

Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage ϕ 1

Joan B. Peris,* Paulina Davis,* José M. Cuevas,*[†] Miguel R. Nebot,[‡] and Rafael Sanjuán*^{†,1}

*Institut Cavanilles de Biodiversitat i Biologia Evolutiva and Departament de Genètica, Universitat de València, 46980 València, Spain,

[†]Centro Superior de Investigación en Salud Pública (CSISP), 46020 València, Spain and [‡]Instituto de Física Corpuscular, CSIC, Universitat de València, 46980 València, Spain

Manuscript received February 4, 2010

Accepted for publication April 1, 2010

ABSTRACT

Empirical knowledge of the fitness effects of mutations is important for understanding many evolutionary processes, yet this knowledge is often hampered by several sources of measurement error and bias. Most of these problems can be solved using site-directed mutagenesis to engineer single mutations, an approach particularly suited for viruses due to their small genomes. Here, we used this technique to measure the fitness effect of 100 single-nucleotide substitutions in the bacteriophage ϕ 1, a filamentous single-strand DNA virus. We found that approximately one-fifth of all mutations are lethal. Viable ones reduced fitness by 11% on average and were accurately described by a log-normal distribution. More than 90% of synonymous substitutions were selectively neutral, while those affecting intergenic regions reduced fitness by 14% on average. Mutations leading to amino acid substitutions had an overall mean deleterious effect of 37%, which increased to 45% for those changing the amino acid polarity. Interestingly, mutations affecting early steps of the infection cycle tended to be more deleterious than those affecting late steps. Finally, we observed at least two beneficial mutations. Our results confirm that high mutational sensitivity is a general property of viruses with small genomes, including RNA and single-strand DNA viruses infecting animals, plants, and bacteria.

MUTATIONAL fitness effects are relevant to many evolutionary processes. For instance, they determine the fraction of mutations that evolves neutrally (OHTA 1992), the amount of genetic variation at the mutation–selection balance (HALDANE 1937), processes of fitness decay, such as Muller’s ratchet (BUTCHER 1995), mutational meltdown (LYNCH *et al.* 1993), or lethal mutagenesis (BULL *et al.* 2007), the ability of organisms to fix beneficial mutations and evolve novel functions (WAGNER 2005), or the origin of sex and recombination (PECK *et al.* 1997; DE VISSER *et al.* 2003). Considerable progress has been made in characterizing mutational fitness effects using model organisms or studying genetic variation in natural populations (EYRE-WALKER and KEIGHTLEY 2007). For instance, mutation–accumulation experiments suggest that the average effect of spontaneous deleterious mutations is 1% or lower (KIBOTA and LYNCH 1996) in *Escherichia coli*, while roughly 90% of engineered gene knockouts are viable (BABA *et al.* 2006) and transposon insertions reduce fitness by 3% or less on average (ELENA *et al.* 1998). In yeast, mutation–accumulation and chemical mutagenesis

experiments have shown that mutations reduce fitness by 1–4% on average in diploid strains (ZEYL and DE VISSER 2001; SZAFRANIEC *et al.* 2003; JOSEPH and HALL 2004). In nematodes most mutations have fitness effects lower than 1% (KEIGHTLEY and CABALLERO 1997; DAVIES *et al.* 1999), in *Drosophila* the average effect of mutations ranges from 0.5 to 3.5% (MUKAI *et al.* 1972; OHNISHI 1977; FERNÁNDEZ and LÓPEZ-FANJUL 1996; FRY *et al.* 1999), and, in humans, most segregating amino acid substitutions have fitness effects lower than 10% (EYRE-WALKER and KEIGHTLEY 2007).

Although mutation–accumulation studies provide valuable information about the average effects of deleterious mutations, their power to infer the entire distribution of mutational effects, including neutral and lethal mutations, is more limited. Also, excluding bias due to selection can be problematic, and the precise location and nature of each mutation is often unknown. On the other hand, studies based on engineering mutations have been generally restricted to large deletions or insertions, which are probably infrequent in nature compared to point mutations. A direct and powerful approach that helps us to solve these difficulties consists of introducing single-nucleotide substitutions by site-directed mutagenesis. Due to their small genome sizes, viruses are excellent systems for achieving this goal. In previous work, this technique has been used for studying mutational fitness effects in several RNA viruses (SANJUÁN *et al.* 2004; CARRASCO *et al.* 2007;

Supporting information available at online <http://www.genetics.org/cgi/content/full/genetics.110.115162/DC1>.

¹Corresponding author: Institut Cavanilles de Biodiversitat i Biologia Evolutiva, University of València, C/Catedrático Agustín Escardino 9, 46980 Paterna, València, Spain. E-mail: rafael.sanjuán@uv.es

DOMINGO-CALAP *et al.* 2009). However, less is known for DNA viruses—but see DOMINGO-CALAP *et al.* (2009). Here, we use this approach to characterize the distribution of mutational fitness effects in the bacteriophage ϕ 1, an inovirus of the bacteriophage ϕ 13 clade, making two important improvements over previous work: first, the number of mutations tested is higher (100) and second, the contribution of experimental error to the observed distribution is explicitly accounted for. We show that one-fifth of single-nucleotide substitutions are lethal, while viable ones reduce fitness by 11% on average and can be described by a heavy-tail two-parameter distribution such as the log-normal. Interestingly, the fraction of beneficial mutations is unexpectedly high. We also compare the average effects of different mutation types and of mutations affecting different genes.

