A Single Nucleotide Change in the polC DNA Polymerase III in Clostridium thermocellum Is Sufficient To Create a Hypermutator Phenotype

Anthony Lanahan,a,b Kamila Zakowicz,a Liang Tian,a,b Daniel G. Olson,a,b Lee R. Lynd,a,b

aThayer School of Engineering at Dartmouth College, Hanover, New Hampshire, USA
bCenter for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

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 reintroduced 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 is hypothesized to function by interfering with metal ion coordination in the PHP (polymerase and histidinol phosphatase) domain, which is 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 cellulose 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.

KEYWORDS  whole-genome sequencing, next-generation sequencing, 5-fluoroorotic acid, 5-FOA, mutation rate, DNA polymerase III, polC, dnaE, Clostridium thermocellum, Hungateiclostridium thermocellum, Ruminiclostridium thermocellum, Acetivibrio thermocellus

Clostridium thermocellum (also known as 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 (1). Its ability to deconstruct crystalline cellulose is mediated by a protein complex called a cellulosome (2), and this system may have applications for deconstruction of other polymers, including plastics (3). 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 by growing them under specified growth conditions for many generations, but experiments can take anywhere from weeks to years (4). 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 has only the DnaE-type
enzyme, and many groups have found mutator mutations in this gene (5–11). 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 (12). Organisms with PolC typically do not have a separate epsilon proofreading subunit and instead rely on proofreading activity of the PHP (polymerase and histidinol phosphatase) domain within the PolC protein. Several hypermutator mutations in the polC gene in Bacillus subtilis have been identified (13–15). A previous ALE experiment identified a polC mutation among many mutations in a strain of C. thermocellum with a hypermutator phenotype (16); 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 reintroduced 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 reintroduction of the mutant gene (17, 18). This process is time-consuming (~2 months per mutation) and requires that deletion of the target gene not be 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 system (19).

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

RESULTS AND DISCUSSION

A CRISPR system successfully introduces point mutations. Initially, we attempted to reintroduce the C669Y mutation into the polC gene using standard homologous recombination techniques (20). 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 (19). 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 those modified by the homology repair template. Colonies were screened for the presence of the polC mutation using a high-resolution melt analysis (HRM) PCR assay (see Fig. S1 in the supplemental material), which identifies mutations by a difference 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 1), 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 17 of 309 colonies (6%) 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 (presumably due to a homologous recombination event between the silent mutation and the target mutation on the homology repair template), indicating that both silent mutations are not necessary to prevent CRISPR targeting.
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* (19) 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 (6% versus 40%) (19).

**Mutation quantification by 5-FOA resistance.** The effect of the *polC<sup>669Y</sup>* mutation was determined with a 5-fluoroorotic acid (5-FOA) resistance assay (Fig. 2). In rich media, the *de novo* pyrimidine biosynthesis pathway is nonessential, and it is thus a neutral site where mutations can accumulate. Mutations that inactivate this pathway create a 5-FOA resistance phenotype (Foar) that can readily be detected, and this assay is frequently used as a measure of mutation rates (21–24). In this experiment, we used 5-FOA resistance to determine the mutation rate of the *polC<sup>669Y</sup>* mutation in *C. thermocellum*.

**TABLE 1 HRM screening results of pooled plasmids**

| Screening step                               | No. in: | Expt 1 | Expt 2 |
|----------------------------------------------|---------|--------|--------|
| Initial colonies                            | 150     | 159    |        |
| Positivity by HRM qPCR                       |         | 2      | 15     |
| Presence of G2004A silent mutation           |         | 2      | 15     |
| Presence of *polC* mutation                  |         | 2      | 15     |
| Presence of G2013T silent mutation           |         | 2      | 14     |

---

**FIG 1** (A) Two-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 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 protospacer-adjacent motif (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. 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. Detailed plasmid maps are available from Addgene.
observed a natural abundance of the Foa’ phenotype at about 2% of wild-type (WT) cells, similar to what we reported previously for *C. thermocellum* (25). Disruption of *pyrF* largely eliminates sensitivity to 5-FOA, with more than 70% of colonies exhibiting the Foa’ phenotype (the reason this number is not 100% is likely a decrease in plating efficiency under the selective condition). Most of the strains harboring the PolC\(^{CC669Y}\) mutation also show an increase in the Foa’ phenotype; however, this was not universally observed. Strain LL1700 has the PolC\(^{CC669Y}\) mutation but did not generate Foa’ mutants at a rate different from that of the WT strain, and strain LL1740 does not have the PolC\(^{CC669Y}\) mutation but generated Foa’ mutants at high frequency.

