DNA breaks are key contributors to the cost of antibiotic resistance

Roberto Balbontín[#1] and Isabel Gordo[#1]

[1] Instituto Gulbenkian de Ciência, Oeiras, Portugal

[#] For correspondence:

E-mail: rbalbontin@igc.gulbenkian.pt and igordo@igc.gulbenkian.pt

ABSTRACT

Bacteria can become resistant to antibiotics by acquiring mutations in genes encoding the physiologically relevant proteins targeted by the drugs. Consequently, resistance mutations can cause growth defects (fitness cost) in the absence of antibiotics[1]. The cost thus, can hinder maintenance and dissemination of resistances, by enabling sensitive bacteria to outcompete resistant clones upon reducing the amount of antibiotics[2]. Besides its paramount importance, the causes of the cost of resistance are poorly understood[3]. Here we show that DNA breaks explain 73% of the variation in the cost caused by resistance mutations affecting transcription, translation, and their coupling, in *Escherichia coli*. We also reveal that the RNase HI, responsible for the specific degradation of R-loops[4], is a key determinant of resistance costs and thus a novel target for antimicrobials specific against resistant bacteria, which we validated using a repurposed drug[5]. Accordingly, we show that lack of RNase HI rapidly drives resistant clones to extinction in polymorphic populations with high resistance levels. These results reveal a key cause of the cost of resistance, and provide a conceptual framework for the development of novel strategies to lower the alarming levels of resistance currently observed in the human microbiome.

Antibiotic resistance (AR) entails a large human and economic burden worldwide[6]. The maintenance and dissemination of AR in bacterial populations depend on the rate at which resistances are acquired and on their effects on bacterial fitness (cost). The cost of AR is influenced by the environment, by interactions between the resistances and their genetic background (epistasis), and by the subsequent acquisition of mutations compensating for fitness defects (compensatory evolution)[2]. Despite its importance, the causes of the cost of AR are not completely understood[3]. Identifying those causes has become, thus, the Holy Grail in the AR field.

Resistance mutations often map to genes encoding the proteins targeted by the antibiotics. These target proteins are typically involved in essential functions, such as transcription, translation, DNA replication, or cell wall biosynthesis. Resistance mutations cause alterations in the structure of the target protein, rendering it insensitive to the drug, but often affecting its function as well[1, 2].

Rifampicin and streptomycin resistance mutations (Rif^R and Str^R) paradigmatically represent resistances to antibiotics targeting transcription and translation, respectively, and have been extensively studied in *Escherichia coli*. Rif^R mutants often show cost[7], commonly associated with alterations in the rates of transcription initiation, elongation, slippage, and termination[8]. Most Str^R mutations also cause a cost[9], generally linked to increased translation fidelity and reduced processivity[9].

Thus, the costs of Rif^R and Str^R mutations are commonly attributed to defects in protein synthesis, either globally[10], or circumscribed to specific functions or regulons[11, 12, 13]. We recently

showed that compensation for the cost of double resistance (Rif^R Str^R) can occur via overexpression of *nusG* and *nusE*[14], which encode the proteins that physically connect the RNA polymerase (RNAP) and the ribosome[15]. This suggests that compensation hinges on reinforcing the coupling between transcription and translation and, consequently, that Rif^R and Str^R mutations disturb this coordination. Both NusG and coupling between transcription and translation prevent spontaneous RNAP backtracking[16, 17], which can cause double-strand DNA breaks (DSBs)[18]. This led us to hypothesize that perturbation of transcription-translation coupling caused by Rif^R and Str^R mutations generate DSBs, which contribute to their cost.

In order to test this hypothesis, we simultaneously measured competitive fitness and activation of the SOS response – a well-known proxy for the occurrence of DSBs[19] – in sensitive *E. coli*, streptomycin resistant strains ($RpsL^{K43N}$, $RpsL^{K43T}$, and $RpsL^{K43R}$), rifampicin resistant strains ($RpoB^{H526L}$, $RpoB^{H526Y}$, and $RpoB^{S531F}$), and double resistant mutants harbouring the nine possible combinations of these resistance alleles. Interestingly, 14 out of the 15 resistant strains show increased SOS activation (Figure 1A), which strongly correlates with the cost of resistance, explaining 73% of its variation (Figure 1B). Thus DSBs can contribute to the cost of AR. To independently confirm the occurrence of DSBs in resistant bacteria, we used a system which permits direct visualization of double-stranded DNA ends[20], combined with the SOS reporter. This allowed us to confirm that resistant mutants indeed show increased DSBs (Extended Data Figure 1, Extended Data Table 1).

Erythromycin targets the 50S ribosomal subunit, affecting translation and its coupling with transcription[21]. Erythromycin resistance mutations (Erm^R) mapping in the genes *rplD* and *rplV* (encoding L4 and L22, respectively) are known to reduce translation elongation rate[22], credibly affecting transcription-translation coupling as well. We isolated Erm^R clones carrying two of these mutations, $RplD^{G66R}$ and $RplV^{\Delta 82\text{-}84}$, and found that both mutants show increased SOS (Figure 1C), demonstrating that mechanistically different perturbations of transcription-translation coupling cause DSBs as well.

We then reasoned that, if DSBs are a major driver of the cost of resistance, the process of compensatory evolution should lead to lower DSBs. To query if this happens, we first measured the cost and SOS induction in the $RpsL^{K43T}$ $RpoB^{H526Y}$ double mutant and in an isogenic strain additionally carrying the most prevalent compensatory mutation found previously: $RpoC^{Q1126K}$ [14]. Supporting our hypothesis, the cost and SOS induction are greatly reduced in the compensated strain (Figure 2A). Further support was given by analysing 9 compensated clones after the propagation of 3 independent populations of $RpsL^{K43T}$ $RpoB^{H526Y}$ double mutants during 15 days, in the absence of antibiotics. All the compensated clones show decreased cost and SOS induction compared to their resistant ancestrals (Figure 2B), demonstrating that compensatory evolution widely targets DSBs. These results thus reinforce the notion that DSBs are an important contributor to the cost of antibiotic resistance mutations.

The cost of resistance is enviroment dependent[2], and specific Rif^R and Str^R mutations show reduced cost in minimal medium supplemented with glucose[12, 23]. We then reasoned that, if DSBs are an important cause of the cost, they should be reduced in an environment where the costs are smaller. In agreement with our hypothesis, resistant mutants show reduced DSBs in minimal medium (Extended Data Figure 2A), and a weaker correlation with the cost (Extended Data Figure 2B), further validating the strong connection between DSBs and cost.

Generation of DSBs by transcription-translation uncoupling has been shown to encompass increased formation of R-loops[18]. We thus hypothesize that deleting the RNase HI, which specifically degrades R-loops[4], would cause an increase in DSBs in strains carrying Rif^R and Str^R mutations. In accordance with our hypothesis, both DSBs and the cost of resistance are greatly exacerbated in the *ΔrnhA* background (Figure 3A, Extended Data Figure 3, Extended Data Table 1, Extended Data Figure 4). Conversely, overproduction of RNase HI ameliorates both phenotypes in a subset of mutants (Extended Data Figure 5); however, strong overproduction is toxic for the cell[24], irrespectively of its genotype (Extended Data Figure 6). These results confirm the involvement of R-loops in the uncoupling-mediated production of DSBs by resistance and underline the importance of RNase HI function for its cost.

We then queried if targeting RNase HI function could be used as a strategy to specifically select against resistant bacteria in polymorphic communities with high frequency of resistance. RNase HI inhibitors are currently studied as antiretrovirals[25]. We tested the effect of a commercially available one,

RHI001, shown to inhibit the activity of the RNase HI protein of *E. coli in vitro*[5], in competitions between sensitive and resistant bacteria. We observed that 500 µM of RHI001 increase the cost of most resistant mutants (Figure 3B, Extended Data Figure 7A). Chemical inhibition was, however, not as effective as genetic removal (Extended Data Figure 7B), as may be expected, since stability, diffusibility across the bacterial envelope, pharmacokinetics and pharmacodynamics of RHI001 *in vivo* are unknown, and potentially suboptimal. Nevertheless, these results suggest that inhibiting RNase HI may be a novel plausible strategy to specifically select against resistant strains coexisting with sensitive bacteria, as long as resistant strains fail to evolve adaptations that abrogate their extinction. In order to test this hypothesis, we propagated a mixture of CFP-labelled sensitive bacteria competing against a pool of five YFP-labelled single resistant mutants ($RpsL^{K43N}$, $RpsL^{K43T}$, $RpoB^{H526L}$, $RpoB^{H526Y}$, and $RpoB^{S531F}$) during 15 days, in the absence of antibiotics. We studied the frequency dynamics of resistant clones under both strong bottlenecks (1:1500), where new adaptive mutations are less likely to spread, and weak bottlenecks (1:50), where propagation of adapted clones is more probable. In parallel, we performed identical propagations, but in strains lacking RNase HI, as a proxy for conditions under optimal inhibition of RNase HI function. We observed that, in the presence of RNase HI, sensitive bacteria initially outcompete resistant clones but, as the propagation progresses, resistant bacteria increase in frequency (likely due to compensation[2]), reaching coexistence (Figure 4, blue lines). Remarkably, in the propagations of strains lacking RNase HI, resistant bacteria were completely outcompeted and disappeared by day 9 (Figure 4, red lines), even under mild bottlenecks (Extended Data Figure 8). Altogether, these results show that targeting RNase HI is a novel promising strategy to selectively eliminate resistant bacteria.

We found that Rif^R and Str^R mutations cause DSBs, which explain 73% of the variation in their cost (Figure 1, Extended Data Figure 1, Extended Data Table 1). Actually, the mutants showing the highest cost and DSBs ($RpsL^{K43N}$ $RpoB^{H526Y}$ and $RpsL^{K43T}$ $RpoB^{H526Y}$) combine alleles causing increased transcription elongation rate[26] and decreased translation rate[27], arguably resulting in maximized uncoupling. Coherently, mutations causing decreased translation elongation rate by affecting a different ribosomal subunit generate DSBs as well (Figure 1C).

