

1 **The molecular chaperone DnaK accelerates protein evolution**

2 José Aguilar-Rodríguez<sup>1,2\*</sup>, Beatriz Sabater-Muñoz<sup>3,4\*</sup>, Víctor Berlanga<sup>4</sup>, David Alvarez-Ponce<sup>5</sup>,

3 Andreas Wagner<sup>1,2,6†</sup>, Mario A. Fares<sup>3,4†</sup>

4 <sup>1</sup>Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich,  
5 Switzerland

6 <sup>2</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland

7 <sup>3</sup>Department of Genetics, Smurfit Institute of Genetics, University of Dublin, Trinity College Dublin,  
8 Dublin, Ireland

9 <sup>4</sup>Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Valencia, Spain

10 <sup>5</sup>Department of Biology, University of Nevada, Reno, USA

11 <sup>6</sup>Santa Fe Institute, Santa Fe, New Mexico, USA

12 \*These authors contributed equally to this work.

13 †Correspondence:

14 Andreas Wagner

15 [andreas.wagner@ieu.uzh.ch](mailto:andreas.wagner@ieu.uzh.ch)

16 Mario A. Fares

17 [mfares@ibmcp.upv.es](mailto:mfares@ibmcp.upv.es), [faresm@tcd.ie](mailto:faresm@tcd.ie)

18 **Keywords:** DnaK, Hsp70 chaperones, molecular chaperones, mutational robustness, experimental  
19 evolution, protein evolution, *Escherichia coli*, evolutionary rates

20

21

## 22 **Abstract**

23 Molecular chaperones, also known as heat-shock proteins, refold misfolded proteins and help other  
24 proteins reach their native conformation. Thanks to these abilities, some chaperones, such as the  
25 Hsp90 protein or the chaperonin GroEL, can buffer the deleterious phenotypic effects of mutations  
26 that alter protein structure and function. Hsp70 chaperones use a chaperoning mechanism different  
27 from Hsp90 and GroEL, and it is not known whether they can also buffer mutations. Here, we show  
28 that they can. To this end, we performed a mutation accumulation experiment in *Escherichia coli*,  
29 followed by whole-genome resequencing. Our sequence data shows that overexpression of the Hsp70  
30 chaperone DnaK increases the tolerance of its clients for nonsynonymous nucleotide substitutions and  
31 nucleotide insertions and deletions. We also show that this elevated mutational buffering on short  
32 evolutionary time scales translates into differences in evolutionary rates on intermediate and long  
33 evolutionary time scales. To this end, we compared the evolutionary rates of DnaK clients and  
34 nonclients using the genomes of *E. coli*, *Salmonella typhimurium*, and 83 other gamma-proteobacterial  
35 species. We find that clients that interact strongly with DnaK evolve faster than weakly interacting  
36 clients. Our results imply that all three major chaperone classes can buffer mutations and affect protein  
37 evolution. They illustrate how an individual protein like a chaperone can have a disproportionate  
38 effect on proteome evolution.

## 39 **Introduction**

40 Molecular chaperones assist proteins in reaching their native conformations, prevent protein  
41 aggregation, and refold misfolded proteins (Young et al. 2004; Hartl and Hayer-Hartl 2009; Hartl et al.  
42 2011). Thanks to these roles, chaperones can restore the native conformation of proteins destabilized  
43 by environmental perturbations, thus providing environmental robustness to organisms coping with  
44 stressful conditions. Because some chaperones can buffer the deleterious effects of mutations that  
45 affect protein folding, they are also a source of mutational robustness (Bogumil and Dagan 2012;  
46 Fares 2015).

47 There are three main chaperone systems, which are the Hsp90 system, the Hsp70 system, and  
48 the Hsp60 system (or chaperonins), of which the bacterial GroEL is a prominent member (Hartl et al.  
49 2011). Overwhelming evidence shows that Hsp90 and GroEL can buffer mutations (Bogumil and

50 Dagan 2012), but whether the same holds for any major chaperone from the Hsp70 system is to our  
51 knowledge unknown.

52 Pioneering work carried out by Rutherford and Lindquist (1998) showed that inhibition of the  
53 chaperone Hsp90 can unveil cryptic genetic variation – genotypic variation without phenotypic  
54 variation – in the fruit fly *Drosophila melanogaster*. Subsequently, similar observations have been  
55 made in the plant *Arabidopsis thaliana* (Queitsch et al. 2002), the yeast *Saccharomyces cerevisiae*  
56 (Cowen and Lindquist 2005) and the fish *Astyanax mexicanus* (Rohner et al. 2013). Further support  
57 was recently provided by Burga et al. (2011) who found that high induction of Hsp90 during  
58 development of the nematode *Caenorhabditis elegans* reduced the penetrance of certain mutations.  
59 Additionally, Lachowiec et al. (2013) found that paralogs of duplicated kinase-coding genes that  
60 encode a substrate of Hsp90 (i.e., a Hsp90 “client”) in *A. thaliana* often evolve faster than nonclient  
61 paralogs. In general, the rate at which non-conservative substitutions – those that alter  
62 physicochemical properties of amino acids – accumulate is especially accelerated in Hsp90 clients  
63 (Pechmann and Frydman 2014).

64 Multiple studies also demonstrate mutational buffering mediated by the bacterial chaperonin  
65 GroEL. For example, Fares et al. (2002) showed that overexpressing GroEL considerably improved  
66 the fitness of *Escherichia coli* strains with a high load of deleterious mutations, a pattern that was also  
67 observed later in *Salmonella typhimurium* (Maisnier-Patin et al. 2005). Moreover, GroEL  
68 overexpression in *E. coli* increases the ability of GroEL client proteins to tolerate mutations (Tokuriki  
69 and Tawfik 2009; Bershtein et al. 2013; Wyganowski et al. 2013; Sabater-Muñoz et al. 2015), as well  
70 as their ability to undergo adaptive evolution (Tokuriki and Tawfik 2009; Wyganowski et al. 2013).  
71 Buffering of destabilizing mutations accelerates the evolutionary rates of GroEL clients (Bogumil and  
72 Dagan 2010; Warnecke and Hurst 2010; Williams and Fares 2010; Pechmann and Frydman 2014).

73 While no Hsp70 chaperone has been directly implicated in mutational buffering, pertinent  
74 circumstantial evidence exists. For example, DnaK – the major bacterial Hsp70 chaperone – is  
75 overexpressed together with GroEL in *S. typhimurium* lineages with deteriorated fitness caused by  
76 accumulating deleterious mutations (Maisnier-Patin et al. 2005). In addition, *D. melanogaster*  
77 populations showing inbreeding depression, where increased homozygosity exposes recessive

78 deleterious mutations, significantly up-regulate the expression of Hsp70 compared to outbred  
79 populations (Pedersen et al. 2005).

