Dobzhansky–Muller model of hybrid dysfunction supported by poor burst-speed performance in hybrid tiger salamanders

B. M. FITZPATRICK
Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA

Abstract
Speciation may result from ‘complementary’ genetic differences that cause dysfunction when brought together in hybrids despite having no deleterious effects within pure species genomes. The theory of complementary genes, independently proposed by Dobzhansky and Muller, yields specific predictions about the genetics of hybrid fitness. Here, I show how alternative models of hybrid dysfunction can be compared using a simple multivariate analysis of hybrid indices calculated from molecular markers. I use the approach to fit models of hybrid dysfunction to swimming performance in hybrid tiger salamander larvae. Poor burst-speed performance is a dysfunction suggesting low vigour and could translate directly into low survival. My analyses show that the Dobzhansky–Muller model fits these data better than heterozygote disadvantage. The approach demonstrated here can be applied to a broad array of nonmodel species, potentially leading to important generalizations about the genetics of hybrid dysfunction.

Keywords:
Ambystoma; biological invasions; genetics; performance; reproductive isolation; speciation

Hybrid dysfunction was once seen as a paradox, posing species and speciation as major challenges to Darwin’s theory of natural selection (Darwin, 1859 Ch. 8; Huxley, 1908 p. 463; Orr, 1996; Coyne & Orr, 2004 Ch. 7). The problem is that natural selection should not allow the evolution of unfit offspring. The paradox was resolved independently by Dobzhansky (1936) and Muller (1940, 1942), who showed that independent evolution within isolated populations can result in genetic incompatibilities between populations without any population having to pass through a low-fitness transitional state (also see Poulton, 1908; Bateson, 1909; Orr, 1996). The simplest formalization of the Dobzhansky–Muller model begins with a species fixed for an $A_1A_1$ genotype at one locus and $B_1B_1$ at another. The species becomes split into two isolated populations. In one isolate, an $A_2$ allele becomes fixed (either by drift or selection; the point is that the genotypes $A_1A_2B_1B_1$ and $A_2A_2B_1B_1$ are not less fit than the ancestral $A_1A_1B_1B_1$). In the other population, $B_2$ becomes fixed. Hybrid genotypes combining the two derived alleles $A_2$ and $B_2$ may suffer low fitness, even sterility or inviability because of interactions between these alleles that evolved without deleterious effects within their native genetic backgrounds. That is, hybrid dysfunction results from ‘complementary’ epistasis between loci $A$ and $B$ (Orr, 1995).

This model is a great example of a theoretical ‘oversimplification’ that has proven extremely useful (e.g. Box, 1976). The Dobzhansky–Muller model has stimulated several advances in the theory of evolution of postzygotic reproductive isolation (e.g. Orr, 1995; Orr & Orr, 1996; Gavrilets, 1997, 2004; Turelli & Orr, 2000; Orr & Turelli, 2001; Kondrashov, 2003). For example, the widespread phenomenon known as Haldane’s rule can be explained if incompatibilities generally occur as recessive or partially recessive effects. Haldane’s rule is the empirical observation that, in animals with sex chromosomes, hybrids of the heterogametic sex are more likely to suffer dysfunction than hybrids of the homogametic sex (Haldane, 1922; Orr, 1997). Under the dominance theory (Muller, 1940, 1942; Orr, 1997; Turelli & Orr, 2000), Haldane’s rule occurs when sex-linked Dobzhansky–Muller incompatibilities are fully or partially recessive. That is, incompatibilities are not fully expressed when one or both loci are heterozygous for alleles from the different species, but they are expressed strongly in the heterogametic sex because loci on the sex
chromosome are hemizygous. The general prediction of the dominance theory (in the notation introduced in the previous paragraph) is that later generation hybrids with genotype $A_1A_2B_1B_2$ should suffer the greatest dysfunction, genotypes $A_1A_2B_2B_2$ and $A_2A_2B_1B_2$ should suffer less than half as much, and the double heterozygote $A_1A_2B_1B_2$ less than half as much dysfunction as the heterozygous–homozygous genotypes (Turelli & Orr, 2000). If this prediction is upheld across different kinds of organisms, it would represent one of the most exciting generalities in evolutionary genetics (Coyne & Orr, 2004).

