Global analysis of carbohydrate utilization by *Lactobacillus acidophilus* using cDNA microarrays

Rodolphe Barrangou*†‡, M. Andrea Azcarate-Peril†, Tri Duong*§, Shannon B. Conners*§, Robert M. Kelly§, and Todd R. Klaenhammer†¶

*Genomic Sciences Graduate Program and Departments of †Food Science and §Chemical Engineering, North Carolina State University, Raleigh, NC 27695

Contributed by Todd R. Klaenhammer, January 5, 2006

The transport and catabolic machinery involved in carbohydrate utilization by *Lactobacillus acidophilus* was characterized genetically by using whole-genome cDNA microarrays. Global transcriptional profiles were determined for growth on glucose, fructose, sucrose, lactose, galactose, trehalose, raffinose, and fructooligosaccharides. Hybridizations were carried out by using a round-robin design, and microarray data were analyzed with a two-stage mixed model ANOVA. Differentially expressed genes were visualized by hierarchical clustering, volcano plots, and contour plots. Overall, only 63 genes (3% of the genome) showed a >4-fold induction. Specifically, transporters of the phosphoenolpyruvate:sugar transferase system were identified for uptake of glucose, fructose, sucrose, and trehalose, whereas ATP-binding cassette transporters were identified for uptake of raffinose and fructooligosaccharides. A member of the LacS subfamily of galactoside-pentose hexuronide translocators was identified for uptake of galactose and lactose. Saccharolytic enzymes likely involved in the metabolism of monosaccharides, disaccharides, and polysaccharides into substrates of glycolysis were also found, including enzymatic machinery of the Leloir pathway. The transcriptome appeared to be regulated by carbon catabolite repression. Although substrate-specific carbohydrate transporters and hydrolyses were regulated at the transcriptional level, genes encoding regulatory proteins CcpA, Hpr, HprK/P, and EI were consistently highly expressed. Genes central to glycolysis were among the most highly expressed in the genome. Collectively, microarray data revealed that coordinated and regulated transcription of genes involved in sugar uptake and metabolism is based on the specific carbohydrate provided. *L. acidophilus*’s adaptability to environmental conditions likely contributes to its competitive ability for limited carbohydrate sources available in the human gastrointestinal tract.

ATP-binding cassette | carbon catabolite repression | fructooligosaccharide | galactoside-pentose hexuronide

A large, diverse, and dynamic microbial community resides in the human gastrointestinal tract (GIT) (1). In particular, the complex intestinal microbial population includes beneficial bacteria, such as bifidobacteria and lactobacilli (2). Among species considered important for human health, a number of documented lactobacilli have been characterized as probiotics (3), which are defined as “live microorganisms which, when administered in adequate amounts confer a health benefit on the host” (4). For such microbes, survival and residence in the GIT relies partially on their ability to survive gastric passage and use available nutrients. *Lactobacillus acidophilus* NCFM (North Carolina Food Microbiology) is a Gram-positive lactic acid bacterium that has the ability to survive in the GIT (5, 6), adhere to human epithelial cells *in vitro* (5, 7), modify fecal flora (6), modulate the host immune response (8), and prevent microbial gastroenteritis (8). Additionally, *L. acidophilus* NCFM has the ability to use nondigestible oligosaccharides, which may also contribute to the organism’s ability to compete in the human GIT (9).

Undigested carbohydrates are a primary source of energy for intestinal microbes residing in the large intestine. Nondigestible oligosaccharides (NDO) consist primarily of plant carbohydrates that are resistant to enzymatic degradation and are not absorbed in the upper intestinal tract. Such dietary compounds eventually reach the large intestine, whereby they are hydrolyzed by a limited range of organisms. As a result, NDO have the ability to selectively modulate the composition of the intestinal microflora (6). NDO, such as raffinose and fructooligosaccharides (FOS), have been shown to selectively promote the growth of probiotic species and are therefore considered prebiotic compounds (10, 11). Prebiotics are defined as nondigestible substances that provide a beneficial physiological effect on the host by selectively stimulating the favorable growth or activity of a limited number of indigenous bacteria (4). Although considerable attention has been devoted to studying modulation of the intestinal flora by prebiotics, the molecular mechanisms involved in uptake and metabolism of those compounds by desirable intestinal microbes remains mostly uncharacterized.

