Novel Escherichia coli active site dnaE alleles with altered base and sugar selectivity

Submit to *Molecular Microbiology* **June 25th, 2021**

RESEARCH ARTICLE

Novel *Escherichia coli* active site *dnaE* alleles with altered base and sugar selectivity

Alexandra Vaisman[1¶], Krystian Łazowski[2¶], Martin A. M. Reijns[3¶], Erin Walsh[1¶], John P. McDonald[1], Kristiniana C. Moreno[1], Dominic R. Quiros[1], Marlen Schmidt[4], Harald Kranz[4], Wei Yang[5], Karolina Makiela-Dzbenska[2] and Roger Woodgate[1*]

[1] Laboratory of Genomic Integrity, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892-3371, USA

[2] Laboratory of DNA Replication and Genome Stability, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland

[3] MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, United Kingdom

[4] Gen-H Genetic Engineering Heidelberg GmbH, Im Neuenheimer Feld 584, 69120 Heidelberg, Germany

[5] Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20814, USA

Running title: *E. coli* DNA polymerase III active site mutants

[¶] AV, KŁ, MR and EW should be considered as joint first authors

[*] Corresponding author:
E-mail: woodgate@nih.gov (RW)
Tel: +1-301-435-4040

Abbreviations-: *Escherichia coli*, *E. coli*; DNA polymerase, pol; Ribonucleotide excision repair, RER; rifampicin, Rif

Key Words: Replicase; Replication fidelity; Ribonucleotide Incorporation; Ribonucleotide Excision Repair; Steric Gate; Mutagenesis

Abstract

The *Escherichia coli dnaE* gene encodes the α-catalytic subunit (pol IIIα) of DNA polymerase III, the cell's main replicase. Like all high-fidelity DNA polymerases, pol III possesses stringent base and sugar discrimination. The latter is mediated by a so-called "steric gate" residue in the active site of the polymerase that physically clashes with the 2'-OH of an incoming ribonucleotide. Our structural modeling data suggests that H760 is the steric gate residue in *E.coli* pol IIIα. To understand how H760 and the adjacent S759 residue help maintain genome stability, we generated DNA fragments in which the codons for H760 or S759 were systematically changed to the other nineteen naturally occurring amino acids and attempted to clone them into a plasmid expressing pol III core (α-θ-ε subunits). Of the possible 38 mutants, only 9 were successfully sub-cloned: 3 with substitutions at H760 and 6 with substitutions at S759. Three of the plasmid-encoded alleles, S759C, S759N and S759T, exhibited mild to moderate mutator activity and were moved onto the chromosome for further characterization. These studies revealed altered phenotypes regarding deoxyribonucleotide base selectivity and ribonucleotide discrimination. We believe that these are the first *dnaE* mutants with such phenotypes to be reported in the literature.

1 | INTRODUCTION

Escherichia coli (*E. coli*) possesses five DNA polymerases, among which pol III holoenzyme (pol III HE), a large asymmetric dimeric macromolecular complex, is the cell's main replicase responsible for chromosome duplication by simultaneous coordinated leading and lagging strand synthesis (reviewed in (Kornberg & Baker, 1992; McHenry, 2003; O'Donnell, 2006; Pomerantz & O'Donnell, 2007; Yao & O'Donnell, 2008; Langston *et al.*, 2009; McHenry, 2011)). Pol III HE consists of 17 subunits [$(\alpha\theta\varepsilon)_2\tau_2\gamma_1\delta\delta'\chi\psi(\beta_2)_2$] encoded by 9 genes expressing the α, β, ε, θ, δ, δ', γ, τ, χ, and ψ polypeptides. The 130 kDa α-subunit polymerase belongs to the C-family of DNA polymerases and is encoded by the *dnaE* gene. The α-subunit is usually found in a tight complex with the 27.5 kDa 3'→5' proofreading exonuclease ε, encoded by the *dnaQ* gene; and the 8.6 kDa subunit θ, encoded by *holE*, which helps stabilize the three-subunit pol III core sub-assembly (Kornberg & Baker, 1992; Kim & McHenry, 1996).

Highly processive pol III HE synthesizes over 48 kb per binding event (Yao & O'Donnell, 2009; Georgescu *et al.*, 2011) and replicates DNA with high speed (up to 1,000 nucleotides per second) (Mok & Marians, 1987; McInerney *et al.*, 2007). As a result, under optimal growth conditions a single holoenzyme is sufficient to complete duplication of the entire 4 Mb *E. coli* genome in ~60 minutes (McInerney & O'Donnell, 2004; Fossum *et al.*, 2007; Reyes-Lamothe *et al.*, 2010).

The fidelity of pol III has been extensively investigated *in vivo*. An important approach that significantly improved our understanding of the molecular mechanisms ensuring accurate replication of the bacterial chromosome is based on genetic selection and screening strains for altered fidelity [reviewed in (Fijalkowska *et al.*, 2012)]. Using

this approach, several attempts were made to isolate mutations in the *dnaE* gene (Sevastopoulos & Glaser, 1977; Maki *et al.*, 1991; Fijalkowska *et al.*, 1993; Fijalkowska & Schaaper, 1993; Oller *et al.*, 1993; Oller & Schaaper, 1994; Hiratsuka & Reha-Krantz, 2000; Sugaya *et al.*, 2002; Vandewiele *et al.*, 2002; Pham *et al.*, 2006b; Yanagihara *et al.*, 2007; Makiela-Dzbenska *et al.*, 2019), as well as genes encoding other pol III HE subunits (Oller et al., 1993; Schaaper, 1993; Fijalkowska & Schaaper, 1996; Schaaper, 1996; Schaaper, 1998; Taft-Benz & Schaaper, 1998; Taft-Benz & Schaaper, 2004; Pham *et al.*, 2006a; Gawel *et al.*, 2008; Gawel *et al.*, 2011) that conferred either antimutator or mutator phenotypes.

The isolation of a series of *dnaE* mutants that result in an altered replication fidelity uncovered an important role of wild-type pol III in contributing to the normally low replication error rates in *E. coli* (Fijalkowska et al., 1993; Fijalkowska & Schaaper, 1993; Schaaper, 1993; Oller & Schaaper, 1994; Schaaper, 1996; Schaaper, 1998). Mapping of the various mutations, which are spread throughout the entire *dnaE* gene, suggested that although some of them are potentially located close to the active site of the enzyme (Kim *et al.*, 1997; Pritchard & McHenry, 1999; Lamers *et al.*, 2006; Parasuram *et al.*, 2018), the effect on replication fidelity was more often indirect. One of the causes of an altered error rate can be a change in a variety of protein–protein interactions within the holoenzyme. For example, the significant *dnaE* mutator activity of the *dnaE173* (E612K) variant results from the reduced ability of the α-subunit to interact with the ε-proofreading subunit, thus disrupting coordination of the extension step mediated by the polymerase with the reverse proofreading step mediated by the exonuclease (Maki *et al.*, 1990; Maki et al., 1991; Mo *et al.*, 1991).

Besides generating base mispairs, DNA polymerases from all kingdoms of life often make mistakes by misincorporating ribonucleotides rather than deoxyribonucleotides. Indeed, due to the considerably greater intracellular concentration of rNTPs compared to dNTPs (up to 1,000-fold in *E. coli*) (Bennett *et al.*, 2009), ribonucleotides are inserted into DNA at substantial levels. *In vitro* experiments using physiological dNTP and rNTP concentrations show that pol III holoenzyme may incorporate up to 1 rNMP every 2.3 kb (Yao *et al.*, 2013), and *in vivo*, the total number of rNMPs per *E. coli* genome has been estimated to be between 190 and 600 (Kouzminova *et al.*, 2017; Cronan *et al.*, 2019; Zatopek *et al.*, 2019), in the absence of RNase HII (encoded by *rnhB*) a key enzyme in ribonucleotide excision repair (RER). Even though rNTPs and dNTPs have the same base-coding potential, ribonucleotide incorporation might affect cellular mutability. This can occur due to the direct changes in polymerase fidelity, either during selection of a nucleotide substrate (rNTP vs. dNTP) (Joyce, 1997; Brown & Suo, 2011; Donigan *et al.*, 2014), or during replication past rNMPs embedded in the template DNA strand (Donigan et al., 2014). In addition, errantly incorporated rNMPs appear to slow the replisome (Yao et al., 2013), which also might affect replication fidelity. Changes in cellular mutability due to rNTP incorporation can also be indirect and caused by the induction of RER. Such an effect has been shown recently in our studies with low-fidelity *E. coli* pol V. Under certain conditions, wild-type pol V promotes considerable levels of spontaneous mutagenesis. However, to our initial surprise, a pol V variant with decreased sugar selectivity resulted in a significant reduction of pol V-dependent mutagenesis. We discovered that this is due to rNMP repair pathways triggered by misincorporated ribonucleotides. The main repair pathway is RER initiated by RNase

HII and completed by high-fidelity pol I-dependent nick translation that simultaneously removes rNMPs but also the pol V-dependent misincorporated dNMPs, effectively resulting in an antimutator effect (Vaisman *et al.*, 2013).

The major mechanism protecting cells from ribonucleotide incorporation is provided by DNA polymerases themselves. In most polymerases, ribose discrimination is determined by a single, so-called “steric gate” residue that not only limits rNTP misincorporation, but also, given its location within the active site, can concurrently influence base selection and overall fidelity (Brown & Suo, 2011; Donigan et al., 2014; Vaisman & Woodgate, 2018; Sassa *et al.*, 2019) as well as catalytic activity (DeLucia *et al.*, 2006). Several studies have also shown that a significant role in rNTP discrimination might also be played by the residue immediately upstream of the steric gate, which controls base substitution fidelity (Nick McElhinny *et al.*, 2010a; Nick McElhinny *et al.*, 2010b; Brown & Suo, 2011; Vaisman *et al.*, 2012; Vaisman & Woodgate, 2018).

We have previously utilized steric gate mutants of pol V to investigate the molecular mechanisms of RER in *E. coli*. These studies led to the unexpected discovery that Nucleotide Excision Repair (NER) participates in ribonucleotide removal (Vaisman et al., 2013). Pol V is a slow and distributive DNA polymerase; we therefore wanted to extend our studies to a more robust and processive polymerase. To do so, we attempted to generate mutations in the cell’s main replicase, pol III, at the putative steric gate residue, H760, or the adjacent S759 residue in the α-catalytic subunit of the polymerase. Here, we present data describing the initial characterization of three S759 mutants with differential impact on phenotypes with regards to base and sugar selectivity. These are the first pol III mutants with such phenotypes and their

characterization provides considerable insights into how *E. coli* normally avoids the catastrophic consequences of high levels of errant deoxyribonucleotide and ribonucleotide incorporation during normal DNA replication.

2 | RESULTS

2.1 | Identification of the steric gate residue in the α-catalytic subunit of *E. coli* pol III

At the time we initiated our studies, there were no high-resolution ternary-complex structures of the *E. coli* α-catalytic subunit with DNA and dNTP substrate. The most detailed structural analysis was that of another C-family polymerase, the α-subunit encoded by *polC* from *Geobacillus kaustophilus* (*G. kaustophilus*) in a ternary complex with DNA and an incoming nucleotide (dGTP) at an atomic resolution of 2.4Å (PDB ID codes 3F2B, 3F2C, and 3F2D) (Evans *et al.*, 2008). Based on this structure, we made a model of the active site of the α-subunit of *E coli* pol III (pol IIIα) (Figure 1) and concluded that the presumptive steric gate residue is H760. This amino acid is thought to be responsible for the prevention of the incorporation of nucleotides with a wrong sugar into DNA, due to the steric clash between the side chain of histidine with the 2'-OH of an incoming ribonucleotide (Figure 1). Support for the idea that H760 plays an important role in sugar selectivity, came from a study by Parasuram *et al*., in which it was independently surmised that the H760 residue contributes to the recognition and interaction with the ribose moiety of the incoming nucleotide (Parasuram et al., 2018).

2.2 | Generation of pol IIIα variants with amino acid substitutions at the steric gate residue H760, or the adjacent residue, S759

The α-subunit replicates the genome in the context of the 17-subunit pol III holoenzyme (Kornberg & Baker, 1992; Pomerantz & O'Donnell, 2007; McHenry, 2011), which can be fractionated into smaller complexes, including pol III core which comprises a tight sub-assembly of α, ε, and θ subunits (Kim & McHenry, 1996). We therefore decided to make mutant variants of the α-subunit that would be expressed in the context of pol III core. To do so, we generated pJM1260 (Table 1, and Figure S1). This vector was designed with several expression and downstream purification options in mind. First, the genes encoding α, θ, and ε subunits (*dnaE, holE,* and *dnaQ*, respectively) were codon optimized for expression in *E. coli* and chemically synthesized (Genscript). The θ-subunit was untagged, whilst ε was His-tagged at its N-terminus, and α was FLAG-tagged at its N-terminus for potential downstream affinity purification. To decrease the possibility that misincorporated bases would be subject to rapid proofreading by the ε-subunit, we also introduced the *dnaQ920* (R56W) mutation into the *dnaQ* gene, which reduces the proofreading activity of the wild-type ε-subunit by ~90% (Taft-Benz & Schaaper, 1998). The vector backbone for pJM1260 is based upon the low-copy vector, pGB2 (~5-copies per cell) (Churchward *et al.*, 1984), so as to reduce any potential overproduction artifacts. However, the vector also contains the strong IPTG-inducible pTrc promoter (de Boer *et al.*, 1983) that could be used to induce the core complex for *in vivo* studies, or downstream purification.

The final α, ε, and θ expression vector, pJM1260, comprises 10,231 bp. Importantly, codons for S759 and H760 of the α-subunit are located in a 453 bp region that is flanked by unique *Bst*BI and *Eag*I sites (Figure S1). We initially planned to synthesize 19 discrete *Bst*BI and *Eag*I fragments in which the steric gate H760 residue was changed to all 19 possible amino acids. However, we considered the possibility that some, or all, of these substitutions may be lethal, given the residue is invariant in all pol IIIα proteins. Therefore, we decided to also independently change the residue adjacent to the steric gate, S759. The precedent for such a change was based upon studies of steric gate mutants of *Saccharomyces cerevisiae* replicative DNA polymerases α, δ, and ε and *E. coli* DNA polymerase V, where changes to the residue adjacent to the steric gate led to altered base selectivity and sugar discrimination (Nick McElhinny et al., 2010a; Nick McElhinny et al., 2010b; Vaisman et al., 2012).

Consequently, we synthesized thirty-eight 453 bp DNA cassettes in which H760, or S759, were individually changed to the 19 other natural amino acids (Genscript). We then attempted to clone the DNA cassette into pJM1260 and were expecting a total of 38 variants. However, despite multiple attempts, we were only able to subclone three H760 variants (*dnaE*_H760F, *dnaE*_H760Q, *dnaE*_H760S) and six S759 variants (*dnaE*_S759A, *dnaE*_S759C, *dnaE*_S759G, *dnaE*_S759N, *dnaE*_S759T, *dnaE*_S759V) into DH5α (Table 1). We assume that our inability to subclone the remaining variants is due to “dominant negative” toxicity caused by the plasmid-encoded mutant *dnaE* variant.

2.3 | Functionality of plasmid encoded *dnaE* variants

The original plasmid clones were isolated from DH5α expressing chromosomal wild-type α- and ε-subunits. To determine if the plasmid-encoded *dnaE* mutant alleles are functionally active, the respective plasmids were transformed into RW1138 lacking pol II, pol IV and pol V (Table 2). In addition, this strain harbors the temperature sensitive *dnaE486* (S885P) allele (Wechsler & Gross, 1971) expressed from the chromosome, enabling it to grow on LB medium at permissive temperature (30°C), but not at non-permissive temperatures (which in this DNA pol II-, pol IV-, and pol V-deficient strain background is >37°C). As expected, wild-type pol IIIα expressed from pJM1260 conferred temperature resistance to the normally temperature sensitive RW1138 strain (Table 3). Similarly, all six S759 alleles were able to confer temperature resistance to the RW1138 strain (Table 3). In contrast, *dnaE*_H760F, *dnaE*_H760Q and *dnaE*_H760S failed to complement the temperature sensitivity of RW1138 (Table 3).

To assay whether the phenotypes of strains harboring the S759 or H760 plasmids corresponded to expression of the mutant pol IIIα subunits, we performed Western blot analysis of the mutant alleles expressed from pJM1260 (in the absence of IPTG induction). Extracts were probed with polyclonal rabbit antisera that we had previously raised to pol III core (unpublished results). This serum does not recognize the low chromosomal levels of the α-subunit, but it does recognize the pJM1260 plasmid encoded wild-type α-subunit (but not ε- or θ-subunits) with high specificity (Figure S2). In contrast, full-length α-subunit was not detected in extracts expressing H760F, H760Q and H760S (unpublished observations). We conclude that *dnaE*_H760F, *dnaE*_H760Q and *dnaE*_H760S plasmid encoded variants are highly unstable, in agreement with the

observation that they were unable to complement the temperature sensitivity of the *dnaE486* allele *in vivo*. In contrast to the H760 variants, all the plasmid-encoded S759 α-subunit variants are readily detectible by Western blots producing signals of approximately similar intensities (Figure S2).

