A peer-reviewed version of this preprint was published in PeerJ on 11 September 2017.

Gerritsen J, Hornung B, Renckens B, van Hijum SAFT, Martins dos Santos VAP, Rijkers GT, Schaap PJ, de Vos WM, Smidt H. 2017. Genomic and functional analysis of *Romboutsia ilealis* CRIB^{T} reveals adaptation to the small intestine. PeerJ 5:e3698 https://doi.org/10.7717/peerj.3698

Genomic and functional analysis of *Romboutsia ilealis* CRIB^T reveals adaptation to the small intestine

Jacoline Gerritsen [1, 2], **Bastian Hornung** [Corresp., 1, 3], **Bernadette Renckens** [4], **Sacha A.F.T. van Hijum** [4, 5], **Vitor A.P. Martins dos Santos** [3, 6], **Ger T. Rijkers** [7, 8], **Peter J. Schaap** [3], **Willem M. de Vos** [1, 9], **Hauke Smidt** [1]

[1] Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands
[2] Winclove Probiotics, Amsterdam, The Netherlands
[3] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands
[4] Nijmegen Centre for Molecular Life Sciences, CMBI, Radboud UMC, Nijmegen, The Netherlands
[5] NIZO, Ede, The Netherlands
[6] LifeGlimmer GmbH, Berlin, Germany
[7] Laboratory for Medical Microbiology and Immunology, St. Antonius Hospital, Nieuwegein, The Netherlands
[8] Department of Science, University College Roosevelt, Middelburg, The Netherlands
[9] Departments of Microbiology and Immunology and Veterinary Biosciences, University of Helsinki, Helsinki, Finland

Corresponding Author: Bastian Hornung
Email address: bastian.hornung@gmx.de

Background. The microbiota in the small intestine relies on their capacity to rapidly import and ferment available carbohydrates to survive in a complex and highly competitive ecosystem. Understanding how these communities function requires elucidating the role of its key players, the interactions among them and with their environment/host.

Methods. The genome of the gut bacterium *Romboutsia ilealis* CRIB^T was sequenced with multiple technologies (Illumina paired end, mate pair and PacBio). The transcriptome was sequenced (Illumina HiSeq) while growing on three different carbohydrate sources and short chain fatty acids were measured via HPLC.

Results. Hence, we present the complete genome of *Romboutsia ilealis* CRIB^T, a natural inhabitant and key player of the small intestine of rats. *R. ilealis* CRIB^T possesses a circular chromosome of 2,581,778 bp and a plasmid of 6,145 bp, carrying 2,351 and eight predicted protein coding sequences, respectively. Analysis of the genome revealed limited capacity to synthesize amino acids and vitamins, whereas multiple and partially redundant pathways for the utilization of different relatively simple carbohydrates are present. Transcriptome analysis allowed pinpointing the key components in the degradation of glucose, L-fucose and fructo-oligosaccharides.

Discussion. This revealed that *R. ilealis* CRIB^T is adapted to a nutrient-rich environment where carbohydrates, amino acids and vitamins are abundantly available and uncovered potential mechanisms for competition with mucus-degrading microbes.

Genomic and functional analysis of *Romboutsia ilealis* CRIB^T reveals adaptation to the small intestine

Jacoline Gerritsen[1,2*], Bastian Hornung[1,3*], Bernadet Renckens[4], Sacha A. F. T. van Hijum[4,5], Vitor A.P. Martins dos Santos[3,6], Ger T. Rijkers[7,8], Peter J. Schaap[3], Willem M. de Vos[1,9] and Hauke Smidt[1]

[1] Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, the Netherlands;
[2] Winclove Probiotics, Hulstweg 11, 1032 LB, Amsterdam, the Netherlands;
[3] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, the Netherlands;
[4] Nijmegen Centre for Molecular Life Sciences, CMBI, Radboud UMC, P.O. Box 9101, 6500 HB, Nijmegen, the Netherlands;
[5] NIZO, Kernhemseweg 2, 6718 ZB, Ede, the Netherlands;
[6] LifeGlimmer GmbH, Markelstrasse 38, Berlin, Germany;
[7] Laboratory for Medical Microbiology and Immunology, St. Antonius Hospital, P.O. Box 2500, 3430 EM, Nieuwegein, the Netherlands;
[8] Department of Science, University College Roosevelt, P.O. Box 94, 4330 AB, Middelburg, the Netherlands;
[9] Departments of Microbiology and Immunology and Veterinary Biosciences, University of Helsinki, P.O. Box 66, 00014, University of Helsinki, Finland

[*] These authors contributed equally to this work
#Corresponding author:
Bastian Hornung[1,3]

Email address: bastian.hornung@gmx.de

Abstract

Background. The microbiota in the small intestine relies on their capacity to rapidly import and ferment available carbohydrates to survive in a complex and highly competitive ecosystem. Understanding how these communities function requires elucidating the role of its key players, the interactions among them and with their environment/host.

Methods. The genome of the gut bacterium *Romboutsia ilealis* CRIB^T was sequenced with multiple technologies (Illumina paired end, mate pair and PacBio). The transcriptome was sequenced (Illumina HiSeq) while growing on three different carbohydrate sources and short chain fatty acids were measured via HPLC.

Results. Hence, we present the complete genome of *Romboutsia ilealis* CRIB^T, a natural inhabitant and key player of the small intestine of rats. *R. ilealis* CRIB^T possesses a circular chromosome of 2,581,778 bp and a plasmid of 6,145 bp, carrying 2,351 and eight predicted protein coding sequences, respectively. Analysis of the genome revealed limited capacity to synthesize amino acids and vitamins, whereas multiple and partially redundant pathways for the utilization of different relatively simple carbohydrates are present. Transcriptome analysis allowed pinpointing the key components in the degradation of glucose, L-fucose and fructo-oligosaccharides.

Discussion. This revealed that *R. ilealis* CRIB^T is adapted to a nutrient-rich environment where carbohydrates, amino acids and vitamins are abundantly available and uncovered potential mechanisms for competition with mucus-degrading microbes.

Introduction

Intestinal microbes live in a complex and dynamic ecosystem, and to survive in this highly competitive environment they have developed close (symbiotic) associations with a diverse array of other intestinal microbes and with their host. This has led to a complex network of host-microbe and microbe-microbe interactions in which the intestinal microbes and the host co-metabolise many substrates (Backhed et al. 2005; Scott et al. 2013). In addition to competition for readily available carbohydrates in the diet, intestinal microbes are able to extract energy from dietary polysaccharides that are indigestible by the host (Flint et al. 2012). Furthermore, intestinal microbes can utilize host-derived secretions (e.g. mucus) as substrates for metabolic processes (Ouwerkerk et al. 2013). In turn, the metabolic activities of the intestinal microbes result in the production of a wide array of compounds of which some are important nutrients for the host. For example, short chain fatty acids (SCFA), the main end-products of bacterial fermentation in the gut, can be readily absorbed by the host and further metabolized as energy sources (Elia & Cummings 2007) (Lange et al. 2015). All together, the metabolic activity of the intestinal microbiota has a major impact on health of the host, and recent studies have indicated an important role for microbial activity in (human) diseases such as inflammatory bowel disease, irritable bowel syndrome and obesity (Gerritsen et al. 2011a; Quigley 2013).

The wide array of microbial genes present in the intestinal tract in addition to the host's own genome provides insight into the complex network of possible host-microbe and microbe-microbe interactions. To this end, it has been estimated that together the human intestinal microbes contribute ~150 times more genes than the human genome (Qin et al. 2010). To be able to maintain themselves in an ecosystem such as the intestinal tract, microbes have adapted or

even specialized in foraging certain niche-specific substrates. However, we have only limited understanding of the spatial and temporal heterogeneity in community composition and activity in the different niches along the length of the intestinal tract. To unravel the functional contribution of specific intestinal microbes to host physiology and pathology, we have to understand their metabolic capabilities at a higher resolution. It is still difficult, however, to associate a functionality in an ecosystem to specific sets of genes and in turn to individual microbial species, and vice versa. To this end, the combination of genome mining and functional analyses with single microbes or simple and defined communities can provide an overall insight in the genetic and functional potential of specific members of the intestinal microbial community (Heinken et al. 2013; Li et al. 2008; Xu et al. 2003).

Little is known about the intestinal microbes adapted to the small intestine (Booijink et al. 2007; van den Bogert et al. 2013b; Zhang et al. 2014). The small intestine is a nutrient-rich environment, and previous studies have shown that the microbial communities in the (human) small intestine are driven by the rapid uptake and conversion of simple carbohydrates (Zoetendal et al. 2012) (Leimena et al. 2013). Community composition and activity in the small intestine is largely determined by host digestive fluids such as gastric acid, bile and pancreatic secretions. In-depth genomic analysis on *Streptococcus* isolates of small intestinal origin has shown that these microbes are adapted to a highly dynamic environment (Van den Bogert et al. 2013a). Here we present the completely sequenced and annotated genome of *Romboutsia ilealis* CRIB^T, a Gram-positive bacterium that was recently isolated from the small intestinal tract of rats (Gerritsen et al. 2014). It was found to be a natural inhabitant of the rat small intestine, specifically of the ileum, and a correlation with improved health status of the rats was observed in an experimental model of acute pancreatitis (Gerritsen et al. 2011b).

Here we describe a model driven analysis of the small intestinal inhabitant *R. ilealis* CRIB^T and provide an overview of the metabolic capabilities and nutritional potential of this organism. These examinations revealed that *R. ilealis* CRIB^T is able to degrade a wide range of relatively simple carbohydrates. In addition, we were able to pinpoint potential mechanisms that enable this organism to survive in the competitive small intestinal environment. These mechanisms include a bile salt hydrolase and an urease, which enhance the organisms ability to handle intestinal conditions, and potential mechanisms for competition for fucose with mucus-degrading microbes.