Lactic acid bacteria can use a variety of nutrients. Specifically, the genomes of lactobacilli and streptococci encode specialized saccharolytic machinery that reflects the nutrient availability in their respective environments (12–15). In particular, the versatile saccharolytic potential of *L. acidophilus* likely reflects its ability to efficiently use energy sources available in the intestinal environment. Although the *L. acidophilus* NCFM genome encodes numerous putative genes potentially involved in the uptake and metabolism of a variety of carbohydrates (16), little information is available regarding their biological functions and expression profiles.

The objective of this study was to identify genes involved in carbohydrate utilization by *L. acidophilus*. Global gene transcription profiles were used to identify uptake systems, catabolic machinery, and regulatory networks involved in the utilization of eight carbohydrates. This work is a comparative global transcriptional analysis of a lactic acid bacterium over a range of carbohydrates.

**Results**

**Differentially Expressed Genes.** Global gene expression patterns obtained from growth on eight different carbohydrates were...
visualized by cluster analysis (17) with Ward’s hierarchical clustering method (Fig. 1), volcano plots (Fig. 2), and contour plots (Fig. 3). Overall, between 23 and 379 genes were differentially expressed between paired treatment conditions (with P values below the Bonferroni correction), representing between 1% and 20% of the genome, respectively (see Fig. 5, which is published as supporting information on the PNAS web site). Although 342 genes (18% of the genome) showed induction levels of ≥2-fold, only 63 genes (3% of the genome) showed induction of ≥4-fold (see Fig. 6, which is published as supporting information on the PNAS web site), indicating that a relatively small number of genes were highly induced. Although overall expression levels of the majority of the genes remained consistent regardless of the growth substrate, select genes showed differential transcription (Fig. 1; see also Fig. 7, which is published as supporting information on the PNAS web site).

In the presence of glucose, ORFs La1679 and La1680 (Figs. 1 and 7) were highly induced compared with other monosaccharides (fructose and galactose) and disaccharides (sucrose, lactose, and trehalose). The induction levels compared with other sugars varied between 3.5 and 6.3 for La1679 and between 3.7 and 4.7 for La1680. La1679 encodes an ATP-binding cassette (ABC) nucleotide binding protein, and La1680 encodes an ABC permease. No solute binding protein was encoded in their vicinity, suggesting a possible role as an exporter rather than an importer. Several genes and operons were specifically repressed by glucose (see Figs. 1 and 7), including ORFs La680–686, which are involved in glycogen metabolism. Because glycogen is metabolized by the cell to store energy, in the presence of the preferred carbon source, such as glucose, energy storage is not necessary. Other genes potentially involved in uptake and hydrolysis of alternative carbohydrate sources were repressed in the presence of glucose.

The three genes of the putative fructose locus, La1777 [FruA, fructose phosphoenolpyruvate:sugar transferase system (PTS) transporter EIIABCFru], La1778 (FruK, phosphofructokinase), and La1779 (FruR, transcription regulator) were differentially expressed (Figs. 1 and 7). Induction levels were up to 3.9, 4.3, and 4.6 for fruA, fruK, and fruR, respectively. These results suggest that fructose is transported into the cell via a PTS transporter

Fig. 1. Hierarchical clustering of gene expression patterns. Gene expression (vertically) during growth on eight carbohydrates (horizontally) is shown colorimetrically. (A) Global gene expression for 1,889 genes. (B) Select genes and operons. Least-squares means represent overall gene expression level: low, blue; high, red. FRU, fructose; GAL, galactose; GLC, glucose; LAC, lactose; RAF, raffinose; SUC, sucrose; TRE, trehalose.

Fig. 2. Volcano plot comparison of gene expression between FOS and raffinose (RAFF). The x axis indicates the differential expression profiles, plotting the fold-induction ratios in a log-2 scale. The y axis indicates the statistical significance of the difference in expression (P value from a t test) in a log10 scale. Genes within the raffinose msm locus are shown in green, genes within the FOS msm locus are shown in blue, and two genes within the trehalose tre locus are shown in red.

