Environment Determines Epistatic Patterns for a ssDNA Virus

S. Brian Caudle,* Craig R. Miller,[†] and Darin R. Rokyta*¹

*Department of Biological Science, Florida State University, Tallahassee, Florida 32306, and [†]Department of Biological Sciences, University of Idaho, Moscow, Idaho 83843

ABSTRACT Despite the accumulation of substantial quantities of information about epistatic interactions among both deleterious and beneficial mutations in a wide array of experimental systems, neither consistent patterns nor causal explanations for these interactions have yet emerged. Furthermore, the effects of mutations depend on the environment in which they are characterized, implying that the environment may also influence epistatic interactions. Recent work with beneficial mutations for the single-stranded DNA bacteriophage ID11 demonstrated that interactions between pairs of mutations could be understood by means of a simple model that assumes that mutations have additive phenotypic effects and that epistasis arises through a nonlinear phenotype–fitness map with a single intermediate optimum. To determine whether such a model could also explain changes in epistatic patterns associated with changes in environment, we measured epistatic interactions for these same mutations under conditions for which we expected to find the wild-type ID11 at different distances from its phenotypic optimum by assaying fitnesses at three different temperatures: 33°, 37°, and 41°. Epistasis was present and negative under all conditions, but became more pronounced as temperature increased. We found that the additive-phenotypes model explained these patterns as changes in the parameters of the phenotype–fitness map, but that a model that additionally allows the phenotypes to vary across temperatures performed significantly better. Our results show that ostensibly complex patterns of fitness effects and epistasis across environments can be explained by assuming a simple structure for the genotype–phenotype relationship.

THE fitness effects of mutations have consistently shown a dependence on the genetic background into which they arise. This dependence, or epistasis, has ramifications throughout evolutionary theory, affecting such fundamental topics as speciation (Dobzhansky 1936; Muller 1939; Coyne and Orr 2004), adaptation (Weinreich *et al.* 2005), the origin and maintenance of sexual reproduction (Kondrashov 1988, 1993; Otto and Feldman 1997; Otto 2009), and even disease emergence (Lalic and Elena 2013). The relevance of epistasis is determined not just by its presence, but also by the particular forms it takes. For example, reciprocal sign epistasis, such as when two mutations that are individually deleterious are beneficial in combination, leads to rugged, multi-peaked fitness landscapes (Poelwijk *et al.* 2011), which can reduce the efficacy of natural selection. Sexual reproduction can be favored over asexual reproduction when deleterious mutations

show synergistic epistasis. In this case, the effects of mutations are more deleterious in combination than alone, and sexually reproducing populations can more easily purge these mutations (Bonhoeffer *et al.* 2004). Studies from numerous systems have shown that the patterns of epistasis are highly variable, and no consistent patterns have emerged (de Visser *et al.* 2011). The reasons for this lack of consistency are so far unclear.

As important as epistasis is to evolutionary theory, its actual underlying causes and molecular mechanisms remain poorly explained (Lehner 2011). Epistasis has been modeled at the DNA-sequence level by assuming that particular sites in a sequence depend on the states at other positions (Kauffman and Levin 1987; Kauffman 1993; Perelson and Macken 1995; Orr 2006; Rokyta *et al.* 2006a). The difficulty with relating such models to empirical systems lies in determining the interdependence among sites, although protein–protein interactions might be a good starting point (Rokyta and Wichman 2009; Lehner 2011). One mechanistic approach to modeling and understanding epistasis is based on metabolic flux theory (Kacser and Burns 1981; Dean *et al.* 1986; Szathmáry 1993). Under such models, the relationship between enzyme function

and flux is concave, such that large changes in enzyme activity can provide only small changes in flux through the system. Such models may be appropriate when selection is acting to increase enzyme function and therefore flux, but much selection on proteins may involve other properties such as stability (DePristo *et al.* 2005), which may be under stabilizing selection instead of directional selection. The presence of sign epistasis (Weinreich *et al.* 2005) within the interactions among mutations indicates something more complex than the diminishing-returns relationship between phenotype and fitness predicted by flux theory, which allows for changes only in the magnitudes of effects. Epistasis also arises in models assuming stabilizing selection around a phenotypic optimum such as Fisher's geometric model (Fisher 1930); such models naturally generate sign and magnitude epistasis and have been found to explain some patterns of epistasis for beneficial mutations (Martin *et al.* 2007; Rokyta *et al.* 2011; Pearson *et al.* 2012).

The environment determines fitness effects of mutations (Remold and Lenski 2004; Lalic *et al.* 2011; Vale *et al.* 2012; Flynn *et al.* 2013) and therefore also affects epistatic patterns (You and Yin 2002; Lehner 2011; Flynn *et al.* 2013; Lalic and Elena 2013; Wang *et al.* 2013). Rokyta *et al.* (2011) used a model based on Fisher's geometric model that assumed additive phenotypic effects for mutations with an intermediate phenotypic optimum to explain the complex epistatic patterns among pairs of individually beneficial mutations for the single-stranded (ss)DNA bacteriophage ID11. The benefits of mutations in combination were always less than predicted assuming additivity for this phage and its mutants, and sign epistasis was common. The model explained these patterns by positing that most single mutations were near or in excess of the optimum and adding the second mutation therefore resulted in a movement away from the optimum. In the present work, we attempted to shift the wild-type genotype, ID11, and its nine previously described beneficial mutations (Rokyta *et al.* 2005, 2008, 2011) closer to and farther from the optimal phenotype by changing the temperature at which fitness was measured, to quantify the changes in fitness effects and patterns of epistatic interactions between mutations. Previous work has shown that close relatives of ID11 in the G4-like group of microvirid coliphages (Rokyta *et al.* 2006b) tend to have fitnesses inversely correlated with temperature (Knies *et al.* 2009). The original mutations were isolated by means of selection at 37°, and we therefore attempted to move the genotypes closer to the optimum by measuring fitnesses at a lower temperature (33°) and farther from the optimum by measuring fitnesses at a higher temperature (41°), leaving the genotypes involved constant. We measured fitness effects and epistatic interactions between mutations at these three temperatures and applied the phenotype–fitness modeling approach of Rokyta *et al.* (2011) and Pearson *et al.* (2012) to determine whether changing patterns of fitness effects and epistatic interactions among environments could be explained by changes in the phenotype–fitness map.

Materials and Methods

Constructing the mutants and fitness assays

The phage isolates used in the present work were the same as those constructed and used by Rokyta *et al.* (2011). The isolation and initial characterization of the nine beneficial mutations for our wild-type ID11 at 37° were described previously (Rokyta *et al.* 2005, 2008). We sequenced the entire genome of each isolate by means of Sanger sequencing to confirm the presence of the mutation(s) of interest and the absence of additional mutations. We measured fitness as the log₂ increase in the phage population per hour on *Escherichia coli* strain C. Briefly, we added ~10⁴–10⁵ phage to a culture of ~10⁸ host cells, allowed growth for 40 min, and then terminated growth by the addition of CHCl₃. Beginning and ending phage concentrations were determined by standard plating assays. For each genotype (wild-type ID11, 9 single mutants, and 18 double mutants), fitness was measured at least five times at each of three temperatures (33°, 37°, and 41°). All statistical analyses and modeling (see below) were conducted with R (R Development Core Team 2006).

