1   MOLECULAR BIOLOGY AND EVOLUTION: In Review

2

3   **First-step Mutations during Adaptation to Thermal Stress Shift the**

4   **Expression of Thousands of Genes Back toward the Pre-stressed State**

5

6   Alejandra Rodríguez-Verdugo[1,#a,b], Olivier Tenaillon[2,3] and Brandon S. Gaut[1,*]

7

8   [1] Department of Ecology and Evolutionary Biology, University of California Irvine,

9   Irvine, United States

10   [2] INSERM, IAME, UMR 1137, F-75018 Paris, France

11   [3] Université Paris Diderot, IAME, UMR 1137, Sorbonne Paris Cité, F-75018

12   Paris, France

13   [#a] Current Address: Department of Environmental Systems Sciences, ETH

14   Zürich, Zürich, Switzerland

15   [#b] Current Address: Department of Environmental Microbiology, Eawag,

16   Dübendorf, Switzerland

17

18   [*] Corresponding author:

19   Brandon S. Gaut

20   e-mail: bgaut@uci.edu

21

Abstract

22

23   The temporal change of phenotypes during the adaptive process remain

24 largely unexplored, as do the genetic changes that affect these phenotypic

25 changes. Here we focused on three mutations that rose to high frequency in the

26 early stages of adaptation within 12 *Escherichia coli* populations subjected to

27 thermal stress (42ºC). All of the mutations were in the *rpoB* gene, which encodes

28 the RNA polymerase beta subunit. For each mutation, we measured the growth

29 curves and gene expression (mRNAseq) of clones at 42ºC. We also compared

30 growth and gene expression to their ancestor under unstressed (37ºC) and

31 stressed conditions (42ºC). Each of the three mutations changed the expression

32 of hundreds of genes and conferred large fitness advantages, apparently through

33 the restoration of global gene expression from the stressed towards the pre-

34 stressed state. Finally, we compared the phenotypic characteristics of one

35 mutant, *I572L*, to two high-temperature adapted clones that have this mutation

36 plus additional background mutations. The background mutations increased

37 fitness, but they did not substantially change gene expression. We conclude that

38 early mutations in a global transcriptional regulator cause extensive changes in

39 gene expression, many of which are likely under positive selection for their effect

40 in restoring the pre-stress physiology.

41

Introduction

42

43   Organisms are often exposed to stressful environments. To cope, they

44   have evolved responses based on the duration of the stress. For example,

45   bacteria that are exposed to increased temperatures display a transient heat-

46   shock response, which involves up-regulation of genes encoding heat stress-

47   proteins (Richter et al. 2010), followed by a period of phenotypic acclimation

48   (Gunasekera et al. 2008). If the environmental stress is sustained over a long

49   period of time, individuals may eventually accumulate mutations that result in

50   long-term adaptation of the population. Although acute responses to stress have

51   been well characterized (Gunasekera et al. 2008), the mechanisms of stress

52   acclimation and adaptation are understood less well (Riehle et al. 2001;

53   Gunasekera et al. 2008).

54   One possible mechanism for adaptation to stressful conditions is genetic

55   change that produces novel traits or new physiological functions, such as

56   antibiotic resistance or the ability to use new metabolic pathways (Blount et al.

57   2012; Quandt et al. 2014). Another mechanism is genetic change that leads with

58   restoration of cellular functions to a pre-stressed state. Rather than creating new

59   functions, restorative mutations revert some aspect of the individuals' altered

60   physiological state back to an unstressed or "normal" state (Carroll and Marx

61   2013). This pattern of restoration has been recently observed during metabolic

62   perturbation and high-temperature adaptation (Fong et al. 2005; Carroll and Marx

3

63   2013; Sandberg et al. 2014). Although both novelty and restoration likely drive

64   long-term adaptation, it is not yet clear which is more frequent.

65   Another issue that remains largely unexplored is the temporal pace of

66   phenotypic change during adaptation to stress. After an environment becomes

67   stressful, the acclimated state becomes the initial phenotypic state upon which

68   natural selection acts. Thereafter, each adaptive mutation moves the population

69   towards a phenotypic optimum (i.e. to a phenotype that best fits the stressful

70   environment (Fisher 1930; Orr 2005)). For historical and methodological reasons

71   (Orr 2005), most evolutionary studies have focused on the end product of

72   adaptation, leaving the early and intermediate steps of adaptation largely

73   unexplored.

74   Fortunately, studies in experimental evolution may provide insight into the

75   sequential magnitude of fitness changes during adaptation. To date, these

76   studies have shown that the first beneficial mutation, or the "first-step mutation",

77   generally confers a large fitness advantage, perhaps because early, large-effect

78   mutations outcompete other beneficial mutations in the population. In contrast,

79   subsequent mutations confer more moderate fitness gains (Chou et al. 2011;

80   Khan et al. 2011), in part due to diminishing returns epistasis (Kryazhimskiy et al.

81   2014).

82   While our knowledge about the temporal trajectory of fitness continues to

83   grow, few studies have examined shifts in phenotypes during the adaptive

84   process. One exception is the study of Fong et al. (2005), which used

85   microarrays to follow the phenotype of gene expression (GE) after a shift in

4

86   growth from glucose to lactate and from glucose to glycerol (Fong et al. 2005).

87   Fong et al. (2005) observed that 39% of genes were differentially expressed

88   during the process of acclimation to glycerol medium; however, the number of

89   differentially expressed genes decreased at an intermediate point of adaptation

90   and also at the endpoint of the experiment. Most of the GE changes during

91   adaptation were restorative (Fong et al. 2005) –i.e., they restored GE to normal,

92   pre-stress levels. Unfortunately, however, the genetic bases of these phenotypic

93   changes were not resolved; it was unclear if the changes in GE during the

94   intermediate point of adaptation were caused by an early adaptive mutation or

95   were caused by combinations of mutations that accumulated during the

96   experiment. Thus, many questions regarding phenotypic adaptation remain

97   unresolved, such as: What are the molecular effects underlying the large fitness

98   advantage conferred by first-step mutations? And, what is the phenotypic

99   contribution of the first-step mutation compared to later steps in an adaptive

100   walk?

101   We have decided to explore these questions based on our recent, large-

102   scale evolution experiment (Tenaillon et al. 2012). In this experiment a strain of

103   *E. coli* B was evolved at 42ºC in 114 replicate populations. After 2000

104   generations, we isolated single clones from each evolved population and

105   identified the genetic changes that had accumulated during the experiment.

106   Overall, the *rpoB* gene, which encodes the β subunit of RNA polymerase

107   (RNAP), was the most mutated gene with a total of 87 mutations in the 114

108   genomes. Three *rpoB* mutations were especially interesting because they

5

109   resulted in amino acid substitutions in the active site of RNAP and conferred

110   rifampicin resistance (Rodríguez-Verdugo et al. 2013). These mutations, all of

111   which are located in codon 572, were driven to high frequency in 12 populations

112   and were also beneficial at high temperature in the low glucose medium

113   (Rodríguez-Verdugo et al. 2013). Furthermore, these mutations typically

114   appeared early in the evolution experiment. In one population, for example, the

115   *rpoB I572L* mutation swept to fixation before 100 generations; it was furthermore

116   shown to confer a large fitness benefit of about 20% (Rodríguez-Verdugo et al.

117   2013). Since these *rpoB* mutations are located in the contact region between the

118   downstream DNA duplex and RNAP (proximal active-site), they could play key

119   roles in modulating the enzyme's activity in all three stages of transcription:

120   initiation, elongation and termination (Ederth et al. 2006).

121   Yet, some questions remain. First, we have not identified the mechanistic

122   bases of their fitness advantage. Second, although we know that these parallel

123   mutations converge on having a fitness advantage at 42ºC (Rodríguez-Verdugo

124   et al. 2013), with disadvantages at low temperatures (Rodríguez-Verdugo et al.

125   2014), we have yet to explore if they converged on other phenotypes. That is, it

126   is still unclear if these mutations confer the same phenotypic effects as *rpoB*

127   mutations found in other regions of RNAP that were also selected in some

128   lineages during the thermal adaptation experiment. In the present study, we

129   address all these questions. Solving these questions is relevant for

130   understanding the adaptive mechanisms prevalent in our experimental conditions

131   and also in other evolutionary experiments, because mutations that affect

6

132   transcriptional regulators (such as RNAP and the Rho termination factor) typically

133   appear in the early stages of stress adaptation (Applebee et al. 2008; Goodarzi

134   et al. 2009; Kishimoto et al. 2010). Numerous observations point to the possibility

135   that highly pleiotropic mutations in transcriptional regulators could be the first

136   step of a general mechanism of adaptation (Fong et al. 2005).

137   With these considerations in mind, we have investigated the timeline of

138   adaptive change during our evolution experiment (fig. 1). To do so, we first

139   describe growth characteristics and GE during acclimation to thermal stress at

140   42⁰C. We then explore first-step mutations in terms of their effect on growth and

141   GE relative to the ancestral strain. More specifically, we explore whether three

142   amino acid substitutions in the RNAP active site converge on the same

143   expression phenotype and also whether their effects are similar to another *rpoB*

144   mutation that affects an amino acid substitution far from the RNAP active site.

145   Finally, we evaluated the phenotypic contribution of one of these first-step

146   mutations relative to the end product of adaptation, as represented by two of

147   clones from the end of the 2000 generation experiment.

148

Results

149   **Results**

150   **Acclimation to thermal stress involved many changes in GE.**

151   To explore the phenotypic effect of high temperature on the ancestor, we

152   first characterized the ancestor's growth at 37⁰C, which we consider to be the

153   pre-stress condition, and at 42⁰C, the stress condition. The ancestor had a

154   significantly longer lag phase and a significantly lower maximum growth rate and

155   final yield when grown at 42ºC than when grown at 37ºC (fig. 2*A* and table 1),

156   reflecting the fact that 42ºC is a stressful temperature.

157   Next, we explored the global GE profile after acclimation to 42ºC by

158   obtaining RNA-seq data from three replicates of the ancestor at the mid-

159   exponential growth phase at both 37ºC and 42ºC. Even under a stringent

160   criterion of significance ($q < 0.001$), we identified 1737 genes that were

161   differentially expressed at 42ºC relative to 37ºC (fig. 2*B* and supplementary

162   dataset S1). Of these differentially expressed genes, 240 genes were highly

163   down-regulated (log_2fold change < -2; blue dots in fig. 2*B*) and 293 genes were

164   highly up-regulated (log_2fold change > 2; red dots in fig. 2*B*) at 42ºC relative to

165   37ºC. Based on an enrichment analysis of GO functional categories, we identified

166   significant down-regulation of genes involved in translation during heat

167   acclimation (GO: 0006412; $P = 4.55*10^{-29}$). Of the genes involved in translation,

168   54 genes transcribed products that were structural constituents of ribosomes

169   (GO:0003735), including *rpl*, *rpm* and *rps* genes (supplementary table S1). Other

170   significantly down-regulated biological processes involved: *i*) amino acid

171   biosynthesis (GO: 0008652; $P = 2.17*10^{-15}$), particularly genes involved in the

172   biosynthesis of methionine (*met* genes); *ii*) biosynthesis of ribonucleosides (GO:

173   0042455; $P = 3.12*10^{-7}$), including purines and pyrimidine (*pur* and *pyr* genes);

174   and *iii*) organelle organization (GO: 0006996; $P = 2.86*10^{-2}$), including *flg* and *flh*

175   genes (supplementary table S1). Surprisingly, the heat-shock inducible genes

176   (Nonaka et al. 2006), were mostly down regulated during heat acclimation, as

8

177 were the subunits of the core RNAP (*rpoA*, *rpoB*, *rpoC* and *rpoZ*) (supplementary

178 table S2).

