TITLE

Natural variation of the drug-binding residues in eukaryotic ribosomes

AUTHORS

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

AFFILIATIONS

[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

CORRESPONDENCE

* To whom correspondence should be addressed:
tom.a.williams@bristol.ac.uk and sergey.melnikov@newcastle.ac.uk

ABSTRACT

Drugs that target eukaryotic ribosomes are becoming increasingly important as research tools and as potential therapies against cancer, fungi and other pathogenic eukaryotes. However, in the absence of comparative studies, we currently do not know how many eukaryotes possess ribosomal drug-binding sites identical to those in humans, and how many significantly differ from humans. Currently, this gap in knowledge is exacerbated by the presence of pseudogenes in eukaryotic genomes, making these comparative analyses challenging due to our inability to discriminate between genuine mutations, pseudogenes and sequencing artifacts. In this study, we resolve this problem by using a novel approach that leverages evolutionary relationships among species. Using this approach, we determine sequence variants for each of the 58 ribosomal drug-binding residues across 8,563 representative eukaryotes, tracing the evolutionary history of these variations from the emergence of eukaryotes 2 billion years ago to their subsequent divergence into distinct lineages. Unexpectedly, we find that yeasts and humans—commonly used as model eukaryotes to study ribosome/drug interactions—differ from most other eukaryotes due to the rRNA substitutions that mainly occur in animals and fungi but are absent in most other eukaryotes. Furthermore, we demonstrate that structural variants previously identified in common pathogens *Leishmania* and *Plasmodium* and viewed as idiosyncratic to a few eukaryotic species, are in fact shared by large clades of eukaryotes. Notably, some eukaryotic lineages exhibit ribosomal drug-binding sites that are more dissimilar to those of humans than humans are to bacteria. Overall, our study provides the most complete overview of ribosomal drug-binding site evolution in eukaryotes—at the levels of single species, single residues, and single drugs—identifying the eukaryotic lineages with structurally distinct ribosomal drug-binding sites compared to those of humans. These findings open new avenues for employing ribosome-targeting drugs as research tools and for the development of lineage-specific inhibitors against eukaryotic parasites.

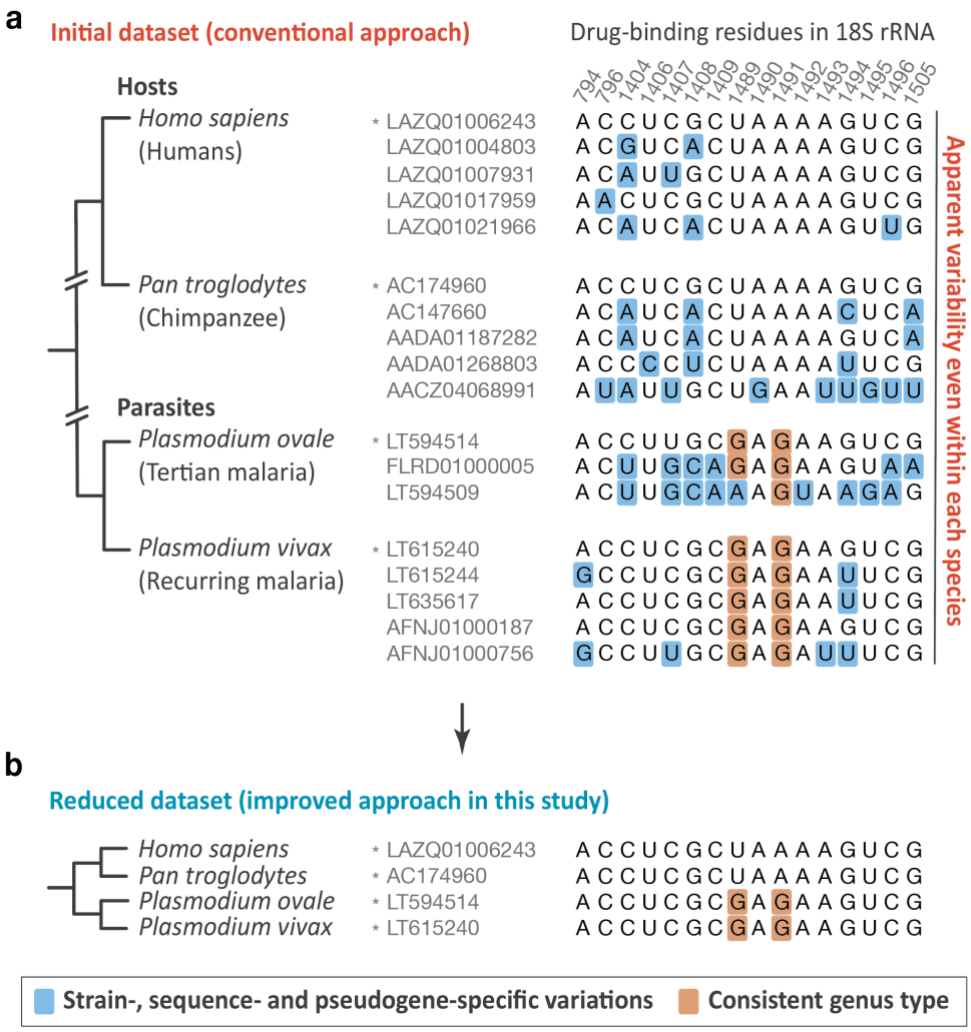


Figure 1 | Evolutionary proximity reveals common variants at drug-binding sites in eukaryotic ribosomes. (**a**) Multiple sequence alignments compare the ribosomal drug-binding residues of the 18S rRNA from a few representative eukaryotes, including humans and chimpanzees, as well as *Plasmodium* parasites. The rRNA sequences are shown as deposited in the SILVA database, which is the most commonly used database for studying rRNA sequences across species. The panel illustrates that the ribosomal drug-binding residues appear to be highly variable even within a single species, but this seeming variability may be attributed to pseudogenes. (**b**) Multiple sequence alignment shows the reduced dataset of 18S rRNA sequences, which includes only those sequences from panel (**a**) that are conserved between immediate neighbors on the tree of life. This approach illustrates the key idea of our study: we reasoned that if a certain rRNA substitution is ancient, it will be present not only in a given species but also in its neighbor on the tree of life (e.g., other eukaryotes from the same genus). By contrast, mutations that are either most recently acquired or false positives (due to misannotations, sequencing errors, or pseudogenes) should be present only in a single sequence, single rRNA operon, single strain, or a single species, but not in neighboring species on the tree of life. This approach helps eliminate pseudogenes and reveal the most common variants of the ribosomal drug-binding sites of eukaryotic ribosomes.

