Genetic variation of polygenic characters and the evolution of genetic degeneracy

S. A. FRANK

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

Keywords:

mutation;
quantitative characters;
robustness.

Abstract

The classical model of mutation–selection balance for quantitative characters sums the effects of individual sites to determine overall character value. I develop an alternative version of this classical model in which character value depends on the averaging of the effects of the individual sites. In this new averaging model, the equilibrium patterns of variance in allelic effects and character values change with the number of sites that affect a character in a different way from the classical model of summing effects. Besides changing the patterns of variance, the averaging model favours the addition of loci to the control of character values, perhaps explaining in part the recent observation of widespread genetic degeneracy.

Introduction

Genetic regulatory networks control the activation and repression of coordinated effects. Quantitative characters are the consequence of the structure of genetic regulatory networks and the parameters that control the dynamics of those networks. The details of regulatory networks are just now being worked out for a few characters (Carroll *et al*., 2001; Davidson, 2001; Ptashne & Gann, 2002). Going forward, we may eventually be able to replace the black box of polygenic control for quantitative characters with a more accurate description of how genes interact to determine phenotype.

At present, no clear generalizations have emerged about the consequences of network structures for polygenic control of quantitative characters. Two ideas – degeneracy and robustness – have been put forward as candidate principles for genetic regulatory networks.

Degeneracy occurs when loss of a gene has little effect because the network retains a backup system (Edelman & Gally, 2001). The observation that knockout mutations often do not have detectable phenotypic consequences (Melton, 1994; Winzeler *et al*., 1999) in the lab has led to the idea that degeneracy (Edelman & Gally, 2001) is a common property of biological networks. Theories of degeneracy face the challenge that natural selection does not easily favour degeneracy because an existing functional system prevents any direct advantage to an organism carrying an extra backup system. Recent theory shows various ways in which backup systems can be exposed to selection and favoured to spread in populations (Nowak *et al*., 1997). This theory relies on protection against complete loss-of-function mutations in Mendelian traits and does not consider degeneracy in quantitative traits under polygenic control.

Robustness occurs in the dynamical control of a complex trait when large changes in the parameters that govern the dynamics have relatively little effect on the quantitative character. For example, Barkai & Leibler (1997) showed that the network structure regulating bacterial chemotaxis provides robust performance with low sensitivity to the individual parameters.

Clearly, robustness is an important property of a well-designed network. However, a robust biological network must accumulate extensive genetic variability because genetic changes have relatively little effect on performance and fitness. No matter how robust the control, genetic variation accumulates until the rate at which natural selection removes variation equals the rate at which mutation adds variation. Most work on robustness of genetic networks has taken an engineering perspective, in which the self-evident property that robust networks must decay by mutation pressure has received little attention.

Correspondence: Steven A. Frank, Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525,USA.
Tel.: +1 949-824-2244; Fax: +1 949-824-2181; e-mail: safrank@uci.edu

These issues of network structure and mutational decay lead to two questions. First, what are the consequences of network architecture for the accumulation of polygenic variation in quantitative characters? Second, what is the optimal network structure to protect against mutational decay?

To make a start on these important questions, I have chosen to study the classical mutation–selection balance model for quantitative traits under stabilizing selection (Kimura, 1965; Lande, 1975; Turelli, 1984; Frank & Slatkin, 1990). I show that by altering the classical model in a minor way, significant differences arise in the equilibrium genetic variance under the balance between mutation and selection.

The classical model sums the effects of individual loci to obtain the deviation of the character value from the average value in the population. Instead, I average the effects of the individual loci. Averaging simply means that I divide the contribution by each locus to the deviation from average value by the number of genes that contribute to the trait.

Averaging has interesting consequences because it creates a negative association between the number of genes affecting a character and the contribution of each gene to character value. The averaging model leads to more genetic variation at each locus because selection per locus is weaker. For character values, the averaging model has the same genetic variance as the summation model when stabilizing selection is relatively strong. When stabilizing selection on each locus is relatively weak, characters controlled by averaging accumulate less genetic variance than characters controlled by summed effects.

I show that under the classical summation model, natural selection favours a reduction in the number of genes that contribute to character value. By contrast, under the averaging model natural selection may often favour an increase in the number of genes that control a character.

Neither the classical summation model nor the averaging model provide a realistic description of genetic regulatory networks and the genetic effects of individual loci within such networks. Rather, different network architectures will combine genetic effects in different ways, some architectures more like the summation model and some more like the averaging model. My analyses show that different architectures have different consequences for the accumulation of genetic variation. This conclusion will help in understanding how different network structures accumulate variation and in how networks may be designed by natural selection and by engineers.

