Comparative genomics and proteomics of Eubacterium maltosivorans : functional identification of trimethylamine methyltransferases and bacterial microcompartments in a human intestinal bacterium with a versatile lifestyle

Feng, Yuan

2022-01

Feng , Y , Thi Phuong Nam Bui , Stams , A J M , Boeren , S , Sanchez-Andrea , I & de Vos , W M 2022 , ' Comparative genomics and proteomics of Eubacterium maltosivorans : functional identification of trimethylamine methyltransferases and bacterial microcompartments in a human intestinal bacterium with a versatile lifestyle ' , Environmental Microbiology , vol. 24 , no. 1 , pp. 517-534 . https://doi.org/10.1111/1462-2920.15886

http://hdl.handle.net/10138/341328
https://doi.org/10.1111/1462-2920.15886

environmental microbiology

Environmental Microbiology (2022) **24**(1), 517–534 doi:10.1111/1462-2920.15886

Comparative genomics and proteomics of *Eubacterium maltosivorans*: functional identification of trimethylamine methyltransferases and bacterial microcompartments in a human intestinal bacterium with a versatile lifestyle

Yuan Feng,[1] Thi Phuong Nam Bui,[1,2] Alfons J. M. Stams,[1,3] Sjef Boeren,[4] Irene Sánchez-Andrea [1] and Willem M. de Vos [1,5*]

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

[2]*Caelus Pharmaceuticals, Amsterdam, The Netherlands.*

[3]*Centre of Biological Engineering, IBB – Institute for Biotechnology and Bioengineering, University of Minho, Campus de Gualtar, Braga, 4710-057, Portugal.*

[4]*Laboratory of Biochemistry, Wageningen University and Research, Stippeneng 4, Wageningen, 6708 WE, The Netherlands.*

[5]*Human Microbiome Research Program, Faculty of Medicine, University of Helsinki, Helsinki, 00014, Finland.*

Summary

***Eubacterium maltosivorans* YI^T is a human intestinal isolate capable of acetogenic, propionogenic and butyrogenic growth. Its 4.3-Mb genome sequence contains coding sequences for 4227 proteins, including 41 different methyltransferases. Comparative proteomics of strain YI^T showed the Wood–Ljungdahl pathway proteins to be actively produced during homoacetogenic growth on H_2 and CO_2 while butyrogenic growth on a mixture of lactate and acetate significantly upregulated the production of proteins encoded by the recently identified *lctABCDEF* cluster and accessory proteins. Growth on H_2 and CO_2 unexpectedly induced the production of two related trimethylamine methyltransferases. Moreover, a set of 16 different trimethylamine methyltransferases together with proteins for bacterial microcompartments were produced during growth and deamination of the quaternary amines, betaine, carnitine and choline. Growth of strain YI^T on 1,2-propanediol generated propionate with propanol and induced the formation of bacterial microcompartments that were also prominently visible in betaine-grown cells. The present study demonstrates that *E. maltosivorans* is highly versatile in converting low-energy fermentation end-products in the human gut into butyrate and propionate whilst being capable of preventing the formation of the undesired trimethylamine by converting betaine and other quaternary amines in bacterial microcompartments into acetate and butyrate.**

Introduction

The human intestinal tract is a natural reservoir of dynamic and diverse microbial communities with important developmental, nutritional and protective functions that are tightly linked to health and disease. The vast majority of the intestinal species are anaerobic bacteria which salvage nutrients and energy by fermenting complex carbohydrates into short-chain fatty acids (SCFAs), other organic acids such as formate and lactate, and gases, including H_2 and CO_2 (Rajilić-Stojanović and de Vos, 2014; Louis and Flint, 2017).

Lactate is another intermediate of carbon metabolism, produced by lactic acid and other abundant bacteria in the human intestinal tract (Smith and Macfarlane, 1997; Barcenilla *et al*., 2000). Lactate accumulation could lead to detrimental consequences, such as acidosis and neurotoxicity whilst increased colonic levels are associated with colitis in patients with inflammatory bowel disease (Hove *et al*., 1994; Hove and Mortensen, 1995; Mortensen and Clausen, 1996; Kowlgi and Chhabra, 2015). Therefore, both isomers of lactate must be efficiently depleted by intestinal microbes. Lactate can be utilized by sulfate- or sulfite-reducing bacteria like *Desulfovibrio*,

Received 28 September, 2021; revised 1 December, 2021; accepted 21 December, 2021. *For correspondence.E-mail willem.devos@wur.nl; Tel. +31 317 483108.

Bilophila and *Clostridium* sp. resulting in elevated hydrogen sulfide levels, which could potentially cause colonic pain and gastrointestinal discomfort in both infants and adults (Chassard *et al*., 2012; Feng *et al*., 2017; Pham *et al*., 2017). Lactate can also be consumed by butyrate-producing bacteria in presence of acetate as isotope labelling studies have identified lactate and acetate as important precursors of butyrate production in mixed gut communities (Bourriaud *et al*., 2005; Muñoz-Tamayo *et al*., 2011). Intestinal bacteria that can perform this low energy metabolic conversion mostly belong to *Clostridium* cluster XIVa, including *Anaerostipes* spp. and *Anaerobutyricum* spp. (previously known as *Eubacterium hallii*) (Duncan *et al*., 2004; Shetty *et al*., 2020). The pathway of butyrate production from lactate and acetate has been recently elucidated in *Anaerobutyricum* and *Anaerostipes* spp. (Shetty *et al*., 2020).

Acetate, propionate and butyrate – the most abundant SCFAs – can fuel the intestinal epithelium cells and confer health benefits via a variety of pathways (Tremaroli and Bäckhed, 2012; Fan and Pedersen, 2021). Moreover, all intestinal SCFAs are known to signal to G-protein coupled receptors, such as free fatty acid receptor 2 (FFAR2) also known as G-protein coupled receptor 43, and have specific health-promoting effects, among others via the production of anorectic gut hormones (Deleu *et al*., 2021). Notably, butyrate-producing bacteria have gained increasing attention as butyrate plays an important role in maintaining gut homeostasis and intestinal epithelial integrity (Hamer *et al*., 2008). Butyrate is also the 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 (Qin *et al*., 2010; Flint *et al*., 2012; De Vos and Nieuwdorp, 2013; Rajilić-Stojanović *et al*., 2013).

Propionate has high affinity for FFAR2 of all SCFAs and its role in promoting health effects as well as increasing satiety has been described (Chambers *et al*., 2015). Propionate can be produced from sugars via various pathways, but several intestinal species can produce this C3 compound from 1,2-propanediol (1,2-PDO), generated from methylpentoses such as fucose or rhamnose (Louis and Flint, 2017). The conversion of 1,2-PDO proceeds through a pathway that generates toxic intermediates that are encapsulated in self-assembling organelles termed bacterial microcompartments (BMCs) that are separated from the cytosol (Kerfeld *et al*., 2018). These BMCs are typically produced when needed and are thought to include a respiratory electron acceptor and a vitamin B12 cofactor (Cheng *et al*., 2011; Thompson *et al*., 2014).

Besides carbohydrate fermentation in the human intestinal tract the metabolism of many other compounds present in our diet is highly relevant since some may generate undesired products (Fan and Pedersen, 2021). These include quaternary amines such as betaine, choline and L-carnitine, which are commonly found in vegetables, fruits, meat and seafood and are well-known bacterial osmoprotectants (Sleator and Hill, 2002; Zeisel *et al*., 2003; Demarquoy *et al*., 2004). Several gut bacteria can convert these quaternary ammonium ions into trimethylamine (TMA), which is converted in the liver into trimethylamine-N-oxide (TMAO) that has been implied in atherosclerosis and cardiovascular risks (Brown and Hazen, 2014). Hence there is considerable interest in understanding the metabolism of these quaternary ammonium ions into compounds that do not lead to TMA or other TMAO precursors.

Recently, we isolated and described a novel human intestinal species *Eubacterium maltosivorans* YI^T affiliated with the *Clostridium* cluster II, which is capable of converting H_2 and CO_2, as well as methanol and formate with weak growth, into acetate (Feng *et al*., 2018). Besides homoacetogenic growth, it was able to ferment a range of monomeric sugars and disaccharides, including maltose, with acetate and butyrate as the main end products. Moreover, *E. maltosivorans* can also convert lactate and acetate into butyrate and 1,2-PDO into propionate. These properties are unique for the gut ecosystem and link three important metabolic processes, homoacetogenesis, propionigenesis and butyrogenesis in a single human species. Some of these features were predicted by genomic characterization of the phylogenetically closely related *Eubacterium limosum* strain ATCC 8486 isolated from human colon (Song *et al*., 2017), strain KIST612 isolated from an anaerobic digester (Roh *et al*., 2011), and strain SA11 obtained from sheep's rumen (Kelly *et al*., 2016), the latter being considered to belong to the new species of *E. maltosivorans* (Feng *et al*., 2018). *Eubacterium limosum* and *E. maltosivorans* strains are capable of using a variety of C1 compounds, such as methanol, CO, as well as H_2 with CO_2, and their demethylating activity is high (Genthner *et al*., 1981; DeWeerd *et al*., 1988). These activities in the intestinal tract have been involved in co-degradation of dietary components such as pectin that generates methanol (Rode *et al*., 1981). Moreover, dietary lignans and (iso) flavonoids that produce methoxylated aromatics can be converted by *E. limosum* strains into enterolactone and a range of other beneficial products (Jin *et al*., 2007; Possemiers *et al*., 2008; Struijs *et al*., 2009). Similarly, vitamin B12 corrinoid-dependent TMA methyltransferases were recently identified in *E. limosum* and found to catalyze the demethylation of carnitine and betaine, thereby limiting the TMA formation and TMAO accumulation

(Koeth *et al*., 2013; Picking *et al*., 2019; Kountz *et al*., 2020). Based on their potential health-promoting properties, *E. limosum*-related bacteria have been suggested to be among the next-generation beneficial microbes (Kanauchi *et al*., 2006), and bacteria related to *E. limosum* have been associated with longevity and health in a cohort of centenarians (Biagi *et al*., 2010).

In the present study, we describe the complete genome sequence of the aforementioned *E. maltosivorans* YI^T, provide a detailed comparative genomic analysis and use a proteome approach to identify the enzymes and pathways involved in acetogenesis (from H_2 and CO_2), butyrogenesis (from lactate and acetate and various sugars) and propionogenesis (from 1,2-PDO) as well as the conversion of quaternary amines (betaine, carnitine and choline), which is associated with the induction of unique TMA methyltransferases and proteins involved in the production of BMCs.

Results and discussion

Complete genome of E. maltosivorans *YI^T and comparative genomics of prophages, CRISPR systems and relevant features*

Eubacterium maltosivorans YI^T affiliates the *Eubacteriaceae* family of *Clostridium* cluster II, sharing the highest 16S rRNA genes similarity to *E. limosum* SA11 (99.7%), *E. limosum* KIST612 (98.1%), *E. callanderi* AMC0717 (97.5%) and *E. limosum* ATCC 8486^T (97.3%), the latter being the type species of the *Eubacterium* genus (Wade, 2006) (Fig. 1). The complete genome of *E. maltosivorans* YI^T consisted of a single circular chromosome of 4 337 501 bp with its general features being listed in Table 1. Many genomic features were shared with the related *E. limosum* strains ATCC 8486^T, KIST612 and SA11, for which complete genomes have been reported (Roh *et al*., 2011; Kelly *et al*., 2016; Song *et al*., 2017).

