Functional constraints on replacing an essential gene with its ancient and modern homologs

Betül Kacar[1,‡,*] Eva Garmendia[2,‡], Nurcan Tunçbağ[3], Dan I. Andersson[2], Diarmaid Hughes[2]

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA
Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden
Department of Health Informatics, Graduate School of Informatics, METU, Ankara, Turkey

‡ Equal contribution
* Address correspondence to: Betül Kacar kacar@fas.harvard.edu

Abstract

The complexity hypothesis posits that network connectivity and protein function are two important determinants of how a gene adapts to and functions in a foreign genome. Genes encoding proteins that carry out essential informational tasks in the cell, in particular where multiple interaction partners are involved, are less likely to be transferable to a foreign organism. Here we investigated the constraints on transfer of a gene encoding a highly conserved informational protein, translation elongation factor Tu (EF-Tu), by systematically replacing the endogenous *tufA* gene in the *Escherichia coli* genome with its extant and ancestral homologs. The extant homologs represented *tuf* variants from both near and distant homologous organisms. The ancestral homologs represented phylogenetically resurrected *tuf* sequences dating from 0.7 to 3.6 bya. Our results demonstrate that all of the foreign *tuf* genes are transferable to the *E. coli* genome, provided that an additional copy of the EF-Tu gene, *tufB,* remains present in the *E. coli* genome. However, when the *tufB* gene was removed, only the variants obtained from the γ-proteobacterial family (extant and ancestral), supported growth. This demonstrates the limited functional interchangability of *E. coli tuf* with its homologs. Our data show a linear correlation between relative bacterial fitness and the evolutionary distance of the extant *tuf* homologs inserted into the *E. coli* genome. Our data and analysis also suggest that the functional conservation of protein activity, and its network interactivity, act to constrain the successful transfer of this essential gene into foreign bacteria.

1

Introduction

31

32   The Complexity Hypothesis assigns the function and the network interactivity of a gene

33   product as two primary factors determining a gene's capacity to successfully transfer and

34   adapt in another genome (Jain, et al. 1999; Sorek, et al. 2007; Wellner, et al. 2007; Cohen, et

35   al. 2011; Kitahara and Miyazaki 2013). Consequently, genes involved in central informational

36   tasks (i.e. replication, transcription and translation) are expected to be less transferable to a

37   foreign genome, in comparison with genes involved in other activities (Smith, et al. 1992; Jain,

38   et al. 2003; Sorek, et al. 2007).

39   Genomic replacement of an essential gene with another homolog may potentially

40   disturb the gene product's function depending on the degree of functional equivalence or

41   compatibility between the two genes (Shen and Huang 1986; Kacar and Gaucher 2013; Blank,

42   et al. 2014; Bershtein, et al. 2015). Alterations in gene dosage resulting from the acquisition of

43   a foreign gene may also perturb cellular homeostasis resulting in a lower transcription rate or

44   an altered global pattern of gene expression, network function and organismal survivability (Ito,

45   et al. 1998; Coulomb, et al. 2005; Drummond, et al. 2005; Durfee, et al. 2008; Zotenko, et al.

46   2008; Lind and Andersson 2013; Bershtein, et al. 2015; Karcagi, et al. 2016).

47   Genes involved in the translation of mRNA into protein perform one of the most crucial

48   informational tasks in the cell, and based on phylogenomic analysis they are expected to be

49   highly resistant to gene transfer (Ciccarelli, et al. 2006; Sorek, et al. 2007). There are examples

50   demonstrating that some ribosomal protein genes can be integrated into foreign genomes

51   under certain conditions (Brochier, et al. 2000; Lind, Berg, et al. 2010; Lind, Tobin, et al. 2010;

52   Knoppel, et al. 2016). However, few studies have systematically tested whether there is a direct

53   correlation between organism fitness and the evolutionary distance between an essential

54   endogenous gene and its substituted (ancestral or extant) homolog.

2

55   Here we focus on the bacterial elongation factor Tu (EF-Tu) protein, encoded by *tuf*,

56   one of the most ancient and highly conserved proteins. EF-Tu has an essential function in the

57   translation machinery by delivering aminoacylated tRNA molecules into the A-site of the

58   ribosome (Miller and Weissbach 1977). EF-Tu is encoded by two genes in *E. coli*, *tufA* and

59   *tufB*, generated by an ancient duplication event thought to be specific to the proteobacterial

60   lineage preceding the Cambrian period (Baldauf, et al. 1996; Lathe and Bork 2001). Expression

61   of EF-Tu is primarily driven by *tufA*, with 66% of cellular EF-Tu expressed from the *tufA* gene

62   (Abdulkarim and Hughes 1996). EF-Tu protein levels in the cell are correlated with cellular

63   fitness and intrinsically regulated in order to maintain growth rate (Tubulekas and Hughes

64   1993a; Brandis, et al. 2016; Kacar, et al. 2016). EF-Tu belongs to the ancient protein repertoire

65   of the cell, evolves slowly and serves as a functional fossil by participating in ancient and

66   conserved functions (Hsiao, et al. 2009). It remains unclear whether *tuf* genes are replaceable

67   by their ancient counterparts or homologs obtained from an extant organism. Answering this

68   question would allow us to explore the limits of interchangeability for the *E. coli tuf* gene, and

69   to ascertain a pattern within and between bacterial lineages across time and divergence. We

70   sought guidance from a methodology referred to as ancestral sequence reconstruction

71   (Pauling and Zuckerkandl 1963; Benner 1995; Thornton 2004; Benner, et al. 2007; Gaucher, et

72   al. 2008; Kacar 2016) and accessed reconstructed ancestral *tuf* variants, as well as modern *tuf*

73   gene sequences in order to observe the patterns of interchangeability among multiple nodes

74   along the EF-Tu phylogenetic tree. We utilized a set of foreign genes representing *E. coli* EF-Tu

75   homologs from closely- and distantly-related bacteria, as well as phylogenetically inferred

76   ancestral EF-Tu proteins dating from 0.7 to 3.6 bya, thus accessing interspecies (modern) and

77   ancestral (paleogenetic) axes. We determined the fitness effects of the introduction of a foreign

78   *tuf* gene into each strain by replacing the native *E. coli tufA* gene with foreign variants, and

79   asked whether these foreign genes could support cell viability when the *tufB* gene was

80   removed from the chromosome. We examined the impact of the *tuf* gene replacements on

81   growth rate and protein levels, as well as on protein function-structure, and assessed the extent

82   of lateral and ancestral phylogenetic distances between the alien gene and host genome that

83   yielded viable organisms.

84   Results

85   **Replacement of the *tufA* gene in *E. coli* reduces relative fitness**.

86   Using genetic recombineering, we generated a set of *E. coli* strains in which the *tufA* gene

87   coding sequence was precisely replaced by the coding sequence of its ancestral and modern

88   homologs (Figure 1, Figure S1A). The 16 *tuf* homologs cover bacterial species from a wide

89   span of taxa (*Yersinia enterocolitica*, *Vibrio cholerae*, *Pseudomonas aeruginosa, Legionella*

90   *pneumophila*, *Bartonella hanselae, Streptococcus pyogenes*, *Bacillus subtilis*, *Thermus*

91   *thermophilus, Mycobacterium smegmatis* and *Thermatoga maritima*) as well as six ancestral

92   sequences that extend deep into the bacterial phylogenetic tree (Figure 1). These six

93   sequences represent ancestral nodes dating from approximately 0.7 bya, back to the last

94   common ancestor of bacterial *tuf*, dated to approximately 3.6 bya (Battistuzzi, et al. 2004;

95   Gaucher, et al. 2008). These homologs of *tuf* encode EF-Tu variants that range in amino acid

96   identity, from 93.9% (*Y. enterocolitica*) to 69.4% (*T. maritima*), relative to *E. coli* EF-Tu (Figure

97   S2). Nucleotide sequences of the ancestral *tuf* genes are shown in Table S4.

98

99   To construct a set of isogenic strains, the endogenous *E. coli tufA* gene was replaced with

100   each of the *tuf* variants while the second endogenous *tuf* gene, *tufB*, remained intact in the

101   genome. All engineered strains in which *tufA* was replaced with a foreign *tuf* gene retained

102   viability (Figure 2). The effects on relative fitness of the foreign *tuf* genes were determined by

103   measuring the exponential growth rates in LB and relating them to that of the isogenic wild-

104   type (carrying native *tufA* and *tufB*) where relative fitness was set to 1.0. The relative fitness of

105   each of the engineered constructs varied from 0.96 down to 0.77 (Figure 2). The relative fitness

106   of *E. coli* in which *tufA* was deleted from the chromosome was 0.7. The similarity in relative

107   fitness between *E. coli* lacking *tufA* and some of the strains carrying foreign *tuf* genes raised

108   the question of whether all of the foreign genes would be capable of supporting viability in the

109   absence of a functioning *tufB* gene.

