

# Point mutations in topoisomerase I alter the mutation spectrum in *E. coli* and impact the emergence of drug resistance genotypes

Amit Bachar<sup>1</sup>, Elad Itzhaki<sup>1</sup>, Shmuel Gleizer<sup>1</sup>, Melina Shamshoom<sup>1</sup>, Ron Milo<sup>1,\*</sup> and Niv Antonovsky<sup>1,2,\*</sup>

<sup>1</sup>Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, 7610001, Israel and

<sup>2</sup>Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, New York, NY 10065, USA

Received May 24, 2019; Revised September 27, 2019; Editorial Decision November 04, 2019; Accepted November 21, 2019

## ABSTRACT

Identifying the molecular mechanisms that give rise to genetic variation is essential for the understanding of evolutionary processes. Previously, we have used adaptive laboratory evolution to enable biomass synthesis from CO<sub>2</sub> in *Escherichia coli*. Genetic analysis of adapted clones from two independently evolving populations revealed distinct enrichment for insertion and deletion mutational events. Here, we follow these observations to show that mutations in the gene encoding for DNA topoisomerase I (*topA*) give rise to mutator phenotypes with characteristic mutational spectra. Using genetic assays and mutation accumulation lines, we find that point mutations in *topA* increase the rate of sequence deletion and duplication events. Interestingly, we observe that a single residue substitution (R168C) results in a high rate of head-to-tail (tandem) short sequence duplications, which are independent of existing sequence repeats. Finally, we show that the unique mutation spectrum of *topA* mutants enhances the emergence of antibiotic resistance in comparison to mismatch-repair (*mutS*) mutators, and leads to new resistance genotypes. Our findings highlight a potential link between the catalytic activity of topoisomerases and the fundamental question regarding the emergence of *de novo* tandem repeats, which are known modulators of bacterial evolution.

## INTRODUCTION

In clonal microorganisms, spontaneous mutations in the DNA sequence are a dominant source of genetic variation (1). The rate and spectrum of mutational events are the outcome of extrinsic environmental conditions (e.g. exposure

to DNA damaging agents) and intrinsic cellular processes (e.g. DNA replication fidelity). As natural selection acts on the genetic variation within a population, identifying mutagenic mechanisms is essential to understanding evolutionary outcomes, such as the emergence of antibiotic resistance (2).

Population genetic models suggest that under strong selective conditions, such as exposure to novel environmental stress, mutator strains with high mutation rates can persist in the evolving population (3,4). Mutators, which have spontaneous mutation rates of up to three times that of the ancestral strain, are frequently observed in laboratory evolution experiments (5–7). A high prevalence of mutator strains is also found in natural bacterial populations, for example, in clinical isolates of *Pseudomonas aeruginosa* from the lungs of cystic fibrosis patients (8), and among pathogenic isolates of *Escherichia coli* and *Salmonella enterica* (9).

The genetic basis of mutator strains is often traced to sequence variation that affects one of the cellular mechanisms related to DNA replication, repair, or maintenance. A partial list of mutator alleles includes the genes of the mismatch repair system, the oxidative lesion 8-oxodG system, and mutations affecting the proofreading capacity of DNA polymerases (1). As the rate and spectrum of mutational events in mutator strains are shaped by the affected cellular mechanism (10), distinct mutational biases can affect the evolutionary dynamics of bacterial populations, as demonstrated in the case of antibiotic resistance (11).

We previously reported on the use of laboratory evolution in chemostats to adapt a metabolically engineered *E. coli* strain toward biomass synthesis from CO<sub>2</sub> (12,13). In two independent adaptive evolution experiments, we noticed distinct biases in the types of mutations that were fixed in the population. Specifically, in clones isolated from the first evolution experiment we found that ≈85% of the detected mutations (28 out of 33 mutations) were short se-

\*To whom correspondence should be addressed. Tel: +972 8 934 4466; Email: niv.anto@gmail.com  
Correspondence may also be addressed to Ron Milo. Email: ron.milo@weizmann.ac.il

quence deletions, with deletion lengths of one to four base pairs (bp) (Supplementary Table S1). In a second independent experiment, sequencing of adapted clones revealed that over 60% of the acquired mutations (14 out of 22 mutations) were base insertions. Notably, all of the insertion mutations were head-to-tail (tandem) sequence duplications (Supplementary Table S1). Tandem sequence duplication mutations are generally considered to arise from ‘slippage’ replication events (14,15), however, the newly formed duplicated sequences were independent of existing sequence repeats. A detailed description regarding the adaptive laboratory evolution process and the characterization of the adapted strains has been previously reported (12,13).

To elucidate the genetic basis underlying the biased mutation spectra that were observed in these two laboratory evolution experiments, we used whole-genome sequencing to identify potential mutator genes in the adapted clones. We found that *topA*, encoding for DNA topoisomerase I, was independently mutated in both of the evolving populations. In clones isolated from the first evolution experiment, where short sequence deletions dominated the mutation spectrum, we identified a point mutation in *topA* (g104c) that leads to a nonsynonymous substitution of an arginine residue (R35P). In the second evolution experiment, where a high frequency of tandem sequence duplication was observed, we found a point mutation (c502t) that results in the nonsynonymous substitution of another arginine residue (R168C).

Topoisomerase I belongs to the ubiquitous family of IA topoisomerases, and is known to relieve helical torsion in the chromosome by introducing transient single-strand breaks in the DNA backbone. Previous studies highlight the involvement of topoisomerases in maintaining genome stability, for example, by preventing the accumulation of DNA:RNA hybrid R-loops and the removal of DNA incorporated ribonucleotides (16,17). Considering the key roles of topoisomerase I in maintaining genome stability, we aimed to test the hypothesis that the point mutations identified in the evolving populations give rise to mutator phenotypes with a highly biased mutational spectrum.

