MANUSCRIPT | WATER RESEARCH

Predicting the performance of chain elongating microbiomes through flow cytometric fingerprinting

Kevin Sabbe[1,2], Liese D'Haen[1], Nico Boon[1,2,#], Ramon Ganigué[1,2,#,*]

[1] Center for Microbial Ecology and Technology (CMET), Ghent University, Coupure Links 653, 9000 Ghent, Belgium

[2] Center for Advanced Process Technology for Urban Resource Recovery (CAPTURE), Frieda Saeysstraat 1, 9052 Ghent, Belgium

[#] Joint senior authors

* Correspondence to: Ramon Ganigué, Ghent University; Faculty of Bioscience Engineering; Center for Microbial Ecology and Technology (CMET); Coupure Links 653; 9000 Gent, Belgium; phone: +32 (0)9 264 59 76; fax: +32 (0)9 264 62 48; E-mail: Ramon.Ganigue@UGent.be; Webpage: www.cmet.UGent.be.

Abstract

As part of the circular bio-economy paradigm shift, waste management and valorisation practices have moved away from sanitation and towards the production of added-value compounds. Recently, the development of mixed culture bioprocess for the conversion of waste(water) to platform chemicals, such as medium chain carboxylic acids, has attracted significant interest. Often, the microbiology of these novel bioprocesses is less diverse and more prone to disturbances, which can lead to process failure. This issue can be tackled by implementing an advanced monitoring strategy based on the microbiology of the process. In this study, flow cytometry was used to monitor the microbiology of lactic acid chain elongation for the production of caproic acid, and assess its performance both qualitatively and quantitatively. Two continuous stirred tank reactors for chain elongation were monitored flow cytometrically for over 336 days. Through community typing, four specific community types could be identified and correlated to both a specific functionality and genotypic diversity. Additionally, the machine-learning algorithms trained in this study demonstrated the ability to predict production rates of, amongst others, caproic acid with high accuracy in the present ($R^2 > 0.87$) and intermediate accuracy in the near future ($R^2 > 0.63$). The identification of specific community types and the development of predictive algorithms form the basis of advanced bioprocess monitoring based on flow cytometry, and have the potential to improve bioprocess control and optimization, leading to better product quality and yields.

Keywords: Caproic acid, machine learning, mixed culture fermentation, monitoring, resource recovery

Highlights:

- Flow cytometry enables real-time and near-future bioprocess performance assessment
- Chain elongation community was translated into unique cytometric fingerprints
- Based on cytometric fingerprints, several community types could be identified
- Each community type reflects a specific reactor performance and community composition
- Machine learning algorithms predict production rates of key fermentation products

1 Introduction

The valorisation of organic waste into added-value bioproducts is key to enable a circular bio-economy (Crognale et al., 2021; Stegmann et al., 2020). Currently, organic waste is mainly valorised as a green energy alternative (*i.e.* biogas) and/or as a soil conditioner (*i.e.* compost) (De Buck et al., 2020; De Groof et al., 2019). However, these products have a rather low economic value (€ 2/tonne food waste as compost, and € 76/tonne organic fraction of municipal solid waste (OFMSW) as biogas), while biogas production through classical anaerobic digestion is a heavily subsidized technology and is becoming less economically sustainable due to decreasing governmental support (De Groof et al., 2019; Kleerebezem et al., 2015; Spirito et al., 2014). Novel bioprocesses for organic carbon recovery are being developed to broaden the spectrum of products that can be generated from waste, as well as the economic revenue derived from them.

Microbial chain elongation for the production of medium-chain carboxylic acids (MCCA) is one of these bioprocesses under development (Angenent et al., 2016; Kang et al., 2022; Kleerebezem et al., 2015; Nzeteu et al., 2022). Compared to the use of biogas, which only finds applications in energy production, MCCA can be used in a myriad of applications. For instance caproic acid (CA), a six-carbon organic acid, can be used as a building block in the chemical industry, as a precursor of bio-fuels, as bio-plasticizer, in flavourings, pharmaceuticals, and as feed and food additive (Angenent et al., 2016; Cavalcante et al., 2017; Spirito et al., 2014). In chain elongation, an electron donor (e.g. lactic acid (LA), ethanol, glucose) is oxidized to acetyl-CoA, which can be combined with an electron acceptor (e.g. acetic acid (AA), butyric acid (BA)) via the reverse beta-oxidation pathway, elongating the carbon chain of the electron acceptor with two carbon atoms (Brodowski et al., 2022; O-Thong et al., 2020; Spirito et al., 2014). Initially, most studies mainly focused on the use of ethanol as electron donor for chain elongation (Grootscholten et al., 2014, 2013; Roghair et al., 2018; Steinbusch et al., 2011). It

has been shown, however, that ethanol supplementation is economically and environmentally unsustainable (Carvajal-Arroyo et al., 2019; Chen et al., 2017b). Therefore, there has been an increased interest in the use of LA as electron donor. LA can easily be produced from carbohydrate-rich waste, such as OFMSW, making it an ideal feedstock for producing MCCA through LA chain elongation (Kucek et al., 2016; Nzeteu et al., 2018).

Most well-established open-culture bioprocesses like anaerobic digestion are quite resilient against disturbances. Bioprocesses like chain elongation, however, often have a less diverse, niche microbiome with a specific functionality that can make these processes more vulnerable to process failure (Agler et al., 2012; Andersen et al., 2017; Brodowski et al., 2022; De Groof et al., 2020; Joshi et al., 2021; Spirito et al., 2018; Steinbusch et al., 2011). Currently, most bioproduction processes are monitored by measuring physicochemical parameters such as pH, product spectra and concentration, biomass concentrations, etc. However, the process microbiome is the catalyst that determines the process output. The microbial community of mixed culture bioprocesses is not often monitored at full scale, and most information on how certain operational conditions and feedstocks affect the microbial community and its performance is derived from DNA sequencing to monitor changes in the genotypic composition of the microbiome and link them to its performance. For instance, Liu et al. (2022) adopted a machine learning approach for predicting chain elongation performance based on sequencing data and Wilson et al. (2021) identified recurrent marine community types (CTs). However, sequencing is still a rather expensive and time-consuming technique that requires weeks to obtain and interpret the results, and fast phenotypic changes of the microbiology cannot be detected at relevant time-scales for mixed-culture bioprocess control (García et al., 2015; Heyse et al., 2021).

Flow cytometry (FC) is a fast and cheap optical single-cell analysis technique for the assessment of the phenotypic diversity of a microbiome within a minutes to hours time-frame (Heyse et al., 2021; Park et al., 2005; Props et al., 2016). It has, furthermore, the potential to be implemented in-line for real-time monitoring (Abu-Absi et al., 2003; Favere et al., 2021, 2020; Hammes et al., 2012). Until now, it has mainly been implemented for monitoring aqueous environments that contain a low microbial abundance and nutrients, such as drinking water facilities (Favere et al., 2021, 2020; Hammes et al., 2012; Prest et al., 2013; Props et al., 2018). More recently, the use of FC is emerging for the detection of disturbances through FC fingerprinting (Props et al., 2018; Rubbens et al., 2021). The application potential of FC has been demonstrated in fermentation processes as well. In pure culture fermentations, for example, FC has been applied to quantify cell densities and to study the morphology and integrity of *Lactobacillus acidophilus* in wine (Narayana et al., 2020; Salma et al., 2013). The application of FC for open fermentation processes and the interpretation of the data can be more challenging due to the high diversity and complexity of the matrix. For monitoring biogas production and waste water treatment, phenotypic sub-communities could be correlated with functionality (Günther et al., 2018, 2012; Koch et al., 2013). In the field of chain elongation, Duber et al. (2018) used FC to assess the metabolic activity of the microbiome based on the cellular redox potential by distinguishing between active, mid-active, and inactive cells.

When dealing with complex data from complex environments, community typing has been proposed as an approach to differentiate between microbiomes with a specific composition and link it with a certain functionality or event. CTs can be based on either the taxonomic composition of the microbiome (e.g. enterotypes in the gut microbiome, in environmental samples, etc.) or the phenotypic composition of the microbiome (e.g. event detection in drinking water systems) (Arumugam et al., 2011; Gabrielli et al., 2021; Props et al., 2018; Wilson et al., 2021). Additionally, flow cytometric fingerprints as a whole (be it in the grid or

Gaussian mixture model (GMM) counts matrix or abstracted metrics thereof, such as Hill-number diversity) could be potentially correlated with specific process parameters to link changes in community structure with process performance. It has been shown that the microbial diversity derived from flow cytometric data correlates well with diversity determined through 16S rRNA sequencing, and that the fingerprint can be used for the predictive modelling of the presence and abundance of certain bacterial taxa via a machine learning approach (Heyse et al., 2021; Props et al., 2016).

In this study we present the use of FC as a tool to monitor LA chain elongation performance and aim to show that: i) FC can be applied as tool for early event detection in a mixed-culture fermentation process via phenotypic community typing; and ii) FC fingerprinting data can be used for the predictive modelling of main process performance indicators (e.g. production rates).

