nature microbiology

LETTERS

PUBLISHED: 26 AUGUST 2016 | ARTICLE NUMBER: 16147 | DOI: 10.1038/NMICROBIOL.2016.147

The essential mycobacterial amidotransferase GatCAB is a modulator of specific translational fidelity

Hong-Wei Su[1†], Jun-Hao Zhu[1,2†], Hao Li[1], Rong-Jun Cai[1], Christopher Ealand[3], Xun Wang[1], Yu-Xiang Chen[1], Masood ur Rehman Kayani[4], Ting F. Zhu[4], Danesh Moradigaravand[5], Hairong Huang[6], Bavesh D. Kana[3] and Babak Javid[1*]

Although regulation of translation fidelity is an essential process[1–7], diverse organisms and organelles have differing requirements of translational accuracy[8–15], and errors in gene translation serve an adaptive function under certain conditions[16–20]. Therefore, optimal levels of fidelity may vary according to context. Most bacteria utilize a two-step pathway for the specific synthesis of aminoacylated glutamine and/or asparagine tRNAs, involving the glutamine amidotransferase GatCAB[21–25], but it had not been appreciated that GatCAB may play a role in modulating mistranslation rates. Here, by using a forward genetic screen, we show that the mycobacterial GatCAB enzyme complex mediates the translational fidelity of glutamine and asparagine codons. We identify mutations in *gatA* that cause partial loss of function in the holoenzyme, with a consequent increase in rates of mistranslation. By monitoring single-cell transcription dynamics, we demonstrate that reduced *gatCAB* expression leads to increased mistranslation rates, which result in enhanced rifampicin-specific phenotypic resistance. Consistent with this, strains with mutations in *gatA* from clinical isolates of *Mycobacterium tuberculosis* show increased mistranslation, with associated antibiotic tolerance, suggesting a role for mistranslation as an adaptive strategy in tuberculosis. Together, our findings demonstrate a potential role for the indirect tRNA aminoacylation pathway in regulating translational fidelity and adaptive mistranslation.

We had shown previously that by genetically enhancing rates of translational error by the expression of mutated tRNAs *in trans* in mycobacteria we could dramatically increase the rates of phenotypic resistance to the first-line antibiotic rifampicin[18]. Mistranslation rates of glutamine to glutamate and asparagine to aspartate in wild-type (WT) mycobacteria were at least tenfold higher[18] than that observed in *Escherichia coli*-K12 (ref. 26) and rose in response to environmental stressors[18]. However, the molecular mechanism underpinning this observation was unresolved, which we hypothesized to be associated with regulation of translational fidelity. To further study this, we devised a forward genetic screen based on two complementary gain-of-function reporters that directly measure mistranslation[18,27] to identify high mistranslator mutants and used these in tandem in an initial selection, followed by a luminescence-based screen of the survivors from the selection (Supplementary Fig. 1). For the selection, we used a reporter strain carrying a point mutation in the aminoglycoside phosphotransferase gene *aph*, which abrogated the ability to confer kanamycin resistance, thus allowing for the relative measurement of asparagine to aspartate mistranslation. Rates of mistranslation were measured by monitoring the relative survival of *Mycobacterium smegmatis* on kanamycin-agar; colonies that emerged were genetic revertants, *de novo* resistant mutants, or mycobacteria with high mistranslation specifically for asparagine to aspartate. After discarding kanamycin-resistant mutants, we screened for *bona fide* high mistranslator mutants using a dual-luciferase assay[18] and identified three mutants with high rates of mistranslation of asparagine to aspartate and glutamine to glutamate (Fig. 1b) but not other errors (Supplementary Fig. 2). Whole-genome sequencing of the mutants identified a number of single nucleotide polymorphisms (SNPs; Supplementary Table 1) and we focused on a single gene, *gatA*, which harboured independent mutations in all three strains (Fig. 1b).

In many bacteria, non-discriminatory glutamyl- and aspartyl-tRNA synthetases naturally misacylate glutamine tRNA to Glu-tRNA^{Gln} and asparagine tRNA to Asp-tRNA^{Asn}. The GatCAB enzyme complex is involved in the transamidation of these naturally misacylated tRNAs to their cognate Gln-tRNA^{Gln} and Asn-tRNA^{Asn} (refs 22,24). Isogenic mutants differing only in their sequence of *gatA* (Supplementary Fig. 3a) also showed high mistranslation rates (Fig. 1c), and complementation of the mutant strains with WT *gatCA* complemented the mistranslation rate phenotype (Supplementary Fig. 4c), verifying that the *gatA* mutations were necessary and individually sufficient to confer the phenotype. Earlier work had shown that EF-Tu discriminates against a number of misacylated tRNAs, including physiologically misacylated Glu-tRNA^{Gln} and Asp-tRNA^{Asn} (refs 2,28), suggesting that even if GatCAB function was suboptimal, these misacylated tRNAs would not be available for translation. However, expression of a non-discriminatory AspRS *in trans* in *E. coli* led to substantial rates of mistranslation[14], suggesting that the discrimination afforded by EF-Tu *in vivo* is possibly overwhelmed when concentrations of misacylated tRNA increase over a

[1]Collaboration Innovation Centre for the Diagnosis and Treatment of Infectious Diseases, School of Medicine, Tsinghua University, Beijing 100084, China. [2]School of Life Sciences, Peking University, Beijing 10087, China. [3]Faculty of Health Sciences, DST/NRF Centre of Excellence for Biomedical TB Research, University of the Witwatersrand, National Health Laboratory Service, Johannesburg 2193, South Africa. [4]Center for Synthetic and Systems Biology, MOE Key Laboratory of Bioinformatics, Collaboration Innovation Centre for the Diagnosis and Treatment of Infectious Diseases, School of Life Sciences, Tsinghua University, Beijing 100084, China. [5]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [6]National Clinical Laboratory on Tuberculosis, Beijing Tuberculosis and Thoracic Tumor Institute, Beijing Chest Hospital, Capital Medical University, Beijing 101149, China. [†]These authors contributed equally to this work. *e-mail: bjavid@gmail.com

