UNIVERSITÄT LEIPZIG

APPLIED MICROBIAL ECOLOGY OF ANAEROBIC REACTOR
MICROBIOMES

Von der Fakultät für Lebenswissenschaften

der Universität Leipzig

genehmigte

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

vorgelegt

von M. Sc. **Bin Liu**

geboren am 02. Oktober 1991 in Jilin, China

Dekan: Prof. Dr. Marc Schönwiesner

Gutachter: Prof. Dr. Hauke Harms
                          Prof. Dr. Diana Z. Sousa

Tag der Verteidigung: 19. February 2021

BIBLIOGRAPHISCHE DARSTELLUNG

Bin Liu

APPLIED MICROBIAL ECOLOGY OF ANAEROBIC REACTOR MICROBIOMES

Fakultät für Lebenswissenschaften

Universität Leipzig

Dissertation

224 Seiten, 333 Literaturangaben, 63 Abbildungen, 24 Tabellen

---

The carboxylate platform is an alternative way of anaerobic digestion to valorise organic wastes or biomass residues, in the form of medium-chain carboxylates (e.g., *n*-caproate and *n*-caprylate). Some anaerobic bacteria are known to produce medium-chain carboxylates through microbial chain elongation. Mixed culture fermentation is commonly employed for the chain elongation processes. However, a systematic investigation of the metabolism and ecological interactions of the chain elongation communities was missing so far. This thesis focuses on the chain elongation communities in closed model ecosystems. First, a model ecosystem was developed by operating a continuous bioreactor with an enriched mixed culture. During long-term reactor operation under constant conditions, the results suggested that the chain-elongating bacteria were outcompeted by butyrate-producing bacteria, leading to the increase of butyrate yield at the cost of *n*-caproate and *n*-caprylate yields. Second, effects of environmental manipulations on chain elongation community assembly and functioning were investigated in the model ecosystems. Shortening the hydraulic retention time shaped the communities towards higher *n*-caproate and *n*-caprylate productivities, which were accurately predicted by using machine learning. The developed machine learning framework to quantitatively predict process performance is transferable to other microbial systems. Increasing pH induced dramatic shifts in the community assembly but exhibited no strong effects on medium-chain carboxylate yields. High functional redundancy was indicated despite the reactors being long-term closed systems. Last, three novel chain-elongating species were isolated, which can convert lactate to *n*-caproate and *iso*-butyrate. The shared metabolic features and genomic diversity of the entire repertoire of chain-elongating species were indicated. In summary, this thesis sparkles new insights into the relationship between chain elongation community diversity and functioning, and it extends the metabolic knowledge of chain elongation bacteria.

"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong. In that simple statement is the key to science".

(Richard Feynman, 1964)

Contents

Abbreviations

ACAT	Acetoacetyl-CoA acetyltransferase (EC 2.3.1.9, EC 2.3.1.16)
ACOCT	Acetyl-CoA:oxalate CoA-transferase (EC 2.8.3.19)
ACP	Acyl carrier protein
ACT	Acyl-CoA thioesterase (EC 3.1.2.20)
ADA	Acetaldehyde dehydrogenase (EC 1.2.1.10)
ADH	Alcohol dehydrogenase (EC 1.1.1.1)
ANI	Average nucleotide identity
ARISA	Automated ribosomal intergenic spacer analysis
ASV	Amplicon sequence variant
AUC	Area under the curve
BCD	Butyryl-CoA dehydrogenase (EC 1.3.8.1)
BLAST	Basic local alignment search tool
BM	Butyryl-CoA:isobutyryl-CoA mutase (EC 5.4.99.13)
BUK	Butyrate kinase (EC 2.7.2.7)
C2	Acetate
C4	*n*-Butyrate
C6	*n*-Caproate
C8	*n*-Caprylate
CDSs	Coding sequences
CE	Chain elongation
CoA	Coenzyme A
CoAT	CoA transferase (EC 2.8.3.-)
COD	Chemical oxygen demand
COG	Clusters of orthologous groups
$^1D/^2D$	Diversity of order one/Diversity of order two
DGGE	Denaturing gradient gel electrophoresis

$^1E/^2E$	Evenness of order one/Evenness of order two
ECH	Enoyl-CoA hydratase (EC 4.2.1.150, EC 4.2.1.55)
EGGNOG	Evolutionary genealogy of genes: non-supervised orthologous groups
ENA	European nucleotide archive
FDR	False discovery rate
GC	Gas chromatography
GTDB	Genome Taxonomy Database
HAD	3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.157, EC 1.1.1.35)
HIB	Hydroxy-isobutyrate
HPLC	High performance liquid chromatography
HRT	Hydraulic retention time
InterPro	Integrative protein signature database
LAB	Lactic acid bacteria
LacP	Lactate permease (TC 2.A.14)
LDH	Lactate dehydrogenase (EC 1.1.1.27)
LME	Linear mixed-effects models
MAGs	Metagenome-assembled genomes
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
MCCs	Medium-chain carboxylates
MICFAM	MicroScope homologous gene families
MMS	Methylmalonate-semialdehyde
MTV-LMM	Microbial Temporal Variability Linear Mixed Model
NCBI	National center for biotechnology information
NGS	Next-generation sequencing
NMDS	Non-metric multidimensional scaling
OD	Optical density
ONT	Oxford Nanopore Technologies
OrthoANIu	Average nucleotide identity by orthology with USEARCH

PCA	Principle component analysis
PCoA	Principle coordinate analysis
PCR	Polymerase chain reactions
PERMANOVA	Permutational multivariate analysis of variance
PFOR	Pyruvate ferredoxin oxidoreductase (EC 1.2.7.1)
PkGDB	Prokaryotic genome database
PTB	Phosphate butyryltransferase (EC 2.3.1.19)
RDA	Redundancy analysis
RRMSE	Relative root mean square error
rRNA	Ribosomal RNA
ROC curve	Receiver operating characteristic curve
SCCs	Short-chain carboxylates
SSCP	Single strand conformation polymorphism
TAN	Total ammonia nitrogen
TGGE	Temperature gradient gel electrophoresis
TrEMBL	Translated EMBL
TRFLP	Terminal restriction fragment length polymorphism
T-RFs	Terminal restriction fragments
VPA	Variation partitioning analysis
WGS	Whole-genome sequencing
XL	Xylanase (EC 3.2.1.8)
XOS	Xylooligosaccharides
XylT	Xylose transporter (EC 7.5.2.10, EC 7.5.2.13)

List of figures

List of tables

Zusammenfassung

Offene Mischkulturen in anaeroben Reaktorsystemen wandeln organische Abfälle oder Biomasse-Rückstände in hauptsächlich kurzkettige Carboxylate mit zwei bis fünf Kohlenstoffatomen um. Kurzkettige Carboxylate können von methanogenen Konsortien in der anaeroben Vergärung in das stark reduzierte Endprodukt Methan umgewandelt werden. Die mikrobielle Kettenverlängerung, z.B. über die reverse β-Oxidation, ist eine alternative Elektronensenke in solchen anaeroben Reaktormikrobiota. In natürlichen Ökosystemen wie dem Pansen sind bestimmte anaerobe Bakterien bekannt, die mittelkettige Carboxylate (z.B. *n*-Caproat und *n*-Caprylat) durch reverse β-Oxidation produzieren. Die Carboxylat-Plattform zielt darauf ab, Kohlenstoff aus Abfallströmen oder Biomasse-Rückständen durch anaerobe Fermentation in Form von mittelkettigen Carboxylaten zurückzugewinnen. Sie bietet die Möglichkeit Chemikalien zu ersetzen, die aus nicht nachhaltigen Quellen wie fossilen Rohstoffen stammen.

Für Verfahren der mikrobiellen Kettenverlängerung wird üblicherweise die Mischkulturfermentation eingesetzt. Die verschiedenen mikrobiellen Gemeinschaften enthalten unterschiedliche funktionelle Gruppen, die an der Hydrolyse und Fermentation der verfügbaren organischen Verbindungen sowie an der Konversion von Intermediaten in mittelkettige Carboxylate beteiligt sind. Im Allgemeinen sind die zugrunde liegenden Stoffwechselwege und ökologischen Wechselwirkungen der mikrobiellen Gemeinschaften nicht gut verstanden. Diese Doktorarbeit behandelt die mikrobielle Kettenverlängerung mit Laktat, insbesondere den Metabolismus und die ökologischen Wechselwirkungen der beteiligten mikrobiellen Gemeinschaften in geschlossenen Modell-Ökosystemen, d.h. in anaeroben Bioreaktoren.

Das erste Kapitel beschreibt die Entwicklung eines Modell-Ökosystems mit reduzierter Komplexität, indem Laktat und Xylan als definierte Kohlenstoffquellen verwendet wurden, um die Substratbedingungen bei der anaeroben Fermentation von Maissilage zu Caproat zu simulieren. Die Fütterung mit definierten Kohlenstoffquellen ermöglichte die Bilanzierung der

Elektronen- und Kohlenstoffflüsse. Durch die Vermeidung einer Kintinuierlichen Inokulation konnten der Metabolismus und die Dynamik dieser vereinfachten mikrobiellen Gemeinschaft gezielter untersucht werden als in offenen Reaktorsystemen. Während eines Langzeit-Reaktorversuchs wurden vier aufeinanderfolgende Stadien beobachtet: Anpassungsphase, Stadium I (Periode hoher Produktivität von mittelkettigen Carboxylaten), Übergangsphase und Stadium II (Periode hoher Butyrat-Produktion). Auf der Basis von 16S rRNA-Amplikonsequenzen und Korrelationen mit Prozessparametern wurden Netzwerkanalysen durchgeführt, um auf potenzielle Stoffwechselfunktionen der beteiligten Spezies und deren ökologische Interaktionen zu schließen. Die Ergebnisse legten nahe, dass der Prozess verschiedene Funktionen der Xylan-Hydrolyse, der Xylose-Fermentation und der Kettenverlängerung mit Laktat als Elektronendonor umfasste. Die geschlussfolgerten Interaktionen, wie die Kooperation zwischen Milchsäurebakterien und kettenverlängernden Bakterien sowie die Konkurrenz zwischen Bakterien, die mittelkettige Carboxylate bilden, und butyratproduzierenden Bakterien, erklärten die Entwicklung der mikrobiellen Gemeinschaft über vier Sukzessionsstadien. In diesem geschlossenen Modell-Ökosystem wurden die kettenverlängernden Bakterien unter konstanten Bedingungen von Butyrat-Produzenten auskonkurriert, was zu einer Erhöhung der Butyrat-Ausbeute auf Kosten der *n*-Caproat- und *n*-Caprylat-Ausbeuten führte.

Im zweiten Abschnitt wurden die Auswirkungen einer Verkürzung der hydraulischen Verweilzeit auf den Aufbau und die Funktion der mikrobiellen Gemeinschaften in den Modell-Ökosystemen getestet, mit dem Ziel, ökophysiologische Funktionen im Hinblick auf die Prozessleistung quantitativ vorherzusagen. Durch die Verkürzung der hydraulischen Verweilzeit von 8 Tagen auf 2 Tage in zwei kontinuierlichen Reaktoren wurden höhere Produktivitäten und Ausbeuten von *n*-Caproat und *n*-Caprylat erreicht. Für die Vorhersage wurde ein Modell durch Anwendung des *Random-Forest*-Ansatzes unter Verwendung von 16S rRNA-Amplikon-Sequenzierungsdaten erstellt. Es wurde eine Genauigkeit von mehr als 90% bei der quantitativen Vorhersage der *n*-Caproat- und *n*-Caprylat-Produktivität erreicht. Vier abgeleitete

Bioindikatoren, die zu den Gattungen *Olsenella*, *Lactobacillus*, *Syntrophococcus* und *Clostridium* IV gehören, deuteten auf deren Relevanz für die höhere Carboxylatproduktivität bei kürzerer hydraulischer Verweilzeit hin. Die Rekonstruktion von Populationsgenomen dieser Bioindikatoren aus Metagenomdaten bestätigte das genetische Potenzial dieser Indikatorspezies, Schlüsselschritte der Carboxylatproduktion durchzuführen. Außerdem wurde eine funktionelle Redundanz bei der Konversion von Xylan und Laktat zu *n*-Butyrat, *n*-Caproat und *n*-Caprylat beobachtet, wobei die relevanten Bioindikatoren in ihrer relativen Abundanz zunahmen. Dies zeigt, dass die beteiligten Stoffwechselwege stark an die Verkürzung der hydraulischen Verweilzeit gekoppelt waren. Im Allgemeinen ist das entwickelte Konzept des maschinellen Lernens zur Identifizierung von Bioindikatoren und zur quantitativen Vorhersage der Prozessleistung auf andere Ökosystemprozesse und mikrobielle Systeme übertragbar, bei denen die Gemeinschaftsdynamik mit Schlüsselfunktionen verknüpft ist.

Das dritte Kapitel beschreibt Untersuchungen zu Auswirkungen des pH-Werts auf den Aufbau und die Funktionsweise der mikrobiellen Gemeinschaften auf der Grundlage der entwickelten Modell-Ökosysteme. Der Anstieg des pH-Wertes von 5,5 auf 6,0 verursachte Schwankungen in den Erträgen von *n*-Butyrat, *n*-Caproat und *n*-Caprylat. Nach der pH-Störung kehrten die Carboxylat-Ausbeuten zu den vorherigen Werten zurück, während sich die mikrobiellen Gemeinschaften in ihrer Zusammensetzung änderten, was als Abnahme der Diversität und Äquität (*Evenness*) und Zunahme des Artenreichtums (*Richness*) beobachtet wurde. Einige zuvor seltene Taxa wurden dominant, was starke selektive Effekte des pH-Werts auf die mikrobielle Diversität widerspiegelt. Durch Anwendung von Aitchison PCA-Clustering, linearen Mischwirkungsmodellen und *Random-Forest*-Klassifizierung wurden die unterschiedlichen pH-Präferenzen der potenziellen Kettenverlängerer *Clostridium* IV und *Clostridium sensu stricto* identifiziert. *Co-occurrence*-Netzwerke für verschiedene pH-Werte zeigten, dass die mit dem Kettenverlängerer *Clostridium* IV kooperierenden Milchsäurebakterien infolge des pH-Anstiegs von *Olsenella* zu *Lactobacillus* wechselten, was die Plastizität des Nahrungsnetzes der mikrobiellen Gemeinschaften zeigt. Im Vergleich zu

den oben genannten Ergebnissen zur Verkürzung der hydraulischen Verweilzeit führte der pH-Anstieg zu dramatischen Verschiebungen in der Zusammensetzung der mikrobiellen Gemeinschaften, zeigte aber keine starken Auswirkungen auf die metabolischen Funktionen im Hinblick auf die Produktion von Carboxylaten mittlerer Kettenlänge. Es zeigte sich eine hohe funktionelle Redundanz, obwohl es sich bei den Reaktoren um langfristig geschlossene Systeme handelte.

Parallel zu den Reaktorexperimenten wurden Reinkulturen von kettenverlängernden Clostridienstämmen isoliert, die drei neue Spezies repräsentieren. Ihre Genome wurden mit Hilfe eines hybriden *Short-* und *Long-Read*-Sequenzierungsverfahrens rekonstruiert. Die drei neuartigen Stämme produzierten *n*-Caproat, *n*-Butyrat, *Iso*-Butyrat und Acetat aus Laktat in Batch-Kultivierung bei pH 5,5. Der genetische Hintergrund der Kettenverlängerung mit Laktat wurde analysiert, wobei die CoA-Transferase als terminales Enzym der reversen β-Oxidation identifiziert wurde. Die Genome weisen eine beträchtliche genetische Heterogenität auf, enthalten jedoch hochkonservierte Gene, die an der Laktatoxidation, der reversen β-Oxidation, der Wasserstoffbildung und einer von zwei Arten von Energieerhaltungssystemen (Rnf und Ech) beteiligt sind. Der genetische Hintergrund der Laktat-basierten Kettenverlängerung in diesen Isolaten und anderen experimentell validierten kettenverlängernden Stämmen wurde durch vergleichende Genomik analysiert. Es zeigte sich, dass das für die Kettenverlängerung spezifische Kerngenom die Wege für die reverse β-Oxidation, Wasserstoffbildung und Energieerhaltung kodiert und gleichzeitig kettenverlängernde Spezies eine beträchtliche Genomheterogenität aufweisen. Weitere Experimente sind erforderlich, um die Stoffwechselwege für die *Iso*-Butyrat-Bildung in diesen Stämmen aufzuklären.

Zusammenfassend lässt sich sagen, dass Modellgemeinschaften für die mikrobielle Kettenverlängerung durch Wechsel von pH-Wert und hydraulischer Verweilzeit in Langzeit-Bioreaktorversuchen angereichert und weiter geformt wurden. Der Metabolismus und die ökologischen Interaktionen der Mikroorganismen, die an der mikrobiellen Kettenverlängerung mit Laktat beteiligt sind, wurden mit Hilfe von 16S rRNA-Amplikonsequenzierung und

Metagenomik in Verbindung mit Netzwerkanalyse, statistischer Modellierung und maschinellem Lernen aufgeklärt, was auch zu neuen Erkenntnissen über die Beziehung zwischen der Diversität der mikrobiellen Gemeinschaften und ihrer Funktionsweise führte. Die Isolierung neuer Spezies erweitert unser Wissen über den Metabolismus der mikrobiellen Kettenverlängerung. Schließlich ist ein besseres Verständnis der Mechanismen, die der Etablierung von Reaktoremikrobiomen zugrunde liegen, der Schlüssel zur Weiterentwicklung von auf Mikrobiota basierenden Biotechnologien.

Summary

Open cultures of anaerobic reactor systems convert organic wastes or biomass residues into mainly short-chain carboxylates with two to five carbon atoms. The short-chain carboxylates can be converted into the highly reduced end product methane by methanogenic consortia in anaerobic digestion. Microbial chain elongation such as via the reverse β-oxidation pathway was found as an alternative electron sink with the same anaerobic reactor microbiota. In natural ecosystems such as rumen microbial ecosystem, some anaerobic bacteria are known to produce medium-chain carboxylates (e.g., *n*-caproate and *n*-caprylate) through reverse β-oxidation. The carboxylate platform aims to recover carbon from waste streams or biomass residues by anaerobic fermentation in the form of medium-chain carboxylates. It has created great opportunities to replace chemicals derived from non-sustainable sources such as fossil feedstock.

Mixed culture fermentation is commonly employed for the chain elongation processes. The diverse microbial chain elongation communities contain different functional groups involved in the processes of hydrolysis and fermentation of available organic compounds as well as the conversion of intermediates to medium-chain carboxylates. In general, the underlying metabolism and ecological interactions of the chain elongation communities are not well understood. This PhD thesis centres on the metabolism and ecological interactions in closed model ecosystems (i.e., anaerobic bioreactors) involved in microbial chain elongation with lactate.

In the first chapter, a model ecosystem with reduced complexity was developed by using lactate and xylan as defined carbon sources to simulate the feedstock conditions of caproate-producing bioreactors operated with corn silage. Feeding defined carbon sources enabled balancing of electron and carbon flows. By preventing continuous inoculation, the simplified community of enrichment cultures allowed to study the metabolic and community dynamics in a clearer manner than open reactor systems. During a long-term reactor experiment, four succession stages including adaptation, stage I (high medium-chain carboxylate-producing period), transition and stage II (high butyrate-

producing period) were observed. Co-occurrence networks of species based on 16S rRNA amplicon sequences and associations with process parameters were analysed to infer potential metabolic functions and microbial interactions. The results suggested that the process included diverse functions of xylan hydrolysis, xylose fermentation and chain elongation with lactate as electron donor. The inferred interactions such as cooperation between lactic acid bacteria and chain-elongating bacteria, as well as competition between medium-chain carboxylate-producing bacteria and butyrate-producing bacteria, resulted in the community development over four succession stages. In this closed model ecosystem, the chain-elongating bacteria were outcompeted by butyrate-producing bacteria under constant conditions, leading to the increase of butyrate yield at the cost of *n*-caproate and *n*-caprylate yields.

The second chapter tested the effects of shortening the hydraulic retention time on the community assembly and functioning in the model ecosystems, aiming to quantitatively predict ecophysiological functions of the microbial communities. For the process performance, higher productivities and yields of *n*-caproate and *n*-caprylate were achieved by reducing the hydraulic retention time from 8 days to 2 days in two continuous reactors. A predictive model was generated by applying the random forest approach using 16S rRNA amplicon sequencing data. More than 90% accuracy in the quantitative prediction of *n*-caproate and *n*-caprylate productivities was achieved. Four inferred bioindicators belonging to the genera *Olsenella*, *Lactobacillus*, *Syntrophococcus* and *Clostridium* IV suggested their relevance to the higher carboxylate productivity at shorter hydraulic retention time. Combined with metagenomics, the recovery of metagenome-assembled genomes of these bioindicators confirmed their genetic potential to perform key steps of carboxylate production. Besides, functional redundancy in the conversion of xylan and lactate to *n*-butyrate, *n*-caproate and *n*-caprylate was revealed, with the relevant bioindicators increasing in relative abundance. Thus, the involved metabolic pathways were strongly coupled to the decrease in hydraulic retention time. In general, the developed machine learning framework to identify bioindicators and to quantitatively predict process performance is transferable to other ecosystem

processes and microbial systems where community dynamics is linked to key functions.

In the third chapter, the effects of pH increase on the chain elongation community assembly and functioning were tested based on the developed model ecosystems. The increase in pH from 5.5 to 6.0 caused fluctuations in the yields of *n*-butyrate, *n*-caproate and *n*-caprylate. After the pH disturbance, the carboxylate yields returned to the previous values while the communities developed to a different state, observed as decrease in diversity and evenness and increase in richness. Some taxa shifted from rare to abundant, reflecting strong selective effects of lower pH values. By applying Aitchison PCA clustering, linear mixed effect models and random forest classification, the different pH preferences of the potential chain elongators *Clostridium* IV and *Clostridium sensu stricto* were identified. By constructing networks for different pH levels, the cooperation of the chain elongator *Clostridium* IV with lactic acid bacteria switches from *Olsenella* to *Lactobacillus* along the pH increase, revealing the plasticity of the food web of chain elongation communities. Compared with the previously observed results of decreasing the hydraulic retention time, pH increase induced dramatic shifts in the community assembly but exhibited no strong effects on community functioning in terms of medium-chain carboxylate production. High functional redundancy was indicated despite the reactors being long-term closed systems.

In parallel to the reactor experiments, pure cultures of chain-elongating clostridial strains were isolated, representing three novel species. Their genomes were assembled using a hybrid short and long read sequencing approach. The three novel strains produced *n*-caproate, *n*-butyrate, *iso*-butyrate and acetate from lactate in batch cultivation at pH 5.5, with the confirmation of their genetic background of lactate-based chain elongation and using CoA transferase as the terminal enzyme. Their genomes show substantial genetic heterogeneity but contain highly conserved genes involved in lactate oxidation, reverse β-oxidation, hydrogen formation and either of two types of energy conservation systems (Rnf and Ech). The genetic background of lactate-based chain elongation in these isolates and other experimentally

validated chain-elongating strains was analysed by comparative genomics. The chain elongation-specific core-genome was indicated to encode the pathways for reverse β-oxidation, hydrogen formation and energy conservation while chain-elongating species displayed substantial genome heterogeneity. Further research is needed to elucidate the pathways for *iso*-butyrate formation in these strains.

In summary, model communities of chain elongation processes were enriched and further shaped by alternations of pH and hydraulic retention time in long-term bioreactor experiments. The metabolism and ecological interactions of reactor microbiota involved in microbial chain elongation with lactate were elucidated by using 16S rRNA amplicon sequencing and metagenomics coupled to network analysis, statistical modelling and machine learning, which also sparkled new insights into the relationship between microbial chain elongation community diversity and functioning. The isolation of novel chain-elongating species further expands our knowledge on the metabolism of chain elongation bacteria. Finally, a better understanding of the rules governing community assembly is key to accelerate the development of microbiota-based biotechnologies.

1 Introduction

1.1 Reactor microbiota

As proposed, the term microbiota refers to the assemblage of microorganisms present in a defined environment (Marchesi and Ravel, 2015). Reactor microbiota used in this study refers to the biologically active environment of reactors, represented as open cultures of anaerobic microbial consortia. Employing microbial communities in bioreactors provides the opportunity to recover valuable resources from organic wastes or biomass residues. Within the reactor microbiota, diverse microbial species may cooperate or compete with each other, presenting a broad metabolic capacity for the utilisation of complex substrates. Understanding the metabolism and ecological interactions of reactor microbiota, and designing effective control strategies for key players hold promise to engineer communities towards the desired bioprocesses, providing a window for potential applications of microbiota-based biotechnologies (Verstraete et al., 2007; Koch et al., 2014; Lawson et al., 2019).

1.2 Carboxylate platform

Using reactor microbiota for producing biogas (mainly methane) in anaerobic digestion is a mature technology that has been developed and widely implemented for decades all over the world (Angenent et al., 2016). While it is a well-known waste-to-energy process, the economics of anaerobic digestion are less favourable due to the low value of biogas (Agler et al., 2011). The carboxylate platform is an alternative way to recover carbon from wastes with anaerobic fermenters, in the form of medium-chain carboxylates (MCCs). Hereafter, their dissociated and undissociated forms are subsumed as carboxylates. MCCs are defined as monocarboxylates of six to twelve carbon atoms, here mostly focusing on *n*-caproate (C6) and *n*-caprylate (C8). MCCs are attractive bio-based products in the context of biorefinery platforms. Currently, MCCs are mainly produced from vegetable oils (e.g., refinery of coconut and palm kernel oils). Due to the limited supply and the high demands in industry, they have relatively high values in the market. The reported prices

of C6 and C8 are 2,880 and 5,060 US dollars per metric ton in 2018, respectively (Scarborough, 2019). Additionally, the solubility of the carboxylate in water decreases along with the increase in carbon chain length. This means the longer-chain carboxylates are easier to extract from water due to their hydrophobic carbon-chains (Angenent et al., 2016). Therefore, biomanufacturing valuable chemicals such as MCCs in a sustainable way can meet the industrial needs and contribute to further evolve current production schemes into a circular economy.

1.3 Microbial chain elongation

The communities of anaerobic digestion contain different trophic groups involved in four main processes: hydrolysis, acidogenesis, acetogenesis and methanogenesis. Short-chain carboxylates (SCCs, two to five carbon atoms) are key intermediates in the conversion of organic wastes or biomass residues to biogas, resulting from the main fermentation of hydrolysis products (Agler et al., 2011). Strategies for inhibiting methanogenesis include the operation at acidic pH in the range of 5.0 to 6.5 and addition of chemical inhibitors (e.g., 2-bromoethanosulfonic acid), which can promote carboxylate production. In anaerobic bacteria, the formation of MCCs from SCCs is known as a metabolic process termed microbial chain elongation (CE). It has been described since the early 1940s, when *Clostridium kluyveri* (Barker et al., 1945) was isolated and characterised for the production of C6. Some specialists in the anaerobic communities are involved in the elongation of fermentation products such as acetate (C2) or *n*-butyrate (C4) to MCCs via reverse β-oxidation. In reverse β-oxidation, acyl-CoA molecules are condensed and reduced, thereby increasing the chain length of carboxylates by two carbon atoms within each cycle (Spirito et al., 2014). As shown in Figure 1.1, initiated with acetyl-CoA, C4, C6 and C8 are all potential end products. In the CE process, ethanol or lactate can serve as electron donor providing energy for coupling acetyl-CoA formation and elongating acyl-CoA units; thioesterase or CoA transferase can act as terminal enzymes resulting in carboxylate formation. Besides acetyl-CoA being proposed as a primary source for reverse β-oxidation, propionyl-CoA can also be elongated for the production of odd-chain products (e.g., *n*-valerate and *n*-

heptanoate) (Jeon et al., 2016). Branched MCCs were also reported to be produced in CE, such as the formation of *iso*-caproate via *iso*-butyrate elongation with ethanol (De Leeuw et al., 2019).
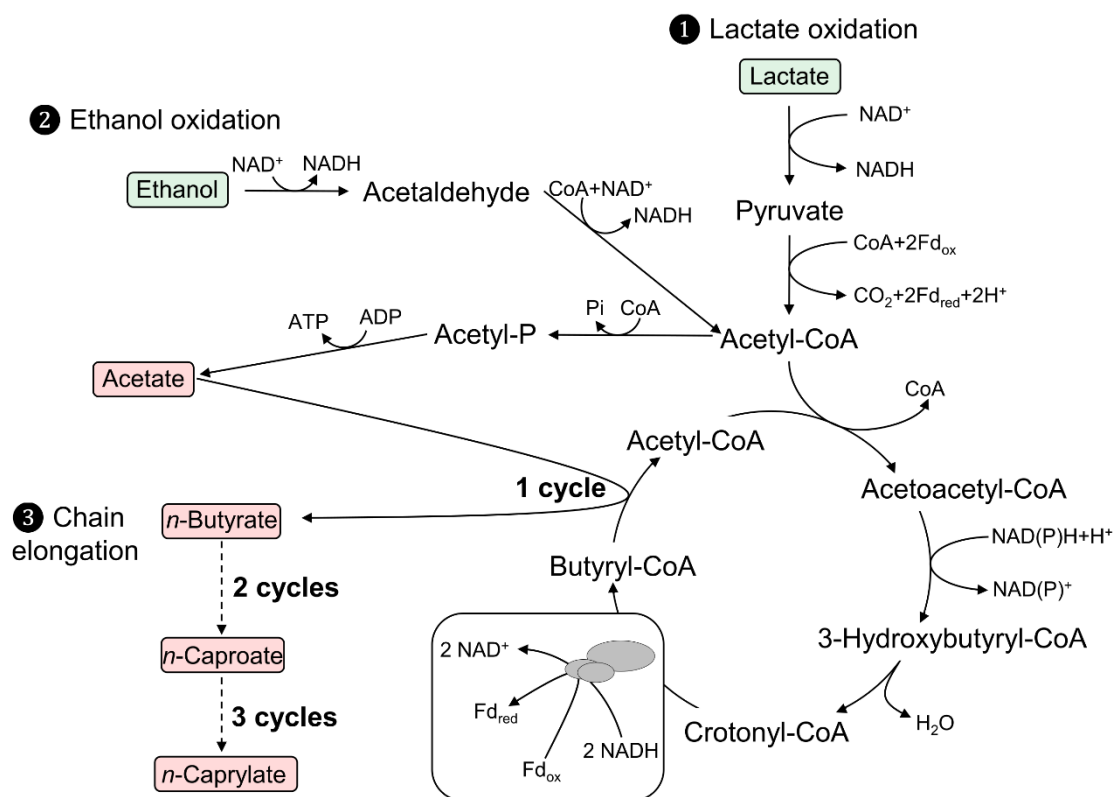
Figure 1.1 The metabolic pathways of reverse β-oxidation with lactate or ethanol as electron donor. Cycle refers to the reverse β-oxidation cycle; the dashed lines represent multi-enzyme reactions between the two indicated molecules.

Energy-rich, reduced molecules such as ethanol and lactate are suitable as electron donor to be oxidised to provide metabolic energy (ATP) via substrate level phosphorylation and reducing equivalents (NADH) for the reverse β-oxidation. Similar to CE with ethanol, CE with lactate can occur without the presence of short-chain carboxylates such as C2 (Zhu et al., 2015), although an extra addition of electron acceptors was reported to increase the production rate of C6 (Zhu et al., 2017). Monosaccharides can also serve as direct electron donors for CE. The chain-elongating bacteria using carbohydrates include: *Megasphaera elsdenii*, *Megasphaera hexanoica*, *Megasphaera indica*, *Pseudoramibacter alactolyticum*, *Caproiciproducens galactitolivorans* (D-

galactitol, sugar alcohol) and *Caproiciproducens* sp. 7D4C2 (Marounek et al., 1989; Willems and Collins, 1996; Kim et al., 2015; Jeon et al., 2017; Esquivel-Elizondo et al., 2020). However, the detailed mechanism of the sugar-driven CE needs to be further studied.

On the one hand, MCCs are the end products of chain-elongating species in the reactor microbiota. On the other hand, the toxicity of MCCs can inhibit the growth of many community members, possibly leading to the collapse of the production of MCCs (Liu et al., 2020a). MCCs cause pH-dependent microbial inhibition that can damage the integrity of bacterial cell membranes (Andersen et al., 2017). For example, *Escherichia coli* was described to be inhibited at an acidic pH of 4.3 whereas not inhibited at pH 7 in the presence of C6 and C8 (Yang et al., 2010; Royce et al., 2013). The inhibition needs to be managed because CE is often performed under acidic conditions to prevent methanogenesis. Effective ways include in-line extraction (pertraction), operation at neutral pH level and maintaining low concentrations of MCCs by attentive organic loading, which should be considered for a stable MCC production.

1.4 Methods for investigating reactor microbiota

A lot of methods have been developed and are still in development for gaining insights into the complexity of microbial communities. Briefly, these can be sorted into culture-independent and culture-dependent techniques. Here, a short overview of those methods employed in this thesis is given.

1.4.1 PCR-based methods

After the DNA extraction procedure, polymerase chain reactions (PCR) are commonly applied for the amplification of 16S rRNA genes (prokaryotic small subunit rRNA genes) or functional marker genes (e.g., *mcrA* for methanogens). Terminal restriction fragment length polymorphism (T-RFLP) is relatively simple to be applied to analyse the microbial community composition and dynamics. Suitable combination of primers and restriction enzymes can be chosen in order to obtain the best resolution at the desired taxonomic level (Talbot et al., 2008).

Other fingerprinting techniques such as denaturing gradient gel electrophoresis (DGGE), single strand conformation polymorphism (SSCP), temperature gradient gel electrophoresis (TGGE) and automated ribosomal intergenic spacer analysis (ARISA) were also commonly used. Fingerprinting cannot provide any sequence-dependent information, but only allows to quantify community diversity and compare community structure between different samples. Combined with cloning and Sanger sequencing of clone libraries, it is possible to identify some members (usually the most abundant ones) of the microbial community.

Next-generation sequencing (NGS) delivers the information of microbiome profiling by directly obtaining it from the presence and abundance of certain sequences, which is fundamentally different from that of fingerprinting. The amplicon NGS circumvents the cloning step that is needed to sequence amplicons based on Sanger technology. The first commercial NGS platform was released by Roche 454 in 2005; with new platforms emerged meanwhile, now the market is dominated by Illumina platforms. For example, Illumina MiSeq and HiSeq can target phylogenetic and functional marker genes, obtaining a higher depth in community analysis than fingerprinting techniques (Cabezas et al., 2015).

1.4.2 Metagenomics

Metagenome analysis provides more detailed genomic information and taxonomic resolution by sequencing all microbial genomes within a sample, and it allows the detection of functional genes and the construction of whole pathways (Knight et al., 2018). Given adequate sequencing depth, it may go deeply into strain-level resolution. Shotgun sequencing is commonly used to profile taxonomic composition and genetic potential of communities, and to recover whole genome sequences. The lack of reference genome data is a main issue of metagenome analyses, particularly for environmental samples with high microbial diversity (Quince et al., 2017). As for the technology itself, long-read sequencing (third generation sequencing, Figure 1.2) holds promise to significantly improve genome reconstruction from metagenomes. PacBio for *de novo* assembly and Oxford Nanopore Technologies (ONT) for portable

sequencing both may successfully occupy the commercial market. The hybrid genome assembly approach that combines the short-read accurate sequencing data (Illumina) with long-read less accurate (PacBio or ONT) sequencing data has shown its ability in the recovery of high-quality genomes (Liu et al., 2020b; Esquivel-Elizondo et al., 2020; Scarborough et al., 2020). At the moment, the complexity of computational analyses and further interpretation are limitations for metagenomics. Compared with 16S rRNA amplicon sequencing analysis, the cost of metagenome analysis including sequencing itself and training computational scientists to analyse the complex datasets is still high.
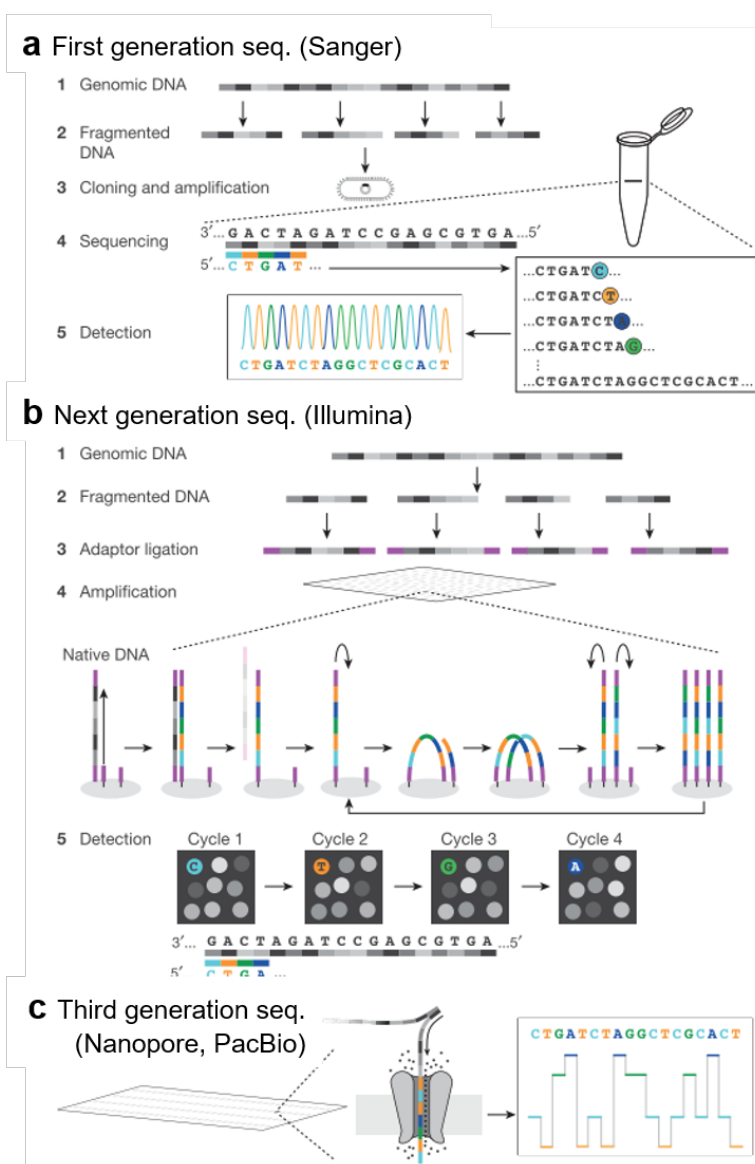
Figure 1.2 Schematic examples of first (**a**), second (**b**) and third (**c**) generation sequencing technologies. Figure was adapted from Shendure et al., 2017.

1.4.3 Culture-dependent methods

The high-throughput DNA-based sequencing approaches generated a countless number of sequences that have not been assigned to any known microorganism (Lagier et al., 2016). We microbiologists should realise that the field of microbial isolation cannot be completely abandoned, because pure cultures remain essential to elucidate the functions of those unknown microorganisms. The important physiological and biochemical features need to be studied by cultivation methods. For pure cultures, defined mixed cultures and self-assembled undefined mixed cultures, special efforts need to be made under anaerobic conditions using systems like batch, continuous, microtiter well plate format and the recently developed lab-on-a chip model (Vrancken et al., 2019). At the moment, it is still challenging to cultivate most of the microorganisms from natural environments under lab conditions, and isolating anaerobes is even more difficult. Culturomics is a promising culture-dependent method, because of the application of high-throughput culture conditions in the investigation of the microbial community, the use of matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) and 16S rRNA amplicon sequencing for the colony identification (Lagier et al., 2016). It can be expected that culturomics will bring an exponential increase of the microbial repertoire in the near future (Lagier et al., 2015). The identification of the rare species and new species might allow to extend our knowledge to levels equivalent to those of sequencing.

1.5 Aims of this study

The carboxylate platform has shown to be an effective way to produce medium-chain carboxylates and it created an opportunity to substitute fossil-based fuels and chemicals. However, characterising the chain elongation processes in existing open reactor systems is a grand challenge, due to the complexity regarding microbial interactions and involved metabolic processes. Systematic investigations of the metabolism and ecological interactions of reactor microbiota involved in lactate-based chain elongation were missing so far. Knowledge on the genetic features and metabolism of lactate-consuming chain-elongating species is limited. The goals of this dissertation were therefore:

1. to develop a closed model ecosystem by operating a continuous bioreactor with an enriched mixed culture, and to clarify how ecological interactions influence the chain elongation process performance and shape the community structure during a long-term reactor operation under constant conditions (**Chapter 2.1**),

2. to investigate how environmental manipulations affect chain elongation community assembly and functioning in the closed model ecosystems during long-term reactor operation (**Chapters 2.2 and 2.3**),

3. to identify the shared metabolic features and genomic diversity of the entire repertoire of experimentally validated chain-elongating bacterial species (**Chapters 2.4 and 2.5**).

2 Research chapters

2.1 Competition between butyrate fermenters and chain-elongating bacteria limits the efficiency of medium-chain carboxylate production

Bin Liu, Sabine Kleinsteuber, Florian Centler, Hauke Harms, Heike Sträuber*

Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

2.1.1 Main text

Competition Between Butyrate Fermenters and Chain-Elongating Bacteria Limits the Efficiency of Medium-Chain Carboxylate Production

Bin Liu, Sabine Kleinsteuber, Florian Centler, Hauke Harms and Heike Sträuber*

Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

Medium-chain carboxylates such as n-caproate and n-caprylate are valuable chemicals, which can be produced from renewable feedstock by anaerobic fermentation and lactate-based microbial chain elongation. Acidogenic microbiota involved in lactate-based chain elongation and their interplay with lactic acid bacteria have not been characterized in detail yet. Here, the metabolic and community dynamics were studied in a continuous bioreactor with xylan and lactate as sole carbon sources. Four succession stages were observed during 148 days of operation. After an adaptation period of 36 days, a relatively stable period of 28 days (stage I) was reached with n-butyrate, n-caproate and n-caprylate productivities of 7.2, 8.2 and 1.8 gCOD L^{-1} d^{-1}, respectively. After a transition period, the process changed to another period (stage II), during which 46% more n-butyrate, 51% less n-caproate and 67% less n-caprylate were produced. Co-occurrence networks of species based on 16S rRNA amplicon sequences and correlations with process parameters were analyzed to infer ecological interactions and potential metabolic functions. Diverse functions including hydrolysis of xylan, primary fermentation of xylose to acids (e.g., to acetate by *Syntrophococcus*, to n-butyrate by *Lachnospiraceae*, and to lactate by *Lactobacillus*) and chain-elongation with lactate (by *Ruminiclostridium* 5 and *Pseudoramibacter*) were inferred from the metabolic network. In stage I, the sub-network characterized by strongest positive correlations was mainly related to the production of n-caproate and n-caprylate. Lactic acid bacteria of the genus *Olsenella* co-occurred with potentially chain-elongating bacteria of the genus *Pseudoramibacter*, and their abundance was positively correlated with n-caproate and n-caprylate concentrations. A new sub-network appeared in stage II, which was mainly related to n-butyrate production and revealed a network of different lactic acid bacteria (*Bifidobacterium*) and potential n-butyrate producers (*Clostridium sensu stricto* 12). The synergy effects between lactate-producing and lactate-consuming bacteria constitute a division of labor cooperation of mutual benefit. Besides cooperation, competition between different taxa determined the bacterial community assembly over

the four succession stages in this resource-limited system. During long-term reactor operation under constant conditions, chain-elongating bacteria were outcompeted by butyrate-producing bacteria, leading to the increase of *n*-butyrate yield at the cost of medium-chain carboxylate yields in this closed model system.

INTRODUCTION

The production of platform chemicals and fuels from renewable resources is a major focus of a circular economy. The carboxylate platform offers the opportunity to sustainably produce bio-based chemicals such as medium-chain carboxylates (MCCs), which are mainly produced from coconut and palm kernel oils (Anneken et al., 2006). MCCs can be widely utilized in agriculture and industry, for example, as precursors for the production of fragrances (Kenealy et al., 1995), antimicrobial agents (Desbois, 2012) and drop-in biofuels (Urban et al., 2017). Besides the multi-functional applications, MCC production in a biorefinery context also meets the requirement of sustainable development because it replaces fossil resources and botanical oils such as palm kernel oil.

Medium-chain carboxylates are monocarboxylates that contain six to twelve carbon atoms. In this study, we subsume their dissociated and undissociated forms as carboxylates, with a main focus on *n*-caproate (C6) and *n*-caprylate (C8). In a process known as chain elongation (CE), intermediates of acidogenesis such as acetate (C2) and *n*-butyrate (C4) can be elongated to MCCs by adding acetyl-CoA in reverse β-oxidation cycles (Spirito et al., 2014). C2 or C4 need to be transformed to acetyl-CoA or butyryl-CoA, respectively, as initial substrate for elongation in the reverse β-oxidation. Thioesterase or coenzyme A transferase can be used as terminal enzymes for MCC production. Ethanol has been well described as electron donor providing energy for coupling acetyl-CoA formation and elongating acyl-CoA units (Seedorf et al., 2008). Besides ethanol, lactate also suits as electron donor for the CE process (Zhu et al., 2015; Kucek et al., 2016; Khor et al., 2017). Feedstocks that are rich in lactate (e.g., ensiled plant biomass) or lactate-precursors (e.g., carbohydrates) are thus promising substrates for the production of MCCs.

Phylogenetically different species have been described as CE bacteria that can produce C6 and even C8. The genera *Clostridium* (*C. kluyveri*), *Eubacterium* (*E. limosum* and *E. pyruvativorans*), *Megasphaera* (*M. elsdenii*, *M. indica*, and *M. hexanoica*), and *Caproiciproducens* (*C. galactitolivorans*) all include chain-elongators (Angenent et al., 2016). The recent discovery of chain-elongating *Ruminococcaceae* bacterium CPB6 (Zhu et al., 2017) suggests that further C6-producers remain to be discovered. To convert complex organic substrates (e.g., corn silage), the joint efforts of different trophic groups in a food web are required. However, the substrate spectrum of pure strains that are able to carry out CE is limited. Therefore, multi-species reactor microbiota can be considered more viable for the utilization of complex substrates due to their broad metabolic capacity. On the one hand, diverse functional groups in the microbial community may cooperate in metabolizing complex substrates like polysaccharides. Recent studies suggested that lactic acid bacteria (LAB) play an important role in lactate-based CE (Andersen et al., 2017; Scarborough et al., 2018b; Lambrecht et al., 2019). However, in such open-culture reactor systems, other intermediates including ethanol may also be produced from hydrolysis and acidogenesis of complex substrates, making it hard to discern the role of lactate. On the other hand, bacterial competition cannot be avoided in a resource-limited bioreactor. From an ecological perspective, the relationships between LAB and CE bacteria are still unexplored.

Aforementioned ecological interactions have been commonly investigated in other engineered microbial ecosystems, such as activated sludge of wastewater treatment plants (Ju and Zhang, 2015) and anaerobic digesters (Ziels et al., 2018). For chain elongation systems, it is not clear how cooperation and competition influence the process performance and shape the structure of the microbial community. To address this question, we studied lactate-based CE in a simplified lab-scale system. To reduce the complexity of a real system such as the anaerobic fermentation of ensiled plant biomass, we applied a model system with sterilized mineral medium containing xylan and lactate as sole carbon sources. We hypothesized that lactate formed *in situ* by sugar fermentation can be converted to MCCs in the CE process. By monitoring the process performance during long-term operation under constant conditions and by investigating the microbial community structure based on 16S rRNA amplicon sequencing, we intended to understand how the community dynamics affects the MCC productivity in our system. By performing network analysis, we aimed to elucidate the ecological interactions between the different functional groups LAB and CE bacteria.

MATERIALS AND METHODS

Growth Medium and Inoculum

The basal medium was modified from a previous study in which lactate was used to produce *n*-caproate (Weimer and Moen, 2013). It contained per liter: 0.054 g $MgCl_2 \cdot 6H_2O$, 0.065 g $CaCl_2 \cdot 2H_2O$, 1.612 g NH_4Cl, 5.470 g KH_2PO_4, 10.415 g K_2HPO_4, 0.032 g Na_2CO_3, 0.030 g cysteine-HCl, 0.5 g yeast extract, 1 mL of vitamin solution (biotin 20 mg/L, folic acid 20 mg/L, pyridoxine 100 mg/L, thiamine 50 mg/L, riboflavin 50 mg/L, nicotinic acid 50 mg/L, calcium pantothenate 50 mg/L, vitamin B_{12} 20 mg/L, *p*-amino benzoic acid 80 mg/L, lipoic acid 50 mg/L), and 1 mL of trace element solution ($FeCl_2 \cdot 4H_2O$ 1.5 g/L, $CuCl_2 \cdot 2H_2O$ 2 mg/L, $CoCl_2 \cdot 6H_2O$ 190 mg/L, $MnCl_2$ 100 mg/L,

$Na_2MoO_4 \cdot 2H_2O$ 36 mg/L, $NiCl_2 \cdot 6H_2O$ 24 mg/L, $Na_2WO_4 \cdot 2H_2O$ 20 mg/L, $Na_2SeO_3 \cdot 5H_2O$ 3 mg/L, $ZnCl_2$ 70 mg/L, H_3BO_3 6 mg/L). The medium was adjusted with 1 M NaOH solution to the operating value of pH 5.5. Lactate and xylan were fed daily as carbon sources to the reactor.

The inoculum was taken from a lab-scale CE reactor fed with lactate-rich corn silage (Lambrecht et al., 2019). This semi-continuous stirred tank reactor with a working volume of 12 L and operated at pH 5.5 had been daily fed with 3 L substrate mix resulting in a substrate retention time of 4 days. The bioreactor microbiota showed a stable performance for producing MCCs. Initially, 1 L fermentation broth taken from the lab-scale CE reactor was sieved (mesh size 2 mm) to remove particles of the corn silage. After the filtration, 875 mL liquid phase was used as inoculum and pumped into the reactor flushed with nitrogen. No chemical agent for the specific inhibition of methanogenesis was applied.

Bioreactor Operation and Sampling

A BioStat-A plus bioreactor (Sartorius AG, Göttingen, Germany) with 1 L working volume was used. The tank reactor was operated at 38 ± 1°C and at a constant stirring rate of 150 rpm. The pH was automatically controlled at 5.5 by addition of 1 M sodium hydroxide solution. For the daily feeding, 1.47 g lactic acid (85%, FCC grade; Sigma Aldrich, St. Louis, MI, United States) diluted in 50 mL deionized water, and 1.25 g water-soluble xylan (more than 95% xylooligosaccharides, from corncob; Roth, Karlsruhe, Germany) dissolved in 75 mL medium were supplied. Once a day, 125 mL effluent was taken before feeding corresponding to a hydraulic retention time (HRT) of 8 d. A gas-tight bag (produced on-site using thermoplastic coated aluminum foil) was used to collect the produced gas or for compensating underpressure in the reactor system. It was connected after a MilliGascounter (MGC-1; Ritter, Bochum, Germany). A buffer bottle was installed between the MGC and the bioreactor preventing the sealing fluid of the MGC-1 to be sucked into the reactor in case of underpressure. A septum was placed in the gas pipe for gas sampling. Gas samples of 1 mL were taken with a syringe flushed with nitrogen and injected into 20-mL gas-tight glass vials that had been flushed with argon for 20 min.

Reactor effluent was used for cell concentration measurement. For other analyses, liquid samples were collected twice per week and centrifuged for 10 min at 20,817 × g (Centrifuge 5417R; Eppendorf, Hamburg, Germany). The supernatants were used for measuring concentrations of xylan, total ammonia nitrogen (TAN), carboxylates and alcohols. Pelleted biomass samples from 50 mL reactor effluent were washed three times with phosphate buffer (PBS, 1.8 g L^{-1} Na_2HPO_4, 0.223 g L^{-1} NaH_2PO_4, 8.5 g L^{-1} NaCl in deionized H_2O pH 7.2; centrifugation at 10,000 × g, 10 min, 10°C) before determination of the cell dry weight. For microbial community analysis, the pelleted cells from 2 mL samples were washed with 100 mM Tris–HCl buffer pH 8.5 and stored at −20°C.

Analyses of Process Parameters

Daily gas production was monitored using MGC-1 and normalized to standard pressure and temperature conditions (101.325 kPa and 273.15 K) as described by Sträuber et al. (2018). Gas composition was analyzed in triplicate for H_2, CO_2, N_2, O_2, and CH_4 by GC according to Urban et al. (2017).

The TAN concentration was monitored twice a week as described previously (Popp et al., 2015).

Concentrations of carboxylates and alcohols were determined by gas chromatography (GC) in triplicate after derivatization of the analytes as previously reported (Urban et al., 2017). Here, 1 mL of 2-ethylbutyric acid was used as the internal standard. For the derivatization, 0.5 mL methanol and 2.5 mL 1 M sulfuric acid were added. Xylan was measured with a modified classical dinitrosalicylic acid reagent method (Miller, 1959). Xylan in the supernatant was acidified with 1 M sulfuric acid and then hydrolyzed at 121°C for 60 min. Before using the reagent, the pH of the hydrolysate was adjusted to be neutral.

The cell mass concentration was determined by measuring the optical density (OD) at 600 nm (spectrophotometer Genesys 10 S, Thermo Scientific Inc., Waltham, MA, United States) and correlated with the cell dry mass. For determining the cell dry mass, the cell pellets were dried at 60°C for 48 h before weighing (six replicates). Considering the microbial community shifts, we calculated a mean correlation coefficient (1 OD_{600} = 0.581 g L^{-1}) based on all cell dry mass measurements except the first measurement (**Supplementary Material B**). The chemical oxygen demand (COD) of microbial biomass was measured with a COD kit (LCK 714, Hach Lange GmbH, Germany) as described by Bonk et al. (2018).

The electron recovery was calculated according to the Eqs 1 and 2.

$$\eta_{e1-} = \frac{q_{C4-C10}}{q_{lac} + q_{xylan}} 100\% \tag{1}$$

$$\eta_{e2-} = \frac{q_{biomass}}{q_{lac} + q_{xylan}} 100\% \tag{2}$$

Where η_{e1-} and η_{e2-} are defined as electron recovery, q_{C4-C10} is the sum of all electrons in the carboxylates (iso-butyrate, n-butyrate, iso-valerate, n-valerate, n-caproate, n-heptanoate, n-caprylate, n-nonanoate, n-decanoate), $q_{biomass}$ is the number of electrons in the cell biomass, q_{lac} and q_{xylan} are the numbers of electrons of the input substrates lactate and xylan.

Microbial Community Analysis

Genomic DNA was extracted from frozen pellets using the NucleoSpin Microbial DNA Kit (Macherey-Nagel, Germany) according to the instructions of the manufacturer. Methods for DNA quantification and quality control were as described before (Lucas et al., 2015). The community dynamics throughout the experiment was studied by terminal restriction fragment length polymorphism (T-RFLP) fingerprinting. For this purpose, bacterial 16S ribosomal RNA (rRNA) genes were amplified by polymerase chain reaction (PCR) using the MyTaq™Mix (Bioline, Germany) and the primers 27f (labeled with phosphoramidite fluorochrome 5-carboxyfluorescein (FAM); 5'-GAG TTT GAT CMT GGY TCA G-3') and 1492r (5'-TAC GGY TAC CTT GTT ACG ACT T-3') (according to Lane, 1991). With a total volume of 12.5 µL, the mixtures of PCR

reaction contained 6.25 µL of MyTaq^{TM} Mix, 3.85 µL of nuclease free water, 0.7 µL of each primer (5 pmol) and 1 µL sample DNA (diluted to 20 ng µL^{-1}). The cycling protocol included an initial denaturation at 95°C for 1 min, followed by 30 cycles of denaturation at 95°C for 15 s, primer annealing at 58°C for 15 s, elongation at 72°C for 10 s, and a final elongation step at 72°C for 15 min. Amplicons were purified using the SureClean Kit (Bioline, Germany) and quantified using a NanoDrop ND 1000 spectral photometer (Thermo Fisher Scientific, United States). For T-RFLP analysis, 80 ng of 16S rRNA amplicons were digested over night at 37°C with 2 U of restriction endonuclease $MspI$ or $RsaI$ (New England Biolabs, Germany). The MapMarker1000 (BioVentures Inc., United States) was applied as fragment size standard. By using capillary electrophoresis with an automatic sequencer (ABI PRISM 3130xl Genetic Analyzer; Applied Biosystems, United States), the terminal restriction fragments (T-RFs) were separated. The electropherograms were analyzed by using the GeneMapper 5 software (Applied Biosystems) and processed by using a script according to Abdo et al. (2006) implemented in R Studio (Version 1.0.143). Low-signal peaks were removed below a threshold of seven times the standard deviation of data sets. T-RFs in the range of 50–1000 bp were included in further analyses.

For analyzing the community composition based on 16S rRNA gene sequences, 25 sampling points representing the different process stages were selected. Amplicon sequencing of 16S rRNA gene fragments was performed on the Illumina MiSeq platform (V3-V4 regions, 2 × 300 bp). The primers 341f (CCT ACG GGN GGC WGC AG) and 785r (GAC TAC HVG GGT ATC TAA KCC) were used for amplification. De-multiplexed sequence data were processed with QIIME2 v2019.1 (Bolyen et al., 2019). Filtering of phiX reads, denoising, merging of paired ends, trimming and chimera detection were done with the plugin of Divisive Amplicon Denoising Algorithm DADA2 (Callahan et al., 2016). The following parameters were used in DADA2: p-trim-left-f 0, p-trim-left-r 0, p-trunc-len-f 250, p-trunc-len-r 200. These were selected by reviewing the Interactive Quality Plot for removing low quality regions of the sequences. Other parameters were used by default. The generated feature table indicates the frequency each amplicon sequence variant (ASV) is observed in each sample. Taxonomic assignment was carried out using a naïve Bayes classifier trained on 16S rRNA gene sequences of the latest Midas database 2.1 (McIlroy et al., 2015). The feature table was rarefied down to the lowest read number (21,214 sequences) for further analyses. Two samples (days 46 and 116) were excluded due to lower read numbers. The de-multiplexed sequence dataset of 25 samples was deposited to the EMBL-EBI database under accession number PRJEB34417.

Statistical Analyses

Non-metric multidimensional scaling (NMDS) was used as ordination technique for dissimilarity matrices based on the T-RFLP profiles including occurrence and relative abundance of terminal restriction fragments (T-RFs). The Bray-Curtis dissimilarity index was used to evaluate the dynamics of the microbial communities (Bray and Curtis, 1957) reflected by the distances between data points. Smaller distances indicate higher similarities of community compositions. Based on the "vegan" R package (Oksanen et al., 2016), the "envfit" algorithm was used to calculate the relationships between abiotic parameters and T-RFLP profiles. The significance threshold was set to 0.01, which was tested by Monto Carlo test with 999 permutations.

Alpha diversity based on T-RFLP and ASV data was evaluated by using the ecological indices including richness, diversity and evenness as described by Lucas et al. (2017). Diversity of order one (D1) and evenness of order one (E1) quantify the diversity and evenness by weighting all present types equally, whereas diversity of order two (D2) and evenness of order two (E2) give more weight to the dominant types than to the rare types.