110

111   Because the *tufB* gene is located between the ribosomal RNA operons *rrnB* and *rrnE,* in a

112   region that is subject to frequent amplification (Anderson and Roth 1981), we asked whether

113   *tufB* was amplified in any of the strains carrying a foreign *tuf* allele. Our expectation was that a

114   less effective, or completely inactive, foreign *tuf* gene, might select for genomes in which *tufB*

115   was amplified as a fitness-compensatory mechanism. Using real-time quantitative PCR (RT-

116   qPCR) we found that the *tufB* region was duplicated or triplicated in 5 of the 16 engineered

117   strains (Figure 2). The 5 strains in which the *tufB* region was amplified are those that carry the

118   most distant *tuf* homologs, compatible with selection for improved fitness.

119

120   **A subset of foreign *tuf* genes support viability**

121   We next asked whether any of the foreign *tuf* genes could support viability in the absence of *E.*

122   *coli tufB*, by attempting to remove the endogenous *tufB* gene from the chromosome of each of

123   the 16 strains carrying a foreign *tuf* gene, as outlined in Figure S1B. In the absence of *tufB*, the

124   only foreign *tuf* sequences that supported viability were those from *Y. enterocolitica, V.*

125   *cholerae, P. aeruginosa*, and AnEF1, the youngest of the ancestral *tuf* genes, at ca. 0.7 bya,

126   (Figure 2), whereas the phylogenetically more distant *tuf* genes did not. The relative fitness of

127   each of the 4 viable strains was assessed by comparing their growth rates to that of the

128   isogenic *E. coli* Δ*tufB* strain (carrying native *tufA*). The relative fitness values ranged from 0.95

129   (*Y. enterocolitica*), down to 0.40 in the case of the *tuf* sequence from *P. aeruginosa* (Figure 3).

5

Fitness of engineered strains correlates with EF-Tu protein levels

A key question is why the EF-Tu gene replacements reduce fitness? Two possibilities (which are not mutually exclusive) are that the foreign genes are sub-optimally expressed (i.e. a concentration problem) or that they are sub-optimal in their function and interaction with the protein synthesis machinery and cellular network (i.e. an activity or toxicity problem). To evaluate the possibility that the foreign genes were sub-optimally expressed we measured EF-Tu abundance by LC-MS/MS in each of the viable strains (Figure 3). This analysis demonstrated that the level of EF-Tu relative to total protein varied significantly between engineered strains carrying different foreign *tuf* genes (Figure 3). With the exception of the ancestral gene, AnEFTU1, there is a good correlation between relative fitness and the concentration of EF-Tu. These data suggest that in most cases at least some of the reduction in relative fitness associated with foreign *tuf* genes is because bacteria do not produce an adequate level of EF-Tu to support fast growth. In the case of AnEFTU1 the reduced fitness may be more closely associated with reduced specific activity and/or toxicity.

Conservation of EF-Tu residues correlates with viability in *E. coli*

The established function of EF-Tu is to deliver aminoacylated tRNAs into the A-site on mRNA-programmed ribosomes in order to drive rapid and accurate protein synthesis. This function requires that EF-Tu must have a sequence and structure that can efficiently interact with GTP, EF-Ts, each of the elongator tRNAs, and with the ribosome. EF-Tu has a highly conserved structure which must be capable of undergoing major structural rearrangements during its functional cycle (Nilsson and Nissen 2005; Yu, et al. 2009). All 17 EF-Tu sequences were aligned with the EF-Tu sequence from *E. coli* (Figure 4) to aid visualization of amino acids differences between EF-Tu's associated with viability versus non-viability. The four viable foreign EF-Tu's (green in Figure 4) show a bias towards a higher conservation of amino acid

155 identity (84% to 94%), relative to *E. coli* EF-Tu, than the twelve non-viable EF-Tu's (82% to

156 69%). It is interesting to note that there is only a small difference in total percent identity

157 between several of the non-viable EF-Tu's, (AnEF2, AnEF3, AnEF4, AnEF6, and *B. hanselae,*

158 80-82% similarity), and the most distantly related viable EF-Tu (*P. aeruginosa,* 84% similarity*)*,

159 suggesting that the loss of viability in these cases might be associated with changes in a few

160 important residues of EF-Tu.

161

162 **Connectivity analysis of EF-Tu reveals an extensive interaction network**

163 Several studies suggest that protein connectivity in a given network might modulate its activity

164 in the cell (Fraser, et al. 2002; Hahn, et al. 2004; Lemos, et al. 2005). To examine this, we

165 retrieved an *E. coli* interactome from the HitPredict database (Patil, et al. 2011) and measured

166 the connectivity of EF-Tu in this network. The connectivity of EF-Tu was quantified by

167 measuring its so-called 'degree centrality'. Degree centrality is defined as the number of

168 interactions that a given node has in its network. The average degree centrality of all proteins in

169 the *E. coli* interactome is 12, whereas the degree centrality of EF-Tu is much higher at 172.

170 Relative to all proteins in the interactome, EF-Tu ranks among the top ten most connected

171 proteins in *E. coli* (Supplementary Table 1). Interestingly, the set of its first interaction partners

172 is enriched in essential proteins (p-value = 3.66 x 10^{-21}), as calculated via the HitPredict

173 Database (Figure S4). Accordingly, this analysis suggests the possibility that, in addition to

174 affecting the specific activity of EF-Tu in protein synthesis, additional deleterious effects on

175 fitness might be due to the foreign *tuf* homologs disturbing the extensive interaction network of

176 EF-Tu.

177

178

7

Discussion

179

180   We generated *E. coli* strains in which the *E. coli tufA* gene was replaced by ancestral

181   and modern homologs of *tuf*, from a broad spectrum of species and ancestral nodes. The

182   origins of the ancestral EF-Tu sequences ranged in age from the Precambrian era,

183   approximately 0.7 bya, back to the last universal common bacterial EF-Tu ancestor,

184   approximately 3.6 bya (Figure 1). We showed that homologs of EF-Tu encoded by *tuf* genes

185   from within the γ-proteobacteria, including one of the reconstructed ancestral node sequences,

186   AnEF1, are functionally active in *E. coli* and support viability when present as the only *tuf* gene

187   in the chromosome (shown in green in Figure 2). In contrast, more distantly related homologs

188   and ancestral sequences were unable to support viability as the sole *tuf* gene. Among the four

189   viable homologs, there was a good correlation (R = 0.938) between phylogenetic distance from

190   *E. coli* EF-Tu and the magnitude of reduced growth fitness for the three homologs from extant

191   species (Figure 2). The exception was the reconstructed ancestral node sequence, AnEF1,

192   where the decrease in relative fitness was much greater than predicted by phylogenetic

193   distance (Figure 2).

194   The correlation between phylogenetic distance from *E. coli* and relative growth fitness for

195   the viable homologs raises the question of whether the underlying cause of the reduced fitness

196   is a reduction in the specific activity of the foreign EF-Tu's and/or a reduction in the amount of

197   EF-Tu produced. By measuring EF-Tu protein concentration as a function of total protein

198   concentration for each viable strain, including *E. coli* carrying only *tufA*, we observed a strong

199   correlation between relative growth fitness and EF-Tu concentration for the each of the four EF-

200   Tu's from extant species (Figure 3). The slope of the correlation suggests that most of the

201   reduced fitness associated with the EF-Tu homologs from the extant species could be attributed

202   to a reduced level of EF-Tu rather than a reduced activity. Consistent with this, experiments

203   have shown that reductions in growth rate are linearly correlated with reductions in cellular EF-

8

204 Tu concentrations (Tubulekas and Hughes 1993b). Once again the exception to this correlation

205 is found for the reconstructed ancestral AnEF1, where a very high concentration of EF-Tu was

206 associated with a low relative fitness (Figure 3) consistent with AnEF1 having a low specific

207 activity. This conclusion is supported by the results of *in vitro* translation, where in a system

208 containing only *T. thermophilus* components (tRNA's, EF-Ts, ribosomes) AnEF1 supported

209 protein synthesis but with only 30% of the activity of native *E. coli* EF-Tu, demonstrating that

210 AnEF1 can participate in peptide synthesis, albeit in a diminished fashion relative to their

211 modern counterpart EF-Tu (Zhou, et al. 2012).

