Finding priority bacterial ribosomes for future structural and antimicrobial research based upon global RNA and protein sequence analysis

Helena B Cooper [1, 2], **Kurt L Krause** [1], **Paul P Gardner** [Corresp. 1]

[1] Department of Biochemistry, University of Otago, Dunedin, New Zealand

[2] Department of Infectious Diseases, Monash University, Melbourne, Victoria, Australia

Corresponding Author: Paul P Gardner
Email address: paul.gardner@otago.ac.nz

Ribosome-targeting antibiotics comprise over half of antibiotics used in medicine, but our fundamental knowledge of their binding sites is derived primarily from ribosome structures from non-pathogenic species. These include *Thermus thermophilus*, *Deinococcus radiodurans* and the archaean *Haloarcula marismortui,* as well as the commensal and sometime pathogenic organism, *Escherichia coli*. Advancements in electron cryomicroscopy have allowed for the determination of more ribosome structures from pathogenic bacteria, with each study highlighting species-specific differences that had not been observed in the non-pathogenic structures. These observed differences suggest that more novel ribosome structures, particularly from pathogens, are required for a more accurate understanding of the level of diversity of the entire bacterial ribosome, with the potential of leading to innovative advancements in antibiotic research. In this study, high accuracy covariance and hidden Markov models were used to annotate ribosomal RNA and protein sequences respectively from genomic sequence, allowing us to determine the underlying ribosomal sequence diversity using phylogenetic methods. This analysis provided evidence that the current non-pathogenic ribosome structures are not sufficient representatives of some pathogenic bacteria, such as *Campylobacter pylori,* or of whole phyla such as Bacteroidota (Bacteroidetes).

1 Title: Finding priority bacterial ribosomes for future structural and
2 antimicrobial research based upon global RNA and protein sequence
3 analysis

5 Helena B. Cooper[1,2], Kurt L. Krause[1] & Paul P. Gardner[1].
6
7 [1]Department of Biochemistry, School of Biomedical Sciences, University of Otago.
8 [2]Department of Infectious Diseases, Central Clinical School, Monash University.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35   **Author Note**
36

37 Keywords: Bacterial Ribosome, Antibiotic Resistance, Phylogeny Analysis, Genomics.

38
39 We have no known conflicts of interest to disclose.
40
41 Correspondence concerning this article should be addressed to:
42

43   Paul P. Gardner, Department of Biochemistry, University of Otago, P. O. Box 56, Dunedin,
44   9054.
45
46   Email: paul.gardner@otago.ac.nz

47   **Abstract**

48

49   Ribosome-targeting antibiotics comprise over half of antibiotics used in medicine, but our
50   fundamental knowledge of their binding sites is derived primarily from ribosome structures from
51   non-pathogenic species. These include *Thermus thermophilus*, *Deinococcus radiodurans* and
52   the archaean *Haloarcula marismortui,* as well as the commensal and sometime pathogenic
53   organism, *Escherichia coli*. Advancements in electron cryomicroscopy have allowed for the
54   determination of more ribosome structures from pathogenic bacteria, with each study
55   highlighting species-specific differences that had not been observed in the non-pathogenic
56   structures. These observed differences suggest that more novel ribosome structures,
57   particularly from pathogens, are required for a more accurate understanding of the level of
58   diversity of the entire bacterial ribosome, with the potential of leading to innovative
59   advancements in antibiotic research. In this study, high accuracy covariance and hidden Markov
60   models were used to annotate ribosomal RNA and protein sequences respectively from
61   genomic sequence, allowing us to determine the underlying ribosomal sequence diversity using
62   phylogenetic methods. This analysis provided evidence that the current non-pathogenic
63   ribosome structures are not sufficient representatives of some pathogenic bacteria, such as
64   *Campylobacter pylori,* or of whole phyla such as Bacteroidota (Bacteroidetes).

65

66   **Significance Statement**

67

68   The growing number of antibiotic resistance pathogenic bacteria are of critical concern to the
69   health profession. Many of the current classes of antibiotics target the bacterial ribosome, the
70   protein making factory for these species. However, much of our knowledge of the bacterial
71   ribosome is based upon non-pathogenic bacteria that are highly divergent from the major
72   pathogens of concern. We have analysed the genetic variation of the RNA and protein
73   components of all available bacterial ribosomes. This has led us to identify the highest priority
74   groups of bacteria that would provide the most benefit most from further analysis of their
75   ribosome structures, from both a medical and evolutionary perspective.

