

1 **MEASURING SELECTION COEFFICIENTS BELOW  $10^{-3}$ : METHOD, QUESTIONS AND**  
2 **PROSPECTS**

3

4 Romain Gallet<sup>1</sup>, Tim F. Cooper<sup>2</sup>, Santiago F. Elena<sup>3,4</sup> and Thomas Lenormand<sup>1\*</sup>

5 \*: corresponding author

6 Affiliations:

7 <sup>1</sup>: CEFE - UMR 5175 1919 route de Mende, F-34293 Montpellier cedex 5, France

8 <sup>2</sup>: Evolution and Ecology Group, Department of Biology and Biochemistry, University of  
9 Houston, Houston, TX 77204, USA

10 <sup>3</sup>: Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Campus UPV CPI 8E, CL.  
11 Ingeniero Fausto Elio s/n, 46022 València, Spain

12 <sup>4</sup>: Santa Fe Institute, 1399 Hyde Road Park, Santa Fe, NM 87501, USA

13

14 Email:

15 rgallet@gmail.com

16 tfcooper@uh.edu

17 sfelena@ibmcp.upv.es

18 thomas.lenormand@cefe.cnrs.fr

19

20 Running title: High precision fitness measures

21 Keywords: mutation fitness effect, mixed model, high throughput data, drift, flow cytometry,  
22 transitivity, epistasis, fluorescence, *E. coli*

23

## ABSTRACT

1  
2 Measuring fitness with precision is a key issue in evolutionary biology, in particular to study  
3 mutations of small effects. It is usually thought that sampling error and drift prevent precise  
4 measurement of very small fitness effects. We circumvented these limits by using a new  
5 combined approach to measure and analyze fitness. We estimated the mutational fitness  
6 effect (MFE) of three independent mini-Tn10 transposon insertion mutations by conducting  
7 competition experiments in large populations of *Escherichia coli* under controlled laboratory  
8 conditions. Using flow cytometry to assess genotype frequencies from very large samples  
9 alleviated the problem of sampling error, while the effect of drift was controlled by using  
10 large populations and massive replication of fitness measures. Furthermore, with a set of  
11 four competition experiments between ancestral and mutant genotypes, we were able to  
12 decompose fitness measures into four estimated parameters that account for fitness effects  
13 of our fluorescent marker ( $\alpha$ ), the mutation ( $\beta$ ), epistasis between the mutation and the  
14 marker ( $\gamma$ ), and departure from transitivity ( $\tau$ ). Our method allowed us to estimate mean  
15 selection coefficients to a precision of  $2 \times 10^{-4}$ . We also found small, but significant epistatic  
16 interactions between the allelic effects of mutations and markers, and confirmed that fitness  
17 effects were transitive in most cases. Unexpectedly, we also detected variation in measures  
18 of  $s$  that were significantly bigger than expected due to drift alone, indicating the existence  
19 of cryptic variation, even in fully controlled experiments. Overall our results indicate that  
20 selection coefficients are best understood as being distributed, representing a limit on the  
21 precision with which selection can be measured, even under controlled laboratory  
22 conditions.

## INTRODUCTION

23  
24 Mutations of small effect can play an important role in evolution, but they are difficult to  
25 measure experimentally because the precision with which fitness effects can be measured is  
26 relatively low (see below for details). For this reason, it remains unclear to what extent  
27 mutations with small beneficial effects contribute to fitness improvements (ORR 2005). It is  
28 also unclear how much deleterious mutations of small effect contribute to the genetic load  
29 and inbreeding depression (BATAILLON and KIRKPATRICK 2000; CHARLESWORTH and CHARLESWORTH  
30 1998). More generally, the existence and influence of mutations of small effect is at the

1 heart of the neutralist-selectionist controversy (e.g. NEI 2005). This debate can only be  
2 addressed experimentally if the precision of fitness measurements is lower than the inverse  
3 of effective population size, which seems beyond reach for large populations (KREITMAN  
4 1996). Finally, a low precision in fitness measures limits the ability to determine whether the  
5 fitness effect of a mutation varies across different environmental or genetic contexts and  
6 adds to other sources of stochasticity (LENORMAND *et al.* 2009) to make it difficult to reliably  
7 predict evolutionary trajectories.

8 Precisely measuring fitness poses technical, conceptual and statistical challenges. The  
9 technical challenge is to set up a technique that allows experiments to be carried out  
10 efficiently. The first major advance was to use 'population cages' with *Drosophila* or other  
11 small animals (starting in the 30's with the work of L'Heritier and Teissier (1937a; 1937b)).  
12 With such devices, environmental conditions are relatively controlled and gene flow can be  
13 eliminated. However drift and indirect selection caused by loci under selection in linkage  
14 disequilibrium with the focal locus, are difficult to account for. The same approach was  
15 applied to microorganisms (DYKHUIZEN and HARTL 1980), which can be made isogenic save for  
16 a focal gene thereby reducing indirect selection due to initial linkage disequilibrium (e.g.  
17 CARRASCO *et al.* 2007; DOMINGO-CALAP *et al.* 2009 for distribution of mutation fitness effects;  
18 ELENA *et al.* 1998; PERIS *et al.* 2010; SANJUAN *et al.* 2004) and can be propagated as large  
19 populations, minimizing the effect of drift relative to selection. They can also be followed  
20 over many generations (DYKHUIZEN and HARTL 1983; LUNZER *et al.* 2002; THATCHER *et al.* 1998).  
21 Long term monitoring increases the ability to detect small differences in fitness between  
22 competing genotypes, but adds the complication that newly arising mutations may perturb  
23 the assay (DYKHUIZEN and HARTL 1983). An important technical issue in all competition  
24 experiments is to determine the frequency of competing genotypes reliably and quickly. In  
25 many cases the idea is to link an easily recognized marker with the gene under scrutiny. It is,  
26 however, important to recognize that a marker can confer a selective difference (a marker  
27 'cost'), which might vary with the genetic background (epistasis) or external environment  
28 (G×E interactions). Finally, inferring allelic selection coefficients against a common reference  
29 strain requires that genotypic fitness is transitive. These potential complications require  
30 adding proper controls to competition experiments.

1 A key conceptual difficulty in measuring the fitness effects of mutations is to distinguish  
2 selection from drift (BEATTY 1984; MILLSTEIN 2008), which is at the heart of several population  
3 cage experiments with *Drosophila* (DOBZHANSKY and PAVLOVSKY 1957). To account for the  
4 effect of drift, a selection coefficient can be defined from the expected change in allele  
5 frequency over one generation (e.g. ROUSSET 2004), which can be estimated from the mean  
6 frequency change in independent competition experiments. Because of drift, replication is  
7 fundamentally necessary to estimate fitness, and the precision of a given fitness measure  
8 must account for the inter-replicate variance. Indeed, it is possible to count all organisms in  
9 an experimental population, so that the genotype frequencies are known without sampling  
10 error. Such an experiment would allow frequency variation to be determined 'exactly', but  
11 would clearly not account for the possibility that drift will cause different outcomes in  
12 different replicates. A further complication is that fitness may vary because of changing  
13 environmental conditions. Fluctuating selection during the course of a competition  
14 experiment or varying selection across replicates of a competition assay can mimic drift  
15 (FELSENSTEIN 1976; LYNCH 1987; O'HARA 2005). If selection varies, and it probably always does  
16 to some extent (BELL 2008a; BELL 2010), measuring selection requires measuring both a mean  
17 *and* a variance (the latter not including sampling error). The remaining variance can be  
18 caused by drift or by heterogeneity in selection, which are difficult to disentangle without  
19 extra information on the effective population size. In summary, measuring selection with  
20 precision requires estimating an expectation over several replicates, so that its variance can  
21 be decomposed into components due to sampling error, drift and variable selection.

22 From a statistical point of view, selection coefficients in the field or in the laboratory are best  
23 estimated by using a fully specified selection model in a likelihood framework (e.g. ARNASON  
24 and LEWONTIN 1991; CLARK 1979; LABBE *et al.* 2009; LENORMAND and RAYMOND 2000; MANLY  
25 1985; OAKESHOTT *et al.* 1983; SACCHERI *et al.* 2008; WILSON *et al.* 1982), which can include drift  
26 if longitudinal data are available (BOLLBACK *et al.* 2008; MANLY 1985; O'HARA 2005). When  
27 selection can be approximated by a continuous process through time in an isolated  
28 population, a simple approach is to regress  $\text{Log}(p/q)$  (where  $p$  and  $q$  represent the  
29 frequencies of the two competitors) over time expressed in units of generations (FISHER  
30 1930). The connection with logistic regression and general linear models is then  
31 straightforward (ARNASON and BARKER 1999) and more appropriate than the use of least

1 squares. However, complications arise in the analysis of time series and correlated error in  
2 repeated measurement through time (ARNASON and BARKER 1999; O'HARA 2005), especially  
3 when both drift and fluctuating selection cause frequency variation. The latter problems can  
4 be important, particularly when analyzing multiple time point series (e.g., arising in long  
5 term population cage or chemostat experiments), although they are rarely taken into  
6 account. Often, replicated experiments are simply pooled, even if significantly different, and  
7 not analyzed to consider variance in the estimates of selection. The development of mixed  
8 models offers an attractive alternative to circumvent this problem and to measure selection  
9 and its variation.

10 We present an approach combining several features to improve and quantify the precision  
11 of fitness measures. First, we use techniques that have proved to be among the most  
12 efficient to measure fitness: competition assay between large populations of *Escherichia coli*  
13 strains to minimize drift and engineered mutations to avoid the problem of indirect  
14 selection. Specifically, we used three genotypes each carrying a single mutation introduced  
15 by the integration of a mini-Tn10 transposon. These mutations were considered neutral,  
16 relative to a common progenitor genotype, in a previous experiment (ELENA *et al.* 1998). We  
17 use two fluorescent markers (ROSENFELD *et al.* 2005) combined with flow cytometry (LUNZER *et*  
18 *al.* 2002) to measure frequency variation with great precision, and thus minimize sampling  
19 error. Other studies have shown the utility of these approaches in measuring genotype  
20 fitness (LEE *et al.* 2009; LUNZER *et al.* 2002; ZHU *et al.* 2005). Key aspects of our approach are  
21 (1) a comprehensive set of four competition assays that enable us to separately estimate  
22 mutational selection coefficients ( $\alpha$ ), the cost of the marker ( $\beta$ ), epistasis between mutation  
23 and marker ( $\gamma$ ), and transitivity ( $\tau$ ). (2) We use short-term batch culture to facilitate massive  
24 replication and to reduce the possibility that *de novo* beneficial mutations will occur. (3) We  
25 analyze the data in an integrated likelihood framework with random effects to partition  
26 sources of variation in our estimates (sampling error versus drift versus variable selection).  
27 Our approach allowed us to estimate both mean *and* variance in selection coefficients at a  
28 precision of 0.02%. This precision allowed us to detect variation in measures of some  
29 mutation selection coefficients that were significantly larger than expected due to drift  
30 alone, indicating the action of some kind of cryptic variation during our competitions. This  
31 finding implies that, in practice, selection coefficients should be considered as being

1 distributed, and that precise measures requires evaluating both the mean *and* the variance  
2 of this distribution. Furthermore, the variance in *s* indicates that some uncontrolled  
3 processes occur in these experiments (cryptic environmental or genetic variation), which  
4 impose a limit to further dissecting the differences seen across replicates. We discuss  
5 implications of these findings and the prospects of this high-throughput method for fitness  
6 measurement.

## 7 MATERIAL AND METHODS

### 8 Strain construction

9 The *E. coli* B strain used in the present study, REL4548, was evolved in Davis Minimal  
10 medium supplemented with 25 µg/mL glucose (DM25) for 10 000 generations as part of a  
11 long-term evolution experiment (ELENA *et al.* 1998).

12 **Insertions of the chromosomal fluorescent markers :** The YFP and CFP genes (provided by  
13 the Yeast Resource Center of the University of Washington) were inserted at the *rhaA* locus  
14 of REL4548 using the a technique developed by Datsenko and Wanner (DATSENKO and  
15 WANNER 2000). Table 1 gives a description of this method as applied to our experiments. A  
16 full description of the method is given in *supplementary material*.

