Autotrophic adaptive laboratory evolution of the acetogen *Clostridium autoethanogenum* delivers the gas-fermenting strain LAbrini with superior growth, products, and robustness

Henri Ingelman[a,*], James K. Heffernan[b,*], Audrey Harris[c,*], Steven D. Brown[c], Kurshedaktar Majibullah Shaikh[a], Asfand Yar Saqib[a], Marina J. Pinheiro[a,1], Lorena Azevedo de Lima[a,2], Karen Rodriguez Martinez[b], Ricardo A. Gonzalez-Garcia[b], Grant Hawkins[c], Jim Daleiden[c], Loan Tran[c], Hunter Zeleznik[c], Rasmus O. Jensen[c], Vinicio Reynoso[c], Heidi Schindel[c], Jürgen Jänes[d], Séan D. Simpson[c], Michael Köpke[c], Esteban Marcellin[b,#], Kaspar Valgepea[a,#]

[a]ERA Chair in Gas Fermentation Technologies, Institute of Technology, University of Tartu, 50411 Tartu, Estonia

[b]Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, 4072 St. Lucia, Australia

[c]LanzaTech Inc., 60077 Skokie, USA

[d]Institute of Molecular Systems Biology, ETH Zürich, 8049 Zürich, Switzerland

*These authors contributed equally to this work.

[1]Present address: Libbs Farmacêutica, 01140-050 São Paulo, Brazil

[2]Present address: Braskem S/A, 13086-530 Campinas, Brazil

[#]Correspondence: Esteban Marcellin, e.marcellin@uq.edu.au; Kaspar Valgepea, kaspar.valgepea@ut.ee

Running title: Autotrophic adaptive laboratory evolution of an acetogen

ABSTRACT

Microbes able to convert gaseous one-carbon (C1) waste feedstocks are increasingly important to transition to the sustainable production of renewable chemicals and fuels. Acetogens are interesting biocatalysts since gas fermentation using *Clostridium autoethanogenum* has already been commercialised. However, most acetogen strains need complex nutrients, display slow growth, and are not robust for routine bioreactor fermentations. In this work, we used three different and independent adaptive laboratory evolution (ALE) strategies to evolve the wild-type *C. autoethanogenum* to grow faster, without yeast extract and to be robust in operating continuous bioreactor cultures. Multiple evolved strains with improved phenotypes were isolated on a minimal medium with one strain, named "LAbrini" (LT1), exhibiting superior performance regarding the maximum specific growth rate, product profile, and robustness in continuous cultures. Whole-genome sequencing of the evolved strains identified 25 mutations. Of particular interest are two genes that acquired seven different mutations across the three ALE strategies, potentially as a result of convergent evolution. Reverse genetic engineering of sporulation-related mutations in genes CLAU_3129 (*spo0A*) and CLAU_1957 recovered all three superior features of our ALE strains through triggering significant proteomic rearrangements. This work provides a robust *C. autoethanogenum* strain to accelerate phenotyping and genetic engineering and to better understand acetogen metabolism, which we named "LAbrini".

Keywords: adaptive laboratory evolution, gas fermentation, chemostat, reverse genetic engineering, whole-genome sequencing.

1. INTRODUCTION

One-carbon (C1) gas (CH_4, CO_2, CO) emissions from burning fossil fuels are the main contributors to climate change (1). Alarmingly, this continuing trend might lead to irreversible climate change and a significant loss of biodiversity (2). To tackle and possibly reverse these horrific trends, a circular bioeconomy with decarbonised energy sectors, sustainable chemical and fuel production, and integrated waste management must replace our linear take-make-waste economy (3). While solar and wind are leading the way for decarbonising energy generation, the improvement of various biocatalysts allowing the "up-cycling" of waste C1 feedstocks is considered one of the key strategies for sustainable production of chemicals and fuels towards achieving carbon neutrality (4).

One attractive route toward recycling waste carbon into products is gas fermentation (5). The ideal biocatalysts for this bioprocess are bacteria termed acetogens. Acetogens use the energy-efficient Wood–Ljungdahl pathway (WLP) to fix carbon oxides (CO_2 and CO) into acetyl-CoA from a wide range of feedstocks (6). Notably, the model-acetogen *Clostridium autoethanogenum* is already used in commercial-scale ethanol production (7) and has been successfully piloted for acetone and isopropanol production (8). However, the lack of non-commercial acetogen strains capable of growing on a minimal medium has slowed the understanding of acetogen metabolism and rational metabolic engineering. The strains currently publicly available uptake gas slowly, grow slowly, require yeast extract (YE), and are challenging to run in continuous bioreactor cultures. Minimal growth medium (e.g. without YE) is required for accurate quantification of carbon distribution and metabolic modelling to understand metabolism and predict phenotypes (9, 10). Faster gas uptake and faster growth would speed up the phenotyping and genetic engineering of acetogens (6, 11). A more robust strain in operating continuous bioreactor cultures (12, 13) would ease generating high-quality steady-state data. Also, the use of various *C. autoethanogenum* strains for academic research complicates cross-comparison between laboratories that using a standard and user-friendly strain could homogenise. A publicly available acetogen strain with the above-mentioned improved features would benefit the academic community enormously.

Adaptive laboratory evolution (ALE) is a powerful and elegant approach for obtaining strains with improved traits. Evolution's reliance on natural selection to enrich mutants with increased fitness allows strain optimisation without requiring *a priori* knowledge of the genetic alterations necessary to effect such changes (14). Genes that accumulate mutations across multiple ALEs are under stronger

71 selection pressure and serve as attractive targets for further genetic engineering studies (15). Such
72 convergent evolution is a fascinating phenomenon where distinct lineages evolve similar traits or
73 functions independently, often in response to analogous environmental challenges. Observing and
74 comparing these adaptations can shed light on the underlying principles of convergent evolution at
75 the molecular and metabolic levels. ALE has recently been used also for engineering and improving
76 autotrophic growth capabilities. For example, ALE was used to convert the natively heterotrophic
77 workhorse microbes *Escherichia coli* (16) and *Pichia pastoris* (17) to CO_2-fixing organisms. In
78 acetogens, ALE has improved tolerance towards methanol in *Sporomusa ovata* (18), cyanide in
79 *Clostridium ljungdahlii* (19), benzene, toluene, and xylenes in *C. autoethanogenum* (20), and acetate
80 in *Clostridium* sp. AWRP (21). ALE was also recently used in *Clostridium carboxidivorans* to expand its
81 product spectrum and improve growth on CO_2+H_2 (22) and in *Eubacterium limosum* for higher CO
82 tolerance and faster growth (23, 24). Additionally, *C. ljungdahlii* and *C. autoethanogenum* have been
83 adapted by continuous bioreactor cultivation to grow without YE (25, 26), but it is not clear if the
84 phenotype was the result of physiological adaptation (i.e. reversible) or evolution (i.e. permanent).
85 In this work, we used three different and independent ALE strategies across three continents to
86 obtain superior *C. autoethanogenum* strains that can grow faster, without YE on minimal medium,
87 and are robust for operating continuous bioreactor cultures. As a result, multiple evolved strains with
88 improved phenotypes were isolated, with one strain – named "LAbrini" (LT1) – exhibiting superior
89 performance in terms of maximum specific growth rate (μ_{max}), product profile, and robustness in
90 bioreactor cultures. Genome sequencing revealed 25 mutations with two genomic "hot-spot" regions
91 across the evolved strains, and reverse genetic engineering of sporulation-related mutations
92 recovered all three superior features of our ALE strains. This work provides the robust *C.*
93 *autoethanogenum* strain LAbrini to the academic community to accelerate phenotyping and genetic
94 engineering and facilitate the generation of high-quality steady-state datasets. Additionally, our work
95 showcases the synergistic benefits of ALE, mutation analysis, and reverse engineering for shaping
96 superior gas-fermenting strains.
97

98 ## 2. MATERIALS AND METHODS

99 ### 2.1. Starting bacterial strain for adaptive laboratory evolution (ALE)

100 The starting strain for all three ALEs (see next section) was *Clostridium autoethanogenum* strain JA1-
101 1 (27) deposited in the German Collection of Microorganisms and Cell Cultures (DSMZ) as DSM 10061
102 and stored as a glycerol stock at –80°C.

103 ### 2.2. Three adaptive laboratory evolution (ALE) workflows

104 ### 2.2.1. University of Tartu (UT) ALE workflow – bottle ALE on CO

105 Glycerol stock of DSM 10061 was revived in PETC-MES medium (28) supplemented with 5 g/L of
106 fructose, 1.5 g/L of YE, and 0.4 g/L of cysteine-HCl·H_2O as the reducing agent. Cells were sub-cultured
107 from exponential growth phase into the same medium but with fructose replaced by CO as the carbon
108 source: this culture was Round 1 of UT ALE (Fig. 1a). Cells were then sub-cultured from exponential
109 growth phase to the same medium with CO (Round 2). Starting from Round 3, each round was carried
110 out in 2–4 replicate bottles. The culture with the fastest growth was sub-cultured to the next round
111 from exponential growth phase. During Rounds 3–5, YE was removed from the medium stepwise: first
112 to 0.5 g/L (Round 3), then to 0.05 g/L (Round 4), and no YE starting from Round 5 of UT ALE. Serial sub-
113 culturing was stopped after 35 rounds when the decrease of maximum specific growth rate (μ_{max})
114 during UT ALE had levelled off (Fig. 1b).

115 Bottle cultivations were carried out with 50 mL medium in 250 mL Schott bottles incubated
116 horizontally at 37°C with orbital shaking at 120 RPM under strictly anaerobic conditions. Culture bottle
117 headspace was pressurised to 140 kPa with a 60% CO and 40% Ar gas mixture (AS Eesti AGA).
118 Additionally, a "swap ALE" was performed where the culture was transferred between the previous
119 CO mixture and syngas (50% CO, 20% CO_2, 20% H_2, 10% Ar; AS Eesti AGA) (Supplementary Fig. 1).
120 Growth was tracked by measuring culture optical density (OD) at 600 nm. Maximum specific growth
121 rate (μ_{max}) was calculated using 3–5 data points from the exponential phase (except for a few cultures
122 using two data points indicated in Fig. 1b) that yielded a correlation coefficient $R^2 > 0.98$ between
123 culture time and natural logarithm (ln) of OD. At the time of sub-culturing, cultures were sampled for
124 extracellular metabolome analysis and occasionally cryo-preserved in glycerol.

125 Glycerol stocks made from Round 26 (110 generations; blue marker in Fig. 1b) and 35 (142
126 generations; yellow marker in Fig. 1b) were used to obtain single colony isolates by plating the mixed
127 populations on Round 1 medium with agar (12 g/L) and incubation at 37°C under autotrophic
128 conditions (same gas mixture as for UT ALE). Ten colonies were screened in bottles under the same
129 conditions as in Round 5 of UT ALE (e.g. CO without YE), and three isolates named UT1 (from Round
130 26) and UT2 and UT3 (from Round 35) were selected for further characterisation (see below) based
131 on μ_{max}, or ethanol or 2,3-butanediol (2,3-BDO) yields per biomass (see Results for details). The latter
132 yields (mmol of product/gram of dry cell weight [gDCW]) were estimated by linear regression between
133 the respective product (mmol/L) and biomass concentrations (gDCW/L) that were measured during
134 exponential growth phase with correlation coefficients $R^2 > 0.95$.

135 ### 2.2.2. University of Queensland (UQ) ALE workflow – bottle and bioreactor ALE on syngas

136 Glycerol stock of DSM 10061 was revived as in Round 1 of UT ALE, and the bottle was pressurised to
137 125 kPa with CO_2. Cells were sub-cultured in single replicate bottles from the exponential growth
138 phase into the same medium but with a gradual reduction of fructose (i.e. 2.5, twice 1, and 0 g/L),
139 with 190 kPa syngas (50% CO, 20% CO_2, 20% H_2, 10% Ar; BOC Australia) replacing the carbon source
140 (Series 1 of UQ ALE). Next, YE was gradually removed from the medium over 8 bottles (i.e. twice 1.5,
141 twice 1, 0.5, twice 0.25, and 0 g/L). The first YE-free culture was cryo-preserved before pausing the
142 ALE experiment (Series 2 of UQ ALE; grey markers in Fig. 1d). The latter glycerol stock was revived in
143 the same medium as in Round 1 of UT ALE and bottle pressurised to 140 kPa with syngas before sub-
144 culturing in YE-free medium 3-times with 190 kPa and 4-times with 140 kPa syngas (Series 3 of UQ
145 ALE; orange markers in Fig. 1d).

146 Bottle cultivations were carried out with either 25 or 100 mL medium (125 or 500 mL bottles) and
147 incubated horizontally at 37°C with orbital shaking at 200 RPM under strictly anaerobic conditions. At
148 the time of sub-culturing, strains were occasionally cryo-preserved. Growth was tracked by measuring
149 culture OD at 600 nm and μ_{max} was calculated using three data points from the exponential phase that
150 yielded a correlation coefficient $R^2 > 0.99$ between culture time and ln of OD.

151 The final culture of Series 3 was used as the inoculum for a bioreactor ALE experiment, where the
152 culture was grown on syngas predominantly under chemostat operation at pH 5 (29). Other
153 parameters (e.g. dilution rate, agitation, gas flow rates) were varied based on culture growth (Series
154 4 of UQ ALE; blue markers in Fig. 1d). After obtaining steady-states at high biomass concentrations
155 and ~2,300 h of fermentation, the bioreactor culture was plated on solid medium (Round 1 of UT ALE
156 medium with 12 g/L agar) and incubated at 37°C under autotrophic conditions to obtain single
157 colonies. Initially, pure CO at 70 kPa was used, but since only a few colonies appeared after 12 days,
158 the headspace was replaced with syngas, and plates were incubated for an additional 7 days, after
159 which 12 colonies were screened in bottles under the same conditions as at the end of Series 2 and 3
160 of UQ ALE (e.g. syngas without YE). Three isolates named UQ1, UQ2, and UQ3 were selected for further
161 characterisation (see below) based on μ_{max}, or ethanol or 2,3-BDO yields per biomass (see Results for
162 details). The latter yields (mmol of product/gDCW) were estimated with individual replicate (single
163 colony revival) product (mmol/L) and biomass concentrations (gDCW/L), measured during the mid-
164 /late-exponential growth phase.

