

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

Anthony Lanahan<sup>1,2</sup>, Kamila Zakowicz<sup>1</sup>, Liang Tian<sup>1,2</sup>, Daniel G. Olson<sup>1,2\*</sup> and Lee R. Lynd<sup>1,2</sup>

<sup>1</sup> Thayer School of Engineering at Dartmouth College, Hanover, NH 03755

<sup>2</sup> Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, 37830

\*To whom correspondence should be addressed: [daniel.g.olson@dartmouth.edu](mailto: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

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

# Introduction

*Clostridium thermocellum* (aka *Acetivibrio thermocellus*, *Hungateiclostridium thermocellum*, and *Ruminiclostridium thermocellum*) is a thermophilic, anaerobic bacterium that can ferment crystalline cellulose to ethanol and has attracted interest as a candidate for cellulosic biofuel production (Lynd et al. 2017). Its ability to deconstruct crystalline cellulose is mediated by a protein complex called a cellulosome (Bayer et al. 2008), and this system may have applications for deconstruction of other polymers, including plastics (Yan et al. 2021). In many cases, however, native properties need to be improved for industrial application.

Adaptive laboratory evolution (ALE) is a commonly used strategy for improving desired properties of strains, but experiments can take anywhere from weeks to years (Sandberg et al. 2019). Increasing the mutation rate of a strain can reduce the duration of ALE experiments. Mutations in DNA polymerase III are known to affect the mutation rate of bacteria. DNA polymerase III is responsible for replication of the bacterial genome. It comes in two major forms, DnaE and PolC. The widely-studied *Escherichia coli* only has the DnaE-type enzyme, and many groups have found mutator mutations in this gene (Vandewiele et al. 2002; Strauss et al. 2000; Maki, Mo, and Sekiguchi 1991; Mo, Maki, and Sekiguchi 1991; Ruiz-Rubio and Bridges 1987; Konrad 1978; Sevastopoulos and Glaser 1977). Typically, mutations that affect the fidelity of the DNA Polymerase III holoenzyme are found in the polymerase domain of DnaE or the separate epsilon proofreading subunit.

The PolC-type enzyme is found primarily in gram positive bacteria with low GC content, and has received much less attention (Timinskas et al. 2014). Organisms with PolC typically do not have a separate epsilon proofreading subunit, and instead rely on proofreading activity of the PHP domain within the PolC protein. Several hypermutator mutations in the *polC* gene in *Bacillus subtilis* have been identified (Barnes et al. 1992) (Gass and Cozzarelli 1973) (Paschalis et al. 2017). A previous ALE experiment identified a *polC* mutation among many mutations in a strain of *C. thermocellum* with a hypermutator phenotype (Holwerda et al. 2020), however the causality of the *polC* mutation was not verified, and the mutation rate was not determined.

In ALE experiments, identifying mutations is only the first step in strain improvement. Mutations have to be subsequently re-introduced so that their effect can be characterized. Previously, it has been difficult to reintroduce point mutations into *C. thermocellum*. Most examples required the deletion of the wild type gene followed by re-introduction of the mutant gene (Zheng et al. 2015; Lo et al. 2015). This process is very time-consuming (~2 months per mutation), and requires that deletion of the target gene is not toxic. Recently, we developed a new CRISPR-based system for introducing point mutations in *C. thermocellum*, based on either the native Type I or heterologous Type II CRISPR systems (Walker et al. 2020).

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*.

## 75 Methods

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

77

Strain/Plasmid	Description	Accession Number <sup>a</sup>	Reference
pLT237	Replicating plasmid with homology region for introducing PolC <sup>C669Y</sup> 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	(Tripathi et al. 2010)
LL1299	LL1004 with <i>hpt</i> and <i>relI</i> deletions	SRX2506395	(Walker et al. 2020)
LL1586	LL1299 with upregulated cas expression using <i>Tsac_0068</i> promoter	SRX4823139	(Walker et al. 2020)
LL1699	Parent strain LL1586; mutant Pol III, WT <i>pyrF</i>	SRX8904888	This work
LL1700	Parent strain LL1586; mutant Pol III, WT <i>pyrF</i>	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 <sup>C669Y</sup> mutation introduced	SRX9642825	This work
LL1743	LL1586, successful CRISPR mutagenesis, PolC <sup>C669Y</sup> mutation introduced	SRX9642826	This work
LL1744	LL1586, successful CRISPR mutagenesis, PolC <sup>C669Y</sup> mutation introduced	SRX9642781	This work
LL1745	LL1586, successful CRISPR mutagenesis, PolC <sup>C669Y</sup> mutation introduced	SRX9642780	This work
LL1746	LL1586, successful CRISPR mutagenesis, PolC <sup>C669Y</sup> 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 <i>polC</i> mutant cell	SAMN20331623	This work
LL1798	LL1742 serial transfer #11 single colony, 128 generations from original <i>polC</i> mutant	SAMN20331624	This work
LL1799	LL1742 serial transfer #22 single colony, 207 generations from original <i>polC</i> mutant cell	SAMN20331625	This work
LL1800	LL1742 serial transfer #33 single colony, 286 generations from original <i>polC</i> mutant cell	SAMN20331626	This work
LL1801	LL1745 serial transfer #1 single colony, 50 generations from original <i>polC</i> mutant cell	SAMN20331627	This work
LL1802	LL1745 serial transfer #11 single colony, 128 generations from original <i>polC</i> mutant cell	SAMN20331628	This work
LL1803	LL1745 serial transfer #22 single colony, 207 generations from original <i>polC</i> mutant cell	SAMN20331629	This work
LL1804	LL1745 serial transfer #33 single colony, 286 generations from original <i>polC</i> mutant cell	SAMN20331630	This work

78 <sup>a</sup> 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.

83

Name	Sequence (5' → 3')	Description
TY113+	ACCTTGATGACACAGAAGAAGGCGATTGA	Plasmid pLT237 confirmation
TY114+	CTAAAAGTCGTTTGTGGTTCA	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	CCAGGAAAGGGCAGTGAAGA	PolC PCR/sequencing primer
1220R	TCAAATTCATAGATTCCCAA	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	GGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATT GTTATCCGCTCACAATT	HA plasmid backbone PCR
BB R2 intF	ACCGCCTTTGAGTGAGCTGATACCG	HA plasmid backbone PCR
BackboneF	TAATAGAAATAATTTTGTTTAACTTTACAAACGGGATT GACTTTTAAAAAAGGATTGATT	HA plasmid backbone PCR
BB F intR	TGCGGCGAGCGGTATCAGCTCACTC	HA plasmid backbone PCR
Backbone Reverse	GGTCGGCATATTAAGGATCCTTGAAGTACTCTTTAATA AAATAAT	Killing plasmid backbone PCR

