COMPARATIVE WHOLE GENOME TRANSCRIPTIONAL RESPONSES OF *RUMINOCOCCUS ALBUS* STRAIN 7 AND 8: A CASE FOR SPECIALIZATION AND NICHE DIFFERENTIATION

BY

INHYUK KWON

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Roderick I. Mackie, Chair
Professor Isaac K. O. Cann
Associate Professor Carin K. Vanderpool
Assistant Professor Jason Ridlon

ABSTRACT

Ruminococcus albus strains are one of dominant fibrolytic bacteria in the rumen that contribute to plant biomass as well as vitamin utilization in host nutrition. To better understanding of host-microbe interactions, it is relevant to establish the model for fiber degradation and vitamin metabolism of the dominant fibrolytic bacteria and investigate their roles in the gut ecosystem. However, the fibrolytic mechanism and vitamin metabolism of *Ruminococcus albus* remain largely unknown. In the current study, comparative genomic and transcriptomic analyses of two different strains 7 and 8 of *R. albus* for plant fiber and folate utilization were used to investigate the conserved and differential mechanism between the two *R. albus* strains.

Through comparative transcriptomic analyses of both strains grown on alkaline peroxide hydrogen treated corn stalk (AHPCS), phosphoric acid swollen cellulose (PASC) and wheat arabinoxylan (WAX), this research demonstrated that the top 5 highly expressed glycoside hydrolase (GH) families, including the versatile GH5, GH9 (Cel9B), GH10, GH11, and GH48 (Cel48A), are the primary GH enzymes employed by both strains of *R. albus* for the hydrolysis of plant cell wall. In addition, the co-expression of these endoglucanases and endoxylanases in response to cellulose and hemicellulose was observed. The previously known adhering mechanism of *R. albus* were transcriptionally analyzed and verified in this research. The genes encoding Pil-like protein or a family 37 carbohydrate binding module (CBM37) domain were highly expressed in both strains during growth on different polysaccharides. Especially, the

significant role of CBM37 in the fiber utilization of *R. albus* was highlighted based on the prevalence of CBM37 domain on the highly expressed GH genes as well as hypothetical genes.

It is notable that distinct strategies between two strains for plant cell wall utilization were proposed in this research. Based on phenotypic, genomic, and transcriptomic evidence, wild type of *R. albus* 8 in rumen appears to preferentially utilize hemicellulose rather than cellulose embedded in the plant cell wall, while *R. albus* 7 prefers to utilize cellulose over hemicellulose. To support this conclusion, *R. albus* 8 utilized more hemicellosic sugars derived from the hydrolysis of AHPCS than *R. albus* 7. More CAZyme genes of *R. albus* 8 responded to WAX than PASC, while those genes of *R. albus* 7 responded to more PASC than WAX. When hemicellulose in AHPCS started to decrease in the culture, *R. albus* 8 down-regulated the expression of genes for sugar transporters and intracellular GH. In contrast, *R. albus* 7 exhibited a sequential expression of sugar transporters and intracellular GH genes, as preferred cellulosic sugars were released from AHPCS after removal of hemicellulose. Notably, we found the putative genes belonging to c-di-GMP regulatory and the accessory gene regulator quorum sensing (Agr QS) systems in *R. albus* 7 and 8. The transcriptional pattern of these genes were in accordance with differential transcriptional pattern of GH genes between both strains and the preferred planktonic growth of strain 8 on AHPCS as opposed to the substrate adherent growth of *R. albus* 7. These results suggest that c-di-GMP and Agr QS systems are implicated not only in biofilm formation of pathogenic bacteria, but also in the fibrolytic systems of commensal bacteria. Supported by the fermentation profile and the growth rate on beechwood xylan together with genomic and transcriptomic evidence, *R. albus* 8 was found to possess a predicted unique phosphoketolase (PK) pathway, which likely enables *R. albus* 8 to catabolize pentose rapidly as

well as conserve energy and costs for enzyme synthesis required for the lower glycolytic sequence. With our proposal for the differential strategies between strains, the co-culture experiment demonstrated that despite a similar fibrolytic mechanism, *R. albus* 7 and 8 could co-exist on complex substrate containing cellulose and hemicellulose.

This research on folate metabolism in *R. albus* 7 and 8 provided genomic evidence for three folate utilization pathways (either *de novo* synthesis, salvage, or both pathways) conserved in the Firmicutes including *R. albus* strains. Through the growth experiments in the presence or absence of folate and para-aminobenzoate (*p*ABA), it was shown that *R. albus* strains 7 and 8 rely on different folate metabolic pathways, *de novo* synthesis or salvage pathway, respectively. In addition, the results of transcriptomic analysis suggest that the folate autotrophic strain, *R. albus* 7, also has an alternative pathway for *p*ABA synthesis and likewise other *Ruminococcus* species lacking the canonical *p*ABA synthetic pathway are likely autotrophs and not auxotrophs.

Notably, the potential long non-coding RNA (lncRNA) loci was identified in the genomes of *R. albus* strains. The putative lncRNA loci consisted of four sequence components; lncRNA, DUF1292 gene, putative 6S RNA, and alcohol dehydrogenase. Based on their transcriptional profiles assessed by RNA-seq and northern blot analyses, it seems likely that the lncRNA loci are involved in the regulatory system related to the stationary phase of cells.

This study provides molecular insight in conserved and differentiated fibrolytic system and folate metabolism between *R. albus* 7 and 8. In addition, the presence of novel lncRNA loci was identified, providing more information on the regulatory mechanism in Gram-positive Firmicutes.

ACKNOWLEDGEMENTS

I would like to thank all of my family members, friends, and colleagues whose contribution made the work presented within this thesis a success. First, I would like to thank Professor Roderick I. Mackie for his selfless dedication to the training of me as an independent scientist and his guidance on life as well as science. I would also like to thank Professor Isaac K. O. Cann for his fruitful suggestions and advice that made this work intriguing and possible. Also, I would like to thank members of my thesis committee including Dr. Carin K. Vanderpool and Jason Ridlon who have always been willing to take time out of their busy schedules to discuss and provide me with insight into my research directions. I also appreciate Dr. Steven L. Daniel for taking his busy time to join and celebrate one of the biggest events in my life.

I have been very fortunate to work with many great people who have contributed to the joyful atmosphere in my laboratory. I will never forget our coffee break every Friday that enables all of people in Mackie's and Cann's laboratories to share an enjoyable experience of life and science. Also, I appreciate all of my colleagues who have supported and encouraged me to overcome many challenges throughout my research. Last, I would like to thank my loving family who have shown me unparalleled support.

TABLE OF CONTENTS

CHAPTER 1.

LITERATURE REVIEW

1.1 Roles of gut microbiota for utilization of complex carbohydrates in mammals

Mammalian genomes lack the enzymes required to deconstruct the structural polysaccharides, cellulose and hemicellulose, present in the plant cell wall. Instead, mammals rely on symbiotic microbes in the gut that are able to degrade dietary plant biomass and convert it to a usable energy source by the host. Large herbivores, such as ruminants and horses, rely considerably on resident gut microbes to gain energy from their primary diet, forage. The major source of energy for ruminants is volatile fatty acids (e.g. acetate, propionate, and butyrate) that are supplied by the ruminal microbes during forage fermentation. In addition, the ruminal bacteria are also utilized as protein source by ruminants after digestion by host digestive enzyme in the lower gut. It has been shown that the metabolic energy produced by gut microbes contributes up to 70 % of total dietary energy for large herbivores (1). Omnivores, notably humans, acquire additional energy from microbial fermentation of non-digestible dietary substrates in large intestine, but the hindgut fermentation products contribute to approximately 10 % of total energy supply to the host (2). Moreover, fermentation products of intestinal microbes play an important role in human health. For example, butyrate is a microbial fermentation product that is the main energy source for the colonic epithelium in human and boosts the intestinal immune system (3, 4). The amounts and types of dietary carbohydrates are known to be primary factors that shape microbial composition and affect their physiological activities in the human intestine (5, 6).

Most of the dietary fiber (plant-derived polysaccharides) enters into the rumen and large intestine in insoluble forms. In spite of a diverse community of microbes in the rumen, only a few microorganisms are capable of degrading the insoluble plant cell wall (7, 8). Other numerous groups of ruminal microbes are non-cellulolytic bacteria that rely on soluble oligosaccharides and polysaccharides, released by the primary fiber degrader for growth (9–11). In spite of capability of degrading cellulose and hemicellulose, some cellulolytic bacterial species are nutritionally specialized to only cellulose (7). For example, one of the predominant cellulolytic ruminal bacteria, *Fibrobacter succinogenes*, breakdown cellulose and hemicellulose in plant cell wall, but it does not transport and utilize the hydrolytic products of xylan (12–14). As a result, the solubilized hemicellulosic polysaccharides are released into the rumen environment and subsequently, utilized by other groups of bacteria. This cross-feeding between microbial species is an important metabolic feature in anaerobic microbial communities that involves fermentation products, such as lactate, ethanol, and hydrogen (15–18).

Recent metagenomic analyses based on 16S rRNA gene sequences have shown a diversity of microbial communities in mammalian gut (10, 11). Although there is compositional variation depending on diets, Bacteroidetes and Firmicutes are the predominant phyla in humans and ruminants. The representative genus in Bacteroidetes phylum is *Bacteroides* species in the human large intestine and *Prevotella* species in the rumen. The Firmicutes also contain gram-variable bacteria, but in the gram-positive bacterial group, Clostridial 16S rRNA sequence based XIVa, IV, and IX clusters are abundant in the gut microbiome.

Despite the fact that numerous microbial species exist in the gut, only a few bacterial species are known to be as specialist cellulolytic bacteria. In the rumen, the primary cellulolytic species are *Fibrobacter succinogenes*, *Ruminococcus albus*, and *Ruminococcus flavefaciens* (19,

20). *F. succinogenes* belongs to a discrete phylum of gram-negative bacteria and the two species of ruminococci are gram-positive Firmicutes belonging to the Clostridial IV cluster. As limited strains have been isolated and physiologically characterized, it is still possible that there may be unknown, but important cellulolytic species in the gut. Cellulolytic species are closely associated with plant surfaces, potentially through biofilm formation (21, 22). From this perspective, the tightly attached bacteria to plant fiber are considered to have the highest potential for fiber degradation. Molecular surveys have reported that there is spatial variation between two dominant phyla, Bacteroidetes and Firmicutes, in the gut. In rumen samples, 16S rRNA genes of Firmicutes are more abundant on the insoluble substrate associated fraction than Bacteroidetes (23). Similarly, in human feces that reflect microbial events in the distal colon, Firmicutes are more abundant in the fiber associated fraction than Bacteroidetes, and those Firmicutes include cluster IV Ruminococci showing the most extensive association with plant biomass (24–26). Among the insoluble associated species, each phylotype was shown to have a differential preference for colonization of particular insoluble polysaccharides in a continuous culture with human fecal inocula (25). The 16S rRNA sequence analysis showed that a different community attached to each substrate (wheat bran, resistant starch, and porcine mucin), including uncultured groups within clostridial cluster XIVa (24). For example, *Ruminococcus bromii* and *Bifidobacterium adolescentis* were the most abundant species on resistant starch containing medium, and *Bifidobacterium bifidum* and uncultured bacteria related to *Ruminococcus lactaris* were the most abundant on mucin.

Based on 16S rRNA gene sequencing, it has been confirmed that the predominant cellulolytic species in the rumen are *F. succinogenes, R. flavefaciens,* and *R. albus,* and their isolated strains have provided detailed information about functional mechanisms of plant cell

wall degradation in the gut through genomic, transcriptomic, and biochemical analyses. In addition, *Ruminococcus* species are also present in the human gut, which suggests that the cellulolytic mechanism deployed by ruminal species can be applied to closely related species inhabiting the human colon (27, 28).

1.2 Plant cell wall degradation by gut microbes

Plant cell wall structure. Plant cell walls consist of recalcitrant cellulose embedded in a hemicellulose matrix and lignin. The major structural component of the plant cell wall is cellulose, consisting of linear chains of β-1,4-linked glucose units. The recalcitrant nature of cellulose is due to the high degree of crystallinity, resulting from an extensive hydrogen bonding network as well as the stability of the glycosidic bonds. The hydrogen-bonds between cellulose microfibrils can be disrupted and form amorphous or even partially soluble cellulose. The amorphous forms are more accessible to enzymatic attack. Hemicellulose refers to polysaccharides of primarily pentose sugars and includes xylan, arabinan, and mannan. The dominant component of hemicellulose is xylan, β-1,4-linked xylose monomers substituted with arabinose, acetate, feruloyl ester, and glucuronic acid moieties. The substituent proportions vary between different xylan sources. For example, wheat arabinoxylan is composed of 66 % xylose and 33 % arabinose, whereas birchwood xylan is composed of 89% xylose, 1 % arabinose, and 8 % glucuronic acids (29, 30). Other components of hemicellulose can consist of different sugars in the backbone, such as xyloglucan (mixed sugar backbone with xylose and glucose) and glucomannan (mixed sugar backbone with mannose and glucose) (31).

Microbial enzymatic components required for plant cell wall breakdown. Plant biomass degradation requires hydrolysis of both cellulose and hemicelluloses and consequently, cellulolytic species need to employ a variety of enzymes, including glycoside hydrolases (GHs), polysaccharide lyases (PLs), and carbohydrate esterases (CEs) (32, 33). GHs are generally composed of catalytic domains and accessory domains that include non-catalytic carbohydrate-binding modules (CBMs), dockerins, and cell surface adhering modules (e.g. sortase motifs). In addition, domains of unknown function are often present in GHs enzymes (34). To date, GHs are classified into 133 families based on amino acid sequence similarity, secondary and tertiary structure, and catalytic mechanism (33, 35). These classifications are described in the Carbohydrate Active enZYme (CAZy) database (http://www.cazy.org/). Most of the GH families exhibit a high degree of substrate specificity (e.g. GH10 and 11 endoxylanases), whereas others seem to be promiscuous (e.g. GH5 enzymes). Based on cleavage site of the polysaccharide chain, GHs are often sorted to two classes, endo- or exo-acting enzymes. Endo-glucanases (EC 3.2.1.4) cleave β-1,4 linkages at random sites within glucan chains. Exo-glucanases (EC 3.2.1.91) bind to the ends of the glucan chain and processively cleave off repeating units of cellobiose (thus, they are often called 'cellobiohydrolase'). There is a mixed type of endoglucanase between endo- and exo-acting enzymes. A processive endoglucanase cleaves β-1,4 linkages anywhere along the glucan chain (a feature of endoglucanase) and then continues to cleave off the ends of the glucan chain processively, similar to exoglucanase (36, 37). Among total 133 families of GH, cellulolytic enzymes are found in GH families 5, 6, 7, 8, 9, 45, 48, 74, and 124 (38). The primary product of these cellulases is cellobiose. It has been shown that both exoglucanases and endoglucanases or processive endoglucanases are required for efficient and full degradation of

cellulose into cellobiose (39). Cellobiohydrolases are generally found in GH6, GH7, or GH48 families and processive endoglucanases have been primarily found in GH9 and GH5 families.

Non-catalytic CBMs are frequently found to be associated with the catalytic CAZyme domains (GH, PL, and CE).They are used in conjunction with catalytic domains to improve activity of the parent enzyme (40, 41). The typical role of CBMs is binding to substrate, which allows the catalytic GH domain to have increased proximity to its target substrate. Another function of CBMs is to modulate activity of its parent enzyme. For example, the CBM3 of GH9 endoglucanases promotes the parent enzyme to cleave the polysaccharide chain processively. In addition, some CBMs have been implicated in location of the appended enzyme to the bacterial cell wall (42, 43).

Compared to cellulose, hemicellulose breakdown requires a more diverse group of enzymes due to the complex polysaccharides contained in hemicelluloses. Hemicellulases can be categorized into two groups. One group of enzymes cleaves the main backbone of polysaccharides, such as xylan, and other group of enzymes cleaves the side chain constituents from the main backbone. In substituted xylan (e.g. wheat arabinoxylan), the backbone of xylose chain is cleaved by endo-xylanases. Both GH10 and GH11 enzymes hydrolyze long polymers of xylan into short oligomers. Subsequently, β -xylosidases cleave xylo-oligosaccharides into xylose monomers. Removal of substituents on xylan backbone requires diverse accessory enzymes, including α -L-arabinofuranosidases (removal of arabinose units), acetylxylan esterases (removal of acetyl groups), α-glucuronidases (removal of glucuronates), and ferulic acid esterases (removal of feruloyl esters) (30, 33). All of these enzymes act synergistically to release monosaccharides that directly enter into fermentation pathways.

Rumen cellulolytic Firmicutes. Two cellulolytic species within ruminal Firmicutes, *Ruminococcus flavefaciens* and *Ruminococcus albus* are Gram-positive bacteria. Both species are able to degrade and ferment both cellulose and hemicellulose in the plant cell wall. Many strains of *R. flavefaciens* show high cellulolytic activity on crystalline cellulose although there is some cellulolytic variation between strains for different types of plant cell wall and cellulose (44). *R. flavefaciens* is known to degrade plant cell wall using a cellulosome-type enzyme complex (10, 11, 38, 45, 46). The cellulosome paradigm has been established through intensive studies on fibrolytic soil bacteria, *Clostridium* species (45, 47). Numerous polysaccharidases are arranged on the surface of *C. thermocellum* through interaction between dockerin and cohesion domains that are present on catalytic enzymes and scaffoldins, respectively. This assembled enzymatic complex mediates cell attachment to plant cell wall. Of the gut cellulolytic Firmicutes, *R. flavefaciens* FD-1 harbors more than 10 putative scaffoldins and more than 200 putative dockerin-bearing proteins. The genes encoding four types of scaffoldins (ScaA, B, C, and E) and cellulose-binding protein (CttA) are clustered together on the genome of *R. flavefaciens*. Of four scaffoldins, ScaA and ScaC contain one or two cohesins that interact with dockerin containing catalytic enzymes, including diverse GHs, CEs, and PLs. The multi-cohesion bearing scaffoldin, ScaB, functions as a platform for anchoring ScaA and ScaC and also interacts with the cell surface anchoring scaffoldin, ScaE. As a result, the enzymatic complex, through dockerin-cohesin interaction on the scaffoldins, enables the cell to bind to and degrade plant cell wall (Fig. 1.1).

However, this cellulosomal system of *R. flavefaciens* is unlikely applied to other predominant cellulolytic species in the rumen, *R. albus*. *R. albus* was isolated together with *R. flavefaciens* as ruminal cocci in 1951. The isolated ruminococci from rumen were classified into

R. albus (white strains) and *R. flavefaciens* (yellow strains) based on the production of yellow pigment. Later, *R. albus* strain 7 was isolated and shown to produce a yellow pigment (48). In addition to differential pigmentation between *R. albus* strains, the yellow strain, *R. albus* 7, exhibited a better capability of degrading cellulose than the white strain, *R. albus* 8 (44, 49). Regarding the cellulosomal system, the genomes of *R. albus* strains do not possess any putative cohesins or scaffoldins that are key components of the cellulosome system, suggesting that this species use a different fibrolytic system. To date, it has been known that *R. albus* adheres to plant fiber using three mechanisms, including a Pil-like protein (CbpC), a glycocalyx, and a unique carbohydrate binding module 37 family (CBM37) (42, 50–53). Research on the adhesion defective mutants of *R. albus* have shown that the mutant strains lack type IV pilin proteins (CbpC for *R. albus* 8 and GP25 for *R. albus* 20), suggesting the Pil-like proteins are involved in adherence of *R. albus* species (50, 54). However, the homologous gene of CbpC in *R. albus* 7 did not respond transcriptionally to cellulose in contrast to *R. flavefaciens* showing strong up-regulation of the orthologous gene on cellulose (55, 56). Other adhesion defective mutants has been isolated from *R. albus* 8 and the mutant lacks two glycoside hydrolases, Cel9B and Cel48A (57). Interestingly, these GHs contain a novel CBM module, which was later classified as the CBM family 37 (52). The CBM37 module is conserved within only *R. albus* species and associated with various CAZymes of *R. albus*. Furthermore, the CBM37 shows a binding affinity to a broad range of polysaccharides as well as the bacterial cell wall (42, 52). Thus, it seems that the CBM37 functions as a cell surface-binding module as well as a substrate-binding module, potentially resulting in the localization of the extracellular catalytic enzymes on the cell surface.

In addition to a distinct fibrolytic system from that of *R. flavefaciens, R. albus* produces bacteriocins that likely function in a competitive manner. Both *R. albus* 7 and 8 are known to

produce a bacteriocin, albusin, which has a growth depression effect on *R. flavefaciens*, which likely account for the dominant population of *R. albus* when it was grown together with *R. flavefaciens* and *F. succinogenes* on cellulose (58–60).

Human gut Firmicutes. In the human gut, especially the colon, the two main families of Firmicutes, the Lachnospiraceae and Ruminococcaceae, are found, accounting for 20 to 30 % of total bacteria (61). Depending on type of non-digestible carbohydrates in diet, there is a variation in the relative abundance between those two families of Firmicutes (6, 62). It is known that high intake of resistant starch increases the abundance of Ruminococcaceae species including *R. bromii*, while high intake of wheat bran promotes Lachnospiraceae including *Roseburia* spp. and *Eubacterium rectale* (6). So far, *Ruminococcus champanellensis* is the only cellulolytic Firmicutes isolated from the human gut (27, 63). The genome of *R. champanellensis* encodes a similar array of putative cellulases to those of the cellulolytic ruminococci in rumen, including GH5, GH9, GH48, and GH74 CAZymes. In addition, *R. champanellensis* possesses various putative dockerin bearing CAZymes and 11 scaffoldins containing cohesins and dockerins, suggesting that it appears to use a cellulosomal system similar to *R. flavefaciens* (28).

Rumen Bacteroidetes (*Prevotella* species). Many highly abundant bacterial species isolated from the mammalian gut cannot degrade intact plant cell wall or crystalline cellulose. Nevertheless, they produce multiple catalytic enzymes that are capable of degrading hemicelluloses (e.g. xylan) and amorphous cellulose. The non-cellulolytic species include ruminal Bacteroidetes (*Prevotella* species) and Firmicutes (*Butyrivibrio fibrisolvens*, *Roseburia* speices, and *Eubacterium rectale*) (10). It appears that these microbes utilize dietary soluble polysaccharides or solubilized polysaccharides released by other cellulolytic microbes (18, 64). The non-cellulolytic Bacteroidetes, *Prevotella bryantii*, was isolated from the rumen and is able

to utilize only soluble xylans and its oligomers (65, 66). It has been known that two gene clusters, encoding GH10 endoxylanase and GH43 β-xylosidase, play an important role in the utilization of xylan by *P. bryantii* (67). These two clusters are induced by a hydrid two component regulator in response to xylo-oligosaccharides. This transcriptional phenomenon is shown to function in the degradation of xylan in *Bacteroides ovatus* (68). In addition, the bulk of xylanase activity is located in the periplasm or membranes, rather than on the cell surface, which is similar to the polysaccharide utilization loci (PUL) system of human gut *Bacteroides* genus within Bacteroidetes phylum (66). Taken together, it seems likely that the *Bacteroides* and *Prevotella* genera within gut Bacteroidetes phylum share a polysaccharide utilizing system (Fig. 1.1).

Recently, the genome of uncultured Bacteroidetes species has been reconstructed using metagenomic analysis from the cow rumen (69). Interestingly, the genome possesses a PUL system for cellulose utilization. Through biochemical analyses, it turns out that the cellulolytic PUL encodes two cellulases (GH5 and GH9) as well as a putative cellobiose phosphorylase (GH94) together with SusCD-like proteins. So far, none of cellulolytic Bacteroidetes species have been isolated and their PUL systems have not shown the cellulolytic activities. Thus, it is a novel finding showing the diverse bacterial mechanism for cellulose degradation in the gut.

Human gut Bacteroidetes. The most common and abundant genera of Bacteroidetes in the human gut are *Bacteroides* and *Prevotella* (70, 71). Because of several isolated strains, their available genome sequences, and genetic tools, the mechanisms of carbohydrate utilization have been intensively investigated in detail in *Bacteroides* species, rather than any other microbial species in human gut. *Bateroides* genus demonstrates the capability of utilizing a broad range of substrates, including dietary polysaccharides as well as host-derived glucans (72–74). To breakdown and utilize those substrates, two polysaccharide utilization systems are employed by

Bacteroides species; starch utilization system (Sus) and Sus-like PUL system. The Sus system of *B. thetaiotaomicron* is composed of eight proteins (SusRABCDEFG). SusCDEFG are localized on the outer membrane and play roles in binding (SusDEF), degrading (SusG), and importing (SusC) of soluble starch into the periplasm (Fig. 1.1). Of those proteins, SusC (a TonB-dependent transporter) and SusD (functioning as binding substrate to the cell surface) are critical components in Sus system because SusC cannot bind the substrate alone and requires SusD to initiate.

The Sus system has been established as a paradigm for glycan utilization, and further expanded to the mechanism for the different polysaccharides utilization in the Bacteroidetes (68). *B. thetaiotaomicron* harbors 88 PULs that contain SusCD like proteins. In addition, the PULs contain CAZymes, required to the hydrolysis of the specific polysaccharides and response regulators that tune the expression level of the PULs. The transcriptomic analyses showed that each PUL is upregulated by specific substrates including xylan, β-glucan, and galactoglucomannan. Such specific responses of PULs are mediated by the hybrid two-component regulators that recognize specific linear oligosaccharides (68). Xylanolytic PULs were investigated in *B. ovatus, B. intestinalis* and ruminal Bacteroidetes such as *P. bryantii*. In *P. bryantii*, the PUL includes *susCD* orthologs, two-component regulators, and endoxylanase (GH10), which were highly expressed in response to wheat arabinoxylan (67, 75). This xylanolytic PUL is found on the genomes of diverse *Prevotella* species as well as *Bacteroides* species, suggesting that this mechanism is likely to apply to xylan-debranching gut Bacteroidetes (67, 76). Similar to *P. bryantii*, two PULs, encoding xylan debranching enzymes and *susCD* othologs, in *B. ovatus* and *B. intestinalis* were strongly up-regulated in response to xylan (68, 75, 77). Taken together, the detailed organization of xylanolytic PULs varies between Bacteroidetes

species, but the overall components of the PULs, such as SusCD-like proteins and hydrolytic enzymes, are well conserved within the Bacteroidetes phylum in human gut and rumen.

Rumen cellulolytic *Fibrobacter succinogenes*. *Fibrobacter succinogenes*, a Gram-negative rod-shaped bacterium, was originally isolated as *Bacteroides succinogenes* from rumen (78). Although *F. succinogenes* is capable of degrading both cellulose and hemicellulose, this bacterium utilizes only cellulosic derived hexose sugars (e. g. cellobiose and cello-oligosaccharides) released from cellulose hydrolysis (13, 79). Thus, it has been hypothesized that the hemicellulolytic activity of this microbe serves as a method to gain access to cellulose embedded in hemicellulose matrix. *F. succinogenes* possesses a large number of GH genes including cellulases and hemicellulases on the genome, but it lacks either putative exo-glucanases (GH6, GH7, or GH48) or processive endo-glucanases (GH5 or GH9), which differ from a typical feature of enzymatic mechanism of cellulolytic ruminococci for the fiber degradation. Thus, it appears that *F. succinogenes* employs a novel cellulolytic system, differing from other cellulolytic Firmicutes. It has been proposed that the glucan chains are channeled into the periplasm and then hydrolyzed into short oligomers in *F. succinogenes* (Fig. 1.1)(79). Neither cellulosomal components (e.g. dockerin and cohesion) nor CBM37 like domains are present on the genome of *F. succinogenes*. Instead, many GH genes including cellulases of *F. succinogenes* contain a highly basic carboxxy-terminal domain (BTD), but its function is unknown (80, 81).

1.3 Sugar metabolism in gut microbes

The hydrolysis of a various polysaccharides in plant cell wall generates a mixture of hexoses, pentoses, and their derivatives. In general, these sugars are catabolized for the generation of energy and conversion to precursors for biosynthetic reactions. There are a variety of pathways for sugar catabolism in bacteria. Herein, I discuss the catabolic pathway for glucose and xylose that are the most abundant form of hexose and pentose in the plant cell wall. Most heterotrophic bacteria, including aerobes and anaerobes, are able to metabolize glucose by the glycolytic pathway that converts glucose into pyruvate. During this process, the free energy is released and conserved in the form of high-energy compounds, ATP and NADH. The most common pathway of sugar metabolism is glycolysis (the Embden-Meyerhof-Parnas pathway; EMP pathway), but some bacteria metabolize glucose through unique pathways that are not found in eukaryotes (e.g. the Entner-Doudoroff pathway and phosphoketolase pathway).

Embden-Meyerhof-Parnas (EMP) and Entner-Doudoroff (ED) pathways. In EMP pathway, the enzymatic reaction steps are described in Fig. 1.2 and the overall reaction for the EMP pathway is given as follows:

$$\text{Glucose} \rightarrow 2\ \text{Pyruvate} + 2\ \text{ATP} + 2\ \text{NADH}$$

Based on the common and unique steps between the EMP and other pathways (e.g. ED pathway and pentose phosphate pathway), the EMP pathway is divided into two parts; the unique part in the upper glycolysis (from glucose-6-phosphate to fructose-1,6-bisphosphate) and the common part in the lower glycolysis (from glyceraldehyde-3-phosphate to pyruvate). In the upper glycolysis pathway, the first step is the phosphorylation of glucose to glucose-6-phosphate using ATP. Subsequently, glucose-6-phosphate is isomerized to fructose-6-phosphate and more

phosphorylated into fructose-1,6-bisphosphate by phosphofructokinase (PFK). PFK is the key enzyme of the EMP pathway. Thus, if a bacterium harbors this enzyme, the bacterium is considered to be a EMP pathway harboring organism. Fructose-1,6-phosphate is cleaved by fructose-1,6-phosphate aldolase into two molecules of triose-phosphate, glyceraldehyde-3-phosphate (G-3-P) and dihydroxyacetone-phosphate (DHAP). DHAP is isomerized to G-3-P and resulting two molecules of G-3-P enter into the lower glycolysis pathway that generates energy. G-3-P is oxidized to 1,3-bisphosphoglycerate coupled with reduction of NAD+ by dehydrogenase. ATP is generated through the conversion of 1,3-biophosphoglycerate to 3-phosphoglycerate by kinase. 3-phosphoglycerate is converted to 2-phosphoglycerate and subsequently, converted to phosphoenolpyruvate (PEP) by mutase and enolase. Additional ATP is generated through the conversion of PEP to pyruvate by kinase.

The Entner-Doudoroff (ED) pathway has only been found in prokaryotes. The ED pathway bearing microorganisms, such as *Pseudomonas* species, do not harbor the key enzyme (PFK) of the EMP pathway. Thus, glucose-6-phosphate is converted to 6-phosphogluconate, coupled with NADP+ reduction, and dehydrated to 2-keto-3-deoxy-6-phosphogluconate (KDPG). Then, KDPG is cleaved by aldolase into pyruvate and G-3-P that is converted into pyruvate through the lower glycolysis pathway (Fig. 1.2). As half molecule of glucose bypasses the energy generating lower glycolysis pathway, the overall reaction for the ED pathway is as follow:

$$\text{Glucose} \rightarrow 2\ \text{Pyruvate} + 1\ \text{ATP} + 1\ \text{NADH} + 1\ \text{NADPH}$$

Compared to the EMP pathway, metabolizing glucose through the ED pathway conserves only one ATP per molecule of glucose. Thus, it has been suggested that the primary function of the ED pathway is to metabolize not glucose, but rather sugar acids like gluconate (82). This

hypothesis seems reasonable, considering that gluconate cannot be metabolized through the EMP pathway and the ED enzymes lacking *Escherichia coli* strains are unable to grow on gluconate (83). However, it cannot be applied to the microorganisms that metabolize glucose through the ED pathway, including *Pseudomonas saccharophilus* and *Zymomonas mobilis* (84). From this perspective, Flamholz et al. (2013) suggested that the energy loss through the ED pathway, relative to the EMP pathway, can be compensated by saving the cost for protein synthesis (85). Through the thermodynamic and kinetic analyses, the authors predicted that the EMP pathway requires several fold more catalytic enzyme production to reach the same glucose conversion rate to pyruvate, as compared to the ED pathway. Thus, the ED pathway likely enables the cell to produce less catalytic enzymes, which allows the ED harboring bacteria to use a similar net energy for the growth with the EMP harboring bacteria.

Pentose phosphate (PP) pathway. Pentose fermenting bacteria can use pentose as energy source using the pentose phosphate (PP) pathway (86). In this pathway, pentoses, such as xylose, arabinose or ribose, are phosphorylated and then, isomerized to xylulose-5-phosphate and ribose-5-phosphate (Fig. 1.2). The pentose-5-phosphates are transformed to fructose-6-phosphate (F-6-P) and glyceraldehyde-3-phosphate (G-3-P) through carbon rearrangement by transketolase and transaldolase. Subsequently, both F-6-P and G-3-P are metabolized through the EMP pathway for the energy generation. The PP pathway is not only used to metabolize pentose, but also used to synthesize essential compounds for the cell growth. Of the intermediates in the PP pathway, ribose-5-phosphate is the precursor for nucleotide synthesis and erythrose-4-phosphate is used to synthesize aromatic amino acids (e.g. phenylalanine, tyrosine, and tryptophan) through the shikimate pathway. This is why about 28 % of glucose is metabolized through the PP pathway when *E. coli* grows on glucose as the sole carbon source (86).

Phosphoketolase (PK) pathway. Phosphoketolases (PK; EC 4.1.2.9, EC 4.1.2.22) are key enzymes of the phosphoketolase pathway for energy conservation of heterofermentative lactobacilli, *Bifidobacterium bifidum*, and *Clostridium acetobutylicum* (86). Depending on substrates, three types of phosphoketolases have been biochemically characterized; xylulose-5-phosphate (Xu-5-P) phosphoketolase, fructose-6-phosphate (F-6-P) phosphoketolase, and xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (XFP). Xu-5-P PK enzyme cleaves Xu-5-P with inorganic phosphate into glyceraldehyde-3-phosphate (G-3P) and acetyl-phosphate, and F-6-P PK enzyme cleaves F-6-P with inorganic phosphate into erythrose-4-phosphate (E-4P) and acetyl-phosphate. XFP enzyme is able to cleave both substrates, Xu-5-P and F-6-P.

PK pathway in heterofermentative lactobacilli. Generally, lactobacilli are classified into two groups based on fermentation products; homofermentative (fermenting lactate only) and heterofermentative (fermenting acetate, ethanol, and lactate) lactobacilli (86). The former group ferments glucose through the EMP pathway, but the latter group uses the PK pathway for fermenting glucose. Heterofermentative *Lactobacillus* species, including *L. brevis*, *L. fermentum*, and *L. mesenteroides*, produce lactate, ethanol, and CO_2 from glucose via the PK pathway (Fig. 1.3). These bacteria do not use the EMP pathway. Instead, they oxidize glucose-6-phosphate (G-6-P) to ribulose-5-phosphate (Ru-5-P), followed by conversion to Xu-5-P by an epimerase. The Xu-5-P specific PK enzyme then, splits Xu-5-P into G-3-P and acetyl phosphate. The former molecule is used to generate ATP through the lower glycolytic pathway, and the latter molecule is reduced to ethanol to regenerate NAD^+ that is required for the glucose oxidation steps. From this PK pathway, heterofermentative lactobacilli produce only one ATP per molecule of glucose. However, during growth on pentose, two molecules of NAD^+ are not reduced, which enables the

cell to use acetyl phosphate for the energy conservation pathway through acetate production. For this reason, heterofermentative bacteria produce one more ATP from pentoses than hexoses.

Bifidum pathway. *Bifidobacterium bifidum* has two phosphoketolases, each active on fructose-6-phosphate (F-6P) and xylulose-5-phosphate (Xu-5-P)(86). This bacterium ferments sugars into acetate and lactate via the bifidum pathway. As shown in Fig. 1.3, F-6-P is cleaved into erythrose-4-phosphate (E-4-P) and acetyl-phosphate, and the E-4-P is rearranged with other F-6P into Xu-5-P which is further metabolized into lactate and acetate as in heterfermentative lactobacilli. From 1 molecule of glucose, *B. bifidum* produces 2.5 ATP using the bifidum pathway, but no energetic advantage from pentose fermentation, which is different from the PK pathway of heterofermentative bacteria.

Phosphoketolase pathway in *Clostridium acetobutylicum*. Of the known cellulolytic bacteria, *C. acetobutylicum* harbors both PP pathways and PK pathway for pentose metabolism (Fig. 1.4) (87–90). In contrast to lactate producing bacteria, the PK pathway is used to metabolize pentoses in *C. acetobutylicum*. The XFP enzyme of *C. acetobutylicum* cleaves Xu-5-P, transformed from pentoses, into G-3-P and acetyl-phosphate, bypassing most of the PP pathway. This feature of the PK pathway results in less ATP production in catabolizing pentose, as compared to the PP pathway. Thus, it was puzzling which pathway is dominant when the cells are grown on pentose. Through the metabolic flux analyses, it turns out that when *C. acetobutylicum* grows on xylose as the sole carbon source, up to 40 % of xylose is metabolized through the PK pathway, and the PP pathway consumes the remaining 60 %. However, the transcriptional analysis showed that the XFP gene was induced by arabinose greater than xylose. Accordingly, when feeding arabinose alone or arabinose plus xylose, the PK pathway is more dominant than the PP pathway. Interestingly, it seems that the PK pathway provides a growth

advantage to the cells. When *C. acetobutylicum* grows on the mixture of pentoses, the cells preferentially utilize arabinose over xylose. Indeed, the growth rate of cells is also higher when grown on arabinose, relative to xylose (87, 91). However, it is still unknown why the PK pathway is differentially regulated between two pentoses and how the PK pathway provides a growth benefit to *C. acetobutylicum*.

1.4 B-Vitamin metabolism of rumen bacteria

Together with vitamin C, B-vitamins are a class of water-soluble vitamins composed of eight vitamins (B1, B2, B3, B5, B6, B9, and B12). They all are essential micronutrients that play important roles in growth metabolism of all living cells. The active form of vitamin B1 (thiamine) is the coenzyme named thiamine pyrophosphate (TPP), which is involved in the enzymatic conversion of pyruvate to acetyl-CoA. Thus, thiamine plays a critical role in the energy (e.g. ATP) generation from carbohydrates. It is also involved in RNA and DNA synthesis, as well as nerve function in human (92). Vitamin B2 (riboflavin) is the precursor of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). Both are essential cofactors that function as electron carriers in energy conservation pathways including the electron transport chain and the TCAcycle, as well as the fatty acid catabolism through beta-oxidation (93). Vitamin B3 (niacin) is essential precursor for two coenzymes: nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). Both coenzymes function as electron carriers involved in various metabolic pathways such as the glycolysis, the TCA cycle, nucleic acid synthesis, et cetera. (94). Vitamin B5 (pantothenic acid) is an essential precursor for coenzyme A that is important for the synthesis of fatty acid, amino acids, ketones, cholesterol,

and phospholipids (95). The active form of vitamin B6 (in three forms of pyridoxine, pyridoxal, and pyridoxamine) is pyridoxal 5'-phosphate (PLP), which functions as a cofactor involved in amino acid, glucose, and lipid metabolism (96). Vitamin B7 (biotin) is an essential coenzyme of four carboxylases: acetyl-CoA carboxylase, propionyl-CoA carboxylase, B-methylcrotonyl-CoA carboxylase, and pyruvate-CoA carboxylase. These four enzymes are involved in various metabolic pathways including fatty acid synthesis, gluconeogenesis, amino acids and cholesterol metabolism (97). The active forms of vitamin B12 (cobalamin) are methylcobalamin and adenosylcobalamin. Both coenzymes are required for enzymatic reactions in the cellular metabolism of carbohydrates, proteins, and lipids (98).

Vitamin B9 (Folic acid) is composed of three molecules (pterin, *p*-aminobenzoate (*p*ABA), and glutamate) to which one-carbon units at various oxidation levels can be attached at the N5 and N10 positions (99). Tetrahydrofolate (THF), a folate derivative, serves as cofactor in one-carbon transfer reaction required for the synthesis of purines, formylmethionyl-tRNA, thymidylate, pantothenate, glycine, serine, and methionine (Matthews RG, 1996). Plants, fungi, and most bacteria make folates de novo, starting from GTP and chorismate, most animals, including humans, lack key enzymes of the synthetic pathway and so folate must be supplied through the diet. However, ruminants in which the rumen is fully functioning are independent of a dietary supply of folate due to biosynthesis by rumen bacteria (100). The rumen bacteria that synthesize folate inside their cells are passed to the abomasum along with digested feed and release their vitamin store to host after digestion by host enzymes. For these reasons, folate synthesis and salvage pathways have been extensively studied in model organisms, such as *Escherichia coli* and Lactobacilli.

The vitamin requirements of the predominant cellulolytic species in the rumen were studied by pinoneers of anaerobic microbial ecology decades ago. A total of seven strains of *Fibrobacter succinogenes* have been studied (101) and all except one had an absolute requirement for biotin. For the exception, strain growth was also stimulated by biotin. The remaining B-vitamins, including folate, did not affect the growth of *F. succinogenes*. The vitamin requirements of Ruminococci were reported by Bryant and Robinson (102) and their requirement differ by strain. All three strains of *R. flavefaciens* had a requirement for biotin and no requirement for folate, similar to *F. succinogenes*. Similarly, all nine strains of *R. albus* studied had an absolute requirement for biotin, but considerable variation in folate requirement was found between individual strains. Some *R. albus* strains, such as *R. albus* 7 and B199, did not require folate for growth, but other strains, such as *R. albus* 20 and $B_3$37, had an absolute requirement for preformed folate. As the genomes of both *R. albus* 7 and 8 have been sequenced, a variety of studies have been conducted to determine their fiber degradation mechanism through genomic, transcriptomic, and proteomic analysis. However, vitamin metabolism of *R. albus* strains is largely unknown, while the vitamin requirement of *R. albus* 8 has not been studied yet.

The folate *de novo* synthesis pathway has the same steps in bacteria and plants, consisting of a pterin branch and a *p*ABA branch. The first reaction of the pterin branch is the conversion of GTP to 7,8-dihydroneopterin triphosphate by GTP cyclohydrolase I (*folE*) (Green JBPM and Matthews RG, 1996). The resulting 7,8-dihydroneopterin is converted into 6-hydroxymethyl-7,8-dihydropterin in three consecutive steps by a specific pyrophosphatase (*folQ*), dihydroneopterin aldoase (*folB*), and hydroxymethyldihydropterin pyrophosphokinase (*folK*) (103). Subsequently, dihydropteroate synthase (*folP*) condenses 6-hydroxymethyl-7,8-dihydropterin with *p*ABA either from *de novo* synthesis or salvage pathway. The resulting dihyropteroate is glutamylated by

dihydrofolate synthase (*folC*), which is reduced to tetrahydrofolate by dihydrofolate reductase (*folA*) (103). In the *p*ABA branch pathway, chorimate is aminated to amonodeoxychorismate by 4-amino-4-deoxychorimate synthase (*pabAB*) and, subsequently, converted to *p*ABA by 4-amino-4-deoxychorismate lyase (*pabC*) (104, 105).

Three kinds of the folate salvage pathways have been studied. The first pathway is known as the intact folate salvage pathway that enables use of dietary or exogenous folate and dihydrofolate (DHF), and DHF reductase (DHFR) reduces these oxidized folates to THF (106). DHFR activity is also required to recycle the DHF that is produced in the thymidylate synthase reaction. The second pathway is known as the pterin salvage pathway studied in *Leishmania* and other tyrpanosomatid parasites. This pathway involves the reduction of fully oxidized pterins to the dihydro- and tetrahydro-pterins by pteridine reductase (107). After these reactions, the dihydro form is used for folate synthesis and tetrahydro-form is used as cofactors for aromatic hydroxylases and other pterin dependent enzymes. In the last salvage pathway, the pterin and *p*ABA-glutamate fragments produced by folate breakdown are recycled for folate synthesis (108). This pathway is known in some bacteria and plants. However, few studies have been conducted and so further biochemical research is required.

1.5 Quorum sensing in Gram-positive bacteria

Many bacteria are known to communicate with other cells using chemical signal molecules- this is termed quorum sensing (109, 110). Quorum sensing (QS) bacteria secrete low molecular weight signaling molecules (autoinducers) whose extracellular concentration increases as a function of increasing cell population. Bacteria monitor the external level of autoinducer and

change behavior by altering gene expression once the concentration of autoinducer reaches at threshold level. These QS system allows bacteria to function as a multicellular organism showing a synchronized behavior on a population wide scale. However, Gram-negative and Gram-positive bacteria employ different types of QS systems.

In Gram-negative bacteria, most of the autoinducers are *N*-acyl-homoserine lactones (HSLs), but Gram-positive bacteria use generally peptides (autoinducing peptides; AIPs). In addition, the extracellular HSL freely diffuses in and out of the cells and binds to the cytoplasmic autoinducer receptor protein, a DNA binding transcriptional activator. This binding event results in allosteric unfolding of the activator, which allows the dimerization and subsequently, activates transcription of the regulated genes (109). In contrast, the extracellular AIPs do not diffuse into the cells, but binds to specific receptors, two component type histidine kinases. Signaling mediated by a phosphorylation cascade activates or represses the target genes including the AIP synthesis genes (110, 111).

In general, it is assumed that QS regulation play a biological role, derived from the canonical QS system of the bioluminescent marine bacterium, *Vibrio fisheri*. When the growth of symbiotic species reaches at high density in the light organ of the Hawaiian squid, they provide the light for the host animal, representing their part of the symbiotic relationship with the host animal (112, 113). However, the biological rationale for QS is not very clear with other bacteria except for biofilm forming pathogenic bacteria. In this review, I describe the regulatory features and biological significance of the well-known *agr*-based QS system in Gram-positive staphylococci.

Agr Quorum sensing system in biofilm associated bacteria, staphylococci. *Staphylococcus aureus* is normally a commensal bacterium in human gut, but it becomes a deadly pathogen once it penetrates into host tissues (114). *S. aureus* exhibits a biphasic behavior depending on cell density. At low cell density, the bacteria produce proteins required for the cell attachment and colonization on surfaces. At high cell density, the bacteria down-regulate these traits and initiate secretion of toxins and proteases, which are required for dissemination and consequently, cause serious human disease (109, 110). Quorum sensing via the accessory gene regulator (agr) system controls this switch in gene expression related to biofilm formation and virulence. The Agr QS locus is composed of four genes; autoinducing peptide (*agrD*), processing and exporting protein (*agrB*), and two-component sensor kinase (*agrC*) and response regulator (*agrA*) (110). The AgrD peptide is processed and secreted in the form of a thiolactone ring (AIP) by AgrB. The AIP binds to and induces phosphorylation of AgrC. Finally, the phosphorylated AgrA binds to the two promoters, P2 and P3, and induces the expression of downstream of genes. The *agrBDCA* is under the control of P2 promoter and consequently is autoactivated by AgrD itself. This results in an exponential increase of AIP level, ensuring that the whole population is changed from the low cell density to high cell density. The other promoter, P3, drives transcription of a regulatory RNA termed RNAIII. It is known that the expressed RNAIII represses genes encoding cell adhesion factors and induces genes encoding virulence factors.

Role of the agr QS system in the biofilm formation of staphylococci has been demonstrated in many studies. Two mechanisms for biofilm formation are known to be under the control of *agr* QS system: One mechanism is dependent on the extracellular polysaccharide PIA, and the other one is presumably dependent on adhesive proteins rather than PIA (110, 115). In both *S. epidermidis* and *S. aureus*, defective mutants showed stronger adherence to polystyrene

and produce more biofilms than *agr* positive strains, suggesting that the Agr system down-regulates the biofilm formation (116). This finding is strongly supported by other studies. During biofilm formation, Yarwood et al. (2004) found the local activation of Agr system in the biofilm of *S. aureus* and subsequently detachment of *agr*-expressing organisms from the biofilm (117). Boles and Horswill (2008) demonstrated that the expression of *agr* genes in *S. aureus* causes dispersion of an established biofilm through the increased production of proteases that have negative effects on biofilm maturation (118). In addition to the biofilm formation and virulence factors, several hundred genes are also controlled by the RNAIII that is activated by AgrA (119). Those genes induced by the RNAIII belong to other transcriptional regulation targets as well as cell metabolism, including nucleic acid metabolism, amino acid metabolism, carbon metabolism, etc. Thus, the *agr* QS system is involved in a complex regulatory network in staphylococci.

1.6 Summary and major goal for thesis research

Mammalian genomes lack the enzymes required to deconstruct the structural polysaccharides, cellulose and hemicellulose, present in plant cell wall. Herbivorous animals rely on the symbiotic microbes in the rumen, caecum or colon to break down cellulose and hemicellulose in their dietary fiber. The resulting productions of short-chain fatty acids and microbial proteins from fibrolytic microbial fermentation play a pivotal role in host nutrition. In contrast, humans rely less on microbial fermentation in large intestine for the acquisition of energy, but their fermentation products from resident microbiota, such as butyrate, play a significant role in host physiology related to health. Thus, for better understanding host-microbe

interaction, it is important and relevant to establish the model for fiber degradation and metabolism of fibrolytic gut microbes.

In the mammalian gut, the dominant fibrolytic bacteria belong to the Firmicutes and Bacteroidetes phyla. Many studies have revealed that the polysaccharide utilization mechanism of Bacteroidetes, including human gut *Bacteroides* and rumen *Prevotella* genera. Through deep genomic, transcriptomic, and biochemical analyses, PUL system has been established as a fibrolytic model paradigm for the mammalian gut Bacteroidetes. However, the fibrolytic Firmicutes have received far less attention than Bacteroidetes. The well-known mechanism for fibrolytic system of Firmicutes is the cellulosomal system. The cellulosomal mechanism has been established through intensive studies on fibrolytic soil bacteria, *Clostridium* species. Recent molecular evidence suggests that some Firmicutes in the mammalian gut, including *Ruminococcus flavefaciens* and *Ruminococcus champanellensis*, also possesses the cellulosomal system. Based on these findings, the cellulosome paradigm has been expanded to the fiber degradation model for rumen and human gut species from soil species and is considered as a representative fibrolytic model for mammalian gut Firmicutes.

It has been proposed that other predominant fibrolytic species, *Ruminococcus albus*, has alternative mechanisms for plant cell wall degradation rather than the cellulosome, including a Pil-like protein, a glycocalyx, and CBM37. As less research have conducted relative to the cellulosomal species, *R. flavefaciens*, it is still insufficient to define the fibrolytic system of *R. albus* and this limits better understanding of their roles in gut ecosystem.

Another role of rumen microbes in host nutrition is a supply of vitamins to the host. Mammals rely on dietary vitamins because they cannot synthesize B vitamins, including folic acid that is an essential growth factor for all bacteria, archaea, and eukaryotes. However,

ruminants in which the rumen is fully functioning are independent of a dietary supply of folate due to biosynthesis by rumen bacteria. Decades ago, the vitamin metabolisms in rumen bacteria were studied by several pioneers of rumen microbiology. Although variation among strains exists, the predominant cellulolytic ruminococci including *R. albus* and *R. flavefaciens* are known to be capable of synthesizing folic acid and consequently contribute to host nutrition for vitamin. However, due to technological limits, their researches were restricted to phenotypic analysis and lacked molecular level information underpinning folate metabolism of rumen microbes.

To date, the many genomes of *Ruminococcus* species have been sequenced and new molecular techniques have developed, which enable us to investigate more fundamental mechanisms of ruminococci for fiber degradation and vitamin utilization at the genomic and transcriptomic level. Thus, the ultimate goal for the current research is to provide deep insight into the mechanisms of the predominant rumen species, *Ruminococcus albus*, regarding their nutritional contribution to host, such as energy and vitamin. Using the distinct strains of *R. albus* for fiber and folic acid utilization reported in previous studies, this research investigates how each *R. albus* strains degrade fiber and utilize folic acid through comparative genomic and transcriptomic analyses.

1.7 References

1. **Van Soest PJ**. 1994. Nutritional ecology of the ruminant, 2nd ed. Cornell University Press, Ithaca.
2. **McNeil NI**. 1984. The contribution of the large intestine to energy supplies in man. Am J Clin Nutr **39**:338–342.
3. **Nicholson J**, **Holmes E**, **Kinross J**, **Burcelin R**, **Gibson G**, **Jia W**, **Petttersen S**. 2012. Host-gut microbiota metabolic interactions **1262**.

4. **Thomas L V**, **Ockhuizen T**, **Suzuki K**. 2014. Exploring the influence of the gut microbiota and probiotics on health: a symposium report. Br J Nutr **112 Suppl** :S1–S18.

5. **De Filippo C**, **Cavalieri D**, **Di Paola M**, **Ramazzotti M**, **Poullet JB**, **Massart S**, **Collini S**, **Pieraccini G**, **Lionetti P**. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A **107**:14691–14696.

6. **Walker AW**, **Ince J**, **Duncan SH**, **Webster LM**, **Holtrop G**, **Ze X**, **Brown D**, **Stares MD**, **Scott P**, **Bergerat A**, **Louis P**, **McIntosh F**, **Johnstone AM**, **Lobley GE**, **Parkhill J**, **Flint HJ**. 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. ISME J **5**:220–230.

7. **Dehority BA**, **Scott HW**. 1967. Extent of Cellulose and Hemicellulose Digestion in Various Forages by Pure Cultures of Rumen Bacteria1. J Dairy Sci **50**:1136–1141.

8. **Jane Morris E**, **Van Gylswyk NO**. 1980. Comparison of the action of rumen bacteria on cell walls from Eragrostis tef. J Agric Sci **95**:313–323.

9. **Dehority BA**. 1991. Effects of microbial synergism on fibre digestion in the rumen. Proc Nutr Soc **50**:149–159.

10. **Flint HJ**, **Bayer EA**, **Rincon MT**, **Lamed R**, **White BA**. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol **6**:121–131.

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

12. **Matte A**, **Forsberg CW**, **Verrinder Gibbins AM**. 1992. Enzymes associated with metabolism of xylose and other pentoses by *Prevotella* (*Bacteroides*) *ruminicola* strains, *Selenomonas ruminantium* D, and *Fibrobacter succinogenes* S85. Can J Microbiol **38**:370–376.

13. **Kobayashi Y**, **Shinkai T**, **Koike S**. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - Review. Folia Microbiol (Praha) **53**:195–200.

14. **Suen G**, **Weimer PJ**, **Stevenson DM**, **Aylward FO**, **Boyum J**, **Deneke J**, **Drinkwater C**, **Ivanova NN**, **Mikhailova N**, **Chertkov O**, **Goodwin LA**, **Currie CR**, **Mead D**, **Brumm PJ**. 2011. The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS One **6**:e18814.

15. **Iannotti EL**, **Kafkewitz D**, **Wolin MJ**, **Bryant MP**. 1973. Glucose fermentation products of *Ruminococcus albus* grown in continuous culture with *Vibrio succinogenes*: changes caused by interspecies transfer of H2. J Bacteriol **114**:1231–1240.

16. **Bryant MP**, **Campbell LL**, **Reddy CA**, **Crabill MR**. 1977. Growth of *Desulfovibrio* in lactate or ethanol media low in sulfate in association with H2 utilizing methanogenic bacteria. Appl Environ Microbiol **33**:1162–1169.

17. **Belenguer A**, **Duncan SH**, **Calder AG**, **Holtrop G**, **Louis P**, **Lobley GE**, **Flint HJ**. 2006. Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. Appl Environ Microbiol **72**:3593–3599.

18. **Flint HJ**, **Duncan SH**, **Scott KP**, **Louis P**. 2007. Interactions and competition within the microbial community of the human colon: Links between diet and health: Minireview. Environ Microbiol **9**:1101–1111.

19. **Krause DO**, **Denman SE**, **Mackie RI**, **Morrison M**, **Rae AL**, **Attwood GT**, **McSweeney CS**. 2003. Opportunities to improve fiber degradation in the rumen: Microbiology, ecology, and genomics. FEMS Microbiol Rev **27**:663–693.

20. **Dehority BA**. 2003. Rumen microbiolgy. Nottingham University Press, Nottingham.

21. **Cheng K-J**, **Stewart CS**, **Dinsdale D**, **Costerton JW**. 1984. Electron microscopy of bacteria involved in the digestion of plant cell walls. Anim Feed Sci Technol **10**:93–120.

22. **Miron J**, **Ben-Ghedalia D**, **Morrison M**. 2001. Invited review: adhesion mechanisms of rumen cellulolytic bacteria. J Dairy Sci **84**:1294–309.

23. **Tajima K**, **Aminov RI**, **Nagamine T**, **Ogata K**, **Nakamura M**, **Matsui H**, **Benno Y**. 1999. Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. FEMS Microbiol Ecol **29**:159–169.

24. **Larue R**, **Yu Z**, **Parisi VA**, **Egan AR**, **Morrison M**. 2005. Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. Environ Microbiol **7**:530–543.

25. **McWilliam Leitch EC**, **Walker AW**, **Duncan SH**, **Holtrop G**, **Flint HJ**. 2007. Selective colonization of insoluble substrates by human faecal bacteria. Environ Microbiol **9**:667–679.

26. **Walker AW**, **Duncan SH**, **Harmsen HJM**, **Holtrop G**, **Welling GW**, **Flint HJ**. 2008. The species composition of the human intestinal microbiota differs between particle-associated and liquid phase communities. Environ Microbiol **10**:3275–3283.