## MATERIALS AND METHODS

### Strains and genomic modifications

An *E. coli* BW25113 strain was used as the parental strain for all further genomic modifications. The adapted strains in which the R35P and R168C *topA* mutations were originally identified have been previously described (12). For the construction of *topA* mutants (referred to in the text as strains R35P and R168C) in the background of the BW25113 strain, we used P1 transduction according to the following procedure. First, P1 lysate was prepared from the  $\Delta$ *sohB780::kan* Keio strain (JW1264-2, referred in the text as *topA*<sup>+</sup> strain), in which a kanamycin resistance marker is located in proximity ( $\approx$ 1 kbp) to the *topA* locus. The lysate was used to transduce the strains containing *topA* mutations, and the transduced cells were plated on kanamycin supplemented LB plates (25  $\mu$ g/ml). We used Sanger sequencing of the *topA* amplicons to screen for transduced clones in which the resistance marker has been introduced into the *sohB* locus without affecting the mutation in the

adjacent *topA* gene. Positive clones, in which the resistance cassette and mutated *topA* locus have been genetically linked, were used as donor strains for a second round of P1 transduction into a BW25113 strain. The short distance between the resistance marker and the *topA* locus (<1 kbp) in comparison to the large size ( $\approx$ 100 kbp) of the DNA fragment delivered to the recipient facilitated the allelic exchange of the mutated *topA* locus to the recipient BW25113 strain. Whole-genome-sequencing was used to validate the introduction of the mutated *topA* genes, along with the  $\Delta$ *sohB780::kan* marker, to an otherwise unperturbed genetic background in the BW25113 strain. The  $\Delta$ *sohB780::kan* Keio strain (JW1264-2), containing the native *topA* sequence, was used as a control strain in all of the genetic assays. *sohB* is a non-essential gene that encodes polypeptide S49 peptidase family protein. No phenotypes related to mutation rate and spectrum are associated with the  $\Delta$ *sohB780::kan* genotype. The strain referred to in the text as *mutS* is the  $\Delta$ *mutS738::kan* from the Keio collection (JW2703-2).

### Drug resistance assays

All assays were performed in 96-well microplates. A single colony from each strain was inoculated into M9 minimal media supplemented with glycerol (2 g/l) as a carbon source and suitable antibiotics (kanamycin 25  $\mu$ g/ml). Following overnight incubation (37°C, 220 RPM), the saturated culture was diluted to a concentration of  $\approx$ 10<sup>5</sup> cells/ml based on OD measurements, and 10  $\mu$ l ( $\approx$ 10<sup>3</sup> cells) were inoculated into 150  $\mu$ l of M9 minimal media supplemented with a limiting amount of glycerol (40 mg/l) to control the overall number of cell divisions. The culture was incubated for 24–48 h, after which 5  $\mu$ l drops from each well were plated on an M9 agar plate supplemented with glycerol (2 g/l) and the counter-selection drug. The assay was repeated twice, using either 2-deoxy-D-galactose (DOG) at a final concentration of 2 mg/ml (Sigma-Aldrich, D4407) or Azidothymidine (AZT) at a final concentration of 5  $\mu$ M (1.3  $\mu$ g/ml) (Sigma-Aldrich, A2169) as the counter-selection drug. The agar plates were incubated for 48 h and resistant colonies were sampled to determine the resistance conferring mutation in each colony. Only a single colony was sampled from each culture to ensure the analysis of independent mutational events. For each sampled colony, we used PCR to amplify and sequence the known genes (*galK*/DOG and *tdk*/AZT) whose loss-of-function rendered the bacteria resistant to the drug. The resulting sequences were compared to the wild-type sequences of *galK* or *tdk* genes and mutations in the resistant colonies were annotated using Geneious software (<http://www.geneious.com>). Assays for D-cycloserine (DCS) resistance were performed as described above, with the following modifications: (a) LB media was used, (b) DCS concentration was 133  $\mu$ g/ml, (c) mutation calling was performed by the whole-genome-sequencing of resistant colonies as described below.

### Mutation accumulation lines

For each strain, a glycerol stock was inoculated in LB media, incubated overnight, and streaked on an LB agar

plate supplemented with kanamycin (25 ug/ml). Isogenic lines were established from a single, well isolated colony. Lines were propagated daily by streaking a random single colony from each line on LB agar plates supplemented with kanamycin (25 ug/ml). The lines were maintained for 48 passages ( $\approx$ 1200 generations) for *topA*<sup>+</sup> and R168C strains or 24 passages ( $\approx$ 600 generations) for strains R35P and R35P *gyrA*. As the growth rates of the different strains varied, different incubation times were needed to obtain comparable generations per passage rates, as determined by CFU counting of typical colonies ( $\approx$ 25 generations/passage). Freezer stocks were sampled from the founding strains and every 25 passages. Whole-genome sequencing was used to determine the mutations accumulated in the final clones, as detailed below. The mutation rate per genome replication was calculated for each mutation type based on the average number of mutations across all lines for each strain.

### Whole-genome-sequencing and mutation calling

A single colony was inoculated in 3 ml of LB media with kanamycin (25 ug/ml) and incubated to saturation. The culture was harvested by centrifugation and genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN). Sequencing libraries were prepared using Nextera tagmentation reactions as described by Baym *et al.* (2015). The final libraries were sequenced using HiSeq 2500 or NextSeq 500 sequencers (Illumina) to yield single reads with a typical coverage of  $\times$ 30 per sample. Demultiplex reads were mapped to the reference of *E. coli* BW25113 (CP009273.1) and mutations were called using breseq pipeline version 0.25 (18). We note that our analysis does not capture certain types of mutations, such as copy number variations.

### Purification of recombinantly expressed TopA variants

His-tagged topoisomerase I proteins (R35P, R168C and wild-type) were expressed in *E. coli* BW25113 transformed with the topoisomerase I encoding plasmid pCAN24-6xHis-TopA obtained from the ASKA library (19). The R35P and R168C encoding mutations were introduced to the coding sequence of *topA* using PCR based site-directed-mutagenesis according to standard protocols. The resulting plasmids, pCAN24-6xHis-TopA:R35P and pCAN24-6xHis-TopA:R168C, were sequenced to validate the mutated sequence. For recombinant protein expression, a single colony of *E. coli* BW25113 transformed either with the native or one of the mutated plasmids, was inoculated in 3 ml of LB media supplemented with chloramphenicol (34 ug/ml). Following overnight incubation (37°C, 220 RPM) the saturated culture was diluted 1:20 into 50 ml of fresh media. Protein over-expression was induced at the mid-exponential phase ( $\approx$ 0.5 OD) by the addition of IPTG (0.5 mM). The cells were harvested and the protein was purified using a His-Spin Protein Miniprep kit (Zymo Research) according to the manufacturer's instructions. Following protein purification, we used an Amicon Ultra 30K centrifugal filter (Merck Millipore) to replace the elution buffer with the plasmid relaxation assay reaction buffer.