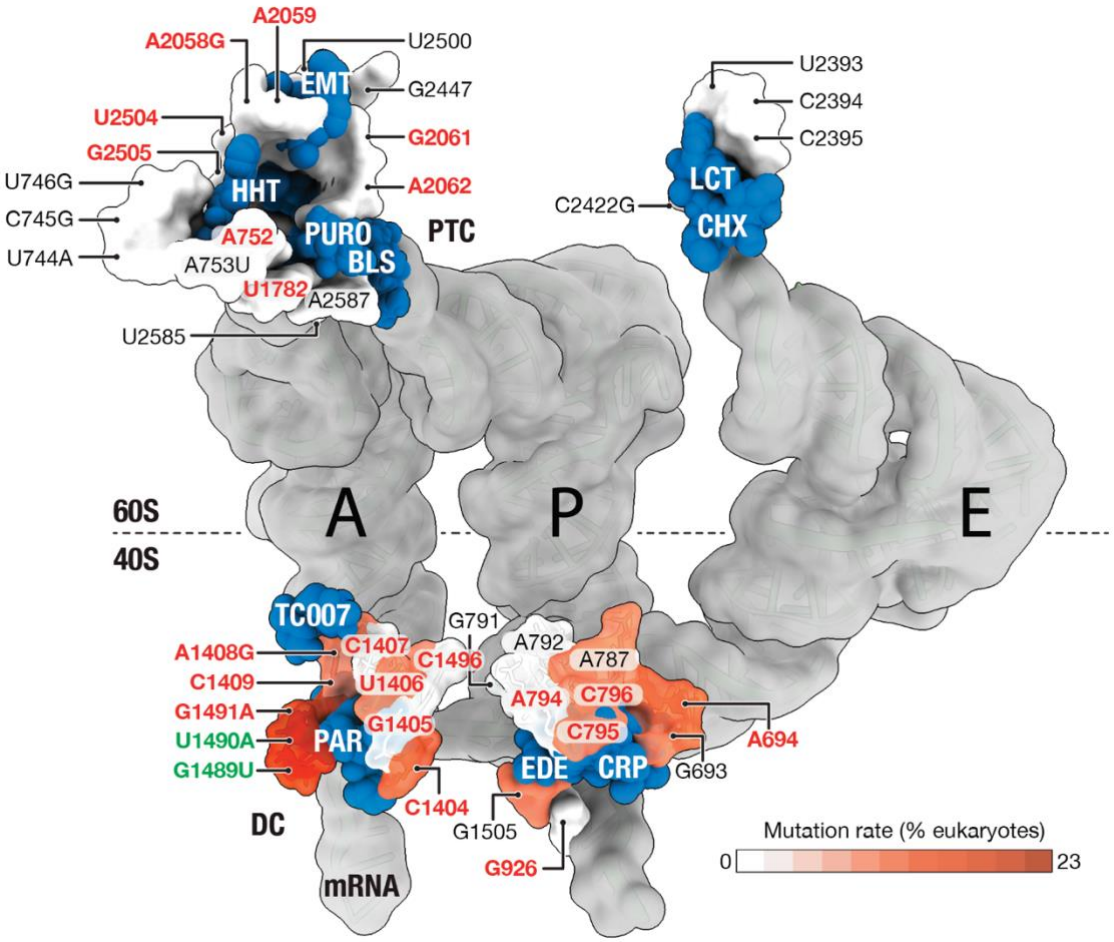


Figure 2 | Most variable drug-binding residues are found in the small ribosomal subunit. Superposed structures highlight the binding sites for ribosome-targeting drugs (in blue) in the ribosome, with each drug-binding residue of eukaryotic ribosomes colored by its conservation across the 8,563 representative eukaryotic sequences analyzed in this study. The figure represents the superposition of multiple PDB structures that capture eukaryotic ribosomes in complex with drugs, as listed in (**Table S1**). The drugs are labeled as follows: EMT – emetine, HHT – homoharringtonine, PURO – puromycin, BLS – blasticidin S, LCT – lactimidomycin, CHX – cycloheximide, PAR – paromomycin, EDE – edeine, CRP - cryptopleurine. The figure shows that, in the large ribosomal subunit, drug-binding residues are highly conserved across eukaryotes, suggesting that all eukaryotes may share a uniform recognition of drugs that target the large ribosomal subunit. However, the drug-binding residues of the small ribosomal subunit are highly variable, both in the decoding center and in the mRNA channel of the ribosome, which illustrates the frequent occurrence of species-specific features of the drug-binding pockets in the small subunit of the eukaryotic ribosome.