17 **Mutant construction:** The three mutants studied here were constructed by Elena *et al.*  
18 (ELENA *et al.* 1998) and were obtained by random single insertions of mini-Tn10 derivative  
19 104 – which contains a tetracycline resistance cassette (KLECKNER *et al.* 1991) – into REL4548.  
20 We chose mutations T63, T103, and T121 from this original collection because they were  
21 identified as neutral using the standard plating method. These mutations were transduced  
22 into REL4548/CFP and REL4548/YFP by P1 transduction, in order to have each mutation  
23 associated with each fluorescent marker. Since P1 transductions were performed between  
24 isogenic strains (except for the marker and the mobilized mutation), the risk of secondary  
25 mutations was low. Transductants were selected on LBA-Tet plates (LB agar plates  
26 supplemented 10 µg/mL Tetracycline). We denote ‘*wc*’ the wild type genotype with CFP  
27 marker (*wc* for wild type cyan), ‘*wy*’ for wild type YFP, ‘*mc*’ for mutant CFP, ‘*my*’ for mutant  
28 YFP.

## 1 **Competition experiments**

2 **Media: Lysogeny broth** (LB) was used for routine molecular work and for reviving strains  
3 from storage (10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract; LB Agar LB + 15 g/L agar).  
4 Davis minimal (DM) medium supplemented with 250  $\mu\text{g}/\text{mL}$  glucose (DM 250) was used for  
5 all competition assays ( $\text{KH}_2\text{PO}_4 \cdot 3\text{H}_2\text{O}$  7 g/L,  $\text{KH}_2\text{PO}_4$  2 g/L,  $(\text{NH}_4)_2\text{SO}_4$  1 g/L, sodium citrate 0.5  
6 g/L; pH was adjusted to 7.0 with HCl or NaOH as necessary). Bottles were weighed before  
7 and after autoclaving and sterile milliQ water was added to compensate for evaporation.  
8 After autoclaving, DM was supplemented with: 2.5 mL glucose 10%, 1 mL  $\text{MgSO}_4^{2-}$  10%, 1 mL  
9 Thiamine (vitamin B1) 0.2%. We call this medium DM250, which is equivalent to the one  
10 used by Lenski (1991), in which the strain REL4548 grew for 10 000 generations, but with 10  
11 times more glucose. Glycerol stocks: All strains were grown overnight and a sample of 750  
12  $\mu\text{L}$  of each culture was mixed to 250  $\mu\text{L}$  of 60% glycerol and kept at  $-80^\circ\text{C}$  for storage.

13 **Culture:** The relative fitness,  $W$ , of each mutant was estimated by measuring the change in  
14 its relative frequency in competition experiments. In order to measure the mutation fitness  
15 effect (MFE) and to control for potential marker effects and epistasis between the mutation  
16 and the marker, we performed four competitions types for each mutant: (a)  $wc/wy$ , (b)  
17  $mc/my$ , (c)  $my/wc$ , and (d)  $mc/wy$ . The rationale for performing all these competitions is  
18 presented below. Competitions were begun by growing the strains to be competed at  $37^\circ\text{C}$   
19 overnight with shaking at 250 rpm in 24 well microtiter plates (Greiner Bio-one 662102 -  
20 suspension culture plates) containing 1 mL/well of DM250. We used DM250 as the growth  
21 medium to obtain large population sizes, which limit the effect of drift, and to facilitate the  
22 measurement of hundreds of thousands cells without having to sample large volumes. To  
23 limit evaporation, each 24-well plate was placed in a two-liter plastic box containing paper  
24 towels soaked with 100 mL water (at the bottom of the box). The next day, 10  $\mu\text{L}$  (100-fold  
25 dilution) of each culture was transferred to a fresh plate and incubated for 24 hours under  
26 identical conditions. On the third day, competitors were mixed at a 1:1 ratio (5  $\mu\text{L}$  of each  
27 competitor) and transferred to a fresh plate under identical conditions. On day four, 20  $\mu\text{L}$  of  
28 each competition was transferred into 10 replicate wells containing 1980  $\mu\text{L}$  of DM250. After  
29 mixing, 1 mL was removed from each well and placed in a plastic test tube at  $4^\circ\text{C}$  for a  
30 subsequent flow cytometry measurement (performed one hour later), while the remaining 1  
31 mL was kept in the microtiter plate to be cultivated under the conditions described above.

1 Finally, on the fifth day, a 100  $\mu$ L sample was taken from each competition, diluted in DM  
2 (not containing glucose, thiamine or  $\text{MgSO}_4^{2-}$ ), and placed in a plastic test tube at 4°C for a  
3 subsequent flow cytometry measurement (performed one hour later). Ten different types of  
4 competitions were performed: *wc vs wy*, *m<sub>T63C</sub> vs m<sub>T63Y</sub>*, *m<sub>T63C</sub> vs wy*, *wc vs m<sub>T63Y</sub>*, *m<sub>T103C</sub> vs*  
5 *m<sub>T103Y</sub>*, *m<sub>T103C</sub> vs wy*, *wc vs m<sub>T103Y</sub>*, *m<sub>T121C</sub> vs m<sub>T121Y</sub>*, *m<sub>T121C</sub> vs m<sub>T121Y</sub>*, *wc vs m<sub>T121Y</sub>*. Each  
6 experimental block consisted of each of these 10 competitions replicated 10-fold. Each  
7 experimental block was repeated at four different dates.

8 **Flow cytometry:** The relative frequency of competitors marked with CFP or YFP was  
9 measured using a Gallios Beckman Coulter flow cytometer at 0 and 24 hours following  
10 mixing of competing genotypes. We decided to separate competitor populations only on the  
11 basis of their fluorescent markers, because CFP and YFP cell populations did not have the  
12 same distribution pattern on forward (FCS) vs. side scatter (SSC) plots. Thresholds were  
13 applied manually (since clustering algorithms often introduce more noise) on the CFP – YFP  
14 plots to determine the boundaries of each population (CFP, YFP, unmarked cells and doubled  
15 marked objects) as shown on Figure 1. These thresholds were the same for all competition  
16 plots because in such a constant environment, cell clusters were always localized in the  
17 same areas of the plot. The frequency of each marker type was calculated using CFP and YFP  
18 population counts only. Unmarked and double-marked populations represented  
19 approximately 0.2 and 1% of the total population, respectively. For simplicity, 'doublets'  
20 (objects composed of two cells) were excluded from our frequency estimates. CC, YY and CY  
21 doublets occur, but only the latter are detected in the C2 population (Figure 1).  
22 Furthermore, doublets may not form at random; doublets with the same color were often  
23 overrepresented (data not shown). Nevertheless, even considering these complications,  
24 ignoring doublets only introduces a bias on  $s$  measures proportional to  $s\varepsilon$ , where  $\varepsilon$  is the  
25 fraction of the CY population (C2 in Figure 1). Under our conditions,  $\varepsilon \approx 1\%$  making this bias  
26 negligible compared to  $s$  (see supplementary materials for details).

## 27 **Precision of frequency measures with cytometry**

28 Our method is based on measuring the relative frequency  $p$  of two competing genotypes at  
29 different time points by counting  $C = 200000$  cells. This large figure, however, still represents



1 a small fraction of the total population and therefore, we estimate frequencies with  
 2 sampling error. The theoretical expectation for this sampling error is

$$3 \quad \sigma_e^2 = p(1 - p)/C.$$

4 If nothing else contributes to measurement error, we should obtain this variance when  
 5 measuring repeatedly the frequency in a given test tube. Preliminary experiments (not  
 6 shown) indicated that much larger error could occur, in particular when test tubes were  
 7 insufficiently mixed. This is an important technical issue and comparing actual measurement  
 8 error to  $\sigma_e^2$  provides an internal check that measurement error is not inflated above the  
 9 sampling error expectation. In the experiments presented here, we used measures of initial  
 10 frequencies ( $p_0$ ) in our replicated competitions to estimate the variance of frequency  
 11 measures performed with cytometry  $\sigma_{obs}^2$ . We found that  $\sigma_{obs}^2/\sigma_e^2$  was 0.94, 1.83, 1.07, and  
 12 0.95 for the four different dates where all the competitions were performed. Except for date  
 13 2, measurement error was very close to that inherent to sampling only. However as shown  
 14 by  $\sigma_{obs}^2$  at date 2 (and other preliminary assays showing more dramatic results), using  
 15 cytometry does not guarantee that measurement error will be low. In particular, thorough  
 16 mixing of test tubes throughout the growth cycle limits cell aggregation and is a crucial step  
 17 in taking advantage of the advantages offered by the cytometric approach (or any other  
 18 approach based on frequency variation).

### 19 **Measure of genotypic fitness**

20 We measured fitness based on a continuous time model  $dp/dt = s p (1 - p)$ , which defines  
 21 selection coefficient ( $s$ ) based on frequency ( $p$ ) variation. This frequency variation was  
 22 measured in the competition experiments described above. For a given competition assay  $k$ ,  
 23 the data is a vector  $\mathbf{n}_k = \{n_{1k}^0, n_{2k}^0, n_{1k}^t, n_{2k}^t\}$  giving the number of genotypes 1 and 2 counted  
 24 at time 0 (beginning) and  $t$  (end of the competition). The log-likelihood of this data given  
 25 initial frequency of genotype 1  $p_{1k}^0$  and selection coefficient  $s_k$  is computed as

$$26 \quad \ln \Pr(\mathbf{n}_k | p_{1k}^0, s_k) = \sum_{i=1,2} \sum_{j=0,t} (n_{ik}^j \ln p_{ik}^j) , \quad (1)$$

27

28 where  $p_{2k}^j = 1 - p_{1k}^j$  and

1

2

$$p_{1k}^t = e^{s_k t} p_{1k}^0 / (p_{2k}^0 + e^{s_k t} p_{1k}^0). \quad (2)$$

3

4 The frequency variation is measured over 24 h. In order to scale fitness measurements ‘per  
 5 generation’, we used the number of cell generations as the time unit. This measure is an  
 6 average over the duration of the competition, which does not contradict the fact that  
 7 conditions change with time in a given assay (e.g. the glucose becomes limiting) because it  
 8 does so similarly in all replicates. Because populations expand by binary fission, we have  $t =$   
 9  $\ln(100)/\ln(2) = 6.6$  in eq. (2). Across replicates of the same competition,  $s_k$  might vary for  
 10 reasons other than sampling error, owing, for example, to drift or to cryptic environmental  
 11 variation. To measure this variation, we used the same logistic regression approach (eqn 1-  
 12 2), but including the assumption that  $s$  was normally distributed  $s \sim N(\bar{s}, \sigma_s)$  among  
 13 replicates. The log-likelihood of this logistic regression with random slope is then

14

$$\ln \Pr(\mathbf{n} | \mathbf{p}_1^0, \bar{s}, \sigma_s) = \sum_k \ln \int_{-\infty}^{\infty} N(\bar{s}, \sigma_s; s) \prod_{i=1,2} \prod_{j=0,t} n_{ik}^j \ln p_{ik}^j ds, \quad (3)$$

16

17 where  $\mathbf{n}$  is the data matrix  $\{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \dots\}$ ,  $\mathbf{p}_1^0$  the vector of all initial frequencies and  
 18  $N(\mu, \sigma; x)$  denotes the probability density function of the Normal distribution with mean  $\mu$   
 19 and standard deviation  $\sigma$ . In all cases, parameters were estimated by maximizing the log-  
 20 likelihood. Support limits for a given estimate were computed within 2 units of log-likelihood  
 21 from the maximum with all other parameters being freely fitted. An equivalent of ‘standard  
 22 error’  $SE_{eq}$  was computed as a quarter of the support range (similarly, 95% confidence  
 23 intervals are  $\pm 1.96$  SE). Computations were done using Mathematica (WOLFRAM RESEARCH  
 24 2008).