212   Why, given that the *tuf* gene regulatory regions are identical in all strains, do the foreign

213 *tuf* coding sequences cause a reduction in the level of EF-Tu produced? Possible explanations

214 include that the nucleotide sequence introduced with the foreign genes affects *tuf* mRNA half-

215 life as previously shown for a mutant of *tuf* (Hammarlöf and Hughes 2008), or that the altered

216 nucleotide sequence reduces translation efficiency, possibly by effects mediated through

217 altered codon usage, or by affecting transcription-translation coupling, as recently shown for

218 *tufB* (Brandis, et al. 2016; Brandis and Hughes 2016).

219   An important question is why some of the not-too-distant foreign homologs are unable

220 to support viability. At the level of total amino acid similarity there is very little separating viable

221 from non-viable EF-Tu sequences, with the boundary falling at approximately 84% similarity to

222 *E. coli* EF-Tu (Figure S2). Given the very high level of conservation of EF-Tu it is possible that

223 the non-viability of some foreign EF-Tu's might be related to the alteration of just one or a few

224 critically important residues. Indeed, many single amino acid substitutions in EF-Tu have been

225 shown to generate protein variants that do not support viability (Abdulkarim, et al. 1991),

226 including one single amino acid substitution in EF-Tu that permits ternary complex formation

227 but abolishes translation activity by preventing ternary complex interaction with the ribosome

228 (Tubulekas and Hughes 1993b). To facilitate an assessment of the amino acid differences

9

229   between viable and non-viable EF-Tu's their amino acid sequences were aligned (Figure 4). Of

230   the 393 residues in *E. coli* EF-Tu, there were 116 residues that were identical in all of the viable

231   EF-Tu homologs, but differed in at least one of the 12 non-viable homologs. Variation at one or

232   more of these residues might explain the difference between EF-Tu viability and non-viability in

233   *E. coli*. The 116 residues were distributed between each of the three structural domains of EF-

234   Tu, with 48 in the G-domain, 38 in domain 2 and 30 in the C-terminal domain 3 (Figure S3-A).

235   Several of the 116 variant residues lie within functionally important regions of EF-Tu including

236   those involved in coordinating GTP hydrolysis, interaction with EF-Ts, and interaction with the

237   ribosome (Kothe, et al. 2004; Kavaliauskas, et al. 2012; Thirup, et al. 2015). For example, in

238   each of the non-viable EF-Tu's there were from 3 – 8 variant residues in regions of the G-

239   domain that are known to be functionally important for coordinating the hydrolysis of GTP on

240   EF-Tu during protein synthesis. These variant residues are located in the Conserved regions I

241   and II, the Effector region, and the Switch I and Switch II loops of the G-domain of EF-Tu

242   (Figure S3-B). The altered residues potentially affecting GTP hydrolysis are at V20, G41, A43,

243   R44, F46, N51, N63, T64, C81, and V88 (Figure 4). In addition, there are alterations in the P-

244   loop, the Switch II region, and parts of domain 3 that are involved in interactions with EF-Ts

245   and in helix D of EF-Tu (residues 144 – 156), that is involved in interactions with EF-Ts and

246   protein L7/L12 on the ribosome (Kothe, et al. 2004; Thirup, et al. 2015). We can only speculate

247   on the exact reason for the non-viability of these EF-Tu homologs. It seems unlikely that it is

248   directly related to defects in binding or hydrolyzing GTP, given that this process involves highly

249   conserved residues and structures and that the EF-Tu's from extant organisms must be

250   capable of supporting viability, including GTP binding and hydrolysis, in their natural system.

251   Similarly, each of the ancestral homologs can support *in vitro* translation, albeit at a low

252   efficiency, arguing that they also can bind and hydrolyze GTP. A similar line of reasoning could

253   also rule out interactions with aa-tRNA's as the cause of non-viability. Perhaps the most

254   plausible reason for non-viability is defective interactions with EF-Ts and/or the ribosome. It

255   seems reasonable to suggest that EF-Tu has co-evolved with EF-Ts and the ribosome, to

256   modulate the efficiency of these interactions in each species. We suggest accordingly that the

257   cause of non-viability for distantly related EF-Tu's is not that they cannot function as enzymes

258   capable of forming a ternary complex and hydrolyzing GTP, but rather that they are defective in

259   one or more of the other important interactions made by EF-Tu, namely, with EF-Ts, with the

260   mRNA-programmed ribosome, and possibly even interactions outside of protein synthesis

261   involving one or more members of EF-Tu's extensive protein interaction network (Figure S4).

262   Co-evolution of EF-Tu with its interaction partners would create a barrier to transfer for EF-Tu's

263   beyond a certain threshold.

264   We previously hypothesized that evolutionary novelties are more likely to be shared

265   between a descendant and its ancient homolog than between two currently existing protein

266   homologs (Kacar and Gaucher 2012). Accordingly, replacing an existing gene with its ancient

267   homolog may have a smaller negative fitness impact on the organism relative to exchanging

268   the native gene with a currently existing homolog. However, functional divergence occurring

269   through time could result in ancestral sequences being so maladapted to the new host cell that

270   a functional organism is all but precluded (Copley 2003). This limitation does not apply only to

271   ancestral genes and it has been suggested that as the number of nodes connecting a protein

272   within its protein-protein interaction network increases, the probability that a protein could be

273   successfully replaced with a homolog will decrease even if there is a functional equivalence

274   between the endogenous gene and the homolog (Jain, et al. 1999). While a careful assessment

275   of candidate ancestral protein properties prior to integration is helpful, in most cases, studying

276   gene-triggered genomic perturbations experimentally through the integration of ancestral

277   genes offers a valuable and complementary alternative to existing methodologies that use

278   extant homologous proteins (Johnsen and Levin 2010; Acevedo-Rocha, et al. 2013; Pal, et al.

279   2014; Hobbs, et al. 2015).

280   How can we identify the specific historical constraints on replacement? We observe

281   that only the ancient EF-Tu representing an ancestor within the γ–proteobacteria, AnEF1 (0.7

282   bya), and the modern EF-Tu homologs from extant γ–proteobacteria, are viable. In contrast, the

283   last common ancestor of the α-, β-, and γ–proteobacteria, AnEF6 (1.3 bya) is non-viable.

284   Accordingly, we speculate that mutational substitutions in EF-Tu occurring between 1.6 bya

285   and 0.7 bya influenced the replaceability of *tuf* genes. These mutations may constrain *tuf*

286   replaceability by disturbing the EF-Tu's functional interaction with other cellular components,

287   ultimately impacting its participation in protein synthesis. Thus, extensive mutational

288   remodeling of interaction partners may be necessary in order to engineer even older ancient *tuf*

289   genes inside the bacteria.

Conclusions

291   We show that foreign *tuf* genes encoding EF-Tu proteins exhibit suboptimal functionality

292   and reduced fitness when introduced into another host. The sub-optimality of the foreign *tuf*

293   genes most likely results from disturbances in interactions directly important for protein

294   synthesis, but suboptimal EF-Tu protein levels and disturbance of other potentially important

295   interactions in the network of EF-Tu might also play a role. The observation that the only *tuf*

296   homologs that supported viability belong to γ-proteobacterial taxon, or an associated ancestral

297   node within the γ-proteobacteria, suggests that there is a relatively stringent "transferability

298   cutoff" i.e., a point in the phylogeny beyond which functional divergence is too great for

299   replacement. For EF-Tu protein this transferability zone is within the ancestral and modern γ-

300   proteobacterial taxon, unlike some ribosomal proteins where constraints on replaceability are

301   less stringent (Condon, et al. 1995; Lind, Tobin, et al. 2010).

