acetate and lactate (Kelly, et al. 2010; Palevich, et al. 2019a; Palevich, et al. 2017).

In some butyrate-forming anaerobes, crotonyl-CoA reduction is linked to electron transport phosphorylation (ETP) via flavin-based electron bifurcating *ech* and *rnf* complexes which act as transmembrane ion pumps (Buckel and Thauer 2013; Herrmann, et al. 2008; Li, et al. 2008; Welte, et al. 2010). A recent analysis of the Hungate1000 dataset (Hackmann and Firkins 2015; Seshadri, et al. 2018; Palevich, et al. 2019b), found that *Pseudobutyrivibrio* and *Butyrivibrio* genomes encode both Ech and Rnf homologues proposed to act in concert with NifJ and Bcd-Etf to form an electrochemical potential and drive ATP synthesis (Gutekunst, et al. 2014; Tremblay, et al. 2013). This allows these rumen bacteria to generate approximately 4.5 ATP/glucose in total, one the highest yields for anaerobic fermentation of glucose (Buckel and Thauer 2013). Given the importance of *eno*, *Pseudobutyrivibrio* and *Butyrivibrio* may be displaying an example of environment-specific evolution by gene loss that warrants further investigation into the alternative pathways that permit ATP generation. The genome sequence of *P. xylanivorans* MA3014 presented here is consistent with the genome architecture of other sequenced *Pseudobutyrivibrio* strains and is a valuable resource for future studies regarding bacterial-driven plant-fibre degradation in ruminants.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Data deposition: The complete genome sequence of *Pseudobutyrivibrio xylanivorans* MA3014 and its annotations are deposited in Genbank under accession numbers CP043028, CP043029 and CP043030.

Acknowledgements

The *P. xylanivorans* MA3014 genome sequencing project was funded by the New Zealand Ministry of Business, Innovation and Employment New Economy Research Fund programme: Accessing the uncultured rumen microbiome, contract number C10X0803. Electron microscopy was conducted with the assistance of the Manawatū Microscopy and Imaging Centre at Massey University, Palmerston North, New Zealand. We thank Sarah Lewis for assistance with the fermentation end product analysis, Abdul Baten and Ron Ronimus for critical review of the manuscript.

Literature Cited

Bryant MP, Small N 1956. The anaerobic monotrichous butyric acid-producing curved rod-shaped bacteria of the rumen. *J Bacteriol*. 72(1): 16-21.

Buckel W, Thauer RK 2013. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. *Biochim Biophys Acta Bioenerg*. 1827(2): 94-113.

Cooper R 1984. Metabolism of methylglyoxal in microorganisms. *Annu Rev Microbiol*. 38(1): 49-68.

Dunne JC, Li D, Kelly WJ, Leahy SC, Bond JJ, Attwood GT, Jordan TW 2011. Extracellular polysaccharide-degrading proteome of *Butyrivibrio proteoclasticus*. *J Proteome Res*. 11(1): 131-142.

Forster R, Teather R, Gong J, Deng S-J 1996. 16*S* rDNA analysis of *Butyrivibrio fibrisolvens*: phylogenetic position and relation to butyrate-producing anaerobic bacteria from the rumen of white-tailed deer. *Lett Appl Microbiol.* 23(4): 218-222.

Forster RJ, Gong J, Teather RM 1997. Group-specific 16*S* rRNA hybridization probes for determinative and community structure studies of *Butyrivibrio fibrisolvens* in the rumen. *Appl Environ Microbiol*. 63(4): 1256-1260.

Gilbert HJ, Knox JP, Boraston AB 2013. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol*. 23(5): 669-677.

Gutekunst K, Chen X, Schreiber K, Kaspar U, Makam S, Appel J 2014. The bidirectional NiFe-hydrogenase in *Synechocystis* sp. PCC 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions. *J Biol Chem*. 289(4): 1930-1937.

277 Hackmann TJ, Firkins JL 2015. Electron transport phosphorylation in rumen butyrivibrios:
278 unprecedented ATP yield for glucose fermentation to butyrate. *Front Microbiol*. 6:622.

279 Harrison PW, Lower RP, Kim NK, Young JP 2010. Introducing the bacterial 'chromid': not a
280 chromosome, not a plasmid. *Trends Microbiol.* 18(4): 141-148.

281 Hazlewood GP, Theodorou MK, Hutchings A, Jordan DJ, Galfre G 1986. Preparation and
282 characterization of monoclonal antibodies to a *Butyrivibrio* sp. and their potential use in the
283 identification of rumen *Butyrivibrios*, using an enzyme-linked immunosorbent assay. *J Gen*
284 *Microbiol.* 132(1): 43-52.