## 25 **Fitness transitivity, allelic fitness and epistasis**

26 To test whether a constant fitness can be attributed to a genotype irrespective of its  
 27 competitor, we performed all possible combinations of competition assays for a given

1 mutant. At a first locus we have the wild type ( $w$ ) and mutant ( $m$ ) alleles. At a second locus  
2 we have two alleles  $c$  and  $y$  (corresponding to the CFP and YFP marker proteins,  
3 respectively). Each competition assay requires competing genotypes to have different alleles  
4 at the marker locus. There are thus four possible combinations: (a)  $wc/wy$ , (b)  $mc/my$ , (c)  
5  $my/wc$ , and (d)  $mc/wy$ . Table 2 indicates the selection coefficient expected in each of these  
6 cases if we assume that the fitness of genotypes  $wc$ ,  $wy$ ,  $mc$  and  $my$  are constant and equal  
7 to  $W_{wc}$ ,  $W_{wy}$ ,  $W_{mc}$ , and  $W_{my}$ , respectively. When measuring the marker effect in the same  
8 background (competitions (a) and (b)), we measured the selection coefficient of the CFP  
9 genotype. Otherwise, we measured the selection coefficient of the mutant genotype against  
10 the wild type (competitions (c) and (d)).

11 Population genetics models usually assume that fitness effects are transitive, *i.e.* that they  
12 could be deduced from some absolute value ranking of the different genotypes. However  
13 this is an assumption that requires evaluation before attributing a selection coefficient to  
14 genotypes. Since competitions (a), (b) and (c) are sufficient to estimate all fitness if they are  
15 transitive, competition (d) can be used to measure departure from transitivity. Specifically,  
16 we introduce a parameter  $\tau$  measuring this departure (see table 2). Further re-  
17 parameterization allows decomposing genotypic fitness into allelic effects and their  
18 interaction (epistasis). We note  $W_{wc} = W_{wy} + \alpha$ ,  $W_{my} = W_{wy} + \beta$ ,  $W_{mc} = W_{wy} + \alpha + \beta + \gamma$ .  $\alpha$  is  
19 the “cost” of the CFP marker,  $\beta$  is the selective effect of the mini-Tn10 mutation and  $\gamma$  is the  
20 epistasis between the two loci.

21 To fit this model, for each mutant, we used eq. (1) summed over the four competition assays  
22 and their replicates, with the parameterization indicated above. Support limits for estimates  
23 were computed within 2 units of log-likelihood all other parameters being freely fitted.

## 24 **Expected amount of drift**

25 In our experiments, population size increases by binary fission. To compute the variance in  
26 frequency introduced by drift, we first determine that each bacteria division increases this  
27 variance by a quantity  $pq/n^2$ , where  $n$  is the population size at the time of this division. We  
28 then sum this variance to the end of population growth:

29

1 
$$\sum_{n=n_i}^{n_f} \frac{pq}{n^2} = pq \left( \frac{1}{n_i} - \frac{1}{n_f} \right) + O \left( \frac{1}{n_i} \right)^2 . \quad (4)$$

2

3 We thus expect the variance of selection coefficients contributed by drift to be

4

5 
$$\sigma_s^2 = \frac{n_f - n_i}{g^2 n_f n_i pq} , \quad (5)$$

6

7 where  $g$  is the number of ‘generations’ (6.6 as explained above over the time course of the  
 8 competition experiment). The variance in frequency change caused by selective differences  
 9 among replicates is  $\text{var}(sgpq) = \sigma_s^2 (gpq)^2$ . In our experiments we have  $n_f$  in the range  
 10  $10^8 - 10^9$  and  $n_i$  100 times less. These population sizes were estimated by serial dilution  
 11 and plating (not shown), and these numbers represent the extreme cases. Thus we expect  $\sigma_s$   
 12 to be between  $10^{-4}$  and  $3 \times 10^{-4}$ . Significantly larger  $\sigma_s$  would indicate that a source of  
 13 variation, in addition to sampling error and drift, contributed to differences among replicate  
 14 competitions (*e.g.* such as random fluctuations in selection coefficients among replicates).

15

## RESULTS

16 We used a flow cytometric approach to measure the fitness of three mutants, each carrying  
 17 a single mutation, that were classified as being ‘neutral’ with conventional methods (ELENA *et*  
 18 *al.* 1998). We did 10 types of competition assays, each replicated 10-fold at each of four  
 19 weeks, giving a total of 400 fitness measures (Figure 2). A standard analysis of deviance (eq.  
 20 1-2) revealed that 96.6% of the deviance was among competition assay types. There were  
 21 significant week (0.6% of the total deviance), week  $\times$  competition (1.3%) and replicate (1.4%)  
 22 effects, although they accounted for a very small fraction of the total deviance. In particular,  
 23 although detectable, the week effect was smaller than the replicate effect (well-to-well  
 24 variation for the same competition during the same week), indicating that the experiments  
 25 were repeatable from one week to another. We also used a simple one-way ANOVA to test  
 26 whether the standard deviation in  $s$  measures among replicates was consistent when  
 27 measured at different weeks. This was the case, although the repeatability was not

1 extremely high. Specifically, we found that 53% of the variance in this standard deviation  
2 was among competitions and 47% within competition between weeks. This variation among  
3 competitions is significantly larger than across dates for the same competition ( $F_{9,30} = 3.7$ ,  $P$   
4  $= 0.003$ ). Repeatability of means and variance at different weeks is crucial for measuring  
5 fitness with precision: it is fairly easy to obtain a very precise measure of frequency change  
6 in a single assay (or even an exact measure if all individuals in the competition are counted  
7 at the beginning and the end), but this is not equivalent to obtaining an accurate measure of  
8 fitness, which must account for inter-replicate variance. As this example shows, analyzing a  
9 very large dataset also provides sufficient statistical power to detect very small biological  
10 effects, but it can also reveal 'nuisance' effects (almost anything tested becoming  
11 'significant'). To cope with these issues, we used an approach quantifying variance  
12 components in a mixed model (Eq. 3 in methods).

13 **Precision of fitness measures:** Competition assays provide a direct measure of the fitness of  
14 one genotype relative to a competing genotype. To determine if the differences we  
15 observed in fitness estimates between replicate competitions was biologically meaningful, as  
16 opposed to a sampling effect, we used a mixed model to directly estimate the amount of  
17 variation in  $s$  ( $\sigma_s$ ) beyond sampling error (eq. 3). This approach provides an estimate of  
18 average selection intensity ( $\bar{s}$ ), a measure of biological heterogeneity in selection among  
19 replicates ( $\sigma_s$ ) and standard errors associated with these two parameters. Beyond estimating  
20 average selection ( $\bar{s}$ ) with some precision, it is important to indicate the magnitude of  
21 variation in  $s$  ( $\sigma_s$ ) and the precision reached to estimate it. Table 3 presents these estimates  
22 for our 10 competition assays. Estimates of  $\bar{s}$  range from 0.00088 (T121 YFP) to -0.024 (T103  
23 CFP). Estimates of  $\sigma_s$  range from 0 to 0.0035, with seven of ten estimates being greater than  
24 zero. In all cases the precision of these estimates is about  $\pm 0.0002$ .

25 **The origin of variation in  $s$  among replicates:** There are four non-exclusive reasons that  
26 changes in the frequency of reference and mutant genotypes during competitions could be  
27 'truly' different among replicates: (1) experimental error unrelated to sampling (e.g.,  
28 pipeting), (2) new mutations occurring in some replicated competitions, (3) drift, and (4)  
29 variation in selection intensity among replicates (e.g., due to cryptic environmental  
30 variations). We consider each possibility in turn.

1 Experimental error is unlikely to be the source of the variation in  $s$  in our experiment. We  
2 repeated each competition type at four dates and  $\sigma_s$  was consistently low and comparable  
3 to the drift expectation in competitions (a) and (b) (Fig. 2). It is unlikely that systematic error  
4 would only occur for some competition types and even more unlikely that this pattern would  
5 be repeatable at different dates. The second hypothesis is that new deleterious or beneficial  
6 mutations, unrelated to the mutation of interest, may occur during the competition and  
7 influence the outcome. The case of deleterious mutations is not really problematic, because  
8 they are unlikely to reach high frequency in a large population and because, if many occur,  
9 they will occur equally in the two competing genotypes. The case of beneficial mutations  
10 may, at first sight, seem trickier. Let us consider a worst-case scenario of the early  
11 occurrence of a beneficial mutation providing a growth advantage of 10% per division. If we  
12 consider the appearance of this mutant at the very start of the pre-culture (*i.e.*, ~17  
13 generations before the start of the competition assay), its frequency at the end of the  
14 competition will be  $< 10^{-5}$  (assuming a competition of 6.6 generations and an effective  
15 population size of  $10^6$  as used in this study), which is too low to have any impact on our  
16 measures. Significant frequency variation (above ~0.02% in our case) would require a  
17 mutation to confer a benefit greater than ~30% (see supplementary materials for details),  
18 which is very unlikely in a strain that has been adapted to the environment for 10,000  
19 generations and for which no such mutations have been identified during the early stages of  
20 this adaptation, when fitness increases were most rapid (BARRICK *et al.* 2009). Moreover,  
21 even if such large effect mutations were available to our strains, they would have to occur  
22 repeatedly in many competitions because our observed  $\text{var}(s)$  is not due to isolated outliers  
23 (Figure 2). We note that some mutation types, notably genomic amplifications, have been  
24 observed to occur at high frequencies and may sometimes confer beneficial effects either  
25 directly or indirectly by increasing the mutational target for new mutations to occur.  
26 However, if these mutations occur at a very high rate, they would occur in both competitors  
27 and thus have a limited effect on  $\text{var}(s)$ . Furthermore, if the occurrence of *de novo* genomic  
28 amplifications were increasing  $\text{var}(s)$  in our experiments, they should do so in all  
29 competitions types, and not only in competitions (c) and (d) (Figure 3). In summary, we  
30 conclude that the rise and spread of new mutations is very unlikely to explain our results.

1 Drift can also cause variation in genotype frequency changes in the different replicates. This  
2 process scales with the inverse of population size and should effectively vanish in very large  
3 populations. In our experiments, we expect  $\sigma_s$  to be between  $10^{-4}$  and  $3 \times 10^{-4}$  if it was  
4 due to drift alone (see material and methods). Our estimates of  $\sigma_s$  (Table 3) varied among  
5 the competition assays. In competition (a) and (b), estimates of  $\sigma_s$  were not different from  
6 the maximum value that would be expected because of drift ( $3 \times 10^{-4}$ , Figure 3). These  
7 competition assays correspond to CFP versus YFP competitions within the same genetic  
8 background. We thus conclude that the cost of expressing the different fluorescent proteins  
9 is not significantly affected by uncontrolled cryptic environmental variation in our  
10 experiments. Other estimates of  $\sigma_s$  (in competitions (c) and (d)) are much larger than the  
11 drift expectation (Figure 3). One possibility is that the effect of drift is greater than expected  
12 from consideration of population size alone. This may be the case if there was substantial  
13 phenotypic diversity in the competing populations so that a subset of the population  
14 contributed disproportionately to population growth. In fact, this explanation seems  
15 unlikely. We find a typical value of  $\sigma_s$  of about 0.001, which would require that  $N_e$  was  
16 reduced to ~9% of the actual population (from eq. 5) (Figure 3). This means that drift can  
17 only explain our observed  $\sigma_s$  if more than 90% of the sampled population is not growing. (In  
18 the most extreme case, less than 1% of the population would have to be growing (T63,  
19 competition c)). Studies performed on *E. coli* populations showing that only a few percent of  
20 the total population were in an “atypical” non-growing physiological state during  
21 exponential population growth (BALABAN *et al.* 2004; LEVIN and ROZEN 2006), support the  
22 conclusion that phenotypic variation is not sufficient to account for variance among  
23 replicates in some of our competitions.

24 In cases where variation is too high to be explained by drift (competitions (c) and (d)),  
25 variation necessarily implies that selection intensity changes slightly among replicates,  
26 perhaps due to environmental variation among replicates. Furthermore,  $\sigma_s$  estimates were  
27 larger for large  $|\bar{s}|$  (Pearson  $r = 0.69$ ), a situation that would be expected when different  
28 competitors have environmental tolerance curves with different slopes (*i.e.* a GxE effect). In  
29 this case environmental variation will not necessary impact both competitors with the same  
30 intensity. Thus, small environmental variations across replicated competitions, can have a  
31 non negligible impact on  $\sigma_s$ . Both the high values of  $\sigma_s$  (compared to the drift expectation)

1 and its pattern of variation (larger in assays with large fitness differences) support the  
2 conclusion that, even under very controlled and standardized conditions, cryptic  
3 environmental variation has a detectable impact on fitness measures.