80 The chaperones from the Hsp70 system are very conserved from bacteria to humans (Powers  
81 and Balch 2013). They play a central role in proteome integrity, and are involved both in co- and post-  
82 translational folding (Hartl et al. 2011). In bacteria, the Hsp70 chaperone DnaK (together with GroEL  
83 and the Trigger Factor) is one of the main molecular chaperones, where it is the central hub in the  
84 chaperone network of the cytosol (Bukau and Walker 1989; Calloni et al. 2012). It interacts with ~700  
85 mostly cytosolic proteins (Calloni et al. 2012). DnaK is highly expressed constitutively, but also stress  
86 inducible, and essential at 42°C (Bukau and Walker 1989; Calloni et al. 2012). During its ATP-  
87 dependent reaction cycle, DnaK interacts with the Hsp40 co-chaperone DnaJ, and the nucleotide  
88 exchange factor GrpE (Hartl et al. 2011). DnaJ determines the client binding specificity of DnaK  
89 (Straus et al. 1990; Hoffmann et al. 1992).

90 Most mutations affecting proteins are neutral or deleterious (Eyre-Walker and Keightley 2007),  
91 and functionally important mutations often destabilize proteins (Tokuriki et al. 2008; Wyganowski et  
92 al. 2013). If DnaK buffers destabilizing mutations, then the deleterious effects of mutations in highly  
93 interacting (strong) clients should be lower than in sporadic (weak) clients, where they should be  
94 lower than in nonclients. In other words, the higher the dependency of a protein's integrity on DnaK,  
95 the higher should be its tolerance to mutations, and the lower the signature of purifying selection that  
96 purges those mutations. With this reasoning in mind, we here use laboratory experiments to evaluate  
97 the effect of DnaK buffering on the evolution of its client proteome on short evolutionary time scales.  
98 We complement our experimental observations with sequence analyses to study the effect of DnaK on  
99 intermediate and long evolutionary time scales.

## 100 **Results**

### 101 ***Experimental evolution of *E. coli* under DnaK overexpression***

102 To study the effect of DnaK overexpression on protein evolution experimentally, we performed  
103 mutation accumulation experiments similar to those we reported recently for the chaperonin GroEL,  
104 but for DnaK overexpression (Sabater-Muñoz et al. 2015). Briefly, we initiated six parallel and  
105 independent clonal lines of evolution, all of which derived from the same hypermutable clone (*E. coli*

106 K12 MG1655  $\Delta mutS$ ) (Bjedov et al. 2007). Cells of the six lines all harbored the plasmid pKJE7,  
107 which contains the operon *dnaK-dnaJ-grpE* under the control of the L-arabinose-inducible *araB*  
108 promoter (Nishihara et al. 1998). We evolved all six lines through repeated single-cell bottlenecks in  
109 the presence of the inducer, to ensure overexpression of DnaK (DnaK<sup>+</sup> lines), as well as of the  
110 cochaperone DnaJ and the nucleotide exchange factor GrpE. Because of the bottlenecks we exposed  
111 the populations to, genetic drift was strong and the efficiency of selection was weak during the  
112 experiment, such that non-lethal mutations are free to accumulate (Barrick and Lenski 2013). We  
113 evolved three of the six clonal lines at 37°C, and the other three at 42°C. The higher temperature  
114 serves to increase the deleterious effect of destabilizing mutations in the bacterial proteome (Bukau  
115 and Walker 1989). At each of the two temperatures, we also evolved a control clonal line founded  
116 from the same parental strain, but carrying a pKJE7-derived plasmid where the operon *dnaK-dnaJ-*  
117 *grpE* is deleted. The two control lines therefore cannot overexpress DnaK, even though their growth  
118 medium contains L-arabinose (DnaK<sup>-</sup> lines).

119 The overexpression of DnaK may be energetically costly, just as is the case for the chaperonin  
120 GroEL (Fares et al. 2002; Sabater-Muñoz et al. 2015). In principle, this cost could favor the  
121 accumulation of mutations that lead to a decrease in the expression of DnaK during the evolution  
122 experiment. However, we observed that the overexpression of DnaK was maintained through the  
123 mutation accumulation experiment at both 37°C and 42°C (Fig. 1, Supplementary Fig. 1). In the  
124 presence of the inducer L-arabinose, all DnaK<sup>+</sup> lines overexpressed DnaK not only at the start of the  
125 evolution experiment, but also at the end, except for a DnaK<sup>+</sup> line evolved at 42°C. However, this loss  
126 of overexpression occurred towards the end of the experiment and even then there was still  
127 overexpression for most of the daily growth cycle of this line (Supplementary Fig. 2). In no line did  
128 we observe overexpression in the absence of the inducer. The control DnaK<sup>-</sup> lines always exhibited  
129 wild-type expression levels of DnaK.

130 Importantly, while all DnaK<sup>-</sup> lines underwent repeated extinctions shortly after 85 single-cell  
131 bottlenecks or approximately 1,870 generations of evolution, the DnaK<sup>+</sup> lines overexpressing DnaK  
132 could be passaged for as long as 180 daily passages or ~3,960 generations. This observation suggests  
133 that overexpressing the chaperone DnaK has increased the robustness of the cells to the accumulation  
134 of deleterious mutations.

### 135 ***Overexpressing DnaK increases the robustness to protein-changing mutations of DnaK clients***

136 Together with the ancestral genome, we sequenced the genomes of all eight clonal lineages after  
137 ~1,870 generations. In the genomes of the evolved DnaK<sup>+</sup> lines, we first studied the incidence of  
138 protein-changing mutations – those that include nonsynonymous nucleotide substitutions as well as  
139 nucleotide insertions and deletions (indels) – among DnaK clients and nonclients. In addition to  
140 nonsynonymous substitutions, we also analyzed indels, since this mutation type, which can affect  
141 protein structure and folding, could in principle be buffered by a molecular chaperone such as DnaK.  
142 In this analysis, we considered as nonclients all proteins from the *E. coli* proteome that are not part of  
143 the set of 671 DnaK clients recently determined by Calloni et al. (2012), and analyzed the lines  
144 evolved at 37°C and 42°C independently. To improve statistical power, we combined mutations across  
145 DnaK<sup>+</sup> lines evolved at the same temperature after classifying them according to whether they affect  
146 DnaK clients or nonclients. In total, we found that the lines evolved at 37°C harbored 74 protein-  
147 changing mutations in clients and 324 mutations in nonclients (Supplementary Table 1). In other  
148 words, 11.03% of clients and 9.73% of nonclients were mutated. At this temperature, although the  
149 proportion of mutated clients is greater than the proportion of mutated nonclients, this difference is not  
150 significant (Fisher's exact test: odds ratio  $F = 1.15$ ,  $P = 0.32$ ). The DnaK<sup>+</sup> lines evolved at 42°C  
151 accumulated 82 mutations in clients and 387 in nonclients (Supplementary Table 1). In this case,  
152 12.22% of clients and 11.63% of clients were affected by mutations. Similarly to our observations at  
153 37°C, we found no evidence for a larger proportion of mutated clients (Fisher's exact test: odds ratio  $F$   
154  $= 1.06$ ,  $P = 0.64$ ). The same held at both temperatures when we controlled for the number of  
155 nonsynonymous sites in clients (502,499 sites) and nonclients (2,692,140 sites) (Fisher's exact test:  
156 odds ratio  $F = 1.22$ ,  $P = 0.13$  at 37°C and  $F = 1.14$ ,  $P = 0.31$  at 42°C).