TY1C F	AAATTTAGGAGGCATATCAAATGGTGAATGGACCAAT AATAATGACTAGAGAAGAAAGAA	Killing plasmid backbone PCR
HA gBlock	ACAGGAAACAGCTATGACCATGATTACGCCGACAATC CTGTAATTGACACTTTGGAACTTTCAAGGCAAATGTTT CCGGAGCTTAAAAACACAACTGGATGTGGTTGCAA AGCATCTTGGTGTTCATTGGAAAACCATCACAGGGCT CTTGACGATGCAAAGGCCTGTGGGGAGATCTTTATCA AATGCCTTGAGATACTGGCTGAAAAGAATGTAAAAAC AATTGATGACATACAAAATGTTTTTGAGGGTTGTTGGA ATTATCAGAAAAGCAAATTCATACCATGCCATTATACTT GTAAAAAATTACGTAGGCCTTAGAAATCTGTACAAGAT TGTGTCCAAGACTCATTTGGAGTATTTTCACAAAAGGC CGAGACTTCCAAGAAAGCTTTTAATGATGCACAGAGA AGGACTCATTTTGGGAAGTGCCTaTGAAGCGGGAGAG CTTTACCGTGCGATTCTTGAAAACAAAGATGAGGACG AAATTTCAAGGATTGTAAAGTTTTACGATTATCTTGAA ATACAGCCTTTGGGCAACAACCGTTTCCTTGTCGAGAG CGGAAAGGTGAACAGTGAGGAAGACTTAAAGAATAT AAACAGAAAAATAGTAGCTTTGGGTGAGAAGTACAAC AAGCCGGTGGTCGCTACCTGCGATGTTTCATTTATGGA TCCGCAGGACGAAGTTTTTCAGAAGAATACTTATGGCG GGGCAAGGTTTTTCCGATGCGGACAACCAGGCACCGT TGTATTTCAGGACCACTGATGAAATGCTGGAAGAGTTT GAATATCTGGGCAGGGAAAAATGCTATGAAGTGGTTG TAACAAATACAAATTTAATTGCGGATATGTGCGAGGA CATACTTCCCATACCGGAAGGAACTTTTCCTCCTAAAAT TGAAGGTGCGGAAGAAGAAATAAGGATGCTTGCCGA AAACAAGGCGAGAGAAATTTATGGCGACCCTCTTCCT GAGATTGTGGAAGAGCGCATGGAAAAGGAGCTTAACT CAATAATAAAAAATGGCTTCTCCGTTATGTATATTATTG CGCAAAAACCTGTATGGAAATCTTTAAGTGACGGCTAT CTTGATAGGGTCGAGGGGATCCGTCGGATCATCCTTTGT TGCCTATAATAGAAATAATTTTGTTT	PolC mutant homology arm gBlock used to construct pLT237
KS1 gBlock1	TTCAAGGATCCTTAATATGCCGACCACGTTGCAATTCC CGTCAAATAATGCATTTTGCAGCCGACGAAACAGGCA AGATAACTGTATTGGCTATAAATGTTTCAGGCAGCGGT ATATTTTGCCTCCCGGTAAAATTAATACAATAAGCTAA AAACTGACGTAGGATAAGCAAAACGGCGCAATTTGA GTTGTAACGTAATATTTTCACTAAAAATAGTAATTATTT CATGTTGTTTTTTTTTAGATTAATTTATAATATAATTTAT	gBlock used to construct pDGO186N-KS1, contains Clo1313_1194 promoter, repeat#1 and spacer#1

	TGTATAAGCAATATCTTAATTATCATTAAAGGGGGAAA AAAAGTGTGTATCGTACCTATGAGGAATTGAACTT TTGGGAAGTGC GTGTGAAGCGGGAGAGCTTTACCG	
KS1 gBlock2	TTTTGGGAAGTGC GTGTGAAGCGGGAGAGCTTTACCG GTTTGTATCGTACCTATGAGGAATTGAACTTTCTGGC TACTCCAAGTTCCTAATGTTATATATGTTTTATAGTAT AAAGCCTTAAGCGAAAATGAATAGGACTGTATAATAA GGTTAAGGCTTTATATTCGCCTGAAGCGTAAAAACGCA ATAGGAAGAATAAAATTTTTGCCGGGGATAGGCCAGC CTTGCAAAGTGGCTGACAAGCGAAGTCTCCCCGCTTCC CCGGCAACAATTTCAAATAGGGAGACAGAAAGGGAG ACGAAATCTCGTATGTTGTGTGGAATTGTGAGCGGAT AACAAATTCACACAGGAAACAGCTATGACCATGATTAC GCCTAATAGAAATAATTTTGTTTAACTTTACAAACGGG ATTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAAG CAGACAAGTAAGCCTCCTAAATCACTTTAGATAAAAA TTTAGGAGGCATATCAAATGGTGAATG	gBlock used to construct pDGO186N-KS1, contains spacer#1, repeat#2 and terminator
KS2 gBlock1	TTCAAGGATCCTTAATATGCCGACCACGTTGCAATTCC CGTCAAATAATGCATTTTGACGCCGACGAAACAGGCA AGATAACTGTATTGGCTATAAATGTTTCAGGCAGCGGT ATATTTTGCTCCCGGTAAATTAATACAATAAGCTAA AAAAGTACGTAGGATAAGCAAAACGGCGCAATTTGA GTTGTAACGTAATATTTTCACTAAAAATAGTAATTATTT CATGTTGTTTTTTTTTAGATTAATTTATAATATAATTTAT TGTATAAGCAATATCTTAATTATCATTAAAGGGGGAAA AAAAGTGTGTATCGTACCTATGAGGAATTGAAACGG GAAGTGC GTGTGAAGCGGGAGAGCTTTACCGTGCG	gBlock used to construct pDGO186N-KS2, contains Clo1313_1194 promoter, repeat#1 and spacer#2
KS2 gBlock2	GGGAAGTGC GTGTGAAGCGGGAGAGCTTTACCGTGC GGTTTGTATCGTACCTATGAGGAATTGAACTTTCTGG CTACTCCAAGTTCCTAATGTTATATATGTTTTATAGTA TAAAGCCTTAAGCGAAAATGAATAGGACTGTATAATA AGGTTAAGGCTTTATATTCGCCTGAAGCGTAAAAACGC AATAGGAAGAATAAAATTTTTGCCGGGGATAGGCCAG CCTTGCAAAGTGGCTGACAAGCGAAGTCTCCCCGCTTC CCCGGCAACAATTTCAAATAGGGAGACAGAAAGGGA GACGAAATCTCGTATGTTGTGTGGAATTGTGAGCGGA TAACAATTTACACAGGAAACAGCTATGACCATGATTA CGCCTAATAGAAATAATTTTGTTTAACTTTACAAACGG GATTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAA	gBlock used to construct pDGO186N-KS2, contains spacer#2, repeat#2 and terminator



	GCAGACAAGTAAGCCTCCTAAATTCACCTTAGATAAAA ATTTAGGAGGCATATCAAATGGTGAATG	
KS3 gBlock1	TTCAAGGATCCTTAATATGCCGACCACGTTGCAATTCC CGTCAAATAATGCATTTTGCAGCCGACGAAACAGGCA AGATAACTGTATTGGCTATAAATGTTTCAGGCAGCGGT ATATTTTGCCTCCCGGTAATAAATAACAATAAGCTAA AAACTGACGTAGGATAAGCAAAACGGCGCAATTTGA GTTGTAACGTAATATTTTCACTAAAAATAGTAATTATTT CATGTTGTTTTTTTTTAGATTAATTTATAATATAATTTAT TGTATAAGCAATATCTTAATTATCATTAAAGGGGAAA AAACTGTTTGTATCGTACCTATGAGGAATTGAAACAG AATCGCACGGTAAAGCTCTCCCGCTTCACACGCAC	gBlock used to construct pDGO186N-KS3, contains Clo1313_1194 promoter, repeat#1 and spacer#3
KS3gBlock2	AGAATCGCACGGTAAAGCTCTCCCGCTTCACACGCACG TTTGATCGTACCTATGAGGAATTGAACTTTCTGGCT ACTCCAAGTTCCTAATGTTATATATGGTTTTATAGTATA AAGCCTTAAGCGAAAATGAATAGGACTGTATAATAAG GTAAAGGCTTTATATTCGCCTGAAGCGTAAAAACGCAA TAGGAAGAATAAAATTTTTGCCGGGGATAGGCCAGCC TTGCAAACTGGCTGACAAGCGAAGTCTCCCGCTTCCC CGGCAACAATTTCAAATAGGGAGACAGAAAGGGAGA CGAAATCTCGTATGTTGTGTGGAATTGTGAGCGGATA ACAATTTACACAGGAAACAGCTATGACCATGATTACG CCTAATAGAAATAATTTTGTTTAACTTTACAAACGGGA TTGACTTTTAAAAAAGGATTGATTCTAATGAAGAAAGC AGACAAGTAAGCCTCCTAAATTCACCTTAGATAAAAAT TTAGGAGGCATATCAAATGGTGAATG	gBlock used to construct pDGO186N-KS3, contains spacer#3, repeat#2 and terminator

