Article

Discoveries

Simplification of ribosomes in bacteria with tiny genomes

Daria D. Nikolaeva[1,3], Mikhail S. Gelfand[2,3,4], Sofya K. Garushyants[3,*]

[1] Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Leninskiye Gory 1-73, Moscow, 119234, Russia;

[2] Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, Moscow, 121205, Russia;

[3] Institute for Information Transmission Problems (Kharkevich Institute), Bolshoy Karetny 19-1, Moscow, 127051, Russia;

[4] Faculty of Computer Science, National Research University Higher School of Economics, Kochnovsky pr. 3, Moscow, 125319, Russia.

* Correspondence: Sofya K. Garushyants garushyants@iitp.ru

Keywords: ribosome; bacteria; evolution; genome reduction; rRNA; ribosomal protein

Abstract

The ribosome is an essential cellular machine performing protein biosynthesis. Its structure and composition are highly conserved in all species. However, some bacteria have been reported to have an incomplete set of ribosomal proteins. We have analyzed ribosomal protein composition in 214 small bacterial genomes (< 1 Mb) and found that although the ribosome composition is fairly stable, some ribosomal proteins may be absent, especially in bacteria with dramatically reduced genomes. The protein composition of the large subunit is less conserved than that of the small subunit. We have identified the set of frequently lost ribosomal proteins and demonstrated that they tend to be situated on the ribosome surface and have fewer contacts to other ribosome components. Moreover, some proteins are lost in an evolutionary correlated manner. The reduction of rRNA is also common in bacteria with tiny genomes with deletions mostly occurring in free loops. Finally, the loss of the anti-Shine-Dalgarno sequence is associated with the genome reduction, the number of lost ribosomal proteins, and with the loss of bL9 and TF.

Introduction

The ribosome is a universal biosynthesis machine present in all eukaryotes and prokaryotes. A bacterial ribosome is comprised of the small (30S) and large (50S) subunits which together form the 70S particle (Kurland 1972; Ramakrishnan 2002). The bacterial ribosome consists of multiple proteins (RP, r-proteins) and three rRNA molecules – 16S in the small subunit, 23S and 5S in the large subunit (Kurland 1972). The main catalytic functions of the ribosome, such as the peptide bond formation, mRNA decoding, and translocation of mRNA and tRNA after the peptide bond formation, are performed by ribosomal RNA (rRNA) (Green & Noller 1997; Nissen et al. 2000; Schmeing et al. 2003). Moreover, rRNA molecules also determine the ribosomal spatial organization providing sites for binding of the ribosomal proteins (Khaitovich et al. 1999). The ribosome of *Escherichia coli* contains 21 proteins in the 30S subunit (bS1–bS21) and 33 proteins in the 50S subunit (uL1–bL36) (Ban et al. 2014; Schuwirth et al. 2005; Kaczanowska & Rydén-Aulin 2007). The role of ribosomal proteins is to stabilize the ribosome and to regulate the ribosomal activity (Aseev and Boni 2011). While the key role in the protein biosynthesis is played by rRNA, the r-protein composition also tends to be conserved in most bacteria (Roberts et al. 2008). Moreover, 33 r-proteins are conserved among different domains of life (Roberts et al. 2008; Smith et al. 2008; Lecompte et al. 2002).

The protein composition of bacterial ribosomes has been studied intensively. The analysis of single-gene knockout mutants of *E. coli* has shown that these bacteria are able to grow without proteins bS6, uS15, bS20, bS21, uL1, bL9, uL11, and bL25, but in most cases the growth rate of these knockout mutants is reduced (Baba et al. 2006). In another study of an *E. coli* knockout collection, nine r-proteins (uL15, bL21, uL24, bL27, uL29, uL30, bL34, and uS17) have been found to be nonessential for survival in experimental conditions (Shoji et al. 2011). A similar study in *Bacillus subtilis* has identified 20 r-proteins non-essential to growth in experimental conditions (Akanuma et al. 2012). Based on this result, the smaller r-proteins were proposed to have been incorporated into the ribosome relatively recently in evolution, and hence to be less essential. A phylogenetic analysis of 995 completely sequenced bacterial genomes has shown that 44 r-proteins are strictly ubiquitous, proteins bS16, bL9, bL19, bL31, bL34, and bL36 are rarely missing, while bS21, S22, bThx, bL25, and uL30 are absent in a large fraction of bacteria (Yutin et al. 2012). A comparative analysis of the translation apparatus of *Mollicutes* has shown that five r-proteins (uS14, S22, bL7, bL25, and bL31) are missing in all studied genomes, uL30 is missing in almost all of them, and bS1 has been lost in seven independent events in different clades (Grosjean et al. 2014). Finally, an analysis of endosymbiotic bacteria with small genomes from a variety of phyla has demonstrated that they lack the largest fraction of r-proteins, as only 17 of 21 small-subunit and 16 of 32 large-subunit r-proteins are universally present in these bacteria (McCutcheon and Moran 2012). For example, *Candidatus* Tremblaya princeps, an endosymbiont of mealybugs (*Pseudococcidae* family), has only 139 genes suggesting little possibility to carry the complete translation apparatus (McCutcheon and Moran 2012). While the published lists of essential r-proteins vary to some extent, bL25, uL30, bL31, bS21, S22, and bThx have been consistently reported to be the least essential.

Less is known about whether there exist universal features common to these non-essential proteins, and how the r-protein loss is linked to the genome reduction. Here, we analyzed the r-protein composition in bacteria with small genomes from a variety of phyla. We have identified a set of frequently lost proteins and found two patterns of evolutionary correlative r-protein loss. A majority of frequently lost proteins have been shown to be located on the ribosome surface and to form less contacts with other ribosome components, compared to universally conserved r-proteins. We also show that the loss of the anti-Shine-Dalgarno sequence is associated with the genome reduction, the number of lost ribosomal proteins, and specifically with the loss of bL9 and TF.

Results

Some ribosomal proteins are missing in bacteria with small genomes

We re-annotated all r-proteins in 214 bacterial strains with small genomes (less than 1 Mbp) from 38 genera of the following phyla: *Proteobacteria*, *Bacteroidetes*, *Spirochaetes*, *Tenericutes*, and *Actinobacteria* (Supplementary Table S1, Materials and Methods). For that, we used 65 Pfam domains: 63 domains for the canonical set of ribosomal proteins, and two domains of the trigger-factor protein (TF), a ribosome-associated chaperone (Supplementary Table S2).

Our results (Supplementary Fig. S1) show that two proteins are absent in almost all considered strains, S22 (all strains) and bThx (present in *Candidatus* Walczuchella monophlebidarum only).

All except 11 r-proteins are lost in at least one strain from our dataset. At that, proteins bL9, uL24, bL25, uL29, uL30, bL32, bL34, bL36, bS21, and TF are lost frequently, as they are absent in at least 19 strains from a variety of phyla. This set later on is called frequently lost proteins. R-proteins of the small subunit are more likely to be retained than the large-subunit r-proteins (Supplementary Fig. S1). The only frequently lost small-subunit protein bS21 is absent in 54 strains from seven genera, while the most frequently lost protein from the large subunit, uL30, is absent in 138 genomes from 18 genera. All frequently lost proteins had been lost independently multiple times in bacteria (Fig. 1, Supplementary Fig. S2). The largest number of r-protein losses was observed in *Candidatus* Tremblaya princeps, *Candidatus* Hodgkinia cicadocola, and *Candidatus* Carsonella rudii, bacteria with the shortest genomes in our dataset.

