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

Anthony Lanahan[1,2], Kamila Zakowicz[1], Liang Tian[1,2], Daniel G. Olson[1,2*] and Lee R. Lynd[1,2]

[1] Thayer School of Engineering at Dartmouth College, Hanover, NH 03755
[2] Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, 37830
*To whom correspondence should be addressed: daniel.g.olson@dartmouth.edu

Keywords

Whole genome sequencing, next-generation sequencing, 5-fluoroorotic acid, 5-FOA, mutation rate, *polC*, *dnaE*, *Hungateiclostridium thermocellum*, *Ruminiclostridium thermocellum*, *Acetivibrio thermocellus*

Abstract

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

Importance

Cellulosic biofuels are a promising approach to decarbonize the heavy duty transportation sector. A longstanding barrier to cost-effective cellulosic biofuel production is the recalcitrance of the material to solubilization. Native cellulose-consuming organisms, such as *Clostridium thermocellum*, are promising candidates for cellulosic biofuel production, however they often need to be genetically modified to improve product formation. One approach is adaptive laboratory evolution. Our findings demonstrate a way to increase the mutation rate in this

30   industrially-relevant organism, which can reduce the time needed for adaptive evolution
31   experiments.

Introduction

33   *Clostridium thermocellum* (aka *Acetivibrio thermocellus, Hungateiclostridium thermocellum,* and
34   *Ruminiclostridium thermocellum*) is a thermophilic, anaerobic bacterium that can ferment
35   crystalline cellulose to ethanol and has attracted interest as a candidate for cellulosic biofuel
36   production (Lynd et al. 2017). Its ability to deconstruct crystalline cellulose is mediated by a
37   protein complex called a cellulosome (Bayer et al. 2008), and this system may have applications
38   for deconstruction of other polymers, including plastics (Yan et al. 2021). In many cases,
39   however, native properties need to be improved for industrial application.
40
41   Adaptive laboratory evolution (ALE) is a commonly used strategy for improving desired
42   properties of strains, but experiments can take anywhere from weeks to years (Sandberg et al.
43   2019). Increasing the mutation rate of a strain can reduce the duration of ALE experiments.
44   Mutations in DNA polymerase III are known to affect the mutation rate of bacteria. DNA
45   polymerase III is responsible for replication of the bacterial genome. It comes in two major
46   forms, DnaE and PolC. The widely-studied *Escherichia coli* only has the DnaE-type enzyme,
47   and many groups have found mutator mutations in this gene (Vandewiele et al. 2002; Strauss et
48   al. 2000; Maki, Mo, and Sekiguchi 1991; Mo, Maki, and Sekiguchi 1991; Ruiz-Rubio and
49   Bridges 1987; Konrad 1978; Sevastopoulos and Glaser 1977). Typically, mutations that affect
50   the fidelity of the DNA Polymerase III holoenzyme are found in the polymerase domain of DnaE
51   or the separate epsilon proofreading subunit.
52
53   The PolC-type enzyme is found primarily in gram positive bacteria with low GC content, and has
54   received much less attention (Timinskas et al. 2014). Organisms with PolC typically do not have
55   a separate epsilon proofreading subunit, and instead rely on proofreading activity of the PHP
56   domain within the PolC protein. Several hypermutator mutations in the *polC* gene in Bacillus
57   subtilis have been identified (Barnes et al. 1992) (Gass and Cozzarelli 1973) (Paschalis et al.
58   2017). A previous ALE experiment identified a *polC* mutation among many mutations in a strain
59   of *C. thermocellum* with a hypermutator phenotype (Holwerda et al. 2020), however the
60   causality of the *polC* mutation was not verified, and the mutation rate was not determined.
61
62   In ALE experiments, identifying mutations is only the first step in strain improvement. Mutations
63   have to be subsequently re-introduced so that their effect can be characterized. Previously, it
64   has been difficult to reintroduce point mutations into *C. thermocellum*. Most examples required
65   the deletion of the wild type gene followed by re-introduction of the mutant gene (Zheng et al.
66   2015; Lo et al. 2015). This process is very time-consuming (~2 months per mutation), and
67   requires that deletion of the target gene is not toxic. Recently, we developed a new CRISPR-
68   based system for introducing point mutations in *C. thermocellum*, based on either the native
69   Type I or heterologous Type II CRISPR systems (Walker et al. 2020).
70

71 In this work, we use the newly-developed CRISPR tools to characterize the effect of a single
72 nucleotide mutation in *polC* on the mutation rate of *C. thermocellum*. The ability to both rapidly
73 create diversity with controllable mutator phenotypes, and characterize the resulting mutations
74 with CRISPR tools dramatically improves our ability to perform ALE on *C. thermocellum*.

Methods

76 Table 1. Strain and plasmids used in this work.
77

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 *polC* 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	*C. thermocellum* DSMZ 1313 wild type strain; WT polC, WT pyrF	NCBI reference sequence NC_017304.1	DSMZ culture collection
LL1005	LL1004 with *pyrF* deletion	not available	(Tripathi et al. 2010)
LL1299	LL1004 with *hpt* and *reIII* deletions	SRX2506395	(Walker et al. 2020)
LL1586	LL1299 with upregulated cas expression using *Tsac*_0068 promoter	SRX4823139	(Walker et al. 2020)
LL1699	Parent strain LL1586; mutant Pol III, WT pyrF	SRX8904888	This work
LL1700	Parent strain LL1586; mutant Pol III, WT pyrF	SRX8904887	This work
LL1738	LL1586, failed CRISPR mutagenesis at *polC* locus	SRX9642234	This work
LL1740	LL1586, failed CRISPR mutagenesis at *polC* locus	SRX9642235	This work

3

LL1741	LL1586, failed CRISPR mutagenesis at *polC* 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 population	SAMN20331617	This work
LL1770	LL1738 serial transfer #33 population	SAMN20331618	This work
LL1793	LL1738 serial transfer #1 single colony, 50 generations from original	SAMN20331619	This work

	CRISPR mutant cell		
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 *polC* mutant cell	SAMN20331623	This work
LL1798	LL1742 serial transfer #11 single colony, 128 generations from original *polC* mutant	SAMN20331624	This work
LL1799	LL1742 serial transfer #22 single colony, 207 generations from original *polC* mutant cell	SAMN20331625	This work
LL1800	LL1742 serial transfer #33 single colony, 286 generations from original *polC* mutant cell	SAMN20331626	This work
LL1801	LL1745 serial transfer #1 single colony, 50 generations from original *polC* mutant cell	SAMN20331627	This work
LL1802	LL1745 serial transfer #11 single colony, 128 generations from original *polC* mutant cell	SAMN20331628	This work
LL1803	LL1745 serial transfer #22 single colony, 207 generations from original *polC* mutant cell	SAMN20331629	This work
LL1804	LL1745 serial transfer #33 single colony, 286 generations from original *polC* mutant cell	SAMN20331630	This work

78   [a] Accession numbers that begin with "SRX" represent samples with WGS performed at JGI,

79   numbers that begin with "SAMN" represent samples with WGS performed at Dartmouth. Both

80   sets are available from the NCBI SRA database.

81

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

5

83

Name	Sequence (5' → 3')	Description
TY113+	ACCTTGATGACACAGAAGAAGGCGATTTGA	Plasmid pLT237 confirmation
TY114+	CTAAAAGTCGTTTGTTGGTTCA	Plasmid pLT237 confirmation
HRM F	CGAGACTTCCAAGAAAGCTTTTAAT	HRM primer
HRM R	GTCCTCATCTTTGTTTTCAAGAATC	HRM primer
1219F	GGAAATTGGTGCGGTAAAGA	PolC HA PCR primer
1219R	CTATCCCTCTTTCGTCAA	PolC HA PCR primer
1219SF	CGTAGGCCTTAGAAATCTGTA	PolC sequencing primer
1219SR	CCGGTATGGGAAGTATGT	PolC sequencing primer
1218F	CCAGGAAAGGGCAGTGGAAGA	PolC PCR/sequencing primer
1220R	TCAAATTTCATAGATTCCCAA	PolC PCR/sequencing primer
LT651	TAATACCTGCAAAGACCC	PolC sequencing primer
LT652	AAGGACAGAAGCAAGGGA	PolC sequencing primer
LT652R	TCCCTTGCTTCTGTCCTT	PolC sequencing primer
1219FR	TCTTTACCGCACCAATTTCC	PolC sequencing primer
1219SRF	ACATACTTCCCATACCGG	PolC sequencing primer
1219RF	TTGACGAAAGAGGGATAG	PolC sequencing primer
LT656R	GGATGTAAGAAAAGGAAAGG	PolC sequencing primer
BackboneR1	GGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATT	HA plasmid backbone PCR
BB R2 intF	ACCGCCTTTGAGTGAGCTGATACCG	HA plasmid backbone PCR
BackboneF	TAATAGAAATAATTTTGTTTAACTTTACAAACGGGATTGACTTTTAAAAAAGGATTGATT	HA plasmid backbone PCR
BB F intR	TGCGGCGAGCGGTATCAGCTCACTC	HA plasmid backbone PCR
Backbone Reverse	GGTCGGCATATTAAGGATCCTTGAACTACTCTTTAATAAAATAAT	Killing plasmid backbone PCR