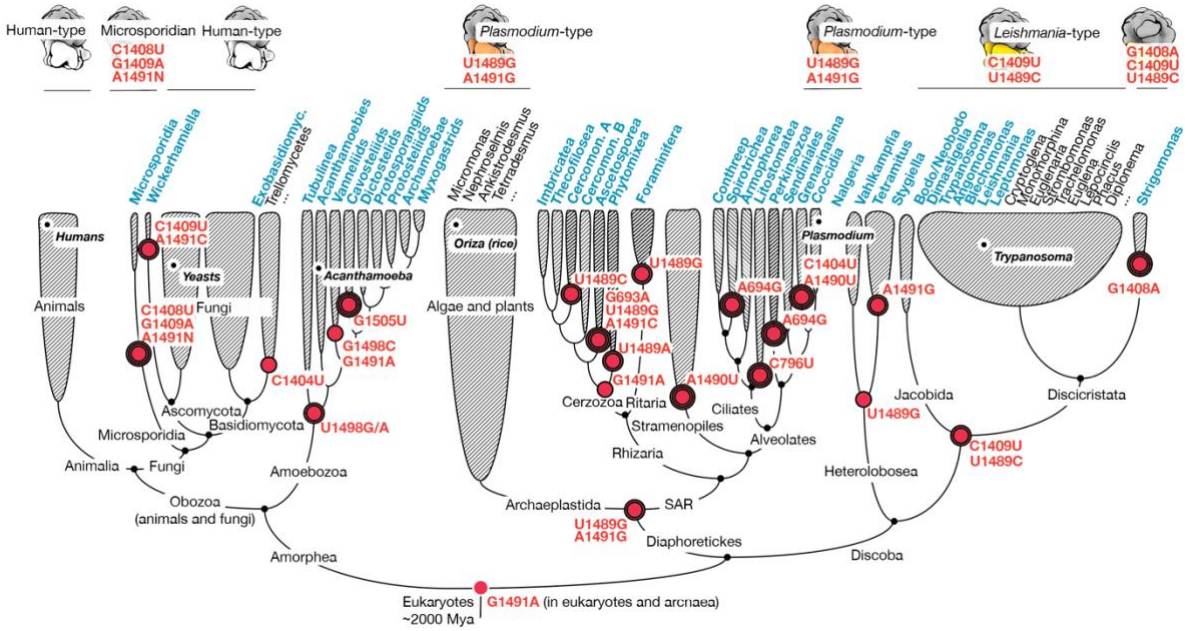
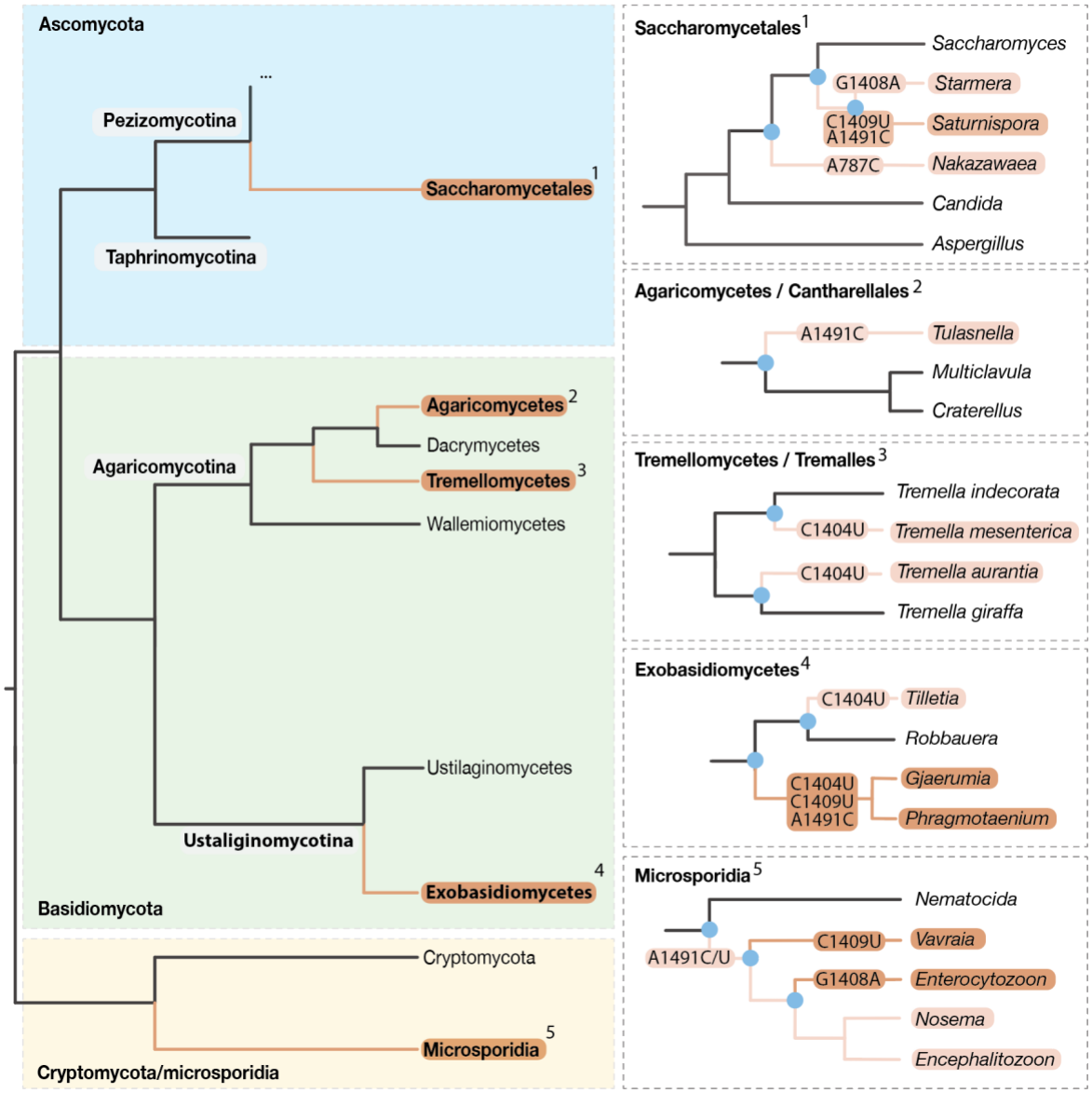


Figure 3 | Most common variants of the drug-binding residues of eukaryotic ribosomes. The eukaryotic tree of life shows the evolutionary history of the naturally occurring variations in the drug-binding residues of the eukaryotic ribosome. Genera highlighted in blue correspond to branches including parasitic species. Red circles indicate evolutionary events related to the acquisition of substitutions in ribosomal drug-binding sites and provide an estimated age of these mutations based on the existing phylogenetic evidence listed in (**Table S3**). Overall, the figure shows that yeasts and humans, traditionally viewed as representative model eukaryotes, in fact bear relatively unusual ribosomal binding site of aminoglycoside antibiotics—due to the residue A1491 compared to G1491 in the majority of eukaryotes from non-animal and non-fungal branches. Therefore, almost all other eukaryotes have structurally distinct binding sites for aminoglycoside antibiotics, where G1491 is present instead of A1491, with some branches (e.g., Discicristata) bearing additional mutations in this and other drug-binding sites of the ribosome (e.g., C1404U, G1408A and C1409U) compared to humans.