285 Henderson G, Cox F, Ganesh S, Jonker A, Young W, Global Rumen Census Collaborators,
286 Janssen, PH 2015. Rumen microbial community composition varies with diet and host, but a
287 core microbiome is found across a wide geographical range. *Sci Rep*. 5(14567): 1-13.

288 Hernandez D, François P, Farinelli L, Østerås M, Schrenzel J 2008. *De novo* bacterial genome
289 sequencing: millions of very short reads assembled on a desktop computer. *Genome Res*.
290 18(5): 802-809.

291 Herrmann G, Jayamani E, Mai G, Buckel W 2008. Energy conservation via electron-transferring
292 flavoprotein in anaerobic bacteria. *J Bacterio*. 190(3): 784-791.

293 Hespell R, Kato K, Costerton J 1993. Characterization of the cell wall of *Butyrivibrio* species.
294 *Can J Microbiol*. 39(10): 912-921.

295 Hungate RE. 1966. The rumen and its microbes. New York, USA: Academic Press.

296 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
297 Markowitz S, Duran C 2012. Geneious Basic: an integrated and extendable desktop software
298 platform for the organization and analysis of sequence data. *Bioinformatics*. 28(12): 1647-
299 1649.

300 Kelly WJ, Leahy SC, Altermann E, Yeoman CJ, Dunne JC, Kong Z, Pacheco DM, Li D, Noel
301 SJ, Moon CD, Cookson AL, Attwood GT 2010. The glycobiome of the rumen bacterium
302 *Butyrivibrio proteoclasticus* B316^{T} highlights adaptation to a polysaccharide-rich
303 environment. *PLoS One*. 5(8): e11942.

304 Kenters N, Henderson G, Jeyanathan J, Kittelmann S, Janssen PH 2011. Isolation of previously
305 uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *J*
306 *Microbiol Methods*. 84(1): 52-60.

307 Kopecný J, Zorec M, Mrazek J, Kobayashi Y, Marinsek-Logar R 2003. *Butyrivibrio hungatei* sp.
308 nov. and *Pseudobutyrivibrio xylanivorans* sp. nov., butyrate-producing bacteria from the
309 rumen. *Int J Syst Evol Microbiol*. 53(1): 201-209.

310 Li F, Hinderberger J, Seedorf H, Zhang J, Buckel W, Thauer RK 2008. Coupled ferredoxin and
311 crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA
312 dehydrogenase/Etf complex from *Clostridium kluyveri*. *J Bacterio*. 190(3): 843-850.

313 Mannarelli BM 1988. Deoxyribonucleic acid relatedness among strains of the species
314 *Butyrivibrio fibrisolvens*. *Int J Syst Bacteriol*. 38(4): 340-347.

Mannarelli BM, Stack RJ, Lee D, Ericsson L 1990. Taxonomic relatedness of *Butyrivibrio*, *Lachnospira*, *Roseburia*, and *Eubacterium* species as determined by DNA hybridization and extracellular-polysaccharide analysis. *Int J Syst Bacteriol.* 40(4): 370-378.

McCartney L, Gilbert HJ, Bolam DN, Boraston AB, Knox JP 2004. Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Anal Biochem*. 326(1): 49-54.

Noel S 2013. Cultivation and community composition analysis of plant-adherent rumen bacteria. *PhD dissertation*. Massey University, Palmerston North, New Zealand.

Palevich N 2016. Comparative genomics of *Butyrivibrio* and *Pseudobutyrivibrio* from the rumen. *PhD dissertation*. Massey University, Palmerston North, New Zealand.

Palevich N 2011. Genome sequencing of rumen bacteria involved in lignocellulose digestion. *MS thesis*. Massey University, Palmerston North, New Zealand.

Palevich N, Kelly WJ, Ganesh S, Rakonjac J, Attwood GT 2019a. *Butyrivibrio hungatei* MB2003 competes effectively for soluble sugars released by *Butyrivibrio proteoclasticus* B316^{T} from growth on xylan or pectin. *Appl Environ Microbiol.* 85(3): e02056-18.

Palevich N, Kelly WJ, Leahy SC, Denman SE, Altermann E, Rakonjac J, Attwood GT 2019b. Comparative genomics of rumen *Butyrivibrio* uncovers a continuum of polysaccharide-degrading capabilities. *Appl Environ Microbiol.* 86(1): e01993-19.

Palevich N, Kelly WJ, Leahy SC, Altermann E, Rakonjac J, Attwood GT 2017. The complete genome sequence of the rumen bacterium *Butyrivibrio hungatei* MB2003. *Stand Genomic Sci*. 12(1): 72.