Materials and methods

Genome sequencing, assembly and annotation

R. ilealis CRIB^T (DSM 25109) was routinely cultured in CRIB medium at 37 °C as previously described (Gerritsen et al. 2014). Genomic DNA extraction was performed as previously described (Van den Bogert et al. 2013a). Genome sequencing was done using 454 Titanium pyrosequencing technology (Roche 454 GS FLX), as well as Illumina (Genome Analyzer II and HiSeq2000) and PacBio sequencing (PacBio RS). Mate-pair data was generated by BaseClear (Leiden, the Netherlands). All other data was generated by GATC Biotech (Konstanz, Germany). The genome was assembled in a hybrid approach with multiple assemblers and scaffolds.

Genome annotation was carried out with an in-house pipeline. Prodigal v2.5 was used for prediction of protein coding DNA sequences (CDS) (Hyatt et al. 2010), InterProScan 5RC7 for protein annotation (Hunter et al. 2012), tRNAscan-SE v1.3.1 for prediction of tRNAs (Lowe &

Eddy 1997) and RNAmmer v1.2 for the prediction of rRNAs (Lagesen et al. 2007). Additional protein function predictions were derived via BLAST identifications against the UniRef50 (Suzek et al. 2007) and Swissprot (UniProt-Consortium 2014) databases (download August 2013). Afterwards the annotation was further enhanced by adding EC numbers via PRIAM version March 06, 2013 (Claudel-Renard et al. 2003). Non-coding RNAs were identified using rfam_scan.pl v1.04, on release 11.0 of the RFAM database (Burge et al. 2013). CRISPRs were annotated using CRISPR Recognition Tool v1.1 (Bland et al. 2007). See the Supplemental Methods in Text S1 for details on the genomic DNA extraction, genome sequencing, assembly, annotation, and metabolic modelling.

Whole-genome transcriptome analysis

R. ilealis CRIB^T was grown in a basal bicarbonate-buffered medium (Stams et al. 1993) supplemented with 16 g/L yeast extract (BD, Breda, The Netherlands) and an amino acids solution as used for the growth of *C. difficile (Karasawa et al. 1995)*. In addition the medium was supplemented with either 0.5 % (w/v) D-glucose (Fisher Scientific Inc., Waltham, MA USA), L-fucose (Sigma-Aldrich, St. Louis, MO, USA) or fructo-oligosaccharide (FOS) P06 (DP 2-4; Winclove Probiotics, Amsterdam, The Netherlands). The final pH of the medium was 7.0. For each condition triplicate cultures were set up.

For RNA-seq analysis, the cells were harvested in mid-exponential phase (OD_{600nm} = 0.25-0.55, ~8-10h incubation) (Table S1).

Total RNA was purified using the RNeasy Mini Kit (QIAGEN GmbH, Hilden, Germany). Depletion of rRNA was performed using the Rib-Zero^{TM} Kit for bacteria (Epicentre Biotechnologies, Madison, WI, USA). Library construction for whole-transcriptome sequencing

(RNA-seq) was done by use of the ScriptSeq™ v2 RNA-seq Library Preparation Kit in combination with ScriptSeq™ Index PCR primers (Epicentre Biotechnologies).

The barcoded cDNA libraries were pooled and sent to GATC Biotech (Konstanz, Germany) where 150 bp sequencing was performed on one single lane using the Illumina HiSeq2500 platform in combination with the TruSeq Rapid SBS (200 cycles) and TruSeq Rapid SR Cluster Kits (Illumina Inc., San Diego, CA, USA). Reads were mapped to the genome with Bowtie2 v2.0.6 (Langmead & Salzberg 2012) using default settings. Details on the RNA-seq raw data analysis can be found in Table S2 and Supplemental Methods in Text S1.

Gene expression abundance estimates and differential expression analysis was performed using Cuffdiff v2.1.1 (Trapnell et al. 2013) with default settings. Differentially expressed genes were determined by pairwise comparison of a given condition to the other three conditions for a total of six pairwise comparisons. Genes were considered significantly differentially expressed when they showed a ≥1.5 log2(fold change) in any of the conditions with a false discovery rate (FDR)-corrected P value (q value) ≤ 0.05 (Tables S3-S6). Principal component analysis was performed with Canoco 5.0 (ter Braak & Smilauer 2012) on log-transformed gene transcript abundances using Hellinger standardization. Gene expression heatmaps were generated based on gene transcript abundances using R v3.1.0 and R-packages svDialogs and gplots.

See the Supplemental Methods in Text S1 for details on the carbohydrate growth experiment and whole genome transcriptome analysis.

Nucleotide sequence accession number

All related data has been deposited at the European Nucleotide Archive. The raw reads for the genome of *R. ilealis* CRIB^T can be accessed via the accession numbers ERR366773,

ERX397233, ERX397242 and ERX339449. The assembly can be accessed under LN555523-LN555524. The RNAseq data has been deposited under the numbers ERS533849- ERS533861.

Results

Genome analysis

Global genome features

R. ilealis CRIB^T contains a single, circular chromosome of 2,581,778 bp and a plasmid of 6,145 bp (Table 1 and Fig. 1). The chromosome contains 2,351 predicted protein CDS, of which 321 were annotated as hypothetical and for 91 only a domain of unknown function could be assigned. The plasmid carries eight predicted protein CDS, of which none was recognized for having a metabolic or replicative function. Furthermore, it appears to be a non-mobilizable plasmid, given that it lacks any known mobilization-associated genes. The overall G+C content of the genome is 27.9 %, which is in good agreement with a G+C content of 28.1 mol% previously determined for *R. ilealis* CRIB^T by HPLC methods (Gerritsen et al. 2014).

INSERT TABLE 1

INSERT FIGURE 1

With a total of 14 copies of the 16S ribosomal RNA (rRNA) gene, *R. ilealis* CRIB^T is among the species with the highest number of 16S rRNA gene copies reported up to this date (Lee et al. 2009). However, it should be noted, that not all 16S rRNA copies in *R. ilealis* CRIB^T are embedded in the conserved 16S-23S-5S rRNA operon structure. This has been reported for other

genomes containing multiple rRNA operons as the result of duplications (Bensaadi-Merchermek et al. 1995; Schwartz et al. 1992).

A COG category could only be assigned to 1,647 of the predicted proteins (70 %) including 372 proteins (16 %) assigned to the categories R (general function prediction only) and S (function unknown) (Fig. S1). This low coverage in function prediction is most likely due to the fact that the COG categorization has not been updated since 2003. With InterProScan a predicted function could be assigned to 82 % of the predicted proteins. Based on the InterPro and PRIAM classifications (Claudel-Renard et al. 2003), an enzymatic function could be predicted for more than 500 proteins.

General metabolic pathways

Analysis of the CDS predicted from the *R. ilealis* CRIB^T genome revealed the presence of a complete set of enzymes for the glycolytic pathway. In line with the anaerobic lifestyle of the organism, enzymes for the oxidative phase of the pentose phosphate pathway could not be detected. Additionally, the genes that encode enzymes involved in the tricarboxylic acid cycle were lacking. Subsequently a metabolic model was constructed with Pathway tools v18.0. A flux balance analysis with the model was performed, suggesting that *R. ilealis* CRIB^T is a mixed acid fermenter as was previously reported (Gerritsen et al. 2014). Predicted end products of fermentation are a mixture of acetate, formate, lactate and ethanol, with the possibility of gas formation (CO_2 and H_2). In addition to ethanol, which can be produced during mixed acid fermentation, 1,2-propanediol was predicted to be formed via the L-fucose degradation pathway. The fermentation end products formate, acetate and lactate are predicted to be produced from pyruvate. No other solvents were predicted to be produced by the metabolic model, however,

low amounts of propionate production (up to 3 mM) were observed repeatedly during *in vitro* growth (Gerritsen et al. 2014). None of the three established pathways for propionate production in the intestinal tract, i.e. the succinate, acrylate or the propanediol pathway (Reichardt et al. 2014), could be identified at the genetic level, but it might be possible that propionate is produced together with formate from threonine fermentation by the activity of acetate kinase in the final step instead of a dedicated propionate kinase, as these enzymes have very similar catabolic mechanisms and carboxylic acid binding sites (Ingram-Smith et al. 2005). No other probable underground reactions (known minor side reactions besides the main reaction catalysed by enzymes) (D'Ari & Casadesus 1998) (Notebaart et al. 2014) for the production of propionate were detected while investigating the BRENDA database (Chang et al. 2015).

The analysis of the genome and the prediction by the model indicated that fermentation is probably the main process for energy conservation in *R. ilealis*. However, the presence of a sulfite reductase gene cluster (CRIB_1284-CRIB_1286) of the dissimilatory *asrC*-type (Dhillon et al. 2005) points at possible anaerobic respiration. Similar siroheme-dependent sulfite reductases are found in many close-relatives of *R. ilealis* such as *Intestinibacter bartlettii*, *Clostridium sordellii* and *C. difficile* (Czyzewski & Wang 2012). Sulfite reduction by *R. ilealis* CRIB^T, and close relatives, has been previously demonstrated *in vitro* (Gerritsen et al. 2014), and increased growth yield and metabolite production was observed in the presence of sulfite for *R. ilealis* CRIB^T (data not shown). In the intestinal tract sulfite is derived from food sources that contain sulfite as preservative, and it has been shown that neutrophils release sulfite as part of the host defence against microbes (Mitsuhashi et al. 1998).

234 ***Metabolism of growth factors and cofactors***

235 Many genes encoding enzymes required for amino acid biosynthesis appeared to be absent in *R.*
236 *ilealis* CRIB^T. Complete pathways are present for the biosynthesis of aspartate, asparagine,
237 glutamate, glutamine and cysteine, using carbon skeletons available from central metabolites or
238 via conversion of other amino acids.

