

1 A single nucleotide change in the *polC*
2 DNA polymerase III in *Clostridium*
3 *thermocellum* is sufficient to create a
4 hypermutator phenotype

5 Anthony Lanahan^{1,2}, Kamila Zakowicz¹, Liang Tian^{1,2}, Daniel G. Olson^{1,2*} and Lee R.
6 Lynd^{1,2}

7 ¹ Thayer School of Engineering at Dartmouth College, Hanover, NH 03755

8 ² Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN,
9 37830

10 *To whom correspondence should be addressed: daniel.g.olson@dartmouth.edu

11 Keywords

12 Whole genome sequencing, next-generation sequencing, 5-fluoroorotic acid, 5-FOA,
13 mutation rate, *DNA polymerase III*, *polC*, *dnaE*, *Clostridium thermocellum*,
14 *Hungateiclostridium thermocellum*, *Ruminiclostridium thermocellum*, *Acetivibrio*
15 *thermocellus*

16 Abstract

17 *Clostridium thermocellum* is a thermophilic, anaerobic, bacterium that natively ferments
18 cellulose to ethanol, and is a candidate for cellulosic biofuel production. Recently, we
19 identified a hypermutator strain of *C. thermocellum* with a C669Y mutation in the *polC*
20 gene, which encodes a DNA polymerase III enzyme. Here we reintroduce this mutation
21 using recently-developed CRISPR tools to demonstrate that this mutation is sufficient to
22 recreate the hypermutator phenotype. The resulting strain shows an approximately 30-
23 fold increase in the mutation rate. This mutation is hypothesized to function by
24 interfering with metal ion coordination in the PHP domain responsible for proofreading.
25 The ability to selectively increase the mutation rate in *C. thermocellum* is a useful tool
26 for future directed evolution experiments.

27 Importance

28 Cellulosic biofuels are a promising approach to decarbonize the heavy duty
29 transportation sector. A longstanding barrier to cost-effective cellulosic biofuel
30 production is the recalcitrance of the material to solubilization. Native cellulose-
31 consuming organisms, such as *Clostridium thermocellum*, are promising candidates for
32 cellulosic biofuel production, however they often need to be genetically modified to
33 improve product formation. One approach is adaptive laboratory evolution. Our findings
34 demonstrate a way to increase the mutation rate in this industrially-relevant organism,
35 which can reduce the time needed for adaptive evolution experiments.

36 Introduction

37 *Clostridium thermocellum* (aka *Acetivibrio thermocellus*, *Hungateiclostridium*
38 *thermocellum*, and *Ruminiclostridium thermocellum*) is a thermophilic, anaerobic
39 bacterium that can ferment crystalline cellulose to ethanol and has attracted interest as
40 a candidate for cellulosic biofuel production (1). Its ability to deconstruct crystalline
41 cellulose is mediated by a protein complex called a cellulosome (2), and this system
42 may have applications for deconstruction of other polymers, including plastics (3). In
43 many cases, however, native properties need to be improved for industrial application.

44

45 Adaptive laboratory evolution (ALE) is a commonly used strategy for improving desired
46 properties of strains by growing them in specified growth conditions for many
47 generations, but experiments can take anywhere from weeks to years (4). Increasing
48 the mutation rate of a strain can reduce the duration of ALE experiments. Mutations in
49 DNA polymerase III are known to affect the mutation rate of bacteria. DNA polymerase
50 III is responsible for replication of the bacterial genome. It comes in two major forms,
51 DnaE and PolC. The widely-studied *Escherichia coli* only has the DnaE-type enzyme,
52 and many groups have found mutator mutations in this gene (5–11). Typically,
53 mutations that affect the fidelity of the DNA Polymerase III holoenzyme are found in the
54 polymerase domain of DnaE or the separate epsilon proofreading subunit.

55

56 The PolC-type enzyme is found primarily in gram positive bacteria with low GC content,
57 and has received much less attention (12). Organisms with PolC typically do not have a
58 separate epsilon proofreading subunit, and instead rely on proofreading activity of the
59 PHP domain within the PolC protein. Several hypermutator mutations in the *polC* gene
60 in *Bacillus subtilis* have been identified (13) (14) (15). A previous ALE experiment
61 identified a *polC* mutation among many mutations in a strain of *C. thermocellum* with a
62 hypermutator phenotype (16), however the causality of the *polC* mutation was not
63 verified, and the mutation rate was not determined.

64

65 In ALE experiments, identifying mutations is only the first step in strain improvement.
66 Mutations have to be subsequently re-introduced so that their effect can be
67 characterized. Previously, it has been difficult to reintroduce point mutations into *C.*
68 *thermocellum*. Most examples required the deletion of the wild type gene followed by re-
69 introduction of the mutant gene (17, 18). This process is very time-consuming (~2
70 months per mutation), and requires that deletion of the target gene is not toxic.
71 Recently, we developed a new CRISPR-based system for introducing point mutations in
72 *C. thermocellum*, based on either the native Type I or heterologous Type II CRISPR
73 systems (19).

74

75 In this work, we use the newly-developed CRISPR tools to characterize the effect of a
76 single nucleotide mutation in *polC* on the mutation rate of *C. thermocellum*. The ability

77 to both rapidly create diversity with controllable mutator phenotypes, and re-introduce
78 the resulting mutations with CRISPR tools, dramatically improves our ability to perform
79 ALE on *C. thermocellum*.

80 Methods

81 Plasmid construction

82 Plasmid construction was performed by isothermal DNA assembly (21), using
83 NEBuilder HiFi DNA Assembly Cloning Kit from NEB. Synthetic DNA was purchased
84 from IDT as gBlocks (Integrated DNA Technologies, Coralville, IA).

85 Growth conditions

86 Routine cultivation was performed using either CTFUD rich medium or MTC-5
87 chemically defined medium (22). Cells were grown at 55°C unless otherwise noted. For
88 solid medium, agar was used at a concentration of 0.8%. Selection for the *cat* marker
89 was performed with 6 µg/ml thiamphenicol. Stock solutions of thiamphenicol were
90 prepared in DMSO at 1000x concentration of 6 mg/ml (Sigma, part number:T0251-5G).
91 Selection for the *neo* marker was performed with 150 µg/ml neomycin, stock solution 50
92 mg/ml in water (Gibco, part number: 21810-031). Selection for *pyrF* mutations was
93 performed with 5-fluoroorotic acid (5-FOA, Zymo Research, part number F9001-5) at a
94 final concentration of 0.5 mg/ml. A 200x 5-FOA stock solution was prepared fresh daily

95 by dissolving 100 mg 5-FOA in 1 ml DMSO. When grown on defined medium, strains
96 with *pyrF* mutations were supplemented with 40 µg/ml uracil (20), a stock solution of 40
97 mg/ml uracil was prepared in 1N NaOH (Sigma, part number: U7050-5g).

98 Two-step CRISPR approach for introducing mutations

99 A two-step CRISPR Type I-B approach was used to introduce the *poIC* mutation, based
100 on our previously described approach (19) with a few modifications. In the two-step
101 approach, first the homology arm plasmid is introduced (primary transformation) and
102 cells are grown to allow homologous recombination to occur. Then a secondary
103 transformation is performed to introduce a killing plasmid to eliminate cells that have not
104 undergone homologous recombination.

105 **Primary transformation with homology template.** Cultures of parent strain LL1586
106 (native Type I-B CRISPR system upregulated by insertion of Tsac_0068 promoter) were
107 grown to mid log phase in 50 ml of CTFUD rich medium at 55°C. The culture was
108 centrifuged, rinsed twice with H₂O, resuspended in H₂O, transformed using
109 electroporation with plasmid pLT237 containing the *poIC* C669Y homology repair
110 template. Transformation was performed using electroporation with a square pulse. The
111 amplitude was 1500 V. A single pulse with a duration of 1.5 ms was applied to a 1 mm
112 cuvette. Cells were removed from the electroporation cuvette and allowed to recover for
113 16 hours in 2 mL CTFUD at 50°C overnight (22). The sample was plated on CTFUD/TM
114 (CTFUD medium with 6 µg/ml thiamphenicol) After 5 days, 10 colonies were picked and
115 pooled into 1 mL CTFUD/TM, and incubated overnight at 55°C. Colonies were pooled to

116 simplify the subsequent subculturing step. The pooled colonies were subcultured twice
117 to provide an opportunity for the homologous recombination events that are selected for
118 in the secondary transformation (19). A 1:20 dilution (50 ul into 1 ml) of the pooled
119 colony culture was prepared in CTFUD/TM medium and grown at 55°C overnight for the
120 first transfer. This was repeated for a second transfer. PCR using primers (TY
121 113+/114+) was used to confirm that plasmid pLT237 was present in the primary
122 culture of 10 pooled colonies, the first transfer culture, and the second transfer culture.
123 All three cultures were all stored at -80°C

124

125 **Secondary Transformation of LL1586/pLT237 cells with plasmids containing**
126 **spacers targeting the wildtype *polC* gene.** The first and second transfer cultures
127 generated after the primary transformation were pooled and 80 uL of the mixture was
128 inoculated into 10 mL of CTFUD and was grown overnight with and without
129 thiamphenicol (6 µg/mL) to determine the effect of growing the secondary
130 transformation with and without antibiotic selection (note: transformation worked better
131 with added antibiotic at this step, see the results section). These cultures were diluted
132 1:10 and grown (+/- thiamphenicol) to mid-log phase (Abs_{600} between 0.5 and 0.8),
133 harvested, and transformed as described above. The harvested cells were transformed
134 with a mixture of killing spacer plasmids (pDGO186N-KS1, pDGO186N-KS2, and
135 pDGO186N-KS3), and with a no-DNA control. The plasmids contained a killing spacer
136 (KS1, KS2, or KS3) as well as a neomycin (*neo*) selection marker. After overnight
137 recovery, a portion of the cells was plated on CTFUD/NEO (150 µg/mL neomycin) and

138 incubated at 55°C. Colonies usually appeared after 3 days, and were picked 1-2 days
139 later, resuspended in 0.5 mL of CTFUD/NEO and grown overnight at 55°C. After
140 neomycin selection, colonies from secondary transformations were screened for the
141 target mutation using qPCR.

142 **HRM qPCR technique for screening point mutations.** The HRM (high-resolution melt
143 analysis) qPCR reaction used a 100 bp amplicon and 25 bp forward and reverse
144 primers that flanked the mutation site (Table 2). For the reaction mixture, 2 uL of
145 bacterial culture was mixed with 10 uL 2X Sso Fast Evagreen qPCR mixture (BioRad
146 USA). The primers were mixed into an equal forward/reverse (F/R) primer mixture and
147 serially diluted to 5 uM. 2 uL of the F/R primer mixture was added along with 6 uL
148 deionized water. The mixture was set to PCR conditions of 95°C/5 min 40X, 95°C/15 s,
149 55°C/1 min. The range of the melt curve was set to 65°C-95°C at a rate of 0.2°C/10 s. The
150 uAnalyze v2 software (23) (24) was used to analyze the raw fluorescence data by
151 making normalized, derivative, and difference plots.