4 **Fitness transitivity:** Population genetic models of selection usually consider fitness effects to  
5 be transitive between competing genotypes. In this view, fitness can be associated with a  
6 given genotype rather than being defined locally relative to particular competitors. (There  
7 are, of course, particular frequency dependent selection schemes that can generate non-  
8 transitive fitness measures (e.g. SINERVO and LIVELY 1996).) Methodologically, transitivity is  
9 also an important assumption in inferring allelic from genotypic fitness effects, as when  
10 using a marker to infer the effect of a mutation. Our experimental design allows us to test  
11 for departures from transitivity because we measured relative fitness in four combinations  
12 of genotypes pairs (see methods). Specifically, to test for transitivity, we need to make three  
13 estimates for a mutation: (1) the allelic cost of the marker, (2) the allelic effect of the  
14 mutation and (3) the epistasis between both. With three competitions, we have three  
15 equations and three unknowns. Thus, adding one competition adds one equation and  
16 provides a means to estimate a departure from consistency (*i.e.*, transitivity) among  
17 competition types. We found that  $\tau$ , a parameter measuring deviations from transitivity, was  
18 not significantly different from zero for competitions involving the T63 and T121 mutants  
19 (LRT; Table 4), meaning that fitness was transitive. By contrast,  $\tau$  was significantly different  
20 from zero for T103, but this departure was quite small ( $\tau = -0.00171 \pm 0.0003$ ) and, more  
21 importantly, very small compared to the fitness differences measured in those competition  
22 assays (Table 4).

23 **Allelic fitness and epistasis:** To test for epistasis between our markers and the focal  
24 mutations, we decomposed genotypic fitness into the allelic effects of the marker and the  
25 mutation, and their interaction (epistasis). The expression of CFP was more costly than YFP  
26 (a 0.4% difference in the wild type) and the allelic effect of mutations was -1.2%, -1.7% and -  
27 0.3% for T63, T103, T121, respectively (Table 4). However, we detected significant  
28 differences between fitness effects of the same mutations when measured in the CFP and  
29 YFP backgrounds, indicating the existence of epistasis between the marker and the three  
30 individual mutations. Even though the strength of epistatic interactions was quite small, it  
31 could represent an important part of the genotypic selection coefficients. For instance, for



1 T63, epistasis was larger than the cost of the marker and represented 43% of the allelic  
2 mutational effect. For the two other mutations, the quantitative importance of epistasis was  
3 much smaller. A caveat to our interpretation of epistasis is that it is possible that we  
4 inadvertently introduced secondary mutations into genotypes during some step required for  
5 strain construction (see methods). Since we observed fitness differences with the three  
6 mutants we tested, the hypothesis of secondary mutation introduction supposes a very high  
7 rate of such mutations during P1 transduction.

## 8 **DISCUSSION**

9 In a large population, even mutations with very small fitness effects can play a role in  
10 the process of adaptation. However, studying them empirically is a significant practical  
11 challenge. Measurement error and drift obviously limit the precision of fitness measures  
12 that can be obtained experimentally. Is it possible to measure selection up to a limit imposed  
13 by the noise produced by sampling error and drift? If not, how close to this limit can we go?  
14 We addressed these questions by performing competition experiments in large *E. coli*  
15 populations (to minimize drift) and by tracking frequency changes using flow-cytometry to  
16 count marked cells (to minimize sampling error).

17 Our experiments are based on short-term batch cultures (6.6 generations). This design has  
18 several convenient features. First, it is a relatively simple experimental set up that can be  
19 massively replicated. Second, it reduces, though does not eliminate, the complication of  
20 newly arising mutations. Third, it entirely accounts for the effect of the marker. Last, it  
21 avoids the complication of using time series data.

22 **Cryptic variation in *s*:** A surprising, and we think important, result was that, for some  
23 competitions types, selection was variable across replicates, probably because of cryptic  
24 environmental variation to which the competing genotypes had different sensitivity.  
25 Although not empirically excluded, the alternative hypothesis that variation in estimates of *s*  
26 was caused by beneficial mutations spreading in a large number of the batch cultures, seems  
27 unlikely for two reasons: (1) such mutations would have to confer a very large benefit  
28 (unlikely to appear in a strain that has evolved in the same environment for 10,000  
29 generations) and (2) adaptive mutations would increase var(*s*) in all competition types, not  
30 only in competitions (c) and (d). Such variation arose despite considerable effort to perform

1 all competitions in precisely controlled conditions. In an absolute sense, this variation was  
2 not large (although much larger than our precision), but it supports the idea that the effect  
3 of mutations can be strongly context-dependent. For instance, it is possible that if our  
4 experiment was performed in a different lab, the  $\bar{s}$  and  $\sigma_s$  might be slightly different  
5 (because of differences in the average environment or in the magnitude of  
6 microenvironmental fluctuations, respectively). *A fortiori*, we expect  $\sigma_s$  to be even larger in  
7 environmentally heterogeneous natural conditions. These observations raise the question of  
8 whether selection coefficients should be described by only their mean values  $\bar{s}$ , or more  
9 appropriately by distributions (with two parameters  $\bar{s}$  and  $\sigma_s$ ), and consequently if mutations  
10 are appropriately described as beneficial, neutral or deleterious, since their effects are  
11 context-dependent, even within controlled laboratory environments. So far, population  
12 genetic models do not typically consider that  $s$  values are distributed, such that one  
13 mutation can have very different fates depending on its  $\sigma_s$ . For instance, the probability of  
14 fixation of a mutation with  $\bar{s} = 0$  and  $\sigma_s > 0$  will not be driven by drift only, as described by  
15 the neutral theory, but will depend on the environmental pattern responsible for  $\sigma_s > 0$  (see  
16 *e.g.* (EWENS 1979)). In any case, much more attention should be paid to variable selection  
17 coefficients and their evolutionary impact. The experimental design and statistical analysis  
18 we propose here, offers an efficient and new approach to do that.

19 **Epistasis with the marker:** In competition experiments with microbes, the neutrality of the  
20 marker is always verified, however, the potential epistatic interactions between the marker  
21 and mutations is not usually systematically investigated. Controlling for this issue requires  
22 switching the markers between backgrounds (in order to perform complementary  
23 competition assays (DYKHUIZEN and HARTL 1980)) and a high level of precision. We found  
24 epistatic interactions between the inserted mutations and the fluorescent marker in all  
25 cases, suggesting that epistatic effects, though perhaps very small, may be common. Such  
26 interactions complicate measures of  $s$  because they require separating the MFE from the  
27 marker cost and the epistatic interactions between them. We note that if we had found that  
28 epistatic effects were of a similar size to (or larger than) the allelic effects it would have  
29 raised the concern that the compared strains may not have been isogenic. This was not the  
30 case in our experiment, nevertheless, we cannot formally exclude the possibility that

1 transformation and P1 transduction manipulations did not introduce any secondary  
2 mutations.

3 **Transitivity:** The assumption of fitness transitivity is made in most population genetic  
4 models that do not specifically include social effects or frequency dependence. This  
5 assumption has been evaluated on several occasions and in different organisms (BELL 2008b;  
6 DE VISSER and LENSKI 2002; GOODMAN 1979; PAQUIN and ADAMS 1983; RICHMOND *et al.* 1975). The  
7 main conclusion is that fitness tends to be transitive unless special social interactions are  
8 present. Like for any ‘null hypothesis’, it is, however, important to realize that the statistical  
9 power of an experiment gives an inherent limit to the detectable departure of transitivity. In  
10 our experiment, we tested this hypothesis and did not find consequential departures from  
11 transitivity in the genotypic fitness measures (Table 4). Given our high statistical power, this  
12 finding represents a strong internal check that our experimental results are robust. The fact  
13 that fitness effects are transitive is also an important result to simplify experiments: without  
14 the need to check for transitivity, only three types of competition need to be performed to  
15 estimate allelic effects and epistasis (instead of four in our design).

16 **Precision of fitness measures and the statistics of selection:** Although sampling error and  
17 drift can make it difficult to measure small fitness effects, replicated measures will tend not  
18 to significantly differ from each other. Consequently, a single fitness value can be  
19 legitimately attributed to a given genotype in a given environment and the precision of this  
20 estimate can be determined as the standard error of the mean fitness effect across  
21 replicates. New high-throughput counting methods alleviate limits due to sampling error and  
22 drift. Here, we show that such methods can reveal that replicated measures differ from one  
23 another for a given genotype in a given environment—i.e. that  $\text{var}(s)$  is significantly greater  
24 than the value expected by drift and sampling error alone. This situation challenges the  
25 simple concept of precision mentioned above. When confronted with this problem, one  
26 approach is to do “as usual” and neglect the observation that replicates differ. In this case,  
27 only providing a mean fitness effect and its standard error do not reflect the actual precision  
28 of the experiment. In particular, it fails to acknowledge that replicates may differ beyond this  
29 standard error. The second approach is to admit that replicates actually differ and represent  
30 different draws in a distribution of fitness effects, even if the environment is supposed  
31 constant. If selection coefficients are distributed, it is thus necessary to measure the mean

1 effect, but also the variance, and possibly higher moments (skewness, kurtosis, etc.) of the  
2 distribution of  $s$  values. The concept of precision in this case must incorporate estimates of  
3 these moments and their standard errors. This is the approach we have taken, introducing  
4 the use of a mixed model that allowed us to decompose sources of variation in frequency  
5 change (sampling error, drift, environmental variation of  $s$ ).

6 We measure fitness based on frequency change as in classical population genetics (ARNASON  
7 and BARKER 1999; DYKHUIZEN and HARTL 1980), which differs from common practices in  
8 experimental evolution (CHEVIN 2010), where fitness is measured as a ratio of growth rates  
9 (e.g. LENSKI *et al.* 1991). We also fully use the information on the individual precision of  
10 fitness estimates (determined by the sampling effort: the number of colonies counted when  
11 plating, or the number of cells counted with flow cytometry), which is not usually reported  
12 with fitness competition experiments. Thus, we can discriminate between sampling error  
13 and other sources of variance across replicates (due to drift, variance in  $s$  etc.), which greatly  
14 enhances the information that can be extracted from the data. By considering these factors,  
15 we were able to measure mean and variance in selection coefficients down to a precision of  
16 0.02% (Tables 3 and 4). Regarding mean selection, this precision represents a ~10-fold  
17 improvement over typical studies using flow cytometry (ALI and YANG 2006; LEE *et al.* 2009;  
18 LUNZER *et al.* 2002) and is comparable to the precision reached in (ZHU *et al.* 2005). More  
19 importantly, as explained above, the massive replication we used also provides a precise  
20 measure of the variance of selection coefficient  $\sigma_s$  ( $\pm 0.02\%$ ).

21 **Neutralist vs. selectionist:** The neutral theory of molecular evolution proposes that the fate  
22 of many mutations is governed by the effect of drift (KIMURA 1983). The development of  
23 precise fitness measures was used in the neutralist / selectionist controversy on proteins in  
24 order to determine if allozymes differed in terms of selection (DYKHUIZEN and HARTL 1980).  
25 Today, this debate has shifted towards smaller fitness effects at the molecular level  
26 (KREITMAN 1996; NEI 2005). The analysis of sequence polymorphism provides different  
27 indirect ways to confront neutralist vs. selectionist expectations. For most mutations, it is  
28 usually thought that there is no alternative to resolve this question. Measuring  $s$  with ever  
29 increasing precision may start to change this perspective and may help to answer some of  
30 the key questions fueling this debate. As we have already seen in our study, mutations  
31 formerly considered as neutral (albeit in a slightly different medium, DM25 in (ELENA *et al.*

1 1998)), actually confer small, but significant, fitness effects. Importantly, even in an  
2 apparently constant environment, their effect is best understood as being distributed, which  
3 complicates straightforward application of discrete classifications (deleterious / neutral /  
4 beneficial), and which would have to be accounted for in theoretical expectations e.g. for  
5 the analysis of sequence polymorphism.

