Systematic and Applied Microbiology 42 (2019) 107–116

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: http://www.elsevier.com/locate/syapm

The diverse and extensive plant polysaccharide degradative apparatuses of the rumen and hindgut *Prevotella* species: A factor in their ubiquity?

Tomaž Accetto*, Gorazd Avguštin

University of Ljubljana, Biotechnical faculty, Animal Science Department, Groblje 3, 1230 Domžale, Slovenia

ARTICLE INFO

Article history:
Received 2 August 2018
Received in revised form 2 October 2018
Accepted 3 October 2018

Keywords:
Prevotella
Rumen
Hindgut
Polysaccharide utilization locus
CAZYme
Metagenome

ABSTRACT

Although the *Prevotella* are commonly observed in high shares in the mammalian hindgut and rumen studies using NGS approach, the knowledge on their actual role, though postulated to lie in soluble fibre degradation, is scarce. Here we analyse in total 23, more than threefold of hitherto known rumen and hindgut *Prevotella* species and show that rumen/hindgut *Prevotella* generally possess extensive repertoires of polysaccharide utilization loci (PULs) and carbohydrate active enzymes targeting various plant polysaccharides. These PUL repertoires separate analysed *Prevotella* into generalists and specialists yet a finer diversity among generalists is evident too, both in range of substrates targeted and in PUL combinations targeting the same broad substrate classes. Upon evaluation of the shares of species analysed in this study in rumen metagenomes we found firstly, that they contributed significantly to total *Prevotella* abundance though much of rumen *Prevotella* diversity may still be unknown. Secondly, the hindgut *Prevotella* species originally isolated in pigs and humans occasionally dominated among the *Prevotella* with surprisingly high metagenome read shares and were consistently found in rumen metagenome samples from sites as apart as New Zealand and Scotland. This may indicate frequent passage between different hosts and relatively low barriers to their successful establishment in rumen versus the hindgut.

Introduction

The rumen is a forestomach with the principal function to supply the animal with short chain fatty acids and microbial protein, which are absorbed through the rumen wall and digested in the abomasum, respectively. Both are the products of a largely bacterial fermentation of ingested plant material, the composition of which may vary according to animal feeding strategy, cattle being one of the most efficient plant fibre utilizers [16]. It became apparent in a recent large scale study [13] which assayed samples from various, mainly domesticated ruminants that some bacterial genera were abundant in rumens all over the world thus constituting the core rumen microbiome. One of the most prominent among these was the genus *Prevotella* of the phylum *Bacteroidetes*. *Prevotella* were detectable already in the rumen samples of three days old animals in a cattle rumen community succession study and reached the astonishing 48% of amplicon NGS reads by two months [17]. The *Prevotella* were also shown to be efficient primary colonizers of ryegrass [28]. The large intestine of humans also harbours the anaerobic microbial community involved in breakdown of plant fibre, which is commonly dominated by the members of bacterial phyla *Bacteroidetes* and *Firmicutes*. Of the *Bacteroidetes* in humans, genus *Prevotella* tends to be dominant in rural populations depending on plant rich diets while members of the genus *Bacteroides* dominate when diet is richer in protein and fat [12]. The *Prevotella* are believed to be important mainly in the non-cellulose plant fibre degradation using the products of polysaccharide utilization loci (PULs) genes to sense, bind and degrade various glycans [26]. The central part of PULs are the paralogues of *Bacteroides thetaiotaomicron* VPI-5482 *susC* and *susD* genes which are always found together and code for the TonB-dependent transporters and carbohydrate-binding lipoproteins respectively [10]. The PUL-encoded polysaccharide utilization apparatus displays specificity for different polysaccharide targets already at the SusD-like protein binding step [19,20]. In the case of xylan and fructan, the SusD-like protein is even specific for their subtypes e.g. glucuronoxylan versus more complex arabinoglucuronoxylan and inulin versus levan respectively [36,43]. The research published till now has

Abbreviations: PUL, polysaccharide utilization locus; CAZYme, carbohydrate acting enzyme.

* Corresponding author.
E-mail address: tomaz.accetto@bf.uni-lj.si (T. Accetto).

https://doi.org/10.1016/j.syapm.2018.10.001
0723-2020/

mainly been focused on xylan and partly glucomannan utilization capabilities of a single *Prevotella* species, *Prevotella bryantii* [7,11,29,37], while less data is available for other three described ruminal species [4]. In fact, the genus *Prevotella* comprises over 50 species in total, the majority of these being of oral origin while at the time of writing seven have been described from rumen and hindgut. Clear evidence exists for the rumen at least, that the described species do not encompass the complete *Prevotella* diversity in this ecosystem, and many new strains have already been cultured and targeted for sequencing [6]. Given the numerical importance and prevalence of *Prevotella* in much researched gut and rumen ecosystems and the almost total lack of insight into diversity of their metabolic capabilities, the analysis of their dominant feature, i.e. the ability to degrade various polysaccharides, is in our opinion more than timely. Recent bioinformatic analysis of predominantly oral *Prevotella* species revealed large differences in glycan degradation capabilities indicating niche specialization: (i) some species were predicted to use starch only, (ii) most oral and urogenital species relied on host glycans, (iii) some oral species were apparently able to live off both host and plant glycans, and finally (iv) certain rumen/hindgut/oral species seemed adapted to use plant polysaccharides only [1]. What is the situation in the larger rumen/hindgut *Prevotella* species sample is the scope of this study. Beside the recognized *Prevotella* species, also the ruminal *Prevotella* strains sequenced by the Hungate 1000 project [6] and pig hindgut strains isolated and sequenced recently in our laboratory were included in this analysis. The repertoires of carbohydrate acting enzymes (CAZYmes), PULs, and ability to utilize various polysaccharides for growth were catalogued. Additionally, the significance of these species in major rumen metagenomes originating from across the world has been assessed.

Material and methods

Strains and genomes

The new pig hindgut *Prevotella* strains are a portion of hitherto undescribed *Bacteroidetes*, which were isolated in a targeted cultivation study, and have been sequenced and annotated as described previously in the case of *Prevotella pectinovora* [32]. Altogether 18 pig gut *Prevotella* strains were sequenced, five of them belonging to *P. pectinovora* [32], and the rest representing novel, still undescribed *Prevotella* species. Additionally *P. bryantii* TC1-1, the only *Prevotella* strain, for which limited genetic manipulation techniques exist and are used in our lab routinely [2,41], was also sequenced. The genomes were deposited at DDBJ/ENA/GenBank under the accessions NPIZ00000000, NPJA00000000, NPJB00000000, NPJC00000000, NPJD00000000, NPJE00000000, NPJF00000000, NPJG00000000, NPJH00000000, NPJI00000000, NPJJ00000000, NPJK00000000, NPJL00000000 and NPJM00000000. *Prevotella* genomes from the Hungate 1000 rumen strains collection were downloaded from the genome portal of the Joint genome institute (http://jgi.doe.gov/) and annotated using Prokka [40]. Strains were classified into species (Table S1) using the Average nucleotide identity (ANI) metric ANIb as implemented in JSpecies [34]. The average nucleotide identity is obtained by first extracting the stretches of similar sequence from a genome pair (in the case of ANIb using blast) and then calculating their average similarity. The ANI is one of primary tools for species description, used particularly to confirm genomic coherence of strains of the proposed species, the other two prerequisites which must be fulfilled being monophyly and phenotypic coherence [38]. When the number of genomes was large (>100), the ANIb was calculated using pyani (https://github.com/widdowquinn/pyani). Rumen *Prevotella* strains KHP1, Ga6B6, AGR2160, MA2016, KHP7, KH1P2, P6B11, HUN102, TC2-28, TF2-5 and NE3005 were a kind gift of the Rumen microbiology team of AgResearch, NZ (http://www.agresearch.co.nz/), while the rumen strains GA33, $B_1$4, TC1-1, TF1-3 and all of the pig gut strains were from our collection. *Bacteroides ovatus* ATCC 8483 and *B. thetaiotaomicron* VPI5482 were obtained from DSMZ, Germany, as DSM 1896 and DSM 2079, respectively. The 16S rRNA genes of all strains were amplified and sequenced to confirm the strain identity.