Uncoupling mediated by resistance mutations generates R-loops (Figure 3, Extended Data Figure 3, Extended Data Table 1, Extended Data Figure 4, Extended Data Figure 5), as uncoupling caused by chemical inhibition of translation[28, 18, 29], and lack of RNase HI exacerbates DSBs in single resistant mutants (Extended Data Figure 4)[30]. Interestingly, single resistance mutations known to have high cost[31] show great DSBs even in the presence of RNase HI (Extended Data Figure 9A), suggesting that extensive uncoupling render it insufficent.

Uncoupling-mediated DSBs can involve increased replication-transcription conflicts[18], which cause both R-loops[32] and DSBs[33]. We observed greater DSBs, and correlation with cost, in rich than in minimal media (Figure 1, Extended Data Figure 2). This might be linked to the fact that replication-transcription conflicts are maximized when cells replicate fast[34] and reduced in minimal medium (Extended Data Figure 9B and C), where bacteria replicate at a slower rate[35]. Another observation supporting the occurrence of replication-transcription conflicts is that, although the cost of Rif^R and

Str^R mutations in rich medium are very similar at 4h and 24h (Extended Data Figure 9D), it is generated in the first 4 hours (comprising exponential growth), while resistant bacteria show advantage between 4 and 24 hours (Extended Data Figure 9E), when growth diminishes. Coherently, Rif^R and/or Nal^R (resistant to nalidixic acid) mutants outcompete sensitive bacteria in aging colonies[36, 37]. Remarkably, both lack of RNase HI and conditions of SOS activation can induce initiation of DNA replication from sites different than *OriC* (constitutive stable DNA replication, cSDR)[38]. Thus, it is plausible that Rif^R and Str^R mutants induce cSDR, favouring replication-transcription conflicts, and causing a feed-forward loop of synergistically deleterious effects, further enhanced by the downregulation of *rnhA* caused by induction of the SOS regulon[39].

We revealed RNase HI as a novel promising target specific against resistant bacteria, showing that its absence causes a very low fitness of resistant bacteria (Figure 3, Extended Data Figure 3, Extended Data Table 1, Extended Data Figure 4), and its chemical inhibition increases the cost of resistance (Extended Data Figure 7). Importantly, long-term propagation experiments demonstrated that removal of RNase HI causes resistant clones to be readily outcompeted to extinction by sensitive bacteria, impeding compensatory evolution (Figure 4, Extended Data Figure 8). This is specially important for pathogens that acquire AR exclusively through mutation, such as *Mycobacterium tuberculosis*, which often carries Rif^R and Str^R mutations[40]. Interestingly, the RNase HI function is essential in its close relative *M. smegmatis*, which led to propose RNase HI as a new antimycobacterial target[41].

A better understanding of the mechanisms affecting AR dissemination is urgently needed[6]. Here we showed that DNA breaks are important determinants of the fitness cost of AR mutations and revealed the RNase HI as a novel promising target for antimicrobial therapy specific against resistant bacteria. Our results underline the importance of determining how the bacterial physiology is affected by antibiotics and resistances, and exemplify how fundamental principles can provide novel strategies to face the global challenge of AR.

METHODS

Bacterial strains, media and growth conditions

All the strains used in this study (Extended Data Table 2) are derivatives of *E. coli* K12 MG1655 (from strains RB266, RB323 or RB324). Fluorescently labelled strains harbour a copy of either YFP or CFP under the control of the lac-regulated promoter P_{LacO-1} inserted either in the *yzgL* pseudogene locus or in the *galK* gene, a deletion comprising the entire *lac* operon (to make constitutive the expression of the fluorescent proteins), and the SOS reporter construction P_{sulA}-*mCherry* inserted in the *ysaCD* pseudogene locus. The SOS reporter fusion was constructed by replacing a *tetA-sacB* selectable/counterselectable marker[42] located upstream from a *mCherry-FRT-aph-FRT* cassette previously inserted in the *ysaCD* locus by the regulatory regions (150bp upstream from the translation initiation site) of *sulA*. Resistant mutants additionally carry different chromosomal alleles conferring antibiotic resistance/s. The fluorescent constructions were generated by Lambda-Red recombineering[43],

208 followed by transference to a clean backgrounds by P1 transduction[44], and subsequent transduction of
209 the resistance alleles. The clean deletion of *rnhA* was constructed by markerless recombineering using
210 a *tetR-*P_{tet}*-ccdB-cat* selection/counterselection cassette as described in Figueroa-Bossi & Bossi[45], and
211 subsequent transfer of the markerless deletion to a clean background by P1 transduction, using as
212 recipient an isogenic strain carrying a deletion of the nearby gene *proB*, which causes proline
213 auxotrophy, and selecting for growth in minimal medium. The presence of each construction/mutation
214 was assessed by PCR-mediated amplification of the corresponding region and sequencing. The strain
215 carrying the Gam-GFP construction (SMR14334[20]) was generously contributed by Professor Susan M.
216 Rosenberg. The strain with the additional copy of the *rnhA* gene inducible by arabinose (RCE442[24])
217 was kindly donated by Dr. Christian J. Rudolph. Derivatives of these strains carrying the P_{sulA}*-mCherry*
218 and different resistance mutations were constructed by P1 trasduction. The plasmid pRB-5 (carrying
219 the *rnhA* gene under the control of a promoter inducible by anhydrotetracycline) was constructed by
220 PCR amplification of the vector pZS*11[46] and the construction *tetR-*$P_{LTetO-1}$*-rnhA* from strain RB1207,
221 subsequent restriction with AatII and HindIII, ligation and electroporation. The construction *tetR-*$P_{LTetO-1}$*-rnhA* in strain RB1207 was made by Lambda-Red recombineering[43], selection/counterselection[45], and
223 a previous $P_{LtetO-1}$*-sfGFP* construction[47]. Cultures were grown in either Lysogeny Broth (LB, Miller
224 formulation)[48] or M9 broth supplemented with 0.4% glucose[49], in 96-well plates incubated at 37°C with
225 shaking (700 r.p.m.) in a Grant-bio PHMP-4 benchtop incubator. In the experiments including the
226 RNase HI inhibitor 2-[[[3-Bromo-5-(2-furanyl)-7-(trifluoromethyl)pyrazolo[1,5-a]pyrimidin-2-
227 yl]carbonyl]amino]-4,5,6,7-tetrahydro-benzo[b]thiophene-3-carboxylic acid ethyl ester (RHI001), the
228 medium was supplemented with 500 µM of RHI001 (Glixx Laboratories Inc., catalog number GLXC-

03982). Solid media was LB containing 1.5% agar, supplemented when necessary with antibiotics at the following concentrations: rifampicin (100 µg/ml), streptomycin (100 µg/ml), ampicillin (100 µg/ml), erythromycin (150 µg/ml), kanamycin (100 µg/ml), chloramphenicol (25 µg/ml).

Competitive fitness/SOS induction assays

The relative fitness (selection coefficient per generation) of each YFP-tagged resistant strain was measured by competitive growth against an isogenic sensitive strain *E. coli* K12 MG1655 constitutively expressing CFP and carrying the SOS reporter. The formula used to calculate the selection coefficient was $\mathbf{s = [\ln(NR_f/Nr_f)-\ln(NR_i/NR_i)]/\ln(NS_f/NS_i)}$, being $\mathbf{NR_i}$ and $\mathbf{NR_f}$ the initial and final number of resistant bacteria, and $\mathbf{NS_i}$ and $\mathbf{NS_f}$ the initial and final number of sensitive bacteria.The competitor strains were first streaked out of their respective frozen vials, then individual colonies were inoculated separately in medium without antibiotics and incubated overnight (approximately 16h); the next morning, the number of cells in each culture was measured by Flow Cytometry, and 10 µl of 1:1 mixtures of YFP and CFP bacteria were added to 140 µl of medium, at an initial number of approximately 10^6 cells. The initial and final frequencies of the strains were obtained by counting their cell numbers in the Flow Cytometer. Generation time was estimated from the doubling time of the reference strain (approximately five generations at 4h, and approximately eight generations at 24h), and the fitness was determined as the average of the independent replicates for each competition. The proportion of SOS-induced bacteria was quantified as the number of either YFP-tagged or CFP-labelled bacteria showing red fluorescence (from the P_{sulA}-*mCherry* SOS reporter fusion) above a

threshold determined by the fluorescence levels of the control strains *lexA (ind-)* (constitutive repression of the SOS response) and *ΔlexA* (constitutive activation of the SOS response)[50]; an illustrative slide is in Supplementary Information. The data represented is the induction level of each mutant normalized with respect to the induction levels of the sensitive it is competing against.

Flow Cytometry

A BD LSR Fortessa™ SORP flow cytometer was used to quantify bacteria, using a 96-well plate High-Throughput Sampler (HTS) and SPHERO fluorescent spheres (AccuCount 2.0 µm blank particles), in order to accurately measure volumes. Bacterial numbers were calculated based on the counts of fluorescently labelled bacteria with respect the known number of beads added to a given volume. The instrument was equipped with a 488 nm laser used for scatter parameters and YFP detection, a 442 nm laser for CFP detection and a 561 nm laser for mCherry detection. Relative to optical configuration, CFP, YPF and mCherry were measured using bandpass filters in the range of 470/20 nm, 540/30 nm and 630/75nm, respectively. The analyser is also equipped with a forward scatter (FSC) detector in a photomultiplier tube (PMT) to detect bacteria. The samples were acquired using FACSDiVa (version 6.2) software, and analysed using FlowJo (version X 10.0.7r2). All Flow Cytometry experiments were performed at the Flow Cytometry Facility of Instituto Gulbenkian de Ciência, Oeiras, Portugal.

Selection for Erm^R bacteria

Fifteen independent colonies of sensitive bacteria were separately inoculated in LB in a 96-well plate, incubated at at 37°C with shaking (700 r.p.m) for 7 hours, and 0.1ml of either independent culture was plated onto a LB agar plate supplemented with 150 µg/ml erythromycin, and incubated at 37°C for 5 days (Erm^R strains grow in the presence of erythromycin, albeit slowly). Colonies able to grow in these plates were streaked onto plates supplemented with 150 µg/ml erythromycin, in order to further assess their *bona fide* resistance, and the *rplD*, and *rplV* genes of the resistant clones were amplified by PCR and sequenced.

