



OPEN

The overlooked evolutionary dynamics of 16S rRNA revises its role as the “gold standard” for bacterial species identification

Oldřich Bartoš^{1✉}, Martin Chmel^{1,2} & Iva Swierczková¹

The role of 16S rRNA has been and largely remains crucial for the identification of microbial organisms. Although 16S rRNA could certainly be described as one of the most studied sequences ever, the current view of it remains somewhat ambiguous. While some consider 16S rRNA to be a variable marker with resolution power down to the strain level, others consider them to be living fossils that carry information about the origin of domains of cellular life. We show that 16S rRNA is clearly an evolutionarily very rigid sequence, making it a largely unique and irreplaceable marker, but its applicability beyond the genus level is highly limited. Interestingly, it seems that the evolutionary rigidity is not driven by functional constraints of the sequence (RNA–protein interactions), but rather results from the characteristics of the host organism. Our results suggest that, at least in some lineages, Horizontal Gene Transfer (HGT) within genera plays an important role for the evolutionary non-dynamics (stasis) of 16S rRNA. Such genera exhibit an apparent lack of diversification at the 16S rRNA level in comparison to the rest of a genome. However, why it is limited specifically and solely to 16S rRNA remains enigmatic.

Ever since 1977, which was a breakthrough year for the invention of the so-called first-generation or Sanger sequencing^{1,2}, the attention of leading evolutionary biologists and taxonomists has been focused on 16S rRNA^{3,4}. 16S, 23S and 5S rRNAs are essential genes that typically constitute a chromosomal rRNA operon⁵. It is the only component that is universal to all self-replicating organisms and its sequences change only slowly over time³. Prior to these advances, it was practically impossible to achieve any valid phylogeny, especially in microbiology, but the advent of both sequencing techniques and the discovery of the potential of 16S rRNA essentially changed the world of microbiology⁶. Since then, (not only) microbiology has built on and, in many ways, relied on 16S rRNA as a universal and reliable marker for species identification and delimitation^{7–13}.

16S rRNA is also widely used in clinical practice, where it has served as a powerful tool for bacterial identification and diagnostics for decades^{14,15}. Although today's clinical practice uses preferentially Mass Spectrometry for diagnostic purposes, especially MALDI TOF instruments^{16,17}, sequencing procedures targeting 16S rRNA remain an important part of the portfolio of microbiology laboratories¹⁵. This is because while MALDI TOF usually provides fast and correct identification, in a significant number of cases it fails to provide any valid information^{15,18}. Whereas 16S rRNA sequencing always gives us at least some idea of the phylogenetic classification of a given organism/pathogen¹⁵.

The reason we began to doubt the specificity of 16S rRNA was the routine identification of an unknown fish pathogen. Both types of analyses were performed, but while the 16S rRNA classification quite clearly identified the sample as a common bacterial species, the MALDI TOF classification failed. We decided to resolve this apparent discrepancy by whole-genome sequencing and detailed characterization of the sample. It turned out as a new species whose average nucleotide identity (ANI) of the genome with the nearest described species was only about 82.5%, while the threshold for describing a new species is reported to be around 95%¹⁹. Therefore, we started to investigate why two evolutionarily well-separated entities essentially share the same copy of 16S rRNA.

Despite that 16S rRNAs are thought to be species-specific²⁰, and the assumptions that genes/molecules involved in complex interactions (such as ribosomes) should not be subject to Horizontal Gene Transfer (HGT)²¹, several studies have reported cases of HGT of 16S rRNAs and some have even evaluated the (in)significance of

¹Military Health Institute, Military Medical Agency, 16200 Prague, Czech Republic. ²Department of Infectious Diseases, First Faculty of Medicine, Charles University and Military University Hospital Prague, 12108 Prague, Czech Republic. ✉email: 124600@seznam.cz

this phenomenon on the viability of a given organism^{20,22,23}. Particularly for this type of sequence, which is typically found in multiple copies in a genome, the availability of a complete genome sequence is crucial for HGT evaluation, which has been facilitated by the advent of third-generation sequencing technologies and platforms²⁴. On the other hand, difficult-to-detect HGT within a bacterial species or between closely related species is probably an important driving force in the evolution of microorganisms²⁵. Recently, some studies have emphasized that 16S rRNA, especially in a phylogenetic reconstruction/estimation context, provides inaccurate results, suggesting the involvement of HGT^{26–28}.

Here we show that the relationship between 16S rRNA and species delimitation and classification is likely to be more complex than previously reported. Furthermore, we focus on the question of why and how it actually resists something that, especially in biology, is considered to be one of the basic definitions of life, i.e. change/evolution²⁹. From this point of view, it seems more appropriate to consider them rather as living fossils carrying information about the origin of the domains of cellular life^{4,30–32}. Further, the role of functional constraints on the evolutionary rigidity of 16S rRNAs has been largely refuted by the study of mitochondrial 16S rRNAs, whose evolutionary dynamics does not significantly differ from those of typical nuclear genes^{33–35}. Therefore, we studied and evaluated the evolutionary dynamics of 16S rRNA in more than 15 bacterial genera comprising over 1,200 species. Further, we also extended this analysis to some representatives of the Archaea and Eukaryotic domains. Next, we documented the manifestations of intra-specific evolutionary dynamics of 16S rRNA in a well-studied species, *Escherichia coli*, whose explanation requires the effective action of HGT and/or concerted evolution.