Richardson A, Calder A, Stewart C, Smith A 1989. Simultaneous determination of volatile and non-volatile acidic fermentation products of anaerobes by capillary gas chromatography. *Lett Appl Microbiol*. 9(1): 5-8.

Seshadri R, Leahy SC, Attwood GT, Teh KH, Lambie SC, Cookson AL, Eloe-Fadrosh EA, Pavlopoulos GA, Hadjithomas M, Varghese NJ, Paez-Espino D, Hungate1000 project collaborators, Perry R, Henderson G, Creevey CJ, Terrapon N, Lapebie P, Drula E, Lombard V, Rubin E, Kyrpides NC, Henrissat B, Woyke T, Ivanova NN, Kelly WJ 2018. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol*. 36(4): 359-367.

Shane BS, Gouws L, Kistner A 1969. Cellulolytic bacteria occurring in the rumen of sheep conditioned to low-protein teff hay. *J Gen Microbiol.* 55(3): 445-457.

Stack RJ 1988. Neutral sugar composition of extracellular polysaccharides produced by strains of *Butyrivibrio fibrisolvens*. *Appl Environ Microbiol*. 54(4): 878-883.

Staden R, Beal KF, Bonfield JK. 1999. The staden package. In: Misener S, Krawetz SA, editors. Bioinformatics methods and protocols. New York: Humana Press. p. 115-130.

Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR 2013. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin: NAD^{+} oxidoreductase essential for autotrophic growth. *mBio*. 4(1): e00406-00412.

354 Van Gylswyk N, Hippe H, Rainey F 1996. *Pseudobutyrivibrio ruminis* gen. nov., sp. nov., a
355 butyrate-producing bacterium from the rumen that closely resembles *Butyrivibrio fibrisolvens*
356 in phenotype. *Int J Syst Evol Microbiol*. 46(2): 559-563.

357 Wang Y, Coleman-Derr D, Chen G, Gu YQ 2015. OrthoVenn: a web server for genome wide
358 comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.*
359 43(W1): W78-W84.

360 Welte C, Krätzer C, Deppenmeier U 2010. Involvement of Ech hydrogenase in energy
361 conservation of *Methanosarcina mazei*. *FEBS J*. 277(16): 3396-3403.

362 Willems A, Amat-Marco M, Collins MD 1996. Phylogenetic analysis of *Butyrivibrio* strains
363 reveals three distinct groups of species within the *Clostridium subphylum* of the gram-positive
364 bacteria. *Int J Syst Bacteriol.* 46(1): 195-199.

365 Willems A, Collins MD. 2009. *Butyrivibrio*. Bryant and Small 1956b, 18, emend. Moore,
366 Johnson and Holdeman 1976, 241^{AL}. In: De Vos P, et al., editors. Bergey's Manual of
367 Systematic Bacteriology. New York: Springer. p. 927-937.

368 Zerbino DR, Birney E 2008. Velvet: algorithms for de novo short read assembly using *de Bruijn*
369 graphs. *Genome Res*. 18(5): 821-829.