The Dobzhansky–Muller model was proposed as an alternative to single-gene speciation in which allele $A_1$ is incompatible with allele $A_2$. This scenario is unlikely because the $A_2A_2$ genotype cannot evolve from the $A_1A_1$ genotype if the heterozygote $A_1A_2$ is sterile or inviable. That is, hybrid dysfunction should not be caused by low fitness of heterozygotes, because the evolutionary pathway between the two species is blocked by natural selection (‘Darwin’s dilemma’; Coyne & Orr, 2004, p. 267). There are two solutions that do allow single-gene speciation to occur. First, if $A_1A_2$ is not completely inviable or sterile, genetic drift may allow lineages to pass through ‘adaptive valleys’ created by low-fitness genotypes (Carson & Templeton, 1984). The primary criticism of this idea is that traversable adaptive valleys must be fairly shallow, that is $h_i$ remains an empirical issue.

The Dobzhansky–Muller model was proposed as an alternative to single-gene speciation in which allele $A_1$ is incompatible with allele $A_2$. This scenario is unlikely because the $A_2A_2$ genotype cannot evolve from the $A_1A_1$ genotype if the heterozygote $A_1A_2$ is sterile or inviable. That is, hybrid dysfunction should not be caused by low fitness of heterozygotes, because the evolutionary pathway between the two species is blocked by natural selection (‘Darwin’s dilemma’; Coyne & Orr, 2004, p. 267). There are two solutions that do allow single-gene speciation to occur. First, if $A_1A_2$ is not completely inviable or sterile, genetic drift may allow lineages to pass through ‘adaptive valleys’ created by low-fitness genotypes (Carson & Templeton, 1984). The primary criticism of this idea is that traversable adaptive valleys must be fairly shallow, that is $h_i$ remains an empirical issue.

The Dobzhansky–Muller model was proposed as an alternative to single-gene speciation in which allele $A_1$ is incompatible with allele $A_2$. This scenario is unlikely because the $A_2A_2$ genotype cannot evolve from the $A_1A_1$ genotype if the heterozygote $A_1A_2$ is sterile or inviable. That is, hybrid dysfunction should not be caused by low fitness of heterozygotes, because the evolutionary pathway between the two species is blocked by natural selection (‘Darwin’s dilemma’; Coyne & Orr, 2004, p. 267). There are two solutions that do allow single-gene speciation to occur. First, if $A_1A_2$ is not completely inviable or sterile, genetic drift may allow lineages to pass through ‘adaptive valleys’ created by low-fitness genotypes (Carson & Templeton, 1984). The primary criticism of this idea is that traversable adaptive valleys must be fairly shallow, that is $h_i$ remains an empirical issue.

The Dobzhansky–Muller model was proposed as an alternative to single-gene speciation in which allele $A_1$ is incompatible with allele $A_2$. This scenario is unlikely because the $A_2A_2$ genotype cannot evolve from the $A_1A_1$ genotype if the heterozygote $A_1A_2$ is sterile or inviable. That is, hybrid dysfunction should not be caused by low fitness of heterozygotes, because the evolutionary pathway between the two species is blocked by natural selection (‘Darwin’s dilemma’; Coyne & Orr, 2004, p. 267). There are two solutions that do allow single-gene speciation to occur. First, if $A_1A_2$ is not completely inviable or sterile, genetic drift may allow lineages to pass through ‘adaptive valleys’ created by low-fitness genotypes (Carson & Templeton, 1984). The primary criticism of this idea is that traversable adaptive valleys must be fairly shallow, that is $h_i$ remains an empirical issue.

The Dobzhansky–Muller model was proposed as an alternative to single-gene speciation in which allele $A_1$ is incompatible with allele $A_2$. This scenario is unlikely because the $A_2A_2$ genotype cannot evolve from the $A_1A_1$ genotype if the heterozygote $A_1A_2$ is sterile or inviable. That is, hybrid dysfunction should not be caused by low fitness of heterozygotes, because the evolutionary pathway between the two species is blocked by natural selection (‘Darwin’s dilemma’; Coyne & Orr, 2004, p. 267). There are two solutions that do allow single-gene speciation to occur. First, if $A_1A_2$ is not completely inviable or sterile, genetic drift may allow lineages to pass through ‘adaptive valleys’ created by low-fitness genotypes (Carson & Templeton, 1984). The primary criticism of this idea is that traversable adaptive valleys must be fairly shallow, that is $h_i$ remains an empirical issue.
analysis of the relationship between individual hybrid phenotypes (such as performance, fertility and viability) and descriptions of individual hybrid genotypes based on ancestry-informative molecular markers (AIMs).

