Engineering *Escherichia coli* to utilize erythritol as sole carbon source

Fang Ba, Xiangyang Ji, Shuhui Huang, Yufei Zhang, Wan-Qiu Liu and Jian Li*

Affiliations:

School of Physical Science and Technology, ShanghaiTech University, Shanghai, 201210, China

*Corresponding author. E-mail: lijian@shanghaitech.edu.cn (J.L.)

Abstract

Erythritol, one of the natural sugar alcohols, is widely used as a sugar substitute sweetener in food industries. Humans and their native gut microbiota lack erythritol catabolic microorganisms to utilize erythritol as carbon source, which inspires us to engineer a model chassis that can respond and utilize erythritol for establishing wide applications in synthetic biology fields. Here, we successfully engineered *Escherichia coli* to respond and utilize erythritol as sole carbon source. First, a heterologous erythritol catabolic gene cluster from *Ochrobactrum spp.* enables *E. coli* to catabolize erythritol into erythrose-4-phosphate and flow into pentose-phosphate pathway. Second, we characterized an erythritol binding transcriptional repressor and its DNA binding site. Then, mRNA transcriptional analysis guided us to build up further metabolic engineering strategies for improving erythritol catabolism. Finally, such engineered *E. coli* strains were applied for *in vitro* simulated intestinal fluid experiments and performed as an erythritol detector to distinguish soda drinks. We anticipate erythritol catabolic *E. coli* strains will expand synthetic biology fields and accelerate applications in biomedical engineering and living therapeutics.

Introduction

The worldwide obesity epidemic and correlative diseases like diabetes, hypertension, and hyperlipidemia encourage people to change their lifestyles by limiting the excessive intake of carbohydrates. Food industries meet people's requirements to produce a series of carbohydrate-reduced foods with sugar alcohols (also called polyols) as additional sweeteners[1]. These compounds contain multiple hydroxyl groups and are usually produced by industrial carbohydrates hydrogenation[2]. Erythritol, the C4 sugar alcohol, naturally exists in small species of vegetables, fruits, mushrooms, and fermented foods[3,4]. It also can be found in human tissues such as semen, lens, cerebrospinal fluid, and serum[1,4,5]. Erythritol industrial manufacturing has been well developed by fermentation processes with engineered lactic acid bacteria and yeast[6,7,8,9,10]. Moreover, erythritol is well-characterized with high stability in a wide range of temperature and pH conditions. Also, it does not take part in Maillard reactions and works as a tremendous non-hygroscopic substance[11]. In general, erythritol works as a stable, cheap, and safe sugar alcohol as a natural food sweetener.

The reason why erythritol is used as a food sweetener because human beings cannot catabolize it. It has been reported that most erythritol can be rapidly absorbed in the small intestine and excreted in urine within 24 h[5,11,12,13,14]. Only a small percent of erythritol flow through intestinal fermentation[13]. However, the human gut microbiota lacks erythritol catabolic microorganisms. Thus, nearly all erythritol cannot be utilized by human beings as energy substance[15,16]. From another aspect, erythritol, such a human-compatible compound, can work as a potential energy substrate by engineered microorganisms that utilize erythritol as energy source, or work as a unique micromolecular compound to regulate human physiological

metabolism.

Up to now, only two distinguishing erythritol degradation pathways have been identified from microorganisms: The first one was discovered in several *Hyphomicrobiales* species, including environmental bacterium *Ochrobactrum*[17], pathogenic bacterium *Brucella*[18,19,20,21,22,23], nitrogen-fixing plant endosymbionts *Rhizobium*[22], and *Sinorhizobium*[24]. These microorganisms commonly uptake erythritol as the initial substrate and yield erythrose-4-phosphate via 5-step catalysis[17]. Erythrose-4-phosphate is the intermediate product of pentose-phosphate pathway and can be utilized as a C4 carbon source. The second pathway was discovered in *Mycolicibacterium smegmatis*[25]. This pathway is simple 3-step catalysis, and the final product is also erythrose-4-phosphate. In general, most other organisms are not able to utilize erythritol as energy source. Meanwhile, synthetic biology takes the core mission of (re)designing and constructing new biological parts, circuits, and systems for useful purposes. As such goals inspired, we purposefully aimed to reprogram a general model microorganism to catabolize erythritol as energy source for potential usage.

To meet this demand, we chose *Escherichia coli*, one of the most used model prokaryote and GRAS (generally recognized as safe) microorganisms as chassis[26,27]. Afterward, we separated and identified an environmental bacterium *Ochrobactrum spp.* from outdoor aerosol and gained an erythritol catabolic gene cluster. Such a 5-step reaction provides *E. coli* capability to utilize erythritol as sole carbon source. Meanwhile, we characterized the erythritol-responded genetic repressor eryD and its DNA-binding site. Furthermore, we did mRNA transcriptional analysis to identify the metabolism of such C4 carbon flux. Relying on the transcriptome results, we designed genetic circuits to facilitate erythritol catabolism and significantly increased the *E.*

coli growth in M9-erythritol medium. For developing applications of such engineered erythritol catabolic microorganism, probiotic *E. coli* Nissle 1917 (EcN) were designed to utilize erythritol, and such catabolism promoted EcN to utilize erythritol in simulated intestinal fluid (SIF) for growth. Finally, our engineered *E. coli* could be used for distinguishing soda drinks whether they contain high-fructose corn syrup (HFCS, which contains sugar) or erythritol (sugar-free sweetener). We anticipate that our erythritol catabolic *E. coli* strains will accelerate synthetic biology with broad potential research and applications such as carbon cycle, agriculture, genetic circuit design, metabolic engineering, biomedical engineering, and living therapeutics.

111 **Results**

112 **Engineered *E. coli* can utilize erythritol as sole carbon source**

113 To gain reliable erythritol catabolic microorganisms from the natural environment, we
114 designed an erythritol-based selective medium derived from M9 minimal medium, one of the
115 chemically defined mediums for *E. coli* incubation. Here we proposed that such erythritol-
116 based selective medium would separate *E. coli*-like gram-negative bacteria that are suitable for
117 further exploration. Hence, we replaced 0.4% glucose (C6 carbon source in M9 minimal
118 medium) with 0.4% erythritol and called it as M9-erythritol medium, while the other
119 components were not changed. After this process, we successfully separated, identified, and
120 characterized an environmental bacterium *Ochrobactrum spp.* from campus aerosol
121 (**Supplementary Fig. 1** and **2**)[28]. Then, ten erythritol catabolism-associated genes were isolated
122 and sequenced from *Ochrobactrum spp.* genome (**Supplementary Fig. 3-5**), DNA sequence
123 alignments suggested that all the genes were almost the same as reported of *Brucella abortus*
124 *2308*[29,30], including erythritol catabolic genes eryA, eryB, eryC, eryH, eryI, putative erythritol
125 binding transcriptional repressor eryD, eryR, and erythritol ABC-transporter genes eryE, eryF,
126 eryG (**Fig. 1a**). Thibault *et al.* has been characterized the 5-step erythritol catabolism proteins
127 *in vitro*, which started with an erythritol kinase eryA to produce D-erythritol 1-phosphate, then
128 the phosphorylated intermediate was oxidized by a membrane FAD-dependent dehydrogenase
129 eryB and yielding D-erythrulose 1-phosphate, which would next be isomerized to L-erythrulose
130 1-phosphate by an isomerase eryC. Next, another isomerase eryH could catalyze phosphate
131 transportation to produce D-erythrulose 4-phosphate, finally, isomerase eryI converted D-
132 erythrulose 4-phosphate to D-erythrose 4-phosphate, which was one of the intermediates in

133 pentose-phosphate pathway[17]. Such C4 phosphate-substrate could be utilized for carbohydrate
134 utilization via glycolysis, nucleotide synthesis, and aromatic amino acids biosynthesis via
135 shikimate pathway (**Fig. 1b**).

136 Thibault *et al.* has been purified all five erythritol catabolism proteins from *E. coli*-based
137 protein expression system and successfully characterized their catalytic functions *in vitro*. On
138 this account, we prospected that the original gene cluster might be functional for *E. coli*
139 erythritol catabolism *in vivo*. Thus, we initially constructed a plasmid containing erythritol
140 catabolism genes and transformed it into *E. coli* MG1655 for testing whether engineered *E. coli*
141 strain could utilize erythritol as carbon source. Surprisingly, such *E. coli* could slowly grow in
142 liquid M9-erythritol medium (**Fig. 1c**), within 72 h incubation, *E. coli* grew as an "S" curve and
143 the OD_{600} value finally reached approximately 0.35. To assuredly determine the erythritol
144 consumption, filtered medium supernatant was collected for high performance liquid
145 chromatography (HPLC) analysis, results from the refractive index detector (RID) indicated
146 that during incubation, erythritol was certainly consumed by engineered *E. coli* as carbon
147 source (**Fig. 1d**).

148 After simply demonstrating erythritol catabolism *in vivo*, we next characterized this gene
149 cluster via other aspects. Firstly, the *in vivo* expression of ten proteins was analyzed by SDS-
150 PAGE and Western-Blot (**Extended Data Fig. 1**). The results showed that except eryC, the
151 other proteins could be well expressed as a high percentage of soluble fractions *in vivo*[17].
152 Secondly, all ten genes' promoter and ribosome binding site (RBS) strength were measured
153 with standardization (**Extended Data Fig. 2** and **3**)[31,32,33]. Thirdly, three erythritol ABC-
154 transporter proteins eryE, eryF, and eryG were predicted and demonstrated their expression,

155 transmembrane phenotype, and localization in *E. coli* cells (**Supplementary Fig. 6** and
156 **Extended Data Fig. 4**). It was worth noting that *E. coli* MG1655 could transport and catabolize
157 erythritol without such ABC-transporter system, which means that *E. coli* might carry native
158 isoenzymes for erythritol transport.
159

160 **EryD is an erythritol binding transcriptional repressor**

161 In 2000, Sangari *et al*. reported that erythritol might act as an inducer and probably bind
162 to eryD in *Brucella abortus 2308* to inhibit its repressor activity[19]. However, it lacked more
163 detailed eryD characterization. Interestingly, our separated erythritol cluster of *Ochrobactrum*
164 *spp.* includes an eryD gene whose DNA sequence was totally aligned with eryD in *Brucella*
165 *abortus 2308*. In addition, sequence analysis indicated eryR, the other putative DNA-binding
166 transcriptional repressor might act similarly to eryD. Hence, characterization of such two
167 proteins should help us figure out their regulation mechanism and expand erythritol-associated
168 synthetic biology applications. To meet this demand, we initially attempted to determine the
169 putative DNA-binding sites within two loci: one was the 732 bp gap between eryE and eryA,
170 and the other is approximately 400 bp gap between eryD and eryR (**Supplementary Fig. 3**).
171 Here we designed an *in vivo* characterization strategy to precisely identified the DNA-binding
172 sites: In general, we supposed the repressor's DNA-binding site was localized after a promoter
173 (generally recognized as a transcriptional unit), thus we brought different putative DNA
174 sequences into a standardized sfGFP generation plasmid, which consisted of a strong RBS, an
175 sfGFP coding sequence, and a strong terminator in iGEM standard backbone pSB1C3. Firstly,
176 we successfully identified two promoters called pEryF-732 (F means forward) and pEryR-732

177 (R means reverse), which are located in the 732 bp gap between eryE and eryA
178 (**Supplementary Fig. 7** and **8**). Secondly, to identify the putative DNA-binding sites, other
179 compatible plasmids were designed to constitutively express eryD or eryR, which consisted of
180 gradient strength constitutive promoters, strong RBS, eryD/eryR coding sequences, and strong
181 terminator with compatible vector. The sfGFP fluorescence of cell pellets and Western-Blot
182 results showed that eryD had a DNA-binding site located in pEryF-732, and eryD expression
183 level could regulate transcription level, but eryR did not (**Supplementary Fig. 9**). Subsequently,
184 promoter 5'-region truncation experiments were performed to precisely identified the 5'-site of
185 eryD DNA-binding site (we called it as erythritol operator, eryO), to achieve this goal, pEryF-
186 732 were truncated along with the gradient and brought into sfGFP generation plasmids
187 (**Supplementary Fig. 10-12**). After that, we determined the front site of eryO. Similarly, 3'-
188 site of eryO were also successfully identified (**Supplementary Fig. 13-15**). Finally, eryO was
189 identified as a 29 bp sequence located between the -35 and -10 region of promoter pEryF-732
190 (**Fig. 2a**). To ensure the reliability of *in vivo* results, electrophoretic gel mobility shift assay
191 (EMSA) was performed in which the FAM-labeled DNA fragments harboring the eryE-eryA
192 732 bp intergenic region were incubated with an increasing amount of eryD. As expected, eryD
193 effectively bound to pEryF-732 promoter (**Fig. 2b**), and a further investigation showed that
194 eryD precisely binds to the 29 bp eryO region (**Supplementary Fig. 16**). In addition, we also
195 confirmed that there was no eryD or eryR DNA-binding sites in the 400 bp gap between eryD
196 and eryR (**Supplementary Fig. 17**). Moreover, we concentrated on the eryD protein
197 characterization. Structure prediction of eryD monomer indicated that eryD was assembled as
198 an N-terminal helix-turn-helix motif for DNA interaction and a C-terminal domain for erythritol

199 regulation. SWISS-MODEL prediction showed that eryD is assembled as a homo-tetramer
200 (**Supplementary Fig. 18**), gel-filtration chromatography and Native-PAGE-based Western-
201 Blot demonstrated this conclusion (**Extended Data Fig. 5**).

202 Next, we demonstrated that eryD could be dynamically regulated by gradient
203 concentration of erythritol (**Fig. 2c-f** and **Supplementary Fig. 19**). For the development of
204 eryD-based genetic circuit design, T7 RNA polymerase-based erythritol induction system (**Fig.**
205 **2g, h** and **Supplementary Fig. 20**), and *E. coli* σ70 factor-based native erythritol induction
206 systems (**Fig. 2i, j** and **Supplementary Fig. 20**) were constructed, performed, and successfully
207 demonstrated their function. In the end, the orthogonality of erythritol-eryD interaction was
208 characterized by a series of generally used carbohydrates and polyols, the results showed that
209 eryD could interact with erythritol with great orthogonality (**Fig. 2k, l**).

210 In summary, we demonstrated that eryD is an erythritol-binding transcriptional repressor,
211 identified its DNA-binding site eryO, demonstrated its concentration-dependent dynamic
212 response, excellent substrate orthogonality, and applied it to genetic circuit design.

213

214 **mRNA transcriptional analysis of engineered *E. coli* erythritol catabolism**

215 To determine erythritol catabolism in engineered *E. coli* strains, we performed comparative
216 mRNA transcriptional analysis on triplicate exponential phase cultures of *E. coli* MG1655 that
217 grew between M9-erythritol medium and M9-glucose medium (**Extended Data Fig. 6**). Totally,
218 1298 genes were significantly different between these two mediums, including 759 genes
219 upregulated, and 539 genes downregulated (**Supplementary Fig. 21**). COG annotations
220 showed that the significant differences were found in genes related to carbohydrate metabolism,

221 amino acid metabolism, energy metabolism, transcription, inorganic ion metabolism, and
222 ribosome-associated translation (**Extended Data Fig. 7**). It was worth noting that most of the
223 genes related to carbohydrate metabolism were overexpressed, which indicates that erythritol
224 was not efficiently utilized as carbon source for engineered *E. coli* than glucose. In addition,
225 nearly all ribosome-associated genes were downregulated, which means that such strain
226 decreased its translation level to cope with such a low-carbon-source environment. Besides,
227 KEGG and GO annotation showed similar conclusions.

228 To take a panoramic view of carbohydrate metabolic pathways of erythritol-utilized *E. coli*
229 MG1655, transcriptomic data were mapped in a schematic diagram including heterologous
230 erythritol degradation pathway, pentose-phosphate pathway, glycolysis, and TCA cycle (**Fig.**
231 **3**). The end product of erythritol catabolism erythrose-4-phosphate can flow into pentose-
232 phosphate pathway via two pathways: one way is transaldolases talA and talB convert
233 erythrose-4-phosphate and β-fructose-6-phosphate into glyceraldehyde-3-phosphate and
234 sedoheptulose-7-phosphate, another way is transketolase tktA and tktB convert erythrose-4-
235 phosphate and xylulose-5-phosphate into β-fructose-6-phosphate and glyceraldehyde-3-
236 phosphate. Transcriptome showed that talA, talB, tktA, and tktB were overexpressed at 5.2, 2.0,
237 1.3, and 5.9 times, which provided metabolic engineering strategies for erythritol catabolism
238 improvement. Several carbohydrate degradation pathways were upregulated, including D-
239 xylose, L-arabinose, lactose, L-rhamnose, and glycerol. However, only galactose and D-
240 mannose degradation pathways were downregulated. In general, the transcriptome comparison
241 between erythritol degradation and glucose degradation was not significantly different in
242 glycolysis and TCA cycle, however, the major distinction was erythrose-4-phosphate

243 downstream conversion in pentose-phosphate pathway.

244 Another concentrated point was that both KEGG and COG annotations indicated much
245 downregulated ribosome-related genes (**Supplementary Information**). In detail, KEGG
246 annotated 0 ups and 52 downs of ribosome, and COG annotated 4 ups and 70 downs of
247 translation, ribosomal structure, and biogenesis. These results suggested that erythritol
248 utilization should cause remarkable reduction of ribosome assembly, which led to suppression
249 of translation and dispensable protein synthesis.