R-protein loss depends on the level of genome reduction

The comparison of the r-protein composition and the genome size revealed a correlation between the genome size and the number of retained ribosomal proteins ($r^2 = 0.6$, $p < 2.2\times10^{-16}$), but the slope is dramatically different for the tiny genomes (Supplementary Fig. S3a). Indeed, for genomes shorter than 350 Kb there is an even stronger correlation between the genome size and the number of r-proteins ($r^2 = 0.7$, $p = 5.7\times10^{-6}$), while the remaining genomes in the dataset show no correlation at all ($p = 0.76$). This pattern holds both for the complete ribosome and the individual subunits (Supplementary Fig. S3bc), the slope being steeper for the large subunit.

Patterns of ribosomal protein loss

112 Ten most frequently lost proteins have been lost independently in different phyla,
113 showing parallel events (Fig. 1, Supplementary Fig. S2). A general tendency is that
114 frequently lost proteins are usually situated at the ribosome surface (Fig. 2). To analyze
115 possible dependencies between frequent losses, we estimated the correlation between the
116 vectors of the protein presence/absence in all strains by the Pagel correlation method (Pagel
117 1994) which allows one to control for the phylogenetic structure of the dataset (Fig. 2a).
118 After the Bonferroni correction two clusters were observed, uL29+TF and
119 bL32+bL34+bL36.
120 One of the observed loss clusters is consistent with the r-proteins location in the
121 ribosome. The correlated loss of proteins uL29 and TF may be explained by the TF position
122 in the ribosomal structure, as the latter is located at the ribosome surface where it is
123 surrounded by r-proteins uL23, uL24, and uL29 (Fig. 2d) directly interacting with uL23 and
124 uL29 (Ferbitz et al. 2004). The second pattern may not be explained by the structural
125 proximity as bL36 is situated far from bL32 and bL34. These r-proteins share a common
126 feature of being deeply immersed within a structure, with bL32 having a surface region,
127 while bL34 and bL36 being immersed completely (Fig. 2bce).
128 In the "onion model" of the ribosome structure and evolution (Hsiao et al. 2009;
129 Petrov et al. 2015), the peptidyl transferase center (PTC) is the core and the most ancient part
130 of the ribosome. According to this model, the farther is an r-protein from the PTC, the later it
131 has been added during ribosome evolution. Indeed, most of the frequently lost proteins are
132 situated in the outer layers of the "onion", except for the r-proteins bL34, bL36, bL25, uL30.
133 However, there is no significant difference between conserved and frequently lost r-proteins
134 in the distribution of distances from the PTC.

135 **Frequently lost r-proteins have less contacts than conserved proteins**

136 To study what features differ between frequently lost and conserved r-proteins, we
137 fitted a logistic regression model to our data, regressing the protein status (either conserved or
138 frequently lost) on the evolutionary rate, protein length, and the number of contacts with all
139 adjacent proteins and rRNA. Only the number of contacts was a significant determinant of
140 the protein status ($p = 0.022$).
141 We further analyzed the differences in the evolutionary rate (Fig. 3a) and the number
142 of contacts (Fig. 3b) between conserved and frequently lost proteins (see Supplementary
143 Table S3). Indeed, frequently lost proteins form significantly less contacts than conserved
144 proteins (means 231 and 478, respectively, $p = 1.4\times10^{-4}$), as the former are mainly situated on

145 the ribosome surface. These findings are supported by measurements of the contacting
146 surface between each of the studied r-proteins and the rest of the ribosome (the Wilcoxon test
147 $p = 2.0\times10^{-24}$ for PDB structure 5H5U that does not contain TF, $p = 8.2\times10^{-5}$ when the value
148 for TF from PDB structure 2D3O is added to data for PDB structure 5H5U). While several
149 PDB structures of the ribosome are available, none of them is complete (i.e. includes the
150 entire ribosome and TF) and for none all parts of the ribosome are resolved. Because of that
151 we considered several PDB structures, however, the results do not depend on the selected
152 structure (Supplementary Fig. S4). However, if the contacting surface values are normalized
153 on the protein's surface, the difference between frequently lost and conserved r-proteins
154 becomes insignificant, and hence the absolute interacting surface is important, and not a
155 relative one.

156 Surprisingly, there was no significant difference in the evolutionary rate between
157 conserved and frequently lost proteins. Moreover, the evolutionary rate was slightly
158 correlated with the protein length ($r^2 = 0.39$, $p = 0.055$), with short proteins evolving slower
159 (Fig. 3c).

160 **Deletions in rRNAs and the loss of r-proteins**

161 To study possible correlations between the loss of r-proteins and their rRNA binding
162 sites, we analyzed 16S and 23S rRNA multiple alignments and identified nine deletion
163 blocks (four in 23S and five in 16S) that had occurred in more than two genera and were
164 longer than five nucleotides in at least one species. All such deletions affect rRNA free loops
165 not involved in the r-protein binding (Supplementary Fig. S5).

166 **Anti-SD loss is frequent among strains from different taxa**

167 The anti-Shine-Dalgarno sequence in 16S rRNA is frequently lost in diverse phyla.
168 Its presence/absence pattern is not strongly linked to the taxonomy, as many phyla have
169 strains both with and without the anti-SD sequence in 16S rRNA (Supplementary Fig. S6).
170 The loss is significantly associated with the genome reduction (one-sided Wilcoxon test $p =$
171 0.034) and with the number of lost r-proteins (one-sided Wilcoxon test $p = 1.04\times10^{-5}$). We
172 have observed a strong correlation of the anti-SD loss with loss of TF and some correlation
173 between the losses of anti-SD and bL9, but with lower confidence (Pagel's test with the
174 Bonferroni correction $p = 7.55\times10^{-5}$ and $p < 0.05$, respectively; see Supplementary Table
175 S4).

Discussion

The bacterial genome size is evolutionarily labile, and radical genome reduction has occurred many times independently throughout the bacterial domain. Such multiple independent genome reduction events allows for a systematic study of the patterns of gene loss associated with them.

While individual r-proteins can be lost in bacteria with genomes of any size, for the tiniest of bacterial genomes, further reduction in the genome size is associated with the loss of generally conserved ribosomal proteins. The fact that ribosomal proteins are among the last to leave a shrinking genome illustrates that they tend to be less dispensable than most proteins. Two r-proteins absent in the largest number of small-genome bacteria are S22 and bThx. Previous studies indicate that these proteins are largely non-essential. Indeed, the gene encoding S22 is mostly expressed at the stationary growth phase, and its deletion does not affect the viability of *E. coli* mutants (Izutsu et al. 2001). bThx is found only in thermophilic bacteria and has been shown to stabilize the organization of RNA elements at the top of the 30S subunit head in *Thermus thermophilus* (Choli et al. 1993; Wimberly et al. 2000), which implies that this protein is only essential for survival at high temperatures. However, the low prevalence of these two proteins may indicate not the ease of their loss, but rather their late arrival. S22 is present only in some enterobacteria (Izutsu et al. 2001), and bThx homologs are only present in thermophilic bacteria from genus *Thermus* (Choli et al. 1993; Wimberly et al. 2000). Therefore, both S22 and bThx seem to have been acquired only by some bacterial groups long after the last common ancestor of bacteria.

Our list of frequently lost proteins is largely consistent with previous observations (Baba et al. 2006; Akanuma et al. 2012; Yutin et al. 2012; Grosjean et al. 2014; Shoji et al. 2011). For example, among six non-ubiquitous r-proteins reported by Yutin et al. 2012, three (bL25, uL30, bS21) are also among the r-proteins identified as frequently lost in our dataset, in addition to two proteins (S22 and bThx) that only occur in some bacterial lineages (see above). However, our focus on bacteria with tiny genomes allowed us to study less dispensable r-proteins. Proteins bL34 and bL36 identified previously as present in almost all genomes (Yutin et al. 2012) are frequently lost in our dataset. While these proteins are generally considered essential, they can be dispensable under certain circumstances. The absence of bL34 affects cell growth, but the cell function can be restored by increasing the magnesium ion flow (Shoji et al. 2011; Akanuma et al. 2014). The absence of bL36 is only essential for cell growth at high temperatures (Ikegami et al. 2005). Moreover, bL36 is

missing in *Bacteroidetes* (Yutin et al. 2012). R-proteins uL24 and uL29 identified here as frequently lost had been lost in at least two independent events, and only in bacteria with drastically reduced genomes. The knockouts of both these proteins are viable (Shoji et al. 2011). Interestingly, r-proteins bL31 and uS14 have been identified as lost in all *Mollicutes* (Grosjean et al. 2014), but we have observed that only few strains of *Mollicutes* lack these proteins. This difference is likely caused by a more sensitive annotation procedure used here.

