2014

From prediction to function using evolutionary genomics: Human-specific ecotypes of *Lactobacillus reuteri* have diverse probiotic functions

Jennifer K. Spinler
Texas Children's Hospital, spinler@bcm.edu

Amrita Sontakke
Baylor College of Medicine

Emily B. Hollister
Baylor College of Medicine

Susan F. Venable
Baylor College of Medicine

Phaik Lyn Oh
University of Nebraska, Lincoln

See next page for additional authors

Authors
Jennifer K. Spinler, Amrita Sontakke, Emily B. Hollister, Susan F. Venable, Phaik Lyn Oh, Miriam A. Balderas, Delphine M.A. Saulnier, Toni-Ann Mistretta, Sridevi Devaraj, Jens Walter, James Versalovic, and Sarah K. Highlander

Genome Biology and Evolution Advance Access published June 19, 2014 doi:10.1093/gbe/evu137

Genome Biology and Evolution

SMBE

From prediction to function using evolutionary genomics: Human-specific ecotypes of *Lactobacillus reuteri* have diverse probiotic functions

Jennifer K. Spinler[1,2,*], Amrita Sontakke[1,2], Emily B. Hollister[1,2], Susan F. Venable[1,2], Phaik Lyn Oh[3], Miriam A. Balderas[4], Delphine M. A. Saulnier[1,2,†], Toni-Ann Mistretta[1,2], Sridevi Devaraj[1,2], Jens Walter[3,‡], James Versalovic[1,2,4], Sarah K. Highlander[4,5,††]

[1]Texas Children's Microbiome Center, Department of Pathology, Texas Children's Hospital, 1102 Bates Ave., Houston, Texas, USA

[2]Department of Pathology & Immunology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas, USA

[3]Department of Food Science and Technology, University of Nebraska, Lincoln, Nebraska, USA

[4]Molecular Virology & Microbiology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas, USA

[5]Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas, USA

[†]Current address: Department of Gastrointestinal Microbiology, German Institute of Human Nutrition, Nuthetal, Germany

[‡]Current address: Departments of Agricultural, Food, & Nutritional Science and Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

[††]Current address: Genomic Medicine, J. Craig Venter Institute, La Jolla, California, USA

*Author for Correspondence: Jennifer K. Spinler, Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, USA, phone: 832-824-8251, fax: 832-825-7211, spinler@bcm.edu

Abstract

The vertebrate gut symbiont *Lactobacillus reuteri* has diversified into separate clades reflecting host origin. Strains show evidence of host adaptation, but how host-microbe co-evolution influences microbial-derived effects on hosts is poorly understood. Emphasizing human-derived strains of *L. reuteri*, we combined comparative genomic analyses with functional assays to examine variations in host interaction among genetically distinct ecotypes. Within clade II or VI, the genomes of human-derived *L. reuteri* strains are highly conserved in gene content and at the nucleotide level. Nevertheless, they share only 70-90% of total gene content, indicating differences in functional capacity. Human-associated lineages are distinguished by genes related to bacteriophages, vitamin biosynthesis, antimicrobial production, and immunomodulation. Differential production of reuterin, histamine, and folate by 23 clade II and VI strains was demonstrated. These strains also differed with respect to their ability to modulate human cytokine production (TNF, MCP-1, IL-1β, IL-5, IL-7, IL-12, and IL-13) by myeloid cells. Microarray analysis of representative clade II and clade VI strains revealed global regulation of genes within the reuterin, vitamin B_{12}, folate, and arginine catabolism gene clusters by the AraC family transcriptional regulator, PocR. Thus, human-derived *L. reuteri* clade II and VI strains are genetically distinct and their differences affect their functional repertoires and probiotic features. These findings highlight the biological impact of microbe:host co-evolution and illustrate the functional significance of subspecies differences in the human microbiome. Consideration of host origin and functional differences at the subspecies level may have major impacts on probiotic strain selection and considerations of microbial ecology in mammalian species.

Keywords

Host-based evolution, reuterin, PocR transcriptional regulation, immunostimulatory, anti-inflammatory, histamine

Introduction

Lactobacillus reuteri are natural residents of mammalian and avian gastrointestinal (GI) tracts as well as the human urogenital tract and breast milk. *L. reuteri* exhibits strain-specific beneficial properties relevant to human health, making it a model organism for studying host:symbiont interactions as well as microbe:host co-evolution. Comparative genomic studies of *Lactobacillus* species performed as part of the Human Microbiome Project highlighted significant heterogeneity within and between species and significant interspecies diversity among strains of *L. reuteri* (Nelson, et al. 2010). Previous phylogenetic studies of *L. reuteri* used amplified-fragment length polymorphism (AFLP) and multi-locus sequence analysis (MLSA) to study more than 100 strains, and these identified host origin as one basis for intraspecies diversity (Oh, et al. 2010), with lineage-specific genomic differences reflecting the niche characteristics in the GI tract of respective hosts (Frese, et al. 2011). Experiments in gnotobiotic mice supported host adaptation of *L. reuteri* strains, as only rodent strains colonized mice efficiently (Frese, et al. 2011). Human-derived *L. reuteri* strains belong to two distinct MLSA clades, designated clade II and VI. Clade II is remarkably specific to humans and it clusters separately from all other clades, while human strains in clade VI are closely related to isolates from chickens (Oh, et al. 2010). Overall, the findings indicated that *L. reuteri* is a host-specific symbiont, and separate lineages within the species suggest that host restriction was maintained over evolutionary time spans, allowing host-driven diversification. The nature of evolutionary processes and their influence on specific host-microbe interactions and health status deserve exploration.

The mutualistic relationship between *L. reuteri* and humans has been validated by numerous studies that have documented the ability of *L. reuteri* to elicit probiotic effects in

different disorders and disease models. *L. reuteri* strains are acid and bile tolerant, produce many of the essential B complex vitamins, notably folate (B_9) (Santos, et al. 2008b), and cobalamin (B_{12}) (Morita, et al. 2008; Santos, et al. 2008a), but also potentially thiamin (B_1) (Saulnier, et al. 2011) and riboflavin (B_2) (Capozzi, et al. 2012). Many strains synthesize the antimicrobial compound, reuterin (Axelsson, et al. 1989) and the organism synthesizes and secretes the anti-inflammatory biogenic amine, histamine (Thomas, et al. 2012). Previous studies examining these probiotic properties and their regulation have indicated that not all strains of *L. reuteri* have equal capacity to express these factors or phenotypes (Jones and Versalovic 2009; Lin, et al. 2008; Spinler, et al. 2008).

In this study, we compared the genomes of ten sequenced *L. reuteri* strains, derived from three host origins, to gain insight into the distinguishing features of human-derived *L. reuteri* strains, and to better understand the beneficial characteristics specific to human clades II and VI. We have complemented these genomic comparisons with functional data that classify the human-derived strains by gene expression, metabolic function, probiotic features, and effects on cytokine production by human immune cells. These data demonstrate clear differences between the two sub-populations of human-derived probiotic *L. reuteri* strains and emphasize the importance of performing studies on the role of the microbiome in an evolutionary context.

Methods

Bacterial strains, human cell lines and culture conditions

The human-derived *L. reuteri* strains and genome statistics are listed in Table 1. Additional strains used for functional assays are listed in Table 2. Bacterial strains were routinely cultured in deMan, Rogosa, Sharpe medium (MRS; Difco, Franklin Lakes, NY) at 37°C in an anaerobic workstation (MACS MG-500, Microbiology International, Frederick, MD) supplied with a mixture of 10% CO_2, 10% H_2, and 80% N_2 for 16-18 h. Strains ATCC 6475::*pocR* and DSM::*pocR* were cultured in the presence of erythromycin (10 μg/mL). Specific culture conditions for individual experiments are detailed throughout. The effect of bacterial supernatants on cytokine production was performed with THP-1 cells (human monocytoid cell line, ATCC TIB-202, ATCC, Manassas, VA) maintained in RPMI (ATCC, Manassas, VA) and 10% heat-inactivated fetal bovine serum (Invitrogen, Carlsbad, CA) at 37°C with 5% CO_2.

Multi-Locus Sequence Analysis

MLSA was performed on 119 *L. reuteri* strains. This analysis was completed using standard techniques as previously described (Oh, et al. 2010). In brief, seven housekeeping genes (D-alanine-D-alanine ligase (*ddl*), D-alanine-D-alanyl carrier protein ligase (*dltA*), DNA gyrase B subunit (*gyrB*), leucyl-tRNA synthetase (*leuS*), phosphoktolase (*pkt*), recombinase (*recA*) and RNA polymerase alpha subunit (*rpoA*)) were PCR amplified from all 119 *L. reuteri* strains, sequenced and included in the MLSA. Clonal Frame software, http://bacteria.stats.ox.ac.uk (Didelot and Falush 2007), which uses a coalescent-based Bayesian method to infer strain relationships, was applied as described previously in Supplementary materials of Oh, *et al.* 2010 (Oh, et al. 2010).

Lactobacillus reuteri genomes and pangenome analysis

Publicly available genome sequences for the ten *L. reuteri* strains (3 complete and 7 draft sequences) used in this study are described in Table 1. Three of the genomes (mlc3, lpuph and ATCC 53608) lacked annotations so the nucleotide sequences were submitted to the Integrated Microbial Genomes (IMG/ER) system (Markowitz, et al. 2010) for gene calling and feature prediction. To maintain consistency, all annotation-based analyses were performed using the IMG annotations for the ten genomes. For draft genomes, we used blastz within Advanced PipMaker (Schwartz, et al. 2000), to obtain ordered and oriented scaffolds or contigs for comparative analyses. The finished genome of JCM 1112 was used as the reference for clade II, ATCC 55730 for clade VI, 100-23 for mlc3 and lpuph, and I5007 (NC_021494) for ATCC 53608. Once scaffolds or contigs were ordered and oriented, all were arranged so the genome sequence began with the start codon for the *dnaA* gene. These were the nucleotide sequences used for comparative analysis and alignments.

Circular blastn-based representations of the ten *L. reuteri* genomes were constructed using BLAST Ring Image Generator (BRIG) (Alikhan, et al. 2011). For the incomplete genomes, ATCC 4659, CF48-3A, 100-23, mlc3, lpuph, and ATCC 53608, the ordered and oriented scaffolds were concatenated to create a single scaffold. These and the complete genome nucleotide sequences were used for input in the order shown in Figure 2 and Supplementary Figure S1, Supplementary Material online. The program passes the genomes to blastn (release 2.2.28+) using default parameters with the first genome as the reference, and returns a graphical image of concentric rings where the density of color at a particular location within a ring indicates nucleotide homology with respect to the reference.

A pangenome analysis was used to describe the *L. reuteri* species as a whole and to gain insight into the host-driven evolution of the species. Genes in all 10 genomes were binned into

operational gene units using the online implementation of CD-HIT-EST (Huang, et al. 2010; Li and Godzik 2006) with default parameters (>90% nucleotide sequence identity spanning the length of the shorter gene in each pair-wise comparison) as described by Hansen, *et al*. (Hansen, et al. 2011). The size of the core genome was estimated utilizing the distribution of shared OGUs among the ten genomes. The function of each OGU was inferred from the annotations obtained from IMG. Higher level functions of were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, et al. 2012) and Cluster of Orthologous Genes (COG (Tatusov, et al. 2003)) assignments obtained from the IMG annotation files.

Average nucleotide identity and gene conservation between genome pairs

We compared the gene conservation at the nucleotide and amino acid levels across the *L. reuteri* genomes to assess diversity within the species. Average nucleotide identity and gene conservation between the 10 *L. reuteri* genomes was calculated using the BLAST release 2.2.28+ (Altschul, et al. 1990), essentially as described previously (Konstantinidis and Tiedje 2005). For ANI, blastn was implemented using default parameters except with the drop-off value of 150 for gapped alignments (-xdrop_gap_final 150) and without query filtering/masking (-dust no). Average nucleotide identity was evaluated for all conserved genes having >60% sequence identity over >70% of the length of the query sequence. Gene content was computed using tblastn, where predicted protein-coding sequences from one *L. reuteri* genome (designated the query genome) were searched against the genomic sequence of a second *L. reuteri* genome (designated the reference genome), utilizing default settings plus the option –db_gencode 11. The tblastn output was filtered by top bit score and e-value per query, then the query hit was used to score conservation, defined as a CDS having >60% sequence identity over >70% of the length of the reference CDS.

Inactivation of *pocR* by insertional mutagenesis and comparative transcriptional analysis of ATCC 6475::*pocR* to DSM 17938::*pocR*

The *pocR* gene was inactivated in DSM 17938 as previously described for strain ATCC 6475 (Santos, et al. 2011). Briefly, the internal fragment of the *pocR* gene was amplified by PCR using primers RB1883F2 5'BHI and RB1883R2 3'ERI (Supplementary Table S1, Supplementary Material online), and inserted into plasmid pORI28 (Russell and Klaenhammer 2001) by directional cloning using standard techniques (Sambrook and Russell 2001) to generate pORIpocR. Temperature-sensitive, site-specific integration of the non-replicating plasmid pORIpocR was carried out as described earlier (Russell and Klaenhammer 2001; Santos, et al. 2011; Walter, et al. 2005). The resulting insertion mutant was designated DSM 17938::*pocR*.

Comparisons of the transcriptome of ATCC 6475::*pocR* and its parent strain were performed using two-color microarrays and published previously [GEO GSE22926 (Santos, et al. 2011)]. The transcriptome of the *pocR* mutant strain, DSM 17938::*pocR*, and its parent strain were compared using two-color microarrays as was that of ATCC 6475::*pocR* (Santos, et al. 2011). In brief, oligonucleotides were designed and synthesized from a draft genome sequence of *L. reuteri* ATCC 55730 (Saulnier, et al. 2011). For expression analysis, stationary phase mRNA was isolated from three biological replicates of the *pocR* mutant and parent strain cultured anaerobically in a semi-defined medium, LDMIII (Jones and Versalovic 2009), as described previously (Saulnier, et al. 2011). Synthesis, labeling, hybridization, and dye-swap comparisons of cDNA were performed also as previously described (Saulnier, et al. 2011). Microarray platform information and data for the DSM 17938 strains are deposited at the NCBI Gene Expression Omnibus (GEO GSE54324; http://www.ncbi.nlm.nih.gov/geo/). All image analysis, normalization, and statistical analysis for the DSM 17938 data was performed as before

(Santos, et al. 2011). Microarray data from the previously published ATCC 6475 data set was compared to the DSM 17938 data set. Only probes common to both arrays were compared.

Production and quantification of reuterin produced from *L. reuteri* strains

Isolation of reuterin from *L. reuteri* supernatants was done as previously described (Spinler, et al. 2008). Briefly, cell pellets were collected at various time points from anaerobic cultures of *L. reuteri* and were washed in sodium phosphate buffer, resuspended to ~1.5 x 10^{10} cells ml^{-1} in glycerol solution then incubated anaerobically at 37°C for 1 hr. The reuterin-containing solution was collected, filter sterilized and stored at 4°C until further analysis. Reuterin-containing solutions were analyzed colorimetrically by absorbance spectroscopy as done previously (Spinler, et al. 2008). A standard curve was generated using 0-6 mM of HPLC quantified reuterin (Spinler, et al. 2008) produced by *L. reuteri* DSM 17938.

Preparation of *L. reuteri* supernatants

L. reuteri supernatants were prepared and analyzed for folate and histamine concentrations, and for cytokine bioassays from human monocytes. Strains were cultured anaerobically for 24 h at 37°C in MRS, and then inoculated into LDMIII at an OD_{600} of 0.1. LDMIII cultures were incubated anaerobically at 37°C for 24 h, then supernatants were collected, filter sterilized using polyvinylidene fluoride membrane filters (0.22 μm pore size, Millipore, Bedford, MA) and stored at -20°C before further processing.

Folate detection from *L. reuteri* supernatants

Supernatants of *L. reuteri* strains were prepared as outlined above and tested for the production of folate by electrochemiluminescence (Wilson, et al. 2005). Folate levels in the filter-sterilized supernatants were determined by a competitive immunoassay using direct

chemiluminescence. Samples were pretreated to release folate conjugates from any endogenous binding proteins in the sample then the free folate was assayed by competition for binding to an acridinium-folate-bound biotin-folate binding protein. Binding was detected using avidin conjugated paramagnetic particles in the solid phase. The inter- and intra-assay coefficients of variation of the assay were <10%.

Histamine production by *L. reuteri* strains and PCR amplification of *hdcP, hdcA, hdcB* and *hisRS2* genes

Supernatants of *L. reuteri* strains were tested for the production of the biogenic amine, histamine. Filter sterilized supernatants were serially diluted in phosphate buffered saline and tested using a histamine ELISA kit (cat# 409010, Neogen, Lansing, MI) as per the manufacturer's instructions. Concentrations of histamine were normalized against culture OD_{600} values. PCR amplification was used to confirm the presence or absence of key genes involved in histamine production by *L. reuteri* strains. The sequence of *hisRS2* was unique from *hisRS1* as to permit discrimination. Genomic DNA was isolated from *L. reuteri* using the MO BIO Ultra Clean Microbial gDNA isolation kit (Carlsbad, CA) as per manufacturer's instructions. Target genes were amplified by PCR using standard techniques (Sambrook and Russell 2001), using primers listed in Supplementary Table S1, Supplementary Material online.

Bioassay of cytokine production from human monocytes

L. reuteri cell-free supernatants were prepared as outlined above and tested for effects on cytokine production by THP-1 cells as previously described (Pena, et al. 2004). Supernatants were vacuum-dried, resuspended in RPMI medium, and normalized by volume to OD_{600} = 1.5. THP-1 cells ($5x10^4$ cells) were stimulated to produce TNF by the addition of 100 ng mL^{-1} Pam_3Cys-SKKKK x 3 HCl (EMC Microcollections, Tüebingen, Germany) as previously

described (Lin, et al. 2008; Pena, et al. 2004). *L. reuteri* supernatants were added to the activated THP-1 cells (5% v/v) in microtiter plates, which were then incubated at 37°C in 5% CO_2 for 3.5 h. THP-1 cells were pelleted (3000 x *g*, 5 min, 4°C), and THP-1 supernatants were assayed using a Human Cytokine/Chemokine-Premixed 14-plex kit (Millipore, Billerica, MA) in a Luminex 100 system (Luminex Corporation, Austin, TX). The 14 analytes tested were: tumor necrosis factor (TNF), granulocyte macrophage colony-stimulating factor (GM-CSF), interferon (INF)-γ, interleukin (IL)-1β, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12, IL-13, and monocyte chemoattractant protein (MCP)-1. Raw data were obtained with MasterPlex CT version 1.2.0.7 and analyzed with MasterPlex QT version 5.0.0.73 (Hitachi MiraiBio, San Francisco, CA).