27. **Chassard C**, **Delmas E**, **Robert C**, **Lawson PA**, **Bernalier-Donadille A**. 2011. *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from human gut microbiota. Int J Syst Evol Microbiol **62**:138–143.

28. **Ben David Y**, **Dassa B**, **Borovok I**, **Lamed R**, **Koropatkin NM**, **Martens EC**, **White BA**, **Bernalier-Donadille A**, **Duncan SH**, **Flint HJ**, **Bayer EA**, **Moraïs S**. 2015. Ruminococcal cellulosome systems from rumen to human. Environ Microbiol.

29. **Kormelink FJM**, **Voragen AGJ**. 1993. Degradation of different [(glucurono)arabino]xylans by a combination of purified xylan-degrading enzymes. Appl Microbiol Biotechnol **38**:688–695.

30. **Dodd D**, **Cann IKO**. 2009. Enzymatic deconstruction of xylan for biofuel production. Glob Change Biol Bioenergy **1**:2–17.

31. **Chesson A**, **Gordon AH**, **Lomax JA**. 1985. Methylation analysis of mesophyll, epidermis, and fibre cell-walls isolated from the leaves of perennial and italian ryegrass. Carbohydr Res.

32. **Himmel ME**, **Xu Q**, **Luo Y**, **Ding S-Y**, **Lamed R**, **Bayer EA**. 2010. Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels **1**:323–341.

33. **Bayer EA**, **Shoham Y**, **Lamed R**. 2013. Lignocellulose-Decomposing Bacteria and Their Enzyme Systems, p. 215–266. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes. Springer Berlin Heidelberg, Berlin.

34. **Bayer EA**, **Shimon LJ**, **Shoham Y**, **Lamed R**. 1998. Cellulosomes-structure and ultrastructure. J Struct Biol **124**:221–34.

35. **Henrissat B**, **Davies G**. 1997. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol **7**:637–644.

36. **Jalak J**, **Kurašin M**, **Teugjas H**, **Väljamä P**. 2012. Endo-exo synergism in cellulose hydrolysis revisited. J Biol Chem **287**:28802–28815.

37. **Wilson DB**, **Kostylev M**. 2012. Cellulase processivity. Methods Mol Biol **908**:93–99.

38. **White BA**, **Lamed R**, **Bayer EA**, **Flint HJ**. 2014. Biomass Utilization by Gut Microbiomes. Annu Rev Microbiol 279–296.

39. **Irwin DC**, **Spezio M**, **Walker LP**, **Wilson DB**. 1993. Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. Biotechnol Bioeng **42**:1002–1013.

40. **Boraston AB**, **Bolam DN**, **Gilbert HJ**, **Davies GJ**. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J **382**:769–781.

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

42. **Ezer A**, **Matalon E**, **Jindou S**, **Borovok I**, **Atamna N**, **Yu Z**, **Morrison M**, **Bayer EA**, **Lamed R**. 2008. Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. J Bacteriol **190**:8220–8222.

43. **Montanier C**, **Money VA**, **Pires VMR**, **Flint JE**, **Pinheiro BA**, **Goyal A**, **Prates JAM**, **Izumi A**, **Stålbrand H**, **Morland C**, **Cartmell A**, **Kolenova K**, **Topakas E**, **Dodson EJ**, **Bolam DN**, **Davies GJ**, **Fontes CMGA**, **Gilbert HJ**. 2009. The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. PLoS Biol **7**:e71.

44. **Krause DO**, **Bunch RJ**, **Smith WJM**, **McSweeney CS**. 1999. Diversity of *Ruminococcus* strains: A survey of genetic polymorphisms and plant digestibility. J Appl Microbiol **86**:487–495.

45. **Doi RH**, **Kosugi A**. 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol **2**:541–551.

46. **Berg Miller ME**, **Antonopoulos DA**, **Rincon MT**, **Band M**, **Bari A**, **Akraiko T**, **Hernandez A**, **Thimmapuram J**, **Henrissat B**, **Coutinho PM**, **Borovok I**, **Jindou S**, **Lamed R**, **Flint HJ**, **Bayer EA**, **White BA**. 2009. Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. PLoS One **4**:e6650.

47. **Bayer EA**, **Belaich J-P**, **Shoham Y**, **Lamed R**. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol **58**:521–554.

48. **Hungate RE**. 1957. Microorganisms in the rumen of cattle fed a constant ration. Can J Microbiol **3**:289–311.

49. **Russell JB**, **Muck RE**, **Weimer PJ**. 2009. Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. FEMS Microbiol Ecol **67**:183–97.

50. **Pegden RS**, **Larson MA**, **Grant RJ**, **Morrison M**. 1998. Adherence of the gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. J Bacteriol **180**:5921–5927.

51. **Morrison M**, **Miron J**. 2000. Adhesion to cellulose by *Ruminococcus albus*: A combination of cellulosomes and Pil-proteins? FEMS Microbiol Lett **185**:109–115.

52. **Xu Q**, **Morrison M**, **Nelson KE**, **Bayer EA**, **Atamna N**, **Lamed R**. 2004. A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. FEBS Lett **566**:11–6.

53. **Weimer PJ**, **Price NPJ**, **Kroukamp O**, **Joubert LM**, **Wolfaardt GM**, **Van Zyl WH**. 2006. Studies of the extracellular glycocalyx of the anaerobic cellulolytic bacterium *Ruminococcus albus* 7. Appl Environ Microbiol **72**:7559–7566.

54. **Rakotoarivonina H**, **Jubelin G**, **Hebraud M**, **Gaillard-Martinie B**, **Forano E**, **Mosoni P**. 2002. Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. Microbiology **148**:1871–80.

55. **Vodovnik M**, **Duncan SH**, **Reid MD**, **Cantlay L**, **Turner K**, **Parkhill J**, **Lamed R**, **Yeoman CJ**, **Miller MEB**, **White BA**, **Bayer EA**, **Marinšek-Logar R**, **Flint HJ**. 2013. Expression of Cellulosome Components and Type IV Pili within the Extracellular Proteome of *Ruminococcus flavefaciens* 007. PLoS One **8**:1–11.

56. **Christopherson MR**, **Dawson J a**, **Stevenson DM**, **Cunningham AC**, **Bramhacharya S**, **Weimer PJ**, **Kendziorski C**, **Suen G**. 2014. Unique aspects of fiber degradation by the ruminal ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. BMC Genomics **15**:1066.

57. **Devillard E**, **Goodheart DB**, **Karnati SKR**, **Bayer EA**, **Lamed R**, **Miron J**, **Nelson KE**, **Morrison M**. 2004. *Ruminococcus albus* 8 Mutants Defective in Cellulose Degradation Are Deficient in Two Processive Endocellulases, Cel48A and Cel9B, Both of Which Possess a Novel Modular Architecture. J Bacteriol **186**:136–145.

58. **Odenyo AA**, **Mackie RI**, **Stahl DA**, **White BA**. 1994. The Use of 16S rRNA-targeted olionucleotide probes to study competition between ruminal fibrolytic Bacteria: development of probes for *Ruminococcus* species and evidence for bacteriocin production. Appl Environ Microbiol **60**:3688–3696.

59. **Chen J**, **Weimer PJ**. 2001. Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. Microbiology **147**:21–30.

60. **Chen J**, **Stevenson DM**, **Weimer PJ**. 2004. Albusin B, a bacteriocin from the ruminal bacterium *Ruminococcus albus* 7 that inhibits growth of *Ruminococcus flavefaciens*. Appl Environ Microbiol **70**:3167–3170.

61. **Eckburg PB**, **Bik EM**, **Bernstein CN**, **Purdom E**, **Dethlefsen L**, **Sargent M**, **Gill SR**, **Nelson KE**, **Relman DA**. 2005. Diversity of the human intestinal microbial flora. Science **308**:1635–1638.

62. **Martínez I**, **Kim J**, **Duffy PR**, **Schlegel VL**, **Walter J**. 2010. Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. PLoS One **5**:e15046.

63. **Cann I**, **Bernardi RC**, **Mackie RI**. 2016. Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm. Environ Microbiol **18**:307–10.

64. **Osborne JM**, **Dehority BA**. 1989. Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by three pure cultures of ruminal bacteria. Appl Environ Microbiol **55**:2247–2250.

65. **BRYANT MP**, **SMALL N**, **BOUMA C**, **CHU H**. 1958. *Bacteroides ruminicola* n. sp. and *Succinimonas amylolytica*; the new genus and species; species of succinic acid-producing anaerobic bacteria of the bovine rumen. J Bacteriol **76**:15–23.

66. **Miyazaki K**, **Martin JC**, **Marinsek-Logar R**, **Flint HJ**. 1997. Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp. brevis) B(1)4. Anaerobe **3**:373–381.

67. **Dodd D**, **Moon YH**, **Swaminathan K**, **Mackie RI**, **Cann IKO**. 2010. Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic bacteroidetes. J Biol Chem **285**:30261–30273.

68. **Martens EC**, **Lowe EC**, **Chiang H**, **Pudlo NA**, **Wu M**, **McNulty NP**, **Abbott DW**, **Henrissat B**, **Gilbert HJ**, **Bolam DN**, **Gordon JI**. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol **9**.

69. **Naas AE**, **Mackenzie AK**, **Mravec J**, **Schückel J**, **Willats WGT**, **Eijsink VGH**, **Pope PB**. 2014. Do rumen Bacteroidetes utilize an alternative mechanism for cellulose degradation? MBio **5**:e01401–14.

70. **Arumugam M**, **Raes J**, **Pelletier E**, **Le Paslier D**, **Yamada T**, **Mende DR**, **Fernandes GR**, **Tap J**, **Bruls T**, **Batto J-M**, **Bertalan M**, **Borruel N**, **Casellas F**, **Fernandez L**, **Gautier L**, **Hansen T**, **Hattori M**, **Hayashi T**, **Kleerebezem M**, **Kurokawa K**, **Leclerc M**, **Levenez F**, **Manichanh C**, **Nielsen HB**, **Nielsen T**, **Pons N**, **Poulain J**, **Qin J**, **Sicheritz-Ponten T**, **Tims S**, **Torrents D**, **Ugarte E**, **Zoetendal EG**, **Wang J**, **Guarner F**, **Pedersen O**, **de Vos WM**, **Brunak S**, **Doré J**, **Antolín M**, **Artiguenave F**, **Blottiere HM**, **Almeida M**, **Brechot C**, **Cara C**, **Chervaux C**, **Cultrone A**, **Delorme C**, **Denariaz G**, **Dervyn R**, **Foerstner KU**, **Friss C**, **van de Guchte M**, **Guedon E**, **Haimet F**, **Huber W**, **van Hylckama-Vlieg J**, **Jamet A**, **Juste C**, **Kaci G**, **Knol J**, **Lakhdari O**, **Layec S**, **Le Roux K**, **Maguin E**, **Mérieux A**, **Melo Minardi R**, **M'rini C**, **Muller J**, **Oozeer R**, **Parkhill J**, **Renault P**, **Rescigno M**, **Sanchez N**, **Sunagawa S**, **Torrejon A**, **Turner K**, **Vandemeulebrouck G**, **Varela E**, **Winogradsky Y**, **Zeller G**, **Weissenbach J**, **Ehrlich SD**, **Bork P**. 2011. Enterotypes of the human gut microbiome. Nature **473**:174–180.

71. **Wu GD**, **Chen J**, **Hoffmann C**, **Bittinger K**, **Chen Y-Y**, **Keilbaugh SA**, **Bewtra M**, **Knights D**, **Walters WA**, **Knight R**, **Sinha R**, **Gilroy E**, **Gupta K**, **Baldassano R**, **Nessel L**, **Li H**, **Bushman FD**, **Lewis JD**. 2011. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. Science (80-).

72. **Salyers AA**, **Vercellotti JR**, **West SEH**, **Wilkins TD**. 1977. Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. Appl Environ Microbiol **33**:319–322.

73. **Sonnenburg JL**, **Xu J**, **Leip DD**, **Chen C-H**, **Westover BP**, **Weatherford J**, **Buhler JD**, **Gordon JI**. 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science **307**:1955–1959.

74. **Martens EC**, **Koropatkin NM**, **Smith TJ**, **Gordon JI**. 2009. Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J Biol Chem **284**:24673–24677.

75. **Dodd D**, **Mackie RI**, **Cann IKO**. 2011. Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol Microbiol **79**:292–304.

76. **Koropatkin NM**, **Cameron EA**, **Martens EC**. 2012. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol **10**:323–335.

77. **Zhang M**, **Chekan JR**, **Dodd D**, **Hong P-Y**, **Radlinski L**, **Revindran V**, **Nair SK**, **Mackie RI**, **Cann I**. 2014. Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc Natl Acad Sci U S A **111**:E3708–17.

78. **Hungate RE**. 1947. Studies on Cellulose Fermentation: III. The Culture and Isolation for Cellulose-decomposing Bacteria from the Rumen of Cattle. J Bacteriol **53**:631–645.

79. **Ransom-Jones E**, **Jones DL**, **McCarthy AJ**, **McDonald JE**. 2012. The Fibrobacteres: An Important Phylum of Cellulose-Degrading Bacteria. Microb Ecol **63**:267–281.

80. **Malburg LM**, **Iyo AH**, **Forsberg CW**. 1996. A novel family 9 endoglucanase gene (celD), whose product cleaves substrates mainly to glucose, and its adjacent upstream homolog (celE) from *Fibrobacter succinogenes* S85. Appl Environ Microbiol **62**:898–906.

81. **Qi M**, **Jun HS**, **Forsberg CW**. 2007. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. Appl Environ Microbiol **73**:6098–6105.

82. **Peekhaus N**, **Conway T**. 1998. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. J Bacteriol.

83. **Sweeney NJ**, **Laux DC**, **Cohen PS**. 1996. *Escherichia coli* F-18 and *E. coli* K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. Infect Immun **64**:3504–3511.

84. **Barton LL**. 2005. Structural and Functional Relationships in Prokaryotes 529–601.

85. **Flamholz A**, **Noor E**, **Bar-Even A**, **Liebermeister W**, **Milo R**. 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci U S A **110**:10039–10044.

86. **Kim BH**, **Gadd GM**. 2008. Bacterial Physiology and Metabolism. Cambridge University Press, Cambridge.

87. **Servinsky MD**, **Germane KL**, **Liu S**, **Kiel JT**, **Clark AM**, **Shankar J**, **Sund CJ**. 2012. Arabinose is metabolized via a phosphoketolase pathway in *Clostridium acetobutylicum* ATCC 824. J Ind Microbiol Biotechnol **39**:1859–1867.

88. **Liu L**, **Zhang L**, **Tang W**, **Gu Y**, **Hua Q**, **Yang S**, **Jiang W**, **Yang C**. 2012. Phosphoketolase pathway for Xylose Catabolism in *Clostridium acetobutylicum* revealed by 13C metabolic flux analysis. J Bacteriol **194**:5413–5422.

89. **Aristilde L**, **Lewis IA**, **Park JO**, **Rabinowitz JD**. 2015. Hierarchy in Pentose Sugar Metabolism in *Clostridium acetobutylicum*. Appl Environ Microbiol **81**:1452–1462.

90. **Sund CJ**, **Liu S**, **Germane KL**, **Servinsky MD**, **Gerlach ES**, **Hurley MM**. 2015. Phosphoketolase flux in *Clostridium acetobutylicum* during growth on L-arabinose. Microbiology **161**:430–40.

91. **Servinsky MD**, **Kiel JT**, **Dupuy NF**, **Sund CJ**. 2010. Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. Microbiology **156**:3478–3491.

92. **Fattal-Valevski A**. 2011. Thiamine (Vitamin B1). J Evid Based Complementary Altern Med **16**:12–20.

93. **Burgess CM**, **Smid EJ**, **van Sinderen D**. 2009. Bacterial vitamin B2, B11 and B12 overproduction: An overview. Int J Food Microbiol **133**:1–7.

94. **Henderson LM**. 1983. Niacin. Annu Rev Nutr **3**:289–307.

95. **Robishaw JD**, **Neely JR**. 1985. Coenzyme A metabolism. Am J Physiol **248**:E1–E9.

96. **Axelrod AE**, **Martin CJ**. 1961. Water-Soluble Vitamins, Part I (Ascorbic Acid, Biotin, Inositol, Nicotinic Acid, Pyridoxin Group). Annu Rev Biochem.

97. **McMahon RJ**. 2002. Biotin in metabolism and molecular biology. Annu Rev Nutr **22**:221–239.

98. **Roth JR**, **Lawrence JG**, **Bobik TA**. 1996. Cobalamin (coenzyme B12): synthesis and biological significance. Annu Rev Microbiol **50**:137–181.

99. **de Crécy-Lagard V**, **El Yacoubi B**, **de la Garza RD**, **Noiriel A**, **Hanson AD**. 2007. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genomics **8**.

100. **McDowell LR**. 2000. Folacin, p. 480–521. *In* Vitamins in animal and human nutritionSecond. Iowa State University Press, Ames.

101. **Scott HW**, **Dehority BA**. 1965. Vitamin requirements of seveal cellulolytic rumen bacteria. J Bacteriol **89**:1169–75.

102. **Bryant MP**, **Robinson IM**. 1961. Some Nutritional Requirements of the Genus *Ruminococcus*. Appl Microbiol **9**:91–95.

103. **Green JBPN**, **Matthews RG**. 1996. Folate biosynthesis, reduction, and polyglutamylation, p. 665–673. *In* Neidhardt, FC (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular BiologySecond. ASM Press, Washington, DC.

104. **Green JM**, **Nichols BP**. 1991. p-Aminobenzoate biosynthesis in *Escherichia coli*: Purification of aminodeoxychorismate lyase and cloning of *pabC*. J Biol Chem **266**:12971–12975.

105. **Parsons JF**, **Jensen PY**, **Pachikara AS**, **Howard AJ**, **Eisenstein E**, **Ladner JE**. 2002. Structure of *Escherichia coli* aminodeoxychorismate synthase: Architectural conservation and diversity in chorismate-utilizing enzymes. Biochemistry **41**:2198–2208.

106. **Hyde JE**. 2005. Exploring the folate pathway in *Plasmodium falciparum*. Acta Trop **94**:191–206.

107. **Nare B**, **Hardy LW**, **Beverley SM**. 1997. The roles of pteridine reductase 1 and dihydrofolate reductase- thymidylate synthase in pteridine metabolism in the protozoan parasite *Leishmania major*. J Biol Chem **272**:13883–13891.

108. **Orsomando G**, **Bozzo GG**, **De La Garza RD**, **Basset GJ**, **Quinlivan EP**, **Naponelli V**, **Rébeillé F**, **Ravanel S**, **Gregory JF**, **Hanson AD**. 2006. Evidence for folate-salvage reactions in plants. Plant J **46**:426–435.

109. **Waters CM**, **Bassler BL**. 2005. Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol **21**:319–346.

110. **Novick RP**, **Geisinger E**. 2008. Quorum sensing in staphylococci. Annu Rev Genet **42**:541–564.

111. **Novick RP**, **Projan SJ**, **Kornblum J**, **Ross HF**, **Ji G**, **Kreiswirth B**, **Vandenesch F**, **Moghazeh S**, **Novick RP**. 1995. The agr P2 operon: An autocatalytic sensory transduction system in *Staphylococcus aureus*. MGG Mol Gen Genet **248**:446–458.

112. **Nealson KH**, **Hastings JW**. 1979. Bacterial bioluminescence: its control and ecological significance. Microbiol Rev **43**:496–518.

113. **Visick KL**, **Foster J**, **Doino J**, **McFall-Ngai M**, **Ruby EG**. 2000. *Vibrio fischeri* lux genes play an important role in colonization and development of the host light organ. J Bacteriol **182**:4578–4586.

114. **Tenover FC**, **Gaynes RP**. 2000. The epidemiology of *Staphylococcus aureus* infections., p. 414–421. *In* Fischetti, VA, Novick, RP, Ferretti, JJ, Portnoy, DA, Rood, JI (eds.), Gram-Positive Pathogens. ASM Press, Washington, DC.

115. **O'Neill E**, **Pozzi C**, **Houston P**, **Smyth D**, **Humphreys H**, **Robinson DA**, **O'Gara JP**. 2007. Association between methicillin susceptibility and biofilm regulation in *Staphylococcus aureus* isolates from device-related infections. J Clin Microbiol **45**:1379–1388.

116. **Vuong C**, **Saenz HL**, **Götz F**, **Otto M**. 2000. Impact of the agr quorum-sensing system on adherence to polystyrene in *Staphylococcus aureus*. J Infect Dis **182**:1688–1693.

117. **Yarwood JM**, **Bartels DJ**, **Volper EM**, **Greenberg EP**. 2004. Quorum Sensing in *Staphylococcus aureus* Biofilms. J Bacteriol **186**:1838–1850.

118. **Boles BR**, **Horswill AR**. 2008. agr-mediated dispersal of *Staphylococcus aureus* biofilms. PLoS Pathog **4**.

119. **Dunman P**, **Murphy E**, **Haney S**. 2001. Transcription Profiling-Based Identification of *Staphylococcus aureus* Genes Regulated by the agr and/or sarA Loci. J Bacteriol **183**:7341–7353.

1.8 Figures

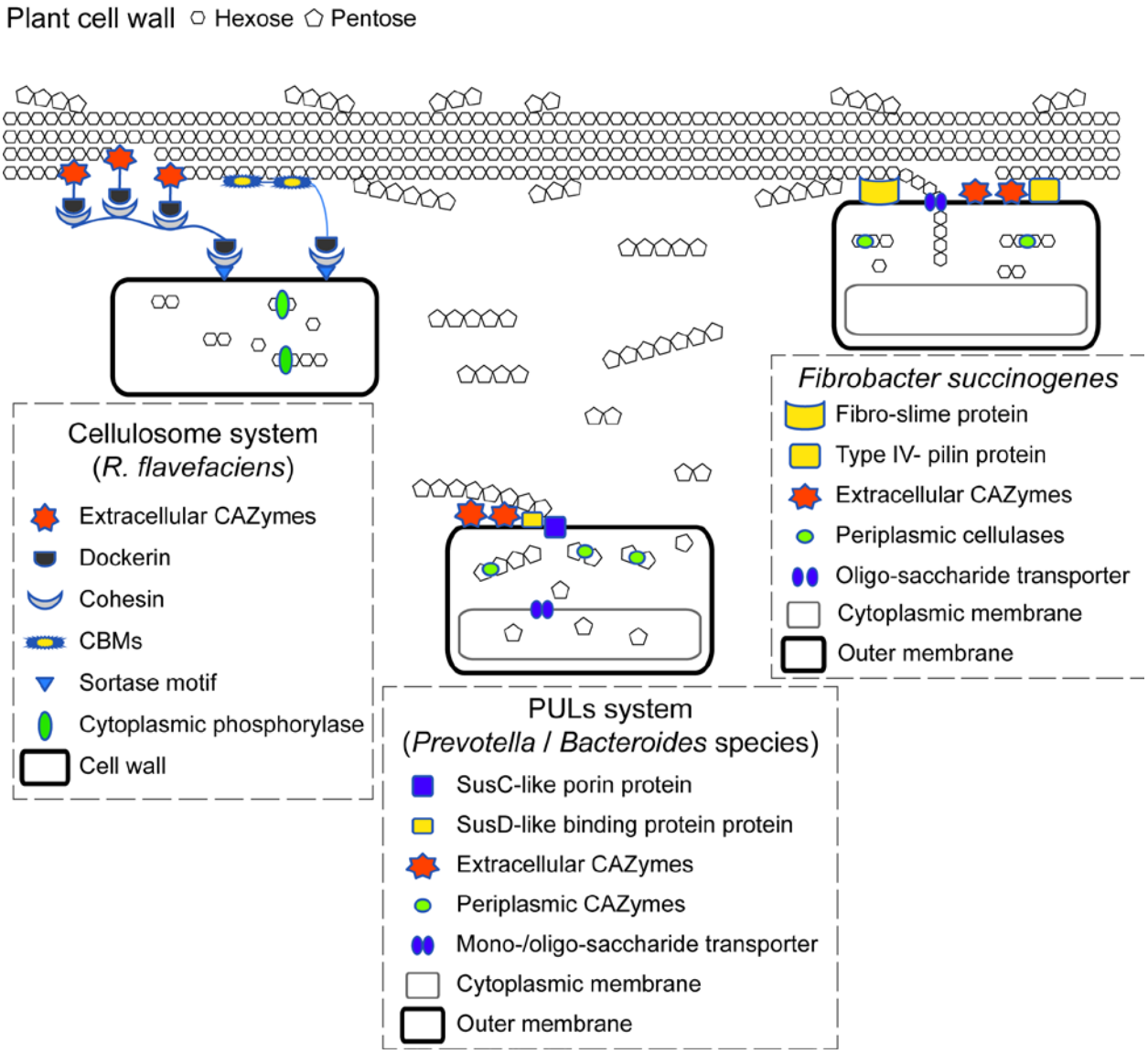


Fig. 1.1. Polysaccharide utilization systems in the rumen characterized in previous studies.

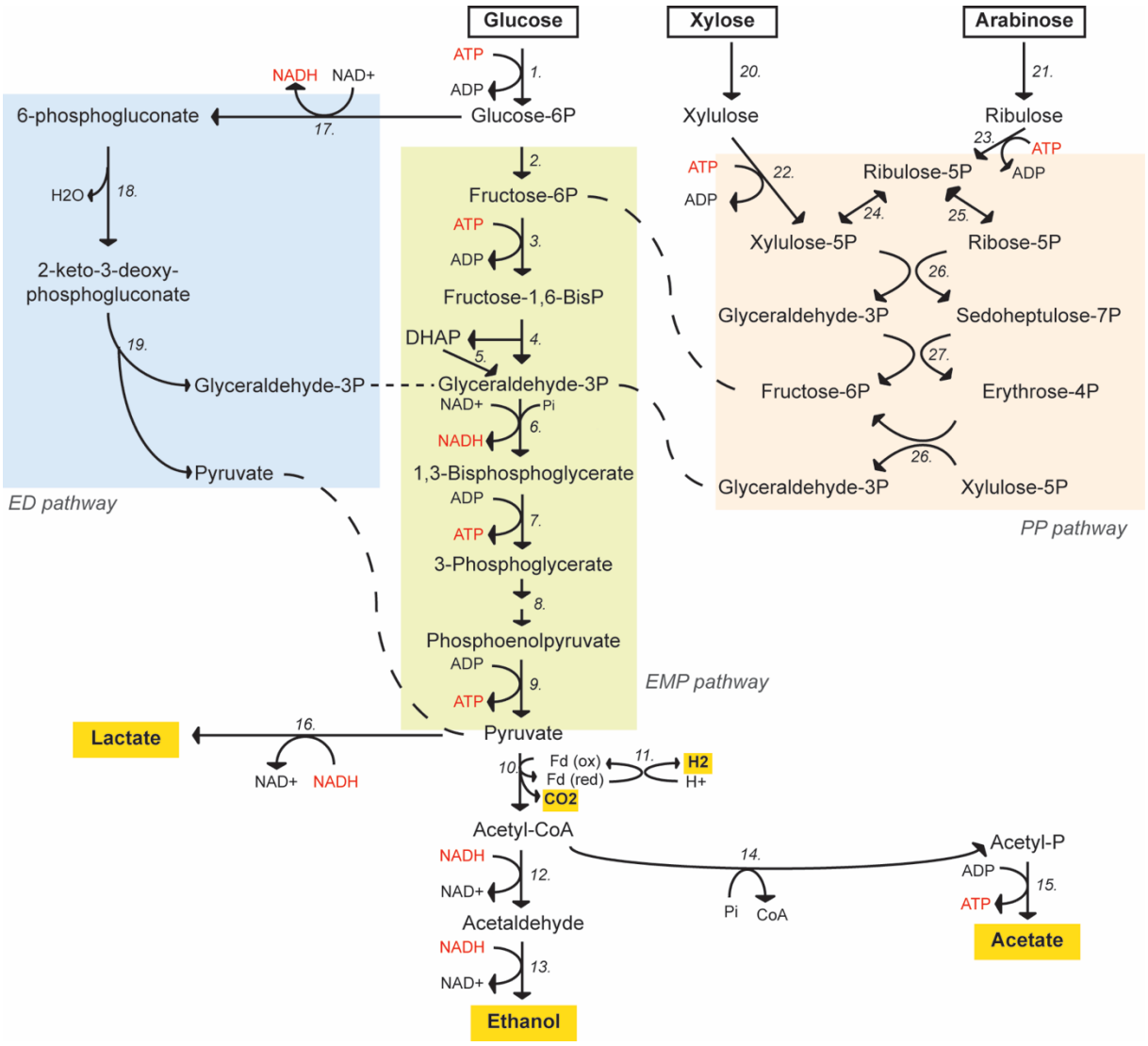


Fig. 1.2. Glycolytic pathway through the EMP, ED, and PP pathways and sugar fermentation pathways in anaerobic bacteria. Enzymes: 1. Glucokinase; 2. Glucose-6-phosphate isomerase; 3. Phosphofructokinase; 4. Fructose-bisphosphate aldolase; 5. Triosephosphate isomerase; 6. Glyceraldehyde3-phosphate dehydrogenase; 7. Phosphoglycerate kinase; 8. Phosphoglycerate mutase and enolase; 9. Pyruvate kinase; 10. Pyruvate ferredoxin oxidoreductase; 11. Ferredoxin dependent hydrogenase; 12. Acetaldehyde dehydrogenase; 13. Alcohol dehydrogenase; 14. Phosphotransacetylase; 15. Acetate kinase; 16. Lactate dehydrogenase; 17. Glucose-6-phosphate dehydrogenase; 18. 6-phosphogluconate dehydratase; 19. 2-keto-3-deoxy-phosphogluconate aldolase; 20. Xylose isomerase; 21. Arabinose isomerase; 22. Xylulose kinase; 23. Ribulose kinase; 24. Epimerase; 25. Isomerase; 26. Transketolase; 27. Transaldolase.

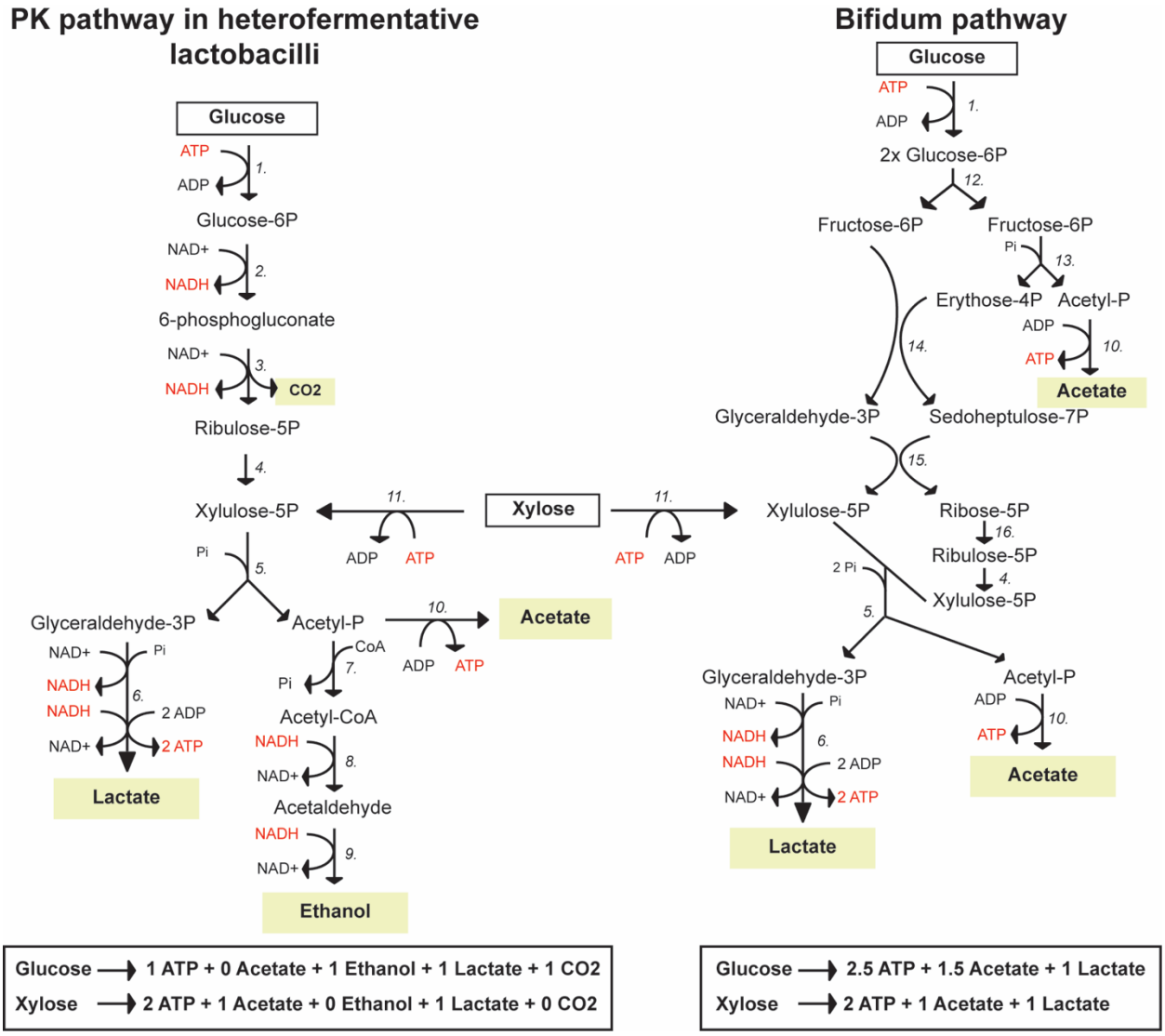


Fig. 1.3. The phosphoketolase pathway in heterofermentative *lactobacillus* and *bifidobacillus* species. 1.glucokinase; 2. glucose-6P dehydrogenase; 3. 6-phosphogluconate dehydrogenase; 4. ribulose-5P-3-epimerase; 5. xylulose-5P phosphoketolase; 6. enzymes of EMP pathway and lactate dehydrogenase; 7. phosphotransacetylase; 8. acetaldehyde dehydrogenase; 9. alcohol dehydrogenase; 10. acetate kinase; 11. xylose isomerase and xylulose kinase; 12. Glucose-6P isomerase; 13. F-6P phosphoketolase; 14. transaldolase; 15. transketolase; 16. ribose-5P isomerase.

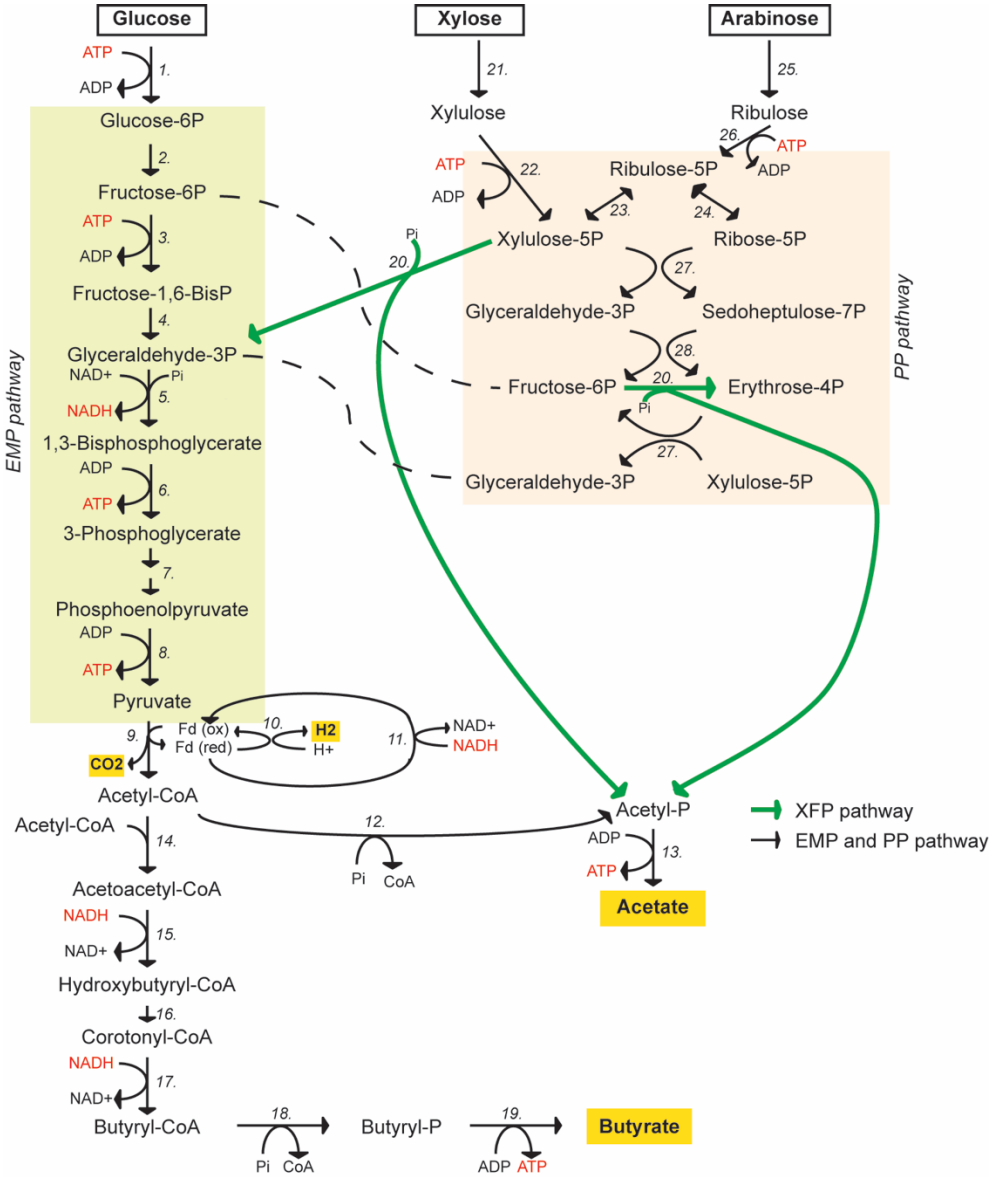


Fig. 1.4. The GMP, PP and XFP pathway in *C. acetobutylicum* during acidogenic growth. Arrows for phosphoketolase reactions are green. Enzymes: 1. Glucokinase; 2. Glucose-6P isomerase; 3. Phosphofructokinase; 4. Fructose-bisphosphate aldolase and triosephosphate isomerase; 5. Glyceraldehyde-3P dehydrogenase; 6. Phosphoglycerate kinase; 7. Phosphoglycerate mutase and enolase; 8. Pyruvate kinase; 9. Pyruvate ferredoxin oxidoreductase; 10. Ferredoxin dependent hydrogenase; 11. NAD+/NADH oxidoreductase; 12. Phosphotransacetylase; 13. Acetate kinase; 14. Thiolase; 15. Hydroxybutyryl-CoA dehydrogenase; 16. Crotonase; 17. Butyryl-CoA dehydrogenase; 18. Phosphotransbutyrylase; 19. Butyrate kinase; 20. Xylulose-5P/fructose-6P phosphoketolase; 21. Xylose isomerase; 22. Xylulose kinase; 23. Epimerase; 24. Isomerase; 25. Arabinose isomerase; 26. Ribulose kinase; 27. Transketolase; 28. Transaldolase.

CHAPTER 2.

COMPARATIVE WHOLE GENOME TRANSCRIPTIONAL RESPONSES OF *RUMINOCOCCUS ALBUS* STRAINS 7 AND 8 GROWN ON COMPLEX POLYSACCHARIDES AND DEFINED POLYSACCHARIDES

2.1 Introduction

The mammalian genomes lack the enzymes required to deconstruct the structural polysaccharides, cellulose and hemicellulose, present in plant cell wall. Herbivorous animals rely on the symbiotic microbes in the rumen, caecum or colon to break down cellulose and hemicellulose in their forage based diet. The resulting production of short-chain fatty acids and microbial proteins from fibrolytic microbial fermentation play a pivotal role in host nutrition. Omnivorous animals, including humans, rely less on microbial fermentation in the large-intestine for the acquisition of nutrients, but their fermentation products from resident microbiota, such as butyrate, play a significant role in host physiology related to health (1–3). Thus, for better understanding of host-microbe interactions, it is relevant to establish the model for fiber degradation and metabolism of fibrolytic gut microbes.

In mammalian gut, although there is compositional variation depending on diet, the dominant fibrolytic bacteria belong to the Bacteroidetes and Firmicutes phyla (4–6). In the human gut, many studies have revealed that the *Bacteroides* species are able to degrade a broad range of polysaccharides, including diet-derived plant polysaccharides (e.g. hemicelluloses and cellulose) and host-derived carbohydrates (e.g. mucin glycans) (7–9). Through deep genomic and biochemical analysis, Polysaccharide Utilization Loci (PULs) has been established as a model paradigm for the fibrolytic mechanism within the Bacteroidetes phylum (10–12). Despite that the

Firmicutes are more abundant than Bacteroidetes in the fiber associated fraction in the human gut, the fibrolytic Firmicutes have received far less attention than Bacteroidetes, and their fibrolytic mechanisms have been inferred from the predominant fibrolytic Firmicutes in the rumen, *Ruminococcus* species (13–15). One of the predominant fibrolytic bacteria in the rumen, *Ruminococcus flavefaciens*, is known to degrade plant cell wall using a cellulosome-type enzyme complex (4, 5, 16–18). The cellulosome paradigm has been established through intensive studies on fibrolytic soil bacteria, *Clostridium* species (16, 18, 19). Numerous polysaccharidases are arranged on the surface of *C. thermocellum* through interaction between dockerin and cohesion domains that are present on catalytic enzymes and scaffoldins, respectively. This assembled enzymatic complex mediates cell attachment to plant cell wall. In *R. flavefaciens*, cohesion- and dockerin-containing proteins have been characterized through genomic, transcriptomic, and biochemical analyses (4, 5, 16, 17). In addition, a recent study reported that the cellulolytic human gut species, *Ruminococcus champanellensis*, harbors cellulosomal components including dockerin and cohesion (20, 21). Consequently, the cellulosome paradigm has been expanded to the fiber degradation model for rumen and human gut species from soil species.

However, the fibrolytic mechanism for *R. albus*, a predominant rumen species, is not well studied. Of the two key components for cellulosome system, only dockerin encoding genes have been found in *R. albus*, suggesting that this species use other fibrolytic mechanism instead of cellulosome (22, 23). Despite a similar repertoire of glycoside hydrolase (GH) families between *R. albus* and *R. flavefaciens*, the cellulose adhering mechanism of *R. albus* is known to be distinct from cellulosome bearing ruminococci, which include three mechanisms; a Pil-like protein (CbpC), a glycocalyx, and carbohydrate binding module 37 family (CBM37) that is found only within *R. albus* species (24–28, 22, 23). Although many carbohydrate-active enzymes

(CAZymes) of *R. albus* species have been biochemically characterized (29–41), it is still insufficient to define the fibrolytic system of *R. albus.* Little molecular evidence has supported how *R. albus* species respond to plant cell wall consisting of hemicellulose and cellulose, what genes are involved in their fibrolytic system, and how they metabolize end products from the hydrolysis of plant cell wall. Various strains of *R. albus* are able to utilize a variety of plant polysaccharides, including cellulose and hemicelluloses. However, cellulolytic capability and pigmentation varies between strains (42–44). Among three strains for which genome sequences are publicly available, the yellow pigmented strain, *R. albus* 7, is known to degrade cellulose better than the non-pigmented strain, *R. albus* 8 (44). Thus, comparison of cellulolytic and hemicellulolytic capability between two distinctive strains, *R. albus* 7 and 8, may unravel the representative fibrolytic system and sugar metabolic pathways employed by *R. albus* species.

Herein, we describe the representative fibrolytic system of one of predominant fibrolytic species, *R. albus*, through comparative genomic, phenotypic, and transcriptomic analyses between *R. albus* 7 and 8. Furthermore, we also proposed that *R. albus* has developed strain specific strategies for fiber degradation and sugar metabolism as a result of differential gene regulation and a unique metabolic enzyme for pentose.

2.2 Materials and methods

Comparative genomic analysis of *R. albus* strain 7 and 8. Genomic sequences of *R. albus* 7 (GenBank accession numbers: NC_014833 for chromosome and NC_014824 to NC_014827 for plasmids) and *R. albus* 8 (NZ_ADKM020000001 to NZ_ADKM02000136) were used for comparative genomic analysis between both strain 7 and 8. Genes were assigned to

Carbohydrate-Active enZYmes (CAZy) families (GHs, PLs, and CEs) if they exhibited significant similarity (e-value < 1×10^{-5}) to biochemically characterized proteins already cataloged in a CAZy family (45), and the dbCAN server was used to verify the CAZy annotations (46). Other catalytic functions of genes were predicted using both the Pfam (47) and the conserved domain database (48) using e-value cutoff of 1×10^{-5}. Signal peptides were predicted using SignalP v4.0 (49). To determine the homologs of CAZy genes in between *R. albus* strain 7 and 8, reciprocal blast was performed and homologs were assumed if two genes each in a different genome of strain find each other as the best hit in the other genome.

Growth of *R. albus* strain 7 and 8 on defined substrates and complex substrate, AHPCS. *R. albus* 7 and 8 were cultured anaerobically at 37 °C in butyl rubber-stoppered Balch tubes using a previously reported defined medium (Table A.1; Kim *et al.*, 2014). The medium was modified to contain 0.4 % (w/v) of defined substrates including cellobiose (Sigma-Aldrich, St. Louis, MO), phosphoric acid swollen cellulose (PASC; Wood, 1971), wheat arabinoxylan (WAX; Megazyme, Bray, Ireland), sugar beet debranched arabinan (Megazyme), Beechwood xylan (Sigma-Aldrich), or 0.5 % (w/v) of complex substrate, alkaline hydrogen peroxide treated corn stalk (AHPCS; Kerley *et al.*, 1985), as the sole carbohydrate source. To determine crystalline cellulolytic activity, 0.4 % (w/v) of Whatman No.1 filter paper was used together with 0.1 % (w/v) of cellobiose as carbohydrate source for both strains. Cellobiose was used to stimulate bacterial growth at initial growth phase and the depletion of cellobiose was measured using high-performance liquid chromatography (HPLC) system described below.

For the single culture experiments, we used the modified 500 ml Corning glass bottles (Pyrex; Corning Glass Works) with 25ml Balch tube (Bellco Glass) attached to neck and side of a glass bottle, which enabled to grow cells in a large volume as well as measure the absorbance

at 600 nm (A_{600nm}) of culture. For the growth experiment with complex substrate, the medium was prepared in 150 ml serum bottle. The cells were adapted to the respective media by culturing for 48 hours three times in succession. Inoculum volume of preculture was 5 % of a total volume of fresh medium and the A_{600nm} values of the medium containing soluble substrates (cellobiose and wheat arabinoxylan) were monitored using a Spectronic 21D spectrophotometer (Milton Roy) and 1 ml of sample was collected anaerobically at each sampling time, followed by centrifugation at 13,000 g for 20 minutes. The resulting pellets were used to quantify the total protein concentration for insoluble substrate containing media (PASC, filter paper, and AHPCS) and supernatants were used to quantify the fermentation products during growth with all different substrates. For the total protein quantification, cell pellets were washed twice with 1% (w/v) KCl and, sequentially, boiled in 1% of CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate) solution for 20 min. Then, total protein was measured using the Bradford method with bovine serum albumin as a standard (53). To measure the A_{600nm} values in liquid fraction of AHPCS containing medium, we used a 25 ml of Balch tube. The insoluble AHPCS completely settled down at the bottom of tube within 30 minute after inoculation. Then, the absorbance value was measured using a Spectronic 21D spectrophotometer. The cellobiose and fermentation products in supernatant were quantitated by Shimadzu HPLC system (Shimadzu, Kyoto, Japan) using an Aminex HPX-87H column (Bio-rad, Hercules, CA, USA) with a refractive index detector (RID-10A).

Quantitative PCR in co-culture experiment. For the co-culture experiment, 20 ml of each strain grown on 0.4 % (w/v) of AHPCS containing medium for 24 hours was mixed together in the anaerobic serum bottles and the mixed pre-culture was inoculated into the fresh medium (0.4 % of AHPCS, w/v). The growth of each strain in co-culture medium was measured using

quantitative PCR (qPCR). Genomic DNA from each time points was used as a template and specific primers were designed to amplify a unique gene that is found in either *R. albus* strain 7 or 8 (Table A.2). The qPCR was performed using the LightCyclerR480 (Roche Diagnostics, Mannheim, Germany) and the SYBR green PCR kit (Qiagen, Hilden, Germany). The thermocycling reaction was performed under the following conditions: 95℃ for 30 s and 40 cycles of 30 s at 95 ℃ and 30 s at 59 ℃. To convert the copy number of gDNA of each strain to the relative absorbance value, the following standard curves were generated (Fig. 2.1).

Sugar analysis of culture medium and substrate residue. The AHPCS (0.4 %, w/v) containing media were prepared in triplicates of Balch tubes. During the bacterial growth, the 10ml of whole culture medium was collected by centrifuging at 13,000g for 20 minutes at different time points. The supernatant was carefully transferred to new 15 ml tubes and the residue was washed five times with ddH_2O and oven-dried at 60 ℃. The water soluble sugars in culture supernatant were displayed by thin layer chromatography (TLC) assay with a mobile phase composed of ethyl acetate, acetic acid, and water (3:2:1) and visualized by spraying with methanolic orcinol and heating at 120 ºC for 5 min. A 600 µl volume of ethanol was added to the 400 µl of supernatant, and the mixture was evaporated through a Savant DNA 120 SpeedVac® concentrator (Savant; Ramsey, MN). The dried product was resuspended in 20 µl of ddH_2O, and 1 µl of each sample was spotted on Silica Gel 60 Å TLC plates (Whatman).

For quantitative analysis of the mono- and oligo-saccharides in the supernatant and dried residue, both liquid and solid samples were hydrolyzed with sulfuric acid (72 %, w/v) according to a previously described method (54). Briefly, 700 µl of each supernatant was mixed with 25 µl of sulfuric acid (72 %, w/v) and the mixture was autoclaved for 1 h. A sugar recovery standard (SRS) comprising the monosaccharides (glucose, xylose, mannose, arabinose, and galactose) was

prepared in the similar way; one aliquot autoclaved with the samples and another aliquot left at room temperature. The solid samples were ball-milled, and incubated with 500 µl of sulfuric acid (72 %, w/v) for 1 h. Then, 14 ml of ddH_2O was added, and subsequently, the diluted mixture was autoclaved for 1 h. A SRS for solid sample was prepared in a similar way. The monosaccharides in sulfuric hydrolysates were quantitated by an ICS-5000^+ HPIC system (Thermo Fisher Scientific, Sunnyvale, CA, USA) using CarboPac PA20 column (Thermo Fisher Scientific, Sunnyvale, CA, USA) with a pulsed-amperometric detector (PAD).

RNA extraction and transcriptomic analysis using RNA sequencing. For RNA extraction experiments, the cells grown on defined substrates were harvested at mid-log phase and the cells grown on AHPCS were harvested at five different growth phases (lag, early-log, mid-log, late-log, stationary phase, inferred from the growth curves), by combining the culture with 2 volumes of RNAprotect® bacterial reagent (Qiagen), followed by centrifugation at 13,000 × *g* for 10 min at room temperature. The resulting cell pellets were stored at – 80 °C until RNA extraction. In the subsequent steps, the cell pellets were treated with lysis buffer (200U/ml of mutanolysin, 150µg/ml of proteinase K, 25mM EDTA, and 0.5% (w/v) SDS) for 30 minute at 55 °C. The total RNA was extracted with the RNeasy mini kit (Qiagen) with the optional on-column DNase treatment step. Then, the total RNA was eluted with DEPC-treated nuclease-free water and stored at – 80 °C until RNA sequencing.

For RNA-Seq analyses, RNA isolated from two biological replicates were used for each growth condition. Bacterial ribosomal RNAs were removed from 10 µg of total RNA with the MicrobExpress kit (LifeTechnologies). The enriched mRNA fraction was converted to RNA-Seq libraries using the TruSeq Stranded RNA Sample Prep kit from Illumina. The barcoded libraries were pooled in equimolar concentration and the pool was analyzed by quantitative PCR and

sequenced on one lane for 101 cycles on a HiSeq2000 using a TruSeq SBS sequencing kit version3. Fastq files were generated and demultiplexed with the bcl2fastq v1.8.4 Conversion Software (Illumina, Inc.). RNAseq library statistics are shown in Table A3 to A.5.

The RNA-Seq data were analyzed using CLC genomics workbench version 5.5.1 from CLC Bio (Cambridge, MA). The genomic sequences of *R. albus* 7 and 8 were used as the reference genome, and RNA-Seq reads were mapped onto the reference sequences using the CLC software. Reads were only assembled if the fraction of the read that aligned with the reference genome was greater than 0.9 and if the read matched other regions of the reference genome at less than 10 nucleotide positions. Then, the RNA-Seq output files were analyzed for statistical significance by using the proportion-based test of Baggerly (55).

Cloning, expression, and purification of the putative GH genes of *R. albus* 7. In this study, the highly expressed genes in *R. albus* 7 that encode putative GH domains were amplified by the Phusion high-fidelity DNA polymerase and cloned into the pET-46b vector (Ek/LIC; Novagen). The ligation product was transformed into E. coli JM109 using heat-shock method. After selection on lysogeny broth (LB) plates supplemented with ampicillin at 100 µg/ml, the colony was picked and cultured in LB medium with same antibiotics at the same concentration. The plasmids were purified from each culture using a QIAprep Spin Miniprep kit (Qiagen), and sequenced to confirm the integrity of the cloned gene at W. M. Keck center for comparative and functional genomics in the University of Illinois at Urbana-Champaign.

For gene expression, the plasmid construct was transformed into E. coli BL21-CodonPlus (DE3) RIPL competent cells using the heat shock method, and cells were grown on LB agar plates supplemented with ampicillin (100µg/ml) and chloramphenicol (50µg/ml). After 16 h

incubation, five colonies were picked and inoculated into LB medium (10 ml) containing ampicillin (100µg/ml) and chloramphenicol (50µg/ml). The cells were grown for 5 to 6 hours at 37 °C with vigorous shaking, and then transferred to 1 liter of LB medium containing ampicillin and chloramphenicol at the same concentrations described above. When the optical density at 600_{nm} of the LB culture reached 0.5, isopropyl-β-D-thiogalactopyranoside (IPTG) (final concentration 0.1 mM) was added into the LB culture to induce gene expression, and the culture was incubated at 16 °C for 16 h.

The cells were harvested by centrifugation, and the cell pellets were resuspended in 30 ml lysis buffer (100 mM HEPES, 500 mM NaCl, 10% glycerol, 0.5 mM Tris(2-chloroethyl) phosphate (TCEP), pH 8.0) and lysed using a French pressure cell (American Instrument Co., Inc., Silver Spring, Md). The cell debris was separated by centrifugation (35,000 g, 30 min, 4 °C). Of ten putative GH genes cloned in this study, the product of Rumal_2606 gene encoding a putative GH48 domain was insoluble and consequently, it was excluded following purification steps and functional assignments. The recombinant proteins were purified by Talon metal affinity resin with binding buffer (100 mM HEPES, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP, 10 mM imidazole, pH 8.0), washing buffer (50 mM HEPES, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP, 10 mM imidazole, pH 7.5), and elution buffer (50 mM HEPES, 500 mM NaCl, 10% glycerol, 0.5 mM TCEP, 300 mM imidazole, pH 7.5). The purified proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and staining with Coomassie brilliant blue G-250 (Fig. 2.2). Then, the proteins were concentrated and dialyzed into a storage buffer (50 mM HEPES, 300 mM NaCl, 10% glycerol, 0.5 mM TCEP, pH7.5) using Amicon Ultra-15 centrifugal filter units. The concentration of each purified protein was determined using the NanoDrop 1000 apparatus from Thermo Fisher Scientific Inc. (Waltham,

MA), based on the molecular mass and extinction coefficients of Rumal_2447 (109 KDa; 248,440 $M^{-1}cm^{-1}$), Rumal_0896 (71 KDa; 159,630 $M^{-1}cm^{-1}$), Rumal_1049 (74 KDa; 213,140 $M^{-1}cm^{-1}$), Rumal_2448 (98 KDa; 246,950 $M^{-1}cm^{-1}$), Rumal_2946 (58 KDa; 94,770 $M^{-1}cm^{-1}$), Rumal_3757 (99 KDa; 204,440 $M^{-1}cm^{-1}$), Rumal_1601 (44 KDa; 77,810 $M^{-1}cm^{-1}$), Rumal_2906 (78 KDa; 76,210 $M^{-1}cm^{-1}$), and Rumal_0076 (73 KDa; 123,190 $M^{-1}cm^{-1}$).

Enzymatic activities of nine glycoside hydrolases on various polysaccharides. The enzymatic activities of the nine CAZymes were determined at 37 °C in 50 mM sodium phosphate buffer (pH6.5) and 150 mM NaCl. Each protein (500nM) was incubated with 0.5 % (w/v) of soluble polysaccharide substrates purchased from Megazyme (Bray, Ireland) or 2 % (w/v) of insoluble polysaccharide substrates. Soluble substrates include carboxymethyl cellulose, barley β-glucan, lupin galactan, larch arabinogalactan, carob galactomannan, konjac glucomannan, tamarind xyloglucan, and wheat arabinoxylan. Insoluble substrates include phosphoric acid swollen cellulose, curdlan (Megazyme), ivory nut mannan (Megazyme), and beechwood xylan (BWX; Sigma-Aldrich, St. Louis, MO). After 16 hours incubation, the reducing ends were quantified using a *para*-hydroxybenzoic acid hydrazide (PAHBAH) assay with glucose as a standard (56).