The classical model of additive effects

In my model, I use the word *sites* for the n pieces of DNA that contribute to the character. In a diploid model, there are $n/2$ loci, with two sites per locus; in a haploid model, there is one site at each locus. I use sites without distinguishing ploidy level because I will add across sites, so it does not matter if there are two sites at one diploid locus or two sites at two haploid loci. Each site has alleles x_i for $i = 1,\ldots,n$ sites, where each x_i is a random variable taking on different values that describe the contribution of the ith site to character value. Note that the word *site* is sometimes used for a variable amino acid position. Here, the number of sites n is NL, where N is the ploidy level and L is the number of loci.

Fisher (1918) developed the classical model for combining the effects of individual sites. Fisher's regression approach used least squares to maximize the proportion of the observed phenotypic variation explained by genotype. For example, consider a phenotypic character z affected by alleles x_i. The different kinds of alleles may be encoded by categorical values, for example, $x_i = 0,1$ when each site has two different allelic forms. Then, by standard regression, we can write $\hat{z} - \bar{z} = \sum_i b_i(x_i - \bar{x}_i)$, where $\hat{z}$ is the expected character value for individuals with a particular genotype. This regression equation can be read as: the expected deviation of a character from its population average equals the sum of the average effect of each site, b_i, multiplied by the deviation of the allelic value from its population average, $x_i - \bar{x}_i$. The average effects of sites are simply the partial regression coefficients obtained by using the standard theory of least squares to fit the regression model. This summing model for character value forms the foundation for measuring heritability and the design of agricultural breeding programs (Falconer & Mackay, 1996).

Classical regression provides the best approach for prediction, which is the aim of heritability measures and breeding programmes. The ubiquity of the summation model made it the standard choice when theoreticians began to study evolutionary models of quantitative characters. However, summation of genetic effects in theoretical models of mutation and selection assume fixed, additive effects for each gene rather than estimating average (additive) effects from a regression model. In other words, the summation of additive effects in models of mutation and selection is based on a loose analogy with regression models of prediction – there is in fact no formal correspondence. My point here is that summing is an arbitrary choice for theoretical models of mutation and selection, based on the classical approach for studies of heritability.

Below, I present a new model for the control of quantitative characters based on the averaging of the effects of individual sites rather than the summing of their effects. Genetic effects will often aggregate in an averaging way when phenotype depends on the frequency of encounter with the protein products of each site. To show the contrast between the standard summation model and my averaging model, it is useful to begin with the classical summation approach for models of mutation and selection.

A balance between the influx of deleterious alleles by mutation and the removal of deleterious alleles by stabilizing natural selection maintains genetic variation in quantitative characters. Under such stabilizing selection, there is an optimum character value, z^*, and characters that deviate from z^* suffer a reduction in fitness. Previous theoretical models (Kimura, 1965; Lande, 1975; Turelli, 1984; Frank & Slatkin, 1990) assume that character value is determined by summing genetic effects, by analogy with the classical regression approach of quantitative genetics. For example, we can write $z = c\sum_i x_i = cX$, where $x_i = 0, \pm1, \pm2,\ldots$, the value of c gives the phenotypic effect of a mutation that transforms an allele to an adjacent type, and X is the sum of the allelic values (Slatkin, 1987; Frank & Slatkin, 1990). Assuming that allelic values are independent, the variance in the character value is $V_z = c^2nV_x$, where V_x is the variance in allelic values for each site.

The standard model of stabilizing selection sets the optimum at $z^* = 0$ and describes fitness losses in terms of the distance from the optimum (Lande, 1975; Turelli, 1984). Fitness is $w(z) = e^{-z^2/2V_s}$, where V_s is inversely proportional to the intensity of selection acting on squared deviations from the optimum. When $z^2/2V_s$ is small, then $w(z) \approx 1-z^2/2V_s = 1-sX^2$, where $s = c^2/2V_s$ is the strength of selection acting on a unit change in the squared sum of allelic values.

