Main Manuscript for

Natural variation in the drug-binding residues of bacterial ribosomes

Chinenye L. Ekemezie[1,&], Lewis I. Chan[1,&], Charlotte R. Brown[1], Karla Helena-Bueno[1], Tom A. Williams[2,*], Sergey V. Melnikov[1,*]

[1] Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK
[2] School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK
[&] Contributed equally
*Corresponding authors

Email:
tom.williams@bristol.ac.uk, sergey.melnikov@newcastle.ac.uk

Author Contributions:
C.L.E., L.I.C. and S.V.M. conceived the study and performed the analysis; C.R.B. assisted with the automation. K.H.B. contributed to the analysis of rRNA variations in species with small genomes. All authors wrote the manuscript.

Competing Interest Statement:
The authors declare that they have no competing interests.

Classification:
Biological sciences

Keywords:
ribosome | ribosome-targeting antibiotics | 16S rRNA | 23S rRNA | ribosome evolution

This PDF file includes:
Main Text
Figures 1 to 4

Abstract

Drugs that target bacterial ribosomes are widely used in modern medicine and veterinary practice to treat bacterial infections and prevent the spread of antimicrobial resistance. However, most studies on targeting ribosomes with drugs are limited to just a few model organisms. Consequently, we do not know whether the ribosomal drug-binding sites observed in model bacteria are as highly conserved across bacteria as is currently implied. In this study, we address this question using a simple but powerful computational pipeline that filters out rare variants and sequencing errors to identify conserved changes at ribosomal drug-binding sites across the bacterial tree of life. This allows us to assess the conservation of 82 individual drug-binding residues of bacterial ribosomes from 8,809 bacterial species—and trace their evolution throughout >4 billion years of bacterial history. Contrary to the common belief in the high conservation of ribosomal drug-binding residues, we find extensive variation among bacterial phyla—and show that approximately 10% of species bear intrinsic ribosomal RNA (rRNA) substitutions that were previously observed only in clinical isolates of drug-resistant bacteria. Our work illustrates that our traditional use of *Escherichia coli* or other model bacteria to understand how bacterial ribosomes are targeted with drugs can be oversimplistic and misleading, as it overlooks widespread lineage-specific variation that make the drug-binding sites of some bacteria more dissimilar to *E. coli* than *E. coli* is to humans. These findings will have numerous implications for the lineage-specific use of ribosome-targeting antibiotics that are currently viewed as universal inhibitors of bacterial protein synthesis.

Significance Statement

Our understanding of how antibiotics target bacterial ribosomes is currently based on the use of a relatively small number of model species—such as *Escherichia coli* or *Thermus thermophilus*—whose ribosomes are widely viewed as generalizable model of bacterial ribosomes. Here, we expose the limitations of this approach and reveal the actual evolutionary diversity of ribosomal drug-binding residues across the entire domain of bacterial species. Our analysis identifies bacterial lineages with structurally dissimilar ribosomal drug-binding sites compared to those in commonly studied bacteria. This raises both concerns and opportunities for using ribosome-targeting antibiotics as lineage-specific rather than universal inhibitors of bacterial protein synthesis. We also present a generalizable, simple, yet effective strategy to assess the evolutionary conservation of any drug-binding site in large and complex datasets of biological sequences.

Main Text

Introduction

When streptomycin, the first ribosome-targeting antibiotic, was discovered in 1942 in the soil bacterium *Streptomyces griseus* (1), its mechanism of action and the molecular structure of ribosomes were unknown. However, in less than a decade, streptomycin and other ribosome-targeting antibiotics discovered throughout 1940s and 1950s became approved for broad use against microbial infections (2). This happened because these antibiotics could effectively inhibit many tested bacteria, including pathogenic species of *Mycobacterium*, *Staphylococcus*, *Salmonella*, *Aeromonas*, *Bacillus*, *Serratia*, *Pseudomonas*, *Clostridium*, and *Proteus* (2–5). Perhaps, in part, due to this success, the actual spectrum of species-specificity for ribosome-targeting antibiotics has never been fully tested. Instead, it was assumed that bacterial ribosomes are highly conserved, and we therefore could use a few model organisms to elucidate the general principles of ribosome drug targeting (6–9).

However, more recent studies revealed that ribosomal drug-binding residues are not strictly conserved across bacteria (10–18). For example, in the thermophilic bacterium *Thermus thermophilus* the naturally occurring substitutions U1782C/U2586C in the 23S rRNA (compared to *Escherichia coli*) confer resistance to tetracenomycin X (13). Similarly, strains of the opportunistic human pathogen *Propionibacteria* bear the naturally occurring 16S rRNA substitution G1491U which conferred intrinsic resistance to aminoglycosides (17). Similarly, some *Mycoplasma* species were reported to carry the A2057G substitution that confers intrinsic resistance to macrolides (18).

These examples showed that bacterial ribosomes can harbour naturally occurring substitutions that confer intrinsic resistance to ribosome-targeting drugs. However, without systematic studies, it remains unclear whether species such as *Thermus*, *Propionibacteria*, and *Mycoplasma* correspond to rare exceptions or rather common variants of ribosomal drug-binding residues in bacterial species. Consequently, we do not know the extent to which rRNA residues which are bound by drugs in model organisms such as *E. coli* are conserved across the bacterial tree of life. How many bacteria possess ribosomal drug-binding sites that are identical to those in *E. coli*? And how many exhibit dissimilarities to *E. coli* and instead resemble the intrinsically resistant species?

Here, we have addressed this question by creating a comprehensive catalogue of naturally occurring substitutions in ribosomal drug-binding residues across the entire bacterial domain of life. To make this analysis possible, we have established an approach that simplifies the analysis of large datasets of biological sequences by selecting a representative sequence for each species that is most similar to its closest neighbours on the tree of life. This approach helps to filter out sequencing or assembly errors, as well as private mutations of limited predictive value at the level of bacterial genera and higher-level clades. This allowed us to assess natural variation at the 82 individual drug-binding residues of the *E. coli* ribosome across 8,809 representative species from all major branches of bacteria.

Results

Evolution-based filtering reveals common variations in ribosomal drug-binding sites. To assess the conservation of drug-binding residues in bacterial ribosomes, we first set out to address the problem of false-positive changes in biological sequences. Previously, datasets of biological sequences, including those of rRNA, were shown to bear spurious mutations that stem from errors of sequencing, assembly or annotation of sequencing data (19, 20). In particular, since many biological sequences are assembled from multiple shorter reads, public repositories

were estimated to contain up to 20% chimeric sequences in which different parts of a given sequence originate from different species instead of a single species. Other common errors include sequencing errors, contamination with DNA from plasmids and viruses, misannotation of species, and using clinical isolates instead of natural strains. As a result, many bacterial species have hundreds of alternative variants of a single gene, making it difficult to distinguish conserved changes fixed in a lineage from rare mutations and sequence errors.

Here we addressed this problem using the public repository of rRNA sequences known as the SILVA dataset NR99 v138.1, which the most complete dataset of non-redundant rRNA sequences comprising 510,508 sequences for 16S rRNA and 95,286 sequences for 23S rRNA (21). We first simplified the SILVA dataset by discarding all rRNA sequences of unclear evolutionary origin, resulting in the removal of approximately two-thirds of the sequences in the original dataset. Next, we excluded sequences that originate from phages, plasmids, and eukaryotic hosts of pathogenic bacteria, as well as those with truncations in ribosomal active centres (see **Methods**). However, even the simplified dataset contained up to 800 alternative rRNA sequences for a single species (**SI_Data 2,3**). For example, for *E. coli* alone, we observed 39 dissimilar variants of the ribosomal drug-binding residues, which stemmed in part from combining sequences from naturally occurring and drug-resistant clinical strains (**Fig 1a**). This illustrates a common obstacle in analysing the evolutionary variations of drug-binding sites across large groups of species.

Therefore, we next further simplified our analysis by using the evolutionary proximity of species on the tree of life. In this approach, we further reduced the SILVA dataset by selecting one representative rRNA sequence per species. To achieve this reduction for each species, we selected the rRNA sequence that was most similar to sequences in the most closely related species in the tree of life (**Fig 1b**). Although this approach discarded sequence variants that were specific to a single sequence or strain of a given species, it minimized the risk of detecting false positives and revealed the most conserved sequence variants in each bacterial genus.

