Experimental evolution of modern *Escherichia coli* harboring an ancient Elongation Factor gene

Betül Kaçar[1,2,3, #a, *], Lily I. Tran[2], Xueliang Ge[4], Suparna Sanyal[4] and Eric A. Gaucher[2, 5]

1. NASA Astrobiology Institute, Mountain View, CA
2. School of Biology, Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA 30332
3. Blue Marble Space Institute of Science, Seattle, WA, USA
4. Department of Cell and Molecular Biology, Uppsala University, Box-596, BMC, 75124, Uppsala, Sweden
5. Petit H. Parker Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332

[#a] Current Address: Harvard University, Department of Organismic and Evolutionary Biology, 26 Oxford Street, Cambridge, MA 02138

* Corresponding author

E-mail: kacar@fas.harvard.edu (BK)

Abstract

Species achieve evolutionary innovations through two major genetic mechanisms, namely regulatory- and structural-level mutations. The ability of populations to evolve involves a balance between selection, genetic drift, epistasis, biochemical and biophysical requirements, thermodynamic properties and other factors. This adaptive diversity begs the question as to whether a restricted pathway governs adaptations or whether multiple pathways are possible to achieve an adaptive response. By combining a unique set of tools drawn from synthetic biology, evolutionary biology and genomics, we experimentally evolved and then characterized the adaptive properties of a modern *E. coli* strain containing a 700 million-year-old reconstructed ancestral Elongation Factor Tu (EF-Tu) gene inserted into its genome for the first time. We then tracked the evolutionary steps taken by the ancient-modern hybrid microorganism through laboratory evolution by monitoring genomic mutations. This study reveals that lineages respond to the ancient gene by increasing the expression levels of the maladapted protein, rather than through direct accumulation of mutations in the open reading frame. In particular, these findings show that the general strategy for the bacteria to adapt to the ancient protein is to accumulate mutations in the *cis*-regulatory region; gene-coding mutations appear to preclude rapid adaptation upon integration of the ancient gene for our system.

Author Summary

Understanding the historical forces that have shaped the evolution of past organisms over time mainly relies on analyzing the behavior of organisms that exist today. This reconciliation requires an evolutionary framework that includes explicit functional links between genomes, natural selection, molecular innovation, phenotypic diversity and adaptation; yet, creating a framework that synthesizes all of these components remains a challenge. Here we make a novel attempt at such a synthesis by combining synthetic biology with natural selection to explore the historical constraints at work in evolutionary processes. In order to study historical pathways and the mechanisms of protein evolution in a complex cellular environment, we directly engineered a synthetic gene representing a 700 million-year-old ancestor of the contemporary elongation factor protein inside a modern *E. coli* strain. We then traced the evolutionary steps of the microorganism harboring this ancient gene by subjecting it to laboratory evolution, directly monitoring any resulting changes within the integrated ancient gene and the rest of the host genome through whole-genome sequencing. Our results demonstrate that an ancient gene can interact with modern cellular machinery, albeit with a cost of decreased fitness, and that lineages respond to the ancient gene by increasing the transcription levels of the maladapted protein. Further development of ancient-modern hybrid model systems has the potential to provide information about fundamental evolutionary processes at work in modern microbes.

Introduction

In the context of Darwinian evolutionary theory, species that persist through time must retain a mechanism that provides functional adaptation in a changing environment. Two major mechanisms orchestrate functional innovation at the genetic level. In the first, mutations accumulating in the non-coding regions influence gene expression, leading to phenotypic change (1-5). In the second, mutational changes in gene coding regions generate structural alteration of proteins and thus lead to functional diversity (6, 7). Whether one mechanism is favored over the other is the subject of ongoing debate in the field of evolutionary biology (8-10). The primary argument against a substantive role for coding mutation-mediated evolution is the assumption that mutations in protein-coding genes have extensive detrimental pleiotropic and fitness effects (11, 12). While there is no doubt that mutations altering protein expression level play an important role in shaping phenotypic diversity, failure to recognize the role of protein level mutations as a source of biological innovation may hamper a comprehensive understanding of phenotypic variation (13-18).

In this study, we describe a novel laboratory system that allows us to experimentally evolve *Escherichia coli* (*E. coli*) bacteria in a manner that appoints regulatory vs. structural mutations for a reconstructed 700 million-year-old Elongation Factor Tu protein (EF-Tu) (S1 Fig). Known as one of the most sequence-conserved proteins in life, EF-Tu constitutes an essential component of the bacterial translation machinery by delivering aminoacylated-tRNAs to the A-site of the ribosome (19). Our experimental system exploits the unique scenario in which *E. coli* bacteria have a paralogous copy of the EF-Tu gene *tufA*, in the form of *tufB*, that frequently recombines

89 with the original copy (20). Each of the EF-Tu genes have their own specific expression
90 machinery, while EF-Tu produced through *tufB* accounts for one-third of the cellular
91 EF-Tu as that produced by the *tufA* gene in bacteria. Their respective encoded proteins
92 differ from one another by a single amino acid in the C-terminus (21-23). The EF-Tu
93 produced by these two genes form 5-10% of the total protein produced in *E. coli*,
94 signifying EF-Tu is a highly abundant protein in the bacterial cell (24). As such, EF-Tu
95 promoters are routinely used to drive recombinant protein expression since they are
96 constitutively active (25).
97
98 In our system, we replaced the *tufB* copy of a laboratory strain of *E. coli* with an
99 ancient EF-Tu variant under the control of the endogenous *tufB* promoter, and we
100 completely removed the *tufA* gene from the bacterial genome (S1 Fig). We
101 hypothesized that a reconstructed ancient protein will be maladapted in a modern
102 genome and that there were two adaptive responses to laboratory evolution. One
103 possible adaptive route would accumulate mutations that alter the encoded protein's
104 structure and function in a manner that recapitulates the historical pathway by which
105 the modern gene evolved from its ancient ancestor. Alternatively, the second route may
106 accumulate mutations in the "weaker" promoter region of *tufB* where the ancient gene
107 is engineered, thus altering the expression of the ancient gene in the modern cell. We
108 further investigated whether *in vivo* analyses into the functionality of ancestral
109 components can be used to discern effects arising from the substituted gene when
110 screened from adaptive responses taken by the host cell to the sub-adapted genetic
111 component (26).
112

113 The engineered EF-Tu represents an ancestral γ-proteobacterium that is inferred
114 to be approximately 700 million-years-old, and has 21 (out of 392) amino acid
115 differences with the modern *E. coli* EF-Tu protein (S2 and S3 Figs). An *in vitro* peptide
116 synthesis assay demonstrates that resurrected EF-Tu can participate in a translation
117 system in which all other necessary components for translation besides EF-Tu are
118 provided by modern *E.* coli, albeit with a lower efficiency than the wild-type (27). The
119 temperature profile of the ancient EF-Tu protein is 39.1 °C, slightly higher than the
120 temperature optimum of the endogenous *E. coli* EF-Tu at 37 °C (28). To examine the
121 adaptation between the ancestral EF-Tu protein and contemporary bacteria, we
122 engineered the ancient gene into the modern bacterial genome. We then evolved the
123 hybrid bacterial populations for 2,000 generations in the laboratory in multiple
124 independent parallel lines, and examined the adaptive response through fitness, whole-
125 genome sequencing, proteomics and biochemical assays.

126 # Results

127 **Rapid fitness improvement of the ancient-modern hybrid bacterial**
128 **populations**

129 Engineering *E. coli* harboring a reconstructed ancestral EF-Tu gene results in
130 two-fold increase in bacterial growth doubling-time and 20% decrease in bacterial
131 fitness relative to the founding strain *E. coli* REL606 (Fig 1, S4 Fig). Adaptation to
132 selective laboratory media by the ancient-modern hybrid populations is shown by the
133 20% increase in fitness values and a rapid drop in growth doubling-time compared to

134 that of native *E. coli* bacteria by 500 generations. The fitness continues to increase by
135 ~5% every 500 generations until stabilizing by generation 1500 (Fig 1, S4 Fig).
136
137 **Fig 1. Overall experimental scheme and the relative change in fitness of the**
138 **ancient-modern hybrid bacteria relative to the modern parent strain.**
139 (A) Engineering a modern *E. coli* strain with an ancient EF-Tu gene followed by
140 laboratory evolution of initially identical six ancient-modern hybrid populations in
141 parallel. (B) Replacement of the endogenous EF-Tu gene with the reconstructed ancient
142 EF-Tu allele significantly reduces the fitness of the ancient-modern hybrid relative to the
143 original strain (black circle), indicated by the black dashed arrow. Hybrid population
144 mean fitness rapidly improved during experimental evolution in minimal glucose
145 medium (blue circles). Error bars show 95 % confidence interval among six replicate
146 populations.
147

148 **Whole Genome Sequencing reveals adaptation of EF-Tu's promoter**

149 Whole genome sequencing of evolved modern (as a control) and ancient-
150 modern hybrid *E. coli* populations was performed periodically throughout the
151 evolutionary course and revealed that the accumulation of parallel mutations in a variety
152 of genes across the populations (Table 1). In particular, five out of six independent
153 lineages have accumulated mutations in *thrT*/*tufB* region, corresponding to a promoter
154 region for *tufA* gene (29); however no mutation on the ancestral EF-Tu gene in any of
155 the six lineages is observed. Shown in Fig 2 are the frequencies of the allelic mutations

located in ancient EF-Tu gene's promoter region, demonstrating that a majority of the mutations accumulated early during the course of the experiment.

Table 1. Parallel mutations for six initially-identical and independently-evolved populations harboring an ancient EF-Tu.

