



Unravelling lactate-acetate and sugar conversion into butyrate by intestinal *Anaerobutyricum* and *Anaerostipes* species by comparative proteogenomics

Sudarshan A. Shetty ¹, Sjeff Boeren,²
Thi P. N. Bui,^{1,3} Hauke Smidt¹ and Willem M. de Vos ^{1,4*}

¹Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands.

²Laboratory of Biochemistry, Wageningen University and Research, Wageningen, The Netherlands.

³No Caelus Pharmaceuticals, Amsterdam, The Netherlands.

⁴Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

Summary

The D- and L-forms of lactate are important fermentation metabolites produced by intestinal bacteria but are found to negatively affect mucosal barrier function and human health. Both enantiomers of lactate can be converted with acetate into the presumed beneficial butyrate by a phylogenetically related group of anaerobes, including *Anaerobutyricum* and *Anaerostipes* spp. This is a low energy yielding process with a partially unknown pathway in *Anaerobutyricum* and *Anaerostipes* spp. and hence, we sought to address this via a comparative genomics, proteomics and physiology approach. We compared growth of *Anaerobutyricum soehngenii* on lactate with that on sucrose and sorbitol. Comparative proteomics revealed complete pathway of butyrate formation from sucrose, sorbitol and lactate. Notably, a gene cluster, *lctABCDEF* was abundantly expressed when grown on lactate. This gene cluster encodes a lactate dehydrogenase (*lctD*), electron transport proteins A and B (*lctCB*), nickel-dependent racemase (*lctE*), lactate permease (*lctF*) and short-chain acyl-CoA dehydrogenase (*lctG*). Investigation of available genomes of intestinal bacteria revealed this new gene cluster to be highly conserved in only *Anaerobutyricum* and *Anaerostipes* spp. Present

study demonstrates that *A. soehngenii* and several related *Anaerobutyricum* and *Anaerostipes* spp. are highly adapted for a lifestyle involving lactate plus acetate utilization in the human intestinal tract.

Introduction

The major fermentation end products in anaerobic colonic sugar fermentations are the short chain fatty acids (SCFAs) acetate, propionate and butyrate. While all these SCFAs confer health benefits, butyrate is used to fuel colonic enterocytes and has been shown to inhibit the proliferation and to induce apoptosis of tumour cells (McMillan *et al.*, 2003; Thangaraju *et al.*, 2009; Topping and Clifton, 2001). Butyrate is also suggested to play a role in gene expression as it is a known inhibitor of histone deacetylases, and recently it was demonstrated that butyrate promotes histone crotonylation *in vitro* (Davie, 2003; Fellows *et al.*, 2018). Related to this ability, exposure to butyrate during differentiation of macrophages was demonstrated to boost antimicrobial activity (Schulthess *et al.*, 2019). In contrast, lactate is also a common end product of anaerobic fermentation in the gut but has no known health benefits. Rather lactate, and notably the D-enantiomer, has been found to be involved in acidosis, reduction of intestinal barrier function in adults and atopic eczema development in children (Ten Bruggencate *et al.*, 2006; Seheult *et al.*, 2017; Wopereis *et al.*, 2018).

Several elegant pioneering studies have shown that butyrate can also be produced by human anaerobes via the conversion of lactate and acetate (Barcenilla *et al.*, 2000; Duncan *et al.*, 2004). This later turned out to be one of major routes for butyrate formation in the gut (Louis and Flint, 2017). However, this relevant metabolic property is only found in a limited number of phylogenetically related species belonging to the genera *Anaerostipes* and *Anaerobutyricum*, the latter being previously known as *Eubacterium hallii* (Shetty *et al.*, 2018). Since the enantiomers D- and L-lactate are end products of fermentation by primary degraders such as *Bifidobacterium* and other (lactic acid) bacteria, these

Received 9 June, 2020; revised 28 September, 2020; accepted 29 September, 2020. *For correspondence. Tel +31317483108; Fax +31317483827. E-mail willem.devos@wur.nl.

are important cross-feeding metabolites in the diet driven trophic chain existing in the intestinal microbiome (Duncan *et al.*, 2004; Deis and Kearsley, 2012; Belzer *et al.*, 2017). *Anaerostipes caccae* and the two *Anaerobutyricum* species, *A. soehngenii* and *A. hallii* were described to be capable of converting D- and L-lactate plus acetate to butyrate via the acetyl CoA pathway (Barcenilla *et al.*, 2000; Duncan *et al.*, 2004; Louis and Flint, 2017). Growth on lactate poses a major energetic barrier because it is an energy-dependent process with low energy yield since the first step in conversion of lactate to pyruvate in the presence of NAD⁺/NADH is an endergonic reaction ($\Delta G^{\circ} = +25$ kJ/mol). Evidence for a molecular mechanism of conversion of lactate to acetate was first demonstrated in the acetogenic model organism *Acetobacterium woodii*, where for every molecules of lactate converted to acetate only 1.5 molecules of ATP were generated (Weghoff *et al.*, 2015). This mechanism, so-called electron confurcation, was suggested to be wide spread in other anaerobic lactate utilizers and hypothesized to be present in intestinal butyrogenic bacteria (Weghoff *et al.*, 2015; Louis and Flint, 2017; Detman *et al.*, 2019). Therefore, using a proteogenomic approach we aimed to investigate whether intestinal bacteria also employ this electron confurcation to convert lactate to butyrate with a focus on *A. soehngenii*. We grew *A. soehngenii* in the presence of three different carbon sources, lactate plus acetate, sucrose and sorbitol. Comparison of proteomic expression data revealed a complete gene cluster, expression of which was induced when *A. soehngenii* was grown in D,L-lactate plus acetate. Investigation of the gene cluster revealed it to be similar to the previously reported gene cluster in *Acetobacterium woodii* involved in the conversion of D,L-lactate to pyruvate, in which an electron transport flavoprotein (EtfAB complex) is active. Extensive search of publicly available bacterial genomes revealed that this gene cluster is highly conserved in several *Anaerobutyricum* and *Anaerostipes* spp. among the butyrate producers from the human intestinal tract. This genomic organization suggests that both *Anaerobutyricum* species and *Anaerostipes caccae* and also a recent human infant isolate *Anaerostipes rhamnosivorans* (Bui *et al.*, 2014) have adapted to a lifestyle involving efficient lactate plus acetate utilization in the human intestinal tract.

Experimental procedures

Bacterial strain and growth media

Anaerobutyricum soehngenii (DSM17630) strain L2-7 was grown routinely in a medium as described previously (Shetty *et al.*, 2018). The composition of the growth

medium was: yeast extract (4.0 g/l), casitone (2.0 g/l), soy peptone (2.0 g/l), NaHCO₃ (4.0 g/l), KH₂PO₄ (0.41 g/l), MgCl₂·6H₂O (0.1 g/l), CaCl₂·2H₂O (0.11 g/l), cysteine-HCL (0.5 g/l), 1 ml hemin (50 mg hemin, 1 ml 1 N NaOH, 99 ml dH₂O), 0.2 ml vitamin K1 solution (0.1 ml vitamin K1, 20 ml 95% EtOH) and trace elements I, trace elements II and vitamin solutions. The trace elements I (alkaline) solution contained the following (mM): 0.1 Na₂SeO₃, 0.1 Na₂WO₄, 0.1 Na₂MoO₄ and 10 NaOH. The trace elements II (acid) solution was composed of the following (mM): 7.5 FeCl₂, 1 H₃BO₄, 0.5 ZnCl₂, 0.1 CuCl₂, 0.5 MnCl₂, 0.5 CoCl₂, 0.1 NiCl₂ and 50 HCl. The vitamin solution had the following composition (g/l): 0.02 biotin, 0.2 niacin, 0.5 pyridoxine, 0.1 riboflavin, 0.2 thiamine, 0.1 cyanocobalamin, 0.1 p-aminobenzoic acid and 0.1 pantothenic acid. This basal medium was supplemented with 30 mM of sodium acetate (termed basal medium with acetate). For routine use, the medium was distributed in 35 ml serum bottles sealed with butyl-rubber stoppers and incubated at 37°C under a gas phase of 1.7 atm (172 kPa) N₂/CO₂ (80:20, v/v). The pH of the medium was adjusted to 7.0 before autoclaving. The vitamin solution was filter sterilized and added to the media bottles after autoclaving.