My study system included an array of hybrid tiger salamander larvae from wild populations. I used molecular data and burst-speed trials to evaluate the relationship between the Turelli–Orr hybrid breakdown score (eqn 1) and hybrid performance. In particular, I compared the Dobzhansky–Muller model vs. heterozygote disadvantage as descriptions of hybrid dysfunction in tiger salamanders, and I estimated the relative magnitudes of the parameters in the Turelli–Orr model to evaluate the predictions of the dominance theory.

**Methods**

**Study system**

The salamanders *Ambystoma californiense* and *A. tigrinum mavortium* hybridize in California, where *A. t. mavortium* was introduced into the range of *A. californiense* (Riley et al., 2003). A diversity of hybrid genotypes can be found in wild populations, and persistent linkage disequilibrium and habitat-dependent invasion success indicate that incomplete reproductive isolation constrains admixture between the two lineages (Fitzpatrick & Shaffer, 2004). These factors make the *Ambystoma* hybrid zone a good system for studying variation among hybrid genotypes.

Theoretical discussions of hybrid dysfunction have usually focused on inviability or sterility. However, sub-lethal and condition-dependent dysfunction are common in hybrids and hybrid zones (e.g. Sage et al., 1986; Grant & Grant, 1998; Schluter, 1998; Ramsey et al., 2003). These negative phenotypes may have important effects on the outcome of natural selection in the wild, even though they do not cause outright death or sterility. There is no reason to think that the genetic basis of such quantitative hybrid dysfunction is qualitatively different from that of hybrid inviability or sterility. Furthermore, sub-lethal dysfunction can be observed, and the individuals expressing it can be genotyped. In contrast, outright inviability must often be inferred from the absence of certain genotypes.

Poor escape performance is a dysfunction that can directly increase the probability of death in the presence of predators, and more generally reflects lack of vigour that may influence successful feeding, interference competition and mating. Associations between burst swimming speed and survival in the face of predators have been demonstrated in sticklebacks (Swain, 1992), tadpoles (Watkins, 1996), and Guppies (Walket et al., 2003). I used burst speed as a measure of escape performance in tiger salamander larvae, and tested for hybrid dysfunction in performance among individuals collected from wild populations as embryos.

**Burst-speed performance**

Tiger salamander eggs were collected in February 2001 from four breeding ponds in the hybrid zone between *Ambystoma californiense* and *A. t. mavortium* (Riley et al., 2003). Pond 1 (32 eggs) is site 52 of Fitzpatrick & Shaffer (2007); pond 2 (36 eggs) is site 56, pond 3 (eight eggs) is site 21, and pond 4 (21 eggs) is site 55. Eggs were maintained in captivity until hatching and hatchlings were raised in a common laboratory environment as described by Fitzpatrick et al. (2003). When larvae were of 2–3 cm total length, they were tested for fast-start performance, measured, and killed for DNA analysis. This size range was large enough to be convenient for filming but not so large that the animals would no longer be vulnerable to predators in the field (primarily predaceous aquatic insects that themselves become prey for larger tiger salamander larvae).

Methods for measuring performance were described by Fitzpatrick et al. (2003). Briefly, each animal was placed in a dish over a 1 cm grid in 3 cm of water. As in nature, all larvae rested on the bottom and therefore ‘ground effects’ were an important part of the environment in all cases. Startle response was elicited by tapping the side of the dish. Escape responses were filmed at 500 frames s$^{-1}$ and each sequence digitized. After smoothing estimates of position over time, I estimated velocity (cm s$^{-1}$) by differentiating each time series.

I used the maximum velocity attained by each individual as its measure of propulsive performance. Each individual was measured only once, leaving within-individual variance as an uncontrolled source of error. I assume that this error is unbiased with respect to genotype so that it may reduce statistical power, but is unlikely to produce erroneous conclusions. It is generally assumed that the startle response of naïve larvae or fry elicits an escape ‘fast start’ at maximal effort; however burst speed might be affected by both physical capability and motivation or fearfulness (Irschick & Losos, 1998; Domenici & Blake, 2000). Thus realized performance (which provides the direct link to the fitness consequences of a particular action) results from morphological, physiological and behavioural factors.

**Molecular markers**

Each larva was assayed for the eight unlinked nuclear markers used by Fitzpatrick & Shaffer (2004, 2007), derived from Voss et al. (2001). These are single-nucleotide differences that distinguish *A. californiense* from *A. t. mavortium* alleles of eight mapped nuclear genes. Each individual was characterized as being homozygous for either species or heterozygous at each marker. I used these marker genotypes to estimate $P_T$, the proportion of each genome derived from introduced *A. t. mavortium*, $P_{CT}$, the proportion of each genome heterozygous, $P_{T}$, the proportion of the genome homozygous for
A. t. mavortium alleles, and \( P_{CC} \) the proportion of the genome homozygous for A. californiense alleles.