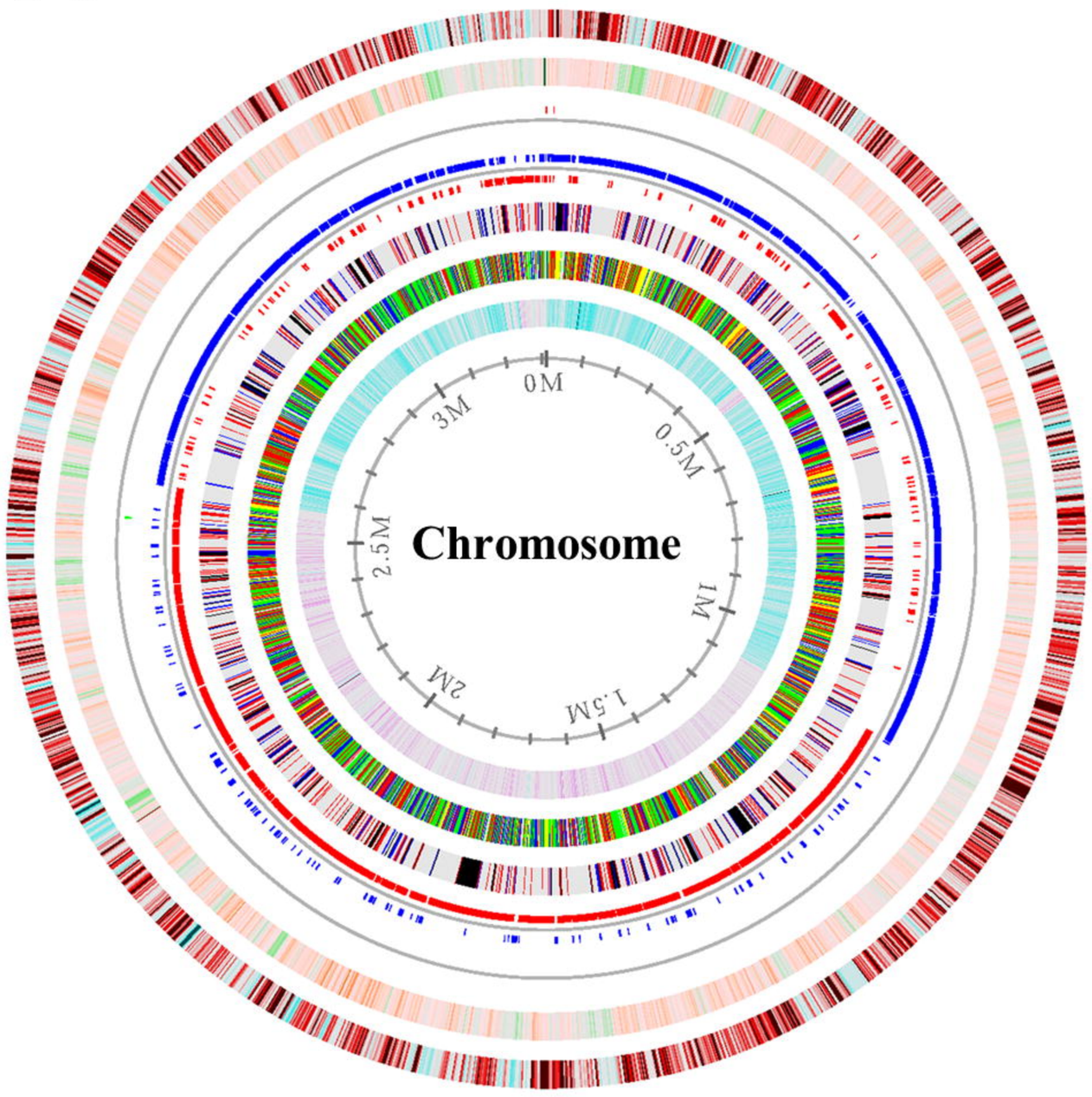
370 Zorec M, Marinšek-Logar R, Čepeljnik T, Nekrep FV 2000. The influence of substrate
371 concentration and growth phase on expression of *Butyrivibrio* sp. Mz^{5} endoxylanases. *Res*
372 *Rep Biotechnical Fac Univ Ljubljana Agric Issue Zootechny*. 76: 199-206.

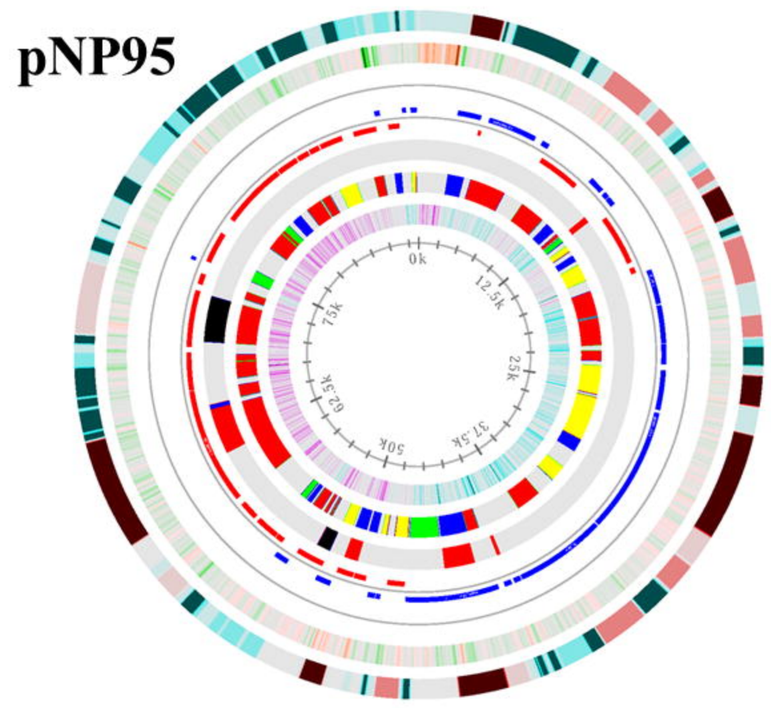
373 **Figure 1.** (A) Genome atlas for *P. xylanivorans* MA3014. The figure represents a circular view
374 of the four replicons that constitute the *P. xylanivorans* MA3014 genome. The key at the right
375 describes the concentric circles within each replicon in the outermost to innermost direction.
376 Circle 1 (innermost circle) indicates GC-skew. Circle 2 shows COG classifications of predicted
377 and annotated open reading fames (ORFs) grouped into five major categories: information
378 storage and processing (yellow); cellular processes and signalling (red); metabolism (green);
379 poorly characterized (blue); uncharacterized or no COG assignment (uncoloured). Circle 3 shows
380 transmembrane helices (TMH) and SignalP domains grouped into four categories: both absent
381 (uncoloured); TMH (red); SignalP (blue); both present (black). Circle 4 indicates ORF
382 orientation in sense (ORF+, blue) or antisense (ORF-, red) directions. Circle 5 shows tRNA
383 (green) and rRNA (red) ribosomal machinery. Circle 6 shows G+C content deviation from the
384 average in either green (low GC spike) or orange (high GC spike). Circle 7 shows BLAST

