bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1
A Novel Efficient L-Lysine Exporter Identified by Functional
2
Metagenomics

3 Sailesh Malla, Eric van der Helm, Behrooz Darbani, Stefan Wieschalka, Jochen Forster,
4 Irina Borodina, Morten O. A. Sommer*
5 *The Novo Nordisk Foundation Center for Biosustainability, Technical University of*
6 *Denmark, Kemitorvet 220, 2800 Kgs. Lyngby, Denmark*
7
8 *Corresponding author:
9 Prof. Morten O. A. Sommer
10 Email: msom@bio.dtu.dk
11

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12 **Abstract**

13 Lack of active export system often limits the industrial bio-based production processes
14 accumulating the intracellular product and hence complexing the purification steps. L-
15 lysine, an essential amino acid, is produced biologically in quantities exceeding two
16 million tons per year; yet, L-lysine production is challenged by efficient export system at
17 high titres during fermentation. To address this issue, new exporter candidates for efficient
18 efflux of L-lysine are needed. Using metagenomic functional selection, we identified 58
19 genes encoded on 28 unique metagenomic fragments from cow gut microbiome library
20 that improved L-lysine tolerance. These genes include a novel putative L-lysine
21 transporter, belonging to a previously uncharacterized EamA superfamily.
22 Characterization using *Xenopus oocyte* expression system as well as an *Escherichia coli*
23 host demonstrates activity as a L-lysine transporter. This novel exporter improved L-
24 lysine tolerance in *E. coli* by 40% and enhanced the specific productivity of L-lysine in
25 an industrial *Corynebacterium glutamicum* strain by 12%. Our approach allows the
26 sequence-independent discovery of novel exporters and can be deployed to increase titres
27 and productivity of toxicity-limited bioprocesses.

28

29 **Keywords**: Amino acid, transporter, *E.coli, C. glutamicum, Xenopus oocytes*

30

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31 **Introduction**

32 The global chemical industry is transitioning from reliance majorly on petrochemical
33 processes to more sustainable bio-based production. This development holds promise to
34 improve the sustainability of the chemical industry while also reducing the overall
35 production costs of chemical products. In order to establish a cost competitive bioprocess,
36 titers of fermentations frequently exceed 100 g/L, which leads to a significant stress on
37 the host organism. Indeed, a majority of industrial bioprocesses are limited in production
38 due to several stresses resulting from high product titers.

39

40 One of the most significant bio-based chemical products is L-lysine. The global bio-based
41 L-lysine production now exceeds 2.5 million tons per year, which is estimated to reach
42 3.0 million tons in 2022 corresponding to 5.6 billion USD of market value according to
43 the current L-lysine market report (Elder, 2019). Industrial L-lysine bioprocesses entirely
44 rely on *Corynebacterium glutamicum* and *Escherichia coli* production strains that achieve
45 titers over 1.2 M (Becker and Wittmann, 2012; Lee and Kim, 2015). It was observed that
46 the L-lysine export rate is inhibited by 50% upon exceeding the extracellular
47 concentration of 400 mM compared to that at 80 mM (Kelle et al., 1996), indicating the
48 substantial inhibition of the L-lysine-specific export in industrial fermentation. Indeed,
49 several studies have demonstrated the benefit of incorporating active efflux systems to
50 address intracellular product accumulation in biobased production (Borodina 2019;
51 Hemberger et al., 2011; Malla et al., 2010). Hence, L-lysine export system is an obvious
52 target to maintain the producer organism at high lysine concentration as well as easing the
53 downstream process. Despite its importance, there are only two identified lysine specific
54 exporters: i) LysE as a member of the lysine efflux permease (LysE; **2.A.75**) family
55 (Bellmann et al., 2001; Vrljic et al., 1999); and, ii) lysine outward permease (LysO or

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

56 YbjE in *E. coli*) (Pathania and Sardesai, 2015). Vrljic et al. (1996) achieved five folds
57 higher lysine export rate upon overexpression of LysE in *C. glutamicum* (Vrljic et al.,
58 1996). Similarly, Yasueda and Gunji have deployed this strategy for ten-fold improvement
59 in L-lysine production by expressing a spontaneously mutated LysE from *C. glutamicum*
60 in *Methylophilis methylotrophus* (Gunji and Yasueda, 2006). In addition to the rational
61 engineering of the existing exporters, there is an urgent demand for new genetic building
62 blocks to further improve L-lysine tolerance and production.
63
64 Functional metagenomic selection is an effective approach to discover novel genes and
65 enzymes due to its ability to access the wide range of genetic elements present in a
66 particular environmental niche (Forsberg et al., 2016; Munck et al., 2015; Sommer et al.,
67 2010, 2009). Hence, we set out to use functional metagenomic selection to identify novel
68 L-lysine transporter candidates from a cow fecal library, with the goal of improving
69 industrial L-lysine production (Fig. 1A). Using C13-labeled L-lysine and mRNA
70 expression of the screened transporter in *Xenopus oocyte*, the transporter candidate was
71 confirmed as L-lysine exporter. Expression of the metagenomic derived L-lysine
72 transporter improved titer and productivity in both Gram-positive and Gram-negative
73 production strains.
74
75 **Materials and methods**
76 **Bacterial strains, growth conditions and chemicals**
77 All bacterial strains, vectors and plasmids used in this study are listed in Table 1. All
78 oligonucleotide primers (synthesized by Integrated DNA Technologies, Inc.) used are
79 presented in Table 2. *E. coli* strains were routinely cultured at 37°C in Luria-Bertani (LB)
80 broth or on agar supplemented with kanamycin (35 μg mL^{-1}) when necessary (hereafter

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

81 referred as LB-km). For the L-lysine production in *E.coli* LB as well as M9 minimal media
82 supplemented with 2 gm/L of yeast extract were used. *C. glutamicum* strains were cultured
83 in modified CGXII medium (Keilhauer et al., 1993) supplemented with 5% sucrose, 1%
84 BHI, 0.5 mM IPTG and kanamycin (when necessary), at 30°C and 250 RPM. Growth
85 media were supplemented with 1.5% agar for plate assays.
86
87 DNA manipulations were carried out following standard protocols (Sambrook and
88 Russell, 2001). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA).
89 Restriction enzymes and T4 DNA ligase were purchased from Fermentas (Denmark) and
90 New England Biolabs (Hertfordshire, UK). DNA sequencing was performed using an
91 automated DNA sequence analyzer.
92
93 **Metagenomic library construction and functional screening for L-lysine**
94 A metagenomic expression library of total DNA extracted from a cow fecal sample was
95 constructed as described previously (Sommer et al., 2009). Briefly, library construction
96 involves i) the isolation of total DNA from 5 g of fecal material using the PowerMax Soil
97 DNA Isolation Kit (Mobio Laboratories Inc.), ii) fragmentation of extracted DNA into
98 pieces of an average size of 2 kb by sonication using a Covaris E210 (Massachusetts,
99 USA) followed by end-repair using the End-It end repair kit (Epicentre), iii) blunt-end
100 ligation into the pZE21 MCS1 expression vector with constitutive promoter pLteo-1 (Lutz
101 and Bujard, 1997) at the *HincII* site using the Fast Link ligation kit (Epicentre), and iv)
102 transformation of the ligated sample into electro-competent *E. coli* top10 cells by the
103 standard method (One Shot® TOP10 Electropcomp™ Cells, Invitrogen).
104

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

105 After electroporation, cells were recovered in 1 ml of SOC medium for 1 h at 37°C, and
106 the library was titered by plating out 1 µl, 0.1 µl and 0.01 µl of recovered cells onto LB-
107 km plates. The insert size distribution was estimated by gel electrophoresis of the PCR
108 products obtained by amplifying inserts using primers annealing to the vector backbone
109 flanking the *Hinc*II site. The average insert size for the library was 1.7 kb. The total size
110 of the metagenomics library was determined by multiplying the average insert size by the
111 number of colony forming units (CFU) per ml ($5x10^8$ bp). The remainder of the recovered
112 cells were inoculated into 10 ml of LB-km liquid media and grown overnight; the library
113 was frozen down in 15% glycerol and stored at -80°C.
114

115 The *E.coli* C4860 strain was used for the functional screening of L-lysine. First, a plasmid
116 prep of the metagenomics library was carried out from the *E. coli* top10 cells harboring
117 the library. Then, 400 ng of metagenomics plasmid DNA was transformed into electro-
118 competent *E. coli* C4860 cells, and the library titer was determined as described above.
119 On the basis of the determined library sizes and the titer of the library, 10^6 cells (i.e., 100
120 µl of the library cells) were plated on LB-km agar supplemented with L-lysine at the
121 selective concentration (14 g/L). Plates were incubated at 37°C, and the growth of colonies
122 (likely lysine-tolerant clones) was assayed after 48-65 h of incubation.
123

124 The metagenomic inserts present in L-lysine-tolerant clones were Sanger sequenced using
125 the pZE21_F and pZE21_R primer pair, which annealed to the vector backbone. The
126 resulting raw sequencing chromatogram files were analyzed and functionally annotated
127 using the deFUME web server (van der Helm et al., 2015):
128 http://www.cbs.dtu.dk//services/deFUME/.
129

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

130 **Minimum inhibitory concentration (MIC) and IC90 determination**

131 For MIC determination, the *E. coli* strains were cultured in LB liquid media at 37°C
132 overnight, and then approximately 1×10^4 cells were inoculated from the overnight cultures
133 into LB liquid media and grown at 30°C and 300 RPM in 96-well microtiter plates
134 containing 150 µl of medium per well. MICs were determined using a logarithmic L-
135 lysine (or chemical) concentration gradient with two-fold serial dilutions. Endpoint
136 absorbance measurements (A_{600nm}) were taken with a plate reader (Synergy H1, BioTek)
137 after 24 or 48 h of incubation and were background-subtracted. Growth inhibition of the
138 *E. coli* strains was plotted against L-lysine (chemical) concentration with a polynomial
139 interpolation between neighboring data points using R software (http://www.r-
140 project.org). The percentage of inhibition was calculated using the formula: $1 - [A_{600nm}$
141 $_{lysine\ (or\ chemical)}/A_{600nm\ control}]$. The inhibitory concentration was defined as the lowest
142 concentration of the chemical that inhibited 90% of the growth of the strain tested (IC90).
143

144 **Growth experiments**

145 Single colonies of the tolerant clone(s) harboring transporter homologues were grown
146 overnight at 37°C with shaking in liquid LB-km medium. The OD values at 600 nm
147 [$(OD)_{600nm}$] were determined by 10-fold dilution. Then, the cultures were diluted to adjust
148 the $(OD)_{600nm}$ to 0.1, and 5 µl of each culture was inoculated in 150 µl fresh media
149 containing L-lysine (0 and 8 g/L) and kanamycin in a 96-well micro-titer plate. The plate
150 was incubated at 37°C for 24 h in an automated spectrophotometer (ELx808, BioTek) that
151 recorded the $(OD)_{630nm}$ at an interval of 60 min. The data were subsequently retrieved and
152 analyzed to determine growth rates. The growth rate data are the average of triplicate
153 experiments, with error bars representing the standard error of the mean (SEM).
154

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

155 **PCR and *E. coli* transformation**

156 PCR was performed in a total volume of 50 μl under the following DNA amplification
157 conditions: 95°C for 5 min, followed by 30 cycles of 95°C for 30 sec, 50–65°C for 30 sec,
158 and 72°C for 1 min, and finally 72°C for 5 min. Electrocompetent or chemically
159 competent *E. coli* cells were transformed with the ligation mixture using a standard
160 protocol and plated onto LB-km agar plates for selection.