2 Materials and methods

2.1 Reactor operation and sampling

Two continuous stirred tank reactors (CSTR), R1 and R2, were inoculated with a pre-adapted LA chain elongating culture coming from previous experiments (Candry et al., 2020). A synthetic medium was used throughout the entire course of the experiment. The medium composition was based on the medium used by Candry et al. (2020) (Table A.1) but the organic carbon sources were adjusted to 231 mM LA, 58.7 mM AA, 7.46 mM propionic acid (PA), 8.93 mM BA and 38.2 mM ethanol, to mimic the carboxylic acid spectrum of the OFMSW. Two multi-port reactors (OCHS Glasgerätebau, Bovenden, Germany) with a working volume of 1000 mL were operated continuously with magnetic stirring. The reactors were operated in a temperature-controlled room at 34 °C at a hydraulic retention time (HRT) of 4 days, except for the start-up phases and the experimental periods where the HRT was temporarily decreased

to 2 days. Since there was no biomass retention, the sludge retention time (SRT) equalled the HRT. The reactors were operated at pH 5.5, controlled with in-line pH controllers (Prominent, Belgium) dosing 2 M HCl. The reactors were sampled every 2 to 3 days for monitoring the biomass concentration and phenotypic composition though FC, along with the concentrations of the organic acids. Sample preparation for flow cytometric measurements was performed immediately after reactor sampling. Besides these unfiltered samples, additional samples were filtered (0.22 µm) and stored at -20 °C for the analysis of the concentration of substrates and products in the reactors. Additionally, samples for 16S rRNA gene amplicon sequencing were prepared by transferring 1 mL of unfiltered reactor sample to 2 mL Micrewtubes® (Simport, Canada) in duplicate. The samples were centrifuged for 5 min at 20.817 g, the supernatant was removed, and the remaining pellets were stored at -20 °C until DNA extraction. At each sampling point, acid consumption and gas production rate were monitored, and the reactor headspace was sampled for the determination of the headspace composition. Throughout the reactor operation, several events occurred, and operational changes were applied that can be considered as disturbances. An overview of these events is shown in Table 1.

Table 1: Overview of the operational periods with the events/disturbances

	Day	Event
Period 1	0	Start up
	49	Increase LA in feed from 12.5 to 20.8 g/L
	65	R1 acidified (malfunctioning stirrer)
Period 2	79	Reinoculate R1 with R2
	180	Lower HRT of R1 from 4 to 2 days
	189	Increased HRT of R1 from 2 to 4 days
	205	R2 acidified (malfunctioning stirrer)
Period 3	224	Mixed R1 and R2
	229	Lowered HRT of both reactors from 4 to 2 days
	257	Increased HRT of both reactors from 2 to 4 days
	273	Reinoculation of R2
	336	End of experiment

2.2 Flow cytometric measurements

Cell concentrations and characteristics were determined via FC. To remove potential aggregates and allow single-cell analysis, 1 mL of the undiluted, unfiltered reactor samples were filtered with 20 µm syringe filters (Filcon, BD Biosciences, Belgium). Subsequently, 1000× dilutions were prepared in phosphate-buffered saline (PBS) in triplicates in 96-well plates, with dilution steps of 10-fold and a final volume of 200 µL. The technical replicates were stained with 1 v/v% SYBR Green I (SG) (100× concentrate in 0.22-µm-filtered dimethyl sulfoxide). After staining, the samples were incubated in the dark at 37 °C for 20 minutes (Props et al., 2016). After incubation, the samples were analysed for 60 seconds or until a total of 50 000 events were measured (ca. 50 µL) with a FACSVerse flow cytometer (BD Biosciences, Belgium) equipped with eight fluorescence detectors (527/32, 783/56, 586/42, 700/54, 660/10, 783/56, 528/45, and 488/45 nm), two scatter detectors, and a blue 20-mW 488-nm laser, a red 40-mW 640-nm laser, and a violet 40-mW 405-nm laser. The flow cytometer was operated with FACSFlow solution (BD Biosciences) as sheath fluid. Instrument performance was verified daily using FACSuite CS&T beads (BD Biosciences).

2.3 Flow cytometry data analysis

2.3.1 Fingerprinting via Gaussian Mixture Modelling and diversity analysis

FC diversity analysis and statistical analysis were performed in RStudio using R (v4.1.0). Flow data was imported via the flowCore package (v2.4.0) and transformed via the arcsine hyperbolic function (Prest et al., 2013). Gates were manually defined on the green (FITC-A) and red (PerCP-Cy5.5-A) fluorescent channels to distinguish background events from the other events. The flow data was cleaned via the FCS_clean() function from the Phenoflow package (v1.1.2), that denoises the flowset using the flowAI package (v1.22.0) by removing events/outliers from the flow data that do not meet the criteria for flow rate, signal acquisition

and dynamic range (Monaco et al., 2016; Props et al., 2016). The data was normalized by dividing the value of the used channels by the maximal value of the green fluorescence signal.

GMMs were generated with the PhenoGMM() function of the Phenoflow package (v1.2.2). The models were generated from the normalized flowdata, downsampled to a sample size of 2000 events. By using the auto_nG argument of the PhenoGMM() function, series of models were generated with an increasing number of mixtures. Based on the Bayesian information criterion (BIC), the models with the optimal number of mixtures were retained (Rubbens et al., 2021). The GMM models of the individual reactors were generated with the determination of the optimal number of mixtures between 5 and 100 with increments of 5, based on the FITC-A, PerCP-Cy5.5-A, SCA-A and FSC-A channels, and were applied to the flowdata with the PhenoMaskGMM() function of Phenoflow (v1.2.2) to determine the events per mixture for each sample. The average number of events of the technical replicates was calculated for each mixture. The Bray-Curtis dissimilarities between samples was determined with the vegdist() function of the vegan package (v2.5.7) and a principle coordinate analysis (PCoA) was performed via the cmdscale() function of the R Stats package (v4.1.1) (Oksanen et al., 2022).

2.3.2 Community typing

Based on the ordination of the samples in a PCoA, samples were clustered. In order to identify CTs based on the phenotypic structure of the microbiome, samples were clustered through k-medoids clustering (Props et al., 2018; Reynolds et al., 2006). Clustering was repeated for 100 bootstrap samples. With each bootstrap, flow cytometric samples were resampled with replacement and rarefying (FCS_resample(), Phenoflow (v1.2.2)). The BC-dissimilarity matrix was obtained and a PCoA analysis performed as described in 3.3.1. The eigenvectors of the PCoA analysis explaining minimal 90% of the variance were determined and retained for further analysis, and k-Medoids clustering was performed with the pam() function (cluster

207 package, v2.1.2) for k clusters ranging between 2 to n-1, where n is the number of samples. For
208 each k, the average silhouette index was determined, and the number of clusters where the
209 average silhouette index was maximal, was retained. Samples were assigned to the most
210 frequent cluster (= CT) that it was assigned to. With the anosim() function of the vegan
211 package (v2.5.7), the ANOSIM statistic "R" was calculated to test whether there was a
212 significant difference between different CTs (Oksanen et al., 2022). Differences between CTs in
213 terms of production rates were assessed via the Kruskal-Wallis rank sum test (stats package,
214 v4.2.1), and pairwise comparisons between the CTs were assessed with the Dunn-test (FSA
215 package, v0.9.3).

216 ### *2.3.3 Training Predictive algorithm: Fingerprint as measure for productivity prediction*

217 Since the production rates of the different intermediates are directly correlated to the activity
218 of the microbiology, a model was trained for the prediction of production rates of the different
219 organic acids based on the relative average cell abundance for each GMM mixture for each
220 timepoint. The data set containing the abundances and calculated production rates was
221 resampled via 5-fold nested cross-validation. The folds were created via the createFolds()
222 function of the caret package (v6.0.90) (Kuhn, 2022). In each fold, 20% of the data was set
223 aside as test set for model validation, whereas the other 80% of the data was used to train the
224 model. Each data point was included in the test set of only one of the five folds. For each fold,
225 a random forest model was trained with the training set through 3 repeats of a 5-fold cross-
226 validation (trainControl() and train() functions of caret package (v6.0.90)). Finally, the models
227 were validated by predicting the production rate of the test set (predict() function of caret
228 package). The model performances were evaluated using the R^2 statistic, which is a value
229 between 0 and 1 that explains the proportion of the variance in the production rate that can
230 be predicted from FC data, and the mean average error (MAE), which is the average deviation
231 between the true and the predicted production rates.

2.4 Amplicon sequencing and 16S rRNA data analysis

The genotypic diversity of the reactor microbiome was assessed using samples stored at -20 °C for DNA analysis. The DNA extraction and quality control involving amplification of the 16S rRNA gene V3-V4 hypervariable regions was performed following the protocol described by Van Landuyt et al. (2022). Library preparation and sequencing was performed at LGC genomics GmbH (Berlin, Germany) on an Illumina Miseq platform with V3 chemistry. Bacterial sequencing data was analysed using R (version 4.0.3) with the methodology described by Van Landuyt et al. (2022).

2.5 Analytical methods

LA, formic acid (FA) and AA were analysed through ion chromatography. Prior to analysis, filtered (0.22 µm) samples were diluted appropriately with distilled water. The ion chromatograph (930 Compact IC Flex, Metrohm, Switzerland) was equipped with a Metrosep organic acids 250/7.8 column, a Metrosep organic acids guard column/4.6 and an 850 IC conductivity detector (Metrohm, Switzerland). The anions were eluted with 1 mM H_2SO_4 at a flow rate of 0.5 mL·min^{-1}. Carboxylic acids (C3-C8, including isoforms of C4-C6) were analysed through gas chromatography. Prior to analysis, filtered (0.22 µm) samples were diluted appropriately with distilled water. 2 mL diluted sample was conditioned with 0.5 mL H_2SO_4 (50%), 400 mg sodium chloride, and 2-methyl hexanoic acids as internal standard for quantification, prior to extraction with 2 mL diethyl ether. The gas chromatograph (GC-2014, Shimadzu®, the Netherlands) was equipped with a flame ionization detector (FID), and a DB-FFAP 123-3232 column (30m x 0.32 mm x 0.25 µm, Agilent, Belgium). The sample (1 µL) was injected at 200 °C, with a split ration of 60 and a purge flow of 3 mL·min^{-1}. The temperature in the column increased from 110 to 165 °C by 6 °C·min^{-1} where it was kept for 2 minutes. The FID had a temperature of 220 °C. Nitrogen was used as carrier gas at a flow rate of 2.49 mL·min^{-1}. The headspace gas sample composition was analysed with a Compact GC4.0 (Global

Analyser Solutions, Breda, The Netherlands), equipped with a Molsieve 5A pre-column and Porabond Q column (CH_4, O_2, H2 and N_2) and a Rt-Q-bond pre-column and column (CO_2, N_2O and H_2S). Concentrations of gases were determined by means of a thermal conductivity detector.