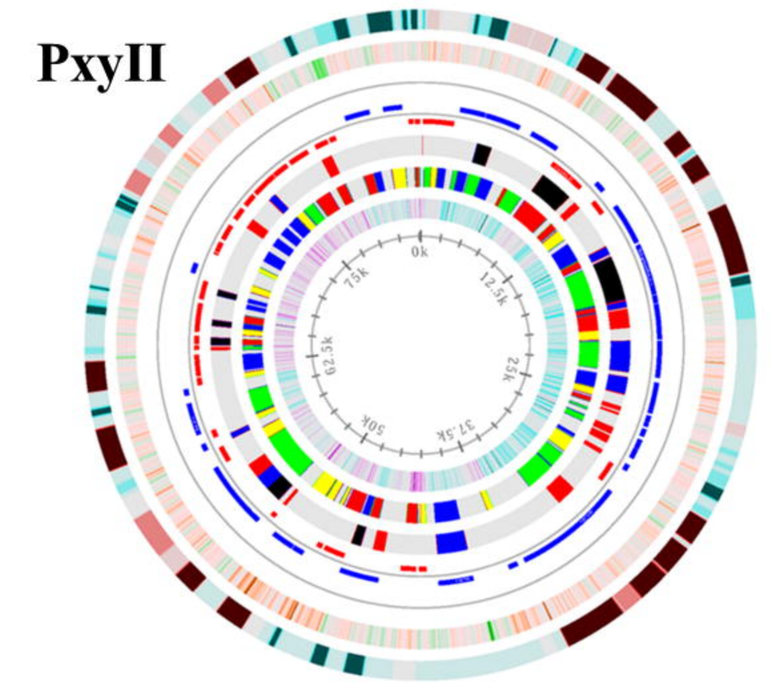
similarities of unique proteins (blue) and highly conserved features (red) relative to sequences in the nonredundant (nr) database. (B-C) Genome synteny analysis. Alignment of the *P. xylanivorans* MA3014 genome against the draft genomes of *P. xylanivorans* Mz 5^T (B) and *P. xylanivorans* NCFB 2399 (C). Whenever the two sequences agree, a colored line or dot is plotted. Units displayed in base-pairs. Color codes: blue, forward sequence, red, reverse sequence. (D) Venn diagram showing the distribution of shared gene families among the *P. xylanivorans* genomes. All *P. xylanivorans* scaffolds with at least a single one-to-one ortholog shared among the genomes were compared.