Genes that accumulated synonymous mutations in at least three out of six laboratory evolved ancient-hybrid populations, and that achieved a frequency of at least 20% between generations 500 and 2000 in the population. *thrT*/*tufB* is the intergenic region between the ancient EF-Tu gene and the *thrT* gene. The bottom three listed genes were specific to the single lineage in which no mutation in the *thrT*/*tufB* region was detected (referred as Rip6). Genes highlighted with an asterisk accumulated mutations in at least one population containing the wild-type EF-Tu gene (control experiment).

Gene	Function	Number of lineages that accumulated a mutation over 20% frequency			
		Generation			
		500	1000	1500	2000
mreB*	cell wall structural complex MreBCD, actin-like component MreB	2	3	3	3
mrdB*	cell wall shape-determining protein	1	1	3	1
thrT/tuf	tRNA-Thr/protein chain elongation factor EF-Tu	3	3	5	5
fadA*	acetyl-CoA acetyltransferase	2	3	3	3
ftsZ	cell division protein FtsZ	1	1	2	4
iclR	DNA-binding transcriptional repressor	-	1	4	3
accC	acetyl-CoA carboxylase	-	1	5	5
pykF*	pyruvate kinase	-	1	3	5
nusA	transcription elongation factor NusA	1	1	1	1
infB	translation initiation factor IF-2	1	1	1	-
hupA	HU, DNA-binding transcriptional regulator, alpha subunit	1	1	1	1

183 ***In vivo* and *in vitro* experiments indicate mutations upregulate ancient EF-**
184 **Tu expression**

185 Whole-cell quantitation of EF-Tu protein levels through shotgun proteomics
186 shows that upon insertion, ancestral EF-Tu produced by *E. coli* decreases by 70%
187 compared to the native strain and increases in most of the lineages upon accumulation
188 of mutations in the *thrT/tufB* region during laboratory evolution (Fig 2, S5 Fig). In order
189 to assess the impact of mutations on expression levels, we individually cloned the
190 evolved promoters into a plasmid expressing a reporter gene, whose fluorescent
191 strength is directly linked to the strength of the evolved promoter region. Results
192 demonstrate that mutations in the promoter region amplify the protein expression levels
193 to varying degrees, ranging from 1.5-20 fold higher (Fig 2). Both *in vivo* and *in vitro* data
194 assign an upregulatory role for these promoter-level mutations. However, the
195 magnitude detected *in vitro* does not directly correlate with the values observed
196 through LC/MS-MS measurements via whole-cell proteomics that suggest an
197 enhancement of 1-2 fold for the mutations. This is expected, however, since the
198 promoter assay is enzymatically-based and thus amplifies signal. Mutations that have
199 lower impact on ancient EF-Tu protein expression relative to a 19 basepair duplication
200 in the *thrT/tufB* region all appear later in the evolutionary experiment. This suggests
201 that, to a first approximation, each population exhibits parallel patterns in response to
202 the ancient EF-Tu (i.e., upregulation of the EF-Tu gene) but that each population may
203 exhibit other unique mechanisms in response to the ancestral gene.
204
205

206 **Fig 2. Analysis of the mutations accumulated in the *cis*-regulatory region *thrT*/*tufB***
207 (A) The *thrT*/*tufB* promoter region in which five of six evolved hybrid populations were
208 found to have accumulated mutations. (B) The allelic frequency of the intergenic
209 mutations per generation per population during laboratory evolution. (C) (Top) Effect of
210 mutations on the expression of a reporter gene with the EF-Tu promoter region
211 measured using luciferase. (Bottom) Relative abundance of ancient EF-Tu protein
212 among evolved hybrid strains using the peak area quantification from MS proteomics
213 data (30, 31). Error bars obtained using Anova/t-test.
214
215
216 Intriguingly, none of the modern lineages evolved in parallel to the ancient-
217 modern populations exhibited mutations in the *tufB* promoter region *thr*/*tufB* or on the
218 *tufB* gene. When over-expressed from a plasmid, ancient EF-Tu protein increases the
219 fitness of the unevolved hybrid strain by 10% relative to wild-type *E. coli*. The 20%
220 decrease of the organismal fitness upon insertion of the ancestral EF-Tu, and the 10%
221 increase of the unevolved ancestral-modern hybrid strain upon overexpression of the
222 ancient EF-Tu suggests that background mutations also play a role in an organism's
223 ability to adapt to selective media; the fitness increase is dependent, albeit not entirely,
224 on the ancient EF-Tu's upregulation in evolved lineages. The total number of
225 synonymous mutations accumulated in each lineage throughout the course of
226 laboratory evolution is shown in S6 Figure.
227

228 **Ancient EF-Tu strains no longer require a protein interaction partner to**
229 **maintain fitness**

230 Amongst the evolved six lineages, one lineage (Rip6) did not accumulate
231 mutations in the *thrT/tufB* region. This lineage contains some mutations that are in
232 parallel with the other lineages but also contains various mutations that are exclusive,
233 namely ribosomal initiation factor protein IF2 (*infB*) and the transcriptional regulation
234 protein NusA (Table 1). The *nusA* gene, which is known to be transcribed by the same
235 operon as *infB* (32) and autoregulates *infB* (33), acquires a 27 base-pair deletion near
236 the C-terminal end of the protein, mutant NusA protein nusAΔ9 (S Table 1). These two
237 mutations appear by Generation 500 of this lineage, each with an abundance of over
238 90% in the population.

239 Upon deletion of its native *nusA* gene, modern *E. coli* experiences a statistically
240 significant fitness decrease of 8%. In contrast to *E. coli*, chromosomal deletion of *nusA*
241 and nusAΔ9 introduces no significant fitness change to the unevolved hybrid bacterial
242 strains, nor to the evolved hybrid Rip6 harboring the mutant allele, respectively. *In vitro*
243 protein binding assays performed through Isothermal Calorimetry measurements
244 indicate that modern EF-Tu and NusA proteins interact with a robust binding constant
245 (Kd) of 14.6 ± 5.2 μM whereas ancient EF-Tu binds to NusA protein two-fold weaker
246 (Fig 3, S7 Fig). On the other hand, nusAΔ9 and ancient EF-Tu exhibit a Kd of 680 ± 66
247 μM, suggesting a complete loss of the interaction caused by at least twenty-fold lower
248 binding to ancient EF-Tu. The interaction between EF-Tu and NusA had no observable
249 effect on dipeptide formation in the ribosome. Additionally, our EF-Tu hydrolysis assays
250 using an EF-Tu H84 variant (34, 35) indicate that NusA does not bind to 70S ribosomes

251 (S8 Fig). Overall these data suggest that the interaction between EF-Tu and NusA does
252 not take place in the ribosome, further, upon replacement of the modern EF-Tu with the
253 ancient counterpart, EF-Tu-NusA interaction is lost and *nusA* accepts mutations that
254 are otherwise deleterious.
255
256 **Fig 3. Fitness change after deletion of *nusA* gene from the ancestral and evolved**
257 **bacterial genome.**
258 (Left) Bacterial constructs with *nusA* knockouts are constructed and competed against
259 the native *E. coli* bacteria for fitness measurement. (Right) The interaction between the
260 native EF-Tu, ancient EF-Tu and NusA variants are measured via *in vitro* Isothermal
261 Calorimetry binding assays.
262

263 # Discussion

264 The ability to design synthetic genes and engineer genomes has provided new
265 methods to characterize phenotypic states and the responses of biological systems to
266 perturbations (36-39). One such method involves replacing native genes with
267 homologous substitutes to investigate the gene and protein functionality of a cell (40). In
268 some cases, however, this approach is limited by divergence between host and donor
269 lineages, which can result in adverse phenotypic responses that preclude production of
270 functional hybrids suitable for study. Substituting the gene sequences of ancestors may
271 lessen these adverse impacts. Moreover, new insights into the functionality of ancestral
272 components may be possible when effects arising from the substituted sequence are
273 screened for functional innovation in artificial or cellular systems (41-50).

274 In this work we described a novel experimental system that combines ancient
275 gene reconstruction with experimental evolution and provided molecular insight into
276 organismal-level adaptation to an ancient gene. We focused on the highly-expressed
277 essential protein EF-Tu, and replaced its native gene with an inferred Precambrian
278 ancestral counterpart. Alteration of contemporary EF-Tu with its ancestor resulted in a
279 drastic decrease in cellular fitness, suggesting that the ancient gene is compatible,
280 albeit maladapted to native bacterial cell conditions. We speculate that this observation
281 is primarily due to the suboptimal function of the ancestral EF-Tu protein in the
282 ribosomal machinery, as demonstrated by previous *in vitro* biochemical assays (27) and
283 also due to the decreased protein levels upon replacement of the modern EF-Tu with its
284 ancestral homolog (S5 Fig). The reconstructed ancient EF-Tu used in this study is
285 codon optimized to function in *E. coli*; however, subtle nucleotide changes may still be
286 affecting its *in vivo* expression levels once the gene is transcribed from the genome
287 under the native *E. coli* promoter. For instance, deletion of *tufB* gene from the *E. coli*
288 genome is known to decrease the EF-Tu abundance by 15% (24). The correlation
289 between the cellular concentration of EF-Tu and organismal fitness (51) provides further
290 evidence of the intertwined role between the level of gene expression and rates of cell
291 growth (52, 53). Indeed, over-expression of ancient EF-Tu protein through an inducible
292 plasmid in the unevolved bacterial ancient-modern hybrid resulted in 10% fitness
293 increase, suggesting that the cellular concentration of the EF-Tu protein influences
294 organismal fitness.
295
296 A rapid fitness increase in populations subjected to experimental evolution
297 demonstrates the fitness constraint introduced on the system by the ancient EF-Tu.
298 Genome sequence analyses reveal that evolved ancient-modern hybrids generate a