250 In addition, after the differentiated gene screening, some distinct upregulated genes and
251 gene clusters were summarized (**Supplementary Fig. 22**). We noticed that the genes of *E. coli*
252 glycolate utilization operon *glcCDEFGBA* were all significantly overexpressed, even glcD was
253 upregulated as 109 times. There was no glycolate or analogue in M9-erythritol medium, so we
254 considered whether there was a glc cluster transcriptional repressor that might respond to
255 erythritol catabolism. In this cluster, glcC is a reported DNA-binding transcriptional repressor
256 that responds to glycolate and acetate[34], hence we did experiments and demonstrated that glcC
257 could respond to erythrose-4-phosphate or its downstream substrates (**Supplementary Fig. 23**).
258 Furthermore, another cluster attracted our attention that the genes of phosphonate uptake and
259 utilization *phn* operon were nearly all upregulated (**Supplementary Fig. 22**)[35], we have
260 characterized that the H-T-H type transcriptional repressor phnP was not responded to erythritol
261 catabolism (data not shown), thus we hypothesized that erythritol utilization may request much
262 phosphonate because erythritol kinase eryA could catalyze erythritol phosphorylation to
263 produce D-erythritol 1-phosphate.

264

265 **Metabolic engineering to improve *E. coli* erythritol catabolism**

266 Up to now, we have successfully demonstrated that engineered *E. coli* can utilize erythritol
267 as sole carbon source, and transcriptome analysis provided much more detail to inspire us on
268 improving erythritol catabolism. As mentioned above, wild type *E. coli* MG1655 with native
269 erythritol catabolism plasmid (pFB147) could grow to stationary phase and OD_{600} reached
270 approximately 0.35 in 60 h (**Fig. 1d**), so we attempt to increase its growth rate and cell density.
271 At first, we tried to screen out a better promoter to improve erythritol catabolism cluster
272 transcription, here we chose a constitutive promoter family with gradient strength to replace
273 native promoter pEryF-732 (**Fig. 4a**), however, there was no clear distinction between pEryF-
274 732 and the other two medium strength promoters J23106 and J23105, interestingly, the
275 strongest promoter J23100 and weak promoter j23114 suppressed growth, even the weakest
276 promoter J23109 could not support growth (**Fig. 4b, c**). These results suggested that a suitable
277 transcription level was sufficient for erythritol catabolism. Next, according to transcriptome
278 analysis, transaldolases talA, talB, and transketolase tktA, tktB were upregulated, so we
279 supposed that overexpression of these genes should increase consumption of erythrose-4-
280 phosphate to accelerate carbon flux and cell growth. To demonstrate our hypothesis, we
281 constructed three compatible plasmids that are able to constitutive express talA-tktB, talB, or
282 tktA (**Fig. 4d, f, h**). Additional annotation is that talA is genetically located next to tktB in the
283 genome, so here we only constructed a talA-tktB plasmid rather than two separated plasmids.
284 Surprisingly, the OD_{600} of talA-tktB overexpressed erythritol catabolism *E. coli* could reach
285 approximately 1.0 after 72 h culture (**Fig. 4e**) with optimal promoter J23114. Similarly, the
286 OD_{600} of talB or tktA overexpressed erythritol catabolism *E. coli* could reach approximately 0.8

287 after 72 h culture (**Fig. 4g, i**). Together, after metabolic engineering, *E. coli* endogenous talA-
288 tktB overexpression could accelerate its erythritol utilization and cell growth (OD_{600} value
289 increase from 0.35 to 1.0 after 72 h culture).
290

291 **Erythritol transport system characterization**

292 As demonstrated above, wild type *E. coli* MG1655 containing erythritol catabolism
293 plasmid (pFB147) was able to utilize erythritol, which means that some native *E. coli* genes
294 may act as erythritol transporter. In *Ochrobactrum spp.*, erythritol ABC-transporter system
295 consisted of eryE, eryF, and eryG[22]. We demonstrated their transmembrane phenotype in *E. coli*
296 Mach1-T1 cells (**Extended Data Fig. 4**), however, when we tried to bring the original cluster
297 into different *E. coli* strains, we found that erythritol ABC-transporter may disturb cell division
298 (**Supplementary Fig. 24**). Therefore, we focused on native genes in *E. coli* MG1655 that assist
299 erythritol transport.

300 One way was to find out homologies of erythritol ABC-transporter (**Extended Data Fig.**
301 **8**), we firstly summarized all carbohydrate or polyol transport genes in *E. coli* MG1655, then
302 BLAST protein-protein alignments were proceeded to screen out homologies of erythritol
303 permease eryG[36,37]. Finally, seven eryG homologous permeases were screened out, including
304 glycerol facilitator glpF, which has been reported to enable *E. coli* to transport erythritol[38].
305 Subsequently, we transform pFB147 into seven knock-out strains to characterize whether these
306 genes could facilitate erythritol transport. Statistical analysis showed that six sugar-ABC
307 transporter systems might facilitate erythritol transport (**Extended Data Fig. 9**). Another way
308 was to rely on the transcriptome result. We summarized 34 upregulated putative carbohydrate

309 transport genes, constructed knock-out strains, and performed the same experiments, however,
310 statistical analysis showed that there were no putative systems to facilitate erythritol transport
311 (**Supplementary Fig. 25**). Together, these results demonstrated that part of *E. coli* native
312 carbohydrate transport systems might facilitate erythritol transport.
313

314 **Erythritol can facilitate growth of engineered probiotic *E. coli* Nissle 1917 in simulated**
315 **intestinal fluid**

316 Following the demonstration, we subsequently attempted to develop and expand its
317 practical applications. Erythritol is one of the noteworthy and widely-used sweeteners in daily
318 foods and drinks, and it also can briefly remain in the human GI tract without digestion, hence
319 we supposed that a kind of simply engineered probiotic which could respond and utilize
320 erythritol might adjust and balance the gut microbiome. Here we focused on the probiotic
321 chassis *E. coli* Nissle 1917 (EcN) which has been widely characterized, engineered, and applied
322 for diagnosis and therapy[27,39,40,41,42]. Firstly, erythritol catabolism plasmid pFB147 was
323 successfully transformed into EcN and endowed its erythritol utilization (**Supplementary Fig.**
324 **24**). Next, such engineered EcN was mixed with simulated intestinal fluid (SIF) in centrifuge
325 tubes to establish *in vitro* anaerobic culture condition at 37°C. After days of incubation, EcN
326 cell density could be calculated by counting colony number on LB-agar plates, these colonies
327 were separated from diluted cell culture (**Fig. 5a**). There were two facts for determination: the
328 initial cell density and erythritol concentration. At the start, we decided to generate a gradient
329 cell density of 10^2, 10^4, 10^6, and 10^8 (**Fig. 5b-e**), with 0.4% erythritol (as the same concentration
330 as M9-erythritol medium). The growth curves showed that engineered EcN could reach a higher

331 cell density than various initial cell densities. 10^2 to 10^6 incubations could reach the highest cell
332 density on day 2, however, 10^8 was on day 1 and declined after day 2, which suggested that
333 high cell density may excessively consume SIF resources. We noticed that the growth curves
334 between 10^4 and 10^6 were exactly similar, so we decided to use the lower density (10^4 EcN) for
335 further experiments. Next, six different erythritol concentrations (0%, 0.01%, 0.1%, 1%, 4%,
336 and 16%) were used in SIF for incubation (**Fig. 5f-k**), we found that 16% erythritol might limit
337 cell growth probably due to high osmosis pressure. The other five erythritol concentrations
338 more or less facilitate engineered EcN growth, the highest cell density reached approximately
339 5×10^9 on day 1 with 0.01% erythritol, and 0.1% erythritol could support EcN alive up to
340 approximately 6×10^8 on day 3. Nonetheless, 0% additional erythritol could also facilitate
341 engineered EcN growth, this was probably due to the remained erythritol in bile salts[43].
342 According to the mentioned incubations, the optimal condition for engineered EcN growth is
343 SIF with additional 0.1% erythritol and 10^4 initial cell density. In summary, we demonstrated
344 that erythritol could facilitate erythritol catabolic engineered EcN growth in SIF.
345

346 **Engineered *E. coli* can distinguish soda drinks**

347 To further expand the applications of erythritol-utilized *E. coli*, we are supposed to build
348 up a strategy for the distinction between soda drinks. In general, there are three types of sweet
349 soda drinks: the first one is sugary drinks that generally contain high-fructose corn syrup
350 (including fructose and glucose) and white sugar (sucrose). The second one is sugar-free drinks
351 with artificial sweeteners, including acesulfame potassium, sucralose, aspartame, saccharin,
352 and neotame. The third one is sugar-free drinks with natural sweeteners, including polyols

353 (erythritol, xylitol, isomaltol, mannitol, sorbitol, allulose, and so on), and other naturally
354 derived components like stevia and monk fruit sweetener. In general, people cannot
355 differentiate sweet sodas by tasting, however, the analysis of the specific components relies on
356 biochemistry analysis by HPLC, GC-MS, and other strategies, which is complicated, costly,
357 and time-consuming[44]. However, our engineered *E. coli* strain may perform as a sweet soda
358 detector. Theoretically, wild type *E. coli* strains can utilize fructose and glucose as carbon
359 sources, and they lack catabolic pathways to utilize sucrose, artificial sweeteners, and some
360 natural sweeteners like erythritol (**Supplementary Fig. 26**). To summarize, we decided to
361 distinguish different sweet sodas by engineered *E. coli* strains.

362 Here, we designed two different *E. coli* strains: one was called Negative Control (NC) due
363 to it consisted of a weak promoter (J23109) that could not support erythritol catabolism, the
364 other one was called Erythritol Detector (ED), which could utilize erythritol (**Fig. 6a**). Next,
365 two strategies were built up via liquid incubation and solid incubation. As for liquid incubation,
366 to prove the concept, four M9 salt-based liquid mediums were composed of additional ddH_2O,
367 fructose, glucose, and erythritol. NC and ED were inoculated into these mediums respectively
368 for days' incubation. As we expected, Both NC and ED could not grow without carbon source
369 (ddH_2O) but could rapidly grow with fructose and glucose, and only ED could slowly grow
370 after 36 h with erythritol but NC could not (**Fig. 6b**). After successfully proving the concept,
371 then we replaced the additional carbon sources by three boughten sodas (10× diluted for use).
372 Soda 1 was a sugar-free drink that contained two artificial sweeteners: acesulfame potassium
373 and sucralose. Soda 2 was a sugary drink with fructose and glucose. Soda 3 was a sugar-free
374 drink that consisted of erythritol and sucralose. After similar incubation, NC could not grow in

375 soda 3, but ED could utilize erythritol and grow in soda 3 (**Fig. 6c**). These results indicated that
376 our engineered *E. coli* strains could easily distinguish different sweet soda drinks via liquid
377 incubation.

378 In another way, a solid incubation strategy was also built up: prepared M9 salt-agar plates
379 with additional carbon source liquid could perform as solid medium to verify whether the *E.*
380 *coli* could utilize such carbon source as carbon source. As expected, the colony of both NC and
381 ED could not be observed without carbon source (ddH_2O) and could be observed with fructose
382 and glucose, only ED could form colony after 24 h with erythritol but NC could not (**Fig. 6d**
383 **and Supplementary Fig. 27**). Then we used the same three sodas to successfully demonstrated
384 that solid incubation strategy could also distinguish different sweet soda drinks (**Fig. 6e**). In
385 summary, we developed simple, visualized, and reliable strategies to verify different soda
386 drinks by engineered erythritol catabolic *E. coli* strains.

387

388 **Discussion**

389 In this project, we demonstrated that engineered *E. coli* could transport and utilize the
390 natural sweetener erythritol as sole carbon source for growth. Firstly, we gained erythritol
391 catabolism gene cluster from a separated environmental bacterium *Ochrobactrum spp.*.
392 Secondly, we successfully characterized erythritol catabolism-related genes from three aspects,
393 including an erythritol-responded transcriptional repressor, an erythritol-catabolic gene cluster,
394 and an erythritol ABC-transporter system. For in-depth exploration, mRNA transcriptional
395 analysis was used to compare distinctions between erythritol and glucose as C4 and C6 carbon
396 sources. Such results inspired us to improve erythritol catabolic efficiency via metabolic

engineering strategies, and we finally remarkably improved the *E. coli* growth in M9-erythritol medium. Moreover, we applied such engineered strains to expand usages in two fields: On the one hand, erythritol catabolism enabled engineered probiotic EcN to utilize erythritol in SIF for increased growth. On the other hand, our engineered *E. coli* performed as "erythritol detector" to distinguish different sweet soda drinks via two simple strategies.

Overall, our work highlights the potential applications of erythritol catabolism, which will be a promising addition to the current synthetic biology field. What is more, we developed a logical research procedure that enables the model chassis to achieve specific goals: Firstly, we raised the question of how to make full use of erythritol to solve problems and expand applications rather than just as a natural sweetener in foods and drinks. Secondly, we set a goal that we want *E. coli* to utilize erythritol as sole carbon source. Thirdly, we acquired resources of separated *Ochrobactrum spp.* and erythritol catabolism-associated genes. Fourthly, proof of concept and optimizations were achieved via different strategies. Finally, some expanded applications might be necessary. In conclusion, such a procedure may inspire synthetic biology researchers to find out scientific topics and solve practical problems.

Looking forward, we expect our research will provide strategies for expanding synthetic biology applications, such *E. coli*-based synthetic capability would build up new approaches for bringing novel abilities to heterologous organisms for useful purposes (**Supplementary Fig. 28, 29**), just as our erythritol catabolism *E. coli* developed connections between lab and daily life. We can imagine that in the near future, sugar-free drinks will not only provide sweet tastes but also balance our intestinal flora, and even treat diseases via engineered probiotics. Furthermore, erythritol will play much more important roles in different fields, including

energy and environment science and engineering, agriculture, food science and technology, materials science and engineering, and biomedical engineering.

441 **Methods**

442 **Strains, vectors, plasmids, primers, and reagents**

443 The details of *E. coli* strains, vectors, plasmids, and primers used in this study are listed in
444 **Supplementary Table 1, 4, 5,** and **6**, respectively. All plasmid sequences are listed in
445 **Supplementary Information**, these plasmids were checked for correctness with Sanger
446 sequencing (GENEWIZ) unless otherwise noted. Oligonucleotides were synthesized by
447 GENEWIZ and listed in Supplementary Table 6. Q5 High-Fidelity DNA Polymerase (New
448 England Biolabs), Phanta Super-Fidelity DNA Polymerase (Vazyme), FastPure Gel DNA
449 Extraction Mini Kit (Vazyme), and ClonExpress Ultra One Step Cloning Kit (Vazyme) were
450 used for molecular cloning. DreamTaq Green PCR Master Mix (Thermo Scientific) was used
451 for colony PCR. Lysogeny Broth (LB) liquid medium contains 10 g tryptone (Oxoid), 5 g yeast
452 extract (Oxoid), and 10 g sodium chloride (Titan/Greagent) in 1 L ddH_2O. LB-agar plate was
453 prepared by adding 15 g agar (Sangon Biotech) per liter LB liquid medium. M9 minimal
454 medium (also M9-glucose medium) contains 200 mL 5× M9 salts (15 g/L KH_2PO_4, 2.5 g/L
455 NaCl, 33.9 g/L Na_2HPO_4, and 5 g/L NH_4Cl), 20 mL 20% w/v glucose, 2 mL 1M $MgSO_4$, and
456 100 μL 1M $CaCl_2$ in 1 L ddH_2O. M9-erythritol medium contains 200 mL 5× M9 salts, 20 mL
457 20% w/v erythritol, 2 mL 1M $MgSO_4$, and 100 μL 1M $CaCl_2$ in 1 L ddH_2O. Antibiotic stocks
458 (1000×) are 100 mg/mL ampicillin, 50 mg/mL kanamycin, and 34 mg/mL chloramphenicol.