Protein purity was validated using polyacrylamide gel electrophoresis and GelCode Blue staining (Thermo Fisher). Purified protein concentration was determined using the bicinchoninic acid assay (Thermo Fisher) according to the manufacturer's protocol.

### Plasmid relaxation assay

Purified enzymes of the same concentrations were diluted serially and assayed for DNA relaxation activity in a final reaction volume of 10 ul with 40 mM Tris-HCl (pH 8.0), 50 mM KCl, 10 mM MgCl<sub>2</sub>, 100 ug/ml BSA, and 800 ng of supercoiled plasmid DNA (pNiv vector, as described in (20)). The vector sequence does not contain designated topoisomerase I cleavage sites. The reaction was incubated at 37°C for 30 min, after which it was terminated by the addition of SDS to a final concentration of 1%. The DNA was electrophoresed in a 0.8% (w/v) agarose gel with TAE buffer for 45 min at 135V. The gel was stained with ethidium bromide, washed twice in DDW for 1 h, and photographed under UV light. We estimated the relative activity according to the defined least quantity of the enzyme required for complete relaxation of negatively supercoiled DNA under the given reaction conditions.

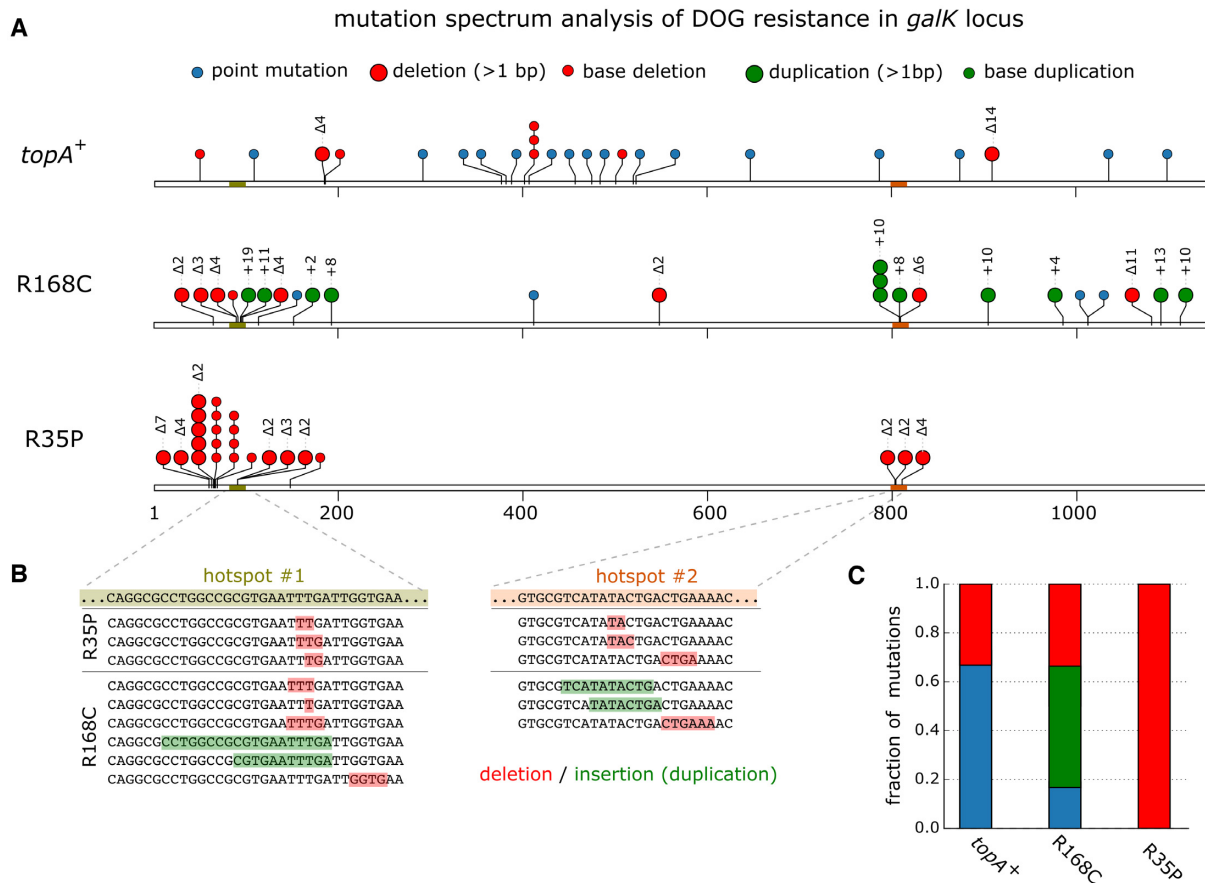
## RESULTS

### R35P and R168C topoisomerase I mutants show biased mutation spectra, enriched in sequence deletions and insertions

To explore the potential effect of topoisomerase I on the spectrum of mutational events, we genetically introduced the mutated *topA* alleles to an *E. coli* BW25113 genetic background (Materials and Methods). Whole-genome sequencing verified the replacement of the native *topA* gene with sequences encoding for either the R35P or the R168C mutated enzyme variants. In addition to the mutations introduced through allelic replacement, in the R35P strain we identified four mutations in genes unrelated to DNA processing which were most probably acquired during early handling of the strain (Supplementary Table S2). In addition, we observed that mutations in *gyrA* or *gyrB* genes, with a positive fitness effect, often emerged in the R35P strain during continuous strain propagation. Compensatory mutations in gyrase genes can reduce the accumulation of DNA torsional stress, and were reported to be essential for the viability of *topA* null mutants (21).