MATERIALS AND METHODS

Bacteriophage and cells: Bacteriophage ϕ 1 and *E. coli* C strain IJ1862 (BULL *et al.* 2004) were originally obtained from James J. Bull (University of Texas). We adapted the phage to our laboratory conditions by performing 80 serial passages at high population sizes as follows: exponentially growing IJ1862 cells at an optical density of $OD_{600} = 0.15$ were inoculated with $\sim 10^5$ particle forming units (pfu) in 0.5 ml of LB medium, incubated in agitation (650 rpm) at 37° in a Thermomixer 24-tube shaker (Eppendorf), and harvested after approximately 2 hr, which corresponded to the late exponential growth phase of the virus ($\sim 10^9$ pfu/ml). Cells were removed by centrifugation and supernatants were aliquoted, stored at -70° , and titrated using LB medium solidified with soft agar (a single cycle of freeze–thawing did not noticeably reduce the viral titer). Initial and final titers were used to calculate growth rates and to adjust sampling times for the next passage accordingly. No significant changes in growth rate were observed during the 80 passages. A single plaque from passage 80 was isolated and used to infect a large culture of IJ1862 cells. Viruses from the supernatant of this culture were stored in several aliquots at -70° and used as the reference virus (wild type) in all fitness assays, whereas the cell pellet was used to isolate the circular double-strand replicative form of the viral DNA using a plasmid isolation kit (Roche). This DNA was used as template for the site-directed mutagenesis reactions. Using a well-adapted phage minimizes the fraction of beneficial mutations and increases the genetic stability of the virus during plaque growth and fitness assays.

Site-directed mutagenesis: Full-length PCR amplicons were obtained from 500 pg of template DNA using Phusion high-fidelity DNA polymerase (New England Biolabs) and a pair of contiguous, divergent, 5'-phosphorylated primers, one of which carried the desired nucleotide substitution. The cycling conditions were 2 min at 98° (initial denaturation), 30 cycles of 30 sec at 98°, 30 sec at 68°, and 4 min at 72°, and a final extension step of 10 min at 72°. The calculated fraction of PCR products carrying undesired mutations is $f = \log(2)\mu Gc = 5.8\%$, where $\mu = 4.4 \times 10^{-7}$ per base per round of copying is the error rate of the enzyme provided by the manufacturer, $G = 6386$ is the genome size, and $c = 30$ the number of PCR cycles. Since primers are contiguous and divergent, PCR products consist of a linear, full-length, genomic DNA. These products were circularized using the Quick T4 ligase (New England Biolabs) and IJ1862 competent cells were transfected by the heat-shock method (42°, 30 sec) in the presence of 100

mm $CaCl_2$. The transfected cells were immediately plated and four individual plaques of each mutant were picked after 20 hr of incubation at 37°, resuspended in LB, and stored at -70° . PCR was performed directly from the resuspended plaques using Phusion DNA polymerase. To check the presence of the mutation and that no additional mutations appeared in the flanking region (~ 500 nt), PCR products were column purified and sequenced using mutagenesis primers that annealed near the target site. Once the mutation was verified, three resuspended plaques were titrated and mixed in equal amounts for fitness assays. Mixing three plaques minimizes the effects of potential additional mutations resulting from PCR errors. Additional details about the site-directed mutagenesis protocol can be found elsewhere (SANJUÁN 2010).

Fitness assays: Approximately 10^5 pfu were inoculated (t_0) into 0.5 ml of LB medium containing exponentially growing IJ1862 cells at an optical density $OD_{600} = 0.15$ and harvested after 2 hr 5 min (t_1) of incubation in agitation (650 rpm) at 37°. Cells were removed by centrifugation and supernatants were stored at -70° . Previous assays showed that the viral titer increases exponentially within this interval. Titers at times t_0 and t_1 were determined and the growth rate (r) was calculated as the increase in log-titer per hour. Relative fitness (W) was defined as the growth rate ratio and the relative fitness effect as $s = W - 1$. Each mutant was assayed in three independent experimental blocks. Three additional blocks were performed for a subset of mutants showing fitness values slightly below 1.0 and for all mutants showing $W > 1$. In each block, 18 mutants and six wild-type replicates were assayed simultaneously. The relative fitness of mutant i was calculated as $W_i = r_i / \bar{r}_0$, where \bar{r}_0 is the average of the six wild-type determinations of the block. This cancels out potential block or day effects. However, using the 108 available $W_0 = r_0 / \bar{r}_0$ values, we detected a significant effect of the position within the block. To minimize this problem, the position of each mutant was changed in each assay. Also, since the bias was roughly linear with the position number, we used the 108 W_0 values to interpolate position effects and to correct W_i values accordingly. Corrected values (reported in the text) were used for inferring the distribution of mutational fitness effects and for testing the effect of each individual mutation. For this latter purpose, we also performed ANOVA tests using uncorrected s -values, accounting for the position within the block as a covariate in the model. This alternative approach yielded the same numbers of neutral, deleterious, and beneficial mutations.

Lethality tests: For some mutants, transfection yielded an anomalously low number of plaques (< 5 vs. typically > 100 for the other mutants). Further, the few clones recovered were nonmutant and thus most likely derived from the template DNA. If this result was repeated, the mutant was classified as a candidate lethal. For these cases, we first performed direct sequencing of the mutagenesis PCR product for the region flanking the target site to verify that the target mutation was present and that no additional changes appeared. Second, we designed primers identical to those used for each mutagenesis reaction except that they did not carry the target substitution (control primers), and we repeated the protocol exactly as above for the mutagenesis and control reactions in parallel using the reference, nonmutated DNA as template in both cases. In all cases, the number of plaques obtained from similar amounts of DNA was much higher (> 100 -fold on average) for controls than for true mutants, thus confirming lethality (supporting information, Figure S1).