Since S759 is located in the active site of the α-catalytic subunit, we anticipated that some pol IIIα variants might exhibit altered fidelity that would be manifested as a spontaneous mutator phenotype. To investigate this possibility, we introduced the plasmids into strain RW1504 (Table 2), which is similar to RW1138, but also carries the proofreading defective *dnaQ920* (R56W) allele on the chromosome and assayed for reversion of the *hisG4*(Oc) allele promoted by the mutant *dnaE* alleles at 30°C (permissive temperature) or 39°C (non-permissive for the chromosomal *dnaE486* allele). The non-permissive temperature in these experiments was lower than in the earlier studies (on LB medium) due to overall lower viability of all strains on the low-histidine minimal medium used to monitor reversion of the *hisG4* allele.

The *dnaE*_S759A, *dnaE*_S759G and *dnaE*_S759V plasmid encoded *dnaE* variants exhibited low levels of spontaneous mutagenesis at both permissive and non-permissive temperatures (Figure 3). In contrast, the *dnaE*_S759C, *dnaE*_S759N and *dnaE*_S759T alleles promoted progressively higher levels of spontaneous mutagenesis at both non-permissive and permissive temperatures (Figure 3). Given that there is no indication of altered fidelity promoted by *dnaE*_S759A, *dnaE*_S759G and *dnaE*_S759V, we chose not to characterize these alleles any further.

2.4 | Moving the S759C, S759N and S759T alleles onto the *E. coli* chromosome

To avoid any possible phenotypic artifacts promoted by the plasmid-encoded FLAG-tagged *dnaE* alleles expressed in the context of pol III core, we decided to move the untagged *dnaE*_S759C, *dnaE*_S759N or *dnaE*_S759T alleles onto the *E. coli* chromosome, where they would be expressed in the context of pol III holoenzyme. To do so, we employed Red/ET recombineering, as previously described by Kim *et al*. (Kim *et al.*, 2014), but with minor changes (see *Experimental procedures* and Table S1). During this process, the respective *dnaE* allele replaced the wild-type *dnaE* gene. We then used conventional P1 transduction protocols to link the *dnaE* alleles to the nearby *yafC727*::Kan allele from JW0198 (*E. coli* Genetic Stock Center), or the *yafC502*::Tn*10* allele from CAG18436 (*E. coli* Genetic Stock Center). Previous studies have shown that *yafC* and *dnaE* are co-transduced with a frequency of ~45% (Fijalkowska et al., 1993; Vandewiele et al., 2002) (Figure S3). Finally, we transduced the respective *dnaE* alleles into the Δ*polB*, Δ*dinB,* and Δ*umuDC* strain, RW1604 (Table 2), so as to avoid any influence of other DNA polymerases on the replication fidelity and/or ribonucleotide incorporation. This strain also harbors the Δ*rnhB782*::Kan^S allele immediately upstream of *dnaE486ts* (Figure S3) and the *dnaQ920* (R56W) allele downstream of *yafC* (Figure S3). The *rnhB-dnaE-yafC-dnaQ* interval is only ~33 kb in length, meaning that all four genes can be co-transduced in a single P1 transduction (linkage of all four genes at one time is ~10%). Indeed, by screening for the appropriate gene marker, we were able to generate a series of *dnaE* strains that were deficient (Δ*rnhB*) or proficient (*rnhB*_wt) for RNase HII-dependent-RER, in the presence of fully active (*dnaQ*_wt) or reduced (*dnaQ920*) proofreading activity of pol III (Table 2).

Interestingly, while we were able to make wild-type *dnaE* and *dnaE*_S759T strains that were Δ*rnhB dnaQ920* using selection for *yafC727*::Kan (Table 2), we were unable to make similar strains carrying the *dnaE*_S759C or *dnaE*_S759N alleles using this same approach. Instead, we used P1 lysates from strains in which Δ*rnhB782*::Kan was first linked to *dnaE*_S759C (EC10540), or *dnaE*_S759N (EC10541), to simultaneously transduce Δ*rnhB782*::Kan and the two S759C/N *dnaE* alleles into the Δ*yafC502*::Tn*10* *dnaQ920* strain, EC10539 (Table 2).

2.5 | Spontaneous mutagenesis promoted by chromosomally encoded *dnaE* variants

The effect of reduced proofreading activity on mutagenesis promoted by wild-type *dnaE*, *dnaE*_S759C, *dnaE*_S759N and *dnaE*_S759T was first investigated by qualitative plate assays that followed the reversion of the *hisG4*(Oc) (Figure S4) and *galK2*(Oc) alleles (Figure S5). These assays reveal that all three *dnaE* alleles expressed from the chromosome confer a mild spontaneous mutator phenotype. In the absence of proofreading, spontaneous mutator activity increased significantly, especially with the *dnaE*_S759N and *dnaE*_S759T alleles.

To more accurately determine effects of the three *dnaE* alleles on potential mutator activity, we used quantitative fluctuation assays to monitor forward mutagenesis to rifampicin resistance (mutations in *rpoB* gene encoding the β-subunit of RNA polymerase). We compared the level of Rif mutagenesis promoted by wild-type *dnaE* and the three *dnaE*_S759 variants in a repair-proficient background, or in a background with altered RER (Δ*rnhB*), or proofreading activity (*dnaQ920*), or both (Δ*rnhB dnaQ920*)

(Table 4). In the repair-proficient background, the *dnaE*_S759N allele displayed a moderate mutator effect, estimated to be ~7-fold higher than wild-type *dnaE*, while *dnaE*_S759T and *dnaE*_S759C alleles were lower mutators (3- and 1.7-fold, respectively). Diminished proofreading (*dnaQ920*) in strains carrying *dnaE* alleles led to a further increase in mutagenesis, resulting in ~3.5 – 66-fold mutator effects compared to the wild-type *dnaE*$^{+}$ strain. Synergistic effects observed between *dnaE* variants and *dnaQ920* allele indicate that replication errors generated by all three mutants are subject to correction by the proofreading activity of pol III. Inactivation of the main RER pathway (Δ*rnhB*) in both proofreading proficient and proofreading deficient backgrounds had no significant effect on mutagenesis promoted by wild-type *dnaE* or any of the three *dnaE* mutants (Table 4).

2.6 | Different mutational spectra for *dnaE*_S759C, S759N and S759T in proofreading deficient *dnaQ920* strains

We were interested in investigating the possibility that the *dnaE*_S759 variants might exhibit altered base substitution specificity in addition to their differing mutator phenotypes. To do so, we analyzed the mutation profiles of the mutant polymerases in a proofreading-deficient (*dnaQ920*) background by determining the spectra of spontaneously arising missense mutations that lead to rifampin resistance. Such an approach has previously been utilized to show that each of *E. coli*'s five DNA polymerases exhibits a unique mutational signature (Garibyan *et al.*, 2003; Wolff *et al.*, 2004; Curti *et al.*, 2009; Makiela-Dzbenska *et al.*, 2011; Vaisman et al., 2013). The strains used in this analysis are proficient for methyl-directed mismatch repair (MMR),

which is known to preferentially target transition mutations for repair (Schaaper & Dunn, 1987). As a consequence, we were expecting most of the mutations to be mismatch repair-insensitive transversions. Indeed, wild-type *dnaE* and *dnaE*_S759C have a high percentage of transversions, 90% and 93%, respectively (Table 5). In contrast, the *dnaE*_S759N and *dnaE*_S759T spectrum exhibited more transitions than transversions (~55% vs. ~45%). The increase in transition mutations with *dnaE*_S759N and *dnaE*_S759T alleles is likely due to the high levels of mutagenesis that overwhelms the mismatch repair machinery (Schaaper & Radman, 1989).

In addition to variability in transitions *vs*. transversions, the pattern of base substitutions in the *rpoB* locus also differ between wild-type *dnaE* and the three variants (Tables 5 and S2 and Figure 3). For example, the predominant mutation in wild-type *dnaE* is AT→CG (~58%, mostly accumulated within hot spots at positions 1687, 1714, and 1715, Figure 3A), yet this mutagenic event comprises just 2.7%, 0.9% and 0.3% in *dnaE*_S759C, *dnaE*_S759N and *dnaE*_S759T strains, respectively (Figure 3B-D). Even in the DNA site especially prone to undergo base changes in all strains tested (position 1714, Figure 4), the types of mutations recovered from wild-type *dnaE* and the three variants were different. In the strain with wild-type *dnaE*, the majority of mutations found at position 1714 were AT→CG transversions, while in other strains they were AT→ TA substitutions (Figure 3). Other notable differences include a dramatic increase in the occurrence of AT→TA transversions in the *dnaE*_S759C strain (~85%, most prominent at three mutagenic peaks at positions 1547, 1577, and 1714, Figure 3B). The types of base changes and mutagenic hot-spots in the *dnaE*_S759N and *dnaE*_S759T strains were very similar to each other (Table 5 and Figure 3C and D), with the exception of

additional CG→TA transitions at positions 1546 and 1691 in the *dnaE*_S759T strain (Figure 3D). We speculate that differences in the types of base substitutions are directly due to the altered misincorporation specificity of the individual *dnaE* allele.

2.7 | Increased ribonucleotide incorporation promoted by *dnaE*_S759C, *dnaE*_S759N and *dnaE*_S759T

To determine the impact of the three S759 variants on ribonucleotide incorporation, we performed alkaline gel electrophoresis of RNase H2-treated genomic DNA (Figure 4). No change in migration was observed for *rnhB*_wt strains due to efficient RER (Figure 4 and S6), whereas RER-deficient Δ*rnhB* strains showed increased fragmentation, indicative of the presence of more genome-embedded ribonucleotides, consistent with previous findings (Kouzminova et al., 2017; Zatopek et al., 2019). For Δ*rnhB* cells expressing the three *dnaE*_S759 variants the number of embedded ribonucleotides was further elevated. The smallest change in the fragmentation pattern was observed for *dnaE*_S759T; *dnaE*_S759C gave an intermediate effect, while the largest effect was observed for *dnaE*_S759N (Figure 4 and S6).

Using the densitometry measurements after alkaline gel electrophoresis, we calculated the frequency of embedded ribonucleotides in genomic DNA from *ΔrnhB E. coli* to be 49 ± 9.6 ribonucleotides per million bases (mean ± SD, n = 8 independent experiments) (Table 6), which is in line with previous experiments using alternative methods that estimated between 20 and 130 embedded ribonucleotides per million bases (Kouzminova et al., 2017; Cronan et al., 2019; Zatopek et al., 2019). Importantly,

our analyses showed significantly increased ribonucleotide incorporation rates as a result of the *dnaE* S759T, S759C and S759N mutations. We estimate the increase relative to wild-type *dnaE* to be 1.8, 2.3 and 8.4-fold respectively, with misincorporation in the *ΔrnhB dnaE*_S759N strain as high as one ribonucleotide every ~2.5 kb (Figure 4 and Table 6).

3 | Discussion

The aim of the current study was to make a series of *E. coli* pol IIIα variants with amino acid substitutions at the presumptive steric gate residue, H760, and the adjacent residue, S759. Given their location in the active site of the enzyme, we anticipated that some would have effects on base and/or sugar selection during replication. We originally envisaged being able to construct 38 "active site" mutants in the pol IIIα subunit. However, after repeated cloning attempts, we were only able to generate 9 new variants. We assume that our inability to make the remaining 29 possible variants is due to synthetic-lethality of the strain when it is transformed with the plasmid-encoded mutant. Three of the novel mutants were located at H760, but Western blots of the α-subunit encoded *dnaE*_H760F, *dnaE*_H760Q and *dnaE*_H760S indicate that they are highly unstable and/or poorly expressed.

All six S759 variants expressed the α-subunit at approximately similar levels and were able to complement the temperature sensitivity of the *dnaE486* allele. Three plasmid-encoded variants, S759C, S759N and S759T also exhibited modest to substantial increases in spontaneous mutagenesis *in vivo.* The alleles were moved to the *E. coli* chromosome to avoid any possible plasmid-encoded phenotypic artifacts and

subjected to a variety of *in vivo* assays to determine their ability to misincorporate nucleotides with an incorrect base or sugar.

The *E. coli* strains used in these studies lack DNA pol II, IV and V, so any replication associated phenotypes can only be attributed to the remaining DNA polymerases: wild-type pol I, or the different pol IIIα variants. We also chose to conduct our studies in a mismatch repair proficient background, so that we could study the effects of the α-subunit S759 mutants on transition and transversion mutagenesis. The studies were enhanced through the comparison of phenotypes of strains with a wild-type repair proficient background to Δ*rnhB* strains lacking RNase HII, which is essential for the majority of RER; and/or a *dnaQ920* strain, which is severely compromised for exonucleolytic 3'$\rightarrow$5' proofreading. Phenotypes in the Δ*rnhB* background were hypothesized to reflect differences in the ability of the *dnaE* variant to incorporate ribonucleotides into the *E. coli* genome, whereas phenotypes in the *dnaQ920* background would reflect differences in accurate/erroneous base selection.

With all three mutant alleles, reduction in the proofreading activity of pol III resulted in a synergistic increase in spontaneous mutagenesis, indicating that errors generated by the mutant polymerases are normally subject to ε-dependent proofreading. Analysis of the genomic DNA fragmentation pattern based on ribonucleotide-induced alkali sensitivity in the Δ*rnhB* background revealed that *dnaE*_S759N and to a lesser extent *dnaE*_S759C and *dnaE*_S759T have reduced ribonucleotide discrimination (Figure 4). However, inactivation of RER by the Δ*rnhB* allele had a minimal effect on levels of spontaneous mutagenesis promoted by the three *dnaE* alleles. Such phenotypes are likely to be expected for the *dnaE*_S759T or *dnaE*_S759C alleles that exhibit limited

ribonucleotide incorporation. However, the fact that there was no difference in the levels of spontaneous mutagenesis in the Δ*rnhB dnaE*_S750N strain implies that RER is unlikely to concomitantly remove dNTPs misincorporated by *dnaE*_S759N.

Based upon the data presented here, we suggest that the three *dnaE* mutants have differential phenotypes regarding base and sugar selection.

***dnaE*_S759C**: This variant exhibited a low spontaneous mutator activity, even in a *dnaQ920* background (Table 4). We therefore conclude that the *dnaE*_S759C variant maintains a high degree of base selectivity. In contrast, analysis of ribonucleotide incorporation (Figure 4 & Table 6), indicates that it incorporates 2.3-fold higher levels of ribonucleotides compared to wild-type *dnaE*, indicating that sugar discrimination in this variant is at least partially compromised.

***dnaE*_S759T**: This variant exhibited a low spontaneous mutator activity in a proofreading proficient (*dnaQ*_wt) background, but a high mutator phenotype in a *dnaQ920* background (Table 4, Figure 3, S4 and S5). This indicates that base errors generated by *dnaE*_S759T are normally efficiently proofread *in vivo*. Furthermore, the spectrum of *rpoB* mutations was unlike that of wild-type *dnaE* (Figure 3) and exhibited a substantial increase in transition mutations (Table 5). We therefore conclude that the *dnaE*_S759T variant has much lower base fidelity than wild-type *dnaE*. Similar to *dnaE*_S759C, ribonucleotide incorporation (Figure 4 & Table 6), is elevated ~1.8-fold compared to wild-type *dnaE*, indicating that in addition to very low base selectivity, sugar discrimination is partially compromised.

***dnaE*_S759N**: This variant exhibited spontaneous mutator activity that was similar to *dnaE*_S759T (Table 4, Figure 3, S4 and S5). However, analysis of ribonucleotide

incorporation promoted by *dnaE*_S759N indicates that sugar selectivity is severely compromised (Figure 4 & Table 6). As a result, the *dnaE*_S759N mutant incorporates ribonucleotides at an ~8.4-fold higher rate than wild-type *dnaE*, which equates to the incorporation of an errant ribonucleotide every ~2.5 kb in the *E. coli* genome. Thus, the *dnaE*_S759N allele is compromised for both base and sugar discrimination.

3.1 | Structural basis for the observed phenotypes of *dnaE*_S759C, *dnaE*_S759N and *dnaE*_S759T

The three *dnaE*_S759 alleles are adjacent to the steric gate that we have identified as H760 (Figure 1). We assume that the various phenotypes observed *in vivo* are due to direct changes in the ability of the α-subunit to misincorporate dNTPs and/or NTPs. The H760 residue is in the "O" helix of the polymerase in the finger domain, both of which undergo conformational changes from an "open" state in the absence of an incoming dNTP to a "closed" state when the polymerase is ready to incorporate a dNTP (Doublie *et al.*, 1999; Evans et al., 2008). S759 butts against the polymerase Palm domain. When S759 is changed to Cys, Thr or Asn, it causes steric clashes with the Palm domain, which contains the catalytic triad and metal ion (Figure 1). The most likely scenario is that the O helix does not close properly on the replicating base pair, thus loosening base selection and/or sugar discrimination.

3.2 | Future considerations

To the best of our knowledge, this is the first time that active site mutants of the α-catalytic subunit of pol III characterized by differential phenotypes regarding base and sugar discrimination as well as different mutational specificity have been reported. We believe that these novel mutants, due to their versatility, provide us with new tools and open new possibilities to study how *E. coli* normally maintains high fidelity replication and avoids the deleterious consequences of errant ribonucleotide misincorporation. Such studies will also need to be accompanied by the determination of the structures of appropriate enzyme–substrate complexes, which should provide an explanation of how modification of the active site architecture affects the substrate specificities and characteristics of each polymerase variant. We also plan to carry out detailed biochemical analysis of the purified α-subunits *in vitro* in the context of α-alone, pol III core and pol III holoenzyme, so as to elucidate novel features of the structural and molecular mechanisms that give rise to the differential phenotypes of the S759 mutants *in vivo*. Last, but not least, we hope that pol III variants characterized by different base and sugar fidelities will help us to determine whether prokaryotic cells employ the same set of repair pathways for cleansing genomic DNA of ribonucleotides incorporated by replicative and translesion DNA polymerases.

