

# Supplementary materials

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

Sudarshan A. Shetty<sup>1</sup>, Sjef Boeren<sup>2</sup>, Thi Phuong Nam Bui<sup>1, 3</sup>, Hauke Smidt<sup>1</sup>, Willem M. de Vos<sup>1, 4\*</sup>

<sup>1</sup>Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands.

<sup>2</sup>Laboratory of Biochemistry, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, The Netherlands.

<sup>3</sup>Caelus Pharmaceuticals, 3474 KG Zegveld, The Netherlands

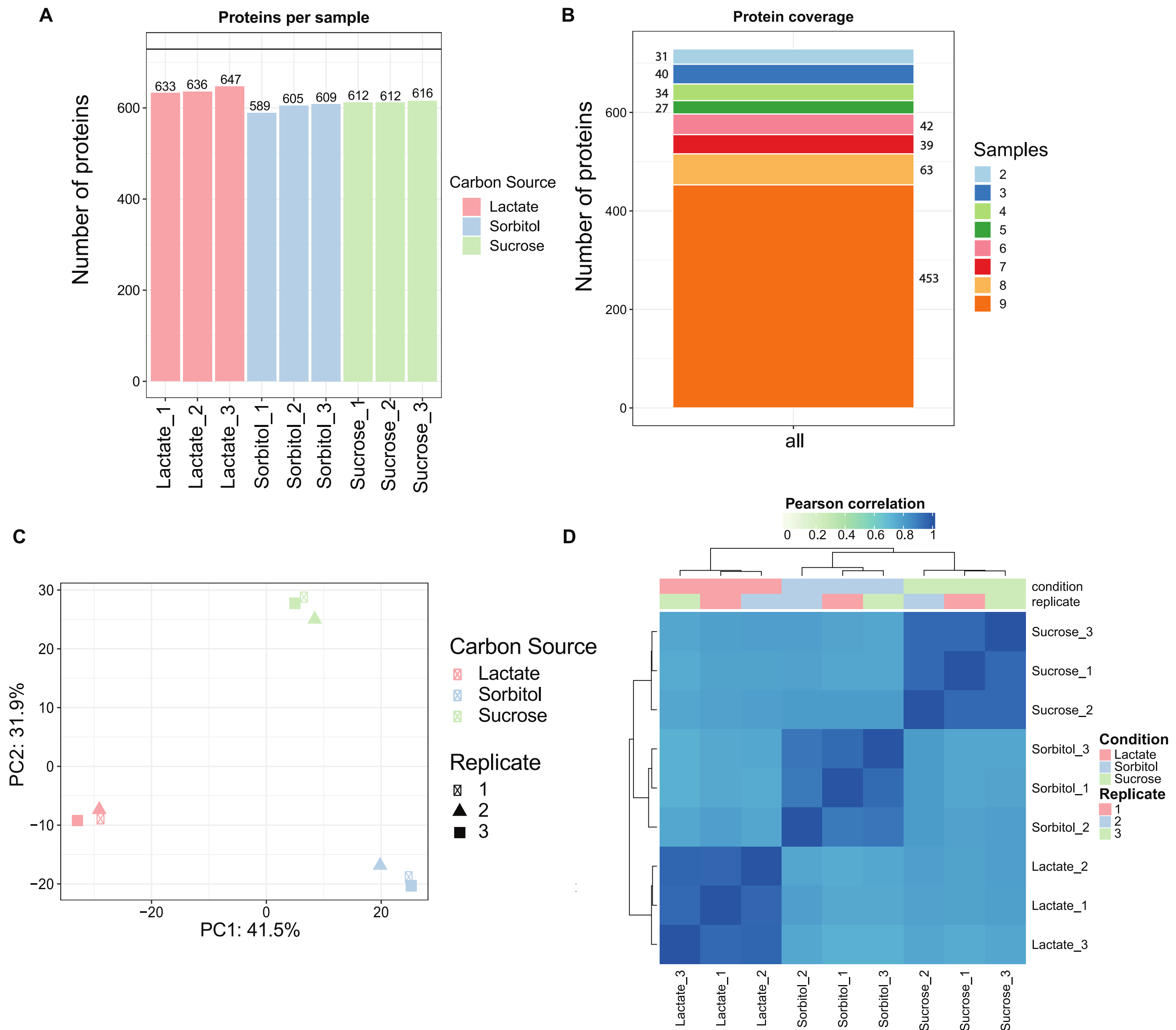
<sup>4</sup>Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