459

460 **Genetic parts and gene clusters**

461 All genetic parts (promoter, RBS, promoter-RBS, terminator, CDS, and protein binding
462 site) and gene clusters used in this study are listed in **Supplementary Table 2,** and **3**. Genetic

463 parts shorter than 100 bp were synthesized within the oligonucleotides (GENEWIZ) and
464 inserted into PCR fragments during molecular cloning. Gene clusters were amplified from the
465 separated *Ochrobactrum spp.* genome, the referenced genome for DNA-DNA alignments was
466 *Brucella anthropi* strain T16R-87 chromosome 2 (GenBank: CP044971.1). Gene clusters were
467 checked for correctness with Sanger sequencing (GENEWIZ), and some of the genetic parts
468 were amplified from gene clusters.
469
470 **Plasmids construction**
471 All plasmids were constructed by Gibson Assembly. In brief, all linear polymerase chain
472 reaction (PCR) products (Q5 High-Fidelity DNA Polymerase, or Phanta Super-Fidelity DNA
473 Polymerase) were extracted by gel extraction and then assembled by using ClonExpress Ultra
474 One Step Cloning Kit (Vazyme). After assembly, the reaction mixture was added to 50 μL
475 competent *E. coli* Mach1-T1 cells, which were incubated overnight on LB-agar plate. Then,
476 DreamTaq Green PCR Master Mix (Thermo Scientific) was used for colony PCR. The possible
477 PCR products were sent to GENEWIZ for Sanger Sequencing.
478
479 **Western-Blot assay**
480 The *E. coli* pellets were collected in 1.5 mL centrifuge tubes at 5000 g and 4°C for 10 min.
481 The supernatant was discarded and the pellets were resuspended with 1 mL 1× phosphate-
482 buffered saline (pH 7.4) and lysed by sonication (Q125 sonicator, Qsonica, 10 s on/off, 50% of
483 amplitude, input energy ~600 Joules). The lysate was then centrifuged at 12000 g and 4°C for
484 10 min. The total (cell lysate, T), soluble (supernatant, S), and pellet (P) fractions were
485 separated by SDS-PAGE or Native-PAGE (Omni-Easy One-Step PAGE Gel Fast Preparation
486 Kit, EpiZyme), followed by wet transferring to PVDF membrane (Bio-Rad) with 1× transfer

487 buffer (25 mM Tris-HCl, 192 mM glycine and 20% v/v methanol in 1 L ddH_2O, pH 8.3). Then,
488 the PVDF membrane was blocked (Protein Free Rapid Blocking Buffer, Epizyme) for 1 h at
489 room temperature. After washing thrice with TBST for each 5 min, 1:10000 (TBST buffer-
490 based) diluted His-Tag Mouse Monoclonal Antibody (Proteintech) solution was added to the
491 membrane and incubated for 1 h at room temperature. After washing thrice with TBST for each
492 5 min, 1:10000 (TBST buffer-based) diluted HRP-Goat Anti-Mouse IgG (H+L) Antibody
493 (Proteintech) solution was added to the membrane and incubated for another 1 h at room
494 temperature. After the last washing with TBST thrice for each 5 min, the membrane was
495 visualized using Omni ECL reagent (EpiZyme) under UVP ChemStudio (analytikjena).
496

497 ***Ochrobactrum spp.* separation, identification, and characterization**

498 25 mL 1× M9-erythritol medium were added into a 50 mL centrifuge tube, the lid was
499 replaced by gauze to block aerosol dust and enable microorganisms to drop into the medium.
500 Such device was stationary-placed outdoors at approximately 25°C for about 7 days. On the
501 seventh day, the medium became turbid, then the mixture was separated onto an M9-erythritol
502 agar plate for picking single colonies at 25°C. After three days, some white colonies were
503 formed and picked up for identification. 16s rRNA was sequenced by primer pair 27F/1492R,
504 cell morphology was characterized by scanning electron microscope (SEM, JSM-7800F Prime)
505 and transmission electron microscope (TEM, JEM-2100 Plus). *Ochrobactrum spp.* genome
506 sequence was referred to *Brucella anthropi* strain T16R-87 chromosome 1 and 2 (GenBank:
507 CP044970.1 and CP044971.1).
508

509 **Erythritol catabolic *E. coli* incubation in M9-erythritol liquid medium**

510 All engineered *E. coli* strains that could utilize erythritol as sole carbon source were
511 cultured and followed the same procedure: Cell pellets from overnight LB culture were
512 collected in 1.5 mL centrifuge tubes at 5000 g and 4°C for 10 min. The supernatant was
513 discarded and the pellets were washed thrice by M9-erythritol medium, then the pellets were
514 resuspended with M9-erythritol and diluted to OD_{600}=1.0 for standardization. Then, the
515 resuspended mixture was inoculated into new M9-erythritol medium (100 mL medium in a 250

516 mL conical flask, or 1 L medium in a 2 L conical flask) as 1:500 v/v for the following incubation,
517 in addition, all antibiotic concentration in M9-erythritol medium was diluted for five times for
518 avoiding growth inhibition in the minimal medium as far as possible.
519

520 **Erythritol HPLC analysis**

521 1 mL M9-erythritol medium (with *E. coli*) cell cultures were collected in 1.5 mL centrifuge
522 tubes at 5000 g and 4°C for 10 min, the supernatants were passed through a 0.22 μm filter to
523 clear cell pellets and other precipitation. Then the samples were loaded for the following HPLC
524 analysis (**Supplementary Fig. 30**).
525 1260 Infinity II Prime LC System (Aligent) was used as HPLC instrument, Hi-Plex H
526 column (Aligent) was used for carbohydrate analysis. The experiment parameters were set as
527 follows: column temperature: 60°C, mobile phase: 5 mM H_2SO_4 solution, flow rate: 0.6
528 mL/min, detector: Refractive Index Detector (RID) at 40°C, sample loading volume: 2 μL,
529 testing time: 20 min for each sample.
530

531 **eryD binding site characterization *in vivo***

532 eryD binding site (eryO) was characterized and located in pEryF-732, which was a 732 bp
533 gap between eryE (5' start) and eryA (3' end). In general, the characterization was performed
534 as *in vivo* incremental truncation process in two steps:
535 Firstly, to determine the 5' site, a series of 100 bp-truncated variant plasmids (pFB186 to
536 pFB191, and pFB210) were constructed and characterized that eryO was located between
537 pEryF-232 to pEryF-132. Secondly, a series of 10 bp-truncated variant plasmids (pFB192 to
538 pFB198, pFB208, and pFB209) were constructed and characterized that eryO was located
539 between pEryF-162 to pEryF-152. Thirdly, a series of 1 bp-truncated variant plasmids (pFB202
540 to pFB207) were constructed and characterized that eryO 5' site was located in pEryF-159.
541 Then, to determine the 3' site, the truncation process was the same as mentioned above,
542 the constructed plasmids (pFB214 to pFB238) were characterized step-by-step. Finally, eryO
543 3' site was located in pEryF-131, and eryO DNA sequence was characterized as a 29 bp DNA
544 sequence of 5'-gaaaaaaaatgcgccatctagaaaatttt-3'.

545

546 **Electrophoretic mobility shift assay (EMSA)**

547 DNA fragment containing the 732 bp eryE-eryA intergenic region (pEryF-732) was
548 amplified with the 5'FAM-labeled primers from plasmid pFB147 and then gel purified. The
549 DNA fragment (200 nM) was incubated with different concentrations of purified eryD in the
550 binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 mM NaCl, 0.1 mM DTT, 10 μg/mL
551 BSA, 5% v/v glycerol). The 20 μL mixtures were incubated for 30 min at room temperature
552 and then mixed with 5 μL of the same buffer supplemented with 50% v/v glycerol and
553 bromophenol blue. Free DNA and eryD–DNA complexes were separated on 7.5%
554 polyacrylamide gels in 1x TAE buffer (4.844 g/L Tris base, 1.21 mL/L acetate acid, 0.372 g/L
555 EDTA disodium salt dihydrate, pH 8.3). The fluorescence of FAM-labeled bands was visualized
556 under UVP ChemStudio (analytikjena).

557

558 **eryD orthogonality characterization**

559 *E. coli* Mach1-T1 with plasmids pFB186 and pFB261 was used for characterization.
560 Inducers (erythritol, glycerol, L-arabinose, D-xylose, D-glucose, D-mannose, D-galactose, D-
561 fructose, L-sorbose, L-rhamnose, D-sucrose, Lactose, Maltodextrin, D-sorbitol, D-mannitol)
562 were dissolved in ddH_2O as different concentration (5%, 1%, 0.1%, 0.01%, w/v) solutions, then
563 the solutions were passed through a 0.22 μm filter.

564 *E. coli* Mach1-T1 was transformed with the relevant plasmids, starter cultures (LB liquid
565 medium containing 50 μg/mL ampicillin and 17 μg/mL chloramphenicol) were inoculated from
566 a single transformant colony and grew overnight at 37°C for 16 h. The next day, 10 μL starter
567 culture was used to inoculate 5 mL LB medium containing 50 μg/mL ampicillin and 17 μg/mL
568 chloramphenicol in test tubes. The cultures were incubated at 37°C with shaking (250 rpm)
569 until OD_{600} reached 0.6-0.8. The cultures were then added with different inducer solutions and
570 incubated for another 16 h at 37°C and 250 rpm.

571 The next day, cell pellets were collected in 1.5 mL centrifuge tubes at 5000 g and 4°C for
572 10 min. The supernatant was discarded and the pellets were resuspended with 1 mL 1×
573 phosphate-buffered saline (pH 7.4). The sfGFP fluorescence of resuspended mixture was

574 measured by microplate reader (Synergy H1, BioTek) and performed with excitation and
575 emission wavelengths at 485 and 528 nm, respectively. All measurements were performed at
576 least in triplicate. Another part of cell pellets was lysed by sonication (Q125 sonicator, Qsonica,
577 10 s on/off, 50% of amplitude, input energy ~600 Joules). The lysate was then centrifuged at
578 12000 g and 4°C for 10 min. The total fraction was separated and analyzed by Western-Blot.
579

580 **eryD structure modeling**

581 *Ochrobactrum spp.* eryD amino acid sequence was the same as eryD (Uniprot ID: Q2YIQ4)
582 from *Brucella abortus* (strain 2308). AlphaFold, Robetta, and SWISS-MODEL were used for
583 structure modeling.
584

585 **Gel-filtration chromatography**

586 To determine the native molecular weight and assembly model of eryD, gel-filtration
587 chromatography was performed using a Superdex 200 Increase 10/300 GL column (GE
588 Healthcare) and standard proteins including Vlm2 (284 kDa) and T7 RNA polymerase (99 kDa).
589 Purified eryD was concentrated in 1× phosphate-buffered saline (pH 7.4).
590

591 **erythritol ABC-transporter system transmembrane phenotype characterization**

592 TMHMM-2.0 was used for transmembrane prediction: eryE was predicted as a single
593 transmembrane protein, and eryG was predicted as a ten-times transmembrane protein. To
594 determine the transmembrane phenotypes, pFB162 and pFB164 were constructed for eryE-
595 sfGFP and eryG-sfGFP expression. After transformation into *E. coli* Mach1-T1, starter cultures
596 (LB liquid medium containing 34 μg/mL chloramphenicol) were inoculated from a single
597 transformant colony and grew overnight at 37°C for 16 h. The next day, 10 μL starter culture
598 was used to inoculate 5 mL LB medium containing 34 μg/mL chloramphenicol in test tubes.
599 The cultures were incubated at 37°C with shaking (250 rpm) until OD_{600} reached 0.6-0.8. The
600 cultures were then added 1% w/v arabinose and incubated continuously for another 6 h at 30°C
601 and 250 rpm. Afterward, cell pellets from 1 mL cell culture were collected in 1.5 mL centrifuge
602 tubes at 5000 g and 4°C for 10 min. The supernatant was discarded and the pellets were

603 resuspended with 1 mL 1× phosphate-buffered saline (pH 7.4). Then the mixture was imaged
604 under laser scanning confocal microscope (FV3000, Olympus).
605
606 **Protein purification**
607 Plasmids pFB161 and pFB167 were used to express and purify eryD and eryR, respectively.
608 These two proteins were expressed and purified by the same protocol. *E. coli* BL21(DE3) was
609 transformed with the relevant plasmids, starter cultures (LB liquid medium containing 100
610 μg/mL ampicillin) were inoculated from a single transformant colony and grew overnight at
611 37°C for 16 h. 5 mL starter culture was used to inoculate 1 L LB medium containing 100 μg/mL
612 ampicillin in a 2 L conical flask. The cultures were incubated at 37°C with shaking (250 rpm)
613 until OD_{600} reached 0.6-0.8. The cultures were then quickly cooled down on ice rapidly to 20°C,
614 and isopropyl-β-D-thiogalactopyranoside (IPTG) was added for induction at a final
615 concentration of 0.5 mM. The cultures then grew for another 16 h at 20°C and 220 rpm.
616 Afterward, cells were harvested by centrifugation (Avanti JXN-26 High-Speed Centrifuge,
617 Beckman Coulter) at 5000 g and 4°C for 10 min, the pellets were resuspended in Buffer A,
618 containing 20 mM sodium phosphate (pH 7.4), 1 M sodium chloride, 1 mM dithiothreitol
619 (DTT), and 50 mM imidazole. The suspension was cooled on ice and then lysed thrice at 1500
620 bar by ultra-high-pressure homogenization (JNBIO). The lysate was then centrifuged at 4°C
621 and 20000 g for 30 min. The supernatant was collected and passed through a 0.22 μm filter.
622 The filtered solution was then purified by Ni^{2+} affinity chromatography using a 1 mL HisTrap
623 FF column (GE Healthcare). The column was equilibrated with 25 mL Buffer A at a constant
624 flow rate of 1mL/min. After column equilibration, the filtered protein solution was loaded,
625 followed by washing with 25 mL Buffer A. Then, bounded proteins were eluted with Buffer B
626 (Buffer A, but with 500 mM imidazole) and collected into 1.5 mL centrifuge tubes with 0.5 mL
627 elution. Each 0.5 mL elution was analyzed by SDS-PAGE. Eluted protein samples were
628 combined in one tube and desalinated using an ultrafiltration tube (Amicon Ultra 3 kDa
629 molecular weight cut-off, Merck/Millipore) with Buffer C (25 mM Tris-HCl (pH 7.5), 1 mM
630 DTT, and 1 M sodium chloride). The desalinated and concentrated protein solution was then
631 mixed with 40% glycerol (v/v in water) with a volume ratio of 1:1. The concentration of the

632 final protein solution was measured at 280 nm after molar attenuation coefficient correction.
633 Purified proteins were flash frozen by liquid nitrogen and stored at -80°C until further use.
634

635 **Protein expression and solubility characterization**

636 Plasmids pFB158 to pFB167 were used to express ten erythritol catabolism-associated
637 genes. *E. coli* strains were transformed with the relevant plasmids, starter cultures (LB liquid
638 medium containing 100 μg/mL ampicillin or 34 μg/mL chloramphenicol) were inoculated from
639 a single transformant colony and grew overnight at 37°C for 16 h. The next day, 10 μL starter
640 culture was used to inoculate 5 mL LB medium containing 100 μg/mL ampicillin or 34 μg/mL
641 chloramphenicol in test tubes. The cultures were incubated at 37°C with shaking (250 rpm)
642 until OD_{600} reached 0.6-0.8. The cultures were then added with inducers (1% w/v arabinose or
643 0.5 mM IPTG) and incubated continuously. Afterward, cell pellets were collected in 1.5 mL
644 centrifuge tubes at 5000 g and 4°C for 10 min. The supernatant was discarded and the pellets
645 were resuspended with 1 mL 1× phosphate-buffered saline (pH 7.4) and lysed by sonication
646 (Q125 sonicator, Qsonica, 10 s on/off, 50% of amplitude, input energy ~600 Joules). The lysate
647 was then centrifuged at 12000 g and 4°C for 10 min. The total (cell lysate, T), soluble
648 (supernatant, S), and pellet (P) fractions were separated and analyzed by SDS-PAGE and
649 Western-Blot.
650

651 **Fluorescence measurement standardization**

652 sfGFP (C-terminal 6×His) was used as fluorescence reporter. pFB286 to pFB291 were
653 performed as standardized sfGFP expression plasmids, these plasmids consisted of four parts
654 within iGEM standardized vector pSB1C3: *E. coli* constitutive promoter library (J23100,
655 J23106, J23105, J23114, J23109, J23113, from strong to weak), strong ribosome binding site
656 B0034, sfGFP, and strong terminator B0015. Other plasmids containing sfGFP were compared
657 to these six standardized plasmids to determine sfGFP expression level. sfGFP fluorescence
658 was measured by microplate reader (Synergy H1, BioTek) and performed with excitation and
659 emission wavelengths at 485 and 528 nm, respectively. All measurements were performed at
660 least in triplicate.

661

662 ***E. coli* genes knock-out**

663 Gene knock-out strategy was derived from Lambda Red recombination. Concretely,
664 plasmids pFB278 to pFB284, and pFB306 to pFB339 were constructed to obtain linear PCR
665 products. The descriptions "homoL" and "homoR" of these plasmids represented "homologous
666 left arm" and "homologous right arm" between the target knock-out gene or gene cluster, these
667 two DNA sequences were designed approximately longer than 300 bp for a higher
668 recombination efficiency. Such linear PCR products (homoL - FRT - KanR - FRT - homoR)
669 were obtained and purified to be used.

670 Wild type *E. coli* MG1655 was firstly transformed with pKD46, on the next day, a single
671 colony was picked up and inoculated in LB medium (with 1% w/v arabinose) and incubated at
672 30°C and 250 rpm for approximately 4-6 h, when OD_{600} reached 0.6-0.8, the cell pellets were
673 collected in 1.5 mL centrifuge tubes at 5000 g and 4°C for 10 min. The supernatant was
674 discarded and the pellets were washed thrice with 4°C ddH_2O. Then 1 μg purified linear PCR
675 products were added to the cell mixture and proceeded electro-transformation (Gene Pulser
676 Xcell system (Bio-Rad) and Gene Pulser/Micropulser Electroporation Cuvettes, 0.1 cm gap
677 (Bio-Rad) were used). Subsequently, colony PCR (primer pair: "homoL" forward primer /
678 "homoR" reverse primer) were performed to test whether the gene or gene cluster was deleted.
679 Finally, the expected knock-out *E. coli* MG1655 strains were continuously incubated for days
680 at 37°C for pKD46 (temperature-sensing plasmid) elimination.

681

682 **Sample preparation and mRNA transcriptome analysis**

683 To process transcriptome analysis, biological triplicate *E. coli* MG1655 (with plasmid
684 pFB147) in both M9-glucose medium (control group) and M9-erythritol medium (experimental
685 group) were performed.