Figures 4 | Many fungi bear derived ribosomal drug-binding sites compared to humans. The schematic structure of the fungal tree of life shows that, after separating from animals, some lineages of fungi, such as Microsporidia and Saccharomycetales, continued to diversify their ribosomal drug-binding residues. In total, of the 1,253 genera of fungi analyzed in this study, 101 (~7%) were found to carry one or more mutations in their ribosomal drug-binding residues compared to humans. As a result, compared to humans, some fungi (e.g. plant parasites *Exobasidiomycetes* or animal parasites Microsporidia) bear more substitutions in their ribosomal drug-binding than a typical bacterium, such as *Escherichia coli*.

INTRODUCTION

Small molecules that target eukaryotic ribosomes are widely used as research tools and experimental medicine to treat cancer and human infections with fungi and other pathogenic eukaryotes (1–3). With more than twenty distinct families of small molecules available for ribosome targeting, we can now arrest or modulate ribosome activity at every major step of its work cycle (4). This allows the use of molecules such as puromycin, anisomycin, emetine, cycloheximide, and harringtonine in ribosome profiling for transcriptome-scale observation of a plethora of biological phenomena that require protein synthesis (3). Ribosome-targeting molecules can be photocaged to arrest protein synthesis in response to a light signal within a desired segment of an organ or a cell, which has been particularly useful in studying neurons during cognition and formation of memory in various model eukaryotes (5–7). In addition, ribosome-targeting drugs like homoharringtonine and ataluren are currently used in clinics to treat cancer and genetic disorders, respectively, with many other ribosome-targeting molecules awaiting their approval in clinical trials (8–10).

This popularity of ribosome-targeting small molecules stems from the assumption that ribosomes are highly conserved across eukaryotes (4,11). This suggests that ribosome-drug interactions observed in organisms like yeasts and humans can be generalized to other eukaryotes. However, growing structural and genomic evidence indicates that certain eukaryotes have structurally dissimilar drug-binding residues compared to those in yeasts and humans. For example, the parasitic fungus *Encephalitozoon cuniculi* was shown to harbor the 18S rRNA substitution A1491U in the aminoglycoside-binding site of the ribosome (12). Similarly, the causative agents of human diseases from the *Leishmania* and *Trypanosoma* genera were shown to exhibit other substitutions in the aminoglycoside-binding site of the ribosome, such as C1409U (13–15). These findings were remarkable because mutations of either A1491 or C1409 in the 16S/18S rRNA of various species were previously characterized as resistance-conferring due to their negative impact on aminoglycosides binding to the ribosome (16,17). Thus, we currently know that ribosomal drug-binding residues are not strictly conserved across eukaryotes, suggesting that resistance (and sensitivity) to ribosome-targeted drugs might vary across the eukaryotic phylogeny. However, in the absence of a large-scale comparative study, it remains unclear whether species like *Encephalitozoon*, *Trypanosoma*, and *Leishmania* represent rare exceptions or rather a common rule among eukaryotes. As a result, we do not know how many eukaryotes have identical ribosomal drug-binding sites to humans, and how many have dissimilar drug-binding sites.

Here, we have answered this question through comparative analysis of ribosomal drug-binding residues in 8,563 eukaryotes from all major branches of this domain of life. To achieve this, we have implemented a novel approach to the analysis to resolve common problems in public sequencing data repositories—such as chimeric sequences, species misannotations and pseudogenes - thereby ameliorating the impact of database errors on our analyses and conclusions (18,19). This allowed us to accurately trace the evolutionary history for every known drug-binding residue of the eukaryotic ribosome throughout 2 billion years of eukaryotic history. The analysis revealed over 60 sequence variants of ribosomal drug-binding residues, including residues characterized as resistance-conferring in model bacteria and eukaryotes. Large families of eukaryotes, including human pathogens, exhibit structurally dissimilar drug binding sites compared to those of humans. The findings can, therefore, guide the accurate application of ribosome-targeting drugs as research tools and may lead to the design of lineage-specific ribosome inhibitors to safely target pathogenic eukaryotes without affecting humans.

RESULTS

Mapping substitutions in ribosomal drug-binding sites during eukaryotic evolution

To assess the conservation of drug-binding sites in eukaryotic ribosomes, we first set out to address the problem of spurious mutations in biological sequences. Previous studies have shown that datasets of biological sequences, including rRNA, may contain errors resulting from incorrect base calls during sequencing, assembly mistakes, or erroneous annotations of sequencing data (18,19). Because many biological sequences are assembled from multiple shorter reads, public repositories were estimated to contain up to 20% chimeric sequences, where different segments of a single sequence originate from distinct species rather than a single source. Other prevalent artifacts include sequencing errors, contamination from DNA of organelles, plasmids, and viruses, and the presence of pseudogenes.

The problem of a high prevalence of pseudogenes is especially acute for eukaryotes in comparison to bacteria. For example, a typical *E. coli* strain is estimated to bear only 10-20 pseudogenes, equating to roughly one pseudogene per 300 genes (20). By contrast, the human genome bears at least 19,724 pseudogenes, which is approximately one pseudogene for each protein-coding gene (21). The pseudogene count is even higher in many other eukaryotes, especially plants, with barley, for example, bearing approximately 35,000 genes but 89,440 pseudogenes (22). As a result, previous large-scale analyses of conservation of ribosomal drug-binding residues are limited to bacterial species (23), leaving a substantial knowledge gap about the variability of these residues across eukaryotes.

To resolve this problem and distinguish signal from noise in eukaryotic rRNA datasets, we focused on changes at drug-binding sites that are present in at least two closely related eukaryotic species. We reasoned that this approach would identify genuine evolutionary changes that fixed in the (relatively) recent past, that are likely to be conserved in the target species, and so have predictive power for guiding the choice and application of ribosome-targeting drugs. By contrast, mutations that are either false positives or very recently acquired should be present only in a single sequence, single rRNA operon, single strain, or a single species, but not in neighboring species on the tree of life. Using this reasoning, we assessed the conservation of 58 rRNA residues—which were previously shown to directly bind ribosome-targeting drugs—in rRNA sequences that were deposited in the SILVA dataset NR99 v138.1, which is the most complete dataset of non-redundant rRNA sequences (**Methods, Figure 1, Table S1**) (24).