Quantification of experimental error: Several precautions were taken to minimize the frequency at which undesired mutations appeared: (i) the template DNA used for all mutagenesis was obtained from a plaque-purified virus to minimize its genetic variation; (ii) we used a high-fidelity DNA

polymerase; and (iii) the region flanking the target substitution was sequenced. Further, as mentioned above, to reduce the potential impact of undesired mutations, three plaques of each mutant were mixed prior to fitness assays. To quantify the error of the method, we obtained 45 clones by the same procedure used for obtaining the mutants, except that the PCR primers did not carry any substitution and measured their fitness relative to the wild type. If the method was fully accurate, all these control clones should be selectively neutral ($s = 0$). Only one clone had a significant fitness defect and sequencing revealed the presence of a small insertion in the primer region. After removing this single case, the mean fitness value of the controls was $\bar{s} = -0.0041$ with variance 0.0008, standard deviation 0.0275, and range -0.0513 – 0.0585 . The means and variances of mutational fitness effects reported in the text were obtained after subtracting the control mean and variance from the total means and variances.

Maximum-likelihood inference of the distribution of mutational fitness effects: Each experimentally observed s -value is determined by the joint action of mutational fitness effects and measurement error. The 44 s -values from the control assays used to calibrate the error (see above) were accurately modeled ($r^2 = 0.992$) by a normal distribution with mean -0.0041 and standard deviation 0.0272. Consistently, a Shapiro–Wilk test indicated that normality could not be rejected for these data ($P = 0.5549$). We convoluted model distributions for mutational fitness effects (exponential, gamma, beta, log-normal, or Weibull) and the above normal to calculate the likelihood of each experimental observation s_i given the model as $P(-s_i; \mathbf{p}) = \int_{-\infty}^{+\infty} f(-s_i - y; \mathbf{p})N(y; \mu, \sigma)dy$, with $\mu = 0.0041$ and $\sigma = 0.0272$, where f is the probability density function (pdf) of the model distribution, \mathbf{p} the parameters of the model, and N the pdf of the normal distribution (the minus preceding s_i was introduced to make the variable positive for deleterious mutations). Alternatively, instead of using the normal distribution, f could be summed over the 44 control data values, an approach that yielded very similar results. The parameters of the model were estimated by maximum likelihood and the total log-likelihood was used to compare models. Distributions defined within the range $[0; \infty]$ were truncated at 1 and normalized accordingly. This correction was not needed for the beta distribution, which is defined in the same range as $-s$ for nonbeneficial mutations. To obtain a standard measure of the goodness of fit of each model, we calculated the correlation coefficient between the cumulative distribution function (cdf) of the maximum-likelihood model and the cdf of the data.

RESULTS

Fitness of single-nucleotide mutants: We obtained 100 viral clones by site-directed mutagenesis, each carrying a different nucleotide substitution (Figure 1). The mutated sites were chosen at random and, consequently, changes were spread through all the genome. Ninety belonged to coding regions and 10 to intergenic regions. Although the substituted nucleotide was not random (most changes were to A), mutations were probably representative of a random set: 70 were transversions and 30 transitions and, among the 90 mutations affecting coding regions, 51 produced amino acid substitutions (missense mutations), 9 produced premature amino acid chain termination (nonsense mutations), and 30 were synonymous. Among the 51 missense mutations, 29 changed the polarity of the amino acid

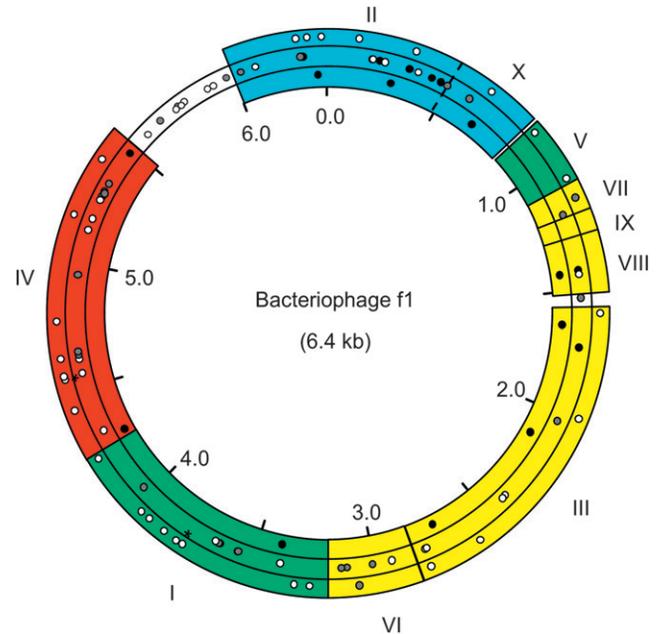


FIGURE 1.—Diagram of the f1 genome showing the 100 single-nucleotide substitutions created by site-directed mutagenesis. Gene names (roman numerals) and genome positions are indicated. Colors indicate broad functional categories (blue, replication; green, particle maturation; yellow, capsid; red, extrusion). Additional details about the function of each gene can be found elsewhere (CALENDAR 2006). Briefly, protein II nicks the viral DNA to allow replication priming, protein X results from in-frame translation of gene II and is required for single-strand DNA accumulation, protein V binds to DNA and collapses the circular genome into a flexible rod, proteins VII and IX are small coat proteins located at the tip of the virion, protein VIII forms the cylinder containing the viral DNA, proteins III and VI are located at the tail of the virion and are involved in termination of virion assembly, protein I is an inner membrane protein that hydrolyzes ATP and promotes capsid morphogenesis, and protein IV forms a channel in the outer membrane to allow virion extrusion. Thinner, noncolored, areas denote intergenic regions, which constitute 9% of the genome. Mutations falling at coding regions are shown in three concentric rings according to whether they are synonymous (outer), missense (central), or nonsense (inner). Open, shaded, and solid circles represent neutral, significantly deleterious, and lethal mutations, respectively. Two significantly beneficial mutations are shown with asterisks. Significance levels were adjusted to 0.05 for multiple tests.