157 While these analyses show that the relative number of mutations in clients is higher than in  
158 nonclients, the fact that this difference is not significant suggest that DnaK clients in *E. coli* do not  
159 have a greater tolerance to protein-changing mutations than nonclients. However, these analyses  
160 ignore a potentially important confounding factor: Clients may be intrinsically more sensitive to  
161 mutations than nonclients. Indeed, it has been shown that nonclient proteins are more soluble and  
162 expose less hydrophobic regions to the solvent than clients, that is client proteins are intrinsically less  
163 stable than nonclient proteins (Calloni et al. 2012). If so, comparing clients and nonclients directly

164 would not be suitable to identify the masking of deleterious mutations by DnaK. The reason is that  
165 while the proportion of mutated clients is not significantly higher than the proportion of mutated  
166 nonclients, this proportion may have increased as a consequence of DnaK overexpression. To evaluate  
167 this possibility, we compared the number of mutations among DnaK clients and nonclients in the  
168 DnaK<sup>+</sup> lines to the same numbers in the DnaK<sup>-</sup> lines (Supplementary Table 1). In the DnaK<sup>-</sup> line  
169 evolved at 37°C, 13% of mutations (19 out of 141) affected DnaK clients. Compared to this expected  
170 proportion when DnaK is not overexpressed, the proportion of mutations in clients in the DnaK<sup>+</sup> lines  
171 is significantly higher than expected (74 out of the total 394 mutations, 18.6%; binomial test:  $P =$   
172 0.002; Fig. 2). Similarly, compared to the DnaK<sup>-</sup> line evolved at 42°C, where 11% of all mutations  
173 affected DnaK clients (15 out of 136 mutations), the DnaK<sup>+</sup> lines showed significantly more mutations  
174 in clients (84 out of 471 total mutations, 17.8%; binomial test:  $P = 1.19 \times 10^{-5}$ ; Fig. 2). These results  
175 suggest that overexpressing DnaK does indeed increase the robustness of its clients to mutations.

176 Temperature itself had no significant effect on the proportion of mutated clients and nonclients  
177 (Fisher's exact test: odds ratio  $F = 1.08$ ,  $P = 0.72$ ). At both temperatures, the number of accumulated  
178 mutations in strong DnaK clients was approximately twice as high as that in weak clients, but this  
179 difference was not significant, likely because of a lack of statistical power (Supplementary Table 1,  
180 Supplementary Information section 2.1). Importantly, this observation is not spuriously explained by a  
181 difference in coding sequence length between strong and weak clients (Wilcoxon rank-sum test,  $P$ -  
182 value = 0.512).

### 183 ***DnaK accelerates protein evolution on intermediate and long evolutionary time scales***

184 We wanted to find out if the DnaK-mediated mutational buffering we observed on the short time  
185 scales of laboratory evolution has also left signatures on longer evolutionary time scales. To this end,  
186 we determined two measures of evolutionary rates for protein-coding genes from gamma-  
187 proteobacteria (Supplementary Table 2). The first, nonsynonymous divergence among one-to-one  
188 orthologs of *E. coli* and *Salmonella enterica*, is relevant for intermediate evolutionary time scales. The  
189 second, protein distance among orthologous proteins found in 85 gamma-proteobacterial genomes  
190 (including *E. coli* and *S. enterica*), is relevant for long time scales. We employ protein distance instead  
191 of nonsynonymous distance because amino acid replacements are less sensitive than nucleotide

192 substitutions to the expected loss of phylogenetic signal between sequences of distantly related taxa.  
193 To assess how strongly a protein depends on DnaK for folding, we used recent experimental  
194 proteomic data which determined how strongly 671 DnaK-interacting proteins interact with DnaK by  
195 measuring the fraction of cellular protein bound to DnaK at 37°C, a property that correlates with  
196 chaperone dependency for folding and maintenance and residence time of the protein on DnaK  
197 (Calloni et al. 2012). We note that this interaction strength is more likely to have remained unchanged  
198 during the divergence of *E. coli* and *S. enterica*, than during the divergence of all the other 83 gamma-  
199 proteobacterial species we analyzed.

200 We find a moderately strong but highly significant positive association between DnaK  
201 dependency and the rate of nonsynonymous substitutions for *S. enterica* and *E. coli* (Spearman's  $\rho =$   
202  $0.367$ ,  $N = 627$ ,  $P\text{-value} < 2.2 \times 10^{-16}$ ; Fig. 3A). This indicates that the stronger the interaction of a  
203 protein with DnaK, the faster the protein evolves. The same pattern is obtained at the larger time  
204 scales of protein distances for 85 gamma-proteobacterial genomes (Spearman's  $\rho = 0.257$ ,  $N = 311$ ,  $P\text{-}$   
205  $\text{value} = 4.4 \times 10^{-6}$ ; Fig. 3B). Gene expression level, which is the most important determinant of protein  
206 evolutionary rates, at least in unicellular organisms (Pál et al. 2001; Drummond et al. 2005), is a  
207 possible confounding factor in this analysis. For example, using codon usage bias (CUB) as a proxy  
208 for gene expression, we observe that genes with higher CUB show lower nonsynonymous divergence  
209 (Spearman's  $\rho = -0.558$ ,  $N = 1014$ ,  $P\text{-value} < 2.2 \times 10^{-16}$ ), protein distance (Spearman's  $\rho = -0.255$ ,  $N =$   
210  $3159$ ,  $P\text{-value} < 2.2 \times 10^{-16}$ ) and DnaK dependency (Spearman's  $\rho = -0.262$ ,  $N = 627$ ,  $P\text{-value} = 2.5 \times 10^{-}$   
211  $^{11}$ ). However, the association between DnaK dependency and evolutionary rate cannot be solely  
212 explained by this confounding factor: A partial correlation analysis shows that the association still  
213 holds after controlling for CUB, both on intermediate time scales (Spearman's  $\rho = 0.295$ ,  $N = 627$ ,  $P\text{-}$   
214  $\text{value} = 1.2 \times 10^{-14}$ ) and long time scales (Spearman's  $\rho = 0.229$ ,  $N = 311$ ,  $P\text{-value} = 3.8 \times 10^{-5}$ ). We use  
215 CUB instead of gene expression data here for two main reasons. First, we can compute CUB for 627  
216 DnaK clients, while expression data is only available for 457 clients. Second, gene expression data has  
217 been measured in just one environment and one strain of *E. coli*, while CUB is the result of selective  
218 pressures imposed by many different environments over long periods of time. However, the  
219 association between evolutionary rate and DnaK dependency still holds after correcting for gene  
220 expression directly (Supplementary Information section 2.2). Together, these results indicate that the