The genome of *E. maltosivorans* YI^T revealed five prophages sequences, of which four are intact and one is incomplete (Fig. 2). The prophages in strain YI^T had similar G + C content: 47.70% (33.3 kb), 41.61% (38.1 kb), 44.95% (53 kb) and 48.59% (44.8 kb) for intact prophage one to four respectively; and 46.47% (22.9 kb) for the incomplete prophage (Fig. S1A). These prophages were unique as no sequences were shared with the single 58.8-kb prophage sequence with a reduced G + C content (40.86%) in strain SA11 that is closest related to strain YI^T (Fig. S1B), indicating exposure to different prophages in human colon and in sheep rumen. In terms of prophage defence systems, both strains possessed CRISPR/Cas system in their genomes. Strain YI^T has one cluster of a 225 bp CRISPR array that contained a 30 bp long direct repeat (5-GTTGAAGCTTAACATGAGATGTATTTAAAT-3) with three spacers. Strain SA11 has a 2328 bp CRISPR system with the same direct repeat with 35 spacers. No spacers were found to be shared between these two strains, indicating previous viral infections were different. Besides, strain SA11 has another 1675 bp CRISPR array, which contained another 30 bp long repeat sequence (5-GTTGAAGTTTAACATGAGATGTATTTAAAT-3) with 25 spacers.

The genome of *E. maltosivorans* YI^T (hereafter *E. maltosivorans*) is predicted to contain 41 *mttB* genes encoding TMA methyltransferases (MttB, COG05598), all of which are predicted to encode proteins in similar size of 458–487 amino acids with one exception of 54 amino acids. These *mttB* genes in *E. maltosivorans* clustered at four major locations in the genome, while in *E. limosum* strains they are more scattered (Fig. 2). Similar to *E. limosum* strains, the *mttB* genes in *E. maltosivorans* are also often associated with genes for cobalamin B12-binding proteins, BCCT (betaine/carnitine/choline transporter) and MFS family transporters and GntR family transcriptional regulators (Kelly *et al*., 2016; Picking *et al*., 2019; Kountz *et al*., 2020).

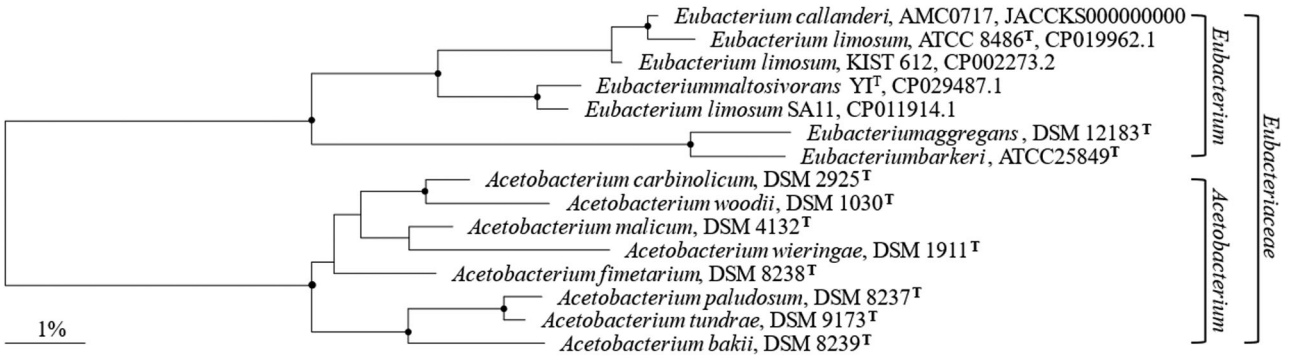

Fig. 1. Phylogenetic affiliation of *E. maltosivorans* YI^T and related *E. limosum* strains. Neighbour-joining tree based on 16S rRNA gene sequence similarities calculated with Jukes–Cantor correction was adapted from Feng *et al*. (2018), with accession numbers of the deposited complete genomes of *E. maltosivorans* YI^T and related *E. limosum* and *E. callanderi* strains.

Table 1. General genome features of *E. maltosivorans* YI^T, *E. limosum* ATCC 8486^T, SA11, KIST612 and *E. callanderi* AMC0717.

Strain	YI CP029487.1	SA11 CP011914.1	ATCC 8486^T CP019962.1	KIST612 CP002273.2	AMC0717 JACCKS000000000
Host	Human	Sheep	Human	Anaerobic digester	Human colonic Mucosa
Genome size (bp)	4 337 501	4 150 332	4 422 837	4 316 707	4 327 770
GC content (%)	47.8	47.4	47.2	47.5	47.5
RNA	77	76	77	73	64
rRNA	16	16	17	16	11
tRNA	61	60	60	57	53
Prophage	5	1	4	4	1
No. subsystems	351	347	354	358	245
No. CDs	4227	4044	4303	4284	4539
No. *mttB* genes	41	39	42	33	31

mttB: TMA methyltransferases.

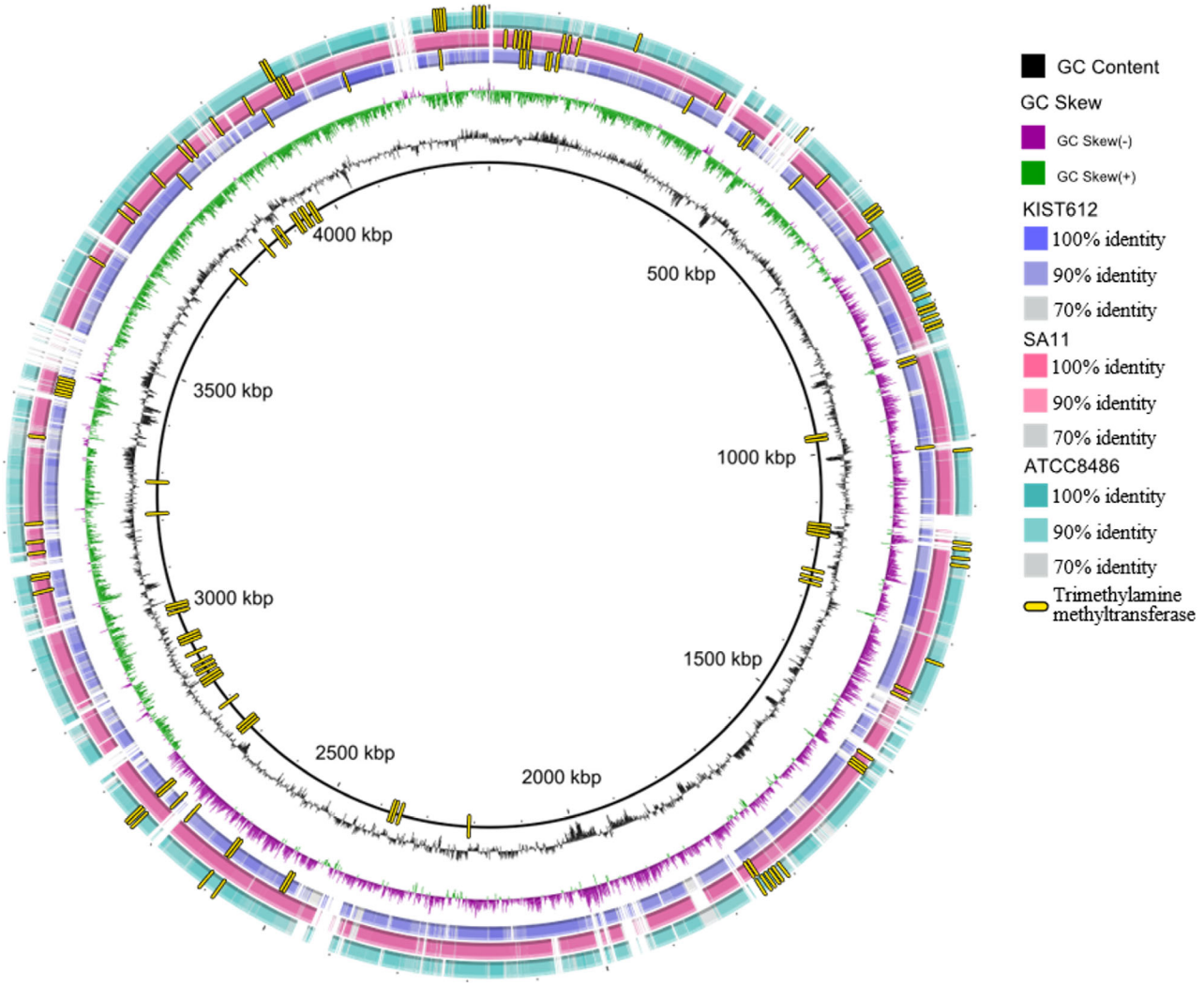

Fig. 2. Genome map of *E. maltosivorans* YI^T compared to the close-related strain KIST612 (purple), SA11 (pink) and ATCC8486T (green), with trimethylamine methyltransferases indicated in each strain (yellow).

Table 2. Substrate utilization and product formation by *E. maltosivorans* at the end exponential phase.

		Production (mM)					
	Consumption substrate (mM)	Acetate	Butyrate	Lactate	Formate	Propionate	Propanol
Glucose	11.6 ± 0.49	13.5 ± 1.16	0.6 ± 0.57	5.4 ± 1.10	5.1 ± 0.15	<0.1	<0.1
Lactate	9.3 ± 0.95	4.7 ± 0.77	3.8 ± 1.0	–	<0.1	–	–
H_2 (CO_2)	45.5 ± 0.85	6.5 ± 0.58	<0.1	<0.1	<0.1	–	–
1,2-PDO	17.5 ± 0.98	<0.1	<0.1	<0.1	<0.1	7.5 ± 0.63	6.6 ± 0.30

Values are average of triplicates and their standard deviation. 1,2-PDO: 1,2-Propanediol.

Comparative proteomics of E. maltosivorans *grown on different carbon sources*

Since *E. maltosivorans* is highly versatile and capable of growing on a variety of different substrates (Feng *et al*., 2018), we focused first on growth with H_2/CO_2 (further referred to as homoacetogenic growth) and lactate supplemented with acetate (further referred to as butyrogenic growth). Next, we tested the growth of *E. maltosivorans* with 1,2-PDO (further referred to as propionogenic growth), as well as demethylation growth of quaternary amines including betaine, carnitine and choline. For this purpose, actively growing *E. maltosivorans* cells pre-cultured in glucose were washed and inoculated in basal medium containing either H_2/CO_2, 20 mM lactate plus 15 mM acetate, 20 mM glucose, or 20 mM 1,2-PDO. When *E. maltosivorans* was grown with 20 mM quaternary amines, an extra 2.5 g L^{-1} PEP was supplied in basal medium. The carbon recovery of homoacetogenic, butyrogenic and propionogenic growth were 57.1%, 88.2%, and 80.5% respectively, while that on its preferred substrate glucose was 72.8% (Table 2). The missing carbon can be attributed to biomass production and intermediates that could not be detected by high-performance liquid chromatography at mid-log phase, such as amino acids.