2.3 Results

Comparative genomic analysis suggests different substrate preferences between *R. albus* 7 and 8. To estimate genotypic variation for fiber degradation between *R. albus* 7 and 8, we analyzed the predicted carbohydrate-active enzymes (CAZymes) encoded by the genomes of both strains, including glycoside hydrolases (GHs), polysaccharide lyases (PLs), and carbohydrate esterases (CEs). The genomes of both strains possess a similar repertoire of

predicted GHs, PLs, and CEs, in terms of the number and variety of CAZy families (Table A.6). Of the well-known cellulase domains, including GH5, GH8, GH9, GH44, GH48, and GH74, both strains possess the same number of these total GH domains on the genome. However, we identified that *R. albus* 8 possesses three more genes encoding GH43 domain for which homologs are not present on the genome of *R. albus* 7 (Fig. 2.3A). The functions of the unique GH43 genes in *R. albus* 8 are predicted as extracellular and intracellular endoarabinanases. In addition, these GH43 genes in *R. albus* 8 are clustered with putative sugar transporters, the intracellular α-L-arabinofuranosidase (Ara51A), and L-arabinose isomerase, which are required for utilization of the end products of arabinan hydrolysis. Considering that the genome sequence of *R. albus* 7 is complete, it is reasonable to conclude that *R. albus* 7 harbors only the ortholog of Ara51A gene (Fig. 2.4). These results suggest that *R. albus* 8 has a potential to utilize a broader range of hemicelluloses, including arabinan, than *R. albus* 7.

***R. albus* 7 and 8 have different capability for utilizing cellulose and hemicelluloses.** To determine the phenotypic variation between *R. albus* 7 and 8 on different defined substrates, both strains were cultured in a defined medium containing 0.4 % (w/v) of microcrystalline cellulose (Whatman filter paper), debranched arabinan, cellobiose, phosphoric acid-swollen cellulose (PASC), or wheat arabinoxylan (WAX). *R. albus* 7 was able to grow on filter paper, while *R. albus* 8 was unable to grow (Fig. 2.5). Conversely, when debranched arabinan used as the sole carbohydrate source, *R. albus* 8 enabled to utilize, but not *R. albus* 7, which is in accordance with our prediction based on comparative genomic analysis (Fig. 2.3B). Next, we tested soluble substrates (cellobiose and WAX) and amorphous cellulose (PASC) that could be utilized by both strains. Both strains were able to grow on cellobiose and PASC with similar growth pattern and production of fermentation products, inferred from the value of standard deviation (Fig. 2.6).

However, when grown on WAX, *R. albus* 7 produced less acetate (12.9 mM) and more ethanol (13.1 mM) than *R. albus* 8 that produced 17.8 mM of acetate and 11.4 mM of ethanol (Fig. 2.6H-I).

To investigate the phenotypic variance between two strains for complex polysaccharides like plant cell wall, both strain 7 and 8 were cultured in a defined medium containing alkaline hydrogen peroxide treated corn stalk (AHPCS). The AHPCS was mainly composed of cellulose (53.7 %) and hemicellulose (23.1 %), which allows both strains to utilize either cellulose or hemicellulose, or both (Table A.7). To examine bacterial growth and adherence to substrates, we quantified the total protein concentration and measured the absorbance values (A_{600nm}) in liquid fraction of bacterial culture once insoluble AHPCS had settled down to the bottom of the Balch tubes. The growth curve of each strain using total cell protein shows that cell number in the *R. albus* 7 culture would be higher than or similar to that in the *R. albus* 8 culture at early and stationary growth phase (Fig. 2.7A). However, the *R. albus* 8 culture had a higher A_{600nm} value than the *R. albus* 7 culture at lag to early-log phases and stationary phase (Fig. 2.7*B*). Especially, the difference of absorbance between two strains increased after 20 hours (late-log) (A_{600nm}: 0.282 for *R. albus* 7 and 0.693 for *R. albus* 8). Compared to the total protein concentration, the different values of A_{600nm} suggest the preferred planktonic growth of *R. albus* 8 as opposed to the substrate adherent growth of *R. albus* 7.

Similar with the fermentation profile on WAX, *R. albus* 7 produced less acetate (15.2 mM) and more ethanol (13.4 mM) than *R. albus* 7 that produced 15.9 mM of acetate and 11.8 mM of ethanol during growth on AHPCS for 48 hours (Fig. 2.7C-D). In addition, *R. albus* 8 produced formate as one of the fermentation products during growth on AHPCS, but not *R. albus* 7. These differential profiles of fermentation products on hemicellulose and hemicellulose

containing plant cell wall suggest that pentose sugars released from hemicellulose hydrolysis were differently metabolized through each strain specific fermentation pathway.

Differential AHPCS degradation and utilization between *R. albus* 7 and 8. To determine the differences in hydrolysis and utilization of individual sugar components in the AHPCS by *R. albus* 7 and 8, each fraction of AHPCS residue and culture supernatant were hydrolyzed using sulfuric acid, and the hydrolyzed monosaccharides in both fractions were quantified. The major monosaccharides in the AHPCS residue were glucose, mainly derived from cellulose, and xylose, derived from hemicellulose. During initial growth for 12 hours, *R. albus* 7 hydrolyzed both cellulose and xylan in AHPCS, while *R. albus* 8 mostly hydrolyzed xylan fraction in AHPCS (Table A.8). Although glucose was the predominant monosaccharide in AHPCS residue, xylose was more abundant than glucose in the culture supernatant, indicating that both strain 7 and 8 prefer to utilize cellulosic sugar than hemicellulosic sugars (Table A.8). During growth for 48 hours, most of xylan in AHPCS was hydrolyzed by both strains as compared to initial amount of xylan at 0 hour (92.4 % in *R. albus* 7, 93.8 % in *R. albus* 8, respectively). However, less xylose was accumulated in the culture supernatant of *R. albus* 8 than *R. albus* 7 (75.2 μM in *R. albus* 7, 54.2 μM in *R. albus* 8, respectively), suggesting more utilization of xylan by *R. albus* 8 than *R. albus* 7 during growth for 48 hours (Table A.8). Consistently, the soluble sugar profiles analyzed by thin layer chromatography (TLC) assay suggest that inferred from sugar standards, the hemicellulosic mono- or di-saccharides were more accumulated in the culture of *R. albus* 7 than *R. albus* 8 (Fig. 2.8).

Comparative transcriptomic analyses of CAZyme genes in *R. albus* 7 and 8 grown on AHPCS and defined substrates. To identify cellular mechanism of *R. albus* 7 and 8 for the utilization of plant cell wall and its major polysaccharide components, cellulose and xylan, we

examined the transcriptomes of both strains at five different growth phases during growth on AHPCS (late-lag; 4h, early-log; 8h, mid-log; 12h, late-log; 20h, and stationary; 36h) and mid-log phase during growth on cellobiose, PASC or WAX. The normalized expression value of each gene is reported as RPKM (reads per kilo base per million mapped reads). During growth on AHPCS, both *R. albus* 7 and 8 expressed a variety of CAZyme genes, including GH, PL, and CE genes. Among a total of the extracellular CAZyme genes predicted by the presence of signal peptide at N-terminus, the predicted or characterized endoglucanases and endoxylanases were predominantly co-expressed (> 2000 RPKM) in both strain during early- to mid-log phases. The top five highly expressed genes encoded GH5, GH9, GH10, GH11, and CE4 domains in *R. albus* 7 and GH5, GH9, GH11, GH48, and CE5 domains in *R. albus* 8, respectively (Fig. 2.9 and 2.10A-B). Interestingly, all of those highly expressed CAZyme genes in both strains encode CBM37 domain only found in the genome of *R. albus* strains, suggesting that *R. albus* species employ a CBM37 dependent hydrolyzing mechanism for plant cell wall, unlike other fibrolytic species.

Of those cellulases (GH5, GH9, and GH48) highly expressed on AHPCS, the GH5 gene in *R. albus* 7 (Rumal_0896) and its homologs in *R. albus* 8(CUS_6389) were constitutively highly expressed on defined polysaccharides, PASC and WAX (Fig. 2.11). We characterized the enzymatic activities of Rumal_0896 on a variety of defined polysaccharides. Based on reducing sugar profiles, the purified gene product of Rumal_0896 showed the hydrolysis activities on both cellulose and hemicellulose, including PASC, ß-glucan, glucomannan, xyloglucan, WAX, and beechwood xylan (Fig. 2.12). These results suggest that both strains rely on the versatile GH5 endoglucanase capable of hydrolyzing cellulose and hemicelluloses during growth on plant cell wall. However, the other cellulase genes encoding GH9 (Rumal_2447) or GH48 (GH2606) in *R.*

albus 7 were induced greater than 2 fold by only PASC, while those homologs (CUS_6909 and CUS_8076) in *R. albus* 8 were induced by both PASC and WAX. Together with highly expressed endoxylanase (Xyl11D; CUS_6323) and its homolog (Rumal_0908), *R. albus* 7 co-expressed the primary cellulases (GH9 and GH48) and hemicellulases (GH11) in response to PASC, whereas *R. albus* 8 displayed this co-expression pattern on only WAX. Among CAZyme genes induced greater than 2 fold on PASC relative to WAX, *R. albus* 7 had 21 more genes transcriptionally responding to PASC than *R. albus* 8 (Fig. 2.13B). However, when we compared CAZyme genes induced by WAX to PASC (> 2 fold), *R. albus* 8 had 17 more genes transcriptionally responding to WAX than *R. albus* 7 (Fig. 2.13D). These distinct transcriptional features of CAZyme genes between two strains suggest that each fraction of cellulose and hemicellulose in plant cell wall induce different array of CAZyme genes employed by *R. albus* 7 and 8. Therefore, each strain would have a different target in plant cell wall for the hydrolysis and utilization.

Transcriptional features of sugar transporters between strains. To elucidate time dependent end product utilization of AHPCS hydrolysis by *R. albus* 7 and 8, we sought to identify the potential transporters for cellobiose, cellooligosaccharides, and xylooligosaccharides through our transcriptomic analysis. Among the predicted sugar ABC transporter genes that highly expressed during growth on AHPCS, we identified candidate genes, based on which had the highest expression level on their predicted substrate, cellobiose, PASC, and WAX (Fig. 2.10C). The predicted xylooligosaccharide transporter genes were highly expressed at lag phase and down-regulated in both strains, as cells grew. Subsequently, the cellobiose transporter genes were up-regulated at early log phase and down-regulated in later growth phases. *R. albus* 7 exhibited sequential expression pattern for the putative cello-oligosaccharides transporter after the

cellobiose transporter, whereas in *R. albus* 8, the cello-oligosaccharides transporter genes were up-regulated together with cellobiose transporter genes at early log-phase and then, down-regulated at later growth phases.

The intracellular GH genes had a similar expression pattern with their cognate substrate transporters (Fig. 2.14). The hemicellulosic sugar processing GH genes, including β-xylosidase, α-L-arabinofuranosidase, and α-glucuronidase, were highly expressed at lag phase and down-regulated later growth phases. Cellobiose phosphorylase gene was highly expressed during lag to early-log phase and down-regulated as cells grew, in agreement with transcriptional responses of the predicted xylooligosaccharide and cellobiose transporters. However, the gene encoding cello-oligosaccharide phosphorylase in *R. albus* 7 was constitutively expressed during growth on AHPCS, whereas in *R. albus* 8, the expression pattern of cellooligosaccharide phosphorylase gene was consistent with cellobiose phosphorylase gene. These results suggest that *R. albus* 7 and 8 utilize xylooligosaccharides at initial cell growth and subsequently, utilize cellobiose, presumably derived from amorphous cellulose region, at early-log phase. Inferred from the sugar analysis of AHPCS residue, the cello-oligosaccharides were likely released from cellulose after early-log phase for *R. albus* 7 and mid-log phase for *R. albus* 8. Accordingly, the genes for cellooligosaccharides utilization in *R. albus* 7 were up-regulated in response to the availability of cellulosic sugars. However, those genes in *R. albus* 8 were not responded to the cellooligosaccharides.

Taken together with comparative genomic, phenotypic and transcriptomic analyses on complex and defined substrates, it seems likely that in spite of similar cellulolytic capability, *R. albus* 8 prefers to utilize hemicellulose, unlike a cellulolytic strain, *R. albus* 7 that utilizes hemicellulose until it can utilize cellulose.

***R. albus* 8 harbors dual pentose metabolic pathways, the PP and PK pathways.** When grown on WAX, *R. albus* 7 produced acetate and ethanol in a ratio of 1.0, while *R. albus* 8 produced acetate and ethanol in a ratio of 1.6. This result led us to examine sugar metabolism of *R. albus* 7 and 8. Both strains harbor all of the genes required to sugar fermentation pathways, including the Embden-Meyerhof-Parnas (EMP) pathway, pentose phosphate (PP) pathway, acetate and ethanol fermentation pathways. Notably, we identified a gene encoding a putative xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (XFP; EC 4.1.2.9, EC 4.1.2.22) on the genome of *R. albus* 8 (Fig. 2.15), which is orthologous to the biochemically characterized XFP (Genbank ID: NP_347971) in the soil cellulolytic bacterium, *Clostridium acetobutylicum* ATCC 824, with 73 % similarity of amino acid sequence (Table A.9). The XFP enzyme catalyzing a xylulose-5-phosphate into glycealdehyde-3-phosphate and acetyl phosphate is a key component of the phosphoketolase (PK) pathway in *C. acetobutylicum* (57, 58). Based on the stoichiometry of the PP pathway and PK pathway for pentose fermentation, the ratio of acetate to ethanol is 1.0 for the PP pathway and 3.0 for the PK pathway (Table A.10). Thus, the profile of fermentation products and the presence of XFP gene in *R. albus* 8 suggests that pentose was metabolized through the PP pathway in *R. albus* 7, while in *R. albus* 8, both PP and PK pathways were used for pentose fermentation with the dominance of PP pathway over PK pathway. To examine the metabolic effect of PK pathway relative to the PP pathway on the growth kinetics of *R. albus* 7 and 8 on pentose, we tried to use homogenous pentose (xylose or arabinose) as a growth substrate, but none of strains were able to use mono-pentose sugar (data not shown). When Beechwood xylan (>90 % xylose residue) was alternatively used in the culture medium, *R. albus* 8 had a higher growth rate than *R. albus* 7 (Fig. 2.16).

Transcriptional pattern of the sugar fermentation pathways. As both strains appeared to metabolize pentose through different pathways, we examined the transcriptional evidence to see which pathway becomes dominant for different substrates. Each pathway contains a unique enzyme, but also shares the entire reaction steps from glyceraldehyde-3-phoshate to acetate fermentation. The XFP is a unique enzyme to the PK pathway, while the transketolase (TK) and transaldolase (TA) are unique to the PP pathway. In addition, the XFP and transketolase enzymes share the substrate, xylulose-5-phosphate. Thus, we postulate that the transcriptional abundance of each XFP and transketolase genes reflects the distribution of metabolic flux of pentose into each pathway. In *R. albus* 7, the putative transketolase and transaldolase genes were up-regulated on WAX (2 fold and 621 fold, respectively), relative to cellobiose, and highly expressed at lag-phase during growth on AHPCS (Fig. 2.17A; Table A.11). In *R. albus* 8, the XFP gene was up-regulated greater than 6 fold, while the transketolase gene was 1.9 fold up-regulated, when grown on WAX, relative to cellobiose (Table A.11). Despite dampened transcriptional response to WAX, the transketolase gene was 3 times more dominant than the XFP gene in the transcriptome of *R. albus* 8, which is in accordance with our prediction based on the profile of fermentation products. However, when *R. albus* 8 was grown on AHPCS, the PK pathway was 2 times more dominant at initial cell growth phase, suggesting that *R. albus* 8 relies on the PK pathway when the cells begin to grow on the plant biomass.

Next, we examined the transcriptional level of genes related to formate fermentation, including the putative pyruvate formate lyase (PFL), PFL activase (AE), pyruvate ferredoxin oxidoreductase (PFOR), hydrogenases (ferredoxin dependent HydA2 and electron-bifurcating HydABC). During growth on AHPCS, *R. albus* 8 had a higher expression level of PFL activate (AE) and higher expressional ratio of PFL: PFOR than *R. albus* 7, in accordance with formate

production profile (Fig. 2.17B). The PFOR reduces the oxidized ferredoxin (Fd_{ox}) and the resulting reduced ferredoxin (Fd_{red}) is used as electron donor for hydrogen production by either HydA2 at high level of hydrogen or HydABC at lower level of hydrogen. During growth on AHPCS, *R. albus* 7 had a higher expression level of both HydA2 and HydABC than *R. albus* 8, with exception of HydA2 at early-log phase (Fig. 2.17B). To support, we measured hydrogen production of both strains and confirmed that *R. albus* 7 produced more hydrogen than *R. albus* 8 during growth on AHPCS (Fig. 2.17C). However, it is unclear why *R. albus* 7 did not produce formate, in spite that the putative PFL genes were highly expressed after mid-log phase.

Co-culture of both R. albus 7 and 8 on AHPCS. With less cellulolytic capability, *R. albus* 8 is unlikely to compete with other cellulolytic species, including *R. albus* 7. However, if *R. albus* 8 targets for hemicellulose less preferred by cellulolytic species in the plant cell wall, it seems feasible that *R. albus* 8 is able to survive and co-exist with *R. albus* 7. We grew both strain 7 and 8 together in a defined medium containing 0.5% (w/v) of AHPCS for 48 hours. To determine the growth of each strain 7 and 8 in co-culture, we performed quantitative PCR (qPCR) that specifically amplified a unique gene present on either strain 7 or 8 genome. After co-incubation, both strains started to grow together on AHPCS, but *R. albus* 7 became more abundant than *R. albus* 8 in the total cell population, composed of the substrate adherent cells and planktonic cells (Fig. 2.18A). In addition, fermentation profile of co-culture was similar with those profiles of *R. albus* 7 single culture on AHPCS, in terms of the acetate: ethanol ratio and formate production (Fig. 2.18B). However, during early growth (0 to 12 hours), the cell number of *R. albus* 8 was increased in liquid fraction and the short length of soluble sugars were less accumulated in the culture, compared to the profile on later growth after 12 hours or on single culture of *R. albus* 7 on AHPCS (Fig. 2.18C-D). After 12 hours, the number of *R. albus* 8 cells in liquid fraction

began to decrease and the short oligosaccharides and monosaccharides started accumulating in the culture. These results suggest that the planktonic cells of *R. albus* 8 utilize soluble sugars derived from the hydrolysis of AHPCS and consequently, *R. albus* 8 could grow together with other cellulolytic stain, *R. albus* 7 on AHPCS.

Differential expression of global gene regulators, cyclic di-GMP and Agr quorum sensing system, in *R. albus* 7 and 8. Since distinct strategies for plant cell wall degradation were identified in between strain 7 and 8, we sought to find a regulatory mechanism on the genomes of both strains underlying two strategies. Based on differential phenotypes between two strains for the substrate attachment during growth on AHPCS, we examined global regulators known to be involved in biofilm formation of bacteria: the bacterial second messenger, cyclic-di-GMP (c-di-GMP), regulatory system and the accessory gene regulator (Agr) quorum sensing system. The cellular level of c-di-GMP is regulated by diguanylate cyclases (DGCs) encoding a GGDEF domain and phosphodiesterses (PDEs) encoding an EAL domain (59). We identified that both strain 7 and 8 possess thirty-eight genes encoding predicted GGDEF, EAL, or both domains (Table A.12). When comparing strains across different growth phases, our transcriptional analysis showed that *R. albus* 7 had more up-regulated (> 2-fold) c-di-GMP regulatory genes and GH genes than *R. albus* 8 throughout growth on AHPCS (Fig. 2.19A-B). When comparing strains across defined substrates, c-di-GMP regulatory genes in *R. albus* 7 were more responded (> 2-fold) to their preferential substrate, cellulose, while c-di-GMP regulatory genes in *R. albus* 8 were more responsive to hemicellulose (Fig. 2.19C-D). These transcriptional patterns were similar to patterns of transcriptional responses of GH genes in both strains, suggesting that there is a correlation between c-di-GMP and GH gene expression. Next, we analyzed the transcriptional level of putative Agr quorum sensing genes, *agrABCD*, in both strains. *R. albus* 8

expressed *agrD* and *argB* genes, encoding autoinducing peptide (AIP) and AIP processing transporter, respectively, dramatically higher than *R. albus* 7 from early to late growth phase (Fig. 2.20; Table A.13). For example, during early-log to stationary phase, *R. albus* 8 expressed *agrB* from 25 to 62 times higher than *R. albus* 7, which is in accordance with phenotypic evidence showing more abundant planktonic cells of *R. albus* 8, relative to *R. albus* 7. Taken together, it seems that there is a strong correlation between transcriptional responses of c-di-GMP and Agr quorum sensing systems.

2.4 Discussion

Plant cell wall degradation by mammalian gut bacteria has been highlighted as a primary contribution to ruminant nutrition and a pivotal role in human health (60, 1, 2, 4, 5). Thus, intensive research has been conducted to identify the fibrolytic mechanism employed by two dominant phyla, Firmicutes and Bacteroidetes. Through deep molecular analyses, the PULs of *Bacteroides* genus has been established as the representative fibrolytic system of Bacteroidetes phylum, while the fibrolytic system of gut Firmicutes is still unclear, except for the cellulosomal paradigm propagated from the soil cellulolytic bacteria, *Clostridium* species (4, 17–19, 61). Recently, it has been reported that the ruminal Firmicutes, *Ruminococcus flavefaciens*, as well as human gut Firmicutes, *Ruminococcus champanellensis* harbors cellulosomal system for fiber degradation, supported by the genomic, transcriptomic, and biochemical analyses (4, 21, 18, 62). Thus, the cellulosomal paradigm seems to be expanded to the mammalian gut from the soil. However, the representative fibrolytic system of other predominant *Rumincoccus* species, *R. albus*, is still unclear due to little molecular evidences that are comparable between *R. albus* strains. In addition, the pentose metabolism of *R. albus* species, has received less attention than

the hexose metabolism, related to interest on their cellulolytic capability. Thurston et al. reported that *R. albus* B199 possessed functional pentose metabolic enzymes, including pentose isomerases, xylulokinase, ribulokinase, transketolase, but not phosphoketolase (63). Accordingly, *R. albus* species is thought to utilize pentose through the PP pathway. As the genome sequences of *R. albus* strains are available, it enables deeper understanding of their sugar metabolism at genomic and transcriptomic levels. Thus, through the comparative genomic, phenotypic, and transcriptomic analyses between *R. albus* strains, we seek to find whether two phenotypically distinct strains, *R. albus* 7 and 8, possess a shared fibrolytic system and sugar metabolism. If so, it would be a representative mechanism of *R. albus* species, or if not then, it would enable us to identify the strain specific fibrolytic mechanism, which likely account for the phenotypic variance between strains and ecological role in gut microbial community at strain level.

Once *R. albus* 7 and 8 started growing on AHPCS, both strains co-expressed a specific array of extracellular hemicellulases and cellulases, encoding GH5, GH9, GH10, GH11, or GH48 domain. Together with the biochemical evidences in the current and previous studies, these five GH families are likely to be primary GH enzymes employed by two strains of *R. albus* for the hydrolysis of plant cell wall. Our transcriptomic data is in accordance with the previous metagenomic research. Dai et al. reported that the top 3 abundant transcripts of cellulases and hemicellulases in cow rumen included GH9, GH5, GH48 and GH10, GH11, GH26, respectively, and these transcripts were primarily synthesized by *Ruminococcus* and *Fibrobacter* genera (64).

Since the adherence of cells to their insoluble substrate is a critical feature as cellulolytic species, we examined the known adhering mechanism of *R. albus* species including Pil-like protein (CbpC), glycocalyx, and CBM37, by a transcriptomic comparison between AHPCS attached strain, *R. albus* 7 and less associated strain, *R. albus* 8, phenotypically demonstrated in

our study. Through an adherence lacking mutant study, the CbpC protein has been implicated in cellulose adherence of *R. albus* 20 (25, 26, 65). In addition, Vodovnik et al. suggested that the orthologs of CbpC is also involved in cellulose attachment of the cellulosomal rumen species, *R. flavefaciens*, by showing up-regulation of a Pil-like protein when grown on cellulose (61). We found that both *R.albus* 7 and 8 harbor the homologs of CbpC, which was highly expressed (>10,000 RPKM) during growth on AHPCS (Table A.14). However, the Pil-like protein seems unlikely to involve direct attachment of cells to cellulose, based on the transcriptome of *R. albus* 8 when grown on AHPCS. We observed that *R. albus* 8 increased planktonic cells after 20 h hours (late-log), but the CbpC homologous gene in *R. albus* 8 was expressed greater at late-log and later growth phases than at early- and mid-log phases (Table A.14). In addition, the homologous genes in both strain 7 and 8 were not significantly induced (< 2 fold) in response to cellulose relative to cellobiose, which is consistent with the transcriptome of *R. albus* 7 grown on Avicel in continuous culture (22). Rakotoarivonina and his colleagues reported that the transcriptional level of CbpC homologous gene in the adhesion-defective mutant strain tended to be even higher than the wild type strain, *R. albus* 20 (66). Taken together with previous studies, our findings suggest that a Pil-like protein may serve as an accessory appendage in adhesion mechanism of *R. albus* species.

We also examined transcriptomic evidence for the role of glycocalyx in cell attachment to the plant cell wall. To date, it has not been identified which genes are involved in the glycocalyx synthesis. Based on the sugar composition in the glycocalyx of *R. albus* 7, it has been proposed that hexose-1-phosphate uridylyltransferase may be one of the candidates for glycocalyx synthetic genes, which was not up-regulated on cellulose, relative to cellobiose (28, 22). We found a gene encoding the putative hexose-1-phosphate uridylyltransferase in both

strains, but its expression level was very low (RPKM < 50) and not responsive to any specific substrates, which was consistent with a previous study (Table A.15-A.16; Christopherson *et al.*, 2014). As glycoside transferases (GTs) are required for the exopolysaccharide biosynthesis of biofilm forming bacteria, it is possible that GT genes may be involved in the synthesis of the glycocalyx (67–70). We identified the putative GT genes in both strains, but none of them were significantly induced by any polysaccharides (Table A.15-A.16). Thus, the adhering system mediated by a glycocalyx seems not to be regulated at the transcriptional level in both *R. albus* strains.

It is important to note that in both strains, most of highly expressed extracellular GH genes on PASC, WAX, and AHPCS encode CBM37 domains rather than dockerins. In addition, both strains harbor unknown function of CBM37 (UF-CBM37) genes that do not encode any known catalytic domains, and those genes were also highly expressed during growth on AHPCS (Fig. 2.21; Table A.17-A.18). More than 80 % of UF-CBM37 genes were predicted to encode extracellular proteins, inferred from the presence of a signal peptide (28 out of 35 genes in *R. albus* 7 and 20 out of 23 genes in *R. albus* 8), suggesting that together with CBM37 bearing GHs, the UF-CBM37 protein may work together outside of cells. Of the top 3 most highly expressed UF-CBM37 genes, we cloned and expressed the Rumal_0897 gene to determine catalytic activity for different polysaccharides, but it displayed no catalytic activity (data not shown). The CBM37 domain has shown a binding affinity to a broad range of polysaccharides, including cellulose and xylan, and even to cell surface (24), which leads to two existing theories for the role of CBM37 in fiber adhesion of *R. albus*. One is that CBM37 may work as a shuttle conveying the CBM37 bearing protein between bacterial cell surface and its target substrate (24). Some of UF-CBM37 genes in *R. albus* 7 contain leucine rich repeat (LRR) domains and it has

been proposed that the extracellular CAZymes may be complexed by UF-CBM37 protein via protein-protein interaction mediated by LRR domains (22). Later hypothesis is also supported by our findings that LRR containing UF-CBM37 genes were highly expressed in both strains during growth on AHPCS (>1000 RPKM at mid-log phase; Fig. 2.22). Interestingly, we also found that some of the highly expressed UF-CBM37 genes on AHPCS did not contain LRR domains, but contains only tandem CBM37 domains. In tandem CBM37 modules, proximal and distal modules are phylogenetically classified into different groups (24, 27). Xu et al. reported that distal CBM37 in tandem CBM37s exhibited a binding affinity to a variety of polysaccharides, whereas the proximal CBM37 bound to the putative ribosomal protein in the cell extract of *R. albus* 8 instead of polysaccharides (27). These results suggest that the proximal CBM37 may function as binding module with other protein. Thus, the UF-tandem CBM37 proteins, together with LRR containing UF-CBM37 proteins, may mediate protein-protein interaction between other extracellular CAZymes, leading to the localization of the CAZymes on the cell surface.

It is very important to note that both strains 7 and 8 apparently have a different capability of degrading crystalline cellulose, despite that both strain 7 and 8 possess all components of fibrolytic system described above. The phenotypic variation between two strains grown on crystalline cellulose implies that additional components must be implicated in the cellulolytic system of *R. albus* 7. One of the potential candidate components is the yellow pigments produced from *R. albus* 7, not *R. albus* 8. The yellow pigment is a distinct phenotypic feature of some cellulolytic microorganisms in the rumen and soil (43, 71, 72). Although the gene encoding the yellow pigment has been yet identified, the role of yellow pigments in cellulose degradation has been reported in rumen bacterium *Ruminococcus flavefaciens*, soil bacterium *Clostridium thermocellum*, and rumen fungi *Orpinomyces joyonii* (72–76). The yellow pigments produced by

those three species showed a strong affinity to both microcrystalline cellulose and endoglucanases. Considering that the highly expressed endoglucanases in *R. albus* possess CBM37 domains, the yellow pigments likely interact with cellulose and CBM37 bearing endoglucanases on the cell surface, resulting in tight adhesion of the yellow pigmenting strains to cellulose.

Considering the different cellulolytic capability between two strains, it seems unlikely that *R. albus* 8 is able to thrive in the rumen where a competition runs on plant biomass between the cellulytic bacteria, including *R. flavefaciens*, *F. succinogenes*, and even *R. albus* 7. Although relatively less abundant than other strains, *R. albus* 7 and SY3, *R. albus* 8 is still detected in the rumen by metagenomics analysis, meaning that *R. albus* 8 co-exists with other cellulolytic species and plays a role as one of member in the microbial community in the rumen (77). Therefore, we explored the growth strategy of *R. albus* 8 in the complex polysaccharide enriched environment, rumen. Our comparative genomic and transcriptomic analyses suggest that *R. albus* 8 intent to utilize hemicellulose less preferred by other cellulolytic species and metabolize pentose sugars rapidly through PK pathway that is absent in other cellulolytic species.

In spite of cellulolytic capability in the pure culture, wild type of *R. albus* 8 in rumen appears to preferentially utilize hemicellulose and amorphous cellulose region rather than cellulose embedded in the plant cell wall. To support, *R. albus* 8 utilized more hemicellosic sugars derived from the hydrolysis of AHPCS than *R. albus* 7 in the current study. Odenyo et al. reported that *R. albus* 8 preferred hemicellulose over cellulose when grown on alkaline hydrogen peroxide-treated wheat straw, whereas *R. flavefaciens* FD-1 did not show any sugar preference (78). Our conclusion is supported by the transcriptomic evidence of cells grown on different polysaccharides. More CAZyme genes, including hemicellulases as well as even cellulases, of *R.*

albus 8 responded to WAX than PASC, while those genes of *R. albus* 7 responded to more PASC than WAX. When both strains were grown on AHPCS, *R. albus* 8 down-regulated the expression of genes for sugar transporters and intracellular GH genes after early-log phase when hemicellulose in the culture started to decrease. In contrast, *R. albus* 7 exhibited a sequential expression of sugar transporters and intracellular GH genes, as preferred cellulosic sugars were released from AHPCS after removal of hemicellulose. Taken together, although *R. albus* 8 was able to utilize both hemicellulose and cellulose in the pure culture, its transcriptomic system appears to target hemicellulose rather than cellulose, while the transcriptomic system of *R. albus* 7 responds to hemicellulose until the cells gain access to cellulose embedded in the plant cell wall.

Despite a similar genotype for the fiber utilization, how do both strains run two distinct strategies? Based on the preferred planktonic growth of *R. albus* 8 on AHPCS as opposed to the substrate adherent growth of *R. albus* 7, we focused on two regulation mechanisms which have been intensively studied in biofilm forming bacteria: the accessory gene regulator (Agr) quorum sensing system and the cyclic-di-GMP (c-di-GMP) regulation system. The Agr quorum sensing (QS) system has been implicated in biofilm-associated function of gram-positive bacteria (79). This system is controlled by an operon composed of *agrABCD* genes. AgrD peptide is processed and secreted by AgrB. These autoinducing peptides are recognized by AgrC and AgrC activates the transcriptional factor (AgrA). Subsequently, AgrA binds to its target promoters. Agr QS is known to control the swich between biofilm formation and virulence behaviors (e.g. toxin secretion and motility). In *Staphylococcus* species, the high level of autoinducing peptides (AgrD) inhibits biofilm formation and stimulates virulence genes, resulting in dispersal of cells from the biofilm (79). Another regulatory system involved in the biofilm formation is C-di-GMP

regulation. C-di-GMP is known as a second messenger in bacteria. C-di-GMP is produced from two molecules of GTP by diguanylate cyclases (DGCs) encoding a GGDEF domain and is broken down by phosphodiesterses (PDEs) encoding an EAL domain (59). Thus, the level of c-di-GMP in the cytoplasm is regulated by these two enzymes. Based on bioinformatical prediction, the c-di-GMP is present in 85 % of all bacteria (80). The role of c-di-GMP has been intensively studied in biofilm associated pathogenic bacteria including *Psuedomonas aeruginosa*, *Salmonella enterica*, and *Vibrio cholerae* (81). Similar with Agr QS, c-di-GMP is also known to control the switch between biofilm formation and motily. In addition, c-di-GMP also controls a broad array of other behaviors including cell cycle propagation, development, fimbrial synthesis, type III secretion, RNA modulation, stress response and virulence (82). For example, the high c-di-GMP level stimulates various biofilm-associated functions, such as the biosynthesis of adhesins and exopolysaccharide matrix in *Pseudomonas* species (59). Taken together, it seems reasonable to postulate that high c-di-GMP and low AgrD stimulate cell attachment and low c-di-GMP and high AgrD level stimulate cell detachment, which would be the potential regulators employed by *R. albus* 7 and 8 for adherence, colonization, and subsequent plant cell wall degradation. In accordance with our prediction, our transcriptional analyses showed that c-di-GMP regulatory genes had a similar transcriptional pattern with CAZyme genes in response to cellulose and hemicellulose. In addition, the Agr QS genes, especially *agrBD*, were highly expressed in the preferred planktonic strain, *R. albus* 8. Taken together, it seems that there is a strong correlation between transcriptional responses of c-di-GMP and Agr quorum sensing systems and the differential strategies employed by *R. albus* 7 and 8 for plant cell wall utilization.

With preference of hemicellulose over cellulose, *R. albus* 8 has achieved the capability of utilizing the second most abundant polysaccharide in the hemicellulose of forages, arabinan (83).

Of predominant cellulolytic ruminal species capable of degrading hemicellulose, *F. succinogenes* and some *R. flavefaciens* strains do not utilize the breakdown products of hemicellulose or utilize them with a low efficiency during growth on the plant cell wall (84–87). Taken together with the inability of *R. albus* 7 to degrade arabinan, *R. albus* 8 preferentially utilizes a broad range of hemicelluloses that are not preferred by other cellulolytic species in order to survive in the rumen.

To compensate less energy yield from pentose relative to hexose through glycolysis, *R. albus* 8 has evolved to catabolize pentose faster than other competitors using the PK pathway. We found that both PP and PK pathways were functional for hemicellulose utilization and the PK pathway became dominant during the initial growth of *R. albus* 8 on AHPCS. The distinct feature of the PK pathway is that pentose bypasses three steps of the PP pathway and two carbons of pentose to bypass the entire EMP pathway to generate one ATP. Thus, it seems feasible that catabolizing pentose through the PK pathway may provide rapid catabolic rate of pentose to *R. albus* 8. In accordance with our hypothesis, *R. albus* 8 had a higher growth rate on beechwood xylan than *R. albus* 7. This result is also supported by previous research on the PK pathway bearing microorganisms. The ruminal fungus, *Aspergillus nidulans*, is known to harbor the PK pathway (88, 89). The mutant strain that over-expressed the PK pathway enhanced a growth rate on xylose (T_d: 219 min) as compared to the wild type (T_d: 260 min), but the growth rate on glucose was not changed between the mutant and wild type strain, suggesting that the PK pathway allows to grow at faster growth rate (90). We were unable to find the putative XFP gene in the available genomes of other cellulolytic strains within *Ruminococcus and Fibrobacter* species. Notably, we found the orthologs in the hemicellulolytic *Firmicutes* in the rumen and human gut, including *Butyrivibrio hungatei*, *Butyrivibrio fibrisolvens*, and *Roseburia intestinalis* (A.A. sequence similarity > 76 %; Table A.19). In accordance with our results, the maximum

growth rate of *B. fibriosolvens* was higher on xylose (T_d: 92 min) than on glucose (T_d: 107 min) in the previous research (91). Therefore, the survival strategy of *R. albus* 8 using the PK pathway would be conversed in the gut Firmicutes that utilize exclusively hemicellulose for growth.

However, considering energy (e.g. ATP) loss through the PK pathway, as compared to the PP pathway, it can be argued that using the PK pathway actually provides a growth advantage to *R. albus* 8. We propose that the general mechanism of the PK pathway, relative to the PP pathway, seems to be similar with the Entner-Doudoroff (ED) pathway, relative to the EMP pathway. In the ED pathway, glucose is converted into 2-keto-3-deoxy-6-phosphogluconate (KDPG) and subsequently, KDPG is cleaved into G-3-P and pyruvate (92). Similar with the PK pathway, one molecule of glucose bypasses a conversion step of dihydroxyacetone phosphate (DHAP) to G-3-P and half of the carbons of glucose bypass the lower EMP pathway (from G-3-P to pyruvate). As a result, the ED pathway generates one less ATP per glucose than the EMP pathway. However, Flamholz et al. suggested that this ATP loss could be compensated by saving the cost for enzyme synthesis required for the lower EMP pathway (93). Through both thermodynamic and kinetic analysis, they predicted that the EMP pathway needs to synthesize several fold more enzymes than the ED pathway to catabolize the same amount of glucose at the same rate. From this perspective, the PK pathway likely compensates the ATP loss by saving the cost for glycolytic protein synthesis. Furthermore, the PK pathway generates fewer reducing equivalents (e.g. NADH) than the PP pathway due to less oxidation of carbon to CO_2 (57, 58). As a result, demands for re-oxidation of electron carriers through ethanol and hydrogen fermentation would be decreased, which enables *R. albus* 8 to change the metabolic flux of acetyl-CoA towards acetate fermentation (1 ATP produced) and

save more cost for the production of enzymes, including electron-bifurcating hydrogenase complex (HydABC) and acetaldehyde/ethanol dehydrogenases.

R. albus 8 appears to rapidly metabolize hemicellulose and then dissociate from plant cell wall to find another hemicellulose. In the current study, we observed an intriguing phenomenon during growth of *R. albus* 8 on insoluble AHPCS. After log phase, *R. albus* 8 increased planktonic cells, while *R. albus* 7 was still attached to the substrate, which is in accordance with the expression pattern of Agr QS genes. Through sugar analysis, we found that after late-log phase, the hemicellulosic sugars in the residue of AHPCS and in the culture medium became depleted. *R. albus* 8 may sense the level of hemicellulose in the plant cell wall and once depleted, the cells repress the expression of genes required for substrate attachment and utilization, which accounts for the increase in planktonic cells. The resulting planktonic cells are able to find new substrate and rapidly metabolize intact hemicellulose. Therefore, *R. albus* 8 may serve as a helper for deconstruction of outer membrane of plant cell wall and consequently, cooperate with the cellulolytic members of the community, such as *R. albus* 7, *R. flavefaciens*, and *F. succinogenes* (Fig. 2.23).

2.5 Conclusion

In the current study, we demonstrate the phenotypic differences for plant cell wall utilization at the strain level. Through the comparative genomic analyses and biochemical enzyme analysis, we found that both strains share a similar mechanism for adherence and fiber deconstruction, which can be a representative fibrolytic system of *R. albus* species. Nevertheless, they displayed distinctly different transcriptional profiles of CAZymes and other metabolic pathways when

grown on polysaccharides (cellulose, hemicellulose, and plant biomass). Our findings strongly indicate that the cellular strategy for plant cell wall utilization varies at strain level of *R. albus* and both strains can be non-competitive, but likely cooperative. Therefore, they are functionally different and non-redundant in the gut ecosystem. This is a strong case for bacterial specialization and niche differentiation that cautions against interpretation of rumen and other gut system at the population and metagenomics level.

2.6 References

1. **Nicholson J, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Petttersen S**. 2012. Host-gut microbiota metabolic interactions **1262**.

2. **Thomas L V, Ockhuizen T, Suzuki K**. 2014. Exploring the influence of the gut microbiota and probiotics on health: a symposium report. Br J Nutr **112 Suppl** :S1–S18.

3. **Hooper L V, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI**. 2001. Molecular analysis of commensal host-microbial relationships in the intestine. Science **291**:881–884.

4. **Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA**. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol **6**:121–131.

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

6. **Flint HJ, Bayer EA.** 2008. Plant cell wall breakdown by anaerobic microorganisms from the mammalian digestive tract. Ann NY Acad Sci **1125**:280–288.

7. **Salyers AA, Vercellotti JR, West SEH, Wilkins TD**. 1977. Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. Appl Environ Microbiol **33**:319–322.

8. **Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, Buhler JD, Gordon JI**. 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science **307**:1955–1959.

9. **Martens EC, Koropatkin NM, Smith TJ, Gordon JI**. 2009. Complex glycan catabolism by the human gut microbiota: The bacteroidetes sus-like paradigm. J Biol Chem **284**:24673–24677.

10. **Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI**. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol **9**.

11. **Dodd D, Mackie RI, Cann IKO**. 2011. Xylan degradation, a metabolic property shared by rumen and human colonic Bacteroidetes. Mol Microbiol **79**:292–304.

12. **Koropatkin NM, Cameron EA, Martens EC**. 2012. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol **10**:323–335.

13. **Larue R, Yu Z, Parisi VA, Egan AR, Morrison M**. 2005. Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. Environ Microbiol **7**:530–543.

14. **McWilliam Leitch EC, Walker AW, Duncan SH, Holtrop G, Flint HJ**. 2007. Selective colonization of insoluble substrates by human faecal bacteria. Environ Microbiol **9**:667–679.

15. **Walker AW, Duncan SH, Harmsen HJM, Holtrop G, Welling GW, Flint HJ**. 2008. The species composition of the human intestinal microbiota differs between particle-associated and liquid phase communities. Environ Microbiol **10**:3275–3283.

16. **Bayer E, Shoham Y, Lamed R**. 2013. Lignocellulose-decomposing bacteria and their enzyme systems, p. 215–266. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, Fourth (eds.), The Prokaryotes. Springer Berlin Heidelberg, Berlin.

17. **White BA, Lamed R, Bayer EA, Flint HJ**. 2014. Biomass utilization by gut microbiomes. Annu Rev Microbiol 279–296.

18. **Doi RH, Kosugi A**. 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol **2**:541–551.

19. **Bayer EA, Belaich J-P, Shoham Y, Lamed R**. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol **58**:521–554.

20. **Chassard C, Delmas E, Robert C, Lawson PA, Bernalier-Donadille A**. 2011. *Ruminococcus champanellensis* sp. nov., a cellulose-degrading bacterium from human gut microbiota. Int J Syst Evol Microbiol **62**:138–143.

21. **Ben David Y, Dassa B, Borovok I, Lamed R, Koropatkin NM, Martens EC, White BA, Bernalier-Donadille A, Duncan SH, Flint HJ, Bayer EA, Moraïs S**. 2015. Ruminococcal cellulosome systems from rumen to human. Environ Microbiol.

22. **Christopherson MR, Dawson JA, Stevenson DM, Cunningham AC, Bramhacharya S, Weimer PJ, Kendziorski C, Suen G**. 2014. Unique aspects of fiber degradation by the ruminal ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. BMC Genomics **15**:1066.

23. **Dassa B, Borovok I, Ruimy-Israeli V, Lamed R, Flint HJ, Duncan SH, Henrissat B, Coutinho P, Morrison M, Mosoni P, Yeoman CJ, White BA, Bayer EA**. 2014. Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. PLoS One **9**:e99221.

24. **Ezer A, Matalon E, Jindou S, Borovok I, Atamna N, Yu Z, Morrison M, Bayer EA, Lamed R**. 2008. Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. J Bacteriol **190**:8220–8222.

25. **Pegden RS, Larson MA, Grant RJ, Morrison M**. 1998. Adherence of the gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. J Bacteriol **180**:5921–5927.

26. **Morrison M, Miron J**. 2000. Adhesion to cellulose by *Ruminococcus albus*: A combination of cellulosomes and Pil-proteins? FEMS Microbiol Lett **185**:109–115.

27. **Xu Q, Morrison M, Nelson KE, Bayer EA, Atamna N, Lamed R**. 2004. A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. FEBS Lett **566**:11–6.

28. **Weimer PJ, Price NPJ, Kroukamp O, Joubert LM, Wolfaardt GM, Van Zyl WH**. 2006. Studies of the extracellular glycocalyx of the anaerobic cellulolytic bacterium *Ruminococcus albus* 7. Appl Environ Microbiol **72**:7559–7566.

29. **Hamura K, Saburi W, Abe S, Morimoto N, Taguchi H, Mori H, Matsui H**. 2012. Enzymatic characteristics of cellobiose phosphorylase from *Ruminococcus albus* NE1 and kinetic mechanism of unusual substrate inhibition in reverse phosphorolysis. Biosci Biotechnol Biochem **76**:812–8.

30. **Iakiviak M, Mackie RI, Cann IKO**. 2011. Functional analyses of multiple lichenin-degrading enzymes from the rumen bacterium *Ruminococcus albus* 8. Appl Environ Microbiol **77**:7541–50.

31. **Ito S, Hamada S, Yamaguchi K, Umene S, Ito H, Matsui H, Ozawa T, Taguchi H, Watanabe J, Wasaki J, Ito S**. 2007. Cloning and sequencing of the cellobiose 2-epimerase gene from an obligatory anaerobe, *Ruminococcus albus*. Biochem Biophys Res Commun **360**:640–5.

32. **Karita S, Morioka K, Kajino T, Sakka K, Shimada K, Ohmiya K**. 1993. Cloning and sequencing of a novel endo-1,4-β-glucanase gene from *Ruminococcus albus*. J Ferment Bioeng **76**:439–444.

33. **Kawahara R, Saburi W, Odaka R, Taguchi H, Ito S, Mori H, Matsui H**. 2012. Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase, β-1,4-mannooligosaccharide phosphorylase. J Biol Chem **287**:42389–99.

34. **Moon YH, Iakiviak M, Bauer S, Mackie RI, Cann IKO**. 2011. Biochemical analyses of multiple endoxylanases from the rumen bacterium *Ruminococcus albus* 8 and their synergistic activities with accessory hemicellulose-degrading enzymes. Appl Environ Microbiol **77**:5157–69.

35. **Nakamura M, Nagamine T, Takenaka A, Aminov RI, Ogata K, Tajima K, Matsui H, Benno Y, Itabashi H**. 2002. Molecular cloning, nucleotide sequence and characteristics of a xylanase gene (*xynA*) from *Ruminococcus albus* 7. Anim Sci J **73**:347–352.

36. **Ohara H, Noguchi J, Karita S, Kimura T, Sakka K, Ohmiya K**. 2000. Sequence of egV and properties of EgV, a *Ruminococcus albus* endoglucanase containing a dockerin domain. Biosci Biotechnol Biochem **64**:80–8.

37. **Ohmiya K, Kajino T, Kato A, Shimizu S**. 1989. Structure of a *Ruminococcus albus* endo-1,4-beta-glucanase gene. J Bacteriol **171**:6771–6775.

38. **Poole DM, Hazlewood GP, Laurie JI, Barker PJ, Gilbert HJ**. 1990. Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes celA and celB. Mol Gen Genet **223**:217–223.

39. **Sawano T, Saburi W, Hamura K, Matsui H, Mori H**. 2013. Characterization of *Ruminococcus albus* cellodextrin phosphorylase and identification of a key phenylalanine residue for acceptor specificity and affinity to the phosphate group. FEBS J **280**:4463–73.

40. **Taguchi H, Hagiwara D, Genma T, Karita S, Kimura T, Sakka K, Ohmiya K**. 2004. Cloning of the *Ruminococcus albus* cel5D and cel9A genes encoding dockerin module-containing endoglucanases and expression of cel5D in *Escherichia coli*. Biosci Biotechnol Biochem **68**:1557–64.

41. **Takano M, Moriyama R, Ohmiya K**. 1992. Structure of a β-glucosidase gene from *Ruminococcus albus* and properties of the translated product. J Ferment Bioeng **73**:79–88.

42. **Krause DO, Bunch RJ, Smith WJM, McSweeney CS**. 1999. Diversity of *Ruminococcus* strains: A survey of genetic polymorphisms and plant digestibility. J Appl Microbiol **86**:487–495.

43. **Hungate RE**. 1957. Microorganisms in the rumen of cattle fed a constant ration. Can J Microbiol **3**:289–311.

44. **Russell JB, Muck RE, Weimer PJ**. 2009. Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. FEMS Microbiol Ecol **67**:183–97.

45. **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B**. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res **42**:490–495.

46. **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**:W445–51.

47. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M**. 2014. Pfam: the protein families database. Nucleic Acids Res **42**:D222–30.

48. **Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH**. 2013. CDD: Conserved domains and protein three-dimensional structure. Nucleic Acids Res **41**:D348–352.

49. **Petersen TN, Brunak S, von Heijne G, Nielsen H**. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **8**:785–6.

50. **Kim JN, Henriksen ED, Cann IKO, Mackie RI**. 2014. Nitrogen utilization and metabolism in *Ruminococcus albus* 8. Appl Environ Microbiol **80**:3095–3102.

51. **Wood TM**. 1971. The cellulase of *Fusarium solani*. Purification and specificity of the -(1-4)-glucanase and the -D-glucosidase components. Biochem J **121**:353–362.

52. **Kerley MS, Fahey GC, Berger LL, Gould JM, Lee Baker F**. 1985. Alkaline hydrogen peroxide treatment unlocks energy in agricultural by-products. Science **230**:820–822.

53. **Bradford MM**. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72**:248–254.

54. **Ibáñez AB, Bauer S**. 2014. Downscaled method using glass microfiber filters for the determination of Klason lignin and structural carbohydrates. Biomass and Bioenergy **68**:75–81.

55. **Baggerly KA, Deng L, Morris JS, Aldaz CM**. 2003. Differential expression in SAGE: Accounting for normal between-library variation. Bioinformatics **19**:1477–1483.

56. **Lever M**. 1972. A new reaction for colorimetric determination of carbohydrates. Anal Biochem **47**:273–279.

57. **Servinsky MD, Germane KL, Liu S, Kiel JT, Clark AM, Shankar J, Sund CJ**. 2012. Arabinose is metabolized via a phosphoketolase pathway in *Clostridium acetobutylicum* ATCC 824. J Ind Microbiol Biotechnol **39**:1859–1867.

58. **Liu L, Zhang L, Tang W, Gu Y, Hua Q, Yang S, Jiang W, Yang C**. 2012. Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by 13C metabolic flux analysis. J Bacteriol **194**:5413–5422.

59. **Boyd CD, O'Toole GA.** 2012. Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. Annu Rev Cell Dev Biol **28**:439–462.

60. **Van Soest PJ**. 1994. Nutritional ecology of the ruminant, 2nd ed. Cornell University Press, Ithaca.

61. **Vodovnik M, Duncan SH, Reid MD, Cantlay L, Turner K, Parkhill J, Lamed R, Yeoman CJ, Miller MEB, White BA, Bayer EA., Marinšek-Logar R, Flint HJ**. 2013. Expression of cellulosome components and type IV pili within the extracellular proteome of *Ruminococcus flavefaciens* 007. PLoS One **8**:1–11.

62. **Berg Miller ME, Antonopoulos DA, Rincon MT, Band M, Bari A, Akraiko T, Hernandez A, Thimmapuram J, Henrissat B, Coutinho PM, Borovok I, Jindou S, Lamed R, Flint HJ, Bayer EA, White BA**. 2009. Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. PLoS One **4**:e6650.

63. **Thurston B, Dawson KA, Strobel HJ**. 1994. Pentose utilization by the ruminal bacterium *Ruminococcus albus*. Appl Environ Microbiol **60**:1087–1092.

64. **Dai X, Tian Y, Li J, Luo Y, Liu D, Zheng H, Wang J, Dong Z, Hu S, Huang L**. 2015. Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. Appl Environ Microbiol **81**:1375–86.

65. **Rakotoarivonina H, Jubelin G, Hebraud M, Gaillard-Martinie B, Forano E, Mosoni P**. 2002. Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. Microbiology **148**:1871–80.

66. **Rakotoarivonina H, Larson MA, Morrison M, Girardeau JP, Gaillard-Martinie B, Forano E, Mosoni P**. 2005. The *Ruminococcus albus* pilA1-pilA2 locus: Expression and putative role of two adjacent pil genes in pilus formation and bacterial adhesion to cellulose. Microbiology **151**:1291–1299.

67. **Gerke C, Kraft A, Sussmuth R, Schweitzer O, Gotz F**. 1998. Characterization of the N-acetylglucosaminyltransferase activity involved in the biosynthesis of the *Staphylococcus epidermidis* polysaccharide intercellular adhesin. J Biol Chem **273**:18586–18593.

68. **Kaplan JB, Velliyagounder K, Ragunath C, Rohde H, Mack D, Knobloch JKM, Ramasubbu N**. 2004. Genes involved in the synthesis and degradation of matrix polysaccharide in *Actinobacillus actinomycetemcomitans* and *Actinobacillus pleuropneumoniae* biofilms. J Bacteriol **186**:8213–8220.

69. **Izano EA, Sadovskaya I, Vinogradov E, Mulks MH, Velliyagounder K, Ragunath C, Kher WB, Ramasubbu N, Jabbouri S, Perry MB, Kaplan JB**. 2007. Poly-N-acetylglucosamine mediates biofilm formation and antibiotic resistance in *Actinobacillus pleuropneumoniae*. Microb Pathog **43**:1–9.

70. **Rehm BHA**. 2010. Bacterial polymers: biosynthesis, modifications and applications. Nat Rev Microbiol **8**:578–592.

71. **Hungate RE**. 1950. the Anaerobic Mesophilic Cellulolytic Bacteria. Bacteriol Rev **14**:1–49.

72. **Ljungdahl LG, Pettersson B, Eriksson KE, Wiegel J**. 1983. A yellow affinity substance involved in the cellulolytic system of *Clostridium thermocellum*. Curr Microbiol **9**:195–199.

73. **Lamed R, Kenig R, Setter E, Bayer EA.** 1985. Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. Enzyme Microb Technol **7**:37–41.

74. **Ljungdahl LG, Coughlan MP, Mayer F, Mori Y, Hon-nami H, Hon-nami K**. 1988. Macrocellulase complexes and yellow affinity substance from *Clostidium thermocellum*., p. 483–500. *In* Wood, WA, Kellogg, ST (eds.), Methods in Enzymology. Part B, Biomass: Cellulose and Hemicellulose. Academic Press, New York.

75. **Kopecný J, Hodrová B**. 1997. The effect of yellow affinity substance on cellulases of *Ruminococcus flavefaciens*. Lett Appl Microbiol **25**:191–196.

76. **Hodrová B, Kopečný J, Káš J**. 1998. Cellulolytic enzymes of rumen anaerobic fungi *Orpinomyces joyonii* and *Caecomyces communis*. Res Microbiol **149**:417–427.

77. **Rozman Grinberg I, Yin G, Borovok I, Berg Miller ME, Yeoman CJ, Dassa B, Yu Z, Mizrahi I, Flint HJ, Bayer EA, White BA, Lamed R**. 2015. Functional phylotyping approach for assessing intraspecific diversity of *Ruminococcus albus* within the rumen microbiome. FEMS Microbiol Lett **362**:1–10.

78. **Odenyo AA, Mackie RI, Fahey GC, White BA**. 1991. Degradation of wheat straw and alkaline hydrogen peroxide-treated wheat straw by *Ruminococcus albus* 8 and *Ruminococcus flavefaciens* FD-1. J Anim Sci **69**:819–826.

79. **Rutherford ST, Bassler BL**. 2012. Bacterial quorum sensing: its role in virulence and possibilities for its control. Cold Spring Harb Perspect Med **2**:1–25.

80. **Galperin MY**. 2004. Bacterial signal transduction network in a genomic perspective. Environ Microbiol.

81. **Hengge R**. 2009. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol **7**:263–273.

82. **Srivastava D, Waters CM**. 2012. A tangled web: regulatory connections between quorum sensing and cyclic Di-GMP. J Bacteriol **194**:4485–93.

83. **Aman P**. 1993. Composition and structure of cell wall polysaccharides in forages., p. 183–200. *In* Jung, HG, Buxton, DR, Hatfield, RD, Ralph, J (eds.), Forage cell wall structure and digestibility. American Society of Agronomy-Crop Science Society of America-Soil Science Society of America, Madison.