TY1C F	AAATTTAGGAGGCATATCAAATGGTGAATGGACCAAT AATAATGACTAGAGAAGAAAGAA	Killing plasmid backbone PCR
HA gBlock	ACAGGAAACAGCTATGACCATGATTACGCCGACAATC CTGTAATTGACACTTTGGAACTTTCAAGGCAAATGTTT CCGGAGCTTAAAAAACACAAACTGGATGTGGTTGCAA AGCATCTTGGTGTTTCATTGGAAAACCATCACAGGGCT CTTGACGATGCAAAGGCCTGTGGGGAGATCTTTATCA AATGCCTTGAGATACTGGCTGAAAAGAATGTAAAAAC AATTGATGACATACAAAATGTTTTTGAGGGTTGTTGGA ATTATCAGAAAGCAAATTCATACCATGCCATTATACTT GTAAAAAATTACGTAGGCCTTAGAAATCTGTACAAGAT TGTGTCCAAGACTCATTTGGAGTATTTTCACAAAAGGC CGAGACTTCCAAGAAAGCTTTTAATGATGCACAGAGA AGGACTCATTTTGGGAAGTGCGTaTGAAGCGGGAGAG CTTTACCGTGCGATTCTTGAAAACAAAGATGAGGACG AAATTTCAAGGATTGTAAAGTTTTACGATTATCTTGAA ATACAGCCTTTGGGCAACAACCGTTTCCTTGTCGAGAG CGGAAAGGTGAACAGTGAGGAAGACTTAAAGAATAT AAACAGAAAAATAGTAGCTTTGGGTGAGAAGTACAAC AAGCCGGTGGTCGCTACCTGCGATGTTCATTTCATGGA TCCGCAGGACGAAGTTTTCAGAAGAATACTTATGGCG GGGCAAGGTTTTTCCGATGCGGACAACCAGGCACCGT TGTATTTCAGGACCACTGATGAAATGCTGGAAGAGTTT GAATATCTGGGCAGGGAAAAATGCTATGAAGTGGTTG TAACAAATACAAATTTAATTGCGGATATGTGCGAGGA CATACTTCCCATACCGGAAGGAACTTTTCCTCCTAAAAT TGAAGGTGCGGAAGAAGAAATAAGGATGCTTGCCGA AAACAAGGCGAGAGAAATTTATGGCGACCCTCTTCCT GAGATTGTGGAAAAGCGCATGGAAAAGGAGCTTAACT CAATAATAAAAAATGGCTTCTCCGTTATGTATATTATTG CGCAAAAACTTGTATGGAAATCTTTAAGTGACGGCTAT CTTGTAGGGTCGAGGGGATCCGTCGGATCATCCTTTGT TGCCTATAATAGAAATAATTTTGTTT	PolC mutant homology arm gBlock used to construct pLT237
KS1 gBlock1	TTCAAGGATCCTTAATATGCCGACCACGTTGCAATTCC CGTCAAATAATGCATTTTGCAGCCGACGAAACAGGCA AGATAACTGTATTGGCTATAAATGTTTCAGGCAGCGGT ATATTTTGCCTCCCGGTAAAATTAATACAATAAGCTAA AAAACTGACGTAGGATAAGCAAAACGGCGCAATTTGA GTTGTAACGTAATATTTTCACTAAAAATAGTAATTATTT CATGTTGTTTTTTTTTTAGATTAATTTATAATATAATTTAT	gBlock used to construct pDGO186N-KS1, contains Clo1313_1194 promoter, repeat#1 and spacer#1

	TGTATAAGCAATATCTTAATTATCATTAAAGGGGGAAA AAAACTGTTTGTATCGTACCTATGAGGAATTGAAACTT TTGGGAAGTGCGTGTGAAGCGGGAGAGCTTTACCG	
KS1 gBlock2	TTTTGGGAAGTGCGTGTGAAGCGGGAGAGCTTTACCG GTTTGTATCGTACCTATGAGGAATTGAAACTTTCTGGC TACTCCAAGTTCCTAATGTTATATATGGTTTTATAGTAT AAAGCCTTAAGCGAAAATGAATAGGACTGTATAATAA GGTTAAGGCTTTATATTCGCCTGAAGCGTAAAAACGCA ATAGGAAGAATAAAATTTTTGCCGGGGATAGGCCAGC CTTGCAAACTGGCTGACAAGCGAAGTCTCCCCGCTTCC CCGGCAACAATTTCAAATAGGGAGACAGAAAGGGAG ACGAAATCTCGTATGTTGTGTGGAATTGTGAGCGGAT AACAATTTCACACAGGAAACAGCTATGACCATGATTAC GCCTAATAGAAATAATTTTGTTTAACTTTACAAACGGG ATTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAAG CAGACAAGTAAGCCTCCTAAATTCACTTTAGATAAAAA TTTAGGAGGCATATCAAATGGTGAATG	gBlock used to construct pDGO186N-KS1, contains spacer#1, repeat#2 and terminator
KS2 gBlock1	TTCAAGGATCCTTAATATGCCGACCACGTTGCAATTCC CGTCAAATAATGCATTTTGCAGCCGACGAAACAGGCA AGATAACTGTATTGGCTATAAATGTTTCAGGCAGCGGT ATATTTTGCCTCCCGGTAAAATTAATACAATAAGCTAA AAAACTGACGTAGGATAAGCAAAACGGCGCAATTTGA GTTGTAACGTAATATTTTCACTAAAAATAGTAATTATTT CATGTTGTTTTTTTTTAGATTAATTTATAATATAATTTAT TGTATAAGCAATATCTTAATTATCATTAAAGGGGGAAA AAAACTGTTTGTATCGTACCTATGAGGAATTGAAACGG GAAGTGCGTGTGAAGCGGGAGAGCTTTACCGTGCG	gBlock used to construct pDGO186N-KS2, contains Clo1313_1194 promoter, repeat#1 and spacer#2
KS2 gBlock2	GGGAAGTGCGTGTGAAGCGGGAGAGCTTTACCGTGC GGTTTGTATCGTACCTATGAGGAATTGAAACTTTCTGG CTACTCCAAGTTCCTAATGTTATATATGGTTTTATAGTA TAAAGCCTTAAGCGAAAATGAATAGGACTGTATAATA AGGTTAAGGCTTTATATTCGCCTGAAGCGTAAAAACGC AATAGGAAGAATAAAATTTTTGCCGGGGATAGGCCAG CCTTGCAAACTGGCTGACAAGCGAAGTCTCCCCGCTTC CCCGGCAACAATTTCAAATAGGGAGACAGAAAGGGA GACGAAATCTCGTATGTTGTGTGGAATTGTGAGCGGA TAACAATTTCACACAGGAAACAGCTATGACCATGATTA CGCCTAATAGAAATAATTTTGTTTAACTTTACAAACGG GATTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAA	gBlock used to construct pDGO186N-KS2, contains spacer#2, repeat#2 and terminator

	GCAGACAAGTAAGCCTCCTAAATTCACTTTAGATAAAA ATTTAGGAGGCATATCAAATGGTGAATG	
KS3 gBlock1	TTCAAGGATCCTTAATATGCCGACCACGTTGCAATTCC CGTCAAATAATGCATTTTGCAGCCGACGAAACAGGCA AGATAACTGTATTGGCTATAAATGTTTCAGGCAGCGGT ATATTTTGCCTCCCGGTAAAATTAATACAATAAGCTAA AAAACTGACGTAGGATAAGCAAAACGGCGCAATTTGA GTTGTAACGTAATATTTTCACTAAAAATAGTAATTATTT CATGTTGTTTTTTTTTAGATTAATTTATAATATAATTTAT TGTATAAGCAATATCTTAATTATCATTAAAGGGGGAAA AAAACTGTTTGTATCGTACCTATGAGGAATTGAAACAG AATCGCACGGTAAAGCTCTCCCGCTTCACACGCAC	gBlock used to construct pDGO186N-KS3, contains Clo1313_1194 promoter, repeat#1 and spacer#3
KS3gBlock2	AGAATCGCACGGTAAAGCTCTCCCGCTTCACACGCACG TTTGTATCGTACCTATGAGGAATTGAAACTTTCTGGCT ACTCCAAGTTCCTAATGTTATATATGGTTTTATAGTATA AAGCCTTAAGCGAAAATGAATAGGACTGTATAATAAG GTTAAGGCTTTATATTCGCCTGAAGCGTAAAAACGCAA TAGGAAGAATAAAATTTTTGCCGGGGATAGGCCAGCC TTGCAAACTGGCTGACAAGCGAAGTCTCCCCGCTTCCC CGGCAACAATTTCAAATAGGGAGACAGAAAGGGAGA CGAAATCTCGTATGTTGTGTGGAATTGTGAGCGGATA ACAATTTCACACAGGAAACAGCTATGACCATGATTACG CCTAATAGAAATAATTTTGTTTAACTTTACAAACGGGA TTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAAGC AGACAAGTAAGCCTCCTAAATTCACTTTAGATAAAAAT TTAGGAGGCATATCAAATGGTGAATG	gBlock used to construct pDGO186N-KS3, contains spacer#3, repeat#2 and terminator

84

Plasmid construction

86   Plasmid construction was performed by isothermal DNA assembly (Gibson 2011), using
87   NEBuilder HiFi DNA Assembly Cloning Kit from NEB. Synthetic DNA was purchased from IDT
88   as gBlocks (Integrated DNA Technologies, Coralville, IA).

Growth conditions

Routine cultivation was performed using either CTFUD rich medium or MTC-5 chemically defined medium (Olson and Lynd 2012). For solid medium, agar was used at a concentration of 0.8%. Selection for the *cat* marker was performed with 6 ug/ml thiamphenicol. Stock solutions of thiamphenicol were prepared in DMSO at 1000x concentration of 6mg/ml (Sigma, part number:T0251-5G). Selection for the *neo* marker was performed with 150 ug/ml neomycin, stock solution 50 mg/ml in water (Gibco, part number: 21810-031). Selection for *pyrF* mutations was performed with 5-fluoroorotic acid (5-FOA, Zymo Research, part number F9001-5) at a final concentration of 0.5 mg/ml. A 200x 5-FOA stock solution was prepared fresh daily by dissolving 100 mg 5-FOA in 1 ml DMSO. When grown on defined medium, strains with *pyrF* mutations were supplemented with 40 ug/ml uracil (Tripathi et al. 2010), a stock solution of 40 mg/ml uracil was prepared in 1N NaOH (Sigma, part number: U7050-5g).

Two-step CRISPR approach for introducing mutations

A two-step CRISPR Type I-B approach was used to introduce the *polC* mutation, based on our previously described approach (Walker et al. 2020) with a few modifications. In the two-step approach, first the homology arm plasmid is introduced (primary transformation) and cells are grown to allow homologous recombination to occur. Then a secondary transformation is performed to introduce a killing plasmid to eliminate cells that have not undergone homologous recombination.