2.6 Calculations

The selectivity of the different carboxylic acids that were produced, was calculated by dividing the net produced concentration of the specific carboxylic acid COD, by the sum of all net produced carboxylic acid COD (eq. (1))

$$Carboxylic\ acid\ selectivity\ (\%\ COD) = \frac{Net\ carboxylic\ acid\ (g\ COD \cdot L^{-1})}{\sum Net\ carboxylic\ acids\ (g\ COD \cdot L^{-1})} \times 100 \quad (1)$$

3 Results

3.1 Chain elongation performance

Two LA chain elongating CSTRs, i.e., R1 and R2 respectively, were operated simultaneously for 336 days at an HRT of 4 days, except for the start-up phases and the experimental periods where HRT was temporarily decreased. During the start-up phase (Period 1, days 0-79), the product spectrum (Figure 1) and headspace composition (Figure A.1) were showing constant variation. From days 0-49, both reactors were fed with a synthetic medium with a reduced LA concentration of 13.30 g COD·L^{-1}. During that period, both reactors produced mostly BA (R1: 8.70 ± 5.50 g COD·L^{-1}, R2: 4.25 ± 1.32 g COD·L^{-1}) and AA (R1: 4.45 ± 2.40 g COD·L^{-1}, R2: 7.64 ± 3.33 g COD·L^{-1}), but a gradual increase of the iso-butyric acid (IBA) concentration was observed over time (max. 5.41 g COD·L^{-1} on day 28 for R1, and 6.55 g COD·L^{-1} on day 47 for R2). From day 49 on, synthetic medium as described in Table A.1 was used, with a LA concentration in the feed to 22.23 g COD·L^{-1}, which led to a decrease in the IBA concentration of both reactors. CA became the dominant product in R2 with concentrations of up to 18.56 g COD·L^{-1}. A similar initial increase in CA was observed in R1, but unfortunately on day 65, the stirring of R1

malfunctioned, causing an overdosage of HCl. No further LA consumption was observed, indicating the loss of activity of R1.

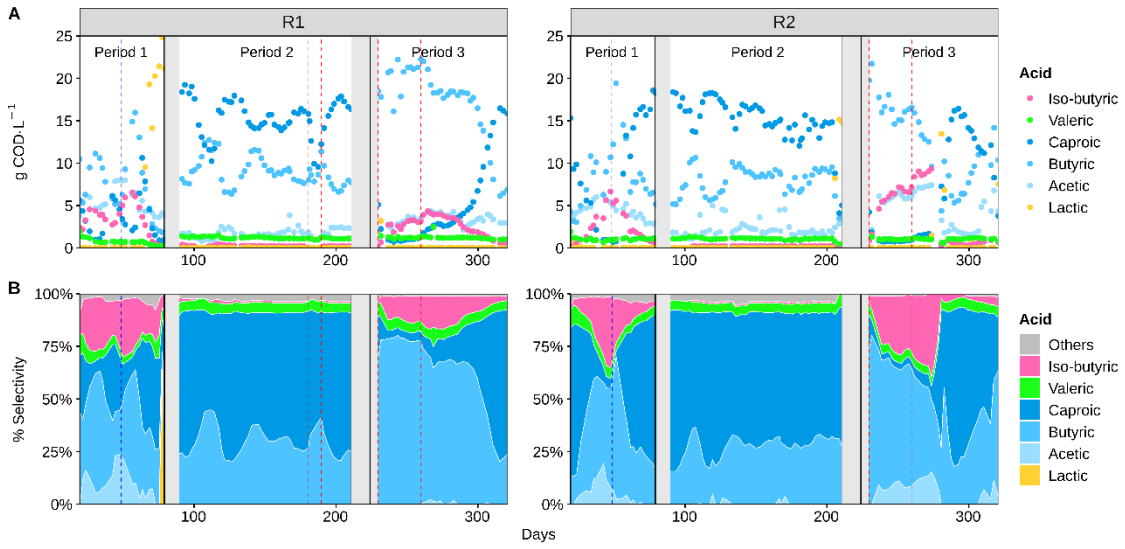


Figure 1: Product concentrations (A) and selectivity (B) of R1 and R2 over time. The black lines indicate the different periods, the blue dashed line indicates the increase in LA in the feed, the red dashed lines indicate a change of the HRT. The 'Others' group encompasses formic, propionic, iso-valeric, iso-caproic, heptanoic and octanoic acid.

Upon reinoculation of R1 with the culture from R2 at the start of Period 2 (days 79-224), activity resumed immediately and from that point on both reactors showed a stable performance from days 89 until 205. During this period, the product spectrum was mainly dominated by CA, BA, and AA in both R1 (CA: 15.23 ± 2.27 g COD·L^{-1}; BA: 9.00 ± 1.72 g COD·L^{-1} ; AA: 1.60 ± 0.57 g COD·L^{-1}) and R2 (CA: 15.42 ± 1.94 g COD·L^{-1}, BA: 7.97 ± 1.55 g COD·L^{-1}, AA: 1.81 ± 0.63 g COD·L^{-1}). From the produced organic acids COD, there was a high selectivity for CA (R1: 61% ± 7%; R2: 64% ± 6%) and the even chain elongation intermediates in general (R1: 92% ± 1%; R2: 92% ± 1%) (Figure 1). Along with the consumption of LA, there was also a co-consumption of AA in both reactors at LA:AA COD ratios of 100:10 ± 3 and 100:7 ± 4 in R1 and R2 respectively, and the fraction of H_2 in the headspace was 27.78% ± 5.95% and 23.02% ± 6.47% respectively. During Period 2, the HRT of R1 was lowered between days 180 to 189 to 2 days to assess the effect of the shorter HRT on the reactor performance, but no significant change in the product spectrum was observed. On day 205, the stirring of R2 malfunctioned,

causing overdosage of HCl. Chain elongation stopped and LA accumulated up to 15.13 g COD·L^{-1} until the process recovered and full LA consumption was observed after 12 days.

On day 224, at the start of Period 3 (days 224-336) the content of both reactors was mixed to obtain two identical microbial communities after the disturbance of R2 and between days 229 and 264 the HRT of both reactors was decreased to 1.90 ± 0.05 and 1.99 ± 0.04 days in R1 and R2, respectively. This resulted in a decrease of the CA production rate to 1.48 ± 1.07 g COD·L^{-1} in R1 and 1.19 ± 0.87 g COD·L^{-1} in R2, and an increase of the IBA concentration to a maximum of 4.35 g COD·L^{-1} in R1 and 9.34 g COD·L^{-1} in R2. Just as in Period 1, a decrease in the H_2 fraction in the headspace was observed (Figure A.1). After increasing the HRT back to 4 days, the CA production gradually restored in R1 to a concentration of 15.87 ± 0.37 g COD·L^{-1}. Since the IBA concentration in R2 did not decrease after increasing the HRT, the reactor was reinoculated on day 273 with stored effluent from R2 and operated in batch mode for 9 days. Although CA was dominant for the two first weeks after the operation in batch mode (15.50 ± 0.80 g COD·L^{-1}), a gradual shift towards BA (20.37 g COD·L^{-1}) and IBA (3.40 g COD·L^{-1}) was observed until the end of the reactor operation.

