



ELSEVIER

Fitness effects of beneficial mutations: the mutational landscape model in experimental evolution

Andrea J Betancourt¹ and Jonathan P Bollback²

The mutational landscape model is a theoretical model describing sequence evolution in natural populations. However, recent experimental work has begun to test its predictions in laboratory populations of microbes. Several of these studies have focused on testing the prediction that the effects of beneficial mutations should be roughly exponentially distributed. The prediction appears to be borne out by most of these studies, at least qualitatively. Another study showed that a modified version of the model was able to predict, with reasonable accuracy, which of a ranked set of beneficial alleles will be fixed next. Although it remains to be seen whether the mutational landscape model adequately describes adaptation in organisms other than microbes, together these studies suggest that adaptive evolution has surprisingly general properties that can be successfully captured by theoretical models.

Addresses

¹Institute for Evolutionary Biology, King's Buildings, University of Edinburgh, Edinburgh, EH9 3JT, UK

²Center for Bioinformatics, University of Copenhagen, Universitetsparken 15, Building 10, Copenhagen DK, DK-2100, Germany

Corresponding author: Betancourt, Andrea J
(abetanco@staffmail.ed.ac.uk)

Current Opinion in Genetics & Development 2006, **16**:618–623

This review comes from a themed issue on
Genomes and evolution
Edited by Chris Tyler-Smith and Molly Przeworski

Available online 19th October 2006

0959-437X/\$ – see front matter

© 2006 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.gde.2006.10.006](https://doi.org/10.1016/j.gde.2006.10.006)

Introduction

Beneficial mutations are exceedingly rare and thus difficult to study in natural populations. In laboratory populations of microbes, however, they are more easily characterized, because they are reasonably common and often have large, easily detected, fitness effects [1]. In fact, experimental evolution studies of microbes have recently attempted to study not just the genetic basis of specific adaptive traits but also *general* characteristics of beneficial mutations [2]. Several of these new studies have focused on describing how beneficial fitness-effects are distributed among mutations: are fitness effects divided evenly among mutations? Or do a few

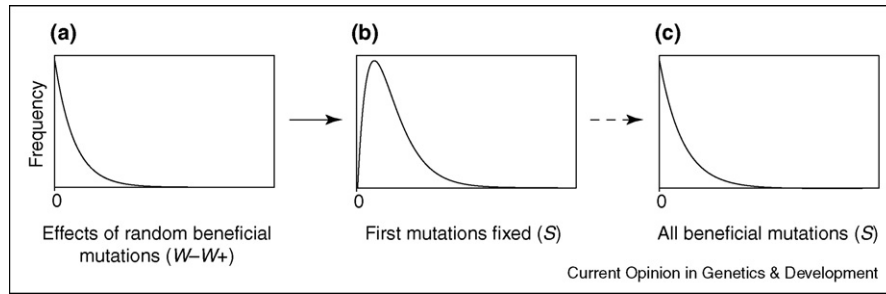
mutations account for most of the possible fitness increase? At the same time, recent theoretical work has begun to generate specific predictions for the distribution of beneficial effects. Consequently, experimentalists can now test the predictions of a particular theoretical model, rather than attempting to infer a distribution of effects directly from laboratory measurements of fitness.

There are actually several sets of beneficial mutations of interest: these include new, random mutations that have not yet experienced either selection or loss due to random genetic drift (Figure 1a); the subset of those beneficial mutations that have survived stochastic loss and become fixed as the first step in adaptation starting from a given wild type (Figure 1b); and all beneficial mutations that have become fixed during the course of adaptation to a new environment. Figure 1 summarizes theoretical expectations for the distribution of effects — given in terms of either selection coefficients or absolute fitness differences — for each set of mutations, which fit either a roughly exponential (Figures 1a and 1c) or a gamma-shaped distribution (Figure 1b). Note that the distributions shown here are derived without incorporating clonal interference, the competition between beneficial mutations that can occur in large populations with restricted recombination. Clonal interference models have been the subject of intense theoretical and experimental work (e.g. [3–6,7]) that is unfortunately beyond the scope of this review, as is other work on the distribution of mutational effects [8]. Instead, we focus here on studies testing the predictions of the mutational landscape model, some of which are shown in Figures 1a and 1c.

