Journal of Phylogenetics & Evolutionary Biology

Wang and Huang, J Phylogenetics Evol Biol 2018, 6:3
DOI: 10.4172/2329-9002.1000204

Research Article Open Access

Assessment of Molecular Markers for Classification of Bacterial Phyla using Topological Dissimilarity of Phylogenetic Trees

Yong Wang[1*] **and Jiao-Mei Huang**[1,2]

[1]*Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan, China*

[2]*University of Chinese Academy of Sciences, Beijing, China*

[*]**Corresponding author:** Yong Wang, Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan, China, Tel: (86) 898-88381062; E-mail: wangy@idsse.ac.cn

Received date: July 17, 2018; **Accepted date:** July 30, 2018; **Published date:** August 03, 2018

Abstract

Single-copy conserved proteins and ribosomal RNA (rRNA) genes are important molecular markers for placement of a new bacterial species into phyla. However, accuracy and consistency of these molecular markers in the classification have not been completely evaluated yet. In this study, 33 highly conserved proteins and three rRNAs were used to construct phylogenetic trees for 19 bacterial phyla. Based on the topological dissimilarity of the trees, formation of taxonomic monophyletic clades of the phyla could be compared among the markers. Our results showed that the trees for conserved proteins and rRNAs are consistent in the classification between the 16S and 23S rRNA genes and ribosomal proteins (r-proteins) (L2, S3, S7, L14, S10 and S12) that are essential to the translation process. To examine the monophyletic sorting efficiency of the markers, phylogenetic clades in the trees were checked for the co-occurrence of taxa from the same phyla. Using translation initiation factor 2, 16S rRNA and 23S rRNA could assign almost all taxa correctly into the monophyletic clades. Taken together, our results suggest that the two rRNAs and several r-proteins may be the candidate molecular markers for accurate classification of bacterial phyla probably due to their involvement in core function of translation process.

Keywords: Molecular marker; Geodesic distance; Bacterial phylum; rRNA gene; Conserved proteins

Introduction

The current, widely accepted framework for bacterial systematics was established based on the similarity of 16S ribosomal RNA (rRNA) genes and other highly conserved genes [1,2]. A bacterial isolate may be classified into different taxonomic groupings by comparison with the sequences in the public databases such as SILVA [3] and RDP [4]. Phylogenomic analysis using concatenated conserved genes is an alternative to locate a bacterium in the tree of life [5,6]. Genomes for novel bacterial phyla were obtained using enrichment cultivation and single amplified genomics. With the growing number of new bacterial phyla in recent years, approaches to the rapid identification and positioning of new taxonomic bacterial groups in a wide range of environments are in high demand. Investigations of the microbial composition of a community with high biodiversity are also challenging because of the lack of comprehensive evaluations of the available markers applied for the taxonomic classification.

A 16S rRNA gene sequence contains both highly conserved regions that can be used for primer design and hyper-variable regions that are for taxonomic positioning of a microorganism [7]. An important issue is that a high copy number of 16S rRNA genes in a certain group of microbes may over-estimate their proportion in the community [1]. Depending on the ecological strategy and genome size, bacteria differ remarkably in the copy number of rRNA operons [8,9]. The wide range of copy numbers, from one up to twelve copies [8], may lead to inaccurate estimates of biodiversity and microbial composition in a sample. An alternative option is to use single-copy conserved genes in bacterial genomes as molecular markers. There are at least 38 universally conserved genes in prokaryotes [5,10]. Their ubiquitous presence in bacterial genomes may permit precise positioning of a bacterial isolate within the systematic phylogeny of all organisms. However, whether the proteins encoded by these conserved genes perform similarly compared to 16S rRNAs as molecular markers in the phylogenetic topology of bacterial phyla remains to be answered. It is possible that the tree of life constructed using some universally conserved proteins differs profoundly from that based on the rRNA genes.