Microscopy

Early exponential cultures were diluted into medium containing the inducer of the Gam-GFP construction (anhydrotetracycline, 25ng/ml) and incubated at 37°C with shaking (240 r.p.m.) for 3h, prior to imaging. Bacterial solutions were then placed onto 1% agarose (in 1X PBS) pads mounted in adhesive frames between the microscope slide and a coverglass. Images were acquired on an Applied Precision DeltavisionCORE system, mounted on an Olympus IX71 inverted microscope, coupled to a Cascade II 1024x1024 EM-CCD camera, using an Olympus 100x 1.4NA Uplan SAPO Oil immersion objective, where GFP and mCherry were imaged with FITC (Ex: 475/28, EM: 528/38) and TRITC (Ex: 542/28, Em: 617/73) fluorescence filtersets, respectively, and DIC optics. Images were deconvoluted

with Applied Precision's softWorx software, and prepared for presentation (cropping smaller fields to facilitate visualization, and false-coloring green and red fluorescent signals) using Fiji/ImageJ.

Long-term propagations of polymorphic populations

The CFP-tagged sensitive (either WT or *ΔrnhA*) and the five YFP-labelled resistant bacterial (either (RpsL^{K43N}, RpsL^{K43T}, RpoB^{H526L}, RpoB^{H526Y}, and RpoB^{S531F} or *ΔrnhA* RpsL^{K43N}, *ΔrnhA* RpsL^{K43T}, *ΔrnhA* RpoB^{H526L}, *ΔrnhA* RpoB^{H526Y}, and *ΔrnhA* RpoB^{S531F}) were streaked individually onto LB agar plates and incubated overnight at 37°C. The next day, three independent colonies from each strain were inoculated separately in LB broth (150 µl per well) in 96-well plate and incubated overnight at 37°C with shaking (700 r.p.m). The next day, bacteria were quantified by Flow Cytometry, and three 1:1:1:1:1:1 mixtures of the sensitive and either resistant bacteria were added to 140 µl of medium, at a initial number of approximately 10^6 cells (ordered in a checkered pattern to avoid cross-contaminations). The initial frequencies of the fluorescent strains were confirmed by Flow Cytometry. Every 24h, during 15 days, 10 µl of bacteria culture was diluted by a factor of 10^{-2}, then transferred to 140 µl fresh LB and allowed to grow for additional 24h, reaching approximately 10^9 cells/ml. In parallel, cell numbers were counted using the Flow Cytometer, in order to measure the frequency of each strain in the mixed population during the experiment, by collecting a sample (10 µl) from the spent culture each day.

Statistical analyses

All analyses were conducted using either Libreoffice Calc version 5.4.1.2 (www.libreoffice.org) software, R version 3.4.4 (www.r-project.org) software via RStudio version 1.1.442 interface (www.rstudio.com) or GraphPad Prism version 7.04 (www.graphpad.com). For each set of competitions ($n \geq 3$), two-tailed unpaired Student's *t*-tests were used. For testing the association between the fitness cost and the level of SOS induction, a linear regression of the s with the logarithm of the fold change in SOS induction of each resistant clone (with respect to the sensitive bacteria it is competing against in the same well), was used.

Data availability

The raw images used for Extended Data Table 1, Extended Data Figure 1, and Extended Data Figure 3 are available in Zenodo (DOI: 10.5281/zenodo.3381746).

REFERENCES

1. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
2. Durão, P., Balbontín, R. & Gordo, I. Evolutionary Mechanisms Shaping the Maintenance of Antibiotic Resistance. *Trends Microbiol.* **26**, 677–691 (2018).
3. Vogwill, T. & MacLean, R. C. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evol. Appl.* **8**, 284–295 (2015).

4. Tadokoro, T. & Kanaya, S. Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J.* **276**, 1482–1493 (2009).
5. Kim, J. *et al.* Identification of two HIV inhibitors that also inhibit human RNaseH2. *Mol. Cells* **36**, 212–218 (2013).
6. Bush, K. *et al.* Tackling antibiotic resistance. *Nat. Rev. Microbiol.* **9**, 894–896 (2011).
7. Jin, D. J. & Gross, C. A. Characterization of the pleiotropic phenotypes of rifampin-resistant *rpoB* mutants of *Escherichia coli*. *J. Bacteriol.* **171**, 5229–5231 (1989).
8. Jin, D. J. & Zhou, Y. N. Mutational analysis of structure-function relationship of RNA polymerase in *Escherichia coli*. *Methods Enzymol.* **273**, 300–319 (1996).
9. Ozaki, M., Mizushima, S. & Nomura, M. Identification and functional characterization of the protein controlled by the streptomycin-resistant locus in E. coli. *Nature* **222**, 333–339 (1969).
10. Qi, Q., Preston, G. M. & MacLean, R. C. Linking system-wide impacts of RNA polymerase mutations to the fitness cost of rifampin resistance in *Pseudomonas aeruginosa*. *mBio* **5**, e01562 (2014).
11. Zhou, Y. N. & Jin, D. J. The *rpoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like 'stringent' RNA polymerases in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2908–2913 (1998).
12. Paulander, W., Maisnier-Patin, S. & Andersson, D. I. The fitness cost of streptomycin resistance depends on *rpsL* mutation, carbon source and RpoS (sigmaS). *Genetics* **183**, 539–546, 1SI-2SI (2009).
13. Pelchovich, G. *et al.* Ribosomal mutations affecting the translation of genes that use non-optimal codons. *FEBS J.* **281**, 3701–3718 (2014).
14. Moura de Sousa, J., Balbontín, R., Durão, P. & Gordo, I. Multidrug-resistant bacteria compensate for the epistasis between resistances. *PLoS Biol.* **15**, e2001741 (2017).

15. Burmann, B. M. *et al.* A NusE:NusG complex links transcription and translation. *Science* **328**, 501–504 (2010).
16. Herbert, K. M. *et al.* E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399**, 17–30 (2010).
17. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010).
18. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146**, 533–543 (2011).
19. Quillardet, P., Huisman, O., D'Ari, R. & Hofnung, M. SOS chromotest, a direct assay of induction of an SOS function in *Escherichia coli* K-12 to measure genotoxicity. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 5971–5975 (1982).
20. Shee, C. *et al.* Engineered proteins detect spontaneous DNA breakage in human and bacterial cells. *eLife* **2**, e01222 (2013).
21. Sedlyarova, N. *et al.* Natural RNA Polymerase Aptamers Regulate Transcription in *E. coli*. *Mol. Cell* **67**, 30-43.e6 (2017).
22. Zaman, S., Fitzpatrick, M., Lindahl, L. & Zengel, J. Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in *Escherichia coli*. *Mol. Microbiol.* **66**, 1039–1050 (2007).
23. Trindade, S., Sousa, A. & Gordo, I. Antibiotic resistance and stress in the light of Fisher's model. *Evol. Int. J. Org. Evol.* **66**, 3815–3824 (2012).
24. Stockum, A., Lloyd, R. G. & Rudolph, C. J. On the viability of *Escherichia coli* cells lacking DNA topoisomerase I. *BMC Microbiol.* **12**, 26 (2012).

25. Wang, X., Gao, P., Menendez-Arias, L., Liu, X. & Zhan, P. Update on Recent Developments in Small Molecular HIV-1 RNase H Inhibitors (2013-2016): Opportunities and Challenges. *Curr. Med. Chem.* **25**, 1682–1702 (2018).
26. Fisher, R. F. & Yanofsky, C. Mutations of the beta subunit of RNA polymerase alter both transcription pausing and transcription termination in the trp operon leader region in vitro. *J. Biol. Chem.* **258**, 8146–8150 (1983).
27. Schrag, S. J. & Perrot, V. Reducing antibiotic resistance. *Nature* **381**, 120–121 (1996).
28. Broccoli, S. *et al.* Effects of RNA polymerase modifications on transcription-induced negative supercoiling and associated R-loop formation. *Mol. Microbiol.* **52**, 1769–1779 (2004).
29. Negro, V. *et al.* RadD Contributes to R-Loop Avoidance in Sub-MIC Tobramycin. *mBio* **10**, (2019).
30. Kogoma, T. *Escherichia coli* RNA polymerase mutants that enhance or diminish the SOS response constitutively expressed in the absence of RNase HI activity. *J. Bacteriol.* **176**, 1521–1523 (1994).
31. Trindade, S. *et al.* Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet.* **5**, e1000578 (2009).
32. Lang, K. S. *et al.* Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell* **170**, 787-799.e18 (2017).
33. Merrikh, H., Zhang, Y., Grossman, A. D. & Wang, J. D. Replication-transcription conflicts in bacteria. *Nat. Rev. Microbiol.* **10**, 449–458 (2012).
34. Merrikh, H., Machón, C., Grainger, W. H., Grossman, A. D. & Soultanas, P. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**, 554–557 (2011).
35. Wang, C. H. & Koch, A. L. Constancy of growth on simple and complex media. *J. Bacteriol.* **136**, 969–975 (1978).

36. Wrande, M., Roth, J. R. & Hughes, D. Accumulation of mutants in ‘aging’ bacterial colonies is due to growth under selection, not stress-induced mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11863–11868 (2008).
37. Katz, S. & Hershberg, R. Elevated mutagenesis does not explain the increased frequency of antibiotic resistant mutants in starved aging colonies. *PLoS Genet.* **9**, e1003968 (2013).
38. Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev. MMBR* **61**, 212–238 (1997).
39. Quiñones, A., Kücherer, C., Piechocki, R. & Messer, W. Reduced transcription of the *rnh* gene in *Escherichia coli mutants* expressing the SOS regulon constitutively. *Mol. Gen. Genet. MGG* **206**, 95–100 (1987).
40. Almeida Da Silva, P. E. A. & Palomino, J. C. Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. *J. Antimicrob. Chemother.* **66**, 1417–1430 (2011).
41. Minias, A. E. *et al.* RNase HI Is Essential for Survival of *Mycobacterium smegmatis*. *PloS One* **10**, e0126260 (2015).
42. Li, X.-T., Thomason, L. C., Sawitzke, J. A., Costantino, N. & Court, D. L. Positive and negative selection using the *tetA-sacB* cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res.* **41**, e204 (2013).
43. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).
44. Lennox, E. S. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1**, 190–206 (1955).

