

# EVOLUTION OF QUANTITATIVE TRAITS

## Introduction

Let's stop and review quickly where we've come and where we're going. We started our survey of quantitative genetics by pointing out that our objective was to develop a way to describe the patterns of phenotypic resemblance among relatives. The challenge was that we wanted to do this for phenotypic traits that whose expression is influenced both by many genes and by the environment in which those genes are expressed. Beyond the technical, algebraic challenges associated with many genes, we have the problem that we can't directly associate particular genotypes with particular phenotypes. We have to rely on patterns of *phenotypic* resemblance to tell us something about how *genetic* variation is transmitted. Surprisingly, we've managed to do that. We now know that it's possible to:

- Estimate the additive effect of an allele.<sup>1</sup>
- Partition the phenotypic variance into genotypic and environmental components and to partition the genotypic variance into additive and dominance components.<sup>2</sup>
- Estimate all of the variance components from a combination of appropriate crossing designs and appropriate statistical analyses.

Now we're ready for the next step: applying all of these ideas to the evolution of a quantitative trait.

---

<sup>1</sup>Actually, we don't know this. You'll have to take my word for it that in certain breeding designs its possible to estimate not only the additive genetic variance and the dominance genetic variance, but also the actual additive effect of "alleles" that we haven't even identified. We'll see a more direct approach soon, when we get to genome-wide associations studies.

<sup>2</sup>I should point out that this is an oversimplification. I've mentioned that we typically assume that we can simply add the effects of alleles across loci, but if you think about how genes actually work in organisms, you realize that such additivity across loci isn't likely to be very common. Strictly speaking there are epistatic components to the genetic variance too, i.e., components of the genetic variance that have to do not with the interaction among alleles at a single locus (the dominance variance that we've already encountered), but with the interaction of alleles at different loci.

## Evolution of the mean phenotype

We're going to focus on how the mean phenotype in a population changes in response to natural selection, specifically in response to viability selection. Before we can do this, however, we need to think a bit more carefully about the relationship between genotype, phenotype, and fitness. Let  $F_{ij}(x)$  be the probability that genotype  $A_iA_j$  has a phenotype smaller than  $x$ .<sup>3</sup> Then  $x_{ij}$ , the genotypic value of  $A_iA_j$  is

$$x_{ij} = \int_{-\infty}^{\infty} x dF_{ij}(x)$$

and the population mean phenotype is  $p^2x_{11} + 2pqx_{12} + q^2x_{22}$ .<sup>4</sup> If an individual with phenotype  $x$  has fitness  $w(x)$ , then the fitness of an individual with genotype  $A_iA_j$  is

$$w_{ij} = \int_{-\infty}^{\infty} w(x) dF_{ij}(x)$$

and the mean fitness in the population is  $\bar{w} = p^2w_{11} + 2pqw_{12} + q^2w_{22}$ .

Now, there's a well known theorem from calculus known as Taylor's theorem. It says that for any function<sup>5</sup>  $f(x)$

$$f(x) = f(a) + \sum_{k=1}^{\infty} \left( \frac{(x-a)^k}{k!} \right) f^{(k)}(a) \quad .,$$

where  $f^{(k)}(a)$  is the  $k$ th derivative of  $f(x)$  evaluated at  $a$ .<sup>6</sup> Using this theorem we can produce an approximate expression describing how the mean phenotype in a population will change in response to selection. Remember that the mean phenotype,  $\bar{x}$ , depends both on the underlying genotypic values and on the allele frequency. So I'm going to write the mean phenotype as  $\bar{x}(p)$  to remind us of that dependency. The phenotype changes from one generation to the next as a result of changes in the frequency of alleles that influence the

---

<sup>3</sup>For those of you who have had probability theory,  $F_{ij}(x)$  is the cumulative distribution for the probability density for phenotype associated with  $A_iA_j$ .