Significant differences of mean carboxylate recoveries were tested by Student's t-test (***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$, $n = 6$). Significant differences of mean biomass recoveries were tested by Mann-Whitney rank sum test (***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$, $n = 6$).

Co-occurrence networks based on 16S rRNA amplicon sequence data and abiotic parameters were inferred by following the protocol of Faust et al. (2015), using the CoNet App (v 1.1.1 beta) (Faust and Raes, 2016). Only ASVs that had > 0.1% relative abundance in more than three samples were included in the analysis to reduce spurious correlations. Relative ASV abundances were converted into absolute mean abundances based on total cell concentrations (gram dry mass per liter). Correlations between ASVs and process parameters (time; biomass; concentrations of C2, C4, C6, C8, and lactate; CO_2 and H_2 content; gas amount) were also considered in the network. Pearson, Spearman, and Kendall correlation coefficients were computed and if at least one method featured a coefficient below -0.75 or above 0.75, an edge connecting the corresponding ASVs or abiotic parameters was added to the network. All networks were visualized and analyzed for topological features in Cytoscape software (v 3.7.1) (Shannon et al., 2003).

RESULTS

Metabolic and Microbial Community Dynamics Over Different Succession Stages

The microbial chain elongation system analyzed in this study was designed to include hydrolysis and primary fermentation. Xylan and lactate were fed daily over a period of 148 days, and the reactor microbiota produced mainly n-butyrate, n-caproate and n-caprylate (**Figure 1A**). The gas was composed mainly of carbon dioxide and hydrogen (**Figure 1C**), with traces of nitrogen and oxygen. No methane was detected in the reactor headspace. The community dynamics analyzed by T-RFLP fingerprinting is shown as NMDS plots (**Figure 2** for $RsaI$ and **Supplementary Figure S1** for $MspI$). From 25 samples analyzed by amplicon sequencing, in total 909,240 sequence reads were obtained, which were assigned to 95 ASVs from high-quality sequence reads. Overall, ASVs were affiliated to three phyla

33

FIGURE 1 | Performance of bioreactor. **(A)** Substrate and product concentrations in the bioreactor during the four succession stages, **(B)** cell concentration (OD_{600} value) in the bioreactor during the four succession stages, and **(C)** daily gas production and gas composition during the four succession stages. Data points for gas production indicate mean values of 3 days. Within 95% confidence intervals, the error bars represent the standard deviation.

(*Firmicutes, Actinobacteria* and *Proteobacteria*), six classes (*Clostridia, Coriobacteriia, Actinobacteria, Erysipelotrichia, Bacilli* and *Alphaproteobacteria*), seven orders, 11 families and 20 genera with at least 0.1% relative abundance for each ASV (**Figure 3**). During this long-term reactor operation, four succession stages – adaptation (days 0–21), stage I (days 22–64), transition (days 65–98) and stage II (days 99–148) – were identified based on carboxylate concentration profiles (**Figure 1A**) and T-RFLP profiles (**Figure 2** and **Supplementary Figure S1**).

In the adaptation stage, a certain share of the carboxylates produced still originated from the liquid inoculum. Large variations in daily gas production and carbon dioxide content were observed (**Figure 1C**). The T-RFLP profiles (**Figure 2**) indicate considerable community shifts from the adaptation stage to the formation of a stable community composition (**Figure 3**). In the adaptation stage, the microbial community was dominated by ASVs assigned to unclassified *Lachnospiraceae*, *Lactobacillus*, *Bifidobacterium*, *Pseudoramibacter*, *Olsenella*, *Aeriscardovia*, *Solobacterium*, and *Atopobium* (**Figure 3**). Hereafter, it took around 2.5 HRTs for the microbiota to adapt to the given process conditions.

After the adaptation, the dominating genera were distinctly different from those in the inoculum, which indicates the highly selective conditions of our reactor system. In contrast to the strong community shifts during adaptation, data points standing for stage I (red), transition (green) and stage II (blue) are less scattered in the NMDS plot, displaying a relatively lower dissimilarity of the community structures within these periods (**Figure 2**). After the transition stage, the microbial community shifted from stage I to stage II as indicated by the vector "Stage." Similar results were obtained by T-RFLP analysis with *Msp*I (**Supplementary Figure S1**).

Alpha diversity metrics shows that richness, diversity of order one (D1) and as well as evenness of order one (E1) based on the T-RFLP data varied more over time than the respective indices based on the ASVs (**Supplementary Figure S2**). This could be due to the limitation of the T-RFLP method, which fails to detect rare sequence types. When focusing on the dominant types (D2 and E2), we observed a clear trend that the diversity (D2) was lower in stage II (mean values; ASV: 4.6, *Rsa*I: 2.5 and *Msp*I: 3.5) compared with stage I (mean values; ASV: 6.1, *Rsa*I: 4.7 and *Msp*I: 5.8). The community in stage II was also less even than that in stage I, as E2 was lower in stage II (mean values; ASV: 0.2, *Rsa*I:

34

FIGURE 2 | Bacterial community dynamics in the four succession stages, illustrated by a non-metric multidimensional scaling (NMDS) plot of T-RFLP profiles (16S rRNA gene amplicons, restriction enzyme *Rsa*I; the plot based on *Msp*I is shown in **Supplementary Figure S1**). Data points are named according to sampling days. Proximity of data points represents community similarity based on the Bray-Curtis index. Colored polygons indicate sampling days of each process stage. The vector shows community shifts within the temporal dynamics ($P < 0.01$, significance calculated by Monte-Carlo test with 999 permutations).

0.2 and *Msp*I: 0.2) compared with stage I (mean values; ASV: 0.3, *Rsa*I: 0.4, and *Msp*I: 0.4).

In stage I, the first period with constant carboxylate production over days 46–64 was observed (**Figure 1A**). With a loading rate of 10.7 gCOD L^{-1} d^{-1} as lactate and 12.1 gCOD L^{-1} d^{-1} as xylan, mean concentrations of 1.0 ± 0.1 gCOD L^{-1} acetate, 7.2 ± 0.7 gCOD L^{-1} *n*-butyrate, 8.2 ± 0.7 gCOD L^{-1} *n*-caproate and 1.8 ± 0.2 gCOD L^{-1} *n*-caprylate were obtained. Additionally, 1.0 ± 0.2 gCOD L^{-1} of lactate and 0.5 ± 0.5 gCOD L^{-1} of xylan were detected during this period. The mean daily gas production was 599.4 ± 85.9 mL d^{-1} in stage I. The gas consisted mainly of CO_2 ($60.4 \pm 1.7\%$) and H_2 ($33.2 \pm 0.7\%$). After the inoculation, still some particles from the seed sludge were retained in the bioreactor, which highly influenced the measured OD values. Therefore, OD values are only shown from day 15 onward (**Figure 1B**) as we assumed that most of the residual particles in the inoculum were washed out by then. The mean cell mass concentration was 0.66 ± 0.02 g dry mass L^{-1}. Here, ASVs assigned to *Syntrophococcus* ($17.2 \pm 5.9\%$), *Lactobacillus* ($15.9 \pm 4.3\%$), *Pseudoramibacter* ($5.9 \pm 2.6\%$), *Olsenella* ($7.9 \pm 2.4\%$), *Aeriscardovia* ($1.9 \pm 0.7\%$), *Solobacterium* ($2.6 \pm 1.7\%$), *Atopobium* ($4.3 \pm 2.5\%$), uncultured *Coriobacteriaceae* ($0.3 \pm 0.2\%$), *Erysipelotrichaceae* UCG 009 ($0.9 \pm 0.6\%$), *Eubacterium nodatum* group ($0.3 \pm 0.2\%$), *Lachnospira* ($0.1 \pm 0.1\%$), unclassified *Erysipelotrichaceae* ($0.1 \pm 0.1\%$), and *Ruminiclostridium* 5 ($42.3 \pm 3.8\%$) (mean relative abundance $\pm$ standard deviation, $n = 7$) predominated,

while ASVs identified as *Bifidobacterium* and unclassified *Lachnospiraceae* were detected below 0.1% of relative abundance. ASVs assigned to *Acetobacter* and *Clostridiales* family XIII UCG 001 were below the detection limit.

From the beginning of the transition interval on day 65, a trend of more C4 production and less C6/C8 production was observed. The daily gas production was not as stable as before. Here, ASVs assigned to *Aeriscardovia* and *Pseudoramibacter* disappeared and *Clostridium sensu stricto* 12 ($0.9 \pm 0.7\%$; $n = 4$) emerged (**Figure 3**).

After the transition period, 46% more C4, 51% less C6 and 67% less C8 were produced compared with stage I. In stage II, we obtained mean concentrations of 1.3 ± 0.3 gCOD L^{-1} acetate, 10.5 ± 1.0 gCOD L^{-1} *n*-butyrate, 4.0 ± 0.3 gCOD L^{-1} *n*-caproate, and 0.6 ± 0.1 gCOD L^{-1} *n*-caprylate. The mean concentration of lactate was 1.6 ± 0.3 gCOD L^{-1}, whereas no xylan was detected in stage II. For comparing with the results of stage I, concentrations over days 130–148 were used for calculating mean concentrations over the last six sampling points in each stage. Remarkably, no propionate was detected since stage I. The daily gas production was 21.1% lower than in stage I, with an average of 473.0 ± 84.3 mL d^{-1}. The contents of CO_2 and H_2 were 62.6 ± 5.0 and $31.3 \pm 5.3\%$, respectively. The fluctuation of the H_2 content was always consistent with the daily gas production. For CO_2, the trend was in the reverse direction throughout stage II. The mean daily hydrogen production was 152.3 ± 48.3 mL d^{-1}, the mean daily carbon dioxide production was 296.2 ± 48.4 mL d^{-1}. Noteworthy, occasional underpressure in stage II was indicated by the sealing fluid of the MGC sucked into the tube toward the reactor. The mean cell mass concentration increased by 42% in stage II up to 0.93 ± 0.07 g dry mass L^{-1}. In stage II, ASVs assigned to *Eubacterium nodatum* group, *Lachnospira*, *Lactobacillus*, *Syntrophococcus* and *Solobacterium* increased in their relative abundance to 1.6 ± 0.8, 1.4 ± 0.7, 18.9 ± 7.6, 39.8 ± 6.9, and $9.4 \pm 3.6\%$ (stage II, $n = 8$), respectively. ASVs of *Atopobium*, *Ruminiclostridium* 5 and *Olsenella* dropped down to abundances of 0.7 ± 0.5, 19.3 ± 4.4, and $7.2 \pm 1.9\%$, respectively. All uncertainties are represented by 95% confidence intervals.

Electron and COD Balances in Stage I and Stage II

The electron recovery indicates the partitioning of electrons from electron donors (xylan and lactate) to acceptors as a result of anabolic and catabolic processes. Taking all compounds analyzed by GC into account and considering a total input of 2.85 mol L^{-1} electron equivalents, the electron balances were similar in stage I ($92 \pm 3\%$) and stage II ($89 \pm 4\%$) (**Supplementary Table S1**). Comparable results were obtained for the COD balances (**Supplementary Table S2**). With an input of 23.48 gCOD L^{-1}, the COD balances in stage I and stage II were 91.2 ± 0.6 and $92.2 \pm 0.8\%$, respectively.

Most of the electrons were recovered in the C4-C10 products in both stages (**Supplementary Table S1** and **Figure 4**). As shown in **Figure 4**, the electron recovery (median values) in these CE products decreased by 12% from stage I to stage II as

35

FIGURE 3 | Microbial community composition profiles in the four succession stages based on amplicon sequencing of 16S rRNA genes. Taxonomic classification of amplicon sequence variants (ASVs) was categorized at the genus level. ASVs with relative abundance lower than 0.1% are classified into "Others."
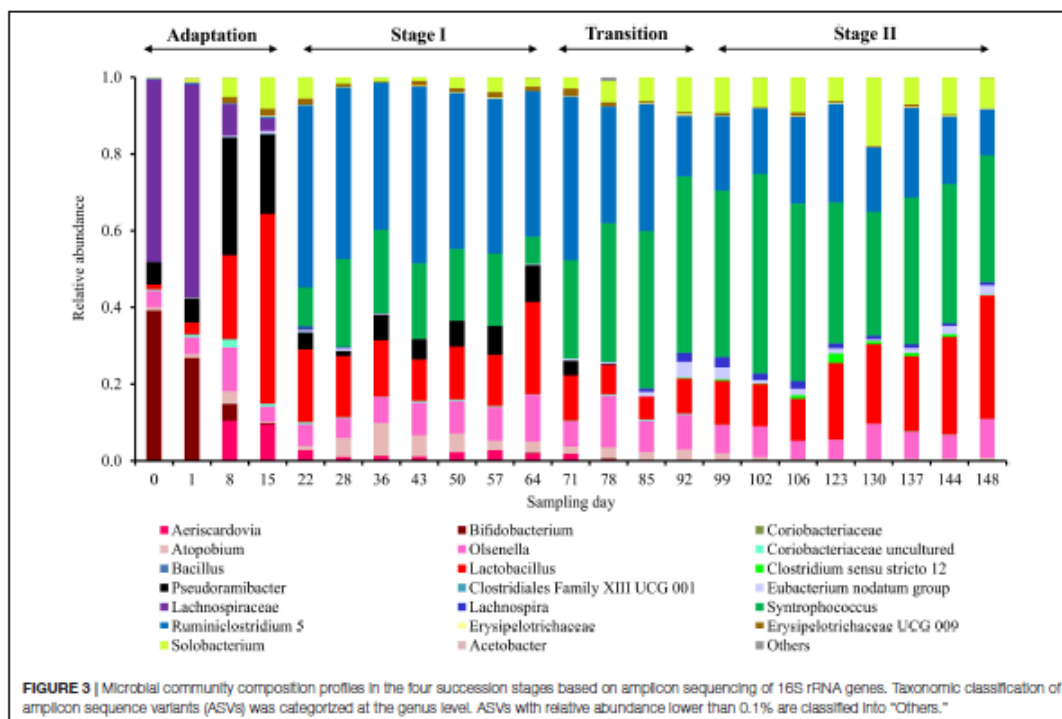
time progressed, which represents a significant difference (two-tailed P-value = 0.007). Moreover, compared with stage II, 80% more of the electrons (median values) from consumed xylan and lactate were channeled to cell biomass in stage II, which also shows a significant difference between the two stages (two-tailed P-value = 0.002). Thus, electron balances indicate that a higher percentage of substrate was directed to cell biomass synthesis and other non-target CE products with progressing operating time.

Network Inference in Stage I and Stage II

We constructed three separate networks to analyze the relationships among microbial taxa or process parameters with the aim to reveal potential functions and ecological interactions within the microbial community in our CE reactor.

The network inferred from data of stage I only (**Figure 5A**) mainly consisted of two co-occurring sub-network modules. The first one was a C6/C8-related sub-network (left) characterized by positive correlations of *Lactobacillus* with unclassified *Coriobacteriaceae*, *Pseudoramibacter* with *Olsenella* and *Pseudoramibacter* with unclassified *Coriobacteriaceae*. C6 production was positively correlated with *Pseudoramibacter*, *Olsenella* and unclassified *Coriobacteriaceae*. The second sub-network (right) was acetate related. C2 production was positively correlated with *Bifidobacterium*, *Clostridiales* family XIII UCG 001, unclassified *Lachnospiraceae* and *Solobacterium*.

Lachnospira, *Eubacterium nodatum* group and *Aeriscardovia* were also involved in positive correlations within this module.

In the network derived on data of stage II only (**Figure 5B**), a new C4-related sub-network module appeared. C4 production positively correlated with *Clostridium sensu stricto* 12, a genus that was positively correlated with *Ruminiclostridium* 5, unclassified *Erysipelotrichaceae* and *Bifidobacterium* here. It is worth to mention that *Atopobium* correlated with C6 and C8 positively in stage II, in which *Pseudoramibacter* had vanished.

In addition, we identified more co-occurrence pairs in a network analysis comprising stage I, transition stage and stage II (**Figure 5C**). Here, *Aeriscardovia* correlated positively with C6 and C8 production, while *Syntrophococcus* correlated with C4 production. Patterns of co-occurrence were detected for *Aeriscardovia* with *Pseudoramibacter* and *Syntrophococcus* with *Clostridium sensu stricto* 12, *Lachnospira*, *Solobacterium*, and *Eubacterium nodatum* group.

DISCUSSION

In our previous studies (Sträuber et al., 2012, 2016, 2018; Lambrecht et al., 2019), ensiled energy crops were used as substrate for MCC production. To simulate the feedstock conditions of an acidogenic fermenter fed with crop silage and producing MCCs in the present study, we selected xylan

FIGURE 4 | Electron recoveries in carboxylates (C4-C10) and cell biomass, represented by 95% confidence intervals. Boxes indicate the interquartile range between the 25th and 75th percentiles, respectively, the red line inside indicates the median value, mean values between both stages show significant differences ($P < 0.05$).
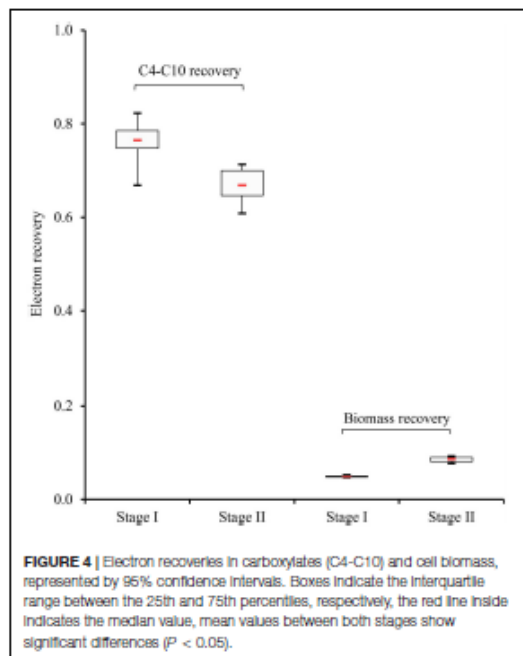
and lactate as model substrates. Feeding such defined carbon sources enabled electron balances and carbon flows. To gain further knowledge on the microbial community development in such a CE model system, we simplified the community by preventing continuous inoculation as it would have occurred in open systems.

Xylan is the major component of hemicellulose in plant cell walls (Badal, 2004). During acidogenic fermentation of corn silage, higher degradation of hemicellulose compared with cellulose was testified both in our previous batch (Sträuber et al., 2012) and continuous (Sträuber et al., 2016) studies. The xylan we used here contained more than 95% xylooligosaccharides (XOS), which is a mixture of oligosaccharides formed from xylose with a polymerization degree ranging from 2 to 10. Lactate is a typical fermentation product of LAB during the ensiling process. Previous studies have shown that the LAB *Bifidobacterium*, which was highly abundant in the inoculum of our CE process but later decreased in abundance, can ferment XOS to lactate and acetate as main products (Okazyki et al., 1990; Falck et al., 2013). *Lactobacillus*, which was highly enriched and became the most abundant LAB in our reactor, can also ferment XOS and produces mainly lactate (Kontula et al., 1998; Ananieva et al., 2012). For the other LAB we detected, such as *Aeriscardovia*, *Atopobium* and *Olsenella*, the ability to hydrolyze XOS has not been demonstrated yet. Therefore, we assume that they benefited from XOS-hydrolyzing bacteria and fermented xylose to produce mainly lactate, acetate and formate (Placidi et al., 2001; Kraatz et al., 2011). Other dominant genera,

such as *Ruminiclostridium* 5, *Solobacterium*, *Syntrophococcus*, *Pseudoramibacter*, *Eubacterium nodatum* group, *Clostridium sensu stricto* 12, and *Lachnospira*, were not yet reported to hydrolyze XOS.

Our study was designed to include hydrolysis and primary fermentation in the CE process. Acetate-producing xylose fermenters such as *Syntrophococcus* (Dore and Bryant, 1990) and butyrate-producing xylose fermenters such as *Lachnospiraceae* (Cotta and Forster, 2006) and *Solobacterium* (Kageyama and Benno, 2000) can be assumed to be involved in primary fermentation. Lactate-utilizing species of the genus *Eubacterium* were reported to produce mainly butyrate (Duncan et al., 2004), which indicates that CE with lactate may also be involved in C4 production. The produced intermediates (C2 and C4) can be elongated to MCCs with lactate as electron donor. We suppose that, besides the lactate fed as substrate, *in situ* lactate formation from xylan as mentioned above contributed to the CE process. The dominant genus *Ruminiclostridium* 5 might have been the potential CE bacteria, as strain CPB6 belonging to this genus was described to catalyze CE with lactate (Zhu et al., 2017). Another genus potentially involved in MCC production is *Pseudoramibacter*, which dominated only in stage I characterized by high C6/C8 concentrations. *Candidatus* P. fermentans was predicted to use lactate as substrate for CE (Scarborough et al., 2019). MCC production directly from xylose without external electron donor, which might also have been a possible process in our reactor, was described for *Candidatus* Weimerbacter bifidus of the family *Lachnospiraceae* (Scarborough et al., 2019).

Overall, our CE process showed diverse functions including hydrolysis of XOS (e.g., by *Lactobacillus*), primary fermentation of xylose to acids (by *Syntrophococcus* for C2, *Lachnospiraceae* for C4, and *Lactobacillus* for lactate) and CE with lactate (by *Eubacterium*, *Ruminiclostridium* 5 and *Pseudoramibacter*).

To identify potential ecological interactions within microbial communities, correlation-based network analysis may help understand the guiding rules of community assembly and decipher the community dynamics (Röttjers and Faust, 2018). Process parameters were also included in the network analyses in some studies on artificial microbial ecosystems (Ju and Zhang, 2015; Ziels et al., 2018) to understand the change of process performance and further to maintain process stability. For our CE microbiota, we detected pairwise relationships among taxa, among abiotic parameters, and between taxa and parameters over time during the different process periods.

Generally, many more positive correlations were observed in stage I than in stage II. In stage I, the higher C6/C8 productivity can be explained by *Pseudoramibacter* that served as a key taxon within the C6/C8-related sub-network. In stage II, the C6/C8-related sub-network was less complex and only determined by *Atopobium*. The C4-related sub-network emerged in stage II with the key taxon *Clostridium sensu stricto* 12, explaining the higher C4 productivity.

Co-occurrences between pairs of phylogenetically distant taxa may suggest bacterial cooperation such as mutualism. As reported by Ju and Zhang (2015), the co-occurrence between ammonia-oxidizing bacteria of the genus *Nitrosomonas* and nitrite-oxidizing bacteria of the genus *Nitrospira* most

FIGURE 5 | Correlation networks for stage I **(A)**, stage II **(B)**, and stage I - transition stage - stage II **(C)**. Pearson, Spearman, and Kendall correlation coefficients were computed for each pair of microbial taxa (rectangular nodes) and abiotic parameters (elliptic nodes) over the respective period using absolute taxon abundances. Edges indicate a coefficient > 0.75 for positive correlations (blue edges) and < −0.75 for negative correlations (red edges). Edge thickness indicates the number of methods supporting the correlation (1 to 3). Shading of the microbial taxon nodes indicates mean relative abundance over the respective period. Note the identical node layout in both networks. To ease comparison, also unconnected nodes are shown that feature correlations in the other network. However, unconnected nodes with zero abundance were removed, like Pseudoramibacter and Aeriscardovia in the network of stage II.

38

likely suggests their mutualistic interactions in activated sludge. In our bioreactor, a typical example was the co-occurrence of *Pseudoramibacter* (phylum *Firmicutes*) and *Olsenella* (phylum *Actinobacteria*) in stage I (**Figure 5A**). Chain elongators such as *Pseudoramibacter* might use lactate released from LAB such as *Olsenella* to produce C6/C8. Lactate-based CE driven by *Olsenella* was recently reported by Lambrecht et al. (2019) for the reactor microbiota that served as inoculum for our reactor. We therefore assume that the C6/C8 sub-network was a key feature of the CE process based on corn silage, which probably persisted after adaptation to the defined carbon sources at least during stage I. The other example of cooperation was shown in the C4-related sub-network of stage II (**Figure 5B**). Here, *Bifidobacterium* (phylum *Actinobacteria*) co-occurred with *Clostridium sensu stricto* 12 (phylum *Firmicutes*), the latter ASV sharing a high similarity (98.3% BLAST identity) with *Clostridium luticellarii*. We propose that *Clostridium sensu stricto* 12 may have used acetate and lactate released from *Bifidobacterium* to produce C4 in our system. *C. luticellarii* was also assumed as the dominant candidate for performing methanol-based CE in the study of de Smit et al. (2019). Most importantly, not only lactate was provided as electron donor for CE, but also removing lactate as the reaction product shifts the reaction equilibrium toward more lactate production. In other words, LAB might increase the availability of energy from such shift. Such synergy between producer and consumer constitutes a division of labor cooperation revealed as mutual benefit (González-Cabaleiro et al., 2015).

By supplying a finite carbon resource in a CE system, it is reasonable to assume that bacterial competition also impacts unignorably the community structure, manifesting the shift of process performance. As reported in the literature, C4-producers like species of the genus *Lachnospira* (Cotta and Forster, 2006) can ferment xylose. In our CE system, the negative correlations between the functional groups of lactate producers (*Olsenella*, *Lactobacillus*) and C4 producers (e.g., *Lachnospira*) (**Figure 5B**) may potentially reflect the competition for the carbon and energy source xylose. This competition might direct the carbon flow more to C4 as observed in stage II. Such negative interactions between different functional groups may have some agreement with previous findings in other biotechnological systems such as wastewater treatment plants (Ju and Zhang, 2015). Likewise, under such resource-limited conditions, widespread competition between taxa crucially structures the microbial community. Finally, although integrating process parameters and absolute biomass can effectively support our hypotheses of bacterial cooperation and competition, the true ecological interactions still need to be validated in culture-dependent experiments with defined synthetic communities of species with known metabolic functions.

During long-term reactor operation, we found that the reactor microbiota self-optimized to yield more biomass at the cost of C6/C8 yields. This indicates that the C6/C8-producing bacteria in our system could not successfully compete with C4-producing bacteria during the battle for the finite resources. This might be different in open systems such as those fed with complex biomass, where new microorganisms including different CE bacteria can enter the system during operation. In our model system, the absolute abundance of potential C6-producers such as *Ruminiclostridium* 5 decreased significantly in stage II, and the genus *Pseudoramibacter* was even completely washed out after the transition period. Instead, other functional groups including potential C2-producers (e.g., *Syntrophococcus*), LAB (*Lactobacillus*) and C4-producers (*Clostridium sensu stricto* 12, *Solobacterium*, *Eubacterium nodatum* group and *Lachnospira*) increased in their absolute abundances. The negative correlations between CE bacteria of the genus *Pseudoramibacter* and other functional groups are shown in **Figure 5C**. Considering the higher C2 and C4 electron recoveries in stage II, we conclude that bacteria of these functional groups (i.e., C2-producers, LAB and C4-producers) captured energy from the substrate more efficiently than C6/C8-producing bacteria for their growth. Our results may confirm the theory that maximum metabolic energy harvest rate for growth can select the microbial catabolic activities in microbial ecosystems (González-Cabaleiro et al., 2015). As for the CE process, longer pathways with higher input of metabolic labor decrease the energy harvest rate. This applies to another recent study on the CE process, which also explained this effect well from the thermodynamics perspective (Wu et al., 2018). For our system, the pattern of competition would be more favored due to the following reasons. First, the enriched members share overlapping metabolic niches and require the same nutrients as so many species are functionally similar. After inoculation of the reactor, no new microorganisms with different metabolic needs or capabilities were brought into this insular community. Second, in our reactor system, all nutrients are well mixed, thus limiting spatial heterogeneity and consequently different niches, but favoring nutrient availability. In both aspects, the conditions in our system differ from those in systems with complex biomass substrate. Moreover, our resource-limited reactor drove selection for favoring bacteria that rapidly grow to take up resources (Maitra and Dill, 2015). As a consequence, community dynamics over time depends on the selection pressures mentioned above (Ghoul and Mitri, 2016). Therefore, competition cannot be avoided when using mixed cultures for producing MCCs. However, the degree of competition might be different in open systems. To what extent our observations could be extrapolated to more complex systems needs to be further tested. Until now, many studies focused mainly on other competing processes such as methanogenesis (Grootscholten et al., 2014), bacterial sulfur reduction (Cavalcante et al., 2016) and the acrylate pathway in lactate-based CE (Kucek et al., 2016) to ensure effective MCC production. Such processes were not observed here. However, the C4-producers should be also realized as a competitor for utilizing carbon sources and other nutrients. In this study, the processes of xylose fermentation to butyrate and lactate-based CE of acetate both contributed to the C4 production. In the mixed microbial fermentation study of Scarborough et al. (2018a), these were also described as the competing processes in CE for producing MCCs.

Another possible explanation is that product inhibition also promoted the community shift. With an operation at pH 5.5, the protonated C6 ($0.72\ g\ L^{-1}$) and C8 ($0.15\ g\ L^{-1}$) in stage I showed inhibitory concentrations comparable to the system reported by Andersen et al. (2017).

Furthermore, other environmental factors such as HRT and pH probably also influenced the community development and the MCC production during the long-term reactor operation. Since the operation conditions were not changed in this study, future experiments could investigate the effects of certain abiotic factors on the CE community assembly by changing them.

Our findings showed that hydrolysis, primary fermentation and CE with lactate functions were all enriched in the reactor microbiota by feeding xylan and lactate as the substrates. Ecological interactions such as cooperation between LAB and CE bacteria, as well as competition between C6/C8-producing bacteria and C4-producing bacteria, resulted in the community development over four succession stages. The higher biomass and C4 yields at the cost of C6/C8 yields may be explained by the ecological interactions discussed above. Additionally, the lower reactor performance in terms of C6/C8 production could be attributed to the loss of diversity in stage II, as the more diverse community in stage I might have a higher capacity to use redundant pathways, resulting in more efficient community functions (Werner et al., 2011). In conclusion, during the long-term reactor operation without tuning any process parameters, the CE reactor microbiota developed toward predominating C4 and biomass production instead of MCC production in our system.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the **Supplementary Material B** and the EMBL European Nucleotide Archive (ENA) under accession number PRJEB34417 (http://www.ebi.ac.uk/ena/data/view/PRJEB34417).

REFERENCES

Abdo, Z., Schüette, U. M. E., Bent, S. J., Williams, C. J., Forney, L. J., Joyce, P., et al. (2006). Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Environ. Microbiol.* 8, 929–938. doi: 10.1111/j.1462-2920.2005.00959.x

Ananieva, M., Mandadzhieva, T., Koliandova, I., and Stoyanovski, S. (2012). Utilization of xylooligosaccharides from different *Lactobacillus* strains. *J. BioSci. Biotech.* 2012, 147–150.

Andersen, S. J., de Groof, V., Khor, W. C., Roume, H., Props, R., Coma, M., et al. (2017). A Clostridium group IV species dominates and suppresses a mixed culture fermentation by tolerance to medium chain fatty acids products. *Front. Bioeng. Biotechnol.* 5:8. doi: 10.3389/fbioe.2017.00008

Angenent, L. T., Richter, H., Buckel, W., Spirito, C. M., Steinbusch, K. J. J., Plugge, C. M., et al. (2016). Chain elongation with reactor microbiomes: open-culture biotechnology to produce biochemicals. *Environ. Sci. Technol.* 50, 2796–2810. doi: 10.1021/acs.est.5b04847

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00336/full#supplementary-material

Anneken, D. J., Both, S., Christoph, R., Fieg, G., Steinberner, U., and Westfechtel, A. (2006). "Fatty acids," in *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. 14 (Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA), 73–116.

Badal, C. S. (2004). "Lignocellulose biodegradation and applications in biotechnology," in *Lignocellulose Biodegradation*, eds B. C. Saha and K. Hayashi (Washington, DC: American Chemical Society), doi: 10.1021/bk-2004-0889.ch001

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Chase, J., Cope, E. K., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bonk, F., Popp, D., Weinrich, S., Sträuber, H., Kleinsteuber, S., Harms, H., et al. (2018). Intermittent fasting for microbes: how discontinuous feeding increases functional stability in anaerobic digestion. *Biotechnol. Biofuels* 11:274. doi: 10.1186/s13068-018-1279-5

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349. doi: 10.2307/1942268

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Cavalcante, W. D. A., Leitão, R. C., Gehring, T. A., Angenent, L. T., and Santaella, S. T. (2016). Anaerobic fermentation for n-caproic acid production: a review. *Process Biochem.* 54, 106–119. doi: 10.1016/j.procbio.2016.12.024

Cotta, M., and Forster, R. (2006). The family Lachnospiraceae, including the genera Butyrivibrio, Lachnospira and Roseburia. *Prokaryotes* 4, 1002–1021. doi: 10.1007/0-387-30744-3_35

de Smit, S. M., de Leeuw, K. D., Buisman, C. J. N., and Strik, D. P. B. T. B. (2019). Continuous n-valerate formation from propionate and methanol in an anaerobic chain elongation open-culture bioreactor. *Biotechnol. Biofuels* 12:132. doi: 10.1186/s13068-019-1468-x

Desbois, P. A. (2012). Potential applications of antimicrobial fatty acids in medicine, agriculture and other industries. *Recent Pat. Antiinfect. Drug Discov.* 7, 111–122. doi: 10.2174/157489112801619728

Dore, J., and Bryant, M. P. (1990). Metabolism of one-carbon compounds by the ruminal acetogen *Syntrophococcus sucromutans. Appl. Environ. Microbiol.* 56, 984–989. doi: 10.1128/AEM.56.4.984-989.1990

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. doi: 10.1128/AEM.70.10.5810

Falck, P., Precha-Atsawanan, S., Grey, C., Immerzeel, P., Stälbrand, H., Adlercreutz, P., et al. (2013). Xylooligosaccharides from hardwood and cereal xylans produced by a thermostable xylanase as carbon sources for *Lactobacillus brevis* and *Bifidobacterium adolescentis. J. Agric. Food Chem.* 61, 7333–7340. doi: 10.1021/jf401249g

Faust, K., Lima-Mendez, G., Lerat, J. S., Sathirapongsasuti, J. F., Knight, R., Huttenhower, C., et al. (2015). Cross-biome comparison of microbial association networks. *Front. Microbiol.* 6:1200. doi: 10.3389/fmicb.2015.01200

Faust, K., and Raes, J. (2016). CoNet app: inference of biological association networks using Cytoscape. *F1000Research* 5:1519. doi: 10.12688/f1000research.9050.2

Ghoul, M., and Mitri, S. (2016). The ecology and evolution of microbial competition. *Trends Microbiol.* 24, 833–845. doi: 10.1016/j.tim.2016.06.011

González-Cabaleiro, R., Ofiţeru, I. D., Lema, J. M., and Rodríguez, J. (2015). Microbial catabolic activities are naturally selected by metabolic energy harvest rate. *ISME J.* 9, 2630–2641. doi: 10.1038/ismej.2015.69

Grootscholten, T. I. M., Strik, D. P. B. T. B., Steinbusch, K. J. J., Buisman, C. J. N., and Hamelers, H. V. M. (2014). Two-stage medium chain fatty acid (MCFA) production from municipal solid waste and ethanol. *Appl. Energy* 116, 223–229. doi: 10.1016/j.apenergy.2013.11.061

Ju, F., and Zhang, T. (2015). Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. *ISME J.* 9, 683–695. doi: 10.1038/ismej.2014.162

Kageyama, A., and Benno, Y. (2000). Phylogenic and phenotypic characterization of some Eubacterium-like isolates from human feces: description of *Solobacterium moorei* gen. nov., sp. nov. *Microbiol. Immunol.* 44, 223–227. doi: 10.1111/j.1348-0421.2000.tb02487.x

Kenealy, W. R., Cao, Y., and Weimer, P. J. (1995). Production of caproic acid by cocultures of ruminal cellulolytic bacteria and *Clostridium kluyveri* grown on cellulose and ethanol. *Appl. Microbiol. Biotechnol.* 44, 507–513. doi: 10.1007/s002530050590

Khor, W. C., Andersen, S., Vervaeren, H., and Rabaey, K. (2017). Electricity-assisted production of caproic acid from grass. *Biotechnol. Biofuels* 10:180. doi: 10.1186/s13068-017-0863-4

Kontula, P., Von Wright, A., and Mattila-Sandholm, T. (1998). Oat bran β-gluco- and xylo-oligosaccharides as fermentative substrates for lactic acid bacteria. *Int. J. Food Microbiol.* 45, 163–169. doi: 10.1016/S0168-1605(98)00156-1

Kraatz, M., Wallace, R. J., and Svensson, L. (2011). *Olsenella umbonata* sp. nov., a microaerotolerant anaerobic lactic acid bacterium from the sheep rumen and pig jejunum, and emended descriptions of *Olsenella, Olsenella uli* and *Olsenella profusa. Int. J. Syst. Evol. Microbiol.* 61, 795–803. doi: 10.1099/ijs.0.022954-0

Kucek, L. A., Nguyen, M., and Angenent, L. T. (2016). Conversion of L-lactate into n-caproate by a continuously fed reactor microbiome. *Water Res.* 93, 163–171. doi: 10.1016/j.watres.2016.02.018

Lambrecht, J., Cichocki, N., Schattenberg, F., Kleinsteuber, S., Harms, H., Müller, S., et al. (2019). Key sub-community dynamics of medium-chain carboxylate production. *Microb. Cell Fact.* 18:92. doi: 10.1186/s12934-019-1143-8

Lane, D. J. (1991). "16S/23S rRNA sequencing," in *Nucleic Acid Techniques in Bacterial Systematics,* eds E. Stackbrandt and M. Goodfellow (Chichester: John Wiley and Sons), 177–203.

Lucas, R., Groeneveld, J., Harms, H., Johst, K., Frank, K., and Kleinsteuber, S. (2017). A critical evaluation of ecological indices for the comparative analysis of microbial communities based on molecular datasets. *FEMS Microbiol. Ecol.* 92:fiw209. doi: 10.1093/femsec/fiw209

Lucas, R., Kuchenbuch, A., Fetzer, I., Harms, H., and Kleinsteuber, S. (2015). Long-term monitoring reveals stable and remarkably similar microbial communities in parallel full-scale biogas reactors digesting energy crops. *FEMS Microbiol. Ecol.* 91:fiv004. doi: 10.1093/femsec/fiv004

Maitra, A., and Dill, K. A. (2015). Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc. Natl. Acad. Sci. U.S.A.* 112, 406–411. doi: 10.1073/pnas.1421138111

McIlroy, S. J., Saunders, A. M., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A. A., et al. (2015). MiDAS: the field guide to the microbes of activated sludge. *Database* 2015:bav016. doi: 10.1093/database/bav062

Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31, 426–428. doi: 10.1021/ac60147a030

Okazaki, M., Fujikawa, S., and Matsumoto, N. (1990). Effect of xylooligosaccharide on the growth of Bifidobacteria. *Bifidobact. Microflora* 9, 77–86. doi: 10.12938/bifidus1982.9.2_77

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2016). *Multivariate Analysis of Ecological Communities in R: vegan Tutorial.* Available online at: https://www.mooreecology.com/uploads/2/4/2/1/24213970/vegantutor.pdf (accessed September 18, 2017).

Placidi, F., Floris, R., Bozzao, A., Romigi, A., Baviera, M. E., Tombini, M., et al. (2001). Complete genome sequence of Atopobium parvulum type strain. *Stand. Genomic Sci.* 57, 534–537. doi: 10.4056/sigs.29547

Popp, D., Schrader, S., Kleinsteuber, S., Harms, H., and Sträuber, H. (2015). Biogas production from coumarin-rich plants-inhibition by coumarin and recovery by adaptation of the bacterial community. *FEMS Microbiol. Ecol.* 91:fiv103. doi: 10.1093/femsec/fiv103

Röttjers, L., and Faust, K. (2018). From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiol. Rev.* 42, 761–780. doi: 10.1093/femsre/fuy030

Scarborough, M. J., Lawson, C. E., Hamilton, J. J., Donohue, T. J., and Noguera, D. R. (2018a). Metatranscriptomic and thermodynamic insights into medium-chain fatty acid production using an anaerobic microbiome. *mSystems* 3:e00221-18. doi: 10.1128/mSystems.00221-18

Scarborough, M. J., Lynch, G., Dickson, M., McGee, M., Donohue, T. J., and Noguera, D. R. (2018b). Increasing the economic value of lignocellulosic stillage through medium-chain fatty acid production. *Biotechnol. Biofuels* 11:200. doi: 10.1186/s13068-018-1193-x

Scarborough, M. J., Myers, K. S., Donohue, T. J., and Noguera, D. R. (2019). Multi-omic analysis of medium-chain fatty acid synthesis by *Candidatus Weimerbacter bifidus,* gen . nov ., sp . nov ., and *Candidatus Pseudoramibacter fermentans,* sp. nov. bioRxiv [Preprint]

Seedorf, H., Fricke, W. F., Veith, B., Brüggemann, H., Liesegang, H., Strittmatter, A., et al. (2008). The genome of Clostridium kluyveri, a strict anaerobe with unique metabolic features. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2128–2133. doi: 10.1073/pnas.0711093105

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Spirito, C. M., Richter, H., Rabaey, K., Stams, A. J. M., and Angenent, L. T. (2014). Chain elongation in anaerobic reactor microbiomes to recover resources from waste. *Curr. Opin. Biotechnol.* 27, 115–122. doi: 10.1016/j.copbio.2014.01.003

Sträuber, H., Bühligen, F., Sabine, K., and Dittrich-Zechendorf, M. (2018). Carboxylic acid production from ensiled crops in anaerobic solid-state fermentation - trace elements as pH controlling agents support microbial chain elongation with lactic acid. *Eng. Life Sci.* 18, 447–458. doi: 10.1002/elsc.201700186

41

Sträuber, H., Lucas, R., and Kleinsteuber, S. (2016). Metabolic and microbial community dynamics during the anaerobic digestion of maize silage in a two-phase process. *Appl. Microbiol. Biotechnol.* 100, 479–491. doi: 10.1007/s00253-015-6996-0

Sträuber, H., Schröder, M., and Kleinsteuber, S. (2012). Metabolic and microbial community dynamics during the hydrolytic and acidogenic fermentation in a leach-bed process. *Energy. Sustain. Soc.* 2:13. doi: 10.1186/2192-0567-2-13

Urban, C., Xu, J., Sträuber, H., dos Santos Dantas, T. R., Mühlenberg, J., Härtig, C., et al. (2017). Production of drop-in fuel from biomass by combined microbial and electrochemical conversions. *Energy Environ. Sci.* 10, 2231–2244. doi: 10.1039/C7EE01303E

Weimer, P. J., and Moen, G. N. (2013). Quantitative analysis of growth and volatile fatty acid production by the anaerobic ruminal bacterium *Megasphaera elsdenii* T81. *Appl. Microbiol. Biotechnol.* 97, 4075–4081. doi: 10.1007/s00253-012-4645-4

Werner, J. J., Knights, D., Garcia, M. L., Scalfone, N. B., Smith, S., Yarasheski, K., et al. (2011). Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4158–4163. doi: 10.1073/pnas.1015676108

Wu, Q., Guo, W., Bao, X., Meng, X., Yin, R., Du, J., et al. (2018). Upgrading liquor-making wastewater into medium chain fatty acid: insights into co-electron donors, key microflora, and energy harvest. *Water Res.* 145, 650–659. doi: 10.1016/j.watres.2018.08.046

Zhu, X., Tao, Y., Liang, C., Li, X., Wei, N., Zhang, W., et al. (2015). The synthesis of n-caproate from lactate: a new efficient process for medium-chain carboxylates production. *Sci. Rep.* 5:14360. doi: 10.1038/srep14360

Zhu, X., Zhou, Y., Wang, Y., Wu, T., Li, X., Li, D., et al. (2017). Production of high-concentration n-caproic acid from lactate through fermentation using a newly isolated *Ruminococcaceae bacterium* CPB6. *Biotechnol. Biofuels* 10:102. doi: 10.1186/s13068-017-0788-y

Ziels, R. M., Svensson, B. H., Sundberg, C., Larsson, M., Karlsson, A., and Yekta, S. S. (2018). Microbial rRNA gene expression and co-occurrence profiles associate with biokinetics and elemental composition in full-scale anaerobic digesters. *Microb. Biotechnol.* 11, 694–709. doi: 10.1111/1751-7915.13264

42

2.1.2 Supplementary information

Figure S1 Bacterial community dynamics in the four succession stages, illustrated by nonmetric multidimensional scaling (NMDS) of T-RFLP profiles (16S rRNA gene amplicons, restriction enzyme *Msp*I). Data points are named according to sampling days. Proximity of data points represents community similarity based on the Bray-Curtis index. Colored polygons indicate sampling days of each succession stage. The vector shows community shifts within the temporal dynamics ($P < 0.01$, significance calculated by Monte-Carlo test with 999 permutations).
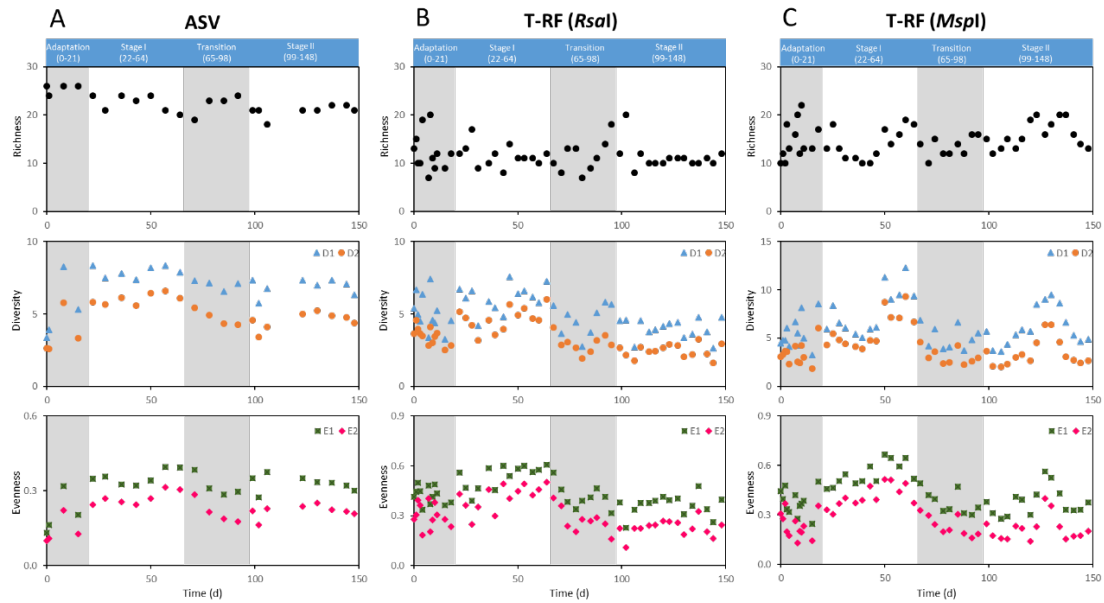
Figure S2 Alpha diversity metrics (richness, diversity and evenness) in the four succession stages based on the relative abundance of amplicon sequence variants (ASV) (**A**) and of terminal restriction fragments (T-RF) of 16S rRNA gene amplicons for restriction enzymes *Rsa*I (**B**) and *Msp*I **(C)**. D1: diversity of order one (q = 1), D2: diversity of order two (q = 2); E1: evenness of order one (q = 1), E2: evenness of order two (q = 2)

Table S1 Electron-equivalent balance in stages I and II showing mean values obtained from different sampling points in both stages, represented by 95% confidence intervals. The electron balance calculation was based on the number of electrons and the molar mass of the reduced compounds.

(mol/L)	Stage I		Stage II	
	Mean	Standard deviation	Mean	Standard deviation
Acetate[1]	0.13138	0.01259	0.16870	0.03952
iso-Butyrate[2]	0.00000	0.00000	0.00068	0.00014
n-Butyrate[3]	0.89421	0.09115	1.28598	0.11784
iso-Valerate[4]	0.00038	0.00009	0.00073	0.00018
n-Valerate[5]	0.00345	0.00026	0.01763	0.00193
n-Caproate[6]	1.00428	0.07569	0.49796	0.04068
n-Heptanoate[7]	0.00124	0.00008	0.00630	0.00028
n-Caprylate[8]	0.22191	0.01219	0.06714	0.00303
n-Nonanoate[9]	0.00011	0.00002	0.00020	0.00004
n-Decanoate[10]	0.00042	0.00010	0.00037	0.00006
Lactate[11]	0.12843	0.03005	0.19604	0.04347
Phenyl acetate[12]	0.00106	0.00147	0.00345	0.00112
Phenyl propionate[13]	0.00000	0.00000	0.00855	0.00175
Ethanol[14]	0.02892	0.00208	0.01566	0.00795
2-Butanol[15]	0.00000	0.00000	0.00017	0.00020
1-Propanol[16]	0.00152	0.00058	0.00370	0.00073
1-Butanol[17]	0.00825	0.00058	0.01835	0.00270
Biomass[18]	0.13511	0.00464	0.23645	0.01400
H_2[19]	0.01850	0.00341	0.01257	0.00584
Xylan[20]	0.04061	0.03988	0.00455	0.01016
Total Output	2.61978	0.07519	2.54884	0.12050
Total Input	2.84848	0	2.84848	0
Electron Balance	92%	3%	89%	4%

Number of electrons (mol mol^{-1}): [1]Acetate 8, [2]*iso*-Butyrate 20, [3]*n*-Butyrate 20, [4]*iso*-Valerate 26, [5]*n*-Valerate 26, [6]*n*-Caproate 32, [7]*n*-Heptanoate 38, [8]*n*-Caprylate 44, [9]*n*-Nonanoate 50, [10]*n*-Decanoate 56, [11]Lactate 12, [12]Phenyl acetate 36, [13]Phenyl propionate 42, [14]Ethanol 12, [15]2-Butanol 24, [16]1-Propanol

18, [17]1-Butanol 24, [18]Biomass 6 × n, [19]H_2 2, [20]Xylan 20 × n; "n" stands for the degree of polymerization. The electron number of cell biomass was calculated by using a theoretical molecular composition; the empirical formula of microorganisms was assumed as $CH_{1.8}O_{0.5}N_{0.2}$ (molar mass = 24.6 g mol^{-1}). *The molar mass* (g mol^{-1}): [1]Acetate 60, [2]*iso*-Butyrate 88, [3]*n*-Butyrate 88, [4]*iso*-Valerate 102, [5]*n*-Valerate 102, [6]*n*-Caproate 116, [7]*n*-Heptanoate 130, [8]*n*-Caprylate 144, [9]*n*-Nonanoate 158, [10]*n*-Decanoate 172, [11]Lactate 90, [12]Phenyl acetate 136, [13]Phenyl propionate 150, [14]Ethanol 46, [15]2-Butanol 74, [16]1-Propanol 60, [17]1-Butanol 74, [18]Biomass 24.6 × n, [19]H_2 1/22.4, [20]Xylan 132 × n; "n" stands for the degree of polymerization.

Table S2 COD balances in stage I and stage II. By using COD kit, one point was selected per week in stage I and stage II, results are shown below. The "Input" means measured results of mineral medium with lactate and xylan added. The "Output" means measured results of effluents. COD balance was calculated from the quotient of "Output" divided by "Input". Within 95% confidence intervals, standard deviation was calculated from three measured values.

	Time (d)	Input (gCOD L^{-1})	Standard deviation	Output (gCOD L^{-1})	Standard deviation	COD balance (%)	Standard deviation
	36	23.48	0.31	20.37	0.14	92.0	0.7
	43	23.48	0.31	19.94	0.09	90.0	0.4
Stage I	50	23.48	0.31	20.25	0.15	91.4	0.7
	57	23.48	0.31	19.94	0.08	90.0	0.3
	64	23.48	0.31	20.38	0.17	92.0	0.8
	120	23.48	0.31	20.52	0.15	92.7	0.7
	127	23.48	0.31	20.02	0.13	90.4	0.6
Stage II	134	23.48	0.31	20.96	0.13	94.6	0.6
	141	23.48	0.31	20.38	0.08	92.0	0.3
	148	23.48	0.31	20.19	0.36	91.2	1.6

2.2 Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture

Bin Liu[1], Heike Sträuber[1], João Saraiva[1], Hauke Harms[1], Sandra Godinho Silva[2], Sabine Kleinsteuber[1]* and Ulisses Nunes da Rocha[1]*

* Authors followed by an asterisk contributed equally to this work

[1]Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

[2]Institute for Bioengineering and Biosciences, Department of Bioengineering, Instituto Superior Técnico Universidade de Lisboa, Lisbon, Portugal

Corresponding Authors:
sabine.kleinsteuber@ufz.de / ulisses.rocha@ufz.de (ordered alphabetically according to last name).

2.2.1 Main text

Abstract

Background: The ability to quantitatively predict ecophysiological functions of microbial communities provides an important step to engineer microbiota for desired functions related to specific biochemical conversions. Here, we present the quantitative prediction of medium-chain carboxylate production in two continuous anaerobic bioreactors from 16S rRNA gene dynamics in enrichment cultures.

Results: By progressively shortening the hydraulic retention time from 8 days to 2 days with different temporal schemes in both bioreactors operated for 211 days, we achieved higher productivities and yields of the target products *n*-caproate and *n*-caprylate. The datasets generated from each bioreactor were applied independently for training and testing in machine learning. A predictive model was generated by employing the random forest algorithm using 16S rRNA amplicon sequencing data. More than 90% accuracy in the prediction of *n*-caproate and *n*-caprylate productivities was achieved. Four inferred bioindicators belonging to the genera *Olsenella*, *Lactobacillus*, *Syntrophococcus* and *Clostridium* IV suggest their relevance to the higher carboxylate productivity at shorter hydraulic retention time. The recovery of metagenome-assembled genomes of these bioindicators confirmed their genetic potential to perform key steps of medium-chain carboxylate production.

Conclusions: Shortening the hydraulic retention time of the continuous bioreactor systems allows to shape the communities with desired chain elongation functions. Machine-learning predictive analytics demonstrates that 16S rRNA amplicon sequencing data can be used to predict the model process performance in a quantitative and accurate way. Characterising and harnessing bioindicators holds promise to manage reactor microbiota towards selection of the target processes, as such taxa can be used to predict ecosystem performance. Our mathematical framework is transferrable to other ecosystem processes and microbial systems where community dynamics is linked to key

functions. The general methodology can be adapted to data types of other functional categories such as genes, transcripts, proteins or metabolites.

Keywords: Predictive biology, carboxylate platform, model ecosystems, reactor microbiota, microbial chain elongation

Background

Microbes form complex communities that play essential roles in ecosystem functioning. Identifying bioindicators derived from community analysis and using them to predict process performance may delineate potential cause-effect relationships with ecosystem functioning [1,2]. The knowledge gained from prediction can be used to generate hypotheses on the role of key species. At ecosystem level, designing effective control strategies for key species holds promise to manage the community towards selection of the target processes, which is crucial for microbiota-based biotechnologies [3-5].

Our goals were to investigate how environmental manipulations affect ecosystem functioning and to predict performance metrics of the quantifiable biological processes by following microbial community dynamics. Model ecosystems offer the opportunity to link microbial diversity and ecosystem functioning in a quantifiable and predictable way [6-8]. Such simplified ecosystems can be still complex regarding microbial interactions and involved metabolic processes [6]. Here, we used anaerobic fermentation reactors as model ecosystems and considered microbial chain elongation (CE) as the quantifiable model ecosystem process. CE is a microbial process that produces medium-chain carboxylates (6 to 8 carbon atoms) through reverse β-oxidation [9]. Recently we enriched a mixed culture that produces *n*-butyrate (C4), *n*-caproate (C6) and *n*-caprylate (C8) from xylan and lactate in a daily-fed reactor system [10], to simulate the feedstock conditions of anaerobic fermentation of ensiled plant biomass [11]. For this bioprocess to be viable, it needs to include diverse functions such as xylan hydrolysis, xylose fermentation and CE with lactate as electron donor. Mixed culture fermentation is characterised by different trophic groups that may cooperate or compete with each other to metabolise complex substrates [9]. Species involved in these interactions can

drive shifts in community structure and function [1]. During the long-term stable reactor operation, the community developed towards predominating C4 and biomass production at the cost of C6/C8 production [10]. We wanted to explore how process parameter changes shape the existing microbiota to optimise the process towards the target products C6 and C8. The current study was conducted on the enriched chain-elongating microbiota in two parallel bioreactors. To promote C6 and C8 production and enrich the functional groups relevant to process performance, we reduced the hydraulic retention time (HRT). HRT refers to the average time soluble compounds reside in the bioreactor. Shortening the HRT is a common operation-based strategy for increasing C6/C8 production [12-16] and a key factor influencing microbial diversity [17]. It is relevant to the microbial growth rate in reactors without biomass retention, and it affects biomass concentration and community composition [18]. Following variations in diversity induced by HRT reduction, we tested if productivity and yield of the target products (C6 and C8) could be predicted by using machine learning. To provide insight into the dynamics of community structure and function, we measured process performance and collected samples for community analysis using high-throughput sequencing of the 16S rRNA gene. Community analysis using 16S rRNA amplicon sequencing data combined with environmental variables can reveal relationships between microbial communities and ecosystem functioning. For example, Werner et al. demonstrated strong relationships between the phylogenetic community structure, reflected by time-resolved 16S rRNA amplicon data, and the methanogenic activity in full-scale anaerobic digesters, by applying constrained ordination [19]. Predictive analytics using machine learning has shown promise in microbiota-based biotechnologies [6,20,21]. We chose the random forest algorithm because it runs efficiently and accurately on high-dimensional datasets with multi-features, and it avoids overfitting, particularly when using different training and test datasets [22]. Our random forest analysis consisted of two parts. First, we performed feature selection identifying Amplicon Sequence Variants (ASVs) that would be relevant to community dynamics caused by HRT reduction. Next, we trained the algorithm with these features (hereafter, HRT bioindicators) that later were used to predict the production of C6 and C8.

Methods

Reactor operation and process monitoring. The inoculum was initially taken from a continuous lab-scale bioreactor that produced C6 and C8 by anaerobic fermentation of lactate-rich corn silage [11]. Enrichment was performed in a reactor that was daily fed with mineral medium (pH 5.5; Additional file 1: Table S1) containing water-soluble xylan (more than 95% xylooligosaccharides, from corncob; Roth, Karlsruhe, Germany) and lactic acid (85%, FCC grade; Sigma Aldrich, St. Louis, USA) as defined carbon sources and produced C4, C6 and C8 over 150 days [10]. For the present study, two 1-L bioreactors (A and B; BIOSTAT® A plus, Sartorius AG, Göttingen, Germany) were filled up with 0.5 L of the enriched culture. Both bioreactors were daily fed with 0.125 L medium containing 1.47 g lactic acid and 1.25 g xylan, without withdrawing effluent. After four days the contents of both bioreactors were mixed by pumping them three times from bioreactor A to B and back while keeping anoxic conditions. Eventually, they were equally distributed to both bioreactors, which is considered the starting point (day 0) of the experiment.