299 complex array of cellular responses. A majority of the evolved lineages accumulate
300 mutations in the promoter location of the ancient *tuf* gene, which leads to an increase in
301 the ancient EF-Tu levels throughout these populations. This result assigns the
302 accumulation of *cis*-regulatory mutations as one of the main mechanisms in adaptation
303 for our laboratory generated system, contrary to what others have observed in studies,
304 where direct accumulation of mutations were repeatedly observed on genes engineered
305 in homologous microbes in laboratory evolution experiments (54, 55). No mutation is
306 observed on the ancient EF-Tu gene-coding region in any of our evolved lineages.
307
308 Understanding the lack of direct accumulation of mutations on the ancient EF-
309 Tu requires a consideration of various possibilities on the fitness effects of the
310 potentially contributing mutations. A suggested background mutation rate of ca.
311 1×10^{-10} to 5×10^{-9} per basepair per generation in *E. coli* (56) corresponding to a mutation
312 rate of ca. 1.8×10^{1} mutations per EF-Tu per generation and ca. 3.6×10^{4} per EF-Tu over
313 the 2000 generations of laboratory evolution. This suggests that every site in the EF-Tu
314 protein has been mutated at least once during the course of our laboratory evolution
315 experiment. Various mechanisms can explain the lack of direct accumulation of non-
316 synonymous mutations on the EF-Tu gene-coding region. The mutations experienced
317 by EF-Tu were detrimental to its essential function in the translation machinery, leading
318 to a non-viable organism. Indeed, considering the important role of EF-Tu in the cell's
319 translational machinery, as well as its role as a hub in a protein-protein interaction
320 network, mutations accumulating directly on the EF-Tu gene can cause cell lethality and
321 thus may not be readily adaptive (45, 57, 58). Relatedly, essential genes evolve slowly
322 (59); increasing the cellular protein level may represent an emergency response from
323 the organism to cope with a drastic alteration introduced by a maladapted protein

central to translation machinery. Second, the genetic integration of the suboptimal EF-Tu may have introduced a rugged adaptive landscape on the genome during the experimental evolution, prohibiting EF-Tu from accumulating a favorable amino acid, beyond the ancestral sequence. Such behavior may suggest a key role for extensive epistasis between EF-Tu's intergenic interactions that acts as a barrier between the combined deleterious effects, and the intermediate states of EF-Tu en route to adaptive solutions (60-64).

Intriguingly, EF-Tu gene amplification was not observed during the course of the laboratory evolution despite the fact that two genes encode for EF-Tu protein in the contemporary *E. coli,* a condition specific to γ-gammaproteobacterial organisms (65). It is possible such amplification could be observed in the furtherance of laboratory generations, or it is likely that the mutations that accumulated in the promoter regions override the need to have an additional copy of EF-Tu. This is an interesting outcome, given that not all the mutations lead to a final EF-Tu concentration in the cell that is equal to the modern *E. coli* EF-Tu concentration. Various mutations that accumulated in the promoter region converged on increasing protein expression levels, including a few mutations that evolved multiple times independently (such as the 19bp duplication in the +94/-21 *thrT/tufB* region in the Rip1 strain), indicating that mutations altering the expression may be more likely and therefore easier to implement in laboratory timescales of observation. It is important to note that the predominance of *cis*-regulatory mutations in our system does not diminish the obvious importance of structural mutations in all cases, but rather indicates that the former class of mutations can make a more extensive contribution on short timescales of adaptation than the

348 latter, particularly for highly-conserved essential hub proteins that have many
349 interaction partners.
350
351 One remaining ancient-modern hybrid evolved lineage, Rip6, deviates from the
352 pattern observed in the majority of the evolved lineages because it did not accumulate
353 mutations in the *cis*-regulatory region. Interestingly, Rip6 maintains an increase in
354 fitness despite no significant change in the EF-Tu protein levels. Our experiments
355 focused on the direct measurement of suppressor mutations specific to this lineage,
356 particularly *nusA* (accumulating a 27 basepair deletion in its C-terminal, which we refer
357 to as nusAΔ⏀ and potential biochemical relationship between these mutations/proteins
358 and EF-Tu activity as well as overall bacterial fitness. Gene knockout and fitness
359 studies show that the deletion of the *nusA* and nusAΔ9 genes does not alter the fitness
360 of the unevolved ancient-modern or evolved Rip6 strains, respectively. On the other
361 hand, deletion of *nusA* from the modern *E. coli* strain decreases its fitness by 8%.
362 Could this decrease be related to perturbations of an interaction between NusA and EF-
363 Tu proteins upon integration of the ancient EF-Tu? Isothermal Calorimetry binding
364 measurements show that ancient EF-Tu protein binds to native NusA protein two-fold
365 weaker than the modern EF-Tu, therefore substitution of the EF-Tu with an ancestral
366 variant diminishes the interaction with NusA. Further, the suppressor mutation
367 accumulated on the *nusA* gene in Rip6 strain does not recover the binding between the
368 NusA and the ancient EF-Tu protein. These mutations are likely instead be a symptom
369 of NusA no longer being functional in the context of an ancient EF-Tu. In addition, one
370 can speculate the lack of interaction between NusA and EF-Tu is important for the
371 adaptation of the ancient EF-Tu in the modern cell, if, especially, the interaction
372 between EF-Tu and NusA proteins did not exist 700 million years ago. To test this

373 hypothesis, we phylogenetically reconstructed an ancient NusA protein representing the
374 ancestor of γ-proteobacteria, and measured the biochemical interaction between the
375 reconstructed ancient nusA and the reconstructed ancient EF-Tu. Our Isothermal
376 Calorimetry measurements did not indicate a detectable interaction between these two
377 ancient proteins, supporting our hypothesis that an ancient interaction between NusA
378 and EF-Tu did not exist. Is NusA-EF-Tu interaction important for EF-Tu's function in the
379 translation machinery? The dipeptide formation assay exhibits no detectible functional
380 role for the NusA-EF-Tu interaction in the ribosome (S7 and S8 Figs). We speculate that
381 upon replacing the endogenous protein, a secondary function for EF-Tu beyond its
382 known functions with the ribosome is altering the strain's fitness. More extensive
383 experiments are required to examine and isolate EF-Tu's alternative roles in the cell to
384 fully explain the observed impacts on cellular fitness.
385
386 Taken together, these findings are indicative of a stringent fine-tuning by the
387 evolved microbe towards the ancient EF-Tu. In addition to decreased EF-Tu protein
388 levels, and the diminished capacity of ancient EF-Tu to participate in the translation
389 system, a third factor to explain the fitness decrease in bacteria is likely to be the
390 perturbation of EF-Tu's interactions with other proteins in the cell. A detailed
391 assessment to dissect the responses that are triggered due to the difference in function
392 of the ancient component and responses due to the laboratory evolution process is
393 necessary. One way to address this would be to use strains that have already
394 undergone adaptation to the laboratory evolution environment and then engineer the
395 bacterial genome with the ancestral variants. This additional step could also allow
396 accumulation of mutations that arise due to the adaptation to the media before the
397 foreign gene is introduced, specifically the identified intragenic mutations (such as *pykF*

398 and *mrdA* genes) that were observed in several other laboratory evolution studies using
399 *E. coli* as well as in our experimental system (66-68). Further, a detailed assessment of
400 genetic level responses is clearly a prerequisite to quantifying the overall impact of the
401 substituted ancestral gene. Such an improvement in genome engineering design would
402 facilitate performing a genome-wide assessment to separate the mutations triggered by
403 the ancient genotype/phenotype from the mutations that are simply due to the
404 adaptation to the laboratory environment.
405
406 Engineering native genomes with ancient genes was recently identified as a
407 challenging experimental approach due to the functional incompatibility of some
408 ancestral genes in modern organisms (42). The substitution of a 700 million-year-old
409 gene into a modern host demonstrates that, despite large changes in the
410 microenvironment, some portion of the gene sequence topography for the host *E. coli*
411 spanning this time interval (and possibly farther into the past) may continue to include
412 functional variants of modern genes that are not yet found in extant organisms.
413
414 On the basis of the data obtained from substitution with the ancestral EF-Tu, is
415 it possible to reveal or retrace the actual historical mechanisms or pathways followed in
416 nature by the various lineages used to construct the ancestral EF-Tu sequences? The
417 short answer appears to be no, since there were no mutations that accumulated on the
418 ancestral gene itself. All mutations appear to relate to the function of the host cell in
419 response to a suboptimal cellular component. Nevertheless, the observed outcome for
420 this particular gene in this particular host demonstrates the possibility that insightful
421 historical information may be accessible through this method. The most consistent
422 drivers of historic mutational change may be macroscopic variables that cannot be

423 readily incorporated into laboratory-scale synthetic evolution experiments, such as the
424 myriad influences of ancient ecological and environmental perturbations induced by
425 novel innovations (69-71). The EF-Tu protein phenotype, though tightly coupled to the
426 optimal growth temperature of its host organism (28), is not likely to have exhibited any
427 changes within its host bacteria over the last 700 million years of its evolution that may
428 in turn be coupled to any other externally measurable changes to the larger biological or
429 geological system. Nevertheless, the system points to the possibility of uncovering, in a
430 relative sense, how recent some protein-protein interactions (such as those between
431 EF-Tu and NusA) might be. Future studies that focus on a particular macroscopic
432 driver, in which this driver is an integral part of the laboratory evolution experimental
433 design, may yield insights into how evolutionary trajectories may have functioned in the
434 past.