179   Among the up-regulated genes, we identified a significant enrichment of

180 genes involved in: *i*) amino acid degradation (GO: 0009063; $P = 6.55*10^{-4}$) such

181 as degradation of arginine (*ast* genes); *ii*) alcohol degradation (GO:0046164; $P =$

182 $5.44*10^{-4}$); *iii*) glycerol metabolic process (GO:0015794; $P = 1.73*10^{-2}$); *iv*)

183 monosaccharide transport (GO: 0015749; $P = 1.11*10^{-2}$); and *v*) cellular

184 response to stress (GO:0033554; $P = 4.60*10^{-2}$). Of the 66 genes that had been

185 identified previously as a component of the general stress response (Weber et al.

186 2005), 57 (86%) were significantly up-regulated at 42ºC in our experiment

187 (supplementary table S2).

188   Overall patterns of GE suggest that the ancestor at 42ºC induced an early

189 stationary phase. To be sure that this GE profile did not result from sampling at

190 the wrong phase of the growth curve, we repeated the experiment. We again

191 sampled three replicates at mid-exponential, repeated RNA sampling, and

192 transcriptome analyses. The results were qualitatively similar between

193 experiments, implying both that the ancestor at 42°C cannot unleash full growth

194 and that the population expands at a low rate in this stressed physiological state.

195 **First-step adaptive mutations changed GE at 42ºC dramatically.**

196   Given insights into the acclimation response at 42ºC, we then investigated

197 the effect of three potential first-step mutations – *rpoB I572F*, *I572L* and *I572N* –

198 on cell growth and GE. Previous estimates of relative fitness indicated that these

199 mutations are advantageous at 42ºC (Rodríguez-Verdugo et al. 2013). We

9

200   complemented this observation by characterizing their growth curves at both

201   42ºC and at 37ºC. The mutants had a significantly longer lag phase and lower

202   final yield at 42ºC compared to their growth at 37ºC (Table 1). Nevertheless, the

203   mutants had a significantly higher maximum growth rate and a significant higher

204   final yield compared to the ancestor grown at 42ºC, reflecting improved fitness

205   under thermal stress (table 1 and fig. 3).

206   We explored two hypotheses about the molecular mechanisms that may

207   underlie the growth improvement of the mutants at 42ºC. Our first hypothesis was

208   the mutants grew better at 42ºC, because the mutated RNAP was more efficient

209   than the ancestral RNAP at transcribing DNA to RNA (Jin et al. 1992; Reynolds

210   2000). Our second, non-exclusive hypothesis was both that the mutated RNAP

211   led directly or indirectly to altered transcription of a set of genes and that this

212   differential expression underlies growth improvement (Conrad et al. 2010).

213   Regarding the first hypothesis, we predicted that the *rpoB* mutations

214   slowed the RNAP complex (Rodríguez-Verdugo et al. 2014), which is otherwise

215   accelerated at high temperatures (Ryals et al. 1982; Mejia et al. 2008). In turn,

216   we reasoned that the reduced speed of RNAP increases both transcription

217   fidelity and termination efficiency (Jin et al. 1988; Zhou et al. 2013), resulting in

218   an overall higher transcription efficiency (Jin et al. 1992). To measure the

219   transcription efficiency of the mutants and the ancestor at 42ºC, we measured

220   the relative abundance of an inducible fluorescent gene (*YFP*) inserted in our

221   strains, at different points post induction (see *Materials and Methods*). Using this

222   method (Reynolds 2000; Brandis et al. 2012), we found that two of the three

10

223   mutants (*I572L* and *I572F*) yielded higher slopes than the ancestor at 42⁰C,

224   consistent with our hypothesis of higher efficiency. However, the assay has wide

225   confidence intervals, and there were therefore no statistical differences in the

226   transcription efficiency of the mutants relative to the ancestor at 42⁰C

227   (supplementary fig. S1). Within the limits of our experiment, we cannot conclude

228   that the growth improvement of the mutants at 42⁰C is caused by higher RNAP

229   efficiency.

230   We also addressed our second hypothesis that the *rpoB* mutations lead to

231   changes in GE with consequent growth improvements at 42⁰C. To test this

232   hypothesis, we obtained RNAseq data from each mutant, based on two

233   replicates per mutant at mid-exponential growth at 42⁰C. We then contrasted the

234   GE profile from each mutant against that of the ancestor grown at 42⁰C.

235   All three mutations displayed hundreds to thousand of differentially

236   expressed genes (q <0.001; fig. 3). The mutant *I572F* had 1332 differentially

237   expressed genes at 42⁰C with 195 highly down-regulated ($\log_2$fold change < -2)

238   and 182 highly up-regulated ($\log_2$fold change > 2) relative to the 42⁰C ancestor.

239   The mutant *I572L* had 598 differentially expressed genes with 126 highly down-

240   regulated and 126 highly up-regulated. Finally, the mutant *I572N* had 1360

241   differentially expressed genes with 233 highly down-regulated genes and 164

242   highly up-regulated genes. To sum: single mutations within codon 572 of *rpo*B

243   lead to numerous, pleiotropic shifts in GE.

11

244 ***rpoB* mutations in codon 572 restored GE back toward the ancestral state.**

245   To investigate the general trend in GE changes from the ancestor to

246 acclimation and then from acclimation to first-step mutations (fig. 1), we used two

247 approaches. The first was Principal Component Analysis (PCA), which

248 simultaneously considered the expression of all genes and all clones. The

249 resultant plot of the first and second components exhibited differences among

250 clones and also clearly indicated that the clones with first-step mutations shifted

251 their gene expression profiles toward the 37ºC ancestor (supplementary fig. S2).

252   Our second approach was to plot the $\log_2$fold expression change during

253 the acclimation response (ancestor grown at 42ºC *vs* ancestor grown at 37ºC, x-

254 axis) against the $\log_2$fold GE change during the adaptive response (mutant

255 grown at 42ºC *vs* ancestor grown at 42ºC, y-axis; fig. 4). If there were no

256 relationship between shifts in GE during acclimation and during the adaptive

257 response, we expected a regression slope of zero; alternatively, if all the genes

258 of a mutant were restored by an adaptive mutation – such that the expression of

259 each genes changed from a stressed state back to a pre-stressed state – the

260 slope of the fitted regression would approach -1.0. For the three mutants we

261 observed a highly significant negative correlation, with slopes of -0.738 for *I572F*,

262 -0.639 for *I572L* and -0.731 for *I572N* (fig. 4*A-C*). All three slopes were highly

263 significant ($p < 0.001$) based on permutation tests (see *Material and Methods*),

264 indicating that the main phenotypic effect of the first-step mutations was to

265 restore global GE toward the pre-stressed state.

12

266   We next examined the GE changes in more detail by characterizing the

267   expression of individual genes into one of four patterns of change that denote the

268   direction of the effect of the mutated RNAP (see *Materials and Methods*,

269   supplementary table S3). First, the expression of an individual gene could be

270   *restored* back toward the pre-stressed state. Second, the expression of a gene

271   could be *reinforced* into an exaggeration of the acclimated state, such that the

272   mutated RNAP produced more transcripts (in the case of up-regulated genes) or

273   fewer transcripts (in the case of down-regulated genes) than the acclimated

274   state. Third, a gene was *unrestored* if the mutated RNAP did not change GE

275   significantly relative to the ancestral RNAP at 42ºC. Finally, GE was *novel* if it

276   was not differentially expressed during the acclimation response, but was

277   expressed significantly differently during the adaptive response. For these

278   definitions, we investigated genes that differed in expression either during

279   acclimation or between acclimation and first-mutation clones (supplementary

280   table S3); as a result, we categorized directional shifts of roughly half of 4202 *E.*

281   *coli* genes.

282   Following this categorization, we observed that most (63%) of the genes

283   that were differentially expressed during acclimation were restored by mutant

284   *I572F* (blue dots in fig. 4*A*; table 2), as expected from the strong overall negative

285   correlation in figure 4*A*. The same was true for mutant *I572N*, for which most

286   (67%) genes were restored (table 2). For mutant *I572L*, we found most genes to

287   be unrestored (68%), but a substantial proportion of genes (32%) had restored

288   expression (table 2). In contrast to restoration, only a small proportion (~3% or

13

289   less) of characterized genes exhibited novel expression in the *rpo*B mutants

290   (table 2). While the statistical power to detect differences likely varies among

291   clones, together the data reinforce the observation that most genes are restored

292   in gene expression by single *rpoB* mutants.

293

294   **An *rpoB* mutation away from the active site of the RNAP also restores GE.**

295   Given that *rpoB* mutations in codon 572 converged in phenotype toward

296   restorative GE, we sought to know if additional mutations in *rpoB* had similar

297   effects. To address this question, we constructed a single mutant, *rpoB I966S*,

298   which alters one of the two parallel α-helices of the Eco flap domain of RNAP

299   (Opalka et al. 2010). The *I966S* mutation was found in 15 of our high-

300   temperature adapted clones (Tenaillon et al. 2012), and its fitness advantage at

301   42ºC was confirmed by competition experiments (Rodríguez-Verdugo et al.

302   2014).

303   We obtained RNAseq data from two replicates of this mutant at 42ºC and

304   performed the same GE analyses. We observed a global pattern of GE change

305   similar to that for the *I572F*, *I572L* and *I572N* mutants. That is, the *I966S*

306   mutation tended to restore GE from the acclimated state toward the pre-stressed

307   state (fig. 4*D*).

308   To explore the convergent effects of the *I966S*, *I572F, I572L* and *I572N*

309   mutations further, we determined the number of genes with restored GE shared

310   among the four mutants. A third of the restored genes were shared between the

311   four mutants, and 86% of the genes were share by at least two mutants (fig. 5),

14

312   indicating a high level of expression convergence. This high level of phenotypic

313   convergence was also highlighted by the pairwise comparisons of differential

314   expression between the mutants (supplementary fig. S3). We performed an

315   enrichment analysis of GO assignments for the 490 restored genes that were

316   shared among the four mutants. Not surprisingly, given the overall pattern of

317   restoration (fig. 4), the restored genes represented the same sets of genes that

318   were enriched for the acclimation response. For example, significantly down-

319   regulated genes during the acclimation response, such as the genes involved in

320   translation (*rpl*, *rpm* and *rps* genes), were significantly up-regulated in the four

321   mutants. This suggests that the different RNAP mutants restored GE towards a

322   growth pattern.