239 The absence of genes to produce branched-chain amino acids (leucine, isoleucine and valine)
240 was also reflected in the absence of branched chain fatty acids in the cell membrane of *R. ilealis*,
241 which is characteristic for the genus *Romboutsia* (Gerritsen et al. 2014). From these observations
242 it can be concluded that *R. ilealis* depends on a number of exogenous amino acids, peptides
243 and/or proteins to fuel protein synthesis. The dependency on an exogenous source of amino acids
244 is reflected by the identification of multiple amino acid transporters, including an
245 arginine/ornithine antiporter, multiple serine/threonine exchangers, a transporter for branched
246 amino acids, and several amino acid symporters and permeases without a predicted specificity.
247 Furthermore, numerous genes have been annotated as protease or peptidase, including several
248 with a signal peptide.

249 *R. ilealis* CRIB^T appears to contain all genes for *de novo* purine and pyrimidine synthesis, as
250 well as for the production of the coenzymes NAD and FAD via salvage pathways from niacin
251 and riboflavin, respectively. While some organic cofactors can be produced by *R. ilealis* CRIB,
252 including siroheme, it mainly relies on salvage pathways (e.g. for lipoic acid) or exogenous
253 sources for the supply of precursors mainly in the form of vitamins (e.g. thiamin, riboflavin,
254 niacin, pantothenate, pyridoxine, biotin, vitamin B12).

255

256 ***Carbohydrate transport and metabolism***

257 As previously reported *R. ilealis* CRIB^T is able to utilize a wide variety of carbohydrates
258 (Gerritsen et al. 2014). Previously good growth of *R. ilealis* on L-fucose, glucose, raffinose and
259 sucrose was described in addition to moderate growth on D-arabinose and D-galactose and weak
260 growth on D-fructose, inulin, lactose, maltose and melibiose. Growth on L-fucose, fructose,
261 galactose, glucose, lactose, maltose, melibiose, raffinose and sucrose was predicted from the
262 genome-scale metabolic model as well. Besides the central glycolytic genes, which are present in
263 a single genomic area, genes encoding the specific carbohydrate degradation enzymes are
264 distributed throughout the genome in gene clusters together with their respective transporters and
265 transcriptional regulator. No separate pathway for the use of D-arabinose could be predicted,
266 however, the L-fucose degradation pathway is likely also used for D-arabinose utilization
267 (LeBlanc & Mortlock 1971). An arabinose transporter, similar to the maltose and sucrose
268 transporters, could be identified. In addition, a gene cluster involved in the degradation of the
269 host-derived carbohydrate sialic acid could be predicted (CRIB_613-CRIB_619) (Almagro-
270 Moreno & Boyd 2009). The structure of this gene cluster is similar to the one identified in *C.*
271 *difficile* (Ng et al. 2013). The ability to degrade the predominantly host-derived carbohydrates L-
272 fucose and sialic acid suggest a role in the utilization of mucin, an abundant host-derived
273 glycoprotein in the intestinal tract (Derrien et al. 2010; Ouwerkerk et al. 2013). However, no
274 growth on mucin was observed (data not shown), which is in line with the lack of a predicted
275 extracellular fucosidase and/or sialidase.

276

277 ***Additional genes of ecological interest***

278 A gene cluster encoding a urease, consisting of three subunits (*ureABC*), and a number of urease
279 accessory genes were identified (CRIB_1381-CRIB_1388). The gene cluster identified in *R.*
280 *ilealis* CRIB is very similar to the urease gene cluster in the genome of *C. sordellii* (Fig. S3), a
281 species in which the urease activity is used to phenotypically distinguish *C. sordellii* strains from
282 *C. bifermentans* strains (Roggentin et al. 1985). Furthermore, a possible ammonium transporter
283 (CRIB_1389) was identified in the genome of *R. ilealis* CRIB^T next to the urease gene cluster.
284 Ureases are nickel-containing metalloenzymes that catalyse the hydrolysis of urea to ammonia
285 and carbon dioxide, and thereby these enzymes allow microbes to use urea as nitrogen source by
286 assimilation via glutamate. They are ubiquitous proteins occurring in diverse organisms (Mobley
287 et al. 1995). In the intestinal environment, where urea is abundantly present (Fuller & Reeds
288 1998), some bacteria use ureases to survive the acidic conditions in the upper part of the
289 intestinal tract as urea hydrolysis leads to a local increase in pH (Rutherford 2014).
290 Another function of ecological interest is the predicted choloylglycine hydrolase. Proteins within
291 the choloylglycine hydrolase family are bile salt hydrolases (BSHs), also known as conjugated
292 bile acid hydrolases (CBAHs), that are widespread among intestinal microbes (Ridlon et al.
293 2006). They are involved in the hydrolysis of the amide linkage in conjugated bile salts,
294 releasing primary bile acids. There is a large heterogeneity among BSHs, for example with
295 respect to their substrate specificity. The BSH of *R. ilealis* CIRB^T was found to be the most
296 similar to the one found in *C. butyricum*. Although the physiological advantages of BSHs for the
297 microbes are not completely understood, it has been hypothesized that they constitute a
298 mechanism to detoxify bile salts and thereby enhance bacterial colonization (Czyzewski & Wang
299 2012).

300

301 **Metabolite and transcriptome analysis**

302 ***Metabolite and transcriptome analysis of R. ilealis CRIB^T during growth on different***
303 ***carbohydrates***

304 To study key pathways predicted to be involved in carbohydrate utilization and their regulation
305 in more detail, a whole-genome transcriptome analysis was performed, focussing on four
306 experimental conditions. Firstly, growth on glucose, a preferred substrate for many microbes
307 present in the intestinal tract, was studied. Secondly the growth on fructans, oligo-
308 /polysaccharides present in many food items, was examined. Previously weak growth on inulin, a
309 polysaccharide consisting of long chains of ß1→2 linked fructose units, was observed (Gerritsen
310 et al. 2014). For this study a shorter fructan (FOS P06, DP2-4) was chosen, because growth on
311 shorter fructans is likely more relevant for microbes living in the small intestine (Zoetendal et al.
312 2012). Thirdly, growth on L-fucose was examined, as growth on this substrate was found to be
313 unique for *R. ilealis* CRIB^T compared to other related microbes. Finally, *R. ilealis* CRIB^T was
314 also grown in the basal medium in the absence of an additional carbon source for comparison
315 (control condition).

316 Based on measurements of optical density and pH during growth (growth characteristics of
317 individual cultures can be found in Table S1), sampling was done in mid-exponential phase (~8-
318 10 h incubation; used for transcriptome analysis) and in stationary phase (24 h incubation), and
319 sugar utilization and fermentation products were measured with HPLC (Table 2). The fact that in
320 neither of the experimental conditions the supplied carbohydrates were depleted and metabolites
321 were still produced during the sampling points at ~8-10 h and 24 h, confirmed that samples
322 obtained for transcriptome analysis at ~8-10 h were taken during exponential growth. In the FOS

cultures an accumulation of extra-cellular fructose was observed. As predicted from the metabolic model, growth on glucose resulted in the production of formate, acetate and lactate (Table 2).

INSERT TABLE 2

Growth on FOS was slightly less than on glucose, however, after 24 h of growth the same fermentation products were observed in similar amounts (Table 2). Growth on L-fucose showed production of 1,2-propanediol instead of lactate. The fact that 1,2-propanediol was observed in one of the control cultures could be explained by the fact that an L-fucose grown culture was used as inoculum for this culture, leading to carry-over of minor amounts of metabolites.

For the genome-wide transcriptome analysis of triplicate cultures grown in the four different conditions (i.e. a total of 12 cultures), a total of 159,250,634 150bp-reads were generated by RNA-seq (overview in Table S2). Principal component analysis of the transcriptomes of the individual cultures showed that the cultures clustered by condition (Fig. 2).

INSERT FIGURE 2

Differential expression of genes involved in carbohydrate degradation and fermentation in R. ilealis CRIB^T

To identify differentially regulated genes, pairwise comparisons were done with cuffdiff (Trapnell et al. 2013) using a cut off of ≥1.5 log2 (fold-change) and q-value ≤0.05. Figure 3 shows a heat map of all differentially regulated genes, exact numbers can be found in Tables S3-6.

INSERT FIGURE 3

INSERT FIGURE 4

The gene cluster involved in glycolysis (CRIB_186-CRIB_191) was most abundantly expressed in the conditions that support the highest growth rates determined by the highest cell density reached in the time period that was measured (glucose, followed by FOS; Fig. 3). This was also reflected in the fact that expression of genes involved in replication such as ribosomal proteins, proteins involved in cell wall biosynthesis and general cell division processes were most strongly expressed during growth in the presence of glucose and to a lesser extent FOS. Other genes involved in the central sugar metabolic pathways (e.g. CRIB_1849, CRIB_140, CRIB_2223, and CRIB_105) were upregulated in these conditions, albeit not significantly differentially regulated. This suggests that these are less tightly regulated at the transcriptional level, probably because they are also involved in other processes than sugar degradation (Commichau et al. 2009). The metabolic model suggests that this is indeed the case, as some of the enzymes produce intermediates which can be consumed by fatty acid biosynthesis and amino acid biosynthesis processes.