435 Conclusion

436 Engineering bacterial genomes with phylogenetically reconstructed genes
437 complements the current technique of genome level alterations of gene and gene
438 clusters with extant homologs, and provides insights into molecular mechanisms of
439 adaptation by accessing historical states of currently existing proteins. The synthetic
440 system described here provides insights into adaptation of foreign genes in bacterial
441 genomes, particularly the role of regulatory vs. structural mutations in adaptive
442 evolution. We designed a method that introduced an ancient variant that is optimal in
443 some ways (presumably essential primary functions such as tRNA binding and/or
444 ribosome binding), and suboptimal in ways that evolved later in the course of evolution
445 (binding to ancillary proteins or fine-tuning epistatic interactions with tRNAs and

446 ribosomes). Our results show that lineages respond to the ancient gene by increasing
447 the expression levels of the maladapted protein, rather than through direct
448 accumulation of mutations on the ancestral gene. We expect our method to uniquely
449 reveal those components and represent a model system for a new area of the design
450 and study of functional synthetic genes in modern microbial systems.

451 # Methods

452 **Bacterial Strains and Culture Conditions**

453 All experiments were done at 37 °C unless stated otherwise. Luria-Broth (LB)
454 media was used as base media for liquid cultures and agar plates, Davis minimal media
455 (DM) was used as base media for experimental evolution and competition assays
456 (supplied with 25 mg/mL of glucose as the sole carbon source), tetrazolium arabinose
457 (TA) agar plates were used as a base media for competition experiments. When
458 required, LB and DM media was supplemented with kanamycin, chloramphenicol and
459 tetracycline antibiotics. All dilutions took place in sterile saline solution. All LB and DM
460 cultures were incubated on a rotary shaker at 200 rpm and 150 rpm, respectively. The
461 gene encoding the ancestral EF-Tu was codon optimized for expression in *E. coli*,
462 chemically synthesized by DNA 2.0 and cloned into a pET15b plasmid as reported
463 previously (72).

464

465

466

467 **Construction of the ancient-modern hybrid strain**

468 Integration of the ancient EF-Tu gene (*Rip*) into the chromosome of *E. coli*
469 strain REL606 (73) was carried out via the λ-red homology recombineering approach as
470 described by Datsenko and Wanner (74). First, linear DNA containing homology
471 sequences of upstream and downstream of *tufA* gene was amplified by PCR, via (5'
472 GTGGTTGCGAAAATCATCGCTAGAATTCCGGGGATCCGTCGACC 3' and 5'
473 TGTAATTAGCCCAGAACTTTAGCAACTGTAGGCTGGAGCTGCTTCG 3'), and pKD13
474 plasmid as template, and then transferred in REL606 cells through electrophoration,
475 together with the temperature sensitive pKD46 plasmid. Recombinants were isolated
476 from LB agar plates containing 50 μg/μL kanamycin at 37 °C, grown in in liquid LB
477 medium containing 50 μg/μL Kanamycin and their genomic DNA was isolated using
478 Promega Wizard Genomic DNA Purification Kit. Confirmation PCR was performed using
479 genomic DNA isolated from colonies as a template, with primers aligning to the
480 chromosome outside of the recombination site (5'
481 CAGGCCGTAATTGAAGCCCGTGGTAAATAAGCC 3') and (5'
482 GAATAATTTATTCGTTCTGACAGTACGAATAAG 3'). Once the successful replacement
483 of *tufA* gene with the kanamycin marker was confirmed via sanger sequencing, the
484 strain was transformed with linear DNA containing homology sequences of upstream
485 and downstream *tufB* flanking the ancient EF-Tu DNA construct SOE-d to a
486 chloramphenicol marker originally amplified from the A007 loxP-Cm-loxP plasmid
487 (Gene Bridges GmbH) via Gibson Assembly. The transformants were selected on LB
488 plates containing 25 μg/μL chloramphenicol and 50 μg/μL kanamycin at 37 °C and the
489 correct insert was screened with primers aligning to the chromosome outside of the
490 recombination site using Fwd primer 5' TCCGTGTCTTAGAGGGACAATCGATG 3'and

491 Rev primer 5' GCAATTAGCTCAGAACTTTTGCTAC 3'. Once confirmed, both the
492 kanamycin and the chloramphenicol markers were removed using pCP20 and 706-Cre
493 plasmids (Gene Bridges GmbH), respectively, followed by the confirmation of the
494 deletions by genomic PCR analysis. Plasmids pKD46 and pCP20 were cured by
495 growing the cultures at 42 °C, the final Δ*tufA*, Δt*ufB*:*Rip* construct was moved into a
496 fresh ancestral strain via bacteriophage P1 transduction . Freezer stocks of the REL606
497 Δ*tufA*, Δ*tufB*:*Rip* were prepared by mixing 50% sterile glycerol and overnight liquid
498 cultures originated from a single colony, in 1:2 ratios. All stocks were stored in -80 °C.
499 Isogenic Ara^+ variants of the REL606 Δ*tufA*, Δ*tufB:Rip* were obtained through gene-
500 gorging protocol (75)(plasmid pJEB12 is kindly donated by Jeff Barrick).

501 **Deletion of the *nusA* gene from the bacterial chromosome**

502 The *nusA* gene from the chromosome of REL606, ancestral REL606 Δ*tufA*,
503 Δ*tufB*:*Rip* and evolved REL606 Δ*tufA*, Δ*tufB:Rip* strain from lineage Rip2 were
504 replaced with a FRT-kan-FRT fragment in the presence of pKD46 helper plasmid as
505 described by Datsenko and Wanner using primers 5' TCCTGCGTGAAGATATGCTG 3'
506 and 5' TCACTTCTTCGCCGATTTCT 3'. PCR amplification of the recombination region
507 and sanger sequencing of this amplified region confirmed the correct replacement and
508 the removal of the selection cassette. The cassette was then removed from the
509 chromosome via pCP20, followed by the curation of pKD46 and pCP20 plasmids at 42
510 °C.

511

512

513 **Growth assays**

514 Saturated overnight cultures were preconditioned by dilution into sterile saline
515 by a 1:100, then again by 1:100 into the DM media, followed by an overnight growth.
516 Preconditioned cultures were diluted 1:100 into the assay medium and 100 μL of this
517 culture was transferred into a 96-well microplate. OD readings were taken at 420 nm
518 every 15 minutes with continuous shaking between readings.

519 **Experimental Evolution**

520 Experimental evolution took place through serial transfers in DM medium
521 supplemented with glucose as carbon source, for 2000 generations (~6.6 generations
522 per day) as described previously (76). Relative fitness was measured using standard
523 Lenski evolution competition assays every 500 generations via competing the evolved
524 populations and selected strains against the wild-type *E. coli* strain REL606 or REL607,
525 and calculated the fitness based on the Malthusian parameter [m=(cdx* f^x)/cd0, where
526 cd0 = count of the competitor on day 0, and cdx = count of the competitor on day x, f =
527 growth of the population over time (x-0). In our minimal media competitions, f = 100
528 because our transfers involve 100-fold dilution.

529 **Fitness measurement of the ancestral strain in the presence of over-**
530 **expressed EF-Tu**

531 Ancient EF-Tu was cloned into a pASK-IBA43 (IBA Life Sciences) vector
532 inducible under a tetracycline promoter using the following primers Forward 5'
533 GTTGGAATTCATGTCTAAAGAAAAGTTTGAACGTAC 3' and Reverse 5'
534 CGGGATCCTCAAGCGATGATTTTCGCAACCAC 3', between the *Xho* and *Nde* sites

535 leading to plasmid pASK-IBA43 harboring the ancient EF-Tu gene. Ligation was
536 confirmed using Forward primer 5' GAGTTATTTTACCACTCCCT 3' and Reverse primer
537 5' CGCAGTAGCGGTAAACG 3'. The plasmid was transferred to REL606 Δ*tufA*,
538 Δ*tufB*:*Rip* cells via electroporation and selected on LB agar plate with chloramphenicol.
539 Five representative colonies were picked, preconditioned in LB media containing 250
540 μM anhydrous tetracycline for 24 hours, followed by a 1:100 dilution into DM media
541 containing glucose. Over-expression of the EF-Tu protein was confirmed through SDS-
542 PAGE gel analysis in comparison to ancestral cells that harbored no plasmid and non-
543 induced plasmid. A REL607 strain was acclimated to the competition environment by
544 separate grown under the same environmental conditions as REL606 Δ*tufA*, Δ*tufB:Rip*
545 harboring pASK-IBA43 with the ancient EF-Tu gene. The competitors were then mixed
546 in equal stoichiometric ratios by diluting into fresh DM medium with glucose containing
547 250 μM anhydrous tetracycline. Samples were plated on tetrazolium arabinose agar
548 plate every 4 hours during the 24-hour competition. The competitions were carried out
549 two times to ensure the precision of fitness estimates.

550 **Whole genome sequencing**

551 Whole genome sequencing was completed for 2000 generations for eight lineages
552 harboring ancient EF-Tu, as well as the wild-type strains. To prepare the sequencing
553 library, we isolated 3 mg of genomic DNA from bacteria grown in 10 mL LB overnight,
554 fragmented and tagged the isolated DNA with specific Illumina adapters using Nextera
555 DNA sample preparation kit. We purified the product using Zymo DNA Clean and
556 Concentrator Kit, dual-indexed the libraries with TruSeq Dual Indexed Sequencing
557 primer sets and ensured the products were pure using an Agilent 2100 BioAnalyzer.