We employed semi-continuous stirred tank reactors for anaerobic fermentation, which were operated at 38 ± 1°C and constantly stirred at 150 rpm. The pH of the reactor broth was automatically controlled at 5.5 by addition of 1 M NaOH. For each bioreactor, the produced gas was collected in a coated aluminium foil bag that also served for compensating underpressure in the reactor system. The bag was connected after a MilliGascounter® (MGC-1; Ritter, Bochum, Germany) that measured on-line the volume of the produced gas. A gas-sample septum was placed in the gas pipe of each bioreactor.

In the beginning, both bioreactors were operated as replicates with an equal HRT of 8 days. For daily feeding, 1.47 g lactic acid and 1.25 g xylan were supplied in mineral medium. After 51 days, we gradually decreased the HRT of bioreactor A from 8 days to 6 days, and further to 4 days and 2 days as shown in Additional file 1: Table S2. Next, we shortened the HRT of bioreactor B from 8 days to 2 days in a fast transition mode and with the same substrate load as in bioreactor A. Considering the effect of time on community assembly, we conducted unequal HRT changes in two bioreactors and aimed to delineate the

model prediction strength with the two different datasets. Finally, both bioreactors were operated at an HRT of 2 days until day 211.

Gas samples were taken through the septum twice per week. Samples for determining cell mass concentrations were collected from the reactor effluent. Concentrations of xylan, carboxylates and alcohols were measured in the effluent supernatants [10]. In total, samples were collected on 59 time points for each bioreactor. At the beginning and the end of the experiment, pelleted biomass from the effluent was used to determine the cell dry mass as previously described [10]. For microbial community analysis, pelleted cells from 2 mL effluent were washed with 100 mM Tris-HCl pH 8.5 and stored at -20˚C until DNA extraction.

Analytical methods. Daily produced gas volume was monitored with the MGC-1 and normalised to standard pressure and temperature [23]. Gas composition (H_2, CO_2, N_2, O_2 and CH_4) was determined by gas chromatography in triplicate [24]. Concentrations of carboxylates and alcohols were analysed in triplicate by gas chromatography [10]. Concentration of xylan was measured by a modified dinitrosalicylic acid reagent method [10]. Cell mass concentration was calculated from optical density (OD) values that were correlated with the cell dry mass [10]. The calculated mean correlation coefficients were 1 OD_{600} = 0.548 g L^{-1} for bioreactor A and 1 OD_{600} = 0.537 g L^{-1} for bioreactor B.

Microbial community analysis. Total DNA was isolated from frozen cell pellets using the NucleoSpin® Microbial DNA Kit (Macherey-Nagel, Düren, Germany). Methods for DNA quantification and quality control were as described before [25]. For high-throughput amplicon sequencing, V3-V4 regions of the 16S rRNA genes were PCR-amplified using primers 341f and 785r [26]. Sequencing was performed on the Illumina Miseq platform (Miseq Reagent Kit v3; 2 × 300 bp). A total of 12,168,404 sequences ranging from 57,612 to 389,963 pairs of reads per sample (mean: 135,205; median: 122,367) were obtained.

The demultiplexed sequence data were processed with the QIIME 2 v2019.7 pipeline [27] using the DADA2 plugin [28]. The DADA2 parameters were set as

follows: trim-left-f 0, trim-left-r 0, trunc-len-f 270, trunc-len-r 230, max-ee 2 and chimera-method consensus. A total of 4,194,700 sequences ranging from 13,518 to 138,498 reads per sample were retained, with a mean of 46,608 reads per sample. The generated feature table indicates the frequency of each ASV clustered at 100% identity. Taxonomic assignment was done with a naïve Bayes classifier trained on 16S rRNA gene sequences of the database MiDAS 2.1 [29], and curated using the RDP Classifier 2.2 with a confidence threshold of 80% [30]. For downstream analyses, ASVs of all samples were rarefied to a sequencing depth of 13,518 reads (rarefaction curve reached the plateau, Additional file 1: Figure S1). We obtained a total of 71 unique ASVs in 90 samples.

Alpha diversity based on rarefied ASV data was evaluated by the observed ASV counts and the Shannon index [31], which were determined using the R package phyloseq v1.30.0 [32]. Dissimilarities in bacterial community composition (beta-diversity) were calculated using Bray-Curtis distance [33] based on rarefied ASV abundances and visualised as nonmetric multidimensional scaling (NMDS) plots. Statistical analyses of beta-diversity results were performed using permutational multivariate analysis of variance (PERMANOVA) [34] in the R package "vegan" (v2.5.6, "adonis" function, Monto-Carlo test with 1000 permutations); P values were adjusted for multiple comparisons using the false discovery rate (FDR) method [35].

Network analysis. The co-occurrence network analysis was performed using the method described by Ju et al. [36]. Briefly, we constructed a correlation matrix by computing possible pairwise Spearman's rank correlations using the rarefied ASV abundances and abiotic parameters (HRT; concentrations of C4, C6 and C8 and lactate; productivities and yields of C4, C6 and C8). Correlation coefficients below -0.7 or above 0.7 and adjusted P-values (FDR method) lower than 0.05 were considered statistically robust. Network visualisation and topological features analysis were conducted in Gephi (v0.9.2) [37].

16S rRNA phylogenetic analysis. The 16S rRNA gene sequences of ASVs were aligned using the SINA alignment algorithm [38] via the SILVA web interface [39]. We additionally used SINA to search and classify the sequences

with the least common ancestor method based on the SILVA taxonomy. For each query sequence, the minimum identity was set to 0.95 and the five nearest neighbours were considered. The tree was reconstructed based on the aligned sequences and their neighbours, with RAxML using the GTRCAT model of evolution. Later only ASV species of this study were kept in the generated tree for an easier viewing. The tree was visualised using iTOL [40].

Metagenomic analysis. Six samples were selected for whole-genome sequencing, which was performed by StarSEQ GmbH (Mainz, Germany), using the Illumina NextSeq 500 system (NEBNext Ultra II FS DNA library prep kit; 2 × 150 bp) with at a minimum of 20 million reads per library generated. Quality checking and reads trimming were performed using metaWRAP (v0.7, raw read QC module) [41] and TrimGalore (v0.4.3) [42]. Reads of human origin were discriminated from microbial reads using BMTagger (v3.101) [43]. All adapters were removed and the resulting reads were assembled using metaSPAdes (v3.11.1) [44]. Paired-end reads were aligned back to the assembly using BWA (v0.7.15, mem algorithm) [45]. Binning of assembled contigs was performed using the metaWRAP modules metaBAT (2.12.1) [46], MaxBin (2.2.4) [47] and CONCOCT (1.0.0) [48]. The metaWRAP-Bin_refinement module was applied to separate the overlaps between two bins. Quality of metagenome-assembled genomes (MAGs) was checked using CheckM (v1.0.7) [49]. MAGs were classified in high or medium quality regarding completeness, contamination, quality score (completeness - 5 × contamination) and strain heterogeneity [50]. The following thresholds were used for high quality: quality score > 50, completeness > 80, contamination < 5 and strain heterogeneity < 50; and for medium quality: quality score > 50, completeness > 50 and contamination < 10. One bin with lower quality was removed from the analysis. The taxonomy was assigned using GTDB-Tk (v0.3.2) [51]. Genome metrics were calculated with the statswrapper tool in the BBTools suite [52]. A phylogenomic tree based on Mash distances was generated with Mashtree (V1.1.2) [53] and visualised in iTOL [40]. Miscellaneous visualisations of the dataset metrics were performed in R with the packages ggplot2 (v3.3.0) and DataExplorer (v0.8.1). Species differentiation was performed using fastANI [54] and aniSplitter.R (http://github.com/felipborim789/aniSplitter/). Genomes were annotated with

Prokka (v1.14.6) [55]. Functional annotation of genes relevant to xylan hydrolysis, xylose fermentation and chain elongation was curated using Swiss-Prot, COG and GenBank [56-58].

Determining bioindicators of HRT changes. The HRT bioindicators were determined using the random forest algorithm (randomForest R package, v4.6-14) [59]. ASV relative abundances were used as features to train and test the random forest classifier. Considering how we replicated the HRT changing mode in both bioreactors (Additional file 1: Table S2), the whole operation period was divided into four sampling intervals: 0-50 days, 51-100 days, 101-140 days and 141-211 days. Based on the results of community analysis, we chose the ASV data of both bioreactors in the sampling intervals of 0-50 days and 141-211 days to determine the HRT bioindicators, and we used data of all samples in the four HRT phases as controls. To evaluate the robustness of the predictions, we trained the classifier with ASV data of one bioreactor and tested in the other bioreactor and vice versa. For random forest classification analysis, importance of the different features (ASVs) was measured by the Gini index (mean decrease in Gini, default in randomForest R package; where larger values indicate a variable to be more important for accurate classification [60]).

The random forest classifier was trained on the training set, with 2,000 trees and 40 variables (with lowest out-of-bag estimated error rates achieved) being selected randomly for each tree. Explained variance (% Var. explained in R) was used to measure the model performance on the training set [59]. We predicted the accuracy by measuring how well the features can classify the HRT phases on the test set [60]. We first computed the feature importance of all 71 ASVs. Then at each step, the ASVs having the smallest importance were eliminated and a new forest was built with the remaining ASVs. For both bioreactors, the features were selected when their Gini scores were higher than 1% of the sum of the Gini scores of all ASVs (Additional file 1: Table S3). Finally, we selected the 15 top-ranked ASVs leading to the model of smallest error rate for classifying the HRT phases of 8 days and 2 days. In each bioreactor, the 15 ASVs that best discriminated between HRT phases were referred to as A-HRT bioindicators or B-HRT bioindicators (bioreactors A and B, respectively). ASVs

common to both sets were defined as HRT bioindicators (workflow of random forest classification in Additional file 1: Figure S2).

Quantitative predictions based on HRT and non-HRT bioindicators. The process parameters specified as concentrations of lactate, C4, C6 and C8, and productivities as well as yields of C4, C6 and C8 were the prediction objects. Here, the relevance of the different ASVs to the prediction was determined by residual sum of squares (IncNodePurity, default in randomForest) for the regressions. Explained variance (% Var. explained in R) was used to measure the model performance on the training set [59]. We predicted the accuracy by measuring how well the features can explain the variance of these process parameters on the test set [60].

We performed the quantitative prediction by applying a two-step regression analysis (workflow in Additional file 1: Figure S3). First, HRT bioindicators were used to predict the data of different process parameters in the sampling intervals of 0-50 days and 141-211 days. Data of all samples in the four HRT phases were considered as controls. Random forest regressors were trained as follows: relative abundance dataset of bioreactor A was used as training set and that of bioreactor B was used as test set and vice versa; 2,000 trees and four out of 11 features were selected randomly for each tree.

Considering community assembly caused by time, we determined the ASVs (non-HRT bioindicators) that could predict the numeric values of each process parameter, using data of samples in the intervals of 0-50 days and 141-211 days. For each process parameter, we started with computing the feature importance of all ASVs and further selected the 15 top-rated ASVs as the bioindicators of this non-HRT parameter. The model was trained as follows: datasets of bioreactors A and B were independently used for training and testing; 2,000 trees and five out of 15 features were selected randomly for each tree. As controls, we used the non-HRT bioindicators of each parameter to predict the corresponding data of all samples in the four HRT phases. The final set of ASVs presented in HRT bioindicators and not in non-HRT bioindicators were considered HRT bioindicators irrespective of time.

Evaluating prediction accuracy. When in both training sets the HRT bioindicators and non-HRT bioindicators explained more than 80% of the variance in a process parameter, we proceeded only with those parameters. To compare the predicted and measured values for these process parameters, we considered the following performance metrics for reflecting the error of the model in predicting consecutive data: relative root mean square error (RRMSE, cutoff < 10%); R squared, slope and intercept of the least squares line of best fit. The final values of RRMSE were averaged among the 100 random forest replicates, with four ASVs for HRT bioindicators and five for non-HRT bioindicators randomly sampled at each replicate.

Results and discussion

Effects of HRT decrease on process performance and microbial diversity. The progressive HRT decrease from 8 to 2 days increased the C6 and C8 productivities and yields in two independent bioreactors (Figure 1). We first shortened the HRT to 6 days and then to 4 days in bioreactor A, which allowed the reactor microbiota to adapt to the new conditions and improved productivities of C4, C6 and C8 (Figure 1a). Further HRT decrease to 2 days confirmed the increasing trend in productivity. At the end of the 2-day HRT period in bioreactor A, we achieved the highest productivities (mmol C L^{-1} d^{-1}) of C4, C6 and C8 up to 115.0, 64.1 and 5.9, respectively. To confirm the observed effects of HRT shortening on the CE process and reactor microbiota, we executed a fast transition mode in bioreactor B and generated a different dataset from the parallel system. Comparable increases in productivity were observed (Figure 1b). We obtained maximum productivities (mmol C L^{-1} d^{-1}) of C4 up to 102.4, C6 up to 62.9 and C8 up to 7.0. The C6 and C8 yields (in terms of C mole product to transferred substrate ratio) increased along with decreasing HRT at the cost of C4 yield. Compared with yields at the 8-day HRT, C6 and C8 yields were higher and the C4 yield was lower in both bioreactors at the 2-day HRT (Figure 1 and Additional file 1: Table S4). Our results suggest that the shorter HRT favoured lactate-based CE producing C6 and C8 over C4 production. C4 can be produced by CE of acetate but also from sugars by butyric acid fermentation [61]. Decreasing the HRT to 2 days led to the

accumulation of lactate and fluctuations of the C4, C6 and C8 production, which lasted longer than 22 HRTs in bioreactor A (Figure 1a). Lactate concentrations were highly correlated with C4 fluctuations (Spearman Rho = -0.90, $P < 0.05$) and C6 concentrations (Rho = -0.89, $P < 0.05$), which reflects how lactate was produced and converted by the reactor microbiota. The HRT reduction resulted in higher gas production and hydrogen content (Additional file 1: Figure S4). Besides, an increase in cell mass production (Additional file 1: Figure S5) suggests a facilitating effect of short HRT on the growth of enriched populations with desirable activities, i.e. more biocatalysts were available in the high C6/C8 production phase.

Decreasing the HRT affected the composition and diversity of the reactor microbiota. Changes in relative abundance of ASVs categorised from phylum to genus between the HRT of 8 days and 2 days are shown in Additional file 1: Figure S6. Alpha diversity metrics showed significantly lower observed ASV counts (pairwise t-test, $P < 0.05$) and higher Shannon index values (pairwise t-test, $P < 0.05$) for HRT of 8 days compared with 2 days (Additional file 1: Figure S7). Beta diversity analysis revealed a significant difference between the communities at different HRTs (PERMANOVA; Pseudo-F = 103.1, $P < 0.001$) but no significant difference between the communities in both reactors at the same HRT (Pseudo-F = 3.3, $P > 0.05$) (Figure 2).

HRT bioindicators predicting process performance. To determine HRT bioindicators, we used HRT of 8 days and 2 days as classes for the random forest classification model and relative abundances of ASVs as the features. To delineate the model prediction strength, we used one reactor dataset to train the model while testing predictions with the other and vice versa. Feature selection based on the random forest classifier with its associated Gini index has shown abilities to identify optimal feature subsets in high-dimensional data [62]. Based on higher than 1% of the mean decrease in Gini scores for both reactors in the prediction accuracy of HRT phases, we selected 15 top-ranked ASVs that would give the best discrimination between HRT phases. The 15 ASVs most relevant to HRT changes were defined as "A- or B-HRT bioindicators", potentially reflecting the key species correlating with

HRT changes in either bioreactor (feature importance in Figure 3). The two bioreactors shared 11 HRT bioindicators.

To answer the question whether HRT bioindicators can be used to predict process performance in terms of C6 and C8 production, we performed a random forest regression analysis in two steps. HRT bioindicators were first chosen as features to train the model. Considering community assembly caused by time, we then determined 15 ASVs most relevant to each non-HRT process parameter (i.e., concentrations of lactate, C4, C6 and C8; productivities and yields of C4, C6 and C8; hereafter, non-HRT bioindicators). Datasets from bioreactors A and B were trained and tested independently. When in both reactors the HRT and non-HRT bioindicators accounted for more than 80% of the variance in a process parameter, we proceeded only with those parameters. In our case, the model could explain more than 80% of the variance in C6 and C8 productivities (Additional file 1: Tables S5-S6).

We evaluated the prediction performance of the model by comparing the predicted and measured values of process parameters. RRMSE was used as the performance metric to reflect the model error in predicting quantitative data of C6/C8 productivity. Our results showed that the C6 and C8 productivities of both bioreactors at the HRT of 8 days and 2 days could be accurately predicted (Figure 4 for HRT bioindicators and Additional file 1: Figure S8 for non-HRT bioindicators). We further tested samples in all HRT phases with HRT and non-HRT bioindicators. The C6 and C8 productivities were also accurately predicted (RRMSE < 6%, Additional file 1: Figures S9-S10). Therefore, we considered HRT bioindicators irrespective of time as the ASVs presented in HRT bioindicators and not in non-HRT bioindicators (feature importance in Additional file 1: Figures S11-S12). Interestingly, the same four ASVs assigned to the genera *Olsenella*, *Lactobacillus*, *Syntrophococcus* and *Clostridium* IV were identified for C6 and C8 productivity (Figure 5). We thus hypothesise that species represented by these four ASVs determined the increased C6/C8 productivities in the CE process manipulated by changing operational conditions – shortening the HRT.

Functional role of HRT bioindicators. Combined with metagenomics, species of HRT bioindicators irrespective of time indicated their roles in driving the catabolism of xylan and lactate to C6/C8 (Figure 6). Among 108 MAGs (dereplicated into 29 species; Figure 7 and Additional file 2), we recovered 12 species with similar phylogenies as the four genera (Table 1). In view of the fermentation process, we annotated the genetic potential for xylan hydrolysis, xylose fermentation and CE with lactate (Additional file 1: Figure S13 and Additional files 3-6). Specifically, *Clostridium* IV species were reported as lactate-based chain-elongating bacteria [63]. Our results suggest that four *Clostridium* IV species (*Acutalibacteraceae* spp. according to GTDB-Tk) can convert lactate to C6/C8. Two *Syntrophococcus* species (*Eubacterium*_H spp. according to EZBioCloud [64]) are potential C6/C8-producers as they hold complete gene sets encoding enzyme complexes that catalyse CE reactions. This genetic potential was also found in genomes of closely related *Syntrophococcus* species (*Eubacterium cellulosolvens* according to EZBioCloud; Additional file 6), which was not described before. Lactate formation from xylose by lactic acid bacteria can enhance CE by providing additional electron donors [23,65-68]. A recent study reported an enriched community dominated by *Lactobacillus* and chain-elongating species, and their co-occurrence suggested lactate produced by *Lactobacillus* to be a key intermediate for C6/C8 production [69]. Network analysis of our previous study [10] revealed the co-occurrence of *Olsenella* with potential chain-elongating species. Species of *Lactobacillus* and *Olsenella* are potential xylose-consuming lactate producers (Figure 6b). Genes encoding xylanases were not found in *Lactobacillus* MAGs but in those assigned to other bioindicators (Figure 6a). Taken together, the delineated synergy effects between these bioindicator species suggest a division of labour with mutual benefits, converting xylan and lactate to C6/C8. A correlation network shows HRT, C6 and C8 productivity being the most highly connected nodes (Additional file 1: Figure S14). Their co-occurrence with ASVs assigned to *Clostridium* IV, *Olsenella* and *Syntrophococcus* indicates strong associations among these taxa, the changed environment and corresponding functions. The predictability of C6 and C8 productivities was relatively poor when using only the four HRT bioindicators irrespective of time (Additional file 1: Figure S15). Besides, we found

redundancy in the main functions of catabolising xylan and lactate to C4, C6 and C8 (Figure 6), with the relevant HRT bioindicators increasing in relative abundances (Additional file 1: Figure S16). Thus, the involved metabolic pathways can be strongly coupled to HRT decreases. The genetic potential overlaps with other distinct taxa of the reactor microbiota, suggesting that HRT bioindicators might be key species of the process, but ecological interactions with other species are critical to ensure the C6/C8 production (functional annotations of xylose fermentation and chain elongation in Additional files 5-6).

Conclusions

Our approach enabled the quantitative prediction of process performance in the anaerobic bioreactor system (Figure 8). In artificial ecosystems with well-controlled conditions (temperature, pH and no immigration of other microbes; Figure 8a), HRT was the most influencing factor controlling community assembly (Figure 8b). However, we cannot exclude the impact of other deterministic factors like microbial interactions within temporal patterns, particularly for such a long-term reactor operation. Effects of compositional stochasticity on community assembly also need to be considered [70,71]. Further studies on these ecological principles will help manage reactor microbiota towards beneficial traits, such as high specificities for C6/C8 production.

The continuous reactor systems with enrichment cultures enabled to select communities with desired CE functions (i.e., high C6 and C8 productivities), and to demonstrate that 16S rRNA amplicon sequencing data can be used to predict CE process performance quantitatively (> 90% accuracy). The described machine learning framework (Figure 8c) may be suitable for other ecosystem processes and more complex communities. For that, it would be necessary to design experiments with (i) sufficient temporal and/or spatial resolution, (ii) parallel sampling for amplicon sequencing data and metadata from desired ecosystem processes, and (iii) correlation of phylogenetic diversity with the ecosystem processes. Our approach was based on phylogenetic diversity that in some ecosystems may correlate with ecosystem processes where microbiota perform key functions. Our general methodology can be adapted to

other data types, such as metagenomes, metatranscriptomes, metaproteomes or metabolomes, and it opens new doors for prediction and hypothesis testing in microbial ecology.

Declarations

Availability of data and materials

All data described in this manuscript are present in the paper and/or the Supplementary material. Both amplicon sequencing data (ERR4158761 to ERR4158850) and metagenome sequencing data (ERR4183110 to ERR4183115) have been deposited to the ENA database under study no. PRJEB38353. The MAGs are publicly available in the ENA under the sample accession no. ERS4594296 to ERS4594324.

Funding

Authors' contributions

B.L., H.S., J.S., S.K. and U.R. designed the study and the experiments. B.L. performed the experiments and analysed the reactor data as well as sequencing data. B.L., J.S. and U.R. performed the machine learning analysis. B.L., H.S., J.S., S.G.S., S.K. and U.R. contributed to data analysis and interpretation. H.H. 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

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol. 2018;16:567–576.

2. de los Reyes FL. Challenges in determining causation in structure-function studies using molecular biological techniques. Water Res. 2010;44:4948–57.

3. Koch C, Müller S, Harms H, Harnisch F. Microbiomes in bioenergy production: from analysis to management. Curr Opin Biotechnol. 2014;27:65–72.

4. Verstraete W, Wittebolle L, Heylen K, Vanparys B, de Vos P, van de Wiele T, et al. Microbial resource management: the road to go for environmental biotechnology. Eng Life Sci. 2007;2:117–26.

5. Kleerebezem R, van Loosdrecht MC. Mixed culture biotechnology for bioenergy production. Curr Opin Biotechnol. 2007;18:207–12.

6. Lawson CE, Harcombe WR, Hatzenpichler R. Common principles and best practices for engineering microbiomes. Nat Rev Microbiol. 2019;17:725–41.

7. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Science. 2018;361:469–74.

8. Zuñiga C, Li CT, Yu G, Al-Bassam MM, Li T, Jiang L, et al. Environmental stimuli drive a transition from cooperation to competition in synthetic phototrophic communities. Nat Microbiol. 2019;4:2184–91.

9. 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.

10. 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.

11. 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.

12. Kucek LA, 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.

13. Kucek LA, Nguyen M, Angenent LT. Conversion of L-lactate into *n*-caproate by a continuously fed reactor microbiome. Water Res. 2016;93:163–71.

14. Duber A, Jaroszynski L, Zagrodnik R, Chwialkowska J, Juzwa W, Ciesielski S, et al. Exploiting the real wastewater potential for resource recovery – *n*-caproate production from acid whey. Green Chem. 2018;20:3790–803.

15. Grootscholten TIM, Steinbusch KJJ, Hamelers HVM, Buisman CJN. Improving medium chain fatty acid productivity using chain elongation by reducing the hydraulic retention time in an upflow anaerobic filter. Bioresour Technol. 2013;136:735–8.

16. Nzeteu CO, Trego AC, Abram F, O'Flaherty V. Reproducible, high-yielding, biological caproate production from food waste using a single-phase anaerobic reactor system. Biotechnol Biofuels. 2018;11:108.

17. Mansfeldt C, Achermann S, Men Y, Walser JC, Villez K, Joss A, et al. Microbial residence time is a controlling parameter of the taxonomic composition and functional profile of microbial communities. ISME J. 2019;13:1589–601.

18. Bonk F, Popp D, Weinrich S, Sträuber H, Becker D, Kleinsteuber S, et al. Determination of microbial maintenance in acetogenesis and methanogenesis by experimental and modelling techniques. Front Microbiol. 2019;10:166.

19. Werner JJ, Knights D, Garcia ML, Scalfone NB, Smith S, Yarasheski K, et al. Bacterial community structures are unique and resilient in full-scale bioenergy systems. Proc Natl Acad Sci USA. 2011;108:4158–63.

20. Oyetunde T, Bao FS, Chen JW, Martin HG, Tang YJ. Leveraging knowledge engineering and machine learning for microbial bio-manufacturing. Biotechnol Adv. 2018;36:1308–15.

21. Lopatkin AJ, Collins JJ. Predictive biology: modelling, understanding and harnessing microbial complexity. Nat Rev Microbiol. 2020;18:507-20.

22. Breiman L. Random forests. Mach Learn. 2001;45:5–32.

23. Sträuber H, Bühligen F, Kleinsteuber S, Dittrich-Zechendorf M. Carboxylic acid production from ensiled crops in anaerobic solid-state fermentation – trace elements as pH controlling agents support microbial chain elongation with lactic acid. Eng Life Sci. 2018;0:447–58.

24. 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.

25. Lucas R, Kuchenbuch A, Fetzer I, Harms H, Kleinsteuber S. Long-term monitoring reveals stable and remarkably similar microbial communities in parallel full-scale biogas reactors digesting energy crops. FEMS Microbiol Ecol. 2015;91:fiv004.

26. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41:e1.

27. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Chase J, Cope EK, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:852–7.

28. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

29. McIlroy SJ, Kirkegaard RH, McIlroy B, Nierychlo M, Kristensen JM, Karst SM, et al. MiDAS 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database. 2017;2017:bax016.

30. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences. Appl Environ Microbiol. 2007;73:5261–7.

31. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27:379–423.

32. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217.

33. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr. 1957;27:325–49.

34. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32–46.

35. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple Testing. J R Stat Soc B. 1995;57:289–300.

36. Ju F, Xia Y, Guo F, Wang Z, Zhang T. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. Environ Microbiol. 2014;16:2421–32.

37. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Int AAAI Conf Weblogs Soc Media. 2009;8:361–2.

38. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.

39. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal

RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188–96.

40. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.

41. Uritskiy G V., Diruggiero J, Taylor J. MetaWRAP – a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.

42. Galore K. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

43. Rotmistrovsky, K. Agarwala R. BMTagger: best match tagger for removing human reads from metagenomics datasets. 2010. Available from: ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/

44. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.

45. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

46. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.

47. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7.

48. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144–6.

49. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.

50. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–42.

51. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36:1925–7.

52. Bushnell B. BBMap short read aligner, and other bioinformatic tools. Available from: http://sourceforge.net/projects/bbmap

53. Katz L, Griswold T, Morrison S, Caravas J, Zhang S, Bakker H, et al. Mashtree: a rapid comparison of whole genome sequence files. J Open Source Softw. 2019;4:1762.

54. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.

55. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

56. Bateman A. UnitProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.

57. 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.

58. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2016;44:D67-72.

59. Liaw A, Wiener M. Classification and regression with random forest. R News. 2002;2:18–22.

60. Huang BFF, Boutros PC. The parameter sensitivity of random forests. BMC Bioinformatics. 2016;17:331.

61. Temudo MF, Mato T, Kleerebezem R, Van Loosdrecht MCM. Xylose anaerobic conversion by open-mixed cultures. Appl Microbiol Biotechnol. 2009;82:231–9.

62. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics. 2009;10:213.

63. 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.

64. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67:1613–7.

65. Xu J, Hao J, Guzman JJL, Spirito CM, Harroff LA, Angenent LT. Temperature-phased conversion of acid whey waste into medium-chain carboxylic acids via lactic acid: no external e-donor. Joule. 2018;2:1–16.

66. Scarborough MJ, Lynch G, Dickson M, McGee M, Donohue TJ, Noguera DR. Increasing the economic value of lignocellulosic stillage through medium-chain fatty acid production. Biotechnol Biofuels. 2018;11:200.

67. Khor WC, Andersen S, Vervaeren H, Rabaey K. Electricity-assisted production of caproic acid from grass. Biotechnol Biofuels. 2017;10:180.

68. Andersen SJ, de Groof V, Khor WC, Roume H, Props R, Coma M, et al. A *Clostridium* group IV species dominates and suppresses a mixed culture fermentation by tolerance to medium chain fatty acids products. Front Bioeng Biotechnol. 2017;5:8.

69. Contreras-Dávila CA, Carrión VJ, Vonk VR, Buisman CNJ, Strik DPBTB. Consecutive lactate formation and chain elongation to reduce exogenous chemicals input in repeated-batch food waste fermentation. Water Res. 2020;1:115215.

70. Chase JM. Stochastic community assembly causes higher biodiversity in more productive environments. Science. 2010;328:1388–91.

71. Ofiţeru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, et al. Combined niche and neutral effects in a microbial wastewater treatment community. Proc Natl Acad Sci USA. 2010;107:15345–50.

Additional files

Additional file 1: Figure S1. Alpha rarefaction curves. **Figure S2.** Workflow of the random forest classification analysis. **Figure S3.** Workflow of a two-step random forest regression analysis. **Figure S4.** Gas production of bioreactors. **Figure S5.** Biomass production of bioreactors. **Figure S6.** Microbial community composition profiles of bioreactors. **Figure S7.** Alpha diversity metrics of bioreactor communities. **Figure S8.** Prediction results of C6 and C8 productivities using non-HRT bioindicators for considering community assembly caused by time. **Figure S9.** Prediction results of C6 and C8 productivities for all samples in the four HRT phases using HRT bioindicators. **Figure S10.** Prediction results of C6 and C8 productivities for all samples in the four HRT phases using non-HRT bioindicators for considering community assembly caused by time. **Figure S11.** Random forest feature importance of A-HRT bioindicators and B-HRT bioindicators used to predict C6 and C8 productivities. **Figure S12.** Random forest feature importance of the non-HRT bioindicators used to predict C6 and C8 productivities. **Figure S13.** Metabolic pathways involved in converting lactate and xylan to *n*-caproate and *n*-caprylate. **Figure S14.** Correlation network of environmental factors, process performance and microbial community. **Figure S15.** Prediction results of C6 and C8 productivities for all samples in the four HRT phases using the four ASVs of HRT bioindicators irrespective of time. **Figure S16.** Reducing HRT

increases abundances of HRT bioindicators driving the catabolism of xylan and lactate to *n*-caproate and *n*-caprylate. **Table S1.** Growth medium used for the reactor operation. **Table S2.** Daily feeding of bioreactors A and B during the four HRT phases. **Table S3.** Gini scores of all ASVs in the classification-based prediction of HRT phases. **Table S4.** Mean carboxylate yields (i.e. C mole product to substrate ratios) at HRTs of 8 days and 2 days (stable production period). **Table S5.** Explained variances of the training set in the regression-based prediction of process parameters using A-HRT bioindicators and B-HRT bioindicators. **Table S6.** Explained variances of the training set in the regression-based prediction of process parameters using non-HRT bioindicators for considering community assembly caused by time.

Additional file 2: Dataset S1. MAGs taxonomy and genome metrics.

Additional file 3: Dataset S2. Functional annotations of xylose fermentation for MAGs with the same taxonomy as HRT bioindicators.

Additional file 4: Dataset S3. Functional annotations of chain elongation for MAGs with the same taxonomy as HRT bioindicators.

Additional file 5: Dataset S4. Functional annotations of xylose fermentation for all MAGs.

Additional file 6: Dataset S5. Functional annotations of chain elongation for all MAGs.

Additional Datasets S1-S5 can be found at:
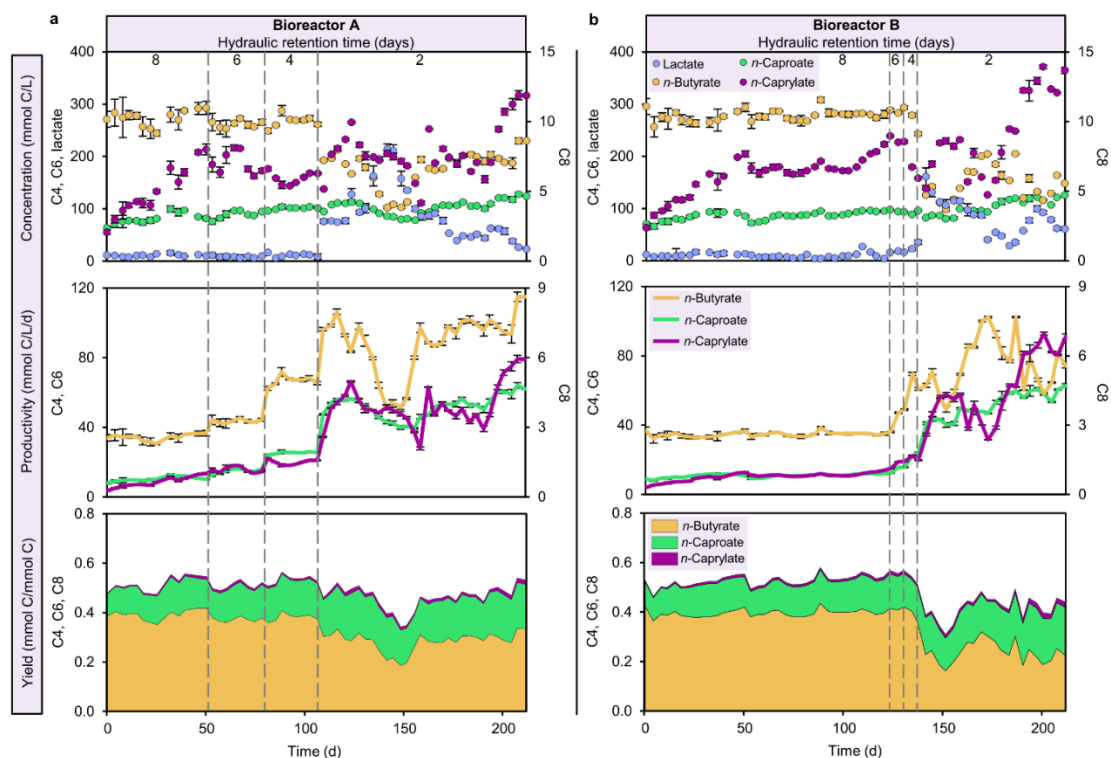
https://github.com/binliu21/carboxylate-platform/tree/SI-chapter-2.2

Figures

Figure 1. Performance of bioreactors. Concentrations of chain elongation products and lactate, as well as productivities and yields of chain elongation products in bioreactors A (**a**) and B (**b**) during the four HRT phases. Chain elongation products: C4, *n*-butyrate; C6, *n*-caproate; C8, *n*-caprylate.
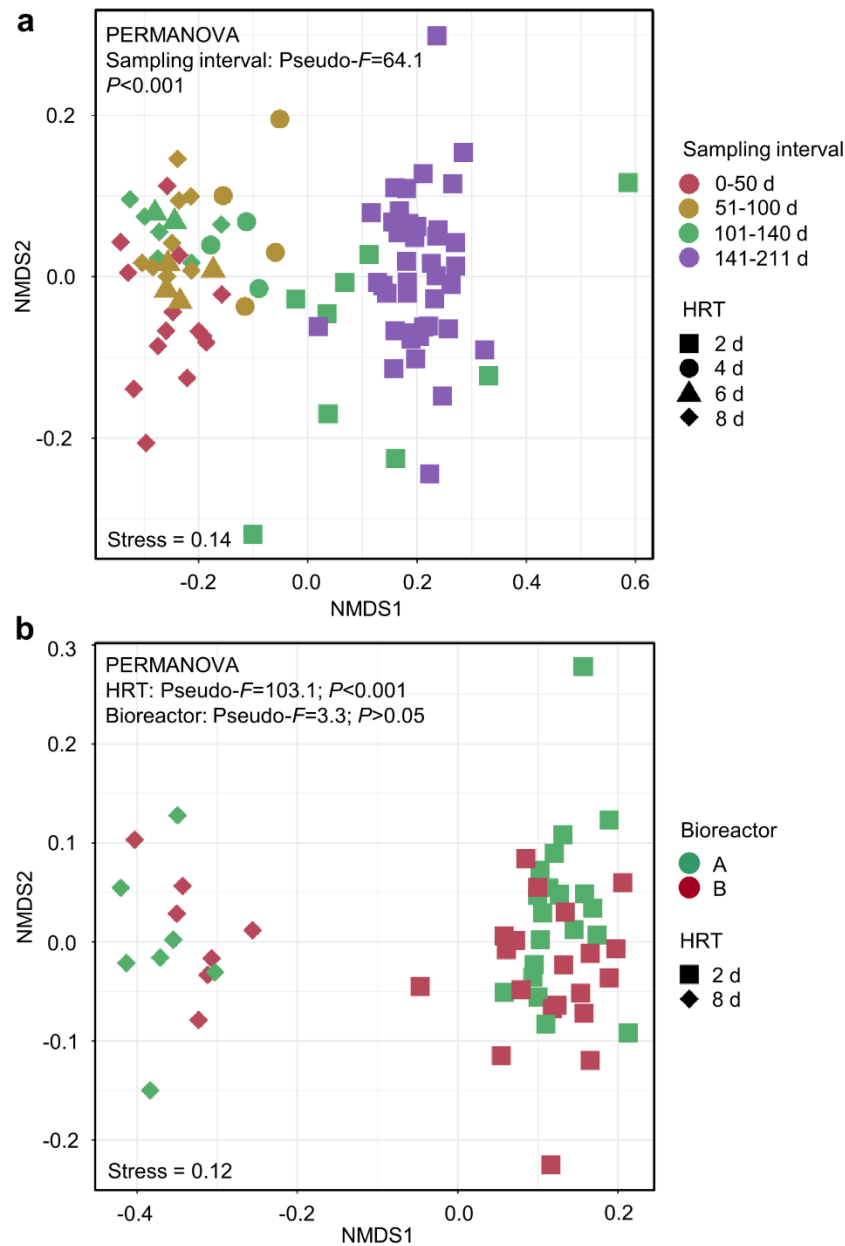
Figure 2. Dissimilarities in bacterial community composition (beta-diversity). Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities of microbial community composition in bioreactors. **a,** All samples in the four HRT phases were considered for dissimilarity calculation. **b,** Samples in the 8-day HRT phase classified to the sampling interval 0-50 days and in the 2-day HRT phase classified to the interval 141-211 days were included.
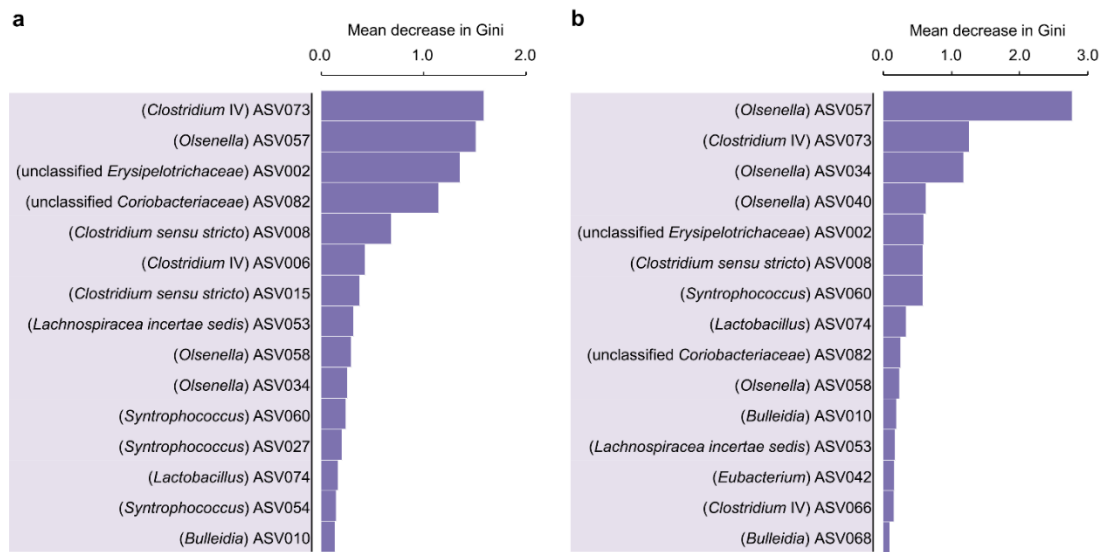
Figure 3. Random forest feature importance of ASVs used to classify the HRT phases (A-HRT bioindicators and B-HRT bioindicators). The top-ranked 15 ASVs reducing the uncertainty in the prediction of HRT phases (HRT of 8 days and 2 days). The order of features (from top to bottom) was based on their mean decrease in Gini scores, according to their ASV abundances distribution, with HRT as the response variable. **a,** Feature importance of A-HRT bioindicators. The ASV importance was calculated using the relative abundance data of bioreactor A as a training set and data of bioreactor B as a test set. **b,** Feature importance of B-HRT bioindicators. Similar to A-HRT bioindicators, ASV importance of B-HRT was calculated using the relative abundance data of bioreactor B as a training set and data of bioreactor A as a test set. The taxonomic classification of ASVs assigned at the genus level is provided in parentheses.
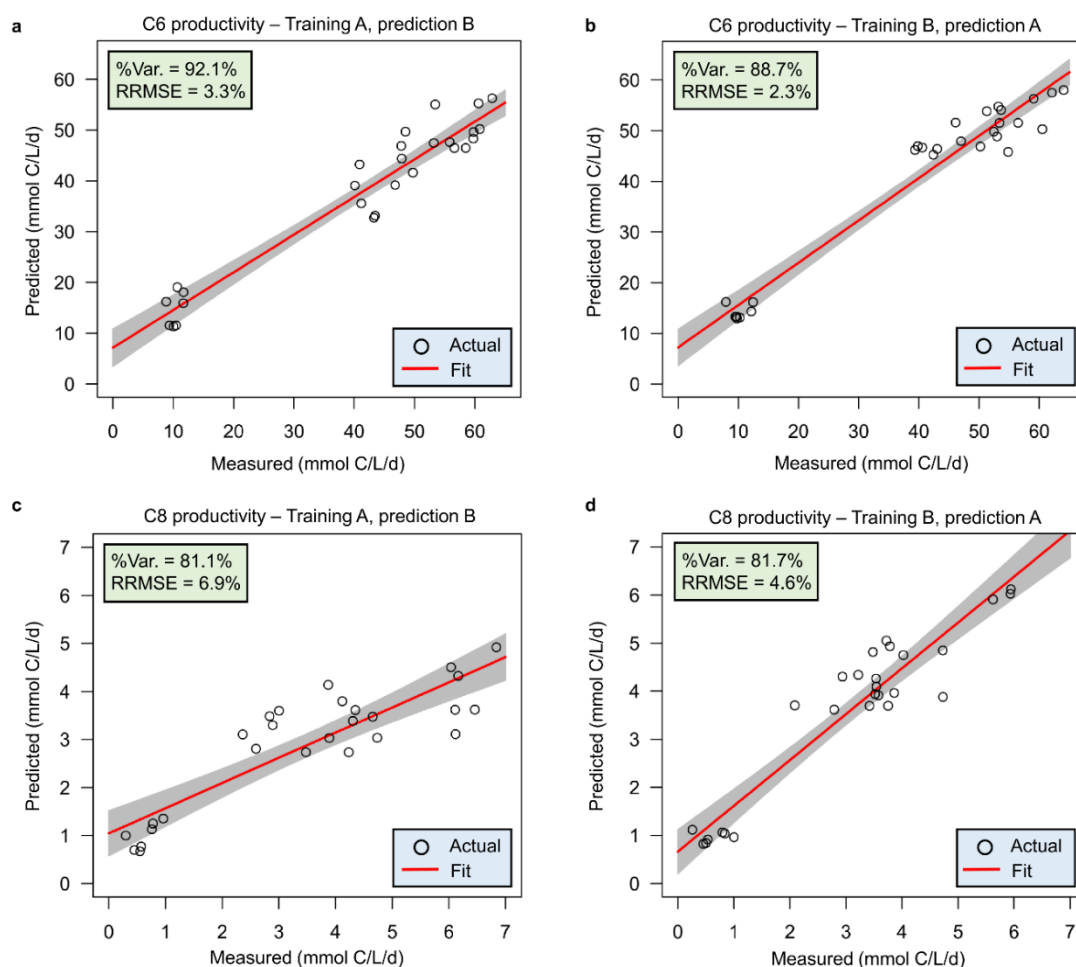
Figure 4. Prediction results of C6 and C8 productivities using HRT bioindicators. a,b, Prediction performance of C6 productivity. **c,d,** Prediction performance of C8 productivity. Results in **a** and **c** were obtained by using relative abundance data of bioreactor A for training the model and data of bioreactor B for testing. Results using the data of bioreactor B for training and bioreactor A for testing are shown in **b** and **d**. The red lines and grey shaded areas depict the best-fit trendline and the 95% confidence interval of the least-squares regression, respectively. C6, *n*-caproate; C8, *n*-caprylate; %Var., explains the variance (%) in C6/C8 productivity of the training set; RRMSE, relative root mean square error.
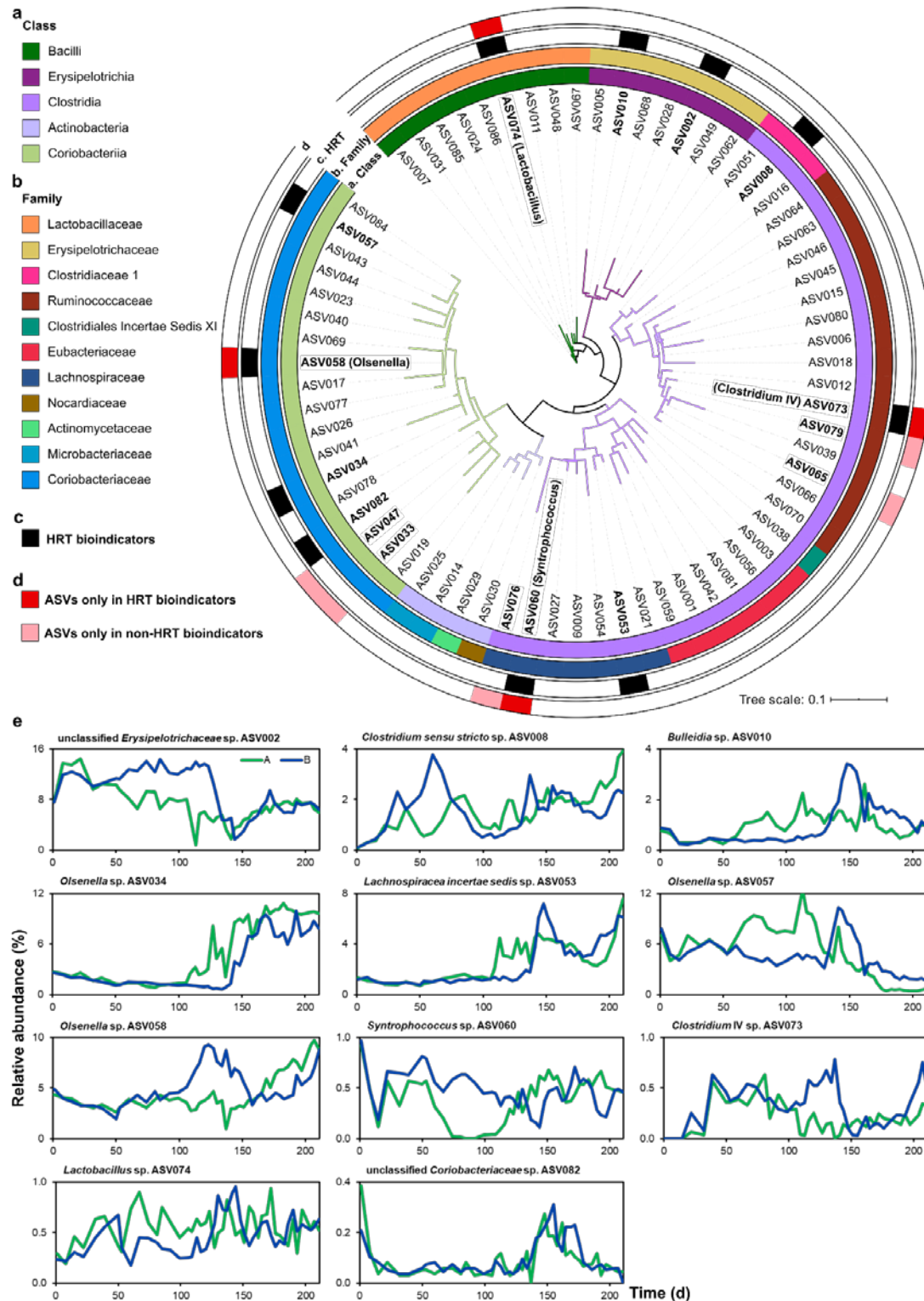
Figure 5. Phylogeny of HRT bioindicators and non-HRT bioindicators for considering community assembly caused by time. a,b, A maximum likelihood 16S rRNA gene tree showing the ASV species based on the rarefied sequencing data. ASVs are coloured according to the class (**a**, first inner ring) and family (**b**, second inner ring). **c**, The third inner ring shows the 11 HRT

bioindicators identified in both reactors for the prediction of HRT phases of 8 days and 2 days. The ASVs identified as HRT bioindicators are shown in bold. Their taxonomic assignments at the genus level are provided in the legend. **d**, The four ASVs of HRT bioindicators irrespective of time are shown in red in the outer ring. The ASVs only present in non-HRT bioindicators of C6/C8 productivity are shown in pink in the outer ring. **e**, Relative abundance dynamics of HRT bioindicators during the whole reactor operation period. In the legend, A and B stand for bioreactors A and B, respectively. The four ASVs (in bold) of HRT bioindicators, irrespective of time, assigned at the genus level are indicated in parentheses. C6, *n*-caproate; C8, *n*-caprylate.
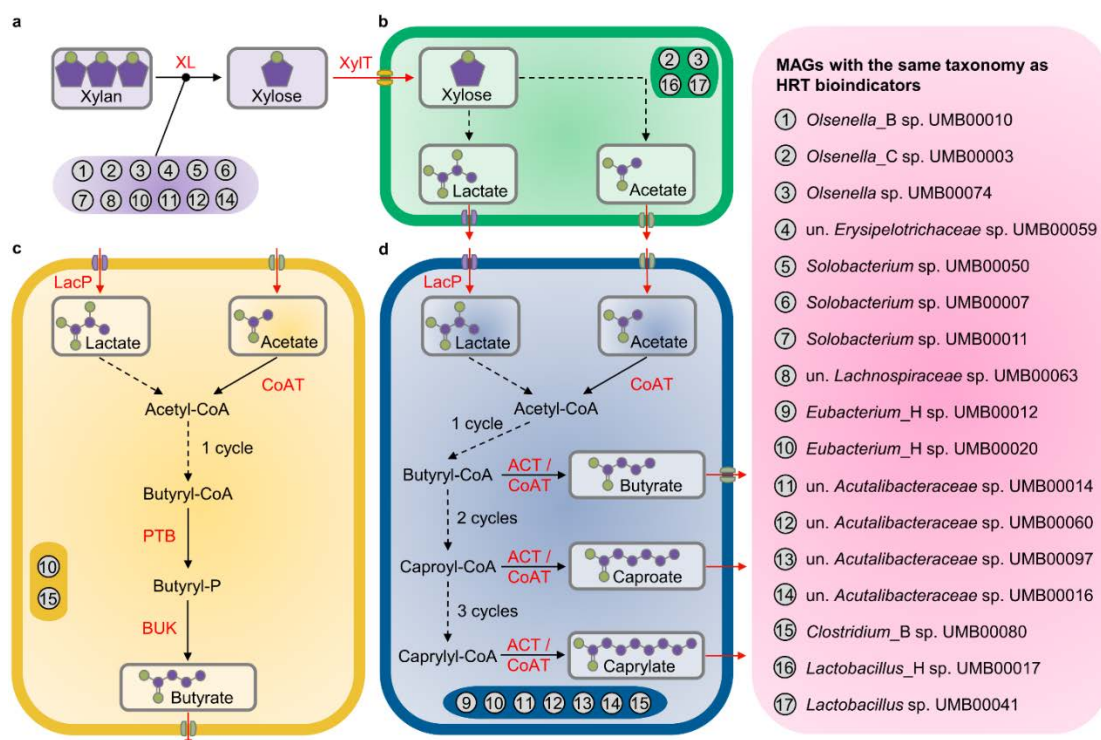
Figure 6. Genetic potential of metagenome-assembled genomes (MAGs) with the same taxonomy as HRT bioindicators driving the catabolism of xylan and lactate to *n*-caproate and *n*-caprylate. These catabolic steps were categorised into four main functions of the anaerobic mixed culture fermentation. **a,** Hydrolysis of xylan. **b,** Xylose fermentation producing acetate and lactate. **c,** Butyrate formation from lactate and acetate. **d,** Chain elongation with lactate as electron donor producing *n*-butyrate, *n*-caproate and *n*-caprylate. Numbers represent the 18 different MAGs with similar phylogenies as the HRT bioindicators at the genus level (details in Table 1). The enzyme abbreviations are provided in red letters next to the pathways (solid lines). Dashed lines represent multi-enzyme reactions between the two indicated molecules. In **(d)**, "cycle" refers to the reverse β-oxidation cycle. The complete metabolic pathways are depicted in Additional file 1: Figure S13. un., unclassified; XL, xylanase (EC 3.2.1.8); XylT, xylose transporter (EC 7.5.2.10, EC 7.5.2.13); LacP, lactate permease (TC 2.A.14); CoAT, butyryl-CoA:acetate CoA-transferase (EC 2.8.3.-); PTB, phosphate butyryltransferase (EC 2.3.1.19); BUK, butyrate kinase (EC 2.7.2.7); ACT, acyl-CoA thioesterase (EC 3.1.2.20).
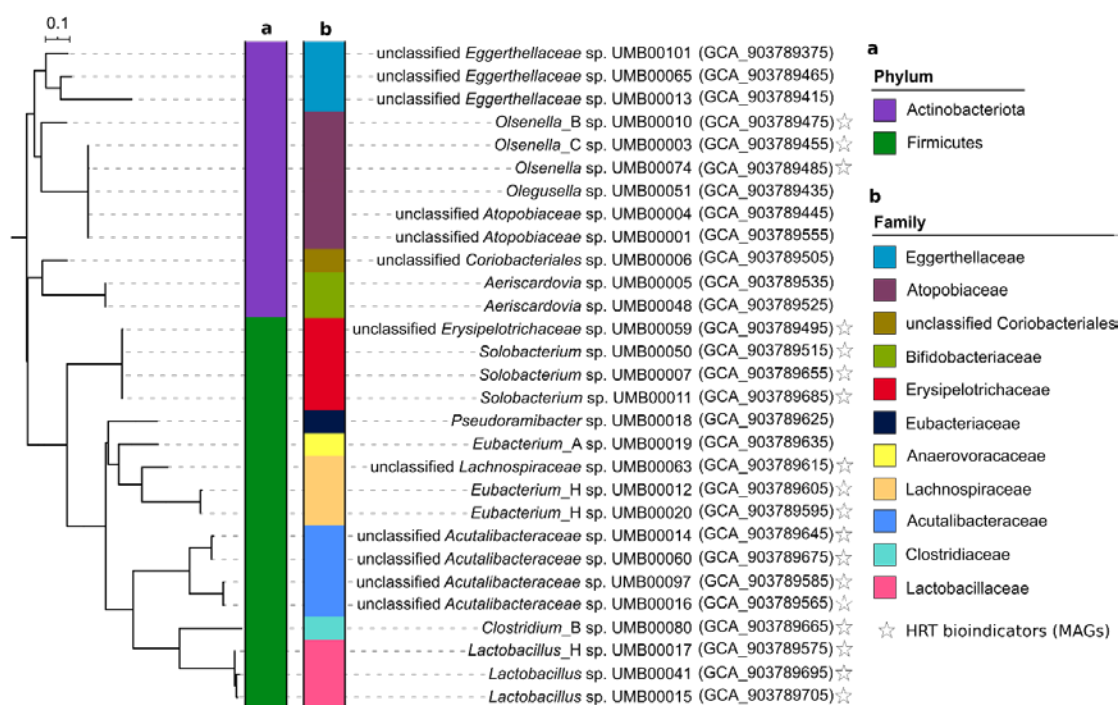
Figure 7. Phylogenetic tree of the recovered metagenome-assembled genomes (MAGs). a,b, A phylogenomic tree based on mash distances showing the MAGs taxonomy determined by GTDB-Tk at phylum (**a**) and family (**b**) levels. A total of 108 MAGs were recovered and differentiated into 29 species based on the ANI values. We defined the representative MAG for each species as that showing high quality. Only the representative MAG for each species is depicted in the tree. The ENA accession numbers of the representative MAGs are shown in parentheses. MAGs with similar phylogenies as HRT bioindicators are indicated by a star.
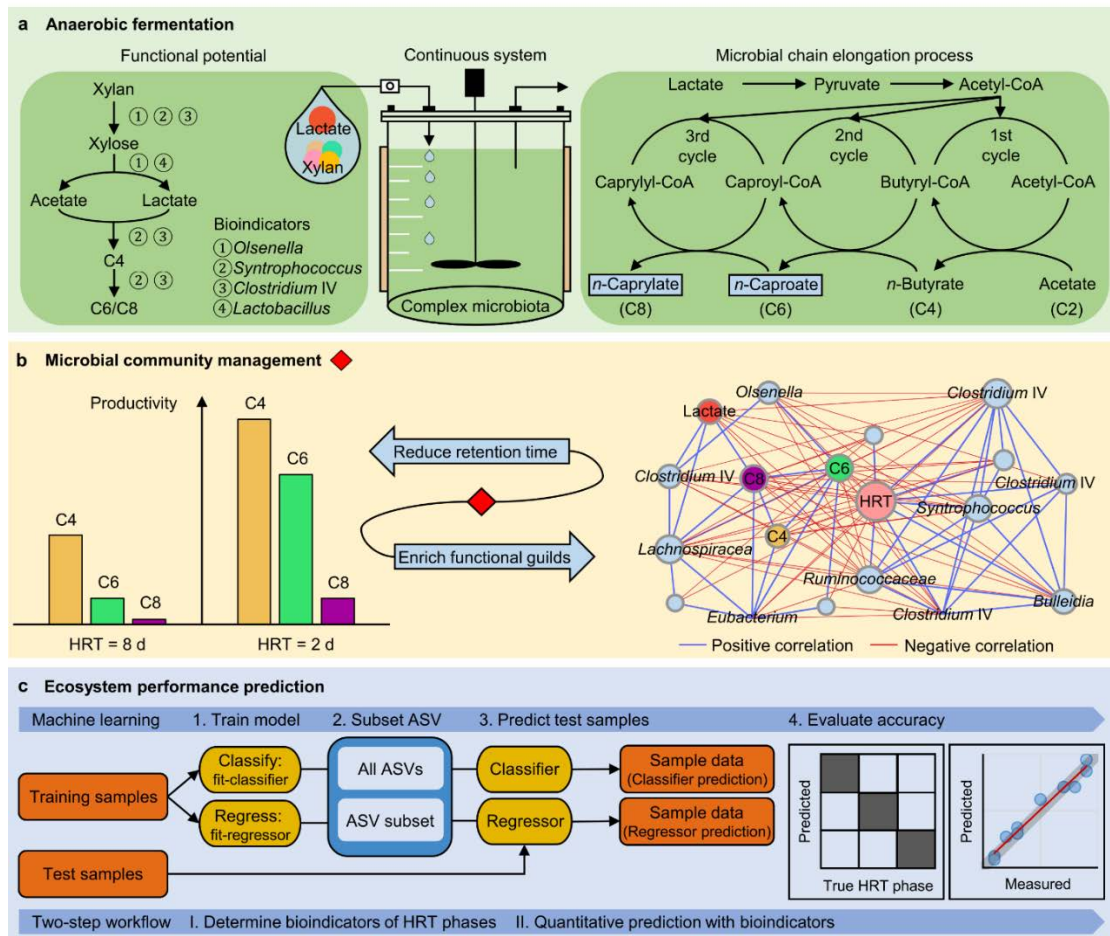
Figure 8. Overview on the quantitative prediction of process performance in the anaerobic bioreactor system. a, Anaerobic mixed culture fermentation of lactate and xylan for the production of *n*-caproate (C6) and *n*-caprylate (C8) by lactate-based chain elongation. Based on the recovery of metagenome-assembled genomes, the left panel shows the bioindicators capable of performing key steps of the fermentation. **b,** Reducing the hydraulic retention time (HRT) as an operation-based strategy to optimise the process performance and to manage the reactor microbiota towards desired functions. Shortening the HRT from 8 days to 2 days enhanced productivities of C4, C6 and C8. The enriched reactor microbiota comprised functional groups involved in xylan hydrolysis, xylose fermentation and chain elongation with lactate, presented by a co-occurrence network of environmental factors (controlled conditions with only reducing the HRT), ecosystem functioning (process performance) and microbial community. The full network is shown in Additional file: Figure S14. **c,** Predicting performance of ecosystem processes with

random forest analysis. We developed a random forest two-step workflow to qualitatively predict the HRT phases and to quantitatively predict carboxylate production by using relative abundance data of the 16S rRNA-derived species (ASVs, Amplicon Sequence Variants).