6 Sampling error, *de novo* beneficial mutations, and drift introduce elements of chance into  
7 fitness competitions, which can limit our ability to measure very small fitness effects. The  
8 effect of these factors can, however, be reduced. Sampling error can be dramatically  
9 decreased using flow cytometry, and the problem of the occurrence of new beneficial  
10 mutations and of drift is reduced by using very short term batch cultures and high  
11 replication. However, a remaining issue is the variation in selection due to micro-  
12 environmental variation across replicated cultures. While this would not be surprising if  
13 replicate measurements had been obtained from different environments (e.g. REMOLD and  
14 LENSKI 2001), that was not the case in our experiment in which all competitions were carried  
15 out in an environment that was kept as consistent as possible. Increased sampling effort,  
16 larger population sizes or longer lasting experiments are not likely to resolve this issue. It is  
17 thus unclear how far precision in fitness measurement can be improved. It all relies on  
18 understanding the sources of variation, and controlling them, whenever possible. Would it  
19 be possible to measure the fitness effects of synonymous mutations or mutations occurring  
20 in non-coding sequences? We cannot answer these questions yet, however, our method is a  
21 step in that direction and it will certainly help to bridge the gap between studies measuring *s*  
22 experimentally and studies inferring *s* from genetic sequences (see EYRE-WALKER and  
23 KEIGHTLEY 2007 for review).

24

25 **Acknowledgements** : We would like to thank E. Flaven, M.-P. Dubois for lab management, C.  
26 Duperray (IRB – Montpellier), C. Mongellaz (IGM – Montpellier) and the Montpellier RIO  
27 Imaging platform, for training RG to flow cytometry, and their help in designing the  
28 cytometer protocols, N. Le Meur for her help with the flowCore R package, L.M. Chevin, P.A.  
29 Gros, G. Martin and F. Rousset for fruitful discussions and insightful comments. We also  
30 thank two anonymous reviewers for useful comments. This work was supported by the ERC  
31 Starting Grant ‘Quantevol’ to T.L. and NSF (IOS-1022373) to T.F.C.

1 **Figures and Tables**

2

Step	Experiment	Description	Achievement
1	PCR	Introduction of the P <sub>A1</sub> promoter in front of the CFP and YFP genes	PCR products called P <sub>A1</sub> -CFP and P <sub>A1</sub> -YFP
2	Cloning	Cloning P <sub>A1</sub> -CFP and P <sub>A1</sub> -YFP in pKD4*	plasmids called pKD4-CFP and pKD4-YFP
3	PCR	pKD4-CFP and pKD4-YFP were used as templates. Primers with 50 bases sequences homologous to the <i>E. coli rhaA</i> gene at their 5' ends were used.	PCR products called <i>rhaA</i> -CFP-Kan and <i>rhaA</i> -YFP-Kan
4	Transformation	The plasmid pKD46* was electroporated into the REL4548 recipient cells	REL4548 carrying pKD46
5	Transformation and homologous recombination	PCR products <i>rhaA</i> -CFP-Kan and <i>rhaA</i> -YFP-Kan were electroporated in REL4548 carrying pKD46.	REL4548 CFP-Kan <sup>R</sup> and REL4548 YFP-Kan <sup>R</sup>
6	Transformation	The plasmid pCP20* was electroporated into REL4548 CFP- Kan <sup>R</sup> and REL4548 YFP-Kan <sup>R</sup>	REL4548 CFP-Kan <sup>R</sup> pCP20 and REL4548 YFP- Kan <sup>R</sup> pCP20
7	Heat Shock	KanR cassette were excised from the REL4548 CFP- Kan <sup>R</sup> and REL4548 YFP- Kan <sup>R</sup> genomes	REL4548 CFP and REL4548 YFP

3 Table 1. Strain construction description. \*from (DATSENKO and WANNER 2000)

4

1

Genotypes	Wild type CFP (wc)	Wild type YFP (wy)	Mutant CFP (mc)
Wild type YFP (wy)	$s_a = W_{wc}/W_{wy} - 1$		
Mutant CFP (mc)		$s_d = W_{mc}/W_{wy} - 1 + \tau$	
Mutant YFP (my)	$s_c = W_{my}/W_{wc} - 1$		$s_b = W_{mc}/W_{my} - 1$

2

3 Table 2. Selection coefficient expected in the different combinations of competition assays:  
4 (a) *wc/wy*, (b) *mc/my*, (c) *my/wc*, and (d) *mc/wy*. Each combination was performed for each  
5 mutant strain.

6

$(\times 10^{-3})$	$\bar{s}$	inf	sup	$SE_{eq}$	$\sigma_s$	inf	sup	$SE_{eq}$
Wild type (a)	-4.13	-3.74	-4.53	0.20	0.80	0.25	1.27	0.26
T63 (b)	1.16	1.58	0.74	0.21	0.59	0.00	1.18	0.29
T63 CFP (c)	-12.23	-11.93	-12.63	0.18	3.51	3.20	3.82	0.16
T63 YFP (d)	-8.28	-7.28	-6.28	0.29	1.49	1.02	1.94	0.23
T103 (b)	-6.48	-6.14	-6.81	0.17	0.00	0.00	0.60	0.15
T103 CFP (c)	-24.26	-24.10	-24.82	0.19	2.84	2.52	3.13	0.15
T103 YFP (d)	-15.55	-14.55	-13.55	0.24	1.11	0.71	1.60	0.22
T121 (b)	-3.60	-3.39	-3.80	0.10	0.00	0.00	0.44	0.11
T121 CFP (c)	-6.40	-5.92	-6.87	0.24	1.15	0.75	1.65	0.23
T121 YFP (d)	0.88	1.88	2.88	0.21	0.87	0.35	1.33	0.25

1

2 Table 3. Estimation of the mean ( $\bar{s}$ ) and standard deviation ( $\sigma_s$ ) of genotypic selection  
3 coefficients per generation in the different competition assays (code in first column). inf and  
4 sup indicates the inferior and superior support limits of the estimates.  $SE_{eq}$  gives a measure  
5 analogous to standard error and equals (sup-inf)/4. All figures are multiplied by 1000.

6



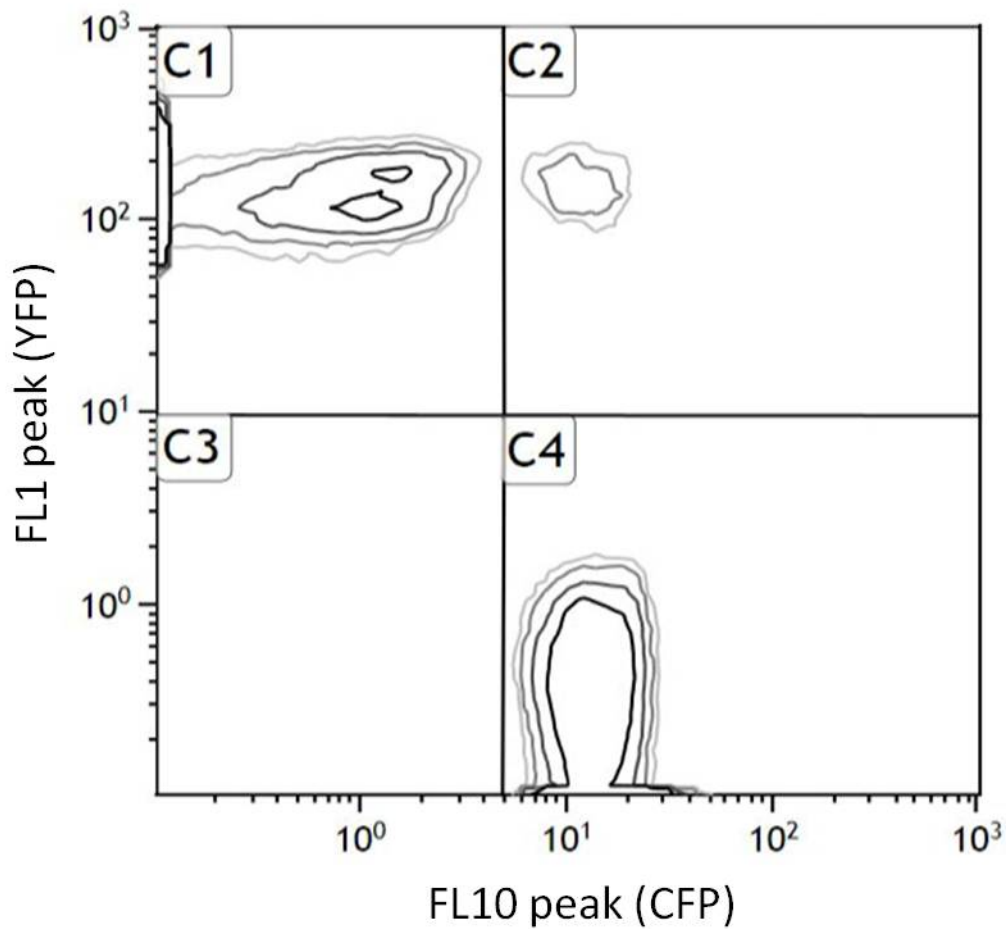
1

parameter	estimate ( $\times 10^{-3}$ )	$SE_{eq}$ ( $\times 10^{-3}$ )	sign.
$\alpha$	-4.13	0.14	***
$\beta_{63}$	-12.27	0.22	***
$\gamma_{63}$	5.28	0.22	***
$\tau_{63}$	-0.13	0.30	ns
$\beta_{103}$	-17.89	0.22	***
$\gamma_{103}$	-2.23	0.22	***
$\tau_{103}$	-1.71	0.30	**
$\beta_{121}$	-2.81	0.22	***
$\gamma_{121}$	0.55	0.22	*
$\tau_{121}$	-0.48	0.30	ns

2

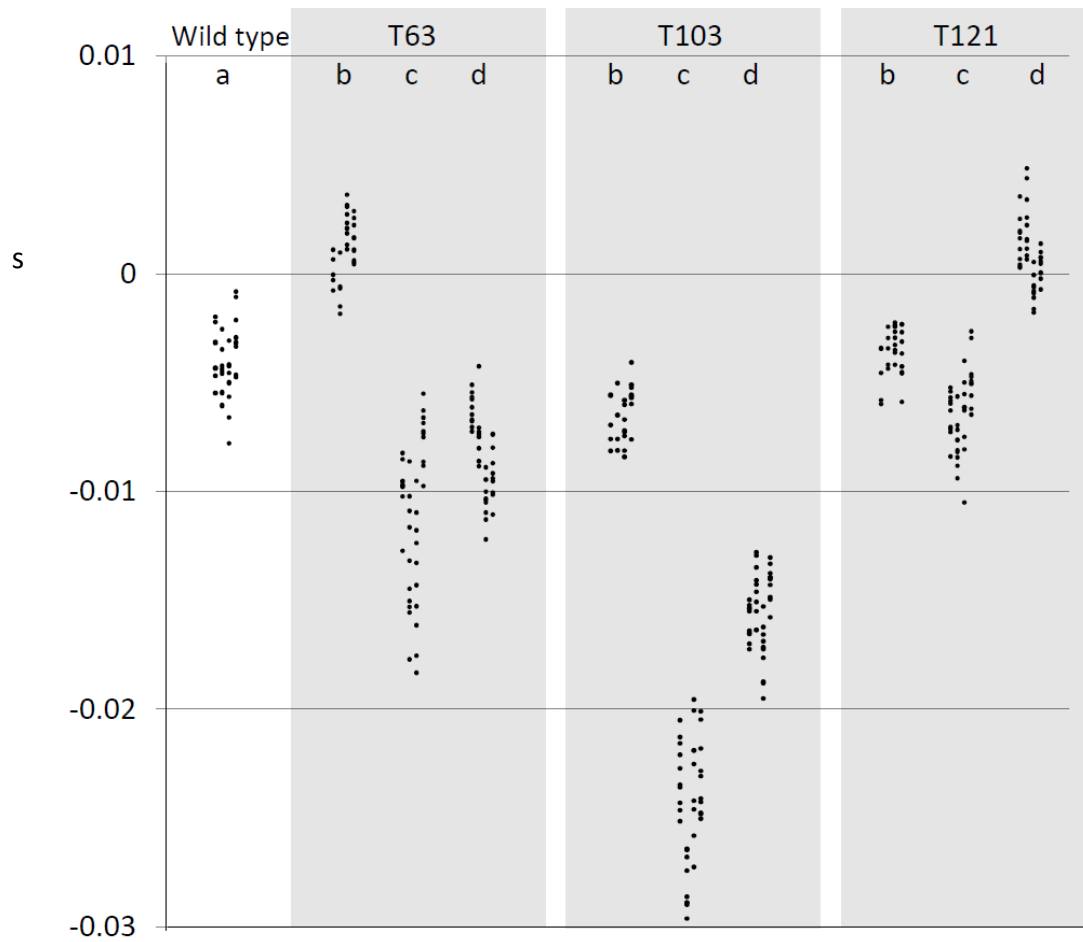
3 Table 4. Estimation of allelic selection coefficients per generation in the different  
4 competition assays (the subscript refers to the mutant T63, T103 or T121).  $\alpha$  is the cost of  
5 the marker (bacteria expressing the fluorescent proteins CFP having a 0.4 % cost relative to  
6 those expressing YFP).  $\beta$  are the allelic effects of the three random mutations.  $\gamma$  are the  
7 epistasis between the random mutations and the marker.  $\tau$  measures departure from  
8 transitive genotypic fitness.  $SE_{eq}$  gives a measure analogous to standard error and is a  
9 quarter of the support range. Significance (sign.) indicates whether the estimates are  
10 different from zero (LRT, \*\*\* < p-value 0.001, \*\* < P-value 0.01, ns: non significant). All  
11 figures are multiplied by 1000.