Fig. 3. Contour plot comparison of gene expression among FOS, raffinose, and trehalose. Shown is a three-way plot of the least-squares means (Lsm) of all of the genes in the presence of FOS (x axis), raffinose (y axis), and trehalose (z axis). In the third dimension (z axis), the gene expression level is coded colorimetrically: blue, low gene expression; red, high gene expression. Also shown are differentially expressed genes: 1437–1442, raffinose msm2 operon; 502–507, FOS msm operon; and 1012–1014, trehalose tre locus.
into fructose-1-phosphate, which the phosphofructokinase FruK phosphorylates into fructose-1,6-bisphosphate.

In the presence of sucrose, the three genes of the sucrose locus were differentially expressed (Figs. 1 and 7), namely, La399 (ScrR, transcription regulator), La400 (ScrB, sucrose-6-phosphate hydrolase), and La401 (ScrA, sucrose PTS transporter EIIBCαE). Compared with glucose, induction levels were up to 3.1, 2.8, and 17.2 for ScrR, ScrB, and ScrA, respectively. La401 in particular showed high induction levels between 8.0 and 17.2 compared with monosaccharides and disaccharides. These results indicate that sucrose could be transported into the cell by a PTS transporter into sucrose-6-phosphate, which is subsequently hydrolyzed into glucose-6-phosphate and fructose by ScrB.

The six genes of the FOS operon were differentially expressed (Figs. 1–3 and 7), namely La502, La503, La504, La506 (MsmEFKAB, ABC transporter), La505 (BfrA, β-fructosidase), and La507 (GtfA, sucrose phosphorylase). Induction levels varied between 15.1 and 40.6 when compared with monosaccharides and disaccharides and between 5.5 and 8.9 when compared with raffinose. These results suggest that FOS is hydrolyzed into fructose and sucrose by the fructosidase. Sucrose is likely subsequently hydrolyzed into fructose and glucose-1-phosphate by the sucrose phosphorylase. In addition to the FOS operon, FOS also induced the fructose operon, the sucrose PTS transporter, the trehalose operon, and an ABC transporter (La1679–La1680).

In the presence of raffinose, the six genes of the raffinose operon were specifically induced (Figs. 1–3 and 7). The raffinose locus consists of La1442, La1441, La1440, La1439 (MsmEFKAB, ABC transporter), La1438 (MelA, α-galactosidase), and La1437 (GtfAβ, sucrose phosphorylase). Induction levels varied between 15.1 and 45.6 compared with all other conditions. Additionally, La1433–La1434 (dihydroxyacetone kinase) and La1436 (glycerol uptake facilitator) were induced between 1.9- and 24.7-fold compared with other conditions.

In the presence of lactose and galactose, 10 genes distributed in two loci were differentially expressed, namely La1463 [LacS, permease of the galactoside-pentose hexuronide (GPH) translocator family], La1462 (LacZ, β-galactosidase), La1461 (con- served hypothetical protein), La1460 (surface protein), La1459 (GalK, galactokinase), La1458 (GalT, galactose-1-phosphate uridylyl transferase), La1457 (GalM, galactose epimerase), La1467–La1468 (LacLM, β-galactosidase large and small subunits), and La1469 (GalE, UDP-glucose epimerase). LacS is similar to GPH permeases previously identified in lactic acid bacteria (18–20). Although LacS contains a EIIA at the carboxy terminus, it is not a PTS transporter. In addition, LacS includes a His at position 560, which might be involved in interaction with HPr, as shown in Streptococcus salivarius (19). In the presence of lactose and galactose, galKTM was induced between 3.7- and 17.6-fold; lacSZ was induced between 2.8- and 17.6-fold; lacL and galE were induced between 2.7 and 29.5 when compared with other carbohydrates not containing galactose, i.e., glucose, fructose, sucrose, trehalose, and FOS. These results suggest that lactose is transported into the cell via the LacS permease. Inside the cell, lactose is hydrolyzed into glucose and galactose by LacZ. Galactose is then phosphorylated by GalK into galactose-1-phosphate and further transformed into UDP-galactose by GalT. UDP-galactose is subsequently glucosylated to UDP-glucose by GalE. UDP-glucose is likely turned into glucose-1-phosphate by La1719 (GalU), which encodes a UDP-glucose phosphorylase that is consistently highly expressed. Finally, the phosphoglucomutase (La687), also highly expressed, likely acts on glucose-1-phosphate to yield glucose-6-phosphate.