84

## 85 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<sub>2</sub>O, resuspended in H<sub>2</sub>O, 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°C 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°C. 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°C 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°C.

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

primary transformation were pooled and 80  $\mu$ L of the mixture was inoculated into 10 mL of CTFUD and was grown overnight with and without thiamphenicol (6  $\mu$ g/mL) to determine the effect of growing the secondary transformation with and without antibiotic selection (note: transformation worked better with added antibiotic at this step, see the results section). These cultures were diluted 1:10 and grown (+/- thiamphenicol) to mid-log phase ( $Abs_{600}$  between 0.5 and 0.8), harvested, and transformed as described above. The harvested cells were transformed both individually with plasmids pDGO186N-KS1, pDGO186N-KS2, pDGO186N-KS3, as well as with a pool of the three plasmids, and with a no-DNA control. The plasmids contained a killing spacer (KS1, KS2, or KS3) as well as a neomycin (*neo*) selection marker. After recovery, a portion of the cells were plated on CTFUD/NEO (150  $\mu$ g/mL neomycin) and incubated at 55°C. Colonies usually appeared after 3 days, and were picked 1-2 days later, resuspended in 0.5 mL of CTFUD/NEO and grown overnight at 55°C. After neomycin selection, colonies from secondary transformations were screened for the target mutation using qPCR.

**HRM qPCR technique for screening point mutations.** The HRM (high-resolution melt analysis) qPCR reaction used a 100 bp amplicon and 25 bp forward and reverse primers that flanked the mutation site (Table 2). For the reaction mixture, 2  $\mu$ L of bacterial culture was mixed with 10  $\mu$ L 2X Sso Fast Evagreen qPCR mixture (BioRad USA). The primers were mixed into an equal forward/reverse (F/R) primer mixture and serially diluted to 5  $\mu$ M. 2  $\mu$ L of the F/R primer mixture was added along with 6  $\mu$ L deionized water. The mixture was set to PCR conditions of 95°C/5 min 40X, 95°C/15 s, 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 uAnalyze v2 software (Dwight, Palais, and Wittwer 2012) (Atmadjaja et al. 2019) was used to analyze the raw fluorescence data by making normalized, derivative, and difference plots.

## 5-FOA resistance test

For the 5-FOA resistance test, 5 to 50  $\mu$ L of a freezer stock, was inoculated into 1.5 mL media, grown for 6-8 hours to mid-log phase (to ensure a rapidly-dividing culture, which is optimal for 5-FOA selection), and 100  $\mu$ L was plated at various dilutions with and without 5-FOA to ensure that there were between 10-200 colonies on each plate. Usually the  $10^{-4}$ ,  $10^{-5}$  and  $10^{-6}$  dilutions had a countable number of colonies for the 5-FOA plates. The plates were incubated at 55°C for 4-6 days, and colonies were counted to determine the fraction of cells exhibiting 5-FOA resistance.

## Mutation accumulation experiment

To determine the mutation rate by mutation accumulation, cultures were serially transferred. In bacterial cells, single colony isolation events usually correspond to about 23-25 generations (Trindade, Perfeito, and Gordo 2010; Kibota and Lynch 1996). For *C. thermocellum*, based on cell volume (0.4  $\mu$ m diameter by 2  $\mu$ m long (Sato et al. 1993)), and colony volume (lenticular shape approximately 1 mm in diameter by 0.4 mm in height), a colony should contain about  $5 \times 10^8$  cells, corresponding to 29 generations. Approximately another 18 generations (1:100 dilution from initial culture of colony pick and 1:2500 dilution subculture for freezer stock preparation) occurred between the initial colony isolation following the introduction of the *polC* mutation and the preparation of the freezer stock. This does not affect the mutations observed in the starting

strains (L1699, LL1700, LL1738, LL1740-LL1746), but needs to be considered for the subsequent single colony isolations. Each serial transfer consisted of a 1:100 dilution (~6.6 generations). After 10 1:100 serial transfers, the 11th transfer was a 1:5000 dilution (~12.3 generations) to provide extra volume for preparing freezer stocks and gDNA for WGS. This was repeated three times. The 11th, 22nd, and 33rd transfers were stored as populations in the freezer, and the presence of the *polC* mutation was confirmed by HRM qPCR.

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

## Whole genome resequencing (WGS) at Dartmouth

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

## Whole genome resequencing (WGS) at JGI

Genomic DNA was submitted to the Joint Genome Institute (JGI) for sequencing with an Illumina MiSeq instrument. Paired-end reads were generated, with an average read length of 150 bp and paired distance of 500 bp. Unamplified libraries were generated using a modified version of Illumina's standard protocol. 100 ng of DNA was sheared to 500 bp using a focused ultrasonicator (Covaris). The sheared DNA fragments were size selected using SPRI beads (Beckman Coulter). The selected fragments were then end repaired, A-tailed and ligated to Illumina compatible adapters (IDT, Inc) using KAPA Illumina library creation kit (KAPA biosystems). Libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then multiplexed into pools for sequencing. The pools were loaded and sequenced on the Illumina MiSeq sequencing platform utilizing a MiSeq Reagent Kit v2 (300 cycle) following a 2 × 150 indexed run recipe.

## WGS data analysis

Read data was analyzed with the CLC Genomic Workbench version 12 (Qiagen Inc., Hilden, Germany). First, reads were trimmed using a quality limit of 0.05 and ambiguity limit of 2. Then 2.5M reads were randomly selected (to avoid errors due to differences in the total number of

reads). Reads were mapped to the reference genome (NC\_017992). Mapping was improved by two rounds of local realignment. The CLC Basic Variant Detection algorithm was used to determine small mutations (single and multiple nucleotide polymorphisms, short insertions and short deletions). Variants occurring in less than 35 % of the reads or fewer than 4 reads were filtered out. The fraction of the reads containing the mutation is presented in XXX . To determine larger mutations, the CLC InDel and Structural Variant algorithm was run. This tool analyzes unaligned ends of reads and annotates regions where a structural variation may have occurred, which are called breakpoints. Since the read length averaged 150 bp and the minimum mapping fraction was 0.5, a breakpoint can have up to 75 bp of sequence data. The resulting breakpoints were filtered to eliminate those with fewer than ten reads or less than 20 % “not perfectly matched.” The breakpoint sequence was searched with the Basic Local Alignment Search Tool (BLAST) algorithm (Altschul et al. 1990) for similarity to known sequences. Pairs of matching left and right breakpoints were considered evidence for structural variations such as transposon insertions and gene deletions. The fraction of the reads supporting the mutation (left and right breakpoints averaged) is presented in Supplementary Table 1. Mutation data from CLC was further processed using custom Python scripts (<https://github.com/danolson1/cth-mutation>).