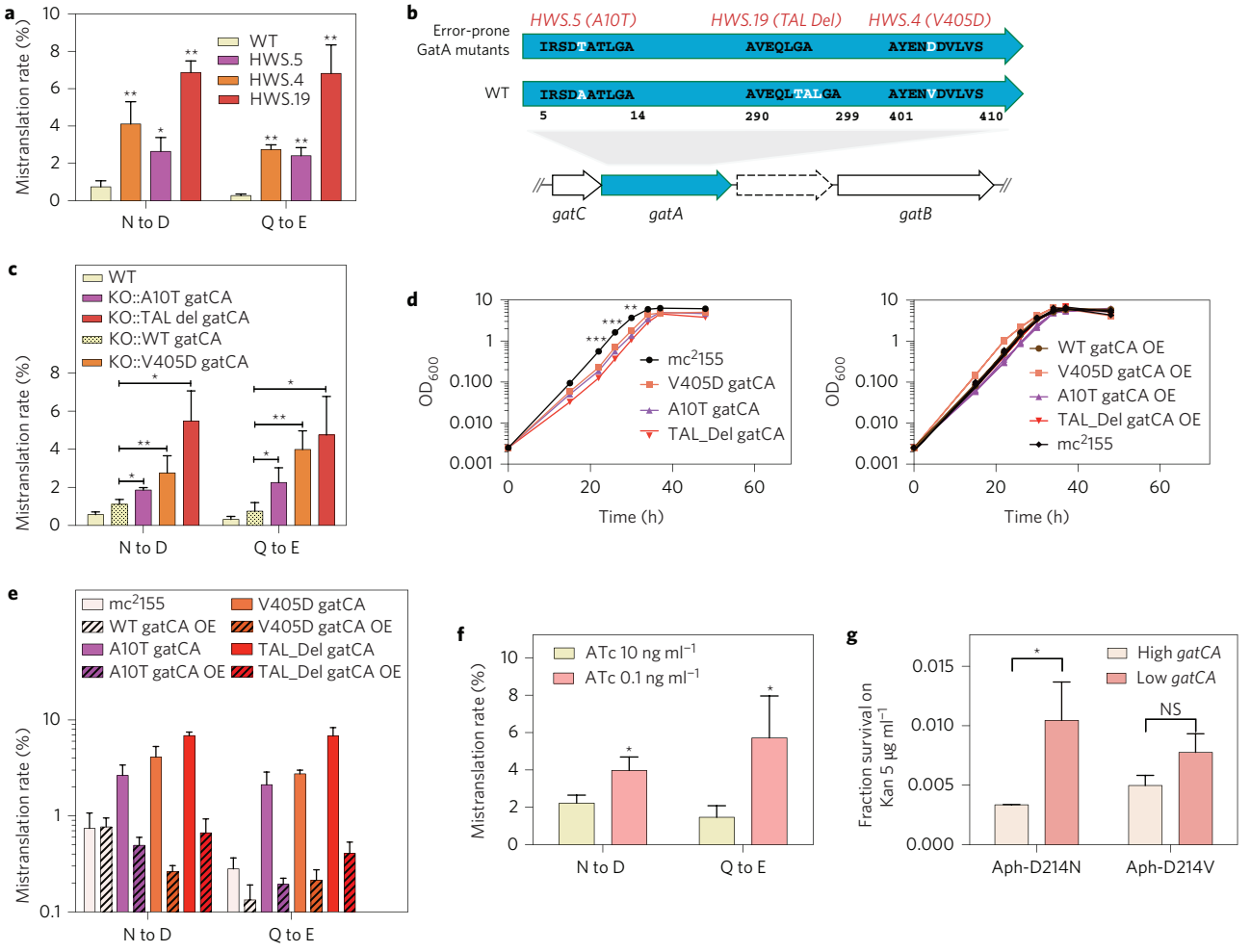

Figure 1 | A forward genetic screen identifies *gatCAB* as a mediator of specific translational fidelity. a, Specific mistranslation rates of isolated mutants compared with WT *M. smegmatis*. **b**, Schematic showing the amino acid changes in GatA in the high mistranslating mutants. **c**, Specific mistranslation rates of *gatA* mutations on an isogenic background. **d**,**e**, Growth (**d**) and mistranslation (**e**) rates of mistranslator mutant strains complemented by the cognate mutant *gatA* allele. **f**, Mistranslation rates of low (ATc 0.1 ng ml^{-1}) and high (ATc 10 ng ml^{-1}) *gatB* expression in a tetracycline-regulated promoter strain. **g**, WT cells expressing high or low *gatCA* as measured by a fluorescent reporter and expressing the Aph-D214N or Aph-D214V mistranslation reporter were isolated by flow cytometry, and mistranslation rates were measured by survival on kanamycin-agar. Data are presented as means of three biological replicates ± s.d. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by *t*-test. Kan, kanamycin.

critical threshold. To determine whether the SNPs caused a loss or aberration of function, we decided to perform complementation of the mutant strain by the cognate mutant allele. Loss of function mutants should be complemented by increased gene dosage, whereas aberrant or gain of function mutants should have their phenotype exacerbated by a second copy of the mutated *gatA* gene. Complementation of the mutant strains by overexpression of the cognate mutant alleles led to a decrease in the mistranslation and growth defect phenotypes (Fig. 1d,e) suggesting that the *gatA* mutations probably cause a partial loss of GatCAB activity, leading to increased abundance of misacylated Glu-tRNA^{Gln} and Asp-tRNA^{Asn}, thus resulting in increased mistranslation. We next assessed if limitation of WT *gatCAB* expression would have the same effect. We created a strain that allowed for regulated *gatB* expression (Supplementary Fig. 3b) and, in this case, depletion of gene expression resulted in increased mistranslation (Fig. 1f). We investigated the expression of *gatCA* in WT cells and found that cells with decreased *gatCA* expression showed increased mistranslation specifically of asparagine to aspartate (Fig. 1g), hence cellular variation of GatCAB abundance appears to regulate mistranslation in mycobacteria.

In the isolated mutant strains, the abundance of both GatA and GatB was decreased (Fig. 2a,b) despite an increase in mRNA expression (Supplementary Fig. 4a), suggesting that the GatCAB complex in the *gatA* mutant strains may be unstable. We therefore examined the abundance and stability of GatA and GatB in both WT and mistranslating strains while protein synthesis was blocked with chloramphenicol (which is bacteriostatic in mycobacteria—not shown). Surprisingly, although GatA abundance was decreased in the *gatA* mutant strains, its levels were stable, whereas GatB (WT in all cases) showed an accelerated decrease in abundance in all three mutant strains (Fig. 2c,d). Furthermore, GatA levels were decreased but not ATc responsive in the *gatB* promoter replacement strain (Supplementary Fig. 3c,d). Complementation of the *gatA* mutant strains with the cognate *gatA* allele restored both GatA and, importantly, GatB abundance (Supplementary Fig. 4b). Mapping of the residues associated with mistranslation (Supplementary Fig. 5) verified that they are surface-exposed and not close to the enzymatic region of GatA, further supporting the notion that the partial loss of function of GatCAB in *gatA* mutant strains may be due to loss of stability of the complex.

We next investigated whether variation in cellular mistranslation rates would result in rifampicin-specific phenotypic resistance (RSPR)[18], which is a distinct form of antibiotic tolerance to

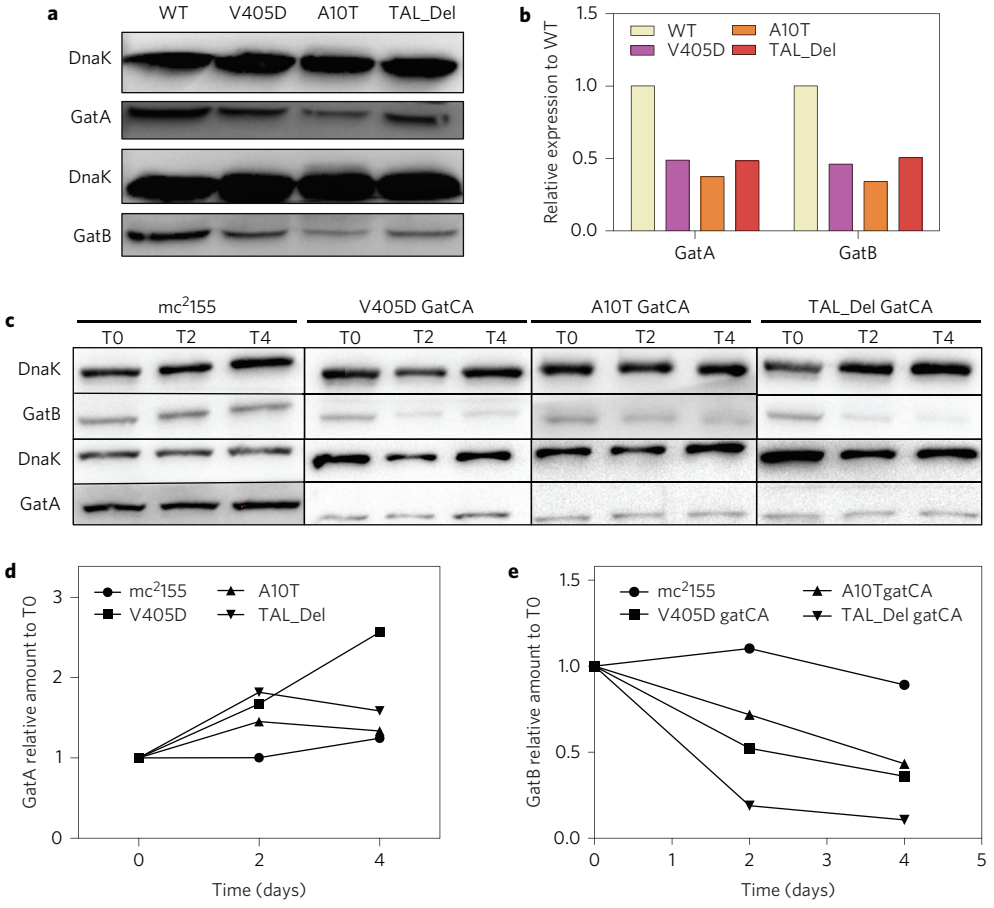

Figure 2 | Mutations in *gatA* result in decreased GatCAB abundance and instability of WT GatB. a,**b**, Western blot (**a**) and quantification (**b**) of GatA and GatB in mistranslator and WT strains. DnaK is used as a loading control. **c**,**d**, Western blot (**c**) and quantification (**d**) of GatA and GatB (**e**) in mistranslator and WT strains at two (T2) and four (T4) days following inhibition of protein synthesis with chloramphenicol. Blots are representative of at least three independent experiments with similar results.