The mutational landscape model

The mutational landscape model was originally conceived as a description of molecular evolution by Maynard-Smith in 1962 [9,10] and was further developed by Gillespie [11], and Kauffman and Levin [12]. This model pictures evolution as a process in which a population moves from a wild type sequence to one of its one-step mutant neighbors (i.e. those mutant sequences that differ from the wild type by a single point mutation), which then becomes the new wild type. In the case of a DNA sequence, for instance, a sequence L nucleotides long would have $3L$ one-step mutant neighbors (because there are three alternative nucleotides at each site) and the sequence can adapt by fixing any of the $3L$ mutant alleles that confer greater fitness (Figure 2). The process is repeated until there are no one-step mutants fitter than the wild type, at which point the population will have

Figure 1



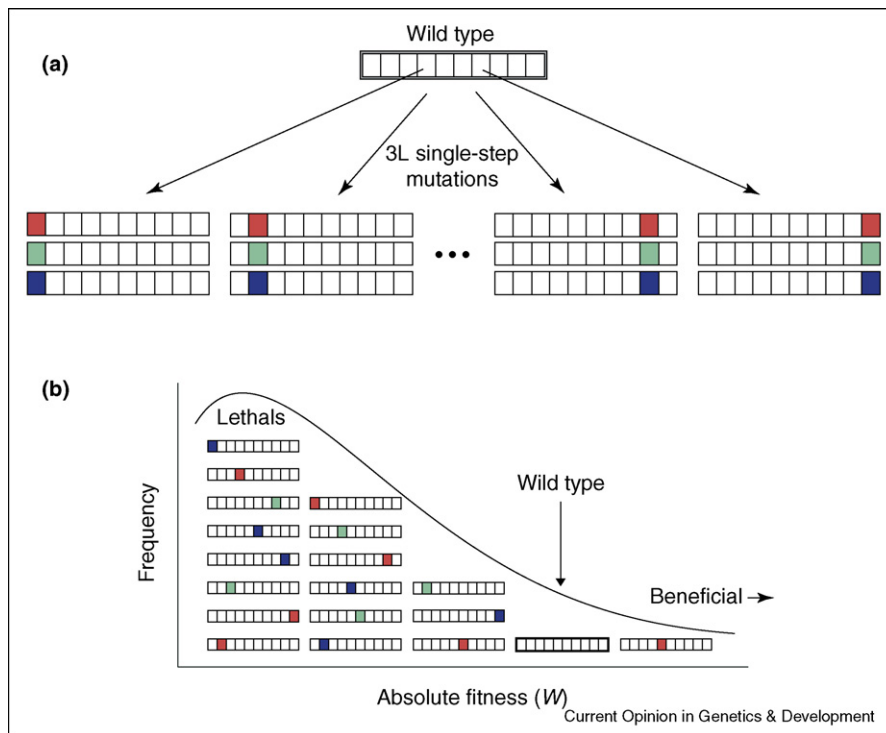
Theoretical distributions for the effects of beneficial mutations. **(a)** The frequency distribution for the effects of random beneficial mutations before any have suffered loss due to genetic drift [14], given in terms of the difference in absolute fitness from the wild type. **(b)** The distribution of the effects of the first mutations fixed [6,25], given in terms of the selection coefficient, S , is shifted to the right owing to genetic drift. **(c)** The distribution of all mutations fixed in the course of adaptation (i.e. the substitutions required to reach a locally optimal sequence [15]) is roughly exponential. Abbreviations: W , absolute fitness.

reached either the fittest possible sequence or a locally optimal one.

The first step in mathematically describing this evolutionary process is to rank the $3L$ mutant sequences — both beneficial and deleterious — by their absolute

fitnesses, with the rank of the best allele being equal to 1. For a beneficial mutation with rank i , there are $i-1$ beneficial mutations. The ranked alleles form a frequency distribution (Figure 2); with enough mutant sequences, this distribution can be approximated by a continuous probability distribution, such as an exponential or gamma.

