

HOW IMPORTANT IS MUTATION FOR SELECTION RESPONSES?

Zhao-Bang Zeng and Hidenori Tachida
Department of Statistics, North Carolina State University,
Raleigh, North Carolina 27695-8203 USA

SUMMARY

The ratio of selection response due to mutation to total response is analyzed in a general genetic model which allows more than two alleles segregating at a locus. Analytic results are obtained for weak selection. For strong selection the ratio is evaluated by simulations. The results show that Hill's prediction (1982, *Genet. Res.* 40: 255-278) of selection response from mutation slightly overestimates the ratio. Significant differences between our and Hill's predictions are likely to occur for relatively strong selection for more than 100 generations. The ratio depends very much on the parameters. For reasonable choices of parameter values the ratio is about 10 to 25% at $t = 25$, and 25 to 50% at $t = 50$, where t is the generation number of selection.

INTRODUCTION

Hill (1982a,b) has proposed a theory to predict the selection response from new mutations. Based on some experimental observations and theoretical analyses, he argued that mutation can make a significant contribution to response to selection when selection continues for more than 20 generations. The basic assumptions of the theory are that: (i) every new mutation occurs at a monomorphic locus so that there will be no more than two alleles segregating in a population; (ii) the distribution of allelic effect difference between new mutant and pre-fixed allele is symmetric and constant, independent of selection effect; and (iii) the rate of mutation and phenotypic variance are constant over generations. From these assumptions the selection response from new mutations, R^M , can be predicted independently from that from the initial genetic variation, R^I , with $R^T = R^I + R^M$, if linkage effects are ignored, and moreover R^M can be predicted as

$$R_t^M = r_1 + r_2 + \dots + r_t \quad (1)$$

where r_t is the expected response from mutants occurring t generations earlier, which gives no limit to selection response. The assumption (ii) implies that the effect of new mutant depends on the effect of pre-fixed allele. However, when allelic effects are treated independent, the response will ultimately reach a limit balanced by selection, mutation, and drift (Zeng, Tachida and Cockerham, 1989). In this paper, we examine the ratio R^M/R^T for this alternative mutation model.

THE MODEL

We assume that there are K alleles at a locus and mutation rate to the i th allele A_i from other alleles is v_i . The total mutation rate is denoted by $u = \sum v_i$. Let p_i and a_i be the frequency and effect of the allele A_i . The change of gene frequency p_i in one generation is

$$\Delta p_i = sp_i(a_i - \sum a_j p_j) + v_i - up_i + \xi_i \quad (2)$$

where $s = \iota/\sigma$, ι is the selection intensity, σ is the phenotypic standard deviation, and ξ_i is the change due to random drift with $\mathcal{E}(\xi_i) = 0$, $\mathcal{E}(\xi_i \xi_j) = -p_i p_j / 2N$, ($i \neq j$), and $\mathcal{E}(\xi_i^2) = p_i(1 - p_i) / 2N$ where \mathcal{E} denotes the expectation and N is the effective population size during selection. Consequently, the transition equations for the first and second moments are

$$\begin{aligned} \mathcal{E}(p_i)' &= (1 + sa_i - u)\mathcal{E}(p_i) + v_i - s \sum_k a_k \mathcal{E}(p_i p_k) \\ \mathcal{E}(p_i p_j)' &= v_i \mathcal{E}(p_j) + v_j \mathcal{E}(p_i) + (1 + sa_i + sa_j - 2u - 1/2N)\mathcal{E}(p_i p_j) \\ &\quad - 2s \sum_k a_k \mathcal{E}(p_i p_j p_k) \\ \mathcal{E}(p_i^2)' &= (2v_i + 1/2N)\mathcal{E}(p_i) + (1 + 2sa_i - 2u - 1/2N)\mathcal{E}(p_i^2) - 2s \sum_k a_k \mathcal{E}(p_i^2 p_k). \end{aligned} \quad (3)$$

We do not intend to give a detailed analysis of this system here. Rather we present some approximations and simulation results based on it.