12

302   Future efforts may involve identifying protein sites that interfere with organismal level

303   function, and epistatically inhibit an ancient protein's function in a descendant organism. Our

304   experiments suggest that a protein like EF-Tu, that is highly conserved, and involved in multiple

305   highly conserved interactions, is so highly optimized and fine-tuned in the host organism that it

306   is essentially irreplaceable with distantly related foreign genes. The degree to which epistatic

307   interactions constrain EF-Tu replaceability and functionality in the cell needs to be studied more

308   to deepen our understanding of the design principles of complex biological systems and to

309   allow us to introduce alterations in modern organisms by genetic engineering and gene

310   replacements.

311   Acknowledgments

321

322

Media and growth conditions

In general, liquid and solid media used were Luria Bertani (LB) media (10 g NaCl -5g in the case of low-salt LB-, 5 g yeast, 10 g triptone) and LA plates (LB with 1.5% of agar). For sacB counter-selections the LB and LA media used had no NaCl and LA media was also supplemented with sucrose 5 %. All incubations were made at 37°C (unless stated otherwise), and liquid cultures were shaken at 200rpm for aeration. The antibiotics used (Sigma-Aldrich, Sweden) had the following final concentrations: kanamycin 50 µg/mL and tetracyclin 15 µg/mL.

Bacterial strains

All strains used in this study are derived from sequenced wild-type *E. coli* K12 strain MG1655 (Hayashi et al, 2006), unless stated otherwise. A list of the strains used in these experiments is shown in Table 1.

Strain construction

The genetic marker [TP22-amilCP_opt-kan-sacB-T0], was inserted into the chromosome to replace either *tufA* or *tufB* by double-stranded DNA lambda-red recombineering (Datsenko & Wanner, 2000; Yu et al., 2000). The lambda-red genes were induced in each strain from the temperature-sensitive pSIM5-tet plasmid by incubation of an over-day culture (OD_{600}: 0.3) at 43°C for 15 minutes. After cooling for 10 minutes in ice, the cells were made electrocompetent by washing in ice-cold water three times. Electroporation of [TP22-amilCP_opt-kan-sacB-T0] PCR product was done with a Gene Pulser (Bio-Rad, USA) by mixing 50 µl of electrocompetent cells and 100 ng of the cassette, with settings 1.8 kV, 25µF and 200 Ω. Cells were recovered in 1mL of low-salt LB at 30 °C overnight with agitation, and after recovery 100 µl of the culture were spread in LA plates containing kanamycin for selection of recombinants.

14

348 Genetic markers were moved between strains by phage-mediated (P1 *virA*) transduction.

349 Lysates of the strains carrying the marker were made by mixing 1 mL of overnight culture

350 containing 5mM $CaCl_2$ with 100 µl of the P1 virA lysate previously made on E. coli MG1655.

351 The bacteria:phage mix was incubated for 10 minutes and then 4mL of soft-agar (LB media +

352 0.8% agar + 5mM $CaCl_2$) were added; this mixture was spread over an LA plate and incubated

353 overnight. To release the bacteriophages the soft-agar was mixed with 4 mL and vortexed. The

354 resultant slurry was centrifuged for 15 minutes at 5000 rpm and the supernatant was filtered

355 through a 0.2 µm filter. Markers were transduced into the desired recipient strain by mixing 100

356 µl of the lysate with 500 µl of an overnight culture containing 5 mM $CaCl_2$. After 10 minutes of

357 incubation, 100 µl of the mixture was spread onto a selective plate.

358

359 Insertions of alien *tuf* genes were made by double-stranded DNA lambda-red recombineering

360 to replace the previously inserted counter-selectable [TP22-amilCP_opt-kan-sacB-T0] marker.

361 To make clean deletions, markers were deleted by single stranded lambda-red recombineering

362 with counter selection for sucrose resistance (Ellis et al., 2001), following the same steps as

363 commented above for the double-stranded DNA lambda-red recombineering.

364

365 **PCR and oligonucleotides**

366 Polymerase Chain Reaction (PCR) was performed on a S1000™ Thermal Cycler (Bio-Rad,

367 USA). Oligonucleotides were designed with the software CLC Main Workbench 7 (CLC bio,

368 Denmark) using the genome of E. coli MG1655 as reference. For generation of the [TP22-

369 amilCP_opt-kan-sacB-T0] cassette, PCR was performed using Phusion® High-Fidelity PCR

370 Master Mix with HF buffer (New England Biolabs, USA) and with the following cycling

371 conditions: 98 °C for 30''; 30 cycles of 98 °C for 10'', 55°C for 30'', 72 °C for 4'; 72 °C for 7'.

372 For routine diagnostic PCR, Fermentas PCR Master Mix (Thermo Scientific, USA) was used

15

373   with the following cycling conditions: 95 ºC for 5'; 30 cycles of 95 ºC for 30'', annealing

374   temperature (T_A) for 30'', 72 ºC for elongation time (E_T)'; 72 ºC for 5'. T_A varied depending on the

375   pair of primers used and E_T was based on the length of the expected product (30'' per kb).

376   Oligonucleotides for construction of strains, PCR and sequencing are shown in Table S1.

377

378   **Preparation of genomic DNA and quantitative real-time PCR**

379   Genomic DNA prepared using the MasterPure™ DNA Purification Kit (Epicentre

380   Biotechnologies, USA) was used to run quantitative real-time PCR. 1 µL gDNA (diluted 1:10,

381   1:100, 1:1000 and 1:10000), 10 µL PerfeCTa SYBR Green FastMix (Quanta BioSciences), 0.6

382   µL of 10 µM forward and reverse primers, and ddH2O were added to a final reaction volume of

383   20 µL. The Eco Real-Time PCR Systems (Illumina) was used for running the PCR. Amplification

384   of the cysG and indT genes was used as standard. The oligonucleotides used as qRT-PCR

385   primers are listed in Table S1.

386

387   **Whole genome sequencing and analysis**

388   Genomic DNA was prepared using the MasterPure™ DNA Purification Kit (Epicentre

389   Biotechnologies, USA). To create libraries of paired-end fragments, Nextera XT Sample

390   Preparation Kit (Illumina, USA) was used following instructions from manufacturer. Sequencing

391   was performed on the Illumina MiSeq, generating 250 bp paired-end reads. Whole genome

392   sequencing data was analyzed using the software CLC Genomic Workbench (CLC bio,

393   Denmark)

394

395

16

Local DNA sequencing

Local sequencing of PCR amplified products was performed at Macrogen Europe sequencing

facilities (Amsterdam, The Netherlands), and data was analyzed using the software CLC Main

Workbench 7 (CLC bio, Denmark).

Growth rates and relative fitness measurements

Growth rate was measured by monitoring the rate of increase in optical density using a

Bioscreen C machine (Oy Growth Curves Ab Ltd, Findland), growing the cultures in a

Honeycomb microtiter plate. To set up an experiment, 0.5 µl of overnight culture from a single

colony was diluted into 1mL of LB medium. For each independent colony, 3 wells of the

microtiter plate were filled with 300 µl of each dilution, as technical replicates. Cultures were

grown for 18h at 37 ºC with continuous shaking and readings of OD_{600} were taken at 5 minutes

intervals. The doubling time (dt) of each strain during exponential growth was estimated over

an interval of 50 minutes beginning at OD_{600}=0.025. Relative fitness was determined in each

case by comparing the doubling time to the average of all the wild-type measurements.

Statistic analysis

All statistical analyses were performed using GraphPad Prism® v6.0c (GraphPad Software,

USA). The significance of differences between fitness costs was calculated using an unpaired

two-tailed t-test.

Proteomics

Cell lysis, digestion and labeling procedures. All cells were placed into Covaris® microTUBE-

15 (Woburn, MA) microtubes with Covaris® TPP buffer. Samples were lysed in an Covaris S220

Focused-ultrasonicator instrument with 125W power over 180s with 10% max peak power.

