Distribution, organization and expression of genes concerned with anaerobic lactate-utilization in human intestinal bacteria

Author list: Paul O. Sheridan¹, Petra Louis¹, Eleni Tsompanidou², Sophie Shaw³, Hermie J. Harmsen², Sylvia H. Duncan¹, Harry J. Flint^{1*}, Alan W. Walker^{1*}

**These authors contributed equally*

¹Gut Health Group, Rowett Institute, University of Aberdeen, Foresterhill, Aberdeen, Scotland, AB25 2ZD, UK

²Medical Microbiology and Infection Prevention, University of Groningen University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands

³Centre for Genome-Enabled Biology and Medicine, 23 St. Machar Drive, Aberdeen, Scotland, AB24 3RY, UK

Corresponding author: Alan W. Walker: alan.walker@abdn.ac.uk

Keywords: Human gut microbiota; lactate-utilizing bacteria; anaerobic metabolism; upregulation by lactate; transcriptomics

Repositories: Novel draft genomes that were generated for this study are available from GenBank under BioProject PRJNA701799. RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10136.

Abstract

Lactate accumulation in the human gut is linked to a range of deleterious health impacts. However, lactate is consumed and converted to the beneficial short chain fatty acids butyrate and propionate by indigenous lactate-utilizing bacteria. To better understand the underlying genetic basis for lactate utilization, transcriptomic analysis was performed for two prominent lactate-utilizing species from the human gut, *Anaerobutyricum soehngenii* and *Coproccoccus catus*, during growth on lactate, hexose sugar, or hexose plus lactate. In *A. soehngenii* L2-7, six genes of the *lct* cluster including NAD-independent D-lactate dehydrogenase (i-LDH) were co-ordinately upregulated during growth on equimolar D and L-lactate (DL-lactate). Upregulated genes included an acyl-CoA dehydrogenase related to butyryl-CoA dehydrogenase, which may play a role in transferring reducing equivalents between reduction of crotonyl-CoA and oxidation of lactate. Genes upregulated in *C. catus* GD/7 included a six-gene cluster (*lap*) encoding propionyl CoA-transferase, a putative lactoyl-CoA epimerase, lactoyl-CoA dehydratase and lactate permease, and two unlinked acyl-CoA dehydrogenase genes that are candidates for acryloyl-CoA reductase. An i-LDH homolog in *C. catus* is encoded by a separate, partial *lct*, gene cluster, but not upregulated on lactate. While *C. catus* converts three mols of DL-lactate via the acrylate pathway to two mols propionate and one mol acetate, some of the acetate can be re-used with additional lactate to produce butyrate. A key regulatory difference is that while glucose partially repressed *lct* cluster expression in *A. soehngenii*, there was no repression of lactate utilization genes by fructose in the non-glucose utilizer *C. catus*. This implies that bacteria such as *C. catus* might be more important in curtailing lactate accumulation in the gut.

Impact statement

Lactate can be produced as a fermentation by-product by many gut bacteria but has the potential to perturb intestinal microbial communities by lowering luminal pH, and its accumulation has been linked to a range of deleterious health outcomes. Fortunately, in healthy individuals, lactate tends not to accumulate as it is consumed by cross-feeding lactate-utilizing bacteria, which can convert it into the beneficial short chain fatty acids butyrate and propionate. Lactate-utilizing gut bacteria are therefore promising candidates for potential development as novel probiotics. However, lactate-utilizers are taxonomically diverse, and the genes that underpin utilization of lactate by these specialized gut bacteria are not fully understood. In this study we used transcriptomics to compare gene expression profiles of *Anaerobutyricum soehngenii* and *Coprococcus catus*, two prominent lactate-utilizing species in the human gut, during growth on lactate alone, sugar alone, or sugar plus lactate. The results revealed strong upregulation of key, but distinct, gene clusters that appear to be responsible for lactate utilization by these, and other, gut bacterial species. Our results therefore increase mechanistic understanding of different lactate utilization pathways used by gut bacteria, which may help to inform selection of optimal lactate-utilizing species for development as novel therapeutics against colonic microbiota perturbations.

Data summary

Novel draft genomes generated for this study have been made available from GenBank (<https://www.ncbi.nlm.nih.gov/bioproject/>) under BioProject number PRJNA701799. RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-10136. Further details of additional existing genomic data that were analyzed in this project are given in Table 1 and Table S2.

Introduction

The human large intestinal microbiota is dominated by obligately anaerobic bacteria, whose growth is largely dependent on the supply of complex carbohydrates and proteins available for fermentation. These substrates are mostly fermented by resident gut anaerobes to short chain fatty acids (SCFAs) and gases, and in the healthy colon the major faecal SCFAs detected are acetate, propionate and butyrate [1]. Lactate is another fermentation product of many anaerobic bacteria that colonize the mammalian gut, either as a major metabolite, as in *Lactobacillus* and *Bifidobacterium* spp., or as an alternative hydrogen sink [2]. However, as the pKa (3.73) for lactic acid is lower than that of other fermentation acids such as acetate (4.76), accumulation of lactate has the potential to dramatically change the gut environment and gut microbiota composition via reduction in prevailing pH [3].

Lactate may therefore help to inhibit the growth of some pathogenic bacteria with poor tolerance for lower pHs [4] and a gut microbiota dominated by lactate-producing bacteria is found in many healthy pre-weaned infants [5]. On the other hand, there are many known deleterious impacts of lactate accumulation in the gut. Indeed, the phenomenon of lactic acidosis, driven by excessive production of lactate following fermentation of easily digestible dietary carbohydrates, is well known in ruminants and is a major problem in animal husbandry [6, 7]. Similarly, in humans, accumulation of toxic D-lactate is life-threatening in short bowel syndrome [8]) and lactate accumulation is also associated with severe colitis [9]. Additionally, lactate can be utilized as a carbon and energy source by certain intestinal pathogens such as *Salmonella* [10] and *Campylobacter* [11], compounding the potential health detriments of lactate accumulation in the gut.

However, in the colon of healthy adult humans, as well as in the non-acidotic rumen, lactate does not accumulate, due the activities of cross-feeding, lactate-utilizing bacteria [12]. Colonic

lactate concentrations are therefore a balance between bacterial production during fermentation and utilization by lactate-utilizing bacteria, which can use lactate to form acetate, propionate or butyrate [13]. Recent research has linked the populations and activities of these lactate-utilizing bacteria with overall fermentation patterns and productivity [14], and a low abundance of commensal lactate-utilizers has also been proposed to make gut microbial communities inherently less stable and more prone to lactate-induced perturbations [15]. As such, lactate-utilizing bacteria have been proposed as promising candidates for development as novel probiotics [15-18].

The ability to utilize lactate as an energy source for growth appears to be limited to relatively few bacterial species among the human intestinal microbiota, although these species are taxonomically diverse and can utilize lactate in different ways. Selective isolation on DL lactate-containing media resulted in the recovery of *Lachnospiraceae* species subsequently identified as *Eubacterium hallii* (since reclassified as the two species *Anaerobutyricum hallii* and *Anaerobutyricum soehngenii* [19]), *Anaerostipes hadrus* and *Anaerostipes caccae* [20, 21]. These isolates produce butyrate from lactate, with net consumption of acetate, suggesting that they initially convert lactate into pyruvate, which is then routed into the butyrate pathway [21]. The lactate to pyruvate conversion is energetically unfavourable and recent evidence in *Acetobacterium woodii* indicates that it is dependent on electron confurcation [22]. Alternative routes for lactate utilization that are known to be important in the rumen [23] result in propionate formation. These include the acrylate pathway found in the *Negativicutes* species *Megasphaera elsdenii* and in at least one member of the *Lachnospiraceae*, *Coprococcus catus* [24], and the succinate pathway that is found among the *Negativicutes* in *Selenomonas ruminantium* and *Veillonella* spp. [24]. Less advantageously, lactate is also a co-substrate for sulfate-reducing bacteria that use it to form acetate and sulfide, the latter of which may be genotoxic [25].

Here we examine the distribution, organization and regulation of genes involved in lactate utilization among dominant representatives of the human intestinal microbiota, based on new isolates of lactate-utilizing species and on newly available genome sequences. In particular, we identify two gene clusters whose transcription is upregulated during growth with lactate as carbon and energy source. One of these corresponds to the *lct* gene cluster of *A. woodii* [22] that was recently identified from proteomic analysis in *A. soehngenii* [26] while the second encodes activities involved in the acrylate pathway of *Coproccoccus catus*. This investigation reveals differences in the regulation, phylogenetic distribution and metabolic function of these two clusters, and provides novel insights into the mechanistic basis of different lactate-utilization strategies used by human gut bacteria.

