Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies

A. Castañeda-García¹*, I. Martín-Blecua³, E. Cebrián-Sastre¹, A. Chiner-Oms², M. Torres-Puente², I. Comas²,³, J. Blázquez¹*

The postreplicative mismatch repair (MMR) is an almost ubiquitous DNA repair essential for maintaining genome stability. It has been suggested that Mycobacteria have an alternative MMR in which NucS, an endonuclease with no structural homology to the canonical MMR proteins (MutS/MutL), is the key factor. Here, we analyze the spontaneous mutations accumulated in a neutral manner over thousands of generations by Mycobacterium smegmatis and its MMR-deficient derivative (ΔnucS). The base pair substitution rates per genome per generation are 0.004 and 0.165 for wild type and ΔnucS, respectively. By comparing the activity of different bacterial MMR pathways, we demonstrate that both MutS/L- and NucS-based systems display similar specificity and mutagenesis bias, revealing a functional evolutionary convergence. However, NucS is not able to repair indels in vivo. Our results provide an unparalleled view of how this mycobacterial system works in vivo to maintain genome stability and how it may affect Mycobacterium evolution.

INTRODUCTION

The maintenance of genome stability and low mutation rates is ensured in cells by different DNA surveillance and correction processes. These include base selection, proofreading, mismatch repair (MMR), base and nucleotide excision repair, recombinational repair, and nonhomologous end joining. The MMR system is a sophisticated DNA repair pathway that detects and removes replication-generated mismatched bases avoiding the acquisition of spontaneous mutations. It also recognizes mismatches in recombination intermediates, inhibiting recombination between nonhomologous DNA sequences (homeologous), and plays important roles in DNA damage responses (1, 2). The canonical MMR pathway includes members of the MutS and MutL protein families, which perform key steps in mismatch correction, and it is highly conserved among the three domains of life with some important exceptions, such as the phylum Actinobacteria and several groups of Archaea (3, 4). Loss of this pathway leads to high rates of mutation (hypermutability) and increased recombination between nonhomologous DNA sequences, disrupting the genetic interspecies barrier (5).

Mycobacteria are a biologically diverse group of bacteria, which includes relevant human pathogens, such as M. tuberculosis, M. leprae, and M. avium, and were defined as deprived of MMR (6). However, we have recently shown that they have a putative noncanonical MMR pathway, whose main factor is the DNA repair protein NucS/EndoMS. NucS has no structural homology with the key canonical MMR factors MutS and MutL (4, 7). According to in silico analysis, the mycobacterial NucS is a putative mismatch-specific endonuclease, required for mutation avoidance and anti-recombination in vivo, both hallmarks of the canonical MMR (4). Archaeal NucS is able to recognize and bind mismatched pair bases in the DNA in vitro, activating its endonuclease activity to promote the mismatch removal (8, 9). In addition, this protein corrects replication errors, acting cooperatively with the sliding clamp of the replisome in Corynebacterium glutamicum (8, 10). Computational studies support an archaean origin for NucS, built upon two distinct domains that suffered a complex evolutionary history of transfers and/or losses (4). Therefore, distinct pathways for MMR seem to have evolved at least twice in nature (4).

Inactivation of NucS in different actinobacterial species leads to the emergence of hypermutator strains with increased spontaneous mutation rates (4, 8, 10). Some bacterial isolates with impaired MMR capacity have been detected in nature, suggesting that modulation of MMR activity can be adaptive (11–13). Similarly, previous studies suggested the existence of naturally occurring M. tuberculosis mutator strains defective in MMR activity, which could be particularly improved for adaptation to antibiotic regimes (4).

However, despite its major importance, the in vivo specificity and mutagenesis bias of this putative alternative MMR remained unknown. Because mutations are the source of variation upon which natural selection acts, analysis of mutations that accumulate in a neutral fashion is essential to understand how bacterial pathogens, such as those belonging to the Mycobacterium genus, evolve virulence and antibiotic resistance. Therefore, in this work, we have performed mutation accumulation (MA) experiments to understand how this putative noncanonical MMR system of mycobacteria works in vivo using both a wild-type strain of M. smegmatis and its noncanonical MMR-deficient isogenic derivative. The results provide a picture of the in vivo noncanonical MMR genome maintenance activity in mycobacteria and the dynamic of mutation across the whole genome over time. A set of reporter plasmids containing specific mutations was also used to reveal the activity of NucS on the correction of each type of mutation detected in the MA evolution. Together, these results allow us to establish definitively the biological role of the mismatch-specific nuclease NucS/EndoMS as guardian of the genome stability and its functional similarity to the canonical MMR.

¹Centro Nacional de Biotecnología–CSIC, Madrid, Spain. ²Instituto de Bio­medicina de Valencia, IBV-CSIC, Valencia, Spain. ³CIBER in Epidemiology and Public Health.
*Corresponding author. Email: acastaneda@cnb.csic.es (A.C.-G.); blazquez@cnb.csic.es (J.B.)
RESULTS

MA experiments
In a recent study, we showed that a *M. smegmatis* nucS-null strain, impaired in the noncanonical MMR pathway, displayed a hypermutable phenotype, with increased rate of spontaneously emerging rifampicin and streptomycin resistances above the wild-type strain (150- and 86-fold, respectively) (4). These results were equivalent to those observed for an MMR-deficient *Escherichia coli* mutant (14, 15).

It is now essential to understand and decipher how mutations emerge across the whole genome in a mycobacterial strain lacking this noncanonical MMR and whether this pathway parallels the activity of the canonical MMR.

Analysis of the mutational profile of *M. smegmatis* mc² 155 wild type and its ΔnucS derivative was carried out by MA experiments and whole-genome sequencing (see Materials and Methods). MA experiments were carried out for 50 weeks with 11 independent lineages for each parental strain, which accounted for 1372 and 1333 generations per line for wild type and nucS each parental strain, which accounted for 1372 and 1333 generations respectively (or a total of 15,095 and 14,662 generations, respectively) (Table 1). Every 7 days, a single colony from each line was restreaked onto a fresh plate, minimizing the effective population size. This type of single-colony MA experiments with a very stringent bottleneck has been demonstrated to minimize selection forces, allowing mutations to accumulate in a neutral unbiased manner (15, 16).

Mutation rates
Sequence analysis of the evolved strains revealed the total number of mutations identified in the whole genome in each strain (Table 1 and Fig. 1A), including base pair substitutions (BPSs) plus insertions/deletions (indels) (Table 2 and Fig. 1, B and C). A total number of 80 mutations were detected (the median value number of mutations per genome is 7) in the wild-type *M. smegmatis* MA lines. The ΔnucS lines accumulated a total of 2444 mutations (the median value of mutations per genome is 215). The average of sequence coverage per line corresponds to >99.9% of the *M. smegmatis* reference genome (GenBank NC_018289.1).

Considering the number of generations along the evolution experiment, global mutation rates can be measured in the whole genome for each division event. Our results indicated that the wild-type mc² strain maintained a low mutation rate, 0.005 mutations per genome per generation (7.58 × 10⁻¹⁰ mutations per nucleotide per generation), while the ΔnucS strain reached 0.167 mutations per genome per generation (238.53 × 10⁻¹⁰ mutations per nucleotide per generation). Therefore, the absence of NucS/EndoMS led to a 31-fold increase in the total mutation rate across the whole genome (Table 1).