## Sanger sequencing

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

## Quantification of mutation rate

The mutation rate was determined based on synonymous mutations, which are generally assumed to have a low effect on fitness (M Kimura 1968). *C. thermocellum* has a genome size of 3,561,619 bp (NC\_017304.1) of which 83% consists of coding regions. For each codon, the number of synonymous single-substitution events were counted, and multiplied by the codon frequency (Kazusa Codon Usage Database) (Nakamura and Tabata 1997), to reveal that 21.4% of all nucleotide positions in *C. thermocellum* coding sequences allow a synonymous mutation. This results in an effective genome size of 632,024 bp. The mutation rate ( $\mu$ , mutations per base pair per generation) is calculated with the following equation:

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

Where m is the number of observed mutations, n is the number of sites analyzed on the 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

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

*polC* gene to ensure that a functional copy was present at each stage of the chromosomal modification. Despite repeated attempts, we were unable to construct the deletion vector due to apparent toxicity in *E. coli*.

We thus pursued an alternative approach, using a recently-developed chromosomal modification technique that co-opts the native Type I CRISPR system in *C. thermocellum* (Walker et al. 2020). This system involves two transformation events (Fig. 1). The first transformation introduces the homology repair template, which introduces the desired point mutation, as well as several silent mutations to prevent spacer recognition. The second transformation introduces the killing spacer module, which targets chromosomes with the wild type *polC* sequence, but not ones modified by the homology repair template. Colonies were screened for the presence of the *polC* mutation using an HRM PCR assay (Supplementary Figure 1), which identifies mutations by a decrease in the melting temperature of mutant PCR amplicons. After transformation with the killing spacer plasmid, colonies were picked, and the *polC* region was analyzed by HRM qPCR. A total of 309 colonies were picked for HRM qPCR screening and 17 mutant candidates were identified (Table 3), Sanger DNA sequencing confirmed the presence of the *polC* mutation in all 17 candidates. Performing the secondary transformation with cells grown in the presence of thiamphenicol resulted in 5-fold higher transformation efficiency. Mutations were subsequently confirmed by Sanger sequencing and whole-genome sequencing, and 15 of 159 colonies (9.4%) had the correct mutant genotype. In most strains with the *polC* mutation, both silent mutations were also present. One strain (LL1746) was missing one of the silent mutations, indicating that both silent mutations are not necessary to prevent CRISPR targeting.

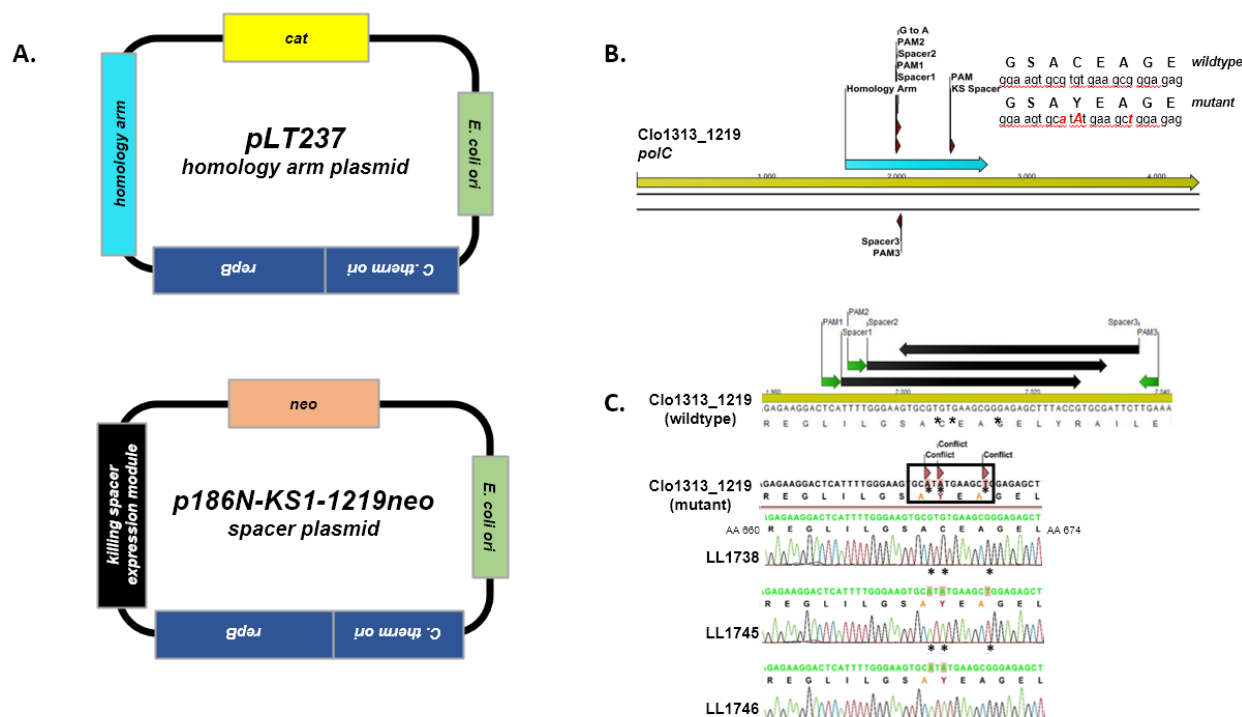
Table 3 HRM screening results

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

The genome editing work was performed at the same time as some of the experiments in our previous publication on CRISPR-based editing in *C. thermocellum* (Walker et al. 2020), and thus does not include several improvements described in that work, such as the incorporation of thermostable recombinases (*exo* and *beta* genes from *Acidithiobacillus caldus*).



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



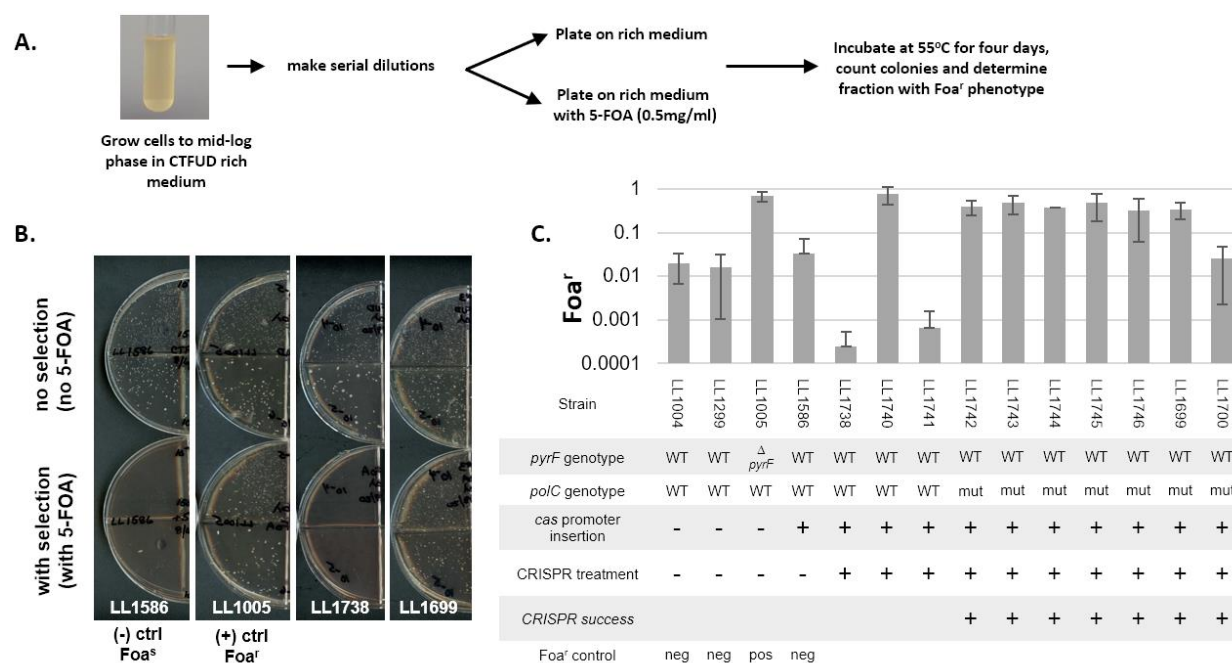
**Figure 1. A) 2-step CRISPR system for introducing mutations in *C. thermocellum*.** Step 1 is transformation with a homology arm plasmid, and step 2 is transformation with a killing spacer cassette plasmid. Each plasmid contains a *C. thermocellum* origin (including the *repB* replication protein),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