rifampicin[29,30] in which genetically susceptible mycobacteria are able to not only survive, but grow in the presence of the drug. This RSPR phenotype is in contradistinction to both classical genetic resistance, which is defined as an increase in MIC, which was not observed in our strains (Supplementary Tables 4 and 5), and non-replicating persistence, in which antibiotic-tolerant cells do not grow in the presence of lethal concentrations of antibiotic. Both the isolated mutants from the forward genetic screen and *gatA* mutations on an isogenic background showed dramatically increased RSPR (Fig. 3a–c), and depletion of *gatB* expression resulted in similar observations (Fig. 3d). Expression of a second copy of the mutated *gatA* gene in the cognate mutant strain complemented the RSPR phenotype (Fig. 3e), verifying that RSPR is due to mistranslation mediated by decreased GatCAB function.

To explore the potential role for *gatA* mutations in tuberculosis disease, we identified a number of SNPs in *gatA* in *Mycobacterium tuberculosis* clinical isolates from the literature[31] and public databases (www.tbdb.org) and made isogenic *gatA* variants of some of these SNPs that occurred in conserved residues (Supplementary Fig. 6 and Supplementary Table 2). We found that some, but not all, of these SNPs conferred detectably high mistranslation and RSPR (Fig. 3f,g and Supplementary Table 3). Finally, we screened a small, archived, strain collection of clinical isolates at the Beijing Chest Hospital (see Methods) for mutations in *gatCAB*. The rifampicin-sensitive strain (283), isolated after inability of standard therapy to sterilize sputum culture by four months, had a mutation in *gatA* (K61N) and showed increased RSPR compared with the original strain (165) isolated from the patient (Fig. 3h,i and Supplementary Fig. 7a) but no increased tolerance to either isoniazid or streptomycin (Supplementary Fig. 7b,c), suggesting that this was not a generalized persister-phenomenon[32–37]. Although strains 165 and 283 had identical mycobacterial interspersed repetitive units (MIRU) typing, whole genome sequencing revealed them to be different strains (176 SNPs, not shown) and therefore the patient had either been re-infected with strain 283 following initiation of therapy or had been infected with two different strains at the time of diagnosis and 283 was the dominant isolated strain at 4 months. To ensure that the mutation in *gatA* was therefore responsible for the observed phenotype, we tested the effect of the Mtb *gatA*-K61N mutation in an isogenic (*M. smegmatis*) background and confirmed that the mutation was sufficient for the observed phenotypes (Fig. 3j,k). We also complemented strain 283 with WT *gatA* and were able to fully complement the RSPR phenotype (Fig. 3l), confirming the causal nature of the mutation in *gatA*. Non-synonymous mutations in *gatA* are over-represented in the *M. tuberculosis* genome as a whole, and in genes associated with drug resistance in clinical isolates (Supplementary Fig. 8), possibly suggesting selection. Taken together, these data suggest that modulation of translational fidelity by mutations in *gatA* may be a source of increased rifampicin phenotypic resistance in circulating *M. tuberculosis* clinical isolates. Further studies should be performed to determine whether mutations in *gatCAB* are over-represented in drug-sensitive clinical isolates associated with prolonged sputum cultures in the context of appropriate therapy.

We measured cellular variation in mistranslation of glutamine to glutamate in single cells using a fluorescent reporter (Supplementary

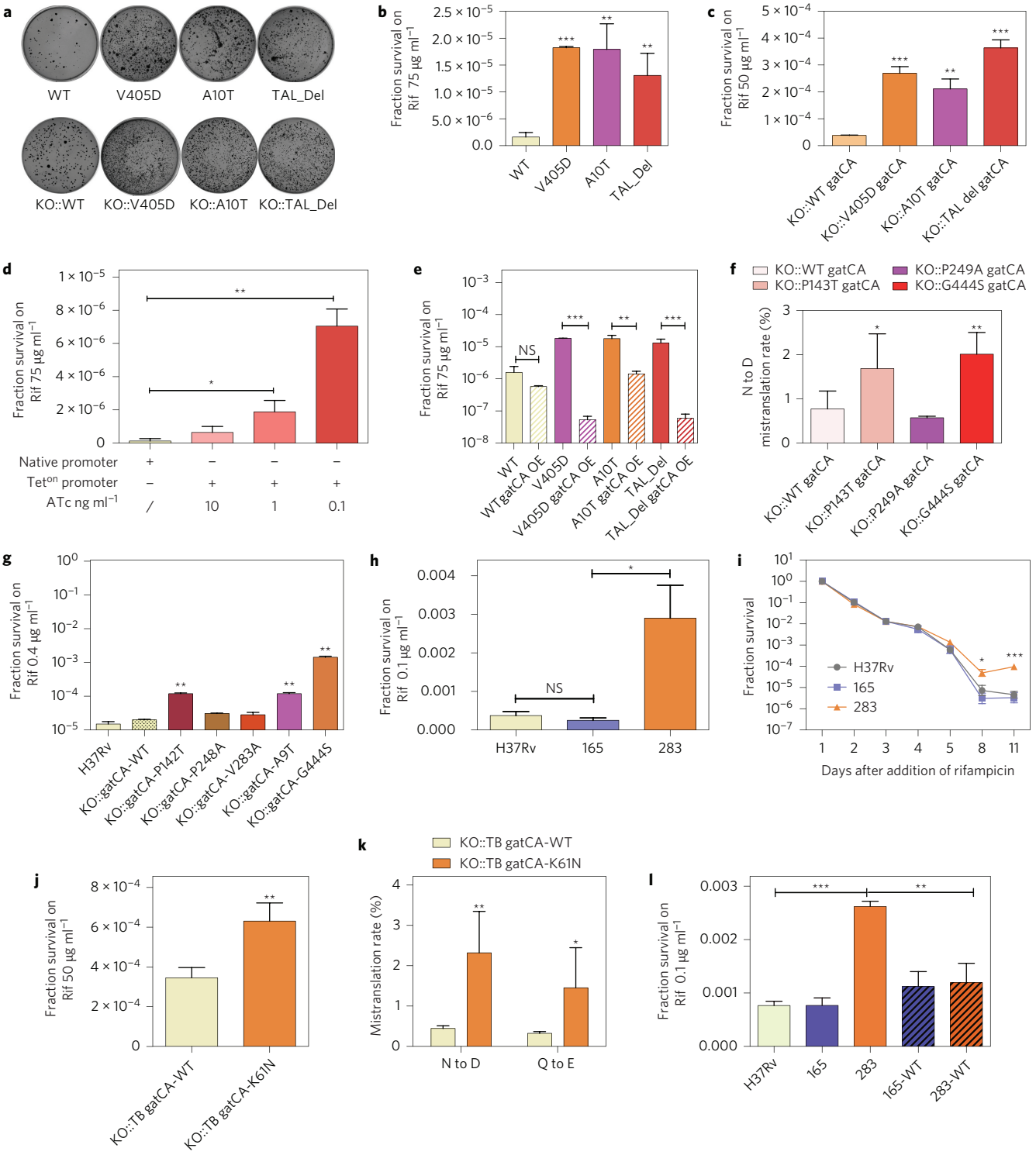

Figure 3 | High mistranslation from *gatA* mutations cause rifampicin-specific phenotypic resistance in *M. tuberculosis*. **a**, Survival and growth on rifampicin-agar of WT and mutant *gatA* strains. **b**,**c**, Quantification of experiments from **a**. **d**, Increased RSPR with decreased expression of *gatB* in a tetracycline-regulated *gatB* promoter strain. **e**, Mistranslator mutants complemented with the cognate mutant *gatA* allele no longer have increased RSPR. **f**, Increased mistranslation of the conserved *gatA* mutations in an isogenic *M. smegmatis* strain measured by the Renilla-Firefly reporter. **g**, RSPR in *gatA* mutants of *M. tuberculosis*. **h**,**i**, RSPR (**h**) and rifampicin-kill curve (**i**) of strains 165 (WT *gatA*) and 283 (*gatA*-K61N). **j**,**k**, RSPR (**j**) and mistranslation rates (**k**) of *M. tuberculosis gatA*-K61N mutation on an isogenic *M. smegmatis gatA::ko* background. **l**, Complementation of strains H37Rv, 165 and 283 with WT *gatCA* results in abolition of the RSPR phenotype in strain 283. Data are presented as means of three biological replicates ± s.d. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; NS, not significant; Rif, rifampicin.

