Lactate-based microbial chain elongation for n-caproate and *iso*-butyrate production: genomic and metabolic features of three novel *Clostridia* isolates

Bin Liu
Helmholtz-Zentrum fur Umweltforschung UFZ
Denny Popp
Helmholtz-Zentrum fur Umweltforschung UFZ
Heike Sträuber
Helmholtz-Zentrum fur Umweltforschung UFZ
Hauke Harms
Helmholtz-Zentrum fur Umweltforschung UFZ
Sabine Kleinsteuber (sabine.kleinsteuber@ufz.de)
Helmholtz-Zentrum fur Umweltforschung UFZ https://orcid.org/0000-0002-8643-340X

Research

Keywords: Carboxylate platform, Medium-chain carboxylates, Branched-chain carboxylates, Novel clostridial species, Comparative genomics

Posted Date: August 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-50186/v1

Version of Record: A version of this preprint was published at Microorganisms on December 11th, 2020. See the published version at https://doi.org/10.3390/microorganisms8121970.

Abstract

Background

The platform chemicals *n*-caproate and *iso*-butyrate can be produced by anaerobic fermentation from agro-industrial residues in a process known as microbial chain elongation. A few chain-elongating species have been discovered to utilize lactate and used to study the physiology of lactate-based chain elongation in pure cultures. Recently we isolated three novel clostridial species (strains BL-3, BL-4 and BL-6) that convert lactate to *n*-caproate and *iso*-butyrate. Here, we analyzed the genetic background of lactate-based chain elongation in these strains and other chain-elongating species by comparative genomics.

Results

All three strains produced *n*-caproate and *iso*-butyrate from lactate, with the highest proportions of *n*-caproate (18%) for BL-6 and *iso*-butyrate (23%) for BL-4 in batch cultivation at pH 5.5. The strains are suggested to represent three novel species based on low similarities with their closest described relatives. The three genomes show low conservation of organization and a relatively small core-genome size (504 out of 6,654 gene families). Including data of another eleven experimentally validated chain-elongating strains, we found that the chain elongation-specific core-genome harbors genes involved in reverse β-oxidation, hydrogen formation and energy conservation, displaying substantial genome heterogeneity. The three new isolates contain the genes for lactate oxidation and a gene cluster encoding enzymes of reverse β-oxidation, including the CoA transferase for the formation of *n*-caproate. Our analysis gave no hints on the isomerization pathway for *iso*-butyrate formation. An operon encoding the Rnf complex was found in BL-3 and BL-4 but not in BL-6, which may instead use the Ech hydrogenase complex for energy conservation. BL-3 and BL-6 were predicted to have genes encoding both the BCD/EtfAB complex and the LDH/EtfAB complex for energy coupling.

Conclusions

The genetic background of lactate-based chain elongation was confirmed in three novel *Clostridia* species that convert lactate to *n*-caproate and *iso*-butyrate. They contain highly conserved genes involved in reverse β-oxidation, hydrogen formation and either of two types of energy conservation systems (Rnf and Ech). Further research is needed to elucidate the mechanism of *iso*-butyrate formation in these strains. Features of the three isolates may be interesting for further applications in *n*-caproate and *iso*-butyrate production.

Background

Speciality chemicals such as *n*-caproate and *iso*-butyrate are valuable products of the carboxylate platform, with a broad range of potential applications in agriculture and industry [1–3]. For example, *n*-caproate can be used as promoter of plant growth and feed additive, or as precursor for the production of biofuels, lubricants and fragrances [1, 4–7]. Currently, *n*-caproate is mainly produced from vegetable oils such as palm kernel oil

[8], though it can be produced from more sustainable feedstocks such as agro-industrial waste by anaerobic fermentation and microbial chain elongation [9, 10]. Compared to linear carboxylates, branched-chain carboxylates such as *iso*-butyrate are of special interest for alternative applications due to their different physical properties, including higher viscosity, higher oxidative stability, and a lower boiling point [11]. For example, *iso*-butyrate can be used for the synthesis of texanol, which is a widely used coalescent for latex paints [2]. Currently, *iso*-butyrate is manufactured by acid-catalyzed Koch carbonylation of propylene, which is derived from fossil feedstock [2]. Microbial production of *iso*-butyrate from organic wastes or biomass residues is a more sustainable alternative, which has been demonstrated by recent studies [12, 13].

The metabolic process to produce *n*-caproate by anaerobic fermentation is called microbial chain elongation, also known as reverse β-oxidation. Several strict anaerobic bacteria are known as chain elongators that use ethanol as electron donor providing reducing equivalents and acetyl-CoA for the elongation of acyl-CoA units, thereby increasing the chain length of carboxylates by two carbons with each cycle [1]. For example, *Clostridium kluyveri* has been well described to elongate short-chain carboxylates (e.g., acetate) to *n*-caproate through reverse β-oxidation with ethanol and acetate as sole carbon and energy sources [14]. The review paper of Angenent et al. highlights the importance of the ethanol-based chain elongation pathway in biotechnology studies [1]. Additionally, chain elongation with lactate is getting increasing attention because some feedstocks (e.g., ensiled plant biomass) are rich in lactate, which is an important intermediate in the anaerobic breakdown of carbohydrates. To date, only few chain-elongating bacteria have been isolated that utilize lactate to produce *n*-caproate, including strains of *Megasphaera elsdenii*, *Megasphaera hexanoica*, *Pseudoramibacter alactolyticus* and *Ruminococcaceae* bacterium CPB6. It has been assumed that the mechanism of chain elongation with lactate is similar to that described for chain elongation with ethanol [10, 15]. However, insufficient knowledge has been generated yet on the physiology of lactate-based chain elongation from pure culture studies, and there is a lack of genome-level information to explore the genetic characteristics shared by chain-elongating bacteria. Previous studies have shown that *iso*-butyrate can be produced in methanol-based chain elongation [3, 12, 13]. The results suggested that *Clostridium luticellarii* might be responsible for the *iso*-butyrate formation during mixed culture fermentation, which was further tested by pure culture study of *C. luticellarii*, showing its ability to convert acetate and methanol to *iso*-butyrate [16]. However, the physiological reason for *iso*-butyrate formation in a chain elongation process has not been fully elucidated, particularly when lactate is the electron donor.

Recently, we reported on a complex bioreactor community that produced *n*-caproate from lactate-rich corn silage [17], and later a mixed culture producing *n*-caproate was enriched with lactate and xylan in a daily-fed bioreactor [18]. To investigate functional key species involved in *n*-caproate formation, we isolated several strains that are capable of converting lactate to *n*-caproate and *iso*-butyrate. For three isolates that turned out to represent novel species according to their 16S rRNA gene sequences, we performed whole genome sequencing and assembled the genomes with a short- and long-read sequencing hybrid approach as recently announced [19]. Further insight into the genomic and metabolic features of these strains may facilitate detailed understanding of lactate-based chain elongation.

The objectives of this study were to investigate the ability of the three strains to form *n*-caproate and *iso*-butyrate from lactate and to discuss these properties based on their complete genome sequences. Batch

experiments were conducted to explore the fermentation profiles with lactate. We performed functional annotation and taxonomic classification of the genomes and elucidated the genetic heterogeneity between the three strains. The shared genetic characteristics deduced from the genomes appear of further biotechnological interest. To analyze the genomic diversity of the entire repertoire of chain-elongating species and to identify the core genes of chain elongation-related pathways and their conservation, we did a comparative genome study by including eleven more genomes of experimentally validated chain-elongating species.

Results And Discussion

Isolation and identification of lactate-consuming strains

After incubation and several transfers of fermentation broth from a corn silage reactor with lactate as substrate, we enriched a mixed culture that produced acetate, *n*-butyrate, *iso*-butyrate and *n*-caproate (Additional file 1, Fig. S1). Isolation of lactate-consuming strains was achieved by plating the mixed culture on complex agar to isolate single colonies. Eleven pure cultures were obtained as confirmed by 16S rRNA gene sequencing. In liquid culture using mineral medium, three strains (designated as BL-3, BL-4 and BL-6) were found to convert lactate to *iso*-butyrate and *n*-caproate. The 16S rRNA gene sequence of BL-3 was 96.8% identical to that of *Clostridium luticellarii* FW431, BL-4 was 93.8% identical to that of *Ruminococcaceae* bacterium CPB6, and BL-6 was 96.3% identical to that of *Clostridium jeddahense* JCD. According to the current species threshold (98.7%) based on 16S rRNA gene identity [20], these three strains can be assumed to represent novel species.

Conversion of lactate to n-caproate and iso-butyrate in batch cultivation

The pure culture batch experiments showed that all three newly isolated strains can convert lactate into acetate, *n*-butyrate, *iso*-butyrate and *n*-caproate (Fig. 1). Started at an initial pH 5.5, the three strains gave different product spectra even though growing in the same mineral medium with lactate as the sole carbon source. Specifically, all three strains produced a large share of acetate (23–43%) and *n*-butyrate (35–57%), whereas propionate and *n*-caprylate were not detected. Based on the final concentrations (mmol C/L), strain BL-6 produced the highest proportion of *n*-caproate (18% for BL-6, 10% for BL-4 and 4% for BL-3) and strain BL-4 produced the highest proportion of *iso*-butyrate (23% for BL-4, 2% for BL-3 and BL-6). As shown in Fig. 1, the *n*-butyrate production rate decreased in cultures of BL-4 and BL-6 after the second spiking with lactate but was constant in the culture of BL-3. Simultaneously, the *iso*-butyrate production rate increased in BL-4 and the *n*-caproate production rate increased in BL-6. This indicates that further chain elongation of *n*-butyrate to *n*-caproate was catalyzed by strain BL-6 while strain BL-4 converted *n*-butyrate to *iso*-butyrate.