The effect of the  $PolC^{C669Y}$  mutation was determined with a 5-FOA resistance assay (Fig. 2). In rich media, the *de novo* pyrimidine biosynthesis pathway is non-essential, and it is thus a neutral site where mutations can accumulate. Mutations that inactivate this pathway create a 5-FOA resistant ( $Foa^r$ ) phenotype that can readily be detected, and this assay is frequently used as a measure of mutation rates (Boeke, La Croute, and Fink 1984; Grogan and Gunsalus 1993; Kondo, Yamagishi, and Oshima 1991; Jacobs and Grogan 1997). In this experiment, we observed a natural abundance of the  $Foa^r$  phenotype at about 2% of WT cells, similar to what

we have reported previously for *C. thermocellum* (Tripathi et al. 2010). Disruption of *pyrF* largely eliminates sensitivity to 5-FOA, with more than 70% of colonies exhibiting the Foa<sup>r</sup> phenotype (the reason this number is not 100% is likely due to a decrease in plating efficiency in the selective condition). Most of the strains harboring the PolC<sup>C669Y</sup> mutation also show an increase in the Foa<sup>r</sup> phenotype, however this was not universally observed. Strain LL1700 has the PolC<sup>C669Y</sup> mutation, but is Foa<sup>s</sup>, and strain LL1740 does not have the PolC<sup>C669Y</sup> mutation, but is Foa<sup>r</sup>.

In order to calculate the mutation rate, the target size needs to be known. We expected to observe mutations primarily in the *pyrF* gene, however Sanger sequencing did not reveal the expected mutations at that locus. Initially we plated on CTFUD with 5-FOA, picked 12 LL1299 (WT) colonies and sequenced the *pyrF* gene, but did not find any mutations. We sequenced the *pyrF* gene from 40 LL1699 (*polC* mutant) colonies picked from 5-FOA plates, and only found one mutation (a single nucleotide insertion resulting in a frame shift). Since *pyrF* mutants exhibit a growth defect that can be complemented with added uracil (Tripathi et al. 2010), we repeated the LL1299 plating experiment on CTFUD with 5-FOA and 40 ug/ml uracil. Of 18 colonies, 5 had mutations (mostly frame-shifts and premature stop codons) at the *pyrF* locus.

Subsequent whole-genome sequencing (Supplementary Table 1) did not identify any mutations associated with *de novo* pyrimidine biosynthesis, and the genetic basis for the Foa<sup>r</sup> phenotype in these strains remains unknown, which prevents an accurate determination of the mutation rate. Instead, we decided to measure mutation accumulation directly, by whole-genome sequencing.

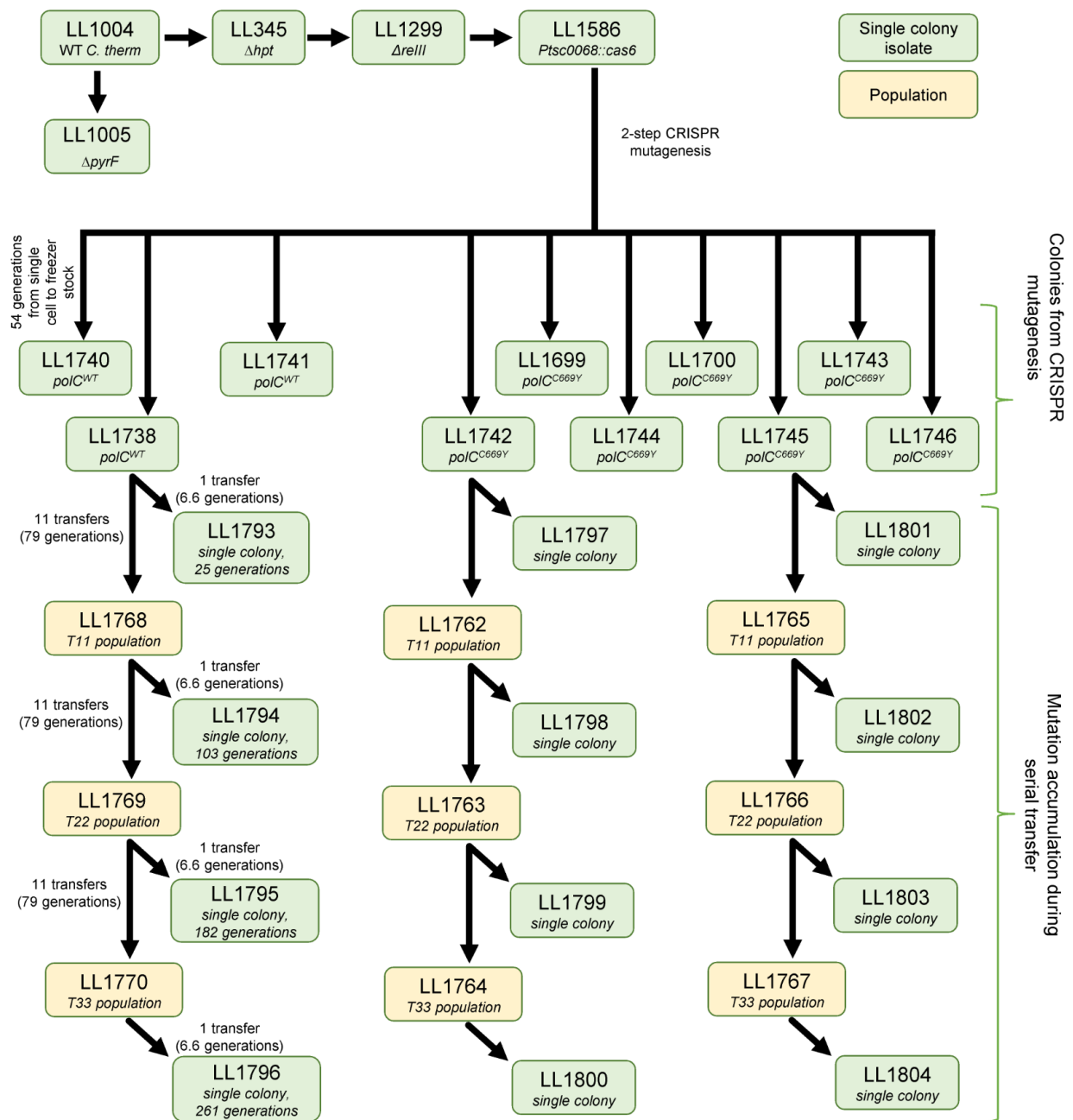


**Figure 2. 5-FOA resistance mutation rate assay. A)** Schematic diagram of cell growth and selective plating. **B)** Dilution plates in the selective (with 5-FOA) and permissive (no 5-FOA) conditions. The parent strain (LL1586) is WT at the *pyrF* locus and Foa<sup>s</sup>. Strain LL1005 has a