3.2 Determining phenotypic community types

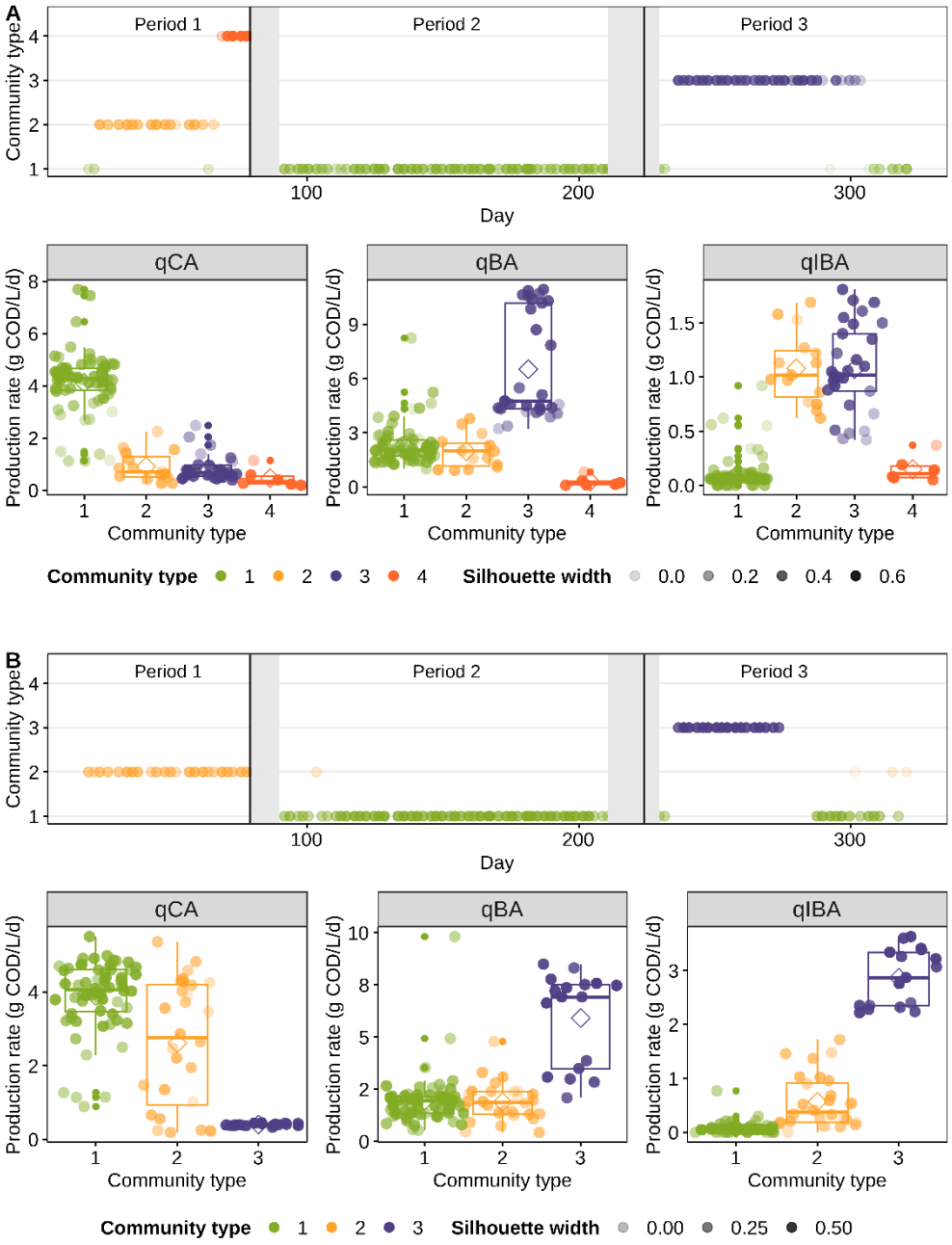
For each reactor, a GMM model was generated from the FC-data for fingerprinting, which resulted in a model with 40 mixtures for R1 and a model with 45 mixtures for R2. Via PCoA analysis based on the Bray-Curtis dissimilarity, followed by k-medoids clustering, CTs were determined in both reactors independently. Four CTs were determined in R1 (ANOSIM R: 0.8485) and three CTs in R2 (ANOSIM R: 0.7976). To assess whether the different CTs in each reactor are showing different functionalities, the CT dynamics were compared in terms of the production rates of CA (qCA), BA (qBA) and IBA (qIBA). Since the conditions for a 1-way ANOVA were not met, the differences between the CTs were assessed via the Kruskal-Wallis rank test

($p < 0.05$), and pairwise comparisons between the CTs were assessed with the Dunn-test ($p < 0.05$). CTs showing similar functionalities in terms of qCA, qBA and qIBA were labelled identically in both reactors, whereas CTs showing different functionalities were labelled uniquely within one reactor.

In terms of qCA, qBA, and qIBA (Figure 2), significant differences (at the 5% significance level) were observed between the determined CTs. In each reactor, there was one CT that showed a significantly higher qCA (4.13 ± 1.22 g $COD \cdot L^{-1} \cdot d^{-1}$ for R1 and 3.88 ± 0.97 g $COD \cdot L^{-1} \cdot d^{-1}$ for R2) compared to the other CTs of the same reactor ($p < 0.05$). In both R1 and R2 this CT was labelled as CT1. Furthermore, in each reactor, two different CTs could be identified that showed significantly higher qIBA, labelled as CT2 (1.08 ± 0.33 g $COD \cdot L^{-1} \cdot d^{-1}$ in R1 and 0.57 ± 0.49 g $COD \cdot L^{-1} \cdot d^{-1}$ in R2) and CT3 (1.07 ± 0.41 g $COD \cdot L^{-1} \cdot d^{-1}$ in R1 and 2.87 ± 0.53 g $COD \cdot L^{-1} \cdot d^{-1}$ in R2). The main difference is that in both reactors CT2 does not have a significantly higher qBA compared to CT1, but CT3 does. Only in R1 there was a fourth CT4 determined, which contains the samples from a crash, where all production rates were very low. It is important to note that, although in R1 and R2 similar CTs were determined in terms of functionality, this does not automatically imply that their community compositions were similar as well.

For each determined CT, samples were analysed via 16s rRNA amplicon sequencing. Sequencing data showed that CTs within each reactor have, besides different metabolic functionalities, also different taxonomic compositions (Figure 3). Additionally, equally labelled CTs in both reactors show similar taxonomic compositions. On genus level, samples from CT1 in both R1 and R2 show relative abundances for *Caproiciproducens* of 97.98% and 97.99%, respectively. In CT2 *Clostridium sensu stricto 12* became more dominant than *Caproiciproducens*. Although there was no significant difference in production rates observed for CT2 between R1 and R2, there was a stronger shift towards *Clostridium sensu stricto 12* in

R2 with abundances of 48.60% in R1 and 76.25% in R2 along with *Caproiciproducens* abundances of 41.38% in R1 and 6.78% in R2. In CT3, however, the abundance of *C. sensu stricto 12* was lower than the abundance of *Caproiciproducens* (77.95% in R1, 84.81% in R2). On ASV level, 3 *Caproiciproducens* ASVs (1, 2 and 3) were dominant in CT1 in both reactors, while only one of the three (ASV1) was dominant in CT3. In CT2, only ASV 2 and 3 were observed in R1, while all three ASVs were observed in R2 but at very low abundances for the *Caproiciproducens* genus. ASV 1 and 2 have respectively a 97.02% and 99.75% BLAST similarity with *Caproicibacterium lactatifermentans,* which is a known CA producer from LA and glucose (Wang et al., 2022). ASV 3 has a 100.00% BLAST similarity with *Ruminococcaceae* bacterium CPB6, which is a known CA producer from LA (Zhu et al., 2017). In R1, CT4 contained the samples of the reactor crash at day 65, of which the community was dominated by *Propionibacterium* spp. (60.70%).



359

360 *Figure 2: Boxplots for the production rates of caproic (qCA), butyric (qBA) and iso-butyric acid (qIBA) per community*
361 *type (CT), and the CT in function of time in R1 (top) and R2 (bottom). The opacity is related to the silhouette width,*
362 *which indicates how well the sample belongs to its designated cluster.*

363

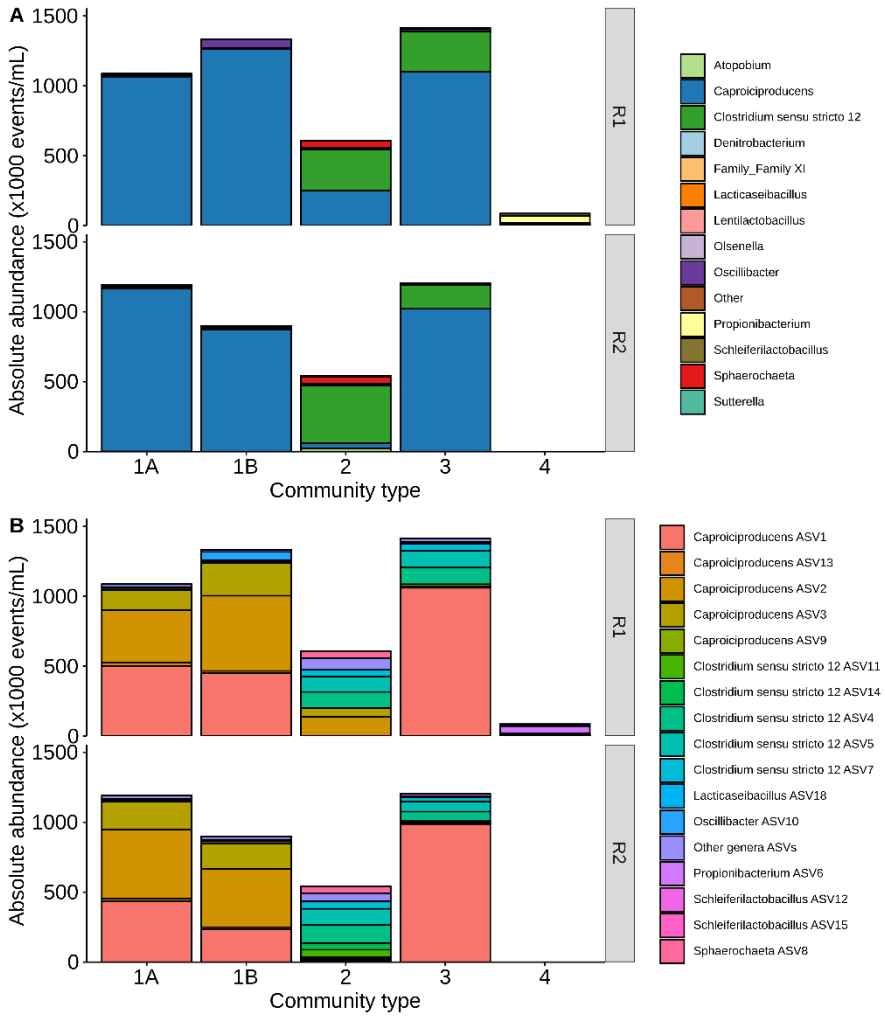


Figure 3: Cell concentration and community composition for each community type of each reactor. CT1 was analysed at two different time points. A) composition on genus level. B) composition on ASV level.