161

162 ***In silico* analysis**

163 The amino acid sequence of MglE (**M**etagenomics **g**ene for **l**ysine **E**xport) was analyzed
164 using Interpro (http://www.ebi.ac.uk/interpro/v), its 2D membrane topology was
165 predicted using Phobius (Käll et al., 2007) and visualized using Protter (Omasits et al.,
166 2014), and the phylogenetic tree was plotted using iTOL. The UniProtKB database
167 (Consortium, 2014) was accessed on 2016-06-01 to query the closest homologs of MglE.
168 Pfam version 29.0 (December 2015) (Finn et al., 2014) was used to extract the LysE
169 family (Pfam id: PF01810) members. The Maximum Likelihood Phylogenetic tree was
170 constructed using CLC Main Workbench 7 with the Neighbor Joining construction
171 method.

172 **Transport assays in *Xenopus* oocytes**

173 The *Xenopus laevis* oocytes were obtained from Ecocyte Bioscience (Germany). Oocytes
174 were kept in Kulori buffer (pH 7.4) and at 18°C. A linear cassette (including T7 promoter,
175 MglE as the gene of interest, and 3´UTR) was amplified from the plasmid pUSER016-
176 T7-MglE using Phusion Hot Start polymerase (ThermoFisher Scientific) and the primers
177 linear_cassette F and linear_cassette R. The linear cassette was used as template for *in-*
178 *vitro* transcription. Capped cRNAs was synthesized using the mMESSAGE
179 mMACHINE® T7 Transcription Kit (AM1344; ThermoFisher). The quality and quantity

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

180 of the RNA were determined by Agilent 2100 Bioanalyzer. For microinjection of cRNA
181 and 13C-labeled L-lysine, we used the RoboInject (Multi Channel Systems, Reutlingen,
182 Germany) automatic injection system (Darbani et al., 2019). Injection needles with
183 opening of 25 µm were used (Multi Channel Systems, MCS GmbH). For expression in
184 oocytes, 25 ng of the *in-vitro* transcribed cRNA for the MglE transporter was
185 microinjected into the oocytes 3 days prior to transport assays. To perform the export
186 assay, 50 nl of the 13C-labeled L-lysine stock solution was injected into the oocytes to
187 obtain estimated internal concentrations of 6 mM, assuming an after-injection dilution
188 factor of 20 (Darbani et al., 2016). Following four washing steps, each batch of 20 oocytes
189 was incubated for 180 min in 90 µl Kulori buffer at pH 5 or pH 7.4. After incubation, 70
190 µl of the medium was collected from each batch with intact oocytes and added onto 70 µl
191 of 60% MeOH before LC-MS analysis. The oocytes were washed four times with Kulori
192 buffer pH 7.4 and intracellular metabolites were extracted in 30% MeOH to analyzed on
193 LC-MS.

194 ***E. coli* double deletion mutant construction**

195 The chromosomal lysine decarboxylases *ldcC* and *cadA* in *E. coli* W3110 were
196 successively knocked out by PCR targeting to create *E.coli* DMLC strain. The gene
197 disruption process was carried out as described previously using an ampicillin-resistant
198 pSIJ8 helper plasmid containing both the λ Red and FLP systems (Jensen et al., 2015).
199 The primer pair IdcC_F2/IdcC_R (situated 64 bp away from the start/stop codon *ldcC*
200 gene) was used to amplify the kanamycin cassette from the genomic DNA of the *ldcC*
201 inframe knocked out Kieo strain b0186 whereas the primer pair CadA_F2/CadA_R
202 (situated 56 bp away from the start/stop codon *cadA* gene) was used to amplify the
203 kanamycin cassette from the genomic DNA of the *cadA* inframe knock-out Kieo strain

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

204 b4131. These PCR products were used to delete *ldcC* and *cadA* in *E. coli* W3110 strain.

205 The detail process of construction of *E. coli* DMLC is given in supplementary materials.

206 **Plasmid construction**

207 *Sub-cloning the metagenomic insert implicated in tolerance phenotypes*

208 The recombinant plasmids pZE-MglE and pZE0-P-MglE-T were constructed as described

209 below. The construction of recombinant plasmids was verified by both restriction

210 mapping and direct nucleotide sequencing of the respective genes in the recombinant

211 plasmids.

212

213 Using the oligonucleotide pair MglE_F/MglE_R and the pZE-RCL-MglE plasmid as

214 template DNA, the exact *orf* of the putative carboxylate/amino acid transporter

215 homologue (referred to as MglE) was amplified. The amplicon was purified with a Qiagen

216 PCR purification kit and digested with the restriction enzymes *Kpn*I and *Hind*III (NEB,

217 UK), followed by ligation using T4 DNA ligase (Fermentas, Denmark) into the

218 corresponding restriction sites of the multiple cloning site (MCS) of the pZE21 vector to

219 construct the recombinant plasmids pZE-MglE.

220

221 To construct the recombinant plasmid pZE0-P-MglE-T, USER cloning was applied. The

222 pZE21-vector backbone without its promoter (pZE0) was amplified using the

223 oligonucleotide pair pZE21_User_F/pZE21_User_R. The putative carboxylate/amino

224 acid transporter homologue and the native promoter and terminator sequences (P-MglE-

225 T fragment) were amplified from the pZE-RCL-MglE plasmid using the primer pair P-

226 MglE-T_User_F/P-MglE-T_User_F. The amplified PCR products were ligated into the

227 recombinant plasmid pZE0-P-MglE-T using the recommended standard USER cloning

228 protocol (NEB, UK).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

229

230 *Codon optimized mglE plasmids construction for oocytes expression study*

231 For functional analysis of MglE in *Xenopus* oocytes, we constructed recombinant plasmid
232 pUSER016-pT7-MglE by USER cloning. The pUSER016 vector backbone was digested
233 with *Pac*I for 18 h and the ends were further nicked with Nt.BbvCI for two hours. The
234 codon optimized *mglE* gene (sequence provided in supplementary data) for *Xenopus*
235 *laevis* was synthetized along with N-terminal linker sequence- GGCTTAA**U** and C-
236 terminal linker sequences -**A**TTAAACC (in the complementary sequence, T is replaced
237 by U for pairing with **A**) including uracil with compatible USER sites easing direct cloning
238 into linearized and nicked pUSER016 vector by USER cloning (USER®, NEB), resulting
239 in plasmid pUSER016-pT7-MglE.

240

241 *Recombinant plasmids with E.coli lysine exporter and homologous genes of mglE.*

242 The lysine exporter, *ybjE,* (900 bp, Genbank accession no. CAQ31402) from *E.coli* BL21
243 (DE3) was amplified using oligonucleotides YbjE-F and YbjE-R. The PCR product of
244 *ybjE* was cloned into pZE21 vector excised with *Kpn*I and *BamH*I restriction enzymes to
245 construct pZE-YbjE expression plasmid.

246

247 Similarly, four *mglE* homologous genes; i) Gene1 (891 bp, 82% identity, Genbank
248 accession no. CDC57518) from *Bacteroides coprophilus* CAG:333, ii) Gene2 (891 bp,
249 78% identity, Genbank accession no. WP_018711839) from *Bacteroides barnesiae*, iii)
250 Gene3 (870 bp, 54% identity, Genbank accession no. CDC66277) from *Bacteroides sp.*
251 CAG:770, iv) Gene4 (912 bp, 43% identity, Genbank accession no. CCZ76555) from
252 *Alistipes finegoldii* CAG, were synthetized along with N-terminal linker sequence -
253 AATTCAT**U** AAAGAGGAGAAAGGTACC and C-terminal linker sequence -

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

254 GTCGACGGTATCG **A**TAAGCTT (in the complementary sequence, T is replaced by U
255 for pairing with **A)** including uracil easing direct user cloning. These four synthetized
256 gene fragments were cloned into the linear PCR fragment of pZE21 vector amplified by
257 using oligonucleotides pZE21_User-F4 and pZE21_User-R4 (Table 2) to construct pZE-
258 Gene1, pZE-Gene2, pZE-Gene3 and pZE-Gene4 recombinant plasmids, respectively.
259
260 **Construction of recombinant industrial *Corynebacterium glutamicum* strains**
261 The full metagenomics insert derived from pZE-RCL-MglE plasmid was amplified using
262 the primer pair RCL-MglE-F/RCL-MglE-R. The resulting PCR product was cloned into
263 pEKEx2 vector, using the *Kpn*I and *Sac*I restriction sites, yielding expression plasmid
264 pEK-RCL-MglE. The newly constructed expression plasmid was introduced into the
265 industrial L-lysine producing strain *C. glutamicum* VL5 by electroporation and
266 concomitant selection on kanamycin containing 2xTY-Agar. Successful construction of
267 the new strain, named *C. glutamicum* VL5/pEK-RCL-MglE, was verified by re-isolation
268 of the plasmid, followed by control digest and PCR on the length of the insert. The same
269 procedure was performed to introduce the empty pEKEx2 expression vector into *C.*
270 *glutamicum* VL5, yielding control strain *C. glutamicum* VL5/pEKEx2.
271
272 **Analysis of L-lysine in supernatant of bacterial culture**
273 A calibration curve of an authentic standard was generated using concentrations ranging
274 from 0.4 to 77 mg/L. The accurate mass of L-lysine from 20-fold diluted samples
275 (supernatants containing L-lysine for quantification) was analyzed using an LC-MS
276 Fusion (Thermo Fisher Scientific, USA) with positive electrospray ionization (ESI+). The
277 final concentration was adjusted for the dilution factor. Bracketing calibration was used
278 for the quantification of the external concentration. For the quantification of L-lysine, an