558 We combined the sets of compatible barcodes (11plex) into a single lane on Illumina
559 HiSeq 2500 Rapid Run flow cell (v1) after QC. Sequencing was in a paired end 2 x 100
560 basepair format (PE100) using TruSeq Rapid SBS reagents. The Breseq (0.23) software
561 was used for the generation and the analysis of the mutations (77).

562 **Luciferase Assay**

563 ***TufB and pBBRlux plasmid cloning.*** The wild-type and mutant (evolved)
564 promoter region of *tufB* gene (EF-Tu protein) was cloned into the pBBRlux plasmid as
565 adapted from Lenz et al. (78). Phusion High-Fidelity DNA polymerase, dNTPs,
566 restriction enzymes (high fidelity), and T4 ligases were all obtained from New England
567 Biolabs. DNA purification materials were purchased from QIAGEN. Promoters were
568 amplified using PCR primers 5'-CAGAATGAAAATCAGGTAGCCGAGTTCCAG-3' and
569 5'-TAGTGATTGCAGCGGTCAGCGTTGTTTTAC-3' and resulted in a 403 basepair
570 product from REL606 *E. coli* in the 4155251-4155654 region of the genome. Restriction
571 sites were subsequently added to the ends of the *tufB* promoter with the following
572 primers: Forward 5'-GATACTAGTCAGAATGAAAATCAGGTAGCCGAGTTCCAG-3' and
573 Reverse 5'-TATGGATCCTAGTGATTGCAGCGGTCAGCGTTGTTTTAC-3' (underlying
574 restriction sites correspond to *Spe*I and *Bam*HI, respectively). The EF-Tu promoter was
575 cloned upstream of the luciferase operon in the pBBRlux plasmid in order to drive
576 transcription. pBBRlux provides chloramphenicol (CMP) resistance.
577

578 ***Scintillation Counts.*** Four experimental constructs: 19 basepair duplication,
579 +94/-21 (Rip and Rip2), +87/-28 (A+87C) (Rip3), +54/-61 (G+54T) (Rip4) and +86/-29
580 (G+86A) (Rip5) and two control constructs: P (no promoter), PΔtuf (wild-type, or

581 unevolved ancestor, *tufB* promoter) were transformed into chemically competent *E. coli*
582 (REL606) cells and incubated at 37 °C for 24 hours on chloramphenicol (CMP) agar
583 plates. A single colony was cultured in LB media containing CMP at 37 °C for 24 hours.
584 A 100 μL aliquot of the overnight culture was diluted one thousand-fold prior to being
585 transferred into a 50 mL Erlenmeyer flask containing 9.9 mL of DM25 media. Cells were
586 grown for ~8.25 hours, or ~5 doublings as monitored by plating (this represents the end
587 of log growth since these cultures reach stationary phase after ~6.6 generations in
588 DM25) and then pelleted. The supernatant was aspirated until 100 μL of media
589 remained, and the pellet was then resuspended in the remaining 100 μL. Scintillation
590 counting was used to quantify the amount of light signal generated by the luciferase
591 pathway. For all six constructs, three readings per sample were averaged for each of
592 the two replicates assayed.
593
594 ***Bacterial Enumeration***. For each construct, a 10μL aliquot was serially diluted
595 50 thousand-fold and 50 μL was plated on agar petri dishes containing CMP.
596 Extrapolation was utilized to determine the total amount of cells in each scintillation
597 assay. Three plates per flask were averaged.
598
599 ***Luciferase Assay Statistical Analysis.*** The luciferase expression per cell was
600 normalized by:

601 $$\frac{\text{total scintillation counts}}{\text{total number of cells}}$$

602
603 Luciferase expression for each construct was subtracted by the amount of luciferase
604 signal from P to eliminate any leaky expression from the pBBRlux vector without

605 promoter and presented as fold-change relative to the amount of luciferase signal from
606 Patuf. A one-way ANOVA with α = 0.05 and a post-hoc Tukey HSD Test were
607 performed against Patuf to determine significant differences.

608 **Protein biochemistry**

609 ***Cloning, expression and purification of modern EF-Tu and ancient EF-Tu***
610 ***proteins.*** Both of the EF-Tu genes were ligated into pET15b plasmid between
611 BamH1/EcoR1 sites, containing an N-term His-Tag with ampicillin-resistance. For
612 expression, the plasmids were transferred in a BL21(DE3) strain, the cells were grown in
613 LB media until OD_{600} reading reached 0.6-0.8 and then induced with 1 mM IPTG for 4
614 hours at 37 °C. The cells were lysed using Bugbuster protein extraction reagent (EMD
615 Millipore) containing benzonase. For purification of the His-tagged protein from the
616 supernatant, the cell debris and membrane fractions were spun at high speed prior to
617 adding onto nitrilotriacetic acid (Ni-NTA) resin gravity-flow columns (Qiagen, Hilden,
618 Germany) at 4 °C that was pre-equilibriated with lysis buffer (50 mM NaH_2PO_4, 300 mM
619 NaCl, 10 mM imidazole, pH 8). The Ni-NTA gravity-flow column was washed twice with
620 lysis buffer containing 20 mM imidazole. His-tagged protein was eluted using elution
621 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 200 mM imidazole, pH 8).
622
623 ***Cloning, expression and purification of NusA proteins.*** Both the wild-type
624 *nusA* and the evolved *nusA* genes were amplified from their host bacterial genome
625 using Forward primer 5'-GTGAAGGTGTCGACGCTGCGTGCGCT-3' and Reverse
626 primer 5'-AGCGCACGCAGCGTCGACACCTTCAC-3'. The amplified DNA was purified
627 through gel extraction, removed from salt and then cloned into a pET15b vector

628 (Novagen) using 5' GGCGACATATGAACAAAGAAATTTTGGC 3' and 5'
629 GGAGCTCGAGTTACGCTTCGTCACCGA 3' primers in between BamH1 and XhoI sites.
630 The plasmids were transferred into a BL21(DE3) strain for expression and induced by
631 IPTG. Cells were lysed by French Press in Buffer A (20 mM Tris-HCl at pH 7.5, 50 mM
632 $MgCl_2$, 200 mM NaCl, 5% glycerol). 25 μM GDP was added in Buffer A for EF-Tu
633 purification. After centrifugation for 30 min at 16,000 rpm (F21-8x50 rotor, Thermo), the
634 supernatant was applied to Ni-NTA column and elute with gradient Buffer B (Buffer A
635 supplied with 500 mM imidazole). To prepare EF-Tu and NusA for Isothermal
636 Calorimetry experiment, the proteins were dialyzed in Buffer C (20 mM Tris-HCl at pH
637 7.5, 50 mM $MgCl_2$, 100 mM KCl) for 16 hours at 4°C.
638
639 ***Isothermal Calorimety Analysis.*** The Isothermal Calorimetry data was
640 measured on a Microcal ITC200 System (GE Healthcare). The syringe was loaded with
641 42 μL of 0.6-1 mM NusA and the sample cell was filled with 10uM EF-Tu. NusA protein
642 was titrated (2.5 μL for each) into EF-Tu with 120 s intervals and the first injection was
643 0.25 μL. The stirring speed was set at 1000 rpm. Control blank experiment was
644 performed by titrating NusA protein into Buffer C (20 mM Tris-HCl at pH 7.5, 50 mM
645 $MgCl_2$, 100 mM KCl).

646 **LC-MS/MS Analysis**

647 ***Sample preparation.*** Whole cell lysate was generated from each ancestral and
648 evolved strains using Bug Buster reagent (EMD Millipore) and following manufacturer's
649 instructions. Total protein was quantified via BCA assay using Pierce BCA protein assay
650 kit (Thermo Fisher Scientific). 30 μg of whole cell lysate was submitted to the

651 Proteomics and Metabolomics Facility at Colorado State University. Samples were
652 processed for in-solution trypsin digestion as previously described (79). Briefly, protein
653 was precipitated out of solution in the presence of 4 volumes of 100% -20oC acetone
654 and then resolubilized in 8M urea, 0.2% ProteaseMAXtm surfactant trypsin enhancer
655 (Promega, Madison, WI). Samples were reduced and alkylated with 5mM dithiothreitol
656 and 5mM iodoacetamide. Trypsin (MS Grade, Thermo Pierce, San Jose, CA) was added
657 at an enzyme to substrate ratio of 1:50 and incubated at 37 °C for 3-hours. Trypsin was
658 deactivated with the addition of 5% trifluoroacetic acid and desalted using C18 OMIX
659 tips (Agilent Technologies, Santa Clara, CA) using manufacturer's instructions. Peptide
660 eluate was dried in a vacuum evaporator and resuspended in 3% acetonitrile/0.1%
661 formic acid at a concentration of approximately 1 μg/μL. Relative Quantitation of EF-Tu
662 proteins were carried out using spectral counting approach. Approximately 2 μg of
663 tryptic digest for each sample was injected using an EASY nanoLC-II system (Thermo
664 Scientific, San Jose, CA). Peptides were purified and concentrated using an on-line
665 enrichment column (EASY-Column, 100 μm ID x 2cm ReproSil-Pur C18). Subsequent
666 chromatographic separation was performed on a reverse phase nanospray column
667 EASY-Column, 3 μm, 75 μm ID x 100mm ReproSil-Pur C18) using a 180-minute linear
668 gradient from 10%-55% buffer B (100% ACN, 0.1% formic acid) at a flow rate of 400
669 nanoliters/min. Peptides were eluted directly into the mass spectrometer (Thermo
670 Scientific Orbitrap Velos). The instrument was operated in Orbitrap-LTQ mode where
671 precursor measurements were acquired in the Orbitrap (60,000 resolution) and MS/MS
672 spectra (top 20) were acquired in the LTQ ion trap with normalized collision energy of
673 35%. Mass spectra were collected over a m/z range of 400-2000 Da using a dynamic
674 exclusion limit of 2 MS/MS spectra of a given peptide mass for 30 s (exclusion duration