PUL and CAZYme repertoires

The genes coding for SusD-like glycan binding proteins were earlier shown to be suitable PUL indicators since they are always found beside the *susC*-like coding genes and produced almost the same results as were obtained in other studies [1]. Furthermore, their phylogeny showed that they separate into neat clusters with homogeneous ligand glycan partners as deduced by transfer of transcriptomic functional studies performed in *Bacteroides* [1]. Thus by finding the *susD* genes both the PUL locations and their putative target glycans may be revealed. The PUL finding workflow was essentially the same as described earlier [1]. Briefly: to find the *susC*-like genes, we used hidden Markov models (hmm) SusD.hmm, SusD-like_2.hmm and SusD-like_3.hmm obtained from protein families database (PFAM) [9] on all predicted proteins of the genome. Protein hmms are probabilistic models derived from multiple sequence alignment of related protein domains summarising the probabilities for a particular amino acid (or insertion) at the particular site of a protein domain. As such they are sensitive tools for general function prediction of amino acid sequence. All vs all SusD-like protein blastp search was then performed and SusD-like protein clusters were subsequently obtained using Markov cluster algorithm (mcl) [8], which essentially divides proteins into clusters based on their similarity. These clusters correspond to PULs with different target glycans as shown recently [1]. For each strain, the presence or absence of SusD-like protein clusters was examined thus yielding strain SusD-like protein/PUL profiles. We then assigned putative substrates to PULs by manually inspecting the PULs using Artemis [5] and comparing them to the experimentally verified PULs described in the related genus *Bacteroides* [25,27] using the blastp search. Protein families data from PFAM, CDD and dbCAN was used to evaluate the blastp results [9,24,50]. The above sequence-based PUL predictions were later tested using growth experiments. The CAZY annotation by dbCAN and ordination analysis using R packages vegan [33] and pvclust [45] was performed as described earlier [1].

Growth experiments

The strains were cultivated in Hungate tubes using the liquid modified DSMZ medium 330 [3] or the rumen fluid reinforced medium M2 [15] in the case of poor growth in M330 (Table S1). The strains were grown to the early stationary phase and then subcultured into media containing: (i) no sugars, (ii) glucose or (iii) one of various hemicellulose, pectin or storage polysaccharide substrates (Table S2). The polysaccharides were added during media preparation to 0.5% while stirring after the initial boiling and the media were then autoclaved. Strain growth was monitored spectrophotometrically at 600 nm for ten days, at least three times in separate experiments. The OD_{600} value of the tube containing no sugars was subtracted for every polysaccharide containing tube at each time point. The average maximum OD_{600} of every substrate was then normalized relative to growth on glucose for each strain by dividing by the maximum OD_{600} obtained on glucose. A heatmap without scaling was then produced in R.

Metagenome read classification

The DNA metagenome datasets of New Zealand methane sheep study [18,42], Scottish cattle methane and large scale metagenome binning study [44,46], plant fibre adherent bacteria study [14] and pig gut gene inventory study [49] were downloaded from EBI (http://www.ebi.ac.uk/). The metagenome reads were then classified by Kraken [48] using k-mer length of 31 and custom library made out of the genomes analysed in this study, genomes representative for different oral *Prevotella* nutritional groups [1], common hindgut *Bacteroidetes* and the *Prevotella* and *Bacteroidetes* metagenome derived genomes (genomic bins) obtained by the Hess and Stewart studies [14,44]. Since Kraken classifies reads on various taxonomical levels, we used Bracken [23] to estimate species abundance from Kraken result.

Results

CAZYme Prevotella *genomes analysis*

Our sample consisted of 50 *Prevotella* genomes encompassing in total eight hindgut, 15 rumen and two free living species according to ANIb metric (Table S1). The pig hindgut strains contributed five yet undescribed species while 11 tentatively novel ruminal species were recognized from the genomes sequenced by Hungate 1000 project [6]. The genomes were mainly of the expected size for the members of the genus *Prevotella*, i.e. 2.5–3.5 Mbp, except for the two pig hindgut strains harbouring slightly larger genomes of around four Mbp. The majority of species possessed high (>35) number of genes encoding the CAZYmes per Mbp (Table S3). Slightly lower was the number in *P. pectinovora* and *Prevotella stercorea* isolated from pig and human feces, respectively, while rumen strain HUN102 showed extensive reduction in CAZYme number. The CAZYme repertoires were compared using non-metric multidimensional scaling which included also oral, urogenital and clinical strains analysed previously [1] (Fig. S1, data in File S1). The rumen strains mainly grouped together with the exception of two strains: HUN102, which presumably belongs to the *Prevotella* group with reduced ability to degrade polysaccharides, and KH1P2 which is similar to oral strains utilizing both host and plant glycans [1]. Hindgut species appear to have rather variable CAZYme repertoires, yet some of them group with the ruminal species, *e.g. P. bryantii* and *Prevotella brevis*. Among the CAZYme families, some are monospecific, that is, they target only one class of substrate, starch or xylan for example. Exploiting this, preliminary predictions of capability to degrade several substrates commonly encountered by hindgut and rumen *Prevotella* can be made (Table S3).

PUL repertoires

In total more than 1100 SusD-like proteins were found to be encoded in genomes of hindgut/rumen/free living strains and were divided by mcl [8] into 157 groups that contained at least 2 members. The numbers encoded per genome varied widely (Table S3). Only moderate correlation among SusD-like protein and CAZYme number encoded per genome was found when all species including oral, urogenital and clinical were analysed (Pearson's r = 0,63), which is lower than reported earlier [1]. This was mainly caused by species harbouring many CAZYmes (>150) and low numbers of SusD-like proteins (<20), exemplified by strains 5-50, BPI-145, TF2-5, 3-122 and TC2-28. When these were omitted, the r was 0.74. The variability of SusD-like protein repertoires as deduced by pvclust and binary distance metric between strains can be seen in Fig. S2A. In general, the SusD-like protein repertoires are in accordance with groups seen on CAZYme NMDS plot (Fig. S1) or pvclust CAZYme dendrogram (Fig. S2B), and the *Prevotella* groups observed earlier [1] are evident. The group that harboured three rumen species in that study is now much expanded and it is noteworthy that it is split into two clusters, one of which includes *P. bryantii* (blue) and the other *Prevotella ruminicola* (green). These two clusters' members resemble quite closely those found in the CAZYme NMDS plot. Also the strains HUN102 and KH1P2 differ from other rumen strains in the manner very similar as seen for CAZYmes. The former may well be an oral strain recently introduced during rumination. Thus, CAZYme and SusD-like protein repertoires appear to be in agreement. Of all the 1100 SusD-like proteins encoded in genomes analysed, we managed to assign putative glycan binding partners and consequently define PULs to 545 of them. Only 35 of these were predicted to bind host glycans. Additional 104 were further identified as putatively binding to plant glycans, but exact nature of the latter is beyond (our) bioinformatic prediction capabilities. Of the top 20 mcl groups (encompassing 420 proteins) 15 are predicted reliably in our opinion: three target storage polysaccharides (starch and fructan), another three xylan, six various pectins, two β-glucan and one α-mannan. Of the remainder, three target unknown plant glycans and the targets of two are completely unknown. Table 1 gives the expected glycan degradation capability of analysed *Prevotella* strains based on the annotated PULs (detailed in Table S5, the SusD-like protein sequences are in Supplementary file 1). Overall, the ability to degrade storage polysaccharides is almost universal as reported earlier for oral species, yet as expected the rumen/hindgut *Prevotella* seem strongly oriented towards hemicellulose and pectin which is in stark contrast to the majority of oral species [1]. Rumen/hindgut *Prevotella* are not uniform, however, as there is much variability in PUL repertoires. Even between *P. bryantii* and *P. ruminicola*, which otherwise seem to be capable to degrade almost all major hemicellulose/pectin substrate classes, the PUL repertoires differ: *P. bryantii* lacks the α-mannan PUL, for example. Even more pronounced is the PUL variability of rumen/hindgut species in the case of *P. brevis*, *P. stercorea*, *P. pectinovora* and the undescribed species represented by the strain 5-64, which all seem to possess a narrow specialist PUL repertoires. We have often found several PULs targeting one substrate class (xylan, β-glucan, gluco/galactomannan, rhamnogalacturonan, homogalacturonan) in the same strain which is a situation similar to the one described in genus *Bacteroides* [25,27,36,43]. On few occasions hitherto unknown PUL diversity was also found: e.g. we identified three starch PULs (Table S5) coding for almost identical CAZYme content (2x GH13, GH97, GH97/CBM20) but very different SusD-like binding proteins which displayed less than 40% amino acid identity (also clearly separated in phylogeny reconstruction dendrograms). While one starch PUL variant was dominant (18 species), the other two were found in six species of mixed rumen and gut origin. These additional starch PULs are not the same as putative accessory starch PULs previously identified [1] of which one is strongly represented also in rumen/gut *Prevotella* (Table S5, mcl cluster 9). Similarly, a third, novel xylan PUL was identified in rumen species represented by the strain BPI-145.