152 5-FOA resistance test

153 For the 5-FOA resistance test, 5 to 50 ul of a freezer stock, was inoculated into 1.5 ml
154 CTFUD media, grown for 6-8 hours to mid-log phase (to ensure a rapidly-dividing
155 culture, which is optimal for 5-FOA selection), and 100 ul was plated at various dilutions
156 with and without 5-FOA to ensure that there were between 10-200 colonies on each
157 plate. Usually the 10^{-4} , 10^{-5} and 10^{-6} dilutions had a countable number of colonies for

158 the 5-FOA plates. The plates were incubated at 55°C for 4-6 days, and colonies were
159 counted to determine the fraction of cells exhibiting 5-FOA resistance.

160 Mutation accumulation experiment

161 To determine the mutation rate by mutation accumulation, cultures were serially
162 transferred. In bacterial cells, single colony isolation events usually correspond to about
163 23-25 generations (25, 26). For *C. thermocellum*, based on cell volume (0.4 μ m
164 diameter by 2 μ m long (27)), and colony volume (lenticular shape approximately 1 mm
165 in diameter by 0.4 mm in height), a colony should contain about 5×10^8 cells,
166 corresponding to 29 generations. Approximately another 18 generations (1:100 dilution
167 from initial culture of colony pick and 1:2500 dilution subculture for freezer stock
168 preparation) occurred between the initial colony isolation following the introduction of
169 the *polC* mutation and the preparation of the freezer stock. This does not affect the
170 mutations observed in the starting strains (L1699, LL1700, LL1738, LL1740-LL1746),
171 but needs to be considered for the subsequent single colony isolations. Each serial
172 transfer consisted of a 1:100 dilution (~6.6 generations). After 10 1:100 serial transfers,
173 the 11th transfer was a 1:5000 dilution (~12.3 generations) to provide extra volume for
174 preparing freezer stocks and gDNA for WGS. This was repeated three times. The 11th,
175 22nd, and 33rd transfers were stored as populations in the freezer, and the presence of
176 the *polC* mutation was confirmed by HRM qPCR.

177

178 Single colonies were isolated from each population to create a population bottleneck to
179 fix mutations. This involved another 1:100 subculture (~6.6 generations), followed by
180 plating on solid medium. A single colony was picked, grown up, and prepared for whole
181 genome sequencing (WGS).

182 Whole genome resequencing (WGS) at Dartmouth

183 Genomic DNA was prepared using the Omega E.Z.N.A. kit following the manufacturer's
184 protocol (Omega Bio-Tek, GA, USA). 500 ng of DNA was used for WGS library
185 preparation using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New
186 England Biolabs, MA, USA). Fractionated, adapter ligated DNA fragments went through
187 5 rounds of PCR amplification and purification. The resulting WGS library was
188 sequenced at the Genomics and Molecular Biology Shared Resource (GMBSR) at
189 Dartmouth. Libraries were diluted to 4 nM, pooled and loaded at 1.8 pM onto a
190 NextSeq500 Mid Output flow cell, targeting 130 million 2x150 bp reads/sample. Base-
191 calling was performed on-instrument using RTA2 and bcls converted to fastq files using
192 bcl2fastq2 v2.20.0.422.

193 Whole genome resequencing (WGS) at JGI

194 Genomic DNA was submitted to the Joint Genome Institute (JGI) for sequencing with an
195 Illumina MiSeq instrument. Paired-end reads were generated, with an average read
196 length of 150 bp and paired distance of 500 bp. Unamplified libraries were generated
197 using a modified version of Illumina's standard protocol. 100 ng of DNA was sheared to

198 500 bp using a focused ultrasonicator (Covaris). The sheared DNA fragments were size
199 selected using SPRI beads (Beckman Coulter). The selected fragments were then end
200 repaired, A-tailed and ligated to Illumina compatible adapters (IDT, Inc) using KAPA
201 Illumina library creation kit (KAPA biosystems). Libraries were quantified using KAPA
202 Biosystem's next-generation sequencing library qPCR kit and run on a Roche
203 LightCycler 480 real-time PCR instrument. The quantified libraries were then
204 multiplexed into pools for sequencing. The pools were loaded and sequenced on the
205 Illumina MiSeq sequencing platform utilizing a MiSeq Reagent Kit v2 (300 cycle)
206 following a 2 × 150 indexed run recipe.

207 WGS data analysis

208 Read data was analyzed with the CLC Genomic Workbench version 12 (Qiagen Inc.,
209 Hilden, Germany). First, reads were trimmed using a quality limit of 0.05 and ambiguity
210 limit of 2. Then 2.5M reads were randomly selected (to avoid errors due to differences in
211 the total number of reads). Reads were mapped to the reference genome (NC_017992).
212 Mapping was improved by two rounds of local realignment. The CLC Basic Variant
213 Detection algorithm was used to determine small mutations (single and multiple
214 nucleotide polymorphisms, short insertions and short deletions). Variants occurring in
215 less than 35 % of the reads or fewer than 4 reads were filtered out. The fraction of the
216 reads containing the mutation is presented in Supplementary Table 1. To determine
217 larger mutations, the CLC InDel and Structural Variant algorithm was run. This tool
218 analyzes unaligned ends of reads and annotates regions where a structural variation

219 may have occurred, which are called breakpoints. Since the read length averaged 150
220 bp and the minimum mapping fraction was 0.5, a breakpoint can have up to 75 bp of
221 sequence data. The resulting break- points were filtered to eliminate those with fewer
222 than ten reads or less than 20 % “not perfectly matched.” The breakpoint sequence was
223 searched with the Basic Local Alignment Search Tool (BLAST) algorithm (28) for
224 similarity to known sequences. Pairs of matching left and right breakpoints were
225 considered evidence for structural variations such as transposon insertions and gene
226 deletions. The fraction of the reads supporting the mutation (left and right breakpoints
227 averaged) is presented in Supplementary Table 1. Mutation data from CLC was further
228 processed using custom Python scripts (<https://github.com/danolson1/cth-mutation>).

229 Sanger sequencing

230 Colony PCR was performed on bacterial cultures and the PCR product purified using
231 the DNA Clean and Concentrator kit (Zymo). Purified PCR products were sequenced at
232 Genewiz (USA).

233 Quantification of mutation rate

234 The mutation rate (base substitution mutation rate per nucleotide site per generation)
235 was determined by considering only synonymous mutations, which are generally
236 assumed to have a low effect on fitness (29). This was done to avoid underestimating
237 the mutation rate, which can occur when lethal mutations are generated that are not
238 observed. . This technique is commonly used for determining mutation rates from

239 whole-genome resequencing data (30) *C. thermocellum* has a genome size of
240 3,561,619 bp (NC_017304.1) of which 83% consists of coding regions. For each codon,
241 the number of synonymous single-substitution events were counted, and multiplied by
242 the codon frequency (Kazusa Codon Usage Database) (31), to reveal that 21.4% of all
243 nucleotide positions in *C. thermocellum* coding sequences allow a synonymous
244 mutation. This results in an effective genome size of 632,024 bp. The mutation rate (μ ,
245 mutations per base pair per generation) is calculated with the following equation:

$$\mu = \frac{m}{nT}$$

246 Where m is the number of observed mutations, n is the number of sites analyzed on the
247 genome, and T is the number of generations (32, 33).

248 Results and Discussion

249 CRISPR system successfully introduces point mutations

250 Initially, we attempted to re-introduce the C669Y mutation into the *polC* gene using
251 standard homologous recombination techniques (22). This required cloning the entire
252 *polC* gene to ensure that a functional copy was present at each stage of the
253 chromosomal modification. Despite repeated attempts, we were unable to construct the
254 deletion vector due to apparent toxicity in *E. coli*.

255

256 We thus pursued an alternative approach, using a recently-developed chromosomal
257 modification technique that co-opts the native Type I CRISPR system in *C.*
258 *thermocellum* (19). This system involves two transformation events (Fig. 1). The first
259 transformation introduces the homology repair template, which introduces the desired
260 point mutation, as well as several silent mutations to prevent spacer recognition. The
261 second transformation introduces the killing spacer module, which targets
262 chromosomes with the wild type *po/C* sequence, but not ones modified by the homology
263 repair template. Colonies were screened for the presence of the *po/C* mutation using an
264 HRM PCR assay (Supplementary Figure 1), which identifies mutations by a difference
265 in the melting temperature of mutant PCR amplicons. After transformation with the
266 killing spacer plasmid, colonies were picked, and the *po/C* region was analyzed by HRM
267 qPCR. A total of 309 colonies were picked for HRM qPCR screening and 17 mutant
268 candidates were identified (Table 3), Sanger DNA sequencing confirmed the presence
269 of the *po/C* mutation in all 17 candidates. Performing the secondary transformation with
270 cells grown in the presence of thiamphenicol resulted in 5-fold higher transformation
271 efficiency. Mutations were subsequently confirmed by Sanger sequencing and whole-
272 genome sequencing, and 17 of 309 colonies (6%) had the correct mutant genotype. In
273 most strains with the *po/C* mutation, both silent mutations were also present. One strain
274 (LL1746) was missing one of the silent mutations (presumably due to a homologous
275 recombination event between the silent mutation and the target mutation on the
276 homology repair template), indicating that both silent mutations are not necessary to
277 prevent CRISPR targeting.

278

279

280

281 The genome editing work was performed at the same time as some of the experiments
282 in our previous publication on CRISPR-based editing in *C. thermocellum* (19), and thus
283 does not include several improvements described in that work, such as the
284 incorporation of thermostable recombinases (*exo* and *beta* genes from *Acidithiobacillus*
285 *caldus*).

286

287 To eliminate CRISPR-mediated restriction, we opted to make two silent mutations in the
288 spacer region, rather than disrupt the PAM sequence (Fig. 1). A benefit of this approach
289 is that we can use a single homology arm plasmid (first transformation) with several
290 killing spacer plasmids (second transformation), however the editing efficiency is slightly
291 lower than what we previously reported (6% vs 40%) (19).