Results and discussion

The performance of 16S rRNA as a marker for the identification of bacterial species

First, to gain insight into how 16S rRNA corresponds to overall divergence at the genome level, we estimated a matrix of evolutionary distances within each single bacterial genus. We used two different methods to estimate evolutionary distances for pairwise comparisons between individual species (see Fig. 1a). First, we used average nucleotide identity (ANI), which is a useful estimate especially for discrimination of closely related species/lineages^{19,36}. In contrast, protein-based phylogenomic Maximum Likelihood distances are more useful for disentangling and inferring deeper evolutionary relationships³⁷. We initially focused on how and whether divergence at the 16S rRNA level actually reflects species boundaries, defined as ~95% divergence in the genome sequence.

To our surprise, immediately after the analysis of the first bacterial genus, it was apparent that the divergence of 16S rRNA between "relatively" closely related species is an essentially non-existent phenomenon (see Fig. 1b, c). Moreover, we instantly identified a case where two species shared a basically identical 16S rRNA variant (>99.9% identity), although at the whole genome level they clearly represented evolutionarily well-separated entities (see Fig. 1b). We show that this is in fact not as rare a phenomenon as one would assume, however, beyond this we show that individual genera show significant differences in the evolutionary dynamics of 16S rRNA (see Fig. 1d). We analyzed 15 major bacterial genera comprising over 1200 species (see Table 1), with each species represented by only one reference sequence (see Supplementary Table 1). Despite such a limited scope, we were able to detect over 175 such cases where two well-differentiated species possess essentially identical copies of 16S rRNA. These analyses were performed within individual bacterial genera, but when we similarly analyzed the data on an inter-generic scale, we found no clear evidence to suggest any recent HGT. These results generally question, at least to some extent, the applicability of 16S rRNA as a species-specific or even strain-specific marker⁹, as well as its suitability for phylogenetic reconstructions of closely related taxa¹².

The rate of evolution of 16S rRNA

The rate of 16S rRNA evolution between relatively closely related species, i.e. ranging from 5% up to 20% divergence, is generally extremely low compared to other genes or the rest of the genome as a whole. 16S rRNA is truly exceptional due its ability to at least seemingly resist something that is perfectly natural in terms of evolutionary biology, namely change²⁹. Virtually identical evolutionary trends can also be observed for 23S rRNA (see Supplementary Fig. 1), but for simplification and its uttermost significance, we will consider only 16S rRNA for the purposes of this text. But how is it possible that this particular sequence maintains a significantly lower mutation rate than the rest of the genome? For a long time, evolutionary biology has been concerned with the concept of Essentiality, which suggests that essential genes, such as 16S rRNA, should evolve slower than more dispensable ones due to increased selection pressure³⁸. However, until recently, attempts to confirm this theory have remained ambiguous, suggesting quite clearly that so-called Essentiality affects the rate of evolution in no fundamental way^{38,39}. Even in this perspective, it is clear that 16S rRNA represents an entirely exceptional example in its evolutionary rigidity.

But what does this rigidity stem from? When, on the one hand, it has been experimentally proven that bacteria can tolerate both recombinant and even foreign copies of 16S rRNA without much difficulty^{20,23}. On the other hand, the 16S rRNA, which is encoded by the mitochondrial genome and is the basis for the formation of the mitochondrion and thus retains the same function as its bacterial counterpart, lacks this unique characteristic^{33,34}. Actually, mitochondrial 16S rRNA even lacks some otherwise conserved structural elements whose role has been taken over by ribosomal proteins encoded in the eukaryotic genome³⁵. Perhaps most interesting is the fact that nuclear mito-ribosomal proteins mirror the increased evolutionary rate of the mitochondrial genome. As a result, nuclear mito-ribosomal proteins evolve more than 10 times faster than cyto-ribosomal proteins, despite being part of the same nuclear genome^{33–35}. These facts show quite clearly that the Essentiality along with the assumed evolutionary constraints it implies do not in themselves provide a satisfactory explanation of the observed phenomenon, i.e. extremely slow evolution rate of 16S rRNA.

