Supporting Information

Conversion of waste microalgae into caproic acid using anaerobic membrane bioreactors without external electron donors

Xingdong Shi [a, b, d], Wei Wei [b,*], Jing Zhao [e], Yawei Li [a, d], Xiaoyin Liu [a, d], Junzeng Xu [a, d], Bing-Jie Ni [c,*]

[a] College of Agricultural Science and Engineering, Hohai University, Nanjing, 211100, China

[b] Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

[c] School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

[d] Jiangsu Province Engineering Research Center for Agricultural Soil-Water Efficient Utilization, Carbon Sequestration and Emission Reduction, Hohai University, Nanjing, 211100, China

[e] School of Minerals and Energy Resources Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

***Corresponding author:**

Prof. Bing-Jie Ni

E-mail: bingjieni@gmail.com

Dr. Wei Wei

E-mail: wei.wei@uts.edu.au

Table S1 The composition of the microalgae

Composition	Content
Carbohydrate	23.4%
Lipid	15.9%
Protein	50.0%
Ash	10.7%

Table S2 List of parameters included in uncertainty and sensitivity analyses

Parameter	Unit	Baseline	Distribution	Spearman's rho		
				MPSP	GWP100	FEC
Blank parameter						
Blank parameter	-	1	0.9, 1.1, uniform	-0.02	-0.02	-0.01
TEA						
Glucoamylase unit price	$/kg	6.16	4.928, 10, uniform	0.06	0.03	0.02
Alpha-amylase unit price	$/kg	6.16	4.928, 10, uniform	0.16	-0.05	-0.05
Plant annual operating days	Day	330	297, 363, uniform	-0.48	0.01	- [b]
Feedstock unit price	$/kg	0.15	0.12, 0.18, uniform	0.25	-	0.01
Natural gas unit price	$/kg	0.2527	0.198, 0.304, uniform	-	-0.01	-
Ethanol unit price [a]	$/kg	1.78	1.424, 2.136, uniform	0.25	-	-
Electricity unit price	$/kWh	0.07	0.067, 0.074, uniform	-0.03	0.23	0.01
Yeast unit price	$/kg	2	1.6, 2.4, uniform	0.03	0.004	0.01
Oleyl alcohol unit price	$/kg	1	0.088, 0.132, uniform	-0.02	-0.02	-0.01
Sulfuric acid unit price	$/kg	0.11	0.088, 0.132, uniform	0.31	0.02	0.01
Ammonium Hydroxide unit price	$/kg	0.204	0.1632, 0.2448, uniform	0.01	0.01	-0.01
NaOH unit price	$/kg	1.01	0.808, 1.212, uniform	0.07	-0.01	-
Pretreatment						
Feedstock capacity	kg/h	5000	4000, 6000, Triangular	0.21	-0.75	-0.20
Conversion						
Acid loading	g/g	1.47	1.1, 1.8, Triangular	0.14	0.07	0.08
Glucoamylase loading	g/g	0.0011	0.0005, 0.002, Triangular	0.02	0.03	-
Alpha-amylase loading	g/g	0.0082	0.005, 0.01, Triangular	0.03	-0.01	-
C6 yield	%	27	10, 50, Triangular	-0.65	-0.55	-0.93
Separation						
Extraction efficiency factor	%	100	50, 150, Triangular	0.05	-0.03	0.02
Facilities						
Boiler efficiency	%	90	85, 95, uniform	-0.07	-0.24	-0.30

[a]: Ethanol unit price was available in the scenario with ethanol as an electron donor

[b]: hyphen means the spearman's rho is quite low than 0.01.

Table S3 Abundance difference in key gene for MCCA and butanol production between control and yeast group