Fig. 9). The relative error rate in *M. smegmatis* was significantly greater than in *E. coli*-K12 (Fig. 4a), which lacks the GatCAB enzyme and the indirect pathway for glutamine and asparagine aminoacyl tRNA synthesis, and is thought to have acquired the direct aminoacylation pathway[38]. Isolation of mycobacterial cells by low and high mistranslation rates showed that individual

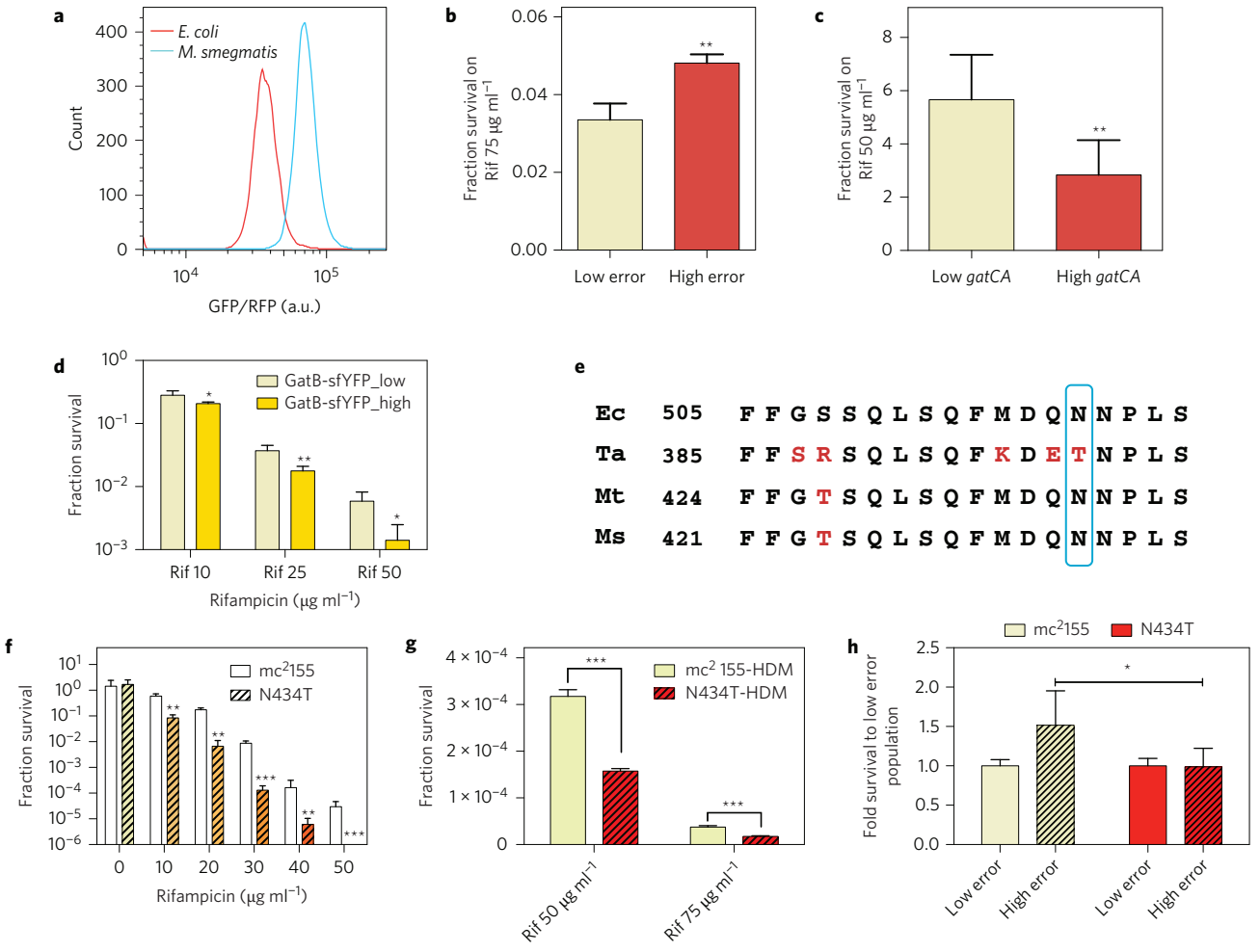

Figure 4 | Mistranslation of a specific residue of RpoB causes RSPR. a, Relative mistranslation of *E. coli* and *M. smegmatis* measured using a single-cell reporter. **b**–**d**, RSPR of cells isolated by flow cytometry with high versus low mistranslation and high versus low *gatCA* expression (**c**) or high versus low GatB expression (**d**). **e**, Alignment of the rifampicin-resistance-determining region of *E. coli* (Ec), *Thermus aquaticus* (Ta), *M. tuberculosis* (Mt) and *M. smegmatis* (Ms). Substitutions that differ from *E. coli* are shown in red, and the mutated residue is boxed. **f**, RSPR of WT (mc^2155) and RpoB-N434T *M. smegmatis*. **g**, RSPR of WT and RpoB-N434T *M. smegmatis* transformed with a plasmid causing specific mistranslation of asparagine to aspartate (HDM). **h**, Cells from WT and RpoB-N434T mycobacteria isolated with high or low mistranslation as measured by the single-cell reporter from **a**. Data are presented as means of three biological replicates ± s.d. except in **a**, which is representative of three separate experiments with each experiment having two replicates. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by *t*-test.

cells with high mistranslation had higher survival and growth on rifampicin (Fig. 4b), as did cells with lower expression of *gatCA* (Fig. 4c). In a strain where GatB was C-terminally tagged with a fluorescent reporter protein (Supplementary Fig. 10), isolation of low versus high GatB expressing cells showed a similar phenotype (Fig. 4d), suggesting that variation in cellular GatCAB is sufficiently limiting to cause RSPR due to adaptive mistranslation. The cellular target of rifampicin is the β subunit of the RNA polymerase. To determine the mechanism by which variation in cellular mistranslation rates may results in RSPR, we examined the previously published structure of RNAP in complex with rifampicin[39]. It has been noted that the asparagine residue N434 (*M. smegmatis* numbering) contributes to rifampicin binding, while not itself having direct contact with the antibiotic[39]. Furthermore, its mutation to aspartate results in rifampicin resistance. However, *Thermus aquaticus*, from which the crystal structure is derived[39], has a naturally occurring substitution to threonine at that position (Fig. 4e), but is still rifampicin-sensitive. Through single-stranded DNA recombineering[40], we made a *M. smegmatis* strain differing to the parent only at that residue (strain RpoB-N434T). This strain had similar resistance to rifampicin (minimum inhibitory concentration (MIC) = 3 μg ml^{-1}, MIC of WT = 2.5 μg ml^{-1}), and MICs to ciprofloxacin and isoniazid equal to those of WT (not shown). However, in RpoB-N434T, residue 434 (threonine) would no longer be mistranslated to N434D by excess misacylated Asp-tRNA^{Asn}, unlike the WT N434 residue. Strain RpoB-N434T had a similar MIC to rifampicin as its parent, but was considerably less phenotypically resistant to the drug (Fig. 4f) as measured by the plate growth assay, especially at higher rifampicin concentrations. Importantly, expression of tRNA^{Asp}, with its anticodon mutated to asparagine, which causes excess mistranslation of asparagine to aspartate[18] (strain HND—high asparagine to aspartate mistranslation), increased RSPR in WT much more than in the RpoB-N434T strain (Fig. 4g) and, unlike WT cells, isolation of high mistranslating cells in the RpoB-N434T strain showed only slightly increased RSPR (Fig. 4h), suggesting that mistranslation of critical residues in RNAP contribute significantly to rifampicin phenotypic resistance.