221 chaperone DnaK affects protein evolution in accordance with the mutational buffering hypothesis.  
222 Importantly, this effect is not only independent of CUB and gene expression, but also of other  
223 biological factors, such as essentiality and number of protein-protein interactions (Supplementary  
224 Information section 2.3, Supplementary Table 3).

225 Finally, while we observe that strong clients evolve faster than weak clients on both  
226 evolutionary time scales, we find that clients evolve more slowly than nonclients (Supplementary  
227 Information section 2.4, Supplementary Fig. 3). This last difference cannot be explained by the  
228 number of protein-protein interactions, by essentiality, or by CUB as confounding factors  
229 (Supplementary Information section 2.4, Supplementary Tables 4 and 5). The reason for this  
230 observation could be that clients are intrinsically less robust to mutations than nonclients due to some  
231 general physicochemical difference. For example, Calloni et al. (2012) found that DnaK clients have  
232 generally low solubility, often belong to heterooligomeric complexes, and are prone to misfolding.

### 233 **Discussion**

234 We show how the constant overexpression of the DnaK-DnaJ-GrpE chaperone system over the course  
235 of a mutation accumulation experiment increases the number of protein-changing mutations –  
236 nonsynonymous substitutions and indels – affecting DnaK clients. Additional evidence of mutational  
237 buffering by DnaK is provided by the observation that only evolving lines overproducing this  
238 chaperone avoid extinction after experiencing more than 85 single-cell bottlenecks. Recently, we  
239 obtained similar results in hypermutable *E. coli* cells evolving in identical conditions but  
240 overproducing the GroEL-GroES chaperonin system (Sabater-Muñoz et al. 2015). There, we observed  
241 that lines evolving with high levels of GroEL were not only less prone to extinction under strong  
242 genetic drift than control lines, but also that they were accumulating significantly more indels and  
243 replacements between amino acids belonging to different physicochemical categories.

244 Additionally, we find that DnaK-mediated mutational buffering has left a trace in DnaK clients  
245 during the divergence of 85 different gamma-proteobacterial species over much longer evolutionary  
246 time scales than those explored in our laboratory evolution experiment. We find that clients that  
247 depend more on DnaK for folding tend to evolve faster than less interacting clients. Similar  
248 chaperone-mediated accelerations of protein evolution have been observed on GroEL clients (Bogumil

249 and Dagan 2010; Warnecke and Hurst 2010; Williams and Fares 2010; Bogumil and Dagan 2012) and  
250 Hsp90 clients (Lachowiec et al. 2013; Pechmann and Frydman 2014). However, we notice that DnaK  
251 clients evolve slower than nonclients. This is likely the result of important physicochemical  
252 differences between clients and nonclients. For example, clients are prone to aggregation and  
253 misfolding (Calloni et al. 2012), which may make them intrinsically less robust to destabilizing  
254 mutations.

255         Despite the great differences in the mechanism of chaperone action between the three major  
256 chaperone families – chaperonins, Hsp90 chaperones and Hsp70 chaperones – (Young et al. 2004;  
257 Hartl et al. 2011; Bogumil and Dagan 2012; Kim et al. 2013), at least some of their members seem to  
258 have qualitatively comparable effects on protein evolution. Elucidating to what extent the buffering  
259 mechanisms of these chaperones differ is an important future direction of enquiry.

260         Thanks to their fostering of mutational robustness, chaperones can facilitate evolutionary  
261 innovations (Rutherford 2003), even though we do not study such innovations here. The increase in  
262 the mutational robustness of a protein caused by chaperone interactions reduces the efficiency of  
263 purifying selection in purging mutations in the protein. Thanks to chaperone-mediated buffering,  
264 many such mutations are neutral and can persist in a population. Importantly, these cryptic genetic  
265 variants may include preadaptive mutations that can generate evolutionary innovations in new  
266 environments (Tokuriki and Tawfik 2009; Wyganowski et al. 2013). To illuminate if and how DnaK  
267 can increase the innovability of its client proteome will also be an interesting subject for future work.

268         In summary, we analysed evolutionary rates of proteins that are subject to DnaK-assisted  
269 folding on short, intermediate, and long evolutionary time scales through a combination of  
270 experimental and comparative approaches. Most of our evidence indicates that the bacterial chaperone  
271 DnaK can buffer mutations in its client proteins, and that these proteins therefore evolve faster than in  
272 the absence of DnaK-mediated folding. This is, to our knowledge, the first demonstration that a  
273 member of the Hsp70 family can buffer the effect of mutations, with long-term consequences on  
274 protein evolution (Bogumil and Dagan 2012). Through its role in protein folding, an individual  
275 chaperone such as DnaK can have a disproportionate effect on proteome evolution, and thus on  
276 genome evolution.

## 277 **Materials and Methods**

### 278 *Sequence data*

279 We obtained the complete genomes of *E. coli* K-12 MG1655 (NC\_000913) and *S. enterica* serovar  
280 Typhimurium LT2 (NC\_003197) from GenBank Genomes (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>).  
281 We also used a dataset from Williams and Fares (2010) that consists of 1092 multiple sequence  
282 alignments of conserved orthologous proteins from 85 gamma-proteobacterial genomes.

### 283 *Orthology*

284 We identified 3159 one-to-one orthologs in *E. coli* and *S. enterica* genomes as reciprocal best hits  
285 (Tatusov et al. 1997) using the Basic Local Alignment Search Tool (BLAST, i.e., BLASTP with an E-  
286 value cut-off of  $10^{-10}$ ). We aligned each pair of orthologous proteins with the Needleman-Wunsch  
287 dynamic programming algorithm, using the Needle program from the EMBOSS package (Rice et al.  
288 2000). We translated the resulting alignments into codon-based nucleotide alignments with PAL2NAL  
289 (Suyama et al. 2006).

