environmental microbiology reports

Environmental Microbiology Reports (2016) **8**(6), 1024–1037 doi:10.1111/1758-2229.12483

Comparative genomics and physiology of the butyrate-producing bacterium *Intestinimonas butyriciproducens*

Thi Phuong Nam Bui,[1]* Sudarshan Anand Shetty,[1] Ilias Lagkouvardos,[2] Jarmo Ritari,[3] Bhawani Chamlagain,[4] François P. Douillard,[3] Lars Paulin,[5] Vieno Piironen,[4] Thomas Clavel,[2] Caroline M. Plugge[1] and Willem M. de Vos[1,3,6]

[1]*Laboratory of Microbiology, Wageningen University, Stippeneng 4, Wageningen, WE 6708, The Netherlands.*

[2]*Technische Universität München, ZIEL Institute for Food and Health, Freising 85354, Germany.*

[3]*Department of Veterinary Biosciences, Division of Microbiology and Epidemiology, University of Helsinki, Helsinki 00014, Finland.*

[4]*Department of Food and Environmental Sciences, University of Helsinki, Helsinki 00014, Finland.*

[5]*DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, Helsinki, Finland.*

[6]*RPU Immunobiology, Department of Bacteriology and Immunology, Faculty of Medicine University of Helsinki, Helsinki 00014, Finland.*

Summary

Intestinimonas **is a newly described bacterial genus with representative strains present in the intestinal tract of human and other animals. Despite unique metabolic features including the production of butyrate from both sugars and amino acids, there is to date no data on their diversity, ecology, and physiology. Using a comprehensive phylogenetic approach,** ***Intestinimomas*** **was found to include at least three species that colonize primarily the human and mouse intestine. We focused on the most common and cultivable species of the genus,** ***Intestinimonas butyriciproducens*****, and performed detailed genomic and physiological comparison of strains SRB521^T and AF211, isolated from the mouse and human gut respectively. The complete 3.3-Mb genomic sequences of both strains were highly similar with 98.8% average nucleotide identity, testifying to their assignment to one single species. However, thorough analysis revealed significant genomic rearrangements, variations in phage-derived sequences, and the presence of new CRISPR sequences in both strains. Moreover, strain AF211 appeared to be more efficient than strain SRB521^T in the conversion of the sugars arabinose and galactose. In conclusion, this study provides genomic and physiological insight into** ***Intestinimonas butyriciproducens*****, a prevalent butyrate-producing species, differentiating strains that originate from the mouse and human gut.**

Received 2 June, 2016; revised 14 September, 2016; accepted 29 September, 2016. *For correspondence. E-mail nam.bui@wur.nl. Tel. +31 317 483116; Fax : +31 317 483829.

Introduction

The gut microbiome is a living organ within the human body that contains highly complex and diverse microbial communities contributing substantially to human health (Zoetendal *et al.*, 2006). While the majority of these communities consists of strictly anaerobic bacteria that are difficult to culture (Flint *et al.*, 2006), an increasing number has successfully been isolated, and presently approximately 1000 cultured species, including eukaryotes and archaea have been reported (Rajilić-Stojanović and de Vos, 2014; Browne *et al.*, 2016). In spite of this, only a small number of these intestinal anaerobes have been characterized at the genomic level and often only draft genomes have been assembled. Complete genomes have the advantage that full genetic repertoires can be studied, and repetitive sequences can be revealed.

The chemical compound butyrate and butyrate-producing bacteria have gained increasing attention as they both play an important role in maintaining gut homeostasis and intestinal epithelial integrity (Hamer *et al.*, 2008). Butyrate is a major substrate for colonocytes and is reported to have anti-carcinogenic, anti-inflammatory, and anti-oxidative effects (Hinnebusch *et al.*, 2002; Klampfer *et al.*, 2003). Interestingly, decreased amounts of butyrogenic bacteria have been observed in inflammatory bowel diseases, colon cancer, and type-2 diabetes (Flint *et al.*, 2012; Qin *et al.*, 2012; de Vos and Nieuwdorp, 2013). Till date several butyrate-producing bacteria have been isolated that mainly belong to *Clostridium* clusters XIVa and IV (gathering many members of the family *Lachnospiraceae* and *Ruminococcaceae* respectively). These bacteria mainly produce butyrate via the acetyl-CoA pathway (Louis

et al., 2010). A recent study based on comparative analysis of metagenomes reported three other pathways leading to butyrate production where amino acids serve as substrates, including lysine and glutamate (Vital *et al*., 2014). All these butyrogenic pathways merged at an energy-generating step where crotonyl-CoA is converted to butyryl-CoA by means of a butyryl-CoA dehydrogenase complex (Bcd/etf). The coupled oxidation of reduced ferredoxin and reduction of NADH from this conversion by the membrane-integrated Rnf complex creates a proton motive force (Li *et al*., 2008). The lysine pathway that includes a specific acetoacetyl-CoA transferase involved in the butyrate formation step was estimated to be the second dominant butyrogenic pathway in the gut. However, no cultured isolates capable of performing this relevant reaction had been described (Vital *et al*., 2014). Using a classical cultivation approach with human stool as inoculant, we isolated strain *Intestinimonas* AF211, which is capable of converting lysine completely and efficiently into butyrate and acetate (Bui *et al*., 2015). This feature is unique for the gut ecosystem and links two important metabolic features, butyrogenesis and amino acid fermentation in the intestinal tract. The complete 16S rRNA gene sequence of strain AF211 revealed more than 99% similarity to *Intestinimonas butyriciproducens* SRB521^T isolated from mouse caecum (Kläring *et al*., 2013), indicating that these two strains belong to the same species within *Clostridium* cluster IV. Of note, *Faecalibacterium prausnitzii*, another butyrogenic species within *Clostridium* cluster IV, is known to be prevalent and abundant in human (Arumugam *et al*., 2011) but not in mice (Xiao *et al*., 2015).

Predictions based on complete genome sequences combined with physiological and other functional studies are powerful tools for improving our understanding of the lifestyle and role of bacteria. A genomic approach can also be used to gain insight into the genetic potential of an organism and assists in predicting its ecological role. Presently, only a few closed genomes of intestinal butyrate-producing bacteria have been characterised, including that of *Faecalibacterium prausnitzii* (Heinken *et al*., 2014) and *Roseburia inulinivorans* (Scott *et al*., 2006). More recently, a partial genome sequence of *Anaerostipes hadrus* PEL 85 was reported (Kant *et al*., 2015) and a number of draft genomes of butyrate-producing bacteria from the *Roseburia/E.rectale* group have been characterised (Sheridan *et al*., 2016). In the present study, we focused on the recently described genus *Intestinimonas* and describe its phylogenetic and host distribution. Moreover, we compare the complete genome sequences of the human isolate *I. butyriciproducens* AF211 and the mouse isolate *I. butyriciproducens* SRB521^T. Finally we describe the physiological characteristics of both strains to confirm their predicted metabolic features and provide explanations of their adaptation to different hosts.

Results and discussion

Diversity and ecology of Intestinimonas

Intestinimonas is a newly described genus with peculiar metabolic functions, but its phylogeny and ecological distribution remain unknown. By utilizing the accumulated sequences available in the Sequence Read Archive (SRA) (Lagkouvardos *et al*., 2016), 9,887 *Intestinimonas*-like 16S rRNA gene amplicons were obtained and their pairwise distances projected in two dimensions using MDS (Fig. 1A). Using full-length substitutes for the four bins containing the most amplicon sequences, enabled further filtering of the sequences that could be confidently classified as members of the genus *Intestinimonas* to 5,732 reads (see Supplementary methods). Clustering those using CROP resulted in the formation of eight none-singleton clusters. The vast majority of OTU sequences (99%) fell into only three major molecular 'species' clusters. The results mirrored the diversity supported by phylogenetic analysis of near full length 16S rRNA genes (Fig. 1B). Ecological assessment revealed that 90% of *I. butyriciproducens*-like sequences were predicted to originate from human samples (Fig. 1B). In contrast, the species represented by clone FJ676152 clustered with reads that originate from various animal samples. Few amplicons, primarily originating from murine gut, mapped to the species represented by clone FJ374227. *I. butyriciproducens* strains have also been isolated from swine faeces, signifying its widespread presence in mammals (Levine *et al*., 2013; Rettedal *et al*., 2014). As *I. butyriciproducens*-like sequences represented a major clade within the genus and were most prevalent in humans, but also present in the murine host, we performed in-depth genomic and physiological comparison of strains AF211 and SRB521^T, isolated from the human and mouse gut respectively.