Lysed cells were digested via FASP digest according to the FASP Filter Aided Sample Prep

17

422   protocol for trypsin digestion, followed by HPLC purification. We used Promega® Sequencing

423   Grade Trypsin/LysC (V5073, Madison, WI) overnight at 38°C. Each sample was submitted for a

424   single LC-MS/MS experiment that was performed on a LTQ Orbitrap Elite (Thermo Fischer)

425   equipped with a Waters (Milford, MA) NanoAcquity HPLC pump or Orbitrap Lumos (Thermo

426   Fischer, San Jose, CA) equipped with EASYLC1000 (Thermo Fischer, San Jose, CA). Peptides

427   were separated onto a 100 μm inner diameter microcapillary trapping column packed first with

428   approximately 5 cm of C18 Reprosil resin (5 μm, 100 Å, Dr. Maisch GmbH, Germany) followed

429   by analytical column ~20 cm of Reprosil resin (1.8 μm, 200 Å, Dr. Maisch GmbH, Germany).

430   Separation was achieved by applying a gradient from 5–27% ACN in 0.1% formic acid over 90

431   min at 200 nl min−1. Electrospray ionization was enabled by applying a voltage of 1.8 kV using

432   a home-made electrode junction at the end of the microcapillary column and sprayed from

433   fused silica pico tips (New Objective, MA). The LTQ Orbitrap Elite/Lumos was operated in data-

434   dependent mode for the mass spectrometry methods. The mass spectrometry survey scan

435   was performed in the Orbitrap in the range of 395–1,800 m/z at a resolution of 6×10^4,

436   followed by the selection of the twenty most intense ions (TOP20) for CID-MS2 fragmentation

437   in the Ion trap using a precursor isolation width window of 2 m/z, AGC setting of 10,000, and a

438   maximum ion accumulation of 200 ms. Singly charged ion species were not subjected to CID

439   fragmentation. Normalized collision energy was set to 35 V and an activation time of 10 ms.

440   Ions in a 10 ppm m/z window around ions selected for MS2 were excluded from further

441   selection for fragmentation for 60 s. The same TOP20 ions were subjected to HCD MS2 event

442   in Orbitrap part of the instrument. The fragment ion isolation width was set to 0.7 m/z, AGC

443   was set to 50,000, the maximum ion time was 200 ms, normalized collision energy was set to

444   27V and an activation time of 1 ms for each HCD MS2 scan.

445

446

447 **Mass spectrometry analysis**

448 Raw data were submitted for analysis in Proteome Discoverer 2.1.0.81 (Thermo Scientific)

449 software. Assignment of MS/MS spectra was performed using the Sequest HT algorithm by

450 searching the data against a protein sequence database including all entries from the users

451 database, and our *E. coli* K12 database as well as other known contaminants such as human

452 keratins and common lab contaminants. Sequest HT searches were performed using a 20 ppm

453 precursor ion tolerance and requiring each peptides N-/C termini to adhere with Trypsin

454 protease specificity, while allowing up to two missed cleavages. 6-plex TMT tags on peptide N

455 termini and lysine residues (+229.162932 Da) was set as static modifications while methionine

456 oxidation (+15.99492 Da) was set as variable modification. A MS2 spectra assignment false

457 discovery rate (FDR) of 1% on protein level was achieved by applying the target-decoy

458 database search. Filtering was performed using a Percolator 64 bit. For quantification, a 0.02

459 m/z window centered on the theoretical m/z value of each the six reporter ions and the

460 intensity of the signal closest to the theoretical m/z value was recorded. Reporter ion intensities

461 were exported in result file of Proteome Discoverer 2.1 search engine as an excel tables.

462

463 **Evolutionary Divergence**

464 Two independent approaches were utilized in order to estimate the evolutionary distance

465 between sequences, both leading the same evolutionary distance estimate output. i) MEGA

466 software: The analyses involved 18 amino acid sequences. Analyses were conducted using the

467 Poisson correction model (Zuckerkandl 1965). All positions containing gaps and missing data

468 were eliminated. There were a total of 385 positions in the final dataset. Evolutionary analyses

469 were conducted in MEGA7 (Kumar, et al. 2016). ii) The branchlenght distances were also

470 calculated via a custom script that used ETE Software (Huerta-Cepas, et al. 2010). The custom

471 python ETE v. 3 python library script is provided under supporting information.

472 **Computational methods**

473 Computational analyses include i) the pairwise and multiple sequence alignments, ii) protein

474 structure analysis, and iii) protein interaction network analysis. Clustal Omega was used to

475 perform multiple sequence alignment with default parameters. Pairwise sequence alignments of

476 the wild type EF-Tu against each ancient and modern EF-Tu proteins were performed by

477 EMBOSS Needle of Clustal Omega with default parameters which uses the Needleman-

478 Wunsch alignment algorithm.

479

480 The protein interactome of *E. coli* was retrieved from the HitPredict database and the network

481 analysis was performed with the network package of Python. The sub-network of EF-Tu

482 interaction partners was visualized with Cytoscape.

483

484

References

Abdulkarim F, Hughes D. 1996. Homologous recombination between the tuf genes of Salmonella typhimurium. *J Mol Biol* 260:506-522.

Abdulkarim F, Tuohy TM, Buckingham RH, Hughes D. 1991. Missense substitutions lethal to essential functions of EF-Tu. *Biochimie* 73:1457-1464.

Acevedo-Rocha CG, Fang G, Schmidt M, Ussery DW, Danchin A. 2013. From essential to persistent genes: a functional approach to constructing synthetic life. *Trends Genet* 29:273-279.

Anderson P, Roth J. 1981. Spontaneous tandem genetic duplications in Salmonella typhimurium arise by unequal recombination between rRNA (rrn) cistrons. *Proc Natl Acad Sci U S A* 78:3113-3117.

Baldauf SL, Palmer JD, Doolittle WF. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc Natl Acad Sci U S A* 93:7749-7754.

Battistuzzi FU, Feijao A, Hedges SB. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* 4:44.

Benner SA. 1995. Reconstructing Ancient Forms of Life. *Journal of cellular biochemistry* 59, Suppl. 19A:200-200.

Benner SA, Sassi SO, Gaucher EA. 2007. Molecular paleoscience: systems biology from the past. *Adv Enzymol Relat Areas Mol Biol* 75:1-132, xi.

Bershtein S, Serohijos AW, Bhattacharyya S, Manhart M, Choi JM, Mu W, Zhou J, Shakhnovich EI. 2015. Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria. *PLoS Genet* 11:e1005612.

Blank D, Wolf L, Ackermann M, Silander OK. 2014. The predictability of molecular evolution during functional innovation. *Proc Natl Acad Sci U S A* 111:3044-3049.

Brandis G, Bergman JM, Hughes D. 2016. Autoregulation of the tufB operon in Salmonella. *Mol Microbiol* 100:1004-1016.

Brandis G, Hughes D. 2016. The Selective Advantage of Synonymous Codon Usage Bias in Salmonella. *PLoS genetics* 12:e1005926.

515   Brochier C, Philippe H, Moreira D. 2000. The evolutionary history of ribosomal protein RpS14:
516   horizontal gene transfer at the heart of the ribosome. *Trends Genet* 16:529-533.

517   Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic
518   reconstruction of a highly resolved tree of life. *Science* 311:1283-1287.

519   Cohen O, Gophna U, Pupko T. 2011. The complexity hypothesis revisited: connectivity rather
520   than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol* 28:1481-1489.

521   Condon C, Squires C, Squires CL. 1995. Control of rRNA transcription in Escherichia coli.
522   *Microbiol Rev* 59:623-645.

523   Copley SD. 2003. Enzymes with extra talents: moonlighting functions and catalytic promiscuity.
524   *Curr Opin Chem Biol* 7:265-272.

525   Coulomb S, Bauer M, Bernard D, Marsolier-Kergoat MC. 2005. Gene essentiality and the
526   topology of protein interaction networks. *Proc Biol Sci* 272:1721-1725.

527   Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed
528   proteins evolve slowly. *Proc Natl Acad Sci U S A* 102:14338-14343.

529   Durfee T, Hansen AM, Zhi H, Blattner FR, Jin DJ. 2008. Transcription profiling of the stringent
530   response in Escherichia coli. *J Bacteriol* 190:1084-1096.

531   Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the
532   protein interaction network. *Science* 296:750-752.

533   Gaucher EA, Govindarajan S, Ganesh OK. 2008. Palaeotemperature trend for Precambrian life
534   inferred from resurrected proteins. *Nature* 451:704-707.

535   Hahn MW, Conant GC, Wagner A. 2004. Molecular evolution in large genetic networks: does
536   connectivity equal constraint? *J Mol Evol* 58:203-211.