686 Cell pellets from overnight LB culture (37°C, 250 rpm for 16 h) were collected in 1.5 mL
687 centrifuge tubes at 5000 g and 4°C for 10 min. The supernatant was discarded and the pellets
688 were washed thrice by M9-glucose medium or M9-erythritol medium, then the pellets were
689 resuspended and diluted to OD_{600}=1.0 for standardization. Then, such mixtures were inoculated

690 into new M9-glucose medium or M9-erythritol medium (1 L medium in a 2 L conical flask,
691 with 7 mg/mL chloramphenicol) as 1:500 v/v for the following incubation.
692 As for M9-glucose medium, when OD_{600} reached 0.6 (mid-log phase, approximately 6 h
693 after inoculation), cell pellets were collected. As for M9-erythritol medium, when OD_{600}
694 reached 0.15 (mid-log phase, approximately 36 h after inoculation), cell pellets were collected.
695 Then the cell pellets were sent to GENEWIZ for further sample analysis.
696

697 **Engineered EcN growth in SIF**

698 Probiotic *E. coli* Nissle 1917 (EcN, Mutaflor) were transformed with plasmid pFB147,
699 such engineered EcN was prepared and diluted in ddH_2O with gradient density (2×10^2, 2×10^4,
700 2×10^6, 2×10^8) for the following experiment. 2× simulated intestinal fluid (SIF) was prepared
701 that followed standard protocol with additional erythritol (0, 0.02%, 0.2%, 2%, 8%, 32%,
702 w/v)[45,46], the solutions were all passed through 0.22 μM filter. 7.5 mL EcN culture and 7.5 mL
703 2× SIF were mixed as 1:1 volume to gain a 15 mL culture. such anaerobic cultures were
704 stationary-placed for incubation for days at 37°C. At each predetermined time point,
705 approximately 100 μL samples were taken and separated (diluted if necessary) on LB-agar plate
706 for cell density calculation. All experiments were performed at least in triplicate.
707

708 **Soda drinks differentiation**

709 *E. coli* MG1655 with plasmids pFB151 and pFB147 were performed as negative control
710 (NC) strain and erythritol detector (ED) strain, respectively. Three kinds of sweet soda drinks
711 were bought from the market. Their ingredients were shown as follows: Soda 1: water, carbon
712 dioxide, citric acid, potassium citrate, acesulfame potassium, sodium benzoate, sucralose,
713 sodium gluconate, and essence. Soda 2: water, high-fructose corn syrup (HFCS, including
714 fructose and glucose), white sugar (sucrose), carbon dioxide, citric acid, potassium citrate,
715 sodium benzoate, sucralose, acesulfame potassium, essence. Soda 3: water, erythritol, carbon
716 dioxide, citric acid, potassium citrate, sodium bicarbonate, sucralose, essence. All sodas were
717 passed through 0.22 μm filter to remove impurities and then shaken to eliminate carbon dioxide.
718 As for liquid incubation strategy, the sodas were 10 times diluted in ddH_2O to decrease the

ingredient concentration, and the diluted sodas were used as different liquid mediums. NC and ED cell pellets from overnight LB culture (37°C, 250 rpm for 16 h) were collected in 1.5 mL centrifuge tubes at 5000 g and 4°C for 10 min. The supernatant was discarded and the pellets were washed thrice by ddH_2O, then the pellets were resuspended and diluted to OD_{600}=1.0 for standardization. Then, 10 μL mixtures were inoculated into 5 mL diluted sodas and incubated at 37°C, 250 rpm for 12 h, 24 h, and 36 h. The solutions were imaged at each time point.

As for solid incubation strategy, firstly, M9 salt agar plates were prepared (10 mL mixture in 10 mm plastic plates). Next, 1 mL filtered sodas (without dilution) were added and separated on the solid medium and waited for desiccation. Next, 10 μL NC or ED cell mixtures were inoculated onto the solid medium and incubated at 37°C for 24 h, and 48 h. The bacterial colonies were observed and imaged at each time point.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (Nos. 31971348 and 32171427). J. L. also acknowledges the starting grant of ShanghaiTech University.

Author contributions

J. L. and F. B. designed the experiments. F. B., X. Y. J., S. H. H., J. C. Z., and Y. F. Z. performed experiments. X. Y. J. helped perform cell culture. S. H. H. performed HPLC analysis. J. C. Z. performed electron microscope imaging. Y. F. Z. helped molecular cloning. F. B. analyzed the data and drafted the manuscript. J. L. and W. Q. L. revised and edited the manuscript. J. L. conceived and supervised the study. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Grembecka, M. Sugar alcohols-their role in the modern world of sweeteners: a review. *Eur. Food Res. Technol.* **241**, 15-16 (2015).
2. Shankar, P., Ahuja, S. & Sriram, K. Non-nutritive sweeteners: review and update. *Nutrition* **29**, 1293-1299 (2013).
3. Bernt, W. O., Borzelleca, J. F., Flamm, G., & Munro, I. C. Erythritol: a review of biological and toxicological studies. *Regul. Toxicol. Pharmacol*. **24**, S191–S197 (1996).
4. Munro, I. C., Berndt, W. O., Borzelleca, J. F., Flamm, G., Lynch, B. S., Kennepohl, E., Bär, E. A., & Modderman, J. Erythritol: an interpretive summary of biochemical, metabolic,

768 toxicological and clinical data. *Food Chem. Toxicol*. **36**, 1139-1174 (1998).
769 5. Oku, T., & Okazaki M. Laxative threshold of sugar alcohol erythritol in human subjects.
770 *Nutr. Res.* **16**, 577-589 (1996).
771 6. Veiga-da-Cunha, M., Santos, H., & Van Schaftingen, E. Pathway and regulation of
772 erythritol formation in *Leuconostoc oenos*. *J Bacteriol*. **175**, 3941-3948 (1993).
773 7. Richter, H., Vlad, D., & Unden, G. Significance of pantothenate for glucose fermentation
774 by *Oenococcus oeni* and for suppression of the erythritol and acetate production. *Arch.*
775 *Microbiol*. **175**, 26-31 (2001).
776 8. Mirończuk, A., Dobrowolski, A., Rakicka, M., Rywińska, A., & Rymowicz, W. Newly
777 isolated mutant of *Yarrowia lipolytica* MK1 as a proper host for efficient erythritol
778 biosynthesis from glycerol. *Process Biochem.* **50**, 61-68 (2015).
779 9. Rymowicz, W., Rywińska, A., & Marcinkiewicz, M. High-yield production of erythritol
780 from raw glycerol in fed-batch cultures of *Yarrowia lipolytica*. *Biotechnol. Lett*. **31**, 377-
781 380 (2009).
782 10. Qiu, X., Xu, P., Zhao, X., Du, G., Zhang, J., & Li, J. Combining genetically-encoded
783 biosensors with high throughput strain screening to maximize erythritol production in
784 *Yarrowia lipolytica*. *Metab. Eng*. **60**, 66-76 (2020).
785 11. de Cock, P. Erythritol. in *Sweeteners and Sugar Alternatives in Food Technology* (eds.
786 O'Donnell, K. & Kearsley, M. W.) 213-241 (Wiley Online Library, 2012).
787 12. Livesey G. Health potential of polyols as sugar replacers, with emphasis on low glycaemic
788 properties. *Nutr. Res. Rev*. **16**, 163-191 (2003).
789 13. Arrigoni, E., Brouns, F., & Amadò, R. Human gut microbiota does not ferment erythritol. *Br.*
790 *J. Nutr*. **94**, 643-646 (2005).
791 14. den Hartog, G. J., Boots, A. W., Adam-Perrot, A., Brouns, F., Verkooijen, I. W., Weseler, A.
792 R., Haenen, G. R., & Bast, A. Erythritol is a sweet antioxidant. *Nutrition* **26**, 449-458
793 (2010).
794 15. Ruiz-Ojeda, F. J., Plaza-Díaz, J., Sáez-Lara, M. J., & Gil, A. Effects of Sweeteners on the
795 Gut Microbiota: A Review of Experimental Studies and Clinical Trials. *Adv Nutr*. **10**, S31-
796 S48 (2019).
797 16. Carocho, M., Morales, P., & Ferreira, I. Sweeteners as food additives in the XXI century:

A review of what is known, and what is to come. *Food Chem. Toxicol*. **107**, 302-317 (2017).

17. Barbier, T., Collard, F., Zúñiga-Ripa, A., Moriyón, I., Godard, T., Becker, J., Wittmann, C., Van Schaftingen, E., & Letesson, J. J. Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in *Brucella*. *Proc. Natl. Acad. Sci. U. S. A*. **111**, 17815-17820 (2014).
18. Lillo, A. M., Tetzlaff, C. N., Sangari, F. J., & Cane, D. E. Functional expression and characterization of EryA, the erythritol kinase of *Brucella abortus*, and enzymatic synthesis of L-erythritol-4-phosphate. *Bioorg. Med. Chem. Lett*. **13**, 737-739 (2003).
19. Sangari, F. J., Agüero, J., & Garcı A-Lobo, J. M. The genes for erythritol catabolism are organized as an inducible operon in *Brucella abortus*. *Microbiology (Reading)* **146**, 487-495 (2000).
20. Sperry, J. F., & Robertson, D. C. Erythritol catabolism by *Brucella abortus*. *J. Bacteriol*. **121**, 619-630 (1975).
21. Rodríguez, M. C., Viadas, C., Seoane, A., Sangari, F. J., López-Goñi, I., & García-Lobo, J. M. Evaluation of the effects of erythritol on gene expression in Brucella abortus. *PLoS One* **7**, e50876 (2012).
22. Geddes, B. A., Hausner, G., & Oresnik, I. J. Phylogenetic analysis of erythritol catabolic loci within the *Rhizobiales* and proteobacteria. *BMC Microbiol*. **13**, 46 (2013).
23. Barbier, T. et al. *Brucella* central carbon metabolism: an update. *Crit. Rev. Microbiol*. **44**, 182-211 (2018).
24. Geddes, B. A., & Oresnik, I. J. Genetic characterization of a complex locus necessary for the transport and catabolism of erythritol, adonitol and L-arabitol in *Sinorhizobium meliloti*. *Microbiology (Reading)* **158**, 2180-2191 (2012).
25. Huang, H., Carter, M. S., Vetting, M. W., Al-Obaidi, N., Patskovsky, Y., Almo, S. C., & Gerlt, J. A. A General Strategy for the Discovery of Metabolic Pathways: d-Threitol, l-Threitol, and Erythritol Utilization in *Mycobacterium smegmatis*. *J. Am. Chem. Soc*. **137**, 14570-14573 (2015).
26. Adams B. L. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. *ACS Synth. Biol*. **5**, 1328-1330 (2016).
27. Reister, M., Hoffmeier, K., Krezdorn, N., Rotter, B., Liang, C., Rund, S., Dandekar, T.,

Sonnenborn, U., & Oelschlaeger, T. A. Complete genome sequence of the gram-negative probiotic *Escherichia coli* strain Nissle 1917. *J. Biotechnol*. **187**, 106-107 (2014).

28. Bathe, S., Achouak, W., Hartmann, A., Heulin, T., Schloter, M., & Lebuhn, M. Genetic and phenotypic microdiversity of *Ochrobactrum spp. FEMS Microbiol. Ecol*. **56**, 272-280 (2006).
29. Chain, P. S., Comerci, D. J., Tolmasky, M. E., Larimer, F. W., Malfatti, S. A., Vergez, L. M., Aguero, F., Land, M. L., Ugalde, R. A., & Garcia, E. Whole-genome analyses of speciation events in pathogenic Brucellae. *Infect. Immun*. **73**, 8353-8361 (2005).
30. Lamontagne, J., Béland, M., Forest, A., Côté-Martin, A., Nassif, N., Tomaki, F., Moriyón, I., Moreno, E., & Paramithiotis, E. Proteomics-based confirmation of protein expression and correction of annotation errors in the *Brucella abortus* genome. *BMC Genomics* **11**, 300 (2010).
31. Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., & Endy, D. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng*. **3**, 4 (2009).
32. Beal, J., Haddock-Angelli, T., Baldwin, G., Gershater, M., Dwijayanti, A., Storch, M., de Mora, K., Lizarazo, M., Rettberg, R., & with the iGEM Interlab Study Contributors. Quantification of bacterial fluorescence using independent calibrants. *PLoS One* **13**, e0199432 (2018).
33. Beal, J. et al. Comparative analysis of three studies measuring fluorescence from engineered bacterial genetic constructs. *PLoS One* **16**, e0252263 (2021).
34. Pellicer, M. T., Fernandez, C., Badía, J., Aguilar, J., Lin, E. C., & Baldom, L. Cross-induction of glc and ace operons of *Escherichia coli* attributable to pathway intersection. Characterization of the glc promoter. *J. Biol. Chem*. **274**, 1745-1752 (1999).
35. Seweryn, P., Van, L. B., Kjeldgaard, M., Russo, C. J., Passmore, L. A., Hove-Jensen, B., Jochimsen, B., & Brodersen, D. E. Structural insights into the bacterial carbon-phosphorus lyase machinery. *Nature* **525**, 68-72 (2015).
36. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**, 3389–3402 (1997).

37. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. NCBI BLAST: a better web interface. *Nucleic Acids Res*. **36**, W5–W9 (2008).
38. Heller, K. B., Lin, E. C., & Wilson, T. H. Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol*. **144**, 274-278 (1980).
39. Sonnenborn U. *Escherichia coli* strain Nissle 1917-from bench to bedside and back: history of a special *Escherichia coli* strain with probiotic properties. *FEMS Microbiol. Lett*. **363**, fnw212 (2016).
40. Pedrolli, D. B., Ribeiro, N. V., Squizato, P. N., de Jesus, V. N., Cozetto, D. A., & Team AQA Unesp at iGEM 2017. Engineering Microbial Living Therapeutics: The Synthetic Biology Toolbox. *Trends Biotechnol*. **37**, 100-115 (2019).
41. Kelly, V. W., Liang, B. K., & Sirk, S. J. Living Therapeutics: The Next Frontier of Precision Medicine. *ACS Synth. Biol*. **9**, 3184-3201 (2020).
42. Cubillos-Ruiz, A., Guo, T., Sokolovska, A., Miller, P. F., Collins, J. J., Lu, T. K., & Lora, J. M. Engineering living therapeutics with synthetic biology. *Nat. Rev. Drug Discov*. **20**, 941-960 (2021).
43. Forker, E. L. Two sites of bile formation as determined by mannitol and erythritol clearance in the guinea pig. *J. Clin. Invest*. **46**, 1189-1195 (1967).
44. Shindou, T., Sasaki, Y., Eguchi, T., Euguchi, T., Hagiwara, K., & Ichikawa, T. Identification of erythritol by HPLC and GC-MS and quantitative measurement in pulps of various fruits. *J. Agric. Food Chem.* **37**, 1474–1476 (1989).
45. Brodkorb, A. et al. INFOGEST static *in vitro* simulation of gastrointestinal food digestion. *Nat. Protoc*. **14**, 991-1014 (2019).
46. Minekus, M. et al. A standardised static *in vitro* digestion method suitable for food - an international consensus. *Food Funct*. **5**, 1113-1124 (2014).

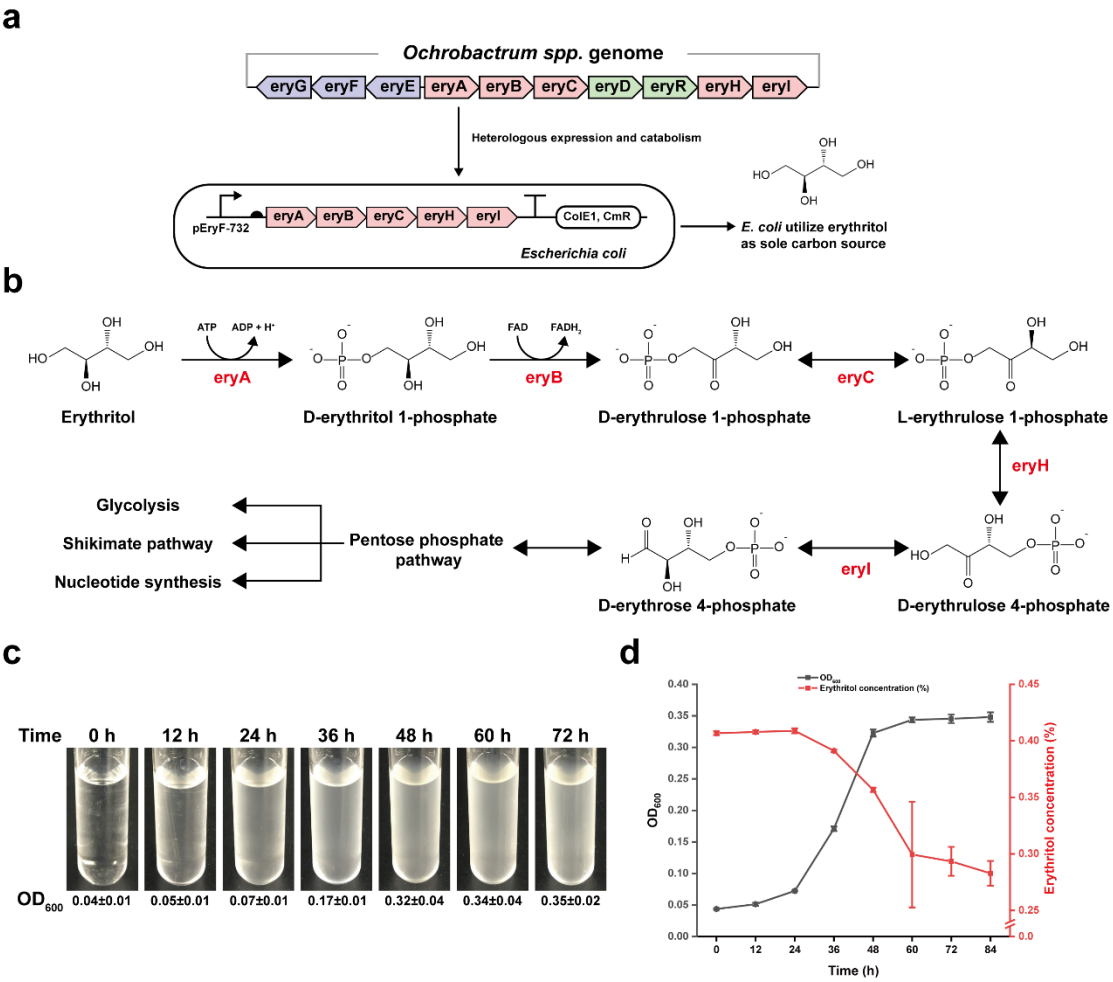


Fig. 1 | Engineering *E. coli* to utilize erythritol as sole carbon source.
(**a**) Screened out *Ochrobactrum spp.* provided erythritol catabolism gene cluster and enable *E. coli* to utilize erythritol as sole carbon source. (**b**) The Erythritol degradation pathway consists of five genes, which catalyze erythritol to D-erythrose 4-phosphate and then flow into pentose phosphate pathway. (**c**) *E. coli* MG1655 harboring plasmid pFB147 could grow in liquid M9-erythritol medium. Each OD_{600} value (average ± standard deviation) was measured with three biological replicates. (**d**) Double Y-axis graph shows *E. coli* growth in M9-erythritol medium and erythritol utilization. Erythritol concentration was measured by HPLC analysis with three biological replicates, the error bar represents the standard deviation (s.d.).