45. Figueroa-Bossi, N. & Bossi, L. Recombineering applications for the mutational analysis of bacterial RNA-binding proteins and their sites of action. *Methods Mol. Biol. Clifton NJ* **1259**, 103–116 (2015).
46. Subramaniam, A. R., Pan, T. & Cluzel, P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2419–2424 (2013).
47. Balbontín, R., Vlamakis, H. & Kolter, R. Mutualistic interaction between *Salmonella enterica* and *Aspergillus niger* and its effects on *Zea mays* colonization. *Microb. Biotechnol.* **7**, 589–600 (2014).
48. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.* **62**, 293–300 (1951).
49. *Maniatis et al. Molecular cloning: a laboratory manual (CSHL Press)*.
50. Lin, L. L. & Little, J. W. Isolation and characterization of noncleavable (Ind-) mutants of the LexA repressor of *Escherichia coli* K-12. *J. Bacteriol.* **170**, 2163–2173 (1988).

AUTHORS CONTRIBUTIONS

R.B. and I.G. conceived the study and designed the experiments, R.B. constructed the strains and performed the experiments, R.B. and I.G. analysed and interpeted the data, R.B. created the figures and wrote the manuscript with input from I.G., I.G. provided resources, funding and supervision.

ACKNOWLEDGEMENTS

We thank Professor Susan M. Rosenberg and Dr. Christian J. Rudolph for kindly providing bacterial strains, and Rui Gardner, Claudia Bispo, Mariana Fernandes, Marta Monteiro and the Flow Cytometry Facility of Instituto Gulbenkian de Ciência for their services and assistance, and Nuno Pimpão Martins and the IGC's Advanced Imaging Facility (AIF-UIC) team for the microscopy technical assistance. This work was funded by the the Marie Sklodowska-Curie Actions (MSCA) with the fellowship 746690-ResistEpist-H2020-MSCA-IF-2016/H2020-MSCA-IF-2016, to R.B., and partially supported by ONEIDA project (LISBOA-01-0145-FEDER-016417) co-funded by FEEI - "Fundos Europeus Estruturais e de Investimento" from "Programa Operacional Regional Lisboa 2020", and by national funds from FCT - "Fundação para a Ciência e a Tecnologia, to I.G. R.B. was also supported by the Fundação para a Ciência e a Tecnologia (FCT) with the fellowship SFRH/BDP/109517/2015. The authors thank Karina Xavier, Leonardo Gastón Guilgur, Pol Nadal Jiménez, Chris Diwo and the members of the Gordo and Xavier labs for critically reading earlier versions of this manuscript.

FIGURES

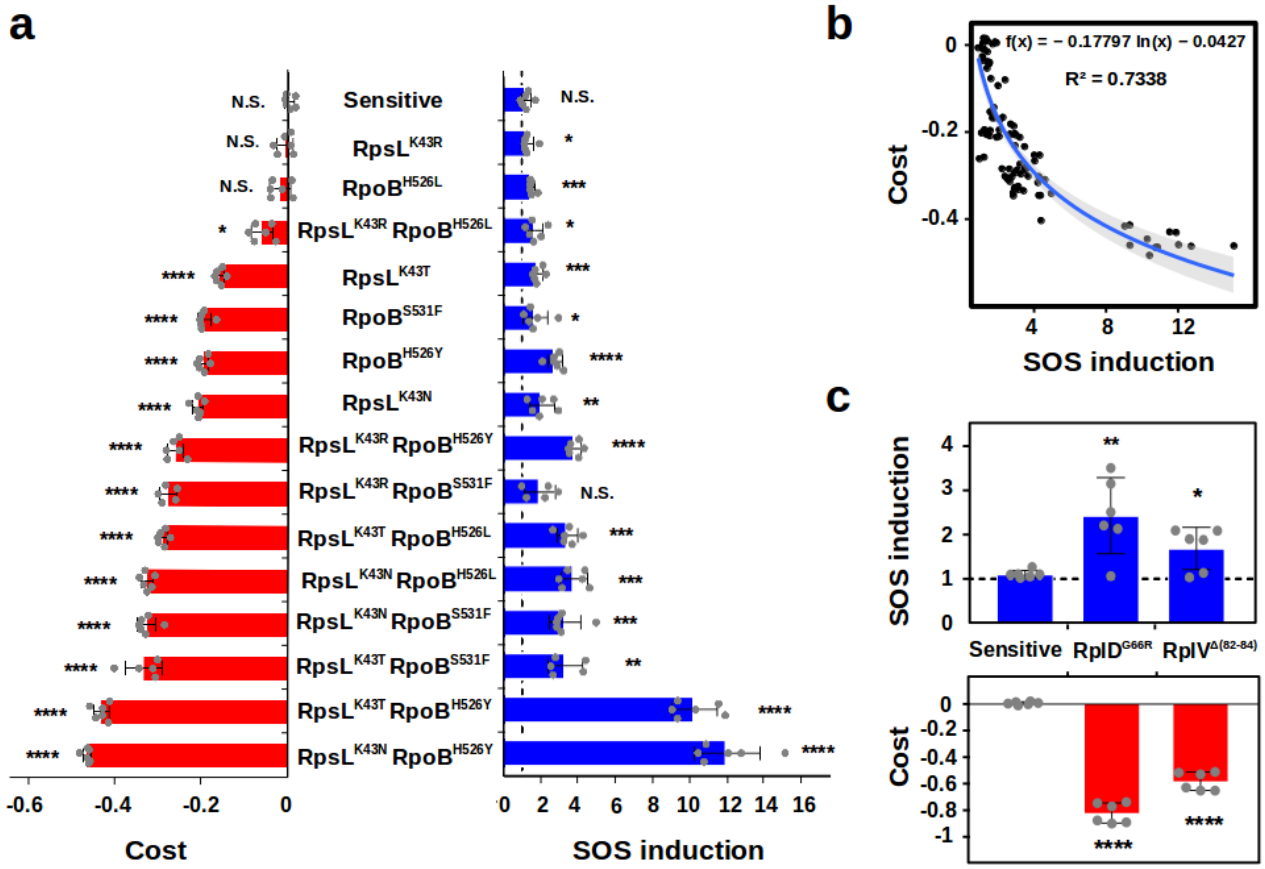


Figure 1. The fitness cost of resistance mutations correlates with the level of induction of the SOS response. a. Selection coefficient per generation, representing the fitness cost (red bars) and fold-change in SOS induction (blue bars) of sensitive bacteria and single and double resistant mutants in LB broth at 4 hours. The strains are ordered from lower to higher fitness cost (top to bottom). The dashed line indicates no SOS induction. Error bars represent mean ± standard deviation (n≥5). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test). **b.** Correlation between the fitness cost (y axis) and the SOS induction (x axis) representing all the data points from (A). The blue line represents the logarithmic regression line, and the grey area represents the 95% confidence interval. **c.** Erm^R mutants show induction of the SOS response. Fold change in SOS induction (blue bars) and selection coefficient per generation, representing the fitness cost (red bars) of sensitive bacteria and Erm^R resistant mutants in LB broth at 4 hours. The dashed line indicates no SOS induction. Error bars represent mean ± standard deviation of independent biological replicates (n=6). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test).

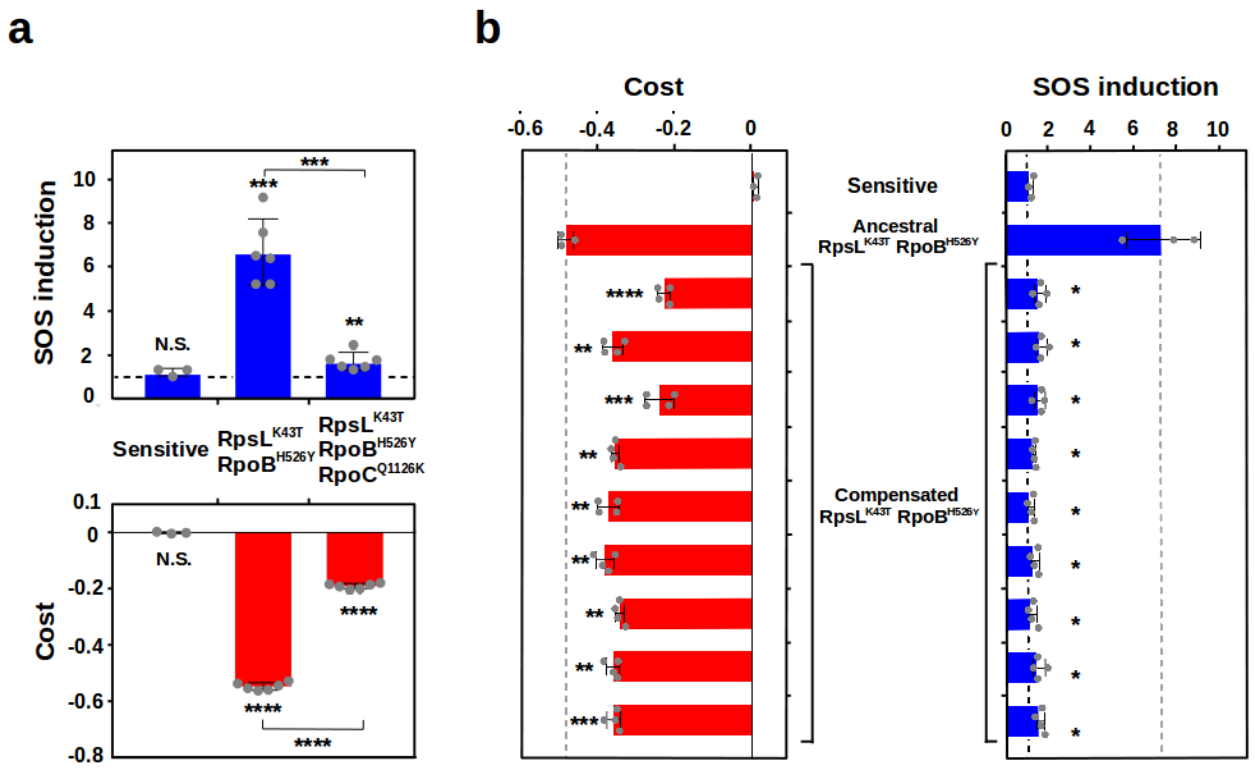


Figure 2. Compensated strains show reduced SOS response. a. Fold-change in SOS induction (blue bars) and selection coefficient per generation, representing the fitness cost (red bars), of sensitive (left) double resistant (center) and double resistant carrying a compensatory mutation (right) bacteria in LB broth at 4 hours. The dashed line indicates no SOS induction. Error bars represent mean ± standard deviation (n≥3). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test). **b.** Selection coefficient per generation, representing the fitness cost (red bars), and fold change in SOS induction (blue bars) of sensitive, ancestral $RpsL^{K43T}$ $RpoB^{H526Y}$ and nine compensated $RpsL^{K43T}$ $RpoB^{H526Y}$ clones in LB broth at 4 hours. The black dashed line indicates no SOS induction. The grey dashed lines mark the cost/SOS of the ancestral double mutant. Error bars represent mean ± standard deviation of independent biological replicates (n≥3). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test).