537   Hammarlöf DL, Hughes D. 2008. Mutants of the RNA-processing enzyme RNase E reverse the
538   extreme slow-growth phenotype caused by a mutant translation factor EF-Tu. *Mol Microbiol*
539   70:1194-1209.

540   Hobbs JK, Prentice EJ, Groussin M, Arcus VL. 2015. Reconstructed Ancestral Enzymes
541   Impose a Fitness Cost upon Modern Bacteria Despite Exhibiting Favourable Biochemical
542   Properties. *J Mol Evol* 81:110-120.

543 Hsiao C, Mohan S, Kalahar BK, Williams LD. 2009. Peeling the onion: ribosomes are ancient
544 molecular fossils. *Mol Biol Evol* 26:2415-2425.

545 Huerta-Cepas J, Dopazo J, Gabaldon T. 2010. ETE: a python Environment for Tree Exploration.
546 *BMC Bioinformatics* 11:24.

547 Ito K, Uno M, Nakamura Y. 1998. Single amino acid substitution in prokaryote polypeptide
548 release factor 2 permits it to terminate translation at all three stop codons. *Proc Natl Acad Sci*
549 *U S A* 95:8165-8169.

550 Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity
551 hypothesis. *Proc Natl Acad Sci U S A* 96:3801-3806.

552 Jain R, Rivera MC, Moore JE, Lake JA. 2003. Horizontal gene transfer accelerates genome
553 innovation and evolution. *Mol Biol Evol* 20:1598-1602.

554 Johnsen PJ, Levin BR. 2010. Adjusting to alien genes. *Mol Microbiol* 75:1061-1063.

555 Kacar B. 2016. Rolling the Dice Twice: Evolving Reconstructed Ancient Proteins in Extant
556 Organisms. In: Ramsey G, Pence CH, editors. Chance in Evolution from Darwin to
557 Contemporary Biology: The University of Chicago Press. p. 264-276.

558 Kacar B, Gaucher EA. 2013. Experimental evolution of protein-protein interaction networks.
559 *Biochem J* 453:311-319.

560 Kacar B, Gaucher EA. 2012. Towards the recapitulation of ancient history in the laboratory:
561 combining synthetic biology with experimental evolution. . In C. A, M. BD, C. O, T. PR editors.
562 Artificial Life 13: Proceedings of the Thirteenth International Conference on the Simulation and
563 Synthesis of Living Systems; Cambridge, MA: MIT Press. p. 11–18.

564 Kacar B, Xueliang G, Sanyal S, Gaucher EA. 2016. Experimental evolution of modern
565 Escherichia coli harboring an ancient Elongation Factor gene. *Biorxiv*:
566 http://dx.doi.org/10.1101/040626.

567 Karcagi I, Draskovits G, Umenhoffer K, Fekete G, Kovacs K, Mehi O, Baliko G, Szappanos B,
568 Gyorfy Z, Feher T, et al. 2016. Indispensability of Horizontally Transferred Genes and Its Impact
569 on Bacterial Genome Streamlining. *Molecular biology and evolution* 33:1257-1269.

570 Kavaliauskas D, Nissen P, Knudsen CR. 2012. The busiest of all ribosomal assistants:
571 elongation factor Tu. *Biochemistry* 51:2642-2651.

572   Kitahara K, Miyazaki K. 2013. Revisiting bacterial phylogeny: Natural and experimental
573   evidence for horizontal gene transfer of 16S rRNA. *Mob Genet Elements* 3:e24210.

574   Knoppel A, Nasvall J, Andersson DI. 2016. Compensating the Fitness Costs of Synonymous
575   Mutations. *Molecular biology and evolution* 33:1461-1477.

576   Kothe U, Wieden HJ, Mohr D, Rodnina MV. 2004. Interaction of helix D of elongation factor Tu
577   with helices 4 and 5 of protein L7/12 on the ribosome. *J Mol Biol* 336:1011-1021.

578   Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
579   Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870-1874.

580   Lathe WC, 3rd, Bork P. 2001. Evolution of tuf genes: ancient duplication, differential loss and
581   gene conversion. *FEBS Lett* 502:113-116.

582   Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene
583   expression levels are coupled in Drosophila and are independently associated with mRNA
584   abundance, protein length, and number of protein-protein interactions. *Mol Biol Evol* 22:1345-
585   1354.

586   Lind PA, Andersson DI. 2013. Fitness costs of synonymous mutations in the rpsT gene can be
587   compensated by restoring mRNA base pairing. *PloS one* 8:e63373.

588   Lind PA, Berg OG, Andersson DI. 2010. Mutational robustness of ribosomal protein genes.
589   *Science* 330:825-827.

590   Lind PA, Tobin C, Berg OG, Kurland CG, Andersson DI. 2010. Compensatory gene
591   amplification restores fitness after inter-species gene replacements. *Mol Microbiol* 75:1078-
592   1089.

593   Miller DL, Weissbach H. 1977. Factors involved in the transfer of aminoacyl-tRNA to the
594   ribosome. In: Weissbach H, Petska , S., editor. Molecular Mechanisms of Protein Biosynthesis.
595   New York. p. 323-373.

596   Nilsson J, Nissen P. 2005. Elongation factors on the ribosome. *Curr Opin Struct Biol* 15:349-
597   354.

598   Pal C, Papp B, Posfai G. 2014. The dawn of evolutionary genome engineering. *Nat Rev Genet*
599   15:504-512.

24

600   Patil A, Nakai K, Nakamura H. 2011. HitPredict: a database of quality assessed protein-protein
601   interactions in nine species. *Nucleic Acids Res* 39:D744-749.

602   Pauling L, Zuckerkandl E. 1963. Chemical Paleogenetics Molecular Restoration Studies of
603   Extinct Forms of Life. *Acta Chemica Scandinavica* 17:S9-S16.

604   Shen P, Huang HV. 1986. Homologous recombination in Escherichia coli: dependence on
605   substrate length and homology. *Genetics* 112:441-457.

606   Smith MW, Feng DF, Doolittle RF. 1992. Evolution by acquisition: the case for horizontal gene
607   transfers. *Trends Biochem Sci* 17:489-493.

608   Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. 2007. Genome-wide
609   experimental determination of barriers to horizontal gene transfer. *Science* 318:1449-1452.

610   Thirup SS, Van LB, Nielsen TK, Knudsen CR. 2015. Structural outline of the detailed
611   mechanism for elongation factor Ts-mediated guanine nucleotide exchange on elongation
612   factor Tu. *J Struct Biol* 191:10-21.

613   Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules.
614   *Nat Rev Genet* 5:366-375.

615   Tubulekas I, Hughes D. 1993a. Growth and translation elongation rate are sensitive to the
616   concentration of EF-Tu. *Mol Microbiol* 8:761-770.

617   Tubulekas I, Hughes D. 1993b. A single amino acid substitution in elongation factor Tu disrupts
618   interaction between the ternary complex and the ribosome. *J Bacteriol* 175:240-250.

619   Wellner A, Lurie MN, Gophna U. 2007. Complexity, connectivity, and duplicability as barriers to
620   lateral gene transfer. *Genome Biol* 8:R156.

621   Yu H, Chan YL, Wool IG. 2009. The identification of the determinants of the cyclic, sequential
622   binding of elongation factors tu and g to the ribosome. *J Mol Biol* 386:802-813.

623   Zhou Y, Asahara H, Gaucher EA, Chong S. 2012. Reconstitution of translation from Thermus
624   thermophilus reveals a minimal set of components sufficient for protein synthesis at high
625   temperatures and functional conservation of modern and ancient translation components.
626   *Nucleic Acids Res* 40:7932-7945.

627   Zotenko E, Mestre J, O'Leary DP, Przytycka TM. 2008. Why do hubs in the yeast protein
628   interaction network tend to be essential: reexamining the connection between the network
629   topology and essentiality. *PLoS Comput Biol* 4:e1000140.