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

279 LC-MS/MS, EVOQ (Bruker, Fremont USA) was used with multiple reaction monitoring
280 (MRM) transition in positive ionisation mode (ESI+), with the quantified transition
281 147→84 (CE 10) and qualifier transition 147→130 (CE 7). The significance of the
282 specific lysine production was calculated using a Student t-test.
283

284 **Results and Discussion**

285 **Identification of metagenomic L-lysine tolerance genes**

286 The gut microbiota from industrial livestock animals are conceivably enriched with the
287 microbes capable to survive at higher L-lysine concentration which is supplied as food
288 additives. A metagenomic library derived from cow feces was constructed and used to
289 find candidate genes that could lead to a higher L-lysine tolerance in *E.coli* C4860 strain.
290 The resulting library was screened for L-lysine tolerance by plating on LB agar plates
291 supplemented with inhibitory concentrations of L-lysine (Fig. 1A) (materials and
292 methods). Colonies appeared only on the plates with the metagenomic library but not on
293 the control plates with *E.coli* C4860 harboring empty vector. Eighty L-lysine tolerant
294 clones were selected for further analysis.
295 Metagenomic inserts present in those L-lysine tolerant clones were PCR amplified,
296 sequenced and annotated using deFUME (van der Helm et al., 2015). In total, 28 unique
297 clones containing 58 complete or partial open reading frames were identified in this
298 analysis. The sequence analysis revealed that resistant clones contained *orf*s homologous
299 to genes encoding L-lysine modification/degradation enzymes, membrane proteins,
300 signaling proteins and hypothetical proteins. (Supplementary Table S1).
301

302 **Selection of transporter candidates**

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

303 For the purpose of improving industrial production strains, novel efflux systems constitute
304 more relevant genetic building blocks than those involved in degradation and
305 modification, which would counteract the objective of L-lysine production. Accordingly,
306 we focused our subsequent analysis on *orfs* encoding potential efflux systems. We
307 identified six unique metagenomic inserts encoding putative membrane proteins, some of
308 which are annotated in Genbank as hypothetical proteins (Supplementary Table S1). Four
309 inserts were selected for further testing based on the absence of putative degradation
310 enzymes flanking the putative transporters on the metagenomic insert.
311
312 We determined the L-lysine IC90 values for each of the four selected metagenomic inserts.
313 The L-lysine IC90 values of the selected metagenomic clones ranged from 10.44 ± 1.277
314 g/L to 14.25 ± 0.415 g/L, corresponding to a more than 40% increase in the L-lysine IC90
315 for the clone harboring pZE-RCL-MglE compared to the empty vector control strain (Fig.
316 1B). All of these strains have similar growth profiles in LB medium, whereas the growth
317 rates varied in the liquid LB supplemented with 8 g/L of L-lysine. At 8 g/L L-lysine
318 supplementation, the growth rate of the highest tolerant metagenomic clone harboring
319 pZE-RCL-MglE was 30% higher than that of the empty vector control (Fig. 1C).
320
321 **Lysine tolerance by MglE in *E. coli* strains**
322 To test whether the pZE-RCL-MglE plasmid carrying the *mglE* gene could confer L-
323 lysine tolerance to industrially relevant *E. coli* strains, we introduced the plasmid into
324 various *E. coli* strains; BL21 (DE3), MG1655, Crooks, W1116, and W3110 and IC90
325 values of L-lysine were determined for all of the constructed *E.coli* strains. Interestingly,
326 the IC90 values of L-lysine for all of these *E. coli* strains were increased upon introduction
327 of the pZE-RCL-MglE plasmid (Fig. 1D). Of particular interest, the L-lysine IC90 was

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

328 increased by 30% in W3110, which is a widely used background strain for L-lysine
329 production (Imaizumi et al., 2005). These data demonstrate that the mechanism of L-
330 lysine tolerance mediated by pZE-RCL-MglE is effective across the industrially relevant
331 *E. coli* strains and accordingly should be generally applicable to industrial *E. coli*-based
332 L-lysine fermentations.
333

334 **Protein sequence analysis of the metagenomic insert carrying MglE**

335 Sequence analysis of the 1.6 kb PCR amplicon from pZE-RCL-MglE metagenomics insert
336 contained an open reading frame *mglE* flanked by native promoter and terminator
337 sequences (Fig. 2A). The MglE protein consists of 297 amino acids, and its closest
338 homolog is a hypothetical protein from *Bacteroides coprophilus* (Genbank accession no.
339 WP_008140691) with 82% identity at the amino acid level. The MglE protein is a member
340 of RhaT/EamA-like transporter family of the drug/metabolite transporter (DMT; **2.A.7**)
341 superfamily. The MglE contains two copies of the EamA domain, which is found in
342 transporters belonging to the EamA family, at 9–143 aa (pfam (Finn et al., 2014) e-value:
343 $3e^{-11}$) and 152–292 aa (pfam e-value $9.6e^{-12}$). The members of the EamA family are
344 diverse, and most of their functions are unknown (Franke et al., 2003). Nevertheless, a
345 few proteins belonging to EamA family are well characterized as exporters such as PecM
346 exports a pigmented compound indigoidine in *Erwinia chrysanthemi* (Rouanet and
347 Nasser, 2001) and YdeD export metabolites of the cysteine pathway in *E. coli* (Daßler et
348 al., 2000). The predicted two-dimensional topology of the MglE transporter possesses six
349 cytoplasmic domains, five periplasmic domains, and ten transmembrane domains, with
350 both the N- and C-terminals in the cytoplasmic region (Fig. 2B).
351

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

352 Phylogenetic analysis shows that the closest MglE homologs (>50% sequence identity)
353 are present in the phylum Bacteroidetes and are annotated as uncharacterized proteins
354 (Fig. 2C). Members of the LysE family (pfam id: PF01810) are mainly (>5 species per
355 phyla) found in the phyla: Proteobacteria, Bacteroidetes, Firmicutes, Actinobacteria. The
356 full list of phyla that contain species with a LysE domain are Proteobacteria: 102,
357 Bacteroidetes: 73, Firmicutes: 45, Actinobacteria: 28, Chloroflexi: 1, Cyanobacteria: 3,
358 Spirochaetes: 1, Verrucomicrobia: 1 and Gemmatimonadetes: 1. Interestingly, nine
359 genomes containing the MglE homolog also contain additional LysE family (pfam id:
360 PF01810, TC code: **2.A.75**) genes (Fig. 2C orange dot). MglE represents a novel L-lysine
361 transporter that could not have been identified by sequence analysis or homology to
362 known L-lysine transporters.
363
364 **Effect of regions in the metagenomics insert on L-lysine tolerance**
365 To investigate the effects of promoter region of the metagenomic clone on lysine
366 tolerance, the recombinant plasmids pZE-MglE (*mglE* cloned into pZE21), and pZE0-P-
367 MglE-T (MglE along with its native promoter and terminator cloned into pZE21 without
368 the pLteo-1 vector promoter) were subcloned from pZE-RCL-MglE plasmid. The
369 constructed plasmids were transformed into *E. coli* C4860, and IC90 values of L-lysine
370 were determined (Fig. 2D). Although these recombinant C4860 strains grew in the same
371 extent in LB media, they showed different L-lysine tolerance. We observed that the IC90
372 for the strain harboring the pZE0-P-MglE-T plasmid was higher than that of the strain
373 harboring the pZE-MglE plasmid, indicating that the native promoter resulted in higher
374 lysine tolerance than the vector promoter. However, the synergistic effect of these two
375 promoters was better still (i.e., the IC90 values of the pZE-RCL-MglE plasmid was higher
376 than those of the strains harboring pZE-MglE or the pZE0-P-MglE-T plasmids). Hence,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

377 the lysine tolerance of the *E. coli* C4860 strain was the highest upon harboring the full
378 metagenomic insert (i.e., pZE-RCL-MglE).
379

380 **MglE exports L-lysine when expressed in *Xenopus* oocytes**

381 For functional analysis, MglE was expressed in *Xenopus* oocytes and the export of C13-
382 labeled L-lysine was measured. The water-injected oocytes were used as negative control.
383 L-lysine export assay was performed by injecting C13-labeled L-lysine to obtain a final
384 cytosolic concentration of 6 mM in the oocytes. The oocytes were incubated in Kulori
385 buffers with two different pH (pH 5 and 7.4) for 3 h, and L-lysine concentrations were
386 measured in the buffer and within the oocytes. MglE-expressing oocytes resulted in 6.6-
387 and 8.5-fold higher extracellular concentrations of C13-labeled L-lysine and natural C12
388 L-lysine respectively than control oocytes, when buffer with pH 5 was used (Fig. 3Ai).
389 The intracellular concentrations of C13-labeled and natural C12 L-lysine were
390 correspondingly 2.1- and 1.4-fold lower than in control oocytes (Fig. 3Ai). On the other
391 hand, there was no significant change in L-lysine quantities when oocytes were incubated
392 at pH 7.4 (Fig. 3Aii). This indicates a proton-gradient dependent export mechanism for
393 MglE transporter. L-lysine resistance (uptake) experiments in *E. coli* also indicated a
394 proton-dependent uptake for L-lysine because L-lysine toxicity was higher at pH 8.5 and
395 10 than at pH 4.5 and 7 (data not shown).
396

397 **MglE assists L-lysine export in *E.coli***

398 To analyze the function of MglE for L-lysine efflux, *E. coli* DMLC strain was constructed
399 by knocking out two lysine decarboxylase genes, constitutive (*ldcC*) and acid-inducible
400 (*cadA*). These deletions should prevent L-lysine degradation, thereby leading to increased
401 L-lysine production (Kikuchi et al., 1997). Then, the pZE-RCL-MglE plasmid and pZE21