(hydrophobic, polar, positively charged, or negatively charged), 10 replaced an aromatic amino acid by an acyclic one or vice versa, and 23 changed the size of the amino acid. Fifteen of the 51 missense mutations affected genes involved in replication, 8 affected maturation genes, 13 affected encapsidation genes, and 15 affected the extrusion gene. These four functions can be more broadly classified as early (replication) and late (maturation, capsid, and extrusion) steps of the infection cycle.

The observed distribution of fitness effects for the 100 mutations is shown in Figure 2. Twenty-one mutations

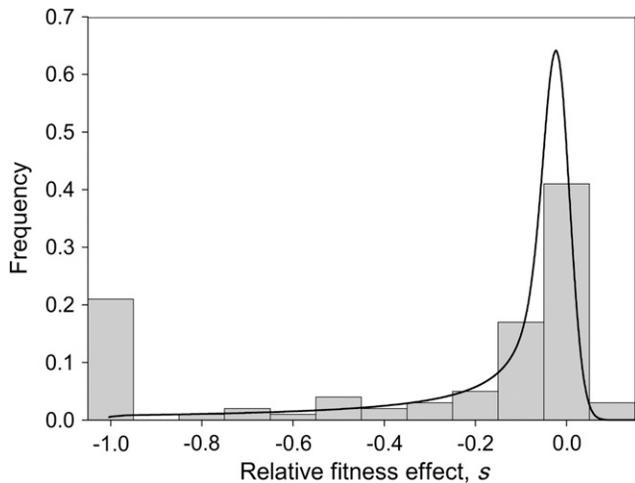


FIGURE 2.—Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage ϕ . The frequency histogram of the 100 mutations is shown. Note that the fitness effect of lethal mutations is $s = -1$. The effect of each individual mutation is provided in Table S1. Shown superimposed is the best-fitting probability density function for viable, non-beneficial mutations (a log-normal distribution with the parameter values shown in Table 3).

were lethal according to the lack of infectivity of the viral DNA. Fitness assays of the 79 nonlethal mutants indicated that 35 were significantly deleterious ($P < 0.05$), 39 did not significantly deviate from neutrality, and 5 were significantly beneficial. After applying the sequential Bonferroni correction to control the false discovery rate for multiple tests of the same hypothesis (BENJAMINI and HOCHBERG 1995), 24 deleterious and 2 beneficial mutations (C3748A and T4541A) remained, whereas 53 did not deviate significantly from neutrality. The mean fitness effect of all viable mutations was $\bar{s}_v = -0.107$ with variance $V(s_v) = 0.037$. After removing the two significantly beneficial mutations, $\bar{s}_v = -0.111$ and $V(s_v) = 0.037$.

All synonymous mutations were viable and their effects did not deviate from neutrality on average (Table 1; Mann–Whitney test: $P = 0.910$) although two (6.7%) were significantly deleterious after correcting for multiple tests (C1161A and G3077A, with s -values -0.102 and -0.114 , respectively). All mutations at intergenic sites were also viable but tended to be deleterious ($P = 0.013$). Missense mutations were strongly deleterious ($P < 0.001$) and 12/51 (24%) were lethal, including 9 that changed the amino acid polarity. The latter were more harmful ($\bar{s} = -0.454$, including lethals) than those not changing the amino acid polarity ($\bar{s} = -0.262$; $P = 0.027$). Similarly, acyclic to aromatic substitutions (or vice versa) were more deleterious ($\bar{s} = -0.601$) than the rest of missense mutations ($\bar{s} = -0.315$; $P = 0.032$). However, fitness effects were apparently the same regardless of whether the mutation changed the size of the amino acid ($P = 0.335$). Finally, all nonsense mutations were lethal.

TABLE 1

Mutational fitness effects (s) for different kinds of mutations

	Mean ^a	Variance ^a	Number ^a	Lethal fraction (%)
Missense	0.371	0.164	51	24
Synonymous	0.003	0.002	30	0
Intergenic	0.138	0.036	10	0
Nonsense	1.000	0.000	9	100
Total	0.294	0.163	100	21

^a For all mutations (including lethals).

Lethal mutations occurred in most genes and thus appeared to be homogeneously spread throughout the genome. However, a more careful examination shows that missense mutations affecting replication genes were more likely to be lethal (7/15) than those affecting other genes (5/36; Fisher exact test, $P = 0.026$). Consistently, the average effect of mutations tended to be more deleterious if initial steps of the infection cycle were affected (Table 2; Spearman correlation, $N = 4$, $P < 0.05$). It is also noteworthy that the two significantly beneficial mutations produced amino acid replacements in the proteins encoded by genes I and IV, which are involved in maturation and extrusion, respectively.

Distribution of mutational fitness effects: We sought to infer the statistical properties of the distribution of mutational fitness effects (excluding lethal and beneficial mutations) by fitting several model distributions to the data. We used a maximum-likelihood procedure, in which the likelihood of each observation given the model was calculated by taking into account measurement error (see MATERIALS AND METHODS). The simplest case considered was the exponential distribution, which has a single parameter. This distribution fit the data reasonably well (Table 3), but we also considered two-parameter models, starting with the gamma distribution, which is a generalization of the exponential. A likelihood-ratio test indicated that the use of this more complex model was justified by the significantly improved fit ($\chi^2 = 47.38$, $P < 0.001$). We considered three additional two-parameter distributions: the beta, the log-normal, and the Weibull distributions (Table 3). The total log-likelihoods of the four models were similar and all fit the observed cumulative density function with a precision above 99%. However, the log-normal and the Weibull distributions, which have heavier tails, were slightly better than the gamma or beta. The maximum likelihood estimates of the log-normal parameters gave a predicted average mutational effect of $\hat{s}_v = -0.122$ and variance $\hat{V}(s_v) = 0.041$. The corresponding inferred probability density function is shown in Figure 2.