Global proteomic profiling of different growth substrates

Actively growing *A. soehngenii* was pre-cultured in medium with 60 mM glucose and inoculated (5%) in 450 ml of media in triplicates (biological) containing 20 mM sucrose, 40 mM D-sorbitol or 80 mM D,L-lactate respectively. In all three conditions, 30 mM of acetate was added in the medium as it is shown to improve growth of *A. soehngenii* (Duncan *et al.*, 2004; Shetty *et al.*, 2018). Cells were harvested at mid-exponential phase. The samples were centrifuged at 4700 r.p.m. for 30 min at 4°C. The cells and supernatant were stored immediately in -80°C until protein extraction. The supernatant was used to measure SCFAs as well as consumption of D,L-lactate, sucrose and sorbitol using Shimadzu Prominence-i LC-2030c high-performance liquid chromatography (HPLC). The column used was Shodex SUGAR SH1011. For the mobile phase, 0.01 N H₂SO₄ was used to which 30 mM of crotonate was added as internal standard.

For whole cell protein extraction, cells were thawed on ice and washed with reduced phosphate buffer solution (PBS with 0.2 mM Titanium (III) citrate) twice before extraction of proteins. The cell pellet was resuspended in SDS-DTT-Tris-Lysis buffer, and cells were disrupted using a French pressure cell (1500 psi). For every sample, the cells were passed through the French pressure cell three times (cell disruption checked under light microscope). After cell disruption, the samples were placed

immediately on ice. The extracted proteins were denatured by heat treatment (95°C for 30 min). Protein quantification was done using the Qubit™ Protein Assay Kit (ThermoFisher Scientific, cat.no. Q33211). A total of 50 µg of proteins plus 5 µl of loading dye was mixed and electrophoresed on 10% SDS gels (10% Mini-PROTEAN® TGX™ Precast Protein Gels) for 40 min at 20 mA. Staining was done using Coomassie Brilliant Blue R250 (2.5 g/l) (in 45% methanol and 10% glacial acetic acid) for 3 h and destained with destaining solution (25% methanol and 10% glacial acetic acid in ultrapure water) overnight.

The gels were treated for reduction and alkylation using 50 mM ammonium bicarbonate, 20 mM beta-mercaptoethanol (pH 8) and 20 mM acrylamide (pH 8). Each lane was cut into three even slices, which were further cut into small pieces of approximately 1–2 mm². Trypsin digestion was performed by adding 50 µl of sequencing grade trypsin (5 ng/µl in 50 mM ammonium bicarbonate) and incubated at room temperature overnight while shaking. The resulting tryptic peptide samples were desalted and concentrated using a µColumn (made with LichroprepC18) and subjected to nanoLC-MS/MS using a Proxeon Easy nanoLC and an LTQ-Orbitrap XL instrument (Lu *et al.*, 2011; Wendrich *et al.*, 2017).

Downstream proteomics data processing, analysis and visualization

Raw data were initially processed using MaxQuant 1.5.2.8 (54) (false discovery rates were set to 0.01 at peptide and protein levels) and additional results filtering (minimally two peptides are necessary for protein identification of which at least one is unique and at least one is unmodified) were performed as described previously (Smaczniak *et al.*, 2012; Cox *et al.*, 2014; Wendrich *et al.*, 2017). Filtered proteinGroups normalized (LFQ) intensities from MaxQuant were further analysed using the DEP bioconductor/R package (Zhang *et al.*, 2018). To avoid including potentially mis-annotated and/or very rare low abundant proteins, only those proteins that were detected in two out of three biological replicates of at least one growth condition were used for further analysis. To avoid infinite or large logarithmic fold changes during pair-wise differential abundance analysis, imputation of missing values was done. Missing data were imputed using random draws from a Gaussian distribution centered around a minimal value to control for proteins missing not at random. Mostly, these proteins were potentially missed in one condition as a consequence of their intensities being below the detection limit. These pre-processed data were then subjected to variance stabilization, and pair-wise differentially abundant proteins were identified using the limma R package (Ritchie

et al., 2015). Here, the tested contrasting conditions were growth on lactate versus sorbitol, lactate versus sucrose and sorbitol versus sucrose for *A. soehngenii*. Visualization was done using R packages, ggplot2 and ggpvr (Wickham, 2011; Kassambara, 2018). The Rmarkdown file with codes to reproduce the analysis of proteomics data obtained from MaxQuant 1.5.2.854 is available at: https://github.com/mibwurrepo/Shetty_et_al_Anaerobutyricum_physiology.

Identification of gene neighbourhood and phylogeny

Gene clusters were identified using the Joint Genome Institute Integrated Microbial Genomes and Microbiome System webserver (Markowitz *et al.*, 2012). The amino acid sequences for EHLA_0974 and EHLA_0978 were searched against the IMG database (Markowitz *et al.*, 2012). The BLASTp hits were limited to only those with an *E*-value cut-off of 1e-10 and at least 60% identity. The amino acid sequences were downloaded in fasta format from the IMG website. The sequences were aligned in MEGA6 using ClustalW. The aligned amino acid sequences were used to reconstruct their phylogenetic relationship based on the Jones-Taylor-Thornton substitution model.

Metagenome data mining

Raw reads (metagenomics) were downloaded for randomly chosen 100 participants from the Human Microbiome Project Data Portal (<https://portal.hmpdacc.org/>). Reads were merged using PEAR v0.9.6 (Zhang *et al.*, 2013). The assembled reads were then queried against protein sequences for *lctABCDEF* using DIAMOND v0.9.14 (settings: *blastx* --threads 6 --max-target-seqs 1 --evalue 0.001) (Buchfink *et al.*, 2014). The output in *.daa format was converted to tab delimited file using diamond view (settings: -f tab). Tab delimited files were imported into R for further processing and visualization (Venables and Smith, 2008; Racine, 2012).

Metatranscriptomics data mining for active expression lactate utilization gene cluster

Publicly available raw reads for 15 metatranscriptomes from a previously published synthetic community study in mice were downloaded from European Nucleotide Archive (ENA) (Kovatcheva-Datchary *et al.*, 2019). The synthetic community consisted of 10 bacterial strains, which included the *A. soehngenii* strain L2-7 used in the present study. Reads were filtered and trimmed using trimmomatic v0.39 (settings: PE SLIDINGWINDOW:4:20 MINLEN:70 -threads 6 HEADCROP:20) (Bolger *et al.*, 2014). Trimmed reads were mapped to

A. soehngenii genome using bowtie2 (v2.3.5.1) (Langmead and Salzberg, 2012). The counting of reads that mapped to the genome features was done using with htseq-count (v0.11.3) (settings: htseq-count --type=CDS --idattr=transcript_id) (Anders *et al.*, 2015). The output from htseq-count was imported into R for further processing (Venables and Smith, 2008; Racine, 2012).

Results and discussion

Fermentation of carbon sources and global proteomic expression profiles

Actively growing *A. soehngenii* cells pre-cultured on glucose were inoculated in basal medium containing 30 mM acetate plus either 20 mM sucrose, 40 mM D-sorbitol or 80 mM D,L-lactate as carbon sources. The initial concentration of each of the substrates was chosen in order to have equal moles of carbon. Notable, we observed higher optical density (O.D₆₀₀) sucrose and sorbitol compared to D,L-lactate during sampling of biomass for proteome extraction (Table 1). Acetate was supplied in the media as it has been previously shown to improve growth of *A. soehngenii* (Duncan *et al.*, 2004; Shetty *et al.*, 2018). Acetate was utilized along with other substrates to produce butyrate as the major end product, confirming its importance in supporting growth of *A. soehngenii* (Table 1). High amounts of formate (15.4 ± 0.4 mM, mean ± SD) were detected from utilization of sucrose (10.9 ± 1.6 mM) while this amount was much lower (5.6 ± 1.7) in the case of sorbitol and no detection in lactate condition (Table 1). The ratio for lactate:acetate:butyrate was 2:1:1.5, which is in agreement with previously reported stoichiometry (Duncan *et al.*, 2004). The carbon recovery was 70%, 64% and 82.3% for lactate, sorbitol and sucrose respectively. The missing carbon can be attributed to biomass production and intermediates that could not be detected by HPLC at mid-log phase.

Next, we compared the total expressed proteome of *Anaerobutyricum soehngenii* when grown on different carbon sources to mid exponential phase. On average,

we detected 617 ± 17 proteins present in the minimum of two biological replicates in each condition, of which 453 were detected in all the three growth conditions (Fig. S1A and B). Principal component analysis of the expressed proteome profiles suggested global differences in protein abundances between the three growth conditions, i.e. D,L-lactate, sucrose and sorbitol (Fig. S1C). Pearson's correlation between replicates was for all conditions > 0.89 (Fig. S1D).