Altogether, the transcriptome of *R. ilealis* CRIB^T grown on FOS was very similar to its transcriptome when grown on glucose (Fig. 2), with only 18 genes significantly upregulated during growth in the presence of FOS compared to glucose (Table S4). Apparent was the upregulation of the gene clusters involved in the transport and degradation of the respective sugars or their derivatives (Fig. 3). In the presence of glucose the glucose-specific PTS system

(CIRB_2017-CRIB_2018) was significantly upregulated, together with its associated transcriptional regulator (CRIB_2019). In turn, in the presence of FOS two clusters predicted to be involved in sucrose degradation (CRIB_148-CRIB_152 and CRIB_1458-1461) were significantly upregulated. The third gene cluster predicted to be involved in sucrose degradation (CRIB_1399-1400) was not significantly regulated during growth on FOS, however, it should be noted that these genes are located in a cluster functionally annotated to melibiose metabolism and are most likely regulated by the transcriptional regulator in this cluster. In addition to the two sucrose degradation clusters, a transport cluster of unknown function (CRIB_1506-CRIB_1509) was upregulated during growth on FOS, albeit only significant compared to growth on glucose. During growth in the presence of L-fucose, the gene cluster predicted to be involved in L-fucose degradation (CRIB_1294-CRIB_1298) was significantly upregulated, including the gene encoding the corresponding transcriptional regulator (CRIB_1299). An overview of the main carbohydrate degradation pathways regulated in the different conditions is given in Figure 4.

During growth on glucose, L-lactate dehydrogenase (CRIB_684) was significantly upregulated, albeit not significant compared to growth on FOS. This enzyme catalyses the reduction of pyruvate resulting in the production of L-lactate and the reoxidation of the NADH formed during glycolysis. Only at time point 24 h lactate was observed (Table 2). This suggests that at time point ~8-10 h the cells were starting to regenerate NAD by upregulating this gene. In the presence of L-fucose, NAD^+ regeneration is achieved via the reduction of lactaldehyde to 1,2-propanediol by lactaldehyde reductase (CRIB_1300), which was upregulated in the presence of L-fucose together with the L-fucose degradation gene cluster. In the spent medium of L-fucose grown cells, 1,2-propanediol was already seen at time point ~8-10 h whereas no lactate production was observed. Another way to regenerate NAD^+ is to reduce pyruvate to ethanol (Fig.

392 4). In the presence of both glucose and FOS an upregulation was seen of the gene encoding the
393 bifunctional aldehyde/alcohol dehydrogenase (CRIB_2231), which converts acetyl-CoA to
394 ethanol. However, in none of the samples ethanol was measured by HPLC analysis.
395
396 ***Expression and regulation of other environmentally relevant processes in R. ilealis CRIB^T***
397 Remarkable was the significant upregulation of a gene cluster related to iron transport
398 (CRIB_892-CRIB-898) during growth on glucose and FOS compared to growth on L-fucose.
399 The significance of this gene cluster for carbohydrate utilization is not known, however, several
400 enzymes could be identified in the genome of *R. ilealis* CRIB^T that use different forms of iron as
401 cofactor, for example the hydrogenases involved in hydrogen metabolism (Calusinska et al.
402 2010), several ferredoxins, and the L-threonine dehydratase (CRIB-426) that was significantly
403 upregulated during growth on L-fucose. As multiple transporters involved in the transport of iron
404 compounds were predicted, it might also be that the uptake of iron provides a competitive
405 advantage to other microbes that are dependent on iron for respiration and other metabolic
406 processes (Kortman et al. 2014).
407 During growth on FOS, a small gene cluster (CRIB_601-CRIB_603) that includes a gene
408 encoding an alternative sigma factor was significantly upregulated. Interestingly, this was also
409 apparent in the control culture that was inoculated with FOS-preconditioned cells. This suggests
410 that in the presence of FOS (or its derivatives sucrose or fructose) transcription is also regulated
411 by RNA polymerase promoter recognition.
412 In the intestinal environment *R. ilealis* CRIB will encounter a wide array of carbohydrates that
413 are either continually or transiently present. Prioritization of carbohydrate utilization is partly
414 achieved at a transcriptional level by the selective expression of genes. The primary mechanism

by which bacteria regulate the utilization of non-preferred carbohydrates in the presence of preferred carbon sources is known as carbon catabolite repression (CCR), a hierarchical system for coordinating sugar metabolism (Deutscher 2008). The fact that, compared to glucose and FOS, L-fucose is utilized by a pathway that does not directly involve fructose-1,6-bisphosphate, a key metabolite in the regulation of CCR of Gram-positive bacteria, made it possible to study CCR by either glucose or FOS. The transcriptome analysis suggests that some genes and operons in *R. ilealis* CRIB^T were indeed subject to CCR in response to the presence of glucose. For example, two gene clusters predicted to be involved in hexuronate metabolism (CRIB_649-CRIB_652 and CRIB_2244-CRIB_2249), pathways that make the use of D-glucuronate and D-galacturonates as sole carbon source possible, were significantly upregulated during growth in the presence of L-fucose compared to growth on glucose (Table S5). In addition, the gene cluster predicted to be involved in sialic acid utilization (CRIB_613-CRIB_616) was downregulated in the presence of glucose as well. Furthermore, when comparing the expression of the gene cluster involved in L-fucose degradation during growth on glucose relative to the growth in the absence of a carbon source (control condition), also this gene cluster appeared to be under CCR in the presence of glucose (Table S5). These results suggest that in *R. ilealis* CRIB^T multiple gene clusters that are involved in the use of alternative carbon sources are subject to CCR. The urease encoding gene cluster (CRIB_1381-CRIB_1388) was significantly upregulated when grown in the absence of an additional carbon source. The fact that this was significant compared to growth on glucose suggests possible CCR of the urease gene cluster, however, upregulation of this gene cluster in the absence of an exogenous carbon source might also be a possible mechanism.

436

437 **Discussion**

438 Gerritsen *et al*. (2011b) have shown by 16S rRNA gene sequence-based analysis that *R. ilealis*
439 CRIB^T is a dominant member of the small intestine microbiota in rats, especially in the ileum.
440 The genomic and transcriptional analysis of *R. ilealis* CRIB^T reported here provides new insights
441 into the genetic and functional potential of this inhabitant of the small intestine. Genomic
442 analysis revealed the presence of metabolic pathways for the utilization of a wide array of
443 'simple' carbohydrates in addition to a multitude of carbohydrate uptake systems that included a
444 series of PTS systems, carbohydrate specific ABC transporters, permeases and symporters.
445 Considering the habitat of *R. ilealis* CRIB^T, we chose to focus on key pathways involved in the
446 utilization of specific diet- and host-derived carbon sources by whole-genome transcriptome
447 analysis.

448 In the intestinal tract, the diet-derived carbohydrates that the host is unable to digest are
449 important sources of energy for many microbes. Here we examined the growth of *R. ilealis*
450 CRIB^T on FOS, a relatively simple oligosaccharide that is indigestible by the host. The
451 transcriptome of *R. ilealis* CRIB^T grown on FOS was very similar to its transcriptome when
452 grown on glucose, a monosaccharide used by the majority of microbes present in the intestinal
453 tract. This is not surprising considering that glucose is in addition to fructose one of the two
454 subunits present in FOS. Differential gene expression analysis demonstrated the apparent FOS-
455 induced upregulation of two separate gene clusters that were predicted to be involved in sucrose
456 degradation. Remarkable was the accumulation of fructose in the culture supernatant during
457 growth of *R. ilealis* CRIB^T on FOS. A simple explanation could be the release of fructose or
458 beta-fructofuranosidase activity after cell lysis. The fact that *R. ilealis* only grows weakly on D-

fructose (Gerritsen et al. 2014), and no transporter specific for fructose could be identified as in close relatives that are able to grow on D-fructose, could explain fructose accumulation. This might indicate a possible symbiotic relationship with other microbes in the close environment, which are most likely able to utilize the excess fructose. Another explanation could be extracellular degradation of FOS, followed by import of sucrose and/or glucose into the cell. Fructan degradation by extracellular enzymes is described for other (intestinal) microbes *(van Hijum et al. 2006)*. However, no extracellular fructansucrase or glucansucrase could be predicted. Furthermore, no new candidates for this activity could be identified via the differential gene expression analysis described here. One possible candidate could be the predicted beta-fructofuranosidase present in the PTS system-containing sucrose degradation gene cluster. Next to the beta-fructofuranosidase-encoding gene, a gene was found to which no function could be assigned, but that was predicted to have a transmembrane region and a domain which could be involved in transport. Given that both loci overlap by a few nucleotides, and that the overlap is within a homopolymer region, it is possible that both loci form one protein due to ribosomal slippage on the homopolymer (Sharma et al. 2014). This could possibly lead to an external membrane-bound enzymatically active protein, which would explain the accumulation of fructose.Future studies with mutant strains might shed more light on the specific contribution of the two predicted sucrose degradation gene clusters to the degradation of FOS, or even longer fructans (inulin), in *R. ilealis* CRIB^T.

Unlike other members of the family *Peptostreptococcaceae*, *R. ilealis* CRIB^T is able to grow on L-fucose, a predominantly host-derived carbon source (Gerritsen et al. 2014). The whole-genome transcriptome analysis confirmed the presence of a functional L-fucose degradation pathway, similar to the pathways previously identified in other intestinal inhabitants such as *E. coli*

(Baldoma & Aguilar 1988), *Bacteroides thetaiotaomicron* (Hooper et al. 1999) and *Roseburia inulinivorans* (Scott et al. 2006). By gene sequence homology a similar pathway was found in *Clostridium perfringens* and the more closely related *C. sordellii* (Fig. S2). L-fucose is a common sugar present within the intestinal environment, as it is a monosaccharide that is an abundant component of many N- and O-linked glycans and glycolipids produced by mammalian cells, including the fucosylated glycans that are found at the terminal positions of mucin glycoproteins (Becker & Lowe 2003). Fucosylated mucin glycoproteins are especially found in the (human) ileum (Robbe et al. 2004; Robbe et al. 2003). For both intestinal commensals and pathogens the ability to utilize L-fucose has been demonstrated to provide a competitive advantage in the intestinal environment (Hooper et al. 1999; Stahl et al. 2011). In *R. ilealis* all enzymes for L-fucose degradation are present in one cluster, however, no fucosidase-encoding gene could be identified, which means that *R. ilealis* is not able to release L-fucose units from fucosylated glycans (e.g. mucin) by itself. Hence, in the intestinal environment *R. ilealis* is dependent on free L-fucose monosaccharides released by other microbes. Furthermore, a gene cluster involved in degradation of sialic acid (Almagro-Moreno & Boyd 2009; Vimr 2013; Vimr et al. 2004) was predicted from the genome, but no extracellular sialidase could be identified similar to *C. difficile* (Ng et al. 2013). This suggests that also for sialic acid, a common residue found in mucin glycoproteins, *R. ilealis* CRIB^T seems to be dependent on the activity of other microbes. Recently it was demonstrated that the host is able to regulate fucosylation of its intestinal epithelial cells in response to pathogen-induced stress (Pickard et al. 2014). This suggests that the ability of microbes to use fucose as an energy source may contribute to the protection of the host against infections by endogenous pathogens.