84. **Dehority BA, Scott HW**. 1967. Extent of cellulose and hemicellulose digestion in various forages by pure cultures of rumen bacteria1. J Dairy Sci **50**:1136–1141.

85. **Matte A, Forsberg CW, Verrinder Gibbins AM**. 1992. Enzymes associated with metabolism of xylose and other pentoses by *Prevotella* (*Bacteroides*) *ruminicola* strains, *Selenomonas ruminantium* D, and *Fibrobacter succinogenes* S85. Can J Microbiol **38**:370–376.

86. **Kobayashi Y, Shinkai T, Koike S**. 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion - Review. Folia Microbiol (Praha) **53**:195–200.

87. **Hespell RB**. 1988. Microbial digestion of hemicelluloses in the rumen. Microbiol Sci **5**:362–365.

88. **Panagiotou G, Kouskoumvekaki I, Jónsdóttir SO, Olsson L**. 2007. Monitoring novel metabolic pathways using metabolomics and machine learning: Induction of the phosphoketolase pathway in *Aspergillus nidulans* cultivations. Metabolomics **3**:503–516.

89. **Thykaer J, Nielsen J**. 2007. Evidence, through C13-labelling analysis, of phosphoketolase activity in fungi. Process Biochem **42**:1050–1055.

90. **Panagiotou G, Anderson MR, Grotkjær T, Regueira TB, Hofmann G, Nielsen J, Olsson L**. 2008. Systems analysis unfolds the relationship between the phosphoketolase pathway and growth in Aspergillus nidulans. PLoS One **3**.

91. **Russell JB, Baldwin RL**. 1978. Substrate preferences in rumen bacteria: evidence of catabolite regulatory mechanisms. Appl Environ Microbiol **36**:319–329.

92. **Kim BH, Gadd GM**. 2008. Bacterial Physiology and Metabolism. Cambridge University Press, Cambridge.

93. **Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R**. 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci U S A **110**:10039–10044.

94. **Grimmler C, Held C, Liebl W, Ehrenreich A**. 2010. Transcriptional analysis of catabolite repression in *Clostridium acetobutylicum* growing on mixtures of d-glucose and d-xylose. J Biotechnol **150**:315–323.

95. **Ibáñez AB, Bauer S**. 2014. Downscaled method using glass microfiber filters for the determination of Klason lignin and structural carbohydrates. Biomass and Bioenergy **68**:75–81.

96. **Zheng Y, Kahnt J, Kwon IH, Mackie RI, Thauer RK**. 2014. Hydrogen formation and its regulation in *Ruminococcus albus*: involvement of an electron-bifurcating [FeFe]-hydrogenase, of a non-electron-bifurcating [FeFe]-hydrogenase, and of a putative hydrogen-sensing [FeFe]-hydrogenase. J Bacteriol **196**:3840–52.

97. **Moon YH, Iakiviak M, Bauer S, Mackie RI, Cann IKO**. 2011. Biochemical analyses of multiple endoxylanases from the rumen bacterium *Ruminococcus albus* 8 and their synergistic activities with accessory hemicellulose-degrading enzymes. Appl Environ Microbiol **77**:5157–69.

2.7 Figures

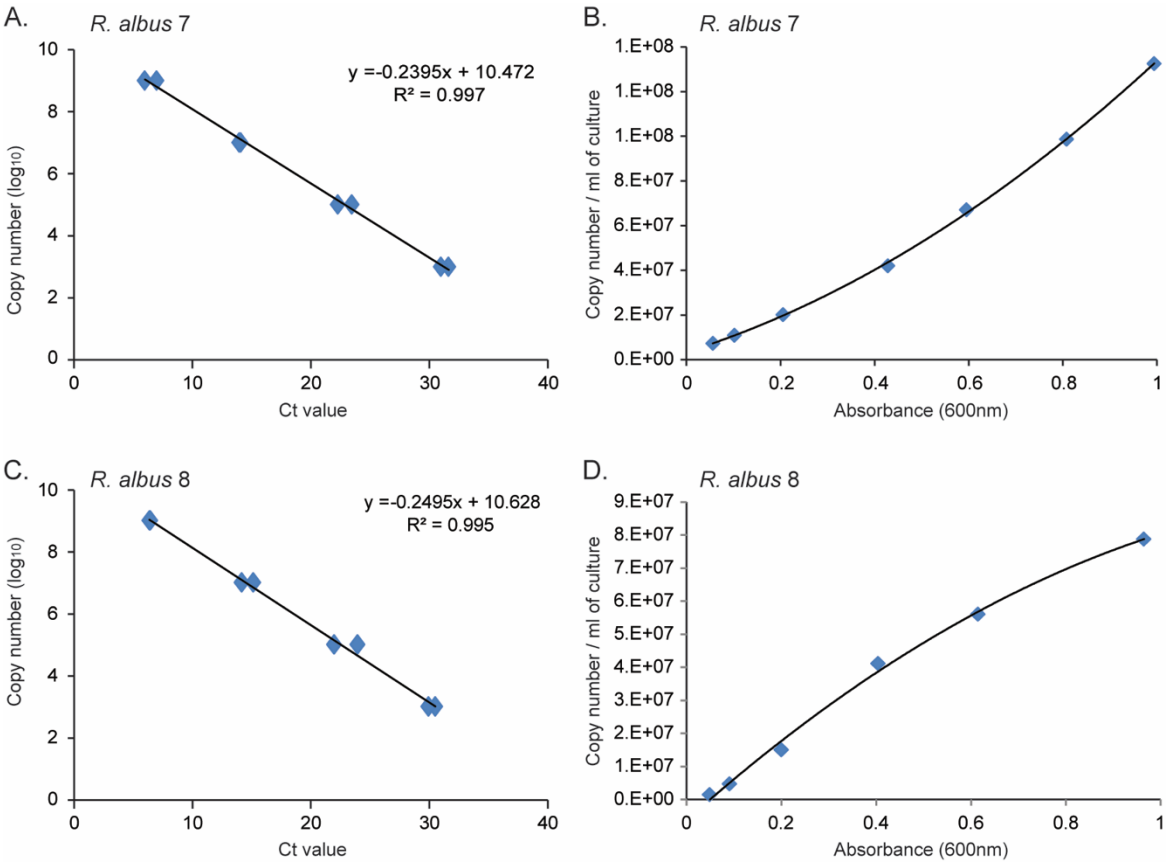


Fig. 2.1 Standard curve for the relative absorbance of each *R. albus* 7 and 8 in co-culture. To convert the copy number of each strain to the relative value, the standard curves were generated from the copy number at different absorbance values.

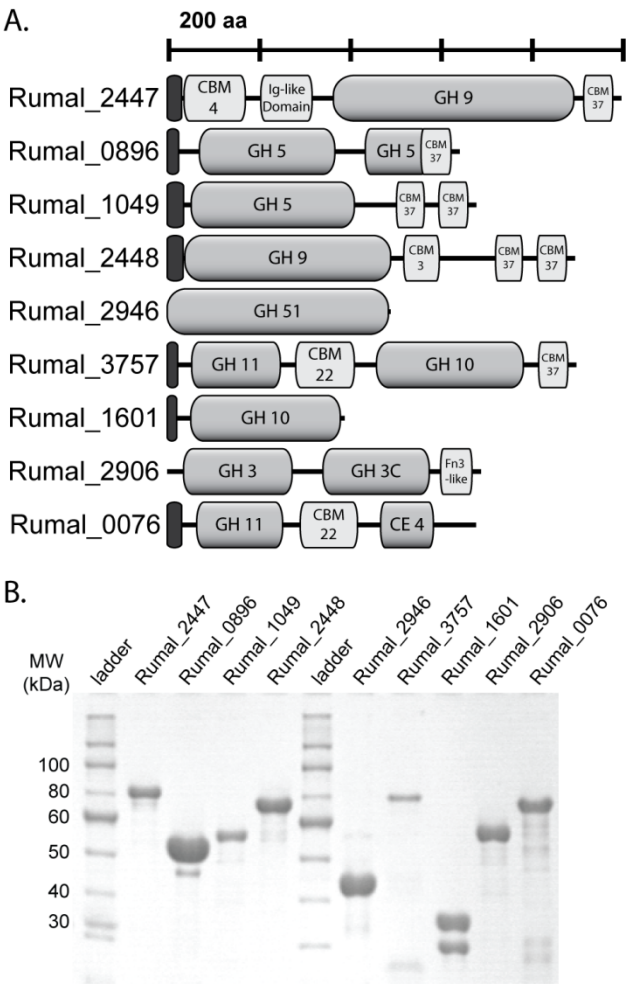


Fig. 2.2 Cloning of glycoside hydrolases from *Ruminococcus albus* 7. (*A*) Domain architecture of the cloned glycoside hydrolases from *R. albus* 7. (*B*) Purification of 9 proteins was performed by Cobalt affinity chromatography, followed by gel filtration. The highly purified proteins were analyzed by 12 % SDS-. Numbers on the left are molecular weights.

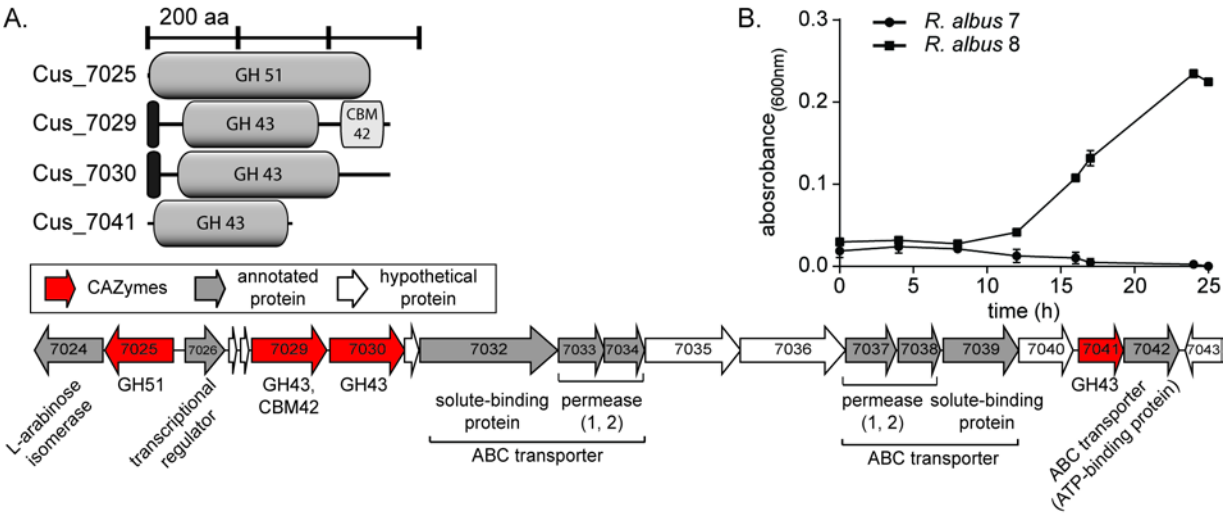


Fig. 2.3 Unique putative arabinan utilization cluster present on *R. albus* 8 genome. (*A*) Putative arabinan utilization cluster present on the *R. albus* 8 genome. Genes were assigned to CAZy families (GHs, PLs, and CEs) if they exhibited significant similarity (*E*-value < 1×10^{-5}) to biochemically characterized proteins already cataloged in a CAZy family. The dbCAN server was used to verify the CAZy annotations, and additional domains were predicted using both Pfam database and the CDD database. Signal peptides were predicted using SignalP v4.0. Arabinofuranosidase activity of Ara51A (CUS_7025) was biochemically reported by Moon and his colleagues (Moon *et al.*, 2011). Determining orthologs of the arabinan utilization cluster in *R. albus* 8 against *R. albus* 7 was conducted using reciprocal BLAST, and no orthologs of genes in the clusters (from Cus_7027 to Cus_7042) are present on *R. albus* 7 genome. (*B*) Growth curve of *R. albus* 7 and 8 on debrached arabinan. Both strain 7 and 8 were grown in a defined medium with 0.4 % (w/v) of debranched arabinan. Data are reported as means ± SD.

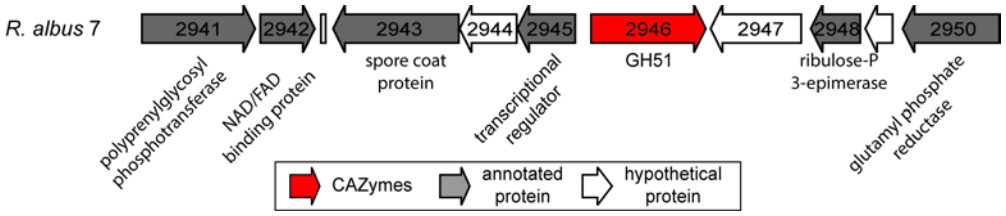


Fig. 2.4 GH51 gene cluster on the genome of *R. albus* 7. The genes are shown as locus tag number. None of genes encoding L-arabinose isomerase, sugar transporters, and GH43 domain are present on the genome of *R. albus* 7.

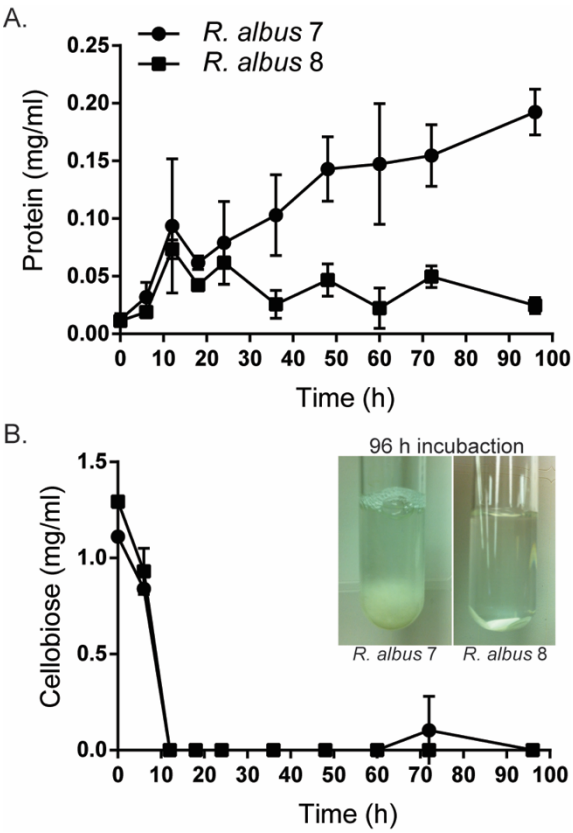


Fig. 2.5 Growth of *R. albus* 7 and 8 on filter paper. Both strains were grown in a defined medium containing 0.1 % of cellobiose and 0.4 % of filter paper as the sole carbohydrate source. (*A*) The protein quantification using the Bradford method was performed to measure the growth of both strains. (*B*) Cellobiose was used to increase cell numbers at the initial growth phase and the concentration of cellobiose in the medium was measured using HPLC. After 10 hours, the cellobiose was completely depleted in the media of both strain 7 and 8, suggesting that later growth of *R. albus* 7 resulted from utilization of filter paper. (*A* and *B*) Data are reported as means ± SD from three biological replicates.

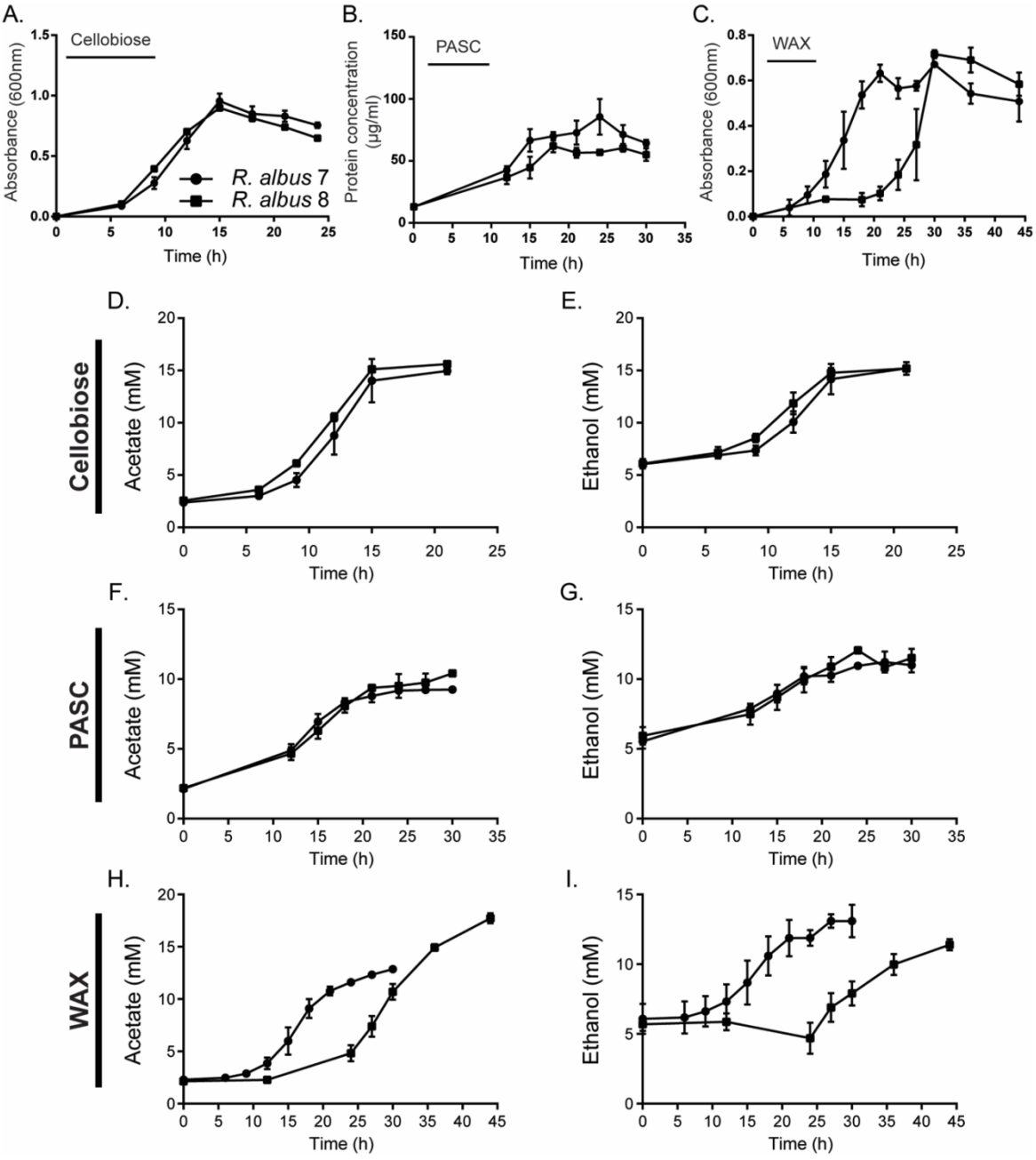


Fig. 2.6 Growth of *R. albus* 7 and 8 on cellobiose, amorphous cellulose and soluble xylan. Both strain 7 and 8 were grown in a defined medium with 0.4 % (w/v) of either cellobiose, phosphoric acid swollen cellulose (PASC) or wheat arabinoxylan (WAX). Growth curves of *R. albus* 7 and 8 were assessed by measuring absorbance at 600 nm for soluble cellobiose (*A*) and WAX (*C*) or by total protein concentration quantified using the Bradford method for insoluble PASC (*B*). The fermentation products, acetate and ethanol, were measured during growth on cellobiose (*C* and *D*), PASC (*E* and *F*), and WAX (*H* and *I*). *Error bars* represent SD for three biological replicates.

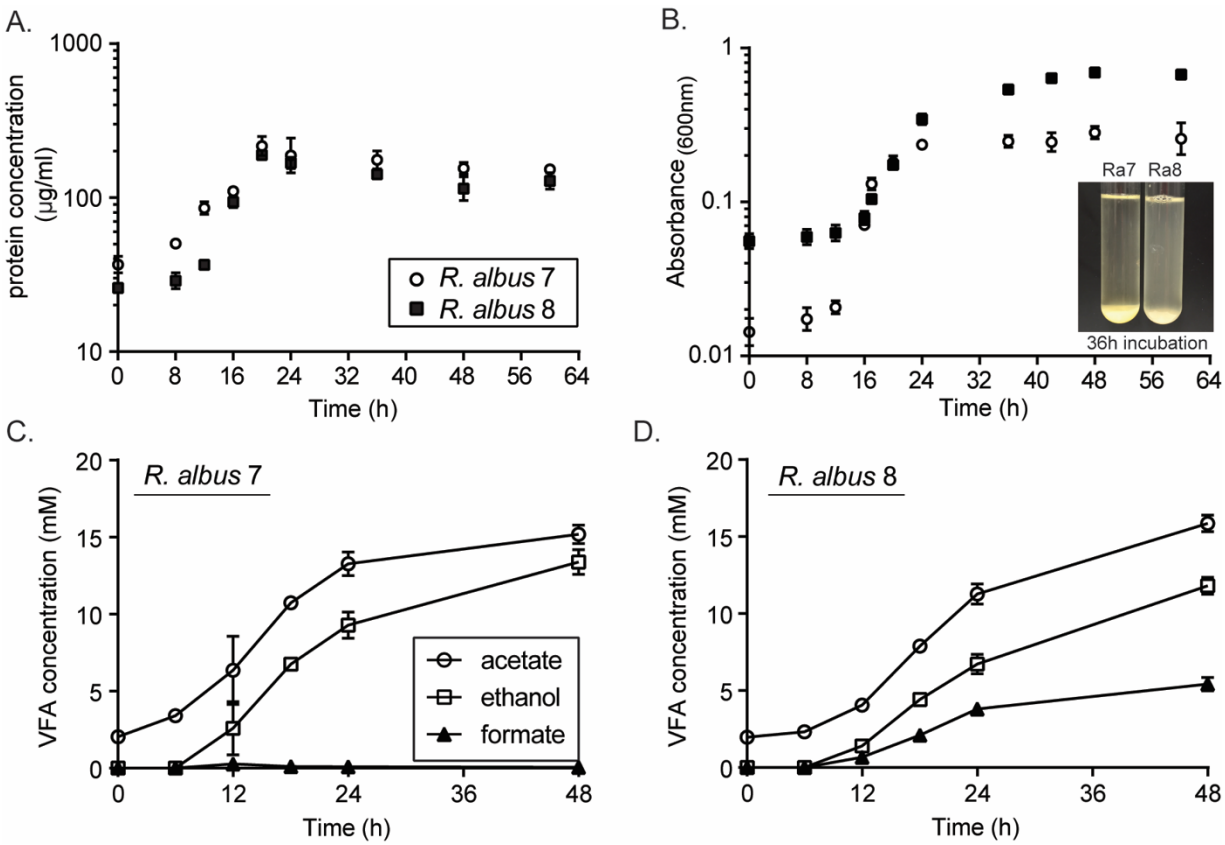


Fig. 2.7 Growth curve and VFA analysis of *R. albus* 7 and 8 on AHPCS. (*A*) Growth curves of *R. albus* 7 and 8 grown on insoluble substrate were assessed by total protein quantification using the Bradford method. (*B*) Growth curve of *R. albus* 7 and 8 on AHPCS assessed by reading the absorbance value at 600nm. The insoluble substrates completely settled down at the bottom 30 minute after inoculation. Then, the absorbance in liquid medium was measured by spectrometer. (*C* and *D*) The fermentation products, acetate, ethanol, and formate, of *R. albus* 7 and 8 were measured using HPLC. Data are reported as means ± SD from three biological replicates.

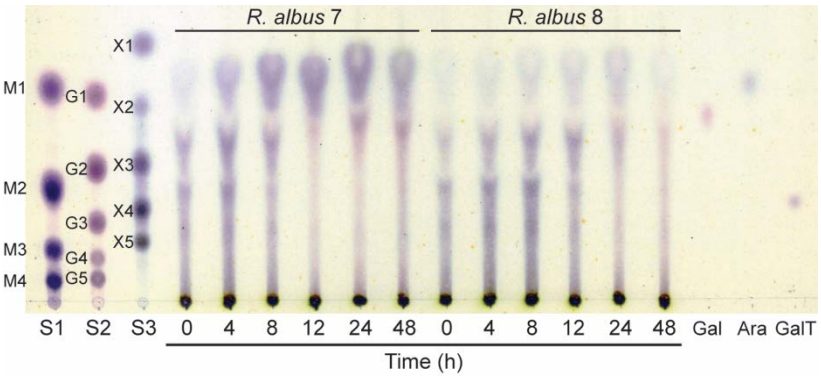


Fig. 2.8 Soluble sugars produced during growth of *R. albus* 7 and 8 on AHPCS. The soluble sugars in supernatant were resolved by thin layer chromatography, followed by staining with methanolic orcinol. Mixture of each monosaccharides (mannose: M1, glucose: G1, and xylose: X1) and their oligomers (manno-oligosaccharides: M2 to M4, cello-oligosaccharides: G2 to G5, xylo-oligosaccharides: X2 to X5), and of galactose (Gal), arabinose (Ara), and galacturonic acid (GalT) were used as standards (S1, S2, and S3).

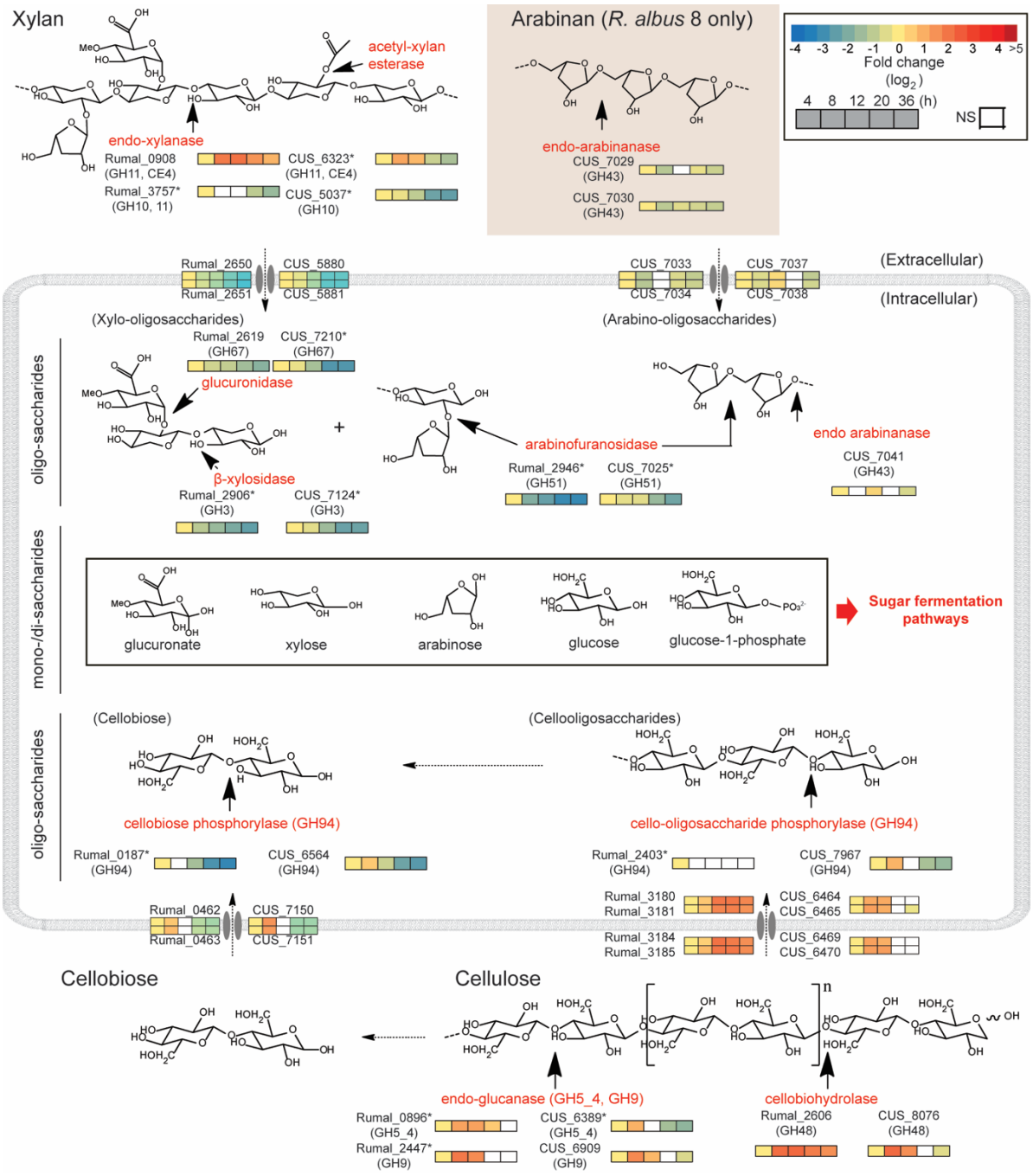


Fig. 2.9 Integrative transcriptional profiles of GH genes in *R. albus* 7 and 8 to utilize polysaccharides in plant cell wall. Schematic view of the degradation of major polysaccharide components in AHPCS by GHs is illustrated. Either putative or biochemically characterized GH genes are shown as their locus tag number (Rumal_# for strain 7 and CUS_# for strain 8, respectively). NS means the fold change in means for two independent experiments was not significantly different ($p \geq 0.05$), as compared with expression level at 4 hour. *Asterisks* denote genes for which biochemical activities have been demonstrated for their cognate gene products in previous studies or confirmed in this study.

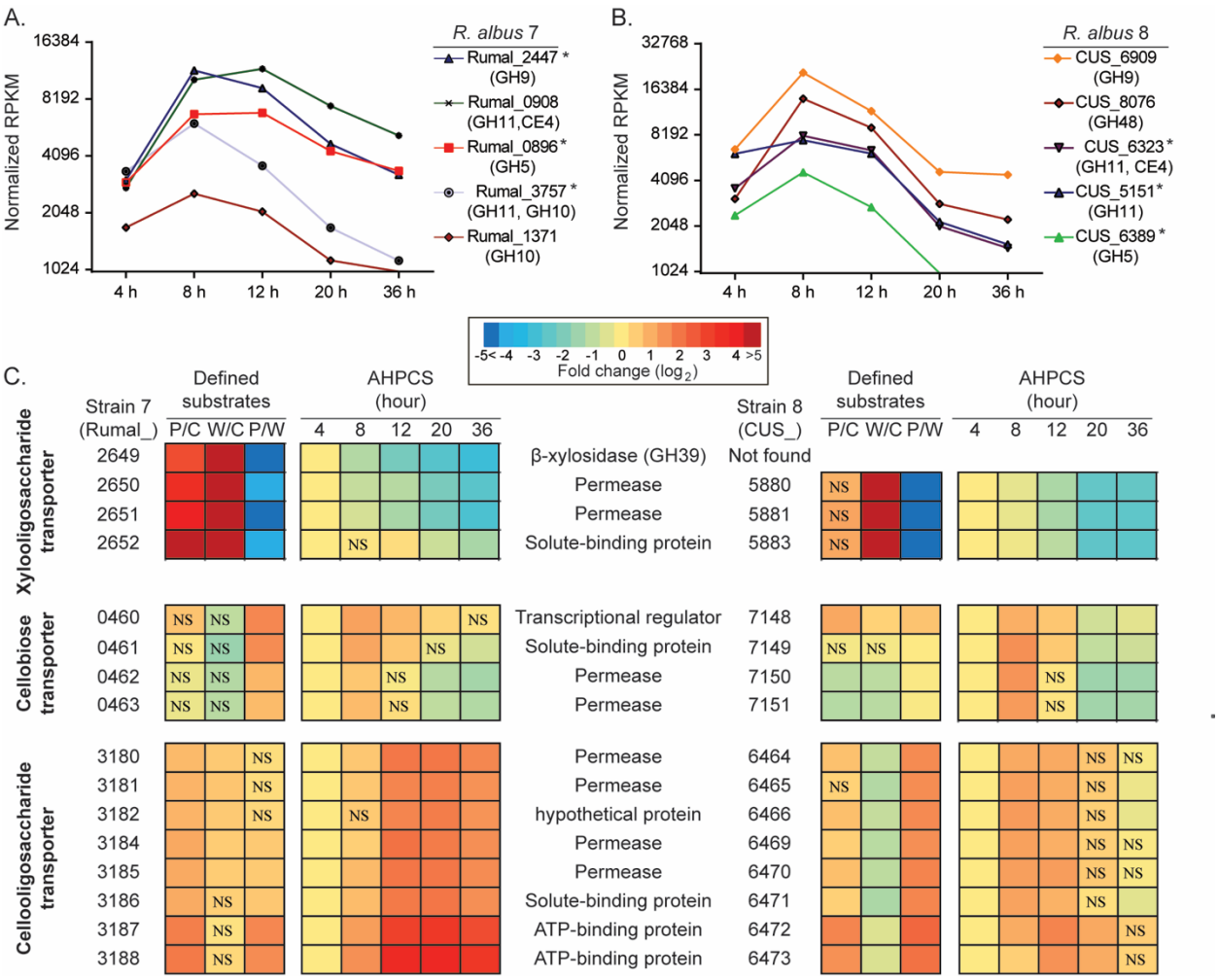


Fig. 2.10 Expression of genes encoding extracellular GHs and putative sugar transporters during growth on AHPCS. During growth on AHPCS, the top five GH genes among highly expressed GH genes (>1000 RPKM) in strain 7 (*A*) and strain 8 (*B*) were plotted. *Asterisks* denote genes for which biochemical activities have been demonstrated for their cognate gene products in previous studies or confirmed in this study.
The expression pattern of predicted sugar transporters are shown as hitmap (*C*). NS means that the fold change in means for two independent experiments was not significantly different (*p* value ≥ 0.05).

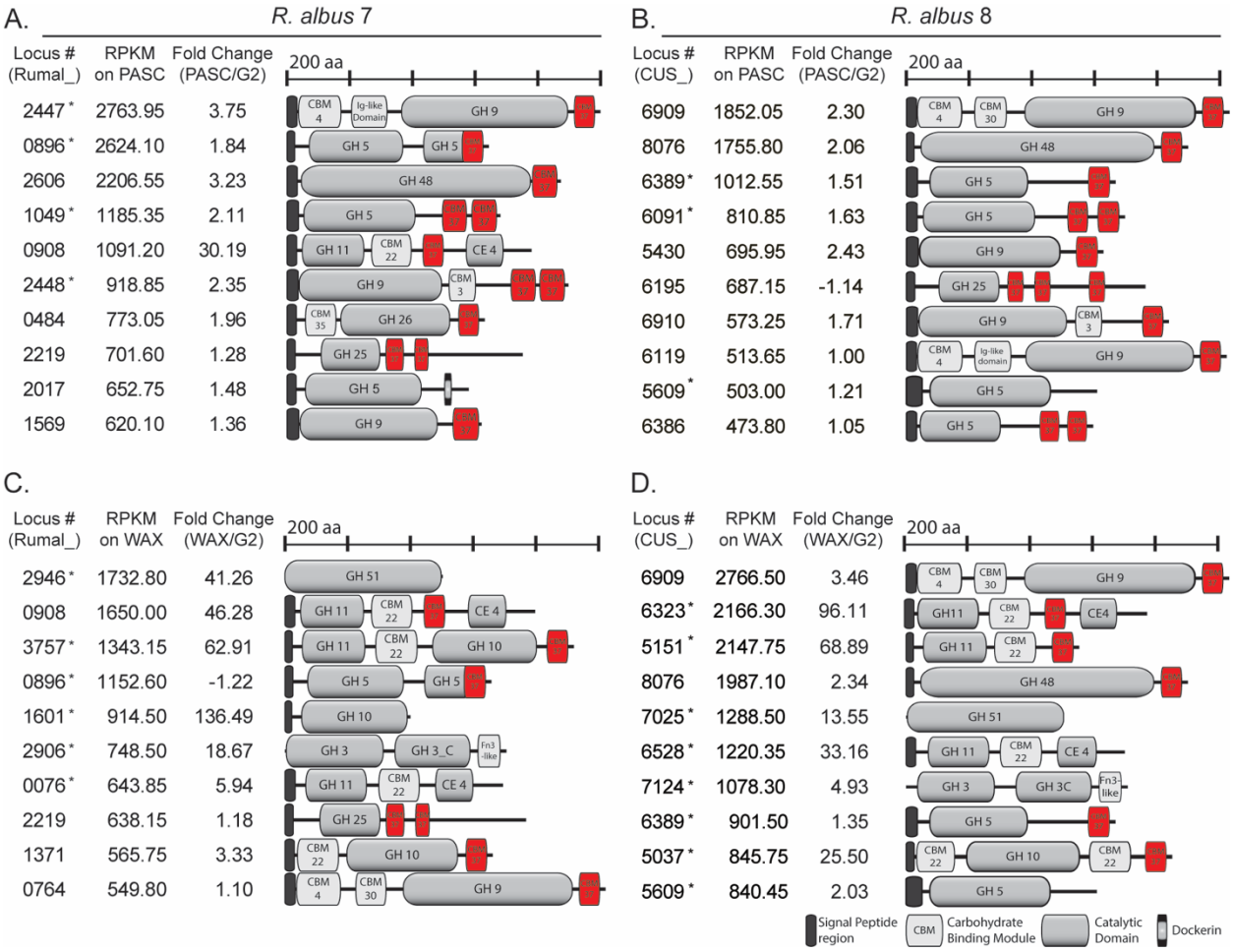


Fig. 2.11 Top ten highly expressed GHs during growth of *R. albus* 7 and 8 with PASC, or WAX. The fold changes were calculated using expression value (RPKM) on either PASC or WAX as compared with expression value on cellobiose (G2). (*A* and *B*) Genes highly expressed during growth of *R. albus* 7 with either PASC (*A*) or WAX (*B*) and listed by magnitude of expression value as RPKM. (*C* and *D*) Genes highly expressed during growth of *R. albus* 8 with either PASC (*C*) or WAX (*D*) and listed by magnitude of expression value as RPKM. *Asterisks* denote genes for which biochemical activities have been demonstrated for their cognate gene products in previous studies or for which enzymatic activities on polysaccharides were confirmed in this study.

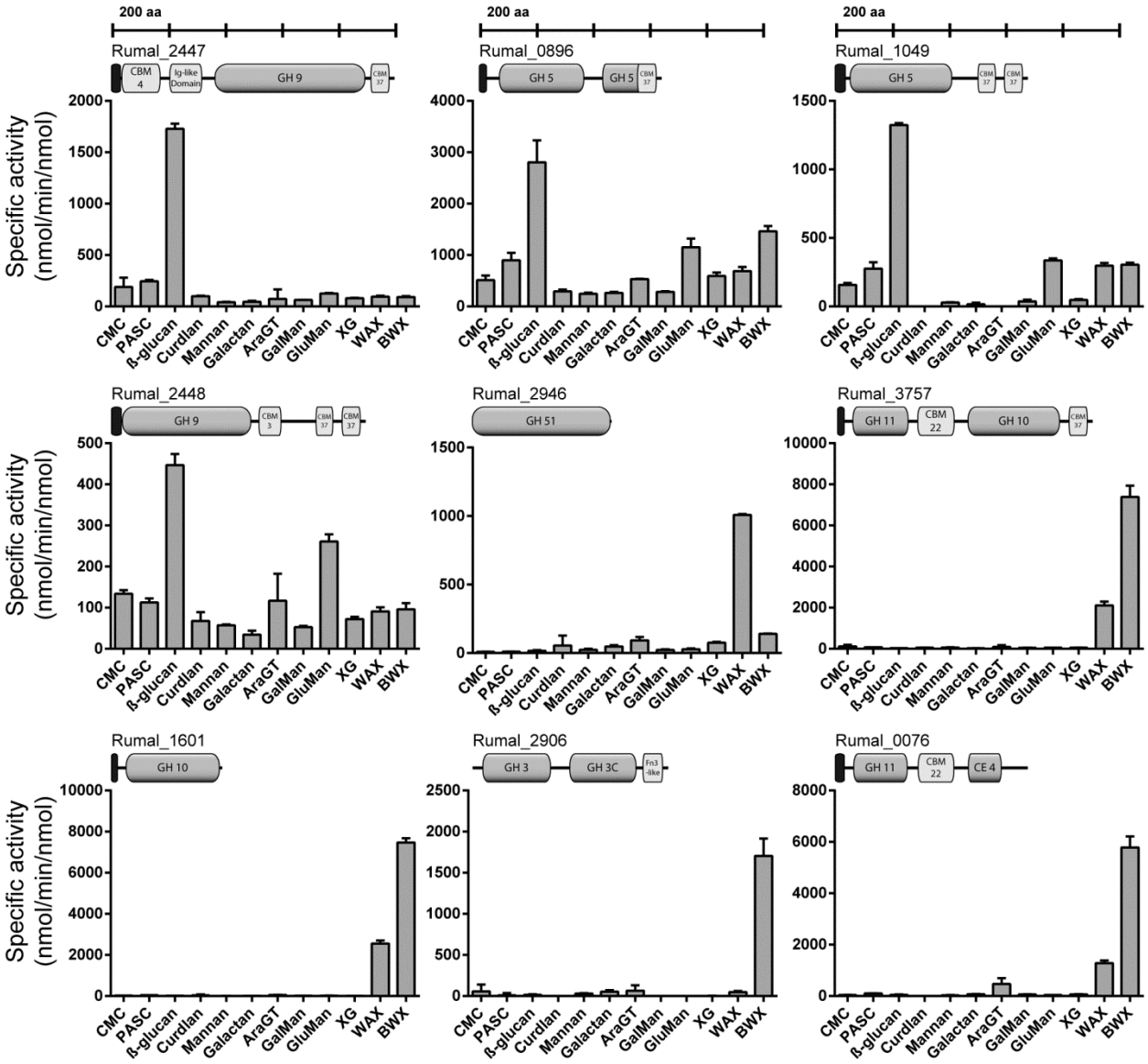


Fig. 2.12 Characterization of nine glycoside hydrolases of *Ruminococcus albus* 7. Specific activities (nmol products released per minute per nmol enzyme) with eight different soluble polysaccharide substrates, including carboxymethyl cellulose (CMC), beta-glucan, galactan, arabinogalactan (AraGT), galactomannan (GalMan), glucomannan (GluMan), xyloglucan (XG), and wheat arabinoxylan (WAX) present at 0.5% (w/v), and four different insoluble polysaccharide substrates, including phosphoric acid swollen cellulose (PASC), curdlan, mannan, and beechwood xylan (BWX), present at 2% (w/v), were reported for nine GH proteins. Reducing sugars released from polysaccharide substrate by GH protein were determined by the *para*-hydroxybenzoic acid hydrazide (PAHBAH) assay using glucose as the reducing end standard. Data are reported as means ± SD from three independent experiments.

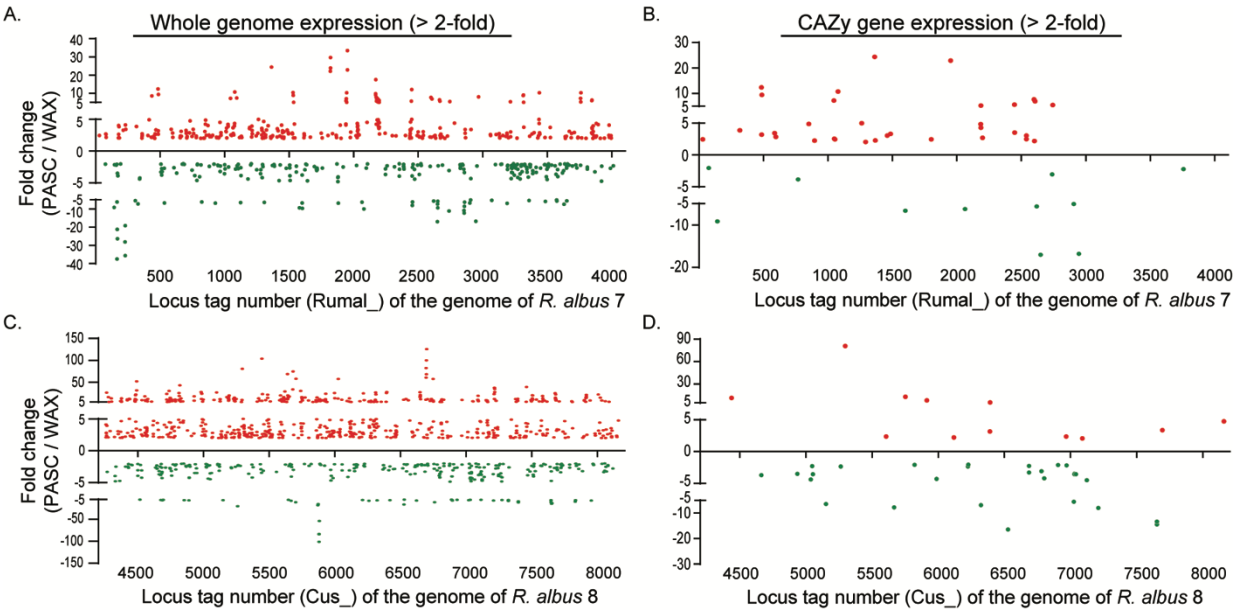


Fig. 2.13 The genes regulated by either PASC or WAX during growth of *R. albus* 7 and 8 with either PASC or WAX. Both strain 7 and 8 were grown in a defined medium containing either cellobiose, PASC or WAX. The total RNA was extracted from the cells at mid log growth phase and used for RNA-seq experiments. The fold changes were calculated using expression value (RPKM) on PASC as compared with expression value on WAX. Each *R. albus* 7 (*A* and *B*) or *R. albus* 8 (*C* and *D*) gene that was induced or repressed greater than 2 fold by PASC relative to WAX is plotted as red dot or green dot, respectively. The CAZyme encoding genes that were induced or repressed at least 2 fold by PASC relative to WAX are shown in (*B*) and (*D*).

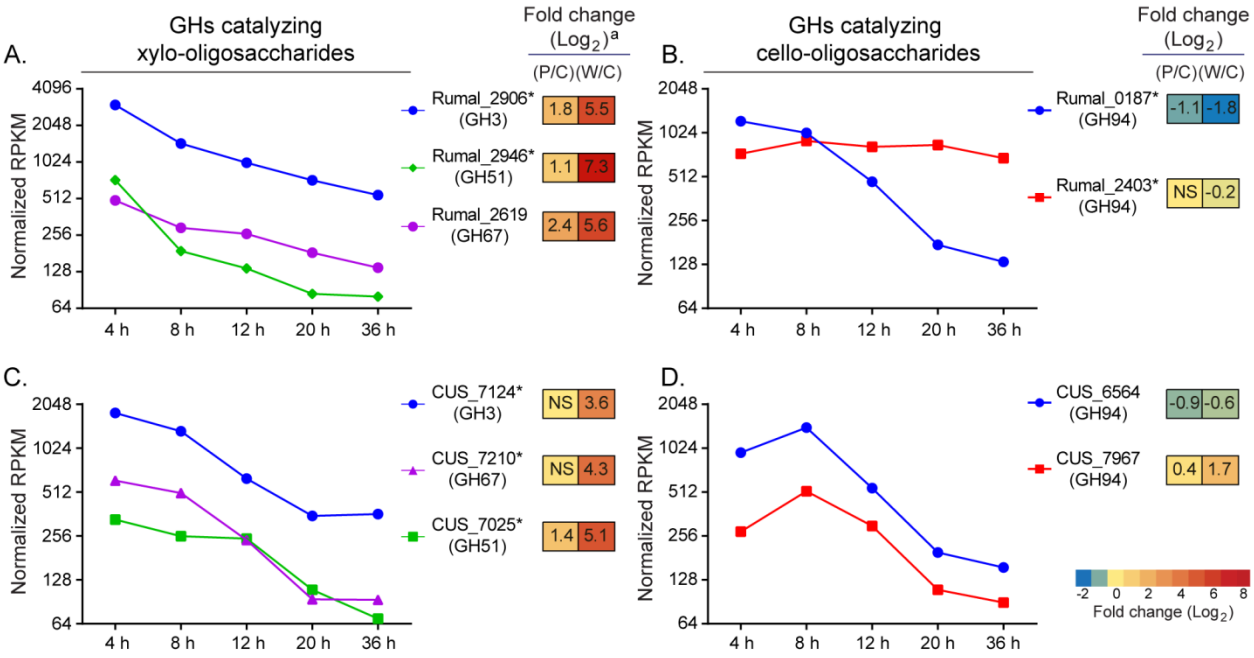


Fig. 2.14 Expression of intracellular GH genes on AHPCS. (*A* and *B*) Expression level of intracellular GH genes of *R. albus* 7. (*C* and *D*) Expression level of intracellular GH genes of *R. albus* 8. The intracellular GH genes that further process xylo-oligosaccharides and cello-oligosaccharidesinto monosaccharides were plotted with their normalized RPKM value on AHPCS and expressional fold changes when grown on defined substrate, either cellobiose (C), phosphoric acid swollen cellulose (P), or soluble wheat arabinoxylan (W).
[a] Both strain 7 and 8 were grown in define medium with either cellobiose, phosphoric acid swollen cellulose, or soluble wheat arabinoxylan as the sole carbohydrate source. RNA was then extracted at mid-log growth phase, and RNA-seq experiments were performed as described under “Experimental Procedures”. NS means that the fold change in means for two independent experiments was not significantly different (p value ≥ 0.05).
Asterisks denote genes for which biochemical activities have been demonstrated for their cognate gene products in previous studies or for which enzymatic activities on polysaccharides were confirmed in this study.

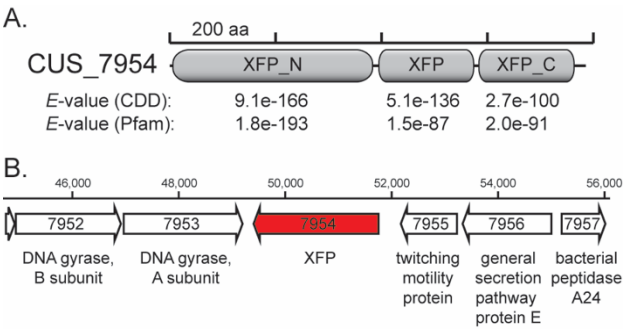


Fig. 2.15 *R. albus* 8 possesses a unique gene encoding xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (XFP). XFP domains in CUS_7954 gene were predicted using both Pfam (Finn *et al.*, 2014) and the Conserved Domain Database (CDD; Marchler-Bauer *et al.*, 2013). The gene encoding xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (xfp) is only present on the genome of *R. albus* 8 within *R. albus* species. *A.* Domain architecture of XFP gene. *B.* XFP gene locus on the genome of *R. albus* 8 shown as locus tag number (CUS_#).

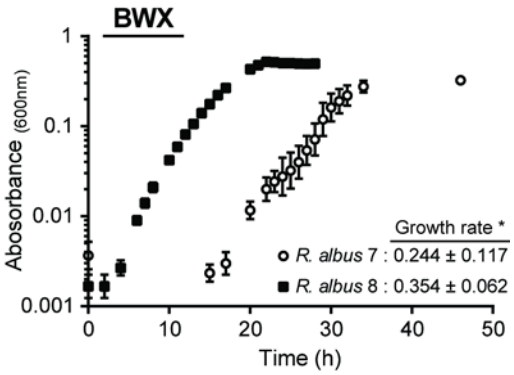
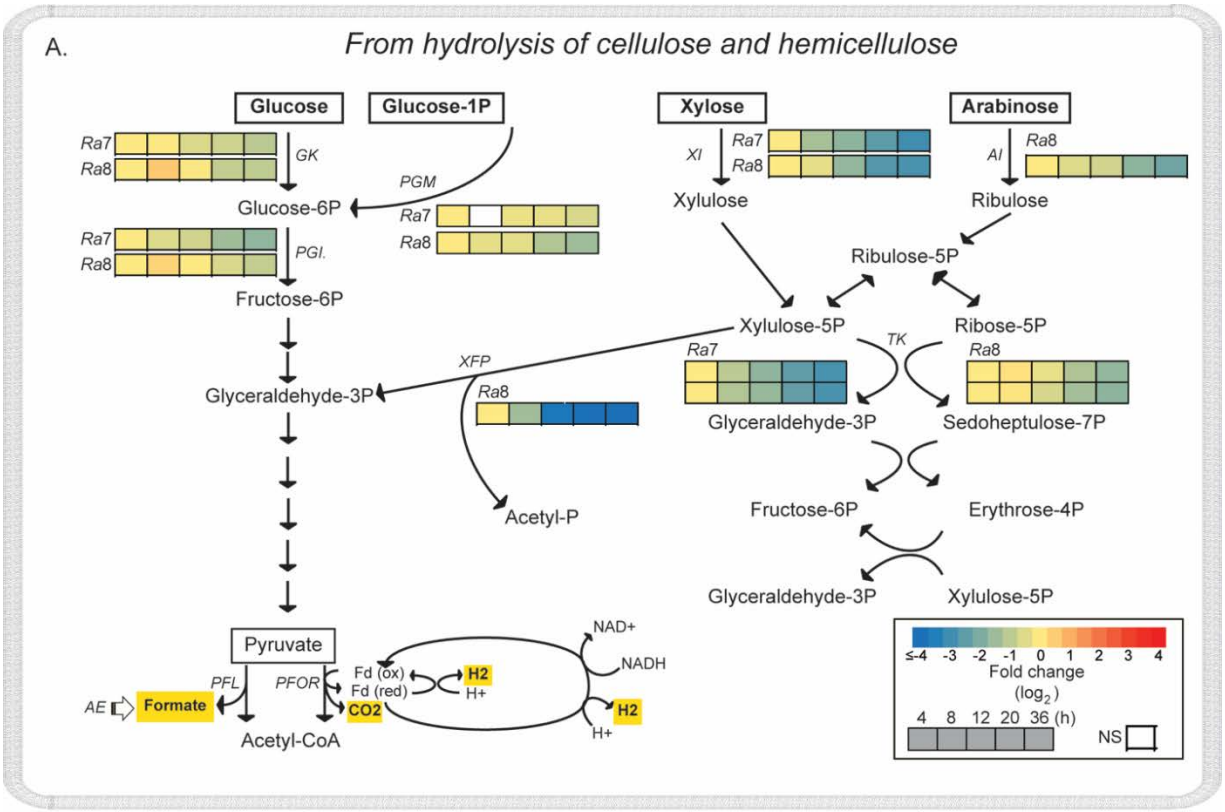


Fig. 2.16 Growth rate of *R. albus* 7 and 8 with Beechwood xylan. Both strains were grown in a defined medium containing 0.4 % (w/v) of Beechwood xylan (BWX). Beechwood xylan obtained from Sigma-Aldrich was composed of mainly xylose residue (> 90% xylose). The growth rate was caluculated using the following equation: $\mu = \ln(At_2/At_1)/(t_2-t_1)$. Data are reported as means ± SD from three biological replicates.



B.

	AE (RPKM)				
	4h	8h	12h	20h	36h
Ra 7	65	54	46	41	37
Ra 8	380	400	432	333	289
	PFL : PFOR[a]				
	4h	8h	12h	20h	36h
Ra 7	0.5 : 1	0.4 : 1	0.8 : 1	1.6 : 1	2.5 : 1
Ra 8	2.3 : 1	1.2 : 1	2 : 1	2.7 : 1	2.9 : 1
Ra7	4h	8h	12h	20h	36h
HydA2	2,962	2,701	2,579	2,599	2,340
HydABC[b]	6,248	4,968	3,443	3,099	2,519
Ra8	4h	8h	12h	20h	36h
HydA2	2,046	3,048	1,696	1,232	635
HydABC	3,714	3,885	2,164	1,451	1,368

C.

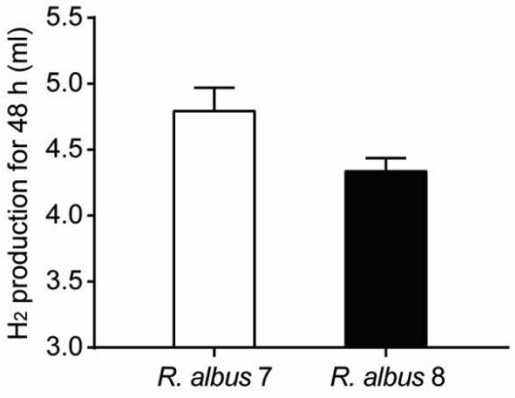


Fig. 2.17 Integrative transcriptional profiles of genes in mono- and di-saccharide utilization pathway of *R. albus* 7 and 8. (*A*) Schematic view of the utilization of monosaccharides through EMP, PP, and XFP pathways is illustrated. The catalytic genes that convert sugars into the intermediates of each EMP, PP, and XFP pathways are shown with their fold changes during growth on AHPCS. Both *R. albus* 7 and 8 possess a gene encoding for L-fucose isomerase domain that was induced greater than 32 fold in *R. albus* 7 and 18 fold in R. albus 8, relative to cellobiose. Compared to other putative xylose isomerase genes in *R. albus* 7 (no xylose isomerase domains has been identified in *R. albus* 8), the putative fucose isomerase gene was expressed greater than 55 ~ 2331 fold on WAX.