Next, we quantified the spectrum of mutational events (i.e. point mutations, insertions and deletions) in the *topA* mutant strains and in a *topA*<sup>+</sup> strain using two independent drug resistance assays (Materials and Methods). As selection agents, we used 2-deoxy-D-galactose (DOG), an inhibitor of galactokinase (encoded by *galK*), and azidothymidine (AZT), an inhibitor of thymidine/deoxyuridine kinase (encoded by *tdk*). Both drugs inhibit the growth of *E. coli* on minimal media via interactions with known molecular targets (22,23). Briefly, we plated cells that were cultured in drug-free media on solid minimal media containing the toxic drug (either DOG or AZT) and a suitable carbon source (Materials and Methods). Resistant mutants that gave rise to colonies were isolated, and the resistance-conferring mutation in





**Figure 1.** Drug resistance assay reveals an enrichment of insertion and deletion mutations in *topA* mutants. (A) Mutation spectrum analysis of DOG resistant colonies arising in R168C and R35P mutant strains or from the *topA*<sup>+</sup> strain. To determine the resistance conferring mutations, we used Sanger sequencing of PCR amplicons of the *galK* locus. Overall, we sampled 24 independent resistant colonies for each genetic background. Notably, only a single colony was sampled from every assay to ensure the analysis of independent mutational events. (B) Resistance conferring mutations arising in the *topA* mutants were mostly localized in two mutation hotspots in *galK*. (C) In contrast to the *topA*<sup>+</sup> strain, where the majority of resistance conferring was due to point mutations, *topA* mutants are highly enriched in deletions (R35P) and tandem sequence duplications (R168C).

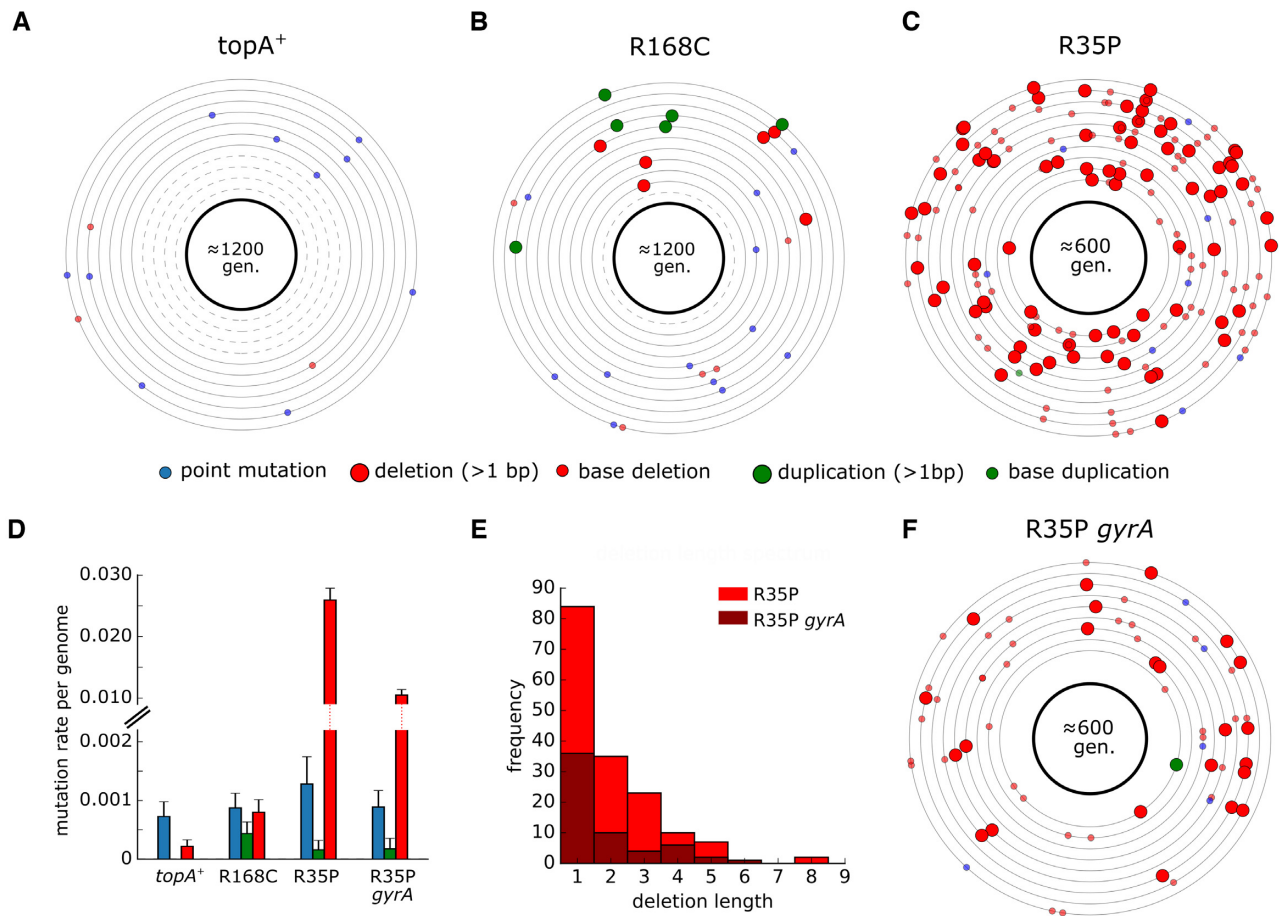
each colony was determined by sequencing the known drug target gene. To ensure the analysis of independent mutational events, only a single colony was sampled from every assay repeat.