targeted *pyrF* deletion and is Foa<sup>r</sup>. Strain LL1738 exhibits the Foa<sup>s</sup> phenotype. Strain LL1699 exhibits the Foa<sup>r</sup> phenotype. **C)** Foa<sup>r</sup> phenotype for wild-type and mutant *polC* strains. The *pyrF* genotype, *polC* genotype, *cas* promoter insertion, and success of CRISPR treatment were determined by Sanger sequencing.

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

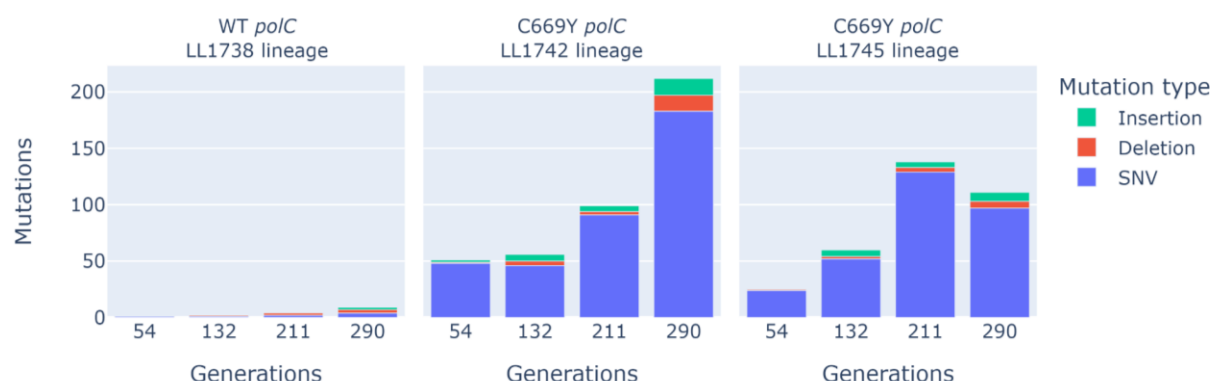


**Figure 3.** Lineages of the strains described in this work. Strain LL1586 was derived from WT *C. thermophilum* (LL1004) by a series of targeted mutations designed to improve the ability to perform targeted genetic modifications. Two-step CRISPR mutagenesis was performed on strain LL1586 to introduce the *PolC<sup>C669Y</sup>* 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<sup>C669Y</sup>* mutation (Fig. 4). Several categories of mutations were identified:

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

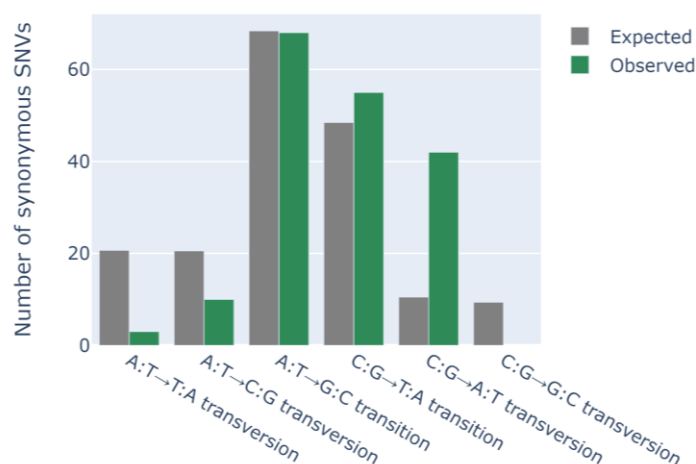
Another category is short (1-3 nucleotide) insertions, deletions, and replacements. These comprise the majority (99%) of observed mutations, and the incidence of these mutations was elevated in PolC mutant strains (Fig. 4).



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

The mutation distribution is not uniform, with an underrepresentation of A:T → T:A and G:C → C:G transversion relative to other types of transitions and transversions, and an overrepresentation of C:G → A:T transversions (Figure 5). This tendency has been observed by others, however a mechanism is not known (Wielgoss et al. 2011; Hershberg and Petrov 2010). Nevertheless, it is important to take this into consideration when designing adaptive laboratory

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



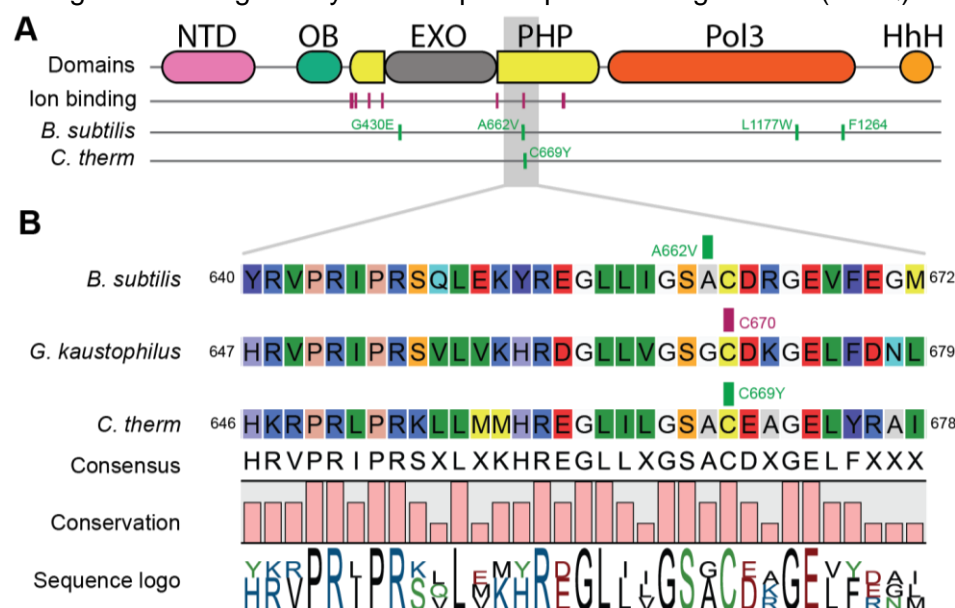
**Figure 5.** Comparison of mutation types. For this analysis, only synonymous SNV mutations from the *polC* mutant strain lineages (strains LL1797, LL1798, LL1799, LL1800, LL1801, LL1802, LL1803, and LL1804) were considered. The wild type strains did not have any synonymous mutations. Expected mutation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

## Quantification of the mutation rate

The simplest way to calculate the mutation rate is to divide the total number of mutations by the total number of generations. However, some mutations are lethal and thus not observed. To correct for this, we can use synonymous SNVs, a subset of mutations that are presumed to be neutral (Motoo Kimura 1984). In the *polC* mutant strains, the median mutation rate is 0.10 mutations per generation or  $1.6 \times 10^{-7}$  per nucleotide. It is not possible to determine the mutation rate for the WT strain using synonymous mutations, since even after 290 generations, no synonymous mutations were observed. We can, however, establish an upper bound of about  $5.5 \times 10^{-9}$  per nucleotide (assuming a single mutation appeared after 290 generations). We can also estimate the upper bound considering all four non-synonymous mutations in strain LL1796 (WT *PolC*, transfer 33). This gives a mutation rate of  $3.9 \times 10^{-9}$  per generation. Many organisms have a mutation rate of about 0.0033 per generation (Drake 1991), which would be  $9.2 \times 10^{-10}$  for an organism with the genome size of *C. thermocellum*. Thus, the *polC* mutation appears to have increased the mutation rate between 30 and 178-fold, and looking at the structure of this 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<sup>C669Y</sup> 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<sup>2+</sup> and Zn<sup>2+</sup>. 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-

sensitive phenotype in *B. subtilis*. The C669 residue in *C. thermocellum* is in the same position as the C670 residue in *G. kaustophilus* (PDB ID: 3F2D) that coordinates the Mn<sup>2+</sup> residue.