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 (20) 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 found only one mutation (a single nucleotide insertion resulting in a frameshift). Since *pyrF* mutants exhibit a growth defect that can be complemented with added uracil (25), we repeated the LL1299 plating experiment on CTFUD with 5-FOA and 40 \(\mu\)g/mL uracil. Of 18 colonies, 5 had mutations (mostly frameshifts and premature stop codons) at the *pyrF* locus.

Subsequent whole-genome sequencing (WGS) (Table S1) did not identify any mutations associated with *de novo* pyrimidine biosynthesis (Clo1313_1262 through Clo1313_1270, which includes *pyrE, pyrB, pyrC, pyrF, carA*, and *pyrD*), and the genetic basis for the Foa’ phenotype in these strains remains unknown, which prevents an accurate determination of the mutation rate. Instead, we decided to measure mutation accumulation directly, by WGS.

**Mutation quantification by whole-genome sequencing.** For the 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 transfers, the popula-
tion 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.

WGS indicated a substantial increase in the rate of mutation accumulation for
strains with the PolCC669Y 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 [26], deletion of rell in strain LL1299 [27],

**FIG 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 PolCC669Y mutation. Ten individual colonies (LL1699, LL1700, LL1738, and LL1740 to -1746) were isolated and sequenced at the polC locus. Of those, three were selected (two with the polC mutation and one without) for the mutation accumulation experiment.
and insertion of a strong constitutive promoter from *Thermoanaerobacterium saccharolyticum* driving the native type I cas operon in strain LL1568 [19]). This also includes transposon insertions. *C. thermocellum* has several native transposon elements (16, 28). Transposons insertions from families IS2, IS10, and IS120 (29) 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 that they appeared during the preparation for genomic DNA extraction for WGS. Mutations in this category constitute 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 to 3 nucleotide) insertions, deletions, and replacements. These make up the majority (99%) of observed mutations, and the incidence of these mutations was elevated in PolC mutant strains (Fig. 4).

The mutation distribution is not uniform, with an underrepresentation of A:T → T:A and G:C → C:G transversions relative to other types of transitions and transversions and an overrepresentation of C:G → A:T transversions (Fig. 5). This tendency has been observed by others; however, a mechanism is not known (30, 31). Nevertheless, it is important to take this into consideration when designing ALE experiments, since this mutational bias affects the likelihood of observing amino acid changes.

**Mutation rate estimation.** 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 single nucleotide variations (SNVs), a subset of mutations that are presumed to be neutral (32). In the polC mutant strains, the median mutation rate is 0.10 mutations per generation or 1.6e−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 5.5e−9 per nucleotide (assuming that a single mutation appeared after 290 generations). We can also estimate the upper bound considering all four nonsynonymous mutations in strain LL1796 (WT PolC, transfer 33). This gives a mutation rate of 3.9e−9 per generation. Many organisms have a mutation rate of about 0.0033 per generation (33), which would be 9.2e−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 (Clo1313_1219) and a DnaE1 (Clo1313_0994), both of which are expressed (16). Many
organisms with PolC, including *C. thermocellum*, do not have an epsilon proofreading subunit. In these organisms, proofreading is mediated either by the embedded exonuclease (EXO) domain or by the PHP domain itself. This domain has been shown to have proofreading activity in *Thermus thermophilus* (34, 35) 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. (12) 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

![Comparison of mutation types](image)

**FIG 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.

![Domain structure of the PolC protein](image)

**FIG 6** Location of the PolC<sup>C669Y</sup> mutation. (A) Domain structure of the PolC protein. The C669Y mutation is located in the PHP 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. (B) 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 *Geobacillus kaustophilus* (PDB ID 3F2D) that coordinates the Mn<sup>2+</sup> residue.
mutation is adjacent to the metal-binding cysteine. It does not have an effect on the mutation rate but does cause temperature instability (13). In E. coli, the dnaE74 mutation (G134R) is in a similar location (5); 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.

