

LONG TERM SELECTION: REPEATABILITY OF RESPONSE IN FINITE
POPULATIONS

Bruce Walsh

Department of Ecology and Evolutionary Biology

BioSciences West,

University of Arizona,

Tucson, Arizona, USA 85721

bwalsh@u.arizona.edu

LONG TERM SELECTION: REPEATABILITY OF RESPONSE IN FINITE
POPULATIONS

Bruce Walsh

University of Arizona, Tucson, Arizona USA

KEYWORDS: Genetic drift, additive genetic variance, mutational variance

INTRODUCTION

With finite population size comes genetic drift --- random change in to random sampling of gametes to form the next generation. Even when selection, genetic drift is still potentially important in that favorable alleles can be lost from a population due to drift. A direct consequence of drift is variation in the response to selection, even among initially identical replicate lines. Under long-term selection, mutation introduces additional variation in response by introducing different mutations in different lines.

A SIMPLE EXAMPLE

Consider a cross between two inbred lines, the first fixed for favorable alleles at two loci and the second fixed for favorable alleles at two different loci. The resulting cross is $AABBccdd \times aabbCCDD$. All F_1 individuals are $AaBbCcDd$ and have favorable alleles at each of the four loci. Genetically identical F_1 replicate lines that are then subjected to identical selection.

If selection is weak on a favorable allele (as would occur if the allele has a small effect on improving the trait of interest), the probability that the allele will be fixed is 50%. Hence, when considering the A locus, 50% of the lines will eventually have the favorable allele (AA) and the other 50% will be aa. The probability that a line will eventually have the favorable alleles at all loci (AABBCCDD) is $(1/2)^4 = 0.0625$. The probabilities of finding favorable alleles at only three, two, or one loci are likewise 0.125, 0.25, and 0.25. One thus expects considerable genetic variation between lines due to differences in the mean performance of the replicate lines. This variation can occur even though the initial lines were completely identical and subject to the same selection. Obviously, slight variations between the genetic compositions and/or variation in the nature of selection on the lines will generally affect line performance.

A final point is that lines identical in performance may in reality have different genetic compositions. For example, two lines each fixed for two favorable alleles may be genetically very different, for example AABBcccc and aabbCCDD. Crossing these fixed lines and selecting allows for the possibility of recovering all four favorable alleles. Thus the amount of genetic variance between fixed lines is considerable and important in breeding. If the between-line genetic variance is high, lines followed by a second round of selection can result in further improvement in the best performing lines. Similarly, if the variance is low, there is less potential for improvement by crossing and further selection.

VARIANCE IN LINE MEANS UNDER PURE DRIFT

The expected variation in the means of lines experiencing only drift can be calculated starting from a common starting point for examining the expected variance in response under

particular, if a trait is determined by a large number of loci each of which is under weak selection (i.e., selection is weak on each), the allelic frequency dynamics are often expected under drift than to those for selection. Consider one locus with alleles Q and q , and suppose (relative to the mean trait value) that the genotypes contribute values of $2a$: $a(1+h)$: 0 to the trait of interest. In the absence of mutation, eventually either allele Q or q is fixed. Under drift, the probability U that allele Q is fixed is just equal to its initial frequency p , and $1-p$ is the probability that allele q is fixed instead. Since ultimately the population will be either QQ or qq , the expected contribution from this locus becomes

$$E[R] = 2a p + 0 * (1-p) = 2 a p \quad (1a)$$

For example, if QQ individuals are, on average, 4 cm taller than qq individuals, and if (say) the initial frequency of allele Q is $p = 0.75$, then the expected contribution from this locus is $2a p = 3$ cm. Recalling that we can write the variance $E(x^2) - \mu^2$, where μ is the mean value and $E(x^2)$ the average value of the variable, the variance about the expected contribution becomes

$$\text{Var}(R) = (2a)^2 p + 0^2 (1-p) - R^2 = 4a^2 p(1-p) \quad (1b)$$

If the alleles show no dominance ($h = 0$ so that the value of Qq is equal to the average of QQ and qq), the additive genetic variance V_A contributed by the locus is $4a^2 p(1-p)$ and hence the variance in response is just twice the additive variance. The phenotypic covariance between a parent and its offspring is $V_A/2$ (1, 2), and one could estimate the expected variance in response under strict drift from the parent-offspring covariance. When dominance or epistasis is present, the relationship between additive variance and variance in response no longer holds.