323   We also determined how many genes of novel GE were shared between

324   the four single mutants: *I572F*, *I572L*, *I572N* and *I966S*. Only 2% of the genes

325   with novel GE were shared between the four mutants, and 33% of the genes

326   were share by at least two mutants (supplementary fig. S4), indicating lower

327   expression convergence for genes with novel GE than for restored genes.

328   In conclusion, the phenotypic convergence observed between the mutants

329   *I572F*, *I572L, I572N* and *I966S* suggests: *i*) restoration of the altered

330   physiological state back toward a pre-stressed state, *ii*) that much of that

331   restoration was achieved by single, highly pleiotropic mutations and *iii*) similar

332   effects result from different amino acid mutations at the same site (codon 572 for

333   mutants *I572F*, *I572L* and *I572N*) or at a different site (codon 966 for mutant

334   *I966S*).

15

Mutations fixed during adaptation contributed few changes in GE.

Following the temporal schematic of this study (fig. 1), our last step was to contrast the GE phenotype of one first-step mutation (*I572L*) to end-products of our adaptation experiment (i.e. clones evolved 2000 generations at 42⁰C; fig. 6*A*). As end-products, we chose two high-temperature adapted clones: clone 27 and clone 97 (clone numbers correspond to reference Tenaillon et al. 2012), which were isolated from two populations in which the mutation *I572L* swept to fixation before 400 generations (Rodríguez-Verdugo et al. 2013). Although sharing the same *rpoB* mutation, these two high-temperature adapted clones differed in their genetic backgrounds (supplementary tables S4 and S5). Clone 27 had two large deletions of 2,896 and 71,416 bp affecting > 65 genes, a 138 bp deletion disrupting the tRNA-Met gene and seven point mutations in seven genes. In contrast, clone 97 had only one 4 bp small deletion, an IS insertion and eight point mutations in eight different genes (Tenaillon et al. 2012).

We again characterized the clones' growth curves at 42⁰C and compared them to both the ancestor grown at 37⁰C, the ancestor grown at 42⁰C and the mutant *I572L* grown at 42⁰C (table 3, fig. 6*A* and supplementary fig. S5). The high-temperature adapted clones 27 and 97 both had significantly shorter lag phases compared to *I572L* mutant. In addition, clone 27 had a significantly higher final yield than *I572L* mutant (table 3). It thus appears that the mutations accumulated after the first-step mutations contribute to better growth at 42⁰C. We note, however, that none of the adapted clones grow as well as the ancestor at

16

357   37ºC, which had a significantly shorter lag phase and a higher final yield (table 3

358   and fig. 6*A*).

359   Knowing that the evolved clones grew more quickly and to higher yields

360   than the single mutant *I572L* at 42ºC, we sought to identify GE changes that

361   could explain their improved growth. We obtained RNA-seq data from each high-

362   temperature adapted clone (two replicates per clone) grown at 42ºC. We then

363   contrasted the GE from each clone against the GE of the first-step mutation

364   *I572L* at 42ºC. To our surprise, we observed very few genes that were

365   differentially expressed (q <0.001): 7 for clone 27 in addition to 56 genes with no

366   measurable GE (fig. 6*B*, supplementary table S4) and 16 for clone 97

367   (supplementary fig. S5). Furthermore, the high-temperature adapted clones

368   maintained the general pattern of restoration back to the ancestral physiological

369   state previously observed for the mutant *I572L* (slope of -0.654 for clone 27 and

370   slope of -0.645 for clone 97; supplementary fig. S6).

371   These observations suggest that the mutations accumulated in later steps

372   of thermal stress adaptation did not substantially change the expression profile

373   caused by the first-step mutation; that is, most detectable GE shifts were caused

374   by the *I572L* mutation and were associated to the restoration of the growth

375   profile. To emphasize this point, we plotted the $\log_2$fold expression change during

376   the first-step adaptive response (mutant *I572L* grown at 42ºC *vs* ancestor grown

377   at 42ºC; supplementary fig. S7) against the $\log_2$fold expression change during

378   the complete adaptive response (high-temperature adapted clone grown at 42ºC

379   *vs* ancestor grown at 42ºC; supplementary fig. S7). We observed a highly

17

380   significant positive correlation (slope of 0.977 for clone 27 and slope of 1.004 for

381   clone 97; supplementary fig. S7*A-B*), confirming high similarity in the overall GE

382   pattern between the mutant *I572L* and the high-temperature adapted clones.

383   Finally, when we contrasted the GE from the two high-temperature adapted

384   clones at 42ºC (clone 97 *vs* clone 27), there were only 15 differentially expressed

385   genes (*q* <0.001; supplementary fig. S7*C* and fig. S3), beyond the 52 genes that

386   were part of the two large deletions in clone 27 and did not have any

387   measureable GE. Therefore the GE profiles for the two high-temperature

388   adapted clones were nearly identical within the limitations of our experiment,

389   despite their differences in genetic background. These observations confirm that

390   the first step mutation *I572L* contributed to most of the changes in GE during

391   high-temperature adaptation.

392   **Discussion**

393   Two aspects of adaptation that have been largely unexplored are the

394   temporal change of phenotypes during the adaptive process and the genetic

395   changes underlying these changes. Here we have focused on the phenotypic

396   effects of first-step mutations during adaptation of *E.coli* to high temperature

397   (42ºC). A significant finding of our study is that single mutations in RNAP led to

398   altered expression of thousands of genes – most of which were differentially

399   expressed during acclimation to 42ºC – and conferred large fitness advantages.

400   A major phenotypic effect of these first-step mutations was to restore global gene

401   expression back towards the pre-stressed state. The GE profile of the ancestor at

402   42ºC revealed that the cells were physiologically stressed, but the first-step

403   mutations restored GE toward an efficient growth profile. Subsequent mutations

404   also increased fitness but did not substantially change GE.

405   **Heat-stress acclimation involves a balance between energy conservation**

406   **and stress resistance.**

407   The acute response to thermal stress, known as the heat-shock response,

408   occurs in diverse organisms (Richter et al. 2010). In *E.coli,* the heat-shock

409   response is transient (on the order of magnitude of minutes) and is characterized

410   by the induction of stress related proteins mediated by the alternative σ^{32} factor

411   (Nonaka et al. 2006). The σ^{32} regulon encodes: *i)* molecular chaperons (i.e. ClpB,

412   DnaK, DnaJ, IbpB, GrpE, GroEL, GroES) that promote protein folding; *ii)*

413   cytosolic proteases (i.e. ClpP, ClpX) that clear misfolded and aggregated

414   proteins; *iii)* metabolic enzymes; iv) DNA/RNA repair enzymes; *v)* regulatory

415   proteins; *vi)* proteins involved in maintaining cellular integrity; and *vii)* proteins

416   involved in transport and detoxification (Nonaka et al. 2006; Richter et al. 2010).

417   Although the heat-shock response has been studied widely, the expression of

418   heat-shock genes after hours or days of thermal stress (i.e. thermal acclimation)

419   is an open question.

420   We have found that most heat-shock induced genes (Nonaka et al. 2006;

421   Richter et al. 2010) were not differentially expressed during acclimation and

422   were, in fact, down regulated. For example, most of the heat-shock genes

423   encoding chaperones, such as *clpB*, *dnaJ*, *dnaK*, *groEL* and *groES*, were down

424   regulated during the acclimation response (supplementary table S2). However,

19

425   one exception is *spy*, which encodes a periplasmic chaperone and was up

426   regulated during acclimation (log$_2$fold change = 5.0; q < 0.001; supplementary

427   dataset S1). Previous studies in *Saccharomyces cerevisiae* and *E.coli* have

428   reported a decrease in the expression of molecular chaperones after ~15

429   minutes of heat-stress induction (Eisen et al. 1998; Zhao et al. 2005; Jozefczuk

430   et al. 2010; Richter et al. 2010). Therefore, it is possible that the heat-shock

431   genes were initially expressed during our experiment, immediately after transfer

432   to 42⁰C but were subsequently down-regulated.

433   Nevertheless, two previous studies have explored the physiological

434   acclimation of *E.coli* to high temperature and reported up-regulation of heat-

435   shock genes (Gunasekera et al. 2008; Sandberg et al. 2014). The discrepancy

436   among studies may be explained by differences in genetic backgrounds (*E.coli*

437   K-12 strain *vs E.coli* B strain), by differences in temperature (42⁰C vs 43⁰C in

438   Gunasekera et al. 2008), by differences in culture conditions (ours favor

439   microaerophiles) and by differences in sampling. For example, in our study

440   bacterial clones were allowed to acclimate at 42⁰C for one day before being

441   sampled the next day during mid-exponential phase at 42⁰C. Thus, the bacteria

442   spent one complete cycle of growth (lag, exponential and stationary phase) at

443   42⁰C, in addition to the ~ half cycle of growth (lag and half exponential phase) at

444   42⁰C (in total ~ 1.5 days). For other studies, the time that the bacteria spent at

445   high temperature before the RNA extraction is unclear (Sandberg et al. 2014).

446   Similarly, we have carefully documented the growth curve of the ancestor at 42⁰C

447   to help ensure that our RNAseq samples originated from the exponential growth

20

448   phase (supplementary fig. S8). Other studies report to have sampled during mid-

449   exponential phase but have not reported growth curves. Under normal conditions

450   growth curves may not be necessary, but the exponential phase can be very

451   short under extreme stress. It is thus possible that previous studies have

452   reported gene expression on slightly different phases of the growth cycle.

453   Microorganisms often resist stressful conditions by modulating GE to limit

454   growth (López-Maury et al. 2008). As a consequence, genes with growth-related

455   functions, which are energy demanding, are down-regulated, allowing a

456   redistribution of resources and energy to the expression of genes related to

457   stress resistance (López-Maury et al. 2008; Jozefczuk et al. 2010; Jin et al.

458   2012). Surprisingly, we have observed a down-regulation of genes encoding

459   different subunits of RNAP during the acclimation response. Assuming that lower

460   expression also reflects protein abundance, this observation implies that the

461   ancestor contains fewer RNAP molecules when grown at 42ºC than when grown

462   at 37ºC.

463   The reduction of RNAP molecules can have important physiological

464   consequences, given that RNAP is limiting for genome-wide transcription

465   (Ishihama 2000). A reduction in the total number of RNAP would limit the

466   transcription rate and favors our hypothesis that mutated RNAP have higher

467   transcriptional efficiency at 42ºC (Rodríguez-Verdugo et al. 2014). Accordingly,

468   we have examined RNAP efficiency and found a slight trend toward higher rates

469   of transcriptional efficiency for two of the three mutations in codon 572 of *rpoB*.