165 #### 2.2.3. LanzaTech (LT ALE) workflow – chemical mutagenesis and bioreactor ALE on syngas

166 Glycerol stock of DSM 10061 was revived in the same medium as in Round 1 of UT ALE and cultured
167 using anaerobic techniques as described previously (8). Chemical mutagenesis was performed by
168 growing cells anaerobically in bottles at 37°C with shaking until mid-exponential growth phase and
169 then exposing cells to the chemical mutagen N-methyl-N'-nitro-N-nitrosoguanidine (NTG) as described

170 before (30), except that after treatment cells were washed with anaerobic medium lacking YE and
171 then transferred and cultured in a chemostat. Bioreactor chemostat gas fermentation with
172 *C. autoethanogenum* has been described before (8) and for selection and adaptation *via* gas
173 fermentation no YE was added to the medium. The culture was grown for approximately 50 days to a
174 final dilution rate (D) of 1 day^{-1} on the gas mixture of 53% CO, 7% H_2, 15% CO_2, and 25% N_2 before
175 single colony isolation and glycerol stock preparation.

176 ## 2.3. Characterisation of single colonies from evolved cultures

177 ### 2.3.1. Bottle experiments

178 We carried out additional bottle experiments to more accurately characterise the single colonies
179 isolated from the evolved cultures. We used the same conditions as for Round 5 in UT ALE (e.g. CO
180 without YE) for UT ALE isolates and LT1 (UT1, UT2, UT3, LT1) and the same conditions as for Series 2
181 in UQ ALE (i.e., syngas without YE) for UQ ALE isolates and LT1 (UQ1, UQ2, UQ3, LT1). We also cultured
182 the ALE starting strain JA1-1 for comparison with YE (Round 2 conditions in UT ALE) and without YE
183 (Round 5 conditions in UT ALE) on CO and syngas with UT ALE cultivation conditions. These cultures
184 were sampled frequently from the exponential phase to enable a more accurate estimation of μ_{max}
185 and product yields. μ_{max} was calculated using 3–12 data points from the exponential phase that yielded
186 a correlation coefficient $R^2 > 0.99$ between culture time and ln(OD). Product yields for UT isolates, LT1,
187 and JA1-1 (mmol of product/gDCW) were estimated as described in section 2.2.1 with correlation
188 coefficients $R^2 > 0.94$. For UQ isolates and LT1 on syngas, product yields were calculated from the
189 difference of two time points in mid-exponential growth.

190 ### 2.3.2. Bioreactor chemostat cultures

191 Cells were grown either on CO (60% CO and 40% Ar; AS Eesti AGA) for UT1, UT2, UT3, and RE2 (see
192 below) or syngas (50% CO, 20% H2, 20% CO2, and 10% Ar; AS Eesti AGA) for UQ1 and LT1 in a
193 chemically defined medium without YE as described before (29). Cells were grown under strictly
194 anaerobic conditions at 37°C and at pH 5 maintained by 2.5 M or 5 M NH_4OH. Chemostat continuous
195 cultures were performed in 1.4 L Multifors bioreactors (Infors AG) at a working volume of 750 mL
196 connected to a Hiden HPR-20-QIC mass spectrometer (Hiden Analytical) for online high-resolution off-
197 gas analysis. The system was equipped with peristaltic pumps; mass flow controllers (MFCs); pH,
198 oxidation-reduction potential (ORP), and temperature sensors. Antifoam (435530; Sigma-Aldrich) was
199 continuously added to the bioreactor at 10 µL/h to avoid foaming.

200 Chemostat cultures were run at D ~0.5, ~1, and ~2 day^{-1} (μ~0.02, ~0.04, and ~0.08 h^{-1}) across
201 various isolates with variable gas-liquid mass transfer (gas flow rate and agitation) to maintain similar
202 steady-state biomass concentrations (Table 1). In total, 33 biologically independent steady-states

203 were achieved after OD, gas uptake and production rates had been stable (<15% variability) for at
204 least three working volumes. Bioreactor off-gas analysis using a mass spectrometer for determination
205 of specific gas uptake (CO and H_2) and production rates (CO_2 and ethanol) (mmol/gDCW/day) and
206 determination of carbon recoveries and balances have been described before (29).

207 ### 2.4. Biomass concentration analysis

208 Biomass concentration in gDCW/L was determined by measuring culture OD at 600 nm and using the
209 correlation coefficient between culture OD and DCW established at 0.23 (UT and LT ALE isolates) or
210 0.21 (UQ ALE isolates) using the methodology described before (31).

211 ### 2.5. Extracellular metabolome analysis

212 Analysis of exo-metabolome was performed using filtered broth samples stored at -20°C until analysis.
213 Organic acids and alcohols were analysed by HPLC (Shimadzu Prominence-I LC-2030 plus system) using
214 a Rezex™ ROA-Organic Acids H+ (8%) 300 ×7.8 mm column (00H-0138-K0; Phenomenex) and a guard
215 column (03B-0138-K0; Phenomenex). Twenty microlitres of the sample were injected using an auto-
216 sampler and eluted isocratically with 0.5 mM H_2SO4 at 0.6 mL/min for 30 min at 45°C. Compounds
217 were detected by a refractive index detector (RID-20A; Shimadzu) and identified and quantified using
218 relevant standards using the software LabSolution (Shimadzu). We note that cells produced 2R,3R-
219 butanediol.

220 ### 2.6. Whole-genome sequencing (WGS) and mutation analysis

221 We performed whole-genome sequencing (WGS) for isolated clones UT1, UT2 and UT3 (UT ALE); UQ1,
222 UQ2, and UQ3 (UQ ALE); LT1 (LT ALE); and also the ALE starting strain JA1-1 (DSM 10061) to ensure
223 the accurate determination of mutations that appeared during ALE experiments. Genomic DNA for
224 sequencing was extracted from exponential phase cultures of UT and UQ isolates using the
225 MasterPure Gram Positive DNA Purification Kit (MGP04100; Biosearch Technologies) or Qiagen
226 DNeasy Powersoil Pro Kit (47016; Qiagen) following the manufacturer's instructions. Genomic DNA
227 for the LT1 strain was extracted with a phenol-chloroform extraction protocol. Purified DNA was
228 quantified for UT and UQ isolates using the NanoDrop™ 1000 instrument (Thermo Scientific) or the
229 Qubit 2.0 instrument (Invitrogen). DNA sequencing was performed using the MiSeq v2 Micro 300
230 cycles sequencing kit (MS-103-1002; Illumina) on the MiSeq sequencer (Illumina) with 2 × 151 bp
231 paired-end dual indexed (2 × 8 bp) reads or using the Nextera DNA Flex Library Preparation Kit
232 (20018705; Illumina) on the NovaSeq6000 sequencer (Illumina) with 2 x 150 bp paired-end reads. For
233 LT1, DNA processing and sequencing was outsourced to third parties to generate 150 bp pair-end

234 reads using the NovaSeq 6000 Illumina platform and continuous long reads from the PacBio Sequel II
235 instrument, which were used to produce a genome assembly (see below).
236 WGS data analysis, including trimming, filtering, mapping assembly, and annotation, was carried
237 out using the tool breseq version 0.37.0 (32) with default settings for detecting missing coverage
238 evidence and new sequence junctions and following parameters: limit fold coverage, 500; require
239 match fraction, 0.95; polymorphism frequency cutoff, 1; polymorphism reject indel homopolymer
240 length, 0; polymorphism reject surrounding homopolymer length, 0; polymorphism minimum variant
241 coverage each strand, 1; polymorphism prediction base quality cutoff, 30; polymorphism minimum
242 total coverage, 2; and polymorphism minimum total coverage each strand, 1. With paired-end read
243 overlaps ignored, the total genome coverage was 50.78 for LT1, 410.67–551.11 for UQ isolates, 88.35–
244 92.70 for UT isolates, and 75.30 for JA1-1. Mean mapping quality was 58.99 for LT1, 58.95–59.07 for
245 UQ isolates, 59.32–59.34 for UT isolates, and 59.3274 for JA1-1. High-quality reads were mapped
246 against the NCBI reference genome assembly ASM148472v1 (GenBank CP012395.1) for *C.*
247 *autoethanogenum* JA1-1 (33). Two mutations were excluded from further analysis as they were
248 considered non-relevant: one mutation was present in all isolates and the starting strain JA1-1, and
249 one mutation was present in JA1-1 only.
250 Additionally, for LT1, the Illumina data was used to polish the PacBio genome assembly with the
251 Pilon v1.24 tool (34). The polished assembly was deemed to be highly accurate based on the
252 observation that no 100% frequency variants were predicted by breseq v0.35.1 (32) using the polished
253 assembly and Illumina data as inputs. The validated LT1 assembly was then deposited to the NCBI
254 genome submission portal (Genbank CP110420) for annotation with the NCBI Prokaryotic Genome
255 Annotation Pipeline. See Supplementary Table 1 for a comparison of strain LT1 (LAbrini) gene
256 identifiers (IDs or locus_tags) with previously published complete assemblies of *C. autoethanogenum*
257 JA1-1 genome sequences (33, 35).

258 ### 2.7. Transcriptome analysis of UT isolate bioreactor chemostat cultures

259 Transcriptome analysis for the steady-state chemostat cultures of UT isolates listed in Table 1 was
260 performed using RNA sequencing (RNA-seq). In addition, the transcriptome of the non-steady-state
261 culture of UT2 at D = 1 day^{-1} without apparent "CO toxicity" (end of culture in Supplementary Fig. 2;
262 see text below for details) was also determined. This enabled the comparison of the transcriptome of
263 the "CO toxicity" steady-state culture of UT2 at D = 1 day^{-1} with a "non-CO toxicity" culture at the
264 same D.
265 Ten millilitres of culture were pelleted by centrifugation (5,000 × *g* for 3 min at 4ºC) and
266 resuspended in 5 mL of RNAlater (76106; Qiagen). Samples were stored at 4ºC overnight, centrifuged
267 (4,000 × *g* for 10min at 4ºC), and pellets stored at -80ºC until RNA extraction. Thawed pellets were

268 resuspended in RLT buffer (74104; Qiagen) containing β-mercaptoethanol and lysed with acid-washed
269 glass beads (G4649; Merck) using the Precellys® 24 instrument with liquid nitrogen cooling (Bertin
270 Technologies). Total RNA was extracted using the RNeasy mini kit (74104, Qiagen) with off-column
271 TURBO™ DNase treatment (AM2239; Invitrogen), followed by purification and enrichment using the
272 RNA Clean and Concentrator™ kit (R1018, Zymo). The efficiency of the total RNA purification and DNA
273 removal was verified using the NanoDrop™ 1000 instrument (Thermo Scientific) and the quality of
274 RNA extracts was checked using the TapeStation 2200 equipment (Agilent Technologies). Total RNA
275 concentration was determined using the Qubit 2.0 instrument (Q32866; Invitrogen). Next, ribosomal
276 RNA (rRNA) was removed using the QIAseq FastSelect –5S/16S/23S Kit (335925; Qiagen) and stranded
277 mRNA libraries were prepared using the QIAseq Stranded RNA Lib Kit (180743; Qiagen). RNA
278 sequencing was performed using the NextSeq MID150 sequencing kit (20024904; Illumina) on the
279 NextSeq500 sequencer (Ilumina) with 2 x 75 bp paired-end dual indexed (2 x 8 bp) reads.
280 The quality of raw NextSeq reads was verified using MultiQC (36) and adapter sequences were
281 trimmed using the Cutadapt Python package (version 2.10; (37)) allowing a minimum read length of
282 35 nt. The resulting high-quality paired-end reads were mapped to the NCBI reference genome
283 assembly ASM148472v1 (GenBank CP012395.1) for *C. autoethanogenum* JA1-1 (33) using the align
284 function within Rsubread package (version 2.4.2; (38)). Then, four .bam files per sample were merged
285 using Samtools (version 1.10; (39)) and genomic features were assigned using the featureCounts
286 functions within Rsubread. Transcript abundances were determined as described before (29). Genes
287 were determined to be differentially expressed by fold-change > 1.5 and q-value < 0.05 after false
288 discovery rate correction (FDR) (40) (Supplementary Table 2). RNA-seq data has been deposited in the
289 NCBI Gene Expression Omnibus repository under accession number GSE225123.

290 ## 2.8. Reverse genetic engineering of mutations in JA1-1

291 The detailed steps of our CRISPR/Cas9-aided homologous recombination workflow for genetic
292 engineering include designing and selecting sgRNAs, plasmid construction and electroporation into *C.*
293 *autoethanogenum* JA1-1, screening and plasmid curing, and confirmation of transformants were
294 described before (28). Briefly, the homology arms (HAs) for the CLAU_3129 (*spo0A*) deletion plasmid
295 were obtained by PCR amplifying the 1 kb upstream and downstream regions of the *spo0A* gene. The
296 derived 5' and 3' HA fragments were fused using overlap extension PCR and moved into the backbone
297 plasmid to create an intermediate plasmid pGFT104. Two unique sgRNAs targeting the *spo0A* gene
298 were selected and cloned into plasmid pGFT104 using a quick-change PCR protocol (41), resulting in
299 plasmids pGFT106 and pGFT107. Nine transformant colonies were screened altogether and the
300 selected isolate named RE1. A similar approach was employed for constructing the plasmid (pGFT088)
301 for generating the SNP in CLAU_1957 (mutation #10 in UT1), wherein the HAs arms were obtained by

302 PCR amplifying the 1kb upstream and downstream regions of the target nucleotide and the SNP was
303 incorporated into the HAs using the quick-change PCR protocol. Only one sgRNA was identified and
304 used where the SNP was part of the PAM sequence, thus disrupting it after homologous recombination
305 to prevent Cas9 from cleaving the target site again. Seven transformant colonies were screened and
306 the selected isolate was named RE2. Supplementary Table 3 lists the primers used and Supplementary
307 Table 4 the strains and plasmids used or constructed in the genetic engineering part of the current
308 study. See Supplementary File 1 (pGFT106), Supplementary File 2 (pGFT107), and Supplementary File
309 3 (pGFT088) for plasmid maps.

310 Growth characterisation of the reverse engineered strains RE1 and RE2 was carried out as
311 described in section 2.3. For autotrophic bottle cultures, strains were grown using same conditions as
312 for Round 5 in UT ALE and for RE1 additionally on syngas. Correlation coefficient R^2 values for μ_{max} and
313 product yield calculations were in the same limits as stated in section 2.3.1 with the exception of the
314 acetate yield for one RE1 CO culture being $R^2 = 0.9$.

315 ## 2.9. Proteome analysis of reverse engineered strains and JA1-1

316 Proteome analysis was carried out from three bioreplicate cultures for each of the two reverse
317 engineered strains RE1 and RE2, and JA1-1. Strains were grown using UT ALE cultivation conditions on
318 syngas and medium of Round 2 in UT ALE (with YE). Cultures were sampled at OD~0.4 and μ_{max} was
319 calculated using 4–5 data points from the exponential phase that yielded a correlation coefficient R^2
320 > 0.98 between culture time and ln of OD.