4 ｜EXPERIMENTAL PROCEDURES

4.1 ｜ Bacterial strains and plasmids

Plasmids used in this study are described in Table 1.

Most of the *E. coli* K-12 strains used in this study are derivatives of RW732 (full genotype: *thr-1 araD139* Δ(*gpt-proA*)*62 lacY1 tsx-33 supE44 galK2 hisG4 rpsL31 xyl-5 mtl-1 argE3 thi-1 sulA211* Δ*umuDC596*::*ermGT* Δ*araD-polB*::Ω Δ*dinB61*::*ble)* (Table 2). All derivatives of RW732 were made by standard methods of P1 transduction using P1*vir* (Table 2).

Where noted, bacteria were grown on LB agar plates containing 20 μg ml^{-1} chloramphenicol; 15 μg ml^{-1} tetracycline; 25 μg ml^{-1} zeocin; 30 μg ml^{-1} kanamycin; 20 μg ml^{-1} spectinomycin; or 100 μg ml^{-1} rifampicin.

4.2 | Construction of a low-copy number plasmid expressing pol III core

The genes encoding the α, ε and θ subunits (*dnaE, dnaQ* and *holE*, respectively) were codon optimized for expression in *E. coli* (Genscript) and synthesized as gene cassettes with appropriate 5' and 3' restriction enzyme sites for subsequent subcloning. The starting vector was the low copy number ampicillin vector pJM975 (Frank *et al.*, 2012). The Lac repressible-IPTG-inducible pTrc promoter (de Boer et al., 1983) was first cloned into pJM975 as a 184 bp *Eco*RI-*Nde*I fragment. Next, the *dnaQ920* gene with an R56W substitution in *dnaQ* was cloned into this vector as a 790 bp *Nde*I-*Bam*HI fragment, to generate pJM1048, which expresses N-terminal His-tagged DnaQ920. *holE* and the 5' end of the FLAG-tagged *dnaE* gene was subsequently subcloned into pJM1048 as a 2,540 bp *Bam*HI-*Bpu*10I fragment. The 3' end of the *dnaE* gene was then subcloned as a 1,351 bp *Bpu*10I-*Xho*I fragment, so as to reconstruct the full-length *dnaE* gene and generate the pol III core destination vector, pJM1260 (Figure S1).

4.3 | Use of a temperature sensitive *dnaE486* strain to determine if plasmid encoded *dnaE*s are functionally active

To determine if the plasmid encoded *dnaE* mutant alleles are functionally active, plasmid DNAs were transformed into RW1138, which harbors the temperature sensitive *dnaE486* allele (S885P) and grown at permissive temperature of 30°C. Transformants were then grown in liquid culture at 30°C overnight. The following morning, serial dilutions of the individual cultures were made and plated on LB agar plates at permissive (30°C) and non-permissive (43°C) temperatures. Plasmid encoded *dnaE* alleles were deemed to be fully functional if equal numbers of viable colonies were obtained at both permissive and non-permissive temperatures.

4.4 | Western blots

RW1138 (*dnaE486*ts) (Table 2) harboring pJM1260 (wild-type *dnaE*), or S759 *dnaE* variants (Table 1) were grown overnight at 30°C in LB medium plus appropriate antibiotics. The next morning, cultures were diluted 1:100 in fresh LB and grown with aeration at 30°C until they reached an OD_{600} of ~0.5. Cells were then; (i) centrifuged; (ii) resuspended in 1X NuPAGE™ LDS sample buffer (Invitrogen, NP0007; 106 mM Tris·HCl, 141 mM Tris base, 2% LDS, 10% glycerol, 0.51 mM EDTA, 0.22 mM SERVA blue G250, 0.175 mM phenol red, pH 8.5) containing 2% β-mercaptoethanol; (iii) immediately frozen in dry ice; (iv) lysed by multiple freeze-thaw cycles; and (v) heated for 10 mins at 70°C. Extracts were applied to a 4-12% NuPAGE Bis-Tris gel (Invitrogen, NP0321). After

separation, proteins were transferred to an Invitrolon PVDF membrane (Invitrogen, LC2005) using standard Western blot protocols. The membrane was incubated overnight with rabbit polyclonal antibodies (1:1,000 dilution) raised against the α-θ-ε subunits of pol III core (Covance, PA). The membrane was then incubated with secondary goat anti-rabbit alkaline phosphatase conjugated antibodies (1:10,000 dilution) (BioRad, 1706518) and visualized using the Tropix CDP-Star assay (Applied Biosystems, T2306). Images were captured on a FluorChem HD2 imaging system (ProteinSimple).

4.5 | His$^+$ and Gal$^+$ reversion assays

Quantitative assays: To assay the effect of plasmid encoded *dnaE* variants on spontaneous mutagenesis, RW1504 was freshly transformed with one of the pJM1260 *dnaE*-variant plasmids described in Table 1 and grown overnight at 30ºC on LB plates containing ampicillin. Five well-separated colonies were picked and inoculated into 5 mls LB/ampicillin medium and grown overnight with shaking at 30ºC. The next day, cultures were harvested by centrifugation and resuspended in an equal volume of SM medium (Sambrook *et al.*, 1989). Aliquots (100 μl) of each culture were plated in triplicate on Davis and Mingioli minimal plates (Davis & Mingioli, 1950) (1% agar, 0.4% glucose, 0.25 μg/ml thiamine, 0.7% potassium hydrogen orthophosphate, 0.2% potassium dihydrogen orthophosphate, 0.1% ammonium sulfate, 0.25% trisodium citrate and 0.01% magnesium sulfate) supplemented with 100 μg ml^{-1} of L-arginine, L-valine, L-leucine, L-threonine, L-leucine, L-proline and 1 μg ml^{-1} L-histidine. Plates were incubated at 30 ºC, or 39 ºC for 4 days, after which time His$^+$ revertant colonies were counted. The data shown are the mean

number of His^+ revertants data obtained from five individual colonies plated in triplicate for each strain.

Qualitative assays: To assay the effect of chromosomally encoded *dnaE* variants on spontaneous mutagenesis, the *dnaE* strain was grown overnight at 37ºC in LB medium containing the appropriate antibiotics. For His^+ reversion assays, the cultures were processed as described above. For Gal^+ reversion assays, overnight cultures were serial diluted in SM medium and ~50 to 100 bacteria plated on MacConkey agar base plates containing 1% galactose. Plates were incubated at 37ºC for 8 days before checking for the appearance of Gal^+ (red) papillae arising from the predominantly Gal^- (pink/orange) colony.

4.7 | Moving *dnaE_S759* alleles to the *E. coli* chromosome

The introduction of the *dnaE*_S759C, *dnaE*_S759N and *dnaE*_S759T alleles into the essential *dnaE* gene of *E. coli* MG1655 was performed according to Kim *et al*., with minor changes (Kim et al., 2014). In a first recombineering step, a linear mutation cassette was introduced via Red/ET recombination using the plasmid pALFIRE (Rivero-Müller *et al.*, 2007) into the chromosomal *dnaE* gene resulting in a hybrid *dnaE* gene ($dnaE_{hybrid}$ in Figure S7) encoding for a fully functional DNA polymerase III α-subunit. The mutation cassette consists of i) a 50 to 100 bp homology arm corresponding to the wild-type *dnaE* gene ($dnaE_{wt}$ in Figure S7), ii) a fragment encoding an alternative nucleotide sequence of the *dnaE* gene ($dnaE_{alt}$ in Figure S7) containing the desired point mutation at position 759, iii) an I-*Sce*I restriction site, iv) a chloramphenicol

selection marker (*cat* in Figure S7), v) a second fragment of the wild-type *dnaE* gene (Δ*dnaE*$_{wt}$) again containing the desired point mutation, followed by a second 50 to 100 bp long homology arm corresponding to the wild-type *dnaE* gene. In a second step, the selection marker was removed via RecA mediated repair using a I-*Sce*I restriction site as the selection strategy as described by Rivero-Müller *et al*., (Rivero-Müller et al., 2007). Finally, all clones were analyzed by DNA sequencing the modified region.

4.8 | Moving chromosomal *dnaE* alleles into *rnhB*$^{+}$/Δ*rnhB* and *dnaQ*$^{+}$/*dnaQ920* strains

Once the respective *dnaE* allele had been successfully moved to the chromosome of the wild-type *E. coli* strain, MG1655, we used conventional P1 transduction protocols to transfer the alleles into various repair-deficient genetic backgrounds (Table 2). To do so, we first linked the respective *dnaE* allele to the nearby *yafC727*::Kan allele from JW0198 (*E. coli* Genetic Stock Center), or the *yafC502*::Tn*10* allele from CAG18436 (*E. coli* Genetic Stock Center). Depending upon the genotype of the recipient strain and existing antibiotic resistance, transductants were either selected for resistance to kanamycin, or tetracycline, and then screened for co-transduction of the respective *dnaE* allele by colony PCR (see below), and/or phenotypic traits, such as conferring temperature resistance to the temperature sensitive *dnaE486ts* parental strain, or a spontaneous mutator phenotype (See Figs. S4 and S5). Due to their close genomic location (Figure S3), *yafC* also co-transduces with *rnhB* and *dnaQ* with high efficiency, and transductants were also screened by colony PCR to ascertain the status of the *rnhB* (*rnhB*_wt vs. Δ*rnhB*) and *dnaQ* (*dnaQ*_wt vs. *dnaQ920*) genes.

4.9 | Colony PCR assay to test for *rnhB*, *dnaE* and *dnaQ* alleles

A sterile pipette tip was used to pick a small quantity of bacteria from the purified P1 transductants and were then subject to PCR amplification. The primers used to amplify *rnhB* were rnhB_F-55 and rnhB_R773 (Table S1). PCR amplification was achieved by denaturation at 95°C for 5 min, followed by 60 cycles of 94°C for 30 s, 1 min at 59°C, 2 min at 72°C, followed by a final extension step at 72°C for 7 min. rnhB_F-55 and rnhB_R773 amplifies 711 bp of intact *rnhB*, 977 bp of Δ*rnhB*::Kan, or 204 bp of Δ*rnhB*::Kan^S.

The primers used to amplify *dnaE486* were EcdnaE486_F2378 and EcdnaE486_R2911 (Table S1). Amplification was achieved by denaturation at 95°C for 5 min, followed by 30 cycles of 94°C for 30 s, 1 min at 55°C, 2 min at 72°C, followed by a final extension step at 72°C for 7 min. The primers amplify a 571 bp region surrounding *dnaE486* [S885 (TCC) → P885 (CCC]. *dnaE486* is cut with *Sma*I/*Xma*I into 293 bp and 278 bp fragments. The primers used to amplify *dna*E_S759 alleles were dnaE_F2059 and dnaE_R2557 (Table S1). Amplification was achieved by denaturation at 95°C for 5 min, followed by 60 cycles of 94°C for 30 s, 1 min at 57°C, 2 min at 72°C, followed by a final extension step at 72°C for 7 min. dnaE_F2059 and dnaE_R2557 amplifies 537 bp surrounding the S759 codon. S759T and S759C both create a new *Bsl*I site. Digestion with *Bsl*I of S759T and S759C PCR amplicons gives a 225 bp fragment and a 299 bp fragment. The *dnaE*_S759N allele does not change a restriction site, so it was confirmed by DNA sequencing.

The primers used to amplify *dnaQ920* were EcdnaQ_F26 and EcdnaQ_R328 (Table S1). Amplification was achieved by denaturation at 95°C for 5 min, followed by 60 cycles of 94°C for 30 s, 1 min at 57°C, 1.5 min at 72°C, followed by a final extension step at 72°C for 7 min. The primers amplify a 341 bp amplicon which gives fragments of 156 bp and 185 bp after digestion with *Pvu*I. *dnaQ920* [R56 (CGG) →W56(TGG)] destroys the *Pvu*I site.

4.10 | Spectra of spontaneous mutations in *rpoB*

The mutation spectra were generated using the *rpoB* mutagenesis assay (Garibyan et al., 2003). A single pair of oligonucleotide primers were used for PCR amplification and a single primer for DNA sequencing because 88% of all *rpoB* mutations are localized in the central 202 bp region of the gene (Garibyan et al., 2003). *E. coli* strains were diluted from a frozen stock cultures so that the initial inoculum contained <1,000 viable cells. For spectral analysis of *rpoB* mutants, several hundred independent LB cultures were grown for 24 h at 37°C in parallel for each strain, and appropriate dilutions were plated on an LB agar plate containing 100 μg/ml rifampicin. Using a pipette tip one colony was picked randomly from each plate to ensure independence of the mutants. About 400 independent Rif resistant colonies were obtained for each strain and subjected to PCR in a 96-well micro-titer plate. An ~1 kb central region of the *rpoB* gene was amplified using the PCR primers RpoB1 and RpoF1 (Table S1) by denaturation at 95°C for 3 min, followed by 30 cycles of 94°C for 30 s, 1 min at 59°C, 2 min at 72°C, followed by a final extension step at 72°C for 7 min. The ~200 bp target region of *rpoB* in each PCR amplicon was sequenced by TACGen (Richmond, CA) Genomics

using WOG923AP01 primer (5′-CAG TTC CGC GTT GGC CTG-3′). Only base-pair substitutions occurring between positions 1,516 and 1,717 of the *rpoB* gene were considered during data analysis. Nucleotide sequences obtained were aligned and analyzed using the ClustalW multiple sequence alignment program (Hinxton, UK).

4.11 | Fluctuation assay for determination of forward mutation rates

For the fluctuation analysis, 15-57 cultures were inoculated with single colonies and grown overnight (~18h) at 37°C. Aliquots (100 µl) of each overnight culture, undiluted or diluted (10-fold), were plated on agar plates containing 100 µg ml^{-1} of rifampicin and incubated for 24-36 hours at 37°C. To determine the colony forming units (CFU), 50 µl of appropriate dilutions of the same cultures were plated on LB plates and incubated for 18-24 hours at 37°C. Mutation rates were calculated using the Maximum Likelihood Estimate (MLE) method (Sarkar *et al.*, 1992; Rosche & Foster, 2000) with a Newton-Raphson-type algorithm modified to account for partial plating, available in a free R package rSalvador (Zheng, 2015; Zheng, 2017). This calculator also computes 95% confidence intervals and employs Likelihood Ratio Test to calculate the statistical significance of the differences between mutation rates of various strains (Zheng, 2016). To account for multiple comparisons, the *P* values were adjusted using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995).

4.12 | Genomic DNA isolation

Genomic DNA used for alkaline gel electrophoresis was isolated from *rnhB*_wt strains; RW1628 (*dnaE*_wild-type), RW1714 (*dnaE*_S759C), RW1612 (*dnaE*_S759N) and RW1610 (*dnaE*_S759T) and Δ*rnhB* strains; RW1630 (*dnaE*_wild-type), RW1736 (*dnaE*_S759C), RW1718 (*dnaE*_S759N) and RW1624 (*dnaE*_S759T) (Table 2), using a previously described method (Ding *et al.*, 2015). In brief, *E. coli* from 1.5 ml of overnight culture was pelleted and resuspended in 200 µl of lysis buffer (2% Triton X-100, 1% SDS, 0.5 M NaCl, 10 mM Tris and 1 mM EDTA, pH 8.0). Cells were lysed by vortexing in the presence of 0.2 ml glass beads (0.4–0.6 mm diameter) and 200 µl of phenol (pH 7.9) for 2 min, then another min after adding 200 µl of TE buffer. After subsequent extractions with 400 µl of phenol:chloroform:isoamylalcohol (25:24:1) and 400 µl of chloroform, DNA was precipitated with ice-cold ethanol. DNA was quantified using the Qubit dsDNA BR Assay (Invitrogen), and quantity and quality checked by agarose gel electrophoresis (0.8%, 1x TAE).

4.13 | Alkaline gel electrophoresis

Genomic DNA was treated with RNase H2 and separated by alkaline gel electrophoresis, essentially as described (Benitez-Guijarro *et al.*, 2018). In brief, genomic DNA (250 ng) was treated with 1 pmol of purified recombinant human RNase H2 (Reijns *et al.*, 2011) and 0.25 µg of DNase-free RNase (Roche) for 1 h at 37°C in 100 µl reaction buffer (60 mM KCl, 50 mM Tris–HCl pH 8.0, 10 mM $MgCl_2$, 0.01%Triton X-100). Nucleic acids were ethanol precipitated and separated by alkaline gel electrophoresis (0.7% agarose, 50 mM NaOH, 1 mM EDTA). After electrophoresis, the gel was neutralized in 0.7 M Tris–HCl pH 8.0, 1.5 M NaCl and stained with SYBR Gold

(Invitrogen). Images were taken using the FLA-5100 imaging system (Fujifilm), and densitometry plots generated using AIDA Image Analyzer (Raytest).