Three of the genes of the putative trehalose locus were also differentially expressed (Figs. 1, 3, and 7). The trehalose locus consists of La1012 (encoding the TreB trehalose PTS trans-
transporter resulted in a loss of the ability to grow on trehalose (21). Similarly, differential expression of the fos operon is consistent with previous work in L. acidophilus indicating that those genes are involved in uptake and catabolism of FOS, induced in the presence of FOS, and repressed in the presence of glucose (9). Additionally, induction of the raffinose msm locus is consistent with previous work in Streptococcus mutans (22) and Streptococcus pneumoniae (23).

A number of lactic acid bacteria take up glucose via a PTS transporter. The EIIMan PTS transporter has the ability to import mannose and glucose (24). The L. acidophilus mannose PTS system is similar to that of Streptococcus thermophilus (24). Specifically, the EIIMan is composed of three proteins, IIABMan, IICMan, and IIDMan, encoded by Laa452 (manL), Laa455 (manM), and Laa456 (manN), respectively (Fig. 4). Whereas most of the carbohydrates examined here specifically induced genes involved in their own transport and hydrolysis, glucose did not. Analysis of the mannose PTS revealed that the genes encoding the EIIABCDMan were consistently highly expressed, regardless of the carbohydrate source (Fig. 1), which is consistent with the important regulatory roles in Gram-positive bacteria established for EIIMan (24, 25). Similarly, EIIABCMan seemed consistently highly expressed (Fig. 1). These expression profiles suggest that glucose and fructose are preferred carbohydrates and that L. acidophilus also is designed for efficient utilization of different carbohydrate sources, as was suggested previously for Lactobacillus plantarum (14).

The genes differentially expressed in the presence of galactose and lactose included a permease (LacS) and the enzyme machinery of the Leloir pathway. Members of the LacS subfamily of GPH translocators have been described in a variety of lactic acid bacteria, including Leuconostoc lactis (26), S. thermophilus (27), S. salivarius (19), Lactobacillus delbrueckii (28), and Lactobacillus helveticus (20). LacS has been reported to have the ability to import both galactose and lactose in select organisms (26, 27). Although the combination of a LacS lactose permease with two β-galactosidase subunits (LacL and LacM) has been described in L. plantarum (14), L. helveticus (20), and Leuconostoc lactis (26), it has never been reported in L. acidophilus. Although constitutive expression of lacS and lacLM has been reported (20, 26), our current results indicate specific induction of the genes involved in uptake and catabolism of both galactose and lactose. Operon organization for galactoside utilization is variable and unstable among Gram-positive bacteria (20, 28–32). Interestingly, even among closely related Lactobacillus species, namely L. johnsonii, L. gasseri, and L. acidophilus, the lactose–galactose locus is not well conserved (Fig. 10, which is published as supporting information on the PNAS web site) (15). It was also recently shown that, within the L. helveticus species, the lac operon is not well conserved (20). Perhaps the presence of mobile elements in the vicinity of those genes is responsible for the instability of this locus (16, 20).

Although it was previously suggested that PTS is the primary sugar transport system of Gram-positive bacteria (12, 25), current microarray data indicate that ABC transport systems also are important. Although PTS transporters are involved in uptake of monosaccharides and disaccharides, those carbohydrates are digested in the upper GIT. In contrast, oligosaccharides reach the lower intestine whereby commensals are likely to compete for more complex and scarce nutrients. Perhaps under such conditions ABC transporters are even more crucial than the PTS, given their apparent roles in the transport of oligosaccharides like FOS and raffinose. In this regard, the ability to use nutrients that are nondigestible by the host has been associated with competitiveness and persistence of beneficial intestinal flora in the colon (13).