630   Zuckerkandl EP, L. . 1965. Evolutionary divergence and convergence in proteins. In: Bryson
631   VV, H.J., editor. Evolving genes and proteins. New York: Academic Press. p. 97-166.
632

633

634

Figures
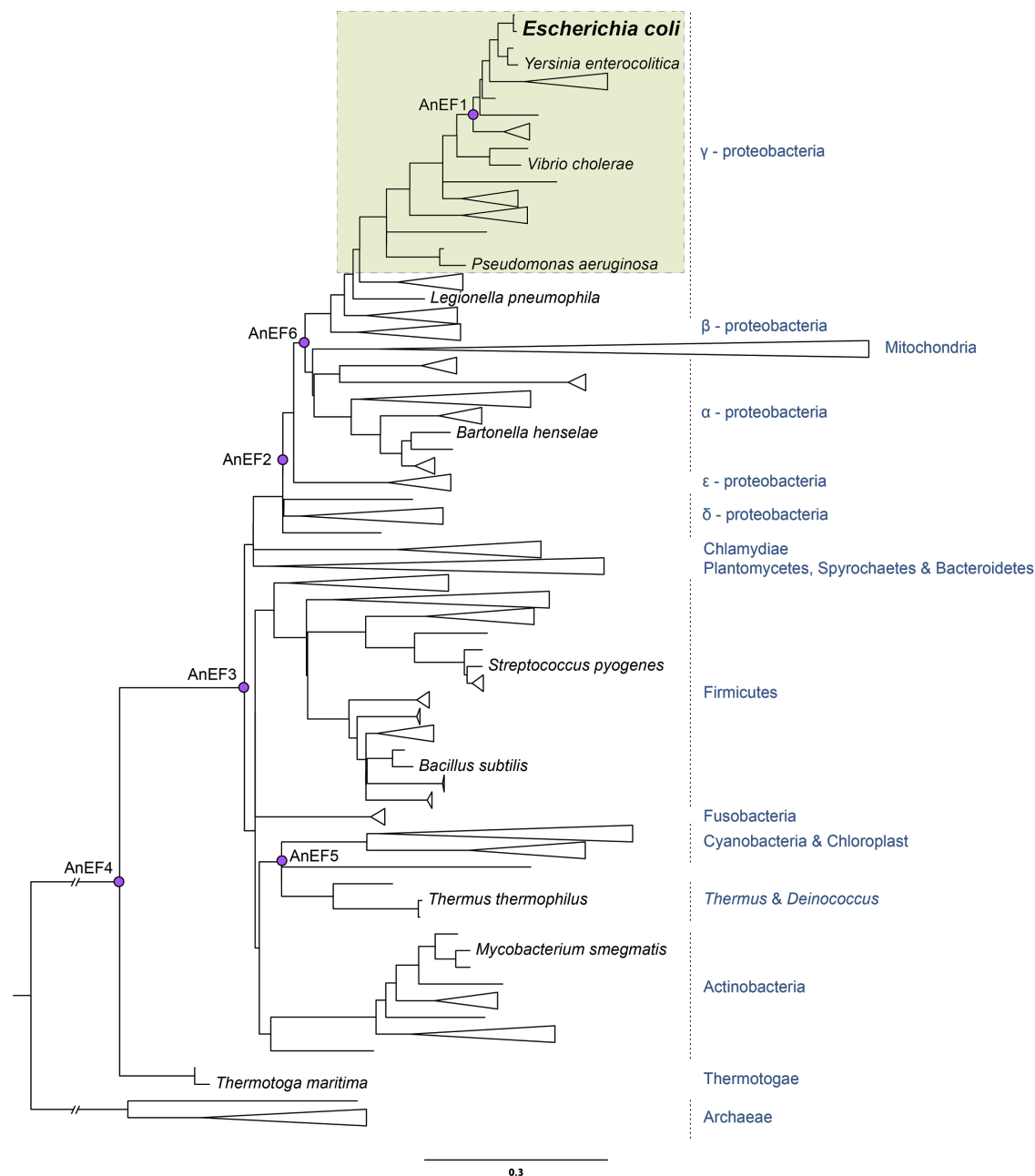
Figure 1. Phylogenetic tree indicating the node and taxa of the ancestral and modern *tuf* (EF-Tu) homologs. Pink circles represent the ancestral EF-Tu nodes. *E. coli* was genetically engineered to carry ancestral or modern homologs of *tuf*, encoding translation elongation factor EF-Tu, replacing the native *E. coli tufA* gene. A shaded box indicates the area of viability of EF-Tu gene exchange. The scale bar expresses units of amino acid substitutions per site. The tree is adapted from (Gaucher et al. 2008, Battistuzzi et al. 2004).

646

647
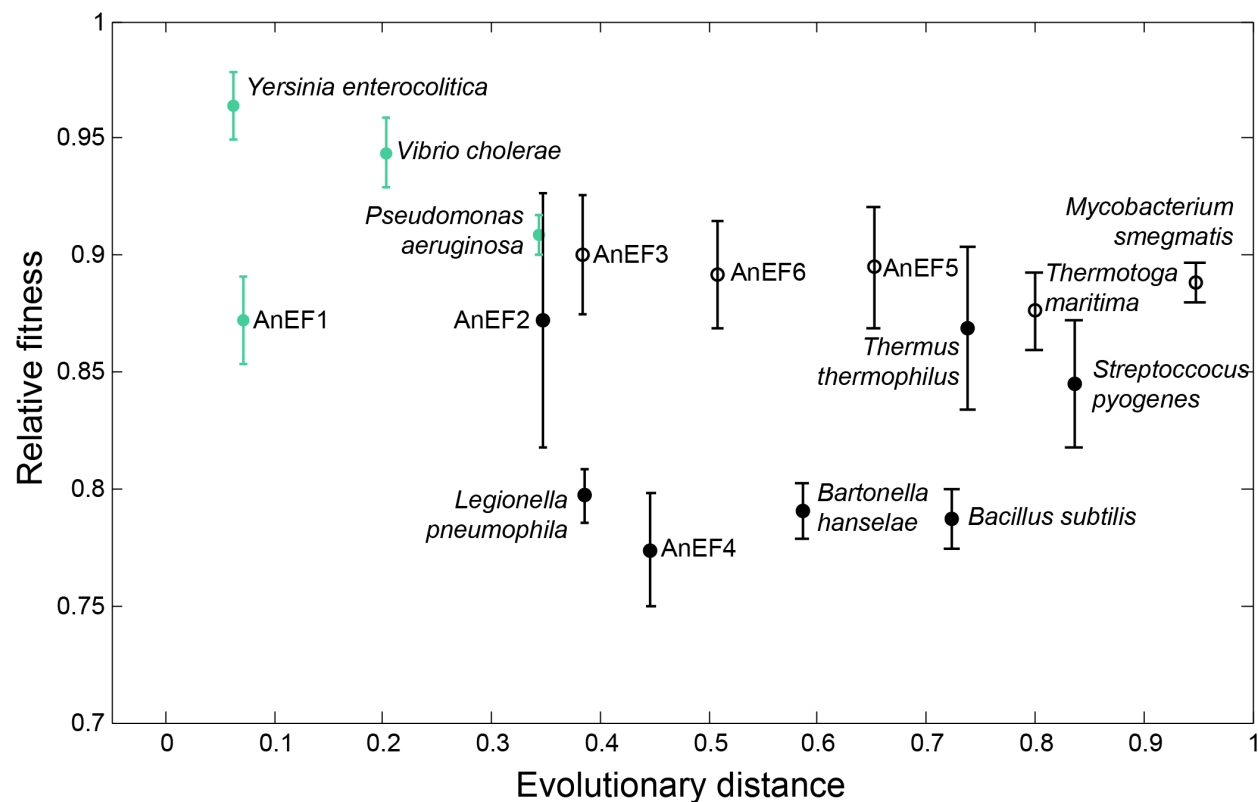
Figure 2. Correlation between relative fitness and evolutionary distance for bacterial strains carrying a foreign *tuf* gene. Relative fitness is shown as a function of evolutionary divergence (see Materials and Methods) of EF-Tu homologs from *E. coli* (1 indicates most different from *E. coli*). Each *E. coli* strain carried a foreign *tuf* gene at the *tufA* location and an intact native *tufB* gene. Fitness was measured as exponential growth rate, relative to *E. coli* wild-type carrying *tufA* and *tufB*. Species names and ancestor notations (AnEF, ancestral EF-Tu) refer to the source of the foreign *tuf* gene sequence in each strain. Strains shown in green carry a foreign *tuf* gene that can support viability even when *E. coli tufB* gene has been deleted. Strains shown in black carry foreign *tuf* genes that do not support viability in the absence of the *E. coli tufB* gene. Empty symbols represent strains where *tufB* region was amplified.