As shown in Figure 1, the sequencing of *galK*, which is the molecular target of DOG, revealed that resistant colonies arising from *topA* mutants exhibit significantly altered mutation spectrum in comparison to the *topA*<sup>+</sup> strain ( $n = 24$  for each strain, Supplementary Table S3). In colonies arising from the R168C mutant strain, we found that 50% of the resistance conferring mutations in *galK* were sequence insertions, composed entirely from *de novo* tandem duplications with lengths of 2–19 bp. In the R35P mutant strain, we found that all of the resistance conferring mutations in *galK* were due to short sequence deletions, with lengths of 1–7 bp. These results were in marked contrast with the point mutation dominated spectrum ( $\approx 66\%$ ) that was observed in resistant colonies arising from the *topA*<sup>+</sup> strain. When we repeated the experiment using AZT as a selection drug, the results were in agreement with the values reported above (Supplementary Figure S1 and Table S3). In addition to the observed differences in the mutation spectra of *topA* mutants in comparison to the *topA*<sup>+</sup> strain, we noted that both R35P

and R168C mutants exhibited a strong localization pattern of mutations with well-defined hotspots (Figure 1B). This contrasts with the uniform spatial distribution of mutations observed in the *topA*<sup>+</sup> strain, where mutations occurred across the *galK* coding sequence without noticeable hotspots. Similarly, a mutational hotspot was also observed in the *topA* mutants when AZT was used as a selection drug, where  $\approx 50\%$  of the resistance conferring mutations were localized in a 20 bp region (Supplementary Figure S1). Comprehensive sequence data for the mutations identified in this experiment can be found in Supplementary Table S3. We computationally compared (using MEME suite 5.0.5) the hotspot mutation regions in *galK* and *tdk* genes to a set of previously characterized topoisomerase I preferred cleavage motifs (24), but did not find that these regions match any of the known sequences.

#### Estimation of genomic mutation rates in R168C and R35P topoisomerase I mutants using mutation accumulation lines

To quantify the genome-wide mutation rate of *topA* mutants, we conducted a mutation accumulation assay in which isogenic replicates were passed through a single-



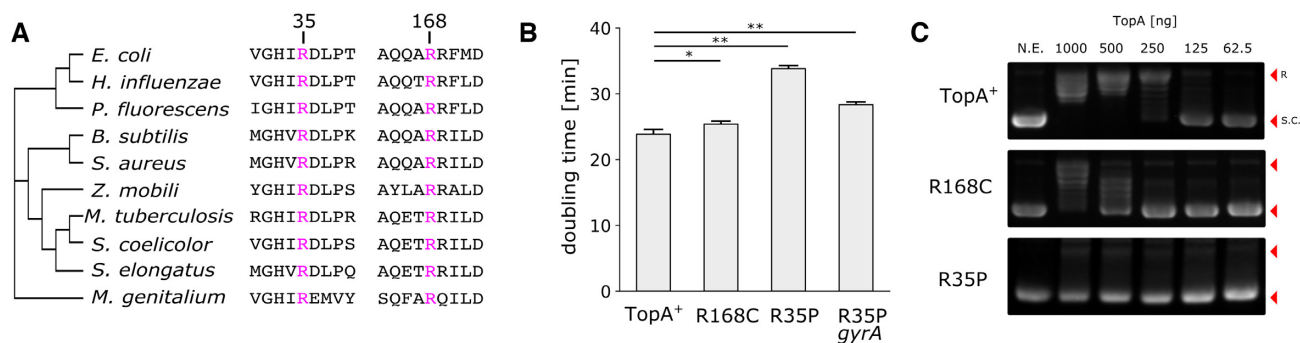
**Figure 2.** Rate and molecular spectrum of mutations in *topA* and wild-type strains as estimated by mutation accumulation lines. (A–C) Each circle represents the genome of a single mutation accumulation line ( $n=11$ ). Mutations that were identified by whole-genome sequencing are positioned on the circle according to their chromosomal coordinates (clockwise). Dashed circles indicate genomes in which no mutation was detected. Detailed lists of mutations can be found in Supplementary Table S4. (D) Average mutation rate per genome replication calculated for each mutation type. Error bars represent SEM. (E) Histogram of short deletion lengths that were identified in R35P mutant strains. (F) Mutations identified in isogenic lines ( $n=9$ ) of the R35P strain that contains a compensatory mutation in *gyrA*. Compensatory mutations in gyrase genes in *topA* mutants were reported to decrease the accumulation of DNA torsional stress and are essential for the viability of *topA* null mutants.

colony bottleneck for 1200 or 600 generations. We used whole-genome sequencing to identify the mutations that accumulated in each line and calculated the rates of deletion, insertion, and point mutations for the different strains (Supplementary Table S4). As shown in Figure 2, our analysis revealed that the rate of short sequence deletion events in the R35P mutant is significantly higher in comparison to the *topA*<sup>+</sup> strain ( $P$ -value  $< 0.001$ ,  $t$ -test) with an  $\approx 100$ -fold change in the mean rate of short deletions.

Since no sequence duplication events were observed in the *topA*<sup>+</sup> lines during our experiment, we cannot directly determine the change in frequency of tandem duplication events. Previously published estimations of sequence insertion rates in wild-type *E. coli* found it to be  $\approx 1 \times 10^{-4}$  mutations per genome per generation (25), a rate which is 5-fold-lower than the rate we observed in the R168C mutant. However, this estimation of the insertion rate in the wild-type strain is dominated by the rate of short sequence insertions (1–3 bp), occurring mainly in homopolymer tracts. As longer insertions, particularly those outside the context of existing sequence repeats, are significantly more rare, we

note that the increase in the rate of *de novo* duplication mutations in the R168C is likely to be significantly higher. In contrast to the marked changes in the frequencies of deletion and duplication events, we found no significant difference in the rate of point mutations between the mutant and the *topA*<sup>+</sup> strains. The point mutation rates measured for all of the strains in our analysis were in agreement with previously reported values of  $\approx 1\text{--}2 \times 10^{-3}$  mutations per genome per generation (25).