Figure 2



The mutational landscape model for a hypothetical DNA sequence. White squares represent the wild type state, and colored squares represent one of the three possible mutations at a site. Starting from the wild type sequence of length L **(a)** there are $3L$ one-step mutant sequences. This distribution of mutations is mapped onto a hypothetical fitness distribution of absolute fitness **(b)**. Most of the single-step mutations fall to the left of the wild type and are deleterious or, as shown on the far left, lethal ($W = 0$). The right side of the distribution from the wild type sequence onward comprises the beneficial single step mutants. Notice that there are many fewer beneficial mutations and that they occupy the extreme right tail of the distribution.

At first, this appears to be a dead end, because very little is known about the shape this distribution should take. Gillespie realized, however, that regardless of the exact shape of the distribution, the wild type sequence almost certainly occurs in its high-fitness tail [11]. Since the tails of many distributions behave similarly, in a manner mathematically described by the extreme-value theory, Gillespie's insight made many biologically interesting problems suddenly mathematically tractable. Extreme-value theory has been used to derive, for instance, the number of adaptive substitutions that occur after an environmental change on a rugged fitness landscape [11] and the probability of parallel adaptive evolution [13]. Recent development of the theory by Orr [14,15] has generated some specific predictions that have been tested in three organisms to date, two viruses and one bacterial species. Two of these studies focus on the distribution of new, random mutations before selection, whereas the third focuses on fixed mutations. We discuss each in turn below.

Empirical tests of the mutational landscape model

The major difficulty in studying random beneficial mutations is in obtaining a sample of mutations unaffected by selection and genetic drift. Sanjuan *et al.* [16**] used a straightforward, brute-force approach: using site-directed mutagenesis in vesicular stomatitis virus, they created 91 mutations, seven of which turned out to be beneficial in replicate competition assays. As expected under the mutational landscape model, the exponential distribution (Figure 1a [14]) produced a good fit to the observed fitness effects of these mutations ($R^2 = 0.888$). However, Sanjuan *et al.* rejected the exponential because a more flexible gamma distribution provided a slightly but significantly better fit ($R^2 = 0.953$; $P = 0.046$). One caveat is that the Sanjuan *et al.* study used a different metric of fitness than the mutational landscape theory work used; in fact, they used the fitness measurement typically used in experimental evolution studies, and not that used in population genetic theory (see the study by Bull *et al.* [1] for an explanation). Because the exponential distribution is only marginally rejected in this study, this slight technical difference in fitness definitions may be of importance. In any case, given the limitations of this kind of experiment (see below), the results qualitatively support the idea that new mutations should have roughly exponential effects.

Kassen and Bataillon [17**] also test for an exponential distribution of the effects of random beneficial mutations, but in *Pseudomonas* bacteria. They used a standard assay for recovering new random antibiotic resistance mutations; if antibiotic resistance is uncorrelated with fitness in other environments, mutations selected for resistance will have randomly distributed fitness effects in other environments. Consistent with this idea, they found that most

of the recovered mutations were not beneficial in the absence of the antibiotic. From 2016 populations, they collected 665 antibiotic-resistant mutants and used this collection of mutations to perform two experiments testing the mutational landscape model. First, the mutational landscape model predicted that the beneficial mutations have exponentially distributed effects. Kassen and Bataillon identified 18 mutations with significantly higher fitness than the wild type in the absence of the antibiotic, and these showed the predicted exponential distribution of effects. A likelihood ratio test showed that a gamma distribution does not provide a significantly better fit ($\text{Log}L_{\text{exp}} = 24.31$ versus $\text{Log}L_{\text{exp}} = 25.45$, $\chi^2 = 2.28$, $P = 0.27$) [17**]. Second, the mutational landscape model predicts a surprising 'invariance' principle: it argues that, regardless of the fitness rank of the wild type allele, beneficial mutations arising from it have the same mean effect, equal to $E(\Delta_1)$, the expected fitness difference between the best and next best allele. In other words, under the mutational landscape model, the average beneficial effect does not become smaller as the rank of the wild type allele improves. To test this principle, Kassen and Bataillon measured fitness effects of 95 of their recovered mutations in four environments. Since the fitness gap between the best and next best allele (Δ_1) is likely to differ for each environment, the mean beneficial effect of the measured mutations should also differ for each environment. Of the 95 mutations measured, the number with beneficial effects varied from 8 to 25, depending on the environment, but they showed the predicted exponential distribution of effects regardless of the environment. Unexpectedly, however, the mean beneficial effect was the same for all four environments; this result might contradict the prediction of the mutational landscape model if in fact Δ_1 does differ between the environment.