Genomic heterogeneity of strains BL-3, BL-4 and BL-6

The genomes of all three isolates were sequenced to better understand the genetic background of their metabolism, particularly of *n*-caproate and *iso*-butyrate formation from lactate. Based on the hybrid genome assembly of short reads (Illumina) and long reads (Oxford Nanopore Technologies), we recently announced high-quality draft genomes of these strains [19]. The genome sizes are depicted in Fig. 2 and detailed in

Table 1. According to the taxonomic classification of GTDB (Genome Taxonomy Database), BL-3 was assigned to *Clostridium*_B (*Clostridiaceae*), whereas BL-4 and BL-6 were assigned to the genera UBA4871 and *Clostridium*_E, respectively, both belonging to the *Acutalibacteraceae*. The number of predicted gene coding sequences (CDSs) ranges from around 2,300 to almost 3,900 in the three genomes (Table 1). For all three genomes, most of the CDSs could be classified in COG (Clusters of Orthologous Groups) functional categories (76% for BL-3, 75% for BL-4 and 73% for BL-6; see details in Additional file 1, Table S1) and EGGNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) categories (86% for BL-3, 85% for BL-4 and 83% for BL-6; see details in Additional file 1, Table S2). Comparative genome analysis revealed a total of 6,654 homologous gene families with 9,508 genes identified in all three genomes and indicates a relatively small core-genome size of 504 homologous gene families (Fig. 2). As for the 2,064 genes conserved in the core-genome, proportions of 27.2%, 20.9% and 19.1% can be considered as core CDSs of strains BL-3, BL-4 and BL-6, respectively. The core CDSs included all necessary genes involved in bioprocesses of lactate oxidation to acetyl-CoA, reverse β-oxidation, hydrogen formation and energy conservation (see Table 2 and details in Additional file 2). According to the pairwise comparison of the three genomes, a few synteny groups on nucleotide level were shared (Additional file 1, Fig. S2), which indicates the low conservation of genome organization and underlines the genomic heterogeneity of the three isolates.

Table 1
Genomic characteristics of all chain elongation strains included in this study

Strain	GTDB taxonomy	Isolation source	Genome size (bp)	GC content (%)	No. of predicted CDSs	Reference
BL-3	*Clostridium*_B	Anaerobic bioreactor	3,855,691	34.32	3,875	[19]
BL-4	*Acutalibacteraceae* UBA4871	Anaerobic bioreactor	2,335,857	42.75	2,323	[19]
BL-6	*Clostridium*_E sp002397665	Anaerobic bioreactor	3,435,529	54.63	3,496	[19]
Megasphaera elsdenii 14–14	*Megasphaera elsdenii*	Human gut	2,504,349	52.75	2,359	[57, 58]
Ruminococcaceae bacterium CPB6	*Acutalibacteraceae* UBA4871 sp002119605	Sludge of a caproate-producing reactor	2,069,994	50.58	2,116	[15, 59]
Megasphaera hexanoica MH	*Caecibacter massiliensis*	Cow rumen	2,877,851	49.00	2,799	[60]
Pseudoramibacter alactolyticus ATCC 23263	*Pseudoramibacter alactolyticus*	Human oral cavity	2,366,982	51.63	2,327	[61, 62]
Candidatus Pseudoramibacter fermentans[a]	*Pseudoramibacter* sp002396065	Anaerobic bioreactor	2,288,358	50.15	2,209	[21]
Clostridium kluyveri DSM 555	*Clostridium*_B *kluyveri*	Canal mud	4,023,800	32.02	4,371	[14]
Caproiciproducens galactitolivorans BS-1	*Acutalibacteraceae* MS4	Anaerobic digester sludge	2,578,839	48.10	2,539	[63, 64]
Eubacterium limosum KIST612	*Eubacterium limosum*	Sheep rumen	4,740,532	46.86	4,605	[61, 65]
Eubacterium pyruvativorans i6	*Eubacterium*_A *pyruvativorans*	Sheep rumen	2,164,212	54.84	1,954	[66, 67]
Candidatus Weimeria bifida[a]	*Lachnospiraceae* UBA2727	Anaerobic bioreactor	2,395,883	45.93	2,477	[21]
Clostridium luticellarii DSM 29923	*Clostridium*_B *luticellarii*	Mud cellar	3,771,178	34.97	3,874	[68, 69]

[a]metagenome-assembled genome (MAG)

Table 2
List of enzymes considered for the manual functional annotation

Predicted function	No.	Enzyme abbreviation	EC number	Enzyme	Reaction
Acetyl-CoA formation	1	LacR	5.1.2.1	Lactate racemase	D-lactate <=> L-lactate
	2	LacP	2.A.14	Lactate permease	
	3	LDH	1.1.1.27	Lactate dehydrogenase	(S)-lactate + NAD^+ → H^+ + NADH + pyruvate
	4	PFOR	1.2.7.1	Pyruvate ferredoxin oxidoreductase	Pyruvate + CoA + 2 oxidized ferredoxin → acetyl-CoA + CO_2 + 2 reduced ferredoxin + 2 H^+
	5	ADH	1.1.1.1	Alcohol dehydrogenase	A primary alcohol + NAD^+ → an aldehyde + NADH
	6	ADA	1.2.1.10	Acetaldehyde dehydrogenase	Acetaldehyde + CoA + NAD^+ → acetyl-CoA + NADH
Reverse β-oxidation	7	ACAT	2.3.1.9, 2.3.1.16	Acetyl-CoA acetyltransferase	2 Acetyl-CoA → acetoacetyl-CoA + CoA
	8	HAD	1.1.1.157, 1.1.1.35	3-Hydroxyacyl-CoA dehydrogenase	Acetoacetyl-CoA + NAD(P)H + H^+ → (S)-3-hydroxybutyryl-CoA + $NAD(P)^+$
	9	ECH	4.2.1.150, 4.2.1.55	Enoyl-CoA hydratase	(S)-3-hydroxybutyryl-CoA → crotonyl-CoA + H_2O
	10	BCD	1.3.8.1	Butyryl-CoA dehydrogenase	Crotonyl-CoA + a reduced electron-transfer flavoprotein → butyryl-CoA + an oxidized electron-Transfer flavoprotein + H^+
	11	EtfAB		Electron transfer flavoprotein A, B	
	12	CoAT	2.8.3.-	Butyryl-CoA:acetate CoA-transferase	Acyl-CoA + acetate → a fatty acid anion + acetyl-CoA
	13	ACT	3.1.2.20	Acyl-CoA thioesterase	An acyl-CoA + H_2O → a carboxylate + CoA + H^+
Energy conservation	14	RnfABCDEG	7.1.1.1	Energy-converting NADH:ferredoxin oxidoreductase	
	15	EchABCDEF		Energy-converting hydrogenase	

Predicted function	No.	Enzyme abbreviation	EC number	Enzyme	Reaction
H_2 formation	16	H2ase	1.12.7.2	Hydrogen:ferredoxin oxidoreductase	2 reduced ferredoxin + 2 H^+ → H_2 + 2 oxidized ferredoxin
Butyrate formation	17	PTB	2.3.1.19	Phosphate butyryltransferase	Butyryl-CoA + phosphate → CoA + butyryl phosphate
	18	BUK	2.7.2.7	Butyrate kinase	ADP + butyryl phosphate → ATP + butyrate
Others	19	BM	5.4.99.13	Butyryl-CoA:isobutyryl-CoA mutase	Butyryl-CoA → *iso*-butyryl-CoA
	20	ACOCT	2.8.3.19	Acetyl-CoA:oxalate CoA-transferase	Acetyl-CoA + oxalate → acetate + oxalyl-CoA
	21	HadABC	4.2.1.157	(R)-2-hydroxyisocaproyl-CoA dehydratase	
	22	CarC	1.3.1.108	Caffeyl-CoA reductase-Etf complex subunit CarC	
	23	HypCDEF		Hydrogenase maturation factor	

Genomic diversity of the reported chain-elongating bacterial strains

In addition to our newly isolated strains, we included eleven strains that have been experimentally validated of microbial chain elongation (Table 1). Two metagenome-assembled genomes (MAGs; *Candidatus* Pseudoramibacter fermentans and *Candidatus* Weimeria bifida) were also included in the comparative genome analysis because their chain elongation traits were evident from metatranscriptome analyses [21]. These 14 obligate anaerobes isolated from various environments all belong to the phylum *Firmicutes*, class *Clostridia* and its closest phylogenetic neighbor – *Negativicutes* (here including species *Megasphaera elsdenii* and *Megasphaera hexanoica*). The genome sizes of the strains range from 2.1 Mbp to 4.7 Mbp, and the GC content varies from 32–55% (Table 1).

We constructed a phylogenomic tree to understand the evolutionary relationships between our isolates and other chain-elongating species (Fig. 3a). The two main branches delineate that strain BL-3 is evolutionary distant from BL-4 and BL-6, as the latter were placed in the other main cluster. BL-3 belongs to a *Clostridiaceae* cluster and is closely related to two chain-elongating species of the genus *Clostridium*: *C. kluyveri* and *C. luticellarii*, with the latter having the highest OrthoANIu (average nucleotide identity by orthology with USEARCH) value of 83.88% to BL-3 (Fig. 3b). The closest chain-elongating relatives of BL-4 and BL-6 are *Ruminococcaceae* bacterium CPB6 and *Caproiciproducens galactitolivorans* BS-1, both affiliated to the family *Acutalibacteraceae* (according to GTDB taxonomy). BL-6 formed a separate cluster

together with *Clostridium jeddahense* and *Clostridium merdae*, for which chain elongation functions have not been described. However, BL-4 and BL-6 have relatively low OrthoANIu values (≤ 75%) and low genome coverages (≤ 25%, referring to the aligned genome fraction) with their closest relatives (Fig. 3b). For all three isolates, the synteny groups on nucleotide level delineate a low conservation of genome organization when aligned to the closest relative.

The number of predicted CDSs in the chain-elongating bacteria ranges from less than 2,000 to more than 4,600 (Table 1), which suggests substantial heterogeneity of their genomes. The pan-genome analysis of the genomes of all 14 strains shows a total of 20,790 homologous gene families with 40,582 genes identified (Fig. 4a). The core-genome presented in all 14 strains consists of only 237 conserved homologous gene families corresponding to 4775 core CDSs, which were distributed in a range of 9–15% for each strain (Fig. 4b). Interestingly, the number of pan-CDSs positively correlated with the genome size, whereas the number of strain-specific CDSs did not correlate with the genome size. For example, *C. kluyveri* DSM 555 holds the second largest genome (4.02 Mbp) with a number of 4288 pan-CDSs, but it has the lowest number of strain-specific CDS (287). The above-mentioned patterns also apply to the comparison of the three isolates as shown in Fig. 4b.

Functional distribution of homologous gene families in the core-genome shows that the majority encode components of well-conserved housekeeping genes for the basic metabolism of bacteria, including DNA and RNA metabolism, protein processing, folding and secretion, cellular processes as well as intermediary and energy metabolism (details in Additional file 3) [22]. The chain elongation specific core-genome also harbors genes involved in reverse β-oxidation, hydrogen formation and energy conservation (Table 2 and details in Additional file 4). These genes are highly conserved in all 14 strains and could be considered hallmarks of chain-elongating bacteria.

Genetic background of chain elongation

To elucidate the genetic background of lactate metabolism and fermentation pathways leading to the formation of *n*-caproate, *n*-butyrate and *iso*-butyrate, we manually curated the functional annotation of genes involved in the following bioprocesses: acetyl-CoA formation from lactate and ethanol, reverse β-oxidation cycle, energy conservation and hydrogen formation. Besides our newly isolated strains, we also included the other eleven chain elongators in this analysis. Especially for those strains reported to use lactate as electron donor, corresponding genes of lactate oxidation were also considered in the manual curation.