### 290 *Evolutionary rates*

291 We estimated the rate of nonsynonymous substitutions ( $d_N$ ) using the program codeml from the  
292 package PAML 4.7 (one-ratio model M0) (Yang 2007). We calculated protein distances for the  
293 gamma-proteobacterial alignments from Williams and Fares (2010), using PROTDIST from the  
294 PHYLIP package (Felsenstein 2005) and the Jones, Taylor and Thornton (JTT) substitution matrix  
295 (Jones et al. 1992). We calculated an average distance for each cluster of orthologous proteins as the  
296 mean of all pairwise distances.

### 297 *Codon usage bias*

298 We computed the Codon Adaptation Index (CAI) using the program CAI from the EMBOSS package  
299 (Rice et al. 2000). We calculated Codon Usage Bias (CUB) for each pair of *E. coli* – *S. enterica*  
300 orthologs as the mean of the CAI values for each pair of orthologs. We used CUB as a proxy for gene  
301 expression.

302

303

304 ***DnaK dependency***

305 We obtained information about DnaK clients from Calloni et al. (2012). This study used quantitative  
306 proteomics to identify 671 DnaK interactors or client proteins. For each of these proteins, the  
307 investigators calculated a relative enrichment factor that indicates the fraction of cellular protein  
308 bound to DnaK at 37°C. We used this measure as a proxy for DnaK dependency.

309 ***Bacterial strains and plasmids***

310 We obtained *E. coli* K-12 substr. MG1655  $\Delta mutS$  from Ivan Matic (Université Paris Descartes,  
311 INSERM U1001, Paris, France) through Jesús Blázquez (Centro Nacional de Biotecnología, CSIC,  
312 Madrid, Spain) (Bjedov et al. 2007). In this *E. coli* strain the gene encoding the protein MutS has been  
313 deleted. This protein is a component of the mismatch repair system that recognizes and binds  
314 mispaired nucleotides so that the mispairing can be corrected by two further repair proteins, MutL and  
315 MutH. The strain MG1655  $\Delta mutS$  has a predicted mutation rate that is 1000-fold higher than the wild  
316 type (Bjedov et al. 2007; Turrientes et al. 2013), which ensures that a sufficient number of mutations  
317 occur during the mutation accumulation experiment. We transformed this strain with the plasmid  
318 pKJE7 (Takara, Cat. #3340), which contains an operon encoding DnaK, and its co-chaperones DnaJ  
319 and GrpE under the regulation of a single promoter inducible by L-arabinose (Nishihara et al. 1998).  
320 We generated a control strain by transforming the same  $\Delta mutS$  strain with a plasmid that lacks the  
321 operon *dnaK-dnaJ-grpE* but is otherwise identical to pKJE7. We refer to this plasmid as pKJE7-  
322 DEL(*dnaK-dnaJ-grpE*). This control plasmid was derived from the plasmid pKJE7 by removal of the  
323 operon *dnaK-dnaJ-grpE* with a restriction digest using *Bam*HI and *Spe*I, followed by religation, after  
324 obtaining permission for plasmid modification from Takara.

325 ***Evolution experiment***

326 We evolved six clonal lines of the hypermutable *E. coli*  $\Delta mutS$  strain containing pKJE7 (DnaK<sup>+</sup> lines)  
327 and two lines containing the control plasmid pKJE7-DEL(*dnaK-dnaJ-grpE*) (DnaK<sup>-</sup> lines) by daily  
328 passaging them through single-cell bottlenecks on solid LB medium (agar plates; Pronadisa #1551 and  
329 #1800) supplemented with 20 µg/mL of chloramphenicol (Sigma-Aldrich #C0378) and 0.2% (w/v) of  
330 L-arabinose (Sigma-Aldrich #A3256). We passaged both the DnaK<sup>-</sup> and DnaK<sup>+</sup> lines during 85 days  
331 or approximately 1,870 generations (assuming ~22 generations per daily growth cycle). We evolved

332 half of the DnaK<sup>+</sup> and DnaK<sup>-</sup> lines under mild heat-stress (42°C) while the other half remained at  
333 37°C.

#### 334 ***Verification of DnaK overexpression***

335 We grew the ancestral and evolved strains (DnaK<sup>+</sup> and DnaK<sup>-</sup>, at 37°C and 42°C) from glycerol stocks  
336 in liquid LB medium supplemented with 20 µg/mL of chloramphenicol in the presence or absence of  
337 the inducer L-arabinose (0.2%). After 24h of growth, we pelleted cells by centrifugation at 12000 rpm.  
338 We resuspended the pelleted cells in 100µl lysis buffer (containing 200 mM Tris-HCl pH6.8, 10mM  
339 DTT, 5% SDS, 50% glycerol). To prepare a crude extract, we first boiled resuspended cells at 95°C  
340 for 15 minutes. After the removal of cell debris by centrifugation, we quantified soluble proteins using  
341 the Bradford method (Bradford 1976). We loaded one microgram of total protein for each sample in  
342 SDS-PAGE gels (12.5% resolving gel). In addition, we loaded onto all gels samples from the ancestral  
343 DnaK<sup>-</sup> and DnaK<sup>+</sup> strains grown in the presence of inducer at 37°C, as controls to facilitate inter-gel  
344 comparisons. We detected DnaK protein by Western blotting using as primary antibody a  
345 mouse monoclonal antibody specific to *E. coli* DnaK (Abcam #ab69617) at a 1:10,000 dilution, and as  
346 secondary antibody a goat polyclonal (alkaline phosphatase-conjugated) antibody specific to mouse  
347 IgG1 (Abcam #ab97237). We scanned membranes after colorimetric detection of conjugated  
348 antibodies with the BCIP<sup>®</sup>/NBT-Purple liquid substrate system (Sigma-Aldrich #BP3679), and used  
349 ImageJ to quantify the intensity of DnaK bands on the Western blots (Schneider et al. 2012). We used  
350 the control samples to normalize abundances, which allow the comparison of DnaK levels across  
351 experiments.

352 We examined the change in DnaK levels along a daily cycle of growth for a DnaK<sup>+</sup> line evolved  
353 at 42°C (line #2, Supplementary Figure 2) that lost overexpression after ~1,870 generations of  
354 mutation accumulation. After 24h of exponential growth at 42°C in liquid LB medium supplemented  
355 with chloramphenicol, we diluted the culture to OD ~0.3, and induced DnaK expression by adding 10  
356 mM of L-arabinose. We allowed the culture to grow for another 24h in the presence of this expression  
357 inducer. Each hour we removed one milliliter of culture and measured the DnaK level following the  
358 protocol described above.

