

The Genetics of Adaptation for Eight Microvirid Bacteriophages

Darin R. Rokyta · Zaid Abdo · Holly A. Wichman

Received: 26 November 2008 / Accepted: 3 July 2009 / Published online: 20 August 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Theories of adaptive molecular evolution have recently experienced significant expansion, and their predictions and assumptions have begun to be subjected to rigorous empirical testing. However, these theories focus largely on predicting the first event in adaptive evolution, the fixation of a single beneficial mutation. To address long-term adaptation it is necessary to include new assumptions, but empirical data are needed for guidance. To empirically characterize the general properties of adaptive walks, eight recently isolated relatives of the single-stranded DNA (ssDNA) bacteriophage ϕ X174 (family Microviridae) were adapted to identical selective conditions. Three of the eight genotypes were adapted in replicate, for a total of 11 adaptive walks. We measured fitness improvement and identified the genetic changes underlying the observed adaptation. Nearly all phages were evolvable; nine of the 11 lineages showed a significant increase in fitness. However, fitness plateaued quickly, and adaptation was achieved through only three substitutions on average. Parallel

evolution was rampant, both across replicates of the same genotype as well as across different genotypes, yet adaptation of replicates never proceeded through the exact same set of mutations. Despite this, final fitnesses did not vary significantly among replicates. Final fitnesses did vary significantly across genotypes but not across phylogenetic groupings of genotypes. A positive correlation was found between the number of substitutions in an adaptive walk and the magnitude of fitness improvement, but no correlation was found between starting and ending fitness. These results provide an empirical framework for future adaptation theory.

Keywords Experimental evolution · Adaptation · Molecular evolution · Virus · Bacteriophage

Introduction

Adaptation theories have seen a recent expansion and elaboration, due in part to the emergence of a new modeling approach. Historically, Fisher's geometric model of adaptation (Fisher 1930) was the most commonly used framework for studying adaptation (e.g., Orr 1998, 2000). This model considers adaptation to a fixed optimum in a continuous, high dimensional phenotypic space. However, adaptation actually happens in the space of DNA sequences (Maynard Smith 1970), and its inherent discreteness and the rarity of beneficial mutations were used by Gillespie (1983, 1984, 1991) as the basis for the mutational landscape model of adaptation. This model has engendered considerable recent interest (Betancourt and Bollback 2006) and has been extended and refined by various authors (Orr 2002, 2003; Rokyta et al. 2006a; Joyce et al. 2008). This work has produced many general predictions,

D. R. Rokyta · H. A. Wichman (✉)
Department of Biological Sciences, University of Idaho,
PO 443051, Moscow, ID 83844-3051, USA
e-mail: hwichman@uidaho.edu

Z. Abdo
Departments of Mathematics and Statistics, University of Idaho,
Moscow, ID, USA

Z. Abdo · H. A. Wichman
Initiative for Bioinformatics and Evolutionary Studies,
University of Idaho, Moscow, ID, USA

Present Address:

D. R. Rokyta
Department of Biological Science, Florida State University,
Tallahassee, FL, USA

some of which have been tested (Rokyta et al. 2005), yet there is a pronounced bias in the predictions: they mostly deal with only a single step in adaptation.

Why does the current theory focus on only one step? The derivations of most single-step results require only rather general assumptions about the distribution of fitness values among genotypes, specifically that this distribution is amenable to extreme value theory (e.g., Orr 2002; Joyce et al. 2008). Furthermore, these assumptions can be (Beisel et al. 2007) and have been (Sanjuán et al. 2004; Kassen and Bataillon 2006; Rokyta et al. 2008; MacLean and Buckling 2009) explored with empirical data. To extend predictions to multiple steps, however, it is necessary to incorporate epistasis into the model, and in particular, a model of the fitness landscape. Several models of fitness landscapes in sequence spaces have been proposed (e.g., Kauffman and Levin 1987; Kauffman 1993; Perelson and Macken 1995), and some have even been implemented within the mutational landscape framework (Gillespie 1983, 1984, 1991; Orr 2002, 2006a; Rokyta et al. 2006a), yet it remains uncertain as to which, if any, of these approaches describes reality. Furthermore, while the single-step results appear to be fairly robust to violations of many of the underlying assumptions, this robustness does not hold for multiple-step predictions (Orr 2006b; Joyce et al. 2008).

Empirical data are needed concerning multiple-step bouts of adaptation, i.e., adaptive walks. The typical predictions made from sequence-based fitness landscape models generally focus on properties of the space that are less sensitive to population dynamics, for example, the number of steps in an adaptive walk, the number of local optima, the number of accessible local optima from a given genotype, and the distribution of optima in the landscape (e.g., Kauffman and Levin 1987; Kauffman and Weinberger 1989; Macken and Perelson 1989; Macken et al. 1991; Kauffman 1993; Perelson and Macken 1995; Stadler and Happel 1999). There are a multitude of different models for adaptive walks, and some of these require parameterization. For example, the *NK* model (Kauffman and Levin 1987; Kauffman 1993), requires the specification of *N*, the number of sites, and *K*, the average number of other sites that affect the fitness contribution of each site. Empirical data are needed to determine which models are consistent with reality and which parameters give reasonable predictions. For example, Kauffman and Weinberger (1989) used empirically estimated walk lengths as a basis for choosing the appropriate value of *K* for the maturation of the immune response. Numerous microbial evolution studies have looked at the adaptation of one or two genotypes (e.g., Crill et al. 2000; Kichler Holder and Bull 2001; Cuevas et al. 2002; Rokyta et al. 2005; Wichman et al. 2005; Perfeito et al. 2007; Pepin and Wichman 2008; Betancourt 2009). Other studies have examined multiple genotypes but

generally lack information concerning the underlying genetics of adaptation (e.g., Travisano et al. 1995; Bull et al. 2004a), which is necessary for a DNA-based model of adaptation. A notable exception is the recent study by Bollback and Huelsenbeck (2009) characterizing the genetics of adaptation across three species of RNA bacteriophages, focusing on rates of parallel evolution. While these studies have been informative for their intended purpose, further experimental data on the genetics of adaptation from multiple genotypes are needed to extend the findings to a more general theoretical characterization of adaptation.