Gene	EC	Control	Yeast	p-values	Effect size	95.0% lower CI	95.0% upper CI
Butanol production							
ADH	1.1.1.1	0.214	0.179	0.009	0.035	0.008	0.062
Hyd	1.12.1.4	0.007	0.009	0.514	-0.002	-0.008	0.004
Nfn	1.18.1.2	0.101	0.090	0.243	0.011	-0.008	0.030
Nfn	1.19.1.1	0.101	0.090	0.261	0.011	-0.008	0.030
ALDH	1.2.1.3	0.074	0.065	0.302	0.009	-0.008	0.025
PFOR	1.2.7.1	0.265	0.298	0.041	-0.033	-0.065	0.000
AOR	1.2.7.5	0.116	0.100	0.109	0.016	-0.004	0.036
PTA	2.3.1.8	0.082	0.085	0.754	-0.003	-0.021	0.015
ACK	2.7.2.1	0.105	0.094	0.234	0.011	-0.008	0.031
MCCA production							
H2ase	1.12.7.2	0.074	0.046	0.000	0.027	0.012	0.043
Ech	1.12.2.1	0.001	0.001	1.000	0.000	-0.002	0.003
ETF	1.5.5.1	0.012	0.012	0.891	0.001	-0.007	0.008
TE	3.1.2.20	0.031	0.050	0.002	-0.019	-0.032	-0.006
CoAT	2.8.3.9	0.024	0.017	0.089	0.008	-0.002	0.017
CoAT	2.8.3.8	0.025	0.017	0.079	0.008	-0.001	0.018
CoAT	2.8.3.1	0.021	0.039	0.000	-0.019	-0.030	-0.008
ADA	1.2.1.10	0.059	0.039	0.003	0.020	0.006	0.034
PTA	2.3.1.8	0.082	0.085	0.754	-0.003	-0.021	0.015
ACK	2.7.2.1	0.105	0.094	0.234	0.011	-0.008	0.031
ADH	1.1.1.1	0.214	0.179	0.009	0.035	0.008	0.062
ACD	1.3.8.8	0.003	0.002	0.787	0.001	-0.003	0.005
ACD	1.3.8.7	0.067	0.065	0.815	0.002	-0.014	0.018
ACD	1.3.8.1	0.100	0.133	0.002	-0.032	-0.053	-0.011
HAD	1.1.1.35	0.048	0.052	0.501	-0.005	-0.019	0.009
HAD	1.1.1.157	0.098	0.094	0.663	0.004	-0.015	0.023
ACS	6.2.1.1	0.103	0.083	0.043	0.019	0.000	0.038
ECH	4.2.1.55	0.004	0.016	0.000	-0.012	-0.019	-0.005
ECH	4.2.1.17	0.123	0.131	0.473	-0.008	-0.030	0.014
ACTA	2.3.1.9	0.195	0.207	0.365	-0.012	-0.039	0.015
ACTA	2.3.1.16	0.036	0.030	0.321	0.006	-0.006	0.017

Table S4 The main TEA output

Item	Value	Item	Value
Net present value (NPV)	75.43 M$	Total direct costs (TDC)	114.97 M$/yr
Total capital investment (TCI)	128.76 M$	Fixed operating cost	5.72 M$/yr
Fixed capital investment (FCI)	122.63 M$	Variable operating cost	18.48 M$/yr
Direct permanent investment	76.64 M$	Annual operating cost	24.20 M$/yr
Return on investment	16.83%	Payback period	5.66 years
Caproic acid minimum selling price ($ kg^{-1})		1.86	
Caproic acid unit production cost per year ($)		8.21 M$	
Annual depreciation		5.47 M$	
Annual sales		51.64 M$	
Annual material cost		18.61 M$	
Annual utility cost		-0.13 M$	

Text S1 GC-MS configuration for organic acid and alcohol determine

The extracted liquid sample was analysed using a GC-2010 system (SHIMMADZ, Japan) coupled to a TQ8040 triple quadrupole mass selective detector (SHIMMADZ, Japan). A total of 1 μl of the extract was injected into the GC, equipped with a capillary column (30 m × 0.32 mm × 0.25 μm, SH-Stabllwax-DA, SHIMMADZU, USA). Carrier gas was helium and the chromatographic conditions were as follows: The analysis followed a temperature program in which the initial temperature was held at 40 °C for 1 minute, increased at a rate of 30 °C per minute to 130 °C, further increased to 170°C at a rate of 10 °C per minute, and then maintained at 170 °C for 2 minutes. Subsequently, the temperature was raised at a rate of 10 °C per minute to 240 °C and held at 240 °C for 1.5 minutes. Split injections with 60 split ratios were used and the first 1.8 min of the analysis were considered as solvent delay and omitted from the final chromatograms. The injection port was maintained at temperatures of 220 °C. Ion source temperature and interface temperature were set at 230 °C and 250 °C respectively. The mass spectrometer was operated in Q3-sim mode.

Text S2 Detailed information for metagenomic sequencing and analysis

Genomic DNA was first extracted from the sludge sample and randomly fragmented. Fragments of a specified average size were selected, followed by end repair, adaptor ligation, PCR amplification, and circularization. Single-stranded circular DNA molecules were then amplified via rolling circle amplification, generating DNA nanoballs (DNBs) containing multiple copies of the DNA. High-quality DNBs were subsequently loaded onto patterned nanoarrays using high-intensity DNA nanochip technology and sequenced via combinatorial Probe-Anchor Synthesis (cPAS).