Topological comparison of the phylogenetic trees is a bottleneck problem, because similarity of taxa distribution on the branches of the trees needs to be quantified. An algorithm was developed to project the node and branch structure of a tree into a multi-dimensional model [11]. As such, the topological dissimilarity between phylogenetic trees can be estimated by the geodesic algorithm. Different from all the traditional algorithms for calculation of Euclidean distance, geodesic can identify the shortest connection in multi-dimensional space between the trees with different topologies [12]. The geodesic distance has also been applied to quantify discrepancies between phylogenetic trees [13,14]. Using the geodesic algorithm, it is possible to compare the performance of rRNA genes and single-copy conserved genes as molecular makers. With the quantitative evaluation, the correlative relationships between different markers in terms of classification consistency can be determined.

In the present study, we evaluated rRNA genes and 33 universally conserved genes with the goal of determining 1) topological difference between phylogenetic trees using 16S, 23S and 5S rRNA genes; and 2) whether universally conserved proteins perform similarly as rRNA genes in classification of bacterial phyla. We also examined the presence of the bacterial species from the same phyla in monophyletic

clades, which further displayed the performance of the molecular markers in capability of sorting the bacterial species precisely.

Materials and Methods

Collection of full-length 16S rRNA genes and corresponding conserved genes

Clusters of Orthologous Groups (COGs) [15] were available in the NCBI database (www.ncbi.nlm.nih.gov/COG/). In August of 2015, the file (cog2003-2014.csv) that contained all the COG IDs was obtained from the COG database. Among the list of COGs, the unique protein IDs of 37 highly conserved COGs (essential bacterial genes listed in supplementary material [5]) (Table S2) for all bacteria were pooled. However, only a small fraction of the bacterial genomes in the NCBI contained all of these COGs and full-length rRNA genes because of incompleteness of the genomes.

To obtain bacterial species with full-length 5S, 16S and 23S rRNA genes and a complete set of the highly conserved genes, all completely sequenced bacterial genomes were downloaded from NCBI in GenBank format. There were a total of 2785 complete genomes in August, 2015. In reference to the protein IDs of the COGs, all the conserved genes were searched in the GenBank files. A total of 505 bacterial species with a complete genome contained at least 33 conserved genes. These bacterial species were sorted into corresponding phyla.

For each of the bacterial phyla, three representative genomes that contain the 33 conserved genes were selected randomly, and the species from different orders were preferred. All rRNA genes (5S, 16S and 23S) were then extracted from the genomes. In the case of multiple copies, only one was retained for analysis. The list of species was provided in Table S3. The proteins of the 33 conserved genes were collected in these genomes. The protein sequences of the conserved genes and the DNA sequences of the rRNA genes were aligned with MUSCLE3.5 individually [16], followed by manual adjustments to delete the alignment positions with gaps for more than 50% of sequences.

Construction of Bayesian phylogenetic inference

The aligned sequences of the conserved genes, and full-length 5S, 23S and 16S rRNA genes were used to reconstruct Bayesian phylogenetic relationships. The best substitution model GTR+Gamma +Invariant for DNA sequences and Blosum62+Gamma+Invariant for proteins were recommended in the output of JModelTest 2.1 [17]. Using these models a Bayesian phylogenetic inference was generated with ten million MCMC chains using BEAST 1.8.1 [18]. With a Burn-in setting of 2500 [19], a consensus tree was produced and posterior probabilities on branch points were then calculated.

AHC of phylogenetic trees using the geodesic distance

Next, the geodesic distance algorithm was used to estimate the dissimilarity of the phylogenetic trees for rRNA genes and conserved proteins. The Bayesian trees for the rRNA genes and conserved proteins were converted to those in Newick format for geodesic analysis using the GTP algorithm [13]. Each tree was treated as a variant and was compared with another. The trees for all rRNA genes (5S, 16S and 23S) and the conserved proteins were pooled for the calculation of the pairwise geodesic distance. A distance matrix was constructed using the pairwise distances after GTP analysis. AHC analysis of the rRNA genes and conserved proteins using the distance matrix was conducted in XLSTAT 2010. The complete linkage model was selected for the agglomerative hierarchical clustering (AHC) analysis.