Lactate oxidation to acetyl-CoA

Lactate can serve as carbon and energy source for chain-elongating bacteria. As shown in Fig. 5, first lactate needs to be transported into the cell, which is facilitated by lactate permease (LacP). Genomes of BL-3 and BL-6 were predicted to harbor the corresponding CDSs, which are located in a gene cluster encoding lactate racemase (LacR) (Fig. 6a and 6c). The gene cluster encoding LacP and LacR was also found in all other lactate-based chain elongators (Fig. 6d-6 h). The fermentation starts with the oxidation of lactate via pyruvate to acetyl-CoA catalyzed by an NAD-dependent lactate dehydrogenase (LDH) and a pyruvate

ferredoxin oxidoreductase (PFOR). All three genomes encode predicted LDH proteins, which are highly similar to each other. Specifically, the BL-3 genome was predicted to have four LDH genes, one of which is located in a gene cluster (Fig. 6a, CDS labels: 11486–11488) comprising also genes for the electron transfer flavoprotein (EtfAB). The BL-4 genome harbors four LDH genes with one located in the gene cluster (Fig. 6b, CDS labels: 2199–2205) encoding membrane-associated energy-converting NADH:ferredoxin oxidoreductase (RnfABCDEG). The BL-6 genome has three LDH genes with one found in a cluster (Fig. 6c, CDS labels: 3216–3223) including genes for butyryl-CoA dehydrogenase (BCD), EtfAB, LacR and LacP. A similar gene cluster (Fig. 6e, CDS labels: 01775–01795) containing genes for LacR, LDH, EtfAB and BCD was found in the genome of *Ruminococcaceae* bacterium CPB6. As for the enzyme PFOR or its synonym pyruvate synthase, all three genomes contain the corresponding genes, enabling the oxidation of pyruvate to acetyl-CoA. Acetyl-CoA then enters the reverse β-oxidation cycles. CDSs for LDH and PFOR were found in all other lactate-based chain-elongating species (Fig. 6d-6 h).

Ethanol oxidation to acetyl-CoA

The ethanol-based chain elongation pathway is well elucidated in *C. kluyveri* [14], and is of particular importance in several biotechnology studies [23–25]. Genome data of BL-3 and BL-6 suggest that they may have the capacity to utilize ethanol as additional or alternative substrate. Small, uncharged molecules like ethanol diffuse through the cytoplasmic membrane and can be oxidized via acetaldehyde to acetyl-CoA. NAD-dependent alcohol dehydrogenase (ADH) and NAD(P)-dependent acetaldehyde dehydrogenase (ADA) catalyze this conversion (Fig. 5). The corresponding CDSs were found in the genomes of BL-3 and BL-6, while BL-4 lacks the gene encoding ADA.

n-*Butyrate and* n-*caproate formation*

Transformation of acetyl-CoA to butyryl-CoA includes three intermediates: acetoacetyl-CoA, 3-hydroxybutyryl-CoA and crotonyl-CoA. The involved enzymes are acetyl-CoA acetyltransferase (ACAT), NAD- and NADP-dependent 3-hydroxyacyl-CoA dehydrogenase (HAD), enoyl-CoA hydratase (ECH) and NAD-dependent butyryl-CoA dehydrogenase complex (BCD/EtfAB) (Fig. 5). The formation of *n*-butyrate further requires butyryl-CoA:acetate CoA transferase (CoAT) to catalyze the reaction of butyryl-CoA and acetate to yield acetyl-CoA and the corresponding fatty acid. Transformation of butyryl-CoA to caproyl-CoA probably happens with the same set of enzymes (ACAT, HAD, ECH and BCD/EtfAB) and a CoAT to remove the CoA from caproyl-CoA, resulting in the formation of *n*-caproate. We came up with the same assumption as described for the ethanol-based chain elongation mechanism of *C. kluyveri* [14] – caproyl-CoA can be a further elongated acyl-CoA when a second analogous cycle proceeds, and CoAT was reported to have a broad substrate specificity [26, 27]. All three genomes contain the genes encoding ACAT, HAD, ECH, BCD, EtfAB and CoAT (Additional file 4 including the summary of all related CDSs). As for BL-3, three sets of ACAT, HAD, ECH, BCD and EtfAB genes are present in the genome, with one cluster encoding CoAT, ACAT, ECH and HAD (Fig. 6a, CDS labels: 13110–13113) as well as one cluster encoding ECH, BCD, EtfAB and HAD (Fig. 6a, CDS labels: 20308–20313), other CDSs are scattered in the genome. As for BL-4, one gene cluster encoding all six enzymes is present in the genome (Fig. 6b, CDS labels: 1867–1873). Two similar clusters were found in the genomes of *Eubacterium limosum* (Fig. 6k, CDS labels: 21760–21785) and *Eubacterium*

pyruvativorans (Fig. 6i, CDS labels: 280031–280037). Another set of HAD, ACAT, ECH, CoAT genes clusters together with acetyl-CoA:oxalate CoA-transferase (ACOCT) and (R)-2-hydroxyisocaproyl-CoA dehydratase (HadABC) genes (Fig. 6b, CDS labels: 1158–1165). The genome of BL-6 harbors two sets of the ACAT, HAD, ECH, BCD and EtfAB genes separated into several sub-clusters, with one comprising genes for HAD, ACAT, ECH, CoAT and HadABC (Fig. 6c, CDS labels: 0555–0562) and two sub-clusters of genes encoding the BCD/EtfAB complex. One set of genes encoding the BCD/EtfAB complex is located in the same cluster with genes for LDH, LacR and LacP (Fig. 6c, CDS labels: 3216–3223) as mentioned above. We found that the genes encoding BCD are in close vicinity to the genes of EtfAB in the genomes of all three isolates (Fig. 6a-6c), which is commonly conserved as a key feature among all genomes of other chain-elongating bacteria (Fig. 6d-6n). Besides CoAT, the acyl-CoA thioesterase (ACT) may also catalyze the formation of *n*-butyrate and *n*-caproate from the terminal acyl-CoA (Fig. 5). Our data suggest that the genome of BL-3 may encode the predicted proteins annotated as thioesterase superfamily proteins. We further compared their protein sequences in all the databases used (see the results in Additional file 5) and confirmed that these thioesterase proteins were not involved in the terminal step of reverse β-oxidation (see CDS labels and final annotations in Additional file 4, sheets BL-3). Genomes of BL-4 and BL-6 both contain the CDSs for ACT (see CDS labels in Additional file 4, sheets BL-4 and BL-6), but presenting a low identity (≤ 40%) to proteins in the databases (see alignment details in Additional files 6 and 7).

Another possible pathway for the *n*-butyrate formation from *n*-butyryl-CoA was identified in the genome of BL-3. As shown in Fig. 5, the generation of butyrate phosphate is catalyzed by phosphate butyryltransferase (PTB). Thereafter, it is converted to butyrate via butyrate kinase (BUK), producing one ATP in this step.

iso-*Butyrate formation*

The formation of *iso*-butyrate as a product of lactate-based chain elongation was experimentally proven in all three isolates. However, the genome analysis did not reveal hints on the assumed pathway, i.e. reversible *n*-butyrate/*iso*-butyrate isomerization [28, 29]. As described by Matthies and Schink [29], the conversion of *n*-butyrate to *iso*-butyrate first requires activation to *n*-butyryl-CoA. Next, the isomerization of *n*-butyryl-CoA via *iso*-butyryl-CoA to *iso*-butyrate is catalyzed by a butyryl-CoA:isobutyryl-CoA mutase (BM) and an isobutyryl-CoA:acetate CoA transferase (CoAT) as shown in Fig. 5. All three genomes lack potential CDSs for BM. The only homologue was found in the genome of BL-3 annotated as methylmalonyl-CoA mutase (see the CDS labels and final annotations in Additional file 4, sheets BL-3), which is known to isomerize (R)-methylmalonyl-CoA to succinyl-CoA, a step involved in the propionate fermentation pathway. These mutases catalyze the rearrangement of carboxyl groups represented as migration to the adjacent carbon atom, in which enzyme activities depend on coenzyme B_{12} [30]. One possible reason for the conversion of *n*-butyrate to *iso*-butyrate is that bacteria can maintain the pool of *iso*-butyrate for synthesizing valine during growth in amino acid-deficient medium [31]. As this isomerization step does not release any free energy, another possible explanation is that bacteria try to overcome inhibition effects of the accumulated *n*butyrate, because the corresponding fatty acid of the unbranched form is more toxic than the branched form. As suggested for a methanol-based chain elongation system [3, 12], the formation of *iso*-butyrate may facilitate bacteria to further obtain energy from the chain elongation process.

Energy conservation and hydrogen formation

As shown in Fig. 5, the cytoplasmic BCD/EtfAB complex catalyzes the transformation of crotonyl-CoA (hexenoyl-CoA) to butyryl-CoA (caproyl-CoA) and simultaneously transfers electrons from NADH to ferredoxin, a mechanism that has been described as flavin-based electron bifurcation [32]. ATP can be produced by the ATP synthase using the ion motive force that is generated by a membrane-associated, proton-translocating ferredoxin:NAD^+ oxidoreductase (Rnf complex) in the oxidation of ferredoxin [33]. The genomes of BL-3 and BL-4 contain the operon arranged as *rnfCDGEAB* encoding the six subunits of the Rnf complex as shown in Fig. 6a and 6b. This gene organization (shown as *rnfBAEGDC* in the other DNA strand) was also found in other genomes of chain-elongating bacteria (Fig. 6d-n). For the genome of BL-6, we could only find four genes for subunits of the Rnf complex during the functional annotation (see CDS labels in the Additional file 4, sheet BL-6), but it contains the CDSs encoding the analogous membrane-associated energy-converting hydrogenase (Ech complex), which was proposed to generate hydrogen for maintaining the cytoplasmic redox balance caused by the oxidation of ferredoxin [34, 35]. As shown in Fig. 6c, CDS labels 2699–2708, a cluster encoding six subunits of the Ech complex and CDSs for the hydrogenase maturation were found. The Ech complex was also identified in the MAG of *Candidatus* Weimeria bifida (Fig. 6m). Additional hydrogenases include hydrogen:ferredoxin oxidoreductase (H2ase), which was found in the genomes of all three isolates, and the bifurcating [Fe-Fe]-hydrogenase (HydABC) using electrons from NADH and reduced ferredoxin, of which no homologous genes were detected (see CDS labels in Additional file 4, sheets BL-3, BL-4 and BL-6).