In the present work, we adapted eight different naturally occurring bacteriophages originally isolated by Rokyta et al. (2006b) to the same standard laboratory culturing conditions. We then characterized the fitness improvement, fitness maxima, and the genetics underlying adaptation. The selected genotypes belonged to four distinct phylogenetic clades, and therefore we were able to compare adaptation both across genotypes as well as across groups. Additionally, the members of one group were adapted in replicate to explore the repeatability of adaptation. By observing a large number of adaptations, we were able to examine general properties of adaptive evolution, such as the relationship between initial and final fitness and the number of substitutions involved in an adaptive walk. Overall these data provide some rough guideposts for future theoretical treatments of adaptive evolution beyond the first step.

Materials and Methods

Bacteriophage Genotypes

We selected eight different genotypes from among the microvirid bacteriophages isolated by Rokyta et al. (2006b) to be adapted to standard laboratory batch culturing. The microvirid coliphages are lytic, single-stranded DNA bacteriophages with genome sizes ranging from ~5.3 to 6.3 kilobases, encoding 11 genes. The prototypical member of this family, ϕ X174, has a generation time of ~12–15 min under standard laboratory conditions. The selected phages and their GenBank accession numbers are as follows: ID12 (DQ079905), ID8 (DQ079898), NC6 (DQ079907), ID2 (DQ079869), NC41 (DQ079890), WA11 (DQ079895), NC28 (DQ079875), WA13 (DQ079873). Genotypes were chosen to represent the range of diversity of the known microvirid phages infecting *Escherichia coli* C (Fig. 1). The fitnesses of 34 of the 42 originally isolated genotypes were measured (data not shown), and seven genotypes were excluded from consideration due to insufficient growth rates. Eight genotypes with the desired

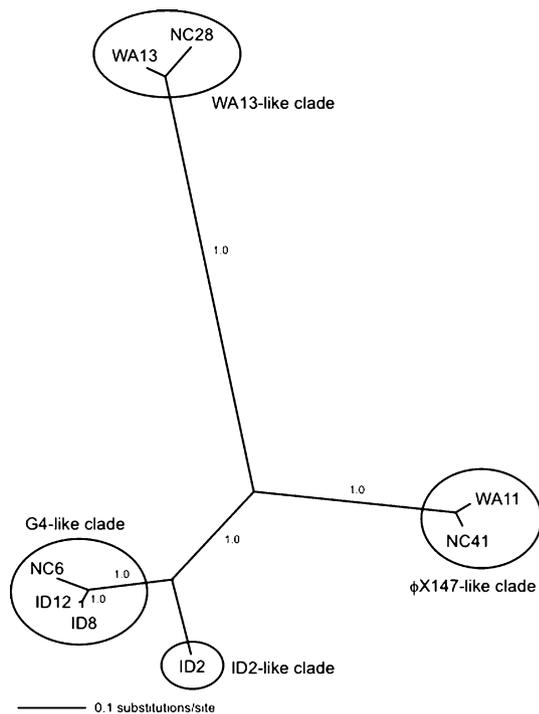


Fig. 1 The maximum a posteriori probability (MAP) phylogeny of the phage genotypes used in the present study based on full genome sequences. Nodal supports are given as posterior clade probabilities. For our analysis, we considered 4 phylogenetic clades: ϕ X174-like, WA13-like, G4-like, and ID2-like

phylogenetic relationship were selected from among those that remained. The first adaptation of a genotype was given an “a” designation (e.g., ID12a). If a genotype was adapted in replicate, the second lineage was labeled “b” (e.g., ID12b). We initiated replicate lineages from independent genetic isolates to preclude spurious parallel evolution due to mutations that may have occurred in plaque growth. A number appending a lineage name indicates a particular flask growth period. For example, ID12a60 indicates the population from the 60th growth period for the ID12a lineage.

Phylogeny Estimation

To illustrate the relationships among the genotypes used in the present study, we conducted a Bayesian phylogenetic analysis. We aligned the nucleotide sequences of the full genomes of the eight genotypes with ClustalW (Thompson et al. 1994). The TrN + G model of DNA substitution was selected using DT-ModSel (Minin et al. 2003), but the slightly more parameter-rich GTR + G model was used for the analysis with MrBayes version 3.0 (Huelsenbeck and Ronquist 2001). This was done for simplicity, although there is evidence that more complex models are preferable in a Bayesian phylogenetic analysis (Huelsenbeck and

Rannala 2004). We used the default priors and a temperature parameter of 0.2 with 4 chains. The chains were run for 1 million generations with the first 100,000 generations discarded as burn-in.

Adaptation Protocol

Our phage passing protocol follows that used by Rokyta et al. (2002). Briefly, our host, *E. coli* C, was allowed to grow to $1\text{--}2 \times 10^8$ cells per ml in phage LB (10 g NaCl, 10 g Bacto tryptone, 5 g yeast extract per liter) supplemented with CaCl_2 to a final concentration of 2 mM in 125 ml flasks shaking at 200 rpm. Phage were added and grown for 40 min (~ 3 generations). The number of phage added was dependent on their fitness as measured during the course of adaptation by monitoring phage concentrations at the end of each growth period. We attempted to maximize the effective population size to reduce waiting times for beneficial mutations while still maintaining a multiplicity of infection (i.e., the ratio of phage to hosts) well below one throughout the growth period. Thus, $10^4\text{--}10^7$ phages were added to initiate each growth period, with less phage being added for high fitness populations and more for low fitness populations. Fitness was measured as phage growth rate and expressed as the \log_2 increase in the total number of phage per hour. All adaptations were performed at 37°C . The host cell densities were kept high relative to phage densities to prevent selection for intra-host competitive ability, thus our environment should select for efficient usage of the host as well as for a reduced latent period (Abedon 1989; Wang et al. 1996). This selection protocol is intended to select for rapid phage replication under standard laboratory conditions. We treated each flask endpoint with chloroform prior to initializing the next growth period to ensure that our adaptations were carried out under the same conditions as our fitness assays. To avoid cross-contamination, adaptations of identical starting genotypes were temporally separated by at least a month. Any cross-contamination of adaptations of different genotypes would have been readily detected through sequencing (see below).