For each of these residues, we traced their evolutionary history and conservation across 8,563 representative eukaryotes from 5,148 distinct genera. Thus, we were able to characterize the evolution of drug-binding residues in eukaryotic ribosomes at the levels of single residue, single species and single ribosome-targeting drug.

Drug-binding sites of the small subunit are highly variable across eukaryotes

We determined the most prevalent rRNA variations across eukaryotes by calculating the overall level of conservation for each drug-binding residue in the ribosome structure (**Figure 2, Supplementary Data 4 and 5**). We observed that the large and small ribosomal subunits have strikingly dissimilar conservation of their drug-binding rRNA nucleotides.

The large ribosomal subunit exhibits an exceptionally high degree of sequence conservation at drug-binding sites: in the eukaryotes analyzed, U2393 and G2422 were the only variable drug-binding residues, having been replaced with As in the pathogenic yeast *Malassezia.* These yeasts are associated with dermatological conditions, including seborrheic dermatitis and tinea versicolor (25).

Other drug-binding residues were conserved across analyzed eukaryotes, pointing to the uniform structure of the drug-binding sites in the large subunit of eukaryotic ribosomes.

In stark contrast, in the small subunit residues, we observed variations for the majority of drug-binding residues, including 15 of the 26 residues in 18S rRNA. Some variations—such as A787C, C796U, and C1404U—were rare, occurring in only 20 to 40 eukaryotes within our dataset. However, other variations, including U1489G/C and A1491G/C/U, were highly prevalent, each observed in nearly 2,000 species. Notably, we observed variations not only in rRNA residues that differ between bacteria and humans but also in residues that are currently viewed as universally conserved due to their conservation across model eukaryotes (e.g. humans) and bacteria (e.g. *E. coli*) (**Figure 2**). As result, we found that approximately 23% of eukaryotes harbor between one and four fixed changes in their drug-binding sites compared to humans.

The evolution of drug-binding residues of the eukaryotic ribosome

We next traced the evolutionary history of variations for each drug-binding residue of eukaryotic ribosomes (**Figure 3**). This allowed us to estimate the evolutionary age of these variations and their distribution across all studied eukaryotes.

Previously, the last common ancestor of modern eukaryotes, LECA, was estimated to have emerged around 1.8-2.3 billion years ago (26,27), with the divergence of stem eukaryotes from archaea occurring somewhat earlier (2.7-2.2 Ga, (27)). Our analysis indicates that some of the key differences between bacterial and eukaryotic ribosomes in terms of drug sensitivity were already established before the divergence of eukaryotes from Archaea. For example, our analysis indicates that G1408 and A1491 in 18S rRNA (compared to A1408 and G1491 in bacterial 16S rRNA) were already established prior to the archaea-eukaryote divergence Importantly, each of these mutations has been characterized as a critical determinant for the species-specific recognition of aminoglycoside antibiotics: they prevent the binding of paromomycin in the ribosomal decoding center due to the steric clash between the rRNA base and the antibiotic (4), leading to approximately 1,000-fold lower ribosome affinity for aminoglycosides (16). Our analysis showed that the major lineages of eukaryotes accumulated additional drug-binding variants during their early diversification.

Specifically, after Discoba separated from the other eukaryotes approximately 1.4 to 1.8 billion years ago (27), it acquired a mutation in the residue U1489. Subsequently, most Discoba, including the branches of Jakobida and Discicristata, also acquired a mutation at C1409. As a result, these species have gained a structurally distinct drug-binding site that bears the C1409U and U1489C mutations compared to humans. Notably, these mutations have been previously characterized in the ribosomal structures of Trypanosoma and Leishmania species as candidates for species-specific targeting of the ribosome with aminoglycoside antibiotics. Our analysis shows that these mutations are evolutionarily ancient and represent a common characteristic not only of *Trypanosoma* and *Leishmania* but of the entire eukaryotic clade of Discoba, which includes human pathogens *Naegleria fowleri*, *Balamuthia mandrillaris*, and many free-living forms.

The TSAR lineage also acquired conserved drug-binding site changes after diverging from other eukaryotes, with the residue U1489G and the residue G1491A, which reverted the sequence of this residue to make it identical to the bacterial variant. As a result, these structural dissimilarities of drug-binding sites have been fixed and shared by all species of this branch, including modern plants, algae, and single-celled eukaryotes from SAR species (that include *Plasmodium* species, among other human pathogens).

The Amorphea branch further split into the ancestor of modern fungi and animals (Obazoa), most of which have retained the ancestral ribosomal state, and the branch of amoeba-bearing Amoebozoa,

which have acquired the mutation of the A1498 base shared by all members of this clade, including human pathogens *Entamoeba*, *Acanthamoeba* and others.

These early mutational events were later followed by the acquisition of numerous additional mutations, resulting in more than 60 unique combinations of the sequence of drug-binding residues among modern eukaryotes (**Figure 3 and Supplementary Data 5&6**).

Certain clades of fungi bear altered ribosomal drug-binding sites compared to humans

Our further analysis of the eukaryotic tree of life showed that, although most animals and fungi share conserved drug-binding residues, certain fungal lineages have gradually acquired multiple changes at these sites compared to humans (**Figure 4**). This was observed, for example, in the deep-branching fungal lineage Microsporidia, all members of which are fungal parasites of animals, including humans (29). In Microsporidia, the initial mutation likely occurred shortly after the split into *Nematocida* species, which retained human-type (that is, ancestral) drug-binding sites, and the remaining microsporidia, which acquired a mutation in the decoding center at residue A1491, which has been identified as a key determinant of species specificity for aminoglycoside antibiotics (16,4). Subsequent evolution of microsporidia has led to additional mutations in the decoding center, including C1489U in the *Vavraia* branch and G1048A in species of *Enterocytozoon*. Hence, most microsporidia exhibit a more dissimilar decoding center compared to humans than humans compared to *E. coli*.