Statistical analysis of microbiological data

Statistical analyses were performed using SPSS 19.0 (IBM SPSS Statisticsm; http://www-01.ibm.com/software/analytics/spss/). Data reported are means from a minimum of three biological replicates for reuterin, histamine, folate, and cytokine assays. Conditions were defined as negative control, positive control, clade II, or clade VI. A one way ANOVA analysis was run between conditions to generate *p*-values. A Tukey HSD (Honestly Significant Difference) algorithm was applied to the ANOVA result to reduce the Type I error rate (false positives). All significance values are based on $p < 0.05$. Folate data were analyzed using the Student's *t*-test with one-tailed distribution and considered statistically significant at a *p*-value $\leq$ 0.05, unless otherwise stated.

Results and Discussion

General features of *L. reuteri* genomes

As of June 2013, the genomes of ten *Lactobacillus reuteri* strains had been sequenced (3 complete genomes and 7 draft sequences), and their features are described in Table 1. Additional *L. reuteri* genomes of strains I5007 and TD1, were deposited in NCBI during the preparation of this manuscript, but were not included in this analysis. The genomes of the strains analyzed here are similar in size (1.9 – 2.3 Mb), G+C content (38.4 – 39.0%), and contain between 1,820 – 2,300 reported predicted protein-coding sequences. The *L. reuteri* strains are isolates from three host types (human, rodent, and pig) and were collected from four geographically distinct regions within Australasia, Europe, North America, and South America. *L. reuteri* is grouped into host-associated clades based on MLSA typing (Figure 1, Table 1), and the ten genomes were assigned to clades I-IV and VI by Oh, et al. (Oh, et al. 2010), with six genomes from human-derived strains belonging to clades II and VI. To visualize differences between the genomes, we generated the nucleotide alignments of the genomes shown in Figure 2 by blastn using the complete clade II genome (JCM 1112) as the reference and the program BLAST Ring Generator (BRIG) (Alikhan, et al. 2011). We also created a clade VI reference alignment, using the completed genome of ATCC 55730 as the reference, which is shown in Supplementary Figure S1, Supplementary Material online.

We conducted a pangenome analysis to better understand the distinct and independent evolution of the *L. reuteri* strains. All predicted protein coding genes present in these genomes (~20,600 genes) were binned into "operational gene units" (OGUs) with other genes having >90% identity using the program CD-HIT (Huang, et al. 2010). Based on this analysis, we predict that the *L. reuteri* pangenome encompasses approximately 3,700 genes. The overall conserved genome is ~1.2 Mb, which represents about 1,230 core OGUs. Strain-specific differences are defined by those genes, within the pangenome, which are not shared among

strains. Although these genes are referred to as "unique", this designation is context-dependent, and it is possible that they may be found in other bacterial species. We have identified 1,200 unique genes (OGUs) in *L. reuteri*, as well as ~1,320 dispensable genes, which are neither unique genes, nor members of the core genome. The predicted pangenome contains approximately 3,700 OGUs, and the number of new OGUs accumulating within the pangenome began to plateau after the addition of eight *L. reuteri* genomes (~16,500 coding sequences), indicating its approach to saturation. The rate and shape of the rarefaction curve (data not shown) predicts that the pangenome should close with approximately 24 *L. reuteri* genomes.

Examining ecological and evolutionary diversity within *L. reuteri*

To explore patterns of evolution and ecology within *L. reuteri*, pair-wise genomic comparisons of the average nucleotide identity (ANI) of each predicted coding sequence (CDS) were compared to the amino acid similarity (gene content) of each CDS (Figure 3). Higher percentage values of both ANI and gene content suggest genomes have evolved similarly and inhabit comparable niches, while lower percentage values indicate genomes that have evolved divergently and adapted to different environments (Konstantinidis, et al. 2006). The ten-genome pair-wise comparison revealed two distinct populations within the *L. reuteri* species and confirmed similar intraspecies diversity revealed earlier with a seven-genome comparison reported by Nelson, *et al.* (Nelson, et al. 2010). Despite having been isolated from disparate hosts, all clades shared ANI of 95% or greater. Utilizing a bacterial species definition of 70% DNA-DNA hybridization (Wayne 1987) or ≥95% ANI (Konstantinidis and Tiedje 2005), the intraspecies divergence observed among our strains does not dictate a new species designation, since all pair-wise comparisons have ≥95% ANI (Figure 3A). Of note, the two human-derived clades were as dissimilar to one another as they were to clades that contained rodent or porcine-

derived strains, suggesting that the two human clades may have evolved separately. Conversely, within-clade ANI values exceed 99% similarity and gene content levels of 92-100% suggesting that genomes within clades II and VI may have undergone clade-specific clonal diversification.

The population structure and the distinct genomic features of *L. reuteri* lineages indicate that these lineages may have been shaped by different evolutionary forces, and we hypothesized that metabolic functions and probiotic phenotypes would differ among the two human-derived *L. reuteri* ecotypes. Here we consider the definition of an ecotype to represent a cohesive group that possesses all the dynamic properties attributed to a species, but one that remains irreversibly separate and ecologically distinct (Cohan and Koeppel 2008). Our findings confirm extremely low sequence variation among strains within the same human ecotype. By MLSA analysis, human-derived *L. reuteri* clade VI strains cluster closely with strains of poultry origin, while clade II strains almost exclusively associate with other human-derived strains (Figure 1). Additionally, our pangenome analysis showed that the majority of the genes that distinguished strains in clades II and VI encode hypothetical proteins (data not shown). These observations are similar to those from *Bacillus anthracis* and *Salmonella enterica* pathovar Typhi genome projects, where strains were considered identical because of >99% ANI, but where further examination revealed the acquisition of a small number of genetic elements capable of delineating distinct ecotypes (Konstantinidis and Tiedje 2005). In this study, we made predictions of function based on genome content using genome sequences of six human-derived *L. reuteri* strains that represent clades II and VI (see Figure 2 and Supplementary Figure S1, Supplementary Material online). To validate the functional differences defining the two human-derived ecotypes, we expanded our study to analyze the clade-specificity of probiotic-associated

traits of 25 human-derived *L. reuteri* strains representing clades II (twelve strains) and clade VI (thirteen strains) (Table 2; Figure 1).

The presence of clade-specific mobile genetic elements support clonal evolution of human-derived *L. reuteri* strains

Mobile genetic elements (such as insertion elements, transposons, and genes associated with bacteriophage) contribute to host specialization of *L. reuteri* (Frese, et al. 2011), and accounted for an additional 126 OGUs specific to human-derived strains. Two complete prophages, unique to the clade II genomes are clearly identified in Figure 2. We assigned new open reading frames (ORFs) and updated the annotations (Supplementary Tables S2a-S2b, Supplementary Material online) as described in the Materials and Methods. ϕLreu1 is about 54 kb in length and is 100% identical among the clade II genomes. It is inserted between a gluconate transporter and the riboflavin synthase alpha and beta subunits, and its structural proteins are most similar to other temperate *Siphoviridae* proteins from *L. plantarum* genomes. The second prophage, ϕLreu2, is about 43.5 kb long, also is 100% conserved and lies between the 50S large subunit protein gene *rpsL* and the beta-galactosidase operon. It contains a potential self-splicing intron element at its terminus. The ϕLreu2 structural proteins suggest that it may also be a member of the *Siphoviridae*, as its capsid, portal and head-tail joining proteins are very similar to those of phage HK97 family phages. The clade VI strains also contain two complete prophages (ϕLreu3 and ϕLreu4) that were revealed on the blastn-based map using ATCC 55730 as the reference genome (Supplementary Figure S1, Supplementary Material online). The annotations for these, also likely *Siphoviridae,* are listed in Supplementary Tables S2c and S2d, Supplementary Material online. Many transposases are present in the genomes, though their types and distributions vary between clades II and VI. Most notably, all of the sequenced clade

II genomes carry 19 copies of the IS*200*/IS*605* family transposase gene while the clade VI genomes have only two. This IS*200*/IS*605* family of insertion elements is of interest because they are the smallest known, lack characteristic inverted repeats, and insert in a site-specific manner (Barabas, et al. 2008). The 100% identity and clade-specific delineation of bacteriophages and transposase genes in these genomes corroborate the notion that the human-derived clade II and VI genomes are distinct clonal groups, and further support the idea that human-derived *L. reuteri* strains have undergone a clade-specific selective sweep important for adapting to the human gastrointestinal tract (Walter, et al. 2011).

Distinct metabolic functions and probiotic phenotypes have evolved to define the two human-derived *L. reuteri* ecotypes.

Probiotic *L. reuteri* produce factors that are beneficial to the mammalian host, aid in host defense, and help to maintain a homeostatic relationship in the gut. Many strains synthesize the antimicrobial compound, reuterin (Axelsson, et al. 1989), and specific strains synthesize and secrete the anti-inflammatory biogenic amine, histamine (Thomas, et al. 2012); each of which will be discussed in detail later. Additionally, lactic acid bacteria create local acidic environments and synthesize essential B complex vitamins.

L. reuteri has an arginine catabolism mechanism that may function to promote local acidic environments and facilitate survival at low pH (Marquis, et al. 1987). An approximately 27 kb gene cluster, which is flanked by an IS*200* insertion element and maps between 484,793 and 511,393 in JCM 1112, carries core and dispensable genes that are observed in multiple clades but have a specific gene order and composition in clades II and VI (Figure 2); updated annotations for these are presented in Supplementary Table S3, Supplementary Material online. The region contains 31 ORFs and carries the *arc* genes responsible for producing enzymes that

catabolize arginine to citrulline and ornithine and ultimately to ammonia, thereby providing a means for *L. reuteri* to thrive in the presence of low pH in the lactic acid environment or during gut transit.

L. reuteri produce many essential B complex vitamins (Capozzi, et al. 2012; Morita, et al. 2008; Santos, et al. 2008a; Santos, et al. 2008b; Saulnier, et al. 2011). Folate (vitamin B_9), specifically, is formed from the precursor molecules, 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (DHPPP) and para-aminobenzoic acid (pABA). Most lactobacilli encode the enzymes required to generate folate from DHPPP and pABA, but not those required for pABA synthesis (Rossi, et al. 2011); thus, they are not predicted to perform *de novo* synthesis of folate (Supplementary Figure S2, Supplementary Material online). Unexpectedly, many of the clade VI *L. reuteri* strains tested do produce folate in the presence of limiting concentrations of pABA, while folate production by clade II strains was not observed under these conditions (Supplementary Figure S2, Supplementary Material online). The organization of the folate genes is similar in both clades, and the proteins have high sequence homology (93-99%). The *folB*, *folK*, *folE*, *thfs* (*folC2)*, *folQ,* and *folP* genes are clustered and map between nucleotides 1,363,448 and 1,369,044 in JCM 1112 (Figure 2; Supplementary Table S4, Supplementary Material online); *folA* and *folC1* are single genes that map at other positions within the genome (Supplementary Table S4, Supplementary Material online).

Production of the antimicrobial compound, reuterin, is enhanced in *L. reuteri* clade VI strains

Most human-derived isolates of *L. reuteri* produce the broad-spectrum antimicrobial compound, reuterin (a mixture of 3-hydroxypropionaldehyde isomers) (Vollenweider, et al. 2003; Walter, et al. 2011), from glycerol in a vitamin B_{12}-dependent process (Axelsson, et al.

1989; Talarico and Dobrogosz 1990). Reuterin production is dependent on the presence of the horizontally-acquired 57 gene cluster, *pdu-cbi-hem-cob* (Figure 4A; Supplementary Table S5, Supplementary Material online)(Morita, et al. 2008), which contains the *pdu* genes that encode enzymes required for glycerol and 1,2-propanediol utilization, linked to the *cbi-hem-cob* genes that encode the proteins required for the biosynthesis of vitamin B_{12}, or cobalamin (Santos, et al. 2008a). In previous work, we observed that strain ATCC 55730 (clade VI) produced three-fold more reuterin than the clade II strains ATCC 6475, ATCC 4659, and ATCC 5679 during the stationary phase of growth (Spinler, et al. 2008). To further examine the strain-specific characteristics of reuterin production, we tested whether reuterin production by human-derived *L. reuteri* was growth phase dependent.

We compared reuterin production in clade VI strain DSM 17938 (a plasmid-cured version of ATCC 55730 (Rosander, et al. 2008)) and ATCC 6475 (clade II), at time points representing logarithmic (4 h), early stationary (8 h), stationary (12 h), and late stationary (24 h) growth phases (Figure 4B). In *L. reuteri* strain DSM 17938, reuterin production increased more than 16-fold from *ca.* 9 mM during logarithmic phase (4 h) to *ca.* 140 mM during stationary phase (12h, Figure 4B). In contrast, in strain ATCC 6475, reuterin production was growth phase independent until late stationary phase and was about one third that of DSM 17938 (Figure 4B). To expand these observations, we measured *in vitro* reuterin production from stationary phase cultures of the 23 additional human-derived *L. reuteri* strains. Statistically significant ($p < 0.05$) clade-specific differences were observed with all but one clade VI strain (MM36-1a). With the exception of this strain, all clade VI strains produced >100 mM reuterin, while all but three of the clade II strains produced less than 50 mM reuterin (Figure 4C).

While the organization of the *pdu-cbi-hem-cob* gene cluster is conserved in all *L. reuteri* genomes that carry it, the genomic islands have distinct clade-specific sequence homology (Figure 2; Supplementary Figure S1, Supplementary Material online). A comparison of the amino acid identity of each predicted ORF from the human strains relative to the corresponding ORF in ATCC 55730 is shown in Figure 4A. Note that all of the CF48-3A protein sequences are 100% identical to those of ATCC 55730. In clade II, high levels of homology between clades (>95%) are restricted to two of the three reuterin biosynthesis proteins (PduC and PduD) plus additional proteins encoded by the *pdu* operon. The predicted vitamin B_{12} biosynthesis proteins encoded by the *pdu-cbi-hem-cob* genomic island are not as similar between clades (61-95% identity) (Figure 4A). The glycerol dehydratase requires vitamin B_{12} as a cofactor for activity, so differences in vitamin B_{12} biosynthetic capacity should also impact the organism's capacity to produce reuterin.

Clade-specific transcriptional regulation by the AraC-like regulator, PocR

Just as pathogenic microbes possess stringent mechanisms for regulating virulence factors, beneficial microbes have evolved specific mechanisms for regulating probiotic functions. A gene encoding one such regulator, PocR, which is a member of the AraC family of transcriptional regulators, is encoded by the *pdu-cbi-hem-cob* gene cluster (Morita, et al. 2008). The 364 amino acid PocR protein encoded by ATCC 55730 and CF48-3A is 100% identical, however the clade VI and clade II proteins are only 80% identical (Figure 4A). In a previous study comparing RNA expression levels from wild-type and a *pocR* insertion mutant in ATCC 6475, we showed that PocR regulates transcription of operons involved in both reuterin and vitamin B_{12} synthesis in the clade II strains JCM 1112 and ATCC 6475 (Santos, et al. 2011). The growth phase dependent production of reuterin observed in clade VI strain DSM 17938

suggested that its production might be transcriptionally regulated. To test this, we created an equivalent *pocR* insertion mutant in DSM 17938 and performed comparative microarray analysis. Inactivation of *pocR* in either ATCC 6475 or DSM 17938 affected expression of genes within the *pdu-cbi-hem-cob* gene cluster as anticipated (Supplementary Figure S3 and Supplementary Tables S6 and S7, Supplementary Material online), but also affected the expression of genes at other loci relevant to probiosis, including genes involved in arginine catabolism and folate production, as listed in Supplementary Tables S6 and S7, Supplementary Material online. Although not part of this study, it should be noted that we previously were able to complement the ATCC 6475::*pocR* mutation with a plasmid-borne copy of the ATCC 6475 *pocR* gene (Santos, et al. 2011) so the insertions are likely non-polar.

When comparing PocR-affected gene expression, the most striking strain-dependent difference was observed for *pocR* itself. PocR is predicted to function by both activating and repressing transcription from cognate promoters (Santos, et al. 2011), and as an AraC family transcriptional regulator it may also regulate its own transcription. Because the *pocR* mutants were gene interruptions, not deletions, it was possible to score *pocR* transcription in the mutants. Under the conditions tested, DSM 17938 (clade VI) PocR did not affect its own expression, however in ATCC 6475, *pocR* transcription was repressed ~38 fold, suggesting that the protein is auto-regulatory in this clade II strain (Supplementary Figure S3 and Tables S6 and S7, Supplementary Material online). These results, along with others (Chen, et al. 1994; Mellin, et al. 2013; Rondon and Escalante-Semerena 1992; Santos 2008), indicate that strain-specific regulatory mechanisms can exist within human-derived *L. reuteri*

Strain-dependent transcriptional regulation by PocR is observed for the major regulons of the 57 gene *pdu-cbi-hem-cob* cluster (Supplementary Figure S3, Supplementary Material online).

As discussed, the *pdu* genes encode distinct functions including the production of the antimicrobial compound, reuterin (Morita, et al. 2008; Talarico and Dobrogosz 1990) and the utilization of glycerol and propanediol as electron acceptors to support growth (Sriramulu, et al. 2008; Talarico, et al. 1990). This is in contrast to other organisms that have two isofunctional vitamin B_{12}-dependent enzymes for dehydrating glycerol or propanediol (Daniel, et al. 1998; Toraya and Fukui 1977; Toraya, et al. 1980). Since the PduCDE enzyme is vitamin B_{12} dependent, upregulation of the *cbi-hem-cob* genes along with the *pdu* genes, as observed in strain DSM 17938 (clade VI), is consistent with elevated reuterin production by this clade (Figure 4). Conversely, the low level of expression of *cbi-hem-cob* genes in ATCC 6475, even in the presence of high levels of *pdu* expression, may explain the reduced reuterin production in this clade (Supplementary Figure S3, Supplementary Material online). Overall, these data suggest that the human-derived strains from clade II and VI differentially exploit the functionality of the *pdu-cbi-hem-cob* cluster, with the former focusing this cluster on energy gain while the latter produce more reuterin.