76

77

Introduction

A global rise of antibiotic resistant bacteria has become an increasingly urgent problem in recent years, with the bacterial ribosome, particularly the ribosomal RNA (rRNA) component, being a common antibiotic target (1–3). Due to the limitations and requirements of X-ray crystallography, antibiotic binding studies initially used extremophiles *Thermus thermophilus* (4), *Deinococcus radiodurans* [a 50S structure] (5, 6) and the archaeon *Haloarcula marismortui* (7), with the pathogenic *Escherichia coli* (8–10) ribosome structure introduced in later years (2, 3, 11). However, improvements in electron cryomicroscopy have allowed more diverse ribosome structures to be analysed more recently, such as from non-pathogenic bacteria *Bacillus subtilis* (12, 13)*, Flavobacterium johnsoniae* (14) and *Mycobacterium smegmatis* (15) (recently renamed "*Mycolicibacterium smegmatis*"), or from pathogens including *Acinetobacter baumannii* (16, 17), *Enterococcus faecalis* (18, 19)*, Listeria monocytogenes* (19, 20)*, Mycobacterium tuberculosis* [a 50S structure] (21), *Pseudomonas aeruginosa* [a 50S structure] (22) and *Staphylococcus aureus* (19, 23, 24). Each pathogen ribosome study highlights species-specific differences in comparison to both non-pathogenic and pathogenic structures, implying that non-pathogenic ribosome structures are not always optimal for inferring antibiotic binding across all bacteria (16, 21–23). As these structural differences could hinder antibiotic research it is important that solved ribosome structures, which form the basis of our understanding of the bacterial ribosome and ribosomal antibiotic binding, are suitably representative of pathogenic species (2, 3).

We further note that the specific target sites of antibiotics correspond to highly conserved, functional ribosome components, and that a survey focusing on these sites will not yield significant diversity. However, we also note that ribosome function has been found to be influenced by less well conserved components of structure (2, 23), and even by elements of structure quite distant from conserved antibiotic binding sites (25). In addition, there is evidence in *S. aureus* and *P. aeruginosa* suggesting that changes to the overall ribosome can lead to antibiotic resistance, in addition to direct mutations in the target sites (22, 25).

Therefore, the aim of this study is to broadly survey ribosome sequence space to determine how representative current ribosome structures are, ranging from the extremophiles (eg: *D. radiodurans, T. thermophilus*) for human pathogenic species (eg: *S. aureus, L. monocytogenes*), when compared to all bacterial species. Using this analysis, we propose new representative bacteria, including some surprising and diverse species, with the potential of yielding valuable new ribosome structural information that could be prioritised for future ribosomal structural studies.

Materials and Methods

Our analysis includes the full-length conserved components of the bacterial ribosome. We have elected not to specifically analyse antibiotic targets (e.g. Spectinomycin targets G1064 and C1992 of 16S rRNA (26)) as these are highly conserved functional components of the ribosome, which therefore carry a limited number of phylogenetically informative characters (27–29). These studies also illustrate the conservation and variation across ribosomal RNA structures.

122 Full-length sequences also allow a global overview of ribosome variation, some of which may
123 change the behaviour of binding sites due to "action at a distance" (2, 23, 25).
124
125 A total of 3,758 bacterial genomes and the *H. marismortui* genome (Accession: AY596297.1)
126 were obtained from the European Nucleotide Archive (30, 31). One representative sequence
127 was retained per species for each rRNA and ribosomal protein sequence, which were filtered to
128 only include genomes with annotations covering 80% of the expected sequence length based
129 upon consensus sequences. If paralogues were present, the sequence with the highest
130 corresponding covariance model and hidden Markov model bit score for the species was used
131 (32).
132
133 Both 16S and 23S rRNA were annotated using barrnap v0.9
134 (https://github.com/tseemann/barrnap), and INFERNAL 1.1.2 was used to create sequence
135 alignments using Rfam v14.3 covariance models (32, 33). Covariance models are the gold-
136 standard for structural RNA alignment (32, 34), and the Rfam models are derived from the CRW
137 database (35), which is the product of 20 years of manual alignment curation, resulting in some
138 of the highest quality ribosomal RNA structural alignments available to date (35, 36).
139
140 Ribosomal protein sequences from 32 universally conserved proteins were identified from six-
141 frame translations of whole genomes, to account for potentially inconsistent or absent genome
142 annotations. The selected sequences were aligned using HMMER v3.1 (hmmer.org), with
143 protein hidden Markov models from Pfam v33.1 (37). Profile HMMs are the gold-standard for
144 protein homology search and alignment (38, 39), Pfam is a widely used and highly curated
145 database of protein domain alignments.
146
147 The advantages of this approach are that our results are up-to-date and consistent across the
148 current genome databases, contain both RNA and protein components of the bacterial ribosome
149 and utilise the most accurate methods for alignment production. Whilst we acknowledge that
150 there are existing ribosomal datasets (e.g. SILVA, GreenGenes, RDB, etc), these do not meet
151 all the above requirements. Although there are alternative ribosomal RNA and protein
152 alignments we could have used (e.g. (28, 40, 41)), their use here would likely produce very
153 similar results.
154
155 Phylogenetic trees for each alignment were generated using the maximum likelihood method
156 from PHYLIP v3.697, with distance matrices of the pairwise distances between species
157 computed in R v4.2.1 using ape v5.6 (42, 43). The distance matrices for each ribosomal gene
158 were summed to create a single unified distance matrix, and this was used for multi-dimensional
159 scaling (MDS) and visualisation (Fig. 1).
160
161 **Results**
162
163 Ribosomal sequence clustering suggests that several bacterial phyla are poorly represented by
164 the currently available ribosome structures. To identify phyla that are poorly represented by
165 solved ribosome structures, two rRNA and 32 universally conserved ribosomal protein