\*Corresponding author: Willem M. de Vos

Email ID: willem.devos@wur.nl

**Competing interests:** The authors declare no competing interests.

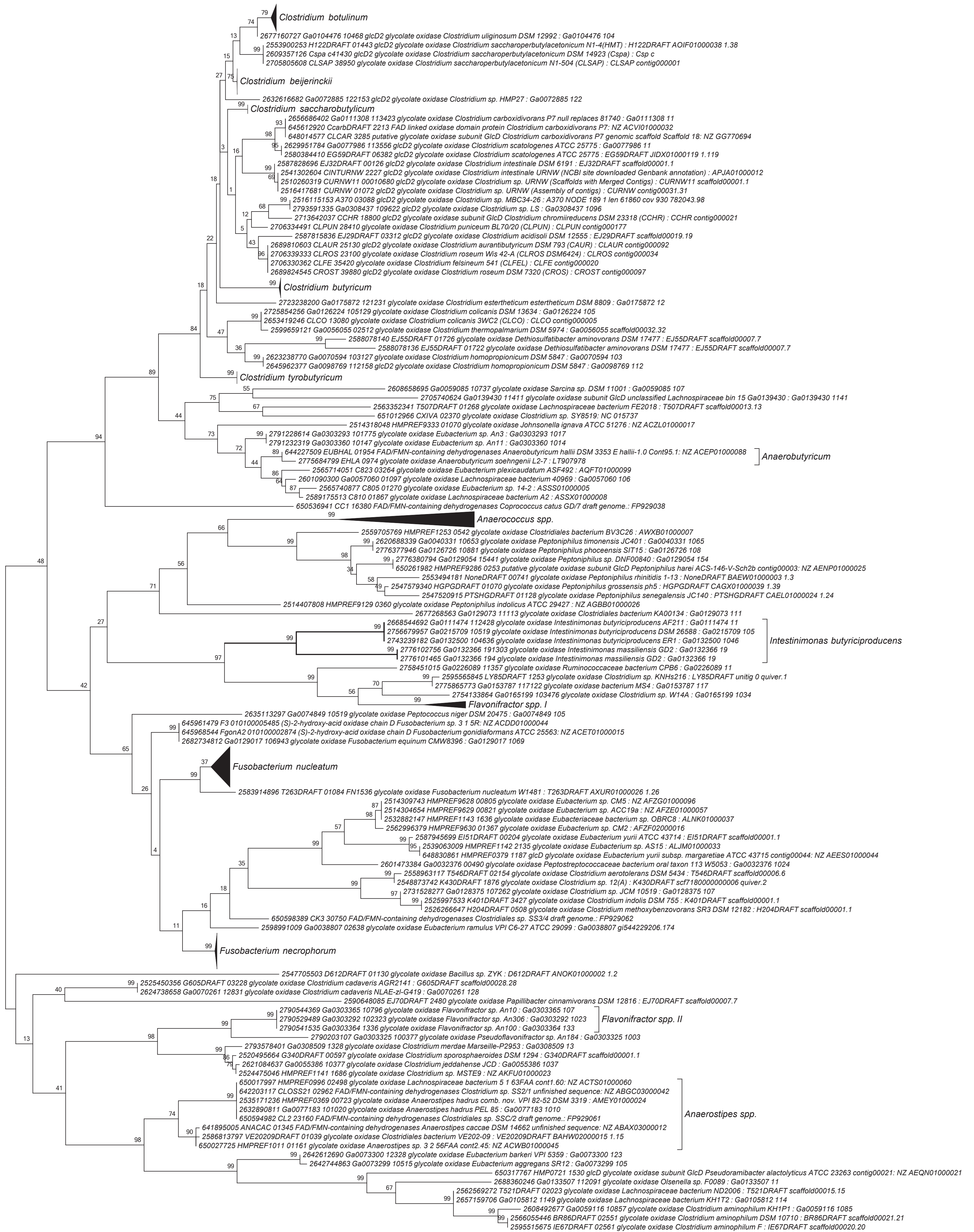
**Running title:** Lactate utilization by *Anaerobutyricum* and related species



**Figure 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. Total 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.