1



2

3 Figure 1. Measuring genotype frequencies with flow cytometry. The fluorescence of each  
4 bacterium was measured in the FL1 (YFP) and FL10 (CFP) channels. Here we show a  
5 representative contour plot. Quadrants represent thresholds delimiting the YFP (C1), CFP  
6 (C4), doubled marked (C2), and unmarked (C3) populations. In this example, populations are  
7 composed of YFP: 110718 (51.03%), CFP: 103996 (47.93%), doubled-marked: 2110 (0.97%),  
8 Unmarked 151 (0.07%) individuals.



1

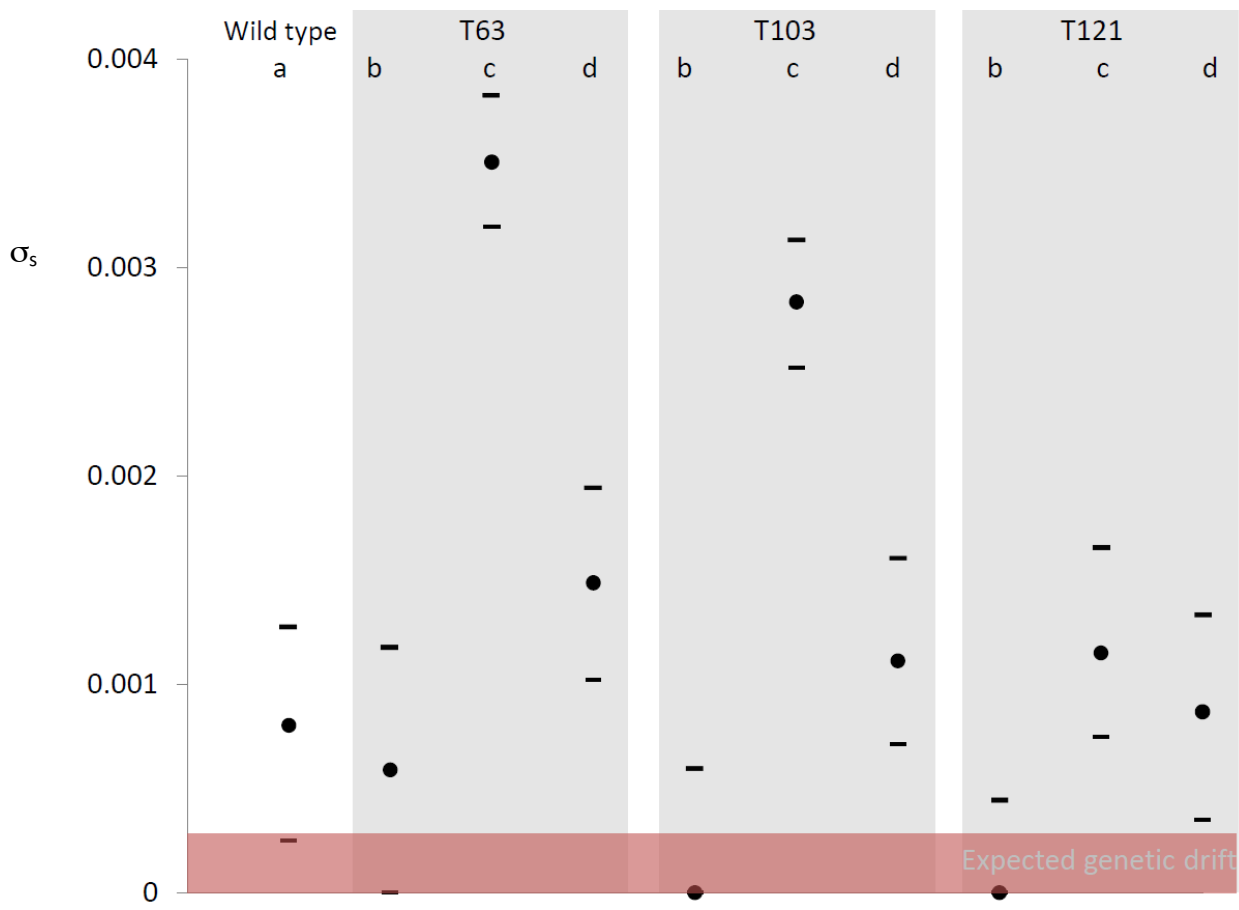
2 Figure 2. Selection coefficients ( $s$ ) of the wild type (REL4548) and mutant strains (REL4548  
 3 T63, T103, or T121) measured in the (a)  $wc/wy$ , (b)  $mc/my$ , (c)  $my/wc$ , and (d)  $mc/wy$   
 4 competitions. Each point represents an  $s$  measure estimated from a single competition  
 5 experiment.  $s$  measures are grouped in lines to show the variance between experiments  
 6 performed at different dates.

7

8

9

10



1  
 2 Figure 3. Variance in selection coefficients ( $\sigma_s$  estimate bold point  $\pm$  support limits indicated  
 3 by bars) of the wild type (REL4548) and mutant strains (REL4548 T63, T103, or T121)  
 4 measured in the (a) *wc/wy*, (b) *mc/my*, (c) *my/wc*, and (d) *mc/wy* competitions. The  
 5 predicted variation expected by genetic drift alone is represented by the shaded line at the  
 6 bottom of the figure.

7

## REFERENCES

- 1
- 2 ALI, A., and O. O. YANG, 2006 A novel small reporter gene and HIV-1 fitness assay. *Journal of*  
3 *Virological Methods* **133**: 41-47.
- 4 ARNASON, E., and J. BARKER, 1999 Analysis of selection in laboratory and field populations, pp.  
5 182-203 in *Evolutionary genetics: from molecules to morphology*, edited by R. SINGH  
6 and C. KRIMBAS. Cambridge University Press, Cambridge.
- 7 ARNASON, E., and R. C. LEWONTIN, 1991 Perturbation-reperturbation test of selection vs.  
8 hitchhiking of the 2 major alleles of esterase-5 in *Drosophila pseudoobscura*. *Genetics*  
9 **129**: 145-168.
- 10 BALABAN, N. Q., J. MERRIN, R. CHAIT, L. KOWALIK and S. LEIBLER, 2004 Bacterial persistence as a  
11 phenotypic switch. *Science* **305**: 1622-1625.
- 12 BARRICK, J. E., D. S. YU, S. H. YOON, H. JEONG, T. K. OH *et al.*, 2009 Genome evolution and  
13 adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**: 1243-U1274.
- 14 BATAILLON, T., and M. KIRKPATRICK, 2000 Inbreeding depression due to mildly deleterious  
15 mutations in finite populations: size does matter. *Genetical Research* **75**: 75-81.
- 16 BEATTY, J., 1984 Chance and natural selection. *Philosophy of Science* **51**: 183-211.
- 17 BELL, G., 2008a *Selection*. Oxford Univ. Press, Oxford.
- 18 BELL, G., 2008b *Selection - The mechanism of Evolution*, NYC, USA.
- 19 BELL, G., 2010 Fluctuating selection: the perpetual renewal of adaptation in variable  
20 environments. *Philosophical Transactions of the Royal Society B-Biological Sciences*  
21 **365**: 87-97.
- 22 BOLLBACK, J. P., T. L. YORK and R. NIELSEN, 2008 Estimation of 2N(e)s from temporal allele  
23 frequency data. *Genetics* **179**: 497-502.
- 24 CARRASCO, P., F. DE LA IGLESIA and S. F. ELENA, 2007 Distribution of fitness and virulence effects  
25 caused by single-nucleotide substitutions in tobacco etch virus. *Journal of Virology*  
26 **81**: 12979-12984.
- 27 CHARLESWORTH, B., and D. CHARLESWORTH, 1998 Some evolutionary consequences of  
28 deleterious mutations. *Genetica* **103**: 3-19.
- 29 CHEVIN, L.-M., 2010 On measuring selection in experimental evolution. *Biology Letters* **Online**  
30 **publication doi: 10.1098/rsbl.2010.0580**.
- 31 CLARK, A., 1979 Effects of interspecific competition on the dynamics of a polymorphism in an  
32 experimental population of *Drosophila melanogaster*. *Genetics* **92**: 1315-1328.
- 33 DATSENKO, K. A., and B. L. WANNER, 2000 One-step inactivation of chromosomal genes in  
34 *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of*  
35 *Sciences of the United States of America* **97**: 6640-6645.
- 36 DE VISSER, J. A. G. M., and R. LENSKI, 2002 Long-term experimental evolution in *Escherichia coli*.  
37 XI. Rejection of non-transitive interactions as cause of declining rate of adaptation.  
38 *BMC Evolutionary Biology* **2**: 19.
- 39 DOBZHANSKY, T., and O. PAVLOVSKY, 1957 An experimental study of interaction between genetic  
40 drift and natural selection. *Evolution* **11**: 311-319.
- 41 DOMINGO-CALAP, P., J. M. CUEVAS and R. SANJUAN, 2009 The fitness effects of random mutations  
42 in single-stranded DNA and RNA bacteriophages. *PLoS Genet* **5**: e1000742.
- 43 DYKHUIZEN, D., and D. HARTL, 1980 Selective neutrality of 6PGD allozymes in *E. coli* and the  
44 effects of genetic background. *Genetics* **96**: 801-817.
- 45 DYKHUIZEN, D. E., and D. L. HARTL, 1983 Selection in chemostats. *Microbiological Reviews* **47**:  
46 150-168.