Fitness Assays

Fitness in our selective environment was measured as the \log_2 increase in total number of phage per hour, which gives the number of phage population doublings per hour. Fitness was measured in conditions identical to our selective environment, except that the initial number of phage was reduced to facilitate titer determination. All populations to be assayed were grown in the standard transfer protocol for 40–60 min to produce a fresh stock. Fitness assays were only performed on stocks that were no more

than a week old. Fitness measures were replicated at least 5 times for the initial and final populations of each lineage.

Sequencing

We sequenced the entire genome of each isolate used to initiate a lineage and each final population. Whole population sequencing yields the average sequence of the population, therefore we should only detect mutations that have fixed or reached high frequency. We also sequenced four isolates from the final population of the NC41a lineage for reasons discussed below.

Statistics

To assess the significance of factors affecting the variation observed in final fitness values, we performed an analysis of variance (ANOVA) using PROC GLM in SAS release 9.1 (SAS Institute Inc., Cary, NC). We treated the four phylogenetic groups (shown in Fig. 1) as fixed effects and genotypes nested within groups as random effects. All other statistical analyses were performed in the R statistical environment (R Development Core Team 2006).

Results and Discussion

Bacteriophage Genotypes

The phage genotypes selected for the present study encompass the known diversity of microvirid bacteriophages that infect *E. coli* C (Fig. 1). Rokyta et al. (2006b) determined the phylogenetic relationships among 42 naturally occurring microvirid phages and five previously identified laboratory strains (φ X174, S13, G4, α 3, and φ K) and delineated at least four distinct groups. One of these groups, the α 3-like phages, had no representatives in the collection of natural isolates described by Rokyta et al., therefore they were not included in the present study. We included three isolates from the G4-like group (ID12, ID8, and NC6), two from the φ X174-like group (NC41 and WA11), and two isolates from the WA13-like group (WA13 and NC28). The phage ID2 is a distant outlier to the G4-like group whose group status was found to be unclear. For our analyses, it was treated as a separate group. The three G4-like isolates were adapted in replicate. For a more complete description of the relationships among these phages, see Rokyta et al. (2006b).

Several of the isolates used to initiate lineages differed slightly from the published sequences. It is unclear whether these differences represent sequencing errors in the original data set described by Rokyta et al. (2006b) or mutations that arose during plaque isolation for the present study. It is

clear, however, that these differences were not sequencing errors in the present study, as they were detected in both the initial and final populations of the affected lineages. The isolate used to initiate the WA13a lineage differed from the published WA13 sequence at position 3561 (G \rightarrow T). The isolate used to begin the WA11a lineage differed from the published WA11 sequence at positions 4946 (G \rightarrow T) and 4948 (G \rightarrow T). Finally, the isolate used to initiate the NC41a lineage differed from the published NC41 sequence at position 1698 (A \rightarrow G). The isolates used to initiate the NC6a and NC6b lineages differed from the published NC6 sequence at five positions. These five mutations were determined to have been errors in the original sequence, and the GenBank file (DQ079907) has been updated accordingly.

Fitness Plateaus

To provide useful information regarding certain properties of adaptive walks, we must be confident that adaptation is as near to completion as technically feasible. Lineages were continued until an apparent fitness plateau had been reached over at least 20 growth periods (\sim 60 generations). The fitness plateau was then confirmed using a one-tailed *t* test with equal variances comparing the approximate fitnesses (based on monitoring the population size at the end of each growth period) of the last ten transfers to the fitnesses of the previous ten at a significance level of $\alpha = 0.05$. These fitness measures are obviously not independent, thus violating one of the assumptions of a *t* test. However, since they are most likely positively correlated, the “true” error variance is larger than that calculated assuming independence, making the test as performed more stringent in identifying equal fitnesses. The number of growth periods per adaptation varied from 30 to 200 (Table 1).

It is possible that despite the constancy of fitness over the last 20 growth periods, adaptation may have been incomplete in some or even all lineages. As discussed more thoroughly by Bull et al. (2004b), our method likely precluded the fixation of mutations with small fitness effects. Using the deterministic equations for frequency change described by Bull et al. (2000), we find that even with 20 growth periods, mutations that increase the growth rate by as much as one doubling per hour may not have had time to reach a frequency of 0.5. This assumes a generation time of 12 min and that the initial frequency of the mutation was 10^{-5} . In a background with a fitness of 25 doublings per hour, this corresponds to a selection coefficient of 0.04. While mutations below a given magnitude of fitness effect may not have had time to fix, the effects of such mutations would be small relative to the average total improvement in fitness, thus having little effect on conclusions based upon

Table 1 A summary of 11 adaptive walks

Lineage	Clade	# of subst. ^a	# of flasks	N_H^b	w_i^c	Std err	w_f^c	Std err	$\Delta w^{c,d}$
ID12a	G4	2	60	1.1×10^5	20.8	0.26	24.6	0.25	3.8*
ID12b	G4	2	70	4.7×10^5	21.1	0.17	24.4	0.35	3.3*
ID8a	G4	5	80	8.3×10^4	15.7	0.32	25.5	0.35	9.8*
ID8b	G4	4	90	5.8×10^5	14.9	0.10	25.6	0.49	10.7*
NC6a	G4	4	70	3.5×10^5	22.4	0.28	26.4	0.25	4.0*
NC6b	G4	3	70	1.0×10^6	22.5	0.16	26.9	0.53	4.4*
ID2a	ID2	9	200	2.1×10^6	8.7	0.51	22.2	0.42	13.5*
NC41a	ϕ X174	0	30	5.0×10^4	26.0	0.23	25.7	0.50	-0.3
WA11a	ϕ X174	1	40	1.5×10^6	21.9	0.36	23.3	0.34	1.4
NC28a	WA13	3	80	5.8×10^4	19.7	0.28	23.9	0.04	4.2*
WA13a	WA13	2	40	6.3×10^5	19.8	0.22	29.1	0.36	9.3*

^a Based on population sequences at the end of adaptation

^b Harmonic mean of population size over the course of adaptation

^c All fitness measures (w) are given in population doublings per hour. The subscripts “i” and “f” indicate “initial” and “final”, respectively

^d *Indicates a significant fitness improvement

final fitness values or fitness improvement. Conclusions based upon the number of substitutions may be more strongly affected, so this unavoidable caveat should be kept in mind.