292 Mutation quantification by 5-FOA resistance

293 The effect of the PolC^{C669Y} mutation was determined with a 5-FOA resistance assay
294 (Fig. 2). In rich media, the *de novo* pyrimidine biosynthesis pathway is non-essential,
295 and it is thus a neutral site where mutations can accumulate. Mutations that inactivate

296 this pathway create a 5-FOA resistant (Foa^r) phenotype that can readily be detected,
297 and this assay is frequently used as a measure of mutation rates (34–37). In this
298 experiment, we observed a natural abundance of the Foa^r phenotype at about 2% of
299 WT cells, similar to what we have reported previously for *C. thermocellum* (20).
300 Disruption of *pyrF* largely eliminates sensitivity to 5-FOA, with more than 70% of
301 colonies exhibiting the Foa^r phenotype (the reason this number is not 100% is likely due
302 to a decrease in plating efficiency in the selective condition). Most of the strains
303 harboring the PolC^{C669Y} mutation also show an increase in the Foa^r phenotype, however
304 this was not universally observed. Strain LL1700 has the PolC^{C669Y} mutation, but did not
305 generate Foa^s mutants at a rate different from the WT strain, and strain LL1740 does
306 not have the PolC^{C669Y} mutation, but generated Foa^r mutants at high frequency.

307

308 In order to calculate the mutation rate, the target size needs to be known. We expected
309 to observe mutations primarily in the *pyrF* gene, however Sanger sequencing did not
310 reveal the expected mutations at that locus. Initially we plated on CTFUD with 5-FOA,
311 picked 12 LL1299 (WT) colonies and sequenced the *pyrF* gene, but did not find any
312 mutations. We sequenced the *pyrF* gene from 40 LL1699 (*polC* mutant) colonies picked
313 from 5-FOA plates, and only found one mutation (a single nucleotide insertion resulting
314 in a frame shift). Since *pyrF* mutants exhibit a growth defect that can be complemented
315 with added uracil (20), we repeated the LL1299 plating experiment on CTFUD with 5-
316 FOA and 40 µg/ml uracil. Of 18 colonies, 5 had mutations (mostly frame-shifts and
317 premature stop codons) at the *pyrF* locus.

318

319 Subsequent whole-genome sequencing (Supplementary Table 1) did not identify any
320 mutations associated with *de novo* pyrimidine biosynthesis (Clo1313_1262 through
321 Clo1313_1270, which includes *pyrE*, *pyrB*, *pyrC*, *pyrF*, *carA*, and *pyrD*), and the genetic
322 basis for the Foa^r phenotype in these strains remains unknown, which prevents an
323 accurate determination of the mutation rate. Instead, we decided to measure mutation
324 accumulation directly, by whole-genome sequencing.

325 Mutation quantification by whole genome sequencing

326 For this mutation accumulation experiment, three strains were selected that had
327 undergone CRISPR mutagenesis: one where the mutagenesis had failed (strain
328 LL1738, WT at *polC*) to serve as a control, and two where the CRISPR mutagenesis
329 had been successful (strains LL1742 and LL1745). Each strain was serially transferred
330 for several generations (Fig. 3). Every 11 generations, the population was stored in the
331 freezer, and a single colony was isolated. Observing the accumulation of mutations over
332 time allows us to estimate the mutation rate.

333 Whole-genome sequencing indicated a substantial increase in the rate of mutation
334 accumulation for strains with the PolC^{C669Y} mutation (Fig. 4). Several categories of
335 mutations were identified:

336 One category is structural rearrangements, which are mutations that involve large (> 3
337 nucleotide) insertions, deletions, or replacements. This includes the targeted genetic

338 modifications (deletion of *hpt* in strain LL345 (38), deletion of *reIII* in strain LL1299 (39),
339 insertion of a strong constitutive promoter from *Thermoanaerobacterium*
340 *saccharolyticum* driving the native Type I cas operon in strain LL1568 (19). This also
341 includes transposon insertions. *C. thermocellum* has several native transposon
342 elements (16, 40). Transposons insertions from families IS2, IS10, and IS120 (41) were
343 observed. All of these insertions, except for two in the LL1769 population, were
344 inherited from the LL1586 parent strain. The two in the LL1769 population were not
345 present in either the subsequent serial transfer (LL1770 population) or the single colony
346 isolate from that transfer (strain LL1795), and we therefore suspect it appeared during
347 the preparation for genomic DNA extraction for whole-genome resequencing. Mutations
348 in this category comprise about 1% of the total number of mutations (7 of 775 for the 12
349 single colony isolates in the mutation accumulation experiment). Since these mutations
350 are not expected to be affected by PolC mutations, they were excluded from
351 subsequent analysis.

352

353 Another category is short (1-3 nucleotide) insertions, deletions, and replacements.
354 These comprise the majority (99%) of observed mutations, and the incidence of these
355 mutations was elevated in PolC mutant strains (Fig. 4).

356 The mutation distribution is not uniform, with an underrepresentation of A:T → T:A and
357 G:C → C:G transversion relative to other types of transitions and transversions, and an

358 overrepresentation of C:G → A:T transversions (Figure 5). This tendency has been
359 observed by others, however a mechanism is not known (30, 42). Nevertheless, it is
360 important to take this into consideration when designing adaptive laboratory evolution
361 experiments, since this mutational bias affects the likelihood of observing amino acid
362 changes.

363 Mutation rate estimation

364 The simplest way to calculate the mutation rate is to divide the total number of
365 mutations by the total number of generations. However, some mutations are lethal and
366 thus not observed. To correct for this, we can use synonymous SNVs, a subset of
367 mutations that are presumed to be neutral (43). In the *polC* mutant strains, the median
368 mutation rate is 0.10 mutations per generation or 1.6e-7 per nucleotide. It is not possible
369 to determine the mutation rate for the WT strain using synonymous mutations, since
370 even after 290 generations, no synonymous mutations were observed. We can,
371 however, establish an upper bound of about 5.5e-9 per nucleotide (assuming a single
372 mutation appeared after 290 generations). We can also estimate the upper bound
373 considering all four non-synonymous mutations in strain LL1796 (WT *PolC*, transfer 33).
374 This gives a mutation rate of 3.9e-9 per generation. Many organisms have a mutation
375 rate of about 0.0033 per generation (44), which would be 9.2e-10 for an organism with
376 the genome size of *C. thermocellum*. Thus, the *polC* mutation appears to have

377 increased the mutation rate between 30 and 178-fold, and looking at the structure of this
378 enzyme suggests a possible mechanism.

379 C669Y mutation may disrupt proofreading activity of *polC*

380 DNA polymerase III comes in two major forms: DnaE, and PolC. The DnaE types are
381 further divided into three subtypes (DnaE1, DnaE2, and DnaE3). *C. thermocellum*
382 contains both a PolC type (Clo1313_1219) and a DnaE1 type (Clo1313_0994), both of
383 which are expressed (16). Many organisms with PolC, including *C. thermocellum*, do
384 not have an epsilon proofreading subunit. In these organisms, proofreading is mediated
385 either by the embedded EXO domain or by the PHP domain itself. This domain has
386 been shown to have proofreading activity in *Thermus thermophilus* (45, 46) The C669Y
387 mutation is located in the PHP domain of PolC (Fig. 6). Exonuclease activity depends
388 on coordination with several metal ions via nine highly conserved residues. (12) The
389 C669Y mutation disrupts one of these residues, which may subsequently disrupt metal
390 ion binding and thus impair proofreading activity.

391 Similar mutations have been observed in *Bacillus* and *E. coli*. In *B. subtilis*, the A662V
392 mutation is adjacent to the metal-binding cysteine. It does not have an effect on the
393 mutation rate, but *does* cause temperature instability. (13) In *E. coli*, the *dnaE74*
394 mutation (G134R) is in a similar location (5), however in *E. coli*, the PHP domain does
395 not have any of the conserved metal-binding residues, and is not thought to have
396 catalytic activity. The observed change in mutation rate (1.8-fold increase) in that

397 organism may be due to changes in binding affinity for the epsilon proofreading subunit
398 (*dnaQ*) instead.

399 Other mutations that may affect the mutation rate

400 In addition to the *polC* mutation, several other mutations were identified that might affect
401 the mutation rate (Supplementary Table 1). Strain LL1803 (transfer 22 in the LL1745
402 lineage) has a mutation in the *mutS* gene (Clo1313_1201) responsible for DNA
403 mismatch repair. Mutations in this gene have been shown to increase the mutation rate
404 in other organisms (47, 48). Strain LL1797 (transfer 1 in the LL1742 lineage) has a
405 mutation in the DNA polymerase III delta subunit (Clo1313_1173). The delta subunit is
406 part of the DNA polymerase III holoenzyme, and is responsible for opening the sliding
407 beta clamp protein to accept a DNA strand for replication (49). Strain LL1800 (transfer
408 33 in the LL1742 lineage) has a mutation in the *polA* polymerase (Clo1313_1334). This
409 polymerase is thought to assist in repairing damaged DNA (50). Several strains
410 (LL1801, LL1803, and LL1804) have mutations in the *lexA* DNA binding protein
411 (Clo1313_2881). All three mutations are in different locations in the gene, and two of
412 them likely eliminate activity (one is a frameshift mutation, the other is a stop codon).
413 The *lexA* gene works in tandem with the *recA* gene to induce the SOS response
414 (programmed DNA repair) (51). Since *lexA* is a repressor of SOS activity, its inactivation
415 by mutation would be expected to lead to constitutive induction of the SOS response,
416 which could alter the mutation rate.

417

418 Hypermutator phenotypes that arise in bacterial populations typically revert to the
419 ancestral mutation rate when maintained in stable conditions (52). The decrease in
420 mutations observed between generations 211 and 290 is most likely explained by the
421 takeover of the LL1745 lineage by a subpopulation with fewer mutations.

422 Conclusions

423 We demonstrate the utility of our recently developed CRISPR/cas system to
424 successfully introduce a PolC^{C669Y} mutation in *C. thermocellum*. We found the HRM
425 technique to be useful for rapidly screening colonies to identify the successful
426 introduction of point mutations. The single C669Y mutation in PolC protein in *C.*
427 *thermocellum* is sufficient to increase the mutation rate about 30-fold. This mutation
428 appears to function by interfering with metal ion coordination in the PHP domain
429 responsible for proofreading. The ability to selectively increase the mutation rate in *C.*
430 *thermocellum* is a useful tool for directed evolution experiments.

431

432

433 Acknowledgements

434 We thank Dr. Česlovas Venclovas for useful discussions related to DNA polymerases.

435

436 Funding was provided by The Center for Bioenergy Innovation, a U.S. Department of
437 Energy Research Center supported by the Office of Biological and Environmental
438 Research in the DOE Office of Science. Whole genome resequencing was performed
439 by the Department of Energy Joint Genome Institute, a DOE Office of Science User
440 Facility, and is supported by the Office of Science of the U.S. Department of Energy
441 under contract number DE-AC02-05CH11231.