In addition, we used mutation accumulation lines to measure the effect of spontaneously arising compensatory mutations on the mutation rate in *topA* mutants. We isolated a clone in which a spontaneous mutation in *gyrA* had emerged in the R35P strain background. Based on our mutation accumulation analysis ( $n=9$ ), we found that the rate of short sequence deletions in this double mutant remained significantly higher than the *topA*<sup>+</sup> strain ( $P$ -value  $< 0.001$ ,  $t$ -test). However, the mutator phenotype was attenuated with an  $\approx 2$ -fold decrease in the deletion rate in comparison to the R35P which lacks the compensatory mutation. Taken together, our results unequivocally show that



**Figure 3.** R168C and R35P mutations affect conserved residues and decrease supercoiling relaxation activity. (A) Multiple sequence alignment of bacterial topoisomerases indicating that the mutated residues (labeled in pink), R35 and R168, are highly conserved (see also Supplementary item 1 for a broader phylogenetic analysis). (B) Doubling times of *topA*<sup>+</sup> strain in comparison to *topA* mutants in LB media. Error bars represent the standard deviation of biological repeats ( $n = 6$ ). \* $P < 0.01$ , \*\* $P < 0.001$  (C) In vitro plasmid relaxation assay shows significant reduction in DNA supercoiling activity in R168C and R35P in comparison to the native enzyme.

R35P and R168C substitutions in topoisomerase I give rise to mutator phenotypes with distinct mutational bias.

### R35P and R168C substitutions affect highly conserved residues in DNA topoisomerase I, lead to slower doubling time, and inhibit DNA relaxation activity

Sequence alignment of 1750 bacterial topoisomerase I homologs shows that both R35P and R168C substitution mutations affect conserved residues in the protein. Our analysis finds that R168 is conserved in all 1750 sequences while R35 is conserved in over 1350 sequences (75%) (Figure 3A and Supplementary item 1). Based on the structural analysis of topoisomerase I, Zhang *et al.* found that R168 participates in a network of ionic and hydrogen bond interactions which hold the DNA substrate in proper conformation for cleavage or re-ligation (26). The same study reports that an activity assay of a mutant enzyme in which these interactions have been perturbed (R168A) shows a significant (>80-fold) loss of relaxation activity when compared to the wild type enzyme.

To determine the fitness effect of these mutations, we compared the growth rate of *topA* mutants to the native strain in LB and in glucose supplemented minimal media (Figure 3B and Supplementary Figure S2). We found that in all cases, mutations in *topA* resulted in fitness cost ( $P$ -value < 0.01 for all mutant strains,  $t$ -test). The effect was most prominent in the R35P mutant, where average doubling time was  $\approx 50\%$  longer in comparison to the *topA*<sup>+</sup> strain. The compensating mutation in *gyrA* in the R35P background reduced the effect but did not eliminate the fitness cost. The smallest fitness cost was observed in the R168C mutant strain with  $\approx 5\%$  increase in doubling time in LB media. Similar fitness costs were observed when the strains were cultured in glucose-limited minimal media (Supplementary Figure S2), indicating that the fitness cost is independent of nutrient availability and of the maximal growth rate.

Next, we tested the effect of R35P and R168C mutations on the catalytic activity of the enzyme using *in vitro* DNA relaxation assay of heterologously expressed topoisomerase I mutant enzymes (Figure 3C). For both of the mutant variants we observed a decrease in plasmid relaxation activity

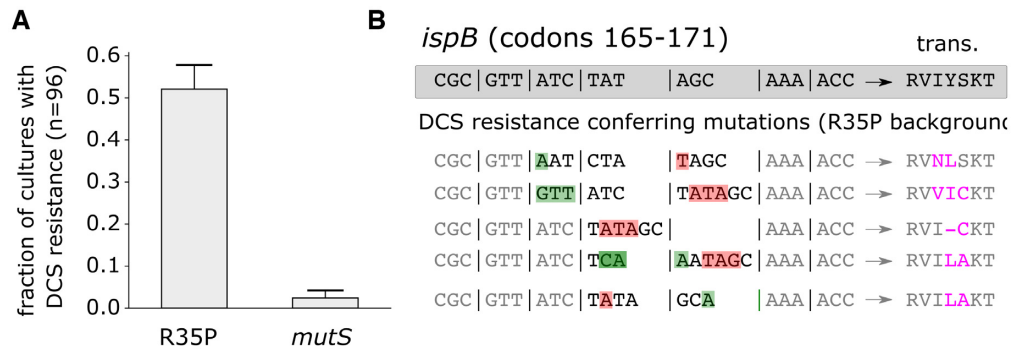
in comparison to the native enzyme. In the R168P mutant, we found that the minimal amount of enzyme required for complete plasmid relaxation was 4-fold higher in comparison to the wild-type enzyme, while in the R35P mutant protein we were not able to detect any plasmid relaxation activity.

### The biased mutation spectra in *topA* mutants impact the emergence of antibiotic resistance

Previous studies have shown that the mutational spectra in mismatch repair *E. coli* mutator strains affect the distributions of beneficial mutations in an antibiotic resistance model system (11). As the mutational spectrum defines accessible beneficial mutations, we sought to experimentally study the effect of the biased mutation spectra observed in *topA* mutants on the emergence of antibacterial resistance. We performed fluctuation assays to compare the rate of spontaneously arising resistance to D-cycloserine (DCS), a broad-spectrum antibiotic, in the R35P mutant strain and in an MMR-deficient mutator strain (*mutS*). Previous studies demonstrated that *mutS* mutators exhibit over 100-fold increase in the frequency of point mutations (27). This value is comparable to the rate of deletion mutations observed in the R35P strain, although the two strains differ in their mutation spectra (point mutations versus short sequence deletions).