Using this approach, we then identified the most common variants of 72 rRNA residues that were previously shown to directly recognize ribosome-targeting drugs (see **Table S1**). In this analysis, we focused only on those residues that were previously shown to recognize drugs in the sequence-depending fashion, via the contacts with the aromatic bases of rRNA (**Table S1**). For each of these residues, we have traced their evolutionary history and assessed their conservation in 8,809 bacterial species. We have summarized our findings in (**SI_Data 4,5**) to provide the most complete information about the conservation of drug-binding sites identified in model species across bacterial diversity—at the level of single nucleotide, single species, and single ribosome-targeting drug.

Variable residues occur in each drug-binding site of bacterial ribosomes. We next asked what are the most and least conserved drug-binding sites in bacterial ribosomes. To answer this, we abandoned the conventional approach where the conservation of rRNA residues is assessed by comparing bulk sets of rRNA sequences, without separating these sequences into different phylogenetic groups (14, 22–24). Instead, we separated rRNA sequences into individual phyla and determined representative profiles of ribosomal drug-binding residues for each phylum (**Figs 2, 3; SI_Data 4,5**). This approach allowed us to assess conservation without the risk of bias toward overrepresented phyla.

We found that our approach gave results that differed from previous estimates (14, 22–24). Earlier estimates made using bulk sets of rRNA sequences suggested that most rRNA residues in the active sites of bacterial ribosomes are highly conserved, with the exception of the decoding center (which is the binding site for aminoglycoside antibiotics). However, our analysis indicates that this phylogeny-unaware approach can be highly misleading because proteobacterial

sequences are highly over-represented in reference databases; that is, reference databases are enriched for sequences that are similar to the *E. coli* model system merely due to evolutionary relatedness rather than functional conservation. Furthermore, some species in this dataset are represented by over 800 sequences, while others by just 20 sequences per entire phylum (**Fig 3**). This creates a strong bias in average conservation scores towards rRNA conservation in Proteobacteria rather than across all bacteria.

By contrast, we showed that when rRNAs from different phyla are compared in a way that treats species and phyla more evenly, they reveal high variability of ribosomal drug-binding residues in bacteria (**Fig 2, 3**). Thus, we found that drug-binding sites of the ribosome are not highly conserved across phyla; instead, variations in rRNA sequence fixed at the species level are present in every known binding site of the ribosome.

Bacterial phyla have dissimilar ribosomal drug-binding sites. We then mapped the identified variations onto the tree of life to assess their evolutionary origin and distribution among species (**Fig 3**). This mapping revealed that bacteria likely began to diversify their ribosomal drug-binding residues before their separation into the modern phyla. The earliest substitution likely occurred at least 3.8 billion years ago (Bya) when the branches of Dictyoglomi and Thermotogae separated from other bacteria, bearing the macrolide-binding residue G746 in the 23S rRNA, while the remaining bacteria (specifically of Terrabacteria) bore U746. Subsequently, bacterial ribosomes continued to diversify their drug-binding residues by accumulating lineage-specific substitutions in 28 rRNA bases that directly bind with ribosome-targeting drugs. As a result, contrary to the idea of highly conserved ribosomal drug-binding sites across bacterial species, we showed that each bacterial phylum has evolved an idiosyncratic set of variations in their ribosomal drug-binding residues (**Fig 3**).

Compared to *E. coli*, the most dissimilar drug-binding residues we observed among the members of Firmicutes and Actinobacteria. Within these phyla, many lineages have acquired ribosomal drug-binding residues that are more different to those of *E. coli* than the differences observed between *E. coli* and humans. This can be seen, for example, in the order Propionibacteriales (Actinobacteria), which includes causative agents of skin infections as well as species used in industrial food processing (**SI_Data 5, 6**). Compared to *E. coli*, these species exhibit up to 15 substitutions in their drug-binding residues. These substitutions are distributed across both ribosomal subunits, particularly in the binding sites for pactamycin, neomycin, thermorubin, and other antibiotics. Thus, some bacterial phyla exhibit highly divergent drug-binding residues compared to those in *E. coli*, whose ribosomes are commonly used as a generalized model of bacterial ribosomes.

We also found that many residues considered representative of bacteria based on structural studies of *E. coli* ribosomes are, in fact, highly atypical for most bacterial species. For example, the non-canonical base pair U1782-U2586 in the 23S rRNA of *E. coli* ribosomes has been shown to recognize the antibiotic tetracenomycin X and confer resistance when mutated to the C1782-C2586 pair. Additionally, *T. thermophilus* ribosomes were shown to naturally carry the C1782-C2586 pair and do not bind tetracenomycin X in co-crystallization experiments (13). Compared to these previous studies, our analysis revealed that the U1782-U2585 pair is highly atypical for species other than proteobacteria, with the majority of species (over 80% of representative organisms) carrying the C1782-C2586 pair. Thus, we showed that some *E. coli* rRNA residues, traditionally viewed as characteristic of bacteria, can be highly atypical for species outside of proteobacteria.

Furthermore, we found that many bacterial phyla have substitutions even in those rRNA bases which are traditionally viewed as universally conserved. For instance, the pactamycin-binding site of the ribosomes is formed by the 16S rRNA residues G693, A694, and A787 that are conserved

between the model bacteria, including *T. thermophilus* and *E. coli*, and model eukaryotes, including yeasts and humans (23, 25). Consistent with this conservation, pactamycin acts as a universal inhibitor of protein synthesis by targeting both bacterial and eukaryotic ribosomes. Previously, substitutions in the pactamycin-binding site have not been observed in nature and were only identified in the archaeon *Halobacterium halobium* that acquired the A694G substitution during experimental laboratory evolution of pactamycin resistance (26). By contrast, we found that variations in the pactamycin-binding site not only occur in nature, but are characteristic of certain Firmicutes, Actinobacteria, Chloroflexi, Spirochaetes and Chlamydiota. In total, we identified natural variations at 15 drug-binding residues that have the same nucleotide state in *E. coli* and humans (**Figure 2**). Thus, we found that many drug-binding residues, which are typically regarded as universally conserved and targeted by "universal inhibitors" of protein synthesis, are, in fact, naturally modified across clades as large as entire phyla.

Overall, our evolutionary analysis showed that bacteria have been diversifying their ribosomal drug-binding residues since the emergence of the first superphyla, leading to highly divergent and phyla-specific variants of drug-binding residues in bacterial ribosomes.

Even closely related bacteria can have dissimilar ribosomal drug-binding sites. Our further analysis showed that, after the separation into phyla, bacteria continued to extensively diversify their ribosomal drug-binding residues. As a result, in addition to phylum-specific substitutions, many smaller clades have accumulated numerous idiosyncratic substitutions in the drug-binding residues (**Fig 3, SI_Data 4,5**).

As might be expected, the highest number of these substitutions were found in pathogenic bacteria with small genomes, particularly *Rickettsia*, *Helicobacter, Bartonella*, *Treponema* and *Mycoplasma* (**Fig 4, SI_Data 4, 5**). Previously, these organisms were shown to accumulate a high number of genetic changes and evolve more rapidly than non-parasitic bacteria due to weaker natural selection associated with parasitic lifestyles (27–29). Here, we found that this rapid evolution extends to the drug-binding residues of the ribosome, leading to significant differences even for species from the same genus. For example, *Mycoplasma* species contain up to 10 substitutions of drug-binding residues compared to each other, making some of these species more dissimilar to each other than they are to *E. coli* from the perspective of drug-binding residues (**Fig 4**).

We also found that some of these variations correlate with the previously observed differences in antibiotic sensitivity among *Mycoplasma* species (18, 30, 31). Specifically, *Mycoplasma pneumoniae*, which has the base G2057 in the macrolide-binding site of the ribosome, has been characterized as highly sensitive to macrolides. For instance, erythromycin can completely inhibit the growth of this species at concentrations as low as just 7 ng/ml. By contrast, *Mycoplasma hominis* (bearing the A2057) and *Mycoplasma fermentas* (A2057) are 5,000 and 18,000-fold more tolerant to erythromycin and exhibit intrinsic resistance to this and other 14- and 15-membered macrolide antibiotics.