Table 1. Summary of metagenome-assembled genomes (MAGs) with the same taxonomy as HRT bioindicators.

HRT bioindicators	Number of MAGs		Taxonomic classification						Representative MAG
	High quality	Medium quality	Phylum	Class	Order	Family	Genus	Species	
Olsenella sp. ASV034	2	3	*Actinobacteriota*	*Coriobacteriia*	*Coriobacteriales*	*Atopobiaceae*	*Olsenella_B*	*Olsenella_B* sp000752675	UMB00010
Olsenella sp. ASV057	4	2	*Actinobacteriota*	*Coriobacteriia*	*Coriobacteriales*	*Atopobiaceae*	*Olsenella_C*	unclassified	UMB00003
Olsenella sp. ASV058	1	0	*Actinobacteriota*	*Coriobacteriia*	*Coriobacteriales*	*Atopobiaceae*	*Olsenella*	unclassified	UMB00074
unclassified *Erysipelotrichaceae* sp. ASV002	4	1	*Firmicutes*	*Bacilli*	*Erysipelotrichales*	*Erysipelotrichaceae*	unclassified	unclassified	UMB00059
	0	1	*Firmicutes*	*Bacilli*	*Erysipelotrichales*	*Erysipelotrichaceae*	*Solobacterium*	unclassified	UMB00050
Bulleidia sp. ASV010	6	0	*Firmicutes*	*Bacilli*	*Erysipelotrichales*	*Erysipelotrichaceae*	*Solobacterium*	*Solobacterium* sp900343155	UMB00007
	5	1	*Firmicutes*	*Bacilli*	*Erysipelotrichales*	*Erysipelotrichaceae*	*Solobacterium*	*Solobacterium* sp900290205	UMB00011
Lachnospiracea incertae sedis ASV053	3	0	*Firmicutes_A*	*Clostridia*	*Lachnospirales*	*Lachnospiraceae*	UBA4285	unclassified	UMB00063
Eubacterium cellulosolvens 6			*Firmicutes_A*	*Clostridia*	*Lachnospirales*	*Lachnospiraceae*	*Eubacterium_H*	*Eubacterium_H cellulosolvens*	
Eubacterium cellulosolvens LD2006			*Firmicutes_A*	*Clostridia*	*Lachnospirales*	*Lachnospiraceae*	*Eubacterium_H*	*Eubacterium_H cellulosolvens_A*	
Syntrophococcus sp. ASV060	5	0	*Firmicutes_A*	*Clostridia*	*Lachnospirales*	*Lachnospiraceae*	*Eubacterium_H*	unclassified	UMB00012
	6	0	*Firmicutes_A*	*Clostridia*	*Lachnospirales*	*Lachnospiraceae*	*Eubacterium_H*	unclassified	UMB00020
Caproiciproducens galactitolivorans BS-1			*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	MS4	unclassified	
	5	0	*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA1033	UBA1033 sp002399935	UMB00014
	1	0	*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA1033	UBA1033 sp002407675	UMB00060
Clostridium IV sp. ASV073	3	0	*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA1033	UBA1033 sp002409675	UMB00097
Ruminococcaceae bacterium CPB6			*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA4871	UBA4871 sp002119605	
	6	0	*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA4871	UBA4871 sp002399445	UMB00016
Clostridium luticellarii DSM29923			*Firmicutes_A*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium_B*	*Clostridium_B luticellarii*	
Clostridium sensu stricto sp. ASV008	3	0	*Firmicutes_A*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium_B*	*Clostridium_B* sp003497125	UMB00080
	6	0	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus_H*	*Lactobacillus_H mucosae*	UMB00017
Lactobacillus sp. ASV074	0	1	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	unclassified	UMB00041
	2	2	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	*Lactobacillus amylovorus*	UMB00015
unclassified *Coriobacteriaceae* sp. ASV082	0	0							

Taxonomy refers to the GTDB (Genome Taxonomy Database) phylogenomic classification. ASVs in bold represent the four HRT bioindicators irrespective of time. Sequence datasets of genomes in red letters were taken from the databases of NCBI and EzBioCloud. These genomes (in red) were used to affiliate the MAGs of *Syntrophococcus*, *Clostridium* IV and *Clostridium sensu stricto*, since their genomes are not available in GTDB. See details of MAGs in Additional file 2: Dataset S1. ASV: amplicon sequencing variant.

2.2.2 Supplementary information
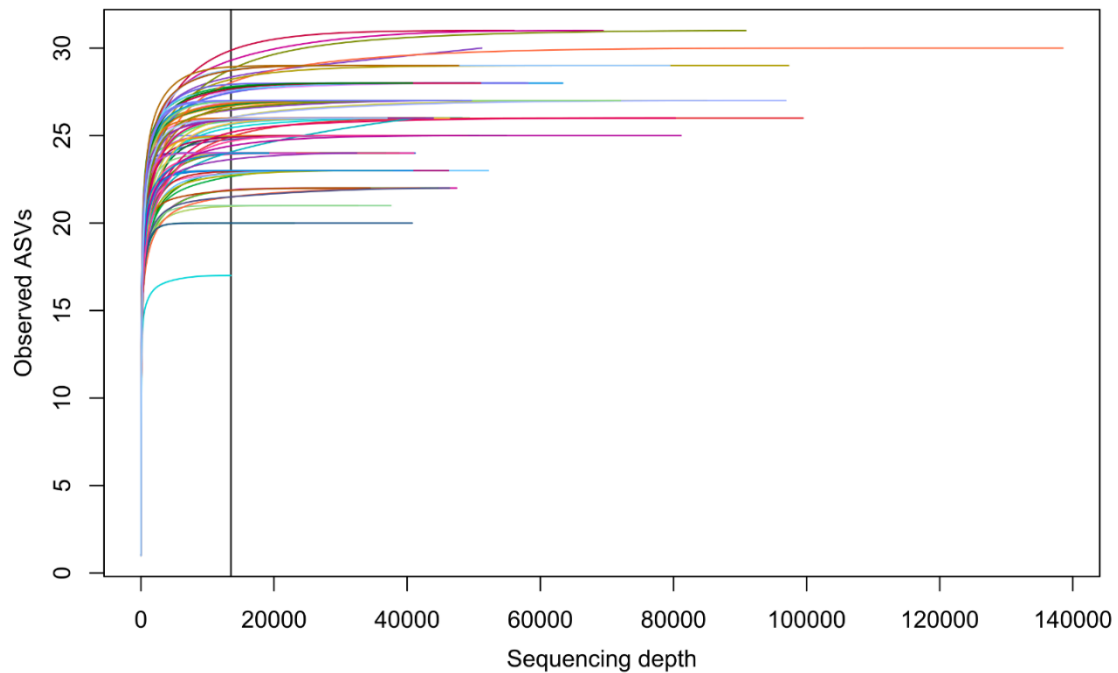
Figure S1. Alpha rarefaction curves. ASVs of all samples were rarefied to an equal sequencing depth of 13,518 reads. Colours represent the different samples.
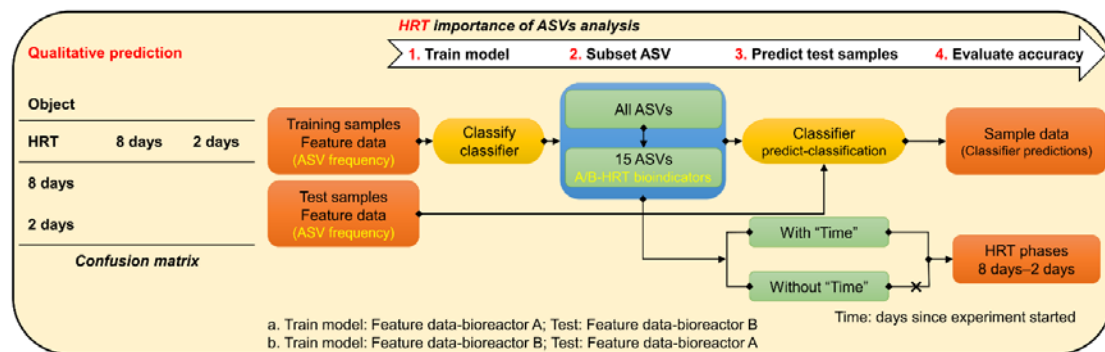
Figure S2. Workflow of the random forest classification analysis. The qualitative prediction of HRT phases of 8 days and 2 days. Letters in yellow show the input (ASV frequency) and output (A/B-HRT bioindicators) of the model.
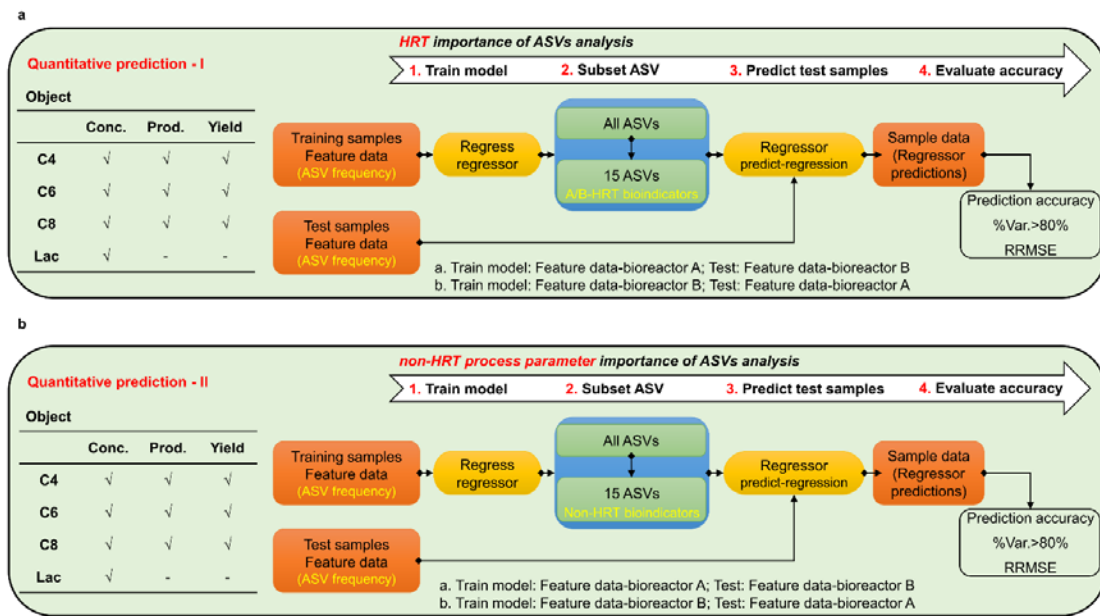
Figure S3. Workflow of a two-step random forest regression analysis. **a,** The first step of quantitative predictions using HRT bioindicators. **b,** The second step of quantitative predictions using non-HRT bioindicators for considering community assembly caused by time. Lac, lactate; Conc., concentration; Prod., productivity; %Var., explains the variance (%) in process parameters of the training set. RRMSE, relative root mean square error.
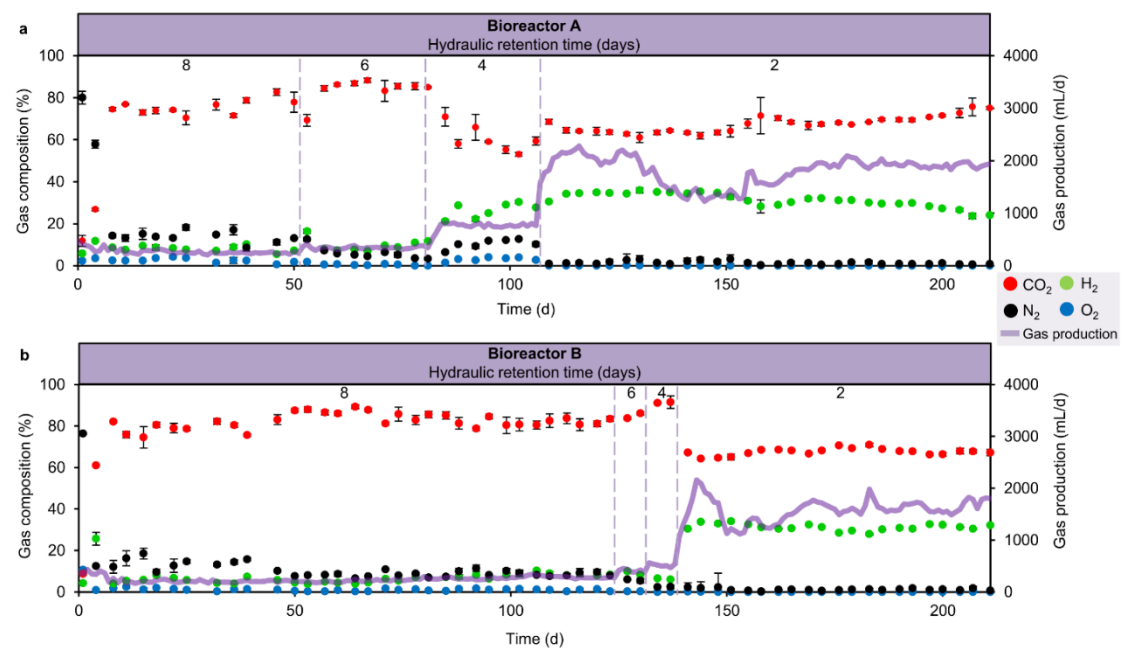
Figure S4. Gas production of bioreactors. Daily gas production and composition in bioreactors A (**a**) and B (**b**), respectively, during the four HRT phases. Error bars indicate the standard deviation.
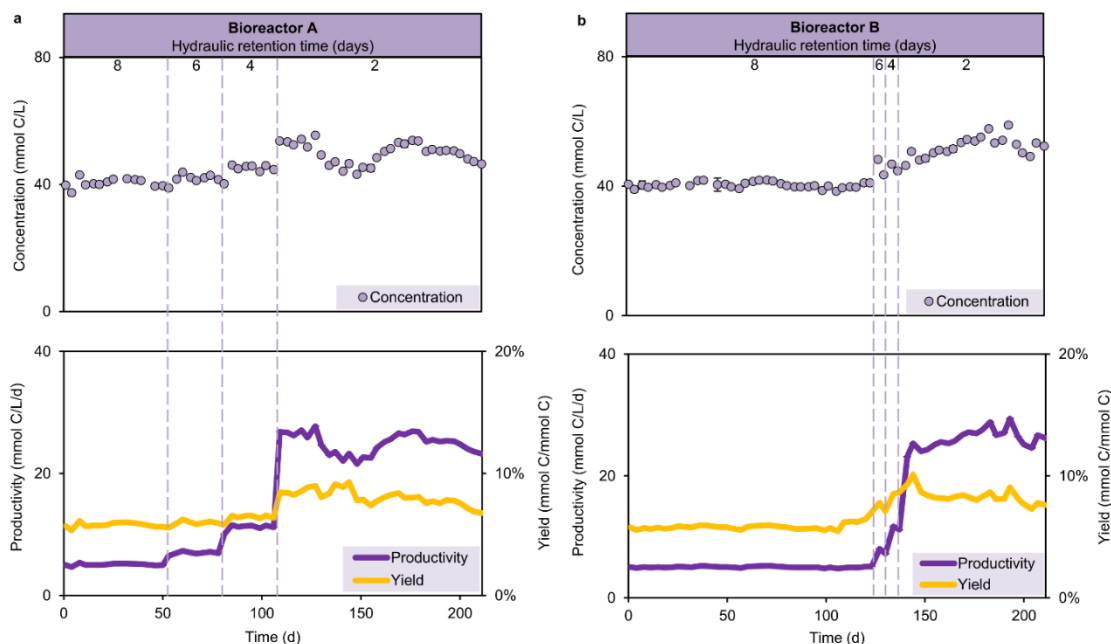
Figure S5. Biomass production of bioreactors. Cell concentration, biomass productivity and biomass yield in bioreactors A (**a**) and B (**b**) during the four HRT phases. The carbon number of cell biomass was calculated by assuming an elemental biomass composition of $CH_{1.8}O_{0.5}N_{0.2}$ (molar mass = 24.6 g/mol). Error bars represent the standard deviation.
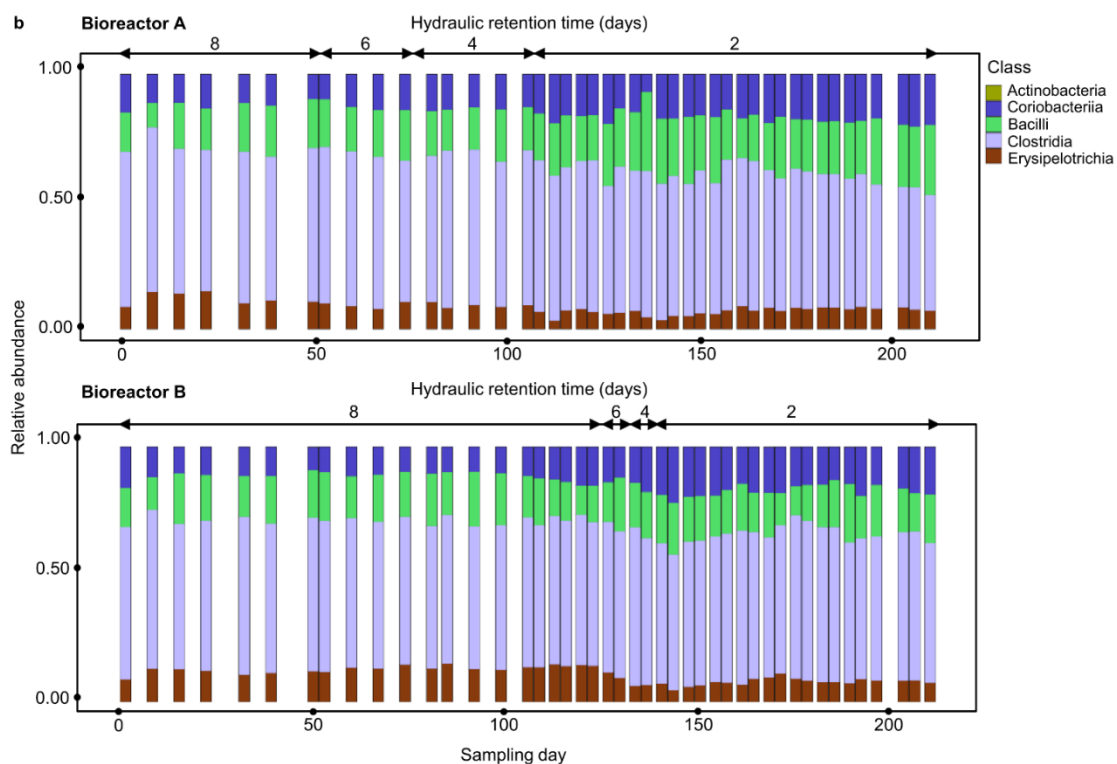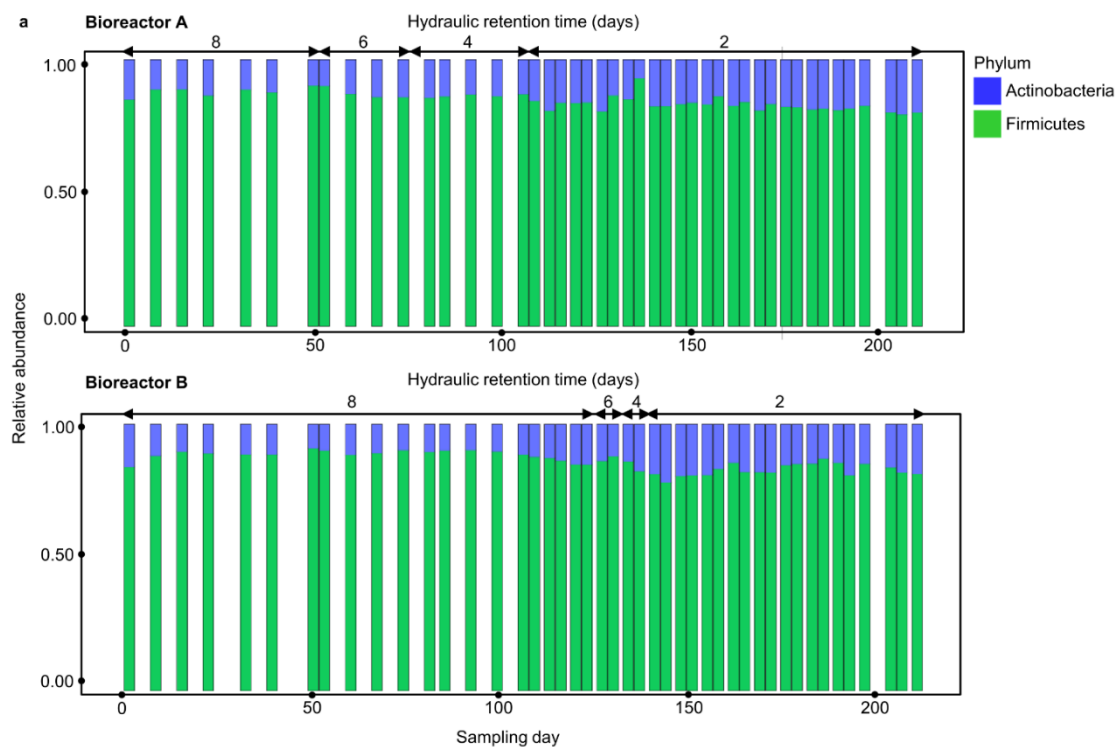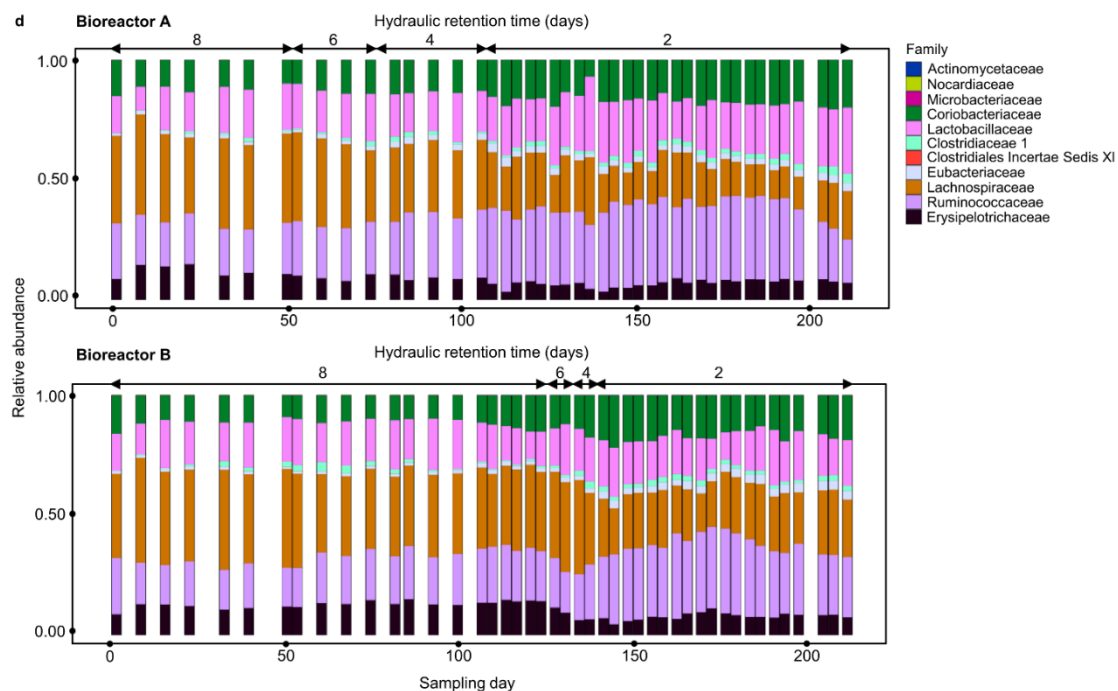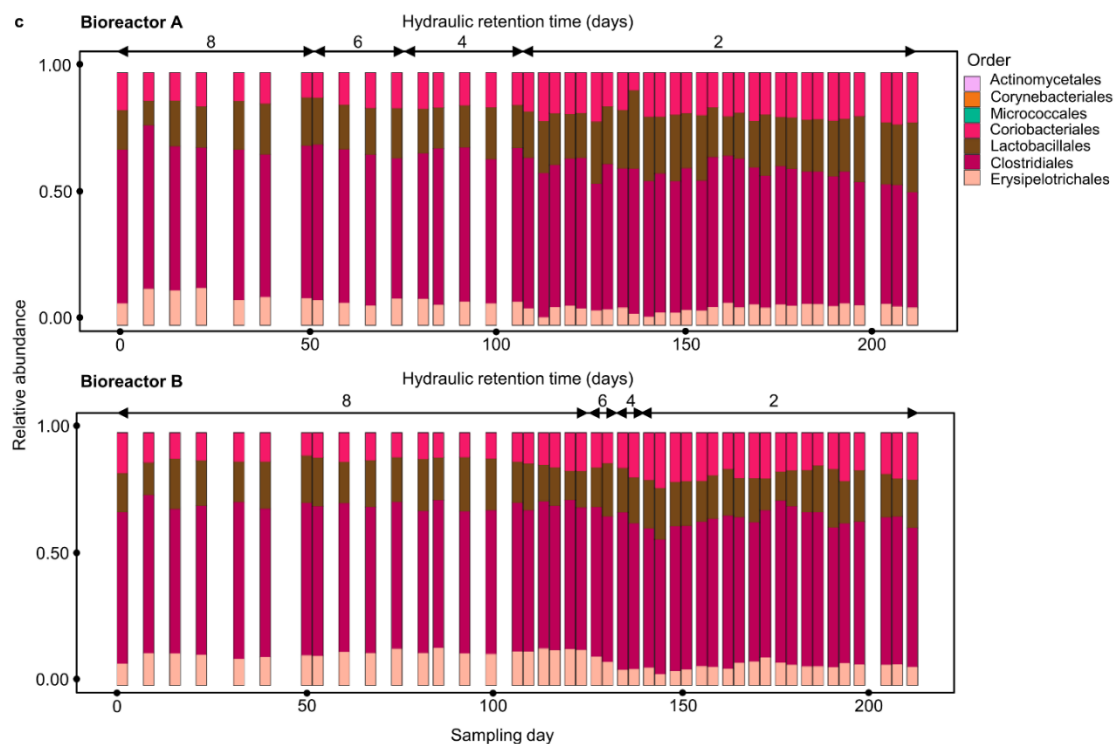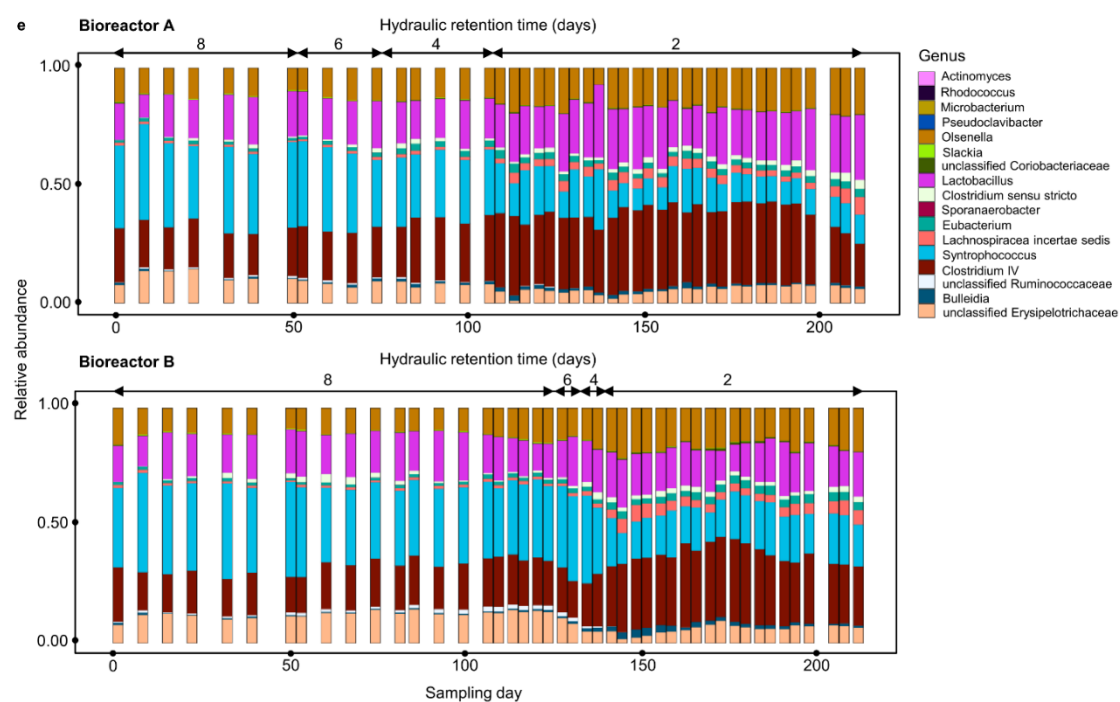
Figure S6. Microbial community composition profiles of bioreactors. Based on amplicon sequencing of 16S rRNA genes, the taxonomic classification of amplicon sequence variants (ASVs) was categorised at the phylum (**a**), class (**b**), order (**c**), family (**d**) and genus (**e**) levels.
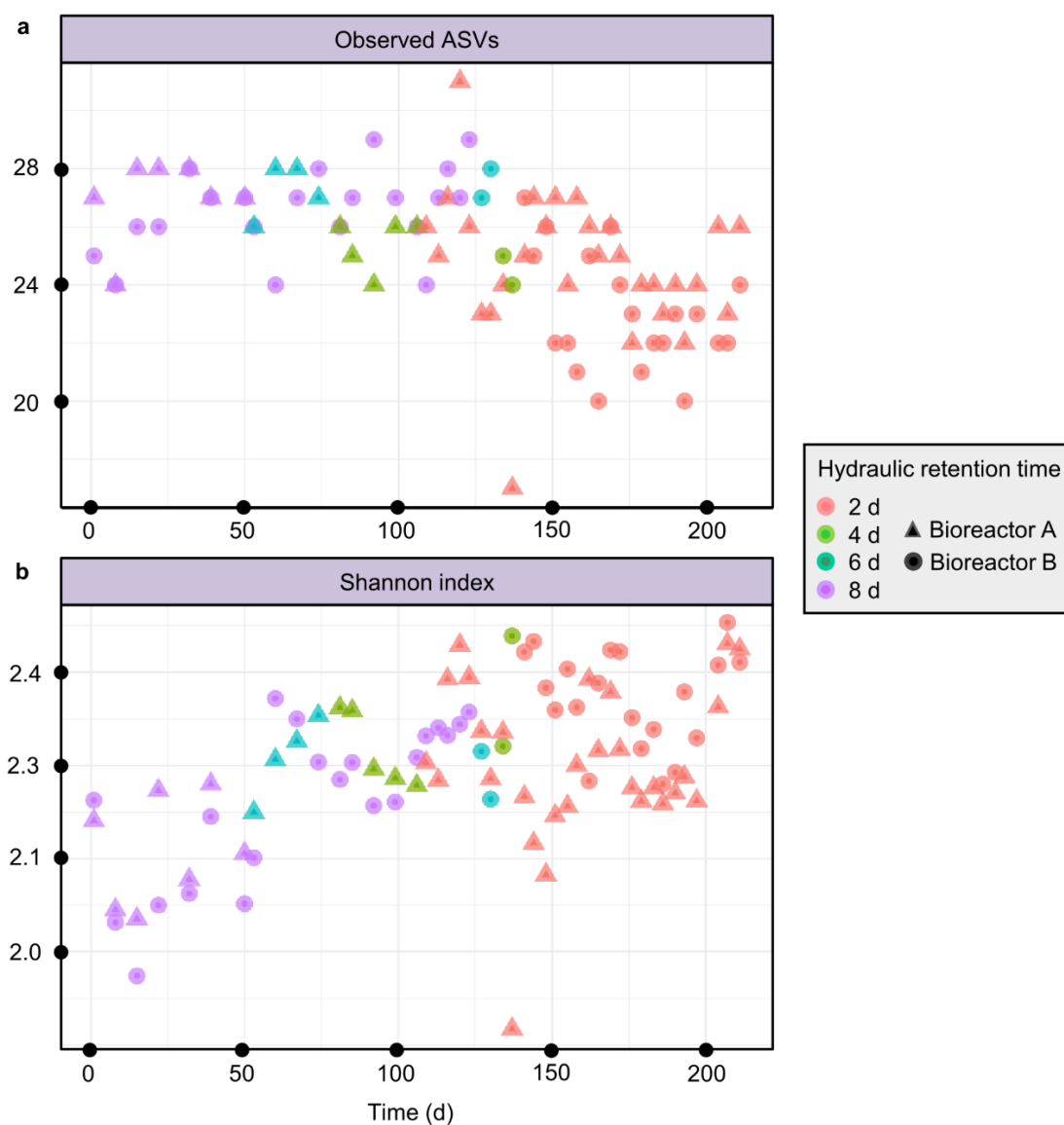
Figure S7. Alpha diversity metrics of bioreactor communities. Based on the relative abundance of ASVs, we calculated the alpha diversity represented by observed ASV counts (**a**) and Shannon index (**b**).

Figure S8. Prediction results of C6 and C8 productivities using non-HRT bioindicators for considering community assembly caused by time. **a,b,** Prediction performance using C6 productivity bioindicators of bioreactors A and B. **c,d,** Prediction performance using C8 productivity bioindicators of bioreactors A and B. Results in a and c were obtained by using the relative abundance data of bioreactor A for training the models and data of bioreactor B for testing. Results using the data of bioreactor B for training and bioreactor A for testing are shown in b and d. %Var., explains the variance (%) in C6/C8 productivity of the training set. RRMSE, relative root mean square error.

Figure S9. Prediction results of C6 and C8 productivities for all samples in the four HRT phases using HRT bioindicators. **a,b,** Prediction performance of C6 productivity. **c,d,** Prediction performance of C8 productivity. Results in **a** and **c** were obtained by using the relative abundance data of bioreactor A for training the models and data of bioreactor B for testing. Results using data of bioreactor B for training and bioreactor A for testing are shown in **b** and **d**. %Var., explains the variance (%) in C6/C8 productivity of the training set. RRMSE, relative root mean square error.

Figure S10. Prediction results of C6 and C8 productivities for all samples in the four HRT phases using non-HRT bioindicators for considering community assembly caused by time. **a,b,** Prediction performance using C6 productivity bioindicators of bioreactors A and B. **c,d,** Prediction performance using C8 productivity bioindicators of bioreactors A and B. Results in **a** and **c** were obtained by using the relative abundance data of bioreactor A for training the models and data of bioreactor B for testing. Results using the data of bioreactor B for training and bioreactor A for testing are shown in **b** and **d**. %Var., explains the variance (%) in C6/C8 productivity of the training set. RRMSE, relative root mean square error.

Figure S11. Random forest feature importance of A-HRT bioindicators and B-HRT bioindicators used to predict C6 and C8 productivities. **a,** Feature importance of A-HRT bioindicators in the prediction of C6 productivity. **b,** Feature importance of B-HRT bioindicators in the prediction of C6 productivity. **c,** Feature importance of A-HRT bioindicators in the prediction of C8 productivity. **d,** Feature importance of B-HRT bioindicators in the prediction of C8 productivity. IncNodePurity, residual sum of squares.
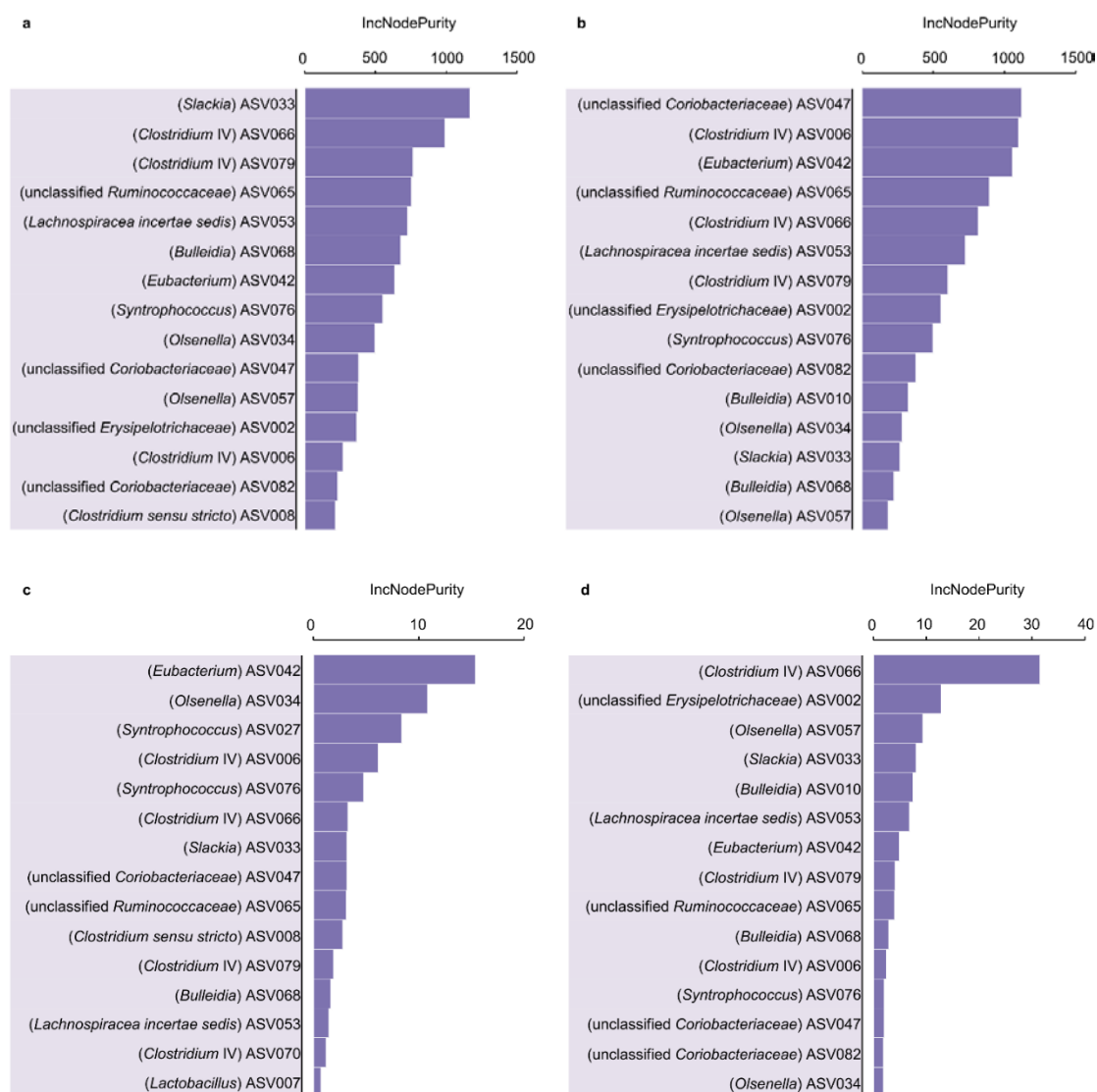
Figure S12. Random forest feature importance of the non-HRT bioindicators used to predict C6 and C8 productivities. **a,b,** The feature importance of C6 productivity bioindicators of bioreactors A and B. **c,d,** The feature importance C8 productivity bioindicators of bioreactors A and B. Relative abundance data of bioreactor A were used as training set and that of bioreactor B as test set (**a,c**); while data of bioreactor B for training and that of bioreactor A for testing (**b,d**). IncNodePurity, residual sum of squares.

a

b

Predicted function	No.	Enzyme abbreviation	EC number	Enzyme
Xylan degradation	1	XL	3.2.1.8	xylanase
Xylose utilization	2	XylT	7.5.2.10;7.5.2.13	xylose transporter
	3	XI	5.3.1.5	xylose isomerase
	4	XK	2.7.1.17	xylulose kinase
	5	R5PE	5.1.3.1	ribulose-5-phosphoate epimerase
	6	R5PI	5.3.1.6	ribose-5-phosphate isomerase
	7	TA	2.2.1.2	transaldolase
	8	TK	2.2.1.1	transketolase
	9	PK	4.1.2.9	D-xylulose 5-phosphate phosphoketolase
Pyruvate transformation	10	PDC	1.2.4.1;2.3.1.2;1.8.1.4	pyruvate dehydrogenase complex
	11	PFOR	1.2.7.-	pyruvate flavodoxin oxidoreductase
	12	PFL	2.3.1.54	formate C-acetyltransferase (pyruvate formate lyase)
	13	ALDH	1.2.1.10	acetaldehyde dehydrogenase
	14	ADH	1.1.1.1	alcohol dehydrogenase
	15	PTA	2.3.1.8	phosphate acetyltransferase
	16	ACK	2.7.2.1	acetate kinase
	17	LDH	1.1.1.27;1.1.1.28	lactate dehydrogenase
	18	PTB	2.3.1.19	phosphate butyryltransferase
	19	BUK	2.7.2.7	butyrate kinase

c

Predicted function	No.	Enzyme abbreviation	EC number	Enzyme
Lactate utilization	1	LacR	5.1.2.1	lactate racemase
	2	LacP	2.A.14	lactate permease
	3	LutABC		lactate utilization protein ABC
Acetyl-CoA formation	4	LDH	1.1.1.27	lactate dehydrogenase
	5	PFOR	1.2.7.1	pyruvate ferredoxin oxidoreductase
Reverse β-oxidation	6	ACAT	2.3.1.9, 2.3.1.16	acetyl-CoA acetyltransferase
	7	HAD	1.1.1.157, 1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
	8	ECH	4.2.1.150, 4.2.1.55	enoyl-CoA hydratase
	9	BCD	1.3.8.1	butyryl-CoA dehydrogenase
	10	EtfAB		electron transfer flavoprotein A,B
	11	CoAT	2.8.3.-	butyryl-CoA:acetate CoA-transferase
	12	ACT	3.1.2.20	acyl-CoA thioesterase
Energy conservation	13	RnfABCDEG	7.1.1.1	energy-converting NADH: ferredoxin oxidoreductase

Figure S13. Metabolic pathways involved in converting lactate and xylan to n-caproate and *n*-caprylate. **a,** The pathways involved in the processes of anaerobic mixed culture fermentation include hydrolysis of xylan (more than 95% xylooligosaccharides), fermentation of xylose and chain elongation with lactate as electron donor. The enzyme abbreviations are provided in red letters next to the pathways (solid lines). **b,** Enzymes of the predicted functions related to

xylan hydrolysis, xylose fermentation and pyruvate transformation. **c,** Enzymes of the predicted functions related to chain elongation with lactate as electron donor. Dashed lines represent multi-enzyme reactions between the two indicated molecules. In (**a**), "cycle" refers to the reverse β-oxidation cycle. The functional annotation of metagenome-assembled genomes (MAGs) with the same taxonomy as HRT bioindicators can be found in Additional file 3: Dataset S2 (for xylan hydrolysis and xylose fermentation) and Additional file 4: Dataset S3 (for chain elongation). The functional annotation of all MAGs can be found in Additional file 5: Dataset S4 (for xylan hydrolysis and xylose fermentation) and Additional file 6: Dataset S5 (for chain elongation).
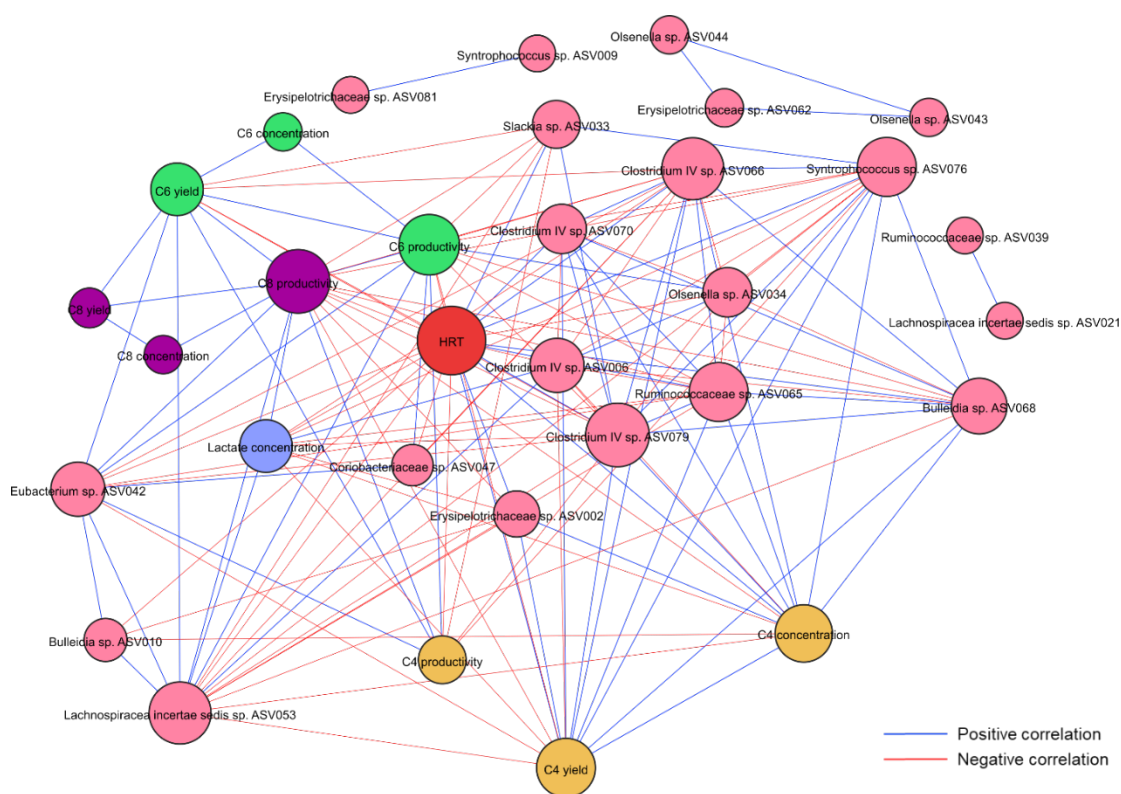
Figure S14. Correlation network of environmental factors, process performance and microbial community. Edges indicate the Spearman coefficient > 0.7 for positive correlations (blue edges) and < 0.7 for negative correlations (red edges). Node size was scaled to represent its degree of connectedness. Here, the environmental factors represent controlled operational parameters with only reducing the HRT, and the process

performance refers to the concentration, productivity and yield of the target products. C4, *n*-butyrate; C6, *n*-caproate; C8, *n*-caprylate.
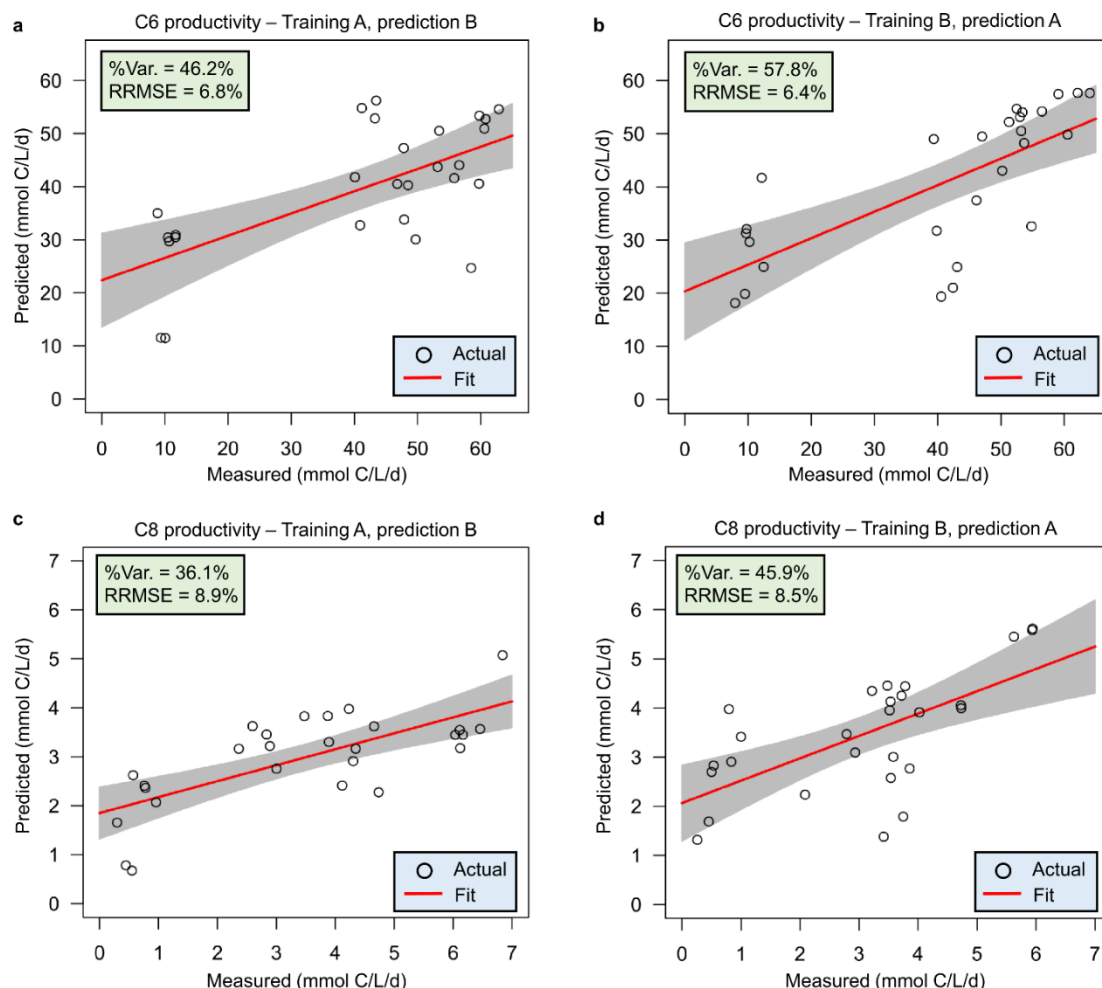
Figure S15. Prediction results of C6 and C8 productivities for all samples in the four HRT phases using the four ASVs of HRT bioindicators irrespective of time. **a,b,** Prediction performance of C6 productivity. **c,d,** Prediction performance of C8 productivity. Results in **a** and **c** were obtained by using the relative abundance data of bioreactor A for training the models and data of bioreactor B for testing. Results using the data of bioreactor B for training and bioreactor A for testing are shown in **b** and **d**. %Var., explains the variance (%) in C6/C8 productivity of the training set. RRMSE, relative root mean square error.

Figure S16. Reducing HRT increases abundances of HRT bioindicators driving the catabolism of xylan and lactate to *n*-caproate and *n*-caprylate. These catabolic steps were categorized into four main functions of the anaerobic mixed culture fermentation. **a,** Hydrolysis of xylan. Relevant HRT bioindicators are *Olsenella* sp. ASV034, *Olsenella* sp. ASV057, *Olsenella* sp. ASV058, unclassified *Erysipelotrichaceae* sp. ASV002, *Bulleidia* sp. ASV010, *Lachnospiracea incertae sedis* ASV053, *Syntrophococcus* sp. ASV060 and *Clostridium* IV sp. ASV073. **b,** Xylose fermentation producing acetate and lactate. Relevant HRT bioindicators are *Olsenella* sp. ASV034, *Olsenella* sp. ASV057, *Olsenella* sp. ASV058 and *Lactobacillus* sp. ASV074. **c,** Butyrate formation from lactate and acetate. Relevant HRT bioindicators are *Syntrophococcus* sp. ASV060 and *Clostridium sensu stricto* sp. ASV008. **d,**

Chain elongation with lactate as electron donor producing *n*-butyrate, *n*-caproate and *n*-caprylate. Relevant HRT bioindicators are *Syntrophococcus* sp. ASV060, *Clostridium* IV sp. ASV073 and *Clostridium sensu stricto* sp. ASV008. The rarefied ASV abundances were calculated using 16S rRNA amplicon sequencing data. Boxes represent the interquartile range between the 25th and 75th percentiles, respectively, the line inside denote the median value, and asterisks indicate significant different mean values (adjusted $P^{***} < 0.001 < ^{**} < 0.01 < ^{*} < 0.05$), which is tested by permutational Student's *t*-test with 9,999 simulations. On the horizontal axis, A and B stand for bioreactors A and B, respectively.

Table S1. Growth medium used for the reactor operation. The medium was anoxic by flushing with nitrogen and adjusted to pH 5.5 with 1 M sodium hydroxide solution.

Concentrations (all components were prepared completely sterile):			
0.054 g/L	$MgCl_2 \times 6\ H_2O$		
0.065 g/L	$CaCl_2 \times 2\ H_2O$		
0.474 g/L	NH_4Cl		
0.5 g/L	Yeast extract		
10.94 g/L	KH_2PO_4		
20.83 g/L	K_2HPO_4		
0.032 g/L	Na_2CO_3		
0.03 g/L	Cysteine-HCl		
1 mL/L Trace element solution		1 mL/L Vitamin solution	
$FeCl_2 \times 4\ H_2O$	1.5 g/L	Biotin	20 mg/L
$CuCl_2 \times 2\ H_2O$	2 mg/L	Folic acid	20 mg/L
$CoCl_2 \times 6\ H_2O$	190 mg/L	Pyridoxine	100 mg/L
$MnCl_2$	100 mg/L	Thiamine	50 mg/L
$Na_2MoO_4 \times 2\ H_2O$	36 mg/L	Riboflavin	50 mg/L
$NiCl_2 \times 6\ H_2O$	24 mg/L	Nicotinic acid	50 mg/L
$Na_2WO_4 \times 2\ H_2O$	20 mg/L	Calcium pantothenate	50 mg/L
$Na_2SeO_3 \times 5\ H_2O$	3 mg/L	Vitamin B_{12}	20 mg/L
$ZnCl_2$	70 mg/L	p-Amino benzoic acid	80 mg/L
H_3BO_3	6 mg/L	Lipoic acid	50 mg/L

Table S2. Daily feeding of bioreactors A and B during the four HRT phases.

Bioreactor			Daily medium feeding				Daily effluent withdrawing (mL)
	HRT (d)	Duration (d)	Lactate (g)	Xylan (g)	Mineral medium (mL)	Deionised anoxic water (mL)	
A	8	0-51	1.47	1.25	11	114	125
	6	52-80	1.96	1.67	15	152	167
	4	81-106	2.94	2.50	23	227	250
	2	107-211	5.88	5.00	45	455	500
B	8	0-123	1.47	1.25	11	114	125
	6	124-130	1.96	1.67	15	152	167
	4	131-137	2.94	2.50	23	227	250
	2	138-211	5.88	5.00	45	455	500

Table S3. Gini scores of all ASVs in the classification-based prediction of HRT phases. The percentage was calculated by dividing the Gini score of certain ASV by the sum of the Gini scores of all ASVs.