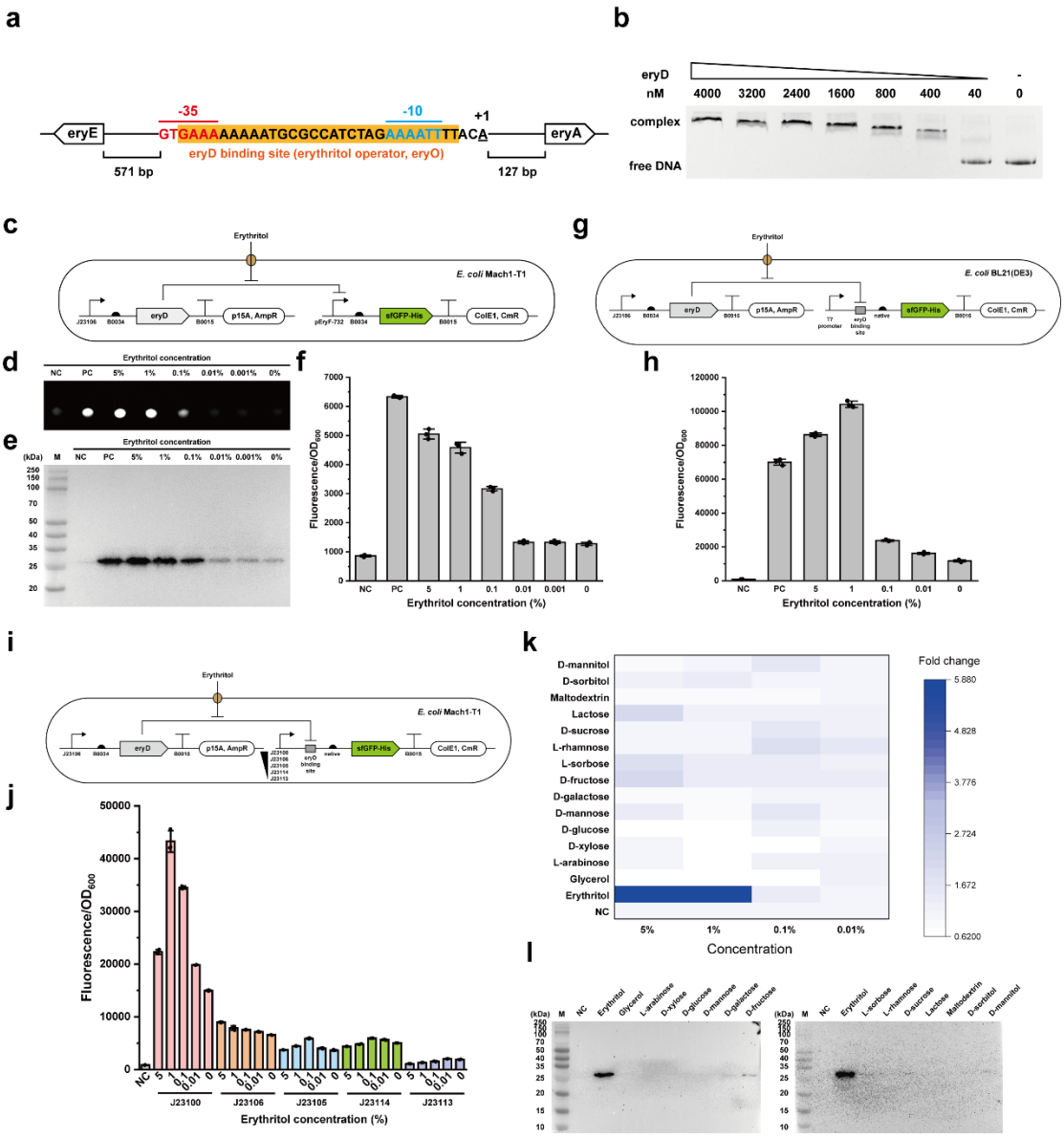


Fig. 2 | erythritol binding transcriptional repressor eryD characterization.

(**a**) DNA sequence of eryD binding site (erythritol operator, eryO, orange highlighted), promoter pEryF-732 sequence (-35, -10, and +1 regions were annotated as red, blue, and black respectively), and their genetic location on erythritol catabolism cluster. (**b**) EMSA analysis of the interaction between eryD and pEryF-732 DNA fragment (5'FAM-labelled). (**c**) Schematic shows two compatible plasmids pFB261 and pFB186 co-existed in *E. coli* Mach1-T1, sfGFP fluorescence indicates the transcription strength of pEryF-732, which may be repressed by eryD. (**d**)-(**f**) Relative sfGFP expression and fluorescence of *E. coli* Mach1-T1 harboring two plasmids in (**c**). Pellets were collected from 1.5 mL overnight LB culture (37°C, 250 rpm for 16 h) with added erythritol, then the pellets were measured as normalized fluorescence with three biological replicates, and Western-Blot analysis (sfGFP with C-terminal 6xHis tag: 27.6 kDa). (**g**) (**h**) Schematic diagram of used strains and plasmids for pT7-eryO operon characterization. pFB276 and pFB261 co-existed in *E. coli* BL21(DE3), pellets were collected from 1.5 mL overnight LB culture (37°C, 250 rpm for 16 h) with added erythritol, then the

pellets were measured as normalized fluorescence with three biological replicates. (**i**) (**j**) Schematic diagram of used strains and plasmids for J231xx-eryO operons characterization. pFB261 and pFB270-pFB275 co-existed in *E. coli* Mach1-T1, pellets were collected from 1.5 mL overnight LB culture (37°C, 250 rpm for 16 h) with added erythritol, then the pellets were measured as normalized fluorescence with three biological replicates. (**k**) (**l**) eryD-erythritol orthogonality test was performed by the same strains and procedure in (**c**). sfGFP expression was analyzed by normalized fluorescence fold change and Western-Blot. All the error bars represent the standard deviation (s.d.).

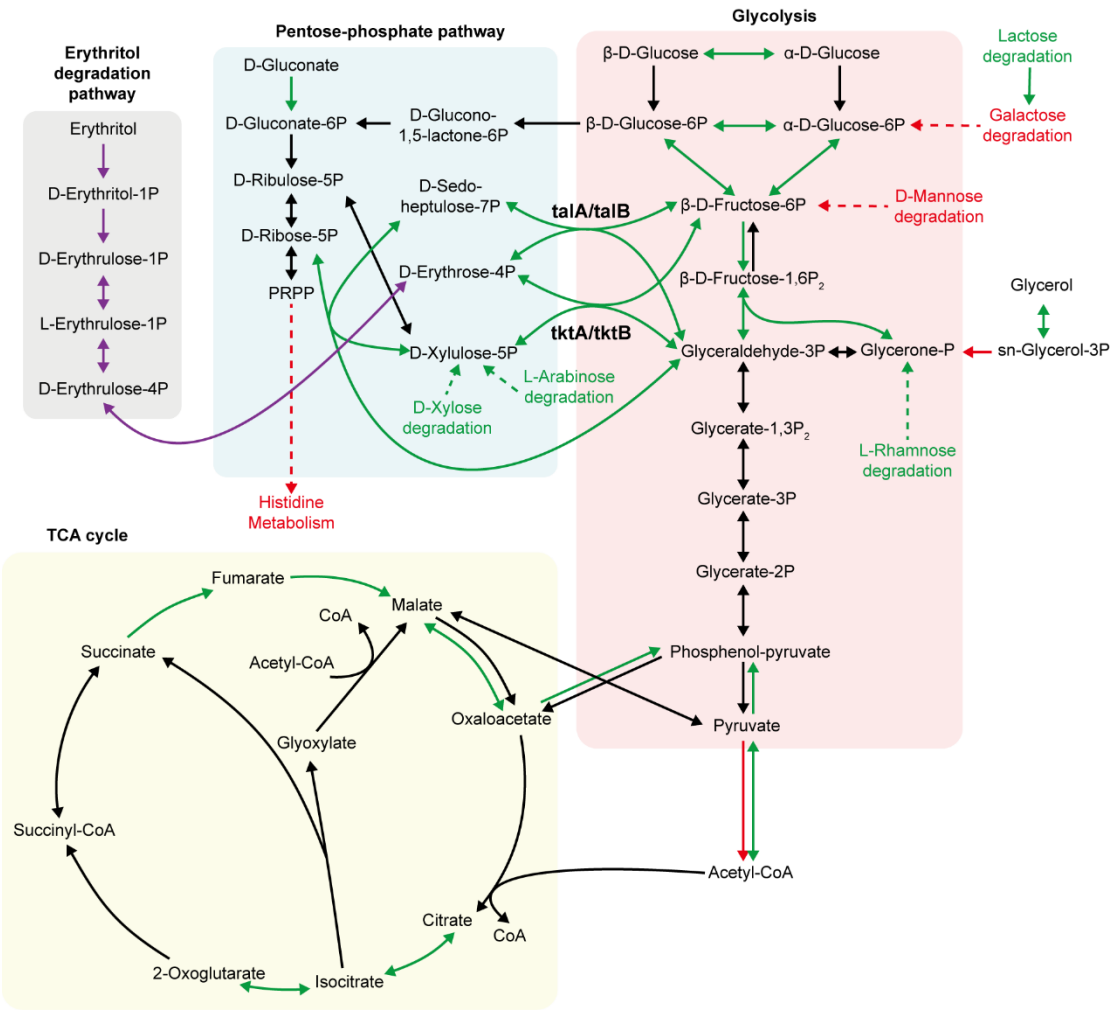


Fig. 3 | Reconstruction of the major carbohydrate metabolic pathways mapped with mRNA transcriptional analysis.

Reconstructed major carbohydrate metabolic pathways of erythritol catabolic *E. coli* MG1655 (harboring pFB147) were based on the relative differential gene expression profiles between two cultures (M9-erythritol as the experimental group, and M9-glucose as the control group). The genes with a significant ($p < 0.05$) differential expression of ≥ 1 log2(fold change) are indicated with arrow (pathway) in green (up-regulated in M9-erythritol medium) or red (down-regulated in M9-erythritol medium), the other genes without significant differential expression are indicated with black arrow (pathway). The details of mRNA transcriptional analysis are shown in the Supplementary Information.

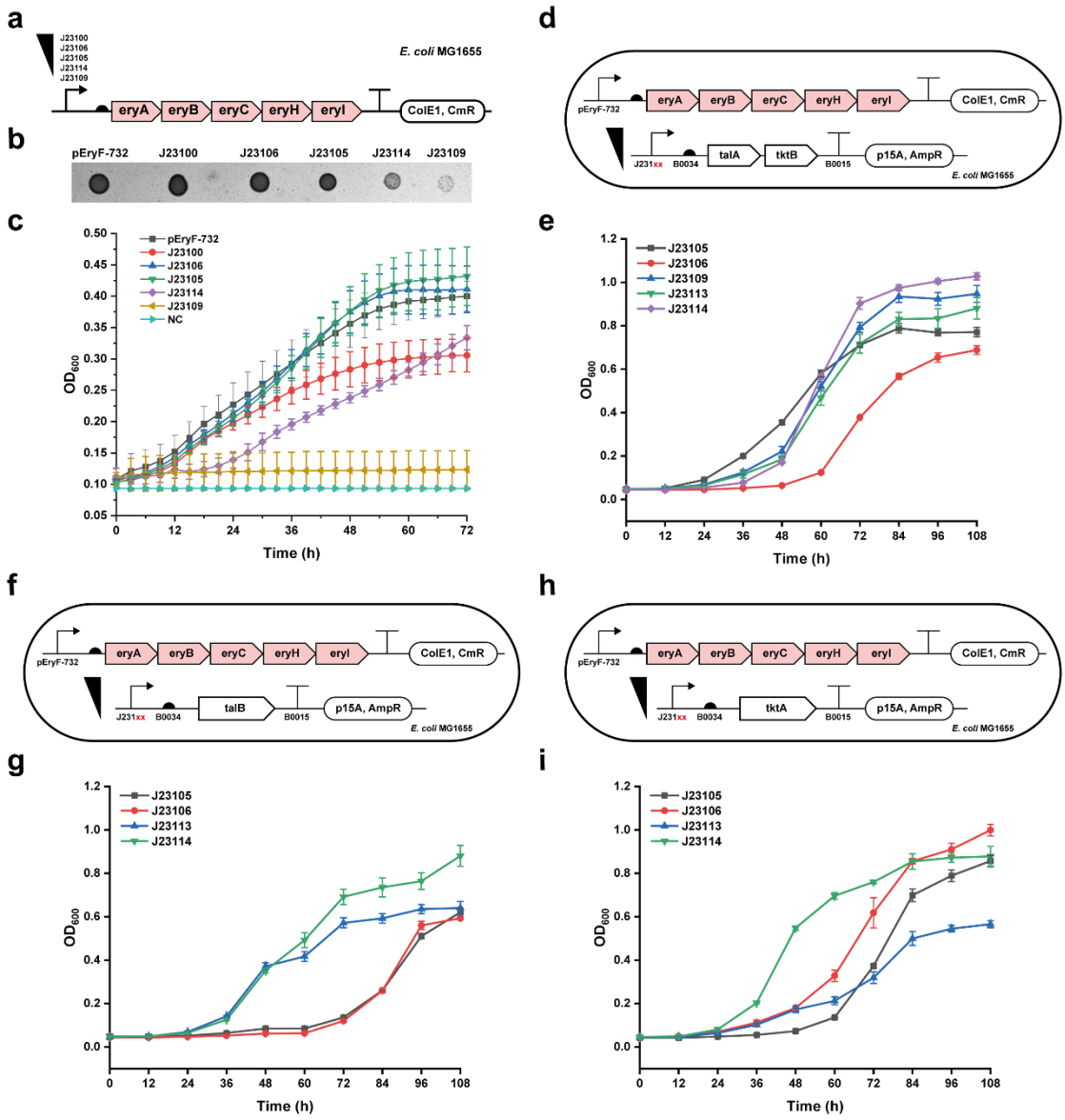


Fig. 4 | Metabolic engineering for improving *E. coli* erythritol catabolism. (**a**) Schematic plasmids of pFB148 to pFB152. Strength-gradient constitutive promoters replaced the native promoter pEryF-732 for erythritol catabolism optimization. (**b**) *E. coli* MG1655 harboring the above-mentioned plasmids grew on M9-erythritol agar plates after 72 h incubation at 37°C. (**c**) Growth curve of *E. coli* MG1655 harboring the above-mentioned plasmids in M9-erythritol liquid mediums (5 mL, 37°C, 250 rpm). (**d**) (**f**) (**h**) *E. coli* MG1655 harboring erythritol degradation plasmid pFB147 and other compatible plasmids. (**d**) (**e**) pFB293-pFB297 co-existed with pFB147 in the strains to constitutively overexpress talA and tktB to explore whether they could accelerate erythritol catabolism. (**f**) (**g**) pFB298-pFB301 co-existed with pFB147 in the strains to constitutively overexpress talB. (**h**) (**i**) pFB302-pFB305 co-existed with pFB147 in the strains to constitutively overexpress tktA. All the measurements were performed with three biological replicates, the error bars represent the standard deviation (s.d.).