Our analysis showed that *Mycoplasma* and other parasitic bacteria tend to accumulate substitutions in the G2057-C2611 and other drug-binding base pairs which often disrupt canonical base pairing. For example, in *Rickettsia and Neorickettsia* species, the aminoglycoside-binding site shows a substitution of the base pair C1409-G1491 to A-G, G-U, or C-A pairs, pointing to substantial structural changes in this drug-binding site of the ribosome. Furthermore, these substitutions are not limited to base-paired residues and involve non-paired residues that use their aromatic bases to form hydrogen bonds with drugs. These include, for example, the pactamycin-binding residues G693 and A694 in the 16S rRNA or the ketolide-binding residues U746 and G748 in the 23S rRNA. Overall, these rRNA variations do not have a particular preference for a single drug binding site; instead, they occur at multiple drug-binding sites within

both ribosomal subunits, including the binding sites for edeine, pactamycin, spectinomycin, thermorubin, and other antibiotics (**Fig 4**).

This finding is remarkable because it challenges the traditional view that phylogenetic distance serves as a reliable predictor of similarity in ribosomal drug-binding sites (15, 23, 32, 33). Our analysis shows that this notion is particularly problematic when applied to pathogenic bacteria with small genomes. These examples of *Mycoplasma* and other pathogens illustrate that species do not necessarily need to be distantly related on the tree of life to display highly dissimilar ribosomal drug-binding sites. Instead, such pronounced dissimilarity can occur even within the same genus, perhaps driven by genetic drift or selection to evade the antimicrobial defences of other organisms in the environment.

Variants described as resistance-conferring in model species are widespread in nature. Finally, we assessed the extent to which the naturally occurring variations identified in our analysis have been previously characterized as resistance-conferring in model bacteria. Our literature review revealed that substitutions of residues directly involved in recognizing ribosome-targeting drugs typically result in a 10- to 80,000-fold increase in tolerance to the corresponding drug (**Tables S5, S6**). Among the most rigorous data were obtained for the 16S rRNA residue C1192, which basepairs with G1064, an interaction that is recognized by the antibiotic spectinomycin (34). Previous studies showed that spectinomycin binding strictly requires the presence of G at position 1064 and C at position 1192 because mutating either of these residues results in up to a 1,000x higher spectinomycin resistance (35–37). Importantly, these high levels of resistance were observed regardless of whether the substitutions maintained base pairing (e.g., C1064-G1192 or A1064-U1192) or not (e.g., U1064-U1192 or G1064-G1192) (35–37). Furthermore, the effect of these substitutions is independent of the organism's identity: for example, the same substitutions, C1192U or C1192G, resulted in comparably high levels of spectinomycin resistance in four different bacteria, including *E. coli*, *Pasteurella multocida*, *Salmonella enterica*, and *Borrelia burgdorferi* (35–37).

However, our analysis revealed that the C1192U and C1192A variants are not restricted to the previously obtained laboratory strains but frequently occur in nature. Particularly, they are common for most members of the Firmicutes, Actinobacteria, and Chloroflexi phyla, suggesting intrinsic spectinomycin resistance in these species (**Fig 3, SI_Data 4,5**). Aside from these C1192 variations, we found that natural variations in 24 drug-binding residues have been characterized as resistance-conferring in model bacterial species (**Tables S5, S6**). Thus, we found that mutations previously thought to be confined to laboratory-engineered strains and clinical isolates of pathogenic bacteria are, in fact, widespread in nature, particularly outside commonly studied lineages such as the Proteobacteria.

Discussion

Variation of ribosomal drug-binding residues in the bacterial domain of life. Prior to our study, the conservation of ribosomal drug-binding residues had never been assessed systematically and instead was estimated using the following two approaches. First, for certain rRNA residues, conservation estimates were derived from sequence alignments, where rRNA sequences were simply grouped into bacterial and eukaryotic, without further separation into phyla and adjustments for sample size (6, 7, 12, 14, 16, 22, 24). While these analyses were instrumental in identifying immutable rRNA bases or major differences between bacteria and eukaryotes, they were heavily biased toward the most common species and phyla, such as Proteobacteria. Alternatively, in a more accurate approach, the variations of drug-binding residues were determined by comparing ribosome structures from different species, including *Deinococcus radiodurans* and *Haloarcula marismortui* (11, 12, 33), or *E. coli* and yeasts or

humans (23, 25). These analyses provided detailed information about the diversity of ribosomal drug-binding residues, but for only a handful of model organisms. As a result, we lacked a comprehensive understanding of the variability of ribosomal drug-binding residues across the bacterial domain of life.

In our work, we addressed this knowledge gap by elucidating the evolutionary variations of individual rRNA drug-binding residues across more than 8,000 representative bacteria. For each drug-binding residue, we traced its evolutionary history from the last bacterial common ancestor to the present day, enabling us to estimate variations not only in currently known species but also in those yet to be discovered, based on their position on the tree of life. Overall, our analysis revealed that bacterial phyla exhibit a great diversity of ribosomal drug-binding residues. This illustrates that the occasional findings of "unusual" ribosome variants in species like *Propionibacterium*, *Thermus*, or *Mycoplasma* are more of a common rule than rare exceptions.

Importantly, this strategy for the analysis of rRNA sequences can be repurposed for any other drug-binding protein or nucleic acid. Currently, the analysis of drug targets is largely dominated by structural studies in a specific model organism, with less attention given to the evolutionary conservation of drug-binding sites across species. This status quo stems in part from the ever-increasing size of sequencing data, which complicates their analyses and separation of useful information from sequencing and annotation mistakes. Our work provides a solution to this problem and helps simplify extremely large datasets, making their analysis feasible, rapid and human-friendly. We anticipate that this approach will become increasingly important due to the exponential accumulation of biological sequences and the need to simplify their size without the loss of essential information.

Compared to *E. coli*, most phyla bear multiple substitutions in ribosomal drug-binding sites. Our work illustrates the limitations of using individual organisms, including *E. coli*, as generalized models of bacterial ribosomes. Specifically, we show that many drug-binding residues of *E. coli* ribosomes, currently viewed as representative and conserved (9, 34, 38), are in fact highly variable across the natural diversity of bacteria. Furthermore, we show that *E. coli* exhibits relatively atypical drug-binding sites for the antibiotics quinupristin, tetracenomycin X, and evernimycin due to the 23S rRNA variations (U1782, U2533, and U2586) that are found in Proteobacteria, but not in most other bacteria. This finding shows that our commonly used division of ribosomes into bacterial and eukaryotic types (11, 23, 39) is incomplete and even misleading, given the substantial diversity of ribosomal drug-binding residues within the bacterial domain of life.

Implications for clinical practice and lineage-specific use of ribosome-targeting drugs. Previously, the discovery of intrinsic resistance in *Propionibacteria* and *Mycoplasma* to certain ribosome-targeting antibiotics has changed the clinical targeting of these human pathogens (15, 17, 18, 30). For example, *M. pneumoniae*, a respiratory mycoplasma, is effectively treated with macrolides, lincosamides, streptogramins, and ketolides. However, *M. hominis*, a genital mycoplasma, is no longer treated with 14- and 15-membered macrolides and ketolides due to its natural resistance (A2058G substitution in the 23S rRNA)—and instead is treated with josamycin and lincosamides (18, 30). Thus, the knowledge of on structural diversity of ribosomal drug-binding sites in a small number of species has led to a paradigm shift from using ribosome-targeting antibiotics in an indiscriminate fashion as universal inhibitors of bacterial protein synthesis to a more accurate selection of ribosome-targeting antibiotics for a given pathogen.