To evaluate the effect of selective pressure in the mutational profile generated in our MA experiments, we studied the proportion of mutations in coding versus noncoding DNA, assuming that the intergenic regions are exposed to lesser selective pressure. When mutations accumulate in a neutral manner, the proportion of mutations in coding versus noncoding DNA fits the proportion between the sizes of coding (genes) versus noncoding sequences (intergenic regions). Coding regions cover 90% of the *M. smegmatis* genome (16). Our MA experiments revealed that 86.9% (wild type) and 92.6% (ΔnucS) of all BPSs are located in the coding regions. Both values are not significantly different from the expected 90% (X² = 16.67, P = 0.95 for the wild type and X² = 0.97, P = 0.95 for ΔnucS) (critical X² = 18.307, DF = 10) (table S1). Moreover, in the absence of selection, the ratio of nonsynonymous to synonymous mutations should not be significantly different from the random expectation, considering that synonymous mutations are relatively neutral. Our MA data show ratios of 2.78 (39/14) and 2.97 (1681/565) for wild-type and ΔnucS lines.

Table 1. Total numbers and rates of mutation identified in MA experiments.

| Strain       | Number of lines | Generations per line | Total generations | Total mutations | Mutation rate per genome* per generation (×10⁻⁶) † | Mutation rate per nucleotide per generation (×10⁻¹⁰) † |
|--------------|-----------------|----------------------|-------------------|---------------|-------------------------------------------------|---------------------------------------------------|
| mc² 155 (wild type) | 11              | 1,372                | 15,095            | 80            | 5.30 ± 1.70                                       | 7.58 ± 2.43                                       |
| ΔnucS        | 11              | 1,333                | 14,662            | 2,444         | 166.69 ± 20.92                                   | 238.53 ± 29.94                                   |

*Reference genome = 6,988,209 nucleotides. †The mean of the mutation rate per line (per genome or per nucleotide) ± 95% confidence limits is shown.

Fig. 1. Whole-genome mutations of the evolved strains. Total number of mutations (A), BPSs (B), and indels (C) of the evolved wild-type *M. smegmatis* mc² 155 (blue circles) and ΔnucS (red circles) strains after performing the MA experiment. The horizontal straight segment represents the median value of mutations in each case.
noncoding ratios were statistically different between wild-type and nucSΔ lines (see above), the coding/
results for both wild-type and MA lines, a total of 18 (Fig. 1C and Table 2), was almost identical to the
number of short insertions and deletions of 1 to 66 bp (most of them ±1 bp), a total of 18 (Fig. 1C and Table 2). The rates of total mutations, BPSs, and in­
dels in our MA experiments.

BPSs and indels
Across the sequenced wild-type M. smegmatis MA lines, a total of 61 BPSs were identified (Fig. 1B and Table 2). The overall BPS rate per nucleotide per generation was 5.78 × 10⁻¹⁰, equivalent to 0.004 per generation. The MA analysis also revealed 19 short insertions and deletions of 1 to 30 base pairs (bp) in length (most of them ±1 bp) (Fig. 1C and Table 2). The indel rate per site per generation was 1.80 × 10⁻¹⁰, and the insertion/deletion ratio was 2.17 (13 insertions and 6 deletions) (Table 2). The rates of total mutations, BPSs, and indels in our MA evolution experiments are similar to the ones reported for M. smegmatis wild-type strain in previous studies (16, 17).

ΔnucS lines accumulated a total of 2426 BPSs (Fig. 1B and Table 2), with an overall mutation rate of 2.36 × 10⁻⁸ per nucleotide per generation or 0.17 per genome per generation. Therefore, loss of nucS led to a ~41-fold increase in BPS rate with respect to the wild-type lines. The number of short insertions and deletions of 1 to 66 bp (most of them ±1 bp), a total of 18 (Fig. 1C and Table 2), was almost identical to the wild-type lines, with an indel rate for the ΔnucS strain of 1.76 × 10⁻¹⁰ per site per generation and an insertion/deletion ratio of 3.5 (14 insertions versus 4 deletions) (Table 2).

The functional context of each BPS (table S1) was identified using the annotated M. smegmatis mc² 155 genome [National Center for Biotechnology Information (NCBI) accession: NC_018289.1]. Al­
though values of BPSs in coding regions are close to the expected results for both wild-type and ΔnucS lines (see above), the coding/ noncoding ratios were statistically different between wild-type and ΔnucS lines [6.62 (53/8) and 12.5 (2246/180), respectively; Fisher’s

### Table 2. Rates of mutation specified by BPSs and indels.

|                | mc² 155 (wild type) |          | ΔnucS            |          |
|----------------|---------------------|----------|------------------|----------|
|                | Number   | Percentage (%) | Mutation rate per genome* per generation (×10⁻³) | Mutation rate per nucleotide per generation (×10⁻³) | Number   | Percentage (%) | Mutation rate per genome* per generation (×10⁻³) | Mutation rate per nucleotide per generation (×10⁻³) |
| **BPSs**       |          |              |                    |          |          |              |                    |          |
| Transitions    | 38       | 62.3       | 4.04 ± 1.44        | 5.78 ± 2.06 | 2426     | 100        | 165.46 ± 20.64   | 236.77 ± 29.53 |
| A>T>G:C       | 10       | 16.4       | 0.66 ± 0.41        | 0.95 ± 0.58 | 1429     | 58.9       | 97.46 ± 10.10    | 139.46 ± 14.46 |
| G>C>A:T       | 28       | 45.9       | 1.85 ± 0.89        | 2.65 ± 1.27 | 959      | 39.5       | 65.40 ± 12.34    | 93.59 ± 17.66  |
| Transversions  | 23       | 37.7       | 1.52 ± 0.60        | 2.18 ± 0.86 | 38       | 1.6        | 2.59 ± 1.20      | 3.71 ± 1.72   |
| A>T>A:T       | 2        | 3.3        | 0.13 ± 0.20        | 0.19 ± 0.28 | 7        | 0.3        | 0.48 ± 0.41      | 0.68 ± 0.58   |
| A>T>C:G       | 6        | 9.8        | 0.40 ± 0.26        | 0.57 ± 0.37 | 6        | 0.2        | 0.41 ± 0.26      | 0.59 ± 0.38   |
| G>C>T>A       | 12       | 19.7       | 0.79 ± 0.51        | 1.14 ± 0.73 | 18       | 0.7        | 1.23 ± 0.72      | 1.76 ± 1.03   |
| G>C>G:C       | 3        | 4.9        | 0.20 ± 0.32        | 0.28 ± 0.45 | 7        | 0.3        | 0.48 ± 0.34      | 0.68 ± 0.49   |
| **Indels**     | 19       | 100        | 1.26 ± 0.54        | 1.80 ± 0.77 | 18       | 100        | 1.23 ± 0.65      | 1.76 ± 0.93   |
| Insertions     | 13       | 68.4       | 0.86 ± 0.48        | 1.23 ± 0.69 | 14       | 77.8       | 0.95 ± 0.68      | 1.37 ± 0.97   |
| Deletions      | 6        | 31.6       | 0.40 ± 0.34        | 0.57 ± 0.48 | 4        | 22.2       | 0.27 ± 0.34      | 0.39 ± 0.49   |

*Reference genome = 6,988,209 nucleotides. †Mutation rates per genome or nucleotide per generation ± 95% confidence limits.

respectively (table S1). These values are very close to the theoretical value previously defined to be 2.60 for M. smegmatis (16), indicating no bias against nonsynonymous mutations. Therefore, we conclude that selection does not affect significantly the distribution of mutations in our MA experiments.