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 (Table S1). 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 (36, 37). 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 (38). 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 (39). 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 and the other is a stop codon). The lexA gene works in tandem with the recA gene to induce the SOS response (programmed DNA repair) (40). Since lexA is a repressor of SOS activity, its inactivation by mutation would be expected to lead to constitutive induction of the SOS response, which could alter the mutation rate.

Hypermutator phenotypes that arise in bacterial populations typically revert to the ancestral mutation rate when maintained under stable conditions (41). The decrease in mutations observed between generations 211 and 290 is most likely explained by the takeover of the LL1745 lineage by a subpopulation with fewer mutations.

Conclusions. We demonstrate the utility of our recently developed CRISPR/Cas system to successfully introduce a PolC<sup>C669Y</sup> mutation into C. thermocellum. 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.

MATERIALS AND METHODS

Strains and plasmids used in this work are listed in Table 2.

Plasmid construction. Plasmid construction was performed by isothermal DNA assembly (42), using a NEBuilder HiFi DNA assembly cloning kit from NEB. Synthetic DNA was purchased from Integrated DNA Technologies (IDT; Coralville, IA) as gBlocks.

Growth conditions. Routine cultivation was performed using either CTFUD rich medium or MTC-5 (medium for thermophilic clostridia) chemically defined medium (20). Cells were grown at 55°C unless otherwise noted. For solid medium, agar was used at a concentration of 0.8%. Selection for the cat marker was performed with 6 μg/mL thiophenicol. Stock solutions of thiophenicol were prepared in dimethyl sulfoxide (DMSO) at a 1,000× concentration of 6 mg/mL (Sigma, no. T0251-5G). Selection for the neo marker was performed with 150 μg/mL neomycin as a 50-mg/mL stock solution in water (Gibco, no. 21810-031). Selection for pyrF mutations was performed with 5-fluoroorotic acid (5-FOA; Zymo Research, no. F9001-S) at a final concentration of 0.5 mg/mL. A 200× 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 μg/mL uracil (25); a stock solution of 40 mg/mL uracil was prepared in 1 N NaOH (Sigma, no. 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 (19) 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.
TABLE 2 Strain and plasmids used in this work