In order to study proteome dynamics of *E. maltosivorans* during growth on different substrates, we conducted three sets of proteomic comparisons in eight growth conditions. The first set included the comparison of homoacetogenic and butyrogenic growth with glucose fermentation, while the second set compared propionogenic growth to maltose fermentation. To identify the potential role of MttB homologues in the butyrogenic conversion of quaternary amines, the third set of proteomes included *E. maltosivorans* cells grown on betaine, carnitine, choline and glucose. From a total of 4227 protein-coding sequences identified in the genome of *E. maltosivorans*, a collective of 1737 non-redundant proteins were identified in the aforementioned eight growth conditions. We observed the canonical proteins involved in the Wood-Ljungdahl Pathway (WLP), the route from acetyl-CoA to butyrate and acetate, as well as acetyl-CoA synthase to be abundantly present in all the proteomes with very low fold-change values within them, suggesting that their corresponding genes are constitutively expressed. Moreover, we identified unique proteome dynamics associated to the growth with different carbon sources as discussed in detail below. Based on the metabolic profiles of the tested substrates, the genomic reconstruction and the results of the comparative proteogenomics experiments that are presented below, we generated a schematic overview of the (homo)acetogenic, propionogenic and butyrogenic conversions that will be discussed further (Fig. 3).

Proteomics-guided identification of the active DL-lactate utilization gene cluster and butyrogenic pathway in E. maltosivorans

Lactate is one of the most important metabolites for the growth of anaerobic bacteria in the human intestinal environment (Duncan *et al*., 2004). It must be efficiently reused to avoid potential negative consequences, such as acidosis, neurotoxicity and cardiac arrhythmia.

During butyrogenic growth on lactate supplied with acetate, various proteins showed highly increased abundance (over 1000-fold compared to homoacetogenic growth and glucose fermentation), including lactate dehydrogenase (Ldh, encoded by CPZ25_RS02230), electron transfer flavoprotein alpha subunit (EtfA, encoded by CPZ25_RS02225), ETF beta subunit (EtfB, encoded by CPZ25_RS02230) and lactate racemase (LctF1, encoded by CPZ25_RS18455) (Fig. 4A; Table S1). The lactate racemase LctF1 ensures both D- and L-lactate promoted the growth of *E. maltosivorans*, which is in line with the substrate utilization since both DL-lactate were provided. In the genome of *E. maltosivorans*, downstream of the *ldh* gene, there is a gene encoding a potential lactate permease (encoded by CPZ25_RS02235), the product of which was not detected in our current proteomic analysis, since it is a likely transmembrane protein. The organization of the DL-lactate utilization gene cluster of *lctABCDEFG* in *E. maltosivorans* is depicted in Fig. 4B.

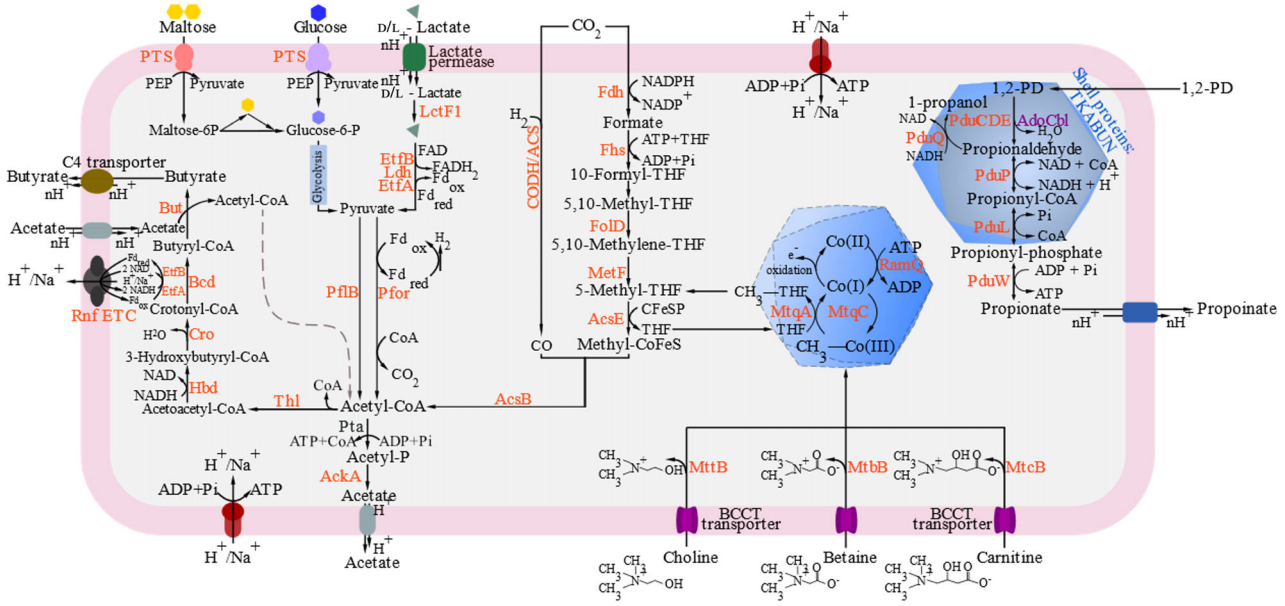

Fig. 3. Schematic overview of the key metabolic features of *E. maltosivorans* based on genomic reconstruction and supported by proteome data. Proteins marked in blue are that detected in the proteome, while those in black are only identified in the genome but not detected in the proteome at any of the growth conditions. SCFA transporters most likely belong to the C4 TRAP family of transporters. The symport is with a proton and *n* indicates the number of protons transported across the cell membrane which may vary under various growth conditions. PTS: phosphoenolpyruvate-dependent sugar phosphotransferase system; Ldh: lactate dehydrogenase; EtfA/EtfB: electron transfer flavoprotein alpha/beta subunit; LctF1: lactate racemase; Pfor: pyruvate:ferredoxin oxidoreductase; PflB: formate acetyltransferase; Pta: acetyl-CoA: phosphate acetyltransferase; Thl: thiolase; Hbd: 3-hydroxybutyryl-CoA dehydrogenase; Cro: crotonase; Bcd: butyryl-CoA dehydrogenase; But, butyryl-CoA: acetate CoA transferase; Fdh: formate dehydrogenase; Fhs: formyl-THF synthetase; FolD: methylene-THF dehydrogenase; MetF: methylene-THF reductase; AcsE: methyltransferase; CODH: carbon monoxide dehydrogenase; ACS: CO dehydrogenase; AscB: acetyl-CoA synthase; MttB: TMA methyltransferases; MtbB: betaine-specific MttB; MtcB: carnitine-specific MttB; MtqA: methylcorrinoid:THF methyltransferase; MtqC: corrinoid protein; RamQ: corrinoid activation enzyme; PduCDELPWQ: propanediol utilization proteins; AdoCbl: adenosylcobalamin.

Identification of the WLP in the proteome of E. maltosivorans *and induction of two highly similar TMA methyltransferases*

Hydrogen is a common gaseous microbial metabolite produced in the intestine through the fermentation of carbohydrates. CO_2 is also produced through a variety of decarboxylation reactions, notably from amino acid fermentations. Homoacetogenic bacteria can convert hydrogen together with CO_2 to acetate through WLP. Therefore, we sought to reconstruct the WLP in *E. maltosivorans* leading to the conversion of H_2/CO_2 to acetate.

The genes encoding proteins of the WLP in *E. maltosivorans* are mainly grouped in three clusters on the chromosome, consisting of a formate dehydrogenase (Fdh) cluster (CPZ25_RS17665-17675), a methyl (CPZ25_RS11820-11850) and a carbonyl branch (CPZ25_RS18530-18570) to form the central intermediate, acetyl-CoA (Supplementary Fig. 3). All of the predicted genes involved in the WLP were detected in the current proteome study, except for MoeA (molybdenum cofactor guanylyltransferase, encoded by CPZ25_RS17670) and a Fdh accessory protein (encoded by CPZ25_RS17675), indicating the activity of WLP in *E. maltosivorans*. However, most of the detected proteins were not higher produced in homoacetogenic growth, when compared with butyrogenic growth cells, suggesting constitutive expression of their genes (Tables S1 and S3). Remarkably, during growth on H_2/CO_2 two highly similar TMA methyltransferases (encoded by CPZ25_RS14065 and CPZ25_RS14915) were highly overproduced (over 150-fold change) compared to growth on glucose or lactate and acetate (Fig. 4A). These two TMA methyltransferases (considered as the MttB prototype) have an identical sequence and hence could not be distinguished from each other. Presently, it is not known whether these two TMA methyltransferases have a specific function in the homoacetogenic metabolism of *E. maltosivorans*.

Growth on maltose induces maltose-specific phosphotransferase and a set of internalin B-like proteins

Starch is a major source of energy in the human diet and is consumed in diverse forms. Resistant starch escapes digestion by glucoamylases in the upper GI tract and is degraded by the combined action of many gut bacteria in the colon, leading to beneficial impacts on colonic function and host health. Maltose is one of the by-products