3.3 Flow cytometry as a tool to predict production rates

The accuracy of the trained random forest regression models for predicting of qCA, qBA and qIBA was assessed by comparing the predicted with the real production rates. The predicted values obtained through 5-fold nested cross-validation showed the model was able to accurately predict qCA (R1: R^2 = 0.87, MAE = 0.48 COD·L^{-1}·d^{-1}; R2: R^2 = 0.89, MAE = 0.40 COD·L^{-1}·d^{-1}), qBA (R1: R^2 = 0.88, MAE = 0.68 COD·L^{-1}·d^{-1}; R2: R^2 = 0.82, MAE = 0.53 COD·L^{-1}·d^{-1}), and qIBA (R1: R^2 = 0.87, MAE = 0.12 COD·L^{-1}·d^{-1}; R2: R^2 = 0.94, MAE = 0.12 COD·L^{-1}·d^{-1}).

375 Finally, to assess whether the FC data could be used to predict future production rates up to 2
376 HRTs in advance, the algorithms were trained for different time intervals between flow
377 cytometric measurement and future production rate. Since samples were mostly taken on
378 Mondays, Wednesdays and Fridays, the time intervals chosen were multiples of 2-3 days. Since
379 operational disturbances cannot be predicted ahead of time based on phenotypic fingerprints,
380 data points that were influenced by operational disturbances were excluded from the dataset
381 for predictions ahead of time, avoiding the prediction of production rates after a disturbance
382 with FC data acquired prior to the occurrence of the disturbance. This was done to ensure the
383 training of the algorithm was not compromised. Increasing the interval led to a gradual
384 decrease of the R^2 values and increase of the MAE (Table A.5), hence leading to a decrease in
385 accuracy of predictions further in the future. Considering qCA in R1, the R^2 decreased from
386 0.87 for real-time predictions, to 0.85 for an interval of 1 HRT, and to 0.80 for 2 HRTs, while the
387 MAE respectively first decreased from 0.48 g $COD \cdot L^{-1} \cdot d^{-1}$, to 0.43 g $COD \cdot L^{-1} \cdot d^{-1}$, and then
388 increased to 0.55 g $COD \cdot L^{-1} \cdot d^{-1}$ (Figure 4). Similar behaviour was observed for qBA and qIBA in
389 both R1 and R2.

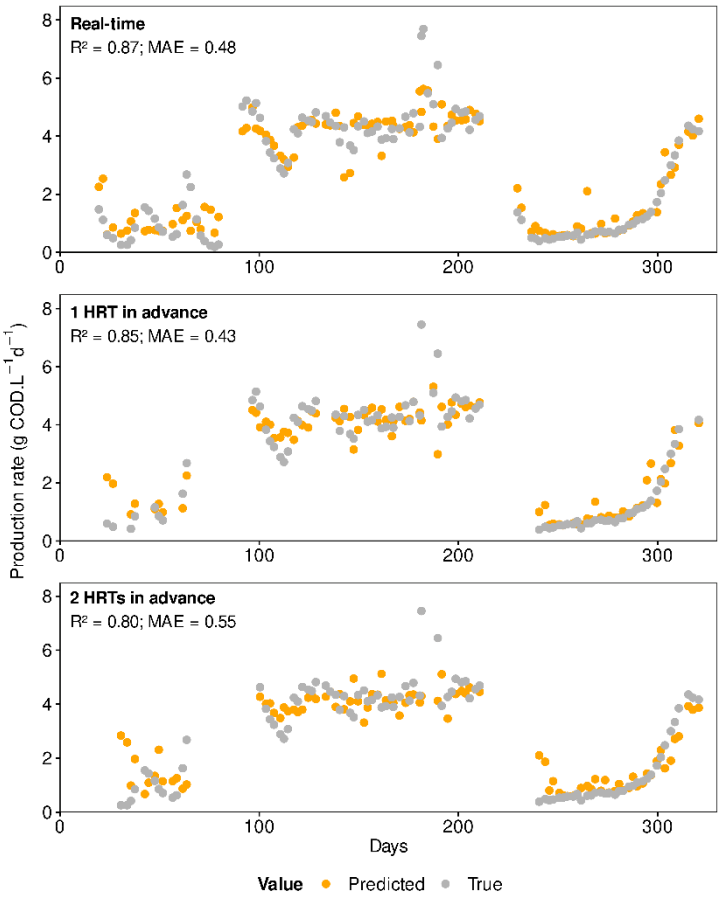


Figure 4: True versus predicted values of the production rates of caproic acid in R1, for different time intervals between flow cytometric measurement and predicted value. Top: real-time prediction; middle: 1 hydraulic retention time (HRT) interval; bottom: 2 HRTs interval.

4 Discussion

4.1 Lactic acid chain elongating microbiome can shift towards dominant BA and IBA production

In both R1 and R2, stable LA chain elongation with high CA selectivities was established. More specifically during Period 2, CA selectivities of 61% ± 7% in R1 and 64% ± 6% in R2 were achieved, which is higher than the values reported in other studies on LA chain elongation (Candry et al., 2020; Carvajal-Arroyo et al., 2019; Duber et al., 2020; Mariën et al., 2022b; Xu et al., 2018). Substantially higher volumetric CA production rates of 17.34 ± 3.53 mmol·L^{-1}·d^{-1} and 16.51 ± 2.19 mmol·L^{-1}·d^{-1} for R1 and R2 respectively, were achieved compared to the study by Candry et al. (2020) (ca. 6.5 mmol·L^{-1}·d^{-1}), where the reactors were operated under the same

operational conditions but no electron acceptor such as AA was present in the feed. Additionally, there was a co-consumption of AA at a molar LA:AA ratio of 100:14 ± 4 in R1 and 100:10 ± 7 in R2, while LA and AA were present in the medium at a 100:25 molar ratio. This indicates that the use of AA as external electron acceptor is beneficial for the chain elongation stability and selectivity, while the absence of AA as electron donor yields a broader spectrum with equal contribution of even and odd chain carboxylic acids (Candry et al., 2020). Several other studies have shown that the presence of an electron acceptor, like AA, enhances the LA chain elongation performance. Brodowski et al. (2022) studied the effect of the presence of external AA in LA chain elongation at different molar LA:AA ratios and found that external AA prevented destabilization of LA chain elongation caused by LA overloading and the formation of odd chain carboxylic acids, and reported approximate CA selectivities of 75%. Mariën et al. (2022b) supplemented AA as an external electron acceptor at a 100:25 molar ratio and also concluded that the addition of external AA was required for sufficient LA conversion during chain elongation.

Apart from stable chain elongation during Period 2, the product spectrum had shifted during the start-up phase in Period 1 and after decreasing the HRT in Period 3. On both occasions higher BA and IBA production were observed compared to Period 2, and a net production of AA was observed instead of a net consumption. The product spectrum might be altered by changing operational parameters such as the HRT. Mariën et al. (2022b) demonstrated that a decreasing HRT can lead to a shift from CA towards BA production in an EGSB reactor, but no net AA production and major shifts towards IBA were reported. Changes in the product spectrum can also be caused by changes in the reactor microbiome. Baleeiro *et al.* (2021) showed that in presence of AA an MCCA producing community can consume H_2 to produce BA, and that the supplementation of H_2 as electron donor improved the production of both IBA

and CA. In both reactors in the current study, the gas production rate and fraction of H_2 significantly decreased when BA and IBA production were dominant.

Organisms such as acetogenic *Clostridia*, can utilize H_2 and CO_2 for acetyl-CoA production via the Wood-Ljungdahl pathway (Bengelsdorf et al., 2018). The amplicon sequencing data showed that the microbial community was dominated by *Caproiciproducens* during stable chain elongation, while during periods of high BA and IBA production rates, *Clostridium sensu stricto 12* was present at high relative abundances. *Clostridium sensu stricto 12 is* closely related to *Clostridium luticellarii*, which is known to be able to produce butyrate and iso-butyrate in the presence of H_2 (Baleeiro et al., 2021; Petrognani et al., 2020). Chen et al. (2017a) showed that exogenous added BA triggered the production of IBA in methanol chain elongation and suggested that isomerization occurred as a detoxifying mechanism for the inhibitory effect of BA. However, Petrognani *et al.* (2020) showed that *C. luticellarii* is able to immediately co-produce both BA and IBA from methanol and AA, indicating that a detoxifying mechanism is not the reason for IBA production in *C. luticellarii*. In the current study, IBA production only occurred gradually and when BA concentrations were elevated compared to values during stable chain elongation.

4.2 Chain elongation community types are connected to both performance and taxonomic changes

FC fingerprinting enabled the detection of different phenotypic CTs in independent reactors based on the cell distribution over the identified GMMs. The assessment of qBA, qIBA and qCA per established CT showed that the different CTs not only have a different phenotypic composition, but also a significantly different metabolism/end-product spectrum (Figure 2). In both reactors, there was a significant difference between the CTs in terms of these production rates, and one particular CT (CT1) could be correlated with high CA production, while two

other CTs (CT2, CT3) could be correlated to elevated BA and IBA production (Figure 2). Props et al. (2018) used a similar approach to determine CTs and link them to contaminations in drinking water, in combination with additional community metrics. It is important to highlight that, in contrast to their study, community typing based on the FC fingerprint did not require additional metrics to achieve a sufficient discriminatory power to differentiate different states of the reactor microbiome performance, more specifically for LA chain elongation.

Besides assessing the phenotypic diversity of the different CTs, 16S rRNA gene sequencing was performed for each CT in each reactor. These results showed that each CT not only has a distinct phenotypic structure and a distinct metabolic functionality, bus also a distinct taxonomic composition (Figure 3). CTs from the different reactors with a similar functionality also show a similar distinct taxonomic composition that is different from the other CTs determined in the reactors. The CTs associated with good CA production (CT1 in both reactors) were dominated by *Caproiciproducens*, which is a known chain elongating organism. *Clostridium sensu stricto 12* was detected in CTs associated with elevated BA and IBA production and has a 100% BLAST similarity to *Clostridium luticellarii*, which is able to produce IBA from methanol and acetate in presence of CO_2 and H_2 (Petrognani et al., 2020). However, conclusions on the metabolism of IBA production and the organisms involved cannot be made, since both genera encompass species that are able to produce both CA and BA.