Whether or not adaptation was truly complete in the sense of exhausting the supply of even small-effect beneficial mutations, we found that adaptation is characterized by a rapid approach to a sustained fitness plateau. This appears to be a general property of adaptation and has been observed repeatedly in experimental evolution studies (Lenski et al. 1991; Lenski and Travisano 1994; Cooper and Lenski 2000; Kichler Holder and Bull 2001; Rokytka et al. 2002; de Visser and Lenski 2002; Elena and Lenski 2003; Bull et al. 2004b; Pepin and Wichman 2008). In our experiments, and assuming a generation time of ~ 12 min, fitness asymptoted within about 600 generations and in some cases within as few as 90.

Evolvability

The capacity of an organism to adapt in the presence of a selective pressure is generally referred to as evolvability (Wagner and Altenberg 1996; Kirschner and Gerhart 1998; McBride et al. 2008). Nine of the 11 experimental lineages responded to selection as evidenced by significant improvement in fitness. All lineages except for NC41a and WA11a, our only representatives of the ϕ X174-like group (Fig. 1), had a significantly higher final fitness than initial fitness (t test, one-sided, unequal variance, $P < 0.01$ Bonferroni corrected; $P > 0.10$ for NC41a and WA11a). Figure 2 shows the initial and final fitnesses of the 11 lineages; Table 1 provides an overall summary of the adaptations. At least six of the eight genotypes were evolvable given our

experimental conditions and the amount of time allowed for a response to selection. The largest fitness gain was achieved by lineage ID2a, which increased its fitness by 13.5 doublings per hour. This corresponds to more than a 500-fold increase in the total number of progeny produced per phage in 40 min, our standard growth period. This genotype, however, began with a much lower fitness than any other genotype and also achieved the lowest final fitness. Lineages NC41a and WA11a did not appear to increase in fitness; the fitness differences between the beginning and end points were -0.3 and 1.4 doublings per hour, respectively. NC41 had the highest initial fitness of all of the genotypes; WA11 had the third highest initial fitness (Table 1). This possibly contributed to their failure to achieve significant fitness improvement. Either these two phages were already at a local fitness optimum for our conditions or adaptive mutations of large to moderate effect are not accessible under these population dynamics. The average improvement in fitness over all 11 lineages was 5.8 doublings per hour, or a 15-fold increase in the number of progeny per phage per growth period.

Bull et al. (2004b) used a similar experimental design and found that only three of eight total genotypes showed large improvements in fitness. The two major differences between our experiments and those of Bull et al. were the particular strain of *E. coli* used as host and the diversity of starting genotypes. However, Bull et al. chose not to focus on questions of evolvability due to the history of their phages, which were all isolated decades ago and maintained in the lab since then. Lab culturing may have pre-adapted some or all of the genotypes to the selective conditions and different genotypes may have been pre-adapted to different extents. In contrast, our phage

genotypes were naïve to laboratory conditions, having never been cultured in the lab beyond that necessary for initial isolation (Rokyta et al. 2006b). Among these naïve genotypes, we found that at least 75% were evolvable, and on average the number of progeny per generation was more than doubled. In a similar study, Bollback and Huelsenbeck (2009) found that all seven of their experimental lineages, which involved three species of RNA bacteriophages, increased in fitness in response to selection for growth at high temperature.

The Number of Substitutions Underlying Adaptation

Adaptation to laboratory conditions required few nucleotide substitutions (Table 1). On average, a lineage reached a fitness plateau after fixing slightly more than three mutations. A single lineage, NC41a, did not fix any mutations, consistent with its failure to significantly increase in fitness. In addition to its final total population sequence, we sequenced four isolates from the final population of the NC41a lineage; all four isolate sequences were identical to the ancestral sequence, suggesting that no mutations had reached appreciable frequency in the population and that the phage NC41 had no capacity to adapt to these conditions in the amount of time allowed. On the other hand, the WA11a lineage fixed a single mutation despite not significantly improving fitness, suggesting that it was in fact evolvable, but the fitness effect of its

substitution was too small to be detected by our assay. The largest number of substitutions acquired during adaptation to a fitness plateau was nine for lineage ID2a. These numbers can only be definitively described as lower bounds on the number of substitutions in a complete adaptive walk. Alternatively, they can be viewed as the number of mutations required to reach high fitness or a fitness plateau.

It is generally not possible to write down an analytical formula for the expected length of an adaptive walk (Orr 2002; Rokyta et al. 2006a). However, Gillespie (1983, 1991) showed that assuming strong selection and weak mutation (see Gillespie 1991), the mean length of an adaptive walk should be

$$L = \frac{1}{2} + \frac{1}{i} + \frac{1}{2} \sum_{k=2}^{i-1} \frac{k+3}{k(k+1)} \quad (1)$$

where i is the fitness rank of the initial allele, implying there are $i - 1$ beneficial mutations available. This assumes, amongst other things (see Orr 2002), that all alleles are mutually accessible. Nonetheless, Orr (2002) found this formula to be a good approximation to simulation results under the more realistic mutational landscape model, which assumes a highly epistatic and rugged landscape. Excluding the NC41a lineage, which apparently had $i = 1$, the mean walk length for the remaining ten adaptations was 3.5 substitutions. By Eq. 1, $3 < L < 4$ corresponds to $8 < i < 20$. Interestingly, the related phage ID11 was found to have $i = 10$ under our selective conditions (Rokyta et al. 2005).