Extreme genome reduction is common not only in endosymbiotic bacteria but also in cell organelles, mitochondria and plastids. The r-protein composition of these organelles has been extensively studied (Yamaguchi and Subramanian 2000; Maier et al. 2013; Petrov et al. 2018). The plastid r-proteins core encoded by all plastids is comprised of only 15 r-proteins of bacterial origin (Maier et al. 2013). This set contains only one of the frequently lost r-proteins reported here, bL36, while other frequently lost r-proteins are frequently absent in plastids (Maier et al. 2013). The protein composition of mitochondria is highly variable across species and is characterized not only by gene loss but also by acquisition of proteins of various origins (Petrov et al. 2018). Among eight large-subunit proteins in our list of frequent losses, six r-proteins (bL9, bL25, uL30, bL32, bL34, bL36) are absent in mitochondria in at least one studied species, but some proteins of bacterial origin are not simply lost but are replaced by non-homologous mitochondria-specific ones. Moreover, while we have observed frequent loss of uL24, in mitochondria this r-protein is not only universally conserved but additionally stabilized by mitochondria-specific mL45 (Petrov et al. 2018). In any case, while the patterns of loss seem superficially similar, plastids, mitochondria, and bacteria with reduced genomes cannot be compared directly, since a large fraction of the organellar genome has been transferred to the nuclear genome. This leads to a different selection regime and the absence of the selective pressure on the proteome size. For example, spinach chloroplast contains 59 r-proteins, of which 53 are orthologous to *E. coli* ones. Of those, some r-protein genes are still retained in the plastid genome, this groups being comprised of about half of small subunit r-proteins (12 out of 25) and eight out of 33 large subunit r-proteins (Yamaguchi and Subramanian 2000; Yamaguchi et al. 2000). Among frequently lost proteins reported here, r-proteins bL25 and bL30 are missing in spinach, two more are encoded in the spinach chloroplast genome (bL32 and bL36), while bS21, bL9, uL24, uL29, and bL34 are encoded by the nuclear genome.

What are the factors that affect the propensity of a protein to be lost? Previously, proteins frequently lost in evolution have been shown to evolve rapidly, have fewer interactions with other proteins, and lower expression levels (Krylov et al. 2003). Although

243 there are some data on ribosomal heterogeneity in bacteria (Byrgazov et al. 2013), the
244 complete ribosome requires all r-proteins, and r-proteins are organized in operons with a
245 variety of regulatory feedback loops providing tight co-regulation (Lemke et al. 2011), that
246 allowed us not to consider the expression level. Taking into account other factors such as the
247 substitution rate, the number of contacts, and the protein length, we observe that only the
248 number of contacts is a significant determinant of the r-protein loss. The number of contacts
249 is associated with exposure of an r-protein on the ribosome surface. Indeed, r-proteins that
250 are frequently lost tend to be located on the ribosome surface, the exceptions being bL34 and
251 bL36 (Fig. 2ce).

252 The order in which r-proteins are assembled in the ribosome complex (Chen &
253 Williamson 2013) provides an indirect evidence of the order in which r-proteins have been
254 incorporated in the ribosome in the course of evolution. Thus, r-proteins that appear only in
255 the late ribosome intermediates should be relatively young. Seven out of ten frequently lost
256 proteins r-proteins – bS21, bL9, bL25, uL29, bL32, bL34 and bL36 – are included only in the
257 late assembly intermediates. Thus, it can be speculated that frequently lost r-proteins have
258 been incorporated in the ribosome late in evolution. This hypothesis is also supported by "the
259 onion model" (Hsiao et al. 2009; Petrov et al. 2015). Most frequently lost proteins in our
260 dataset are located at the farthest layers from the PTC and thus, they should have appeared
261 the latest in the evolution. Surprisingly, r-proteins bL25, bL34, and bL36 are closer to the
262 PTC then other frequently lost r-proteins, however they are included in the late assembly
263 intermediates (Chen & Williamson 2013). These proteins possibly appeared late in evolution,
264 but then migrated deeper in the ribosome structure. Another interesting question is whether r-
265 proteins are lost at random or there are certain patterns of ribosome simplification, where the
266 loss of one protein leads to the loss of others. Our dataset contained many genomes with
267 incomplete sets of r-proteins, that allowed us to study the patterns of common protein loss.

268 One such pattern is the loss of proteins around the ribosome exit tunnel (TF, and
269 uL29). The trigger-factor (TF) is a chaperone associated with the ribosome exit channel
270 (Hoffman et al. 2010), and is the most frequently lost protein in the pattern. TF is directly
271 interacting with uL23 and uL29 (Ferbitz et al. 2004). We show that the loss of uL29 is
272 correlated with the loss of TF, while the loss of uL23 is not associated with TF. The loss of
273 this functional block is surprising, as even highly reduced genomes tend to retain chaperones
274 (McCutcheon and Moran 2012).

275 One more pattern is the correlated loss of bl32, bL34, and bL36. Two proteins in this

276 pattern – bL32 and bL34 – are positioned near r-proteins forming the exit tunnel, uL22,
277 uL23, uL24, and uL29 (Nikolay et. al 2015), whereas bL36 is far from bL32 and bL34, so
278 this pattern may not be explained just by the spatial proximity. The only common structural
279 feature of these three proteins is that they all are buried deeply in the structure. At that, little
280 is known about the functional role of these r-proteins. bL32 together with bL17 determine the
281 LSU assembly path and without these two r-proteins the subunit core becomes unstable
282 (Davis et. al 2016; Nikolay et. al 2018). bL36 has been reported to be highly conserved in
283 Bacteria, but lacking in Archaea and Eukarya (Maeder and Draper 2005). bL36-deficient
284 ribosomes exhibited disruptions of the rRNA tertiary structure, suggesting that bL36 is
285 important for the 23S rRNA organization and stabilization (Maeder and Draper 2005).
286 Hence, this entire group may be associated with the 23S rRNA stabilization.

287 Among other frequently lost proteins, uL30 is highly conserved in Archaea and
288 Eukarya (Lecompte et al. 2002), where it is thought to be essential for the selenocysteine
289 recognition (Chavatte et al. 2005). The role of this protein in bacteria is not clear. bL25 has
290 been proposed to be essential for interaction with r-protein uL16, the latter being necessary
291 for the ribosome stability (Anikaev et al. 2016). Although the loss of bL25 is tolerated in *E.*
292 *coli*, mutant bacteria have a reduced growth rate (Baba et al. 2006). bS21 is required for the
293 recognition of native templates, and its function resembles the function of bS1 (Van Duin
294 and Wijnands 1981). R-protein bL9 reduces translation frameshifting (Dunkle et al. 2010).
295 uL24 is an LSU assembly initiator protein, together with uL3 (Nikolay et. al 2015). TF and
296 uL29 discussed in detail above are associated with protein folding. Thus, the loss of these
297 proteins should reduce the ribosome fidelity and overall efficacy of protein translation.