Apart from the BCD/EtfAB complex, the predicted EtfAB-containing complexes for energy coupling may also include the LDH/EtfAB complex. The redox potential of the pyruvate/lactate pair (E_0' = -190 mV) is much higher than that of the NAD^+/NADH pair (E_0' = 320 mV), which introduces a thermodynamic bottleneck of the lactate oxidation coupled to NAD^+ reduction. Our annotation results show that strains BL-3, BL-6 and *Ruminococcaceae* bacterium CPB6 have LDH genes next to EtfAB genes (Fig. 6a, CDS labels: 11486–11488; Fig. 6c, CDS labels: 3217–3220; Fig. 6e, CDS labels: 01780–01790). Therefore, similar like the mode of lactate metabolism in the strict anaerobic acetogen *Acetobacterium woodii*, we assume that the LDH/EtfAB complex of these species can also use flavin-based electron confurcation to solve the energetic enigma: driving electron flow from lactate to NAD^+ at the cost of exergonic electron flow from reduced ferredoxin to NAD^+ [33, 36].

The manually curated annotation of all above-mentioned CDSs in the genomes of other lactate-based chain-elongating strains are provided in Additional files 8–12 (CDSs predicted with all methods integrated in the MicroScope platform).

Sugar metabolism of chain-elongating bacteria

We further analyzed the genetic potential of the three isolates to metabolize monosaccharides (xylose, glucose and fructose) to fermentation products. Strains BL-3 and BL-6 both can potentially ferment xylose and fructose, but not glucose; BL-4 cannot utilize any of these sugars (Additional file 13). BL4 has a smaller genome (2.3 Mbp) than BL-3 (3.9 Mbp) and BL-6 (3.4 Mbp), and such small genomes with poor sugar fermentation pathways are also characteristic of several other chain-elongating strains such as *Ruminococcaceae* bacterium CPB6, *C. kluyveri* DSM555, *C. galactitolivorans* BS1 and *E. pyruvativorans* i6,

which may reflect the specific ecological niche of chain-elongating bacteria with small genomes and streamlined carbohydrate metabolism.

Conclusions

Our results suggest three novel *Clostridia* species (BL-3, BL-4 and BL-6) that are able to convert lactate to *n*-caproate and *iso*-butyrate in batch cultivation, with the confirmation of their genetic background of lactate-based chain elongation and using CoA transferase as the terminal enzyme. Further research is needed to elucidate the mechanism of *iso*-butyrate formation in these strains. By comparative genome analysis including further eleven experimentally validated chain-elongating bacteria, we found a substantial genetic heterogeneity but highly conserved genes related to chain elongation, hydrogen formation and energy conservation, which can be considered hallmarks of chain-elongating bacteria. Based on the genomic features, chain-elongating species may contain two types of energy conservation systems in the re-oxidation of reduced ferredoxin – proton-translocating ferredoxin:NAD^+ oxidoreductase (Rnf complex) and energy-converting hydrogenase (Ech complex). Besides the proposed BCD/EtfAB complex for flavin-based electron bifurcation, energy coupling may also include the LDH/EtfAB complex in the oxidation of lactate and the supply of acetyl-CoA for chain elongation. Overall, this study investigated the genomic and metabolic features of three novel chain-elongating isolates, which could be interesting for further applications in *n*-caproate and *iso*-butyrate production.

Methods

Enrichment, isolation and identification of lactate-consuming strains

Anaerobic fermentation broth from a caproate-producing reactor (38 °C, pH 5.5 and hydraulic retention time of 4 d) fed with corn silage was initially taken as the inoculum. Serum bottles (120 mL) with 45 mL mineral medium [18] containing 5 g/L lactic acid (initial pH 5.5) were inoculated with 5 mL of the sieved reactor broth (mesh size 2 mm). After replacing the headspace by N_2/CO_2 (80:20 in a volume ratio, 100 kPa), the bottles were statically incubated at 37 °C in the dark. Liquid samples were collected every two weeks at the beginning, and later lactic acid was replenished when it had been consumed. Four successive transfers (1:10 dilution in fresh medium) were done spanning more than 700 days.

A single bottle of the fourth transfer was used to isolate lactate-consuming strains. The culture was plated on complex agar (medium DSM104c with additional 5 g/L lactic acid) and incubated in an anaerobic chamber at 37 °C for two weeks. Colonies were picked and re-streaked three times for purification, and then transferred to liquid mineral medium bottles to determine their product spectrum. Further, the isolates that produced *iso*-butyrate and *n*-caproate were identified by Sanger sequencing of the 16S rRNA gene (details in Additional file 1). Based on 16S rRNA gene identity with their closest relatives, potential new species including the isolates designated as strains BL-3, BL-4 and BL-6 were selected for whole genome sequencing.

Lactate utilization in batch cultivation

Batch cultures of isolates BL-3, BL-4 and BL-6 were run in mineral medium with lactate as sole carbon source as described above. The bottles were inoculated with 5 mL seed cultures (optical density at 600 nm ~ 2), which were routinely cultivated in a complex medium (DSM 104c with extra 5 g/L of lactic acid added). The pH was adjusted to 5.5 with 1 M NaOH or 1 M H_2SO_4 after adding 50 mM lactic acid (85%, FCC grade; Sigma Aldrich, St. Louis, USA) to the bottles. The cultivation bottles were statically incubated at 37 °C. Liquid samples were collected twice per week. After one week, lactic acid (75 mM) was added again to each bottle, and adjusted to pH 5.5 accordingly. The batch tests in this study were carried out in duplicate.

Analytical techniques

Liquid samples of the cultures were centrifuged for 10 min at 20,817 × *g* (Centrifuge 5417R; Eppendorf, Hamburg, Germany). Acetate, lactate, propionate, *iso*-butyrate, *n*-butyrate, *n*-valerate, *n*-caproate, *n*-caprylate and ethanol concentrations of the supernatant were determined in triplicate by high performance liquid chromatography (Shimadzu Corporation, Kyoto, Japan) equipped with a refractive index detector RID-10A and a HiPlex H column together with a pre-column (Agilent Technologies) as previously described [37].

Gene prediction and annotation

We sequenced the genomes of the three isolates with the Oxford Nanopore Technologies MinION and the Illumina NextSeq platforms, and three complete genomes were constructed using a hybrid assembly approach as described previously [19]. Prediction and functional annotation of CDSs was accomplished by the MicroScope automatic annotation pipeline [38]. Automatic annotations of selected CDSs were manually curated by comparing the protein sequences with the PkGDB, Swiss-Prot, TrEMBL, COG, EGGNOG, FIGfams and InterPro databases [38–43], and by using the following methods: MaGe/Curated annotation, Syntonome RefSeq, Similarities SwissProt, Similarities TrEMBL, UniFIRE SAAS, UniFIRE UniRules, PRIAM EC number, FigFam, InterProScan and PsortB. COGNiTOR [44] was used to classify the CDSs into COG functional categories. CDSs classification into EGGNOG (v4.5.1) was performed by eggNOG-mapper v1.0.3 [41]. All these databases and tools are integrated in the MicroScope platform as described by Vallenet et al. [38]. Genomes of *Clostridium jeddahense* JCD, *Ruminococcaceae* bacterium CPB6, *Clostridium merdae* Marseille-P2935, *Megasphaera elsdenii* 14–14, *Eubacterium pyruvativorans* i6, *Megasphaera hexanoica* MH, *Caproiciproducens* sp. NJN-50, *Caproiciproducens galactitolivorans* BS1, *Eubacterium limosum* KIST612, *Candidatus* Weimeria bifida, *Candidatus* Pseudoramibacter fermentans and *Pseudoramibacter alactolyticus* ATCC 23263 were submitted to the MicroScope platform. The genome annotation of these strains available in the MicroScope PkGDB database was done by following the same procedures.

Phylogenetic analysis and taxonomic classification

Phylogenetic analysis of 16S rRNA gene sequences was performed on the Phylogeny.fr platform [45]. According to the Nucleotide BLAST (Basic Local Alignment Search Tool) comparison results against the rRNA/ITS databases (16S ribosomal RNA sequences (Bacteria and Archaea)) of NCBI (National Center for Biotechnology Information) [46], the ten hits with the highest BLAST score for each isolate were selected. The 16S rRNA gene sequences of all selected strains were aligned using MUSCLE v3.8.31 with default

settings [47]. After alignment, Gblocks v0.91b was used to remove ambiguous regions (i.e. containing gaps and/or poorly aligned) as described by Castresana [48]. The phylogenetic tree was reconstructed using the maximum likelihood method contained in PhyML v3.1 [49, 50]. Robustness of tree topology was assessed by 100 bootstrap replicates. Finally, the tree was visualized by using TreeDyn v198.3 [51]. Besides the taxonomic classification of the genomes in MicroScope, GTDB-Tk v1.0.2 was used for taxonomic assignment to GTDB [52], and the corresponding NCBI taxonomy.

A phylogenomic tree of strains BL-3, BL-4 and BL-6 and other chain-elongating bacteria was calculated based on genomic similarity. The genomic similarity was estimated using Mash [53], which computes the distance between two genomes. This distance D is correlated to the average nucleotide identity (ANI) like: $D \approx 1\text{-ANI}$. A neighbor-joining tree was constructed, which displays clustering annotations. This clustering was calculated from all-pairs distances ≤ 0.06 ($\approx 94\%$ ANI), corresponding to the ANI standard to define a species group. The Louvain method for community detection was used for computing the clustering [54]. The ANI (OrthoANIu value) comparison of the genomes of the isolates to related genomes was calculated by an ANI calculator, which improved the original OrthoANI (Average Nucleotide Identity by Orthology) algorithm by applying USEARCH instead of BLAST as described by Yoon et al. [55].

Default settings were used for all tools unless otherwise specified.

Pan-genome analysis

The interface Comparative Genomics of the MicroScope platform was employed to analyze the pan-genome, core-genome and variable genome for our newly sequenced genomes and for all the available genomes of chain-elongating bacteria in the comparison. The MicroScope homologous gene families (MICFAM, protein sequence pairs with at least 80% amino-acid identity and 80% alignment coverage) [56] were considered for these analyses.