Methods

Isolation of new strains of lactate utilizing bacteria

New strains of *Anaerobutyricum soehngenii* (HTF-83D) and *Anaerostipes hadrus* (HTF-920, HTF-146, HTF-370 and HTF-412) were isolated from human faecal samples by dilution and culturing of single colonies on the anaerobic medium YCFA [27] supplemented with glucose. Taxonomic identification of strains was carried out by 16S rRNA gene sequencing and BLASTn querying each sequence against the NCBI 16S rRNA gene reference database. New strains of *Anaerobutyricum* were further classified as either *A. hallii* or *A. soehngenii* by average nucleotide identity (ANI) and average amino acid identity (AAI) using nucmer [28] and CompareM (<https://github.com/dparks1134/CompareM>), respectively (Table S1).

Bacterial strains, growth conditions and genomes

The bacterial strains and genomes used in this work, including both new and previously isolated strains, are described in Table 1 and Table S2. Routine culturing of bacterial strains was in anaerobic M2GSC medium [29] in 7.5 ml aliquots in Hungate tubes, sealed with butyl rubber septa (Bellco Glass). Growth experiments were carried out in basal YCFA medium [27] supplemented with 35 mM lactate and/or 11 mM glucose (or fructose for *C. catus* since this species is incapable of growing on glucose as sole carbon source). All cultures were incubated anaerobically without agitation at 37°C, using the anaerobic methods described previously by Bryant [30].

Species	Strain	Donor	Genome Accession	Contigs	CDS number	Genome Reference	Strain Reference
<i>Anaerobutyricum soehngenii</i>	L2-7	1	GCA_900209925	1	3005	[31]	[32]
	SL6/1/1	2	JAFIQP000000000	498	2900	This study	[33]
	HTF-83D	3	JAFIQO00000000	262	2809	This study	This study
<i>Anaerobutyricum hallii</i>	SM6/1	2	JAFIQO00000000	341	3202	This study	[33]
	DSM3353	4	GCA_900209925	175	2829	Unpublished	[34]
<i>Anaerostipes hadrus</i>	HTF-920	7	N/A	N/A	N/A	N/A	This study
	SS2/1	2	GCA_000154545	34	3022	Unpublished	[33]
	SSC/2	2	GCA_000210695	120	3170	Unpublished	[33]
<i>Anaerostipes caccae</i>	DSM14662	8	GCA_000154305	26	3460	Unpublished	[32]
<i>Coproccoccus catus</i>	GD/7	9	GCA_000210555	1	3261	Unpublished	[33]

Table 1. New and existing strains used for *in vitro* cultivation studies in this work, and their corresponding genomes. “N/A” indicates no genome available for this strain, which was used for *in vitro* work only. The full list of strains used for genomic-based analyses is shown in Table S2.

DNA isolation, genome sequencing and annotation

Genomic DNA was extracted with the Ultraclean Microbial DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). The DNA concentration and purity were measured using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Qubit double-stranded DNA (dsDNA) HS and BR assay kits (Life Technologies, Carlsbad, CA, USA). One nanogram of bacterial DNA was used for library preparation. The DNA library was prepared using the Nextera XT library preparation kit with the Nextera XT v2 index kit

(Illumina, San Diego, CA, USA). The library fragment length was aimed at fragments with a median size of 575 bases and was assessed with the Genomic DNA ScreenTape assay with the 2200 Tape-Station system (Agilent Technologies, Waldbronn, Germany). Subsequently, the library was sequenced on an Illumina MiSeq sequencer, using a 2x250 (500v2) cartridge, using the MiSeq reagent kit v2 generating 250-bp paired-end reads, (Illumina, San Diego, CA, USA). MiSeq data were processed with MiSeq control software v2.4.0.4 and MiSeq Reporter v2.4 (Illumina, San Diego, CA, USA). The resulting fastq files were filtered using TrimGalore v0.4.0 (<https://github.com/FelixKrueger/TrimGalore>), removing one nucleotide off the 3' end (--trim1) and removing both pairs of the paired end reads if one read did not pass filtering (--paired). Filtered fastq files were assembled using the idba_ud assembler v1.1.3 [35] using standard settings. Newly generated genome sequences are available from GenBank (BioProject PRJNA701799), with individual accession numbers shown in Table 1 and Table S2. Additional sequences were retrieved from GenBank (accession numbers in Table S2). New genome sequences were annotated using Prokka v1.11 [36] and publicly available sequences were reannotated in the same manner to ensure uniformity between annotation methods. Genome completeness was estimated using CheckM v1.1.3 [37].

Genome analyses

Lactate utilization loci were searched for using the Hidden Markov Model (HMM) profiles of conserved domains found in three-component L-lactate hydrogenase (PF02754, PF02589, PF13183, PF11870), FAD-dependent D-lactate dehydrogenase (PF01565, PF02913, PF09330, PF12838, PF02754), FMN-dependent L-lactate dehydrogenase (PF01070, PF00173), NAD-dependent D-lactate dehydrogenase (PF00389, PF02826), NAD-dependent L-lactate dehydrogenase (PF02866, PF00056), lactate permease (PF02652), lactate racemase (PF09861)

and ETF (PF00766, PF01012) using hmmsearch in HMMER3 [38]. Butyrate and propionate producing pathways were detected by BLASTP against UniProt [39] and KEGG [40] with a threshold of 60% identity using the marker genes butyrate kinase and butyryl-CoA:acetate CoA-transferase as indicators of butyrate production, and propanediol utilization protein (PduP), lactoyl CoA dehydratase alpha-subunit (LcdA) and methyl-malonyl-CoA decarboxylase alpha-subunit (MmdA) as indicators of propionate production through the propanediol, acrylate and succinate pathways, respectively [24]. Rnf complexes were detected by query against amino acid sequences of the experimentally proven Rnf complexes of *Clostridium ljungdahlii* and *A. woodii* [41, 42].

All genes containing the lactate permease Pfam [43] domain PF02652 in UniProt (June 2020) [39] were downloaded and clustered into groups of >70% similarity using CD-HIT v4.8.1 [44]. Representatives of each cluster were aligned using MAFFT L-INS-I v7.407 [45]. The acyl-CoA dehydrogenase in the *lct* cluster (L2-7_01909) was used as a BLASTP query against *C. catus* GD/7 and the *A. hadrus* genomes, homologs were aligned using MAFFT L-INS-i [45] and a HMM profile was created from the alignment using hmmbuild in HMMER3 [38]. This profile was then queried against the genomes of known lactate utilizers using hmmsearch (-T 80) to detect divergent members of the acyl-CoA dehydrogenase protein superfamily. Matching sequences were aligned using MAFFT L-INS-i [45] and the alignment was manually inspected for likely mis-annotations (none were found). For both the lactate permease and acyl-CoA dehydrogenase alignments, spurious sequences and poorly aligned regions were removed with trimAl (automated1, resoverlap 0.55 and seqoverlap 60) [46] and maximum likelihood trees were constructed with IQ-TREE [47] using the best fitting protein model predicted in ModelFinder [48]. Branch supports were computed using the SH-aLRT test [49] for the lactate permease tree and ultrafast bootstraps for the acyl-CoA dehydrogenase tree. The tree figures were generated using iTOL [50].

Differential gene expression analysis

Fresh YCFA broth supplemented with 11 mM monosaccharide (glucose for *A. soehngenii* and fructose for *C. catus*), 35 mM lactate or both 11 mM monosaccharide and 35 mM lactate were inoculated 1:75 with an overnight grown culture of either *A. soehngenii* L2-7 or *C. catus* GD/7. These cultures were grown to OD₆₅₀ 0.3 before cultures were centrifuged to cell pellets (5 min, 1,200 g, 20 °C). The supernatant was transferred to a new tube for SCFA and monosaccharide analysis, and the cell pellet was resuspended in 500 µl of RNeasy lysis buffer (Invitrogen™), left at room temperature for 5 mins and frozen at -70 °C. Total RNA was extracted using the RNeasy PowerMicrobiome Kit (QIAGEN), following the manufacturer's instructions. This included a bead beating step for lysing Gram-positive cell walls. Ribosomal RNA was depleted using the Illumina Ribo-Zero rRNA removal kit. Libraries were prepared using the Illumina TruSeq stranded mRNA library kit and sequenced at the Centre for Genome-Enabled Biology and Medicine (CGEBM) at The University of Aberdeen using a High Output 1x75 kit on the Illumina NextSeq platform producing 75 bp single end reads. Reads were combined from two runs producing between 14237482 and 16008520 reads per sample. These transcriptomic datasets are available from the ArrayExpress database at EMBL-EBI under accession number E-MTAB-10136. Quality of raw reads was assessed using FastQC v0.11.3 [51] and sequences were filtered using TrimGalore v0.4.0. Between 14141677 and 15921037 sequences remained per sample, which, on average, was 99% of the raw reads. The appropriate reference genomes for *A. soehngenii* L2-7 and *C. catus* GD/7 (Table S2) were indexed in preparation for alignment using the HISAT2 build function (version 2.1.0) [52]. Filtered reads were then aligned to the appropriate reference genome using HISAT2 v2.1.0 [52] with default settings. Alignments were then converted to BAM format and sorted using SAMtools v1.2 [53]. Reads aligned to gene regions were counted using featureCounts (subread package version 5.0-pl) [54] with settings to split multi-mapping reads as a fraction between aligned regions. Genes with

significant differential expression ($FDR < 0.05$) between conditions were identified using edgeR v3.16.5 [55], utilizing the GLM function, as more than two growth conditions were used in this study.