Strain growth in media with various plant polysaccharides

The growth on both M330 and M2 media containing no sugars was very poor with maximum OD_{600} less than 0.1 and 0.2, respectively, while all strains achieved OD_{600} of above 1 for at least one substrate. Only strains HUN102 and 5-64 grew to OD_{600} of 0.25 without the added carbohydrates. The strains from the genus *Bacteroides* which served as controls behaved as reported previously [27] for all substrates tested. The average growth and predicted degradation capabilities are shown in Table 2. Overall, we find good correspondence (84% accurate) between bioinformatic prediction and observed growth, especially for starch and

Table 1
Overview of the PUL repertoires found in studied *Prevotella* strains. The strains grouped by a thick border are representatives of the same species or yet undescribed species, while others are the sole representatives of their species. ++, more than one SusD-like protein for specific substrate present, effectors (CAZYmes, regulator, transporter) forming clearly recognizable PUL in at least one case; +, as ++ but only one SusD-like protein; +/−, either SusD-like protein or clearly recognizable effectors (anywhere in the genome) are present, but not both; blank, no SusD-like protein or effectors.

substrate / strain	starch	fructan	xylan	β-glucan	xyloglucan	gal/glu mannan	α-mannan (yeast)	homogalacturonan	rhamnogalacturonan	arabinan	arabinogalactan	dextran	host glycans
3-122	+	+	+/-	+				++	+			+/-	
P. brevis ATCC 19188	+	+	+/-				+	++	+				
P. brevis P6B11	+	+	+/-				+	++	+				
4-51	+								+/-			+	++
5-108	+								+/-			+	++
4-67	+								+/-			+	++
5-64	+								+/-			+	++
5-119	+	+						++	+				
5-125	+	+						++	+				
4-65	+	+						++	+				
4-76	+	+						++	+				
5-60	+	+						++	+				
P. albensis DSM 11370	+	+	++	+						++		+	+/-
AGR2160	+	+		+	+	+		+		+			
P. copri DSM 18205	+	+	++	++	+			++		+			
P. oryzae DSM 17970	+		++	++				++	++	+	+/-	+	+/-
P. paludivivens DSM 17968	+	+	++	++		+		++	++	++	++	+	
5-92	+	+	+	+	+	+		+	+	+			
2-180	+	++	++	+	+	+		+	+	++			
TC2-28	++	++	+	+	+	+/-		++	+				
TF2-5	+	+	++	+				+	+	+			
3-120	+/-		++			+		+	+				
4-98	+/-		++			+		+	+				
3-92	+/-		++			+		+	+				
4-119	+/-		++			+		+	+				
5-126	+/-		++			+		+	+				
5-50	+/-		++			+		+	+				
P. bryantii B14	+	+	++		+	+		++	++	++	+	+	
P. bryantii C21a	+	+	++		+	+		++	++	++	+		+/-
P. bryantii TC1-1	+	++	++		+	+		++	++	++	+		+/-
FB3001	+	+	++		+	+		++	++	++	+		+/-
BPI-162	+	++	++	++		+	++	++	++	+	+		+
BPI-34	+	++	++	++		+	++	++	++	+	+		+
KHP1	+	++	++	++			++	++	++	++	+		+
P. ruminicola 23	+	+	++	++			++	++	++	+	++		+
Ga6B6	+	+	++	++		+	++	++	++	+	++		+
RM4	+	+	++	++			++	++	++	+	+		+
MA2016	+	++	++	++	+	+	++	++	++	+	+		+
NE3005	+	++	++	+		+/-		+	+	+	++		
P6B1	+	+	++	++		+	++	++	++	+	+		
P6B4	+	+	++	++		+/-	++	++	++	+	+		
KHP7	+	+	++	++			++	+	+		++		+
FD3004	+	+	++	++			++	+	+		++		+
LC2012	+	+	++	++				+			++		+
TC2-24	+	+	++	++				+			++		+
BP1-145		+	+					+	+				
BP1-148		+	+					+	+				
HUN102	+	+		+									
KH1P2	+	++		++			++	+				+	+
P. stercorea DSM 18206	+	+											

xylan, whereas there were false negative results observed for β-glucan and especially xyloglucan. This may be the result of one of the following: (i) protein sequences are not conserved enough to transfer the data from *Bacteroides* transcriptomic profiling or the PUL is fragmented in *Prevotella*, (ii) some PULs of *Prevotella* may not be present in *Bacteroides*, and (iii) there are no monospecific CAZYme families whose members would act solely on these substrates to use as markers. On the other hand, false positives were rare implying the PUL annotation was conservative and they were concentrated mainly in the α-mannan PUL. In this case, four strains

Table 2
Assessment of the bioinformatic prediction accuracy. The first column in each substrate class indicates bioinformatic prediction according to the same legend as in Table 1 (for readability – substitutes blank here), while an average growth (OD_{600}) is given in the second column (see Table S6 for details). The second column may also state further details of substrates used in growth experiments. The growth results deviating from predictions are underlined and in bold. *P. bryantii* TF1-3 was included as a control, because of growth variability of *P. bryantii* TC1-1 in medium containing beechwood xylan.

substrate / strain	starch	soluble	fructan	inulin	xylan	beechwood/ arabino	β-glucan		xyloglucan		gal/glu mannan		α-mannan		homo galacturonan		rhamno galacturonan	
3-122	+	1.12	+	0.47	+/-	0/0.86	+	0.63	-	0	-	0/0	-	0	++	0.16	+	0.64
P. brevis ATCC19188	+	1.42	+	1.32	+/-	0.13/0.12	-	0	-	0	-	0/0	+	1.39	++	0.29	+	1.05
P. brevis P6B11	+	1.61	+	0.85	+/-	0.08/0.14	-	0	-	0	-	0/0	+	1.42	++	0.36	+	0.83
5-64	+	0.99	-	**1.34**	-	0/0	-	0	-	0	-	0/0	-	0	-	**0.14**	+/-	0
4-65	+	1.01	+	1.01	-	0/0	-	0	-	0	-	0/0	-	0	++	**0**	+	0.22
AGR2160	+	1.58	+	1.59	-	0/0	+	1.76	+	1.74	+	1.61/1.27	-	0	+	1.1	-	**0.43**
2-180	+	1.33	++	1.09	++	0.92/1.17	+	1.21	+	0.94	+	1.49/1.06	-	0	+	0.83	+	0.69
TC2-28	++	1.01	++	1.1	+	0/0.84	+	0.77	+	0.7	+/-	0.67/0.71	-	0	++	0.7	+	0.75
TF2-5	+	1.6	+	1.58	++	0.99/1.42	+	1.53	-	**1.27**	-	**1.5/1.01**	-	0	+	**0**	+	**0**
5-50	+/-	0	-	**0.68**	++	0.68/1.16	-	**1.15**	-	**1.1**	+	1.12/0.8	-	0	+	0.67	+	0.19
P. bryantii B14	+	1.54	+	1.35	++	0.67/1.24	-	**1.48**	+	1.43	+	0.58/1.05	-	0	++	1.09	++	0.81
P. bryantii TC1-1	+	1.24	++	1.39	++	0/0.98	-	**1.44**	+	1.15	+	1.12/0.98	-	0	++	0.83	++	0.87
P. bryantii TF1-3		1.42		1.53		1.07/1.45		1.71		1.63		1.32/1.23		0		1.02		0.98
KHP1	+	1.66	++	1.72	++	1.03/1.42	++	1.75	-	**1.8**	-	**0.2/1.32**	++	**0**	+	0.96	++	0.21
Ga6B6	+	1.73	+	1.34	++	0.77/1.34	++	1.83	-	**1.76**	+	0.25/1.06	++	**0**	+	0.67	++	0.25
MA2016	+	1.42	++	1.55	++	1.12/1.28	++	1.63	+	1.5	+	1.25/1.12	++	**0**	++	0.64	++	0.19
NE3005	+	1.24	++	1.05	++	0.23/0.64	+	0.9	-	**0.73**	+/-	0.7/0.22	-	0	+	0.43	+	0.28
KHP7	+	1.69	+	1.63	++	0.96/1.53	++	1.96	-	**1.59**	-	**0.48/0.85**	++	**0**	+	0.59	+	0.33
HUN102	+	0.94	+	1.54	-	0/0.23	+	**0**	-	0	-	0/0	-	0	-	**0.18**	-	0
KH1P2	+	0.97	++	0.95	-	0/0	++	0.92	-	**0.59**	-	**0.92/0.71**	++	0.21	+	0.16	-	0
B. thetaiotaomicron		0.85		0.91		0		0		0		0/0		0.7		0.61		0.33
B. ovatus		0.86		0.78		0.31/0.78		0.74		0.66		0.76/0.48		0.25		0.63		0.66