ANALYSIS AND DISCUSSION

First, when selection is weak ($N\iota\sigma_a/\sigma \rightarrow 0$, where σ_a^2 is the variance of allelic effects), an approximate solution of (3) can be found. This is obtained by expanding the solution in terms of s (formally it should be derived in terms of $2Ns\sigma_a$ in unit of time of $2N$ generations)

$$\begin{aligned} \mathcal{E}(p_{it}) &= s^0 f_{i0}(t) + s^1 f_{i1}(t) + s^2 f_{i2}(t) + \dots \\ \mathcal{E}(p_{it} p_{jt}) &= s^0 g_{ij0}(t) + s^1 g_{ij1}(t) + s^2 g_{ij2}(t) + \dots \end{aligned} \quad (4)$$

We need $f_{i1}(t)$ to obtain the selection response and for this it is necessary to solve for $f_{i0}(t)$ and $g_{ij0}(t)$. By equating the coefficients of s^0 on both sides of the equation for $\mathcal{E}(p_i)$ with the initial condition $f_{i0}(0) = p_{i0}$, the initial frequency of the allele A_i , it is found $f_{i0}(t) = p_{i*} + (p_{i0} - p_{i*})\rho^t$, where $p_{i*} = v_i/u$ and $\rho = (1 - u)$. Similarly, after obtaining $g_{ij0}(t)$ and $g_{ii0}(t)$, $f_{i1}(t)$ is found to be

$$\begin{aligned} f_{i1}(t) &= \frac{4N}{1 + \theta} p_{i*} \alpha_i (1 - \rho^t) + \frac{\theta}{2 + \theta} (p_{i0} \alpha_i + p_{i*} \beta_i - 2p_{i*} \alpha_i) t \rho^t \\ &\quad + \frac{4N}{2 + \theta} (p_{i0} \beta_i + \frac{\theta^2}{(1 + \theta)(2 + \theta)} p_{i*} \alpha_i - \frac{\theta}{2 + \theta} (p_{i0} \alpha_i + p_{i*} \beta_i)) (\rho^t - \lambda^t) \end{aligned} \quad (5)$$

where $\lambda = (1 - 2u - 1/2N)$, $\theta = 4Nu$, $\alpha_i = a_i - \sum_j a_j p_{j*}$, and $\beta_i = a_i - \sum_j a_j p_{j0}$. By assuming that a_i 's are identically independently distributed with $\mathcal{E}(a_i) = \mathcal{E}(a_i a_j) = 0$, ($i \neq j$), and $\mathcal{E}(a_i^2) = \sigma_a^2$, the selection response from the locus, $\mathcal{E}(R_i^T) = 2\mathcal{E}(\sum a_i p_{it})$, ignoring the terms of $O(s^2)$, is then

$$\begin{aligned} \mathcal{E}(R_i^T) &= \frac{8Ns\sigma_a^2}{1 + \theta} (1 - q_*) (1 - \rho^t) + \frac{4\theta s\sigma_a^2}{2 + \theta} (q_* - q_{0*}) t \rho^t \\ &\quad + \frac{8Ns\sigma_a^2}{2 + \theta} (Q_* - Q_0 + \frac{2\theta}{2 + \theta} (q_{0*} - q_*) (\rho^t - \lambda^t)) \end{aligned} \quad (6)$$

where $q_* = \sum p_{i*}^2$, $q_{0*} = \sum p_{i0} p_{i*}$, $Q_0 = \sum p_{i0}^2$, and $Q_* = (1 + \theta q_*) / (1 + \theta)$ (Cockerham, 1984; Cockerham and Tachida, 1987). Note that in this model the prediction of $\mathcal{E}(R_t^M)$ [$= \mathcal{E}(R_t^T) - \mathcal{E}(R_t^I)$] generally depends on $\mathcal{E}(R_t^I)$. When there is no mutation ($u = 0$), (6) reduces to

$$\mathcal{E}(R_t^I) = 4N_s \sigma_a^2 (1 - Q_0) (1 - (1 - 1/2N)^t), \quad (7)$$

the conventional weak selection prediction of selection response from the initial genetic variation. As $t \rightarrow \infty$, (6) approaches the selection limit discussed in Zeng, Tachida and Cockerham (1989). Also by letting $Q_0 = 1$, (6) predicts the selection response from a locus fixed in the base population.