- 1 ELENA, S. F., L. EKUNWE, N. HAJELA, S. A. ODEN and R. E. LENSKI, 1998 Distribution of fitness effects  
2 caused by random insertion mutations in *Escherichia coli*. *Genetica* **102-3**: 349-358.
- 3 EWENS, W. J., 1979 *Mathematical population genetics*, Berlin, Heidelberg, New York.
- 4 EYRE-WALKER, A., and P. D. KEIGHTLEY, 2007 The distribution of fitness effects of new mutations.  
5 *Nature Reviews Genetics* **8**: 610-618.
- 6 FELSENSTEIN, J., 1976 The theoretical population genetics of variable selection and migration.  
7 *Annual Review of Genetics* **10**: 253-280.
- 8 FISHER, R. A., 1930 *The genetical theory of natural selection*. Oxford University Press, Oxford.
- 9 GOODMAN, D., 1979 Competitive hierarchies in laboratory drosophila. *Evolution* **33**: 207-219.
- 10 KIMURA, M., 1983 *The neutral theory of molecular evolution*. Cambridge university press,  
11 Cambridge, UK.
- 12 KLECKNER, N., J. BENDER and S. GOTTESMAN, 1991 Uses of transposons with emphasis on Tn10.  
13 *Methods in Enzymology* **204**: 139-180.
- 14 KREITMAN, M., 1996 The neutral theory is dead. Long live the neutral theory. *BioEssays* **18**:  
15 678-683.
- 16 L'HERITIER, P., and G. TEISSIER, 1937a Elimination of the mutant forms in drosophila  
17 populations. Case of drosophila with the "ebony gene". *Comptes Rendus Des Seances*  
18 *De La Societe De Biologie Et De Ses Filiales* **124**: 882-884.
- 19 L'HERITIER, P., and G. TEISSIER, 1937b Elimination of the mutant forms in drosophila  
20 populations. The case of drosophila with the "bar gene". *Comptes Rendus Des*  
21 *Seances De La Societe De Biologie Et De Ses Filiales* **124**: 880-882.
- 22 LABBE, P., N. SIDOS, M. RAYMOND and T. LENORMAND, 2009 Resistance Gene Replacement in the  
23 Mosquito *Culex pipiens*: Fitness Estimation From Long-Term Cline Series. *Genetics*  
24 **182**: 303-312.
- 25 LEE, M. C., H. H. CHOU and C. J. MARX, 2009 Asymmetric, bimodal trade-offs during adaptation  
26 of methylobacterium to distinct growth substrates. *Evolution* **63**: 2816-2830.
- 27 LENORMAND, T., and M. RAYMOND, 2000 Analysis of clines with variable selection and variable  
28 migration. *The American Naturalist* **155**: 70-82.
- 29 LENORMAND, T., D. ROZE and F. ROUSSET, 2009 Stochasticity in evolution. *Trends in Ecology &*  
30 *Evolution* **24**: 157-165.
- 31 LENSKI, R. E., M. R. ROSE, S. C. SIMPSON and S. C. TADLER, 1991 Long-term experimental evolution  
32 in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American*  
33 *Naturalist* **138**: 1315 - 1341.
- 34 LEVIN, B. R., and D. E. ROZEN, 2006 Opinion - Non-inherited antibiotic resistance. *Nature*  
35 *Reviews Microbiology* **4**: 556-562.
- 36 LUNZER, M., A. NATARAJAN, D. E. DYKHUIZEN and A. M. DEAN, 2002 Enzyme kinetics, substitutable  
37 resources and competition: From biochemistry to frequency-dependent selection in  
38 *lac*. *Genetics* **162**: 485-499.
- 39 LYNCH, M., 1987 The consequences of fluctuating selection for isozyme polymorphisms in  
40 *Daphnia*. *Genetics* **115**: 657-669.
- 41 MANLY, B. F. J., 1985 *The statistics of natural selection*. Chapman and Hall Ltd, London.
- 42 MILLSTEIN, R. L., 2008 Distinguishing drift and selection empirically: "The great snail debate" of  
43 the 1950s. *Journal of the History of Biology* **41**: 339-367.
- 44 NEI, M., 2005 Selectionism and neutralism in molecular evolution. *Molecular Biology and*  
45 *Evolution* **22**: 2318-2342.

- 1 O'HARA, R. B., 2005 Comparing the effects of genetic drift and fluctuating selection on  
2 genotype frequency changes in the scarlet tiger moth. *Proceedings of the Royal*  
3 *Society B-Biological Sciences* **272**: 211-217.
- 4 OAKESHOTT, J. G., S. R. WILSON and J. B. GIBSON, 1983 An attempt to measure selection  
5 coefficients affecting the alcohol-dehydrogenase polymorphism in *Drosophila*  
6 *melanogaster* populations maintained on ethanol media. *Genetica* **61**: 151-159.
- 7 ORR, H. A., 2005 The genetic theory of adaptation: A brief history. *Nature Reviews Genetics* **6**:  
8 119-127.
- 9 PAQUIN, C., and J. ADAMS, 1983 Frequency of fixation of adaptive mutations is higher in  
10 evolving diploid than haploid yeast populations. *Nature* **302**: 495-500.
- 11 PERIS, J. B., P. DAVIS, J. M. CUEVAS, M. R. NEBOT and R. SANJUAN, 2010 Distribution of fitness  
12 effects caused by single-nucleotide substitutions in bacteriophage  $\phi$ 1. *Genetics* **185**:  
13 603-609.
- 14 REMOLD, S. K., and R. E. LENSKI, 2001 Contribution of individual random mutations to  
15 genotype-by-environment interactions in *Escherichia coli*. *Proceedings of the*  
16 *National Academy of Sciences of the United States of America* **98**: 11388-11393.
- 17 RICHMOND, R. C., M. E. GILPIN, S. PEREZSALAS and F. J. AYALA, 1975 Search for emergent  
18 competitive phenomena - Dynamics of multispecies drosophila systems. *Ecology* **56**:  
19 709-714.
- 20 ROSENFELD, N., J. W. YOUNG, U. ALON, P. S. SWAIN and M. B. ELWITZ, 2005 Gene regulation at the  
21 single-cell level. *Science* **307**: 1962-1965.
- 22 ROUSSET, F., 2004 *Genetic structure and selection in subdivided populations*. Princeton  
23 University Press, Princeton.
- 24 SACCHERI, I. J., F. ROUSSET, P. C. WATTS, P. M. BRAKEFIELD and L. M. COOK, 2008 Selection and gene  
25 flow on a diminishing cline of melanic peppered moths. *Proceedings of the National*  
26 *Academy of Sciences of the United States of America* **105**: 16212-16217.
- 27 SANJUAN, R., A. MOYA and S. F. ELENA, 2004 The distribution of fitness effects caused by single-  
28 nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of*  
29 *Sciences of the United States of America* **101**: 8396-8401.
- 30 SINERVO, B., and C. M. LIVELY, 1996 The rock-paper-scissors game and the evolution of  
31 alternative male strategies. *Nature* **380**: 240 - 243.
- 32 THATCHER, J. W., J. M. SHAW and W. J. DICKINSON, 1998 Marginal fitness contributions of  
33 nonessential genes in yeast. *Proceedings of the National Academy of Sciences of the*  
34 *United States of America* **95**: 253-257.
- 35 WILSON, S. R., J. G. OAKESHOTT, J. B. GIBSON and P. R. ANDERSON, 1982 Measuring selection  
36 coefficients affecting the alcohol-dehydrogenase polymorphism in *Drosophila*  
37 *melanogaster*. *Genetics* **100**: 113-126.
- 38 WOLFRAM RESEARCH, I., 2008 *Mathematica*, pp. in *Version 7.0*, Champaign, IL
- 39 ZHU, G. P., G. B. GOLDING and A. M. DEAN, 2005 The selective cause of an ancient adaptation.  
40 *Science* **307**: 1279-1282.
- 41
- 42

## SUPPLEMENTARY MATERIALS

1

### 2 **Insertions of the chromosomal fluorescent markers**

3 The YFP and CFP genes were inserted at the *rhaA* locus of REL4548 using the one-step  
4 inactivation of chromosomal genes technique developed by Datsenko & Wanner (DATSENKO  
5 and WANNER 2000). pZA32-YFP and pZE1R-CFP - plasmids kindly provided by Dr Michael  
6 Elowitz - were used as templates.

7 The first step of this construction was to introduce the P<sub>A1</sub> (bacteriophage  $\lambda$  promoter)  
8 upstream of the CFP and YFP genes. Because the YFP and CFP sequences (wild-type codons,  
9 developed by University of Washington Yeast Resource Center) only differ by 20 nucleotides  
10 (out of 717 bp), the same primers could be used for both PCRs. The primer P-BspHI-promA1-  
11 YC/FP for (109 bases) is composed of a 5' TCTC tail (increasing digestion efficiency by BspHI),  
12 a BspHI restriction site, the P<sub>A1</sub> promoter containing a ribosome binding site (rbs) and a 20 bp  
13 sequence homologous to the YFP and CFP genes 5' extremity (see supplementary materials  
14 for primer sequences and PCR mix). The P-Clal-YC/FP rev primer is composed of a 5' TCTC  
15 tail, a Clal restriction site and a 20 bp sequence homologous to the YFP and CFP genes 3'  
16 extremity. The resulting PCR products were named P<sub>A1</sub>-CFP and P<sub>A1</sub>-YFP.

17 The second step consisted in cloning the P<sub>A1</sub>-CFP and P<sub>A1</sub>-YFP in pKD4 (DATSENKO and WANNER  
18 2000). P<sub>A1</sub>-CFP and P<sub>A1</sub>-YFP were digested by BspHI and Clal, and then gel purified. pKD4 was  
19 first digested by BspHI and Clal, and then treated with shrimp alkaline phosphatase (SAP).  
20 The SAP catalyzes the release of 5'- and 3'-phosphate groups from DNA, precluding the  
21 ligation of the linearized plasmid with the remaining pKD4 BspHI-Clal fragments. Finally, P<sub>A1</sub>-  
22 CFP and P<sub>A1</sub>-YFP were cloned in pKD4 by using a T4 ligase. Ligation products called pKD4-CFP  
23 and pKD4-YFP were electroporated into DH5- $\alpha$   $\lambda$ -pir, and plated on LB plates supplemented  
24 with 50  $\mu$ g/mL Kanamycin (LBA-Kan). Fluorescent colonies were purified by streaking on  
25 fresh LBA-Kan plates, and the next day, a colony was inoculated in liquid LB-Kan.

26 After isolation with Qiagen miniprep kit, pKD4-YFP and pKD4-CFP plasmids were used as  
27 templates in a second PCR in which the promoter sequence, the fluorescent genes and the  
28 Kanamycin cassette carried by pKD4, were amplified (see DATSENKO and WANNER 2000 for  
29 more details). The primers used in this second PCR - p-pKD4-RhaAH1-2863-2888 for (84



1 bases) and p-pKD4p2H2-1496-1477rev (79 bases) – contained either a BamHI or a Sall  
 2 restriction site at their 5' extremities (restriction sites not used in this study), fifty bases  
 3 homologous of the *E. coli rhaA* locus and the 21 or 27 bases homologous to the template at  
 4 their 3' extremities. The resulting PCR products were electroporated into REL4548  
 5 competent cells carrying the thermosensitive plasmid pKD46 (DATSENKO and WANNER 2000)  
 6 and transformed cells spread on LBA-Kan (incubation at 37°C in order to get rid of pKD46).  
 7 Fluorescent recombinants were streaked on LBA-Kan for purification, and on LBA-Amp (50  
 8 µg/ml Ampicilin) to check for the loss of pKD46. PCRs were carried out on recombinant  
 9 clones with two sets of primers (with a forward primer p-pKD4-2863-2888 specific to the  
 10 insert, or p-RhaA-for specific to the *rhaA* gene, and the reverse primer RhaA-1262-1243-rev  
 11 specific to the *rhaA* gene) to check that the insertion of CFP-Kan and YFP-Kan cassettes  
 12 occurred at the right locus.

13 The excision of the kanamycin resistance cassette, corresponding to the last step of this  
 14 cloning, was performed by the electroporation and induction of the plasmid pCP20  
 15 (DATSENKO and WANNER 2000). Transformants were first cultivated at 30°C on LBA-Amp  
 16 plates. The flipase production (allowing the recombination of the FRT sequences surrounding  
 17 the Kanamycin cassette) was induced by streaking colonies from the LBA-Amp plates, on LB  
 18 plates at 42°C. To check for Kanamycin resistance cassette removal and loss of pCP20,  
 19 colonies were streaked on LBA, LBA-Kan and LBA-Amp plates. PCRs on the *rhaA* locus were  
 20 performed on 4548 CFP Kan<sup>S</sup> Amp<sup>S</sup> and 4548 YFP Kan<sup>S</sup> Amp<sup>S</sup> clones as well as on 4548, 4548  
 21 CFP-Kan and 4548 YFP-Kan. Both the Kan<sup>S</sup> associated with a fluorescent phenotype and the  
 22 size of the PCR products confirmed the removal of the kanamycin cassette and the insertion  
 23 of the fluorescent genes.