The remaining fungal lineages were further split into several clades, including Ascomycota and Basidiomycota, with specific branches within these clades acquiring mutations in both the aminoglycoside-binding site and the mRNA channel. These branches include the fungi *Trellomycetes* and *Exobasidomycetes* (C1409U), which encompass pathogenic organisms, such as *Tilletia horrida* (C1404U), a common rice pathogen, and *Gjaerumia minor* (C1404U, C1407U, A1494C), implicated in keratitis of humans. Overall, among the 1,201 fungal species analyzed in our study, we found approximately 7% to exhibit dissimilar ribosomal drug-binding sites compared to humans, due to variations in the 18S rRNA bases A787, C1408, C1409 and A1491 (*E. coli* numbering is used throughout the manuscript. Please refer to Table S2 for the correspondence between species) within the decoding center and the mRNA channel of eukaryotic ribosomes.

Thus, we found that most fungi have identical ribosomal drug-binding residues compared to those in humans, suggesting that ribosome-targeting drugs are unlikely to serve as broad-spectrum antifungal agents. However, species from certain fungal clades, particularly microsporidia, possess multiple mutations in their drug-binding residues compared to humans, indicating the potential for their safe lineage-specific targeting.

DISCUSSION

Animals and fungi possess distinct ribosomal drug-binding sites compared to most other eukaryotes

In this study, we have assessed the conservation of ribosomal drug-binding residues across the eukaryotic domain of life. We determined the natural variations for each of the 58 ribosomal drug-binding residues in 8,563 representative eukaryotes, enabling us to identify lineages with dissimilar residues compared to humans. One major finding of our work is that yeasts and humans, traditionally used as model organisms for studying ribosome-targeting drugs in eukaryotes, exhibit rather unusual variants of drug-binding residues compared to most other eukaryotes. Specifically, yeasts and humans, along with most Amorphea species, appear to have kept the ancestral sequence variant of the drug-binding sites, whereas the other branches of eukaryotes have acquired substitutions in the

drug-binding residues G1408, C1409, A1491 of the 18S rRNA. Importantly, some of these substitutions have reverted the rRNA sequences of eukaryotic rRNA to its bacterial-type variants (e.g. in the 18S rRNA residue G1491 in Diaphoretickes).

One important implication of this finding is how we currently study eukaryotic ribosome targeting with small molecules. From the perspective of drug sensitivity, ribosomes from different organisms are separated into two major groups: bacterial-type and eukaryotic-type (11). This separation is based on the overall protein content and the presence of rRNA expansion segments, allowing the use of organisms like *E. coli* or *T. thermophilus* as representative bacteria and organisms like yeasts or humans as representative eukaryotes (4). However, our work shows that this division, while helpful in many studies of eukaryote-specific ribosomal proteins and rRNA expansion segments, is incomplete and even misleading when applied to the drug-binding residues of the eukaryotic ribosome. For instance, some groups of eukaryotes, such as SAR, share more similarities in their drug-binding residues with bacteria than with humans due to the convergent evolution of the 18S rRNA variation A1491G. Additionally, some others, like microsporidia, are highly dissimilar to both bacteria and humans. Instead, our work shows that the simplistic division of ribosomes into bacterial and eukaryotic is mostly accurate only when applied to the large ribosomal subunit. By contrast, a range of different residues, and so potentially drug-binding sites and sensitivities, are found across eukaryotes in the small subunit of the ribosome, including (roughly) an animal/fungi-type, *Leishmania*-type, and *Plasmodium*-type, among many others.

Many of the naturally occurring variations in ribosomal drug-binding sites predate the origin of antibiotic-producing bacteria

Since when have eukaryotic ribosomes started to diversify their drug-binding residues? And what were the evolutionary forces driving this diversification of ribosomal drug-binding residues? Our mapping of the rRNA sequence variants on the tree of life allows us to gain insights into both of these questions. According to this analysis, substitutions at the drug-binding sites of eukaryotic ribosomes began to accumulate from the origin of the first eukaryotic branches >1.3 Ga (27), with some mutations mapping to the stem lineages of major groups including Amorphea, Discoba and TSAR. This means that common 18S rRNA variants such as A1491G in SAR, and G1409A and C1489U in Discoba species, emerged significantly earlier than the antibiotic-producing genus *Streptomyces*, which is responsible for most natural ribosome-targeting drugs known to date and is estimated to be 382 million years old (30). This analysis suggests that the evolution of drug-binding residues in nature, at least during early eukaryotic evolution, differs from their evolution in clinical settings and has likely been driven by factors other than the existence of antibiotics in the environment.

Many naturally occurring rRNA variants presenting concerns and opportunities for species-specific targeting of eukaryotic ribosomes

Previous studies showed that mutations even in a single rRNA drug-binding residue can confer drug resistance by reducing drug-affinity as much as 1,000 fold (31). In this study, we showed that some of these variations, previously observed in clinical isolates of human pathogens or laboratory-engineered resistant strains, are widely present in nature.

Our analysis, along with the previously obtained experimental studies, suggests that some of the naturally occurring variations are likely neutral. Specifically, variations of residues G1489U and U1490A have been characterized as neutral for ribosome targeting by aminoglycoside drugs, as demonstrated through ribosome mutagenesis in *Mycobacterium smegmatis* and *E. coli* (32–34).

However, other variations, including 13 residues of the 18S rRNA, have been shown to alter the shape of the drug-binding pocket and define species-specific ribosome targeting with drugs (red labels in **Figure 2** and **Table S4**). For example, the variations A1408G and G1491A in the ribosomal decoding site have been characterized as determinants of species-specific binding of aminoglycoside antibiotics, allowing for the safe targeting of bacterial ribosomes (A1408 and G1491) without affecting ribosomes in humans (G1408 and A1491) (35). Experiments in yeasts showed that when eukaryotic ribosomes are mutated to introduce the bacteria-specific variant G1491 in the 18S rRNA (instead of A1491), they develop aminoglycoside sensitivity similar to *E. coli*, evidenced by a 60-fold higher sensitivity to the antibiotic paromomycin. Similarly, mutating the G1408 site to the bacteria-type A1408 resulted in a ~200-fold increase in sensitivity of yeast ribosomes to the antibiotics neomycin and kanamycin A (16). Our analysis reveals that these variants are not limited to bacteria, but are also prevalent across eukaryotes, naturally occurring in up to 23% of eukaryotic species.