True positives	18	17	12	11	6	9	3	15	15
False positives	0	0	0	**1**	0	0	**4**	2	**1**
True negatives	1	0	7	4	6	6	12	0	2
False negatives	0	**2**	0	**3**	**7**	**4**	0	**2**	**1**

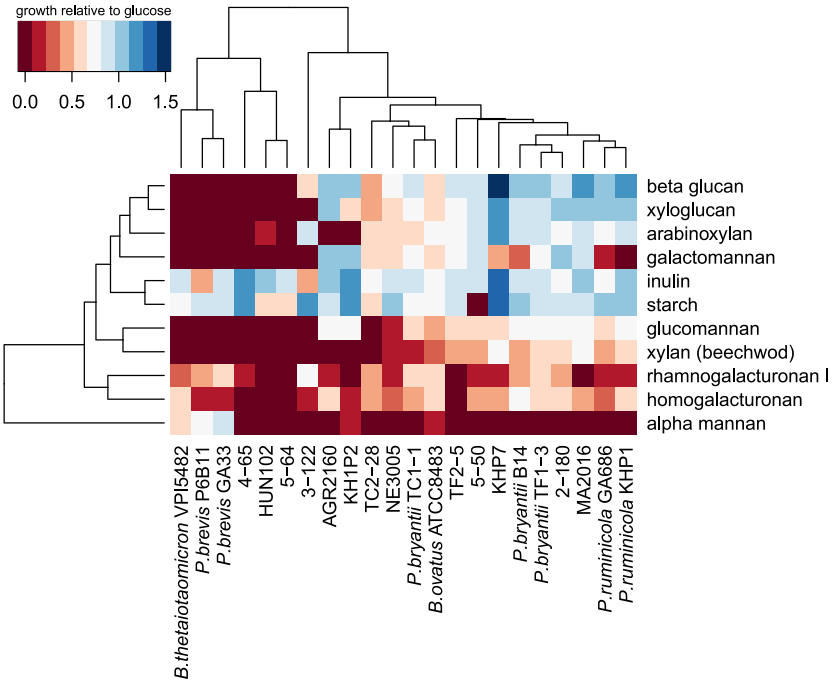

Fig. 1. Growth of *Prevotella* and *Bacteroides* strains measured as OD_{600} in media containing various polysaccharide substrates normalized to growth in glucose (value of 1 indicates the same OD_{600}, i.e., 100%).

were predicted to degrade α-mannan based on the presence of two or even three PULs, yet failed to grow on yeast α-mannan. We propose that these species may in fact use mannan different from the one found in yeast, possibly α-mannan of ruminal origin, e.g. from anaerobic fungi. Growth using various substrates was normalized to growth on glucose and is presented in Fig. 1 (data in Table S6). As expected, storage polysaccharides were utilized almost universally with the (predicted) exception of strain 5-50. Overall, strains were divided into two groups: one which could not utilize any of the hemicelluloses tested and the other whose members were gener-

ally proficient, yet at the same time diverse in the hemicellulose utilization. Both groups were able to utilize pectin, but none of the strains grew on pectin as well as seen on hemicelluloses or glucose. Beechwood xylan also proved to be harder to utilize than arabinoxylan as all the strains grew poorer in the former and some strains even showed no growth on it while growing well on arabinoxylan (TC2-28, 3-122, TC1-1). Finally, some diversity was evident within the species in *P. bryantii* despite almost identical repertoires of CAZYmes and SusD-like proteins.

Presence of Prevotella *species analysed in this study in rumen metagenomes*

We were interested how common the described and tentative new *Prevotella* species of rumen and hindgut origin (our sample) may be in the rumen: whether they are detectable but rare, or attain shares comparable to those seen for dominant bacteria. To approximate this, we used metagenome data from several recent rumen and pig studies [18,42,44,46,49]. In the Stewart cattle study, hundreds of metagenome data derived bacterial genomes were retrieved with high completeness. 67 of them, which were identified as belonging to the genus *Prevotella* [44], were used as markers for highly abundant bacteria. By using the ANIb metric we noted that some of these genomes actually belonged to the same species thus resulting in 40 putative novel uncultured *Prevotella* species, while two of the genomes actually belonged to gut dwelling *Prevotella copri* and the undescribed rumen species represented by *Prevotella* strain NE3005 from our sample (Table S7). We then classified the metagenome reads using Kraken/Bracken against the custom library made of our sample, "Stewart *Prevotella* genomes", and common gut *Bacteroidetes* of the genera *Bacteroides* (including rumen strains), *Parabacteroides*, *Odoribacter*, *Barnesiella* and *Alistipes*. The total results are in Table S8 while the summary can be seen in Table 3 and Fig. 2. Several conclusions can be made: (i) the "Stewart *Prevotella* genomes" are commonly found in high proportions in all of the above metagenome studies and may therefore be a good representation of hitherto uncultured *Prevotella* diversity (primarily) in the rumen. (ii) the *Prevotella* share in all metagenomes is significant and the range of medians is 7.9–11.3% with maxima up to around 30% of reads in some samples. The medians obtained for *Prevotella* in rumen are roughly one fifth relative to those obtained by Stewart et al. [44] using the same Kraken approach for *Bacteria* and *Archaea* (including all of the Hungate1000 genomes and 913 of their genomic bins) on their own metagenome data and on data of the New Zealand sheep study [18,42]. The *Prevotella* shares are notably low, around 2%, in Chinese pig gut samples (excluding the Ba Ring breed) where antibiotics were added in daily feed [49]. (iii) the ratio of metagenome reads classified to cultured *Prevotella* strains (our sample) versus the uncultured "Stewart *Prevotella* genomes" is around 0.3 for rumen and 0.66 for the sheep metagenome implying that although information gained from our sample is meaningful, much *Prevotella* functional diversity may still be uncovered even in the rumen. (iv) Although *Prevotella* species that were isolated (or their genomes computationally derived) from the rumen and gut dominate their respective metagenomes, there are metagenome samples, especially in the Stewart et al. study [44], where reads classified to pig hindgut strains isolated in 2014 in Slovenia predominate and may reach more than 7%. These strains (5-50, 3-122, but *P. copri*, too) are also consistently found in rumen samples of New Zealand sheep [18,42] and Scottish cattle [46]. Furthermore, the "Stewart *Prevotella* genomes" RUG457 and RUG782, which may be genomes of the same species according to ANIb and provisionally termed R10 here, are actually much more common in the pig metagenome samples [49] than in the rumen where they were collected. Interestingly, the reads belonging to genomes RUG101, RUG753 and R8 affiliated RUG073 and RUG541, are sporadically found in both the pig hindgut and rumen metagenomes in shares over 1%. A simple control was carried out for rumen sample ERR2027904 [44], where 7.7 and 1.2 percent of all reads classified to pig and human hindgut species represented by strain 5-50 and *P. copri* respectively. The reads recruited to these genomes were indicated from the Kraken result using kraken-translate of the Kraken package [48], extracted from the metagenome reads and mapped to *P. copri* DSM18205 and 5-50 genomes using Bowtie 2 [21]. The metagenome reads mapped evenly along the genomes indicating that high classification percentages indeed did not arise from some sort of multiplying horizontal transfer element but are rather the reflection of the ubiquity of the whole genome in the sample.