359

360 ***Whole-genome resequencing***

361 We sequenced the genomes of the DnaK<sup>-</sup> and DnaK<sup>+</sup> lines after 85 single-cell bottlenecks, an addition  
362 to the ancestral  $\Delta mutS$  strain from which both the DnaK<sup>+</sup> and DnaK<sup>-</sup> lines were derived.

363 Specifically, for the evolved lines we performed paired-end Illumina whole-genome  
364 sequencing. For DNA extraction, we used the QIAmp DNA mini kit (Qiagen, Venlo [Pays Bas],  
365 Germany) in a QiaCube automatic DNA extractor using bacterial pellets obtained from approximately  
366 10 mL cultures. We constructed multiplexed DNaseq libraries from each clonal evolution line using  
367 the TrueSeq DNA polymerase chain reaction-free HT sample preparation kit (Illumina). We  
368 performed paired-end sequencing on an Illumina HiSeq2000 platform, using a 2 × 100 cycles  
369 configuration.

370 We performed whole-genome sequencing of the ancestral strain using paired-end and shotgun  
371 454 sequencing (Roche/454 Life Sciences, Branford, CT). We performed a *de novo* assembly using  
372 the GS De Novo Assembler version 2.6 (Roche), and combined the mapped and *de novo* assemblies  
373 into a single assembly using in-house scripts. We examined reads to confirm indels using ssaha and  
374 breseq v0.24rc4 (Deatherage and Barrick 2014), and transferred annotations from *E. coli* MG1655  
375 using RATT (Otto et al. 2011).

376 We converted sequencing reads converted from Illumina quality scores into Sanger quality  
377 scores. Subsequently, we used the breseq v 0.24rc4 (version 4) pipeline (Deatherage and Barrick  
378 2014) for aligning the Illumina reads to our *E. coli* parental genome and for identifying single  
379 nucleotide polymorphisms and indels using bowtie2 (Langmead and Salzberg 2012). We performed  
380 individual runs of breseq, with junction prediction disabled but otherwise default parameters, for the  
381 ancestral sequence, as well as for each of the evolved lines.

382

383

384

385

386

387 **Acknowledgments**

388 The authors thank Xiaoshu Chen and Jianzhi Zhang for kindly providing us with the gene expression  
389 data. This work was supported by the Forschungskredit program of the University of Zurich (grant  
390 FK-14-076 to J.A.), the Swiss National Science Foundation (grant 31003A\_146137 to A.W.), the  
391 University Priority Research Program in Evolutionary Biology at the University of Zurich (to A.W.),  
392 the Science Foundation Ireland (grant 12/IP/1673 to M.A.F), and the Spanish Ministerio de Economía  
393 y Competitividad (grant BFU2012-36346 to M.A.F.).

394

395

396

397

398

399

400

401

402

403

404

405

406

407

## 408 **References**

- 409 Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. *Nat. Rev. Genet.*  
410 14:827–839.
- 411 Bershtein S, Mu W, Serohijos AWR, Zhou J, Shakhnovich EI. 2013. Protein quality control acts on  
412 folding intermediates to shape the effects of mutations on organismal fitness. *Mol. Cell* 49:133–  
413 144.
- 414 Bjedov I, Dasgupta CN, Slade D, Le Blastier S, Selva M, Matic I. 2007. Involvement of *Escherichia*  
415 *coli* DNA polymerase IV in tolerance of cytotoxic alkylating DNA lesions in vivo. *Genetics*  
416 176:1431–1440.
- 417 Bogumil D, Dagan T. 2010. Chaperonin-dependent accelerated substitution rates in prokaryotes.  
418 *Genome Biol. Evol.* 2:602–608.
- 419 Bogumil D, Dagan T. 2012. Cumulative impact of chaperone-mediated folding on genome evolution.  
420 *Biochemistry* 51:9941–9953.
- 421 Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of  
422 protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248–254.
- 423 Bukau B, Walker GC. 1989. Cellular defects caused by deletion of the *Escherichia coli* dnaK gene  
424 indicate roles for heat shock protein in normal metabolism. *J. Bacteriol.* 171:2337–2346.
- 425 Burga A, Casanueva MO, Lehner B. 2011. Predicting mutation outcome from early stochastic  
426 variation in genetic interaction partners. *Nature* 480:250–253.
- 427 Calloni G, Chen T, Schermann SM, Chang H-C, Genevaux P, Agostini F, Tartaglia GG, Hayer-Hartl  
428 M, Hartl FU. 2012. DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Rep.*  
429 1:251–264.
- 430 Cowen LE, Lindquist S. 2005. Hsp90 potentiates the rapid evolution of new traits: drug resistance in  
431 diverse fungi. *Science.* 309:2185–2189.
- 432 Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from  
433 next-generation sequencing data using breseq. *Methods Mol. Biol.* 1151:165–188.
- 434 Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins  
435 evolve slowly. *Proc. Natl. Acad. Sci. U. S. A.* 102:14338–14343.
- 436 Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat. Rev.*  
437 *Genet.* 8:610–618.
- 438 Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria: groEL  
439 buffers against deleterious mutations. *Nature* 417:398. Available from:  
440 [http://personales.upv.es/sfelena/Nature Fares et al.pdf](http://personales.upv.es/sfelena/Nature%20Fares%20et%20al.pdf)
- 441 Fares MA. 2015. The origins of mutational robustness. *Trends Genet.* 31:373–381.
- 442 Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6.
- 443 Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in protein folding and proteostasis.  
444 *Nature* 475:324–332.
- 445 Hartl FU, Hayer-Hartl M. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat.*  
446 *Struct. Mol. Biol.* 16:574–581.
- 447 Hoffmann HJ, Lyman SK, Lu C, Petit MA, Echols H. 1992. Activity of the Hsp70 chaperone  
448 complex--DnaK, DnaJ, and GrpE--in initiating phage lambda DNA replication by sequestering  
449 and releasing lambda P protein. *Proc. Natl. Acad. Sci. U. S. A.* 89:12108–12111.
- 450 Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from  
451 protein sequences. *Comput. Appl. Biosci.* 8:275–282.
- 452 Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. 2013. Molecular chaperone functions in  
453 protein folding and proteostasis. *Annu. Rev. Biochem.* 82:323–355.