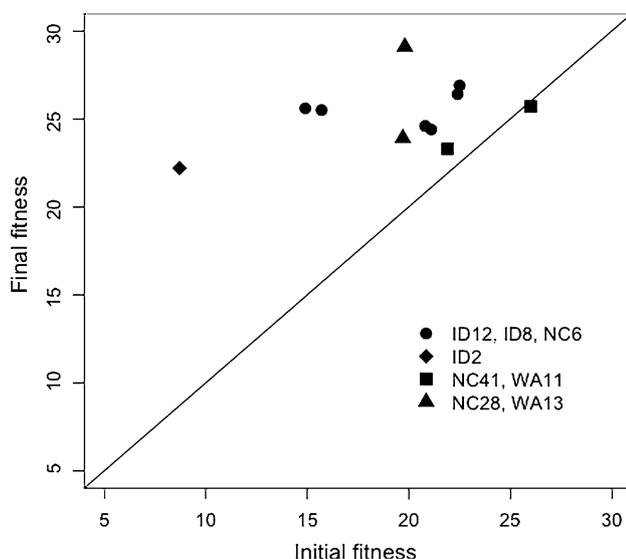


Fig. 2 Fitness improvement for all 11 experimental lineages. The diagonal line corresponds to a failure to increase fitness. The magnitude of fitness improvement is indicated by vertical displacement above the line. Phylogenetic groups (see Fig. 1) are designated with different symbols. Both initial and final fitnesses have standard errors less than 0.55 and are measured in population doublings per hour

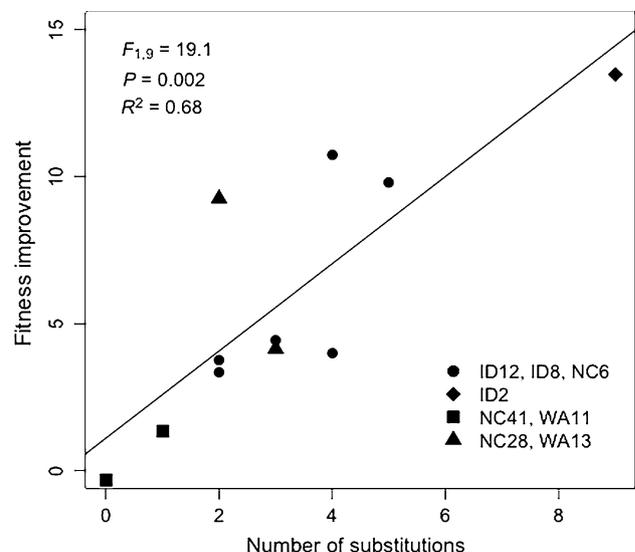


Fig. 3 Fitness improvement and the number of mutations fixed during adaptation. A significant positive correlation was found between the number of substitutions and the magnitude of fitness improvement for an adaptive walk ($R^2 = 0.68$, $F_{1,9} = 19.1$, $P = 0.002$). The regression line is shown. Phylogenetic groups (see Fig. 1) are designated with different symbols

Assuming a similar value of *i* for the eight phages used in the present study, our walk lengths are roughly consistent with the mutational landscape model and thus a highly epistatic and rugged fitness landscape.

Larger fitness gains required more substitutions (Fig. 3). There was a significant positive correlation between the number of observed substitutions and the magnitude of fitness improvement ($R^2 = 0.68$, $F_{1,9} = 19.1$, $P = 0.002$). The average effect over all 35 observed substitutions was a fitness increase of 1.8 doublings per hour. On average, each mutation approximately doubled the number of offspring produced in 40 min.

Parallel Evolution

Parallel evolution occurred at a high rate across the three pairs of adaptations beginning from identical genotypes (Table 2). For the ID12a and ID12b lineages, two of the four substitutions were parallel events. Six of nine substitutions in the ID8a and ID8b lineages were identical. NC6a and NC6b had two of seven total substitutions in common. Therefore, the percentage of parallel changes in replicate adaptations ranged from approximately 30% up to 67%. Parallel changes appear to be a common feature of viral adaptation (Bull et al. 1997; Wichman et al. 1999, 2000;

Table 2 The genetic details of 11 adaptive walks

Genotype	Clade	Genome size	Nucleotide position ^a	Δnuc	Amino acid position ^b	Δaa
ID12a60	G4	5529	1970	A → G	C84	N → S
			4919	G → A	H135	G → D
ID12b70	G4	5529	2101	C → T	D42	Silent
			4919	G → A	H135	G → D
ID8a80	G4	5540	1958	A → G	C80	N → S
			3628	C → T	F355	P → S
			4493	A → G	G171	T → A
			4770	C → T	H82	H → Y
			4930	G → A	H135	G → D
ID8b90	G4	5540	1958	A → G	C80	N → S
			3628	C → T	F355	P → S
			4493	A → G	G171	T → A
			4951	G → A	H142	G → D
NC6a70	G4	5529	2418	C → T	D148(E89)	A(L) → V(F)
			4168	A → G	G66	N → S
			5033	G → A	H173	G → E
			5071	G → A	H186	A → T
NC6b70	G4	5529	2418	C → T	D148(E89)	A(L) → V(F)
			4482	A → G	G171	T → A
			4919	G → A	H135	G → D
ID2a200	ID2	5486	858	A → G	A266	D → G
			1570	C → T	A503(B99)	I(R) → I(W)
			1783	C → T	C22(K49)	H(S) → Y(L)
			2087	C → T	D39	Silent
			3264	G → A	F239	R → H
			3932	T → C	G2	F → L
			4882	C → T	H135	S → F
			4902	G → A	H142	G → S
			5070	C → A	H198	Q → K
			NC41a30	φX174	5386	None
WA11a40	φX174	5387	3027	T → C	H33	F → L
NC28a80	WA13	6065	2903	G → T	D129(E69)	A(K) → S(N)
			3904	A → G	F272	Silent
			5116	T → C	H10	V → A
WA13a40	WA13	6068	2523	T → C	D2	S → P
			5289	C → T	H67	P → S

^a Nucleotide positions follow the published ancestral sequences

^b Amino acid positions consist of the gene name (A–H, J, K) and amino acid residue number. Amino acid positions given in parentheses indicate that the substitution affected a region encoding overlapping genes

Cuevas et al. 2002; Rokyta et al. 2005). Consistent with our results, Bollback and Huelsenbeck (2009) found that the mean frequency of parallel changes among replicate lineages of the RNA bacteriophages MS2 and NL95 were 47 and 42.6%, respectively. Orr (2005) showed that the probability of parallel evolution for a single step is given by $P = 2/i$, where again i is the fitness rank of the initial genotype. Thus, parallel evolution can occur at a high rate when the number of possible beneficial mutations is low, although there is currently no formula analogous to Orr's for multiple steps in adaptation.