The third test of the mutational landscape model, by Rokyta *et al.* [18**], evaluates its ability to predict which beneficial allele will be the next one fixed. Given the fitness ranks of the beneficial alleles (again, where 1 is the rank of the best allele and i is the rank of the wild type), the theory predicts that the next allele fixed will have rank j , with a mean probability as follows:

$$E[P_{ij}] = \frac{1}{i-1} \sum_{k=j}^{i-1} \frac{1}{k} \quad [15]$$

Essentially, this expectation results from the fact that, in the absence of clonal interference, each allele has an independent probability of fixation that depends only on the size of its beneficial effect (S) relative to other beneficial alleles. Therefore, the probability that allele j will be the next fixed is as follows:

$$\frac{S_j}{\sum_{k=1}^{i-1} S_k} \quad [11]$$

With extreme value theory, the mutational landscape model can predict these relative effects for alleles given their fitness ranks.

To test the accuracy of these predictions, Rokytá *et al.* [18**] propagated 20 populations of a natural isolate bacteriophage — similar to the laboratory strain Φ X174 — until whole-genome sequencing showed that each population had fixed a single beneficial mutation. A total of nine amino acid substitutions were fixed in the 20 populations, with the most frequent fixed in six populations. After measuring and ranking the fitness of each allele, they compared the number of times each was fixed among the 20 populations with the theoretical expectation based on its rank. The model performed surprisingly poorly: for example, the fittest allele (with $j = 1$) was fixed in only 1 out of 20 populations (observed $P_{10,1} = 0.05$) versus the expected 6 (expected $P_{10,1} = 0.314$).

Why did the model perform so poorly? The problem might not be that the model fails to predict the effects of beneficial alleles, but instead that the model assumes equal mutation rates. Instead, mutation rates are known to be biased, with transitions being more common than transversions. Thus, the fittest allele in this experiment might have been fixed infrequently because it occurred rarely, as it happened to be a G→T transversion. Incorporating estimates of actual mutation rates substantially improved the fit of the model ($P = 0.10$ for the equal mutation rates version of the model; $P = 0.49$ for the mutation-adjusted model). In fact, the mutation-adjusted version of the model performed almost as well as a different model that also incorporated information about the fitness of each allele and the demography of the experimental populations ($P = 0.67$) [19]. In other words, the general properties of molecular evolution captured by the mutational landscape model explain the results of Rokytá *et al.*'s experiment reasonably well, independent of most of the particular details of the experimental system [18**].

Despite the impressive amount of work involved in these studies, each has its limitations and drawbacks. In cases where the identities of the beneficial mutations are known, the number of mutations studied might have been too few to ensure a representative sample or a powerful test of the model. In cases where the identities of mutations are not known, there is no guarantee that the observed distribution is not confounded with non-independent measurements performed on the same mutations. Both kinds of study also suffer from two limitations common to most experimental evolution work. First, fitness measurements are noisy, and so there is limited power to distinguish between different distributions of fitness effects. Second, since every method for detecting beneficial mutations has some limit to its resolution,

mutations with small effects may have been preferentially missed; a large excess of small effect mutations might go undetected. To date, only Rokytá *et al.* have made a convincing case for qualitatively rejecting the predictions of the model, although the discrepancy was easily reconciled by incorporating a biologically realistic mutation process.