Transcription profiles of genes differentially expressed in the conditions tested indicated that most carbohydrate uptake systems and their respective sugar hydrodases were specifically induced by their substrate. Moreover, genes within those inducible loci were down-regulated in the presence of glucose, and cre sequences (33) were identified in their promoter–operator regions (Fig. 11, which is published as supporting information on the PNAS web site). Together, these results indicate regulation of carbohydrate uptake and metabolism at the transcription level and implicate the involvement of a global regulatory system compatible with carbon catabolite repression (CCR), which controls transcription of proteins involved in the transport and catabolism of carbohydrates. CCR is a mechanism widely distributed amongst Gram-positive bacteria, including lactobacilli (34); mediated in cis by catabolite responsive elements (33, 35); and in trans by repressors of the LacI family, which is responsible for transcriptional repression of genes encoding unnecessary saccharolytic components in the presence of preferred substrates (25, 35–37). This regulatory mechanism allows cells to coordinate carbohydrate utilization and focus primarily on preferred energy sources. CCR is based on several key proteins, namely HPr (La639, ptsH), EI (La640, ptsI), CcpA (La431, ccpA), etc.
and HPk/P (La676, ptsK), all of which are encoded in \textit{L. acidophilus} (16).

The PTS is characterized by a phosphate transfer cascade involving PEP, EI, HPk, and EIABC, whereby a phosphate is ultimately transferred to the carbohydrate substrate (25, 38). HPk, an important component of CCR, is regulated via phosphorylation by EI and HPk/P (39). Although the phosphorylation cascade suggests regulation at the protein level, several studies report transcriptional modulation of ccpA and ptsH. In \textit{S. thermophilus}, CcpA production is induced by glucose (27). In several bacteria, the carbohydrate source modulates ptsH transcription levels (40). In contrast, expression levels of ccpA, ptsH, ptsI, and ptsK did not vary widely in the presence of different carbohydrates in \textit{L. acidophilus} (Fig. 1). These results are consistent with regulation via phosphorylation at the protein level. Similar results have been reported for ccpA expression levels in \textit{Lactobacillus pentosus} (34) and for ptsH transcription in \textit{S. thermophilus} (24).

In summary, a variety of carbohydrate uptake systems were identified and characterized, including PTS, ABC, and GHP transporters. The uptake and catabolic machinery is highly regulated at the transcriptional level, suggesting that the \textit{L. acidophilus} transcriptome is flexible, dynamic, and designed for efficient carbohydrate utilization. Differential gene expression indicated the presence of a global CCR regulatory network, where key genes were consistently highly expressed, suggesting regulation at the protein level rather than the transcriptional level. Collectively, \textit{L. acidophilus} appears able to adapt its metabolic machinery to fluctuating carbohydrate sources available in the environment. In particular, ABC transporters of the MsmEFGK family involved in uptake of FOS and raffinose likely play an important role in the ability of \textit{L. acidophilus} to adapt its metabolic machinery to fluctuating carbohydrate sources available in the environment. In particular, ABC transporters of the MsmEFGK family involved in uptake of FOS and raffinose likely play an important role in the ability of \textit{L. acidophilus} to adapt its metabolic machinery to fluctuating carbohydrate sources available in the environment.

### Materials and Methods

#### Bacterial Strains and Media.

The strain used in this study is \textit{L. acidophilus} NCFM (NCK56) (16). Cultures were aerobically propagated at 37°C in a semisynthetic medium described in ref. 9. The carbohydrates added were glucose (dextrose) (Sigma), fructose (Sigma), sucrose (Sigma), FOS (raftilose P95; Orafti, Tienen, Belgium), raffinose (Sigma), lactose (Fisher), galactose (Sigma), or trehalose (Sigma). Cells were harvested and RNA was isolated as described in ref. 42.

##### RNA Isolation.

Total RNA was isolated using TRIzol (GIBCO/BRL) by following the manufacturer’s instructions. \textit{L. acidophilus} cells were inoculated into semisynthetic medium supplemented with 1% (wt/vol) select sugars and propagated to mid-log phase (OD_{660} \approx 0.6). Cells were harvested and RNA was isolated as described in ref. 42.

##### Microarray Fabrication.

A whole-genome cDNA microarray was used for global gene expression analysis (42). The microarray contained triplicate spots of 1,889 cDNA PCR products amplified from genomic DNA, as described in ref. 42. PCR amplicons were spotted on GAPS II aminosilane-coated glass slides (Corning) by using an Affymetrix 417 Arrayer, and slides were processed as described in refs. 42 and 43. Details for the microarray platform are available at the National Center for Biotechnology Information Gene Expression Omnibus database (GEO platform GPL1401).