Differential proteomes of *A. soehngenii* during growth on sorbitol and sucrose

Sucrose is a common substrate in the distal part of the upper intestinal tract and the ability to utilize sucrose is one of the differentiating features for *A. soehngenii* compared to *A. hallii* (Shetty *et al.*, 2018). Sugar alcohols are commonly used as replacement of glucose for controlling the blood glucose levels in humans. One such common sugar alcohol is sorbitol, which is metabolized slowly by the human body and hence can be available as a carbon and energy source for bacteria in the large intestine (Deis and Kearsley, 2012). Therefore, we sought to identify the active proteins in the metabolic pathway that leads to conversion of sorbitol and sucrose to butyrate and formate.

In total, we identified 23 proteins that were differentially expressed (adjusted *p*-value = 0.05, log₂-fold change, log₂FC = 1.5) when *A. soehngenii* was grown in presence of sorbitol compared to sucrose. Among these, seven were overexpressed and 16 were less expressed when cells were grown on sorbitol compared to sucrose (Fig. 1A). Further inspection of the expressed proteome revealed that *A. soehngenii* has an inducible operon for sorbitol uptake. The genes encoding the phosphotransferase system (PTS) for sorbitol are located between the *gutM* gene coding for the glucitol/sorbitol operon activator and *licR* encoding the lichenan operon transcriptional antiterminator (Fig. 1B). The significantly higher (log₂FC > 6) abundance of these proteins during growth on sorbitol when compared to that on sucrose suggests that this gene cluster is coordinately expressed

Table 1. Substrate utilization and fermentation end products for *A. soehngenii*.

	Optical density (O.D ₆₀₀)	Consumption (mM)		Production (mM)	
		Substrate	Acetate	Butyrate	Formate
D,L-lactate	0.75 ± 0.04	46.6 ± 2.8	19.3 ± 0.2	19.4 ± 2.7	ND
Sorbitol	8.0 ± 1.0	25.0 ± 3.3	15.0 ± 3.9	16.3 ± 3.2	5.6 ± 1.7
Sucrose	5.9 ± 0.57	10.9 ± 1.6	5.3 ± 3.2	18.2 ± 1.1	15.4 ± 0.4

The amounts of substrate and end products were measured for the mid-exponential samples used for proteomics. The values (mean ± SD) represent mean and standard deviation of biological triplicates for each condition. Carbon recovery for D,L-lactate was 70%, for sorbitol it was 64% and for sucrose it was 82.3% at the time of sampling at mid-exponential phase. The amount of formed CO₂ is calculated based on assumption that 1 mol of C6 compound utilized releases 2 mol of CO₂ or 1 mol of formate, hence 1 mol of lactate releases only 1 mol of CO₂ (Bui *et al.*, 2019). ND: not detected.

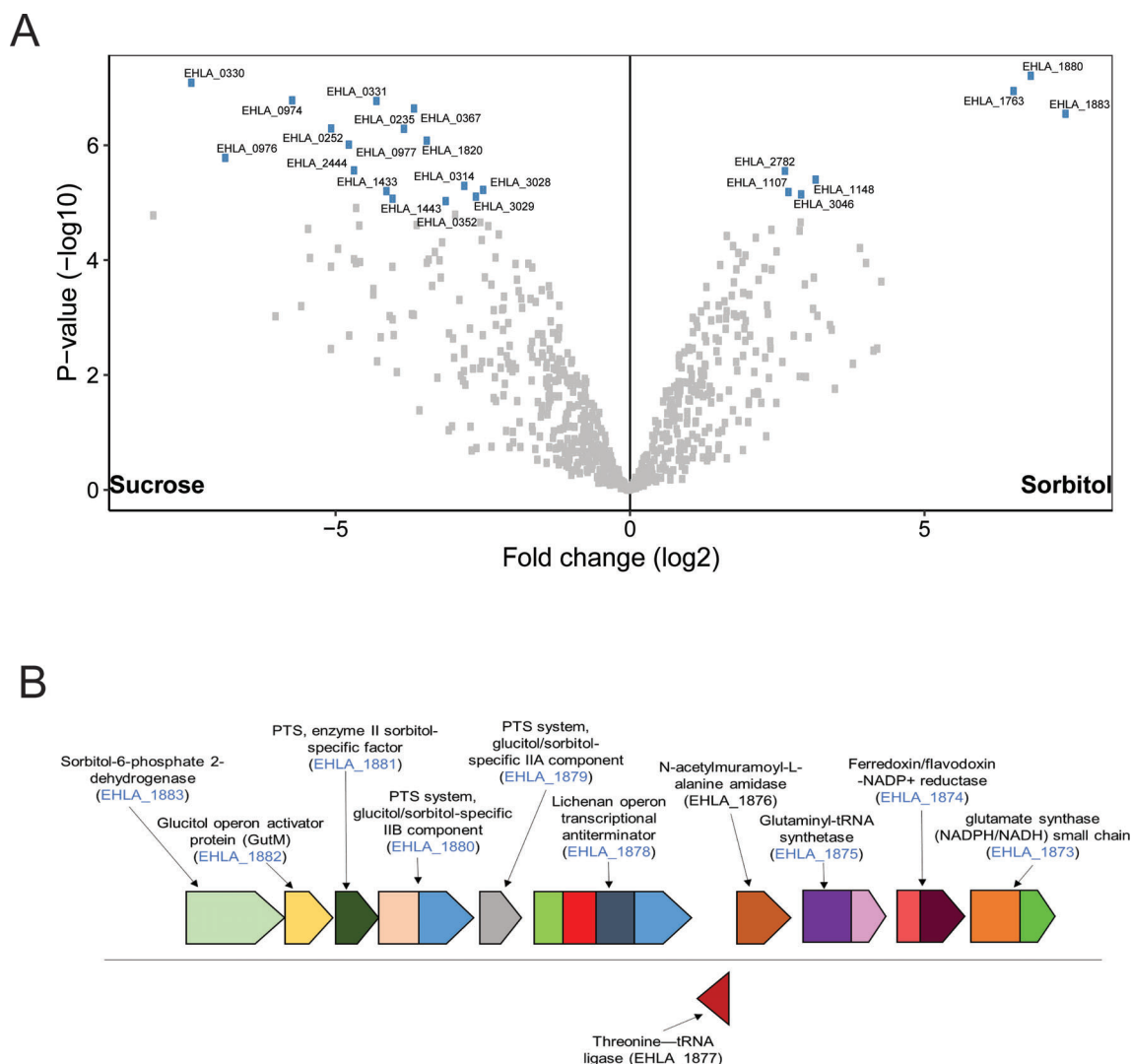


Fig. 1. Comparison of *A. soehngenii* expressed proteomes when grown on sorbitol or sucrose. **A.** Volcano plot depicting the differential protein expression of *A. soehngenii* when grown on sucrose or sorbitol. The significantly different (adjusted p -value <0.05) proteins are labelled with corresponding locus tags. **B.** Organization of gene cluster involved in sorbitol utilization in the genome of *A. soehngenii*. The locus tags detected in the proteome are coloured in blue.

and highly induced in the presence of sorbitol. The promoter region is located before the gene EHLA_1883 that encodes sorbitol-6-phosphate dehydrogenase, while the terminator is located immediately after the transcriptional antiterminator protein EHLA_1878.