Short chain fatty acid, monosaccharide and ethanol analysis

SCFA concentrations were measured by gas chromatography (GC) as described previously (Richardson *et al.*, 1989). In brief, following derivatization of the samples using N-tert-butyltrimethylsilyl-N-methyltrifluoroacetamide, the samples were analysed using a Hewlett Packard gas chromatograph (GC) fitted with a silica capillary column using helium as the carrier gas. L-lactate concentration was assessed by Lactate Reagent (Trinity biotech) using a Konelab 30 chemistry analyser. D-lactate was calculated as the difference between total lactate and L-lactate. Methanol, ethanol, propanol, butanol and pentanol concentrations were also measured by GC using a ZB WAX column. Glucose and fructose concentration were assessed using the glucose hexokinase and fructose hexokinase assays for Konelab.

Results

Lactate utilization in human colonic bacteria

Some anaerobic colonic bacteria are capable of utilizing lactate as a carbon and energy source. This process can generate the SCFAs butyrate, propionate or acetate depending on the species [21, 24]. Previous studies have been based on only a small number of available strains, but the isolation and genome sequencing here of new strains of lactate-utilizing species allowed for more in-depth analysis. Each strain was grown in culture provided with glucose (or fructose in the case of *C. catus* GD/7), DL-lactate, or a combination of the two as energy sources (Figure

1). All nine strains identified as either *Anaerostipes* or *Anaerobutyricum* spp. showed the ability to consume lactate and acetate with production mainly of butyrate, as previously reported [21]. This contrasts with production mainly of propionate from lactate by *C. catus* (Figure 1). The *A. hadrus* strains tested showed little or no ability to utilize L-lactate, whereas the other species utilized both isomers (Figure 1B). Lactate-utilization was repressed by the presence of hexose sugar in the growth medium to varying extents for the different *Anaerobutyricum* and *A. hadrus* strains, with utilization of both lactate stereoisomers particularly repressed by the presence of glucose for *A. soehngeni* strains L2-7 and SL6/1/1 (Figure 1B). Interestingly, although propionate was the main SCFA produced by *C. catus* GD/7 when grown on lactate some butyrate and acetate production was also observed, indicating the presence of these two SCFAs as intermediate- or end-products of lactate utilization in this organism.

Figure 1. Fermentation profiles of lactate-utilizing bacteria grown on sugar, on lactate, or a combination of both. Panel A shows the net production (to the right of the figures), or consumption (to the left of the figures) of various short chain fatty acids by lactate-utilizing bacteria grown on either hexose sugar alone, lactate alone, or a combination of both. Cultures were incubated anaerobically for 24 h in media supplemented with 11 mM glucose, 35 mM lactate or both. Glucose was substituted with 11 mM fructose for *C. catus* GD/7. Iso-butyrate, valerate, iso-valerate, and succinate were not detected in any of the samples. Panel B shows net consumption or production of L- or D-lactate by the tested strains, with or without simultaneous exposure to hexose sugars in the culture medium. Error bars are the SEM of three biological replicates. There was no remaining glucose or fructose in any culture, with the exception of *A. soehngenii* L2-7 (2 mM, SEM 2.0) and SL6/1/1 (6 mM, SEM 3.2). Expanded data are shown in Table S3.

285

286 *Changes in gene expression during growth on DL-lactate*

287 To investigate the genetic control and regulation of lactate utilization, transcriptomic analysis
 288 was performed on two prevalent lactate-utilizing species from the human gut, *A. soehngenii*
 289 and *C. catus*. This involved assessing the differential expression of genes during growth on a
 290 hexose sugar (glucose for *A. soehngenii* and fructose for *C. catus*), on DL-lactate or on a
 291 combination of sugar plus DL-lactate (Figure S1). Growth was carefully monitored in each of
 292 six replicate experiments for each strain to ensure sampling during mid-exponential growth
 293 phase (OD₆₅₀=0.3). RNA was extracted from cell pellets and the culture supernatant was
 294 analysed for their SCFA profiles and remaining hexose (Figure 2). In agreement with the
 295 results shown in Figure 1, in the absence of glucose *A. soehngenii* L2-7 had utilized much of
 296 the lactate at the time of sampling, but lactate utilization was significantly repressed in the
 297 presence of glucose ($P < 0.001$). In the absence of hexose sugar *C. catus* GD/7 had also utilized
 298 much of the lactate, but utilization was slightly lower in the presence of fructose ($P < 0.01$).
 299 Sugar was still present in the growth media initially supplemented with it (Figure 2), indicating
 300 there was potential for repression at the point of RNA harvesting.

301

Figure 2. Fermentation end products of *A. soehngenii* L2-7 and *C. catus* GD/7 during growth on lactate, a monosaccharide alone or on both in combination. Cultures were harvested at an OD₆₅₀ reading of 0.3. Iso-butyrate, valerate, iso-valerate and succinate were not detected in any of the samples. Error bars are the SEM of six biological replicates.

The results from the experiments with *A. soehngenii* L2-7 showed differences in overall transcriptomic profile were observed during growth on lactate compared to glucose alone or a combination of glucose and lactate (Figure 3). Conversely, little overall difference in transcriptomic profile was observed between glucose and the combination of glucose and lactate, with only 24 differentially expressed genes as opposed to 1427 between glucose and lactate as a sole carbon source. In *C. catus* GD/7, the overall transcriptomic profile was again strongly influenced by the growth substrate (Figure 3). In *C. catus* GD/7, however, and in contrast to *A. soehngenii* L2-7, lactate-induced genes were not simply repressed by the presence of monosaccharide (fructose), with all three transcriptome profiles differing from each other.

Figure 3. Volcano plots of log fold gene expression change vs. $-\log_{10}$ of the false discovery rate value and multidimensional scaling plots showing the gene expression distribution of samples during growth on lactate, monosaccharide or a combination of the two. Genes with a significant FDR value <0.05 in the volcano plots are highlighted in red. A more detailed overview is given in Table S4. “Combined” indicates cultures incubated with both lactate and hexose.

<i>A. soehngeni</i> L2-7		log2-fold change		
		G v L	G v LG	L v LG
L2-7_01905	Lactate permease (<i>lctE</i>)	-14.3	-4.4	10.0
L2-7_01906	D-iLDH (<i>lctD</i>)	-14.2	-3.7	10.5
L2-7_01907	ETF-B (<i>lctB</i>)	-14.1	-3.5	10.6
L2-7_01908	ETF-A (<i>lctC</i>)	-14.2	-3.7	10.5
L2-7_01909	Acyl-CoA dehydrogenase (<i>lctG</i>)	-14.1	-3.7	10.4
L2-7_01910	Lactate racemase (<i>lctF</i>)	-14.0	-3.5	10.5
L2-7_01911	LutR transcriptional regulator (<i>lctA</i>)	-3.0	NoDiff	2.5
L2-7_02157	Aquaporin family protein	-6.3	-9.6	-3.3
L2-7_02158	Racemase	-5.3	-9.0	-3.7
L2-7_02594	Putative cyclic lactone autoinducer peptide	-5.2	NoDiff	5.8
L2-7_02595	Hypothetical protein	-4.9	NoDiff	4.9
L2-7_02596	RNA polymerase subunit sigma-70	-5.7	NoDiff	6.0