Phenotypic characterization of I. butyriciproducens strains

Physiological and biochemical characterization of *I. butyriciproducens* strains AF211 and SRB521^T using API Rapid 32A, API 20NE, and API ZYM kits revealed that strain SRB521^T but not AF211 possessed N-acetyl-β-glucosaminidase activity (Supporting Information Table S1). Interestingly, analysis of cellular fatty acid composition, an important chemotaxonomic characteristic for microbial classification, revealed differences between the two strains (Supporting Information Table S2). More than 20% content difference was observed in major fatty

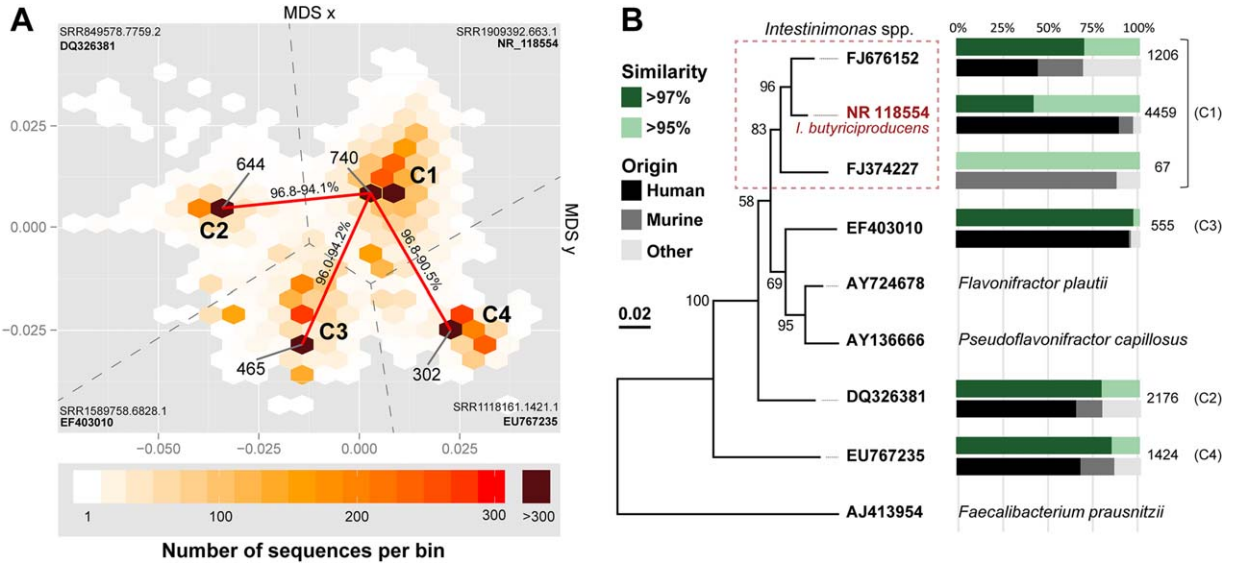


Fig. 1. 16S rRNA-based diversity and host distribution of *Intestinimonas*.
A. MDS plot of pairwise distances between *Intestinimonas*-like sequences extracted from SRA. Sequences were binned in hexagons color-coded according to the number of sequences they contained. The reads formed separated clusters (C1-C4) around bins containing over 300 sequences each. Similarity distances between representative short amplicon reads and clones from the three neighbouring cores and the one containing the reference 16S rRNA sequence of *I. butyriciproducens* (cluster C1) are shown next to the red lines connecting them, respectively; the corresponding sequence accession for each cluster core are given in the corresponding corner. The grey dash lines indicate midpoints between cores after assignment of each point (OTU) to its closest center.
B. Maximum likelihood tree of *Intestinimonas*-related 16S rRNA gene sequences and representatives of most abundant amplicon reads (Fig 1A) together with selected neighbouring genera and outgroup. Bootstrap support is shown next to branches, with branch lengths measured in the number of substitutions per site. The box with red dashed-line indicates members of the genus *Intestinimonas*. The number of unique amplicon sequences mapped to each clone is shown on the right hand side.

acids, such as $C_{14:00}$; iso-$C_{19:1}$ I and $C_{18:00}$. The cellular fatty acids have a major role in maintaining the integrity and viability of microbial cells, and may change in composition dependent on the environmental and growth conditions. Recently, differences in intra-species cellular fatty acid composition in *Lactobacillus plantarum* strains were found to be associated with the origin of isolation (Garmasheva *et al*., 2015). Strains AF211 and SRB521^T were isolated from different hosts; the observed variations in cellular fatty acids may reflect the different environmental conditions in their natural habitats.

In the gut environment, bacteria need to resist several stress conditions. One of these is bile, which can act as a surfactant and has the ability to affect the phospholipids and proteins of cell membranes (Begley *et al*., 2006). Bacteria are known to tolerate bile via two mechanisms: one is by using bile salt hydrolases (BSH) to hydrolyse bile and the other is by using efflux pumps to remove bile that passes through the outer membrane (Begley *et al*., 2006; Jones *et al*., 2008). The genomes of both AF211 and SRB521^T lacked the coding sequence for canonical BSHs but indicated the presence of genes for several MDR efflux pumps that could aid in resistance to bile. *In vitro* investigation of bile tolerance showed growth of strains AF211 and SRB521^T after 3 days of incubation in 0.1% bile (w/v) (Supporting Information Fig. S1) but no growth was observed in the presence of higher concentrations. This may be due to the absence of a bile salt hydrolase gene in both strains. While strain AF211 had a shorter lag time and grew to a higher OD than strain SRB521^T without bile the differences between doubling times with and without bile were most pronounced (an increase of 17 and 20.4 h for strain AF211 and SRB521^T respectively) (Supporting Information Fig. S1). The reduced bile inhibition of strain AF211 suggests that it to be better adapted to bile than the mouse strain although it may also reflect different adaptation to bile components since these are known to differ between mouse and man (Sayin *et al*., 2013).

Comparative genomics - general features of the I. butyriciproducens genomes

I. butyriciproducens strains AF211 and SRB521^T both have a single circular chromosome with similar genome features (Fig. 2A). No plasmid sequences were detected in either of the genomes. The genome size of AF211 is 112,795 bp larger than that of SRB521^T, indicating larger genetic potential in the human isolate that is also reflected in a higher number of coding sequences i.e. 3,363 vs. 3,268. Furthermore, the average nucleotide identity (ANI) between the two isolates confirmed that they belong to the same species. Additional genomic

A

	AF211	SRB521^T
Host	Human	Mouse
Accession number	CP011307	CP011524
Genome Size (Base pairs)	3,376,476	3,263,681
GC content (%)	59.1	59.4
RNA genes	57	58
Number of Coding Sequences	3363	3268
Number of Subsystems	348	338
Average nucleotide Identity (ANI)	98.80 %	98.80 %

B

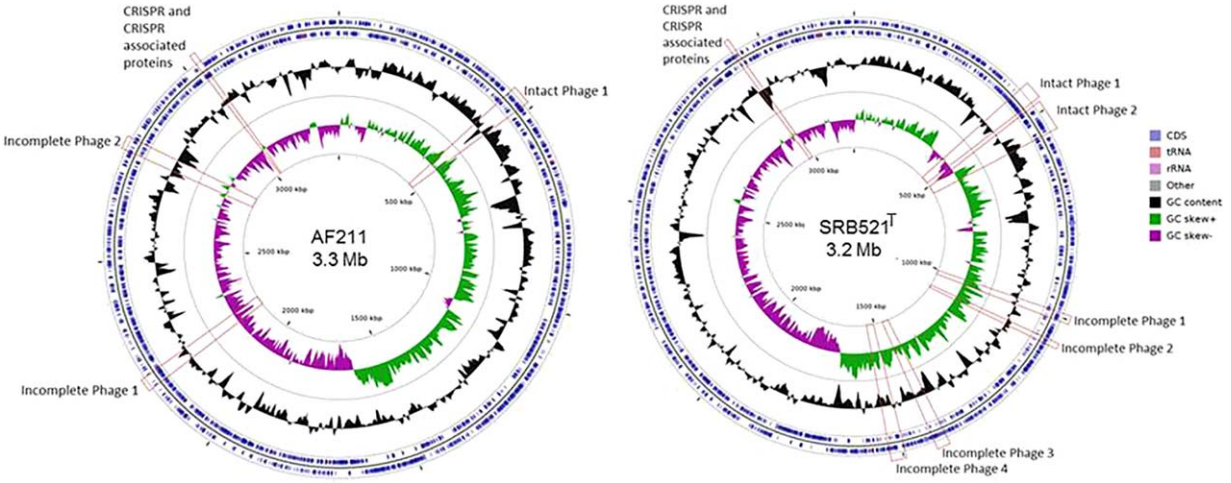


Fig. 2. General genome features (A) and genome maps (B) of *Intestinimonas butyriciproducens* isolates. From the outer circle inward, coding regions are marked on the first two rings: genes encoded on the positive strand (outside) and if encoded on the negative strand (inside). The third ring (black) shows local GC content. The innermost graph shows the CG skew. Phage regions and CRISPR associated genes are marked within red lines.

features such as the distribution of GC content, GC-skew, and location of genes including tRNA, rRNA, prophage genes are indicated in Fig. 2B.