Interesting was the identification of a urease gene cluster in *R. ilealis* CRIB^T, which appeared to be induced in carbon source limiting circumstances. Urea in the intestinal tract is derived from the breakdown of amino acids. *Helicobacter pylori* is a well-known example where urease activity contributes to the survival of the bacterium in the acidic environment of the stomach (Marshall et al. 1990). For some of the urease-positive bacteria, this enzyme has been shown to act as a virulence factor as it is responsible for urea hydrolysis that leads to increased pH and ammonia toxicity (Rutherford 2014). However, for commensal intestinal bacteria ureases can probably function as colonization factors as well, as they contribute in general to acid resistance and thereby play a role in gastrointestinal survival (Marshall et al. 1990). Urea is released into all parts of the intestinal tract via diffusion from the blood, but it has been reported that pancreatic excretions and bile are a main route of entry (Bergner et al. 1986). We have not been able to demonstrate urease activity in *R. ilealis* CRIB^T during growth in the presence of urea (data not shown). However, different mechanisms for the expression of urease have been identified in other microbes: constitutive, inducible by urea, or controlled by nitrogen source availability (Mobley et al. 1995). For *C. perfringens* for example, the urease activity, which is plasmid borne, was shown to be only expressed in nitrogen-limiting conditions (Dupuy et al. 1997). The increased urease gene expression by *R. ilealis* CRIB^T observed in the control condition, in the absence of an additional carbohydrate, suggests an alternative mechanism for regulation of urease gene expression.

Conclusions

We are just starting to elucidate the composition and function of the microbial communities in the mammalian small intestine. Recently we have reported the isolation and characterization of *R. ilealis* CRIB^T from the small intestine of a rat (Gerritsen et al. 2014). In rats this species was

identified to be a dominant member of the ileal microbiota (Gerritsen et al. 2011b). Here we applied a holistic systems biology approach, involving several fields of wet and dry biology, to study *R. ilealis* CRIB^T, a natural and abundant inhabitant of the small intestinal tract of rats. Its ability to use host sugars that are liberated by other microbes suggests that *R. ilealis* CRIB^T is in competition with mucus-degrading microbes, like *B. thetaiotaomicron* or *Akkermansia muciniphila*. In conclusion, *R. ilealis* is a species that is able to utilize an array of carbohydrates using different and partially redundant pathways. In contrast, it has only limited ability to *de novo* synthesize amino acids and vitamins, and hence the organism shows an adaption to a nutrient-rich environment in which carbohydrates and exogenous sources of amino acids and vitamins are abundantly available. A deeper investigation of key players in the intestinal tract like *R. ilealis* and others will lead to a better understanding of how the microbial communities in us function as a whole. The more we understand how each organism works, and how they interact, the better we get an insight into these environments and can predict how nutrition will influence our health and well-being.

Acknowledgements

The authors thank Hans Heilig (Wageningen University, Laboratory of Microbiology) for his help with genomic DNA isolations and Jasper Koehorst (Wageningen University, Laboratory of Systems and Synthetic Biology) for his help with the genome annotation.

References

Almagro-Moreno S, and Boyd EF. 2009. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol Biol* 9:118.

Backhed F, Ley RE, Sonnenburg JL, Peterson DA, and Gordon JI. 2005. Host-bacterial mutualism in the human intestine. *Science* 307:1915-1920.

Baldoma L, and Aguilar J. 1988. Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic-anaerobic regulation of L-lactaldehyde dissimilation. *Journal of bacteriology* 170:416-421.

Becker DJ, and Lowe JB. 2003. Fucose: biosynthesis and biological function in mammals. *Glycobiology* 13:41R-53R.

Bensaadi-Merchermek N, Salvado JC, Cagnon C, Karama S, and Mouches C. 1995. Characterization of the unlinked 16S rDNA and 23S-5S rRNA operon of *Wolbachia pipientis*, a prokaryotic parasite of insect gonads. *Gene* 165:81-86.

Bergner H, Simon O, Zebrowska T, and Munchmeyer R. 1986. Studies on the secretion of amino acids and of urea into the gastrointestinal tract of pigs. 3. Secretion of urea determined by continuous intravenous infusion of 15N-urea. *Arch Tierernahr* 36:479-490.

Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, and Hugenholtz P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209.

Booijink CC, Zoetendal EG, Kleerebezem M, and de Vos WM. 2007. Microbial communities in the human small intestine: coupling diversity to metagenomics. *Future microbiology* 2:285-295.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, and Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* 41:D226-232.

Calusinska M, Happe T, Joris B, and Wilmotte A. 2010. The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* 156:1575-1588.

Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, Sensen CW, and Schomburg D. 2015. BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res* 43:D439-446.

Claudel-Renard C, Chevalet C, Faraut T, and Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* 31:6633-6639.

Commichau FM, Rothe FM, Herzberg C, Wagner E, Hellwig D, Lehnik-Habrink M, Hammer E, Volker U, and Stulke J. 2009. Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics* 8:1350-1360.

Czyzewski BK, and Wang DN. 2012. Identification and characterization of a bacterial hydrosulphide ion channel. *Nature* 483:494-497.

D'Ari R, and Casadesus J. 1998. Underground metabolism. *BioEssays* 20:181-186.

Derrien M, van Passel MW, van de Bovenkamp JH, Schipper RG, de Vos WM, and Dekker J. 2010. Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* 1:254-268.

Deutscher J. 2008. The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* 11:87-93.

Dhillon A, Goswami S, Riley M, Teske A, and Sogin M. 2005. Domain evolution and functional diversification of sulfite reductases. *Astrobiology* 5:18-29.

Dupuy B, Daube G, Popoff MR, and Cole ST. 1997. *Clostridium perfringens* urease genes are plasmid borne. *Infection and immunity* 65:2313-2320.

Elia M, and Cummings JH. 2007. Physiological aspects of energy metabolism and gastrointestinal effects of carbohydrates. *Eur J Clin Nutr* 61 Suppl 1:S40-74.

Flint HJ, Scott KP, Duncan SH, Louis P, and Forano E. 2012. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3:289-306.

Fuller MF, and Reeds PJ. 1998. Nitrogen cycling in the gut. *Annu Rev Nutr* 18:385-411.

Gerritsen J, Fuentes S, Grievink W, van Niftrik L, Tindall BJ, Timmerman HM, Rijkers GT, and Smidt H. 2014. Characterization of *Romboutsia ilealis* gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus *Clostridium* into the genera *Romboutsia* gen. nov., *Intestinibacter* gen. nov.,

Terrisporobacter gen. nov. and *Asaccharospora* gen. nov. *International journal of systematic and evolutionary microbiology* 64:1600-1616.

Gerritsen J, Smidt H, Rijkers GT, and de Vos WM. 2011a. Intestinal microbiota in human health and disease: the impact of probiotics. *Genes Nutr* 6:209-240.

Gerritsen J, Timmerman HM, Fuentes S, van Minnen LP, Panneman H, Konstantinov SR, Rombouts FM, Gooszen HG, Akkermans LM, Smidt H, and Rijkers GT. 2011b. Correlation between protection against sepsis by probiotic therapy and stimulation of a novel bacterial phylotype. *Applied and environmental microbiology* 77:7749-7756.

Heinken A, Sahoo S, Fleming RM, and Thiele I. 2013. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* 4:28-40.

Hooper LV, Xu J, Falk PG, Midtvedt T, and Gordon JI. 1999. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A* 96:9833-9838.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJ, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, and Yong SY. 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res* 40:D306-312.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, and Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.

Ingram-Smith C, Gorrell A, Lawrence SH, Iyer P, Smith K, and Ferry JG. 2005. Characterization of the acetate binding pocket in the *Methanosarcina thermophila* acetate kinase. *Journal of bacteriology* 187:2386-2394.

Karasawa T, Ikoma S, Yamakawa K, and Nakamura S. 1995. A defined growth medium for clostridium difficile. *Microbiology* 141:371-375.

Kortman GA, Raffatellu M, Swinkels DW, and Tjalsma H. 2014. Nutritional iron turned inside out: intestinal stress from a gut microbial perspective. *FEMS Microbiol Rev*.

Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, and Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100-3108.

Lange K, Hugenholtz F, Schols H, Kleerebezem M, Smidt H, Müller M, and Hooiveld GJEJ. 2015. Comparison of the effects of five dietary fibers on mucosal transcriptional profiles and luminial micribiota composition and SCFA concentrations in murine colon. *Molecular Nutrition & Food Research* in press.

Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-359.

LeBlanc DJ, and Mortlock RP. 1971. Metabolism of D-arabinose: a new pathway in *Escherichia coli*. *Journal of bacteriology* 106:90-96.

Lee ZM, Bussema C, 3rd, and Schmidt TM. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* 37:D489-493.