24

## 25 **Primer sequences**

Name	Sequence	Constructio n step	Reference
P-BspHI- pA1-	TCTCT <b>CATG</b> ATTATCAAAAAGAGTATT <b>GACTT</b> AAAGTCT AACCTATAG <b>GATACTT</b> ACAGCCATCGAGAG <i>Gattaaagag</i>	PCR Introductio n of	This study

YC/FP for	<i>gagaaa</i> GGCGAAATGCGTAAAGGAGAAGAAC (109 bases) In blue : BspHI restriction site In black: P <sub>A1</sub> In pink: Ribosome binding site In green, CFP/YFP homologous sequence	promoter pA1 upstream the CFP or YFP cassette	
P-ClaI- YC/FP rev	TTCATCGATTTATTTGTATAGTTCATCCATGCC In blue : ClaI restriction site In green, CFP/YFP homologous sequence	PCR 1 Introduction of promoter pA1 upstream the CFP or YFP cassette	This study
p-pKD4- <i>rha</i> AH1- 2863- 2888 for	GAGTCGACGGCGCTGCGCCAACCTTGATCGTTTACCCGT TTCAATGCACTGCTGGCAGGGAAGCATTATCAGGGTT ATTGTCTC (84 bases) In blue : Sall restriction site In red: <i>rhaA</i> homologous sequence In green: pKD4-CFP (or pKD4-YFP) homologous sequence	PCR 2 Creating a linear fragment for homologous recombination	This study
p- pKD4p2H 2-1496- 1477rev	ACGGATCCACCACGCGGCAATGCGGTTGATAGAGGCA TCGAAGAAGTCAAGGCCGATATCCTCCTTAGTTCCTATT CCG (79 bases) In blue : BamHI restriction site In red: <i>rhaA</i> homologous sequence In green: pKD4-CFP (or pKD4-YFP) homologous	PCR 2 Creating a linear fragment for homologous	This study

	sequence	s recombinati on	
p- <i>rhaA</i> -for	GACCACTC AACTGGAACAGGCC	PCR 3 Checking for CFP/YFP insertion in the <i>rhaA</i> locus	This study
p-pKD4- 2863- 2888 for	GAAGCATTATCAGGGTTATTGTCTC	PCR 3 Checking for CFP/YFP insertion in the <i>rhaA</i> locus	This study
<i>rhaA</i> - 1262- 1243-rev	CCACGCTGGCTCAAATCGC	PCR 3 Checking for CFP/YFP insertion in the <i>rhaA</i> locus	This study

1

2

1

2 **PCR reactions:**

3

4 PCR 1 - Introduction of promoter PA1 upstream the CFP or YFP cassette

5 Plasmid pZE1R-CFP or pZA32-YFP 1  $\mu$ L

6 Primer P-BspHI-pA1-YC/FP for 2  $\mu$ L

7 Primer P-ClaI-YC/FP rev 2  $\mu$ L

8 Pfu 10X buffer 2  $\mu$ L

9 dNTPs 1  $\mu$ L

10 Pfu 0.5  $\mu$ L

11 H<sub>2</sub>O 11.5  $\mu$ L

12 Program : Hybridization at 52°C for 1 min, elongation at 72°C for 2min, 25 cycles.

13

14 PCR 2 - Creating a linear fragment for homologous recombination

15 Plasmid pKD4-CFP (or pKD4-YFP) 2  $\mu$ L

16 Primer p-pKD4-RhaAH1-2863-2888 for 1  $\mu$ L

17 Primer pKD4p2H2-1496-1477rev 1  $\mu$ L

18 Phusion 2X mastermix 10  $\mu$ L

19 H<sub>2</sub>O 6  $\mu$ L

20 Program : Hybridization at 60°C for 1 min, elongation at 72°C for 3min, 25 cycles.

21

22 PCR 3 - checking CFP and YFP cassette insertions and Kanamycin cassette removal

23 Colony inoculated in the PCR mix, with a toothpick.

24 Primer forward 1  $\mu$ L

1	Primer reverse	1 $\mu$ L
2	Taq 10X buffer	2 $\mu$ L
3	MgCl <sub>2</sub>	1.5 $\mu$ L
4	dNTP	1 $\mu$ L
5	Taq Goldstar Red	0.2 $\mu$ L
6	H <sub>2</sub> O	13.3 $\mu$ L

7

8 Program : Hybridization at 60°C for 1 min, elongation at 72°C for 3min 15 sec, 30 cycles.

9

## 10 **Effect of “doublets” on coefficient of selection estimates**

11 CY particles (C2), CC and YY doublet particles may pose a problem for accurate analysis of  
 12 marker proportions. We have found that “singlet” and “doublet” particles can be nicely  
 13 separated on a FSC TOF (time of flight) - FSC Peak plot. The frequency of YY (C1 region), CY  
 14 (C2 region), and CC (C4 region) can thus be estimated. If cells were associated randomly in a  
 15 doublet, when C and Y are initially introduced at a ratio of 1:1, we would expect to have C1 =  
 16 0.25, C2 = 0.5 and C3 = 0.25 within doublets. Most of the time, what we observe is C1 = 0.33,  
 17 C2 = 0.33 and C3 = 0.33. This means that cells are not associated randomly in a doublet, but  
 18 that they have more chance to be with a cell of the same color. Our hypothesis for such  
 19 observation is that when entering stationary phase, a fraction of the daughter cell  
 20 population might stay attached to each other. In Rang et al 2003, they observed that  
 21 *“Cultures of the bacteria most affected by GFP exhibited a proportion of elongated cells,*  
 22 *which suggests that GFP production could interfere with cell division in these strains”*, which  
 23 is another plausible explanation for our observation.

24 Now, the question is to evaluate the bias introduced by ignoring doublets. We can introduce  
 25 some notation to be specific. There are several unknowns: the fraction of doublets (either  
 26 CC, YY or CY, only the latter being measured in the C2 region). Let's denote it  $\phi$ . Then there is  
 27 the frequency of C bacteria ( $p$ ). Last, we can consider that doublets may not form by the  
 28 random association of two bacteria (e.g. two C bacteria may be more likely to form a

1 doublets than a pair of C and Y bacteria, as explained above). We can introduce a departure  
 2 from random association  $F$  (akin to a departure from Hardy Weinberg proportion). It is then  
 3 straightforward to compute the frequency of each type of particle (C, Y, CC, YY, CY). We  
 4 measure frequency by the ratio  $(C+CC)/(C+CC+Y+YY)$ , which introduces a bias from the true  $p$   
 5 value equal to

$$6 \quad -(1 - F)(1 - p)p(1 - 2p)\phi + O(\phi)^2.$$

7 This bias can be expressed in terms of  $\epsilon$ , the proportion of the C2 population (defined as  
 8  $CY/(C+CC+CY+YY+Y)$ ). It is simply

$$-\epsilon(p - 1/2)$$

9 which is very small, especially when  $p$  is close to  $1/2$ . In our case, average initial frequency is  
 10  $0.493 \pm 0.015$  and  $\epsilon < 1\%$ , which corresponds to a bias close to  $10^{-4}$ . The bias made on the  
 11 frequency change,  $\Delta p$ , during the competition is different and it is this bias that is most  
 12 relevant to estimating the strength of selection. Following the same approach and assuming  
 13  $F$  and  $\phi$  are approximately constant during the competition, it is

$$\frac{1}{2}\Delta p(1 - F)\phi$$

14

15 In other words the bias on the frequency change is proportional to itself. Expressing this in  
 16 terms of  $\epsilon$  and the intensity of selection per generation (with  $\Delta p = s g p q$ ), we obtain  
 17  $g s \epsilon p q$ . The bias on selection coefficient is therefore equal to  $s \epsilon$ . Thus, it is always very small  
 18 compared to  $s$ , provided the fraction of C2 population remains  $\ll 1$ . Correcting for doublets  
 19 would be probably necessary when  $\epsilon$  is greater than a few % (which, in our experiments, it is  
 20 not). Under those circumstances, it would be important to have a clear experimental  
 21 understanding of the origin of doublets particles (and measures of  $F$  and  $\phi$  as we did). Last,  
 22 it is important to underline that these slight biases do not introduce errors, so that  
 23 estimations of variances remain unaltered.

## 24 **Impact of de novo beneficial mutations during the assay**

25

1 Beneficial mutations should have an implausibly high selection coefficient to impact fitness  
2 measures made in our experiments. The computation is different if we consider between  
3 week variance or within week (between replicates) variance.

#### 4 **Computation between weeks**

5 The pre-cultures start from a frozen glycerol stock with a sample size of  $N_{gly}$  around  $10^5$ .  
6 Importantly, all replicates within a week use the same pre-culture and only experiments  
7 performed in different weeks use different pre-cultures (*i.e.* different subsamples of the  
8 same glycerol stock). Seventeen generations later (around 10 generations during pre-  
9 culture, and seven more during a second round of pre-culture removing any trace of  
10 glycerol), the competition starts for 6.6 more generations. During the competition, the  
11 presence of a beneficial mutation with selection coefficient  $s_e$  may distort the frequency  
12 change measured in our experiment. If competitions always start with this beneficial  
13 mutation, little variance among competitions will be generated. On the contrary, variance in  
14  $s$  among weeks will occur if the mutation is only present in some weeks. Hence, the worst  
15 case situation occurs when there is around  $\frac{1}{2}$  chances to sample the beneficial mutation  
16 from the glycerol. Computing this probability from a Poisson distribution, we find that the  
17 worst case is a frequency in the glycerol of around

18

$$19 \quad p_{gly} = \text{Log}[2] / N_{gly}.$$

20

21 Starting from this initial frequency, the beneficial mutation will end up at the end of the  
22 preculture after 17 generations at a frequency

23

$$24 \quad p_o = p_{gly} \text{Exp}[s_e 17]$$

25

26 (from logistic growth). Let's assume that the beneficial mutation occurred in a YFP cell. In  
27 our experiments,  $p_{yfp}$  –the frequency of the YFP cells—is always close to  $\frac{1}{2}$ . During the

1 competition itself, the beneficial mutation will cause an extra change  $\Delta p_e$  in frequency of the  
2 YFP background

3

$$4 \quad \Delta p_e = s_e p_o (1 - p_{yfp}).$$

5

6 The bias per generation in the estimated selection coefficient is then

7

$$8 \quad s_{bias} = \Delta p_e / (p_{yfp} (1 - p_{yfp}) \text{ 6.6}).$$

9

10 Solving for the selection coefficient  $s_e$  causing a bias in selection at least as large as our  
11 precision (*i.e.* of the order of  $s_{bias} = 2 \cdot 10^{-4}$ ), we obtain  $s_e = 0.33$ . We are unaware of adaptive  
12 changes as high as this in the ecological situation of the growth environments we used. In  
13 particular, none has been found in the Lenski experiment. Moreover, our strain has already  
14 evolved 10,000 generations, which is the period where the strongest adaptive changes have  
15 already been fixed (Barrick et al., 2009). Indeed, the fitness of the relevant population  
16 increased by less than 30% in 20,000 generations following the time point at which REL4548  
17 was isolated (Barrick et al., 2009).

### 18 **Computation within weeks**

19 Each replicated competition within a week starts from the same preculture and a large  
20 inoculum ( $10^6$  cells). The worst case is when the beneficial mutation occurs early in half of  
21 the competitions. Since we sample  $N_{start} = 5 \cdot 10^5$  cells of a given competitor, it means that

22

$$23 \quad p_{start} = \text{Log}[2] / N_{start}$$

24



1 Using the exact same computation as above, except that the beneficial has only 6.6  
2 generation to change in frequency, we obtain  $s_e = 476$ , which is clearly unrealistic.  
3 Importantly, we observe inflated  $\text{var}(s)$  within weeks of the same order of magnitude as  
4 between weeks, so that it is quite clear that adaptive changes are unable to explain our  
5 results.