<sup>4</sup>I know that the integral looks horrible, but that horrible looking integral is just a mathematical way of saying what I've repeated for the last several lectures: The genotypic value for a genotype is just the mean phenotype associated with that genotype.

<sup>5</sup>Actually there are restrictions on the functions to which it applies, but we can ignore those restrictions for our purposes.

<sup>6</sup>If that mumbo jumbo doesn't mean anything to you, don't worry about it. Just remember that what we're about to do relies on an approximation and that the approximation works best with the difference between  $s$  and  $a$  is small.

phenotype, assuming that the environmental effects on phenotypes don't change.

$$\begin{aligned}\bar{x}(p') &= \bar{x}(p) + (p' - p) \left( \frac{d\bar{x}}{dp} \right) + o(p^2) \\ \bar{x}(p) &= p^2 x_{11} + 2pqx_{12} + q^2 x_{22} \\ \frac{d\bar{x}(p)}{dp} &= 2px_{11} + 2qx_{12} - 2px_{12} - 2qx_{22} \\ &= 2 \{ (px_{11} + qx_{12} - \bar{x}/2) - (px_{12} + qx_{22} - \bar{x}/2) \} \\ &= 2(\alpha_1 - \alpha_2) \\ \bar{x}(p') &\approx \bar{x}(p) + (p' - p) (2(\alpha_1 - \alpha_2)) \\ \Delta\bar{x} &= (\Delta p) (2(\alpha_1 - \alpha_2))\end{aligned}$$

In other words, the change in mean phenotype from one generation to the next depends first on how much the frequency of the  $A_1$  allele changes and second on the difference between the additive effect of  $A_1$  and  $A_2$ .

Now you need to remember (from lo those many weeks ago) that

$$p' = \frac{p^2 w_{11} + pqw_{12}}{\bar{w}} \quad .$$

Thus,

$$\begin{aligned}\Delta p &= p' - p \\ &= \frac{p^2 w_{11} + pqw_{12}}{\bar{w}} - p \\ &= \frac{p^2 w_{11} + pqw_{12} - p\bar{w}}{\bar{w}} \\ &= p \left( \frac{pw_{11} + qw_{12} - \bar{w}}{\bar{w}} \right) \quad .\end{aligned}$$

Now,<sup>7</sup> let's do a linear regression of fitness on phenotype. After all, to make any further progress, we need to relate phenotype to fitness, so that we can use the relationship between phenotype and genotype to infer the change in allele frequencies, from which we will infer

---

<sup>7</sup>Since we're having so much fun with mathematics why should we stop here?

the change in mean phenotype.<sup>8</sup> From our vast statistical knowledge, we know that the slope of this regression line is

$$\beta_1 = \frac{\text{Cov}(w, x)}{\text{Var}(x)}$$

and its intercept is

$$\beta_0 = \bar{w} - \beta_1 \bar{x} \quad .$$

Let's use this regression equation to determine the fitness of each genotype. This is only an approximation to the true fitness,<sup>9</sup> but it is adequate for many purposes.

$$\begin{aligned} w_{ij} &= \int_{-\infty}^{\infty} w(x) dF_{ij}(x) \\ &\approx \int_{-\infty}^{\infty} (\beta_0 + \beta_1 x) dF_{ij}(x) \\ &= \beta_0 + \beta_1 x_{ij} \\ \bar{w} &= \beta_0 + \beta_1 \bar{x} \quad . \end{aligned}$$

If we substitute this into our expression for  $\Delta p$  above, we get

$$\begin{aligned} \Delta p &= p \left( \frac{pw_{11} + qw_{12} - \bar{w}}{\bar{w}} \right) \\ &= p \left( \frac{p(\beta_0 + \beta_1 x_{11}) + q(\beta_0 + \beta_1 x_{12}) - (\beta_0 + \beta_1 \bar{x})}{\bar{w}} \right) \\ &= p\beta_1 \left( \frac{px_{11} + qx_{12} - \bar{x}}{\bar{w}} \right) \\ &= p\beta_1 \left( \frac{\alpha_1 - \bar{x}/2}{\bar{w}} \right) \\ &= p\beta_1 \left( \frac{\alpha_1 - (p\alpha_1 + q\alpha_2)}{\bar{w}} \right) \\ &= \frac{pq\beta_1(\alpha_1 - \alpha_2)}{\bar{w}} \quad . \end{aligned}$$