Primary transformation with homology template. Cultures of parent strain LL1586 (native Type I-B CRISPR system upregulated by insertion of Tsac_0068 promoter) were grown to mid log phase in 50 ml of CTFUD rich medium at 55°C. The culture was centrifuged, rinsed twice with H_2O, resuspended in H_2O, transformed using electroporation with plasmid pLT237 containing the *polC* C669Y homology repair template. Transformation was performed using electroporation with a square pulse. The amplitude was 1500 V. A single pulse with a duration of 1.5 ms was applied to a 1 mm cuvette. Cells were removed from the electroporation cuvette and allowed to recover for 16 hours in 2 mL CTFUD at 50℃ overnight (Olson and Lynd 2012). The sample was plated on CTFUD/TM (CTFUD medium with 6 ug/ml thiamphenicol) and recovered for 18-24 hours at 55°C. After 5 days, 10 colonies were picked and pooled into 1 mL CTFUD/TM, and incubated overnight at 55℃. The pooled colonies were subcultured twice to provide an opportunity for the homologous recombination events that are selected for in the secondary transformation (Walker et al. 2020). A 1:20 dilution (50 ul into 1 ml) of the pooled colony culture was prepared in CTFUD/TM medium and grown at 55℃ overnight for the first transfer. This was repeated for a second transfer. PCR using primers (TY 113+/114+) was used to confirm that plasmid pLT237 was present in the primary culture of 10 pooled colonies, the first transfer culture, and the second transfer culture. All three cultures were all stored at -80℃.

Secondary Transformation of LL1586/pLT237 cells with plasmids containing spacers targeting the wildtype *polC* gene. The first and second transfer cultures generated after the

10

129   primary transformation were pooled and 80 uL of the mixture was inoculated into 10 mL of
130   CTFUD and was grown overnight with and without thiamphenicol (6 ug/mL) to determine the
131   effect of growing the secondary transformation with and without antibiotic selection (note:
132   transformation worked better with added antibiotic at this step, see the results section). These
133   cultures were diluted 1:10 and grown (+/- thiamphenicol) to mid-log phase (Abs$_{600}$ between 0.5
134   and 0.8), harvested, and transformed as described above. The harvested cells were
135   transformed both individually with plasmids pDGO186N-KS1, pDGO186N-KS2, pDGO186N-
136   KS3, as well as with a pool of the three plasmids, and with a no-DNA control. The plasmids
137   contained a killing spacer (KS1, KS2, or KS3) as well as a neomycin (*neo*) selection marker.
138   After recovery, a portion of the cells were plated on CTFUD/NEO (150 ug/mL neomycin) and
139   incubated at 55℃. Colonies usually appeared after 3 days, and were picked 1-2 days later,
140   resuspended in 0.5 mL of CTFUD/NEO and grown overnight at 55℃. After neomycin selection,
141   colonies from secondary transformations were screened for the 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 primers that
144   flanked the mutation site (Table 2). For the reaction mixture, 2 uL of bacterial culture was mixed
145   with 10 uL 2X Sso Fast Evagreen qPCR mixture (BioRad USA). The primers were mixed into an
146   equal forward/reverse (F/R) primer mixture and serially diluted to 5 uM. 2 uL of the F/R primer
147   mixture was added along with 6 uL deionized water. The mixture was set to PCR conditions of
148   95℃/5 min 40X, 95℃/15 s, 55℃/1 min. The range of the melt curve was set to 65℃-95℃ at a rate
149   of 0.2℃/10 s. The uAnalyze v2 software (Dwight, Palais, and Wittwer 2012) (Atmadjaja et al.
150   2019) was used to analyze the raw fluorescence data by making normalized, derivative, and
151   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 media,
154   grown for 6-8 hours to mid-log phase (to ensure a rapidly-dividing culture, which is optimal for 5-
155   FOA selection), and 100 ul was plated at various dilutions with and without 5-FOA to ensure that
156   there were between 10-200 colonies on each plate. Usually the 10^{-4}, 10^{-5} and 10^{-6} dilutions had
157   a countable number of colonies for the 5-FOA plates. The plates were incubated at 55℃ for 4-6
158   days, and colonies were counted to determine the fraction of cells exhibiting 5-FOA resistance.

159   # Mutation accumulation experiment

160   To determine the mutation rate by mutation accumulation, cultures were serially transferred. In
161   bacterial cells, single colony isolation events usually correspond to about 23-25 generations
162   (Trindade, Perfeito, and Gordo 2010; Kibota and Lynch 1996). For *C. thermocellum*, based on
163   cell volume (0.4 um diameter by 2 um long (Sato et al. 1993)), and colony volume (lenticular
164   shape approximately 1 mm in diameter by 0.4 mm in height), a colony should contain about 5e8
165   cells, corresponding to 29 generations. Approximately another 18 generations (1:100 dilution
166   from initial culture of colony pick and 1:2500 dilution subculture for freezer stock preparation)
167   occurred between the initial colony isolation following the introduction of the *polC* mutation and
168   the preparation of the freezer stock. This does not affect the mutations observed in the starting

169   strains (L1699, LL1700, LL1738, LL1740-LL1746), but needs to be considered for the

170   subsequent single colony isolations. Each serial transfer consisted of a 1:100 dilution (~6.6

171   generations). After 10 1:100 serial transfers, the 11th transfer was a 1:5000 dilution (~12.3

172   generations) to provide extra volume for preparing freezer stocks and gDNA for WGS. This was

173   repeated three times. The 11th, 22nd, and 33rd transfers were stored as populations in the

174   freezer, and the presence of the *polC* mutation was confirmed by HRM qPCR.

175

176   Single colonies were isolated from each population to create a population bottleneck to fix

177   mutations. This involved another 1:100 subculture (~6.6 generations), followed by plating on

178   solid medium. A single colony was picked, grown up, and prepared for whole genome

179   sequencing (WGS).

180 Whole genome resequencing (WGS) at Dartmouth

181   Genomic DNA was prepared using the Omega E.Z.N.A. kit following the manufacturer's protocol

182   (Omega Bio-Tek, GA, USA). 500 ng of DNA was used for WGS library preparation using the

183   NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA).

184   Fractionated, adapter ligated DNA fragments went through 5 rounds of PCR amplification and

185   purification. The resulting WGS library was sequenced at the Genomics and Molecular Biology

186   Shared Resource (GMBSR) at Dartmouth. Libraries were diluted to 4 nM, pooled and loaded at

187   1.8 pM onto a NextSeq500 Mid Output flow cell, targeting 130 million 2x150 bp reads/sample.

188   Base-calling was performed on-instrument using RTA2 and bcls converted to fastq files using

189   bcl2fastq2 v2.20.0.422.

190 Whole genome resequencing (WGS) at JGI

191   Genomic DNA was submitted to the Joint Genome Institute (JGI) for sequencing with an

192   Illumina MiSeq instrument. Paired-end reads were generated, with an average read length of

193   150 bp and paired distance of 500 bp. Unamplified libraries were generated using a modified

194   version of Illumina's standard protocol. 100 ng of DNA was sheared to 500 bp using a focused

195   ultrasonicator (Covaris). The sheared DNA fragments were size selected using SPRI beads

196   (Beckman Coulter). The selected fragments were then end repaired, A-tailed and ligated to

197   Illumina compatible adapters (IDT, Inc) using KAPA Illumina library creation kit (KAPA

198   biosystems). Libraries were quantified using KAPA Biosystem's next-generation sequencing

199   library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified

200   libraries were then multiplexed into pools for sequencing. The pools were loaded and

201   sequenced on the Illumina MiSeq sequencing platform utilizing a MiSeq Reagent Kit v2 (300

202   cycle) following a 2 × 150 indexed run recipe.

203 WGS data analysis

204   Read data was analyzed with the CLC Genomic Workbench version 12 (Qiagen Inc., Hilden,

205   Germany). First, reads were trimmed using a quality limit of 0.05 and ambiguity limit of 2. Then

206   2.5M reads were randomly selected (to avoid errors due to differences in the total number of

207 reads). Reads were mapped to the reference genome (NC_017992). Mapping was improved by
208 two rounds of local realignment. The CLC Basic Variant Detection algorithm was used to
209 determine small mutations (single and multiple nucleotide polymorphisms, short insertions and
210 short deletions). Variants occurring in less than 35 % of the reads or fewer than 4 reads were
211 filtered out. The fraction of the reads containing the mutation is presented in XXX . To determine
212 larger mutations, the CLC InDel and Structural Variant algorithm was run. This tool analyzes
213 unaligned ends of reads and annotates regions where a structural variation may have occurred,
214 which are called breakpoints. Since the read length averaged 150 bp and the minimum mapping
215 fraction was 0.5, a breakpoint can have up to 75 bp of sequence data. The resulting break-
216 points were filtered to eliminate those with fewer than ten reads or less than 20 % "not perfectly
217 matched." The breakpoint sequence was searched with the Basic Local Alignment Search Tool
218 (BLAST) algorithm (Altschul et al. 1990) for similarity to known sequences. Pairs of matching left
219 and right breakpoints were considered evidence for structural variations such as transposon
220 insertions and gene deletions. The fraction of the reads supporting the mutation (left and right
221 breakpoints averaged) is presented in Supplementary Table 1. Mutation data from CLC was
222 further processed using custom Python scripts (https://github.com/danolson1/cth-mutation).

Sanger sequencing

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

Quantification of mutation rate

227 The mutation rate was determined based on synonymous mutations, which are generally
228 assumed to have a low effect on fitness (M Kimura 1968). *C. thermocellum* has a genome size
229 of 3,561,619 bp (NC_017304.1) of which 83% consists of coding regions. For each codon, the
230 number of synonymous single-substitution events were counted, and multiplied by the codon
231 frequency (Kazusa Codon Usage Database) (Nakamura and Tabata 1997), to reveal that 21.4%
232 of all nucleotide positions in *C. thermocellum* coding sequences allow a synonymous mutation.
233 This results in an effective genome size of 632,024 bp. The mutation rate ($\square$, mutations per
234 base pair per generation) is calculated with the following equation:

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

235 Where m is the number of observed mutations, n is the number of sites analyzed on the
236 genome, and T is the number of generations (Kucukyildirim et al. 2020; Lynch et al. 2016).