| Strain/plasmid | Description* | Addgene or accession numberb | Reference |
|----------------|--------------|-----------------------------|-----------|
| pLT237         | Replicating plasmid with homology region for introducing \(\text{polC}^{\text{C669Y}}\) mutation. Confers thiamphenicol resistance. | Addgene 174300 | This work |
| pDG10186N-KS1  | Replicating plasmid with sgRNA cassette containing a killing spacer 1 targeting the wild-type \(\text{polC}\) gene. | Addgene 174301 | This work |
| pDG10186N-KS2  | Same as above, with killing spacer 2 | Addgene 174302 | This work |
| pDG10186N-KS3  | Same as above, with killing spacer 3 | Addgene 174303 | This work |
| LL1004         | \(C.\) thermocellum DSMZ 1313 wild-type strain; WT \(\text{polC}, \text{WT pyrF}\) | NCBI reference sequence NC_017304.1 | DSMZ culture collection |
| LL1005         | LL1004 with pyrF deletion | Not available | 25 |
| LL1299         | LL1004 with hpt deletion to allow for 8AZH selection and \(\text{reII}\) deletion to improve transformation efficiency | SRX2506395 | 19 |
| LL1586         | LL1299 with upregulated cas expression using Tsac_0068 promoter | SRX4823139 | 19 |
| 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 \(\text{polC}\) locus | SRX9642234 | This work |
| LL1740         | LL1586, failed CRISPR mutagenesis at \(\text{polC}\) locus | SRX9642235 | This work |
| LL1741         | LL1586, failed CRISPR mutagenesis at \(\text{polC}\) locus | SRX9642649 | This work |
| LL1742         | LL1586, successful CRISPR mutagenesis, \(\text{polC}^{\text{C669Y}}\) mutation introduced | SRX9642825 | This work |
| LL1743         | LL1586, successful CRISPR mutagenesis, \(\text{PolC}^{\text{C669Y}}\) mutation introduced | SRX9642826 | This work |
| LL1744         | LL1586, successful CRISPR mutagenesis, \(\text{PolC}^{\text{C669Y}}\) mutation introduced | SRX9642781 | This work |
| LL1745         | LL1586, successful CRISPR mutagenesis, \(\text{PolC}^{\text{C669Y}}\) mutation introduced | SRX9642780 | This work |
| LL1746         | LL1586, successful CRISPR mutagenesis, \(\text{PolC}^{\text{C669Y}}\) 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 |
| LL1773         | LL1738 serial transfer 1 single-colony | SAMN20331619 | This work |
| LL1793         | LL1738 serial transfer 11 single-colony | SAMN20331620 | This work |
| LL1794         | LL1738 serial transfer 22 single-colony | SAMN20331621 | This work |
| LL1795         | LL1738 serial transfer 33 single-colony | SAMN20331622 | This work |
| LL1796         | LL1738 serial transfer 3 single-colony | SAMN20331623 | This work |
| LL1797         | LL1742 serial transfer 1 single-colony | SAMN20331624 | This work |
| LL1798         | LL1742 serial transfer 11 single-colony | SAMN20331625 | This work |
| LL1799         | LL1742 serial transfer 22 single-colony | SAMN20331626 | This work |
| LL1800         | LL1742 serial transfer 33 single-colony | SAMN20331627 | This work |
| LL1801         | LL1745 serial transfer 1 single-colony | SAMN20331628 | This work |
| LL1802         | LL1745 serial transfer 11 single-colony | SAMN20331629 | This work |
| LL1803         | LL1745 serial transfer 22 single-colony | SAMN20331630 | This work |

*a sgRNA, single guide RNA; 8AZH, 8-azahypoxanthine.

bAccession numbers that begin with "SRX" represent samples with WGS performed at JGI; numbers that begin with "SAMN" represent samples with WGS performed at Dartmouth. Both sets are available from the NCBI SRA database.

(i) Primary transformation with homology template. Cultures of parent strain LL1586 (native type I-8 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 water, resuspended in water, and 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 1,500 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 h in 2 mL CTFUD at 50°C overnight (20). The sample was plated on CTFUD-TM (CTFUD medium with 6 \(\mu\)g/mL thiamphenicol). After 5 days, 10 colonies were picked, pooled in 1 mL CTFUD-TM, and incubated overnight at 55°C. Colonies were pooled to simplify the subsequent subculturing step. The pooled colonies were subcultured twice to provide an opportunity for the homologous recombination events that are selected for in the
secondary transformation (19). A 1:20 dilution (50 µl into 1 mL) of the pooled colony culture was prepared in CFUD-TM medium and grown at 55°C overnight to generate 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 stored at ~80°C.

(ii) Secondary transformation of LL1586/pLT237 cells with plasmids containing spacers targeting the wild-type polC gene. The first and second transfer cultures generated after the primary transformation were pooled, and 80 µl of the mixture was inoculated into 10 mL of CFUD and was grown overnight with and without thiamphenicol (6 µg/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 Results). These cultures were diluted 1:10, grown (with and without thiamphenicol) to mid-log phase (A_600 between 0.5 and 0.8), harvested, and transformed as described above. The harvested cells were transformed with a mixture of killing spacer plasmids (pDG0186N-KS1, pDG0186N-KS2, and pDG0186N-KS3) 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 overnight recovery, a portion of the cells was plated on CFUD-NEO (150 µg/mL neomycin) and incubated at 55°C. Colonies usually appeared after 3 days and were picked 1 to 2 days later, resuspended in 0.5 mL of CFUD-NEO, and grown overnight at 55°C. After neomycin selection, colonies from secondary transformations were screened for the target mutation using qPCR.