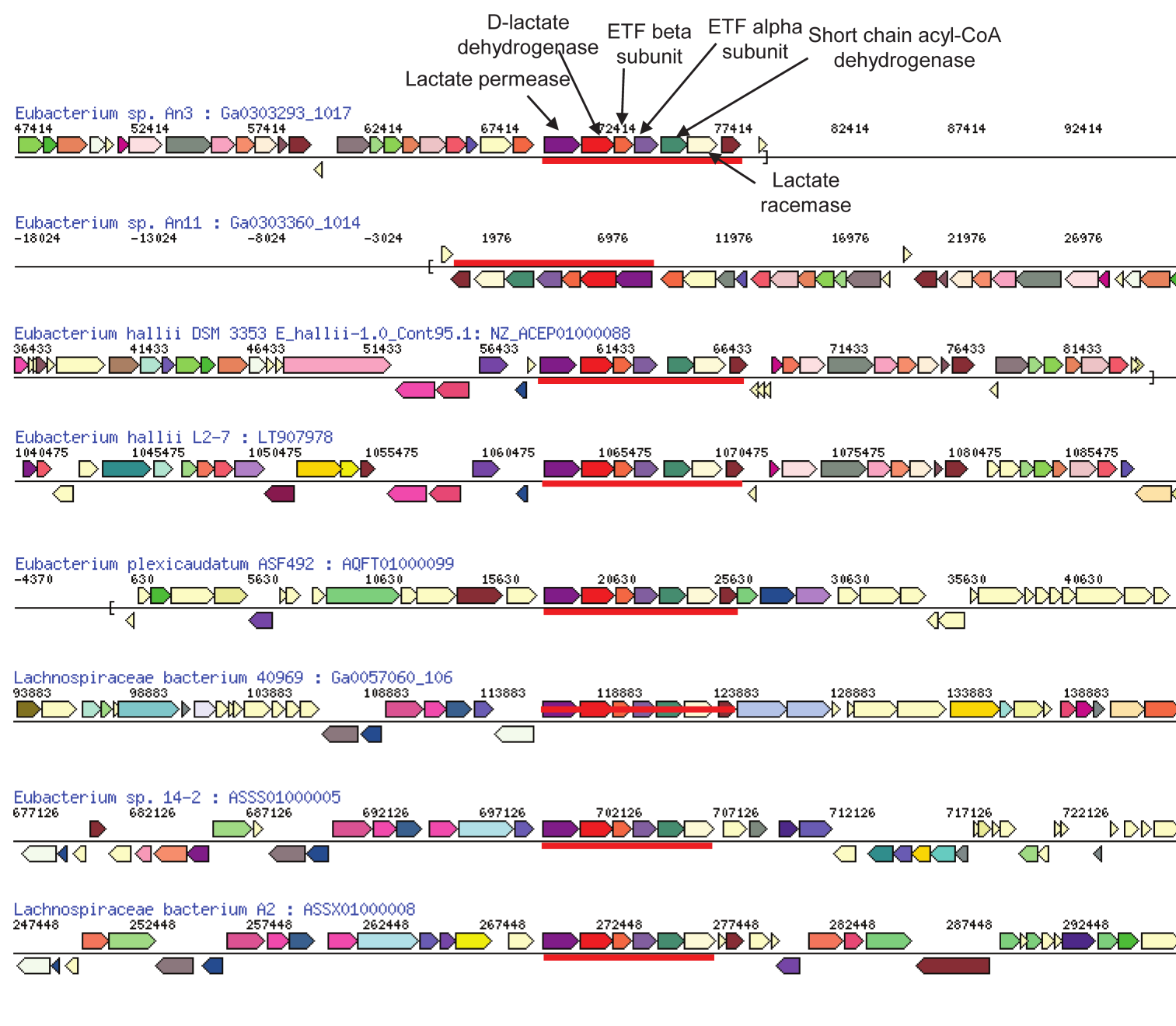


**Figure S2:** Maximum likelihood phylogenetic tree of 285 lactate dehydrogenase homologues. Phylogenetic tree of 285 LDH homologs identified by searching against 55,499 isolate genomes of the IMG/ER database (as of 2 October 2018). Labels represent the IMG gene id, IMG annotation for the gene product, 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.