321 Sampling, sample preparation, LC-MS/MS analysis using data-independent acquisition (DIA), and
322 DIA MS data analysis were performed as described before (28). Briefly, cell pellets were treated with
323 a chaotrope-based lysis buffer and bead beating, lysates alkylated and processed, and proteins
324 digested. Peptides were then analysed on a Ultimate 3500 RSLCnano system (Dionex) and a Q-Exactive
325 HF (Thermo Fisher Scientific) mass spectrometer. Analysis of DIA MS data was performed using the
326 DIA-NN software suite (42) and the protein sequence database of assembly ASM148472v1 for *C.*
327 *autoethanogenum* JA1-1 (NCBI Genbank CP012395.1) (33). This allowed us to confidently quantitate
328 31,449 peptides and 2,257 proteins across all samples, and 27,030 peptides and 2,152 proteins on
329 average within each sample after removing shared peptides from analysis.

330 Protein expression fold-changes (FC) with q-values between RE1 and JA1-1, and RE2 and JA1-1
331 were determined using the software Perseus (43) with Student's T-test. Only proteins quantified with
332 at least two peptides (2,215 across all samples) were used to ensure higher quantification accuracy.
333 Proteins were considered to be differentially expressed by a FC > 2 and a q-value < 0.05 after FDR
334 correction (40). Differentially expressed proteins are presented in Supplementary Table 5. Proteomics
335 data have been deposited to the ProteomeXchange Consortium

336 (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (44) with the data
337 set identifier PXD047330.

338 2.10. Sporulation tests for reverse engineered strains and JA1-1

339 Autotrophic cultures for sporulation tests were the same as described for proteome analysis while
340 heterotrophic cultures were grown in 10 mL of YTF (28) in 50ml serum bottles and incubated
341 horizontally at 37°C with orbital shaking at 120 RPM under strictly anaerobic conditions. Samples from
342 the stationary phase during ~40 and ~20 days of autotrophic and heterotrophic incubation,
343 respectively, were observed on glass slips covered with an agarose pad using phase-contrast
344 microscopy (Zeiss Observer Z1) at ×1000 magnification. Heterotrophic cultures were also plated
345 before and after heat-shock (bottles stored for 10 min in a 80°C water bath) on YTF agar (28) and
346 plates were photographed after colonies appeared in 7-8 days.

347 3. RESULTS

348 3.1. Autotrophic adaptive laboratory evolution (ALE) of *Clostridium autoethanogenum* JA1-
349 1 for improved growth phenotypes

350 Three different autotrophic ALE strategies were employed across three facilities (Fig. 1a; see Materials
351 and Methods for details) to improve the wild-type *Clostridium autoethanogenum* JA1-1 strain's
352 growth phenotype. For the ALE experiment at the University of Tartu (UT ALE), serial propagation of
353 batch bottle cultures was employed for 142 generations with the gradual removal of YE at the start.
354 For the ALE experiment done at The University of Queensland (UQ ALE), firstly, serial propagation of
355 batch cultures in serum bottles for 107 generations was used with the gradual removal of YE at the
356 start. After that, the evolved population was transferred into a continuous bioreactor culture, where
357 it was grown for another 109 generations without YE. Moreover, for the third ALE experiment done
358 at LanzaTech, Inc. (LT ALE), cells from a bottle culture were exposed to the chemical mutagen NTG
359 after which the culture was transferred into a continuous bioreactor culture and grown without YE for
360 ~50 days. The UT ALE experiment was conducted on CO, while the LT and UQ ALE experiments were
361 done on syngas (CO, CO_2, H_2).

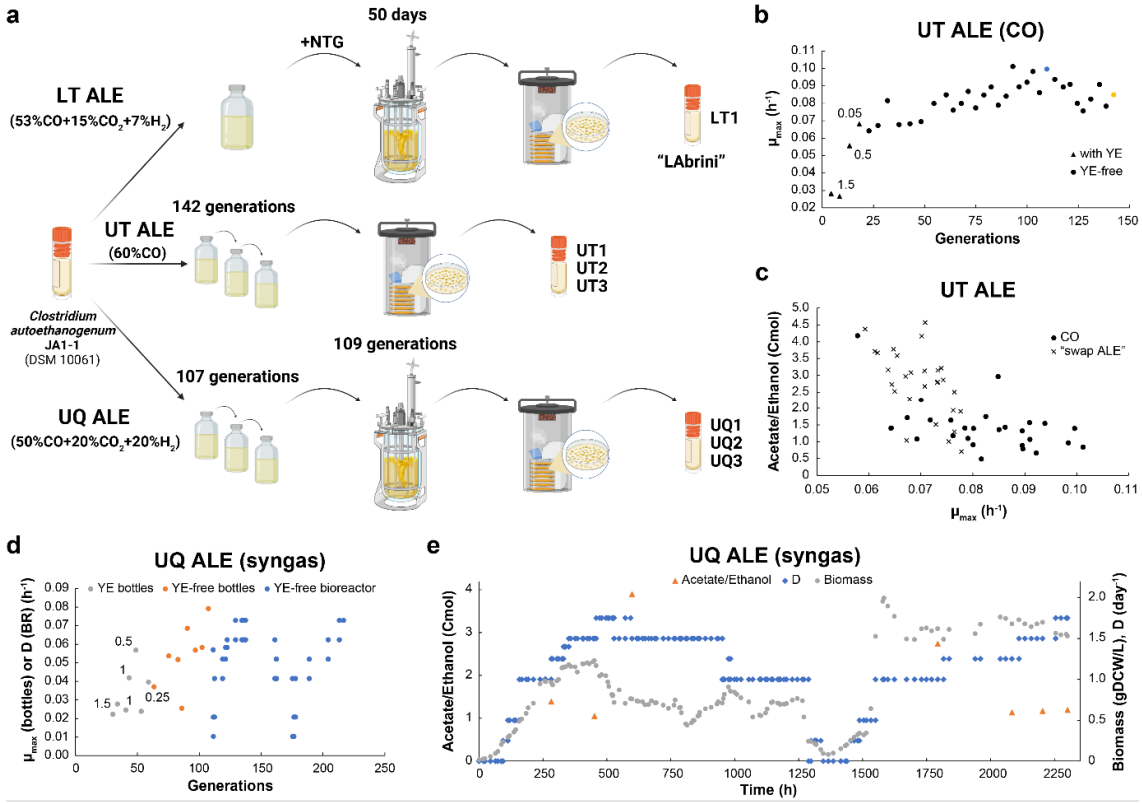


Fig. 1. Autotrophic adaptive laboratory evolution (ALE) of *Clostridium autoethanogenum* JA1-1 for improved growth phenotypes executed with three strategies across three facilities. (a) Overview of the different ALE strategies employed. See text for details. Created with BioRender.com. (b) Change of maximum specific growth rate (μ_{max}) with batch culture generations during UT ALE on CO. Blue and yellow circles denote cultures used for isolating single colonies (see text for details). For cultures with generations ~43, ~48, ~96, and ~105, μ_{max} was calculated using two data points from the exponential phase instead of using 3–5 data points as for the rest. (c) Change of acetate/ethanol (Cmol) with μ_{max} in combined batch data from UT (UT ALE and UT "swap ALE"; see text for details). (d) Change of μ_{max} in batch or dilution rate (D) in bioreactor (BR) continuous cultures with culture generations during UQ ALE on syngas. (e) Change of acetate/ethanol, biomass concentration (biomass), and D over fermentation time during the continuous culture of UQ ALE on syngas. Numbers next to triangles (panel b) or circles (panel d) denote yeast extract (YE) concentrations. gDCW, gram of dry cell weight.

In UT ALE, serial propagation of CO bottle cultures allowed the YE in the media to be removed stepwise by the fourth subculture (Fig. 1b). Notably, a sudden ~2-fold increase of μ_{max} was observed during YE removal for the third sub-culture (first bottle with reduced YE, 0.5 g/L). The population reached its peak μ_{max} of 0.10 h^{-1} (~4.6-fold higher compared to the start) by the ~93rd generation, after which a declining μ_{max} trend was observed (Supplementary Fig. 3). This phenomenon was previously observed when carrying out ALE under stable conditions (45, 46). We also tested if higher μ_{max} values could be reached when transferring the ALE culture between CO and syngas ("UT swap ALE"), but this did not occur (Supplementary Fig. 1). Intriguingly, combined data from UT (UT ALE and "UT swap ALE") revealed a declining trend of the acetate/ethanol ratio with faster growth (Fig. 1c).

383 In UQ ALE, the experiment was started with four fructose batch cultures and then transferred to
384 syngas bottles with stepwise removal of YE (Fig. 1d). The increase of μ_{max} during YE removal was not
385 as abrupt as in UT ALE but the increasing trend was clear and the final batch bottle culture with ~107
386 generations reached a μ_{max} of 0.08 h^{-1} (~3.2-fold higher compared to first syngas bottle). Then, the
387 evolved population was transferred to a chemostat bioreactor system where additional selective
388 pressure was applied by changing the dilution rate (D; i.e. μ under steady-state) for ~109 generations.
389 We obtained a stable culture at high biomass concentrations and D = 1.8 day^{-1} (μ = 0.08 h^{-1}) and the
390 UQ ALE bioreactor culture seemed to excrete less acetate compared to ethanol with faster growth
391 above D = 1 day^{-1} (Fig. 1e).
392 In LT ALE, *C. autoethanogenum* was first treated with the chemical mutagen NTG during growth
393 with YE after which the culture was used to inoculate a continuous bioreactor culture to a final D of 1
394 day^{-1} without YE for ~50 days (Fig. 1a). For all the three ALE strategies, ALEs were terminated once
395 significant changes in evolving population phenotypes were no longer observed (e.g. μ_{max} and
396 maximum biomass in bottle cultures, gas data in continuous cultures).

397 ### 3.2. Isolated ALE clones exhibit enhanced growth characteristics and metabolite production
398 patterns

399 Next, we plated the evolved mixed populations to isolate single clones for detailed characterisation
400 and for evaluating the level of phenotypic heterogeneity within the mixed populations. For UT ALE,
401 we isolated clones from both the culture with peak μ_{max} of ~0.10 (blue marker in Fig. 1b) and the last
402 ALE bottle with lower μ_{max} (yellow marker in Fig. 1b). Ten colonies were isolated for both UT and UQ
403 ALE, respectively, for preliminary comparison of growth and by-product patterns in batch cultures
404 with the same gas and media as in ALE. Notably, all isolates could be grown without the presence of
405 YE, thus fulfilling the first aim of our work. For UT ALE, we expected that isolates from the peak μ_{max}
406 population (blue marker) would grow faster than isolates from the last slower-growing population
407 (yellow marker): this was not, however, observed (Supplementary Fig. 4a). Variability in μ_{max} of UQ
408 ALE isolate batch cultures was also seen (Supplementary Fig. 4b). Similarly, phenotypic heterogeneity
409 in terms of by-product patterns within isolates of both UT and UQ ALE was observed (Supplementary
410 Fig. 4c-d). In particular, some isolates of UT ALE showed lower acetate/ethanol ratios, while three
411 isolates from UQ ALE with high ethanol production were detected. As a result, three isolates from
412 both UT (UT1, UT2, UT3) and UQ ALE (UQ1, UQ2, UQ3) were selected for further detailed
413 characterisation based on μ_{max} and intriguing metabolite production patterns in the preliminary screen
414 (indicated with asterisks on Supplementary Fig. 4). From LT ALE, one isolate termed LT1 was selected
415 for further study.

We grew the seven selected isolates in autotrophic batch bottle cultures without YE with frequent sampling to more accurately compare growth and by-product patterns. UT and UQ ALE isolates were grown on the gas they were evolved on, CO and syngas, respectively; the isolate obtained through chemical mutagenesis – LT1 – was grown on both. For comparison, the starting wild-type strain for all ALE strategies – *C. autoethanogenum* JA1-1 – was also cultured on both gases but with YE supplementation as we could not observe growth on CO without YE after 520 h. On CO, isolate UT3 had the highest μ_{max} (0.08 h^{-1}) among UT ALE while LT1 showed the fastest growth on CO with a μ_{max} of 0.10 h^{-1} (Fig. 2a). In comparison, the starting ALE strain JA1-1 showed a μ_{max} of 0.02 h^{-1} (with YE) that is ~2- to ~4-fold lower than all the evolved isolates tested on CO. During growth on syngas, UQ1, UQ3, and LT1 showed similar growth (μ_{max} ~0.08 h^{-1}) that was ~1.3- and ~3-fold higher than for UQ2 and JA1-1 (with YE), respectively (Fig. 2b). All evolved isolates grew faster than the starting strain, thus fulfilling the second aim of our work of obtaining faster-growing strains.

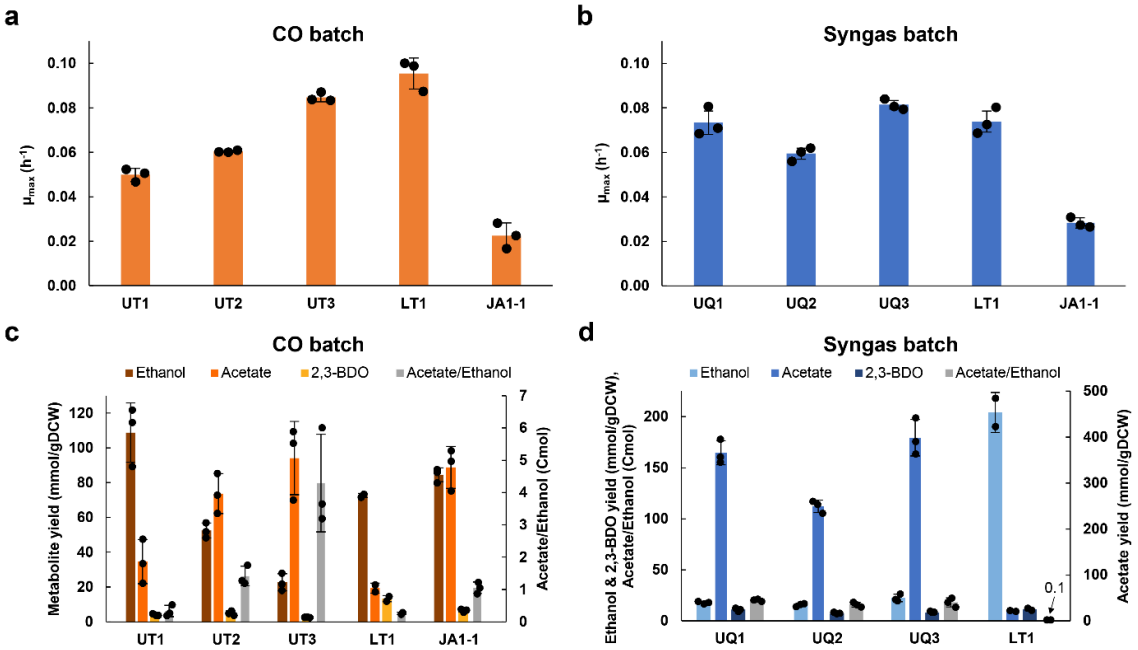


Fig. 2. Autotrophic characterisation of selected isolates from UT, UQ, and LT ALE, and the wild-type *C. autoethanogenum* starting strain of ALE (JA1-1) in batch cultures. (a) Maximum specific growth rate (μ_{max}) of UT and LT ALE isolates and JA1-1 in CO batch cultures. (b) μ_{max} of UQ and LT ALE isolate LT1 and JA1-1 in syngas batch cultures. (c) Production yields (mmol/gDCW) of growth by-products and acetate/ethanol (Cmol) of UT and LT ALE isolate LT1 and JA1-1 in CO batch cultures. (d) Production yields (mmol/gDCW) of growth by-products and acetate/ethanol of UQ and LT ALE isolate LT1 in syngas batch cultures. All data for UT, UQ, and LT1 are from yeast extract (YE)-free medium, JA1-1 data are from medium with YE. Bars show average ± standard deviation between bioreplicates, except for LT1 production yields (two bioreplicates). gDCW, gram of dry cell weight; 2,3-BDO, 2,3-butanediol.