166 phylogeny trees were initially created using a representative sequence for each species. To
167 confirm that these universally conserved ribosomal proteins were truly conserved in our dataset,
168 the trees were filtered on whether they contained more than 95% of isolates from each
169 taxonomy class in the dataset. Taxonomy class was used over phyla to take into account
170 variability within larger, overrepresented phyla such as Pseudomonadota (Proteobacteria) (41).
171 This resulted in ribosomal proteins uS2, uS4, uL23, uL24 and uL30 being removed because
172 they were completely absent in Epsilonproteobacteria and uL15 being removed as it was only
173 present in 16% of Epsilonproteobacteria genomes. As assessing individual ribosomal
174 sequences is unlikely to reflect structural variation, the remaining two rRNA and 26 ribosomal
175 protein phylogenetic trees were combined into one distance matrix, resulting in a final dataset of
176 1,396 species. This combined tree was then reduced using multidimensional scaling (MDS),
177 allowing us to visualise the most dissimilarities between species (41).

178
179 (The resulting MDS plot shows four main clusters; an Alphaproteobacteria cluster and a
180 Betaproteobacteria plus Gammaproteobacteria cluster illustrate the non-monophyletic nature of
181 Pseudomonadota (Proteobacteria) (41), one is dominated by Actinomycetota (Actinobacteria)
182 and the fourth consists of the remaining bacterial phyla and Proteobacterial classes (Fig. 1). The
183 Pseudomonadota (Proteobacteria) cluster follows observations made in previous studies, such
184 as Beta and Gammaproteobacteria being closely related, and Deltaproteobacteria and
185 Oligoflexia being more related to non-Proteobacteria (41, 44). Due to these non-monophyletic
186 properties, Pseudomonadota (Proteobacteria) classes will be treated as individual phyla for the
187 rest of this study (41, 44). Bacteroidota (Bacteroidetes), Acidithiobacilla and
188 Epsilonproteobacteria formed slightly isolated groups away from the larger clusters, suggesting
189 that these could become more defined clusters if their sample sizes were larger (Fig. 1).
190 Alphaproteobacteria is the only cluster without a solved structure, implying the presence of
191 phyla specific variation that has not been captured by current structures (Fig. 1). Therefore, it is
192 unlikely that all the bacterial phyla are well represented by the available ribosome structures,
193 given that the solved structures in the multiple phyla cluster group together, instead of being
194 evenly distributed throughout the clusters (Fig. 1).

195
196 **Evaluation of current structures indicates that *Bacillus subtilis* is the most**
197 **representative.** To evaluate whether current ribosome structures from non-pathogens are
198 sufficiently representative of the available sequenced bacterial ribosomes, phylogenetic
199 distances were calculated from the summed distance matrix between 13 published solved
200 structures (Supplementary Table 2) and 1,385 bacterial ribosomal sequences available in this
201 study (11–16, 18, 21–23). The solved structure with the lowest recorded distance for each
202 species is considered to be the most representative and assumes that species with similar
203 primary sequences will form similar tertiary structures. As the minimum distance to the nearest
204 solved structure increases, it becomes more likely that current structures are not suitably
205 representative. These underrepresented species should be prioritised for future ribosome
206 structural studies. Overall, *B. subtilis* was considered to be the most representative structure for
207 382 species*,* followed by *A. baumannii* for 264 species and 167 for *M. smegmatis* (Fig. 1). *D.*
208 *radiodurans*, *T. thermophilus* and unsurprisingly the archaeon *H. marismortui* were the three
209 least representative structures analysed and were not representative of any pathogenic species

210 (Fig. 1), implying that the structures from these three species are becoming less representative
211 with the introduction of ribosome structures from less divergent bacteria (11). *M. tuberculosis*
212 was also representative for a small number of species, which is likely due to the presence of *M.*
213 *smegmatis*, given that these two bacteria are closely related – effectively splitting the
214 Actinomycetota (Actinobacteria) (Fig. 1).
215
216 This observation does not imply that one structure is necessarily sufficient to represent a
217 complete phyla or class, with *E. coli* and *P. aeruginosa* representing 149 and 110 species
218 respectively, accounting for most Gammaproteobacteria (Fig. 1). Whilst these structures are
219 very similar, they do capture some important differences (22). Each solved structure tends to
220 solely be representative of the phyla it originated from, with *M. smegmatis* as the most
221 representative of Actinomycetota (Actinobacteria), and both *E. coli* and *P. aeruginosa*
222 representing Gammaproteobacteria exclusively. However, *B. subtilis* and *A. baumannii* were not
223 representative of only their respective phyla, as the Gammaproteobacterium *A. baumannii* is
224 representative of both Alpha- and Betaproteobacteria and the Bacillota (Firmicutes) *B. subtilis*
225 was the most representative for the remaining phyla without a ribosome structure (Table 1).
226 Therefore, we hypothesise that having at least one representative structure per phylum, or
227 preferably one per class, would allow the majority of bacteria to be sufficiently represented.
228