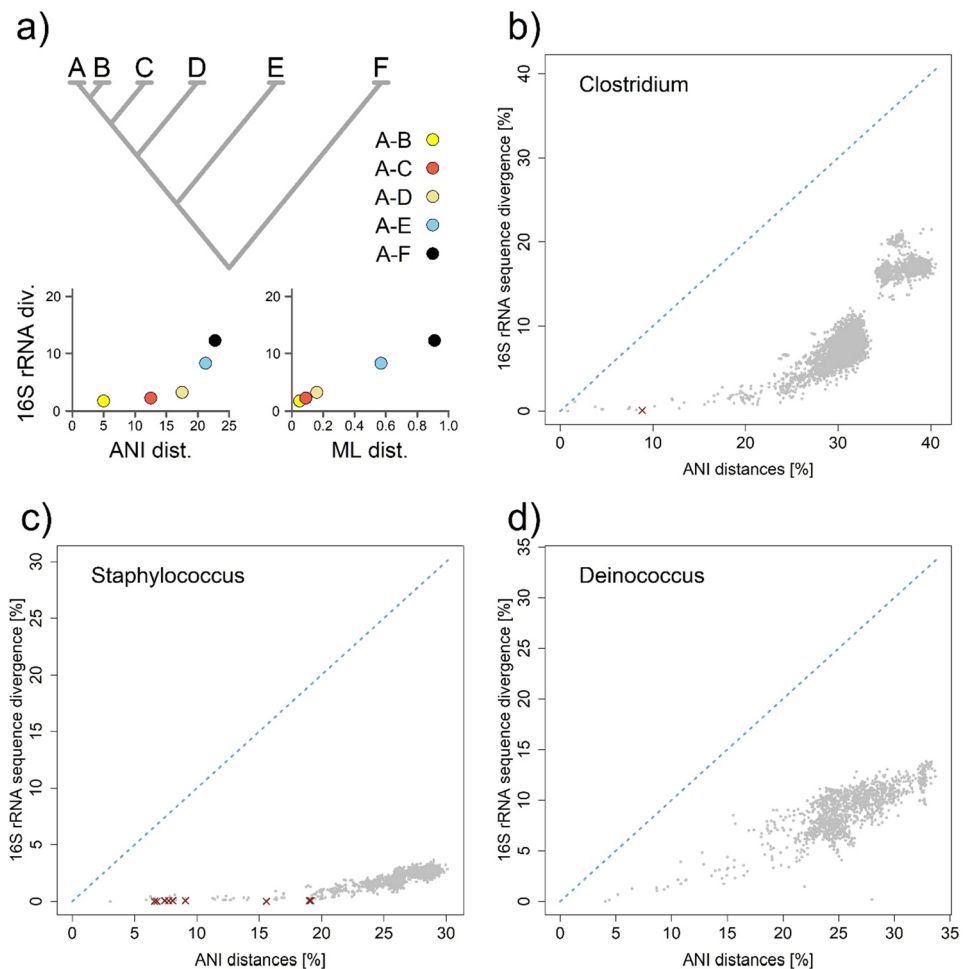


Figure 1. Comparison of evolutionary (substitution) rates in selected bacterial genera as a function of Average Nucleotide Identity (ANI) estimates: **(a)** schematic representation of a phylogenetic tree with varying degrees of evolutionary distances to the hypothetical species A. All species pairs involving species A are shown in illustrative plots comparing 16S rRNA sequence divergence in response to genomic divergence. We measured genomic diversity in two different ways, using either average nucleotide identity (ANI) or maximum likelihood (ML) phylogenetic distances. While ANI is useful for distinguishing (relatively) closely related species or isolates, whereas ML distances are a more appropriate measure for macro-evolutionary scales. **(b)** *Clostridium*; **(c)** *Staphylococcus*; **(d)** *Deinococcus*. Individual species-to-species comparisons are shown in gray, dark crosses indicate species pairs that share essentially the same copy of 16S rRNA (identity > 99.9%), despite being evolutionarily well-separated entities. Graphs of all taxa studied can be reviewed in Supplementary Fig. 3.

Differences between bacterial genera

One of the unique features of 16S rRNA is that it is typically found in multiple copies per genome. Copy number is usually a relatively stable characteristic of a given genus, but there are notable differences between genera, and a typical bacterial species such as *E. coli* contains about 7 copies on average⁴⁰. In theory, genes with multiple copies, such as 16S rRNA, should be subject to significantly stronger negative selection than single-copy genes⁴¹. Thus, we tested whether the rate of 16S rRNA evolution depends on the number of copies of 16S rRNA contained in the average species of each genus. In this comparison, we also included the prokaryotic genus *Thermococcus*, which is the only one of the selected genera to contain only a single copy of 16S rRNA. Unlike the first analysis, where we were mainly interested in separating and visualizing closely related species around the imaginary (5%) species boundary, this time we used Maximum Likelihood (ML) distances, which are a more reliable measure especially in relation to deeper phylogenetic relationships (see Fig. 2). Interestingly, it turned out that the copy number of 16S rRNA has no obvious effect on the rate of its evolution (at least in bacteria) (see Fig. 2a), but it led to two somewhat unexpected findings.

The first and more obvious finding is that there are significant differences between the genera, but they do not correspond at all to the copy-number hypothesis, which is in line with an earlier study⁴². The second finding is that, despite quite reasonable expectations, we have observed a rather significant deviation from the linearity of mutation accumulation (evolution) over time in some genera (compare Fig. 2b with c, d). The expectation of linearity follows both from the so-called Neutral theory of molecular evolution^{43,44} and from actual experimental data^{39,45}. Since these are very conserved sequences, the expected linear relationship cannot be disrupted even

Genus/taxon	Species	HGT pairs	16S rRNA copies	SCO genes
<i>Clostridium</i>	102	1	4 (1; 9,5)	86
<i>Streptococcus</i>	106	1	4 (1; 5)	311
<i>Bacillus</i>	107	30	4,5 (1; 10)	98
<i>Corynebacterium</i>	137	0	4 (1; 4)	197
<i>Bartonella</i>	43	2	2 (1; 2)	369
<i>Staphylococcus</i>	57	9	6 (5; 6)	795
<i>Burkholderia</i>	44	7	4 (1; 6)	1141
<i>Rhizobium</i>	80	52	1 (1; 3)	804
<i>Acinetobacter</i>	83	0	4 (1; 7)	756
<i>Mycobacterium</i>	119	9	1 (1; 1)	421
<i>Vibrio</i>	123	1	8 (1; 11)	245
<i>Nocardia</i>	99	2	1 (1; 1)	535
<i>Leptospira</i>	69	62	1 (1; 2)	1068
<i>Deinococcus</i>	54	0	1 (1; 3)	518
<i>Legionella</i>	16	0	4 (3; 4)	404
Subtotal	1240	176	NA	NA
<i>Thermococcus</i>	31	3	1 (1; 1)	720
<i>Aves</i>	41	0	NA	147
<i>Actinopterygii</i>	35	0	NA	115
Total	1348	179	NA	NA