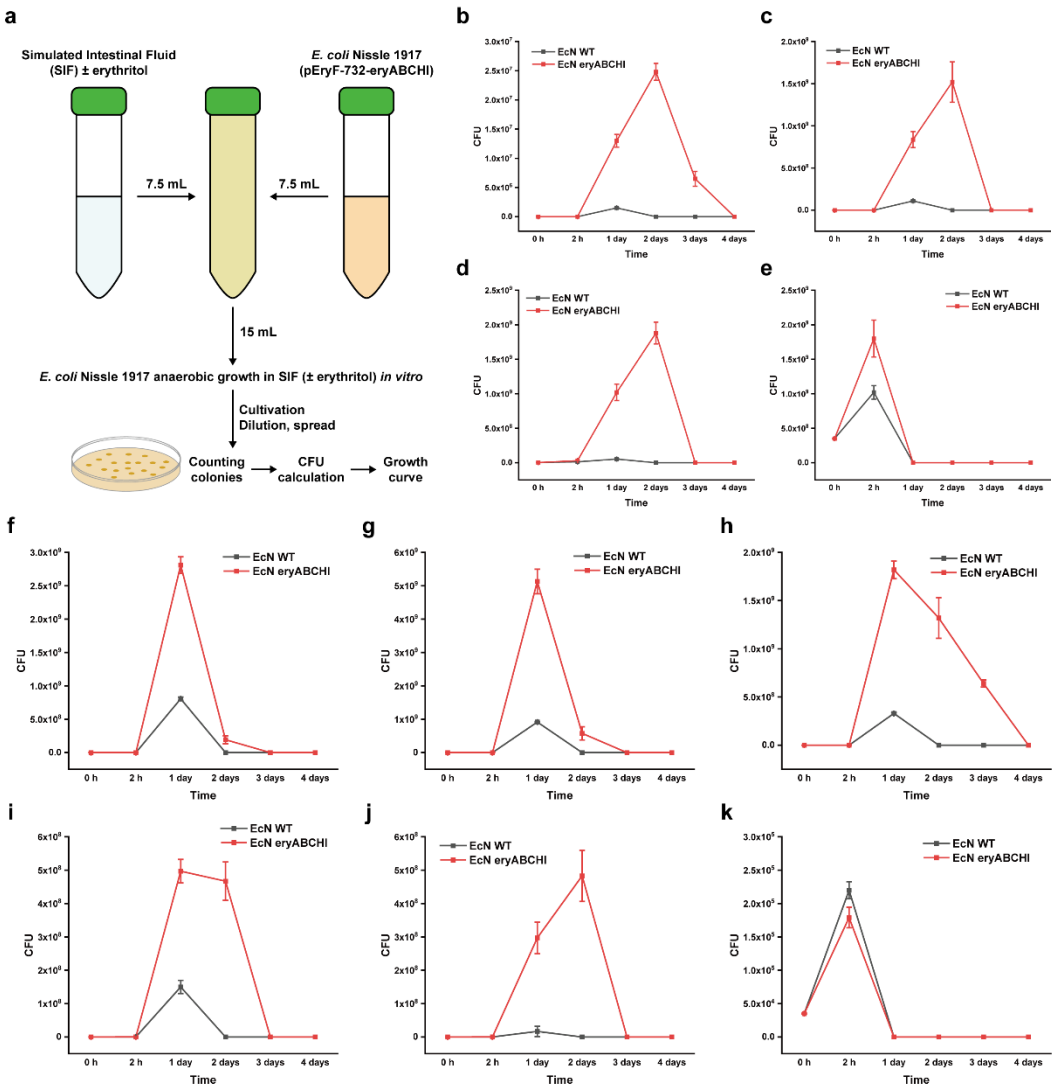


Fig. 5 | Erythritol can facilitate the growth of engineered probiotic *E. coli* Nissle 1917. (**a**) Schematic workflow of sample preparation, incubation, and measurement, the mixture was incubated at 37°C without shaking. (**b**)-(**e**) Growth curves with different inoculated cell densities of 10^2, 10^4, 10^6, and 10^8, respectively, erythritol was added as 0.4% in SIF. (**f**)-(**k**) Growth curves with different erythritol concentrations of 0%, 0.01%, 0.1%, 1%, 4%, and 16% in SIF, respectively, all the initial inoculated cell densities were the same as 10^4. "EcN WT" means wild type *E. coli* Nissle 1917 without plasmids, "EcN eryABCHI" means wild type *E. coli* Nissle 1917 harboring pFB147. All the measurements were performed with three biological replicates, the error bars represent the standard deviation (s.d.).

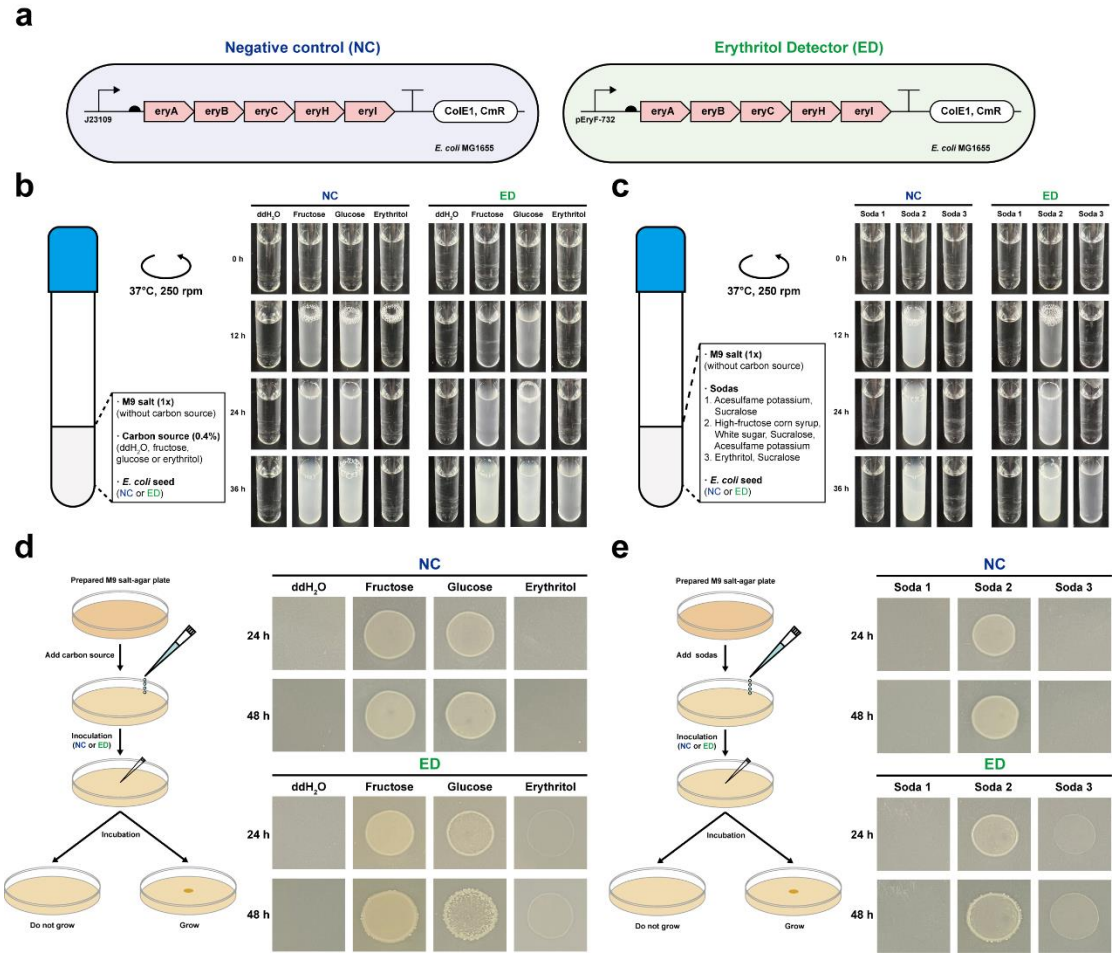
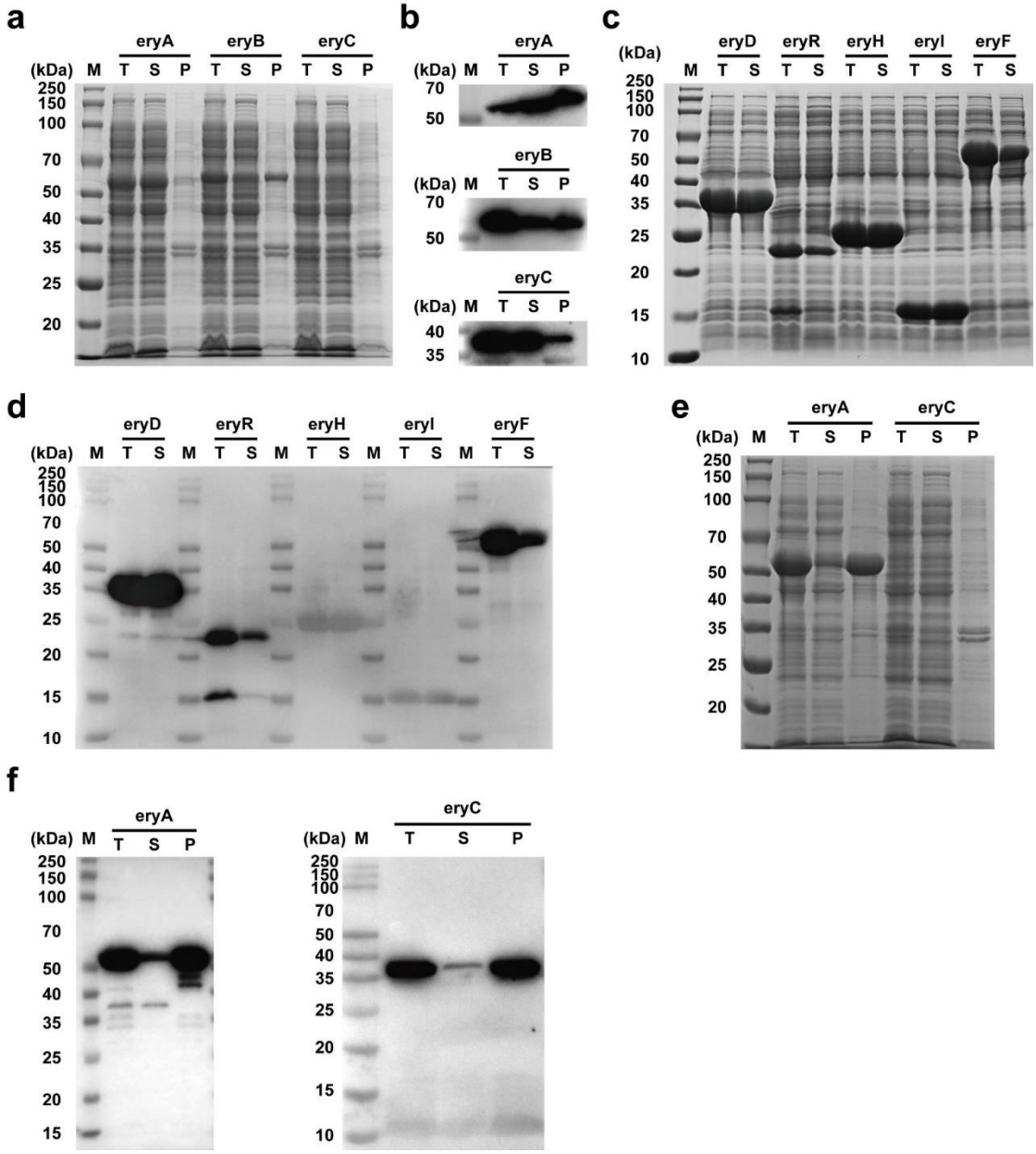


Fig. 6 | Engineered *E. coli* can distinguish soda drinks.

(**a**) Schematic of two *E. coli* MG1655 strains were used as negative control (NC, with pFB151 that cannot utilize erythritol as sole carbon source), and erythritol detector (ED, with pFB147 that can utilize erythritol as sole carbon source). (**b**) (**c**) The liquid incubation strategy was performed to distinguish synthetic M9-based carbohydrate mediums and soda drinks. (**d**) (**e**) The solid incubation strategy was also performed for distinguishing synthetic M9-based carbohydrate mediums (the same as (**b**)) and soda drinks (the same as (**c**)). All samples are repeated by three biological replicates with similar results.



Extended Data Fig. 1 | erythritol catabolism-associated genes *in vivo* expression. (**a**) SDS-PAGE analysis of gene expression of eryA (with N-terminal 6xHis tag, 56.2 kDa), eryB (with N-terminal 6xHis tag, 57.2 kDa), and eryC (with N-terminal 6xHis tag, 35.8 kDa). All these three genes were expressed by plasmids pFB158, pFB159, and pFB160 respectively. *E. coli* Mach1-T1 with plasmids were initially inoculated with overnight culture (5 mL LB medium at 37°C, 250 rpm for 16 h) as a volume ratio of 1:500, and then incubated in 5 mL new LB medium at 37°C, 250 rpm, when OD_{600} reached 0.6, 1% (w/v) arabinose was added and incubated for another 16 h at 20°C, 250rpm. Then the cell pellets were resuspended (equal volume) with 1x phosphate-Buffered Saline (pH 7.4) and lysed by sonication, the total cell lysate "T", supernatant "S" and pellets "P" were analyzed by SDS-PAGE. The result shows that eryA and eryB could be well expressed with acceptable percentage of soluble parts, however, eryC protein bands could not be observed on the gel. It is worth noting that we initially tried to construct the IPTG-inducted pET22b-eryA/eryB/eryC plasmids, however,

we could not get the correct colonies, the possible reason was that eryA/eryB/eryC leakage might cause cell toxicity[7], and IPTG-LacI inducible systems always lead to unexpected protein expression leakage. Thus, we constructed the arabinose-araC inducible systems that weaken leakage and finally, we gained the correct plasmids (pFB158, pFB159, and pFB160).

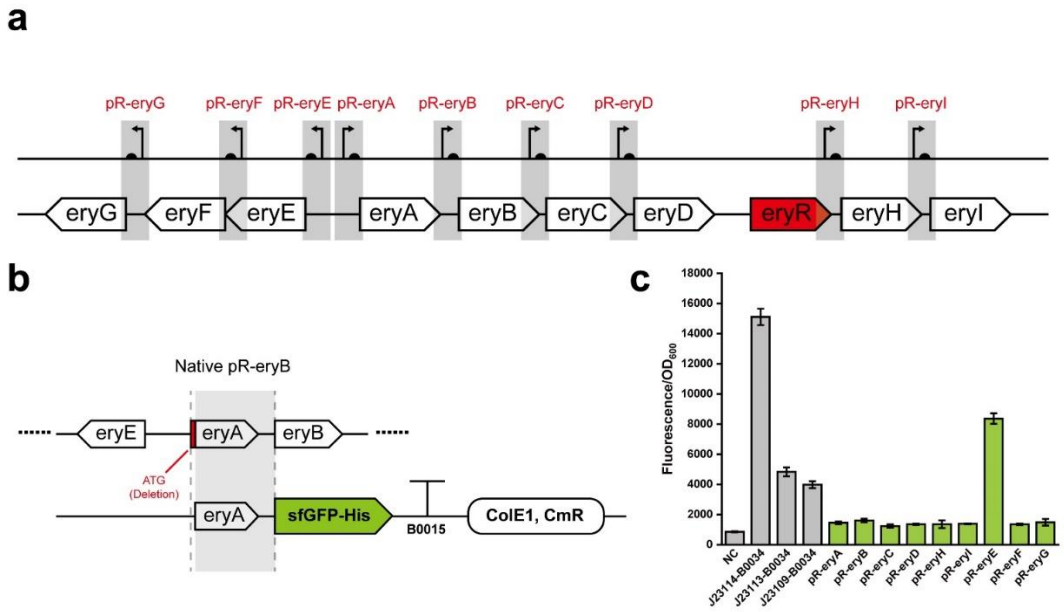
(**b**) Western-Blot analysis of the same samples in (**a**), the result indicated that eryC could be expressed with acceptable percentage of soluble parts, however, the total expression amount is not enough to be directly observed on SDS-PAGE gel.

(**c**) SDS-PAGE analysis of gene expression of eryD (with N-terminal 6xHis tag, 34.5 kDa), eryR (with N-terminal 6xHis tag, 24.8 kDa), eryH (with N-terminal 6xHis tag, 28.8 kDa), eryI (with N-terminal 6xHis tag, 16.6 kDa), eryF (with N-terminal 6xHis tag, 56.2 kDa). All these five genes were expressed by plasmids pFB161, pFB167, pFB165, pFB166, and pFB163 respectively. *E. coli* BL21(DE3) with plasmids were initially inoculated with overnight culture (5 mL LB medium at 37°C, 250 rpm for 16 h) as a volume ratio of 1:500, and then incubated in 5 mL new LB medium at 37°C, 250 rpm, when OD_{600} reached 0.6, 0.5 mM IPTG was added and incubated for another 16 h at 20°C, 250rpm. Then the cell pellets were resuspended (equal volume) with 1x phosphate-Buffered Saline (pH 7.4) and lysed by sonication, the total cell lysate “T”, supernatant “S” and pellets “P” were analyzed by SDS-PAGE. The result shows that all of the five proteins could be well expressed with acceptable percentage of soluble parts and well observed on the gel.

(**d**) Western-Blot analysis of the same samples in (**c**).

(**e**) SDS-PAGE analysis of gene expression of eryA and eryC. All the procedures were the same as (**a**) except the arabinose-associated incubation: after 1% (w/v) arabinose addition, the medium was incubated for another 4 h at 37°C, 250rpm rather than 16 h at 20°C, 250rpm.

(**f**) Western-Blot analysis of the same samples in (**e**). All the results of (**e**) and (**f**) indicated that in this induction process, both eryA and eryC could be expressed with nearly all insoluble parts. Moreover, eryC could not be expressed a large amount in either high or low temperatures.

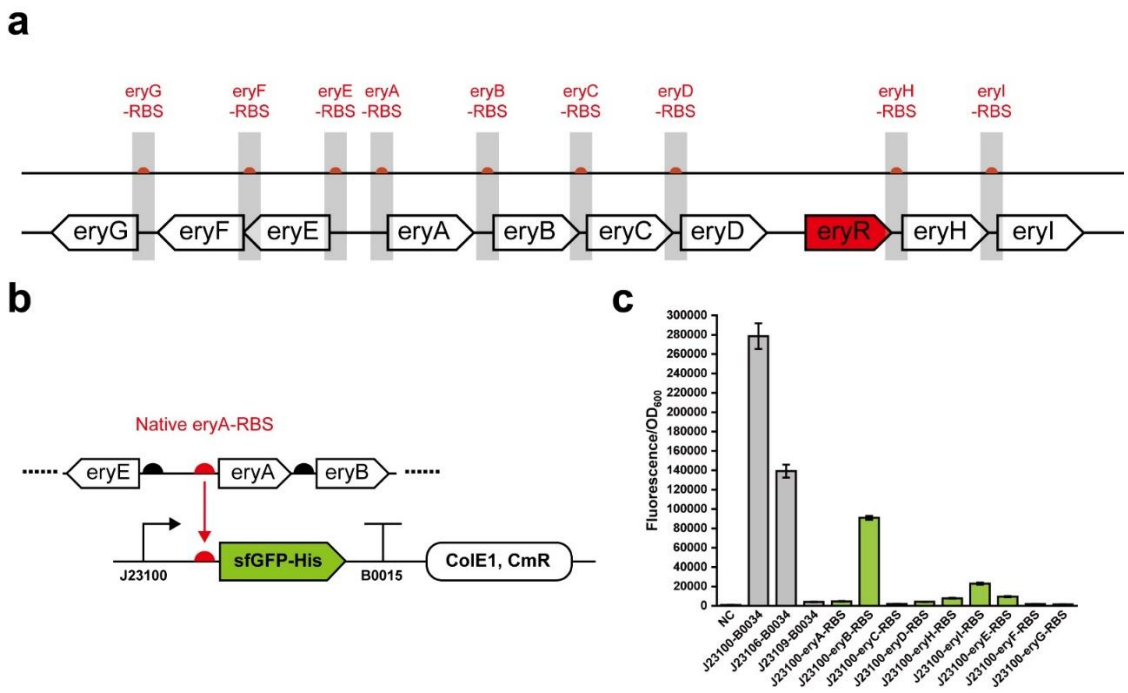


Extended Data Fig. 2 | Native promoter-RBS (pR-) composite part characterization in original erythritol cluster.