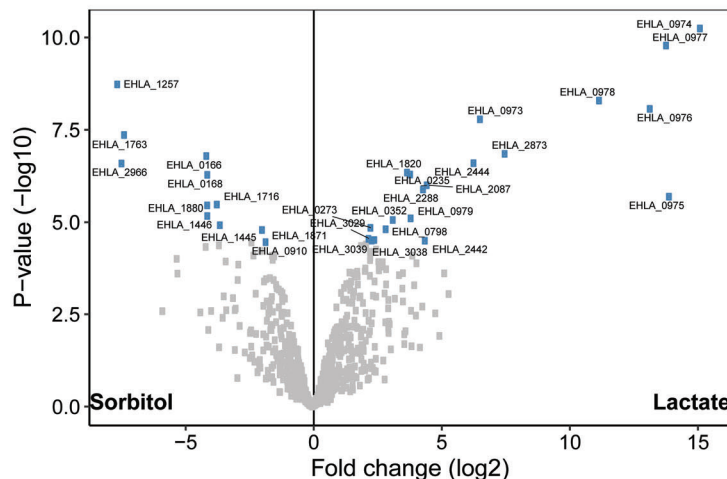
Comparison of the expressed proteome of cells grown on sucrose and sorbitol further revealed that the product of EHLA_0352, bacterial extracellular solute-binding protein (IMG-annotation: carbohydrate ABC transporter substrate-binding protein, CUT1 family (TC 3.A.1.1.-)) had a ten-fold higher abundance in cells grown on sucrose (Table S1). We identified two locus tags EHLA_0193 and EHLA_0331 encoding a protein-Npi-phosphohistidine-sugar PTS system. The protein encoded by EHLA_0331 was less abundant ($\log_2\text{FC} > 4$,

adjusted p -value = 0.039) in sorbitol compared to sucrose. Also the protein encoded by EHLA_0193 was observed at lower abundance ($\log_2\text{FC} > 8$, adjusted p -value = 0.074), albeit not significantly, but we observed that it was $>5 \log_2$ fold higher in abundance in cells grown on sucrose compared to cells grown either on D,L-lactate or sorbitol (Table S1). Upstream of the gene encoding the sucrose PTS is a LacI-type HTH domain containing gene (EHLA_0192), the product of which was not detected in the proteome data but is likely a transcriptional regulator controlling the expression of the sucrose-PTS gene. Formate was not detected lactate condition and investigation of proteomics data revealed that a pyruvate formate lyase [EHLA_2966 formate C-acetyltransferase and EHL_2967 (formate-C-acetyltransferase)-activating enzyme] is

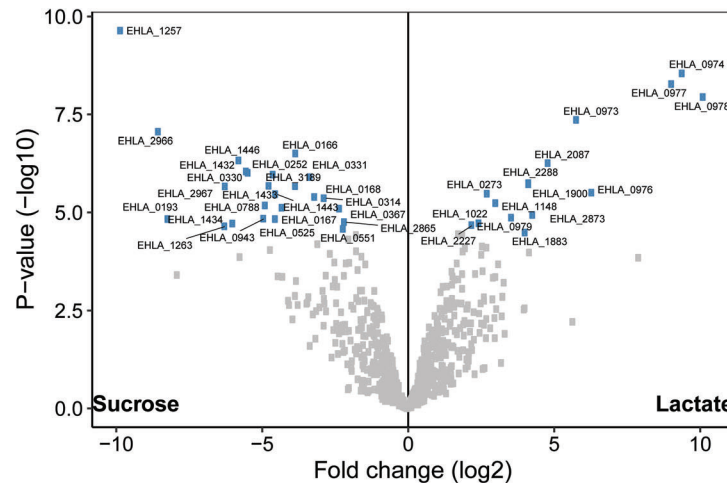
significantly less abundant in cells grown on lactate than on sucrose or sorbitol. Next, we did a BLASTp search for the active proteins involved in sucrose utilization in the publicly available genome of *A. hallii* DSM3353. The products encoded by the *A. soehngenii* genes with locus tags EHLA_0192 and EHLA_0193 were not identified in *A. hallii*, while a homologue of the protein encoded by the gene with locus tag EHLA_0195 was identified, which

shared 52% identity with a 4- α -glucanotransferase in *A. hallii*. Additionally, the gene encoding maltose 6'-phosphate phosphatase (EHLA_0194) did not have a homologue in the genome of *A. hallii*. Based on our findings, we hypothesize that genes with the locus tags EHLA_0192, EHLA_0193, EHLA_0194 and EHLA_0195 are likely responsible for the ability of *A. soehngenii* to utilize sucrose. These genes are not (completely) present in

A



B



EHLA_1257	Phosphopyruvate_hydration
EHLA_2966	Formate_C-acetyltransferase
EHLA_1446	AMP-dependent_synthetase/ligase
EHLA_0235	Glycerol_kinase
EHLA_0166	Acyl-CoA_dehydrogenase
EHLA_1820	OmpR/PhoB-type_DNA-binding_domain_profile
EHLA_0352	Bacterial_extracellular_solute-binding_protein
EHLA_2444	Protein-Npi-phosphohistidine-sugar_phosphotransferase
EHLA_3029	Aldehyde_dehydrogenase_family
EHLA_3038	Propanediol/glycerol_dehydratase_medium_subunit
EHLA_3039	Propanediol_dehydratase
EHLA_0168	Electron_transfer_flavoprotein_alpha/beta-subunit_N-terminal
EHLA_1445	Methyltransferase_domain
EHLA_0910	Phosphoglycerate_kinase
EHLA_0976	Electron_transfer_flavoprotein_alpha/beta-subunit_N-terminal
EHLA_0798	Acetate_kinase
EHLA_1871	IMP_dehydrogenase
EHLA_2873	Periplasmic_binding_protein_domain
EHLA_2442	Carbohydrate_kinase_PfkB
EHLA_0975	Electron_transfer_flavoprotein_alpha/beta-subunit_N-terminal
EHLA_1716	Alkyl_hydroperoxide_reductase_subunit_C/Thiol_specific_antioxidant
EHLA_0273	Pyruvate:ferredoxin_oxidoreductase_core_domain_II
EHLA_0979	GntR_bacterial_regulatory_protein_HTH_signature
EHLA_0974	CO_dehydrogenase_flavoprotein-like_FAD-binding_subdomain_2
EHLA_2288	Domain_of_unknown_function_DUF2088
EHLA_0977	Short-chain_acyl-CoA_dehydrogenase
EHLA_0973	LctP_transporter_lactate_permease(LctP)_family
EHLA_2087	Phosphoribosyltransferase_domain
EHLA_1763	Oxidoreductase_family_NAD-binding_Rossmann_fold
EHLA_0978	Domain_of_unknown_function_DUF2088
EHLA_1880	Protein-Npi-phosphohistidine-sugar_phosphotransferase
EHLA_0193	Protein-Npi-phosphohistidine-sugar_phosphotransferase
EHLA_0330	1-PFK_hexose_kinase_1-phosphofructokinase_family
EHLA_1263	FeoA_domain
EHLA_0252	BMC_domain
EHLA_0943	Hypothetical_protein
EHLA_0525	Shikimate_kinase
EHLA_1432	AAA+ ATPase_domain
EHLA_1434	Adenosinetriphosphatase
EHLA_1433	Adenosinetriphosphatase
EHLA_1443	Alpha/Beta_hydrolase_fold
EHLA_2967	[Formate-C-acetyltransferase]-activating_enzyme
EHLA_0788	P-type_cation-transporting_ATPase_superfamily_signature
EHLA_0331	Protein-Npi-phosphohistidine-sugar_phosphotransferase
EHLA_3189	4-alpha-glucanotransferase
EHLA_0367	Starch_synthase_[glycosyl-transferring]/glycogen(starch)_synthase
EHLA_0167	Electron_transfer_flavoprotein_alpha/beta-subunit_N-terminal
EHLA_0314	6-phosphofructokinase
EHLA_2865	Tryptophan_synthase
EHLA_0551	Pyridoxal_5-phosphate_synthase_subunit_PdxS_[pdxS]
EHLA_2227	Lactaldehyde_reductase
EHLA_1022	Glycerol_dehydrogenase
EHLA_1148	PBP2_GlnP
EHLA_1900	2-C-methyl-D-erythritol_2,4-cyclodiphosphate_synthase
EHLA_1883	Short-chain_dehydrogenase/reductase_conserved_site

Fig. 2. Comparison of *A. soehngenii* expressed proteomes when grown on D,L-lactate compared to sorbitol or sucrose. The significantly different (adjusted *p*-value <0.05) proteins are labelled with corresponding locus tags. A. Volcano plot depicting the differential protein expression of *A. soehngenii* when grown on sorbitol or D,L-lactate. B. Volcano plot depicting the differential protein expression of *A. soehngenii* when grown on sucrose or D,L-lactate. The proteins encoded by genes with the locus tags are indicated.