Correlated fitness landscapes

At least one study suggests that the mutational landscape model might not adequately describe evolution in all molecules. Following a quasi-empirical approach pioneered by Fontana *et al.* [20], Cowperthwaite *et al.* [21*] studied the effects of random mutations that improve the thermostability of a given RNA secondary structure. Using this approach, which is much less labor-intensive than the empirical work outlined above, they studied beneficial mutations arising from 6959 wild type sequences. The resulting distribution of effects was far from the predicted exponential distribution, with a substantial excess of very small effect mutations. There are, however, some potential limitations of this test: RNA stability might not be a good proxy for fitness, and the short RNA molecule studied might be atypical. But it is also possible that the discrepancy stems from a potentially nonbiological assumption of the mutational landscape model: the exponential prediction comes from a version of the model that assumes a maximally rugged fitness landscape (i.e. one in which the fitness of the wild type sequence is uncorrelated with the fitness of its one-step neighbors). In reality, of course, a wild type sequence and its mutant alleles can show any degree of correlation, and in the case of thermostability of RNA structures, wild type and mutant fitnesses are known to be correlated [20].

Orr [22*] recently investigated one version of a correlated landscape in which the sequence is divided into blocks that each independently and additively affect fitness. Given that single mutations change the fitness of only one block, mutant and wild type alleles have similar (correlated) fitnesses. This early work on correlated fitness landscapes suggests that some predictions of the mutational landscape model are robust to some amount of correlation in the landscape. In particular, the beneficial fitness effects of random mutations on correlated landscapes should still be exponentially distributed. However, this version of the model has more stringent requirements for the length of the molecule: in order for extreme value theory to work, there must be enough mutant sequences per block to ensure that the wild type sequence occurs in the high-fitness tail of roughly continuous distribution. In other words, the model might not perform well for molecules that are both correlated and short, such as the 76-nucleotide RNA molecule studied by Cowperthwaite *et al.* [21*].

Conclusions

The empirical and theoretical work summarized here represents an important first step toward understanding the general properties of the genetics of adaptation. Although the particular biology of a system almost certainly affects the detailed picture, the mutational landscape model appears to provide a good overall description of the distribution of mutational effects. But one key frustration with the theory, particularly in relation to the mutational landscape work, is that it does not adequately describe evolution in most experimental evolution populations. Although these populations provide the best available testing ground for the theory, they typically experience more mutations and stronger selection than current versions of the theory assume. The experimental work, in turn, has suffered from technical limitations: for the exact genetic basis of adaptation to be known, the organism used must be easily resequenced. To date, such studies have been restricted to simple viruses, which might differ in some systematic way from more complex organisms. Fortunately, advances in DNA sequencing technology might soon make similar high-resolution experiments possible in organisms other than viruses [23,24].