These markers represent a small sample of the genome and should not be thought of as candidates or QTL markers. Because of the high level of linkage disequilibrium in hybrid tiger salamander populations (Fitzpatrick & Shaffer, 2004), they can be expected to provide a good estimate of the overall genomic ancestry of an individual because the sampled loci are correlated with a good estimate of the overall genomic ancestry of an individual (Parra et al., 1998; Redden et al., 2006). Estimating genomic heterozygosity based on a few markers is unreliable in large panmictic populations where unlinked loci segregate independently (DeWoody & DeWoody, 2005; but see Aparicio et al., 2007). However, in admixed populations, statistical associations (linkage disequilibria) between unlinked loci decay slowly and can be maintained by immigration and by selection against hybrids (Long, 1991; Barton & Gale, 1993; Barton, 2000; Pritchard et al., 2000). When linkage disequilibrium is high across the genome, a few randomly chosen markers can be used to estimate genome-wide ancestry and heterozygosity (Parra et al., 1998; Redden et al., 2006). For example, a principal components analysis of allele frequencies for this set of markers in the tiger salamander hybrid zone (Fitzpatrick & Shaffer, 2007) found that 94.2% of the variance among ponds could be explained by the first principal component and all unlinked nuclear markers were strongly correlated with that component (variable–variate correlations > 0.97). Linkage disequilibria within ponds in this system are probably maintained by a combination of selection and dispersal between ponds with disparate allele frequencies (Fitzpatrick & Shaffer, 2004, 2007).

Statistical analyses

The predictor variables in the Turelli–Orr model (eqn 1) are products of estimated proportions from multinomial distributions (each individual has potentially unique \( P_{TT} \), \( P_{CT} \), and \( P_{CC} \). As multinomial proportions covary (they must sum to 1), estimates of their products must be corrected for bias. From the definition of covariance, the expectation of a product of random variables is the product of the means plus the covariance (Lynch & Walsh, 1998, p. 817). Similarly, the expectation of the square of a random variable is the mean plus the variance. From the covariances and variances of multinomial proportions, estimators for \( \hat{P}_2 = \hat{P}_{TT} \hat{P}_{CT} \), \( \hat{P}_1 = \hat{P}_{CT} (\hat{P}_{CC} + \hat{P}_{TT}) \) and \( \hat{P}_0 = \hat{P}_{CT}^2 \), are as follows:

\[
\begin{align*}
\hat{P}_2 &= \hat{P}_{CC} \hat{P}_{TT} + \frac{-\hat{P}_{CT} \hat{P}_{TT}}{k} \\
\hat{P}_1 &= \hat{P}_{CT} \hat{P}_{TT} + \hat{P}_{CT} \hat{P}_{CC} + \frac{-\hat{P}_{CT} \hat{P}_{TT} - \hat{P}_{CT} \hat{P}_{CC}}{k} \\
\hat{P}_0 &= \hat{P}_{CT}^2 + \frac{\hat{P}_{CT} (1 - \hat{P}_{CT})}{k}
\end{align*}
\]

where \( k \) is the number of markers (eight in this case).

I used Akaike’s information criterion (AIC and \( \Delta \zeta \); Burnham & Anderson, 2004) to compare the fit of six models to the data:

Model 1: \( BS_{hybrid} = \text{Constant} - (B_2 \hat{P}_2 + B_1 \hat{P}_1 + B_0 \hat{P}_0) \)

Model 2: \( BS_{hybrid} = \text{Constant} - (B_2 \hat{P}_2 + B_1 \hat{P}_1 + B_0 \hat{P}_0) + B_4 \hat{P}_T \)

Model 3: \( BS_{hybrid} = \text{Constant} - B_5 \hat{P}_{CT} \)

Model 4: \( BS_{hybrid} = \text{Constant} - (B_5 \hat{P}_{CT} + B_4 \hat{P}_T) \)

Model 5: \( BS_{hybrid} = \text{Constant} - (B_5 \hat{P}_{CT} + B_4 \hat{P}_T^2) \)

Model 6: \( BS_{hybrid} = \text{Constant} - (B_5 \hat{P}_{CT} + B_4 \hat{P}_T^2) + B_7 \hat{P}_T \)

(3)