##### cDNA Preparation and Microarray Hybridization.

For each hybridization, two total RNA samples (25 µg each) were amino-allyl-labeled by reverse transcription using random hexamers (Invitrogen) as primers in the presence of amino-allyl dUTP (Sigma), by a SuperScript II reverse transcriptase (Invitrogen), as described in refs. 41 and 42. Labeled cDNA samples were subsequently coupled with either Cy3 or Cy5 N-hydroxysuccinimidy-dyes (Amersham Pharmacia Biosciences), and purified by using a PCR purification kit (Qiagen). The resulting samples were hybridized onto microarray slides and further processed as described in ref. 42 and according to The Institute for Genomic Research protocol (43). Hybridizations were performed according to a single round-robin design, so that all possible direct pair-wise comparisons were conducted. With eight different sugars, a total of 28 hybridizations were performed (see Fig. 12, which is published as supporting information on the PNAS web site). Each treatment was labeled seven times, and every other treatment was labeled with either Cy3 or Cy5 four and three times, respectively. Microarray data details are available at the National Center for Biotechnology Information (GEO series GSE2577).

##### Microarray Data Collection and Analysis.

Microarray images were acquired with a ScanArray 4000 microarray scanner (Packard Bioscience, which is now PerkinElmer). Signal fluorescence, including spot and background intensities was subsequently quantified and assigned to genomic ORFs using QUANTARRAY 3.6 (PerkinElmer). Raw data were imported into SAS (SAS Institute), compiled, background-corrected, log2-transformed, and subjected to a mixed model of analysis of variance (SAS proc mixed) with two sequential linear models (44) outlined in the Supporting Materials and Methods, which is published as supporting information on the PNAS web site. ANOVA mixed models have proven successful at analyzing microarray data (41, 44–50). The array and spot effects were treated as random effects, whereas dye and treatment effect were considered fixed effects. The resulting difference between least-square estimates for two different treatments is analogous to a log2-transformed ratio of gene expression between those two treatments. Differences were calculated between all pairs of treatments for each gene, and a measure of statistical significance was obtained from a t test using these differences and their associated standard errors. A Bonferroni correction was applied to account for bias due to multiple tests by dividing the desired level of significance (\( \alpha = 0.05 \)) by the total number of comparisons performed (54,781), as used in refs. 49 and 50. Thus, the corrected \( \alpha \)-positive rate was \( \alpha = 9.12 \times 10^{-7} \) corresponding to a \( \log_{10}(P\ \text{value}) = 6.04 \) (10^{-6.04}). All \( P \) values that fell below \( \alpha \) were considered statistically significant. Volcano plots of log2-transformed fold changes (induction ratios) versus log2-transformed \( P \) values (statistical significance) and three-way plots (contour plots) of individual treatment effects were used to visualize contrasts between treatments and statistical significance of the results. Global transcriptional patterns were visualized by using Ward’s method of hierarchical clustering in JMP 5.0 (SAS Institute) with least-squares mean estimates and their standardized counterparts as input.

##### Real-Time QRT-PCR.

Five genes differentially expressed in microarray experiments were selected for real-time QRT-PCR experiments to validate the induction levels measured. These genes were selected for their broad expression range (least-squares mean between –1.52 and +3.87) and their induction levels between sugars (fold induction up to 35). Also, the selected genes were correlated functionally with carbohydrate utilization. The selected genes were \( \beta \)-fructosidase (La105), trehalose PTS (La1012), glyceral uptake facilitator (La1436), \( \beta \)-galactosidase (La1467), and ABC transporter (La1679). Experiments were conducted with a QRT-PCR thermalycler (I-cycler; Bio-Rad) in combination with a QuantiTect SYBR Green PCR kit (Qiagen). Six carbohydrates samples were included, namely glucose, fructose, sucrose, FOS, lactose, and galactose. Each set of samples was analyzed in triplicate. The RNA samples used in QRT-PCR experiments were identical to those used in microarray experiments.
We thank E. Durmaz, Dr. O. McAuliffe, Dr. G. Gibson, and Dr. D. Nielsen for insightful discussions and technical help; R. Brierley and Dr. B. Sosinsky for assistance with equipment for microarray work; Dr. Eric Altermann for help with digital images; and Dr. J. Bruno-Barea and Dr. H. Hassan for help with QRT-PCR. This work was sponsored by Dairy Management Inc., the Southeast Dairy Foods Research Center, and Danisco USA. R.B. and T.D. are recipients of Integrative Graduate Education and Research Traineeship fellowships from the National Science Foundation.