675 of 90 s). Compound lists of the resulting spectra were generated using Xcalibur 2.2
676 software (Thermo Scientific) with a S/N threshold of 1.5 and 1 scan/group.
677
678 ***Data Analysis – Spectral Counting.*** *Database searching* Tandem mass spectra
679 were extracted, charge state deconvoluted and deisotoped by ProteoWizard version
680 3.0. All MS/MS samples were analyzed using Mascot (Matrix Science, London, UK;
681 version 2.3.02). Mascot was set up to search the Uniprot_e_coli_custom_reverse
682 database (Updated August 2014, 8750 entries) (80) assuming the digestion enzyme
683 trypsin, allowing up to 3 missed cleavages. Mascot was searched with a fragment ion
684 mass tolerance of 0.80 Da and a parent ion tolerance of 20 PPM. Oxidation of
685 methionine M (+15.99) and carbamidomethyl of cysteine C (+57) were specified in
686 Mascot as variable modifications. Scaffold (version Scaffold_4.3.4, Proteome Software
687 Inc., Portland, OR) was used to validate MS/MS based peptide and protein
688 identifications. Peptide identifications were accepted if they could be established at
689 greater than 69.0% probability to achieve an FDR less than 0.1% by the Scaffold Local
690 FDR algorithm. Protein identifications were accepted if they could be established at
691 greater than 99.0% probability to achieve an FDR less than 1.0% and contained at least
692 2 identified peptides (81, 82). Protein probabilities were assigned by the Protein Prophet
693 algorithm (83). Proteins that contained similar peptides and could not be differentiated
694 based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.
695 Binary comparisons were created in separate Scaffold files comparing wild-type *E. coli*
696 REL606 and unevolved ancestor harboring the ancient protein and the evolved lineages
697 tested (biological replicates n=3) to Strain/Treatment Group (each n=3). Biological
698 samples were organized into Categories based on strain type. Each Category had 3
699 biological replicates. Normalization of spectral counts was not applied based on these

700 criteria: An equal amount of sample from each replicate was loaded into the mass
701 spectrometer and there was no deviation in processing and the number of spectra
702 between samples is closely similar (% CV < 5% between biological replicates). Spectral
703 counting uses the sum of the MS/MS spectra assigned to each protein as a measure of
704 abundance (84). A t-test was performed on Total Spectral Counts for each MS sample
705 using the embedded algorithm in Scaffold v 4.3.4. Proteins with P-values less than 0.05
706 are excluded in calculation of fold changes compared to *E. coli* REL606.

707 Acknowledgements

708 We gratefully acknowledge Neerja Hajela for providing the *E. coli* REL606 strain,
709 members of the Rich Lenski laboratory for providing feedback during the earlier phase
710 of this work, Deepak Unni for his assistance during the Breseq analyses, Jennifer Zhang
711 for her assistance in the laboratory, Istem Fer for her assistance with R, and Brian
712 Hammer for providing the pBBRlux plasmid. Zachary Blount and Seth Childers
713 provided helpful discussion and comments on this manuscript.

714 References

715 1. Wray GA. The evolutionary significance of cis-regulatory mutations. Nature
716 reviews Genetics. 2007;8(3): 206-216.

717 2. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory
718 of morphological evolution. Cell. 2008;134(1): 25-36.

719 3. Carroll SB. Homeotic genes and the evolution of arthropods and chordates.
720 Nature. 1995;376(6540): 479-485.

721 4. Jacob F. Evolution and Tinkering. Science. 1977;196(4295): 1161-1166.

722 5. Zuckerkandl E. Molecular pathways to parallel evolution: I. Gene nexuses and
723 their morphological correlates. Journal of molecular evolution. 1994;39(6): 661-678.

724 6. Lynch M. The frailty of adaptive hypotheses for the origins of organismal
725 complexity. Proc Natl Acad Sci U S A. 2007;104: 8597-8604.

726 7. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins.
727 Journal of molecular biology. 1961;3: 318-356.

728 8. Nadimpalli S, Persikov AV, Singh M. Pervasive variation of transcription factor
729 orthologs contributes to regulatory network evolution. PLoS genetics. 2015;11(3):
730 e1005011.

731 9. Long A, Liti G, Luptak A, Tenaillon O. Elucidating the molecular architecture of
732 adaptation via evolve and resequence experiments. Nature reviews Genetics.
733 2015;16(10): 567-582.

734 10. Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of
735 adaptation. Evolution. 2007;61(5): 995-1016.

736 11. Prud'homme B, Gompel N, Carroll SB. Emerging principles of regulatory
737 evolution. Proc Natl Acad Sci U S A. 2007;104: 8605-8612.

738 12. Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic
739 evolution? Evolution. 2008;62(9): 2155-2177.

740 13. Voordeckers K, Pougach K, Verstrepen KJ. How do regulatory networks evolve
741 and expand throughout evolution? Current opinion in biotechnology. 2015;34: 180-188.

742 14. Lynch VJ, Wagner GP. Resurrecting the role of transcription factor change in
743 developmental evolution. Evolution. 2008;62(9): 2131-2154.

744 15. Wagner A. The origins of evolutionary innovations. A theory of transformative
745 change in living systems.: 1st ed. Oxford: Oxford University Press; 2011.

746 16. Orr HA. The genetic theory of adaptation: a brief history. Nature reviews
747 Genetics. 2005;6(2): 119-127.

748 17. Martin A, Orgogozo V. The Loci of repeated evolution: a catalog of genetic
749 hotspots of phenotypic variation. Evolution; international journal of organic evolution.
750 2013;67(5): 1235-1250.

751 18. Wilke CO. Molecular clock in neutral protein evolution. BMC genetics. 2004;5:
752 25.

753 19. Kavaliauskas D, Nissen P, Knudsen CR. The busiest of all ribosomal assistants:
754 elongation factor Tu. Biochemistry. 2012;51(13): 2642-2651.

755 20. Abdulkarim F, Hughes D. Homologous recombination between the tuf genes of
756 Salmonella typhimurium. Journal of molecular biology. 1996;260(4): 506-522.

757 21. van der Meide PH, Vijgenboom E, Talens A, Bosch L. The role of EF-Tu in the
758 expression of *tufA* and *tufB* genes. European journal of biochemistry / FEBS.
759 1983;130(2): 397-407.

760 22. Zengel JM, Lindahl L. A secondary promoter for elongation factor Tu synthesis
761 in the str ribosomal protein operon of Escherichia coli. Molecular & general genetics.
762 1982;185(3): 487-492.

763 23. Van Delft JH, Schmidt DS, Bosch L. The tRNA-tufB operon transcription
764 termination and processing upstream from *tufB*. Journal of molecular biology.
765 1987;197(4): 647-657.

766 24. Schnell R, Abdulkarim F, Kalman M, Isaksson LA. Functional EF-Tu with large C-
767 terminal extensions in an E-coli strain with a precise deletion of both chromosomal tuf
768 genes. FEBS letters. 2003;538(1-3): 139-144.

769 25. Kim DW, Uetsuki T, Kaziro Y, Yamaguchi N, Sugano S. Use of the human
770 elongation factor 1 alpha promoter as a versatile and efficient expression system. Gene.
771 1990;91(2): 217-223.

772 26. Kacar B, Gaucher EA. Towards the Recapitulation of Ancient History in the
773 Laboratory: Combining Synthetic Biology with Experimental Evolution. Proceedings of
774 the Artificial Life 2012;13: 11-18.

775 27. Zhou Y, Asahara H, Gaucher EA, Chong S. Reconstitution of translation from
776 Thermus thermophilus reveals a minimal set of components sufficient for protein
777 synthesis at high temperatures and functional conservation of modern and ancient
778 translation components. Nucleic acids research. 2012;40(16): 7932-7945.

779 28. Gromiha MM, Oobatake M, Sarai A. Important amino acid properties for
780 enhanced thermostability from mesophilic to thermophilic proteins. Biophysical
781 Chemistry. 1999;82(1): 51-67.

782 29. Lee JS, An G, Friesen JD, Fill NP. Location of the tufB promoter of E. coli:
783 cotranscription of tufB with four transfer RNA genes. Cell. 1981;25(1): 251-258.

784 30. Gerster S, Kwon T, Ludwig C, Matondo M, Vogel C, Marcotte EM, et al.
785 Statistical approach to protein quantification. Molecular & cellular proteomics.
786 2014;13(2): 666-677.

787 31. Kristensen AR, Foster LJ. High throughput strategies for probing the different
788 organizational levels of protein interaction networks. Molecular bioSystems. 2013;9(9):
789 2201-2212.

790 32. Shazand K, Tucker J, Grunberg-Manago M, Rabinowitz JC, Leighton T. Similar
791 organization of the nusA-infB operon in Bacillus subtilis and Escherichia coli. Journal of
792 bacteriology. 1993;175(10): 2880-2887.

793 33. Nakamura Y, Plumbridge J, Dondon J, Grunberg-Manago M. Evidence for
794 autoregulation of the nusA-infB operon of Escherichia coli. Gene. 1985;36(1-2): 189-
795 193.

796 34. Scarano G, Krab IM, Bocchini V, Parmeggiani A. Relevance of histidine-84 in the
797 elongation factor Tu GTPase activity and in poly(Phe) synthesis: its substitution by
798 glutamine and alanine. FEBS letters. 1995;365(2-3): 214-218.