(iii) HRM qPCR technique for screening point mutations. The HRM qPCR used a 100-bp amplicon and 25-bp forward and reverse primers that flanked the mutation site (Table 3). For the reaction mixture, 2 µl of bacterial culture was mixed with 10 µl 2× Sso Fast EvaGreen qPCR mixture (Bio-Rad, USA). The primers were mixed into an equal forward/reverse (F/R) primer mixture and serially diluted to 5 µM. Two microliters of the F/R primer mixture was added along with 6 µl deionized water. The mixture was set to PCR conditions of 95°C for 5 min and then 40 cycles of 95°C for 15 s and 55°C for 1 min. The range of the melting curve was set to 65°C to 95°C at a rate of 0.2°C/10 s. uAnalyze v2 software (43, 44) 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 µl of a freezer stock was inoculated into 1.5 mL CFUD medium and grown for 6 to 8 h to mid-log phase (to ensure a rapidly dividing culture, which is optimal for 5-FOA selection), and 100 µl was plated at various dilutions with and without 5-FOA to ensure that there were between 10 and 200 colonies on each plate. Usually the 10⁻³, 10⁻⁴, and 10⁻⁵ dilutions had a countable number of colonies for the 5-FOA plates. The plates were incubated at 55°C for 4 to 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 to 25 generations (45, 46). For C. thermocellum, based on cell volume (0.4 µm in diameter by 2 µm long [47]) and colony volume (lenticular shape, approximately 1 mm in diameter by 0.4 mm in height), a colony should contain about 5e8 cells, corresponding to 29 generations. Approximately another 18 generations (1:100 dilution from initial colony of colony pick and 1:2,500 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, and 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 1:5,000 dilution (~12.3 generations) to provide extra volume for preparing freezer stocks and genomic DNA (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, and prepared for whole-genome sequencing (WGS).

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). Five hundred nanograms 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 2 × 150-bp reads/sample. Base-calling was performed on-instrument using RTA2 and bc1 converted to fastq files using bc1fastq2 v2.20.0.422.

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. Single-amplified libraries were generated using a modified version of Illumina’s standard protocol. One hundred nanograms of DNA was sheared to 500 pM using a focused ultrasonicator (Covaris). The sheared DNA fragments were size selected using solid-phase reversible immobilization (SPRI) beads (Beckman Coulter). The selected fragments were then end repaired, A-tailed, and ligated to Illumina-compatible adapters (IDT) using a KAPA Illumina library creation kit (KAPA Biosystems). Libraries were quantified using KAPA Biosystems’ next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then
| Name   | Sequence (5'→3') | Description |
|--------|------------------|-------------|
| TY113+ | ACCTTGATGACACAGAAGAAGGCGATTTGA | Plasmid pLT237 confirmation |
| TY114+ | CTAAAGTCGTTTGTTGTC | Plasmid pLT237 confirmation |
| HRM F  | CGAGAATCCAGAAGATCCTTAAT | HRM primer |
| HRM R  | GTCTCTACTTGTGGTTTCCAAGAATC | HRM primer |
| 1219F  | GGAATTGGTGCCGTTAAGA | PoIC HA PCR primer |
| 1219R  | CTATGCCCTTCTGTGCCA | PoIC HA PCR primer |
| 1219SF | CTAGGCGCTTGAAGAACTGTA | PoIC sequencing primer |
| 1219SR | CCGGTATGGGAAGTATGT | PoIC sequencing primer |
| 1218F  | CCAGGAAAGGGCAGTGGAAGA | PoIC sequencing primer |
| 1220R  | TCAAATTTCATAGATTCCCAA | PoIC sequencing primer |
| LT651  | TAATACCTGCAAAGACCC | PoIC sequencing primer |
| LT652  | AAGAGCAGAAGAAGGGA | PoIC sequencing primer |
| LT652R | TCTCTGCTTTCTGTCTT | PoIC sequencing primer |
| 1219FR | TCTTTACCGCACCAATTTCCC | PoIC sequencing primer |
| 1219SFR | ACATACTTCCCATACCGG | PoIC sequencing primer |
| 1219FR | TTGACGAAAGAGGGAAGAT | PoIC sequencing primer |
| LT656R | GGATGTAAGAAAAGGAAAGG | PoIC sequencing primer |
| BackboneR1 | GGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCCAACAATT | HA plasmid backbone PCR |
| BB R intF | ACCGCCTTTTGGATGCTGCTGACG | HA plasmid backbone PCR |
| BackboneF | TAATAGAAATAATTTTTATATTCTTCAAAACCGGATTTGACGTTTTAAATGCCGGAAGTAT | HA plasmid backbone PCR |
| Backbone Reverse | GGTCCGATATTAAGGATCCTTGACCTACTCTTTAATAAAATAT | Killing plasmid backbone PCR |
| TY1 C F | AAATATTAGGACGCTACATCAAAATGGGATGATGAAAATAAATAGGATGAG | PoIC mutant homology arm gBlock used to construct pLT237; contains Clo1313_1194 promoter (underlined), repeat 1 (bold), and spacer 1 (italic and underlined) |
| KS1 gBlock1 | TTCAGGATCTCTATATCTGGCACCGGGGGAAGTCTTCTCTTGTAATTTGTATCGGGAAGTAAAAAACAAAGAATGCAAACTTTGCTGGTGGTTGAAGGTGATGGGATGGGGAAAGT | gBlock used to construct pDGO186N-KS1; contains Cio1313_1194 promoter (underlined), repeat 1 (bold), and spacer 1 (italic and underlined) |
| KS1 gBlock | JTTTTGGAGAATCCTGGGTAAGGAGGAATCTTCTCTGCAACCTG ATATCGGGAAGTAAAAAACAAAGAATGCAAACTTTGCTGGTGGTTGAAGGTGATGGGATGGGGAAAGT | gBlock used to construct pDGO186N-KS1; contains spacer 1 (italic and underlined), repeat 2 (bold), and terminator (underlined) |