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

402 empty vector were transformed into *E. coli* DMLC, yielding *E. coli* DMLC/pZE-RCL-
403 MglE and *E. coli* DMLC/pZE21, respectively. Subsequently, these strains along with *E.*
404 *coli* W3110/pZE21, were cultured in LB media as well as M9 minimal media
405 supplemented with yeast extract. The bacterial growth and extracellular L-lysine
406 concentrations were measured after 24 h. We found that in LB media the absolute
407 extracellular L-lysine titer was increased from 744 mg/L to 806 mg/L in presence of the
408 metagenomic insert consisting MglE operon i.e. 8.3% higher L-lysine production in the
409 *E. coli* DMLC/pZE-RCL-MglE strain as compared to *E. coli* DMLC/pZE21 in LB.
410 However, the biomass of the strain expressing MglE was decreased by about 50% as
411 compared to the control strain (Fig. 3Bi). Furthermore, the L-lysine production was
412 analysed in minimal media supplementing yeast extract. In minimal media, the
413 extracellular L-lysine accumulation was increased from 6 mg/L to 30 mg/L ($p<0.0001$, t-
414 test) whereas there was 12% reduced in biomass upon expressing the MglE metagenomic
415 fragment (Fig. 3Bii). The difference in extracellular L-lysine concentration among the
416 strain expressing MglE and the control strain is prominent in minimal media ($p<0.0001$*,
417 t-test) than in LB media ($p=0.1997$, t-test). Hence, the above results demonstrate that
418 MglE assists exporting L-lysine from *E. coli.*
419
420 The reduced biomass might be due to the combined effect of expression of membrane
421 protein in high copy number and active export of L-lysine by MglE which could have
422 resulted in less available energy for the cell to build biomass. It is worth noting that a big
423 proportion of total cellular energy-demand is for the transport machinery which has
424 accordingly been under evolutionary selection towards a higher energetic efficiency
425 (Darbani et al., 2018).
426

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

427 **Improvement of L-lysine productivity in industrial *C. glutamicum* strain**

428 Expression of a functionally active gene(s) from one strain to another often requires
429 special optimizations. For the practical application in industrial L-lysine bioprocesses,
430 retaining the activity of the discovered exporter into the production host, *C. glutamicum*,
431 is very essential. To test whether the MglE can boost the L-lysine bioprocesses, we cloned
432 and expressed it in an industrially used *C. glutamicum* L-lysine production strain.
433

434 The full metagenomic insert carrying the MglE along with its native promoter and
435 terminator was amplified from the pZE-RCL-MglE plasmid and cloned into pEKEx2
436 expression vector to construct pEK-RCL-MglE expression plasmid. The constructed
437 plasmid was introduced into *C. glutamicum* VL5 strain (Materials and methods), a
438 producer strain currently used for industrial L-lysine production. Using sucrose as a
439 carbon source, L-lysine production from the constructed *C. glutamicum* recombinant
440 strain was analyzed at 30 h. Unlike in *E.coli* strains, expression of the full metagenomics
441 insert harboring MglE did not have any growth inhibition effect as the vector control and
442 the recombinant strains both have nearly equal biomass. Notably, upon incorporation of
443 the MglE operon into *C. glutamicum*, the specific L-lysine productivity was improved by
444 12 ± 0.07% relative to the empty vector control (Fig. 3C). By expressing the operon in
445 this highly efficient producer strain, we showed not only the functionality of the MglE in
446 a Gram-positive species but also the benefit of the efflux system on an already highly
447 optimized producer strain.
448

449 **Comparison of L-lysine tolerance conferred by MglE and its homologous genes**

450 The YbjE transporter from *E. coli* is one of the functionally characterized lysine-specific
451 exporters. Based upon the homology search, four genes in the NCBI database having

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

452 different homologies (ranging from 43% to 82% in amino acid level) with MglE were also
453 selected and gene synthetized. To find out the lysine tolerance of MglE as compared to
454 that of YbjE and the four selected MglE homologous proteins (with 82, 78, 54 and 43%
455 amino acid sequence identity), recombinant expression plasmids pZE-YbjE, pZE-Gene1,
456 pZE-Gene2, pZE-Gene3 and pZE-Gene4 were constructed. The constructed plasmids,
457 pZE-MglE and pZE21 vector control were introduced into *E.coli* C4860 and *E.coli* DMLC
458 strains and determined the IC90 of L-lysine for the constructed recombinant strains as
459 described in early experiments.
460
461 Interestingly, we found that the expression of MglE provided similar L-lysine tolerance
462 in *E.coli* C4860 strain as that of YbjE whereas in *E.coli* DMLC strain, MglE displayed
463 even higher tolerance than YbjE. On the other hand, all of the four MglE homologs
464 showed lower lysine tolerance than MglE in both strains (Fig. 4).
465
466 **Conclusion**
467 We deployed functional metagenomic selections to identify novel genetic building blocks
468 encoding product efflux systems. Through applying this approach to the problem of L-
469 lysine toxicity at high concentration, we discovered MglE, which is confirmed as a novel
470 type of L-lysine exporter that belongs to the EamA superfamily. We demonstrate the
471 benefits of MglE in both Gram-negative (*E. coli*) and Gram-positive (*C. glutamicum*)
472 strains and show that MglE expression improves the productivity of an industrial L-lysine
473 production strain. If incorporated into industrial-scale bioprocesses, this discovery has the
474 potential to enhance L-lysine production by about 12%, representing an increased profit
475 on the order of 200-500 million USD per year. Our approach of using functional
476 metagenomics to identify novel transporters is independent of prior knowledge of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

477 transporter gene sequences and can be generally applied to most of the toxic
478 compounds/chemicals production by fermentation.
479

480 **Protein accession numbers**

481 The sequence information of full metagenomics insert harboring *mglE* is deposited in
482 the NCBI GenBank database with the accession number KU708839.

483

484 **Acknowledgements**

485 The research leading to these results was funded by the Novo Nordisk Foundation (Grant
486 Agreement no. NNF10CC1016517), and the European Union Seventh Framework
487 Programme (FP7-KBBE-2013-7-single-stage) under grant agreement no. 613745,
488 Promys. This work was further supported by the European Union Seventh Framework
489 Programme-ITN (FP7/2012/317058 to E.v.d.H). Stefan Wieschalka acknowledges
490 funding from Højteknologifonden. IB acknowledges the financial support the European
491 Research Council under the European Union's Horizon 2020 research and innovation
492 programme (YEAST-TRANS project, Grant Agreement no. 757384). We would like to
493 thank Vitalys I/S for providing the VL5 strain for lab-scale testing, Christian Munk for
494 helping with R-script to calculate IC90 values.

495 **Competing financial interests**

496 The authors declare no competing financial interests.

497 **Author contributions**

498 M.O.A.S. and S.M. conceived the study; S.M. designed and performed *E.coli*
499 experiments, B.D and I.B. conceived the experiments on *Xenopus oocytes* and B.D.
500 performed experiments. S.W. performed *C. glutamicum* experiments, E.v.d.H performed

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

501 *in silico* analysis. M.O.A.S., I.B., and J.F. led the research teams. M.O.A.S. and S.M.

502 drafted the manuscript and the other authors contributed to the writing.

503

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

504 **References**
505

506 Archer, C.T., Kim, J.F., Jeong, H., Park, J., Vickers, C.E., Lee, S., Nielsen, L.K., 2011.
507 The genome sequence of E. coli W (ATCC 9637): comparative genome analysis
508 and an improved genome-scale reconstruction of E. coli. BMC Genomics 12, 9.
509 https://doi.org/10.1186/1471-2164-12-9

510 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. a,
511 Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of Escherichia coli K-12
512 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems
513 biology 2, 1–11. https://doi.org/10.1038/msb4100050

514 Becker, J., Wittmann, C., 2012. Systems and synthetic metabolic engineering for amino
515 acid production - the heartbeat of industrial strain development. Current Opinion in
516 Biotechnology 23, 718–726. https://doi.org/10.1016/j.copbio.2011.12.025

517 Bellmann, A., Vrljić, M., Patek, M., Sahm, H., Kramer, R., Eggeling, L., 2001.
518 Expression control and specificity of the basic amino acid exporter LysE of
519 Corynebacterium glutamicum. Microbiology 147, 1765–1774.

520 Blattner, F.R., Plunkett, G.I., Bloch, a. C., Perna, T.N., Burland, V., Riley, M., Collado-
521 Vides, J., Glasner, D.J., Rode, K.C., Mayhew, F.G., Gregor, J., Davis, W.N.,
522 Kirkpatrick, a. H., Goeden, a. M., Rose, J.D., Mau, B., Shao, Y., 1997. The
523 Complete Genome Sequence of Escherichia coli K-12. Science 277, 1453–1462.
524 https://doi.org/10.1126/science.277.5331.1453

525 Borodina I. Understanding metabolite transport gives an upper hand in strain
526 development. Microb Biotechnol. 2019;12(1):69-70.

527 Cock, P.J. a, Antao, T., Chang, J.T., Chapman, B. a, Cox, C.J., Dalke, A., Friedberg, I.,
528 Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.L., 2009. Biopython: freely
529 available Python tools for computational molecular biology and bioinformatics.
530 Bioinformatics (Oxford, England) 25, 1422–3.
531 https://doi.org/10.1093/bioinformatics/btp163

532 Consortium, T.U., 2014. UniProt: a hub for protein information. Nucleic Acids Research
533 43, D204–D212. https://doi.org/10.1093/nar/gku989

534 Darbani B, Kell DB, Borodina I. Energetic evolution of cellular Transportomes. BMC
535 Genomics. 2018, 19(1):418.

536 Darbani, B., Motawia, M.S., Olsen, C.E., Nour-Eldin, H.H., Møller, B.L., Rook, F.,
537 2016. The biosynthetic gene cluster for the cyanogenic glucoside dhurrin in
538 Sorghum bicolor contains its co-expressed vacuolar MATE transporter. Scientific
539 Reports 6, 1–8. https://doi.org/10.1038/srep37079

540 Darbani B, Stovicek V, van der Hoek SA, Borodina I. Engineering energetically
541 efficient transport of dicarboxylic acids in yeast Saccharomyces cerevisiae.Proc
542 Natl Acad Sci U S A. 2019;116(39):19415-19420.