442 Additional whole genome resequencing was carried out in the Genomics and Molecular
443 Biology Shared Resource (GMBSR) at Dartmouth which is supported by NCI Cancer
444 Center Support Grant 5P30CA023108

445

446 Lee R. Lynd is a cofounder of the Enchi corporation, a start-up company focusing on
447 cellulosic ethanol production using *Clostridium thermocellum*. There are no other
448 competing interests.

449

450 Bibliography

- 451 1. Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS,
452 Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. *Curr Opin*
453 *Biotechnol* 45:202–211.
- 454 2. Bayer EA, Lamed R, White BA, Flint HJ. 2008. From cellulosomes to

- 455 cellulomics. Chem Rec 8:364–377.
- 456 3. Yan F, Wei R, Cui Q, Bornscheuer UT, Liu Y-J. 2021. Thermophilic whole-cell
457 degradation of polyethylene terephthalate using engineered *Clostridium*
458 *thermocellum*. Microb Biotechnol 14:374–385.
- 459 4. Sandberg TE, Salazar MJ, Weng LL, Palsson BO, Feist AM. 2019. The
460 emergence of adaptive laboratory evolution as an efficient tool for biological
461 discovery and industrial biotechnology. Metab Eng 56:1–16.
- 462 5. Vandewiele D, Fernández de Henestrosa AR, Timms AR, Bridges BA, Woodgate
463 R. 2002. Sequence analysis and phenotypes of five temperature sensitive mutator
464 alleles of dnaE, encoding modified alpha-catalytic subunits of *Escherichia coli*
465 DNA polymerase III holoenzyme. Mutat Res 499:85–95.
- 466 6. Strauss BS, Roberts R, Francis L, Pouryazdanparast P. 2000. Role of the *dinB*
467 gene product in spontaneous mutation in *Escherichia coli* with an impaired
468 replicative polymerase. J Bacteriol 182:6742–6750.
- 469 7. Maki H, Mo JY, Sekiguchi M. 1991. A strong mutator effect caused by an amino
470 acid change in the alpha subunit of DNA polymerase III of *Escherichia coli*. J Biol
471 Chem 266:5055–5061.
- 472 8. Mo JY, Maki H, Sekiguchi M. 1991. Mutational specificity of the *dnaE173* mutator
473 associated with a defect in the catalytic subunit of DNA polymerase III of
474 *Escherichia coli*. J Mol Biol 222:925–936.
- 475 9. Ruiz-Rubio M, Bridges BA. 1987. Mutagenic DNA repair in *Escherichia coli*. XIV.

- 476 Influence of two DNA polymerase III mutator alleles on spontaneous and UV
477 mutagenesis. *Mol Gen Genet* 208:542–548.
- 478 10. Konrad EB. 1978. Isolation of an *Escherichia coli* K-12 *dnaE* mutation as a
479 mutator. *J Bacteriol* 133:1197–1202.
- 480 11. Sevastopoulos CG, Glaser DA. 1977. Mutator action by *Escherichia coli* strains
481 carrying *dnaE* mutations. *Proc Natl Acad Sci USA* 74:3947–3950.
- 482 12. Timinskas K, Balvočiūtė M, Timinskas A, Venclovas Č. 2014. Comprehensive
483 analysis of DNA polymerase III α subunits and their homologs in bacterial
484 genomes. *Nucleic Acids Res* 42:1393–1413.
- 485 13. Barnes MH, Hammond RA, Kennedy CC, Mack SL, Brown NC. 1992. Localization
486 of the exonuclease and polymerase domains of *Bacillus subtilis* DNA polymerase
487 III. *Gene* 111:43–49.
- 488 14. Gass KB, Cozzarelli NR. 1973. Further genetic and enzymological
489 characterization of the three *Bacillus subtilis* deoxyribonucleic acid polymerases. *J*
490 *Biol Chem* 248:7688–7700.
- 491 15. Paschalis V, Le Chatelier E, Green M, Nouri H, Képès F, Soutanas P, Janniere L.
492 2017. Interactions of the *Bacillus subtilis* DnaE polymerase with replisomal
493 proteins modulate its activity and fidelity. *Open Biol* 7.
- 494 16. Holwerda EK, Olson DG, Ruppertsberger NM, Stevenson DM, Murphy SJL,
495 Maloney MI, Lanahan AA, Amador-Noguez D, Lynd LR. 2020. Metabolic and
496 evolutionary responses of *Clostridium thermocellum* to genetic interventions aimed

- 497 at improving ethanol production. *Biotechnol Biofuels* 13:40.
- 498 17. Zheng T, Olson DG, Tian L, Bomble YJ, Himmel ME, Lo J, Hon S, Shaw AJ, van
499 Dijken JP, Lynd LR. 2015. Cofactor Specificity of the Bifunctional Alcohol and
500 Aldehyde Dehydrogenase (AdhE) in Wild-Type and Mutant *Clostridium*
501 *thermocellum* and *Thermoanaerobacterium saccharolyticum*. *J Bacteriol*
502 197:2610–2619.
- 503 18. Lo J, Zheng T, Hon S, Olson DG, Lynd LR. 2015. The bifunctional alcohol and
504 aldehyde dehydrogenase gene, *adhE*, is necessary for ethanol production in
505 *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. *J*
506 *Bacteriol* 197:1386–1393.
- 507 19. Walker JE, Lanahan AA, Zheng T, Toruno C, Lynd LR, Cameron JC, Olson DG,
508 Eckert CA. 2020. Development of both type I-B and type II CRISPR/Cas genome
509 editing systems in the cellulolytic bacterium *Clostridium thermocellum*. *Metab Eng*
510 *Commun* 10:e00116.
- 511 20. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool
512 JD, Warner AK, Rajgarhia VB, Lynd LR, Hogsett DA, Caiazza NC. 2010.
513 Development of *pyrF*-based genetic system for targeted gene deletion in
514 *Clostridium thermocellum* and creation of a *pta* mutant. *Appl Environ Microbiol*
515 76:6591–6599.
- 516 21. Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. *Meth*
517 *Enzymol* 498:349–361.
- 518 22. Olson DG, Lynd LR. 2012. Transformation of *Clostridium thermocellum* by

- 519 electroporation. *Meth Enzymol* 510:317–330.
- 520 23. Dwight ZL, Palais R, Wittwer CT. 2012. uAnalyze: web-based high-resolution DNA
521 melting analysis with comparison to thermodynamic predictions. *IEEE/ACM Trans*
522 *Comput Biol Bioinform* 9:1805–1811.
- 523 24. Atmadjaja AN, Holby V, Harding AJ, Krabben P, Smith HK, Jenkinson ER. 2019.
524 CRISPR-Cas, a highly effective tool for genome editing in *Clostridium*
525 *saccharoperbutylacetonicum* N1-4(HMT). *FEMS Microbiol Lett* 366.
- 526 25. Trindade S, Perfeito L, Gordo I. 2010. Rate and effects of spontaneous mutations
527 that affect fitness in mutator *Escherichia coli*. *Philos Trans R Soc Lond B Biol Sci*
528 365:1177–1186.
- 529 26. Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to
530 overall fitness in *E. coli*. *Nature* 381:694–696.
- 531 27. Sato K, Tomita M, Yonemura S, Goto S, Sekine K, Okuma E, Takagi Y, Hon-Nami
532 K, Saikit T. 1993. Characterization of and Ethanol Hyper-production by *Clostridium*
533 *thermocellum* I-1-B. *Biosci Biotechnol Biochem* 57:2116–2121.
- 534 28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
535 search tool. *J Mol Biol* 215:403–410.
- 536 29. Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–626.
- 537 30. Wielgoss S, Barrick JE, Tenailon O, Cruveiller S, Chane-Woon-Ming B, Médigue
538 C, Lenski RE, Schneider D. 2011. Mutation Rate Inferred from Synonymous
539 Substitutions in a Long-Term Evolution Experiment with *Escherichia coli*. G3

- 540 (Bethesda) 1:183–186.
- 541 31. Nakamura Y, Tabata S. 1997. Codon-anticodon assignment and detection of
542 codon usage trends in seven microbial genomes. *Microb Comp Genomics* 2:299–
543 312.
- 544 32. Kucukyildirim S, Behringer M, Williams EM, Doak TG, Lynch M. 2020. Estimation
545 of the Genome-Wide Mutation Rate and Spectrum in the Archaeal Species
546 *Haloferax volcanii*. *Genetics* 215:1107–1116.
- 547 33. Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL.
548 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev*
549 *Genet* 17:704–714.
- 550 34. Boeke JD, La Croute F, Fink GR. 1984. A positive selection for mutants lacking
551 orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid
552 resistance. *Mol Gen Genet* 197:345–346.
- 553 35. Grogan DW, Gunsalus RP. 1993. *Sulfolobus acidocaldarius* synthesizes UMP via
554 a standard de novo pathway: results of biochemical-genetic study. *J Bacteriol*
555 175:1500–1507.
- 556 36. Kondo S, Yamagishi A, Oshima T. 1991. Positive selection for uracil auxotrophs of
557 the sulfur-dependent thermophilic archaeobacterium *Sulfolobus acidocaldarius* by
558 use of 5-fluoroorotic acid. *J Bacteriol* 173:7698–7700.
- 559 37. Jacobs KL, Grogan DW. 1997. Rates of spontaneous mutation in an archaeon
560 from geothermal environments. *J Bacteriol* 179:3298–3303.

- 561 38. Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden
562 JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC. 2011. High ethanol titers from
563 cellulose by using metabolically engineered thermophilic, anaerobic microbes.
564 *Appl Environ Microbiol* 77:8288–8294.
- 565 39. Riley LA, Ji L, Schmitz RJ, Westpheling J, Guss AM. 2019. Rational development
566 of transformation in *Clostridium thermocellum* ATCC 27405 via complete
567 methylome analysis and evasion of native restriction-modification systems. *J Ind*
568 *Microbiol Biotechnol* 46:1435–1443.
- 569 40. Zverlov VV, Klupp M, Krauss J, Schwarz WH. 2008. Mutations in the scaffoldin
570 gene, *cipA*, of *Clostridium thermocellum* with impaired cellulosome formation and
571 cellulose hydrolysis: insertions of a new transposable element, IS1447, and
572 implications for cellulase synergism on crystalline cellulose. *J Bacteriol* 190:4321–
573 4327.
- 574 41. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the
575 reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34:D32-6.
- 576 42. Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased
577 towards AT in bacteria. *PLoS Genet* 6:e1001115.
- 578 43. Kimura M. 1984. *The Neutral Theory of Molecular Evolution*. Cambridge University
579 Press.
- 580 44. Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based
581 microbes. *Proc Natl Acad Sci USA* 88:7160–7164.