Model 1 is a simple Turelli–Orr model. The coefficients, \( B_i \), estimate the composite parameters \( n_i \) in the Turelli–Orr model (eqn 1), that is, the product of the number of pairwise incompatibilities and the average effect of each type of incompatibility. Model 2 allows for an additive difference between the parental species by estimating a coefficient, \( B_4 \), for the relationship between maximum burst-speed velocity (BS) and the proportion of the genome derived from A. t. mavortium (\( P_T \)). Model 3 allows burst speed to depend only on heterozygosity (\( P_{CT} \)), with model 4 including the additive term. Models 5 and 6 fit a quadratic function of heterozygosity. The quadratic term could represent dominance \( \times \) dominance epistasis or a nonlinear relationship between the number of underdominant loci and performance. Models 1 and 2 correspond to the Dobzhansky–Mueller model where hybrid performance is a function of interactions between genes. Models 3 and 4 correspond to a nonepistatic model where hybrid performance is influenced by heterozygosity (dominance interactions between alleles of the same locus). Note that in the Turelli–Orr model, a simple effect of heterozygosity is inseparable from interactions involving heterozygous loci because \( P_{CT} (P_{TT} + P_{CC}) = P_{CT} - P_{CT}^2 \). That is, the epistatic Turelli–Orr model includes a nonlinear function of heterozygosity. An alternative treatment, using linear and quadratic functions of heterozygosity (\( P_{CT} \)) and ancestry (\( P_T \)) following Lynch (1991), is shown in Appendix S1 (see Supplementary material).

I fitted the models via iteratively reweighted least squares using the generalized linear model (glm) function in R 2.3.0 (http://www.r-project.org). I included terms for the source pond and the snout–vent length (SVL) of each individual to account for potential effects of these factors on burst speed. After choosing a model based on AIC, I used bootstrap distributions to estimate confidence limits for the fitted parameter values.

Traditional estimates of confidence intervals for regression coefficients do not account for error in the independent variables. The genomic proportions, \( P_{TT}, P_{CC}, P_T \) and \( P_{CT} \), were estimated with error due to sampling the genome. Therefore, I used two bootstrap approaches to incorporate this sampling variance into assessments of confidence in my parameter estimates. The standard nonparametric bootstrap (Efron, 1979) simulates the
The same triangular space can be described by these proportions define a ternary coordinate system. The easiest way to see this is by comparing the importance of rigorous evaluation of confidence intervals. The typical frequentist interpretation of nonparametric bootstrap distributions is that an estimate is significantly different from zero if 95% of its bootstrap distribution is greater or less than zero. The fraction of the Bayesian bootstrap distribution that is greater or less than zero.

In both cases, I bootstrapped across molecular markers and across individuals for each of 10,000 replicates. Bootstrapping across markers addresses error in individual estimates of $P_{CC}$, $P_{CT}$ and $P_{TT}$ due to sampling only a fraction of each salamander's genome. This is in the same spirit as bootstrapping across characters in phylogenetic analysis (Felsenstein, 1985) and results in each individual having associated with it a distribution (bootstrap or posterior) rather than a single-point estimate of the summary statistics $P_{ij}$. Bootstrapping across individuals addresses errors caused by sampling a finite number of salamanders from each population. As each standard or Bayesian bootstrap replicate includes both sampling processes (sampling markers and sampling individuals) the resulting distributions incorporate uncertainty from both kinds of sampling error. The typical frequentist interpretation of nonparametric bootstrap distributions is that an estimate is significantly different from zero if 95% of its bootstrap distribution is greater or less than zero. The fraction of the Bayesian bootstrap distribution that is greater than zero is taken as an estimate of the posterior probability that the parameter is greater than zero relative to an uninformative prior (i.e. that the parameter is equally likely to be greater or less than zero).