Results

Comparison of classification sensitivity between conserved proteins and rRNA genes

From complete genomes deposited in the NCBI, we collected a total of 56 representative bacterial species of 23 bacterial phyla (subphyla of Proteobacteria were treated as phyla). The three rRNA genes were extracted from the genomes along with 33 COGs (conserved proteins) for reconstruction of phylogenetic trees of the bacterial phyla separately. Using geodesic algorithm, the topological similarity of the trees was quantified and exhibited by AHC. In the AHC result, four clusters were below the primary merging dissimilarity level at 2.1 (Figure 1).

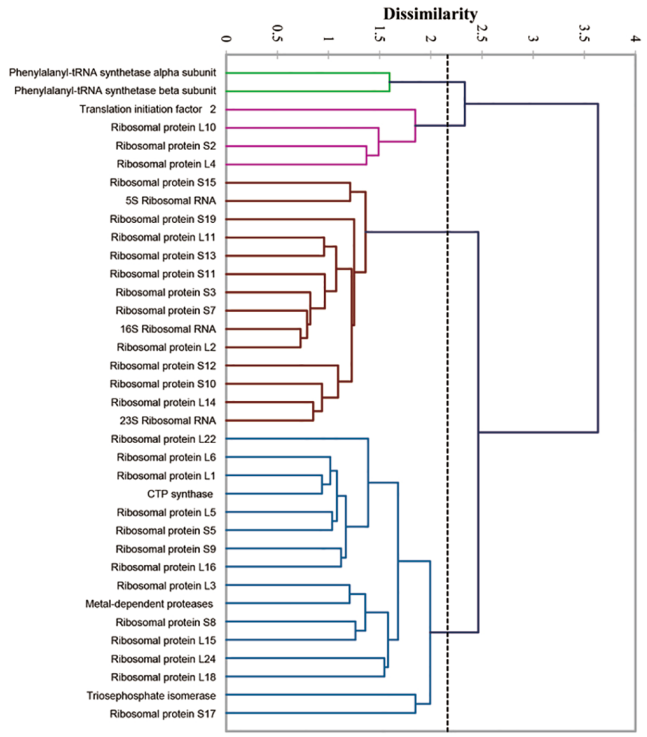

Figure 1: Dissimilarity and AHC of phylogenetic trees for bacterial phyla. The phylogenetic trees for 56 bacterial species using rRNA genes (5S, 16S and 23S) and 33 conserved genes were evaluated in terms of their topological dissimilarity, which was demonstrated in the AHC clustering. The clusters below the primary merging dissimilarity level (dotted line) were shown with different colors.

The first cluster consisted of two subunit genes for phenyl-tRNA synthetase; the neighboring cluster was composed of translation initiation factor 2 and three r-proteins. The remaining proteins and rRNA genes were grouped into two big clusters in our AHC result. The three rRNA genes were located in one cluster, whereas their nearest

neighbors were different. The trees constructed using 5S rRNA genes exhibited the smallest dissimilarity to those for r-proteins S15 and S19, whereas the 16S rRNA genes exhibited affinity to r-proteins L2, S7 and S3. The topological dissimilarity also displayed a closer relationship between 23S rRNA genes and r-proteins L14, S10 and S12. Excluding r-proteins L2 and L14, all these r-proteins were affiliated with the 30S ribosomal subunit. In another large cluster, 13 r-proteins were neighbors of CTP synthase, metal dependent protease (COG0533) and triosephosphate isomerase.