Raw sequencing reads were filtered to remove adapter sequences, contaminants, and low-quality reads using SOAPnuke [1] with the parameters -n 0.001 -l 20 -q 0.5 --adaMis 3 --minReadLen 150. To ensure the accuracy of metagenomic analysis by preventing interference from the yeast genome, Bowtie 2 was used to remove the genome of *Saccharomyces cerevisiae*, obtained from NCBI (GCF_000146045.2)[2]. The resulting clean reads were co-assembled to contigs using metaSPAdes with default settings [3]. MetaPhlan 4 was used to identify the taxonomic assignment [4]. The gene functions were further characterized through eggNOG-mapper [5]. Gene ontology (GO), KEGG pathways, and CAZy family information were compiled using custom scripts [6].

In terms of genome-centric metagenomic analysis, the initial binning of contigs was performed with CONCOCT, MaxBin2, and metaBAT2, followed by refinement using the Bin_refinement module of metaWRAP [7], applying thresholds of ≥50% completeness and ≤10% contamination to generate the final metagenome-assembled genomes (MAGs). Genome quality was assessed with CheckM [8], and CoverM was used to map MAGs to the clean reads to determine their relative abundances in each sample [9]. Taxonomic assignment of all MAGs was performed using

the Genome Taxonomy Database, assisted by GTDB-Tk (v2.1.0) referencing the R207_v2 release [10]. MAGs were annotated under bacterial or archaeal modes using Prokka [11]. The function genes identified was used via METABOLIC (v4.0) [12].

Text S3 Detailed information for life cycle analysis (LCA) and techno-economic analysis (TEA)

System description

The microalgae biorefinery for producing caproic acid was designed to treat 5000 kg dry microalgae biomass per hour. Incoming dry microalgae feedstocks were first pre-processed to meet the quality specifications, such as size, ash, and moisture. The feed is normalized to approximately 4 wt.% total solids by adding water. Pretreatment uses sulfuric acid, nominally 1.47 g H_2SO_4 per g dry microalgae on a 93 wt% solution basis, at elevated temperature (121 °C) to hydrolysis biomass. The pretreated slurry is neutralized with ammonium hydroxide, brought to approximately 55 °C, and conditioned for saccharification. Conditioned slurry receives glucoamylase and alpha amylase, with dose rates scaled to microalgae throughput, in two saccharification stages operated near 55 °C and 90 °C, respectively. Sodium hydroxide is metered in two split additions to maintain pH during hydrolysis. The hydrolysate undergoes solid liquid separation to remove insoluble biomass components prior to fermentation. Fermentation is carried out at approximately 37 °C and mildly acidic pH (pH 5) with optional yeast addition and recycling to emulate a membrane bioreactor configuration, targeting yeast recovery near 95%. The model represents conversion to a distribution of products that include acetic, propionic, butyric, valeric, caproic, heptanoic, and caprylic acids, with caproic acid as the primary product. Fermentation performance, including titter, yield, and product distribution, is parameterized through specifications that enable either relative factors or absolute yield settings referenced to the incoming biomass mass flow. Produced organic acids are extracted from the clarified broth with oleyl alcohol (200 kg/h) using a multistage mixer settler train. The organic phase is purified by a distillation sequence tuned to recover and isolate C4 to C8 acids while returning solvent to the extraction loop. A little of fresh oleyl alcohol will be complemented to maintain the total flow of extractant is 200 kg/h. Final products include

butyric, caproic, heptanoic, and caprylic acid streams for revenue accounting. A combined waste stream comprising extraction raffinate, fermentation bleed, and separation residues is treated via anaerobic digestion with explicit biogas generation followed by a high-rate wastewater treatment train, and the treated water is routed for reuse. Site utilities include a cooling tower, a chilled water package, a process water centre, air distribution, a heat exchange network for heat integration, and a boiler turbogenerator sized to satisfy onsite heat and electricity demands, with net electricity import or export recorded for analysis. All process modeling, mass and energy tracking, equipment sizing, and cost correlations were implemented in BioSTEAM with a project specific thermodynamic and chemical property set. The full flowsheet is instantiated through a system factory, and custom unit models are provided for pretreatment, saccharification, fermentation, extraction, anaerobic digestion, and ancillary operations. The distribution of products in the yeast group was estimated to be 1.0% ethanol, 0.4% butanol, 10.0% acetic acid, 0.9% propionic acid, 18.0% butyric acid, 1.0% valeric acid, 27.0% caproic acid, 0.6% heptanoic acid, and 4.0% caprylic acid. In the control system, the corresponding values were 1.0% ethanol, 0.2% butanol, 10.0% acetic acid, 0.9% propionic acid, 8.6% butyric acid, 1.0% valeric acid, 8.0% caproic acid, 0.4% heptanoic acid, and 0.5% caprylic acid.