Additional Files

Additional file 1: Supplementary Methods, Tables and Figures

Supplementary Methods Sanger sequencing of 16S rRNA genes. **Table S1** COG (Clusters of Orthologous Groups) classification. **Table S2** EGGNOG (Evolutionary Genealogy of Gene: Non-supervised Orthologous Groups) classification. **Fig. S1** Fermentation products of the enrichment culture (a single bottle of the fourth transfer) during growth on lactate. Mean values of three measurements are given and error bars represent the standard deviation. **Fig. S2** Pairwise comparison of the conservation of the synteny groups in the three new isolates. Strand conservations are depicted in purple and strand inversions in blue. The synton size corresponds to at least three genes. **Fig. S3** Maximum likelihood tree of the three new strains and closest relatives based on 16S rRNA gene sequences. Bootstrap values above 50% are shown at the node. Accession numbers of 16S rRNA gene sequences are given in parentheses. Scale bar = 8% nucleotide substitution.

Additional file 2: Results of pan-genome analysis of the three isolates

Additional file 3: Results of pan-genome analysis of all chain-elongating strains

Additional file 4: Annotation summary of all chain-elongating strains

Additional file 5: Functional annotations of strain BL-3

Additional file 6: Functional annotations of strain BL-4

Additional file 7: Functional annotations of strain BL-6

Additional file 8: Functional annotations of *Megasphaera elsdenii*

Additional file 9: Functional annotations of *Ruminococcaceae* bacterium CPB6

Additional file 10: Functional annotations of *Megasphaera hexanoica*

Additional file 11: Functional annotations of *Pseudoramibacter alactolyticus*

Additional file 12: Functional annotations of *Candidatus* Pseudoramibacter fermentans

Additional file 13: Functional annotations of genes involved in sugar fermentation for all chain-elongating strains

Abbreviations

ACAT: acetoacetyl-CoA acetyltransferase; ACOCT: acetyl-CoA:oxalate CoA-transferase; ACT: acyl-CoA thioesterase; ADA: NAD(P)-dependent acetaldehyde dehydrogenase; ADH: NAD-dependent alcohol dehydrogenase; ANI: average nucleotide identity; BCD: butyryl-CoA dehydrogenase; BLAST: basic local alignment search tool; BM: butyryl-CoA:isobutyryl-CoA mutase; BUK: butyrate kinase; CDS: coding sequence; CoA: coenzyme A; CoAT: CoA transferase; COG: clusters of orthologous groups; ECH: enoyl-CoA hydratase; EGGNOG: evolutionary genealogy of genes: non-supervised orthologous groups; GTDB: Genome Taxonomy Database; HAD: NAD- and NADP-dependent 3-hydroxyacyl-CoA dehydrogenase; InterPro: the integrative protein signature database; LDH: lactate dehydrogenase; MAG: metagenome-assembled genome; MICFAM: MicroScope homologous gene families; NCBI: National Center for Biotechnology Information; OrthoANIu: average nucleotide identity by orthology with USEARCH; PFOR: pyruvate ferredoxin oxidoreductase; PkGDB: prokaryotic genome database; PTB: phosphate butyryltransferase; TrEMBL: translated EMBL

Declarations

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional files. The 16S rRNA gene sequences and the genome sequences of the three isolates have been deposited in the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/view/) under BioProject nos. PRJEB39379 and PRJEB36835, respectively.

Funding

This work was funded by the China Scholarship Council (# 201606350010), the BMBF – German Federal Ministry of Education and Research (# 031B0389B, # 01DQ17016, and # 031A317) and the Helmholtz Association (Program Renewable Energies). The funding agencies had neither influence on the design of the study, the collection, analysis and interpretation of the data nor the writing of the manuscript.

Authors' contributions

BL carried out isolation and physiological characterization of the strains as well as manual curation of the genome annotation and drafted the manuscript. BL and DP performed genome sequencing and genome assembly. BL, DP, HS, and SK contributed to data analysis and interpretation. HH contributed to the discussion of the results. All authors critically contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Ute Lohse for her technical assistance in DNA extraction and Sanger sequencing. Cloud computing facilities used for the analysis of the amplicon were provided by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Angenent LT, Richter H, Buckel W, Spirito CM, Steinbusch KJJ, Plugge CM, et al. Chain elongation with reactor microbiomes: open-culture biotechnology to produce biochemicals. Environ Sci Technol. 2016;50:2796–810.
2. Zhang K, Woodruff AP, Xiong M, Zhou J, Dhande YK. A synthetic metabolic pathway for production of the platform chemical isobutyric acid. ChemSusChem. 2011;4:1068–70.
3. de Leeuw KD, de Smit SM, van Oossanen S, Moerland MJ, Buisman CJN, Strik DPBTB. Methanol-based chain elongation with acetate to *n*–butyrate and isobutyrate at varying selectivities dependent on pH. ACS Sustain Chem Eng. 2020;8:8184–94.
4. Scalschi L, Vicedo B, Camañes G, Fernandez-Crespo E, Lapeña L, González-Bosch C, et al. Hexanoic acid is a resistance inducer that protects tomato plants against *Pseudomonas syringae* by priming the

jasmonic acid and salicylic acid pathways. Mol Plant Pathol. 2013;14:342–55.

5. Urban C, Xu J, Sträuber H, dos Santos Dantas TR, Mühlenberg J, Härtig C, et al. Production of drop-in fuel from biomass by combined microbial and electrochemical conversions. Energy Environ Sci. 2017;10:2231–44.
6. Evans NP, Collins DA, Pierson FW, Mahsoub HM, Sriranganathan N, Persia ME, et al. Investigation of medium chain fatty acid feed supplementation for reducing *Salmonella typhimurium* colonization in turkey poults. Foodborne Pathog Dis. 2017;14:531–6.
7. Kenealy WR, Cao Y, Weimer PJ. Production of caproic acid by cocultures of ruminal cellulolytic bacteria and *Clostridium kluyveri* grown on cellulose and ethanol. Appl Microbiol Biotechnol. 1995;44:507–13.
8. Anneken DJ, Both S, Christoph R, Fieg G, Steinberner U, Westfechtel A. Fatty acids. *Ullmann's Encyclopedia of Industrial Chemistry*. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA. 2006;14:73–116.
9. Agler MT, Wrenn BA, Zinder SH, Angenent LT. Waste to bioproduct conversion with undefined mixed cultures: the carboxylate platform. Trends Biotechnol. 2011;29:70–8.
10. Wu Q, Bao X, Guo W, Wang B, Li Y, Luo H, et al. Medium chain carboxylic acids production from waste biomass: Current advances and perspectives. Biotechnol Adv. 2019;37:599–615.
11. de Leeuw KD, Buisman CJN, Strik DPBTB. Branched medium chain fatty acids: *iso*-caproate formation from *iso*-butyrate broadens the product spectrum for microbial chain elongation. Environ Sci Technol. 2019;53:7704–13.
12. Chen WS, Huang S, Strik DPBTB, Buisman CJN. Isobutyrate biosynthesis via methanol chain elongation: converting organic wastes to platform chemicals. J Chem Technol Biotechnol. 2017;92:1370–9.
13. Huang S, Kleerebezem R, Rabaey K, Ganigué R. Open microbiome dominated by *Clostridium* and *Eubacterium* converts methanol into *i*-butyrate and *n*-butyrate. Appl Microbiol Biotechnol. 2020;104:5119–31.
14. Seedorf H, Fricke WF, Veith B, Brüggemann H, Liesegang H, Strittmatter A, et al. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc Natl Acad Sci USA. 2008;105:2128–33.
15. Zhu X, Zhou Y, Wang Y, Wu T, Li X, Li D, et al. Production of high-concentration *n*-caproic acid from lactate through fermentation using a newly isolated *Ruminococcaceae* bacterium CPB6. Biotechnol Biofuels. 2017;10:102.
16. van Brabant P. Understanding bio-isomerisation during methanol fermentation. M.Sc. Thesis, Ghent University; 2019; Catalog number: rug01:002785053.
17. Lambrecht J, Cichocki N, Schattenberg F, Kleinsteuber S, Harms H, Müller S, et al. Key sub-community dynamics of medium-chain carboxylate production. Microb Cell Fact. 2019;18:92.
18. Liu B, Kleinsteuber S, Centler F, Harms H, Sträuber H. Competition between butyrate fermenters and chain-elongating bacteria limits the efficiency of medium-chain carboxylate production. Front Microbiol. 2020;11:336.
19. Liu B, Popp D, Sträuber H, Harms H, Kleinsteuber S. Draft genome sequences of three *Clostridia* isolates involved in lactate-based chain elongation. Microbiol Resour Announc. 2020; accepted.

20. Erko S, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today. 2006;33:152–5.
21. Scarborough MJ, Myers KS, Donohue TJ, Noguera DR. Medium-chain fatty acid synthesis by "*Candidatus* Weimeria bifida" gen. nov., sp. nov., and "*Candidatus* Pseudoramibacter fermentans" sp. nov. Appl Environ Microbiol. 2020;86:e02242-19.
22. Gil R, Silva FJ, Pereto J, Moya A. Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev. 2004;68:518–37.
23. Agler MT, Spirito CM, Usack JG, Werner JJ, Angenent LT. Chain elongation with reactor microbiomes: upgrading dilute ethanol to medium-chain carboxylates. Energy Environ Sci. 2012;5:8189.
24. Kucek L, Spirito CM, Angenent LT. High *n*-caprylate productivities and specificities from dilute ethanol and acetate: chain elongation with microbiomes to upgrade products from syngas fermentation. Energy Environ Sci. 2016;9:3482–94.
25. Grootscholten TIM, Steinbusch KJJ, Hamelers HVM, Buisman CJN. Chain elongation of acetate and ethanol in an upflow anaerobic filter for high rate MCFA production. Bioresour Technol. 2013;135:440–5.
26. Stadtman ER. The coenzyme A transphorase system in *Clostridium kluyveri*. J Biol Chem. 1953;203:501–12.
27. Stadtman ER. Functional group of coenzyme A and its metabolic relations, especially in the fatty acid cycle: discussion. Fed Proc. 1953;12:692–3.
28. Tholozan JL, Samain E, Grivet JP. Isomerization between *n*-butyrate and isobutyrate in enrichment cultures. FEMS Microbiol Lett. 1988;53:187–91.
29. Matthies C, Schink B. Reciprocal isomerization of butyrate and isobutyrate by the strictly anaerobic bacterium strain WoG13 and methanogenic isobutyrate degradation by a defined triculture. Appl Environ Microbiol. 1992;58:1435–9.
30. Barker HA. Coenzyme B_{l2}-dependent mutases causing carbon chain rearrangements. Enzym. 1972;6:509–37.
31. Allison MJ. Production of branched-chain volatile fatty acids by certain anaerobic bacteria. Appl Environ Microbiol. 1978;35:872–7.
32. Buckel W, Thauer RK. Flavin-based electron bifurcation, a new mechanism of biological energy coupling. Chem Rev. 2018;118:3862–86.
33. Buckel W, Thauer RK. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Biochim Biophys Acta – Bioenerg. 2013;1827:94–113.
34. Hedderich R, Forzi L. Energy-converting [NiFe] hydrogenases: more than just H_2 activation. J Mol Microbiol Biotechnol. 2006;10:92–104.
35. Schuchmann K, Müller V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. Nat Rev Microbiol. 2014;12:809–21.
36. Weghoff MC, Bertsch J, Müller V. A novel mode of lactate metabolism in strictly anaerobic bacteria. Environ Microbiol. 2015;17:670–7.