L2-7_02597	Helix-turn-helix domain-containing protein	-5.7	NoDiff	5.8
L2-7_00821	Hnd complex, subunit A	-4.33	NoDiff	5.12
L2-7_00822	Hnd complex, subunit C	-4.07	NoDiff	4.56
L2-7_00823	Hnd complex, subunit D	-3.98	NoDiff	4.38
L2-7_00377	Rnf complex, subunit C	-2.90	NoDiff	2.91
L2-7_00378	Rnf complex, subunit D	-3.08	NoDiff	2.98
L2-7_00379	Rnf complex, subunit G	-3.20	NoDiff	3.13
L2-7_00380	Rnf complex, subunit E	-3.33	NoDiff	3.25
L2-7_00381	Rnf complex, subunit A	-3.27	NoDiff	3.22
L2-7_00382	Rnf complex, subunit B	-3.24	NoDiff	3.15

<i>C. catus</i> GD/7		log2-fold change		
		F v L	F v LF	L v LF
GD-7_01094	Propionyl CoA transferase (<i>lapA</i>)	-10.3	-9.2	1
GD-7_01095	Lactoyl CoA epimerase (<i>lapB</i>)	-11.8	-10.7	1.1
GD-7_01096	Lactoyl CoA dehydratase subunit (<i>lapC</i>)	-12.6	-11.4	1.2
GD-7_01097	Lactoyl CoA dehydratase subunit (<i>lapD</i>)	-13.1	-11.9	1.2
GD-7_01098	Lactoyl CoA dehydratase subunit (<i>lapE</i>)	-13	-11.9	1.2
GD-7_01099	Lactate permease (<i>lapF</i>)	-13	-11.9	1.2
GD-7_00065	3-hydroxybutyryl-CoA dehydratase	-8.6	-6.2	2.5
GD-7_00066	H ⁺ /gluconate symporter and related permeases	-6.9	-4.6	2.3
GD-7_00760	Amidohydrolase family protein	-7.6	-4.4	3.2
GD-7_00761	MFS transporter	-8.3	-4.7	3.6
GD-7_01087	Acyl-CoA dehydrogenase	-9.6	-6	3.6
GD-7_01088	ABC transporter substrate-binding protein	-6.6	-3.5	3.1
GD-7_01551	Acyl-CoA dehydrogenase	-11.7	-10.8	1

326

327 **Table 2. Selected genes with high expression differences during growth on glucose (G),**
328 **fructose (F), lactate (L) or a combination of substrates (LG/LF).** All transcripts are listed
329 that exhibited >5 log2-fold changes in expression, together with expression changes for linked
330 genes within the same gene cluster. Changes of less than 5-fold are listed for the Rnf and Hnd
331 genes clusters in *A. soehngenii* because of their metabolic relevance. All listed genes have an
332 FDR value of <0.05. NoDiff = No significant difference in expression. Expanded to all genes
333 in Tables S5-S10.

In *A. soehngensis* L2-7, eleven transcripts showed >5 Log2-fold increase in expression during growth on lactate alone compared with growth on glucose alone, and increases were lower when glucose was present alongside lactate (Table 2). Six transcripts (L2-7_01905-01910) that showed the highest coordinate upregulation ($\log_2 > 14$) were encoded by the *lct* gene cluster that includes an NAD-independent lactate dehydrogenase (iLDH), lactate permease, ETF electron transport factor (alpha and beta subunits), lactate racemase and an acyl-CoA dehydrogenase (Table 2). The linked regulatory gene *lctA* showed a lower amplitude of induced expression. A second cluster of three genes that showed the same behaviour, also at lower amplitude, included an autoinducer, RNA polymerase sigma factor and HTH regulatory protein. Two further transcripts, from a second lactate racemase and closely linked aquaporin genes, were also highly induced by lactate, but were not repressed by glucose. The components of the Rnf complex, which couples the oxidation of reduced ferredoxin to the reduction of NAD [41], was highly induced by lactate and repressed by glucose, as were the components of the Hnd complex, which catalyzes the reduction of NADP in the presence of molecular H₂ to yield NADPH [56, 57].

In *C. catus* GD/7, seven genes showed >10 Log2 fold increased expression during growth on lactate alone compared with fructose alone (Table 2). Six of these were encoded by a cluster of linked genes, but this cluster showed no sequence similarity with the *lct* cluster of *A. soehngensis* described above and comprised genes (GD-7_01094-01099) encoding three components of lactoyl-CoA dehydratase, propionyl-CoA transferase, a methyl-malonyl-CoA epimerase homolog and a lactate permease (Table 2). These genes encode most of the activities required for the acrylate pathway to convert lactate into propionate, consistent with previous evidence using isotopically labelled lactate showing that this bacterium uses the acrylate pathway for lactate utilization [24]. This locus is thus a cluster encoding lactate utilization via

the acrylate pathway (*lap*). It seems likely that the methyl-malonyl-CoA epimerase homolog (*lapB*) functions as a lactoyl-CoA epimerase, catalyzing the reversible conversion of L-lactoyl-CoA to D-lactoyl-CoA. The upregulated lactate permease gene was one of four encoded at separate locations in the *C. catus* GD/7 genome (Table S11). These six genes, together with two unlinked genes (GD-7_01551, GD-7_01087) encoding acyl-CoA dehydrogenases, were also highly upregulated on lactate when fructose was present. The two acyl-CoA dehydrogenases are candidates for propionyl-CoA dehydrogenase (acrylate reductase) activity.

Wider occurrence of lactate-utilization gene clusters in the genomes of human colonic bacteria

Homologs of the putative lactate utilization genes identified in *A. soehngenii* L2-7 and *C. catus* GD/7 were also detected in other known lactate-utilizing bacteria from the human colonic microbiota and other environments (Figure 4A & B, Table S12 & S13). Homologs of the upregulated *lct* genes from *A. soehngenii* L2-7 were present in the two species of *Anaerobutyricum*, *Anaerostipes caccae*, *Eubacterium limosum*, *C. catus* and *Megasphaera elsdenii* (Table S12) (although as noted above they were upregulated only moderately in *C. catus* (Table S8)). The genome for *A. hadrus* DSM3319 lacked the racemase gene, consistent with the observation that it consumed only the D isomer of lactate (Figure 1B). *Lct* homologs were also present in the *Acetobacterium woodii* gene cluster, but this cluster does not contain the putative acyl-CoA dehydrogenase (ACoAD) gene *lctG*.

Homologs of the six clustered *lap* cluster genes were detected in *Megasphaera elsdenii*, another species known to use the acrylate pathway for lactate utilization, although the percentage identity was very low for the lactate permease gene (Table S13). Five of the six *lap* cluster genes, were detected in *Anaerotignum propionicum*, excluding *lapB*. This supports the

prediction of *lapB* as a lactoyl-CoA epimerase, as *Anaerotignum propionicum* cannot utilize L-lactate [58].

In most butyrate-producing *Firmicutes* species from the human colon, most or all of the genes encoding the central butyrate pathway are present in a single gene cluster, although the butyryl acetate:CoA transferase is separately encoded (Louis & Flint 2009). This was found to be true for *A. soehngenei* but not for *C. catus* GD/7, where the key functions are encoded in at least three separate locations in the genome (Table S14). While the crotonase and beta-hydroxybutyrate dehydrogenase genes are linked in *C. catus* GD/7, two candidate thiolase genes are unlinked. Meanwhile, the butyryl-CoA dehydrogenase (BCD) in the *C. catus* GD/7 genome (GD-7_01569) is situated downstream of the i-LDH (GD-7_01568) and upstream of two ETF genes (GD-7_01570, 01571) (Figure 4A, Table S14). These appear to be homologous with the *lctBCFG* genes of *A. soehngenei*, although these *lct* genes were not upregulated by lactate in *C. catus*.

A. *lct* cluster genes

A. soehngenii (L2-7_01905-01911)
A. hallii
A. caccae
A. hadrus
A. woodii
C. catus

B. *lap* cluster genes

C. catus (GD-7_01094-01099)
M. elsdenii
A. propionicum

Figure 4. Organization of lactate utilization loci observed in (A) *A. soehngenii* and (B) *C. catus* in other previously described lactate-utilizing bacteria. The black line in the *A. propionicum* *lap* cluster indicates that *lapF* gene is present in a different region of the genome. Lactate permease (Lperm), NAD-independent D-lactate dehydrogenase (D-iLDH), electron transfer flavoprotein (ETF α & β), acyl-CoA dehydrogenase (ACoAD), lactate racemase (Lrce), lactate operon transcriptional regulator (Regulator) propionyl-CoA transferase (PCoAT), lactoyl CoA epimerase (LCoAE), lactoyl-CoA transferase (LCoAD, act, α & β). Expanded to more lactate utilizing bacteria genomes in Tables S12-13.