VARIANCE IN RESPONSE UNDER DRIFT AND SELECTION

When selection occurs, the above expressions for the expected mean a response still hold, provide we replace p by U, the probability of fixation. Substituting U for p in the above expressions gives

$$E [R] = 2 a U + 0 (1-U) = 2 a U$$

and

$$\text{Var}(R) = (2a)^2 U + 0^2 (1-U) - R^2 = 4 a^2 U(1-U)$$

-----> FIGURE ONE APPROXIMATELY HERE <-----

Figure 1 plots the variance in response as a function of U. Note that its maximum value (a^2) at $U = 1/2$. Note from the figure that a mild selection may actually increase the variance in response relative to that under drift alone. If U is closer to 1/2 than is p, this results in an increased variance. For example, if $p = 0.1$, then under drift $\text{Var}(R) = 4 a^2 p(1-p) = 0.36 a^2$. If selection increases the probability of Q being fixed to $U = 0.4$, then $\text{Var}(R) = 4 a^2 U(1-U) = 0.96 a^2$. Selection on Q becomes stronger, the variance in response becomes significantly greater than that under drift alone. Conceptually, this makes sense in that if selection favors the most favorable alleles, then there will be little variation among the alleles that are fixed. Consequently when there is strong selection underlying loci, there is little to be gained by crossing and reselection.

COMPUTING U, THE PROBABILITY OF FIXATION UNDER DRIFT AND SELECTION

To obtain the probability U of fixation of allele Q , we need an expression for selection coefficients to fitnesses on genotypes and an expression relating how selection on a trait translates into fitnesses at the underlying loci. Suppose the fitness values are

Genotype	QQ	Qq	qq
Fitness	$1 + 2s$	$1 + s$	1
Character value	$1 + 2a$	$1 + a$	1

For these fitnesses, Kimura (3) showed that if Q starts at frequency p , the probability of fixation is

$$U = [1 - \text{Exp}(-4Nsp)] / [1 - \text{Exp}(-4Ns)] \quad (3)$$

where N is the population size and $\text{Exp}(x)$ is the exponential function. If $s = 0$, $U = p$, and the allele behaves as if it is neutral.

It remains to obtain s as a function of the selection on the trait at the individual alleles. For a locus underlying a trait under selection, the selection coefficient s is approximately

$$s = (a / \sigma_z) i \quad (4)$$

σ_z is the square root of phenotypic variance of the trait, and i is the selection coefficient on the character (for example, if the upper 20, 10, or 5% of the population are selected, i is 1.7, and 2.0). Details are given in Walsh and Lynch (4). Thus, $U(p)$ is significantly different from p when $2N(1-p)(a / \sigma_z) i$ is significantly large. The parameters are population (N), individual allelic effect (a / σ_z), and selection (i) coefficient.

allele is largely controlled by drift. Using Equation (3) and (4), the probability that a favorable allele will exceed 0.7 when

$$N | i | p (| a | / \sigma_z) > 0.5 \quad (5)$$

Thus if the product of initial allele frequency and standardized allele effect is small, many favorable alleles will be lost. Conversely, if Equation (5) is satisfied, favorable alleles are fixed, and the resulting variance in response in crosses between replicate lines are unlikely to generate significant subsequent selection.

THE EFFECTS OF MUTATION

In addition to the variance introduced by drift, additional variation is generated by mutation. Let V_M denote the amount of variation introduced per generation by mutation. Under a pure drift model, the increase in between-line variance from drift for n lines separated for t generations approaches $2 t N V_D$, which can be substantial. Crosses between replicate lines that have been separated for t generations thus show considerable potential for additional response. Hill (5) predicts that between-line divergence when selection on the underlying loci is strong relative to drift is

CONCLUSIONS

Considerable variation between the performance of selected replicate lines is expected from both drift and mutation. If selection on any underlying locus is strong, significant variation in response is expected. Crosses between such replicate lines are unlikely to show significant genetic variation for further selection. If selection is weak on any locus, there is little between-line variation, and selection on crosses is unlikely to result in significant improvement.

REFERENCES

1. Falconer, D. ; Mackay, T. F. C. Introduction to Quantitative Genetics. Longman: Essex, 1996.
2. Lynch, M.; Walsh, B. Genetics and Analysis of Quantitative Traits. Sinauer Associates: Sunderland, MA, 1998.
3. Kimura, M. Some problems of stochastic processes in genetics. *Journal of Theoretical Biology* 1957, 28, 882--901.
4. Walsh, B.; Lynch, M. Evolution and Selection of Quantitative Traits. *Evolutionary Ecology* (2003) are online at http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.htm
5. Hill, W. G. Fixation probabilities of mutant genes with artificial selection. *Genetics, Selection, Evolution* 1985, 17, 351--358.

FIGURE CAPTION

Figure 1. The variance in response (in units of a^2) as a function of the number of alleles fixed.