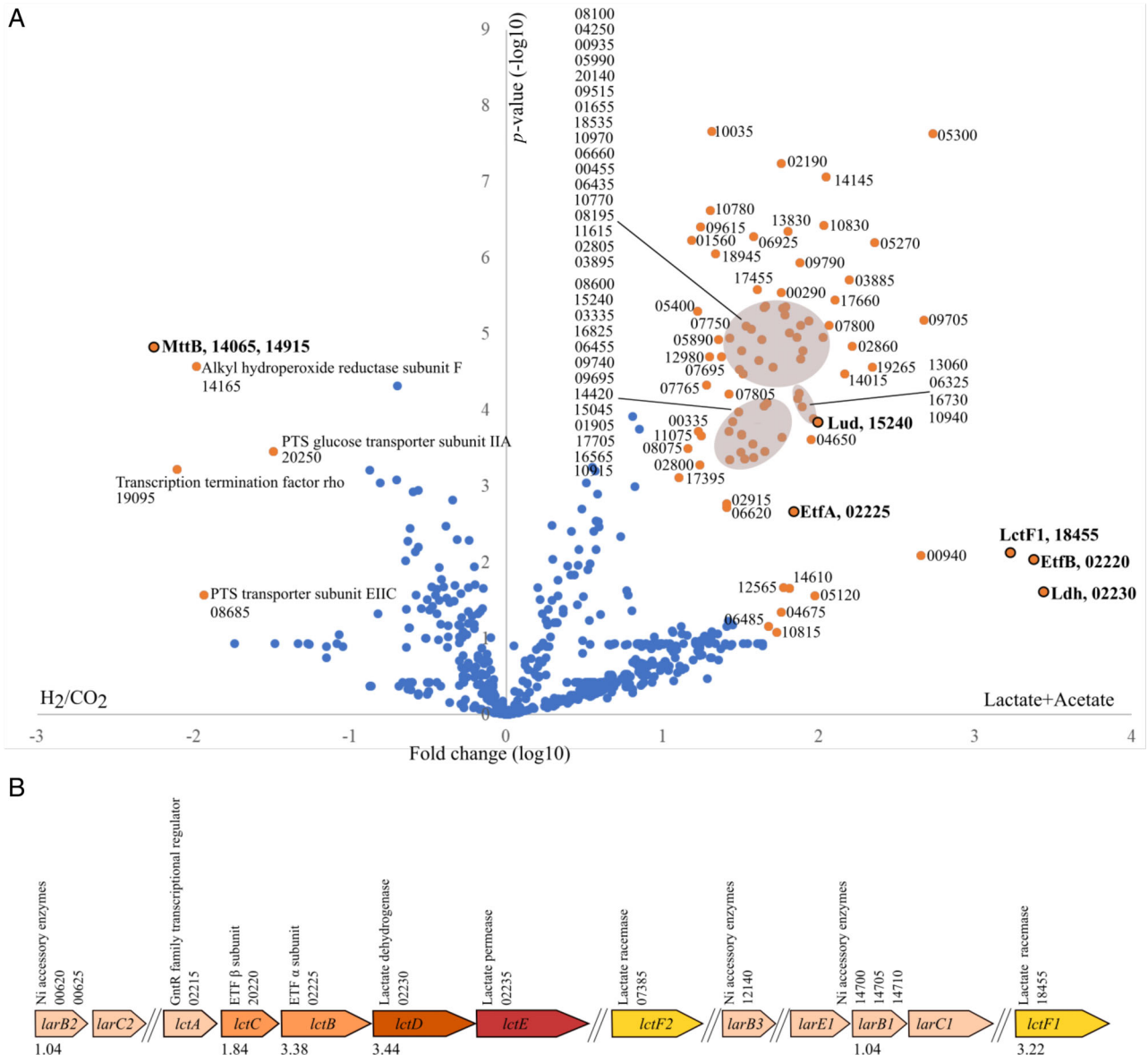

Fig. 4. Differentially abundant proteins in *E. maltosivorans* cells grown on lactate/acetate and hydrogen (A) and organization of gene clusters involved in lactate utilization in the genome of *E. maltosivorans* (B). Proteins that are significantly different in abundance between the two conditions in the proteome are coloured in orange, with the locus tags of the genes encoding for these proteins indicated. The locus tag for the genes encoded in *E. maltosivorans* starts with CPZ25_RS, which is omitted to avoid repetition in the figures. For clarity, proteins related to lactate utilization and the MttB are labelled in bold, and a complete list with product names for locus tags is given in Table S1. EtfA: electron transfer flavoprotein alpha; EtfB: electron transfer flavoprotein beta; Ldh: lactate dehydrogenase; LctF: lactate racemase; Lud: lactate utilizing domain; MttB: trimethylamine methyltransferase.

from the degradation of resistant starch that in the colon is mainly converted by *Ruminococcus bromii* (Ze *et al*., 2012). The ability to utilize maltose is one of the differentiating features for *E. maltosivorans* in respect to *E. limosum*. Therefore, we sought to identify the proteins involved in the metabolic pathway that converts maltose to acetate and butyrate by comparing the proteomes of *E. maltosivorans* grown on maltose and 1,2 PDO.

Genome analysis predicted that maltose uptake in *E. maltosivorans* depends on a maltose phosphoenolpyruvate phosphotransferase (Mal-PTS) system encoded by *malP* (CPZ25_RS08305). The phosphorylated maltose-6-phosphate is then degraded by a maltose-6-phosphate hydrolase (encoded by *malA*, CPZ25_RS08300) into glucose and glucose-6-phosphate, which both enter the glycolysis pathway and lead to production of acetate and butyrate. Lastly, there are two potential maltose operon regulators encoded by *malQ* and *malR* (CPZ25_RS08290 and CPZ25_RS08295 respectively) that may control the expression of the PTS genes and are located

in the proximity of the genes for the maltose utilization cluster (Fig. 5B).

The proteins encoded by the *malPQR* cluster of genes had significantly higher abundance in the proteome of cells grown on maltose compared to propionogenic growth (Fig. 5A). Among these proteins, the products of *malR* and *malP* were uniquely detected in the cells grown with maltose, indicating that the Mal-PTS is specifical for transporting maltose. Remarkably, we also noticed that the products of genes closely located around the maltose utilization cluster were coordinately expressed and highly induced in the presence of maltose (Fig. 5A). Among these genes, CPZ25_RS08280, CPZ25_RS08315, CPZ25_RS08325 and CPZ25_RS08330 encode large, over 1000 amino acid residue proteins. The protein products of CPZ25_RS08280, CPZ25_RS08325 and CPZ25_RS08330 share ∼40%–72% similarities to each other and their functions remain unknown although their genes were found to be conserved in all related *E. limosum* genomes. All proteins have signalP-predicted N-terminal signal sequences and those encoded by locus tags CPZ25_RS08315 and CPZ25_RS08320 have considerable similarity to internalin B, a well-studied *Listeria monocytogenes* surface protein involved in human cell binding (Parida *et al*., 1998). Moreover, these both contain remnants of sortase recognition sequences in their C-terminus. Of note, the *E. maltosivorans* genome is predicted to encode two class B sortases (CPZ25_RS01590 and CPZ25_RS01785), a protein that is known to covalently couple this kind of proteins to the cell envelope (Ilangovan *et al*., 2001; Marraffini *et al*., 2006). Whether this is also the case with these internalin B-like proteins remains to be confirmed since cell-envelope binding usually precludes detection in the cytoplasmic proteome as we report here.

Growth on 1,2-PDO induces proteins involved in BMCs

To further study the versatility of *E. maltosivorans*, we grew it on 1,2-PDO resulting in the production of equal amounts of propionate and 1-propanol (Fig. S4). Interestingly, we did not observe accumulation of intermediate propionaldehyde, which is known as a toxic compound to the cell. Therefore, we hypothesized that propionaldehyde could be sequestered within an organelle so-called a BMC which is a distinct mechanism to avoid lethal effects while growing on 1,2-PDO or other substrates with toxic intermediates (Chowdhury *et al*., 2014). This is in line with the detection of a total 27 *pdu* genes encoding BMC production and degradative enzymes in genome of *E. maltosivorans*, with 17 of which are closely located in a cluster. We found homologues of major (PduABB') and minor (PduKNTU) BMC shell proteins (Fig. 5D1). This confirms earlier predictions for the presence of a PDU1 type BMC encoded by multiple loci in the genome of *E. limosum* KIST612 (Axen *et al*., 2014).

Of note, an ethanolamine utilization (*eutJ*) homologue is located among the *pdu* cluster, which is a part of ethanolamine BMC and encodes a chaperone whose specific function is currently unknown (Kofoid *et al*., 1999). When the proteome of 1,2-PDO-grown cells was compared with those grown on maltose, all these BMC shell proteins were detected at a significantly higher abundance and PduABKNU, as well as EutJ, were significantly overproduced (Fig. 5A). The formation of BMCs during growth with 1,2-PDO was confirmed by analyzing cells grown on this substrate by transmission electron microscopy (TEM) and comparing these with cells grown on glucose (Fig. 5D).

Growth on quaternary amines is associated with induction of BMCs in E. maltosivorans

In our previous analysis we found that *E. maltosivorans* can grow on betaine (Feng *et al*., 2018). To further test other quaternary amines, we grew *E. maltosivorans* on betaine, carnitine, or choline (each 20 mM) on a medium with yeast extract and peptone (each 0.25%). Compared to the negative control (without substrate), *E. maltosivorans* produced 5 and 8 mM acetate when grown on carnitine and betaine respectively, and small amounts of butyrate productions (1–2 mM) were observed from both substrates (Fig. S5). However, choline did not generate these SCFA above background level but resting cells appeared to adapt to this substrate as evident from the determined proteome (see below).

Proteome analysis of *E. maltosivorans* grown on the quaternary amines and glucose detected 812 proteins with 420 as core proteins detected in all conditions. Importantly, 11 out of the 27 proteins encoded by the *pdu* cluster were detected to be higher expressed when *E. maltosivorans* was grown on quaternary amines than on glucose, especially the genes involved in cobalamin assimilation (encoded by *pduS*) and the formation of the BMC shells (encoded by *pduABN*, Table 3). Functional identification of BMCs was provided by TEM image analysis of *E. maltosivorans* cells grown on betaine that clearly revealed BMC structures that were similar in number and structure as those visible in cells grown on 1,2-PDO but absent in glucose-grown cells (Fig. 5D2). The involvement of BMCs in the degradation of quaternary amines such as choline has been earlier established for *Desulfovibrio alaskensis*, *Proteus mirabilis* and an uropathogenic *Escherichia coli* (Kuehl *et al*., 2014; Jameson *et al*., 2016; Herring *et al*., 2018). However, these bacteria all contain choline TMA lyase and its activating enzyme (CutCD) generates TMA and acetaldehyde. The