Pan- and core genome

The pan-genome of *I. butyriciproducens* was calculated to contain 4,762 coding DNA sequences (CDS), of which 2,612 CDS (2,603,042bp) were considered as core. The average number of CDS predicted per genome was approximately 3,427, indicating that an average of 76.21% of the genome of each isolate was part of the core. The overall subsystem level core and accessory genome features were further analysed. This revealed that approximately 22.8% of the genome sequences were unique to one of the isolates signifying the strain level differences in the genomic content of AF211 and SRB521^T. The additional coding capacity in the AF211 was mostly assigned to genes coding for hypothetical proteins, mobile element proteins, and restriction enzymes that were located on genomic islands (Supporting Information Fig. S2).

Flexible genomic regions in I. butyriciproducens

To identify whether genomic features of *I. butyriciproducens* AF211 and SRB521^T could be linked to the different origin of the strains, a detailed analysis of their genomes was made. As already evident from differences in GC skew (Fig. 2), the two genomes carry a variety of genomic rearrangements including large inversions and integration of foreign regions of DNA (Fig. 3).

Alignment of the genomes of AF211 and SRB521^T revealed the presence of 16 locally collinear blocks (LCBs) with several regions of inversion and rearrangement (Fig. 3A). Moreover, an obvious X-pattern was

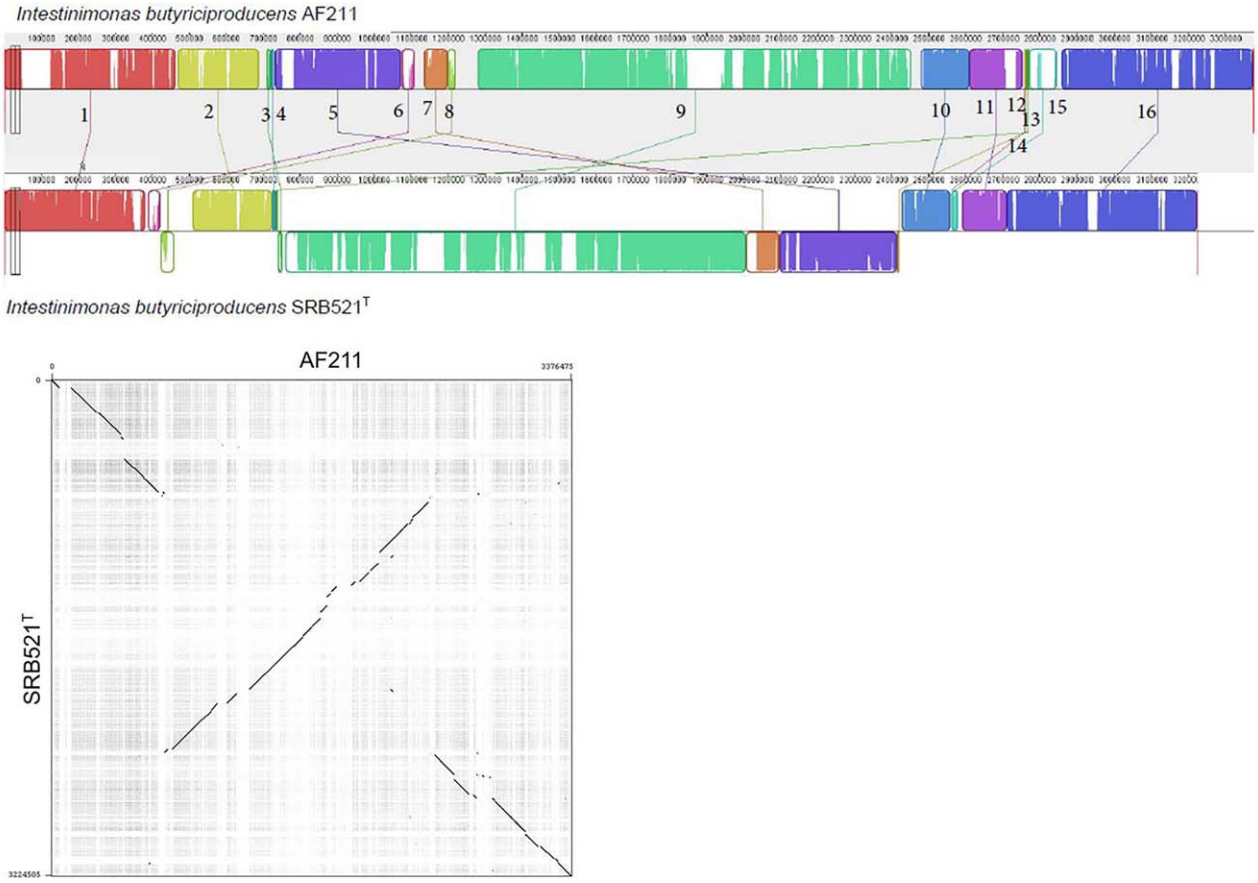


Fig. 3. Alignment of the genome sequences of *I. butyriciproducens* strain AF211 and SRB521^T.
A. The genome of strain AF211 was used as reference for global alignment using progressive MAUVE. The numbers indicate the locally collinear blocks (LCBs) that were identified in both the genomes. Conserved and highly related regions are coloured, and low-identity unique regions are in white (colourless). LCBs below the mid-line in strain SRB521^T are inverted relative to the strain AF211 sequence.
B. Dot plot alignments of AF211 and SRB521^T using Gepard. SRB521^T is on the y-axis and AF211 is on the x-axis and the genomes are starting with the origin of replication at position 1. The black lines represent regions of similarity while the breaks in this line (syntenic line) represent regions of genomic variations at a given locus between the two genomes.

observed when the genomes of AF211 and SRB521^T were compared in a dot blot, indicating a symmetric chromosomal inversion, involving approximately one third of the genome around the origin of replication (Fig. 3B). Not surprisingly, genes for a site-specific recombinase (SRB521_00768) and a probable integrase/recombinase (SRB521_002435c) flanking the inverted region in genome of SRB521^T were detected. One of the LCBs, which includes a site-specific recombinase gene (SRB521_00768), is located at 738168-739784 bp in the mouse isolate; the corresponding LCB in the human isolate with the same gene (AF_02777) is located at around 2770592-2772208 bp, signifying the role of this gene in genomic rearrangement.

In addition, we searched the genomes for the presence of genomic islands to identify non-self-mobilising elements that code for proteins with diverse functions. The AF211 genome contained 24 genomic islands (GIs) while SRB521^T had 18 (Supporting Information Fig. S2A). These include the already noted (incomplete) prophages (see Fig. 2 and Supporting Information Fig. S3) and series of other predicted gene clusters that are part of the GIs (Supplementary data 1). Remarkably, the majority of the genes located on the GI's encode hypothetical proteins in both AF211 (56%) and SRB521^T (63%).

Prophages and CRISPR systems in I. butyriciproducens

There has been considerable attention for the presence and function of bacteriophages in the human intestine. Specifically, the intestinal mucosa has been identified as a niche for bacteriophages, potentially indicating their role in maintaining microbiota dynamics (Barr *et al.*, 2013; Wootton, 2013). As bacteriophage diversity and composition may vary between mouse and man, it was of interest to analyse the natural defence systems of the two *I. butyriciproducens* strains, including the analysis of prophages known to provide superinfection immunity.

Comparative analysis of the genomes of AF211 and SRB521^T demonstrated major differences in the prophage elements. The genome of AF211 revealed one intact and 2 incomplete prophages while that of SRB521^T was predicted to have two intact and 4 incomplete prophages (see Supporting Information Fig. S3). Prophage integration sites in the genomes of both strains were found to have significantly different G + C content (Fig. 2A). In AF211, the prophages had varying G + C content, intact prophage 1 and incomplete prophage 1 had a higher G + C content of 61.77% and 62.42% respectively vs. 54.95% for incomplete prophage 2 when compared to the overall G + C content of 59.4% for the entire genome of AF211. In case of SRB521^T, intact prophage 1 had a G + C content of 59.29%, while that of intact prophage 2 was 64.15%. Additionally, in the SRB521^T genome, two of the incomplete prophages had higher G + C content (63.84% and 64.10%), while for the other two it was lower, 55.66% and 54.90% respectively. These observed differences suggest that the genes were acquired from both high and low G + C content bacterial taxa. The comparison of prophage sequences between the two genomes revealed that the two strains do not share any prophage sequences, indicating exposure to different prophages in human and mouse gut. This observation was also supported by the CRISPR/Cas system search in the two genomes. Both strains possessed clustered regularly interspaced short palindromic repeats (CRISPR/Cas system) in their genomes. The CRISPR array involves associated proteins to confer resistance to bacteriophages. The CRIPSR arrays contained the same 33 bp long direct repeat (5′-ATTTCAATCCACGCCCCCCGTGTGGGGGGC-GAC-3′). The number of spacers was 58 and 21 for strain AF211 and SRB521^T respectively. Interestingly, only two spacers were shared between the two strains (Supporting Information Fig. S2B). Blast analysis of the spacer regions revealed no sequence with complete identity to phage database entries, which can be attributed to the lack of phage sequences in Genbank Phage, and ACLAME genes databases. However, spacer 11 of the CRISPR system in strain SRB521^T targeted the phage minor tail protein in prophage reported in *Nocardia farcinica* IFM 10152. Interestingly, spacer 46 of the CRISPR system in strain AF211 appeared to target phages that made up 18% of total virus like particle (VLP) contig sequences in metagenomic DNA from human fecal samples (Minot *et al*., 2011). Harbouring an immune system against this highly abundant viral family could be beneficial for survival of AF211 in the human gut.