Testing for additivity of phenotypic effects

The analysis of this data set under the gamma model was conducted analogously to previous work (Rokyta *et al.* 2011; Pearson *et al.* 2012) with slight modifications for the three-temperature experimental regime. The gamma model assumes a gamma-shaped relationship between phenotype (x) and expected fitness (g),

$$g(x, \alpha, \beta, A, B) = A \left(\frac{x}{\beta} \right)^{\alpha-1} e^{-x/\beta} + B, \quad (1)$$

where α is the shape parameter, β is the scale parameter, A controls the height of the curve, and B shifts the entire curve in the fitness dimension. The model assumes that one phenotype explains fitness and fitness effects. Deviations between observed and expected fitness are assumed to be normally distributed with mean zero and variances of $\sigma_{\text{singles}}^2$ and $\sigma_{\text{doubles}}^2$ for singles and doubles, respectively. Separate variances were used for two reasons: (1) the fitting algorithm treats singles and doubles differently so that singles tend to fall closer to the gamma function (using one variance could allow a highly deviant double to drive parameter estimation) and (2) we expected more variance among doubles because their fitnesses are affected by two mutations. The model assumes that mutations shift phenotypes in an additive fashion. If x_0 is the phenotype of the wild type and Δx_i and Δx_j are the phenotypic effects of mutations i and j , then the expected phenotype of the double mutant is just $x_{ij} = x_0 + \Delta x_i + \Delta x_j$. The values of x_0 , Δx_i , and Δx_j therefore predict x_{ij} , and x_{ij} predicts the fitness of double mutant ij by means of Equation 1. Because we imputed phenotypes and their scale is arbitrary (phenotypes x and the scale β always appear as a ratio in the gamma function), we reduced the dimensionality of the problem by setting $\beta = 100$ throughout our analyses. All

phenotypes and phenotypic effects were constrained to have positive values.

We compared three models of increasing complexity to explain our data: a null model and two nested manifestations of the gamma model. In the null model, the fitness of each genotype at each temperature was assumed to come from a normal distribution with a mean μ_t for that temperature and variance σ^2 . In the simpler, one-phenotype gamma model (G1), each genotype (wild type, singles, and doubles) had a single fixed phenotype across all three temperatures. Each temperature was assumed to have a distinct phenotype–fitness map on the same phenotypes. In G1, three sets of gamma parameters and one set of phenotypes (wild type plus the single mutants) were therefore estimated. In the more complex three-phenotype gamma model (G3), not only did temperature affect phenotype function (each temperature has its own gamma function), but also the phenotypes themselves were also free to vary at each temperature. In G3, three sets of gamma parameters and three sets of phenotypes were estimated. To fit the two models, we used an iterative least-squares approach described previously (Rokyta *et al.* 2011; Pearson *et al.* 2012). For the G1 model, we began with an initial set of gamma parameters that we applied to all three temperatures. Given these three (initially identical) functions, the phenotype step of the algorithm identified the set of phenotypes that minimized the squared deviations of observed fitness values from expected fitness values. Deviations were summed over wild type, single mutants, and double mutants. Having obtained phenotypes, the gamma step of the algorithm fixed these phenotypes and found the set of gamma parameters that minimized the squared deviations. In this step, we optimized the gamma parameters at each temperature independently. We continued to iterate between the phenotype step and the gamma step until the algorithm either arrived at a fixed set of parameter values (phenotypes and gamma parameters) or began cycling through the same set of values. For the G3 model we used the same approach, except that the phenotypes were free to vary between temperatures. Initiating these algorithms at different gamma parameters does not always yield the same parameter estimates or the same level of fit; the likelihood surface has many local peaks. We addressed this problem by initiating the algorithm at 25 different gamma curves that represent a wide range of shapes and heights (we used the same set as that in Pearson *et al.* 2012). Point estimates were obtained by taking the best fit across the set of 25 starting points.

For model selection, we used a likelihood-ratio framework. We first tested whether the simpler G1 model was significantly better than an intercept-only null model and, subsequently, whether G3 was better than G1. To perform the first test, we calculated the log-likelihood of the data under the null and under G1 and took the differences: $\Lambda_{G1} = \ln L_{G1} - \ln L_{\text{null}}$. We then obtained the distribution of the Λ_{G1} statistic by bootstrap. Specifically, we randomized the data sets consistent with the null model where, within each temperature, single-mutant fitnesses were randomized and then double-mutant

fitnesses were independently randomized. Randomization excluded wild type. This randomization breaks all associations between the single mutants and the double mutants consistent with the null model's assumption that such associations are not meaningful. We chose to randomize rather than to simulate new data under the null model to make the analysis robust to violations of the normality assumptions. We generated 100 such randomized data sets, fitted each to each gamma model, and calculated Λ_{G1}^* . The fractions of exceedances, $\Lambda_{G1}^* > \Lambda_{G1}$, provided the approximate *P*-values associated with G1. To determine whether G3 was a significant improvement over G1, we used a likelihood-ratio test by redefining the simpler G1 model as the null. We first calculated $\Lambda_{G3-G1} = \ln L_{G3} - \ln L_{G1}$ for the real data. We then simulated 100 data sets under the null model. To do this, we used the three estimated gamma functions and the imputed phenotypes to determine the expected fitnesses of all genotypes. From these, we simulated normal deviates for all single- and double-mutant genotypes with mean zero and variances as estimated, $\hat{\sigma}_{\text{singles}}^2$ and $\hat{\sigma}_{\text{doubles}}^2$. Each simulated data set was fitted to each gamma model, and Λ_{G3-G1}^* was calculated. The *P*-value associated with support for the more complex G3 model was approximated as the fraction of instances where $\Lambda_{G3-G1}^* > \Lambda_{G3-G1}$. Note that we used parametric bootstrap here rather than randomization (as before) because it is not possible to randomize the data consistent with the G1 model, but it is possible to simulate under it. Although 100 replicates could result in low precision of our *P*-value estimates, we were limited by the extreme computational burden associated with fitting the gamma models.

R^2 values were calculated as $1 - \text{SS}_{\text{Res}}/\text{SS}_{\text{Total}}$, where SS_{Total} and SS_{Res} are the total and residual sums of squares. SS_{Total} was calculated by summing the squared deviations of each genotype's fitness at a given temperature from the mean fitness at that temperature followed by summing over the three temperatures. SS_{Res} was calculated as the sum of squared deviations between the observed fitness values and those predicted under the model. Log-likelihoods under the gamma models were calculated by taking the deviations of observed from predicted fitness under the model for each genotype at each temperature, determining their densities in the normal distribution, taking logarithms, and summing. The normal distribution was assumed to have mean zero and variance estimated from the deviates themselves. The log-likelihood under the null (intercept-only) model was calculated by taking each genotype at each temperature, determining the deviation between its observed fitness and the mean fitness at that temperature, finding the probability density of this value in the normal, taking the logarithm, and summing over all data points. The variance of the normal was estimated from all deviates (pooled across temperatures and for singles and doubles). Note that our results do not depend strongly on subtle issues in how the likelihoods were calculated such as, for example, the use of separate variances for the single and double mutants in the gamma model. Significance was evaluated by means of bootstrap techniques in which the exact

Table 1 Nine mutations beneficial to the ssDNA bacteriophage ID11

Label	Protein function	Protein name	Aa position	Δ Aa	Nuc position	Δ Nuc
A	DNA binding	J	15	A→V	2520	C→T
B	DNA binding	J	20	V→L	2534	G→T
C	Coat	F	3	V→F	2609	G→T
D	Coat	F	314	A→V	3543	C→T
E	Coat	F	322	N→S	3567	A→G
F	Coat	F	355	P→S	3665	C→T
G	Coat	F	416	M→I	3850	G→A
H	Coat	F	419	T→A	3857	A→G
I	Coat	F	421	D→G	3864	A→G

The nine beneficial mutations used in this study affect two different viral proteins: the DNA-binding protein J and the major coat protein F. Positions are based on the published genome sequence of ID11 (GenBank accession no. AY751298). Nuc, nucleotide; Δ Nuc, nucleotide change; Aa, amino acid; Δ Aa, amino acid change.

same likelihood calculations were performed on the real and randomized or simulated data sets.