**Graphical interpretation**

Beyond comparing the fit of the Dobzhansky–Muller model (models 1 and 2) vs. heterozygote disadvantage (models 3 and 4), more detailed inferences about the genetics of hybrid dysfunction can be obtained from the relative magnitudes of the fitted parameters (hence the importance of rigorous evaluation of confidence intervals). The easiest way to see this is by comparing performance surfaces. First, it is important to note that the space of possible genotypes described by genomic proportions is triangular. Because $P_{CC} + P_{CT} + P_{TT} = 1$, these proportions define a ternary coordinate system. The same triangular space can be described by $P_T$ and $P_{CT}$: pure native ($P_T = 0$) and pure introduced ($P_T = 1$) individuals have zero heterozygosity, individuals such as F1 hybrids with $P_{CT} = 1$ must have $P_T = 0.5$, but individuals with $P_T = 0.5$ may have any value of heterozygosity. Plotting performance or fitness as a surface over this triangular genotype space allows us to visualize different models of hybrid fitness (Fig. 1). If fitness or performance is a linear function of ancestry or heterozygosity, then the surface is a triangular section of a plane (Fig. 1a–c). If dysfunction is caused by incompatibilities, the surface is curved (Fig. 1d–f). If incompatibilities are recessive ($h_1$ and $h_0 = 0$ in eqn 1), only recombinant hybrids with $P_{CT} \neq 0$ and $P_{CC} \neq 0$ express dysfunction (Fig. 1d). If incompatibilities are additive ($h_0 = 1/2h_1 = 1/2h_1$), then all hybrids with intermediate ancestry have the same expected breakdown (Fig. 1e). This occurs because even though heterozygous incompatibilities are less severe than homozygous incompatibilities, maximally recombinant homozygotes ($P_{CT} = 0$, $P_{CC} = P_{TT} = 0.5$, so maximum $P_{TT}P_{CC} = 0.25$) express, on average, only 1/4 of the incompatibilities expressed in the F1 ($P_{CT} = P_{CT}^2 = 1$). By the same logic, F1 hybrids will have the greatest expected dysfunction if incompatibilities are dominant ($h_0 = h_1 = h_2$) because they fully express all possible incompatibilities, but curvature of the surface still distinguishes this case from simple heterozygote disadvantage (Fig. 1c,f).

**Results**

Model 1, the simple Dobzhansky–Muller model (including terms for SVL and source pond), was the best description of the data according to AIC (Table 1). Adding an extra term for an additive species difference (model 2) did not significantly increase the likelihood, and models 3 and 4, invoking heterozygote disadvantage rather than interlocus incompatibilities, had substantially lower likelihoods (Table 1). According to the $\Delta_1$ statistics in Table 1, the models invoking heterozygote disadvantage (3 and 4) had significantly less support than the models invoking epistasis (1 and 2). Models 5 and 6, with quadratic functions of heterozygosity, were also unsupported (Table 1). Table 2 and Fig. 2 illustrate the fit of the Dobzhansky–Muller model to the data. The fitted performance surface is most consistent with additivity of Dobzhansky–Muller incompatibilities (Figs 1e and 2d). The $B_i$ estimates in Table 2 support this interpretation: $B_i \approx 1/2B_1 \approx 1/4B_2$. However, the estimated $B_0$ is not statistically distinguishable from zero (with 5% type I error), therefore we cannot formally reject a recessive Dobzhansky–Muller model where heterozygous–heterozygous incompatibilities cause no dysfunction. Nevertheless, almost 95% of the posterior (Bayesian bootstrap) distribution of $B_0$ is greater than zero, leading me to prefer the hypothesis that $B_0 > 0$ if there is no a priori reason to consider complete recessivity the preferred or null hypothesis. An alternative Bayesian perspective on recessivity vs. dominance of incompatibilities is to compare the fraction of the posterior density with $B_0/B_1$ greater or less than 1/2. From the Bayesian bootstrap distribution, the fractions are 0.43 and 0.57, respectively, providing little reason to prefer either hypothesis.
There was no statistically significant variation among source ponds not accounted for by variation among genotypes (Table 2). The relationship between SVL and maximum burst speed was significant (longer larvae tended to be faster, Table 2), however, leaving SVL out of the model did not greatly affect the parameter estimates for the genetic variables ($B_2 = 0.0414$, $B_1 = 0.0188$, $B_0 = 0.0103$), nor their statistical significance.

Bootstrapping over markers only or over individuals only allowed an assessment of the relative influence of sampling markers vs. individuals on overall standard errors. Standard errors of the coefficients among bootstrap replicates when only markers were resampled (SE$_K$; Table 2) were much smaller than standard errors among bootstrap replicates when only individuals were resampled (SE$_N$; Table 2). Furthermore, bootstrap standard errors when both markers and individuals were resampled were not much different from the individuals-only standard errors (Table 2). The similarity between inferences from parametric $P$-values (which do not account for error caused by sampling markers) and bootstrap support values (Table 2) also indicate that the sampling error caused by using a few markers to estimate genome-wide summary statistics is not very large for this data set.

There were no significant associations between burst speed and individual marker genotypes. Separate one-way ANOVAs of burst speed among genotypes never supported rejecting the null hypothesis (with $\alpha = 0.05$).