Our data support the notion that the evolutionary retention of the indirect aminoacylation pathway and GatCAB by most bacterial species may allow for specific fine-tuning of translational fidelity while avoiding the perils of error catastrophe associated with low-fidelity ribosomes[13]. We speculate that, at least in mycobacteria

and potentially in other bacterial species, the indirect aminoacylation pathway may play a role in the generation of phenotypic diversity via errors in gene translation.

Methods

Bacterial strains and culture. WT *M. smegmatis* mc^2-155 (ref. 41) and derivatives thereof (Supplementary Table 7) were cultured in Middlebrook 7H9 media supplemented with 0.2% glycerol, 0.05% Tween-80, 10% ADS (albumin-dextrose-salt) and antibiotics, as appropriate. *M. tuberculosis*-H37Rv and derivatives or other *M. tuberculosis* strains were cultured in 7H9 media supplemented with 0.2% glycerol, 0.05% Tween-80, 10% OADC (oleic acid, albumin, dextrose and catalase) and antibiotics as appropriate. *E. coli* strains DH5α and BL21 were maintained in lysogeny broth (LB). If not otherwise noted, cells were grown and maintained at 37 °C with shaking.

Clinical isolates. The Biobank resource of the Beijing Chest Hospital archives frozen clinical isolates from patients (Clinical Database and Sample Bank of Tuberculosis of Beijing). Dr Huang's laboratory has IRB approval (2014-36-1) for interrogation of these isolates. All isolates were de-identified and were not linked to clinical outcome other than by the date of the sample and time within the clinical course that the samples were obtained. We asked to search the anonymized sample set (of ∼20 isolates) from patients with prolonged (>4 months) sputum culture positivity but whose isolate was deemed fully drug-sensitive by drug-susceptibility testing. The *gatCAB* locus from DNA extracted from these isolates was sequenced and we found 1 isolate from 20 (strain 283) with a mutation in *gatA* (K61N). We then asked for information regarding the presenting clinical isolate from that patient (strain 165). Both strains 165 and 283 were revived and grown under standard conditions for the assays described in the following.

Forward genetic screen. A total of 1×10^7 colony forming units (c.f.u.) of WT *M. smegmatis* mc^{2}155 transformed with a plasmid expressing a mutated *aph* gene-D214N (ref. 18) were spread on each of ten LB-agar plates with 10 μg ml^{-1} kanamycin. Colonies that grew on the low-dose kanamycin plates were inoculated into 7H9 with 25 μg ml^{-1} kanamycin for identification of genetic kanamycin resistance. Those clones that were not resistant to high-dose kanamycin were selected for the secondary screen, which employed Renilla-Firefly dual luciferase reporters[18]. These reporters were transformed into candidate strains and specific mistranslation rates were measured to identify high mistranslator mutant strains. The original reporter plasmid was cured from candidates before further experiments.

Phenotypic resistance on agar plate assay. Stationary phase (optical density at 600 nm, $OD_{600nm} = 2.0$) *M. smegmatis* mc^{2}155 cultured in 7H9 was pelleted and re-suspended with culture media. The c.f.u.s were enumerated by plating aliquots of multiple tenfold serial dilution onto LB agar containing antibiotics of interest. Fractional survival was calculated as c.f.u.s growing on selective media divided by c.f.u.s on non-selective media. Usually, multiple dilutions were subjected to one antibiotic concentration to ensure quantifiable plates. The same method was applied for *M. tuberculosis* H37Rv, with the exception that all bacteria were plated on 7H10 or 7H11 agar media.

Dual luciferase mistranslation assay. The mistranslation assay was carried out as previously described[18]. Briefly, strains expressing reporters were grown to late log phase, then diluted 1:20 into 7H9 medium, followed by induction of the reporter with anhydrotetracycline (ATc, 50 ng ml^{-1}). After 6 h induction, cells were collected and lysed by passive lysis buffer provided by a dual luciferase kit (Promega), and luminescence was measured by a Fluoroskan Ascent FL luminometer with 1,000 ms as integration time according to the manufacturer's instructions (Promega).

Single-cell mistranslation reporter. A mutated green fluorescent protein (GFP) (E222Q) was constructed by site-directed mutagenesis using standard methodology and found to exhibit <1% fluorescence of the WT GFP protein. This mutant protein was fused to WT monomeric red fluorescent protein (mRFP) with a flexible GGSGGGGSGGGSSGG linker between the two fluorescent proteins and cloned into an episomal vector (pTet) under control of a tetracycline-inducible promoter. Relative mistranslation rates were measured by comparing the ratio of green to red fluorescence by flow cytometry, with a high ratio indicating a high cellular mistranslation rate.

***M. smegmatis gatCA* isogenic mutant construction.** The *gatCAB* genes are essential and therefore cannot be deleted using allelic exchange. The *M. smegmatis gatCA* operon with 125 bp upstream of the coding region was cloned by primers Msm gatCA-F and Msm gatCA-R into vector pML1342 (Addgene), which integrates into the mycobacterial L5 phage integration site. Zeo-F and Zeo-R were used to clone the zeocin marker and replace the hygromycin marker of pML1342 to make vector pZML1342. After transforming the vector containing different genotype *gatCA* operons (WT; A10T; V405D; TAL_Del) into mc^{2}155 to make a *gatCA* merodiploid strain, the strain was transformed with the recombineering plasmid pNIT(kan)::RecET::sacB plasmid, a kind gift from the Rubin laboratory. A streptomycin-resistant cassette flanking the 550 bp upstream and 534 bp downstream region of *gatCA* was used as an allele exchange substrate (AES). Fusion PCR was used to make the AES. US550-F and US550-R were used to amplify the US550; DS534-F and DS534-R were used to amplify the DS534; and StrepR-F and StrepR-R were used to amplify the streptomycin-resistant cassette. The three dsDNA fragments were ligated together through fusion PCR, and the entire product was ligated into a blunt end cloning vector pJet1.2 (ThermoScientific). This plasmid was used as a template to amplify the AES by PCR. A PCR purification kit (Qiagen) was used to purify the double-stranded DNA AES for transformation into the recombination competent mycobacterial strain[42].

The strain was grown to $OD_{600nm} = 0.4$, and expression of *recET* was induced with 10 μM isovaleronitrile (IVN) for 5 h. Once RecET was induced, competent cells were made by standard methods and transformed with 2 μg of AES. The cells were allowed to recover for 3 h and spread on streptomycin 25 μg ml^{-1} LB-agar plates. After three days, colonies were screened for the right recombinants and confirmed by Southern blotting.

***M. smegmatis gatB* knockdown strain construction.** A regulated promoter strain for *gatA* could not be made, possibly due to polar effects on downstream *gatB* expression. We therefore chose to make a tet-regulated *gatB* promoter strain. The pSES suicide vector[42] was modified to include a tetracycline-on operator and *tetR* gene, and then the first 550 bp of the *gatB* homologous sequence was used to make an ATc-inducible *gatB* expression strain. Plasmid (1 μg) was mixed with 400 μl mc^{2}155 competent cells and transformed by electroporation using an electroporator set to 2,500 V, 25 μF and 1,000 Ω. The transformants were allowed to recover for 3 h and the cells were spread on LB-agar plates supplemented with 50 μg ml^{-1} hygromycin. The recombinants were screened by PCR. There was a clear ATc-dependent growth phenotype (Supplementary Fig. 3e) associated with a small increase in GatB protein levels (Supplementary Fig. 3d), but the strain could not be fully complemented, even with high concentrations of ATc (not shown).