Frequent sampling also enabled us to accurately determine production yields (mmol/gDCW) for the detected growth by-products acetate, ethanol, and 2,3-BDO. Notably, the lowest acetate/ethanol

440 ratios were detected on CO for isolates LT1 (0.27) and UT1 (0.32) that were ~4- and ~3-fold lower than
441 for JA1-1 (1.05) (Fig. 2c). While UT1 showed the highest ethanol production yield with the lowest μ_{max}
442 of the evolved strains, the fastest growing UT ALE isolate UT3 showed the highest acetate production
443 yield and acetate/ethanol of 4.14 (Fig. 2a and c). The production yield of 2,3-BDO was the highest for
444 LT1, making the strain overall the most favourable on CO in terms of reduced by-products (Fig. 2c).
445 Importantly, LT1 demonstrated preferable metabolite production also during growth on syngas in
446 comparison with UQ isolates (Fig. 2d): it showed a >6-fold higher ethanol production yield with the
447 lowest acetate production yield, translating into a low acetate/ethanol ratio of 0.10. UQ1 and LT1
448 showed the highest 2,3-BDO production yields. Interestingly, higher acetate production yields for UT
449 and UQ isolates correlate well with higher μ_{max}. At the same time, LT1 achieves fast growth with
450 significantly lower acetate production (Supplementary Fig. 5). These extracellular metabolomics data
451 support the notion that evolving acetogens in the presence of C1 substrates trigger rearrangements
452 in central carbon metabolism (18).

453 ### 3.3. Bioreactor continuous gas fermentation cultures show the superiority of isolate LT1

454 We next tested the performance and robustness of all UT ALE isolates, LT1, and UQ1 in continuous
455 gas fermentation bioreactor as this is the industrially relevant condition (7). Steady-state continuous
456 cultures also provide the necessary data to describe the cell's physiological state (e.g., μ, gas uptake,
457 by-products, carbon distribution) (47). In total, 33 biologically independent steady-states were
458 achieved across chemostat cultures for strains grown on CO or syngas at D ranging from ~0.5 to ~2
459 day^{-1} (D = μ at steady-state) (Table 1).
460

461 **Table 1.** Summary of 33 steady state chemostat cultures of UT, LT1, and UQ1 isolates.

Strain	Feed gas	Gas flow rate (mL/min)	Agitation (RPM)	D (day^{-1})	Biomass concentration (gDCW/L)	Number of replicates
UT2	CO	35	690	1.00 ± 0.00	0.52 ± 0.03	3
UT3*	CO	28	570	0.49 ± 0.01	1.03 ± 0.15	3
UT1	CO	28	570	0.53 ± 0.00	1.13 ± 0.06	3
UT2	CO	24	570	0.52 ± 0.01	0.88 ± 0.09	4
LT1	Syngas	50	675	0.99 ± 0.04	1.41 ± 0.09	4
LT1	Syngas	72	800	1.97 ± 0.11	1.50 ± 0.07	4
LT1	Syngas	35	525	0.50 ± 0.06	0.42 ± 0.05	2
UQ1	Syngas	35	525	0.50 ± 0.06	0.47 ± 0.05	2
RE2	CO	21	570	0.50 ± 0.02	0.97 ± 0.12	4
RE2	CO	45	680	0.96 ± 0.01	1.66 ± 0.07	4

462 D, dilution rate; gDCW, gram of dry cell weight. Data are average ± standard deviation between bioreplicates. Syngas
463 composition - 50% CO, 20% CO_2, 20% H_2, 10% Ar. All steady-states are biologically independent. Asterisk denotes that steady-
464 state were reached only for gas uptake and production rates.
465
466 We initially aimed to run all strains at D = 1 day^{-1} but failed after multiple attempts to obtain
467 steady-state at D = 1 day^{-1} on CO for UT2 at CO uptake rates (COUR) of ~1200 mmol/L/day and biomass
468 concentrations of ~1.5 gDCW/L. We did obtain steady-state at D = 1 day^{-1} for UT2 at COUR ~400
469 mmol/L/day, and interestingly CO uptake could only be increased after strongly decreasing the CO%
470 in the bioreactor headspace through changing gas-liquid mass transfer (Supplementary Fig. 2;
471 between ~380 to ~440 h), but no steady-state was obtained. Thus UT2 might have struggled due to
472 CO toxicity. For this "CO toxicity" steady-state of UT2 at D = 1 day^{-1} (Table 1; up to ~380 h on
473 Supplementary Fig. 2), transcriptome analysis revealed a starkly different transcriptional profile
474 compared to all other cultures (Supplementary Fig. 6). The transcriptional profile was also very distinct
475 from the unstable "non-CO toxicity" phase of the same culture operated at same D (from ~480 h on
476 Supplementary Fig. 2). Notably, we detected strong up-regulation of numerous hypothetical, putative
477 membrane, and sporulation-related proteins and down-regulation of the HytA-E/FdhA complex
478 components (Supplementary Table 2). After these challenges with UT2 cultures and gas uptake with
479 UQ1 at D = 1 day^{-1}, we used D = 0.5 day^{-1} to compare strains across gases.
480 For UT3 though, we could obtain steady-state data at D = 0.5 day^{-1} only regarding gas uptake and
481 production since acetate and ethanol production were highly sensitive to any operational activities to
482 the culture (e.g. halting agitation for a volume check, sampling 1% of culture volume). Furthermore,
483 adjustment of gas-liquid mass transfer was needed even after obtaining target agitation and gas flow
484 (Fig. 3). In striking comparison, operating LT1 cultures was straightforward. Thus we additionally
485 achieved steady-states at D = 1 and 2 day^{-1} at high COUR and biomass concentrations (Fig. 3). The
486 COUR for LT1 growing on syngas at D = 1 day^{-1} is triple of the steady-state value of UT2 on CO
487 (Supplementary Fig. 7). Moreover, faster gas-liquid mass transfer ramping and higher Ds also meant
488 that two steady-states were achieved for LT1 with half the time it took to achieve one steady-state for
489 UT3 (Fig. 3). The robustness of LT1 is highly relevant for the acetogen gas fermentation field as it makes
490 it significantly less challenging to operate continuous gas fermentation cultures to obtain high-quality
491 data. Thus a robust strain for bioreactor continuous gas fermentation was obtained that fulfils the
492 third aim of our work.
493

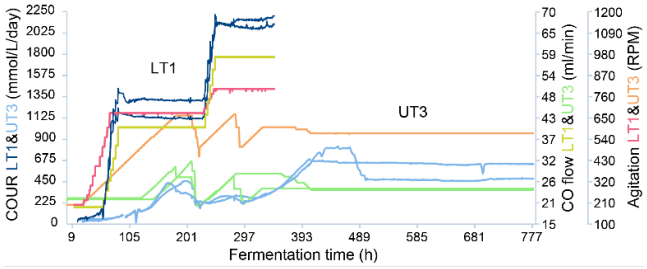


Fig. 3. Robustness of strain LT1 in terms of operation of bioreactor continuous gas fermentation cultures. LT1 data are for the chemostats that achieved steady-states at specific growth rates (μ) of 1 and 2 day^{-1}. See Table 1 for more details on LT1 and UT3 chemostats. COUR, CO uptake rate (mmol/L/day); RPM, rounds per minute.

A comparison of steady-state data from the bioreactor runs demonstrates that various metabolic rearrangements took place during the three ALE strategies (Fig. 4). Different from the batch data (Fig. 2), UT strains showed similar growth characteristics in CO chemostats (D = 0.5 day^{-1}) except for 2,3-BDO production (Fig. 4). While the acetate concentration for UT2 was lower most likely due to lower biomass concentration (Fig. 4a), specific by-product production rates (Fig. 4b) and specific gas rates (Fig. 4c) showed similar values, except for the specific production rate of 2,3-BDO ($q_{2,3\text{-BDO}}$; mmol/gDCW/day). Chemostat results were different from batch data also when comparing UQ1 and LT1 in syngas chemostats at D = 0.5 day^{-1} (Fig. 4). The two strains showed very similar performance at the same D both in terms of by-product production and gas uptake and production. Notably, LT1 outperformed all strains concerning acetate/ethanol ratios by showing values <1 at D = 1 and 2 day^{-1} (Fig. 4a). The ability to achieve higher biomass concentrations for LT1 D = 1 and 2 day^{-1} cultures also translated into the highest ethanol and 2,3-BDO levels (Fig. 4a) and due to faster growth also to higher specific production rates (q_{EtOH} and $q_{2,3\text{-BDO}}$) (Fig. 4b). LT1 cultures at D = 1 and 2 day^{-1} also show the highest carbon flux to the reduced ethanol and 2,3-BDO products across the various strains and steady-states (Fig. 5). Similarly to the latter strain, the specific H_2 uptake rate (q_{H2}) for LT1 did not increase proportionally with D (Fig. 4c), suggesting another avenue for improving *C. autoethanogenum* performance through ALE, i.e. improving H_2 uptake.

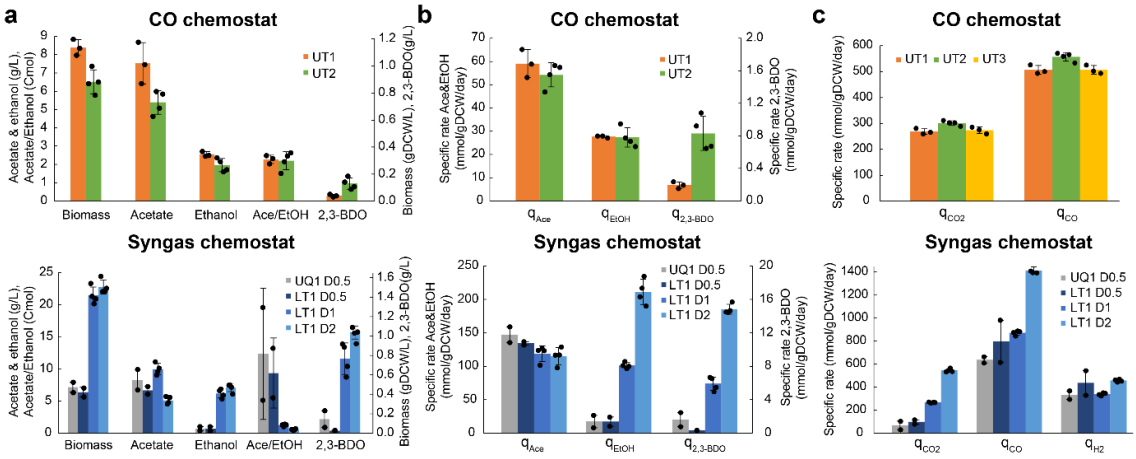


Fig. 4. Characterisation of selected isolates from UT, UQ, and LT ALE in autotrophic chemostat cultures. (a) Biomass and by-product concentrations and acetate/ethanol (Ace/EtOH). (b) Specific by-product production rates (mmol/gDCW/day). (c) Specific gas uptake (CO, H_2) and production (CO_2) rates. CO chemostat data for UT1 and UT2 are from dilution rate (D) 0.5 day^{-1}; for syngas chemostat data, the number following D denotes D value in day^{-1}. UT3 reached steady-state only regarding gas uptake and production. Bars show average ± standard deviation between four (UT2, LT1 D1, LT1 D2), three (UT1), or two (UQ1 D0.5, LT1 D0.5) bioreplicates. See Table 1 for details. gDCW, gram of dry cell weight; 2,3-BDO, 2,3-butanediol; q, specific rate.

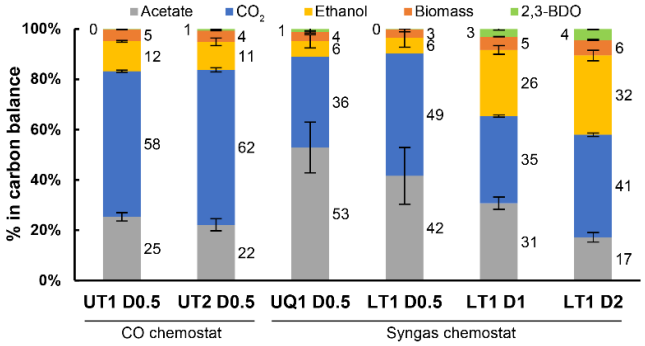


Fig. 5. Carbon balances of selected isolates from UT, UQ, and LT ALE in autotrophic chemostat cultures. The number following D (dilution rate) denotes D value in day^{-1}. Carbon flow into CO_2 for UQ1 D0.5 and LT1 D0.5 were calculated by subtracting measured carbon flows into other products and biomass from 100%. Carbon recoveries for UT strains and LT D1 and D2 were normalised to 100% to have a fair comparison of carbon distributions between different conditions. Bars show average ± standard deviation between four (UT2, LT1 D1, LT1 D2), three (UT1), or two (UQ1 D0.5, LT1 D0.5) bioreplicates. See Table 1 for details. 2,3-BDO, 2,3-butanediol.