The metagenomes analysed above were derived from cow digesta [44,46] and sheep rumen content obtained using stomach intubation [18], which both apparently contain liquid and particulate fractions. In contrast, the Hess et al. study [14] obtained cattle rumen metagenomes of the switchgrass adherent microbes after three day incubation. In that study, five *Bacteroidetes* genomic bins of high coverage, second only by small margin to two *Clostridiales* bins, were discovered [14]. We estimated the share of these genomic bins in the Hess metagenome reads and compared them to shares of *Prevotella* from our sample and other *Bacteroidetes* as above. We found that the rumen tentative species represented by strains LC2012 and 5-50 from our sample and one of the "Stewart *Prevotella* genomes", RUG751, had comparable shares suggesting they may be efficient switchgrass late colonizers (Table S9).

Discussion

While polysaccharides of the plant feed entering the rumen are complex and diverse, it seems plausible, given the number of species found in the rumen, that capability for its degradation is found in many bacterial species (is redundant) and mainly competitive interactions may shape the polysaccharide degrading microbial community [47]. On the other hand, the wealth of different monomers, linkages and substituents making up the polysaccharides and further their inhomogeneity across plant taxa and their tissues are being appreciated lately along with the specialization of the enzymatic systems breaking them down [26]. Thus, not only competition but also niche specialization may drive the genome evolution of the major rumen and hindgut plant glycan degraders, such as *Prevotella*. We observed in fact much diversity in glycan degrading potential among *Prevotella* genomes analysed in this study on several levels. First, the classic specialist/generalist dichotomy is exhibited by *P. brevis*, *P. pectinovora*, *P. stercorea* and tentative species represented by strain 5-64 versus other *Prevotella* species where the former cannot use any of the hemicelluloses but only storage polysaccharides, pectin or yeast mannan. The generalist *Prevotella* species were exceedingly competent at hemicellulose breakdown, frequently growing on all substrates tested (apart from α-mannan) and the growth occasionally surpassed that seen on glucose (Fig. 1). Also, the ability to degrade xyloglucan, which is rare in genus *Bacteroides* [22], was routinely seen in growth experiments while less so in bioinformatic prediction of xyloglucan PULs (probably owing to an unsure annotation of a single PUL, see Table S5, unidentified plant glycan sheet, 8. mcl cluster). Second, there was variation among generalists also, e.g. strain AGR2160 was not predicted and it actually did not grow on xylan while being capable of using many other hemicelluloses. Similarly, tentative new species represented by strains 5-50 and BPI-145 lacked capability to degrade starch, which is otherwise almost universally used by members of the genus *Prevotella*. Third, several PULs dedicated to the degradation of one major substrate class were often present in a singular species, yet the combination of PULs differed between species, e.g. *P. bryantii* and *P. ruminicola* possess markedly differ-

Table 3
The mean, median and maximum percentages of classified metagenomic reads in the samples of three major rumen and a pig metagenome study. The hindgut and rumen *Prevotella* are coloured brown and green respectively while the rest are the metagenome derived *Prevotella* genomes obtained by Stewart et al. [44] which were delineated into species using ANIb (R1–R18). (For interpretation of the references to colour in this table legend, the reader is referred to the web version of this article.)

	Wallace et al., 2015 cattle rumen		Shi et al., 2014 sheep rumen		Stewart et al., 2018 cattle rumen		Xiao et al., 2016 pig feces		Wallace	Shi	Stewart	Xiao
	average	median	average	median	average	median	average	median	max	max	max	max
Prevotella	10.16	9.71	8.67	8.12	13.10	11.29	8.66	7.90	20.71	13.63	28.07	33.24
P. copri	0.11	0.02	0.02	0.02	0.12	0.02	2.41	1.21	0.57	0.03	1.77	13.55
2-180	0.03	0.03	0.04	0.04	0.11	0.03	0.60	0.32	0.04	0.05	1.03	4.58
3-122	0.04	0.04	0.06	0.04	0.32	0.19	0.55	0.37	0.06	0.32	1.68	2.40
4-65 (*P. pectinovora*)	0.02	0.02	0.03	0.03	0.02	0.02	0.61	0.40	0.02	0.04	0.05	2.94
5-50	0.06	0.05	0.06	0.06	0.35	0.04	0.92	0.59	0.08	0.10	7.66	9.73
5-64	0.04	0.04	0.04	0.04	0.05	0.05	0.18	0.13	0.05	0.05	0.16	0.66
5-92	0.02	0.02	0.04	0.04	0.03	0.03	0.55	0.38	0.03	0.05	0.05	3.96
P. stercorea	0.03	0.02	0.04	0.04	0.03	0.02	0.42	0.37	0.04	0.05	0.05	1.56
P. albensis	0.06	0.05	0.03	0.03	0.03	0.02	0.02	0.01	0.17	0.05	0.07	0.06
P. brevis	0.12	0.10	0.15	0.14	0.11	0.10	0.02	0.02	0.26	0.21	0.25	0.04
P. bryantii	0.07	0.08	0.31	0.09	0.11	0.09	0.03	0.03	0.11	2.22	0.65	0.10
P. ruminicola	0.31	0.32	0.52	0.40	0.30	0.26	0.02	0.02	0.61	1.08	0.88	0.05
AGR2160	0.06	0.03	0.02	0.02	0.06	0.02	0.02	0.01	0.18	0.03	0.51	0.16
HUN102	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.03
KH1P2	0.01	0.01	0.04	0.04	0.02	0.02	0.01	0.01	0.02	0.08	0.05	0.05
MA2016	0.05	0.05	0.11	0.10	0.05	0.05	0.01	0.01	0.08	0.26	0.11	0.03
NE3005	0.12	0.10	0.20	0.16	0.22	0.17	0.01	0.01	0.24	0.84	1.46	0.02
TC2-28	0.16	0.15	0.54	0.37	0.25	0.19	0.01	0.01	0.34	1.52	0.72	0.02
TF2-5	0.08	0.08	0.46	0.39	0.18	0.15	0.01	0.01	0.16	1.52	1.07	0.04
BPI-145, BPI-148	0.39	0.21	0.14	0.14	0.32	0.27	0.01	0.01	1.01	0.27	1.65	0.03
KHP7, FD3004	0.06	0.06	0.07	0.08	0.06	0.07	0.01	0.01	0.12	0.12	0.12	0.02
LC2012, TC2-24	0.05	0.05	0.07	0.07	0.08	0.07	0.01	0.01	0.12	0.11	0.56	0.02
P6B1, P6B4	0.10	0.10	0.12	0.11	0.09	0.08	0.01	0.01	0.18	0.20	0.19	0.02
R1	0.07	0.07	0.12	0.12	0.21	0.12	0.03	0.03	0.10	0.27	0.69	0.06
R10	0.13	0.07	0.04	0.04	0.17	0.06	0.91	0.55	0.32	0.06	1.75	4.07
R12	0.40	0.40	1.07	0.50	0.60	0.42	0.01	0.01	0.85	4.62	3.41	0.02
R13	0.05	0.02	0.02	0.02	0.06	0.01	0.01	0.01	0.20	0.03	1.10	0.06
R14	0.48	0.24	0.25	0.16	0.77	0.30	0.02	0.01	1.84	1.13	6.53	0.13
R15	0.44	0.28	0.60	0.43	0.65	0.47	0.01	0.01	1.58	2.03	2.87	0.03
R16	0.33	0.22	0.76	0.44	0.65	0.38	0.01	0.01	0.67	2.39	5.92	0.03
R17	0.21	0.14	0.33	0.32	0.50	0.36	0.01	0.01	0.46	0.58	2.95	0.03
R18	0.97	0.15	0.04	0.04	0.97	0.07	0.06	0.03	3.83	0.05	12.53	0.82
R2	0.31	0.16	0.17	0.13	0.16	0.08	0.02	0.02	1.19	0.57	1.09	0.04
R3	0.15	0.14	0.24	0.19	0.29	0.24	0.02	0.02	0.26	0.75	1.26	0.07
R4	0.06	0.05	0.09	0.08	0.09	0.04	0.02	0.02	0.13	0.20	0.59	0.10
R5	0.53	0.35	0.07	0.07	0.49	0.11	0.09	0.05	1.41	0.12	4.67	0.67
R6	0.03	0.03	0.16	0.08	0.10	0.05	0.01	0.01	0.05	0.98	0.64	0.03
R7	0.07	0.07	0.06	0.06	0.12	0.08	0.01	0.01	0.14	0.11	0.81	0.02
R8	0.36	0.04	0.03	0.03	0.29	0.03	0.14	0.02	1.67	0.04	5.92	1.53
R9	0.20	0.06	0.02	0.02	0.21	0.02	0.04	0.01	0.52	0.02	7.66	0.70
RUG069	0.09	0.07	0.11	0.12	0.22	0.10	0.01	0.01	0.28	0.19	3.57	0.02
RUG101	0.01	0.01	0.01	0.01	0.15	0.01	0.08	0.01	0.03	0.02	2.86	1.01
RUG106	0.06	0.02	0.03	0.02	0.03	0.02	0.01	0.01	0.15	0.07	0.20	0.03
RUG116	0.62	0.30	0.11	0.11	0.81	0.50	0.01	0.01	2.09	0.18	3.58	0.02
RUG137	0.13	0.05	0.03	0.03	0.07	0.02	0.03	0.01	0.43	0.04	0.59	0.24
RUG212	0.05	0.02	0.01	0.01	0.06	0.01	0.03	0.01	0.27	0.01	1.03	0.64
RUG273	0.06	0.03	0.19	0.10	0.21	0.08	0.01	0.01	0.18	1.33	1.91	0.03
RUG301	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.23	0.02
RUG430	0.04	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.12	0.02	0.07	0.05
RUG473	0.90	0.43	0.19	0.13	0.92	0.56	0.01	0.01	2.64	1.17	5.53	0.02
RUG536	0.00	0.00	0.02	0.01	0.01	0.01	0.00	0.00	0.01	0.03	0.04	0.01
RUG647	0.14	0.11	0.14	0.14	0.20	0.17	0.01	0.01	0.36	0.23	0.90	0.02
RUG751	0.03	0.01	0.05	0.02	0.05	0.02	0.01	0.01	0.10	0.27	0.65	0.02
RUG753	0.57	0.19	0.02	0.02	0.57	0.05	0.31	0.02	2.77	0.03	5.73	3.50
RUG759	0.44	0.21	0.04	0.04	0.13	0.05	0.03	0.02	1.10	0.05	1.12	0.28
RUG765	0.08	0.07	0.07	0.07	0.10	0.08	0.01	0.01	0.20	0.14	0.73	0.01
RUG827	0.04	0.03	0.16	0.12	0.10	0.06	0.01	0.01	0.11	0.46	1.03	0.03
RUG838	0.04	0.04	0.07	0.05	0.04	0.03	0.00	0.00	0.08	0.17	0.48	0.01
RUG839	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.08	0.02	0.10	0.01
hRUG878	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.02	0.05	0.01	0.08	0.15
hRUG905	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.09
hRUG908	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.11