***M. tuberculosis gatCA* isogenic mutant construction.** This was performed as described previously[43]. The H37Rv *gatCA* regions 2,000 bp upstream and downstream were amplified with primers TBUS2000-F/TBUS20000-R and TBDS2000-F/TBDS2000-R, respectively. These were ligated into the p2NIL suicide vector. The *sacB* and *lacZ* gene were then cut from pGOAL17 and ligated into p2NIL::US2000::DS2000 to make the *gatCA* deletion construct- pΔTB gatCA. For complementation (prior to deletion, to make a merodiploid), the L5 integration vector (pZML1342) containing different genotype *gatCA* operons (P142T, P248A, V283A) was used. Different genotype genes were made by site-directed mutagenesis using primers described in Supplementary Table 6. Two 2 μg suicide vectors were transformed into the H37Rv strain and selected on 7H11 plates supplemented with kanamycin 25 μg ml^{-1} and X-gal 40 μg ml^{-1}. Blue colonies were screened for single cross-over by Southern blot. A *gatCA* merodiploid strain was then made in the positive single cross-over strain by standard transformation protocols in *M. tuberculosis*. The cells were grown to stationary phase (OD_{600} of ∼3) and spread onto 7H11 plates containing X-gal and 2% sucrose after tenfold serial dilution. White colonies were screened for double cross-over strains with the original locus knocked out and verified by Southern blotting.

Single-stranded DNA recombineering. WT *M. smegmatis* mc^{2}155 was transformed with plasmid pKM402 (gift from K. Murphy), which expresses the mycobacteriophage-derived recombinases *recT*. This strain was grown to log phase (OD_{600} of ∼0.4) and *recT* expression induced with 200 ng ml^{-1} ATc. When the OD_{600nm} reached ∼1.0, cells were collected and rendered electrocompetent by standard methods. Because we wished to make strain RpoB-N434T, this strain could not be made by selection. We enriched for DNA competent cells by co-transforming 1 μg of the specific oligonucleotide (N434T-short-lagging) with 300 ng of a hygromycin-marked plasmid (pTet-RenFF). Following transformation, cells were recovered overnight before plating on hygromycin-LB agar, and the resulting transformants were screened by mismatch amplification mutant assay (MAMA)-PCR for positive recombinants[40].

***M. smegmatis gatB-sfYFP* fluorescent reporter strain construction.** A strain with a tagged native *gatA* could not be constructed (not shown) either due to polar effects on downstream *gatB* or because the protein could not tolerate fluorescent tagging. An *M. smegmatis* strain expressing *gatB* with C-terminal tagged super-folder YFP was therefore constructed with homologous recombineering. Linear dsDNA for recombineering was constructed by overlap PCR amplification. Briefly, a 500 bp DNA sequence flanking the last 500 bp upstream of the stop codon as well as 3′UTR of *gatB* coding DNA sequence (CDS) was amplified with primer pairs *gatB*_CCDS_F/*gatB*_sfYFP_R and *gatB*_DNS_F/*gatB*_DNS_R. Histidine-tagged sfYFP was amplified from pMV261_msfYFP (gift from E.J. Rubin), and Zeocin resistant markers flanked with *loxP* sites were amplified from pKM Zeo-lox plasmid. Overlapping sequences were introduced into each fragment through PCR primers for further fusion with overlap PCR. The 2.4 kbp final PCR product was cloned into PCR2.1 plasmid by TOPO cloning and sequence verified to avoid PCR errors. Correct sequences were further amplified by KOD hotstart high-fidelity polymerase,

and dialysed. An *M. smegmatis* strain with a nitrite-inducible RecET plasmid (pNit_RecET-sacB-kan) was cultured in 7H9 broth supplemented with 10% ADC, 0.2% glycerol, 0.05% Tween-80 and 25 μg ml^{-1} kanamycin. RecET recombinase expression was induced with 10 μM isovaleronitrile (IVN) overnight and electrocompetent cells were prepared according to a standard protocol. Cells were electroporated with 1 ug of purified and dialysed PCR product, recovered for 4 h at 37 °C in 7H9 medium before plating on LB solid medium supplemented with 20 μg ml^{-1} zeocin. Candidate recombinants were verified with PCR as well as western blot, with both mouse monoclonal anti-His tag antibody and rabbit polyclonal anti-GatB antibody. Bona fide sfYFP tagged strains were then streaked on LB solid medium containing 20 μg ml^{-1} zeocin, 24 colonies were further patched on LB solid medium containing 25 μg ml^{-1} kanamycin, and colonies that failed to grow on kanamycin-containing plates were used for downstream applications.

MAMA-PCR. Hygromycin-resistant transformants from the oligonucleotide recombineering reaction (see above) were screened by MAMA-PCR for recombinants with the N434T substitution in RpoB. Individual colonies were patched and then grown to the stationary phase: a 0.5 ml volume was boiled for 20 min to liberate genomic DNA, the cells were spun down and the supernatant was used as a template for subsequent PCR reactions. Invitrogen Platinum Taq DNA polymerase was used in MAMA-PCR. Positives were verified by Sanger sequencing of the RpoB rifampicin-resistance-determining region. Once the positive clone was identified, the pKM402 and pTet-RenFF plasmids were cured by growing the strain to stationary phase and then patching onto kanamycin and hygromycin plates: a kanamycin-sensitive, hygromycin-sensitive clone was selected for further evaluation in subsequent studies.

Genomic DNA extraction and Southern blot. Genomic DNA was extracted by standard phenol chloroform extraction. Culture aliquots (5 ml) were spun down and re-suspended in 500 μl Tris-EDTA (TE) buffer. After boiling in 80 °C for 1 h, protease K and SDS were added and incubated at 60 °C for 1 h. Then 10% hexadecyl trimethyl ammonium bromide (CTAB) and 5 M NaCl were added and incubated for 20 min at 60 °C, and the result was frozen at −80 °C. Following overnight freezing, the genomic DNA was extracted by phenol chloroform. Southern blot was performed with the DIG DNA labelling and detection kit (Roche). Genomic DNA (1 μg) was digested with corresponding enzymes overnight and the digest run on a 1% agarose gel. The DNA was transferred onto Hybond N membrane by electrophoresis with 600 mA for 2 h at 4 °C. The DNA was crosslinked to the membrane by an ultraviolet crosslinker. The hybridized membrane was probed with a probe synthesized by the PCR DIG probe synthesis kit (Roche) at 54 °C overnight. The membrane was washed, blocked and probed by antibody according to the manufacturer's instructions.

Reverse-transcriptase quantitative PCR. RNA was extracted by Trizol reagent and bead-beating, then 25 ml of OD_{600nm} = 0.8 culture was spun down and TRIZOL reagent added to re-suspend the pellet. The mixture was transferred into 2 ml ribolysing tubes. Cell lysis was performed for 40 s four times, with 90 s intervals on ice every time. The sample was incubated at room temperature for 5 min and then centrifuged for 10 min to discard cell debris. Chloroform (0.2 ml) was added to the supernatant and shaken for 15 s vigorously and then centrifuged at 13,000 r.p.m. for 15 min at 4 °C. The aqueous (topmost) layer was transferred to a new tube, 0.5 ml isopropanol was added, and the mixture was incubated for 2 h at −20 °C. This was then centrifuged at 13,000 r.p.m. for 15 min at 4 °C. The supernatant was decanted and the pellet washed once with 1 ml 70% ethanol (made in diethyl pyrocarbonate (DEPC) water). After ethanol removal, the pellet was allowed to air dry briefly (~5 min) and re-suspended in 100 ul DEPC treated water. Following this, a Qiagen RNA extraction kit was used to purify again, then Turbo Dnase was used to digest genomic DNA. Biorad iScript supermix was used to perform reverse transcription PCR and iQ SYBR Green Super Mix was used for quantitative PCR. *MysA* was used as the reference gene, and the primers are described in Supplementary Table 1. All PCR reactions were performed in triplicate. The $2^{-\Delta\Delta Ct}$ calculation was used to calculate the relative expression level of mutants to WT.

Western blot. Western blot was performed using standard methods. Rabbit polyclonal antibodies to *M. smegmatis* GatA and GatB were raised by immunizing rabbits with peptides from the respective proteins (Thermo). Specificity was confirmed by performing western blotting against lysates of *M. smegmatis* mc^{2}155 over-expressing *gatCAB* as a positive control and *E. coli* lysates as a negative control. To control for lane loading, membranes were also probed with a monoclonal antibody against mycobacterial DnaK (TS29, Abcam). Band integrated density was quantified by ImageJ.