So where are we now?<sup>10</sup> Let's substitute this result back into the equation for  $\Delta \bar{x}$ . When we do we get

$$\Delta \bar{x} = (\Delta p) (2(\alpha_1 - \alpha_2))$$

---

<sup>8</sup>Whew! That was a mouthful.

<sup>9</sup>Specifically, we are implicitly assuming that the fitnesses are adequately approximated by a linear function of our phenotypic measure.

<sup>10</sup>You don't have to tell me where you *wish* you were. I can reliably guess that it's not here.

$$\begin{aligned}
&= \left( \frac{pq\beta_1(\alpha_1 - \alpha_2)}{\bar{w}} \right) (2(\alpha_1 - \alpha_2)) \\
&= 2pq\alpha^2 \left( \frac{\beta_1}{\bar{w}} \right) \\
&= V_a \left( \frac{\beta_1}{\bar{w}} \right) .
\end{aligned}$$

This is great if we've done the regression between fitness and phenotype, but what if we haven't?<sup>11</sup> Let's look at  $\text{Cov}(w, x)$  in a little more detail.

$$\begin{aligned}
\text{Cov}(w, x) &= p^2 \int_{-\infty}^{\infty} xw(x) dF_{11}(x) + 2pq \int_{-\infty}^{\infty} xw(x) dF_{12}(x) \\
&\quad + q^2 \int_{-\infty}^{\infty} xw(x) dF_{22}(x) - \bar{x}\bar{w} \\
&= p^2 \left( \int_{-\infty}^{\infty} xw(x) dF_{11}(x) - x_{11}\bar{w} + x_{11}\bar{w} \right) \\
&\quad + 2pq \left( \int_{-\infty}^{\infty} xw(x) dF_{11}(x) - x_{12}\bar{w} + x_{12}\bar{w} \right) \\
&\quad + q^2 \left( \int_{-\infty}^{\infty} xw(x) dF_{22}(x) - x_{22}\bar{w} + x_{22}\bar{w} \right) \\
&\quad - \bar{x}\bar{w} \\
&= p^2 \left( \int_{-\infty}^{\infty} xw(x) dF_{11}(x) - x_{11}\bar{w} \right) \\
&\quad + 2pq \left( \int_{-\infty}^{\infty} xw(x) dF_{11}(x) - x_{12}\bar{w} \right) \\
&\quad + q^2 \left( \int_{-\infty}^{\infty} xw(x) dF_{22}(x) - x_{22}\bar{w} \right) .
\end{aligned}$$

Now

$$\begin{aligned}
\int_{-\infty}^{\infty} xw(x) dF_{ij}(x) - x_{ij}\bar{w} &= \bar{w} \left( \int_{-\infty}^{\infty} \frac{xw(x)}{\bar{w}} dF_{ij}(x) - x_{ij} \right) \\
&= \bar{w}(x_{ij}^* - x_{ij}) ,
\end{aligned}$$

where  $x_{ij}^*$  refers to the mean phenotype of  $A_i A_j$  after selection. So

$$\begin{aligned}
\text{Cov}(w, x) &= p^2 \bar{w}(x_{11}^* - x_{11}) + 2pq \bar{w}(x_{12}^* - x_{12}) + q^2 \bar{w}(x_{22}^* - x_{22}) \\
&= \bar{w}(\bar{x}^* - \bar{x}) ,
\end{aligned}$$

---

<sup>11</sup>Hang on just a little while longer. We're almost there.