Factors produced by *L. reuteri* strains differentially affect cytokine production by human myeloid cells

We quantified eleven human cytokines produced by stimulated THP-1 monocytoid cells in response to treatment with cell-free supernatants from twelve clade II and thirteen clade VI strains (Figure 5). The effects were generally clade-specific and consistent with the prior literature regarding *L. reuteri* and immunomodulation (Pena, et al. 2005). Clade II strains can be considered anti-inflammatory based on aggregate cytokine responses (Figure 5) and the established correlation between TNF inhibition and suppression of intestinal inflammation by *L. reuteri* (Hemarajata, et al. 2013; Pena, et al. 2005). In addition to suppressing TNF, most clade

II strains suppressed monocyte chemoattractant protein-1 (MCP-1), IL-1β, and interleukin-12 (IL-12). Suppression of the pro-inflammatory cytokines TNF, MCP-1, IL-1β, and IL-12 by clade II strains is consistent with the known ability of *L. reuteri* strain ATCC 6475 to suppress intestinal inflammation.

In contrast to the suppression of pro-inflammatory cytokines by clade II, supernatants derived from clade VI strains stimulated IL-7, IL-12, and IL-13 production ($p < 0.05$ versus media control and clade II strains, Figure 5). Clade VI strains are considered immunostimulatory based on these aggregate cytokine responses and they may serve the host by promoting the development and maintenance of an active mucosal immune surveillance system. Most clade VI members (but not ATCC 55730 and CF48-3A) suppressed IL-5 production (Figure 5); these findings appear to be inconsistent with the aggregate cytokine responses. However, IL-5 can play a prominent role at sites distal to the intestine and may modulate immune responses at other body sites. IL-5 is a cytokine that promotes eosinophil differentiation from bone marrow precursor cells, and acts as a potent eosinophil chemoattractant (Akuthota and Weller 2012). Eosinophilic inflammation contributes to distinct disease phenotypes of the skin, lungs, and sinuses, and suppression of eosinophilic inflammation could be beneficial in decreasing atopic and allergic diseases away from the intestine. Oral administration of *L. reuteri* ATCC 55730 (clade VI) can modulate IFN-γ and IL-4 cytokine expression at sites distal from the intestine in patients with atopic dermatitis (Miniello, et al. 2010), and can reduce the incidence of IgE-associated eczema at 2 years of age (Abrahamsson, et al. 2007; Forsberg, et al. 2013). Others have shown that oral administration of *L. reuteri* modulates allergic airway responses in mice (Forsythe, et al. 2007; Karimi, et al. 2009).

Histamine production by *L. reuteri* clade II strains corresponds with anti-inflammatory properties.

The anti-inflammatory properties of *L. reuteri* ATCC 6475 (clade II) have been extensively characterized, and some have recently been linked to its ability to produce histamine from histidine (Hemarajata, et al. 2013; Thomas, et al. 2012). In *L. reuteri,* and other lactobacilli, histamine is produced by a pyruvoyl-dependent histidine decarboxylase (HdcA) (Thomas, et al. 2012). The *hdcA* gene is linked to *hdcB*, encoding a protein of unknown function that is presumed, by association, to have a role in the decarboxylation system (Le Jeune, et al. 1995; van Poelje and Snell 1990). A gene encoding a histidine-histamine antiporter (*hdcP*) (Lucas, et al. 2005) lies immediately upstream of *hdcAB* in the clade II strains and they are directly linked to a second histidyl-tRNA synthetase (*hisRS2*) gene that shares limited homology with the core *hisRS* gene (Figure 6A). The predicted HisRS2 protein contains the catalytic residues required for histidyl-tRNA synthesis (Sissler, et al. 1999) so it is likely that this protein contributes to the pool of charged tRNAs available for protein synthesis under conditions of histidine depletion in this organism, which is a histidine auxotroph.

In JCM 1112, the histidine decarboxylase locus maps between coordinates 1923821 and 1928862 (Supplementary Table S8, Supplementary Material online) and it occurs at similar chromosomal positions in the genomes of DSM 20016, ATCC 4659, and ATCC 6475, but the genes are not present in the genomes of ATCC 55730 or CF48-3A. PCR analysis of additional nonsequenced clade VI strains revealed the occasional presence of the gene cluster in this clade (see below). Database searches revealed the presence of complete histidine decarboxylase gene clusters on *Lactobacillus* plasmids pLRI01 (GenBank NC_021504), pLF01 (GenBank AVAB01000110), and an unstable 80 kb plasmid in *L. hilgardii* IOEB 0006 (Lucas, et al. 2005);

as well as in several *Lactobacillus* genomes: *L. buchneri* B301 (Martin, et al. 2005), *L. saerimneri* 30a (Romano, et al. 2013), *L. sakei* LTH 2076 (GenBank DQ13288) and *L. vaginalis* ATCC 49540 (GenBank NZ_ACGV0). The histidine decarboxylase gene (*hdcA*) is also found in the incomplete genomes of *L. fructivorans* KCTC 3543 (GenBank NZ_AEQY) and *L. rossiae* DSM 15814 (GenBank NZ_AUAW).

Bioassays of the 25 human-derived *L. reuteri* strains show that the majority of clade II strains produced histamine (Figure 6B), and conditioned media from histamine producers suppressed TNF production from human monocytes by more than 50% relative to controls (Figure 5). All *L. reuteri* strains, including those in clade VI, were tested by PCR for the presence of the *hdcP*, *hdcA*, *hdcB*, and *hisRS2* genes (Figure 6B). As expected, all histamine producing strains in both clades contained the entire gene cluster and non-producing strains lacked the histidine decarboxylase gene and *hdcB*. Clade II strains SR-11 and SR-14 were $hdcP^{+}hdcA^{-}hdcB^{-}hisRS2^{+}$ suggesting that they may have incurred an internal deletion within the gene cluster; the presence of this deletion was not confirmed. Clade VI strains M81R43 and MV36-2a both yielded a positive signal for the *hisRS2* gene but were negative for the remainder of the histidine decarboxylase gene cluster. Two clade VI strains, M45R2 and MM36-1a, are $hdcP^{+}hdcA^{+}hdcB^{+}hisRS2^{+}$ and produce histamine (Figure 6B), but do not suppress TNF production (Figure 5). Perhaps these strains produce histamine antagonists or other factors that counteract this anti-inflammatory phenotype. While histamine production by *L. reuteri* clade II strains has clearly been linked to the TNF-suppressive phenotype, other clade specific factors may be responsible for the effects on additional anti-inflammatory cytokines observed here.

Do distinct probiotic functions reflect the evolutionary history of strains?

Our analyses demonstrate that strains of clade II and VI differ substantially in their functional attributes, which is likely to reflect their distinct ecology and the symbiotic relationships that they maintain with their hosts. Currently, we can only speculate on how these differences evolved. The majority of clade II isolates were isolated from human fecal samples, and these strains do not cluster with isolates from other host species (Oh et al., 2000), suggesting that clade II represents the autochthonous *L. reuteri* population in the human intestinal tract. In contrast, in clade VI, human isolates cluster tightly with isolates from poultry (Oh et al., 2010), and since strains from this cluster are extremely rare in human fecal samples (but very common in poultry fecal samples), they might be allochthonous to humans originating from poultry (Frese, et al. 2011; Oh, et al. 2010). Alternatively, if we consider that the majority of human-derived clade VI strains are indeed autochthonous to humans, the population genetic structure suggests that migration to humans was very recent. In either case, it is likely that clade VI strains have evolved with poultry for most of their evolutionary history. Our findings support such distinct evolutionary trajectories.

The ecological niches occupied by *L. reuteri* strains in the GI tracts of chickens and humans are highly distinct. *L. reuteri* is one of the most abundant *Lactobacillus* spp. in the crop of chickens where it forms biofilms on the stratified squamous epithelium of the crop (Abbas Hilmi, et al. 2007). Accordingly, genome comparisons by microarray analysis revealed that genome content of clade VI strains is more similar to rodent and pig strains, reflecting similarities in lifestyle, probably related to similarities in biofilm formation in the proximal gastrointestinal tract of these animals. In contrast, the human gastrointestinal tracts do not contain stratified squamous epithelia, and *Lactobacillus* biofilms have not been demonstrated in the human gut (Walter 2008). This is consistent with the genome content of clade II strains,

which have deleted virtually all large surface proteins annotated as adhesins and the two EPS clusters present in *L. reuteri* strains from rodents (Frese et al., 2011).

Following the theory that clade II strains are autochthonous to humans, they might focus the function of the PduCDE enzyme on propanediol utilization and energy generation to support growth in the human gut where simple sugars are limited. Propanediol is produced by gut microbes as a fermentation byproduct of rhamnose and fucose (Badia, et al. 1985). These simple sugars are components of glycoconjugates found in intestinal epithelial cells and plant cell walls (Bry, et al. 1996), as well as in breast milk oligosaccharides (Castanys-Munoz, et al. 2013). Sriramulu, *et al.* (Sriramulu, et al. 2008) have shown that DSM 20016 (clade II) shows a growth advantage over rodent-derived *L. reuteri* 100-23 (that lacks the *pdu-cbi-hem-cob* cluster) when propanediol is provided in the growth medium, further supporting the idea that clade II strains preferentially utilize the propanediol dehydratase activity of PduCDE.

Furthermore, the cytokine bioassays reveal that most clade II strains are associated with suppression of pro-inflammatory cytokines produced by macrophages, while most clade VI strains are associated with a general pro-inflammatory cytokine increase, and specifically suppress IL-5. Clade VI strains may play an important role in fostering development and maintenance of robust and fully mature immune responses by the gut microbiome. Clade II strains may serve an equally important role in helping the host regulate and resolve immune responses, and suppress potentially deleterious effects of mucosal inflammation. Since the immune cells used in our assays are of human origin, the findings might reflect the distinct relationships of the clades with the human host.

Conclusions

Host-driven evolution of *L. reuteri* has resulted in ecotypes that specialize towards particular host species (Frese, et al. 2011; Oh, et al. 2010). We show that human strains from two distinct ecotypes differ markedly in bacterial functions that are likely to influence the interrelationships of these strains with the human host (Figure 7), indicating that the evolutionary trajectory of a gut microbe influences bacterial traits that can be beneficial to its host. The most striking differences between the two ecotypes are the amounts of the antimicrobial compound reuterin produced, and the remarkable differences in intestinal immunomodulatory capacity between the clades. Clade II strains can be considered immunosuppressive and clade VI strains can be considered immunostimulatory. The contrasting immunomodulatory effects of these two separate clades provide a prominent example how different lineages within a single bacterial species can benefit the host by different mechanisms. The different genomic characteristics of clades II and VI evident in the ANI and gene content analysis (Figure 3B) indicate that the evolution of the two human-derived *L. reuteri* ecotypes was influenced by separate and distinct environmental pressures (and potentially by two different hosts). Most importantly, the evolutionary process impacted how strains interact with host cells. Thus, our work provides novel information about host-microbe interrelationships as it shows that distinct evolutionary paths within the same species are likely to determine how gut microbes impact their host, which is likely important for their health effects. Important lineage distinctions within a single bacterial species have recently been described in other organisms and include the distinction of two different clades of *Wolbachia* infecting the same population of *Drosophila* (Ellegaard, et al. 2013), subclades of freshwater isolates of the Alphaproteobacteria SAR11 that are unique and distinct from saltwater clades (Zaremba-Niedzwiedzka, et al. 2013), and the identification of four distinct clades of *Chlamydia trachomatis* associated with four distinct disease types (Joseph, et

al. 2012). These and our work are concrete examples of the evolution of discreet ecotypes within a bacterial species and address the "bacterial species concept" (Rossello-Mora and Amann 2001). These findings suggest that evolutionary criteria could be helpful for the selection of probiotic strains. Understanding the genomic basis of the probiotic characteristics associated with distinct lineages of *L. reuteri* will ultimately allow us to predict and assign groups of probiotic *L. reuteri* strains best suited to prevent or treat various classes of diseases.

Abbreviations

AFLP, amplified fragment length polymorphism; ANI, average nucleotide identity; Arg, arginine; ATCC, American Type Culture Collection; BRIG, BLAST Ring Image Generator; CDS, coding sequence; COG, Cluster of Orthologous Genes; DHF, dihydrofolate; DHP, 7,8-dihydropteroate; DHPPP, 6-hydroxymethyl-7,8-dihydropterin pyrophosphate; DSM, German Collection of Microorganisms and Cell Cultures; fol, folate; GI, gastrointestinal; GM-CSF, granulocyte macrophage colony-stimulating factor; hdc, histidine decarboxylase; hisRS, histidyl-tRNA synthetase; IL, interleukin; IMG/ER, Integrated Microbial Genomes/Expert Review; JCM, Japan Collection of Microorganisms; Kyoto Encyclopedia of Genes and Genomes, KEGG; Mb, megabase; MCP, monocyte chemoattractant protein; MGH, Massachusetts General Hospital; MLSA, multi-locus sequence analysis; MRS, deMan, Rogosa, Sharpe; NCBI, National Center for Biotechnology Information; OGU, operational gene unit; ORF, open reading frame; pABA, para-aminobenzoic acid; Pdu, propanediol utilization; THF, tetrahydrofolate; TLR2, Toll-like receptor 2; TNF, tumor necrosis factor

Competing interests

JV and JW received unrestricted research support from BioGaia AB.

Author's contributions

JKS and SKH conceived the project. JKS generated the *pocR* insertion mutants, completed experimental design and data analysis for all microbial assays. JKS and SKH designed and JKS, EBH, and SKH performed comparative genomic analyses. JKS, SKH, and JW wrote the manuscript and generated figures. SKH updated annotations. AS performed experiments for the following assays: folate, end-point reuterin, histamine, THP-1 bioassay, and end-point PCR. SV completed the Luminex assay. MAB carried out reuterin time course experiments. DMAS performed microarray experiments, and TAM contributed microarray data analysis. PLO and JW provided additional human-derived *L. reuteri* strains and provided MLSA analysis. SD performed folate detection. JV and SKH provided funding for the project. All authors reviewed and approved the manuscript.

Acknowledgements

We thank Eamonn Connolly (BioGaia AB, Stockholm) for providing several human-derived *L. reuteri* strains listed in Table 2, William Borowski for initial amino acid sequence analysis, and Carissa Thomas and Peera Hemarajata for sharing their expertise in the TNF and histamine assays. We would like to acknowledge the gracious efforts of Sabeen Raza and Karen Prince for assistance with figure illustrations.

This work was supported by the National Institutes of Health (NIH), National Cancer Institute [grant number U01 CA170930 to JV]; by the National Institute of Diabetes, Digestive and Kidney Diseases [grant numbers UH3 DK083990 to JV, P30 DK56338, and R01 DK065075 to JV]; by the National Center for Complementary and Alternative Medicine [grant number R01 AT004326 to JV]; by the NIH Common Fund and National Human Genome Research Institute

[grant numbers U54 HG003273 to JV, U54 HG004973 to SKH]; and by BioGaia AB (Stockholm, Sweden) to JV.

References

Abbas Hilmi HT, Surakka A, Apajalahti J, Saris PE. 2007. Identification of the most abundant *Lactobacillus* species in the crop of 1- and 5-week-old broiler chickens. Appl Environ Microbiol. 73: 7867-7873. doi: 10.1128/AEM.01128-07

Abrahamsson TR, et al. 2007. Probiotics in prevention of IgE-associated eczema: a double-blind, randomized, placebo-controlled trial. J Allergy Clin Immunol. 119: 1174-1180. doi: 10.1016/j.jaci.2007.01.007

Akuthota P, Weller PF. 2012. Eosinophils and disease pathogenesis. Semin Hematol. 49: 113-119. doi: 10.1053/j.seminhematol.2012.01.005

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 12: 402. doi: 10.1186/1471-2164-12-402

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215: 403-410. doi: 10.1016/S0022-2836(05)80360-2

Axelsson LT, Chung TC, Dobrogosz WJ, Lindgren SE. 1989. Production of a broad spectrum antimicrobial substance by *Lactobacillus reuteri*. Microb Ecol Health Dis. 2: 131-136.

Badia J, Ros J, Aguilar J. 1985. Fermentation mechanism of fucose and rhamnose in *Salmonella typhimurium* and *Klebsiella pneumoniae*. J Bacteriol. 161: 435-437.

Barabas O, et al. 2008. Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. Cell. 132: 208-220. doi: 10.1016/j.cell.2007.12.029

Bry L, Falk PG, Midtvedt T, Gordon JI. 1996. A model of host-microbial interactions in an open mammalian ecosystem. Science. 273: 1380-1383.

Capozzi V, Russo P, Duenas MT, Lopez P, Spano G. 2012. Lactic acid bacteria producing B-group vitamins: a great potential for functional cereals products. Appl Microbiol Biotechnol. 96: 1383-1394. doi: 10.1007/s00253-012-4440-2

Castanys-Munoz E, Martin MJ, Prieto PA. 2013. 2'-fucosyllactose: an abundant, genetically determined soluble glycan present in human milk. Nutr Rev. 71: 773-789. doi: 10.1111/nure.12079

Chen P, Andersson DI, Roth JR. 1994. The control region of the *pdu/cob* regulon in *Salmonella typhimurium*. J Bacteriol. 176: 5474-5482.

Cohan FM, Koeppel AF. 2008. The origins of ecological diversity in prokaryotes. Curr Biol. 18: R1024-1034. doi: 10.1016/j.cub.2008.09.014

Daniel R, Bobik TA, Gottschalk G. 1998. Biochemistry of coenzyme B_{12}-dependent glycerol and diol dehydratases and organization of the encoding genes. FEMS Microbiol Rev. 22: 553-566.