DISCUSSION

Previous work has shown that the distribution of mutational fitness effects can vary across species (EYRE-

TABLE 2
Mutational fitness effects (s) of missense mutations for different functional categories

	Genes	Mean ^a	Variance ^a	Number ^a	Lethal fraction
Replication	II, X	0.535	0.214	15	47%
Maturation	I, V	0.362	0.205	8	25%
Capsid	III, VI, VII, VIII, IX	0.305	0.130	13	15%
Extrusion	IV	0.270	0.113	15	7%
Total		0.371	0.164	51	24%

^a For all mutations (including lethals).

WALKER and KEIGHTLEY 2007), implying that results from a given model system cannot be readily extrapolated to distantly related groups and that more empirical information is therefore needed. Here, we have carried out a detailed characterization of the fitness effects of single-nucleotide substitutions in a filamentous single-strand DNA phage. It is interesting to see how these data compare to those obtained with other viruses using the same method. In vesicular stomatitis virus (VSV), tobacco etch virus (TEV), and bacteriophage Q β , more than one-third of single-nucleotide substitutions are lethal, while nonlethal ones reduce fitness by 10–13% on average (SANJUÁN *et al.* 2004; CARRASCO *et al.* 2007; DOMINGO-CALAP *et al.* 2009; SANJUÁN 2010). A recent study has shown that, in the DNA bacteriophage Φ X174, the lethal fraction is 20% and the mean effect size of viable ones is 13% (DOMINGO-CALAP *et al.* 2009). The results obtained with Φ X174 are very similar to the ones obtained with f1 despite that fact that these are two phylogenetically unrelated single-strand DNA phages. Also, the constancy of average fitness effect size across RNA and ssDNA viruses (10–13%) is remarkable (SANJUÁN 2010). Interestingly, the lethal fraction is lower in f1 and Φ X174 than in the RNA viruses VSV, TEV, or Q β , yet still remarkably high. The comparison with nonviral systems

is more problematic because of the variety of methods used and the different types of mutations tested in each study. Despite this limitation, previous work has established that RNA and single-strand DNA viruses show low tolerance to mutation compared with organisms such as *E. coli*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and others (ELENA *et al.* 2006). This is consistent with the fact that small and compact genomes as those of many viruses encode few or no mechanisms of robustness at the molecular level as, for example, alternative metabolic pathways, genetic redundancy, or modularity (WAGNER 2005; ELENA *et al.* 2006).

We have classified viable mutants as deleterious, neutral, or beneficial according to whether their fitness values significantly deviates from that of the wild type. However, it is difficult to distinguish mutations of small effect from strictly neutral ones. Even after performing additional replicates for small-effect mutants, most effect sizes lower than 0.05 were not significantly different from zero. It is also noteworthy that, since neutrality depends not only on the selection coefficient but also on the effective population size (KIMURA 1983; OHTA 1992), determining *s*-values and modeling their statistical distribution is probably more informative than classifying mutations as deleterious or neutral on the basis of their

TABLE 3
Statistical models describing mutational fitness effects for 77 viable, nonbeneficial, mutations

Model	Probability density function ^a	Parameter estimates ^b	Log-likelihood ^c	Goodness of fit ^d
Exponential	$f(-s; \lambda) = \frac{\lambda e^{-\lambda s}}{1 - e^{-\lambda}}$	$\lambda = 8.144$ (6.867;9.564)	35.11	0.957
Gamma	$f(-s; \alpha, \beta) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1} e^{-s/\beta}$	$\alpha = 0.344$ (0.276;0.425) $\beta = 0.387$ (0.262;0.676)	58.80	0.993
Beta	$f(-s; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1} (1-s)^{\beta-1}$	$\alpha = 0.315$ (0.257;0.385) $\beta = 2.265$ (1.661;3.017)	58.35	0.992
Log-normal	$f(-s; \mu, \sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi} \operatorname{erfc}(\mu/\sigma\sqrt{2})} e^{-(\ln s - \mu)^2/2\sigma^2}$	$\mu = -2.895$ (-3.848;-0.916) $\sigma = 3.473$ (2.550;4.691)	60.97	0.994
Weibull	$f(-s; k, \lambda) = \frac{k}{\lambda(1 - e^{-\lambda})^k} \left(\frac{s}{\lambda}\right)^{k-1} e^{-(s/\lambda)^k}$	$\lambda = 0.079$ (0.049;0.161) $k = 0.440$ (0.358;0.530)	60.32	0.994

^a The constraint $0 \leq -s \leq 1$ was imposed ($0 < -s \leq 1$ for the Log-normal) and probability density functions were normalized accordingly.

^b Obtained by maximum likelihood, accounting for experimental error; confidence intervals correspond to an increase of one log-likelihood unit.

^c Calculated from probability densities. Since the latter can be larger than 1, the log-likelihood can be positive.

^d Squared correlation coefficient (r^2) between the observed and predicted cumulative distribution functions.

selection coefficient. Concerning the shape of the distribution, viable mutations of small effect are more abundant than those of large effect in all biological systems examined so far (ELENA *et al.* 1998; DAVIES *et al.* 1999; SANJUÁN *et al.* 2004; COWPERTHWAITTE *et al.* 2005; POON and CHAO 2005; BABA *et al.* 2006; CARRASCO *et al.* 2007; EYRE-WALKER and KEIGHTLEY 2007; DOMINGO-CALAP *et al.* 2009; SANJUÁN 2010). A very simple model that captures this essential feature is the exponential distribution, but the conclusion emerging from several studies is that the observed coefficients of variation are larger than predicted by this model and that the observed distributions tend to have sharper peaks and heavier tails. These properties are better accounted for by the log-normal or the Weibull distributions. Another point worth mentioning is that in all previous studies, the contribution of experimental error to the shape of the observed distribution was unaccounted for, whereas here this has been made possible by carrying out appropriate controls and implementing a more powerful statistical analysis.