799 35. Daviter T, Wieden HJ, Rodnina MV. Essential role of histidine 84 in elongation
800 factor Tu for the chemical step of GTP hydrolysis on the ribosome. Journal of molecular
801 biology. 2003;332(3): 689-699.

802 36. Pal C, Papp B, Posfai G. The dawn of evolutionary genome engineering. Nature
803 reviews Genetics. 2014;15(7): 504-512.

804 37. Bell G. Experimental macroevolution. Proceedings Biological sciences / The
805 Royal Society. 2016;283(1822).

806 38. Endy D. Foundations for engineering biology. Nature. 2005;438(7067): 449-453.

807 39. Esvelt KM, Wang HH. Genome-scale engineering for systems and synthetic
808 biology. Molecular systems biology. 2013;9: 641.

809 40. Yadav VG, De Mey M, Lim CG, Ajikumar PK, Stephanopoulos G. The future of
810 metabolic engineering and synthetic biology: towards a systematic practice. Metabolic
811 engineering. 2012;14(3): 233-241.

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

814 42. Hobbs JK, Prentice EJ, Groussin M, Arcus VL. Reconstructed Ancestral
815 Enzymes Impose a Fitness Cost upon Modern Bacteria Despite Exhibiting Favourable
816 Biochemical Properties. Journal of molecular evolution. 2015;81(3-4):110-120.

817 43. Kacar B. Rolling the Dice Twice: Evolving Reconstructed Ancient Proteins in
818 Extant Organisms. 2015. Preprint. Available: BioRxiv:http://dx.doi.org/10.1101/027003.
819 Accessed 21 September 2015.

820 44. Gaucher EA, Das UK, Miyamoto MM, Benner SA. The crystal structure of eEF1A
821 refines the functional predictions of an evolutionary analysis of rate changes among
822 elongation factors. Molecular Biology And Evolution. 2002;19(4): 569-573.

823 45. Kacar B, Gaucher EA. Experimental evolution of protein-protein interaction
824 networks. Biochem J. 2013;453(3): 311-319.

825 46. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, et al.
826 Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms
827 underlying evolutionary innovation through gene duplication. PLoS biology. 2012;10(12):
828 e1001446.

829 47. Voordeckers K, Verstrepen KJ. Experimental evolution of the model eukaryote
830 Saccharomyces cerevisiae yields insight into the molecular mechanisms underlying
831 adaptation. Curr Opin Microbiol. 2015;28: 1-9.

832 48. Lunzer M, Milter SP, Felsheim R, Dean AM. The biochemical architecture of an
833 ancient adaptive landscape. Science. 2005;310(5747): 499-501.

834 49. Cole MF, Gaucher EA. Exploiting models of molecular evolution to efficiently
835 direct protein engineering. Journal of molecular evolution. 2011;72(2): 193-203.

836 50. Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the
837 functional synthesis. Nature Reviews Genetics. 2007;8(9): 675-688.

838 51. Tubulekas I, Hughes D. Growth and translation elongation rate are sensitive to
839 the concentration of EF-Tu. Molecular microbiology. 1993;8(4): 761-770.

840 52. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of
841 cell growth and gene expression: origins and consequences. Science. 2010;330(6007):
842 1099-1102.

843 53. Weisse AY, Oyarzun DA, Danos V, Swain PS. Mechanistic links between cellular
844 trade-offs, gene expression, and growth. Proc Natl Acad Sci U S A. 2015;112(9): 1038-
845 1047.

846 54. Counago R, Chen S, Shamoo Y. In vivo molecular evolution reveals biophysical
847 origins of organismal fitness. Molecular Cell. 2006;22(4): 441-449.

848 55. Pena MI, Van Itallie E, Bennett MR, Shamoo Y. Evolution of a single gene
849 highlights the complexity underlying molecular descriptions of fitness. Chaos.
850 2010;20(2): 026107.

851 56. Lenski RE, Winkworth CL, Riley MA. Rates of DNA sequence evolution in
852 experimental populations of Escherichia coli during 20,000 generations. Journal of
853 molecular evolution. 2003;56(4): 498-508.

854 57. Goldman AD, Samudrala R, Baross JA. The evolution and functional repertoire
855 of translation proteins following the origin of life. Biology direct. 2010;5: 15.

856 58. Pereira-Leal JB, Levy ED, Teichmann SA. The origins and evolution of functional
857 modules: lessons from protein complexes. Philosophical transactions of the Royal
858 Society of London Series B, Biological sciences. 2006;361(1467): 507-517.

859 59. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more
860 evolutionarily conserved than are nonessential genes in bacteria. Genome research.
861 2002;12(6): 962-968.

862 60. Gong LI, Bloom JD. Epistatically interacting substitutions are enriched during
863 adaptive protein evolution. PLoS genetics. 2014;10(5): e1004328.

864 61. Kvitek DJ, Sherlock G. Reciprocal sign epistasis between frequently
865 experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS
866 genetics. 2011;7(4): e1002056.

867 62. Lunzer M, Golding GB, Dean AM. Pervasive cryptic epistasis in molecular
868 evolution. PLoS genetics. 2010;6(10): e1001162.

869 63. Bridgham JT, Ortlund EA, Thornton JW. An epistatic ratchet constrains the
870 direction of glucocorticoid receptor evolution. Nature. 2009;461(7263): 515-578.

871 64. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Microbial evolution. Global
872 epistasis makes adaptation predictable despite sequence-level stochasticity. Science.
873 2014;344(6191): 1519-1522.

874 65. Lathe WC, 3rd, Bork P. Evolution of tuf genes: ancient duplication, differential
875 loss and gene conversion. FEBS letters. 2001;502(3): 113-116.

876 66. Conrad TM, Lewis NE, Palsson BO. Microbial laboratory evolution in the era of
877 genome-scale science. Molecular systems biology. 2011;7: 509.

878 67. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome
879 evolution and adaptation in a long-term experiment with Escherichia coli. Nature.
880 2009;461(7268): 1243-1274.

881 68. Herron MD, Doebeli M. Parallel evolutionary dynamics of adaptive diversification
882 in Escherichia coli. PLoS biology. 2013;11(2): e1001490.

883 69. Hedges SB, Blair JE, Venturi ML, Shoe JL. A molecular timescale of eukaryote
884 evolution and the rise of complex multicellular life. BMC evolutionary biology. 2004;4: 2.

885 70. Menge BAS, J. P. Community regulation: Variation in disturbance, competition
886 and predation in relation to environmental stress and recruitment. American Naturalist.
887 1987;130: 730-757.

888 71. Lambeck K, Esat TM, Potter EK. Links between climate and sea levels for the
889 past three million years. Nature. 2002;419(6903): 199-206.

890 72. Gaucher EA, Govindarajan S, Ganesh OK. Palaeotemperature trend for
891 Precambrian life inferred from resurrected proteins. Nature. 2008;451(7179): 704-707.

892 73. Blount ZD. The unexhausted potential of E. coli. eLife. 2015;4.

893 74. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in
894 Escherichia coli K-12 using PCR products. P Natl Acad Sci USA. 2000;97(12): 6640-
895 6645.

896 75. Herring CD, Glasner JD, Blattner FR. Gene replacement without selection:
897 regulated suppression of amber mutations in Escherichia coli. Gene. 2003;311: 153-
898 163.

899 76. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics
900 and genetic bases of adaptation. Nature reviews Genetics. 2003;4(6): 457-469.

901 77. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved
902 microbes from next-generation sequencing data using breseq. Methods in molecular
903 biology. 2014;1151: 165-188.

904 78. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. The small
905 RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi
906 and Vibrio cholerae. Cell. 2004;118(1): 69-82.

907 79. Schauer KL, Freund DM, Prenni JE, Curthoys NP. Proteomic profiling and
908 pathway analysis of the response of rat renal proximal convoluted tubules to metabolic
909 acidosis. American journal of physiology Renal physiology. 2013;305(5): 628-640.

910 80. Elias JE, Gygi SP. Target-decoy search strategy for mass spectrometry-based
911 proteomics. Methods in molecular biology. 2010;604: 55-71.

912 81. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to
913 estimate the accuracy of peptide identifications made by MS/MS and database search.
914 Analytical chemistry. 2002;74(20): 5383-5392.

915 82. Kall L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides
916 identified by tandem mass spectrometry using decoy databases. J Proteome Res.
917 2008;7(1): 29-34.

918 83. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying
919 proteins by tandem mass spectrometry. Analytical chemistry. 2003;75(17): 4646-4658.