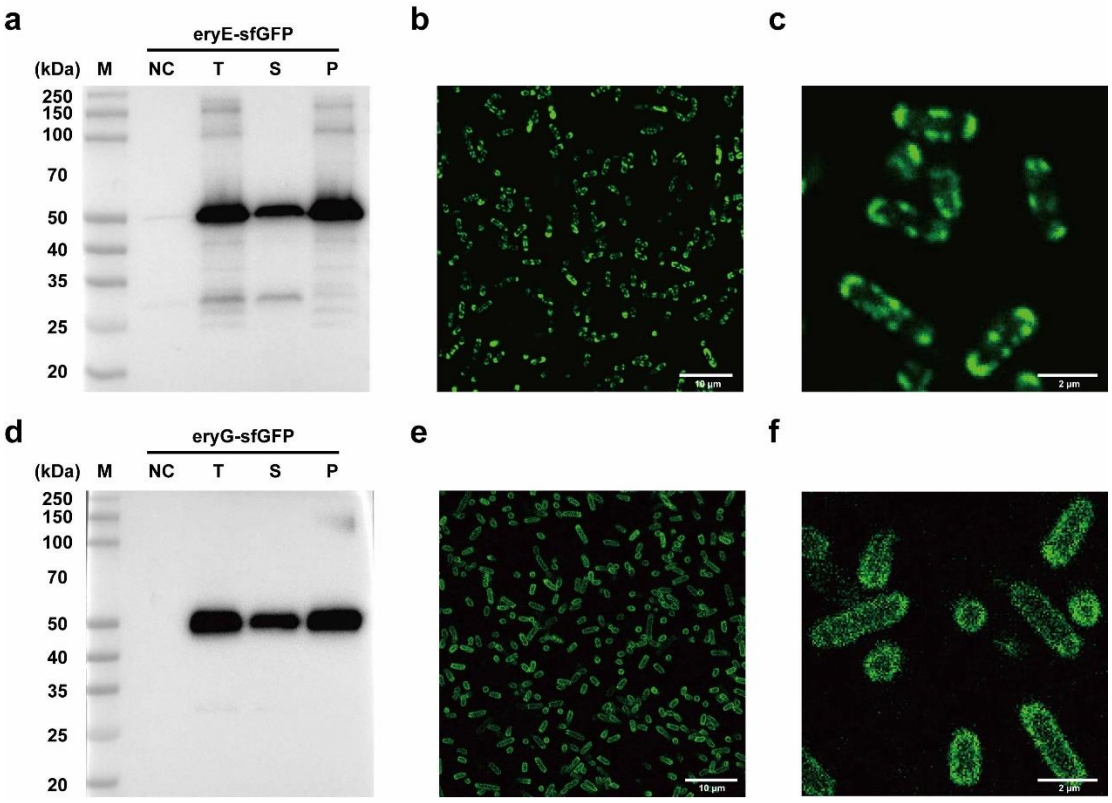
(**a**) Schematic promoter-RBS (pR-) composite part localization. It is worth mentioning that we did not successfully characterize the function of eryR, thus here we did not test pR-eryR.

(**b**) Schematic procedure of "pR-ery" reporter plasmids construction (pFB177 to pFB185). Here we illustrate the construction of pR-eryB-sfGFP reporter plasmid as an example. The DNA sequence between "eryA CDS except initiation codon ATG" and "the last nucleotide before eryB CDS" was considered as pR-eryB. This part was then assembled with the following sfGFP (6xHis) and terminator B0015 in vector pSB1C3 (pFB178).

(**c**) Normalized fluorescence result of all nine "pR-ery" parts. All measurements were performed in *E. coli* Mach1-T1 with three biological replicates. "NC" represents *E. coli* Mach1-T1 without plasmids. The other three gray values correspond to pFB291, pFB290, and pFB289 respectively to act as the positive control (iGEM standard biological parts). The results indicate that all nine composite parts were not strong enough.



Extended Data Fig. 3 | Native RBS characterization in original erythritol cluster.
(**a**) Schematic RBS localization. It is worth mentioning that we did not successfully characterize the function of eryR, thus here we did not test eryR-RBS.
(**b**) Schematic procedure of "ery-RBS" reporter plasmids construction (pFB168 to pFB176). Here we illustrate the construction of J23100-eryA-RBS-sfGFP reporter plasmid as an example. The proposed "eryA-RBS" sequence (approximately 40 bp) was assembled with promoter J23100, sfGFP (6xHis), and terminator B0015 in vector pSB1C3 (pFB168).
(**c**) Normalized fluorescence result of all nine RBS parts. All measurements were performed in *E. coli* Mach1-T1 with three biological replicates. "NC" represents *E. coli* Mach1-T1 without plasmids. The other three gray values correspond to pFB286, pFB288, and pFB289 respectively to act as the positive control (iGEM standard biological parts). The results indicate that "eryB-RBS" was a lit strong compared with strong ribosome binding site B0034, but the other eight were not strong enough.



Extended Data Fig. 4 | eryE-sfGFP and eryG-sfGFP characterization.

(**a**) Western-Blot analysis of eryE-sfGFP(6xHis) (50.4 kDa). *E. coli* Mach1-T1 with plasmid pFB162 were initially inoculated with overnight culture (5 mL LB medium at 37°C, 250 rpm for 16 h) as a volume ratio of 1:500, and then incubated in 5 mL new LB medium at 37°C, 250 rpm, when OD_{600} reached 0.6, 1% (w/v) arabinose was added and incubated for another 6 h at 30°C, 250rpm. Then the cell pellets were resuspended (equal volume) with 1x phosphate-Buffered Saline (pH 7.4) and lysed by sonication, the total cell lysate "T", supernatant "S" and pellets "P" were analyzed by Western-Blot. "NC" represents the total cell lysate of *E. coli* Mach1-T1 with plasmid pFB162 and without arabinose induction.

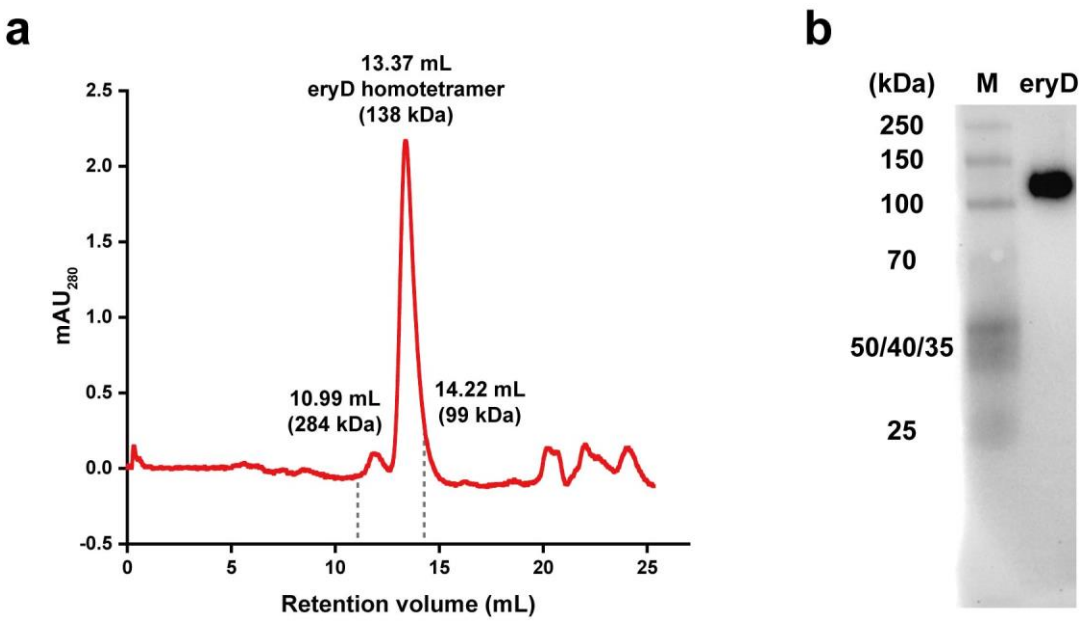
(**b**) Confocal laser scanning microscope image of *E. coli* Mach1-T1 that express eryE-sfGFP, green border indicates the transmembrane localization of eryE-sfGFP.

(**c**) Partially enlarged view of the image in (**b**).

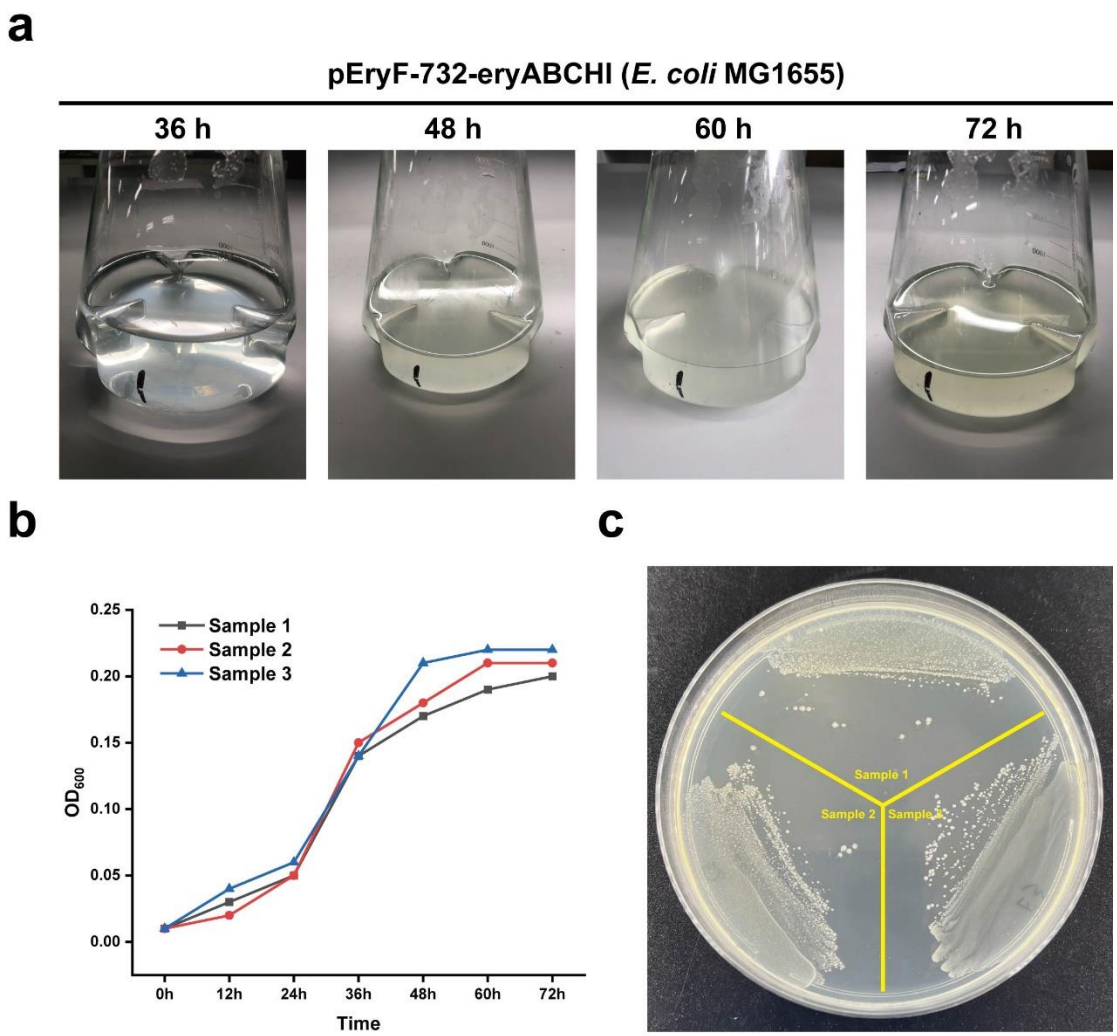
(**d**) Western-Blot analysis of eryG-sfGFP(6xHis) (64.6 kDa). *E. coli* Mach1-T1 with plasmid pFB164 were initially incubated in 5 mL LB medium at 37°C, 250rpm, when OD_{600} reached 0.6, 1% (w/v) arabinose was added and incubated for another 6 h at 30°C, 250rpm. Then the cell pellets were resuspended (equal volume) with 1x phosphate-buffered saline (pH 7.4) and lysed by sonication, the total cell lysate "T", supernatant "S" and pellets "P" were analyzed by Western-Blot. "NC" represents the total cell lysate of *E. coli* Mach1-T1 with plasmid pFB164 and without arabinose induction.

(**e**) Confocal laser scanning microscope image of *E. coli* Mach1-T1 that express eryG-sfGFP, green border indicates the transmembrane localization of eryG-sfGFP.

(**f**) Partially enlarged view of the image in (**e**).



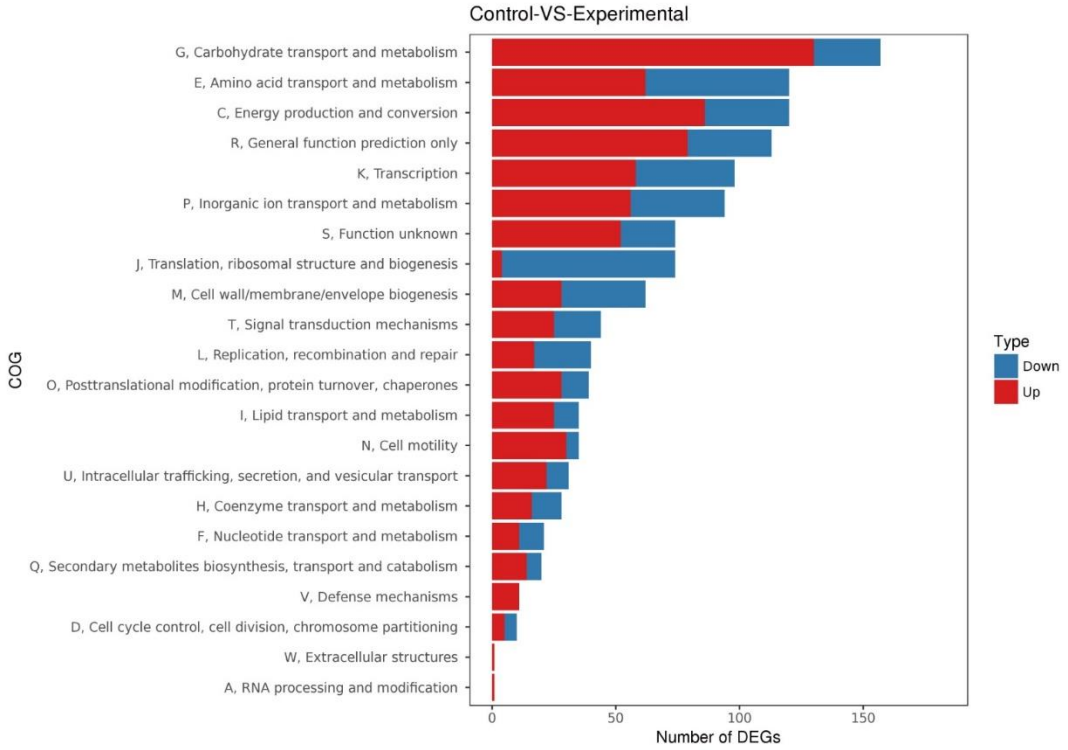
Extended Data Fig. 5 | eryD assembled as homotetramer.
(**a**) Gel-filtration chromatography of the purified eryD(6xHis) with Superdex 200 increase 10/300 GL column. eryD homotetramer was eluted as 13.37 mL (red peak), the other two purified enzymes in our group were used as standard: Vlm2 (284 kDa) was eluted at 10.99 mL, and T7 RNA polymerase (99 kDa) was eluted at 14.22 mL.
(**b**) Native-PAGE gel electrophoresis-based Western-Blot analysis of the purified eryD(6xHis). the expected homotetramer band (approximately 138 kDa) was shown between the two markers of 150 kDa and 100 kDa.