The evolved strain LT1 fulfilled all the aims of our work to create an improved model-acetogen strain available to the public: it grows without complex nutrients on minimal medium, shows high gas uptake enabling faster growth, and is robust in terms of operating continuous bioreactor gas fermentation cultures. Additionally, the metabolic feature of high carbon flux into reduced growth by-products makes LT1 an attractive chassis for metabolic engineering. We have thus deposited the LT1 strain at the DSMZ strain collection (DSM 115981) for public use and propose to make it the model-

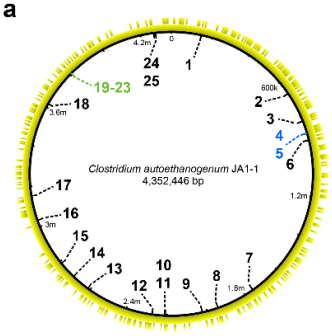
538 acetogen strain named *C. autoethanogenum* LAbrini (i.e. Laboratory Adapted Abrini). The improved
539 characteristics of LAbrini hold promise for speeding up phenotyping and engineering of acetogens. For
540 example, we currently map genotype-phenotype links for every gene in LAbrini using a systems-level
541 approach.

542 ### 3.4. Genetic mutations associated with improved phenotypes

543 We next used whole-genome DNA sequencing (WGS) to identify genetic mutations during the three
544 ALE strategies, which may have led to different phenotypes. WGS data also allows the comparison of
545 genetic variability between the evolved strains. We sequenced the starting strain JA1-1 and seven key
546 evolved strains (UT1, UT2, UT3, UQ1, UQ2, UQ3, and LT1) and performed mutation analysis after high-
547 quality genome assemblies (see Material and Methods). The starting strain for our three ALE strategies
548 was sequenced to ensure we reliably detected genetic mutations that appeared during ALE instead of
549 relying only on the published genome sequences for *C. autoethanogenum* JA1-1 (33, 35).

550 In total, 25 mutations were detected across the evolved strains, with 21 affecting protein-coding
551 sequences (CDSs) (Fig. 6; Supplementary Table 6). 17 of the 25 mutations were single nucleotide
552 polymorphisms (SNPs). Eight were deletions, while four being ≥ 168 bp (Fig. 6b). Two hot-spots were
553 detected in the genome that accrued mutations during multiple ALE strategies: mutations #19–23
554 around genome position 548,800 (all in the same CDS) and mutations #4–5 around position 3,469,400
555 (both in the same CDS) (Fig. 6). These two genes are examples of convergent evolution under stronger
556 autotrophic selection pressure across different ALE strategies. Principal component analysis (PCA) of
557 all the WGS data confirmed that strains isolated from the same population (from UT or UQ ALE) are
558 closely related (Supplementary Fig. 8). Comparison of the genomes of the superior strain LT1 and
559 starting strain JA1-1 showed similar but not identical values: length of 4,349,910 nt, 3,942 CDSs, and
560 31.1% GC content for LT1; and 4,352,446 nt; 3,964 CDSs; 31.1% GC content for JA1-1.

a



b

Strain	Mut#	Position	Mutation	FC	AAC	Gene ID	Description of protein product
LT1	1.	4,203,213	Δ2,717 bp	CDS		CLAU_3832-3830	Pyrimidine-nucleoside phosphorylase (3832); phosphopentomutase (3831); nucleoside transporter domain protein (3830)
	5.	3,469,046	C→A	CDS	A11E	CLAU_3129	Sporulation transcriptional activator Spo0A
	13.	1,744,062	G→T	CDS	*401L	CLAU_1579	Carbon-monoxide dehydrogenase (acceptor)
	14.	1,651,616	G→T	CDS	S207I	CLAU_1506	Ribonuclease PH
	25.	81,706	A→G	NC			
UT shared	3.	3,533,188	G→A	CDS	E194K	CLAU_3194	Histidine kinase
	9.	2,333,941	G→T	NC			
	16.	1,323,289	T→C	NC			
UT1	10.	2,150,981	G→A	CDS	G94D	CLAU_1957	Two component transcriptional regulator winged helix family
	19.	549,451	Δ231 bp	CDS		CLAU_0471	Amino acid permease
UT2	11.	2,146,875	(A)9→8	CDS		CLAU_1953	Transcription-repair coupling factor
	23.	548,344	G→T	CDS	E392*	CLAU_0471	Amino acid permease
UT3	7.	2,605,812	G→A	CDS	G87R	CLAU_2375	Putative signal transduction protein with CBS and DRTGG domain-containing protein
	22.	548,364	Δ527 bp	CDS		CLAU_0471	Amino acid permease
UQ shared	4.	3,469,766	C→A	CDS	P251H	CLAU_3129	Sporulation transcriptional activator Spo0A
	12.	2,083,514	(CAGCAC) 2→1	CDS		CLAU_1904	Ribosomal protein S12
	24.	86,954	(T)8→7	NC			
UQ1	6.	3,426,938	Δ168 bp	CDS		CLAU_3080-3079	Pseudouridine synthase (3080); lipoprotein signal peptidase (3079)
	15.	1,565,636	G→T	CDS	K56N	CLAU_1423	Ferric uptake regulator Fur family
	17.	1,200,575	C→A	CDS	A54E	CLAU_1076	Cobalamin synthesis protein cbiG
	18.	733,832	(18-bp) 5→2	CDS		CLAU_0649	Putative secreted protein containing bacterial Ig-like domain
	21.	548,603	T→G	CDS	C305W	CLAU_0471	Amino acid permease
UQ2	**8.**	**2,414,456**	**A→G**	**CDS**	**T25A**	**CLAU_2202**	**Transcriptional regulator AbrB family**
	21.	548,603	T→G	CDS	C305W	CLAU_0471	Amino acid permease
UQ3	2.	3,683,989	C→A	CDS	S195Y	CLAU_3345	Methyl-accepting chemotaxis protein
	8.	**2,414,456**	**A→G**	**CDS**	**T25A**	**CLAU_2202**	**Transcriptional regulator AbrB family**
	20.	549,237	G→A	CDS	W94*	CLAU_0471	Amino acid permease

Fig. 6. Whole-genome sequencing of the selected isolates from UT, UQ, and LT ALE and identified mutations relative to the wild-type *C. autoethanogenum* starting strain of ALE (JA1-1). (a) Distribution of the identified 25 mutations across the circular genome of JA1-1. The number denotes mutation shown in panel b (column Mut#) and colour matches the shading of mutation in panel b (exceptions are black numbers with white shading in panel b). Yellow layers are protein-coding sequences (CDS) on forward (inner) and reverse (outer) strands. (b) Details of the 25 mutations identified in the seven ALE strains. Mutations with colour shading appeared in the same genes across multiple ALE strategies, i.e. convergent mutations, while the non-convergent mutation in bold font (#8) appeared in two isolates within one ALE strategy (details shown only for the first occurrence); shading colours match colour of numbers (Mut#) on panel a. The bases within the parentheses in column "Mutation" denote the repeat sequence that decreased from the number of repeats indicated on the left of the arrow (→) to the number indicated on the right. See Supplementary Table 6 for additional details on mutations. Mut#, mutation number; FC, feature class; AAC, amino acid change; Gene ID, gene identifier; SNP, single nucleotide polymorphism; Δ, deletion; NC, non-coding; *, stop codon.

The first hot-spot at ~548,800 position is a proposed amino acid permease (CLAU_0471; function predicted based on conserved domain PotE) that was affected by two deletions and three different SNPs: one 231 bp deletion in UT1 and one 527 bp deletion in UT3; two hits in UQ isolates and one in UT2 (mutations #19–23 in Fig. 6b and Supplementary Table 6). Interestingly, nucleotides in the ALE starting strain JA1-1 at these three SNP positions of CLAU_0471 are fully conserved across homologs of Clostridia species (Supplementary Fig. 9). The second hot-spot at position ~3,469,400 is the sporulation transcription factor *spo0A* (CLAU_3129) that was affected by a G-to-T SNP at two different locations in UQ and LT ALE (mutations #4–5 in Fig. 6b and Supplementary Table 6). Similarly to the mutations in the first hot-spot (Supplementary Fig. 9), nucleotides in the starting strain JA1-1 at both

584 positions of the second hot-spot are highly conserved across homologs of Clostridia species
585 (Supplementary Fig. 10). UQ1 acquired a 168 bp deletion (mutation #6) across genes encoding for a
586 lipoprotein signal peptidase (CLAU_3079) and a pseudouridine synthase (CLAU_3080) that are
587 involved in sporulation (48) and pyrimidine metabolism, respectively (Fig. 6b and Supplementary Table
588 6). Interestingly, the latter gene was upregulated under butanol stress in a *spo0A* overexpression
589 mutant strain in *Clostridium acetobutylicum* (49). Pyrimidine metabolism seems to have played an
590 important role in improving strain phenotypes in our ALE approaches as the large 2,717 bp deletion in
591 LT1 (mutation #1) resulted in the deletion of two genes (CLAU_3830 and 3831) and truncation of one
592 gene (CLAU_3832) that are associated with pyrimidine metabolism (Fig. 6b and Supplementary Table
593 6).

594 Among the other spatially dispersed mutations in the genome, the mutation potentially most
595 relevant to the improved phenotype of LT1 is a SNP (mutation #13) in the carbon monoxide
596 dehydrogenase gene AcsA (CLAU_1579) that encodes for the CODH subunit of the CODH/acetyl-CoA
597 synthase (ACS) protein complex (Fig. 6b and Supplementary Table 6). Mutation #13 removes the AcsA
598 stop codon, resulting in a putative translational fusion product. Interestingly, while NTG treatment
599 during LT ALE did not result in significantly more mutations for LT1 than what appeared during UT and
600 UQ ALE, all but one mutation appeared in unique genes compared to UT and UQ strains (Fig. 6b). The
601 only CDS mutation (#3) present in all UT ALE strains was found in a predicted histidine kinase
602 (CLAU_3194), which is interesting as some histidine kinases regulate sporulation and metabolism in
603 Clostridia (50). Notably, one SNP appeared during UQ ALE in a usually highly conserved position in
604 Clostridia species (Supplementary Fig. 11) - mutation #8 in an AbrB family transcriptional regulator
605 (CLAU_2202) (Fig. 6b and Supplementary Table 6). Interestingly, CLAU_2202 is among the top 100
606 abundant proteins in *C. autoethanogenum* (51). Other potentially relevant mutations in CDSs
607 appeared in a ribosomal protein S12 (mutation #12), and transcription-associated genes (mutations
608 #10, #11, and #15). A homolog of the Ferric uptake regulator of Fur family (CLAU_1423) that acquired
609 mutation #15 in our work – the putative transcriptional regulator PerR – has acquired numerous
610 mutations within the ALE experiments compiled in the ALEdb database (52). Future advances in
611 quantifying genotype-to-phenotype links in acetogens will assist in interpreting the genetic changes
612 we detected and improve experimental evidence-based (vs. computationally derived) annotation of
613 *C. autoethanogenum* genome.

614 ### 3.5. Reverse genetic engineering recovers superior features of our ALE strains

615 The causality of mutations observed during ALE can be evaluated by reconstructing the genetic
616 modifications in the starting strain of ALE, i.e. using reverse genetic engineering. Observation of
617 convergent evolution by appearance of mutations in *spo0A* during both LT and UQ ALE (Fig. 6b) is likely

618 significant due to the role of *spo0A* as a master regulator of sporulation and solventogenesis in
619 Clostridia (53–56). Interestingly, mutation #10 observed during UT ALE (in UT1) occurred in a two-
620 component transcriptional regulator CLAU_1957 (based on GenBank CP012395.1 annotation; (33))
621 that could be potentially involved in sporulation processes (e.g. phosphorelay, sporulation activator;
622 (57)) based on predictions using protein sequence or structure homology by DeepFRI (58), ProteinInfer
623 (59), and UNIPROT database (60) (Supplementary Table 7; UniProtKB U5RXC5). Furthermore, this
624 mutation is predicted to have occurred in a response regulator receiver domain (one of the two
625 predicted domains) (61). Therefore, we targeted to mutate *spo0A* or CLAU_1957 in our ALE starting
626 strain JA1-1 to investigate whether alterations to cellular sporulation processes could, at least
627 partially, replicate the improved phenotypes obtained during our ALE: faster growth, growth without
628 complex nutrients, and robustness in continuous culture operation.

629 We used CRISPR/Cas9-aided homologous recombination (28) for reverse genetic engineering of
630 JA1-1. Mutations #4 and #5 occurred across LT and UQ ALE (Fig. 6b) in the linker domain between the
631 activator domains and in the receiver domain of Spo0A, respectively (62). Mutations in the same
632 domains in other Clostridia have led to the same phenotype as *spo0A* deletion, i.e. non-sporulating
633 cells (63–65). We thus aimed to investigate the effects of both mutations #4 and #5 through *spo0A*
634 deletion. Both of the two sgRNAs we used yielded 100% gene deletion efficiency as *spo0A* was deleted
635 in all screened colonies (Supplementary Fig. 12 and Supplementary File 4 and 5); this strain was named
636 RE1. Mutation #10 was a SNP in CLAU_1957, and we identified only one potential sgRNA that would
637 realize the G-to-A mutation and disrupt the PAM sequence, thus preventing Cas9 from cleaving the
638 target site again after successful homologous recombination. This sgRNA yielded 100% gene-editing
639 efficiency as all screened colonies showed the desired SNP (Supplementary File 6 and 7); this strain
640 was named RE2. Strikingly, RE1 autotrophic batch bottle cultures without YE showed >4-fold faster
641 growth than JA1-1 with YE and similar values on both CO and syngas as LT1 (Fig. 7a and Fig. 2). We
642 cultured RE2 on CO, i.e. same gas feed as used for UT1 characterization (ALE strain with same SNP as
643 RE2), and also this reverse engineered strain displayed significantly faster growth (>3-fold) without YE
644 compared to JA1-1 with YE and similar μ_{max} to UT1 (Fig. 7a and Fig. 2).