No.	ASV name	Bioreactor A Mean decrease in Gini	Percentage	ASV name	Bioreactor B Mean decrease in Gini	Percentage
1	ASV073	1.59	16.0%	ASV057	2.77	27.8%
2	ASV057	1.51	15.1%	ASV073	1.26	12.6%
3	ASV002	1.36	13.7%	ASV034	1.18	11.8%
4	ASV082	1.15	11.5%	ASV040	0.63	6.3%
5	ASV008	0.68	6.9%	ASV002	0.59	5.9%
6	ASV006	0.43	4.3%	ASV008	0.59	5.9%
7	ASV015	0.37	3.8%	ASV060	0.58	5.8%
8	ASV053	0.31	3.2%	ASV074	0.34	3.4%
9	ASV058	0.29	3.0%	ASV082	0.26	2.6%
10	ASV034	0.26	2.6%	ASV058	0.24	2.4%
11	ASV060	0.24	2.5%	ASV010	0.19	1.9%
12	ASV027	0.20	2.0%	ASV053	0.17	1.8%
13	ASV074	0.17	1.7%	ASV042	0.17	1.7%
14	ASV054	0.15	1.5%	ASV066	0.16	1.6%
15	ASV010	0.14	1.4%	ASV068	0.10	1.0%
16	ASV007	0.11	1.1%	ASV070	0.09	0.9%
17	ASV047	0.10	1.0%	ASV076	0.08	0.8%
18	ASV065	0.09	0.9%	ASV065	0.08	0.8%
19	ASV076	0.08	0.8%	ASV079	0.06	0.6%
20	ASV046	0.07	0.7%	ASV038	0.05	0.5%
21	ASV042	0.07	0.7%	ASV047	0.05	0.5%
22	ASV066	0.07	0.7%	ASV025	0.05	0.5%
23	ASV038	0.07	0.7%	ASV007	0.05	0.5%
24	ASV033	0.06	0.6%	ASV084	0.04	0.4%
25	ASV070	0.06	0.6%	ASV006	0.03	0.3%
26	ASV030	0.05	0.5%	ASV078	0.03	0.3%
27	ASV079	0.04	0.4%	ASV033	0.03	0.3%
28	ASV040	0.04	0.4%	ASV027	0.03	0.3%
29	ASV049	0.03	0.3%	ASV031	0.02	0.2%
30	ASV026	0.03	0.3%	ASV015	0.02	0.2%
31	ASV051	0.02	0.2%	ASV041	0.01	0.1%
32	ASV068	0.02	0.2%	ASV059	0.01	0.1%
33	ASV025	0.02	0.2%	ASV030	0.01	0.1%
34	ASV031	0.02	0.2%	ASV046	0.00	0.0%
35	ASV069	0.01	0.1%	ASV049	0.00	0.0%
36	ASV084	0.01	0.1%	ASV016	0.00	0.0%
37	ASV085	0.01	0.1%	ASV051	0.00	0.0%
38	ASV014	0.01	0.1%	ASV012	0.00	0.0%
39	ASV019	0.01	0.1%	ASV001	0.00	0.0%
40	ASV003	0.00	0.0%	ASV003	0.00	0.0%
41	ASV018	0.00	0.0%	ASV005	0.00	0.0%
42	ASV011	0.00	0.0%	ASV009	0.00	0.0%
43	ASV081	0.00	0.0%	ASV011	0.00	0.0%
44	ASV001	0.00	0.0%	ASV014	0.00	0.0%
45	ASV005	0.00	0.0%	ASV017	0.00	0.0%
46	ASV009	0.00	0.0%	ASV018	0.00	0.0%
47	ASV012	0.00	0.0%	ASV019	0.00	0.0%
48	ASV016	0.00	0.0%	ASV021	0.00	0.0%
49	ASV017	0.00	0.0%	ASV023	0.00	0.0%
50	ASV021	0.00	0.0%	ASV024	0.00	0.0%
51	ASV023	0.00	0.0%	ASV026	0.00	0.0%
52	ASV024	0.00	0.0%	ASV028	0.00	0.0%
53	ASV028	0.00	0.0%	ASV029	0.00	0.0%
54	ASV029	0.00	0.0%	ASV039	0.00	0.0%
55	ASV039	0.00	0.0%	ASV043	0.00	0.0%
56	ASV041	0.00	0.0%	ASV044	0.00	0.0%
57	ASV043	0.00	0.0%	ASV045	0.00	0.0%
58	ASV044	0.00	0.0%	ASV048	0.00	0.0%
59	ASV045	0.00	0.0%	ASV054	0.00	0.0%
60	ASV048	0.00	0.0%	ASV056	0.00	0.0%
61	ASV056	0.00	0.0%	ASV062	0.00	0.0%
62	ASV059	0.00	0.0%	ASV063	0.00	0.0%
63	ASV062	0.00	0.0%	ASV064	0.00	0.0%
64	ASV063	0.00	0.0%	ASV067	0.00	0.0%
65	ASV064	0.00	0.0%	ASV069	0.00	0.0%
66	ASV067	0.00	0.0%	ASV071	0.00	0.0%
67	ASV071	0.00	0.0%	ASV077	0.00	0.0%
68	ASV077	0.00	0.0%	ASV080	0.00	0.0%
69	ASV078	0.00	0.0%	ASV081	0.00	0.0%
70	ASV080	0.00	0.0%	ASV085	0.00	0.0%
71	ASV086	0.00	0.0%	ASV086	0.00	0.0%

Table S4. Mean carboxylate yields (i.e. C mole product to substrate ratios) at HRTs of 8 days and 2 days (stable production period).

Bioreactor	HRT (d)	Duration (d)	C4 (mmol C/mmol C)	C6 (mmol C/mmol C)	C8 (mmol C/mmol C)
A	8	0-51	39.1	11.6	0.7
	2	162-211	26.8	16.9	1.3
B	8	0-123	39.6	12.3	0.9
	2	193-211	22.3	18.7	2.0

Table S5. Explained variances of the training set in the regression-based prediction of process parameters using A-HRT bioindicators and B-HRT bioindicators. Features (ASVs) explaining more than 80% of the variance in a process parameter are indicated in green. A and B stand for bioreactors A and B, respectively.

Training set	Predicted variable		Set of ASVs	Number of ASVs	Explained variance (%)
Bioreactor A	concentration	C4	ASV_all	71	89.37 ± 0.48
			A-HRT bioindicators	15	87.81 ± 0.54
		C6	ASV_all	71	67.66 ± 0.84
			A-HRT bioindicators	15	68.06 ± 0.84
		C8	ASV_all	71	58.43 ± 0.91
			A-HRT bioindicators	15	59.30 ± 1.00
		Lactate	ASV_all	71	77.30 ± 0.79
			A-HRT bioindicators	15	78.20 ± 0.77
	productivity	C4	ASV_all	71	88.61 ± 0.41
			A-HRT bioindicators	15	88.94 ± 0.35
		C6	ASV_all	71	93.80 ± 0.45
			A-HRT bioindicators	15	95.87 ± 0.23
		C8	ASV_all	71	80.91 ± 0.78
			A-HRT bioindicators	15	86.62 ± 0.51
	yield	C4	ASV_all	71	85.39 ± 0.58
			A-HRT bioindicators	15	80.81 ± 0.73
		C6	ASV_all	71	73.81 ± 0.64
			A-HRT bioindicators	15	54.62 ± 0.98
		C8	ASV_all	71	60.50 ± 0.81
			A-HRT bioindicators	15	56.84 ± 0.92
Bioreactor B	concentration	C4	ASV_all	71	79.49 ± 0.49
			B-HRT bioindicators	15	78.61 ± 0.63
		C6	ASV_all	71	73.94 ± 0.54
			B-HRT bioindicators	15	77.11 ± 0.50
		C8	ASV_all	71	62.37 ± 0.75
			B-HRT bioindicators	15	63.64 ± 0.83
		Lactate	ASV_all	71	68.64 ± 0.85
			B-HRT bioindicators	15	72.02 ± 0.66
	productivity	C4	ASV_all	71	65.90 ± 0.63
			B-HRT bioindicators	15	67.91 ± 0.58
		C6	ASV_all	71	93.92 ± 0.36
			B-HRT bioindicators	15	93.44 ± 0.37
		C8	ASV_all	71	84.37 ± 0.68
			B-HRT bioindicators	15	85.65 ± 0.58
	yield	C4	ASV_all	71	68.67 ± 0.82
			B-HRT bioindicators	15	70.99 ± 0.49
		C6	ASV_all	71	73.57 ± 0.54
			B-HRT bioindicators	15	78.17 ± 0.41
		C8	ASV_all	71	65.75 ± 0.83
			B-HRT bioindicators	15	67.19 ± 0.93

Table S6. Explained variances of the training set in the regression-based prediction of process parameters using non-HRT bioindicators for considering community assembly caused by time. Features (ASVs) explaining more than 80% of the variance in a process parameter are indicated in green. A and B stand for bioreactors A and B, respectively.

Training set	Predicted variable		Set of ASVs	Number of ASVs	Explained variance (%)
Bioreactor A	concentration	C4	ASV_all	71	89.37 ± 0.48
			A-C4c_bioindicators	15	91.69 ± 0.25
		C6	ASV_all	71	67.66 ± 0.84
			A-C6c_bioindicators	15	70.72 ± 0.57
		C8	ASV_all	71	58.43 ± 0.91
			A-C8c_bioindicators	15	63.95 ± 0.54
		Lactate	ASV_all	71	77.30 ± 0.79
			A-LACc_bioindicators	15	79.38 ± 0.53
	productivity	C4	ASV_all	71	88.61 ± 0.41
			A-C4p_bioindicators	15	88.83 ± 0.26
		C6	ASV_all	71	93.80 ± 0.45
			A-C6p_bioindicators	15	95.64 ± 0.20
		C8	ASV_all	71	80.91 ± 0.78
			A-C8p_bioindicators	15	83.20 ± 0.48
	yield	C4	ASV_all	71	85.39 ± 0.58
			A-C4y_bioindicators	15	87.07 ± 0.36
		C6	ASV_all	71	73.81 ± 0.64
			A-C6y_bioindicators	15	74.73 ± 0.40
		C8	ASV_all	71	60.50 ± 0.81
			A-C8y_bioindicators	15	61.85 ± 0.66
Bioreactor B	concentration	C4	ASV_all	71	79.49 ± 0.49
			B-C4c_bioindicators	15	81.50 ± 0.33
		C6	ASV_all	71	73.94 ± 0.54
			B-C6c_bioindicators	15	75.15 ± 0.36
		C8	ASV_all	71	62.37 ± 0.75
			B-C8c_bioindicators	15	64.35 ± 0.57
		Lactate	ASV_all	71	68.64 ± 0.85
			B-LACc_bioindicators	15	71.16 ± 0.51
	productivity	C4	ASV_all	71	65.90 ± 0.63
			B-C4p_bioindicators	15	67.90 ± 0.41
		C6	ASV_all	71	93.92 ± 0.36
			B-C6p_bioindicators	15	95.25 ± 0.24
		C8	ASV_all	71	84.37 ± 0.68
			B-C8p_bioindicators	15	85.90 ± 0.53
	yield	C4	ASV_all	71	68.67 ± 0.82
			B-C4y_bioindicators	15	71.17 ± 0.46
		C6	ASV_all	71	73.57 ± 0.54
			B-C6y_bioindicators	15	74.70 ± 0.35
		C8	ASV_all	71	65.75 ± 0.83
			B-C8y_bioindicators	15	67.84 ± 0.52

2.3 Effects of pH increase on microbial chain elongation and community dynamics in closed bioreactor ecosystems

In preparation for submission

2.3.1 Main text

Abstract

As for microbial chain elongation in engineered systems, it is unclear how alterations of pH can affect the abundance of key players, the responses of microbial interactions and the community functioning in terms of medium-chain carboxylate yields. Here, we explored its effect on the community dynamics measured by sequencing 16S rRNA genes in continuous anaerobic bioreactors. Increasing pH from 5.5 to 6.0 caused fluctuations in the yields of *n*-caproate and *n*-caprylate. After the pH disturbance, the yields returned to the previous values while the communities developed to a different state, observed as decrease in diversity and evenness and increase in richness. By applying Aitchison PCA clustering, linear mixed effect models and random forest classification, the different pH preferences of potential chain elongators *Clostridium* IV and *Clostridium sensu stricto* were identified. By constructing networks for different pH levels, the cooperation of the chain elongator *Clostridium* IV with lactic acid bacteria switches from *Olsenella* to *Lactobacillus* along the pH increase, revealing the plasticity of the food web of chain elongation communities. pH increase induced dramatic shifts in the community composition but exhibited no strong effects on medium-chain carboxylate yields. High functional redundancy was indicated despite the reactors being long-term closed systems.

Introduction

In microbial ecology, it is important to understand the main environmental factors driving the deterministic processes of microbial community assembly and functioning [1–3]. Ecological selection exerted by abiotic and biotic factors in deterministic processes influences the growth rates and interactions between community members, and thereby determines the composition and functioning of microbial communities [4–7]. For engineered systems, pH is frequently indicated as a key parameter influencing microbial diversity and consequently shaping the system for a specific function [8–12].

Here, we explored the effects of pH increase based on previously developed model ecosystems, which use anaerobic fermentation reactors with enriched mixed cultures under well-controlled abiotic conditions [13,14]. By preventing continuous inoculation, such closed systems are simplified but still relatively complex considering microbial interactions and metabolic processes. Enrichment cultures can maintain their functional stability in a self-assembled manner, which is challenging for synthetic communities [15]. In our model ecosystem, we focus on the process of microbial chain elongation (CE) to produce the carboxylates *n*-butyrate (C4), *n*-caproate (C6) and *n*-caprylate (C8) [13]. Using xylan and lactate to simulate the feedstock conditions of anaerobic fermentation of ensiled plant biomass [16], the lactate-based CE coupled with *in situ* lactate formation holds promise to valorise organic wastes or biomass residues within the carboxylate platform [17]. Efficient and stable CE processes rely on trophic relationships among community members with diverse functions, in our case including xylan hydrolysis, xylose fermentation and lactate-driven CE, presenting diverse and parallel pathways for substrate metabolism in a food web [13].

Next-generation sequencing data (e.g., 16S rRNA amplicon sequencing) allow us to capture the dynamics of entire communities with a high phylogenetic resolution over long-term experimentation [7]. Such microbiome datasets are generated inherently with compositional bias in sparse sequencing count data because of the fixed capacity of sequencing instruments, and the hypothesis behind is that sequencing counts the number of molecules relevant to the

bacteria in the population [18]. It is acknowledged that only a proportion of counts per feature (e.g., amplicon sequencing variants – ASV) is available, thus compositional approaches were developed to avoid the common pitfalls in the analysis of relative abundance data. As reviewed by Gloor et al. [18], the workflow normally starts with a log-ratio transformation, then different tools of distancing for clustering and ordination, multivariate comparison and correlation analysis are included [19–21]. The latter is commonly unveiled by association network algorithms, inferring non-random co-occurrence patterns between community members and assessing microbial responses to environmental changes. In this study, both standard microbiome analysis and the compositional replacements were implemented to achieve statistically more robust results.

Besides the changed environmental factor pH, time needs to be considered as an important component in this long-term experimental study. Regularly frequent sampling with replicates over long time in microbial communities gives insight into their stability, development or response to and recovery from perturbation [9,22,23]. Linear mixed-effects models (LME) and variations thereof are commonly used for modelling time-resolved 16S rRNA amplicon data, thereby identifying temporal microbial interaction patterns [24,25]. We hypothesised that the pH value predominantly determines the assembly of CE reactor microbiota, but the impact of time needs to be disentangled by applying LME. The identified taxa with their temporal patterns are key to understand their roles in community assembly. Feature selection using random forest classification was performed to denote bioindicators of pH changes, and the genetic potential of these bioindicator taxa was investigated by functional annotation of the accessible metagenome-assembled genomes (MAGs) [14]. As for CE, it is still unclear how the different microorganisms interact and what conditions they thrive in. In this context, pH can be a critical parameter that affects these relationships, and ultimately the end products of CE. In this study, we focused on the effects of pH increase from three aspects: (i) the abundance of identified key players, particularly potential chain-elongating species, (ii) the responses of microbial interactions, and (iii) the CE community functioning represented mainly by the production of C4, C6 and C8.

Materials and methods

Reactor operation and sampling

The inoculum originated from a lab-scale caproate-producing bioreactor that was operated with corn silage [16]. We enriched the microbial community in a 1-L bioreactor (BIOSTAT® A plus, Sartorius AG, Göttingen, Germany) fed with mineral medium containing xylan and lactic acid over 150 days [13]. The enriched community producing C4, C6 and C8 was further selected by reducing the hydraulic retention time (HRT) in two parallel BIOSTAT bioreactors for almost one year [14]. The present study continued to shape the reactor microbiota by increasing the pH, with a fixed HRT of four days. Before starting the experiment, the microbial communities of both bioreactors (A and B) were equally distributed by pumping the content from A to B and back while keeping it anoxic.

The reactor configuration was similar as before [13], with both bioreactors operated at 38 ± 1°C, constantly stirred at 150 rpm and the pH automatically controlled by addition of 5 M NaOH. For daily feeding, 2.94 g lactic acid and 2.50 g water-soluble xylan were supplied in 0.25 L mineral medium. The starting pH was 5.5 in both bioreactors A and B. After 42 days, we increased the pH of bioreactor A from 5.5 to 6.0, and further to 6.5 from day 112 to day 238. By considering the effect of time on community assembly, a different temporal scheme of pH increase was applied in reactor B (pH 5.5: day 0-144, pH 6.0: day 145-214, pH 6.5: day 215-238).

Gas samples and liquid samples of reactor effluent were collected twice per week. The effluent was centrifuged and the supernatant was used for measuring concentrations of xylan, carboxylates and alcohols. Pelleted cells were stored at -20˚C for DNA-based community analysis [13].

Analytical methods

Daily gas production was monitored on-line with a gas counter as described [26]. Gas composition was determined in triplicate for H_2, CO_2, N_2 and O_2 by gas chromatography [27]. Concentrations of carboxylates and alcohols were

analysed in triplicate by gas chromatography, and xylan was measured by a modified dinitrosalicylic acid reagent method as reported [13]. At the beginning and the end of the experiment, cell mass concentration was calculated from optical density (OD) values correlated with cell dry mass [13], with mean correlation coefficients of 1 OD_{600} = 0.641 g L^{-1} for bioreactor A and 1 OD_{600} = 0.632 g L^{-1} for bioreactor B.

Total DNA was isolated from frozen pellets using the NucleoSpin Microbial DNA Kit (Macherey-Nagel, Düren, Germany). Methods for DNA quality control and quantification were reported before [28]. 16S rRNA genes were PCR-amplified using primers 341f and 785r [29], with high-throughput amplicon sequencing performed on the Illumina Miseq platform (Miseq Reagent Kit v3, 2 × 300 bp).

Microbiome data analysis

The QIIME 2 v2020.2 pipeline [30] with DADA2 plugin [31] was applied to demultiplex sequences, filter phiX reads, denoising, merging read pairs, trimming and removing chimeras of the sequences. A total of 6,855,572 sequences ranging from 21,389 to 66,272 pairs of reads per sample were obtained, with a median of 50,439 in 136 samples. A feature table was created indicating the frequency of each ASV clustered at 100% identity. ASVs with frequencies lower than two in less than three samples were filtered out before further analyses. Taxonomy was assigned with a naïve Bayes classifier trained on the database MiDAS 2.1 [32] and curated with RDP Classifier 2.2 [33] (confidence threshold: 80%). The filtered ASV table was rarefied to a depth of 21,389 reads for the downstream analyses (rarefaction curve reached the plateau, Figure S1). A total of 97 unique ASVs were retained.

Alpha diversity based on rarefied ASV data was evaluated by using the ecological indices including diversity, evenness and richness as described [34]. The indices of order one (^{1}D and ^{1}E) quantify the diversity and evenness by weighting all ASVs equally, while the indices of order two (^{2}D and ^{2}E) give more weight to the dominant taxa. Considering the compositional nature of the high-throughput sequencing data [18], we analysed the data with standard approaches and their compositional replacements. For dissimilarities in

community composition (beta diversity), we used both Bray-Curtis distance-based principle coordinate analysis (PCoA) [35], and Aitchison principle component analysis (PCA) via DEICODE that is robust to data sparsity [19]. The QIIME 2 plugin Qurro [36] was used to visualise and explore feature rankings in the produced DEICODE biplot. PERMANOVA ("adonis" function in R vegan package, v2.5.6; 999 permutations) [20] was used for statistical analyses of beta-diversity, with P values adjusted by the false discovery rate (FDR) method [37].

Statistical analysis of effects of pH increase on reactor microbiota time series

A redundancy analysis-based variation partitioning analysis (VPA) was used to provide a quantification of the relative contribution of individual process parameters (pH and time) and their interactive effects on temporal variation in microbial community composition. VPA was performed using "varpart" function in R package vegan. For each process parameter, we performed a partial Mantel test to examine its correlation with community composition represented by Aitchison and Bray-Curtis distances, independent of time (9999 permutations) using vegan.

The QIIME 2 plugin q2-longitudinal with default settings was used to construct the LME for regression analyses involving dependent data [25]. Random intercepts models (REML method) were used to track longitudinal changes of metrics including alpha- and beta-diversity and ASV abundances in microbial communities. In brief, pH and time were designated as fixed effects and bioreactor as a random effect, whereat values represent samples of a random collection. The response variables are the following metrics: ^1D, ^2D, ^1E, ^2E, richness, PC1 of Aitchison or Bray-Curtis and ASV abundance.

The Microbial Temporal Variability Linear Mixed Model (MTV-LMM) was used to identify autoregressive taxa and predict their relative abundances at later time points [24]. The model assumes that the temporal changes in relative abundance of ASVs are a time-homogenous high-order Markov process. To select the core time-dependent taxa, MTV-LMM was applied to each individual

pH level, which generated a temporal kinship matrix representing the similarity between every pair of normalised ASV abundances (a given time for a given individual) across time. A concept of time-explainability was introduced to quantify the temporal variance explained by the microbial community in previous time points.

Random forest (RF) classification

Supervised classification of pH levels on community compositions was performed using QIIME 2 q2-sample-classifier with default settings [38]. Rarefied ASV data were used as features to train and test the classifier. First, a nested cross-validation of the RF model was applied to overview the classification of the pH levels for all samples. For model optimisation, a second layer of cross validation (outer loop) was incorporated to split the dataset into training and test sets five times, and therefore each sample ended up in a test set once. During each iteration of the outer loop, the training set is split again five times in an inner loop to optimise parameter settings for estimation of that fold. Five different final models were trained, with each sample receiving a predicted value. The overall accuracy was calculated by comparing the predicted values to the true values.

Next we performed a feature selection by randomly picking 80% of the samples to train a RF classifier, and the remaining 20% of the samples were used to test classification accuracy of the classifier. K-fold cross-validation ($K = 5$) was performed during automatic feature selection and parameter optimisation steps to tune the model. As determined by using recursive feature elimination, the most important features that maximised model accuracy were selected. Model accuracy and predictions were based on the classifier that utilised the reduced feature set.

Network analysis

Co-occurrence networks based on rarefied ASV data and process parameter data were inferred by using the FlashWeave v0.16 implemented in Julia [21]. FlashWeave uses the centred log-ratio approach for the correction of compositional microbial abundances, and it infers direct associations. Three

networks were constructed from the data of the three individual pH levels, which featured a correlation coefficient below -0.5 or above 0.5. Another network was constructed from the entire data of all pH levels. All networks were visualised in Cytoscape v3.8.0 [39], with topological features analysed.

Results

Fluctuation and recovery of reactor performance

The pH increase from 5.5 to 6.0 caused fluctuations in the concentrations of CE products and lactate, which were not observed upon further increase to pH 6.5 (Figure 1). First we applied this pH increase in bioreactor A, which immediately presented an increased concentration (mmol C/L) of C8 up to 29.1, corresponding to a yield (C mole product to the transferred substrate ratio) of 5.2, and a relatively stable yield of C6 (mean of 16.0 ± 1.5 at pH 6.0). Lactate accumulated to a concentration of 147.5 while C4 concentration dropped to 69.1, with a yield of 12.1 (Figure 1a). The pH increase caused no effects on the fast consumption of xylan (Figure S2). Soon afterwards the accumulated lactate was consumed and C4 returned to the previous concentration of 273.9 on day 95 at pH 6.0. Interestingly, a further increase to pH 6.5 did not result in any fluctuations of the production of C4, C6 and C8 (Figure 1a). Later we replicated the pH increase from 5.5 to 6.5 in bioreactor B to confirm the observed impacts of pH increase on CE. With a longer operation at pH 5.5 for 144 days, comparable fluctuations in concentrations of lactate, C4 and C8 were observed, but presenting a delay of 38 days after the pH increase to 6.0. Concentrations of lactate, C4, C6 and C8 were relatively stable when bioreactor B was operated at pH 6.5. No propionate was detected in both bioreactors. The pH increase also resulted in fluctuations of daily gas production and gas composition including CO_2 and H_2 (Figure S3). Besides, a general upward trend of cell mass production suggests a facilitating effect of higher pH on the growth of enriched populations with CE functions (Figure S4).

Emergence of rare species and development in microbial community composition

As shown in Figure S5, the relative abundance of ASVs categorised from phylum to genus varied along with the pH gradients in such closed microbial systems. For example, the genera *Actinomyces* and *Prevotella* became apparent at pH 6.5, along with increasing abundance of *Clostridium sensu stricto* and decreasing abundances of *Clostridium* IV and *Eubacterium* (Figure S5e). After the pH increase, alpha diversity metrics showed decreases in diversity (^1D) and evenness (^1E), but increase in richness (Figure 2; similar results for ^2D and ^2E as shown in Figure S6). We used LME models to test whether these indices were impacted by pH and time. Three separate LME models were fitted to examine ^1D, ^1E and richness across pH gradients because the trajectories appear nearly linear. Diversity was significantly impacted by pH ($P < 0.001$) and time ($P < 0.001$), indicating that diversity was reduced much stronger by a factor of 6.188 by pH increase than by time with a factor of 0.209 (Table S1). Evenness and richness were also significantly associated with pH and time, although pH presented much stronger impacts on both indices (Tables S2-S3).

Beta diversity analysis including Aitchison distance-based PCA and Bray-Curtis distance-based PCoA revealed that the bacterial communities differed significantly between the three pH levels (PERMANOVA; $P < 0.001$) (Figure 3a and Figure S7). ASVs belonging to *Clostridium* IV, *Oscillibacter*, *Olsenella* and *Syntrophococcus* were strongly associated with the communities of pH 5.5 and 6.0 while *Clostridium sensu stricto* sp. ASV009 was most strongly associated with the community of pH 6.5 (Figure 3a). Based on the association with dissimilarities in community composition, ASVs represented by *Clostridium* IV sp. ASV008 (the lowest ranked taxa) and *Clostridium sensu stricto* sp. ASV009 (the highest ranked taxa) correspond to the most influential taxa driving the Aitchison PCA clustering (Figure 3b). Fitting LME models to their dynamics in relative abundance (Figure 3c), results showed that the relative abundance of ASV008 was significantly impacted by pH ($P < 0.001$) and time ($P = 0.002$), whereas only pH ($P < 0.001$) significantly impacted the abundance of ASV009, and time exhibited no significant effects ($P = 0.091$) (Tables S4-S5). In both cases, pH presented a much stronger impact than time. By applying LME models, we examined how beta diversity changed over time in each bioreactor

(Figure 3d-3e). Results indicated that pH was the most influencing factor, although time had significant effects as well (Tables S6-S7). The impact of pH on microbial community assembly was further confirmed by partial Mantel test. We correlated the time-corrected dissimilarities of community composition with pH, and the results show their strong, significant correlations based on Aitchison distance ($r_m = 0.61$, $P < 0.001$) and Bray-Curtis distance ($r_m = 0.72$, $P < 0.001$) (Table 1). Evaluating the overall contributions of pH and time by VPA, they together could explain 61% of the microbial community variations based on Bray-Curtis (Figure S8). 24% and 3% variations were independently explained by pH and time, respectively. These results support those inferred from the LME models.

pH bioindicators and time-dependent taxa

Overall, the nested cross-validation of RF classification represented an accuracy of 97.8% in prediction of the pH levels for all 136 samples (Figure S9), by using ASV data to follow community composition dynamics. We carried out recursive feature elimination with cross-validation, 18 most important features were selected that gave a perfect discrimination between three pH levels (Figure 4). These ASVs were defined as pH bioindicators, belonging to the genera *Clostridium* IV, *Syntrophococcus*, *Lactobacillus*, *Olsenella*, *Bulleidia*, *Clostridium sensu stricto*, *Eubacterium*, *Lachnospiraceae incertae sedis*, *Sporanaerobacter* and *Actinomyces* (Figure 4b). Among these pH bioindicators, four showed increases while 14 showed decreases in abundance along the pH increase. Interestingly, the most influential ASVs driving the Aitchison PCA clustering were also exhibited as pH bioindicators, including the abundant taxa *Clostridium* IV sp. ASV008 and *Clostridium sensu stricto* sp. ASV009 (Figure 4b).

By using MTV-LMM, we identified time-dependent taxa, whose abundance can be predicted based on the previous community composition. In this longitudinal study, 32, 25 and 40 ASVs were predicted to be significantly ($P < 0.05$) affected by the past composition of the community at pH 5.5, 6.0 and 6.5, respectively, with the time-explainability ranging from 17% to 80%, 17% to 83% and 13% to 96%, respectively (Figure S10). Inferring microbial interactions of these taxa

deserves more attention in understanding the dynamics of composition and functions of the community.

Microbial interaction patterns

Partial Mantel test showed the community composition significantly correlated with process performance and the changing conditions (Table 1). We consequently constructed an overall network and three separate networks for each pH level, to discern the succession of microbial interactions and to reveal potential metabolic functions. After the disturbance of pH increase to 6.0, more nodes and edges, as well as higher average clustering coefficient and heterogeneity were found, suggesting that the overall interaction intensity was higher at pH 6.5 (Table S8). In agreement with Aitchison PCA analysis, pH was significantly correlated with pH bioindicators ASV008 and ASV009 (Figure S11). Changes of interaction patterns over pH were shown in Figure 5. At the family level, *Ruminococcaceae* co-occurred with *Lachnospiraceae* and *Erysipelotrichaceae* at all pH levels, while it co-occurred with *Coriobacteriaceae* only at pH 5.5 (e.g., *Clostridium* IV sp. ASV090 with *Olsenella* sp. ASV049) and with *Actinomycetaceae* only at pH 6.5 (*Clostridium* IV sp. ASV037 with *Actinomyces* sp. ASV019), and with *Lactobacillus* at pH 6.0 and 6.5. *Clostridiaceae* 1 co-occurred with *Clostridiales Incertae Sedis* XI (*Clostridium sensu stricto* sp. ASV009 with *Sporanaerobacter* sp. ASV029) and *Erysipelotrichaceae* only after the pH increase to 6.0. *Erysipelotrichaceae* presented positive correlations with *Lactobacillus* at pH 6.0 and 6.5, where its negative correlation with *Coriobacteriaceae* vanished. Interestingly, the positive correlation of *Erysipelotrichaceae* (*Bulleidia* sp. ASV004) with *Lachnospiraceae* (*Syntrophococcus* sp. ASV001) vanished at pH 6.0. The positive correlation between C6 yield and *Eubacterium* sp. ASV015 was presented in the networks of overall, pH 5.5 and pH 6.0, but not pH 6.5 (Figures S11 and 5). In general, more relatively strong correlations ($|r| > 0.5$) emerged at pH 6.5, including the negative correlation of *Prevotella* sp. ASV041 with *Bulleidia* sp. ASV017.

Discussion

pH niches of chain elongators *Clostridium* IV and *Clostridium sensu stricto*

We recovered MAGs of *Clostridium* IV and *Clostridium sensu stricto* presenting the genetic potential of CE in our previous study [14], the communities of which were further shaped by gradual pH increase. Based on the statistically robust results of Aitchison PCA clustering coupled with LME models and RF classification, a clear conclusion can be drawn: mildly acidic pH values (lower than 6.0) are favourable for *Clostridium* IV while more neutral pH 6.5 is suitable for *Clostridium sensu stricto*. MAGs classification of *Clostridium* IV sp. ASV008 showed highest similarity to the lactate-based chain elongator *Ruminococcaceae* bacterium CPB6 (*Acutalibacteraceae* UBA4871 according to the Genome Taxonomy Database [40]), which was described to prefer mildly acidic pH (i.e. 5.5 - 6.0) and to suffer from low growth rates and long lag times at pH values above 6.0 [41]. MAGs of *Clostridium sensu stricto* sp. ASV009 showed highest similarity to *Clostridium luticellarii* (*Clostridium*_B *luticellarii* according to the Genome Taxonomy Database [40]), which presented optimal growth at pH 6.5 [42] and the ability of CE [43–46], but it needs to be investigated whether this species uses lactate as electron donor for CE. Particularly, these two ASVs represent time-dependent taxa that are key to understand the community assembly and can be used to characterise the temporal trajectories of the community. The pH preferences of *Clostridium* IV ASV008 and *Clostridium sensu stricto* ASV009 tie together with concepts in niche theory suggesting that microorganisms are able to live within a designated range of pH values, and outside this range, they are outcompeted by other, better adapted organisms [47]. Due to the growth optima of different populations, alteration of pH is an important tool to shape and control CE reactor microbiota.

pH as a key determinant of microbial community assembly

Regular and sufficiently dense sampling with replicates is crucial to capture compositional patterns of communities inferred from time series data [7,22].

Microbial interaction is a main factor affecting such time-dependent patterns. Given that pH had a much stronger association with community assembly than time did, we concluded that pH increase mainly determined variations in microbial interactions along pH gradients. Our former studies indicated that lactate-based CE driven by *Olsenella* is an essential feature when maintaining the pH at 5.5 [13,14]. Along with increasing pH, the lactic acid bacteria *Olsenella* cooperating with the chain elongator *Clostridium* IV was replaced by *Lactobacillus*. Both genera are xylose-consuming lactate producers according to the functional annotation of their MAGs (Table 2). An enriched community dominated by the co-occurrence of CE species and *Lactobacillus* was reported in a recent study [48], which also suggested inherent benefits of *in situ* lactate formation in CE [13]. The shift in a mutualism of lactate-consuming chain elongator and lactate producer revealed the plasticity of the CE microbiota food web. Additionally, the co-occurrence of phylogenetically closely related taxa may indicate their overlapping metabolic niches, such as the appearance of *Lactobacillus* spp. ASV003 and ASV011, *Syntrophococcus* spp. ASV001 and ASV013, *Clostridium* IV spp. ASV002 and ASV005 at all pH ranges.

As suggested by the storage effect, dormant rare taxa can germinate and become dominant under proper conditions [49,50]. In this study, the increase in richness can be explained by an abundance shift of some taxa from rare to abundant, reflecting strong inhibition effects of lower pH on these taxa. With the increased number of microbial interactions and increasing interaction intensity strongly coupled to these taxa at higher pH, the factor pH shaping the community assembly was revealed by considering the growth and interactions of community members in such long-term closed systems.

Besides, indirect effects of pH cannot be ignored. At higher pH, the concentrations of undissociated carboxylic acids (e.g., C6/C8) are lower, which are known growth inhibitors of the CE community members [11,51–54]. The effects of pH on proton concentration changes and CO_2-HCO_3^- equilibrium in a lactate conversion (e.g., 3 lactate$^-$ + 2 H_2O → caproate$^-$ + 3 HCO_3^- + H^+ + 2 H_2) can cause the actual Gibbs energy change and further the energy release during CE with lactate [55].

Community changes do not always affect community functioning

We assumed that increase in pH would induce shifts in the community and consequently in community functioning. However, unlike in a complex, open CE system [16], increasing pH exhibited no strong effects on CE community functioning, which means those changes in community composition did not necessarily lead to the improvement in carboxylate production during long-term reactor operation. This agrees with the rare associations we observed for the ASVs to process parameters in the networks. Without introducing new microorganisms by inoculation, the emergence of rare species indicated high functional redundancy despite the reactors being closed systems. The reactor performance returned to the previous state after the fluctuation in carboxylate production along pH gradients, reflecting those coexisting rare taxa can increase functional resilience to environmental disturbances. Interestingly, the disturbance caused dramatic but transient increases of C6 and C8 yields. How to maintain such disturbance effects needs to be investigated systematically. Indeed, improving functional redundancy deserves equal importance as maximising the carboxylate production, because the presented parallel pathways of substrate conversion are essential to guarantee the functional stability during perturbation [9,11,56].

Other studies reported that low pH values favour CE in mixed culture fermentation, mainly because higher pH would support competing processes, such as methanogenesis [57] and the acrylate pathway for propionate production [10,58]. Results showed that our communities possibly lack these pathways and the corresponding functional species. Although pH is not an effective tool to manage the reactor microbiota towards beneficial traits of high specificities of C6/C8 production, it is clear that effects of alterations in pH are not universal for all reactor microbiota simply because of the difference in initial community composition.

Data availability

All data described in this manuscript are present in the paper and/or the Supplementary material. Amplicon sequencing data (ERR4450775 to ERR4450910) have been deposited to the ENA database under study no. PRJEB39808.

Supplementary information

Supplementary Figures and Tables: Figures S1-S11 and Tables S1-S8.

References

1. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Science. 2018;361:469–74.

2. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev. 2013;77:342–56.

3. Antwis RE, Griffiths SM, Harrison XA, Aranega-Bou P, Arce A, Bettridge AS, et al. Fifty important research questions in microbial ecology. FEMS Microbiol Ecol. 2017; 93:1–10.

4. Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK. Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. ISME J. 2018;12:1072–83.

5. Fargione J, Brown CS, Tilman D. Community assembly and invasion: An experimental test of neutral versus niche processes. Proc Natl Acad Sci USA. 2003;100:8916–20.

6. Chesson P. Mechanisms of maintenance of species diversity. Annu Rev Ecol Syst. 2000;31:343–66.

7. Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, et al. Signatures of ecological processes in microbial community time series. Microbiome. 2018;6:120.

8. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat Microbiol. 2019;4:1183–95.

9. Werner JJ, Knights D, Garcia ML, Scalfone NB, Smith S, Yarasheski K, et al. Bacterial community structures are unique and resilient in full-scale bioenergy systems. Proc Natl Acad Sci USA. 2011;108:4158–63.

10. Kucek LA, Nguyen M, Angenent LT. Conversion of L-lactate into n-caproate by a continuously fed reactor microbiome. Water Res. 2016;93:163–71.

11. Agler MT, Werner JJ, Iten LB, Dekker A, Cotta MA, Dien BS, et al. Shaping reactor microbiomes to produce the fuel precursor n-butyrate from pretreated cellulosic hydrolysates. Environ Sci Technol. 2012;46:10229–38.

12. 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.

13. 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.

14. Liu B, Sträuber H, Saraiva J, Harms H, Silva SG, Kleinsteuber S, da Roche UN. Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture. Submitted.

15. Lawson CE, Harcombe WR, Hatzenpichler R. Common principles and best practices for engineering microbiomes. Nat Rev Microbiol. 2019;17:725–41.

16. 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.

17. 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.

18. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. Front Microbiol. 2017;8:2224.

19. Martino C, Morton JT, Marotz CA, Thompson LR, Tripathi A, Knight R et al. A novel sparse compositional technique reveals microbial perturbations. MSystems. 2019;4:e00016-19.

20. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32–46.

21. Tackmann J, Rodrigues JFM, von Mering C. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. Cell Syst. 2019;9:286-296.

22. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. Metagenomics meets time series analysis: Unraveling microbial community dynamics. Curr Opin Microbiol. 2015;25:56–66.

23. Ryo M, Aguilar-Trigueros CA, Pinek L, Muller LAH, Rillig MC. Basic principles of temporal dynamics. Trends Ecol Evol. 2019;34:723–33.

24. Shenhav L, Furman O, Briscoe L, Thompson M, Silverman JD, Mizrahi I, et al. Modeling the temporal dynamics of the gut microbial community in adults and infants. PLoS Comput Biol. 2019;15:e1006960.

25. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, et al. q2-longitudinal: Longitudinal and paired-sample analyses of microbiome data. mSystems. 2018;3:e00219-18.

26. Sträuber H, Bühligen F, Kleinsteuber S, Dittrich-Zechendorf M. Carboxylic acid production from ensiled crops in anaerobic solid-state fermentation - trace elements as pH controlling agents support microbial chain elongation with lactic acid. Eng Life Sci. 2018;0:447–58.

27. 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.

28. Lucas R, Kuchenbuch A, Fetzer I, Harms H, Kleinsteuber S. Long-term monitoring reveals stable and remarkably similar microbial communities in parallel full-scale biogas reactors digesting energy crops. FEMS Microbiol Ecol. 2015;91:fiv004.

29. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 2013;41:e1.

30. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Chase J, Cope EK, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:852–7.

31. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

32. McIlroy SJ, Kirkegaard RH, McIlroy B, Nierychlo M, Kristensen JM, Karst SM, et al. MiDAS 2.0: An ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database. 2017;2017:1–9.

33. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences. Appl Environ Microbiol. 2007;73:5261–7.

34. Lucas R, Groeneveld J, Harms H, Johst K, Frank K, Kleinsteuber S. A critical evaluation of ecological indices for the comparative analysis of microbial

communities based on molecular datasets. FEMS Microbiol Ecol. 2017;92:fiw209.

35. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr. 1957;27:325–49.

36. Fedarko MW, Martino C, Morton JT, González A, Rahman G, Marotz CA, et al. Visualizing 'omic feature rankings and log-ratios using Qurro. NAR Genomics Bioinforma. 2020;2:lqaa023.

37. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple Testing. J R Stat Soc B. 1995;57:289–300.

38. Bokulich N, Dillon M, Bolyen E, Kaehler B, Huttley G, Caporaso J. q2-sample-classifier: machine-learning tools for microbiome classification and regression. J Open Source Softw. 2018;3:934.

39. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

40. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996.

41. 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.

42. 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.

43. van Brabant, P. Understanding bio-isomerisation during methanol fermentation. 2019. M.Sc. Thesis, Ghent University; Catalog number: rug01:002785053.

44. 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.

45. de Smit SM, de Leeuw KD, Buisman CJN, Strik DPBTB. Continuous *n*-valerate formation from propionate and methanol in an anaerobic chain elongation open-culture bioreactor. Biotechnol Biofuels. 2019;12:132.

46. 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.

47. Morton JT, Sanders J, Quinn RA, McDonald D, Gonzalez A, Vázquez-Baeza Y, et al. Balance trees reveal microbial niche differentiation. mSystems. 2017;2:e00162-16.

48. Contreras-Dávila CA, Carrión VJ, Vonk VR, Buisman CNJ, Strik DPBTB. Consecutive lactate formation and chain elongation to reduce exogenous chemicals input in repeated-batch food waste fermentation. Water Res. 2020;1:115215.

49. Chesson PL, Warner RR. Environmental variability promotes coexistence in lottery competitive systems. Am Nat. 1981;117:923–43.

50. Lennon JT, Jones SE. Microbial seed banks: The ecological and evolutionary implications of dormancy. Nat Rev Microbiol. 2011;9:119–30.

51. Andersen SJ, de Groof V, Khor WC, Roume H, Props R, Coma M, et al. A *Clostridium* group IV species dominates and suppresses a mixed culture fermentation by tolerance to medium chain fatty acids products. Front Bioeng Biotechnol. 2017;5:8.

52. Ge S, Usack JG, Spirito CM, Angenent LT. Long-term *n*-caproic acid production from yeast-fermentation beer in an anaerobic bioreactor with continuous product extraction. Environ Sci Technol. 2015;49:8012–21.

53. Royce LA, Liu P, Stebbins MJ, Hanson BC, Jarboe LR. The damaging effects of short chain fatty acids on *Escherichia coli* membranes. Appl Microbiol Biotechnol. 2013;97:8317–27.

54. Jarboe LR, Royce LA, Liu P. Understanding biocatalyst inhibition by carboxylic acids. Front Microbiol. 2013;4:272.

55. Kleerebezem R, van Loosdrecht MCM. A generalized method for thermodynamic state analysis of environmental systems. Crit Rev Environ Sci Technol. 2010;40:1–54.

56. Hashsham SA, Fernandez AS, Dollhopf SL, Dazzo FB, Hickey RF, Tiedje JM, et al. Parallel processing of substrate correlates with greater functional stability in methanogenic bioreactor communities perturbed by glucose. Appl Environ Microbiol. 2000;66:4050–7.

57. Grootscholten TIM, Strik DPBTB, Steinbusch KJJ, Buisman CJN, Hamelers HVM. Two-stage medium chain fatty acid (MCFA) production from municipal solid waste and ethanol. Appl Energy. 2014;116:223–9.

58. Candry P. Kinetics and microbial ecology of chain elongation for production of medium-chain carboxylic acids. 2020. PhD thesis. Ghent University, Belgium.

Figures
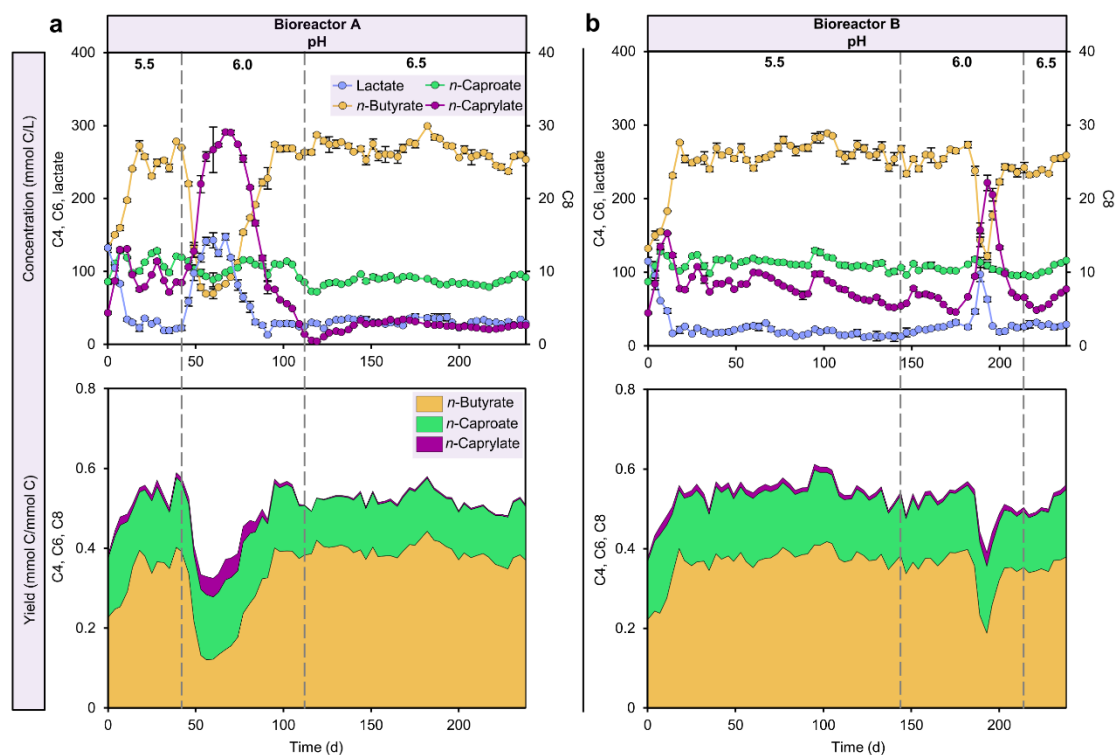
Figure 1. Performance of bioreactors. Concentrations of chain elongation products and lactate, as well as yields of chain elongation products in bioreactors A (**a**) and B (**b**) at three pH levels. Chain elongation products: C4, *n*-butyrate; C6, *n*-caproate; C8, *n*-caprylate.
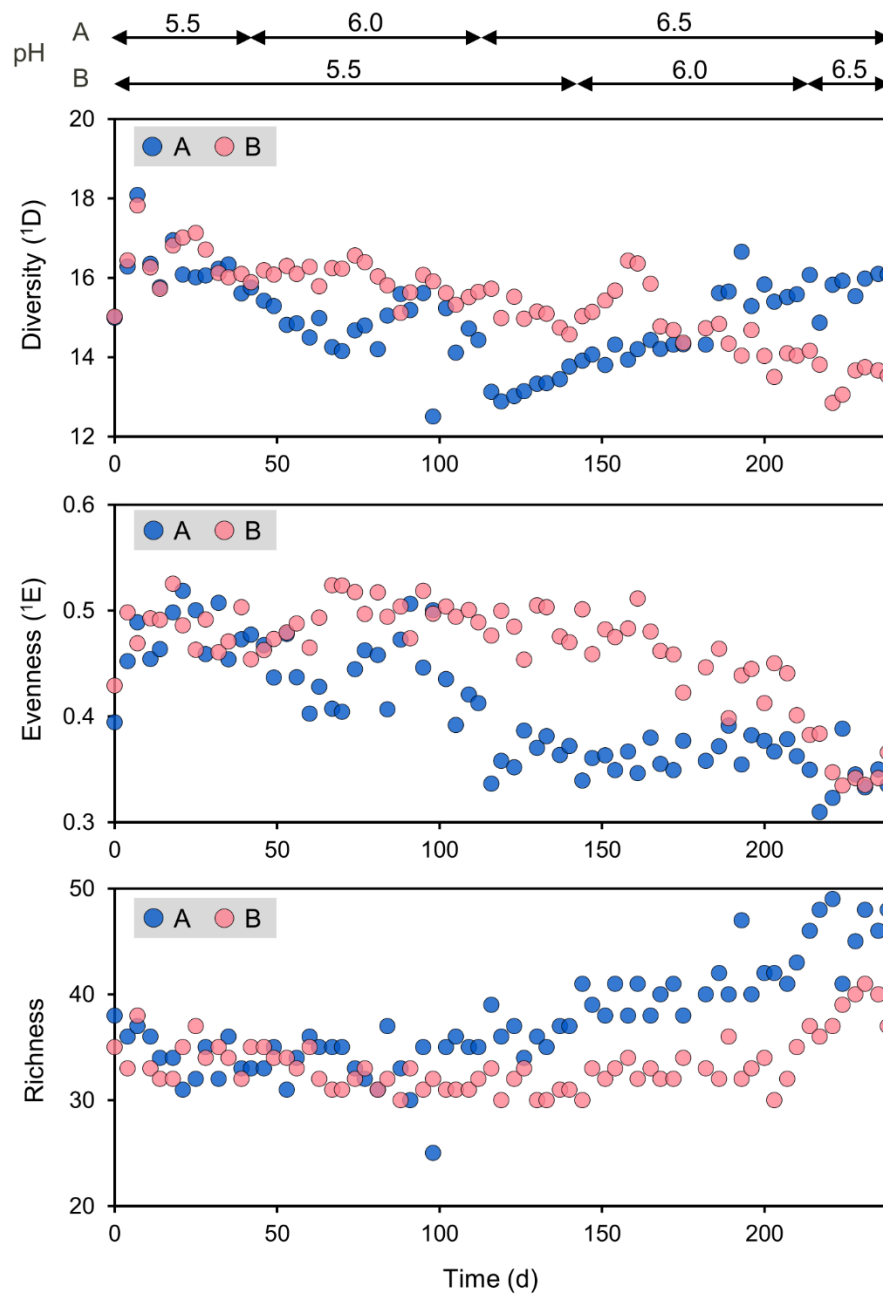
Figure 2. Longitudinal changes in alpha diversity at three pH levels. Based on the relative abundance of ASVs, we calculated the alpha diversity represented by diversity of order one (^{1}D) (**a**), evenness of order one (^{1}E) (**b**) and richness (**c**). Diversity and evenness of order one were quantified by weighting all present types equally. Results of linear mixed-effects model for alpha diversity can be found in Tables S1-S3. A and B stand for bioreactors A and B.
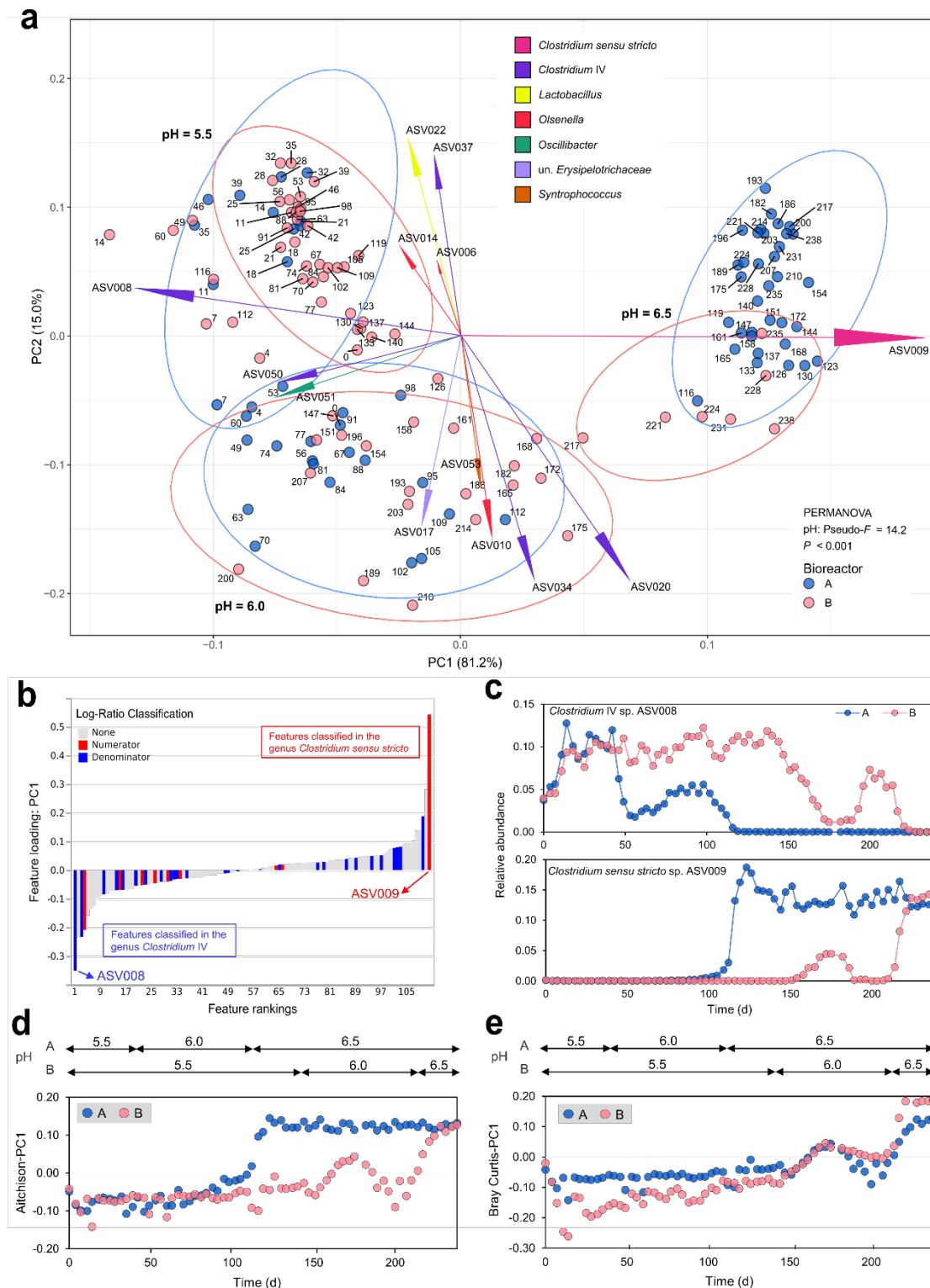
Figure 3. Effects of pH increase and time on bacterial community composition. a, A variance-based compositional principle component analysis (PCA) biplot based on Aitchison distance. Dots are named according to sampling days. Ellipses of 95% confidence intervals were added to each individual pH levels of the bioreactors. The size of an ASV arrow indicate the

strength of the relationship of that ASV to community composition. ASVs are coloured by family. **b,** ASV ranks estimated from Aitchison distance-based PCA (PC1) with *Clostridium* IV and *Clostridium sensu stricto* highlighted. **c,** Longitudinal changes in relative abundances of *Clostridium* IV sp. ASV008 and *Clostridium sensu stricto* sp. ASV009 at the three pH levels. Results of linear mixed-effects model can be found in Tables S4-S5. **d,e,** Longitudinal changes in bacterial community composition at the three pH levels, based on Aitchison (**d**) and Bray-Curtis (**e**) dissimilarities. Results of linear mixed-effects model can be found in Tables S6-S7.
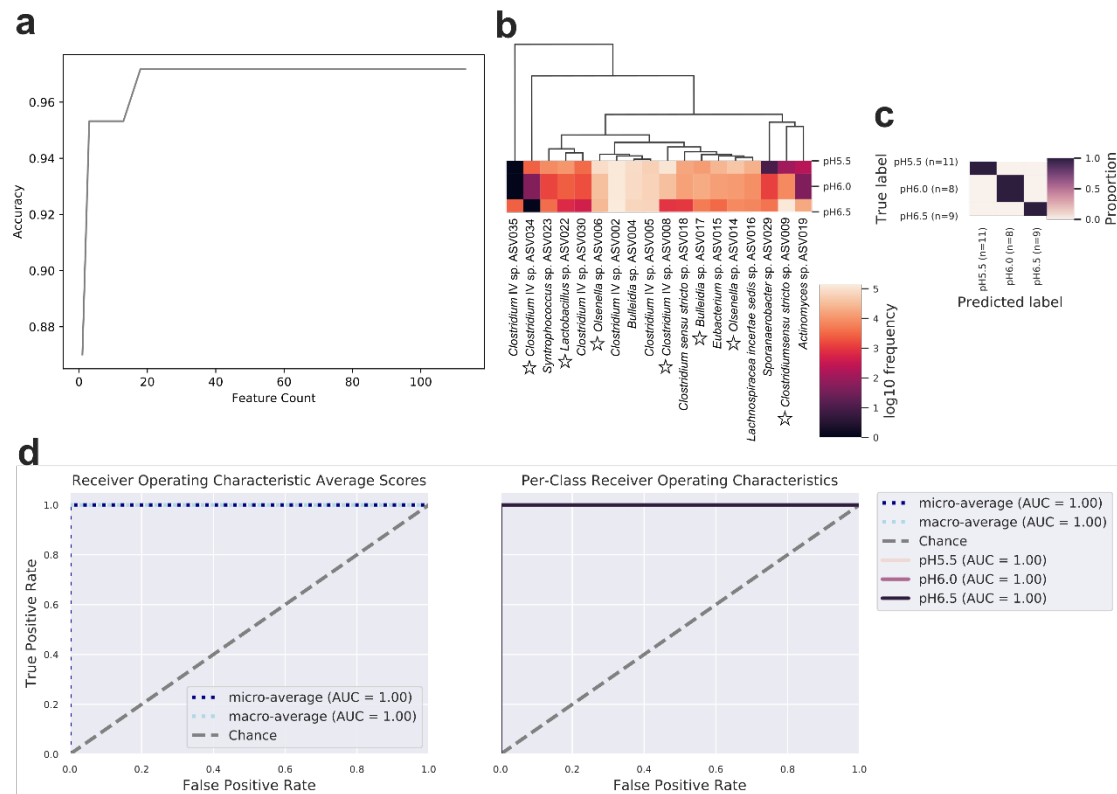
Figure 4. pH bioindicators determined by random forest classification accurately predict the different pH levels. a, A recursive feature elimination plot illuminates the model accuracy changes as a function of ASV count. The top-ranked 18 ASVs (pH bioindicators) that maximise accuracy are automatically selected for optimising the model, based on their mean decrease in Gini scores, according to their ASV abundance distribution, with pH as the response variable. **b,** A heatmap shows dynamics of the mean abundance of

pH bioindicators at the different pH levels. ASVs shown in Aitchison PCA biplot are indicated by a star. **c,** Confusion matrix for the optimal classifier of samples at different pH levels. The classifier was trained on the randomly picked 80% of the samples, which was then tested on the remaining 20%. Overall accuracy was calculated by comparing the predicted values to the true values. **d,** The Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) curves represent the classification accuracy of random forest. The ROC curve plots the relationship between the true positive rate and the false positive rate at various threshold settings. The AUC indicates the probability that the classifier ranks a randomly chosen sample of the given class higher than other classes. The random chance is represented as a diagonal line extending from the lower-left to the upper-right corner. In addition to show the ROC curves for each class, average ROCs and AUCs were calculated. "Micro-averaging" calculates metrics globally by averaging across each sample; hence class imbalance impacts this metric. "Macro-averaging" gives equal weight to the classification of each sample.
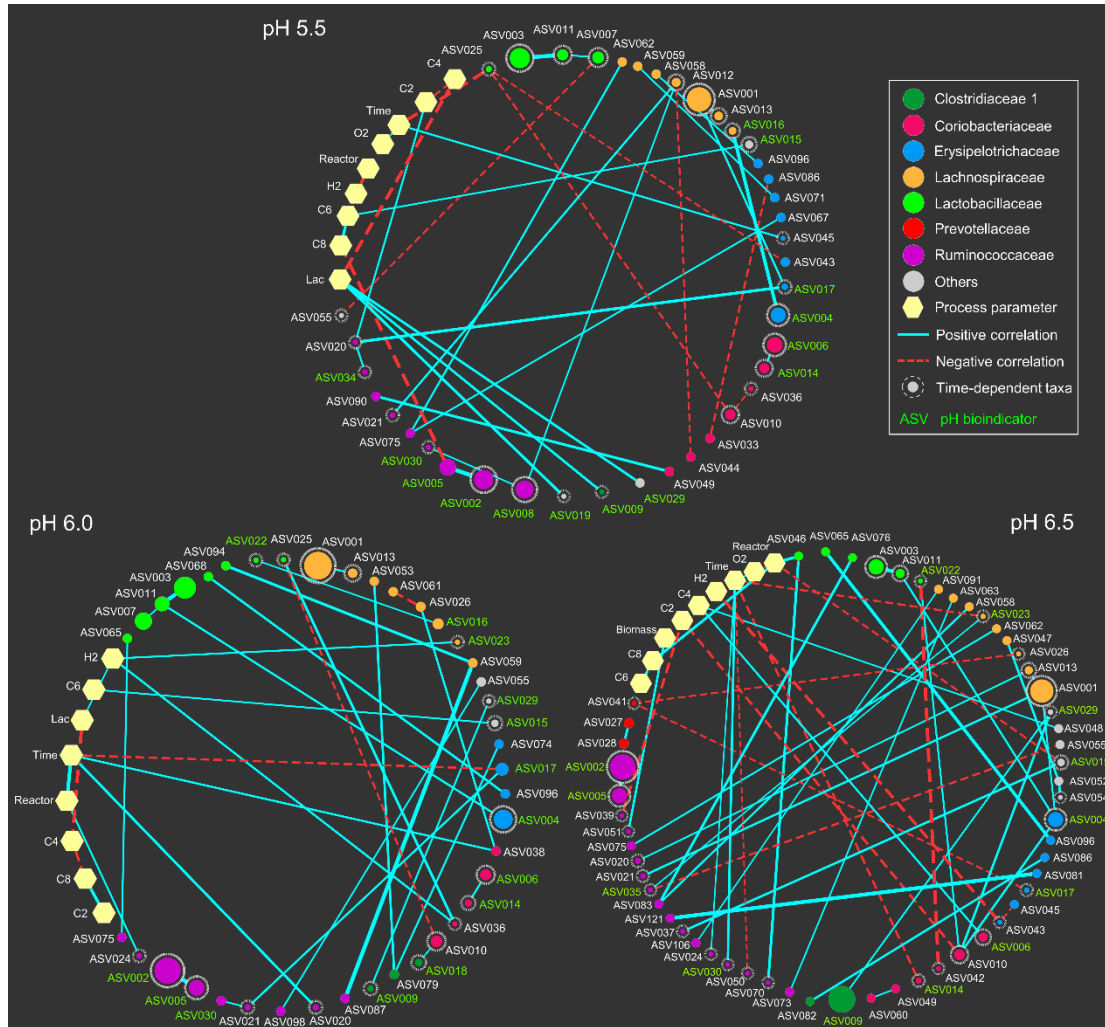
Figure 5. Co-occurrence networks for the three individual pH levels. Edges indicate a coefficient > 0.5 for positive correlations and < -0.5 for negative correlations. Edge thickness reflects the strength of the correlation. The size of each ASV node is proportional to the mean relative abundance over the corresponding pH level. ASV nodes are coloured and grouped by family. ASV nodes with grey dash borders are those time-dependent taxa of each individual pH level, whose abundance can be predicted based on the previous microbial community composition. pH bioindicators identified by random forest classification are shown with green letters. "Others" include the ASVs belonging to families *Eubacteriaceae* (ASV015), *Actinomycetaceae* (ASV019), *Clostridiales Incertae Sedis* XI (ASV029), *Microbacteriaceae* (ASV048), *Veillonellaceae* (ASV052, ASV054) and *Nocardiaceae* (ASV055). Lac, lactate concentration; C2, acetate yield; C4, *n*-butyrate yield; C6, *n*-caproate yield; C8, *n*-caprylate yield.