4.14 | Quantification of genome-embedded ribonucleotides

Numeric analysis was performed in R (version 3.5.2) and Microsoft Excel 2016. Plotting and statistical tests were carried out in GraphPad Prism (version 9.1.1). The number of genome-embedded ribonucleotides was estimated using a combination of previously described methods (Reijns *et al.*, 2012; Uehara *et al.*, 2018). Starting from the densitometric histograms per lane after alkaline gel electrophoresis, background intensity was uniformly subtracted and smoothed by fitting the smooth.spline function with 40 degrees of freedom in R. Peaks in the molecular weight marker lanes (NEB Quick-Load 1 kb Extend DNA ladder) were identified under supervision and the linear model lm(y~log(x)) fitted to produce electrophoretic distance (y; mean of all marker lanes per gel) to fragment size (x) calibration curves. The resulting model was then applied to calculate the fragment size (sz) per electrophoresis distance interval, and fragment count per interval (n_{sz}) estimated as $n_{sz} = I_{sz} / sz$, with I_{sz} the densitometric intensity for the interval. To avoid fragment count errors resulting from noise near the bottom of the alkaline gel, where small changes in staining intensity would result in relatively large changes in the inferred number of small fragments, a cut-off electrophoretic distance (d_{max}) was introduced, with the corresponding molecular weight as the minimum fragment size (sz_{min}). The sum of n_{sz} in all intervals (Σn_{sz}) down to d_{max} and the sum across the same intervals for $sz \cdot n_{sz}$ ($\Sigma(sz \cdot n_{sz})$) were then used to determine a preliminary estimate of mean fragment size for each sample: $\overline{sz} = \Sigma(sz \cdot n_{sz})/$

Σn_{sz}, and a correction applied for small fragments migrated beyond d_{max}, to give a corrected mean fragment size $\overline{sz}_{corr} = \overline{sz} \cdot \exp(-sz_{min} / \overline{sz})$. For a genome of size G ($9.28 \cdot 10^6$ nt for MG1655) the number of breakpoints for each sample was then calculated as $N = G / \overline{sz}_{corr}$. Additional break points (i.e., an estimate of the number of embedded ribonucleotides per genome) in *ΔrnhB* DNA were computed as $Nribo = N_{\Delta rnhB} - N_{WT}$. As there was no significant difference in the corrected mean fragment sizes for the different *rnhB*+ strains, N_{WT} was taken as the mean of N for all *rnhB*+ samples per gel. To determine the statistical significance of differences in $\overline{sz}_{corr}$ or Nribo between samples across 6-8 independent experiments, an unpaired 2-sided t-test with Welch's correction was performed.

4.15 | Molecular Modeling

Although structures of *E. coli* pol IIIα in the apo and DNA-complex form have been reported (Lamers et al., 2006; Fernandez-Leiro *et al.*, 2015), a ternary-complex structure with DNA and incoming dNTP is not available yet. The ternary complex structure of the C-family DNA polymerase from *G. kaustophilus* (PDB: 3F2C) (Evans et al., 2008) offers the best resolution (2.5 Å) view of the catalytic center engaging in DNA synthesis and is a relevant model for *E. coli* pol IIIα because of the conserved amino acid sequence in the region (24% identity and 39% similarity). Indeed, the apo structure of *E. coli* DNA pol IIIα (PDB: 2HNH) (Lamers et al., 2006) was superimposable with the catalytic core of *G. kaustophilus* PolC, which includes residues 825 to 1,102 encompassing the palm and thumb domains. The structure superposition confirms the

sequence alignment, and a model of the *E. coli* DNA pol IIIα ternary complex was thus generated as shown in Figure 1.

Acknowledgements

Funding for this work was provided by the National Institute of Child Health and Human development/ National Institutes of Health Intramural Research Program (RW); National Institute of Diabetes and Digestive and Kidney Diseases/ National Institutes of Health Intramural Research Program (WY); National Science Centre, Poland - 2015/18/M/NZ3/00402 (KMD); 2019/35/N/NZ1/03402 (KŁ); and UK Medical Research Council Human Genetics Unit core grant (MRC, U127580972) to Andrew P. Jackson (MAMR). We thank Nicholas Ashton, LGI/NIH, for statistical analysis presented in Figure 2; and Alan O'Callaghan, MRC-HGU, for advice on the use of R. We also thank Iwona Fijalkowska and the late Piotr Jonczyk (Polish Academy of Science) for helpful discussions and critical reading of the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest with the content of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Orcid IDs:
Alexandra Vaisman: 0000-0002-2521-1467
Erin Walsh: 0000-0003-0299-5315
John P. McDonald: 0000-0003-2482-148X
Martin A. M. Reijns: 0000-0002-5048-2752
Krystian Łazowski: 0000-0001-5163-7026
Kristiniana C. Moreno: None
Dominic R. Quiros: 0000-0003-2613-5192
Marlen Schmidt: None
Harald Kranz: None
Wei Yang: 0000-0002-3591-2195
Karolina Makiela-Dzbenska: 0000-0002-1224-7444
Roger Woodgate: 0000-0002-2521-1467

AUTHOR CONTRIBUTIONS

Conceptualization: JPM, RW

Data Curation: AV, MAMR, KŁ, KMD

Formal Analysis: AV, EW, MAMR, KŁ, KMD

Funding Acquisition: KMD, KŁ, WY, RW

Investigation: AV, KŁ, MAMR, EW, JPM, KCM, DRQ, MS, WY, KMD, RW

Visualization: AV, EW, MAMR, KŁ, KMD, WY

Writing-Original Draft: AV, EW, JPM, MAMR, WY, KMD, RW

Writing-Review and Editing: AV, KŁ, MAMR, EW, JPM, KCM, DRQ, MS, HK, WY, KMD, RW

Abbreviated Summary

The accurate replication of every living organisms' genome is achieved by high fidelity DNA polymerases. The enzymes not only need to choose the nucleotide with the correctly paired base (G, A, T or C), but also the right sugar (deoxyribonucleotide, not ribonucleotide). We report here, the construction and characterization of three novel active site mutants of the α-catalytic subunit of *E. coli* DNA polymerase III that have altered phenotypes with regard to base and sugar discrimination.

REFERENCES

Benitez-Guijarro, M., Lopez-Ruiz, C., Tarnauskaite, Z., Murina, O., Mian Mohammad, M., Williams, T.C., Fluteau, A., Sanchez, L., Vilar-Astasio, R., Garcia-Canadas, M., Cano, D., Kempen, M.H., Sanchez-Pozo, A., Heras, S.R., Jackson, A.P., Reijns, M.A. & Garcia-Perez, J.L., (2018) RNase H2, mutated in Aicardi-Goutieres syndrome, promotes LINE-1 retrotransposition. *EMBO J* **37**.

Benjamini, Y. & Hochberg, Y., (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**: 289-300.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J. & Rabinowitz, J.D., (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* **5**: 593-599.

Brown, J.A. & Suo, Z., (2011) Unlocking the sugar "steric gate" of DNA polymerases. *Biochemistry* **50**: 1135-1142.

Churchward, G., Belin, D. & Nagamine, Y., (1984) A pSC101-derived plasmid which shows no sequence homology to other commonly used cloning vectors. *Gene* **31**: 165-171.

Cronan, G.E., Kouzminova, E.A. & Kuzminov, A., (2019) Near-continuously synthesized leading strands in *Escherichia coli* are broken by ribonucleotide excision. *Proc Natl Acad Sci U S A* **116**: 1251-1260.

Curti, E., McDonald, J.P., Mead, S. & Woodgate, R., (2009) DNA polymerase switching: effects on spontaneous mutagenesis in *Escherichia coli*. *Mol Microbiol* **71**: 315-331.

Davis, B.D. & Mingioli, E.S., (1950) Mutants of *Escherichia coli* requiring methionine or vitamin B12. *J Bacteriol* **60**: 17-28.

de Boer, H.A., Comstock, L.J. & Vasser, M., (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci U S A* **80**: 21-25.

DeLucia, A.M., Chaudhuri, S., Potapova, O., Grindley, N.D. & Joyce, C.M., (2006) The properties of steric gate mutants reveal different constraints within the active sites of Y-family and A-family DNA polymerases. *J Biol Chem* **281**: 27286-27291.

Ding, J., Taylor, M.S., Jackson, A.P. & Reijns, M.A., (2015) Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. *Nat Protoc* **10**: 1433-1444.

Donigan, K.A., McLenigan, M.P., Yang, W., Goodman, M.F. & Woodgate, R., (2014) The steric gate of human DNA polymerase ι regulates ribonucleotide incorporation and deoxyribonucleotide fidelity. *J Biol Chem* **289**: 9136-9145.

Doublie, S., Sawaya, M.R. & Ellenberger, T., (1999) An open and closed case for all polymerases. *Structure* **7**: R31-35.

Evans, R.J., Davies, D.R., Bullard, J.M., Christensen, J., Green, L.S., Guiles, J.W., Pata, J.D., Ribble, W.K., Janjic, N. & Jarvis, T.C., (2008) Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc Natl Acad Sci U S A* **105**: 20695-20700.

Fernandez-Leiro, R., Conrad, J., Scheres, S.H. & Lamers, M.H., (2015) cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *Elife* **4**.

Fijalkowska, I.J., Dunn, R.L. & Schaaper, R.M., (1993) Mutants of *Escherichia coli* with increased fidelity of DNA replication. *Genetics* **134**: 1023-1030.

Fijalkowska, I.J. & Schaaper, R.M., (1993) Antimutator mutations in the α subunit of *Escherichia coli* DNA polymerase III: identification of the responsible mutations and alignment with other DNA polymerases. *Genetics* **134**: 1039-1044.

Fijalkowska, I.J. & Schaaper, R.M., (1996) Mutants in the Exo I motif of *Escherichia coli dnaQ*: defective proofreading and inviability due to error catastrophe. *Proc Natl Acad Sci U S A* **93**: 2856-2861.

Fijalkowska, I.J., Schaaper, R.M. & Jonczyk, P., (2012) DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair. *FEMS Microbiol Rev* **36**: 1105-1121.

Fossum, S., Crooke, E. & Skarstad, K., (2007) Organization of sister origins and replisomes during multifork DNA replication in *Escherichia coli*. *EMBO J* **26**: 4514-4522.

Frank, E.G., McDonald, J.P., Karata, K., Huston, D. & Woodgate, R., (2012) A strategy for the expression of recombinant proteins traditionally hard to purify. *Analytical Biochemistry* **429**: 132-139.

Garibyan, L., Huang, T., Kim, M., Wolff, E., Nguyen, A., Nguyen, T., Diep, A., Hu, K., Iverson, A., Yang, H. & Miller, J.H., (2003) Use of the *rpoB* gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair* **2**: 593-608.

Gawel, D., Jonczyk, P., Fijalkowska, I.J. & Schaaper, R.M., (2011) dnaX36 Mutator of *Escherichia coli*: effects of the τ subunit of the DNA polymerase III holoenzyme on chromosomal DNA replication fidelity. *J Bacteriol* **193**: 296-300.

Gawel, D., Pham, P.T., Fijalkowska, I.J., Jonczyk, P. & Schaaper, R.M., (2008) Role of accessory DNA polymerases in DNA replication in *Escherichia coli*: analysis of the *dnaX36* mutator mutant. *J Bacteriol* **190**: 1730-1742.

Georgescu, R.E., Kurth, I. & O'Donnell, M.E., (2011) Single-molecule studies reveal the function of a third polymerase in the replisome. *Nat Struct Mol Biol* **19**: 113-116.

Hiratsuka, K. & Reha-Krantz, L.J., (2000) Identification of *Escherichia coli dnaE* (*polC*) mutants with altered sensitivity to 2',3'-dideoxyadenosine. *J Bacteriol* **182**: 3942-3947.

Joyce, C.M., (1997) Choosing the right sugar: how polymerases select a nucleotide substrate. *Proc Natl Acad Sci U S A* **94**: 1619-1622.

Kim, D.R. & McHenry, C.S., (1996) *In vivo* assembly of overproduced DNA polymerase III. Overproduction, purification, and characterization of the α, α-ε, and α-ε-θ subunits. *J Biol Chem* **271**: 20681-20689.

Kim, D.R., Pritchard, A.E. & McHenry, C.S., (1997) Localization of the active site of the α subunit of the *Escherichia coli* DNA polymerase III holoenzyme. *J Bacteriol* **179**: 6721-6728.

Kim, J., Webb, A.M., Kershner, J.P., Blaskowski, S. & Copley, S.D., (2014) A versatile and highly efficient method for scarless genome editing in *Escherichia coli* and *Salmonella enterica*. *BMC Biotechnol* **14**: 84.

Kornberg, A. & Baker, T.A., (1992) *DNA replication*. W.H. Freeman & Co, New York.

Kouzminova, E.A., Kadyrov, F.F. & Kuzminov, A., (2017) RNase HII saves *rnhA* mutant *Escherichia coli* from R-Loop-associated chromosomal fragmentation. *J Mol Biol* **429**: 2873-2894.

Lamers, M.H., Georgescu, R.E., Lee, S.G., O'Donnell, M. & Kuriyan, J., (2006) Crystal structure of the catalytic α subunit of *E. coli* replicative DNA polymerase III. *Cell* **126**: 881-892.

Langston, L.D., Indiani, C. & O'Donnell, M., (2009) Whither the replisome: emerging perspectives on the dynamic nature of the DNA replication machinery. *Cell Cycle* **8**: 2686-2691.

Maki, H., Akiyama, M., Horiuchi, T. & Sekiguchi, M., (1990) Molecular mechanisms of replicational fidelity in *Escherichia coli*. *Basic Life Sci* **52**: 299-308.

Maki, H., Mo, J.Y. & Sekiguchi, M., (1991) A strong mutator effect caused by an amino acid change in the α subunit of DNA polymerase III of *Escherichia coli*. *J Biol Chem* **266**: 5055-5061.

Makiela-Dzbenska, K., Jonczyk, P., Schaaper, R.M. & Fijalkowska, I.J., (2011) Proofreading deficiency of Pol I increases the levels of spontaneous *rpoB* mutations in *E. coli*. *Mutat Res* **712**: 28-32.
Makiela-Dzbenska, K., Maslowska, K.H., Kuban, W., Gawel, D., Jonczyk, P., Schaaper, R.M. & Fijalkowska, I.J., (2019) Replication fidelity in *E. coli*: Differential leading and lagging strand effects for *dnaE* antimutator alleles. *DNA Repair* **83**: 102643.
McHenry, C.S., (2003) Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences. *Mol Microbiol* **49**: 1157-1165.
McHenry, C.S., (2011) DNA replicases from a bacterial perspective. *Annu Rev Biochem* **80**: 403-436.
McInerney, P., Johnson, A., Katz, F. & O'Donnell, M., (2007) Characterization of a triple DNA polymerase replisome. *Mol Cell* **27**: 527-538.
McInerney, P. & O'Donnell, M., (2004) Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. *J Biol Chem* **279**: 21543-21551.
Mo, J.Y., Maki, H. & Sekiguchi, M., (1991) Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. *J Mol Biol* **222**: 925-936.
Mok, M. & Marians, K.J., (1987) The *Escherichia coli* preprimosome and DNA B helicase can form replication forks that move at the same rate. *J Biol Chem* **262**: 16644-16654.
Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundstrom, E.B., Johansson, E., Chabes, A. & Kunkel, T.A., (2010a) Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol* **6**: 774-781.
Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundstrom, E.B., Burgers, P.M., Johansson, E., Chabes, A. & Kunkel, T.A., (2010b) Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A* **107**: 4949-4954.
O'Donnell, M., (2006) Replisome architecture and dynamics in *Escherichia coli*. *J Biol Chem* **281**: 10653-10656.
Oller, A.R., Fijalkowska, I.J. & Schaaper, R.M., (1993) The *Escherichia coli galK2* papillation assay: its specificity and application to seven newly isolated mutator strains. *Mutat Res* **292**: 175-185.
Oller, A.R. & Schaaper, R.M., (1994) Spontaneous mutation in *Escherichia coli* containing the *dnaE911* DNA polymerase antimutator allele. *Genetics* **138**: 263-270.
Parasuram, R., Coulther, T.A., Hollander, J.M., Keston-Smith, E., Ondrechen, M.J. & Beuning, P.J., (2018) Prediction of active site and distal residues in *E. coli* DNA polymerase III α polymerase activity. *Biochemistry* **57**: 1063-1072.
Pham, P.T., Zhao, W. & Schaaper, R.M., (2006a) Mutator mutants of *Escherichia coli* carrying a defect in the DNA polymerase III θ subunit. *Mol Microbiol* **59**: 1149-1161.
Pham, P.T., Zhao, W. & Schaaper, R.M., (2006b) Mutator mutants of *Escherichia coli* carrying a defect in the DNA polymerase III τ subunit. *Mol Microbiol* **59**: 1149-1161.
Pomerantz, R.T. & O'Donnell, M., (2007) Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol* **15**: 156-164.
Pritchard, A.E. & McHenry, C.S., (1999) Identification of the acidic residues in the active site of DNA polymerase III. *J Mol Biol* **285**: 1067-1080.
Reijns, M.A., Bubeck, D., Gibson, L.C., Graham, S.C., Baillie, G.S., Jones, E.Y. & Jackson, A.P., (2011) The structure of the human RNase H2 complex defines key interaction interfaces relevant to enzyme function and human disease. *J Biol Chem* **286**: 10530-10539.
Reijns, M.A., Rabe, B., Rigby, R.E., Mill, P., Astell, K.R., Lettice, L.A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., Devenney, P.S., Sexton, D., Grimes, G., Holt, I.J., Hill, R.E., Taylor, M.S., Lawson, K.A., Dorin, J.R. & Jackson, A.P., (2012) Enzymatic removal