Results and Discussion

CRISPR system successfully introduces point mutations

239 Initially, we attempted to re-introduce the C669Y mutation into the *polC* gene using standard
240 homologous recombination techniques (Olson and Lynd 2012). This required cloning the entire

13

241   *polC* gene to ensure that a functional copy was present at each stage of the chromosomal

242   modification. Despite repeated attempts, we were unable to construct the deletion vector due to

243   apparent toxicity in *E. coli*.

244

245   We thus pursued an alternative approach, using a recently-developed chromosomal

246   modification technique that co-opts the native Type I CRISPR system in *C. thermocellum*

247   (Walker et al. 2020). This system involves two transformation events (Fig. 1). The first

248   transformation introduces the homology repair template, which introduces the desired point

249   mutation, as well as several silent mutations to prevent spacer recognition. The second

250   transformation introduces the killing spacer module, which targets chromosomes with the wild

251   type *polC* sequence, but not ones modified by the homology repair template. Colonies were

252   screened for the presence of the *polC* mutation using an HRM PCR assay (Supplementary

253   Figure 1), which identifies mutations by a decrease in the melting temperature of mutant PCR

254   amplicons. After transformation with the killing spacer plasmid, colonies were picked, and the

255   *polC* region was analyzed by HRM qPCR. A total of 309 colonies were picked for HRM qPCR

256   screening and 17 mutant candidates were identified (Table 3), Sanger DNA sequencing

257   confirmed the presence of the *polC* mutation in all 17 candidates. Performing the secondary

258   transformation with cells grown in the presence of thiamphenicol resulted in 5-fold higher

259   transformation efficiency. Mutations were subsequently confirmed by Sanger sequencing and

260   whole-genome sequencing, and 15 of 159 colonies (9.4%) had the correct mutant genotype. In

261   most strains with the *polC* mutation, both silent mutations were also present. One strain

262   (LL1746) was missing one of the silent mutations, indicating that both silent mutations are not

263   necessary to prevent CRISPR targeting.

264

265   Table 3 HRM screening results

266

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

267

268

269   The genome editing work was performed at the same time as some of the experiments in our

270   previous publication on CRISPR-based editing in *C. thermocellum* (Walker et al. 2020), and

271   thus does not include several improvements described in that work, such as the incorporation of

272   thermostable recombinases (*exo* and *beta* genes from *Acidithiobacillus caldus*).

273

14

274   To eliminate CRISPR-mediated restriction, we opted to make two silent mutations in the spacer
275   region, rather than disrupt the PAM sequence (Fig. 1). A benefit of this approach is that we can
276   use a single homology arm plasmid (first transformation) with several killing spacer plasmids
277   (second transformation), however the editing efficiency is slightly lower than what we previously
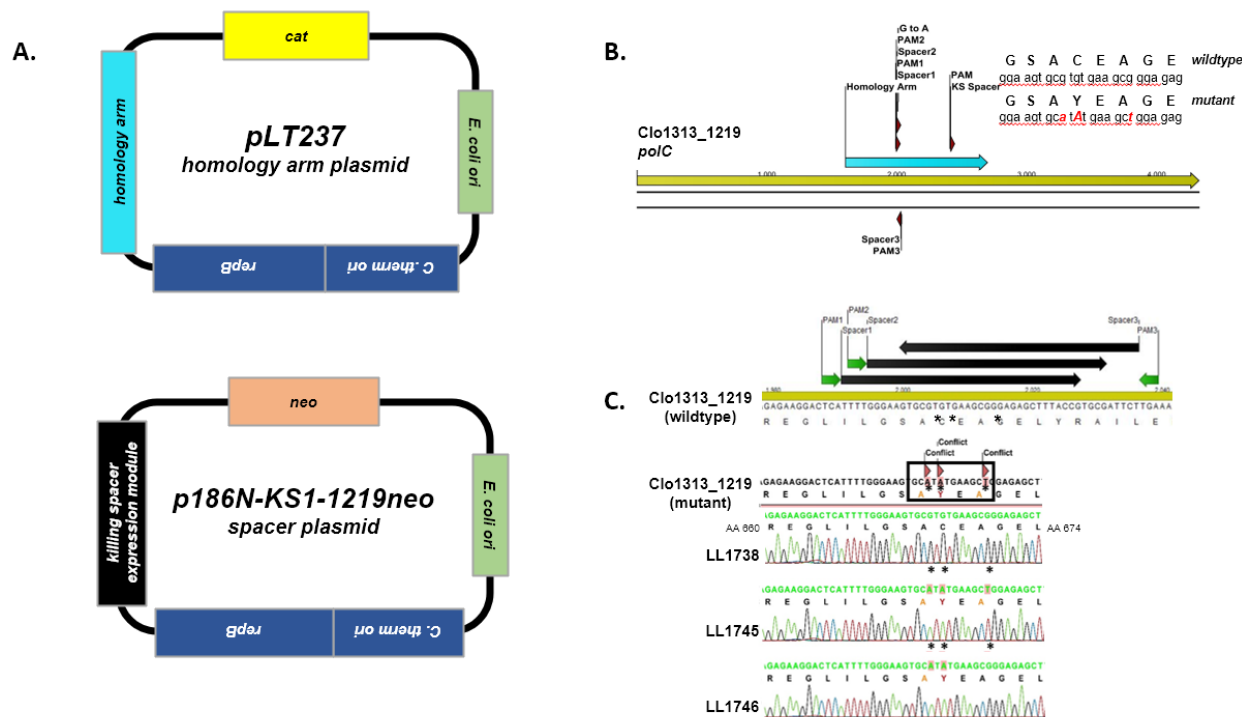278   reported (10% vs 40%) (Walker et al. 2020).

279
280   **Figure 1. A) 2-step CRISPR system for introducing mutations in *C. thermocellum*.** Step 1 is
281   transformation with a homology arm plasmid, and step 2 is transformation with a killing spacer cassette
282   plasmid. Each plasmid contains a *C. thermocellum* origin (including the *repB* replication protein), an *E.*
283   *coli* origin, and either a chloramphenicol (*cat*) or a neomycin (*neo*) selection marker. **B) The *polC* region**
284   **targeted for mutations.** The homology arm is shown at the start of the target sequence. The spacer
285   cassette is composed of three spacers oriented in the forward (KS1 or KS2) or reverse (KS3) directions.
286   Spacer regions were chosen to be immediately downstream of a TTN or TCD PAM sequence. The three
287   target mutations are also shown: two silent mutations (to disrupt CRISPR targeting) and the targeted
288   point mutation to create a cysteine to tyrosine amino acid change at position 669. **C) Confirmation of**
289   **mutagenesis.** The *polC* mutation targets were analyzed by Sanger sequencing. Strain LL1738 is an
290   example of failed mutagenesis, and displays a wild-type sequence. Strain LL1745 is an example of
291   successful mutagenesis, exhibiting all three targeted mutations. In strain LL1746, the *polC* mutation was
292   introduced successfully with only one of the two silent mutations.
293
294   The effect of the PolC[C669Y] mutation was determined with a 5-FOA resistance assay (Fig. 2). In
295   rich media, the *de novo* pyrimidine biosynthesis pathway is non-essential, and it is thus a
296   neutral site where mutations can accumulate. Mutations that inactivate this pathway create a 5-
297   FOA resistant (Foa[r]) phenotype that can readily be detected, and this assay is frequently used
298   as a measure of mutation rates (Boeke, La Croute, and Fink 1984; Grogan and Gunsalus 1993;
299   Kondo, Yamagishi, and Oshima 1991; Jacobs and Grogan 1997). In this experiment, we
300   observed a natural abundance of the Foa[r] phenotype at about 2% of WT cells, similar to what

301 we have reported previously for *C. thermocellum* (Tripathi et al. 2010). Disruption of *pyrF* largely
302 eliminates sensitivity to 5-FOA, with more than 70% of colonies exhibiting the Foa^r phenotype
303 (the reason this number is not 100% is likely due to a decrease in plating efficiency in the
304 selective condition). Most of the strains harboring the PolC^{C669Y} mutation also show an increase
305 in the Foa^r phenotype, however this was not universally observed. Strain LL1700 has the
306 PolC^{C669Y} mutation, but is Foa^s, and strain LL1740 does not have the PolC^{C669Y} mutation, but is
307 Foa^r.
308
309 In order to calculate the mutation rate, the target size needs to be known. We expected to
310 observe mutations primarily in the *pyrF* gene, however Sanger sequencing did not reveal the
311 expected mutations at that locus. Initially we plated on CTFUD with 5-FOA, picked 12 LL1299
312 (WT) colonies and sequenced the *pyrF* gene, but did not find any mutations. We sequenced the
313 *pyrF* gene from 40 LL1699 (*polC* mutant) colonies picked from 5-FOA plates, and only found
314 one mutation (a single nucleotide insertion resulting in a frame shift). Since *pyrF* mutants exhibit
315 a growth defect that can be complemented with added uracil (Tripathi et al. 2010), we repeated
316 the LL1299 plating experiment on CTFUD with 5-FOA and 40 ug/ml uracil. Of 18 colonies, 5 had
317 mutations (mostly frame-shifts and premature stop codons) at the *pyrF* locus.
318
319 Subsequent whole-genome sequencing (Supplementary Table 1) did not identify any mutations
320 associated with *de novo* pyrimidine biosynthesis, and the genetic basis for the Foa^r phenotype
321 in these strains remains unknown, which prevents an accurate determination of the mutation
322 rate. Instead, we decided to measure mutation accumulation directly, by whole-genome
323 sequencing.
324