Congruency of phyla in phylogenetic trees

The accuracy of the taxonomic sorting by rRNA genes and conserved proteins may be visualized based on the counting of congruent species from the same phyla in the phylogenetic trees. We counted the occurrence of clades in which species from the same phyla were grouped, and plotted a black-white map for all phylogenetic markers (Figure 2).

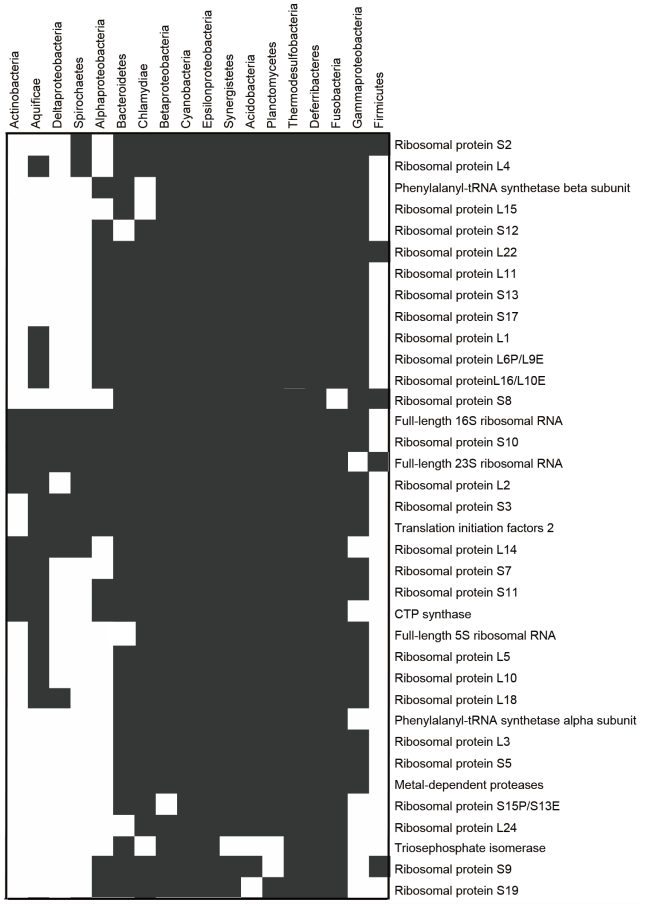

Figure 2: Congruence of bacterial phyla among phylogenetic trees. Taxa from the same phyla were examined 462 in the phylogenetic trees reconstructed using rRNAs and 33 conserved genes. If the monophyletic cluster was present in a phylogenetic tree for a phylum, a black square is depicted.

The full-length 16S and 23S rRNA genes and r-protein S10 almost clustered all species from the same phyla into one clade. The species from only one phylum could not be grouped. For the 16S rRNA gene and r-protein S10, the species from Firmicutes were split into different clades. The performance of other markers was even worse in terms of the monophyletic grouping of species from the same phyla (Figure 2). This might be discerned by the low posterior probabilities on the branches of the phylogenetic trees for these markers, particularly for those branches approaching to the root of the trees. Our result may trigger a debate over selection of optimal markers for species delimitation. Except for 16S rRNA gene, 23S rRNA gene and some essential r-proteins, most of the known proteins probably were not qualified for precise assignment of a bacterial species to a taxonomic rank at a high level.

Surprisingly, translation initiation factor 2 could also accurately set the boundaries between individual phyla, and only species from Actinobacteria and Firmicutes did not remain in the same clades (Figure 2). However, the phylogenetic relationships of the species exhibited a high dissimilarity between translation initiation factor 2 (IF2) and the 16S rRNA genes. The AHC distance of the two clusters in which they was over the minimum emerging threshold line (Figure 1), indicating that the phylogenetic relationships of the bacterial phyla as revealed by the two markers differed remarkably. For all molecular markers, the taxa from Firmicutes, Actinobacteria and Spirochaetes were most difficult to be clustered into monophyletic clades (Figure 2). Our result demonstrated that the species from the class Clostridia could not be grouped with the other two from Firmicutes by most markers, and these species were responsible for the divergence to a high degree. This finding indicates that the phylogenetic depth of Clostridia can be reflected only by full-length 23S rRNA and four molecular markers.