543 Daßler, T., Maier, T., Winterhalter, C., Böck, A., 2000. Identification of a major
544 facilitator protein from Escherichia coli involved in efflux of metabolites of the
545 cysteine pathway. Molecular Microbiology 36, 1101–1112.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

546 https://doi.org/10.1046/j.1365-2958.2000.01924.x

547 Elder, M. World markets for fermentation ingredients, BCC Research: Market Research
548 Reports. 2018. FOD020E. Available online: http://www.bccresearch.com

549 Eikmanns, B.J., Kleinertz, E., Liebl, W., Sahm, H., 1991. A family of Corynebacterium
550 glutamicum/Escherichia coli shuttle vectors for cloning, controlled gene
551 expression, and promoter probing. Gene 102, 93–98. https://doi.org/10.1016/0378-
552 1119(91)90545-M

553 Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger,
554 A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta,
555 M., 2014. Pfam: the protein families database. Nucleic Acids Research 42, D222–
556 D230. https://doi.org/10.1093/nar/gkt1223

557 Forsberg, K.J., Patel, S., Witt, E., Wang, B., Ellison, T.D., Dantas, G., 2016.
558 Identification of genes conferring tolerance to lignocellulose-derived inhibitors by
559 functional selections in soil metagenomes. Applied and Environmental
560 Microbiology 82, 528–537. https://doi.org/10.1128/AEM.02838-15

561 Franke, I., Resch, A., Daßler, T., Maier, T., Bock, A., 2003. YfiK from Escherichia coli
562 Promotes Export of O -Acetylserine and Cysteine. Journal of bacteriology 185,
563 1161–1166. https://doi.org/10.1128/JB.185.4.1161

564 Gunji, Y., Yasueda, H., 2006. Enhancement of l-lysine production in methylotroph
565 Methylophilus methylotrophus by introducing a mutant LysE exporter. Journal of
566 Biotechnology 127, 1–13. https://doi.org/10.1016/j.jbiotec.2006.06.003

567 Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E.,
568 Baba, T., Wanner, B.L., Mori, H., Horiuchi, T., 2006. Highly accurate genome
569 sequences of Escherichia coli K-12 strains MG1655 and W3110. Molecular
570 Systems Biology 2. https://doi.org/10.1038/msb4100049

571 Hemberger, S., Pedrolli, D.B., Stolz, J., Vogl, C., Lehmann, M., Mack, M., 2011. RibM
572 from Streptomyces davawensis is a riboflavin/roseoflavin transporter and may be
573 useful for the optimization of riboflavin production strains. BMC Biotechnology
574 11, 119. https://doi.org/10.1186/1472-6750-11-119

575 Imaizumi, A., Takikawa, R., Koseki, C., Usuda, Y., Yasueda, H., Kojima, H., Matsui,
576 K., Sugimoto, S.-I., 2005. Improved production of L-lysine by disruption of
577 stationary phase-specific rmf gene in Escherichia coli. Journal of biotechnology
578 117, 111–8. https://doi.org/10.1016/j.jbiotec.2004.12.014

579 Jensen, S.I., Lennen, R.M., Herrgård, M.J., Nielsen, A.T., 2015. Seven gene deletions in
580 seven days: Fast generation of Escherichia coli strains tolerant to acetate and
581 osmotic stress. Scientific Reports 5, 17874. https://doi.org/10.1038/srep17874

582 Käll, L., Krogh, A., Sonnhammer, E.L.L., 2007. Advantages of combined
583 transmembrane topology and signal peptide prediction-the Phobius web server.
584 Nucleic Acids Research 35, 429–432. https://doi.org/10.1093/nar/gkm256

585 Keilhauer, C., Eggeling, L., Sahm, H., 1993. Isoleucine synthesis in Corynebacterium
586 glutamicum: Molecular analysis of the ilvB-ilvN-ilvC operon. Journal of

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

587 Bacteriology 175, 5595–5603.

588 Kelle, R., Laufer, B., Brunzema, C., Weuster-Botz, D., Krämer, R., Wandrey, C., 1996.
589 Reaction engineering analysis of L-lysine transport by Corynebacterium
590 glutamicum. Biotechnology and Bioengineering 51, 40–50.
591 https://doi.org/10.1002/(SICI)1097-0290(19960705)51:1<40::AID-
592 BIT5>3.0.CO;2-0

593 Kikuchi, Y., Kojima, H., Tanaka, T., Takatsuka, Y., Kamio, Y., 1997. Characterization
594 of a second lysine decarboxylase isolated from Escherichia coli . Journal of
595 bacteriology 179, 4486–4492.

596 Lee, S.Y., Kim, H.U., 2015. Systems strategies for developing industrial microbial
597 strains. Nature Biotechnology 33, 1061–1072. https://doi.org/10.1038/nbt.3365

598 Lutz, R., Bujard, H., 1997. Independent and tight regulation of transcriptional units in
599 escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.
600 Nucleic Acids Research 25, 1203–1210. https://doi.org/10.1093/nar/25.6.1203

601 Malla, S., Niraula, N.P., Liou, K., Sohng, J.K., 2010. Self-resistance mechanism in
602 Streptomyces peucetius: Overexpression of drrA, drrB and drrC for doxorubicin
603 enhancement. Microbiological Research 165, 259–267.

604 Munck, C., Albertsen, M., Telke, A., Ellabaan, M., Nielsen, P.H., Sommer, M.O. a.,
605 2015. Limited dissemination of the wastewater treatment plant core resistome.
606 Nature Communications 6, 8452. https://doi.org/10.1038/ncomms9452

607 Omasits, U., Ahrens, C.H., Müller, S., Wollscheid, B., 2014. Protter: Interactive protein
608 feature visualization and integration with experimental proteomic data.
609 Bioinformatics 30, 884–886. https://doi.org/10.1093/bioinformatics/btt607

610 Pathania, A., Sardesai, A. a., 2015. Distinct paths for basic amino acid export in
611 Escherichia coli : YbjE (LysO) mediates export of L-lysine. Journal of Bacteriology
612 197, 2036–2047. https://doi.org/10.1128/JB.02505-14

613 Rouanet, C., Nasser, W., 2001. The PecM protein of the phytopathogenic bacterium
614 Erwinia chrysanthemi, membrane topology and possible involvement in the efflux
615 of the blue pigment. Journal of Molecular Microbiology and Biotechnology 3, 309–
616 318.

617 Sambrook, J., Russell, D.W., 2001. Molecular Cloning : A Laboratory Manual, Third.
618 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

619 Sommer, M.O. a, Church, G.M., Dantas, G., 2010. A functional metagenomic approach
620 for expanding the synthetic biology toolbox for biomass conversion. Molecular
621 systems biology 6. https://doi.org/10.1038/msb.2010.16

622 Sommer, M.O. a, Dantas, G., Church, G.M., 2009. Functional characterization of the
623 antibiotic resistance reservoir in the human microflora. Science (New York, N.Y.)
624 325, 1128–31. https://doi.org/10.1126/science.1176950

625 van der Helm, E., Geertz-Hansen, H.M., Genee, H.J., Malla, S., Sommer, M.O.A., 2015.
626 deFUME: Dynamic exploration of functional metagenomic sequencing data. BMC

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

627 Research Notes 8, 328–330. https://doi.org/10.1186/s13104-015-1281-y

628 Vrljic, M., Garg, J., Bellmann, A., Wachi, S., Freudi, R., Malecki, M., Sahm, H.,
629 Kozina, V., Eggeling, L., Saier, M., 1999. The LysE Superfamily : Topology of the
630 Lysine Exporter LysE of Corynebacterium glutamicum , a Paradyme for a Novel
631 Superfamily of Transmembrane Solute Translocators Real-Time PCR
632 Bioinformatics and Data. Journal of Molecular Microbiology and Biotechnology 1,
633 327–336.

634 Vrljic, M., Sahm, H., Eggeling, L., 1996. A new type of transporter with a new type of
635 cellular function : L -lysine export from Corynebacterium glutamicum. Molecular
636 Microbiology 22, 815–826.

637

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

638 **Table legends**

639 **Table 1.** Bacterial strains and plasmids used in this study.

640 **Table 2.** Oligonucleotides used in this study.

641

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

642 **Table 1.** Bacterial strains and plasmids used in this study.

Strains/Plasmids	Description	Source/reference
Strains		
Escherichia coli		
DH5α	General cloning host	Invitrogen
Top10	Highly-efficient host for cloning and plasmid propagation	Invitrogen
BL21(DE3)	*ompT hsdT hsdS* (r_B^- m_B^-) *gal* (DE3)	Novagen
C4860	Host for phage PhiX174, sup-, lambda-, F-. (now known as DSM 13127)	DSMZ
MG1655	F^- l^- ilvG$^-$ rfb-50 rph-1	(Blattner et al., 1997)
W3110	F^- lambda$^-$ IN(rrnD-rrnE)1 rph-1	(Hayashi et al., 2006)
W1116	strain W and ATCC 9637	(Archer et al., 2011)
Crooks	*E. coli* ATCC 8739	DSMZ
BL21/pZE21	*E.coli* BL21 carrying pZE21	This study
C4860/pZE21	*E.coli* C4860 carrying pZE21	This study
MG1655/pZE21	*E.coli* MG1655 carrying pZE21	This study
W3110/pZE21	*E.coli* W3110 carrying pZE21	This study
W1116/pZE21	*E.coli* W1116 carrying pZE21	This study
Crooks/pZE21	*E.coli* Crooks carrying pZE21	This study
BL21(DE3)/pZE-RCL-MglE	*E.coli* BL21 carrying pZE-RCL-MglE	This study
C4860/pZE-RCL-MglE	*E.coli* C4860 carrying pZE-RCL-MglE	This study
MG1655/pZE-RCL-MglE	*E.coli* MG1655 carrying pZE-RCL-MglE	This study
W3110/pZE-RCL-MglE	*E.coli* W3110 carrying pZE-RCL-MglE	This study
W1116/pZE-RCL-MglE	*E.coli* W1116 carrying pZE-RCL-MglE	This study
Crooks/pZE-RCL-MglE	*E.coli* Crooks carrying pZE-RCL-MglE	This study
C4860/pZE-MglE	*E.coli* C4860 carrying pZE-MglE	This study
C4860/pZE0-P-MglE-T	*E.coli* C4860 carrying pZE0-P-MglE-T	This study
C4860/pZE-YbjE	*E.coli* C4860 carrying pZE-YbjE	This study
C4860/pZE-Gene1	*E.coli* C4860 carrying pZE-Gene1	This study
C4860/pZE-Gene2	*E.coli* C4860 carrying pZE-Gene2	This study
C4860/pZE-Gene3	*E.coli* C4860 carrying pZE-Gene3	This study
C4860/pZE-Gene4	*E.coli* C4860 carrying pZE-Gene4	This study
W3110/pSIJ8	*E.coli* W3110 carrying pSIJ8	This study
W3110:: ΔldcC/pSIJ8	*ldcC* deletion mutant of *E.coli* W3110 carrying pSIJ8	This study
W3110::ΔldcC.ΔcadA/pSIJ8	*ldcC* and *cadA* deletion mutant of *E.coli* W3110 (DMLC) carrying pSIJ8	This study
DMLC	*ldcC* and *cadA* deletion mutant of *E.coli* W3110	This study
DMLC/pZE21	*E.coli* DMLC carrying pZE21	This study
DMLC/pZE-RCL-MglE	*E.coli* DMLC carrying pZE-RCL-MglE	This study
Kieo strain		
b0186	*ldcC* replaced by kanamycin cassette *E.coli* K-12 BW25113	(Baba et al., 2006)
b4131	*cadA* replaced by kanamycin cassette in *E.coli* K-12 BW25113	(Baba et al., 2006)
Corynebacterium. glutamicum		
VL5	Industrial L-lysine production strain of *C .glutamicum*	Vitalys
VL5/pEKEx2	*C. glutamicum* VL5 carrying pEKEx2	This study
VL5/pEK-RCL-MglE	*C. glutamicum* VL5 carrying pEK-RCL-MglE	This study
Plasmids and vectors		
pZE21	*E.coli* expression vector with pLtet promoter, pBR322 origin of replication. and kanamycin cassette	(Lutz and Bujard, 1997)
pSIJ8	pKD46 based vector with rhaRS-prha-FLP and ampillicin cassette	(Jensen et al., 2015)
pUSER016	pNB1u based vector with T7 promoter for in vitro transcription	(Darbani et al., 2016)
pEKEx2	IPTG inducible *C. glutamicum* expression vector with kanamycin cassette from pUC4 K; Ptrc, lacI, pUC18 mcs, and pBL1 origin of replications.	(Eikmanns et al., 1991)
pZE-RCL-MglE	pZE21 carrying L-lysine resistant metagenomic insert with MglE operon from cow fecal sample	This study
pZE-MglE	pZE21 carrying *mglE* gene	This study
pZE0-P-MglE-T	Promoter less pZE21 carrying native promoter and *mglE* gene and native terminator region	This study
pUSER016-pT7-MglE	Codon optimized MglE gene cloned into pUSER016 for functional studies in *Xenopus oocytes*	This study
pEK-RCL-MglE	pEKEx2 carrying L-lysine resistant metagenomic insert with MglE operon from cow fecal sample	This study
pZE-YbjE	pZE21 carrying *ybjE* gene from *E.coli* BL21(DE3)	This study
pZE-Gene1	pZE21 carrying gene from *Bacteroides coprophilus* CAG:333 having 82% identity with *mglE*	This study
pZE-Gene2	pZE21 carrying codon optimized gene from *Bacteroides barnesiae* having 78% identity with *mglE*	This study
pZE-Gene3	pZE21 carrying gene from *Bacteroides* sp. CAG:770 having 54% identity with *mglE*	This study
pZE-Gene4	pZE21 carrying codon optimized gene from *Alistipes finegoldii* CAG:68 having 43% identity with *mglE*	This study