- 582 45. Stano NM, Chen J, McHenry CS. 2006. A coproofreading Zn(2+)-dependent
583 exonuclease within a bacterial replicase. *Nat Struct Mol Biol* 13:458–459.
- 584 46. Lapenta F, Montón Silva A, Brandimarti R, Lanzi M, Gratani FL, Vellosillo
585 Gonzalez P, Perticarari S, Hochkoepler A. 2016. *Escherichia coli* DnaE
586 Polymerase Couples Pyrophosphatase Activity to DNA Replication. *PLoS ONE*
587 11:e0152915.
- 588 47. Luan G, Cai Z, Gong F, Dong H, Lin Z, Zhang Y, Li Y. 2013. Developing
589 controllable hypermutable *Clostridium* cells through manipulating its methyl-
590 directed mismatch repair system. *Protein Cell* 4:854–862.
- 591 48. Willems RJ, Top J, Smith DJ, Roper DI, North SE, Woodford N. 2003. Mutations in
592 the DNA mismatch repair proteins MutS and MutL of oxazolidinone-resistant or -
593 susceptible *Enterococcus faecium*. *Antimicrob Agents Chemother* 47:3061–3066.
- 594 49. Turner J, Hingorani MM, Kelman Z, O'Donnell M. 1999. The internal workings of a
595 DNA polymerase clamp-loading machine. *EMBO J* 18:771–783.
- 596 50. Hernández-Tamayo R, Oviedo-Bocanegra LM, Fritz G, Graumann PL. 2019.
597 Symmetric activity of DNA polymerases at and recruitment of exonuclease ExoR
598 and of PolA to the *Bacillus subtilis* replication forks. *Nucleic Acids Res* 47:8521–
599 8536.
- 600 51. Butala M, Zgur-Bertok D, Busby SJW. 2009. The bacterial LexA transcriptional
601 repressor. *Cell Mol Life Sci* 66:82–93.
- 602 52. Couce A, Caudwell LV, Feinauer C, Hindré T, Feugeas J-P, Weigt M, Lenski RE,

603 Schneider D, Tenaillon O. 2017. Mutator genomes decay, despite sustained
604 fitness gains, in a long-term experiment with bacteria. Proc Natl Acad Sci USA
605 114:E9026–E9035.

606

607 Figure Captions

608 **Figure 1. A) 2-step CRISPR system for introducing mutations in *C. thermocellum*.** Step 1 is
609 transformation with a homology arm plasmid, and step 2 is transformation with a killing spacer cassette
610 plasmid. Each plasmid contains a *C. thermocellum* origin (including the *repB* replication protein), an *E.*
611 *coli* origin, and either a chloramphenicol (*cat*) or a neomycin (*neo*) selection marker. **B) The *polC* region**
612 **targeted for mutations.** The spacer cassette is composed of three spacers oriented in the forward (KS1
613 or KS2) or reverse (KS3) directions. Spacer regions were chosen to be immediately downstream of a TTN
614 or TCD PAM sequence. The three target mutations are also shown: two silent mutations (to disrupt
615 CRISPR targeting) and the targeted point mutation to create a cysteine to tyrosine amino acid change at
616 position 669. The *polC* mutation targets were analyzed by Sanger sequencing. Strain LL1738 is an
617 example of failed mutagenesis, and displays a wild-type sequence. Strain LL1745 is an example of
618 successful mutagenesis, exhibiting all three targeted mutations. In strain LL1746, the *polC* mutation was
619 introduced successfully with only one of the two silent mutations. Detailed plasmid maps are available
620 from Addgene.

621 **Figure 2. 5-FOA resistance mutation rate assay. A)** Schematic diagram of cell growth
622 and selective plating. **B)** Dilution plates in the selective (with 5-FOA) and permissive (no
623 5-FOA) conditions. The parent strain (LL1586) is WT at the *pyrF* locus and Foa^S. Strain
624 LL1005 has a targeted *pyrF* deletion and is Foa^r. Strain LL1738 exhibits the Foa^S
625 phenotype. Strain LL1699 exhibits the Foa^r phenotype. **C)** Foa^r phenotype for wild-type

626 and mutant *polC* strains. The *pyrF* genotype, *polC* genotype, *cas* promoter insertion,
627 and success of CRISPR treatment were determined by Sanger sequencing.

628 **Figure 3.** Lineages of the strains described in this work. Strain LL1586 was derived
629 from WT *C. thermocellum* (LL1004) by a series of targeted mutations designed to
630 improve the ability to perform targeted genetic modifications. Two-step CRISPR
631 mutagenesis was performed on strain LL1586 to introduce the PolC^{C669Y} mutation. Ten
632 individual colonies (LL1699, LL1700, LL1738, LL1740-1746) were isolated and
633 sequenced at the *polC* locus. Of those, three were selected (two with the *polC* mutation,
634 and one without) for the mutation accumulation experiment.

635 **Figure 4.** Accumulation of mutations identified by whole-genome resequencing. Each
636 transfer is a 1:100 dilution. For each lineage, mutations present in the parent strain were
637 filtered out. SNV stands for “single nucleotide variation.” The data used to generate this
638 figure is presented in detail in Supplementary Table 1.

639 **Figure 5.** Comparison of mutation types. For this analysis, only synonymous SNV
640 mutations from the *polC* mutant strain lineages (strains LL1797, LL1798, LL1799,
641 LL1800, LL1801, LL1802, LL1803, and LL1804) were considered. The wild type strains
642 did not have any synonymous mutations. Expected mutation frequency was determined
643 based on the codon frequency and the probability that a mutation results in a
644 synonymous mutation.

645 **Figure 6.** Location of the PolC^{C669Y} mutation. **Panel A** shows the domain structure of
646 the PolC protein. The C669Y mutation is located in the PHP (polymerase and histidinol

647 phosphatase) domain. Other domains include N-terminal domain (NTD), oligonucleotide
648 binding (OB), exonuclease (EXO), polymerase core (Pol3), and tandem helix-hairpin-
649 helix motif (HhH). The PHP domain contains eight highly conserved residues (magenta)
650 that coordinate binding with the metal ions, Mn²⁺ and Zn²⁺. Mutations known to affect
651 polymerase fidelity in *B. subtilis* are also indicated. **Panel B** shows a detailed view of
652 the region surrounding the C669Y mutation. The C669Y mutation is adjacent to an
653 A662Y mutation observed to cause a temperature-sensitive phenotype in *B. subtilis*.
654 The C669 residue in *C. thermocellum* is in the same position as the C670 residue in *G.*
655 *kaustophilus* (PDB ID: 3F2D) that coordinates the Mn²⁺ residue.

656

657

658

659 Tables

660 Table 1. Strain and plasmids used in this work.

Strain/Plasmid	Description	Accession Number ^a	Reference
pLT237	Replicating plasmid with homology region for introducing PolC ^{C669Y} mutation. Confers thiamphenicol resistance.	Addgene 174300	This work

pDGO186N-KS1	Replicating plasmid with sgRNA cassette containing a killing spacer #1 targeting the wild type <i>polC</i> gene.	Addgene 174301	This work
pDGO186N-KS2	Same as above, with killing spacer #2	Addgene 174302	This work
pDGO186N-KS3	Same as above, with killing spacer #3	Addgene 174303	This work
LL1004	<i>C. thermocellum</i> DSMZ 1313 wild type strain; WT <i>polC</i> , WT <i>pyrF</i>	NCBI reference sequence NC_017304.1	DSMZ culture collection
LL1005	LL1004 with <i>pyrF</i> deletion	not available	(20)
LL1299	LL1004 with <i>hpt</i> deletion to allow for 8AZH selection and <i>reIII</i> deletion to improve transformation efficiency	SRX2506395	(19)
LL1586	LL1299 with upregulated <i>cas</i> expression using <i>Tsac_0068</i> promoter	SRX4823139	(19)
LL1699	Parent strain LL1586; mutant Pol III, WT <i>pyrF</i>	SRX8904888	This work

LL1700	Parent strain LL1586; mutant Pol III, WT pyrF	SRX8904887	This work
LL1738	LL1586, failed CRISPR mutagenesis at <i>polC</i> locus	SRX9642234	This work
LL1740	LL1586, failed CRISPR mutagenesis at <i>polC</i> locus	SRX9642235	This work
LL1741	LL1586, failed CRISPR mutagenesis at <i>polC</i> locus	SRX9642649	This work
LL1742	LL1586, successful CRISPR mutagenesis, PolC ^{C669Y} mutation introduced	SRX9642825	This work
LL1743	LL1586, successful CRISPR mutagenesis, PolC ^{C669Y} mutation introduced	SRX9642826	This work
LL1744	LL1586, successful CRISPR mutagenesis, PolC ^{C669Y} mutation introduced	SRX9642781	This work
LL1745	LL1586, successful CRISPR mutagenesis, PolC ^{C669Y} mutation introduced	SRX9642780	This work

LL1746	LL1586, successful CRISPR mutagenesis, PolC ^{C669Y} mutation introduced	SRX9642779	This work
LL1762	LL1742 serial transfer #11 population	SAMN20331610	This work
LL1763	LL1742 serial transfer #22 population	SAMN20331611	This work
LL1764	LL1742 serial transfer #33 population	SAMN20331612	This work
LL1765	LL1745 serial transfer #11 population	SAMN20331613	This work
LL1766	LL1745 serial transfer #22 population	SAMN20331614	This work
LL1767	LL1745 serial transfer #33 population	SAMN20331615	This work
LL1768	LL1738 serial transfer #11 population	SAMN20331616	This work
LL1769	LL1738 serial transfer #22	SAMN20331617	This work

	population		
LL1770	LL1738 serial transfer #33 population	SAMN20331618	This work
LL1793	LL1738 serial transfer #1 single colony, 50 generations from original CRISPR mutant cell	SAMN20331619	This work
LL1794	LL1738 serial transfer #11 single colony, 128 generations from original CRISPR mutant cell	SAMN20331620	This work
LL1795	LL1738 serial transfer #22 single colony, 207 generations from original CRISPR mutant cell	SAMN20331621	This work
LL1796	LL1738 serial transfer #33 single colony, 286 generations from original CRISPR mutant cell	SAMN20331622	This work
LL1797	LL1742 serial transfer #1 single colony, 50 generations from original <i>po/C</i> mutant cell	SAMN20331623	This work
LL1798	LL1742 serial transfer #11 single colony, 128 generations from original <i>po/C</i> mutant	SAMN20331624	This work

LL1799	LL1742 serial transfer #22 single colony, 207 generations from original <i>po/C</i> mutant cell	SAMN20331625	This work
LL1800	LL1742 serial transfer #33 single colony, 286 generations from original <i>po/C</i> mutant cell	SAMN20331626	This work
LL1801	LL1745 serial transfer #1 single colony, 50 generations from original <i>po/C</i> mutant cell	SAMN20331627	This work
LL1802	LL1745 serial transfer #11 single colony, 128 generations from original <i>po/C</i> mutant cell	SAMN20331628	This work
LL1803	LL1745 serial transfer #22 single colony, 207 generations from original <i>po/C</i> mutant cell	SAMN20331629	This work
LL1804	LL1745 serial transfer #33 single colony, 286 generations from original <i>po/C</i> mutant cell	SAMN20331630	This work

661 ^a Accession numbers that begin with “SRX” represent samples with WGS performed at
662 JGI, numbers that begin with “SAMN” represent samples with WGS performed at
663 Dartmouth. Both sets are available from the NCBI SRA database.