Discussion

In this study, our result indicates coherence of some r-proteins and rRNA genes in the taxonomic sorting accuracy of the bacterial phyla. Interestingly, the three rRNAs are proximate to different r-protein neighbors. Although we used protein sequences of r-proteins and DNA sequences of rRNA genes for the construction of the trees, similar sorting effect was exhibited by this study. This means that equivalent informative loci for the classification were loaded in the r-proteins and neighboring rRNAs regardless of proteins or DNA sequences. This is perhaps attributable to their functional association. We noticed that all of the ribosomal proteins adjacent to the rRNA genes in the AHC result were highly related to translation functions (Table S1). For example, r-protein L2 is a primary rRNA binding protein that is critical for the association of 30S and 50S subunits and for tRNA binding and peptide bond formation [20]. The trees for r-proteins S3 and S7 were also highly similar to that of the 16S rRNA gene, which highlights their importance as r-protein markers for sorting and grouping of bacterial phyla (Figure 1). The two r-proteins bind directly to the head and lower part of the 30S subunit, respectively. The former is involved in unwinding the helical structure of mRNA and is able to position an mRNA in the translation machinery [21]. R-protein S7 is located just above the cleft and decoding site, and it is one of two protein components responsible for assembly initiation of the bacterial small subunit ribosome [22]. S7 is also a major protein component that had been shown to cross-link with tRNA molecules bound at A and P sites [23,24]. As one of the principal regulatory elements, S7 can also control r-protein synthesis by the translational feedback mechanism [25]. Moreover, it can stabilize the decoding center of the 16S rRNA [26]. Therefore, S3 and S7 are heavily involved in translation processes. Their similarity in phylogenetic topology to the 16S rRNA gene

probably stems from numerous contacts between these r-proteins and 16S rRNA. Experimental evidence is required to verify the closer functional association between r-proteins and 16S rRNA.

In the present study, the phylogenetic tree for 23S rRNA revealed a short distance to those trees for L14, S10 and S12 r-proteins. They are another group of molecular markers with high reliability for taxonomic grouping of the phyla. Similarly, the close functional relationship between 23S rRNA and the r-proteins justifies the topological similarity of their trees. L14 is located at the interface of the small and large subunits, together with L2 [27]. Binding of translation silencing factor RsfA on L14 will result in the termination of translation [28]. The ribosomal structure at a resolution of 3.3A° showed that S12 interacts with 23S rRNA and serves as a critical part of the decoding center by modulating tRNA selection in response to streptomycin [28]. S10 is an anti-termination apparatus in the 70S ribosome [29]. It is regulated by r-protein L4 [30], a factor that initiates the assembly of the large subunit [31]. It is interesting that 23S rRNA grouped with S10 rather than L4 in the topological comparison of the phylogenetic trees (Figure 1). This observation implies that some parts of S10 co-evolved with 23S rRNA sites that may form a decoding center. However, further evidence is needed to support this hypothesis. Although L4 is critical for the assembly of the large subunit, our findings indicate that it is not congruent evolutionarily with the 23S rRNA gene.

The 5S rRNA transfers information and coordinates different functional centers in the ribosome [32]. The structure of the 50S ribosomal subunit suggests that it binds to r-proteins L5, L18 and L25 [33]. The topological distance of the phylogenetic trees showed that 5S rRNA was not in the same cluster as L5 and L18 but was closer to the 16S and 23S rRNAs (Figure 1). This result again indicates that structural proximity is not a prerequisite for phylogenetic congruency. The 5S rRNA potentially functions as more than a coordinator and it is likely that an unknown functional importance resulted in its grouping with the rRNA genes and other essential r-proteins.