## Other mutations that may affect the mutation rate

In addition to the *polC* mutation, several other mutations were identified that might affect the mutation rate. Strain LL1803 (transfer 22 in the LL1745 lineage) has a mutation in the *mutS* gene (Clo1313\_1201) responsible for DNA mismatch repair. Mutations in this gene have been shown to increase the mutation rate in other organisms (Luan et al. 2013; Willems et al. 2003). Strain LL1797 (transfer 1 in the LL1742 lineage) has a mutation in the DNA polymerase III delta subunit (Clo1313\_1173). The delta subunit is part of the DNA polymerase III holoenzyme, and is responsible for opening the sliding beta clamp protein to accept a DNA strand for replication (Turner et al. 1999). Strain LL1800 (transfer 33 in the LL1742 lineage) has a mutation in the *polA* polymerase (Clo1313\_1334). This polymerase is thought to assist in repairing damaged DNA (Hernández-Tamayo et al. 2019). Several strains (LL1801, LL1803, and LL1804) have mutations in the *lexA* DNA binding protein (Clo1313\_2881). All three mutations are in different locations in the gene, and two of them likely eliminate activity (one is a frameshift mutation, the other is a stop codon). The *lexA* gene works in tandem with the *recA* gene to induce the SOS response (programmed DNA repair) (Butala, Zgur-Bertok, and Busby 2009). Since *lexA* is a repressor of SOS activity, its inactivation by mutation would be expected to lead to constitutive induction of the SOS response. Although the effect of these mutations has not been tested, at least some of them are likely to be anti-mutator alleles. Hypermutator phenotypes that arise in bacterial populations typically revert to the ancestral mutation rate when maintained in stable conditions (Couce et al. 2017).

## Conclusions

We demonstrate the utility of our recently developed CRISPR/cas system to successfully introduce a PolC<sup>C669Y</sup> mutation in *C. thermocellum*. This is the first example of the use of a CRISPR system to introduce a *novel* mutation identified by ALE (the Walker et al. 2020 paper demonstrated proof-of-concept by putting a stop codon in the *pyrF* gene). We found the HRM technique to be useful for rapidly screening colonies to identify the successful introduction of point mutations. The single C669Y mutation in PolC protein in *C. thermocellum* is sufficient to increase the mutation rate about 30-fold. 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 directed evolution experiments.

## Acknowledgements

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



Funding was provided by The Center for Bioenergy Innovation, a U.S. Department of Energy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. Whole genome resequencing was performed by the Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and is supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231. Additional whole genome resequencing was carried out in the Genomics and Molecular Biology Shared Resource (GMBSR) at Dartmouth which is supported by NCI Cancer Center Support Grant 5P30CA023108

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

## Bibliography

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

Experiment with Bacteria.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (43): E9026–35. <https://doi.org/10.1073/pnas.1705887114>.

Drake, J W. 1991. “A Constant Rate of Spontaneous Mutation in DNA-Based Microbes.” *Proceedings of the National Academy of Sciences of the United States of America* 88 (16): 7160–64. <https://doi.org/10.1073/pnas.88.16.7160>.

Dwight, Zachary L, Robert Palais, and Carl T Wittwer. 2012. “UAnalyze: Web-Based High-Resolution DNA Melting Analysis with Comparison to Thermodynamic Predictions.” *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9 (6): 1805–11. <https://doi.org/10.1109/TCBB.2012.112>.

Gass, K B, and N R Cozzarelli. 1973. “Further Genetic and Enzymological Characterization of the Three *Bacillus Subtilis* Deoxyribonucleic Acid Polymerases.” *The Journal of Biological Chemistry* 248 (22): 7688–7700.

Gibson, Daniel G. 2011. “Enzymatic Assembly of Overlapping DNA Fragments.” *Methods in Enzymology* 498: 349–61. <https://doi.org/10.1016/B978-0-12-385120-8.00015-2>.

Grogan, D W, and R P Gunsalus. 1993. “*Sulfolobus Acidocaldarius* Synthesizes UMP via a Standard de Novo Pathway: Results of Biochemical-Genetic Study.” *Journal of Bacteriology* 175 (5): 1500–1507. <https://doi.org/10.1128/jb.175.5.1500-1507.1993>.

Hernández-Tamayo, Rogelio, Luis M Oviedo-Bocanegra, Georg Fritz, and Peter L Graumann. 2019. “Symmetric Activity of DNA Polymerases at and Recruitment of Exonuclease ExoR and of PolA to the *Bacillus Subtilis* Replication Forks.” *Nucleic Acids Research* 47 (16): 8521–36. <https://doi.org/10.1093/nar/gkz554>.

Hershberg, Ruth, and Dmitri A Petrov. 2010. “Evidence That Mutation Is Universally Biased towards AT in Bacteria.” *PLoS Genetics* 6 (9): e1001115. <https://doi.org/10.1371/journal.pgen.1001115>.

Holwerda, Evert K, Daniel G Olson, Natalie M Ruppertsberger, David M Stevenson, Sean J L Murphy, Marybeth I Maloney, Anthony A Lanahan, Daniel Amador-Noguez, and Lee R Lynd. 2020. “Metabolic and Evolutionary Responses of *Clostridium Thermocellum* to Genetic Interventions Aimed at Improving Ethanol Production.” *Biotechnology for Biofuels* 13 (March): 40. <https://doi.org/10.1186/s13068-020-01680-5>.

Jacobs, K L, and D W Grogan. 1997. “Rates of Spontaneous Mutation in an Archaeon from Geothermal Environments.” *Journal of Bacteriology* 179 (10): 3298–3303. <https://doi.org/10.1128/jb.179.10.3298-3303.1997>.

Kibota, T T, and M Lynch. 1996. “Estimate of the Genomic Mutation Rate Deleterious to Overall Fitness in *E. Coli*.” *Nature* 381 (6584): 694–96. <https://doi.org/10.1038/381694a0>.

Kimura, Motoo. 1984. *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Kimura, M. 1968. “Evolutionary Rate at the Molecular Level.” *Nature* 217 (5129): 624–26. <https://doi.org/10.1038/217624a0>.

Kondo, S, A Yamagishi, and T Oshima. 1991. “Positive Selection for Uracil Auxotrophs of the Sulfur-Dependent Thermophilic Archaeobacterium *Sulfolobus Acidocaldarius* by Use of 5-Fluoroorotic Acid.” *Journal of Bacteriology* 173 (23): 7698–7700. <https://doi.org/10.1128/jb.173.23.7698-7700.1991>.