Under our experimental conditions, we find that DCS resistant colonies emerged in  $\approx 50\%$  of R35P *topA* mutant cultures, but only in  $\approx 3\%$  of the *mutS* cultures (Figure 4A,  $n = 96$ ). Whole-genome sequencing of independently arising resistant colonies from R35P strain cultures ( $n = 5$ ) revealed that all of the analyzed clones acquired mutations in *ispB*, an essential gene in the biosynthesis of isoprenoid quinones. Specifically, we observed the repeated occurrence of complex mutations, combining base insertions and deletions, in a localized hotspot in the *ispB* gene. Sanger sequencing of the *ispB* gene verified that in all five occurrences, these mutations impacted 2–3 amino acids within the coding sequence (Figure 4B and Supplementary Table S5). In contrast, resistant colonies arising from cultures of the *mutS* strain ( $n = 5$ ) were dominated by point mutations with no single locus that was mutated in more than two clones (Supplemen-





**Figure 4.** Cycloserine resistance conferring mutations in *ispB* observed in a *topA* mutant (R35P) but not in a MMR mutator (*mutS*). (A) Drug resistance fluctuation assay of R35P *topA* and *mutS* (*topA*+) mutants. Error bars represent the standard deviation of three experimental repeats, with 96 cultures per experiment. (B) Whole genome sequencing of cycloserine resistant colonies arising in the R35P *topA* mutant revealed complex mutations in *ispB* ( $n = 5$ ). In all five resistant clones, at least two residues in the protein sequence were affected by a combination of sequence deletions (red) and insertions (green).

tary Table S5). The marked differences in DCS resistance rates and genotypes between the *mutS* and *topA* mutators strengthen previous observations regarding the key role of the mutational spectra on the emergence of antibiotic resistant phenotypes.

## DISCUSSION

Identifying the cellular processes that give rise to mutagenesis is crucial for the mechanistic understanding of genetic variation. The methyl-directed mismatch repair system (MMR) is a major DNA repair pathway for correcting base-base mismatches or small insertion-deletion loops generated by replication errors. The MMR complex is capable of recognizing mismatch pairing, discriminating between the newly synthesized strand and the template, removing the mismatched base from the new strand, and correctly filling the gap through repair synthesis. *E. coli* with mutations affecting components of the mismatch repair system (e.g. *mutS*), exhibit a strong mutator phenotype with over a 100-fold increase in the rate of point mutations. MMR deficient strains also show an increase in the frequency of insertion and deletion in comparison to the wild-type mutation rate; however, these are mostly limited to the gain or loss of a single nucleotides in homopolymer runs, or the loss of di-/tri- nucleotide in microsatellite repeats (28). This mutational pattern points towards the most common mechanism for the introduction of frameshift mutations, which are usually generated by polymerase slippage or unequal crossover during the replication of repeat sequences (14).

In contrast, here we show that mutations perturbing the relaxation activity of DNA topoisomerase I in *E. coli* lead to mutator phenotypes with deletion and tandem-duplication enriched mutation spectra, which are independent of existing repeat sequences. As such, it is unlikely that these mutations arise from polymerase replication errors. Potential alternative sources of the mutagenesis observed in *topA* mutants may include the direct involvement of the enzyme, a secondary mechanism linked to the accumulation of torsional strain due to the impaired activity of mutants. Other models involving additional enzymes and processes are possible. In *E. coli*, disruption of *topA* leads to excessive nega-

tive supercoiling and the accumulation of RNA:DNA hybrid R-loops and the stabilization of non B-DNA forms (29). These non-canonical DNA structures are suggested to stimulate genetic instability, as was recently demonstrated by the accumulation of deletion, insertion and complex mutations following dCas9-induced R-loop formation in yeast (30). This highlights the loss of DNA supercoiling homeostasis as a potential indirect mechanism for the mutator phenotype in *topA* strains. Alternatively, it is possible that failure to complete the catalytic cycle in the topoisomerase mutated enzymes directly leads to mutagenic events. Potential mechanisms of mutagenesis due to catalytic impairment include the formation of a stable cleavage complex leading to replication fork collapse, or through the introduction of single-strand DNA breaks. For example, it was recently demonstrated that the repair of adjacent single-strand breaks in plants can lead to the formation of tandem sequence duplications (31).

An additional characteristic of the mutation spectrum observed in *topA* mutants is the localization of mutations in sequence hotspots, as observed in all three resistance assays (Figure 1, Supplementary Figure S1 and Figure 4). Although we were not able to identify common sequence motifs in these regions, or high similarity to preferred known topoisomerase I cleavage sites, it is possible that the mutated enzymes exhibit sequence preference that is different from cleavage sites previously described for the wild-type enzyme. Alternatively, it is possible that an intrinsic feature of the sequence renders it more sensitive to the accumulation of excessive negative supercoiling or the formation of non-canonical DNA structures that can lead to mutagenic events. While our results unequivocally identify mutations in *topA* as the genetic basis of mutator phenotypes in *E. coli*, further studies are required to elucidate the molecular details underlying the difference in mutation rates and spectra between R35P and R168C *topA* mutants.

Fixation of *topA* mutants had been previously observed in multiple laboratory evolution experiments (7). For example, in a long-term evolution experiment in *E. coli*, mutations in *topA* with positive fitness effects emerged in multiple populations (32). However, to the best of our knowledge, this is the first description of a *topA* mutator phenotype with an accelerated rate of tandem duplications in *E. coli*. Tan-

dem sequence repeats are a well-documented source of genetic instability that plays a significant role in microbial evolution (33). While several molecular mechanisms pertaining to tandem repeat expansion and contraction have been described (14, 34), the mechanism underlying *de novo* formation of tandem repeats from non-repeating sequences remains unknown. Our findings highlight a link between the catalytic activity of DNA topoisomerase I and the emergence of *de novo* tandem repeats as ubiquitous source for genetic instability.

## DATA AVAILABILITY

Sequence data from whole-genome sequencing of laboratory evolution isolates, mutation accumulation lines, and DCS resistance mutants that support the findings of this study is accessible at European Nucleotide Archive [accessions PRJEB13306 and PRJEB35213].

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors wish to thank Yinon Bar-On, Dr. Dan Davidi, Roe Ben Nissan, Prof. Rotem Sorek and Prof. Zvi Livneh for the helpful discussions.

## FUNDING

Yeda-Sela Center for Basic Research, European Research Council [NOVCARBFIX 646827]; Israel Science Foundation [740/16]; Beck-Canadian Center for Alternative Energy Research; R.M. is the Charles and Louise Gartner professional chair; N.A. is supported by the Human Frontier Science Program [LT000546/2018-L]. Funding for open access charge: Yeda-Sela Center for Basic Research, European Research Council [NOVCARBFIX 646827]; Israel Science Foundation [740/16]; Beck-Canadian Center for Alternative Energy Research; R.M. is the Charles and Louise Gartner professional chair; N.A. is supported by the Human Frontier Science Program [LT000546/2018-L].  
*Conflict of interest statement.* None declared.