664

665 Table 2. Primers and synthetic DNA (gBlocks) used in this work.

Name	Sequence (5' → 3')	Description
TY113+	ACCTTGATGACACAGAAGAAGGCGATTTGA	Plasmid pLT237 confirmation
TY114+	CTAAAAGTCGTTTGTGGTTCA	Plasmid pLT237 confirmation
HRM F	CGAGACTTCCAAGAAAGCTTTTAAT	HRM primer
HRM R	GTCCTCATCTTTGTTTTCAAGAATC	HRM primer
1219F	GGAAATTGGTGCGGTAAAGA	PoIC HA PCR primer
1219R	CTATCCCTCTTCGTCAA	PoIC HA PCR primer
1219SF	CGTAGGCCTTAGAAATCTGTA	PoIC sequencing primer
1219SR	CCGGTATGGGAAGTATGT	PoIC sequencing primer
1218F	CCAGGAAAGGGCAGTGGGAAGA	PoIC PCR/sequencing primer
1220R	TCAAATTCATAGATTCCCAA	PoIC PCR/sequencing primer
LT651	TAATACCTGCAAAGACCC	PoIC sequencing primer
LT652	AAGGACAGAAGCAAGGGA	PoIC sequencing primer
LT652R	TCCCTTGCTTCTGTCCTT	PoIC sequencing primer

1219FR	TCTTTACCGCACCAATTTCC	PolC sequencing primer
1219SRF	ACATACTTCCCATAACCGG	PolC sequencing primer
1219RF	TTGACGAAAGAGGGATAG	PolC sequencing primer
LT656R	GGATGTAAGAAAAGGAAAGG	PolC sequencing primer
BackboneR 1	GCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA ATTGTTATCCGCTCACAATT	HA plasmid backbone PCR
BB R2 intF	ACCGCCTTTGAGTGAGCTGATACCG	HA plasmid backbone PCR
BackboneF	TAATAGAAATAATTTTGTTTAACTTTACAAACGGG ATTGACTTTTAAAAAAGGATTGATT	HA plasmid backbone PCR
BB F intR	TGCGGCGAGCGGTATCAGCTCACTC	HA plasmid backbone PCR
Backbone Reverse	GGTCGGCATATTAAGGATCCTTGA ACTACTCTTTA ATAAAATAAT	Killing plasmid backbone PCR
TY1C F	AAATTTAGGAGGCATATCAAATGGTGAATGGACC AATAATAATGACTAGAGAAGAAAGAA	Killing plasmid backbone PCR
HA gBlock	ACAGGAAACAGCTATGACCATGATT <u>ACGCCGACA</u>	PolC mutant homology arm gBlock used to construct

	<p><u>ATCCTGTAATTGACACTTTGGAAC TTTCAAGGCAA</u></p> <p><u>ATGTTTCCGGAGCTTAAAAACACAACTGGATGT</u></p> <p><u>GGTTGCAAAGCATCTTGGTGTTTCATTGGAAAACC</u></p> <p><u>ATCACAGGGCTCTTGACGATGCAAAGGCCTGTGG</u></p> <p><u>GGAGATCTTTATCAAATGCCTTGAGATACTGGCTG</u></p> <p><u>AAAAGAATGTAAAAACAATTGATGACATACAAAA</u></p> <p><u>TGTTTTGAGGGTTGTTGGAATTATCAGAAAGCAA</u></p> <p><u>ATTCATACCATGCCATTATACTTGTA AAAAATTAC</u></p> <p><u>GTAGGCCTTAGAAATCTGTACAAGATTGTGTCCAA</u></p> <p><u>GACTCATTGGAGTATTTTCACAAAAGGCCGAGAC</u></p> <p><u>TTCCAAGAAAGCTTTTAATGATGCACAGAGAAGG</u></p> <p><u>ACTCATTGGGAAGTGCaTATGAAGctGGAGAG</u></p> <p><u>CTTACCGTGCGATTCTTGAAAACAAAGATGAGG</u></p> <p><u>ACGAAATTTCAAGGATTGTAAAGTTTTACGATTAT</u></p> <p><u>CTTGAAATACAGCCTTTGGGCAACAACCGTTTCCT</u></p> <p><u>TGTCGAGAGCGGAAAGGTGAACAGTGAGGAAGA</u></p> <p><u>CTTAAAGAATATAAACAGAAAAATAGTAGCTTTG</u></p> <p><u>GGTGAGAAGTACAACAAGCCGGTGGTCGCTACCT</u></p> <p><u>GCGATGTTCAATTCATGGATCCGCAGGACGAAGTT</u></p> <p><u>TTCAGAAGAATACTTATGGCGGGGCAAGGTTTTT</u></p>	<p>pLT237</p> <p>Pol C homology region in italics and underlined</p> <p>G to A nucleotide point mutation resulting in C to Y amino acid change in bold and large font</p> <p>G to A and G to T silent nucleotide point mutations to facilitate HRM qPCR screen in lowercase and bold</p>
--	---	--

	<p><u>CCGATGCGGACAACCAGGCACCGTTGTATTTTCAG</u></p> <p><u>GACCACTGATGAAATGCTGGAAGAGTTTGAATAT</u></p> <p><u>CTGGGCAGGGAAAAATGCTATGAAGTGGTTGTAA</u></p> <p><u>CAAATACAAATTTAATTGCGGATATGTGCGAGGA</u></p> <p><u>CATACTTCCCATAACCGGAAGGAACTTTTCTCCTAA</u></p> <p><u>AATTGAAGGTGCGGAAGAAGAAATAAGGATGCT</u></p> <p><u>TGCCGAAAACAAGGCGAGAGAAATTTATGGCGAC</u></p> <p><u>CCTCTTCTGAGATTGTGAAAAGCGCATGGAAA</u></p> <p><u>AGGAGCTTAACTCAATAATAAAAAATGGCTTCTCC</u></p> <p><u>GTTATGTATATTATTGCGCAAAAACCTGTATGGAA</u></p> <p><u>ATCTTTAAGTGACGGCTATCTTGTAGGGTCGAGG</u></p> <p><u>GGATCCGTCGGATCATCCTTTGTTGCCTATAATAG</u></p> <p>AAATAATTTTGTTT</p>	
<p>KS1 gBlock1</p>	<p><u>TTCAAGGATCCTTAATATGCCGACCACGTTGCAAT</u></p> <p><u>TCCCGTCAAATAATGCATTTTGCAGCCGACGAAAC</u></p> <p><u>AGGCAAGATAACTGTATTGGCTATAAATGTTTCA</u></p> <p><u>GGCAGCGGTATATTTTGCCTCCCGTAAAATTAAT</u></p> <p><u>ACAATAAGCTAAAAAACTGACGTAGGATAAGCAA</u></p> <p><u>AACGGCGCAATTTGAGTTGTAACGTAATATTTTCA</u></p>	<p>gBlock used to construct pDGO186N-KS1, contains Clo1313_1194 promoter, repeat#1 and spacer#1</p> <p>promoter is underlined</p> <p>repeat#1 in bold</p>

	<p><u>CTAAAATAGTAATTATTTTCATGTTGTTTTTTTTTA</u></p> <p><u>GATTAATTTATAATATAATTTATTGTATAAGCAAT</u></p> <p><u>ATCTTAATTATCATTAAAGGGGGAAAAAACTGTT</u></p> <p><u>TGTATCGTACCTATGAGGAATTGAAAC<u>TTTTGGG</u></u></p> <p><u>AAGTGC GTGAAGCGGGAGAGCTTTACCG</u></p>	<p>spacer#1 in italics and underlined</p>
<p>KS1 gBlock2</p>	<p><u>TTTTGGGAAGTGC GTGAAGCGGGAGAGCTTTA</u></p> <p><u>CCGGTTTGTATCGTACCTATGAGGAATTGAAAC</u></p> <p><u>TTCTGGCTACTCCAAGTTCCTAATGTTATATATGGT</u></p> <p><u>TTTATAGTATAAAGCCTTAAGCGAAAATGAATAG</u></p> <p><u>GACTGTATAATAAGGTTAAGGCTTTATATTCGCCT</u></p> <p><u>GAAGCGTAAAAACGCAATAGGAAGAATAAAATTT</u></p> <p><u>TTGCCGGGGATAGGCCAGCCTTGCAAACTGGCTG</u></p> <p><u>ACAAGCGAAGTCTCCCCGCTTCCCCGGCAACAATT</u></p> <p><u>TCAAATAGGGAGACAGAAAGGGAGACGAAATCT</u></p> <p>CGTATGTTGTGTGGAATTGTGAGCGGATAACAAT</p> <p>TTCACACAGGAAACAGCTATGACCATGATTACGCC</p> <p>TAATAGAAATAATTTTGTTTAACTTTACAAACGGG</p> <p>ATTGACTTTTAAAAAAGGATTGATTCTAATGAAGA</p> <p>AAGCAGACAAGTAAGCCTCCTAAATCACTTTAGA</p>	<p>gBlock used to construct pDGO186N-KS1, contains spacer#1, repeat#2 and terminator</p> <p>spacer#1 in italics and underlined</p> <p>repeat#2 in bold</p> <p>terminator is underlined</p>