Leimena MM, Ramiro-Garcia J, Davids M, Van den Bogert B, Smidt H, Smid EJ, Boekhorst J, Zoetendal EG, Schaap PJ, and Kleerebezem M. 2013. A comprehensive metatranscriptome analysis pipeline and its validation using human small intestine microbiota datasets. *BMC Genomics* 14.

Li M, Wang B, Zhang M, Rantalainen M, Wang S, Zhou H, Zhang Y, Shen J, Pang X, Wei H, Chen Y, Lu H, Zuo J, Su M, Qiu Y, Jia W, Xiao C, Smith LM, Yang S, Holmes E, Tang H, Zhao G, Nicholson JK, Li L, and Zhao L. 2008. Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci U S A* 105:2117-2122.

Lowe TM, and Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955-964.

Marshall BJ, Barrett LJ, Prakash C, McCallum RW, and Guerrant RL. 1990. Urea protects *Helicobacter* (*Campylobacter*) *pylori* from the bactericidal effect of acid. *Gastroenterology* 99:697-702.

Mitsuhashi H, Nojima Y, Tanaka T, Ueki K, Maezawa A, Yano S, and Naruse T. 1998. Sulfite is released by human neutrophils in response to stimulation with lipopolysaccharide. *J Leukoc Biol* 64:595-599.

654 Mobley HL, Island MD, and Hausinger RP. 1995. Molecular biology of microbial ureases. *Microbiol Rev*
655 59:451-480.
656 Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer
657 BC, Monack DM, and Sonnenburg JL. 2013. Microbiota-liberated host sugars facilitate post-
658 antibiotic expansion of enteric pathogens. *Nature* 502:96-99.
659 Notebaart RA, Szappanos B, Kintses B, Pal F, Gyoerkei A, Bogos B, Lazar V, Spohn R, Csoergoe B, Wagner
660 A, Ruppin E, Pal C, and Papp B. 2014. Network-level architecture and the evolutionary potential
661 of underground metabolism.pdf. *Proc Natl Acad Sci U S A* 111:11762-11767.
662 Ouwerkerk JP, de Vos WM, and Belzer C. 2013. Glycobiome: bacteria and mucus at the epithelial
663 interface. *Best Pract Res Clin Gastroenterol* 27:25-38.
664 Pickard JM, Maurice CF, Kinnebrew MA, Abt MC, Schenten D, Golovkina TV, Bogatyrev SR, Ismagilov RF,
665 Pamer EG, Turnbaugh PJ, and Chervonsky AV. 2014. Rapid fucosylation of intestinal epithelium
666 sustains host-commensal symbiosis in sickness. *Nature* 514:638-641.
667 Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T,
668 Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan
669 M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-
670 Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S,
671 Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Meta HITC, Bork P, and
672 Ehrlich SD. 2010. A human gut microbial gene catalogue established by metagenomic
673 sequencing. *Nature* 464:59-65.
674 Quigley EM. 2013. Gut bacteria in health and disease. *Gastroenterol Hepatol (N Y)* 9:560-569.
675 Reichardt N, Duncan SH, Young P, Belenguer A, McWilliam Leitch C, Scott KP, Flint HJ, and Louis P. 2014.
676 Phylogenetic distribution of three pathways for propionate production within the human gut
677 microbiota. *The ISME journal* 8:1323-1335.
678 Ridlon JM, Kang DJ, and Hylemon PB. 2006. Bile salt biotransformations by human intestinal bacteria. *J*
679 *Lipid Res* 47:241-259.
680 Robbe C, Capon C, Coddeville B, and Michalski JC. 2004. Structural diversity and specific distribution of
681 O-glycans in normal human mucins along the intestinal tract. *Biochem J* 384:307-316.
682 Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, and Michalski JC. 2003. Evidence of
683 regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the
684 intestinal tract. *The Journal of biological chemistry* 278:46337-46348.
685 Roggentin P, Gutschker-Gdaniec G, Schauer R, and Hobrecht R. 1985. Correlative properties for a
686 differentiation of two *Clostridium sordellii* phenotypes and their distinction from *Clostridium*
687 *bifermentans*. *Zentralblatt fur Bakteriologie, Mikrobiologie, und Hygiene Series A, Medical*
688 *microbiology, infectious diseases, virology, parasitology* 260:319-328.
689 Rutherford JC. 2014. The emerging role of urease as a general microbial virulence factor. *PLoS Pathog*
690 10:e1004062.
691 Schwartz JJ, Gazumyan A, and Schwartz I. 1992. rRNA gene organization in the Lyme disease spirochete,
692 *Borrelia burgdorferi*. *Journal of bacteriology* 174:3757-3765.
693 Scott KP, Gratz SW, Sheridan PO, Flint HJ, and Duncan SH. 2013. The influence of diet on the gut
694 microbiota. *Pharmacol Res* 69:52-60.
695 Scott KP, Martin JC, Campbell G, Mayer CD, and Flint HJ. 2006. Whole-genome transcription profiling
696 reveals genes up-regulated by growth on fucose in the human gut bacterium "*Roseburia*
697 *inulinivorans*". *Journal of bacteriology* 188:4340-4349.
698 Sharma V, Prere MF, Canal I, Firth AE, Atkins JF, Baranov PV, and Fayet O. 2014. Analysis of tetra- and
699 hepta-nucleotides motifs promoting -1 ribosomal frameshifting in Escherichia coli. *Nucleic Acids*
700 *Res* 42:7210-7225.
701 Stahl M, Friis LM, Nothaft H, Liu X, Li J, Szymanski CM, and Stintzi A. 2011. L-fucose utilization provides
702 *Campylobacter jejuni* with a competitive advantage. *Proc Natl Acad Sci U S A* 108:7194-7199.
703 Stams AJ, Van Dijk JB, Dijkema C, and Plugge CM. 1993. Growth of syntrophic propionate-oxidizing
704 bacteria with fumarate in the absence of methanogenic bacteria. *Applied and environmental*
705 *microbiology* 59:1114-1119.
706 Suzek BE, Huang H, McGarvey P, Mazumder R, and Wu CH. 2007. UniRef: comprehensive and non-
707 redundant UniProt reference clusters. *Bioinformatics* 23:1282-1288.

ter Braak CFJ, and Smilauer P. 2012. *Canoco Reference Manual and User's Guide: Software for Ordination,Version 5.0.* Ithaca, NY, USA: Microcomputer Power.

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, and Pachter L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 31:46-53.

UniProt-Consortium. 2014. Activities at the universal protein resource (UniProt). *Nucleic Acids Res* 42:D191-D198.

Van den Bogert B, Boekhorst J, Herrmann R, Smid EJ, Zoetendal EG, and Kleerebezem M. 2013a. Comparative genomics analysis of *Streptococcus* isolates from the human small intestine reveals their adaptation to a highly dynamic ecosystem. *PloS one* 8:e83418.

van den Bogert B, Erkus O, Boekhorst J, de Goffau M, Smid EJ, Zoetendal EG, and Kleerebezem M. 2013b. Diversity of human small intestinal *Streptococcus* and *Veillonella* populations. *FEMS Microbiol Ecol* 85:376-388.

van Hijum SA, Kralj S, Ozimek LK, Dijkhuizen L, and van Geel-Schutten IG. 2006. Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. *Microbiol Mol Biol Rev* 70:157-176.

Vimr ER. 2013. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiol* 2013:816713.

Vimr ER, Kalivoda KA, Deszo EL, and Steenbergen SM. 2004. Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* 68:132-153.

Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, and Gordon JI. 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299:2074-2076.

Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma ZS, and Shi P. 2014. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *The ISME journal* 8:881-893.

Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CC, Troost FJ, Bork P, Wels M, de Vos WM, and Kleerebezem M. 2012. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *The ISME journal* 6:1415-1426.

Figure 1(on next page)

Circular map of the *R. ilealis* CRIB^T genome

Both chromosome and non-mobilizable plasmid are shown. For the chromosome tracks from inside to outside are as follows: 1, GC skew; 2, G+C content; 3, RNAs [rRNAs (blue), tRNAs (orange) and ncRNAs (purple)]; 4, all predicted protein CDS [with predicted function (light-blue), hypothetical proteins and proteins to which only a domain of unknown function could be assigned (grey)]; 5, genes or gene clusters of interest [(mobile genetic elements (red), Cas proteins (pink), urease gene cluster (yellow), choloylglycine hydrolase (black), gene clusters involved in carbohydrate utilization (green)]. For the plasmid tracks from inside to outside are as follows: 1, GC skew; 2, G+C content; 3, all predicted CDS.