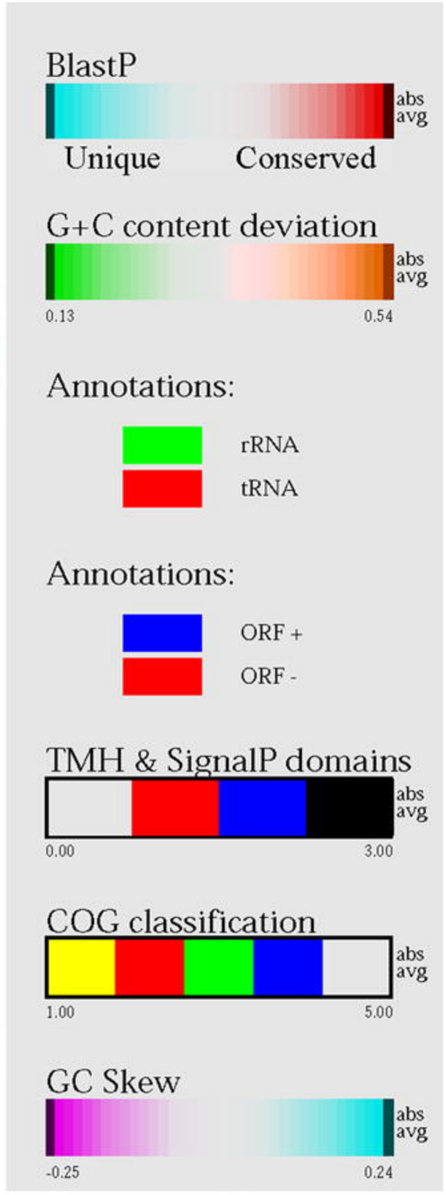
Figure 2. (A-C) Electron micrographs of *P. xylanivorans* MA3014. (A-B) Scanning EMs of MA3014 cells adherence to the surface (A) and exposed end (B) of NDF plant material, at 5,000 × magnification. (C) Transmission EM of negatively stained MA3014 cells grown in liquid medium at 10,000 × magnification. Arrows indicate the presence of glycogen inclusions. (D) Fermentation pathways in rumen *Pseudobutyrivibrio* and *Butyrivibrio*. Abbreviations: Bcd-Etf, butyryl-CoA dehydrogenase/electron transferring flavoprotein; Ech, *E. coli* hydrogenase-3-type hydrogenase; Fd, ferredoxin; Fd_{ox}, oxidized Fd; Fd_{red}, reduced Fd; Glo, glyoxalase; MsgA, methylglyoxal synthase; NAD, nicotinamide adenine dinucleotide; NAD_{ox}, oxidized NAD; NAD_{red}, reduced NAD; NifJ, nitrogen fixation J; Rnf, *Rhodobacter* nitrogen fixation; ATPase = F_0F_1-ATPsynthase.