TEA configuration

The minimum product selling price was computed using discounted cash flow rate of return analysis. Unless stated otherwise, the financial basis assumed 10 percent internal rate of return, 330 operating days per year, MACRS 7 depreciation for process equipment with steam and power at MACRS 20, a corporate income tax of 21 percent, a construction schedule of 8 percent, 60 percent, and 32 percent, working capital at 5 percent of fixed capital investment, and a financing structure with 40 percent debt at 8 percent interest over 10 years. Inside and

outside battery limit facilities, including cooling, chilled water, air, process water, and the boiler turbogenerator, are included in capital and operating costs. By product revenues and net electricity import or export are credited or charged consistently within the operating ledger.

LCA configuration

Attributional life cycle assessment was performed at the biorefinery fence line with a functional unit of 1 kg of the main product, caproic acid by default. Inventories of input chemicals, direct emissions, and net electricity were taken directly from simulated stream tables and utility ledgers. Impact characterization focused on 100-year global warming potential and fossil energy consumption. Category factors were compiled from literature and applied to material and energy flows. The framework separates contributions from feedstock, microalgae, materials such as acids, bases, enzymes, solvent, net electricity, and direct non biogenic emissions, with optional end of life carbon accounting for products.

Uncertainty analysis

Uncertainty was propagated via Monte Carlo simulation with Latin hypercube sampling. Sampled parameters included fermentation performance, extraction and distillation efficiencies, reagent and solvent requirements, microalgae composition and price, and key financial items. Outputs included the minimum product selling price, 100-year global warming potential, and fossil energy consumption. Global sensitivities were quantified using Spearman's rank correlation coefficients computed from the simulation ensemble. All parameter ranges, distributions, and system configurations, including a baseline with yeast recycle, no yeast variants, and an ethanol supplemented case, are specified in the accompanying analysis scripts.

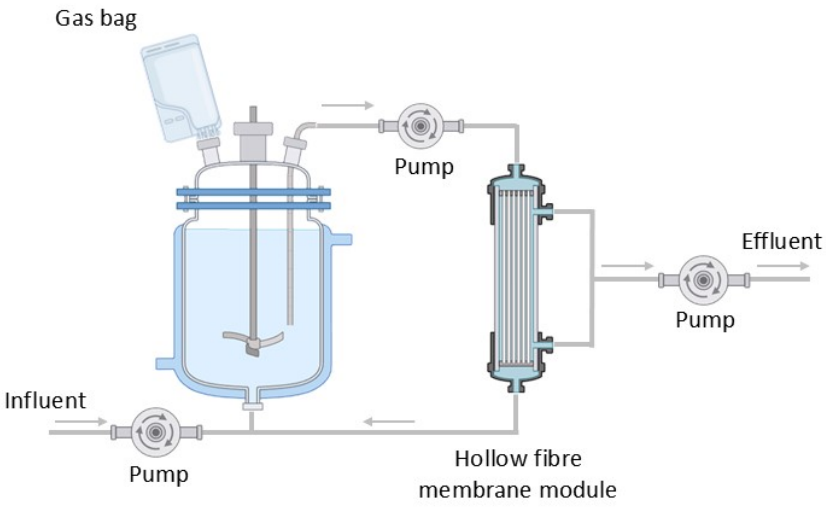

Fig. S1 Schematic diagram of anaerobic membrane bioreactor setups.