Mutational spectrum and bias
The mutational spectrum of M. smegmatis wild type revealed by MA experiments indicates a high proportion of transitions, comprising 62.3% (38 transitions) of all BPSs (Table 2). The most common type of substitution is the G:C>A:T mutation, representing 45.9% of total changes (28 of 61 total) and 73.7% of all transitions (28 of 38) (Table 2). However, the proportion of transversions reached more than one­third of all the identified BPSs, with 37.7% (23 transversions). Both the spectrum and the transition/transversion ratio of 1.65 are very similar to the previously described for wild-type M. smegmatis (16).

The absence of the noncanonical MMR leads to a substantial change in the spectrum of BPSs. The effect of loss of the noncanonical MMR was to considerably increase the number of both types of transitions. ΔnucS lines accumulated 2388 BPSs with 98.4% and 38 transversions (1.6%), with a transition/transversion ratio of 62.8, 38-fold higher than that of wild-type lines. Moreover, if we focus only on transitions, the preferred NucS target, the mutation rate per genome per generation increased ~65-fold in the ΔnucS strain (162.86 versus 2.52; Tables 2 and 3). This indicated that this repair system acts preferentially recognizing and repairing transition errors. In the ΔnucS strain, A:T>G:C and G:C>A:T events rose markedly when compared to those detected in the wild type (1429 versus 10 A:T>G:C and 959 versus 28 G:C>A:T, respectively). However, only minor or no changes can be seen in the four types of transversions when wild-type and ΔnucS strains were compared (Tables 2 and 3).

A shift in the BPS bias occurred in the absence of the noncanonical MMR. While the wild-type strain presented a mutational bias for G:C>A:T, it changed to A:T>G:C in the ΔnucS strain. In the NucS-deficient strain, A:T>G:C BPSs represented 58.9% of all changes (1429 of...
2426) and 59.8% of the observed transitions (1429 of 2388). This suggests that one of the activities of the noncanonical MMR seems to control the increase in genomic G:C content (see Discussion).

**DNA strand bias**

*M. smegmatis* genome is divided into two replicores (right and left) that replicate in opposite directions, starting at the origin of replication and ending at the terminus in the Ter region, which is located at *dif* sites (18). The leading and lagging strands are switched in the two replicores relative to the conventional 5'-3' numbering system. Assuming that most mutations are the result of replication errors, the A:T->G:C base substitutions should result from C mispaired with a template A or G mispaired with a template T.

Table S2 shows that the most prominent BPS in Δ*nucS*, A:T->G:C transitions, occurred almost twice as likely with A rather than with T (444 As versus 235 Ts) templating the lagging strand (right replicore). However, A:T->G:C transitions occurred almost twice as likely with T rather than with A (494 Ts versus 256 As) templating the leading strand (left replicore). Regarding G:C->A:T transitions, they were almost twice as likely to occur with C rather than with G (289 Cs versus 171 Ts) templating the lagging strand; the opposite happened in the leading strand (309 Gs versus 190 Cs). Therefore, NucS seems to balance the asymmetry of mutations detected in each DNA strand as seen in canonical MMR (15).

**Analysis of the mutation rate of BPSs and indels with integrative reporter plasmids**

Sequence analysis of the evolved strains in MA experiments indicated changes in the profile of mutations between the wild-type and Δ*nucS* strains. The ability of NucS to recognize and repair specific errors in vivo can be studied in detail through the design of genetic reporters containing each type of point mutation. These reporters allow us to measure the mutation rate of BPSs and indels selectively to establish the repair specificity of NucS inside the cells.

A set of integrative reporter plasmids with a modified *aph* gene marker (kanamycin phosphotransferase gene) was generated by site-directed mutagenesis to measure the rate of each specific mutation. While the intact *aph* marker conferred kanamycin resistance (Kan^R_), the mutated *aph* versions were all kanamycin susceptible (Kan^S_) (see Materials and Methods and the Supplementary Materials). Six reporter plasmids (named pMV-aph numbers 1 to 6) containing all six types of BPSs were constructed by introducing specific point mutations at the CAT codon (encoding the key residue histidine His^152_) in pMV361. Two plasmids, numbers 1 and 2, contained two different transitions, while the other four (numbers 3 to 6) have all four possible transversions. A scheme of the mutated plasmids with inactive *aph* and the resulting sequences with the amino acid changes are shown in Fig. 2.

By kanamycin resistance reversion assays, we measured the rates of each base substitution mutation in the wild-type and *nucS*-defective strains (see Materials and Methods). All base substitutions occurred at a very low rate in the wild-type strain, most of them around 10^{-9} mutations per cell per generation, although, as expected, both transitions (A:T->G:C and G:C->A:T) were the most frequent mutations, while some others were especially rare or undetectable (e.g., A:T->T:A transversion) (Fig. 3 and table S3).

Elimination of the nuclease NucS led to a significant increase in both types of transitions (16.94-fold and 12.28-fold for A:T->G:C and G:C->A:T, respectively) when compared to the wild-type mutation rates (Fig. 3 and table S3). A much lesser effect of NucS was detected in the mutation rates of transversions. All transversions showed no increase or a slight increase in mutation rates (less than two to three times) in the Δ*nucS* strain, keeping similar rates to the ones detected in a wild-type strain (Fig. 3). These results indicate that NucS is able to target and repair specifically the transitions that emerge in the genome, displaying a much lesser role on the repair of transversions.

In addition to the BPS reporters, a set of frameshift mutation reporters at the *aph* gene was designed to analyze the activity of NucS on different insertions/deletions. Three plasmid reporters were created with +1-, +2-, and +3-bp frameshifts in the CAT codon.

---

**Fig. 2. Assembly of reporter plasmids.** Integrative plasmid constructs used to measure Kan^R_ reversion events are shown. Modified bases are shown in green, and indels are shown in orange.

**Fig. 3. Mutation rates for Kan^R_ reversion events.** Rates of spontaneous mutations per cell per generation conferring kanamycin resistance for *M. smegmatis mc^2_155* (wild type) (blue) and its Δ*nucS* derivative (red) are shown. Ninety-five percent confidence intervals are shown. The limit of detection was around 10^{-10}, as inferred by observed mutation rates of A:T->T:A transversion.
of the proper base pair could reconstruct a functional gene in these frameshift in the aph gene (position 624–627). Two reporter plasmids (pMV-aph numbers 7 and 8) were made to analyze indels through the insertion or deletion of a G base, respectively, in this stretch of 4 Gs to generate a frameshift in the aph gene (Fig. 2). Likewise, only deletion or insertion of the proper base pair could reconstruct a functional gene in these frameshift-carrying reporters. When integrated in the M. smegmatis chromosome, each indel construction allowed us to measure the rate of a single base pair deletion or insertion (see Materials and Methods). Our data showed that both strains, wild type and ΔnucS, exhibited similar rates of 1-bp insertion/deletion (Fig. 3 and table S3). These results indicated that the noncanonical MMR pathway is unable to repair single base pair deletion or insertion in M. smegmatis.