Table 1. Partial Mantel tests showing significant correlations between the time-corrected dissimilarities of microbial community composition and process parameters.

Process parameter	Aitchison distance		Bray-Curtis distance	
	r_m[a]	P[b]	r_m	P
pH	0.61	< 0.001	0.72	< 0.001
Conc. C2[c]	0.27	< 0.001	0.18	< 0.001
Conc. C4	0.07	0.013	-0.01	0.569
Conc. C6	0.29	< 0.001	0.48	< 0.001
Conc. C8	0.25	< 0.001	0.16	< 0.001
Conc. lactate	0.02	0.258	0.01	0.401
Conc. biomass	0.16	< 0.001	0.11	0.002
Yield C2	0.27	< 0.001	0.15	< 0.001
Yield C4	0.09	0.004	0.00	0.448
Yield C6	0.38	< 0.001	0.40	< 0.001
Yield C8	0.22	< 0.001	0.13	0.003
Yield biomass	0.09	0.001	0.06	0.037
O_2	0.43	< 0.001	0.44	< 0.001
CO_2	0.14	< 0.001	0.18	< 0.001
H_2	0.19	< 0.001	0.16	< 0.001
Time	0.14	< 0.001	0.33	< 0.001

[a]r_m, the correlation coefficient based on partial Mantel test, in which time was controlled. The permutation test compares the original r_m to r_m computed in 9999 random permutations.

[b]The reported P value is one-tailed.

[c]Conc., concentration

Table 2. Metagenome-assembled genomes (MAGs) with the same taxonomy as ASVs.

ASV	Taxonomic classification[a]						Representative MAG	Accession number[b]
	Phylum	Class	Order	Family	Genus	Species		
Clostridium IV sp. ASV008	*Firmicutes_A*	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA1033	UBA1033 sp002399935	UMB00014	GCA_903789645
	Firmicutes_A	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA1033	UBA1033 sp002407675	UMB00060	GCA_903789675
	Firmicutes_A	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA1033	UBA1033 sp002409675	UMB00097	GCA_903789585
	Firmicutes_A	*Clostridia*	*Oscillospirales*	*Acutalibacteraceae*	UBA4871	UBA4871 sp002399445	UMB00016	GCA_903789565
Clostridium sensu stricto sp. ASV009	*Firmicutes_A*	*Clostridia*	*Clostridiales*	*Clostridiaceae*	*Clostridium_B*	*Clostridium_B* sp003497125	UMB00080	GCA_903789665
Olsenella sp. ASV049	*Actinobacteriota*	*Coriobacteriia*	*Coriobacteriales*	*Atopobiaceae*	*Olsenella_B*	*Olsenella_B* sp000752675	UMB00010	GCA_903789475
	Actinobacteriota	*Coriobacteriia*	*Coriobacteriales*	*Atopobiaceae*	*Olsenella_C*	unclassified	UMB00003	GCA_903789455
	Actinobacteriota	*Coriobacteriia*	*Coriobacteriales*	*Atopobiaceae*	*Olsenella*	unclassified	UMB00074	GCA_903789485
Lactobacillus spp. ASV046, ASV065, ASV076	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus_H*	*Lactobacillus_H mucosae*	UMB00017	GCA_903789575
	Firmicutes	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	unclassified	UMB00041	GCA_903789695
	Firmicutes	*Bacilli*	*Lactobacillales*	*Lactobacillaceae*	*Lactobacillus*	*Lactobacillus amylovorus*	UMB00015	GCA_903789705

[a]Taxonomy refer to the Genome Taxonomy Database (GTDB) phylogenomic classification
[b]Accession numbers refer to the European Nucleotide Archive (ENA)

2.3.2 Supplementary information
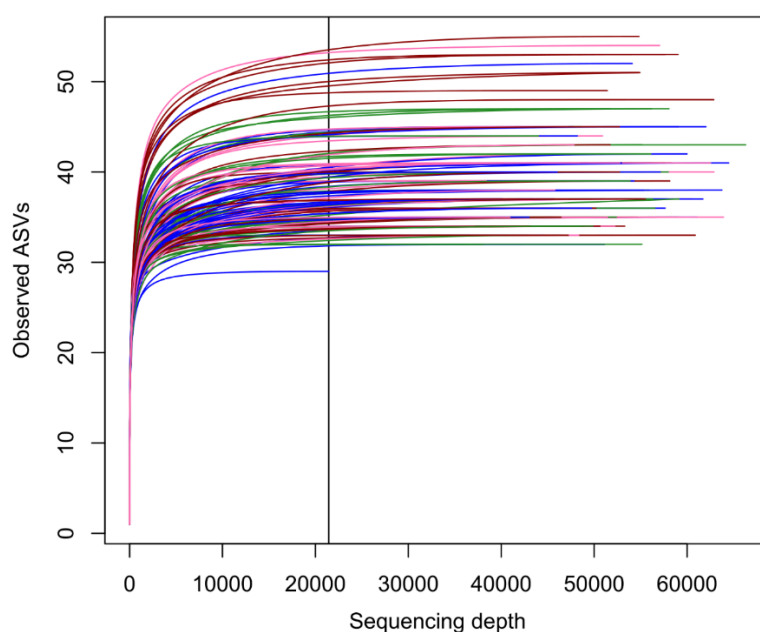
Figure S1. Alpha rarefaction curves. ASVs of all samples were rarefied to an equal sequencing depth of 21,389 reads. Colours represent the different samples.
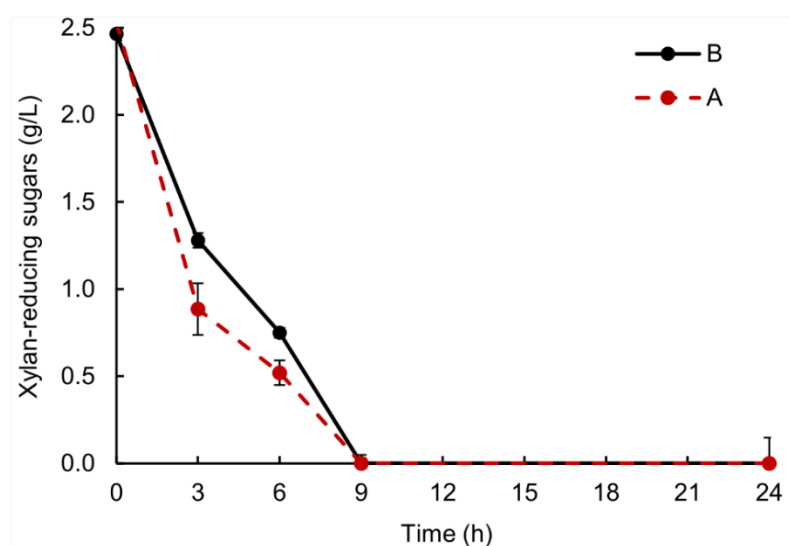
Figure S2. Daily consumption of xylan in bioreactors. During the fluctuations at pH 6.0 (day 67), an intensive sampling shows the fed water-soluble xylan was fast consumed in both bioreactors. A and B stand for bioreactors A and B. Error bars represent the standard deviation.
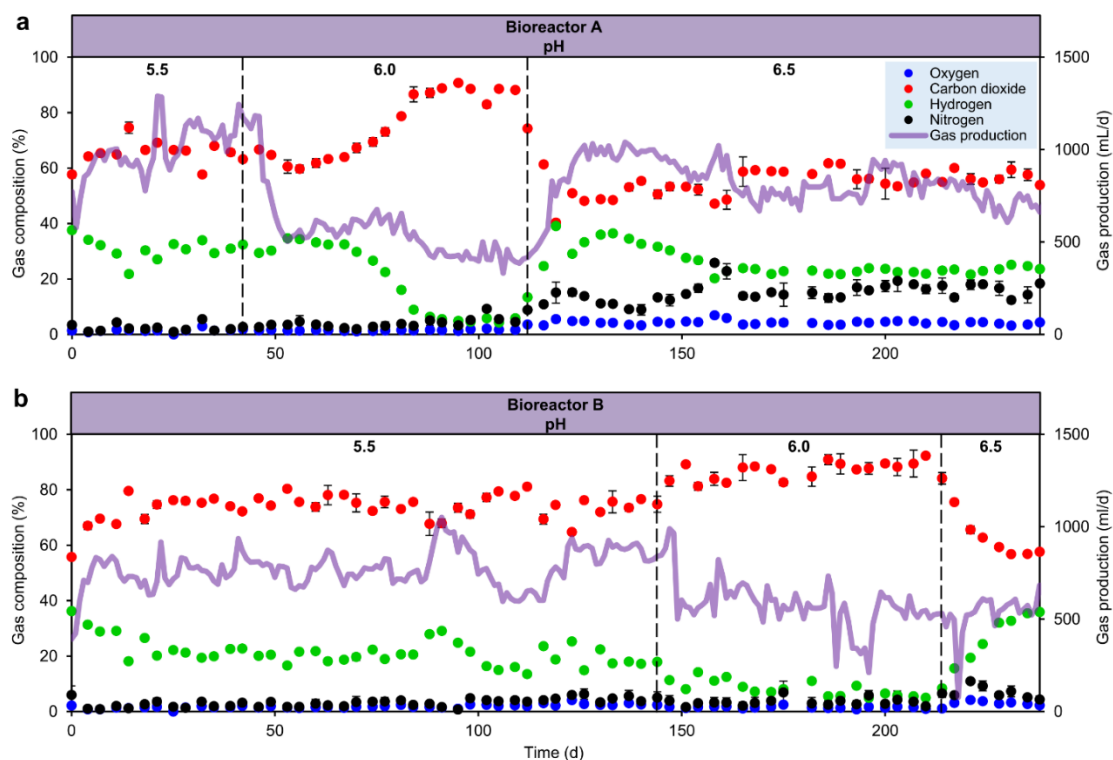
Figure S3. Gas production of bioreactors. Daily gas production and composition in bioreactors A (**a**) and B (**b**), respectively, at three pH levels. Error bars indicate the standard deviation.
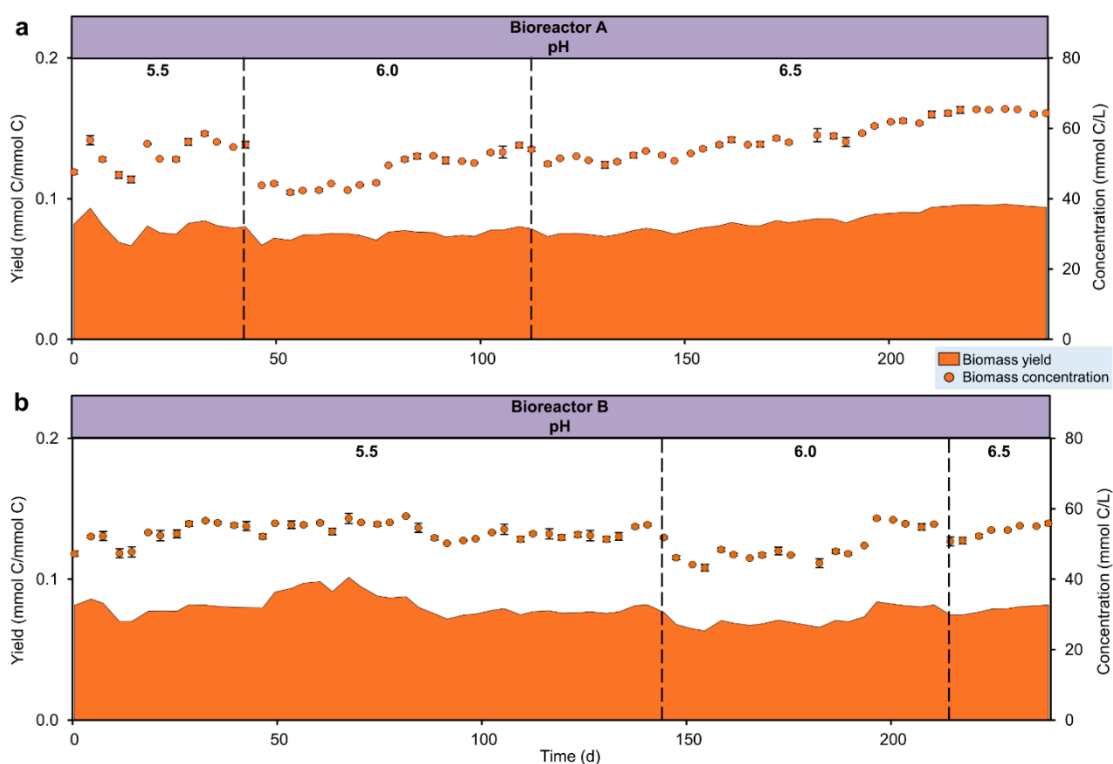
Figure S4. Biomass production of bioreactors. Cell concentration and biomass yield in bioreactors A (**a**) and B (**b**) at three pH levels. The carbon number of cell biomass was calculated by assuming an elemental biomass composition of $CH_{1.8}O_{0.5}N_{0.2}$ (molar mass = 24.6 g mol^{-1}). Error bars represent the standard deviation.
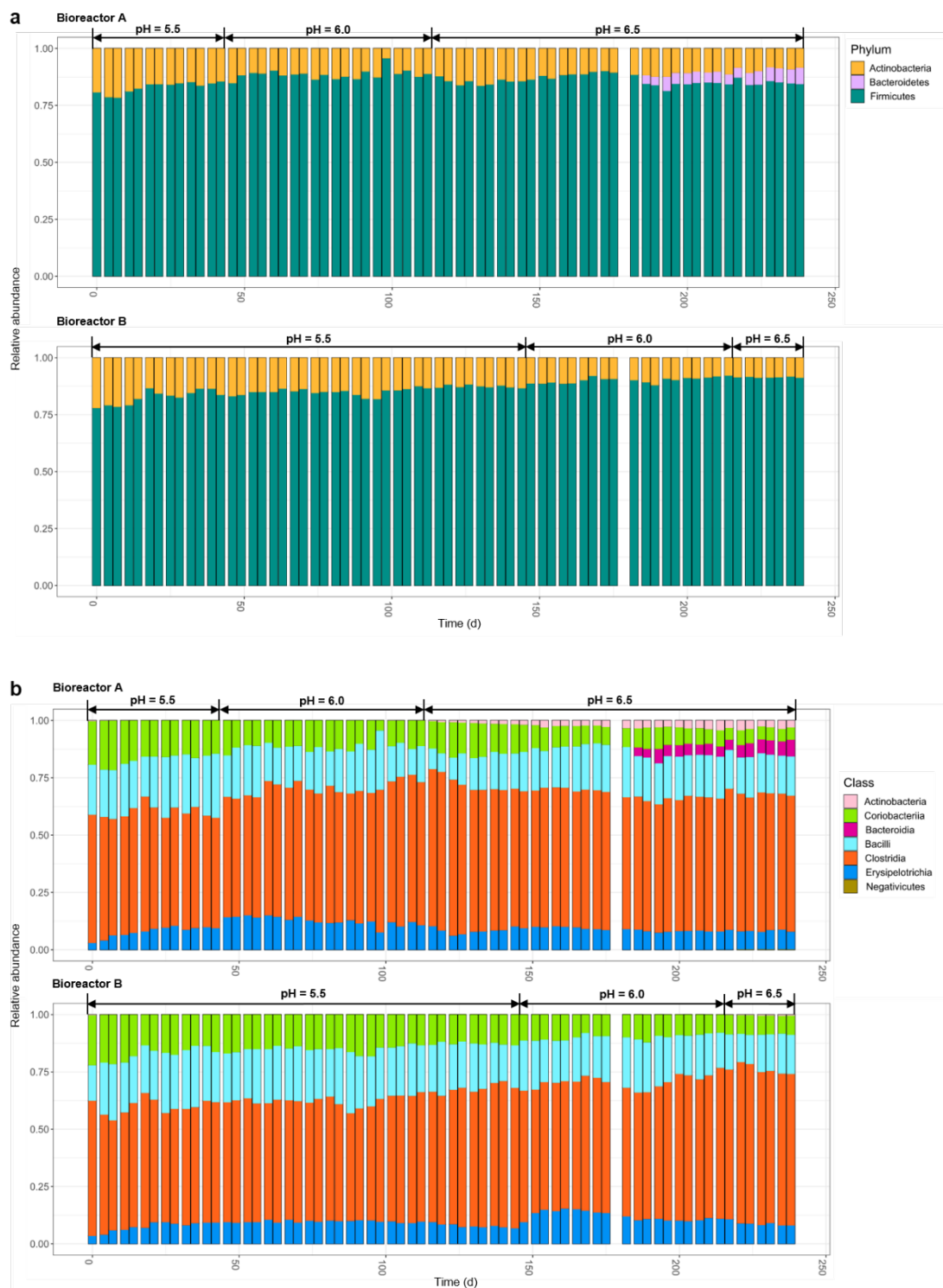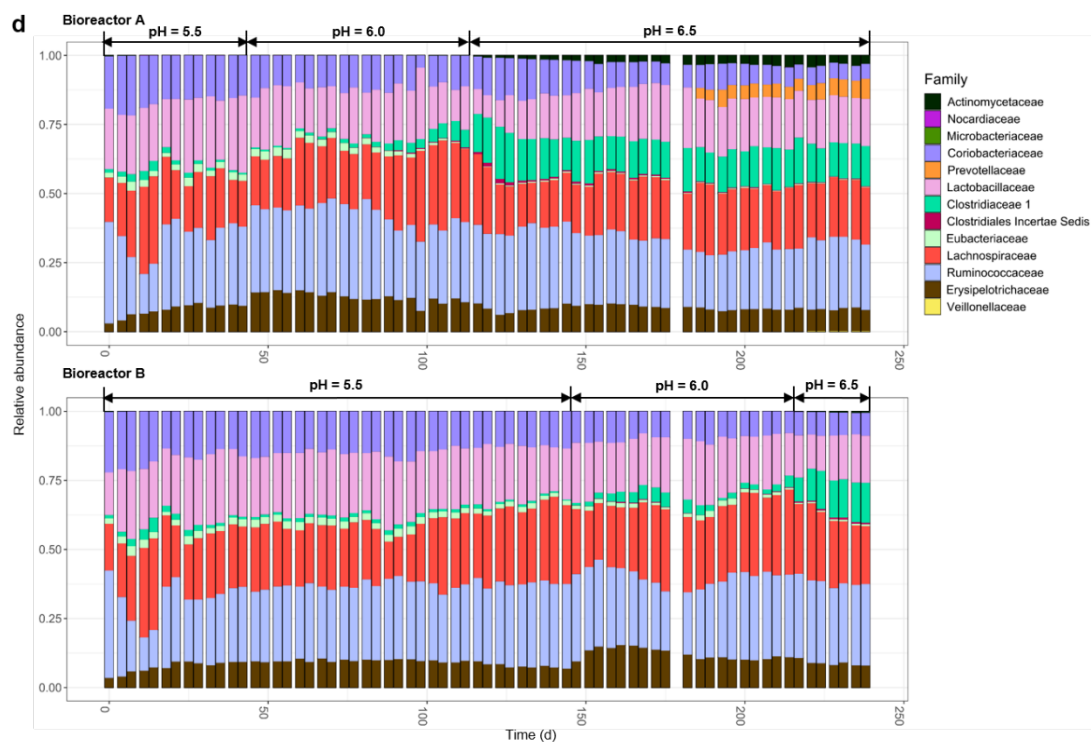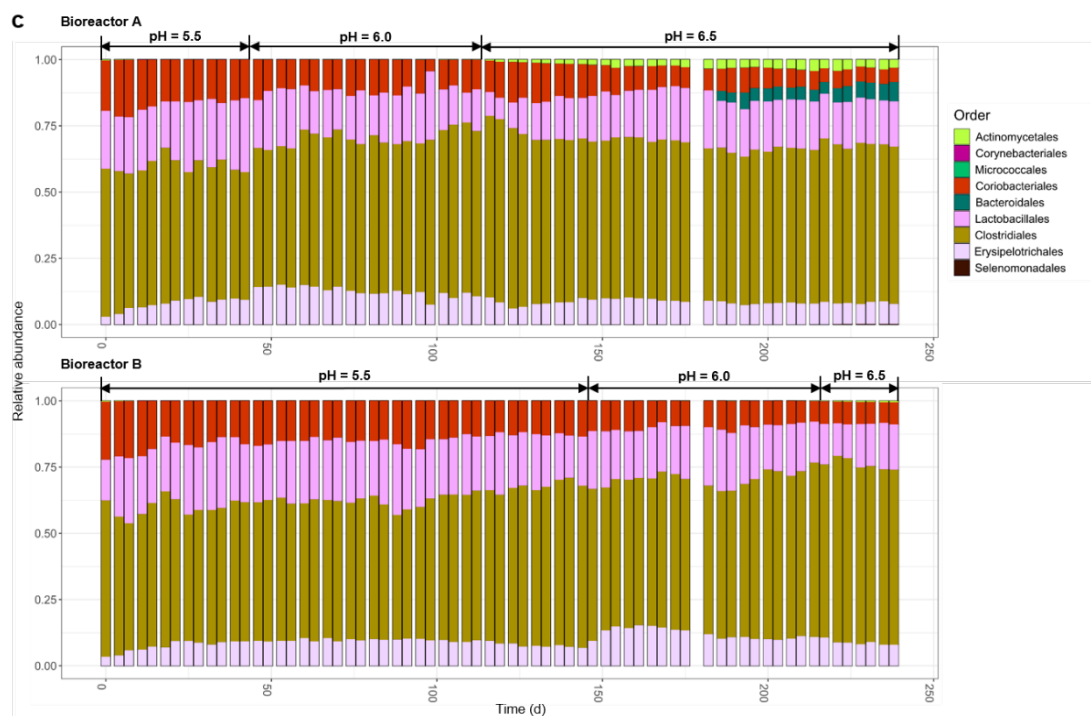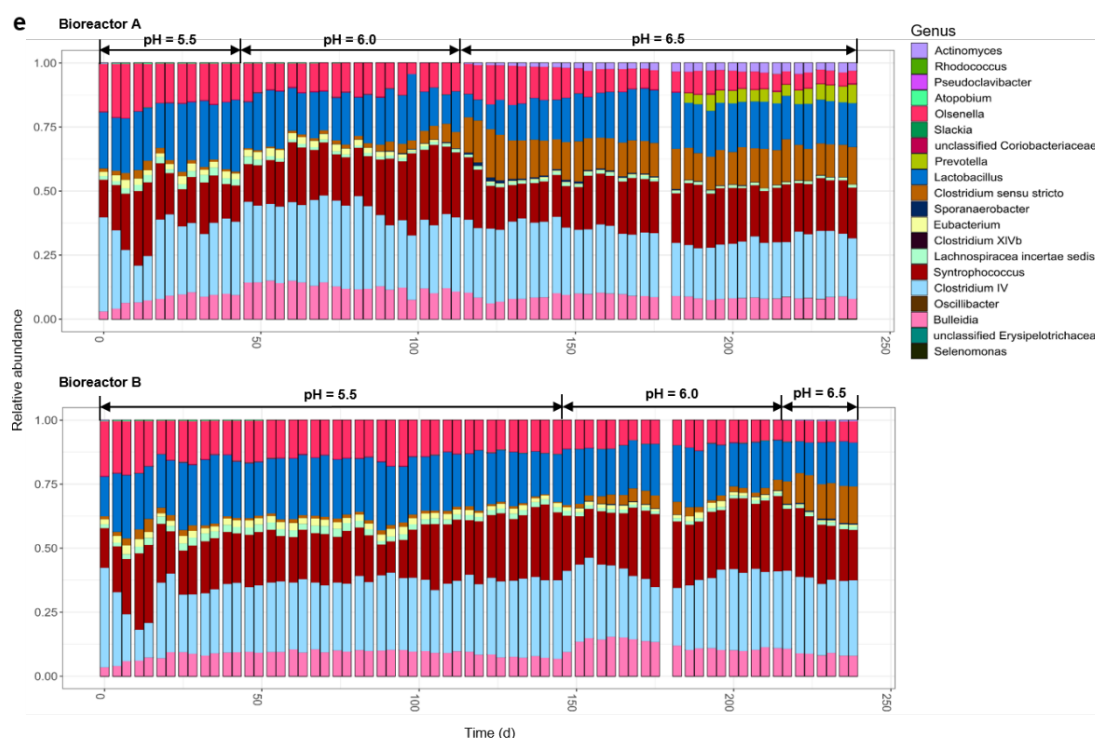
Figure S5. Microbial community composition profiles of bioreactors.
Based on amplicon sequencing of 16S rRNA genes, the taxonomic classification of amplicon sequence variants (ASVs) was categorised at the phylum (**a**), class (**b**), order (**c**), family (**d**) and genus (**e**) levels.
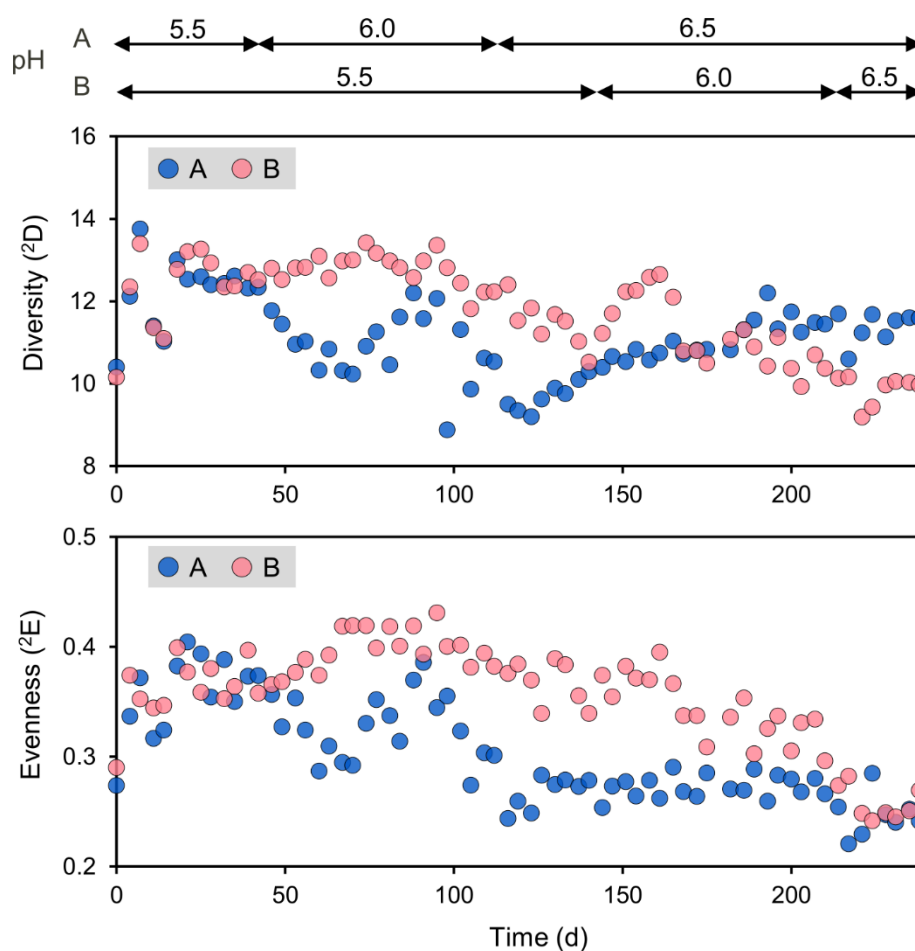
Figure S6. Longitudinal changes in diversity and evenness of order two of bioreactor communities. Based on the relative abundance of ASVs, we calculated the alpha diversity represented by diversity of order two (^2D) and evenness of order two (^2E), which give more weight to the dominant types than to the rare types. A and B stand for bioreactors A and B.
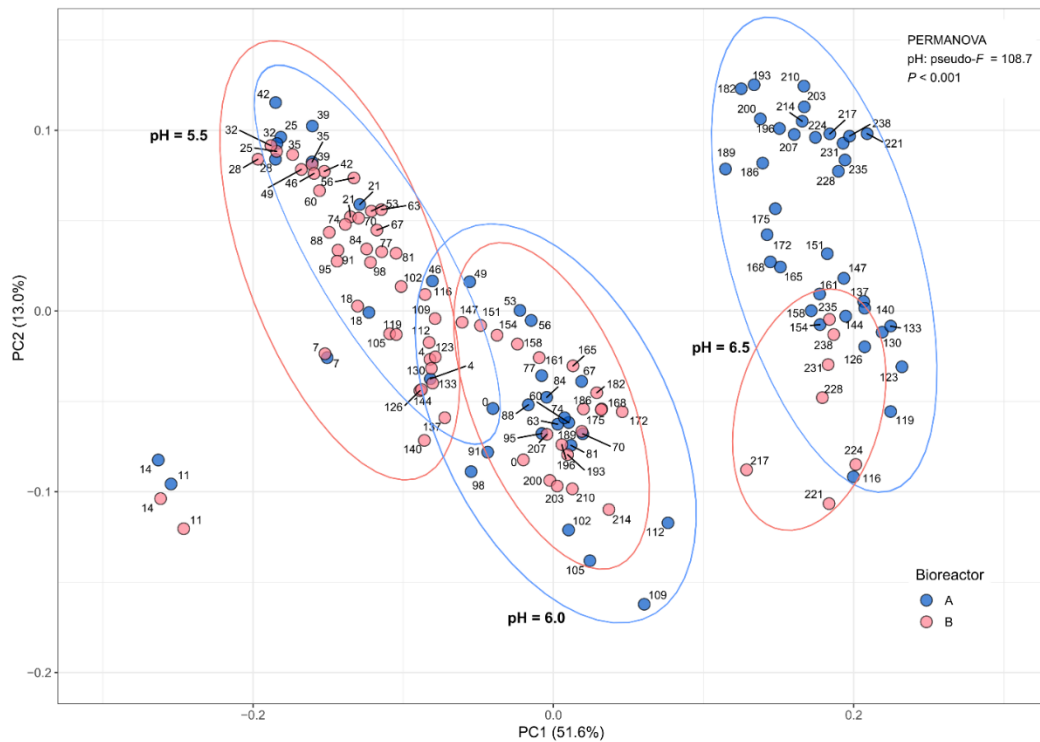
Figure S7. Dissimilarities in bacterial community composition (beta-diversity). Principal coordinates analysis (PCoA) based on Bray-Curtis dissimilarities of microbial community composition in bioreactors. Dots are named according to sampling days. Ellipses of 95% confidence intervals were added to each individual pH levels of the bioreactors.
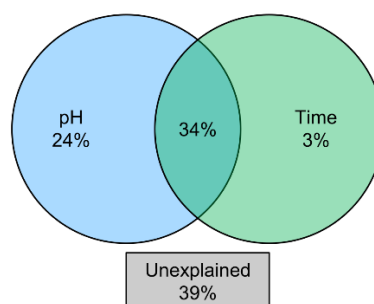
Figure S8. Variation partitioning analysis (VPA) showing the relative importance of pH and time on microbial community variations. VPA was used with redundancy analysis (RDA), and multiple partial RDAs were ran to determine the partial, linear effect of each explanatory matrix in the response data. Numbers represent adjusted coefficients of determination (Adj. R^2 values).

a,

	pH5.5	pH6.0	pH6.5	Overall Accuracy
pH5.5	1.0000	0.0000	0.0000	
pH6.0	0.0513	0.9487	0.0000	
pH6.5	0.0000	0.0238	0.9762	
Overall Accuracy				0.9779
Baseline Accuracy				0.4044
Accuracy Ratio				2.4182
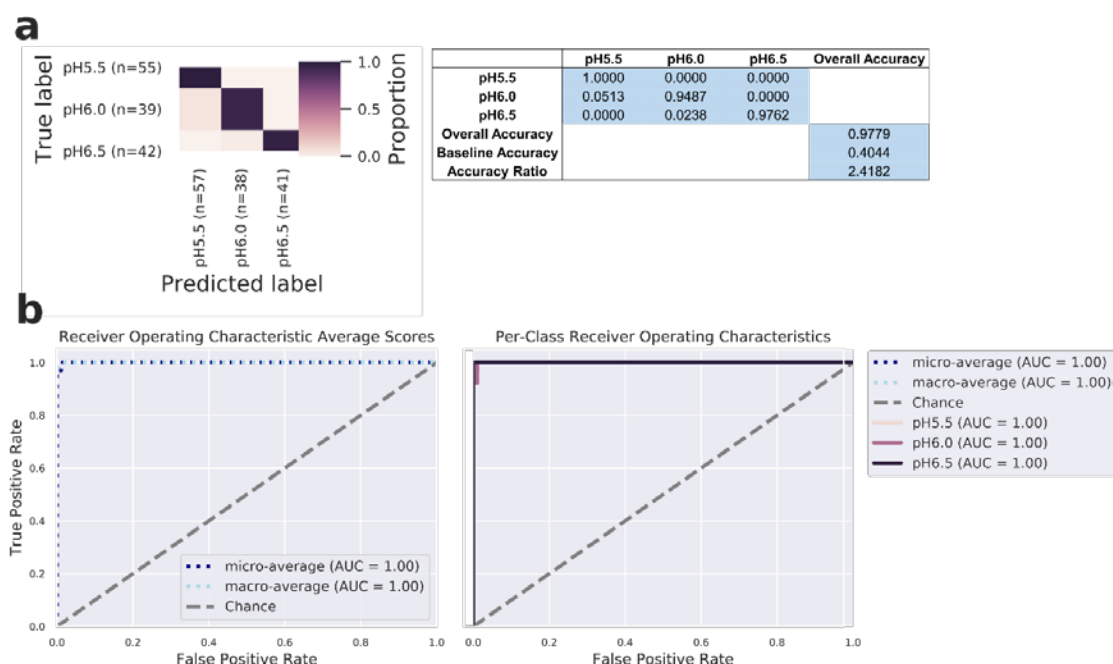
Figure S9. Nested cross-validation of random forest classification in the prediction of pH levels for each sample. a, Confusion matrix for the random forest classifier of all samples at three pH levels. For model optimisation, two layers of K-fold ($K = 5$) cross validation was incorporated to split the dataset into training and test set. Five different final models were trained, each sample received a predicted value and feature importance scores were averaged across each iteration. Overall accuracy was calculated by comparing the predicted values to the true values. **b,** The Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) curves represent the classification accuracy of the random forest. The ROC curve plots the relationship between the true positive rate and the false positive rate at various threshold settings. The AUC indicates the probability that the classifier ranks a randomly chosen sample of the given class higher than other classes. The random chance is represented as a diagonal line extending from the lower-left to the upper-right corner. In addition to show the ROC curves for each class, average ROCs and AUCs were calculated. "Micro-averaging" calculates metrics globally by averaging across each sample; hence class imbalance impacts this metric. "Macro-averaging" gives equal weight to the classification of each sample.
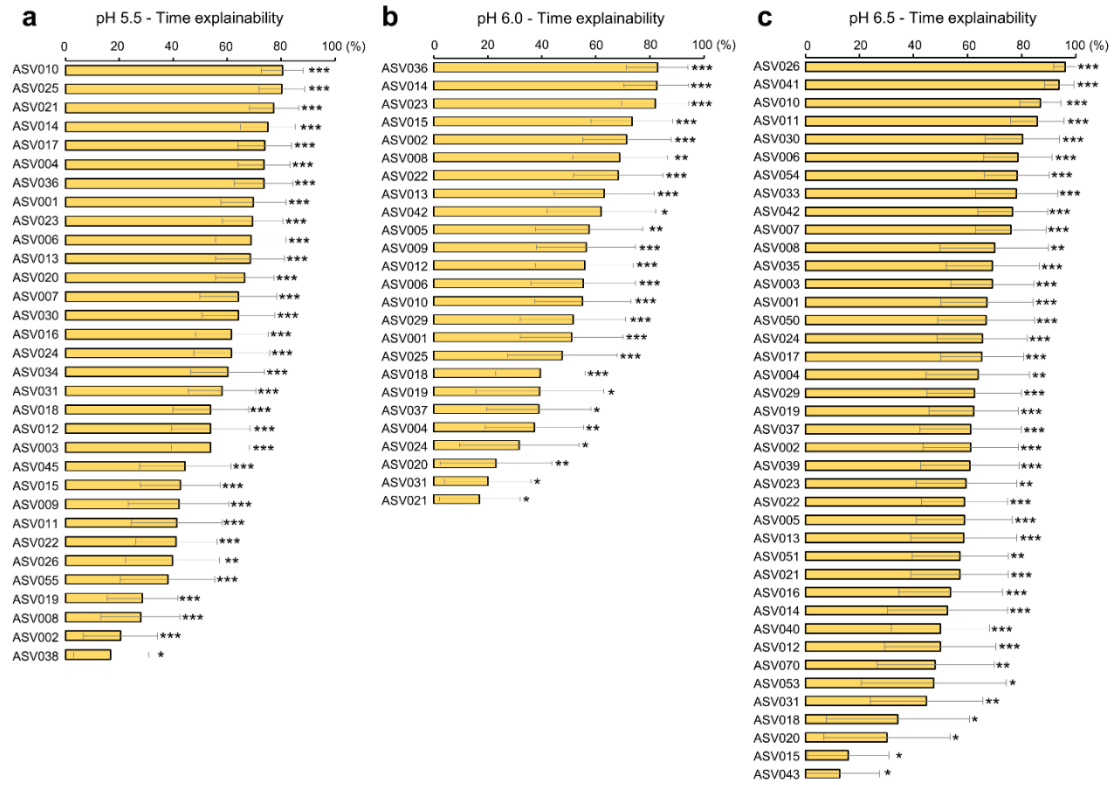
Figure S10. The core time-dependent taxa of individual pH levels. Using relative abundance data of ASVs of both bioreactors, a Microbial Temporal Variability Linear Mixed Model (MTV-LMM) was applied to identify time-dependent taxa of each individual pH level, whose abundance can be predicted based on the previous microbial community composition. As described, the time-explainability is denoted as the temporal variance explained by the microbial community in the previous time points. The time-explainability P-values: P*** < 0.001 < ** < 0.01 < * < 0.05.
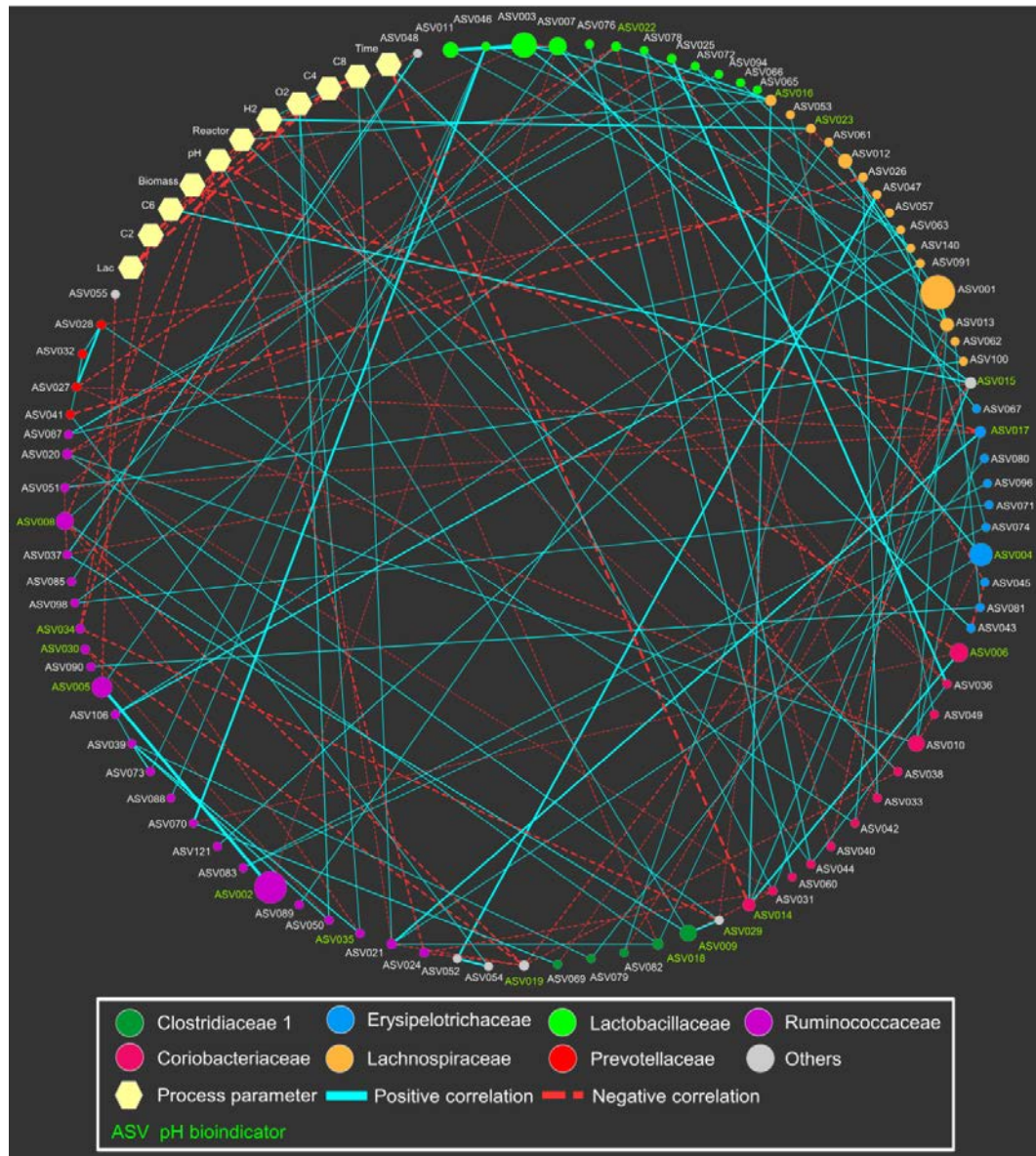
Figure S11. Co-occurrence network for the entire period of reactor operation. Edges indicate the significant (*P* < 0.05) correlations. Edge thickness reflects the strength of the correlation. Size of each ASV node is proportional to the mean relative abundance over the whole period. ASV nodes are coloured and grouped by family. "Others" include the ASVs belonging to families *Eubacteriaceae* (ASV015), *Actinomycetaceae* (ASV019), *Clostridiales Incertae Sedis* XI (ASV029), *Microbacteriaceae* (ASV048), *Veillonellaceae* (ASV052, ASV054) and *Nocardiaceae* (ASV055). pH bioindicators identified by random forest classification are shown with green letters. Lac, lactate concentration; C2, acetate yield; C4, *n*-butyrate yield; C6, *n*-caproate yield; C8, *n*-caprylate yield.

Table S1. Linear mixed-effects model results for diversity of order one (^{1}D). We consider time and pH as the fixed effects, and bioreactor as the random effect.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	50.883	7.718	6.593	< 0.001
Time	-0.209	0.021	-9.743	< 0.001
pH	-6.188	1.327	-4.663	< 0.001
Time:pH	0.035	0.004	9.777	< 0.001
Var. pH[a]	0.396	1.071		
Var. Bioreactor [T.B][b]	0.217			
Cov. (pH, bioreactor)[c]	0.953			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S2. Linear mixed-effects model results for evenness of order one (^{1}E). We consider time and pH as the fixed effects, and bioreactor as the random effect.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	0.808	0.128	6.322	< 0.001
Time	0.002	0.001	2.709	0.007
pH	-0.061	0.022	-2.720	0.007
Time:pH	< -0.001	< 0.001	-2.741	0.006
Var. pH[a]	< 0.001	0.002		
Var. Bioreactor [T.B][b]	< 0.001			
Cov. (pH, bioreactor)[c]	< 0.001			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S3. Linear mixed-effects model results for Richness. We consider time and pH as the fixed effects, and bioreactor as the random effect.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	85.190	12.733	6.690	< 0.001
Time	0.674	0.073	9.179	< 0.001
pH	9.045	2.227	4.062	< 0.001
Time:pH	0.116	0.012	9.344	< 0.001
Var. pH[a]	0.466	0.620		
Var. Bioreactor [T.B][b]	1.024			
Cov. (pH, bioreactor)[c]	13.738			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S4. Linear mixed-effects model results for the relative abundance of *Clostridium* IV sp. ASV008 at the different pH levels. We consider time and pH as the fixed effects, and bioreactor as the random effect.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	0.499	0.108	4.599	< 0.001
Time	0.002	0.001	3.119	0.002
pH	-0.077	0.019	-4.123	< 0.001
Time:pH	< -0.001	< 0.001	-2.864	0.004
Var. pH[a]	< 0.001	0.008		
Var. Bioreactor [T.B][b]	< -0.001			
Cov. (pH, bioreactor)[c]	< 0.001			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S5. Linear mixed-effects model results for the relative abundance of *Clostridium sensu stricto* sp. ASV009 at the different pH levels. We consider time and pH as the fixed effects, and bioreactor as the random effect.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	-0.081	0.220	-4.003	< 0.001
Time	0.001	< 0.001	1.691	0.091
pH	0.156	0.037	4.252	< 0.001
Time:pH	< -0.001	< 0.001	-1.804	0.071
Var. pH[a]	0.002	0.114		
Var. Bioreactor [T.B][b]	< -0.001			
Cov. (pH, bioreactor)[c]	< 0.001			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S6. Linear mixed-effects model results for microbial community composition that is represented by the PC1 from the Aitchison distance-based principal component analysis.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	-1.033	0.254	-4.059	< 0.001
Time	0.001	0.001	1.683	0.092
pH	0.169	0.043	3.909	< 0.001
Time:pH	< -0.001	< 0.001	-1.387	0.165
Var. pH[a]	0.002	0.070		
Var. Bioreactor [T.B][b]	-0.001			
Cov. (pH, bioreactor)[c]	< 0.001			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S7. Linear mixed-effects model results for microbial community composition that is represented by the PC1 from the Bray-Curtis distance-based principal coordinate analysis.

Variable or parameter	Coefficient	Standard error	Z-score	*P* value
(Intercept)	-2.216	0.235	-9.428	< 0.001
Time	0.004	0.001	3.161	0.002
pH	0.369	0.041	9.081	< 0.001
Time:pH	-0.001	< 0.001	-2.952	0.003
Var. pH[a]	< 0.001	0.014		
Var. Bioreactor [T.B][b]	< -0.001			
Cov. (pH, bioreactor)[c]	0.001			

[a]Variance of pH

[b]Variance of bioreactor [treatment of bioreactor B]

[c]Covariance of pH and bioreactor (random intercept)

Table S8. Summary statistics of networks.

Dataset	No. of Nodes	No. of Edges	avgN[a]	avgCC[b]	Density	Heterogeneity	Centralisation
Entire	100	151	3.256	0.062	0.038	0.469	0.057
pH 5.5	70	77	2.612	0.039	0.054	0.512	0.074
pH 6.0	60	63	2.455	0.029	0.057	0.447	0.062
pH 6.5	86	99	2.528	0.078	0.036	0.559	0.065

[a]avgN, average number of neighbours

[b]avgCC, average clustering coefficient

2.4 Draft genome sequences of three *Clostridia* isolates involved in lactate-based chain elongation

Bin Liu, Denny Popp, Heike Sträuber, Hauke Harms, Sabine Kleinsteuber #

Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

Correspondence to Sabine Kleinsteuber: sabine.kleinsteuber@ufz.de

AMERICAN SOCIETY FOR MICROBIOLOGY

Microbiology® Resource Announcements

Draft Genome Sequences of Three *Clostridia* Isolates Involved in Lactate-Based Chain Elongation

Bin Liu,[a] ⓘ Denny Popp,[a] Heike Sträuber,[a] Hauke Harms,[a] ⓘ Sabine Kleinsteuber[a]

[a]Department of Environmental Microbiology, Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany

ABSTRACT Hitherto, few species have been reported to convert lactate to *n*-caproate. Here, we report the high-quality draft genomes of three *Clostridia* strains isolated on lactate as the sole carbon source. The genomes were assembled using a hybrid short- and long-read sequencing approach. The genes involved in lactate-based chain elongation were identified.

Recently, we reported reactor microbiota that produce *n*-caproate from corn silage by anaerobic fermentation (1). To enrich the lactate-consuming bacteria involved in this process, anaerobic batch cultures in liquid mineral medium with lactate as the sole carbon source were inoculated with sieved reactor broth (mesh size, 2 mm; inoculation ratio, 1:10) from a lab-scale continuous stirred tank reactor (fermenting corn silage at 38°C; pH, 5.5; hydraulic retention time, 4 days) and incubated at 37°C. Pure strains from single colonies were isolated on agar medium DSM 104c with 5 g/liter lactate; their fermentation products were analyzed in liquid culture. Strains were identified by PCR and Sanger sequencing using 16S rRNA-specific primers (2). Three caproate-producing isolates designated BL-3, BL-4, and BL-6 represented new species based on their 16S rRNA gene sequences (3) and were selected for whole-genome sequencing (WGS).

Genomic DNA was extracted from the cell pellets using a NucleoSpin microbial DNA kit (Macherey-Nagel, Germany). WGS was performed with both long and short reads to obtain accurate sequences and complete scaffolds. Short-read sequencing using the Illumina NextSeq 500 system (NEBNext Ultra II FS DNA library prep kit; 2 × 150 bp) was performed by StarSEQ GmbH (Mainz, Germany). FASTQ data generation, demultiplexing, and adapter trimming of the raw sequencing reads were automatically performed by the Illumina software. The sequence quality was analyzed using FastQC v0.11.9 (4). For long-read sequencing, the library was prepared using the ligation sequencing kit (1D SQK-LSK109) and the native barcoding kit (1D EXP-NBD104) on an R9.4 SpotON flow cell with a MinION Mk1B device from Oxford Nanopore Technologies (ONT; UK). MinION was controlled with MinKNOW v3.1.19 (ONT). Base calling with the high-accuracy model and demultiplexing were accomplished by Guppy v3.1.5 (ONT) using default parameters. Porechop v0.2.3 (5) was used to trim adapters, applying default parameters, with additional internal adapter removal using a 90% identity threshold. Long-read sequencing of BL-3, BL-4, and BL-6 produced 465,840 reads (1.1 Gb of data, 270× coverage, N_{50} value of 9 kb), 186,991 reads (1.2 Gb, 500× coverage, N_{50} value of 11 kb), and 75,620 reads (0.5 Gb, 145× coverage, N_{50} value of 12 kb), respectively. Short-read sequencing of BL-3, BL-4, and BL-6 generated 2,272,799 reads (6.9 Gb, 180× coverage), 2,172,274 reads (6.6 Gb, 284× coverage), and 1,574,086 reads (4.8 Gb, 140× coverage), respectively.

Hybrid *de novo* genome assembly based on short and long reads was performed using Unicycler v0.4.8 with default parameters (6). The genome assembly of strain BL-3 resulted in seven contigs. For strains BL-4 and BL-6, single circular contigs were assembled. Putative gene coding sequences (CDSs) were identified and annotated

Address correspondence to Sabine Kleinsteuber, sabine.kleinsteuber@ufz.de.

149

TABLE 1 Genome features of isolates BL-3, BL-4, and BL-6

Feature	Data for strain:		
	BL-3	BL-4	BL-6
Genus assignment[a]	*Clostridium_B*	*UBA4871*	*Clostridium_E*
BioSample accession no.	SAMEA6567123	SAMEA6567124	SAMEA6567125
Run accession no.	ERR3959273, ERR3959415	ERR3959274, ERR3959416	ERR3959275, ERR3959417
WGS or chromosome accession no.	CADDXC010000000	LR778134	LR778135
Genome size (bp)	3,855,691	2,335,857	3,435,529
GC content (%)	34.32	42.75	54.63
No. of contigs	7	1	1
Completeness[b] (%)	98.6	97.9	98.0
Contamination[b] (%)	1.0	0.3	1.3
No. of CDSs	3,867	2,319	3,480
No. of tRNA genes	67 (21 types)	54 (20 types)	63 (21 types)
No. of rRNA genes (5S, 16S, 23S)	14 (4, 5, 5)	9 (3, 3, 3)	9 (3, 3, 3)
No. of miscellaneous RNA genes	524	24	32
No. of tmRNA[c] genes	1	1	1

[a] Genus assignment refers to the Genome Taxonomy Database (8) phylogenomic classification.
[b] Genome completeness and contamination as calculated by CheckM (9).
[c] tmRNA, transfer-messenger RNA.

using the MicroScope automatic annotation platform via external submission (7). The genome sizes, GC contents, numbers of predicted CDSs, and genome quality parameters are listed in Table 1. The new isolates and their genomes are valuable resources for exploring the metabolic features of chain-elongating bacteria.

Data availability. The sequence data are available in the European Nucleotide Archive (ENA) database under accession number PRJEB36835; see Table 1 for the BioSample, WGS or chromosome, and run accession numbers.

ACKNOWLEDGMENTS

REFERENCES

1. Lambrecht J, Cichocki N, Schattenberg F, Kleinsteuber S, Harms H, Müller S, Sträuber H. 2019. Key sub-community dynamics of medium-chain carboxylate production. Microb Cell Fact 18:92. https://doi.org/10.1186/s12934-019-1143-8.
2. Lane DJ. 1991. 16S/23S rRNA sequencing, p 177–203. *In* Stackbrandt E, Goodfellow M (ed), Nucleic acid techniques in bacterial systematics, John Wiley and Sons, Chichester, United Kingdom.
3. Stackebrandt E, Ebers J. 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 8:152–155.
4. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
5. Wick R. 2017. Porechop. http://github.com/rrwick/Porechop.
6. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
7. Vallenet D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, Burlot L, Bussell X, Fouteau S, Gautreau G, Lajus A, Langlois J, Planel R, Roche D, Rollin J, Rouy Z, Sabatet V, Médigue C. 2020. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. Nucleic Acids Res 48:D579–D589. https://doi.org/10.1093/nar/gkz926.
8. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.
9. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

2.5 Three novel *Clostridia* isolates produce *n*-caproate and *iso*-butyrate from lactate: comparative genomics of chain-elongating bacteria

Bin Liu [1], Denny Popp [1], Nicolai Müller [2], Heike Sträuber [1], Hauke Harms [1] and Sabine Kleinsteuber [1,*]

[1] Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, 04318 Leipzig, Germany; liu.bin@ufz.de (B.L.); denny.popp@ufz.de (D.P.); heike.straeuber@ufz.de (H.S.); hauke.harms@ufz.de (H.H.); sabine.kleinsteuber@ufz.de (S.K.)

[2] Department of Biology, University of Konstanz, D-78457 Konstanz, Germany; nicolai.mueller@uni-konstanz.de (N.M.)