In this study, not all r-proteins were included in the evaluation. Although some of the r-proteins are also critical in the decoding process, they are not as conserved as the 33 genes in this study. An example is r-protein S1, which also mediates the initiation of translation by unwinding the secondary structure of mRNA and positioning it in the decoding channel [34]. However, r-protein S1 was not consistently present in all bacterial phyla, which excluded their possibility as molecular markers.

Recently, one study took advantage of this method to quantitatively compare phylogenetic trees reconstructed using 38 conserved bacterial genes [5]. The pairwise geodesic distance revealed that the topology of the tree for IF2 is highly similar to the concatenated marker sequences [5]. The result in the present study also implies the importance of IF2 as a molecular marker. However, our result indicates that usage of the IF2 for phylogenetic studies may result in a different bacterial systematics, compared with 16S rRNA.

For Spirochaetes, a recent work has revealed a large genetic distance among different classes [35]. A large number of genetic variations in species from Spirochaetes have probably blurred the informative sites that are useful for the correct taxonomic assignment of different Spirochaetes classes. In summary, our results indicate that partial rRNAs and most r-proteins lack sufficient informative content for completely distinguishing these taxa at the phylum level. Some new phyla, such as Deferribacteres and Planctomycetes, lack a sufficient number of sequenced genomes at lower taxonomic levels. Thus, it may be easier to form a monophyletic core than those with representatives from different classes. Moreover, random selection of the taxa and alignment accuracy rendered difficulties in phylogenetic coherence for taxa from the same phyla in phylogenetic trees.

A considerable percentage of the submissions of proteins and genes to the public databases such as the NCBI are not associated with ascertained taxonomic information. This situation could be improved until the complete genomes of previously undefined phyla were revealed as wrongly assigned taxa. Recently, several novel phyla were discovered and their complete genomes were released [36,5]. This provides an opportunity to further evaluate the molecular markers.

Conclusion

In this study, we examined conserved genes and rRNA genes in terms of their sensitivity and efficiency for splitting bacterial species into corresponding phyla. Several r-proteins and full-length rRNAs may be desirable molecular markers in future studies. Not all markers provided a phylogenetic topology that was consistent with that based on 16S rRNA, suggesting the presence of multiple nomenclature systems in the Bacteria domain. To be cautious, we should develop the current 16S rRNA-based relationships between phyla. The markers suggested in this study require further evaluation in studies of environmental communities and metagenomes as more new phyla and unculturable bacteria are discovered.

Author contributions

Y.W. wrote the manuscript. Y.W. and J.M.H. analyzed the data.

Competing interests

The authors declare no competing financial interests.

Acknowledgements

This study was supported by the National Science Foundation of China No. 31460001 and No. 41476104. This work was also supported by the Strategic Priority Research Program of Chinese Academy of Sciences (CAS) No. XDB06010201 and awards from the Institute of Deep Sea Science and Engineering of CAS (SIDSSE-201206 and SIDSSE-201305) and the National Key Research and Development Program of China (2016YFC0302500).

References

1. Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF, et al. (2007) Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Env Microbiol 73: 278-288.
2. Santos SR, Ochman H (2004) Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environ Microbiol 6: 754-759.
3. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590-D596.
4. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37: 141-145.
5. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499: 431-437.