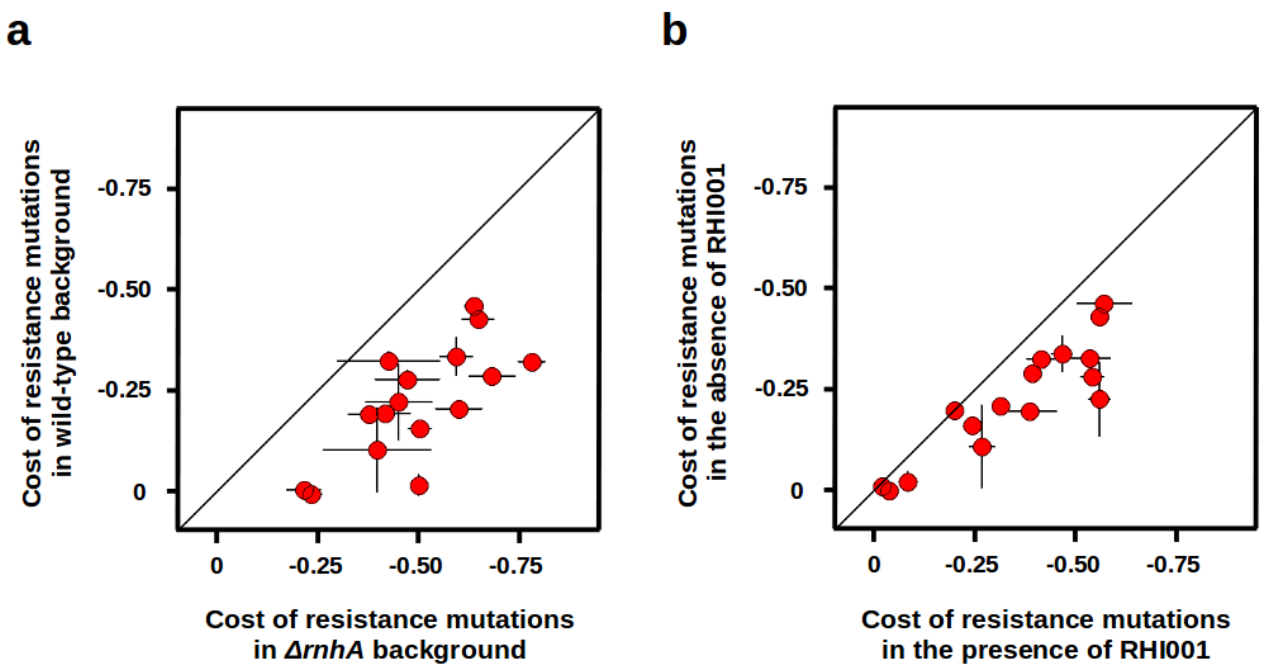


Figure 3. The lack of RNase HI function greatly increases the cost of resistance. a. Correlation between the fitness cost of resistance mutations in wild-type (y axis) or *ΔrnhA* backgrounds (x axis). The black line represents the linear regression if the costs were identical. **B.** Correlation between the fitness cost of resistance mutantions in the absence of the RNase HI inhibitor (y axis) or in its presence (x axis). The black line represents the linear regression if the costs were identical. The values corresponding to the wild-type background in panel **a** and to absence of RHI001 in panel **b** are those shown in Figure 1A. Error bars represent mean ± standard deviation of independent biological replicates (n≥3).

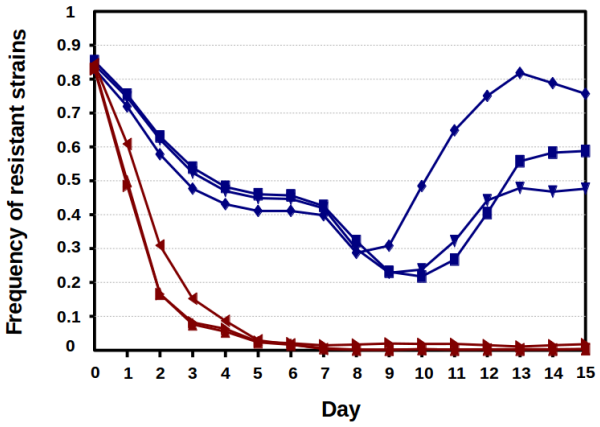


Figure 4. Lack of RNase HI favours outcompetition of resistant mutants by sensitive bacteria. Frequency of single resistant mutants during three independent long-term competitions against sensitive bacteria either in a genetic background including RNase HI (blue lines) or in a *ΔrnhA* background (red lines), imposing a strong bottleneck (1:1500 dilutions).

Extended Data Table 1. Percentage of cells showing Gam-GFP foci in sensitive and resistant bacteria either in wild-type or *ΔrnhA* backgrounds. Except in the *Δdam* positive control (in which over 100 cells sufficed to provide an illustrative example), at least 1000 cells per group were analysed.