Supporting Figure 6
Supporting Figure 7
Supporting Figure 8
Supporting Figure 9
Supporting Figure 10
L. gasseri

L. johnsonii

L. acidophilus

lacS lactose-proton symporter
lacLM beta-galactosidase
galE galactose epimerase
galK galactokinase
galT galactose-1P uridyl transferase

Supporting Figure 11
<table>
<thead>
<tr>
<th>Probed</th>
<th>cre1/2</th>
<th>Overmatch</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>La400</td>
<td>cre1</td>
<td>TGataaaCGgttgaCA</td>
<td>-72 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>AGataaCGctttaCA</td>
<td>-17 bp</td>
</tr>
<tr>
<td>La401</td>
<td>cre1</td>
<td>TGaataCGtttatCA</td>
<td>-48 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>TAaaagCGttttaCA</td>
<td>-17 bp</td>
</tr>
<tr>
<td>La452</td>
<td>cre1</td>
<td>TAaaagCGgttttCA</td>
<td>-27 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>TGaaagCGatataTA</td>
<td>-172 bp</td>
</tr>
<tr>
<td></td>
<td>cre3</td>
<td>TAaaagCGttttcCA</td>
<td>-140 bp</td>
</tr>
<tr>
<td></td>
<td>cre4</td>
<td>TTcaaaCGttttCA</td>
<td>-14 bp</td>
</tr>
<tr>
<td>La680</td>
<td>cre1</td>
<td>AGtaagCGctttCC</td>
<td>-40 bp</td>
</tr>
<tr>
<td>La1012</td>
<td>cre1</td>
<td>TGtgatCGctttCA</td>
<td>-82 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>TGaaaCGcttttAT</td>
<td>-15 bp</td>
</tr>
<tr>
<td>La1013</td>
<td>cre1</td>
<td>ATaaagCGtttttCA</td>
<td>-155 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>TGaaagCGatcCA</td>
<td>-88 bp</td>
</tr>
<tr>
<td>La1442</td>
<td>cre1</td>
<td>AGaataCGcaattAA</td>
<td>-69 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>TGaaagCGcttaAA</td>
<td>-38 bp</td>
</tr>
<tr>
<td>La1459</td>
<td>cre1</td>
<td>TGaaaCGatgtaCA</td>
<td>-27 bp</td>
</tr>
<tr>
<td>La1460</td>
<td>cre1</td>
<td>GAtggaCGaataTA</td>
<td>-22 bp</td>
</tr>
<tr>
<td>La1461</td>
<td>cre1</td>
<td>AGgtatCGtcatCT</td>
<td>-103 bp</td>
</tr>
<tr>
<td>La1463</td>
<td>cre1</td>
<td>AAaattCGctttCT</td>
<td>-36 bp</td>
</tr>
<tr>
<td>La1465</td>
<td>cre1</td>
<td>AAaattCGtaagTA</td>
<td>-27 bp</td>
</tr>
<tr>
<td>La1467</td>
<td>cre1</td>
<td>TAaaagCGttttCA</td>
<td>-32 bp</td>
</tr>
<tr>
<td>La1469</td>
<td>cre1</td>
<td>TGtaatCGatttCA</td>
<td>-21 bp</td>
</tr>
<tr>
<td>La1684</td>
<td>cre1</td>
<td>AGttttCGgacaAC</td>
<td>-61 bp</td>
</tr>
<tr>
<td></td>
<td>cre2</td>
<td>AGaaatCGctttCA</td>
<td>-25 bp</td>
</tr>
</tbody>
</table>

**Supporting Figure 12**
Supporting Figure 5