## REFERENCES

- Miller, J.H. (1996) Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu. Rev. Microbiol.*, **50**, 625–643.
- Chopra, I., O'Neill, A.J. and Miller, K. (2003) The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Resist. Updat.*, **6**, 137–145.
- Wielgoss, S., Barrick, J.E., Tenaillon, O., Wisner, M.J., Dittmar, W.J., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R.E. and Schneider, D. (2013) Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 222–227.
- Good, B.H. and Desai, M.M. (2016) Evolution of mutation rates in rapidly adapting asexual populations. *Genetics*, **204**, 1249–1266.
- Sniegowski, P.D., Gerrish, P.J. and Lenski, R.E. (1997) Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature*, **387**, 703–705.
- Bachmann, H., Starrenburg, M.J.C., Molenaar, D., Kleerebezem, M. and van Hylckama Vlieg, J.E.T. (2012) Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res.*, **22**, 115–124.
- Phaneuf, P.V., Gosting, D., Palsson, B.O. and Feist, A.M. (2019) ALEdb 1.0: a database of mutations from adaptive laboratory evolution experimentation. *Nucleic Acids Res.*, **47**, D1164–D1171.
- Oliver, A., Cantón, R., Campo, P., Baquero, F. and Blázquez, J. (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*, **288**, 1251–1254.
- LeClerc, J.E., Li, B., Payne, W.L. and Cebula, T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*, **274**, 1208–1211.
- Foster, P.L., Lee, H., Popodi, E., Townes, J.P. and Tang, H. (2015) Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E5990–9.
- Couce, A., Guelfo, J.R. and Blázquez, J. (2013) Mutational spectrum drives the rise of mutator bacteria. *PLoS Genet.*, **9**, e1003167.
- Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N. *et al.* (2016) Sugar Synthesis from CO<sub>2</sub> in *Escherichia coli*. *Cell*, **166**, 115–125.
- Herz, E., Antonovsky, N., Bar-On, Y., Davidi, D., Gleizer, S., Prywes, N., Noda-Garcia, L., Lyn Frisch, K., Zohar, Y., Wernick, D.G. *et al.* (2017) The genetic basis for the adaptation of *E. coli* to sugar synthesis from CO<sub>2</sub>. *Nat. Commun.*, **8**, 1705.
- Levinson, G. and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.*, **4**, 203–221.
- Zhou, K., Aertsen, A. and Michiels, C.W. (2014) The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol. Rev.*, **38**, 119–141.
- Cho, J.-E., Kim, N. and Jinks-Robertson, S. (2015) Topoisomerase I-dependent deletions initiated by incision at ribonucleotides are biased to the non-transcribed strand of a highly activated reporter. *Nucleic Acids Res.*, **43**, 9306–9313.
- Usongo, V. and Drolet, M. (2014) Roles of type IA topoisomerases in genome maintenance in *Escherichia coli*. *PLoS Genet.*, **10**, e1004543.
- Deatherage, D.E. and Barrick, J.E. (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol. Biol.*, **1151**, 165–188.
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H. and Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF A rchive): Unique Resources for Biological Research. *DNA Res.*, **12**, 291–299.
- Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., Liebermeister, W., Flamholz, A., Noor, E., Amram, S. *et al.* (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.*, **41**, e98.
- Stockum, A., Lloyd, R.G. and Rudolph, C.J. (2012) On the viability of *Escherichia coli* cells lacking DNA topoisomerase I. *BMC Microbiol.*, **12**, 26.
- Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. and Copeland, N.G. (2005) Simple and highly efficient BAC recombineering using *galK* selection. *Nucleic Acids Res.*, **33**, e36.
- Shelat, N.Y., Parhi, S. and Ostermeier, M. (2016) A positive selection for nucleoside kinases in *Escherichia coli*. *PLoS One*, **11**, e0162921.
- Annamalai, T., Dani, N., Cheng, B. and Tse-Dinh, Y.-C. (2009) Analysis of DNA relaxation and cleavage activities of recombinant *Mycobacterium tuberculosis* DNA topoisomerase I from a new expression and purification protocol. *BMC Biochem.*, **10**, 18.
- Lee, H., Popodi, E., Tang, H. and Foster, P.L. (2012) 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.
- Zhang, Z., Cheng, B. and Tse-Dinh, Y.-C. (2011) Crystal structure of a covalent intermediate in DNA cleavage and rejoining by *Escherichia coli* DNA topoisomerase I. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 6939–6944.
- Foster, P.L., Niccum, B.A., Popodi, E., Townes, J.P., Lee, H., Mohammedismail, W. and Tang, H. (2018) Determinants of base-pair



- substitution patterns revealed by whole-genome sequencing of DNA mismatch repair defective *Escherichia coli*. *Genetics*, **209**, 1029–1042.
28. Long,H., Miller,S.F., Williams,E. and Lynch,M. (2018) Specificity of the DNA Mismatch repair system (MMR) and mutagenesis bias in bacteria. *Mol. Biol. Evol.*, **35**, 2414–2421.
  29. Drolet,M. (2006) Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.*, **59**, 723–730.
  30. Laughery,M.F., Mayes,H.C., Pedroza,I.K. and Wyrick,J.J. (2019) R-loop formation by dCas9 is mutagenic in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **47**, 2389–2401.
  31. Schiml,S., Fauser,F. and Puchta,H. (2016) Repair of adjacent single-strand breaks is often accompanied by the formation of tandem sequence duplications in plant genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 7266–7271.
  32. Crozat,E., Philippe,N., Lenski,R.E., Geiselmann,J. and Schneider,D. (2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics*, **169**, 523–532.
  33. Treangen,T.J., Abraham,A.-L., Touchon,M. and Rocha,E.P.C. (2009) Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.*, **33**, 539–571.
  34. Bichara,M., Wagner,J. and Lambert,I.B. (2006) Mechanisms of tandem repeat instability in bacteria. *Mutat. Res.*, **598**, 144–163.