of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**: 1008-1022.
Reyes-Lamothe, R., Sherratt, D.J. & Leake, M.C., (2010) Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science* **328**: 498-501.
Rivero-Müller, A., Lajic, S. & Huhtaniemi, I., (2007) Assisted large fragment insertion by Red/ET-recombination (ALFIRE)-an alternative and enhanced method for large fragment recombineering. *Nucleic Acids Res* **35**: e78.
Rosche, W.A. & Foster, P.L., (2000) Determining mutation rates in bacterial populations. *Methods* **20**: 4-17.
Sambrook, J., Fritsch, E.F. & Maniatis, T., (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
Sarkar, S., Ma, W.T. & Sandri, G.H., (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica* **85**: 173-179.
Sassa, A., Yasui, M. & Honma, M., (2019) Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. *Genes Environ* **41**: 3.
Schaaper, R.M., (1993) The mutational specificity of two *Escherichia coli dnaE* antimutator alleles as determined from *lacI* mutation spectra. *Genetics* **134**: 1031-1038.
Schaaper, R.M., (1996) Suppressors of *Escherichia coli mutT*: anitimutators for DNA replication errors. *Mutat Res* **350**: 17-23.
Schaaper, R.M., (1998) Antimutator mutants in bacteriophage T4 and *Escherichia coli*. *Genetics* **148**: 1579-1585.
Schaaper, R.M. & Dunn, R.L., (1987) Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc Natl Acad Sci U S A* **84**: 6220-6224.
Schaaper, R.M. & Radman, M., (1989) The extreme mutator effect of *Escherichia coli mutD5* results from saturation of mismatch repair by excessive DNA replication errors. *EMBO J* **8**: 3511-3516.
Sevastopoulos, C.G. & Glaser, D.A., (1977) Mutator action by *Escherichia coli* strains carrying *dnaE* mutations. *Proc Natl Acad Sci U S A* **74**: 3497-3450.
Sugaya, Y., Ihara, K., Masuda, Y., Ohtsubo, E. & Maki, H., (2002) Hyper-processive and slower DNA chain elongation catalysed by DNA polymerase III holoenzyme purified from the *dnaE173* mutator mutant of *Escherichia coli*. *Genes Cells* **7**: 385-399.
Taft-Benz, S.A. & Schaaper, R.M., (1998) Mutational analysis of the 3'➛5' proofreading exonuclease of *Escherichia coli* DNA polymerase III. *Nucleic Acids Res* **26**: 4005-4012.
Taft-Benz, S.A. & Schaaper, R.M., (2004) The θ subunit of *Escherichia coli* DNA polymerase III: a role in stabilizing the epsilon proofreading subunit. *J Bacteriol* **186**: 2774-2780.
Uehara, R., Cerritelli, S.M., Hasin, N., Sakhuja, K., London, M., Iranzo, J., Chon, H., Grinberg, A. & Crouch, R.J., (2018) Two RNase H2 mutants with gifferential rNMP processing activity reveal a threshold of ribonucleotide tolerance for embryonic development. *Cell Rep* **25**: 1135-1145 e1135.
Vaisman, A., Kuban, W., McDonald, J.P., Karata, K., Yang, W., Goodman, M.F. & Woodgate, R., (2012) Critical amino acids in *Escherichia coli* responsible for sugar discrimination and base-substitution fidelity. *Nucleic Acids Res* **40**: 6144-6157.
Vaisman, A., McDonald, J.P., Huston, D., Kuban, W., Liu, L., Van Houten, B. & Woodgate, R., (2013) Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair. *PLoS Genet* **9**: e1003878.
Vaisman, A. & Woodgate, R., (2018) Ribonucleotide discrimination by translesion synthesis DNA polymerases. *Crit Rev Biochem Mol Biol* **53**: 382-402.
Vandewiele, D., Fernández de Henestrosa, A.R., Timms, A.R., Bridges, B.A. & Woodgate, R., (2002) Sequence analysis and phenotypes of five temperature sensitive mutator alleles

of *dnaE*, encoding modified α-catalytic subunits of *Escherichia coli* DNA polymerase III holoenzyme. *Mutat Res* **499**: 85-95.
Wechsler, J.A. & Gross, J.D., (1971) *Escherichia coli* mutants temperature-sensitive for DNA synthesis. *Mol Gen Genet* **113**: 273-284.
Wolff, E., Kim, M., Hu, K., Yang, H. & Miller, J.H., (2004) Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli rpoB* to assess the role of polymerase IV in spontaneous mutation. *J Bacteriol* **186**: 2900-2905.
Yanagihara, F., Yoshida, S., Sugaya, Y. & Maki, H., (2007) The *dnaE173* mutator mutation confers on the α subunit of *Escherichia coli* DNA polymerase III a capacity for highly processive DNA synthesis and stable binding to primer/template DNA. *Genes Genet Syst* **82**: 273-280.
Yao, N.Y. & O'Donnell, M., (2008) Replisome dynamics and use of DNA trombone loops to bypass replication blocks. *Mol Biosyst* **4**: 1075-1084.
Yao, N.Y. & O'Donnell, M., (2009) Replisome structure and conformational dynamics underlie fork progression past obstacles. *Curr Opin Cell Biol* **21**: 336-343.
Yao, N.Y., Schroeder, J.W., Yurieva, O., Simmons, L.A. & O'Donnell, M.E., (2013) Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc Natl Acad Sci U S A* **110**: 12942-12947.
Zatopek, K.M., Potapov, V., Maduzia, L.L., Alpaslan, E., Chen, L., Evans, T.C., Jr., Ong, J.L., Ettwiller, L.M. & Gardner, A.F., (2019) RADAR-seq: A RAre DAmage and Repair sequencing method for detecting DNA damage on a genome-wide scale. *DNA Repair* **80**: 36-44.
Zheng, Q., (2015) A new practical guide to the Luria-Delbrück protocol. *Mutat Res* **781**: 7-13.
Zheng, Q., (2016) Comparing mutation rates under the Luria-Delbrück protocol. *Genetica* **144**: 351-359.
Zheng, Q., (2017) rSalvador: An R Package for the Fluctuation Experiment. *G3 (Bethesda)* **7**: 3849-3856.

Table 1. Plasmids used in this study.

Plasmid	Relevant Characteristics	Source or Reference
pJM1260	Low-copy-number plasmid expressing codon optimized pol III core (α, θ, ε)	This study
pJM1260-*dnaE*_H760F	As pJM1260, but expressing *dnaE*_H760F	This study
pJM1260-*dnaE*_H760Q	As pJM1260, but expressing *dnaE*_H760Q	This study
pJM1260-d*naE*_H760S	As pJM1260, but expressing *dnaE*_H760S	This study
pJM1260-*dnaE*_S759A	As pJM1260, but expressing *dnaE*_S759A	This study
pJM1260-*dnaE*_S759C	As pJM1260, but expressing *dnaE*_S759C	This study
pJM1260-*dnaE*_S759G	As pJM1260, but expressing *dnaE*_S759G	This study
pJM1260-*dnaE*_S759N	As pJM1260, but expressing *dnaE*_S759N	This study
pJM1260-*dnaE*_S759T	As pJM1260, but expressing *dnaE*_S759T	This study
pJM1260-*dnaE*_S759V	As pJM1260, but expressing *dnaE*_S759V	This study
pALFIRE	Plasmid encoding for Red α/β and RecA expressed from the arabinose promoter and encoding for the I-*Sce*I restriction enzyme expressed from the anhydrotetracycline promoter.	(Rivero-Müller et al., 2007)

Table 2. *E. coli* strains used in this study.

Strain	Relevant Genotype	Source or Reference
P640	*dnaE* S759T	GeneBridges/Gen-H
P648	*dnaE* S759N	GeneBridges/Gen-H
P685	*dnaE* S759C	GeneBridges/Gen-H
JW0198	Δ*yafC727*::Kan	*E. coli* Genetic Stock Center
CAG18436	Δ*yafC502*::*Tn10*	*E. coli* Genetic Stock Center
RW1606	*dnaE* S759T Δ*yafC727*::Kan	P640 x P1. JW0198
RW1608	*dnaE* S759N Δ*yafC727*::Kan	P648 x P1. JW0198
RW1712	*dnaE* S759C Δ*yafC727*::Kan	P685 x P1. JW0198
RW1692	*dnaE* S759N Δ*yafC502*::Tn*10*	P640 x P1. CAG18436
RW1720	*dnaE* S759T Δ*yafC502*::Tn*10*	P648 x P1. CAG18436
RW1722	*dnaE* S759C Δ*yafC502*::Tn*10*	P685 x P1. CAG18436
RW1138[a]	*dnaE486ts* Δ*yafC502*::Tn*10*	LGI[b] stocks
RW1494[a]	Δ*rnhB782*::Kan *dnaE486ts*	LGI stocks
RW1504[a]	*rnhB*_wt *dnaE486ts* Δ*yafC502*::Tn*10 dnaQ920*	LGI Stocks
RW1604[a]	Δ*rnhB782 dnaE486ts* Δ*yafC502*::Tn*10 dnaQ920*	LGI Stocks
RW1726[a]	Δ*rnhB782*::Kan *dnaE486ts dnaQ920*	LGI Stocks
RW1628[a]	*rnhB*_wt *dnaE*_wt *dnaQ*_wt	LGI Stocks
RW1610[a]	*rnhB*_wt *dnaE* S759T Δ*yafC727*::Kan *dnaQ*_wt	RW1604 x P1. RW1606
RW1612[a]	*rnhB*_wt *dnaE* S759N Δ*yafC727*::Kan *dnaQ*_wt	RW1604 x P1. RW1608
RW1714[a]	*rnhB*_wt *dnaE* S759C Δ*yafC727*::Kan *dnaQ*_*wt*	RW1604 x P1. RW1712
RW1614[a]	*rnhB*_wt *dnaE*_wt Δ*yafC727*::Kan *dnaQ920*	RW1604 x P1. JW0198
RW1616[a]	*rnhB*_wt *dnaE* S759T Δ*yafC727*::Kan *dnaQ920*	RW1604 x P1. RW1606
RW1618[a]	*rnhB*_wt *dnaE* S759N Δ*yafC727*::Kan *dnaQ920*	RW1604 x P1. RW1608
RW1716[a]	*rnhB*_wt *dnaE* S759C Δ*yafC727*::Kan *dnaQ920*	RW1604 x P1. RW1712
RW1630[a]	Δ*rnhB782*::Kan *dnaE*_wt *dnaQ*_wt	LGI Stocks
RW1620[a]	Δ*rnhB782 dnaE*_wt Δ*yafC727*::Kan *dnaQ920*	RW1604 x P1. JW0198
RW1624[a]	Δ*rnhB782 dnaE* S759T Δ*yafC727*::Kan *dnaQ*_wt	RW1604 x P1. RW1606
RW1626[a]	Δ*rnhB782 dnaE* S759T Δ*yafC727*::Kan *dnaQ920*	RW1604 x P1. RW1606
RW1718[a]	Δ*rnhB782*::Kan *dnaE* S759N Δ*yafC502*::Tn*10 dnaQ*_wt	RW1494 x P1. RW1692
RW1736[a]	Δ*rnhB782*::Kan *dnaE* S759C Δ*yafC502*::Tn*10 dnaQ*_wt	RW1726 x P1. RW1722
EC7344	*dnaQ920* Δ*yafC502*::Tn*10*	LDRGS[c] Stocks
EC10539[a]	*dnaQ920* Δ*yafC502*::Tn*10*	RW1628 x P1. EC7344
EC10540[a]	Δ*rnhB782*::Kan *dnaE* S759N	RW1628 x P1. RW1718
EC10541[a]	Δ*rnhB782*::Kan *dnaE* S759C	RW1628 x P1. RW1736
EC10544[a]	Δ*rnhB782*::Kan *dnaE* S759C Δ*yafC502*::Tn*10 dnaQ920*	EC10539 x P1. EC10541
EC10545[a]	Δ*rnhB782*::Kan *dnaE* S759N Δ*yafC502*::Tn*10 dnaQ920*	EC10539 x P1. EC10540

[a:] *thr-1 Δ(argF-lac)169 tsx-33 supE44 galK2 hisG4 rpsL31 xyl-5 mtl-1 argE3 thi-1 sulA211 Δ(umuDC)596*::*ermGT ΔdinB61*::*ble ΔaraD-polB*::Ω

[b:] Laboratory of Genomic Integrity

[c:] Laboratory of DNA Replication and Genome Stability

Table 3. Viability of strains expressing *dnaE* steric gate variants from pJM1260[a].

***dnaE* variant**	**10^{-7} CFU/ml**	
	30ºC	**43ºC**
*dnaE*_wt	103 ± 7	107 ± 13
*dnaE*_H760F	234 ± 2	0 ± 0
*dnaE*_H760Q	314 ± 87	0 ± 0
*dnaE*_H760S	22 ± 13	0 ± 0
*dnaE*_S759A	83 ± 14	79 ± 19
*dnaE*_S759C	150 ± 34	158 ± 23
*dnaE*_S759G	149 ± 22	151 ± 20
*dnaE*_S759N	22 ± 4	18 ± 6
*dnaE*_S759T	41 ± 5	34 ± 5
*dnaE*_S759V	47 ± 20	40 ± 14

[a]: Viability assays were performed using *E. coli* RW1138 (Table 2), which in the absence of a functional *dnaE* gene, grows at 30°C, but not at 37°C, or higher. CFU; Colony forming unit. The values reported in the table are the average number of colonies obtained from three independent experiments (four plates each) ± standard error of the mean.

Table 4. Mutation rates of spontaneous rifampicin resistance in *dnaE* strains proficient- or deficient- in DnaQ and/or RNase HII activity.

Genotype	Rif mutation rate x 10^9			
	***dnaQ*_wt *rnhB*_wt**	***dnaQ920* *rnhB*_wt**	***dnaQ*_wt Δ*rnhB***	***dnaQ920* Δ*rnhB***
*dnaE*_wt	2.04 (1.30 – 2.96)	12.8 (10.3 – 15.5)	1.49 (0.92 – 2.23)	11.0 (8.6 – 13.6)
*dnaE*_S759C	3.44 (2.86 – 4.08)	44.8 (39.8 – 49.7)	2.43 (1.66 – 3.36)	32.8 (26.5 – 39.2)
*dnaE*_S759N	14.0 (11.6 – 16.3)	854 (757 – 953)	12.4 (9.5 – 15.4)	805 (731 – 876)
*dnaE*_S759T	6.11 (4.55 – 7.81)	263 (222 – 303)	7.47 (5.70 – 9.38)	225 (187 – 266)

Spontaneous *rpoB* mutation rates were measured in wild-type, *dnaQ920*, and *ΔrnhB* genetic backgrounds. The mutation rates and 95% confidence intervals (in brackets) were calculated as described in Materials & Methods (Zheng, 2017), with the population sizes of n=15–57 cultures for each strain.

Table 5. **Mutational changes in *rpoB* leading to rifampicin resistance of *E. coli* *dnaQ920* strains expressing *dnaE_wt* and *dnaE*_S759 variants**

bp change	***dnaE*_wt**	***dnaE*_S759C**	***dnaE*_S759N**	***dnaE*_S759T**
CG→GC	2 (0.7%)	5 (1.3%)	0 (0%)	0 (0%)
CG→AT	8 (2.6%)	15 (4.0%)	0 (0%)	5 (1.5%)
CG→TA	9 (3%)	12 (3.2%)	104 (31.8%)	106 (30.8%)
AT→TA	89 (29.5%)	316 (85.2%)	140 (42.8%)	152 (44.2%)
AT→CG	174 (57.6%)	10 (2.7%)	3 (0.9%)	1 (0.3%)
AT→GC	20 (6.6%)	13 (3.5%)	80 (24.5%)	80 (23.3%)
Transitions	29 (10.6%)	25 (6.7%)	184 (56.3%)	186 (54.1%)
Transversions	273 (90.4%)	346 (93.3%)	143 (43.7%)	158 (45.9%)
Total	302	371	327	344

Data shown in brackets are number of particular base substitutions calculated as a percent of total mutations, or the number of transitions or transversions calculated as a percent of total mutations.

Table 6. Ribonucleotides embedded in the genome of Δ*rnhB* strains expressing wild-type *dnaE* or *dnaE*_S759 variants[a]

Genotype	rN per genome	rN per Mb	Kb per rN	Fold difference[b]	p-value[c]
***dnaE*_wt**	457 ± 89	49 ± 9.6	21.1 ± 4.7	1.00	1.00
***dnaE*_S759C**	1049 ± 122	113 ± 13	9.0 ± 1.1	2.30	< 0.0001
***dnaE*_S759N**	3854 ± 589	415 ± 63	2.5 ± 0.4	8.44	< 0.0001
***dnaE*_S759T**	803 ± 167	86 ± 18	12.0 ± 2.3	1.76	0.0003

[a] Numbers shown: mean ± standard deviation of n = 6-8 independent measurements; based on an *E. coli* genome of 4.64 Mbp (i.e. 9.28 Mb)

[b] Difference between mean values for wild-type *dnaE* and individual S759 variants

[c] p-values calculated using unpaired 2-sided t-test with Welch's correction for rN per genome relative to *dnaE*_wt

Figure legends

Figure 1. A model of the catalytic center of *E. coli* DNA pol IIIα in a complex with DNA and dNTP substrate. The model was generated using the ternary complex structure of *G. kaustophilus* PolC (PDB: 3F3C) (Evans et al., 2008). (See details described in Materials and Methods). The secondary structures of the palm and finger domains that surround the catalytic center are shown in pink and blue, which are covered by a semi-transparent molecular surface. The DNA template and primer strands are shown in golden with the last nucleotide in the primer strand and incoming dNTP shown as "sticks". The active site residues D401, D403 and D555 (in the palm domain) are shown as pink sticks with red oxygen atoms. The α-helix containing residues forming the steric gate (in the finger domain) is highlighted in dark blue, and the H760 and S759 residues analyzed in this manuscript are colored in magenta, while

other key residues are colored blue. Dark blue and red colors in all stick models represent nitrogen and oxygen atoms, respectively. H760 directly contacts the deoxyribose of the incoming dNTP and forms the steric gate while S759 snuggly fits in a shallow pocket of the palm domain. The position of the 2'-OH and its close proximity with H760 is marked by a collision sign ().