Based on our findings, we anticipate a similar revision for a much broader range of bacterial species. Our work showed that approximately one in ten representative bacteria carry rRNA substitutions that have been observed in clinical isolates of drug-resistant human pathogens

(SI_Data 4,5). This finding may have important implications for clinical practice because our findings suggest that the intrinsic resistance previously observed in *Mycoplasma*, *Propionibacterium* and *Thermus* species is likely much more prevalent among non-model bacteria due to the presence of identical or even more structurally divergent drug-binding sites in these species compared to *E. coli*. Our evolutionary analyses have two caveats worth considering. The first is that the phenotypic impact of the observed changes at drug-binding sites might, in some cases, be compensated by changes at other sites close in the three-dimensional structure of the molecule; thus, the precise effects on drug resistance of the lineage-specific variants we identified will require further study. Second, our analysis is - inevitably - focused on variation at drug-binding sites that have been characterised in *E. coli*. Given the extensive variation we have documented at these sites, it seems reasonable to suggest that the position of drug-binding residues is also likely to change over evolutionary time. Identifying new drug-binding sites in divergent bacterial lineages might be achieved in future work through experimental characterisation or, perhaps, computational docking analyses (40). Nonetheless, our work presents the first comprehensive overview of the structural variability of ribosomal drug-binding sites in bacteria which may inform a more accurate and personalized choice of ribosome-targeting drugs for a given pathogen.

Furthermore, our study showed that ribosomal drug-binding sites in many bacterial clades, including pathogenic bacteria, are more distinct from those in *E. coli* than in *E. coli* compared to humans. This finding is particularly remarkable considering that substitutions of rRNA nucleotides can be sufficient for species-specificity of ribosome recognition by antibiotics (41). For example, previous mechanistic studies have revealed that the substitutions A1408G and G1491A in the 16S rRNA of bacterial ribosomes compared to those in eukaryotes allow to use selectively target bacterial ribosomes with aminoglycosides while causing minimal to no harm to animals (23, 25, 42). Similarly, the substitutions A2055C and U2572A in the bacterial (compared to eukaryotic) 23S rRNA render species-specific ribosome targeting with the “A-site cleft” inhibitors, including tiamulin in bacteria and bruceantin in eukaryotes (16). Therefore, our finding that bacterial species possess numerous rRNA substitutions compared to each other suggests that species-specific targeting of ribosomes is plausible not only at the level of domains of life (bacteria vs. eukaryotes) but also at the level of bacterial phyla and even smaller bacterial clades. We anticipate that these findings will pave the way for the development of more precise ribosome-targeting antibiotics that target specific pathogenic bacteria with minimal harm to commensal microbes.

Materials and Methods

Annotating ribosomal drug-binding residues. To annotate the drug-binding residues of the ribosome, we used the Protein Data Bank (43) to retrieve previously determined ribosome structures in complex with 35 clinically relevant drugs (34, 44–67), which corresponded to 21 chemical families of ribosome-targeting molecules (**Table S1**). We then used PyMOL (version 2.5.0) (68) to select ribosomal drug-binding residues located within hydrogen bond distance from the drug, with a distance cut-off of 3.3 Å for structures containing hydrogen atoms and 4.3 Å for structures lacking hydrogen atoms. To identify ribosomal residues that bind ribosome-targeting drugs in a sequence-dependent manner, we then excluded residues that bind drugs with either their sugar moiety or phosphate backbone of rRNA nucleotides. This allowed us to select only those rRNA residues that bind drugs using their nucleotide bases (**Table S2**).

Creating a high-fidelity dataset of rRNA sequences for evolutionary analyses. To enable accurate analysis of rRNA sequences, we have used the public repository of rRNA sequences, SILVA, particularly the dataset NR99 v138.1 that contained 510,508 non-redundant 16S rRNA sequences corresponding to 40,239 species (17,609 genera) ("SSU dataset") and 95,286 23S rRNA sequences from 14,849 species (7,837 genera) ("LSU dataset") from all domains of life

(21). To filter out false-positive sequence variations, we have devised a 10-step data reduction strategy to bypass the errors in sequence datasets that stem from automated annotation of sequencing data and sequencing mistakes, such as chimeric sequences (where a single sequence is assembled from fragments from different species), sequences with misannotated phylogeny, or sequences from non-genomic DNA and pseudogenes (**Table S3 and S4**). In this reduction strategy, the initial SILVA datasets were first reduced by eliminating sequences from non-bacterial species as well as from unidentified bacteria for which we could not confirm the identity of an organism. This included sequences whose headers contained the words "unidentified", "uncultured", "metagenome", "cluster" and “sp.”, as well as sequences whose names began with a lowercase letter. The obtained datasets were then further reduced by eliminating rRNA sequences that are encoded by plasmids and viruses. We then removed sequences of apparently poor sequencing quality, based on the presence of 'ambiguous' nucleotides represented by symbols such as "R", "X", "N", and "S" instead of the standard set of RNA bases “A”, “G”, “C”, and “U”. This resulted in the 23S rRNA dataset containing 56,154 sequences from 5,726 bacterial species (from 1,678 genera) and the 16S rRNA dataset containing 94,486 sequences from 14,132 of bacterial species (from 2,967 genera) (**SI_Data S2 and S3**).

Assessing of drug-binding residues in bacterial rRNA. To enable the analysis of rRNA conservation across bacterial species, the reduced sequence libraries of 16S and 23S rRNA were used for multiple sequence alignments. To reduce gaps and alignment errors, the datasets were first separated into sequences from individual bacterial phyla, and then the multiple sequence were produced using MAFFT v7.4908 (69) with default settings (FFT-NS-2) for each phylum.

To assess the conservation of drug-binding residues, the aligned sequences of 16S and 23S rRNA were reduced to the sequences of drug-binding residues only (**SI_Data S4, S5**). To remove potential “false-positive” sequence variations that originate from truncations, the errors of sequencing, alignment or genome assembly, we removed sequences that contained truncations in the ribosomal drug-binding sites.

The resulting dataset was further reduced to a single sequence per organism using the following rule: whenever the same organism contained several sequence variants, we used only one variant, which was identical (or the most similar) to the sequences of the adjacent members on the tree of life. This allowed us to identify mutations that were a common characteristic of a clade of species and eliminate sequencing and annotation artifacts, as well as mutations that were characteristic of an individual strain (**SI_Data S4, S5**).

Mapping the conservation of bacterial ribosomal drug-binding residues on ribosome structure. To highlight conservation of the drug-binding residues in the structure of the bacterial ribosome, we used ChimeraX v1.6.19. First, we generated the profile sequence of the ribosomal drug-binding residues for each bacterial phylum in our dataset (shown in **SI_Data 5,6**). We then aligned the profile sequences and calculated the percentage of conservation for each drug-binding residue in the aligned phyla-specific profile sequences. The resulting scores were then shown in (**Figure 2**).

Assessing the conservation of drug-binding residues in rRNA operons of the same species. To assess potential variability of rRNA drug-binding residues between rRNA operons from the same species, we have repeated our analysis of rRNA-binding residues by using the manually curated dataset of rRNA operons rrnDB (version 5.8) (70) instead of the SILVA dataset. Our analysis showed that the drug-binding residues of rRNA encoded by different operons of the same species remain conserved, thereby excluding the possible variability of antimicrobial resistance between rRNA produced by different operons of the same bacterial species.

Tracing the evolutionary origin of variations in bacterial ribosomal drug-binding sites. To trace the evolutionary history of variations in the ribosomal drug-binding residues, we used 16S rRNA sequences listed in (**SI_Data 5,6**), we used the dated bacterial tree generated using Mega11 with default settings (71), rooting the tree based on recent analyses (72, 73). The relative age of each mutation was estimated based on a published bacterial timetree (73) or the original studies of *Mycoplasma* (74, 75), *Helicobacter* (27, 76) and *Rickettsiales* (77).

Acknowledgments

We thank members of Melnikov lab for their help in the preparation of this manuscript and Heath Murray, Katarzyna Mickiewicz, Robert Hirt and Claudia Schneider (Newcastle University, UK) for their critical comments. This work was funded by the MRC Discovery Medicine North Doctoral Training Partnership (MR/N013840/1 to C.L.E.), the BBSRC UK Doctoral Training Partnership (BB/T008695/1 to C.R.B. and L.I.C.), the Newcastle University NUORS 2021 Award to K.H-B., and the Royal Society (RGS/R2/202003) to S.V.M.