Table 1. List of genera/taxa included in this study. For each taxon we report: the number of species; the number of detected horizontal gene transfers (HGT), i.e. species pairs sharing the same variant of 16S rRNA (> 99.9% identity); median of copy number of 16S rRNA with interquartile ranges; number of detected single-copy orthologous (SCO) genes used to estimate the phylogenetic distances.

by substitution saturation⁴⁶. We suggest that the most likely explanation for this phenomenon, barring a mistake, may be a relatively high level of elusive HGT⁴², especially among closely related species, which stabilizes a particular 16S rRNA phenotype/genotype at the genus level. This is particularly evident in the aforementioned prokaryotic genus *Thermococcus*, which, despite expectations, shows one of the lowest diversifications at the 16S rRNA level. And it is mainly extremophilic organisms in which HGT is often mentioned not only as a means of adaptation, but especially as a necessity for maintaining genome integrity^{25,47}. In general, it is known that the susceptibility of different bacterial genera to HGT varies and is probably related to their life-history traits⁴⁸.

Comparison of bacteria with vertebrates

To assess the potential role of HGT, we needed to obtain data from organisms for which it can be ruled out a priori. We therefore chose representatives of vertebrates, i.e. fish (*Actinopterygii*) and birds (*Aves*), which are characterized by relatively small genomes⁴⁹. Unlike bacteria, vertebrates have two functional equivalents of bacterial 16S rRNA, the first being eukaryotic cytoplasmic 18S rRNA and the second being mitochondrial 16S rRNA.

In contrast to some bacterial genera, the evolutionary (substitution) rate of 18S rRNA gene of the vertebrates follows the expected linear model (see Fig. 3a), i.e. linear regression fits the data better than a second-order polynomial model. The evolutionary rate of mito-ribosomal 16S rRNA is much higher and the only thing that slows it down, at least optically, is substitution saturation (see Fig. 3a). Figure 3b shows an overall comparison of 16S/18S rRNA evolution models from selected bacterial genera as well as data from vertebrates. In summary, in organisms for which we have ruled out a priori the effective action of HGT, we can state that evolution (mutation accumulation) proceeds linearly in evolutionary time, i.e. in accordance with the theoretical expectations. Quite surprisingly, the substitution rate of 18S rRNA is higher than that of the bacterial genera examined, despite the fact that 18S rRNA is considered to provide only low-level taxonomic resolution in vertebrates⁵⁰. Instead, mitochondrial rRNA is often used for accurate identification at the species level⁵¹.

Intraspecific 16S rRNA evolutionary dynamics

The most convincing evidence of HGT and/or concerted evolution was obtained by analyzing ~ 3700 'complete' genomes of *Escherichia coli* isolates. We found that there are at least three distinct 16S rRNA variants/genotypes within these genomes that possess the following key characteristics: (1) they are clearly distinguishable from each other; (2) they are not linked to specific phylogenetic lineages; (3) and, most importantly, we are able to find genomes in which only one variant is represented, but also those in which these variants are combined in different ways.

We screened thousands of genomes and looked for variability within 16S rRNA copies of individual genomes. In most cases, the variability assessed by BLAST resulted mainly from deletion-affected and hence possibly non-functional copies. However, within *E. coli* isolates, we thus revealed that these genomes host at least three well-defined 16S rRNA variants (see Fig. 4a). We selected three groups of representative genomes for which we were able to confirm that all (seven) copies of 16S rRNA clearly matched only one of the identified reference