Didelot X, Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics. 175: 1251-1266. doi: 10.1534/genetics.106.063305

Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SG. 2013. Comparative genomics of *Wolbachia* and the bacterial species concept. PLoS Genet. 9: e1003381. doi: 10.1371/journal.pgen.1003381

Forsberg A, Abrahamsson TR, Bjorksten B, Jenmalm MC. 2013. Pre- and post-natal *Lactobacillus reuteri* supplementation decreases allergen responsiveness in infancy. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 43: 434-442. doi: 10.1111/cea.12082

Forsythe P, Inman MD, Bienenstock J. 2007. Oral treatment with live *Lactobacillus reuteri* inhibits the allergic airway response in mice. Am J Respir Crit Care Med. 175: 561-569. doi: 10.1164/rccm.200606-821OC

Frese SA, et al. 2011. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. PLoS Genet. 7: e1001314. doi: 10.1371/journal.pgen.1001314

Hansen EE, et al. 2011. Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. Proc Natl Acad Sci U S A. 108 Suppl 1: 4599-4606. doi: 10.1073/pnas.1000071108

Hemarajata P, et al. 2013. *Lactobacillus reuteri*-Specific Immunoregulatory Gene (RsiR) Modulates Histamine Production and Immunomodulation by *Lactobacillus reuteri*. J Bacteriol. doi: 10.1128/JB.00261-13

Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 26: 680-682. doi: 10.1093/bioinformatics/btq003

Jones SE, Versalovic J. 2009. Probiotic *Lactobacillus reuteri* biofilms produce antimicrobial and anti-inflammatory factors. BMC Microbiol. 9: 35. doi: 10.1186/1471-2180-9-35

Joseph SJ, et al. 2012. Population genomics of *Chlamydia trachomatis*: insights on drift, selection, recombination, and population structure. Mol Biol Evol. 29: 3933-3946. doi: 10.1093/molbev/mss198

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40: D109-114. doi: 10.1093/nar/gkr988

Karimi K, Inman MD, Bienenstock J, Forsythe P. 2009. *Lactobacillus reuteri*-induced regulatory T cells protect against an allergic airway response in mice. Am J Respir Crit Care Med. 179: 186-193. doi: 10.1164/rccm.200806-951OC

Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci. 361: 1929-1940. doi: 10.1098/rstb.2006.1920

Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 102: 2567-2572. doi: 10.1073/pnas.0409727102

Le Jeune C, Lonvaud-Funel A, ten Brink B, Hofstra H, van der Vossen JM. 1995. Development of a detection system for histidine decarboxylating lactic acid bacteria based on DNA probes, PCR and activity test. J Appl Bacteriol. 78: 316-326.

Li W, Godzik A. 2006. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22: 1658-1659. doi: 10.1093/bioinformatics/btl158

Lin YP, Thibodeaux CH, Pena JA, Ferry GD, Versalovic J. 2008. Probiotic *Lactobacillus reuteri* suppress proinflammatory cytokines via c-Jun. Inflamm Bowel Dis. 14: 1068-1083. doi: 10.1002/ibd.20448

Lucas PM, Wolken WA, Claisse O, Lolkema JS, Lonvaud-Funel A. 2005. Histamine-producing pathway encoded on an unstable plasmid in *Lactobacillus hilgardii* 0006. Appl Environ Microbiol. 71: 1417-1424. doi: 10.1128/AEM.71.3.1417-1424.2005

Markowitz VM, et al. 2010. The integrated microbial genomes system: an expanding comparative analysis resource. Nucleic Acids Res. 38: D382-390. doi: 10.1093/nar/gkp887

Marquis RE, Bender GR, Murray DR, Wong A. 1987. Arginine deiminase system and bacterial adaptation to acid environments. Appl Environ Microbiol. 53: 198-200.

Martin MC, Fernandez M, Linares DM, Alvarez MA. 2005. Sequencing, characterization and transcriptional analysis of the histidine decarboxylase operon of *Lactobacillus buchneri*. Microbiology. 151: 1219-1228. doi: 10.1099/mic.0.27459-0

Mellin JR, et al. 2013. A riboswitch-regulated antisense RNA in *Listeria monocytogenes*. Proc Natl Acad Sci U S A. 110: 13132-13137. doi: 10.1073/pnas.1304795110

Miniello VL, et al. 2010. *Lactobacillus reuteri* modulates cytokines production in exhaled breath condensate of children with atopic dermatitis. J Pediatr Gastroenterol Nutr. 50: 573-576. doi: 10.1097/MPG.0b013e3181bb343f

Morita H, et al. 2008. Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. DNA Res. 15: 151-161. doi: 10.1093/dnares/dsn009

Nelson KE, et al. 2010. A catalog of reference genomes from the human microbiome. Science. 328: 994-999. doi: 10.1126/science.1183605

Oh PL, et al. 2010. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. ISME J. 4: 377-387. doi: 10.1038/ismej.2009.123

Pena JA, et al. 2004. Genotypic and phenotypic studies of murine intestinal lactobacilli: species differences in mice with and without colitis. Appl Environ Microbiol. 70: 558-568.

Pena JA, et al. 2005. Probiotic *Lactobacillus* spp. diminish *Helicobacter hepaticus*-induced inflammatory bowel disease in interleukin-10-deficient mice. Infect Immun. 73: 912-920. doi: 10.1128/IAI.73.2.912-920.2005

Romano A, et al. 2013. Genome sequence of *Lactobacillus saerimneri* 30a (formerly *Lactobacillus* sp. strain 30a), a reference lactic acid bacterium strain producing biogenic amines. Genome Announc. 1. doi: 10.1128/genomeA.00097-12

Rondon MR, Escalante-Semerena JC. 1992. The *poc* locus is required for 1,2-propanediol-dependent transcription of the cobalamin biosynthetic (*cob*) and propanediol utilization (*pdu*) genes of *Salmonella typhimurium*. J Bacteriol. 174: 2267-2272.

Rosander A, Connolly E, Roos S. 2008. Removal of antibiotic resistance gene-carrying plasmids from *Lactobacillus reuteri* ATCC 55730 and characterization of the resulting daughter strain, *L. reuteri* DSM 17938. Appl Environ Microbiol. 74: 6032-6040. doi: 10.1128/AEM.00991-08

Rossello-Mora R, Amann R. 2001. The species concept for prokaryotes. FEMS Microbiol Rev. 25: 39-67.

Rossi M, Amaretti A, Raimondi S. 2011. Folate production by probiotic bacteria. Nutrients. 3: 118-134. doi: 10.3390/nu3010118

Russell WM, Klaenhammer TR. 2001. Efficient system for directed integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* chromosomes via homologous recombination. Appl Environ Microbiol. 67: 4361-4364.

Sambrook J, Russell DW. 2001. Molecular cloning: a laboratory manual. In. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Santos F 2008. Vitamin B_{12} synthesis in *Lactobacillus reuteri*. [Ph.D.]. [Wageningen, The Netherlands]: Wageningen University.

Santos F, et al. 2011. Functional identification in *Lactobacillus reuteri* of a PocR-like transcription factor regulating glycerol utilization and vitamin B_{12} synthesis. Microb Cell Fact. 10: 55. doi: 10.1186/1475-2859-10-55

Santos F, et al. 2008a. The complete coenzyme B_{12} biosynthesis gene cluster of *Lactobacillus reuteri* CRL1098. Microbiology. 154: 81-93. doi: 10.1099/mic.0.2007/011569-0

Santos F, Wegkamp A, de Vos WM, Smid EJ, Hugenholtz J. 2008b. High-Level folate production in fermented foods by the B_{12} producer *Lactobacillus reuteri* JCM1112. Appl Environ Microbiol. 74: 3291-3294. doi: 10.1128/AEM.02719-07

Saulnier DM, et al. 2011. Exploring metabolic pathway reconstruction and genome-wide expression profiling in *Lactobacillus reuteri* to define functional probiotic features. PLoS One. 6: e18783. doi: 10.1371/journal.pone.0018783

Schwartz S, et al. 2000. PipMaker--a web server for aligning two genomic DNA sequences. Genome Res. 10: 577-586.

Sissler M, et al. 1999. An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. Proc Natl Acad Sci U S A. 96: 8985-8990.

Spinler JK, et al. 2008. Human-derived probiotic *Lactobacillus reuteri* demonstrate antimicrobial activities targeting diverse enteric bacterial pathogens. Anaerobe. 14: 166-171. doi: 10.1016/j.anaerobe.2008.02.001

Sriramulu DD, et al. 2008. *Lactobacillus reuteri* DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. J Bacteriol. 190: 4559-4567. doi: 10.1128/JB.01535-07

Talarico TL, Axelsson LT, Novotny J, Fiuzat M, Dobrogosz WJ. 1990. Utilization of glycerol as a hydrogen acceptor by *Lactobacillus reuteri*: Purification of 1,3-Propanediol:NAD oxidoreductase. Appl Environ Microbiol. 56: 943-948.

Talarico TL, Dobrogosz WJ. 1990. Purification and characterization of glycerol dehydratase from *Lactobacillus reuteri*. Appl Environ Microbiol. 56: 1195-1197.

Tatusov RL, et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 4: 41. doi: 10.1186/1471-2105-4-41

Thomas CM, et al. 2012. Histamine derived from probiotic *Lactobacillus reuteri* suppresses TNF via modulation of PKA and ERK signaling. PLoS One. 7: e31951. doi: 10.1371/journal.pone.0031951

Toraya T, Fukui S. 1977. Immunochemical evidence for the difference between coenzyme-B_{12}-dependent diol dehydratase and glycerol dehydratase. Eur J Biochem. 76: 285-289.

Toraya T, Kuno S, Fukui S. 1980. Distribution of coenzyme B_{12}-dependent diol dehydratase and glycerol dehydratase in selected genera of *Enterobacteriaceae* and *Propionibacteriaceae*. J Bacteriol. 141: 1439-1442.

van Poelje PD, Snell EE. 1990. Cloning, sequencing, expression, and site-directed mutagenesis of the gene from *Clostridium perfringens* encoding pyruvoyl-dependent histidine decarboxylase. Biochemistry. 29: 132-139.

Vollenweider S, Grassi G, Konig I, Puhan Z. 2003. Purification and structural characterization of 3-hydroxypropionaldehyde and its derivatives. J Agric Food Chem. 51: 3287-3293. doi: 10.1021/jf021086d

Walter J. 2008. Ecological role of lactobacilli in the gastrointestinal tract: Implications for fundamental and biomedical research. Appl Environ Microbiol. 74: 4985-4996. doi: 10.1128/AEM.00753-08

Walter J, Britton RA, Roos S. 2011. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proc Natl Acad Sci U S A. 108 Suppl 1: 4645-4652. doi: 10.1073/pnas.1000099107

Walter J, et al. 2005. A high-molecular-mass surface protein (Lsp) and methionine sulfoxide reductase B (MsrB) contribute to the ecological performance of *Lactobacillus reuteri* in the murine gut. Appl Environ Microbiol. 71: 979-986. doi: 10.1128/AEM.71.2.979-986.2005

Wayne LG. 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol. 37: 463-464.

Wilson DH, Williams G, Herrmann R, Wiesner D, Brookhart P. 2005. Issues in immunoassay standardization: the ARCHITECT Folate model for intermethod harmonization. Clinical chemistry. 51: 684-687. doi: 10.1373/clinchem.2004.042358

Zaremba-Niedzwiedzka K, et al. 2013. Single-cell genomics reveal low recombination frequencies in freshwater bacteria of the SAR11 clade. Genome Biol. 14: R130. doi: 10.1186/gb-2013-14-11-r130

Figures

Fig. 1. Phylogenetic analysis of *L. reuteri* isolates derived from four hosts.

The genealogy of 119 *L. reuteri* strains based on analysis of MLSA sequences. Tree branches are color coded by host origin: green = rodent, red = pig, blue = human, yellow = poultry. The 25 strains included in the functional analyses are indicated by closed black circles (●), and strains with sequenced genomes are indicated by a closed black circle with a yellow center (◉). Reuterin production is designated as follows: ▲ high reuterin production, ▽ low reuterin production. Strains that produce high amounts of folate are indicated by closed green circles (●), and strains producing histamine are indicated by closed red circles (●).

Fig. 2. Genomic comparison of ten *L. reuteri* genomes to clade II strain JCM 1112.

Circular representation of ten *L. reuteri* genomes based on nucleotide homology with respect to JCM 1112. Diagram represents blastn results of each genome against JCM 1112 with results rendered using the BRIG program (Alikhan, et al. 2011). Each genome is color coded as indicated by the legend. Relative shading density (from darker to lighter) within each circle represents relative levels of nucleotide homology. White regions indicate regions with no identity to the reference. Features of interest are annotated.

Fig. 3. Ecological and genetic diversity within the *L. reuteri* species.

Conserved genes (y-axis) versus average nucleotide evolutionary distance (x-axis) plot for ten sequenced *L. reuteri* genomes isolated from 3 mammalian hosts. Each data point represents a whole-genome comparison between two genomes. A) Pair-wise comparisons between all ten genomes. Intra-clade comparisons of human-derived strains are circled near the origin. B) Pair-wise comparisons of human-derived MLSA II and MLSA VI *L. reuteri* genomes.

Fig. 4. The *pdu-cbi-hem-cob* gene cluster and clade-specific reuterin production.

A) CDS map of the 57 gene *pdu-cbi-hem-cob* cluster that encodes the PocR regulator (shown in red), genes *pduCDE* responsible for production of the antimicrobial reuterin, Pdu proteins required for 1,2 propanediol and glycerol utilization, and Cbi, Hem and Cob proteins required for synthesis of vitamin B_{12}, as per Morita, *et al*. (Morita, et al. 2008). Below the CDS map are heat maps representing percent amino acid identities of each protein within the *pdu-cbi-hem-cob* gene cluster of four MLSA II *L. reuteri* genomes relative to the amino acid sequences of MLSA VI proteins from strains ATCC 55730 and CF48-3A. B) Growth curve of strains ATCC 6475 (clade II) and DSM 17938 (clade VI) in MRS medium at 37°C under anaerobic conditions (left); growth phase production of reuterin by strain ATCC 6475 (clade II) and DSM 17938 (clade VI) (right). C) Reuterin production at stationary phase (12h) from 25 human-derived *L. reuteri* strains. Strains with sequenced genomes are indicated in bold with an asterisk (*). Results are expressed as the mean ± SD, n=3, and significant differences between group means as determined by one-way ANOVA (p <0.05) are indicated by the placement of different letters above each bracketed group.

Fig. 5. The effects of *L. reuteri* secreted factors on cytokine profiles of human immune cells.

Inflammatory cytokine production by TLR2-activated THP-1 cells in the presence of culture supernatants from *L. reuteri* clade II or clade VI strains. Supernatants from THP-1 cells were assayed using a Human Cytokine/Chemokine-Premixed 14-plex kit in a Luminex 100 system. All values are presented on a 0 to 1 scale, indicating a max/min normalization per analyte and are displayed as heatmaps with red indicating values closer to the maximum (1) and green indicating values closer to minimum (0).

Fig. 6. Relative production of histamine by human-derived *L. reuteri* clades.

A) CDS map of the *hdcP-hdcA-hdcB-hisRS2* gene cluster in clade II *L. reuteri* strains. B) Histamine in conditioned media by 25 human-derived *L. reuteri* strains was measured by ELISA and compared to media control. Results of end point PCR tests for the four histidine decarboxylase cluster genes are summarized: √ indicates correct PCR product observed; - indicates no PCR product observed. All values are presented on a 0 to 1 scale, indicating a max/min normalization per analyte and are displayed as heatmaps with red indicating values closer to the maximum (1) and green indicating values closer to minimum (0).

Fig. 7. Functional illustration of human-derived *L. reuteri* ecotypes.

L. reuteri strains from both clades promote acidic environments, produce the antimicrobial reuterin, synthesize essential vitamins, and generate immunomodulatory compounds that effect immune signaling in the host. The differences associated with each ecotype are illustrated here.

Tables

Table 1. Characteristics of sequenced *L. reuteri* strains.

Strain Name (Alternate)	MLSA Clade[a]	Host (Body Site)	Genome Length (bp)	ORFs	Scaffolds	GenBank Accession Number(s)
JCM 1112	II	Human (GI Tract)	2,039,414	1820	1	NC_010609
DSM 20016 (F275)	II	Human (GI Tract)	1,999,618	1900	1	NC_009513
ATCC PTA 4659 (MM2-3)	II	Human (Breast Milk)	1,943,466	2044	95	NZ_GG693756-6850
ATCC PTA 6475 (MM4-1A)	II	Human (Breast Milk)	2,067,914	2095	7	NZ_ACGX02000001-007
ATCC 55730 (SD2112)	VI	Human (Breast Milk)	2,264,399	2246	1	NC_015697

CF48-3A	VI	Human (GI Tract)	2,032,595	2164	92	NZ_GG693664-755
100-23	III	Rat (GI Tract)	2,305,557	2181	2	NZ_AAPZ02000001-002
mlc3	III	Mouse (GI Tract)	2,018,630	1986	126[b]	AEAW00000000
lpuph	I	Mouse (GI Tract)	2,116,621	2066	127[b]	AEAX00000000
ATCC 53608 (1063)	IV	Pig (GI Tract)	1,968,532	1864	13	NZ_FR854361-373

[a]As determined by Oh, *et al*. (Oh, et al. 2010)

[b]Number of contigs

Table 2. Human-derived *L. reuteri* strains used in this study.

Strain Name (Alternate)	MLSA Clade[a]	Isolation Site	Source
DSM 20016	II	Feces	DSMZ[b]
ATCC PTA 4659 (MM2-3)	II	Breast Milk	BioGaia AB
ATCC PTA 5289 (FJ1)	II	Oral Cavity	BioGaia AB
ATCC PTA 6475 (MM4-1A)	II	Breast Milk	BioGaia AB
CF15-6	II	Feces	BioGaia AB
CF4-6g	II	Feces	BioGaia AB
JCM 1112	II	Feces	JCM[c]
LMS11-1	II	Feces	MGH[d]
LMS11-3	II	Feces	MGH
MM3-1a	II	Breast Milk	BioGaia AB
SR-11	II	Stomach	O'Toole, PW
SR-14	II	Stomach	O'Toole, PW

ATCC 55730 (SD2112)	VI	Breast Milk	BioGaia AB
CF48-3A	VI	Feces	BioGaia AB
CF6-2a	VI	Feces	BioGaia AB
DSM 17938	VI	Breast Milk	BioGaia AB
M27U15	VI	Breast Milk	BioGaia AB
M45R2	VI	Breast Milk	BioGaia AB
M81R43	VI	Breast Milk	BioGaia AB
MF14-C	VI	Feces	BioGaia AB
MF2-3	VI	Feces	BioGaia AB
MM34-4a	VI	Breast Milk	BioGaia AB
MM36-1a	VI	Breast Milk	BioGaia AB
MV36-2a	VI	Vagina	BioGaia AB
MV4-1a	VI	Vagina	BioGaia AB
ATCC 6475::*pocR*	II	Breast Milk	(Santos, et al. 2011)
DSM 17938::*pocR*	VI	Breast Milk	This study

[a]As determined by Oh, *et al.* (Oh, et al. 2010)

[b]Deutsche Sammlung von Mikroorganismen und Zellkulturen

[c]Japan Collection of Microorganisms

[d]Microbiology Laboratories, Massachusetts General Hospital

Supplementary Material

Supplementary figures S1-S3 and tables S1-S8 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Supplementary Table S1. Primers used in this study.