Genotype	$A_1A_1$	$A_1A_2$	$A_2A_2$
Phenotype	1.303	1.249	0.948

Table 1: A simple example to illustrate response to selection in a quantitative trait.

where  $\bar{x}^*$  is the population mean phenotype after selection. In short,<sup>12</sup> combining our equations for the change in mean phenotype and for the covariance of fitness and phenotype and remembering that  $\beta_1 = \text{Cov}(w, x)/\text{Var}(x)$ <sup>13</sup>

$$\begin{aligned}\Delta\bar{x} &= V_a \left( \frac{\bar{w}(\bar{x}^* - \bar{x})}{\bar{w} V_p} \right) \\ &= h_N^2 (\bar{x}^* - \bar{x})\end{aligned}$$

$\Delta\bar{x} = \bar{x}' - \bar{x}$  is referred to as the response to selection and is often given the symbol  $R$ . It is the change in population mean between the parental generation (before selection) and the offspring generation (before selection).  $\bar{x}^* - \bar{x}$  is referred to as the selection differential and is often given the symbol  $S$ . It is the difference between the mean phenotype in the parental generation before selection and the mean phenotype in the parental generation after selection. Thus, we can rewrite our final equation as

$$R = h_N^2 S \quad .$$

This equation is often referred to as the *breeders equation*.

## A Numerical Example

To illustrate how this works, let's examine the simple example in Table 1.

Given these phenotypes,  $p = 0.25$ , and  $V_p = 0.16$ , it follows that  $\bar{x} = 1.08$  and  $h_N^2 = 0.1342$ . Suppose the mean phenotype after selection is 1.544. What will the phenotype be among the newly born progeny?

$$\begin{aligned}S &= \bar{x}^* - \bar{x} \\ &= 1.544 - 1.08 \\ &= 0.464\end{aligned}$$

---

<sup>12</sup>We finally made it.

<sup>13</sup>You also need to remember that  $\text{Var}(x) = V_p$ , since they're the same thing, the phenotypic variance.

Genotype	$A_1A_1$	$A_1A_2$	$A_2A_2$
Frequency	$p^2$	$2pq$	$q^2$
Fitness	$w_{11}$	$w_{12}$	$w_{22}$
Additive fitness value	$2\alpha_1$	$\alpha_1 + \alpha_2$	$2\alpha_2$

Table 2: Fitnesses and additive fitness values used in deriving Fisher's Fundamental Theorem of Natural Selection.

$$\begin{aligned}
\Delta\bar{x} &= h_N^2 S \\
&= (0.1342)(0.464) \\
&= 0.06 \\
\bar{x}' &= \bar{x} + \Delta\bar{x} \\
&= 1.08 + 0.06 \\
&= 1.14
\end{aligned}$$

## Fisher's Fundamental Theorem of Natural Selection

Suppose the phenotype whose evolution we're interested in following is fitness itself.<sup>14</sup> Then we can summarize the fitnesses as illustrated in Table 2.

Although I didn't tell you this, a well-known fact about viability selection at one locus is that the change in allele frequency from one generation to the next can be written as

$$\Delta p = \left(\frac{pq}{2\bar{w}}\right) \left(\frac{d\bar{w}}{dp}\right) \quad .$$

Using our new friend, Taylor's theorem, it follows immediately that

$$\bar{w}' = \bar{w} + (\Delta p) \left(\frac{d\bar{w}}{dp}\right) + \left(\frac{(\Delta p)^2}{2}\right) \left(\frac{d^2\bar{w}}{dp^2}\right) \quad .$$

Or, equivalently

$$\Delta\bar{w} = (\Delta p) \left(\frac{d\bar{w}}{dp}\right) + \left(\frac{(\Delta p)^2}{2}\right) \left(\frac{d^2\bar{w}}{dp^2}\right) \quad .$$

---

<sup>14</sup>The proof of the fundamental theorem that follows is due to C. C. Li [?]