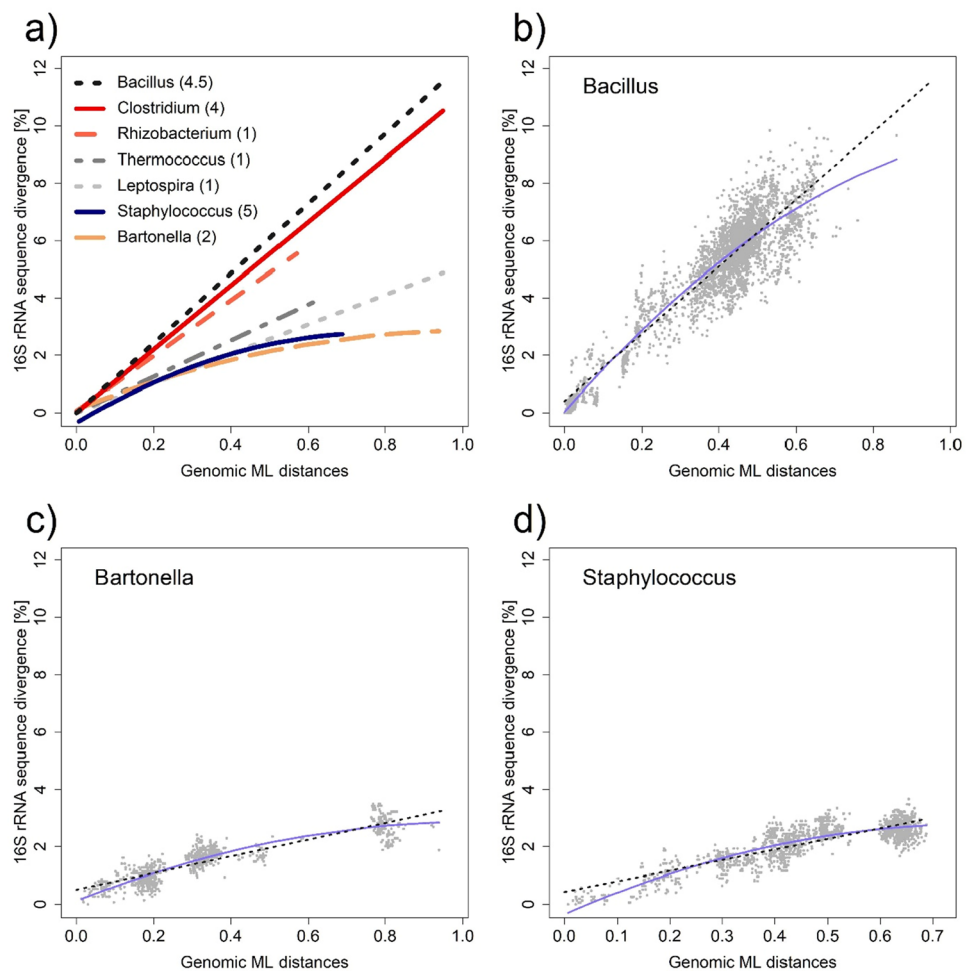


Figure 2. Comparison of evolutionary (substitution) rates in selected bacterial genera as a function of genomic maximum likelihood (ML) evolutionary distances: **(a)** comparison of selected (bacterial) genera, each represented by a linear regression or second order polynomial model. For all linear regressions in this graph, we deliberately set the intercept to zero for better readability. We preferred the polynomial regression model where it fit the data significantly better than linear regression, both statistically and especially visually. Numbers in parentheses represent the median 16S rRNA copy number for a given bacterial genus. Note that we did not observe any apparent difference in the rate of mutation accumulation at the level of 16S rRNA between genera with higher or lower copy number of this sequence/gene. **(b)** *Bacillus*, **(c)** *Bartonella*, **(d)** *Staphylococcus*. Individual species-to-species comparisons are shown in gray. The black dashed line represents the fitted linear regression model, while the blue solid line represents the second-order polynomial function/model. Graphs of all taxa studied can be reviewed in Supplementary Fig. 4.

variants. We then estimated a phylogenomic tree that clearly showed that the possession of one or another variant is essentially randomly distributed across the entire phylogeny (see Fig. 4b). Therefore, we can exclude that these are specific variants linked to phylogenetically lower taxa, e.g. subspecies. While the simple presence of multiple variants could be explained, for example, as Incomplete Lineage Sorting (ILS)⁵², the presence of genomes/strains with(out) a mixture of variants cannot be satisfactorily explained without the involvement of HGT or efficient concerted evolution. In fact, we identified virtually all possible combinations of the full genome set of 16S rRNA copies involving a combination of variants A and B, as well as B and C (see Fig. 4c, d). Although in this case we cannot rule out that this is intraspecific variation and not a manifestation of HGT, these data demonstrate the power and importance of concerted evolution for the evolutionary dynamics of 16S rRNA.

In fact, this variability at the 16S rRNA level has been described previously and its significance appears to be biologically relevant⁵³; but see⁵⁴. In particular, transcriptional upregulation of a variant referred to as rssh 16S rRNA (referred to here as variant B) is associated with a general stress response, activation of starvation related gene pathways and increased antibiotic resistance⁵³. While transcription and not translation is considered the major controlling element of gene expression, it turns out that even translation and therefore ribosomes as such can play an important regulatory role^{53,55}.

The rssh 16S rRNA variant has been described in the model strain K-12, in which it typically occurs only in single copy⁵³. Therefore, we examined genomic metadata deposited at NCBI to determine whether the