We found that 2 of the 100 mutations examined were significantly beneficial even after correcting for multiple tests. As discussed above, the actual fraction might be higher, since mutations of small effect are difficult to detect. High fractions of beneficial mutations have been reported previously in a low-fitness artificial recombinant of VSV (4%) or in debilitated mutation-accumulation lines of bacteriophage Φ X174 (18%) (SANJUÁN *et al.* 2004; SILANDER *et al.* 2007). Not surprisingly, these populations can undergo rapid adaptive or compensatory evolution but, in contrast, our fl genotype had been maintained at high population sizes under the same laboratory conditions for 80 passages without experiencing significant changes in fitness. Other previous estimates of the fraction of beneficial mutations in DNA genomes for moderately well adapted genotypes are 1–2% in bacteriophage Φ X174 (SILANDER *et al.* 2007), <0.5% in *E. coli* (ELENA *et al.* 1998), and 3% in yeast after correcting for multiple tests (THATCHER *et al.* 1998). In light of these data, we speculate that in DNA microorganisms, adaptive evolution is mainly limited by mutational supply, as opposed to RNA viruses. In the latter, failure to reach the evolutionary optimum might be more often caused by an excessive mutational load rather than by a poor supply of beneficial mutations. Consistently, recent work with bacteriophage T7 strongly suggests that artificially increasing the mutation rate results in faster adaptation (SPRINGMAN *et al.* 2009). Therefore, the roughly constant mutation rate of 0.003 per genome and round of copying found in most microorganisms (DRAKE 1991; DRAKE *et al.* 1998) might be generally suboptimal in terms of adaptation.

Finally, our results can also help elucidate other basic evolutionary questions. For instance, the neutral theory of molecular evolution (KIMURA 1983), as well as adaptive evolution models based on extreme value theory

(ORR 2003), depend critically on the relative abundance of deleterious, neutral, and beneficial mutations.

We thank Jim Bull for supplying the virus and the cell line, and Loles Catalán and Concha Hueso for technical assistance. This work was financially supported by grant BFU2008-03978/BMC from the Spanish MICIIN to R.S., the Ramón y Cajal program to R.S., the Juan de la Cierva program to J.M.C., and grant MD-01485 from the National Institutes of Health to P.D.

LITERATURE CITED

- BABA, T., T. ARA, M. HASEGAWA, Y. TAKAI, Y. OKUMURA *et al.*, 2006 Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**: 2006.
- BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.* **57**: 289–300.
- BULL, J. J., M. R. BADGETT, R. SPRINGMAN and I. J. MOLINEUX, 2004 Genome properties and the limits of adaptation in bacteriophages. *Evolution* **58**: 692–701.
- BULL, J. J., R. SANJUÁN and C. O. WILKE, 2007 Theory of lethal mutagenesis for viruses. *J. Virol.* **81**: 2930–2939.
- BUTCHER, D., 1995 Muller's ratchet, epistasis and mutation effects. *Genetics* **141**: 431–437.
- CALENDAR, R. L., 2006. *The Bacteriophages*. Oxford University Press, UK.
- CARRASCO, P., F. DE LA IGLESIA and S. F. ELENA, 2007 Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco etch virus. *J. Virol.* **81**: 12979–12984.
- COWPERTHWAITTE, M. C., J. J. BULL and L. A. MEYERS, 2005 Distributions of beneficial fitness effects in RNA. *Genetics* **170**: 1449–1457.
- DAVIES, E. K., A. D. PETERS and P. D. KEIGHTLEY, 1999 High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**: 1748–1751.
- DE VISSER, J. A., J. HERMISSON, G. P. WAGNER, M. L. ANCEL, H. BAGHERI-CHAICHIAN *et al.*, 2003 Perspective: evolution and detection of genetic robustness. *Evolution* **57**: 1959–1972.
- DOMINGO-CALAP, P., J. M. CUEVAS and R. SANJUÁN, 2009 The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. *PLoS Genet.* **5**: e1000742.
- DRAKE, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. USA* **88**: 7160–7164.
- DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH and J. F. CROW, 1998 Rates of spontaneous mutation. *Genetics* **148**: 1667–1686.
- ELENA, S. F., L. EKUNWE, N. HAJELA, S. A. ODEN and R. E. LENSKI, 1998 Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* **102/103**: 349–358.
- ELENA, S. F., P. CARRASCO, J. A. DARÓS and R. SANJUÁN, 2006 Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* **7**: 168–173.
- EYRE-WALKER, A., and P. D. KEIGHTLEY, 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**: 610–618.
- FERNÁNDEZ, J., and C. LÓPEZ-FANJUL, 1996 Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. *Genetics* **143**: 829–837.
- FRY, J. D., P. D. KEIGHTLEY, S. L. HEINSOHN and S. V. NUZHIDIN, 1999 New estimates of the rates and effects of mildly deleterious mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **96**: 574–579.
- HALDANE, J. B. S., 1937 The effect of variation on fitness. *Am. Nat.* **71**: 337–349.
- JOSEPH, S. B., and D. W. HALL, 2004 Spontaneous mutations in diploid *Saccharomyces cerevisiae*: more beneficial than expected. *Genetics* **168**: 1817–1825.
- KEIGHTLEY, P. D., and A. CABALLERO, 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **94**: 3823–3827.
- KIBOTA, T. T., and M. LYNCH, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**: 694–696.
- KIMURA, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