Recalling that  $\bar{w} = p^2w_{11} + 2p(1-p)w_{12} + (1-p)^2w_{22}$  we find that

$$\begin{aligned}
\frac{d\bar{w}}{dp} &= 2pw_{11} + 2(1-p)w_{12} - 2pw_{12} - 2(1-p)w_{22} \\
&= 2[(pw_{11} + qw_{12}) - (pw_{12} + qw_{22})] \\
&= 2[(pw_{11} + qw_{12} - \bar{w}/2) - (pw_{12} + qw_{22} - \bar{w}/2)] \\
&= 2[\alpha_1 - \alpha_2] \\
&= 2\alpha \quad ,
\end{aligned}$$

where the last two steps use the definitions for  $\alpha_1$  and  $\alpha_2$ , and we set  $\alpha = \alpha_1 - \alpha_2$ . Similarly,

$$\begin{aligned}
\frac{d^2\bar{w}}{dp^2} &= 2w_{11} - 2w_{12} - 2w_{12} + 2w_{22} \\
&= 2(w_{11} - 2w_{12} + w_{22})
\end{aligned}$$

Now we can plug these back into the equation for  $\Delta\bar{w}$ :

$$\begin{aligned}
\Delta\bar{w} &= \left\{ \left( \frac{pq}{2\bar{w}} \right) \left( \frac{d\bar{w}}{dp} \right) \right\} \left( \frac{d\bar{w}}{dp} \right) + \frac{\left\{ \left( \frac{pq}{2\bar{w}} \right) \left( \frac{d\bar{w}}{dp} \right) \right\}^2}{2} [2(w_{11} - 2w_{12} + w_{22})] \\
&= \left\{ \left( \frac{pq}{2\bar{w}} \right) (2\alpha) \right\} (2\alpha) + \left\{ \left( \frac{pq}{2\bar{w}} \right) (2\alpha) \right\}^2 (w_{11} - 2w_{12} + w_{22}) \\
&= \frac{2pq\alpha^2}{\bar{w}} + \frac{p^2q^2\alpha^2}{\bar{w}^2} (w_{11} - 2w_{12} + w_{22}) \\
&= \frac{V_a}{\bar{w}} \left\{ 1 + \frac{pq}{2\bar{w}} (w_{11} - 2w_{12} + w_{22}) \right\} \quad ,
\end{aligned}$$

where the last step follows from the observation that  $V_a = 2pq\alpha^2$ . The quantity  $\frac{pq}{2\bar{w}}(w_{11} - 2w_{12} + w_{22})$  is usually quite small, especially if selection is not too intense.<sup>15</sup> So we are left with

$$\Delta\bar{w} \approx \frac{V_a}{\bar{w}} \quad .$$

## Selection on multiple traits

So far we've studied only the evolution of a single trait, e.g., height or weight. But organisms have many traits, and they evolve at the same time. How can we understand their

---

<sup>15</sup>Notice that it's exactly equal to 0 if the fitness of the heterozygote is exactly intermediate. In that case, all of the variance in fitness is additive.

simultaneous evolution? The basic framework of the quantitative genetic approach was first outlined by Russ Lande and Steve Arnold [?].

Let  $z_1, z_2, \dots, z_n$  be the phenotype of each character that we are studying. We'll use  $\bar{\mathbf{z}}$  to denote the vector of these characters before selection and  $\bar{\mathbf{z}}^*$  to denote the vector after selection. The selection differential,  $\mathbf{s}$ , is also a vector given by

$$\mathbf{s} = \bar{\mathbf{z}}^* - \bar{\mathbf{z}} \quad .$$

Suppose  $p(\mathbf{z})$  is the probability that any individual has phenotype  $\mathbf{z}$ , and let  $W(\mathbf{z})$  be the fitness (absolute viability) of an individual with phenotype  $\mathbf{z}$ . Then the mean absolute fitness is

$$\bar{W} = \int W(\mathbf{z})p(\mathbf{z})d\mathbf{z} \quad .$$

The relative fitness of an individual with phenotype  $\mathbf{z}$  can be written as

$$w(\mathbf{z}) = \frac{W(\mathbf{z})}{\bar{W}} \quad .$$

Using relative fitnesses the mean relative fitness,  $\bar{w}$ , is 1. Now

$$\bar{\mathbf{z}}^* = \int \mathbf{z}w(\mathbf{z})p(\mathbf{z})d\mathbf{z} \quad .$$