A

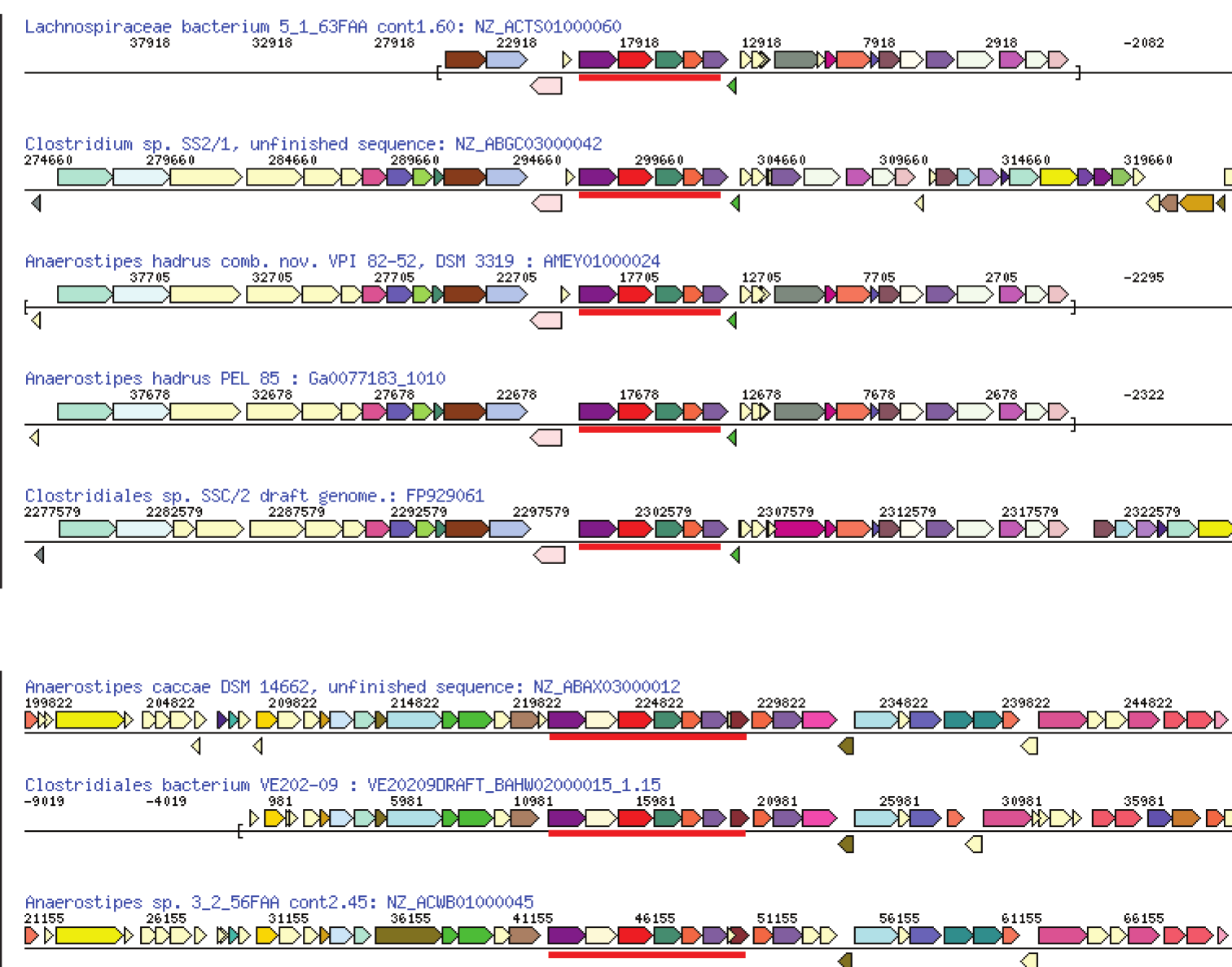
Anaerobutyricum group



B

Group I  
No racemase

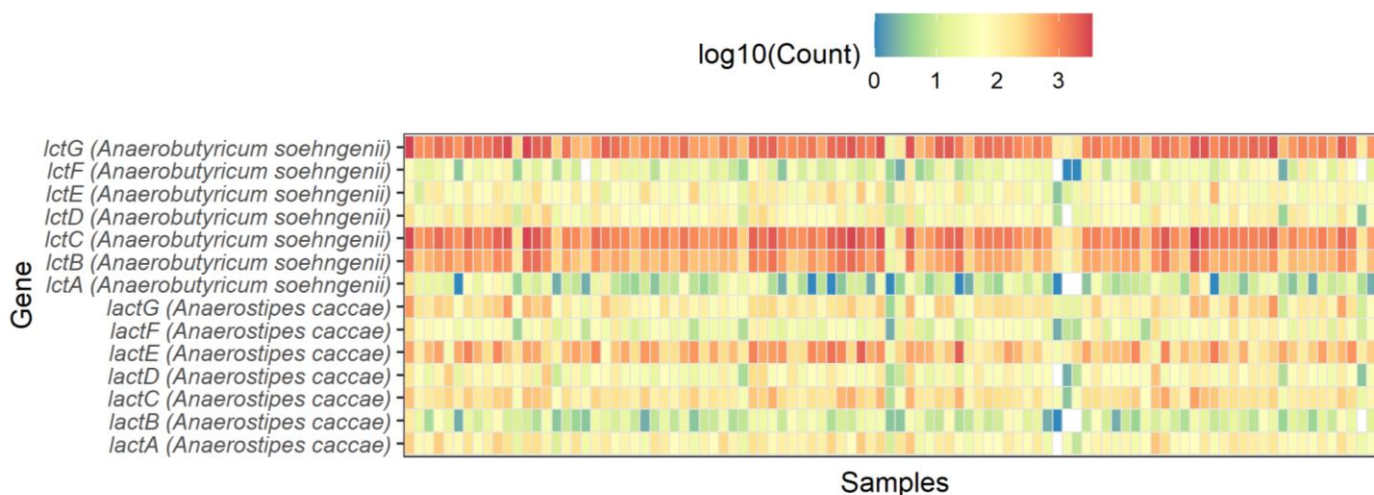
Anaerostipes group

Group II  
With racemase

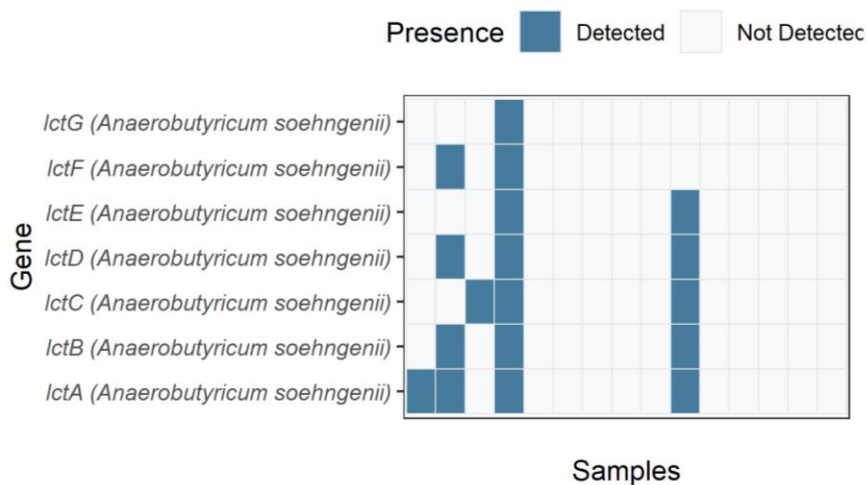
**Figure S3: Comparison of the neighbourhood of genes involved in lactate utilization in genomes of *A. soehngeni* and related lactate utilizing bacteria.** The red line highlights the gene cluster involved in lactate utilization. Neighbourhoods of genes in other genomes with the same top cluster of 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 organisation as *A. soehngeni* lactate utilization gene cluster. B] Shows selected bacterial genomes that share similar gene organisation as *Anaerostipes caccae* lactate utilization gene cluster.

A

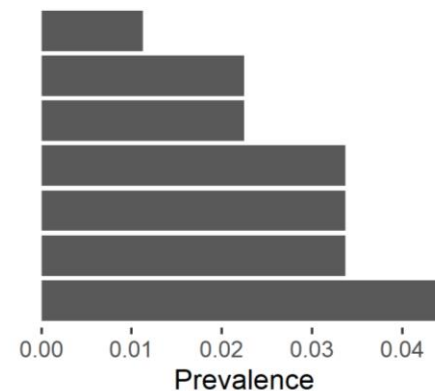
Human gut metagenome (HMP), n = 100



B

Metatranscriptomes from  
Simplified intestinal microbiota in a  
mouse model, n = 15

C

**Supplementary figure S4: Mining of publicly available metagenomes and metatranscriptomes.**

A] Abundance of *lctABCDFG* protein homologues in human gut metagenomes from the human microbiome project. B] Detection and prevalence of transcripts for *lctABCDFG* gene cluster of *Anaerobutyricum soehngenii* in Simplified Intestinal Microbiota (SIM) colonized mouse model.