- 454 Lachowiec J, Lemus T, Thomas JH, Murphy PJM, Nemhauser JL, Queitsch C. 2013. The Protein  
455 Chaperone HSP90 Can Facilitate the Divergence of Gene Duplicates. *Genetics* 193:1269–1277.
- 456 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–  
457 359. Available from: <http://dx.doi.org/10.1038/nmeth.1923>
- 458 Maisnier-Patin S, Roth JR, Fredriksson A, Nyström T, Berg OG, Andersson DI. 2005. Genomic  
459 buffering mitigates the effects of deleterious mutations in bacteria. *Nat. Genet.* 37:1376–1379.
- 460 Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids:  
461 differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of  
462 an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.*  
463 64:1694–1699.
- 464 Otto TD, Dillon GP, Degraeve WS, Berriman M. 2011. RATT: Rapid Annotation Transfer Tool.  
465 *Nucleic Acids Res.* 39:1–7.
- 466 Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158:927–  
467 931.
- 468 Pechmann S, Frydman J. 2014. Interplay between Chaperones and Protein Disorder Promotes the  
469 Evolution of Protein Networks. *PLoS Comput. Biol.* 10:e1003674.
- 470 Pedersen KS, Kristensen TN, Loeschcke V. 2005. Effects of inbreeding and rate of inbreeding in  
471 *Drosophila melanogaster*- Hsp70 expression and fitness. *J. Evol. Biol.* 18:756–762.
- 472 Powers ET, Balch WE. 2013. Diversity in the origins of proteostasis networks - a driver for protein  
473 function in evolution. *Nat. Rev. Mol. Cell Biol.* 14:237–248.
- 474 Queitsch C, Sangster T a, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature*  
475 417:618–624.
- 476 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software  
477 Suite. *Trends Genet.* 16:276–277.
- 478 Rohner N, Jarosz DF, Kowalko JE, Yoshizawa M, Jeffery WR, Borowsky RL, Lindquist S, Tabin CJ.  
479 2013. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in  
480 cavefish. *Science.* 342:1372–1375.
- 481 Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature*  
482 396:336–342.
- 483 Rutherford SL. 2003. Between genotype and phenotype: protein chaperones and evolvability. *Nat.*  
484 *Rev. Genet.* 4:263–274.
- 485 Sabater-Muñoz B, Prats-Escriche M, Montagud-Martínez R, López-Cerdán A, Toft C, Aguilar-  
486 Rodríguez J, Wagner A, Fares MA. 2015. Fitness Trade-Offs Determine the Role of the  
487 Molecular Chaperonin GroEL in Buffering Mutations. *Mol. Biol. Evol.* 32:2681–2693.
- 488 Schneider C a, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis.  
489 *Nat. Methods* 9:671–675.
- 490 Straus D, Walter W, Gross C a. 1990. DnaK, DnaJ, and GrpE heat shock proteins negatively regulate  
491 heat shock gene expression by controlling the synthesis and stability of sigma 32. *Genes Dev.*  
492 4:2202–2209.
- 493 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments  
494 into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
- 495 Tatusov RL, Koonin E, Lipman DJ. 1997. A genomic perspective on protein families. *Science*  
496 278:631–637.
- 497 Tokuriki N, Stricher F, Serrano L, Tawfik DS. 2008. How Protein Stability and New Functions Trade  
498 Off. *PLoS Comput. Biol.* 4:e1000002.
- 499 Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme  
500 evolution. *Nature* 459:668–673.

501 Turrientes M-C, Baquero F, Levin BR, Martínez J-L, Ripoll A, González-Alba J-M, Tobes R,  
502 Manrique M, Baquero M-R, Rodríguez-Domínguez M-J, et al. 2013. Normal Mutation Rate  
503 Variants Arise in a Mutator (Mut S) *Escherichia coli* Population. *PLoS One* 8:e72963.

504 Warnecke T, Hurst LD. 2010. GroEL dependency affects codon usage--support for a critical role of  
505 misfolding in gene evolution. *Mol. Syst. Biol.* 6:340.

506 Williams TA, Fares MA. 2010. The effect of chaperonin buffering on protein evolution. *Genome Biol.*  
507 *Evol.* 2:609–619.

508 Wyganowski KT, Kaltenbach M, Tokuriki N. 2013. GroEL/ES buffering and compensatory mutations  
509 promote protein evolution by stabilizing folding intermediates. *J. Mol. Biol.* 425:3403–3414.

510 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–  
511 1591.

512 Young JC, Agashe VR, Siegers K, Hartl FU. 2004. Pathways of chaperone-mediated protein folding in  
513 the cytosol. *Nat. Rev. Mol. Cell Biol.* 5:781–791.

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

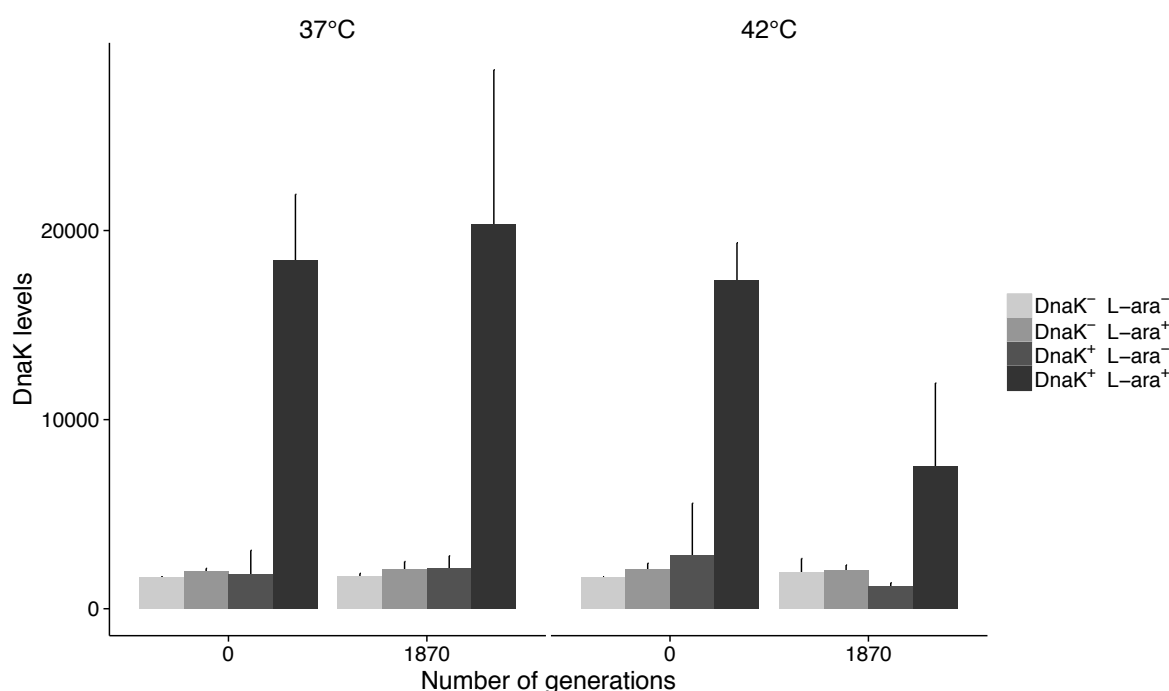
539

540

541

542 **Figures**

543



544

545

546 **Figure 1. DnaK abundance at the end of the mutation accumulation experiment.** We measured