Protein stability by western blot assay. The mycobacterial strains were grown in 50 ml supplemented 7H9 medium to mid-log (OD_{600} = 1) phase. Chloramphenicol (final concentration 300 μg ml^{-1}) was added to the cultures at that time (time = 0) to inhibit protein synthesis. Aliquots (10 ml) of bacterial culture were taken at indicated time points and put on ice. Cells were lysed using bead beating, and the protein concentrations of lysates were determined using a Bradford assay. Equal protein concentrations were loaded onto SDS–polyacrylamide gel electrophoresis gels, and membranes were probed by western blot as described already.

Flow cytometry of testing mistranslation level of *E.coli* and *M. smegmatis*. The *E. coli* BL21 strain expressing pET28a-GFP_E222Q_Linker_RFP was grown in LB medium and *M. smegmatis* mc^{2}155 expressing pUVtetOR-GFP_E222Q_Linker_RFP was grown in 7H9 to the desired density (OD_{600} = 0.5–1). Expression of the reporter was induced by dilution of 100 μl original culture into LB medium and 7H9 medium with isopropyl beta-D-1-thiogalactopyrasonide (IPTG) (1 μM) and ATc (50 ng ml^{-1}) for *E. coli* and *M. smegmatis*, respectively, then grown for another 3 h. After 3 h, 1 ml culture was spun down and washed once with PBS, then the cells were re-suspended in 1 ml PBS. The sample GFP and RFP signals were analysed by flow cytometry (BD LSR Fortessa) and the data were presented as a histogram of GFP/RFP ratio.

Flow sorting experiments. Fluorescence-activated cell sorting (FACS) was carried out using a BD FACSAria special order research product (SORP). In general, bacterial samples were cultured to the desired density (OD_{600nm} = 0.4–0.8 unless specified) and then washed with PBS. Single cells were isolated by syringe filtering with 5.0 μm filter units. Multiple forward and side scatter gates were used to exclude debris and cellular aggregates. GFP was excited with a 488 nm laser and monitored with a fluorescein isothiocyanate (FITC) filter with customized settings, and mRFP or mCherry was excited with a 561 nm laser and monitored with the Texas-Red filter. sfYFP was excited at 488 nm and monitored with a PE filter with customized settings.

Different subpopulations were sorted into sterile Eppendorf tubes loaded with 200 μl 7H9 media. In general, following isolation there was a 750 μl total suspension volume with 5×10^5 bacilli of the desired subpopulation. To test the rifampicin-specific phenotypic resistance of each subpopulation, a 100 μl suspension of the sorted population was plated onto LB agar containing 50 and 75 μg ml^{-1} rifampicin. The remaining suspension was diluted 100-fold, and 100 μl of this was plated on LB agar to allow precise quantitation of fraction survival. All plating experiments were carried out in triplicates.

To evaluate *gatCA* and GatB expression variation and corresponding rifampicin tolerance, *M. smegmatis* mc^{2}155 expressing pSE100::P_{gatCA}mCherry or gatB_sfYFP were grown to an OD_{600nm} of 0.4–0.8, and prepared as described already. The highest 10% and lowest 10% GFP or YFP expression subpopulations were gated and sorted into 7H9 medium and plated as above.

To evaluate single-cell mistranslation variation and corresponding rifampicin tolerance, the *M. smegmatis* mc^{2}155 strain expressing pUVtetOR-::GFP_E222Q_linker_RFP was cultured to OD_{600nm} 0.2–0.3, and induced with 50 ng ml^{-1} ATc for 3 h at 37 °C with shaking. Induced culture were then prepared as above. Besides regular gate settings, the level of mistranslation was evaluated by the ratio of GFP to mRFP. A small proportion of cells showed extremely high or low mRFP expression and, to avoid dramatic ratio variation due to extreme mRFP signals (both low or high), an extra gate was used to exclude extreme mRFP signals. Subpopulations with the highest 10% and lowest 10% GFP/mRFP ratio were sorted into 7H9 and examined as described above.

To evaluate the mistranslation level of the *gatCA* expression variation population, we constructed a strain expressing Aph-D214N or Aph-D214V reporter in L5 site (pZML1342-Aph-D214N and pZML1342-Aph-D214V). This strain was transformed with the pSE100 vector containing *gatCA* promoter driven mCherry and sigA promoter driven GFP (P_{gatCA}-mCherry; P_{sigA}-GFP) as a reporter to monitor *gatCA* transcription activity. Based on the strain expressing these two reporters (mc^{2}155::Aph-D214N-pSE100-P_{gatCA}-mCherry/P_{sigA}-GFP; mc^{2}155::Aph-D214V-pSE100-P_{gatCA}-mCherry/P_{sigA}-GFP), the culture was grown to an OD_{600nm} of ~2, and 1 ml cells were washed once with PBS, followed by resuspension of the cells into 1 ml PBS. A total of 1×10^7 cells were analysed and the high *gatCA* (5%) and low *gatCA* (5%) expressing subpopulations (according to the mCherry signal) were sorted, and ~1×10^6 c.f.u.s were spread on kanamycin 5 μg ml^{-1} LB-agar plates for testing the mistranslation levels of N to D and V to D.

Protein structure modelling. The structural model of *M. tuberculosis* GatCAB was performed using the SWISS-MODEL online server. The sequences of *M. smegmatis* GatA and GatB were used as the input and aligned against the crystal structure of *Aquifex aeolicus* GatCAB (PDB 3H0M). Output models were ranked with Qmean4 and GQME scores and the ones with highest scores were used to construct the model. The three structures were docked together using the programme CHIMERA (http://www.cgl.ucsf.edu/chimera/). Sites of residues in GatA mutated in this study were highlighted in red.

Sequencing and data analysis. Genomic DNA for each of the strains was extracted from a cohort of logarithmically growing cells using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). The quality of the extracted DNA was assessed using agarose gel, and about 50 ng of each sample was used to synthesize a sequencing library according to the manufacturer's protocol. Briefly, the DNA was sheared using Covaris S220 ultrasonicator system (Life Technologies), size selected on an agarose gel, adapter ligated, and amplified by PCR. The amplified library was quantified using a Qubit fluorometer (Life Technologies) and the quality was assessed using an

Agilent 2100 Bioanalyzer. Libraries were pooled together and loaded on the flow cell for sequencing using HiSeq2000 (Illumina). The average insert size was ~500 bp and the planned read length was 100 bp.

The overall quality of the raw sequencing data was increased by the removal of contamination by adapter sequences, trimming of low-quality bases and removal of ambiguous bases. High-quality reads were used for alignment with the reference genome (GenBank accession no. NC_008596.1) using BWAtools (version 0.7.5a)[44]. The chances of wrongly mapped reads and a false-positive SNP were minimized by adapting a stringent criterion; that is, reads and SNPs with mapping and SNP quality scores of less than 30 were discarded. The alignments were used to mark and exclude the duplicates from further analysis. SNPs were called using SAMtools (version 0.1.19)[45], and annotated using VCFtools (version 4.0)[46] and SnpEff (version 4.1)[47].

Analysis of SNP frequency in *gatA* in clinical isolates. The rate of non-synonymous mutations (dN/gene length) in *gatA* was compared with other genes in a published set of 25 clinical isolates that adequately represent global diversity[48]. To compare with genes associated with drug resistance, a set of ~70 genes identified with drug resistance were used as comparators[49].

Statistical analysis. All experiments were performed at least three times on separate days (that is, independent experiments). Data representative of the results, with at least three biological replicates where relevant, are shown in the figures and/or tables. None of the samples or experiments were blinded. Means between two sample sets were compared by unpaired, two-directional Student's *t*-tests using graphpad prism or Excel unless otherwise specified. $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$ were considered statistically significant results.