* Correspondence: sabine.kleinsteuber@ufz.de; Tel: +49-341-235-1325
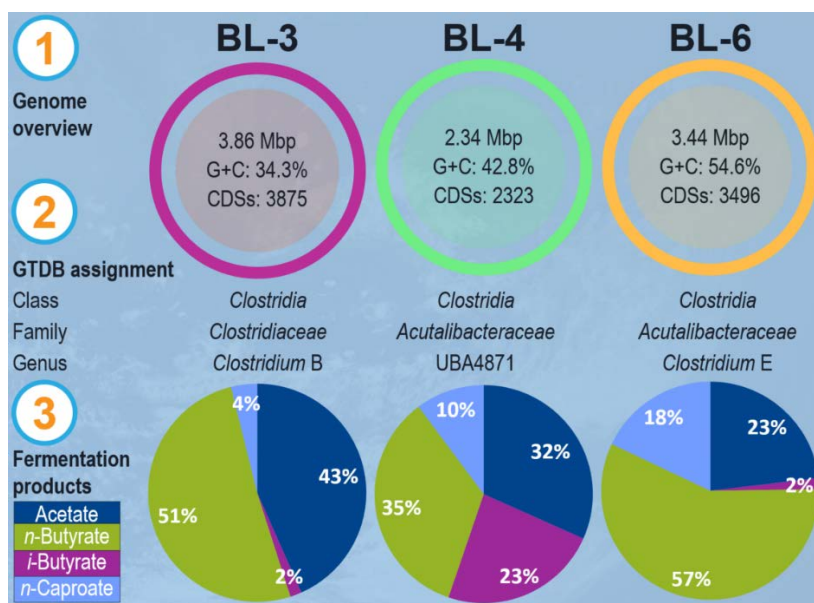
2.5.1 Main text

Abstract: 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. Few lactate-consuming chain-elongating species have been isolated and knowledge on their shared genetic features is still limited. Recently we isolated three novel clostridial strains (BL-3, BL-4 and BL-6) that convert lactate to *n*-caproate and *iso*-butyrate. Here, we analysed the genetic background of lactate-based chain elongation in these isolates and other chain-elongating species by comparative genomics. The three strains produced *n*-caproate, *n*-butyrate, *iso*-butyrate, and acetate from lactate, with the highest proportions of *n*-caproate (18%) for BL-6 and of *iso*-butyrate (23%) for BL-4 in batch cultivation at pH 5.5. The three genomes show low conservation of organisation and a relatively small core-genome size. They contain highly conserved genes involved in lactate oxidation, reverse β-oxidation, hydrogen formation and either of two types of energy conservation systems (Rnf and Ech). Including genomes of another eleven experimentally validated chain-elongating strains, we found that the chain elongation-specific core-genome encodes the pathways for reverse β-oxidation, hydrogen formation and energy conservation, while displaying substantial genome heterogeneity. Metabolic features of these isolates may be interesting for biotechnological applications in *n*-caproate and *iso*-butyrate production.

Keywords: Novel clostridial species; Carboxylate platform; Medium-chain carboxylates; Branched-chain carboxylates; Anaerobic fermentation; Reverse β-oxidation

1. Introduction

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 as 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. Some strictly 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. highlighted 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 utilise 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 product spectrum of the three new lactate-consuming strains and to give insights into their metabolism based on their genomes. Batch experiments were conducted to explore the fermentation profiles with lactate. Functional genome annotation and phylogenomic analysis aimed at elucidating the genetic background of *n*-caproate and *iso*-butyrate production and the genetic heterogeneity between the three strains. To analyse 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 performed a comparative genome analysis by including eleven more genomes of experimentally validated chain-elongating species.

2. Materials and Methods

2.1 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 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 **Supplementary Methods**). 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.

2.2 Lactate utilisation 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 and 0.05% yeast extract as described above. The bottles were inoculated with 5 mL seed cultures (optical density at 600 nm [OD_{600}] ~ 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 the pH was adjusted to 5.5 accordingly. All batch tests were carried out in duplicate. For further investigation of the growth of isolate BL-4, anoxic, bicarbonate buffered

freshwater medium at pH 7.3 reduced with cysteine was used. The basal medium consisted of NaCl (1 g/L), $MgCl_2$ (0.4 g/L), KH_2PO_4 (0.2 g/L), NH_4Cl (0.25 g/L), KCl (0.5 g/L), $CaCl_2$ (0.15 g/L) and $Na_2SO_4 \times 10 \, H_2O$ (0.16 g/L) and was autoclaved for at least 30 min at 121°C and 1 bar overpressure in a Widdel-flask. After cooling to room temperature under a stream of N_2/CO_2 (80:20), a separately autoclaved solution of $NaHCO_3$ was added to a final concentration of 30 mM. Then each 1 mL of trace element solution SL13, 7-Vitamin solution and selenite-tungstate solution were added per liter medium (modified after [20-23]). Finally, the medium was amended with 0.4 mg/L resazurin as a redox indicator and filter-sterilized cysteine-HCl (3 mM final concentration) as reducing agent. In case the redox indicator of the medium did not turn colorless within 30 min of stirring under N_2/CO_2, 25 µM to 50 µM titanium(III)-nitrilotriacetic acid was added from a filter-sterilized stock solution to aid in establishing reduced conditions. This was the case for all pH 7.3-media used in this study. After the medium turned colorless, the pH of the medium was adjusted to pH 7.3 and the medium was thereafter dispensed into the cultivation vessels under N_2/CO_2. Where indicated, 0.05% yeast extract was added as an additional source of vitamins and amino acids. Strain BL-4 was cultivated in 25-mL tubes closed with rubber stoppers and filled with 10 mL medium at 37°C. The OD_{600} was monitored over time with a Camspec tube photometer as described before [24].

2.3 Analytical techniques

Liquid samples of the batch 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 (HPLC; 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 [25]. For further investigation of the growth of BL-4, HPLC samples were withdrawn with syringes and needles, acidified with 20 µL of 1 M H_2SO_4, centrifuged to remove cells, and the supernatant analysed by refractive index

detection after separation on a Rezex RHM monosaccharide column with 30 mM sulfuric acid at 40°C as described [24].

2.4 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 coding sequences (CDSs) was accomplished by the MicroScope automatic annotation pipeline [26]. Automatic annotations of selected CDSs were manually curated by comparing the protein sequences with the PkGDB, Swiss-Prot, TrEMBL, COG (Clusters of Orthologous Groups), EGGNOG. (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups), FIGfams and InterPro databases [26–31] 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 [32] 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 [29]. All these databases and tools are integrated in the MicroScope platform as described by Vallenet et al. [26]. 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* BS-1, *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.

2.5 Phylogenetic analysis and taxonomic classification

Phylogenetic analysis of 16S rRNA gene sequences was performed on the Phylogeny.fr platform [33]. According to the Nucleotide BLAST (Basic Local Alignment Search Tool) comparison result against the rRNA/ITS databases

(16S ribosomal RNA sequences (Bacteria and Archaea)) of NCBI (National Center for Biotechnology Information) [34], 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 [35]. After alignment, Gblocks v0.91b was used to remove ambiguous regions (i.e. containing gaps and/or poorly aligned) as described by Castresana [36]. The phylogenetic tree was reconstructed using the maximum likelihood method contained in PhyML v3.1 [37,38]. Robustness of tree topology was assessed by 100 bootstrap replicates. Finally, the tree was visualised by using TreeDyn v198.3 [39]. Besides the taxonomic classification of the genomes in MicroScope, GTDB-Tk v1.0.2 was used for taxonomic assignment to GTDB (Genome Taxonomy Database) [40] and the corresponding NCBI taxonomy.

A phylogenomic tree of strains BL-3, BL-4, BL-6 and other chain-elongating bacteria was calculated based on genomic similarity. The genomic similarity was estimated using Mash [41], 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 with clustering annotations was constructed. 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 [42]. The ANI (OrthoANIu value) comparison of the genomes of the isolates to related genomes was calculated by an ANI calculator that improved the original OrthoANI (Average Nucleotide Identity by Orthology) algorithm by applying USEARCH instead of BLAST as described by Yoon et al. [43].

Default settings were used for all tools unless otherwise specified.

2.6 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) [44] were considered for these analyses.

2.7 Data availability

All data generated or analysed during this study are included in this published article and its additional files. The full-length 16S rRNA gene sequences of the three isolates have been deposited in the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/browser/home) under BioProject PRJEB39379, with the accession numbers LR861112, LR861113, and LR861114. The genome data of the three isolates have been deposited in ENA under BioProject PRJEB36835, with Whole Genome Sequencing or Chromosome accession numbers CADDXC010000000, LR778134, and LR778135.

3. Results and Discussion

3.1 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 (**Figure 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 (Erko and Ebers, 2006), these three strains can be assumed to represent novel species and were consequently selected for whole genome sequencing.

3.2 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 (**Figure 1**). Started at an initial pH 5.5, the three strains displayed 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% to 43%) and *n*-butyrate (35% to 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 2% for BL-6). As shown in **Figure 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 might convert *n*-butyrate to *iso*-butyrate.

3.3 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 genomes of these strains with CheckM completeness of 98.6%, 97.9% and 98% and contamination of 1.0%, 0.3% and 1.3% for BL-3, BL-4 and BL-6, respectively [19]. The genome sizes are depicted in **Figure 2** and detailed in **Table 1**. According to the taxonomic classification of GTDB, BL-3 was assigned to the genus *Clostridium*_B (*Clostridiaceae*), whereas BL-4 and BL-6 were assigned to the genera UBA4871 and *Clostridium*_E, respectively, both belonging to the *Acutalibacteraceae* (*Ruminococcaceae* according to the NCBI taxonomy). The number of predicted gene 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 functional categories (76% for BL-3, 75% for BL-4 and 73% for BL-6; see details in **Table S1**) and EGGNOG categories (86% for BL-3, 85% for BL-4 and 83% for BL-6; see

details in **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 (**Figure 2**). As for the 2,064 genes conserved in the core-genome, proportions of 27.2%, 20.9% and 19.1% can be considered core CDSs of strains BL-3, BL-4 and BL-6, respectively. The core CDSs include 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 **Supplemental file 2**). According to the pairwise comparison of the three genomes, a few synteny groups on nucleotide level are shared (**Figure S2**), which indicates the low conservation of genome organisation and underlines the genomic heterogeneity of the three isolates.

3.4 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 [46]. 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% to 55% (**Table 1**).

We constructed a phylogenomic tree to understand the evolutionary relationships between our isolates and other chain-elongating species (**Figure 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 (**Figure 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 (**Figure 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 revealed a total of 20,790 homologous gene families with 40,582 genes identified (**Figure 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% to 15% for each strain (**Figure 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 CDSs). The above-mentioned patterns also apply to the comparison of the three isolates as shown in **Figure 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 **Supplemental file 3**) [60]. The chain elongation-specific core-genome also comprises genes involved in reverse β-oxidation, hydrogen formation and energy conservation (**Table 2** and

details in **Supplemental file 4**). These genes are highly conserved in all 14 strains and can be considered hallmarks of chain-elongating bacteria.

3.5 Genetic basis of lactate conversion to n-caproate and iso-butyrate

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.

3.5.1 Lactate oxidation to acetyl-CoA

Lactate can serve as carbon and energy source for chain-elongating bacteria. As shown in **Figure 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) (**Figures 6a and 6c**). The gene cluster encoding LacP and LacR was also found in all other lactate-based chain elongators (**Figures 6d-6h**). 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 that 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 (**Figure 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 (**Figure 6b**, CDS labels: 2199-2205) encoding the membrane-associated energy-converting NADH:ferredoxin oxidoreductase (RnfABCDEG). The BL-6 genome has three LDH genes with one found in a cluster (**Figure 6c**, CDS labels: 3216-3223) including genes for butyryl-CoA dehydrogenase (BCD), EtfAB, LacR and

LacP. A similar gene cluster **(Figure 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 (**Figure 6d-6h**).

3.5.2 Ethanol oxidation to acetyl-CoA

The ethanol-based chain elongation pathway is well elucidated in *C. kluyveri* [14] and of particular biotechnological importance as shown in several studies [61–63]. Genome data of BL-3 and BL-6 suggest that these strains are capable of utilizing 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 (**Figure 5**). The corresponding CDSs were found in the genomes of BL-3 and BL-6, but not in the BL-4 genome.

3.5.3 *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) (**Figure 5**). The formation of *n*-butyrate further requires butyryl-CoA:acetate CoA transferase (CoAT) to catalyse the reaction of butyryl-CoA and acetate to yield acetyl-CoA and the corresponding fatty acid. Transformation of butyryl-CoA to caproyl-CoA may happen 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 [64,65]. All three genomes contain the genes encoding ACAT, HAD, ECH, BCD, EtfAB and CoAT (**Supplemental 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 (**Figure 6a**, CDS labels: 13110-13113) as well as one cluster encoding ECH, BCD, EtfAB and HAD (**Figure 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 (**Figure 6b**, CDS labels: 1867-1873). Two similar clusters were found in the genomes of *Eubacterium limosum* (**Figure 6k**, CDS labels: 21760-21785) and *Eubacterium pyruvativorans* (**Figure 6i**, CDS labels: 280031-280037). Another set of HAD, ACAT, ECH and CoAT genes clusters together with the genes for acetyl-CoA:oxalate CoA-transferase (ACOCT) and (R)-2-hydroxyisocaproyl-CoA dehydratase (HadABC) (**Figure 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 (**Figure 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 (**Figure 6c**, CDS labels: 3216-3223) as mentioned above. We found that the genes encoding BCD are located in close vicinity to the genes of EtfAB in the genomes of all three isolates (**Figure 6a-6c**), which is commonly conserved as a key feature among all genomes of other chain-elongating bacteria (**Figure 6d-6n**).

Besides CoAT, the acyl-CoA thioesterase (ACT) may also catalyse the formation of *n*-butyrate and *n*-caproate from the terminal acyl-CoA (**Figure 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 **Supplemental 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 **Supplemental file 4**, sheets BL-3). Genomes of BL-4 and BL-6 both contain

the CDSs for ACT (see CDS labels in **Supplemental file 4**, sheets BL-4 and BL-6), but presenting a low identity (≤ 40%) to proteins in the databases (see alignment details in **Supplemental files 6-7**). Further experiments are required to assess the functionality of these CDSs and if the predicted enzymes play a role as terminal enzymes in reverse β-oxidation. A recent study on lactate-based chain elongation in mixed cultures using guild-based metabolic models suggested that butyrate is formed by CoAT, whereas caproate and caprylate are formed by ACT [66]. As pointed out by the authors, this might depend on the organisms, and the affinities of CoAT and ACT enzymes for different chain lengths need to be assessed.

Besides CoAT and ACT, a third pathway potentially contributing to *n*-butyrate formation from *n*-butyryl-CoA was identified in the genome of BL-3. As illustrated in **Figure 5**, a phosphate butyryltransferase (PTB) forms butyryl phosphate that is further converted to butyrate by a butyrate kinase (BUK). The latter step leads to the formation of one ATP, in contrast to the CoAT route, which conserves energy in the form of acetyl-CoA. The PTB/BUK route might favor butyrate production at the cost of caproate yield, i.e. butyrate is not further elongated due to acetyl-CoA shortage and possibly due to higher growth rates. In our previous study on a mixed culture growing on xylan and lactate under constant conditions [18], co-occurrence network analysis predicted a *Clostridium sensu stricto* (closely related to *C. luticellarii*) as key butyrate producer that outcompeted caproate producers as reflected by higher microbial biomass production and a drop in caproate and caprylate concentrations. The lack of BUK genes in the genomes of strains BL-4 and BL-6 is consistent with the previously reported progressive loss of BUK genes found in some clostridial lineages [67]. From the biotechnological perspective, strains BL-4 and BL-6 seem to be more beneficial than BL-3 as they yield more caproate and less acetate compared with strain BL-3. However, detailed experiments are required to characterise the kinetics of lactate conversion and product formation in the strains under different growth conditions and in pure and mixed culture settings.

3.5.4 *iso*-Butyrate formation

The formation of *iso*-butyrate as a product of lactate-based chain elongation was experimentally proven in all three isolates. The genome analysis revealed hints on the assumed pathway, i.e. reversible *n*-butyrate/*iso*-butyrate isomerization [68,69]. As described by Matthies and Schink [69], the conversion of *n*-butyrate to *iso*-butyrate first requires activation to *n*-butyryl-CoA. Next, the isomerisation of *n*-butyryl-CoA via *iso*-butyryl-CoA to *iso*-butyrate is catalysed by a butyryl-CoA:isobutyryl-CoA mutase (BM) and an isobutyryl-CoA:acetate CoA transferase (CoAT) as shown in **Figure 5**. At the first glance, none of the three genomes seems to encode a BM, but we found a BM homologue in the genome of BL-3 that might have been misannotated as methylmalonyl-CoA mutase. As reported by Cracan et al. [70], the fusion protein IcmF (isobutyryl-CoA mutase fused) composed of the small subunit of BM, a GTPase domain and the large subunit of BM has been widely misannotated as methylmalonyl-CoA mutase in other bacterial genomes. CDSs for a putative IcmF were found in the genomes of BL-3 and of the *iso*-butyrate producer *C. luticellarii* (see the CDS labels in **Supplemental file 4**). A CoA transferase gene located next to these CDSs may confirm the prediction function in isomerisation. BMs catalyse the rearrangement of carboxyl groups as migration to the adjacent carbon atom, in which enzyme activities depend on coenzyme B_{12} [71]. One possible reason for the conversion of *n*-butyrate to *iso*-butyrate is that bacteria can maintain the pool of *iso*-butyrate for synthesising valine during growth in amino acid-deficient medium [72]. As this isomerisation 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 CE process [3,12], the formation of *iso*-butyrate may facilitate bacteria to further obtain energy from chain elongation.

The genomes of BL-4 and BL-6 lack CDSs for BM, but the formation of *iso*-butyrate from lactate is also conceivable via methylmalonyl-CoA and methylmalonate-semialdehyde, representing a reverse process of anaerobic *iso*-butyrate degradation by *Desulfococcus multivorans* [73]. At first sight, not all candidate genes predicted for this hypothetical pathway were found in strain BL-4 (**Figure S4**,) and other reported *iso*-butyrate-producing CE species

(**Supplemental file 4**), thus physiological experiments are needed to elucidate the mechanism of *iso*-butyrate formation in CE strains. In order to find indications of the presence of the anticipated methylmalonyl-CoA pathway, strain BL-4 was cultivated with 50 mM sodium succinate (**Figure 7**). The culture reached an OD_{600} of around 0.2 while concomitantly consuming 39 mM succinate and producing propionate (37 mM) and minor amounts of acetate (4.2 mM), formate (0.3 mM), *iso*-butyrate (0.2 mM), butyrate (0.1 mM) and 1-propanol (0.8 mM). Therefore, succinate was decarboxylated to propionate in an almost 1:1 stoichiometric ratio. The latter reaction, to our knowledge, is only catalysed with the enzymes of propionic acid fermentation, i.e. via methylmalonyl-CoA as an intermediate. This indicates that BL-4 has the enzymes necessary for the conversion of organic acids to propionyl-CoA and could theoretically produce *iso*-butyrate through a reversal of the *iso*-butyrate degradation pathway in *Desulfococcus multivorans* [73].

We hypothesise that pyruvate derived from lactate oxidation is carboxylated to oxaloacetate with concomitant decarboxylation of methylmalonyl-CoA to propionyl-CoA by a transcarboxylase. The genes for a transcarboxylase could not be identified at first sight. However, a BLAST-search of the amino acid sequence of the genes of the respective enzyme complex in *Propionibacterium freudenreichii* DSM 20271 against the genome of BL-4 revealed three potential homologs. The three major methylmalonyl-CoA carboxyltransferase subunits of *P. freudenreichii* DSM 20271 12S, 5S and 1.3S (IMG-locus tags Ga0077868_111809, Ga0077868_111810 and Ga0077868_111807) are similar to a carboxyltransferase (CLOSBL4_v1_1895, 33% identities), an oxaloacetate decarboxylase (CLOSBL4_v1_1897, 52% identities) and a glutaconyl-CoA decarboxylase subunit gamma (CLOSBL4_v1_1896, 39% identities) respectively, and similarly arranged in one gene cluster. These genes therefore possibly constitute a methylmalonyl-CoA transcarboxylase. Yet, a gene candidate for a methylmalonyl-CoA mutase could not be identified. As a consequence of the ability to decarboxylate succinate to propionate, strain BL-4 might also be able to convert lactate to propionyl-CoA, which in turn could be carboxylated to methylmalonate-semialdehyde (MMS). MMS could then be reduced to 3-hydroxy-isobutyrate (3-HIB), which then might be activated to 3-

hydroxyisobutyryl-CoA (3-HIB-CoA) by a CoA-transferase. The pathway could proceed with the dehydration of 3-HIB-CoA to 3-enoyl-isobutyryl-CoA (a.k.a. methylacrylyl-CoA) and reduction of the latter to isobutyryl-CoA. Finally, *iso*-butyrate could be produced either by another CoA-transferase or by phosphorylation and dephosphorylation by a phosphotransferase and an *iso*-butyrate kinase. The genes responsible for the conversion of propionyl-CoA to *iso*-butyrate could not be completely identified in the genome of strain BL-4. However, inferring from the fact that valine is degraded to acetate and *iso*-butyrate, strain BL-4 should at least have the biochemical machinery for the conversion of *iso*-butyrate to 3-hydroxyisobutyrate and methylmalonyl-CoA and vice versa (**Figure 8**) [74]. Otherwise, the production of acetate from valine cannot be easily explained. Acetate was always produced in media with 0.05% yeast extract (4.2 mM acetate during growth with succinate, **Figure 7**) and could therefore result from the degradation of other organic compounds in yeast extract. However, acetate concentrations in valine-grown cultures were twice as high (9 mM, **Figure 8b**). Possibly, valine could also be co-fermented in a Stickland-reaction, i.e. fermentation of pairs of amino acids such as valine and glycine, yet this would also lead to accumulation of amounts of *iso*-butyrate in a 2:1 acetate to *iso*-butyrate ratio, which was not the case (15 mM *iso*-butyrate produced, **Figure 8b**). It is hence questionable where the reducing equivalents derived from valine oxidation to *iso*-butyrate ended up and possibly, these reducing equivalents were used to generate the various other side products present in the valine-grown cultures (**Figure 8b**). Alternatively, pyruvate, and subsequently acetate, could be produced by the enzymes of the valine biosynthesis pathway acting in reverse, i.e. acetohydroxy-acid synthase (ilvB, CLOSBL4_v1_0646), acetolactate synthase (ilvH, CLOSBL4_v1_0647) and acetohydroxy-acid isomeroreductase (ilvC, CLOSBL4_v1_0648). Yet, it is doubtful whether the thermodynamic equilibrium allows for such a reversal of these enzyme reactions as the latter pathway usually favours valine production and as at least the reaction of acetohydroxy-acid synthase is irreversible [75].

A comprehensive metabolic pathway of lactate conversion to *iso*-butyrate is not available to date for strain BL-4 and the former might be a combined variation of the known pathways of propionic acid fermentation and branched chain

amino acid degradation. It appears that *iso*-butyrate is only formed in large amounts, when butyrate accumulation levels out and might also depend on the pH of the culture (**Figure 1**). Moreover, the amount of *iso*-butyrate formed is too high to be explained by degradation of branched-chain amino acids alone. The proposed methylmalonyl-CoA pathway could be a plausible explanation for *iso*-butyrate production from lactate, yet it remains enigmatic why strain BL-4 does not convert lactate into propionate as end-product by classical propionic acid fermentation instead of *iso*-butyrate, i.e. the question remains what are the advantages of proceeding degradation to the level of *iso*-butyrate.

3.5.5 Energy conservation and hydrogen formation

As shown in **Figure 5**, the cytoplasmic BCD/EtfAB complex catalyses 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 [76]. 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 [77]. 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 **Figures 6a and 6b**. This gene organization (shown as *rnfBAEGDC* in the other DNA strand) was also found in other genomes of chain-elongating bacteria (**Figures 6d-n**). For BL-6, we could only find four genes for subunits of the Rnf complex during the functional annotation (see CDS labels in the **Supplemental 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 [78,79]. The Ech uses reduced ferredoxin as electron donor and reduces protons, not NAD$^+$ like Rnf. As shown in **Figure 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 (**Figure 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 **Supplemental 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 (**Figure 6a**, CDS labels: 11486-11488; **Figure 6c**, CDS labels: 3217-3220; **Figure 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^+ [77,80].

The manually curated annotation of all above-mentioned CDSs in the genomes of other lactate-based chain-elongating strains is provided in **Supplemental file 8.**

4. Conclusions

Our results suggest three novel *Clostridia* species, represented by the strains 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 pathways for *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, the genomic and metabolic features of the three novel chain-elongating isolates might be interesting for further research and biotechnological applications with regard to *n*-caproate and *iso*-butyrate production.

Supplementary Materials: Supplemental file 1: Supplementary Methods, Tables S1. COG (Clusters of Orthologous Groups) classification, Table S2. EGGNOG (Evolutionary Genealogy of Gene: Non-supervised Orthologous Groups) classification, Figure S1. Fermentation products of the enrichment culture (a single bottle of the fourth transfer) during growth on lactate, Figure S2. Pairwise comparison of the conservation of the synteny groups in the three new isolates, Figure S3. Maximum likelihood tree of the three new strains and closest relatives based on 16S rRNA gene sequences. Figure S4. Hypothetical *iso*-butyrate-producing lactate degradation pathway with the independence of *iso*-butyrate-CoA-mutase. Supplemental file 2: Data file S1, Results of pan-genome analysis of the three isolates. Supplemental file 3: Data file S2, Results of pan-genome analysis of all chain-elongating strains. Supplemental file 4: Data file S3, Annotation summary of all chain-elongating strains. Supplemental file 5: Data file S4, Functional annotations of strain BL-3. Supplemental file 6: Data file S5, Functional annotations of strain BL-4. Supplemental file 7: Data file S6, Functional annotations of strain BL-6. Supplemental file 8: Data file S7, Functional annotations of other lactate-based chain-elongating strains.

Data files S1-S8 can be found at: https://github.com/binliu21/carboxylate-platform/tree/SI-chapter-2.5

Author Contributions: Conceptualization, B.L., H.S. and S.K.; methodology, B.L., D.P., N.M., H.S. and S.K.; formal analysis, B.L. and D.P.; investigation, B.L.; data curation, B.L., D.P., H.S., N.M. and S.K.; writing—original draft

preparation, B.L.; writing—review and editing, D.P., H.S., N.M., H.H. and S.K.; visualisation, B.L.; supervision, H.S., H.H. and S.K.; project administration, S.K.; All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Angenent, L.T.; Richter, H.; Buckel, W.; Spirito, C.M.; Steinbusch, K.J.J.; Plugge, C.M.; Strik, D.P.B.T.B.; Grootscholten, T.I.M.; Buisman, C.J.N.; Hamelers, H.V.M. Chain elongation with reactor microbiomes: Open-culture biotechnology to produce biochemicals. *Environ. Sci. Technol.* **2016**, *50*, 2796–2810, doi:10.1021/acs.est.5b04847.

2. Zhang, K.; Woodruff, A.P.; Xiong, M.; Zhou, J.; Dhande, Y.K. A synthetic metabolic pathway for production of the platform chemical isobutyric acid. *ChemSusChem.* **2011**, *4*, 1068–1070, doi:10.1002/cssc.201100045.

3. Leeuw, K.D. De; Smit, S.M. De; Oossanen, S. Van; Moerland, M.J.; Buisman, C.J.N.; Strik, D.P.B.T.B. Methanol-based chain elongation with acetate to *n*-butyrate and isobutyrate at varying selectivities dependent on pH. *ACS Sustain. Chem. Eng* **2020**, *8*, 8184–8194, doi:10.1021/acssuschemeng.0c00907.

4. Scalschi, L.; Vicedo, B.; Camañes, G.; Fernandez-Crespo, E.; Lapeña, L.; González-Bosch, C.; García-Agustín, P. 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*, 342–355, doi:10.1111/mpp.12010.

5. Urban, C.; Xu, J.; Sträuber, H.; dos Santos Dantas, T.R.; Mühlenberg, J.; Härtig, C.; Angenet, L.T.; Harnisch, F. Production of drop-in fuel from biomass by combined microbial and electrochemical conversions. *Energy Environ. Sci.* **2017**, *10*, 2231–2244, doi:10.1039/C7EE01303E.

6. Evans, N.P.; Collins, D.A.; Pierson, F.W.; Mahsoub, H.M.; Sriranganathan, N.; Persia, M.E.; Karnezos, T.P.; Sims, M.D.; Dalloul, R.A. Investigation of medium chain fatty acid feed supplementation for reducing *Salmonella typhimurium* colonization in turkey poults. *Foodborne Pathog. Dis.* **2017**, *14*, 531–536, doi:10.1089/fpd.2016.2273.

7.	Kenealy, W.R.; Cao, Y.; Weimer, P.J. Production of caproic acid by cocultures of ruminal cellulolytic bacteria and *Clostridium kluyveri* grown on cellulose and ethanol. *Appl. Microbiol. Biotechnol.* **1995**, *44*, 507–513, doi:10.1007/s002530050590.

8.	Anneken, D. J., Both, S., Christoph, R., Fieg, G., Steinberner, U., and Westfechtel, A. Fatty Acids. *Ullmann's Encycl. Ind. Chem.* **2006**, *14*, 73–116.

9.	Agler, M.T.; Wrenn, B.A.; Zinder, S.H.; Angenent, L.T. Waste to bioproduct conversion with undefined mixed cultures: The carboxylate platform. *Trends Biotechnol.* **2011**, *29*, 70–78, doi:10.1016/j.tibtech.2010.11.006.

10.	Wu, Q.; Bao, X.; Guo, W.; Wang, B.; Li, Y.; Luo, H.; Wang, H.; Ren, N. Medium chain carboxylic acids production from waste biomass: Current advances and perspectives. *Biotechnol. Adv.* **2019**, *37*, 599–615, doi:10.1016/j.biotechadv.2019.03.003.

11.	de Leeuw, K.D.; Buisman, C.J.N.; Strik, D.P.B.T.B. Branched medium chain fatty acids: *Iso*-caproate formation from *iso*-butyrate broadens the product spectrum for microbial chain elongation. *Environ. Sci. Technol.* **2019**, *53*, 7704–7713, doi:10.1021/acs.est.8b07256.

12.	Chen, W.S.; Huang, S.; Strik, D.P.B.T.B.; Buisman, C.J.N. Isobutyrate biosynthesis via methanol chain elongation: converting organic wastes to platform chemicals. *J. Chem. Technol. Biotechnol.* **2017**, *92*, 1370–1379, doi:10.1002/jctb.5132.

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*, 5119–5131, doi:10.1007/s00253-020-10551-w.

14.	Seedorf, H.; Fricke, W.F.; Veith, B.; Brüggemann, H.; Liesegang, H.; Strittmatter, A.; Miethke, M.; Buckel, W.; Hinderberger, J.; Li, F.; et al. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic

features. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 2128–2133, doi:10.1073/pnas.0711093105.

15. Zhu, X.; Zhou, Y.; Wang, Y.; Wu, T.; Li, X.; Li, D.; Tao, Y. Production of high-concentration *n*-caproic acid from lactate through fermentation using a newly isolated *Ruminococcaceae* bacterium CPB6. *Biotechnol. Biofuels* **2017**, *10*, 102, doi:10.1186/s13068-017-0788-y.

16. van Brabant P. Understanding bio-isomerisation during methanol fermentation. *M.Sc. Thesis*, Ghent University, Belgium, 2019.

17. Lambrecht, J.; Cichocki, N.; Schattenberg, F.; Kleinsteuber, S.; Harms, H.; Müller, S.; Sträuber, H. Key sub-community dynamics of medium-chain carboxylate production. *Microb. Cell Fact.* **2019**, *18*, 92, doi:10.1186/s12934-019-1143-8.

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, doi:10.3389/fmicb.2020.00336.

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**, *9*, e00679-20, doi: 10.1128/MRA.00679-20.

20. Müller, N.; Griffin, B.M.; Stingl, U.; Schink, B. Dominant sugar utilizers in sediment of Lake Constance depend on syntrophic cooperation with methanogenic partner organisms. *Environ. Microbiol.* **2008**, *10*, 1501–1511, doi:10.1111/j.1462-2920.2007.01565.x.

21. Widdel, F.; Kohring, G.W.; Mayer, F. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids - III. Characterization of the filamentous gliding *Desulfonema limicola* gen. nov. sp. nov., and *Desulfonema magnum* sp. nov. *Arch. Microbiol.* **1983**, *134*, 286–294, doi:10.1007/BF00407804.

22.   Widdel, F.; Bak, F. *Gram-Negative Mesophilic Sulfate-Reducing Bacteria*. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer KH (eds) The Prokaryotes. Springer Verlag, Berlin, **1992**; pp. 3352-3378.

23.   Widdel, F.; Pfennig, N. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids - I. Isolation of new sulfate-reducing bacteria enriched with acetate from saline environments. Description of *Desulfobacter postgatei* gen. nov., sp. nov. *Arch. Microbiol.* **1981**, *129*, 395–400, doi:10.1007/BF00406470.

24.   Patil, Y.; Junghare, M.; Müller, N. Fermentation of glycerol by *Anaerobium acetethylicum* and its potential use in biofuel production. *Microb. Biotechnol.* **2017**, *10*, 203–217, doi:10.1111/1751-7915.12484.

25.   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, doi:10.1186/2192-0567-2-13.

26.   Vallenet, D.; Calteau, A.; Dubois, M.; Amours, P.; Bazin, A.; Beuvin, M.; Burlot, L.; Bussell, X.; Fouteau, S.; Gautreau, G.; 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–D589, doi:10.1093/nar/gkz926.

27.   Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515, doi:10.1093/nar/gky1049.

28.   Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E. V.; Krylov, D.M.; Mazumder, R.; Smirnov, S.; Nikolskaya, A.N.; et al. The COG database: An updated vesion includes eukaryotes. *BMC Bioinformatics* **2003**, *4,* 41, doi:10.1186/1471-2105-4-41.

29.   Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; Von Mering, C.; Bork, P. Fast genome-wide functional annotation through

orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122, doi:10.1093/molbev/msx148.

30.   Meyer, F.; Overbeek, R.; Rodriguez, A. FIGfams: Yet another set of protein families. *Nucleic Acids Res.* **2009**, *37*, 6643–6654, doi:10.1093/nar/gkp698.

31.   Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L.; et al. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, 211–215, doi:10.1093/nar/gkn785.

32.   Tatusov, R.L. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36, doi:10.1093/nar/28.1.33.

33.   Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J.F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465–W469, doi:10.1093/nar/gkn180.

34.   Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T.L. NCBI BLAST: a better web interface. Nucleic Acids Res. *2008, 36,* W5–W9, doi:10.1093/nar/gkn201.

35.   Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.

36.   Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552, doi:10.1093/oxfordjournals.molbev.a026334.

37.   Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 696–704, doi:10.1080/10635150390235520.

38.   Anisimova, M.; Gascuel, O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **2006**, *55*, 539–552, doi:10.1080/10635150600755453.

39.   Chevenet, F.; Brun, C.; Bañuls, A.L.; Jacq, B.; Christen, R. TreeDyn: Towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **2006**, *7*, 439, doi:10.1186/1471-2105-7-439.

40.   Chaumeil, P.-A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **2019**, *36*, 1925–1927, doi:10.1093/bioinformatics/btz848.

41.   Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **2016**, *17*, 132, doi:10.1186/s13059-016-0997-x.

42.   Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *10*, P10008, doi:10.1088/1742-5468/2008/10/P10008.

43.   Yoon, S.H.; Ha, S. min; Lim, J.; Kwon, S.; Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **2017**, *110*, 1281–1286, doi:10.1007/s10482-017-0844-4.

44.   Vallenet, D.; Calteau, A.; Cruveiller, S.; Gachet, M.; Lajus, A.; Josso, A.; Mercier, J.; Renaux, A.; Rollin, J.; Rouy, Z.; et al. MicroScope in 2017: An expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res.* **2017**, *45*, D517–D528, doi:10.1093/nar/gkw1101.

45.   Erko, S.; Ebers, J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today* **2006**, *33*, 152–155.

46.   Scarborough MJ, Myers KS, Donohue TJ, N.D. 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, doi:10.1128/AEM.02242-19.

47.   Stanton, T.B.; Humphrey, S.B. 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*, 3874–3882, doi:10.1128/AEM.69.7.3874-3882.2003.

48.   Weimer, P.J.; Moen, G.N. Quantitative analysis of growth and volatile fatty acid production by the anaerobic ruminal bacterium *Megasphaera elsdenii* T81. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 4075–4081, doi:10.1007/s00253-012-4645-4.

49.   Tao, Y.; Zhu, X.; Wang, H.; Wang, Y.; Li, X.; Jin, H.; Rui, J. Complete genome sequence of *Ruminococcaceae* bacterium CPB6: A newly isolated culture for efficient *n*-caproic acid production from lactate. *J. Biotechnol.* **2017**, *259*, 91–94, doi:10.1016/j.jbiotec.2017.07.036.

50.   Jeon, B.S.; Kim, S.; Sang, B.I. *Megasphaera hexanoica* sp. Nov., a medium-chain carboxylic acid-producing bacterium isolated from a cow rumen. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 2114–2120, doi:10.1099/ijsem.0.001888.

51.   Willems, A.; Collins, M.D. 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*, 1083–1087, doi:10.1099/00207713-46-4-1083.

52.   Holdeman, L. V.; Cato, E.P.; Moore, W.E.C. Amended description of *Ramibacterium alactolyticum* Prevot and Taffanel with proposal of a neotype strain. *Int. J. Syst. Bacteriol.* **1967**, *17*, 323–341, doi:10.1099/00207713-17-4-323.

53.   Kim, B.C.; Jeon, B.S.; Kim, S.; Kim, H.; Um, Y.; Sang, B.I. *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*, 4902–4908, doi:10.1099/ijsem.0.000665.

54.   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, doi:10.1128/MRA.00346-19.

55.   Genthner, B.R.; Davis, C.L.; Bryant, M.P. Features of rumen and sewage sludge strains of *Eubacterium limosum*, a methanol- and H_2-CO_2-utilizing species. *Appl. Environ. Microbiol.* **1981**, *42*, 12–19, doi:10.1007/s11011-014-9623-3.

56.   Wallace, R.J.; McKain, N.; McEwan, N.R.; Miyagawa, E.; Chaudhary, L.C.; King, T.P.; Walker, N.D.; Apajalahti, J.H.A.; Newbold, C.J. *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*, 965–970, doi:10.1099/ijs.0.02110-0.

57.   Wallace, R.J.; Chaudhary, L.C.; Miyagawa, E.; McKain, N.; Walker, N.D. Metabolic properties of *Eubacterium pyruvativorans*, a ruminal 'hyper-ammonia-producing' anaerobe with metabolic properties analogous to those of *Clostridium kluyveri*. *Microbiology* **2004**, *150*, 2921–2930, doi:10.1099/mic.0.27190-0.

58.   Wang, Q.; Wang, C.D.; Li, C.H.; Li, J.G.; Chen, Q.; Li, Y.Z. *Clostridium luticellarii* sp. nov., isolated from a mud cellar used for producing strong aromatic liquors. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 4730–4733, doi:10.1099/ijsem.0.000641.

59.   Poehlein A, Bremekamp R, Lutz VT, Schulz LM, D.R. Draft genome sequence of the butanoic acid-producing bacterium *Clostridium luticellarii* DSM 29923, used for strong aromatic Chinese liquor production. *Genome Annoucements* **2018**, *6,* e00556-18, doi:10.1128/genomeA.00504-17.

60.   Gil, R.; Silva, F.J.; Pereto, J.; Moya, A. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 518–537, doi:10.1128/MMBR.68.3.518.

61.   Agler, M.T.; Spirito, C.M.; Usack, J.G.; Werner, J.J.; Angenent, L.T. Chain elongation with reactor microbiomes: upgrading dilute ethanol to medium-chain carboxylates. *Energy Environ. Sci.* **2012**, *5*, 8189, doi:10.1039/c2ee22101b.

62.   Kucek, L.; Spirito, C.M.; Angenent, L.T. 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*, 3482–3494, doi:10.1039/C6EE01487A.

63.   Grootscholten, T.I.M.; Steinbusch, K.J.J.; Hamelers, H.V.M.; Buisman, C.J.N. Chain elongation of acetate and ethanol in an upflow anaerobic filter for high rate MCFA production. *Bioresour. Technol.* **2013**, *135*, 440–445, doi:10.1016/j.biortech.2012.10.165.

64.   Stadtman, E.R. The coenzyme A transphorase system in *Clostridium kluyveri. J. Biol. Chem.* **1953**, *203*, 501–512.

65.   Stadtman, E.R. Functional group of coenzyme A and its metabolic relations, especially in the fatty acid cycle: *Discussion. Fed. Proc.* **1953**, *12*, 692–693.

66.   Louis, P.; Duncan, S.H.; McCrae, S.I.; Millar, J.; Jackson, M.S.; Flint, H.J. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J. Bacteriol.* **2004**, *186*, 2099–2106, doi:10.1128/JB.186.7.2099-2106.2004.

67.   Matthew J. Scarborough, Joshua J. Hamilton, Elizabeth A. Erb, Timothy J. Donohue, D.R.N. Diagnosing and predicting mixed-culture fermentations with unicellular and guild-based metabolic models. *mSystems* **2020**, *5*, e00755-20, doi:10.1128/mSystems.00755-20.

68.   Tholozan, J.L.; Samain, E.; Grivet, J.P. Isomerization between *n*-butyrate and isobutyrate in enrichment cultures. *FEMS Microbiol. Lett.* **1988**, *53*, 187–191, doi:10.1016/0378-1097(88)90441-7.

69.   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*, 1435–1439, doi:10.1128/AEM.58.5.1435-1439.1992.

70.   Cracan, V.; Padovani, D.; Banerjee, R. IcmF is a fusion between the radical B_{12} enzyme isobutyryl-CoA mutase and its G-protein chaperone. *J. Biol. Chem.* **2010**, *285*, 655–666, doi:10.1074/jbc.M109.062182.

71.   Barker, H.A. Coenzyme B_{l2}-Dependent mutases causing carbon chain rearrangements. *Enzym.* **1972**, *6*, 509–537, doi:10.1016/S1874-6047(08)60049-9.

72.   Allison, M.J. Production of branched-chain volatile fatty acids by certain anaerobic bacteria. *Appl. Environ. Microbiol.* **1978**, *35*, 872–877, doi:10.1128/AEM.35.5.872-877.1978.

73.   Stieb, M.; Schink, B. Anaerobic degradation of isobutyrate by methanogenic enrichment cultures and by a *Desulfococcus multivorans* strain. *Arch. Microbiol.* **1989**, *151*, 126–132, doi:10.1007/BF00410965.

74.   Marshall, V.D.; Sokatch, J.R. Regulation of valine catabolism in *Pseudomonas putida*. *J. Bacteriol.* **1972**, *110*, 1073–1081, doi:10.1128/JB.110.3.1073-1081.1972.

75.   Gollop, N.; Damri, B.; Barak, Z.; Chipman, D.M. Kinetics and mechanism of acetohydroxy acid synthase isozyme III from *Escherichia coli*. *Biochemistry* **1989**, *28*, 6310–6317, doi:10.1021/bi00441a024.

76.   Buckel, W.; Thauer, R.K. Flavin-based electron bifurcation, a new mechanism of biological energy coupling. *Chem. Rev.* **2018**, *118*, 3862–3886, doi:10.1021/acs.chemrev.7b00707.

77.	Buckel, W.; Thauer, R.K. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na⁺ translocating ferredoxin oxidation. *Biochim. Biophys. Acta - Bioenerg.* **2013**, *1827*, 94–113, doi:10.1016/j.bbabio.2012.07.002.

78.	Hedderich, R.; Forzi, L. Energy-converting [NiFe] hydrogenases: More than just H$_2$ activation. *J. Mol. Microbiol. Biotechnol.* **2006**, *10*, 92–104, doi:10.1159/000091557.

79.	Schuchmann, K.; Müller, V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* **2014**, *12*, 809–821, doi:10.1038/nrmicro3365.

80.	Weghoff, M.C.; Bertsch, J.; Müller, V. A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ. Microbiol.* **2015**, *17*, 670–677, doi:10.1111/1462-2920.12493.

81.	Sträuber, H.; Bühligen, F.; Kleinsteuber, S.; Dittrich-Zechendorf, M. Carboxylic acid production from ensiled crops in anaerobic solid-state fermentation - trace elements as pH controlling agents support microbial chain elongation with lactic acid. *Eng. Life Sci.* **2018**, *0*, 447–458, doi:10.1002/elsc.201700186.

82.	Lane, D.J. *16S/23S rRNA sequencing*. *In* Stackbrandt E and Goodfellow M (Eds.), Nucleic acid techniques in bacterial systematics, John Wiley and Sons, Chichester. **1991**; pp. 177-203.

Figures
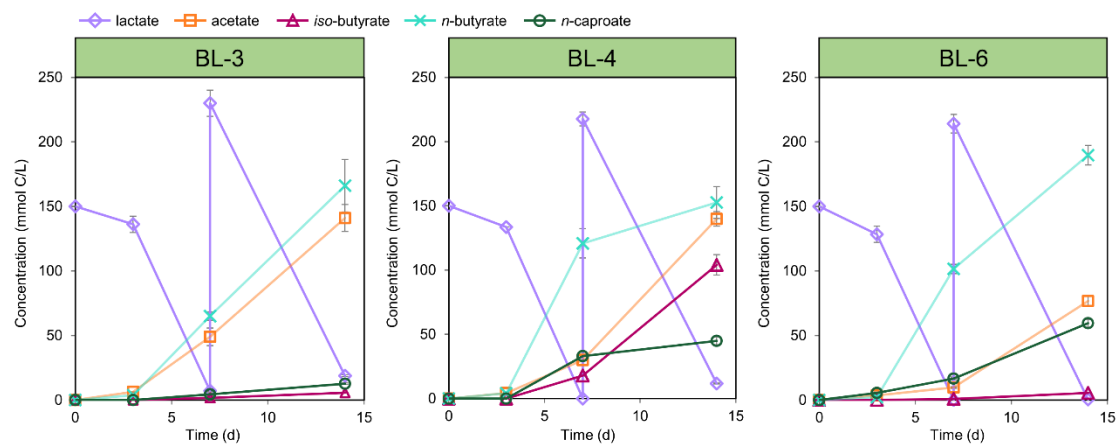
Figure 1. Fermentation products of strains BL-3, BL-4 and BL-6 during growth on lactate. 75 mM lactic acid was added to each bottle on day 7. Mean values of six measurements of duplicate batch cultures are given and error bars represent the standard deviation.

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.

185
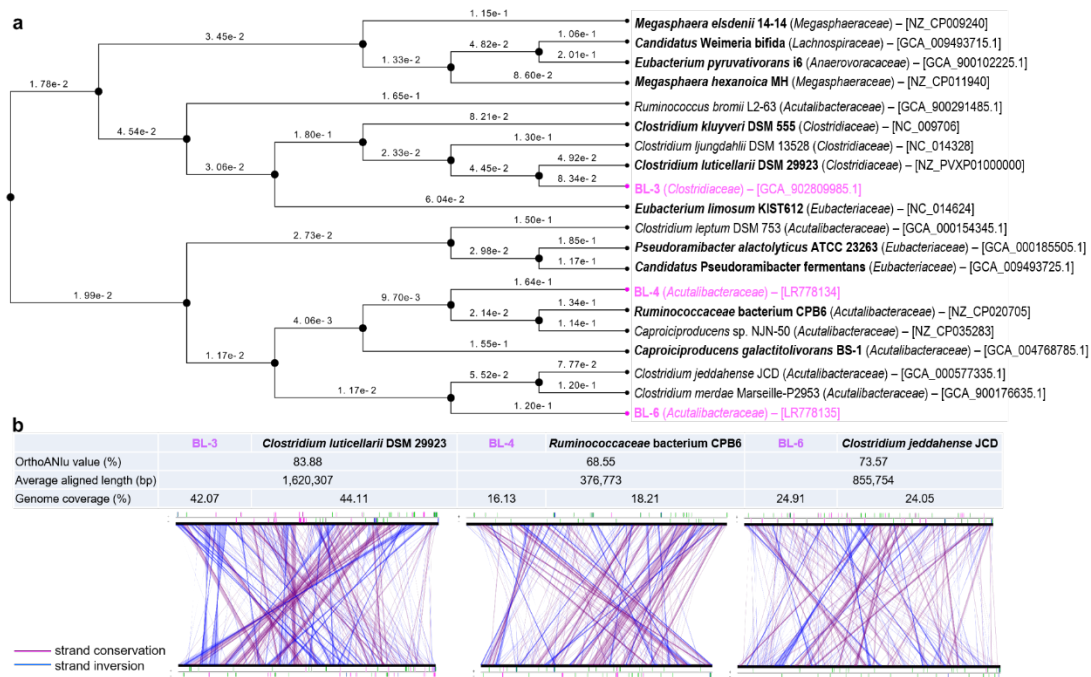
Figure 3. Phylogenomic analysis of the three isolates. **(a)** Neighbor-joining tree showing the genome similarity between 14 chain-elongating bacterial strains. The newly isolated strains are highlighted in pink and all experimentally validated chain-elongating strains are indicated in bold. Additional related species based on 16S rRNA phylogenetic analysis were included (see the phylogenetic tree in Figure S3). GTDB taxonomic assignments at the family level are shown in parentheses. The NCBI/ENA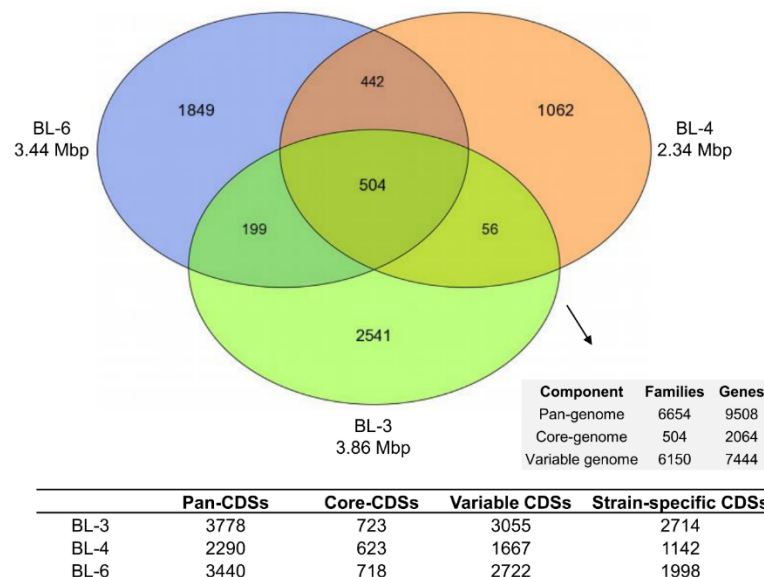 accession numbers of the genomes are shown in brackets. Distances indicated at the branches correlate to the average nucleotide identity (ANI) according to: $D \approx 1\text{-ANI}$. **(b)** USEARCH OrthoANI comparison for strains BL-3, BL-4 and BL-6 to related genomes. The line plots give an overview of the conservation of synteny groups on nucleotide level. Strand conservations are depicted in purple and strand inversions in blue. The synton size was selected with higher than three genes for the analysis.
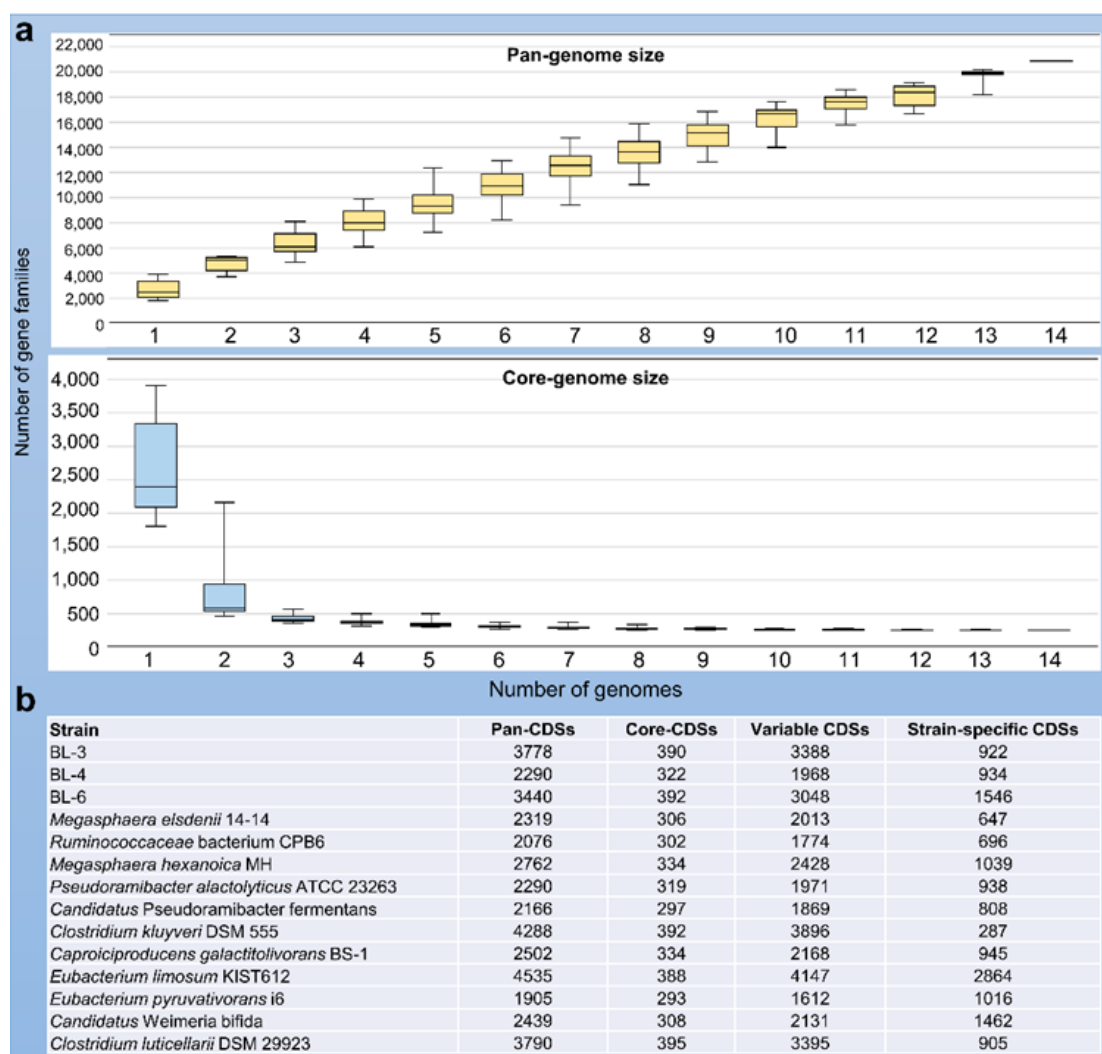
Figure 4. Pan-genome analysis of the 14 chain-elongating bacterial strains. **(a)** Pan-genome and core-genome sizes and their changes for the increasing genome set. Families refer to the MicroScope homologous gene families (MICFAM), in which the protein-coding genes share at 80% of amino acid sequence identity and 80% of alignment coverage. **(b)** Summary of gene counts for each strain. CDS: gene coding sequence.
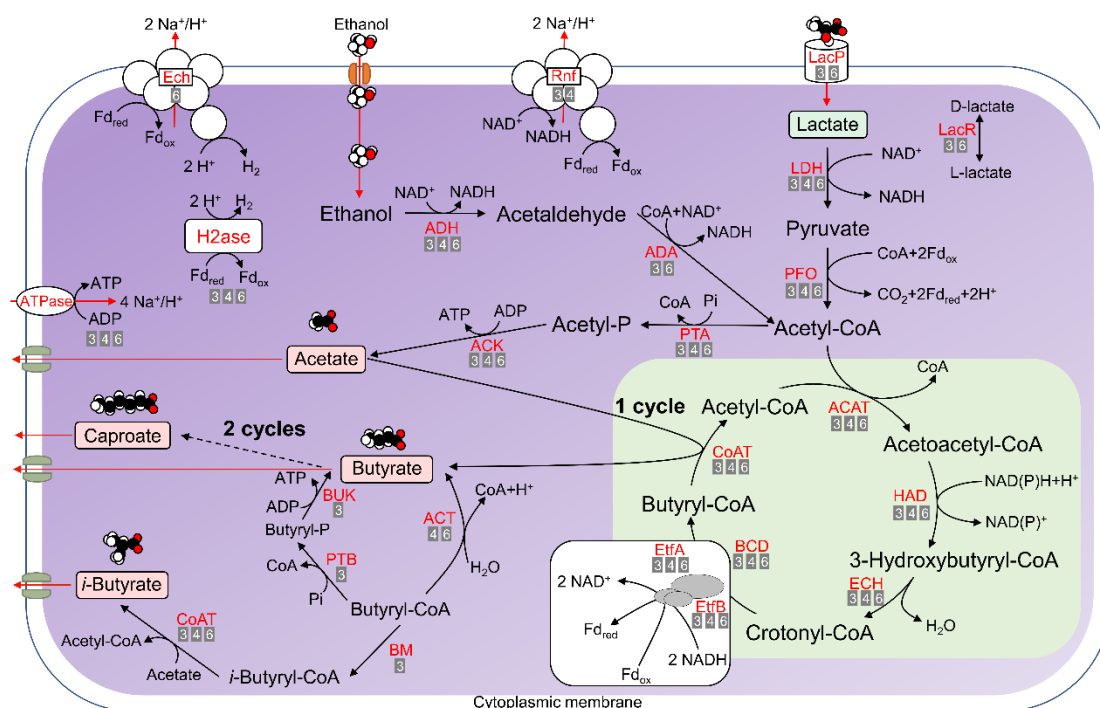
Figure 5. Metabolic pathways involved in lactate-based or ethanol-based chain elongation and production of acetate, *n*-butyrate, *iso*-butyrate and *n*-caproate as predicted from the genome annotation of strains BL-3, BL-4 and BL-6. Enzyme abbreviations (see Table 2 for full names) are provided in red letters next to the pathways (solid lines). The numbers below the enzyme names indicate the strains that were predicted to harbour 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 dashed line represents multi-enzyme reactions between the two indicated molecules, and "cycle" refers to the reverse β-oxidation. The conversion of the terminal acyl-CoA to the corresponding fatty acid can be catalysed by CoAT or alternatively by ACT as shown at the example of butyrate. A third way of butyrate formation from butyryl-CoA proceeds via PTB and BUK. The predicted pathway of *iso*-butyrate formation via isomerisation of *n*-butyryl-CoA by BM is shown; an alternative hypothetical pathway for *iso*-butyrate formation from lactate is depicted in **Figure S4** (Supplemental File 1).
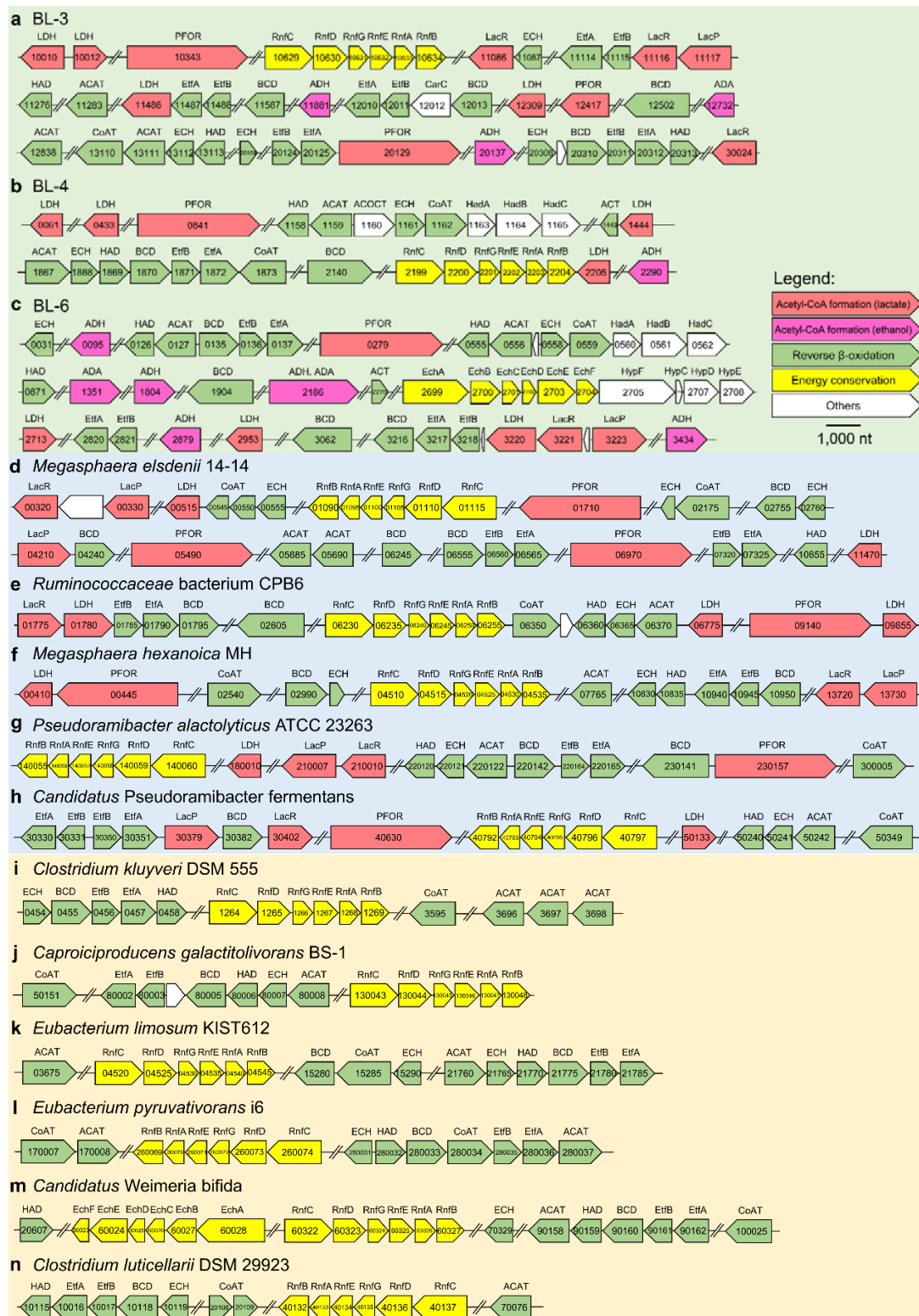
Figure 6. Arrangement of predicted CDSs in genomes of strains BL-3 **(a)**, BL-4 **(b)**, BL-6 **(c)**, other bacterial strains reported of chain elongation with lactate **(d-h),** and with other reduced substrates **(i-n)**. Numbers in the arrows denote the corresponding CDS labels. Abbreviations above the arrow refer to the enzyme names (see Table 2 for full names). Scale bar: 1,000 nucleotides (nt).
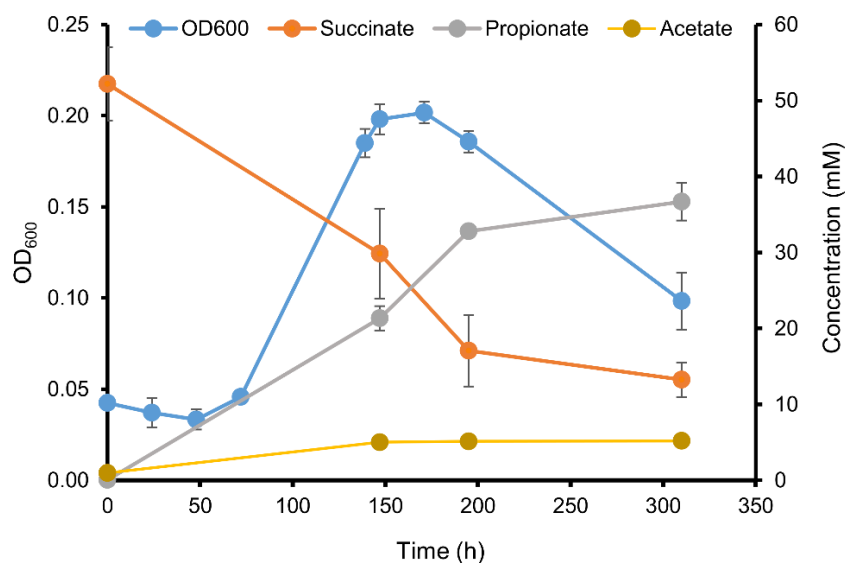
Figure 7. Fermentation kinetics of strain BL-4 during growth on 50 mM succinate and 0.05% yeast extract. Shown are mean values of triplicates. Error bars represent the standard deviation. Some error bars are smaller than symbol size. Small amounts (< 2 mM) of formate, butyrate, *iso*-butyrate, and 1-propanol were also formed, but omitted in the figure for clarity.