229
230
231
232   Introducing new ribosome structures shows that an Epsilonproteobacteria representative, such
233   as a *Campylobacter jejuni*, should be a high priority for further structural work. To simulate the
234   effect of having at least one representative ribosome structure per phyla, we selected one
235   proposed structure per phyla which had the smallest average phylogenetic distance to all other
236   members in the respective phyla. The species that were initially considered to be potential
237   representatives were selected based on how dissimilar they were to a species with a solved
238   ribosome structure. These were selected by determining the phylogenetic distance between
239   each isolate to a solved structure on the combined rRNA and ribosomal protein phylogenetic
240   tree, and selecting the structure with the shortest distance as the representative for that
241   individual. Based on the distribution of these distances (Supplementary Fig. 1), species with a
242   distance to a solved structure greater than the lower quartile (11.41) were considered for the
243   new ribosome structure analysis, as these species are most likely to be poorly represented by
244   current structures. The available pathogens with a minimum distance above this threshold were
245   prioritised due to their relevance in health research. The introduction of new proposed structures
246   resulted in a decrease in the average minimum distance to a solved structure across each
247   phylum, implying that having at least one structure per phylum is a suitable sampling strategy
248   for improving representation (Table 1). Of the ten highest priority structures, only
249   Epsilonproteobacteria, Chlamydiota (Chlamydiae) and Chlorobiota (Chlorobi) had an average
250   minimum distance per phyla below the lower quartile threshold, after the proposed structures
251   were introduced (Table 1). This suggests that the other phyla listed are more diverse and may
252   require additional structures to capture the remaining variation, or that a representative non-
253   pathogenic ribosome would be a more beneficial representative. *Campylobacter jejuni* was
254   identified as the highest priority structure (Table 1)*,* as it was the most representative
255   Epsilonproteobacteria pathogen, had the largest average minimum distance observed across all
256   phyla and is a WHO priority pathogen (1). However, the selected ribosome for solving does not
257   specifically need to be from a pathogenic strain, with either an attenuated lab strain (21, 23), or
258   another species from the same genus (15, 18) being appropriate alternatives (Fig. 1).
259
260
261
262   **Discussion**
263
264   Our results reveal significant phylogenetic gaps in the existing cohort of solved ribosome
265   structures. Filling these gaps will help identify clade-specific features of isolated bacterial clades
266   that lack representative ribosome structures. This is important for the analysis of ribosome
267   evolution, pathogen research, and potentially for antibiotic development.
268
269   There are two limitations with this approach to ribosome structure prioritisation: there is a bias
270   towards species with available genomes, which are enriched in pathogenic species rather than
271   being the most evolutionarily representative (45). In addition, our approach prioritises the most
272   novel ribosomes within each major bacterial clade, with the risk of over-emphasising the most

273 diverse ribosomes. This is somewhat mitigated by selecting the most closely related pathogenic
274 species as representative targets.
275
276 While having at least one representative ribosome structure per phyla is expected to capture the
277 bulk of structural variation (Table 1), there is no guarantee that primary sequence differences
278 result in significant changes at the tertiary level. A prominent example of this is *H. marismortui*
279 which, although it is an archaean ribosome structure, has been used as an alternative to
280 bacterial counterparts (11), despite having no close relatives at the primary level (Fig. 1).
281 Another consideration is that the structural variation observed may be species-specific, rather
282 than shared across a phylum, with species-specific differences having been observed between
283 *P. aeruginosa* and *A. baumannii*, yet no phyla-specific differences when compared to *E. coli* and
284 other available structures (16, 22). However, these three structures were observed to be
285 relatively divergent from each other based upon the MDS analysis (Fig. 1), which reinforces the
286 *in vivo* observation that *P. aeruginosa* and *A. baumannii* ribosomes are more similar to each
287 other than to *E. coli,* suggesting that phylogenetic analyses have the potential to reflect
288 structural variation (16).
289
290 Our list of high-priority bacterial ribosomes reveals that a number of bacterial clades are poorly
291 represented by the existing solved structures (Table 1, Supplementary Table 1). Working down
292 this list gives an opportunity to identify the most variant ribosomes based upon sequence
293 analysis, which in many instances will reflect diverse ribosome structures. These will allow the
294 development of more specific classes of antibiotics, as well as provide further information
295 regarding the specifics of ribosome structure evolution.
296
297 **Data Availability**
298
299 All data described was last accessed in September 2022. Custom scripts, details of parameters,
300 and dependencies used, and the accessions for all downloaded data and the resulting curated
301 alignments and trees are available on GitHub: https://github.com/helena-bethany/ribosomal-
302 variation
303
304 **Funding**
305

306 This work was supported by the Department of Biochemistry (University of Otago) as an
307 Honours year research project and by an Australian Government Research Training Program
308 (RTP) Scholarship.