Of the lactate-utilizing bacteria analyzed, only *Veillonella parvula*, *Bacillus subtilis* and *Campylobacter jejuni* appear to lack the Rnf complex, and all genomes except *C. jejuni* and *Shewanella oneidensis* encoded genes indicative of butyrate and/or propionate production (Figure 5). For bacteria that did not encode L-iLDHs, the lack of a lactate racemase corresponded with an inability to utilize L-lactate. As described earlier, the *lct* and *lap* clusters

410 use different mechanisms of lactate utilization and are present in a different range of organisms.

411 Additionally, *Shewanella oneidensis* uses a third mechanism of lactate utilization, involving a

412 very different lactate dehydrogenase [59]. Homologs of this three-component L-lactate

413 hydrogenase are present in the genomes of *B. subtilis*, *C. jejuni*, *E. coli* and *V. parvula* (Figure

414 5). Other types of NAD-independent L- and D-lactate dehydrogenases have also been found in

415 organisms from all three domains of life [60]. Thus, there appear to be several divergent

416 paradigms of lactate utilization in bacteria. However, all of these mechanisms appear to require

417 members of the lactate permease superfamily and most possess some member of the acyl-CoA

418 dehydrogenase superfamily. Therefore, the diversity of these proteins warranted detailed

419 further investigation.

Phylogenetic analysis of 17653 lactate permeases from bacteria, archaea and eukaryotes revealed five families of lactate permease: LP-I and LP-II, which were present exclusively in bacteria; LP-III and LP-IV, which were present in bacteria and archaea; and LP-V which was present exclusively in eukaryotes. The lactate permeases identified in the transcriptomics work presented here belong to LP-IV, which was further divided into five subfamilies (A-E). Three of the four lactate permease genes in the *C. catus* GD/7 genome, including the copy upregulated on lactate, clustered with the *Anaerotignum* and *S. oneidensis* lactate permeases in LP-IV-C (Figure 6). The fourth clustered with the lactate-induced *A. soehngenii* lactate permease in LP-IV-E. This subfamily notably also contains the lactate permeases of *Anaerostipes* spp., *Megasphaera* spp., *Veillonella* spp., *Acetobacterium* spp. and *E. limosum*.

Lactate permeases are also notably absent from the genomes of the gut bacteria, *Eubacterium rectale* A1-86, *Roseburia intestinalis* L1-82, *Roseburia hominis* A2-183, *Coprococcus eutactus* L2-50 and *Faecalibacterium praunsnitzii* A2-165, which do not utilize lactate [21], thus indicating a role for lactate permease as a marker gene of lactate utilization, despite disparate mechanisms of lactate oxidation.

Figure 6. Phylogeny of lactate permease families. A) Families of lactate permease, and B) Sub-families of LP-IV. The conserved lactate permease domain PF02652 was detected in 17653 protein sequences from Uniprot (June 2020). These proteins were formed into 1627 clusters of 70% identity. A maximum-likelihood tree was constructed from the representative sequences of these clusters using the LG+F+R8 model. All branches shown have >70% support of 1000 SH-*alrt* replicates. Proteins were clustered into protein families and sub-families using the average branch length distances to the leaves in the phylogenetic tree. The tree was rooted using minimal ancestor deviation. Lactate permeases of *C. catus* GD/7 and *A. soehngenii* L2-7 are listed underneath the subfamilies of which they belong to. The upregulated lactate permeases detected in the preceding transcriptomics work in this study are indicated in red font. Representative genes of each family of lactate permease and each sub-family of LP-IV are listed in Table S15 and Table S16, respectively.

Phylogeny of Acyl-CoA dehydrogenase superfamilies

The acyl-CoA dehydrogenase in the *lct* cluster (L2-7_01909) and its homologs in the *C. catus* and *A. hadrus* genomes were used to create a HMM profile to detect divergent members of this protein superfamily in the genomes of other lactate-utilizing bacteria. These genes were found

to be ubiquitous in lactate-utilizing bacteria, with many genomes possessing more than one copy (Table 3). Analysis of these genes revealed that they can be divided into a multitude of clades (Figure 7), many of which, such as the experimentally proven [69] caffeyl-CoA reductase of *A. woodii*, are the only representatives of their clade in our dataset. One clade contained the butyryl-CoA dehydrogenase present in the gene cluster responsible for butyrate production from carbohydrates [70]. This clade also contained the lactate induced *lct* cluster acyl-CoA dehydrogenase, indicating that it too is a butyryl-CoA dehydrogenase. The highly lactate-induced acyl-CoA dehydrogenases of *C. catus* (GD-7_01551 and GD-7_01087) form separate clades (acyl-CoA dehydrogenase families 2 and 4, respectively) with genes from several other species. No genes of these clades have been experimentally characterized, complicating functional prediction. Additionally, genes from *C. catus*, *A. soehngenii* and *A. caccae* form a clade with the experimentally proven [71] acryloyl reductase of *Anaerotignum propionicum*.

A. soehngenii L2-7 and *C. catus* GD/7 possess four and five acyl-CoA dehydrogenases, respectively (Figure 7, Table 3). Two of the *A. soehngenii* L2-7 gene products group with butyryl-CoA dehydrogenase (BCD) enzymes, including the BCD in the central butyrate pathway (L2-7_00754) and the upregulated *lctG* product (L2-7_01919). This strongly suggests that the lactate-induced *lctG* gene product is an alternative BCD enzyme that links reduction of crotonyl-CoA to the oxidation of lactate when lactate is being utilized.

Figure 7. Phylogeny of Acyl-CoA dehydrogenase genes present in lactate-utilizing bacteria. Branch validation was performed using 1000 ultrafast bootstrap replicates and a hill-climbing nearest neighbor interchange (NNI) search was performed to reduce the risk of overestimating branch supports. Branches with less than 90% UFBoot support were collapsed. Genes labelled “Caffeyl-CoA reductase” and “Acryloyl reductase” were experimentally verified by Bertsch *et al.* [69] and Hetzel *et al.* [71], respectively, and the gene labelled “Butyryl-CoA dehydrogenase” is present in the butyrate production loci as described previously [70]. This maximum-likelihood tree was reconstructed using the LG+G4 model. *Anaerobutyricum soehngenii* L2-7 and *Coproccoccus catus* GD/7 genes are represented in red and green font, respectively. Asterisks indicate genes that were highly upregulated by lactate in the transcriptomics work presented in this study.

Acyl-CoA dehydrogenase families						
Species	1	2	3	4	5	# Genomes queried
<i>Coprococcus catus</i>	2	1 (1)	0	1	1 (1)	1
<i>Anaerobutyricum soehngenii</i>	0/1	0	0	0/1	2 (1)	3
<i>Anaerobutyricum hallii</i>	0	0	0	0	2	5
<i>Anaerostipes caccae</i>	1	1	1	2	1	1
<i>Anaerostipes</i> sp. 3-2-56FAA	1	1	1	2	1	1
<i>Anaerostipes hadrus</i>	0	0/1	1	0/1/2	1	12
<i>Eubacterium limosum</i>	0	0	0	0/1	0/2	3
<i>Megasphaera elsdenii</i>	0	0	2	0	0	1
<i>Escherichia coli</i>	0	0	0	1	0	1
<i>Anaerotignum propionicum</i>	1	1	0	0	1	1

Table 3. Acyl CoA dehydrogenase families of lactate-utilizing bacteria. These enzymes are ubiquitous in lactate-utilizing bacteria, with some genomes possessing multiple genes. Many of these genes form five distinct gene families (Figure 7). The number genes of a given family that were upregulated during growth on lactate in *A. soehngenii* L2-7 and *C. catus* GD/7 is shown in parentheses. Black slashes indicate differences in gene copy number between queried genomes of the same species. Expanded to include more lactate-utilizing bacteria genomes in Table S17.

Discussion

Lactate can be utilized by many aerotolerant bacteria through its conversion to pyruvate via D- or L- lactate oxidases [60, 72]. This reaction is not available to obligate anaerobes however, and different mechanisms are therefore required in the absence of oxygen. As noted previously, anaerobic conversion of lactate to pyruvate is energetically unfavourable [22] and this ability appears to have a limited phylogenetic distribution. Here we have identified two different six-gene clusters (*lct* and *lap*) whose transcriptional expression is highly upregulated during growth on lactate as the sole added carbon and energy source. These two gene clusters correspond to two entirely different mechanisms for lactate utilization. The upregulated cluster in *A. soehngenii* corresponds to five genes of the *lct* gene cluster identified previously in

Acetobacterium woodii, which is an acetogen that has been shown to convert lactate to pyruvate via a mechanism involving electron confurcation [22]. Products of this *lct* gene cluster were recently detected by proteomic analysis [26], and the cluster reported in the related lactate-utilizing species *A. hallii*, *A. rhamnosivorans* and *Anaerostipes caccae*. Five of the six clustered genes are also present in *Anaerostipes hadrus*, but all strains of this species with sequenced genomes examined here lacked the lactate racemase (*lctF*), thus explaining why they can only utilize D-lactate.