A. hallii, which cannot grow on sucrose (Shetty *et al.*, 2018).

Proteomics-guided identification of the active D,L-lactate utilization gene cluster

Several intestinal bacteria such as those belonging to the genera *Bifidobacterium* and *Lactobacillus* are known lactate producers. Lactate is one of the major cross-feeding metabolites that are converted to butyrate by butyrate producing bacteria such as *A. soehngenii* (Duncan *et al.*, 2004). All known bacteria belonging to the genera *Anaerobutyricum* and *Anaerostipes* have been shown to convert lactate and acetate into butyrate but only *A. soehngenii* and *A. hallii* as well as *Anaerostipes caccae* and *Anaerostipes rhamnosivorans* have been described to utilize both D- and L- forms of lactate

(Duncan *et al.*, 2004). However, to the best of our knowledge, the active genes involved in butyrogenesis from D,L-lactate and acetate have not been investigated in detail for members of *Anaerobutyricum* and related genera. The expressed proteome of *A. soehngenii* revealed that 31 and 39 proteins were significantly increased when grown on D,L-lactate versus sorbitol and D,L-lactate versus sucrose, respectively ($\log_2FC \geq 1.5$, adjusted p -value = 0.05). A total of 15 proteins were shared between the two comparisons (including those encoded by genes with locus tags EHLA_0973, EHLA_0974, EHLA_0976, EHLA_0977, EHLA_0978, EHLA_0979) that were induced by growth on D,L-lactate. Proteins with significantly higher abundance during growth on D,L-lactate included lactate permease, lactate dehydrogenase, electron transfer flavoprotein beta subunit, electron transfer flavoprotein alpha subunit, short-chain acyl-CoA

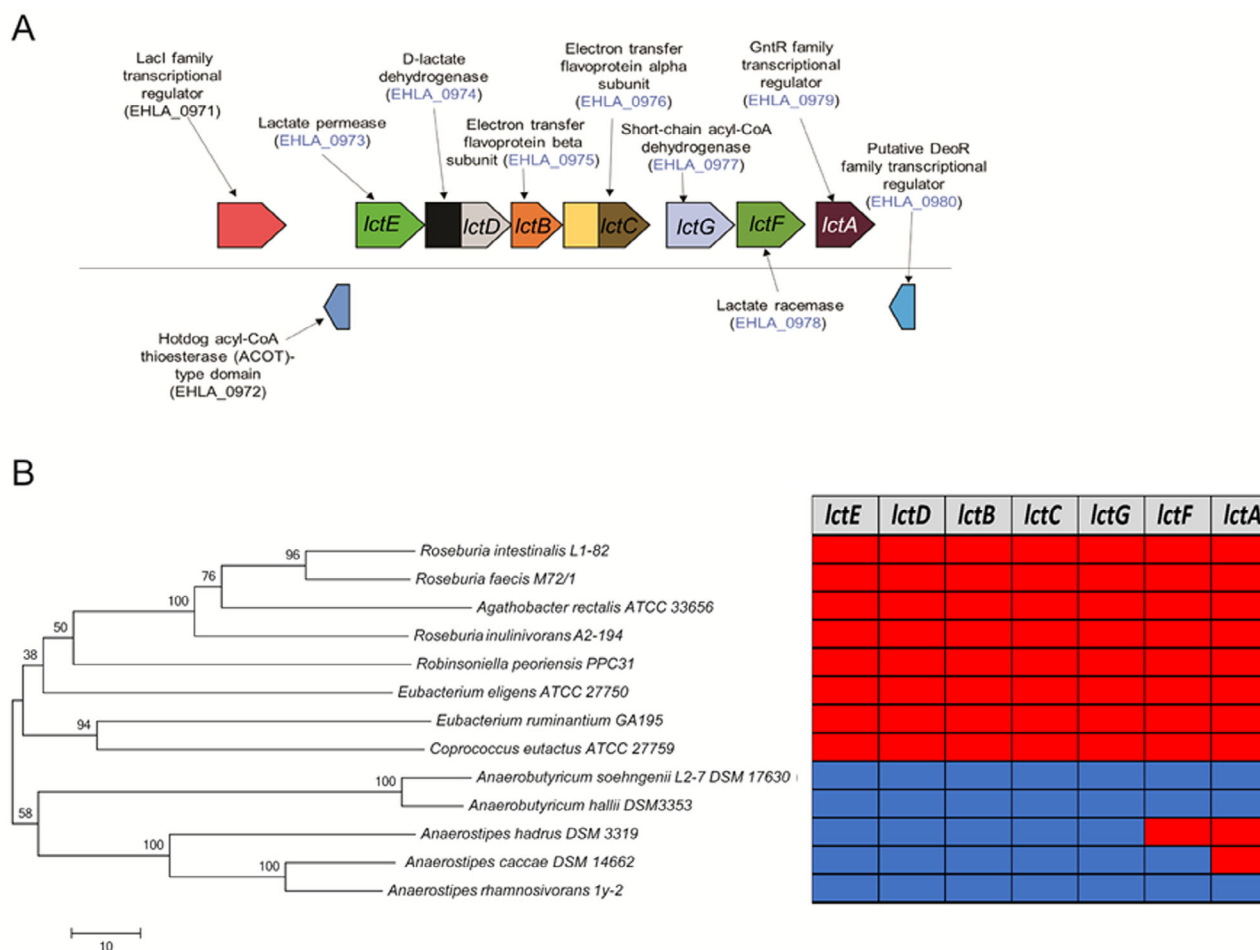


Fig. 3. D,L-lactate utilization gene cluster. A. Organization of gene cluster involved in D,L-lactate utilization in the genome of *A. soehngenii*. The locus tags detected in the proteome are coloured in blue. B. Phylogenetic tree based on the 16S rRNA gene sequences for *A. soehngenii* and closely related species. The evolutionary history was inferred using the neighbour-joining method based on 1000 bootstrap replicates. The amino acid sequences for D-lactate dehydrogenase (EHLA_0974), lactate permease (EHLA_0973) and nickel-dependent lactate racemase (EHLA_0978) were searched against the genomes of closely related species. Presence in genome is indicated by blue-coloured cells, absence by red-coloured cells.

dehydrogenase and lactate racemase (encoded by genes with locus tags EHLA_0973 - EHLA_0978) (Fig. 2A and B). Remarkably, lactate dehydrogenase (encoded by the gene with locus tag EHLA_0974) co-occurs with EtfA and EtfB, which suggests the presence of a D-lactate dehydrogenase/electron-transferring flavoprotein (LDH/Etf) complex allowing growth on lactate, which is a low energy substrate (Weghoff *et al.*, 2015). Interestingly, a gene coding for a short-chain acyl-CoA dehydrogenase (*lctG*), homologous to that coding for butyryl-CoA dehydrogenase, was identified to be located downstream of the gene encoding lactate dehydrogenase, and corresponding gene products were found produced together with other lactate-specific proteins. The organization of the new D,L-lactate utilization gene cluster, *lctABCDEFGH* is depicted in Fig. 3A.

Comparison of genomes of *Anaerobutyricum* and related species

Next, we searched for homologues of genes encoding lactate permease, lactate dehydrogenase and lactate racemase (encoded by genes with the locus tags EHLA_0973, EHLA_0974 and EHLA_0978 respectively) in the genomes of bacteria closely related to *Anaerobutyricum* (Fig. 3B). *Anaerobutyricum hallii*, *Anaerostipes caccae* and *Anaerostipes rhamnosivorans*

have homologues of genes coding for lactate dehydrogenase, lactate racemase and lactate permease. *Anaerobutyricum rhamnosivorans*, which was previously isolated from infant faeces and known to have the ability to utilize D,L-lactate and acetate, has all the genes of the *lct* operon as those found in *Anaerobutyricum* species (Bui *et al.*, 2014). However, the genome of *A. hadrus* lacked the gene encoding lactate racemase, which is consistent with its inability to utilize L-lactate (Fig. 3B) (Duncan *et al.*, 2004; Allen-Vercoe *et al.*, 2012). We further expanded our search to other bacterial genomes present in the IMG database (Markowitz *et al.*, 2012). The lactate racemase encoded in the genome of *Anaerobutyricum* is predicted to be a nickel-dependent enzyme (Desguin *et al.*, 2014). Further support for the nickel dependency derived from the presence of a set of auxiliary enzymes, namely *larE* [pyridinium-3,5-biscarboxylic acid mononucleotide sulfurtransferase (EC:4.4.1.37), EHLA_1081], *larC1* [pyridinium-3,5-bisthiocarboxylic acid mononucleotide nickel chelataze (EC:4.99.1.12), EHLA_1342] and *larC2* [pyridinium-3,5-biscarboxylic acid mononucleotide synthase (EC:2.5.1.143), EHLA_1341], predicted to be involved in nickel incorporation were detected in the proteome but encoded in a scattered way in the genome. A BLASTp search for amino acid sequences similar to the product of the gene with locus tag EHLA_0978 revealed their