309
310
311
312

313 **Acknowledgements**

314

315 The authors would like to thank Steven Gregory (University of Rhode Island) and Gerwald Jogl

316 (Brown University) for their discussions and advice regarding this project.

317

318 **References**

319 1. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C,

320   Kahlmeter G, Kluytmans J, Carmeli Y, Ouellette M, Outterson K, Patel J, Cavaleri M, Cox

321   EM, Houchens CR, Grayson ML, Hansen P, Singh N, Theuretzbacher U, Magrini N, WHO

322   Pathogens Priority List Working Group. 2018. Discovery, research, and development of

323   new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet

324   Infect Dis 18:318–327.

325 2. Lin J, Zhou D, Steitz TA, Polikanov YS, Gagnon MG. 2018. Ribosome-Targeting

326   Antibiotics: Modes of Action, Mechanisms of Resistance, and Implications for Drug Design.

327   Annual Review of Biochemistry https://doi.org/10.1146/annurev-biochem-062917-011942.

328 3. Wilson DN. 2014. Ribosome-targeting antibiotics and mechanisms of bacterial resistance.

329   Nat Rev Microbiol 12:35–48.

330 4. Polikanov YS, Melnikov SV, Söll D, Steitz TA. 2015. Structural insights into the role of rRNA

331   modifications in protein synthesis and ribosome assembly. Nat Struct Mol Biol 22:342–344.

332 5. Harms JM, Wilson DN, Schluenzen F, Connell SR, Stachelhaus T, Zaborowska Z, Spahn

333   CMT, Fucini P. 2008. Translational regulation via L11: molecular switches on the ribosome

334   turned on and off by thiostrepton and micrococcin. Mol Cell 30:26–38.

335 6. Kaminishi T, Schedlbauer A, Fabbretti A, Brandi L, Ochoa-Lizarralde B, He C-G, Milón P,

336   Connell SR, Gualerzi CO, Fucini P. 2015. Crystallographic characterization of the ribosomal

337   binding site and molecular mechanism of action of Hygromycin A. Nucleic Acids Res

338   43:10015–10025.

339   7.   Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of

340   the large ribosomal subunit at 2.4 A resolution. Science 289:905–920.

341   8.   Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H. 2015.

342   Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-

343   EM. Nature 520:567–570.

344   9.   Noeske J, Wasserman MR, Terry DS, Altman RB, Blanchard SC, Cate JHD. 2015. High-

345   resolution structure of the Escherichia coli ribosome. Nat Struct Mol Biol 22:336–341.

346   10.   Watson ZL, Ward FR, Méheust R, Ad O, Schepartz A, Banfield JF, Cate JH. 2020.

347   Structure of the bacterial ribosome at 2 Å resolution. Elife 9.

348   11.   Fox GE. 2010. Origin and evolution of the ribosome. Cold Spring Harb Perspect Biol

349   2:a003483.

350   12.   Sohmen D, Chiba S, Shimokawa-Chiba N, Innis CA, Berninghausen O, Beckmann R, Ito K,

351   Wilson DN. 2015. Structure of the Bacillus subtilis 70S ribosome reveals the basis for

352   species-specific stalling. Nat Commun 6:6941.

353   13.   Crowe-McAuliffe C, Takada H, Murina V, Polte C, Kasvandik S, Tenson T, Ignatova Z,

354   Atkinson GC, Wilson DN, Hauryliuk V. 2021. Structural Basis for Bacterial Ribosome-

355   Associated Quality Control by RqcH and RqcP. Mol Cell 81:115–126.e7.

356   14.   Jha V, Roy B, Jahagirdar D, McNutt ZA, Shatoff EA, Boleratz BL, Watkins DE, Bundschuh

357   R, Basu K, Ortega J, Fredrick K. 2021. Structural basis of sequestration of the anti-Shine-

358   Dalgarno sequence in the Bacteroidetes ribosome. Nucleic Acids Res 49:547–567.

359   15.   Hentschel J, Burnside C, Mignot I, Leibundgut M, Boehringer D, Ban N. 2017. The

360   Complete Structure of the Mycobacterium smegmatis 70S Ribosome. Cell Rep 20:149–

361   160.

362   16. Morgan CE, Huang W, Rudin SD, Taylor DJ, Kirby JE, Bonomo RA, Yu EW. 2020. Cryo-

363   electron Microscopy Structure of the Acinetobacter baumannii 70S Ribosome and

364   Implications for New Antibiotic Development. MBio 11.

365   17. Zhang Z, Morgan CE, Bonomo RA, Yu EW. 2021. Cryo-EM Determination of Eravacycline-

366   Bound Structures of the Ribosome and the Multidrug Efflux Pump AdeJ of Acinetobacter

367   baumannii. MBio 12:e0103121.

368   18. Murphy EL, Singh KV, Avila B, Kleffmann T, Gregory ST, Murray BE, Krause KL, Khayat R,

369   Jogl G. 2020. Cryo-electron microscopy structure of the 70S ribosome from Enterococcus

370   faecalis. Sci Rep 10:16301.