ent PULs that target homogalacturonan and rhamnogalacturonan (Table S5). Finally, in the case of starch degradation, there is usually only one PUL per *Prevotella* genome, yet we have found three variants encoding very different SusD-like starch binding proteins. Regarding the last two points, it was shown previously that PUL variants may be specialized for specific representatives of broad substrate classes, for example inulin vs levan in the case of fructan, and glucuronoxylan vs more complex arabinoglucoronxylan for xylan. This specificity is conferred jointly by SusD-like binding proteins and degradative enzymes [36,43]. Thus, there is plenty of evidence for niche specialization already on the genomic level. It has to be stressed, however, that an additional layer of niche spe-

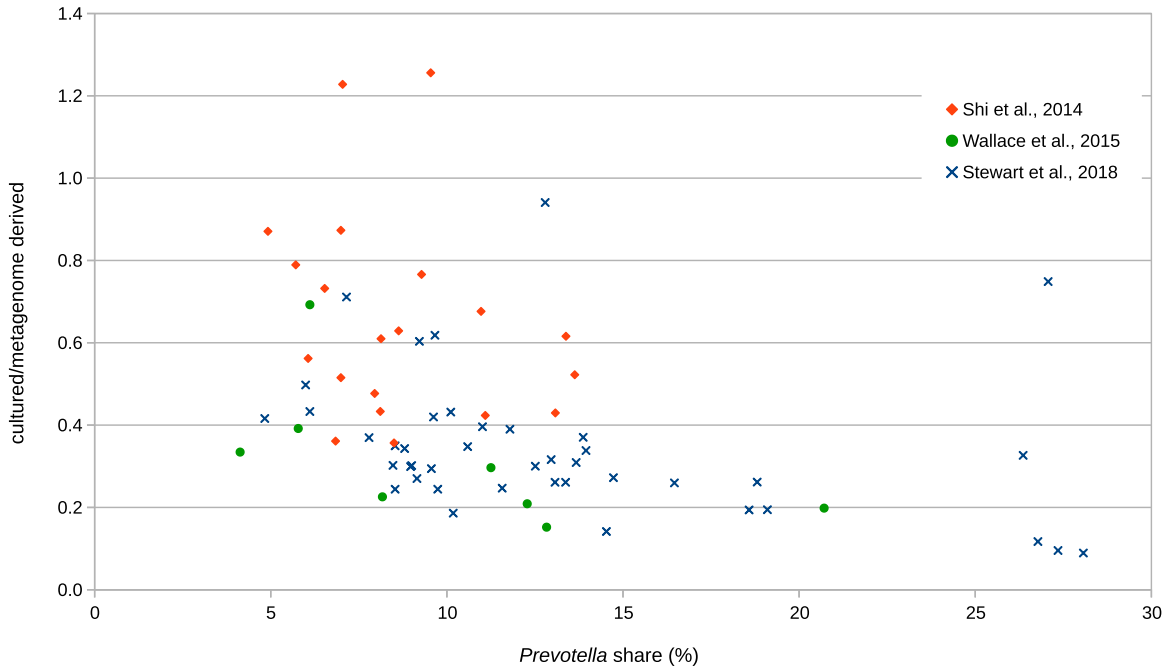

Fig. 2. Overview of the shares of metagenome reads classified as *Prevotella* in the samples of three major rumen studies [42,44,46] originating in Scotland and New Zealand. The metagenome derived *Prevotella* genomes obtained in the study of Stewart et al. [44] are present in all three studies and are usually predominating (vertical axis).

cialization exists: the prioritization of some glycans over others as has been observed in *B. thetaiotaomicron* [35], which is likely to exist also in *Prevotella*. Taken together this suggests that there is solid foundation for many *Prevotella* species to coexist in the rumen or gut at the same time. Interestingly, although the generalist species (*P. ruminicola* and species represented by strains BPI-145, TC2-28 and TF2-5) dominated among cultured *Prevotella* in the studied rumen metagenomes, the specialist *P. brevis* was not far behind. Yet the metagenome data also suggests that although *Prevotella* species analysed in this study may be an important component of rumen microbiota and the knowledge gained from them useful, much of the *Prevotella* diversity remains unexplored as the cultured *Prevotella* were outnumbered according to Kraken read classification by a an average factor of 3.3 and 1.5 in cattle and sheep metagenomes respectively by the *Prevotella* genomic bins recovered by the Stewart study (Fig. 2) [44]. As mentioned in the "Results" section, the latter may indeed correspond to real species waiting to be isolated and together with the cultured ones make up one fifth (regarding the medians) of the total classifiable *Bacteria* and *Archaea* in two major rumen studies [42,44] by using the Kraken approach. Interestingly, this number is comparable to the *Prevotella* shares obtained in the worldwide rumen 16S rRNA gene metagenomic sequencing [13]. *Prevotella* of our sample and one of the "Stewart *Prevotella* genomes" were also common in the switchgrass adherent bacteria metagenome [14] adding weight to a notion that some *Prevotella* may in fact colonise the insoluble fibre [28]. Interestingly, the dominant species in the adherent metagenome, represented by strains 5-50, LC2012 and RUG751, were not dominant in other rumen metagenomes analysed, which were derived mostly from (though probably not exclusively) the rumen fluid. The major finding of this study is also the apparent ease of transfer and sporadic establishment in the cattle rumen of *Prevotella* species originally isolated in the hindgut of humans and pigs. *P. copri* and the strains isolated in 2014 in Slovenia from pig feces, for example, were among or even the most dominant *Prevotella* species in some cattle metagenome samples reaching almost 10% of all reads in one case. Further, one of the metagenome bins found by Stewart et al. [44] actually belongs to *P. copri* and another is much more abundant in pig metagenome samples [49] than in the rumen. That the transfer of *Prevotella* and its relatives may proceed efficiently among the anaerobic habitats they usually inhabit in the same host was showed in fact indirectly many years ago by the Salyers' group while monitoring *tetQ* alleles from rumen and oral *Prevotella* and colonic *Bacteroides* [31]. There, the practically identical alleles were found in oral *Prevotella intermedia* and colonic *Bacteroides fragilis* strain "*with ease, using relatively small number of isolates*" as the authors reported. Thus, if animals are cohoused, e.g. pigs and cattle, the transmission may be possible between pig hindgut and rumen, explaining the above findings. In the case of human species *P. copri* this raises concern as there may not exist any major species barriers in spreading of the antimicrobial resistance genes between humans, pigs and domestic ruminants via the *Bacteroidales* species and their numerous conjugative transposons [39].