37. Sträuber H, Schröder M, Kleinsteuber S. Metabolic and microbial community dynamics during the hydrolytic and acidogenic fermentation in a leach-bed process. Energy Sustain Soc. 2012;2:13.
38. Vallenet D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, et al. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. Nucleic Acids Res. 2019;48:D579–89.
39. Bateman A. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.
40. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4:41.
41. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34:2115–22.
42. Meyer F, Overbeek R, Rodriguez A. FIGfams: yet another set of protein families. Nucleic Acids Res. 2009;37:6643–54.
43. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: The integrative protein signature database. Nucleic Acids Res. 2009;37:211–5.
44. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.
45. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36:W465–9.
46. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36:W5–9.
47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
48. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.
49. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52:696–704.
50. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 2006;55:539–52.
51. Chevenet F, Brun C, Bañuls AL, Jacq B, Christen R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics. 2006;7:439.
52. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with Genome Taxonomy Database. Bioinfomatics. 2019;36:1925–7.
53. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:132.
54. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008;10:P10008.

55. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110:1281–6.
56. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, et al. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res. 2017;45:D517–28.
57. Stanton TB, Humphrey SB. Isolation of tetracycline-resistant *Megasphaera elsdenii* strains with novel mosaic gene combinations of tet(O) and tet(W) from swine. Appl Environ Microbiol. 2003;69:3874–82.
58. Weimer PJ, Moen GN. Quantitative analysis of growth and volatile fatty acid production by the anaerobic ruminal bacterium *Megasphaera elsdenii* T81. Appl Microbiol Biotechnol. 2013;97:4075–81.
59. Tao Y, Zhu X, Wang H, Wang Y, Li X, Jin H, et al. Complete genome sequence of *Ruminococcaceae* bacterium CPB6: a newly isolated culture for efficient *n*-caproic acid production from lactate. J Biotechnol. 2017;259:91–4.
60. Jeon BS, Kim S, Sang BI. *Megasphaera hexanoica* sp. nov., a medium-chain carboxylic acid-producing bacterium isolated from a cow rumen. Int J Syst Evol Microbiol. 2017;67:2114–20.
61. Willems A, Collins MD. Phylogenetic relationships of the genera *Acetobacterium* and *Eubacterium* sensu stricto and reclassification of *Eubacterium alactolyticum* as *Pseudoramibacter alactolyticus* gen. nov., comb. nov. Int J Syst Bacteriol. 1996;46:1083–7.
62. Holdeman LV, Cato EP, Moore WEC. Amended description of *Ramibacterium alactolyticum* Prevot and Taffanel with proposal of a neotype strain. Int J Syst Bacteriol. 1967;17:323–41.
63. Kim BC, Jeon BS, Kim S, Kim H, Um Y, Sang BI. *Caproiciproducens galactitolivorans* gen. nov., sp. nov., a bacterium capable of producing caproic acid from galactitol, isolated from a wastewater treatment plant. Int J Syst Evol Microbiol. 2015;65:4902–8.
64. Bengelsdorf FR, Poehlein A, Daniel R, Dürre P. Genome sequence of the caproic acid-producing bacterium *Caproiciproducens galactitolivorans* BS-1^T (JCM 30532). Microbiol Resour Announc. 2019;8:e00346-19.
65. Genthner BR, Davis CL, Bryant MP. Features of rumen and sewage sludge strains of *Eubacterium limosum*, a methanol- and H_2-CO_2-utilizing species. Appl Environ Microbiol. 1981;42:12–9.
66. Wallace RJ, McKain N, McEwan NR, Miyagawa E, Chaudhary LC, King TP, et al. *Eubacterium pyruvativorans* sp. nov., a novel non-saccharolytic anaerobe from the rumen that ferments pyruvate and amino acids, forms caproate and utilizes acetate and propionate. Int J Syst Evol Microbiol. 2003;53:965–70.
67. Wallace RJ, Chaudhary LC, Miyagawa E, McKain N, Walker ND. Metabolic properties of *Eubacterium pyruvativorans*, a ruminal “hyper-ammonia-producing” anaerobe with metabolic properties analogous to those of *Clostridium kluyveri*. Microbiology. 2004;150:2921–30.
68. Wang Q, Wang CD, Li CH, Li JG, Chen Q, Li YZ. *Clostridium luticellarii* sp. nov., isolated from a mud cellar used for producing strong aromatic liquors. Int J Syst Evol Microbiol. 2015;65:4730–3.
69. Poehlein A, Bremekamp R, Lutz VT, Schulz LM, Daniel R. Draft genome sequence of the butanoic acid-producing bacterium *Clostridium luticellarii* DSM 29923, used for strong aromatic Chinese liquor production. Genome Announc. 2018;6:e00556-18.

Figures

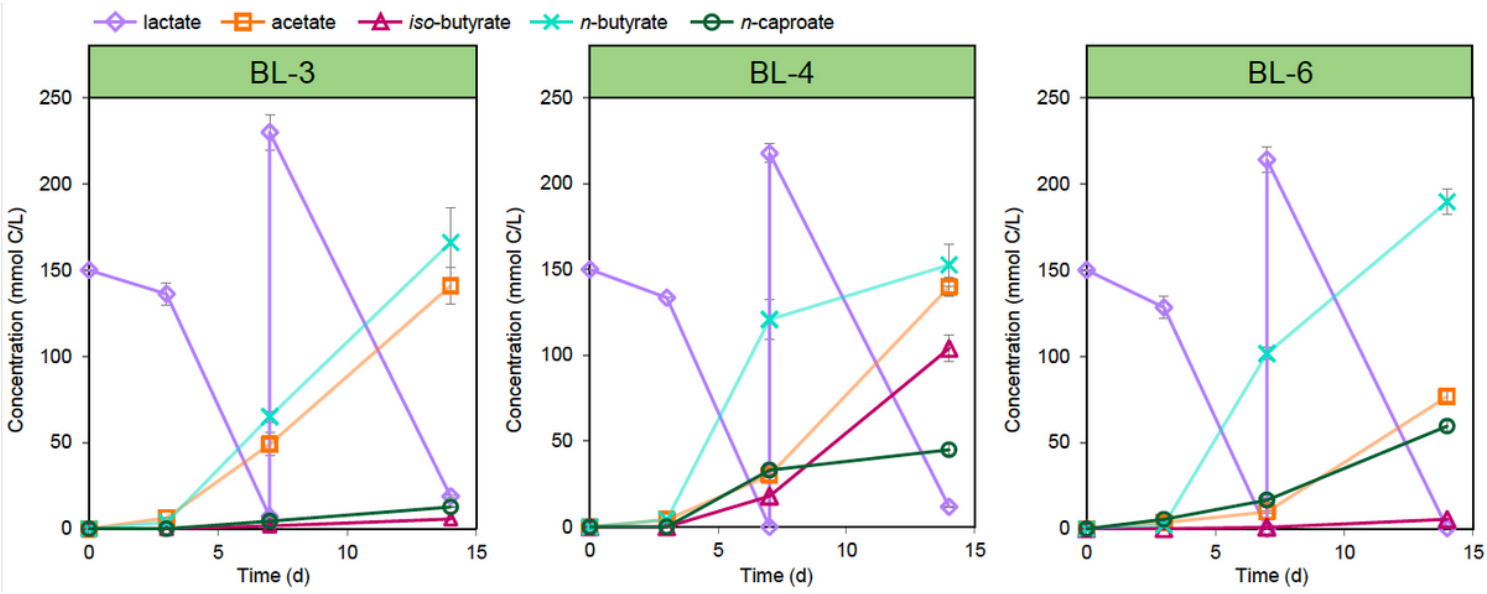
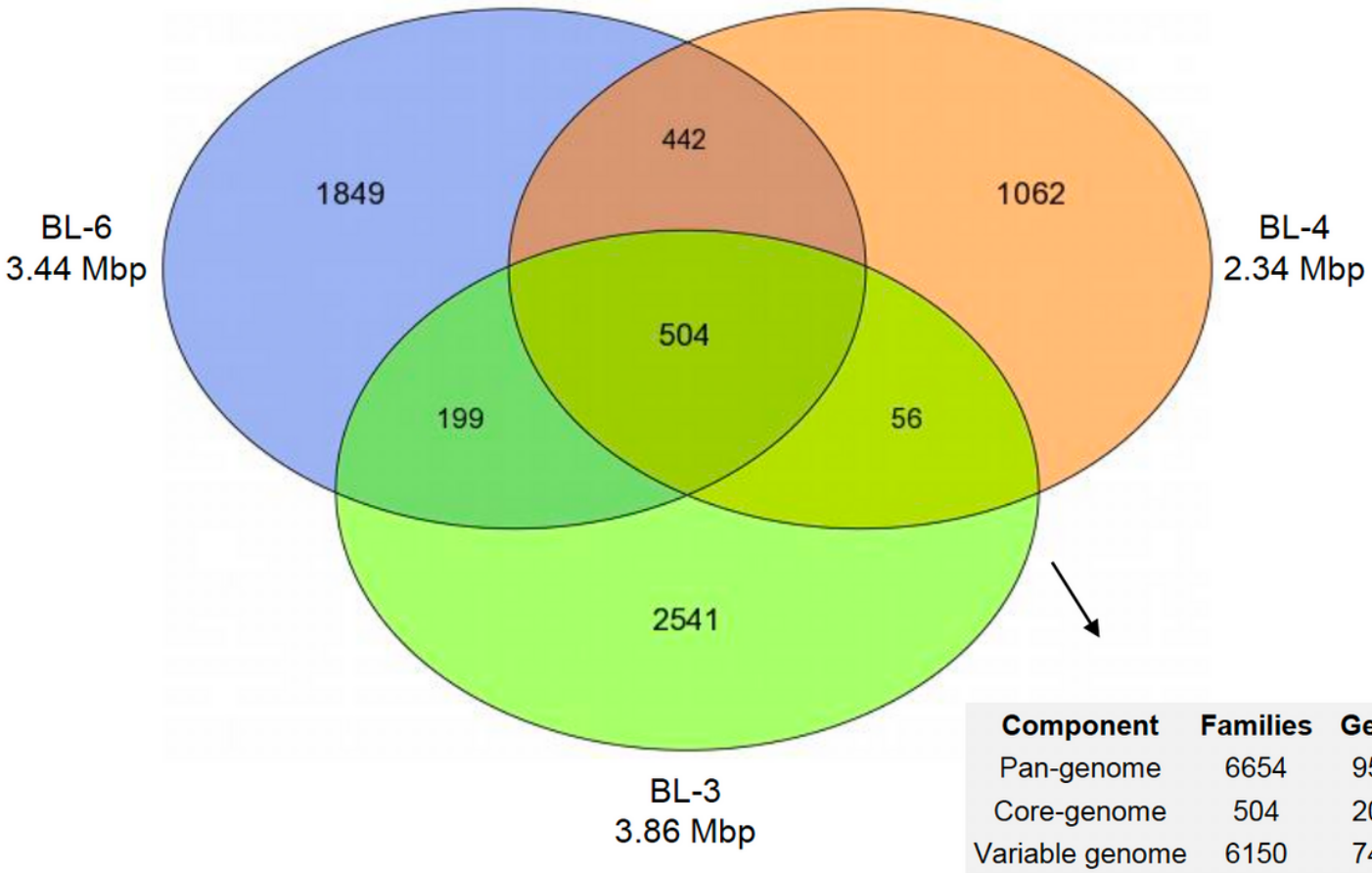


Figure 1

Fermentation products of strains BL-3, BL-4 and BL-6 during growth on lactate. 75 mM lactic acid was added to each bottle on day 7. Mean values of six measurements of duplicate batch cultures are given and error bars represent the standard deviation.