- LYNCH, M., R. BÜRGER, D. BUTCHER and W. GABRIEL, 1993 The mutational meltdown in asexual populations. *J. Hered.* **84**: 339–344.
- MUKAI, T., S. I. CHIGUSA, L. E. METTLER and J. F. CROW, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**: 335–355.
- OHNISHI, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. III. Heterozygous effect of polygenic mutations. *Genetics* **87**: 547–556.
- OHTA, T., 1992 The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**: 263–286.
- ORR, H. A., 2003 The distribution of fitness effects among beneficial mutations. *Genetics* **163**: 1519–1526.
- PECK, J. R., G. BARREAU and S. C. HEATH, 1997 Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* **145**: 1171–1199.
- POON, A., and L. CHAO, 2005 The rate of compensatory mutation in the DNA bacteriophage Φ X174. *Genetics* **170**: 989–999.
- SANJUÁN, R., 2010 Mutational fitness effects in RNA and ssDNA viruses: common patterns revealed by site-directed mutagenesis studies. *Phil. Trans. Roy. Soc. B* (in press).
- SANJUÁN, R., A. MOYA and S. F. ELENA, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. USA* **101**: 8396–8401.
- SILANDER, O. K., O. TENAILLON and L. CHAO, 2007 Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* **5**: e94.
- SPRINGMAN, R., T. KELLER, I. MOLINEUX and J. J. BULL, 2009 Evolution at a high imposed mutation rate: adaptation obscures the load in phage T7. *Genetics* **184**: 221–232.
- SZAFRANIEC, K., D. M. WŁOCH, P. SŁIWA, R. H. BORTS and R. KORONA, 2003 Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast *Saccharomyces cerevisiae*. *Genet. Res.* **82**: 19–31.
- THATCHER, J. W., J. M. SHAW and W. J. DICKINSON, 1998 Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl. Acad. Sci. USA* **95**: 253–257.
- WAGNER, A., 2005 *Robustness and Evolvability in Living Systems*. Princeton University Press, Princeton, NJ.
- ZEYL, C., and J. A. G. M. DE VISSER, 2001 Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* **157**: 53–61.

Communicating editor: J. LAWRENCE

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.109.115162/DC1>

Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage ϕ 1

Joan B. Peris, Paulina Davis, José M. Cuevas, Miguel R. Nebot and Rafael Sanjuán

Copyright © 2010 by the Genetics Society of America

DOI: 10.1534/genetics.109.115162

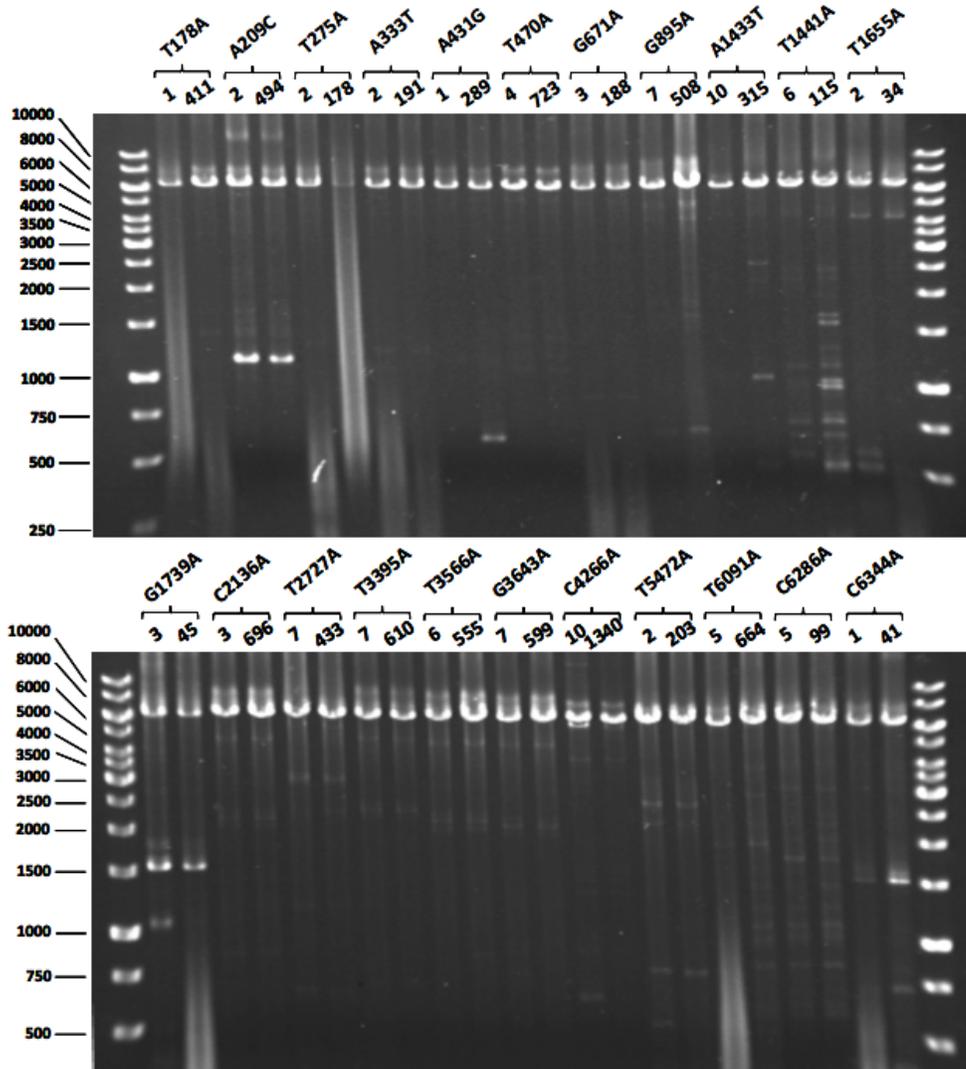


FIGURE S1.—Confirmation of lethal mutations. Site-directed mutagenesis reactions and control reactions (using non-mutagenic but otherwise identical primers) were loaded in 0.8% agarose gels and stained with ethidium bromide. For each candidate mutation, the left and right lanes contain the mutagenesis and control reactions, respectively. Bands of the expected size (6.4 kb) were obtained in all cases (plus additional, non-specific, lower molecular weight bands which did not interfere with transfections). The number of plaques obtained after transfection of *E. coli* cells with these PCR products are shown below the brackets.