References

1. A. Schatz, E. Bugle, S. A. Waksman, Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. *Experimental Biology and Medicine* **55**, 66–69 (1944).
2. P. R. Burkholder, Antibiotics: The exploitation of microbial antagonisms is having a challenging impact on medicine and society. *Science* **129**, 1457–1465 (1959).
3. S. A. Waksman, H. A. Lechevalier, Neomycin, a new antibiotic active against streptomycin-resistant bacteria, including tuberculosis organisms. *Science* **109**, 305–307 (1949).
4. R. R. Willcox, Orally administered antibiotics and the venereal diseases. *Nature* 466–468 (1950).
5. T. H. Haight, M. Finland, Laboratory and clinical studies on erythromycin. *New England Journal of Medicine* **247**, 227–232 (1952).
6. D. E. Brodersen, *et al.*, The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. *Cell* **103**, 1143–1154 (2000).
7. A. P. Carter, *et al.*, Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**, 340–348 (2000).
8. D. N. Wilson, Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol* **12**, 35–48 (2014).
9. H. Paternoga, *et al.*, Structural conservation of antibiotic interaction with ribosomes. *Nat Struct Mol Biol* 1–13 (2023). https://doi.org/10.1038/s41594-023-01047-y.
10. J. Harms, *et al.*, High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679–688 (2001).
11. F. Schlünzen, *et al.*, Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
12. D. Tu, G. Blaha, P. B. Moore, T. A. Steitz, Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell* **121**, 257–270 (2005).
13. I. A. Osterman, *et al.*, Tetracenomycin X inhibits translation by binding within the ribosomal exit tunnel. *Nature Chemical Biology* **16**, 1071–1077 (2020).
14. H. David-Eden, A. S. Mankin, Y. Mandel-Gutfreund, Structural signatures of antibiotic binding sites on the ribosome. *Nucleic Acids Research* **38**, 5982–5994 (2010).
15. S. Akshay, *et al.*, Phylogenetic sequence variations in bacterial rRNA affect species-specific susceptibility to drugs targeting protein synthesis. *Antimicrobial Agents and Chemotherapy* **55**, 4096–4102 (2011).
16. G. Gürel, G. Blaha, P. B. Moore, T. A. Steitz, U2504 determines the species specificity of the a-site cleft antibiotics: The structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. *Journal of Molecular Biology* **389**, 146–156 (2009).
17. U. Hoeffler, H. L. Ko, G. Pulverer, Antimicrobial susceptibility of *Propionibacterium acnes* and related microbial species. *Antimicrob Agents Chemother* **10**, 387–394 (1976).

18. S. Pereyre, *et al.*, Mutations in 23S rRNA account for intrinsic resistance to macrolides in Mycoplasma hominis and Mycoplasma fermentans and for acquired resistance to macrolides in M. hominis. *Antimicrob Agents Chemother* **46**, 3142–3150 (2002).
19. K. E. Ashelford, N. A. Chuzhanova, J. C. Fry, A. J. Jones, A. J. Weightman, At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**, 7724–7736 (2005).
20. J. S. Johnson, *et al.*, Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature communications* **10**, 5029 (2019).
21. C. Quast, *et al.*, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012).
22. C. R. Woese, *et al.*, Conservation of primary structure in 16S ribosomal RNA. *Nature* **254**, 83–86 (1975).
23. N. Garreau De Loubresse, *et al.*, Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513**, 517–522 (2014).
24. S. Smit, J. Widmann, R. Knight, Evolutionary rates vary among rRNA structural elements. *Nucleic Acids Research* **35**, 3339–3354 (2007).
25. I. Prokhorova, *et al.*, Aminoglycoside interactions and impacts on the eukaryotic ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **114** (2017).
26. A. S. Mankin, Pactamycin resistance mutations in functional sites of 16 S rRNA. *Journal of molecular biology* **274**, 8–15 (1997).
27. J.-F. Tomb, *et al.*, The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* **388**, 539–547 (1997).
28. J. O. Andersson, S. G. Andersson, Genome degradation is an ongoing process in Rickettsia. *Molecular biology and evolution* **16**, 1178–1191 (1999).
29. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* **10**, 13–26 (2012).
30. P. M. Furneri, G. Rappazzo, M. P. Musumarra, G. Tempera, L. S. Roccasalva, Genetic basis of natural resistance to erythromycin in Mycoplasma hominis. *Journal of Antimicrobial Chemotherapy* **45**, 547–548 (2000).
31. T. Sato, H. Higuchi, S. Yokota, Y. Tamura, Mycoplasma bovis isolates from dairy calves in Japan have less susceptibility than a reference strain to all approved macrolides associated with a point mutation (G748A) combined with multiple species-specific nucleotide alterations in 23S rRNA. *Microbiology and Immunology* **61**, 215–224 (2017).
32. D. N. Wilson, J. M. Harms, K. H. Nierhaus, F. Schlünzen, P. Fucini, Species-specific antibiotic-ribosome interactions: implications for drug development. *FASEB Journal* **386**, 1239–1252 (2005).
33. G. M. Blaha, Y. S. Polikanov, T. A. Steitz, Elements of ribosomal drug resistance and specificity. *Curr Opin Struct Biol* **22**, 750–758 (2012).
34. M. A. Borovinskaya, S. Shoji, J. M. Holton, K. Fredrick, J. H. D. Cate, A steric block in translation caused by the antibiotic spectinomycin. *ACS Chem. Biol.* **2**, 545–552 (2007).
35. C. Kehrenberg, S. Schwarz, Mutations in 16S rRNA and ribosomal protein S5 associated with high-level spectinomycin resistance in Pasteurella multocida. *Antimicrobial agents and chemotherapy* **51**, 2244–2246 (2007).
36. M. O'connor, A. E. Dahlberg, Isolation of spectinomycin resistance mutations in the 16S rRNA of Salmonella enterica serovar Typhimurium and expression in Escherichia coli and Salmonella. *Current microbiology* **45**, 0429–0433 (2002).
37. M. F. Brink, G. Brink, M. Verbeet, H. A. de Boer, Spectinomycin interacts specifically with the residues G 1064 and C 1192 in 16S rRNA, thereby potentially freezing this molecule into an inactive conformation. *Nucleic acids research* **22**, 325–331 (1994).
38. M. A. Borovinskaya, *et al.*, Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nature structural & molecular biology* **14**, 727–732 (2007).
39. S. Melnikov, *et al.*, One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* **19**, 560–567 (2012).
40. S. Forli, *et al.*, Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols* **11**, 905–919 (2016).
41. K. L. Triman, A. Peister, R. A. Goel, Expanded versions of the 16S and 23S ribosomal RNA mutation databases (16SMDBexp and 23SMDBexp). *Nucleic acids research* **26**, 280–284 (1998).
42. H. Fan-Minogue, D. M. Bedwell, Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. *RNA* **14**, 148–157 (2008).
43. H. M. Berman, *et al.*, The protein data bank. *Nucleic acids research* **28**, 235–242 (2000).