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

643 **Table 2.** Oligonucleotides used in this study.

Primers	Oligonucleotide sequences (5'-3')	Restriction site
pZE21_F	ATCAGTGATAGAGATACTGAGCAC	
pZE21_R	TTTCGTTTTATTTGATGCCTCTAG	
MglE_F	AAA<u>*GGTACC*</u>ATGAGAAATTTAAGTAAAAAAGC	*KpnI*
MglE_R	TTA<u>*AAGCTT*</u>CTATCTCTTTGTTACTGAAATCAT	*HindIII*
RCL-MglE-F	CGG<u>*GGTACC*</u>AAGAGGAGAAAGGTACCGGGC	*KpnI*
RCL-MglE-R	C<u>*GAGCTC*</u>CCGGGCTGCAGGAATTCG	*SacI*
YbjE_F	AAA <u>*GGTACC*</u> ATGTTTTCTGGGCTGTTAATCA	*KpnI*
YbjE_R	TAT <u>*GGATCC*</u> TTACGCAGAGAAAAAGGCGAT	*BamHI*
linear_cassette F	TGCAAGGCGATTAAGTTGGGTAACGC	
linear_cassette R	CCATGATTACGCCAAGCTATTTAGGTGACAC	
For User Cloning		
pZE21_User_F	ATAAGACGG**U** ATCGATAAGCTTGATATCGAATTCC	
pZE21_User_R	AGGGTACCC**U** CGAGGTGAAGACGAAAGGGCCTCG	
P-MglE-T_User_F	AGGGTACCC**U** GAAGTCAAGCATCTCAAAAAACTAC	
P-MglE-T_User_R	ACCGTCTTA**U** TTACGATTTTACTACGGAGTATTAA	
pZE21_User_F4	ATAAGCT**U** GATATCGAATTCCTGCAGCCC	
pZE21_User_R4	AATGAAT**U** CGGTCAGTGCGTCCTGCTGAT	
For gene deletion and confirmation		
IdcC_F1	TCAGCGCCTGATGAGCTACG	
IdcC_F2	AGTTCTGAAAAAGGGTCACTTC	
IdcC_R	TCGCAATATGGTGAACCTGTT	
CadA_F1	TGAAGTACTCCCAGATTTGGATC	
CadA_F2	CGGCTGTGAGGGTGTTTTCA	
CadA_R	TTAATTAAAAGTATTTTCCGAGGCTCC	

644 Restriction sites are indicated by <u>underlined</u> and *italics*.

645

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

646 **Figure legends**

647 **Figure 1. Functional metagenomic screening and Lysine tolerance of the screened metagenomics**
648 **clones.**

649 A. Functional metagenomic screening for lysine transporters**.** Cow fecal metagenomic DNA library
650 construction and its functional selection for L-lysine exporters.
651 B. L-lysine IC90 values for the selected *E. coli* C4860 resistant clones harboring putative transporters.
652 C. The growth rates of *E. coli* C4860 harboring pZE-RCL-MglE and pZE21 (control) in LB-Km and LB-
653 Km supplemented with 8 g/L of L-lysine.
654 D. Improved L-lysine tolerance (IC90 values) by expression of metagenomics insert carrying MglE
655 transporter in industrially relevant *E. coli* strains.
656

657 **Figure 2. *In silico* analysis of the metagenomic insert contained in pZE-RCL-MglE plasmid and L-**
658 **lysine tolerance of its subclones**

659 A. The metagenomic insert present in the pZE-RCL-MglE plasmid is shown. P: promoter region, MglE:
660 membrane bound transporter protein, and T: terminator sequence.
661 B. The membrane topology of the MglE transporter protein showing 6 cytoplasmic regions, 5 periplasmic
662 regions, and 10 transmembrane α-helices. The 15 BLASTp hits with a >50% sequence identity were
663 aligned to MglE and the conservation (bitscore) for each residues was calculated using the Biopython
664 package (Cock et al., 2009).
665 C. A Maximum Likelihood phylogenetic protein tree containing all the significant (sequence identity
666 >50%) BLASTp hits in the UniProtKB database against MglE. The percentage identity of the 15 hits
667 is shown as horizontal bars. The presence of a LysE family member in the genomes is shown with an
668 orange dot. The highest homology with MglE is found in *Bacteroides coprophilus* CAG:333, with 82%
669 sequence identity.
670 D. The strategy for subcloning the MglE transporter with or without its native promoter and the resultant
671 L-lysine tolerance phenotypes for the *E. coli* C4860 strain are shown. The biomass of the strains (at 24
672 h) cultured in LB media was also shown in terms of OD in closed circles.
673

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.30.071142; this version posted May 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

674 **Figure. 3.** Functional characterization of MglE transporter activity in *Xenopus* oocytes, *E.coli* and *C.*
675 *glutamicum*

676 A. L-lysine export assay of MglE transporter in *Xenopus* oocytes in Kulori buffer i) at pH 5 and ii) at
677 pH 7.4. The bars represent lysine contents in the oocytes or the buffer (means, ± Std., 3-4 biological
678 replicates each involving 20 oocytes). Significant changes are in comparison with the controls
679 marked by asterisks (** $p < 0.01$; Fishers one-way ANOVA).
680 B. The extracellular L-lysine concentration of reference *E. coli* W3110 and its isogenic mutant strain
681 DMLC along with the cell OD_{600nm} values confirming the exporter activity of the MglE protein in
682 i) LB media and ii) in M9 minimal media supplemented with yeast extract (* $p<0.0001$, t-test).
683 C. Fold enhancement of specific productivity of L-lysine by MglE protein in the industrial *C.*
684 *glutamicum* VL5 strain compared to the empty vector control. Error bars are s.e.m, * denotes p =
685 0.023, n=4).

686

687 **Figure. 4**. Comparison of lysine tolerance (presented in relative IC90 values) displayed by MglE, YbjE
688 (Lysine specific exporter from *E.coli*) and four synthetic MglE homologs. The biomass of the strains (at 24
689 h) cultured in LB media was also shown in closed circles.

690

A

construction
of metagenomic library

metagenomic
DNA source

fragmented
metagenomic
DNA

plasmid
expression library

transformation
and cell library

function selection
for L-lysine exporter

lysine

native lysine
exporter

plasmid
containing
MglE

E. coli

exporter from
metagenomic
DNA (MglE)

functional selection
on agar plate supplemented
with L-lysine

B

IC_{90} L-lysine [gL^{-1}]

16
14
12
10
8
0

CGML1
CGML2
CGML3
RCL-MglE
pZE21

C

in LB media

LB + 8 [gL^{-1}] L-lysine

growth rate [h -1]