***Δdam* (positive control) 21.28%**

Sensitive 0.72%

***rpsL (K43N)* 0.73%**

***rpoB (H526Y)* 0.96%**

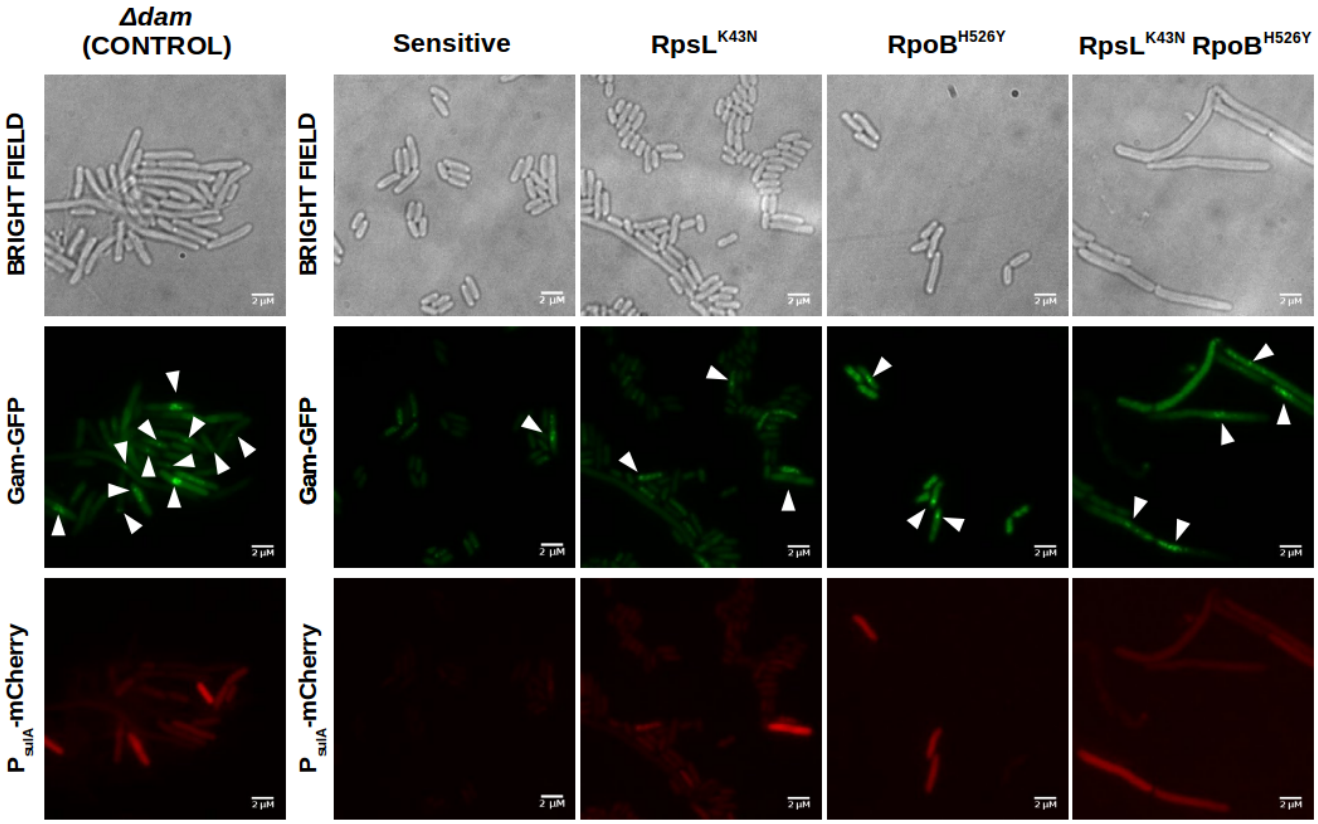
***rpsL (K43N) rpoB (H526Y)* 10.71%**

***ΔrnhA* (sensitive) 9.27%**

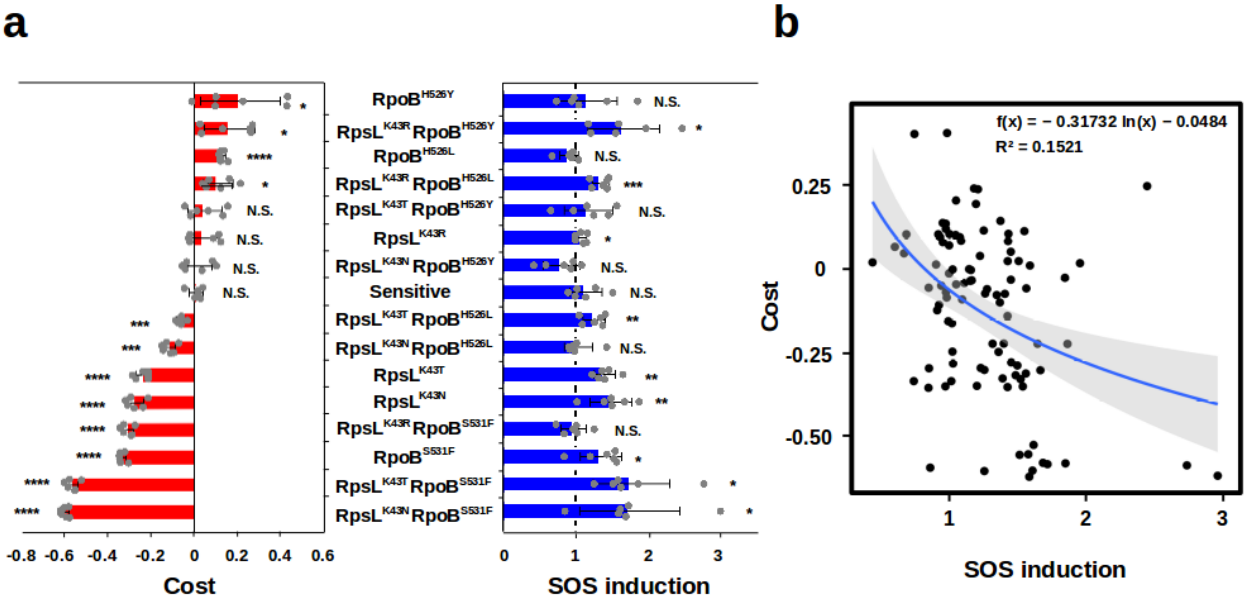
***ΔrnhA rpsL (K43N)* 33.92%**

***ΔrnhA rpoB (H526Y)* 10.02%**

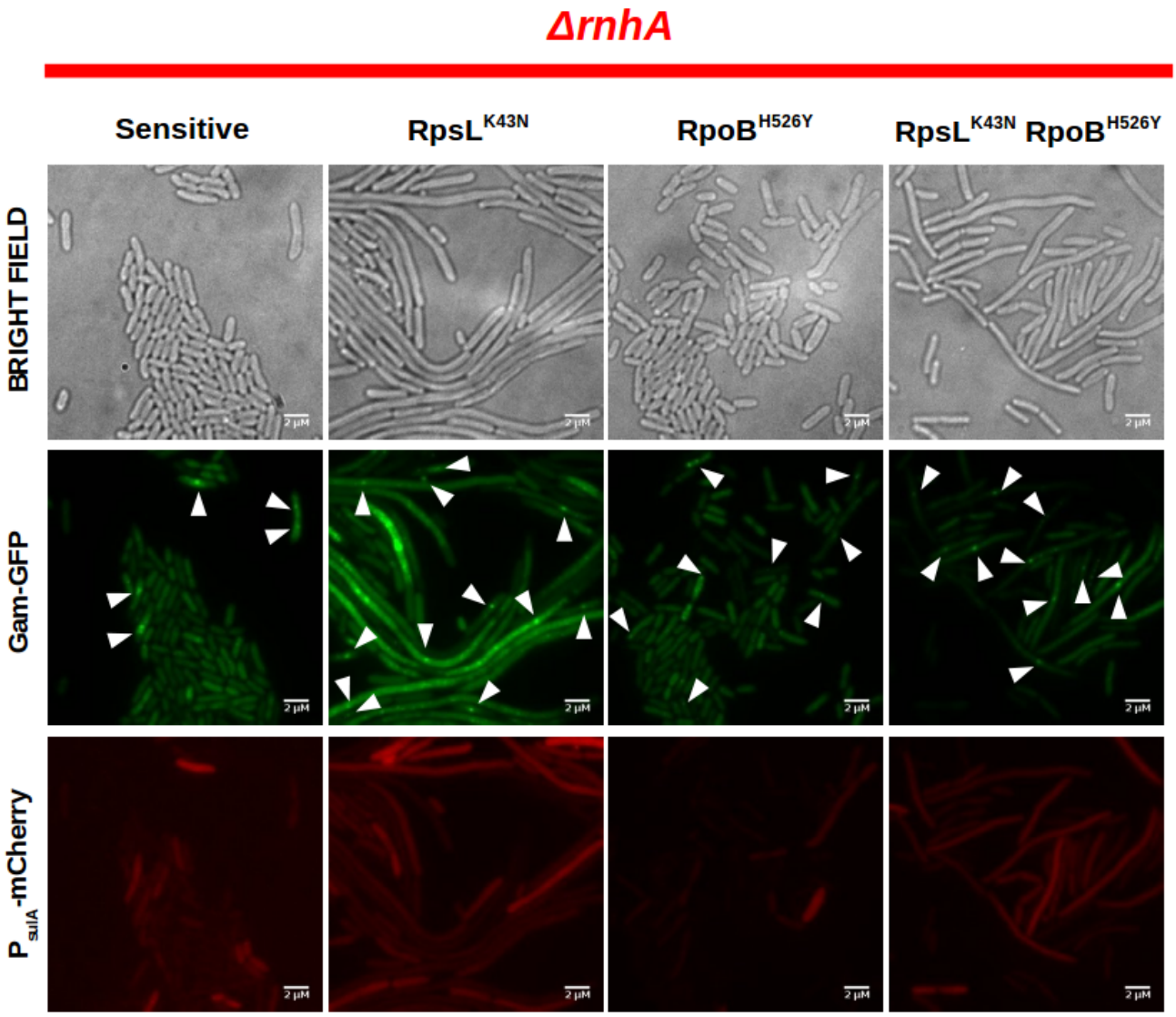
***ΔrnhA rpsL (K43N) rpoB (H526Y)* 49.90%**



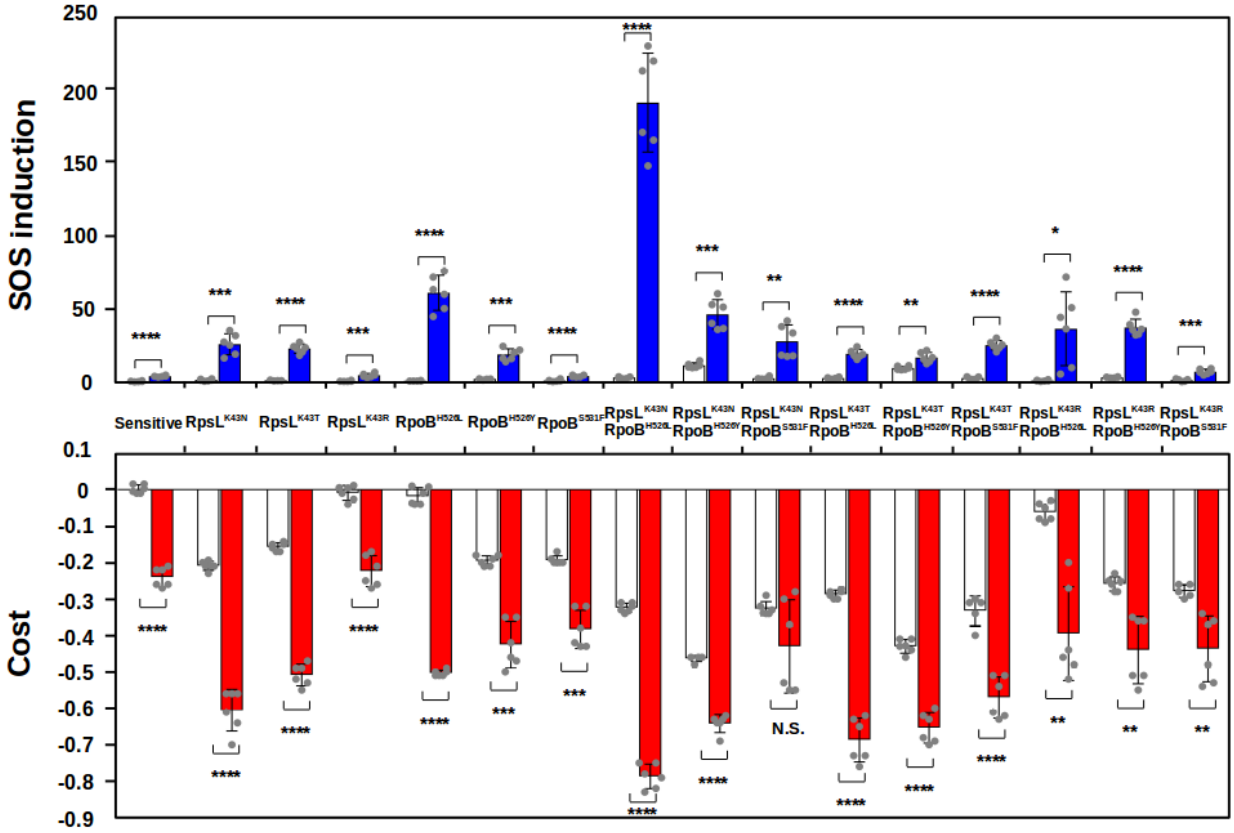
Extended Data Figure 1. Visualization of DNA breaks and SOS induction of sensitive, single and double resistant strains by epifluorescence microscopy. Bacterial cells show a disperse faint green fluorescence distributed along the cytoplasm, unless DSBs are present; upon generation of DBS, the disperse fluorescence concentrates in bright fluorescent foci (central panels, false-colored in green). Red fluorescence (bottom panels, false-colored in red) is absent until the SOS response is activated due to the presence of DSBs. Cell elongation (top panels) is a well-known phenotype derived from the inhibition of cell division by SOS-regulated proteins.