Acknowledgements

We thank K Dyer, P Haddrill, A Orr, D Presgraves and M Przeworski for critical reading of the manuscript. This work was supported by funding from the Royal Society to AJB and a grant from the Danish Medical Research Council (FSS) to JPB.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Bull JJ, Badgett MR, Wichman HA: **Big-benefit mutations in a bacteriophage inhibited with heat.** *Mol Biol Evol* 2000, **17**:942-950.
2. Elena SF, Lenski RE: **Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.** *Nat Rev Genet* 2003, **4**:457-469.
3. Miralles R, Gerrish PJ, Moya A, Elena SF: **Clonal interference and the evolution of RNA viruses.** *Science* 1999, **285**:1745-1747.
4. Gerrish PJ, Lenski RE: **The fate of competing beneficial mutations in an asexual population.** *Genetica* 1998, **102-103**:127-144.
5. de Visser JA, Rozen DE: **Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*.** *Genetics* 2006, **172**:2093-2100.
6. Rozen DE, de Visser JA, Gerrish PJ: **Fitness effects of fixed beneficial mutations in microbial populations.** *Curr Biol* 2002, **12**:1040-1045.
7. Hegreness M, Shores N, Hartl D, Kishony R: **An equivalence principle for the incorporation of favorable mutations in asexual populations.** *Science* 2006, **311**:1615-1617.
The authors argue that, in large populations that experience many beneficial mutations, clonal interference eliminates all but a small subset of mutations that tend to have large effects. Given that the distribution of these successful mutations has a small variance, the authors reason that evolution in these populations can be described as if although only mutations with a single effect had occurred. Using a technologically innovative approach, they follow beneficial substitutions in 72 populations and show that the results can be described using only two effective parameters, a constant effect size for beneficial mutations, and the rate at which they occur.
8. Barrett RDH, MacLean RC, Bell G: **Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations.** *Biology Letters* 2006, **2**:236-238.
The authors examined the fitness of 96 bacterial populations, 68 of which were putatively fixed for a single adaptive allele. The results show that most of these populations fixed mutations with intermediate effects, but do not fit the theoretical expectation well, probably because selection was stronger than assumed by the theory.
9. Smith JM: **Natural selection and the concept of a protein space.** *Nature* 1970, **225**:563-564.
10. Maynard Smith J: **The limitations of molecular evolution.** In *The Scientist Speculates*. Edited by Good IJ. Heinemann; 1962:252-256.
11. Gillespie JH: **Molecular evolution over the mutational landscape.** *Evolution Int J Org Evolution* 1983, **38**:1116-1129.
12. Kauffman S, Levin S: **Towards a general theory of adaptive walks on rugged landscapes.** *J Theor Biol* 1987, **128**:11-45.
13. Orr HA: **The probability of parallel evolution.** *Evolution Int J Org Evolution* 2005, **59**:216-220.
14. Orr HA: **The distribution of fitness effects among beneficial mutations.** *Genetics* 2003, **163**:1519-1526.
15. Orr HA: **The population genetics of adaptation: the adaptation of DNA sequences.** *Evolution Int J Org Evolution* 2002, **56**:1317-1330.
16. Sanjuan R, Moya A, Elena SF: **The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus.** *Proc Natl Acad Sci USA* 2004, **101**:8396-8401.
Part of a larger study examining the effects of random mutations — both beneficial and deleterious — generated by site-directed mutagenesis in vesicular stomatitis virus. This brute-force approach has led to several interesting results, detailed in this and in companion studies [26,27].
17. Kassen R, Bataillon T: **Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria.** *Nat Genet* 2006, **38**:484-488.
This work tests the two predictions of the mutational landscape model using a large sample of beneficial mutations in *Pseudomonas* bacteria.
18. Rokyta DR, Joyce P, Caudle SB, Wichman HA: **An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus.** *Nat Genet* 2005, **37**:441-444.
This study tests the ability of two different models — one of them being the mutational landscape model — to predict how often each of a set of known beneficial mutations will be fixed in replicate populations.
19. Wahl LM, Krakauer DC: **Models of experimental evolution: the role of genetic chance and selective necessity.** *Genetics* 2000, **156**:1437-1448.
20. Fontana W, Stadler PF, Bornberg-Bauer EG, Griesmacher T, Hofacker IL, Tacker M, Tarazona P, Weinberger ED, Schuster P: **RNA folding and combinatorial landscapes.** *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 1993, **47**:2083-2099.
21. Cowperthwaite MC, Bull JJ, Meyers LA: **Distributions of beneficial fitness effects in RNA.** *Genetics* 2005, **170**:1449-1457.
This quasi-empirical study tests the ability of the mutational landscape model to predict the effects of random mutations on the thermostability of RNA structures, and finds an unexpected large excess of mutations with small effects.
22. Orr HA: **The population genetics of adaptation on correlated fitness landscapes: the block model.** *Evolution Int J Org Evolution* 2006, **60**:1113-1124.
This theoretical study re-examines earlier predictions of the mutational landscape model, but on a correlated fitness landscape. In general, most predictions of the earlier work turn out to be robust, except for those regarding adaptive walks.

23. Gresham D, Ruderfer DM, Pratt SC, Schacherer J, Dunham MJ, Botstein D, Kruglyak L: **Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray.** *Science* 2006, **311**:1932-1936.
24. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al.*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376-380.
25. Kimura M: *The neutral theory of molecular evolution.* Cambridge, UK: Cambridge University Press; 1983
26. Rafael Sanjuán R, Moya A, Elena SF: **The contribution of epistasis to the architecture of fitness in an RNA virus.** *Proc Natl Acad Sci USA* 2004, **101**:15372-15379.
27. Rafael Sanjuán R, Cuevas JM, Moya A, Elena SF: **Epistasis and the adaptability of an RNA virus.** *Genetics* 2005, **170**:1001-1008.