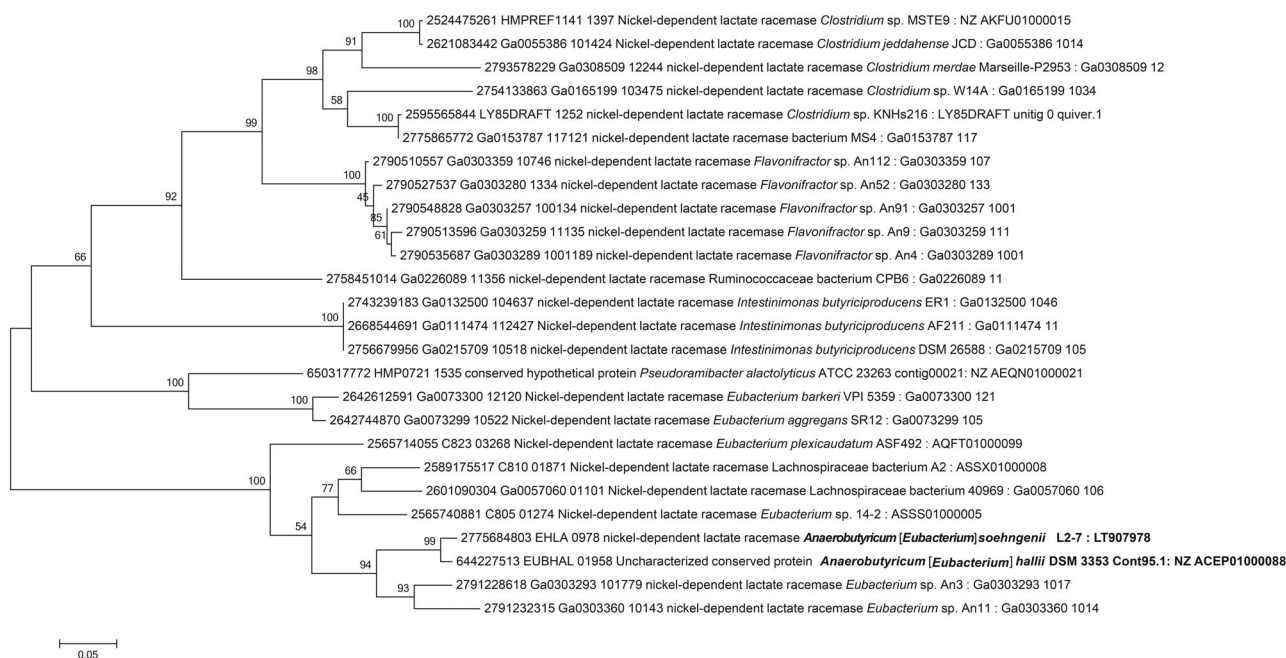


Fig. 4. Maximum likelihood phylogenetic tree of 26 lactate racemase homologues. Phylogenetic tree of 26 lactate racemase homologues identified by searching against 55 499 isolate genomes of the IMG/ER database (as of 2 October 2018). Labels represent the IMG gene ID, locus tag, IMG annotation for the gene product, taxonomic identity, strain name, and assembly and/or contig/scaffold and NCBI/DDJB/ENA accession number of genomes which contains this gene. The numbers on branches represent the bootstrap values from 1000 replicates. See methods for details on calculation.

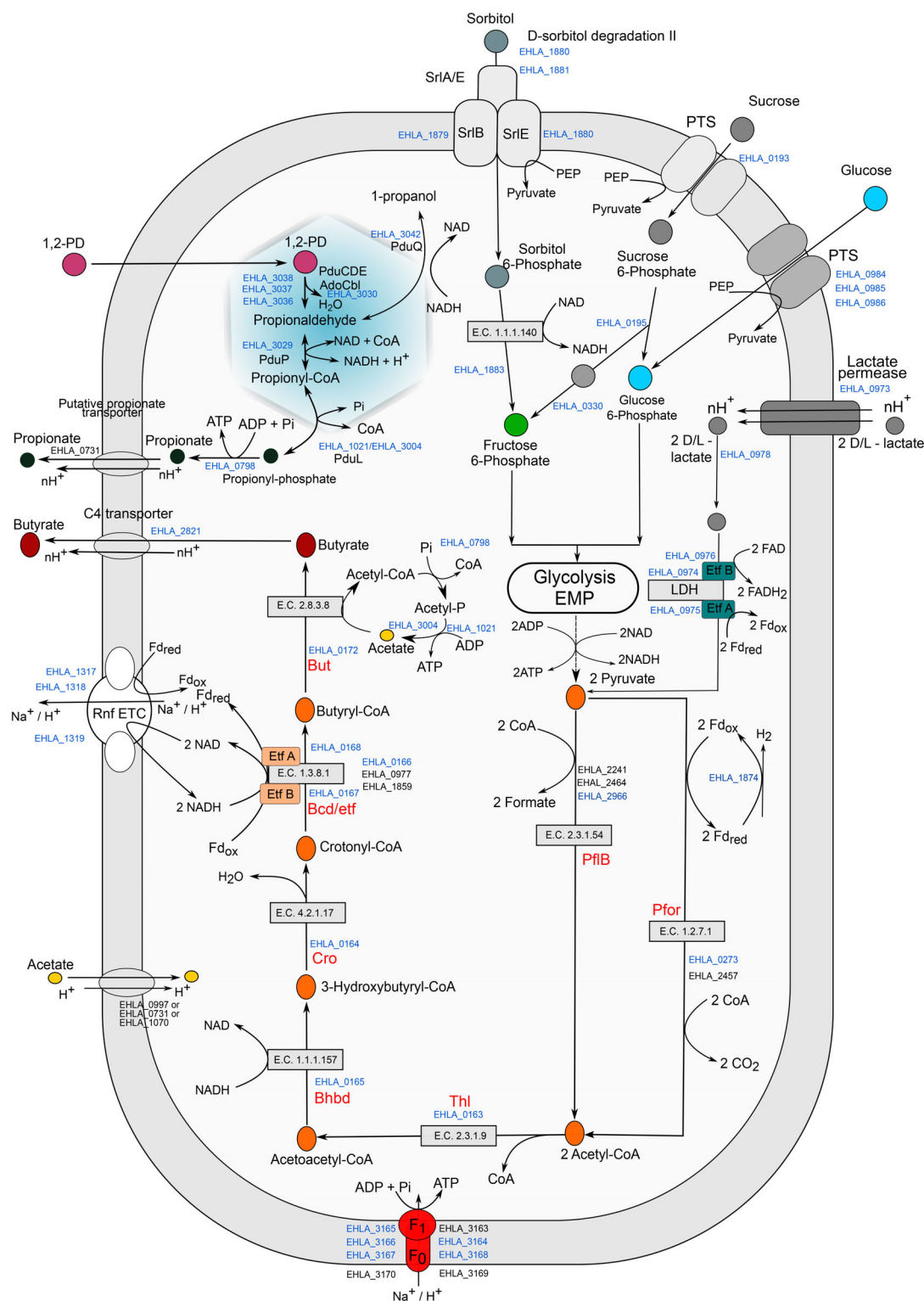


Fig. 5. Overview of the key metabolic routes of butyrate and propionate formation in *A. soehngenii*. Relevant locus tags with the prefix EHLA_ are indicated and those in blue relate to gene products that are detected in proteome, while those in black are only identified in genome but not detected in the proteome at any of the growth conditions. The SCFA transporters most likely belong to the C4 TRAP family of transporters. The symport is with a proton and *n* indicates the number of protons transported across the cell membrane which may vary under various growth conditions.

presence in genomes of members of genera *Intestinimonas* and *Flavonifractor*, as well as in *Eubacterium barkeri*, *E. aggregans*, *Pseudoaeromonas alactolyticus*, *Clostridium merdae*, *C. jeddahense* and related species. A phylogenetic analysis suggests that lactate racemase (encoded by the gene with locus tag EHLA_0978) in *A. soehngenii* shares similarity to the lactate racemases present in related *E. barkeri*, *E. aggregans*, *Intestinimonas* spp. and *Flavonifractor* (Fig. 4). *Eubacterium aggregans* (isolated from methanogenic bioreactors) and *Intestinimonas* spp. are known for their ability to convert lactate plus acetate to butyrate (Tahar Mechichi and Woo, 1998; Kläring et al., 2013) (Bui et al 2016), although the growth of the human gut isolate *I. butyriciproducens* AF211 is relatively slow on lactate and acetate (Bui et al., 2015). *Flavonifractor* species are closely related to *Intestinimonas* and can potentially utilize lactate and acetate to produce butyrate.