658

659

660
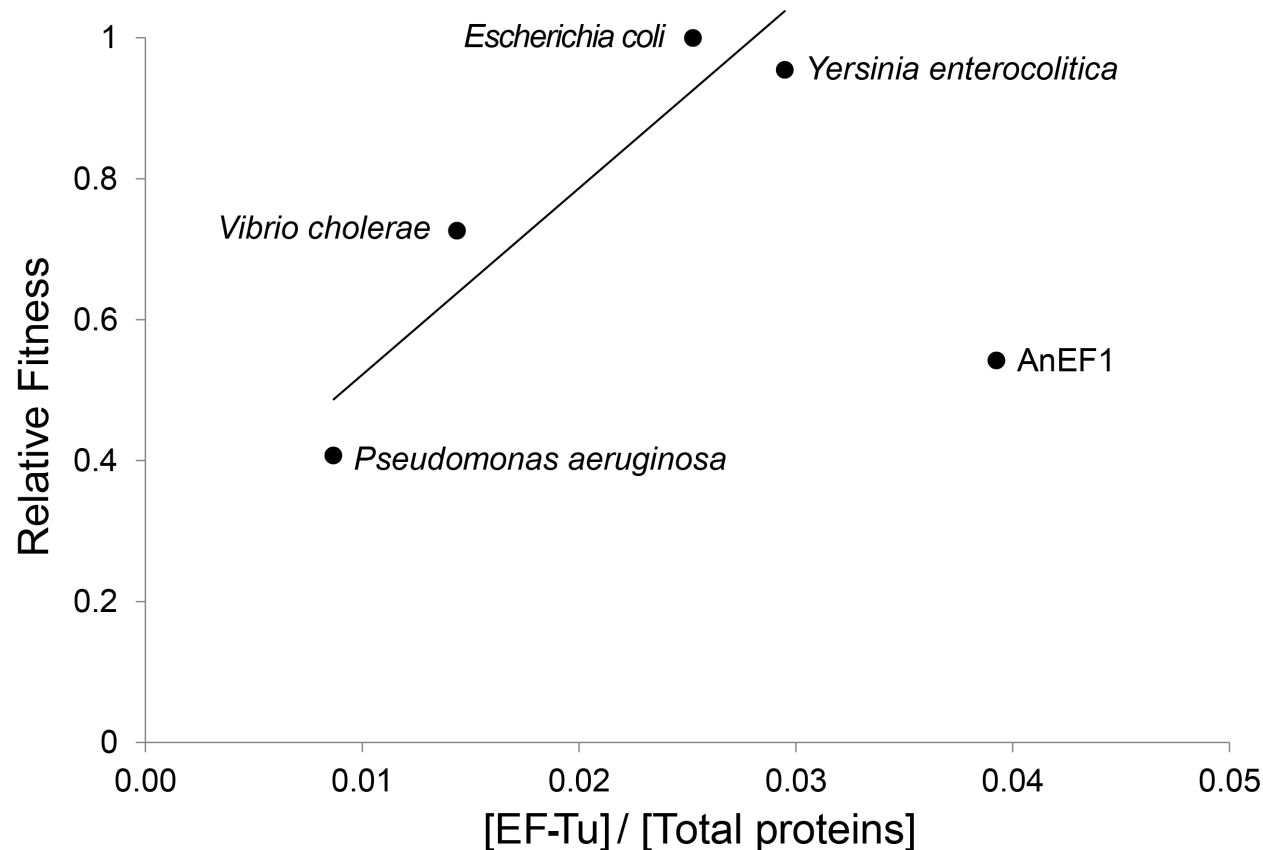
Figure 3. Relative fitness of strains carrying a single *tuf* gene, at the *tufA* locus, as a function of EF-Tu protein produced. EF-Tu concentration was normalized to the total protein concentration. Species names indicate the species origin of the only *tuf* gene present. (R=0.938).

661

662

663

664

665

666

667

668

669

670
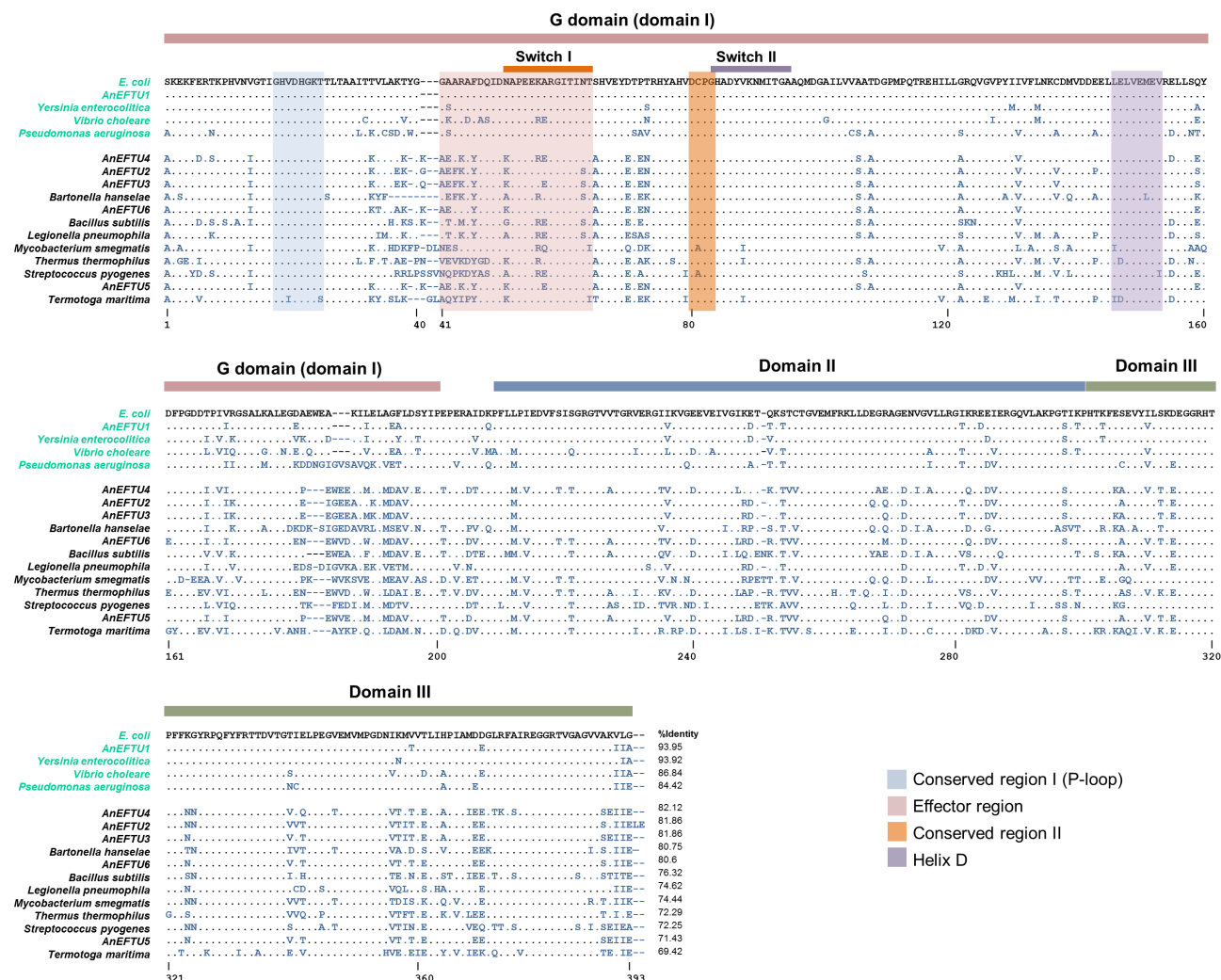
Figure 4. Alignment of EF-Tu protein sequences encoded by foreign *tuf* variants

EF-Tu sequences were aligned using Clustal W. Sequences labeled in green indicate that the foreign *tuf* gene supports viability as the only *tuf* gene in the genome. Sequences labeled in black indicate that the strain requires *E. coli tufB* for viability.