(Continued on next page)
## Table 3 (Continued)

| Name          | Description                                                                                       |
|---------------|---------------------------------------------------------------------------------------------------|
| **KS2 gBlock1** | gBlock used to construct pDGO186N-KS2; contains Clol313_1194 promoter (underlined), repeat 1 (bold), and spacer 2 (italic and underlined) |
| **KS2 gBlock2** | gBlock used to construct pDGO186N-KS2; contains spacer 2 (italic and underlined), repeat 2 (bold), and terminator (underlined) |
| **KS3 gBlock1** | gBlock used to construct pDGO186N-KS3; contains Clol313_1194 promoter (underlined), repeat 1 (bold), and spacer 3 (italic and underlined) |
| **KS3 gBlock2** | gBlock used to construct pDGO186N-KS3; contains spacer 3 (italic and underlined), repeat 2 (bold), and terminator (underlined) |
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 were analyzed with the CLC Genomic Workbench version 12 (Qiagen, Hilden, Germany). First, reads were trimmed using a quality limit of 0.05 and an ambiguity limit of 0.2. Then, 2.5 million reads were randomly selected (to avoid errors due to differences in the total number of reads). Reads were mapped to the reference genome (NC_017304.1). 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 Table S1. 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 10 reads or less than 20% “not perfectly matched.” The breakpoint sequence was searched with the Basic Local Alignment Search Tool (BLAST) algorithm (48) 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 Table S1. Mutation data from CLC were further processed using custom Python scripts (https://github.com/danolson1/cht-mutation).

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

Quantification of mutation rate. The mutation rate (base substitution mutation rate per nucleotide site per generation) was determined by considering only synonymous mutations, which are generally assumed to have a small effect on fitness (49). This was done to avoid underestimating the mutation rate, which can occur when lethal mutations are generated that are not observed. This technique is commonly used for determining mutation rates from WGS data (30). *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 was determined and multiplied by the codon frequency (Kazusa Codon Usage Database) (50), 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 (μ [mutations per base pair per generation]) is calculated as 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 (51, 52).

SUPPLEMENTAL MATERIAL
Supplemental material is available online only.
SUPPLEMENTAL FILE 1, PDF file, 2.4 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.4 MB.