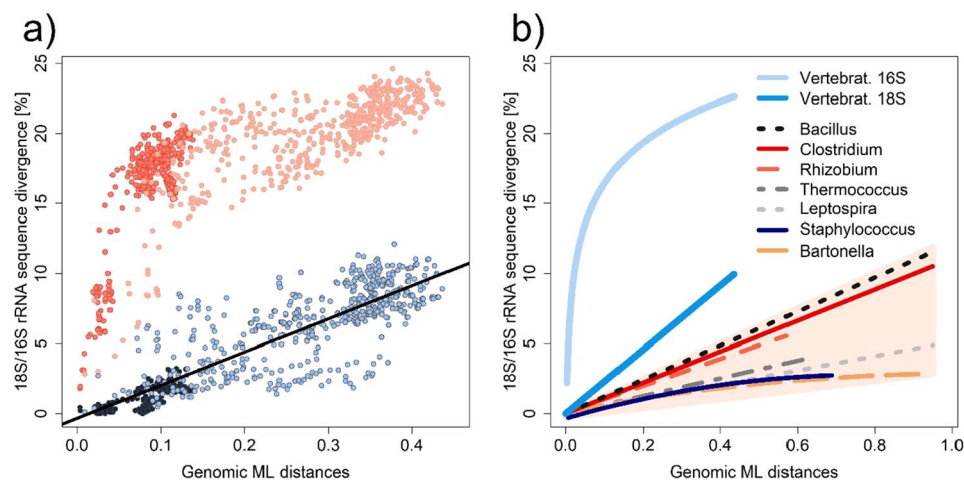


Figure 3. Comparison of evolutionary (substitution) rates between vertebrate eukaryotic 18S rRNA and mitochondrial 16S rRNA: **(a)** Data for 18S rRNA are shown in shades of blue, whereas mitochondrial 16S rRNA are shown in shades of red. The darker shade always represents the group of birds (Aves), while the lighter shade represents the group of selected fish (Actinopterygii). Linear regression is shown for the 18S rRNA data (black). The data are shown separately in Supplementary Fig. 5; **(b)** comparison of evolutionary (substitution) rates between selected bacterial genera and vertebrate eukaryotic organisms, respectively their 18S rRNA and mitochondrial 16S rRNA. The bacterial genera are the same as in Fig. 2a and are shown on a beige background. Eukaryotic 18S rRNAs are represented by the linear model in bright blue, whereas mitochondrial 16S rRNAs are represented by a logarithmic regression model shown in light blue.

documented accumulation of copies of this particular 16S rRNA variant is associated with any specific location or phenotype. We found that enrichment of this particular 16S rRNA variant is tightly associated with the enterohemorrhagic *E. coli* (EHEC) serotype O157:H7 (see Supplementary Fig. 2), which has been reported in many countries worldwide⁵⁶. EHEC serotype O157:H7 was first recognized in 1982 and is considered a major public health concern as the causative agent of hemorrhagic colitis and life-threatening hemolytic uremic syndrome in humans⁵⁶. Transmission usually occurs through consumption of contaminated food or water, and therefore serotype O157:H7 is considered a food-borne pathogen⁵⁶. Interestingly, serotype O157:H7 has been described as having a survival capacity far exceeding that of common commensal strains, allowing it to survive the harsh conditions frequently encountered in the human food chain⁵⁷.

Conclusions

In this study, we have clearly demonstrated why it is not reasonable to rely on 16S rRNA as a species-specific marker. Undoubtedly, 16S rRNA remains a valuable marker due to its unique properties, but we must be aware of its limitations. Tools that overcome these limitations already exist today, such as Metagenomic Shotgun Sequencing⁵⁸, but their wider adoption outside the high-end research environment cannot be expected in the near future.

However, we have shown that although 16S rRNA can be considered one of the most sequenced and studied sequences, its unique evolutionary dynamics has long been overlooked. We have demonstrated that both Horizontal Gene Transfer (HGT) and concerted evolution can play a significant role in the evolution of 16S rRNA.

While the effects of HGT on the evolutionary dynamics of 16S rRNA at the inter-specific level seem to be clear, the role of concerted evolution is somewhat ambiguous. In theory, the effective action of concerted evolution at the species level should counteract inter-specific HGT⁵⁹, at least unless other processes such as selection or molecular drive are involved^{59,60}. On the other hand, 16S rRNA sequences of closely related taxa are similar to such an extent that their fixation/loss can be driven by genetic drift (chance) alone.

However, why HGT is so remarkably exhibited at the level of 16S rRNA remains enigmatic. Especially considering the fact that we have ruled out that this could be related to its most obvious feature, i.e. its multiple representation in the genome.