The growth of the *Prevotella* strains on various polysaccharides was in agreement (not complete, see results) with the bioinformatic prediction thus largely validating it. PULs, which seemed to be involved in yeast mannan degradation as judged from their *Bacteroides* counterparts have probably evolved for a distinct type of this glycan in ruminal *Prevotella*. In addition, *P. pectinovora*, which contains many PULs and numerous CAZYmes supposedly involved in pectin breakdown (Tables S3 and S5), grew surprisingly poorly on rhamnogalacturonan and not at all on homogalacturonan prompting that wider range of substrates should be used to assess the degradation potential. The bioinformatic prediction is thus apparently not precise enough in certain cases as some PULs may be very specific. The results presented above are an attempt to gain a broader perspective on rumen/hindgut *Prevotella* species polysaccharide degradative ability, which is in our opinion one of major foundations why these somewhat less researched bacteria feature prominently in almost every gut or rumen study using NGS. Yet, given the above *Prevotella* adeptness to grow well on various substrates which may actually greatly overlap that of the *Bacteroides* species, other factors favouring one or the other genus in anaerobic habitats may play key roles in determining their relative shares, e.g. like bile acids in the hindgut [30]. In addition, these results may guide more detailed and focused analyses of *Prevotella* PULs

and given that the feed entering the rumen is essentially not predigested as in the case of the hindgut, PULs targeting novel substrates may still be uncovered in *Prevotella*.

Acknowledgements

This work was supported by the Slovenian Research Agency (research project P4-0097). Some *Prevotella* strains were obtained from the AGresearch (NZ) as stated in methods section. For them, the following applies: cultures were obtained from the Hungate1000 culture collection hosted by AgResearch Limited and kindly provided by Sinead Leahy and Kerri Reilly. The Hungate1000 project was funded by the New Zealand Government in support of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases. The Hungate1000 community sequencing project (CSP612) was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.syapm.2018.10.001.

References

[1] Accetto, T., Avguštin, G. (2015) Polysaccharide utilization locus and CAZYme genome repertoires reveal diverse ecological adaptation of *Prevotella* species. Syst. Appl. Microbiol. 38 (7), 453–461, http://dx.doi.org/10.1016/j.syapm.2015.07.007.

[2] Accetto, T., Avguštin, G. (2011) Inability of *Prevotella bryantii* to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in *Bacteroidetes*. PLoS One 6 (8), e22914, http://dx.doi.org/10.1371/journal.pone.0022914.

[3] Accetto, T., Avgustin, G. (2007) Studies on *Prevotella* nuclease using a system for the controlled expression of cloned genes in *P. bryantii* TC1-1. Microbiol. Read. Engl. 153 (Pt 7), 2281–2288, http://dx.doi.org/10.1099/mic.0.2007/005850-0.

[4] Avgustin, G., Wallace, R.J., Flint, H.J. (1997) Phenotypic diversity among ruminal isolates of *Prevotella ruminicola*: proposal of *Prevotella brevis* sp. nov., *Prevotella bryantii* sp. nov., and *Prevotella albensis* sp. nov. and redefinition of *Prevotella ruminicola*. Int. J. Syst. Bacteriol. 47 (2), 284–288, http://dx.doi.org/10.1099/00207713-47-2-284.

[5] Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B.G., Parkhill, J., Rajandream, M.-A. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics (Oxford, England) 24 (23), 2672–2676, http://dx.doi.org/10.1093/bioinformatics/btn529.

[6] Creevey, C.J., Kelly, W.J., Henderson, G., Leahy, S.C. (2014) Determining the culturability of the rumen bacterial microbiome. Microb. Biotechnol. 7 (5), 467–479, http://dx.doi.org/10.1111/1751-7915.12141.

[7] Dodd, D., Moon, Y.-H., Swaminathan, K., Mackie, R.I., Cann, I.K.O. (2010) Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic *Bacteroidetes*. J. Biol. Chem. 285 (39), 30261–30273, http://dx.doi.org/10.1074/jbc.M110.141788.

[8] van Dongen, S., Abreu-Goodger, C. (2012) Using MCL to extract clusters from networks, in: van Helden, J., Toussaint, A., Thieffry, D. (Eds.), Bacterial Molecular Networks, vol. 804, Springer New York, New York, NY, pp. 281–295.

[9] Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M. (2014) Pfam: the protein families database. Nucleic Acids Res. 42 (D1), D222–D230, http://dx.doi.org/10.1093/nar/gkt1223.

[10] Foley, M.H., Cockburn, D.W., Koropatkin, N.M. (2016) The Sus operon: a model system for starch uptake by the human gut *Bacteroidetes*. Cell. Mol. Life Sci. 73 (14), 2603–2617, http://dx.doi.org/10.1007/s00018-016-2242-x.

[11] Gasparic, A., Martin, J., Daniel, A.S., Flint, H.J. (1995) A xylan hydrolase gene cluster in *Prevotella ruminicola* B(1)4: sequence relationships, synergistic interactions, and oxygen sensitivity of a novel enzyme with exoxylanase and beta-(1,4)-xylosidase activities. Appl. Environ. Microbiol. 61 (8), 2958–2964.

[12] Gorvitovskaia, A., Holmes, S.P., Huse, S.M. (2016) Interpreting *Prevotella* and *Bacteroides* as biomarkers of diet and lifestyle. Microbiome 4, 15, http://dx.doi.org/10.1186/s40168-016-0160-7.

[13] Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Global Rumen Census Collaborators, Janssen, P.H. (2015) Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Sci. Rep. 5, 14567, http://dx.doi.org/10.1038/srep14567.

[14] Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z., Rubin, E.M. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331 (6016), 463–467, http://dx.doi.org/10.1126/science.1200387.

[15] Hobson, P.N. (1969) Rumen bacteria, in: Norris, J.R., Ribbons, D.W. (Eds.), Methods in Microbiology, vol. 3, Academic Press, pp. 133–149.

[16] Hofmann, R.R. (1989) Evolutionary steps of ecophysiological adaptation and diversification of ruminants: a comparative view of their digestive system. Oecologia 78 (4), 443–457, http://dx.doi.org/10.1007/BF00378733.

[17] Jami, E., Israel, A., Kotser, A., Mizrahi, I. (2013) Exploring the bovine rumen bacterial community from birth to adulthood. ISME J. 7 (6), 1069–1079, http://dx.doi.org/10.1038/ismej.2013.2.

[18] Kamke, J., Kittelmann, S., Soni, P., Li, Y., Tavendale, M., Ganesh, S., Janssen, P.H., Shi, W., Froula, J., Rubin, E.M., Attwood, G.T. (2016) Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. Microbiome 4, 56, http://dx.doi.org/10.1186/s40168-016-0201-2.

[19] Koropatkin, N., Martens, E.C., Gordon, J.I., Smith, T.J. (2009) Structure of a SusD homologue, BT1043, involved in mucin *O*-glycan utilization in a prominent human gut symbiont. Biochemistry 48 (7), 1532–1542, http://dx.doi.org/10.1021/bi801942a.