![Fig. 1 Hypothetical performance surfaces on the ternary coordinate system defined by the proportion of the genome homozygous for *Ambystoma tigrinum maverum* alleles ($P_{TT}$), the proportion homozygous for *A. californiense* alleles ($P_{CC}$), and the proportion of the genome that is heterozygous ($P_{CT}$).](image)

Table 1 Comparison of alternative models for the genetics of variation in burst speed performance among hybrid tiger salamander larvae.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>$F_{DF}$</th>
<th>$P$</th>
<th>Log-likelihood</th>
<th>AIC</th>
<th>$\Delta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3419</td>
<td>6.6067</td>
<td>&lt;0.0001</td>
<td>379.498</td>
<td>-740.996</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.3476</td>
<td>5.8609</td>
<td>&lt;0.0001</td>
<td>379.917</td>
<td>-739.833</td>
<td>1.115</td>
</tr>
<tr>
<td>3</td>
<td>0.2262</td>
<td>5.6285</td>
<td>0.00015</td>
<td>372.271</td>
<td>-730.542</td>
<td>10.454</td>
</tr>
<tr>
<td>4</td>
<td>0.2365</td>
<td>4.6459</td>
<td>0.00037</td>
<td>372.289</td>
<td>-728.578</td>
<td>12.418</td>
</tr>
<tr>
<td>5</td>
<td>0.2525</td>
<td>5.0669</td>
<td>0.00016</td>
<td>373.317</td>
<td>-730.633</td>
<td>10.363</td>
</tr>
<tr>
<td>6</td>
<td>0.2534</td>
<td>4.3187</td>
<td>0.00037</td>
<td>373.378</td>
<td>-728.757</td>
<td>12.239</td>
</tr>
</tbody>
</table>

$F$-statistics and associated $P$-values test each model against the null model of no association between maximum burst speed and any independent variable. AIC is the Akaike information criterion; a lower (i.e. more negative) AIC represents a better fit for the data. $\Delta_i$ is the rescaled AIC ($AIC_i$ of model $i - AIC$ of best model in the set); models with $\Delta_i \leq 2$ have substantial support and models with $\Delta_i > 10$ have essentially no support (Burnham & Anderson, 2004).
that CC homozygotes, TT homozygotes and CT heterozygotes had the same average burst speed (detailed results available from the author). Taken together with the nonparametric bootstrap results, this shows that the patterns in Table 2 and Fig. 2 are unlikely to be driven by a single locus. Rather, genotypes for a single marker are uninformative outside the context of the multilocus genotype.

**Discussion**

My analyses show that the Dobzhansky–Muller model explains slow burst speed of hybrid tiger salamander larvae better than heterozygote disadvantage. Though rejection of heterozygote disadvantage may not be surprising in the light of existing data (Coyne & Orr, 2004), its theoretical plausibility (Nei et al., 1983) means that it cannot be dropped from consideration. The Dobzhansky–Muller model has been studied primarily in the context of hybrid sterility and inviability of model organisms such as *Drosophila* (Coyne & Orr, 2004). My approach in this study illustrates how the Dobzhansky–Muller model can be evaluated in nonmodel systems and applied to sub-lethal and condition-dependent hybrid dysfunction in addition to sterility and inviability. The power of the Dobzhansky–Muller model to account for...
poor swimming performance in hybrid tiger salamander larvae strongly supports the generality of ‘conspecific epistasis’ as an important outcome of genome evolution and cause of reproductive isolation (Orr & Presgraves, 2000).

The apparent additivity of incompatibilities contributing to hybrid dysfunction in burst speed is somewhat at odds with previous data and theory on the Dobzhansky–Muller model. Inviability and sterility are thought to result most often from recessive or partially recessive incompatibilities in Drosophila and Lycoserpionis (Presgraves, 2003; Tao & Hardl, 2003; Coyne & Orr, 2004; Moyle & Graham, 2005), though dominant incompatibilities are not unheard of (Brideau et al., 2006; Sweigart et al., 2006) and Willett (2006) showed complex patterns of additive × dominance epistasis among three electron transport loci in Tigropus. Recessivity of incompatibilities is an important part of the dominance theory explanation of Haldane’s rule (the tendency for the heterogametic sex to suffer greater hybrid dysfunction in the F1 generation; Turelli & Orr, 2000). Tiger salamanders do not have degenerate sex chromosomes (Cuny & Malacinski, 1985), therefore they are not subject to Haldane’s rule. However, the importance of additive, recessive and dominant incompatibilities is general as it relates to the relative fitness of highly heterozygous genotypes such as F1 hybrids vs. highly recombinant genotypes with many homozygous–homozygous interactions (Fig. 1). As noted by Coyne & Orr (2004), the value of the dominance theory and other hypotheses is not simply in explaining Haldane’s rule per se, but in suggesting broad generalities about the evolution of hybrid dysfunction in very different kinds of organisms.