470   However, none of these differences were supported statistically (supplementary

21

471 fig. S1). A limitation in RNAP molecules could also indirectly affect bacterial

472 growth (Jin et al. 2012). For example, when growth conditions are unfavorable,

473 RNAP molecules are released from the rRNA operons, thereby reducing rRNA

474 synthesis (i.e. reducing growth) so that more RNAP molecules become free and

475 available for genome-wide transcription (Jin et al. 2012). Therefore, the reduction

476 in growth at 42°C may be explained in part by the down-regulation of genes

477 involved in translation and ribosome biogenesis but also by a potential limitation

478 of RNAP.

479   Other down-regulated genes involved in energy demanding processes

480 include genes associated with the biosynthesis of amino acids, nucleotides,

481 ribonucleotides and constituents of the flagella (supplementary table S1). A

482 similar pattern of expression has been observed during acute exposure to

483 thermal stress (Jozefczuk et al. 2010). In addition, metabolomic studies have

484 reported a sharp decline in the levels of nucleotides in *E.coli* cultures exposed to

485 heat stress (Jozefczuk et al. 2010; Ye et al. 2012). Therefore, our study suggests

486 that the pattern of energy conservation mediated through down-regulation of

487 energy demanding processes not only occurs during the acute response to

488 stress but also occurs during the acclimation response.

489   We hypothesize that some of the resources and energy are redistributed to

490 express genes controlled by the sigma S factor or σ^S, the master regulator of the

491 stress response (Battesti et al. 2011). Using previous observations (Weber et al.

492 2005; Keseler et al. 2013), we generated a list of 66 genes expressed under

493 several stress conditions (including high temperature) and under regulation of σ^S.

494   Of these, 62 genes (94%) were significantly up-regulated during acclimation to

495   42ºC (supplementary table S2). σ^S induced genes were related to: *i*) the

496   synthesis molecules responsible to deal with the detrimental effects of stress, *ii*)

497   transport systems, and *iii*) the production of metabolic enzymes, mostly involved

498   in the central energy metabolism (Weber et al. 2005).

499   In conclusion, the ancestor acclimates to thermal stress by launching

500   alternative GE programs (López-Maury et al. 2008). One involves the down-

501   regulation of growth-related pathways, potentially resulting in energy

502   conservation. The other involves the up-regulation of stress genes that are

503   involved in repair and metabolic adjustments to high temperature.

504   **Restoration of the ancestral physiological state is advantageous.**

505   This study has shown that first-step mutations dramatically altered GE from

506   the acclimated state and acted primarily to restore GE toward the pre-stressed

507   state (fig. 3). Previous studies have also reported that small deletions in RNAP

508   change GE of hundreds of genes (Conrad et al. 2010). Taken together, these

509   observations suggest that *rpoB* mutations have important downstream effects

510   that affect complex networks of interacting genes and their products (i.e. global

511   regulators of GE or "hubs"; (Barabasi and Oltvai 2004)). Our study also adds to

512   previous observations that global regulators are a primary target of natural

513   selection in microbial evolution experiments (Philippe et al. 2007; Hindre et al.

514   2012).

515   At least two pieces of evidence suggest that the restoration of GE was

516   advantageous in our experiment, leading to the rapid fixation of *rpoB* mutations in

23

517 evolved populations (Rodríguez-Verdugo et al. 2013). First, restoration in GE

518 may be advantageous because it "reactivates" some of the growth-related

519 functions that were down-regulated during the acclimation response. The GO

520 categories related to growth, such as translation and ribosomal biogenesis, were

521 significantly enriched in the four mutants relative to the acclimated state. The up-

522 regulation of growth-related genes may explain why the mutants have higher

523 maximum growth rates and higher final yields than the ancestor (fig. 3), as well

524 as higher relative fitness values (Rodríguez-Verdugo et al. 2013; Rodríguez-

525 Verdugo et al. 2014)).

526   Second, the general trend of restoration occurred in parallel in all four *rpo*B

527 mutants (*I572F, I572L, I572N* and *I966S*); such phenotypic convergence is

528 commonly interpreted as sign of adaptive evolution (Christin et al. 2010). Not

529 only did we observe convergence in global GE (fig. 4), but we also observed

530 substantial overlap of genes with restored expression (fig. 5). Surprisingly, these

531 observations extend not only to three different mutations in the *rpoB* codon (572)

532 but also to a mutation (*I966S*) that is not in the RNAP active site (fig. 4*D* and fig.

533 5). It is unclear whether the effects of these RNAP mutants are direct or indirect.

534 For a direct effect, the mutated RNAPs would have had new intrinsic binding

535 affinities for promoters throughout the genome. Alternatively because RNAP is a

536 sensor of the physiological state of the cell (Ishihama 2000), it may induce

537 different GE programs. For example, under amino acid starvation RNAP modifies

538 recruitment of sigma factors through its interaction with the alarmone ppGpp, and

539 it consequently induces the stringent response program (Chatterji and Ojha

540   2001). The fact that mutations with similar effects are not limited to the active site

541   of the RNAP favors the possibility of indirect effects, but further experiments are

542   needed to test this hypothesis.

543   It is worth noting that restoration to the non-stress phenotype is not complete,

544   because first-step mutations differ markedly from the 37ºC ancestor in GE

545   dynamics. For example, each of the four *rpoB* mutations differ from the ancestor

546   in the expression of a minimum of 392 genes (*I572L*) and as many as 688 genes

547   (*I572N*). The important point is that first-step mutations restore hundreds of

548   genes toward non-stress expression dynamics but substantial differences remain

549   relative to the wild type phenotype. These differences may explain, in part,

550   differences in growth characteristics between first-step mutations and the

551   ancestor at 37ºC (supplementary fig. S9).

552   In contrast to restorative changes in GE, we find that few novel changes in

553   GE occurred in parallel among the four *rpo*B mutants (supplementary fig. S4).

554   These observations add to a growing literature that suggests restorative changes

555   in GE are a general trend in laboratory adaptation experiments, but novel (and/or

556   reinforced) changes are less common (Fong et al. 2005; Carroll and Marx

557   2013)(Sandberg et al. 2014); although see (Szamecz et al. 2014).

558   **Pleiotropy and compensation**

559   The four *rpoB* mutations (*I572F*, *I572L*, *I572N* and *I966S*) were

560   advantageous in the conditions of our experiment, but they also had pleiotropic

561   effects, such as fitness trade-offs at low temperatures (Rodríguez-Verdugo et al.

562   2014). Previous studies have shown that mutations in global regulators often

563   have maladaptive side effects (Hindre et al. 2012). We hypothesized, therefore,

564   that later adaptive mutations compensate for maladaptive side effects of highly

565   pleiotropic first-step mutations (Hindre et al. 2012).

566   To contrast first-step mutations to the end-products of our adaptation

567   experiment (fig. 1), we compared the *I572L rpoB* mutation to two clones (97 and

568   27) that included the *I572L* mutation. Each of the two clones differed from the

569   single-mutant background by ~10 mutations; these mutations included both point

570   mutations and more complex genetic changes, such as a large deletion. Based

571   on these contrasts, we have found that the first-step mutation *I572L* contributed

572   most of the GE variation during thermal stress adaptation (987 differentially

573   expressed genes), while later mutations contributed fewer changes in GE (63

574   and 16 differentially expressed genes in clones 27 and 97, respectively; fig. 6

575   and supplementary fig. S5). Interestingly, these few changes in GE may

576   contribute to significant changes in growth parameters (table 3). Therefore, the

577   number of differentially expressed genes was not proportional to fitness gains.

578   This "disconnect" between the number of genes differentially expressed

579   and the magnitude of the fitness advantage might be caused by compensatory

580   changes to the pleiotropic side effects of mutation *I572L*. Under this model, we

581   presume that some of the hundred of differentially expressed genes in *I572L*

582   have beneficial effects, while others have deleterious effects, netting an overall

583   beneficial change in GE. If true, it is likely that subsequent mutations change the

584   expression of only few genes, but most of these changes are beneficial. For

585   example, the large deletion in clone 27 contained genes involved in iron

586   acquisition (*fep* and *ent* operons; Crosa and Walsh 2002), and copper and silver

587   efflux system (*cus* operons; Long et al. 2010). Costly, non-functional pathways

588   are often "shut-down" in order to save energy that would be otherwise used to

589   produce unnecessary proteins and metabolisms (Cooper et al. 2001; Lewis et al.

590   2010). Therefore the large deletion of 71 kb in length may be an energetic

591   benefit, explaining its occurrence in 35 high-temperature adapted clones

592   (Tenaillon et al. 2012).

593   Clone 97, which lacked the large deletions of clone 27, displayed fewer

594   changes in gene expression than clone 27. In this clone, one of the significantly

595   down-regulated genes ($q < 0.001$) was the *rmf* gene, which encodes the

596   ribosome modulation factor (RMF). RMF has been associated with decreased

597   translation activity and is expressed during slow growth conditions, such as

598   stationary phase (Polikanov et al. 2012). Therefore, the down-regulation of *rmf*,

599   occurring in parallel in clone 27 and 97, might increase protein synthesis and

600   thus growth.

601   Finally, we observed that the genes involved in flagellum-dependent cell

602   motility (e.g. *flg* genes) were highly down-regulated during the acclimation

603   response ($\log_2$fold < -3 and $q < 0.001$) but were again up-regulated in the single

604   mutants *I572F* and *I572L* ($q < 0.001$; supplementary dataset S1). In the

605   conditions of our evolution experiment — a well-mixed environment lacking

606   physical structure — motility seems unnecessary. Given that the biosynthesis of

607   flagella is costly (Soutourina and Bertin 2003), reducing the expression *flg* genes

608   might be beneficial (Cooper et al. 2003; Fong et al. 2005). We therefore posit that

27

609   the restoration in GE of *flg* genes might be costly and have deleterious effects.

610   Interestingly, some of the *flg* genes were down-regulated ($q < 0.05$) in clone 97

611   when compared to the first-step mutant *I572L* (supplementary table S6).

612   Therefore, later adaptive mutations might contribute to the fine-tuning of GE by

613   compensating the side effects of restoration. We note, however, that we have not

614   yet identified the mutation that causes the down-regulation of *flg* genes in clone

615   97. That being said, the up-regulation of flagellar genes after restorative shifts in

616   GE and the fixation of subsequent mutations that down-regulate them has been

617   observed previously (Sandberg et al. 2014).