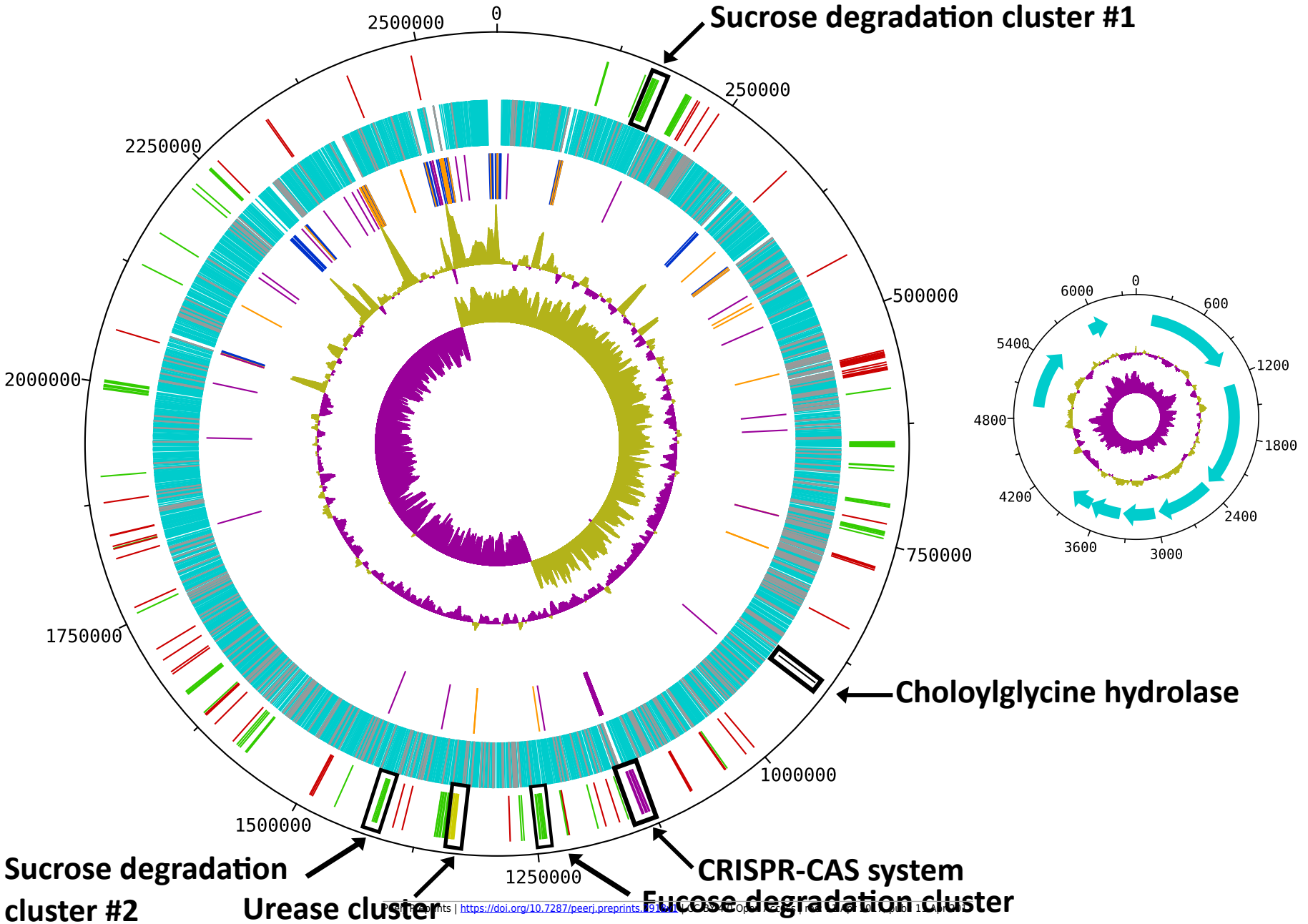
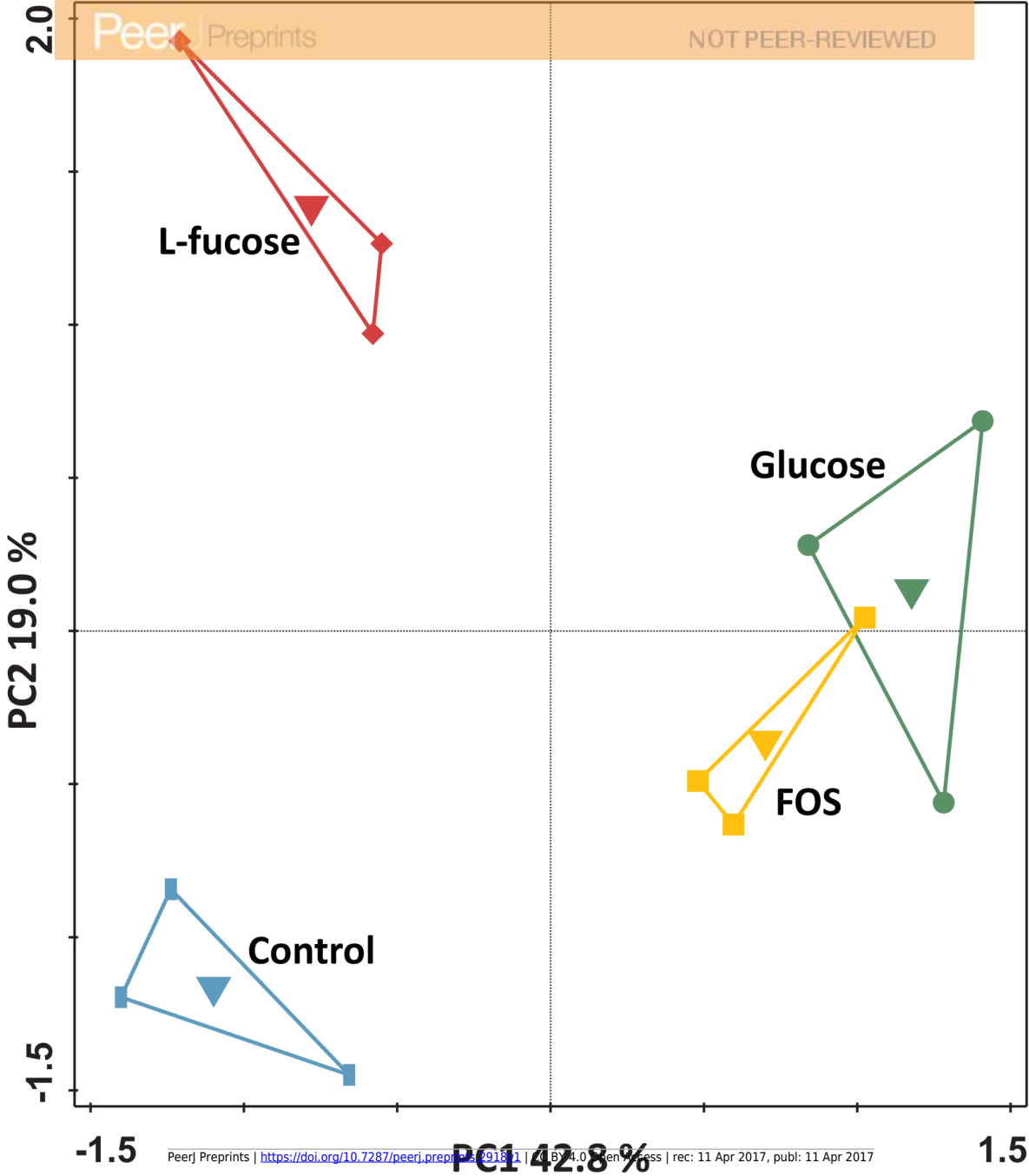


Figure 2(on next page)

Principal component analysis of the transcriptomes of *R. ilealis* CRIB^T grown on different carbohydrates (glucose, FOS and L-fucose) or in the absence of an additional carbon source (control).

First and second ordination axes are plotted, explaining 42.8 % and 19.0 % of the variability in the data set, respectively. Individual transcriptomes are symbol-coded by experimental condition: glucose (circles), FOS (squares), L-fucose (diamonds) and control (rectangles). The experimental conditions were used as supplementary variables as well and could explain 62.9 % of the variation.

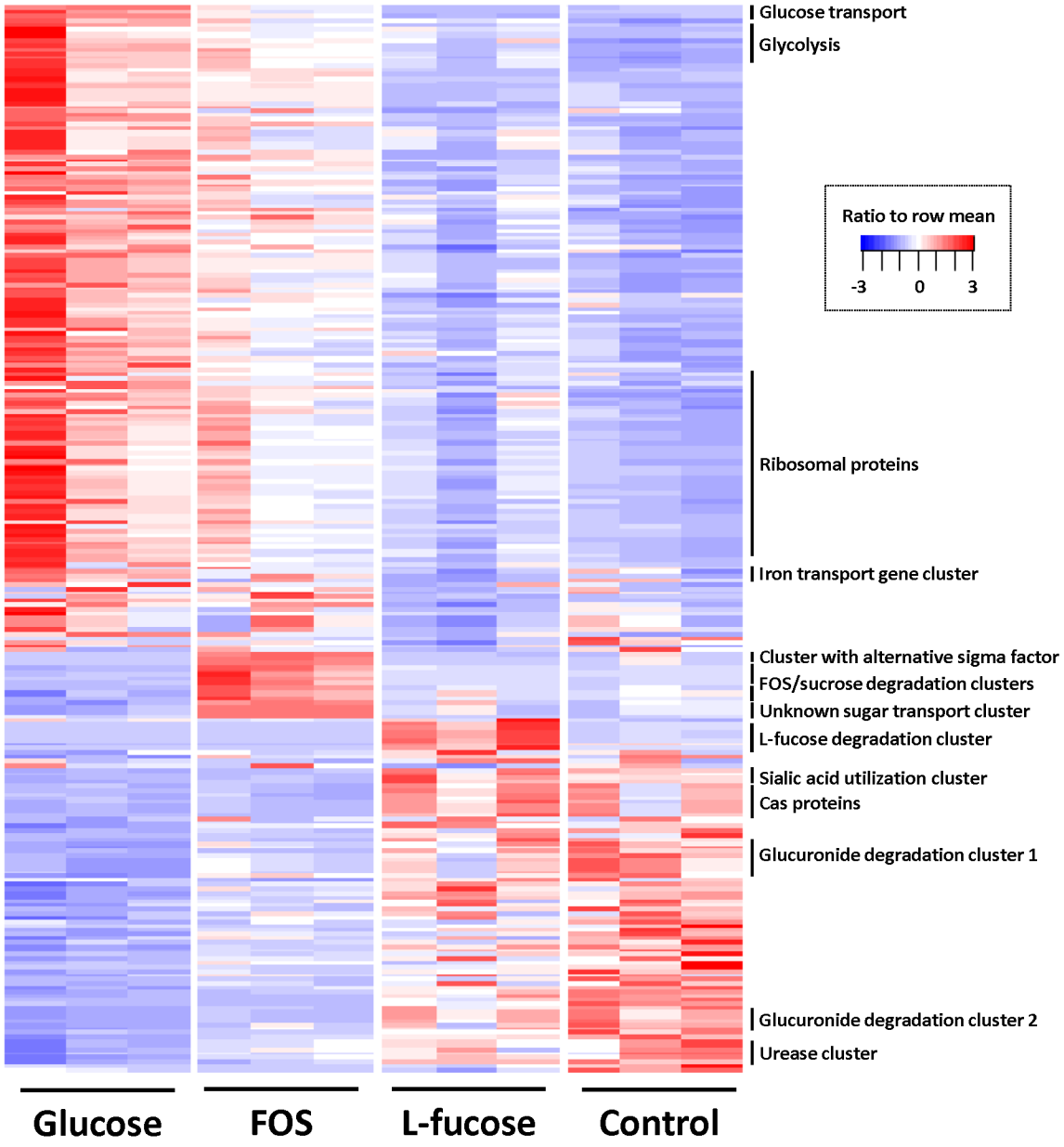


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2918v1 | CC BY 4.0 Open Access | rec: 11 Apr 2017, publ: 11 Apr 2017

Figure 3(on next page)

Heatmap of genes differentially expressed in at least one of the four conditions (≥1.5 log2(fold change) and q value ≤0.05).