In a study performed by Liu et al. (2020) aiming at CA production via LA chain elongation, BA production was positively correlated with the abundance of *Clostridium sensu stricto 12*. Mariën et al. (2022b), investigated the effect of the volumetric loading rate on LA chain elongation and showed that shorter HRTs led to a shift towards BA and higher abundances of *Clostridium sensu stricto 12*. This genus was also observed in other studies on MCCA production (Baleeiro et al., 2021; Mariën et al., 2022a). In particular, Baleeiro *et al.* (2021)

476 reported that *Clostridium sensu stricto 12* thrived when LA and ethanol were supplemented,
477 and that its abundance could be positively correlated with non-methanogenic hydrogen
478 consumption. de Leeuw et al. (2020) hypothesized that *Clostridium sensu stricto 12* has the
479 ability to isomerize BA and IBA in both directions at a mildly acidic pH (5.2 and 5.5) during
480 methanol chain elongation. The addition of BA or IBA to the process lead to an equilibrium of
481 0.69 IBA:0.31 BA. In a study from Mariën et al (2022a) on LA chain elongation, IBA production
482 was detected in the system until tryptone was added. At that point, the relative abundance of
483 *Clostridium sensu stricto 12* OTU (also present up until that point) decreased to < 1%. These
484 findings indicate that *Clostridium sensu stricto 12* detected in the present study might be
485 responsible for the high BA production during Periods 1 and 2 and its isomerization to IBA,
486 along with the decrease in H_2 production potentially due to acetogenic hydrogen consumption.

487 ### *4.3 FC as a tool to predict future reactor performance*

488 This study has also shown that process performance parameters (i.e., production rates) can be
489 predicted based on cytometric fingerprints. The generated models had high R^2-values (> 0.80)
490 for the production rates of BA, IBA and CA, indicating that the FC data and the used
491 fingerprinting approach provides sufficient information for the prediction of production rates.
492 Since microbial communities are dynamic and changes in the community composition can
493 have an effect on the process performance later in time, the models were also trained for the
494 prediction of production rates in future time. Predictions were made for up to two times the
495 HRT in the future, which resulted in models with still an intermediate overall prediction
496 performance (R^2 > 0.63 for qCA; > 0.47 for qBA; > 0.79 for qIBA). The further in time the
497 predictions were made, the lower the R^2-statistic and the larger the MAE. This shows that
498 based on the cytometric fingerprint it is possible to assess instantaneous process performance
499 with high prediction performance, and to predict short-term future performance with an
500 intermediate level of confidence. While training of the models generated in this research was

limited to the FC data obtained during a defined period of reactor operation, future data would likely allow for improved predictions both in the present and the short-term future. A similar approach was used by Heyse et al. (2021) where models were trained for the prediction of the presence and abundance of bacterial taxa from FC data, where the abundance of top 50 OTUs could be predicted with R^2-values of 0.35 ± 0.24, ranging between 0.00 and 0.81, and demonstrating the potential of FC in microbiome management in aquaculture. In analogy with the current study, Liu et al. (2022) recently adopted a machine learning approach for the prediction of ecophysiological functions – i.e. the C4, C6, and C8 yields and concentrations – based on taxonomic data obtained through 16S rRNA amplicon sequencing. Although the chain elongation process performance could be predicted quantitatively with a > 90% accuracy, this approach relying on 16S rRNA amplicon sequencing is more expensive and time consuming than FC, while instant analysis and data interpretation with a sufficient temporal resolution is key for a novel monitoring strategy. Since obtaining and analysing sequencing results is time consuming, it is not possible to steer the process in time. Additionally, the timely detection of changes in the process performance will rely on the sampling frequency and the HRT of the system. While changes in the phenotypic structure of the microbiome can be detected instantly via FC, key organisms can still be detected via sequencing until there is sufficient wash-out, even when they are not active anymore. FC is therefore not only a tool for a novel monitoring strategy, but it is also fast enough to be implemented as steering strategy. FC is furthermore a cheaper technology, with the potential to be implemented in-line for RT-FCM (Abu-Absi et al., 2003; Favere et al., 2021, 2020; Hammes et al., 2012). Finally, if a crash of the system can be predicted, one might be able to adjust the feeding rate, add inoculum or stop the process to prevent excessive economic and energy losses. In contrast to the conventional approach of monitoring bioprocesses, which primarily focuses on measuring physicochemical parameters such as product concentrations, the method presented in this study offers a

distinct advantage in terms of predictive power. By employing this method, operators are able to assess process performance proactively, enabling them to identify potential issues before significant fluctuations in the concentrations of the desired product, such as caproic acid, are observed.

4.4 Final considerations and future outlook

The current study is a first approach for the use of FC as monitoring tool of a reactor microbiome, demonstrating that it can be a powerful tool to monitor LA chain elongation and, more broadly, the performance of mixed culture bioprocesses. In this study, stable chain elongation performance (i.e., R1: 4.13 ± 1.22 g $COD \cdot L^{-1} \cdot d^{-1}$ and R2: 3.88 ± 0.97 $COD \cdot L^{-1} \cdot d^{-1}$) was observed when *Caproiciproducens* was the dominant genus (> 93% relative abundance). Yet, it can be prone to disturbances by which the process performance and product spectrum might change, as seen in Periods 1 and 3 of the reactors operation where the product spectrum shifted towards IBA. While the exact mechanism of IBA production in the studied system is not yet fully understood, FC coupled with cell sorting could be used in future research. Bacteria belonging to GMM mixtures that strongly contributed to the prediction of the production rates of interest in this research, could then be isolated using a cell sorter, and their metabolism further studied.

The present work demonstrated that FC-based community typing is a good approach to assess the microbial health and performance of a mixed culture fermentation system. While the CT itself gives information on the production rates and product profile, the silhouette width could in future research indicate how strongly the reactor microbiome is associated with its designated CT, allowing monitoring the transition from one CT and metabolism to another. Furthermore, future work should also try to expand this approach to other LA chain elongation systems to assess whether the same CTs can be identified or if others arise, and the application

potential in other bioprocesses could be tested, provided that a constant cytometric fingerprint of a stable process can be established. It is important to note that this approach becomes challenging if the genotypic diversity varies significantly over time and multiple metastable conditions or alternative stable microbiomes with the same function exist. The current study also demonstrated that the predictive regression model established here is reliable and can provide a quantitative assessment/prediction of the fitness of the process. This method has the advantage that it does not rely on taxonomic data for its prediction. The algorithm can also be trained to predict ahead of time, so that the effect of certain changes in the fingerprint can be anticipated (here up to two times the hydraulic retention time in advance with reasonable accuracy; $R^2 > 0.63$ for CA).

Both approaches require a sufficiently large dataset encompassing the most commonly observed process performances and functionalities to cover the different performance states. Establishing these systems with only a limited training set may lead to limited predictive capacity or biased predictions when a certain CT emerges that has not been used in training the ML algorithm. It is for the further development therefore recommended to incorporate machine learning techniques that enable incremental training of both the predictive models and the identification of CTs, as this will allow for the monitoring system to continuously adapt and improve. However, in case of community typing the different identified CTs need to be defined and fixed as reference states, so that future samples can be compared.

Compared to sequencing, FC measurements are faster (minutes vs. days) and cheaper than sequencing, and changes in phenotypes are faster detected than changes in genotypes (Props et al., 2016). While FC monitoring was done manually in this study, it can in practice be implemented in-line for real-time flow cytometric measurement (RT-FCM). During RT-FCM, reactor samples are automatically diluted, stained, incubated and measured, increasing the

temporal resolution and decreasing the workload, which increases its advantages compared to sequencing (Abu-Absi et al., 2003; Favere et al., 2021, 2020; Hammes et al., 2012). When complex feedstocks are used, performing FC analysis can become challenging due to the presence of organic and inorganic impurities that can generate significant background signals, and the presence of organisms inherent to the feedstock. To overcome this issue, the effect of the constant influx of these organisms on the process needs to be studied on the one hand, while on the other hand either the substrate needs to be treated to remove the impurities, or the flow cytometric protocol needs to be improved so that the background signals can be differentiated from the other cells.

5 Conclusion

This study investigated the application potential of FC as a monitoring tool in mixed culture bioprocesses, more specifically LA chain elongation to CA, for which two CSTRs were operated and monitored for a period of 336 days, and during which stable chain elongation with high CA selectivity was alternated with periods of dominant BA and IBA production. Two monitoring approaches involving FC fingerprinting were studied. In a first approach, several CTs could be identified based on the fingerprint. Each CT could be linked to a specific functionality related to the production rates of the key fermentation products (i.e., qCA, qBA and qIBA) on the one hand, and to a distinct taxonomic composition on the other hand. CTs associated with a high qCA were dominated by *Caproiciproducens*, while dominant qBA and qIBA could be associated with the presence of *C. sensu stricto 12*, indicating competition between the two genera. In a second approach, FC data could be used to train random forest regression models, which enabled real-time predictions of the production rate of, amongst others, the key fermentation product CA, with a high accuracy ($R^2 > 0.87$) and up to 2 HRTs in the future with a reasonable accuracy ($R^2 > 0.63$). For the further optimisation of the studied approaches, future studies

should focus on the application in fermentation systems with more complex substrates, and the implementation of incremental learning to allow the monitoring system to continuously adapt and improve. The insights obtained from this study have demonstrated the application potential flow cytometry as a monitoring tool for fermentation processes, and its predictive power compared to conventional monitoring strategies relying on physicochemical parameters. This paves the way for further advancements in this approach, facilitating enhanced bioprocess control and optimisation. Consequently, it promotes improved process stability, enhanced product quality, and increased yields.