618   **Concluding remarks**

619   Mutations in global regulators of GE are observed recurrently in laboratory

620   evolution experiments (Applebee et al. 2008; Goodarzi et al. 2009; Kishimoto et

621   al. 2010). It is not always clear if these mutations represent the first step of an

622   adaptive walk or later step, but at least in the case of *rpoB* and *rpoC* mutations, it

623   seems they are often first-step mutations (Herring et al. 2006). Therefore, the

624   pattern that we have observed in our study may not be specific to our system but

625   instead a more general phenomenon. Based on our results, we propose a

626   general, two-step adaptive process. First, a mutation affecting global

627   transcriptional regulator appears in the population and changes global GE. This

628   expression change is mostly restorative, so that the stressed state moves toward

629   a pre-stressed (ancestral) state. These changes in GE confer a high advantage,

630   promoting the rapid fixation of the mutation in the population. Once the cell

631   recovers its "normal" homeostatic state, other mutations accumulate and

28

632   contribute to novel functions (fine-tuning of adaptive traits) or/and compensate for

633   the side effects of the first-step pleiotropic mutation. Future directions would be to

634   confirm this pattern by performing time course studies of GE (including heat

635   shock and acclimation response, as well as all the intermediate steps of

636   adaptation) coupled with genomic data.

637   ## Materials and Methods

638   ### Growth conditions

639   Unless otherwise noted, the culture conditions used for the physiological

640   assays (growth curves, transcription efficiency and RNA-seq assays), were the

641   same used during the high-temperature evolution experiment (Tenaillon et al.

642   2012). Briefly, strains were revived in LB and incubated at 37ºC with constant

643   shaking (120 rpm). Overnight cultures were diluted 10^4-fold into 10 ml Davis

644   minimal medium supplemented with 25 µg/ml glucose (DM25) and incubated 1 d

645   at 37ºC to allow the strains to acclimate to the culture conditions. The following

646   day, we transferred 100 µl of the overnight culture in 9.9 ml of fresh DM25 (100-

647   fold dilution) and we incubated them at 42ºC for one day to allow the strains to

648   acclimate to high temperature.

649   ### Growth curves

650   Acclimated strains were grown at the assay temperature (either 37ºC or

651   42ºC) and the densities were measured approximately every hour during the lag

652   phase and every ~30 minutes during the exponential phase. Population densities

653   were measured using an electronic particle counter (Coulter Counter model

654   Multisizer 3 equipped with a 30-µm-diameter aperture tube). To measure density,

655   50 µl of culture was diluted in 9.9 mL Isoton II diluent (Beckman Coulter), and 50

656   µl of the resulting dilution was counted electronically. To estimate the maximum

657   growth rate, μ_{max}, we fitted a linear regression to the natural logarithm of the cell

658   density over time during the exponential phase (i.e. linear part) using the *lm*

659   function in R version 3.0.2 (Team 2013, supplementary fig. S8). Three estimates

660   of μ_{max} were obtained for each genotype. The duration of the lag phase, τ, was

661   defined as the time elapsed before the beginning of the exponential phase. The

662   final yield was estimated from the cell counts at the end of the exponential phase.

663   **RNA extraction and preparation for sequencing**

664   To investigate the GE profile prior to thermal stress, we grew the ancestor

665   (REL1206), previously acclimated to the growth conditions, at 37ºC until it

666   reached mid-exponential phase (three replicates). To investigate the GE profile

667   during acclimation, we grew the ancestor at 42º, previously acclimated to 42ºC

668   (two replicates), until mid-exponential phase. The remaining strains (single first-

669   step mutants and high-temperature adapted clones) were grown at 42ºC until

670   mid-exponential phase, with two replicates for each.

671   Briefly, bacterial cultures were grown in DM25 medium until they reached

672   mid-exponential phase, which we determined by electronic counts. 80 ml of

673   culture was filtered through a 0.2 µm cellulose membrane (Life Science,

674   Germany). Cells, concentrated in the filter, were stabilized with Qiagen RNA-

675   protect Bacteria Reagent and pellet for storage at -80ºC prior to RNA extraction.

676   Total RNA was extracted using RNeasy Mini Kit (Qiagen). Total RNA was DNase

30

677 treated using Turbo DNA-free kit (Ambion) and rRNA was depleted using the

678 Ribo-Zero rRNA Removal kit for Gram-Negative Bacteria (Epicentre

679 Biotechnologies, Medion, WI, USA). cDNA library was constructed using TruSeq

680 RNA v2 kit (Illumina, San Diego, CA, USA). Libraries were multiplexed 8-fold and

681 sequenced on an Illumina Hiseq 2000 platform. 100-bp single-end reads were

682 generated.

683

684 **mRNAseq data analyses**

685 Reads were mapped to the *E.coli* B REL606 genome reference

686 (CP000819.1) using *bwa* 0.6.2 (Li and Durbin 2009), using default parameters

687 (http://bio-bwa.sourceforge.net/bwa.shtml). Only unique, perfectly matching

688 reads to the 4204 annotated coding regions were retained for further analyses

689 (supplementary table S7). Differential expression analysis was performed using

690 the DESeq R package (Anders and Huber 2010). We used the *P*-values adjusted

691 by the Benjamini and Hochberg approach (*q* values), which controls for false

692 discovery rate. Genes with *q* less than 0.001 were considered significantly

693 differentially expressed. For the GO term enrichment analyses, we used the

694 Enrichment analysis tool from http://geneontology.org/page/go-enrichment-

695 analysis.

696 Differentially expressed genes (DEG) were classified in one of the four

697 categories (restored, reinforced, unrestored, and novel) based on the contrasts

698 for which they were significant and the direction and value of their fold change

699 (supplementary table S3). We also examined overall trends in GE shifts by

700   plotting two ratios that share an underlying parameter (fig. 4 and supplementary

701   fig. S6), which creates an inherent negative bias for correlation values. We

702   therefore tested the strength of correlation by randomization; we choice the three

703   independent parameters at random from the dataset, constructed ratios for the

704   observed number of genes, estimated a correlation coefficient, and repeated the

705   process 1000 times to achieve a distribution of correlation values that reflect the

706   inherent bias.

707   **Construction of the fluorescently labeled strains for the transcription**

708   **efficiency assay**

709   We generated a ~4 kb-long linear DNA fragment carrying the *CAT*, *tetR*

710   and *YFP* genes from an *E.coli* strain carrying the *CAT:tetR:YFP* genomic

711   cassette (Fehér et al. 2012) kindly provided by Csaba Pál (Biological Research

712   Center, Szeged, Hungary). We amplified the *CAT;tetR;YFP* cassette by PCR

713   using the primers ARV19 and ARV20 (supplementary table S8) and *Pfu* DNA

714   polymerase (Promega). The purified PCR product was integrated into the

715   ancestral strain REL1206 carrying the pKD46 recombineering plasmid as

716   previously described by Datsenko and Wanner (Datsenko and Wanner 2000).

717   Briefly, the ancestral strain carrying the pKD46 plasmid was grown overnight at

718   30ºC in 5 ml of LB with 100 µg/ml of ampicillin. The overnight culture was 100

719   fold-diluted in 100 ml of LB with ampicillin and 1 mM L-arabinose (Sigma) and

720   grown at 30ºC to an OD_{600} of 0.6. Electrocompetent cells were made by washing

721   the culture 5 times with ice-cold water. ~200 ng of linear DNA was electroporated

722   into 25 µl of cells. After electroporation, 1 ml of LB was added, and the cells were

723   incubated at 30ºC for 2 h with shaking, then plated 100 µl on LB agar plates with

724   chloramphenicol (20 µg/ml). We selected a single colony and purified it in LB

725   agar plate containing chloramphenicol. Correct integration was verified by PCR

726   using primers ARV34 and ARV35 (supplementary table S8).

727   **Transcription efficiency assay**

728   To measure the transcription efficiency of the ancestor and the mutants

729   we performed a quantitative reverse transcriptase RT-PCR assay (Reynolds

730   2000; Brandis et al. 2012). In brief, we grew the fluorescently labeled strains in

731   the same culture conditions previously described except that we supplemented

732   the DM medium with 100 µg/ml glucose (DM100). We confirmed the advantage

733   of these mutations to high temperature despite the higher amount of glucose

734   based on growth curves of the ancestor and the mutants at 42ºC in DM100.

735   Acclimated cultures were grown at 42.2ºC until they reached mid-exponential

736   phase. We took 1 ml of uninduced cells (sample T_0) and stabilized them in

737   RNAprotect Bacteria Reagent (Qiagen). Immediately after, we induced cells by

738   adding 10µl of anhydotetracycline (66 µg/ml) to the medium. Samples of 1 ml of

739   culture were taken at 1, 2, 3 and 4 min after induction (samples T_1, T_2, T_3 and T_4)

740   and stabilized with RNA protect Bacteria Reagent and pelleted for storage at -

741   80ºC prior to RNA extraction. Total RNA was prepared using RNeasy Mini Kit

742   (Qiagen). RNA was DNase treated using Turbo DNA-free kit (Ambion). We used

743   300 ng of DNA-free RNA to produce cDNA with the High Capacity cDNA

744   Reverse Transcription Kit (Applied Biosystems). From each reverse transcribed

745   product we quantified the abundance of cDNA of the gene reference *gst* (Gst,

746   glutathione transferase; Pfaffl 2001) and the target gene YFP (Yellow fluorescent

747   protein) using Fast SYBR Green Master Mix (Applied Biosystems) quantified on

748   a Stratagene MX3005P QPCR System (Agilent Technologies). For each PCR

749   reaction we used 0.625 µM forward primer (ARV48 for *gst* or ARV50 for YFP,

750   supplementary table S8), 0.625 µM reverse primer (ARV49 for *gst* or ARV51 for

751   YFP, supplementary table S8), 3 µl CDNA template, 4.5 µl RNase-free water and

752   10 µl of Fast SYBR Green Master Mix, to have a final reaction volume of 20 µl.

753   To control for the intra-assay variation (repeatability), we prepared three

754   replicates of each reaction. The PCR thermal cycling conditions were 95ºC for 20

755   sec followed by 40 cycles of 95ºC for 3 sec and 60ºC for 30 sec.

756   The efficiency of the amplifications for each pair of primers was determined

757   from a standard curve using the formula $E = 10^{[-1/s]}$, where *s* is the slope of the

758   standard curve (Pfaffl 2001). To calculate the *Relative Expression Ratio* (i.e. the

759   relative change in GE of the target gene YFP normalized to the reference gene

760   *gst* and relative to the uninduced control sample T_0), we used the mathematical

761   model for relative quantification in real-time RT-PCR developed by Pfaffl, 2001

762   (Pfaffl 2001; Brandis et al. 2012). Transcription efficiency was calculated as the

763   slope of the fitted linear regression between the *Relative Expression Ratio*

764   against time, based on three replicates.

765   **Acknowledgments**

References

775 Anders S, Huber W. 2010. Differential expression analysis for sequence count

776   data. *Genome Biol* 11:R106.

777 Applebee MK, Herrgard MJ, Palsson BO. 2008. Impact of individual mutations on

778   increased fitness in adaptively evolved strains of *Escherichia coli. J*

779   *Bacteriol* 190:5087–5094.

780 Barabasi A-L, Oltvai ZN. 2004. Network biology: understanding the cell's

781   functional organization. *Nat Rev Genet* 5:101–113.

782 Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress

783   response in *Escherichia coli. Annu Rev Microbiol* 65:189–213.