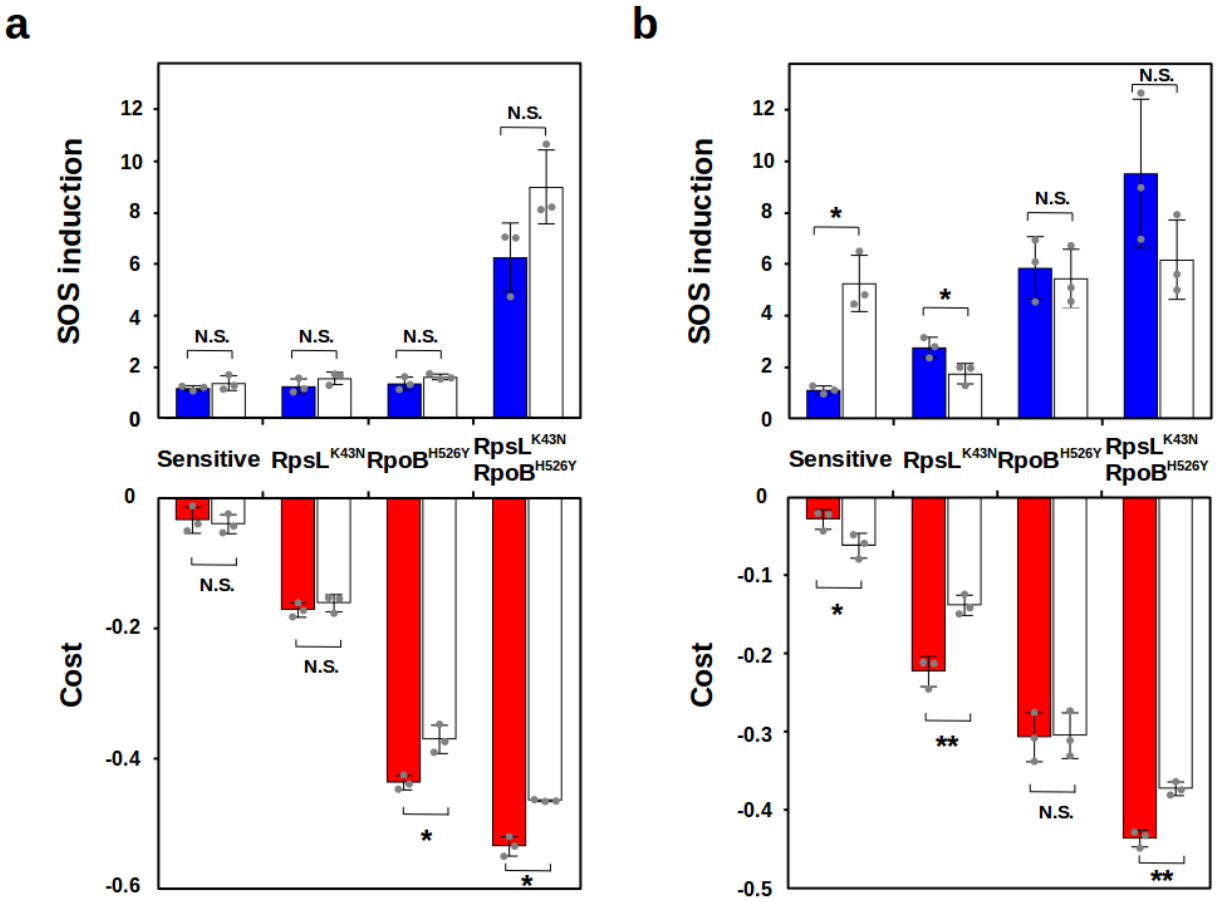
Extended Data Figure 2. The SOS response is reduced in minimal medium supplemented with glucose. **a.** Selection coefficient per generation, representing the fitness cost (red bars), and fold change in SOS induction (blue bars) of sensitive bacteria and single and double resistant mutants in minimal medium at 8 hours. The strains are ordered from lower to higher fitness cost (top to bottom). The dashed line indicates no SOS induction. Error bars represent mean ± standard deviation of independent biological replicates (n=6). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test). **b.** Correlation between the fitness cost (y axis) and the SOS induction (x axis) representing all the data points from panel **a**. The blue line represents the logarithmic regression line, and the grey area represents the 95% confidence interval.



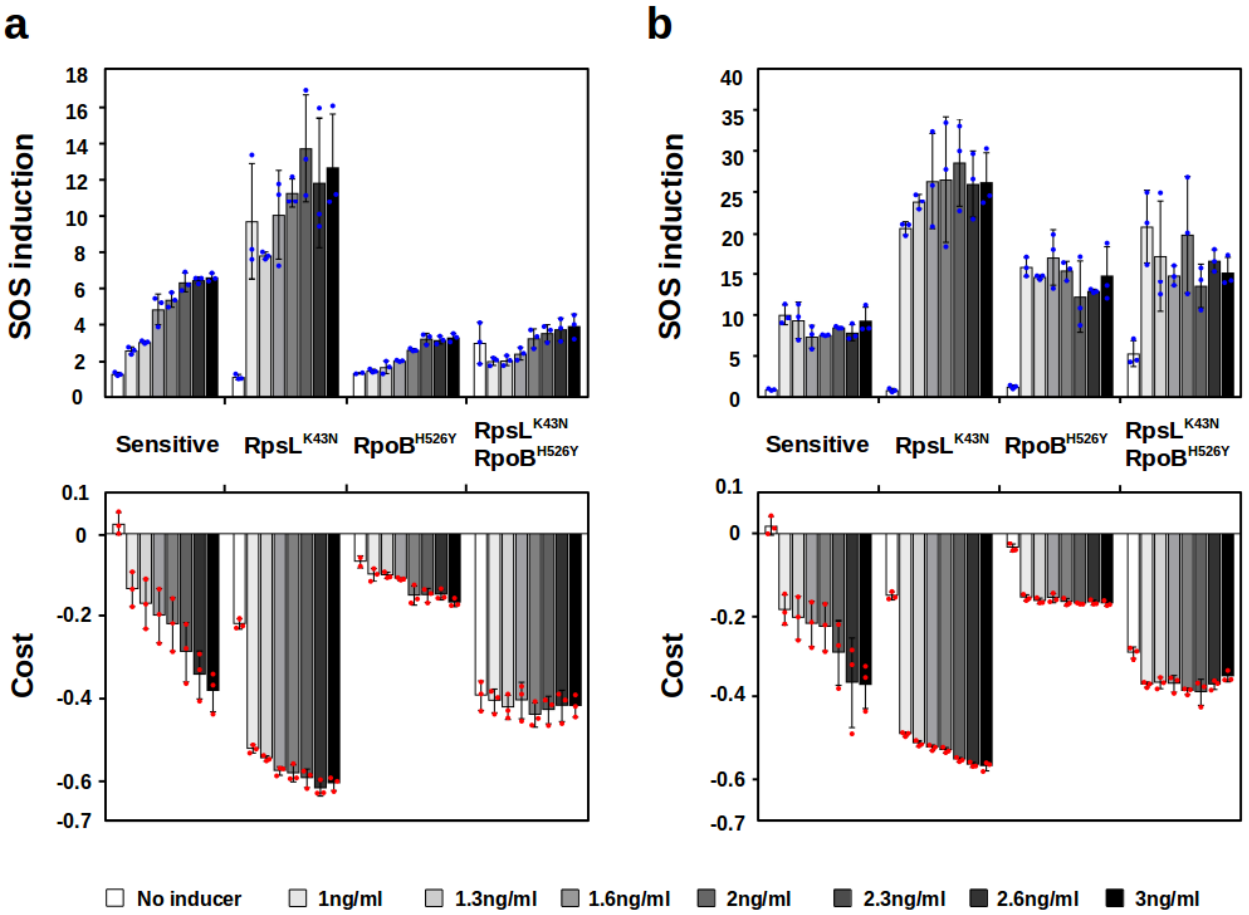
Extended Data Figure 3. Visualization of DNA breaks and SOS induction of sensitive, single and double resistant strains in a *ΔrnhA* background by epifluorescence microscopy. Bacterial cells show a disperse faint green fluorescence distributed along the cytoplasm, unless DSBs are present; upon generation of DSBs, the disperse fluorescence concentrates in bright fluorescent foci (central panels, false-colored in green). Red fluorescence (bottom panels, false-colored in red) is absent until the SOS response is activated due to the presence of DSBs. Cell elongation (top panels) is a well known phenotype derived from cell division inhibition by SOS-regulated proteins.



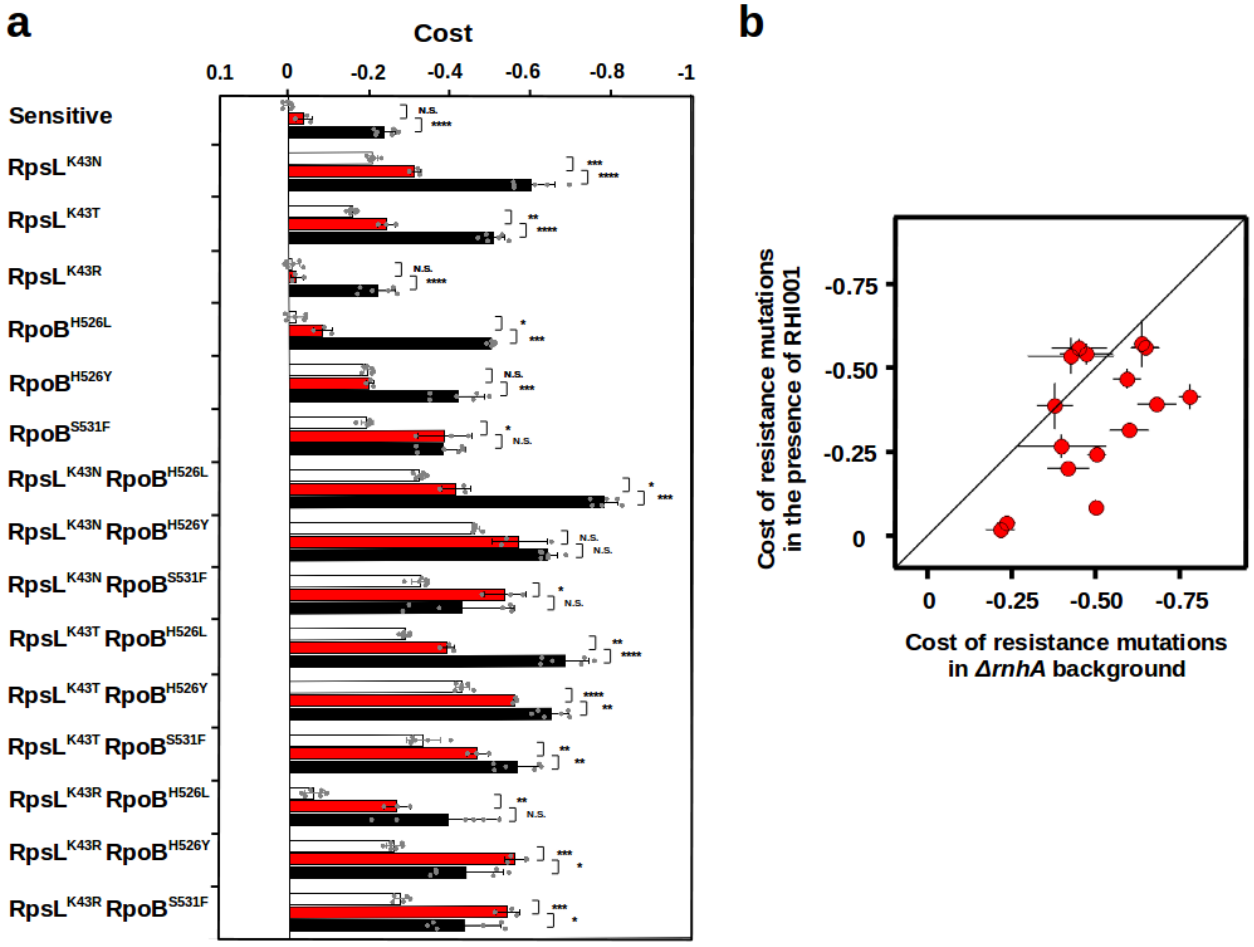
Extended Data Figure 4. Deletion of *rnhA* (encoding RNase HI) greatly increases both the fitness cost and the SOS induction. Fold change in SOS induction (blue bars) and selection coefficient per generation, representing the fitness cost (red bars), of single and double resistant strains in a *ΔrnhA* background. The white bars represent the corresponding values in bacteria with an intact *rnhA* gene (data from the experiments shown in Figure 1), for comparison. Error bars represent mean ± standard deviation of independent biological replicates (n=6). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test).