Colour coding by ratio to row mean. Key gene clusters are indicated.

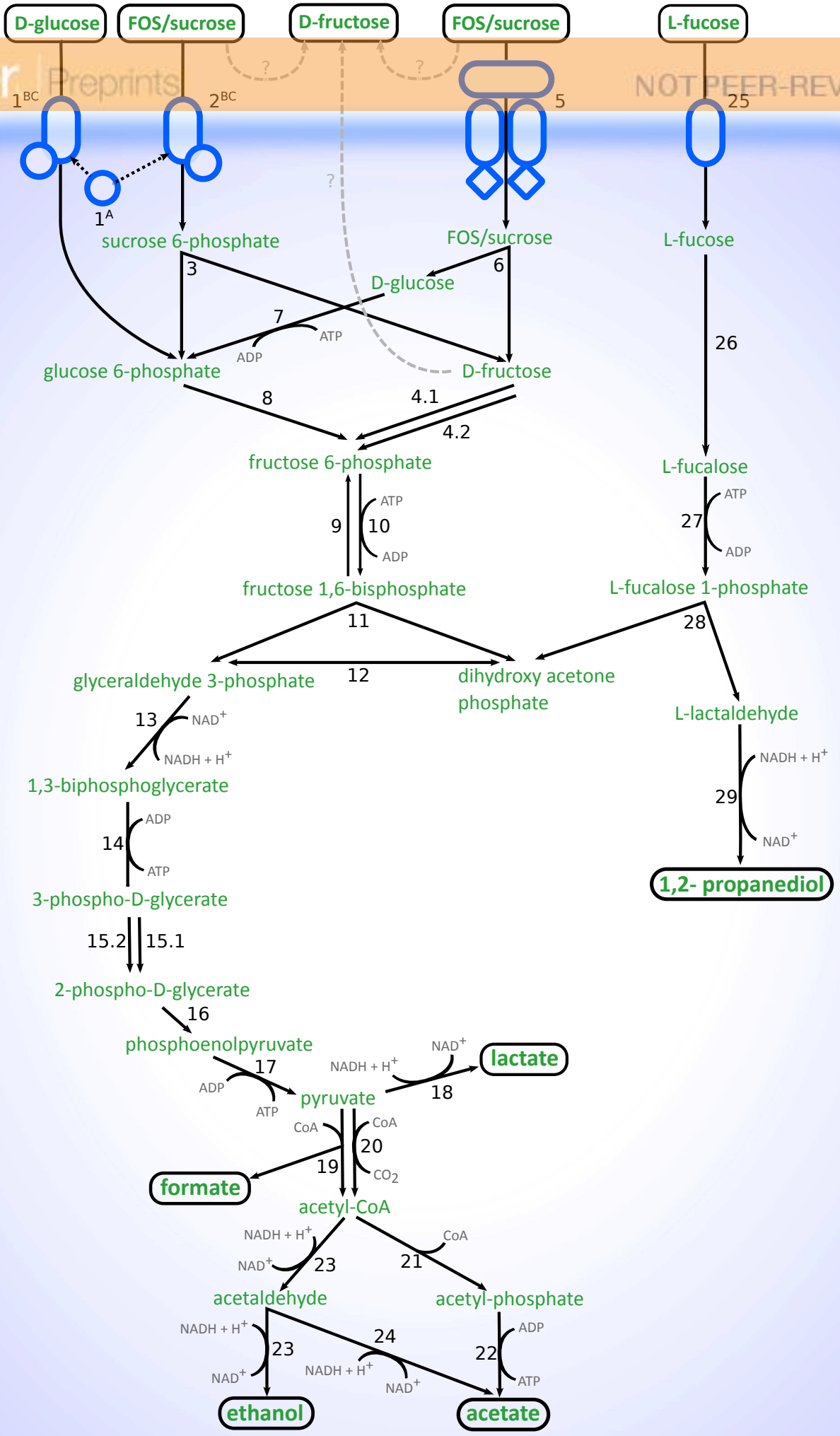


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2918v1 | CC BY 4.0 Open Access | rec: 11 Apr 2017, publ: 11 Apr 2017

Figure 4(on next page)

Schematic overview of the pathways involved in degradation of glucose, FOS and L-fucose in *R. ilealis* CRIB^T

1^A; PTS system glucose-specific EIIA component (CRIB_2018); 1^{BC}, PTS system glucose-specific EIIBC component (CRIB_2017); 2^{BC}, PTS system sucrose-specific EIIBC component (CRIB_1461); 3, ß-fructofuranosidase with RDD family protein (CRIB_1459 and CRIB_1460); 4, fructokinase (CRIB_152 and CRIB_1458); 5; ABC-type transporter (CRIB_148-CRIB_150); 6, ß-fructofuranosidase (CRIB_151); 7, glucokinase (CRIB_1849); 8, glucose 6-phosphate isomerase (CRIB_140); 9, fructose 1,6-bisphosphatase (CRIB_45 and CRIB_2020); 10, 6-phosphofructokinase ; (CRIB_104); 11, fructose-bisphosphate aldolase (CRIB_2223); 12, triosephosphate isomerase (CRIB_189); 13, glyceraldehyde-3-phosphate dehydrogenase (CRIB_187); 14, phosphoglycerate kinase; 15, phosphoglycerate mutase (CRIB_1223) and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (CRIB_190); 16, enolase (CRIB_191); 17, pyruvate kinase (CRIB_105); 18, L-lactate dehydrogenase (CRIB_684); 19, formate acetyltransferase (CRIB_2141); 20, pyruvate-flavodoxin oxidoreductase (CRIB_2021); 21, phosphate acetyltransferase (CRIB_2171); 22, acetate kinase (CRIB_1927); 23, bifunctional aldehyde-alcohol dehydrogenase (CRIB_2231); 24, fatty aldehyde dehydrogenase (CRIB_2231); 25, L-fucose permease (CRIB_1294); 26, L-fucose isomerase (CRIB_1298); 27, L-fuculokinase (CRIB_1297); 28, L-fuculose phosphate aldolase (CRIB_1297); 29, lactaldehyde reductase (CRIB_1300).



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2918v1 | CC BY 4.0 Open Access | rec: 11 Apr 2017, publ: 11 Apr 2017

Table 1(on next page)

General features of the *R. ilealis* CRIB^T genome.

1

	Chromosome	Plasmid
Size (bp)	2,581,778	6,145
G+C content (%)	27.9	29.3
Protein CDS	2,351	8
Pseudogenes	12	0
Coding density	1.10	1.02
Average gene size (bp)	899	531
rRNA genes		
16S rRNA genes	14	0
23S rRNA genes	14*	0
5S rRNA genes	14	0
tRNAs	109	0
ncRNAs	28	0
CRISPR repeats	1*71	0

2 * An additional 23S rRNA gene is expected in one of the gaps.

Table 2 (on next page)

Fermentation end products of *R. ilealis* CRIB^T produced during growth on different carbohydrates (glucose, FOS or L-fucose) or in basal medium in the absence of a carbon source (control condition)

Samples were obtained during mid-exponential phase (~8-10 h incubation; used for transcriptome analysis) and in stationary phase (24 h incubation). For the control cultures, fermentation products are shown for the individual cultures separating the carbohydrates used for preconditioning of the inoculum. For the three other conditions, values represent means of triplicate cultures with standard deviations. N.D., not detected.

1

	Formate		Acetate		Propionate		Lactate		1,2-propanediol	
	8-10 h	24 h	8-10 h	24 h	8-10 h	24 h	8-10 h	24 h	8-10 h	24 h
Control: basal medium	3.2	7.7	2.0	6.2	2.0	2.2	N.D.	N.D.	N.D.	N.D.
(glucose inoc.)	4.5	9.2	2.4	7.4	2.4	2.9	N.D.	N.D.	N.D.	N.D.
(FOS inoc.)	4.8	10.8	2.3	9.8	2.3	3.0	N.D.	N.D.	1.0	1.0
(L-fucose inoc)										
Basal medium + glucose (5 % w/v)	4.4±1.2	28.2±4.3	1.0±0.9	16.3±2.2	1.0±0.9	1.3±0.1	N.D.	3.0±0.7	N.D.	N.D.
Basal medium + FOS (5 % w/v)	4.7±0.6	27.3±2.5	1.4±0.0	17.7±1.4	1.4±0.0	1.6±0.1	N.D	2.5±0.3	N.D.	N.D.
Basal medium + L-fucose (5 % w/v)	6.7±0.1	19.5±3.6	2.8±0.1	16.3±2.9	2.8±0.1	2.8±0.4	N.D.	N.D.	1.3±0.1	7.7±1.4

2