Our analysis confirms that the *lct* gene cluster is present in the newly sequenced strains of *A. soehngenii*, *A. hallii* and *A. hadrus* described here. Notably, however, the *lct* cluster in all these strains encodes an acyl-CoA dehydrogenase specified by the *lctG* gene that is lacking from the *A. woodii* cluster (Schoelmerich, Katsyv et al. 2018). Since the *lctG* gene product is related to, but distinct from, the BCD gene encoded by the central butyrate pathway cluster [70], it seems likely that it is an alternative butyryl-CoA dehydrogenase (BCD) that acts to reduce crotonyl-CoA when lactate is the growth substrate (Figure 8). While the conversion of pyruvate to acetyl-CoA can produce enough reduced ferredoxin to balance the oxidation of lactate via the D-iLDH/ETF complex [73], redox balance and net energy production are also linked to the operation of the butyrate cycle (Figure 8). The upregulation by lactate of an orthologous enzyme catalyzing the BCD reaction, encoded by *lctG*, may therefore play a critical role. The exact relationships between the co-ordinately upregulated i-LDH, BCD and ETF proteins are not known for *A. soehngenii*, although it has been suggested that in *C. butyricum* the LDH-ETF and BCD-ETF complexes act separately [74].

It has been suggested that the molecular hydrogen producing-Hnd complex is involved in balancing redox states during lactate oxidation in *Desulfovibrio* species [57], and its induction during growth of *A. soehngenii* on lactate indicates a likely mechanism for the lactate oxidation-coupled hydrogen production observed in this species [21]. However, resolution of

in both the *lap* and *lct* clusters, their sequences are only distantly related to each other. In addition, it seems likely that one of the two unlinked acyl-CoA dehydrogenases that were also upregulated by lactate must correspond to propionyl-CoA dehydrogenase that, along with ETF proteins, carries out the acryloyl-CoA reductase reaction [71].

Operation of the acrylate pathway is normally assumed to result in the conversion of three mol lactate to two mol propionate and one mol of acetate (Figure 9A) [2]. The acetate arises through conversion of lactate to pyruvate via i-LDH, with the reducing equivalents balancing those needed to reduce two acrylate to two propionate. Our results however indicate the formation of some butyrate, in addition to propionate, during growth on lactate. This can be explained readily if we assume that some of the acetate produced is converted along with additional lactate to butyrate (Figure 9B), as happens in *A. soehngeni*. The finding that the i-LDH gene is linked to BCD and ETF genes in *C. catus* GD/7 appears consistent with this proposal. An unexplained feature is the 3-OH butyryl-CoA dehydratase transcript (GD-7_00065) that was upregulated by lactate, although this might possibly be an ortholog of the second β -hydroxybutyryl-CoA dehydrogenase gene (GD-7_01384) that is linked to crotonase in *C. catus* GD/7.

Figure 9. Proposed mechanisms of (A) propionate and acetate formation from lactate and (B) butyrate formation from acetate and lactate in *C. catus* GD/7. NAD-independent D-

lactate dehydrogenase (D-iLDH), electron transfer flavoprotein (ETF), ferredoxin (Fd), thiolase (THL), β -hydroxybutyryl-CoA dehydrogenase (BHBD) and crotonase (CRO) and butyryl-CoA dehydrogenase (BCD). Locus tags are in italics and are colored based on log2 fold change of their expression during growth on lactate versus fructose. Asterisks indicate most likely candidate genes.

Homologs of *lap* genes were detected in a few other bacterial species, but notably were present in *Megasphaera elsdenii*, which is also known to use the acrylate pathway for lactate utilization. Indeed, *M. elsdenii* is known to convert lactate to butyrate under conditions of carbon-limited growth [66]. It is not known whether the same applies in *C. catus*. *M. elsdenii* also encodes activities involved in the formation of propionyl-CoA from pyruvate via malate and succinate [75].

Two other anaerobic lactate-utilizing species, *Veillonella parvula* and *Anaerotignum propionicum*, showed high levels of sequence identity with several genes of the *lct* cluster, but lacked a close homolog of the D-iLDH. The latter species however also carries homologs for most of the *lap* cluster and utilizes lactate via the acrylate pathway [76], whereas *V. parvula* encodes a three component L-lactate dehydrogenase, homologous to that of *S. oneidensis* (Figure 5), and uses the ‘randomizing’ succinate pathway rather than the acrylate pathway in producing propionate [2].

In conclusion, we have shown that two distinct routes for anaerobic lactate fermentation in obligate anaerobes, via pyruvate to butyrate and via acrylate to propionate, both involve major transcriptional upregulation of the key genes involved (the *lct* cluster in *A. soehningii* and the *lap* cluster in *C. catus*) in response to lactate. There was however a clear difference in the impact of hexose sugars on gene expression. Whereas *A. soehningii* *lct* expression was subject to partial repression by glucose, this was not the case for *C. catus* acrylate genes with fructose.

The lack of repression of lactate utilization genes by hexose in *C. catus* may give this and other species with similar metabolism, such as *Megasphaera elsdenii*, particularly important roles in controlling lactate concentrations within gut communities under conditions where there are significant concentrations of free carbohydrates that could be used as alternative energy sources. Indeed, these two species appear to be implicated in accounting for differences in animal productivity that are associated with lactate utilization in the ruminants [14]. The contribution of *Anaerobutyricum* species to lactate utilization in the human colon, where they are relatively abundant within the microbiota, is more likely to be influenced by the availability of sugars as alternative sources of energy. These considerations may be of relevance when selecting strains or strain combinations of lactate-utilizing bacteria intended for use as therapeutic probiotics with the aim of enhancing the stability of the microbial community.

Author statements

Authors and contributors

POS carried out laboratory and *in silico* experiments/data analysis. PL, SHD, HJF and AWW also contributed to experimental design, data interpretation and visualization. ET and HJM generated essential strain data. SS contributed to formal analysis and visualization. POS, PL, SHD, HJF and AWW wrote the manuscript. All co-authors commented on advanced drafts of the manuscript, and contributed edits to the final draft.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

PL, SHD, HJF, AWW and the Rowett Institute received core financial support from the Scottish Government Rural and Environmental Sciences and Analytical Services (SG-RESAS). The work was also partly funded by Chr. Hansen, a global bioscience company that develops natural ingredient solutions for the food, nutritional, pharmaceutical and agricultural industries, who supported POS and ET. Chr. Hansen exerted no influence on results obtained and presented in this manuscript. This also does not alter our commitment to sharing data and materials.

Acknowledgements

The authors would like to thank Donna Henderson (Rowett Institute, University of Aberdeen) for carrying out gas chromatography analysis of short chain fatty acids, and also thank the Centre for Genome-Enabled Biology and Medicine (CGEBM) at the University of Aberdeen for carrying out the Illumina-based transcriptomics sequencing. They would also like to acknowledge the support of the Maxwell computer cluster funded by the University of Aberdeen.

References

- [1]. **Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. (2012).** The role of the gut microbiota in nutrition and health. *Nature reviews Gastroenterology & hepatology* **9**, 577.
- [2]. **Gottschalk, G. (1979).** *Bacterial Metabolism*. Verlag: New York, Heidelberg, Berlin.: Springer.
- [3]. **Duncan, S. H., Louis, P., Thomson, J. M. & Flint, H. J. (2009).** The role of pH in determining the species composition of the human colonic microbiota. *Environ Microbiol* **11**, 2112-2122.
- [4]. **McWilliam Leitch, E. & Stewart, C. (2002).** Susceptibility of *Escherichia coli* O157 and non-O157 isolates to lactate. *Lett Appl Microbiol* **35**, 176-180.

655 [5]. **Stewart, C. J., Ajami, N. J., O'Brien, J. L. & other authors (2018).** Temporal
656 development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**,
657 583-588.

658 [6]. **Slyter, L. L. (1976).** Influence of acidosis on rumen function. *J Anim Sci* **43**, 910-929.

659 [7]. **Russell, J. B. & Hino, T. (1985).** Regulation of lactate production in *Streptococcus*
660 *bovis*: a spiraling effect that contributes to rumen acidosis. *J Dairy Sci* **68**, 1712-1721.