Results and Discussion

Mutational effects increase with temperature

Our main objective was to determine how epistatic interactions between pairs of mutations (*i.e.*, genotype-by-genotype interactions) are affected by changes in the environment (*i.e.*, genotype-by-genotype-by-environment interactions). We must first set our null expectations in the absence of epistasis by determining how the individual fitness effects of mutations depend on the environment (*i.e.*, genotype-by-environment interactions or plasticity). The nine mutations (Table 1) were identified for and by means of their beneficial effects for the bacteriophage ID11 at 37° (Rokyta *et al.* 2005, 2008). Knies *et al.* (2009) showed that the optimal temperatures for ID11 and its close relatives were <37°. We therefore expected that our wild-type ID11 would be closer to its optimal temperature at a lower temperature (33°) and farther from its optimum at a higher temperature (41°) compared to 37°. Previous work showing diminishing-returns epistasis for beneficial mutations (Bull *et al.* 2000; MacLean *et al.* 2010; Chou *et al.* 2011; Khan *et al.* 2011; Pearson *et al.* 2012; Sousa *et al.* 2012) suggests that effect sizes will decrease for genotypes closer to the optimum, assuming that fitnesses are comparable across temperatures. We therefore expected fitness effects of the single mutations to increase with temperature as the fitness of the wild type decreased.

The fitness of the wild-type ID11 decreased as the temperature increased from 33° to 41° from 20.05 to -4.08 doublings per hour (Table 2, Table 3, and Table 4), consistent with movement away from the optimum and the results of Knies *et al.* (2009). The mean effects of the nine single mutations increased with increasing temperature (Figure 1 and Figure 2A). Overall, the mutations were beneficial as expected under the conditions in which they were identified, but deleterious at a lower temperature. At a higher temperature, the mutations were more beneficial than under the original isolation conditions. The mean effect at 33° was -1.26 ± 0.55 doublings per hour (plus or minus the standard error). The mean effect

at 37° (the original temperature at which the mutations were isolated) was 3.49 ± 0.45 doublings per hour, and the mean effect at 41° was 9.37 ± 0.96 doublings per hour. All three possible comparisons showed significant differences ($P < 0.001$ on the basis of two-sided, Welch two-sample *t*-tests with a Bonferroni correction for three tests). Diminishing-returns epistasis appears to be common in bacterial (MacLean *et al.* 2010; Chou *et al.* 2011; Khan *et al.* 2011; Sousa *et al.* 2012; Flynn *et al.* 2013) and viral (Bull *et al.* 2000; Pearson *et al.* 2012) systems. This pattern has typically been observed by changing the background fitness by means of modifying the genetic background, while here we reduced fitness by altering the environment. Our complementary result therefore generalizes the pattern to include environmentally induced fitness changes (similar to the results of Chou *et al.* 2009).

Note that the fitness of the wild-type genotype at 41° was -4.08 ± 0.61 doublings per hour (Table 4) and that this negative fitness value sets the magnitude of fitness effects for this temperature. While absolute fitness can never be less than zero (*i.e.*, an organism cannot produce negative offspring), growth rates or Malthusian fitnesses such as our measurements can take on values less than zero in a meaningful manner. A growth rate of zero implies that the population is remaining at a constant size, and a negative value indicates that the population is declining. In the context of phage growth, a value of zero could indicate a failure to initiate infection and stability in the environment or infection with just enough reproduction to maintain a constant population size. Our observed negative value could likewise indicate that the wild type is simply degrading in the environment or initiating infection with little or no production of progeny. Either way, the measured value accurately reflects the change in population size of the wild type and the estimated fitness effects for mutations relative to the wild type and therefore accurately reflects the selective advantages they confer.

A number of studies have examined genotype-by-environment interactions for individual mutations in microbial or viral systems. For the ssDNA phage ϕ X174, which is a close relative of our wild-type phage ID11 (Rokyta *et al.* 2006b), Vale *et al.* (2012) found that, for 36 random mutations, fitness effects

Table 2 Fitnesses and fitness effects of all genotypes tested at 33°

Genotype	Fitness	<i>n</i>	ΔW_{wt}	ΔW_{add}	ΔW_1	ΔW_2	ε_{ij}
ID11	20.05 ± 0.45	13	—	—	—	—	—
A	21.41 ± 0.34	5	1.36	—	—	—	—
B	19.05 ± 0.35	5	-1.00	—	—	—	—
C	18.20 ± 0.30	5	-1.85	—	—	—	—
D	19.67 ± 0.44	5	-0.38	—	—	—	—
E	17.59 ± 0.37	5	-2.46	—	—	—	—
F	19.16 ± 0.41	6	-0.90	—	—	—	—
G	16.65 ± 0.43	7	-3.40	—	—	—	—
H	20.61 ± 0.25	6	0.56	—	—	—	—
I	16.76 ± 0.39	5	-3.29	—	—	—	—
AB	17.62 ± 0.30	5	-2.43	0.35	-1.43	-3.79	-2.79
AD	17.17 ± 0.45	6	-2.88	0.97	-2.50	-4.24	-3.86
AF	17.81 ± 0.31	5	-2.25	0.46	-1.35	-3.61	-2.71
AG	17.93 ± 0.58	5	-2.12	-2.05	1.28	-3.48	-0.08
AH	20.39 ± 0.29	6	0.34	1.92	-0.22	-1.02	-1.58
BD	16.46 ± 0.14	5	-3.60	-1.39	-3.22	-2.60	-2.21
BE	16.29 ± 0.38	5	-3.77	-3.47	-1.30	-2.77	-0.30
BG	15.83 ± 0.56	6	-4.23	-4.41	-0.82	-3.23	0.18
BH	19.74 ± 0.30	5	-0.32	-0.45	-0.88	0.68	0.13
BI	15.14 ± 0.17	5	-4.91	-4.29	-1.62	-3.91	-0.62
CD	15.81 ± 0.25	5	-4.24	-2.23	-3.86	-2.39	-2.01
CE	13.39 ± 0.29	5	-6.66	-4.31	-4.20	-4.81	-2.35
CH	18.65 ± 0.15	5	-1.40	-1.29	-1.96	0.45	-0.11
DF	18.12 ± 0.49	7	-1.93	-1.28	-1.03	-1.55	-0.65
DH	18.62 ± 0.49	7	-1.43	0.17	-1.99	-1.05	-1.60
DI	14.39 ± 0.27	5	-5.66	-3.67	-2.37	-5.28	-1.99
EH	16.92 ± 0.24	5	-3.13	-1.90	-3.69	-0.67	-1.23
EI	15.67 ± 0.54	6	-4.38	-5.75	-1.09	-1.92	1.37

Fitnesses are given as the average plus or minus the standard error. *n* gives the number of replicate assays for each genotype. The fitness effect relative to the wild type is designated by ΔW_{wt} . ΔW_{add} gives the fitness effect expected on the assumption that the effects of single mutations were additive, ΔW_1 gives the effect of adding the first mutation in the genotype name into the background of the second, and ΔW_2 gives the effect of adding the second mutation in the genotype name into the background of the first. The deviation from additivity, ε_{ij} for the combination of mutations *i* and *j*, was calculated according to Equation 2.

were different but positively correlated when their effects were measured on two closely related hosts. Remold and Lenski (2004) showed that the fitness effects of 18 random insertions varied across two resource environments in *E. coli*, and Lalíć *et al.* (2011) showed that viral host species affected the fitness effects of 20 mutations for tobacco etch potyvirus (TEV). In all of these cases, the vast majority of the mutations considered were deleterious or neutral under most conditions. Similar to these previous studies, we found that the environment had a significant effect on the fitness effects of mutations, but our work was unique in several respects. Although we were dealing with a much smaller set of mutations, these mutations were all beneficial under their original isolation conditions and were surprisingly more beneficial under one of the two new conditions we considered (Figure 1 and Figure 2A). That deleterious mutations remain deleterious across environments is not unexpected, but that many mutations could retain their benefits across environments implies that, even if environment alters the effects of mutations, it may not always drastically alter the course of adaptive evolution. This result may also reflect the fact that our environmental variable is continuous rather than categorical;