In addition to parallel evolution across replicates, an appreciable number of substitutions occurred in parallel across genotypes (Table 2). Lineage NC6b actually shared more parallel substitutions with ID8a than with its own replicate lineage, NC6a. Half of the substitutions in these two lineages were parallel changes. Four different genotypes (lineages ID12a/b, ID8a, NC6b, and ID2a) had substitutions at residue 135 in gene H (H135), though the amino acid change for lineage ID2a was different than in the other lineages (S → F instead of G → D). The evolved amino acid residue in the other lineages, however, was inaccessible through a single mutation from the ancestral codon of ID2a. Even if it were beneficial to the ID2 genotype, it is highly unlikely for anything other than single-mutation neighbors to be explored under our experimental conditions (Gillespie 1984). Residue G171 had parallel changes in two genotypes, ID8a/b and NC6b. Two genotypes, ID8b and ID2a, both had substitutions at position H142, though the amino acid change was different. In this case, both amino acid changes were accessible through a single mutation from both ancestral genotypes. At the genome level, NC6 differs from ID12 and ID8 at approximately 7% of its nucleotide sites; ID2 differs from ID12, ID8, and NC6 at approximately 20% of its nucleotide sites. ID12 and ID8 differ at 1.5% of their sites. Thus, parallel evolution occurred across genotypes that differ by as much as 7% and identical sites responded to the same selective pressure across a sequence distance of 20%.

Parallel evolution across different genotypes has been noted before in the phages ϕ X174 and S13 (Wichman et al. 2000), though the genotypes involved only differed by 2.1% at the nucleotide level. Most reports of parallel evolution involve replicates of the same genotype, and while this tells us much about the constraints and limitations on adaptation for that genotype, parallel evolution across different genotypes tells us something about epistasis and hence the fitness landscape. For these mutations, there is no evidence of sign epistasis (Weinreich et al. 2005), i.e., whether these mutations are beneficial, deleterious or neutral does not appear to depend on the genetic background. This phenomenon is indicative of a smooth fitness landscape.

The Genetics of Adaptation

Nearly all substitutions observed in our adaptations were missense changes (Table 2). Of the 35 total substitutions, only three were synonymous in all coding frames. The phage ϕ X174 and the other microvirid phages have extensive overlap in their coding regions, such that ~20% of the genome encodes more than one protein. Therefore, more than 90% of the substitutions affected the amino acid sequence of their respective ancestral genotypes. Five substitutions fell within regions coding two gene products; four of these were nonsynonymous in both reading frames, and a single substitution was synonymous in one reading frame and nonsynonymous in the other.

The majority of substitutions occurred within the genes encoding the phage capsid proteins (F, G, H, and J). Fourteen substitutions affected protein H, the pilot protein, which is involved in binding host lipopolysaccharide (LPS) during adsorption (Suzuki et al. 1999; Inagaki et al. 2003) and enters the host with the phage genome (Hayashi et al. 1988). Five nonsynonymous substitutions affected protein G, the major spike protein, which is also involved in binding the host LPS (Inagaki et al. 2003). However, these mutations all lie at or near the physical interface between the G protein and the major capsid protein (F), based on the ϕ X174 and G4 structures (McKenna et al. 1992, 1994, 1996), suggesting an effect on something other than attachment. Four substitutions affected protein F, the major capsid protein. One was synonymous, and the remaining three affected two different sites, both of which are internal to the F protein. Mutations in gene F have been found to alter host range (Crill et al. 2000) and have also been shown to be important in temperature range and capsid stability (Bull et al. 2000); our results are more consistent with the latter.

The remainders of the observed substitutions were in non-capsid genes (Table 2). Four substitutions affected protein C, which is involved in initiating the switch from double-stranded phage DNA replication to the production of single-stranded phage genomes (Hayashi et al. 1988). Six substitutions affected the nucleotide sequence of the external scaffolding protein, gene D. Two of these were silent, and three were also nonsynonymous changes in gene E, which codes for the lysis protein. The ID2a lineage had two substitutions affecting protein A, which is involved in phage DNA replication (Hayashi et al. 1988). One of these two substitutions was also nonsynonymous in gene B, which encodes the internal scaffolding protein (Hayashi et al. 1988).

Fitness Limits

Initial fitness in our selective environment did not predict the maximum fitness attained in that environment. We

found no significant correlation between initial fitness and final fitness ($R^2 = 0.16$, $F_{1,9} = 1.71$, $P = 0.22$). Thus, the evolutionary history of a genotype, in terms of initial fitness, does not predetermine its ability to adapt to a new environment. A similar result was described by Travisano et al. (1995) for *E. coli*; they found little effect of population history on the ultimate outcome of adaptation. However, as noted above, larger fitness gains do tend to require more substitutions, and thus lower fitness genotypes may still be at a disadvantage in terms of the rate of adaptation.