661 [8]. **Kaneko, T., Bando, Y., Kurihara, H., Satomi, K., Nonoyama, K. & Matsuura, N.**
662 **(1997).** Fecal microflora in a patient with short-bowel syndrome and identification of
663 dominant lactobacilli. *J Clin Microbiol* **35**, 3181-3185.

664 [9]. **Vernia, P., Caprilli, R., Latella, G., Barbetti, F., Magliocca, F. M. & Cittadini, M.**
665 **(1988).** Fecal lactate and ulcerative colitis. *Gastroenterology* **95**, 1564-1568.

666 [10]. **Gillis, C. C., Hughes, E. R., Spiga, L. & other authors (2018).** Dysbiosis-associated
667 change in host metabolism generates lactate to support *Salmonella* growth. *Cell host &*
668 *microbe* **23**, 54-64. e6.

669 [11]. **Thomas, M. T., Shepherd, M., Poole, R. K., van Vliet, A. H., Kelly, D. J. &**
670 **Pearson, B. M. (2011).** Two respiratory enzyme systems in *Campylobacter jejuni* NCTC
671 11168 contribute to growth on L-lactate. *Environ Microbiol* **13**, 48-61.

672 [12]. **Counotte, G., Lankhorst, A. & Prins, R. (1983).** Role of DL-lactic acid as an
673 intermediate in rumen metabolism of dairy cows. *J Anim Sci* **56**, 1222-1235.

674 [13]. **Belenguer, A., Duncan, S. H., Calder, A. G., Holtrop, G., Louis, P., Lobley, G. E. &**
675 **Flint, H. J. (2006).** Two routes of metabolic cross-feeding between *Bifidobacterium*
676 *adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol*
677 **72**, 3593-3599.

678 [14]. **Shabat, S. K. B., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S.,**
679 **Miller, M. E. B., White, B. A., Shterzer, N. & Mizrahi, I. (2016).** Specific microbiome-
680 dependent mechanisms underlie the energy harvest efficiency of ruminants. *The ISME*
681 *journal* **10**, 2958-2972.

682 [15]. **Wang, S. P., Rubio, L. A., Duncan, S. H. & other authors (2020).** Pivotal Roles for
683 pH, Lactate, and Lactate-Utilizing Bacteria in the Stability of a Human Colonic Microbial
684 Ecosystem. *mSystems* **5**, 10.1128/mSystems.00645-20.

685 [16]. **Van Immerseel, F., Ducatelle, R., De Vos, M. & other authors (2010).** Butyric acid-
686 producing anaerobic bacteria as a novel probiotic treatment approach for inflammatory bowel
687 disease. *J Med Microbiol* **59**, 141-143.

688 [17]. **Perraudieu, F., McMurdie, P., Bullard, J. & other authors (2020).** Improvements to
689 postprandial glucose control in subjects with type 2 diabetes: a multicenter, double blind,
690 randomized placebo-controlled trial of a novel probiotic formulation. *BMJ Open Diabetes*
691 *Res Care* **8**, 10.1136/bmjdrc-2020-001319.

- [18]. **Gilijamse, P. W., Hartstra, A. V., Levin, E. & other authors (2020).** Treatment with *Anaerobutyricum soehngenii*: a pilot study of safety and dose–response effects on glucose metabolism in human subjects with metabolic syndrome. *npj Biofilms and Microbiomes* **6**, 1-10.
- [19]. **Shetty, S. A., Zuffa, S., Bui, T. P. N., Aalvink, S., Smidt, H. & 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.
- [20]. **Allen-Vercoe, E., Daigneault, M., White, A., Panaccione, R., Duncan, S. H., Flint, H. J., O'Neal, L. & Lawson, P. A. (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.
- [21]. **Duncan, S. H., Louis, P. & 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.
- [22]. **Weghoff, M. C., Bertsch, J. & Müller, V. (2015).** A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ Microbiol* **17**, 670-677.
- [23]. **Counotte, G. H., Prins, R. A., Janssen, R. H. & Debie, M. J. (1981).** Role of *Megasphaera elsdenii* in the Fermentation of dl-[2-C]lactate in the Rumen of Dairy Cattle. *Appl Environ Microbiol* **42**, 649-655.
- [24]. **Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., Leitch, C. M., Scott, K. P., Flint, H. J. & Louis, P. (2014).** Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal* **8**, 1323-1335.
- [25]. **Marquet, P., Duncan, S. H., Chassard, C., Bernalier-Donadille, A. & Flint, H. J. (2009).** Lactate has the potential to promote hydrogen sulphide formation in the human colon. *FEMS Microbiol Lett* **299**, 128-134.
- [26]. **Shetty, S. A., Boeren, S., Bui, T. P. N., Smidt, H. & de Vos, W. M. (2020).** Unravelling lactate-acetate and sugar conversion into butyrate by intestinal *Anaerobutyricum* and *Anaerostipes* species by comparative proteogenomics. *Environ Microbiol*
- [27]. **Lopez-Siles, M., Khan, T. M., Duncan, S. H., Harmsen, H. J., Garcia-Gil, L. J. & Flint, H. J. (2012).** Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl Environ Microbiol* **78**, 420-428.
- [28]. **Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004).** Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12.
- [29]. **Miyazaki, K., Martin, J. C., Marinsek-Logar, R. & Flint, H. J. (1997).** Degradation and Utilization of Xylans by the Rumen Anaerobe *Prevotella bryantii*(formerly *P. ruminicola* subsp. *brevis*) B 1 4. *Anaerobe* **3**, 373-381.

- [30]. **Bryant, M. (1972).** Commentary on the Hungate technique for culture of anaerobic bacteria. *Am J Clin Nutr* **25**, 1324-1328.
- [31]. **Shetty, S. A., Ritari, J., Paulin, L., Smidt, H. & De Vos, W. M. (2017).** Complete Genome Sequence of Eubacterium hallii Strain L2-7. *Genome Announc* **5**, 10.1128/genomeA.01167-17.
- [32]. **Barcenilla, A., Pryde, S. E., Martin, J. C., Duncan, S. H., Stewart, C. S., Henderson, C. & Flint, H. J. (2000).** Phylogenetic Relationships of Butyrate-Producing Bacteria from the Human Gut. *Appl Environ Microbiol* **66**, 1654-1661.
- [33]. **Louis, P., Duncan, S. H., McCrae, S. I., Millar, J., Jackson, M. S. & Flint, H. J. (2004).** Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* **186**, 2099-2106.
- [34]. **Holdeman, L. V. & Moore, W. (1975).** New Genus, Coprococcus, Twelve New Species, and Emended Descriptions of Four Previously Described Species of Bacteria from Human Feces. *Int J Syst Evol Microbiol* **25**, 98-98.
- [35]. **Peng, Y., Leung, H. C., Yiu, S. & Chin, F. Y. (2012).** IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420-1428.
- [36]. **Seemann, T. (2014).** Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069.
- [37]. **Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. (2015).** CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055.
- [38]. **Eddy, S. R. (1998).** Profile hidden Markov models. *Bioinformatics* **14**, 755-763.
- [39]. **UniProt Consortium (2019).** UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515.
- [40]. **Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004).** The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**, D277-80.
- [41]. **Tremblay, P. L., Zhang, T., Dar, S. A., Leang, C. & Lovley, D. R. (2012).** The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD⁺ oxidoreductase essential for autotrophic growth. *mBio* **4**, e00406-12.
- [42]. **Biegel, E., Schmidt, S. & Müller, V. (2009).** Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*. *Environ Microbiol* **11**, 1438-1443.
- [43]. **El-Gebali, S., Mistry, J., Bateman, A. & other authors (2019).** The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432.

766 [44]. **Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. (2012).** CD-HIT: accelerated for clustering
767 the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152.

768 [45]. **Katoh, K. & Toh, H. (2008).** Recent developments in the MAFFT multiple sequence
769 alignment program. *Briefings in bioinformatics* **9**, 286-298.

770 [46]. **Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. (2009).** trimAl: a tool
771 for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**,
772 1972-1973.

773 [47]. **Nguyen, L., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. (2015).** IQ-TREE: a
774 fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol*
775 *Biol Evol* **32**, 268-274.

776 [48]. **Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A. & Jermiin, L. S.**
777 **(2017).** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature*
778 *methods* **14**, 587.

779 [49]. **Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O.**
780 **(2010).** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing
781 the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

782 [50]. **Letunic, I. & Bork, P. (2007).** Interactive Tree Of Life (iTOL): an online tool for
783 phylogenetic tree display and annotation. *Bioinformatics* **23**, 127-128.