Table 3 Fitnesses and fitness effects of all genotypes tested at 37°

Genotype	Fitness	<i>n</i>	ΔW_{wt}	ΔW_{add}	ΔW_1	ΔW_2	ε_{ij}
ID11	15.18 ± 0.20	14	—	—	—	—	—
A	19.07 ± 0.19	5	3.89	—	—	—	—
B	19.34 ± 0.43	5	4.15	—	—	—	—
C	19.36 ± 0.56	6	4.18	—	—	—	—
D	18.62 ± 0.49	7	3.44	—	—	—	—
E	16.84 ± 0.36	5	1.65	—	—	—	—
F	18.58 ± 0.37	5	3.39	—	—	—	—
G	21.02 ± 0.26	6	5.84	—	—	—	—
H	18.62 ± 0.42	5	3.44	—	—	—	—
I	16.60 ± 0.28	5	1.42	—	—	—	—
AB	18.66 ± 0.25	5	3.47	8.04	-0.68	-0.42	-4.57
AD	18.82 ± 0.37	5	3.63	7.33	0.19	-0.26	-3.69
AF	18.07 ± 0.56	5	2.88	7.28	-0.51	-1.01	-4.40
AG	22.50 ± 0.25	5	7.32	9.73	1.48	3.43	-2.41
AH	20.29 ± 0.29	5	5.10	7.33	1.66	1.21	-2.23
BD	17.77 ± 0.33	5	2.59	7.59	-0.85	-1.56	-5.00
BE	15.58 ± 0.53	6	0.40	5.81	-1.25	-3.75	-5.41
BG	17.31 ± 0.54	7	2.12	9.99	-3.72	-2.03	-7.87
BH	19.47 ± 0.43	5	4.29	7.59	0.85	0.14	-3.30
BI	16.52 ± 0.48	7	1.34	5.57	-0.08	-2.81	-4.23
CD	17.52 ± 0.36	6	2.34	7.62	-1.10	-1.84	-5.28
CE	11.56 ± 0.47	6	-3.62	5.83	-5.27	-7.80	-9.46
CH	19.49 ± 0.28	5	4.31	7.62	0.87	0.13	-3.31
DF	19.30 ± 0.31	5	4.12	6.83	0.73	0.68	-2.72
DH	18.48 ± 0.43	5	3.30	6.88	-0.14	-0.14	-3.58
DI	15.40 ± 0.34	5	0.21	4.86	-1.21	-3.23	-4.64
EH	16.54 ± 0.44	5	1.35	5.10	-2.09	-0.30	-3.74
EI	12.74 ± 0.35	5	-2.44	3.07	-3.86	-4.09	-5.51

Fitnesses are given as the average plus or minus the standard error. *n* gives the number of replicate assays for each genotype. The fitness effect relative to the wild type is designated by ΔW_{wt} . ΔW_{add} gives the fitness effect expected on the assumption that the effects of single mutations were additive, ΔW_1 gives the effect of adding the first mutation in the genotype name into the background of the second, and ΔW_2 gives the effect of adding the second mutation in the genotype name into the background of the first. The deviation from additivity, ε_{ij} for the combination of mutations *i* and *j*, was calculated according to Equation 2.

the difference between growing at two different temperatures may be fundamentally different from that between growing on, for example, two different hosts.

Epistasis varies across environments

Before describing our results for epistatic patterns, we must first establish appropriate terminology. We begin with epistasis itself, which we define as departure from additive fitness effects for mutations. Antagonistic epistasis refers to cases where the effect of mutations in combination is less than expected, and synergistic epistasis refers to cases where the effect is larger. This terminology has been used when mutations are all either beneficial or deleterious (Sanjuán *et al.* 2004). In our results, we have cases that cross these standard boundaries of strictly beneficial or strictly deleterious mutations, so we also refer to cases where the expected (additive) effect and observed effects are of opposite sign as antagonistic (Figure 3), because such pairs are either less beneficial or less deleterious than expected. Positive epistasis refers to cases where the fitness of the mutations in combination is higher than expected, and negative epistasis refers to cases where fitness is lower than expected (Figure 3). We

Table 4 Fitnesses and fitness effects of all genotypes tested at 41°

Genotype	Fitness	<i>n</i>	ΔW_{wt}	ΔW_{add}	ΔW_1	ΔW_2	ε_{ij}
ID11	-4.08 ± 0.61	16	—	—	—	—	—
A	0.08 ± 0.59	7	4.16	—	—	—	—
B	9.22 ± 0.19	5	13.30	—	—	—	—
C	3.00 ± 0.16	5	7.08	—	—	—	—
D	8.07 ± 0.25	5	12.15	—	—	—	—
E	6.08 ± 0.30	5	10.18	—	—	—	—
F	7.27 ± 0.29	15	11.36	—	—	—	—
G	5.61 ± 0.36	6	9.69	—	—	—	—
H	5.30 ± 0.09	5	9.38	—	—	—	—
I	2.89 ± 0.33	5	6.98	—	—	—	—
AB	8.78 ± 0.16	5	12.86	17.46	-0.44	8.70	-4.60
AD	9.33 ± 0.33	5	13.42	16.32	1.27	9.26	-2.90
AF	10.98 ± 0.35	5	15.07	15.52	3.71	10.91	-0.45
AG	8.60 ± 0.46	7	12.68	13.86	2.99	8.52	-1.17
AH	9.14 ± 0.37	5	13.23	13.55	3.85	9.07	-0.32
BD	12.17 ± 0.40	6	16.25	25.45	4.10	2.95	-9.20
BE	5.90 ± 0.17	5	9.98	23.48	-0.20	-3.32	-13.49
BG	9.66 ± 0.34	6	13.74	23.00	4.05	0.44	-9.25
BH	9.76 ± 0.36	5	13.84	22.69	4.46	0.54	-8.85
BI	6.63 ± 0.26	5	10.71	20.28	3.73	-2.59	-9.57
CD	8.95 ± 0.17	5	13.03	19.24	0.88	5.95	-6.21
CE	2.98 ± 0.29	6	7.06	17.26	-3.12	-0.02	-10.20
CH	5.56 ± 0.33	5	9.65	16.47	0.27	2.57	-6.82
DF	11.81 ± 0.22	5	15.89	23.51	4.53	3.74	-7.62
DH	10.40 ± 0.56	5	14.49	21.54	5.11	2.34	-7.05
DI	7.30 ± 0.36	5	11.38	19.13	4.40	-0.77	-7.75
EH	5.42 ± 0.44	5	9.50	19.56	0.12	-0.68	-10.06
EI	3.29 ± 0.26	5	7.37	17.15	0.40	-2.81	-9.78

Fitnesses are given as the average plus or minus the standard error. *n* gives the number of replicate assays for each genotype. The fitness effect relative to the wild type is designated by ΔW_{wt} . ΔW_{add} gives the fitness effect expected on the assumption that the effects of single mutations were additive, ΔW_1 gives the effect of adding the first mutation in the genotype name into the background of the second, and ΔW_2 gives the effect of adding the second mutation in the genotype name into the background of the first. The deviation from additivity, ε_{ij} for the combination of mutations *i* and *j*, was calculated according to Equation 2.

measured the strength of epistasis by means of a deviation from additivity (Sanjuán *et al.* 2004; Rokyta *et al.* 2011)

$$\varepsilon_{ij} = \Delta W_{ij} - (\Delta W_i + \Delta W_j), \quad (2)$$

where ΔW_i is the fitness effect of single mutant *i*, and ΔW_{ij} is the effect of combining mutations *i* and *j*. Because our fitness is measured as a Malthusian fitness (*i.e.*, a growth rate), we expected effects to add together in double mutants in the absence of epistasis, which corresponds to multiplicative effects if we switched to a more standard Wrightian fitness. Under our terminology, $\varepsilon > 0$ corresponds to positive epistasis, and $\varepsilon < 0$ corresponds to negative epistasis. For strictly beneficial mutations, $\varepsilon < 0$ also indicates antagonistic epistasis, and $\varepsilon > 0$ indicates synergistic epistasis. For deleterious mutations, the signs are reversed for antagonistic and synergistic epistasis (but not for positive and negative epistasis). We classified cases for which the signs of ΔW_{ij} and $(\Delta W_i + \Delta W_j)$ differed as antagonistic epistasis.