Supplementary Tables S2a – S2d. Prophage ϕLreu1 - ϕLreu4 Gene Annotations.

Supplementary Table S3. Arginine Catabolism Gene Cluster Annotations.

Supplementary Table S4. Folate Biosynthesis Gene Annotations.

Supplementary Table S5a-S5b. The *pdu-cbi-hem-cob* Gene Cluster Annotations.

Supplementary Table S6. Microarray expression data for selected genes in clade II strain *L. reuteri* ATCC 6475 wild-type versus PocR mutant.

Supplementary Table S7. Microarray expression data for selected genes in clade VI strain *L. reuteri* DSM17938 wild-type versus PocR mutant.

Supplementary Table S8. Histidine Decarboxylase Gene Cluster for *L. reuteri* Clade II.

Supplementary Fig. S1. Genomic comparison of ten *L. reuteri* genomes to strain ATCC 55730.

Circular representation of ten *L. reuteri* genomes based on nucleotide homology with respect to ATCC 55730. Diagram represents blastn results of each genome against ATCC 55730 with results rendered using the BRIG program (Alikhan, et al. 2011). Each genome is color coded as indicated by the legend. Relative shading density (from darker to lighter) within each circle represents relative levels of nucleotide homology. White regions indicate regions with no identity to the reference. Features of interest are annotated.

Supplementary Fig. S2. Folate production by human-derived *L. reuteri* clades.

A) Schematic overview of para-aminobenzoic acid (pABA)-dependent dihydrofolate (DHF) and tetrahydrofolate (THF) biosynthesis by *L. reuteri.* Precursor 6-hydroxymethyl-7,9-dihydropterin pyrophosphate (DHPPP) is synthesized from GTP through the action of FolE, FolQ, FolB, and FolK, and then is condensed with externally acquired pABA by FolP to form DHP (7,8-dihydropteroate), which is glutamylated via FolC1 to produce DHF (dihydrofolate). DHF can be further glutamylated to form DHF polyglutamate or can be converted to THF via ThfS (FolC2) and to THF polyglutamate by FolC1. B) Folate production in filtered supernatants of human-derived clade II or VI *L. reuteri* strains cultured in a defined medium. The results are expressed as the mean ± SD, n=3, * p <0.05 relative to medium control. Group means of clade VI assays were significantly different than those clade II group means and when compared to control medium alone. Strains with sequenced genomes are indicated in bold with an asterisk (*).

Supplementary Fig. S3. Heatmap of PocR-regulated genes in the *pdu-cbi-hem-cob* gene cluster from ATCC 6475 and DSM 17938.

CDS map of the 57 gene *pdu-cbi-hem-cob* cluster that encodes the PocR regulator (shown in red), genes *pduCDE* responsible for production of the antimicrobial reuterin, Pdu proteins required for 1,2 propanediol and glycerol utilization, and Cbi, Hem and Cob proteins required for synthesis of vitamin B_{12}, as per Morita, *et al*. (Morita, et al. 2008). Below the CDS map are heat maps representing fold changes (>1.5) of the wild-type strain with respect to the isogenic *pocR* mutant strain (WT/*pocR* mutant) with an adjusted p-value < 0.05. Values having a p-value > 0.05 are indicated as NS. Data for three genes was undetectable (ND).

. 1.

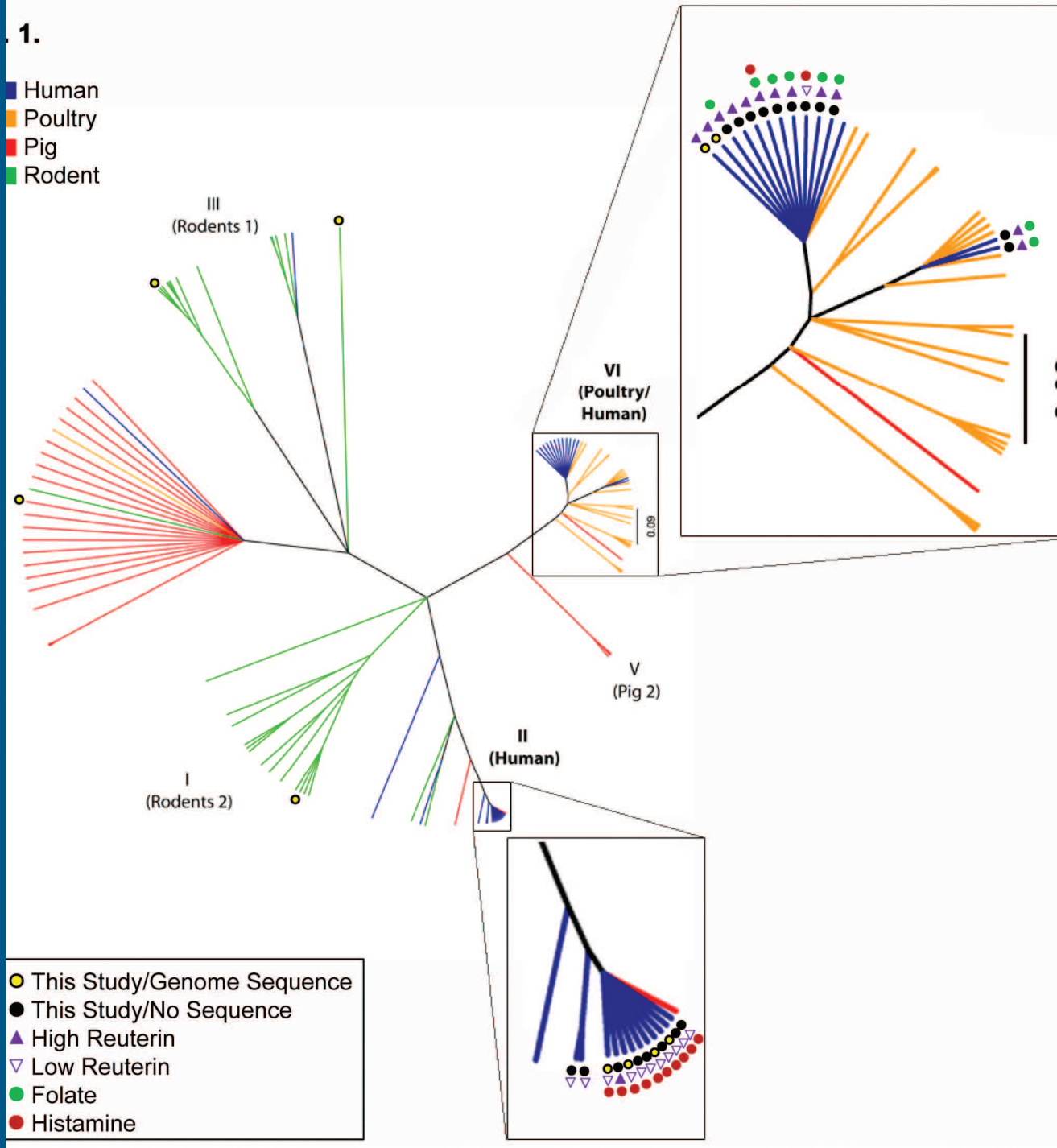
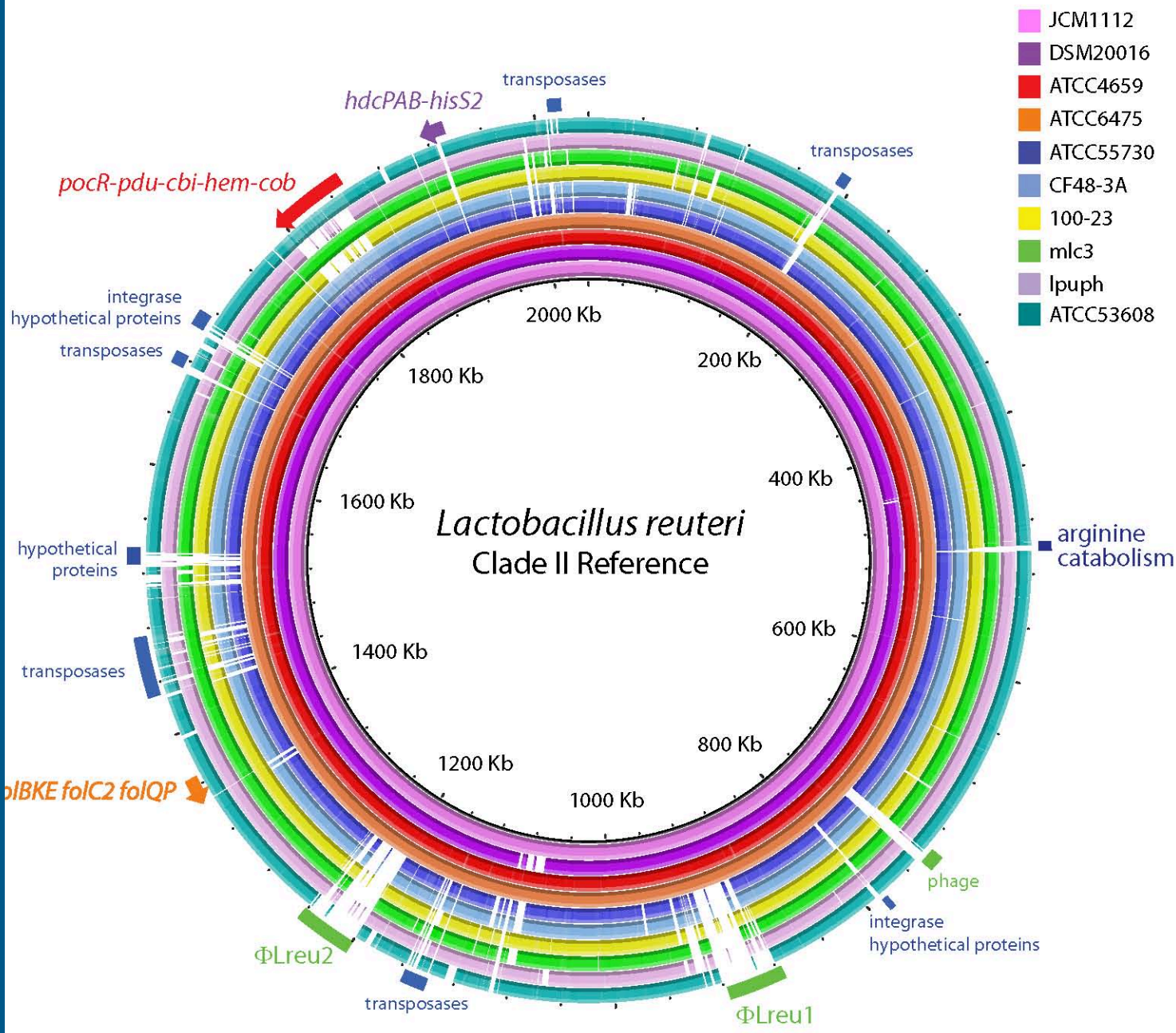
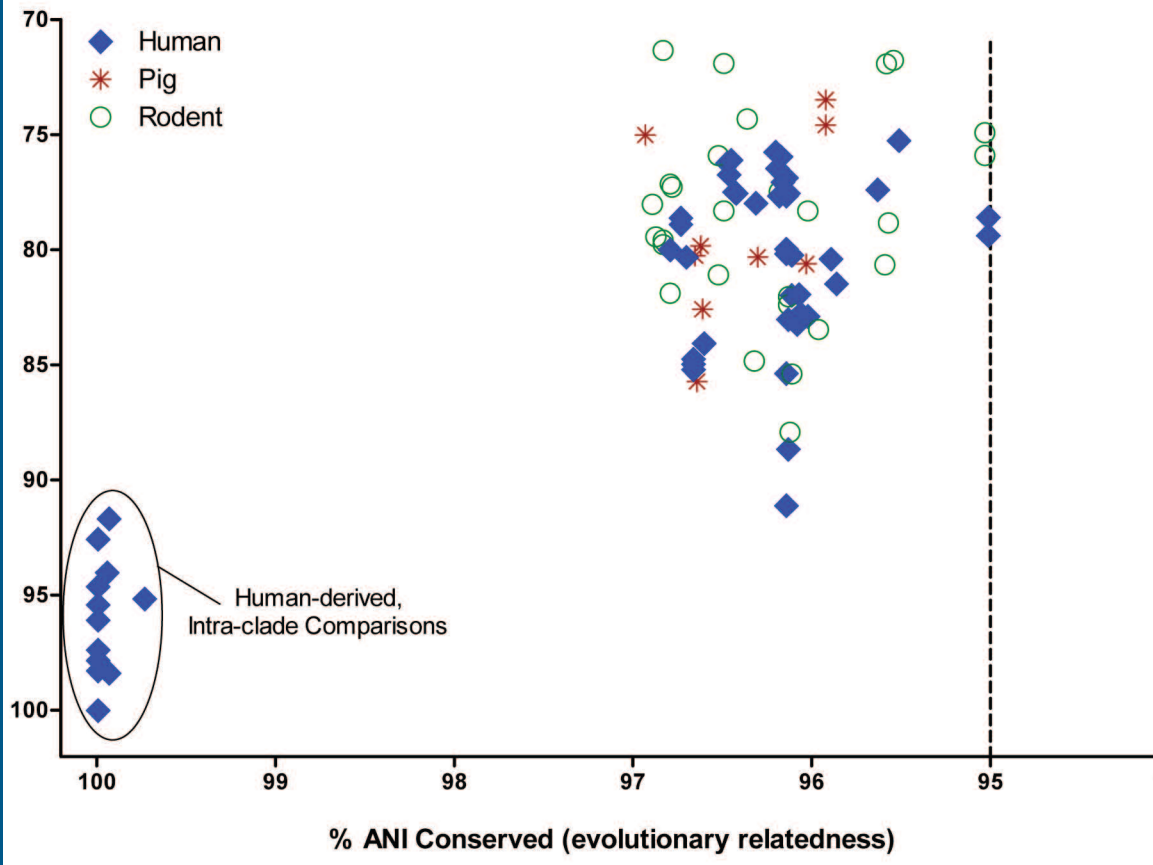
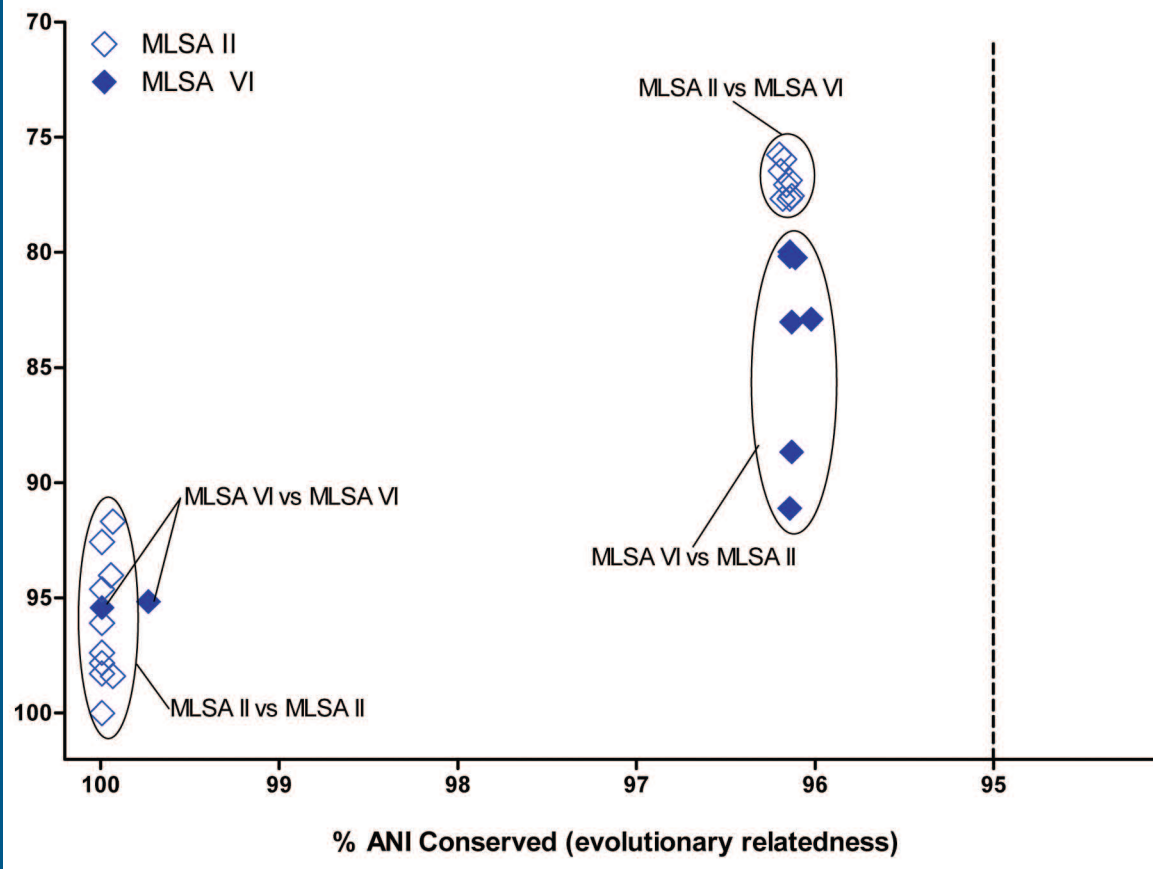

Human
Poultry
Pig
Rodent
III
(Rodents 1)
VI
(Poultry/
Human)
0.09
V
(Pig 2)
II
(Human)
I
(Rodents 2)
This Study/Genome Sequence
This Study/No Sequence
High Reuterin
Low Reuterin
Folate
Histamine