Recall that  $Cov(X, Y) = E(X - \mu_x)(Y - \mu_y) = E(XY) - \mu_x\mu_y$ . Consider

$$\begin{aligned} \mathbf{s} &= \bar{\mathbf{z}}^* - \bar{\mathbf{z}} \\ &= \int \mathbf{z}w(\mathbf{z})p(\mathbf{z})d\mathbf{z} - \bar{\mathbf{z}} \\ &= E(w, \mathbf{z}) - \bar{w}\bar{\mathbf{z}} \quad , \end{aligned}$$

where the last step follows since  $\bar{w} = 1$  meaning that  $\bar{w}\bar{\mathbf{z}} = \bar{\mathbf{z}}$ . In short,

$$\mathbf{s} = Cov(w, \mathbf{z}) \quad .$$

That should look familiar from our analysis of the evolution of a single phenotype.

If we assume that all genetic effects are additive, then the phenotype of an individual can be written as

$$\mathbf{z} = \mathbf{x} + \mathbf{e} \quad ,$$

where  $\mathbf{x}$  is the additive genotype and  $\mathbf{e}$  is the environmental effect. We'll denote by  $\mathbf{G}$  the matrix of genetic variances and covariances and by  $\mathbf{E}$  the matrix of environmental variances and covariances. The matrix of phenotype variances and covariances,  $\mathbf{P}$ , is then given by<sup>16</sup>

$$\mathbf{P} = \mathbf{G} + \mathbf{E} \quad .$$

---

<sup>16</sup>Assuming that there are no genotype  $\times$  environment interactions.

Now, if we're willing to assume that the regression of additive genetic effects on phenotype is linear<sup>17</sup> and that the environmental variance is the same for every genotype, then we can predict how phenotypes will change from one generation to the next

$$\begin{aligned}\bar{\mathbf{x}}^* - \bar{\mathbf{x}} &= \mathbf{GP}^{-1}(\bar{\mathbf{z}}^* - \bar{\mathbf{z}}) \\ \bar{\mathbf{z}}' - \bar{\mathbf{z}} &= \mathbf{GP}^{-1}(\bar{\mathbf{z}}^* - \bar{\mathbf{z}}) \\ \Delta\bar{\mathbf{z}} &= \mathbf{GP}^{-1}\mathbf{s}\end{aligned}$$

$\mathbf{GP}^{-1}$  is the multivariate version of  $h_N^2$ . This equation is also the multivariate version of the breeders equation.

But we have already seen that  $\mathbf{s} = Cov(w, z)$ . Thus,

$$\beta = \mathbf{P}^{-1}\mathbf{s}$$

is a set of partial regression coefficients of relative fitness on the characters, i.e., the dependence of relative fitness on that character alone holding all others constant.

Note:

$$\begin{aligned}s_i &= \sum_{j=1}^n \beta_j P_{ij} \\ &= \beta_1 P_{i1} + \cdots + \beta_i P_{ii} + \cdots + \beta_n P_{in}\end{aligned}$$

is the total selective differential in character  $i$ , including the indirect effects of selection on other characters.

## An example: selection in a pentastomid bug

94 individuals were collected along shoreline of Lake Michigan in Parker County, Indiana after a storm. 39 were alive, 55 dead. The means of several characters before selection, the trait correlations, and the selection analysis are presented in Table 3.