As reported by Rokyta *et al.* (2011), the average deviation from additivity at 37° for the nine mutations was $\bar{\varepsilon}_{37} = -4.52 \pm 0.43$ doublings per hour (plus or minus the standard error). This corresponds to antagonistic and negative epistasis.

Our current results revealed that the average deviation at 33° was $\bar{\varepsilon}_{33} = -1.24 \pm 0.31$ doublings per hour, and the deviation at 41° was $\bar{\varepsilon}_{41} = -6.96 \pm 0.88$ doublings per hour. All three means were significantly less than zero ($P < 0.01$ with two-sided one-sample *t*-tests and a Bonferroni correction for three tests); epistasis was significantly negative in all three environments. The deviations from additivity were significantly different among temperatures ($F_{2,51} = 23.53$, $P \ll 0.001$). We therefore found strong evidence that epistasis was present in all three environments and that its patterns were different across environments. Such differences in patterns of epistasis among different hosts have been observed for the RNA virus TEV (Lalić and Elena 2013). In our case, the deviations from additivity increased in magnitude with temperature, indicating that epistasis (as defined by Equation 2) is becoming more pronounced and negative as the effects are measured in worse environments for our wild-type phage (Figure 2B). This result is in contrast to the results of Wang *et al.* (2013), who showed that, for two beneficial mutations for *E. coli*, epistasis was more negative in genetic backgrounds with higher growth rates.

The average deviations from additivity are useful measures for characterizing the overall patterns of epistatic interactions, but their interpretation is not simple. The average deviation was by far the smallest at 33°, but this might not necessarily indicate that epistasis is weaker at this temperature. The effects of mutations at this lower temperature were much smaller (Figure 1 and Figure 2A) and therefore the expectations for their effects in combination are smaller (Figure 3). This observation might explain the difference in the average magnitude of the deviation from additivity across temperatures. If we normalize the deviations from additivity by the average magnitudes of the effects of the mutations individually, we get $\bar{\varepsilon}_{33}/|\bar{\Delta W}_{33}| = -0.98$, $\bar{\varepsilon}_{37}/|\bar{\Delta W}_{37}| = -1.29$, and $\bar{\varepsilon}_{41}/|\bar{\Delta W}_{41}| = -0.74$. Under this normalized measure, epistasis is strongest at 37° and weakest at 41°. Because this normalized measure of epistasis involves a ratio of two nonindependent values, standard tests of significance were not applicable. We instead used a bootstrap approach with 10,000 replicates where, in each replicate, double mutants were resampled with replacement, the set of single mutants that appeared in this random set of doubles was used to calculate the mean effect sizes at each temperature, ε_{ij} values for each double at each temperature were normalized by dividing by the mean effect size, and then normalized values were averaged. For each replicate, the same set of randomly selected genotypes was used across temperatures. We ranked the temperatures by their magnitudes of normalized deviations; rank 1 therefore refers to the temperature with the most negative normalized deviation. The proportions of the bootstrap samples in which a given temperature achieved ranks 1, 2, and 3, respectively, were 33°, 0.17, 0.81, 0.02; 37°, 0.83, 0.17, 0.00; and 41°, 0.00, 0.02, 0.98. Under this normalized measure, epistasis was clearly weakest at 41° because it ranked lowest in 98% of the samples. Epistasis appears to be strongest at 37°, but 33° was still ranked

Figure 1 Fitness effects increased with temperature. The dashed lines designate the mean fitness effects at each temperature. The fitness of the wild-type genotype, ID11, declined with increasing temperature (Table 2, Table 3, and Table 4), while the average benefit of the mutations increased. Note that the rank order of the genotypes changes across temperatures. Values are given in units of population doublings per hour. Error bars designate plus or minus one standard error. The letters correspond to the mutational identities described in Table 1.

highest in 17% of samples. Under both the normalized and the nonnormalized measures, we found significant variation in epistasis across temperatures, but whether epistasis was strongest or weakest at 41° depends on whether our null expectations for additivity involve absolute or relative fitness effects.

Beyond these magnitude differences, qualitative differences exist depending on the terminology we employ (Figure 3). The patterns for 37° and 41° are qualitatively the same. Under both conditions, all of the mutations were beneficial, and all of the deviations from additivity were negative (Figure 3). Epistasis was therefore negative and antagonistic for all pairs of mutations. The fitnesses of the double mutants were less than expected and the fitness effects in combination were smaller than expected. At 33°, epistasis is also predominantly negative; the fitnesses of the double mutants were less than expected (Figure 3). Most of the mutations were deleterious at 33°, so the fitnesses of the double mutants were less than expected because the deleterious effects in combination were generally larger than expected. The fitness effects therefore showed synergistic epistasis at 33°. In summary, the mutations under all conditions showed negative epistasis. When the effects of mutations were beneficial, they showed antagonistic epistasis, and they showed synergistic epistasis when they were deleterious.

Empirical measures of epistatic patterns have varied substantially among systems (Bonhoeffer *et al.* 2004). Beneficial mutations have predominantly shown negative and antagonistic epistasis (Sanjuán *et al.* 2004; Chou *et al.* 2011; Khan *et al.* 2011; Kvitek and Sherlock 2011; Rokyta *et al.* 2011; Wang *et al.* 2013) as predicted by theory based on Fisher's geometric model (Martin *et al.* 2007). Most information about epistasis for individual mutations comes from deleterious mutations, in part because they are easier to identify, but also because their epistatic patterns have implications for the evolution and maintenance of sexual reproduction (Kondrashov 1988, 1993). In particular, sexual reproduction will be favored with synergistic epistasis for deleterious mutations and antagonistic for beneficial mutations, both of which

correspond to negative epistasis (Bonhoeffer *et al.* 2004). When epistasis is negative, recombination mediated through sexual reproduction allows deleterious mutations to be more effectively eliminated from populations. Deleterious mutations have been found to show both antagonistic (Burch and Chao 2004; Sanjuán *et al.* 2004; Sanjuán 2006; Pepin and Wichman 2007; Trindade *et al.* 2009) and synergistic epistasis (Sanjuán *et al.* 2004). Our results, demonstrating nearly strictly negative epistasis for pairs of mutations under conditions where they were both beneficial and deleterious, are consistent with theories of the origin and maintenance of sexual reproduction based on epistatic interactions between mutations and their role in the elimination of deleterious mutations from populations.

Sign epistasis

Negative and positive epistasis are relevant to the evolution of sex, but sign epistasis determines the accessibility of different adaptive outcomes and therefore the predictability, repeatability, and limitations of evolution (Weinreich *et al.* 2005, 2006; Poelwijk *et al.* 2011). Sign epistasis refers to cases where the effect of a mutation—in particular whether it is beneficial or deleterious—depends on its genetic context. For our 9 single mutations and 18 double mutants at 37°, Rokyta *et al.* (2011) found evidence for decompensatory epistasis, which is a special case of sign epistasis for which mutations that are beneficial become deleterious in some genetic contexts. In the context of building double mutants from pairs of mutants that are individually beneficial, we can imagine three different outcomes for the fitness of a double mutant relative to its two constituent single mutants (Figure 4):

1. The fitness of the double mutant could be higher than both single-mutant fitnesses, which would indicate that each mutation retains a fitness benefit in the presence of its pair, and we have no evidence for sign epistasis. In this case, either genetic pathway linking the wild type to the double mutant involves only fitness improvement.
2. The fitness of the double mutant could be less than both single-mutant fitnesses, which would indicate that adding the second of the two mutations always comes at a fitness

Figure 2 As fitness-effect sizes increased with temperature, epistasis became more negative. (A) Fitness effects changed from predominantly negative to entirely positive as the temperature increased from 33° to 41°. The dashed line distinguishes deleterious from beneficial fitness effects. Fitnesses were measured as population doublings per hour. (B) Epistasis became more negative as temperature increased from 33° to 41°. The dashed line indicates additivity of mutational effects. Deviations from additivity were calculated according to Equation 2 and are given in units of doublings per hour. Values for the same mutant are connected by solid lines.

cost regardless of the order. For this case, the pathways linking the wild type to the double mutant always involve a fitness loss.