**Comparison of NucS versus MutSL repair activities**

We have performed a comparative analysis of the in vivo repair capacities of eight different bacterial MMRs (two harboring NucS and six with MutSL), including M. smegmatis, as calculated from MA experiments (Table 3 and table S4). The higher activity on transitions than transversions is a general feature shared by canonical and noncanonical MMR systems, being especially relevant in M. smegmatis. M. smegmatis NucS has very poor activity on all types of transversions, particularly on A:T→G:C (only 2.87% are repaired) and G:C→T:A (35.25% are repaired), leading to a weak 1.7-fold increase when NucS was eliminated (Table 3 and table S4). C. glutamicum, another NucS-harboring bacterium, also exhibited very poor activity on all types of transversions (8, 21). In contrast, most MutSL-harboring species retained some activity, in most cases higher than 70% on transversions, with increases of 7 to 12 times when MutSL activity was eliminated. The case of Deinococcus radiodurans seems to be special as its ΔmutS-deficient derivative showed extremely low increase in mutation rate (only fourfold), suggesting that the low MMR activity may be compensated by other highly efficient alternative DNA repair pathways (21). In summary, M. smegmatis MMR function is similar to that of other bacterial MutSL-based MMR on transitions, although it displays clear unique features such as its inability to correct small indels and its exceptionally low activity on transversions.

**DISCUSSION**

Mutation is a major evolutionary force able to generate genetic variation, and it is the substrate for the complex process of evolution. However, most organisms keep very low mutation rates, probably as a result of selective pressure against the cost of deleterious mutations, reaching a finely tuned balance (22–24). Our MA experimental data estimate a spontaneous mutation rate of 0.005 mutations per genome per generation for the wild-type M. smegmatis strain (7.58 × 10⁻¹⁰ per nucleotide per generation), similar to previous reports in mycobacteria (16) and other free-living bacteria (15, 21, 25–27). The mutation rate detected here is close to the value of ~0.003 mutations per genome per generation proposed for all microbes by Drake (24). Our results thus indicate that a fine-tuned mutation rate could be an evolved trait also in mycobacteria, like in other prokaryotes (24).

Several DNA repair mechanisms remove mutations to reduce the mutation rate, especially highlighting the essential role of the MutS-MutL-based MMR pathway. Mycobacteria, including M. smegmatis, lack canonical MMR enzymes (MutS and MutL) (4, 6). Nevertheless, we and others have recently suggested that they have a noncanonical MMR pathway, whose key factor is the DNA repair protein NucS, with a different evolutionary origin and no structural homology to the canonical MMR enzymes MutS and MutL (4). In our MA experiment, the nucS-deficient strain accumulated a large number of mutations, displaying a mutation rate of ~0.16 per genome per generation or 2.38 × 10⁻⁸ per site per generation (>30-fold higher than that of the wild-type strain).

NucS mainly targets BPSs, but according to our results, it is unable to recognize and repair indels. Along the MA evolution, BPS rate reached 0.16 mutations per genome per generation in the ΔnucS strain, more than 40-fold higher than the one detected in the wild-type strain. The rate of indels was similar in both strains. Moreover, indel reporter plasmids confirmed that +1-bp insertion or −1-bp deletion occurs at very low rates with similar levels in both strains, reflecting the lack of repair activity of NucS on indels. This seems to be a unique characteristic of the mycobacterial MMR system, as most bacterial canonical MMRs are able to efficiently repair short indels (21).

When the type of BPSs accumulated along MA experiment was studied, a clear mutational bias for transitions was revealed in both wild-type and ΔnucS strains (A:T→G:C and G:C→A:T). The loss of NucS led to a notable ~65-fold increase in the transition rate, while the transversion rate remained almost unchanged (increasing less than twofold). The transition/transversion ratio changed from 1.65 in the wild-type to 62.8 in the nucS-deficient strain (38-fold higher). While NucS corrects transitions very efficiently, its activity on transversions is extremely poor. In this sense, canonical MMR-bearing bacteria tend to have lower activity on transversions than on transitions, as seen in previous MA experiments (Table 3 and table S4). However, a remarkable feature of the M. smegmatis NucS system, shared by NucS-bearing bacteria (see data from C. glutamicum in Table 3), is the extremely low activity on transversions and also the lack of activity on short indels.

The set of six BPSs reporter plasmids revealed which type of point mutations can be repaired by NucS. The study of the mutation rate of each specific mutation confirmed that both transitions (A:T→G:C and G:C→A:T) are the preferred mutations corrected by NucS, strongly increasing their rate when this repair protein is removed. No or weak increase was detected for all the four types of transversions in the nucS-deficient strain.

When the mutational spectrum was analyzed, a change in the mutational bias was detected along the MA evolution. While the wild-type strain accumulated G:C→A:T transitions, the mutational profile of the ΔnucS variant was dominated by A:T→G:C transitions. It suggests that, although NucS could repair very efficiently both types of mismatches, A:T→G:C is the predominant substrate corrected by the noncanonical MMR, as it becomes the most abundant mutation when NucS is absent. The mutational spectrum in the absence of the noncanonical MMR suggests that this system prevents further increases in the already G:C-rich mycobacterial genome, in agreement with previous reports that proposed a bias that drives genomes toward greater A:T content (15, 28–35). Together, our data strongly support the role of NucS as a repair protein able to remove mutations and protect the integrity of the genome, confirming that the NucS-based noncanonical MMR is a specific DNA repair mechanism that corrects transitions in the mycobacterial genome.

A model for DNA repair mechanisms that cooperate to avoid and remove mutations can be proposed in mycobacteria. The mycobacterial
Table 3. Mutation rates in wild-type and MMR-deficient strains from bacterial species as obtained from MA datasets.