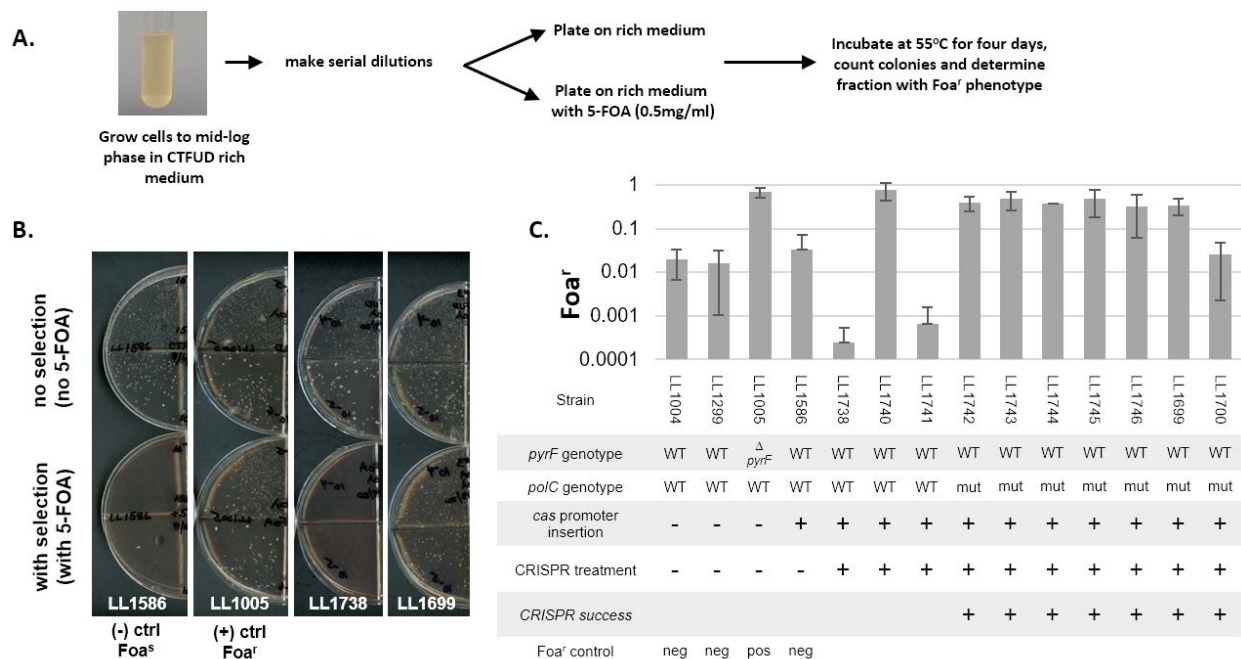| Strain	LL1004	LL1299	LL1005	LL1586	LL1738	LL1740	LL1741	LL1742	LL1743	LL1744	LL1745	LL1746	LL1699	LL1700
pyrF genotype	WT	WT	Δ*pyrF*	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
polC genotype	WT	WT	WT	WT	WT	WT	WT	mut	mut	mut	mut	mut	mut	mut
cas promoter insertion	-	-	-	+	+	+	+	+	+	+	+	+	+	+
CRISPR treatment	-	-	-	-	+	+	+	+	+	+	+	+	+	+
CRISPR success								+	+	+	+	+	+	+
Foa^r control	neg	neg	pos	neg										

325
326 **Figure 2**. **5-FOA resistance mutation rate assay. A)** Schematic diagram of cell growth and
327 selective plating. **B)** Dilution plates in the selective (with 5-FOA) and permissive (no 5-FOA)
328 conditions. The parent strain (LL1586) is WT at the *pyrF* locus and Foa^s. Strain LL1005 has a

16

329   targeted *pyrF* deletion and is Foa^r. Strain LL1738 exhibits the Foa^s phenotype. Strain LL1699

330   exhibits the Foa^r phenotype. **C)** Foa^r phenotype for wild-type and mutant *polC* strains. The *pyrF*

331   genotype, *polC* genotype, *cas* promoter insertion, and success of CRISPR treatment were

332   determined by Sanger sequencing.

333   For this mutation accumulation experiment, three strains were selected that had undergone

334   CRISPR mutagenesis: one where the mutagenesis had failed (strain LL1738, WT at *polC*) to

335   serve as a control, and two where the CRISPR mutagenesis had been successful (strains

336   LL1742 and LL1745). Each strain was serially transferred for several generations (Fig. 3). Every

337   11 generations, the population was stored in the freezer, and a single colony was isolated.

338   Observing the accumulation of mutations over time allows us to estimate the mutation rate.

339
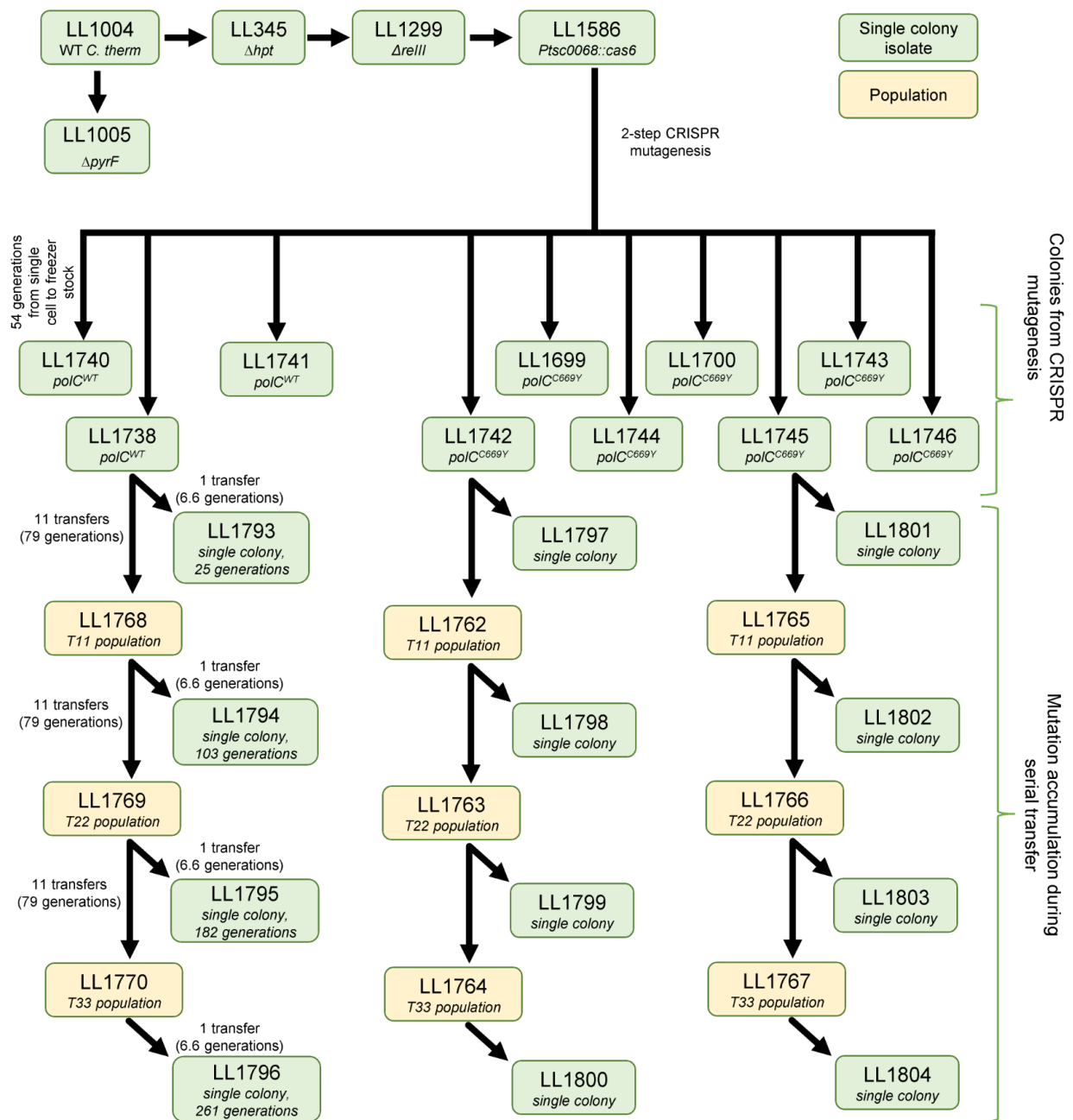
Figure 3. Lineages of the strains described in this work. Strain LL1586 was derived from WT *C. thermocellum* (LL1004) by a series of targeted mutations designed to improve the ability to perform targeted genetic modifications. Two-step CRISPR mutagenesis was performed on strain LL1586 to introduce the PolC^{C669Y} mutation. Eight individual colonies were isolated and sequenced at the *polC* locus.

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

18

349   One category is structural rearrangements, which are mutations that involve large (> 3
350   nucleotide) insertions, deletions, or replacements. This includes the targeted genetic
351   modifications (deletion of *hpt* in strain LL345 (Argyros et al. 2011), deletion of *relII* in strain
352   LL1299 (Riley et al. 2019), insertion of a strong constitutive promoter from
353   *Thermoanaerobacterium saccharolyticum* driving the native Type I cas operon in strain LL1568
354   (Walker et al. 2020). This also includes transposon insertions. *C. thermocellum* has several
355   native transposon elements (Zverlov et al. 2008; Holwerda et al. 2020). Transposons insertions
356   from families IS2, IS10, and IS120 (Siguier et al. 2006) were observed. All of these insertions,
357   except for two in the LL1769 population, were inherited from the LL1586 parent strain. The two
358   in the LL1769 population were not present in either the subsequent serial transfer (LL1770
359   population) or the single colony isolate from that transfer (strain LL1795), and we therefore
360   suspect it appeared during the preparation for genomic DNA extraction for whole-genome
361   resequencing. Mutations in this category comprise about 1% of the total number of mutations (7
362   of 775 for the 12 single colony isolates in the mutation accumulation experiment). Since these
363   mutations are not expected to be affected by PolC mutations, they were excluded from
364   subsequent analysis.
365
366   Another category is short (1-3 nucleotide) insertions, deletions, and replacements. These
367   comprise the majority (99%) of observed mutations, and the incidence of these mutations was
368   elevated in PolC mutant strains (Fig. 4).
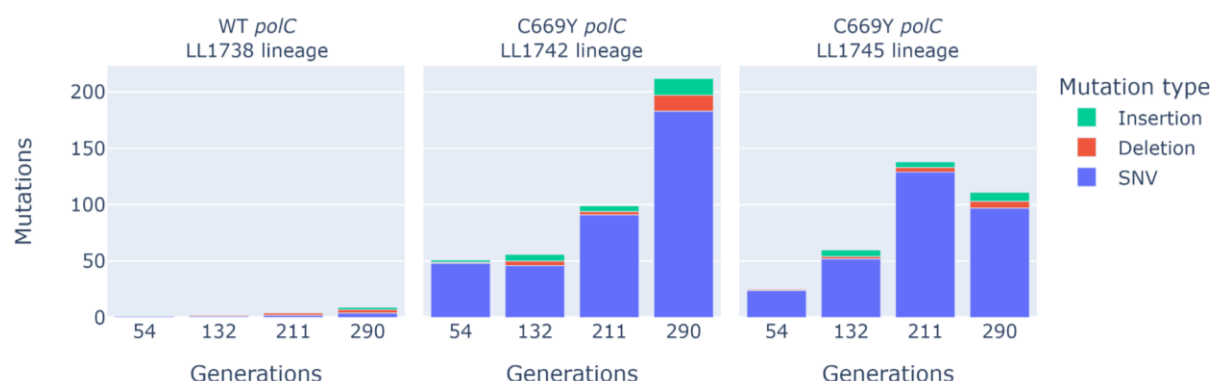369