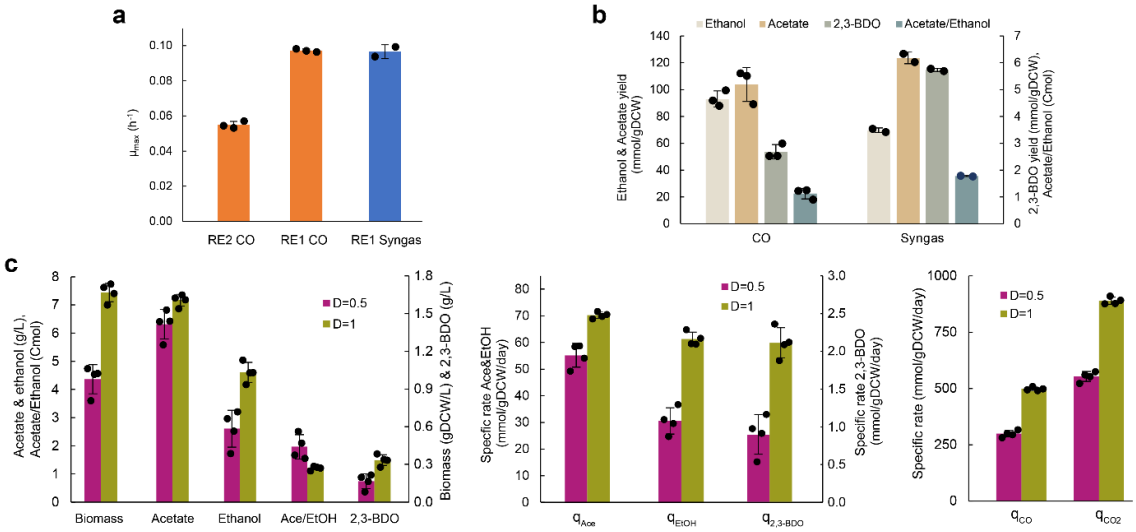


Fig. 7. Autotrophic characterization of reverse engineered strains RE1 and RE2 in batch and chemostat cultures. (a) Maximum specific growth rate (μ_{max}) in bottle batch cultures with yeast extract (YE)-free medium. (b) Production yields (mmol/gDCW) of growth by-products and acetate/ethanol (Ace/EtOH) of strain RE1 in bottle batch cultures. (c) RE2 growth characteristics in CO chemostats at dilution rate (D) 0.5 and 1 day^{-1}. See Fig. 4 legend for parameter details. Bars show average ± standard deviation between four (panel c), three (RE2 CO, RE1 CO, and CO data on panel b), or two (RE1 Syngas and Syngas data on panel b) bioreplicates. gDCW, gram of dry cell weight; 2,3-BDO, 2,3-butanediol; q, specific rate.

Quantification of by-product patterns of RE1 in autotrophic bottle cultures (Fig. 7b) showed similar yields to JA1-1 on CO (see Fig. 2c; growth with YE) and higher acetate/ethanol ratios compared to LT1 (see Fig. 2c and d) while ethanol and acetate production did not coincide in RE1. The performance of RE2 was tested in bioreactor continuous cultures as we experienced significant difficulties with growing UT strains to high biomass concentrations on CO (see section 3.3 above). We efficiently operated the culture at D = 0.5 day^{-1} and achieved steady-state at high gas uptake rates and biomass concentrations at D = 1 day^{-1} (double of UT and UQ strains; see Table 1) (Fig. 7c). The robustness of RE2 in continuous cultures was surprising since just one SNP was introduced into the JA1-1 background. Despite the significant difference in bioreactor operation between UT1 and RE2 cultures, the two strains displayed somewhat similar growth characteristics (e.g. biomass concentration, specific by-product production rates, specific gas rates) at D = 0.5 day^{-1} except >4-fold higher 2,3-BDO production for RE2 (Fig. 4 and 7c). Comparison of RE2 performance at D = 1 day^{-1} is only possible with LT1 data (Table 1). RE2 showed lower production of acetate and ethanol (Fig. 4 and 7c), though LT1 cultures were fed with syngas (RE2 with CO). This is also reflected in the carbon balance where significantly more carbon is diverted to CO_2 for RE2 (Supplementary Fig. 13 and Fig. 5), which is consistent with another strain of *C. autoethanogenum* diverting more carbon to CO_2 on CO compared to syngas (9). Higher D led to similar effects for RE2 and LT1 (Fig. 4 and 7c). In conclusion,

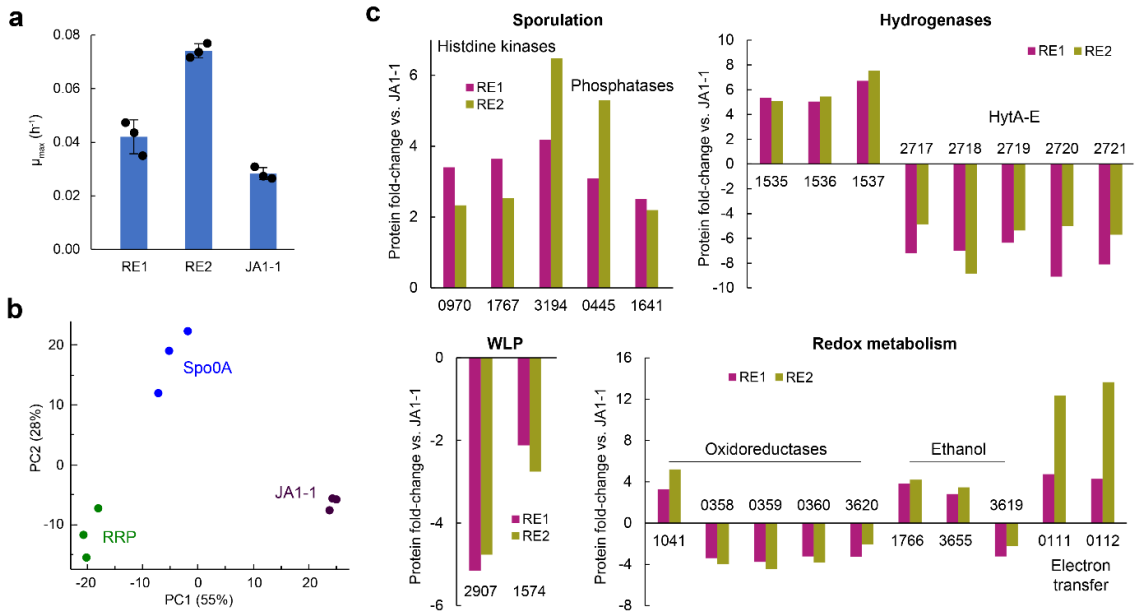
671 reconstructing mutations in two sporulation-related genes was sufficient to recover all three superior
672 features of our ALE strains.

673 We tested next if the reverse engineered mutations in sporulation-related genes might have
674 affected sporulation in the strains. For this, we cultured RE1, RE2, and JA1-1 both autotrophically and
675 heterotrophically and used microscopy to detect potential differences in sporulation from stationary
676 phase samples. We also plated heterotrophic cultures before and after a heat-shock with the aim to
677 observe potential heat-resistant spores. Neither microscopy (Supplementary Fig. 14 and 15) nor
678 plating (Supplementary Fig. 16) identified any spores or sporulating cells for any of the strains
679 (including JA1-1). Although the authors who isolated *C. autoethanogenum* JA1-1 report it as a spore-
680 former (27), a recent work also failed to detect JA1-1 forming spores (66). The latter and our
681 observations are supported by a review on Clostridial sporulation concluding that sporulation is rarely
682 observed in acetogens (62).

683 Mutations in sporulation-related genes have led to significant effects on Clostridial phenotypes
684 other than sporulation (62). We thus profiled the proteomic response of JA1-1 to the reverse
685 engineered mutations as functional genomics has been informative for interpreting genotype-
686 phenotype links in engineered strains. For this, JA1-1, RE1, and RE2 were grown in syngas bottle
687 cultures with YE in the media as JA1-1 does not grow without YE. We note that these proteome
688 expression patterns are affected by variable μ_{max} values (Fig. 8a) and potential differences in
689 consumption of YE components between the strains. Our analysis identified 284 differentially
690 expressed proteins (fold-change > 2 and q-value < 0.05) between RE1 and JA1-1, and 514 between
691 RE2 and JA1-1 (Supplementary Table 5). Proteomic responses to either mutation were similar in
692 general (Fig. 8b; Supplementary Fig. 17) and shared 205 differentially expressed proteins potentially
693 due to the 66-fold down-regulation of Spo0A in RE2 (Fig. 8c; Supplementary Table 5). Notably,
694 peptides for five sporulation-related proteins highly expressed in JA1-1 (CLAU_1942, 1946, 3117,
695 3233, 3248) were not detected in either engineered strain (also not for Spo0A in RE1). At the same
696 time, three histidine kinases and two phosphatases potentially linked to Spo0A regulation were up-
697 regulated in both RE1 and RE2 (Fig. 8c), including CLAU_3194 that was mutated in all UT isolates
698 (mutation #3 in Fig. 6b). Phosphate metabolism was further affected by 12 differentially expressed
699 kinases not shared between the two strains (Supplementary Table 5).

700 Regarding functionalities specifically relevant for acetogens, repression of the electron-bifurcating
701 hydrogenase HytA-E complex was potentially compensated by induction of hydrogenase CLAU_1535
702 and its associated proteins in both strains (Fig. 8c). In addition, a formate dehydrogenase (CLAU_2907)
703 and the methylene-THF dehydrogenase (CLAU_1574) in the WLP were down-regulated in both strains.
704 Also redox metabolism was markedly affected by both mutations as oxidoreductases, proteins

705 involved in ethanol metabolism and electron transfer were differentially expressed in both strains
706 compared to JA1-1 (Fig. 8c). The necessary transcriptional changes needed to recover the superior
707 phenotypes of our ALE strains through reverse engineering could potentially have been realised
708 through 2-to-10-fold up-regulation of seven and/or 2-to-6-fold down-regulation of two transcriptional
709 regulators shared for both strains (Supplementary Table 5). Mutations in either sporulation-related
710 gene thus lead to significant proteomic rearrangements in *C. autoethanogenum*.



711

712 **Fig. 8.** Proteomic responses of the wild-type *C. autoethanogenum* starting strain of ALE (JA1-1) to the reversed engineered
713 mutations in syngas bottle cultures. (a) Maximum specific growth rate (μ_{max}) for reverse engineered strains RE1, RE2, and
714 JA1-1 with yeast extract (YE) in medium. Bars show average ± standard deviation between three bioreplicates. (b) Principal
715 component analysis (PCA) of protein MS intensities for the three bioreplicate cultures of RE1, RE2, and JA1-1. PC, principal
716 component. (c) Potential key proteins differentially expressed (fold-change > 2 and q-value < 0.05) for both RE1 vs JA1-1 and
717 RE2 vs JA1-1 comparisons. Protein IDs are preceded with CLAU_. See Supplementary Table 5 for data.

718

719 ## 4. DISCUSSION

720 While gas fermentation using *C. autoethanogenum* has been deployed commercially for ethanol
721 production (7), non-commercial acetogen strains struggle to grow on minimal medium, show slow gas
722 uptake and growth, and are challenging to run in continuous bioreactor cultures. To overcome these
723 limitations, we used three different and independent ALE strategies in this work to obtain strains with
724 improved phenotypes of which the strain named LAbrini (LT1) showed superior performance in
725 acetogen gas fermentation. We characterised the evolved strains using bottle and bioreactor cultures,
726 gas analysis, metabolomics, transcriptomics, and DNA sequencing. Finally, reverse genetic engineering
727 of sporulation-related mutations recovered all three superior features of our ALE strains.

728 During one of the ALEs (UT), we noticed a declining μ_{max} during the last ⅓ of the ALE (Fig. 1b). This
729 implies that evolutionary pressures other than towards growing faster were involved in this ALE
730 approach of serial batch culture propagation. Similar trends have been seen before (45, 46).
731 Additionally, UT ALE populations showed a declining trend of acetate/ethanol with faster growth (Fig.
732 1c). Some isolates of the ALE populations of this work displayed faster growth coupled with higher
733 acetate production (Supplementary Fig. 5). These observations at isolate and population levels suggest
734 that cells can find various metabolic routes across different ALE approaches to achieve faster growth,
735 as noted before (67, 68). We hypothesise that cross-feeding between subpopulations took place
736 similar to previous bacterial ALEs (69–72). Future co-culturing experiments of isolates with different
737 phenotypes (e.g. high and low acetate producers) might shed light on these possibilities.
738 Characterisation of evolved strains in batch and chemostat cultures revealed that batch bottle
739 cultures might fail to predict performance in continuous cultures. For example, UT1 and UT2 showed
740 nearly identical q_{EtOH} and acetate/ethanol ratios in chemostats after very different results in batch
741 cultures and only LT1 could thrive above D = 0.5 day^{-1} while most evolved strains showed similar μ_{max}
742 in batch cultures, which was > 0.5 day^{-1}. Furthermore, efforts to operate UT1 chemostats at higher D
743 and gas-liquid mass transfer seemed to have been compromised by CO toxicity (see below), which
744 was not present in batch cultures with lower gas-liquid mass transfer. These results emphasise the
745 importance of testing novel strains (e.g. acquired using ALE or metabolic engineering) in industrially-
746 relevant continuous fermentation conditions (7). We also note that the isolation of LT1 from a
747 chemostat culture after NTG mutagenesis could have contributed towards LT1 superior performance
748 in our bioreactor continuous cultures. Thus, isolating evolved strains from operational conditions
749 similar to those where the strain would eventually be used for scientific investigation or industrial
750 exploitation could be valuable. When comparing our different ALE approaches, the combination of
751 NTG with selective pressure during continuous cultivation in LT ALE proved to be the most successful
752 strategy. NTG treatment presumably introduced high genetic diversity, setting the stage for
753 subsequent selection under operational conditions. Indeed, LT1 was superior to UQ ALE isolates that
754 were also obtained from chemostats but without prior treatment with a mutagen. These results
755 support the effectiveness of NTG-induced mutagenesis combined with selective pressure in microbes
756 (73, 74).
757 Transcriptome analysis of the potential "CO toxicity" state of UT1 chemostat at D = 1 day^{-1}
758 revealed a distinct transcriptional profile compared to other cultures, including UT1 in a "non-CO
759 toxicity" but unstable culture at D = 1 day^{-1}. UT1 in the "CO toxicity" state showed strong up-regulation
760 of numerous hypothetical and sporulation-related proteins (Supplementary Table 2). These proteins
761 serve as promising candidates for potential overexpression to play a role in mitigating CO toxicity.

762 Intriguingly, we also observed ~2-fold down-regulation of components of the HytA-E/formate
763 dehydrogenase (FdhA) complex (75), which plays a key role in CO_2 reduction to formate in *C.*
764 *autoethanogenum* using NADPH and reduced ferredoxin (Fd) or H_2 (29, 75, 76). Wang and co-workers
765 who purified the complex hypothesised that it might also protect cells from over-reduction of NADP
766 and Fd pools during growth on CO (75). They propose that lowered CO oxidation increases the CO
767 concentration in cells that inhibits CO_2 reduction to formate by the complex that lowers reoxidation
768 of Fd. The solution for avoiding this over-reduction state is NADPH and Fd reoxidation by HytA-E
769 through proton reduction to H_2. Our results are consistent with their hypothesis as we detected H_2
770 production when trying to improve COUR of the UT1 "CO toxicity" culture by decreasing gas-liquid
771 mass transfer, i.e. CO oxidation (Supplementary Fig. 2). We thus hypothesise that such regulation
772 could be linked to the down-regulation of HytA-E/FdhA complex component genes. The metabolism
773 of *C. autoethanogenum* thus displays reactional plasticity in choosing electron sinks to drive CO_2
774 reduction as proposed before (75, 77). Such flexibility would realise growth advantages under
775 changing environmental conditions.