| Type of mutation | This study | Previous studies† |
|------------------|------------|-------------------|
|                  | Mycobacterium smegmatis | Corynebacterium glutamicum | Escherichia coli | Deinococcus radiodurans‡ | Bacillus subtilis | Pseudomonas fluorescens | Vibrio cholerae | Vibrio fischeri |
| wt               | &Delta;nucS | Increase§ | wt | &Delta;nucL | Increase | wt | &Delta;nutS | Increase | wt | &Delta;mutS | Increase | wt | &Delta;mutS | Increase | wt | &Delta;mutS | Increase | wt | &Delta;mutS | Increase | wt | &Delta;mutS | Increase | wt | &Delta;mutS | Increase |
| Total BPSs       | 5.8        | 236.8 | 40.9 | 8.9 | 461.3 | 51.7 | 2 | 274.7 | 137 | 5 | 18.6 | 3.7 | 3.3 | 331 | 101.1 | 0.9 | 286.9 | 308.2 | 1.1 | 90.6 | 84.9 | 2 | 65.4 | 322 |
| Transitions      | 3.6        | 233.1 | 64.7 | 6.2 | 457.8 | 73.4 | 1.1 | 268.4 | 238.2 | 3.2 | 17.7 | 5.6 | 2.5 | 321.7 | 130.7 | 0.6 | 283.3 | 481.1 | 0.6 | 89.1 | 137.1 | 0.9 | 64.1 | 681.8 |
| A&T→G:C         | 0.9        | 139.5 | 147.1 | 2.3 | 205.4 | 90.6 | 0.4 | 192.9 | 457.5 | 1.1 | 6.1 | 5.4 | 1.2 | 158.3 | 128.2 | 0.2 | 144.4 | 754.7 | 0.1 | 44.5 | 302.8 | 0.3 | 38.7 | 1456.6 |
| G>C→A:T         | 2.7        | 93.6 | 35.3 | 4 | 252.4 | 63.5 | 0.7 | 75.6 | 107.1 | 2 | 11.6 | 5.7 | 1.2 | 163.4 | 133.3 | 0.4 | 138.9 | 349.3 | 0.5 | 44.6 | 88.7 | 0.7 | 253.7 | 376.2 |
| Transversions    | 2.2        | 3.7 | 1.7 | 2.7 | 3.5 | 1.3 | 0.9 | 6.3 | 7.1 | 1.8 | 0.9 | 0.5 | 0.8 | 9.3 | 11.4 | 0.3 | 3.7 | 10.7 | 0.4 | 1.5 | 3.6 | 1.1 | 13.5 | 12.4 |
| A→T→T:A         | 0.2        | 0.7 | 3.6 | 0.3 | 0 | 0 | 0.1 | 2.4 | 16.2 | 0.2 | 0.3 | 1.2 | 0.3 | 3.9 | 14.8 | 0.01 | 0.7 | 59.9 | 0.05 | 0.4 | 7.6 | 0.04 | 5.8 | 137.7 |
| A→T→C→G        | 0.6        | 0.6 | 1 | 1 | 2.6 | 2.6 | 0.3 | 1.7 | 5.2 | 1.1 | 0.3 | 0.3 | 0.3 | 2.7 | 10.6 | 0.1 | 0.4 | 3.7 | 0.1 | 0.3 | 2 | 0.2 | 0.9 | 3.9 |
| G>C→T:A         | 1.1        | 1.8 | 1.5 | 1 | 0.9 | 0.9 | 0.3 | 1.7 | 6.5 | 0.4 | 0 | 0 | 0.2 | 1 | 4.5 | 0.1 | 0.9 | 6.6 | 0.2 | 0.4 | 1.8 | 0.7 | 49 | 6.8 |
| G>C→C→G        | 0.3        | 0.7 | 2.4 | 0.4 | 0 | 0 | 0.1 | 0.5 | 3.5 | 0.1 | 0.3 | 2.5 | 0.1 | 1.8 | 23.4 | 0.1 | 1.7 | 18.2 | 0.05 | 0.5 | 11.5 | 0.1 | 18 | 19.4 |
| Total indels     | 1.8        | 1.8 | 1 | 0.9 | 0 | 0 | 0.2 | 51.7 | 286.3 | 0.2 | 4.7 | 21.4 | 1.2 | 52.3 | 44.3 | 0.1 | 24.7 | 209.5 | 0.2 | 24.2 | 142.2 | 0.6 | 57.6 | 101.1 |
| Insertions       | 12         | 1.4 | 1.1 | 0.4 | 0 | 0 | 0 | 0.1 | 2.2 | 19 | 0.3 | 31.9 | 111.4 | 0.03 | 22.8 | 687.1 | 0.04 | 12.5 | 323.2 | 0.2 | 25.2 | 120.6 |
| Deletions        | 0.6        | 0.4 | 0.7 | 0.5 | 0 | 0 | 0.1 | 2.5 | 24.2 | 0.9 | 20.4 | 22.9 | 0.1 | 1.9 | 22.6 | 0.1 | 11.7 | 89 | 0.4 | 3.2 | 89.8 |
| Overall          | 7.6        | 238.5 | 31.5 | 9.8 | 461.3 | 46.9 | 2.2 | 326.4 | 149.4 | 5.2 | 23.3 | 4.5 | 4.5 | 383.3 | 86.1 | 1 | 311.6 | 297.1 | 1.2 | 114.8 | 92.8 | 2.6 | 712.3 | 273.7 |

*Values have been adjusted to one decimal when possible. †All values have been calculated according to the method applied in our study taking the original data from the following sources: C. glutamicum (8), E. coli (15), B. subtilis (21, 26), D. radiodurans (25), P. fluorescens (21), V. cholerae (27), and V. fischeri (27). ‡Data from chromosomal mutations. §Fold increase [MMR-deficient versus wild-type (wt) strain mutation rates]. ¶Blank spaces: No data available.
DNA polymerase DnaE1 has a very accurate and faithful activity during replication, with a very low error rate (17). DNA polymerase DnaE1 contains a PHP domain with 3'-5' exonuclease activity, able to correct very efficiently both base pair mutations and indels during DNA replication. The absence of PHP exonuclease activity led to 7 to 11 mutations per genome per generation, a 2300- to 3700-fold increase over the wild type in a *M. smegmatis* mutant accumulation assay (17). While short indels can be almost suppressed by the DNA polymerase exonuclease activity, base pair mutations require an additional correction mechanism: the NucS-based noncanonical MMR. NucS is able to remove the remaining base pair mutations (mainly transitions) to reach wild-type basal mutation rates. Moreover, it has been shown that mismatches inhibit DNA polymerase activity in mycobacteria (17, 36). Hence, NucS activity could also be required for the replisome to facilitate DNA polymerization. Further studies are needed to shed light on the interplay between the DNA polymerase and the mismatch-specific nuclease NucS at the molecular level.

Our results provide an exceptional view of how the noncanonical MMR pathway works in vivo to maintain the low mutation rate and GC content of mycobacterial genome. Notably, despite the absence of structural homology and the different evolutionary origin (4), this mycobacterial noncanonical MMR system displays function and bias similar to the canonical MMR, in terms of BPSs (transitions), in other organisms. This reveals a clear example of functional evolutionary convergence of two nonrelated MMR pathways toward similar repair functions and highlights the importance of MMR for the stability of genomes. However, the similarities and differences (such as the inability to correct small indels and its very poor activity on transversions) between NucS and canonical MMR systems reveal that this alternative DNA repair pathway has its own identity and unique properties.

Last, hypermutable bacterial pathogens, very often associated with defects in MMR components, are frequently isolated and pose a serious risk in many clinical infections (37, 11). In addition, the existence of hypermutable *M. tuberculosis* strains deficient in NucS activity has been suggested (4). Because *M. tuberculosis* acquires antibiotic resistance exclusively by mutations, the knowledge of mechanisms that influence the generation of mutations may allow the development of new strategies to predict and combat antibiotic resistance in this deadly pathogen.