Fig. 2.17 (Cont.)
In other study, the orthologous gene annotated as L-fucose isomerase in *C. acetobutylicum* was induced by D-xylose so that it has been proposed as xylose isomerase (Grimmler *et al.*, 2010). Taken together, we postulate a putative fucose isomerase gene would be a xylose isomerase as shown in this figure. *NS* means the fold change in means for two independent experiments was not significantly different ($p \geq 0.05$), as compared with expression level at 4 hour.
(*B*) Expression level of genes involved in conversion of pyruvate to acetyl-CoA shown as RPKM value.
[a] Expression ratio of total RPKM of PFL genes and PFOR gene. Two PFL genes (Rumal_1173, Rumal_1175 and CUS_4465, CUS_4467) and one PFOR gene (Rumal_0032 and CUS_6517) are present on both genomes.
[b] A sum of RPKM of each HydA, HydB, and HydC gene in the HydABC cluster.
(*C*) The amount of H_2 in the headspace of each *R. albus* 7 and 8 culture after 48 hours incubation with AHPCS. Data are reported as means ± S. D. from three biological replicates.
Abbreviations: *GK*, glucokinase; *PGI*, glucose-6P isomerase; *XI*, xylose isomerase; *AI*, arabinose isomerase; *XFP*, xylulose-5P/fructose-6P phosphoketolase; *TK*, transketolase; *PFL*, pyruvate formate lyase; *PFOR*, pyruvate ferredoxin oxidoreductase; *AE*, PFL activating enzyme.

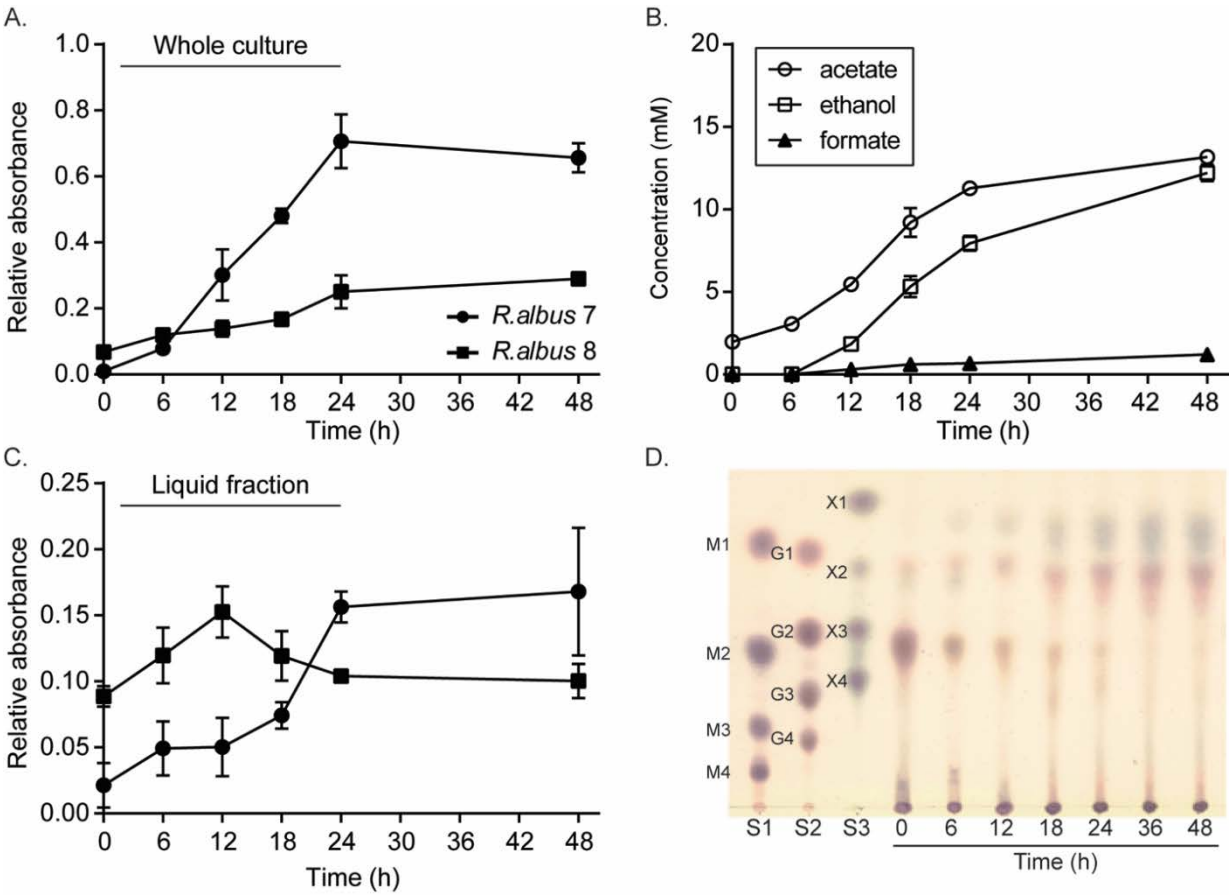


Fig. 2.18 Co-culture of *R. albus* 7 and 8 on AHPCS. Both strain 7 and 8 were grown together with AHPCS as the sole carbohydrate source, and the supernatants and substrate residues of bacterial culture were collected at 0, 4, 8, 12, 16, and 24 h after inoculation. To measure the population of each strain, the qPCR was performed, and the resulting copy number was converted into the relative absorbance value by standard curve. (*A*) Growth curve of each strain 7 and 8 on AHPCS in the whole culture (*B*) The accumulation of fermentation products during co-growth of *R. albus* 7 and 8. (C) The relative absorbance of each strain in liquid fraction. (*D*) Profile of soluble sugars in co-culture medium of *R. albus* 7 and 8. The soluble sugars in supernatant were resolved by thin layer chromatography. Mixture of each monosaccharides (mannose: M1, glucose: G1, and xylose: X1) and their oligomers (manno-oligosaccharides: M2 to M4, cello-oligosaccharides: G2 to G4, xylo-oligosaccharides: X2 to X4) were used as standards (S1, S2, and S3).

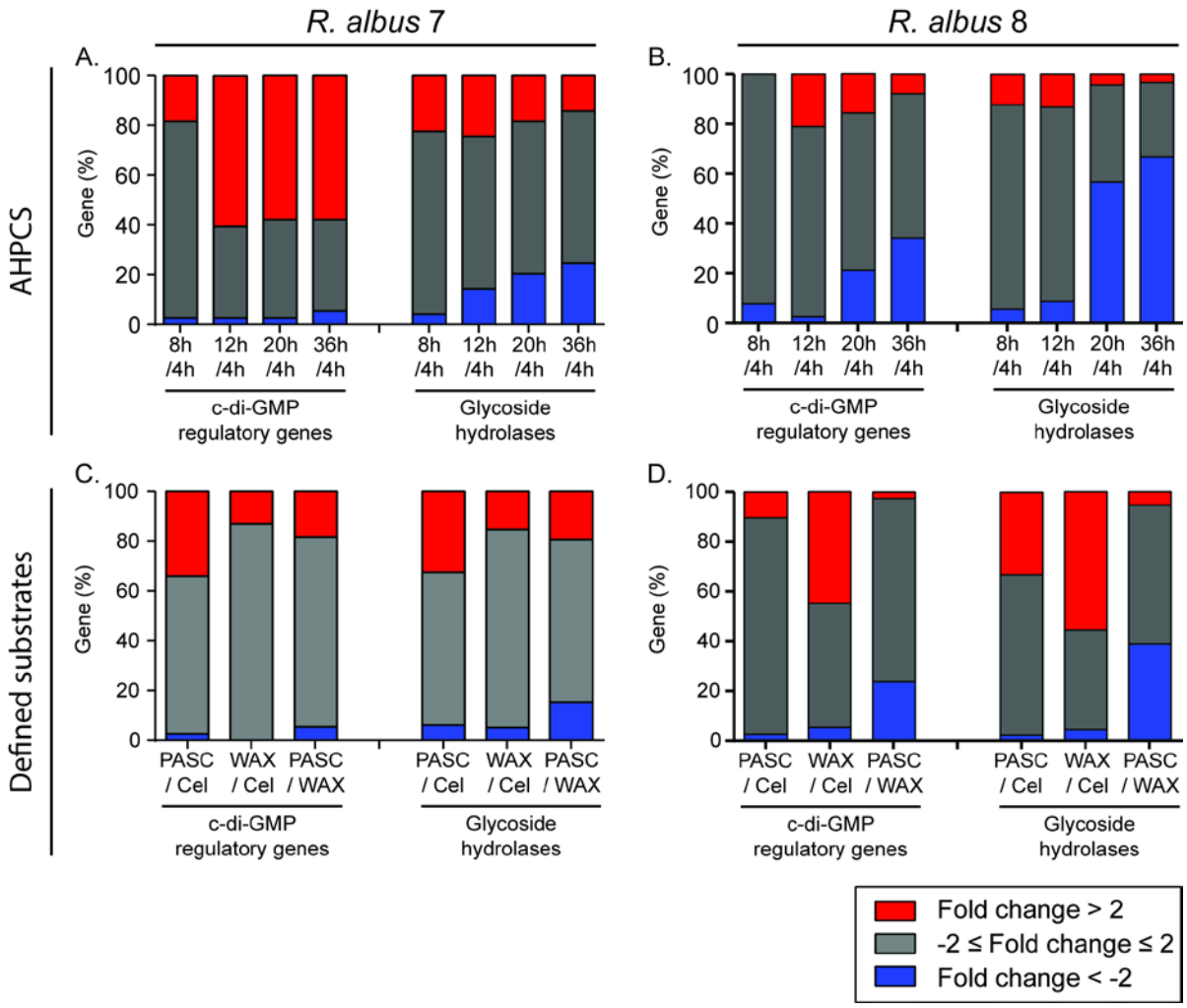


Fig. 2.19 Transcriptional pattern of cyclic di-GMP regulatory genes and GH genes during growth of *R. albus* 7 and 8 on AHPCS and defined substrates, cellobiose (Cel), PASC, or WAX. Each red and blue colored proportion means the percentage of genes up- or down- regulated greater than 2 fold in total number of c-di-GMP regulatory genes and GH genes.

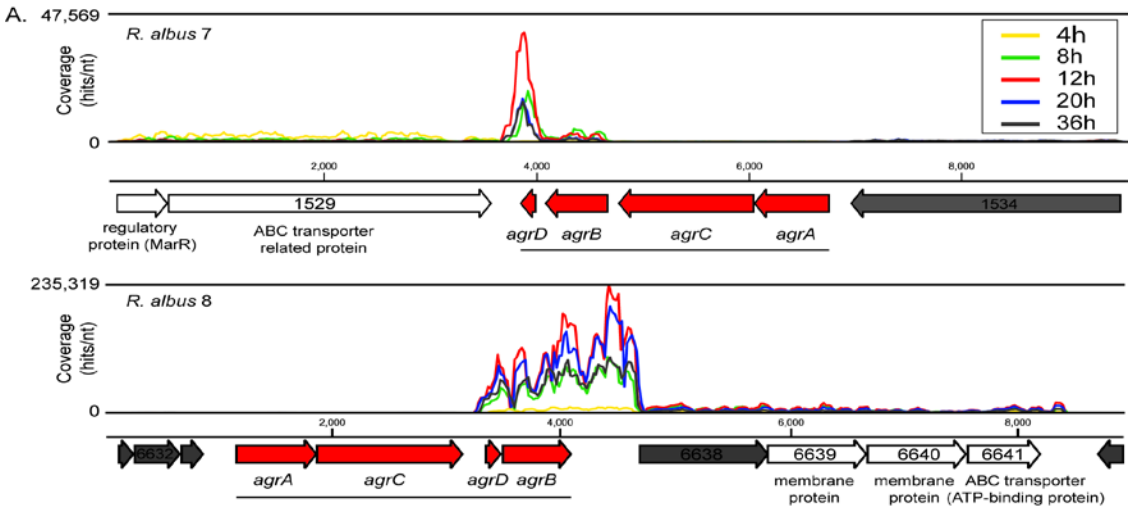


Fig. 2.20 RNA-seq coverage of Agr quorum sensing operon in *R. albus* 7 and 8 grown on AHPCS.

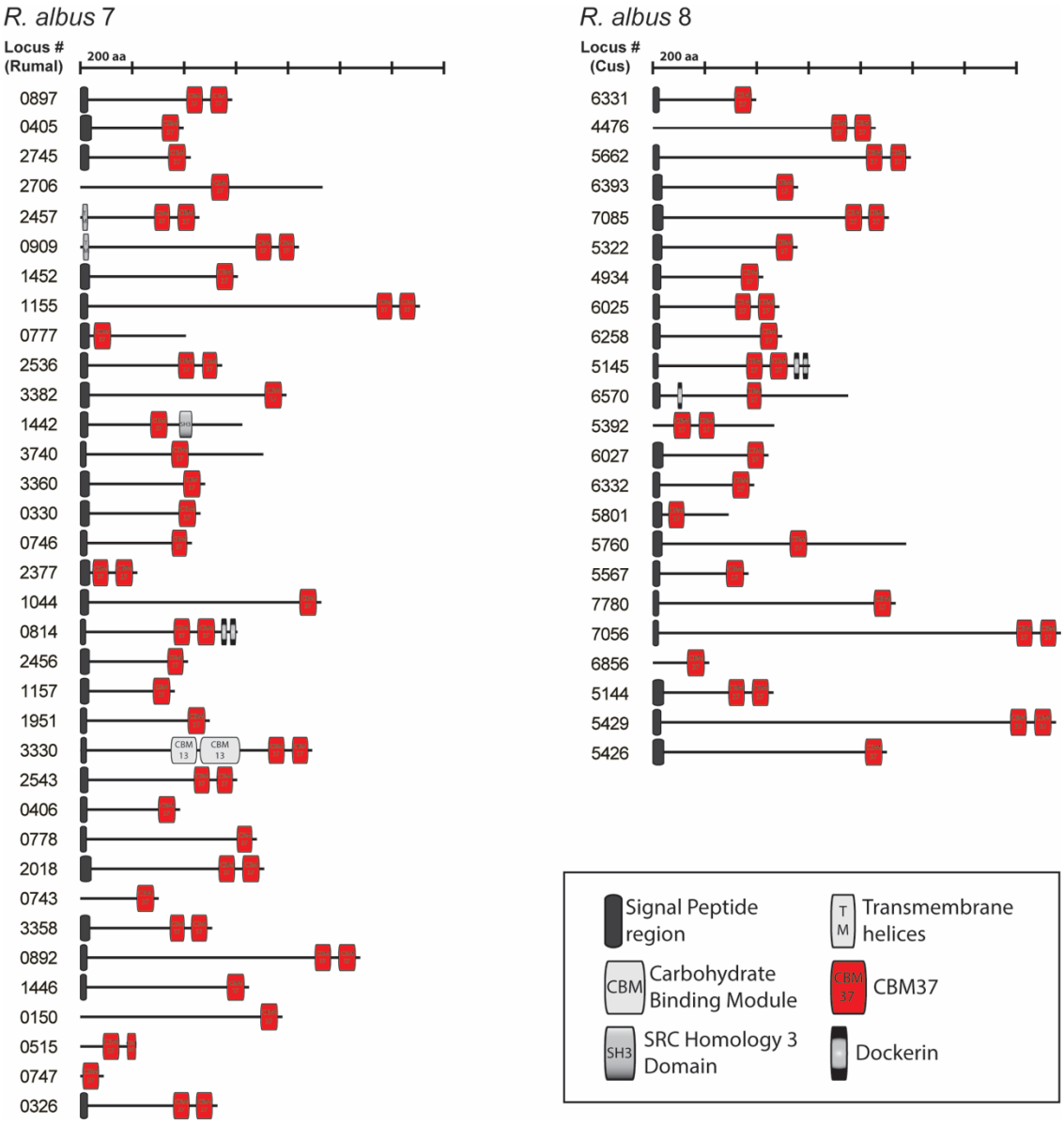


Fig. 2.21 Unknown function of CBM37 genes encoded on the genome of *R. albus* 7 and 8.

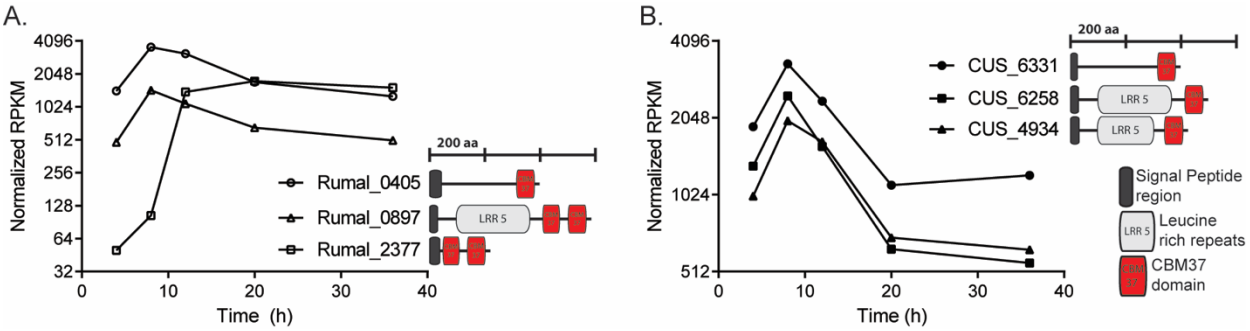


Fig. 2.22 Top 3 highly expressed genes among unknown function of CBM37 genes in *R. albus* 7 and 8 during growth on AHPCS.

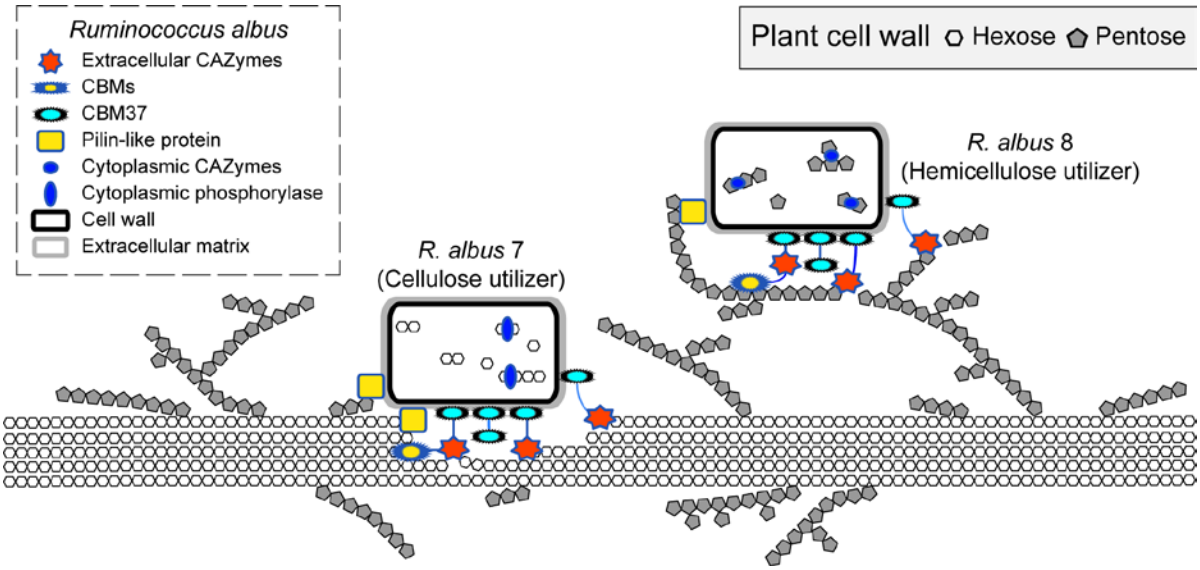


Fig. 2.23 Proposed plant cell wall degrading mechanisms of rumen bacteria.

CHAPTER 3.

COMPARATIVE GENOMICS-DRIVEN ANALYSIS OF FOLATE AND *p*-AMINOBENZOATE METABOLISM IN *RUMINOCOCCUS* STRAINS

3.1 Introduction

Folate (vitamin B_9) is an essential growth factor for animals, plants, and microorganisms. Tetrahydrofolate (THF), a folate derivative, serves as cofactors in one-carbon transfer reactions required for the synthesis of purines, formylmethionyl-tRNA, thymidylate, pantothenate, glycine, serine, and methionine (1). Plants, fungi, and most bacteria make folates de novo, starting from GTP and chorismate, and most animals, including humans, lack key enzymes of the synthetic pathway and therefore folate must be supplied through the diet. However, ruminants in which the rumen is fully functioning are independent of a dietary supply of folate due to biosynthesis by rumen bacteria (2, 3). The rumen bacteria that synthesize folate pass into the abomasum along with digested feed and release their vitamins into the host intestine.

The vitamin requirements of *Ruminococcus albus*, one of the predominant cellulolytic and hemicellulolytic rumen bacteria, were reported by Bryant and Robinson (4). All nine strains studied had an absolute requirement for biotin, but considerable variation in folate requirement existed between individual strains. Some *R. albus* strains, such as *R. albus* 7 and B199, did not require folate for growth, but other strains, such as *R. albus* 20 and $B_3$37, had an absolute requirement for exogenous folate. The genomes of both *R. albus* 7 and 8 have been sequenced, and various studies have been conducted to determine their fiber degradation mechanism through genomic, transcriptomic, and proteomic analyses (5–10). However, vitamin metabolism of *R.*

albus strains remains largely unknown, while the vitamin requirements of *R. albus* 8 have not been studied.

The folate *de novo* synthesis and salvage pathways have been extensively studied in model organisms, such as *Escherichia coli* and Lactobacilli. Folate is composed of three components; a pterin core converted from GTP, *p*-aminobenzoate (*p*ABA), and a glutamate moiety. In the canonical pathway, a total of ten catalytic enzymes are required for the *de novo* synthesis of folate, FolEQBK for the pterin branch, PabABC for the *p*ABA branch, and FolPCA for the synthesis of tetrahydrofolate (THF) using three precursors (GTP, *p*ABA, and glutamate). For the folate uptake or salvage pathway, some bacterial species require FolT, a folate-binding protein that interacts with the energy-coupling factor (ECF) transporter for folate (11, 12). This unique ECF transport system is found mainly in Firmicutes that salvage rather than synthesize folate.

The first reaction of the pterin branch is the conversion of GTP to 7,8-dihydroneopterin triphosphate by GTP cyclohydrolase I (FolE) (13). The resulting 7,8-dihydroneopterin is converted into 6-hydroxymethyl-7,8-dihydropterin in three consecutive steps by a specific pyrophosphatase (FolQ), dihydroneopterin aldolase (FolB), and hydroxymethyldihydropterin pyrophosphokinase (FolK) (13). Subsequently, dihydropteroate synthase (FolP) condenses 6-hydroxymethyl-7,8-dihydropterin with *p*ABA either from the *de novo* synthesis or the salvage pathway. The resulting dihydropteroate is glutamylated by dihydrofolate synthase (FolC), which is reduced to tetrahydrofolate by dihydrofolate reductase (FolA) (13). In the *p*ABA branch pathway, chorismate produced from the shikimate pathway is aminated to 4-aminodeoxychorismate by 4-amino-4-deoxychorimate synthase (PabAB) and, subsequently, cleaved into *p*ABA and pyruvate by 4-amino-4-deoxychorismate lyase (PabC) (14, 15). Other

alternative or atypical synthesis pathways for folate in bacteria, archaea, and protozoa have been reported (16–18), but little research has been conducted with rumen bacteria. Thus, it is relevant to establish whether the folate synthesizing mechanism in model bacteria is applicable to rumen bacteria.

In this paper, we performed a comparative genomic analysis for folate metabolism in *Ruminococcus* species as well as other Firmicutes whose genome sequences have been deposited in the publicly available databases. Next, we examined the vitamin requirement of *R. albus* strain 7 and 8, and showed that this was in agreement with our prediction based on genomic analysis. Then, we demonstrated that the predicted genes in the folate biosynthetic pathway of *R. albus* 7 responded transcriptionally to the demand for folate. In addition, we report the transcriptional responses of global genes for folate biosynthesis in *R. albus* 7 and suggest that non canonical *Ruminococcus* species have a potential for *de novo* synthesis of *p*ABA through an alternative pathway.

3.2 Materials and methods

Comparative genomic analysis for folate metabolism between *R. albus* 7 and 8. Genomic sequences of *R. albus* 7 (GenBank accession numbers: NC_014833 for chromosome and NC_014824 to NC_014827 for plasmids), *R. albus* 8 (NZ_ADKM020000001 to NZ_ADKM02000136), and *R. albus* SY3 (JEOB01000001 to JEOB01000004) were used for comparative genomic analysis among *R. albus* strains. Initial identification of the putative genes in the folate biosynthetic pathway in *Ruminococcus* species and other THF riboswitch bearing bacterial strains was performed using the RAST server (19). Annotation of the predicted genes

was verified using the Conserved Domain Database (CDD) (20), Pfam database (21), and the Enzyme Function Initiative server (EFI) that assigns the predicted function of genes (22), based on the amino acid similarity of a query gene (*E*-value < 1×10^{-5}) to biochemically characterized proteins in the Protein Data Bank (PDB) (23). The tetrahydrofolate (THF) riboswitches present on the genomes of *R. albus* strains and other bacterial strains were predicted by Rfam database (24).

Growth of *R. albus* 7 and 8 with different vitamins. *R. albus* 7 and 8 were cultured anaerobically at 37 °C in butyl rubber-stoppered Balch tubes using a previously reported chemically defined medium with a 100 % CO_2 gas phase (25). The medium contained 0.4 % (w/v) of cellobiose as the sole carbohydrate source and ammonium sulfate 0.4 % (w/v) as the sole nitrogen source (Table 3.1). To determine folate and *p*ABA requirement of *R. albus* 7 and 8, either 10 µg of filter sterilized folate or *p*ABA was separately added into 10 ml of fresh medium after autoclaving. To exclude the transfer of vitamin residue from the seed inoculum, the cells were grown in the respective medium for 24 hours and successively transferred to fresh medium three times. Subsequently, 0.2 ml of pre-grown culture was inoculated into 10 ml of fresh medium. After inoculation, the absorbance at 600 nm was monitored every hour using a Spectronic 21D spectrophotometer (Milton Roy, Warminster, PA). During log growth phase, the growth rate (µ) and doubling time (*G*) was calculated using the following equation: $\mu = \ln(At_2/At_1)/(t_2-t_1)$ and $G = \ln(2)/\mu$, respectively (*A*: absorbance value at 600nm, *t*: incubation time).

RNA extraction and transcriptional analysis by RNA sequencing. For transcriptional analysis, *R. albus* 7 cells grown with different vitamins were harvested at log phase (A_{600nm} of 0.5 for cultures containing *p*ABA and folate and A_{600nm} of 0.2 for cultures lacking *p*ABA and folate),

followed by centrifugation at 13,000 × *g* for 10 min at 4 °C. The resulting cell pellets were stored at –80 °C until RNA extraction. In the subsequent steps, the cell pellets were treated with lysis buffer (200U/ml of mutanolysin, 200 µg/ml of lysozyme, 150µg/ml of proteinase K, 25mM EDTA, and 0.5% (w/v) SDS) for 30 minute at 55 °C (26). The total RNA was extracted with the RNeasy mini kit (Qiagen) with the optional on-column DNase treatment step. Then, the total RNA was eluted with DEPC-treated nuclease-free water and stored at –80 °C until RNA sequencing. For RNA-Seq analyses, RNA isolated from two biological replicates were used for each growth condition. Bacterial ribosomal RNAs were removed from 10 µg of total RNA with the MicrobExpress kit (LifeTechnologies). The enriched mRNA fraction was converted to RNA-Seq libraries using the TruSeq Stranded RNA Sample Prep kit from Illumina. The barcoded libraries were pooled in equimolar concentration and the pool was analyzed by quantitative PCR and sequenced on one lane for 101 cycles on a HiSeq2000 using a TruSeq SBS sequencing kit version3. Fastq files were generated and demultiplexed with the bcl2fastq v1.8.4 Conversion Software (Illumina, Inc.). RNAseq library statistics are shown in Table 3.2. The RNA-Seq data was analyzed using CLC genomics workbench version 5.5.1 from CLC Bio (Cambridge, MA). RNA-Seq reads were mapped onto the genome of *R. albus* 7 as the reference sequence using CLC software. Reads were only assembled if the fraction of the read that aligned with the reference genome was greater than 0.9 and if the read matched other regions of the reference genome at less than 10 nucleotide positions. Then, the RNA-Seq output files were analyzed for statistical significance by using the proportion-based test of Baggerly (27).

Potential candidates for *p*ABA biosynthesis in *R. albus* 7. Phenylpropionic acid (PPA), phenylacetic acid (PAA), benzoic acid (BA), L-tyrosine, 4-hydroxylbenzoic acid, and 3-(4-hydroxyphenyl)propionic acid were individually prepared in distilled water at a concentration of

5 mM. After filter-sterilization, 50 µl of each solution was added into 10 ml of medium lacking folate, *p*ABA, PPA, and PAA to provide a final concentration of 25 µM. To determine the effect of carbon source on the cell aggregation of *R. albus* 7 in the presence of folate or *p*ABA, 0.4 % (w/v) of wheat arabinoxylan (WAX) was used as the sole carbohydrate source in the defined medium described above.

3.3 Results

Genomic analysis for folate metabolism of *Ruminococcus* species and other Firmicutes. We performed genomic analysis for folate metabolism of *R. albus* strain 7, 8, and SY3 for which genome sequences were available in the NCBI database. We sought orthologous genes to the *folEQBKPCA* genes of the canonical folate biosynthetic pathway in these genomes. The closed genome of *R. albus* 7 was predicted to encode the putative THF biosynthetic genes, *folEQBKPCA* (Table 3.3; Fig. 3.1A). However, the *folT* gene required in the folate salvage pathway was absent from the genome. In contrast, the orthologous genes to *folQBKPA* with the exception of *folE* were missing on the draft genome sequence of *R. albus* 8 and instead, the putative *folT* gene was present (Table 3.3; Fig. 3.1B). *R. albus* SY3 harbors both the folate biosynthetic and salvage pathway on the draft genome sequence (Table 3.3; Fig. 3.1C). These results suggest that there is likely to be considerable variation for folate metabolism at the strain level of *R. albus* (Fig. 3.2). Here, we expanded our comparative genomic analysis to the *Ruminococcus* genus level by analyzing the available genome sequences of *Ruminococcus* species isolated from the rumen and human gut. All strains analyzed in this study were predicted to possess either folate biosynthetic genes (*folEQBKPCA*) or salvage gene (*folT*) or both (Fig. 3.3). However, all *Ruminococcus* species that harbor a putative folate biosynthetic pathway lack

orthologs for the canonical *p*ABA synthetic genes, i.e., *pabABC*, suggesting that the potential folate biosynthetic species of *Ruminococcus* possess non-orthologous genes to *pabABC* or require preformed *p*ABA for folate biosynthesis.

As the THF riboswitch is known to sense THF level and regulate the transcription and translation of down-stream genes, we performed genomic analysis to identify the predicted THF riboswitch and their down-stream genes in the 13 genomes of *Ruminococcus* strains and 45 complete genomes of Firmicutes in the Rfam database. We found that 52 of 58 genomes possessed predicted THF riboswitches, which are mostly located upstream of *folT* and only 4 of 52 sequences are present upstream of *folE* or *folC* genes (Fig. 3.3). The prevalence of THF riboswitch sequence upstream of *folT* suggests that most THF riboswitch-bearing organisms would sense THF level and subsequently regulate the folate salvage pathway through a THF riboswitch. Interestingly, despite having similar potential for folate biosynthesis, the genes regulated by the THF riboswitch in *R. albus* and *R. flavefaciens* vary at the strain level (Table 3.4; Fig. 3.3). For example, all *R. albus* strain 7 and SY3 and *R. flavefaciens* strains FD-1 and 007c were predicted to possess the folate biosynthetic pathway. The THF riboswitches were present upstream of *folT* genes in *R. albus* SY3 and *R. flavefaciens* 007c, while in the genomes of *R. albus* 7 and *R. flavefaciens* FD-1 the riboswitches are upstream of the *folC* genes. This positional difference of the THF riboswitch between strains suggests that each strain differs in their transcriptional response to the THF level.

Growth studies confirm that *R. albus* 7 is folate autotroph, while *R. albus* 8 is a folate auxotroph. To determine the folate and *p*ABA requirements for growth, we cultured both strains in a defined medium containing either folate, *p*ABA or without the additions. *R. albus* 7 grew on either folate or *p*ABA containing medium at a similar growth rate (Fig. 3.4A and C), indicating

that *R. albus* 7 is able to synthesize folate via the *de novo* synthetic pathway. Interestingly, *R. albus* 7 was able to grow on the medium lacking even *p*ABA in the absence of folate although it grew slower and to a lower absorbance value (600nm) than when either folate or *p*ABA was supplied. This result suggests that *R. albus* 7 can synthesize *p*ABA *de novo* through an undetermined pathway or folate through an alternative *p*ABA independent pathway. *R. albus* 8 was unable to grow on the medium lacking folate, but grew on the medium containing folate or folate with *p*ABA (Fig. 3.4B and C). These results confirm that *R. albus* 8 is unable to synthesize folate *de novo* using *p*ABA and consequently, *R. albus* 8 has an absolute requirement for preformed folate in the medium, which is in accordance with our prediction based on the genomic analysis.

Transcriptional responses of folate biosynthetic genes to a lack of folate and *p*ABA. To investigate if the predicted folate metabolic pathways are transcriptionally functional and to understand the potential mechanism for the growth without exogenous folate and *p*ABA, a whole genome RNA sequencing (RNA-seq) approach was used in this study. The autotrophic strain, *R. albus* 7, was grown in the presence or absence of both folate and *p*ABA. Total RNA was extracted from cells at the mid-exponential phase. The normalized expression value of each gene is reported as RPKM (reads per kilo base per million mapped reads). During growth of *R. albus* 7 with both folate and *p*ABA relative to the absence of both vitamins in the medium, a total of 672 genes were induced greater than 2 fold, and 199 genes were repressed greater than 2 fold ($p < 0.05$; Table 3.5). Of those up-regulated genes, only 13 genes were significantly up-regulated greater than 10 fold, which included genes related to not only folate related metabolism, but also an antioxidant defense system (Table 3.6). We examined the specific transcriptional responses of genes related to folate metabolism in *R. albus* 7. Of seven genes in the folate synthetic pathway,

genes in the folate synthetic cluster including *folQ*, *folP*, *folK*, and *folA* did not respond transcriptionally, while the putative *folE* and one of two *folC* genes were up-regulated greater than 3 fold (Table 3.7; Fig. 3.5). Although both Rumal_1336 and Rumal_2417 encode a putative *folC* domain, the THF riboswitch controlled gene (Rumal_1336) was 15.7-fold up-regulated, while the non-riboswitch controlled *folC* gene (Rumal_2417) did not respond transcriptionally to a lack of both folate and *p*ABA, suggesting that gene expression is controlled by the THF riboswitch in response to the exogenous level of folate and *p*ABA.

The *de novo* synthesis of folate requires three precursors, GTP, glutamate, and *p*ABA. A variety of metabolic pathways, such as glycolysis, pentose phosphate (PP) pathway, TCA cycle, ammonium assimilation, and purine synthesis pathways, are required for the synthesis of those precursors from cellobiose and ammonia used as the main carbon and nitrogen sources by the Ruminococci. We determined the transcriptional responses of the metabolic genes related to folate synthesis. During exponential growth without exogenous folate and *p*ABA, in general, the metabolic pathways tended to be up-regulated toward the supply of precursors for folate synthesis (Table 3.8-3.10; Fig. 3.5). Transcriptional analysis showed the up-regulation of genes in the central sugar metabolic pathways connecting to glutamate and GMP synthesis through the purine metabolic pathway and ammonia assimilation. Consistently, genes involved in one carbon metabolism that are required for purine metabolism were up-regulated (Table 3.11). Based on our transcriptomic analysis, it seems likely that folate was synthesized through the canonical synthetic pathway, thereby requiring *p*ABA as one of precursors.

Potential precursor for *p*ABA *de novo* synthesis in *R. albus* 7. We sought to find the potential biosynthetic pathway for *p*ABA, based on our transcriptional data. The shikimate pathway produces chorismate, the precursor for the canonical synthesis pathway for *p*ABA and aromatic

amino acids. Three genes in this pathway were up-regulated greater than 2 fold, including putative *aroA*, *aroD*, and *aroE* (Table 3.12; Fig. 3.6). In the downstream of the shikimate pathway, we also found that the genes required for the synthesis of the aromatic amino acids including tyrosine and phenylalanine, but not tryptophan, were also up-regulated (Table 3.13; Fig. 3.6). If chorismate was used to synthesize *p*ABA, the potential non-orthologous genes to *pabABC* would be required to have glutamine amidotransferase and lyase activities. However, all predicted amidotransferase genes that we identified had a very low expression level and were not significantly ($p \geq 0.05$) induced in response to a lack of both vitamins, suggesting that the chorismate would be unlikely the precursor for *p*ABA. As the synthetic genes for tyrosine and phenylalanine were up-regulated, we postulated that *R. albus* 7 might have an alternative *p*ABA synthesis pathway using a tyrosine or benzene ring bearing compound as precursor. Since phenylpropionic acid (PPA) and phenylacetic acid (PAA) that contain a benzene ring are known as essential micronutrients for *R. albus* to utilize cellulose and therefore, these two compounds were supplied in the medium for growth of cells in this study (28–31). We reasoned that it was biochemically feasible that tyrosine, PPA, or PAA could be converted into *p*ABA through a beta-oxidation reaction at the side carbon chain (Fig. 3.7). To examine these hypothetical pathways, we tested six compounds as *p*ABA precursors, including PPA, PAA, L-tyrosine, 3-(4-hydroxyphenyl)propionate, benzoate, and 4-hydroxybenzoate. *R. albus* 7 was able to grow on the folate and *p*ABA lacking medium in the absence of both PPA and PAA (Fig. 3.8A-C right panel). Interestingly, when cells were grown on the cellobiose medium, a lack of PPA caused the cells to form aggregates in liquid medium (Fig. 3.8i-iii left panel). However, when the sole carbon source in the medium was substituted with hemicellulose, wheat arabinoxylan (WAX), *R. albus* 7 did not aggregate in the absence of PPA (Fig. 3.8D left panel and iv right panel). Next, we

cultured *R. albus* 7 on cellobiose plus PPA containing medium with the other test compounds. However, there was no growth stimulatory effect of tyrosine, 3-(4-hydroxyphenyl) propionate, benzoate, and 4-hydroxybenzoate on *R. albus* 7 (Fig. 3.8E left panel).

3.4 Discussion

Folate and its derivatives are essential cofactors because the reduced form, tetrahydrofolate, is involved in one carbon unit transfer reaction in purine and amino acid synthesis in all organisms (13, 32, 33). Thus, folate should be supplied to the cell through either *de novo* synthesis or uptake of extracellular folate. Mammals, including humans, cannot synthesize folate *de novo* and consequently rely on the dietary folate, while ruminants are independent of the dietary folate due to a supply from rumen bacteria (2, 3). *Ruminococcus* species are predominant fibrolytic bacterial species in the rumen and other gut ecosystem and have been studied primarily for their fibrolytic mechanism (34). However, their contribution to the vitamin nutrition of the host, especially folate, has received far less attention, relative to the carbohydrate nutrition of host. In previous studies, *R. albus* and *R. flavefaciens* showed considerable variation in folate requirement between strains, indicating that each strain within *Ruminococcus* genus possesses a differential capability of folate synthesis and consequently makes a different contribution to the folate nutrition of host (4, 35). Although genome sequences are now available for several strains of *Ruminococcus* species, the molecular mechanisms for their folate metabolism are still unknown.

Based on our comparative genomic analyses for *Ruminococcus* species and other THF riboswitch bearing Firmicutes, we predicted that two major distinct pathways for folate

utilization are conserved in these bacterial genomes, folate *de novo* synthetic and salvage pathways, or both. Despite the presence of the folate biosynthetic pathway, most of THF riboswitch controlled genes in those genomes are involved in the transport of exogenous folate (*folT*). Only four strains in this study encode the THF riboswitch upstream of genes in the folate biosynthetic pathway. It has been reported that the THF riboswitch either terminates transcription or initiates translation of downstream mRNA in response to THF or *p*ABA binding (36, 37). Our transcriptomic work showed that only the THF riboswitch bearing *folC* gene was up-regulated in *R. albus* 7, while the non-THF riboswitch bearing *folC* gene did not respond to a lack of folate and *p*ABA. Thus, the level of THF or *p*ABA is likely sensed by THF riboswitch and consequently would control either the folate synthesis pathway or salvage pathway of THF riboswitch-bearing Firmicutes, including *Ruminococcus* species.

In this study, we demonstrated that *R. albus* 7 was able to grow with *p*ABA as well as folate and *R. albus* 8 was able to grow only with folate addition. Interestingly, all isolated species within the *Ruminococcus* genus lack the canonical *p*ABA biosynthetic pathway, but our results demonstrate that *R. albus* 7 is independent of even dietary *p*ABA for the growth in the absence of dietary folate. It is possible that *R. albus* 7 circumvents the folate requirement for growth or synthesize folate through a *p*ABA independent pathway. It has been reported that the addition of the purines (e.g. guanine, adenine, xanthine, or inosine) restored growth of *Streptococcus faecalis* R and *Lactococcus lactis* in the absence of folate (38, 39). However, our chemically defined growth medium did not contain any purines. In addition, the genes involved in the folate synthesis pathway composed of pterin branch, glutamate synthesis, and GMP synthesis, were up-regulated during growth in the absence of both vitamins. Thus, it seems likely that *R. albus* 7 synthesizes folate through the canonical pathway and consequently must possess an alternative

*p*ABA synthetic pathway. In the canonical pathway conserved in many bacteria and eukaryotes, *p*ABA is synthesized from chorismate in two steps. The hybrid PabAB or individual PabA and PabB proteins convert chorismate to 4-amino-4-deoxychorismate (ADC) through glutamine amidotransferase and aminase reactions. The resultant ADC is cleaved into *p*ABA and pyruvate by ADC lyase (PabC). We were unable to find any orthologous genes to *pabABC* in the available genomes of the *Ruminococcus* genus. In *R. albus* 7, the best BLAST hits of the reference PabAB proteins of *E. coli* K012 and *L. lactis* MG1363 are the tryptophan synthesis genes, anthranilate phosphoribosyltransferase (*trpD*) and anthranilate synthase (*trpE/G*). Thus, it is possible to assume that in *Ruminococcus* species, the tryptophan synthase is mis-annotated and functions as *p*ABA synthase because both enzymes require the same substrate (chorismate) and perform similar enzymatic reactions. In the tryptophan synthetic pathway, chorismate is converted to 2-aminobenzoic acid (anthranilate) by TrpEG enzymes, while the chorismate is converted to 4-aminobenzoic acid (4-ABA) by PabABC in *p*ABA synthetic pathway (14, 40). It has been reported that the folate synthetic operon contains a glutamine amidotransferase (PabA/TrpG) which is involved in both biosynthesis of *p*ABA and anthranilate (41, 42). However, both putative *trpD* and *trpE/G* genes in *R. albus* 7 were not induced in response to a lack of exogenous *p*ABA and the ortholog of *pabC* gene, essential for the final *p*ABA synthesis reaction step, is still missing in the complete genome sequence of *R. albus* 7. *R. albus* 7 possesses a gene encoding a putative YceG domain (Rumal_1954), which is annotated as 4-amino-4-deoxychorismate lyase (PabC) by genome sequence uploader group to NCBI database. However, there is no biochemical evidence for the lyase activity of YceG protein and in addition it did not respond to a lack of exogenous folate and *p*ABA. In support of this possibility, Pfam (PF02618) and Interpro (IPR003770) database report that a number of genes encoding YceG have been mis-

annotated as the PabC protein due to proximity to PabC (21, 43). Taken together, it seems unlikely that *R. albus* 7 possesses orthologs to the canonical *p*ABA synthesis genes.

Other alternative pathways for *p*ABA synthesis have been suggested in some bacterial species and archaea that do not possess the orthologs of the *pabABC* genes. A single enzyme of the COG5424 family (pyrroloquinoline quinone biosynthesis protein C; PqqC) in *C. trachomatis and N. europaea* showed the capability of *p*ABA synthesis using an unknown substrate (44, 45). In the methanogenic archaeon, *Methanococcus maripaludis*, *p*ABA is derived from an intermediate in the shikimate pathway, 3-dehydroquinate, and not from chorismate (16). However, we were unable to find an orthologous gene to the PqqC gene in *Ruminococcus* species and downstream of shikimate pathway, while aromatic amino acid (phenylalanine and tyrosine) synthesis pathway were up-regulated in the absence of *p*ABA. Thus, we postulated that *R. albus* 7 possesses an alternative *p*ABA synthesis pathway using phenolic ring compounds, such as tyrosine and PPA. Phenylalanine has a similar molecular structure to tyrosine. However, we excluded phenylalanine since it has no side chain at fourth carbon in benzene ring, making the *p*ABA synthetic reaction much less favorable than tyrosine. In addition, if phenylalanine is used as a precursor for the *p*ABA synthesis, PPA and benzoate tested in this study would have a similar stimulatory effect on the growth of *R. albus* 7, since both compounds are predicted as intermediates in our hypothetical pathway. Although none of the six compounds had a growth stimulatory effect like *p*ABA, it is possible that the tested compounds are not transported into cells or the potential pathway for *p*ABA synthesis would be independent of the six compounds.

It is notable that a lack of PPA in *R. albus* 7 culture caused the cells to form aggregates in the cellobiose medium. PPA has been reported to improve cellulose utilization by *R. albus* 7 and 8 (28, 29, 31, 46). Later, PPA was shown to increase transcription of the pilA1 gene that is

involved in the attachment of *R. albus* 8 to cellulose (47). When cellobiose was substituted with WAX, the aggregation phenomenon was not observed, suggesting that PPA would be implicated in extracellular matrix that is induced by the end product of cellulose hydrolysis, cellobiose (8, 48), but not pentose based polysaccharide such as WAX. Future work is still required to define the biological role of PPA in fibrolytic gut bacteria.

An additional intriguing feature of the transcriptome of *R. albus* 7 was the high induction of an antioxidant enzyme cluster, consisting of peroxiredoxin, a transcriptional regulator, thioredoxin, and thioredoxin-disulfide reductase. The putative peroxiredoxin and thioredoxin genes were highly up-regulated (57 and 11 fold, respectively) in this study (Fig. 3.9). Peroxiredoxins are ubiquitous peroxidase enzymes identified in many bacteria, archaea, and eukaryotes (49, 50). After reduction of hydrogen peroxide by peroxiredoxin, resulting oxidized peroxiredoxin is regenerated by thioredoxin (49, 51). Thus, both enzymes, together with superoxide dismutase and catalase, have been known as components of a redox system involved in balancing redox state and extensively studied for their role in defense mechanism against oxidative stress (49, 52). It has been reported that folate deficiency induces oxidative stress and subsequently antioxidant enzymes in animal cells (53–55). However, the effect of folate deficiency on the oxidative stress is still unknown in bacteria. To our knowledge, this is the first transcriptional analyses result showing folate deficiency-induced oxidative stress response in the anaerobic bacterium, *R. albus* 7. It is unclear why the oxidative stress response genes were strongly induced under anaerobic growth condition. A possible explanation is that the deficiency of folate and *p*ABA would generate an abnormal redox state, such as imbalance of NAD(P)+/NAD(P)H or FAD+/FADH involved in the folate and one carbon metabolism pathways, and subsequently, the abnormal redox state would be controlled by the transcriptional

regulator in the peroxiredoxin operon. This explanation is supported by the highly up-regulated genes in one carbon pool metabolism in the absence of folate and *p*ABA, including NAD(P)H dependent 5,10-methylenetetrahydrofolate reductases. The highly up-regulated peroxiredoxin operon contains a putative transcriptional regulator gene, which is predicted as MarR family protein. The MarR type regulators are found in many bacteria and archaea (56–58). It is known that MarR protein regulates gene expression required for resistance to antibiotics, organic solvents, and oxidative stress agents (56–58). Taken together with the transcriptional evidence, it appears that the MarR type regulator would be a strong candidate for the transcriptional regulator in response to stress induced by folate deficiency.

In this study, we demonstrated that differences in folate metabolism (either *de novo* biosynthesis, salvage, or both pathways) are present at the species level and even at the strain level of bacterial species, leading us to address a fundamental question; why does folate metabolism in Fimicutes vary with species and strains during bacterial evolution? Many cases of the prevalence of auxotrophs for essential metabolites such as vitamins and amino acids have been reported between species and strains within the bacterial and eukaryotic kingdoms (59–61). According to the "Black Queen" hypothesis proposed by Morris and colleagues, where members of a community lose the ability to perform functions whose products are available from the environment, auxotrophs presumably arise from autotrophic ancestors as a result of the loss of essential biosynthetic function to reduce the metabolic burden when the corresponding metabolite is available in their habitat or produced by neighboring commensal organisms (62, 63). This hypothesis is supported by empirical observations showing that inactivation and/or deletion of certain metabolic biosynthetic genes provide an increase in fitness to the engineered bacterial strains (60, 64). Applying this concept to the variation of folate metabolic pathways in

R. albus species, strain 7, 8, and SY3 possessing either of the three folate utilization pathways would have evolved from an autotrophic ancestor harboring a complete folate biosynthetic pathway. Considering the potential saving of biosynthetic costs for synthesis of GTP, glutamate and production of the long (six enzymatic steps) biosynthetic pathway, it seems reasonable that in the rumen where exogenous folate is present in sufficient amounts (rumen fluid level 80-186 ng/ml; Ragaller *et al.*, 2009), auxotrophic strains subsequently arise to utilize exogenous folate by acquisition of the transport or salvage pathway (*folT*). For comparison, levels of folate in the synthetic growth medium (1 ug/ml) are ca. 10 fold higher. As a result of this selection, the lineage capable of actively transporting folate lost the redundant genetic material for folate biosynthesis. This loss of genetic material or adaptive genome streamlining is thought to occur more frequently in nutrient rich or constant environments such as the intestinal tract (60, 63, 66). Indeed, the presence of three different folate utilization pathways in the strains of *R. albus* suggests that at least the metabolic differentiation for folate in *R. albus* species would have occurred following the speciation event.

3.5 Conclusion

This study provides the first integrative analysis of genomic, phenotypic, and transcriptomic results for folate and *p*ABA metabolism of the mammalian gut species, *Ruminococcus albus*. We provide genomic evidence for three folate utilization pathways (either *de novo* synthesis, salvage, or both pathways) conserved in the Firmicutes including *R. albus* strains. We demonstrated that at the strain level, *R. albus* strains 7 and 8 rely on different folate metabolic pathways, *de novo* synthesis or salvage pathway, which would imply a differential role

of each bacterial strain in the gut and host nutrition as either autotroph or auxotroph, respectively. In addition, our findings suggest that the folate autotrophic strain, *R. albus* 7, has an alternative pathway for *p*ABA synthesis and likewise other *Ruminococcus* species lacking the canonical *p*ABA synthetic pathway could not be *p*ABA auxotrophs, but autotrophs. Taken together, our findings provide molecular insight in folate metabolism of *R. albus* 7 and 8, which contributes to an understanding and functional model for vitamin metabolism of the microbial community in the mammalian gut.

3.6 References

1. **Matthews RG**. 1996. One-carbon metabolism, p. 600–611. *In* Neidhardt, FC (ed.), In Escherichia coli and Salmonella: Cellular and Molecular Biology. Volimes I and II.Second. ASM Press, Washington DC.

2. **Wolin MJ, Miller TL, Stewart CS**. 1997. The Rumen Microbial Ecosystem, p. 467–491. *In* Hobson, PN, Stewart, CS (eds.), Second edi. Blackie Academic & Professional, London.

3. **McDowell LR**. 2000. Folacin, p. 480–521. *In* Vitamins in animal and human nutritionSecond. Iowa State University Press, Ames.

4. **Bryant MP, Robinson IM**. 1961. Some Nutritional Requirements of the Genus Ruminococcus. Appl Microbiol **9**:91–95.

5. **Dassa B, Borovok I, Ruimy-Israeli V, Lamed R, Flint HJ, Duncan SH, Henrissat B, Coutinho P, Morrison M, Mosoni P, Yeoman CJ, White BA, Bayer EA**. 2014. Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. PLoS One **9**:e99221.

6. **Christopherson MR, Dawson J a, Stevenson DM, Cunningham AC, Bramhacharya S, Weimer PJ, Kendziorski C, Suen G**. 2014. Unique aspects of fiber degradation by the ruminal ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. BMC Genomics **15**:1066.

7. **Ezer A, Matalon E, Jindou S, Borovok I, Atamna N, Yu Z, Morrison M, Bayer EA, Lamed R**. 2008. Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. J Bacteriol **190**:8220–8222.

8. **Devillard E, Goodheart DB, Karnati SKR, Bayer EA, Lamed R, Miron J, Nelson KE, Morrison M**. 2004. *Ruminococcus albus* 8 mutants defective in cellulose degradation Are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. J Bacteriol **186**:136–145.

9. **Moon YH, Iakiviak M, Bauer S, Mackie RI, Cann IKO**. 2011. Biochemical analyses of multiple endoxylanases from the rumen bacterium *Ruminococcus albus* 8 and their synergistic activities with accessory hemicellulose-degrading enzymes. Appl Environ Microbiol **77**:5157–69.

10. **Iakiviak M, Mackie RI, Cann IKO**. 2011. Functional analyses of multiple lichenin-degrading enzymes from the rumen bacterium *Ruminococcus albus* 8. Appl Environ Microbiol **77**:7541–50.

11. **Eudes A, Erkens GB, Slotboom DJ, Rodionov DA, Naponelli V, Hanson AD**. 2008. Identification of genes encoding the folate- and thiamine-binding membrane proteins in firmicutes. J Bacteriol **190**:7591–7594.

12. **Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD, Eitinger T**. 2009. A novel class of modular transporters for vitamins in prokaryotes. J Bacteriol **91**:42–51.

13. **Green JBPN, Matthews RG**. 1996. Folate biosynthesis, reduction, and polyglutamylation, p. 665–673. *In* Neidhardt, FC (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular BiologySecond. ASM Press, Washington, DC.

14. **Green JM, Nichols BP**. 1991. p-Aminobenzoate biosynthesis in *Escherichia coli*: Purification of aminodeoxychorismate lyase and cloning of *pabC*. J Biol Chem **266**:12971–12975.

15. **Parsons JF, Jensen PY, Pachikara AS, Howard AJ, Eisenstein E, Ladner JE**. 2002. Structure of *Escherichia coli* aminodeoxychorismate synthase: Architectural conservation and diversity in chorismate-utilizing enzymes. Biochemistry **41**:2198–2208.

16. **Porat I, Sieprawska-Lupa M, Teng Q, Bohanon FJ, White RH, Whitman WB**. 2006. Biochemical and genetic characterization of an early step in a novel pathway for the biosynthesis of aromatic amino acids and p-aminobenzoic acid in the archaeon *Methanococcus maripaludis*. Mol Microbiol **62**:1117–31.

17. **Dittrich S, Mitchell SL, Blagborough AM, Wang Q, Wang P, Sims PFG, Hyde JE**. 2008. An atypical orthologue of 6-pyruvoyltetrahydropterin synthase can provide the missing link in the folate biosynthesis pathway of malaria parasites. Mol Microbiol **67**:609–18.

18. **Kuratsu M, Hamano Y, Dairi T**. 2010. Analysis of the *Lactobacillus* metabolic pathway. Appl Environ Microbiol **76**:7299–7301.

19. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

20. **Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH**. 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res **41**:D348–52.

21. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M**. 2014. Pfam: the protein families database. Nucleic Acids Res **42**:D222–30.

22. **Gerlt J a, Allen KN, Almo SC, Armstrong RN, Patricia C, Cronan JE, Dunaway-mariano D, Imker HJ, Matthew P, Minor W, Poulter CD, Raushel FM, Sali A, Brian K**. 2011. The enzyme function initiative. Biochemistry **50**:9950–9962.

23. **Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE**. 2000. The Protein Data Bank. Nucleic Acids Res **28**:235–242.

24. **Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD**. 2015. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res **43**:D130–7.

25. **Kim JN, Henriksen ED, Cann IKO, Mackie RI**. 2014. Nitrogen utilization and metabolism in *Ruminococcus albus* 8. Appl Environ Microbiol **80**:3095–102.

26. **Rakotoarivonina H, Larson MA, Morrison M, Girardeau JP, Gaillard-Martinie B, Forano E, Mosoni P**. 2005. The *Ruminococcus albus* pilA1-pilA2 locus: Expression and putative role of two adjacent pil genes in pilus formation and bacterial adhesion to cellulose. Microbiology **151**:1291–1299.

27. **Baggerly KA, Deng L, Morris JS, Aldaz CM**. 2003. Differential expression in SAGE: accounting for normal between-library variation. Bioinformatics **19**:1477–1483.

28. **Hungate RE, Stack RJ**. 1982. Phenylpropanoic Acid: Growth factor for *Ruminococcus albus*. Appl Environ Microbiol **44**:79–83.

29. **Stack RJ, Hungate RE, Opsahl WP**. 1983. Phenylacetic acid stimulation of cellulose digestion by *Ruminococcus albus* 8. Appl Environ Microbiol **46**:539–544.

30. **Stack RJ, Cotta MA**. 1986. Effect of 3-phenylpropanoic Acid on growth of and cellulose utilization by cellulolytic ruminal bacteria. Appl Environ Microbiol **52**:209–10.

31. **Morrison M, Mackie RI, Kistner A**. 1990. 3-phenylpropanoic acid improves the affinity of *Ruminococcus albus* for cellulose in continuous culture. Appl Environ Microbiol **56**:3220–3222.

32. **Cossins EA, Chen L**. 1997. Folates and one-carbon metabolism in plants and fungi. Phytochemistry **45**:437–452.