370
371   **Figure 4**. Accumulation of mutations identified by whole-genome resequencing. Each transfer is
372   a 1:100 dilution. For each lineage, mutations present in the parent strain were filtered out. SNV
373   stands for "single nucleotide variation." The data used to generate this figure is presented in
374   detail in Supplementary Table 1.
375
376   The mutation distribution is not uniform, with an underrepresentation of A:T → T:A and G:C →
377   C:G transversion relative to other types of transitions and transversions, and an
378   overrepresentation of C:G → A:T transversions (Figure 5). This tendency has been observed by
379   others, however a mechanism is not known (Wielgoss et al. 2011; Hershberg and Petrov 2010).
380   Nevertheless, it is important to take this into consideration when designing adaptive laboratory

381 evolution experiments, since this mutational bias affects the likelihood of observing amino acid
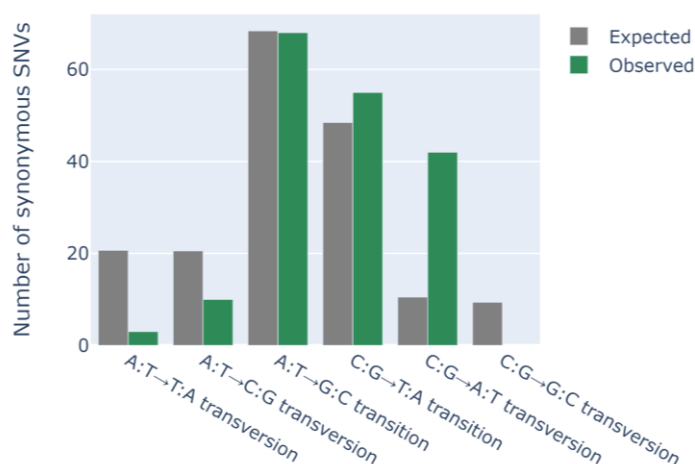382 changes.

383
384 **Figure 5**. Comparison of mutation types. For this analysis, only synonymous SNV mutations
385 from the *polC* mutant strain lineages (strains LL1797, LL1798, LL1799, LL1800, LL1801,
386 LL1802, LL1803, and LL1804) were considered. The wild type strains did not have any
387 synonymous mutations. Expected mutation frequency was determined based on the codon
388 frequency and the probability that a mutation results in a synonymous mutation.
389

Quantification of the mutation rate

391 The simplest way to calculate the mutation rate is to divide the total number of mutations by the
392 total number of generations. However, some mutations are lethal and thus not observed. To
393 correct for this, we can use synonmyous SNVs, a subset of mutations that are presumed to be
394 neutral (Motoo Kimura 1984). In the *polC* mutant strains, the median mutation rate is 0.10
395 mutations per generation or 1.6e-7 per nucleotide. It is not possible to determine the mutation
396 rate for the WT strain using synonymous mutations, since even after 290 generations, no
397 synonymous mutations were observed. We can, however, establish an upper bound of about
398 5.5e-9 per nucleotide (assuming a single mutation appeared after 290 generations). We can
399 also estimate the upper bound considering all four non-synonymous mutations in strain LL1796
400 (WT PolC, transfer 33). This gives a mutation rate of 3.9e-9 per generation. Many organisms
401 have a mutation rate of about 0.0033 per generation (Drake 1991), which would be 9.2e-10 for
402 an organism with the genome size of *C. thermocellum*. Thus, the *polC* mutation appears to have
403 increased the mutation rate between 30 and 178-fold, and looking at the structure of this
404 enzyme suggests a possible mechanism.

C669Y mutation may disrupt proofreading activity of *polC*

DNA polymerase III comes in two major forms: DnaE, and PolC. The DnaE types are further divided into three subtypes (DnaE1, DnaE2, and DnaE3). *C. thermocellum* contains both a PolC type (Clo1313_1219) and a DnaE1 type (Clo1313_0994), both of which are expressed (Holwerda et al. 2020). Many organisms with PolC, including *C. thermocellum*, do not have an epsilon proofreading subunit. In these organisms, proofreading is mediated either by the embedded EXO domain or by the PHP domain itself. This domain has been shown to have proofreading activity in *Thermus thermophilus* (Stano, Chen, and McHenry 2006). The C669Y mutation is located in the PHP domain of PolC (Fig. 6). Exonuclease activity depends on coordination with several metal ions via nine highly conserved residues. (Timinskas et al. 2014) The C669Y mutation disrupts one of these residues, which may subsequently disrupt metal ion binding and thus impair proofreading activity.

Similar mutations have been observed in *Bacillus* and *E. coli*. In *B. subtilis*, the A662V mutation is adjacent to the metal-binding cysteine. It does not have an effect on the mutation rate, but *does* cause temperature instability. (Barnes et al. 1992) In *E. coli*, the *dnaE74* mutation (G134R) is in a similar location (Vandewiele et al. 2002), however in *E. coli*, the PHP domain does not have any of the conserved metal-binding residues, and is not thought to have catalytic activity. The observed change in mutation rate (1.8-fold increase) in that organism may be due to changes in binding affinity for the epsilon proofreading subunit (*dnaQ*) instead.
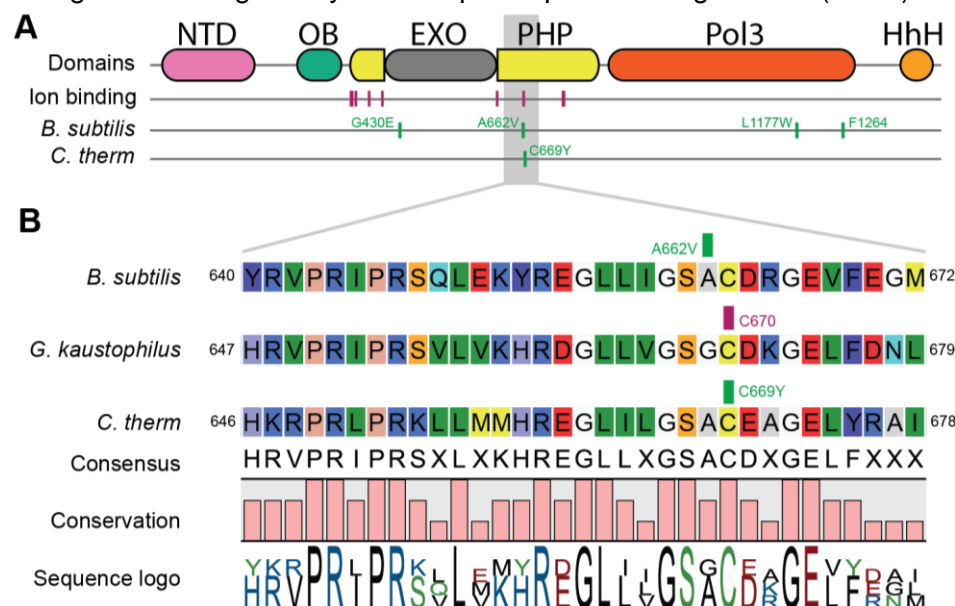

Figure 6. Location of the PolC[C669Y] mutation. **Panel A** shows the domain structure of the PolC protein. The C669Y mutation is located in the PHP (polymerase and histidinol phosphatase) domain. Other domains include N-terminal domain (NTD), oligonucleotide binding (OB), exonuclease (EXO), polymerase core (Pol3), and tandem helix-hairpin-helix motif (HhH). The PHP domain contains eight highly conserved residues (magenta) that coordinate binding with the metal ions, Mn^{2+} and Zn^{2+}. Mutations known to affect polymerase fidelity in *B. subtilis* are also indicated. **Panel B** shows a detailed view of the region surrounding the C669Y mutation. The C669Y mutation is adjacent to an A662Y mutation observed to cause a temperature-

433   sensitive phenotype in *B. subtilis*. The C669 residue in *C. thermocellum* is in the same position
434   as the C670 residue in *G. kaustophilus* (PDB ID: 3F2D) that coordinates the Mn^{2+} residue.

Other mutations that may affect the mutation rate

436   In addition to the *polC* mutation, several other mutations were identified that might affect the
437   mutation rate. Strain LL1803 (transfer 22 in the LL1745 lineage) has a mutation in the *mutS*
438   gene (Clo1313_1201) responsible for DNA mismatch repair. Mutations in this gene have been
439   shown to increase the mutation rate in other organisms (Luan et al. 2013; Willems et al. 2003).
440   Strain LL1797 (transfer 1 in the LL1742 lineage) has a mutation in the DNA polymerase III delta
441   subunit (Clo1313_1173). The delta subunit is part of the DNA polymerase III holoenzyme, and is
442   responsible for opening the sliding beta clamp protein to accept a DNA strand for replication
443   (Turner et al. 1999). Strain LL1800 (transfer 33 in the LL1742 lineage) has a mutation in the
444   *polA* polymerase (Clo1313_1334). This polymerase is thought to assist in repairing damaged
445   DNA (Hernández-Tamayo et al. 2019). Several strains (LL1801, LL1803, and LL1804) have
446   mutations in the *lexA* DNA binding protein (Clo1313_2881). All three mutations are in different
447   locations in the gene, and two of them likely eliminate activity (one is a frameshift mutation, the
448   other is a stop codon). The *lexA* gene works in tandem with the *recA* gene to induce the SOS
449   response (programmed DNA repair) (Butala, Zgur-Bertok, and Busby 2009). Since *lexA* is a
450   repressor of SOS activity, its inactivation by mutation would be expected to lead to constitutive
451   induction of the SOS response. Although the effect of these mutations has not been tested, at
452   least some of them are likely to be anti-mutator alleles. Hypermutator phenotypes that arise in
453   bacterial populations typically revert to the ancestral mutation rate when maintained in stable
454   conditions (Couce et al. 2017).