Konrad, E B. 1978. “Isolation of an *Escherichia Coli* K-12 DnaE Mutation as a Mutator.” *Journal of Bacteriology* 133 (3): 1197–1202. <https://doi.org/10.1128/JB.133.3.1197-1202.1978>.

Kucukyildirim, Sibel, Megan Behringer, Emily M Williams, Thomas G Doak, and Michael Lynch.



2020. "Estimation of the Genome-Wide Mutation Rate and Spectrum in the Archaeal Species *Haloferax Volcanii*." *Genetics* 215 (4): 1107–16. <https://doi.org/10.1534/genetics.120.303299>.

Lo, Jonathan, Tianyong Zheng, Shuen Hon, Daniel G Olson, and Lee R Lynd. 2015. "The Bifunctional Alcohol and Aldehyde Dehydrogenase Gene, *AdhE*, Is Necessary for Ethanol Production in *Clostridium Thermocellum* and *Thermoanaerobacterium Saccharolyticum*." *Journal of Bacteriology* 197 (8): 1386–93. <https://doi.org/10.1128/JB.02450-14>.

Luan, Guodong, Zhen Cai, Fuyu Gong, Hongjun Dong, Zhao Lin, Yanping Zhang, and Yin Li. 2013. "Developing Controllable Hypermutable *Clostridium* Cells through Manipulating Its Methyl-Directed Mismatch Repair System." *Protein & Cell* 4 (11): 854–62. <https://doi.org/10.1007/s13238-013-3079-9>.

Lynch, Michael, Matthew S Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W Kelley Thomas, and Patricia L Foster. 2016. "Genetic Drift, Selection and the Evolution of the Mutation Rate." *Nature Reviews. Genetics* 17 (11): 704–14. <https://doi.org/10.1038/nrg.2016.104>.

Lynd, Lee R., Xiaoyu Liang, Mary J Biddy, Andrew Allee, Hao Cai, Thomas Foust, Michael E Himmel, Mark S Laser, Michael Wang, and Charles E Wyman. 2017. "Cellulosic Ethanol: Status and Innovation." *Current Opinion in Biotechnology* 45 (June): 202–11. <https://doi.org/10.1016/j.copbio.2017.03.008>.

Maki, H, J Y Mo, and M Sekiguchi. 1991. "A Strong Mutator Effect Caused by an Amino Acid Change in the Alpha Subunit of DNA Polymerase III of *Escherichia Coli*." *The Journal of Biological Chemistry* 266 (8): 5055–61.

Mo, J Y, H Maki, and M Sekiguchi. 1991. "Mutational Specificity of the DnaE173 Mutator Associated with a Defect in the Catalytic Subunit of DNA Polymerase III of *Escherichia Coli*." *Journal of Molecular Biology* 222 (4): 925–36. [https://doi.org/10.1016/0022-2836\(91\)90586-u](https://doi.org/10.1016/0022-2836(91)90586-u).

Nakamura, Y, and S Tabata. 1997. "Codon-Anticodon Assignment and Detection of Codon Usage Trends in Seven Microbial Genomes." *Microbial & Comparative Genomics* 2 (4): 299–312. <https://doi.org/10.1089/omi.1.1997.2.299>.

Olson, Daniel G, and Lee R Lynd. 2012. "Transformation of *Clostridium Thermocellum* by Electroporation." *Methods in Enzymology* 510: 317–30. <https://doi.org/10.1016/B978-0-12-415931-0.00017-3>.

Paschalis, Vasileios, Emmanuelle Le Chatelier, Matthew Green, Hamid Nouri, François Képès, Panos Soultanas, and Laurent Janniere. 2017. "Interactions of the *Bacillus Subtilis* DnaE Polymerase with Replisomal Proteins Modulate Its Activity and Fidelity." *Open Biology* 7 (9). <https://doi.org/10.1098/rsob.170146>.

Riley, Lauren A, Lexiang Ji, Robert J Schmitz, Janet Westpheling, and Adam M Guss. 2019. "Rational Development of Transformation in *Clostridium Thermocellum* ATCC 27405 via Complete Methylome Analysis and Evasion of Native Restriction-Modification Systems." *Journal of Industrial Microbiology & Biotechnology* 46 (9–10): 1435–43. <https://doi.org/10.1007/s10295-019-02218-x>.

Ruiz-Rubio, M, and B A Bridges. 1987. "Mutagenic DNA Repair in *Escherichia Coli*. XIV. Influence of Two DNA Polymerase III Mutator Alleles on Spontaneous and UV Mutagenesis." *Molecular & General Genetics : MGG* 208 (3): 542–48.

<https://doi.org/10.1007/BF00328153>.

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

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

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

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

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

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

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

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

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

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

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

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

Cameron, Daniel G Olson, and Carrie A Eckert. 2020. "Development of Both Type I-B and Type II CRISPR/Cas Genome Editing Systems in the Cellulolytic Bacterium *Clostridium Thermocellum*." *Metabolic Engineering Communications* 10 (June): e00116. <https://doi.org/10.1016/j.mec.2019.e00116>.

Wielgoss, Sébastien, Jeffrey E Barrick, Olivier Tenaillon, Stéphane Cruveiller, Béatrice Chane-Woon-Ming, Claudine Médigue, Richard E Lenski, and Dominique Schneider. 2011. "Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With *Escherichia Coli*." *G3 (Bethesda, Md.)* 1 (3): 183–86. <https://doi.org/10.1534/g3.111.000406>.

Willems, Rob J, Janetta Top, Derek J Smith, David I Roper, Sarah E North, and Neil Woodford. 2003. "Mutations in the DNA Mismatch Repair Proteins MutS and MutL of Oxazolidinone-Resistant or -Susceptible *Enterococcus Faecium*." *Antimicrobial Agents and Chemotherapy* 47 (10): 3061–66. <https://doi.org/10.1128/AAC.47.10.3061-3066.2003>.

Yan, Fei, Ren Wei, Qiu Cui, Uwe T Bornscheuer, and Ya-Jun Liu. 2021. "Thermophilic Whole-Cell Degradation of Polyethylene Terephthalate Using Engineered *Clostridium Thermocellum*." *Microbial Biotechnology* 14 (2): 374–85. <https://doi.org/10.1111/1751-7915.13580>.

Zheng, Tianyong, Daniel G Olson, Liang Tian, Yannick J Bomble, Michael E Himmel, Jonathan Lo, Shuen Hon, A Joe Shaw, Johannes P van Dijken, and Lee R Lynd. 2015. "Cofactor Specificity of the Bifunctional Alcohol and Aldehyde Dehydrogenase (AdhE) in Wild-Type and Mutant *Clostridium Thermocellum* and *Thermoanaerobacterium Saccharolyticum*." *Journal of Bacteriology* 197 (15): 2610–19. <https://doi.org/10.1128/JB.00232-15>.

Zverlov, Vladimir V, Martina Klupp, Jan Krauss, and Wolfgang H Schwarz. 2008. "Mutations in the Scaffoldin Gene, CipA, of *Clostridium Thermocellum* with Impaired Cellulosome Formation and Cellulose Hydrolysis: Insertions of a New Transposable Element, IS1447, and Implications for Cellulase Synergism on Crystalline Cellulose." *Journal of Bacteriology* 190 (12): 4321–27. <https://doi.org/10.1128/JB.00097-08>.