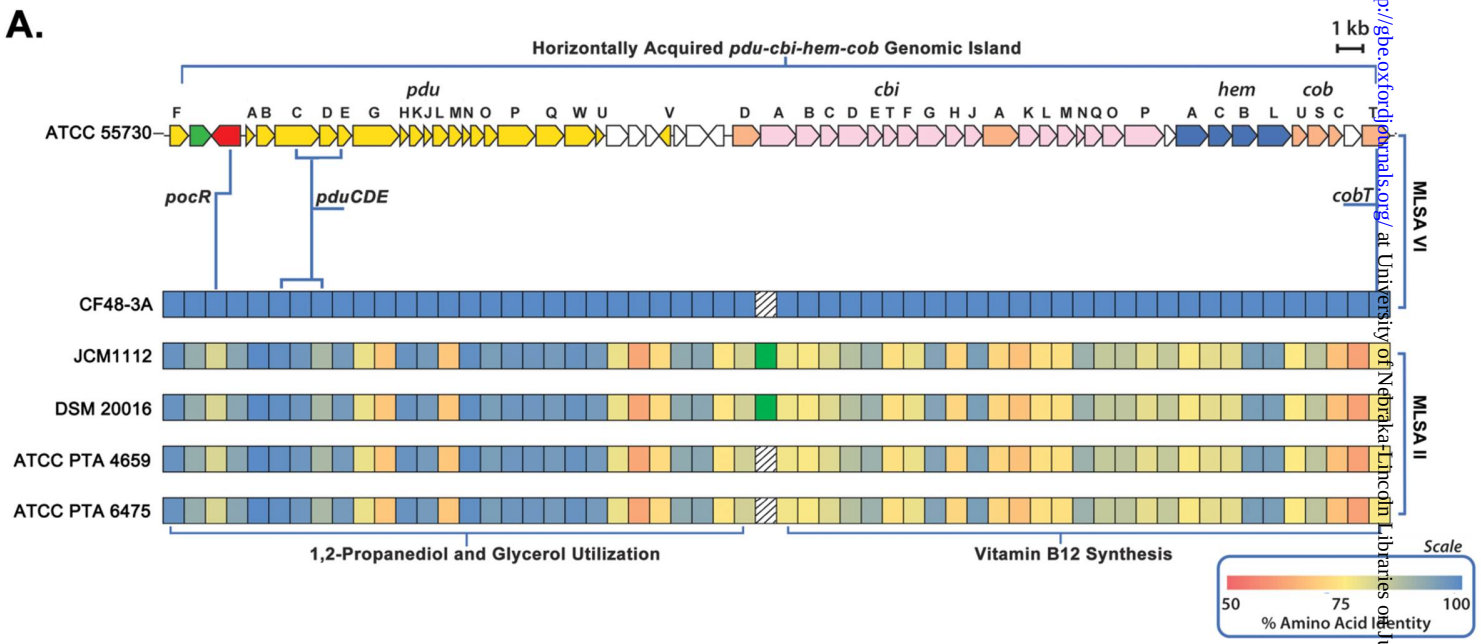
Fig 2

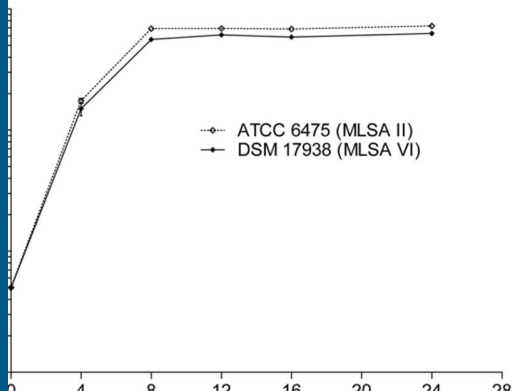

JCM1112
DSM20016
ATCC4659
ATCC6475
ATCC55730
CF48-3A
100-23
mlc3
lpuph
ATCC53608
transposases
hdcPAB-hisS2
transposases
pocR-pdu-cbi-hem-cob
integrase
hypothetical proteins
transposases
2000 Kb
200 Kb
1800 Kb
400 Kb
1600 Kb
Lactobacillus reuteri
Clade II Reference
hypothetical
proteins
arginine
catabolism
600 Kb
transposases
1400 Kb
800 Kb
1200 Kb
1000 Kb
folBKE folC2 folQP
phage
integrase
hypothetical proteins
ΦLreu2
transposases
ΦLreu1

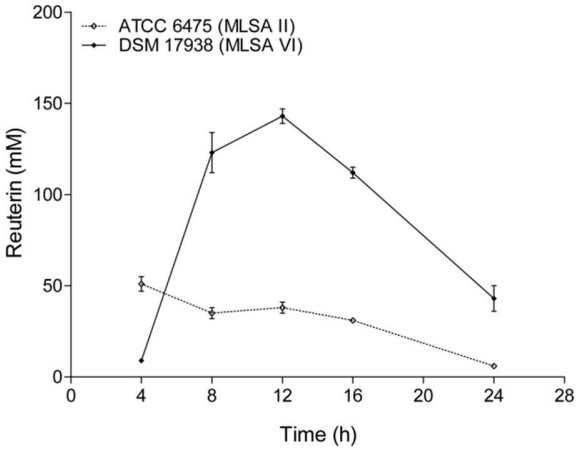
3.


Human
Pig
Rodent
Human-derived,
Intra-clade Comparisons
70
75
80
85
90
95
100
100
99
98
97
96
95
% ANI Conserved (evolutionary relatedness)


MLSA II
MLSA VI
MLSA II vs MLSA VI
MLSA VI vs MLSA VI
MLSA VI vs MLSA II
MLSA II vs MLSA II
70
75
80
85
90
95
100
100
99
98
97
96
95
% ANI Conserved (evolutionary relatedness)







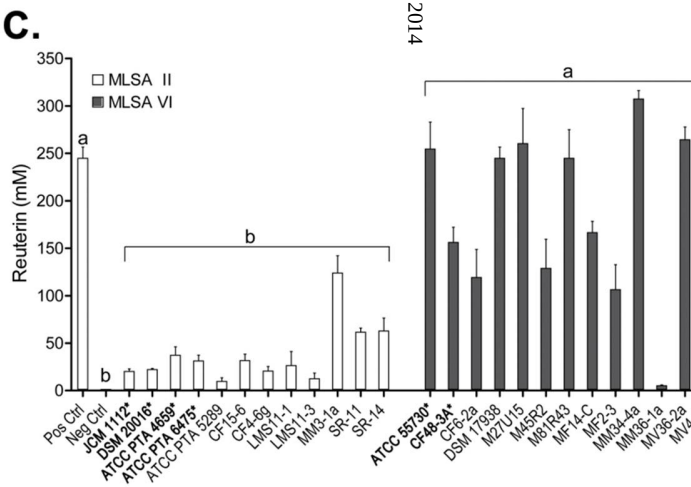
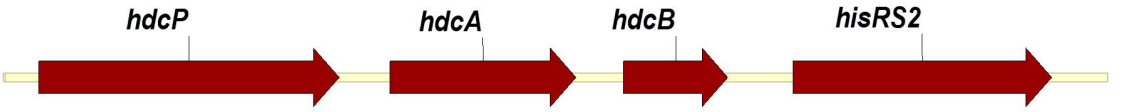


Fig. 5

Human-derived *Lactobacillus reuteri*		Cytokines Assayed by Luminex											
		TNF-α	MCP-1	IL-8 (CXCL8)	IL-7	IL-13	IL-5	GM-CSF	IL-6	IL-12(p70)	IL-10	IFN-γ	IL-1β
Control	LDM Media	0.649	0.653	0.552	0.444	0.291	1.000	0.706	0.297	0.694	0.157	0.352	0.287
MLSA Clade II	DSM 20016	0.148	0.121	0.337	0.267	0.000	0.937	0.430	0.532	0.111	0.836	0.696	0.314
	ATCC PTA 4659	0.083	0.157	0.412	0.385	0.939	0.103	0.610	0.450	0.902	0.389	0.432	0.223
	ATCC PTA 5289	0.158	0.134	0.625	0.348	0.032	0.983	0.648	0.608	0.124	0.431	0.679	0.213
	ATCC PTA 6475	0.144	0.270	0.429	0.250	0.449	1.000	0.575	0.546	0.888	0.546	0.080	0.971
	CF15-6	0.135	0.262	0.347	0.304	0.044	0.925	0.401	0.408	0.123	0.089	0.699	0.133
	CF4-6g	0.143	0.131	0.623	0.330	0.047	0.983	0.607	0.661	0.117	0.526	0.707	0.213
	JCM 1112	0.048	0.303	0.222	0.095	0.361	0.914	0.327	0.145	0.873	0.177	0.000	0.244
	LMS11-1	0.125	0.192	0.593	0.413	0.042	0.937	0.708	1.000	0.130	1.000	0.772	0.251
	LMS11-3	0.117	0.164	0.569	0.312	0.032	0.925	0.653	0.807	0.111	0.815	0.713	0.227
	MM3-1a	0.135	0.302	0.386	0.278	0.026	0.943	0.506	0.403	0.103	0.107	0.735	0.153
	SR-11	0.612	0.102	0.719	0.623	0.152	0.943	0.832	0.486	0.155	0.318	1.000	0.243
	SR-14	0.506	0.125	0.588	0.646	0.337	0.695	0.774	0.422	0.516	0.284	0.923	0.039
MLSA Clade VI	ATCC 55730	0.703	0.669	0.626	0.530	0.487	1.000	0.760	0.302	0.856	0.221	0.382	0.385
	CF48-3A	0.916	1.000	0.625	0.674	0.361	0.948	0.903	0.317	0.920	0.212	0.730	0.699
	CF6-2a	0.475	0.514	0.495	0.670	0.947	0.057	0.544	0.290	0.896	0.324	0.579	0.214
	DSM 17938	0.711	0.714	0.627	0.717	0.938	0.040	0.625	0.372	0.906	0.335	0.528	0.410
	M27U15	0.572	0.173	0.601	0.826	0.912	0.000	0.749	0.336	1.000	0.371	0.673	0.028
	M45R2	0.584	0.151	0.577	0.754	0.913	0.057	0.714	0.297	0.939	0.336	0.620	0.008
	M81R43	0.823	0.942	0.683	0.905	0.932	0.017	0.940	0.408	0.897	0.376	0.890	0.450
	MF14-C	0.739	0.801	0.487	0.717	0.903	0.040	0.568	0.259	0.891	0.389	0.589	0.278
	MF2-3	0.801	0.149	1.000	1.000	1.000	0.132	1.000	0.357	0.944	0.372	0.877	0.181
	MM34-4a	0.534	0.440	0.607	0.646	0.947	0.052	0.583	0.293	0.889	0.312	0.487	0.205
	MM36-1a	0.588	0.162	0.635	0.769	0.912	0.057	0.815	0.331	0.935	0.364	0.656	0.025
	MV36-2a	0.609	0.683	0.460	0.767	0.937	0.000	0.590	0.336	0.930	0.306	0.746	0.425
	MV4-1a	0.470	0.256	0.775	0.749	0.917	0.029	0.845	0.390	0.891	0.370	0.628	0.191

ig. 6

A.



B.

Human-derived *Lactobacillus reuteri*		Histamine*	Presence of Histidine Decarboxylase Related Genes†			
			hdcP	*hdcA*	*hdcB*	*hisS2*
MLSA Clade II	DSM 20016	0.625	√	√	√	√
	ATCC PTA 4659	0.971	√	√	√	√
	ATCC PTA 5289	0.932	√	√	√	√
	ATCC PTA 6475	1.000	√	√	√	√
	CF15-6	0.469	√	√	√	√
	CF4-6g	0.844	√	√	√	√
	JCM 1112	0.544	√	√	√	√
	LMS11-1	0.612	√	√	√	√
	LMS11-3	0.487	√	√	√	√
	MM3-1a	0.660	√	√	√	√
	SR-11	0.191	√	-	-	√
	SR-14	0.008	√	-	-	√
MLSA Clade VI	ATCC 55730	0.000	-	-	-	-
	CF48-3A	0.003	-	-	-	-
	CF6-2a	0.001	-	-	-	-
	DSM 17938	0.000	-	-	-	-
	M27U15	0.004	-	-	-	-
	M45R2	0.445	√	√	√	√
	M81R43	0.000	-	-	-	√
	MF14-C	0.000	-	-	-	-
	MF2-3	0.192	-	-	-	-
	MM34-4a	0.001	-	-	-	-
	MM36-1a	0.487	√	√	√	√
	MV36-2a	0.000	-	-	-	√
	MV4-1a	0.005	-	-	-	-

Fig. 7

Fig. S1

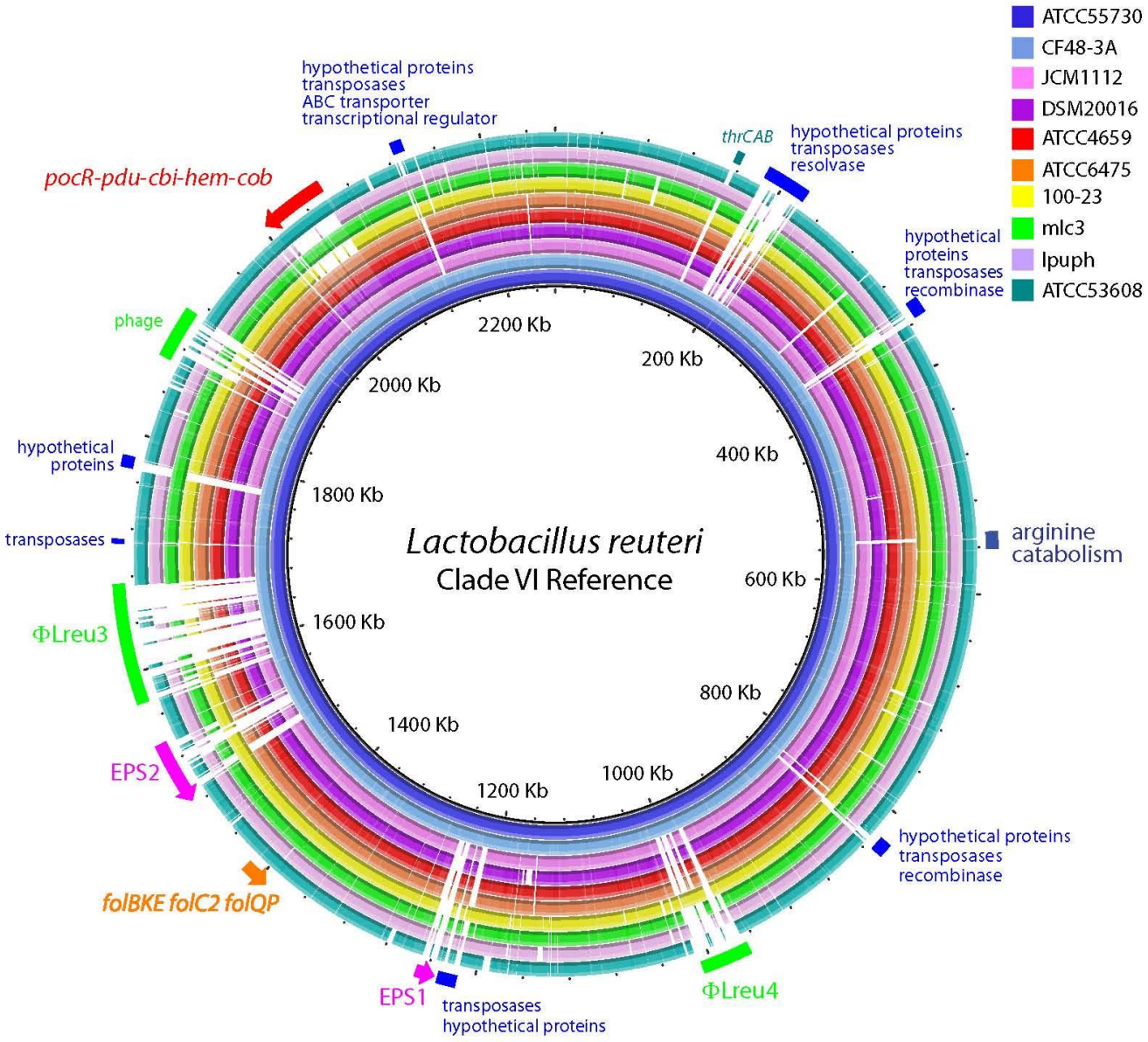

ATCC55730
CF48-3A
JCM1112
DSM20016
ATCC4659
ATCC6475
100-23
mlc3
lpuph
ATCC53608
hypothetical proteins
transposases
ABC transporter
transcriptional regulator
pocR-pdu-cbi-hem-cob
thrCAB
hypothetical proteins
transposases
resolvase
hypothetical
proteins
transposases
recombinase
phage
hypothetical
proteins
transposases
ΦLreu3
EPS2
folBKE folC2 folQP
EPS1
transposases
hypothetical proteins
ΦLreu4
hypothetical proteins
transposases
recombinase
arginine
catabolism
2200 Kb
200 Kb
2000 Kb
400 Kb
1800 Kb
600 Kb
1600 Kb
800 Kb
1400 Kb
1000 Kb
1200 Kb
Lactobacillus reuteri
Clade VI Reference