Conclusions

456   We demonstrate the utility of our recently developed CRISPR/cas system to successfully
457   introduce a PolC^{C669Y} mutation in *C. thermocellum*. This is the first example of the use of a
458   CRISPR system to introduce a *novel* mutation identified by ALE (the Walker et al. 2020 paper
459   demonstrated proof-of-concept by putting a stop codon in the *pyrF* gene). We found the HRM
460   technique to be useful for rapidly screening colonies to identify the successful introduction of
461   point mutations. The single C669Y mutation in PolC protein in *C. thermocellum* is sufficient to
462   increase the mutation rate about 30-fold. This mutation appears to function by interfering with
463   metal ion coordination in the PHP domain responsible for proofreading. The ability to selectively
464   increase the mutation rate in *C. thermocellum* is a useful tool for directed evolution experiments.
465
466

Acknowledgements

469

Bibliography

484 Altschul, S F, W Gish, W Miller, E W Myers, and D J Lipman. 1990. "Basic Local Alignment
485   Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-
486   2836(05)80360-2.
487 Argyros, D Aaron, Shital A Tripathi, Trisha F Barrett, Stephen R Rogers, Lawrence F Feinberg,
488   Daniel G Olson, Justine M Foden, et al. 2011. "High Ethanol Titers from Cellulose by Using
489   Metabolically Engineered Thermophilic, Anaerobic Microbes." *Applied and Environmental
490   Microbiology* 77 (23): 8288–94. https://doi.org/10.1128/AEM.00646-11.
491 Atmadjaja, Aretha N, Verity Holby, Amanda J Harding, Preben Krabben, Holly K Smith, and
492   Elizabeth R Jenkinson. 2019. "CRISPR-Cas, a Highly Effective Tool for Genome Editing in
493   Clostridium Saccharoperbutylacetonicum N1-4(HMT)." *FEMS Microbiology Letters* 366 (6).
494   https://doi.org/10.1093/femsle/fnz059.
495 Barnes, M H, R A Hammond, C C Kennedy, S L Mack, and N C Brown. 1992. "Localization of
496   the Exonuclease and Polymerase Domains of Bacillus Subtilis DNA Polymerase III." *Gene*
497   111 (1): 43–49. https://doi.org/10.1016/0378-1119(92)90601-k.
498 Bayer, Edward A, Raphael Lamed, Bryan A White, and Harry J Flint. 2008. "From Cellulosomes
499   to Cellulosomics." *Chemical Record (New York, N.Y.)* 8 (6): 364–77.
500   https://doi.org/10.1002/tcr.20160.
501 Boeke, Jef D., Francois La Croute, and Gerald R. Fink. 1984. "A Positive Selection for Mutants
502   Lacking Orotidine-5′-Phosphate Decarboxylase Activity in Yeast: 5-Fluoro-Orotic Acid
503   Resistance." *Molecular & General Genetics : MGG* 197 (2): 345–46.
504   https://doi.org/10.1007/BF00330984.
505 Butala, M, D Zgur-Bertok, and S J W Busby. 2009. "The Bacterial LexA Transcriptional
506   Repressor." *Cellular and Molecular Life Sciences* 66 (1): 82–93.
507   https://doi.org/10.1007/s00018-008-8378-6.
508 Couce, Alejandro, Larissa Viraphong Caudwell, Christoph Feinauer, Thomas Hindré, Jean-Paul
509   Feugeas, Martin Weigt, Richard E Lenski, Dominique Schneider, and Olivier Tenaillon.
510   2017. "Mutator Genomes Decay, despite Sustained Fitness Gains, in a Long-Term

511   Experiment with Bacteria." *Proceedings of the National Academy of Sciences of the United*
512   *States of America* 114 (43): E9026–35. https://doi.org/10.1073/pnas.1705887114.
513 Drake, J W. 1991. "A Constant Rate of Spontaneous Mutation in DNA-Based Microbes."
514   *Proceedings of the National Academy of Sciences of the United States of America* 88 (16):
515   7160–64. https://doi.org/10.1073/pnas.88.16.7160.
516 Dwight, Zachary L, Robert Palais, and Carl T Wittwer. 2012. "UAnalyze: Web-Based High-
517   Resolution DNA Melting Analysis with Comparison to Thermodynamic Predictions."
518   *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9 (6): 1805–11.
519   https://doi.org/10.1109/TCBB.2012.112.
520 Gass, K B, and N R Cozzarelli. 1973. "Further Genetic and Enzymological Characterization of
521   the Three Bacillus Subtilis Deoxyribonucleic Acid Polymerases." *The Journal of Biological*
522   *Chemistry* 248 (22): 7688–7700.
523 Gibson, Daniel G. 2011. "Enzymatic Assembly of Overlapping DNA Fragments." *Methods in*
524   *Enzymology* 498: 349–61. https://doi.org/10.1016/B978-0-12-385120-8.00015-2.
525 Grogan, D W, and R P Gunsalus. 1993. "Sulfolobus Acidocaldarius Synthesizes UMP via a
526   Standard de Novo Pathway: Results of Biochemical-Genetic Study." *Journal of Bacteriology*
527   175 (5): 1500–1507. https://doi.org/10.1128/jb.175.5.1500-1507.1993.
528 Hernández-Tamayo, Rogelio, Luis M Oviedo-Bocanegra, Georg Fritz, and Peter L Graumann.
529   2019. "Symmetric Activity of DNA Polymerases at and Recruitment of Exonuclease ExoR
530   and of PolA to the Bacillus Subtilis Replication Forks." *Nucleic Acids Research* 47 (16):
531   8521–36. https://doi.org/10.1093/nar/gkz554.
532 Hershberg, Ruth, and Dmitri A Petrov. 2010. "Evidence That Mutation Is Universally Biased
533   towards AT in Bacteria." *PLoS Genetics* 6 (9): e1001115.
534   https://doi.org/10.1371/journal.pgen.1001115.
535 Holwerda, Evert K, Daniel G Olson, Natalie M Ruppertsberger, David M Stevenson, Sean J L
536   Murphy, Marybeth I Maloney, Anthony A Lanahan, Daniel Amador-Noguez, and Lee R
537   Lynd. 2020. "Metabolic and Evolutionary Responses of Clostridium Thermocellum to
538   Genetic Interventions Aimed at Improving Ethanol Production." *Biotechnology for Biofuels*
539   13 (March): 40. https://doi.org/10.1186/s13068-020-01680-5.
540 Jacobs, K L, and D W Grogan. 1997. "Rates of Spontaneous Mutation in an Archaeon from
541   Geothermal Environments." *Journal of Bacteriology* 179 (10): 3298–3303.
542   https://doi.org/10.1128/jb.179.10.3298-3303.1997.
543 Kibota, T T, and M Lynch. 1996. "Estimate of the Genomic Mutation Rate Deleterious to Overall
544   Fitness in E. Coli." *Nature* 381 (6584): 694–96. https://doi.org/10.1038/381694a0.
545 Kimura, Motoo. 1984. *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
546 Kimura, M. 1968. "Evolutionary Rate at the Molecular Level." *Nature* 217 (5129): 624–26.
547   https://doi.org/10.1038/217624a0.
548 Kondo, S, A Yamagishi, and T Oshima. 1991. "Positive Selection for Uracil Auxotrophs of the
549   Sulfur-Dependent Thermophilic Archaebacterium Sulfolobus Acidocaldarius by Use of 5-
550   Fluoroorotic Acid." *Journal of Bacteriology* 173 (23): 7698–7700.
551   https://doi.org/10.1128/jb.173.23.7698-7700.1991.
552 Konrad, E B. 1978. "Isolation of an Escherichia Coli K-12 DnaE Mutation as a Mutator." *Journal*
553   *of Bacteriology* 133 (3): 1197–1202. https://doi.org/10.1128/JB.133.3.1197-1202.1978.
554 Kucukyildirim, Sibel, Megan Behringer, Emily M Williams, Thomas G Doak, and Michael Lynch.