ACKNOWLEDGMENTS
We thank Ceslovas Venclovas for useful discussions related to DNA polymerases.

Funding was provided by The Center for Bioenergy Innovation, a U.S. Department of Energy (DOE) Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. WGS was performed by the DOE Joint Genome Institute, a DOE Office of Science User Facility, and is supported by the Office of Science of the DOE under contract number DE-AC02–05CH11231. Additional WGS was carried out in the Genomics and Molecular Biology Shared Resource (GBMBSR) at Dartmouth, which is supported by NCI Cancer Center Support Grant S30CA23108.

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.

REFERENCES
1. Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. Curr Opin Biotechnol 45:202–211. https://doi.org/10.1016/j.copbio.2017.03.008.
2. Bayer EA, Lamed R, White BA, Flint HJ. 2008. From cellulosomes to cellu-lycosomics. Chem Rev 8:364–377. https://doi.org/10.1021/cr078160).
3. Yan F, Wei R, Cui Q, Bornscheuer UT, Liu Y-J. 2021. Thermophilic whole-cell degradation of polyethylene terephthalate using engineered *Clostridium thermocellum*. Microbiol Biotechnol 14:374–385. https://doi.org/10.1111/1751-7915.13580.
4. Sandberg TE, Salazar MJ, Weng LL, Palsson BO, Feist AM. 2019. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. Metab Eng 56:1–16. https://doi.org/10.1016/j.menb.2019.08.004.
5. Vandewiele D, Fernández de Henestrosa AR, Timms AR, Bridges BA, Woodgate R. 2002. Sequence analysis and phenotypes of five tempera-ture sensitive mutator alleles of dnaE, encoding modified alpha-catalytic subunits of *Escherichia coli* DNA polymerase III holoenzyme. Mutat Res 499:85–95. https://doi.org/10.1016/s0027-5107(01)00268-8.
25. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR, Hogsett DA, Caiazzo NC. 2010. Development of pyrG-based genetic system for targeted gene deletion in Clostridium thermocellum and creation of a ptl mutant. Appl Environ Microbiol 76:6591–6599. https://doi.org/10.1128/AEM.01484-10.

26. Argysos DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazzo NC. 2011. High ethanol titer from cellulose by using metabolically engineered thermophilic, aerobic microorganisms. Appl Environ Microbiol 77:8288–8294. https://doi.org/10.1128/AEM.00646-11.

27. Riley LA, Ji L, Schmitz RJ, Westpheling J, Guas AM. 2019. Rational development of transformation in Clostridium thermocellum ATCC 27405 via complete methylome analysis and evasion of native restriction-modification systems. J Ind Microbiol Biotechnol 46:1435–1443. https://doi.org/10.1007/s10295-019-02218-x.

28. Zverlov VV, Klupp M, Krauss J, Schwarz WH. 2008. Mutations in the scaffoldin gene, ciaP, of Clostridium thermocellum with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS1443, and implications for cellulase synergism on crystalline cellulose. J Bacteriol 190:4321–4327. https://doi.org/10.1128/JB.190.10.3298-3303.1997.

29. Siguier P, Perochon J, Lestrade M, Mahillon J. 2006. IseeF: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.

30. Wielgoss S, Barrick JE, Tenaillon O, Cuveller S, Chane-Woon-Ming B, Médigue C, Lenski RE, Schneider D. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with Escherichia coli strain G3. (Bethesda) https://doi.org/10.1122/1.411836.6.

31. Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet 6:e1001115. https://doi.org/10.1371/journal.pgen.1001115.

32. Kimura M. 1984. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, United Kingdom.

33. Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci U S A 88:7160–7164. https://doi.org/10.1073/pnas.88.16.7160.

34. Zverlov VV, Klupp M, Krauss J, Schwarz WH. 2008. Mutations in the scaffoldin gene, ciaP, of Clostridium thermocellum with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS1443, and implications for cellulase synergism on crystalline cellulose. J Bacteriol 190:4321–4327. https://doi.org/10.1128/JB.190.10.3298-3303.1997.