0.8
0.6
0.4
0.2
0.0

0.4
0.3
0.2
0.1
0.0

CGML1
CGML2
CGML3
RCL-MglE
pZE21

CGML1
CGML2
CGML3
RCL-MglE
pZE21

D

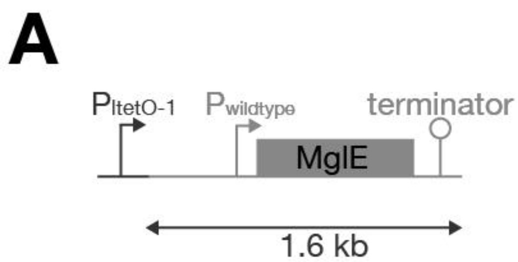
W1116
Crooks
W3110
MG1655
BL21
C4860

pZE-RCL-MglE
pZE21

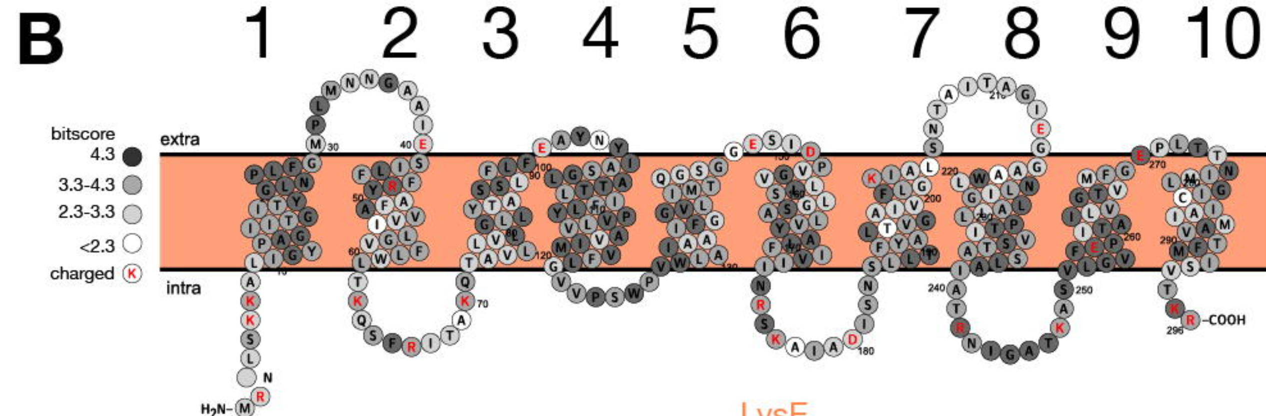
0
5
10
15
20

IC_{90} L-lysine [gL^{-1}]