3. The fitness of the double mutant could instead fall between the fitnesses of the two single mutants. In this case, the order matters; the second mutation would be beneficial or deleterious depending on which was added first. One pathway is all uphill on the fitness landscape, but the other has a fitness valley.

The latter two cases provide evidence for sign epistasis; if the fitness of the double mutant is less than the fitness of at least one of the singles, sign epistasis is present. To test for the presence of sign epistasis, we therefore need test only whether the fitness of the double mutant is less than the fitness of its constituent single mutant with the highest fitness for the two data sets, 37° and 41°, for which the mutations were all beneficial individually. Using one-sided Welch two-sample *t*-tests with a Bonferroni correction for 18 tests, we found that six double mutants (BE, BG, BI, CE, DI, and EI; $P < 0.05$) showed evidence for sign epistasis at 37°, as reported by Rokyta *et al.* (2011) (Figure 4). Similarly, for 41°, we found evidence of sign epistasis for only four double mutants (BE, BI, CE, and EI; $P < 0.05$).

At 33°, two of the single mutations had average effects greater than zero and seven had average effects less than zero (Table 2). The mutations were not strictly beneficial (or deleterious) as they were at the other two temperatures (Table 3 and Table 4). To test for evidence of sign epistasis for the mutants at 33°, we therefore need new criteria. For the one double mutant composed of two beneficial mutations (AH), we can use the same criterion as for 37° and 41°; if the fitness of the double is less than the fitness of the largest-effect single, we must have sign epistasis. For doubles composed of two deleterious mutations (BD, BE, BG, BI, CD, CE, DF, DI, and EI), if the fitness of the double is greater than the fitness of the lowest-fitness single mutant, we must have

sign epistasis. For the eight cases for which one mutation was beneficial and the other was deleterious (AB, AD, AF, AG, BH, CH, DH, and EH), we would need the double-mutant fitness to either be greater than that of the beneficial mutant or less than that of the deleterious mutant. For each double mutant, we selected the appropriate one-sided Welch two-sample *t*-test and used a Bonferroni correction for 18 tests. For the doubles comprising a beneficial and a deleterious single mutant, we tested whether the fitness of the double mutant was less than that of the deleterious mutant, because in all cases the fitness of the double was closer to that of the deleterious mutant. Of all of the 18 double mutants at 33°, only one (AD; $P < 0.05$) showed evidence for sign epistasis.

Relative to the results for 37°, the results for 33° showed significantly fewer cases of sign epistasis ($P = 0.04$, Fisher's exact test), but the results for 41° did not show significantly fewer cases ($P = 0.36$, Fisher's exact test). Sign epistasis was therefore most pronounced under the conditions used to select for the original nine mutations, 37°, and least pronounced at 33°, where the wild type is most fit (Figure 4).

Additivity of phenotypic effects

We demonstrated above that, for our genotype and nine mutations, the effects of mutations (Figure 1 and Figure 2A) and their epistatic interactions (Figure 2B and Figure 3) depend, in an apparently complex manner, on the environment. For our phage genotypes, epistasis was predominantly negative, but synergistic when mutations were deleterious and antagonistic when mutations were beneficial (Figure 3). The average magnitude of the deviation from additivity (Equation 2) increased significantly with temperature (Figure 2B), as did the average magnitudes of the effects of the individual mutations (Figure 2A). When normalized for mutational effect sizes, the deviation from additivity was actually greatest at our intermediate temperature. Sign epistasis was detectable at all temperatures, but more common at higher

Figure 3 Epistasis was predominantly negative and antagonistic. Negative epistasis dominated at all temperatures, and the average magnitude of the deviation from additivity increased with increasing temperature. Most pairs of mutations showed antagonistic epistasis at 37° and 41° but showed synergistic epistasis at 33°.

temperatures (Figure 4). We have provided a characterization of how the patterns of epistasis change with environment while keeping the genotypes involved constant, but, as simple patterns, they provide little insight into why epistasis changes as it does for temperature and no basis for predicting changes under new conditions.

Rokyta *et al.* (2011) and Pearson *et al.* (2012) used a simple, one-dimensional version of Fisher's geometric model (Fisher 1930) to explain patterns of epistasis for beneficial mutations. Both studies assumed that mutations were affecting a single phenotype, the phenotypic effects of the mutations were additive across genetic backgrounds, and phenotypes were mapped to fitness by means of a gamma function. Some evidence, but certainly not all evidence (Lehner 2011), suggests that underlying biophysical or biochemical effects of mutations could be additive in some cases (DePristo *et al.* 2005) and therefore epistasis at the level of fitness effects could arise from nonlinearity in the mapping from phenotype to fitness. The gamma function is flexible and amounts to the assumption of stabilizing selection around an optimal phenotype; it also allows for asymmetry in the phenotype–fitness relationship. Because phenotypes are unknown and unmeasured, they must be treated as missing data and imputed. Because the same mutations appear many times in different contexts, they can be estimated without overparameterization. Pearson *et al.* (2012) used this type of model to explain how the effects of two of our mutations, A and F, change when they are introduced into phages that are closely related to our original wild type, ID11. They found that the effects of the

two mutations were highly context dependent but could be explained by assuming that their phenotypic effects were context independent (*i.e.*, additive). Rokyta *et al.* (2011) showed that the pattern of antagonistic and decompensatory epistasis for our nine beneficial mutations at 37° could also be explained by an additive phenotypic effects gamma model. Decompensatory epistasis arose because many of the single mutations exceeded the phenotypic optimum, causing the addition of another mutation to drive the genotype even farther from the optimum.

Given the success of the additive phenotypic effects model in explaining epistasis for a set of mutations and genotypes under a single environmental condition, we applied it to the present problem of understanding how epistatic patterns depend on the environment in which fitness is measured. To do so, we must first consider what aspects of the model can change with temperature. The simplest model would assume that only the relationship between phenotype and fitness changes with temperature, while the phenotypes remain constant. We refer to this model as G1 because each genotype has one phenotype that is retained across environments. Alternatively, we can allow both the phenotype–fitness relationship and the phenotypes to change with temperature. We refer to this more complicated model as G3. The G1 model gave an overall $R^2 = 0.51$ (Figure 5). The difference in log-likelihood scores between G1 and an intercept-only null model, which assumes that fitnesses at each temperature are normally distributed with a separate mean for each temperature, was $\Lambda_{G1} = 29.98$. We conducted a randomization test to assess significance by randomizing the single-mutant and double-mutant fitnesses separately and reestimating Λ_{G1} for each of 100 randomized data sets. None of the 100 randomized data sets produced a value as large as our observed value (all were <26.49), indicating that G1 was a significant improvement over the null ($P < 0.01$). The G1 model most effectively explained the patterns at 33°, with $R^2_{33} = 0.61$, $R^2_{37} = 0.35$, and $R^2_{41} = 0.55$ (Figure 5). At 33°, all of the single and double mutants fell to the right of the peak, indicating that their predominantly deleterious effects could be explained by their overshooting the phenotypic optimum. Even though the pattern for epistasis at 33° is predominantly negative and synergistic (Figure 3), the patterns are much closer to additivity than for the other two temperatures, and this is reflected in the nearly linear decline in fitness with phenotype in the fitted model (Figure 5). At 41°, all of the single mutants fell to the left of the optimum and all therefore increased fitness relative to the wild-type genotype. Most of the double mutants had phenotypes near or slightly past the optimum, which explains the small observed amount of sign epistasis. The curvature of the map around the optimum explains why the doubles showed negative and antagonistic epistasis. Finally, at 37°, the patterns were intermediate to those at the other two temperatures. Most of the single mutants had phenotypes very near the optimum, and the double mutants therefore were pushed farther off of the optimum, generating extensive sign, negative,