555   2020. "Estimation of the Genome-Wide Mutation Rate and Spectrum in the Archaeal
556   Species Haloferax Volcanii." *Genetics* 215 (4): 1107–16.
557   https://doi.org/10.1534/genetics.120.303299.
558 Lo, Jonathan, Tianyong Zheng, Shuen Hon, Daniel G Olson, and Lee R Lynd. 2015. "The
559   Bifunctional Alcohol and Aldehyde Dehydrogenase Gene, AdhE, Is Necessary for Ethanol
560   Production in Clostridium Thermocellum and Thermoanaerobacterium Saccharolyticum."
561   *Journal of Bacteriology* 197 (8): 1386–93. https://doi.org/10.1128/JB.02450-14.
562 Luan, Guodong, Zhen Cai, Fuyu Gong, Hongjun Dong, Zhao Lin, Yanping Zhang, and Yin Li.
563   2013. "Developing Controllable Hypermutable Clostridium Cells through Manipulating Its
564   Methyl-Directed Mismatch Repair System." *Protein & Cell* 4 (11): 854–62.
565   https://doi.org/10.1007/s13238-013-3079-9.
566 Lynch, Michael, Matthew S Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W Kelley
567   Thomas, and Patricia L Foster. 2016. "Genetic Drift, Selection and the Evolution of the
568   Mutation Rate." *Nature Reviews. Genetics* 17 (11): 704–14.
569   https://doi.org/10.1038/nrg.2016.104.
570 Lynd, Lee R., Xiaoyu Liang, Mary J Biddy, Andrew Allee, Hao Cai, Thomas Foust, Michael E
571   Himmel, Mark S Laser, Michael Wang, and Charles E Wyman. 2017. "Cellulosic Ethanol:
572   Status and Innovation." *Current Opinion in Biotechnology* 45 (June): 202–11.
573   https://doi.org/10.1016/j.copbio.2017.03.008.
574 Maki, H, J Y Mo, and M Sekiguchi. 1991. "A Strong Mutator Effect Caused by an Amino Acid
575   Change in the Alpha Subunit of DNA Polymerase III of Escherichia Coli." *The Journal of
576   Biological Chemistry* 266 (8): 5055–61.
577 Mo, J Y, H Maki, and M Sekiguchi. 1991. "Mutational Specificity of the DnaE173 Mutator
578   Associated with a Defect in the Catalytic Subunit of DNA Polymerase III of Escherichia
579   Coli." *Journal of Molecular Biology* 222 (4): 925–36. https://doi.org/10.1016/0022-
580   2836(91)90586-u.
581 Nakamura, Y, and S Tabata. 1997. "Codon-Anticodon Assignment and Detection of Codon
582   Usage Trends in Seven Microbial Genomes." *Microbial & Comparative Genomics* 2 (4):
583   299–312. https://doi.org/10.1089/omi.1.1997.2.299.
584 Olson, Daniel G, and Lee R Lynd. 2012. "Transformation of Clostridium Thermocellum by
585   Electroporation." *Methods in Enzymology* 510: 317–30. https://doi.org/10.1016/B978-0-12-
586   415931-0.00017-3.
587 Paschalis, Vasileios, Emmanuelle Le Chatelier, Matthew Green, Hamid Nouri, François Képès,
588   Panos Soultanas, and Laurent Janniere. 2017. "Interactions of the Bacillus Subtilis DnaE
589   Polymerase with Replisomal Proteins Modulate Its Activity and Fidelity." *Open Biology* 7
590   (9). https://doi.org/10.1098/rsob.170146.
591 Riley, Lauren A, Lexiang Ji, Robert J Schmitz, Janet Westpheling, and Adam M Guss. 2019.
592   "Rational Development of Transformation in Clostridium Thermocellum ATCC 27405 via
593   Complete Methylome Analysis and Evasion of Native Restriction-Modification Systems."
594   *Journal of Industrial Microbiology & Biotechnology* 46 (9–10): 1435–43.
595   https://doi.org/10.1007/s10295-019-02218-x.
596 Ruiz-Rubio, M, and B A Bridges. 1987. "Mutagenic DNA Repair in Escherichia Coli. XIV.
597   Influence of Two DNA Polymerase III Mutator Alleles on Spontaneous and UV
598   Mutagenesis." *Molecular & General Genetics : MGG* 208 (3): 542–48.

599   https://doi.org/10.1007/BF00328153.

600 Sandberg, Troy E, Michael J Salazar, Liam L Weng, Bernhard O Palsson, and Adam M Feist.
601   2019. "The Emergence of Adaptive Laboratory Evolution as an Efficient Tool for Biological
602   Discovery and Industrial Biotechnology." *Metabolic Engineering* 56 (August): 1–16.
603   https://doi.org/10.1016/j.ymben.2019.08.004.

604 Sato, Kanji, Minoru Tomita, Sotaro Yonemura, Shingo Goto, Kenji Sekine, Emiko Okuma,
605   Yoshio Takagi, Koyu Hon-Nami, and Takashi Saikit. 1993. "Characterization of and Ethanol
606   Hyper-Production by *Clostridium Thermocellum* I-1-B." *Bioscience, Biotechnology, and*
607   *Biochemistry* 57 (12): 2116–21. https://doi.org/10.1271/bbb.57.2116.

608 Sevastopoulos, C G, and D A Glaser. 1977. "Mutator Action by Escherichia Coli Strains
609   Carrying DnaE Mutations." *Proceedings of the National Academy of Sciences of the United*
610   *States of America* 74 (9): 3947–50. https://doi.org/10.1073/pnas.74.9.3947.

611 Siguier, P, J Perochon, L Lestrade, J Mahillon, and M Chandler. 2006. "ISfinder: The Reference
612   Centre for Bacterial Insertion Sequences." *Nucleic Acids Research* 34 (Database issue):
613   D32-6. https://doi.org/10.1093/nar/gkj014.

614 Stano, Natalie M, Joe Chen, and Charles S McHenry. 2006. "A Coproofreading Zn(2+)-
615   Dependent Exonuclease within a Bacterial Replicase." *Nature Structural & Molecular*
616   *Biology* 13 (5): 458–59. https://doi.org/10.1038/nsmb1078.

617 Strauss, B S, R Roberts, L Francis, and P Pouryazdanparast. 2000. "Role of the DinB Gene
618   Product in Spontaneous Mutation in Escherichia Coli with an Impaired Replicative
619   Polymerase." *Journal of Bacteriology* 182 (23): 6742–50.
620   https://doi.org/10.1128/jb.182.23.6742-6750.2000.

621 Timinskas, Kestutis, Monika Balvočiūtė, Albertas Timinskas, and Česlovas Venclovas. 2014.
622   "Comprehensive Analysis of DNA Polymerase III α Subunits and Their Homologs in
623   Bacterial Genomes." *Nucleic Acids Research* 42 (3): 1393–1413.
624   https://doi.org/10.1093/nar/gkt900.

625 Trindade, Sandra, Lilia Perfeito, and Isabel Gordo. 2010. "Rate and Effects of Spontaneous
626   Mutations That Affect Fitness in Mutator Escherichia Coli." *Philosophical Transactions of*
627   *the Royal Society of London. Series B, Biological Sciences* 365 (1544): 1177–86.
628   https://doi.org/10.1098/rstb.2009.0287.

629 Tripathi, Shital A, Daniel G Olson, D Aaron Argyros, Bethany B Miller, Trisha F Barrett, Daniel M
630   Murphy, Jesse D McCool, et al. 2010. "Development of PyrF-Based Genetic System for
631   Targeted Gene Deletion in Clostridium Thermocellum and Creation of a Pta Mutant."
632   *Applied and Environmental Microbiology* 76 (19): 6591–99.
633   https://doi.org/10.1128/AEM.01484-10.

634 Turner, J, M M Hingorani, Z Kelman, and M O'Donnell. 1999. "The Internal Workings of a DNA
635   Polymerase Clamp-Loading Machine." *The EMBO Journal* 18 (3): 771–83.
636   https://doi.org/10.1093/emboj/18.3.771.

637 Vandewiele, Dominique, Antonio R Fernández de Henestrosa, Andrew R Timms, Bryn A
638   Bridges, and Roger Woodgate. 2002. "Sequence Analysis and Phenotypes of Five
639   Temperature Sensitive Mutator Alleles of DnaE, Encoding Modified Alpha-Catalytic
640   Subunits of Escherichia Coli DNA Polymerase III Holoenzyme." *Mutation Research* 499 (1):
641   85–95. https://doi.org/10.1016/s0027-5107(01)00268-8.

642 Walker, Julie E, Anthony A Lanahan, Tianyong Zheng, Camilo Toruno, Lee R Lynd, Jeffrey C

643   Cameron, Daniel G Olson, and Carrie A Eckert. 2020. "Development of Both Type I-B and
644   Type II CRISPR/Cas Genome Editing Systems in the Cellulolytic Bacterium Clostridium
645   Thermocellum." *Metabolic Engineering Communications* 10 (June): e00116.
646   https://doi.org/10.1016/j.mec.2019.e00116.
647   Wielgoss, Sébastien, Jeffrey E Barrick, Olivier Tenaillon, Stéphane Cruveiller, Béatrice Chane-
648   Woon-Ming, Claudine Médigue, Richard E Lenski, and Dominique Schneider. 2011.
649   "Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution
650   Experiment With Escherichia Coli." *G3 (Bethesda, Md.)* 1 (3): 183–86.
651   https://doi.org/10.1534/g3.111.000406.
652   Willems, Rob J, Janetta Top, Derek J Smith, David I Roper, Sarah E North, and Neil Woodford.
653   2003. "Mutations in the DNA Mismatch Repair Proteins MutS and MutL of Oxazolidinone-
654   Resistant or -Susceptible Enterococcus Faecium." *Antimicrobial Agents and Chemotherapy*
655   47 (10): 3061–66. https://doi.org/10.1128/AAC.47.10.3061-3066.2003.
656   Yan, Fei, Ren Wei, Qiu Cui, Uwe T Bornscheuer, and Ya-Jun Liu. 2021. "Thermophilic Whole-
657   Cell Degradation of Polyethylene Terephthalate Using Engineered Clostridium
658   Thermocellum." *Microbial Biotechnology* 14 (2): 374–85. https://doi.org/10.1111/1751-
659   7915.13580.
660   Zheng, Tianyong, Daniel G Olson, Liang Tian, Yannick J Bomble, Michael E Himmel, Jonathan
661   Lo, Shuen Hon, A Joe Shaw, Johannes P van Dijken, and Lee R Lynd. 2015. "Cofactor
662   Specificity of the Bifunctional Alcohol and Aldehyde Dehydrogenase (AdhE) in Wild-Type
663   and Mutant Clostridium Thermocellum and Thermoanaerobacterium Saccharolyticum."
664   *Journal of Bacteriology* 197 (15): 2610–19. https://doi.org/10.1128/JB.00232-15.
665   Zverlov, Vladimir V, Martina Klupp, Jan Krauss, and Wolfgang H Schwarz. 2008. "Mutations in
666   the Scaffoldin Gene, CipA, of Clostridium Thermocellum with Impaired Cellulosome
667   Formation and Cellulose Hydrolysis: Insertions of a New Transposable Element, IS1447,
668   and Implications for Cellulase Synergism on Crystalline Cellulose." *Journal of Bacteriology*
669   190 (12): 4321–27. https://doi.org/10.1128/JB.00097-08.