371   19. Crowe-McAuliffe C, Murina V, Turnbull KJ, Kasari M, Mohamad M, Polte C, Takada H,

372   Vaitkevicius K, Johansson J, Ignatova Z, Atkinson GC, O'Neill AJ, Hauryliuk V, Wilson DN.

373   2021. Structural basis of ABCF-mediated resistance to pleuromutilin, lincosamide, and

374   streptogramin A antibiotics in Gram-positive pathogens. Nat Commun 12:3577.

375   20. Koller TO, Turnbull KJ, Vaitkevicius K, Crowe-McAuliffe C, Roghanian M, Bulvas O,

376   Nakamoto JA, Kurata T, Julius C, Atkinson GC, Johansson J, Hauryliuk V, Wilson DN.

377   2022. Structural basis for HflXr-mediated antibiotic resistance in Listeria monocytogenes.

378   Nucleic Acids Res 50:11285–11300.

379   21. Yang K, Chang J-Y, Cui Z, Li X, Meng R, Duan L, Thongchol J, Jakana J, Huwe CM,

380   Sacchettini JC, Zhang J. 2017. Structural insights into species-specific features of the

381   ribosome from the human pathogen Mycobacterium tuberculosis. Nucleic Acids Res

382   45:10884–10894.

383   22. Halfon Y, Jimenez-Fernandez A, La Rosa R, Espinosa Portero R, Krogh Johansen H,

384   Matzov D, Eyal Z, Bashan A, Zimmerman E, Belousoff M, Molin S, Yonath A. 2019.

385   Structure of Pseudomonas aeruginosa ribosomes from an aminoglycoside-resistant clinical

386   isolate. Proc Natl Acad Sci U S A 116:22275–22281.

387 23. Eyal Z, Matzov D, Krupkin M, Wekselman I, Paukner S, Zimmerman E, Rozenberg H,

388   Bashan A, Yonath A. 2015. Structural insights into species-specific features of the

389   ribosome from the pathogen Staphylococcus aureus. Proc Natl Acad Sci U S A

390   112:E5805–14.

391 24. Halfon Y, Matzov D, Eyal Z, Bashan A, Zimmerman E, Kjeldgaard J, Ingmer H, Yonath A.

392   2019. Exit tunnel modulation as resistance mechanism of S. aureus erythromycin resistant

393   mutant. Sci Rep 9:11460.

394 25. Belousoff MJ, Eyal Z, Radjainia M, Ahmed T, Bamert RS, Matzov D, Bashan A,

395   Zimmerman E, Mishra S, Cameron D, Elmlund H, Peleg AY, Bhushan S, Lithgow T, Yonath

396   A. 2017. Structural Basis for Linezolid Binding Site Rearrangement in the Staphylococcus

397   aureus Ribosome. MBio 8.

398 26. Brink MF, Brink G, Verbeet M, de Boer HA. 1994. Spectinomycin interacts specifically with

399   the residues G 1064 and C 1192 in 16S rRNA, thereby potentially freezing this molecule

400   into an inactive conformation. Nucleic Acids Res 22:325–331.

401 27. Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG,

402   Harvey SC, Fox GE, Wartell RM, Hud NV, Williams LD. 2014. Evolution of the ribosome at

403   atomic resolution. Proc Natl Acad Sci U S A 111:10251–10256.

404 28. Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD. 2018. Translation: The

405   Universal Structural Core of Life. Mol Biol Evol 35:2065–2076.

406 29. Tirumalai MR, Rivas M, Tran Q, Fox GE. 2021. The Peptidyl Transferase Center: a Window

407   to the Past. Microbiol Mol Biol Rev 85:e0010421.

408   30. Amid C, Alako BTF, Balavenkataraman Kadhirvelu V, Burdett T, Burgin J, Fan J, Harrison

409   PW, Holt S, Hussein A, Ivanov E, Jayathilaka S, Kay S, Keane T, Leinonen R, Liu X,

410   Martinez-Villacorta J, Milano A, Pakseresht A, Rahman N, Rajan J, Reddy K, Richards E,

411   Smirnov D, Sokolov A, Vijayaraja S, Cochrane G. 2020. The European Nucleotide Archive

412   in 2019. Nucleic Acids Res 48:D70–D76.

413   31. Baliga NS, Bonneau R, Facciotti MT, Pan M, Glusman G, Deutsch EW, Shannon P, Chiu Y,

414   Weng RS, Gan RR, Hung P, Date SV, Marcotte E, Hood L, Ng WV. 2004. Genome

415   sequence of Haloarcula marismortui: a halophilic archaeon from the Dead Sea. Genome

416   Res 14:2221–2234.

417   32. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches.