Figure 4 Fitness landscapes show that sign epistasis was most common in the original environment (37°). The plots show the two possible mutational pathways linking the wild-type ID11 to each double mutant. Single mutations are listed below each plot with the higher-fitness mutant above the lower-fitness mutant. Paths shown as dashed lines indicate the presence of sign epistasis. Fitness values are shown in units of doublings per hour.

and antagonistic epistasis. Nonetheless, even though the model fitted significantly better than the null model, it did not do a particularly good job of explaining the variation in the data. The more complex G3 model gave an overall $R^2 = 0.81$ (Figure 6). The difference in log-likelihoods between G3 and G1 was $\Delta_{G3-G1} = 71.5$. We used parametric bootstrapping to assess significance; only 5 of the 100 parametric bootstrap data sets simulated under G1 yielded a Δ_{G3-G1} this large or larger, indicating that G3 was likely a better explanation of our data than G1 ($P \approx 0.05$). Compared to G1, the fit of the model across temperatures was better (as expected) and more uniform, with $R^2_{33} = 0.84$, $R^2_{37} = 0.78$, and $R^2_{41} = 0.82$. The major patterns described above for G1 were nonetheless largely retained by G3. The single mutations tended to exceed the phenotypic optimum at 33°, be near the optimum at 37°, and be shy of the optimum at 41°. The improvement in fit between models G3 and G1 implies that, in addition to the phenotype–fitness map, the phenotype is influenced by environment.

The additive phenotype models with gamma phenotype–fitness maps can clearly explain the patterns of fitness effects and epistatic interactions observed in our measurements across three temperatures. The comparison between G1, which

assumes constant phenotypes across temperatures, and the normal null, which makes no constraints on phenotypes, suggests that imposing additive phenotypes and stabilizing selection explains a significant portion of the variation in our data. The epistatic patterns change across temperatures in a manner consistent with simply shifting the shape, but not the form, of the phenotype–fitness relationship. The phenotype–fitness relationship determines much of the change across temperature under G1. The rejection of G1 in favor of G3, however, suggests that some of the difference across temperatures is attributable to differences in the genotype–phenotype relationship across temperatures. This may in part be due to our assumption of a single phenotypic dimension. In reality, these mutations probably affect a number of phenotypes that contribute to fitness.

Given that neither the phenotype–fitness functions nor the phenotypes were observed, one might fairly ask whether our models, though statistically significant, have a connection to biological reality. To explore whether our gamma models captured aspects of the underlying biology in our system, we determined whether the models performed poorly when we might expect them to on the basis of structural considerations. We reasoned that pairs of mutations affecting amino acid sites

Figure 5 The phenotype–fitness maps for three temperatures with constant phenotypes across temperatures. The G1 model assumes a single phenotype for each genotype with a gamma phenotype–fitness map that changes with temperature. The overall fit of the model was $R^2 = 0.51$, and we were able to reject a null model, assuming that fitnesses are normally distributed with different means for each temperature by means of a randomization test ($P < 0.01$). Note that at 33°, all of the singles were to the right of the peak and at 41°, they were all to the left of the peak.

closer together in the capsid structure should be less likely to show additive phenotypic effects as assumed by both G1 and G3. We therefore compared distances between α -carbons at mutational sites in the double mutants against the average (signed) deviation from the model predictions across temperatures for G1 (Figure 5) and G3 (Figure 6). We found a significant positive correlation for both G1 ($P = 0.014$, $R^2 = 0.32$) and G3 ($P = 0.008$, $R^2 = 0.36$). In particular, mutational pairs close together in the structure, such as CE and EI, tended to fall below the fitted curve; this nonrandom association suggests that the model successfully identified a real biological signal. In fact, the CE pair, which was nearly always

the most extreme outlier for both models, is composed of two mutations affecting adjacent sites in the protein structure. It is hard to imagine how these mutations could have independent phenotypic effects, and the models clearly failed to fit them (Figure 5 and Figure 6).

Our modeling results suggest that an intermediate phenotypic optimum is important in determining the fitnesses of mutants and the epistatic interactions between them. Our nine mutations were primarily located along interfaces between the major coat protein subunits in the final capsid structure, suggesting a role in assembly kinetics. Viral assembly kinetics theory (Zlotnick 1994, 2003; Zlotnick and Stray 2003) has shown

Figure 6 The phenotype–fitness maps for three temperatures with independent phenotypes for each temperature. The G3 model assumes that each temperature has its own gamma phenotype–fitness map and phenotypes. The overall fit of the model was $R^2 = 0.81$, and we were able to reject the G1 model in favor of the G3 model with $P = 0.05$ by means of parametric bootstrapping.

that proper assembly requires intersubunit binding energies in a narrow range of intermediate strength. If binding energies are too low, assembly does not begin or proceeds slowly. If they are too high, assembly gets stuck in kinetic traps, leading to the formation of primarily incomplete assembly intermediates. Faster, more efficient assembly would certainly lead to higher overall growth rates, and the energetics of the assembly reactions should depend intimately on temperature. Within this framework, our results (Figure 5 and Figure 6) suggest that the primary molecular cause of epistasis might shift from kinetic traps at low temperature, where the wild type is near the optimal binding energy, to slower assembly of intermediates at high temperature, where the wild type is far below the optimum.

Our models therefore had both predictive power and a relation to the underlying biology of our experimental system. Other approaches to modeling epistasis exist, however, and it is worth asking whether they would be viable alternatives for explaining our data. Rokyta *et al.* (2011) showed that our gamma model performs as well as the model developed by Martin *et al.* (2007) for the data at 37°, which is not surprising because both models make the same underlying biological assumptions of stabilizing selection on additive phenotypes. Both models are simply implementations of Fisher's geometric model. Highly epistatic, random landscape models such as those assumed in implementations of the mutational landscape model (Gillespie 1984; Orr 2002) can be rejected outright because, under those models, the fitness effects of double mutants would have the same probability of being beneficial as a random mutation, which should be quite low, but most of the double mutants were beneficial at 37° and 41°.

Our modeling results provide simple, intuitive explanations for the complex patterns of epistatic interactions for our nine mutations and support for Fisher's geometric model (Fisher 1930). Like our simple models, Fisher's geometric model assumes additive phenotypic effects with some intermediate, optimal phenotype. The model is often implemented with a Gaussian phenotype–fitness relationship in a high-dimensional phenotypic space (as opposed to our single dimension) and has been used successfully in theoretical investigations of the effects of mutations (Fisher 1930; Martin and Lenormand 2008), epistasis (Martin *et al.* 2007), the evolution of sex (Peck *et al.* 1997), dominance (Manna *et al.* 2011), the cost of complexity (Orr 2000), and development (Rice 1990) (among other topics). Because of statistical limitations, our models and those of Rokyta *et al.* (2011) and Pearson *et al.* (2012) were constrained to a single phenotypic dimension, but otherwise captured the major assumption of Fisher's geometric model—the additivity of phenotypic effects of mutations. Rokyta *et al.* (2011) showed that such a model can explain patterns of epistasis between individual mutations, and Pearson *et al.* (2012) showed that such a model can explain epistasis between mutations and their broader genetic context. Here we have shown that a Fisher-like model can be extended to explain the interplay between epistatic patterns and the environment.

Acknowledgments

Funding for this work was provided by the U.S. National Institutes of Health (NIH) (NIH R01 GM099723 to D.R.R. and NIH R01 GM076040 to C.R.M.).