Figure 2. Quantitative His$^+$ mutagenesis assays in RW1504 expressing S759 and H760 mutants. Strains were grown overnight at 30°C in appropriate antibiotics. Aliquots were harvested by centrifugation and resuspended in an equal volume of SM buffer. 100 μl of the overnight culture was spread on each low-histidine minimal plate and incubated at either 30°C, or 39°C, for four days, after which time, His$^+$ revertants were counted. Symbols represent average counts for individual biological replicates (n = 3 - 6). Error bars represent one standard deviation. Unpaired, two-tailed *t* tests were used to assess statistical significance between the mean colony counts for strains expressing wild-type *dnaE* or *dnaE* variant, at 30°C or 39°C. * = $p < 0.05$, ** = $p < 0.01$. We did not detect a statistically significant difference in colony count between cultures grown at 30°C or 39°C.

Figure 3. Spectra of spontaneous mutations in the *rpoB* locus in a *dnaQ920* proofreading-deficient background. (**A**) Wild-type *dnaE*, **(B)** *dnaE*_S759C, (**C**) *dnaE*_S759N, or (**D**) *dnaE*_S759T. The types of base-pair substitutions observed in the *rpoB* gene that result in rifampicin resistance are color coded as shown in the figure. The arrows indicate mutagenic hot spots. The numbers in brackets next to the name of the

dnaE allele refer to the number of mutants identified / number of mutants assayed. A more detailed spectral analysis can be found in Table S2.

Figure 4. **Increased ribonucleotide incorporation by *dnaE* steric gate mutants.**

(**a**) High molecular weight genomic DNA isolated from *rnhB*_wt and ∆*rnhB E. coli* with wild-type or *dnaE* variants separated by TAE agarose gel electrophoresis (**b**) RNase H2-treated genomic DNA separated by alkaline gel electrophoresis (representative of ≥6 independent experiments). (**c**) Densitometric intensity plots for the gel shown in panel B show greater fragmentation in the Δ*rnhB* strains, indicating higher numbers of genome-embedded ribonucleotides. (**d)** Densitometry plots were used to calculate the number of ribonucleotides per Δ*rnhB* genome relative to *rnhB*_wt strains, showing significantly increased levels in *dnaE*_S759T (1.8-fold), *dnaE*_S759C (2.3-fold) and *dnaE*_S759N strains (8.4-fold) compared to *dnaE*_wt. Individual data points indicate values from n = 6-8 independent experiments, with bars and error bars indicating mean ± SD. Unpaired 2-sided t-test with Welch's correction; **, p <0.01; ***, p < 0.001; ****, p < 0.0001.

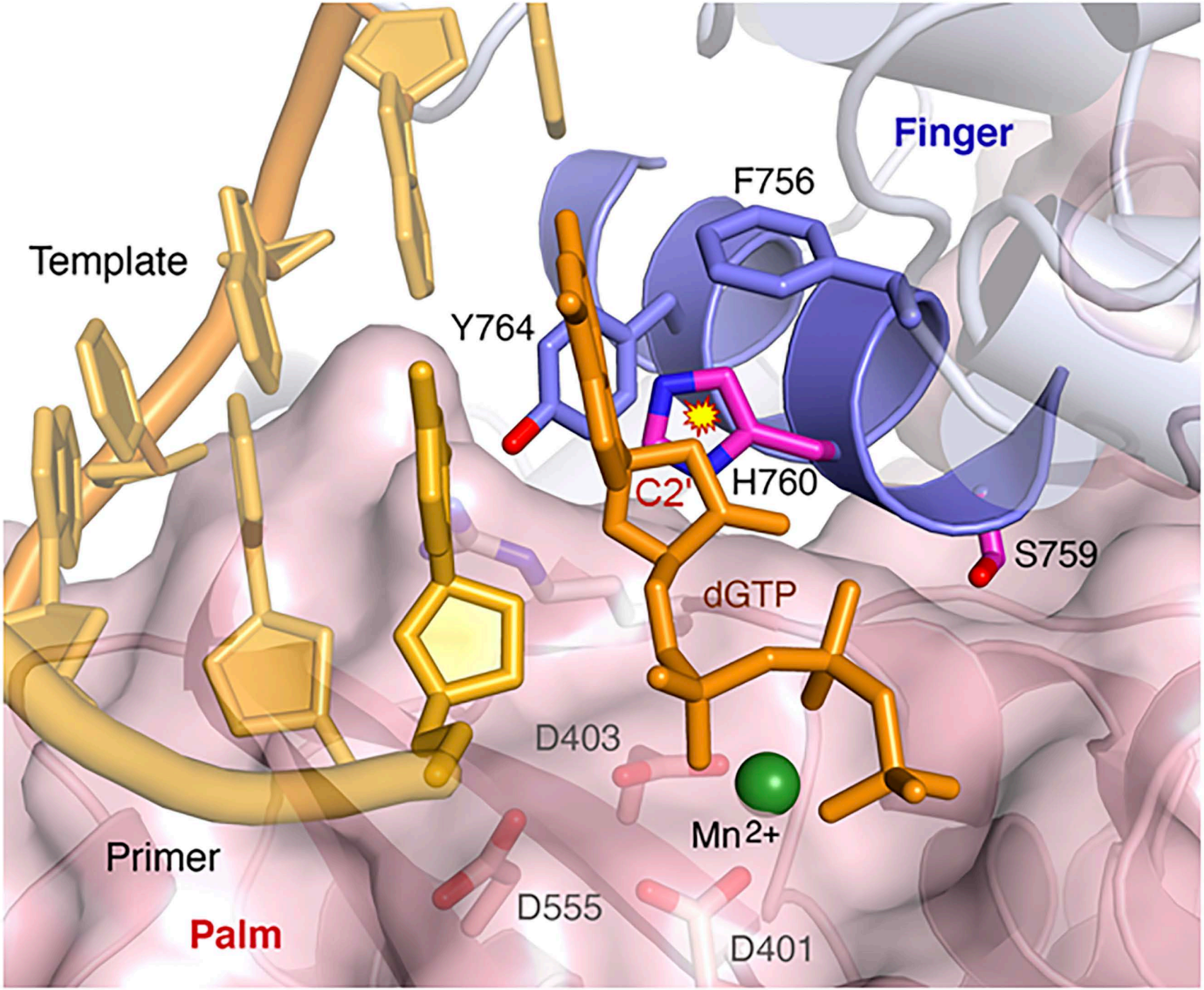

Finger
F756
Template
Y764
C2'
H760
S759
dGTP
D403
Mn2+
Primer
D555
Palm
D401

Figure 1

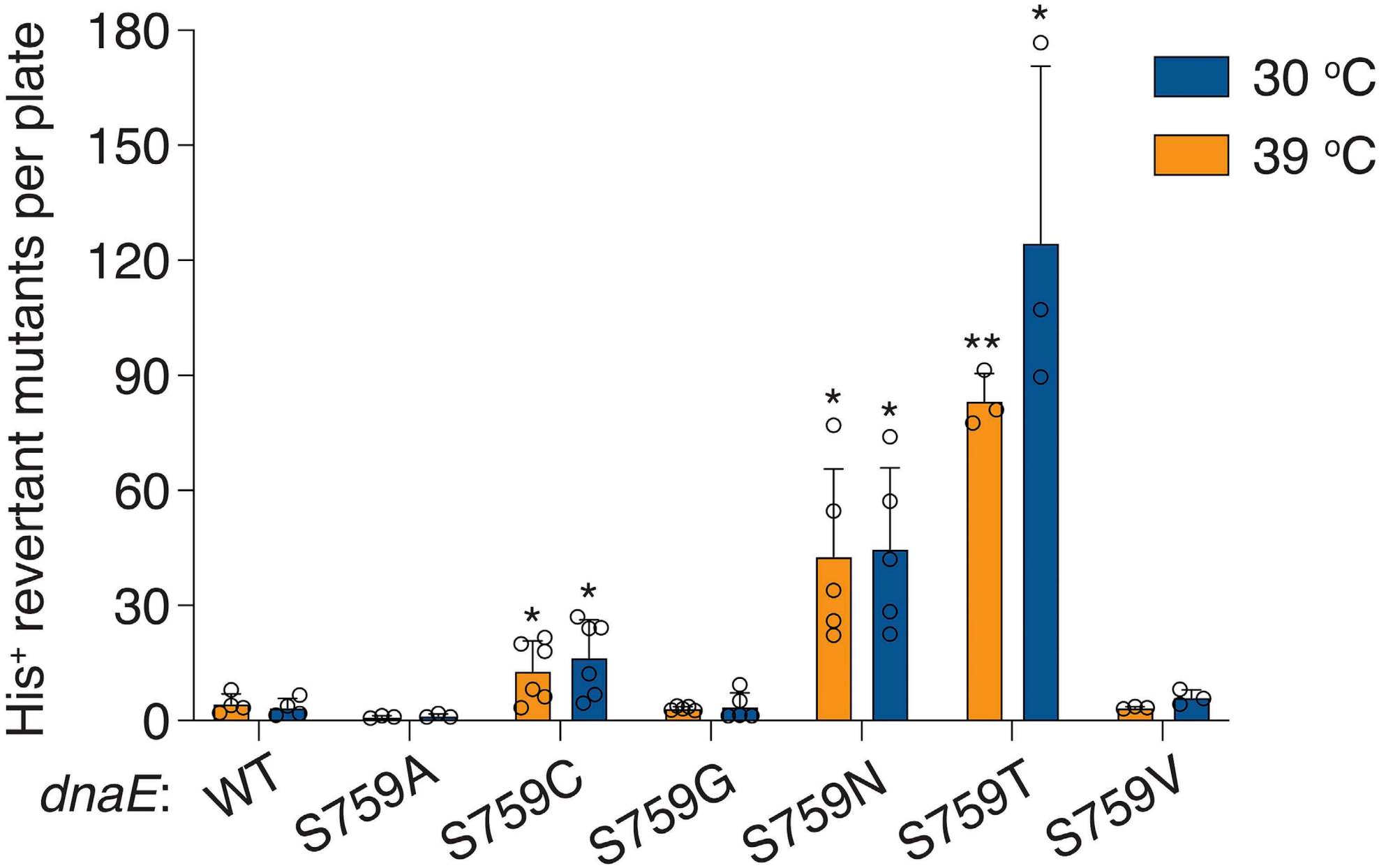

His+ revertant mutants per plate
180
150
120
90
60
30
0
30 °C
39 °C
*
*
*
*
*
**
dnaE:
WT
S759A
S759C
S759G
S759N
S759T
S759V

Fig.2

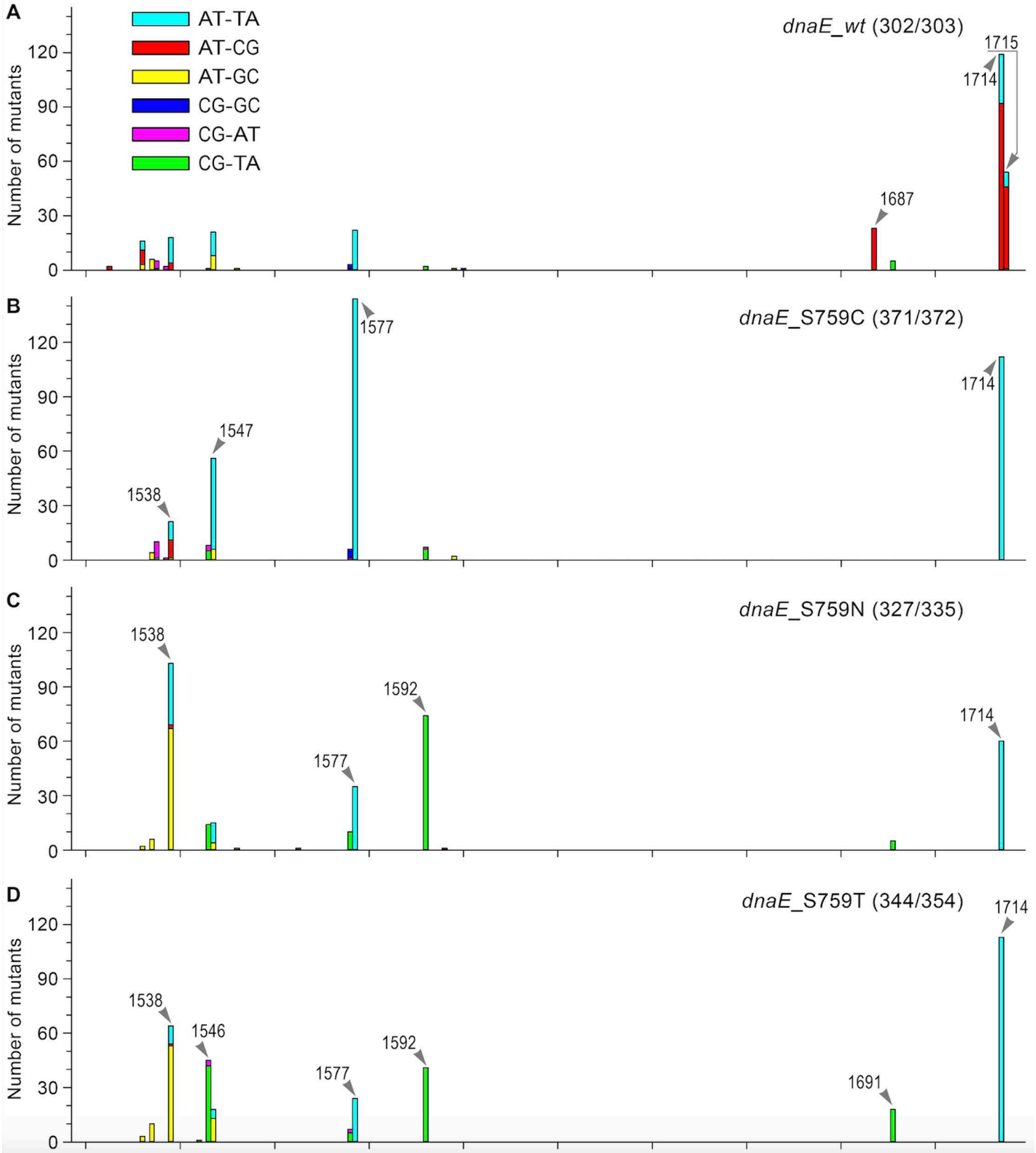

A
AT-TA
AT-CG
AT-GC
CG-GC
CG-AT
CG-TA
dnaE_wt (302/303)
Number of mutants
0
30
60
90
120
1714
1715
1687
B
dnaE_S759C (371/372)
1577
1714
1547
1538
C
dnaE_S759N (327/335)
1538
1592
1714
1577
D
dnaE_S759T (344/354)
1714
1538
1546
1592
1577
1691

Fig. 3

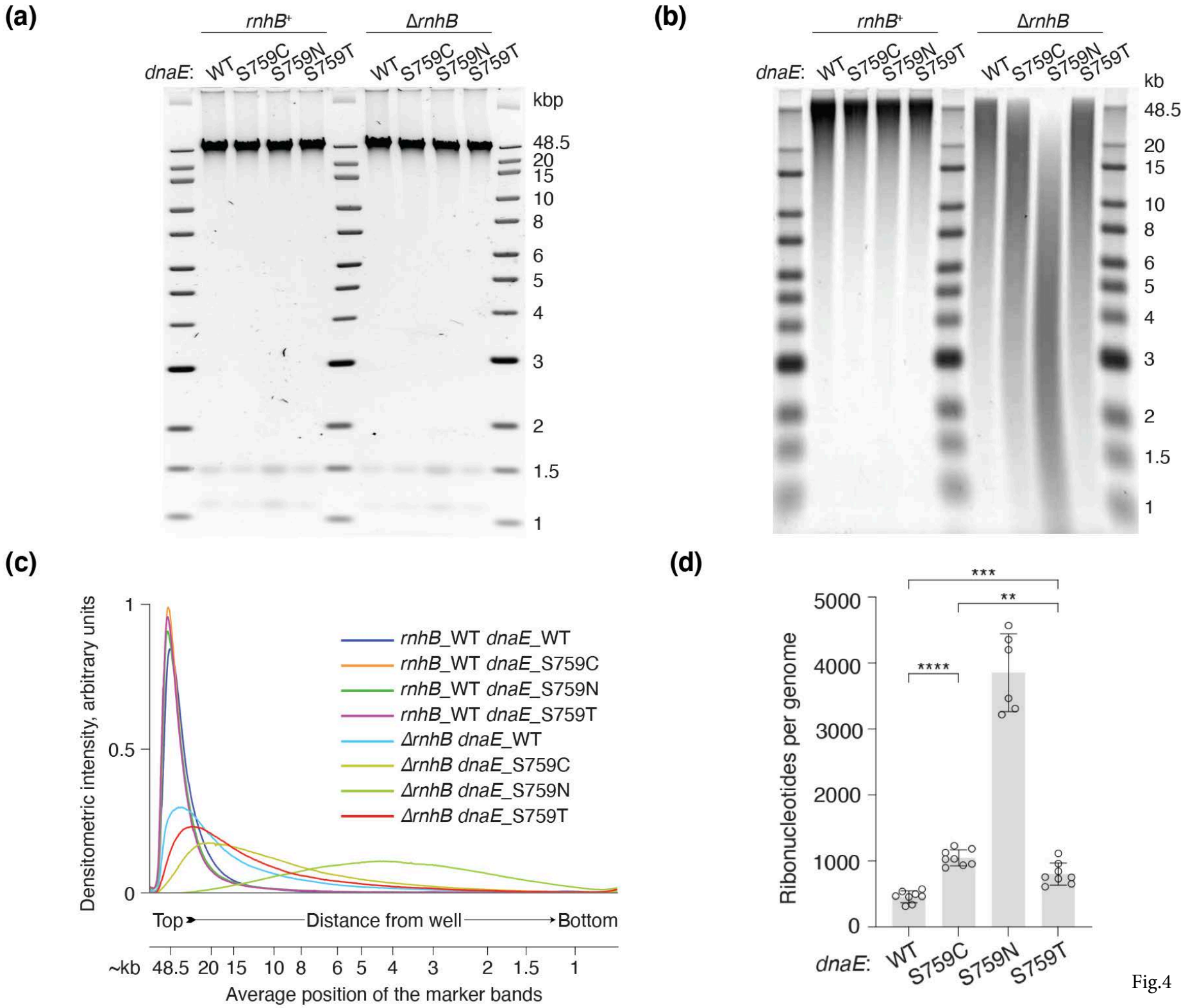
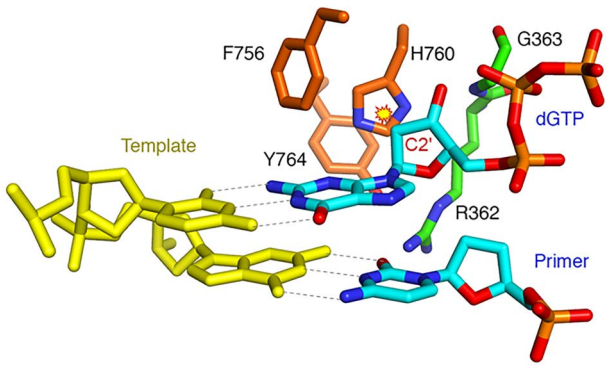