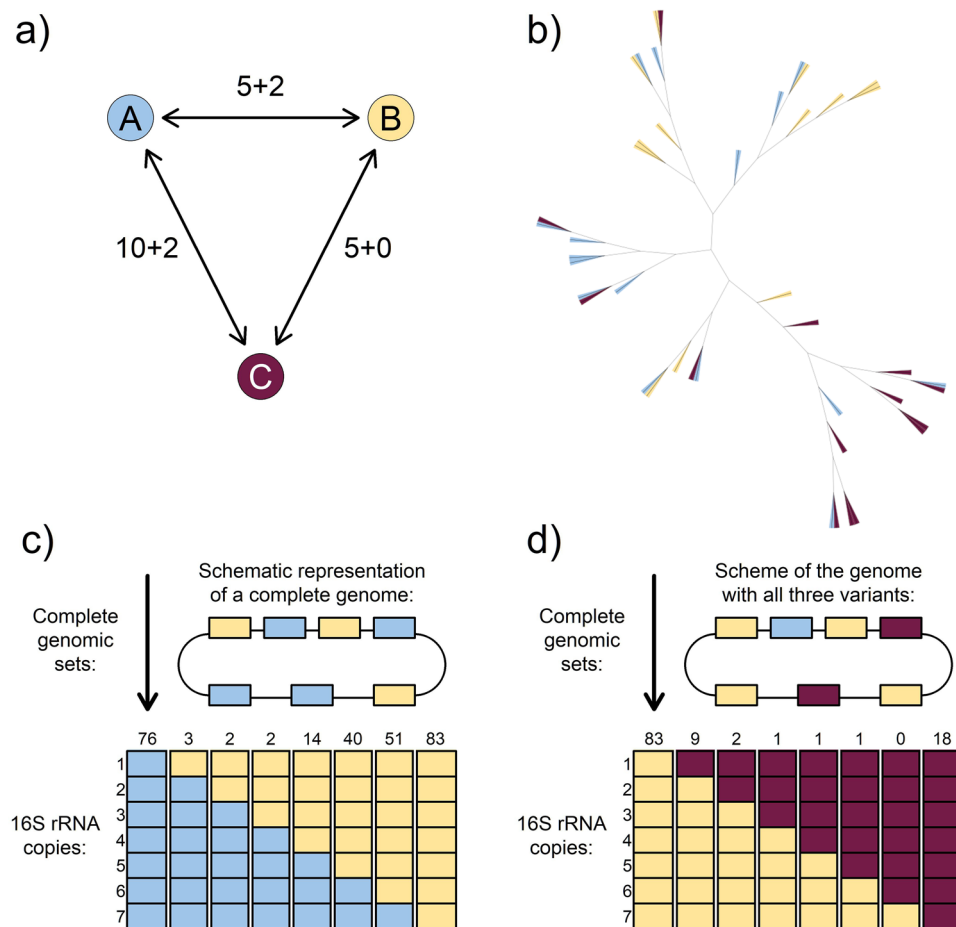


Figure 4. (a) Schematic representation of discovered variants of 16S rRNA in *Escherichia coli* genomes. The numbers along the arrows indicate the number of mismatches + gap openings. (b) Schematic phylogenomic tree of *Escherichia coli* genomes in which exclusively one of the 16S rRNA reference variants was identified. (c) Schematic representation of detected genomes with mixed representation of variants A and B. Numbers above the columns indicate the number of detected genomes with a given ratio/number of variants. (d) Schematic representation of detected genomes with mixed representation of variants B and C.

Data availability

This study was based on the analyses of publicly available genomic data from NCBI repositories, their accession numbers are provided in Supplementary Tables 2, 3 and 4.

Received: 13 November 2023; Accepted: 12 April 2024

Published online: 20 April 2024

References

1. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**, 5463–5467 (1977).
2. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).
3. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* **74**, 5088–5090 (1977).
4. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proc. Natl. Acad. Sci.* **87**, 4576–4579 (1990).
5. Anda, M. *et al.* Bacteria can maintain rRNA operons solely on plasmids for hundreds of millions of years. *Nat. Commun.* **14**(1), 7232 (2023).
6. Olsen, G. J. & Woese, C. R. Ribosomal RNA: A key to phylogeny. *FASEB J.* **7**, 113–123 (1993).
7. Teng, F. *et al.* Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci. Rep.* **8** (2018).
8. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10** (2019).
9. Jeong, J. *et al.* The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. *Sci. Rep.* **11** (2021).
10. Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D. H. & Soo, R. M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J.* **15**, 1879–1892 (2021).