The column labeled  $s$  is the selective differential for each character. The column labeled  $s'$  is the *standardized* selective differential, i.e., the change measured in units of standard deviation rather than on the original scale.<sup>18</sup> A multiple regression analysis of fitness versus phenotype on the original scale gives estimates of  $\beta$ , the direct effect of selection on that

---

<sup>17</sup>And we were willing to do this when we were studying the evolution of only one trait, so why not do it now?

<sup>18</sup>To measure on this scale the data is simply transformed by setting  $y_i = (x_i - \bar{x})/s$ , where  $x_i$  is the raw score for the  $i$ th individual,  $\bar{x}$  is the sample mean for the trait, and  $s$  is its standard deviation.

Character	Mean before selection	standard deviation		
head	0.880	0.034		
thorax	2.038	0.049		
scutellum	1.526	0.057		
wing	2.337	0.043		

  

	head	thorax	scutellum	wing
head	1.00	0.72	0.50	0.60
thorax		1.00	0.59	0.71
scutellum			1.00	0.62
wing				1.00

  

Character	$s$	$s'$	$\beta$	$\beta'$
head	-0.004	-0.11	$-0.7 \pm 4.9$	$-0.03 \pm 0.17$
thorax	-0.003	-0.06	$11.6 \pm 3.9^{**}$	$0.58 \pm 0.19^{**}$
scutellum	-0.16*	-0.28*	$-2.8 \pm 2.7$	$-0.17 \pm 0.15$
wing	-0.019**	-0.43**	$-16.6 \pm 4.0^{**}$	$-0.74 \pm 0.18^{**}$

Table 3: Selection analysis of pentastomid bugs on the shores of Lake Michigan.

trait. A multiple regression analysis of fitness versus phenotype on the transformed scale gives the standardized direct effect of selection,  $\beta'$ , on that trait.

Notice that the selective differential<sup>19</sup> for the thorax measurement is negative, i.e., individuals that survived had smaller thoraces than those that died. But the *direct* effect of selection on thorax is strongly positive, i.e., all other things being equal, an individual with a large was more likely to survive than one with a small thorax. Why the apparent contradiction? Because the thorax measurement is positively correlated with the wing measurement, and there's strong selection for decreased values of the wing measurement.

## Cumulative selection gradients

Arnold [?] suggested an extension of this approach to longer evolutionary time scales. Specifically, he studied variation in the number of body vertebrae and the number of tail vertebrae in populations of *Thamnophis elegans* from two regions of central California. He found relatively little vertebral variation within populations, but there were considerable differences

<sup>19</sup>The cumulative effect of selection on the change in mean phenotype.

	body	tail
body	35.4606	11.3530
tail	11.3530	37.2973

Table 4: Genetic variance-covariance matrix for vertebral number in central Californian garter snakes.

in vertebral number between populations on the coast side of the Coast Ranges and populations on the Central Valley side of the Coast Ranges. The consistent difference suggested that selection might have produced these differences, and Arnold attempted to determine the amount of selection necessary to produce these differences.

## The data

Arnold collected pregnant females from two local populations in each of two sites in northern California 282 km apart from one another. Females were collected over a ten-year period and returned to the University of Chicago. Dam-offspring regressions were used to estimate additive genetic variances and covariances of vertebral number.<sup>20</sup> Mark-release-recapture experiments in the California populations showed that females with intermediate numbers of vertebrae grow at the fastest rate, at least at the inland site, although no such relationship was found in males. The genetic variance-covariance matrix he obtained is shown in Table 4.

## The method

We know from Lande and Arnold’s results that the change in multivariate phenotype from one generation to the next,  $\Delta\bar{\mathbf{z}}$ , can be written as

$$\Delta\bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta} \quad ,$$

where  $\mathbf{G}$  is the genotypic variance-covariance matrix,  $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{s}$  is the set of partial regression coefficients describing the direct effect of each character on relative fitness.<sup>21</sup> If we are willing to assume that  $\mathbf{G}$  remains constant, then the total change in a character subject to selection for  $n$  generations is

$$\sum_{k=1}^n \Delta\bar{\mathbf{z}} = \mathbf{G} \sum_{k=1}^n \boldsymbol{\beta} \quad .$$

---

<sup>20</sup>1000 progeny from 100 dams.