	TAAAAATTTAGGAGGCATATCAAATGGTGAATG	
KS2 gBlock1	<p><u>TTCAAGGATCC</u><u>TTAATATGCCGACCACGTTGCAAT</u></p> <p><u>TCCCGTCAAATAATGCATTTTGCAGCCGACGAAAC</u></p> <p><u>AGGCAAGATAACTGTATTGGCTATAAATGTTTCA</u></p> <p><u>GGCAGCGGTATATTTTGCCTCCCGGTAAAATTAAT</u></p> <p><u>ACAATAAGCTAAAAAACTGACGTAGGATAAGCAA</u></p> <p><u>AACGGCGCAATTTGAGTTGTAACGTAATATTTTCA</u></p> <p><u>CTAAAAATAGTAATTATTTTCATGTTGTTTTTTTTTA</u></p> <p><u>GATTAATTTATAATATAATTTATTGTATAAGCAAT</u></p> <p><u>ATCTAATTATCATTAAAGGGGGAAAAAACT</u>GTT</p> <p>TGTATCGTACCTATGAGGAATTGAAAC<u>GCGGAAG</u></p> <p><u>TGCGTGTGAAGCGGGAGAGCTTACCGTGCG</u></p>	<p>gBlock used to construct pDGO186N-KS2, contains Clo1313_1194 promoter, repeat#1 and spacer#2</p> <p>promoter is underlined</p> <p>repeat#1 in bold</p> <p>spacer#2 in italics and underlined</p>
KS2 gBlock2	<p><u>GGGAAGTGCGTGTGAAGCGGGAGAGCTTACCG</u></p> <p><u>TGCGGTTTGTATCGTACCTATGAGGAATTGAAAC</u></p> <p><u>TTTCTGGCTACTCCAAGTTCCTAATGTTATATATGG</u></p> <p><u>TTTTATAGTATAAAGCCTTAAGCGAAAATGAATAG</u></p> <p><u>GACTGTATAATAAGGTTAAGGCTTTATATTCGCCT</u></p> <p><u>GAAGCGTAAAAACGCAATAGGAAGAATAAAATTT</u></p>	<p>gBlock used to construct pDGO186N-KS2, contains spacer#2, repeat#2 and terminator</p> <p>spacer#2 in italics and underlined</p>

	<p><u>TTGCCGGGGATAGGCCAGCCTTGCAA</u>ACTGGCTG</p> <p><u>ACAAGCGAAGTCTCCCCGCTTCCCCGGCAACAATT</u></p> <p><u>TCAAATAGGGAGACAGAAAGGGAGACGAAATCT</u></p> <p>CGTATGTTGTGTGGAATTGTGAGCGGATAACAAT</p> <p>TTCACACAGGAAACAGCTATGACCATGATTACGCC</p> <p>TAATAGAAATAATTTTGTTTAACTTTACAAACGGG</p> <p>ATTGACTTTTAAAAAAGGATTGATTCTAATGAAGA</p> <p>AAGCAGACAAGTAAGCCTCCTAAATTCACCTTAGA</p> <p>TAAAATTTAGGAGGCATATCAAATGGTGAATG</p>	<p>repeat#2 in bold</p> <p>terminator is underlined</p>
<p>KS3</p> <p>gBlock1</p>	<p><u>TTCAAGGATCCTTAATATGCCGACCACGTTGCAAT</u></p> <p><u>TCCCGTCAAATAATGCATTTTGCAGCCGACGAAAC</u></p> <p><u>AGGCAAGATAACTGTATTGGCTATAAATGTTTCA</u></p> <p><u>GGCAGCGGTATATTTTGCCTCCCGGTAAAATTAAT</u></p> <p><u>ACAATAAGCTAAAAAACTGACGTAGGATAAGCAA</u></p> <p><u>AACGGCGCAATTTGAGTTGTAACGTAATATTTTCA</u></p> <p><u>CTAAAAATAGTAATTATTTTATGTTGTTTTTTTTTA</u></p> <p><u>GATTAATTTATAATATAATTTATTGTATAAGCAAT</u></p> <p><u>ATCTTAATTATCATTAAAGGGGGAAAAAACTGTT</u></p> <p>TGTATCGTACCTATGAGGAATTGAAAC<u>AGAATCG</u></p>	<p>gBlock used to construct pDGO186N-KS3, contains Clo1313_1194 promoter, repeat#1 and spacer#3</p> <p>promoter is underlined</p> <p>repeat#1 in bold</p> <p>spacer#3 in italics and underlined</p>

	<u>CACGGTAAAGCTCTCCCGCTTCACACGCAC</u>	
KS3gBlock2	<u>AGAATCGCACGGTAAAGCTCTCCCGCTTCACACGC</u> <u>ACGTTTGTATCGTACCTATGAGGAATTGAAACTT</u> <u>TCTGGCTACTCCAAGTTCCTAATGTTATATATGGTT</u> <u>TTATAGTATAAAGCCTTAAGCGAAAATGAATAGG</u> <u>ACTGTATAATAAGGTTAAGGCTTTATATTCGCCTG</u> <u>AAGCGTAAAAACGCAATAGGAAGAATAAAATTTT</u> <u>TGCCGGGGATAGGCCAGCCTTGCAAACCTGGCTGA</u> <u>CAAGCGAAGTCTCCCGCTTCCCGGCAACAATTT</u> <u>CAAATAGGGAGACAGAAAGGGAGACGAAATCTC</u> GTATGTTGTGTGGAATTGTGAGCGGATAACAATT TCACACAGGAAACAGCTATGACCATGATTACGCCT AATAGAAATAATTTTGTTTAACTTTACAAACGGGA TTGACTTTTAAAAAAGGATTGATTCTAATGAAGAA AGCAGACAAGTAAGCCTCCTAAATTCACTTTAGAT AAAAATTTAGGAGGCATATCAAATGGTGAATG	gBlock used to construct pDGO186N-KS3, contains spacer#3, repeat#2 and terminator spacer#3 in italics and underlined repeat#3 in bold terminator is underlined

666

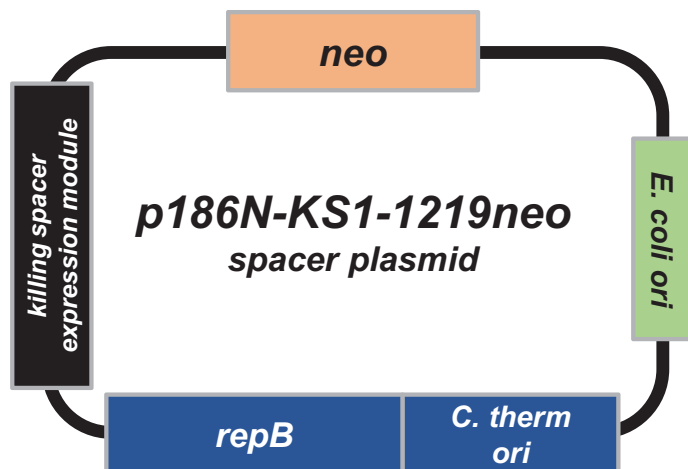
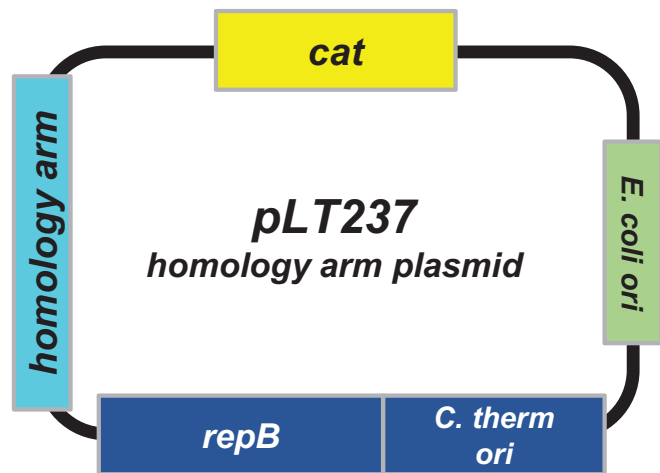
667

668 Table 3 HRM screening results of pooled plasmids

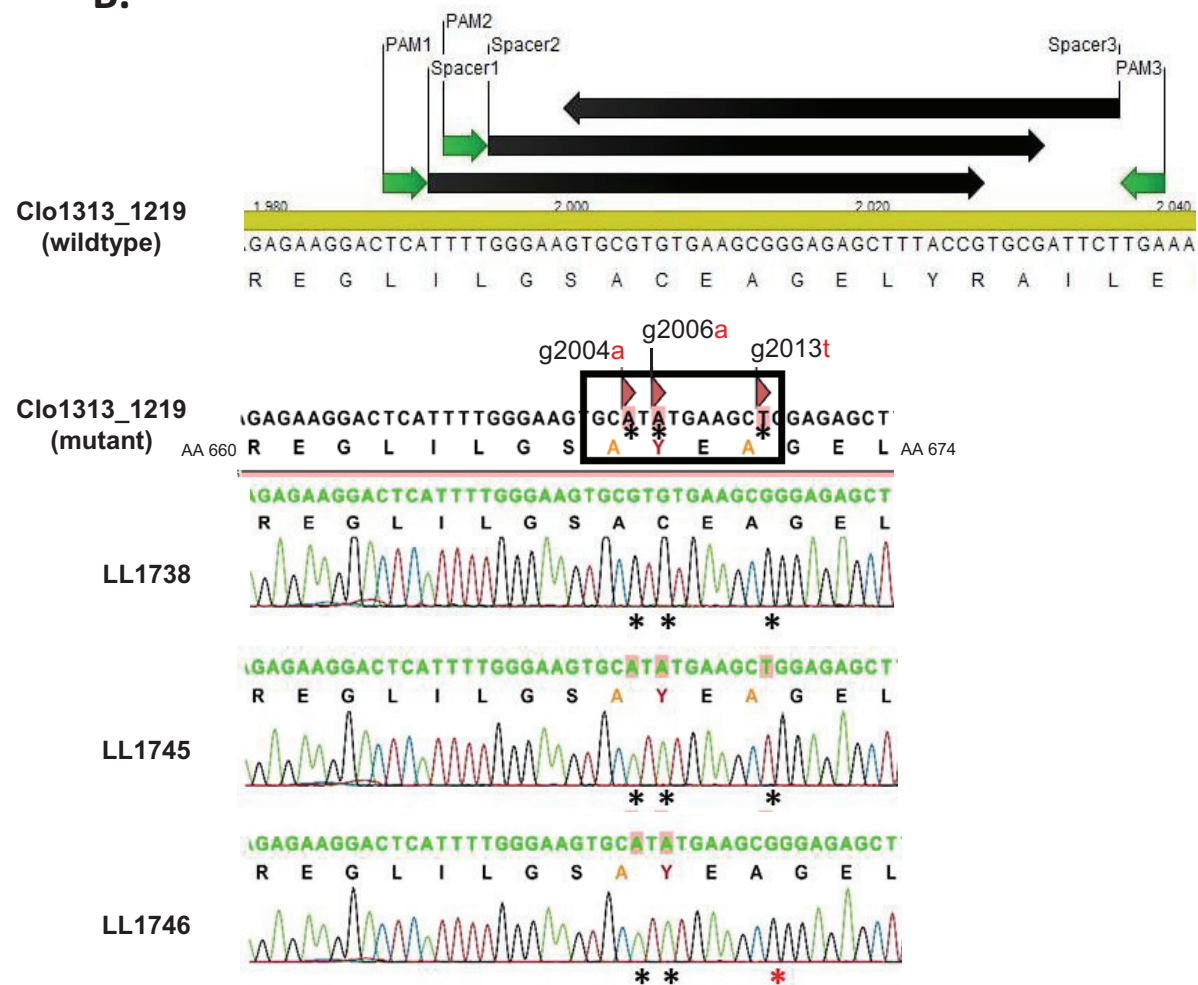
Screening step	Experiment 1	Experiment 2
Initial colonies	150	159
Positive by HRM qPCR	2	15
G2004A silent mutation present	2	15
<i>polC</i> mutation present	2	15
G2013T silent mutation present	2	14