298 Surprisingly, all deletions in 23S and 16S rRNAs happened in free loops meaning
299 that the loss of ribosomal proteins was not accompanied by the reduction of their respective
300 binding sites on the ribosomal RNAs. However, the shortening of rRNA is a tendency shared
301 also by mitochondrial rRNA — the SSU mt-rRNA is 40% shorter than bacterial 16S rRNA,
302 while LSU mt-rRNA represents a half of bacterial 23S rRNA (Anderson et al. 1981).
303 Interestingly, deletions in 23S rRNA that have been found in our dataset are located in
304 helices h10, h63, h79, and h98, which have been recently reported to be lost frequently in mt-
305 rRNA of various species (Petrov et al. 2018). Degeneration and losses of a particular element
306 of 16S rRNA, the anti-Shine-Dalgarno sequence, have been reported in various species
307 including symbionts with small genomes (Lim et al. 2012). The anti-SD loss is polyphyletic,
308 as the sequence may be present or absent in representatives of all considered phyla,

309 consistent with previous observations (Amin et al. 2018). While the loss of the anti-SD
310 sequence is a common event among intracellular symbionts, some symbionts with reduced
311 genomes, such as *B. aphidicola*, retain anti-SD (Lim et al. 2012). Nevertheless, we have
312 observed that the loss of anti-SD is significantly associated with the genome reduction and
313 the loss of ribosomal proteins. Moreover, we see an association between the loss of anti-SD
314 and the absence of frequently lost ribosomal proteins bL9 and TF. Surprisingly, none of them
315 are situated near the 3' end of 16S rRNA, where the anti-SD sequence is located. Thus, the
316 association between the absence of these proteins and the anti-SD sequence may not be
317 explained by direct interactions, and changes in structural dynamics should be taken into
318 account.

319 In conclusion, proteins are not lost at random, and there are at least two distinct
320 patterns of loss of spatially adjacent proteins. Overall, frequently lost proteins have less
321 contacts, are located on the ribosome surface, and have been incorporated in the ribosome
322 late in evolution. That is consistent with loss not affecting essential ribosome functions,
323 especially in symbiotic bacteria that live in stable host environments.

324 Materials and Methods

325 Dataset of small bacterial genomes

326 The list of all bacterial species with complete genomes not exceeding 1 Mbp was
327 compiled from the IMG/M database (Chen et al. 2017). The genomic data for all strains of
328 these species (214 genomes in total, Supplementary Table S1) were downloaded from the
329 NCBI FTP; files with protein sequences in February 2017
330 (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/), files with RNA sequences in
331 September 2017 (ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/).

332 The genomes of some strains in the selected species exceeded 1 Mb, but were still
333 retained in the dataset to check whether the gene loss tendency is consistent throughout the
334 species.

335 Annotation of protein domains

336 Ribosomal proteins and trigger factor (TF) were re-annotated in the downloaded
337 genomes (the complete list of proteins is given in Supplementary Table S2) using HMM
338 matrices from the Pfam-A database (Finn et al. 2016). The HMM-profiles for the selected
339 domains were scanned against each genome with the HMMER software (Mistry et al. 2013).

340 A protein was considered to be present in a given genome if the respective domain had a hit
341 in the HMMER search with E-value < 0.001; only the best such hit was retained for further
342 analysis. If the HMMER domain bias had the same order of magnitude as the similarity
343 score, additional filtering was performed as described below.
344 For each domain we created a multiple alignment of proteins with confident HMMer
345 predictions (available at https://github.com/darianick/ribo-simpler). Protein alignments were
346 built using the GUIDANCE2 Server (Penn et al. 2010) with the MAFFT program for
347 building multiple sequence alignments. A candidate protein with a high domain bias value
348 was added to the domain alignment and considered as present if it fitted the multiple
349 sequence alignment keeping the GUIDANCE alignment confidence score above the
350 recommended threshold (0.8). If the confidence score was below the threshold, the protein
351 was removed from the dataset.
352 Two-domain proteins (uL2, uL5, bL7/L12, bL9, uL11, uS4, uS5, TF) were
353 considered to be present if both domains were found.

354 **Identification of independent r-protein losses**

355 To map the phyletic patterns of r-protein losses, we constructed the maximum
356 likelihood tree of concatenated multiple alignments of all conserved r-proteins (see below)
357 using PhyML v. 3.1 (Guindon et al. 2010) with the default parameters and 100 bootstrap
358 replicates (available at https://github.com/darianick/ribo-simpler). Prior to the tree
359 construction all columns containing gaps were removed from the concatenated alignment.
360 This tree was used to calculate the number of independent losses of each protein.

361 **Estimation of the evolutionary rate**

362 The evolutionary rate of a protein was defined as the average branch length in the
363 respective phylogenetic tree. The maximum likelihood phylogenetic tree was built for each
364 ribosomal protein using PhyML with 100 bootstrap replicates (LG+G model was utilized
365 with tree topology and branch lengths optimised ("-o tl" parameters) and with gamma
366 parameter identified for each protein alignment using ProtTest (version 3.4.2) (Darriba et al.
367 2011)). The evolutionary rate for a protein was compared to the average rates for the set of
368 conserved proteins (uL5, uL14, uL15, bL20, uS5, uS7, uS8, uS9, uS11, uS13, and uS17). For
369 each studied protein, the corresponding rates for every conserved protein from the set were
370 calculated in the subtree with the same set of species as in the tree of the studied protein. To
371 compare the results for proteins with different phyletic profiles, we calculated the relative

372 evolutionary score, i.e. the protein's evolutionary rate divided by the corresponding average
373 rate for the set of conserved r-proteins. For two-domain proteins (see Supplementary Table
374 S5) the same gamma parameters and, subsequently, the same evolutionary rates were
375 obtained, with the exception of ribosomal protein bL9 - for this protein we utilize an average
376 evolutionary rate between the two evolutionary rates for domains.

377 **Detection of patterns of ribosomal protein loss**

378 The p-values for pairwise correlations between losses of r-proteins were estimated
379 using Pagel's test for correlated evolution (Pagel 1994) with R-package phytools v0.6-99
380 (Revell 2012) based on vectors of ribosomal protein presence/absence in all studied bacterial
381 strains and the phylogenetic tree of conserved r-proteins (see above). The resulting p-values
382 after the Bonferroni correction were organized in a heatmap with significant hits grouped
383 together and representing patterns of evolutionary correlated r-protein loss. The obtained
384 results are not always stable (e. g. bS21 vs. uL24, uL30 vs. bL32, bS21 vs. bL36 on Fig. 2a)
385 due to the presence of long branches on the tree.

386 **Calculation of the number of contacts**

387 The number of contacts was estimated by measuring pairwise distances between
388 atoms in ribosome PDB structures (PDB ID: 5H5U and 2D3O) (Ma et al. 2016; Schlünzen et
389 al. 2005). Atoms were defined as contacting if the distance between them was at most 5Å.
390 All protein self-contacts were disregarded. When multiple atoms of a studied protein were
391 connected with the same atom in the ribosome structure, only one such contact was retained.
392 The pairwise distances were calculated with the PyMOL Molecular Graphics System,
393 Version 1.8 Schrödinger, LLC and script pairwise_dist.py. For each protein the total number
394 of contacts was calculated. The contacting surface between each studied r-protein and the
395 remaining ribosome was calculated using the "get_area" function in PyMOL by the
396 following formulae:

$$A+(R-A)-R,$$

397 where A is a r-protein surface, R is a surface of a complete ribosome, R - A is a surface of
398 the complete ribosome lacking r-protein A.

399 The distance between r-proteins and PTC was measured using "distance" function in
400 PyMOL with a parameter "mode = 4" to obtain the distance between centroids of r-proteins
401 and PTC. PTC was set as residues 2063-2663 of 23S rRNA.