Received 18 December 2015; accepted 19 July 2016; published 26 August 2016

References

1. Bullwinkle, T. J. *et al.* Oxidation of cellular amino acid pools leads to cytotoxic mistranslation of the genetic code. *Elife* **3,** e02501 (2014).
2. LaRiviere, F. J., Wolfson, A. D. & Uhlenbeck, O. C. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* **294,** 165–168 (2001).
3. Lee, J. W. *et al.* Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443,** 50–55 (2006).
4. Liu, Y. *et al.* Deficiencies in tRNA synthetase editing activity cause cardioproteinopathy. *Proc. Natl Acad. Sci. USA* **111,** 17570–17575 (2014).
5. Lu, J., Bergert, M., Walther, A. & Suter, B. Double-sieving-defective aminoacyl-tRNA synthetase causes protein mistranslation and affects cellular physiology and development. *Nat. Commun.* **5,** 5650 (2014).
6. Reynolds, N. M., Lazazzera, B. A. & Ibba, M. Cellular mechanisms that control mistranslation. *Nat. Rev. Microbiol.* **8,** 849–856 (2010).
7. Zaher, H. S. & Green, R. Quality control by the ribosome following peptide bond formation. *Nature* **457,** 161–166 (2009).
8. Leng, T., Pan, M., Xu, X. & Javid, B. Translational misreading in *Mycobacterium smegmatis* increases in stationary phase. *Tuberculosis* **95,** 678–681 (2015).
9. Li, L. *et al.* Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in *Mycoplasma* parasites. *Proc. Natl Acad. Sci. USA* **108,** 9378–9383 (2011).
10. Ling, J., O'Donoghue, P. & Soll, D. Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nat. Rev. Microbiol.* **13,** 707–721 (2015).
11. Netzer, N. *et al.* Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* **462,** 522–526 (2009).
12. Reynolds, N. M. *et al.* Cell-specific differences in the requirements for translation quality control. *Proc. Natl Acad. Sci. USA* **107,** 4063–4068 (2010).
13. Ribas de Pouplana, L., Santos, M. A., Zhu, J. H., Farabaugh, P. J. & Javid, B. Protein mistranslation: friend or foe? *Trends Biochem. Sci.* **39,** 355–362 (2014).
14. Ruan, B. *et al.* Quality control despite mistranslation caused by an ambiguous genetic code. *Proc. Natl Acad. Sci. USA* **105,** 16502–16507 (2008).
15. Yadavalli, S. S. & Ibba, M. Selection of tRNA charging quality control mechanisms that increase mistranslation of the genetic code. *Nucleic Acids Res.* **41,** 1104–1112 (2013).
16. Bezerra, A. R. *et al.* Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proc. Natl Acad. Sci. USA* **110,** 11079–11084 (2013).
17. Fan, Y. *et al.* Protein mistranslation protects bacteria against oxidative stress. *Nucleic Acids Res.* **43,** 1740–1748 (2015).
18. Javid, B. *et al.* Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. *Proc. Natl Acad. Sci. USA* **111,** 1132–1137 (2014).
19. Lee, J. Y. *et al.* Promiscuous methionyl-tRNA synthetase mediates adaptive mistranslation to protect cells against oxidative stress. *J. Cell Sci.* **127,** 4234–4245 (2014).
20. Miranda, I. *et al. Candida albicans* CUG mistranslation is a mechanism to create cell surface variation. *MBio* **4,** e00285-13 (2013).
21. Bailly, M., Blaise, M., Lorber, B., Becker, H. D. & Kern, D. The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. *Mol. Cell* **28,** 228–239 (2007).
22. Curnow, A. W. *et al.* Glu-tRNAGln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94,** 11819–11826 (1997).
23. Huot, J. L. *et al.* Gln-tRNAGln synthesis in a dynamic transamidosome from *Helicobacter pylori*, where GluRS2 hydrolyzes excess Glu-tRNAGln. *Nucleic Acids Res.* **39,** 9306–9315 (2011).
24. Nakamura, A., Yao, M., Chimnaronk, S., Sakai, N. & Tanaka, I. Ammonia channel couples glutaminase with transamidase reactions in GatCAB. *Science* **312,** 1954–1958 (2006).
25. Silva, G. N. *et al.* A tRNA-independent mechanism for transamidosome assembly promotes aminoacyl-tRNA transamidation. *J. Biol. Chem.* **288,** 3816–3822 (2013).
26. Manickam, N., Nag, N., Abbasi, A., Patel, K. & Farabaugh, P. J. Studies of translational misreading *in vivo* show that the ribosome very efficiently discriminates against most potential errors. *RNA* **20,** 9–15 (2014).
27. Wong, S. Y. *et al.* Functional role of methylation of G518 of the 16S rRNA 530 loop by GidB in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **57,** 6311–6318 (2013).
28. Roy, H., Becker, H. D., Mazauric, M. H. & Kern, D. Structural elements defining elongation factor Tu mediated suppression of codon ambiguity. *Nucleic Acids Res.* **35,** 3420–3430 (2007).
29. Adams, K. N. *et al.* Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell* **145,** 39–53 (2011).
30. Aldridge, B. B. *et al.* Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science* **335,** 100–104 (2012).
31. Comas, I. *et al.* Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* **45,** 1176–1182 (2013).
32. Dhar, N. & McKinney, J. D. Microbial phenotypic heterogeneity and antibiotic tolerance. *Curr. Opin. Microbiol.* **10,** 30–38 (2007).
33. Fridman, O., Goldberg, A., Ronin, I., Shoresh, N. & Balaban, N. Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513,** 418–421 (2014).
34. Lewis, K. Persister cells. *Annu. Rev. Microbiol.* **64,** 357–372 (2010).
35. Maisonneuve, E. & Gerdes, K. Molecular mechanisms underlying bacterial persisters. *Cell* **157,** 539–548 (2014).
36. Shah, D. *et al.* Persisters: a distinct physiological state of *E. coli*. *BMC Microbiol.* **6,** 53 (2006).
37. Shan, Y., Lazinski, D., Rowe, S., Camilli, A. & Lewis, K. Genetic basis of persister tolerance to aminoglycosides in *Escherichia coli*. *MBio* **6,** e00078-15 (2015).
38. Siatecka, M., Rozek, M., Barciszewski, J. & Mirande, M. Modular evolution of the Glx-tRNA synthetase family—rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. *Eur. J. Biochem.* **256,** 80–87 (1998).
39. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* **104,** 901–912 (2001).
40. van Kessel, J. C. & Hatfull, G. F. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol. Microbiol.* **67,** 1094–1107 (2008).
41. Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. Jr Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4,** 1911–1919 (1990).
42. Siegrist, M. S. *et al.* Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. *Proc. Natl Acad. Sci. USA* **106,** 18792–18797 (2009).
43. Parish, T. & Stoker, N. G. Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis* tlyA plcABC mutant by gene replacement. *Microbiology* **146,** 1969–1975 (2000).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
45. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
46. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).
47. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6,** 80–92 (2012).
48. Gagneux, S. *et al.* Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 2869–2873 (2006).
49. Koser, C. U. *et al.* Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *New Engl. J. Med.* **369,** 290–292 (2013).

Acknowledgements

This work was funded in part by the Bill and Melinda Gates Foundation (OPP1109789) and start-up funds from Tsinghua University to B.J. B.J. and T.F.Z. are Tsinghua-Janssen scholars. The authors thank S. Fortune, E. Rubin, M. Chao and P. Lehner for their critical reading of the manuscript, and C. Koser and P. Abel Zur Wiesch for helpful discussions. The authors also acknowledge technical assistance from the microbial sorting facility at Peking University, and thank the Clinical Database and Sample Bank of Tuberculosis of

Beijing (D131100005313012) of the National Clinical Lab on Tuberculosis, Beijing Chest Hospital, for access to their strain collection.

Author contributions

B.J. conceived and oversaw the design and implementation of the project. H.W.S., J.H.Z., C.E., X.W. and H.L. designed and performed the research and analysed the data. R.J.C. and Y.X.C. made and provided reagents. M.K., T.F.Z. and D.M. analysed whole-genome sequencing data. H.H., B.D.K. and B.J. analysed data. H.W.S., J.H.Z. and B.J. wrote the paper with input from the other authors.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.J.

Competing interests

The editors note that one of the individuals acknowledged for critical reading of the manuscript, M.C., as well as being a former colleague of the corresponding author, is an editor on the staff of *Nature Microbiology*, but was not in any way involved in the journal review process.