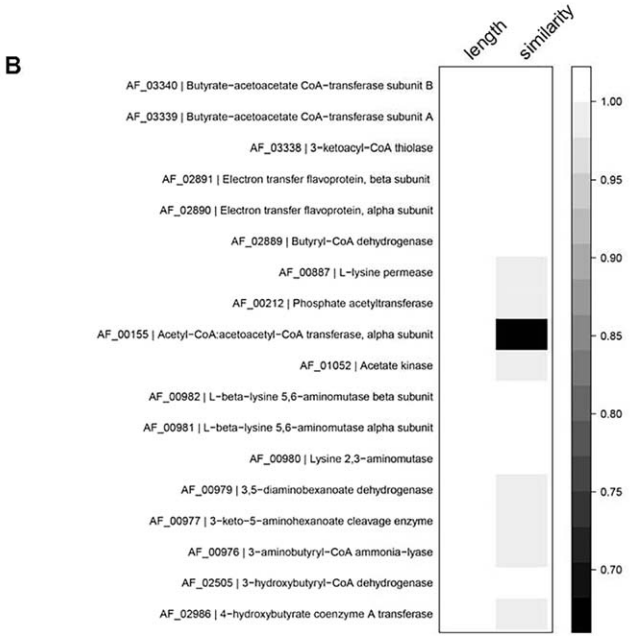
Butyrogenesis from sugars and lysine

As the intestine of men and mice differ considerably in the composition, load, and kinetics of dietary components, it was of importance to address the metabolic capacity of the two *I. butyriciproducens* strains. Genomic analysis revealed the presence of a complete glycolytic pathway while the pentose phosphate and the tricarboxylic acid pathways were incomplete in both strains. *In vitro* tests confirmed a mixed acid fermentation using variety of carbohydrate substrates, which resulted in production of butyrate as a major product and ethanol, isobutyrate or lactate as minor metabolites. The two strains grew poorly in hexose sugars but growth was much improved by the addition of acetate (Fig. 4A). A similar feature was previously described in *Faecalibacterium prausnitzii;* the presence of acetate increased energy harvest from glucose by this species (Heinken *et al*., 2014) via stimulation of the acetyl-CoA pathway. This pathway has a butyryl-CoA dehydrogenase (Bcd) electron-transferring flavoprotein (Etf) complex that generates a proton motive force via a membrane-integrated Rnf complex (Li *et al*., 2008). Comparison of the genes invovled in the acetyl-CoA pathway between the two genomes revealed high similarity in butyryl-CoA:acetate CoA transferases which are the key enzymes in butyrate formation (Fig. 4B). Neither of two strains was capable of utilizing polysaccharides (inulin, starch, mucin and pectin) in vitro although the genomes contained a number of glycoside hydrolases. Strain AF211 was characterized by a higher number of genes related to auxiliary activites, carbohydrate esterases, glycoside hydrolases, and glycoside transferases (Supporting Information Fig. S4). Genes coding for sugar phosphorylation such as hexokinase, galactosekinase, and xylulosekinase were detected, demonstrating potential to grow on hexose. Indeed, laboratory tests showed good growth on galactose, arabinose, glucose and some growth on fructose, maltose and lactose in presence of 10mM acetate and 2 g/l yeast extract but mannitol, cellobiose, raffinose, xylose, D-mannose, saccharose or sorbitol did not support growth (Fig. 4A). Butyrate production was higher in cultures of strain AF211 compared to SRB521^T when grown on galactose and arabinose, suggesting the higher effeciency in sugar conversions of the human isolate over the mouse isolate.

The amino acid lysine has been found as substrate for butyrogenesis in *I. butyriciproducens* AF211 (Bui *et al*., 2015). Strain SRB521^T showed similar capability of producing butyrate and acetate from lysine (Supporting Information Fig. S5) and the genome of this strain also coded for the entire lysine pathway (Fig. 4C). Genes involved in the lysine pathway in strain SRB521^T and AF211 were compared (Fig. 4B) and found to have the same length and high nucleotide identities (above 95%), except for the acetyl-CoA: acetoacetate CoA transferase gene (about 70%) (Fig. 4B,C). Thus, the lysine pathways could be a common feature of *I.*

A

Substrate	Strain	Consumption (mM)		Production (mM)			
		Glucose	Acetate	Lactate	Isobutyrate	Ethanol	Butyrate
GALACTOSE ACETATE	AF 211	0.3	4.95	0	1.20	1.93	6.83
	SRB 521^T	0.6	1.19	0.4	0.89	0	3.05
FRUCTOSE ACETATE	AF 211	0	1.58	0.1	1.01	1.04	3.47
	SRB 521^T	0.7	1.07	0	0.91	0	3.42
MALTOSE ACETATE	AF 211	0	0.76	0.1	0.97	0	2.40
	SRB 521^T	1.3	0.54	0	0	0	1.70
LACTOSE ACETATE	AF 211	1.15	1.06	0	1.05	0	2.75
	SRB 521^T	1.08	0.50	0.46	0	0	2.00
ARABINOSE ACETATE	AF 211	0.61	5.16	0.35	0.84	0.99	7.30
	SRB 521^T	0.89	0.12	0.44	2.63	0	2.11



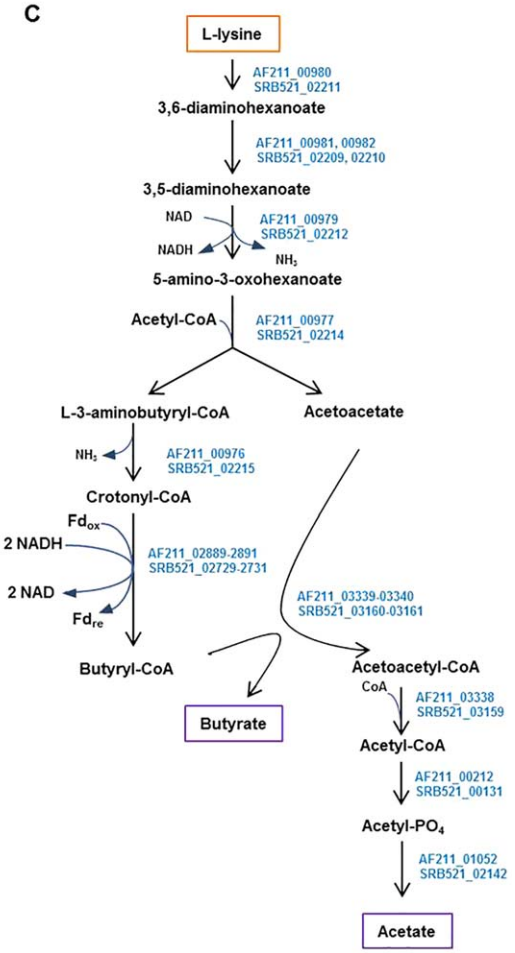


Fig. 4. Butyrogenesis in strain AF211 and SRB521^T.
A. Fermentation profiles of AF211 and SRB521^T in 10 mM sugars and 10 mM acetate for 2 weeks incubation. Product formation is calculated in mM. The highlighted values indicated discrepancy in butyrate formation and acetate consumption. Values were mean of duplicates and standard deviation was below 10%.
B. Genes involved in butyrate formation in strain AF211 and strain SRB521^T. The dark colour indicates low similarity of sequences of strain SRB521^T as compared to that of strain AF211.
C. Lysine conversion pathway in both strains: locus tags are indicated in blue.

butyriciproducens species, irrespective of the host environment. This unique capacity of *I. butyriciproducens* to grow on lysine positions this strain at the cross-road between protein metabolism and butyrogenesis in the gut ecosystem. Of note, a small amount of iso-butyrate was formed during sugar fermentation, which was associated with the supplementation of yeast extract. We considered the possibility of the Stickland reaction (Stickland, 1934). However, we found that growth on yeast extract alone did not give any iso-butyrate production and neither of the two strains grew in a mixture of amino acids (valine, serine, asparagine, threonine, isoleucine, aspartate, glutamate, methionine, citrulline, glycine). Iso-butyrate can also be formed via the branched chain amino acid pathway, as has been observed in a marine anaerobe although in this case no energy is produced and this pathway is assumed to prolong starvation survival (Harwood and Canale-Parola, 1981). A similar pathway may explain the observed iso-butyrate production in *I. butyriciproducens*, since both strains contain the genes for this pathway, including a branched chain amino acid aminotransferase, a phosphotransbutyrylase and several copies of oxidoreductases and kinases.