402 **rRNA analysis**

403 16S and 23S rRNA alignments (available at https://github.com/darianick/ribo-
404 simpler) were built using SINA Alignment Service (Pruesse et al. 2012), and then all
405 common gaps were removed. We considered rRNA deletions if they occurred in more than
406 two genera and were longer than five nucleotides in at least one species. Two deletion blocks
407 where deletions were present at the same set of strains and separated by not more than five
408 nucleotides in the reference sequence were considered as a single deletion block. As a
409 reference we selected 23S and 16S rRNA sequences from PDB ID: 5H5U.

410 A strain was defined as having the anti-Shine-Dalgarno sequence if there was an anti-
411 SD motif CCUCCU at the 3'-end of the strain's 16S rRNA.

412 **Statistical analysis and data visualization**

413 All statistical analyses were performed using R. We used the logistic regression
414 (function glm, family=binomial()) to find determinants of r-proteins loss. The correlation test
415 was performed in R with the cor.test() function. The Wilcoxon(-Mann-Whitney) unpaired
416 rank sum test was performed using "wilcox.exact()" function from "exactRankTests" R
417 package. The dependency between the total number of r-proteins and the genome length
418 (Supplementary Fig. S3) was built in R using the ggplot2 package. Curve smoothing was
419 performed using the "loess" method. Phylogenetic trees were visualized with the iTOL server
420 (Letunic and Bork 2011).

421 Acknowledgements

422 The project was initiated with Sofya Sherstneva and Anna Ergemlidze at the Summer
423 School of Molecular and Theoretical Biology for high school students, supported by the
424 Zimin Foundation. We thank Arthur Zalevsky (FBB MSU) and Georgii Bazykin (IITP,
425 Skoltech) for useful discussions and suggestions. This work was supported by the Russian
426 Science Foundation (grant 18-14-00358).

427 References

428 Akanuma G et al. 2012. Inactivation of Ribosomal Protein Genes in Bacillus subtilis Reveals
429 Importance of Each Ribosomal Protein for Cell Proliferation and Cell Differentiation. J.
430 Bacteriol. 194:6282–6291. doi: 10.1128/JB.01544-12.

431 Akanuma G et al. 2014. Defect in the Formation of 70S Ribosomes Caused by Lack of

Ribosomal Protein L34 Can Be Suppressed by Magnesium. J. Bacteriol. 196:3820–3830. doi: 10.1128/JB.01896-14.

Amin, M. R., Yurovsky, A., Chen, Y., Skiena, S., & Futcher, B. 2018. Re-annotation of 12,495 prokaryotic 16S rRNA 3'ends and analysis of Shine-Dalgarno and anti-Shine-Dalgarno sequences. PloS one, 13(8), e0202767. doi: 10.1371/journal.pone.0202767

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L.,Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., et al. 1981. Sequence and organization of the human mitochondrial genome. Nature 290, 457 – 465. doi: 10.1038/290457a0

Anikaev AY, Isaev AB, Korobeinikova AV, Garber MB, Gongadze GM. 2016. Role of protein L25 and its contact with protein L16 in maintaining the active state of *Escherichia coli* ribosomes *in vivo*. Biochemistry Moscow. 81:19–27. doi: 10.1134/S0006297916010028.

Aseev LV, Boni IV. 2011. Extraribosomal functions of bacterial ribosomal proteins. Mol Biol. 45:739. doi: 10.1134/S0026893311050025.

Baba T et al. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2:2006.0008. doi: 10.1038/msb4100050.

Ban, Nenad, et al. 2014. A new system for naming ribosomal proteins. Current opinion in structural biology 24: 165-169. doi: 10.1016/j.sbi.2014.01.002.

Bokov K, Steinberg SV. 2009. A hierarchical model for evolution of 23S ribosomal RNA. Nature. 457:977–980. doi: 10.1038/nature07749.

Byrgazov, K., Vesper, O., & Moll, I. 2013. Ribosome heterogeneity: another level of complexity in bacterial translation regulation. Current opinion in microbiology, 16(2), 133-139. doi: 10.1016/j.mib.2013.01.009.

Chavatte L, Brown BA, Driscoll DM. 2005. Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. Nature Structural & Molecular Biology. 12:408–416. doi: 10.1038/nsmb922.

Chen I-MA et al. 2017. IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic Acids Res. 45:D507–D516. doi: 10.1093/nar/gkw929.

Chen SS, Williamson JR. 2013. Characterization of the Ribosome Biogenesis Landscape in E. coli Using Quantitative Mass Spectrometry. Journal of Molecular Biology. 425:767–779. doi: 10.1016/j.jmb.2012.11.040.

Choli T, Franceschi F, Yonath A, Wittmann-Liebold B. 1993. Isolation and Characterization of a New Ribosomal Protein from the Thermophilic Eubacteria, *Thermus thermophilus,*

467 *T. aquaticus* and *T. flavus*. Biological Chemistry Hoppe-Seyler. 374:377–384. doi:
468 10.1515/bchm3.1993.374.1-6.377.
469 Darriba, Diego, et al. 2011. ProtTest 3: fast selection of best-fit models of protein evolution.
470 Bioinformatics 27.8: 1164-1165. doi: 10.1093/bioinformatics/btr088.
471 Davis, J. H., Tan, Y. Z., Carragher, B., Potter, C. S., Lyumkis, D., & Williamson, J. R.
472 (2016). Modular assembly of the bacterial large ribosomal subunit. Cell, 167(6), 1610-
473 1622. doi: https://doi.org/10.1016/j.cell.2016.11.020.
474 Dunkle JA, Xiong L, Mankin AS, Cate JHD. 2010. Structures of the Escherichia coli
475 ribosome with antibiotics bound near the peptidyl transferase center explain spectra of
476 drug action. PNAS. 107:17152–17157. doi: 10.1073/pnas.1007988107.
477 Ferbitz L et al. 2004. Trigger factor in complex with the ribosome forms a molecular cradle
478 for nascent proteins. Nature. 431:590–596. doi: 10.1038/nature02899.
479 Finn RD et al. 2016. The Pfam protein families database: towards a more sustainable future.
480 Nucleic Acids Res. 44:D279–D285. doi: 10.1093/nar/gkv1344.
481 Hoffmann A, Bukau B, Kramer G. 2010. Structure and function of the molecular chaperone
482 Trigger Factor. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research.
483 1803:650–661. doi: 10.1016/j.bbamcr.2010.01.017.
484 Hsiao, C., Mohan, S., Kalahar, B. K., & Williams, L. D. 2009. Peeling the onion: ribosomes
485 are ancient molecular fossils. Molecular Biology and Evolution, 26(11), 2415-2425.
486 doi:10.1093/molbev/msp163.
487 Green R, Noller HF. 1997. Ribosomes and Translation. Annu. Rev. Biochem. 66:679–716.
488 doi: 10.1146/annurev.biochem.66.1.679.
489 Grosjean H et al. 2014. Predicting the Minimal Translation Apparatus: Lessons from the
490 Reductive Evolution of Mollicutes Gojobori, T, editor. PLoS Genetics. 10:e1004363.
491 doi: 10.1371/journal.pgen.1004363.
492 Guindon S et al. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood
493 Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 59:307–321. doi:
494 10.1093/sysbio/syq010.
495 Ikegami A, Nishiyama K, Matsuyama S, Tokuda H. 2005. Disruption of rpmJ Encoding
496 Ribosomal Protein L36 Decreases the Expression of secY Upstream of the spc Operon
497 and Inhibits Protein Translocation in Escherichia coli. Bioscience, Biotechnology, and
498 Biochemistry. 69:1595–1602. doi: 10.1271/bbb.69.1595.
499 Izutsu K et al. 2001. Escherichia coli ribosome-associated protein SRA, whose copy number
500 increases during stationary phase. J. Bacteriol. 183:2765–2773. doi:

501 10.1128/JB.183.9.2765-2773.2001.
502 Kaczanowska M, Rydén-Aulin M. 2007. Ribosome Biogenesis and the Translation Process in
503 Escherichia coli. Microbiol. Mol. Biol. Rev. 71:477–494. doi: 10.1128/MMBR.00013-
504 07.
505 Khaitovich P, Mankin AS, Green R, Lancaster L, Noller HF. 1999. Characterization of
506 functionally active subribosomal particles from Thermus aquaticus. PNAS. 96:85–90.
507 doi: 10.1073/pnas.96.1.85.
508 Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene Loss, Protein Sequence
509 Divergence, Gene Dispensability, Expression Level, and Interactivity Are Correlated in
510 Eukaryotic Evolution. Genome Res. 13:2229–2235. doi: 10.1101/gr.1589103.
511 Kurland CG. 1972. Structure and Function of the Bacterial Ribosome. Annual Review of
512 Biochemistry. 41:377–408. doi: 10.1146/annurev.bi.41.070172.002113.
513 Lecompte O, Ripp R, Thierry J-C, Moras D, Poch O. 2002. Comparative analysis of
514 ribosomal proteins in complete genomes: an example of reductive evolution at the
515 domain scale. Nucleic Acids Res. 30:5382–5390.
516 Lemke, J. J., Sanchez-Vazquez, P., Burgos, H. L., Hedberg, G., Ross, W., & Gourse, R. L.
517 2011. Direct regulation of Escherichia coli ribosomal protein promoters by the
518 transcription factors ppGpp and DksA. Proceedings of the National Academy of
519 Sciences, 108(14), 5712-5717.
520 Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of
521 phylogenetic trees made easy. Nucl. Acids Res. 39:W475–W478. doi:
522 10.1093/nar/gkr201.
523 Lim, K., Furuta, Y., & Kobayashi, I. 2012. Large variations in bacterial ribosomal RNA
524 genes. Molecular biology and evolution, 29(10), 2937-2948. doi:
525 10.1093/molbev/mss101.
526 Ma C et al. 2017. Mechanistic insights into the alternative translation termination by ArfA
527 and RF2. Nature. 541:550–553. doi: 10.1038/nature20822.
528 Maeder, C., & Draper, D. E. (2005). A small protein unique to bacteria organizes rRNA
529 tertiary structure over an extensive region of the 50 S ribosomal subunit. Journal of
530 molecular biology, 354(2), 436-446. doi: 10.1016/j.jmb.2005.09.072.
531 Maier, U. G., Zauner, S., Woehle, C., Bolte, K., Hempel, F., Allen, J. F., & Martin, W. F.
532 2013. Massively convergent evolution for ribosomal protein gene content in plastid and
533 mitochondrial genomes. Genome biology and evolution, 5(12), 2318-2329. doi: 10.1093/
534 gbe/evt181.

535 McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nature
536 Reviews Microbiology. 10:13–26. doi: 10.1038/nrmicro2670.
537 Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search:
538 HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41:e121–
539 e121. doi: 10.1093/nar/gkt263.
540 Nikolay, Rainer & Bruck, David & Achenbach, John & Nierhaus, Knud. 2015. Ribosomal
541 Proteins: Role in Ribosomal Functions. 10.1002/9780470015902.a0000687.pub4.
542 Nikolay, R., Hilal, T., Qin, B., Mielke, T., Bürger, J., Loerke, J., ... & Spahn, C. M.
543 (2018). Structural visualization of the formation and activation of the 50S ribosomal
544 subunit during in vitro reconstitution. Molecular cell, 70(5), 881-893. doi:
545 10.1016/j.molcel.2018.05.003
546 Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The Structural Basis of Ribosome
547 Activity in Peptide Bond Synthesis. Science. 289:920–930. doi:
548 10.1126/science.289.5481.920.
549 Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the
550 comparative analysis of discrete characters. Proceedings of the Royal Society of London.
551 Series B: Biological Sciences, 255(1342), 37-45. doi:
552 https://doi.org/10.1098/rspb.1994.0006
553 Penn O et al. 2010. GUIDANCE: a web server for assessing alignment confidence scores.
554 Nucleic Acids Research. 38:W23–W28. doi: 10.1093/nar/gkq443.
555 Petrov, A. S., Gulen, B., Norris, A. M., Kovacs, N. A., Bernier, C. R., Lanier, K. A., ... &
556 Williams, L. D. 2015. History of the ribosome and the origin of translation. Proceedings
557 of the National Academy of Sciences, 112(50), 15396-15401.
558 doi:10.1073/pnas.1509761112.
559 Petrov, A. S., Wood, E. C., Bernier, C. R., Norris, A. M., Brown, A., & Amunts, A. (2018).
560 Structural Patching Fosters Divergence of Mitochondrial Ribosomes. Molecular biology
561 and evolution, 36(2), 207-219. doi: 10.1093/molbev/msy221.
562 Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple sequence
563 alignment of ribosomal RNA genes. Bioinformatics. 28:1823–1829. doi:
564 10.1093/bioinformatics/bts252.
565 Ramakrishnan V. 2002. Ribosome Structure and the Mechanism of Translation. Cell.
566 108:557–572. doi: 10.1016/S0092-8674(02)00619-0.
567 Revell, L. J. 2012 phytools: An R package for phylogenetic comparative biology (and other
568 things). Methods Ecol. Evol., 3, 217-223. doi: 10.1111/j.2041-210X.2011.00169.x

Roberts E, Sethi A, Montoya J, Woese CR, Luthey-Schulten Z. 2008. Molecular signatures of ribosomal evolution. PNAS. 105:13953–13958. doi: 10.1073/pnas.0804861105.

Schlünzen F et al. 2005. The Binding Mode of the Trigger Factor on the Ribosome: Implications for Protein Folding and SRP Interaction. Structure. 13:1685–1694. doi: 10.1016/j.str.2005.08.007.

Schmeing TM, Moore PB, Steitz TA. 2003. Structures of deacylated tRNA mimics bound to the E site of the large ribosomal subunit. RNA. 9:1345–1352. doi: 10.1261/rna.5120503.

Schuwirth BS et al. 2005. Structures of the Bacterial Ribosome at 3.5 Å Resolution. Science. 310:827–834. doi: 10.1126/science.1117230.

Shoji S, Dambacher CM, Shajani Z, Williamson JR, Schultz PG. 2011. Systematic chromosomal deletion of bacterial ribosomal protein genes. J Mol Biol. 413:751–761. doi: 10.1016/j.jmb.2011.09.004.

Smith TF, Lee JC, Gutell RR, Hartman H. 2008. The origin and evolution of the ribosome. Biol Direct. 3:16. doi: 10.1186/1745-6150-3-16.

Van Duin J, Wijnands R. 1981. The Function of Ribosomal Protein S21 in Protein Synthesis. European Journal of Biochemistry. 118:615–619. doi: 10.1111/j.1432-1033.1981.tb05563.x.

Wimberly BT et al. 2000. Structure of the 30S ribosomal subunit. Nature. 407:327–339. doi: 10.1038/35030006.