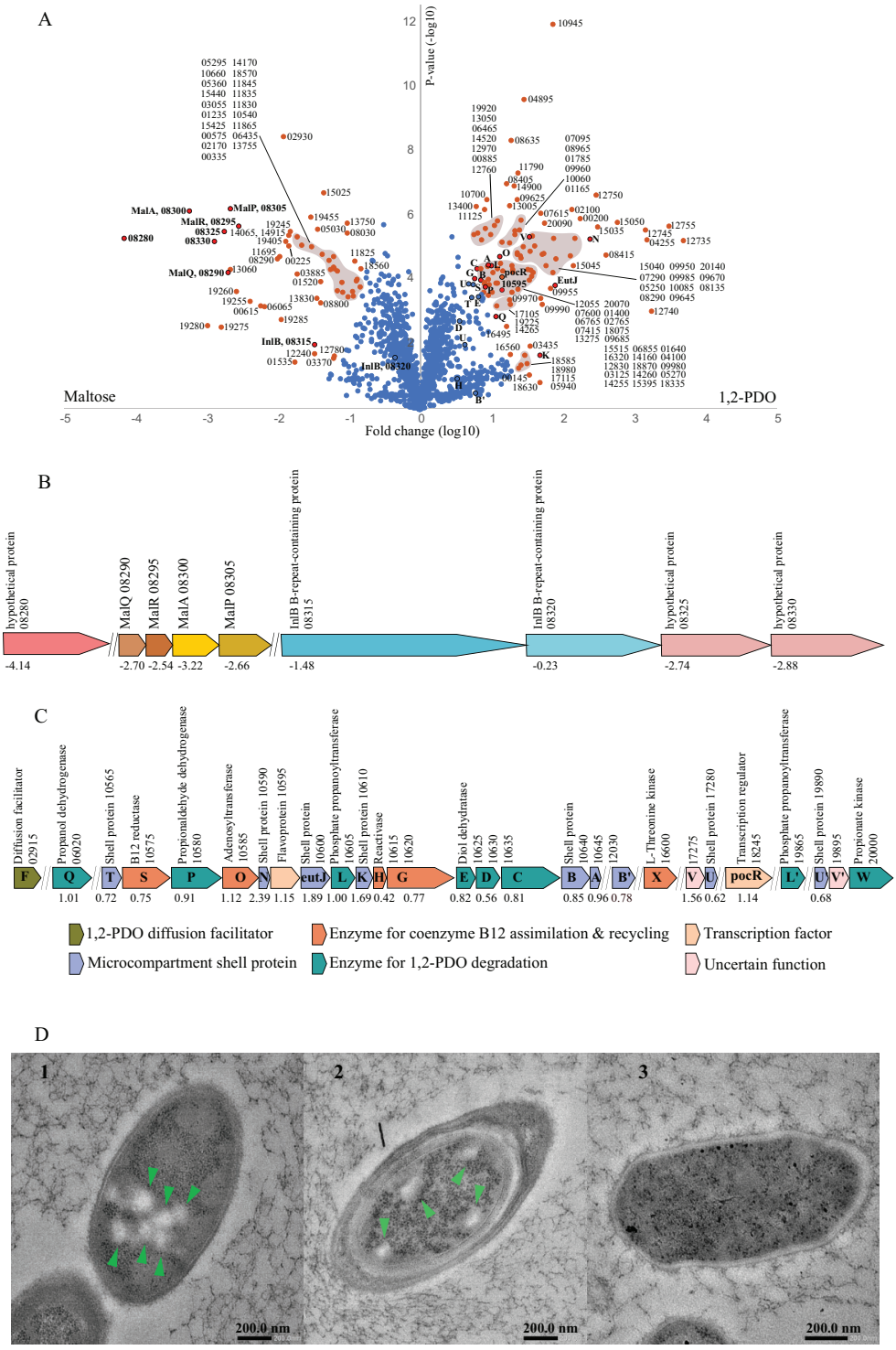

Fig. 5. Comparative analysis of proteomics (A), prediction of gene clusters (B & C) and presence of BMCs (D) in *E. maltosivorans* grown on 1,2-PDO and maltose.

A. Differentially abundant *E. maltosivorans* proteins in cells grown on 1,2-PDO versus on maltose. Proteins that are significantly different in abundance between the two conditions in the proteome are coloured in orange, with the locus tags of the genes encoding for these proteins indicated. The locus tag for the genes encoded in *E. maltosivorans* starts with CPZ25_RS, which is omitted to avoid repetition in the figures. For clarity, locus tags related with maltose utilization and microcompartments formation are labelled in bold, and a complete list with product names for locus tags is given in Table S4. B, C. Operon of the genes involved in maltose transportation and 1,2-PDO degradation in the genome of *E. maltosivorans* with colour coded predicted functions. D. TEM visualization of microcompartments in *E. maltosivorans* when grown with 1,2-PDO (1), betaine (2) and negative control glucose (3). Green arrowheads point to microcompartments structures, with the black scale bars.

Table 3. Proteins involved in bacterial microcompartments in *E. maltosivorans* were higher expressed when grown with betaine (BE), carnitine (CA) and choline (CH) than with glucose (Glu).

Locus tag	Protein	Fold change induction		
		BE vs. Glu	CA vs. Glu	CH vs. Glu
10565	Polyhedral organelle shell protein PduT	2.7	3.4	1.9
10575	Cobalamin reductase, PduS	14.8	13.5	10.7
10580	Propionaldehyde dehydrogenase, PduP	2.4	3.2	2.5
10590	Shell protein, PduN	27.5	1.0	1.0
10620	Glycerol dehydratase reactivase subunit alpha	1.0	1.8	1.5
10625	PD dehydratase small subunit PduE	4.7	12.3	6.6
10630	PD dehydratase medium subunit PduD	5.4	12.3	4.3
10635	PD dehydratase large subunit PduC	9.1	13.2	9.1
10640	Microcompartment protein PduB	12.6	18.2	13.2
10645	Microcompartment protein PduA	6.2	6.5	2.3
19865	Phosphate propanoyl transferase, PduL	1.8	2.3	1.1

PD: Propanediol. The locus tag for the genes encoded in *E. maltosivorans* starts with CPZ25_RS, which is omitted to avoid repetition in the table. Fold change induction was calculated by dividing the protein abundances in the different growth conditions.

BMCs are involved to sequester acetaldehyde to prevent toxicity, and subsequently metabolize acetaldehyde to ethanol and acetate (Herring *et al*., 2018). On the other hand, the resulting TMA is excreted in the gut lumen and cannot be used as a nitrogen or carbon source by most of the gut microbes. TMA then can enter the bloodstream, and once transported to the liver it is converted to TMAO predominantly by flavin monooxygenase 3. High serum levels of TMAO have been shown to promote formation of atherosclerotic plaques in a mouse model (Koay *et al*., 2021). Furthermore, serum levels of TMAO were significantly correlated with the incidence of heart attack and stroke (Wang *et al*., 2011). Moreover, it is a common uremic toxin (Velasquez *et al*., 2016). In contrast to the TMA-producing species, *E. maltosivorans* deaminates betaine and other quaternary amines in a process that likely involves a BMC. Quaternary amines are thought to be actively transported into the cell by BCCT transporter family proteins that also likely export the generated deaminated products (Ziegler *et al*., 2010). In addition, based on earlier discoveries in another strain (Müller *et al*., 1981), recent studies have elucidated the biochemical pathway of quaternary amines in *E. limosum* ATCC 8486^T into acetate and butyrate during demethylation of betaine, choline and carnitine into dimethylglycine, dimethylethanolamine and norcarnitine respectively (Müller *et al*., 1981; Picking *et al*., 2019; Kountz *et al*., 2020). In this way, one of three methyl groups from quaternary amines is transferred to MtqC (a corrinoid protein), together with a cob(I)alamin by MttB homologues, forming methyl-Co(III)-MtqC. The N,N-dimethyl compound is thus produced and released. The methyl group is subsequently transferred to tetrahydrofolate (THF) by a methylcorrinoid:THF methyltransferase (MtqA). The resulting methyl-THF enters the WLP, leading to cell biomass, and acetate/butyrate and CO_2 production. Co(I) MtqC can be adventitiously oxidized to catalytically inactive cobalt(II), which is reduced to active cobalt(I) by the ATP and reducing power-dependent activity of RamQ (a corrinoid activation enzyme) (Picking *et al*., 2019; Kountz *et al*., 2020). Based on genomic reconstructions a similar pathway is expected to operate in *E. maltosivorans* (Fig. 3). Moreover, the present proteomic and TEM data provide further support for the fact that BMCs play a role in this unique pathway, rationalize the observation that the WLP genes are rather constitutively expressed as needed for growth on quaternary amines, and explain the requirement for the high number of MttB homologues predicted from the genomes of *E. maltosivorans* and related species (Table 1; see below). In general, the BMCs are thought to protect the cell from toxic metabolic intermediates, prevent unwanted side reactions and may enhance flux through a multistep pathway (Kerfeld *et al*., 2018). In absence of further data, we hypothesize that the role of a BMC in the demethylation of quaternary amines by *E. maltosivorans* may be to encapsulate the vitamin B12-depending demethylation cascade and contain the appropriate TMA methyltransferase within the BMC.

Demethylation of quaternary amines is associated with induction of TMA methyltransferases in E. maltosivorans

Next, we focused on the expressed products of *mttB* genes in *E. maltosivorans* when induced by the growth with different quaternary amines and the other tested substrates (see Fig. 3). Among the total of 41 MttB homologues predicted by the genome of *E. maltosivorans*, 16 were detected to be present in the proteomes of cells grown on the quaternary amines. Of note, most of these detected MttB homologues were detected in the cells grown at all quaternary amines, except for protein

Table 4. Differential expression of 16 MttB homologues and other components of a corrinoid-dependent methyltransferase in *E. maltosivorans* when grown with betaine (BE), carnitine (CA), choline (CH) or glucose (Glu).

		Fold change induction		
Locus tag	Protein (potential function)	BE vs. Glu	CA vs. Glu	CH vs. Glu
04610	MttB protein	15.8	16.8	16.8
05865	Betaine-specific MttB protein	**114.5**	29.8	1.0
06255	MttB protein	33.9	6.6	0.7
12920	MttB protein	6.6	5.3	4.4
12935	MttB protein	60.3	19.1	1.0
13475	Betaine-specific MttB protein	**25.7**	1.0	1.0
13640	MttB protein	5.9	2.6	3.0
13755; 18250	Betaine-specific MttB protein	**1148.2**	3.2	17.2
14265	MttB protein	22.4	10.7	14.6
13805	Carnitine-specific MttB protein	1.0	**933.2**	1.0
14255	MttB protein	5.9	6.3	5.1
14915; 14065	Carnitine-specific MttB protein	3.1	**575.4**	9.6
17920	MttB protein	1.3	3.5	5.4
12890	Choline-specific MttB protein	1.0	1.0	**61.5**
11850	MtqA	5.6	2.2	0.9
13765	MFS transporter	70.8	1.0	1.0
14190	MtqC	0.5	0.04	0.0
18535	RamQ	4.2	1.8	0.9

MFS: Major facilitator superfamily protein; RamQ: corrinoid activation enzyme; MtqA: methylcorrinoid:tetrahydrofolate methyltransferase; MtqC: a corrinoid protein. The locus tag for the genes encoded in *E. maltosivorans* starts with CPZ25_RS, which is omitted to avoid repetition in the table. Fold change induction was calculated by dividing the protein abundances in the different growth conditions. Bold entries reflect putative TMA methyltransferases involved in the utilization of betaine, carnitine or choline.

products of CPZ25_RS12890 and CPZ25_RS13475, and CPZ25_RS13805, which was only detected when *E. maltosivorans* was grown on choline, betaine and carnitine respectively (Table 4).

A recent proteome study of *E. limosum* grown on proline betaine showed high expression of only one of the MttB homologues, in line with earlier finding that this demethylating enzyme systems seem to be very specific (Müller *et al*., 1981). The MttB enzymes catalyzed THF methylation of proline betaine specifically, forming the key intermediate methyl-THF in the WLP (Picking *et al*., 2019). Later, Kountz *et al*. (2020) also revealed a carnitine-specific MttB protein, namely, MtcB, which catalyzes specifically L-carnitine demethylation. We found the homologues of MtpB and MtcB in the genome of *E. maltosivorans*, with 98.3% and 77.0% amino acid similarities (encoded by CPZ25_RS13805 and CPZ25_RS13245 respectively) to those in *E. limosum*. In addition, we identified MtqA and MtqC homologues in the genome of *E. maltosivorans*, sharing high similarity (>96% protein identity) to those of *E. limosum*.