920 84. Paoletti AC, Washburn MP. Quantitation in proteomic experiments utilizing mass
921 spectrometry. Biotechnology & genetic engineering reviews. 2006;22: 1-19.
922

923

924

925

926

927

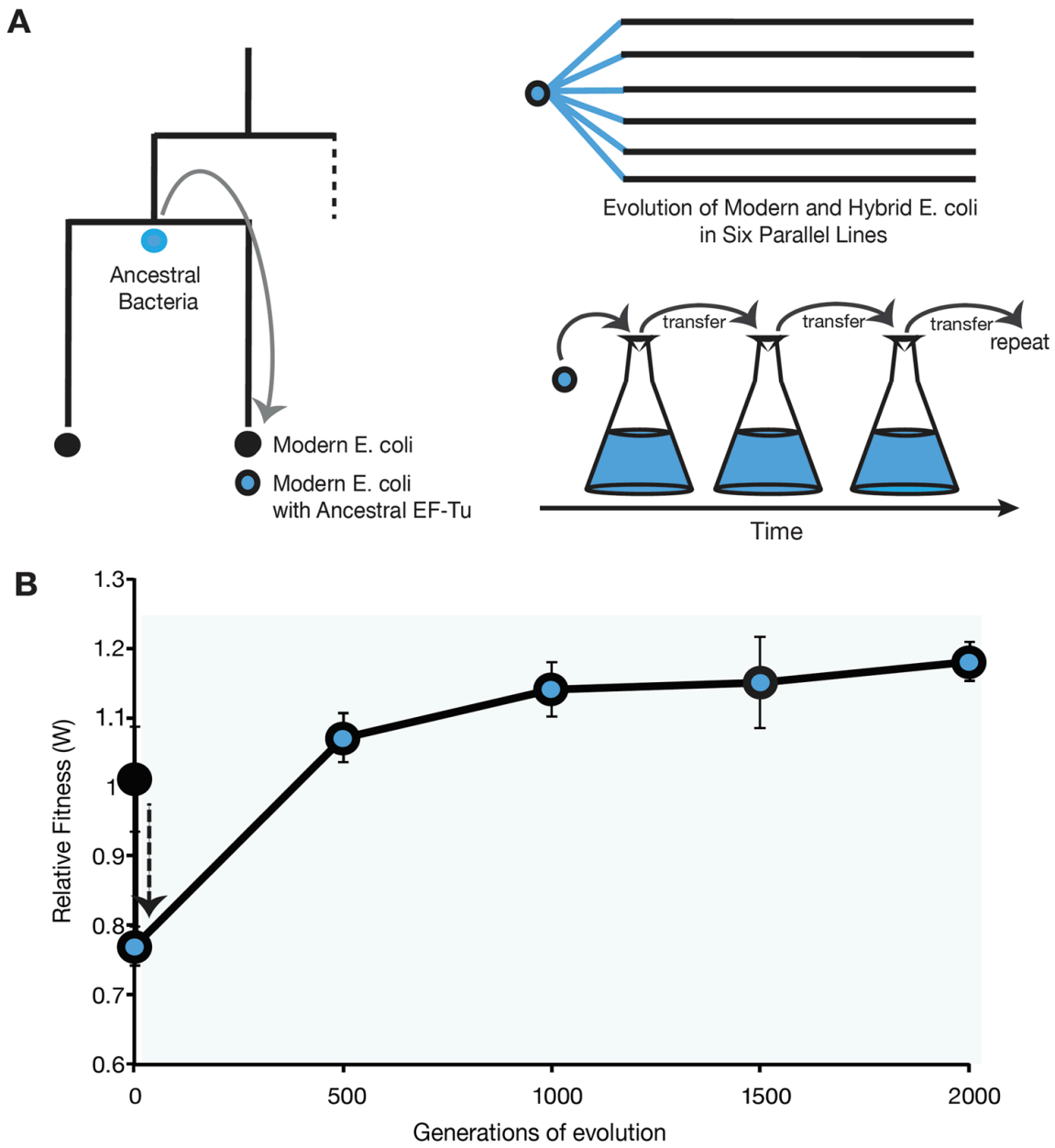
928

929

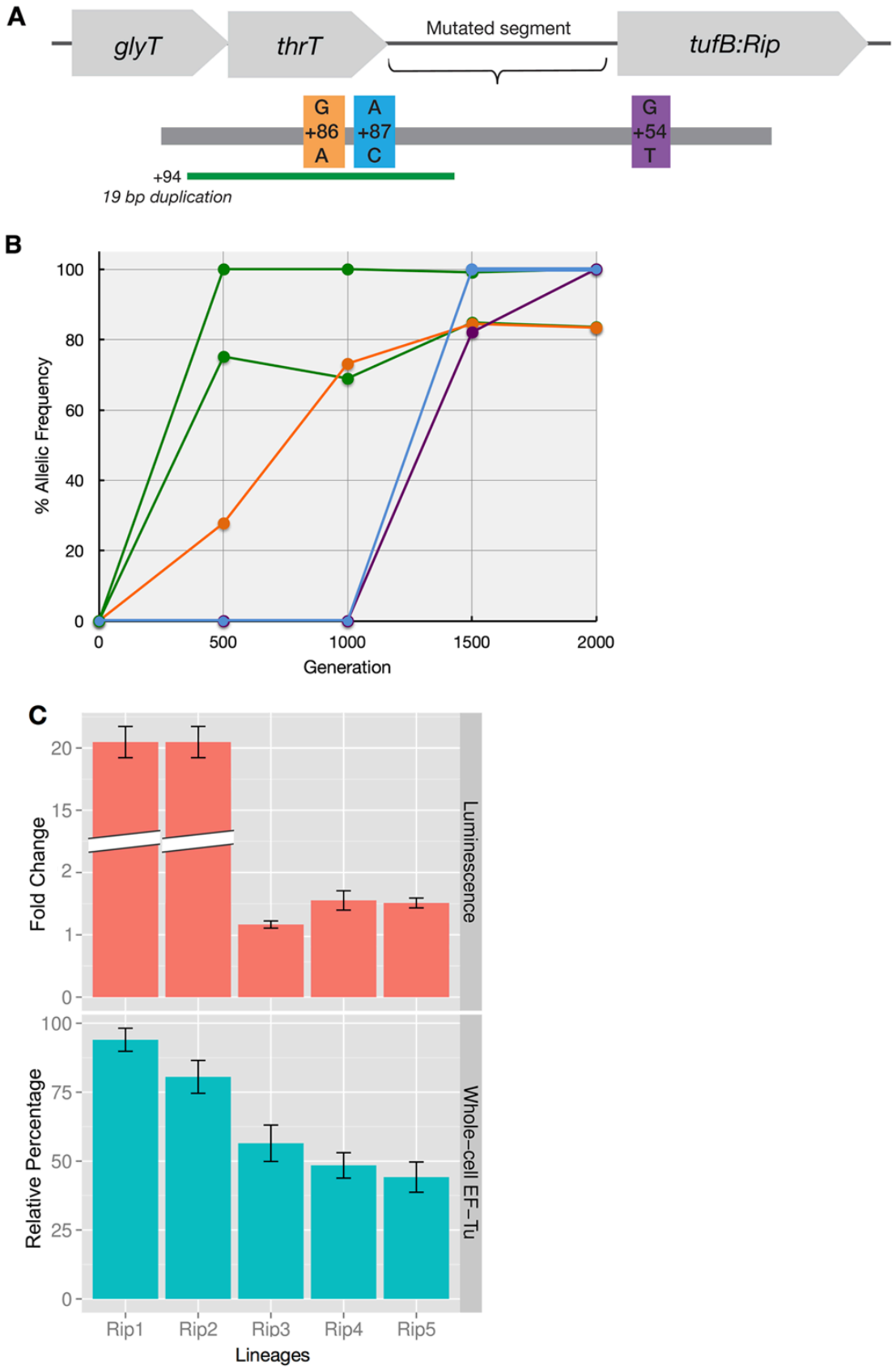
930
931

Figures

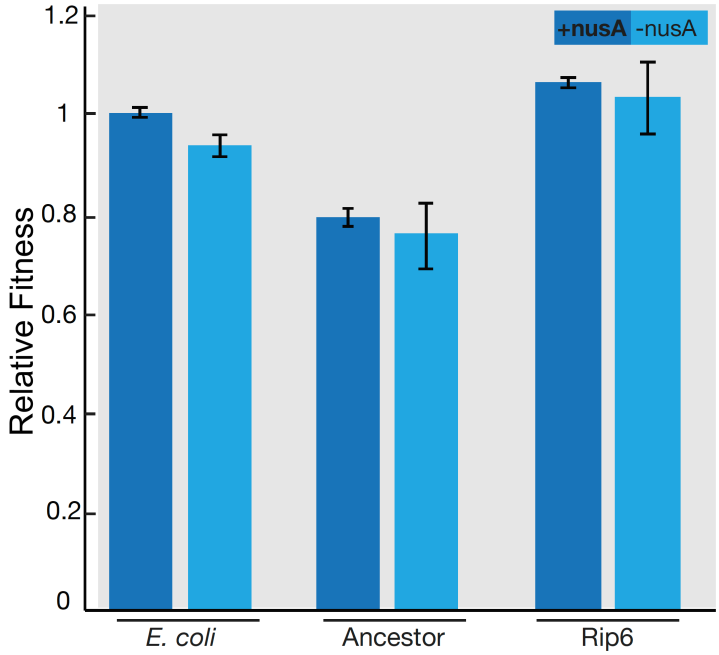
Fig1.

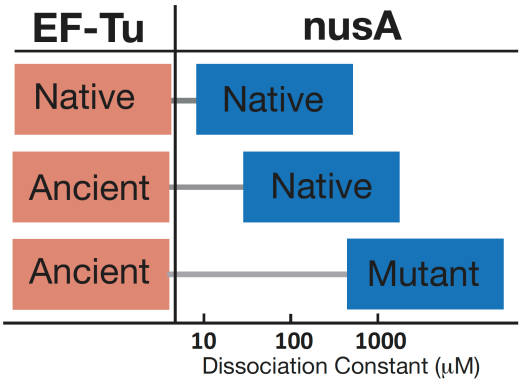


939 **Fig2.**


A
glyT
thrT
Mutated segment
tufB:Rip
G
+86
A
A
+87
C
G
+54
T
+94
19 bp duplication
B
% Allelic Frequency
100
80
60
40
20
0
0
500
1000
1500
2000
Generation
C
Fold Change
20
15
2
1
0
Luminescence
Relative Percentage
100
75
50
25
0
Whole-cell EF-Tu
Rip1
Rip2
Rip3
Rip4
Rip5
Lineages

951 **Fig3.**





952

953

954

955

956

957

958

959

960

961

962

963

964

965