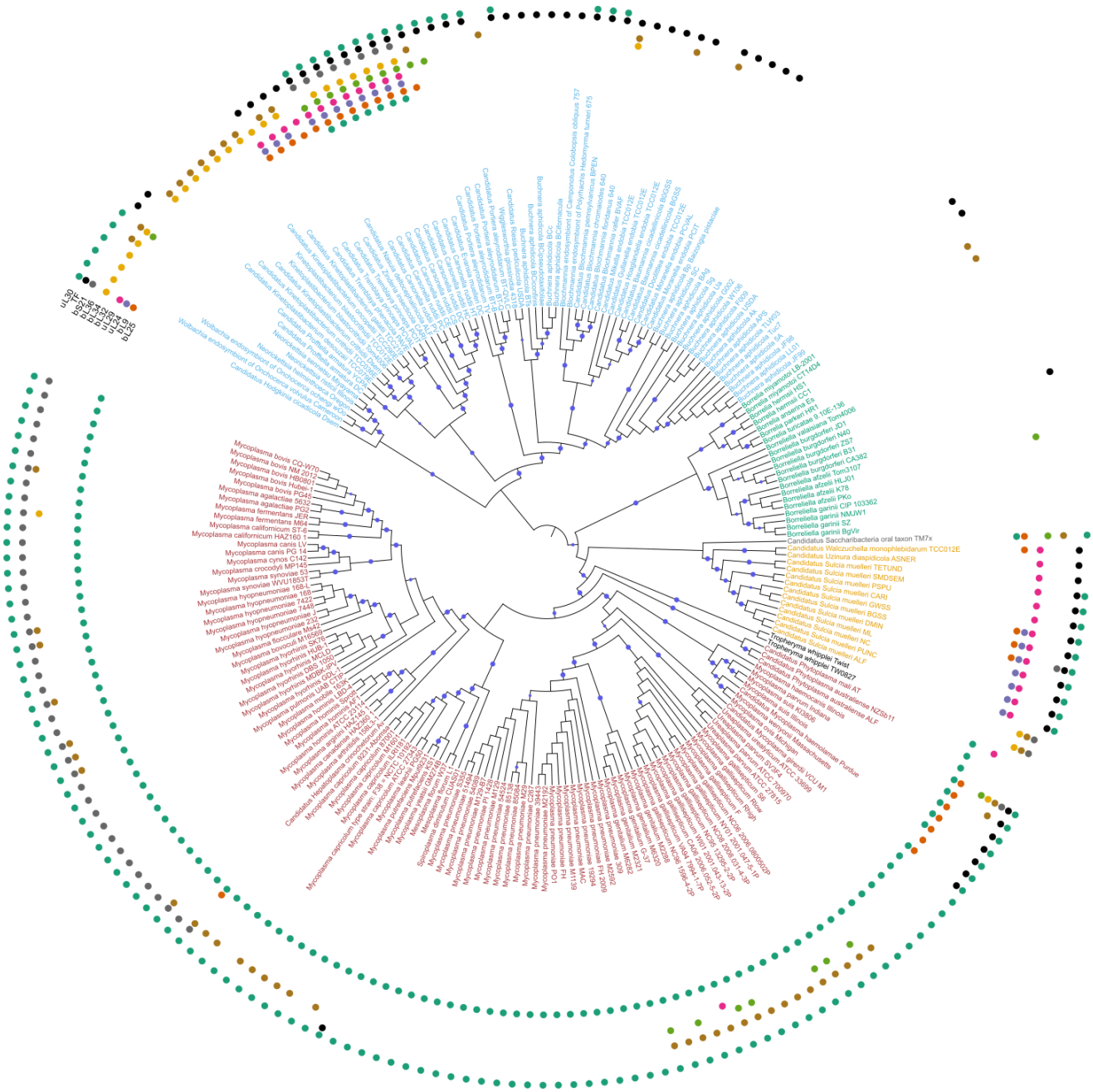
Yamaguchi K., von Knoblauch K., Subramanian A.R. 2000. The plastid ribosomal proteins: identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). J. Biol. Chem. 275, 28455–28465. doi: 10.1074/jbc.M004350200.

Yamaguchi, K., & Subramanian, A. R. 2000. The Plastid Ribosomal Proteins IDENTIFICATION OF ALL THE PROTEINS IN THE 50 S SUBUNIT OF AN ORGANELLE RIBOSOME (CHLOROPLAST). Journal of Biological Chemistry, 275(37), 28466-28482. doi: 10.1074/jbc.M005012200.

Yutin N, Puigbò P, Koonin EV, Wolf YI. 2012. Phylogenomics of Prokaryotic Ribosomal Proteins. PLOS ONE. 7:e36972. doi: 10.1371/journal.pone.0036972.

Figures

Figure 1. Maximum likelihood phylogenetic tree of analyzed bacterial species, constructed by concatenated alignment of conserved r-proteins. The tree was constructed by PhyML with 100 bootstrap replicates. In this representation branch lengths are ignored (the tree with branch lengths is provided as Supplementary Fig. S2). Bootstrap values in the range 0.9–1 are shown by blue circles. The absence of one of ten frequently lost proteins (bL9, bL21, L24, bL25, uL29, bL32, bL34, bL36, bS21, TF) in a strain is marked by a colored circle. Leaves are colored by the phyla: *Actinobacteria* – black, *Bacteroidetes* – yellow, *Proteobacteria* – blue, *Spirohaetes* – green, *Tenericutes* – red, unclassified bacteria – grey.

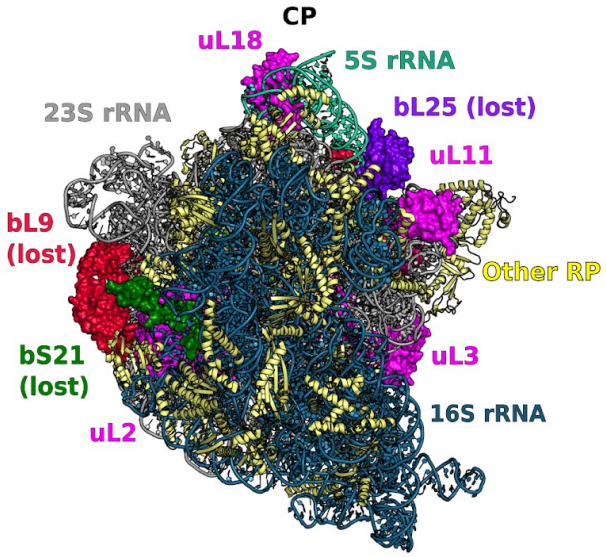


607 **Figure 2. Patterns of r-protein loss.** (a) Pagel's test of correlated evolution between vectors
608 of ribosomal protein presence/absence in bacterial strains with the Bonferroni correction.
609 Insignificant p-values are colored blue, the significant ones are colored red (the more
610 significant, the more intense). Only ten frequently lost proteins are considered. (b) The crown
611 view of the 70S ribosome (the typical crown view position is marked using several ribosomal
612 proteins (magenta), the position of the central protuberance and 5S rRNA (light green) and
613 the position of 16S rRNA (dark blue)) and various perspectives (c,e) of the ribosome
614 showing positions of frequently lost proteins. PDB ID: 5H5U (*Escherichia coli*) (Ma et al.
615 2017). (d) 23S rRNA and r-proteins which are connected with trigger factor. PDB ID: 2D3O
616 (*Deinococcus radiodurans*) (Schlünzen et al. 2005). Protein labels are of the same color as
617 the respective proteins in the structure. In (c), (e) only frequently lost ribosomal proteins are
618 labeled.