### Genomic DNA preparation and whole-genome resequencing analysis

The parental strains and the evolved lines (11 wild-type–derived strains and 11 Δ*nucS*-derived strains) were analyzed by whole-genome sequencing. DNA was extracted using the standard protocol for a high-quality preparation of mycobacterial genomic DNA (38). DNA concentration and purity were measured using a Qubit 3.0 fluorometer (Life Technologies) and a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Whole-genome sequencing libraries were constructed with the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA) following the manufacturer’s instructions. Sequencing was performed on the Illumina MiSeq instrument using a MiSeq V3 sequencing kit to obtain 300-bp paired-end reads. The obtained Illumina reads were filtered with the fastp tool to remove low-quality bases. After that, the filtered reads were aligned to the reference genome (*M. smegmatis* reference genome GenBank NC_018289.1) with the short-read alignment tool BWA. Potential duplicates were removed with Picard tools. The variant calling was performed with VarScan. A single-nucleotide polymorphism (SNP) was called if (i) we have at least 20 reads supporting the genomic position, and (ii) it was found at a frequency of 0.9 or higher and (iii) was not found near an indel region (10 bp) or in regions of high density of SNPs (≥3 SNPs in 10-bp window). Indels were called and filtered (Fisher Strand > 200, Strand Odds Ratio > 3, RMSMappingQuality < 40, ReadPosRankSum < -20, and Depth Coverage < 20) with Genome Analysis Toolkit (GATK). Single variants and indels were annotated by using SnPEFF. The average of sequence coverage per line corresponded to >99.9% of the *M. smegmatis* reference genome. Mutations were identified according to the detected changes in the DNA sequence between the evolved strains and the parental strain in each case. BPSs were annotated with the type of BPS and the position in the reference genome. BPSs were determined to affect coding DNA (synonymous or nonsynonymous BPSs) or intergenic noncoding DNA.

### MATERIALS AND METHODS

#### Bacterial strains and media

The *M. smegmatis* wild-type reference strain mc² 155 (American Type Culture Collection, 700084) and a noncanonical MMR strain (Δ*nucS*) were used in this study. The noncanonical MMR strain is a derivative of *M. smegmatis* mc² 155 with an in-frame deletion of the target gene *nucS* (MSMEG_4923) (4). *M. smegmatis* strains were grown in Middlebrook 7H9 or 7H10 with 0.5% glycerol, 0.05% Tween 80, and 10% OADC (Oleic Albumin Dextrose Catalase) enrichment.

#### MA experiments

*M. smegmatis* MA independent lines were evolved in parallel starting from the parental strains, wild-type mc² 155, and its noncanonical MMR-deficient (Δ*nucS*) derivative. Each strain generated a total number of 11 MA lines able to evolve independently for 350 days (50 weeks). Evolution of the MA lines allowed the accumulation of spontaneous mutations at random in their bacterial genomes. The evolution of each line was performed on Middlebrook 7H10 solid agar to generate isolated colonies. Every week, a well-isolated single colony (the farthest from the inoculum site) from each MA line was transferred by streaking to a new plate. Each line passed through a total number of ~1300 to 1400 cell divisions (with an average number of generations per passage estimated to be ~26 to 27 cell divisions). The number of generations per passage was estimated by measuring the number of viable cells in wild-type and Δ*nucS* colonies of different size, assuming that each colony was generated by a single cell. Colonies were excised from the agar plates, resuspended in saline solution (0.9% NaCl supplemented with 0.05% Tween 80), and plated on 7H10 petri dishes by serial dilutions. In this way, by measuring the diameter of colonies from MA lines, we could determine the number of cells per colony. The number of generations (*n*) was then calculated by *n* = log₂*N*, with *N* being the number of cells per colony. MA lines were incubated at 37°C under aerobic conditions. When all the evolved MA lineages reached 50 weeks of growth in solid medium without selective pressure, all lineages were saved and stored at −80°C. The procedure used for this experiment ensures that each line had a single-cell bottleneck in every passage to fresh medium. Mutations accumulated in an effectively neutral fashion minimizing selective pressure and bias.

#### Reporter plasmid design and construction

A set of six BPS reporter plasmids was constructed by introducing base pair changes: two transitions (A:T→G:C and G:C→A:T) and four transversions (G:C→T:A, G:C→C:G, A:T→C:G, and A:T→T:A). All point mutations were designed and constructed at position 394–396 bp of...
the \textit{aph} gene, conferring kanamycin resistance (\textit{\text{Kan}}^R) in the pMV361 integrative plasmid (39), to inactivate the \textit{Kan}^R phenotype. In addition, two frameshift reporter plasmids were made by the insertion or deletion of 1 bp (a G nucleotide) at position 624–627 bp in a stretch of GGGG in the \textit{aph} gene. The susceptibility phenotype to kanamycin of the resulting plasmids was analyzed by calculating the minimal inhibitory concentration (MIC) of kanamycin in broth. A hygromycin-resistant (\textit{Hyg}^R) gene was cloned in the reporter vectors to generate a functional antibiotic marker to select plasmid transformation and integration. All the changes in the sequence were carried out by site-directed mutagenesis using the primers listed in table S5.

\textbf{Principles of the design and evaluation of reporter plasmids}

When a pMV361 plasmid with an intact \textit{aph} gene was introduced in mycobacteria, cells produce an active enzyme able to modify and inactivate this antibiotic by phosphorylation, conferring kanamycin resistance. However, when an essential histidine at position 132 (His\textsuperscript{132}) was replaced by any other amino acid, the resulting enzyme lost its activity. Mutated \textit{aph} gene was inactive and unable to confer \textit{Kan}^R when it was inserted in the \textit{M. smegmatis} chromosome.

The set of plasmids with specific BPSs or indels was integrated in the \textit{M. smegmatis} chromosome at the unique \textit{att} site in both wild-type and \textit{nucS} strains and selected using hygromycin. The MIC of kanamycin for \textit{M. smegmatis} strains that carried the integrated plasmids was very low (MIC, 2 \mu g/ml). By contrast, the intact \textit{aph} gene conferred a high level of kanamycin resistance (MIC, 128 \mu g/ml) in both wild-type and its \textit{nucS} derivative. All the \textit{aph} mutated alleles therefore are inactive in mycobacteria as the result of a single specific BPS or indel in each reporter plasmid.

The \textit{Kan}^R phenotype can be recovered exclusively by a specific BPS that restores a functional wild-type \textit{aph} gene affecting only the previously modified base. When integrated in the mycobacterial chromosome, each base substitution and indel construction allowed us to measure the rate of each spontaneous point mutation at this specific site. The \textit{aph} reversion system is capable of detecting very low rates of each type of mutation.

We have verified by sequencing that the \textit{Kan}^R colonies were generated by reversion at a specific point previously modified by site-specific mutagenesis and not by additional chromosomal mutations. High-level \textit{Kan}^R-conferring mutations result by modification of the 16S ribosomal RNA (\textit{rRNA}), a structural component of the ribosome, through \textit{rrn} gene mutations in mycobacteria. However, \textit{M. smegmatis} harbors two copies of the \textit{rrn} gene (\textit{rrn}A and \textit{rrn}B), and it is not able to acquire high-level \textit{Kan}^R by a single-allele \textit{rrn} mutation. As a result, high-level \textit{Kan}^R (50 \mu g/ml) cannot be reached by spontaneous mutation at a detectable level (mutation rate, \(\leq 1 \times 10^{-10}\)).