Overall, our findings present both concerns and opportunities for using drugs to target eukaryotic ribosomes. While eukaryotic ribosomes are traditionally regarded as conserved molecular machines, our research reveals that their drug-binding sites often exhibit significant variability across different eukaryotic lineages, thus necessitating caution and additional analyses prior to applying these drugs to non-model eukaryotes. On the opportunity front, our work suggests the possibility for lineage-specific ribosome-targeting drugs across a wide variety of eukaryotes. Notably, we demonstrate that structural variants identified in *Leishmania* and *Plasmodium*, previously considered idiosyncratic to rare eukaryotic lineages, are in fact common characteristics of entire eukaryotic supergroups. What's more, we show that some lineages of eukaryotes have more dissimilar ribosomal drug-binding sites compared to humans than humans do compared to bacteria. Therefore, because eukaryotic pathogens pose significant and emerging global health challenges, our work suggests that novel therapeutics can be found among new generations of ribosome inhibitors that target lineage-specific substitutions in the ribosomal RNA.

METHODS

Locating the drug-binding residues in eukaryotic ribosomes

To identify the ribosomal residues that are directly involved in drug recognition, we analyzed 29 previously determined structures deposited in the Protein Data Bank (36) in which eukaryotic ribosomes were bound to a representative of each family of ribosome-targeting drugs (**Table S1**). Using ChimeraX, we selected non-hydrogen atoms of the ribosome located within 4.3 Å from a drug. These sets of atoms were then reduced to only include atoms that belonged to amino acid side chains of ribosomal proteins and aromatic bases of rRNA. This allowed us to determine the coordinates of each ribosomal residue that makes a sequence-dependent contact with ribosome-targeting drugs, as well as the numbering correspondence between conserved rRNA residues of *E. coli*, *S. cerevisiae* and *H. sapience* ribosomes (**Table S2**).

Creating a high-fidelity dataset of rRNA sequences for evolutionary analyses

To enable accurate analysis of rRNA sequences, we used the SILVA database and have devised a multi-step data reduction strategy in order to bypass the notorious issues of automated data annotation and sequencing, including the occurrence of chimeric sequences, misannotated phylogeny, non-genomic DNA and pseudogenes (**Supplementary Figure 1**). In brief, as our starting point we used the SILVA's dataset NR99 v138.1, which contains 510,508 16S/18S rRNA sequences corresponding to 78,104 unique "species" (including metagenomic variants, unknown species, species candidates and other ambiguously defined species) ("SSU dataset") and 95,286 sequences of 23S/5.8S rRNA from 20,471 unique "species" ("LSU dataset") from all domains of life (24). We first reduced these SILVA datasets by eliminating sequences from non-eukaryotic species as well as from unidentified eukaryotes 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 organelles. We then removed all sequences of apparently poor sequencing quality, based on the presence of 'ambiguous' sites represented by symbols such as "R", "X", "N", and "S" instead of the standard set of RNA bases "A", "G", "C", and "U". The sequences from each of these datasets were initially aligned using MAFFT v7.490 with default settings (FFT-NS-2) (37) during the quality control analysis. The sequences that passed the quality control were then re-extracted from the original SILVA dataset of aligned rRNA sequences and are shown in (**Supplementary Data 2 and 3**).

Assessing the conservation of the ribosomal drug-binding residues

To simplify the assessment of conservation of drug-binding residues of the eukaryotic ribosome, we trimmed the aligned rRNA sequences to only the drug-binding residues. This file was then manually inspected for quality control, and the rRNA with apparent truncations in the drug-binding sites (based on the presence of gaps in the place of drug-binding residues) were removed from our further analysis. The remaining sequences were then reordered according to the location of species on the tree of life. 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 (e.g. other eukaryotic species from the same genus). 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 or an individual rRNA operon (**Supplementary Data 4 and 5**).

Reconstructing the evolutionary history of mutations in ribosomal drug-binding sites

To trace the evolutionary history of sequence variations in the ribosomal drug-binding residues, we used 16S rRNA sequences listed in (**Supplementary Data 5**) to generate a phylogenetic tree. The tree was generated using default settings on Mega11 (38). The tree was then colored by conservation of ribosome-drug-binding residues by highlighting species and lineages with the human-type and non-

human-type sequences of ribosomal drug-binding residues (**Figures 3,4**). The relative age of each mutation was estimated using TimeTree (39) or the original studies cited where appropriate in the manuscript.

ACKNOWLEDGEMENTS

We would like to thank Robert Hirt, Claudia Schneider, Wyatt Yue and Heath Murray (all Newcastle University, UK) and Erik C. Böttger (University of Zurich, Switzerland) for their comments on the manuscript and Yuri Polikanov (University of Illinois at Chicago, USA) for encouraging this study. This work was supported by the BBSRC Doctoral Training Program (BB/T008695/1 to L.I.C. and C.R.B.), the MRC Discovery Medicine North Doctoral Training Partnership (MR/N013840/1 to C.L.E.) and the Newcastle University Academic Track Fellowship (to S.M.).

REFERENCES

1. Dmitriev SE, Vladimirov DO, Lashkevich KA. A Quick Guide to Small-Molecule Inhibitors of Eukaryotic Protein Synthesis. Biochemistry Moscow. 2020 Nov 1;85(11):1389–421.

2. Olombrada M, Lázaro-Gorines R, López-Rodríguez JC, Martínez-del-Pozo Á, Oñaderra M, Maestro-López M, et al. Fungal ribotoxins: a review of potential biotechnological applications. Toxins. 2017;9(2):71.

3. Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. Nature reviews Molecular cell biology. 2015;16(11):651–64.

4. Garreau De Loubresse N, Prokhorova I, Holtkamp W, Rodnina MV, Yusupova G, Yusupov M. Structural basis for the inhibition of the eukaryotic ribosome. Nature. 2014 Sep 25;513(7519):517–22.

5. Flexner LB, Gambetti P, Flexner JB, Roberts RB. Studies on memory: distribution of peptidyl-puromycin in subcellular fractions of mouse brain. Proc Natl Acad Sci USA. 1971 Jan;68(1):26–8.