Antibiotic resistance of I. butyriciproducens

The ability of antibiotics to modulate the gut ecosystem is well known. Strain AF211 and SRB521^T showed a

A

Antibiotics	*I. butyriciproducens* SRB521^T (ug/ml)	*I. butyriciproducens* AF211 (ug/ml)
Ciprofloxacin	>32	>32
Cefotaxime	0.064	0.064-0.05
Erythromycin	1	0.75-1
Oxacillin	0.38-0.5	0.38
Teicoplanin	0.032	0.047
Tetracycline	>256	>256
Tobramycin	1.5-2	2
Vancomycin	0.75	0.75
Chloramphenicol	0.75	0.75
Sulfamethoxazole	>32	> 32

B

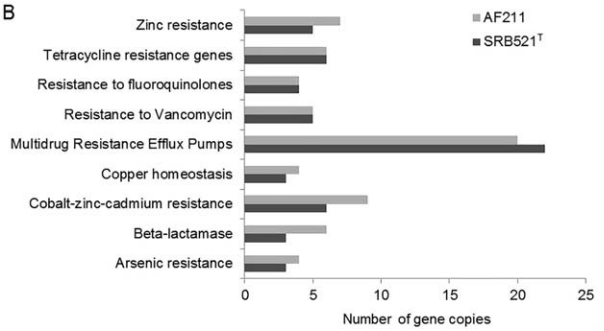


Fig. 5. Resistance genes in strain AF211 and SRB521^T.
A. Numbers of resistance genes in the genomes of *I. butyriciproducens* AF211 and SRB521^T.
B. Minimum inhibitory concentration of the 2 strains (μg/ml).

similar sensitivity profile to the tested antibiotics (Fig. 5A). Several resistance mechanisms appear to be encoded by the genomes of the two strains (Fig. 5B). Multidrug resistance efflux pumps were abundantly present with >20 copies in both strains. These are non-specific transport systems used to export (or import) toxic compounds in many bacteria and involved in housekeeping functions (Poelarends *et al*., 2000). Antibiotic resistance analysis using the E-test showed highest minimum inhibitory concentration (MIC) on tetracycline (>256 μg/ml) for both strains and lowest MIC on teicoplanin (0.032 μg/ml for AF211 and 0.047 μg/ml for SRB521^T). Results for tetracycline in the present study were higher from that reported earlier for strain SRB521^T (Kläring *et al*., 2013) as the E-test was done in different media and incubation condition that might result in different MIC values (Huys *et al*., 2002). Tetracycline is a protein synthesis inhibitor that binds to the 30S subunit of ribosomes. The observed tetracycline resistance may be due to the potential production of a TetW protein that shares 99% identity with TetW-like proteins in *Roseburia, Bifidobacterium* and some other intestinal strains. In addition, the presence of multiple efflux pumps may also contribute to tetracycline resistance (Fig. 5), as demonstrated in *Pseudomonas aeruginosa* (Li *et al*., 1994) and *E. coli* (Okusu *et al*., 1996) and subject to activation by a transcriptional regulator TetR, which was also found to be predicted from the genomes of both *I.butyriciproducens* strains. The two strains were sensitive to teicoplanin, vancomycin, cefotaxime, and oxacillin, which are inhibitors of cell wall synthesis in Gram-positive bacteria, but they showed relatively high MIC values with ciprofloxacin and sulfamethoxazole (Fig. 5B). Although (remnants of) some vancomycin resistance genes (*vanB, vanW, vanS* and *vanR*) were detected in the two genomes, these were not sufficient to provide vancomycin resistance, which requires a complex operon with at least five genes (Kruse *et al*., 2014). A set of genes predicted to code for resistance to inorganic compounds including arsenic resistance protein, cobalt-zinc-cadmium resistance protein, and copper resistance protein, were also found as part of the defence system of both strains.

Oxidative stress, transport systems, and vitamin production in I. butyriciproducens

The human intestine is considered as an anaerobic ecosystem, especially the large intestine, even though there is a gradient of oxygen from the epithelial layer outward (Albenberg *et al*., 2014). The reduction of oxygen results in production of superoxide radicals, hydrogen peroxide and hydroxyl radicals that are extremely toxic for cells (Fridovich, 1995). Hence, it is important for intestinal microbes that possess defence systems to prevent accumulation of these reactive oxygen species. Moreover, the oxidation of substrates and various xenobiotics might generate toxic oxidizing compounds in the environment, which intestinal anaerobes need to deal with for survival (Brioukhanov and Netrusov, 2004). Several genes predicted to have protective role under oxidative stress were detected in both strains (Supporting Information Table S3). Among them, CoA-disulfide reductase, rubrerythrin, superoxide reductase, and alkyl hydroperoxide reductase subunit C-like protein were detected in the proteome of strain AF211 (Bui *et al* 2015), indicative of active functions of these proteins. Superoxide reductase (EC 1.15.1.2) and manganese superoxide dismutase (EC 1.15.1.1) aid in lowering superoxide concentration and this process is carried out in the cytosol (Sheng *et al*., 2014) while catalase (EC 1.11.1.6) prevents accumulation of hydrogen peroxide. These enzymatic activities have been tested in strict anaerobes including Clostridia*, Bacteroides* spp., acetogens, sulfate reducers, and methanogens, indicating a wide distribution among anaerobes (Brioukhanov and Netrusov, 2004). It has been reported that cell starvation also results in the synthesis of antioxidative defence systems apart from O_2 (Rocha and Smith, 1997). Genes encoding the protein rubrerythrin were most abundantly present in the genomes of either strain (6-8 copies). Rubrerythrin has a protective effect under oxidative

stress, although its mechanism of action still remains unknown (Mukhopadhyay *et al*., 2007). In addition, a gene for the peroxide stress regulator PerR was found in both genomes. Interestingly, the alkyl hydroperoxide reductase subunit C (AhpC) was found to be encoded by strain AF211 but not SRB521^T. AhpC is involved in organic peroxide detoxification (Rocha and Smith, 1999). The presence of different oxidative stress defence systems in the two anaerobic strains enhances their survival in the gut environment where there is always a gradient of oxygen toward lumen.

Membrane transporters that facilitate the exchange of ions, nutrients and metabolites with the environment are essential for bacterial metabolism. A total of 102 genes were annotated as membrane transporters in the genome of strain AF211, and 86 in strain SRB521^T for the same mechanisms of transport system (Supporting Information Fig. S6). This suggests that strain AF211 has an improved overall capability of exchanging molecules with its environment. For both strains, ECF class transporters and ABC transporters were the most prevalent groups (10 to 30 copies) while symporter and antiporter were the least abundant ($<$2 copies) (Supporting Information Fig. S6). The analysis showed that the genome of strain AF211 harboured more copies of genes for ECF class transporters (30 versus 24), ABC transporters (47 versus 43), Nickel/Cobalt transporters (9 versus 4) and TRAP transporters (8 versus 6) than strain SRB521^T, and this might result in better capability of transporting nutrients. The ECF transporters are present in various microbial lineages and responsible for transporting vitamins (Rodionov *et al*., 2009). ABC transporters for oligopeptides, dipeptides and branched-chain amino acids were found, indicating the capability of taking up these compounds, which corresponds with our previous report of growth on different protein-derived sources (Bui *et al*., 2015). Galactose and multiple sugar ABC transporters were predicted from the genomes of both strains. Although sugar phosphotransferase systems (PTS) are often present in Gram-positive bacteria, no PTS system was found in the genomes.

Vitamin B12 is an essential cofactor for bacterial and host metabolism and has been suggested to modulate the intestinal microbe-host symbiosis (Degnan *et al*., 2014). This vitamin is produced by a selected group of intestinal microbes and hence there is microbial competition for uptake of this metabolite (Allen and Stabler, 2008). We showed that strain AF211 is capable of producing pseudo-vitamin B12 (Supporting Information Fig. S7). Although we did not test for vitamin B12 synthesis in strain SRB521^T, genomic analysis of strain SRB521^T showed an identical vitamin B12 synthesis pathway to that of strain AF211, indicating the capability of strain SRB521^T to produce a pseudo-vitamin B12 (data not shown). Some genes in the corresponding metabolic pathway appeared to be missing, but this can be attributed to the shortcomings of annotations currently available in present genome databases. It has been demonstrated in *Clostridium sticklandii* that lysine 5,6-aminomutase is a B12-dependent enzyme (Chang and Frey, 2000). This enzyme also appeared as a key player in lysine fermentation by *I. butyriciproducens* (Bui *et al*., 2015), indicating that this strain produced vitamin B12 to support its metabolic conversion of lysine into butyrate. In addition, several other conversions also require vitamin B12 as cofactor, including the conversion of 1,2-propanediol to propionate, ethanolamine to acetaldehyde and ammonia, and glycerol to 3-hydroxypropionaldehyde (the antimicrobial reuterin) (Garsin, 2010; Engels *et al*., 2016). This could also apply to *I. butyriciproducens* although initial experiments showed no growth in the bicarbonate buffered medium on glycerol, 1,2-propanediol or ethanolamine. We did observe that both *I. butyriciproducens* strains encode a vitamin B12 ABC transporter comprising the membrane proteins BtuF (SRB521_01945, SRB521_01948 and AF_01368c, AF_01372c) and BtuC (AF_01370c, SRB521_01947), as well as the iron B12 siderophore hemin periplasmic substrate-binding component. This suggests that the *I. butyriciproducens* strains have the ability to exchange B12 with the surrounding environment and hence are contributing to the intestinal microbe-host symbiosis.