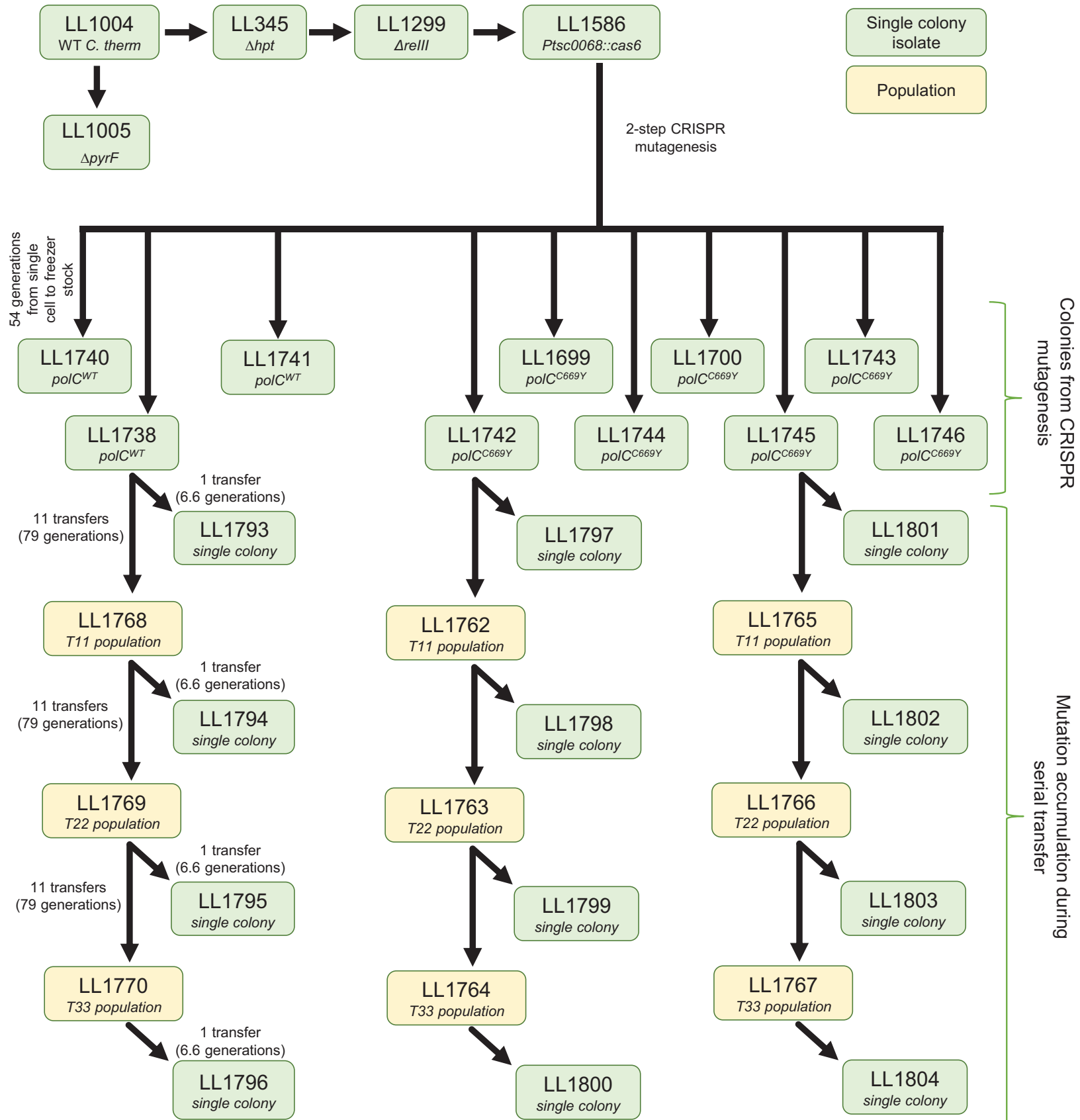
669

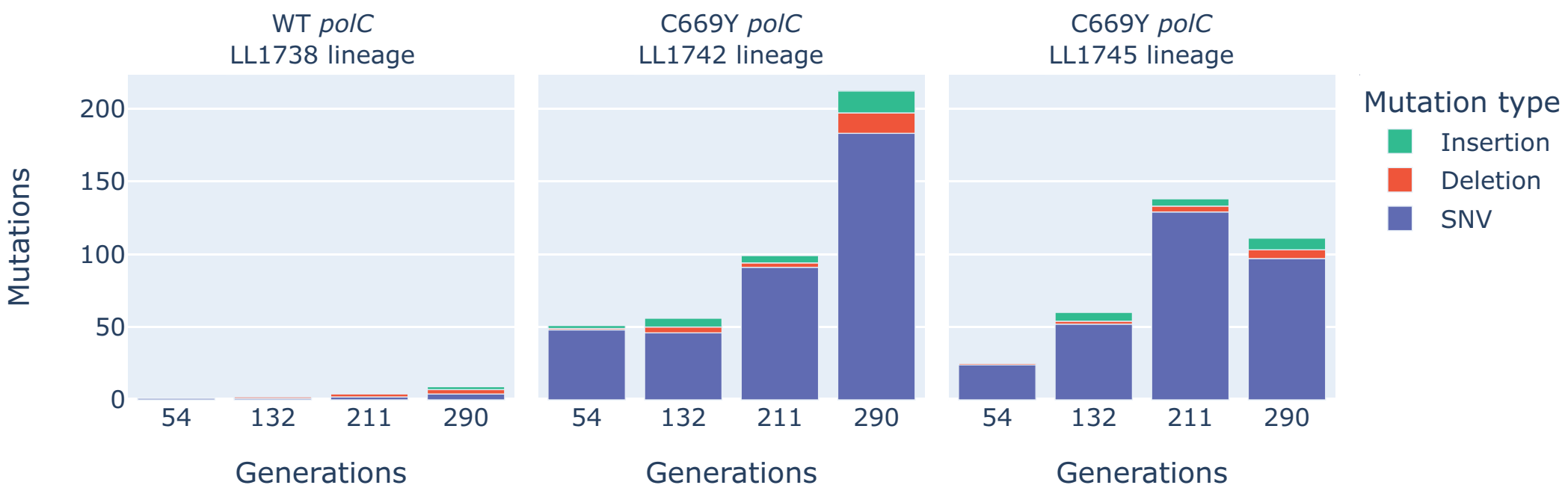
A.



B.

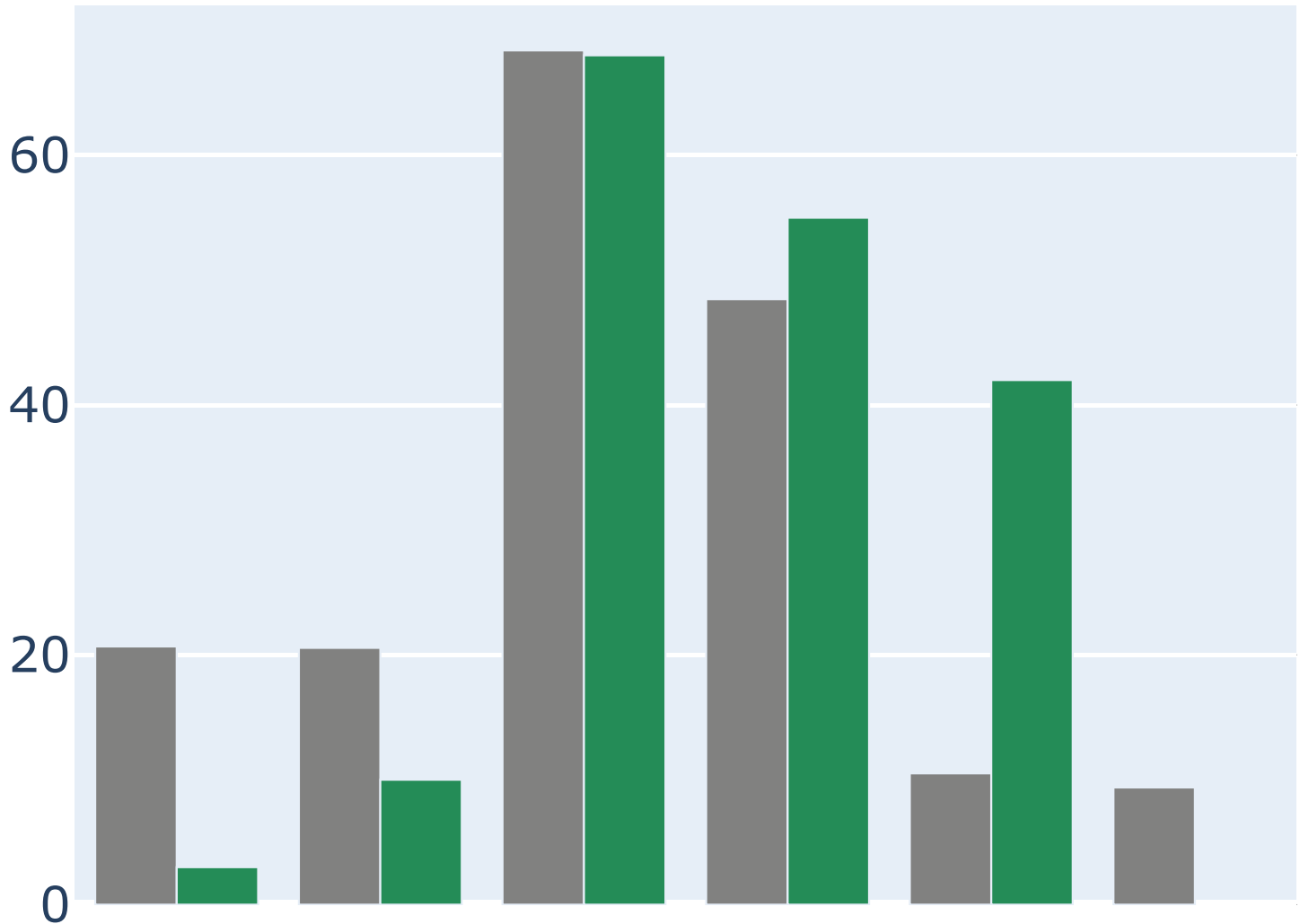






Number of synonymous SNVs

Expected
Observed



A:T to T:A transversion
A:T to C:G transversion
A:T to G:C transversion
C:G to T:A transition
C:G to A:T transition
C:G to G:C transversion