A

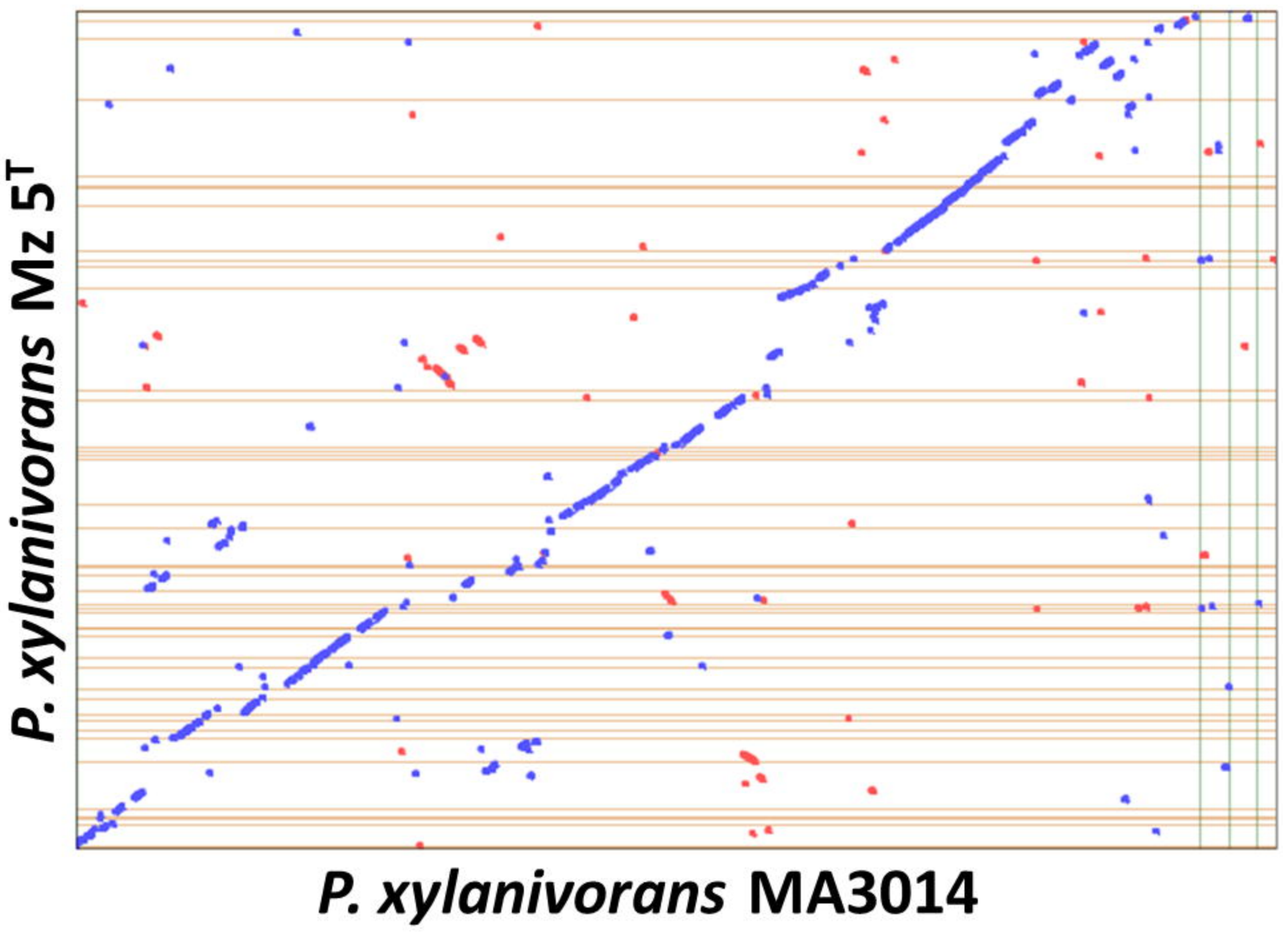




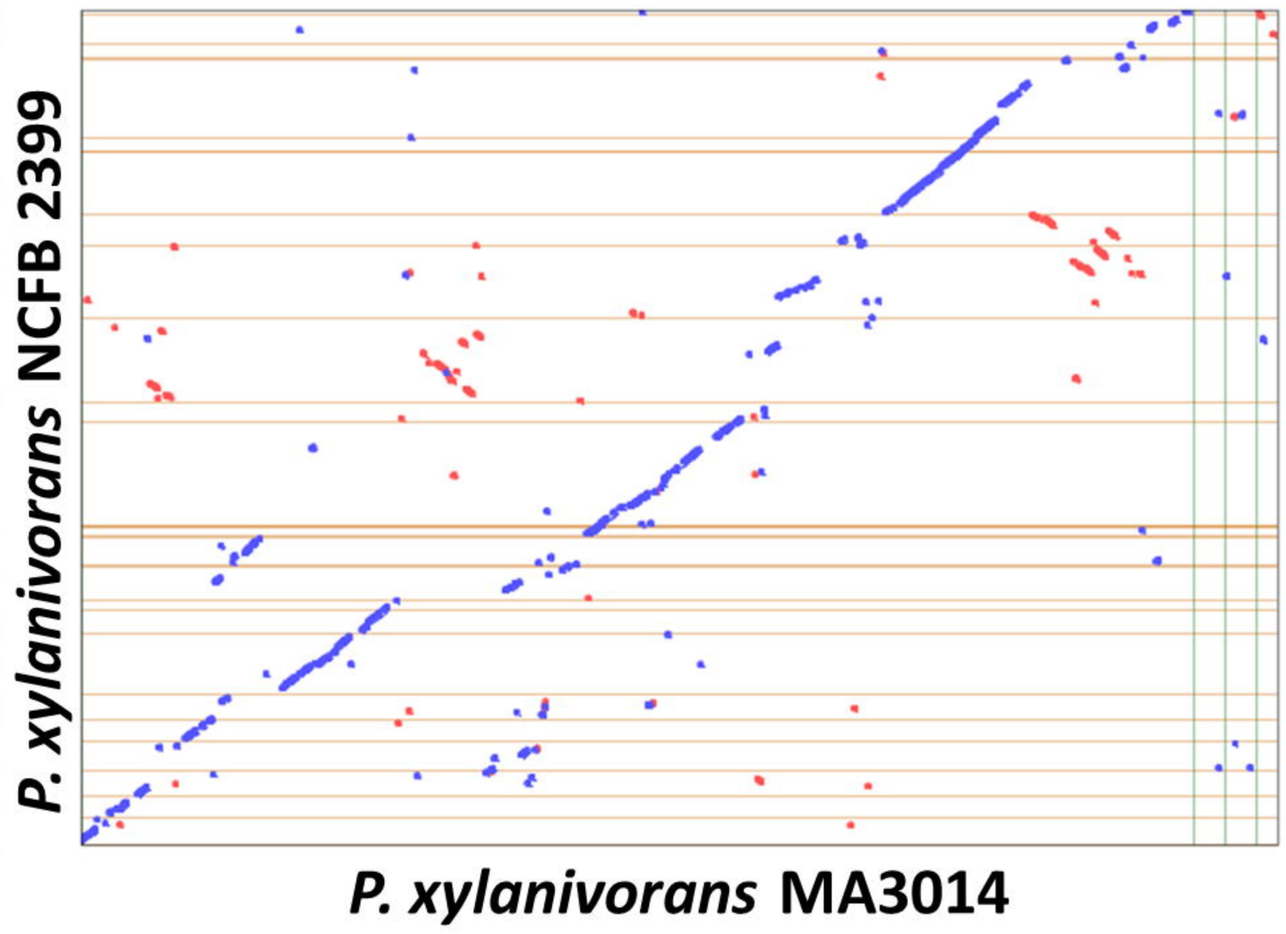




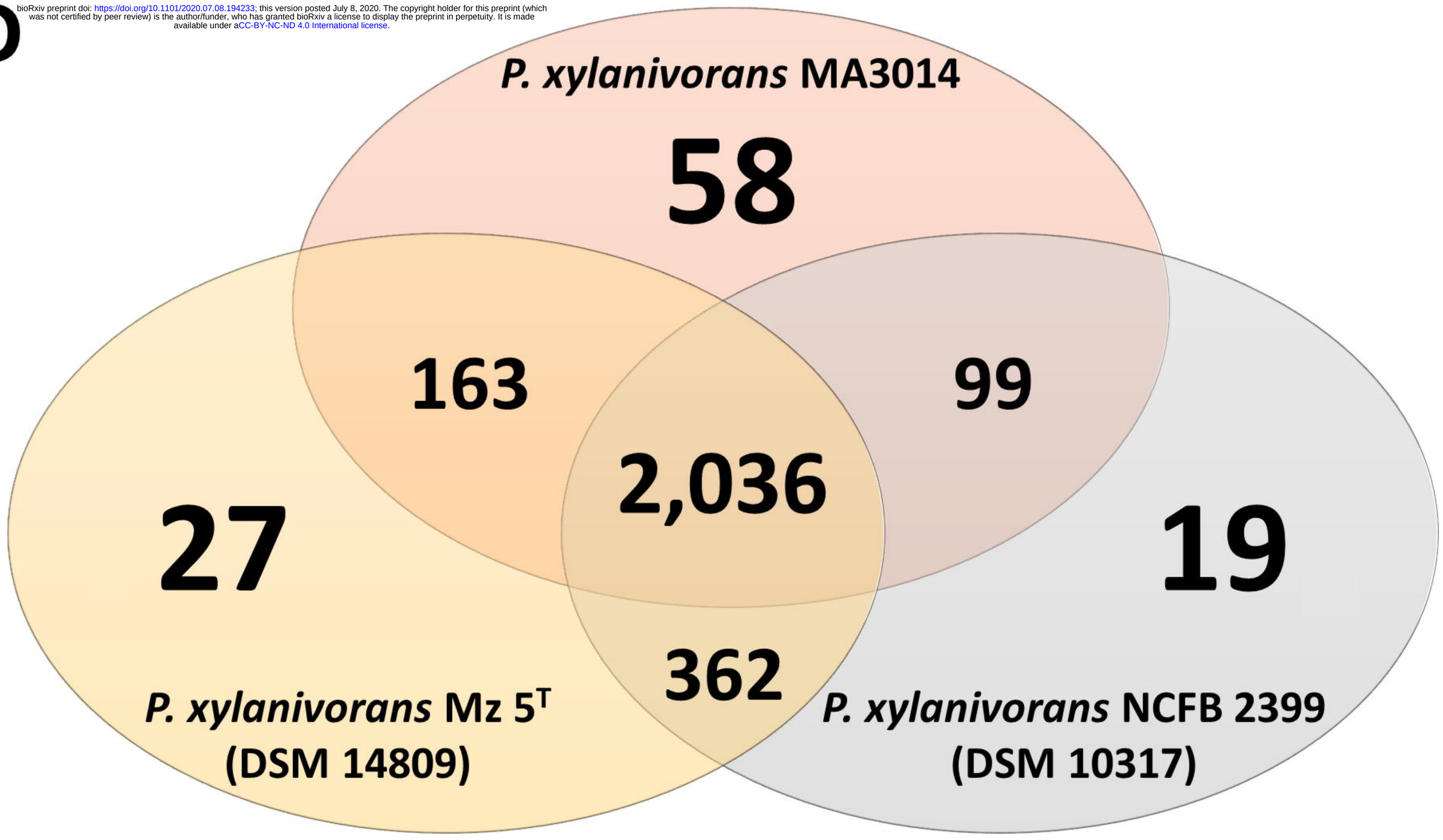
B



C



D



A

10 µm

B

10 µm

C

500 nm

D

Oligosaccharide Breakdown and Uptake

D-Glucose

Galacturonate

DKI

Glucuronate

Uronic Acids Metabolism

KDG

KDGP

Rnf

Ech

ATPase

Methylglyoxal Shunt

D-Lactate

GloA/B

Methylglyoxal

MgsA

DHAP

D-Fru-1,6-biP

GAP

2-P-glycerate

racemase

L-lactate

LdhA

Ldh

Enolase (eno)

PEP

Pyruvate

Formate

Acetate

NifJ

Acetyl-CoA

Butyrate Metabolism

Butyrate

Butyryl-CoA

Bcd-Etf

Crotonyl-CoA

3-HB-CoA

Acetoacetyl-CoA