6. Ropelewski AJ, Nicholas HB, Mendez RRG (2010) MPI-PHYLIP: Parallelizing Computationally Intensive Phylogenetic Analysis Routines for the Analysis of Large Protein Families. Plos ONE 5: e13999.
7. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, et al. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 82: 6955-6959.
8. Fogel GB, Collins CR, Li J, Brunk CF (1999) Prokaryotic genome size and SSU rDNA copy number: estimation of microbial relative abundance from a mixed population. Microb Ecol 38: 93-113.
9. Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of Bacteria. Appl Environ Microbiol 66: 1328-1333.
10. Wu DY, Jospin G, Eisen JA (2013) Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of Bacteria and Archaea and their major subgroups. PloS ONE 8: e77033.
11. Billera LJ, Holmes SP, Vogtmann K (2001) Geometry of the space of phylogenetic trees. Adv Appl Math 27: 733-767.
12. Weyenberg G, Yoshida R (2016) Phylogenetic tree distances. In: Encyclopedia of Evolutionary Biology. Academic Press, Oxford 285-290.
13. Owen M (2011) Computing geodesic distances in tree space. SIAM J Discr Math 25: 1506-1529.
14. Yang B, Wang Y, Qian PY (2016) Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics 17: 135.
15. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28: 33-36.
16. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.
17. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature Meth 9: 772.
18. Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.
19. Condamine FL, Nagalingum NS, Marshall CR, Morlon H (2015) Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. BMC Evol Biol 15: 65.
20. Rippa V, Cirulli C, Di Palo B, Doti N, Amoresano A, et al. (2010) The ribosomal protein L2 interacts with the RNA polymerase α subunit and acts as a transcription modulator in Escherichia coli. J Bacteriol 192: 1882-1889.
21. Takyar S, Hickerson RP, Noller HF (2005) mRNA helicase activity of the ribosome. Cell 120: 49-58.
22. Hosaka H, Nakagawa A, Tanaka I, Harada N, Sano K, et al. (1997) Ribosomal protein S7: a new RNA-binding motif with structural similarities to a DNA architectural factor. Structure 5: 1199-1208.
23. Döring T, Mitchell P, Osswald M, Bochkariov D, Brimacombe R (1994) The decoding region of 16S RNA; a cross-linking study of the ribosomal A,P and E sites using tRNA derivatized at position 32 in the anticodon loop. EMBO J 13: 2677-2685.
24. Sylvers LA, Kopylov AM, Wower J, Hixson SS, Zimmermann RA (1992) Photochemical cross-linking of the anticodon loop of yeast tRNAPhe to 30S-subunit protein S7 at the ribosomal A and P sites. Biochimie 74: 381-389.
25. Dean D, Yates JL, Nomura M (1981) Identification of ribosomal protein S7 as a repressor of translation within the str operon of E. coli. Cell 24: 413-419.
26. Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, et al. (2000) Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102: 615-623.
27. Gao H, Sengupta J, Valle M, Korostelev A, Eswar N, et al. (2003) Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. Cell 113: 789-801.
28. Vila-Sanjurjo A, Lu Y, Aragonez JL, Starkweather RE, Sasikumar M, et al. (2007) Modulation of 16S rRNA function by ribosomal protein S12. Biochimica Biophysica Acta 1769: 462-471.
29. Das A, Ghosh B, Barik S, Wolska K (1985) Evidence that ribosomal protein S10 itself is a cellular component necessary for transcription antitermination by phage λ N protein. Proc Natl Acad Sci USA 82: 4070-4074.
30. Worbs M, Huber R, Wahl MC (2000) Crystal structure of ribosomal protein L4 shows RNA-binding sites for ribosome incorporation and feedback control of the S10 operon. EMBO J 19: 807-818.
31. Nierhaus KH (1991) The assembly of prokaryotic ribosomes. Biochimie 73:739-755.
32. Dinman JD (2005) 5S rRNA: Structure and function from head to toe. Int J Biomed Sci 1: 2-7.
33. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289: 905-920.
34. Duval M, Korepanov A, Fuchsbauer O, Fechter P, Haller A, et al. (2013) Escherichia coli ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. PLoS Biol 11: e1001731.
35. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, et al. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Rev Microbiol 12: 635-645.
36. Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, et al. (2013) Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nature Comm 4: 2120.
37. Nikolay R, van den Bruck D, Achenbach J, Nierhaus KH (2015) Ribosomal proteins: Role in ribosomal functions. eLS 4: a0000687.