Three genotypes (ID12, ID8, and NC6) were adapted in replicate, and the maximum fitnesses attained in these replicates were not significantly different. A one-way ANOVA showed that while the final fitnesses differed significantly among the three replicated genotypes ($F_{2,28} = 17.09$, $P \ll 0.01$), there was no significant effect of replicates within genotypes ($F_{3,28} = 0.32$, $P = 0.81$). Despite adapting through different substitutions, the same fitnesses were ultimately reached. Thus, at least for these three phages, differences in the underlying genetics of adaptation did not constrain the evolutionary outcome. A similar pattern holds for initial fitness values: there were significant differences among the genotypes ($F_{2,28} = 418.40$, $P \ll 0.01$), but not between replicates ($F_{3,28} = 1.85$, $P = 0.16$). For simplicity, the “b” replicates will be excluded from the remaining analyses.

We performed a hierarchical mixed-effects ANOVA on the final fitness values, adjusting for unbalanced sampling structure, to assess the differences between phylogenetic groups (Fig. 1) and evaluate the significance of variation due to genotypes within groups. The four groups were treated as fixed effects and genotypes within groups were treated as random effects. We found no significant differences in the mean fitnesses of the different phylogenetic groups ($F_{3,4.01} = 1.06$, $P = 0.46$), but genotypes contribute significantly to the within group variation ($F_{4,36} = 37.91$, $P \ll 0.01$). An analysis of the residuals did not show any significant deviation from normality or equality of variance. The above analysis treats the phage ID2 as its own phylogenetic group. Excluding this phage had no significant effect on the analysis. In terms of uncorrected nucleotide sequence distance between groups, the largest average genomic pairwise distance is about 0.40 (ID2 or the G4-like phages versus the WA13-like phages) and the smallest is approximately 0.20 (ID2 versus the G4-like phages).

In a related study, Bull et al. (2004b) adapted eight bacteriophages to conditions similar to ours. Their phages also fell into four distinct groups, except that these groups differed drastically in genome composition and size. They found evidence for both global and local constraints on adaptation. Global constraints consist of biological limitations of the phages, whereas local constraints consist of

variational limitations. Our phages all fall within one of the groups considered by Bull et al., and thus our study is on a much finer scale. Still, we know that phylogenetically distinct microvirid phages differ, for example, in the host factors used in replication (Baas 1985), and thus the potential exists for global constraints on achievable fitness. However, this was not observed. The variation in final fitness seems to be due solely to local constraints, i.e., to differences in the accessibility of high fitness genotypes.

Conclusions

Through the observation of 11 adaptive walks in microvirid bacteriophages, we have described some general properties of adaptation which should provide guidance for the development of multiple-step theories of adaptive molecular evolution to complement previous single-step theories (e.g., Orr 2002; Joyce et al. 2008). We have found that for our phages, most (75%) naïve genotypes are able to adapt to new conditions and that this adaptation occurs quickly (~ 90 –600 generations) and requires only a few substitutions. Parallel evolution is rampant, accounting for approximately 1/3 to 2/3 of substitutions across replicates, and occurs even across genotypes that differ by as much as 7% at the nucleotide level. Furthermore, the same sites responded to selection in genotypes differing by $\sim 20\%$. We found no evidence of historical constraints, in terms of starting fitness, on fitness limits, but because greater fitness improvements take more substitutions, there is potentially a constraint on the rate of adaptation. Replicate adaptations always proceeded through different genetic pathways, but achieved the same final fitnesses. Finally, we found no evidence for global constraints on adaptation across the level of diversity considered in the present study, only local constraints.

Acknowledgments The authors would like to thank S. B. Caudle for laboratory assistance and G. W. Daughdrill and J. J. Bull for discussions. This work was supported by NIH Grant Number P20 RR16454 from the BRIN Program of the National Center for Research Resources and NCRR COBRE Grant P20RR16448.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abedon ST (1989) Selection for bacteriophage latent period length by bacterial density: a theoretical examination. *Microb Ecol* 18:79–88
- Baas PD (1985) DNA replication of single-stranded *Escherichia coli* DNA phages. *Biochim Biophys Acta* 825:111–139