Conclusions

The 16S rRNA amplicon metadata search revealed that *I. butyriciproducens* is the most common known species within the *Intestinimonas* genus and is most prevalent in humans. Comparative genomic and physiological analysis of *I. butyriciproducens* strains AF211 and SRB521^T isolated from the human and mouse intestine, respectively, confirmed the so far unique capability of using lysine for growth as a major feature of the species. Lysine is abundant in the gut as it is part of human diet and is released via microbial activity. In healthy subjects lysine is usually excreted in the faeces and hence its conversion by *Intestinimonas butyriciproducens* prevents this loss. Moreover, *Intestinimonas butyriciproducens* converts lysine into butyrate, which is an energy source for colonocytes, and releases ammonium, which can be used as nitrogen source by bacteria and neutralises acidity created by SCFA production. Both strains of *I. butyriciproducens* contained 2 butyrate synthesis pathways, where both lysine and simple sugars can serve as energy source. However, physiological data suggested that AF211 is more efficient in utilizing human-specific sugars such as arabinose and galactose in presence of acetate. In conclusion, the present study provides

detailed genomic and physiological insight into newly discovered *I. butyriciproducens* strains with unique metabolic capabilities and differentiates two strains of this species with different host origins.

Experiment procedures

Strains and growth condition

I. butyriciproducens strain AF211 was obtained from a stool sample of a healthy individual (Bui *et al*., 2015) and *I. butyriciproducens* strain SRB521^T was isolated from the caecum of a mouse (Kläring *et al*., 2013). Bacteria were routinely grown in anaerobic Reinforced Clostridium Medium (RCM, Difco) in 120-ml serum bottles sealed with butyl-rubber stoppers at 37°C under a gas phase of 1.7 atm of N_2/CO_2 (80: 20, v/v).

Metadata search of Intestinimonas *spp. using 16S rRNA gene amplicons*

To explore the ecological distribution and sequence supported diversity of the genus *Intestinimonas* we followed a modified version of a method previously described (Lagkouvardos *et al*., 2014; Kläring *et al*., 2015) and now available through the IMNGS web platform at www.imngs.org (Lagkouvardos *et al*., 2016). In short, the 16S rRNA gene of the type strain of *I. butyriciproducens* strain SRB521^T (NR_118554) was used for a similarity query (95%) against all 16S rRNA sequences available from the SRA in IMNGS (build 1508). The obtained sequences were filtered, aligned and used for calculation of species-like clusters (approx. 3% similarity) using CROP (Hao *et al*., 2011). The detailed method is available as supplementary materials.

Genome sequencing and assembly

Cells of strain AF211 and SRB521^T grown in RCM (overnight cultures) were used for genomic DNA extraction performed using ZR Fungal/Bacteria DNA MiniPrep kit (ZYMO, USA) according to manufacturer's instructions. Sequencing effort included the use of two complementary platforms based on the Illumina and PacBio technologies.

Genome sequencing of two 15 kb libraries using with PacBio RS II instrument using P4/C2 chemistry (Pacific Biosciences, Menlo Park CA, USA) at the DNA sequencing and genomics laboratory, Institute of Biotechnology, University of Helsinki, Finland. Data processing and filtering was done with PacBio SMRT analysis pipeline v2.2 and the Hierarchical Genome Assembly Process (HGAP) protocol (http://www.pacb.com/devnet/). Filtering was run with minimum subread length 500 and polymerase read length quality 0.80. For pre-assembly, settings were: minimum seed read length 7000, split target into chunks 1 and alignment candidate per chunk 24. Assembly was performed by the Celera V1 assembler with parameters genome size 3,000,000, target coverage 30, overlapper error rate 0.06, overlapper mini length 40 and overlapper K-mer 14. Assembly polishing was done with Quiver.

In addition, genome sequences were also collected using a HiSeq2000. This resulted in 3,202,992 and 2,759,530 paired reads for AF211 and SRB521^T, respectively, which were used to correct for possible PacBio sequencing errors. For the HiSeq paired-end reads, all rRNA reads were removed with SortMeRNA v1.9 using default settings except for an increased error value of 20% for the adaptors, and using the reverse complement of the adaptors as well (Kopylova *et al*., 2012). Quality trimming was performed with PRINSEQ Lite v0.20.0 with a minimum sequence length of 40 and a minimum quality of 30 on both ends and as mean quality of read (Schmieder and Edwards, 2011). For assembly, Ray v2.3 was used (Boisvert *et al*., 2012). Duplicate contigs in both the genome assemblies of AF211 and SRB521^T were discarded if they had a hit with at least 99% sequence identity within a bigger contig, which spanned at least 98% of the contig query length. Furthermore, contigs with length $<$500 bp were discarded. Next, Illumina contigs were aligned to the PacBio assembly using promer algorithm from the MUMmer 3.0 software package (Kurtz *et al*., 2004) with the parameters show-tiling -c -R -i 50. The alignment was processed in R (http://www.r-project.org/) using in-house scripts to extract the Illumina sequence and combined with PacBio assembly at gap positions to produce the final closed genome sequence. A circular element was detected within this assembly, based on the BLASTP results of the predicted proteins (e-value 0.0001), and this added to the final assembly result. Scaffolding of the contigs was done with SSPACE-LongRead and the PacBio CCS reads using default settings (Boetzer and Pirovano, 2014). The annotation was done by RAST server (Aziz *et al*., 2008). Functional prediction of proteins was verified manually by BLASTing the amino acid sequences in Pfam (Finn *et al*., 2014), Brenda, Interpro (Hunter *et al*., 2012) and Uniprot databases. For a global overview of the genome features, the circular genome map was created using the virtual machine CGView Comparison Tool (CCT) (Grant *et al*., 2012).

Specific annotations

The core genome of *I. butyriciproducens* was calculated using Spine and the pan-genome was identified using AGEnt (Ozer *et al*., 2014). For genomes of two strains, specific annotations for CRISPR genes was done using CRISPRfinder (Grissa *et al*., 2007). Target phages for

CRISPR were identifies using CRISPRtarget using Genbank Phage, and ACLAME genes databases with an E-value cut-off of 0.1 (Biswas *et al*., 2013). The CRISPRTarget tool identifies target phages using different databases and the search algorithm has its own scoring system with a score of 20 which was kept as default minimum as suggested by the authors (Biswas *et al*., 2013). All the spacers targeting phages in both $SRB521^T$ and AF211 had an acceptable score of 21. Prophage genes in the genomes were annotated using PHAST (PHAge Search Tool) (Zhou *et al*., 2011). The PHAST uses a scoring system based on three methods (i) Presence of known phage genes/proteins; (ii) $>50\%$ genes/proteins in the region are related to a known phage; and (iii) $<50\%$ genes/proteins in the region are related to a known phage (Zhou *et al*., 2011). In general, if a given phage region's total score is <70, it is marked as incomplete; between 70 to 90, it is marked as questionable; if >90, it is marked as intact. The carbohydrate active enzymes (CAZymes) were annotated using the amino acid sequences with default search parameters (Park *et al*., 2010). Genomic islands were identified using online IslandViewer and a standalone program i.e. SeqWord Gene Island Sniffer that identifies genomic islands based on the analysis of oligonucleotide usage variations in DNA sequences (Bezuidt *et al*., 2009; Langille and Brinkman, 2009; Dhillon *et al*., 2013).

Phenotypic characterization

To investigate carbohydrate assimilation and detection of enzyme activities, API 20NE, API rapid 32, API ZYM were used with overnight-grown cells of both strains. The Gram reaction was determined using standard methods (Plugge *et al*., 2000). In-vivo DNA-DNA hybridization was done at the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany), as described previously (Ley *et al*., 1970) with some modifications (Huss *et al*., 1983).

For cellular fatty acid analysis, cells of both strains were grown in RCM medium for 5 days before being centrifuged (10,000 rpm) for 10 min at 4°C. Pellets were stored at −80°C and later used to extract cellular fatty acids as described previously (Kämpfer and Kroppenstedt, 1996). Products were analysed by Agilent model 6890N gas 199 chromatography (MIDI Sherlock, Newark, N.J.) as described previously (Miller, 1982; Kämpfer and Kroppenstedt, 1996).

Bile tolerance was tested in RCM supplemented with OX-Bile (0.1; 0.2; 0.3; 0.4; 0.5% (w/v); SIGMA ALDRICH). Growth was determined via OD measurement and product formation using HPLC as described previously (van Gelder *et al*., 2012).