784 Blount ZD, Barrick JE, Davidson CJ, Lenski RE. 2012. Genomic analysis of a key

785   innovation in an experimental *Escherichia coli* population. *Nature* 489:513–

786   518.

787 Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory

788   mutations in rifampicin-resistant RNA polymerase. *Mol Microbiol* 85:142–

789   151.

790   Carroll SM, Marx CJ. 2013. Evolution after introduction of a novel metabolic

791   pathway consistently leads to restoration of wild-type physiology. *PLoS*

792   *Genet.* 9:e1003427.

793   Chatterji D, Ojha AK. 2001. Revisiting the stringent response, ppGpp and

794   starvation signaling. *Curr Opin Microbiol* 4:160–165.

795   Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ. 2011. Diminishing returns

796   epistasis among beneficial mutations decelerates adaptation. *Science*

797   332:1190–1192.

798   Christin P-A, Weinreich DM, Besnard G. 2010. Causes and evolutionary

799   significance of genetic convergence. *Trends Genet* 26:400–405.

800   Conrad TM, Frazier M, Joyce AR, Cho BK, Knight EM, Lewis NE, Landick R,

801   Palsson BO. 2010. RNA polymerase mutants found through adaptive

802   evolution reprogram *Escherichia coli* for optimal growth in minimal media.

803   *Proc Natl Acad Sci USA* 107:20500–20505.

804   Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression

805   after 20,000 generations of evolution in *Escherichia coli. Proc Natl Acad Sci*

806   *USA.* 100:1072–1077.

807   Cooper VS, Schneider D, Blot M, Lenski RE. 2001. Mechanisms causing rapid

808   and parallel losses of ribose catabolism in evolving populations of

809   *Escherichia coli* B. *J Bacteriol* 183:2834–2841.

810   Crosa JH, Walsh CT. 2002. Genetics and assembly line enzymology of

811   siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* 66:223–249.

812   Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in

813   Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA.*

814   97:6640–6645.

815   Ederth J, Mooney RA, Isaksson LA, Landick R. 2006. Functional interplay

816   between the jaw domain of bacterial RNA polymerase and allele-specific

817   residues in the product RNA-binding pocket. *J Mol Biol* 356:1163–1179.

818   Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and

819   display of genome-wide expression patterns. *Proc Natl Acad Sci USA.*

820   95:14863–14868.

821   Fehér T, Bogos B, Méhi O, Fekete G, Csörgő B, Kovács K, Pósfai G, Papp B,

822   Hurst LD, Pál C. 2012. Competition between transposable elements and

823   mutator genes in bacteria. *Mol Biol Evol.* 29:3153–3159.

824   Fisher RA. 1930. The genetical theory of natural selection. Univ. Press.

825   Fong SS, Joyce AR, Palsson BØ. 2005. Parallel adaptive evolution cultures of

826   Escherichia coli lead to convergent growth phenotypes with different gene

827   expression states. *Genome Res* 15:1365–1372.

828   Goodarzi H, Hottes AK, Tavazoie S. 2009. Global discovery of adaptive

829   mutations. *Nature Methods* 6:581–583.

830   Gunasekera TS, Csonka LN, Paliy O. 2008. Genome-wide transcriptional

831   responses of Escherichia coli K-12 to continuous osmotic and heat stresses.

832   *J Bacteriol* 190:3712–3720.

833   Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert

834   TJ, Blattner FR, Van den Boom D, Cantor CR. 2006. Comparative genome

835   sequencing of Escherichia coli allows observation of bacterial evolution on a

836   laboratory timescale. *Nat Genet* 38:1406–1412.

837   Hindre T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial

838   adaptation through in vivo and in silico experimental evolution. *Nat Rev*

839   *Microbiol.* 10:352–365.

840   Ishihama A. 2000. Functional modulation of *Escherichia coli* RNA polymerase.

841   *Annu Rev Microbiol.* 54:499–518.

842   Jin DJ, Walter WA, Gross CA. 1988. Characterization of the termination

843   phenotype of rifampicin-resistant mutants. *J Mol Biol* 202:245–253.

844   Jin DJ, Burgess RR, Richardson JP, Gross CA. 1992. Termination efficiency at

845   rho-dependent terminators depends on kinetic coupling between RNA

846   polymerase and rho. *Proc Natl Acad Sci USA.* 89:1453–1457.

847   Jin DJ, Cagliero C, Zhou YN. 2012. Growth rate regulation in Escherichia coli.

848   *FEMS Microbiol Rev* 36:269–287.

849   Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros☐Inostroza A,

850   Steinhauser D, Selbig J, Willmitzer L. 2010. Metabolomic and transcriptomic

851   stress response of *Escherichia coli*. *Mol Syst Biol* 6:364.

852   Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S,

853   Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M.

854   2013. EcoCyc: fusing model organism databases with systems biology.

855   *Nucleic Acids Res* 41:D605–D612.

856   Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative epistasis

857   between beneficial mutations in an evolving bacterial population. *Science*

858   332:1193–1196.

859   Kishimoto T, Iijima L, Tatsumi M, Ono N, Oyake A, Hashimoto T, Matsuo M,

860   Okubo M, Suzuki S, Mori K. 2010. Transition from positive to neutral in

861   mutation fixation along with continuing rising fitness in thermal adaptive

862   evolution. *PLoS genetics* 6:e1001164.

863   Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Global epistasis makes

864   adaptation predictable despite sequence-level stochasticity. *Science*

865   344:1519–1522.

866   Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD,

867   Adkins JN, Schramm G, Purvine SO, Lopez☐Ferrer D. 2010. Omic data

868   from evolved *E. coli* are consistent with computed optimal growth from

869   genome☐scale models. *Mol Syst Biol* 6:390.

870   Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–

871   Wheeler transform. *Bioinformatics* 25:1754–1760.

872   Long F, Su C-C, Zimmermann MT, Boyken SE, Rajashankar KR, Jernigan RL,

873   Edward WY. 2010. Crystal structures of the CusA efflux pump suggest

874   methionine-mediated metal transport. *Nature* 467:484–488.

875   López-Maury L, Marguerat S, Bähler J. 2008. Tuning gene expression to

876   changing environments: from rapid responses to evolutionary adaptation.

877   *Nat Rev Genet* 9:583–593.

878   Mejia YX, Mao HB, Forde NR, Bustamante C. 2008. Thermal probing of *E-coli*

879   RNA polymerase off-pathway mechanisms. *J Mol Biol* 382:628–637.

880   Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA. 2006. Regulon

881   and promoter analysis of the *E. coli* heat-shock factor, σ32, reveals a

882   multifaceted cellular response to heat stress. *Genes Dev* 20:1776–1789.

883   Opalka N, Brown J, Lane WJ, Twist K-AF, Landick R, Asturias FJ, Darst SA.

884   2010. Complete structural model of *Escherichia coli* RNA polymerase from a

885   hybrid approach. *PLoS biology* 8:e1000483.

886   Orr HA. 2005. The genetic theory of adaptation: A brief history. *Nat Rev Genet*

887   6:119–127.

888   Pfaffl MW. 2001. A new mathematical model for relative quantification in real-

889   time RT–PCR. *Nucleic Acids Res* 29:e45–e45.

890   Philippe N, Crozat E, Lenski RE, Schneider D. 2007. Evolution of global

891   regulatory networks during a long☐term experiment with *Escherichia coli.*

892   *BioEssays* 29:846–860.

893   Polikanov YS, Blaha GM, Steitz TA. 2012. How hibernation factors RMF, HPF,

894   and YfiA turn off protein synthesis. *Science* 336:915–918.

895   Quandt EM, Deatherage DE, Ellington AD, Georgiou G, Barrick JE. 2014.

896   Recursive genomewide recombination and sequencing reveals a key

897   refinement step in the evolution of a metabolic innovation in *Escherichia*

898   *coli. Proc Natl Acad Sci USA* 111:2217–2222.

899   Team RC. 2013. *R: A language and environment for statistical computing.*

900   Reynolds MG. 2000. Compensatory evolution in rifampin-resistant Escherichia

901   coli. *Genetics* 156:1471–1481.

902   Richter K, Haslbeck M, Buchner J. 2010. The heat shock response: life on the

903   verge of death. *Mol Cell* 40:253–266.

904   Riehle MM, Bennett AF, Long AD. 2001. Genetic architecture of thermal

905   adaptation in *Escherichia coli. Proc Natl Acad Sci USA* 98:525–530.

906   Rodríguez-Verdugo A, Carrillo-Cisneros D, González-González A, Gaut BS,

907   Bennett AF. 2014. Different tradeoffs result from alternate genetic

908   adaptations to a common environment. *Proc Natl Acad Sci USA*

909   111:12121–12126.

910   Rodríguez-Verdugo A, Gaut BS, Tenaillon O. 2013. Evolution of *Escherichia coli*

911   rifampicin resistance in an antibiotic-free environment during thermal stress.

912   *BMC Evol Biol* 13:50.

913   Ryals J, Little R, Bremer H. 1982. Temperature dependence of RNA synthesis in

914   *Escherichia coli. J Bacteriol* 151:879–887.

915   Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ,

916   Palsson BO, Sommer M, Feist AM. 2014. Evolution of *Escherichia coli* to

917   42° C and subsequent genetic engineering reveals adaptive mechanisms

918   and novel mutations. *Mol Biol Evol* 31:2647–2662.

919   Soutourina OA, Bertin PN. 2003. Regulation cascade of flagellar expression in

920   Gram□negative bacteria. *FEMS Microbiol Rev* 27:505–523.

921   Szamecz B, Boross G, Kalapis D, Kovács K, Fekete G, Farkas Z, Lázár V,

922   Hrtyan M, Kemmeren P, Koerkamp MJAG. 2014. The genomic landscape of

923   compensatory evolution. *PLoS Biol* 12:e1001935.

924   Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD,

925   Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science*

926   335:457–461.

927   Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R. 2005. Genome-wide

928   analysis of the general stress response network in *Escherichia coli*: σS-

929   dependent genes, promoters, and sigma factor selectivity. *J Bacteriol*
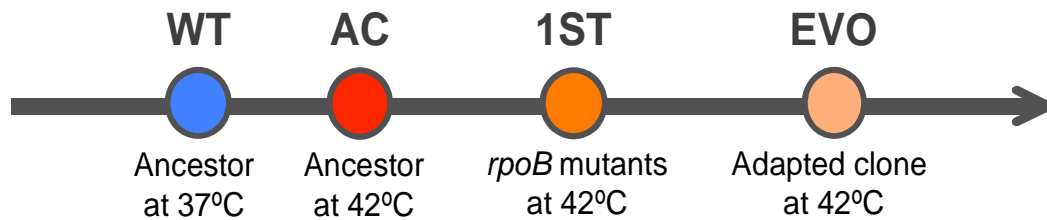
930   187:1591–1603.