- Beisel CJ, Rokyta DR, Wichman HA, Joyce P (2007) Testing the extreme value domain of attraction for distributions of beneficial fitness effects. *Genetics* 176:2441–2449
- Betancourt AJ (2009) Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics* 181:1535–1544
- Betancourt AJ, Bollback JP (2006) Fitness effects of beneficial mutations: the mutational landscape model in experimental evolution in a virus. *Curr Opin Gen Dev* 16:618–623
- Bollback JP, Huelsenbeck JP (2009) Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics* 181:225–234
- Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ (1997) Exceptional convergent evolution in a virus. *Genetics* 147:1497–1507
- Bull JJ, Badgett MR, Wichman HA (2000) Big-benefit mutations in a bacteriophage inhibited with heat. *Mol Biol Evol* 17:942–950
- Bull JJ, Badgett MR, Springman R, Molineux IJ (2004a) Genome properties and the limits of adaptation in bacteriophages. *Evolution* 58:692–701
- Bull JJ, Pfennig DW, Wang IN (2004b) Genetic details, optimization and phage life histories. *Trends Ecol Evol* 19:76–82
- Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736–739
- Crill WD, Wichman HA, Bull JJ (2000) Evolutionary reversals during viral adaptation to alternating hosts. *Genetics* 154:27–37
- Cuevas JM, Elena SF, Moya A (2002) Molecular basis of adaptive convergence in experimental populations of RNA viruses. *Genetics* 162:533–542
- de Visser JAGM, Lenski RE (2002) Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol Biol* 2:19
- Elena SF, Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Gen* 4:457–469
- Fisher RA (1930) *The genetical theory of natural selection*. Oxford University Press, Oxford, UK
- Gillespie JH (1983) A simple stochastic gene substitution model. *Theor Popul Biol* 23:202–215
- Gillespie JH (1984) Molecular evolution over the mutational landscape. *Evolution* 38:1116–1129
- Gillespie JH (1991) *The causes of molecular evolution*. Oxford University Press, New York
- Hayashi M, Aoyama A, Richardson DL, Hayashi MN (1988) Biology of the bacteriophage ϕ X174. In: Calendar R (ed) *The bacteriophages*, vol 2. Plenum Press, New York, NY, pp 1–71
- Huelsenbeck JP, Rannala B (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol* 53:904–913
- Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755
- Inagaki M, Kawaura T, Wakashima H, Kato M, Nishikawa S, Kashimura N (2003) Different contributions of the outer and inner R-core residues of lipopolysaccharide to the recognition by spike H and G proteins of bacteriophage ϕ X174. *FEMS Microbiol Lett* 226:221–227
- Joyce P, Rokyta DR, Beisel CJ, Orr HA (2008) A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. *Genetics* 180:1627–1643
- Kassen R, Bataillon T (2006) Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature Genet* 38:484–488
- Kauffman SA (1993) *The origins of order*. Oxford University Press, New York
- Kauffman S, Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* 128:11–45
- Kauffman SA, Weinberger ED (1989) The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J Theor Biol* 141:211–245
- Kichler Holder K, Bull JJ (2001) Profiles of adaptation in two similar viruses. *Genetics* 159:1393–1404
- Kirschner M, Gerhart J (1998) Evolvability. *Proc Natl Acad Sci USA* 95:8420–8427
- Lenski RE, Travisano M (1994) Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci USA* 91:6808–6814
- Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* 138:1315–1341
- Macken CA, Perelson AS (1989) Protein evolution on rugged landscapes. *Proc Natl Acad Sci USA* 86:6191–6195
- Macken CA, Hagan PS, Perelson AS (1991) Evolutionary walks on rugged landscapes. *SIAM J Appl Math* 51:799–827
- MacLean RC, Buckling A (2009) The distribution of fitness effects of beneficial mutations in *Pseudomonas aeruginosa*. *PLoS Genetics* 5:e1000406
- Maynard Smith J (1970) Natural selection and the concept of a protein space. *Nature* 225:563–564
- McBride RC, Ogbunugafor CB, Turner PE (2008) Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evol Biol* 8:231
- McKenna R, Xia D, Willingmann P, Ilag LL, Krishnaswamy S, Rossmann MG, Olson NH, Baker TS, Incardona NL (1992) Atomic structure of single-stranded DNA bacteriophage ϕ X174 and its functional implications. *Nature* 355:137–14
- McKenna R, Ilag LL, Rossmann MG (1994) Analysis of the single-stranded DNA bacteriophage ϕ X174, refined at a resolution of 3.0 Å. *J Mol Biol* 237:517–543
- McKenna R, Bowman BR, Ilag LL, Rossmann MG, Fane BA (1996) Atomic structure of the degraded procapsid particle of the bacteriophage G4: induced structural changes in the presence of calcium ions and functional implications. *J Mol Biol* 256:736–775
- Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny. *Syst Biol* 52:1–10
- Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949
- Orr HA (2000) Adaptation and the cost of complexity. *Evolution* 54:13–20
- Orr HA (2002) The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56:1317–1330
- Orr HA (2003) The distribution of fitness effects among beneficial mutations. *Genetics* 163:1519–1526
- Orr HA (2005) The probability of parallel evolution. *Evolution* 59:216–220
- Orr HA (2006a) The distribution of beneficial fitness effects among beneficial mutations in Fisher's geometric model of adaptation. *J Theor Biol* 238:279–285
- Orr HA (2006b) The population genetics of adaptation on correlated fitness landscapes: the block model. *Evolution* 60:1113–1124
- Pepin KM, Wichman HA (2008) Experimental evolution and genome sequencing reveal variation in levels of clonal interference in large populations of bacteriophage ϕ X174. *BMC Evol Biol* 8:85
- Perelson AS, Macken CA (1995) Protein evolution on partially correlated landscapes. *Proc Natl Acad Sci USA* 92:9657–9661
- Perfeito L, Fernandes L, Mota C, Gordo I (2007) Adaptive mutations in bacteria: high rate and small effects. *Science* 317:813–815
- R Development Core Team (2006) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>

- Rokyta D, Badgett MR, Molineux IJ, Bull JJ (2002) Experimental genomic evolution: extensive compensation for loss of DNA ligase activity in a virus. *Mol Biol Evol* 19:230–238
- Rokyta DR, Joyce P, Caudle SB, Wichman HA (2005) An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nature Genet* 37:441–444
- Rokyta DR, Beisel CJ, Joyce P (2006a) Properties of adaptive walks on uncorrelated landscapes under strong selection and weak mutation. *J Theor Biol* 243:114–120
- Rokyta DR, Burch CL, Caudle SB, Wichman HA (2006b) Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol* 188:1134–1142
- Rokyta DR, Beisel CJ, Joyce P, Ferris MT, Burch CL, Wichman HA (2008) Beneficial fitness effects are not exponential for two viruses. *J Mol Evol* 67:368–376
- Sanjuán R, Moya A, Elena SF (2004) The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci USA* 101:8396–8401
- Stadler PF, Happel R (1999) Random field models for fitness landscapes. *J Math Biol* 38:435–478
- Suzuki R, Inagaki M, Karita S, Kawaura T, Kato M, Nishikawa S, Kashimura N, Morita J (1999) Specific interaction of fused H protein of bacteriophage ϕ X174 with receptor lipopolysaccharides. *Virus Res* 60:95–99
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nuc Acids Res* 22:4673–4680
- Travisano M, Mongold JA, Bennett AF, Lenski RE (1995) Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* 267:87–90
- Wagner GP, Altenberg L (1996) Complex adaptations and the evolution of evolvability. *Evolution* 50:967–976
- Wang IN, Dykhuizen DE, Slobodkin LB (1996) The evolution of phage lysis timing. *Evol Ecol* 10:545–558
- Weinreich DM, Watson RA, Chao L (2005) Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–1174
- Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ (1999) Different trajectories of parallel evolution during viral adaptation. *Science* 285:422–424
- Wichman HA, Scott LA, Yarber CD, Bull JJ (2000) Experimental evolution recapitulates natural evolution. *Philos Trans R Soc London B* 355:1677–1684
- Wichman HA, Millstein J, Bull JJ (2005) Adaptive molecular evolution for 13,000 phage generations: a possible arms race. *Genetics* 170:19–31