\textbf{Reversion assay and estimation of mutation rates}

Reporter plasmids containing BPSs or indels were transformed in the \textit{mc}\textsuperscript{2} 155 wild-type and \textit{nucS} strains by electroporation. All the constructions were integrated at the unique \textit{att} site in the \textit{M. smegmatis} genome and selected using hygromycin (50 \mu g/ml). The analysis of the mutation rates of each specific mutation was developed by a reversion assay, measuring the acquisition of \textit{Kan}^R by the emergence of each specific type of mutation able to restore a functional wild-type \textit{aph} gene. The reversion of each mutation was verified by polymerase chain reaction amplification and sequencing of the \textit{aph} gene from five independent colonies from each test. We previously verified that no \textit{Kan}^R colonies were generated by mutations in the absence of the \textit{aph} gene by plating cultures of the strains \textit{mc}\textsuperscript{2} 155 wild type and \textit{nucS} harboring plasmid pMV361-Hyg-R on plates with kanamycin (see also the Supplementary Materials).

Mutation rates were experimentally addressed using Luria-Delbrück fluctuation assays. For each specific mutation, 16 to 24 independent cultures derived from \textit{mc}\textsuperscript{2} 155 and \textit{nucS} strains (carrying the proper reporter plasmid) were grown on 10 ml of \textit{7H9} broth and plated in \textit{7H10} solid medium with kanamycin (50 \mu g/ml) to determine the number of mutants. Viable cells were estimated by plating serial dilutions in \textit{7H10} solid medium without any antibiotic, and plates were incubated for 3 to 5 days at 37\degree C. Once viable cells and reversion mutants arose, colony counts were carried out to obtain the number of viable cells and mutants for each type of mutation. The distribution of mutant counts between parallel cultures was used to calculate mutation rates. The estimated number of mutations per culture (\textit{m}) and 95\% confidence intervals were calculated using the maximum likelihood estimator applying the \textit{Newton.LD.plating}, \textit{newton.LD.plating}, and \textit{confint.LD.plating} functions (plating functions account for differences in plating efficiency) implemented in the package rSalvador (http://eeeeric.com/rSalvador/) (Q. Zheng, R. Salvador: An essay. R package version 1.7). Last, mutation rates were calculated by dividing \textit{m} by the total number of generations, assumed to be roughly equal to the average final number of cells. After calculating mutation rates and 95\% confidence intervals, a comparative analysis was carried out between wild-type (\textit{mc}\textsuperscript{2} 155) and \textit{nucS}-lacking (\textit{nucS}) strains.

\section*{SUPPLEMENTARY MATERIALS}

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/7/eaa4453/DC1

Table S1. Genomic context of BPSs.

Table S2. Distribution of BPSs in the replicores.

Table S3. Mutation rates and 95\% confidence limits for KanR reversion events.

Table S4. Efficiency of BPS and indel repair of different bacterial MMRs obtained from MA experiments.

Table S5. Oligonucleotides used to construct BPSs and indel reporter plasmids.