Additional file can be found in the online version of the paper.

Data availability

The raw fastq files that served as a basis for the bacterial community analysis were deposited in the National Center for Biotechnology Information (NCBI) database (Accession number PRJNA942594). All other data can be made available upon request.

Acknowledgements

K.S. is supported by the VLAIO (Flemish Agency for Innovation & Entrepreneurship) via a Baekeland Ph.D. fellowship (HBC.2019.2194) and by Organic Waste Systems nv. R.G. is supported by the Special Research Fund of Ghent University (grant number BOF19/STA/044).

6 References

Abu-Absi, N.R., Zamamiri, A., Kacmar, J., Balogh, S.J., Srienc, F., 2003. Automated flow cytometry for acquisition of time-dependent population data. Cytometry 51A, 87–96. https://doi.org/10.1002/cyto.a.10016

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

Andersen, S.J., De Groof, V., Khor, W.C., Roume, H., Props, R., Coma, M., Rabaey, K., 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. https://doi.org/10.3389/fbioe.2017.00008

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., 2016. Chain Elongation with Reactor Microbiomes: Open-Culture Biotechnology To Produce Biochemicals. Environ. Sci. Technol. 50, 2796–2810. https://doi.org/10.1021/acs.est.5b04847

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., Weissenbach, J., Ehrlich, S.D., Bork, P., 2011. Enterotypes of the human gut microbiome. Nature 473, 174–180. https://doi.org/10.1038/nature09944

Baleeiro, F.C.F., Kleinsteuber, S., Sträuber, H., 2021. Hydrogen as a Co-electron Donor for Chain Elongation With Complex Communities. Front. Bioeng. Biotechnol. 9, 650631. https://doi.org/10.3389/fbioe.2021.650631

Bengelsdorf, F.R., Beck, M.H., Erz, C., Hoffmeister, S., Karl, M.M., Riegler, P., Wirth, S., Poehlein, A., Weuster-Botz, D., Dürre, P., 2018. Bacterial Anaerobic Synthesis Gas (Syngas) and CO 2 + H 2 Fermentation, in: Advances in Applied Microbiology. Elsevier, pp. 143–221. https://doi.org/10.1016/bs.aambs.2018.01.002

Brodowski, F., Łężyk, M., Gutowska, N., Oleskowicz-Popiel, P., 2022. Effect of external acetate on lactate-based carboxylate platform: Shifted lactate overloading limit and hydrogen co-production. Sci. Total Environ. 802, 149885. https://doi.org/10.1016/j.scitotenv.2021.149885

Candry, P., Radić, L., Favere, J., Carvajal-Arroyo, J.M., Rabaey, K., Ganigué, R., 2020. Mildly acidic pH selects for chain elongation to caproic acid over alternative pathways during lactic acid fermentation. Water Res. 186, 116396. https://doi.org/10.1016/j.watres.2020.116396

Carvajal-Arroyo, J.M., Candry, P., Andersen, S.J., Props, R., Seviour, T., Ganigué, R., Rabaey, K., 2019. Granular fermentation enables high rate caproic acid production from solid-free thin stillage. Green Chem. 21, 1330–1339. https://doi.org/10.1039/C8GC03648A

Cavalcante, W. de A., Leitão, R.C., Gehring, T.A., Angenent, L.T., Santaella, S.T., 2017. Anaerobic fermentation for n-caproic acid production: A review. Process Biochem. 54, 106–119. https://doi.org/10.1016/j.procbio.2016.12.024

Chen, W.-S., Huang, S., Strik, D.P., Buisman, C.J., 2017a. Isobutyrate biosynthesis via methanol chain elongation: converting organic wastes to platform chemicals: Isobutyrate

663 production from organic waste using a mixed culture fermentation. J. Chem. Technol.
664 Biotechnol. 92, 1370–1379. https://doi.org/10.1002/jctb.5132
665 Chen, W.-S., Strik, D.P.B.T.B., Buisman, C.J.N., Kroeze, C., 2017b. Production of Caproic Acid from
666 Mixed Organic Waste: An Environmental Life Cycle Perspective. Environ. Sci. Technol.
667 51, 7159–7168. https://doi.org/10.1021/acs.est.6b06220
668 Crognale, S., Braguglia, C.M., Gallipoli, A., Gianico, A., Rossetti, S., Montecchio, D., 2021. Direct
669 Conversion of Food Waste Extract into Caproate: Metagenomics Assessment of Chain
670 Elongation Process. Microorganisms 9, 327.
671 https://doi.org/10.3390/microorganisms9020327
672 De Buck, V., Polanska, M., Van Impe, J., 2020. Modeling Biowaste Biorefineries: A Review. Front.
673 Sustain. Food Syst. 4, 11. https://doi.org/10.3389/fsufs.2020.00011
674 De Groof, V., Coma, M., Arnot, T., Leak, D.J., Lanham, A.B., 2019. Medium Chain Carboxylic Acids
675 from Complex Organic Feedstocks by Mixed Culture Fermentation. Molecules 24, 398.
676 https://doi.org/10.3390/molecules24030398
677 De Groof, V., Coma, M., Arnot, T.C., Leak, D.J., Lanham, A.B., 2020. Adjusting Organic Load as a
678 Strategy to Direct Single-Stage Food Waste Fermentation from Anaerobic Digestion to
679 Chain Elongation. Processes 8, 1487. https://doi.org/10.3390/pr8111487
680 de Leeuw, K.D., de Smit, S.M., van Oossanen, S., Moerland, M.J., Buisman, C.J.N., Strik,
681 D.P.B.T.B., 2020. Methanol-Based Chain Elongation with Acetate to n-Butyrate and
682 Isobutyrate at Varying Selectivities Dependent on pH. ACS Sustain. Chem. Eng. 8, 8184–
683 8194. https://doi.org/10.1021/acssuschemeng.0c00907
684 Duber, A., Jaroszynski, L., Zagrodnik, R., Chwialkowska, J., Juzwa, W., Ciesielski, S., Oleskowicz-
685 Popiel, P., 2018. Exploiting the real wastewater potential for resource recovery – *n* -
686 caproate production from acid whey. Green Chem. 20, 3790–3803.
687 https://doi.org/10.1039/C8GC01759J
688 Duber, A., Zagrodnik, R., Chwialkowska, J., Juzwa, W., Oleskowicz-Popiel, P., 2020. Evaluation of
689 the feed composition for an effective medium chain carboxylic acid production in an
690 open culture fermentation. Sci. Total Environ. 728, 138814.
691 https://doi.org/10.1016/j.scitotenv.2020.138814
692 Favere, J., Buysschaert, B., Boon, N., De Gusseme, B., 2020. Online microbial fingerprinting for
693 quality management of drinking water: Full-scale event detection. Water Res. 170,
694 115353. https://doi.org/10.1016/j.watres.2019.115353
695 Favere, J., Waegenaar, F., Boon, N., De Gusseme, B., 2021. Online microbial monitoring of
696 drinking water: How do different techniques respond to contaminations in practice?
697 Water Res. 202, 117387. https://doi.org/10.1016/j.watres.2021.117387
698 Gabrielli, M., Turolla, A., Antonelli, M., 2021. Bacterial dynamics in drinking water distribution
699 systems and flow cytometry monitoring scheme optimization. J. Environ. Manage. 286,
700 112151. https://doi.org/10.1016/j.jenvman.2021.112151
701 García, F.C., Alonso-Sáez, L., Morán, X.A.G., López-Urrutia, Á., 2015. Seasonality in molecular and
702 cytometric diversity of marine bacterioplankton: the re-shuffling of bacterial taxa by
703 vertical mixing: Seasonality in diversity of marine bacteria. Environ. Microbiol. 17, 4133–
704 4142. https://doi.org/10.1111/1462-2920.12984
705 Grootscholten, T.I.M., Steinbusch, K.J.J., Hamelers, H.V.M., Buisman, C.J.N., 2013. Improving
706 medium chain fatty acid productivity using chain elongation by reducing the hydraulic
707 retention time in an upflow anaerobic filter. Bioresour. Technol. 136, 735–738.
708 https://doi.org/10.1016/j.biortech.2013.02.114
709 Grootscholten, T.I.M., Strik, D.P.B.T.B., Steinbusch, K.J.J., Buisman, C.J.N., Hamelers, H.V.M.,
710 2014. Two-stage medium chain fatty acid (MCFA) production from municipal solid waste
711 and ethanol. Appl. Energy 116, 223–229.
712 https://doi.org/10.1016/j.apenergy.2013.11.061

Günther, S., Becker, D., Hübschmann, T., Reinert, S., Kleinsteuber, S., Müller, S., Wilhelm, C., 2018. Long-Term Biogas Production from Glycolate by Diverse and Highly Dynamic Communities. Microorganisms 6, 103. https://doi.org/10.3390/microorganisms6040103

Günther, S., Koch, C., Hübschmann, T., Röske, I., Müller, R.A., Bley, T., Harms, H., Müller, S., 2012. Correlation of Community Dynamics and Process Parameters As a Tool for the Prediction of the Stability of Wastewater Treatment. Environ. Sci. Technol. 46, 84–92. https://doi.org/10.1021/es2010682