TABLE S1.

Single-nucleotide mutants of bacteriophage F1 obtained by site-directed mutagenesis

Nucleotide substitution	Gene	Amino acid substitution	Relative fitness effect, s ± SEM
T117A	II	None	0.006 ± 0.009
T178A	II	L194I	-1
G181A	II	V195I	-0.064 ± 0.017
A209C	II	Q204P	-1
C231A	II	S211R	-0.009 ± 0.038
T275A	II	L226STOP	-1
A333T	II	E245D	-1
T336A	II	None	-0.002 ± 0.039
C337A	II	None	0.024 ± 0.027
C370A	II	L258I	0.029 ± 0.051
A431G	II	D278G	-1
T470A	II	L290Q	-1
A499T	II	N301Y	-0.406 ± 0.021
G601A	II	G335S	-0.162 ± 0.021
T657A	II	None	0.008 ± 0.021
G671A	II	W358STOP	-1
A878G	V	None	-0.039 ± 0.021
C1088A	V	None	0.033 ± 0.021
C1161A	VII	None	-0.102 ± 0.016
G1205A	VII	R33K	-0.216 ± 0.008
A1433T	VIII	I45F	-1
T1441A	VIII	Y47STOP	-1
C1449A	VIII	A50E	-0.01 ± 0.035
T1544A	Intergenic	None	-0.514 ± 0.029
T1605A	III	None	-0.005 ± 0.009
T1655A	III	L26STOP	-1
G1739A	III	C54Y	-1
T2004A	III	None	-0.044 ± 0.026
A2045T	III	N156I	-0.25 ± 0.046
C2136A	III	Y186STOP	-1
A2408T	III	D277V	-0.046 ± 0.034
C2423A	III	A282E	-0.059 ± 0.057
T2592A	III	None	0.018 ± 0.016
T2727A	III	Y383STOP	-1
G2791A	III	V405I	0.002 ± 0.022
C2795A	III	A406D	-0.084 ± 0.053
C2799A	III	None	-0.026 ± 0.017

T2937A	VI	F28I	-0.052 ± 0.007
T3016A	VI	L54H	-0.521 ± 0.136
G3077A	VI	None	-0.114 ± 0.02
C3120A	VI	L89I	-0.129 ± 0.017
T3142A	VI	F96Y	-0.603 ± 0.029
G3264A	I	None	-0.007 ± 0.041
C3324A	I	None	-0.008 ± 0.041
C3389A	I	S66Y	-0.085 ± 0.035
T3395A	I	L67STOP	-1
T3564A	I	F123L	-0.663 ± 0.041
T3566A	I	L124H	-1
G3643A	I	D150N	-1
A3649G	I	I152V	-0.130 ± 0.004
G3664A	I	V157I	0.022 ± 0.025
C3748A	I	Q185K	0.074 ± 0.016
T3771A	I	None	-0.012 ± 0.05
G3795A	I	None	0.055 ± 0.024
T3855A	I	None	0.003 ± 0.024
T3927A	I	None	0.039 ± 0.022
T3966A	I	None	-0.007 ± 0.028
T4025A	I	V277D	-0.113 ± 0.011
A4218G	I	None	0.019 ± 0.035
C4266A	IV	S16STOP	-1
T4304A	IV	S29T	0.031 ± 0.017
A4417G	IV	None	0.034 ± 0.035
A4537C	IV	None	-0.032 ± 0.027
T4541A	IV	S108T	0.067 ± 0.016
A4547C	IV	N110H	0.044 ± 0.018
G4552A	IV	None	0.023 ± 0.018
T4603A	IV	N128K	-0.070 ± 0.02
T4615A	IV	None	0.001 ± 0.02
C4620A	IV	T134N	-0.788 ± 0.032
A4634G	IV	N139D	-0.694 ± 0.014
T4747A	IV	None	0.045 ± 0.017
C4938A	IV	S240Y	-0.280 ± 0.048
G5120A	IV	E301K	-0.096 ± 0.079
G5158A	IV	None	-0.001 ± 0.025
G5168A	IV	V317I	-0.006 ± 0.03
G5253A	IV	S345N	-0.037 ± 0.052
A5265C	IV	Q349P	-0.210 ± 0.071
G5271A	IV	S351N	-0.119 ± 0.027
A5286T	IV	N356I	-0.371 ± 0.058

A5322G	IV	D368G	-0.526 ± 0.053
G5389A	IV	None	0.015 ± 0.02
T5472A	IV	V418D	-1
C5575A	Intergenic	None	-0.03 ± 0.035
C5649A	Intergenic	None	-0.458 ± 0.091
C5727A	Intergenic	None	-0.019 ± 0.013
A5740C	Intergenic	None	-0.146 ± 0.053
C5753A	Intergenic	None	-0.008 ± 0.032
C5771A	Intergenic	None	-0.053 ± 0.024
G5876A	Intergenic	None	-0.105 ± 0.075
A5903G	Intergenic	None	0.030 ± 0.028
C5966A	Intergenic	None	-0.082 ± 0.042
G6028A	II	R8Q	-0.251 ± 0.026
G6090A	II	V29I	-0.006 ± 0.044
T6091A	II	V29E	-1
T6266A	II	None	-0.051 ± 0.018
A6276G	II	I91V	-0.152 ± 0.042
C6286A	II	S94Y	-1
T6308A	II	None	0.044 ± 0.031
C6344A	II	C112STOP	-1
T6362A	II	None	-0.012 ± 0.012