In apparent contrast with the results obtained for *E. limosum* (Picking *et al*., 2019; Kountz *et al*., 2020), the *E. maltosivorans* proteome showed multiple overproduced MttB proteins for each of the quaternary amines betaine and carnitine as well as choline, though we only analyzed resting cells for the latter (Table 4). For instance, MttB homologues encoded by CPZ25_RS13805 and 18250 were >300 times higher produced when grown on betaine than when grown on carnitine, while MttB homologue encoded by CPZ25_RS13805 was >900 times overproduced when grown with carnitine than grown with betaine. Remarkably, although choline did not seem to support the growth of *E. maltosivorans*, the proteomes of the cells were affected, as one MttB homologue (encoded by CPZ25_RS12890) was uniquely produced when cells were incubated in the presence of choline. Based on our present proteome data, we hypothesize that candidate TMA methyltransferases specific for betaine are CPZ25_RS05865, 13475, 18250 and 13755, specific for carnitine are CPZ25_RS13805, 14065 and 14915, and CPZ25_RS12890 specific for choline.

To fully picture the 41 MttB homologues encoded by the genome of *E. maltosivorans*, a Maximum-likelihood phylogenetic tree was constructed (Fig. 6). The amino acid sequence of these *E. maltosivorans* MttB proteins varied considerably, as when compared to the prototype MttB (CPZ25_RS14065 and 14915), these vary from 100% to approximately 27% in sequence identity. Through our current proteome and genome studies, we identified 16 active MttB homologues induced by growth or incubation with three different quaternary amines and proposed specific substrates for eight MttB homologues. The substrates for the other MttB homologues in *E. maltosivorans* remain unknown but we speculate that these may be involved in growth on other C1 compounds like methanol, CO or formate. Moreover, these may be involved in the ability to demethylate other substrates

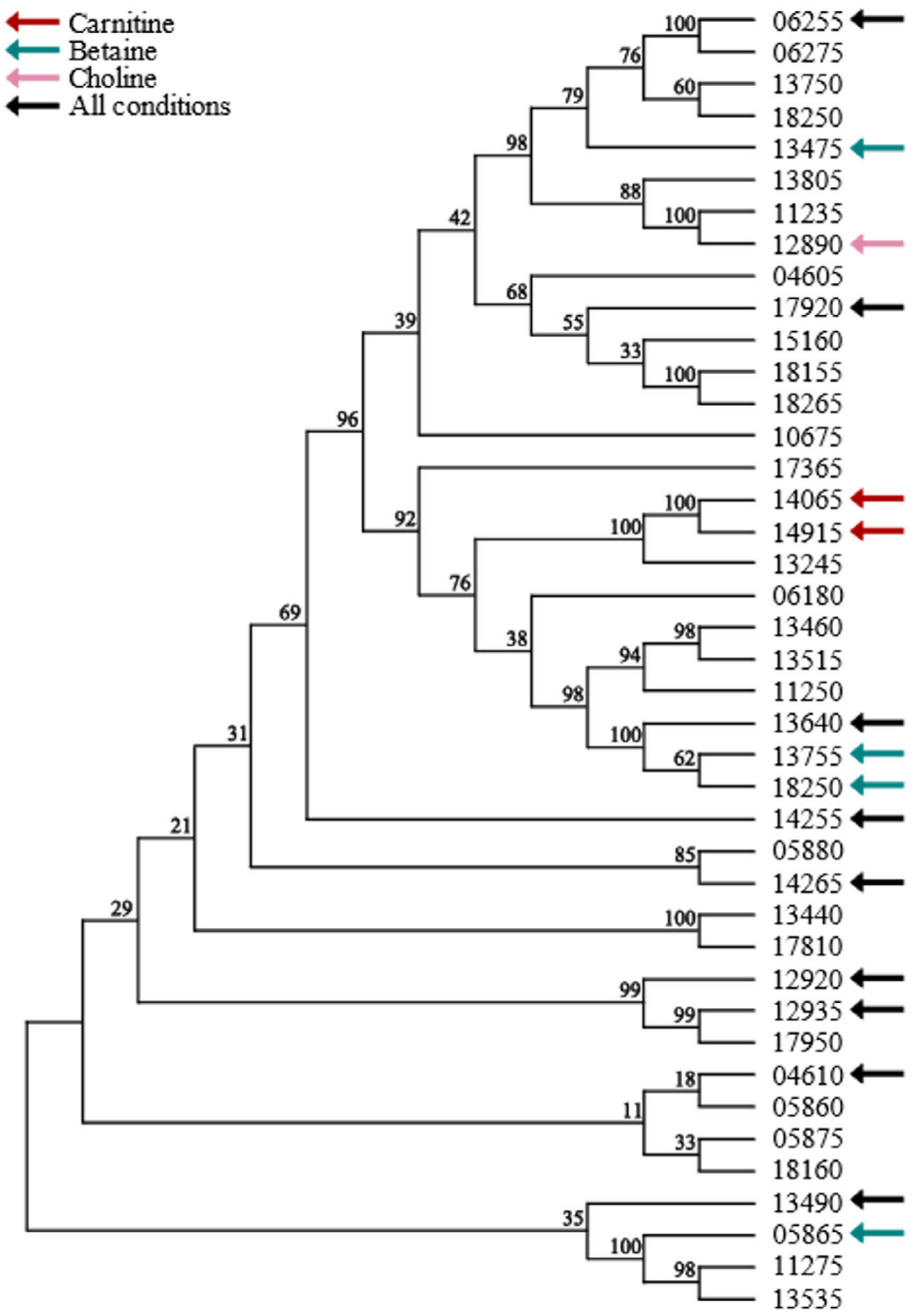

Fig. 6. Maximum likelihood phylogenetic tree of 41 trimethylamine methyltransferase (MttB) homologues in *E. maltosivorans*. Label represents locus tag of the gene product, which in *E. maltosivorans* starts with CPZ25_RS. To avoid repetition of the prefix in the figure, the locus tags are represented only by the specific identifier. The numbers on branches represent the bootstrap values from 1000 replicates. The proteins detected in the proteomes described here are marked with arrows and the growth conditions are indicated.

such as dietary lignans and (iso)flavonoids, which have been linked to possible health benefits and longevity (Jin *et al*., 2007; Possemiers *et al*., 2008; Struijs *et al*., 2009).

Conclusions

Our genomic and proteomic analyses revealed key metabolic features of *E. maltosivorans* (Fig. 2). *E. maltosivorans* has lactate utilization gene clusters, Ldh/Etf complex and butyryl-CoA dehydrogenase to convert low energy compounds D,L-lactate with acetate into butyrate. *E. maltosivorans* also catalyzes the conversion of 1,2-PDO to propionate and propanol and by proteome analysis and TEM, we confirmed that microcompartments were formed during the growth of *E. maltosivorans* with 1,2-PDO.

Genome analysis predicted *E. maltosivorans* to produce 41 trimethylamine methyltransferases (MttB). In our current proteome studies, 16 of these were detected when strain YI^T was grown with the quaternary amines betaine, carnitine and choline. In addition, two MttB proteins were produced when the cells were grown under

homoacetogenic conditions. Substrate-specific MttB proteins for betaine, carnitine and choline were proposed based on the highly abundant protein products of *E. maltosivorans* compared to that grown with glucose. The deamination of betaine and other quaternary amines might also involve a microcompartment, possibly due to the vitamin B12-dependent nature of these MttB homologues. By this pathway, TMA production thus can be reduced in the human gut ecosystem and consequently leading to a lower TMAO level in the blood serum. The present study demonstrates that *E. maltosivorans* is highly versatile in converting low-energy fermentation end-products in the human gut into health-promoting SCFA and may be a useful therapeutic strain to prevent the production of the undesired TMAO.

Experimental procedures

Growth medium and conditions

Cultivation of *E. maltosivorans* YI^T (DSM 105863^T) was handled as previously described (Feng *et al*., 2018). Briefly, it was grown based on anoxic basal medium (Stams *et al*., 1993) supplemented with 1 g L^{-1} yeast extract and 20 mM of each substrate separately. Eight different substrates were selected to conduct the comparative proteomics study for *E. maltosivorans*: homoacetogenic growth with H_2/CO_2 (80:20 vol./vol.), butyrogenic growth with 20 mM lactate with 15 mM acetate, fermentation growth with 20 mM glucose or 20 mM maltose, propionogenic growth with 20 mM 1,2-PDO, and demethylate growth on either 20 mM carnitine, betaine, or choline. Each condition was run in triplicates. For growth on carnitine, betaine and choline (20 mM each), the growth medium was amended with 2.5 g L^{-1} peptone and 2.5 g L^{-1} yeast extract. Each condition was run in duplicates. Except otherwise indicated, cultures were incubated statically at 37°C in the dark without shaking.

Genomic DNA preparation, assembly and annotation

Exponentially growing cells on maltose (50 ml) were harvested by centrifuging at 4°C for genomic DNA isolation. MasterPure Gram-positive DNA purification kit (Epicentre) was used to extract the genomic DNA of *E. maltosivorans* YI^T. The quality of the extracted DNA was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). The genome was sequenced using both PacBio RSII instrument and Illumina. Both Illumina paired-end reads and PacBio subreads were used as input for the assembler spades (v.3.11.1) with a custom kmer setting of 21,31,51,81,91 and coverage cutoff set to auto. After removing scaffolds smaller than 500 bp, one scaffold was obtained of 43 375 050 bp (0 N's). Illumina paired-end reads were mapped back to the assembly using bowtie2 (v2.3.3.1). The resulting sam files were converted with samtools (v1.3.1). The indexed bam file was used as input for Pilon (v1.22) which resulted in a 4 337 503 bp (N's) polished assembly. The complete genome sequence of *E. maltosivorans* YI^T (DSM 105863^T) was deposited at GenBank/EMBL-EBI under the accession number CP029487.1, and proteins were annotated by the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova *et al*., 2016; Haft *et al*., 2018; Li *et al*., 2021). The complete genomes of the three known *E. limosum* strains, including the type strain ATCC8486^T, strain KIST612 and strain SA11, were obtained from NCBI with accession numbers of CP019962.1, NC_014624.2 and CP011914.1 respectively.

Specific annotations

The core genome of the *E. maltosivorans* and *E. limosum strains* was identified using Spine and the pan-genome was identified using AGEnt (Ozer *et al*., 2014). CRISPR genes of the four strains were annotated using CRISPRfinder (Grissa *et al*., 2007). 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). Prophage sequences were detected using PHASTER (Arndt *et al*., 2016). In general, if a given phage region's total score is <70, it is recognized as incomplete; between 70 and 90, it is recognized as questionable; if >90, it is intact. Genomic islands were identified using online IslandViewer version 4 (Bertelli *et al*., 2017).

Proteomics

For proteomics analysis, *E. maltosivorans* cells cultivated with different substrates as described above were harvested at middle or late exponential growth phase by centrifugation for 20 min at 4700*g* at 4 °C (Supplementary Fig. 2). The pellets were washed with 1 ml PBS and immediately stored at −80°C until protein extraction. In total, we conducted three sets of proteomic analyses, with the first set comparing homoacetogenic and butyrogenic growth to glucose fermentation in biological triplicates, the second comparing propionogenic growth to maltose fermentation in biological triplicates and the last comparing growth with quaternary amines to glucose fermentation in biological duplicates.

