Machine learning identifies bioindicators for predicting caproate and caprylate productivity in microbial chain elongation

Bin Liu*, Heike Sträuber, João Saraiva, Hauke Harms, Ulisses Nunes da Rocha, Sabine Kleinsteuber

*Presenter, liu.bin@ufz.de

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

HIGHLIGHTS:

- Productivity and yield of *n*-caproate and *n*-caprylate were enhanced by reducing the hydraulic retention time.
- Bioindicators of hydraulic retention time were inferred from 16S rRNA amplicon sequence variants (ASVs) by machine learning using a random forest approach.
- The bioindicators quantitatively predicted the productivity of *n*-caproate and *n*-caprylate with more than 90% accuracy.

BACKGROUND: The ability to quantitatively predict the chain elongation process in microbial communities provides a stepping stone to engineer microbiomes for producing desired biochemicals such as medium-chain carboxylates[1,2]. Before engineering strategies such as changing operating conditions can be built to improve process performance, we need to analyze whether the quantifiable processes can be predicted by following microbial community dynamics. Here, we present the quantitative prediction of a lactate-based chain elongation process in bench-scale continuous bioreactors from community dynamics by machine learning. Recently, we enriched an undefined mixed culture capable of producing *n*-butyrate (C4), *n*-caproate (C6) and *n*-caprylate (C8) from xylan and lactate in a daily-fed bioreactor[3]. We found that the community developed toward predominating C4 production at the cost of C6 and C8 yields. Since C6 and C8 were the target products, it was relevant to manipulate the reactor microbiome for promoting the chain elongation process and thus optimizing C6/C8 productivity. To this end, we conducted a long-term study on the enriched chain-elongating microbiome in two parallel bioreactors continuously fed with lactate and xylan in mineral medium for 211 days.

RESULTS & DISCUSSION: We progressively reduced the hydraulic retention time (HRT) from 8 d to 2 d with different changing modes in both bioreactors. Comparing the performance at the two HRTs, the productivities at HRT of 2 d increased 2.5-fold, 5.6-fold and 7.2-fold for C4, C6 and C8, respectively; and the yields increased 1.5-fold and 2.4-fold for C6 and C8, whereas the C4 yield decreased 1.6-fold. Decreasing the HRT affected the composition and diversity of reactor microbiomes. Beta diversity analysis indicated a significant difference between the communities at HRT of 8 d

and 2 d ($P < 0.001$) but no significant difference between the communities in both reactors at the same HRT ($P > 0.05$). We hypothesized that the HRT reduction induced variations in community diversity, which could be used to predict the productivity and yield of carboxylates by machine learning. Our random forest analysis consisted of two parts. First, we did the feature selection for ASVs that were relevant to community dynamics caused by HRT reduction. The two bioreactors shared 11 HRT bioindicators (ASVs). Next, we trained the random forest algorithm with these bioindicators that were used later to predict C6/C8 productivity and yield. We reached more than 90% accuracy in the quantitative prediction of C6/C8 productivity (Figure 1). Four ASVs assigned to the genera *Olsenella*, *Lactobacillus*, *Syntrophococcus* and *Clostridium* IV were denoted as bioindicators of C6/C8 productivity. The inferred bioindicators may delineate their relevance to the enhanced C6/C8 productivity in the chain elongation process, manipulated by HRT decline. These species might be involved in the lactate-based chain elongation producing C6 and C8, e.g. species affiliated to *Clostridium* IV have been described as lactate-based chain elongation bacteria[4], while lactate formation by lactic acid bacteria (*Olsenella* and *Lactobacillus*) promoted the chain elongation process[3,5,6].

CONCLUSION: The quantifiable chain elongation process can be accurately predicted from 16S rRNA ASV dynamics with machine learning. If experiments with sufficient temporal and/or spatial resolution can be carried out, such a framework could be adapted for other ecosystem processes and more complex communities.
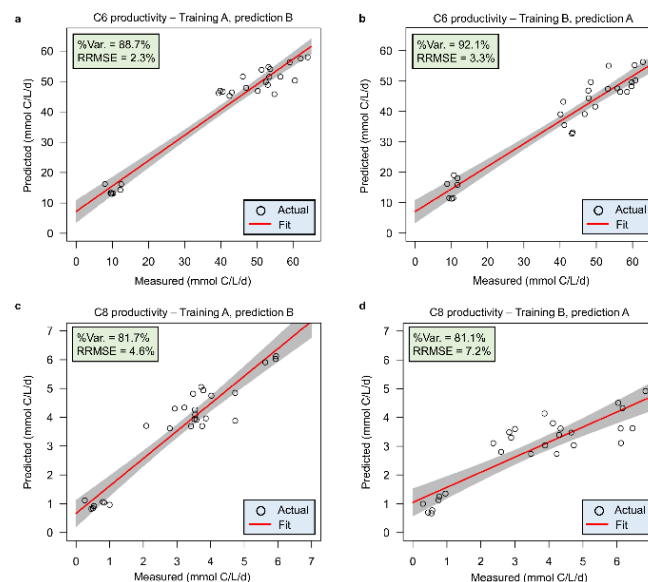
Figure 1. Prediction results of C6/C8 productivity using HRT bioindicators. a,b, Prediction performance of C6 productivity. c,d, Prediction performance of C8 productivity. In a and c, we used relative abundance data of bioreactor A as training set and relative abundance data of bioreactor B as test set. In b and d, relative abundance data of bioreactor B were used for training and of bioreactor A for testing. %Var., explained the target variance (%) of the training set. RRMSE, relative root mean square error.

REFERENCES

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

2.   Zhou, J. Predictive microbial ecology. *Microb. Biotechnol.* 2, 154–156 (2009).

3.   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*. 11, 336 (2020).

4.   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* 10, 102 (2017).

5.   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*. 18, 92 (2019).

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