Component	Families	Genes
Pan-genome	6654	9508
Core-genome	504	2064
Variable genome	6150	7444

	Pan-CDSs	Core-CDSs	Variable CDSs	Strain-specific CDSs
BL-3	3778	723	3055	2714
BL-4	2290	623	1667	1142
BL-6	3440	718	2722	1998

Figure 2

Genomic heterogeneity of strains BL-3, BL-4 and BL-6. Venn diagram showing the shared and unique gene families of the three isolates, and numbers of CDSs presenting the pan-genome and core-genome as well as variable and strain-specific genes. Families refer to the MicroScope homologous gene families (MICFAM), in which the protein-coding genes share at least 80% amino acid sequence identity and 80% alignment coverage.

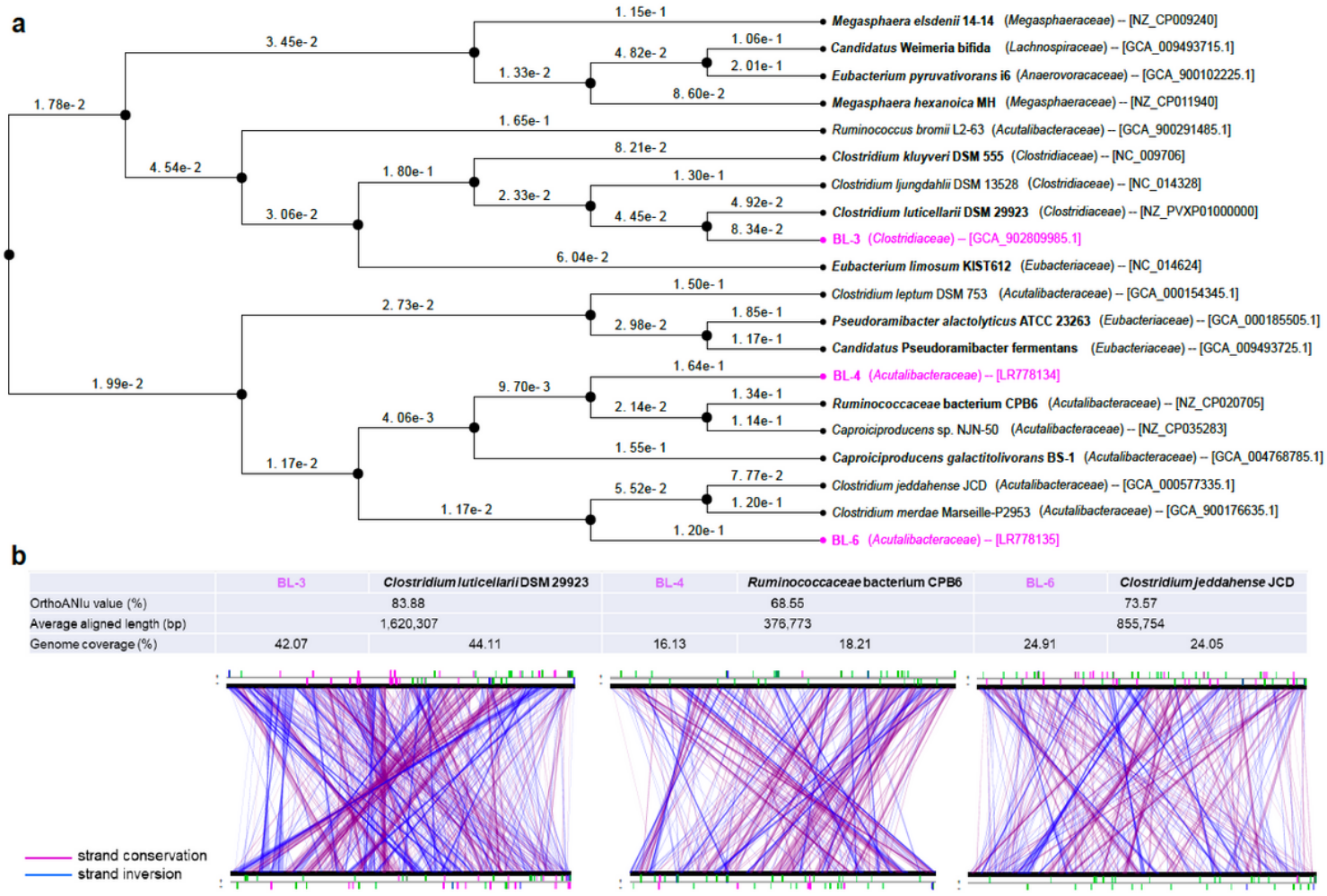
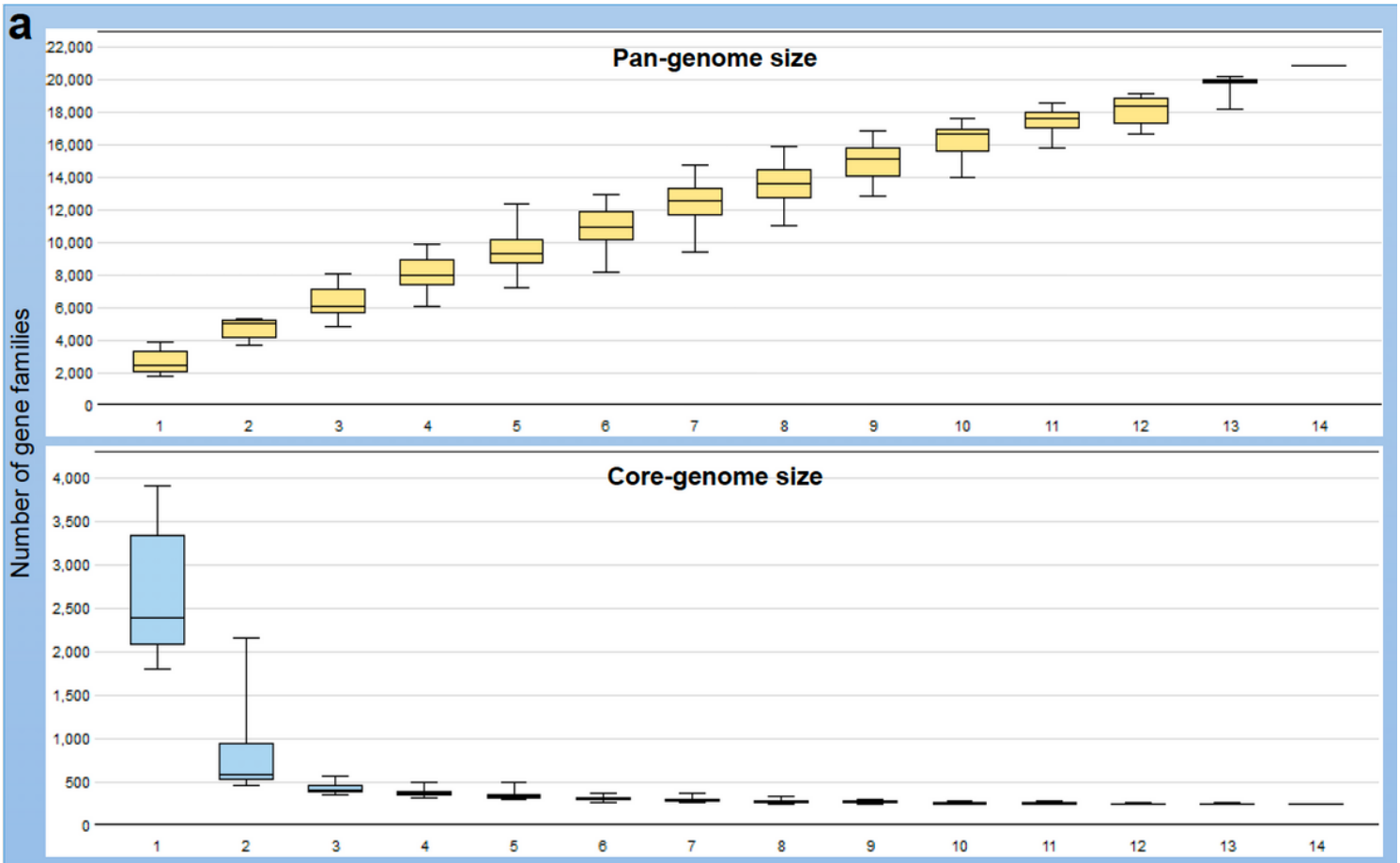


Figure 3

Phylogenomic analysis of the three isolates. (a) Neighbor-joining tree showing the genome similarity between 14 chain-elongating bacterial strains. The newly isolated strains are highlighted in pink and all experimentally validated chain-elongating strains are indicated in bold. Additional related species based on 16S rRNA phylogenetic analysis were included (see the phylogenetic tree in Additional file 1, Fig. S3). GTDB taxonomic assignments at the family level are shown in parentheses. The accession numbers of the genomes are shown in brackets. Distances indicated at the branches correlate to the average nucleotide identity (ANI) according to: D ≈ 1-ANI. (b) USEARCH OrthoANI comparison for strains BL-3, BL-4 and BL-6 to related genomes. The line plots give an overview of the conservation of synteny groups on nucleotide level. Strand conservations are depicted in purple and strand inversions in blue. The synton size corresponds to at least three genes.