418   Bioinformatics 29:2933–2935.

419   33. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A,

420   Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding

421   RNA families. Nucleic Acids Res 46:D335–D342.

422   34. Freyhult EK, Bollback JP, Gardner PP. 2007. Exploring genomic dark matter: a critical

423   assessment of the performance of homology search methods on noncoding RNA. Genome

424   Res 17:117–125.

425   35. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N,

426   Madabusi LV, Müller KM, Pande N, Shang Z, Yu N, Gutell RR. 2002. The comparative RNA

427   web (CRW) site: an online database of comparative sequence and structure information for

428   ribosomal, intron, and other RNAs. BMC Bioinformatics 3:2.

429   36. Gutell RR, Lee JC, Cannone JJ. 2002. The accuracy of ribosomal RNA comparative

430   structure models. Curr Opin Struct Biol 12:301–310.

431   37.   El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson

432   LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE,

433   Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–

434   D432.

435   38.   Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

436   39.   Söding J, Remmert M. 2011. Protein sequence comparison and fold recognition: progress

437   and good-practice benchmarking. Curr Opin Struct Biol 21:404–411.

438   40.   Doris SM, Smith DR, Beamesderfer JN, Raphael BJ, Nathanson JA, Gerbi SA. 2015.

439   Universal and domain-specific sequences in 23S-28S ribosomal RNA identified by

440   computational phylogenetics. RNA 21:1719–1730.a new view

441   41.   Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN,

442   Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson

443   R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.

444   42.   Felsenstein J. 2009. PHYLIP (Phylogeny Inference Package) version 3.7a. Distributed by

445   the author. http://www.evolution.gs.washington.edu/phylip.html. Department of Genome

446   Sciences, University of Washington.

447   43.   Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and

448   evolutionary analyses in R. Bioinformatics 35:526–528.

449   44.   Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, Loy A, Naganuma T,

450   Nakai R, Whitman WB, Hahn MW, Kuever J, Hugenholtz P. 2020. Proposal to reclassify the

451   proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum

452   Thermodesulfobacteria into four phyla reflecting major functional capabilities. Int J Syst

453   Evol Microbiol 70:5972–6016.

454   45. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L,

455   Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D'haeseleer P,

456   Zemla A, Singer M, Lapidus A, Nolan M, Copeland A, Han C, Chen F, Cheng J-F, Lucas S,

457   Kerfeld C, Lang E, Gronow S, Chain P, Bruce D, Rubin EM, Kyrpides NC, Klenk H-P, Eisen

458   JA. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature

459   462:1056–1060.