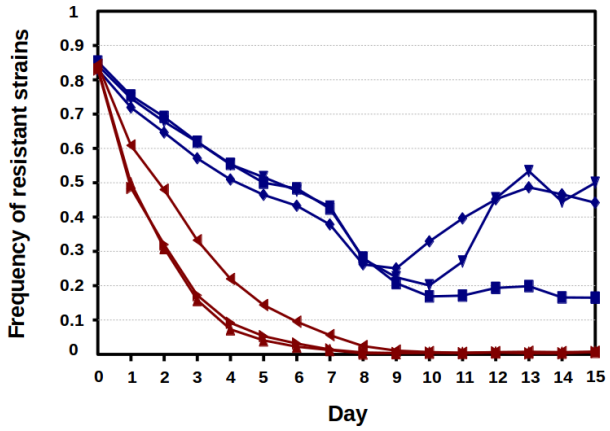
463 **Extended Data Figure 5. Mild overproduction of RNase HI ameliorates both cost and SOS induction.** Fold change in
464 SOS induction (top) and selection coefficient per generation, representing the fitness cost (bottom), of sensitive bacteria
465 (left), two single resistant mutants (center) and the double mutant resulting from combining these two alleles (right) at 4h
466 (**a**) or 24h (**b**) in a background harbouring an additional chromosomal copy of the gene encoding the RNase HI (*rnhA*)
467 under the control of the arabinose promoter, either in the absence (red bars) or the presence (white bars) of arabinose. Error
468 bars represent mean ± standard deviation of independent biological replicates (n=3). N.S. non-significant; $*P < 0.05$; $**P <$
469 0.01; $***P < 0.001$; $****P < 0.0001$ (two-tailed Student's *t* test).



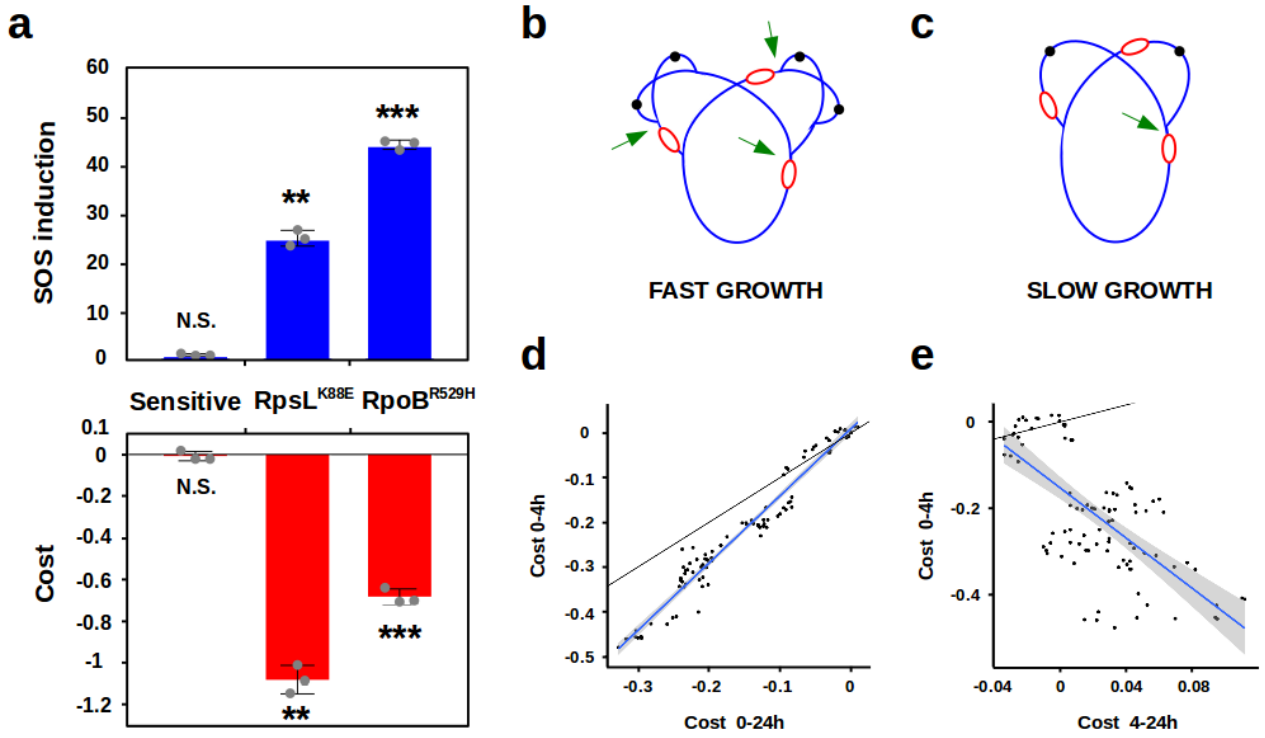
Extended Data Figure 6. Strong overproduction of RNase HI worsen both cost and SOS induction in all the backgrounds. Fold change in SOS induction (top) and selection coefficient per generation, representing the fitness cost (bottom), of sensitive bacteria (left), two single resistant mutants (center) and the double mutant resulting from combining these two alleles (right) at 4h (**a**) or 24h (**b**) in a background harbouring a plasmid carrying a copy of the RNase HI gene (*rnhA*) under the control of the a promoter inducible by anhydrotetracycline, either in the absence (blue/red bars) or the presence of different concentrations of the inducer (grayscale, from white to black bars). The dashed line indicates no cost/SOS induction. Error bars represent mean ± standard deviation of independent biological replicates (n≥2). N.S. non-significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$ (two-tailed Student's *t* test).



Extended Data Figure 7. Chemical inhibition RNase HI cause an increase in the fitness cost of resistant bacteria. a. Selection coefficient, representing the fitness cost of sensitive bacteria and different single and double resistant strains in the presence of the RNase HI inhibitor RHI001 (red bars). The corresponding values in the absence of the inhibitor (data from the experiments shown in Figure 1) or in a *ΔrnhA* background (data from the experiments shown in Figures 3A and Extended Data Figure 4) are represented as white and black bars, respectively, for comparison. The dashed line indicates no cost. Error bars represent mean ± standard deviation (n=3). N.S. non-significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (two-tailed Student's *t* test). **b.** Correlation between the fitness cost of resistance mutantions in the presence of the RNase HI inhibitor (y axis) or in the *ΔrnhA* background (x axis). The black line represents the linear regression if the costs were identical. The values corresponding to the presence of RHI001 and the *ΔrnhA* background are those shown in panel **a** and Extended Data Figure 4, respectively. Error bars represent mean ± standard deviation of independent biological replicates (n≥3).



Extended Data Figure 8. Lack of RNase HI favours outcompetition of resistant mutants by sensitive bacteria. Frequency of single resistant mutants during three independent long-term competitions against sensitive bacteria either in a genetic background including RNase HI (blue lines) or in a *ΔrnhA* background (red lines), imposing a mild bottleneck (1:50 dilutions).



Extended Data Figure 9. a. Costly single mutants show increased SOS induction. Fold change in SOS induction (blue bars) and selection coefficient per generation, representing the fitness cost (red bars), of sensitive bacteria (left) and two costly single resistant mutants (center and right). Error bars represent mean ± standard deviation of independent biological replicates (n=3). N.S. non-significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$ (two-tailed Student's *t* test). **Costs are expressed during fast growth**. Schematic representation of multi-fork DNA replication of *E. coli* during fast growth (**b**) or slow growth (**c**). Black dots represent the origin of replication, and red lines represent transcription forks; green arrows mark regions of potential conflicts between replication and transcription forks. **d.** Correlation between the fitness cost between time 0 and 4h (y axis) and the fitness cost between time 0 and 24h (y axis), in LB broth (data from the experiments shown in Figure 1). The black line represents the linear regression if the costs were identical. The blue line represents the linear regression line, and the grey area represents the 95% confidence interval. **e.** Correlation between the fitness cost between time 0 and 4h (y axis) and the fitness cost between time 4h and 24h (y axis), in LB broth (data from the experiments shown in Figure 1). The black line represents the linear regression if the costs were identical. The blue line represents the linear regression line, and the grey area represents the 95% confidence interval.