b

Strain	Pan-CDSs	Core-CDSs	Variable CDSs	Strain-specific CDSs
BL-3	3778	390	3388	922
BL-4	2290	322	1968	934
BL-6	3440	392	3048	1546
Megasphaera elsdenii 14-14	2319	306	2013	647
Ruminococcaceae bacterium CPB6	2076	302	1774	696
Megasphaera hexanoica MH	2762	334	2428	1039
Pseudoramibacter alactolyticus ATCC 23263	2290	319	1971	938
Candidatus Pseudoramibacter fermentans	2166	297	1869	808
Clostridium kluyveri DSM 555	4288	392	3896	287
Caproiciproducens galactitolivorans BS-1	2502	334	2168	945
Eubacterium limosum KIST612	4535	388	4147	2864
Eubacterium pyruvativorans i6	1905	293	1612	1016
Candidatus Weimeria bifida	2439	308	2131	1462
Clostridium luticellarii DSM 29923	3790	395	3395	905

Figure 4

Pan-genome analysis of the 14 chain-elongating bacterial strains. (a) Pan-genome and core-genome sizes and their changes for the increasing genome set. Families refer to the MicroScope homologous gene families (MICFAM), in which the protein-coding genes share at 80% of amino acid sequence identity and 80% of alignment coverage. (b) Summary of gene counts for each strain. CDS: gene coding sequence

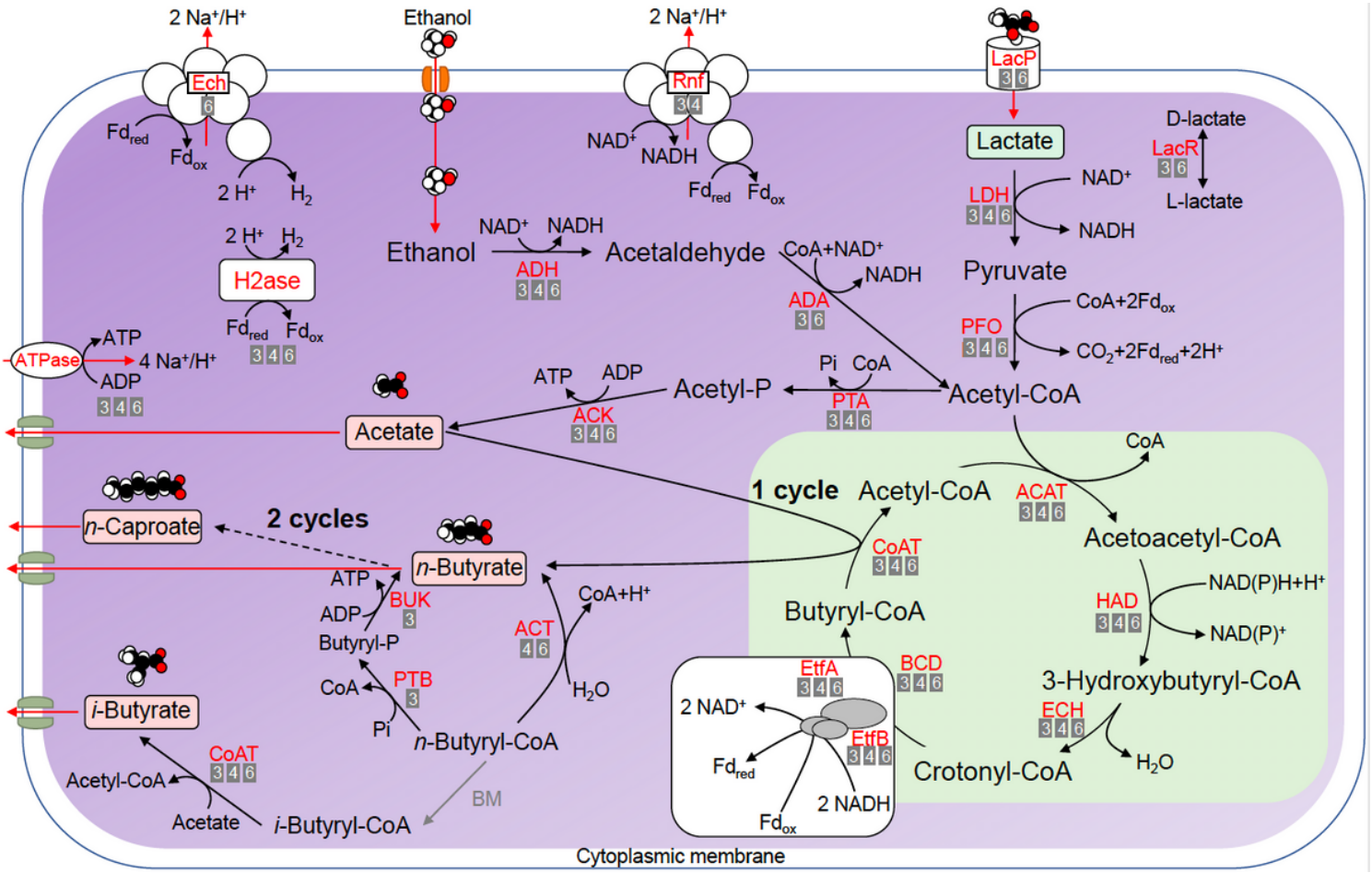


Figure 5

Metabolic pathways involved in lactate fermentation and chain elongation. Pathways of the formation of acetyl-CoA from ethanol were also included. Enzyme abbreviations (see Table 2 for full names) are provided in red letters next to the pathways (solid lines). The numbers below the enzyme names indicate which strains were predicted to harbor the corresponding CDSs, i.e. “3” refers to strain BL-3, “4” refers to strain BL-4 and “6” refers to strain BL-6. The hypothetical pathway related to the formation of iso-butyrate is shown as gray line, and the enzyme that was not found to be encoded in the genomes is indicated in gray. The dashed line represents multi-enzyme reactions between the two indicated molecules, and “cycle” refers to the reverse β-oxidation cycle.

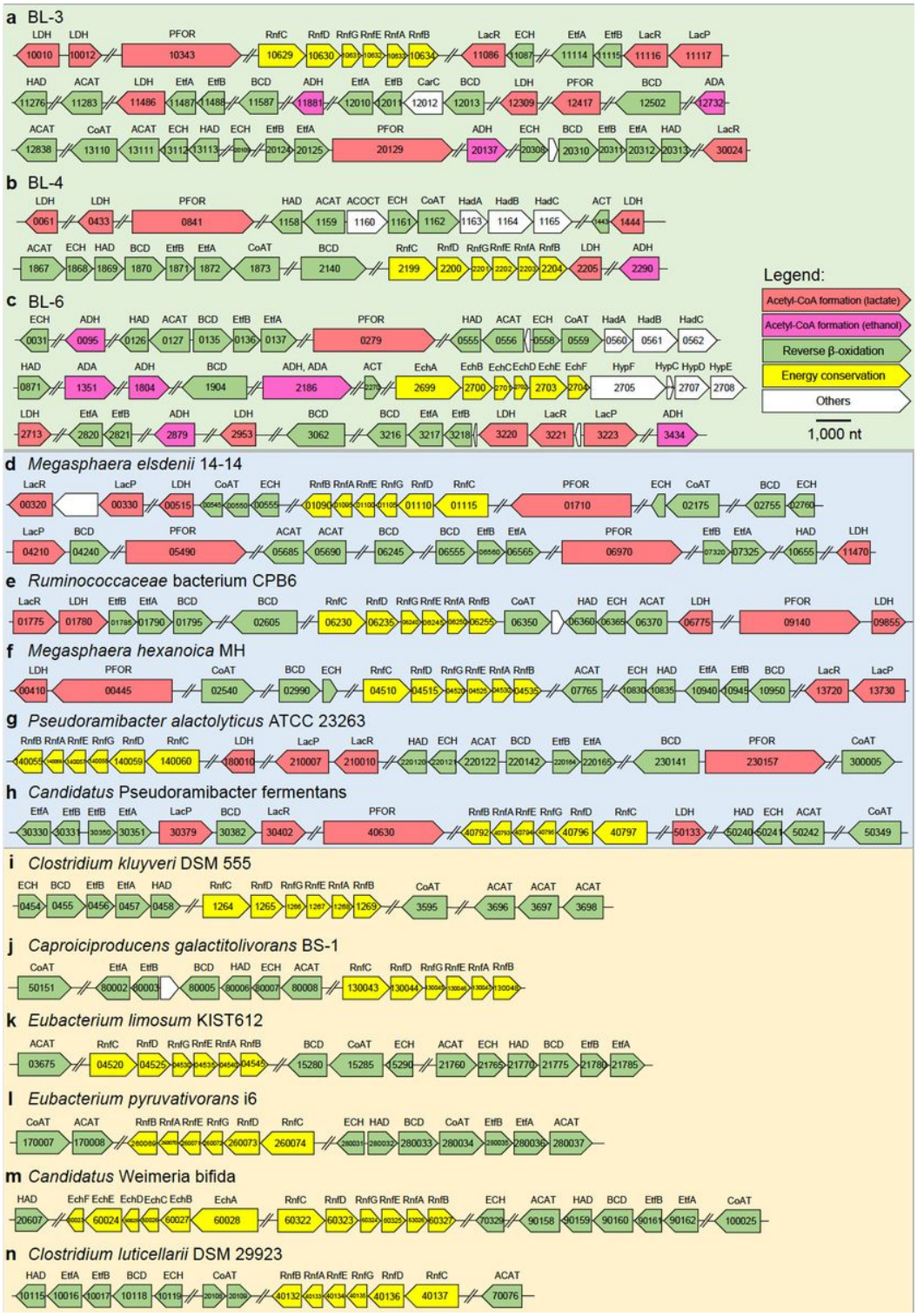


Figure 6

Arrangement of predicted CDSs in genomes of strains BL 3 (a), BL-4 (b), BL-6 (c), other bacterial strains reported of chain elongation with lactate (d-h), and with other reduced substrates (i-n). Numbers in the arrows denote the corresponding CDS labels. Abbreviations above the arrow refer to the enzyme names (see Table 2 for full names). Scale bar: 1,000 nucleotides (nt)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile13.xlsx
- Additionalfile12.xlsx
- Additionalfile11.xlsx
- Additionalfile10.xlsx
- Additionalfile9.xlsx
- Additionalfile8.xlsx
- Additionalfile7.xlsx
- Additionalfile6.xlsx
- Additionalfile5.xlsx
- Additionalfile4.xlsx
- Additionalfile3.xlsx
- Additionalfile2.xlsx
- Additionalfile1.docx