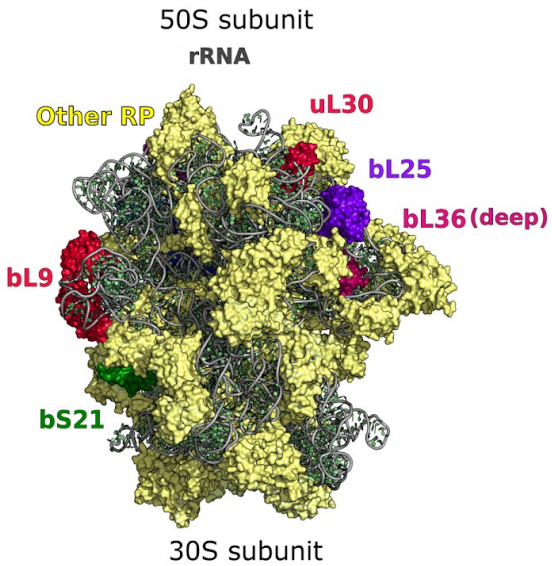
a

	bL36	bL34	bL32	uL24	bL9	uL29	TF	uL30	bS21	bL25
bL25	2.5e-08	2.8e+01	1.0e+00	9.6e+00	1.1e+01	2.4e+01	7.0e-01	4.6e+00	8.7e-02	0.0e+00
bS21	3.7e-04	2.1e+01	3.3e-29	2.4e-02	2.4e+00	1.0e+01	1.2e+00	8.5e+00	0.0e+00	8.7e-02
uL30	1.2e-06	2.6e+01	1.2e-01	6.8e+00	7.0e+00	4.6e+00	3.2e+00	0.0e+00	8.5e+00	4.6e+00
TF	1.1e+00	1.5e+01	3.5e-28	7.6e-01	5.1e-01	1.1e-06	0.0e+00	3.2e+00	1.2e+00	7.0e-01
uL29	1.2e+00	3.2e+01	2.9e-28	1.2e-01	1.4e+00	0.0e+00	2.4e-06	4.6e+00	1.0e+01	2.4e+01
bL9	3.9e+01	2.5e+01	2.8e-23	8.0e-01	0.0e+00	1.4e+00	5.1e-01	7.0e+00	2.4e+00	1.1e+01
uL24	1.4e+01	8.8e+00	2.5e-28	0.0e+00	8.0e-01	1.2e-01	7.6e-01	6.8e+00	7.0e-02	9.6e+00
bL32	1.5e-23	6.7e-27	0.0e+00	9.7e-19	5.9e-25	2.3e-27	1.8e-28	5.7e-17	2.1e-29	1.0e+00
bL34	1.3e-02	0.0e+00	5.7e-22	8.8e+00	2.5e+01	3.2e+01	1.5e+01	2.6e+01	2.1e+01	2.8e+01
bL36	0.0e+00	1.3e-02	4.8e-25	1.4e+01	3.4e+01	1.2e+00	1.1e+00	1.4e-06	1.3e+00	2.0e-08

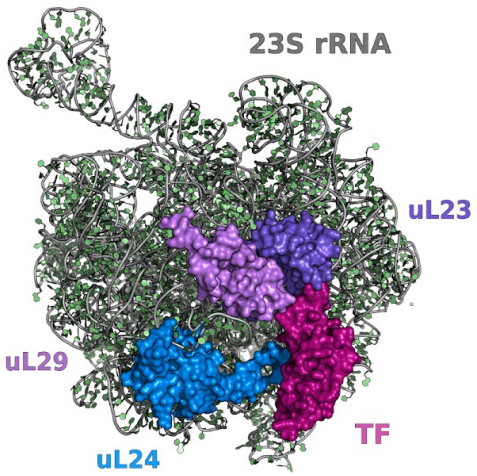
b



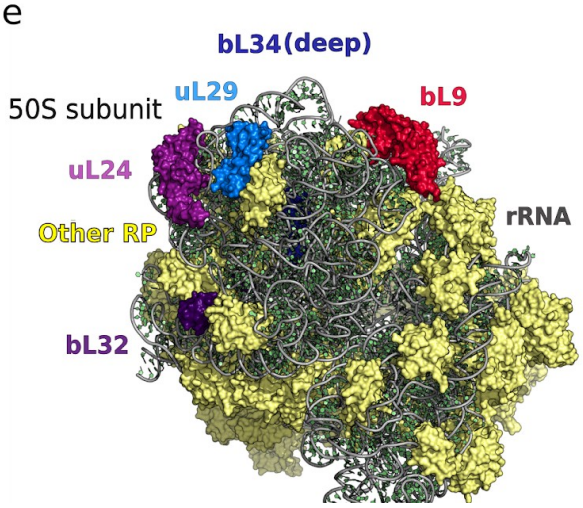
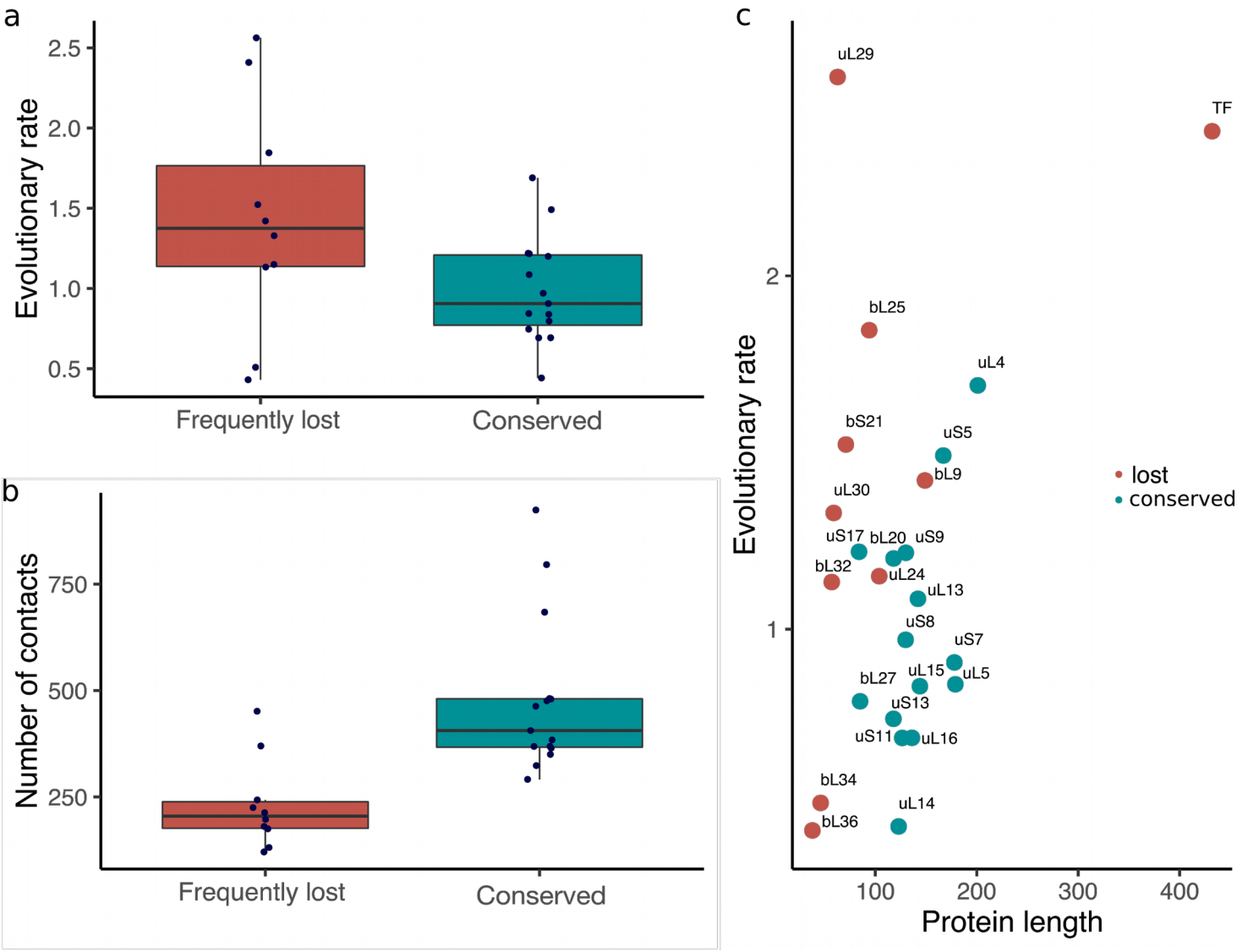
c



d



e



619 **Figure 3. Determinants of the r-protein loss.** The differences in the evolutionary rate (a)
620 and the number of contacts (b) between the conserved (blue) and frequently lost (red)
621 ribosomal proteins. (c) The dependency between the evolutionary rate and protein length
622 (number of amino acid residues) for both conserved and frequently lost ribosomal proteins.