547 the abundance of the chaperone DnaK for both DnaK<sup>+</sup> and DnaK<sup>-</sup> lines evolved for approximately

548 1,870 generations at 37°C and 42°C. For comparison, we also measured the abundance of the

549 chaperone in the ancestral DnaK<sup>+</sup> and DnaK<sup>-</sup> strains at both temperatures. We determined DnaK levels

550 in the presence and absence of the inducer L-arabinose (L-ara<sup>+</sup> and L-ara<sup>-</sup>, respectively), as described

551 in Material and Methods (“Verification of DnaK overexpression”), via the intensity of the DnaK band

552 in a Western blot. The evolved lines did not lose the ability to overexpress DnaK in the presence of the

553 inducer L-arabinose except for a DnaK<sup>+</sup> line evolved at 42°C (line #2), which explains the decrease in

554 the average DnaK abundance at the end of the evolution experiment. However, this loss of

555 overexpression occurred late in the evolution experiment, and it is not even complete for most of the

556 daily growth cycle of this line (Supplementary Figure 2). The height of the bars indicates mean DnaK

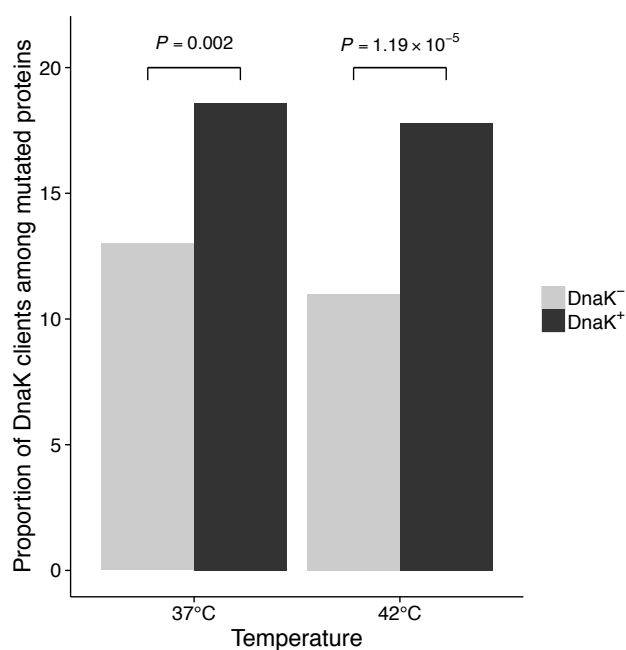
557 abundance across two experimental replicates per strain and condition. Error bars represent one

558 standard deviation of the mean.

559

560

561



562

563

564 **Figure 2. Higher proportion of mutated DnaK clients among all mutated proteins after ~1,870**

565 **generations of mutation accumulation.** The proportion of mutated DnaK clients is significantly

566 higher in DnaK<sup>+</sup> lines that overexpress DnaK than in the control DnaK<sup>-</sup> lines that do not express the

567 chaperone at such high levels. This is observed both for lines evolved at 37°C and 42°C. We consider

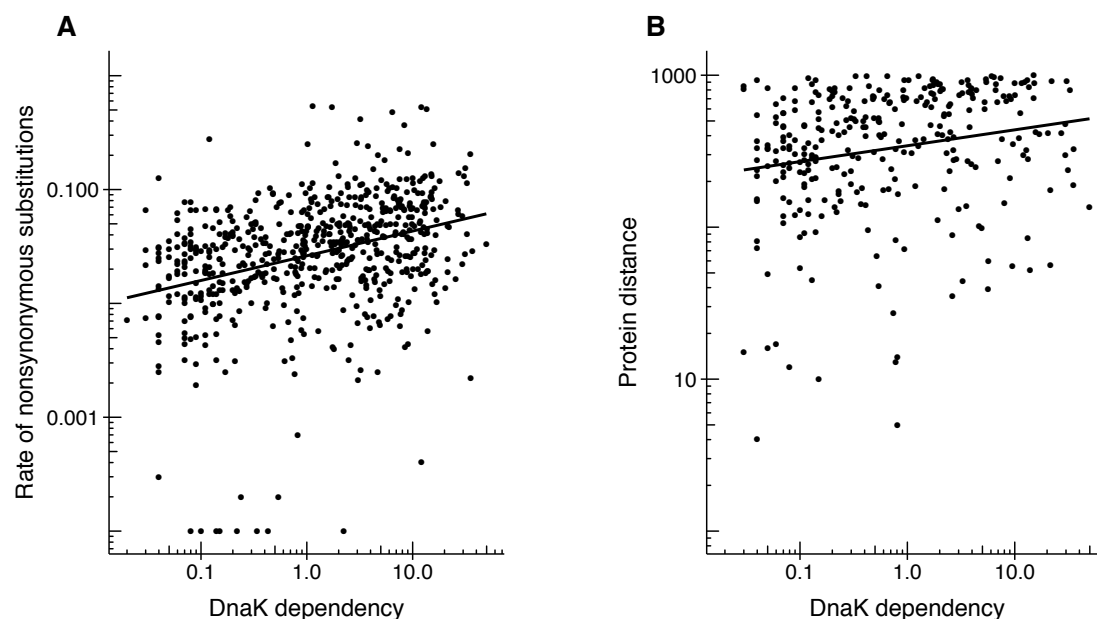
568 a protein as mutated if it has a nonsynonymous substitution and/or an indel. Statistical significance

569 was evaluated with a binomial test.

570

571

572



573  
574 **Figure 3. DnaK accelerates protein evolution on intermediate and long evolutionary time scales.**  
575 Scatter-plots showing the relationship between DnaK dependency (calculated as a relative enrichment  
576 factor that indicates the fraction of cellular protein bound to DnaK at 37°C, horizontal axis) and the  
577 degree of divergence over (A) intermediate time scales, measured as nonsynonymous divergence  
578 (Spearman's  $\rho = 0.367$ ,  $N = 627$ ,  $P$ -value  $< 2.2 \times 10^{-16}$ ), and (B) long time scales, measured as protein  
579 distance (Spearman's  $\rho = 0.257$ ,  $N = 311$ ,  $P$ -value  $= 4.4 \times 10^{-6}$ ) (vertical axes). Solid lines represent the  
580 best fit to the points. Note the logarithmic scale on both axes.

581  
582  
583  
584  
585  
586  
587  
588  
589