33. **Scott JM, Weir DG**. 1998. Folic acid, homocysteine and one-carbon metabolism: a review of the essential biochemistry. J Cardiovasc Risk **5**:223–227.

34. **Krause DO, Dalrymple BP, Smith WJ, Mackie RI, McSweeney CS**. 1999. 16S rDNA sequencing of *Ruminococcus albus* and *Ruminococcus flavefaciens*: Design of a signature probe and its application in adult sheep. Microbiology **145**:1797–1807.

35. **Scott HW, Dehority BA**. 1965. Vitamin requirements of several cellulolytic rumen bacteria. J Bacteriol **89**:1169–75.

36. **Ames TD, Rodionov DA, Weinberg Z, Breaker RR**. 2010. A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. Chem Biol **17**:681–5.

37. **Trausch JJ, Batey RT**. 2014. A disconnect between high-affinity binding and efficient regulation by antifolates and purines in the tetrahydrofolate riboswitch. Chem Biol **21**:205–216.

38. **Samuel CE, D'Ari L, Rabinowitz JC**. 1970. Evidence against the folate-mediated formylation of formyl-accepting methionyl transfer ribonucleic acid in *Streptococcus faecalis* R. J Biol Chem **245**:5115–5121.

39. **Wegkamp A, van Oorschot W, de Vos WM, Smid EJ**. 2007. Characterization of the role of para-aminobenzoic acid biosynthesis in folate production by *Lactococcus lactis*. Appl Environ Microbiol **73**:2673–81.

40. **Walsh CT, Liu J, Rusnak F, Sakaitani M**. 1990. Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. Chem Rev **90**:1105–1129.

41. **Kane JF, Holmes WM, Jensen RA**. 1972. Metabolic interlock. The dual function of a folate pathway gene as an extra-operonic gene of tryptophan biosynthesis. J Biol Chem **247**:1587–1596.

42. **Slock J, Stahly DP, Han CY, Six EW, Crawford IP**. 1990. An apparent *Bacillus subtilis* folic acid biosynthetic operon containing pab, an amphibolic *trpG* gene, a third gene required for synthesis of para-aminobenzoic acid, and the dihydropteroate synthase gene. J Bacteriol **172**:7211–7226.

43. **Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJA, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD**. 2015. The InterPro protein families database: The classification resource after 15 years. Nucleic Acids Res **43**:D213–D221.

44. **Satoh Y, Kuratsu M, Kobayashi D, Dairi T**. 2014. New gene responsible for para-aminobenzoate biosynthesis. J Biosci Bioeng **117**:178–83.

45. **Adams NE, Thiaville JJ, Proestos J, Juárez-Vázquez AL, Mccoy AJ, Barona-Gómez F**. 2014. Promiscuous and adaptable enzymes fill " holes " in the tetrahydrofolate pathway in *Chlamydia* species. MBio **5**:e01378–14.

46. **Stack RJ, Cotta MA**. 1986. Effect of 3-phenylpropanoic acid on growth of and cellulose utilization by cellulolytic ruminal bacteria **52**:209–210.

47. **Pegden RS, Larson MA, Grant RJ, Morrison M**. 1998. Adherence of the Gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. J Bacteriol **180**:5921–5927.

48. **Thurston B, Dawson KA, Strobel HJ**. 1993. Cellobiose versus glucose utilization by the ruminal bacterium *Ruminococcus albus*. Appl Environ Microbiol **59**:2631–2637.

49. **Perkins A, Nelson KJ, Parsonage D, Poole LB, Karplus PA**. 2015. Peroxiredoxins: guardians against oxidative stress and modulators of peroxide signaling. Trends Biochem Sci **40**:435–445.

50. **Wood ZA, Schroder E, Robin Harris J, Poole LB**. 2003. Structure, mechanism and regulation of peroxiredoxins. Trends Biochem Sci **28**:32–40.

51. **Rhee SG, Woo HA, Kil IS, Bae SH**. 2012. Peroxiredoxin functions as a peroxidase and a regulator and sensor of local peroxides. J Biol Chem **287**:4403–4410.

52. **Miller RA, Britigan BE**. 1997. Role of oxidants in microbial pathophysiology. Clin Microbiol Rev **10**:1–18.

53. **Chern CL, Huang RFS, Chen YH, Cheng JT, Liu TZ**. 2001. Folate deficiency-induced oxidative stress and apoptosis are mediated via homocysteine-dependent overproduction of hydrogen peroxide and enhanced activation of NF-κB in human Hep G2 cells. Biomed Pharmacother **55**:434–442.

54. **Lan W, Guhaniyogi J, Horn MJ, Xia JQ, Graham B**. 2007. A density-based proteomics sample fractionation technology: Folate deficiency-induced oxidative stress response in liver and brain. J Biomol Tech **18**:213–225.

55. **Kao T-T, Chu C-Y, Lee G-H, Hsiao T-H, Cheng N-W, Chang N-S, Chen B-H, Fu T-F**. 2014. Folate deficiency-induced oxidative stress contributes to neuropathy in young and aged zebrafish--implication in neural tube defects and Alzheimer's diseases. Neurobiol Dis **71**:234–44.

56. **Alekshun MN, Levy SB**. 1999. The mar regulon: Multiple resistance to antibiotics and other toxic chemicals. Trends Microbiol **7**:410–413.

57. **Egland PG, Harwood CS**. 1999. BadR, a new MarR family member, regulates anaerobic benzoate degradation by *Rhodopseudomonas palustris* in concert with AadR, an Fnr family member. J Bacteriol **181**:2102–2109.

58. **Wu RY, Zhang RG, Zagnitko O, Dementieva I, Maltzev N, Watson JD, Laskowski R, Gornicki P, Joachimiak A**. 2003. Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor. J Biol Chem **278**:20240–20244.

59. **Helliwell KE, Wheeler GL, Leptos KC, Goldstein RE, Smith AG**. 2011. Insights into the evolution of vitamin B12 auxotrophy from sequenced algal genomes. Mol Biol Evol **28**:2921–33.

60. **D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, Kost C**. 2014. Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution (N Y) **68**:2559–2570.

61. **Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ**. 2013. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci U S A **110**:20338–43.

62. **Morris JJ, Lenski RE, Zinser ER**. 2012. The Black Queen hypothesis: Evolution of dependencies through adaptive gene loss. MBio **3**:e00036–12–e00036–12.

63. **Morris JJ**. 2015. Black Queen evolution: The role of leakiness in structuring microbial communities. Trends Genet **31**:475–482.

64. **Lee M-C, Marx CJ**. 2012. Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet **8**:e1002651.

65. **Ragaller V, Hüther L, Lebzien P**. 2009. Folic acid in ruminant nutrition: a review. Br J Nutr **101**:153–64.

66. **Wolf YI, Koonin E V.** 2013. Genome reduction as the dominant mode of evolution. BioEssays **35**:829–837.

3.7 Tables and figures

Table 3.1 Anaerobic medium for culturing *Ruminococcus albus* strains 7 and 8

Ingredients	Concentration in Medium (mg/L)
Cellobiose	4000
$(NH_4)_2SO_4$	4000
Resazurin (0.1% w/v solution)	1 ml
Sodium bicarbonate	4000
Hemin	0.1
p-Aminobenzoic acid (*p*ABA)[a]	1
Folic acid[a]	1
Mineral Solution	
K_2HPO_4	300
KH_2PO_4	300
NaCl	600
$CaCl_2 \cdot 2H_2O$	60
$MgSO_4 \cdot 7H_2O$	60
Pfennig's Trace Elements Solution	
EDTA	0.5
$ZnSO_4 \cdot 7H_2O$	0.1
$MnCl_2 \cdot 4H_2O$	0.03
H_3BO_3	0.03
$CoCl_2 \cdot 6H_2O$	0.2
$CuCl_2 \cdot 2H_2O$	0.01
$FeCl_2 \cdot 4H_2O$	1.5
$NiCl_2 \cdot 6H_2O$	0.02
$Na_2MoO_4 \cdot 2H_2O$	0.03
Na_2SeO_3	0.01
Basic vitamin B Solution	
Thiamine-HCl	5
Ca-D-pantothenate	10
Nicotinamide	10
Riboflavin	5
Pyridoxine-HCl	10
Biotin	0.5
Vitamin B_{12}	0.5
VFA Solution	
Acetic acid (99.7% w/v)	0.137 ml
Propionic acid (99.5% w/v)	0.06 ml
Butyric acid (99% w/v)	0.04 ml
Isobutyric acid (99% w/v)	0.01 ml
2-methylbutyric acid (98% w/v)	0.01 ml
n-valeric acid (99% w/v)	0.01 ml
Isovaleric acids (99% w/v)	0.01 ml
3-Phenylpropionic acid (99% w/v)	3.75
Phenylacetic acid (99% w/v)	3.40
Cysteine Sulfide Solution	
Cysteine-HCl	500
$Na_2S \cdot 9H_2O$	500

[a] Either folic acid, *p*ABA, or none was added to the basic vitamin-containing medium.

Table 3.2 RNA-seq mapping table

RNAseq Sample ID	Total reads (avg. length)	Reads after Trimming[a] (avg. length)	Uniquely mapped reads[b] (%)	Non-specifically mapped reads (%)	Unmapped reads (%)
Both Vit-1	13,884,440 (100 nt)	13,849,013 (95 nt)	10,871,957 (78.5 %)	245,864 (1.8 %)	2,731,192 (19.7 %)
Both Vit-2	13,271,722 (100 nt)	13,232,202 (95 nt)	10,297,173 (77.8 %)	188,137 (1.4 %)	2,746,892 (20.8 %)
None-1	24,502,587 (100 nt)	24,344,150 (99 nt)	19,944,552 (81.9 %)	361,000 (1.5 %)	4,038,598 (16.6 %)
None-2	34,633,184 (100 nt)	34,470,450 (99 nt)	28,653,325 (83.1 %)	488,248 (1.4 %)	5,328,877 (15.5 %)

[a] Sequencing reads were trimmed using the CLC genomics Workbench program (ver. 6.5.1) with a quality score limit of 0.05 and a maximum number of ambiguities of 2.
[b] The trimmed reads were mapped to the *R. albus* 7 genome with annotation using the CLC genomic Workbench program with a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, a maximum number of 0.8, and a maximum number of hits for a read of 10.

Table 3.3 The putative folate synthesis genes on the genome of *R. albus* 7 and 8

R. albus 7

Locus # (Rumal)	Predicted function[a]	GenBank Accession	EFI server (*E*-value)	Pfam (*E*-value)	CDD (*E*-value)
0212	*folE*	ADU20769	6.90E-22	6.30E-58	7.68E-77
1922	*folQ*	ADU22416	0.008	1.30E-09	4.71E-06
1923	*folP*	ADU22417	7.60E-53	1.20E-64	1.50E-108
1924	Bifunctional *folB*/*folK*	ADU22418	2.30E-38	5.70E-41	3.19E-62
1336	*folC*	ADU21851	6.60E-37	1.40E-10	2.22E-105
2417	*folC*	ADU22897	1.50E-38	1.70E-10	5.48E-133
1925	*folA*	ADU22419	8.90E-19	6.70E-47	8.57E-57

R. albus 8

Locus # (Cus_)	Gene name	GenBank Accession	EFI server (*E*-value)	Pfam (*E*-value)	CDD (*E*-value)
5138	*folE*	EGC04024	7.70E-24	7.60E-60	3.86E-77
5042	*folC*	EGC02520	4.70E-35	6.10E-11	1.98E-112
5043	*folT*	EGC02525	3.70E-04	1.20E-11	9.65E-33
6073	*folT*	EGC02818	7.20E-06	1.10E-18	2.86E-16

R. albus SY3

Locus # (RaSY3_)	Gene name	GenBank Accession	EFI server (*E*-value)	Pfam (*E*-value)	CDD (*E*-value)
11620	*folE*	EXM38954	3.6E-23	4.2E-59	4.13E-76
16765	*folQ*	EXM37958	N/A	1.8E-10	1.16E-04
16760	*folP*	EXM37957	1.5E-53	4.8E-81	2.03E-93
16755	Bifunctional *folB*/*folK*	EXM37956	9.18E-32	2.5E-30	7.4E-42
07295	*folC*	EXM39633	5.7E-36	2.5E-07	2.37E-129
00625	*folC*	EXM40435	9.2E-36	2.2E-09	8.17E-120
16750	*folA*	EXM37955	9.6E-18	1.5E-46	1.57E-59
00620	*folT*	EXM40434	2.3E-4	4.4E-18	2.06E-28

[a] The function of genes was predicted using the Conserved Domain Database (CDD) and Pfam database if they exhibited significant amino acid sequence similarity (*E*-value < 1×10^{-5}) to biochemically characterized proteins in the Protein Data Bank (PDB). The EFI server was used to verify the CDD and Pfam annotations.

[b] No matches of query sequence found.

Table 3.4 Variation in the THF riboswitch controlled genes among *Ruminococcus* strains[a]

Species	Genome ID / location site	Downstream gene (Locus tag ID)	Genbank ID	Predicted Annotation (CDD)
R. albus 7	CP002403 /1531889-1531987	Rumal_1336	ADU21851	Folypolyglutamate synthase (FolC)
R. albus SY3	JEOB01000001 /150490-150602	RaSY3_00620	EXM40434	ECF transporter S component, folate family (FolT)
R. albus 8	ADKM02000092 /26627-26516	CUS_5043	EGC02525	ECF transporter S component, folate family (FolT)
R. albus 8	ADKM02000086 /59361-59461	CUS_6073	EGC02818	ECF transporter S component, folate family (FolT)
R. flavefaciens FD-1	ACOK01000004 /41711-41613	RflaF_ 010100001289	WP_009982980	Folypolyglutamate synthase (FolC)
R. flavefaciens 007c	ATAX01000028 /265603-265718	RF007C_15165	EWM52951	ECF transporter S component, folate family (FolT)

[a] Tetrahydrofolate (THF) riboswitches were predicted by Rfam database.

Table 3.5 Whole genome transcriptional response in *R. albus* 7 when grown on folate and *p*ABA lacking medium compared to both vitamins supplied in the medium

Global gene expression	Number of gene
Total number of gene	4032
Significantly up- or down-regulated genes ($p<0.05$)	1624
Up-regulated genes greater than 2 fold	672
Down-regulated genes greater than 2 fold	199

Table 3.6 Genes induced greater than 10-fold in *R. albus* 7 during growth in the absence of folate and *p*ABA, compared to the presence of folate and *p*ABA

Locus tag #	Predicted Annotation	Fold change	*p*-value
Rumal_0454	hypothetical protein	186.5	3.40E-11
Rumal_2402	hypothetical protein	74.1	3.66E-04
Rumal_3171	Peroxiredoxin	56.8	3.20E-04
Rumal_0673	hypothetical protein	45.2	8.88E-15
Rumal_3172	MarR family transcriptional regulator	35.4	6.95E-05
Rumal_2909	hypothetical protein	24.6	0
Rumal_1440	pseudo gene	20.5	1.01E-03
Rumal_2444	5- methyltetrahydropteroyltriglutamate /homocysteine S-methyltransferase	18.6	0
Rumal_2445	5,10-methylenetetrahydrofolate reductase	18.4	0
Rumal_1280	AMP-dependent synthetase and ligase	16.0	0
Rumal_1336	FolC bifunctional protein	15.7	0
Rumal_2896	degV family protein	13.9	0
Rumal_2104	family 1 extracellular solute-binding protein	12.9	1.65E-13
Rumal_3173	thioredoxin	11.1	9.55E-15
Rumal_2686	hypothetical protein	11.1	1.66E-03
Rumal_1541	hypothetical protein	10.4	3.50E-11
Rumal_0132	hypothetical protein	10.2	3.67E-05
Rumal_3854	dTDP-4-dehydrorhamnose 3,5-epimerase	10.0	5.12E-04

Table 3.7 Transcriptional response of the folate biosynthetic pathway in *R. albus* 7 to a lack of exogenous folate and *p*ABA**,** compared to the presence of folate and *p*ABA

			Normalized RPKM	
Locus tag (Rumal_#)	Annotation	Fold Change	Both	None
1922	Metal dependent phosphohydrolase (*folQ*)	-1.7	373.1	220.9
1923	Dihydropteroate synthase (*folP*)	1.8	159.3	283.8
1924	2-amino-4-hydroxy-6- hydroxymethyldihydropteridine pyrophosphokinase (*folK*)	1.4	134.7	189.5
1925	Dihydrofolate reductase region (*folA*)	NS[a]	254.9	241.7
0212	GTP cyclohydrolase I (*folE*)	3.4	36.3	124.7
1336	FolC bifunctional protein (*folC*)	15.7	26.0	406.5
2417	FolC bifunctional protein (*folC*)	1.8	53.5	98.4

[a] NS means fold change was not statistically significant ($p \geq 0.05$).

Table 3.8 Transcriptional response of genes involved in the central sugar metabolism to a lack of exogenous folate and *p*ABA in *R. albus* 7

			Normalized RPKM	
Locus tag (Rumal_#)	Predicted function	Fold Change	Both	None
EMP pathway				
2867	Glucokinase	NS[a]	229.7	181.8
2138	Glucose-6-phosphate isomerase	1.5	558.4	808.1
0431	6-phosphofructokinase	NS	371.5	300.3
1631	6-phosphofructokinase	3.2	69.9	223.0
2486	6-phosphofructokinase	1.2	1158.7	1415.6
2608	Fructose-bisphosphate aldolase	2.3	814.8	1871.9
0090	Triosephosphate isomerase	NS	712.1	937.6
627	Glyceraldehyde-3-phosphate dehydrogenase	NS	5465.6	5102.5
1028	Phosphoglycerate kinase	NS	1549.2	1550.5
0087	phosphoglycerate mutase	1.5	283.4	411.9
0088	Phosphoglycerate mutase	-2.0	876.8	449.2
2834	Enolase	NS	38.1	24.8
3948	Pyruvate kinase	4.6	58.0	265.1
PP pathway				
2948	Ribulose-phosphate 3-epimerase	NS	437.6	449.7
2220	Ribose-5-phosphate isomerase	4.4	147.0	652.0
1983	Transketolase, N-terminal section	1.4	442.4	603.5
1982	Transketolase, C-terminal section	NS	564.5	651.3
1602	Transaldolase	NS	9.4	5.1
TCA cycle				
0561	Citrate synthase	3.7	359.5	1318.6
3111	Aconitate hydratase	2.2	215.1	479.3
3116	Isocitrate dehydrogenase, NADP-dependent	1.5	554.0	816.1
0456	Malate dehydrogenase	1.7	199.8	331.1
0350	Fumarase, alpha subunit	1.9	42.9	81.7
3058	Fumarase, beta subunit	NS	933.1	976.5
2250	Fumarate reductase/succinate dehydrogenase	NS	3.3	5.6
2347	Fumarate reductase/succinate dehydrogenase	NS	88.4	100.8
3120	Fumarate reductase/succinate dehydrogenase	2.6	61.3	159.7
Anaplerotic pathway				
0626	Oxaloacetate decarboxylase	2.2	590.6	1282.2
1789	PEP carboxykinase	NS	2834.7	2591.2

[a] NS means fold change was not statistically significant ($p \geq 0.05$).

Table 3.8 (Cont.)

Locus tag (Rumal_#)	Predicted function	Fold Change	Normalized RPKM Both	Normalized RPKM None
Conversion of pyruvate to acetyl-coA				
1173	Pyruvate formate-lyase	1.3	1179.1	1582.7
1175	Pyruvate formate-lyase	-2.1	6496.2	3123.2
0032	Pyruvate ferredoxin oxidoreductase	2.2	1220.4	2705.6
3407	HydA2, ferredoxin-dependent hydrogenase	NS[a]	2163.1	2021.4
2964	HydA, electron-bifurcating hydrogenase	1.4	1272.1	1720.3
2965	HydB, electron-bifurcating hydrogenase	1.2	690.1	837.9
2966	HydC, electron-bifurcating hydrogenase	-1.6	1149.1	707.7
Fermentation pathway				
0279	Phosphate acetyltransferase	2.0	349.9	696.2
1651	Acetate kinase	2.6	478.0	1257.1
3401	Acetaldehyde dehydrogenase /Alcohol dehydrogenase	1.3	1810.8	2343.5
3118	Alcohol dehydrogenase	2.7	24.7	67.1
3178	Acetaldehyde dehydrogenase	1.3	1559.4	2096.6

[a] NS means fold change was not statistically significant ($p \geq 0.05$).

Table 3.9 Transcriptional response of genes involved in the purine metabolism to a lack of exogenous folate and *p*ABA in *R. albus* 7

Locus tag (Rumal_#)	Annotation	Fold Change	Normalized RPKM	
			Both	None
IMP synthesis				
1783	Ribose-phosphate pyrophosphokinase	2.4	282.8	667.0
2632	Phosphoribosyltransferase	NS[a]	10.0	49.9
0259	Amidophosphoribosyltransferase	NS	460.0	362.3
3942	Amidophosphoribosyltransferase	NS	6.6	3.8
0267	Phosphoribosylamine/glycine ligase	1.6	352.7	553.8
0265	Phosphoribosylglycinamide formyltransferase	NS	56.0	42.5
1739	Phosphoribosylformylglycinamidine synthase	1.3	351.6	467.1
0260	Phosphoribosylformylglycinamidine cyclo-ligase	NS	700.1	512.2
0256	Phosphoribosylaminoimidazole carboxylase, catalytic subunit	-3.9	398.9	102.3
0258	Phosphoribosylaminoimidazole-succinocarboxamide synthase	NS	398.6	502.5
2935	Adenylosuccinate lyase	2.9	119.3	348.7
0271	Phosphoribosylaminoimidazolecarboxamide formyltransferase	NS	441.9	428.0
0277	Inosine monophosphate cyclohydrolase-like protein	1.8	220.4	403.3
GTP synthesis				
3203	Inosine-5-monophosphate dehydrogenase	5.1	355.9	1827.8
1611	GMP synthase large subunit	2.3	397.2	896.3
1808	Guanylate kinase	1.2	269.4	320.9
3948	Pyruvate kinase	4.6	58.0	265.1
ATP synthesis				
1765	Adenylosuccinate synthetase	1.6	503.5	792.5
2935	Adenylosuccinate lyase	2.9	119.3	348.7
0057	Adenylate kinase	-2.3	2748.3	1217.1
3948	Pyruvate kinase	4.6	58.0	265.1

[a] NS means fold change was not statistically significant ($p \geq 0.05$).

Table 3.10 Transcriptional responses of genes involved in the nitrogen metabolism to a lack of exogenous folate and *p*ABA in *R. albus* 7

			Normalized RPKM	
Locus tag (Rumal_#)	Annotation	Fold Change	Both	None
2461	Glutamate dehydrogenase (*gdh*)	3.0	275.8	824.7
2615	Glutamine synthetase catalytic subunit (*glnA*)	1.2	77.1	90.6
0216	Glutamine synthetase catalytic subunit (*glnN*)	1.6	337.1	529.8

Table 3.11 Transcriptional responses of genes involved in the one carbon pool metabolism by folate in *R. albus* 7 to a lack of exogenous folate and *p*ABA, compared to the presence of folate and *p*ABA

			Normalized RPKM	
Locus tag (Rumal_#)	Annotation	Fold Change	Both	None
2224	Formate-tetrahydrofolate ligase	2.1	262.5	553.9
3726	Homocysteine S-methyltransferase	1.4	200.8	272.1
2050	Glycine hydroxymethyltransferase	2.0	773.5	1540.4
1959	Thymidylate synthase, flavin-dependent	1.1	298.6	327.0
3207	Methylenetetrahydrofolate dehydrogenase (NADP(+))	1.4	189.2	254.7
2445	5,10-methylenetetrahydrofolate reductase	18.4	142.3	2617.4
2778	5,10-methylenetetrahydrofolate reductase	7.2	97.8	700.8

Table 3.12 Transcriptional response of the Shikimate pathway to a lack of exogenous folate and *p*ABA in *R. albus* 7

			Normalized RPKM	
Locus tag (Rumal_#)	Annotation	Fold Change	Both	None
0234	DAHP synthase (*aroA*)	2.6	37.5	99.0
0379	DHQ synthase-I (*aroB*)	1.8	22.1	38.8
0442	Dehydroquinate dehydratase (*aroC*)	1.5	97.8	143.7
0018	Shikimate dehydrogenase (*aroD*)	3.8	45.2	172.6
0139	Shikimate kinase (*aroEI*)	2.2	36.3	80.3
0380	EPSP synthase (*aroF*)	NS[a]	36.6	31.7
0386	Chorimate synthase (*aroG*)	1.9	198.7	383.7

[a] NS means fold change was not statistically significant ($p \geq 0.05$).

Table 3.13 Transcriptional responses of the biosynthesis pathways for aromatic amino acids to a lack of exogenous folate and *p*ABA in *R. albus* 7

			Normalized RPKM	
Locus tag (Rumal_#)	Annotation	Fold Change	Both	None
Trytophan synthesis				
1716	Anthranilate phosphoribosyltransferase	NS	168.0	127.5
1717	Anthranilate synthase component I	NS	187.2	167.6
1862	Phosphoribosylanthranilate isomerase	-3.1	132.1	42.6
1715	Indole-3-glycerol-phosphate synthase	NS	153.5	65.9
1717	Anthranilate synthase component I	NS	187.2	167.6
1712	Tryptophan synthase subunit alpha	NS	509.2	254.9
1713	Tryptophan synthase subunit beta	NS	165.3	158.8
Tyrosine synthesis				
3158	Chorismate mutase	3.0	152.1	454.8
1172	Prephenate dehydrogenase	1.8	146.3	265.5
0224	Histidinol-phosphate aminotransferase	2.0	62.4	124.2
Phenyl-alanine synthesis				
3158	Chorismate mutase	3.0	152.1	454.8
2960	Prephenate dehydratase	2.1	128.8	265.3
0224	Histidinol-phosphate aminotransferase	2.0	62.4	124.2

[a] NS means fold change was not statistically significant ($p \geq 0.05$).

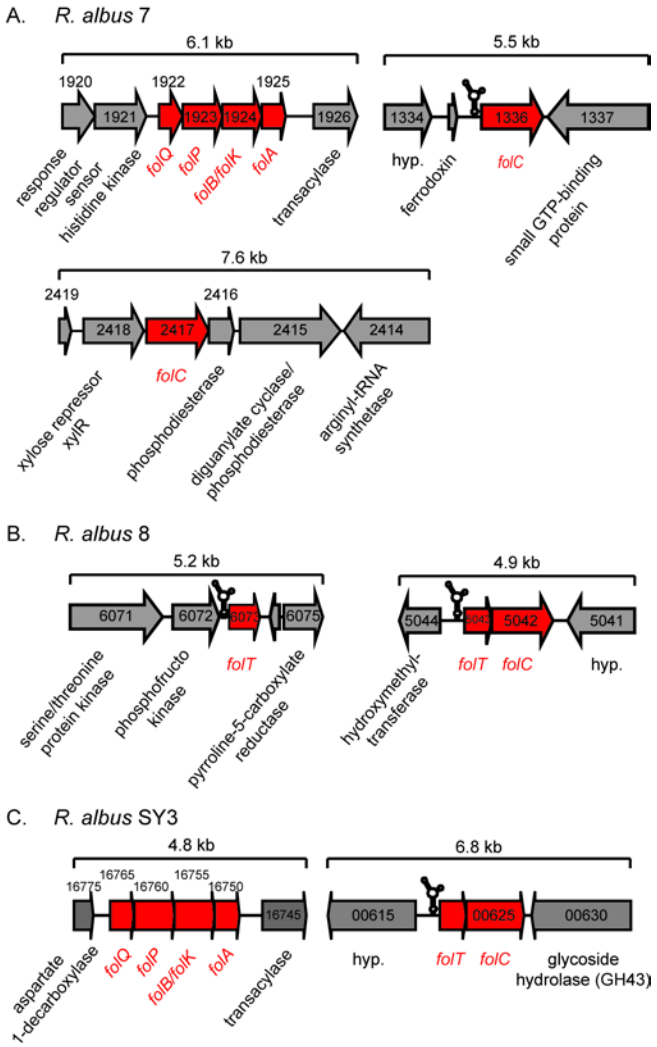


Fig. 3.1 Genes involved in either the folate biosynthetic or salvage pathways of *R. albus* strain 7 (*A*), strain 8 (*B*) and strain SY3 (*C*). Four tetrahydrofolate (THF) riboswitches were predicted in *R. albus* strains (one upstream of *folC* gene in strain 7, two upstream of *folT* genes in strain 8, and one upstream of *folT* gene in strain SY3, respectively).

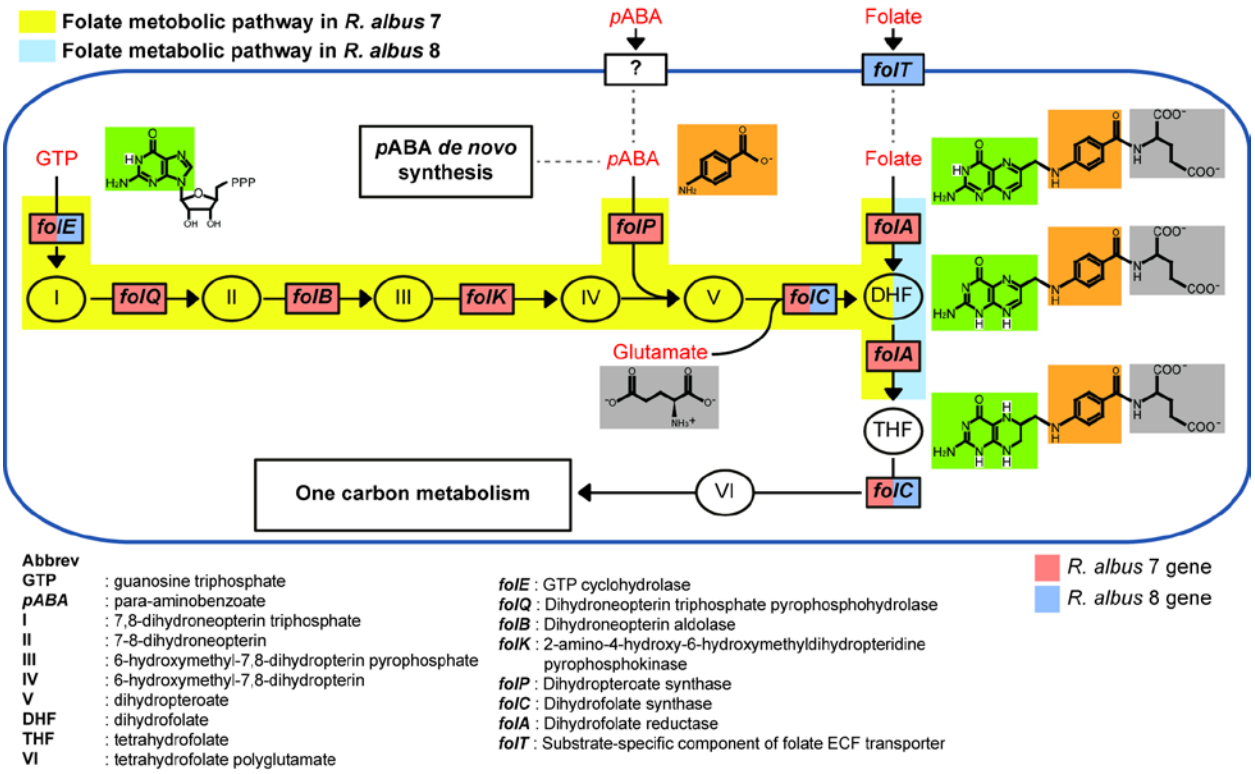


Fig. 3.2 Schematic diagram of *de novo* folate biosynthetic and salvage pathways. The yellow line presents the predicted folate biosynthetic pathway by condensation of the pterin branch with *p*ABA and glutamate in *R. albus* 7, while the blue line presents the predicted direct folate salvage pathway through FolT dependent ECF transporter in *R. albus* 8. In this Figure, genes present on the strain 7 genome are shown in red, while genes present on the strain 8 genome are shown in blue. 'Abbrev' defines the chemical names described in the figure above.

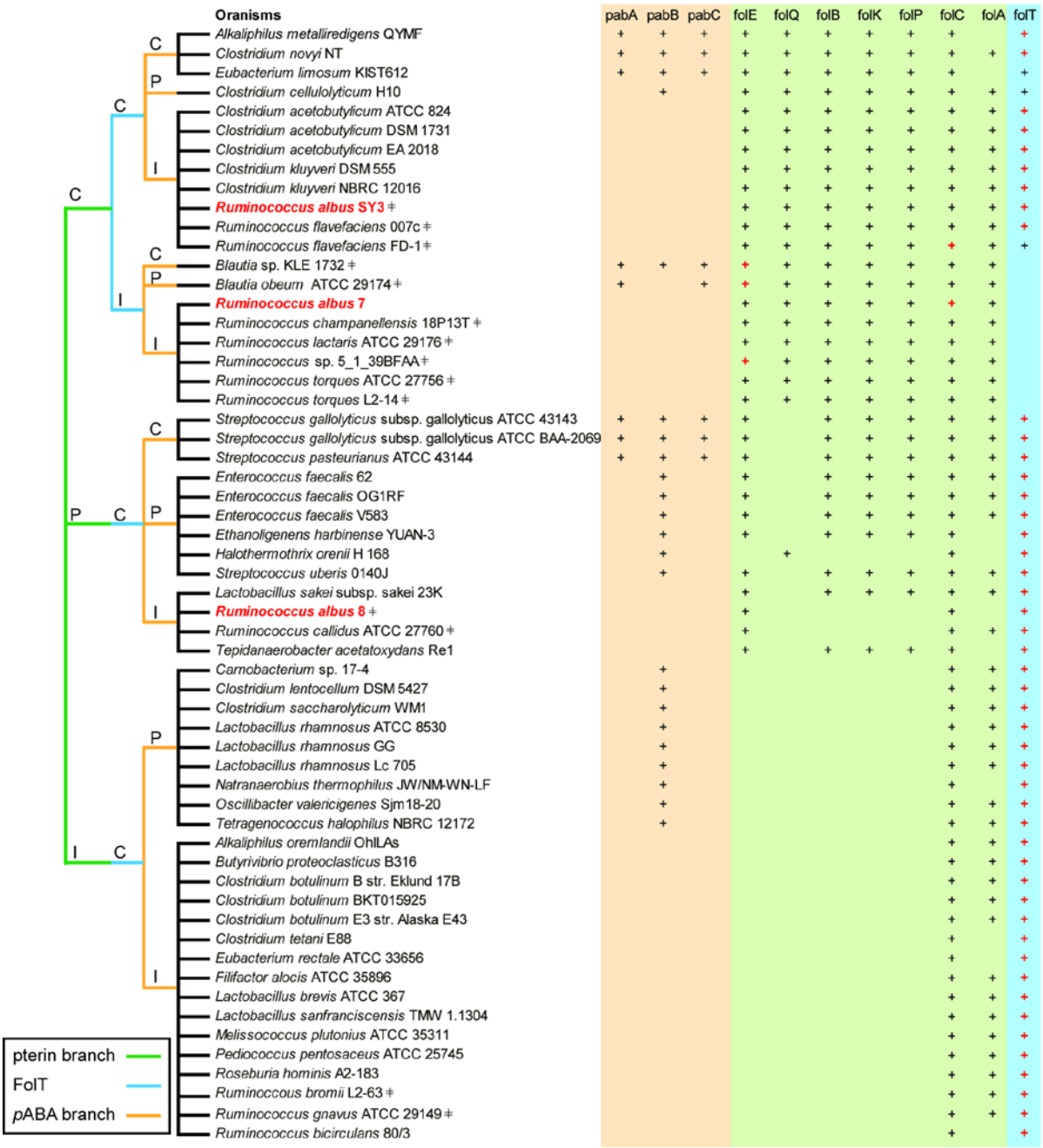
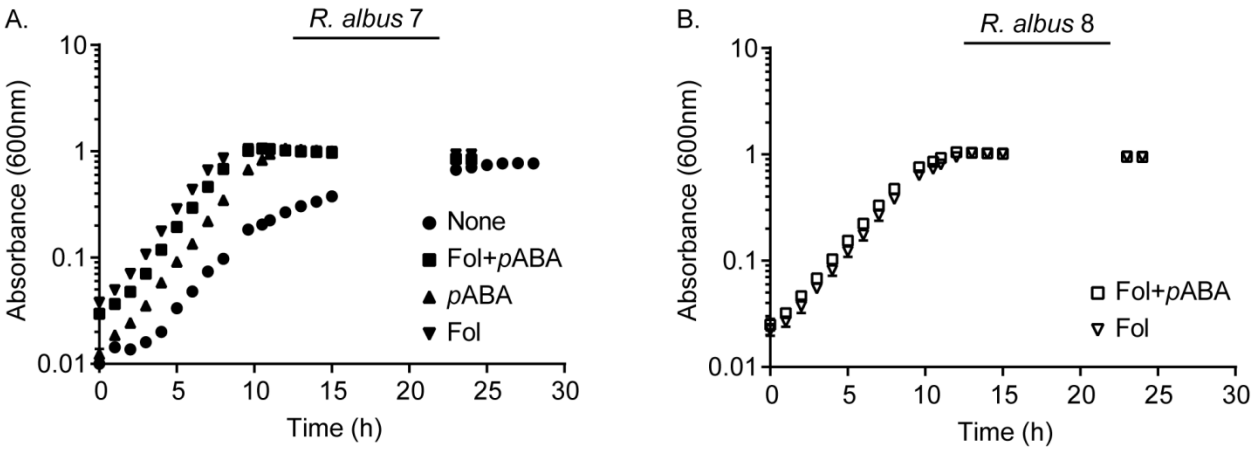


Fig. 3.3 Presence of putative genes for the folate biosynthetic and salvage pathways in isolated strains of *Ruminococcus* and other bacterial strains encoding a THF riboswitch. A THF riboswitch controlled gene on bacterial genome sequences was predicted using the Rfam database. Among the strains on Rfam database except *Ruminococcus* strains, only strains that have complete genome sequences were analyzed and shown in order of the presence of each folate biosynthesis, salvage, and *p*ABA synthesis pathways. 'C' means the genome sequences of strains possessing complete biosynthetic pathway, 'P' means the genome sequences of strains possessing partial biosynthetic pathway, and 'I' means the genome sequences of strains missing most of genes in the pathway. Red cross represents the genes bearing a THF riboswitch at upstream region. ‡ Denotes draft genome sequence.



C.

	R. albus 7			*R. albus* 8		
	Max *A* (600nm)	Growth rate (h^{-1})	Doubling time (h)	Max *A* (600nm)	Growth rate (h^{-1})	Doubling time (h)
None	0.78 ± 0.02	0.36 ± 0.07	2.01 ± 0.45		No growth	
Fol+ *p*ABA	1.06 ± 0.00	0.48 ± 0.05	1.47 ± 0.16	1.04 ± 0.01	0.41 ± 0.05	1.72 ± 0.19
*p*ABA	1.07 ± 0.02	0.45 ± 0.04	1.56 ± 0.14		No growth	
Fol	1.05 ± 0.02	0.47 ± 0.04	1.48 ± 0.13	1.06 ± 0.01	0.40 ± 0.03	1.76 ± 0.13

Fig. 3.4 Growth curve of *R. albus* 7 and 8 grown in defined medium supplemented with either folate, *p*ABA or both. Absorbance at 600 nm (A_{600nm}) values were measured every hour. Growth curves of *R. albus* 7 and *R. albus* 8 are shown in panel (*A*) and (*B*), respectively. *Error bars* represent standard deviations for three biological replicates. The maximum absorbance value, growth rate and doubling time of *R. albus* 7 and 8 are presented in panel (*C*).

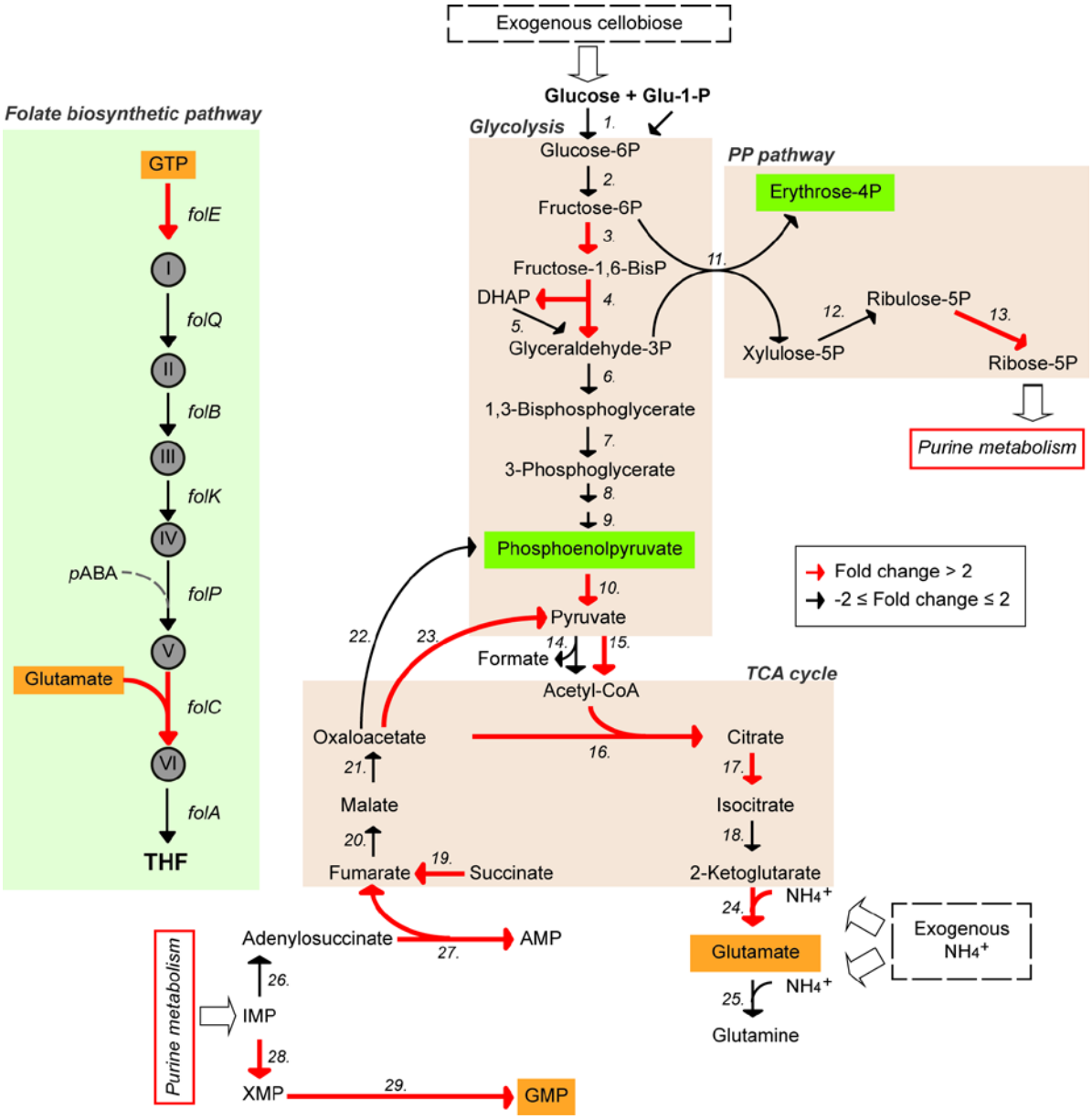


Fig. 3.5 Integrative transcriptional profiles of the central sugar metabolic pathways and folate biosynthesis pathway in *R. albus* 7 grown in the presence or absence of folate and *p*ABA. The locus tag number and the expression level of predicted catalytic genes in the pathway are shown in Table 3.8 to 3.10. The precursors for the shikimate and folate synthesis pathways are highlighted in green and orange, respectively. The genes up-regulated greater than 2 fold in the absence of both vitamins, relative to the presence of both vitamins, are shown as *red arrows*. Enzyme name: *1.* Glucokinase and phosphoglucomutase; *2.* Glucose-6-phosphate isomerase; *3.* 6-phosphofurctokinase; *4.* Fructose-bisphosphate aldolase; *5.* Triosephosphate isomerase; *6.* Glyceraldehyde-3-phosphate dehydrogenase; *7.* Phosphoglycerate kinase; *8.* Phosphoglycerate mutase; *9.* Enolase; *10.* Pyruvate kinase; *11.* Transketolase; *12.* Ribulose-5-phosphate 3-epimerase; *13.* Ribose-5-phosphate isomerase; *14.* Pyruvate formate lyase; *15.* Pyruvate ferredoxin oxidoreductase; *16.* Citrate synthase; *17.* Aconitate hydratase; *18.* Isocitrate dehydrogenase; *19.* Fumarate reductase/succinate dehydrogenase; *20.* Fumarase; *21.* Malate dehydrogenase; *22.* PEP carboxykinase; *23.* Oxaloacetate decarboxylase; *24.* Glutamate dehydrogenase; *25.* Glutamine synthetase; *26.* Adenylosuccinate synthetase; *27.* Adenylosuccinate lyase; *28.* Inosine-5-monophosphate dehydrogenase; *29.* GMP synthase.

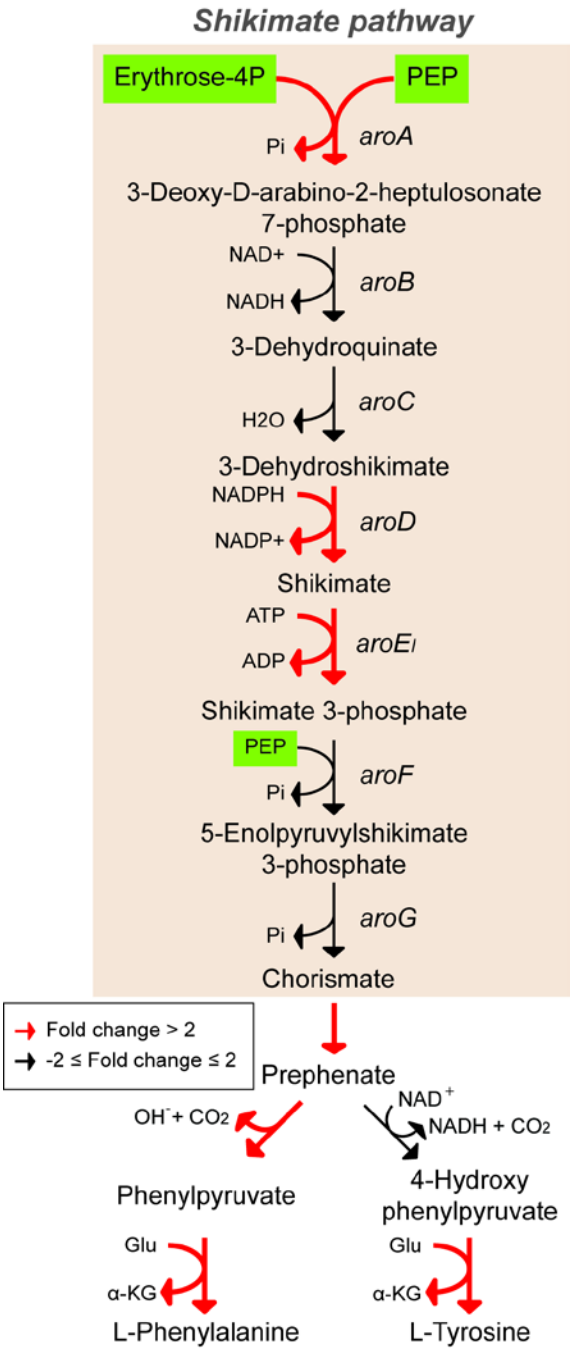


Fig. 3.6 Integrative transcriptional profiles of the biosynthesis pathway for chorismate, phenylalanine and tyrosine in *R. albus* 7 grown in the presence or absence of folate and *p*ABA. The locus tag number and the expression level of predicted catalytic genes in the pathway are shown in Table 3.12 and 3.13. The precursors for the shikimate and folate synthesis pathways are highlighted in green and orange, respectively. The genes up-regulated greater than 2 fold in the absence of both vitamins, relative to the presence of both vitamins, are shown as *red arrows*.

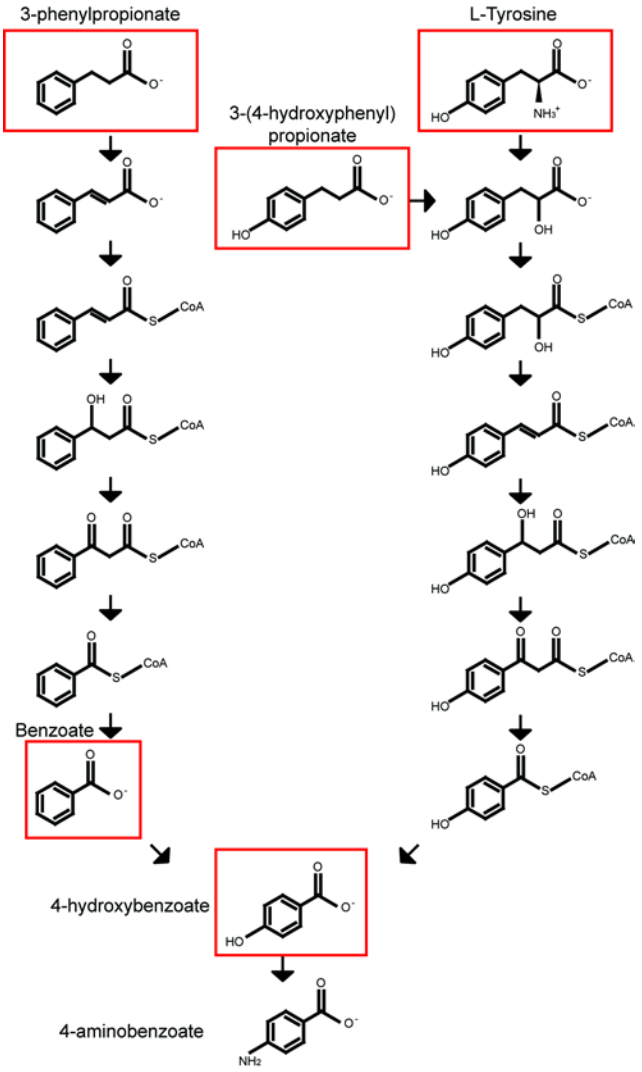


Fig. 3.7 Hypothetical pathways for *de novo* synthesis of *p*ABA using 3-phenylpropionate or tyrosine as precursor. Red highlighted molecules were exogenously added to a bacterial culture medium and tested for the growth stimulating effects on *R. albus* 7 in the absence of folate and *p*ABA.

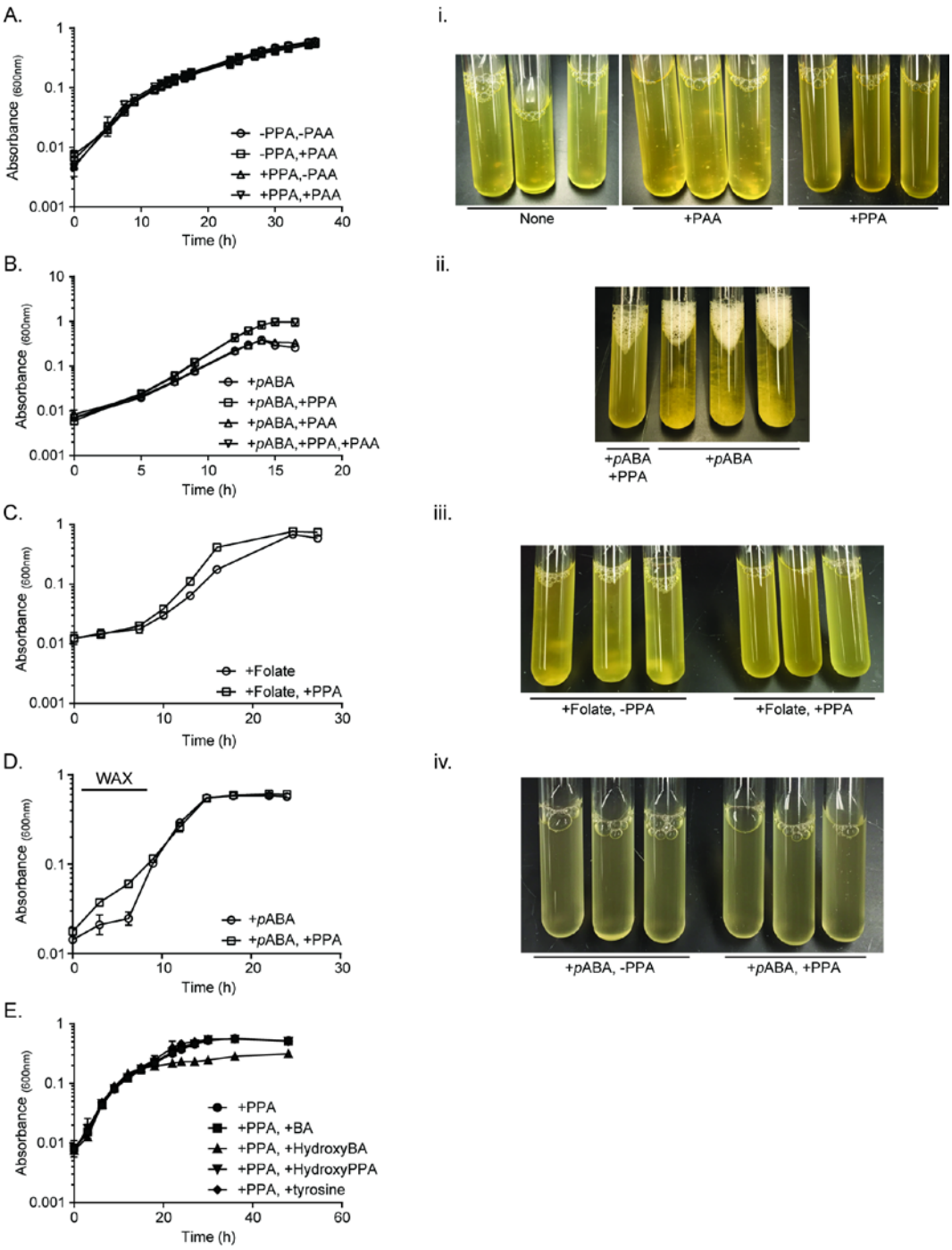
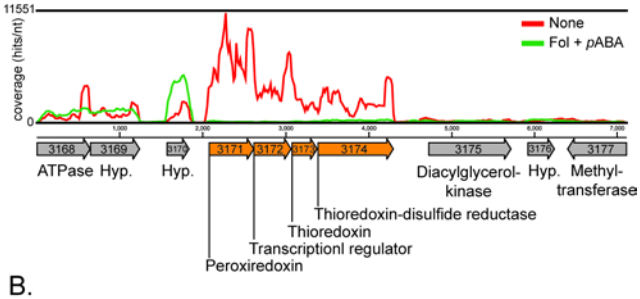


Fig. 3.8 Growth curve of *R. albus* 7 in a defined medium with a potential precursor for *p*ABA synthesis. *R. albus* 7 was grown in a defined medium containing either 0.4 % (w/v) of cellobiose (*A*, *B*, *C*, and *E*) or 0.4 % (w/v) of wheat arabinoxylan (WAX; *D*) as the sole carbohydrate source in the absence of folate and *p*ABA. After autoclaving, filter-sterilized potential precursor for *p*ABA synthesis was added in the medium, bringing the final concentration to 25 µM. Abbreviation: phenylpropionic acid (PPA), phenylacetic acid (PAA), benzoic acid (BA), tyrosine, hydroxylbenzoic acid (hydroxyBA), and hydroxyphenypropionic acid (HydroxyPPA)

A.



B.

Locus ID (Rumal_#)	GenBank ID	Fold change	*p*-value	RPKM (None)	RPKM (Both)
3171	ADU23635	56.8	3.2E-04	2,435	43
3172	ADU23636	35.4	7.0E-05	1,220	35
3173	ADU23637	11.1	9.6E-15	680	61
3174	ADU23638	7.5	2.4E-07	973	129

Fig. 3.9 RNA-seq coverage and fold change for the putative peroxiredoxin cluster in *R. albus* 7. (*A*) A detailed view of the nucleotide coverage for the putative peroxiredoxin cluster in *R. albus* 7 is shown. (*B*) The fold change and expression levels (RPKM: reads per kilo base per million mapped reads) of the cluster were calculated from two biological replicates in each experimental group (both folate and *p*ABA supplied or none supplied).

CHAPTER 4.