6. Eberwine J, Miyashiro K, Kacharmina JE, Job C. Local translation of classes of mRNAs that are targeted to neuronal dendrites. Proc Natl Acad Sci U S A. 2001 Jun 19;98(13):7080–5.

7. Ko T, Oliveira MM, Alapin JM, Morstein J, Klann E, Trauner D. Optical Control of Translation with a Puromycin Photoswitch. J Am Chem Soc. 2022 Nov 30;144(47):21494–501.

8. Zhou DC, Zittoun R, Marie JP. Homoharringtonine: an effective new natural product in cancer chemotherapy. Bulletin du cancer. 1995;82(12):987–95.

9. Peltz SW, Morsy M, Welch EM, Jacobson A. Ataluren as an Agent for Therapeutic Nonsense Suppression. Annu Rev Med. 2013 Jan 14;64(1):407–25.

10. Ryan NJ. Ataluren: First Global Approval. Drugs. 2014 Sep;74(14):1709–14.

11. Melnikov S, Ben-Shem A, Garreau de Loubresse N, Jenner L, Yusupova G, Yusupov M. One core, two shells: bacterial and eukaryotic ribosomes. Nat Struct Mol Biol. 2012 Jun;19(6):560–7.

12. Nicholson D, Salamina M, Panek J, Helena-Bueno K, Brown CR, Hirt RP, et al. Adaptation to genome decay in the structure of the smallest eukaryotic ribosome. Nat Commun. 2022 Feb 1;13(1):591.

13. Shalev M, Kondo J, Kopelyanskiy D, Jaffe CL, Adir N, Baasov T. Identification of the molecular attributes required for aminoglycoside activity against Leishmania. Proceedings of the National Academy of Sciences. 2013 Aug 13;110(33):13333–8.

14. Hashem Y, Des Georges A, Fu J, Buss SN, Jossinet F, Jobe A, et al. High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature. 2013;494(7437):385–9.

15. Shalev-Benami M, Zhang Y, Rozenberg H, Nobe Y, Taoka M, Matzov D, et al. Atomic resolution snapshot of Leishmania ribosome inhibition by the aminoglycoside paromomycin. Nat Commun. 2017 Nov 17;8(1):1589.

16. Fan-Minogue H, Bedwell DM. Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. RNA. 2008 Jan 1;14(1):148–57.

17. Hobbie SN, Kaiser M, Schmidt S, Shcherbakov D, Janusic T, Brun R, et al. Genetic Reconstruction of Protozoan rRNA Decoding Sites Provides a Rationale for Paromomycin Activity against Leishmania and Trypanosoma. PLoS Negl Trop Dis. 2011 May 24;5(5):e1161.

18. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol. 2005 Dec;71(12):7724–36.

19. Edgar RC. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ. 2018;6:e4652.

20. Lawrence JG, Hendrix RW, Casjens S. Where are the pseudogenes in bacterial genomes? Trends in microbiology. 2001;9(11):535–40.

21. Torrents D, Suyama M, Zdobnov E, Bork P. A genome-wide survey of human pseudogenes. Genome research. 2003;13(12):2559–67.

22. Prade VM, Gundlach H, Twardziok S, Chapman B, Tan C, Langridge P, et al. The pseudogenes of barley. The Plant Journal. 2018 Feb;93(3):502–14.

23. David-Eden H, Mankin AS, Mandel-Gutfreund Y. Structural signatures of antibiotic binding sites on the ribosome. Nucleic Acids Research. 2010 Oct 1;38(18):5982–94.

24. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 2012;41(D1):D590–6.

25. Velegraki A, Cafarchia C, Gaitanis G, Iatta R, Boekhout T. Malassezia infections in humans and animals: pathophysiology, detection, and treatment. PLoS pathogens. 2015;11(1):e1004523.

26. Strassert JF, Irisarri I, Williams TA, Burki F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nature Communications. 2021;12(1):1879.

27. Mahendrarajah TA, Moody ER, Schrempf D, Szánthó LL, Dombrowski N, Davín AA, et al. ATP synthase evolution on a cross-braced dated tree of life. Nature Communications. 2023;14(1):7456.

28. Weiss LM. Microsporidiosis. Hunter's tropical medicine and emerging infectious diseases. 2020;825–31.

29. McDonald BR, Currie CR. Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus Streptomyces. mBio. 2017 Jun 6;8(3):10.1128/mbio.00644-17.

30. Blaha GM, Polikanov YS, Steitz TA. Elements of ribosomal drug resistance and specificity. Curr Opin Struct Biol. 2012 Dec;22(6):750–8.

31. Vila-Sanjurjo A, Squires CL, Dahlberg AE. Isolation of kasugamycin resistant mutants in the 16 S ribosomal RNA of Escherichia coli. Journal of molecular biology. 1999;293(1):1–8.

32. Akshay S, Bertea M, Hobbie SN, Oettinghaus B, Shcherbakov D, Böttger EC, et al. Phylogenetic sequence variations in bacterial rRNA affect species-specific susceptibility to drugs targeting protein synthesis. Antimicrobial Agents and Chemotherapy. 2011 Aug 17;55(9):4096–102.

33. Polikanov YS, Osterman IA, Szal T, Tashlitsky VN, Serebryakova MV, Kusochek P, et al. Amicoumacin A inhibits translation by stabilizing mRNA interaction with the ribosome. Molecular cell. 2014;56(4):531–40.

34. Prokhorova I, Altman RB, Djumagulov M, Shrestha JP, Urzhumtsev A, Ferguson A, et al. Aminoglycoside interactions and impacts on the eukaryotic ribosome. Proc Natl Acad Sci USA [Internet]. 2017 Dec 19 [cited 2024 Sep 9];114(51). Available from: https://pnas.org/doi/full/10.1073/pnas.1715501114

35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic acids research. 2000;28(1):235–42.

36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 2013;30(4):772–80.

37. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Molecular biology and evolution. 2021;38(7):3022–7.

38. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, et al. TimeTree 5: an expanded resource for species divergence times. Molecular Biology and Evolution. 2022 Aug 1;39(8):msac174.