Hammes, F., Broger, T., Weilenmann, H.-U., Vital, M., Helbing, J., Bosshart, U., Huber, P., Peter Odermatt, R., Sonnleitner, B., 2012. Development and laboratory-scale testing of a fully automated online flow cytometer for drinking water analysis. Cytometry A 81A, 508–516. https://doi.org/10.1002/cyto.a.22048

Heyse, J., Schattenberg, F., Rubbens, P., Müller, S., Waegeman, W., Boon, N., Props, R., 2021. Predicting the Presence and Abundance of Bacterial Taxa in Environmental Communities through Flow Cytometric Fingerprinting. mSystems 6, e00551-21. https://doi.org/10.1128/mSystems.00551-21

Joshi, S., Robles, A., Aguiar, S., Delgado, A.G., 2021. The occurrence and ecology of microbial chain elongation of carboxylates in soils. ISME J. 15, 1907–1918. https://doi.org/10.1038/s41396-021-00893-2

Kang, S., Kim, H., Jeon, B.S., Choi, O., Sang, B.-I., 2022. Chain elongation process for caproate production using lactate as electron donor in Megasphaera hexanoica. Bioresour. Technol. 346, 126660. https://doi.org/10.1016/j.biortech.2021.126660

Kleerebezem, R., Joosse, B., Rozendal, R., Van Loosdrecht, M.C.M., 2015. Anaerobic digestion without biogas? Rev. Environ. Sci. Biotechnol. 14, 787–801. https://doi.org/10.1007/s11157-015-9374-6

Koch, C., Fetzer, I., Schmidt, T., Harms, H., Müller, S., 2013. Monitoring Functions in Managed Microbial Systems by Cytometric Bar Coding. Environ. Sci. Technol. 130108105239000. https://doi.org/10.1021/es3041048

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

Kuhn, M., 2022. caret: Classification and regression training (manual).

Liu, B., Kleinsteuber, S., Centler, F., Harms, H., Sträuber, H., 2020. Competition Between Butyrate Fermenters and Chain-Elongating Bacteria Limits the Efficiency of Medium-Chain Carboxylate Production. Front. Microbiol. 11, 336. https://doi.org/10.3389/fmicb.2020.00336

Liu, B., Sträuber, H., Saraiva, J., Harms, H., Silva, S.G., Kasmanas, J.C., Kleinsteuber, S., Nunes da Rocha, U., 2022. Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture. Microbiome 10, 48. https://doi.org/10.1186/s40168-021-01219-2

Mariën, Q., Candry, P., Hendriks, E., Carvajal-Arroyo, J.M., Ganigué, R., 2022a. Substrate loading and nutrient composition steer caproic acid production and biofilm aggregation in high-rate granular reactors. J. Environ. Chem. Eng. 10, 107727. https://doi.org/10.1016/j.jece.2022.107727

Mariën, Q., Ulčar, B., Verleyen, J., Vanthuyne, B., Ganigué, R., 2022b. High-rate conversion of lactic acid-rich streams to caproic acid in a fermentative granular system. Bioresour. Technol. 355, 127250. https://doi.org/10.1016/j.biortech.2022.127250

Monaco, G., Chen, H., Poidinger, M., Chen, J., de Magalhães, J.P., Larbi, A., 2016. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. Bioinformatics 32, 2473–2480. https://doi.org/10.1093/bioinformatics/btw191

763 Narayana, S.K., Mallick, S., Siegumfeldt, H., van den Berg, F., 2020. Bacterial Flow Cytometry and
764 Imaging as Potential Process Monitoring Tools for Industrial Biotechnology.
765 Fermentation 6, 10. https://doi.org/10.3390/fermentation6010010
766 Nzeteu, C.O., Trego, A.C., Abram, F., O'Flaherty, V., 2018. Reproducible, high-yielding, biological
767 caproate production from food waste using a single-phase anaerobic reactor system.
768 Biotechnol. Biofuels 11, 108. https://doi.org/10.1186/s13068-018-1101-4
769 Nzeteu, C. orline, Coelho, F., Trego, A.C., Abram, F., Ramiro-Garcia, J., Paulo, L., O'Flaherty, V.,
770 2022. Development of an enhanced chain elongation process for caproic acid
771 production from waste-derived lactic acid and butyric acid. J. Clean. Prod. 338, 130655.
772 https://doi.org/10.1016/j.jclepro.2022.130655
773 Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B.,
774 Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker,
775 B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A.,
776 FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D.,
777 Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F., Weedon, J.,
778 2022. vegan: Community ecology package (manual).
779 O-Thong, S., Zhu, X., Angelidaki, I., Zhang, S., Luo, G., 2020. Medium chain fatty acids production
780 by microbial chain elongation: Recent advances, in: Advances in Bioenergy. Elsevier, pp.
781 63–99. https://doi.org/10.1016/bs.aibe.2020.04.002
782 Park, H.-S., Schumacher, R., Kilbane, J.J., 2005. New method to characterize microbial diversity
783 using flow cytometry. J. Ind. Microbiol. Biotechnol. 32, 94–102.
784 https://doi.org/10.1007/s10295-005-0208-3
785 Petrognani, C., Boon, N., Ganigué, R., 2020. Production of isobutyric acid from methanol by
786 *Clostridium luticellarii*. Green Chem. 22, 8389–8402.
787 https://doi.org/10.1039/D0GC02700F
788 Prest, E.I., Hammes, F., Kötzsch, S., van Loosdrecht, M.C.M., Vrouwenvelder, J.S., 2013.
789 Monitoring microbiological changes in drinking water systems using a fast and
790 reproducible flow cytometric method. Water Res. 47, 7131–7142.
791 https://doi.org/10.1016/j.watres.2013.07.051
792 Props, R., Monsieurs, P., Mysara, M., Clement, L., Boon, N., 2016. Measuring the biodiversity of
793 microbial communities by flow cytometry. Methods Ecol. Evol. 7, 1376–1385.
794 https://doi.org/10.1111/2041-210X.12607
795 Props, R., Rubbens, P., Besmer, M., Buysschaert, B., Sigrist, J., Weilenmann, H., Waegeman, W.,
796 Boon, N., Hammes, F., 2018. Detection of microbial disturbances in a drinking water
797 microbial community through continuous acquisition and advanced analysis of flow
798 cytometry data. Water Res. 145, 73–82. https://doi.org/10.1016/j.watres.2018.08.013
799 Reynolds, A.P., Richards, G., de la Iglesia, B., Rayward-Smith, V.J., 2006. Clustering Rules: A
800 Comparison of Partitioning and Hierarchical Clustering Algorithms. J. Math. Model.
801 Algorithms 5, 475–504. https://doi.org/10.1007/s10852-005-9022-1
802 Roghair, M., Liu, Y., Strik, D.P.B.T.B., Weusthuis, R.A., Bruins, M.E., Buisman, C.J.N., 2018.
803 Development of an Effective Chain Elongation Process From Acidified Food Waste and
804 Ethanol Into n-Caproate. Front. Bioeng. Biotechnol. 6, 50.
805 https://doi.org/10.3389/fbioe.2018.00050
806 Rubbens, P., Props, R., Kerckhof, F.-M., Boon, N., Waegeman, W., 2021. PhenoGMM: Gaussian
807 Mixture Modeling of Cytometry Data Quantifies Changes in Microbial Community
808 Structure. mSphere 6, e00530-20. https://doi.org/10.1128/mSphere.00530-20
809 Salma, M., Rousseaux, S., Sequeira-Le Grand, A., Alexandre, H., 2013. Cytofluorometric
810 detection of wine lactic acid bacteria: application of malolactic fermentation to the
811 monitoring. J. Ind. Microbiol. Biotechnol. 40, 63–73. https://doi.org/10.1007/s10295-
812 012-1200-3

Spirito, C.M., Marzilli, A.M., Angenent, L.T., 2018. Higher Substrate Ratios of Ethanol to Acetate Steered Chain Elongation toward *n* -Caprylate in a Bioreactor with Product Extraction. Environ. Sci. Technol. 52, 13438–13447. https://doi.org/10.1021/acs.est.8b03856

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

Stegmann, P., Londo, M., Junginger, M., 2020. The circular bioeconomy: Its elements and role in European bioeconomy clusters. Resour. Conserv. Recycl. X 6, 100029. https://doi.org/10.1016/j.rcrx.2019.100029

Steinbusch, K.J.J., Hamelers, H.V.M., Plugge, C.M., Buisman, C.J.N., 2011. Biological formation of caproate and caprylate from acetate: fuel and chemical production from low grade biomass. Energy Env. Sci 4, 216–224. https://doi.org/10.1039/C0EE00282H

Van Landuyt, J., Kundu, K., Van Haelst, S., Neyts, M., Parmentier, K., De Rijcke, M., Boon, N., 2022. 80 years later: Marine sediments still influenced by an old war ship. Front. Mar. Sci. 9, 1017136. https://doi.org/10.3389/fmars.2022.1017136

Wang, H., Gu, Y., Zhao, D., Qiao, Z., Zheng, J., Gao, J., Ren, C., Xu, Y., 2022. Caproicibacterium lactatifermentans sp. nov., isolated from pit clay used for the production of Chinese strong aroma-type liquor. Int. J. Syst. Evol. Microbiol. 72. https://doi.org/10.1099/ijsem.0.005206

Wilson, J.M., Chamberlain, E.J., Erazo, N., Carter, M.L., Bowman, J.S., 2021. Recurrent microbial community types driven by nearshore and seasonal processes in coastal Southern California. Environ. Microbiol. 23, 3225–3239. https://doi.org/10.1111/1462-2920.15548

Xu, J., Hao, J., Guzman, J.J.L., Spirito, C.M., Harroff, L.A., 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, 280–295. https://doi.org/10.1016/j.joule.2017.11.008

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