To prepare the protein samples for the first set proteomes, an in-stage-tip sample preparation method was used (Kulak *et al*., 2014). The harvested washed pellets were thawed on ice and 2 mg wet weight was

transferred and resuspended in 0.5 ml of 100 mM Tris/HCl at pH 8. The suspension was sonicated six times at 30% amplitude in cycles of 30 s pulse and 30 s rest on ice by a Branson sonifier SFX150 equipped with a 3 mm tip (Branson, Carouge, CH). Proteins in the supernatant were collected by centrifugation at 10 000*g* for 10 min at 4°C and transferred to low binding Eppendorf microtubes. Disulphide bonds of the proteins were reduced by incubating in 20 mM dithiothreitol (DDT) in a thermomixer at 500 rpm at 45°C for 15 min. When the mixture cooled down to room temperature, AcrylAmide was added up to 25 mM to alkylate the reduced cysteines in the dark at room temperature for 30 min. Followed by centrifugation at 14 000*g* for 2 min, 70–80 µl protein extract (ca. 40 µg protein) was collected and loaded onto a C18 stage-tip using a double membrane (Empore C18, 3M), as described by Sánchez-Andrea *et al*. (2020). After a clean-up with 150 µl 5% acetonitrile (AcNi) in 50 mM ammonium bicarbonate buffer (ABC), stage tips were moved to clean 0.5 ml low binding microcentrifuge tubes and 500 ng of trypsin in 50 mM ABC was added on top to perform the enzymatic digestion of the proteins at room temperature overnight. The enzymatic digestion was stopped by adding 75 µl 0.1% formic acid and all liquid was eluted into the low binding tubes. In the end, hydrophobic peptides were eluted from the membrane manually with 5 µl 50% AcNi/50% 0.1% formic acid. Finally, after drying most of the AcNi and water in a concentrator, the final volume was made up to 50 µl with 0.1% formic acid. The peptide samples were analyzed as described by Sánchez-Andrea *et al*. (2020) using the same machine and setup. LCMS runs with all MSMS spectra obtained in first set proteome comparisons were analyzed with MaxQuant 1.6.0.1 as previous as well (Sánchez-Andrea *et al*., 2020).

For the second and the third set proteomes, around 10 mg harvested pellets were thawed, resuspended into 100 mM Tris pH 8 and sonicated using the same procedure as described for the first set. After sonication, for the protein determination, samples were centrifuged at 10 000*g* for 10 min at 4°C and the protein concentration was measured with Qubit™ Protein Assay Kit (Thermo Scientific). 40 µg protein of each (not centrifuged) sample was transferred to a 2 ml low binding tube. Proteins were reduced with 15 mM DDT at 45°C for 30 min, unfolded with 6 M urea and alkylated with 20 mM AcrylAmide at 21°C for 30 min. The pH was adjusted to 7 with 10% trifluoroacetic acid (TFA) and proteins were cleaned up by using a PAC method (Batth *et al*., 2019) using speedbeads and washing the proteins with 1 ml 70% ethanol and then 1 ml 100% acetonitrile. Digestion was conducted with 100 µl 0.5 pg μl^{-1} trypsin in 50 mM ABC at room temperature overnight. To stop the digestion, 10% TFA in H_2O was added to the samples until the pH dropped to 3. After filtering the peptide samples, volumes were adjusted to 50 µl with 0.1% formic acid.

5 µl of peptide samples were loaded directly onto a 0.10 × 250 mm ReproSil-Pur 120 C18-AQ 1.9 µm beads analytical column (prepared in-house) at a constant pressure of 825 bar (flow rate of circa 700 nl min^{-1}) with 1 ml L^{-1} HCOOH in water and eluted at a flow of 0.5 µl min^{-1} with a 50 min linear gradient from 9% to 34% acetonitrile in water with 1 ml L^{-1} formic acid with a Thermo EASY nanoLC1000. An electrospray potential of 3 kV was applied directly to the eluent via a stainless-steel needle fitted into the waste line of the micro cross that was connected between the micro cross and the analytical column. Full scan positive mode FTMS spectra were measured between *m/z* 380 and 1400 on an Exploris 480 at resolution 60 000. MS and MSMS AGC targets were set to 300% and 100% respectively or maximum ion injection times of 50 ms (MS) and 30 ms (MSMS) were used. HCD fragmented (Isolation width 1.2 *m/z*, 28% normalized collision energy) MSMS scans in a cycle time of 1.1 s were recorded for the most abundant 2–5+ charged peaks in data-dependent mode (Resolution 15 000, threshold 2e4, 15 s exclusion duration for the selected *m/z* ± 10 ppm).

LCMS runs with all MSMS spectra obtained were analysed with MaxQuant 1.6.3.4 (Cox and Mann, 2008). For all three sets of proteome analysis, an *E. maltosivorans* sequence database (NCBI accession number 72721) was used together with a contaminants database that contains sequences of common contaminants including Trypsins (P00760, bovin and P00761, porcin) and human keratins [Keratin K22E (P35908), Keratin K1C9 (P35527), Keratin K2C1 (P04264) and Keratin K1CI (P35527)]. Propionamide (C) set as a fixed modification, while variable modifications were set for Protein N-terminal acetylation and M oxidation which were completed by non-default settings for de-amidation of N and Q, the maximum number of modifications per peptide was 5. The label-free quantification (LFQ), as well as the match between runs options, was enabled. De-amidated peptides were allowed to be used for protein quantification and all other quantification settings were kept default. Filtering was done in MaxQuant by using a False Discovery Rate of 1% on protein and peptide level and on the MaxQuant result table by only accepting proteins with at least two identified peptides of which at least one should be unique and at least one should be unmodified. Reversed hits were deleted from the MaxQuant output. No extra filtering was applied after grouping the samples. We carefully checked the data and saw that protein groups with just one measured value all have log-fold changes below 0.6 and log *p*-values lower than 1. These are therefore never considered to be significantly different. The normal logarithm was taken from protein LFQ

MS1 intensities as obtained from MaxQuant. Zero 'Log LFQ' values were replaced by a value of 4.4 (A value slightly lower than the lowest measured value) to make sensible ratio calculations possible. Relative protein quantitation of the first and second set proteomes was done with Perseus by applying a two-sample *T*-test using the 'LFQ intensity' columns obtained with FDR set to 0.05 and S0 set to 1. Total non-normalized protein intensities corrected for the number of measurable tryptic peptides (intensity-based absolute quantitation) intensity were, after taking the normal logarithm, used for plotting on the *y*-axis in a protein ratio versus abundance plot. All three mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno *et al*., 2016) partner repository with the dataset identifiers PXD028586, PXD028574 and PXD028575.

Transmission electron microscopy

Biomass of *E. maltosivorans* YI^T grown with 1,2-PDO, betaine, or glucose was obtained by centrifugation (5 min, 4°C at 14 000*g*) from late exponential cultures. Biomass were resuspended and fixed in 1 ml 2.5% (vol./vol.) glutaraldehyde in 0.1 M sodium cacodylate buffer (pH 7.2) for 1 h at room temperature. After rinsing in the same buffer, biomass was resuspended and incubated in 100 μl of 3% gelatin in 0.1 M phosphate at room temperature for 20 min till gelatin solidified. The solidified specimens were incubated in 0.1 M sodium cacodylate buffer for 15 min at room temperature to become loose from tube wall. The specimens were infiltrated by ethanol and were then embedded in resin for 8 h at 70°C. Thin sections (<100 nm) of polymerized resin samples were obtained with microtomes. After staining with 2% (wt./vol.) aqueous uranyl acetate, the samples were analyzed with a Jeol 1400 plus TEM.

Acknowledgements

We thank Jelmer Vroom and Dr. Iame Alves Guedes for assistance with the TEM pictures, as well as Dr. Sudarshan Shetty and Dr. Bastian Hornung for useful suggestions on specific genome annotation. This work was supported by the SIAM Gravitation Grant 024.002.002 of the Netherlands Organization for Scientific Research.

References

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**: W16–W21.

Axen, S.D., Erbilgin, O., and Kerfeld, C.A. (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol* **10**: e1003898.

Barcenilla, A., Pryde, S.E., Martin, J.C., Duncan, S.H., Stewart, C.S., Henderson, C., and Flint, H.J. (2000) Phylogenetic relationships of butyrate-producing bacteria from the human gut. *Appl Environ Microbiol* **66**: 1654–1661.

Batth, T.S., Tollenaere, M.X., Rüther, P., Gonzalez-Franquesa, A., Prabhakar, B.S., Bekker-Jensen, S., *et al*. (2019) Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol Cell Proteomics* **18**: 1027–1035.

Bertelli, C., Laird, M.R., Williams, K.P., Simon Fraser University Research Computing Group, Lau, B.Y., Hoad, G., *et al*. (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* **45**: W30–W35.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., *et al*. (2010) Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One* **5**: e10667.

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

Bourriaud, C., Robins, R.J., Martin, L., Kozlowski, F., Tenailleau, E., Cherbut, C., and Michel, C. (2005) Lactate is mainly fermented to butyrate by human intestinal microfloras but inter-individual variation is evident. *J Appl Microbiol* **99**: 201–212.

Brown, J.M., and Hazen, S.L. (2014) Meta-organismal nutrient metabolism as a basis of cardiovascular disease. *Curr Opin Lipidol* **25**: 48.

Chambers, E.S., Viardot, A., Psichas, A., Morrison, D.J., Murphy, K.G., Zac-Varghese, S.E.K., *et al*. (2015) Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* **64**: 1744–1754.

Chassard, C., Dapoigny, M., Scott, K.P., Crouzet, L., Del'Homme, C., Marquet, P., *et al*. (2012) Functional dysbiosis within the gut microbiota of patients with constipated-irritable bowel syndrome. *Aliment Pharmacol Ther* **35**: 828–838.

Cheng, S., Sinha, S., Fan, C., Liu, Y., and Bobik, T.A. (2011) Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1, 2-propanediol degradation by Salmonella. *J Bacteriol* **193**: 1385.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O., and Bobik, T.A. (2014) Diverse bacterial microcompartment organelles. *Microbiol Mol Biol Rev* **78**: 438.

Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–1372.

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

Deleu, S., Machiels, K., Raes, J., Verbeke, K., and Vermeire, S. (2021) Short chain fatty acids and its producing organisms: an overlooked therapy for IBD? *EBioMedicine* **66**: 103293.

Demarquoy, J., Georges, B., Rigault, C., Royer, M.-C., Clairet, A., Soty, M., *et al*. (2004) Radioisotopic determination of L-carnitine content in foods commonly eaten in Western countries. *Food Chem* **86**: 137–142.

DeWeerd, K.A., Saxena, A., Nagle, D.P., Jr., and Suflita, J. M. (1988) Metabolism of the 18O-methoxy substituent of 3-methoxybenzoic acid and other unlabeled methoxybenzoic acids by anaerobic bacteria. *Appl Environ Microbiol* **54**: 1237.