Fig. S2

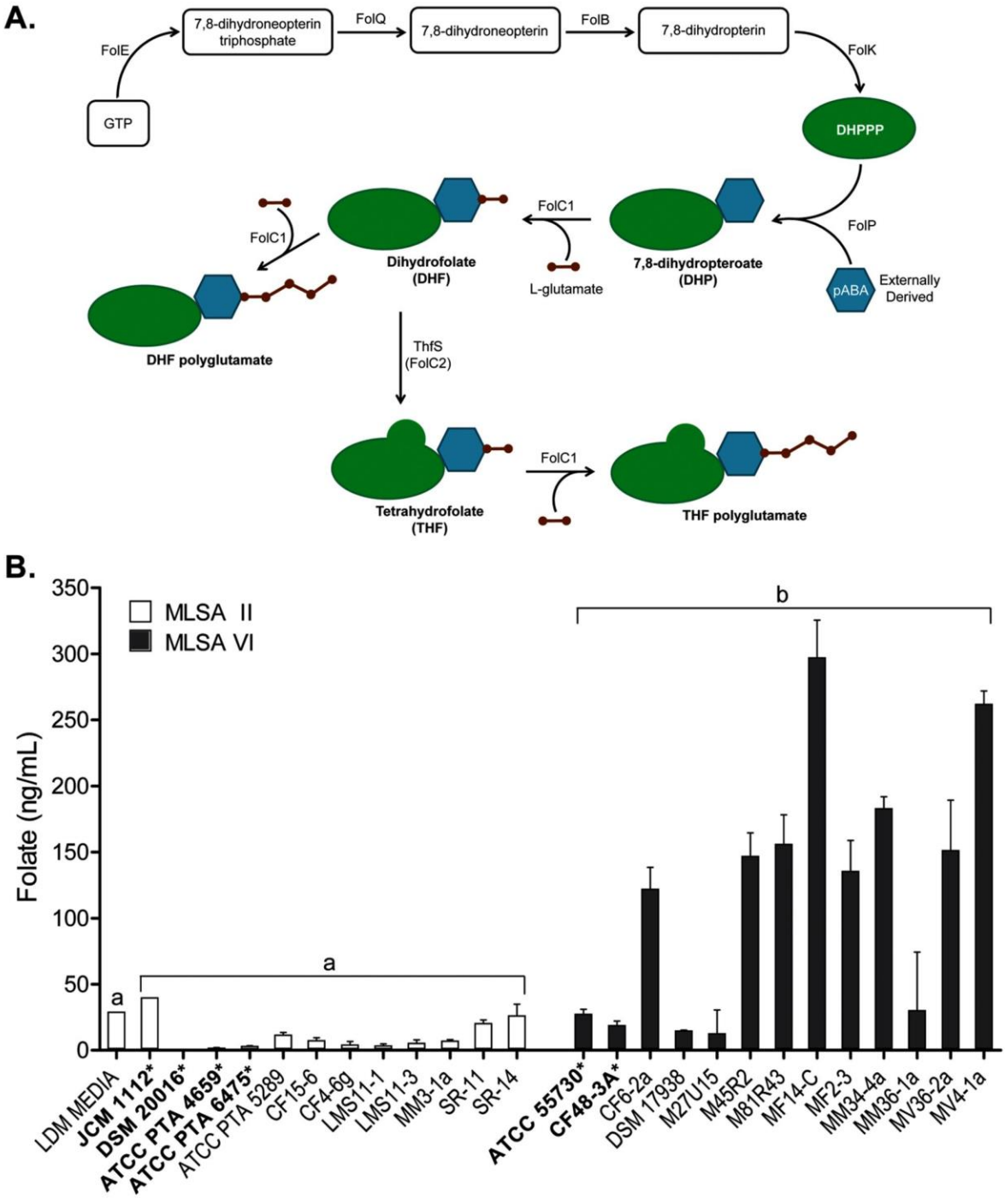
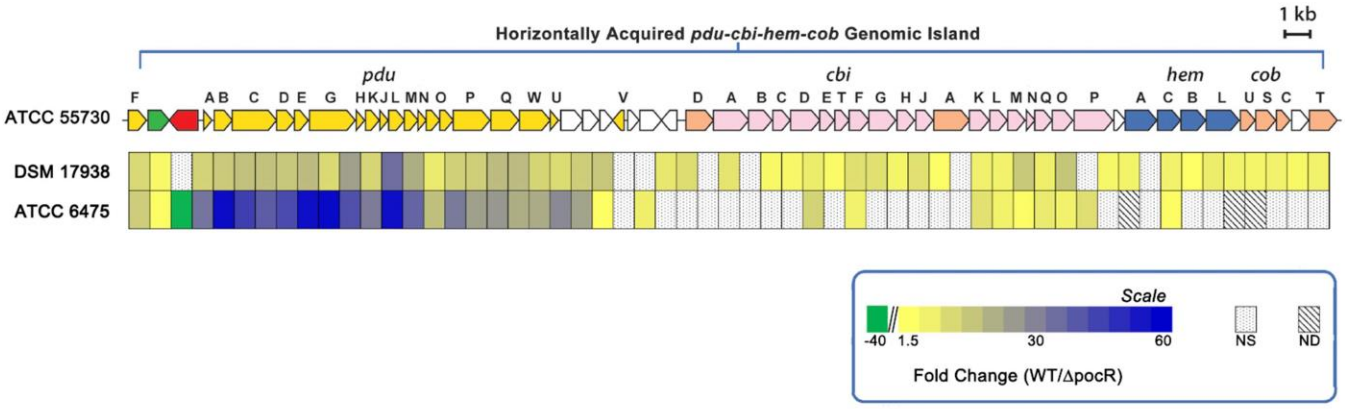

A.
GTP
FolE
7,8-dihydroneopterin triphosphate
FolQ
7,8-dihydroneopterin
FolB
7,8-dihydropterin
FolK
DHPPP
FolP
pABA
Externally Derived
7,8-dihydropteroate (DHP)
FolC1
L-glutamate
Dihydrofolate (DHF)
FolC1
DHF polyglutamate
ThfS (FolC2)
Tetrahydrofolate (THF)
FolC1
THF polyglutamate
B.
MLSA II
MLSA VI
Folate (ng/mL)
350
300
250
200
150
100
50
0
a
a
b
LDM MEDIA
JCM 1112*
DSM 20016*
ATCC PTA 4659*
ATCC PTA 6475*
ATCC PTA 5289
CF15-6
CF4-6g
LMS11-1
LMS11-3
MM3-1a
SR-11
SR-14
ATCC 55730*
CF48-3A*
CF6-2a
DSM 17938
M27U15
M45R2
M81R43
MF14-C
MF2-3
MM34-4a
MM36-1a
MV36-2a
MV4-1a

Fig. S3


1 kb
Horizontally Acquired pdu-cbi-hem-cob Genomic Island
pdu
cbi
hem
cob
F AB C DE G HKJLMNO P Q WU V D A BCDETFG HJ A KLMNQO P A CBL USC T
ATCC 55730
DSM 17938
ATCC 6475
Scale
-40 1.5 30 60
NS
ND
Fold Change (WT/ΔpocR)

Supplementary Table S1. Primers used in this study.

Primer Name	Sequence (5'-3')	Target Gene	Reference
RB1883F2 5'BHI	TGACGGATCCTAAGTGTGACTGGCACTGTTGTAGATTGTAATACGC	*pocR* (DSM 17938)	This Study
RB1883R2 3'ERI	TGACGATTCGTCAACATCAATATGTGAGGCATCAAG	*pocR* (DSM 17938)	This Study
hdcP F3	TGGACATTGTGCATATTCCTATTCC	*hdcP*	This Study
hdcP R	GCGCTTCCTATTCCTTACAATCC	*hdcP*	This Study
hdcA F	GGATTGTAAGGAATAGGAAGCGC	*hdcA*	This Study
hdcA R	GCTCCACATCCACCTACCTTA	*hdcA*	This Study
hdcB F	GCTGTTTGTGTACCACGATGTA	*hdcB*	This Study
hdcB R	CCTAAGGTAAAGTGTTCCACCG	*hdcB*	This Study
hisS F	ATTTCGTGATGGGGAGCGTT	*hisS2*	This Study
hisS R	TGCGGATAATTGAAGCCCATTA	*hisS2*	This Study

Supplementary Table S2a. Prophage ϕLreu1 Gene Annotations[a]

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
878463	879728	-	421	integrase/recombinase
880043	880195	+	50	transposase
880196	881053	-	285	hypothetical protein
881074	881595	-	173	hypothetical protein
881609	882013	-	134	conserved hypothetical *Lactobacillus* phage protein
882067	882471	-	134	conserved hypothetical protein
882488	882823	-	111	transcriptional regulator/phage repressor
882984	883199	+	71	transcriptional regulator/phage repressor
883218	883991	+	257	phage antirepressor
884003	884131	+	42	hypothetical protein
884131	884253	+	40	hypothetical protein
884267	884470	+	67	hypothetical protein
884467	884715	+	82	possible transcriptional regulator
884693	884896	+	67	hypothetical protein
884911	885087	+	58	hypothetical protein
885087	885359	+	90	transcriptional regulator
885352	886281	+	309	RecT family recombinase
886265	887089	+	274	conserved hypothetical protein
887094	887951	+	285	phage replication protein
887944	888273	+	109	hypothetical protein
888270	888641	+	123	hypothetical protein
888638	888730	+	30	hypothetical protein
888717	888995	+	92	hypothetical protein
888992	889180	+	62	hypothetical protein
889180	889332	+	50	hypothetical protein
889376	889570	+	64	hypothetical protein
889570	889830	+	86	hypothetical protein
889830	890171	+	113	hypothetical protein
890171	890359	+	62	hypothetical protein
890367	890627	+	86	hypothetical protein
890627	890938	+	103	hypothetical protein
891001	891144	+	47	hypothetical protein
891226	891648	+	140	possible phage recombinase/resolvase
891661	891945	+	94	conserved hypothetical protein
892596	892802	+	68	hypothetical protein
892890	893558	+	222	conserved hypothetical protein
893558	893983	+	141	conserved hypothetical protein
893997	894776	+	259	possible ParB family nuclease
894793	895296	+	167	hypothetical protein
895271	896566	+	431	phage terminase, large subunit
896566	898227	+	553	phage portal protein
898227	899177	+	316	phage head morphogenesis protein

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
899188	899433	+	81	hypothetical protein
899563	900216	+	217	phage scaffold protein
900231	901307	+	358	phage capsid protein
901320	901685	+	121	conserved hypothetical *Lactobacillus* phage protein
901685	901999	+	104	conserved hypothetical *Lactobacillus* phage protein
901989	902549	+	186	phage head-tail joining protein
902558	902968	+	136	conserved hypothetical *Lactobacillus* phage protein
902971	903618	+	215	phage tail protein
903638	904183	+	181	conserved hypothetical *Lactobacillus* phage protein
904276	904455	+	59	hypothetical protein
904459	908109	+	1216	phage tail tape measure protein
908106	908993	+	295	conserved hypothetical phage protein
908990	911221	+	743	M23 family peptidase/possible phage lysin
911205	914786	+	1193	conserved hypothetical protein
914779	915051	+	90	hypothetical protein
915097	915366	+	89	conserved hypothetical *Lactobacillus* phage protein
915379	916524	+	381	probable lipase
916527	917636	+	369	conserved hypothetical *Lactobacillus* phage protein
917648	918052	+	134	hypothetical protein
918045	918164	+	39	hypothetical protein
918203	918754	+	183	hypothetical protein
918793	919962	+	389	hypothetical protein
919977	920426	+	149	phage holin protein
920416	921615	+	399	N-acetylmuramoyl-L-alanine amidase
921687	921854	-	55	conserved hypothetical protein
921867	923396	-	509	glucokinase
923624	924157	+	177	GNAT family acetyltransferase
924421	924525	-	34	hypothetical protein
924665	924874	-	69	*Lactobacillus* conserved hypothetical protein
925668	927581	+	637	amidase
927663	927938	+	91	*Lactobacillus* conserved hypothetical protein
927999	928490	-	163	conserved hypothetical *Lactobacillus* phage protein
928800	930332	+	510	type I restriction modification system DNA methyltransferase subunit HsdM
930322	930909	+	195	type I restriction modification system DNA specificity subunit HsdS
930830	931465	-	211	type I restriction modification system DNA specificity subunit HsdS
931508	932482	+	324	integrase

[a]Coordinates are based on the nucleotide sequence of JCM 1112 (GenBank NC_01609.1).

Supplementary Table S2b. Prophage ϕLreu2 Gene Annotations[a]

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
1239954	1241084	+	376	phage integrase
1239634	1239951	+	105	*Lactobacillus* conserved hypothetical protein
1238962	1239471	+	169	conserved hypothetical protein
1238393	1238833	+	146	conserved hypothetical protein
1237943	1238380	+	145	phage repressor
1237561	1237764	-	67	phage repressor
1236739	1237512	-	257	phage antirepressor
1236449	1236727	-	92	*Lactobacillus* conserved hypothetical protein
1236326	1236448	-	40	hypothetical protein
1236183	1236314	-	43	hypothetical protein
1235663	1236169	+	168	hypothetical protein
1235425	1235610	-	61	hypothetical protein
1235159	1235389	-	76	hypothetical protein
1234177	1235166	-	329	phage RecT family recombinase
1233369	1234193	-	274	conserved hypothetical protein
1232507	1233364	-	285	phage replication protein
1232185	1232514	-	109	hypothetical protein
1231817	1232188	-	123	hypothetical protein
1231728	1231820	-	30	hypothetical protein
1231463	1231741	-	92	hypothetical protein
1231278	1231466	-	62	hypothetical protein
1231126	1231278	-	50	hypothetical protein
1230888	1231082	-	64	hypothetical protein
1230628	1230888	-	86	hypothetical protein
1230287	1230628	-	113	hypothetical protein
1230099	1230287	-	62	hypothetical protein
1229831	1230091	-	86	hypothetical protein
1229520	1229831	-	103	hypothetical protein
1229314	1229457	-	47	hypothetical protein
1228810	1229232	-	140	*Lactobacillus* conserved hypothetical protein
1228513	1228797	-	94	hypothetical protein
1227062	1227904	-	280	hypothetical protein
1226331	1226789	-	152	IS*200* family transposase
1226026	1226172	-	48	Lactobacillus conserved hypothetical protein
1225492	1225671	-	59	Lactobacillus conserved hypothetical protein
1224888	1225424	-	178	phage endonuclease
1224270	1224746	-	158	phage terminase small subunit protein
1223810	1224247	+	145	hypothetical protein
1223640	1223813	+	57	hypothetical protein
1221685	1223574	-	629	phage terminase large subunit protein
1220308	1221498	-	396	phage portal protein
1219590	1220321	-	243	S14 family ClpP protease

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
1218425	1219600	-	391	phage major capsid protein
1218037	1218405	-	122	phage DNA packaging protein
1217718	1218068	-	116	phage head-tail joining/adaptor protein
1217300	1217716	-	138	phage head-tail joining protein
1216920	1217303	-	127	phage tail protein
1216205	1216915	-	236	phage tail protein
1215757	1216143	-	128	*Lactobacillus* conserved hypothetical protein
1215533	1215682	-	49	hypothetical protein
1211701	1215531	-	1276	phage tail tape measure protein
1210841	1211686	-	281	phage tail protein
1209082	1210827	-	581	glycoside hydrolase/peptidase/lysin
1205055	1209134	-	1359	hypothetical protein
1204820	1205062	-	80	hypothetical protein
1204508	1204777	-	89	conserved hypothetical *Lactobacillus* phage protein
1203350	1204495	-	381	conserved hypothetical *Lactobacillus* phage protein
1202238	1203347	-	369	conserved hypothetical *Lactobacillus* phage protein
1201822	1202226	-	134	hypothetical protein
1201710	1201829	-	39	hypothetical protein
1201120	1201671	-	183	hypothetical protein
1199993	1201081	-	362	hypothetical protein
1199652	1199978	-	108	hypothetical protein
1199284	1199655	-	123	hypothetical protein
1198095	1199294	-	399	N-acetylmuramoyl-L-alanine amidase
1197209	1197328	+	39	hypothetical protein
1196399	1196857	-	152	IS*200* family transposase

[a]Coordinates are based on the nucleotide sequence of JCM 1112 (GenBank NC_01609.1).

Supplementary Table S2c. Prophage ϕLreu3 Gene Annotations[a]

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
462928	463611	+	227	integrase
463616	464374	+	252	ISChy4 transposase
464301	464597	-	98	hypothetical protein HMPREF0538_20464
464600	465334	-	244	type II restriction-modification system modification subunit
465378	465590	-	70	hypothetical protein HMPREF0538_20466
465619	468228	-	869	hypothetical protein HMPREF0538_20467
468234	469532	-	432	hypothetical protein HMPREF0538_20468
469910	470368	+	152	hypothetical protein HMPREF0538_20469
470451	470654	+	67	hypothetical protein HMPREF0538_20470
470638	470955	+	105	hypothetical protein HMPREF0538_20471
470952	472091	+	379	*Lactobacillus* conserved phage protein
472069	472644	+	191	phage-associated protein
472700	474634	+	644	DNA polymerase
474701	475105	+	134	hypothetical protein HMPREF0538_20475
475108	477363	+	751	P4 family phage protein
477411	477857	+	148	VRR-NUC domain protein
477838	479193	+	451	SNF2 domain protein
479190	479657	+	155	probable restriction endonuclease
479806	480183	+	125	HNH endonuclease domain protein
480303	480845	+	180	hypothetical protein HMPREF0538_20481
480845	482071	+	408	DNA (cytosine-5-)-methyltransferase
482144	482764	+	206	hypothetical protein HMPREF0538_20483
482767	482973	+	68	hypothetical protein HMPREF0538_20484
483104	484072	+	322	transposase
484096	485715	+	539	prophage protein
485744	487009	+	421	HK97 family portal protein
487006	487677	+	223	S14 family peptidase ClpP
487696	488874	+	392	HK97 family phage protein
488891	489169	+	92	hypothetical protein HMPREF0538_20490
489169	489552	+	127	bacteriophage head-tail adaptor
489539	489952	+	137	holin
489949	490098	+	49	hypothetical protein HMPREF0538_20493
490120	490239	+	39	hypothetical protein HMPREF0538_20494
490306	490827	+	173	lysozyme
491094	491327	+	77	hypothetical protein HMPREF0538_20496
491363	493024	+	553	integrase/recombinase
493017	494594	+	525	Integrase/recombinase
494764	495615	+	283	ABC transporter ATP-binding protein
495578	496468	+	296	ABC transporter
496471	496923	+	150	LytR/AlgR family transcriptional regulator
496925	497311	+	128	hypothetical protein HMPREF0538_20502

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
497439	497561	-	40	hypothetical protein HMPREF0538_20503
497928	498164	-	78	hypothetical protein HMPREF0538_20504
498189	501305	-	1038	type I site-specific deoxyribonuclease
501360	502517	-	385	type I restriction/modification specificity protein
502514	503482	-	322	integrase/recombinase
503550	504209	+	219	type I restriction-modification system specificity subunit
504206	504736	-	176	hypothetical protein HMPREF0538_20509
504726	506264	-	512	type I restriction-modification system DNA-methyltransferase
506261	506482	-	73	hypothetical protein HMPREF0538_20511
506494	509106	-	870	hypothetical protein HMPREF0538_20512
509111	510496	-	461	hypothetical protein HMPREF0538_20513
510807	511325	+	172	sigma-70 family protein
511407	511628	+	73	hypothetical protein HMPREF0538_20515
511612	511947	+	111	hypothetical protein HMPREF0538_20516
511944	513080	+	378	phage protein
513061	513636	+	191	hypothetical protein HMPREF0538_20518
513692	515626	+	644	DNA-directed DNA polymerase
515713	516099	+	128	hypothetical protein HMPREF0538_20520
516102	518357	+	751	P4 family prophage protein
518405	518851	+	148	VRR-NUC domain protein
518832	520187	+	451	SNF2 domain protein
520184	520651	+	155	probable restriction endonuclease
520799	521176	+	125	HNH endonuclease domain protein
521299	521841	+	180	hypothetical protein HMPREF0538_20526
521841	523070	+	409	DNA (cytosine-5-)-methyltransferase
523144	523776	+	210	hypothetical protein HMPREF0538_20528
523769	523975	+	68	hypothetical protein HMPREF0538_20529
524041	525642	+	533	prophage protein
525670	525831	+	53	hypothetical protein HMPREF0538_20531
525972	526133	-	53	lipoprotein
526190	527446	+	418	HK97 family portal protein
527443	528105	+	220	S14 family peptidase ClpP
528126	529304	+	392	HK97 family major capsid protein
529317	529595	+	92	hypothetical protein HMPREF0538_20536
529596	529976	+	126	bacteriophage head-tail adaptor
529966	530388	+	140	holin
530486	530674	+	62	hypothetical protein HMPREF0538_20539
530736	532289	+	517	phage integrase/recombinase
532276	532680	+	134	phage integrase/recombinase
532667	534253	+	528	phage integrase/recombinase

[a]Coordinates are based on the nucleotide sequence of ATCC 55730 (GenBank NC_015697).