Literature Cited

- Bonhoeffer, S., C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, 2004 Evidence for positive epistasis in HIV-1. *Science* 306: 1547–1550.
- Bull, J. J., M. R. Badgett, and H. A. Wichman, 2000 Big-benefit mutations in a bacteriophage inhibited with heat. *Mol. Biol. Evol.* 17: 942–950.
- Burch, C. L., and L. Chao, 2004 Epistasis and its relationship to canalization in the RNA virus $\phi 6$. *Genetics* 167: 559–567.
- Chou, H.-H., J. Berthet, and C. J. Marx, 2009 Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. *PLoS Genet.* 5: e1000652.
- Chou, H.-H., H.-C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190–1192.
- Coyne, J. A., and H. A. Orr, 2004 *Speciation*. Sinauer Associates, Sunderland, MA.
- de Visser, J. A. G. M., T. F. Cooper, and S. F. Elena, 2011 The causes of epistasis. *Proc. Biol. Sci.* 278: 3617–3624.
- Dean, A. M., D. E. Dykhuizen, and D. L. Hartl, 1986 Fitness as a function of β -galactosidase activity in *Escherichia coli*. *Genet. Res.* 48: 1–8.
- DePristo, M. A., D. M. Weinreich, and D. L. Hartl, 2005 Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* 6: 678–687.
- Dobzhansky, T., 1936 Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21: 113–135.
- Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- Flynn, K. M., T. F. Cooper, F. B. G. Moore, and V. S. Cooper, 2013 The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. *PLoS Genet.* 9: e1003426.
- Gillespie, J. H., 1984 Molecular evolution over the mutational landscape. *Evolution* 38: 1116–1129.
- Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. *Genetics* 97: 639–666.
- Kauffman, S., and S. Levin, 1987 Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128: 11–45.
- Kauffman, S. A., 1993 *The Origins of Order*. Oxford University Press, New York.
- Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332: 1193–1195.
- Knies, J. L., J. G. Kingsolver, and C. L. Burch, 2009 Hotter is better and broader: thermal sensitivity of fitness in a population of bacteriophages. *Am. Nat.* 173: 419–430.
- Kondrashov, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* 336: 435–440.
- Kondrashov, A. S., 1993 Classification of hypotheses on the advantage of amphimixis. *J. Hered.* 84: 372–387.
- Kvitek, D. J., and G. Sherlock, 2011 Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 7: e1002056.
- Lalić, J., and S. F. Elena, 2013 Epistasis between mutations is host-dependent for an RNA virus. *Biol. Lett.* 9: 20120396.
- Lalić, J., J. M. Cuevas, and S. F. Elena, 2011 Effect of host species on the distribution of mutational fitness effects for an RNA virus. *PLoS Genet.* 7: e1002378.

- Lehner, B., 2011 Molecular mechanisms of epistasis within and between genes. *Trends Ecol. Evol.* 27: 323–331.
- MacLean, R. C., G. G. Perron, and A. Gardner, 2010 Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* 186: 1345–1354.
- Manna, F., G. Martin, and T. Lenormand, 2011 Fitness landscapes: an alternative theory for the dominance of mutation. *Genetics* 189: 923–937.
- Martin, G., and T. Lenormand, 2008 The distribution of beneficial and fixed mutation fitness effects close to an optimum. *Genetics* 179: 907–916.
- Martin, G., S. F. Elena, and T. Lenormand, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* 39: 555–560.
- Muller, H. J., 1939 Reversibility in evolution considered from the standpoint of genetics. *Biol. Rev. Camb. Philos. Soc.* 14: 261–280.
- Orr, H. A., 2000 Adaptation and the cost of complexity. *Evolution* 54: 13–20.
- Orr, H. A., 2002 The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56: 1317–1330.
- Orr, H. A., 2006 The population genetics of adaptation on correlated fitness landscapes: the block model. *Evolution* 60: 1113–1124.
- Otto, S. P., 2009 The evolutionary enigma of sex. *Am. Nat.* 174: S1–S14.
- Otto, S. P., and M. W. Feldman, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* 51: 134–147.
- Pearson, V. M., C. R. Miller, and D. R. Rokytá, 2012 The consistency of beneficial fitness effects of mutations across diverse genetic backgrounds. *PLoS ONE* 7: e43864.
- Peck, J. R., G. Barreau, and S. C. Heath, 1997 Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* 145: 1171–1199.
- Pepin, K. M., and H. A. Wichman, 2007 Variable epistatic effects between mutations at host recognition sites in ϕ X174 bacteriophage. *Evolution* 61: 1710–1724.
- Perelson, A. S., and C. A. Macken, 1995 Protein evolution on partially correlated landscapes. *Proc. Natl. Acad. Sci. USA* 92: 9657–9661.
- Poelwijk, F. J., S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans, 2011 Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* 272: 141–144.
- R Development Core Team, 2006 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Remold, S. K., and R. E. Lenski, 2004 Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. *Nat. Genet.* 36: 423–426.
- Rice, S. H., 1990 A geometric model for the evolution of development. *J. Theor. Biol.* 143: 319–342.
- Rokytá, D. R., and H. A. Wichman, 2009 Genic incompatibilities in two hybrid bacteriophages. *Mol. Biol. Evol.* 26: 2831–2839.
- Rokytá, D. R., P. Joyce, S. B. Caudle, and H. A. Wichman, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Genet.* 37: 441–444.
- Rokytá, D. R., C. J. Beisel, and P. Joyce, 2006a Properties of adaptive walks on uncorrelated landscapes under strong selection and weak mutation. *J. Theor. Biol.* 243: 114–120.
- Rokytá, D. R., C. L. Burch, S. B. Caudle, and H. A. Wichman, 2006b Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J. Bacteriol.* 188: 1134–1142.
- Rokytá, D. R., C. J. Beisel, P. Joyce, M. T. Ferris, C. L. Burch *et al.*, 2008 Beneficial fitness effects are not exponential for two viruses. *J. Mol. Evol.* 67: 368–376.
- Rokytá, D. R., P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel *et al.*, 2011 Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genet.* 7: e1002075.
- Sanjuán, R., 2006 Quantifying antagonistic epistasis in a multifunctional RNA secondary structure of the Rous sarcoma virus. *J. Gen. Virol.* 87: 1595–1602.
- Sanjuán, R., A. Moya, and S. F. Elena, 2004 The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. USA* 101: 15376–15379.
- Sousa, A., S. Magalhães, and I. Gordo, 2012 Cost of antibiotic resistance and the geometry of adaptation. *Mol. Biol. Evol.* 29: 1417–1428.
- Szathmáry, E., 1993 Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. *Genetics* 133: 127–132.
- Trindade, S., A. Sousa, K. B. Xavier, F. Dionisio, M. G. Ferreira *et al.*, 2009 Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet.* 5: e1000578.
- Vale, P. F., M. Choisy, R. Froissart, R. Sanjuán, and S. Gandon, 2012 The distribution of mutational fitness effects of phage ϕ X174 on different hosts. *Evolution* 66: 3495–3507.
- Wang, Y., C. D. Arenas, D. M. Stoebel, and T. F. Cooper, 2013 Genetic background affects epistatic interactions between two beneficial mutations. *Biol. Lett.* 9: 20120328.
- Weinreich, D. M., R. A. Watson, and L. Chao, 2005 Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59: 1165–1174.
- Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312: 111–114.
- You, L., and J. Yin, 2002 Dependence of epistasis on environment and mutation severity as revealed by *in silico* mutagenesis of phage T7. *Genetics* 160: 1273–1281.
- Zlotnick, A., 1994 To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* 241: 59–67.
- Zlotnick, A., 2003 Are weak protein-protein interactions the general rule in capsid assembly? *Virology* 315: 269–274.
- Zlotnick, A., and S. J. Stray, 2003 How does your virus grow? Understanding and interfering with virus assembly. *Trends Biotechnol.* 21: 536–542.

Communicating editor: J. Lawrence