Growth on carbohydrates and amino acids

Intestinimonas butyriciproducens strains were grown in 50 ml anaerobic bicarbonate buffered medium (Stams *et al*., 1993) with 2 g/l yeast extract. Carbohydrates were added from 1M sterile stock solutions (10 mM final concentration). Tested carbohydrates were glucose, fructose, maltose, mannitol, cellobiose, lactose, raffinose, xylose, D-mannose, saccharose, galactose, D,L- arabinose and sorbitol. The polysaccharides starch, inulin, pectin, and mucin were added before autoclaving and used at a final concentration of 0.5% (w/v). Valine, serine, asparagine, threonine, isoleucine, aspartate, glutamate, methionine, citrulline, glycine were added to make 10 mM at final concentration from 1M stock filter-sterile solutions. The growth in the mixture of amino acid was performed in the bicarbonate buffered medium containing 0.1 g/l yeast extract. Samples were collected at different time points and fermentation profiles were analysed using GC and HPLC analysis as described previously (van Gelder *et al*., 2012). A Thermo Scientific Spectra HPLC system was used to detect organic acids including lactate, formate, acetate, propionate, isobutyrate, butyrate and alcohols including methanol, ethanol, 1,2 propanediol, 2-propanol and butanol. The HPLC system was equipped with a Agilent Metacarb 67H 300 × 6.5 mm column kept at 45°C and running with 5 mM H_2SO_4 as eluent and a refractive index detector.

Antibiotic susceptibility tests

The E-test was done to identify minimal inhibitory concentrations (MICs) according to the manufacturer's protocol (bioMérieux, France). Both strains were pre-grown in RCM broth (overnight cultures) and 50 μl was spread on RCM agar plates (1.5% w/v agar) until the agar surface was dry and the liquid was absorbed by the agar. Two E-test strips were placed on one plate to test each antibiotic. Results were noted according to manufacturer's protocol (bioMérieux, France). Antibiotics tested included ciprofloxacin, cefotaxime, erythromycin, oxacillin, teicoplanin, tetracycline, tobramycin, vancomycin and sulfamethoxazole. The concentration range was 0.016–256 μg/ml for chloramphenicol, oxacillin, tetracycline, tobramycin and vancomycin, and 0.016–32 μg/ml for ciprofloxacin, cefotaxime, erythromycin, teicoplanin and sulfamethoxazole. MIC values were recorded directly from the strips after 48 h and rechecked after 4 days. Duplicate experiments were performed for each antibiotic.

Vitamin B12 analysis

Cell pellet (0.5 g) was suspended in 10 ml of extraction buffer (pH 4.5; 8.3 mM sodium hydroxide and 20.7 mM acetic acid) containing one mg of sodium cyanide. Following heat extraction (100°C; 30 min) and cooling on

ice, the extract was recovered by centrifugation (6900 *g*; 10 min). Five ml of the extract was purified using an immunoaffinity column (Easy-Extract; R-Biopharma, Glasgow, UK) according to the manufacturer's instructions, and finally recovered in 500 μl of water. To identify the vitamin form, the extract was analysed using UHPLC (Waters Inc.; USA) and quadrupole time-of-flight mass spectrometer (Q-TOF; Synapt, G2-Si, Waters) as described previously (Deptula *et al*., 2015).

Acknowledgement

This research was partly supported by ERC Advanced Grant 250172 - MicrobesInside from the European Research Council and the Netherlands Organization for Scientific Research (Spinoza Award and SIAM Gravity Grant 024.002.002) to WMdV. TC received financial support from the German Research Foundation (grant CL481/2-1).

Conflict of Interest

There are no relevant conflicts of interest.

References

Albenberg, L., Esipova, T.V., Judge, C.P., Bittinger, K., Chen, J., Laughlin, A., *et al*. (2014) Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota in humans and mice. *Gastroenterology* **147**: 1055–1063.

Allen, R.H., and Stabler, S.P. (2008) Identification and quantitation of cobalamin and cobalamin analogues in human feces. *Am J Clin Nutr* **87**: 1324–1335.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., *et al*. (2011) Enterotypes of the human gut microbiome. *Nature* **473**: 174–180.

Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., *et al*. (2008) The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., *et al*. (2013) Bacteriophage adhering to mucus provide a non–host-derived immunity. *Proc Natl Acad Sci U S A* **110**: 10771–10776.

Begley, M., Hill, C., and Gahan, C.G.M. (2006) Bile salt hydrolase activity in probiotics. *Appl Environ Microbiol* **72**: 1729–1738.

Bezuidt, O., Lima-Mendez, G., and Reva, O. (2009) SeqWord gene island sniffer: A program to study the lateral genetic exchange among bacteria. *World Acad Sci Eng Technol* **58**: 1169–11274.

Biswas, A., Gagnon, J.N., Brouns, S.J., Fineran, P.C., and Brown, C.M. (2013) CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol* **10**: 817–827.

Boetzer, M., and Pirovano, W. (2014) SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**: 211.

Boisvert, S., Raymond, F., Godzaridis, É., Laviolette, F., and Corbeil, J. (2012) Ray Meta: Scalable de novo metagenome assembly and profiling. *Genome Biol* **13**: R122–R122.

Brioukhanov, A.L., and Netrusov, A.I. (2004) Catalase and superoxide dismutase: Distribution, properties, and physiological role in cells of strict anaerobes. *Biochemistry* **69**: 949–962.

Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., *et al*. (2016) Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**: 543–546.

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.

Chang, C.H., and Frey, P.A. (2000) Cloning, sequencing, heterologous expression, purification, and characterization of adenosylcobalamin-dependentd-Lysine 5,6-aminomutase from *Clostridium sticklandii*. *J Biol Chem* **275**: 106–114.

de Vos, W.M., and Nieuwdorp, M. (2013) Genomics: A gut prediction. *Nature* **498**: 48–49.

Degnan, P.H., Taga, M.E., and Goodman, A.L. (2014) Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab* **20**: 769–778.

Deptula, P., Kylli, P., Chamlagain, B., Holm, L., Kostiainen, R., Piironen, V., *et al*. (2015) BluB/CobT2 fusion enzyme activity reveals mechanisms responsible for production of active form of vitamin B_{12} by *Propionibacterium freudenreichii*. *Microb Cell Fact* **14**: 186.

Dhillon, B.K., Chiu, T.A., Laird, M.R., Langille, M.G., and Brinkman, F.S. (2013) IslandViewer update: Improved genomic island discovery and visualization. *Nucleic Acids Res* **41**: W129–W132.

Engels, C., Ruscheweyh, H.J., Beerenwinkel, N., Lacroix, C., and Schwab, C. (2016) The common gut microbe *Eubacterium hallii* also contributes to intestinal propionate formation. *Front Microbiol* **7**: 713.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., *et al*. (2014) Pfam: The protein families database. *Nucleic Acids Res* **42**: D222–D230.

Flint, H.J., N.A., Logan, H.M.L.-S., and Oyston, P.C.F. (2006) *The Significance of Prokaryote Diversity in the Human Gastrointestinal Tract. Prokaryotic Diversity*. Cambridge University Press, UK.

Flint, H.J., Scott, K.P., Louis, P., and Duncan, S.H. (2012) The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* **9**: 577–589.

Fridovich, I. (1995) Superoxide radical and superoxide dismutases. *Annu Rev Biochem* **64**: 97–112.

Garmasheva, I., Vasyliuk, O., Kovalenko, N., Ostapchuk, A., and Oleschenko, L. (2015) Intraspecies cellular fatty acids heterogeneity of *Lactobacillus plantarum* strains isolated from fermented foods in Ukraine. *Lett Appl Microbiol* **61**: 283–292.

Garsin, D.A. (2010) Ethanolamine utilization in bacterial pathogens: Roles and regulation. *Nat Rev Microbiol* **8**: 290–295.

Grant, J.R., Arantes, A.S., and Stothard, P. (2012) Comparing thousands of circular genomes using the CGView Comparison Tool. *BMC Genomics* **13**: 202.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007) CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35**: W52–W57.

Hamer, H.M., Jonkers, D., Venema, K., Vanhoutvin, S., Troost, F.J., and Brummer, R.J. (2008) Review article: The role of butyrate on colonic function. *Aliment Pharmacol Ther* **27**: 104–119.

Hao, X., Jiang, R., and Chen, T. (2011) Clustering 16S rRNA for OTU prediction: A method of unsupervised Bayesian clustering. *Bioinformatics* **27**: 611–618.

Harwood, C.S., and Canale-Parola, E. (1981) Branched-chain amino acid fermentation by a marine *Spirochete*: Strategy for starvation survival. *J Bacteriol* **148**: 109–116.

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.

Hinnebusch, B.F., Meng, S., Wu, J.T., Archer, S.Y., and Hodin, R.A. (2002) The effects of short-chain fatty acids on human colon cancer cell phenotype are associated with histone hyperacetylation. *J Nutr* **132**: 1012–1017.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., *et al*. (2012) InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Res* **40**: 4725.

Huss, V.A.R., Festl, H., and Schleifer, K.H. (1983) Studies on the spectrophotometric determination of DNA hybridization from renaturation rates. *Syst Appl Microbiol* **4**: 184–192.

Huys, G., D'haene, K., and Swings, J. (2002) Influence of the culture medium on antibiotic susceptibility testing of food-associated lactic acid bacteria with the agar overlay disc diffusion method. *Lett Appl Microbiol* **34**: 402–406.

Jones, B.V., Begley, M., Hill, C., Gahan, C.G.M., and Marchesi, J.R. (2008) Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci U S A* **105**: 13580–13585.