In the stepwise mutation model (Slatkin, 1987), μ is the probability in each generation that an allele mutates to an adjacent allelic class. The equilibrium variance in allelic values, V_x, and the character variance, V_z, depend on the relative intensity of selection and mutation. Approximations can be obtained for the two bounding conditions $s >> \mu$ and $s << \mu$. When $s >> \mu$, we have approximately (Turelli, 1984; Frank & Slatkin, 1990)

$$V_x = \mu/s = 2\mu V_s/c^2 \quad (1)$$

$$V_z = c^2n\mu/s = 2n\mu V_s. \quad (2)$$

When $s << \mu$ (Kimura, 1965; Lande, 1975; Frank & Slatkin, 1990),

$$V_x = \sqrt{\mu/2s} = \sqrt{\mu V_s/c^2} \quad (3)$$

$$V_z = c^2n\sqrt{\mu/2s} = n\sqrt{c^2\mu V_s}. \quad (4)$$

Averaging genetic effects

Now suppose that the character depends on the average of the allelic values rather than the sum, that is, $z = (c/n)X$. The results for the previous model are changed by two opposing forces. First, averaging rather than summing reduces the character variance by n^2, $V_z = c^2V_x/n$. Second, averaging reduces the strength of selection on each site by n^2 relative to the summing model because the contribution of each site is reduced by $1/n$ and selection changes by the square of character value. With these two factors, we can write the equilibrium variances under the averaging model. For $s/n^2 >> \mu$,

$$V_x = n^2\mu/s = 2n^2\mu V_s/c^2 \quad (5)$$

$$V_z = c^2n\mu/s = 2n\mu V_s. \quad (6)$$

Compared with the summing model, the variance of allelic effects at each site is much larger – weighted by an additional factor n^2 because of the reduced selection per site. However, the total character variance remains the same because the averaging process reduces the total character variance by the same factor as the increase in the variance per site.

When $s/n^2 << \mu$,

$$V_x = \sqrt{n^2\mu/2s} = \sqrt{n^2\mu V_s/c^2} \quad (7)$$

$$V_z = c^2\sqrt{\mu/2s} = \sqrt{c^2\mu V_s}. \quad (8)$$

In this case, averaging increases allelic variances per site by a factor of n. This makes the total character variance independent of the number of sites under the averaging model.

We could also consider sets of characters that combine in a summing or averaging way to make an aggregate character that affects fitness. For example, a single character in an averaging suite of characters may have high character variance in the same way that a single site has high variance when contributing in an averaging way to a character.

Consequences of the genetic control of characters

We can now study how natural selection affects the number of sites that determine a character. Suppose that some individuals have n sites affecting the character and other individuals have $n + a$ sites affecting the character. If the average fitness of those individuals with $n + a$ sites is greater than the average fitness of those individuals with n sites, then the population evolves towards control of the character by the higher number of sites. In particular, the condition for the spread of additional sites affecting the character is $E(\hat{w} - \tilde{w}) > 0$, where the hat denotes the class with $n + a$ sites and the tilde denotes the class with n sites.

Under the summing model, adding more sites increases the variance of the character and reduces fitness, so selection favours a reduction in the number of sites contributing to character value. Under the averaging model, if the two classes with n sites and $n + a$ sites are both at mutation–selection equilibrium, then there are two cases. First, for relatively strong selection, $s/n^2 >> \mu$, the class with more sites has higher character variance and lower fitness because character variance rises linearly with the number of sites. Selection therefore favours a reduction in the number of sites controlling the character. Second, for relatively weak selection, $s/n^2 << \mu$, character variance is independent of the

number of sites. Thus, the number of sites is a neutral character that can drift higher or lower.

The time to achieve mutation–selection equilibrium can be long because mutation is a weak force. The competition between the classes with n and $n + a$ sites may often depend on the process that generates the extra a sites and the dynamics of mutation and selection within each class. For example, consider the case of the averaging model with relatively weak selection, $s/n^2 << \mu$. Suppose a population initially has n sites at mutation–selection equilibrium. A single individual then adds a sites to the control of the character z, for a total of $n + a$ sites. The initial success of the additional sites depends on whether those extra sites occur in a relatively successful genotype.

If the extra sites do increase initially, they will recombine with the other n sites, reducing linkage disequilibrium with those sites. The new sites will also mutate, accumulate genetic variation, and evolve so that their average allelic value moves toward the optimum of zero. For a diploid case with n sites and $n/2$ loci, adding another locus sets $a = 2$. Under the assumption that the new a sites are in approximate linkage equilibrium with the original n sites, the condition for the increase of an additional diploid locus is $n(V_y + 2\bar{y}^2) < 2(n + 1)V_x$, where V_x is the variance in allelic effects at the original n sites, V_y is the variance in allelic effects at the new a sites, and $\bar{y}$ is the average allelic value at the new sites (see Appendix).