\section*{REFERENCES AND NOTES}

1. E. Friedberg, G. Walker, S. Wolfram, R. Wood, R. Schultz, T. Ellenberger, \textit{DNA Repair and Mutagenesis} (American Society of Microbiology, ed. 2, 2006).
2. R. R. Iyer, A. Pluciennik, V. Burdett, P. L. Modrich, DNA mismatch repair: Functions and mechanisms. \textit{Chem. Rev.} \textbf{106}, 302–323 (2006).
3. J. A. Eisen, P. C. Hanawalt, A phylogenomic study of DNA repair genes, proteins, and processes. \textit{Mutat. Res.} \textbf{435}, 171–213 (1999).
4. A. Castraveda-García, A. I. Prieto, J. Rodríguez-Beltrán, N. Alonso, D. Cantillon, C. Costas, L. Pérez-Lago, E. D. Zegeye, M. Herranz, P. Pluciennik, T. Jonjum, D. García De Viedma, M. Paget, S. J. Waddell, A. M. Rojas, A. J. Doherty, J. Blázquez. A non-canonical mismatch repair pathway in prokaryotes. \textit{Nat. Commun.} \textbf{8}, 14246 (2017).
5. I. Matic, C. Rayssiguier, M. Radman, Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. \textit{Cell} \textbf{80}, 507–515 (1995).
6. V. Mizrahi, S. J. Andersen, DNA repair in \textit{Mycobacterium tuberculosis}. What have we learnt from the genome sequence? Mol. Microbiol. \textbf{29}, 1331–1339 (1998).
7. B. Ren, J. Kuhn, L. Meslet-Cladière, J. Brieffaux, C. Norais, R. Lavigne, D. Flament, R. Laddenstein, H. Myllykallio, Structure and function of a novel endonuclease acting on branched DNA substrates. \textit{EMBO J.} \textbf{28}, 2479–2489 (2009).
8. N. Takemoto, I. Numata, M. Su’etsugu, T. Miyoshi-Akiyama, Bacterial EndoMS/NucS acts as a clamp-mediated mismatch endonuclease to prevent asymmetric accumulation of replication errors. \textit{Nucleic Acids Res.} \textbf{46}, 6152–6165 (2018).
9. S. Ishino, Y. Nishi, S. Oda, T. Uemori, T. Sagara, N. Takatsu, T. Yamagami, T. Shirai, Y. Ishino, Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. \textit{Nucleic Acids Res.} \textbf{44}, 2977–2986 (2016).
10. S. Ishino, S. Skouloulis, H. Kudo, C. ‘\textit{Hermitte-Stead}, A. Es-Sadik, J.-C. Lambry, Y. Ishino, H. Myllykallio, Activation of the mismatch-specific endonuclease EndoMS/NucS by the replication clamp is required for high fidelity DNA replication. \textit{Nucleic Acids Res.} \textbf{46}, 6206–6217 (2018).
11. A. Oliver, R. Cantón, P. Campo, F. Baquero, J. Blázquez, High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science 288, 1251–1253 (2000).
12. J. E. LeClerc, B. Li, W. L. Payne, T. A. Cebula, High mutation frequencies among Escherichia coli and Salmonella pathogens. Science 274, 1208–1211 (1996).
13. I. Matic, M. Radman, F. Taddei, B. Picard, C. Doit, E. Bingen, E. Denamur, J. Elion, J. E. Leclerc, T. A. Cebula, Highly variable mutation rates in commensal and pathogenic Escherichia coli. Science 273, 1833–1834 (1997).
14. E. C. Cox, Bacterial mutator genes and the control of spontaneous mutation. Annu. Rev. Genet. 10, 135–156 (1976).
15. H. Lee, E. Pappodi, H. Tang, P. L. Foster, Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc. Natl. Acad. Sci. U.S.A. 109, E2774–E2783 (2012).
16. S. Kucukyildirim, H. Long, W. Sung, T. G. Doak, M. Lynch, The rate and spectrum of spontaneous mutations in Mycobacterium smegmatis, a bacterium naturally devoid of the postreplicative mismatch repair pathway. JG 6, 2157–2163 (2016).
17. J. M. Rock, U. F. Lang, M. R. Chase, C. B. Ford, E. R. Gerrick, R. Gawande, M. Coscollà, S. Gagneux, S. M. Fortune, M. H. Lamers, DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader. Nat. Genet. 47, 677–681 (2015).
18. A. Cascioferro, F. Boldrin, A. Serafini, R. Provvedi, G. Palù, R. Manganelli, Xer site-specific recombination, an efficient tool to introduce unmarked deletions into mycobacteria. Appl. Envir. Microbiol. 76, 5312–5316 (2010).
19. C. G. Cupples, M. Cabrera, C. Cruz, J. H. Miller, A set of lacZ mutations in Escherichia coli that allow rapid detection of specific frameshift mutations. Genetics 125, 275–280 (1990).
20. T. Seier, D. R. Padgett, G. Zillerberg, V. A. Sutera, N. Toha, S. T. Lovett, Insights into mutagenesis using Escherichia coli chromosomal lacZ strains that enable detection of a wide spectrum of mutational events. Genetics 188, 247–262 (2011).
21. H. Long, S. F. Miller, E. Williams, M. Lynch, Specificity of the DNA mismatch repair system (MMR) and mutagenesis bias in bacteria. Mol. Biol. Evol. 35, 2414–2421 (2018).
22. M. Kimura, On the evolutionary adjustment of spontaneous mutation rates. Genet. Res. 9, 23–34 (1967).
23. M. Lynch, Evolution of the mutation rate. Trends Genet. 26, 345–352 (2010).
24. J. W. Drake, A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U.S.A. 88, 7160–7164 (1991).
25. H. Long, S. Kucukyildirim, W. Sung, E. Williams, H. Lee, M. Ackerman, T. G. Doak, H. Tang, M. Lynch, Background mutational features of the radiation-resistant bacterium Deinococcus radiodurans. Mol. Biol. Evol. 32, 2383–2392 (2015).
26. W. Sung, M. S. Ackerman, J. F. Gout, S. F. Miller, E. Williams, P. L. Foster, M. Lynch, Asymmetric context-dependent mutation patterns revealed through mutation-accumulation experiments. Mol. Biol. Evol. 32, 1672–1683 (2015).
27. M. M. Dillon, W. Sung, R. Sebra, M. Lynch, V. S. Cooper, Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in Vibrio cholerae and Vibrio Fischeri. Mol. Biol. Evol. 34, 93–109 (2017).
28. L. Garibyan, T. Huang, M. Kim, E. Wolff, A. Nguyen, T. Nguyen, A. Diep, K. Hu, A. Iverson, H. Yang, J. H. Miller, Use of the rolA gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. DNA Repair 2, 593–608 (2003).
29. E. Becket, L. Tse, M. Yung, A. Cosico, J. H. Miller, Polynucleotide phosphorylase plays an important role in the generation of spontaneous mutations in Escherichia coli. J. Bacteriol. 194, 5613–5620 (2012).
30. C. Rewinski, M. G. Marinus, Mutation spectrum in Mycobacterium smegmatis, a bacterium defective in mismatch correction: The nature of in vivo DNA replication errors. Proc. Natl. Acad. Sci. U.S.A. 84, 6220–6224 (1987).
31. R. M. Schaaper, R. L. Dunn, 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. 105, 17826–17833 (2008).
32. P. A. Lind, D. I. Andersson, Whole-genome mutational biases in bacteria. Proc. Natl. Acad. Sci. U.S.A. 105, 17826–17833 (2008).
33. K. I. Baldi, E. P. C. Rocha, E. J. Feil, The temporal dynamics of slightly deleterious mutations in Escherichia coli and Shigella spp. Mol. Biol. Evol. 26, 345–355 (2009).
34. R. Hershberg, D. A. Petrov, Evidence that mutation is universally biased towards AT in bacteria. PLOS Genet. 6, e1001115 (2010).
35. F. Hildebrand, A. Meyer, A. Eyre-Walker, Evidence of selection upon genomic GC-content in bacteria. PLOS Genet. 6, e1001107 (2010).
36. S. Baños-Mateos, A. M. M. Van Roon, U. F. Lang, S. L. Maslen, J. M. Skehel, M. H. Lamers, High-fidelity DNA replication in Mycobacterium tuberculosis relies on a trinucleic zinc finger MMR and molecular spectrum of spontaneous mutations in Mycobacterium smegmatis, a bacterium defective in mismatch correction: The nature of in vivo DNA replication errors. Proc. Natl. Acad. Sci. U.S.A. 105, 17826–17833 (2008).
37. A. Giraud, I. Matic, M. Radman, M. Fons, F. Taddei, Mutator bacteria as a risk factor in treatment of infectious diseases. Antimicrob. Agents Chemother. 46, 863–865 (2002).
38. J. T. Beisile, M. G. Sonnenberg, in Mycobacteria Protocols. Methods in Molecular Biology, T. Parish, N. G. Stoker, Eds. (Humana Press, ed. 1, 1998), pp. 31–44.
39. C. K. Stover, V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. B. Snapper, R. G. Barletta, W. R. Jacobs Jr., B. R. Bloom, New use of BCG for recombinant vaccines. Nature 351, 456–460 (1991).

Acknowledgments: We thank J. Rodríguez-Beltrán and J. Ramón Valverde for helping us with statistics and J. L. Martínez for useful comments on the manuscript. Funding: This work was supported by Plan Nacional de I + D + i 2013–2016 and the Instituto de Salud Carlos III, Subdirección General de Redes y Centros de Investigación Cooperativa, Ministerio de Economía, Industria y Competitividad, Spanish Network for Research in Infectious Diseases (REIPI RD16/0016/0009)—cofinanced by the European Development Regional Fund “A Way to Achieve Europe”—by Operative Program Intelligent Growth 2014–2020 and grants FIS PI17/00159 from the Instituto de Salud Carlos III and SAF2015–72793-EXP from the Spanish Ministry of Science and Competitiveness (MINECO)-FEDER. E.C.-S. was the recipient of a PP15 doctoral research fellowship (F118/000336) cofinanced by the Instituto de Salud Carlos III and the European Social Fund. I.C., A.C.-O., and M.T.-P. are funded by projects of the S81 European Research Council (ERC) (638553-TB-ACCELERATE) and Ministerio de Economía y Competitividad research grant SAF 2016–77346-R. Author contributions: A.C.-G. designed the project, designed and performed experiments, made M. smegmatis strains, performed MA sequence analysis, and wrote the manuscript. J.B. designed the project, directed the experimental work, obtained funds, and wrote the manuscript. I.M.-B. constructed reporter plasmids, measured mutation rates, and performed statistical analysis. E.C.-S. performed part of the MA experiments and statistical analysis. I.C., M.T.-P., and A.C.-O. performed sequencing and MA sequence analysis. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 19 June 2019 Accepted 25 November 2019 Published 12 February 2020 10.1126/sciadv.aay4453

Citation: A. Castañeda-García, I. Martín-Blecua, E. Cebrián-Sastre, A. Chiner-Oms, M. Torres-Puente, I. Comas, J. Blázquez, Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies. Sci. Adv. 6, eaay4453 (2020).