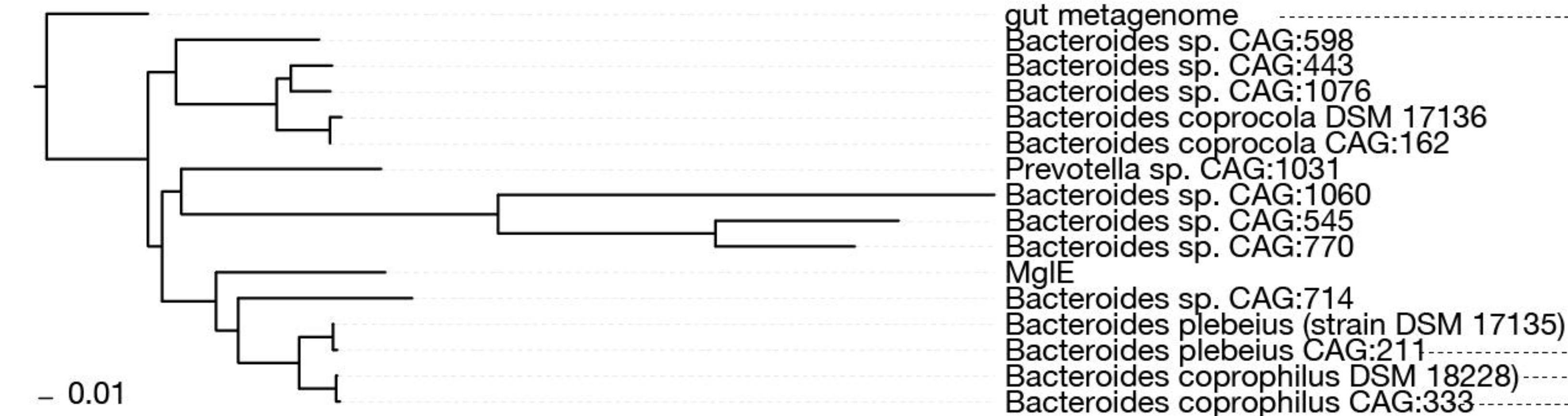
A

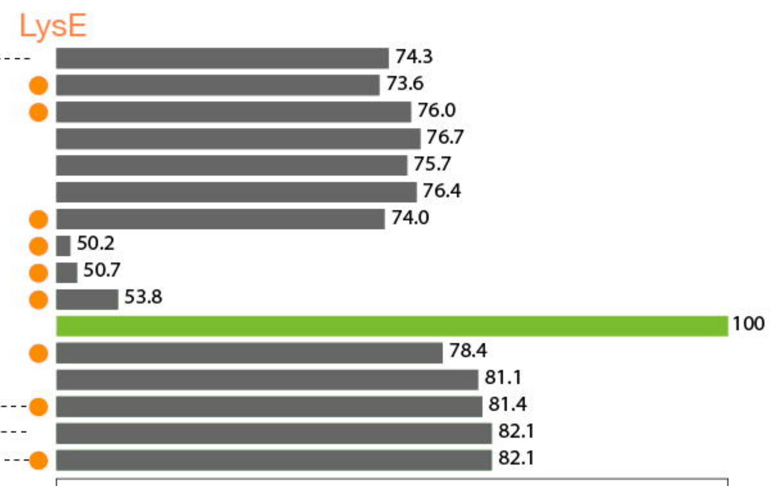


B

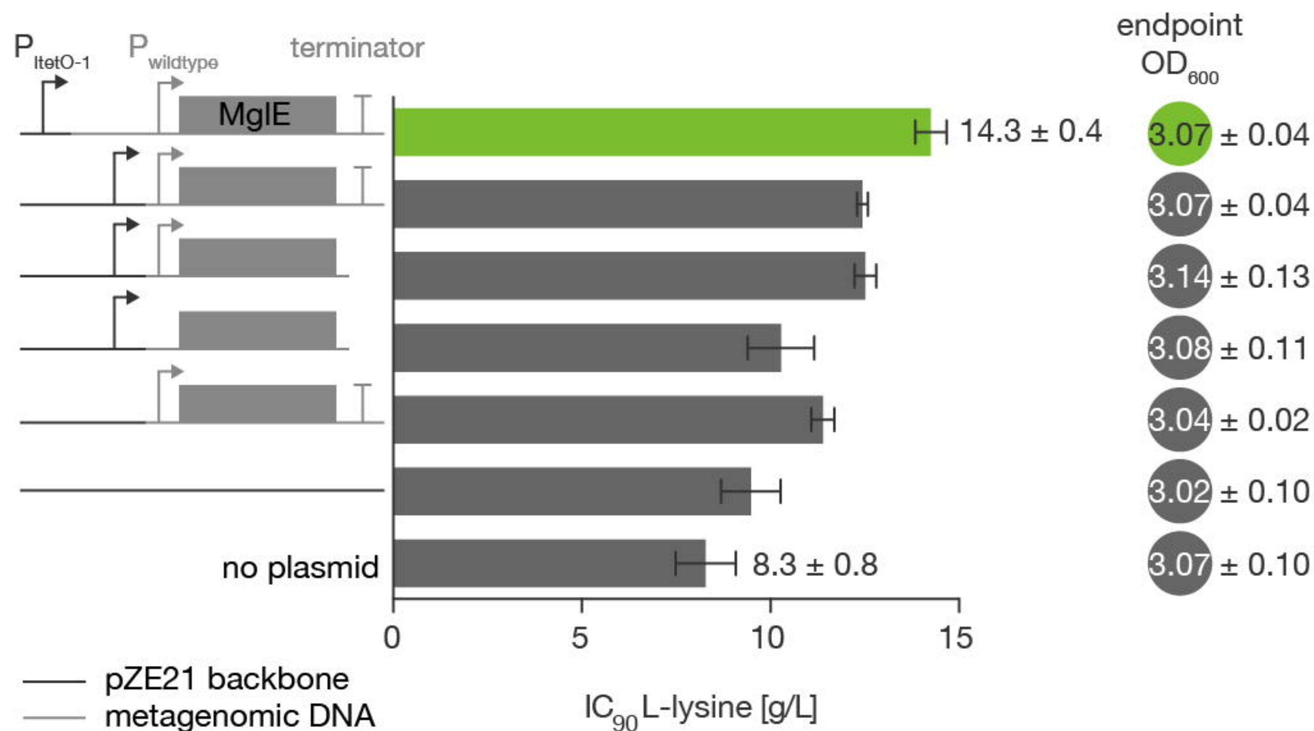


C

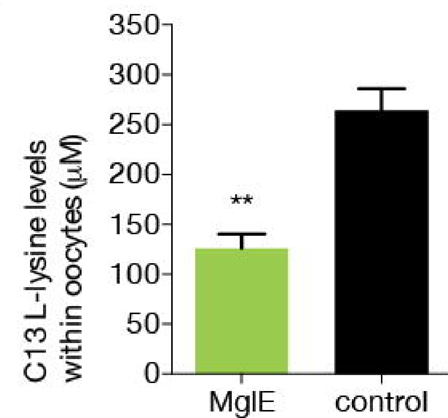


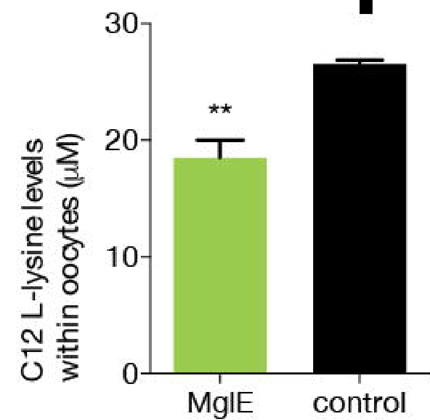


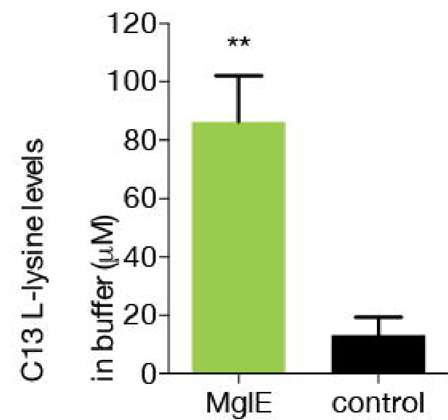
D

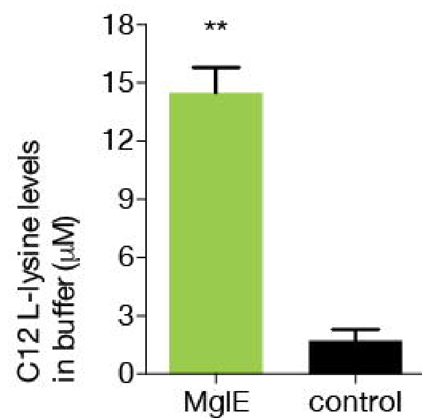


A i) pH 5.0

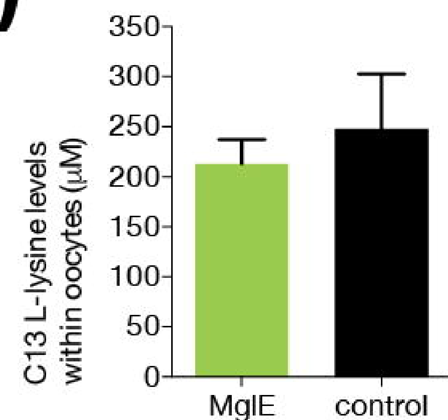


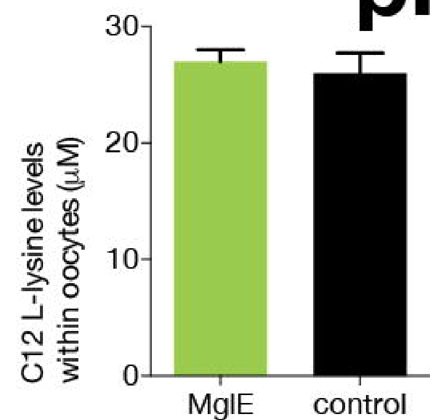


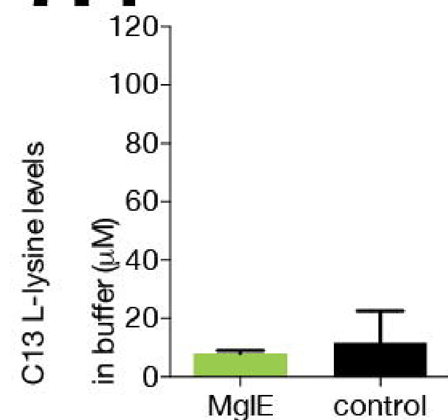


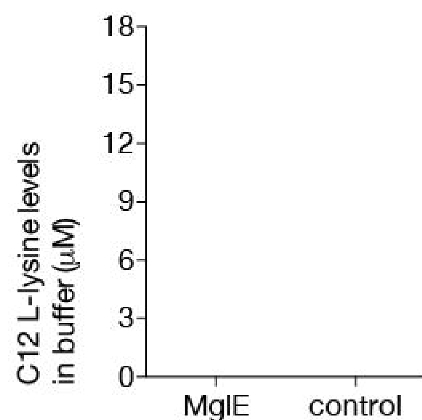


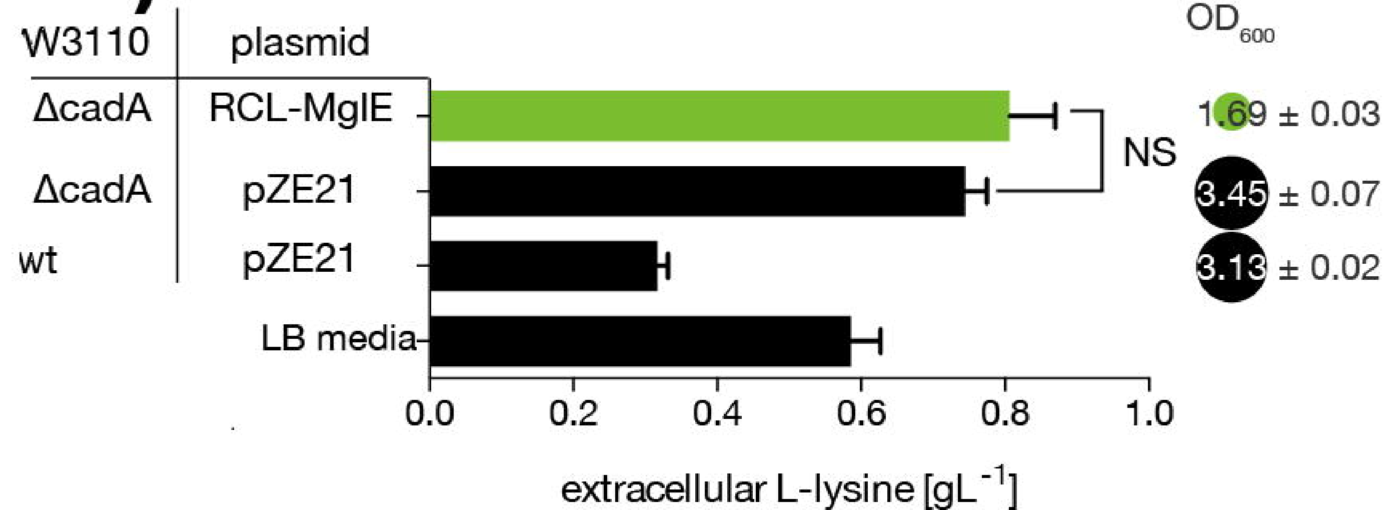
ii) pH 7.4

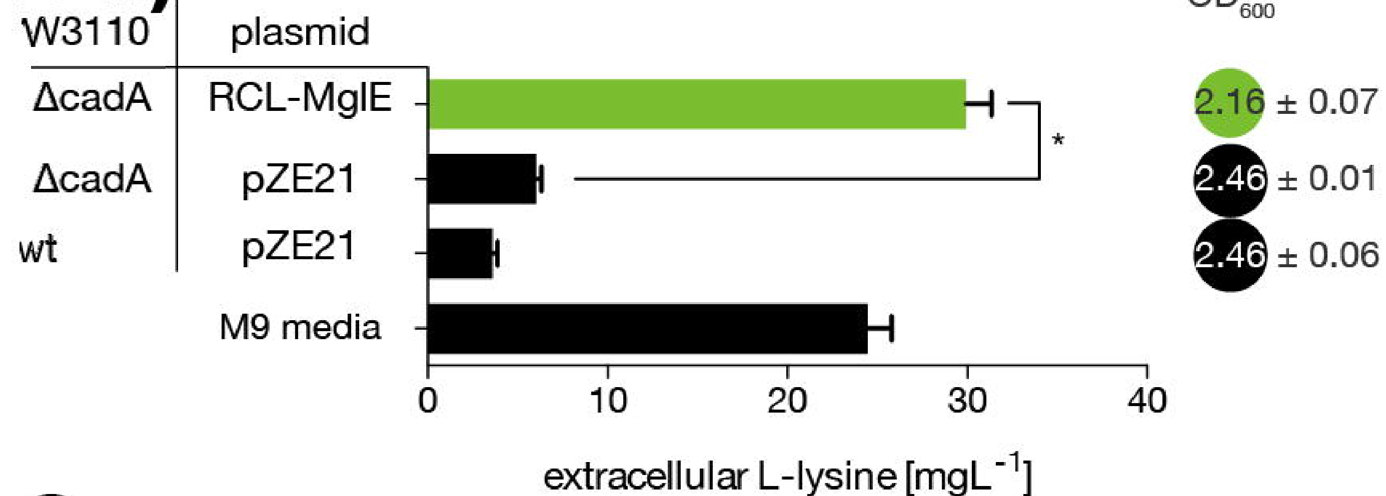


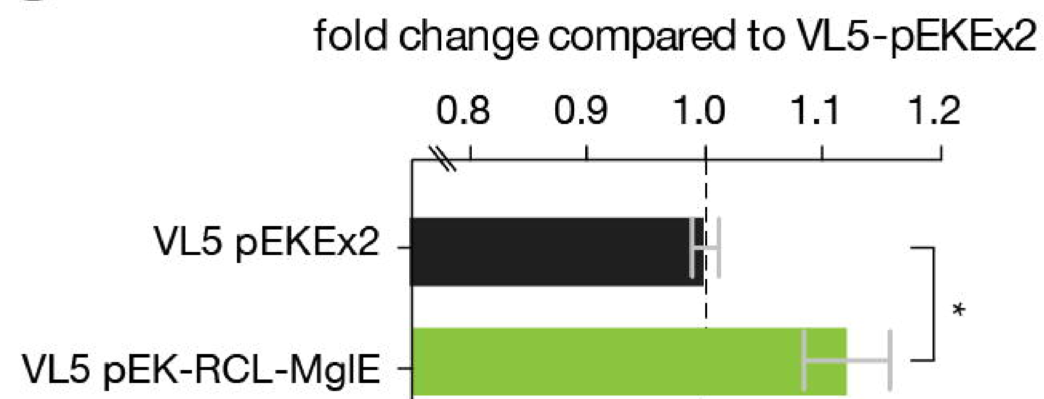


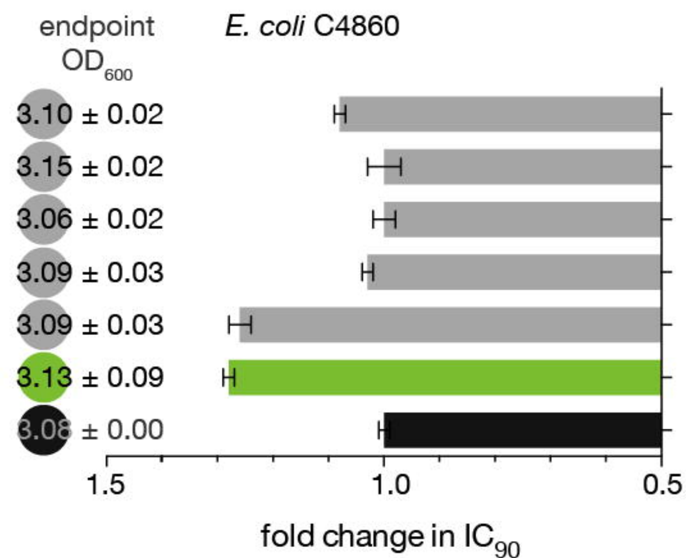











endpoint OD600
E. coli C4860
3.10 ± 0.02
3.15 ± 0.02
3.06 ± 0.02
3.09 ± 0.03
3.09 ± 0.03
3.13 ± 0.09
3.08 ± 0.00
1.5
1.0
0.5
fold change in IC90
E. coli W3110 ΔldcC ΔcadA
pZE-Gene4 (Alistipes finegoldii)
pZE-Gene3 (Bacteroides sp. CAG:770)
pZE-Gene2 (Bacteroides barnesiae)
pZE-Gene1 (Bacteroides coprophilus CAG:333)
pZE-YbjE
pZE-MglE
pZE21
0.5
1.0
1.5
fold change in IC90
endpoint OD600
3.32 ± 0.02
3.31 ± 0.02
3.30 ± 0.02
3.34 ± 0.01
3.44 ± 0.01
3.36 ± 0.02
3.33 ± 0.02