Supplementary Table S2d. Prophage ϕLreu4 Gene Annotations[a]

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
2099733	2099939	+	68	hypothetical protein HMPREF0538_22063
2099951	2100718	+	255	phage antirepressor
2100730	2100858	+	42	hypothetical protein HMPREF0538_22065
2100858	2100980	+	40	hypothetical protein HMPREF0538_22066
2100995	2101201	+	68	DNA-binding protein
2101198	2101377	+	59	hypothetical protein HMPREF0538_22068
2101413	2101643	+	76	hypothetical protein HMPREF0538_22069
2101636	2101908	+	90	6-phospho-beta-glucosidase
2101912	2102664	+	250	Erf family protein
2102667	2103347	+	226	hypothetical protein HMPREF0538_22072
2103337	2103765	+	142	single-strand binding protein
2103778	2104608	+	276	phage protein
2104624	2105421	+	265	DNA replication protein
2105424	2105576	+	50	chaperone DnaJ
2105580	2105840	+	86	hypothetical protein HMPREF0538_22077
2105830	2106351	+	173	hypothetical protein HMPREF0538_22078
2106364	2106921	+	185	hypothetical protein HMPREF0538_22079
2106998	2107447	+	149	hypothetical protein HMPREF0538_22080
2107440	2107718	+	92	hypothetical protein HMPREF0538_22081
2107715	2108086	+	123	oxygen-independent coproporphyrinogen III oxidase
2108101	2109231	-	376	IS*30* family transposase
2109293	2109490	+	65	XRE family transcriptional regulator
2109480	2109722	+	80	hypothetical protein HMPREF0538_22085
2109691	2110077	+	128	hypothetical protein HMPREF0538_22086
2110190	2110384	+	64	hypothetical protein HMPREF0538_22087
2110488	2110631	+	47	hypothetical protein HMPREF0538_22088
2110663	2110854	+	63	hypothetical protein HMPREF0538_22089
2110964	2111479	+	171	hypothetical protein HMPREF0538_22090
2112304	2112783	+	159	XRE family transcriptional regulator
2112773	2113207	+	144	hypothetical protein HMPREF0538_22092
2113753	2113932	+	59	probable phage-associated protein
2113925	2114203	+	92	hypothetical protein HMPREF0538_22094
2114447	2114617	+	56	hypothetical protein HMPREF0538_22095
2114643	2114882	-	79	hypothetical protein HMPREF0538_22096
2114838	2115695	+	285	HNH endonuclease domain protein
2115879	2116346	+	155	phage terminase small subunit
2116343	2118052	+	569	phage terminase large subunit
2118033	2118395	+	120	hypothetical protein HMPREF0538_22100
2118558	2119817	+	419	phage portal protein
2119756	2120409	+	217	phage head maturation protease
2120410	2121546	+	378	phage protein

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
2121688	2122101	+	137	phage protein
2122004	2122357	+	117	hypothetical protein HMPREF0538_22105
2122341	2122715	+	124	phage protein
2122712	2123131	+	139	phage protein
2123131	2123760	+	209	phage major tail protein
2123813	2124154	+	113	glutamyl-tRNA synthetase
2124163	2124333	+	56	hypothetical protein HMPREF0538_22110
2124350	2128852	+	1500	TP901 family phage tail tape measure protein
2128856	2129698	+	280	hypothetical protein HMPREF0538_22112
2129773	2132241	+	822	phage lysin
2132254	2132616	+	120	hypothetical protein HMPREF0538_22114
2132616	2134490	+	624	phage minor head protein
2134502	2135287	+	261	hypothetical protein HMPREF0538_22116
2135302	2135700	+	132	hypothetical protein HMPREF0538_22117
2135944	2136324	+	126	hypothetical protein HMPREF0538_22118
2136485	2136961	+	158	hypothetical protein HMPREF0538_22119
2136974	2137900	+	308	endolysin
2138047	2138205	+	52	hypothetical protein HMPREF0538_22121

[a]Coordinates are based on the nucleotide sequence of ATCC 55730 (GenBank NC_015697).

Supplementary Table S3. Arginine Catabolism Gene Cluster Annotations[a]

New Annotation Coordinates				
Gene Start	**Gene Stop**	**Strand**	**Protein (aa)**	**Protein Function Prediction**
484793	485251	-	152	IS*200* family transposase
485472	486311	-	279	transcriptional regulator
486564	487571	+	335	ornithine carbamoyltransferase ArcB
487588	488520	+	310	carbamate kinase ArcC
488564	489478	-	304	DMT drug/metabolite transporter
489562	490143	-	193	M10 family zinc metallopeptidase
490419	490787	+	122	probable secreted cupredoxin domain protein
490787	491074	+	95	probable cupredoxin domain protein
491074	493011	+	645	P-type ATPase superfamily copper transporter
493130	493606	+	158	*Lactobacillus* conserved hypothetical protein
493835	494164	-	109	*Lactobacillus* conserved hypothetical protein
494395	495783	+	462	CPA1 family monovalent Na+/H+ antiporter
496429	496611	-	60	conserved hypothetical protein
496613	496789	-	58	hypothetical protein
497061	497399	-	112	transposase
497426	498307	-	293	transposase
499483	499848	-	121	hypothetical protein
499845	499976	-	43	hypothetical protein
500400	500801	-	133	*Lactobacillus* conserved hypothetical protein
500818	501219	-	133	hypothetical protein
501268	502227	-	319	hypothetical protein
502244	502468	-	74	hypothetical protein
502468	502938	-	156	hypothetical protein
503095	503859	-	254	possible ribonuclease H
503856	504455	-	199	conserved hypothetical protein
504681	505274	-	197	hypothetical protein
505465	506484	+	339	beta-lactamase/penicillin binding protein
506692	507924	+	410	arginine deiminase ArcA
508035	508496	+	153	ArgR family transcriptional regulator
508517	509938	+	473	APC family amino acid polyamine organocation transporter
509996	511393	+	466	APC family amino acid polyamine organocation transporter

[a]Coordinates are based on the nucleotide sequence of JCM 1112 (GenBank NC_01609.1).

Supplementary Table S4. Folate biosynthesis gene annotations[*]

Gene	Protein Function Prediction	Strand	Protein (aa)	ATCC 55730[a]		JCM 1112[b]	
				Start	Stop	Start	Stop
folC1	folypolyglutamate synthase	+	437	1764152	1765465	578388	579701
folA	dihydrofolate reductase	+	162	2068933	2069421	853517	854005
folP	dihydropteroate synthase	-	387	277203	278366	1364163	1364975
folQ	dihydroneopterin triphosphate diphosphatase	-	195	278368	278955	1365328	1365915
thfS (*folC2*)	tetrahydrofolate synthase	-	419	278945	280204	1365905	1367164
folE	GTP cyclohydrolase I	-	192	280191	280769	1367151	1367729
folK	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase	-	170	280751	281263	1367711	1368223
folB	dihydroneopterin aldolase	-	111	281266	281601	1368226	1368561

[*]Coordinates are based on the nucleotide sequence of [a]ATCC 55730 (GenBank NC_015697) and [b]JCM 1112 (GenBank NC_01609.1).

Supplementary Table S5a. Clade VI *pdu-cbi-hem-cob* gene cluster annotations[a]

Start	Stop	Strand	Protein (aa)	Gene	Protein Function Prediction
919827	920534	-	235	*pduF*	MIP family glycerol uptake facilitator protein
918921	919754	-	277	*eutJ*	ethanolamine utilization protein
917746	918840	+	364	*pocR*	AraC-family transcriptional regulator
917212	917490	-	92	*pduA*	carbon dioxide concentrating mechanisms, carboxysome shell protein
916398	917114	-	238	*pduB*	propanediol utilization protein
914697	916373	-	558	*pduC*	glycerol dehydratase, large subunit
913969	914679	-	236	*pduD*	glycerol dehydratase, medium subunit
913440	913955	-	171	*pduE*	glycerol dehydratase, small subunit
911565	913412	-	615	*pduG*	propanediol dehydratase reactivation protein
911219	911599	-	126	*pduH*	propanediol dehydratase reactivation protein
910642	911211	-	189	*pduK*	propanediol utilization protein
910342	910629	-	95	*pduJ*	carbon dioxide concentrating mechanisms, carboxysome shell protein
909675	910319	-	214	*pduL*	propanediol utilization protein
909141	909644	-	167	*pduM*	propanediol utilization protein
908881	909153	-	90	*pduN*	carbon dioxide concentrating mechanisms, carboxysome shell protein
908273	908860	-	195	*pduO*	propanediol utilization protein
907797	908270	-	157	$pduO_{bis}$	propanediol utilization protein
906367	907794	-	475	*pduP*	CoA-dependent propionaldehyde dehydrogenase
905228	906367	-	379	*pduQ*	propanediol dehydrogenase
904017	905204	-	395	*pduW*	acetate kinase
903654	904001	-	115	*pduU*	propanediol utilization protein
902794	903588	-	264		conserved hypothetical protein
902014	902646	-	210		phosphoglycerate mutase
901457	902017	-	186		hypothetical protein
901036	901464	+	142	*pduV*	propanediol utilization protein
900479	900928	-	149		flavodoxin
899616	900464	-	282		flavoprotein
898973	899557	+	194		putative ATP:cob(I)alamin adenosyltransferase
897545	898633	-	362	*cobD*	L-threonine-O-3-phosphate decarboxylase
896184	897548	-	454	*cbiA*	cobyrinic acid A, C-diamide synthase
895228	896187	-	319	*cbiB*	cobalamin biosynthesis protein
894536	895222	-	228	*cbiC*	precorrin-8X methylmutase
893407	894564	-	385	*cbiD*	cobalamin biosynthesis protein
892811	893416	-	201	*cbiE*	precorrin-6Y C5,15-methyltransferase

Start	Stop	Strand	Protein (aa)	Gene	Protein Function Prediction
892264	892818	-	184	*cbiT*	precorrin-8W decarboxylase
891488	892258	-	256	*cbiF*	precorrin-4 C11-methyltransferase
890430	891485	-	351	*cbiG*	cobalamin biosynthesis protein
889693	890418	-	241	*cbiH*	precorrin-3B C17-methyltransferase
888938	889696	-	252	*cbiJ*	precorrin-6X reductase
887554	888948	-	464	*cobA*	uroporphyrinogen-III C-methyltransferase
886785	887561	-	258	*cbiK*	cobalt chelatase
886081	886779	-	232	*cbiL*	precorrin-2 C20-methyltransferase
885335	886078	-	247	*cbiM*	cobalamin biosynthesis protein
885015	885338	-	107	*cbiN*	cobalt ABC transporter permease component
884312	884989	-	225	*cbiQ*	cobalt ABC transporter permease component
883493	884302	-	269	*cbiO*	cobalt ABC transporter ATP-binding protein
881886	883403	-	505	*cbiP*	adenosylcobyric acid synthase
881419	881889	-	156	*cysG*	putative siroheme synthase
880151	881416	-	421	*hemA*	glutamyl-tRNA reductase
879244	880161	-	305	*hemC*	porphobilinogen deaminase
878267	879238	-	323	*hemB*	delta-aminolevulinic acid dehydratase
876952	878283	-	443	*hemL*	glutamate-1-semialdehyde 2,1-aminotransferase
876300	876890	-	196	*cobU*	adenosylocobinamide-phosphate guanylyltransferase
875530	876291	-	253	*cobS*	cobalamin-5'-phosphate synthase
874943	875533	-	196	*cobC*	alpha-ribazole-5'-phosphate phosphatase
874219	874929	-	236		type I site-specific deoxyribonuclease
873165	874217	-	350	*cobT*	nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase

[a]Coordinates are based on the nucleotide sequence of ATCC 55730 (GenBank NC_015697).

Supplementary Table S5b. Clade II *pdu-cbi-hem-cob* gene cluster annotations[a]

Start	Stop	Strand	Protein (aa)	Gene	Protein Function Prediction
1844734	1845441	-	235	*pduF*	MIP family glycerol uptake facilitator protein
1843816	1844649	-	277	*eutJ*	ethanolamine utilization protein
1842696	1843775	+	359	*pocR*	AraC-family transcriptional regulator
1842161	1842442	-	93	*pduA*	carbon dioxide concentrating mechanisms, carboxysome shell protein
1841347	1842063	-	238	*pduB*	propanediol utilization protein
1839646	1841322	-	558	*pduC*	glycerol dehydratase, large subunit
1838918	1839628	-	236	*pduD*	glycerol dehydratase, medium subunit
1838390	1838905	-	171	*pduE*	glycerol dehydratase, small subunit
1836509	1838359	-	616	*pduG*	propanediol dehydratase reactivation protein
1836163	1836546	-	127	*pduH*	propanediol dehydratase reactivation protein
1835586	1836155	-	189	*pduK*	propanediol utilization protein
1835283	1835573	-	96	*pduJ*	carbon dioxide concentrating mechanisms, carboxysome shell protein
1834610	1835254	-	214	*pduL*	propanediol utilization protein
1834075	1834578	-	167	*pduM*	propanediol utilization protein
1833815	1834093	-	92	*pduN*	carbon dioxide concentrating mechanisms, carboxysome shell protein
1833207	1833794	-	195	*pduO*	propanediol utilization protein
1832731	1833204	-	157	$pduO_{bis}$	propanediol utilization protein
1831295	1832728	-	477	*pduP*	CoA-dependent propionaldehyde dehydrogenase
1830156	1831277	-	373	*pduQ*	propanediol dehydrogenase
1828948	1830132	-	394	*pduW*	acetate kinase
1828585	1828932	-	115	*pduU*	propanediol utilization protein
1827718	1828512	-	264		conserved hypothetical protein
1826977	1827621	-	214		phosphoglycerate mutase
1826423	1826980	-	185		hypothetical protein
1826002	1826430	+	142	*pduV*	propanediol utilization protein
1825446	1825895	-	149		flavodoxin
1824582	1825424	-	280		flavoprotein
1823968	1824534	+	188		putative ATP:cob(I)alamin adenosyltransferase
1822871	1823839	-	322		transposase
1821438	1822526	-	362	*cobD*	L-threonine-O-3-phosphate decarboxylase
1820074	1821438	-	454	*cbiA*	cobyrinic acid A, C-diamide synthase
1819118	1820077	-	319	*cbiB*	cobalamin biosynthesis protein

Start	Stop	Strand	Protein (aa)	Gene	Protein Function Prediction
1818429	1819112	-	227	*cbiC*	precorrin-8X methylmutase
1817297	1818448	-	383	*cbiD*	cobalamin biosynthesis protein
1816698	1817300	-	200	*cbiE*	precorrin-6Y C5,15-methyltransferase
1816151	1816705	-	184	*cbiT*	precorrin-8W decarboxylase
1815373	1816134	-	253	*cbiF*	precorrin-4 C11-methyltransferase
1814315	1815370	-	351	*cbiG*	cobalamin biosynthesis protein
1813577	1814302	-	241	*cbiH*	precorrin-3B C17-methyltransferase
1812822	1813580	-	252	*cbiJ*	precorrin-6X reductase
1811438	1812832	-	464	*cobA*	uroporphyrinogen-III C-methyltransferase
1810666	1811445	-	259	*cbiK*	cobalt chelatase
1809951	1810664	-	237	*cbiL*	precorrin-2 C20-methyltransferase
1809221	1809967	-	248	*cbiM*	cobalamin biosynthesis protein
1808913	1809224	-	103	*cbiN*	cobalt ABC transporter permease component
1808215	1808892	-	225	*cbiQ*	cobalt ABC transporter permease component
1807400	1808203	-	267	*cbiO*	cobalt ABC transporter ATP-binding protein
1805820	1807325	-	501	*cbiP*	adenosylcobyric acid synthase
1805347	1805805	-	152	*cysG*	putative siroheme synthase
1804080	1805345	-	421	*hemA*	glutamyl-tRNA reductase
1803173	1804090	-	305	*hemC*	porphobilinogen deaminase
1802196	1803167	-	323	*hemB*	delta-aminolevulinic acid dehydratase
1800881	1802176	-	431	*hemL*	glutamate-1-semialdehyde 2,1-aminotransferase
1800228	1800818	-	196	*cobU*	adenosylocobinamide-phosphate guanylyltransferase
1799458	1800219	-	253	*cobS*	cobalamin-5'-phosphate synthase
1798871	1799461	-	196	*cobC*	alpha-ribazole-5'-phosphate phosphatase
1798124	1798834	-	236		type I site-specific deoxyribonuclease
1797070	1798140	-	356	*cobT*	nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase

[a]Coordinates are based on the nucleotide sequence of JCM 1112 (GenBank NC_01609.1).

Supplementary Table S8. *L. reuteri* Clade II Histidine Decarboxylase Gene Cluster[a]

Start	Stop	Strand	Protein (aa)	Gene	Protein Function Prediction
1927408	1928862	-	484	*hdcP*	APC family amino acid-polyamine-organocation transporter
1926216	1927151	-	311	*hdcA*	histidine decarboxylase
1925451	1925975	-	174	*hdcB*	conserved hypothetical protein HdcB
1923821	1925122	-	433	*hisRS2*	histidyl-tRNA synthase

[a]Coordinates are based on the nucleotide sequence of JCM1112 (GenBank NC_010609).