IDENTIFICATION OF LONG NON-CODING RNA LOCI CONSERVED IN *RUMINOCOCCUS ALBUS* STRAINS: A POTENTIAL TRANSCRIPTIONAL REGULATOR RELATED TO STATIONARY PHASE

4.1 Introduction

Many transcriptional studies based on RNAseq have identified a significant number of transcripts that do not code for proteins in bacteria and eukaryotes, including mammals (1–5). In bacteria, non-coding RNAs (ncRNAs) are generally small (50-250 nucleotides; nt)(6), and therefore they are often termed small RNAs (sRNAs). To date, many sRNAs have been characterized with diverse regulatory functions. The most-studied class of regulatory sRNAs activate or repress translation or the stability of their target transcripts by base pairing, in response to various environmental signals (reviewed in reference 7). Bifunctional sRNAs regulate their targets through base pairing interaction as well as production of a small protein encoded downstream of the base-pairing region (8). The last class is protein-binding sRNAs that sequester transcriptional or translational regulators and subsequently inhibit their activities. Unlike small sized ncRNAs in bacteria, considerably longer ncRNAs between 200 nt to even 118 kb in length have been exclusively found in eukaryotes, including mammals (5, 9). The functions of long ncRNAs (lncRNAs) are still unclear in eukaryotes and are relatively unknown and undocumented in bacteria. Most of lncRNAs are transcribed by RNA polymerase II, like messenger RNA (mRNA); however, they do not go through the subsequent translation steps. To date, lncRNAs have been known to regulate their target genes through generally five different interactions depending on the association sites of lncRNAs with their target gene (i.e. promoter,

gene body, enhancer or intervening regions vs sense or antisense). In bacteria, some characterized sRNAs are longer than general sRNAs (e.g. CsrB RNA: 350 nt and RNase P RNA: 400 nt)(5, 10). Nevertheless, to our knowledge ncRNAs longer than 1 kb have yet to be reported in bacteria, while lncRNAs longer than 1 kb are commonly found in eukaryotes (5). As a result, the roles of lncRNAs in the regulatory system of eukaryotes have received extensive attention. It is, therefore, important to identify the potentially functional lncRNAs in bacteria. Such an analysis will help determine whether the lncRNAs mediated regulatory systems are exclusively employed by eukaryotes, and furthermore provide a better understanding of ncRNAs mediated regulatory systems in bacteria. Herein, we report the unique lncRNA loci conserved in *R. albus* strains and their growth dependent expression pattern based on integrated genomic and transcriptomic analyses.

4.2 Material and methods

Identification of conserved long noncoding RNA in *R. albus* strains. Genomic sequences of *R. albus* 7 (GenBank accession numbers: NC_014833 for chromosome and NC_014824 to NC_014827 for plasmids), *R. albus* 8 (NZ_ADKM020000001 to NZ_ADKM02000136), and *R. albus* SY3 (JEOB01000001 to JEOB01000004) were used for comparative genomic analysis of conserved long noncoding RNA (lncRNA) in *R. albus* strains. The functional prediction of the neighboring genes of lncRNA was performed using the RAST server (11), Pfam database (12), and Rfam database (13). To identify the lncRNA expressed in *R. albus* 7 and 8 during growth on alkaline hydrogen peroxide treated corn stalk (AHPCS), the unmapped RNA-seq reads to the

annotated reference genome were re-mapped onto the reference sequence without annotation using the CLC genomics workbench version 5.5.1 from CLC Bio (Cambridge, MA).

Computational analysis of lncRNA region. Tandem repeats in lncRNA region in *R. albus* strains 7, 8, and SY3 were analyzed using an online software, Tandem repeats finder (14). Transcription promoters and terminators of lncRNA region were predicted using the PePPER and ARNold webservers, respectively (15, 16).

RNA extraction and Northern blot hybridization. *R. albus* strains 7 and 8 were grown in a biochemically defined medium containing AHPCS. When both strains reached stationary phase (24 hours incubation), inferred from previous experiments, the cells of strain 7 and 8 were harvested by centrifugation at 13,000 × *g* for 10 min at 4°C. The resulting cell pellets were stored at – 80 °C until RNA extraction. In the subsequent steps, the cell pellets were treated with lysis buffer (200U/ml of mutanolysin, 200 µg/ml of lysozyme, 150µg/ml of proteinase K, 25mM EDTA, and 0.5% (w/v) SDS) for 30 minute at 55 °C (17). The total RNA was extracted with the RNeasy mini kit (Qiagen) with the optional on-column DNase treatment step. Then, the total RNA was eluted with DEPC-treated nuclease-free water and stored at – 80 °C until RNA sequencing.

Total RNA (1µg per lane) of *R. albus* 7 or strain 8 was separated in 1.2 % (w/v) of agarose/formaldehyde gels, followed by transfer onto BrightStar®-Plus membranes (Ambion, Austin, Texas, USA) by capillary transfer using the NorthernMax® blotting and hybridization kit (Ambion). The RNA probe for the detection of lncRNA was synthesized and biotinylated using the MAXIscript® kit (Ambion) with primers listed in Table 4.1. After overnight hybridization with a biotin-labelled probe, the targeted lncRNA was visualized using the Biotin Chromogenic

Detection kit (Fermentas, Burlington, Ontario, Canada) according to the manufacturer's instructions.

Computational analysis of DUF1292 domain in *R. albus* strains 7, 8, and SY3. In order to obtain a better functional prediction of the DUF1292 domain, the structures of the DUF1292 encoding proteins were modeled by the Phyre2 protein fold recognition server (18) using amino acid sequences of Rumal_3117 in strain 7, CUS_6629 in strain 8, and RASY3_05270 in strain SY3. Pfam database was used to identify the orthologous genes within the bacterial kingdom.

4.3 Results and discussion

The novel ncRNA region highly expressed in *R. albus* 7 and 8 during growth on AHPCS. After mapping the RNA sequencing (RNA-seq) reads of strain 7 and 8 grown on AHPCS to the annotated region of the reference genome, we found that non-mapped reads were highly abundant (> 35 % of a total transcripts) in the transcriptomes of *R. albus* 7 and 8 at late-log and stationary phases (Table 4.2). It is possible that this was the results of a contaminant RNA during bacterial culture or library preparation. However, when we re-mapped those non-mapped RNA-seq reads to the reference sequence without annotation, most of reads mapped onto the reference sequence (> 93 % of a total non-mapped reads) (Table 4.3), showing high expression of non-coding RNAs (ncRNA) in *R. albus* 7 and 8 during late-log to stationary phase on AHPCS. Interestingly, we found that the abundant ncRNAs were primarily expressed from a specific region on the genomes of *R. albus* strain 7 and 8 (Fig. 4.1-4.2). The long ncRNA (lncRNA), inferred from the size of the non-coding region (> 500 nt), was clustered together with the DUF1292 gene, a non-coding region and a putative alcohol dehydrogenase in *R. albus* strains 7

and 8. The non-coding region in strain 7 is annotated as 6S RNA (Rumal_R0067; Gene ID: 10079406) in the NCBI database based on Rfam prediction (Score: 37.67). The 6S RNA (or SsrS RNA) is a ncRNA that binds Eσ70 and consequently inhibits RNA polymerase during stationary phase of *Escherichia coli* (19, 20). In accordance with the predicted function of 6S RNA, the expression pattern of Rumal_R0067 gene in strain 7 and its homologous region in strain 8 responded in a growth dependent manner, similar to the lncRNA (Fig. 4.3). Since the lncRNA loci, composed of lncRNA, DUF1292 gene, 6S RNA, and alcohol dehydrogenase, was conserved in strains 7 and 8, we sought to find the homologous region in *R. albus* strain SY3 based on its genome sequence in the NCBI database. As shown in Fig. 4.4, all three strains possess a very similar lncRNA loci arranged in the same gene order, suggesting that the lncRNA loci is conserved in *R. albus* strains.

Sequence feature of the lncRNAs in *R. albus* strains. The transcription of lncRNA in *R. albus* 8 terminated at the end of contig00066 (Accession: ADKM02000093); therefore, we sought to find the missing region in the draft genome sequence of *R. albus* 8 by gap sequencing. The lncRNA region in strain 7 contained three repeat sequences that are of 216 bp size. Since we found that the beginning of contig00128 (Accession: ADKM02000120) in strain 8 contained a long non-coding region of four repeat sequences (Fig. 4.5), we designed primers to amplify the region between the end of contig00066 and the beginning of contig00128 (Fig. 4.6A). The amplicon size was approximately 3.5kb (Fig. 4.6B), which is larger than the size of the lncRNA region in strains 7 and SY3 (~ 2 kb). After sequencing the amplicon, it was discovered that the intergenic region between CUS_6629 and CUS_5685 genes in strain 8 contains a predicted transposase (0.8 kb) (Fig. 4.7), suggesting that the insertion event of the transposable element into the genome of strain 8 occurred after speciation. It is notable that the lncRNA regions of

strains 7, 8, and SY3 possessed similar genomic architecture characterized by tandem repeat sequences. The tandem repeat sequences in lncRNA is considered to be one of distinct features from other ncRNA (21, 22). In mammalian genomes, tandem repeat sequences together with transposon elements are more commonly found in lncRNA sequences compared to protein-coding genes (21, 22). It has been proposed that these elements assist the function of lncRNA potentially through base pairing with other RNAs bearing similar repeat sequences or through a yet unknown mechanism (23, 24). Taken together, it seems that these elements of lncRNA are conserved not only in eukaryotes, but also in bacteria.

Transcriptionally functional lncRNAs in *R. albus* 7 and 8. In addition to the RNAseq based transcriptomic analysis, we verified the presence of transcripts of lncRNA in *R. albus* 7 and 8 grown on AHPCS using northern blot analysis with probes that either detected the tandem repeat sequence or the predicted promoter region in lncRNA (Fig. 4.7). To determine the direction of the transcription of lncRNA, either sense or antisense probe was hybridized with total RNA extracted from strains 7 and 8 grown for 24 hours on AHPCS. The lncRNA in strain 7 was hybridized with a sense probe while the lncRNA in strain 8 was hybridized with an antisense probe (Fig. 4.8), which is in accordance with the predicted transcriptional direction based on the location of their cognate promoters. The predicted transcript sizes from the transcription promoter to the terminators in strain 7 and 8 were 1715 bp and 1593 bp, respectively. The northern blot analysis showed that the detected sizes of transcript in strain 7 by all three probes were similar to the predicted transcript size of the lncRNA. Interestingly, the size of lncRNA transcript in strain 8 differed from those in strain 7. In strain 8, the detected size of transcript by the probe binding to promoter region upstream of the transposase was similar to the predicted size of transposase containing non-coding region (1.7 kb), while the tandem repeat detecting

probe hybridized to a shorter size of transcript (0.8 kb), which is similar in size to the total length of the four tandem repeat sequences downstream of the transposase. This result suggests that the intergenic region between CUS_6629 and CUS5685 is transcriptionally partitioned by the integrated transposase and that the transcript of lncRNA downstream of transposase is likely processed to small transcripts containing the tandem repeat elements.

A potential ncRNA regulator, DUF1292 protein. Inferred from the expression pattern of the lncRNA loci throughout the growth on AHPCS, it seems likely that the lncRNA loci are involved in the transcriptional regulation of cell metabolism related to stationary phase. Of two protein-coding genes in lncRNA loci, the DUF1292 gene was the most highly expressed in both strains at early-log phase (8h) and subsequently repressed as cells grew. However, the putative alcohol dehydrogenase gene had a low transcriptional level throughout the growth on AHPCS (Table 4.4), and ethanol was accumulated in both strains cultures during stationary phase in the previous experiment. Taken together, the putative alcohol dehydrogenase gene in the lncRNA loci appears not to be transcriptionally functional. To gain a better insight into the function of the lncRNA loci, we investigated the highly expressed DUF1292 gene in the lncRNA loci. Since none of protein encoding DUF1292 has been characterized to date, we predicted the protein structure of DUF1292 gene using the Phyre2 server. The Phyre2 server generates a protein structure model based on the amino acid sequence and matches it with the known protein structures in their library (18). The predicted structures of DUF1292 gene in strains 7, 8, and SY3 consisted of two alpha-helices and two beta-sheets (Fig. 4.9). Despite low confidence score and sequence identity, the best hit protein structure for DUF1292 predicted by the Phyre2 was a double-strand RNA binding domain (Fig. 4.9 left panel). Since the neighboring putative 6S RNA gene was predicted to contain several stem-loop structures (Fig. 4.10), it seems possible that the

DUF1292 interacts with the transcript of the putative 6S RNA. However, the expression pattern of DUF1292 genes during growth of *R. albus* 7 and 8 on AHPCS were opposite to the increased pattern of the neighboring ncRNAs (Table 4.4), suggesting that the DUF1292 protein interacts and inhibits the transcription and/or function of the neighboring lncRNA and/or putative 6S RNA. Since ncRNA-mediated regulatory mechanism, in relation to the stationary phase of growth, is poorly understood in Gram-positive Firmicutes, we sought to find a reference mechanism in Gram-negative bacteria that have been relatively well-characterized for ncRNA regulation. In *E. coli*, the Csr (carbon storage regulator) system is known to regulate carbon metabolism and diverse traits including motility, quorum sensing, and biofilm development (reviewed in references 7 and 25). The central component of the Csr system, CsrA, is an RNA-binding protein that represses metabolic processes related to stationary phase, such as glycogen synthesis and gluconeogenesis, and facilitates metabolic processes related to exponential phase, such as glycolysis and motility. The CsrA activity is antagonized by two ncRNAs, i.e., CsrB and CsrC. The CsrD protein mediates the degradation of CsrB and CsrC by RNase E (26). Thus, the decrease in activity of CsrD results in facilitating the functions of CsrB and CsrC, leading to the stimulation of pathways related to stationary phase. The orthologous genes to Csr regulatory genes in Gram-negative bacteria are not found in *R. albus* strains. In addition, DUF1292 genes are predominantly present in the genomes of Gram-positive Firmicutes, and not in the organisms known to possess the Csr system, including *E. coli*, *Pseudomonas aeruginosa*, and *Vibrio cholerae* (Fig. 4.11). It is possible that the protein encoding DUF1292 domain may function as a unique regulator of Firmicutes targeting ncRNAs that are highly expressed during late log-stationary phase.

In summary, we have determined that lncRNA loci, consisting of four sequence components (lncRNA, DUF1292 gene, putative 6S RNA, and alcohol dehydrogenase), is conserved in *R. albus* strains. Based on their transcriptional profiles assessed by RNA-seq and northern blot analyses, it seems likely that the lncRNA loci are involved in the regulatory system related to the stationary phase of the *R. albus* cells. The predicted protein model for the DUF1292 gene suggests that the DUF1292 is a unique ncRNAs regulator exclusively employed by Firmicutes. To the best of our knowledge, this is the first report that demonstrates the lncRNA in bacteria with a growth dependent transcriptional pattern. For better understanding of the regulatory mechanism in bacteria, more research is required to determine the functional characterization of lncRNAs.

4.4 References

1. **Montzka Wassarman K, Zhang A, Storz G**. 1999. Small RNAs in *Escherichia coli*. Trends Microbiol **7**:37–45.

2. **Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR**. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science **296**:916–9.

3. **Gottesman S**. 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet **21**:399–404.

4. **Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES**. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature **458**:223–227.

5. **Cech TR, Steitz JA**. 2014. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. Cell **157**:77–94.

6. **Vogel J, Wagner EGH**. 2007. Target identification of small noncoding RNAs in bacteria. Curr Opin Microbiol **10**:262–270.

7. **Bobrovskyy M, Vanderpool CK**. 2013. Regulation of Bacterial Metabolism by Small RNAs Using Diverse Mechanisms. Annu Rev Genet **47**:209–232.

8. **Wadler CS, Vanderpool CK**. 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci **104**:20454–20459.

9. **Bonasio R, Shiekhattar R**. 2014. Regulation of Transcription by Long Noncoding RNAs. Annu Rev Genet **48**:433–455.

10. **Altuvia S**. 2007. Identification of bacterial small non-coding RNAs: experimental approaches. Curr Opin Microbiol **10**:257–261.

11. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

12. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M**. 2014. Pfam: the protein families database. Nucleic Acids Res **42**:D222–30.

13. **Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD**. 2015. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res **43**:D130–7.

14. **Benson G**. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27**:573–580.

15. **De Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J**. 2012. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. BMC Genomics **13**:1–10.

16. **Naville M, Ghuillot-Gaudeffroy A, Marchais A, Gautheret D**. 2014. ARNold: A web tool for the prediction of Rho-independent transcription terminators. RNA Biol **8**:11–13.

17. **Rakotoarivonina H, Larson MA, Morrison M, Girardeau JP, Gaillard-Martinie B, Forano E, Mosoni P**. 2005. The *Ruminococcus albus* pilA1-pilA2 locus: Expression and putative role of two adjacent pil genes in pilus formation and bacterial adhesion to cellulose. Microbiology **151**:1291–1299.

18. **Mezulis S, Yates CM, Wass MN, E Sternberg MJ, Kelley LA**. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc **10**.

19. **Brownlee GG**. 1971. Sequence of 6S RNA of *E. coli*. Nat New Biol **229**:147–149.

20. **Wassarman KM, Storz G**. 2000. 6S RNA Regulates *E. coli* RNA Polymerase Activity. Cell **101**:613–623.

21. **Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP**. 2011. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. Cell **147**:1537–1550.

22. **Kelley D, Rinn J**. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol **13**:R107.

23. **Gong C, Maquat LE**. 2011. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. Nature **470**:284–288.

24. **Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest ARR, Carninci P, Biffo S, Stupka E, Gustincich S**. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature **491**:454–457.

25. **Repoila F, Darfeuille F**. 2009. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. Biol Cell **101**:117–131.

26. **Suzuki K, Babitzke P, Kushner SR, Romeo T**. 2006. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes Dev **20**:2605–17.

4.5 Tables and figures

Table 4.1 Primers used in this study

Name	Probe direction	Sequence (5' to 3')[a]
PCR amplification of DNA templates for biotinylated RNA probe generation		
R. albus 7		
Ra7_Rep_For	Sense	TAATACGACGACTCACTATAGGGTATGCCATTGGAACCTTTTCCTTTCGGTAGATTAACGC
Ra7_Rep_Rev		AATTGGTTTGACTGACGTCAATAATGTACGGCAGGAAATGCCGCATATCTTATAATAGGGAG
Ra7_Ext_For	Sense	TAATACGACGACTCACTATAGGGGCTCCGGAACTAAGCTGTATTTATCTTCGGGATCAC
Ra7_Ext_Rev		GCAGCTTAGCTTCCGAGCTAAAGGGAGCGTCC
Ra7_Prmt_For	Sense	TAATACGACTCACTATAGGGTCTTTCAAGATCTCCACGAGCCCTG
Ra7_Prmt_Rer		CGTTGTTATCGCCGGGTCTTGGTAATAAAGCTGATAGTTCCTTAAAATGAG
R. albus 8		
Ra8_Rep_For	Antisense	GGTTCGCCGCACAAATTATATAAGGAGTGGTGCATTATGGTAATGC
Ra8_Rep_Rev		TAATACGACTCACTATAGGGCACCGCCGAAAAGTCGGGATAAAAAACC
Ra8_Prmt_For	Antisense	GCACGTTGTTACCGTTGGGTCTTGATATTAAAGCCGATAGTTCC
Ra8_Prmt_Rev		TAATACGACTCACTATAGGGCAAGAAATGTACGAGCCCTGAACACTC
Primers for gap sequencing between contig_66 and contig_128 in the draft genome sequence of *R. albus* 8		
Contig_66_For		CAACATTCTCCCCCAGAAGTCAAGTAGC
Contig_128_Rev		GTAACTCGGCAACCCTCTTACGATCC

[a] Underlined sequence denotes additional sequence for T7 transcription

Table 4.2 RNA-seq mapping results of *R. albus* 7 and 8 during growth on AHPCS

RNAseq Sample ID	Unmapped reads (%)[a]	RNAseq Sample ID	Unmapped reads (%)
Ra7_1_4h	15.8 %	Ra8_1_4h	16.8 %
Ra7_2_4h	16.0 %	Ra8_2_4h	19.1 %
Ra7_1_8h	21.0 %	Ra8_1_8h	16.3 %
Ra7_2_8h	17.6 %	Ra8_2_8h	15.8 %
Ra7_1_12h	36.0 %	Ra8_1_12h	26.4 %
Ra7_2_12h	34.7 %	Ra8_2_12h	29.6 %
Ra7_1_20h	48.6 %	Ra8_1_20h	36.2 %
Ra7_2_20h	49.3 %	Ra8_2_20h	41.5 %
Ra7_1_36h	49.8 %	Ra8_1_36h	36.3 %
Ra7_2_36h	47.4 %	Ra8_2_36h	40.4 %

[a] Percentage proportion of unmapped reads in a total sequencing reads to annotated region of genome sequences of *R. albus* 7 and 8
[b] Reads were trimmed using CLC Genomics Workbench v5.5.1 with a quality score limit of 0.05 and maximum number of ambiguities of 2

Table 4.3 Mapping of non-mapped reads in RNAseq mapping against reference genome without annotation

RNAseq Sample ID	Total reads	Mapped reads (%)[a]	Not mapped reads (%)	RNAseq Sample ID	Total reads	Mapped reads (%)	Not mapped reads (%)
Ra7_1_4h	3,738,738	3,559,663 (95.2 %)	179,075 (4.8 %)	Ra8_1_4h	3,839,000	3,631,065 (94.6 %)	207,935 (5.4 %)
Ra7_2_4h	3,945,463	3,761,329 (95.3 %)	184,134 (4.7 %)	Ra8_2_4h	4,525,628	4,242,075 (93.7 %)	283,553 (6.3 %)
Ra7_1_8h	4,950,471	4,831,568 (97.6 %)	118,903 (2.4 %)	Ra8_1_8h	3,734,229	3,676,794 (98.5 %)	57,435 (1.5 %)
Ra7_2_8h	4,504,867	4,379,632 (97.2 %)	125,235 (2.8 %)	Ra8_2_8h	3,556,735	3,517,143 (98.9 %)	39,592 (1.1 %)
Ra7_1_12h	8,390,257	8,286,494 (98.8 %)	103,763 (1.2 %)	Ra8_1_12h	6,059,752	5,995,916 (99.0 %)	63,836 (1.0 %)
Ra7_2_12h	7,599,086	7,496,731 (98.6 %)	102,355 (1.4 %)	Ra8_2_12h	6,206,911	6,145,019 (99.0 %)	61,892 (1.0 %)
Ra7_1_20h	11,050,452	10,972,893 (99.3 %)	77,559 (0.7 %)	Ra8_1_20h	8,187,565	8,148,485 (99.5 %)	39,080 (0.5 %)
Ra7_2_20h	11,004,157	10,904,852 (99.1 %)	99,305 (0.9 %)	Ra8_2_20h	9,886,491	9,838,202 (99.5 %)	48,289 (0.5 %)
Ra7_1_36h	11,534,590	11,450,714 (99.3 %)	83,876 (0.7 %)	Ra8_1_36h	8,049,230	8,008,819 (99.5 %)	40,411 (0.5 %)
Ra7_2_36h	10,442,988	10,345,198 (99.1 %)	97,790 (0.9 %)	Ra8_2_36h	9,032,114	8,962,841 (99.2 %)	69,273 (0.8 %)

[a] Unmapped reads from RNAseq mapping to the reference genome with annotation were re-mapped to the *R. albus* 7 and 8 genome without annotation using CLC Genomics Workbench v5.5.1 with a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and maximum number of hits for a read of 10.

Table 4.4 Transcriptional profile of genes in lncRNA loci in *R. albus* 7 and 8 during growth on AHPCS

		Normalized RPKM					*p*-value			
Feature ID	Annotation	4h	8h	12h	20h	36h	8 vs 4 h	12 vs 4 h	20 vs 4 h	36 vs 4 h
R.albus 7										
Rumal_3117	Hypothetical protein (DUF1292)	1503.9	1598.3	971.6	925.0	862.2	7.3E-01	0.0E+00	0.0E+00	0.0E+00
Rumal_3118	Alcohol dehydrogenase	91.2	61.5	35.1	24.8	25.4	7.7E-04	2.3E-13	0.0E+00	0.0E+00
R. albus 8										
CUS_6628	Alcohol dehydrogenase	97.7	78.3	68.8	34.5	27.6	2.0E-02	8.3E-04	0.0E+00	0.0E+00
CUS_6629	Hypothetical protein (DUF1292)	652.8	760.8	690.7	444.9	396.5	5.0E-01	8.3E-01	3.0E-02	7.9E-03

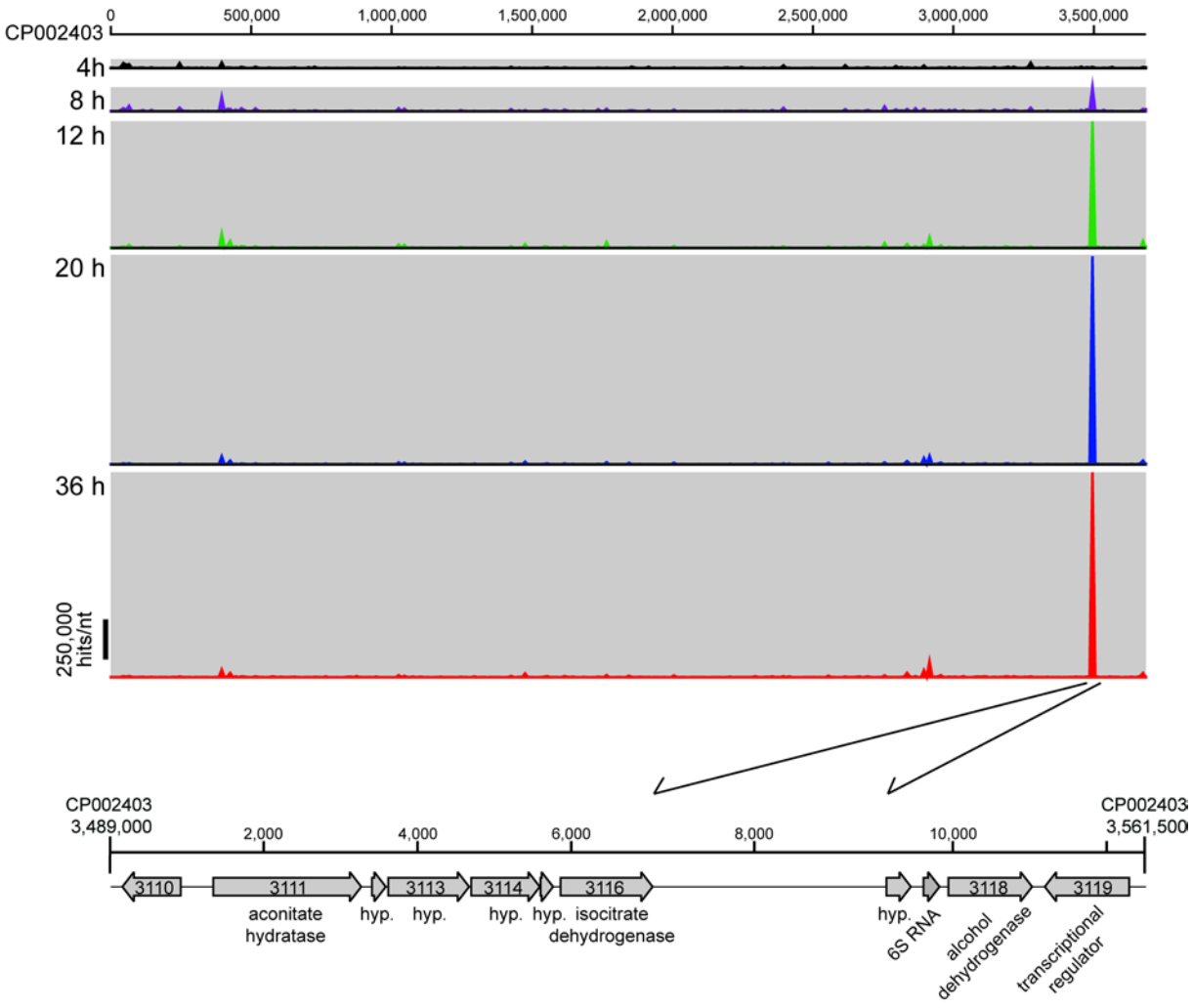


Fig. 4.1 Expression of ncRNA during growth of *R. albus* 7 on AHPCS.

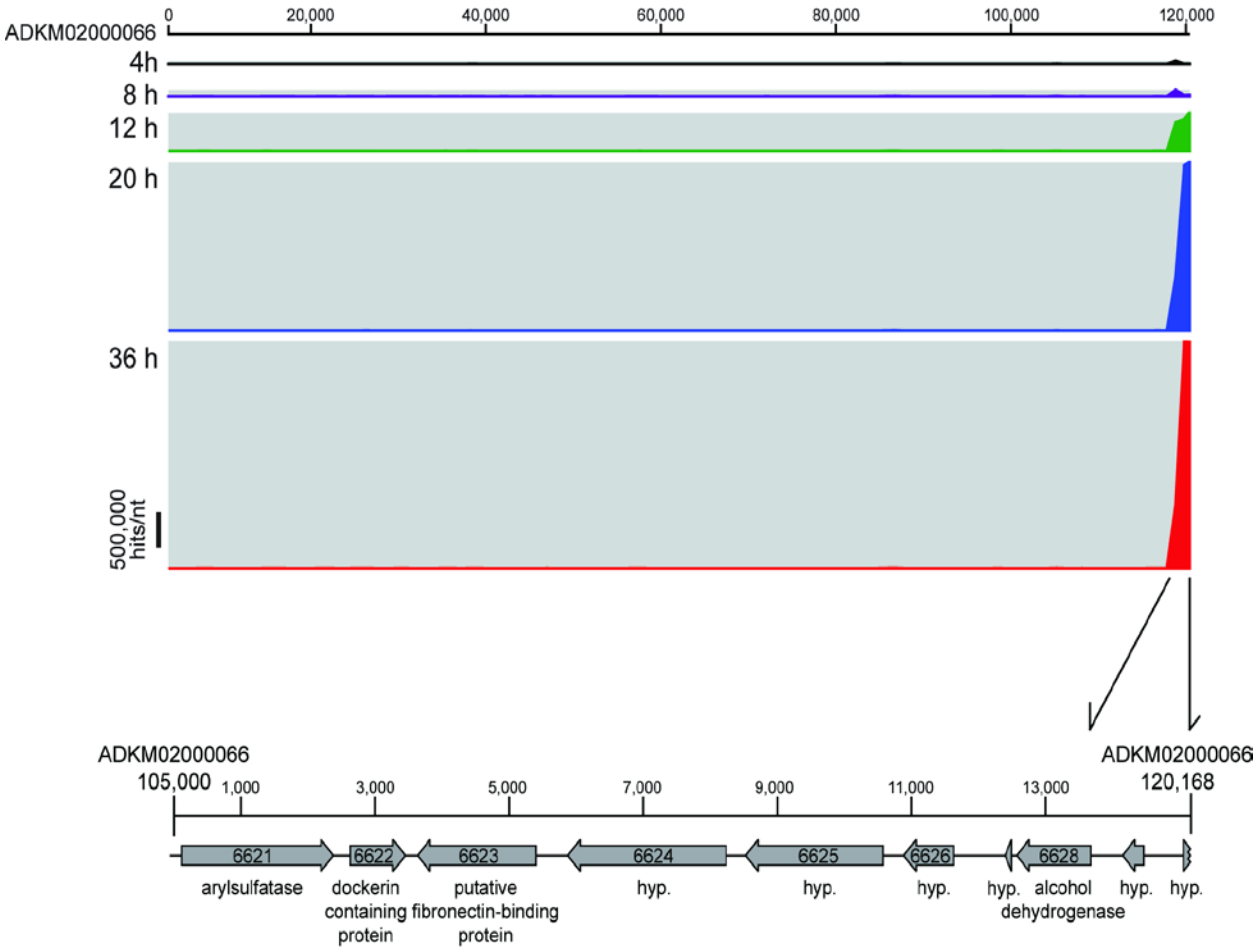


Fig. 4.2 Expression of ncRNA during growth of *R. albus* 8 on AHPCS.

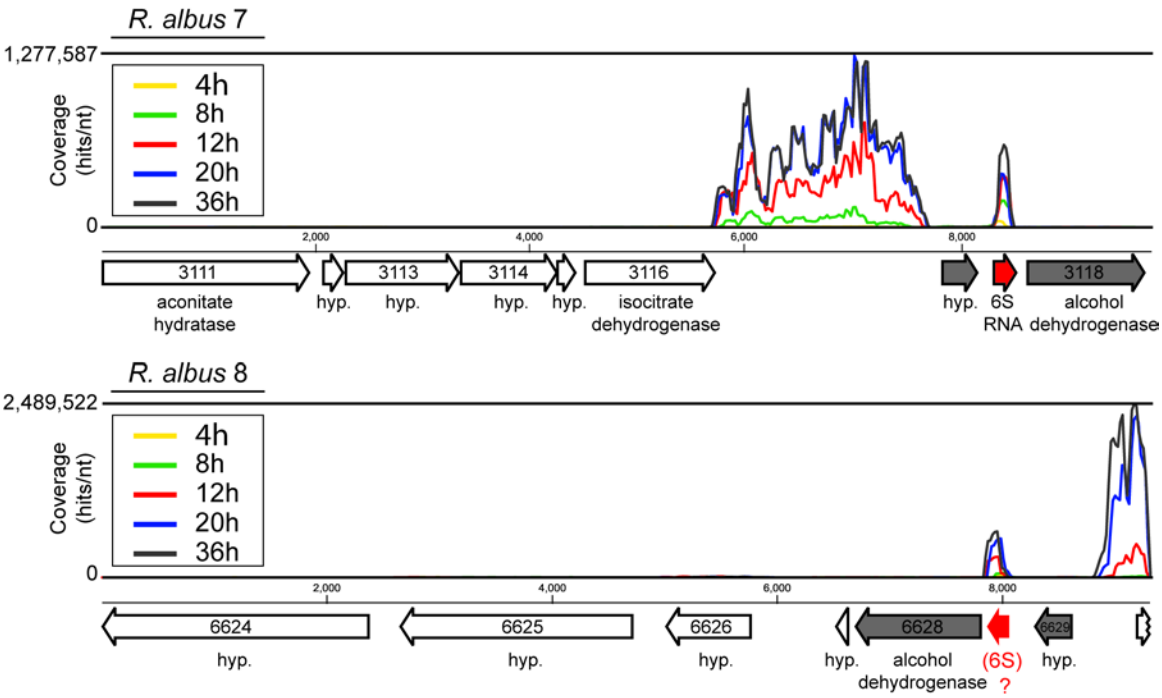


Fig. 4.3 RNA coverage of predominantly expressed ncRNA region in *R. albus* 7 and 8 during growth on AHPCS.

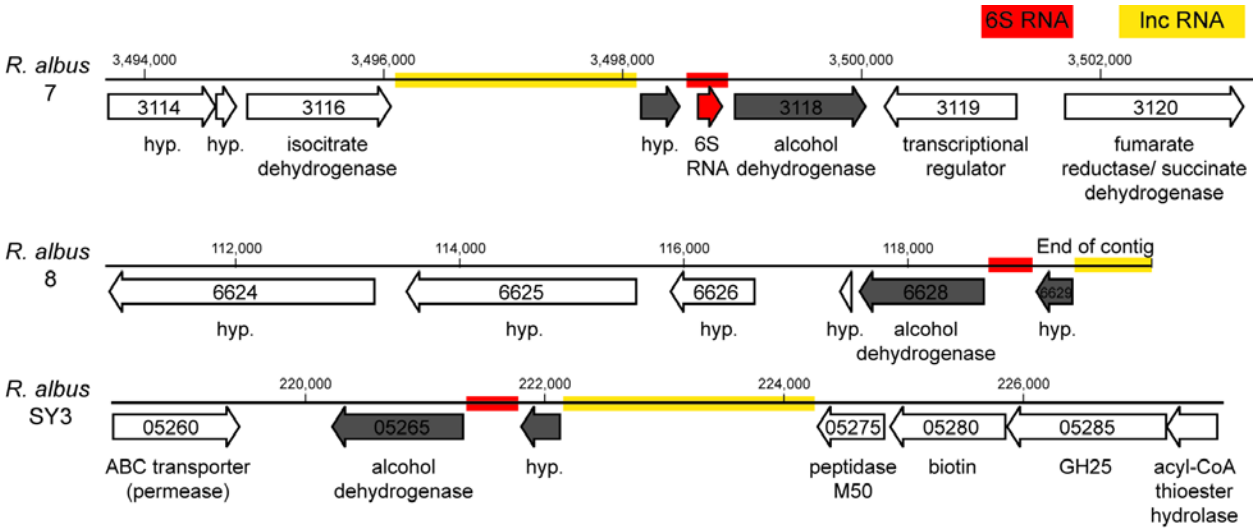
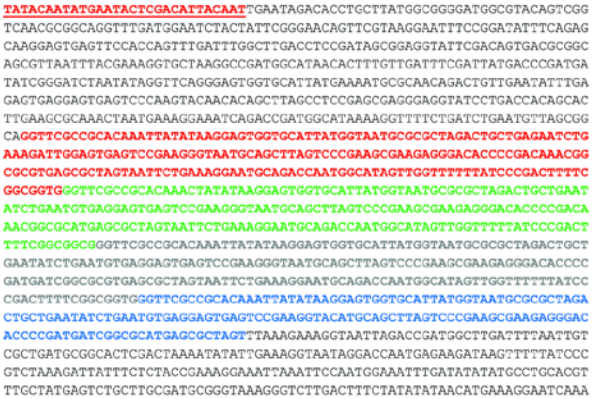


Fig. 4.4 Conserved long ncRNA region in *R. albus* strains 7, 8, and SY3.

R. albus 7

R. albus 8



R. albus SY3

Fig. 4.5 Tandem repeat sequences in the lncRNA region of *R. albus* strains 7, 8, and SY3. The top and bottom underlined sequences denote predicted transcription promoter and terminator, respectively. The red, green, gray or blue colored regions denote the tandem repeat sequences in the lncRNA region.

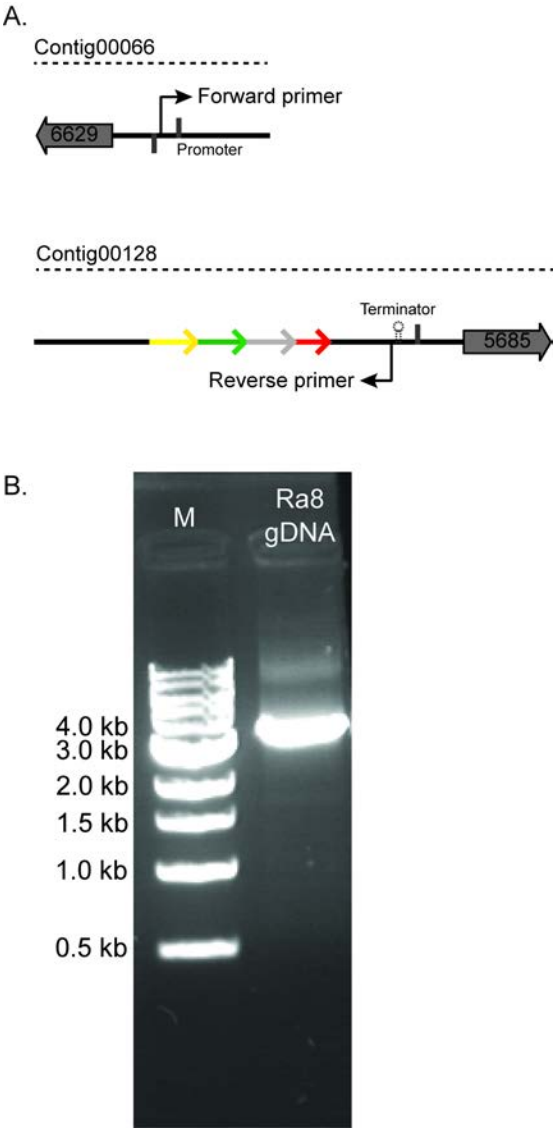


Fig. 4.6 Gap sequencing between contig00066 and contig00128 in the draft genome sequence of *R. albus* 8. *A*. Primer binding sites for contig00066 and contig00128. *B*. The PCR amplification size using the primer set above.

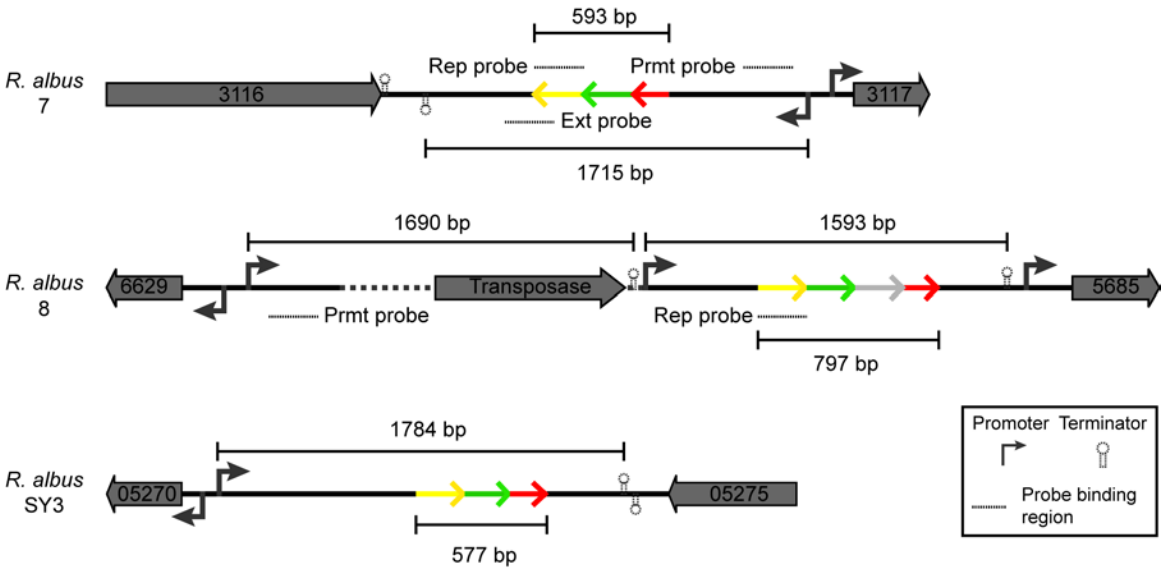


Fig. 4.7 Predicted transcription promoters, terminators, and tandem repeat sequences in lncRNA region of *R. albus* strains 7, 8, and SY3. Probes for northern blot analysis were designed to bind to either downstream of promoter (Prmt), tandem repeat sequence (Rep), or extended tandem repeat (Ext) in the transcripts of strain 7 and 8.

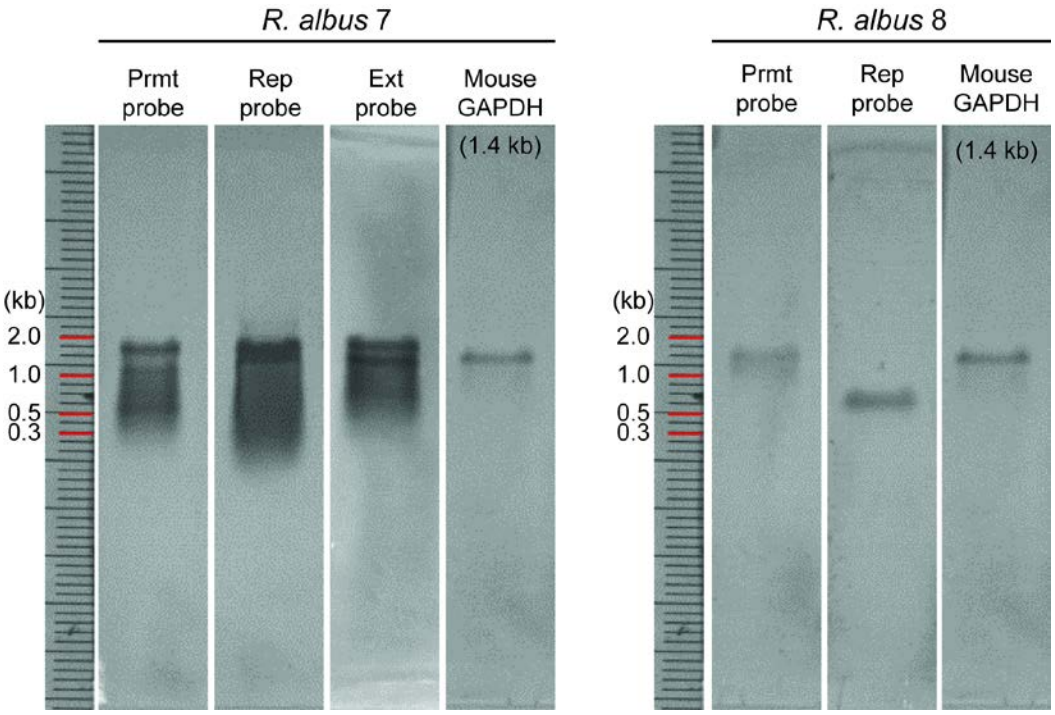


Fig. 4.8 Detection of the transcripts of lncRNAs in *R. albus* 7 and 8 using Northern blot hybridization. Sense probes were used to detect lncRNA of strain 7, while antisense probes were used to detect lncRNA of strain 8. The probe binding sites to the transcripts are illustrated in Fig. 6.

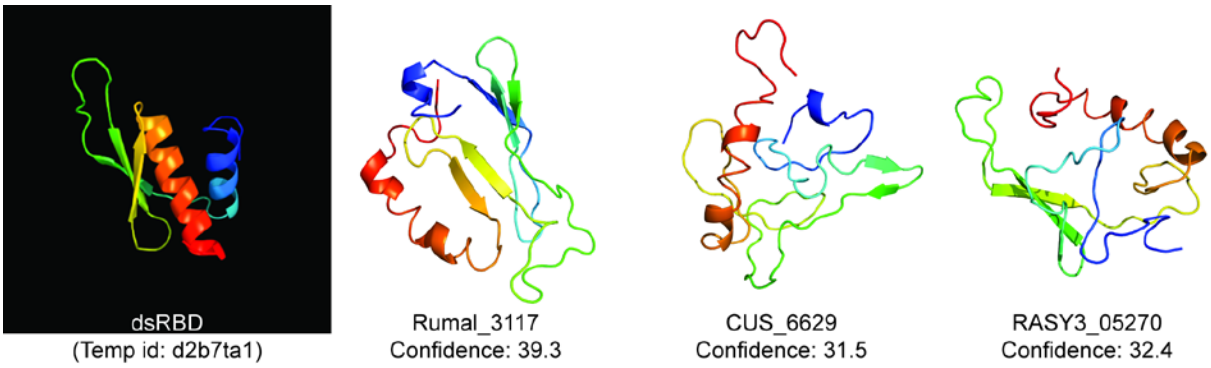


Fig. 4.9 Predicted model for DUF1292 genes in *R. albus* strains 7 (Rumal_#), 8 (CUS_#), and SY3 (RASY3_#). The Phyre2 server was used for protein modeling of DUF1292 genes in three strains (26). The 'Confidence' represents the probability that the match between the query sequence and the template sequence is a true homology. Best hit domain of DUF1292 genes were a double-stranded RNA-binding domain (dsRBD) shown in left panel.

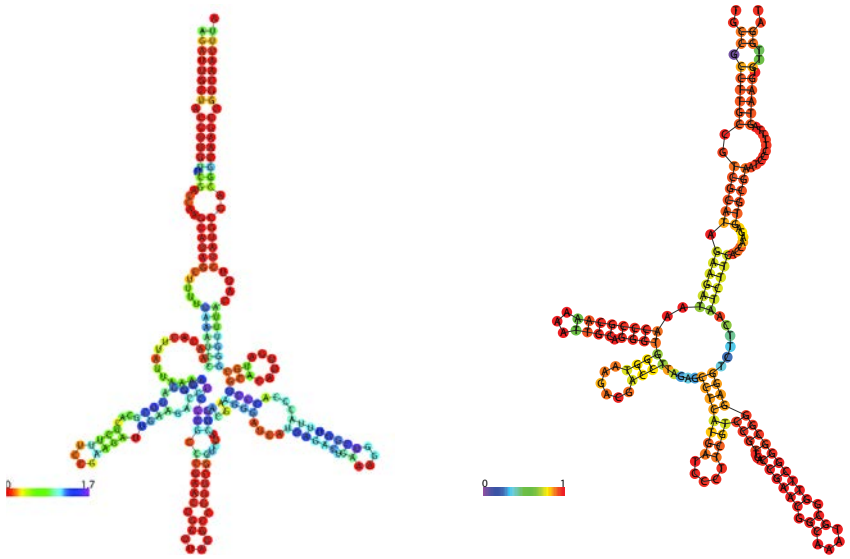


Fig. 4.10 Predicted secondary structure of 6S RNA in *R. albus* 7 and 8. The secondary structure for Rumal_R0067 gene in strain 7 (left panel) and its homologous gene in strain 8 (right panel) were predicted using the RNAfold server (27).

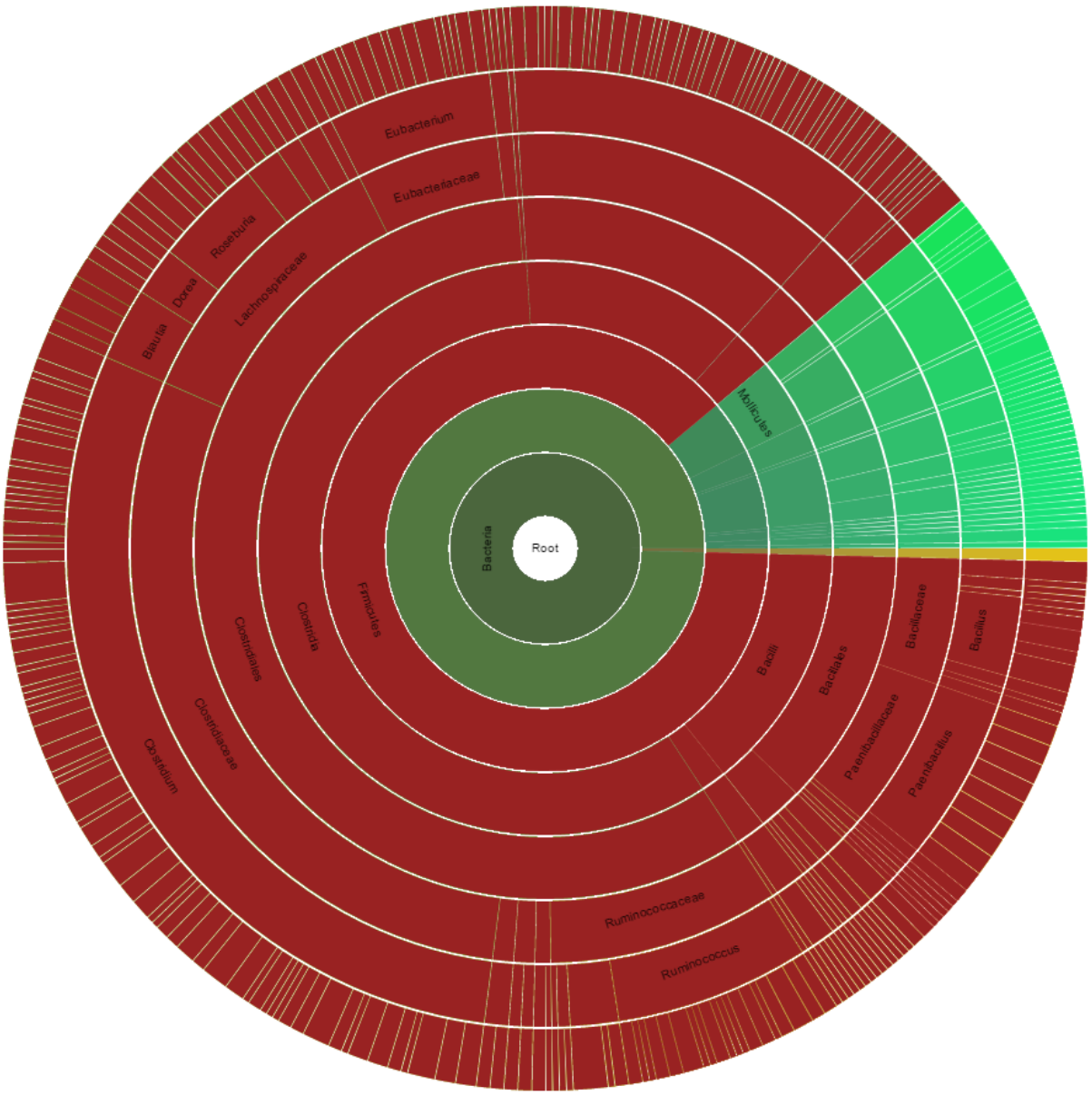


Fig. 4.11 Prevalence of DUF1292 domain in the genomes of Firmicutes. Among 488 bacterial sequences encoding DUF1292 in Pfam database, 432 sequences are present on the genome sequences of Firmicutes.

CHAPTER 5.

DISCUSSION AND CONCLUSION

5.1 Introduction

Herbivorous animals such as ruminants rely on the symbiotic microbes in the rumen to break down cellulose and hemicellulose contained in their dietary fiber. The resulting production of short-chain fatty acids from carbohydrate microbial fermentation and microbial protein synthesized in the rumen play a pivotal role in host nutrition (1–3). Another role of rumen microbes in host nutrition is the supply of water-soluble vitamins (especially B group vitamins) to the host. Despite a lack of biosynthetic pathways for B vitamins in the host genome, they are independent of a dietary supply of B vitamins due to biosynthesis by rumen bacteria (4, 5). Thus, for better understanding of host-microbe interactions, it is important and relevant to establish the model for fiber degradation and vitamin metabolism of fibrolytic gut microbes and investigate their roles in the gut ecosystem.

Of the dominant fibrolytic Firmicutes in the rumen, the fibrolytic mechanism of *Ruminococcus albus* is less well known than the other dominant *Ruminococcus* species, *R. flavefaciens* that is known to possess the cellulosomal system (6, 7). Furthermore, even species harboring the well-known cellulosomal or PULs system display differential fibrolytic activities at the strain level (8–10), which raises the question: **can these fibrolytic models characterized in some specific strains be applied to the other strains within the cognate species and even to other species based on similarity of genome sequence?** To explore this fundamental question, this research was conducted using integrated comparative genomic, phenotypic, and

transcriptomic analyses of two strains of *R. albus*, *R. albus* 7 and 8, focusing on the conservative and strain specific mechanisms for fiber and folate utilization between strains.

5.2 Conservative fibrolytic system in two distinct strains, *R. albus* 7 and 8.

Bacterial fibrolytic systems generally consist of two components; cell attachment to insoluble substrate and substrate degradation by carbohydrate active enzymes (CAZymes). The previous genomic analysis reported that three strains of *R. albus* (7, 8, and SY3) possess a similar repertoire of CAZyme genes, in terms of the number (122 domains in strain 7, 114 domains in strain 8, and 124 domains in strain SY3) and variety (46 families in strain 7, 43 families in strain 8, and 42 families in strain SY3) of CAZy families (11). In addition to the similar genetic profile of CAZymes, this research also identified that both strains possess same number of genes encoding Pil-like protein (CbpC; n=1), Cel9B (n=1) and Cel48A (n=1), and that both strains possess similar number of CBM37 bearing CAZyme genes as well as hypothetical genes. Despite a variety of CAZyme families present on both genomes, the five CAZyme genes encoding the versatile GH5, GH9 (Cel9B), GH10, GH11, and GH48 (Cel48A) domains were transcriptionally predominant (>1000 RPKM) in both strain 7 and 8 during growth on alkaline hydrogen peroxide treated corn stalk (AHPCS), phosphoric acid swollen cellulose (PASC), and wheat arabinoxylan (WAX). These results suggest that these five GH families are likely to be the primary GH enzymes employed by two strains of *R. albus* for the hydrolysis of plant cell wall polysaccharides. Consistently, Dai et al. (2015) reported that the top 3 abundant transcripts of cellulases and hemicellulases in cow rumen included GH9, GH5, GH48 and GH10, GH11, GH26, respectively, and these transcripts were predicted to be primarily synthesized by *Ruminococcus* and *Fibrobacter* genera (12).

It is important to note that in both strains of *R. albus*, most of highly expressed extracellular GH genes on PASC, WAX, and AHPCS encode CBM37 domains rather than dockerins. In addition, unknown function of CBM37 (UF-CBM37) genes that do not encode any known or predictable catalytic domains were also highly expressed in both strains. These catalytic or non- catalytic CBM37 genes were predicted to be extracellular proteins, inferred from the presence of a signal peptide, suggesting that together with CBM37 bearing GHs, the UF-CBM37 protein may work together after secretion to the external medium. The CBM37 domain has been shown to bind a broad range of polysaccharides, including cellulose and xylan, and even to the cell surface (13). Some of UF-CBM37 genes in *R. albus* 7 contain leucine rich repeat (LRR) domains and it has been proposed that the extracellular CAZymes may be complexed by UF-CBM37 protein via protein-protein interaction mediated by LRR domains (14). Their concept is also supported by our findings that LRR containing UF-CBM37 genes were highly expressed in both strains during exponential growth on AHPCS. Interestingly, we also found that some of the highly expressed UF-CBM37 genes on AHPCS did not contain LRR domains, but contains only tandem CBM37 domains (Fig. S15). In tandem CBM37 modules, proximal and distal modules are phylogenetically classified into different groups (13, 15). Xu et al. (2004) reported that distal CBM37 in tandem CBM37s exhibited a binding affinity to a variety of polysaccharides, whereas the proximal CBM37 bound to the putative ribosomal protein in the cell extract of *R. albus* 8 instead of polysaccharides (15). These results suggest that the proximal CBM37 may function as a binding module with other proteins. Thus, the UF-tandem CBM37 proteins, together with LRR containing UF-CBM37 proteins, may mediate protein-protein interaction between other extracellular CAZymes, leading to the localization of the CAZymes on the surface of *R. albus*, unlike the cellulosomal system.