784 [51]. **Andrews, S. (2010).** FastQC: a quality control tool for high throughput sequence data.
785 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

786 [52]. **Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. (2019).** Graph-based
787 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**,
788 907-915.

789 [53]. **Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,**
790 **Abecasis, G. & Durbin, R. (2009).** The sequence alignment/map format and SAMtools.
791 *Bioinformatics* **25**, 2078-2079.

792 [54]. **Liao, Y., Smyth, G. K. & Shi, W. (2014).** featureCounts: an efficient general purpose
793 program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930.

794 [55]. **Robinson, M. D., McCarthy, D. J. & Smyth, G. K. (2010).** edgeR: a Bioconductor
795 package for differential expression analysis of digital gene expression data. *Bioinformatics*
796 **26**, 139-140.

797 [56]. **de Luca, G., de Philip, P., Rousset, M., Belaich, J. P. & Dermoun, Z. (1998).** The
798 NADP-Reducing Hydrogenase of *Desulfovibrio fructosovorans*: Evidence for a Native
799 Complex with Hydrogen-Dependent Methyl-Viologen-Reducing Activity. *Biochem Biophys*
800 *Res Commun* **248**, 591-596.

801 [57]. **Malki, S., Saimmaime, I., De Luca, G., Rousset, M., Dermoun, Z. & Belaich, J. P.**
802 **(1995).** Characterization of an operon encoding an NADP-reducing hydrogenase in
803 *Desulfovibrio fructosovorans*. *J Bacteriol* **177**, 2628-2636.

804 [58]. **Kuchta, R. D. & Abeles, R. H. (1985).** Lactate reduction in *Clostridium propionicum*.
805 Purification and properties of lactyl-CoA dehydratase. *J Biol Chem* **260**, 13181-13189.

806 [59]. **Pinchuk, G. E., Rodionov, D. A., Yang, C. & other authors (2009).** Genomic
807 reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously
808 uncharacterized machinery for lactate utilization. *Proc Natl Acad Sci U S A* **106**, 2874-2879.

809 [60]. **Jiang, T., Gao, C., Ma, C. & Xu, P. (2014).** Microbial lactate utilization: enzymes,
810 pathogenesis, and regulation. *Trends Microbiol* **22**, 589-599.

811 [61]. **Chai, Y., Kolter, R. & Losick, R. (2009).** A widely conserved gene cluster required for
812 lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation. *J Bacteriol*
813 **191**, 2423-2430.

814 [62]. **Dong, J. M., Taylor, J. S., Latour, D. J., Iuchi, S. & Lin, E. C. (1993).** Three
815 overlapping lct genes involved in L-lactate utilization by *Escherichia coli*. *J Bacteriol* **175**,
816 6671-6678.

817 [63]. **Dym, O., Pratt, E. A., Ho, C. & Eisenberg, D. (2000).** The crystal structure of D-
818 lactate dehydrogenase, a peripheral membrane respiratory enzyme. *Proc Natl Acad Sci U S A*
819 **97**, 9413-9418.

820 [64]. **Genthner, B. R., Davis, C. L. & Bryant, M. P. (1981).** Features of rumen and sewage
821 sludge strains of *Eubacterium limosum*, a methanol- and H₂-CO₂-utilizing species. *Appl*
822 *Environ Microbiol* **42**, 12-19.

823 [65]. **Kelly, W. J., Henderson, G., Pacheco, D. M. & other authors (2016).** The complete
824 genome sequence of *Eubacterium limosum* SA11, a metabolically versatile rumen acetogen.
825 *Standards in genomic sciences* **11**, 26.

826 [66]. **Prabhu, R., Altman, E. & Eiteman, M. A. (2012).** Lactate and acrylate metabolism by
827 *Megasphaera elsdenii* under batch and steady-state conditions. *Appl Environ Microbiol* **78**,
828 8564-8570.

829 [67]. **Gillis, C. C., Winter, M. G., Chanin, R. B., Zhu, W., Spiga, L. & Winter, S. E.**
830 **(2019).** Host-Derived Metabolites Modulate Transcription of *Salmonella* Genes Involved in l-
831 Lactate Utilization during Gut Colonization. *Infect Immun* **87**, 10.1128/IAI.00773-18. Print
832 2019 Apr.

833 [68]. **Seeliger, S., Janssen, P. H. & Schink, B. (2002).** Energetics and kinetics of lactate
834 fermentation to acetate and propionate via methylmalonyl-CoA or acrylyl-CoA. *FEMS*
835 *Microbiol Lett* **211**, 65-70.

836 [69]. **Bertsch, J., Parthasarathy, A., Buckel, W. & Muller, V. (2013).** An electron-
837 bifurcating caffeyl-CoA reductase. *J Biol Chem* **288**, 11304-11311.

- [70]. **Louis, P. & Flint, H. J. (2009).** Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett* **294**, 1-8.
- [71]. **Hetzel, M., Brock, M., Selmer, T., Pierik, A. J., Golding, B. T. & Buckel, W. (2003).** Acryloyl-CoA reductase from *Clostridium propionicum*: An enzyme complex of propionyl-CoA dehydrogenase and electron-transferring flavoprotein. *European journal of biochemistry* **270**, 902-910.
- [72]. **Gibello, A., Collins, M. D., Dominguez, L., Fernandez-Garayzabal, J. F. & Richardson, P. T. (1999).** Cloning and analysis of the L-lactate utilization genes from *Streptococcus iniae*. *Appl Environ Microbiol* **65**, 4346-4350.
- [73]. **Müller, V., Chowdhury, N. P. & Basen, M. (2018).** Electron bifurcation: a long-hidden energy-coupling mechanism. *Annu Rev Microbiol* **72**, 331-353.
- [74]. **Detman, A., Mielecki, D., Chojnacka, A., Salamon, A., Błaszczuk, M. K. & Sikora, A. (2019).** Cell factories converting lactate and acetate to butyrate: *Clostridium butyricum* and microbial communities from dark fermentation bioreactors. *Microbial cell factories* **18**, 36.
- [75]. **Chen, L., Shen, Y., Wang, C., Ding, L., Zhao, F., Wang, M., Fu, J. & Wang, H. (2019).** *Megasphaera elsdenii* lactate degradation pattern shifts in rumen acidosis models. *Frontiers in microbiology* **10**, 162.
- [76]. **Johns, A. (1952).** The mechanism of propionic acid formation by *Clostridium propionicum*. *Microbiology* **6**, 123-127.

Supplementary information

Table S1. Classifying new *Anaerobutyricum* genomes by average nucleotide identity and average amino acid identity.

Table S2. Genome information of strains included in this study. Includes genomes completeness, CDS number and references.

Table S3. Short chain fatty acid profiles following growth of lactate-utilizing bacteria on different carbon sources (numerical data for experiments presented in Figure 1).

Table S4. Overview showing number of differentially expressed genes during the transcriptomics experiments in each growth condition.

Table S5. Differentially expressed genes in *Anaerobutyricum soehngenii* L2-7 during growth on glucose versus lactate

Table S6. Differentially expressed genes in *Anaerobutyricum soehngenii* L2-7 during growth on glucose versus glucose and lactate

Table S7. Differentially expressed genes in *Anaerobutyricum soehngenii* L2-7 during growth on lactate versus glucose and lactate

Table S8. Differentially expressed genes in *Coprococcus catus* GD/7 during growth on fructose versus lactate

Table S9. Differentially expressed genes in *Coprococcus catus* GD/7 during growth on fructose versus fructose and lactate

Table S10. Differentially expressed genes in *Coprococcus catus* GD/7 during growth on lactate versus fructose and lactate

881 **Table S11.** Lactate utilization HMM (full sequence bit score ≥ 80) in *Coprococcus catus* GD/7
882 and *Anaerobutyricum soehngenii* L2-7

883 **Table S12.** Comparison of *Anaerobutyricum soehngenii* L2-7 lactate utilization locus to that
884 of other known lactate utilizers

885 **Table S13.** Comparison of *Coprococcus catus* GD/7 lactate utilization locus to that of other
886 known lactate utilizers

887 **Table S14.** Comparison of *Coprococcus catus* GD/7 butyrate production loci to those of other
888 known lactate utilizers

889 **Table S15.** Representative genes present in each lactate permease protein family

890 **Table S16.** Representative genes present in each sub-family of LP-IV

891 **Table S17.** Abundance of acyl-CoA dehydrogenase families 1-5 in known lactate utilizers

892 **Figure S1.** Transcriptomic experimental workflow

893

894