Duncan, S.H., Louis, P., and Flint, H.J. (2004) Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* **70**: 5810–5817.

Fan, Y., and Pedersen, O. (2021) Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* **19**: 55–71.

Feng, Y., Stams, A.J.M., de Vos, W., and Sánchez-Andrea, I. (2017) Enrichment of sulfidogenic bacteria from the human intestinal tract. *FEMS Microbiol Lett* **364**. Available from: https://doi.org/10.1093/femsle/fnx028

Feng, Y., Stams, A.J.M., Sánchez-Andrea, I., and De Vos, W.M. (2018) *Eubacterium maltosivorans* sp. nov., a novel human intestinal acetogenic and butyrogenic bacterium with a versatile metabolism. *Int J Syst Evol Microbiol* **68**: 3546–3550.

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.

Genthner, B.R., Davis, C.L., and Bryant, M.P. (1981) Features of rumen and sewage sludge strains of Eubacterium limosum, a methanol- and H_2-CO_2-utilizing species. *Appl Environ Microbiol* **42**: 12–19.

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.

Haft, D.H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., *et al*. (2018) RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* **46**: D851–D860.

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.

Herring, T.I., Harris, T.N., Chowdhury, C., Mohanty, S.K., and Bobik, T.A. (2018) A bacterial microcompartment is used for choline fermentation by *Escherichia coli* 536. *J Bacteriol* **200**: e00764-17.

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.

Hove, H., and Mortensen, P.B. (1995) Colonic lactate metabolism and D-lactic acidosis. *Dig Dis Sci* **40**: 320–330.

Hove, H., Nordgaard-Andersen, I., and Mortensen, P.B. (1994) Effect of lactic acid bacteria on the intestinal production of lactate and short-chain fatty acids, and the absorption of lactose. *Am J Clin Nutr* **59**: 74–79.

Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R.T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus. *Proc Natl Acad Sci U S A* **98**: 6056–6061.

Jameson, E., Fu, T., Brown, I.R., Paszkiewicz, K., Purdy, K. J., Frank, S., and Chen, Y. (2016) Anaerobic choline metabolism in microcompartments promotes growth and swarming of P roteus mirabilis. *Environ Microbiol* **18**: 2886–2898.

Jin, J.-S., Zhao, Y.-F., Nakamura, N., Akao, T., Kakiuchi, N., and Hattori, M. (2007) Isolation and characterization of a human intestinal bacterium, *Eubacterium* sp. ARC-2, capable of demethylating arctigenin, in the essential metabolic process to enterolactone. *Biol Pharm Bull* **30**: 904–911.

Kanauchi, O., Fukuda, M., Matsumoto, Y., Ishii, S., Ozawa, T., Shimizu, M., *et al*. (2006) Eubacterium limosum ameliorates experimental colitis and metabolite of microbe attenuates colonic inflammatory action with increase of mucosal integrity. *World J Gastroenterol* **12**: 1071.

Kelly, W.J., Henderson, G., Pacheco, D.M., Li, D., Reilly, K., Naylor, G.E., *et al*. (2016) The complete genome sequence of *Eubacterium limosum* SA11, a metabolically versatile rumen acetogen. *Stand Genomic Sci* **11**: 26.

Kerfeld, C.A., Aussignargues, C., Zarzycki, J., Cai, F., and Sutter, M. (2018) Bacterial microcompartments. *Nat Rev Microbiol* **16**: 277–290.

Klampfer, L., Huang, J., Sasazuki, T., Shirasawa, S., and Augenlicht, L. (2003) Inhibition of interferon γ signaling by the short chain fatty acid Butyrate11Montefiore Medical Center New Research Initiative Award to LK and the American Cancer Society Institutional Research Grant to LK (ACS IRG# 98-274-01), UO1 CA88104 (to LA), and P3. *Mol Cancer Res* **1**: 855–862.

Koay, Y.C., Chen, Y.-C., Wali, J.A., Luk, A.W.S., Li, M., Doma, H., *et al*. (2021) Plasma levels of trimethylamine-N-oxide can be increased with 'healthy' and 'unhealthy' diets and do not correlate with the extent of atherosclerosis but with plaque instability. *Cardiovasc Res* **117**: 435–449.

Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T., *et al*. (2013) Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* **19**: 576–585.

Kofoid, E., Rappleye, C., Stojiljkovic, I., and Roth, J. (1999) The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J Bacteriol* **181**: 5317.

Kountz, D.J., Behrman, E.J., Zhang, L., and Krzycki, J.A. (2020) MtcB, a member of the MttB superfamily from the human gut acetogen *Eubacterium limosum*, is a cobalamin-dependent carnitine demethylase. *J Biol Chem* **295**: 11971–11981.

Kowlgi, N.G., and Chhabra, L. (2015) D-lactic acidosis: an underrecognized complication of short bowel syndrome. *Gastroenterol Res Pract* **2015**: 476215.

Kuehl, J.V., Price, M.N., Ray, J., Wetmore, K.M., Esquivel, Z., Kazakov, A.E., *et al*. (2014) Functional genomics with a comprehensive library of transposon mutants for the sulfate-reducing bacterium *Desulfovibrio alaskensis* G20. *MBio* **5**: e01041-14.

Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**: 319–324.

Li, W., O'Neill, K.R., Haft, D.H., DiCuccio, M., Chetvernin, V., Badretdin, A., *et al*. (2021) RefSeq: expanding the

prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* **49**: D1020–D1028.

Louis, P., and Flint, H.J. (2017) Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* **19**: 29–41.

Marraffini, L.A., Dedent, A.C., and Schneewind, O. (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* **70**: 192–221.

Mortensen, P.B., and Clausen, M.R. (1996) Short-chain fatty acids in the human colon: relation to gastrointestinal health and disease. *Scand J Gastroenterol* **31**: 132–148.

Müller, E., Fahlbusch, K., Walther, R., and Gottschalk, G. (1981) Formation of N, N-dimethylglycine, acetic acid, and butyric acid from betaine by Eubacterium limosum. *Appl Environ Microbiol* **42**: 439–445.

Muñoz-Tamayo, R., Laroche, B., Walter, É., Doré, J., Duncan, S.H., Flint, H.J., and Leclerc, M. (2011) Kinetic modelling of lactate utilization and butyrate production by key human colonic bacterial species. *FEMS Microbiol Ecol* **76**: 615–624.

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**: 1–17.

Parida, S.K., Domann, E., Rohde, M., Müller, S., Darji, A., Hain, T., *et al*. (1998) Internalin B is essential for adhesion and mediates the invasion of Listeria monocytogenes into human endothelial cells. *Mol Microbiol* **28**: 81–93.

Pham, V.T., Lacroix, C., Braegger, C.P., and Chassard, C. (2017) Lactate-utilizing community is associated with gut microbiota dysbiosis in colicky infants. *Sci Rep* **7**: 1–13.

Picking, J.W., Behrman, E.J., Zhang, L., and Krzycki, J.A. (2019) MtpB, a member of the MttB superfamily from the human intestinal acetogen *Eubacterium limosum*, catalyzes proline betaine demethylation. *J Biol Chem* **294**: 13697–13707.

Possemiers, S., Rabot, S., Espín, J.C., Bruneau, A., Philippe, C., González-Sarrías, A., *et al*. (2008) *Eubacterium limosum* activates isoxanthohumol from hops (*Humulus lupulus* L.) into the potent phytoestrogen 8-prenylnaringenin in vitro and in rat intestine. *J Nutr* **138**: 1310–1316.

Qin, J., Li, R., Raes, J., Burgdorf, K.S., Manichanh, C., Nielsen, T., *et al*. (2010) A human gut microbial gene catalogue established by metagenomic sequencing: commentary. *Nature* **11**: 28.

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.

Rajilić-Stojanović, M., Shanahan, F., Guarner, F., and de Vos, W.M. (2013) Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm Bowel Dis* **19**: 481–488.

Rode, L.M., Genthner, B.R.S., and Bryant, M.P. (1981) Syntrophic association by cocultures of the methanol-and CO_2-H_2-utilizing species *Eubacterium limosum* and pectin-fermenting *Lachnospira multiparus* during growth in a pectin medium. *Appl Environ Microbiol* **42**: 20.

Roh, H., Ko, H.-J., Kim, D., Choi, D.G., Park, S., Kim, S., *et al*. (2011) Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J Bacteriol* **193**: 307.

Sánchez-Andrea, I., Guedes, I.A., Hornung, B., Boeren, S., Lawson, C.E., Sousa, D.Z., *et al*. (2020) The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. *Nat Commun* **11**: 5090.

Shetty, S.A., Boeren, S., Bui, T.P.N., Smidt, H., and De Vos, W.M. (2020) Unravelling lactate-acetate and sugar conversion into butyrate by intestinal Anaerobutyricum and Anaerostipes species by comparative proteogenomics. *Environ Microbiol* **22**: 4863–4875.

Sleator, R.D., and Hill, C. (2002) Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev* **26**: 49–71.

Smith, E., and Macfarlane, G.T. (1997) Dissimilatory amino acid metabolism in human colonic bacteria. *Anaerobe* **3**: 327–337.

Song, Y., Shin, J., Jeong, Y., Jin, S., Lee, J.K., Kim, D.R., *et al*. (2017) Determination of the genome and primary transcriptome of syngas fermenting *Eubacterium limosum* ATCC. *Sci Rep* **7**: 13694.

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.

Struijs, K., Vincken, J., and Gruppen, H. (2009) Bacterial conversion of secoisolariciresinol and anhydrosecoisolariciresinol. *J Appl Microbiol* **107**: 308–317.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., *et al*. (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* **44**: 6614–6624.

Thompson, M.C., Crowley, C.S., Kopstein, J., Bobik, T.A., and Yeates, T.O. (2014) Structure of a bacterial microcompartment shell protein bound to a cobalamin cofactor. *Acta Crystallogr Sect F Struct Biol Commun* **70**: 1584–1590.

Tremaroli, V., and Bäckhed, F. (2012) Functional interactions between the gut microbiota and host metabolism. *Nature* **489**: 242–249.

Velasquez, M.T., Ramezani, A., Manal, A., and Raj, D.S. (2016) Trimethylamine N-oxide: the good, the bad and the unknown. *Toxins (Basel)* **8**: 326.

Vizcaíno, J.A., Csordas, A., Del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., *et al*. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**: D447–D456.

Wade, W.G. (2006) The genus *Eubacterium* and related genera. *Prokaryotes* **4**: 823–835.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., Dugar, B., *et al*. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**: 57–63.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J* **6**: 1535–1543.

Zeisel, S.H., Mar, M.-H., Howe, J.C., and Holden, J.M. (2003) Concentrations of choline-containing compounds and betaine in common foods. *J Nutr* **133**: 1302–1307.

Ziegler, C., Bremer, E., and Krämer, R. (2010) The BCCT family of carriers: from physiology to crystal structure. *Mol Microbiol* **78**: 13–34.

Supporting Information

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

Appendix S1: Supporting Information.