(a)
rnhB+
ΔrnhB
dnaE: WT S759C S759N S759T WT S759C S759N S759T
kbp
48.5
20
15
10
8
6
5
4
3
2
1.5
1
(b)
rnhB+
ΔrnhB
dnaE: WT S759C S759N S759T WT S759C S759N S759T
kb
48.5
20
15
10
8
6
5
4
3
2
1.5
1
(c)
Densitometric intensity, arbitrary units
1
0.5
0
rnhB_WT dnaE_WT
rnhB_WT dnaE_S759C
rnhB_WT dnaE_S759N
rnhB_WT dnaE_S759T
ΔrnhB dnaE_WT
ΔrnhB dnaE_S759C
ΔrnhB dnaE_S759N
ΔrnhB dnaE_S759T
Top
Distance from well
Bottom
~kb 48.5 20 15 10 8 6 5 4 3 2 1.5 1
Average position of the marker bands
(d)
Ribonucleotides per genome
5000
4000
3000
2000
1000
0
***
**
****
dnaE: WT S759C S759N S759T

Fig.4



Graphical Abstract

molecular microbiology

Supplementary Information

Novel *Escherichia coli* active site *dnaE* alleles with altered base and sugar selectivity

Alexandra Vaisman, Krystian Łazowski, Martin A. M. Reijns, Erin Walsh, John P. McDonald, Kristiniana C. Moreno, Dominic R. Quiros, Marlen Schmidt, Harald Kranz, Wei Yang, Karolina Makiela-Dzbenska and Roger Woodgate

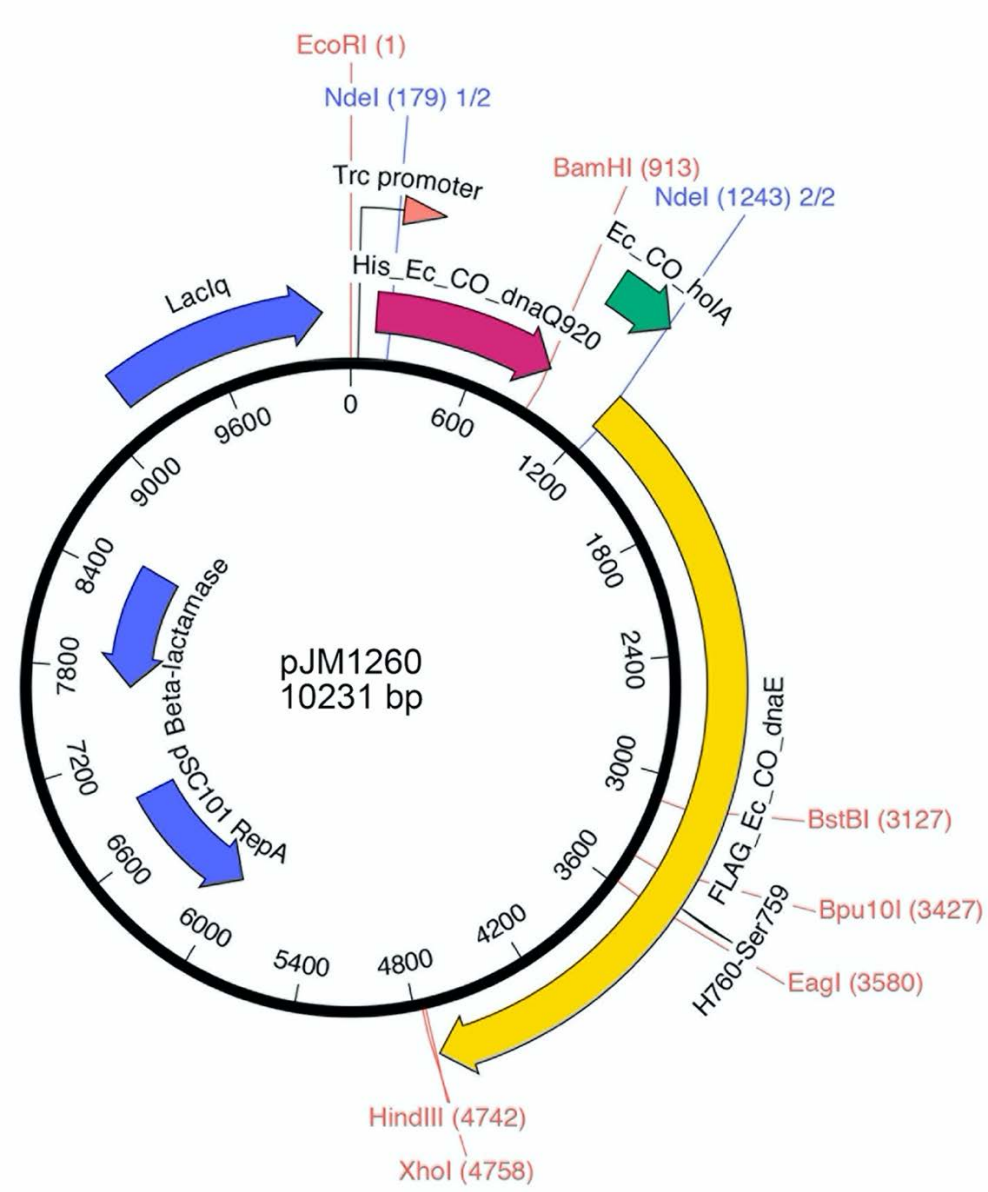


Figure S1. Cartoon of the pol III core vector, pJM1260. This 10,231 bp low-copy number plasmid expresses N-terminal HIS-tagged ε, untagged θ and N-terminal FLAG-tagged α. Steric gate substitutions can be generated by subcloning a chemically synthesized DNA fragment containing the desired substitution into the unique *Bst*BI and *Eag*I restriction enzyme sites.

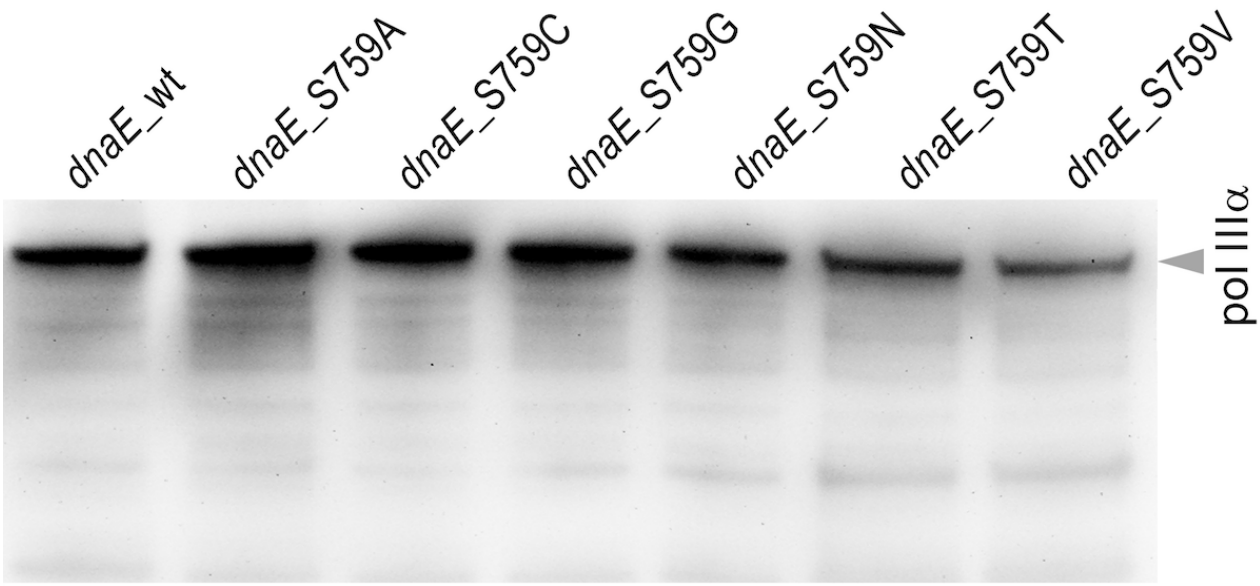


Figure S2. Western blot of the plasmid encoded α-subunit of pol III core expressed from pJM1260 and derivatives. The α-subunit of pol III core expressed from plasmid pJM1260 and variants were detected in whole cell extracts from *E. coli* RW1138 using polyclonal rabbit antibodies raised against pol III core. The position of full-length pol IIIα is indicated on the right-hand side of the image.

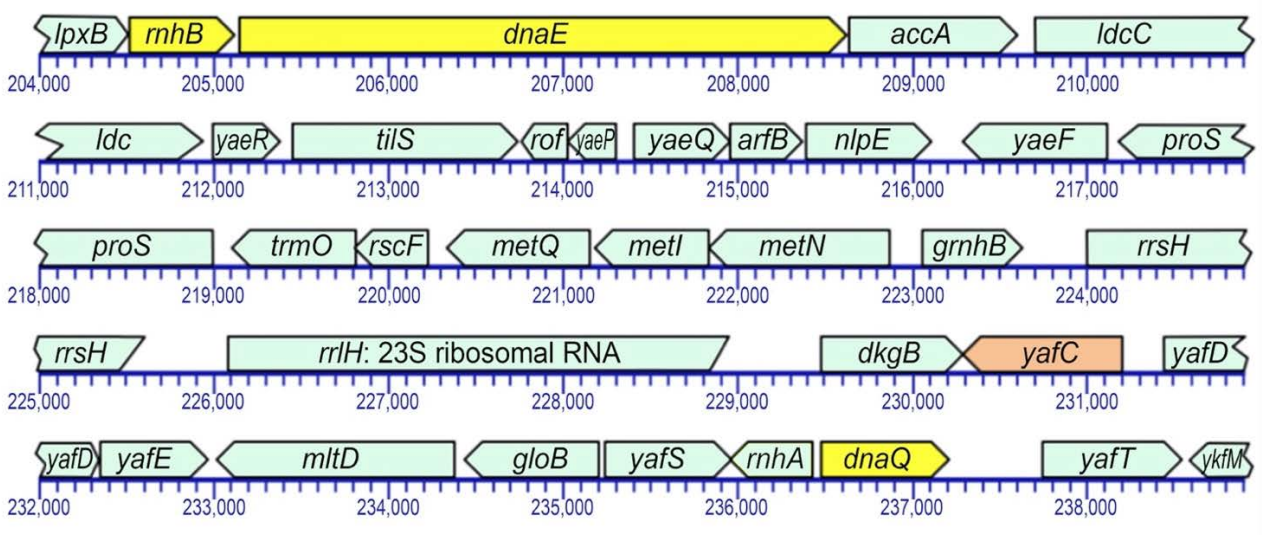


Figure S3. Schematic structure of the *lpxB-yklM* interval of the *E. coli* chromosome. The position of *yafC*, which was marked with kanamycin, or tetracycline resistance, for selection purposes, is shown in pale orange. The position of *dnaE, dnaQ, and rnhB* genes are highlighted in yellow. Co-transduction of *yafC* with *dnaE* was estimated to be ~50%; linkage to *dnaQ920* was ~44%; and to *rnhB* it was ~32%. All four genes can be co-transduced at one time with a frequency of ~10%.

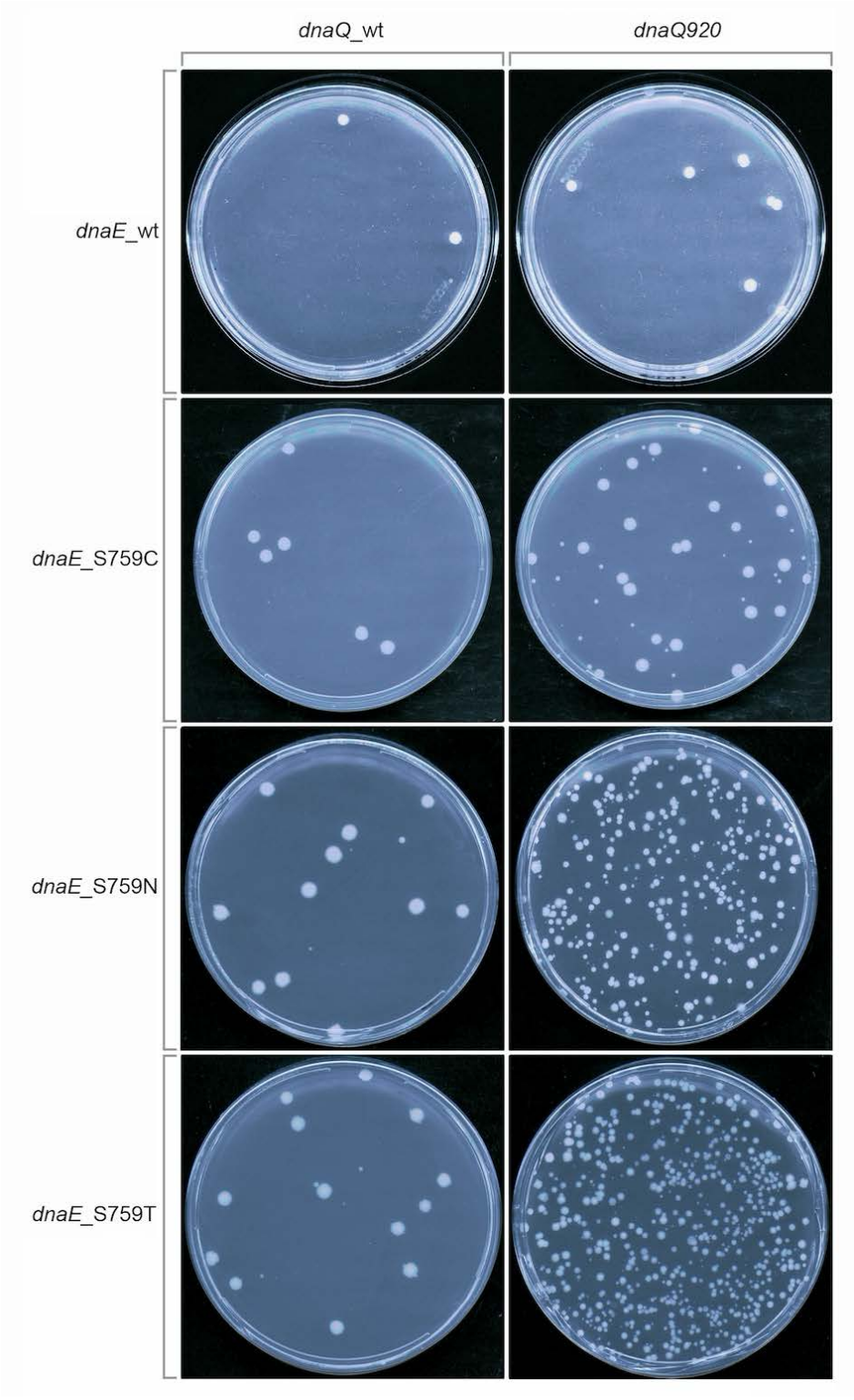


Figure S4. Qualitative plate assay to measure spontaneous mutator activity of wild-type *dnaE* and *dnaE*_S759 variants. The images shown are representative of

multiple repeat experiments. Overnight cultures grown in appropriate antibiotics were harvested by centrifugation, then resuspended in SM buffer and 100 μl of the SM solution. spread on minimal low histidine plates and incubated at 37 °C for 4 days. His^+ revertants grow up as white colonies on the background “mist” of the His^- parental strain. These experiments reveal that in a *dnaQ*$^+$ strain the three *dnaE* variants are mild-mutators compared to wild-type *dnaE.* Mutator activity increases significantly in *dna920* strains, with strong mutator effects observed with *dnaE*_S759N and *dnaE*_S759T.