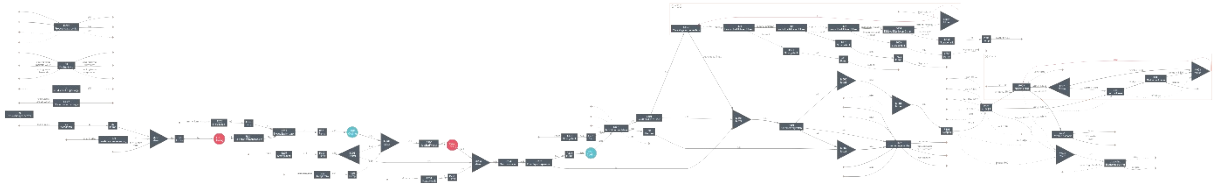

Fig. S2 The complete process flow diagram of microalgae biorefinery for producing caproic acid

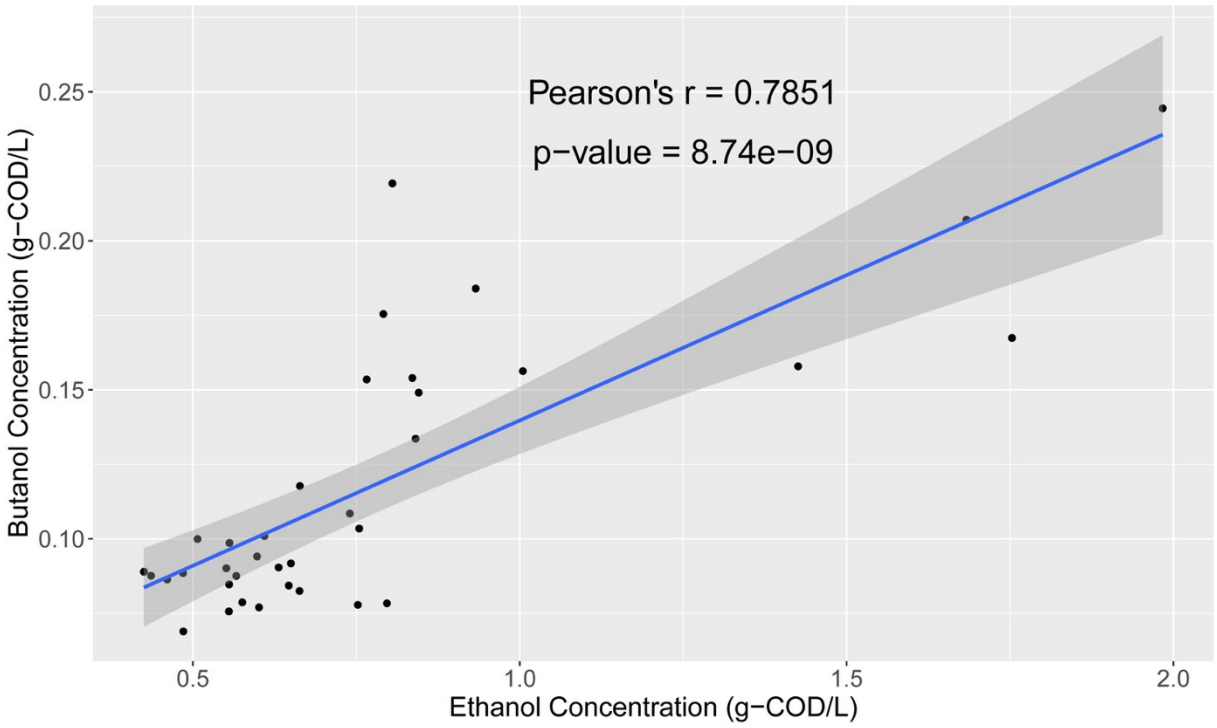

Fig. S3 The correction between ethanol concentration and butanol concentration.