Two issues that may affect the reliability of inferences based on the statistical approach used in this study are the relationship between individual genotype and the Turelli–Orr hybrid breakdown score, and the relationship between the breakdown score and the phenotype of dysfunction. First, recall that $E(S)$ is the expected breakdown score for genotypes sharing the same $P_{T}$, $P_{C}$, and $P_{CT}$. The variance around that expectation depends on the diversity of genotypes sharing the same summary statistics and the distribution of effects of incompatibilities. The $h_i$ parameters in the Turelli–Orr model are average effects per pairwise incompatibility, and their use in calculating $E(S)$ may be justified only if their magnitudes and variances are small, and if most dysfunction is due to pairwise incompatibilities rather than more complex interactions (Turelli & Orr, 2000). The fact that hybrid dysfunction accumulates gradually with genetic divergence in most groups of animals studied (Coyne & Orr, 2004; Bolnick & Near, 2005) suggests that many incompatibilities of relatively small effect usually contribute to hybrid breakdown, and therefore Turelli & Orr’s (2000) $E(S)$ is probably valid. The relative importance of pairwise vs. more complex incompatibilities is unknown, but pairwise incompatibilities are probably most common in the early stages of divergence, whereas more complex incompatibilities are likely to arise as more differences accumulate (Orr, 1995).

Secondly, Turelli & Orr (2000) did not specify a function linking $E(S)$ to fitness or performance, requiring only that it be monotonic. I have reported results for the simplest, linear function. However, it is easy to fit an exponential or logistic relationship by log or logit transformations, respectively, of the dependent variable (for the logit, the variable should be first transformed so that it ranges from just over zero to just less than one). Applying those transformations to my data did not substantially affect the fit between model and data; the best fit in both cases was model 1 with $R^2 = 0.353$ for the exponential and $R^2 = 0.303$ for the logit. Further, in both cases the fitted performance surface was indistinguishable from Fig. 2d, and parameter estimates were consistent with $B_0 = 1/2B_1 \approx 1/4B_2$. Thus, at least for the range of genotypes and burst speeds in this study, the main conclusions are not sensitive to the function linking hybrid dysfunction to the Turelli–Orr breakdown score; the Dobzhansky–Muller model explains the data better than heterozygote disadvantage and incompatibilities appear to be additive, on average.

An alternative framework for studying the genetics of hybrid fitness was described by Lynch (1991); also see Hill, 1982; Lynch & Walsh, 1998; Rundle & Whitlock, 2001). Lynch related phenotypes of hybrids to the quantitative genetic concepts of additive effects, dominance, additive × additive epistasis, dominance × dominance epistasis and additive × dominance epistasis using a polynomial function of ancestry and heterozygosity. As shown in Appendix S1, the Turelli–Orr model can be expressed as a special case of Lynch’s model and estimates of $B_i$ (eqn 2) calculated from fitted coefficients of Lynch’s model. Comparison of the Lynch and Turelli–Orr frameworks leads to two interesting insights. First, when Dobzhansky–Muller incompatibilities are additive, hybrid dysfunction is attributed to additive × additive epistasis in Lynch’s framework. Secondly, the ‘dominance effect’ in Lynch’s framework is affected by between-locus incompatibilities unless incompatibilities are exactly additive, i.e. ‘dominance’ is not necessarily a simple combination of single-locus dominance effects. In any case, analysis of my data using Lynch’s framework agrees with the results in Table 2 (Appendix S1).

I have illustrated a simple approach to estimating parameters of the Turelli–Orr formulation of the Dobzhansky–Muller model using variables estimable from molecular marker data from natural hybrids. This allows the study of the genetics of hybrid dysfunction to extend to nonmodel organisms and components of dysfunction other than inviability and infertility. My results from tiger salamander burst-speed performance support the Dobzhansky–Muller model of genic incompatibilities as the cause of hybrid dysfunction (or at least abnormal trait expression) and suggest that most incompatibilities are...
References


Supplementary material

The following supplementary material is available for this article:

Appendix S1 The Tuvelli–Orr model as a special case of Lynch’s model of hybrid fitness. This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1402-9101.2007.01448.x

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 28 February 2007; revised 24 July 2007; accepted 27 August 2007

Performance in hybrid salamanders 351