[20] Koropatkin, N.M., Martens, E.C., Gordon, J.I., Smith, T.J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16 (7), 1105–1115, http://dx.doi.org/10.1016/j.str.2008.03.017.

[21] Langmead, B., Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357–359, http://dx.doi.org/10.1038/nmeth.1923.

[22] Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinter, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C., Brumer, H. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. Nature 506 (7489), 498–502, http://dx.doi.org/10.1038/nature12907.

[23] Lu, J., Breitwieser, F.P., Thielen, P., Salzberg, S.L. (2017) Bracken: estimating species abundance in metagenomics data. PeerJ Comput. Sci. 3, e104, http://dx.doi.org/10.7717/peerj-cs.104.

[24] Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Lu, S., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Bryant, S.H. (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 41 (D1), D348–D352, http://dx.doi.org/10.1093/nar/gks1243.

[25] Martens, E.C., Chiang, H.C., Gordon, J.I. (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4 (5), 447–457, http://dx.doi.org/10.1016/j.chom.2008.09.007.

[26] Martens, E.C., Kelly, A.G., Tauzin, A.S., Brumer, H. (2014) The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. J. Mol. Biol. 426 (23), 3851–3865, http://dx.doi.org/10.1016/j.jmb.2014.06.022.

[27] Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., Gordon, J.I. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. 9 (12), e1001221, http://dx.doi.org/10.1371/journal.pbio.1001221.

[28] Mayorga, O.L., Kingston-Smith, A.H., Kim, E.J., Allison, G.G., Wilkinson, T.J., Hegarty, M.J., Theodorou, M.K., Newbold, C.J., Huws, S.A. (2016) Temporal metagenomic and metabolomic characterization of fresh perennial ryegrass degradation by rumen bacteria. Front. Microbiol. 7, http://dx.doi.org/10.3389/fmicb.2016.01854.

[29] Miyazaki, K., Miyamoto, H., Mercer, D.K., Hirase, T., Martin, J.C., Kojima, Y., Flint, H.J. (2003) Involvement of the multidomain regulatory protein XynR in positive control of xylanase gene expression in the ruminal anaerobe *Prevotella bryantii* B(1)4. J. Bacteriol. 185 (7), 2219–2226.

[30] Nakayama, J., Watanabe, K., Jiang, J., Matsuda, K., Chao, S.-H., Haryono, P., Laongkham, O., Sarwoko, M.-A., Sujaya, I.N., Zhao, L., Chen, K.-T., Chen, Y.-P., Chiu, H.-H., Hidaka, T., Huang, N.-X., Kiyohara, C., Kurakawa, T., Sakamoto, N., Sonomoto, K., Tashiro, K., Tsuji, H., Chen, M.-J., Leelavatcharamas, V., Liao, C.-C., Nitisinprasert, S., Rahayu, E.S., Ren, F.-Z., Tsai, Y.-C., Lee, Y.-K. (2015) Diversity in gut bacterial community of school-age children in Asia. Sci. Rep. 5, 8397, http://dx.doi.org/10.1038/srep08397.

[31] Nikolich, M.P., Hong, G., Shoemaker, N.B., Salyers, A.A. (1994) Evidence for natural horizontal transfer of tetQ between bacteria that normally colonize humans and bacteria that normally colonize livestock. Appl. Environ. Microbiol. 60 (9), 3255–3260.

[32] Nograšek, B., Accetto, T., Fanedl, L., Avguštin, G. (2015) Description of a novel pectin-degrading bacterial species *Prevotella pectinovora* sp. nov., based on its phenotypic and genomic traits. J. Microbiol. Seoul Korea 53 (8), 503–510, http://dx.doi.org/10.1007/s12275-015-5142-0.

[33] Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. 2015 Vegan: Community Ecology Package. R Package Version 2.2-1.

[34] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106 (45), 19126–19131, http://dx.doi.org/10.1073/pnas.0906412106.

[35] Rogers, T.E., Pudlo, N.A., Koropatkin, N.M., Bell, J.S.K., Moya Balasch, M., Jasker, K., Martens, E.C. (2013) Dynamic responses of *Bacteroides* thetaiotaomicron during growth on glycan mixtures. Mol. Microbiol. 88 (5), 876–890, http://dx.doi.org/10.1111/mmi.12228.

[36] Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé, A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J., Bolam, D.N. (2015) Glycan complexity dictates microbial resource allocation in the large intestine. Nat. Commun. 6, 7481, http://dx.doi.org/10.1038/ncomms8481.

[37] Rosewarne, C.P., Pope, P.B., Cheung, J.L., Morrison, M. (2014) Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum *Bacteroidetes*. J. Ind. Microbiol. Biotechnol. 41 (3), 601–606, http://dx.doi.org/10.1007/s10295-013-1395-y.

[38] Rosselló-Móra, R., Amann, R. (2015) Past and future species definitions for *Bacteria* and *Archaea*. Syst. Appl. Microbiol. 38 (4), 209–216, http://dx.doi.org/10.1016/j.syapm.2015.02.001.

[39] Salyers, A.A., Gupta, A., Wang, Y. (2004) Human intestinal bacteria as reservoirs for antibiotic resistance genes. Trends Microbiol. 12 (9), 412–416, http://dx.doi.org/10.1016/j.tim.2004.07.004.

[40] Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069, http://dx.doi.org/10.1093/bioinformatics/btu153.

[41] Seničar, L., Accetto, T. (2015) The 5′ untranslated mRNA region base content can greatly affect translation initiation in the absence of secondary structures in *Prevotella bryantii* TC1-1. FEMS Microbiol. Lett. 362 (1), 1–4, http://dx.doi.org/10.1093/femsle/fnu023.

[42] Shi, W., Moon, C.D., Leahy, S.C., Kang, D., Froula, J., Kittelmann, S., Fan, C., Deutsch, S., Gagic, D., Seedorf, H., Kelly, W.J., Atua, R., Sang, C., Soni, P., Li, D., Pinares-Patiño, C.S., McEwan, J.C., Janssen, P.H., Chen, F., Visel, A., Wang, Z., Attwood, G.T., Rubin, E.M. (2014) Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. Genome Res. 24 (9), 1517–1525, http://dx.doi.org/10.1101/gr.168245.113.

[43] Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., Sonnenburg, J.L. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141 (7), 1241–1252, http://dx.doi.org/10.1016/j.cell.2010.05.005.

[44] Stewart, R.D., Auffret, M.D., Warr, A., Wiser, A.H., Press, M.O., Langford, K.W., Liachko, I., Snelling, T.J., Dewhurst, R.J., Walker, A.W., Roehe, R., Watson, M. (2018) Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat. Commun. 9 (1), 870, http://dx.doi.org/10.1038/s41467-018-03317-6.

[45] Suzuki, R., Shimodaira, H. (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics (Oxford, England) 22 (12), 1540–1542, http://dx.doi.org/10.1093/bioinformatics/btl117.

[46] Wallace, R.J., Rooke, J.A., McKain, N., Duthie, C.-A., Hyslop, J.J., Ross, D.W., Waterhouse, A., Watson, M., Roehe, R. (2015) The rumen microbial metagenome associated with high methane production in cattle. BMC Genomics 16, 839, http://dx.doi.org/10.1186/s12864-015-2032-0.

[47] Weimer, P.J. (2015) Redundancy, resilience, and host specificity of the ruminal microbiota: implications for engineering improved ruminal fermentations. Front. Microbiol. 6, http://dx.doi.org/10.3389/fmicb.2015.00296.

[48] Wood, D.E., Salzberg, S.L. (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46, http://dx.doi.org/10.1186/gb-2014-15-3-r46.

[49] Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., Liang, S., Pedersen, A.Ø., Kjeldsen, N.J., Liu, C., Maguin, E., Doré, J., Pons, N., Chatelier, E.L., Prifti, E., Li, J., Jia, H., Liu, X., Xu, X., Ehrlich, S.D., Madsen, L., Kristiansen, K., Rogel-Gaillard, C., Wang, J. (2016) A reference gene catalogue of the pig gut microbiome. Nat. Microbiol. 1 (12), 16161, http://dx.doi.org/10.1038/nmicrobiol.2016.161.

[50] Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40 (W1), W445–W451, http://dx.doi.org/10.1093/nar/gks479.