Assuming that the new sites have an average value near the optimum, $\bar{y} \approx 0$, a sufficient condition for the increase of the new locus is approximately $V_y < 2V_x$. This is an easy condition to satisfy because the variance at the new sites, V_y, should not be twice the variance at the old sites, V_x. Thus, selection favours a steady increase in the number of sites controlling the character value. As each additional site spreads through the population, the equilibrium character variance remains constant because, for averaging control and $s/n^2 << \mu$, the character variance is independent of the number of sites.

Conclusions

Allelic effects will often aggregate in an averaging way when phenotype depends on the frequency of encounter with the protein products of each site. This sort of phenotypic averaging probably happens in many cases. However, the relative occurrence of summing, averaging, and other more complex patterns of aggregating effects can only be determined by empirical study. The value of the theory here is to show clearly that this is an important problem for understanding the nature of quantitative variation. This understanding has consequences for the evolution of the genetic control of characters and for the distribution of allelic effects and character values in complex, polygenic traits.

Acknowledgments

I thank M. A. Nowak for helpful comments. National Science Foundation grant DEB–0089741 and National Institutes of Health grant AI24424 support my research.

References

Barkai, N. & Leibler, S. 1997. Robustness in simple biochemical networks. *Nature* **387**: 913–917.

Carroll, S.B., Grenier, J.K. & Weatherbee, S.D. 2001. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell, Oxford.

Davidson, E.H. 2001. *Genomic Regulatory Systems: Development and Evolution*. Academic Press, San Diego, CA.

Edelman, G.R. & Gally, J.A. 2001. Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. USA* **98**: 13763–13768.

Falconer, D.S. & Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edn. Longman, London.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh* **52**: 399–433.

Frank, S.A. & Slatkin, M. 1990. The distribution of allelic effects under mutation and selection. *Genet. Res.* **55**: 111–117.

Kimura, M. 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Natl. Acad. Sci. USA* **54**: 731–736.

Lande, R. 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* **26**: 221–235.

Melton, D.W. 1994. Gene targeting in the mouse. *BioEssays* **16**: 633–638.

Nowak, M.A., Boerlijst, M.C., Cooke, J. & Maynard Smith, J. 1997. Evolution of genetic redundancy. *Nature* **388**: 167–171.

Ptashne, M. & Gann, A. 2002. *Genes and Signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Slatkin, M. 1987. Heritable variation and heterozygosity under a balance between mutations and stabilizing selection. *Genet. Res.* **50**: 53–62.

Turelli, M. 1984. Heritable genetic variation via mutation–selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Pop. Biol.* **25**: 138–193.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**: 901–906.

Received 04 June 2002; accepted 11 June 2002

Appendix

The criteria that favour the addition of an extra diploid locus to the control of a trait are derived as follows. Fitness is $w = 1 - z^2/2V_s$. Denote the fitness of genotypes with n sites as $\bar{w}$ and the fitness of genotypes with $n + a$ sites as $\hat{w}$. The additional sites spread if $E(\hat{w} - \bar{w}) > 0$, which is equivalent to $E(\hat{z}^2) < E(\bar{z}^2)$. Expanding this condition with the definition of z gives

$$(c^2/(n+a)^2)\left(nV_x + E\left(\sum_{j=1}^{a} y_j\right)^2\right) < c^2V_x/n,$$

where I have assumed that the n original sites have means $\bar{x} = 0$ and each has a distribution independent of the other sites. This condition simplifies to $nE(\sum_{j=1}^{a} y_j)^2 < (2na + a^2)V_x$. The new sites with values y_j may not have had time to evolve toward the optimum $\bar{y} = 0$. We can take account of deviations in mean values by writing $y_j = \bar{y}_j + \delta_j$, where δ describes the random deviations of y about its mean and thus δ has a mean of zero.

The transition to add an additional diploid locus, with $a = 2$, is $nE(y_1+y_2)^2 < 4(n+1)V_x$. Here y_1 and y_2 are alternative alleles at the same diploid locus and would therefore likely have the same distribution and, under random mating, would have independent distributions. Thus, $E(y_1 + y_2)^2 = 2V_y + 4\bar{y}^2$, giving the condition $n(V_y + 2\bar{y}^2) < 2(n+1)V_x$. Assuming $\bar{y} \approx 0$, a sufficient condition is $V_y < 2V_x$, as in the text.