44. H. Demirci, *et al.*, A structural basis for streptomycin-induced misreading of the genetic code. *Nat Commun* **4**, 1355 (2013).
45. Y. S. Polikanov, *et al.*, Amicoumacin A inhibits translation by stabilizing mRNA interaction with the ribosome. *Molecular cell* **56**, 531–540 (2014).
46. M. Shalev-Benami, *et al.*, Atomic resolution snapshot of Leishmania ribosome inhibition by the aminoglycoside paromomycin. *Nat Commun* **8**, 1589 (2017).
47. E. Svidritskiy, C. Ling, D. N. Ermolenko, A. A. Korostelev, Blasticidin S inhibits translation by trapping deformed tRNA on the ribosome. *Proceedings of the National Academy of Sciences* **110**, 12283–12288 (2013).
48. C. Davidovich, *et al.*, Induced-fit tightens pleuromutilins binding to ribosomes and remote interactions enable their selectivity. *Proceedings of the National Academy of Sciences* **104**, 4291–4296 (2007).
49. N. B. Olivier, *et al.*, Negamycin induces translational stalling and miscoding by binding to the small subunit head domain of the Escherichia coli ribosome. *Proceedings of the National Academy of Sciences* **111**, 16274–16279 (2014).
50. A. I. Cocozaki, *et al.*, Resistance mutations generate divergent antibiotic susceptibility profiles against translation inhibitors. *Proceedings of the National Academy of Sciences* **113**, 8188–8193 (2016).
51. M. A. Borovinskaya, *et al.*, Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nature structural & molecular biology* **14**, 727–732 (2007).
52. K. Tsai, *et al.*, Structural basis for context-specific inhibition of translation by oxazolidinone antibiotics. *Nature structural & molecular biology* **29**, 162–171 (2022).
53. M. A. Borovinskaya, S. Shoji, K. Fredrick, J. H. Cate, Structural basis for hygromycin B inhibition of protein biosynthesis. *Rna* **14**, 1590–1599 (2008).
54. A. Schedlbauer, *et al.*, Structural characterization of an alternative mode of tigecycline binding to the bacterial ribosome. *Antimicrobial agents and chemotherapy* **59**, 2849–2854 (2015).
55. J. L. Hansen, T. M. Schmeing, P. B. Moore, T. A. Steitz, Structural insights into peptide bond formation. *Proceedings of the National Academy of Sciences* **99**, 11670–11675 (2002).
56. T. M. Schmeing, K. S. Huang, D. E. Kitchen, S. A. Strobel, T. A. Steitz, Structural insights into the roles of water and the 2′ hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Molecular cell* **20**, 437–448 (2005).
57. D. Nicholson, T. A. Edwards, A. J. O'Neill, N. A. Ranson, Structure of the 70S ribosome from the human pathogen Acinetobacter baumannii in complex with clinically relevant antibiotics. *Structure* **28**, 1087–1100 (2020).
58. J. A. Dunkle, L. Xiong, A. S. Mankin, J. H. Cate, Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proceedings of the national academy of sciences* **107**, 17152–17157 (2010).
59. S. Arenz, *et al.*, Structures of the orthosomycin antibiotics avilamycin and evernimicin in complex with the bacterial 70S ribosome. *Proceedings of the National Academy of Sciences* **113**, 7527–7532 (2016).
60. J. Noeske, *et al.*, Synergy of streptogramin antibiotics occurs independently of their effects on translation. *Antimicrobial agents and chemotherapy* **58**, 5269–5279 (2014).
61. F. Schluenzen, *et al.*, The antibiotic kasugamycin mimics mRNA nucleotides to destabilize tRNA binding and inhibit canonical translation initiation. *Nature structural & molecular biology* **13**, 871–878 (2006).
62. D. Bulkley, F. Johnson, T. A. Steitz, The antibiotic thermorubin inhibits protein synthesis by binding to inter-subunit bridge B2a of the ribosome. *Journal of molecular biology* **416**, 571–578 (2012).
63. L. Zhang, *et al.*, The structural basis for inhibition of ribosomal translocation by viomycin. *Proceedings of the National Academy of Sciences* **117**, 10271–10277 (2020).
64. J. L. Hansen, *et al.*, The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Molecular cell* **10**, 117–128 (2002).
65. R. E. Stanley, G. Blaha, R. L. Grodzicki, M. D. Strickler, T. A. Steitz, The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nature structural & molecular biology* **17**, 289–293 (2010).
66. J. M. Harms, *et al.*, Translational regulation via L11: molecular switches on the ribosome turned on and off by thiostrepton and micrococcin. *Molecular cell* **30**, 26–38 (2008).
67. G. Gürel, G. Blaha, P. B. Moore, T. A. Steitz, U2504 determines the species specificity of the a-site cleft antibiotics: The structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. *Journal of molecular biology* **389**, 146–156 (2009).
68. Schrodinger, The PyMOL Molecular Graphics System, Version 1.8. (2015).
69. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772–780 (2013).
70. J. A. Klappenbach, P. R. Saxman, J. R. Cole, T. M. Schmidt, rrndb: the ribosomal RNA operon copy number database. *Nucleic acids research* **29**, 181–184 (2001).

71. K. Tamura, G. Stecher, S. Kumar, MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution* **38**, 3022–3027 (2021).
72. G. A. Coleman, *et al.*, A rooted phylogeny resolves early bacterial evolution. *Science* **372**, eabe0511 (2021).
73. A. A. Davín, *et al.*, An evolutionary timescale for Bacteria calibrated using the Great Oxidation Event. (2023). https://doi.org/10.1101/2023.08.08.552427.
74. C. M. Fraser, *et al.*, The Minimal Gene Complement of *Mycoplasma genitalium*. *Science* **270**, 397–404 (1995).
75. J. I. Glass, *et al.*, The complete sequence of the mucosal pathogen Ureaplasma urealyticum. *Nature* **407**, 757–762 (2000).
76. C. F. Prada, M. A. Casadiego, C. C. Freire, Evolution of Helicobacter spp: variability of virulence factors and their relationship to pathogenicity. *PeerJ* **10**, e13120 (2022).
77. B. Liu, Y.-S. Ren, C.-Y. Su, Y. Abe, D.-H. Zhu, Pangenomic analysis of Wolbachia provides insight into the evolution of host adaptation and cytoplasmic incompatibility factor genes. *Frontiers in microbiology* **14**, 1084839 (2023).