35. Lapenta F, Montón Silva A, Brandimarti R, Lanzi M, Grifani FL, Vellissio Gonzalez P, Perticarari S, Hochkoeppler A. 2016. Escherichia coli DNA polymerase couples pyrophosphatase activity to DNA replication. PLoS One 11:e0152915. https://doi.org/10.1371/journal.pone.0152915.

36. Luan G, Cai Z, Gong F, Dong H, Lin Z, Zhang Y, Li Y. 2013. Developing controllable hypermutable Clostridium cells through manipulating its methyl-directed mismatch repair system. Protein Cell 4:854–862. https://doi.org/10.1123/s11328-013-0750-9.

37. Willems RJ, Top J, Smith DJ, Roper DI, North SE, Woodford N. 2003. Mutations in the DNA mismatch repair proteins MutS and MutL of oxazolidinone-resistant, multiply-resistant or susceptible Enterococcus faecium. Antimicrob Agents Chemother 47:1045–1049. https://doi.org/10.1128/AAC.47.5.1045-1049.2003.

38. Turnerr J, Hingorani MM, Kelman Z, O’Donnell M. 1999. The internal workings of a DNA polymerase clamp-loading machine. EMBO J 18:771–783. https://doi.org/10.1093/emboj/18.3.771.

39. Hernández-Tamayo R, Oviedo-Bocanegra LM, Fritz G, Graumann PL. 2019. Symmetric activity of DNA polymerases at recruitment of DNA replication complexes. Nature 567:237–241. https://doi.org/10.1038/s41586-019-1315-8.

40. Butala M, Zgu‚ Bertok D, Busby SJW. 2009. The bacterial LexA transcriptional repressor. Cell Mol Life Sci 66:82–93. https://doi.org/10.1007/s00018-008-3878-6.

41. Couce A, Caudwell LV, Feanauer C, Hindre T, Feugeas JP, Weigt M, Lenski RE, Schneider D, Tenaillon O. 2017. Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. Proc Natl Acad Sci U S A 114:E9026–E9036. https://doi.org/10.1073/pnas.1705887114.

42. Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498:349–361. https://doi.org/10.1016/B978-0-12-385120-8.00015-2.

43. Dwight ZL, Palais R, Wittwer CT. 2012. uAnalyze: web-based high-resolution DNA melting analysis with comparison to thermodynamic predictions. IEEE/ACM Trans Comput Biol Bioinform 9:1805–1811. https://doi.org/10.1109/TCBB.2012.112.

44. Atmadjaja AN, Holby V, Harding AJ, Krabben P, Smith HK, Jenkinson ER. 2015. CRISPR/Cas genome editing systems in the cellulolytic bacterium Clostridium thermocellum. Metab Eng Commun 10:e00116.https://doi.org/10.1016/j.mec.2019.e00116.
45. Trindade S, Perfeito L, Gordo I. 2010. Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. Philos Trans R Soc Lond B Biol Sci 365:1177–1186. https://doi.org/10.1098/rstb.2009.0287.

46. Kibota TT, Lynch M. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature 381:694–696. https://doi.org/10.1038/381694a0.

47. Sato K, Tomita M, Yonemura S, Goto S, Sekine K, Okuma E, Takagi Y, Hon-Nami K, Saikit T. 1993. Characterization of and ethanol hyper-production by *Clostridium thermocellum* I-1-B. Biosci Biotechnol Biochem 57:2116–2121. https://doi.org/10.1271/bbb.57.2116.

48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

49. Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217:624–626. https://doi.org/10.1038/217624a0.

50. Nakamura Y, Tabata S. 1997. Codon-anticodon assignment and detection of codon usage trends in seven microbial genomes. Microb Comp Genomics 2:299–312. https://doi.org/10.1089/omi.1.1997.2.299.

51. Kucukyildirim S, Behringer M, Williams EM, Doak TG, Lynch M. 2020. Estimation of the genome-wide mutation rate and spectrum in the archaeal species *Haloferax volcanii*. Genetics 215:1107–1116. https://doi.org/10.1534/genetics.120.303299.

52. Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet 17:704–714. https://doi.org/10.1038/nrg.2016.104.