5.3 Proposed additional component in the cellulolytic system of the yellow pigmented strain of *R. albus*

The research described in this thesis demonstrates that both strains 7 and 8 apparently have a different capability of degrading crystalline cellulose, despite the findings that both strain 7 and 8 possess all the components of the fibrolytic system described above. The differential phenotypic variation exhibited between the two strains grown on crystalline cellulose implies that additional components must be implicated in the cellulolytic system of *R. albus* 7. One of the potential candidate components is the yellow pigment produced by *R. albus* 7, but not *R. albus* 8. The yellow pigment is a distinct phenotypic feature of some cellulolytic microorganisms in the rumen and soil (16–18). Although the gene encoding the yellow pigment has been yet identified, the role of yellow pigments in cellulose degradation has been reported in the rumen bacterium *Ruminococcus flavefaciens*, soil bacterium *Clostridium thermocellum*, and rumen fungi *Orpinomyces joyonii* (18–22). The yellow pigments produced by those three species showed a strong affinity for both microcrystalline cellulose and endoglucanases. Considering that the highly expressed endoglucanases in *R. albus* possess CBM37 domains, the yellow pigments likely interact with cellulose and CBM37 bearing endoglucanases on the cell surface, resulting in tight adhesion of the yellow pigmenting strain of *R. albus* species to cellulose. This mechanism could apply to other cellulolytic Firmicutes that produce the yellow pigment, so it is still relevant to explore the role of yellow pigment in the cellulolytic system of bacteria.

5.4 Distinct strategy for fiber utilization between strains of *R. albus*

In addition to the similar genomic potential for fiber degradation, nine most highly expressed GH genes in strain 7 grown on PASC and WAX were cloned, expressed, and

characterized for the enzymatic activities on different polysaccharides. Compared to the homologous GHs that were biochemically characterized in the previous studies (23, 24), the catalytic activities of those nine GHs in strain 7 were similar to their homologs in strain 8, in terms of the substrate range and the amount of reducing sugars, suggesting that both strains 7 and 8 possess a similar genetical and catalytical potential for the fiber degradation. However, the genes encoding CAZymes or substrate adhering components transcriptionally differ between strains in response to cellulose and hemicellulose. In strain 7, genes encoding predicted CAZymes were more transcriptionally responsive to PASC than WAX, while strain 8 had more genes transcriptionally responsive to WAX than PASC. Together with their differential ability for degrading filter paper and arabinan between strains, this research proposes that strain 7 likely prefers cellulose, while strain 8 prefers hemicellulose contained in the plant cell wall. To support our hypothesis, both strains were grown on AHPCS (delignified corn stalk) composed of cellulose and hemicellulose. During growth on AHPCS, strain 7 had more substrate adherent cells and utilized less hemicellulosic sugars than strain 8 that exhibited preferred planktonic growth and hemicellulosic sugar. In addition to this phenotypic variation, the time dependent transcriptomic analyses suggest that *R. albus* 7 subsequently responds to and utilizes xylooligosaccharides, cellobiose, and cellooligosaccharides as the degradation process from outside hemicellulose to inside cellulose core in plant cell wall takes place, while *R. albus* 8 preferentially responds to and utilizes hemicellulose over cellulose. Taken together, it is reasonable to conclude that despite similar genomic contents, each strain of *R. albus* uses a different strategy for the utilization of cellulose and hemicellulose in the plant cell wall through differential regulatory mechanism. With this model for the differential strategies between strains, it is possible for both strains to co-exist and co-grow on complex substrate containing

hemicellulose and cellulose. To support, this research demonstrates that *R. albus* 7 and 8 grew together on AHPCS, especially at late-log to early log phases when hemicellulose in AHPCS is abundant in the culture.

5.5 Potential regulatory system for fiber degradation in *R. albus* strains

In addition to distinct strategies for plant cell wall degradation between strain 7 and 8, the research presented in this thesis identifies the potential regulatory systems underlying the strategies used by strain 7 and 8. Based on differential phenotypes between two strains for substrate attachment during growth on AHPCS, we examined global regulators known to be involved in biofilm formation of bacteria: the bacterial second messenger, cyclic-di-GMP (c-di-GMP), regulatory system and the accessory gene regulator (Agr) quorum sensing system. The high level of c-di-GMP and the low level of autoinducing peptide in the accessory gene regulator (Agr) quorum sensing system are known to facilitate biofilm formation of some pathogenic bacteria (reviewed in reference 25). The putative c-di-GMP regulatory genes (diguanylate cyclase and phosphodiesterase) and Agr genes (*agrABCD*) were identified in both strain 7 and 8. Intriguingly, those c-di-GMP regulatory genes identified in this research were conserved with CAZyme genes in both strains of *R. albus* as well as other cellulolytic Ruminococci, including *R. flavefaciens* and *R. champanellensis*. Transcriptomic analyses showed that there was a strong correlation between those two regulatory system and fibrolytic system in *R. albus* 7 and 8. Strain 7 had more c-di-GMP regulatory genes transcriptionally responsive to cellulose, while strain 8 had more c-di-GMP regulatory genes transcriptionally responsive to hemicellulose, which was similar to the transcriptional pattern with CAZyme genes in response to cellulose and

hemicellulose. In addition, the Agr QS genes, especially *agrBD*, were highly expressed in *R. albus* 8 that preferred the planktonic niche compared to the adherent phenotype preferred *R. albus* 7. Taken together, these findings suggests that both the c-di-GMP regulatory and Agr quorum sensing system could be potential regulators for the fibrolytic system in *R. albus* 7 and 8 and furthermore, that c-di-GMP and Agr QS systems are implicated not only in biofilm formation of pathogenic bacteria, but also in adherence to insoluble fibrous substrate and the fibrolytic system of commensal fiber degrading bacteria.

5.6 Unique phosphoketolase pathway potentially conferring competitive fitness to hemicellulolytic strains in the gut

Through the glycolysis and pentose phosphate (PP) pathways bacteria can conserve more ATP from hexose than pentose per molecule. Furthermore, considering more energy conservation through phosphorolytic cleavage of cellooligosaccharides relative to pentose polymers, the hemicellulose preference seems unlikely to confer competitive fitness to *R. albus* 8 against other cellulolytic bacteria, including *R. flavefaciens*, *F. succinogenes*, and even *R. albus* 7. Although relatively less abundant than *R. albus* 7 and SY3, *R. albus* 8 is still detected in the rumen by metagenomics analysis (26), meaning that *R. albus* 8 co-exists with other cellulolytic species and plays a role as a member of the microbial community in the rumen. Notably, this research identified that the hemicellulose preferring strain, *R. albus* 8, possessed a unique phosphoketolase (PK) pathway, while the cellulose preferring strain, *R. albus* 7 did not contain the PK pathway. Strain 8 possesses a putative xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (XFP) gene, which is orthologous to the biochemically characterized XFP gene

in *Clostridium acetobutylicum* ATCC 824 (27, 28). The XFP enzyme catalyzes the conversion of xylulose-5-phosphate into glyceraldehyde-3-phosphate and acetyl phosphate, which is a key component of the phosphoketolase (PK) pathway in *C. acetobutylicum*. Based on the profiles of fermentation products and the XFP gene expression on during growth on PASC, WAX, and AHPCS, pentose is likely to be metabolized through the PP pathway in *R. albus* 7, while in *R. albus* 8, both PP and PK pathways are used for pentose fermentation with the transcriptionally dominance of the PK pathway over the PP pathway in late-log phase on AHPCS. This finding raised the question of how can the PK pathway confer competitive fitness to *R. albus* 8? Flamholz et al. (2013) proposed that although the Entner-Doudoroff (ED) pathway generates one less ATP per glucose than the glycolysis pathway, this ATP loss could be compensated by saving the cost for enzyme synthesis required for the lower glycolysis pathway (29). In the ED pathway, half of the carbons of glucose bypass the lower glycolytic sequence (from glyceraldehyde-3-phosphate to pyruvate), which is similar to the PK pathway in which two carbons of pentose bypass the lower glycolysis and fermentation pathways (from glyceraldehyde-3-phosphate to acetyl-phosphate). Using this rationale, the PK pathway may enable *R. albus* 8 to catabolize pentose rapidly as well as conserve energy and costs for enzyme synthesis required for the lower glycolytic sequence, which could confer competitive fitness to *R. albus* 8 in comparison to other cellulolytic bacteria, including *R. albus* 7. In supporting this concept it was shown that strain 8 had a higher growth rate than strain 7 when grown on beechwood xylan consisting of mostly xylose residues (>90%). We also identified that some strains belonging to other characterized hemicellulolytic bacterial groups in the gut, such as *Butyrivibrio fibrisolvens* and *Roseburia intestinalis*, also possess the putative PK pathway, suggesting that the PK pathway is likely to be

a unique sugar metabolic pathway conferring competitive fitness to hemicellulolytic bacteria, including *R. albus* 8.

5.7 Folate and *p*ABA metabolism in *R. albus* strains

The vitamin metabolism of *R. albus* strains remains largely unknown, while the vitamin requirements of *R. albus* 8 have not been studied. Our research on folate metabolism in *R. albus* 7 and 8 provided genomic evidence for three folate utilization pathways (either *de novo* synthesis, salvage, or both pathways) conserved in the Firmicutes including *R. albus* strains. Through the growth experiments in the presence or absence of folate and para-aminobenzoate (*p*ABA), this research demonstrated that at the strain level, *R. albus* strains 7 and 8 rely on different folate metabolic pathways where we show that *R. albus* 7 is a folate autotroph while *R. albus* 8 is a folate auxotroph. These findings indicate that each strain performs a different ecological role in ruminal folate metabolism. In addition, the results of transcriptomic analysis suggest that the folate autotrophic strain, *R. albus* 7, also has an alternative pathway for *p*ABA synthesis and likewise other *Ruminococcus* species lacking the canonical *p*ABA synthetic pathway are likely autotrophs and not auxotrophs. It is important to note that the variation in folate metabolic pathways between strains is not only present in *R. albus* species, but also in other Gram-positive Firmicutes, which leads us to address a fundamental question about this functional diversity for folate metabolism observed in microbial ecology. According to the "Black Queen" hypothesis proposed by Morris and colleagues (2012, 2015), where members of a community lose the ability to perform functions whose products are available from the environment, auxotrophs presumably arise from autotrophic ancestors as a result of the loss of essential biosynthetic function to reduce the metabolic burden when the corresponding metabolite is available in their

habitat or produced by neighboring commensal organisms (30, 31). Applying this concept to the variation of folate metabolic pathways in *R. albus* species, strain 7, 8, and SY3 possessing either of the three folate utilization pathways would have evolved from an autotrophic ancestor harboring a complete folate biosynthetic pathway. Considering the potential saving of biosynthetic costs for synthesis of GTP, glutamate and production of the long (six enzymatic steps) biosynthetic pathway, it seems reasonable that in the rumen where exogenous folate is present in sufficient amounts, auxotrophic strains subsequently arise to utilize exogenous folate by acquisition of the transport or salvage pathway. As a result of this selection, the lineage capable of actively transporting folate lost the redundant genetic material for folate biosynthesis. This loss of genetic material or adaptive genome streamlining is thought to occur more frequently in nutrient rich or constant environments such as the intestinal tract (31–33).

5.8 Identification of long non-coding RNA loci in *R. albus* strains

Through transcriptomic analyses assessed by RNA sequencing, we identified a potential long non-coding RNA (lncRNA) loci in the genomes of strain 7 and 8. The putative lncRNA loci consisted of four sequence components (lncRNA, DUF1292 gene, putative 6S RNA, and alcohol dehydrogenase), which is conserved in *R. albus* strains 7, 8, and SY3. Based on their transcriptional profiles assessed by RNA-seq and northern blot analyses, it seems likely that the lncRNA loci are involved in the regulatory system related to the stationary phase of cells. Furthermore, the orthologous genes to the DUF1292 gene were predominantly present in the Firmicute genomes. The protein structure of DUF1292 was predicted to be similar to double strand RNA-binding protein, which imply that the DUF1292 may be a unique ncRNAs regulator exclusively employed by Firmicutes. To the best of our knowledge, this is the first report that

demonstrates a transcriptionally functional lncRNA in bacteria with a growth dependent expression pattern.

5.9 Conclusion

In closing, our findings provide molecular insight into conserved and differentiated fibrolytic system and folate metabolism demonstrated by *R. albus* 7 and 8. These findings strongly indicate that the cellular strategy for plant cell wall and folate utilization varies at strain level of *R. albus*, which means that each strain can be assigned to different niches and different ecological function in the rumen. Therefore, each strain within this bacterial species can be non-competitive, but likely cooperative by playing different ecological roles in the gut. This is a strong case for bacterial specialization and niche differentiation, and suggests that interpretation and functional model construction of rumen and other gut system at the population and metagenomics level is oversimplified and incorrect. In addition, this research provides molecular evidence for the presence of lncRNA in Gram-positive Firmicutes. For better understanding of the regulatory mechanism in bacteria, more research is required to determine the functional characterization of lncRNAs.

5.10 References

1. **Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI**. 2001. Molecular analysis of commensal host-microbial relationships in the intestine. Science **291**:881–884.

2. **Nicholson J, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Petttersen S**. 2012. Host-gut microbiota metabolic interactions **1262**.

3. **Thomas LV, Ockhuizen T, Suzuki K**. 2014. Exploring the influence of the gut microbiota and probiotics on health: a symposium report. Br J Nutr **112 Suppl** :S1–S18.

4. **Wolin MJ, Miller TL, Stewart CS**. 1997. The Rumen Microbial Ecosystem, p. 467–491. *In* Hobson, PN, Stewart, CS (eds.), Second edi. Blackie Academic & Professional, London.

5. **McDowell LR**. 2000. Folacin, p. 480–521. *In* Vitamins in animal and human nutrition. Second edi. Iowa State University Press, Ames.

6. **Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA**. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol **6**:121–131.

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

8. **Krause DO, Bunch RJ, Smith WJM, McSweeney CS**. 1999. Diversity of *Ruminococcus* strains: A survey of genetic polymorphisms and plant digestibility. J Appl Microbiol **86**:487–495.

9. **Robert C, Chassard C, Lawson PA, Bernalier-Donadille A**. 2007. *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. Int J Syst Evol Microbiol **57**:1516–1520.

10. **McNulty NP, Wu M, Erickson AR, Pan C, Erickson BK, Martens EC, Pudlo NA, Muegge BD, Henrissat B, Hettich RL, Gordon JI**. 2013. Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome. PLoS Biol **11**:e1001637.

11. **Dassa B, Borovok I, Ruimy-Israeli V, Lamed R, Flint HJ, Duncan SH, Henrissat B, Coutinho P, Morrison M, Mosoni P, Yeoman CJ, White BA, Bayer EA**. 2014. Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. PLoS One **9**:e99221.

12. **Dai X, Tian Y, Li J, Luo Y, Liu D, Zheng H, Wang J, Dong Z, Hu S, Huang L**. 2015. Metatranscriptomic analyses of plant cell wall polysaccharide degradation by microorganisms in the cow rumen. Appl Environ Microbiol **81**:1375–86.

13. **Ezer A, Matalon E, Jindou S, Borovok I, Atamna N, Yu Z, Morrison M, Bayer EA, Lamed R**. 2008. Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. J Bacteriol **190**:8220–8222.

14. **Christopherson MR, Dawson JA, Stevenson DM, Cunningham AC, Bramhacharya S, Weimer PJ, Kendziorski C, Suen G**. 2014. Unique aspects of fiber degradation by the ruminal ethanologen *Ruminococcus albus* 7 revealed by physiological and transcriptomic analysis. BMC Genomics **15**:1066.

15. **Xu Q, Morrison M, Nelson KE, Bayer EA, Atamna N, Lamed R**. 2004. A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. FEBS Lett **566**:11–6.

16. **Hungate RE**. 1950. The anaerobic mesophilic cellulolytic bacteria. Bacteriol Rev **14**:1–49.

17. **Hungate RE**. 1957. Microorganisms in the rumen of cattle fed a constant ration. Can J Microbiol **3**:289–311.

18. **Ljungdahl LG, Pettersson B, Eriksson KE, Wiegel J**. 1983. A yellow affinity substance involved in the cellulolytic system of *Clostridium thermocellum*. Curr Microbiol **9**:195–199.

19. **Lamed R, Kenig R, Setter E, Bayer EA.** 1985. Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. Enzyme Microb Technol **7**:37–41.

20. **Ljungdahl LG, Coughlan MP, Mayer F, Mori Y, Hon-nami H, Hon-nami K**. 1988. Macrocellulase complexes and yellow affinity substance from *Clostidium thermocellum*., p. 483–500. *In* Wood, WA, Kellogg, ST (eds.), Methods in Enzymology. Part B, Biomass: Cellulose and Hemicellulose. Academic Press, New York.

21. **Kopecný J, Hodrová B**. 1997. The effect of yellow affinity substance on cellulases of *Ruminococcus flavefaciens*. Lett Appl Microbiol **25**:191–196.

22. **Hodrová B, Kopečný J, Káš J**. 1998. Cellulolytic enzymes of rumen anaerobic fungi *Orpinomyces joyonii* and *Caecomyces communis*. Res Microbiol **149**:417–427.

23. **Moon YH, Iakiviak M, Bauer S, Mackie RI, Cann IKO**. 2011. Biochemical analyses of multiple endoxylanases from the rumen bacterium *Ruminococcus albus* 8 and their synergistic activities with accessory hemicellulose-degrading enzymes. Appl Environ Microbiol **77**:5157–69.

24. **Iakiviak M, Mackie RI, Cann IKO**. 2011. Functional analyses of multiple lichenin-degrading enzymes from the rumen bacterium *Ruminococcus albus* 8. Appl Environ Microbiol **77**:7541–50.

25. **Srivastava D, Waters CM**. 2012. A tangled web: Regulatory connections between quorum sensing and cyclic Di-GMP. J Bacteriol **194**:4485–4493.

26. **Rozman Grinberg I, Yin G, Borovok I, Berg Miller ME, Yeoman CJ, Dassa B, Yu Z, Mizrahi I, Flint HJ, Bayer EA, White BA, Lamed R**. 2015. Functional phylotyping approach for assessing intraspecific diversity of *Ruminococcus albus* within the rumen microbiome. FEMS Microbiol Lett **362**:1–10.

27. **Liu L, Zhang L, Tang W, Gu Y, Hua Q, Yang S, Jiang W, Yang C**. 2012. Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by 13C metabolic flux analysis. J Bacteriol **194**:5413–5422.

28. **Servinsky MD, Germane KL, Liu S, Kiel JT, Clark AM, Shankar J, Sund CJ**. 2012. Arabinose is metabolized via a phosphoketolase pathway in *Clostridium acetobutylicum* ATCC 824. J Ind Microbiol Biotechnol **39**:1859–1867.

29. **Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R**. 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci U S A **110**:10039–10044.

30. **Morris JJ, Lenski RE, Zinser ER**. 2012. The Black Queen hypothesis: Evolution of dependencies through adaptive gene loss. MBio **3**:e00036–12–e00036–12.

31. **Morris JJ**. 2015. Black Queen evolution: The role of leakiness in structuring microbial communities. Trends Genet **31**:475–482.

32. **Wolf YI, Koonin E V.** 2013. Genome reduction as the dominant mode of evolution. BioEssays **35**:829–837.

33. **D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, Kost C**. 2014. Less is more: Selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution (NY) **68**:2559–2570.

APPENDIX A: Supplemental tables for Chapter 2

Table A.1 Anaerobic medium for culturing *R. albus* strain 7 and 8

Ingredient	Concentration in stock (per liter)	Concentration in Media (per liter)
Carbohydrate source		4 g or 5 g[a]
Ammonium sulfate		4 g
Hemin		0.25 mg
Resazurin	0.1% (w/v)	1 ml
Sodium carbonate		4 g
Mineral Solution		50 ml
K_2HPO_4	6 g	
KH_2PO_4	6 g	
NaCl	12 g	
$CaCl_2 \cdot 2H_2O$	1.2 g	
$MgSO_4 \cdot 7H_2O$	1.2 g	
$(NH_4)_2SO_4$	12 g	
Pfennig's Trace Elements Solution		1 ml
EDTA	0.5 g	
$ZnSO_4 \cdot 7H_2O$	0.1 g	
$MnCl_2 \cdot 4H_2O$	0.03 g	
H_3BO_3	0.03 g	
$CoCl_2 \cdot 6H_2O$	0.2 g	
$CuCl_2 \cdot 2H_2O$	0.01 g	
$FeCl_2 \cdot 4H_2O$	1.5 g	
$NiCl_2 \cdot 6H_2O$	0.02 g	
$Na_2MoO_4 \cdot 2H_2O$	0.03 g	
Na_2SeO_3	0.01 g	
Schaefer Vitamin B Solution		10 ml
Thiamine-HCl	20 mg	
Ca-D-pantothenate	20 mg	
Nicotinamide	20 mg	
Riboflavin	20 mg	
Pyridoxine-HCl	20 mg	
Para-aminobenzoic acid	1 mg	
Biotin	0.5 mg	
Folic acid	0.125 mg	
Vitamin B_{12}	0.2 mg	
Tetrahydrofolic acid	0.125 mg	
VFA Solution (ml/L)		10 ml
Acetic acid	13.7 ml	
Propionic acid	6 ml	
Butyric acid	3.68 ml	
Isobutyric acid	1.1 ml	
2-methylbutyric acid	0.94 ml	
n-valeric acid	1.1 ml	
Isovaleric acids	1.1 ml	
Phenylacetate	340 mg	
Phenylpropionate	375 mg	
Cysteine Sulfide Solution (ml/L)		20 ml
NaOH	1.25 g	
Cysteine-HCl	5 g	
$Na_2S.9H_2O$	5 g	

[a] Either cellobiose (4g) or PASC (4g) or WAX (4g) or AHPCS (5g) was added into medium as defined carbohydrate sources.

Table A.2 Primer list for qPCR in co-culture experiment

Target gene	Type	Sequence	Size (bp)	Tm (°C)	GC (%)	Amplicon size
R. albus 7 Rumal_2649 (GH39)	Forward	5' CGATGCTGAGTGGATACAGAAG 3'	22	62.2	50	100 bp
	Reverse	5' CTCGGGAAATTCAACGGTATAGT 3'	23	62.2	43.5	
R. albus 8 CUS_6296 (GH105)	Forward	5' GGCACTACAGCGGAAAGTTA 3'	20	62.1	50	103 bp
	Reverse	5' CTGCCGCTTTGTTTGCATAG 3'	20	62.2	50	

Table A.3 RNAseq table of *R. albus* strain 7 and 8 grown on defined substrates, cellobiose (G2), phosphoric acid swollen cellulose (PASC), and wheat arabinoxylan (WAX)

RNAseq Sample ID (*R. albus* 7)	Total reads (avg. length)	Reads after Trimming[a] (avg. length)	Uniquely mapped reads[b] (%)	Non-specifically mapped reads (%)	Unmapped reads (%)
G2_1	13884440 (100 nt)	13884423 (95.4 nt)	10884216 (78.4%)	246422 (1.8%)	2753785 (19.8%)
G2_2	13271722 (100 nt)	13271697 (95.3 nt)	10310290 (77.7%)	188779 (1.4%)	2772628 (20.9%)
PASC_1	14248635 (100 nt)	14248606 (95 nt)	9474246 (66.5%)	209414 (1.5%)	4564946 (32.0%)
PASC_2	19983757 (100 nt)	19983724 (95 nt)	12899129 (64.5%)	315624 (1.6%)	6768971 (33.9%)
WAX_1	14312298 (100 nt)	14312281 (95.5 nt)	11083337 (77.4%)	229221 (1.6%)	2999723 (21.0%)
WAX_2	13335050 (100 nt)	13335026 (95.3 nt)	10002369 (75.0%)	243412 (1.8%)	3089245 (23.2%)
RNAseq Sample ID (*R. albus* 8)	Total reads (avg. length)	Reads after Trimming[a] (avg. length)	Uniquely mapped reads[b] (%)	Non-specifically mapped reads (%)	Unmapped reads (%)
G2_1	11101424 (100 nt)	11101400 (93.8 nt)	8491613 (76.5%)	14498 (0.1%)	2595289 (23.4%)
G2_2	9900094 (100 nt)	9900078 (94.1 nt)	7409361 (74.8%)	13543 (0.1%)	2477174 (25.0%)
PASC_1	12321261 (100 nt)	12321234 (93.9 nt)	8307217 (67.4%)	13340 (0.1%)	4000677 (32.5%)
PASC_2	7949411 (100 nt)	7949398 (94.2 nt)	5601618 (70.5%)	9009 (0.1%)	2338771 (29.4%)
WAX_1	11742722 (100 nt)	11724707 (93.9 nt)	8820790 (75.2%)	16627 (0.1%)	2905290 (24.8%)
WAX_2	7101989 (100 nt)	7101976 (93.9 nt)	5324747 (75.0%)	10066 (0.1%)	1767163 (24.9%)

[a] Reads were trimmed using CLC Genomics Workbench v5.5.1 with a quality score limit of 0.05 and maximum number of ambiguities of 2

[b] Reads were mapped to the *Ruminococcus albus* genome using CLC Genomics Workbench v5.5.1 with a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and maximum number of hits for a read of 10.

Table A.4 RNAseq mapping results of *R. albus* 7 grown on AHPCS

RNAseq Sample ID	Total reads (avg. length)	Reads after Trimming[a] (avg. length)	Uniquely mapped reads[b] (%)	Non-specificall mapped reads (%)	Unmapped reads (%)
Ra7_1_4h	23,671,352 (100 nt)	23,647,170 (99.9 nt)	19,607,620 (82.9 %)	300,812 (1.3 %)	3,738,738 (15.8 %)
Ra7_2_4h	24,664,042 (100 nt)	24,620,534 (99.9 nt)	20,360,774 (82.7 %)	314,297 (1.3 %)	3,945,463 (16.0 %)
Ra7_1_8h	23,645,746 (100 nt)	23,617,412 (99.9 nt)	18,182,125 (77.0 %)	484,816 (2.1 %)	4,950,471 (21.0 %)
Ra7_2_8h	25,673,585 (100 nt)	25,636,165 (99.9 nt)	20,539,769 (80.1 %)	591,529 (2.3 %)	4,504,867 (17.6 %)
Ra7_1_12h	23,357,375 (100 nt)	23,335,851 (99.9 nt)	14,382,676 (61.6 %)	562,918 (2.4 %)	8,390,257 (36.0 %)
Ra7_2_12h	21,945,634 (100 nt)	21,920,175 (99.9 nt)	13,796,122 (62.9 %)	524,967 (2.4 %)	7,599,086 (34.7 %)
Ra7_1_20h	22,780,551 (100 nt)	22,758,649 (99.9 nt)	11,256,046 (49.5 %)	452,151 (2.0 %)	11,050,452 (48.6 %)
Ra7_2_20h	22,344,255 (100 nt)	22,315,496 (99.9 nt)	10,802,576 (48.4 %)	508,763 (2.3 %)	11,004,157 (49.3 %)
Ra7_1_36h	23,203,443 (100 nt)	23,156,864 (99.9 nt)	11,161,208 (48.2 %)	461,066 (2.0 %)	11,534,590 (49.8 %)
Ra7_2_36h	22,082,227 (100 nt)	22,047,210 (99.9 nt)	11,157,394 (50.6 %)	446,828 (2.0 %)	10,442,988 (47.4 %)

[a] Reads were trimmed using CLC Genomics Workbench v5.5.1 with a quality score limit of 0.05 and maximum number of ambiguities of 2

[b] Reads were mapped to the *R. albus* 7 genome with annotation using CLC Genomics Workbench v5.5.1 with a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and maximum number of hits for a read of 10.

Table A.5 RNAseq mapping results of *R. albus* 8 grown on AHPCS

RNAseq Sample ID	Total reads (avg. length)	Reads after Trimming[a] (avg. length)	Uniquely mapped reads[b] (%)	Non-specifically mapped reads (%)	Unmapped reads (%)
Ra8_1_4h	22,833,478 (100 nt)	22,812,422 (99.9 nt)	18,915,697 (82.9 %)	57,725 (0.3 %)	3,839,000 (16.8 %)
Ra8_2_4h	23,773,956 (100 nt)	23,751,750 (99.9 nt)	19,164,964 (80.7 %)	61,158 (0.3 %)	4,525,628 (19.1 %)
Ra8_1_8h	22,957,201 (100 nt)	22,936,311 (99.9 nt)	19,155,817 (83.5 %)	46,265 (0.2 %)	3,734,229 (16.3 %)
Ra8_2_8h	22,539,094 (100 nt)	22,516,573 (99.9 nt)	18,914,245 (84.0 %)	45,593 (0.2 %)	3,556,735 (15.8 %)
Ra8_1_12h	23,018,947 (100 nt)	22,993,203 (99.9 nt)	16,896,712 (73.5 %)	36,739 (0.2 %)	6,059,752 (26.4 %)
Ra8_2_12h	20,983,054 (100 nt)	20,958,924 (99.9 nt)	14,718,875 (70.2 %)	33,138 (0.2 %)	6,206,911 (29.6 %)
Ra8_1_20h	22,622,493 (100 nt)	22,594,984 (99.9 nt)	14,385,901 (63.7 %)	21,518 (0.1 %)	8,187,565 (36.2 %)
Ra8_2_20h	23,861,439 (100 nt)	23,832,809 (99.9 nt)	13,923,593 (58.4 %)	22,725 (0.1 %)	9,886,491 (41.5 %)
Ra8_1_36h	22,197,689 (100 nt)	22,170,057 (99.9 nt)	14,100,963 (63.6 %)	19,864 (0.1 %)	8,049,230 (36.3 %)
Ra8_2_36h	22,369,699 (100 nt)	22,345,394 (99.9 nt)	13,295,059 (59.5 %)	18,221 (0.1 %)	9,032,114 (40.4 %)

[a] Reads were trimmed using CLC Genomics Workbench v5.5.1 with a quality score limit of 0.05 and maximum number of ambiguities of 2

[b] Reads were mapped to the *R. albus* 8 genome with annotation using CLC Genomics Workbench v5.5.1 with a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and maximum number of hits for a read of 10.

Table A.6 Comparison of the predicted GHs, PLs, and CEs genes of *R. albus* strain 7 and 8

GH family	2	3	4	5	8	9	10	11	13	16	23	25	26	27	28	30	31	36	39	43	44	48	51	53	67	74	77	94	95	98	99	105	109	113	114	124	130	Total
R. albus 7	3	5	1	14	1	8	5	5	5	2	1	5	8	2	1	2	1	3	1	6	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	2	99
R. albus 8	3	6	1	15	1	7	3	3	5	1	1	5	6	1	1	1	1	2	0	9	1	1	1	1	1	2	1	2	2	0	0	2	1	1	0	0	2	90

PL family	1	9	10	11	Total
R. albus 7	3	1	1	2	7
R. albus 8	2	1	1	2	6

CE family	1	2	3	4	7	8	9	10	12	Total
R. albus 7	4	4	2	5	1	2	1	2	3	24
R. albus 8	2	5	3	5	1	3	1	4	3	27

The grey color means that same numbers of CAZy genes are present on the genomes of both strains. The orange color means that more genes are present on the genome of *R. albus* 7, and the blue color means that more genes are present on the genome of *R. albus* 8.

Table A.7 Chemical analysis of alkaline hydrogen peroxide treated cornstalk (AHPCS)

total glucan [%]	cellulosic glucan [%]	hemicellulosic glucan [%]	xylan [%]	arabinan [%]	acetyl [%]	Klason lignin [%]
58.30	53.708	4.587	17.18	1.22	0.16	2.02

Polysaccharides composition of AHPCS was determined by a previously described method (Ibáñez and Bauer, 2014).

Table A.8 Monosaccharides composition of dry residue and culture supernatant in culture during growth on AHPCS

		Monosaccharides in dry residue (mM ± SD)					
	Incubation time (h)	Glucose	Disappearance	Xylose	Disappearance	Arabinose	Disappearance
R. albus 7	0	16.27 ± 0.51	0.0 %	4.49 ± 0.26	0.0 %	0.31 ± 0.00	0.0 %
	12	13.20 ± 0.67	18.9 %	2.86 ± 0.15	36.3 %	0.16 ± 0.02	48.4 %
	24	5.92 ± 0.31	63.6 %	0.93 ± 0.07	79.3 %	0.06 ± 0.01	80.6 %
	48	2.53 ± 0.11	84.4 %	0.34 ±0.02	92.4 %	0.04 ± 0.00	87.1 %
R. albus 8	0	15.37 ± 0.49	0.0 %	4.38 ± 0.17	0.0 %	0.30 ± 0.01	0.0 %
	12	15.30 ± 0.51	0.5 %	3.60 ± 0.12	17.8 %	0.21 ± 0.01	30.0 %
	24	7.09 ± 0.65	53.9 %	1.39 ± 0.16	68.3 %	0.08 ± 0.00	73.3 %
	48	2.12 ± 0.07	86.2 %	0.27 ± 0.02	93.8 %	0.05 ± 0.00	83.3 %
		Monosaccharides in supernatant (µM ± SD)					
	Incubation time (h)	Glucose	Disappearance	Xylose	Disappearance	Arabinose	Disappearance
R. albus 7	0	13.19 ± 0.44	0.0 %	74.90 ± 2.17	0.0 %	11.12 ± 0.28	0.0 %
	12	14.07 ± 1.45	-6.7 %	91.14 ± 22.23	-21.7 %	14.51 ± 0.88	-30.5 %
	24	26.88 ± 0.47	-103.8 %	89.38 ± 3.31	-19.3 %	15.83 ± 0.32	-42.4 %
	48	30.11 ± 0.24	-128.3 %	75.21 ± 1.65	-0.4 %	15.45 ± 0.51	-38.9 %
R. albus 8	0	11.71 ± 0.88	0.0 %	64.32 ± 5.11	0.0 %	10.73 ± 0.79	0.0 %
	12	9.97 ± 0.20	14.9 %	78.10 ± 1.80	-21.4 %	13.84 ± 0.48	-29.0 %
	24	14.26 ± 0.38	-21.8 %	54.84 ± 1.87	14.7 %	14.21 ± 0.44	-32.4 %
	48	18.64 ± 0.13	-59.2 %	54.22 ± 3.76	15.7 %	14.07 ± 0.52	-31.1 %

Values are reported from the mean for three biological replicates. Disappearance value was calculated using comparison of sugar amount measured at 0 h with disappeared sugar amount in each fraction at each sampling time.

Table A.9 PSI-blast results of XFP gene of *R. albus* 8 to previously characterized XFP genes

Bacterial species	Genbank ID	AA seq. identity	AA seq. identity (%)	*E*-value
Clostridium acetobutylicum ATCC 824	NP_347971	Pubmed ID: 22922942, 22865845 25527534	73	0
Bifidobacterium breve	ADF97524	Pubmed ID: 20693675	49	0
Bifidobacterium longum	3AI7_A	Pubmed ID: 20674574	48	0
Bifidobacterium lactis	CAC29121	Pubmed ID: 11292814	46	0
Lactobacillus pentosus	XPKA_LACPE	Pubmed ID: 11823225	58	0

Table A.10 Predicted metabolic output during glucose and pentose metabolism via the EMP, PPP, and XFP pathway in *R. albus* 7 and 8

Catabolic pathway	Strain	H_2 level	G-3-P formed	NADH - EMP	Reduced ferredoxin[a]	Acetate production	Ethanol production	CO_2 production	H_2 production	Max net ATP[b]
5 Glucose -EMP	*R. albus* 7 and 8	Very low	10	10	10	10	0	10	20	20
		Very high	10	10	10	5	5	10	10	15
6 Pentose -PPP	*R. albus* 7 and 8	Very low	10	10	10	10	0	10	20	20
		Very high	10	10	10	5	5	10	10	15
6 Pentose -XFP	*R. albus* 8	Very low	6	6	6	12	0	6	12	18
		Very high	6	6	6	9	3	6	6	15

[a] Theoretical maximum number was calculated, based on the assumption that all generated pyruvates are converted to acetyl-CoA by pyruvate ferredoxin oxidoreductase.

[b] Maximum yield of net ATP from sugar fermentation was calculated, based on the assumption that all reduced ferredoxins were re-oxidized by electron-bifurcating hydrogenase at low hydrogen level and ferredoxin-dependent hydrogenase at high hydrogen level (Zheng *et al.*, 2014).

Table A.11 The expression of unique genes in each PP and PK pathways during growth of *R. albus* 7 and 8 on AHPCS

		Normalized RPKM on AHPCS					Normalized RPKM at mid-log phase		
Annotated function	Pathway	4h	8h	12h	20h	36h	Cellobiose	PASC	WAX
R. albus 7									
Transketolase[a]	PP pathway	2002	930	629	364	278	539	424	1078
Transaldolase	PP pathway	2589	1642	1371	625	490	10	200	6241
R. albus 8									
D-xylulose 5-phosphate /D-fructose 6-phosphate phosphoketolase	XFP pathway	2241	820	174	66	54	41	24	266
Transketolase[a]	PP pathway	985	1088	728	396	298	426	403	808
Transaldolase	PP pathway	5257	3188	2328	1636	2105	7	32	2473
A ratio of XFP to Transketolase:		2.3	0.8	0.2	0.2	0.2	0.1	0.1	0.3

[a] Both strains possess two genes encoding N-terminal section of transketolase and C-terminal section of transketolase, respectively. The RPKM value of transketolase is shown as the average of RPKM between two genes.

Table A.12 The number of predicted diguanylate cyclases (encoding GGDEF domain), diguanylate phosphodiesterases (encoding EAL domain), and diguanylate cyclae/phosphodiesterase (encoding both GGDEF and EAL domain) on the genome of *Ruminococcus* species and other plant cell wall degrading bacteria

Phylum	Genus	species	strain	GGDEF	EAL	GGDEF & EAL	total	Genome status	Size (Mb)	GC (%)	Gene
Firmicutes	*Ruminococcus*	*albus*	7	24	6	8	38	Complete	4.49	41.2	3872
		albus	8	29	9	0	38	draft	4.05	46.6	3899
		albus	SY3	27	5	0	32	draft	4.29	44.7	3790
		flavefaciens	FD-1	20	0	4	24	draft	4.57	45.0	3722
		flavefaciens	17	3	0	0	3	draft	3.45	44.9	2934
		champanellensis	18P13	9	0	0	9	draft	2.57	53.3	2154
		torques	L2-14	5	0	0	5	draft	3.34	41.1	2838
		bromii	L2-63	1	0	0	1	draft	2.25	41.4	1852
Firmicutes	*Blautia*	*obeum*	A2-162	9	2	0	11	draft	3.76	42.6	3200
	Ruminiclostridium	*thermocellum*	ATCC27405	6	0	1	7	Complete	3.84	39.0	3363
		thermocellum	DSM1313	8	0	1	9	Complete	3.56	39.1	3102
	Clostridium	*cellulolyticum*	H10	3	0	1	4	Complete	4.07	37.4	3569
Fibrobacteres	*Fibrobacter*	*succinogenes*	S85	22	1	4	27	Complete	3.84	48.1	3188
Bacteroidetes	*Prevotella*	*ruminicola*	23	0	0	0	0	Complete	3.62	47.7	2860

Table A.13 The expression level of Agr quorum sensing genes in *R. albus* 7 and 8 during growth on AHPCS

Time	*R. albus* 7 (RPKM)		*R. albus* 8 (RPKM)	
	agrB	*agrD*	*agrB*	*agrD*
4h (lag)	106	87	2,976	1,251
8h (early-log)	941	1,638	23,298	7,134
12h (mid-log)	1,104	8,318	43,196	11,181
20h (late-log)	560	2,973	35,328	10,794
36h (stationary)	397	2,521	22,444	8,262

Table A.14 The expression of a Pil-like protein cluster in *R. albus* 7 and 8 during growth on AHPCS and defined substrates

			Normalized RPKM mean between two replicates							
Function	*R. albus* 7 locus #	Annotation	4h	8h	12h	20h	36h	Cellobiose	PASC	WAX
Pilin locus	Rumal_0365	type IV pilin (CbpC homolog)	19,053	38,939	46,646	36,007	31,786	27,101	48,934	36,501
	Rumal_0366	type IV pilin (Cus_7104 homolog)	224	409	653	794	850	463	829	665
	Rumal_0367	protein-export membrane protein SecD	1,208	977	447	261	201	1,154	670	1,361
	Rumal_0368	protein-export membrane protein SecF	1,269	1,147	532	291	260	2,121	1,250	2,399
			Normalized RPKM mean between two replicates							
Function	*R. albus* 8 locus #	Annotation	4h	8h	12h	20h	36h	Cellobiose	PASC	WAX
Pilin locus	CUS_7103	pilA protein (CpbC)	10,671	10,243	24,755	40,960	29,870	18,802	18,695	63,844
	CUS_7104	type IV pilin	663	557	1,308	1,646	1,475	1,196	2,089	1,952
	CUS_7105	export membrane protein SecD	840	775	339	179	173	972	333	378
	CUS_7106	export membrane protein SecF	1,098	972	475	230	225	1,282	439	626

Table A.15 Transcriptional level of genes encoding putative glycoside transferase and hexose-1-P uridylyltransferase in *R. albus* 7

			Normalized RPKM					Normalized RPKM			*p*-value					
Locus tag #	Predicted annotation	Genbank ID	4h	8h	12h	20h	36h	Cellobiose (G2)	PASC	WAX	8 vs 4 h	12 vs 4 h	20 vs 4 h	36 vs 4 h	PASC vs G2	WAX vs G2
Rumal_0027	GT2	ADU20590	58	66	96	85	85	87	110	96	0.38	0.00	0.01	0.01	0.04	0.47
Rumal_0300	GT2	ADU20857	78	134	202	157	134	51	81	63	0.00	0.00	0.00	0.00	0.00	0.08
Rumal_0395	GT2	ADU20950	362	256	141	101	91	108	47	106	0.03	0.00	0.00	0.00	0.00	0.95
Rumal_0396	GT2	ADU20951	361	243	124	82	73	111	55	101	0.04	0.00	0.00	0.00	0.00	0.59
Rumal_0436	GT2	ADU20990	4	2	3	4	18	2	1	1	0.26	0.50	0.81	0.28	0.54	0.24
Rumal_2484	GT2	ADU22964	120	128	85	70	58	114	76	69	0.45	0.00	0.00	0.00	0.00	0.00
Rumal_2820	GT2	ADU23289	333	323	216	175	149	168	126	179	0.70	0.00	0.00	0.00	0.00	0.38
Rumal_3528	GT2	ADU23970	52	54	50	51	48	33	23	27	0.84	0.55	0.53	0.28	0.04	0.23
Rumal_3607	GT2	ADU24048	132	173	241	109	128	247	182	331	0.50	0.00	0.05	0.55	0.00	0.03
Rumal_3858	GT2	ADU24282	4	5	5	2	8	4	7	6	0.69	0.55	0.29	0.42	0.12	0.44
Rumal_3861	GT2	ADU24285	3	4	4	2	5	3	5	3	0.58	0.63	0.63	0.51	0.47	1.00
Rumal_0397	GT4	ADU20952	394	254	144	104	83	149	80	126	0.00	0.00	0.00	0.00	0.00	0.19
Rumal_0398	GT4	ADU20953	320	195	121	87	73	106	49	85	0.00	0.00	0.00	0.00	0.00	0.15
Rumal_0404	GT4	ADU20959	188	269	271	164	148	221	255	168	0.00	0.00	0.01	0.00	0.15	0.01
Rumal_3595	GT4	ADU24036	196	151	100	60	62	120	96	125	0.01	0.00	0.00	0.00	0.09	0.81
Rumal_3599	GT4	ADU24040	188	223	236	100	116	266	108	433	0.67	0.01	0.00	0.00	0.00	0.07
Rumal_3852	GT4	ADU24276	3	5	5	3	8	3	3	3	0.20	0.33	0.75	0.19	0.90	0.72
Rumal_3862	GT4	ADU24286	4	4	4	2	6	2	5	2	0.70	0.92	0.34	0.55	0.12	0.72
Rumal_0248	GT5	ADU20805	116	150	279	213	170	33	92	53	0.20	0.00	0.00	0.00	0.00	0.00
Rumal_1847	GT5	ADU22345	165	202	250	209	181	188	278	249	0.01	0.00	0.02	0.84	0.00	0.01
Rumal_3052	GT28	ADU23517	316	214	182	186	149	257	206	267	0.00	0.00	0.00	0.00	0.00	0.55
Rumal_0466	GT35	ADU21019	181	215	191	171	145	79	83	70	0.36	0.86	0.12	0.00	0.82	0.32
Rumal_2782	GT35	ADU23251	124	136	121	104	89	61	63	57	0.71	0.47	0.01	0.00	0.92	0.69
Rumal_1775	GT51	ADU22273	469	402	394	308	329	1444	962	1876	0.30	0.00	0.00	0.00	0.00	0.04
Rumal_2095	Galactose-1-P uridylyltransferase	ADU22584	124	100	95	92	87	38	39	37	0.03	0.00	0.00	0.00	0.95	0.93

Table A.16 Transcriptional level of genes encoding putative glycoside transferase and hexose-1-P uridylyltransferase in *R. albus* 8

			Normalized RPKM					Normalized RPKM			*p*-value					
Locus tag #	Predicted GT family	Genbank ID	4h	8h	12h	20h	36h	Cellobiose (G2)	PASC	WAX	8 vs 4 h	12 vs 4 h	20 vs 4 h	36 vs 4 h	PASC vs G2	WAX vs G2
CUS_4414	GT2	EGC04861	115	107	213	112	57	0	0	0	0.29	0.00	0.59	0.00	1.00	1.00
CUS_5984	GT2	EGC02360	227	303	237	109	101	231	230	358	0.00	0.69	0.00	0.00	0.26	0.00
CUS_6070	GT2	EGC02770	185	124	93	55	51	137	118	154	0.00	0.00	0.00	0.00	0.02	0.11
CUS_6174	GT2	EGC04758	55	50	32	18	16	31	25	42	0.31	0.00	0.00	0.00	0.16	0.05
CUS_6341	GT2	EGC01683	188	181	93	51	35	113	70	63	0.46	0.00	0.00	0.00	0.00	0.00
CUS_6342	GT2	EGC01665	221	201	110	52	39	120	40	59	0.23	0.00	0.00	0.00	0.00	0.00
CUS_6523	GT2	EGC03906	41	49	59	51	40	30	23	22	0.37	0.09	0.40	0.19	0.10	0.18
CUS_7423	GT2	EGC04606	50	44	40	13	11	69	47	89	0.24	0.09	0.00	0.00	0.00	0.02
CUS_4803	GT4	EGC04484	588	505	446	200	124	702	337	505	0.04	0.00	0.00	0.00	0.00	0.00
CUS_4804	GT4	EGC04498	527	485	336	154	115	537	382	313	0.26	0.00	0.00	0.00	0.00	0.00
CUS_4806	GT4	EGC04503	361	311	244	115	95	538	511	356	0.02	0.00	0.00	0.00	0.11	0.00
CUS_5332	GT4	EGC04381	92	114	114	44	26	25	16	16	0.13	0.23	0.00	0.00	0.03	0.07
CUS_6333	GT4	EGC01659	314	342	304	142	117	188	180	105	0.57	0.60	0.00	0.00	0.33	0.00
CUS_6339	GT4	EGC01644	251	235	110	62	45	138	72	77	0.21	0.00	0.00	0.00	0.00	0.00
CUS_6340	GT4	EGC01650	233	217	104	55	46	130	67	61	0.26	0.00	0.00	0.00	0.00	0.00
CUS_5970	GT5	EGC02409	51	123	218	175	113	92	122	152	0.00	0.00	0.01	0.05	0.05	0.00
CUS_7757	GT5	EGC01898	121	124	121	76	61	192	241	100	0.87	0.80	0.00	0.00	0.08	0.00
CUS_4341	GT27	EGC02980	98	106	243	117	54	0	0	0	0.66	0.00	0.68	0.00	1.00	1.00
CUS_6554	GT28	EGC03406	308	327	241	151	115	218	156	247	0.71	0.00	0.00	0.00	0.00	0.04
CUS_5705	GT35	EGC01563	508	390	176	71	61	342	156	200	0.00	0.00	0.00	0.00	0.00	0.00
CUS_6587	GT35	EGC03478	144	199	123	62	49	69	54	48	0.01	0.17	0.00	0.00	0.02	0.01
CUS_7775	GT51	EGC01901	1016	859	830	515	388	1336	933	1076	0.00	0.00	0.00	0.00	0.00	0.00
CUS_4474	GT81	EGC01942	348	250	185	99	89	260	127	228	0.00	0.00	0.00	0.00	0.00	0.06
CUS_6182	UDP-Hexose-1-P uridylyl transferase	EGC04839	165	130	87	60	60	29	43	36	0.00	0.00	0.00	0.00	0.05	0.23

Table A.17 The expression of unknown function of CBM37 genes in *R. albus* 7 during growth on AHPCS

			Normalized RPKM				
CBM family	Locus tag	Accession #	4h	8h	12h	20h	36h
37	Rumal_0405	ADU20960	1428.9	3591.8	3132.8	1722.8	1276.1
37	Rumal_0897	ADU21424	485.3	1451.1	1092.3	658.8	505.5
37	Rumal_0909	ADU21435	589.2	776.7	486.3	344.7	273.0
37	Rumal_1155	ADU21675	1074.5	653.8	274.6	137.7	120.7
37	Rumal_2745	ADU23214	262.9	645.1	617.3	453.1	367.4
37	Rumal_2457	ADU22937	292.9	440.9	548.3	469.9	455.7
37	Rumal_1044	ADU21570	213.8	423.6	281.8	150.9	108.2
37	Rumal_0777	ADU21311	127.7	384.9	328.4	270.4	214.9
37	Rumal_2543	ADU23018	168.3	311.1	172.9	95.0	75.5
37	Rumal_3740	ADU24175	106.2	271.9	225.7	116.6	99.2
37	Rumal_2536	ADU23011	98.4	238.8	280.5	224.3	187.5
37	Rumal_3382	ADU23841	56.3	231.6	210.6	125.5	122.1
37	Rumal_0330	ADU20887	212.8	217.2	182.5	124.2	111.0
37	Rumal_1442	ADU21950	56.5	211.0	180.8	129.3	102.5
37	Rumal_2456	ADU22936	186.6	209.6	139.3	79.9	68.7
37	Rumal_0814	ADU21345	78.3	196.6	175.2	176.6	153.2
37	Rumal_1452	ADU21958	152.7	184.9	101.0	34.2	34.0
37	Rumal_2706	ADU23178	356.4	176.1	78.8	26.6	29.3
37	Rumal_2018	ADU22511	44.7	176.1	120.0	52.1	44.2
37	Rumal_3360	ADU23820	81.0	172.7	191.2	110.5	102.4
37	Rumal_0778	ADU21312	74.2	162.9	92.6	59.2	48.8
37	Rumal_1951	ADU22444	69.2	155.6	167.2	133.9	143.4
37	Rumal_0406	ADU20961	57.8	128.3	118.4	68.8	57.9
37	Rumal_2377	ADU22857	50.0	103.5	1399.3	1754.2	1529.7
37	Rumal_3358	ADU23819	74.9	83.2	49.3	50.3	58.6
37	Rumal_0892	ADU21419	30.1	68.7	67.5	55.3	49.6
37	Rumal_0746	ADU21284	39.4	54.3	51.0	38.4	41.7
37	Rumal_1446	ADU21954	14.3	18.3	14.2	9.6	11.0
37	Rumal_0743	ADU21281	14.1	17.9	21.6	16.6	27.3
37	Rumal_1157	ADU21677	13.4	10.1	8.4	4.1	8.0
37	Rumal_0150	ADU20712	12.9	9.7	6.8	4.1	11.3
37	Rumal_0515	ADU21067	7.7	4.4	6.8	4.1	10.8
37	Rumal_0747	ADU21285	4.0	1.1	1.7	0.9	3.2
37	Rumal_0326	ADU20883	1.0	0.4	1.2	1.5	5.3
13, 37	Rumal_3330	ADU23792	37.1	50.2	40.0	27.6	29.9

Table A.18 The expression of unknown function of CBM37 genes in *R. albus* 8 during growth on AHPCS

			Normalized RPKM				
CBM family	Locus tag	Accession #	4h	8h	12h	20h	36h
37	CUS_4476	EGC01945.1	7.1	46.9	448.2	254.4	80.6
37	CUS_4934	EGC04726.1	1010.3	1995.1	1649.6	694.8	622.9
37	CUS_5144	EGC04022.1	13.4	5.2	8.7	1.8	1.6
37	CUS_5145	EGC04014.1	219.9	369.2	499.4	243.0	169.4
37	CUS_5322	EGC04411.1	436.7	482.8	373.8	215.3	152.5
37	CUS_5392	EGC01068.1	106.0	180.5	254.5	110.5	80.0
37	CUS_5426	EGC02492.1	0.0	0.0	0.0	0.0	0.0
37	CUS_5429	EGC02463.1	0.0	0.0	0.0	0.0	0.0
37	CUS_5567	EGC04257.1	142.0	258.6	178.6	81.8	82.7
37	CUS_5662	EGC02136.1	0.0	0.0	0.0	0.0	0.0
37	CUS_5760	EGC02255.1	62.7	55.8	61.3	21.5	12.0
37	CUS_5801	EGC02252.1	114.4	38.9	30.0	13.4	8.7
37	CUS_6025	EGC02760.1	81.1	105.6	84.6	33.1	31.4
37	CUS_6027	EGC02784.1	82.0	83.9	50.4	16.2	20.9
37	CUS_6258	EGC04131.1	1324.7	2490.8	1580.2	627.0	552.7
37	CUS_6331	EGC01667.1	1890.6	3331.2	2378.5	1115.7	1218.4
37	CUS_6332	EGC01693.1	247.2	208.6	133.2	56.2	85.0
37	CUS_6393	EGC01673.1	168.0	226.5	255.5	157.1	95.3
37	CUS_6570	EGC03432.1	386.2	371.0	211.0	111.5	105.4
37	CUS_6856	EGC03160.1	54.3	18.0	20.8	5.5	5.0
37	CUS_7056	EGC01237.1	22.8	15.3	44.9	18.7	9.9
37	CUS_7085	EGC02602.1	869.9	1092.1	780.6	460.2	380.4
37	CUS_7780	EGC01841.1	18.4	34.4	81.3	22.2	7.6

Table A.19 PSI-blast result of XFP gene of *R. albus* 8

Bacterial species	Genbank ID	AA seq. identity (%)	*E*-value
Ruminococcus albus 8	EGC03698	100.0	0
Ruminococcus sp. N15.MGS-57	WP_046439888	90.8	0
Ruminococcus sp. CAG:579	CDA72973	86.7	0
Eubacterium sp. CAG:248	CDB66887	82.8	0
Firmicutes bacterium CAG:41	CDB94315	82.3	0
Eubacterium sp. CAG:252	CDB69115	82.8	0
Eubacterium sp. CAG:86	CCX84702	82.4	0
Eubacterium sp. CAG:76	CDF09696	82.7	0
Eubacterium sp. CAG:38	CDE36108	81.9	0
Clostridium sp. CAG:122	CCZ41637	81.2	0
Roseburia sp. CAG:18	CCZ79117	80.1	0
Roseburia intestinalis L1-82	EEV01821	80.0	0
Roseburia	CDA56372	80.0	0
Lachnospiraceae bacterium AC2014	WP_034219286	79.0	0
Butyrivibrio	WP_026656520	78.6	0
Butyrivibrio sp. LB2008	WP_026514362	78.4	0
Butyrivibrio hungatei	WP_027209225	78.4	0
Lachnospiraceae bacterium NK4A179	WP_022784606	79.0	0
Clostridium sp. KLE 1755	ERI66602	80.3	0
Firmicutes bacterium CAG:94	CDD29974	78.5	0
Clostridiales bacterium VE202-27	WP_025488577	80.1	0
Ruminococcus sp. CAG:9	CDD78970	78.5	0
Subdoligranulum variabile DSM 15176	EFB77568	78.4	0
Subdoligranulum variabile	WP_040917362	78.4	0
Dorea longicatena	WP_028087269	78.4	0
Pseudobutyrivibrio ruminis	WP_028243311	77.3	0
Dorea longicatena CAG:42	CDE17103	78.2	0
Lachnospiraceae bacterium AC2031	WP_027440906	77.1	0
Dorea longicatena	EDM62349	78.2	0
Coprococcus	CDB85317	78.3	0
Oscillibacter ruminantium	WP_040663351	77.3	0
Eubacterium xylanophilum	WP_026835763	79.2	0
Pseudobutyrivibrio sp. MD2005	WP_028234503	76.7	0
Clostridium sp. CAG:7	CCY43738	77.5	0
Butyrivibrio sp. CAG:318	CDC38768	76.9	0
Oscillibacter valericigenes	BAK98480	76.6	0
Firmicutes bacterium CAG:227	CDC91934	76.6	0

Table A.19 (Cont.)

Bacterial species	Genbank ID	AA seq. identity (%)	*E*-value
Roseburia sp. CAG:380	CDC94423	76.0	0
Roseburia sp. CAG:309	CDD35913	77.9	0
Clostridium sp. CAG:277	CDE68000	75.5	0
Firmicutes bacterium CAG:114	CCY25958	75.2	0
Clostridium sp. CAG:75	CCZ52238	75.4	0
Roseburia sp. CAG:197	CDA25989	75.6	0
Clostridium viride	WP_035139223	75.2	0
Butyrivibrio sp. NC3005	WP_026504437	75.5	0
Butyrivibrio fibrisolvens	WP_022754901	76.7	0
Butyrivibrio fibrisolvens	WP_027205600	76.6	0
Firmicutes bacterium CAG:238	CDA90055	75.0	0
Butyrivibrio fibrisolvens	WP_022757821	76.3	0
Clostridium sp. CAG:307	CDE26872	74.7	0