Figure 8. Fermentation kinetics of strain BL-4 during growth on 50 mM L-valine and 0.05% yeast extract. Shown are mean values of triplicates. Error bars represent the standard deviation. Some error bars are smaller than symbol size. **a:** optical density at 600 nm **b:** difference of fermentation products identified and quantified by HPLC ($t_{end} - t_0$ values).

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

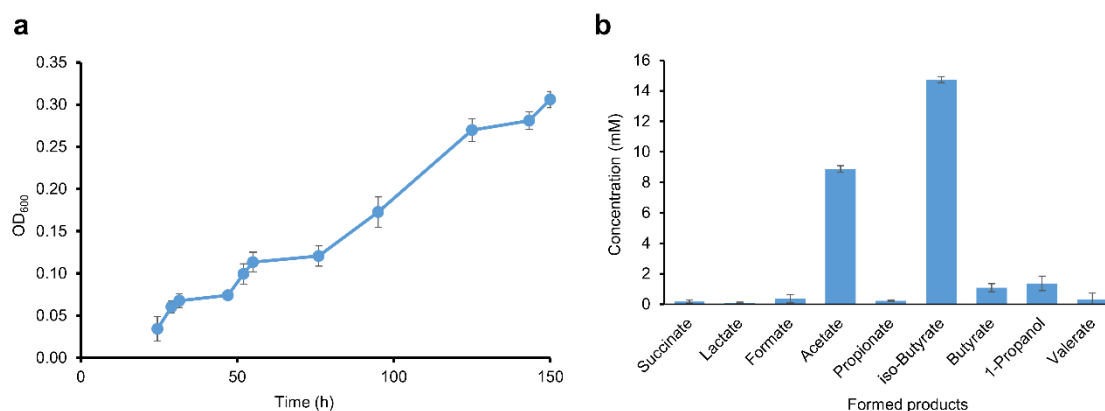| Strain	GTDB taxonomy	Isolation source	Genome size (bp)	GC content (%)	No. of predicted CDSs	Reference
BL-3	*Clostridium_B*	Anaerobic bioreactor	3,855,691	34.32	3,875	[19]
BL-4	*Acutalibacteraceae* UBA4871	Anaerobic bioreactor	2,335,857	42.75	2,323	[19]
BL-6	*Clostridium_E* sp002397665	Anaerobic bioreactor	3,435,529	54.63	3,496	[19]
Megasphaera elsdenii 14-14	*Megasphaera elsdenii*	Human gut	2,504,349	52.75	2,359	[47,48]
Ruminococcaceae bacterium CPB6	*Acutalibacteraceae* UBA4871 sp002119605	Sludge of a caproate-producing reactor	2,069,994	50.58	2,116	[15,49]
Megasphaera hexanoica MH	*Caecibacter massiliensis*	Cow rumen	2,877,851	49.00	2,799	[50]
Pseudoramibacter alactolyticus ATCC 23263	*Pseudoramibacter alactolyticus*	Human oral cavity	2,366,982	51.63	2,327	[51,52]
Candidatus Pseudoramibacter fermentans[a]	*Pseudoramibacter* sp002396065	Anaerobic bioreactor	2,288,358	50.15	2,209	[46]
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	[53,54]
Eubacterium limosum KIST612	*Eubacterium limosum*	Sheep rumen	4,740,532	46.86	4,605	[51,55]
Eubacterium pyruvativorans i6	*Eubacterium_A pyruvativorans*	Sheep rumen	2,164,212	54.84	1,954	[56,57]
Candidatus Weimeria bifida[a]	*Lachnospiraceae* UBA2727	Anaerobic bioreactor	2,395,883	45.93	2,477	[46]
Clostridium luticellarii DSM 29923	*Clostridium_B luticellarii*	Mud cellar	3,771,178	34.97	3,874	[58,59]

[a] metagenome-assembled genome (MAG)

Table 2. List of enzymes considered for the manual functional annotation

Predicted function	No.	Enzyme abbreviation	EC number	Enzyme
Acetyl-CoA formation	1	LacR	5.1.2.1	Lactate racemase
	2	LacP	2.A.14	Lactate permease
	3	LDH	1.1.1.27	Lactate dehydrogenase
	4	PFOR	1.2.7.1	Pyruvate ferredoxin oxidoreductase
	5	ADH	1.1.1.1	Alcohol dehydrogenase
	6	ADA	1.2.1.10	Acetaldehyde dehydrogenase
Reverse β-oxidation	7	ACAT	2.3.1.9, 2.3.1.16	Acetyl-CoA acetyltransferase
	8	HAD	1.1.1.157, 1.1.1.35	3-Hydroxyacyl-CoA dehydrogenase
	9	ECH	4.2.1.150, 4.2.1.55	Enoyl-CoA hydratase
	10	BCD	1.3.8.1	Butyryl-CoA dehydrogenase
	11	EtfAB		Electron transfer flavoprotein A,B
	12	CoAT	2.8.3.-	Butyryl-CoA:acetate CoA-transferase
	13	ACT	3.1.2.20	Acyl-CoA thioesterase
Energy conservation	14	RnfABCDEG	7.1.1.1	Energy-converting NADH:ferredoxin oxidoreductase
	15	EchABCDEF		Energy-converting hydrogenase
H2 formation	16	H2ase	1.12.7.2	Hydrogen:ferredoxin oxidoreductase
Butyrate formation	17	PTB	2.3.1.19	Phosphate butyryltransferase
	18	BUK	2.7.2.7	Butyrate kinase
Others	19	BM	5.4.99.13	Butyryl-CoA:isobutyryl-CoA mutase
	20	ACOCT	2.8.3.19	Acetyl-CoA:oxalate CoA-transferase
	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

2.5.2 Supplementary information

## Supplementary methods: Sanger sequencing of 16S rRNA genes

Genomic DNA was extracted from fresh cell pellets of the isolates and purified using the NucleoSpin Microbial DNA kit (Macherey-Nagel, Germany). The concentration and quality of DNA were determined by NanoDrop™ UV-Vis spectrophotometer (NanoDrop™ ONE, Thermo Scientific, Waltham, USA) and by agarose gel electrophoresis. Amplification of bacterial 16S rRNA genes by PCR using MyTaq™ Mix (Bioline, Germany) and sequencing were carried out as described previously [81], with few modifications. For almost complete sequencing of 16S rRNA genes, sequencing primers 27f, 357f, 519r, 530f, 927r, 1104r, 1114f and 1492r were used [82]. Amplicons were purified using the SureClean Kit (Bioline, Germany) and quantified using the NanoDrop. The DNA sequence analysis software Sequencher® v5.4.6 (Gene Codes Corporation, Ann Arbor, MI USA) was used for trimming and aligning the forward and reverse sequences and assembling contigs. The sequences were compared against the National Center for Biotechnology Information (NCBI) rRNA/ITS databases (16S ribosomal RNA sequences (Bacteria and Archaea)) using the nucleotide BLAST (Basic Local Alignment Search Tool) web interface [34].

Table S1. COG (Clusters of Orthologous Groups) classification

Process	Class ID	Description	BL-3 CDS	BL-3 %	BL-4 CDS	BL-4 %	BL-6 CDS	BL-6 %
Cellular processes and signaling	D	Cell cycle control, cell division, chromosome partitioning	65	1.68	32	1.38	44	1.26
Cellular processes and signaling	M	Cell wall/membrane/envelope biogenesis	171	4.41	116	4.99	129	3.69
Cellular processes and signaling	N	Cell motility	71	1.83	9	0.39	68	1.95
Cellular processes and signaling	O	Posttranslational modification, protein turnover, chaperones	86	2.22	47	2.02	74	2.12
Cellular processes and signaling	T	Signal transduction mechanisms	200	5.16	82	3.53	146	4.18
Cellular processes and signaling	U	Intracellular trafficking, secretion, and vesicular transport	66	1.70	30	1.29	59	1.69
Cellular processes and signaling	V	Defense mechanisms	72	1.86	61	2.63	65	1.86
Cellular processes and signaling	W	Extracellular structures	3	0.08	6	0.26	1	0.03
Information storage and processing	B	Chromatin structure and dynamics	1	0.03	1	0.04	1	0.03
Information storage and processing	J	Translation, ribosomal structure and biogenesis	161	4.15	148	6.37	153	4.38
Information storage and processing	K	Transcription	324	8.36	196	8.44	287	8.21
Information storage and processing	L	Replication, recombination and repair	246	6.35	113	4.86	259	7.41
Metabolism	C	Energy production and conversion	271	6.99	120	5.17	168	4.81
Metabolism	E	Amino acid transport and metabolism	389	10.04	241	10.37	306	8.75
Metabolism	F	Nucleotide transport and metabolism	75	1.94	56	2.41	64	1.83
Metabolism	G	Carbohydrate transport and metabolism	177	4.57	126	5.42	263	7.52
Metabolism	H	Coenzyme transport and metabolism	131	3.38	52	2.24	109	3.12
Metabolism	I	Lipid transport and metabolism	86	2.22	61	2.63	70	2.00
Metabolism	P	Inorganic ion transport and metabolism	218	5.63	105	4.52	182	5.21
Metabolism	Q	Secondary metabolites biosynthesis, transport and catabolism	70	1.81	23	0.99	41	1.17
Poorly characterized	R	General function prediction only	483	12.46	271	11.67	349	9.98
Poorly characterized	S	Function unknown	250	6.45	154	6.63	175	5.01

Table S2. EGGNOG (Evolutionary Genealogy of Gene: Non-supervised Orthologous Groups) classification

Process	Class ID	Description	BL-3		BL-4		BL-6	
			CDS	%	CDS	%	CDS	%
Cellular processes and signaling	D	Cell cycle control, cell division, chromosome partitioning	33	0.85	25	1.08	28	0.80
Cellular processes and signaling	M	Cell wall/membrane/envelope biogenesis	168	4.34	134	5.77	126	3.60
Cellular processes and signaling	N	Cell motility	46	1.19	4	0.17	37	1.06
Cellular processes and signaling	O	Posttranslational modification, protein turnover, chaperones	91	2.35	48	2.07	65	1.86
Cellular processes and signaling	T	Signal transduction mechanisms	143	3.69	67	2.88	105	3.00
Cellular processes and signaling	U	Intracellular trafficking, secretion, and vesicular transport	38	0.98	23	0.99	37	1.06
Cellular processes and signaling	V	Defense mechanisms	72	1.86	56	2.41	67	1.92
Cellular processes and signaling	W	Extracellular structures	1	0.03	0	0	0	0
Information storage and processing	B	Chromatin structure and dynamics	1	0.03	1	0.04	1	0.03
Information storage and processing	J	Translation, ribosomal structure and biogenesis	157	4.05	147	6.33	146	4.18
Information storage and processing	K	Transcription	267	6.89	159	6.84	230	6.58
Information storage and processing	L	Replication, recombination and repair	221	5.70	111	4.78	261	7.47
Metabolism	C	Energy production and conversion	267	6.89	112	4.82	150	4.29
Metabolism	E	Amino acid transport and metabolism	315	8.13	199	8.57	261	7.47
Metabolism	F	Nucleotide transport and metabolism	76	1.96	59	2.54	65	1.86
Metabolism	G	Carbohydrate transport and metabolism	129	3.33	92	3.96	204	5.84
Metabolism	H	Coenzyme transport and metabolism	109	2.81	38	1.64	93	2.66
Metabolism	I	Lipid transport and metabolism	75	1.94	56	2.41	62	1.77
Metabolism	P	Inorganic ion transport and metabolism	158	4.08	89	3.83	185	5.29
Metabolism	Q	Secondary metabolites biosynthesis, transport and catabolism	41	1.06	14	0.60	19	0.54
Poorly characterized	S	Function unknown	979	25.26	559	24.06	782	22.37
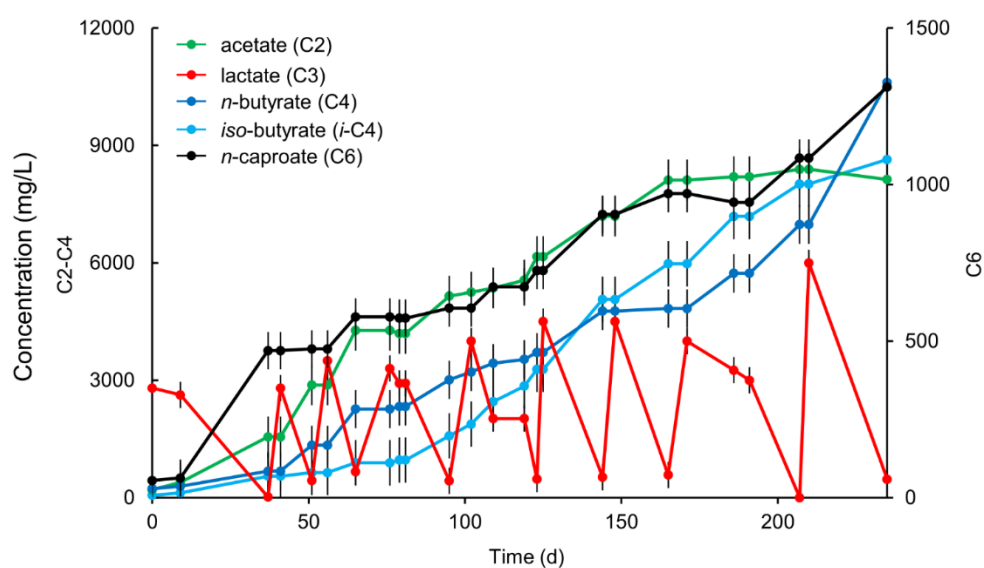
Figure S1. Fermentation products of the enrichment culture (a single bottle of the fourth transfer) during growth on lactate. Mean values of three measurements are given and error bars represent the standard deviation.
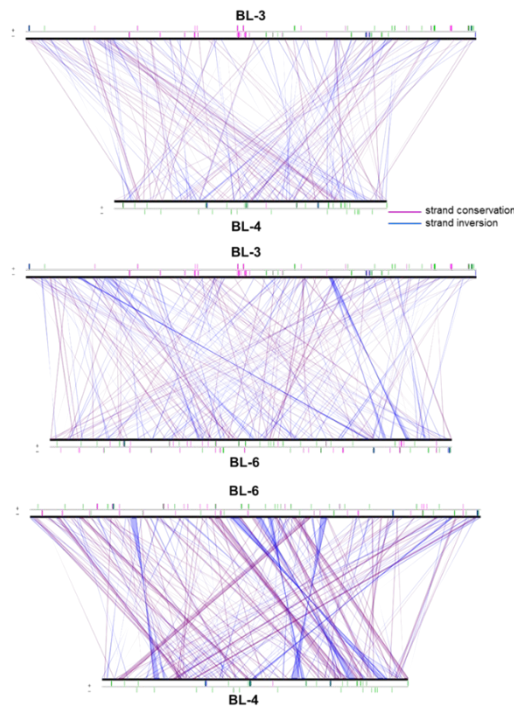
Figure S2. Pairwise comparison of the conservation of the synteny groups in the three new isolates. Strand conservations are depicted in purple and strand inversions in blue. The synton size was selected higher than three genes.
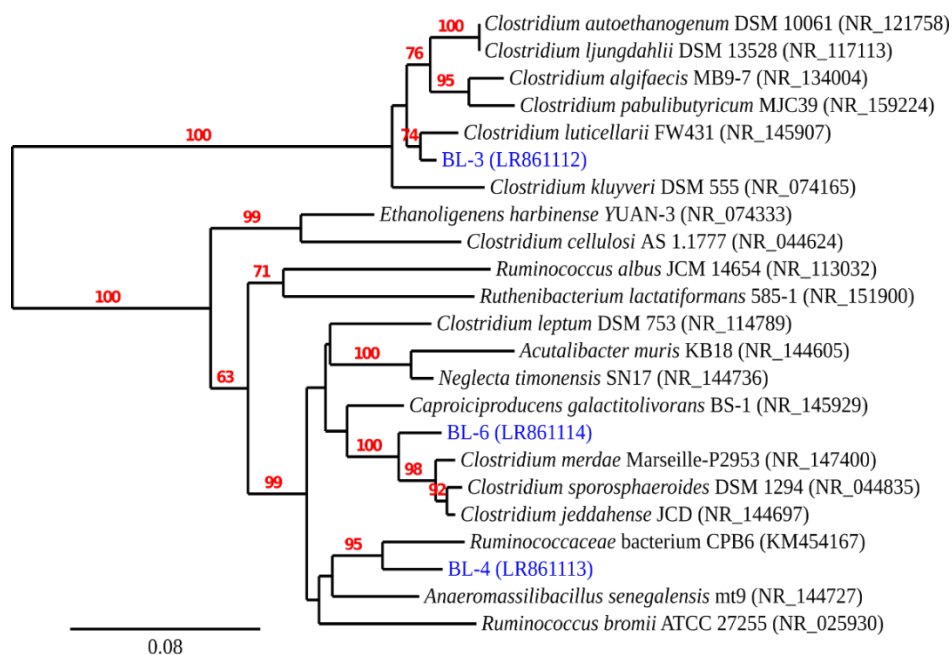
Figure S3. Maximum likelihood tree of the three new strains and closest relatives based on 16S rRNA gene sequences. Bootstrap values above 50% are shown at the node. GenBank or European Nucleotide Archive (ENA) accession numbers of 16S rRNA sequences are presented in parentheses. Scale bar = 8% nucleotide substitution.

Figure S4. Hypothetical *iso*-butyrate-producing lactate degradation pathway with the independence of *iso*-butyrate-CoA-mutase. Numbers represent the locus tags of the predicted genes for strain BL-4. See the corresponding CDSs in Supplementary file 4.

3 General discussion

3.1 Understanding microbial community assembly in model ecosystems

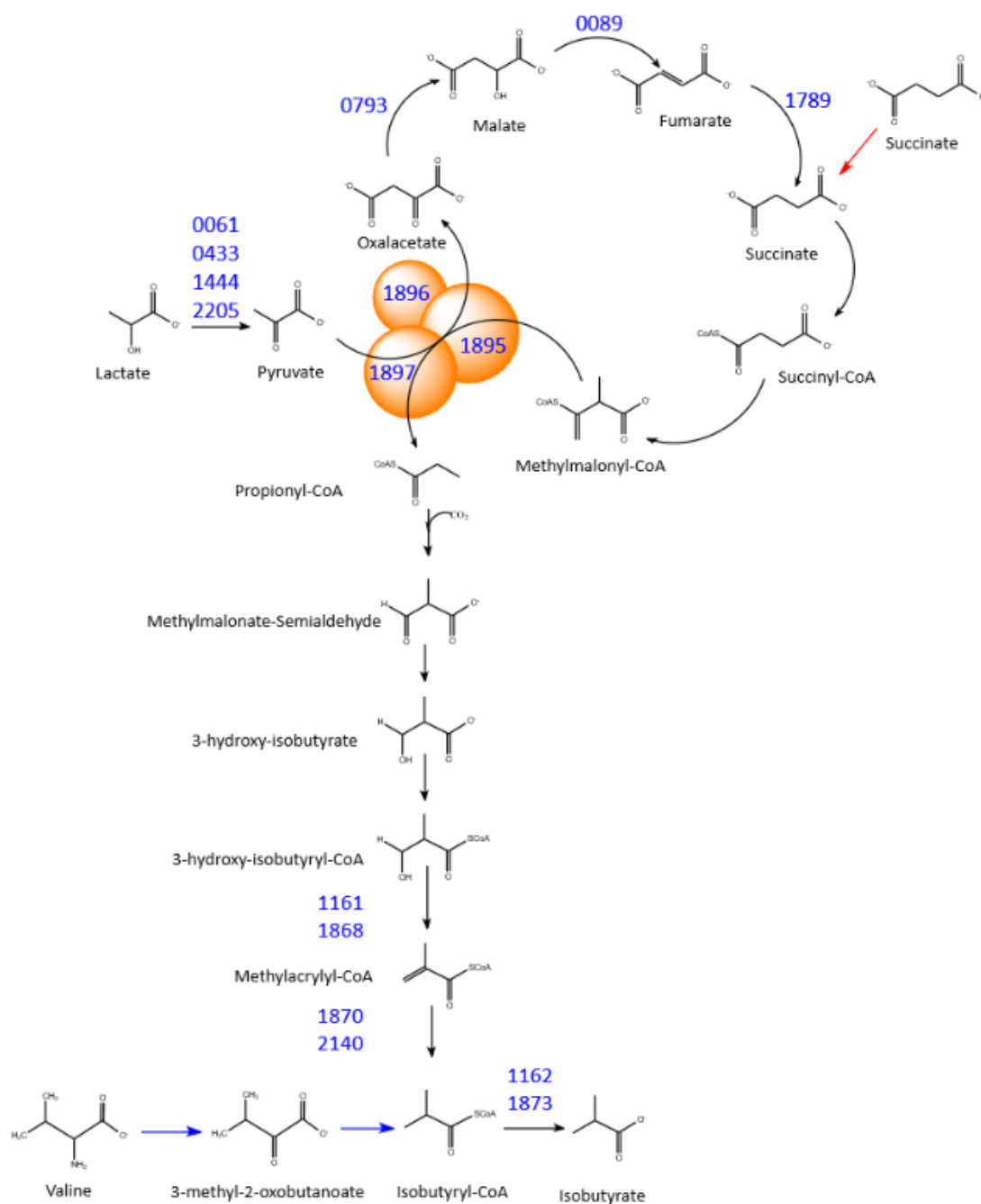The interactions of microorganisms shape the composition and function of the microbial community in an ecosystem (Zengler and Zaramela, 2018). The ecosystem also consists of various species that interact with the environment, making it more difficult to explore. Currently, we still lack a deeper understanding of the rules governing microbial community assembly (Faust, 2019). The contributions of deterministic and stochastic processes in the assembly of microbial communities are generally accepted, but characterising these processes in natural systems is a grand challenge (Wu et al., 2019a). In engineered systems such as bioreactors, by developing ecosystem models, we could have the opportunity to understand the complexity of community assembly. In this PhD thesis, by enriching self-assembled chain-elongating microbiota in continuous reactors, I explored microbial ecosystem models with a top-down approach. Here, the top-down approach refers to look at the enriched consortia as a whole while the bottom-up approach focuses more on synthetic communities. Through ecological selection, the top-down approach is used to understand how the manipulation of environmental factors (e.g., alterations of pH and hydraulic retention time) would force the existing microbiota to reassemble in order to maximise the growth yield. Afterwards, different mathematical models can be applied to capture community functioning by representing different functional groups of species, and to infer the potential reasons in the explanation of the microbial community dynamics. In other words, these investigations by using 16S rRNA gene analysis or metagenomics coupled to modelling or machine learning are valuable for generating different hypotheses in the field of microbial ecology, although most theories are in a rational way transferred from ecology of macroorganisms, such as the community assembly mechanisms controlling diversity patterns (Zhou and Ning, 2017; Ning et al., 2019). On the one hand, with rapid advances in sequencing technologies, categorising microbial diversity becomes relatively easy. On the

other hand, due to the dynamics of microbial ecosystems with numerous interacting species, it is still not possible to disentangle the factors controlling community assembly. For our chain-elongating reactor systems with highly enriched mixed cultures, a number of around 100 taxa (ASVs) is more than enough for any confirmed conclusions drawn from such advanced data analyses (Chapters 2.1, 2.2 and 2.3). As it is common to almost all kind of top-down studies, the ecological questions cannot be answered clearly (Prosser et al., 2007).

Going back to the starting point in understanding microbial community assembly, I anticipate boldly that defined mixed cultures with a reasonable number of species under well-controlled conditions in many replicates hold the promise for a holistic mechanistic understanding of community assembly (Vrancken et al., 2019). Most of the synthetic communities used for studying their dynamics are based on simple consortia with selected model microorganisms for a specific environment. There are some limitations for the approach of bottom-up design. First, it trades clarity by the sacrifices of reflecting the realism of the communities. According to my knowledge, no truly representative systems have been developed that exactly mimic *in situ* environments. Second, most microorganisms relevant for environmental processes resist cultivation, not to mention the low availability of well-documented strains regarding their ecophysiology and metabolism (Lawson et al., 2019). The CE process is a good example. Not many chain-elongating bacterial strains have been isolated for utilising lactate, ethanol or monosaccharides (Chapter 2.5). Other mysteries of CE include the unclearness of reverse β-oxidation pathway regarding the substrate fluxes and flexibility, as well as enzymology (e.g., the missing of trans-2-enoyl-CoA reductase in the chain elongator *Ruminococcaceae* bacterium CPB6; and terminal enzymes for controlling the carbon length) and the pathway using acyl carrier protein (ACP) as opposed to CoA in the fatty acid biosynthesis (Liao et al., 2016; Tao et al., 2017; Han et al., 2018). The knowledge in understanding membrane and enzyme processes of the toxicity of carboxylic acids is missing. The difficult maintenance of cultures is also challenging to construct synthetic communities.

Given the truth that many species need to be characterised, at the moment, exploration in the simplified and well-controlled systems is the most viable option to start unveiling the metabolic and ecological interactions within community members. Many studies have shown promise in this way, including the classical co-cultivation of syntrophic communities (Stams and Plugge, 2009), the competitive exclusion of the Gause principle (Connell and Sousa, 1983), the huge numbers of settings of defined communities showing the initial evenness of the community influences ecosystem function (Wittebolle et al., 2009) and a groundbreaking study of dynamics of three interacting species in closed ecosystems with many replicates, showing the effects of geometric random walks in a system are clearly contingent (Hekstra and Leibler, 2012). As for CE, the proposed mutualism of chain-elongating species and lactic acid bacteria (e.g., *Olsenella* and *Lactobacillus*) was mentioned in many CE studies (Contreras-Dávila et al., 2020; Liu et al., 2020a), but this symbiotic interaction has never been proved systemically in co-cultivation or any bottom-up designs. Sequencing of microbial communities is really convenient but its generated data are overwhelming compared with cultivation studies, which are important to prove those findings from sequencing studies.

High-throughput cultivation techniques would be a perfect match for parallel tests in defined mixed culture fermentation. The system of mini-bioreactor arrays is one of the examples, in which a continuous-flow mode with reactors miniaturised in volume and operated up to 48 in parallel was developed (Auchtung et al., 2015). A future combination with microsensor devices that determine chemical properties, community structure and functioning and ecophysiological parameters can be developed as an in-line monitoring system. The acquired high quality data would be important for any kind of model development to test and prove ecological theories. Besides the high-throughput feature, affordability and easy usability also need be considered. In this case, droplet-based microfluidics is also promising in the characterisation of microbial communities (Kehe et al., 2019).

3.2 Linking microbial community structure to functioning

A critical topic in ecology is to understand the relationship between diversity and ecosystem functioning (Prosser et al., 2007). How are microbial diversity and interactions relevant to ecosystem functioning, stability and resilience? For this PhD thesis, I found that decreasing the hydraulic retention time can shape the reactor microbiota resulting in the improvement of CE process performance in terms of C6/C8 productivity (Chapter 2.2). However, when testing the increase in pH gradient (Chapter 2.3), the reactor performance returned to the previous state while the communities developed to a different state (i.e., decreases in diversity and evenness but increase in richness). Since the CE bioreactors are engineered systems, I assumed that there would be a strong link between the reactor microbiota and its functions regarding C6/C8 production. Thus, it is important to understand the complexity of the microbial communities with parallel pathways of an enormous number of coexisting but taxonomically distinct species, and the changes to communities are not necessary reflected by the most relevant functioning (Agler et al., 2012b; Vanwonterghem et al., 2016; Louca et al., 2017). That is to say, the taxonomic composition of communities appears decoupled from functional composition. Then coming up with the open questions: how to completely decouple certain functions from the community assembly of metabolically overlapping microorganisms with functional redundancy, in particular of those in resource competition with the target processes, and what determines the degree of functional redundancy in microbial systems? Acquiring more knowledge on physiological, genomic and metabolic features of species themselves is of particular importance in answering these questions. In addition, functional redundancy is critical for maintaining the stability of community functioning against environmental perturbations. Hewn in stone is that functional redundancy always exists in microbial open systems (Louca et al., 2018). To be practical, promoting functional redundancy in systems dealing with organic wastes or biomass residues is an important aspect in microbial resource management.

3.3 Moving from intriguing science to real-world practice – Microbiota-based biotechnology

Ecological theories are particularly crucial to provide predictions on microbial community functioning. Making it predictable is an important step to leverage fundamental scientific principles and quantitative design to manage the communities performing desired functions. Microbiota-based biotechnologies are relevant to the fields of environmental technology, agriculture, human health, energy and many others (Figure 3.1).
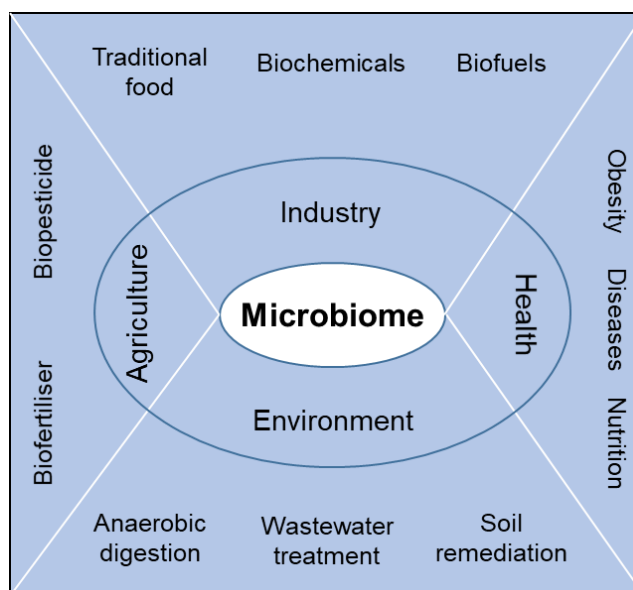

Figure 3.1 The application of microbiota. Figure was adapted from Jiang et al., 2017.

Biotechnological processes using pure culture fermentation are economically attractive due to their high titers and production rates (Angenent et al., 2020). However, considering the utilisation of waste streams or biomass residues, mixed culture fermentation is a wiser option with inherent advantages. It can be operated in nonsterile conditions to perform multiple functions with robust redundancy in the utilisation of complex substrates. Anaerobic digestion is the most successful application to date aiming at producing renewable energy in the form of methane. Now it is evolving to the production of MCCs (e.g., C6 and C8), which are ways more valuable than methane. However, the low product concentration and selectivity are still a challenge for any downstream processes.

To my knowledge, for open-culture biotechnology, manipulation of environmental conditions of the reactor microbiota is the only way for process control. In this context, process design with spatial organisation can be a promising way to reduce the complexity of reactor microbiota, aiming at steering microbial pathways towards the production of the corresponding desired products. For example, the study of Xu et al., 2018 demonstrated a temperature-phased bioreactor system in converting acid whey into MCCs via lactate. Without the addition of external electron donors, they integrated lactic acid fermentation in a first phase (thermophilic and mildly acidic conditions) with CE in a second (mesophilic conditions), showing lactate-based CE within carboxylate platform to be a promising waste stream recovery strategy. This PhD thesis also emphasised the importance of *in situ* lactate formation in the lactate-based CE process, by taking initiatives to valorise carbohydrate-rich waste streams in a maximal way.

Another example for the exploration of spatial organisation came to the defined mixed culture fermentation in a biofilm membrane reactor system (Shahab et al., 2020). By adopting strategies of synthetic ecology and process engineering, they constructed an artificial food chain with cross-kingdom microbial consortia including an aerobic fungus, lactic acid bacteria and lactate-based chain-elongating bacteria, which can convert the complex substrate lignocellulose to valuable platform chemicals such as butyrate, valerate and caproate. This groundbreaking study successfully presented the possibility of engineering stable and controllable synthetic communities with the utilisation of their ecological niches.

Despite our broad scientific interest in knowing microbiota from multiple environments, a way to synthesise all the gained knowledge into best practice would advance the microbiota-based biotechnologies to diminish our dependence on fossil resources. Given our nascent knowledge of microbial ecology at the moment, it is a long and hard way of microbiota engineering. Thus, working together with multidisciplinary experts is essential to turn the challenge into opportunities and to finally realise the dream of a circular economy.

4 References

Agler, M. T., Werner, J. J., Iten, L. B., Dekker, A., Cotta, M. A., Dien, B. S., et al. (2012). Shaping reactor microbiomes to produce the fuel precursor *n*-butyrate from pretreated cellulosic hydrolysates. *Environ. Sci. Technol.* 46, 10229–10238.

Agler, M. T., Wrenn, B. A., Zinder, S. H., and Angenent, L. T. (2011). Waste to bioproduct conversion with undefined mixed cultures: The carboxylate platform. *Trends Biotechnol.* 29, 70–78.

Andersen, S. J., de Groof, V., Khor, W. C., Roume, H., Props, R., Coma, M., et al. (2017). A *Clostridium* group IV species dominates and suppresses a mixed culture fermentation by tolerance to medium chain fatty acids products. *Front. Bioeng. Biotechnol.* 5, 8.

Angenent, L.T., Magdalena, J. A., Jeon, B. S., and Joseph G. U. (2020). Eco-mimicry opens new doors for bioprocess engineers. *Joule* 4, 2074–2077.

Angenent, L. T., Richter, H., Buckel, W., Spirito, C. M., Steinbusch, K. J. J., Plugge, C. M., et al. (2016). Chain elongation with reactor microbiomes: Open-culture biotechnology to produce biochemicals. *Environ. Sci. Technol.* 50, 2796–2810.

Auchtung, J. M., Robinson, C. D., and Britton, R. A. (2015). Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). *Microbiome* 3, 42.

Barker, H. A., Kamen, M. D., and Bornstein, B. T. (1945). The synthesis of butyric and caproic acids from ethanol and acetic acid by *Clostridium kluyveri*. *Proc. Natl. Acad. Sci. USA*. 31, 373–381.

Cabezas, A., de Araujo, J. C., Callejas, C., Galès, A., Hamelin, J., Marone, A., et al. (2015). How to use molecular biology tools for the study of the anaerobic digestion process? *Rev. Environ. Sci. Biotechnol.* 14, 555–593.

Connell, J. H., and Sousa, W. P. (1983). On the evidence needed to judge ecological stability or persistence. *Am. Nat.* 121, 789–824.

Contreras-Dávila, C. A., Carrión, V. J., Vonk, V. R., Buisman, C. N. J., and Strik, D. P. B. T. B. (2020). Consecutive lactate formation and chain elongation to reduce exogenous chemicals input in repeated-batch food waste fermentation. *Water Res.* 1, 115215.

De Leeuw, K. D., Buisman, C. J. N., and Strik, D. P. B. T. B. (2019). Branched medium chain fatty acids: *iso*-caproate formation from *iso*-butyrate broadens the product spectrum for microbial chain elongation. *Environ. Sci. Technol.* 53, 7704–7713.

Esquivel-Elizondo, S., Ba, C., Temovska, M., and Jeon, B. S. (2020). The isolate *Caproiciproducens* sp. 7D4C2 produces *n*-caproate at mildly acidic conditions from hexoses: genome and rBOX comparison with related strains and chain-elongating bacteria. Preprint. Available at: https://doi.org/10.1101/2020.07.19.210914.

Faust, K. (2019). Towards a better understanding of microbial community dynamics through high-throughput cultivation and data integration. *mSystems* 4, e00101-19.

Han, W., He, P., Shao, L., and Lü, F. (2018). Metabolic interactions of a chain elongation microbiome. *Appl. Environ. Microbiol.* 84, e01614-18.

Hekstra, D. R., and Leibler, S. (2012). Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* 149, 1164–1173.

Jeon, B. S., Choi, O., Um, Y., and Sang, B. I. (2016). Production of medium-chain carboxylic acids by *Megasphaera* sp. MH with supplemental electron acceptors. *Biotechnol. Biofuels* 9, 129.

Jeon, B. S., Kim, S., and Sang, B. I. (2017). *Megasphaera hexanoica* sp. nov., a medium-chain carboxylic acid-producing bacterium isolated from a cow rumen. *Int. J. Syst. Evol. Microbiol.* 67, 2114-2120.

Jiang, L. L., Zhou, J. J., Quan, C. S., and Xiu, Z. L. (2017). Advances in industrial microbiome based on microbial consortium for biorefinery. *Bioresour. Bioprocess.* 4, 11.

Kehe, J., Kulesa, A., Ortiz, A., Ackerman, C. M., Thakku, S. G., Sellers, D., et al. (2019). Massively parallel screening of synthetic microbial communities. *Proc. Natl. Acad. Sci. USA*. 116, 12804–12809.

Kim, B. C., Jeon, B. S., Kim, S., Kim, H., Um, Y., and Sang, B. I. (2015). *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.* 65, 4902-4908.

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018). Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422.

Koch, C., Müller, S., Harms, H., and Harnisch, F. (2014). Microbiomes in bioenergy production : From analysis to management. *Curr. Opin. Biotechnol.* 27, 65–72.

Lagier, J.-C., Khelaifi, S., Alou, M. T., Ndongo, S., Delerce, J., Vitton, V., et al. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* 1, 16203.

Lagier, J. C., Hugon, P., Khelaifia, S., Fournier, P. E., La Scola, B., and Raoult, D. (2015). The rebirth of culture in microbiology through the example of culturomics to study human gut microbiota. *Clin. Microbiol. Rev.* 28, 237–264.

Lawson, C. E., Harcombe, W. R., and Hatzenpichler, R. (2019). Common principles and best practices for engineering microbiomes. *Nat. Rev. Microbiol.* 17, 725–741.

Liao, J. C., Mi, L., Pontrelli, S., and Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat. Rev. Microbiol.* 14, 288–304.

Liu, B., Kleinsteuber, S., Centler, F., Harms, H., and Sträuber, H. (2020a). Competition between butyrate fermenters and chain-elongating bacteria limits the efficiency of medium-chain carboxylate production. *Front. Microbiol.* 11, 336.

Liu, B., Popp, D., Sträuber, H., Harms, H., Kleinsteuber, S. (2020b). Draft genome sequences of three *Clostridia* isolates involved in lactate-based chain elongation. *Microbiol. Resour. Announc.* 9, e00679-20.

Louca, S., Jacques, S. M. S., Pires, A. P. F., Leal, J. S., Srivastava, D. S., Parfrey, L. W., et al. (2017). High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* 1, 0015.

Louca, S., Polz, M. F., Mazel, F., Albright, M. B. N., Huber, J. A., O'Connor, M. I., et al. (2018). Function and functional redundancy in microbial systems. *Nat. Ecol. Evol.* 2, 936–943.

Marchesi, J. R., and Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome* 3, 1–3.

Marounek, M., Fliegrova, K., and Bartos, S. (1989). Metabolism and some characteristics of ruminal strains of *Megasphaera elsdenii*. *Appl. Environ. Microbiol.* 6, 1570-3.

Ning, D., Deng, Y., Tiedje, J. M., and Zhou, J. (2019). A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl. Acad. Sci. USA.* 116, 16892–16898.

Prosser, J. I., Bohannan, B. J. M. B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., et al. (2007). The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* 5, 384–392.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844.

Royce, L. A., Liu, P., Stebbins, M. J., Hanson, B. C., and Jarboe, L. R. (2013). The damaging effects of short chain fatty acids on *Escherichia coli* membranes. *Appl. Microbiol. Biotechnol.* 97, 8317–8327.

Scarborough, M. J. (2019). Harnessing microbiomes to close the carbon cycle: production of beneficial chemicals from complex renewable feedstocks. *Ph.D.* Thesis, University of Wisconsin-Madison, U.S.A.

Scarborough, M. J, Myers, K. S., Donohue, T. J., and Noguera, D. (2020). Medium-chain fatty acid synthesis by "*Candidatus* Weimeria bifida" gen. nov., sp. nov., and "*Candidatus* Pseudoramibacter fermentans" sp. nov. *Appl. Environ. Microbiol.* 86, e02242-19.

Shahab, R. L., Brethauer, S., Davey, M. P., Smith, A. G., Vignolini, S., Luterbacher, J. S., et al. (2020). A heterogeneous microbial consortium producing short-chain fatty acids from lignocellulose. *Science* 369, eabb1214.

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., et al. (2017). DNA sequencing at 40: Past, present and future. *Nature* 550, 345–353.

Spirito, C. M., Richter, H., Rabaey, K., Stams, A. J. M., and Angenent, L. T. (2014). Chain elongation in anaerobic reactor microbiomes to recover resources from waste. *Curr. Opin. Biotechnol.* 27, 115–122.

Stams, A. J. M., and Plugge, C. M. (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat. Rev. Microbiol.* 7, 568–577.

Talbot, G., Topp, E., Palin, M. F., and Massé, D. I. (2008). Evaluation of molecular methods used for establishing the interactions and functions of microorganisms in anaerobic bioreactors. *Water Res.* 42, 513–537.

Tao, Y., Zhu, X., Wang, H., Wang, Y., Li, X., Jin, H., et al. (2017). Complete genome sequence of *Ruminococcaceae* bacterium CPB6: A newly isolated culture for efficient *n*-caproic acid production from lactate. J. Biotechnol. 259, 91–94.

Vanwonterghem, I., Jensen, P. D., Rabaey, K., and Tyson, G. W. (2016). Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion. *Environ. Microbiol.* 18, 3144–3158.

Verstraete, W., Wittebolle, L., Heylen, K., Vanparys, B., de Vos, P., van de Wiele, T., et al. (2007). Microbial Resource Management: The road to go for environmental biotechnology. *Eng. Life Sci.* 2, 117–126.

Vrancken, G., Gregory, A. C., Huys, G. R. B., Faust, K., and Raes, J. (2019). Synthetic ecology of the human gut microbiota. *Nat. Rev. Microbiol.* 17, 754–763.

Willems, A., and Collins, M. (1996). Phylogenetic relationships of the genera *Acetobacterium* and *Eubacterium sensu stricto* and reclassification of

Eubacterium alactolyticum as *Pseudoramibacter alactolyticus* gen. nov., comb. nov. *Appl. Environ. Microbiol.* 46, 1083-7.

Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., et al. (2009). Initial community evenness favours functionality under selective stress. *Nature* 458, 623–626.

Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., et al. (2019). Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat. Microbiol.* 4, 1183–1195.

Xu, J., Hao, J., Guzman, J. J. L., Spirito, C. M., Harroff, L. A., and Angenent, L. T. (2018). Temperature-phased conversion of acid whey waste into medium-chain carboxylic acids via lactic acid: no external e-donor. *Joule* 2, 1–16.

Yang, J., Hou, X., Mir, P. S., and McAllister, T. A. (2010). Anti-*Escherichia coli* O157:H7 activity of free fatty acids under varying pH. *Can. J. Microbiol.* 56, 263–267.

Zengler, K., and Zaramela, L. S. (2018). The social network of microorganisms - How auxotrophies shape complex communities. *Nat. Rev. Microbiol.* 16, 383–390.

Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* 81, e00002-17.

Zhu, X., Tao, Y., Liang, C., Li, X., Wei, N., Zhang, W., et al. (2015). The synthesis of *n*-caproate from lactate: a new efficient process for medium-chain carboxylates production. *Sci. Rep.* 5, 14360.

Zhu, X., Zhou, Y., Wang, Y., Wu, T., Li, X., Li, D., et al. (2017). Production of high-concentration *n*-caproic acid from lactate through fermentation using a newly isolated *Ruminococcaceae* bacterium CPB6. *Biotechnol. Biofuels* 10, 102.

5 Appendix

5.1 Declaration of authorship

I, Bin Liu, hereby declare that:

- I have written this thesis autonomously with my own ideas and judgements. No other sources than the stated quotations from other works were used. Full reference if their work have been provided in the proper way.

- All persons involved as co-authors for the material for my thesis are stated including their respective contribution (Chapter 5.2).

- No other persons than stated above were involved in the preparation of the thesis, in particular, no PhD consultants were used, and no third party has received direct or indirect financial benefits in goods and services for a contribution to this thesis.

- This thesis in an equal or similar form has not been submitted to an academic institution to obtain a doctoral degree or any other academic degree, and has not been published yet.

- No further doctoral examination process has taken or is taking place.

- I am the exclusive author of the chapter 2.3 "Effects of pH increase on microbial chain elongation and community dynamics in closed bioreactor ecosystems".

Bin Liu

Leipzig, 12 November 2020

5.2 Coauthor contributions

Author contribution statement, Bin Liu
Applied microbial ecology of anaerobic reactor microbiomes

Author contribution statement:

Title: Competition between butyrate fermenters and chain-elongating bacteria limits the efficiency of medium-chain carboxylate production

Journal: Frontiers in Microbiology

Authors: Bin Liu, Sabine Kleinsteuber, Florian Centler, Hauke Harms, Heike Sträuber

Bin Liu:

- study and experiment design
- data analysis and interpretation
- reactor operation
- T-RFLP
- amplicon sequencing
- writing of manuscript

Sabine Kleinsteuber:

- study and experiment design
- revision of manuscript

Florian Centler:

- network analysis
- revision of manuscript

Hauke Harms:

- revision of manuscript

Heike Sträuber (senior author):

- study and experiment design
- revision of manuscript

Bin Liu

Heike Sträuber

Author contribution statement:

Title: **Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture**

Journal: Submitted to Microbiome

Authors: Bin Liu, Heike Sträuber, João Saraiva, Hauke Harms, Sandra Godinho Silva, Sabine Kleinsteuber, Ulisses Nunes da Rocha

Bin Liu:

- study and experiment design
- data analysis and interpretation
- reactor operation
- machine learning analysis
- amplicon sequencing
- writing of manuscript

Heike Sträuber:

- study and experiment design
- data interpretation
- revision of manuscript

João Saraiva:

- study and experiment design
- data interpretation
- machine learning analysis
- revision of manuscript

Hauke Harms:

- revision of manuscript
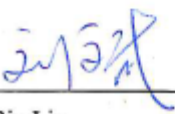
Sandra Godinho Silva:

- data analysis and interpretation
- revision of manuscript

Ulisses Nunes da Rocha:

- study and experiment design
- data interpretation
- machine learning analysis
- revision of manuscript

Sabine Kleinsteuber (senior author):

- study and experiment design
- data interpretation
- revision of manuscript

Bin Liu

Sabine Kleinsteuber

Author contribution statement:

Title: **Draft genome sequences of three *Clostridia* isolates involved in lactate-based chain elongation**

Journal: Microbiology Resource Announcements

Authors: Bin Liu, Denny Popp, Heike Sträuber, Hauke Harms, Sabine Kleinsteuber

Bin Liu:

- study and experiment design
- strain isolation
- genome sequencing and assembly
- data interpretation
- writing of manuscript

Denny Popp:

- study and experiment design
- genome sequencing and assembly
- data interpretation
- revision of manuscript

Heike Sträuber:

- study and experiment design
- data interpretation
- revision of manuscript

Hauke Harms:

- revision of manuscript

Sabine Kleinsteuber (senior author):

- study and experiment design
- data interpretation
- revision of manuscript

Bin Liu

Sabine Kleinsteuber

Author contribution statement:

Title: **Three novel *Clostridia* isolates produce *n*-caproate and *iso*-butyrate from lactate: comparative genomics of chain-elongating bacteria**

Journal: Submitted to Microorganisms

Authors: Bin Liu, Denny Popp, Nicolai Müller, Heike Sträuber, Hauke Harms, Sabine Kleinsteuber

Bin Liu:

- study and experiment design
- strain isolation and physiological characterization
- genome sequencing and assembly
- data analysis and interpretation
- writing of manuscript

Denny Popp:

- study and experiment design
- genome sequencing and assembly
- data analysis and interpretation
- revision of manuscript

Nicolai Müller:

- study and experiment design
- data analysis and interpretation
- revision of manuscript

Heike Sträuber:

- study and experiment design
- data analysis and interpretation
- revision of manuscript

Hauke Harms:

- revision of manuscript

Sabine Kleinsteuber (senior author):

- study and experiment design
- data analysis and interpretation
- revision of manuscript

Bin Liu   Sabine Kleinsteuber

5.3 Curriculum Vitae

5.4 List of publications and conference contributions

Publications

Liu, B., Sträuber, H., Saravia, J.P., Harms, H., Sandra, G.S., Kleinsteuber, S. # and da Rocha, U. N. # (2020). Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture. *Submitted*; (#Corresponding authors)

Liu, B., Popp, D., Sträuber, H., Harms, H. and Kleinsteuber, S. # (2020). Lactate-based microbial chain elongation for *n*-caproate and *iso*-butyrate production: genomic and metabolic features of three novel *Clostridia* isolates. *Submitted*

Liu, B., Popp, D., Sträuber, H., Harms, H. and Kleinsteuber, S. # (2020). Draft genome sequences of three Clostridia isolates involved in lactate-based chain elongation. *Microbiology Resource Announcements* Vol. 9, No. e00679-20.

Liu, B., Kleinsteuber, S., Centler, F., Harms, H. and Sträuber, H. # (2020). Competition between butyrate fermenters and chain-elongating bacteria limits the efficiency of medium-chain carboxylate production. *Frontiers in Microbiology*, Vol. 11, No. 336.

The following articles, not included in this thesis, also were published during my PhD study:

Lian, S., Nikolausz, M., Nijenhuis, I., da Rocha, U. N., **Liu, B.,** Corrêa, F.B., Saraiva, J.P. and Richnow, H.H. (2020). Biotransformation of hexachlorocyclohexanes contaminated biomass for energetic utilization demonstrated in continuous anaerobic digestion system. *Journal of Hazardous Materials*, Vol. 384, No. 121448.

Meng, X. #, **Liu, B. #,** Zhang, H., Wu, J., Yuan, X. and Cui Z. (2019). Co-composting of the biogas residues and spent mushroom substrate: physicochemical properties and maturity assessment. *Bioresource Technology*, Vol. 276, pp. 281–287. (# These authors contributed equally to the work)

Meng, X., **Liu, B.,** Chen X., Luo, X., Yuan, X., Wang, X., Zhu, W., Wang, H. and Cui, Z. (2018). Effect of pig manure on the chemical composition and microbial diversity during co-composting with spent mushroom substrate and rice husks. *Bioresource Technology*, Vol. 251, pp. 22-30.

Conference contributions

10/2020, "Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture", International Chain Elongation Conference 2020 – virtual, organised by Wageningen University & Research. (oral presentation)

03/2020, "Lactate-based microbial chain elongation for *n*-caproate production: genomic and metabolic features of three novel *Clostridiales* isolates", 6th Joint Conference of the DGHM & VAAM, Leipzig, Germany. (poster)

07/2019, "Resilience of chain-elongating reactor microbiota upon temperature perturbation", International Workshop on Valorization of Agricultural Residues via Anaerobic Digestion: from Biogas to Carboxylates, DBFZ, Leipzig, Germany. (oral presentation)

06/2019, "Resilience of chain-elongating reactor microbiota upon temperature perturbation", 16th IWA World Congress on Anaerobic Digestion, Delft, The Netherlands. (oral presentation)

05/2019, "The best bet to boost reactor microbiomes not being lazy", HIGRADE Conference 2019, UFZ – PhD Conference, Leipzig, Germany. (poster)

05/2019, "The best bet to boost reactor microbiomes not being lazy", 31st Chinese-German Chemical Association Annual Conference, Leipzig, Germany. (poster)

08/2018, "Dynamics of a chain-elongating reactor microbiome producing medium-chain fatty acids from lactic acid and xylan", 17th International Symposium on Microbial Ecology (ISME17), Leipzig, Germany. (poster)

06/2018, "Chain elongation with lactate producing medium-chain carboxylates by a semi-continuously fed anaerobic reactor microbiome", 2nd International Conference on Anaerobic Digestion Technology, Chiang Mai, Thailand. (oral presentation)

04/2018, "Dynamics of a chain elongating reactor microbiome producing medium-chain carboxylates", Annual Conference 2018 of the Association for General and Applied Microbiology (VAAM), Wolfsburg, Germany. (poster)

10/2017, "Chain elongation with lactate producing medium-chain carboxylates by a semi-continuously fed anaerobic reactor microbiome", 2017 Biotechnology Symposium – Innovative Molecular Compounds, Leipzig University (BIOCITY), Germany. (poster)

08/2015, "A critical insight of the dynamics in biogas reactor microbial communities", 9th China New Energy International Forum – biomass section, Beijing, China. (poster)

5.5 Acknowledgements

My PhD is not the effort of one person, but with so many contributors involved. Surely there can be an exhaustive list of people who helped me during these four years. To those not mentioned by name below, I appreciate you.

First, I would like to thank my supervisors Prof. Dr. Hauke Harms, Dr. Sabine Kleinsteuber and Dr. Heike Sträuber. From my application for CSC scholarship to the arrival in Leipzig and until this moment, all of you have given me huge support. Your supervision, encouragement and patience helped me to build the confidence to be good in science.

Many thanks to my colleagues and former members of the group of Microbiology of Anaerobic Systems in the Department of Environmental Microbiology (UMB). It has been my pleasure to work with you. Sincere thanks to colleagues of the research groups "Systems Biology of Microbial Communities", headed by Dr. Florian Centler, and "Microbial Data Science", headed by Dr. Ulisses Nunes da Rocha, and I do enjoy all the discussions and collaborations. We surely need to keep in touch in the future. Moreover, I want to give my gratitude to Dr. Denny Popp, who is always patient and helpful.

Special thanks to the colleagues from DBFZ for their support in the lab.

Last, to my parents and friends who have always believed in me and encouraged me to stick to my dreams.