A BLASTp search for amino acid sequences of lactate dehydrogenase (encoded by the gene with locus tag EHLA_0974) revealed a widespread distribution of homologues in the IMG genome database (Fig. S2). As with lactate racemase, the phylogenetically related (based on 16S rRNA gene) *Anaerobutyricum* and *Anaerostipes* did not share high identity with each other with respect to lactate-dehydrogenase-encoding genes and were phylogenetically placed in two distinct clusters. However, further analysis of neighbourhood genes indicated that both the *Anaerobutyricum* and *Anaerostipes* lactate dehydrogenase gene is followed by genes coding for electron transfer proteins (Fig. S3). Mining of the publicly available human gut metagenomes part of the human microbiome project revealed high abundance and prevalence of the *lctABCDEF* genes (Huttenhower et al., 2012) (Fig. S4). To investigate whether the *lctABCDEF* gene cluster in *A. soehngenii* was active under in vivo conditions, we analysed the previously published metatranscriptomics data from a simplified intestinal microbiota (SIM)-colonized mice (Kovatcheva-Datchary et al., 2019). The

SIM consisted of 10 human intestinal strains, including *A. soehngenii* strain L2-7. Although in the SIM study the abundance of *A. soehngenii* was low, all the genes of the *lctABCDEF* gene cluster were found to be actively transcribed in one of the samples (Table S2). This considerable expression of the *lctABCDEF* genes (1.7% of genes of strain L2-7) in a SIM-colonized mouse model testifies for their functionality in vivo. Overall, these observations provide evidence for the adaptation of *A. soehngenii*, *A. hallii* and *Anaerostipes caccae* to utilize an important but low-energy cross-feeding metabolite, D,L-lactate.

Concluding remarks

The genomic and proteomic data presented in this study allowed elucidating key metabolic features of *A. soehngenii* (Fig. 5). We identified an inducible operon encoding a sorbitol transporter, which was abundant in the expressed proteome of *A. soehngenii* when grown in presence of sorbitol. The ability to utilize sucrose is likely conferred by two proteins that are not encoded in the genome of the phylogenetically related *A. hallii*, viz. a sucrose PTS and LacI-type HTH domain, which were highly induced in presence of sucrose. The specialized ability of *A. soehngenii* to convert D,L-lactate and acetate to butyrate is conferred by the new genomic organization of a lactate utilization gene cluster. Although initial steps of glycolysis require ATP, the overall conversion of glucose to pyruvate is exergonic (Table 2, Eq. 1). However, growth on lactate poses a major energetic barrier because lactate has to be converted to pyruvate in an NADH-generating and endergonic process (Table 2, Eq. 2a). The free energy resulting from the coupling with oxidation of ferredoxin has been calculated as -9.5 kJ/mol (see Eq. 2b; Weghoff et al., 2015).

The bacterial species, *Anaerostipes caccae*, *Anaerostipes rhamnosivorans* and the two *Anaerobutyricum* species, *A. soehngenii* and *A. hallii*, are capable of converting D,L-lactate plus acetate to butyrate. Other common intestinal bacteria such as

Table 2. Calculated Gibbs' free energy for conversion of glucose to pyruvate and lactate to pyruvate.

Reaction	ΔG^0
Equation 1: Glucose + 2Pi + 2ADP + 2NAD ⁺ → 2 Pyruvate + 2ATP + 2NADH + 2H ⁺ + 2H ₂ O	-88 kJ/mol
Equation 2a: Lactate + NAD ⁺ ↔ Pyruvate + NADH	+25 kJ/mol
Equation 2b as proposed by Weghoff et al. (2015): Lactate + Fd ²⁻ + 2NAD ⁺ → Pyruvate + Fd + 2NADH	-9.5 kJ/mol
Equation 3: 4 Lactate + 2 Acetate → 3 Butyrate + 4CO ₂ + 2H ₂ + 2H ₂ O	-139 kJ/mol
Equation 4: Glucose + 2H ₂ O → Butyrate + 2CO ₂ + 4H ₂	-249.96 kJ/mol

The calculations were done following the information reported previously (Thauer et al., 1977).

Faecalibacterium prausnitzii, *Roseburia intestinalis* and *Agathobacter rectalis* are known to consume acetate for the production of butyrate from sugars (Duncan *et al.*, 2002; Duncan and Flint, 2008; Heinken *et al.*, 2014). As a consequence, a community with co-occurrence of acetate utilizers will have high competition for acetate, which poses another challenge for the efficient utilization of lactate (along with the low energy yield) by *Anaerobutyricum* and *Anaerostipes* spp. Genome analysis demonstrated that *A. hallii*, *A. soehngenii*, *Anaerostipes caccae* and *Anaerostipes rhamnosivorans* have well-adapted lactate utilization gene clusters that encode all necessary proteins [transcriptional regulator, transporter, racemase (two copies), LDH/Etf complex, a homologue of butyryl-CoA dehydrogenase] that were also detected in high amounts in the expressed proteome data. In conclusion, data presented here has revealed how *Anaerobutyricum soehngenii* and related butyrate producers can overcome the energetic barrier to utilize D,L-lactate to produce butyrate.

Acknowledgement

We thank Dr. Irene Sanchez Andrea for useful discussions on anaerobic metabolism and thermodynamics during the course of the study and Ton van Gelder for technical support. This research was partly supported by the Netherlands Organization for Scientific Research, Spinoza Award and SIAM Gravity Grant 024.002.002 to WMdV and the UNLOCK project NRGWI.obrug.2018.005 to HS.

References

- Allen-Vercoe, E., Daigneault, M., White, A., Panaccione, R., Duncan, S.H., Flint, H.J., *et al.* (2012) *Anaerostipes hadrus* comb. nov., a dominant species within the human colonic microbiota; reclassification of *Eubacterium hadrum* Moore *et al.* 1976. *Anaerobe* **18**: 523–529.
- Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.
- Barcenilla, A., Pryde, S.E., Martin, J.C., Duncan, S.H., Stewart, C.S., Henderson, C., and Flint, H.J. (2000) Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* **66**: 1654–1661.
- Belzer, C., Chia, L.W., Aalvink, S., Chamlagain, B., Piironen, V., Knol, J., and de Vos, W.M. (2017) Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *MBio* **8**: e00770–e00717.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Buchfink, B., Xie, C., and Huson, D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**: 59.
- Bui, T.P.N., de Vos, W.M., and Plugge, C.M. (2014) *Anaerostipes rhamnosivorans* sp. nov., a human intestinal, butyrate-forming bacterium. *Int J Syst Evol Microbiol* **64**: 787–793.
- Bui, T.P.N., Ritari, J., Boeren, S., de Waard, P., Plugge, C.M., and de Vos, W.M. (2015) Production of butyrate from lysine and the Amadori product fructoselysine by a human gut commensal. *Nat Commun* **6**: 10062.
- Bui, T.P.N., Schols, H.A., Jonathan, M., Stams, A.J., de Vos, W.M., and Plugge, C.M. (2019) Mutual metabolic interactions in co-cultures of the intestinal *Anaerostipes rhamnosivorans* with an acetogen, methanogen, or pectin-degrader affecting butyrate production. *Front Microbiol* **10**: 2449.
- Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**: 2513–2526.
- Davie, J.R. (2003) Inhibition of histone deacetylase activity by butyrate. *J Nutr* **133**: 2485S–2493S.
- Deis, R.C., and Kearsley, M.W. (2012) Sorbitol and mannitol. In *Sweeteners and Sugar Alternatives in Food Technology*, pp. 331–346. West Sussex, England: John Wiley & Sons, Ltd.
- Desguin, B., Goffin, P., Viaene, E., Kleerebezem, M., Martin-Diaconescu, V., Maroney, M.J., *et al.* (2014) Lactate racemase is a nickel-dependent enzyme activated by a widespread maturation system. *Nat Commun* **5**: 3615.
- Detman, A., Mielecki, D., Chojnacka, A., Salamon, A., Błaszczak, M.K., and Sikora, A. (2019) Cell factories converting lactate and acetate to butyrate: *Clostridium butyricum* and microbial communities from dark fermentation bioreactors. *Microb Cell Fact* **18**: 36.
- Duncan, S.H., and Flint, H.J. (2008) Proposal of a neotype strain (A1-86) for *Eubacterium rectale*. Request for an opinion. *Int J Syst Evol Microbiol* **58**: 1735–1736.
- Duncan, S.H., Hold, G.L., Barcenilla, A., Stewart, C.S., and Flint, H.J. (2002) *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int J Syst Evol Microbiol* **52**: 1615–1620.
- Duncan, S.H., Louis, P., and Flint, H.J. (2004) Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* **70**: 5810–5817.
- Fellows, R., Denizot, J., Stellato, C., Cuomo, A., Jain, P., Stoyanova, E., *et al.* (2018) Microbiota derived short chain fatty acids promote histone crotonylation in the colon through histone deacetylases. *Nat Commun* **9**: 105.
- Heinken, A., Khan, M.T., Paglia, G., Rodionov, D.A., Harmsen, H.J.M., and Thiele, I. (2014) Functional metabolic map of *Faecalibacterium prausnitzii*, a beneficial human gut microbe. *J Bacteriol* **196**: 3289–3302.
- Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., *et al.* (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207.
- Kassambara, A. (2018) ggpubr: “ggplot2” based publication ready plots. *R package version 01.7*.
- Kläring, K., Hanske, L., Bui, N., Charrier, C., Blaut, M., Haller, D., *et al.* (2013) *Intestinimonas butyriciproducens* gen. nov., sp. nov., a butyrate-producing bacterium from the mouse intestine. *Int J Syst Evol Microbiol* **63**: 4606–4612.