Figures

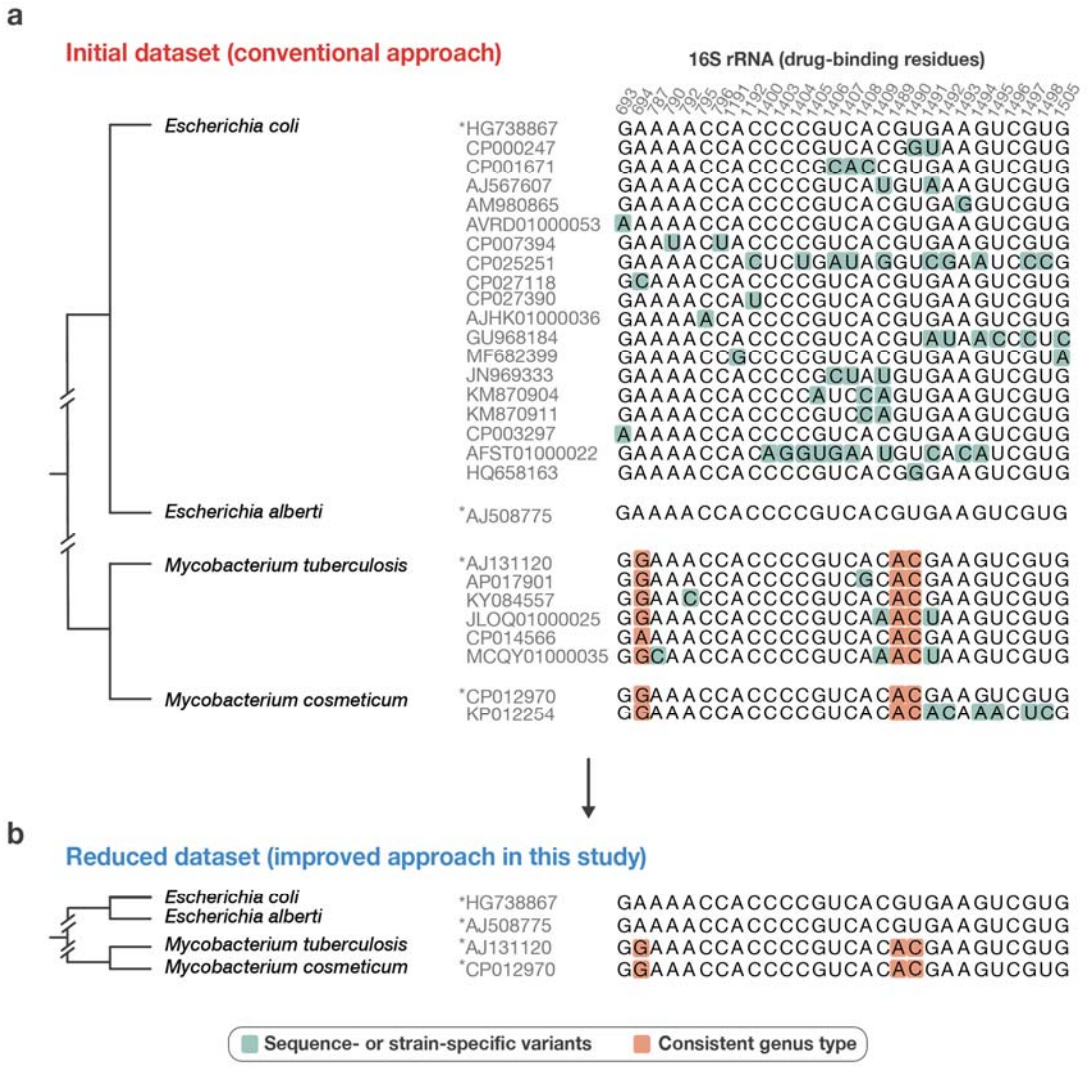

Figure 1. Evolution-based filtering reveals common natural variants of the drug-binding residues of bacterial ribosomes. (**a**) A fragment of aligned 16S rRNA sequences from the SILVA dataset (NR99 v138.1) illustrates the apparent variability of drug-binding residues in bacterial ribosomes among representative members of *Escherichia* and *Mycobacterium* species. Notably, the extensive number of sequences in current biological repositories suggests that *E. coli* alone contains at least 19 dissimilar variants of drug-binding residues. This high apparent degree of variability within a single organism complicates the identification of common variations in drug-binding residues and raises questions about the authenticity of these variants—specifically, how many of them are sequencing errors or strain-specific variants and how many are fixed in species or larger bacterial clades. (**b**) A reduced alignment of rRNA sequences demonstrates our methodology in which we represent each species by a single rRNA sequence, by filtering out the variants that are not shared (at least) between sister rRNA sequences on the tree of life. This approach markedly simplifies the analysis of large libraries of rRNA sequences. Although this method omits variations that are specific to individual sequences or individual bacterial strains, it effectively minimizes stochastic errors, making it possible to identify natural variations that are shared by bacterial genera or broader clades. We have applied this approach

to assess conservation of drug-binding residues in 510,508 rRNA sequences from 8,809 bacterial species.

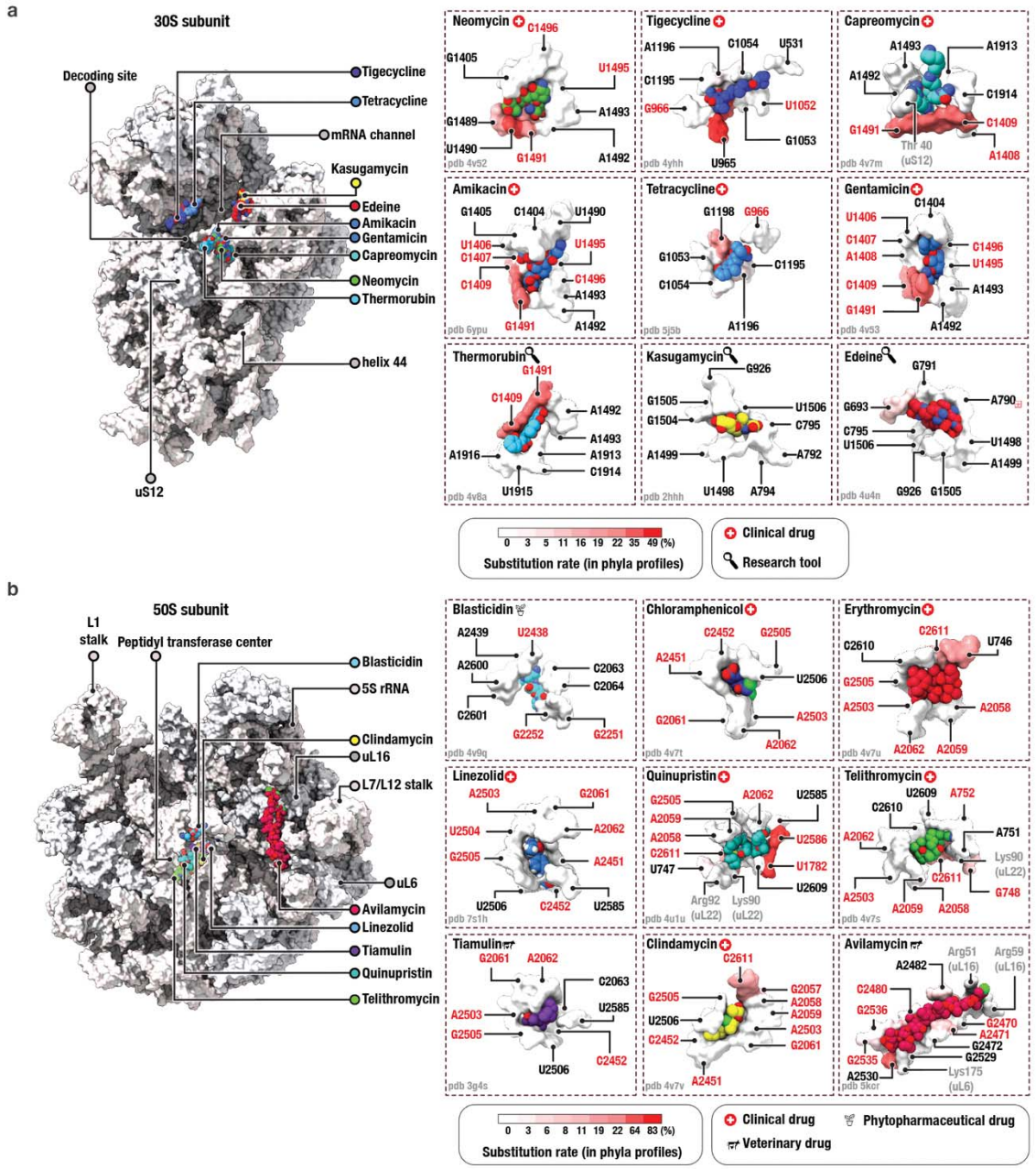

Figure 2 | Highly variable residues occur in each drug-binding site of bacterial ribosomes. Structures of ribosomal subunits (on the left) and 18 representative drug-binding pockets (on the right) illustrate the conservation of individual drug-binding residues in the 16S and 23S rRNA across bacterial phyla. Each ribosome-targeting drug is shown as coloured blobs, and each drug-binding residue is depicted as a surface coloured by conservation across phyla. The figure shows that contrary to the common assumption that drug-binding sites are highly conserved, drug-binding pockets vary greatly across bacterial phyla. This fact becomes apparent when the analysis of rRNA sequences from different species is adjusted to the size of each bacterial

phylum. This adjustment eliminates the bias of the public repositories of rRNA sequences towards overrepresented proteobacteria. Overall, this figure illustrates that bacterial ribosomes are not uniform but possess highly variable ribosomal drug-binding sites.