Extended Data Fig. 6 | Samples preparation for mRNA transcriptional analysis.
Wild type *E. coli* MG1655 strain with plasmid pFB147 was incubated in 5 mL LB medium (with 34 μg / mL chloramphenicol) for overnight culture (37°C, 250 rpm for 16 h), on the second day, 5 mL cell culture was washed with 1x phosphate-Buffered Saline (pH 7.4) by three times and then resuspended by 5 mL M9-erythritol liquid medium, then the resuspended cell mixture was inoculated into 1 L M9-erythritol liquid medium (with 6 μg / mL chloramphenicol) as 1:200 v/v in 2 L shaking flask. After that, the 1 L cell mixture was incubated for 72 h (37°C, 250 rpm), such process was repeated by three biological replicates. After these pre-experiments, we found that at 36 h, the cell reached the middle logarithmic phase (see panel **b**). Afterward, we again processed the same experiments (as three biological replicates) and collect the cell pellets at 36 h as three parallel RNA-seq samples (experimental group samples).
(**a**) The cell cultures in pre-experiments reached stationary phase at 72 h.
(**b**) Cell growth curve (OD_{600}-time) of pre-experiment samples. In such a situation, the middle logarithmic phase was reached at 36 h and the stationary phase was reached at 72 h.

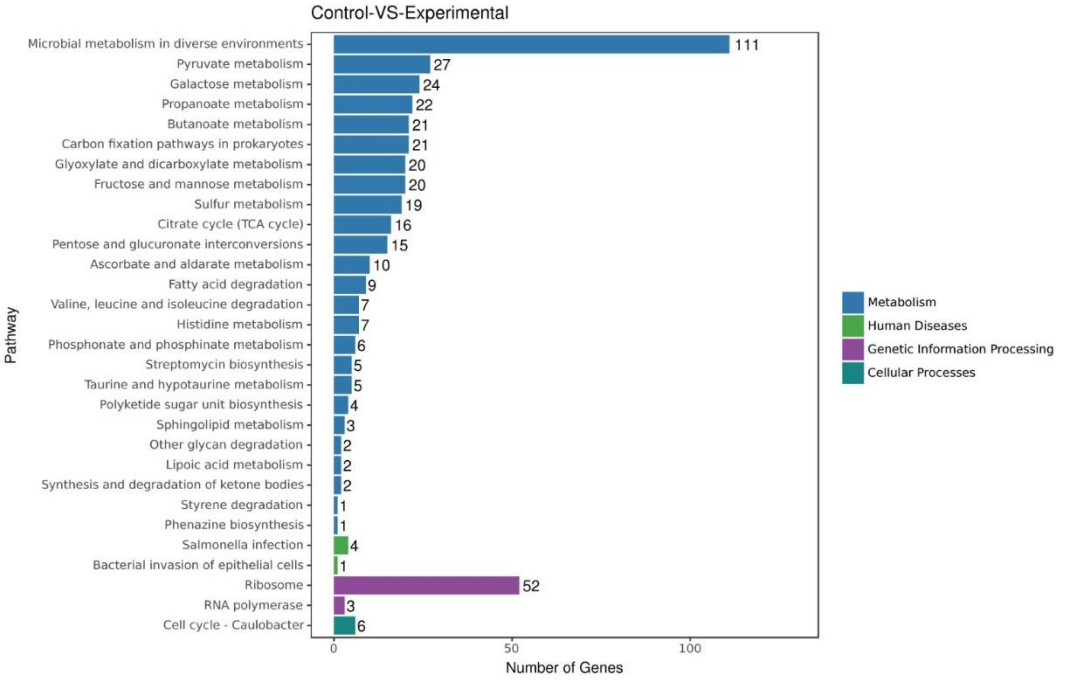
(**c**) The cell cultures (samples in pre-experiment, at 72 h) were picked and separated on LB-agar plates (with 34 μg / mL chloramphenicol) to test the cell activity (overnight culture at 37°C for 16 h). The colonies were considered normal and did not lose plasmid pFB147.

a

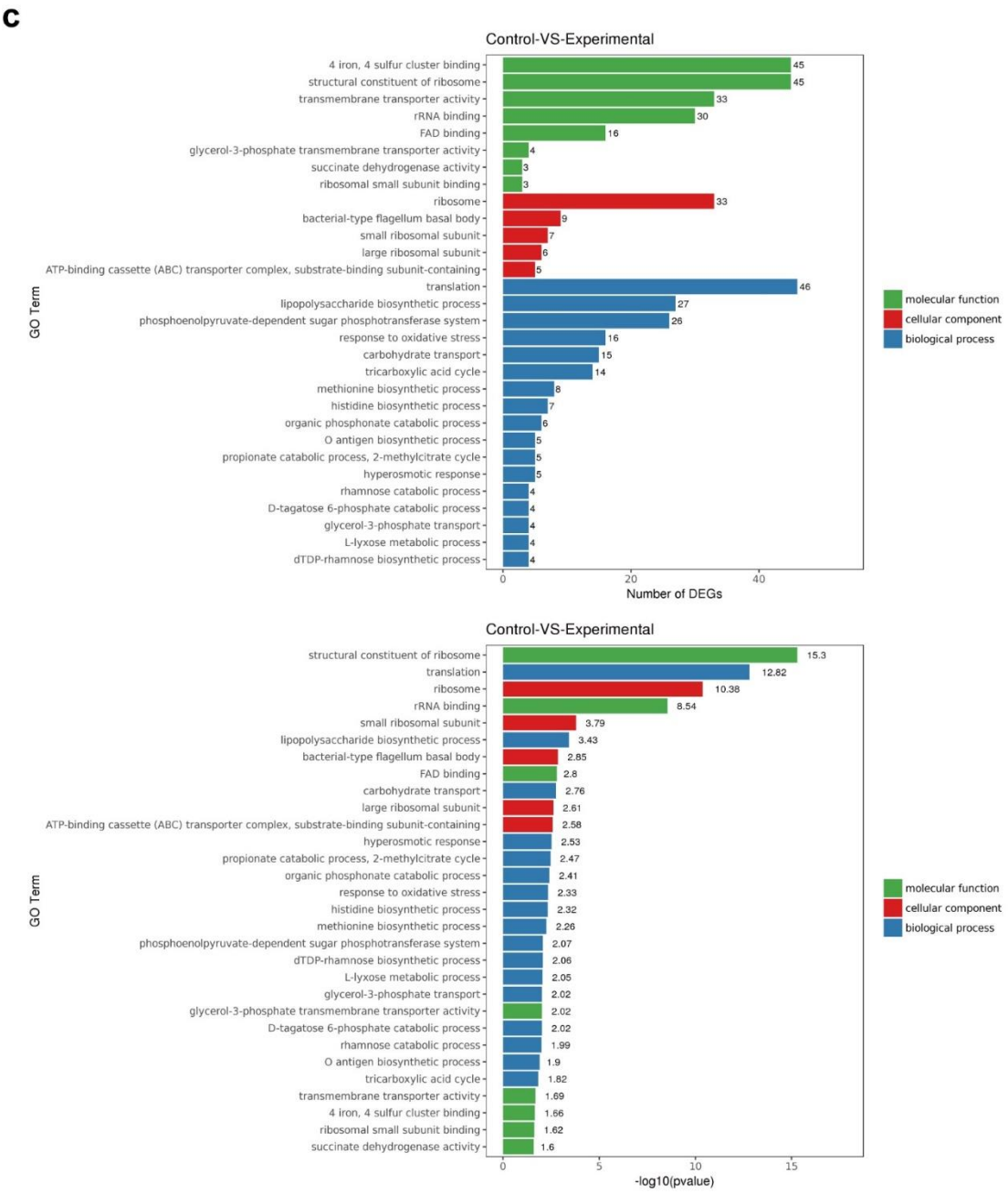

Control-VS-Experimental
G, Carbohydrate transport and metabolism
E, Amino acid transport and metabolism
C, Energy production and conversion
R, General function prediction only
K, Transcription
P, Inorganic ion transport and metabolism
S, Function unknown
J, Translation, ribosomal structure and biogenesis
M, Cell wall/membrane/envelope biogenesis
T, Signal transduction mechanisms
L, Replication, recombination and repair
O, Posttranslational modification, protein turnover, chaperones
I, Lipid transport and metabolism
N, Cell motility
U, Intracellular trafficking, secretion, and vesicular transport
H, Coenzyme transport and metabolism
F, Nucleotide transport and metabolism
Q, Secondary metabolites biosynthesis, transport and catabolism
V, Defense mechanisms
D, Cell cycle control, cell division, chromosome partitioning
W, Extracellular structures
A, RNA processing and modification
COG
0
50
100
150
Number of DEGs
Type
Down
Up

1209

b

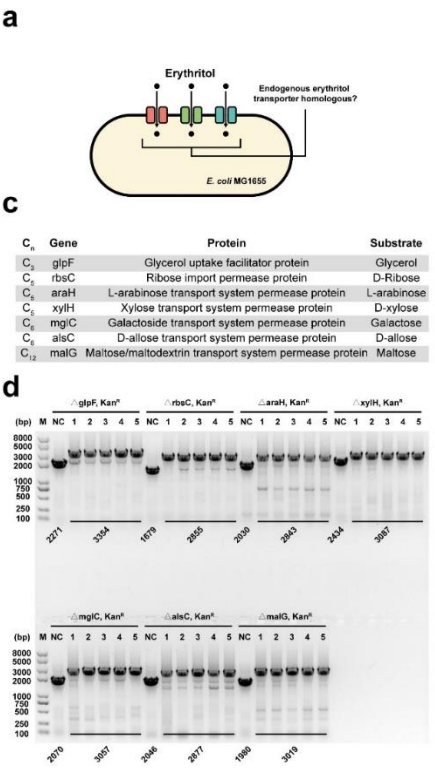

Control-VS-Experimental
Microbial metabolism in diverse environments 111
Pyruvate metabolism 27
Galactose metabolism 24
Propanoate metabolism 22
Butanoate metabolism 21
Carbon fixation pathways in prokaryotes 21
Glyoxylate and dicarboxylate metabolism 20
Fructose and mannose metabolism 20
Sulfur metabolism 19
Citrate cycle (TCA cycle) 16
Pentose and glucuronate interconversions 15
Ascorbate and aldarate metabolism 10
Fatty acid degradation 9
Valine, leucine and isoleucine degradation 7
Histidine metabolism 7
Phosphonate and phosphinate metabolism 6
Streptomycin biosynthesis 5
Taurine and hypotaurine metabolism 5
Polyketide sugar unit biosynthesis 4
Sphingolipid metabolism 3
Other glycan degradation 2
Lipoic acid metabolism 2
Synthesis and degradation of ketone bodies 2
Styrene degradation 1
Phenazine biosynthesis 1
Salmonella infection 4
Bacterial invasion of epithelial cells 1
Ribosome 52
RNA polymerase 3
Cell cycle - Caulobacter 6
Pathway
0
50
100
Number of Genes
Metabolism
Human Diseases
Genetic Information Processing
Cellular Processes

1210



Extended Data Fig. 7 | mRNA transcriptional analysis

Transcriptome results were analyzed via (**a**) COG, (**b**) KEGG, and (**c**) GO databases. The common different genes were associated with carbohydrate metabolism, amino acid metabolism, and ribosome. In detail, carbohydrate metabolism was commonly up-regulated, however, ribosome and translation genes were down-regulated, which indicated that erythritol-utilized *E. coli* tended to uptake more easily utilized carbohydrates, also suppressed ribosome assembly and protein translation process to resist such a poor carbohydrate environment.



b

C_n	Gene	Protein	Substrate
C_3	glpF	Glycerol uptake facilitator protein	Glycerol
C_5	rbsA	Ribose import ATP-binding protein	D-Ribose
	rbsB	Ribose import binding protein	
	rbsC	Ribose import permease protein	
C_5	araF	L-arabinose-binding periplasmic protein	L-arabinose
	araG	Arabinose import ATP-binding protein	
	araH	L-arabinose transport system permease protein	
C_5	araE	Arabinose-proton symporter	L-arabinose
C_5	xylF	D-xylose-binding periplasmic protein	D-xylose
	xylG	Xylose import ATP-binding protein	
	xylH	Xylose transport system permease protein	
C_5	xylE	D-xylose-proton symporter	D-xylose
C_5/C_6	rhaT	L-rhamnose-proton symporter	Lyxose/L-rhamnose/L-xylose/mannose
C_6	crr	PTS system glucose-specific EIIA component	D-glucose
	ptsG	PTS system glucose-specific EIICB component	
	ptsH	Phosphocarrier protein	
	ptsI	Phosphoenolpyruvate-protein phosphotransferase	
C_6	fruA	PTS system fructose-specific EIIB'BC component	D-frucose
	fruB	Multiphosphoryl transfer protein	
C_6	galP	Galactose-proton symporter	Galactose
C_6	mglA	Galactose/methyl galactoside import ATP-binding protein	Galactose
	mglB	D-galactose-binding periplasmic protein	
	mglC	Galactoside transport system permease protein	
C_6	manX	PTS system mannose-specific EIIAB component	D-Mannose
	manY	PTS system mannose-specific EIIC component	
	manZ	PTS system mannose-specific EIID component	
C_6	fucP	L-fucose-proton symporter	L-fucose
C_6	alsA	D-allose import ATP-binding protein	D-allose
	alsB	D-allose-binding periplasmic protein	
	alsC	D-allose transport system permease protein	
C_{12}	lacY	Lactose permease	Lactose
C_{12}	malE	Maltose/maltodextrin-binding periplasmic protein	Maltose
	malF	Maltose/maltodextrin transport system permease protein	
	malG	Maltose/maltodextrin transport system permease protein	
	malK	Maltose/maltodextrin import ATP-binding protein	
C_{12}	malX	PTS system maltose-specific EIICB component	Maltose
C_{12}	treB	PTS system trehalose-specific EIIBC component	Trehalose
C_{12}	melB	Melibiose carrier protein	Melibiose
C_6 (polyol)	srlA	PTS system glucitol/sorbitol-specific EIIC component	Sorbitol
	srlB	PTS system glucitol/sorbitol-specific EIIA component	
	srlE	PTS system glucitol/sorbitol-specific EIIB component	
C_6 (polyol)	mtlA	PTS system mannitol-specific EIICBA component	Mannitol
	cmtA	PTS system mannitol-specific cryptic EIICB component	
	cmtB	Mannitol-specific cryptic phosphotransferase enzyme IIA component	

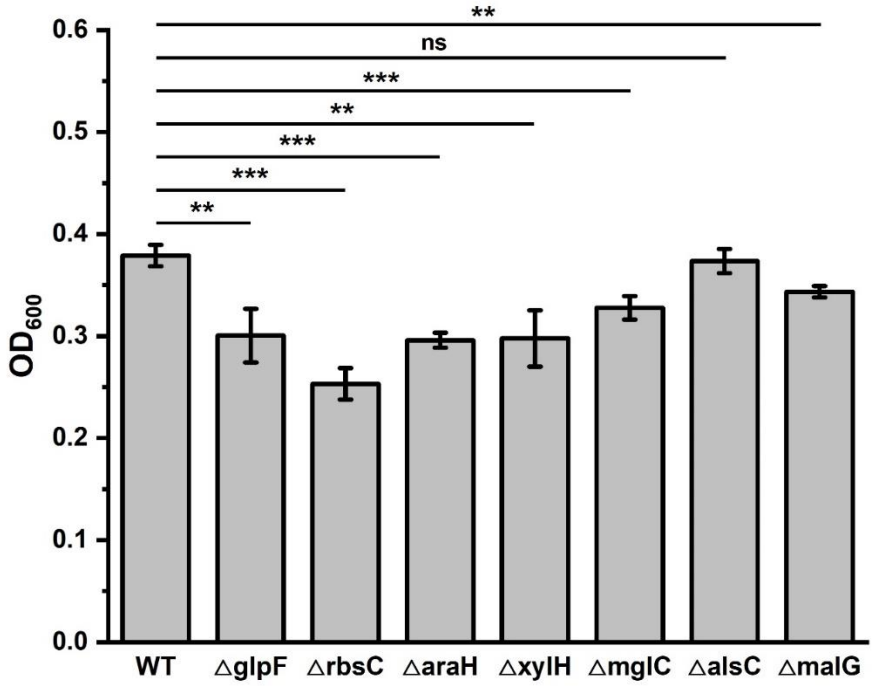
Extended Data Fig. 8 | Potential endogenous erythritol transporter homologous screening.

(**a**) Schematic diagram of screening *E. coli* MG1655 native transporter which could facilitate erythritol transportation.

(**b**) List of all carbohydrate transportation genes and proteins in *E. coli* MG1655.

(**c**) Screened seven carbohydrate transporter-associated genes. glpF has been reported that is able to transport erythritol. The other six genes are somewhat homologous to erythritol ABC transporter gene eryG by BLAST protein-protein alignment analysis.

(**d**) Seven genes' mutated *E. coli* MG1655 strains construction by Lambda-Red recombination. For each of the seven mutation strains, "NC" means the wild type *E. coli* MG1655 genome PCR product by primer pair "forward left-homologous-arm" and "reverse right-homologous-arm", and the other five PCR products were the desired mutation strains by the same primer pair. The KanR genes were not deleted by pCP20.



Extended Data Fig. 9 | Several eryG homologous permeases in *E. coli* may facilitate erythritol catabolism.

BLAST protein-protein alignment results showed that there were seven *E. coli* native carbohydrate ABC-transporter permeases that were homologous to erythritol permease eryG, we constructed 7 different knock-out *E. coli* MG1655 strains by Lambda-Red recombination. All these 7 strains and WT (wild type *E. coli* MG1655) strain were then transformed with pFB147 to catalyze erythritol. After that, all strains were incubated (M9-erythritol medium, 37°C, no shaking) in 96-well plates for 72 h (stationary phase). OD_{600} values were measured each 12 h. This diagram showed the OD_{600} values at 72 h. All values were measured as three biological replicates, the error bar represents the standard deviation (s.d.). Student's t-tests are used for statistical analysis, and $p < 0.05$ indicates statistical significance (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). In conclusion, glpF, rbsC, araH, xylH, mglC, and malG knock-out strains suppressed their growth rate, in another word, these transport systems could facilitate erythritol transport.