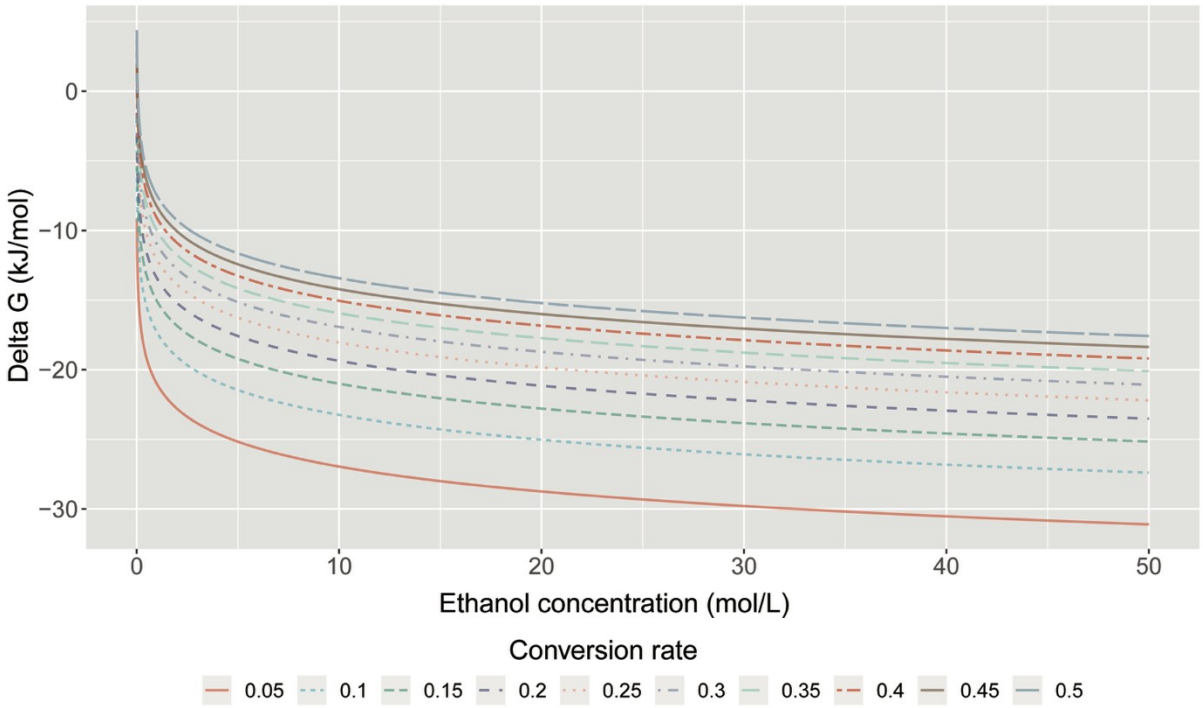

Fig. S4 The Gibbs free energy changes for microbial butanol production at different ethanol concentration and conversion rate. This simulation was performed under the condition of pH 5, 37 °C and 2 g L^{-1} butyric acid as substrate.

Reference

1 Y. Chen, Y. Chen, C. Shi, Z. Huang, Y. Zhang, S. Li, Y. Li, J. Ye, C. Yu, Z. Li, X. Zhang, J. Wang, H. Yang, L. Fang and Q. Chen, *GigaScience*, 2018, **7**, gix120.

2 B. Langmead and S. L. Salzberg, *Nat. Methods*, 2012, **9**, 357–359.

3 S. Nurk, D. Meleshko, A. Korobeynikov and P. A. Pevzner, *Genome Res.*, 2017, **27**, 824–834.

4 A. Blanco-Míguez, F. Beghini, F. Cumbo, L. J. McIver, K. N. Thompson, M. Zolfo, P. Manghi, L. Dubois, K. D. Huang, A. M. Thomas, W. A. Nickols, G. Piccinno, E. Piperni, M. Punčochář, M. Valles-Colomer, A. Tett, F. Giordano, R. Davies, J. Wolf, S. E. Berry, T. D. Spector, E. A. Franzosa, E. Pasolli, F. Asnicar, C. Huttenhower and N. Segata, *Nat. Biotechnol.*, 2023, **41**, 1633–1644.

5 A. Hernández-Plaza, D. Szklarczyk, J. Botas, C. P. Cantalapiedra, J. Giner-Lamia, D. R. Mende, R. Kirsch, T. Rattei, I. Letunic, L. J. Jensen, P. Bork, C. von Mering and J. Huerta-Cepas, *Nucleic Acids Res.*, 2023, **51**, D389–D394.

6 X. Shi, W. Wei, L. Wu, Y. Huang and B.-J. Ni, *Appl. Environ. Microbiol.*, 2024, **90**, e01250-23.

7 G. V. Uritskiy, J. DiRuggiero and J. Taylor, *Microbiome*, 2018, **6**, 158.

8 D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson, *Genome Res.*, 2015, **25**, 1043–1055.

9 S. T. N. Aroney, R. J. P. Newell, J. N. Nissen, A. P. Camargo, G. W. Tyson and B. J. Woodcroft, *Bioinformatics*, DOI:10.1093/bioinformatics/btaf147.

10 P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz and D. H. Parks, *Method. Biochem. Anal.*, 2022, **38**, 5315–5316.

11 T. Seemann, *Method. Biochem. Anal.*, 2014, **30**, 2068–2069.

12 Z. Zhou, P. Q. Tran, A. M. Breister, Y. Liu, K. Kieft, E. S. Cowley, U. Karaoz and K. Anantharaman, *Microbiome*, 2022, **10**, 33.