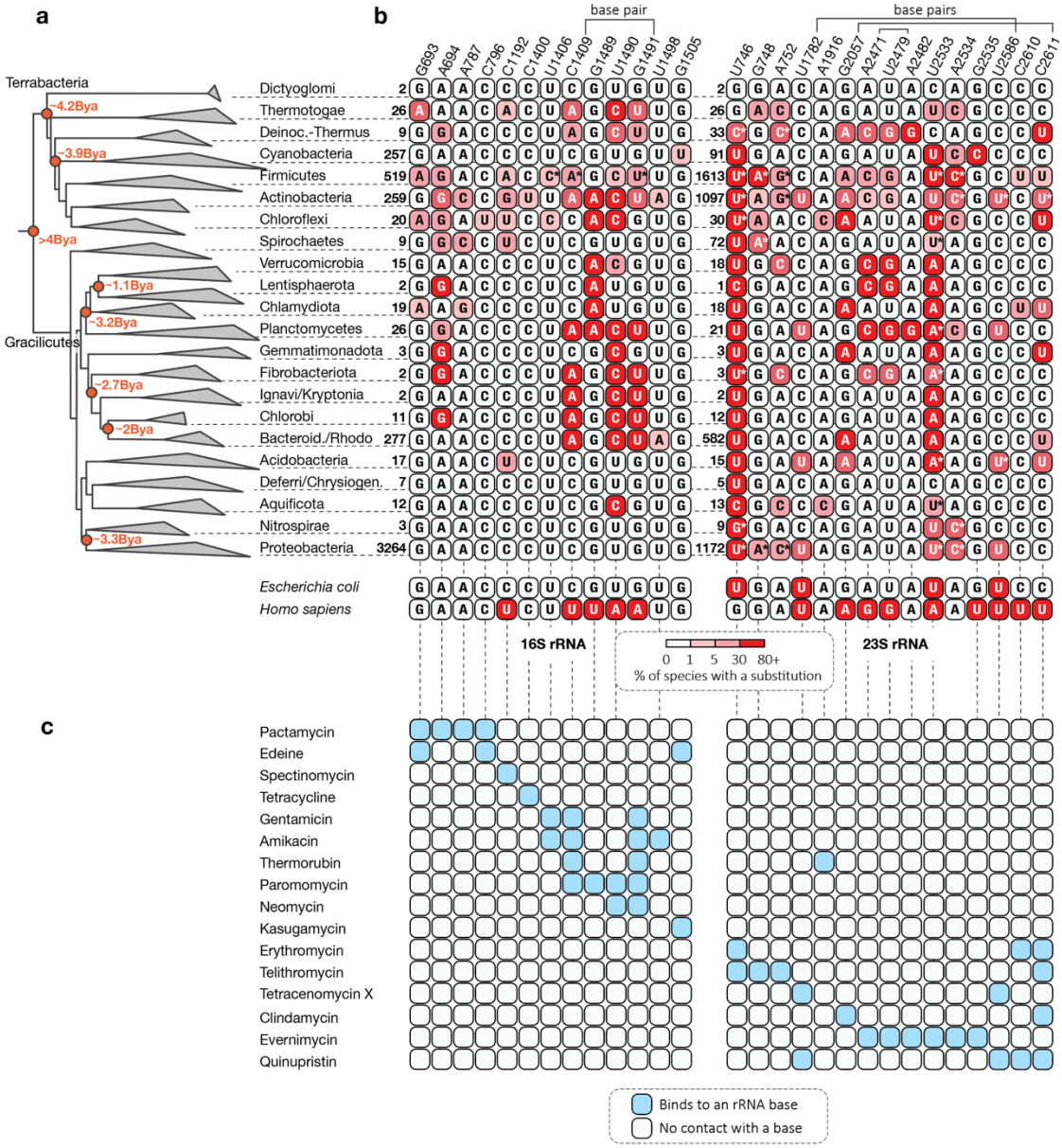

Figure 3 | Bacterial phyla possess conserved changes in ribosomal drug-binding sites. (**a**) The bacterial tree of life illustrates the evolutionary diversification of ribosomal drug-binding residues (**b**) and shows which ribosome-targeting drugs bind to specific residues in rRNA (**c**). Numbers in (**b**) correspond to the total count of representative rRNA sequences that were used to assess substitution profiles of drug-binding residues in rRNA. For simplicity, (**b, c)** display only those rRNA residues that form direct contacts between ribosome-targeting antibiotics and their variable bases (as opposed to the invariant backbone). This panel provides only a simplified summary of variations, while more comprehensive information regarding 8,809 bacterial species can be found in (**SI_Data S4, S5**). Asterisks indicate sites that have multiple types of substitutions (e.g. G to U and A instead of G to U). Overall, the figure shows that most bacterial phyla exhibit

multiple substitutions in their ribosomal drug-binding residues compared to *E. coli* and to one another. Some of these substitutions likely occurred at least 3 billion years ago (Bya), indicating their ancient origin and widespread distribution across bacteria.

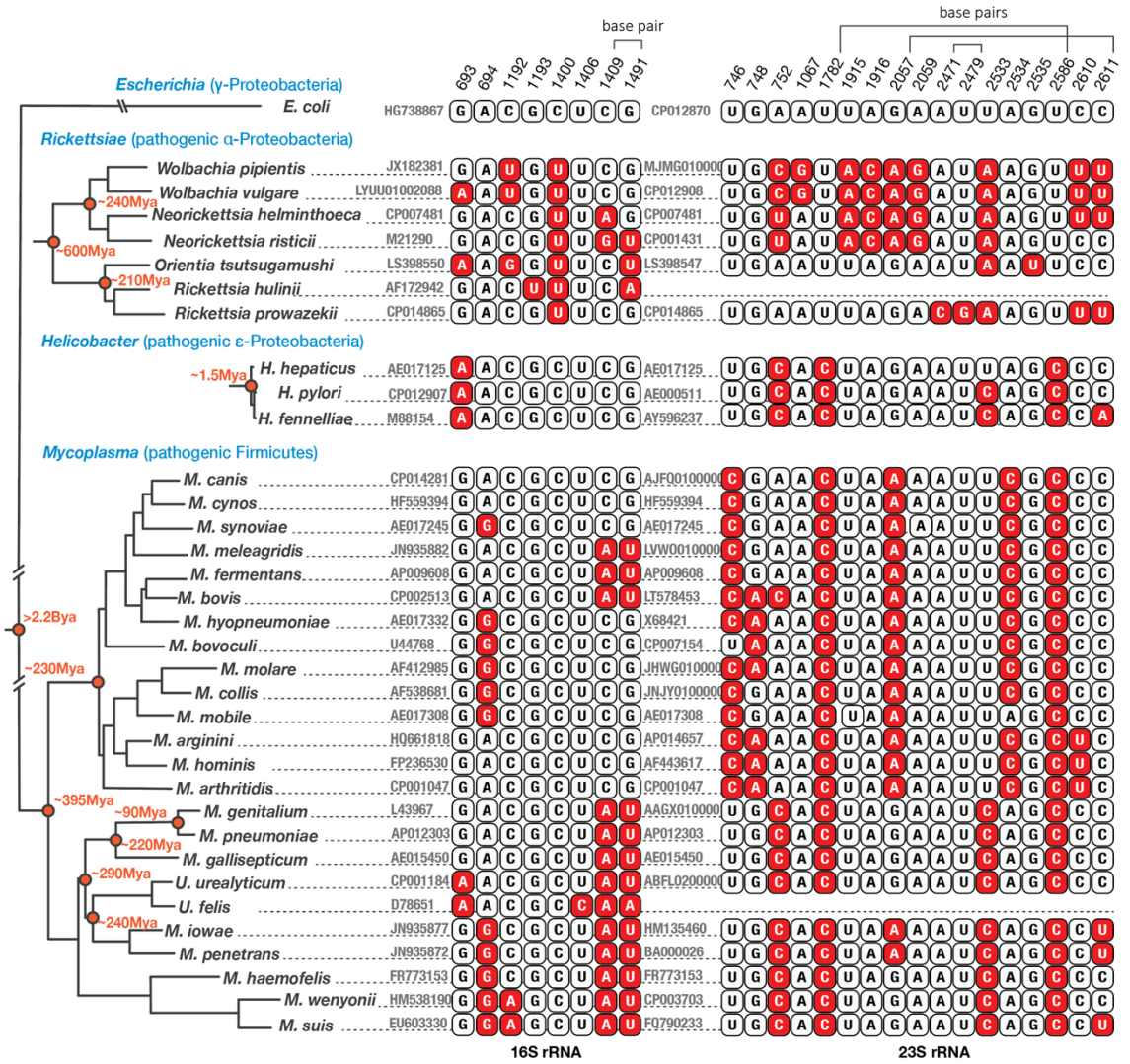

Figure 4 | Closely related bacteria can have highly divergent ribosomal drug-binding sites. A phylogenetic tree and multiple sequence alignment illustrate the natural variations in ribosomal drug-binding residues of pathogenic bacteria *Mycoplasma*, *Helicobacter*, and *Rickettsiae*. The estimated ages of bacterial lineages in millions of years (Mya) are derived from genomic and fossil analyses of the corresponding branches (see **Methods**). Asterisks denote two *Mycoplasma* species (*M. hominis* and *M. pneumoniae*) where the natural divergence of rRNA residues G/A2057 was previously shown to confer intrinsic antibiotic resistance, as observed in an ~18,000x higher tolerance to 14- and 15-membered macrolides (18). Overall, this analysis reveals that drug-binding sites can diverge rapidly among even closely-related bacteria.