931   Ye Y, Zhang L, Hao F, Zhang J, Wang Y, Tang H. 2012. Global metabolomic

932   responses of *Escherichia coli* to heat stress. *J Proteome Res* 11:2559–

933   2566.

934   Zhao K, Liu M, Burgess RR. 2005. The global transcriptional response of

935   *Escherichia coli* to induced σ32 protein involves σ32 regulon activation

936   followed by inactivation and degradation of σ32 in vivo. *J Biol Chem*

937   280:17758–17768.

938   Zhou YN, Lubkowska L, Hui M, Chen S, Strathern J, Jin DJ, Kashlev M. 2013.

939   Isolation and characterization of RNA Polymerase rpoB mutations that alter

940   transcription slippage during elongation in *Escherichia coli*. *J Biol Chem*

941   288:2700–2710.

942

943 **Figures**

944

WT	AC	1ST	EVO
Ancestor at 37ºC	Ancestor at 42ºC	*rpoB* mutants at 42ºC	Adapted clone at 42ºC

945 **Fig 1. Thermal stress adaptation timeline.** We characterized four states during

946 thermal stress adaptation: a wild-type (WT) state, which corresponds to the

947 ancestor grown at 37ºC; an acclimatized (AC) state, which corresponds to the

948 ancestor acclimatized at 42ºC; a first-step adaptive (1ST) state, which

949 corresponds to the *rpoB* single mutants; and an evolved (EVO) state, which

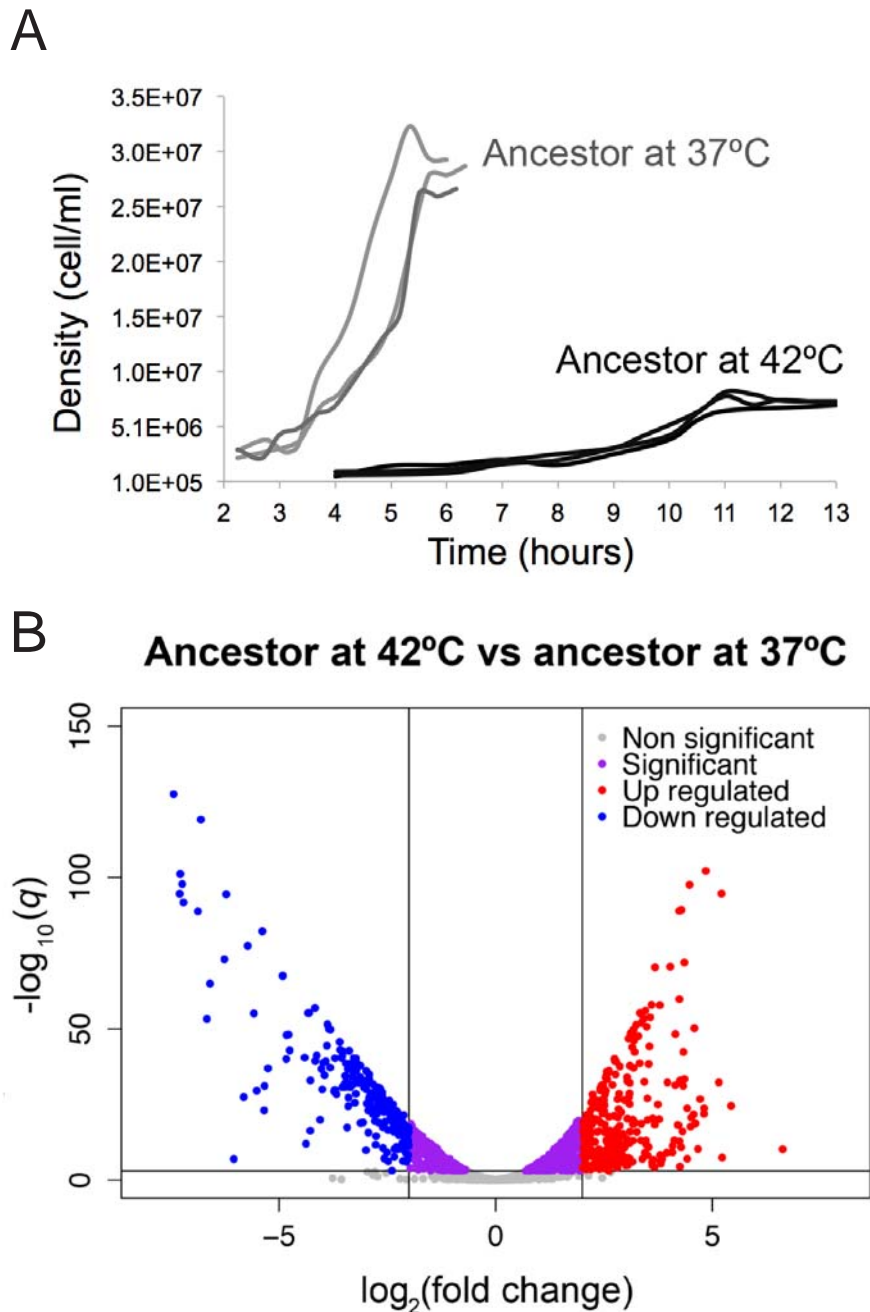950 corresponds to the high-temperature adapted clones.

951

Fig 2. Phenotypic characterization of the ancestor at 37⁰C and 42⁰C. (A)

Growth curves of the ancestor grown at 37⁰C and 42⁰C (three replicates at each

temperature). (B) Volcano plot showing the global differential expression of

genes (represented as dots) between the ancestor grown at 42⁰C compare to the

ancestor grown at 37⁰C. Colors represent status with respect to 2-fold expression

958   difference, represented by two vertical lines, and a significance at $q = 0.001$,
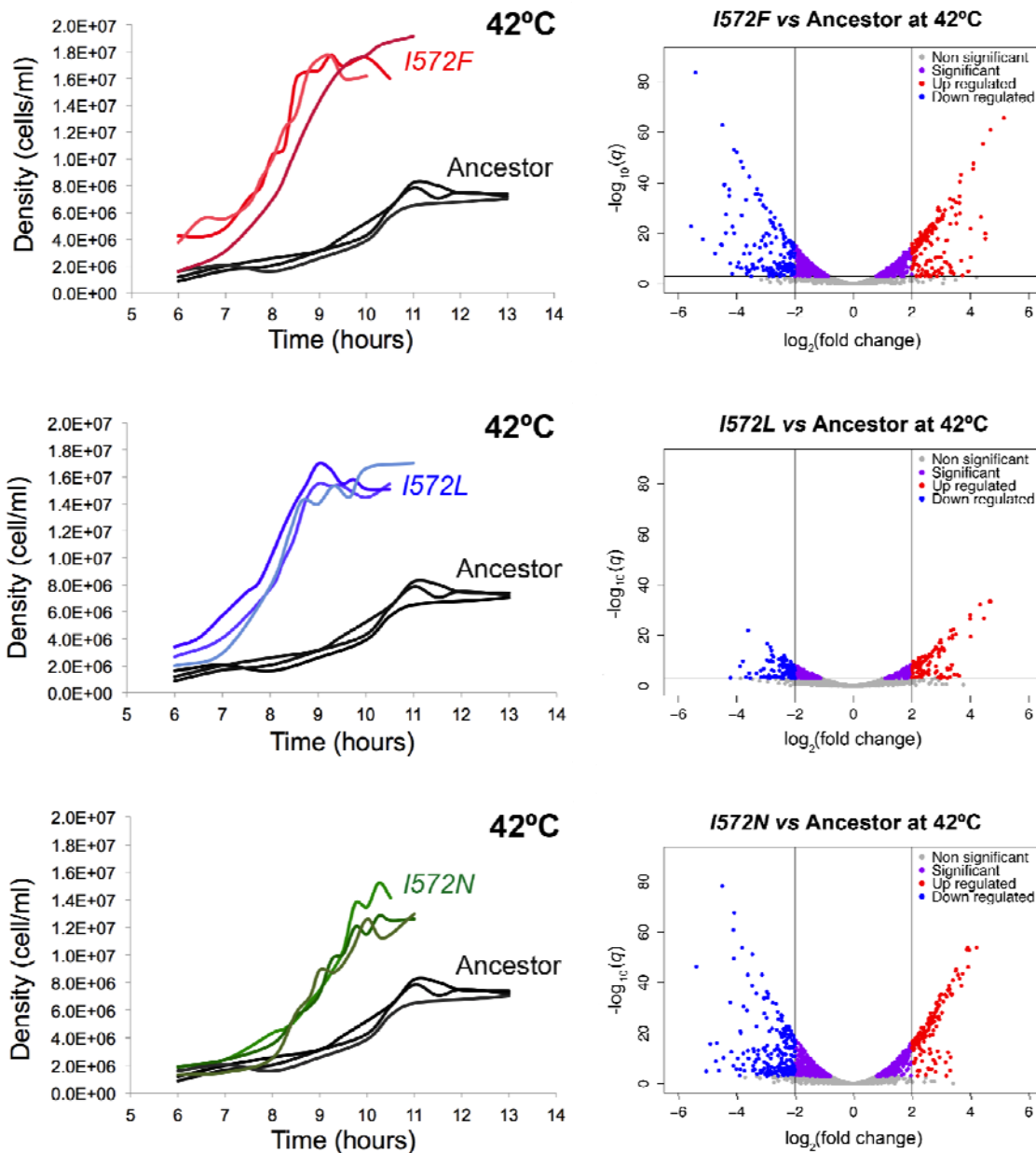
959   represented by an horizontal line.