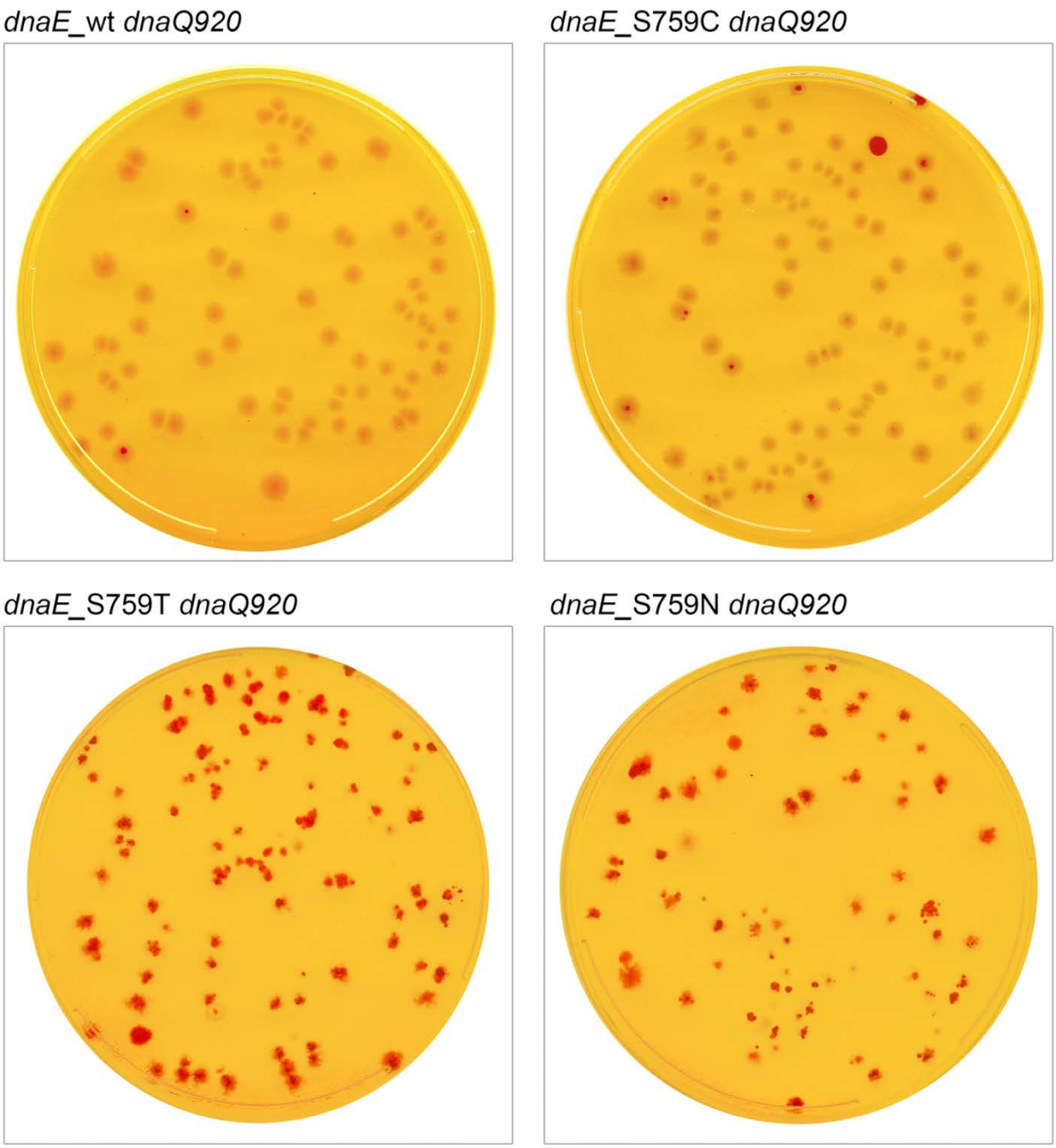


Figure S5. Qualitative papillation assay to measure spontaneous mutator activity of wild-type *dnaE* and S759 variants in *dnaQ920* strains. The images shown are representative of multiple repeat experiments. Overnight cultures grown in appropriate antibiotics were serially diluted up to 10^6-fold in SM buffer. 50 - 100 μl of the diluted

cultures were plated on MacConkey agar base containing 1% galactose and incubated at 37 °C for 8 days. Under these conditions, the Gal^- parental strain appears pink/orange against the background. Bacteria that revert to Gal^+ are able to metabolize the galactose in the medium and appear as bright red papillae against the pink/orange strain background. As expected, the wild-type *dnaE dnaQ920* strain (RW1614) exhibited almost no indication of Gal^+ papillation. The *dnaE*_S759C *dnaQ920* strain (RW1716) gave a handful of colonies with papillae. In dramatic contrast, virtually all colonies exhibited significant Gal^+ papillation with the *dnaE*_S759T or *dnaE*_S759N alleles in the *dnaQ920* background (strains RW1616 and RW1618 respectively).

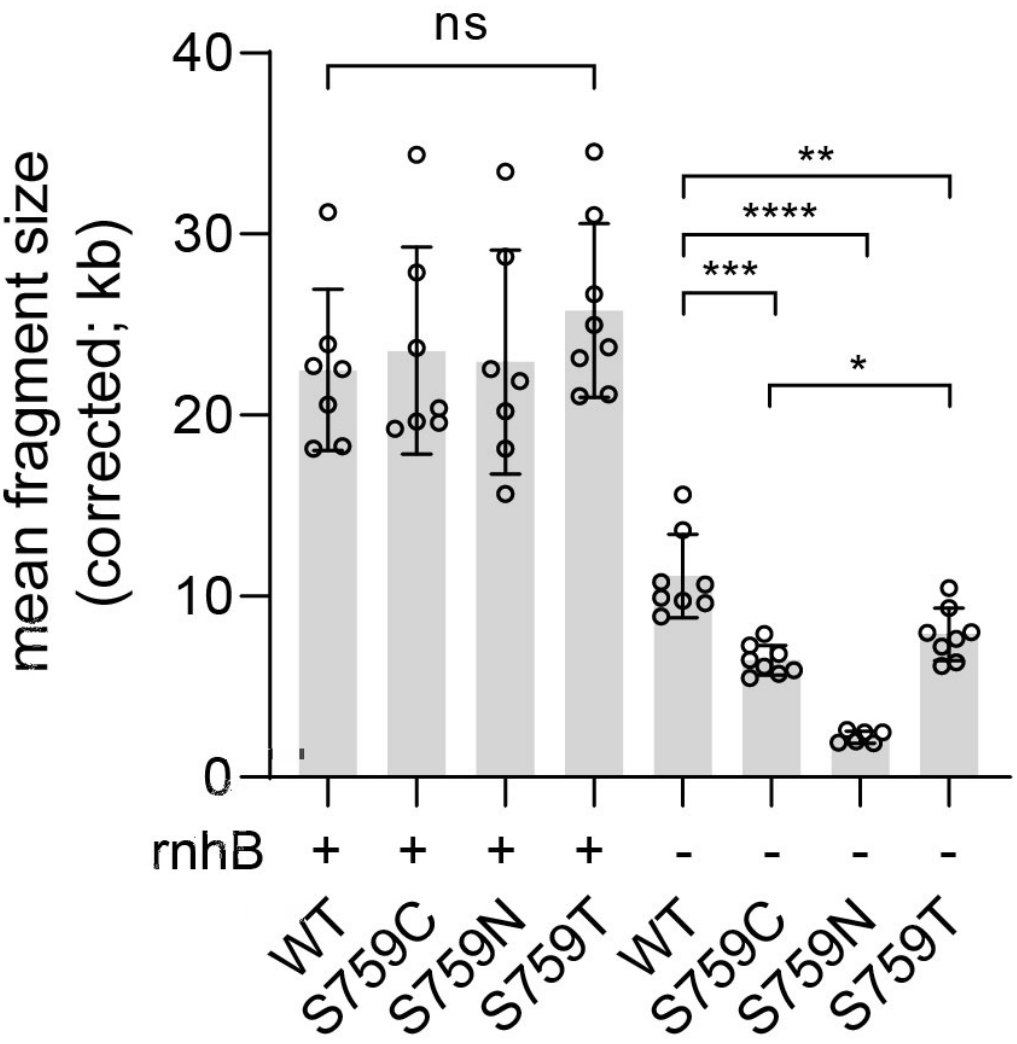


Figure S6. Increased fragmentation of RNase H2-treated genomic DNA from Δ*rnhB*, but not *rnhB*_WT strains expressing *dnaE* mutants. Densitometry plots after alkaline gel electrophoresis of RNase H2-treated genomic DNA were used to calculate the corrected mean fragment size for each sample per gel. This was then used to determine the number of genome-embedded ribonucleotides for the Δ*rnhB* strains relative to the *rnhB*$^{+}$ strains (see *Experimental procedures*). Individual data points indicate values from n = 6-8 independent experiments, with bars and error bars indicating mean ± SD. Unpaired 2-sided t-test with Welch's correction; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$; ns, not significant.

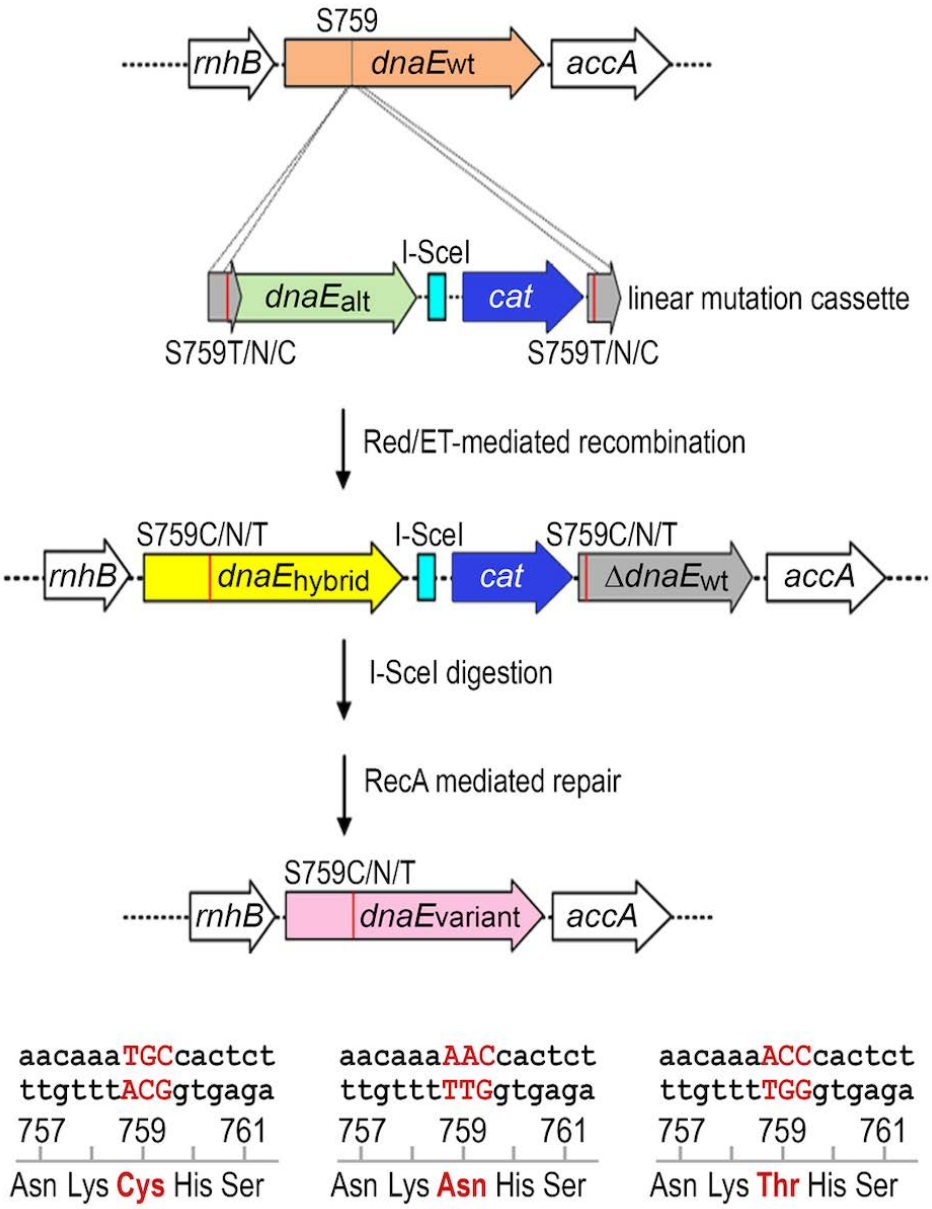


Figure S7. Genome modification strategy to generate the desired point mutations at position S759 in the essential *dnaE* gene. First, a linear mutation cassette was inserted into the chromosomal *dnaE* gene (wt) resulting in a hybrid *dnaE* gene (hybrid) encoding for a fully functional DNA polymerase III α-subunit. The selection marker was removed by a I-*Sce*I restriction and subsequent RecA-mediated repair forming the original endogenous *dnaE* allele including the desired point mutations.

S1 Table. PCR primers

Name	Sequence	Source
rnhB_F55	TTCCGTGAACTGCATCAGCA	Lofstrand
rnhB_R773	GCCATCGATCATCGAGTAGT	Lofstrand
EcdnaE486_F2378	GAC CGC CGA TAT GGA CAA	Lofstrand
EcdnaE486_R2911	GAC ATA ACG CTC AAT CTC TTT	Lofstrand
dnaE_F2059	CTACGGCATTATCCTGTATC	Lofstrand
dnaE_R2557	AAGTAGCCGCCTTTATTACG	Lofstrand
EcdnaQ_F26	ACTGCAATTACACGCCAGAT	Lofstrand
EcdnaQ_R328	CGCTTAAGCAACGAAAACTC	Lofstrand
rpoB1	CACA**C**GGCATCTGGTTGATACG	Lofstrand
rpoF1	TGGCGAAATGGCGGAAAAC	Lofstrand
PCR1-up-S759T	GAACTGGCGATGAAAATCTTCGACCTGGTGGAGAAATTCGC TGGTTACGGATTTAACAAAACCCACTCTGCGGCCTATGCGC TGGTGTCATACCAGACC	BioSpring
PCR2-down-S759T	GGATAGTGCGCTTTCAGCCATAACGTTTGATATGACACCAAA GCATAGGCCGCAGAGTGGGTTTTGTTAAATCCGTTTACGCC CCGCCCTGCCACTCATC	BioSpring
PCR1-up-S759N	GAACTGGCGATGAAAATCTTCGACCTGGTGGAGAAATTCGC TGGTTACGGATTTAACAAAAATCACTCTGCGGCCTATGCGCT GGTGTCATACCAGACC	BioSpring
PCR2-down-S759N	GGATAGTGCGCTTTCAGCCATAACGTTTGATATGACACCAAA GCATAGGCCGCAGAGTGATTTTTGTTAAATCCGTTTACGCCC CGCCCTGCCACTCATC	BioSpring
PCR1-up-S759C	GAACTGGCGATGAAAATCTTCGACCTGGTGGAGAAATTCGC TGGTTACGGATTTAACAAATGCCACTCTGCGGCCTATGCGC TGGTGTCATACCAGACC	BioSpring
PCR2-down-S759C	GGATAGTGCGCTTTCAGCCATAACGTTTGATATGACACCAAA GCATAGGCCGCAGAGTGGCATTTGTTAAATCCGTTTACGCC CCGCCCTGCCACTCATC	BioSpring
cp1	CTGTATCAGGAACAGGTCATG	BioSpring
cp2	GACGATCAAATGCACCTGAC	BioSpring
cp3	GGAAGTAGCCGCCTTTATTAC	BioSpring
cp4	GATCCCGCTGGATGATAAG	BioSpring
cp5	CCATCTAATACCACCTGCTC	BioSpring
cp6	CTGGAGTGAATACCACGAC	BioSpring

Table S2. **Spectrum of spontaneous mutations generated in the *rpoB* gene in MMR-proficient strains.**

	Position	bp change	*dnaE*$^+$	S759C	S759N	S759T
	1525	AT→CG	2			
	1532	AT→TA	5			
[	1532	AT→CG	8			
	1532	AT→GC	3		2	3
	1534	AT→GC	6	4	6	10
	1535	CG→AT	4	9		
[	1535	CG→TA	1	1		
	1537	CG→AT	2	1		
	1538	AT→TA	14	10	34	10
[	1538	AT→CG	4	10	2	1
	1538	AT→GC		1	67	53
	1544	AT→GC				1
	1546	CG→AT		3		3
[	1546	CG→TA	1	5	14	42
	1547	AT→TA	13	50	11	5
[	1547	AT→GC	8	6	4	13
	1552	AT→GC	1		1	
	1565	CG→TA			1	
	1576	CG→GC	2	5		
[	1576	CG→AT	1	1		2
	1576	CG→TA			10	5
	1577	AT→TA	22	144	35	24
	1592	CG→AT		1		
[	1592	CG→TA	2	6	74	41
	1596	AT→CG			1	
	1598	AT→GC	1	2		
	1600	CG→AT	1			
	1687	AT→CG	23			
	1691	CG→TA	5		5	18
	1714	AT→TA	27	112	60	113
[	1714	AT→CG	92			
	1715	AT→TA	8			
[	1715	AT→CG	45			
	1715	AT→GC	1			
	Total		302	371	327	344

[a] The data are the number of mutants found for each type of base substitution at a particular position

[b] The numbering system originates from Garibyan *et al*., (DNA Repair. 2003; **2**: 593–608), where the A of the ATG initiation codon is #1. Brackets indicate different types of mutations at the same nucleotide.