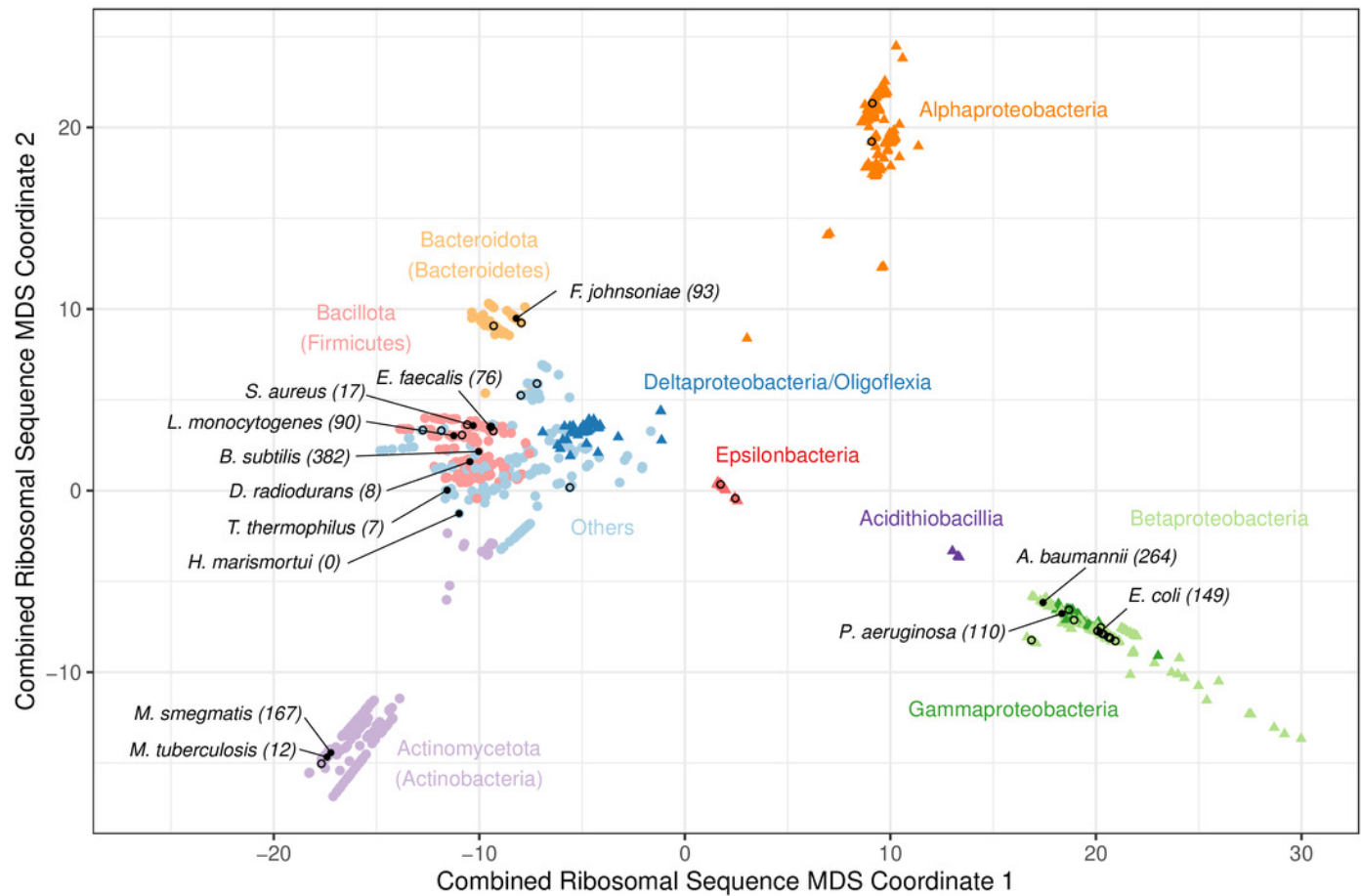
460

461 **Captions:**

462 **Table 1.** The ten highest priority ribosome structures to solve, with priority given to pathogenic
463 species - Choices were based on the average minimum distances to a solved structure prior to
464 the introduction of each proposed structure (See Supplementary Table 1 for an extended list).
465
466 **Fig. 1.** A multiple-dimensional scaling (MDS) plot for combined phylogenetic distances from
467 16S, 23S rRNA and 26 universally conserved ribosomal proteins (N=1,396). All
468 Pseudomonadota (Proteobacteria) are plotted as triangles and coloured by class, whereas other
469 included phyla are plotted as circles and coloured individually. Species with a solved ribosome
470 structure (N=13), or known pathogenic bacteria without a structure (N=38) have been
471 highlighted with a filled or hollow black circle respectively. The number of bacteria that consider
472 each species with solved structure to be representative, based on the shortest phylogenetic
473 distance to a solved structure, has been labelled along with the species' names. The full list of
474 species, MDS coordinates, the solved structures that were considered to be most representative
475 and the corresponding minimum distance are available on Github.

Figure 1

A multiple-dimensional scaling (MDS) plot for combined phylogenetic distances from 16S, 23S rRNA and 26 universally conserved ribosomal proteins.

A multiple-dimensional scaling (MDS) plot for combined phylogenetic distances from 16S, 23S rRNA and 26 universally conserved ribosomal proteins (N=1,396). All Proteobacteria are coloured by class, and other included phyla are coloured individually to highlight taxonomic relationships with clusters. Species with a solved ribosome structure (N=13), or known pathogenic bacteria without a structure (N=38) have been highlighted with a black dot. The number of bacteria that consider each species with solved structure to be representative, based on the minimum phylogenetic distance, has been labelled along with the species' names. The full list of species, MDS coordinates, the solved structures that were considered to be most representative and the corresponding minimum distance are available on Github.

The ten highest priority ribosome structures to solve, with priority given to pathogenic species.

Choices were based on the average minimum distances to a solved structure prior to the introduction of each proposed structure (See Supplementary Table 1 for an extended list).

1   **Table 1.** The ten highest priority ribosome structures to solve, with priority given to pathogenic
2   species - Choices were based on the average minimum distances to a solved structure prior to
3   the introduction of each proposed structure (See Supplementary Table 1 for an extended list).
4

Proposed target species	Phylum	Min distance to closest structure	Avg min distance for phylum*
Campylobacter jejuni	Epsilonproteobacteria	41.07 (*B. subtilis*)	42.24 (10.11)
Chlamydophila trachomatis	Chlamydiota (Chlamydiae)	40.25 (*B. subtilis*)	39.58 (8.27)
Opitutaceae bacterium	Verrucomicrobiota (Verrucomicrobia)	38.24 (*B. subtilis*)	37.89 (16.00)
Borrelia recurrentis	Spirochaetota (Spirochaetes)	38.00 (*B. subtilis*)	39.41 (20.50)
Singulisphaera acidiphila	Planctomycetota (Planctomycetes)	37.81 (*B. subtilis*)	39.54 (21.48)
Brucella melitensis	Alphaproteobacteria	37.73 (*A. baumannii*)	38.78 (19.26)
Chlorobium limicola	Chlorobiota (Chlorobi)	37.52 (*F. johnsoniae*)	37.34 (5.64)
Ureaplasma urealyticum	Mycoplasmatota (Tenericutes)	35.05 (*L. monocytogenes*)	32.16 (26.65)
Leptospirillum ferriphilum	Nitrospirota (Nitrospirae)	33.96 (*B. subtilis*)	32.63 (16.34)
Persephonella marina	Aquificota (Aquificae)	33.29 (*B. subtilis*)	34.06 (14.81)

6   * The first value is the average minimum distance prior to the proposed structures being selected, and the value in brackets is the
7   recalculated average distance after the proposed structures have been incorporated.