960

961

962   **Fig 3. Phenotypic characterization of the mutants compared to the ancestor**

963   **at 42°C.** The left side of the figure shows the growth curves of the ancestor and

45

964   the mutants grown at 42ºC (three replicates per each genotype). The right side of

965   the figure shows the volcano plots showing the global differential expression of

966   genes (represented as dots) between the mutants grown at 42ºC and the

967   ancestor grown at 42ºC. Colors represent status with respect to 2-fold expression

968   difference, represented by two vertical lines, and a significance at $q = 0.001$,

969   represented by an horizontal line.
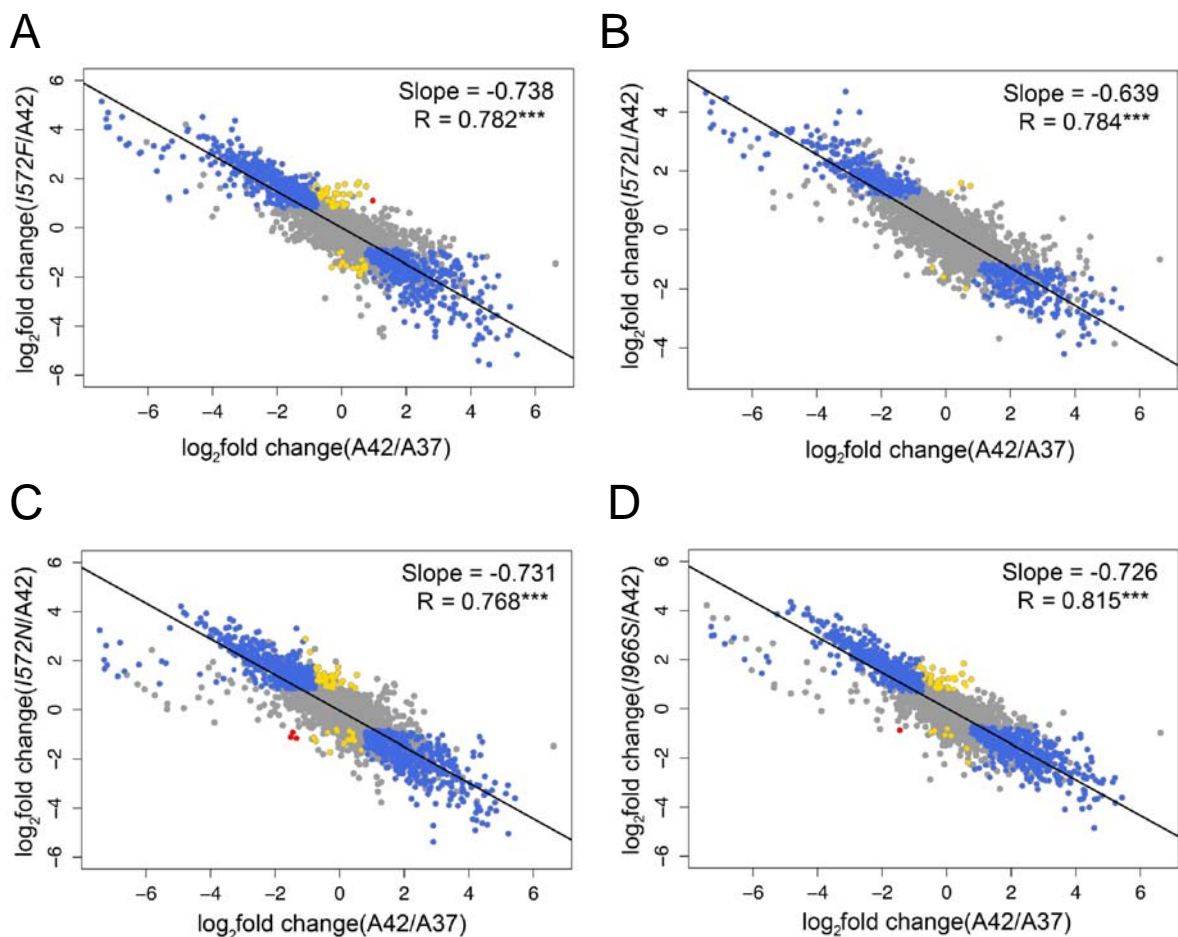
970

971

972   **Fig 4. Global changes in GE during the acclimation and adaptive response.**

973   In all the graphs the *x*-axis represents the acclimation changes (ancestor grown

974   at 42ºC *vs* ancestor grown at 37ºC). The *y*-axis represents the changes at

975   different steps of the adaptive walk: first-step adaptive mutations (*I572F, I572L,*

976   *I572N*) and a single mutant (*I966S*) *vs* ancestor grown at 42ºC. Changes in

977   expression were categorized and colored as follows: restored (blue), reinforced

978   (red) and novel (yellow). Both unrestored and uninformative genes are colored in

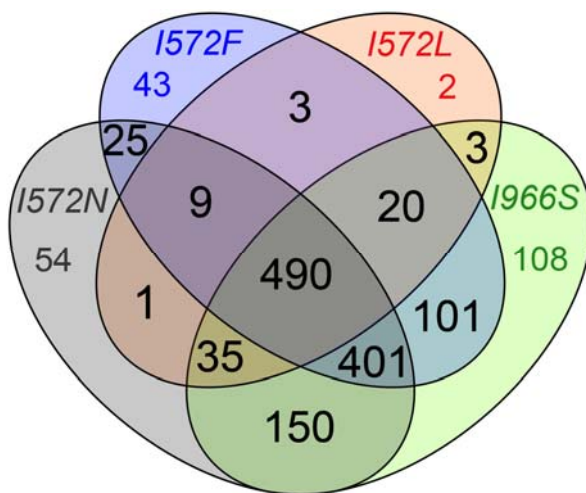979   grey. The black line represents the linear regression fitted to the dots in each

980   graph.

981

982

983   **Fig 5. Convergence of genes with restored expression in single mutants.**

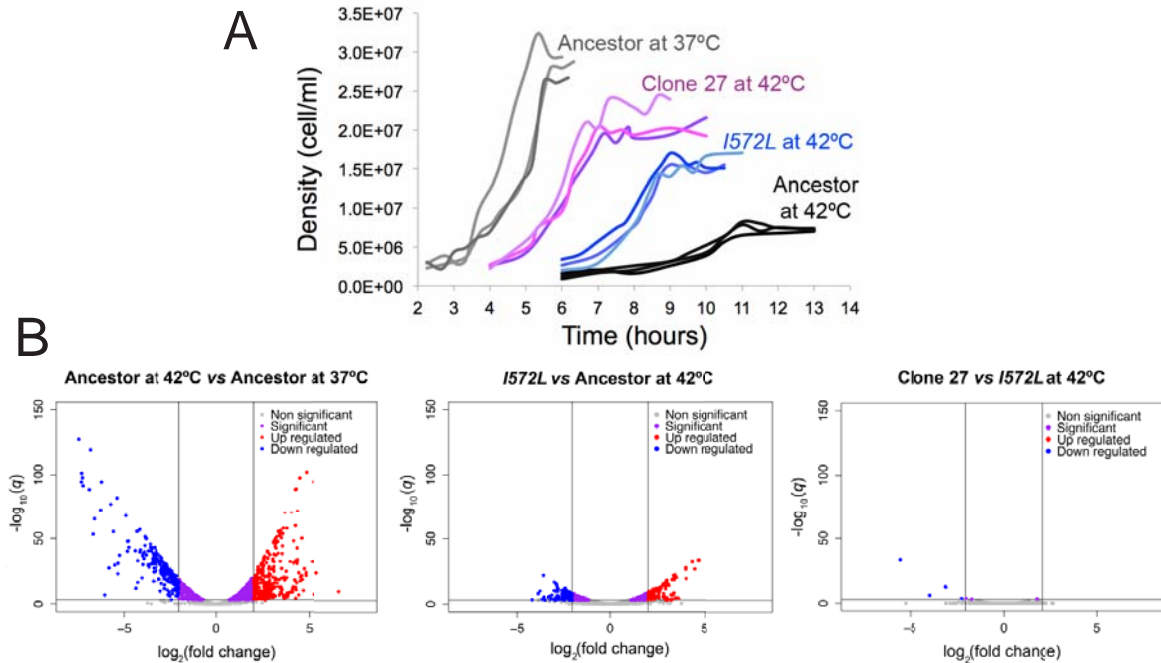984   Number of shared genes with restored expression.

985

986

987 **Fig 6. Phenotypic changes during thermal stress adaptation of clone 27.** (A)

988 Growth improvements during the heat stress adaptive walk. (B) Changes in GE

989 during the acclimation and adaptive response.

990

991 **Tables**

992 **Table 1. Growth parameters of the first-step mutants and the ancestor.**

Strain	μ_{max}[1] 37ºC	μ_{max}[1] 42ºC	P[2]	τ[3] 37ºC	τ[3] 42ºC	P[2]	Yield[4] 37ºC	Yield[4] 42ºC	P[2]
Ancestor	0.756	0.418	**0.002**	3.4	6.7	**0.035**	$2.79*10^7$	$7.21*10^6$	**0.001**
SE	(0.018)	(0.033)		(0.1)	(0.7)		$(8.15*10^5)$	$(1.46*10^5)$	
I572F	0.876	0.734	0.233	3.4	6.8	**0.004**	$2.44*10^7$	$1.72*10^7$	**0.008**
SE	(0.085)	(0.021)		(0.4)	(0.4)		$(1.08*10^6)$	$(6.44*10^5)$	
P[5]	0.293	**0.002**		1.000	0.846		0.063	**0.003**	
I572L	0.898	0.809	0.526	3.2	6.8	**0.005**	$2.72*10^7$	$1.57*10^7$	**0.003**
SE	(0.090)	(0.092)		(0.4)	(0.2)		$(9.86*10^5)$	$(4.24*10^5)$	
P[5]	0.252	**0.039**		0.713	0.829		0.615	**0.001**	
I572N	1.124	0.622	0.140	3.4	8.5	**0.001**	$3.05*10^7$	$1.28*10^7$	**0.005**
SE	(0.219)	(0.067)		(0.2)	(0.0)		$(1.76*10^6)$	$(5.95*10^5)$	
P[5]	0.235	0.074		0.908	0.111		0.274	**0.008**	

993 [1] Maximum growth rate with standard error in parenthesis.

48

994   [2] Significance value from a two-sample t-test. The null hypothesis is that the mean at 37ºC and

995   the mean at 42ºC are equal.

996   [3] Duration of the lag phase with standard error in parenthesis.

997   [4] Final yield (cells/ml) with standard error in parenthesis.

998   [5] Significance value from a two-sample t-test. The null hypothesis is that the mean of the ancestor

999   and the mean of the mutant are equal. Numbers in bold correspond to *P*-values <0.05.

1000

1001   **Table 2. Classification of the genes into four patterns of expression**

1002   **change.**

Category	I572F	I572L	I572N	I966S
1) Restored	1092	563	1165	1308
2) Reinforced	1	0	3	1
3) Unrestored	644	1174	569	428
4) Novel	59	6	53	39

1003

1004   **Table 3. Growth parameters of the high-temperature adapted clones 27 and**

1005   **97 at 42ºC compare to the mutant *I572L* and the ancestor at 37ºC and 42ºC**

1006   **(values in table 1).**

Strain	μ_{max} [1]	τ [2]	Yield [3]
27	0.769 (0.041)	4.7 (0.3)	$2.09*10^7$ ($1.22*10^6$)
97	0.794 (0.063)	4.7 (0.3)	$1.87*10^7$ ($2.20*10^6$)
Comparison	**Significance** [4]		
27 *vs* ancestor at 42ºC	**0.003**	0.076	**0.007**
97 *vs* ancestor at 42ºC	**0.013**	0.076	**0.034**
27 *vs* I572L at 42ºC	0.716	**0.011**	**0.039**
97 *vs* I572L at 42ºC	0.651	**0.011**	0.302
27 *vs* ancestor at 37ºC	0.801	**0.049**	**0.011**
97 *vs* ancestor at 37ºC	0.613	**0.049**	**0.040**

1007   [1] Maximum growth rate with standard error in parenthesis.

1008   [2] Duration of the lag phase with standard error in parenthesis.

1009   [3] Final yield (cells/ml) with standard error in parenthesis.

1010   [4] Significance value from a two-sample t-test. The null hypothesis is that the means are equal.