776 Our ALE experimental data analysis confirmed previous observations that different ALE
777 approaches often lead to similarly improved phenotypes but through various underlying genetic
778 changes (67–69, 71, 72). Interestingly, a proposed amino acid permease (CLAU_0471) was affected in
779 all the isolated strains in two different ALE approaches (UT and UQ) by two distinct deletions and three
780 SNPs. It is unclear why a proposed amino acid permease was under such strong selective pressure
781 when the complex nutrient YE was removed from the medium. The other gene that was mutated
782 during two of the ALEs (LT and UQ) was *spo0A* (CLAU_3129) which is considered a master regulator of
783 sporulation and solventogenesis in Clostridia (53–56). Down-regulation or inactivation of *spo0A* in *C.*
784 *acetobutylicum* causes a reduction in sporulation and increased solvent production (78, 79). However,
785 in other Clostridia Spo0A can improve fitness (65) and regulate metabolic and virulence factors outside
786 the sporulation process (80). Interestingly, mutation #10 in a predicted phosphorelay regulator
787 (CLAU_1957) and mutation #3 in a proposed histidine kinase (CLAU_3194) could be linked to affect
788 sporulation in UT1 as latter regulators are involved in phosphorylation circuits involving histidine
789 kinases (81) and as similar kinases regulate sporulation and solventogenesis in *Clostridium beijerinckii*
790 and *C. acetobutylicum* (50, 82). Additionally, both histidine kinases (53, 54) and CLAU_1957 (83) could
791 affect Spo0A directly. Mutations in two-component sensor systems composed of histidine kinases and
792 response regulators that appeared in ALE experiments with other bacteria were postulated to be
793 related to general adaptation to the growth environment (84). Sporulation could have also been
794 affected during UQ ALE (in UQ1) by truncation of a lipoprotein signal peptidase (CLAU_3079), which
795 is involved in sporulation of *Clostridium difficile* (48) and deletion of the gene pseudouridine synthase

796 (CLAU_3080) that was upregulated in a *spo0A* overexpression mutant strain in *C. acetobutylicum* (49).
797 Notably, protein expression of CLAU_3080 was up-regulated 3-fold in our RE2 strain (Supplementary
798 Table 5). Both reverse engineered mutations triggered widespread proteomic rearrangements
799 demonstrating their effects beyond the sporulation process. Similarly, altered expression of genes in
800 Clostridial sporulation networks regulate gene expression across various metabolic functions (54, 62).
801 We hypothesise that mutations in the above-described genes could have emerged from selective
802 pressure towards faster growth since suppressing sporulation functionalities would save cells valuable
803 energy to grow faster, for instance, through conserving a phosphate group (54). This is consistent with
804 our reverse engineering results of strains with mutations in *spo0A* or CLAU_1957 growing significantly
805 faster than wild-type *C. autoethanogenum*. The specific regulatory mechanisms governing the
806 proteome changes that realised the superior phenotypes remain to be elucidated.

807 The CODH/ACS complex realises CO oxidation/CO_2 reduction and acetyl-CoA synthesis (85), and
808 its activity is essential for the autotrophic growth of *C. autoethanogenum* (86). The *acsA* (CLAU_1579)
809 stop codon removal in strain LT1 has been reported in an independent study that observed
810 translational readthrough at the protein level (86). The *C. autoethanogenum* AcsA protein is uniquely
811 truncated compared to other acetogens. The other potential key genetic modification in LT1 seems to
812 be associated with pyrimidine metabolism evidenced by the large 2,717 bp deletion affecting three
813 genes in pyrimidine metabolism. This could realise a faster supply of RNA monomers, allowing higher
814 translation rates and faster growth. Thus, the superior performance of LAbrini in continuous
815 bioreactor cultures and its ability to grow in the absence of YE is due to its genotype and the selection
816 strategy employed in its development. While growth without YE has been reported for acetogens
817 before (25, 26), it is unclear if these phenotypes resulted from physiological adaptation or evolution.
818 Further studies are required to link mutations to its phenotype, key structural/functional attributes of
819 the ACS/CODH enzyme complex, and regulation.

820

821 ## CONCLUSIONS

822 Our work is important for two reasons. First, we offer a superior model-acetogen strain – LAbrini
823 – for the community that grows fast without complex nutrients, is robust for continuous gas
824 fermentation cultures, and shows high carbon flux to reduced by-products. These strain characteristics
825 will speed up both phenotyping and metabolic engineering of acetogens. Second, our comprehensive
826 ALE approach with three strategies offers insights into convergent evolution trajectories in acetogens
827 and highlights novel engineering targets for improved phenotypes. Further work is needed to validate
828 the causality of other mutations uncovered in our ALEs by reconstructing mutations of interest, both

829 in combination and in isolation, and assaying their contribution to observed phenotypes. Future ALE
830 experiments with LAbrini should potentially be designed to select for improved H_2 uptake, CO
831 tolerance, and by-product tolerance.
832

833 ## DECLARATION OF INTEREST

834 LanzaTech has interest in commercial gas fermentation with *C. autoethanogenum*. AH, SB, GH, JD, LT,
835 HZ, ROJ, VR, HS, SDS, and MK are employees of LanzaTech.
836

837 ## ACKNOWLEDGEMENTS

838 This work was funded by the European Union's Horizon 2020 research and innovation programme
839 under grant agreement N810755 and the Estonian Research Council's grant agreement PSG289.
840 Australian Government funding through its investment agency, the Australian Research Council,
841 towards the ARC Centre of Excellence in Synthetic Biology (CE200100029) is gratefully acknowledged.
842 We acknowledge support from the Australian Centre for Ecogenomics for Whole Genome Sequencing
843 of UQ strains. We thank Kristina Reinmets and Ugochi Jennifer Nwaokorie for their valuable comments
844 on the manuscript, Shilpa Nagaraju for help with Supplementary Table 1, and Nastassia Shtaida for
845 help with microscopy.
846

847 ## DATA AVAILABILITY STATEMENT

848 RNA sequencing data have been deposited in the NCBI Gene Expression Omnibus repository under
849 accession number GSE225123. Proteomics data have been deposited to the ProteomeXchange
850 Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (44)
851 with the data set identifier PXD047330.The genome assembly of strain LT1 ("LAbrini") has been
852 deposited to the NCBI genome submission portal with reference Genbank CP110420.
853

854 ## AUTHOR CONTRIBUTIONS

855 Conceptualization: HI, JKH, AH, EM, SB and KV; Methodology: HI, JKH, AH, SB, KMS, AYS, MJP, LAL,
856 KRM, RAGG, GH, JD, LT, HZ, ROJ, VR, HS, JJ, MK, EM, and KV; Formal analysis: HI, JKH, AYS, MJP, LAL,
857 KRM, SB, VR, HS, JJ, and KV; Investigation: HI, JKH, AH, SB, KMS, AYS, MJP, LAL, KRM, RAGG, GH, JD,
858 LT, HZ, ROJ, and KV; Resources: SDS, MK, EM, and KV; Writing – Original Draft: HI and KV; Writing –
859 Review & Editing: HI, JKH, AH, SB, KMS, AYS, MJP, LAL, KRM, RAGG, LT, HZ, ROJ, VR, JJ, SDS, MK, EM,
860 and KV; Supervision: MK, EM, and KV; Project Administration: MK, EM, and KV; Funding Acquisition:
861 SDS, MK, EM, and KV.
862

863 REFERENCES

864 1. IEA, 2022. World Energy Outlook 2022. https://www.iea.org/reports/world-energy-outlook-
865 2022

866 2. IPCC, 2021. Climate Change 2021: the Physical Science Basis. Contribution of Working Group
867 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

868 3. Kircher M. 2021. Bioeconomy – present status and future needs of industrial value chains. N
869 Biotechnol 60:96–104.

870 4. Zhang J, Liang Guo, Cong Gao, Song W, Wu J, Liu L, Chen Xiulai. 2022. Metabolic engineering
871 strategies for microbial utilization of C1 feedstocks. Syst Microbiol Biomanufacturing 2022 3: 122–136.

872 5. Fackler N, Heijstra BD, Rasor BJ, Brown H, Martin J, Ni Z, Shebek KM, Rosin RR, Simpson SD,
873 Tyo KE, Giannone RJ, Hettich RL, Tschaplinski TJ, Leang C, Brown SD, Jewett MC, Köpke M. 2021.
874 Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical
875 Production from Gas Fermentation. Annu Rev Chem Biomol Eng 12:439–470.

876 6. Pavan M, Reinmets K, Garg S, Mueller AP, Marcellin E, Köpke M, Valgepea K. 2022. Advances
877 in systems metabolic engineering of autotrophic carbon oxide-fixing biocatalysts towards a circular
878 economy. Metab Eng 71:117–141.

879 7. Köpke M, Simpson SD. 2020. Pollution to products: recycling of 'above ground' carbon by gas
880 fermentation. Curr Opin Biotechnol 65:180–189.

881 8. Liew FE, Nogle R, Abdalla T, Rasor BJ, Canter C, Jensen RO, Wang L, Strutz J, Chirania P, De
882 Tissera S, Mueller AP, Ruan Z, Gao A, Tran L, Engle NL, Bromley JC, Daniell J, Conrado R, Tschaplinski
883 TJ, Giannone RJ, Hettich RL, Karim AS, Simpson SD, Brown SD, Leang C, Jewett MC, Köpke M. 2022.
884 Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale.
885 Nat Biotechnol 40:335–344.

886 9. Valgepea K, de Souza Pinto Lemgruber R, Abdalla T, Binos S, Takemori N, Takemori A, Tanaka
887 Y, Tappel R, Köpke M, Simpson SD, Nielsen LK, Marcellin E. 2018. H_2 drives metabolic rearrangements
888 in gas-fermenting *Clostridium autoethanogenum*. Biotechnol Biofuels 11:55.

889 10. Valgepea K, Loi KQ, Behrendorff JB, Lemgruber R de SP, Plan M, Hodson MP, Köpke M, Nielsen
890 LK, Marcellin E. 2017. Arginine deiminase pathway provides ATP and boosts growth of the gas-
891 fermenting acetogen *Clostridium autoethanogenum*. Metab Eng 41:202–211.

892 11. Bourgade B, Minton NP, Islam MA. 2021. Genetic and metabolic engineering challenges of C1-
893 gas fermenting acetogenic chassis organisms. FEMS Microbiol Rev fuab008:1–20.

894 12. Mahamkali V, Valgepea K, Lemgruber R de SP, Plan M, Tappel R, Köpke M, Simpson SD, Nielsen
895 LK, Marcellin E. 2020. Redox controls metabolic robustness in the gas-fermenting acetogen
896 *Clostridium autoethanogenum*. Proc Natl Acad Sci 117:13168–13175.

897 13. Klask C, Kliem-Kuster N, Molitor B, Angenent LT. 2020. Nitrate Feed Improves Growth and
898 Ethanol Production of *Clostridium ljungdahlii* With CO_2 and H_2, but Results in Stochastic Inhibition
899 Events. Front Microbiol 11:2019.12.15.877050.
900 14. Sandberg TE, Salazar MJ, Weng LL, Palsson BO, Feist AM. 2019. The emergence of adaptive
901 laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. Metab
902 Eng 56:1–16.
903 15. Dragosits M, Mattanovich D. 2013. Adaptive laboratory evolution - principles and applications
904 for biotechnology. Microb Cell Fact 12:64.
905 16. Gleizer S, Ben-Nissan R, Bar-On YM, Antonovsky N, Noor E, Zohar Y, Jona G, Krieger E,
906 Shamshoum M, Bar-Even A, Milo R. 2019. Conversion of *Escherichia coli* to Generate All Biomass
907 Carbon from CO_2. Cell 179:1255-1263.e12.
908 17. Gassler T, Sauer M, Gasser B, Egermeier M, Troyer C, Causon T, Hann S, Mattanovich D, Steiger
909 MG. 2019. The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph
910 capable of growth on CO_2. Nat Biotechnol 38:210–216.
911 18. Tremblay PL, Höglund D, Koza A, Bonde I, Zhang T. 2015. Adaptation of the autotrophic
912 acetogen *Sporomusa ovata* to methanol accelerates the conversion of CO_2 to organic products. Sci
913 Rep 5:1–11.
914 19. Oswald F, Zwick M, Omar O, Hotz EN, Neumann A. 2018. Growth and product formation of
915 *Clostridium ljungdahlii* in presence of cyanide. Front Microbiol 9:1–12.
916 20. Piatek P, Olsson L, Nygård Y. 2020. Adaptation during propagation improves *Clostridium*
917 *autoethanogenum* tolerance towards benzene, toluene and xylenes during gas fermentation.
918 Bioresour Technol Reports 12:100564.
919 21. Kwon SJ, Lee J, Lee HS. 2022. Metabolic changes of the acetogen *Clostridium* sp. AWRP
920 through adaptation to acetate challenge. Front Microbiol 13:982442.
921 22. Antonicelli G, Ricci L, Tarraran L, Fraterrigo Garofalo S, Re A, Vasile NS, Verga F, Pirri CF, Menin
922 B, Agostino V. 2023. Expanding the product portfolio of carbon dioxide and hydrogen-based gas
923 fermentation with an evolved strain of *Clostridium carboxidivorans*. Bioresour Technol 387:129689.
924 23. Kang S, Song Y, Jin S, Shin J, Bae J, Kim DR, Lee J-K, Kim SC, Cho S, Cho BK. 2020. Adaptive
925 Laboratory Evolution of *Eubacterium limosum* ATCC 8486 on Carbon Monoxide. Front Microbiol
926 11:402.
927 24. Jin S, Kang S, Bae J, Lee H, Cho B-K. 2022. Development of CO gas conversion system using
928 high CO tolerance biocatalyst. Chem Eng J 449:137678.
929 25. Martin ME, Richter H, Saha S, Angenent LT. 2015. Traits of selected *Clostridium* strains for
930 syngas fermentation to ethanol. Biotechnol Bioeng 113:531–539.