<sup>21</sup> $\mathbf{P}$  is the phenotypic variance-covariance matrix and  $\mathbf{s}$  is the vector of selection differentials.

Thus,  $\sum_{k=1}^n \beta$  can be regarded as the cumulative selection differential associated with a particular observed change, and it can be estimated as

$$\sum_{k=1}^n \beta = \mathbf{G}^{-1} \sum_{k=1}^n \Delta \bar{\mathbf{z}} \quad .$$

## The results

The overall difference in vertebral number between inland and coastal populations can be summarized as:

$$\begin{aligned} \text{body}_{\text{inland}} - \text{body}_{\text{coastal}} &= 16.21 \\ \text{tail}_{\text{inland}} - \text{tail}_{\text{coastal}} &= 9.69 \end{aligned}$$

Given the estimate of  $\mathbf{G}$  already obtained, this corresponds to a cumulative selection gradient between inland and coastal populations of

$$\begin{aligned} \beta_{\text{body}} &= 0.414 \\ \beta_{\text{tail}} &= 0.134 \end{aligned}$$

Applying the same technique to looking at the differences between populations within the inland site and within the coastal site we find cumulative selection gradients of

$$\begin{aligned} \beta_{\text{body}} &= 0.035 \\ \beta_{\text{tail}} &= 0.038 \end{aligned}$$

for the coastal site and

$$\begin{aligned} \beta_{\text{body}} &= 0.035 \\ \beta_{\text{tail}} &= -0.004 \end{aligned}$$

for the inland site.

## The conclusions

“To account for divergence between inland and coastal California, we must invoke cumulative forces of selection that are 7 to 11 times stronger than the forces needed to account for differentiation of local populations.”

Furthermore, recall that the selection gradients can be used to partition the overall response to selection in a character into the portion due to the direct effects of that character alone and the portion due to the indirect effects of selection on a correlated character. In this case the overall response to selection in number of body vertebrae is given by

$$\mathbf{G}_{11}\beta_1 + \mathbf{G}_{12}\beta_2 \quad ,$$

where  $\mathbf{G}_{11}\beta_1$  is the direct effect of body vertebral number and  $\mathbf{G}_{12}\beta_2$  is the indirect effect of tail vertebral number. Similarly, the overall response to selection in number of tail vertebrae is given by

$$\mathbf{G}_{12}\beta_1 + \mathbf{G}_{22}\beta_2 \quad ,$$

where  $\mathbf{G}_{22}\beta_2$  is the direct effect of tail vertebral number and  $\mathbf{G}_{12}\beta_1$  is the indirect effect of body vertebral number. Using these equations it is straightforward to calculate that 91% of the total divergence in number of body vertebrae is a result of direct selection on this character. In contrast, only 51% of the total divergence in number of tail vertebrae is a result of direct selection on this character, i.e., 49% of the difference in number of tail vertebrae is attributable to indirect selection as a result of its correlation with number of body vertebrae.

## The caveats

While the approach Arnold suggests is intriguing, there are a number of caveats that must be kept in mind in trying to apply it.

- This approach assumes that the  $\mathbf{G}$  matrix remains constant.
- This approach cannot distinguish strong selection that happened over a short period of time from weak selection that happened over a long period of time.
- This approach *assumes* that the observed differences in populations are the result of selection, but populations isolated from one another will diverge from one another even in the absence of selection simply as a result of genetic drift.
  - Small amount of differentiation between populations within sites could reflect relatively recent divergence of those populations from a common ancestral population.
  - Large amount of differentiation between populations from inland versus coastal sites could reflect a more ancient divergence from a common ancestral population.

## **Creative Commons License**

These notes are licensed under the Creative Commons Attribution License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.