Kämpfer, P., and Kroppenstedt, R.M. (1996) Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Canadian J Microbiol* **42**: 989–1005.

Kant, R., Rasinkangas, P., Satokari, R., Pietilä, T.E., and Palva, A. (2015) Genome sequence of the butyrate-producing anaerobic bacterium *Anaerostipes hadrus* PEL 85. *Genome Announc* **3**(2):e00224–15.

Klampfer, L., Huang, J., Sasazuki, T., Shirasawa, S., and Augenlicht, L. (2003) Inhibition of Interferon γ signaling by the short chain fatty acid butyrate. *Mol Cancer Res* **1**: 855–862.

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.

Kläring, K., Just, S., Lagkouvardos, I., Hanske, L., Haller, D., Blaut, M., *et al*. (2015) *Murimonas intestini* gen. nov., sp. nov., an acetate-producing bacterium of the family *Lachnospiraceae* isolated from the mouse gut. *Int J Syst Evol Microbiol* **65**: 870–878.

Kopylova, E., Noé, L., and Touzet, H. (2012) SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Kruse, T., Levisson, M., de Vos, W.M., and Smidt, H. (2014) vanI: A novel d-Ala-d-Lac vancomycin resistance gene cluster found in *Desulfitobacterium hafniense*. *Microb Biotechnol* **7**: 456–466.

Kurtz, S., Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.

Lagkouvardos, I., Weinmaier, T., Lauro, F.M., Cavicchioli, R., Rattei, T., and Horn, M. (2014) Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. *ISME J* **8**: 115–125.

Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., and Clavel, T. (2016) IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep* **6**: 33721.

Langille, M.G., and Brinkman, F.S. (2009) IslandViewer: An integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25**: 664–665.

Levine, U.Y., Looft, T., Allen, H.K., and Stanton, T.B. (2013) Butyrate-producing bacteria, including mucin degraders, from the swine intestinal tract. *Appl Environ Microbiol* **79**: 3879–3881.

Ley, J.D., Cattoir, H., and Reynaerts, A. (1970) The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**: 133–142.

Li, F., Hinderberger, J., Seedorf, H., Zhang, J., Buckel, W., and Thauer, R.K. (2008) Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*. *J Bacteriol* **190**: 843–850.

Li, X.Z., Livermore, D.M., and Nikaido, H. (1994) Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: Resistance to tetracycline, chloramphenicol, and norfloxacin. *Antimicrob Agents and Chemother* **38**: 1732–1741.

Louis, P., Young, P., Holtrop, G., and Flint, H.J. (2010) Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ Microbiol* **12**: 304–314.

Miller, L.T. (1982) Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* **16**: 584–586.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., *et al*. (2011) The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res* **21**: 1616–1625.

Mukhopadhyay, A., Redding, A.M., Joachimiak, M.P., Arkin, A.P., Borglin, S.E., Dehal, P.S., *et al*. (2007) Cell-wide responses to low-oxygen exposure in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* **189**: 5996–6010.

Okusu, H., Ma, D., and Nikaido, H. (1996) AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (Mar) mutants. *J Bacteriol* **178**: 306–308.

Ozer, E.A., Allen, J.P., and Hauser, A.R. (2014) Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics* **15**: 737.

Park, B.H., Karpinets, T.V., Syed, M.H., Leuze, M.R., and Uberbacher, E.C. (2010) CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* **20**: 1574–1584.

Plugge, C.M., Zoetendal, E.G., and Stams, A.J.M. (2000) *Caloramator coolhaasii* sp. nov., a glutamate-degrading, moderately thermophilic anaerobe. *Int J Syst Evol Microbiol* **50**: 1155–1162.

Poelarends, G.J., Mazurkiewicz, P., Putman, M., Cool, R.H., Veen, H.W., and Konings, W.N. (2000) An ABC-type multidrug transporter of *Lactococcus lactis* possesses an exceptionally broad substrate specificity. *Drug Resist Updat* **3**: 330–334.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., *et al*. (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**: 55–60.

Rajilić-Stojanović, M., and de Vos, W.M. (2014) The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev* **38**: 996–1047.

Rettedal, E.A., Gumpert, H., and Sommer, M.O.A. (2014) Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nat Commun* **5**: 4714.

Rocha, E.R., and Smith, C.J. (1997) Regulation of *Bacteriodes fragilis* katB mRNA by oxidative stress and carbon limitation. *J Bacteriol* **179**: 7033–7039.

Rocha, E.R., and Smith, C.J. (1999) Role of the alkyl hydroperoxide reductase (ahpCF) gene in oxidative stress defense of the obligate anaerobe *Bacteroides fragilis*. *J Bacteriol* **181**: 5701–5710.

Rodionov, D.A., Hebbeln, P., Eudes, A., ter Beek, J., Rodionova, I.A., Erkens, G.B., *et al*. (2009) A novel class of modular transporters for vitamins in prokaryotes. *J Bacteriol* **191**: 42–51.

Sayin, S.I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H.U., Bamberg, K., *et al*. (2013) Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab* **17**: 225–235.

Schmieder, R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.

Scott, K.P., Martin, J.C., Campbell, G., Mayer, C.D., and Flint, H.J. (2006) Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "*Roseburia inulinivorans*". *J Bacteriol* **188**: 4340–4349.

Sheng, Y., Abreu, I.A., Cabelli, D.E., Maroney, M.J., Miller, A.F., Teixeira, M., and Valentine, J.S. (2014) Superoxide dismutases and superoxide reductases. *Chem Rev* **114**: 3854–3918.

Sheridan, P.O., Martin, J.C., Lawley, T.D., Browne, H.P., Harris, H.M.B., Bernalier-Donadille, A., *et al*. (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic *Firmicutes*. *Microb Genomics* **2**: 000043.

Stams, A.J.M., Van Dijk, J.B., Dijkema, C., and Plugge, C.M. (1993) Growth of syntrophic propionate-oxidizing bacteria with fumarate in the absence of methanogenic bacteria. *Appl Environ Microbiol* **59**: 1114–1119.

van Gelder, A.H., Aydin, R., Alves, M.M., and Stams, A.J.M. (2012) 1,3-Propanediol production from glycerol by a newly isolated *Trichococcus* strain. *Microb Biotechnol* **5**: 573–578.

Vital, M., Howe, A.C., and Tiedje, J.M. (2014) Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**: e00889.

Wootton, L. (2013) Phage biology: A new barrier at mucosal surfaces. *Nat Rev Microbiol* **11**: 430–431.

Xiao, L., Feng, Q., Liang, S., Sonne, S.B., Xia, Z., Qiu, X., *et al*. (2015) A catalog of the mouse gut metagenome. *Nat Biotechnol* **33**: 1103–1108.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011) PHAST: A fast phage search tool. *Nucleic Acids Res* **39**: W347–W352.

Zoetendal, E.G., Vaughan, E.E., and De Vos, W.M. (2006) A microbial world within us. *Mol Microbiol* **59**: 1639–1650.

Supporting Information

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

Table S1. Physiological and biochemical comparison of *I. butyriproducens* strain AF211 and SRB521^T. The experiments were performed by using API Rapid 32 (A), API 20NE (B), API ZYM (C). +: positive response; -: negative response; w: weak response; vw: very weak response.

Table S2. Cellular fatty acid composition of the strain AF211 and SRB521^T. Data were obtained in the present study. All strains were grown in RCM for 5 days at 37°C. Values are percentages of total cellular fatty acids.

Table S3. Number of genes associated to oxidative stress in AF211 and SRB521^T

Fig. S1. Growth of strains AF211 and SRB521^T in the presence and absence of 0.1 % bile. Strains were grown in RCM and RCM plus 0.1 % bile (Oxgall). The doubling times of strain AF211 with and without bile are 6.6 hours and 23.6 hours (17 hours difference) while those of strain SRB521^T are 11.7 hours and 32.1 hours (20.4 hours difference), respectively. Values are mean of duplicates. Error bars are standard deviations.

Fig. S2. Genomic islands and spacer comparison. A. Positions of Genomic islands in both AF211 and SRB521^T as predicted by SeqWord Sniffer. B. Spacer comparison.

Fig. S3. Prophage regions detected in *I. butyriciproducens* strain AF211 and SRB521^T genome sequences.

Fig. S4. CAZyme families detected in AF211 and SRB521^T

Fig. S5. Lysine fermentation by strain AF211 and SRB521^T. Product formation was determined after 2 days of growth in bicarbonate buffered medium (mM). Orange and green boxes show the results from duplicate experiments of strain AF211 and SRB521^T, respectively. 15 mM L-lysine was added as substrate.

Fig. S6. Annotated transport systems in the genomes of AF211 and SRB521^T

Fig. S7. Pseudovitamin B12 production by *Intestinimonas butyriciproducens* AF211. UHPLC-UV (361 nm) chromatogram (A) of the immunoaffinity purified cell extract of AF211 and the mass spectrum (B) of the pseudovitamin B12 peak eluting at 3.16 min on performing MS/MS of the protonated molecular ions in a quadrupole time-of-flight (Q-TOF) mass spectrometer.