931 26. Schulz S, Molitor B, Angenent LT. 2023. Acetate augmentation boosts the ethanol production
932 rate and specificity by *Clostridium ljungdahlii* during gas fermentation with pure carbon monoxide.
933 Bioresour Technol 369:128387.
934 27. Abrini J, Naveau H, Nyns E-J. 1994. *Clostridium autoethanogenum*, sp. nov., an anaerobic
935 bacterium that produces ethanol from carbon monoxide. Arch Microbiol 161:345–351.
936 28. Nwaokorie UJ, Reinmets K, de Lima LA, Pawar PR, Shaikh KM, Harris A, Köpke M, Valgepea K.
937 2023. Deletion of genes linked to the C1-fixing gene cluster affects growth, by-products, and proteome
938 of *Clostridium autoethanogenum*. Front Bioeng Biotechnol 11:717.
939 29. Valgepea K, de Souza Pinto Lemgruber R, Meaghan K, Palfreyman RW, Abdalla T, Heijstra BD,
940 Behrendorff JB, Tappel R, Köpke M, Simpson SD, Nielsen LK, Marcellin E. 2017. Maintenance of ATP
941 Homeostasis Triggers Metabolic Shifts in Gas-Fermenting Acetogens. Cell Syst 4:505-515.e5.
942 30. Otte B, Grunwaldt E, Mahmoud O, Jennewein S. 2009. Genome shuffling in *Clostridium diolis*
943 DSM 15410 for improved 1,3-propanediol production. Appl Environ Microbiol 75:7610–7616.
944 31. Peebo K, Valgepea K, Nahku R, Riis G, Õun M, Adamberg K, Vilu R. 2014. Coordinated
945 activation of PTA-ACS and TCA cycles strongly reduces overflow metabolism of acetate in *Escherichia*
946 *coli*. Appl Microbiol Biotechnol 98:5131–5143.
947 32. Deatherage DE, Barrick JE. 2014. Identification of Mutations in Laboratory-Evolved Microbes
948 from Next-Generation Sequencing Data Using breseq, p. 165–188. In Methods in molecular biology
949 (Clifton, N.J.).
950 33. Humphreys CM, McLean S, Schatschneider S, Millat T, Henstra AM, Annan FJ, Breitkopf R,
951 Pander B, Piatek P, Rowe P, Wichlacz AT, Woods C, Norman R, Blom J, Goesman A, Hodgman C, Barrett
952 D, Thomas NR, Winzer K, Minton NP. 2015. Whole genome sequence and manual annotation of
953 *Clostridium autoethanogenum*, an industrially relevant bacterium. BMC Genomics 16:1085.
954 34. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman
955 J, Young SK, Earl AM. 2014. Pilon: An integrated tool for comprehensive microbial variant detection
956 and genome assembly improvement. PLoS One 9:e112963.
957 35. Brown SD, Nagaraju S, Utturkar S, De Tissera S, Segovia S, Mitchell W, Land ML, Dassanayake
958 A, Köpke M. 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the
959 genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant
960 Clostridia. Biotechnol Biofuels 7:40.
961 36. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: Summarize analysis results for
962 multiple tools and samples in a single report. Bioinformatics 32:3047–3048.
963 37. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
964 reads. EMBnet.journal 17:10.

965 38. Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better
966 for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47:e47–e47.
967 39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.
968 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
969 40. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful
970 Approach to Multiple Testing. J R Stat Soc Ser B 57:289–300.
971 41. Liu H, Naismith JH. 2008. An efficient one-step site-directed deletion, insertion, single and
972 multiple-site plasmid mutagenesis protocol. BMC Biotechnol 8:91.
973 42. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. 2019. DIA-NN: neural networks and
974 interference correction enable deep proteome coverage in high throughput. Nat Methods 17:41–44.
975 43. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus
976 computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13:731–740.
977 44. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S,
978 Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaíno JA. 2022.
979 The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences.
980 Nucleic Acids Res 50:D543–D552.
981 45. Wiser MJ, Ribeck N, Lenski RE. 2013. Long-Term Dynamics of Adaptation in Asexual
982 Populations. Science 342:1364–1367.
983 46. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, Szubin R, Palsson BO,
984 Feist AM. 2015. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth
985 of *Escherichia coli* K-12 MG1655 on glucose minimal medium. Appl Environ Microbiol 81:17–30.
986 47. Adamberg K, Valgepea K, Vilu R. 2015. Advanced continuous cultivation methods for systems
987 microbiology. Microbiol 161:1707–1719.
988 48. Fimlaid KA, Jensen O, Donnelly ML, Francis MB, Sorg JA, Shen A. 2015. Identification of a Novel
989 Lipoprotein Regulator of *Clostridium difficile* Spore Germination. PLOS Pathog 11:e1005239.
990 49. Alsaker KV, Spitzer TR, Papoutsakis ET. 2004. Transcriptional Analysis of *spo0A* Overexpression
991 in *Clostridium acetobutylicum* and Its Effect on the Cell's Response to Butanol Stress. J Bacteriol
992 186:1959–1971.
993 50. Xin X, Cheng C, Du G, Chen L, Xue C. 2020. Metabolic Engineering of Histidine Kinases in
994 *Clostridium beijerinckii* for Enhanced Butanol Production. Front Bioeng Biotechnol 8:214.
995 51. Valgepea K, Talbo G, Takemori N, Takemori A, Ludwig C, Mahamkali V, Mueller AP, Tappel R,
996 Köpke M, Simpson SD, Nielsen LK, Marcellin E. 2022. Absolute Proteome Quantification in the Gas-
997 Fermenting Acetogen *Clostridium autoethanogenum*. 7:e00026-22.

998 52. Phaneuf PV, Gosting D, Palsson BO, Feist AM. 2019. Aledb 1.0: A database of mutations from
999 adaptive laboratory evolution experimentation. Nucleic Acids Res 47:D1164–D1171.
1000 53. Steiner E, Dago AE, Young DI, Heap JT, Minton NP, Hoch JA, Young M. 2011. Multiple orphan
1001 histidine kinases interact directly with Spo0A to control the initiation of endospore formation in
1002 *Clostridium acetobutylicum*. Mol Microbiol 80:641–654.
1003 54. Paredes CJ, Alsaker KV, Papoutsakis ET. 2005. A comparative genomic view of clostridial
1004 sporulation and physiology. Nat Rev Microbiol 3:969–978.
1005 55. Dürre P. 2016. Physiology and sporulation in clostridium. Bact Spore From Mol to Syst 313–
1006 329.
1007 56. Kolek J, Diallo M, Vasylkivska M, Branska B, Sedlar K, López-Contreras AM, Patakova P. 2017.
1008 Comparison of expression of key sporulation, solventogenic and acetogenic genes in *C. beijerinckii*
1009 NRRL B-598 and its mutant strain overexpressing *spo0A*. Appl Microbiol Biotechnol 101:8279–8291.
1010 57. Hoch JA. 1993. Regulation of the phosphorelay and the initiation of sporulation in *Bacillus*
1011 *subtilis*. Annu Rev Microbiol 47:441–465.
1012 58. Gligorijević V, Renfrew PD, Kosciolek T, Leman JK, Berenberg D, Vatanen T, Chandler C, Taylor
1013 BC, Fisk IM, Vlamakis H, Xavier RJ, Knight R, Cho K, Bonneau R. 2021. Structure-based protein function
1014 prediction using graph convolutional networks. Nat Commun 12:1–14.
1015 59. Sanderson T, Bileschi ML, Belanger D, Colwell LJ. 2023. ProteInfer, deep neural networks for
1016 protein functional inference. Elife 12:e80942.
1017 60. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R,
1018 Bye-A-Jee H, Cukura A, Denny P, Dogan T, Ebenezer TG, Fan J, Garmiri P, da Costa Gonzales LJ, Hatton-
1019 Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Joshi V, Jyothi D, Kandasaamy S, Lock A, Luciani
1020 A, Lugaric M, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Mishra A, Moulang K, Nightingale
1021 A, Pundir S, Qi G, Raj S, Raposo P, Rice DL, Saidi R, Santos R, Speretta E, Stephenson J, Totoo P, Turner
1022 E, Tyagi N, Vasudev P, Warner K, Watkins X, Zaru R, Zellner H, Bridge AJ, Aimo L, Argoud-Puy G,
1023 Auchincloss AH, Axelsen KB, Bansal P, Baratin D, Batista Neto TM, Blatter MC, Bolleman JT, Boutet E,
1024 Breuza L, Gil BC, Casals-Casas C, Echioukh KC, Coudert E, Cuche B, de Castro E, Estreicher A, Famiglietti
1025 ML, Feuermann M, Gasteiger E, Gaudet P, Gehant S, Gerritsen V, Gos A, Gruaz N, Hulo C, Hyka-
1026 Nouspikel N, Jungo F, Kerhornou A, Le Mercier P, Lieberherr D, Masson P, Morgat A, Muthukrishnan
1027 V, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Poux S, Pozzato M, Pruess M, Redaschi N, Rivoire C, Sigrist
1028 CJA, Sonesson K, Sundaram S, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Huang H, Laiho K,
1029 McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Zhang J. 2023. UniProt: the Universal
1030 Protein Knowledgebase in 2023. Nucleic Acids Res 51:D523–D531.

1031 61. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P,
1032 Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA,
1033 Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman
1034 A. 2023. InterPro in 2022. Nucleic Acids Res 51:D418–D427.
1035 62. Diallo M, Kengen SWM, López-Contreras AM. 2021. Sporulation in solventogenic and
1036 acetogenic clostridia. Appl Microbiol Biotechnol 105:3533–3557.
1037 63. Sandoval NR, Venkataramanan KP, Groth TS, Papoutsakis ET. 2015. Whole-genome sequence
1038 of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased
1039 butanol production and superior growth. Biotechnol Biofuels 8:227.
1040 64. DiCandia MA, Edwards AN, Jones JB, Swaim GL, Mills BD, McBride SM. 2022. Identification of
1041 Functional Spo0A Residues Critical for Sporulation in *Clostridioides difficile*. J Mol Biol 434:167641.
1042 65. Humphreys JR, Debebe BJ, Diggle SP, Winzer K. 2023. *Clostridium beijerinckii* strain
1043 degeneration is driven by the loss of Spo0A activity. Front Microbiol 13:1075609.
1044 66. Piatek P, Humphreys C, Raut MP, Wright PC, Simpson S, Köpke M, Minton NP, Winzer K. 2022.
1045 Agr Quorum Sensing influences the Wood-Ljungdahl pathway in *Clostridium autoethanogenum*. Sci
1046 Rep 12:411.
1047 67. Fong SS, Joyce AR, Palsson B. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead
1048 to convergent growth phenotypes with different gene expression states. Genome Res 15:1365–1372.
1049 68. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome
1050 evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243–1247.
1051 69. Helling RB, Vargas CN, Adams J. 1987. Evolution of *Escherichia coli* during growth in a constant
1052 environment. Genetics 116:349–358.
1053 70. Rozen DE, Lenski RE. 2000. Long-Term Experimental Evolution in *Escherichia coli*. VIII.
1054 Dynamics of a Balanced Polymorphism. Am Nat 155:24–35.
1055 71. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular
1056 evolution over 60,000 generations. Nature 551:45–50.
1057 72. Treves DS, Manning S, Adams J. 1998. Repeated evolution of an acetate-crossfeeding
1058 polymorphism in long-term populations of *Escherichia coli*. Mol Biol Evol 15:789–797.
1059 73. Lee DH, Feist AM, Barrett CL, Palsson B. 2011. Cumulative number of cell divisions as a
1060 meaningful timescale for adaptive laboratory evolution of *Escherichia coli*. PLoS One 6:e26172.
1061 74. Bennett RK, Gregory GJ, Gonzalez JE, Har JRG, Antoniewicz MR, Papoutsakis ET. 2021.
1062 Improving the Methanol Tolerance of an *Escherichia coli* Methylotroph via Adaptive Laboratory
1063 Evolution Enhances Synthetic Methanol Utilization. Front Microbiol 12:638426.

1064 75. Wang S, Huang H, Kahnt J, Mueller AP, Köpke M, Thauer RK. 2013. NADP-Specific Electron-
1065 Bifurcating [FeFe]-Hydrogenase in a Functional Complex with Formate Dehydrogenase in *Clostridium*
1066 *autoethanogenum* Grown on CO. J Bacteriol 195:4373–4386.
1067 76. Mock J, Zheng Y, Mueller AP, Ly S, Tran L, Segovia S, Nagaraju S, Köpke M, Dürre P, Thauer RK.
1068 2015. Energy conservation associated with ethanol formation from H_2 and CO_2 in *Clostridium*
1069 *autoethanogenum* involving electron bifurcation. J Bacteriol 197:2965–2980.
1070 77. Lemaire ON, Jespersen M, Wagner T. 2020. CO_2-Fixation Strategies in Energy Extremophiles:
1071 What Can We Learn From Acetogens? Front Microbiol 11:1–8.
1072 78. Li Q, Chen J, Minton NP, Zhang Y, Wen Z, Liu J, Yang H, Zeng Z, Ren X, Yang J, Gu Y, Jiang W,
1073 Jiang Y, Yang S. 2016. CRISPR-based genome editing and expression control systems in *Clostridium*
1074 *acetobutylicum* and *Clostridium beijerinckii*. Biotechnol J 11:961–972.
1075 79. Harris LM, Welker NE, Papoutsakis ET. 2002. Northern, morphological, and fermentation
1076 analysis of *spo0A* inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. J Bacteriol
1077 184:3586–3597.
1078 80. Talukdar PK, Olguín-Araneda V, Alnoman M, Paredes-Sabja D, Sarker MR. 2015. Updates on
1079 the sporulation process in *Clostridium* species. Res Microbiol 166:225–235.
1080 81. Loomis WF, Shaulsky G, Wang N. 1997. Histidine kinases in signal transduction pathways of
1081 eukaryotes. J Cell Sci 110:1141–1145.
1082 82. Xu M, Zhao J, Yu L, Tang IC, Xue C, Yang ST. 2015. Engineering *Clostridium acetobutylicum* with
1083 a histidine kinase knockout for enhanced n-butanol tolerance and production. Appl Microbiol
1084 Biotechnol 99:1011–1022.
1085 83. Fujita M, Losick R. 2003. The master regulator for entry into sporulation in *Bacillus subtilis*
1086 becomes a cell-specific transcription factor after asymmetric division. Genes Dev 17:1166–1174.
1087 84. Lim HG, Eng T, Banerjee D, Alarcon G, Lau AK, Park MR, Simmons BA, Palsson BO, Singer SW,
1088 Mukhopadhyay A, Feist AM. 2021. Generation of *Pseudomonas putida* KT2440 Strains with Efficient
1089 Utilization of Xylose and Galactose via Adaptive Laboratory Evolution. ACS Sustain Chem Eng 9:11512–
1090 11523.
1091 85. Lemaire ON, Wagner T. 2021. Gas channel rerouting in a primordial enzyme: Structural
1092 insights of the carbon-monoxide dehydrogenase/acetyl-CoA synthase complex from the acetogen
1093 *Clostridium autoethanogenum*. Biochim Biophys Acta - Bioenerg 1862:148330.
1094 86. Liew F, Henstra AM, Winzer K, Köpke M, Simpson SD, Minton NP. 2016. Insights into CO_2
1095 Fixation Pathway of *Clostridium autoethanogenum* by Targeted Mutagenesis. MBio 7:e00427-16.