11. Dueholm, M. K. D. *et al.* MiDAS 4: A global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. *Nat. Commun.* **13** (2022).
12. Zhang, J. *et al.* Phylogenetic analysis of *Arthrospira* strains from Ordos based on 16S rRNA. *Sci. Rep.* **12** (2022).
13. Chen, Z. *et al.* Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. *mSystems* **4**, 13 (2019).
14. Shang, S., Chen, G., Wu, Y., Du, L. & Zhao, Z. Rapid diagnosis of bacterial sepsis with PCR amplification and microarray hybridization in 16S rRNA gene. *Pediatr. Res.* **58**, 143–148 (2005).
15. Church, D. L. *et al.* Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. *Clin. Microbiol. Rev.* **33**(4), 10–1128 (2020).
16. Cuénod, A., Foucault, F., Pflüger, V. & Egli, A. Factors associated with MALDI-TOF mass spectral quality of species identification in clinical routine diagnostics. *Front. Cell. Infect. Microbiol.* **11** (2021).
17. Alizadeh, M. *et al.* MALDI-TOF mass spectroscopy applications in clinical microbiology. *Adv. Pharmacol. Pharmaceut. Sci.* **2021**, 1–8 (2021).
18. Singhal, N., Kumar, M., Kanaujia, P. K. & Virdi, J. S. MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Front. Microbiol.* **6** (2015).
19. Lee, I., Ouk Kim, Y., Park, S. C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **66**, 1100–1103 (2016).
20. Miyazaki, K. & Tomariguchi, N. Occurrence of randomly recombined functional 16S rRNA genes in *Thermus thermophilus* suggests genetic interoperability and promiscuity of bacterial 16S rRNAs. *Sci. Rep.* **9**, 11233 (2019).
21. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc. Natl. Acad. Sci.* **96**, 3801–3806 (1999).
22. Tian, R. M., Cai, L., Zhang, W. P., Cao, H. L. & Qian, P.-Y. Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene. *Genome Biol. Evol.* **7**, 2310–2320 (2015).
23. Kitahara, K., Yasutake, Y. & Miyazaki, K. Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **109**, 19220–19225 (2012).
24. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* **19**, R227–R240 (2010).
25. Colnaghi, M., Lane, N. & Pomiankowski, A. Repeat sequences limit the effectiveness of lateral gene transfer and favored the evolution of meiotic sex in early eukaryotes. *Proc. Natl. Acad. Sci.* **119**, 220504119 (2022).
26. Kitahara, K. & Miyazaki, K. Revisiting bacterial phylogeny. Mobile genetic. *Elements* **3**, e24210 (2013).
27. Hassler, H. B. *et al.* Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. *Microbiome* **10**, 104 (2022).
28. Caudill, M. T. & Brayton, K. A. The use and limitations of the 16S rRNA sequence for species classification of anaplasma samples. *Microorganisms* **10**, 605 (2022).
29. Higgs, P. G. Chemical evolution and the evolutionary definition of life. *J. Mol. Evol.* **84**, 225–235 (2017).
30. Roberts, E., Sethi, A., Montoya, J., Woese, C. R. & Luthey-Schulten, Z. Molecular signatures of ribosomal evolution. *Proc. Natl. Acad. Sci.* **105**, 13953–13958 (2008).
31. Hsiao, C., Mohan, S., Kalahar, B. K. & Williams, L. D. Peeling the onion: Ribosomes are ancient molecular fossils. *Mol. Biol. Evol.* **26**, 2415–2425 (2009).
32. Bowman, J. C., Petrov, A. S., Frenkel-Pinter, M., Penev, P. I. & Williams, L. D. Root of the tree: The significance, evolution, and origins of the ribosome. *Chem. Rev.* **120**, 4848–4878 (2020).
33. Pietromonaco, S. F., Hessler, R. A. & O'Brien, T. W. Evolution of proteins in mammalian cytoplasmic and mitochondrial ribosomes. *J. Mol. Evol.* **24**, 110–117 (1986).
34. O'Brien, T. Properties of human mitochondrial ribosomes. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **55**, 505–513 (2003).
35. Petrov, A. S. *et al.* Structural patching fosters divergence of mitochondrial ribosomes. *Mol. Biol. Evol.* **36**, 207–219 (2018).
36. Kim, M., Oh, H. S., Park, S. C. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **64**, 346–351 (2014).
37. Baldauf, S. L. Phylogeny for the faint of heart: A tutorial. *Trends Genet.* **19**, 345–351 (2003).
38. Koonin, E. V. Systemic determinants of gene evolution and function. *Mol. Syst. Biol.* **1** (2005).
39. Alvarez-Ponce, D., Sabater-Muñoz, B., Toft, C., Ruiz-González, M. X. & Fares, M. A. Essentiality is a strong determinant of protein rates of evolution during mutation accumulation experiments in *Escherichia coli*. *Genome Biol. Evol.* **8**, 2914–2927 (2016).
40. Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* **186**, 2629–2635 (2004).
41. Mano, S. & Innan, H. The evolutionary rate of duplicated genes under concerted evolution. *Genetics* **180**, 493–505 (2008).
42. Espejo, R. T. & Plaza, N. Multiple ribosomal RNA operons in bacteria; their concerted evolution and potential consequences on the rate of evolution of their 16S rRNA. *Front. Microbiol.* **9**, 1232 (2018).
43. Kimura, M. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* **11**, 247–270 (1968).
44. Ohta, T. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst.* **23**, 263–286 (1992).
45. Bosshard, L., Peischl, S., Ackermann, M. & Excoffier, L. Dissection of the mutation accumulation process during bacterial range expansions. *BMC Genomics* **21** (2020).
46. Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. An index of substitution saturation and its application. *Mol. Phylogenet. Evol.* **26**, 1–7 (2003).
47. van Wolferen, M., Ajon, M., Driessen, A. J. M. & Albers, S.-V. How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles* **17**, 545–563 (2013).
48. Vos, M. The evolution of bacterial pathogens in the Anthropocene. *Infect. Genet. Evol.* **86**, 104611 (2020).
49. Sessions, S. K. Genome size. In *Brenner's Encyclopedia of Genetics*. 301–305 <https://doi.org/10.1016/b978-0-12-374984-0.00639-2> (2013).
50. Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F. & Taberlet, P. DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biol. Lett.* **10**(9), 20140562 (2014).
51. Sato, Y., Miya, M., Fukunaga, T., Sado, T. & Iwasaki, W. MitoFish and MiFish pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Mol. Biol. Evol.* **35**(6), 1553–1555 (2018).
52. Scornavacca, C. & Galtier, N. Incomplete lineage sorting in mammalian phylogenomics. *Syst. Biol.* **syw082** (2016).
53. Kurylo, C. M. *et al.* Endogenous rRNA sequence variation can regulate stress response gene expression and phenotype. *Cell Rep.* **25**(1), 236–248 (2018).
54. Ferretti, M. B. & Karbstein, K. Does functional specialization of ribosomes really exist?. *RNA* **25**(5), 521–538 (2019).
55. Barna, M. Ribosomes take control. *Proc. Natl. Acad. Sci.* **110**(1), 9–10 (2013).
56. Lim, J. Y., Yoon, J. W. & Hovde, C. J. A brief overview of *Escherichia coli* O157: H7 and its plasmid O157. *J. Microbiol. Biotechnol.* **20**(1), 5 (2010).
57. Ameer, M. A., Wasey, A., & Salen, P. *Escherichia coli* (*E. coli* O157 H7) (2018).

58. Nearing, J. T., Comeau, A. M. & Langille, M. G. I. Identifying biases and their potential solutions in human microbiome studies. *Microbiome* **9**, 113 (2021).
59. Santoyo, G. & Romero, D. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* **29**(2), 169–183 (2005).
60. Dover, G. Molecular drive. *Trends Genet.* **18**(11), 587–589 (2002).

Acknowledgements

This study was supported by the Czech Ministry of Defence, project MO1012.

Author contributions

O.B. and M.C. wrote the main manuscript text. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-59667-3>.

Correspondence and requests for materials should be addressed to O.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024