- Kovatcheva-Datchary, P., Shoaie, S., Lee, S., Wahlström, A., Nookaew, I., Hallen, A., et al. (2019) Simplified intestinal microbiota to study microbe-diet-host interactions in a mouse model. *Cell Rep.* **26**: 3772–3783.
- Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* **9**: 357.
- Louis, P., and Flint, H.J. (2017) Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* **19**: 29–41.
- Lu, J., Boeren, S., De Vries, S., Van Valenberg, H., Vervoort, J., and Hettinga, K. (2011) Filter-aided sample preparation with dimethyl labeling to identify and quantify milk fat globule membrane proteins. *J Proteomics* **75**: 34–43.
- Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* **40**: D115–D122.
- McMillan, L., Butcher, S.K., Pongracz, J., and Lord, J.M. (2003) Opposing effects of butyrate and bile acids on apoptosis of human colon adenoma cells: differential activation of PKC and MAP kinases. *Br J Cancer* **88**: 748–753.
- Racine, J.S. (2012) *RStudio: A Platform-Independent IDE for R and Sweave*. *J Appl Econ* **27**: 167–172.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47–e47.
- Schulthess, J., Pandey, S., Capitani, M., Rue-Albrecht, K.C., Arnold, I., Franchini, F., et al. (2019) The short chain fatty acid butyrate imprints an antimicrobial program in macrophages. *Immunity* **50**: 432–445.
- Seheult, J., Fitzpatrick, G., and Boran, G. (2017) Lactic acidosis: an update. *Clin Chem Lab Med* **55**: 322–333.
- Shetty, S.A., Zuffa, S., Bui, T.P.N., Aalvink, S., Smidt, H., and De Vos, W.M. (2018) Reclassification of *Eubacterium hallii* as *Anaerobutyricum hallii* gen. nov., comb. nov., and description of *Anaerobutyricum soehngenii* sp. nov., a butyrate and propionate-producing bacterium from infant faeces. *Int J Syst Evol Microbiol* **68**, 3741–3746. <https://doi.org/10.1099/ijsem.0.003041>.
- Smaczniak, C., Li, N., Boeren, S., America, T., Van Dongen, W., Goerdayal, S.S., et al. (2012) Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues. *Nat Protoc* **7**: 2144.
- Tahar Mechichi, M.L., and Woo, H. (1998) *Eubacterium aggregans* sp. nov., a new homoacetogenic bacterium from olive mill wastewater treatment digester. *Anaerobe* **4**: 283–291.
- Ten Bruggencate, S.J., Bovee-Oudenhoven, I.M., Lettink-Wissink, M.L., Katan, M.B., and van der Meer, R. (2006) Dietary fructooligosaccharides affect intestinal barrier function in healthy men. *J Nutr* **136**: 70–74.
- Thangaraju, M., Cresci, G.A., Liu, K., Ananth, S., Gnanaprakasam, J.P., Browning, D.D., et al. (2009) GPR109A is a G-protein-coupled receptor for the bacterial fermentation product butyrate and functions as a tumor suppressor in colon. *Cancer Res* **69**: 2826–2832.
- Thauer, R.K., Jungermann, K., and Decker, K. (1977) Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol Rev* **41**: 100.
- Topping, D.L., and Clifton, P.M. (2001) Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol Rev* **81**: 1031–1064.
- Venables, W.N., and Smith, D.M. (2008) *An Introduction to R*, (100). Vienna, Austria: Network Theory 2008.
- Weghoff, M.C., Bertsch, J., and Müller, V. (2015) A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ Microbiol* **17**: 670–677.
- Wendrich, J.R., Boeren, S., Möller, B.K., Weijers, D., and De Rybel, B. (2017) In vivo identification of plant protein complexes using IP-MS/MS. In *Plant Hormones*, (pp. 147–158) New York, NY: Springer.
- Wickham, H. (2011) ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* **3**: 180–185.
- Wopereis, H., Sim, K., Shaw, A., Warner, J.O., Knol, J., and Kroll, J.S. (2018) Intestinal microbiota in infants at high risk for allergy: effects of prebiotics and role in eczema development. *J Allergy Clin Immunol* **141**: 1334–1342.
- Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2013) PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* **30**: 614–620.
- Zhang, X., Smits, A.H., van Tilburg, G.B., Ovaa, H., Huber, W., and Vermeulen, M. (2018) Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat Protoc* **13**: 530.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Supporting information Significantly different proteins identified for each pairwise comparison.

Table S2. Identification of IctABCDEFGF cluster of *A. soehngenii* L2-7 in cecum metatranscriptomes from mice colonized with a simplified intestinal microbiota (SIM).

Fig. S1. Overview of the global proteomic comparison.

A. Number of proteins detected in each of the biological triplicates for different carbon sources.

B. Coverage of proteins detected in all the samples. Totally nine samples were processed for proteomics, for each carbon source there were biological triplicates (i.e. 3×3), and the numbers given on either side of the bar plot represent the coverage of proteins in different number of samples.

C. Principal components analysis of demonstrating the variation in protein expression profiles under different growth conditions.

D. Correlation heatmap depicting the replicability between biological triplicates for each growth condition.

Fig. S2. Maximum likelihood phylogenetic tree of 285 lactate dehydrogenase homologues. Phylogenetic tree of 285 LDH homologs identified by searching against 55 499 isolate genomes of the IMG/ER database (as of 2 October 2018). Labels represent the IMG gene ID, IMG annotation for the gene product, taxonomic identity, strain name, an assembly and/or contig/scaffold, which contains this gene. Where several genomes for a species were present, the branch of the

tree was collapsed for clarity. The numbers on branches represent the bootstrap values from 1000 replicates. See methods for details on calculation.

Fig. S3. Comparison of the neighbourhood of genes involved in lactate utilization in genomes of *A. soehngenii* and related lactate-utilizing bacteria.

The red line highlights the gene cluster involved in lactate utilization. Neighbourhoods of genes in other genomes with the same top cluster of orthologs (COG) hit and roughly same matching length are shown below using the IMG gene neighbourhood search. Genes of the same colour (except light yellow) are from the same orthologous group (top COG hit).

A. Shows selected bacterial genomes that share similar gene organization as *A. soehngenii* lactate utilization gene cluster.

B. Shows selected bacterial genomes that share similar gene organization as *Anaerostipes caccae* lactate utilization gene cluster.

Fig. S4. Mining of publicly available metagenomes and metatranscriptomes.

A. Abundance of *lctABCD*FG protein homologues in human gut metagenomes from the human microbiome project.

B. Detection and prevalence of transcripts for *lctABCD*FG gene cluster of *Anaerobutyricum soehngenii* in Simplified Intestinal Microbiota (SIM)-colonized mouse model.