The ratio of selection response due to mutation to total response, $\mathcal{E}(R_t^M) / \mathcal{E}(R_t^T)$, is presented in Table 1 for a model of infinite alleles, $K \rightarrow \infty$ ($p_{i*} \rightarrow 0$, $\sum p_{i*}^2 \rightarrow 0$), and an initial neutral equilibrium population with an effective size N_0 (thus $Q_0 = 1 / (1 + 4N_0u)$). This ratio depends on parameters N_0, N, u and t , but the table shows that for given t the ratio is mainly a function of N_0 . The effects of N and u on the ratio are relatively small. For N this is because it does not influence the initial and new mutation genetic variances. An increase of N would increase both $\mathcal{E}(R_t^I)$ and $\mathcal{E}(R_t^M)$ because of the decrease of drift effect, but with $\mathcal{E}(R_t^I)$ slightly more increased. Contrarily u influences both the initial genetic variance and the new genetic variance in a similar manner.

TABLE 1
 $\mathcal{E}(R_t^M) / \mathcal{E}(R_t^T) \times 100$ under weak selection from (6) and (7)

N_0	N	u	$t = 25$	$t = 50$	$t = 75$	$t = 100$	$t = 150$	$t = 200$
50	20	10^{-4}	11.65	22.73	32.36	40.52	52.95	61.51
100			6.14	12.77	19.29	25.43	36.09	44.54
250			2.60	5.56	8.77	12.08	18.61	24.60
500			1.21	2.76	4.52	6.42	10.40	14.32
1000			0.56	1.35	2.29	3.35	5.66	8.06
250	10	10^{-4}	2.87	6.45	10.31	14.11	21.04	26.97
	50		2.43	4.94	7.52	10.14	15.37	20.47
	100		2.37	4.73	7.07	9.38	13.90	18.26
250	20	10^{-3}	1.80	4.54	7.91	11.62	19.07	25.78
		10^{-5}	2.68	5.66	8.85	12.11	18.53	24.42
		10^{-6}	2.69	5.67	8.85	12.12	18.51	24.40

The value of N_0 can be inferred from estimates of the initial genetic variance $\sigma_g^2 = 2m(1 - Q_0)\sigma_a^2$ and the mutation variance $\sigma_m^2 = 4mu\sigma_a^2$, where m is the number of loci that could mutate. When the initial population is at the neutral equilibrium, $\sigma_g^2 / \sigma_m^2 = 2N_0 / (1 + 4N_0u) \rightarrow 2N_0$, if $4N_0u$ is small. For example, for bristle number in *Drosophila*, σ_m^2

is usually estimated at around $0.001\sigma_e^2$ (Hill, 1982b), where σ_e^2 is the environmental variance. Taking $\sigma_g^2/\sigma_e^2 = 0.5$ (i.e., 0.33 initial heritability) would indicate N_0 to be about 250.

Note that the ratios in Table 1 are for weak selection. As selection force increases, the ratio increases. To determine the effects of selection on the ratio, we performed simulations using the pseudosampling method (Kimura and Takahata, 1983; see also Zeng and Cockerham, 1990). Allelic effects were assumed to be normally distributed. Initial gene frequencies were assumed to result from a neutral equilibrium population or from a cross between two inbred lines. In the neutral equilibrium case, given N_0 and u , Ewens' sampling method (with sample size 100) was applied to sample the initial number and frequencies of alleles. Whenever a mutation occurs, a new allele is introduced. In the inbred cross case, we assume that the inbred lines have been separated for τ generations from a common equilibrium population with $q_0 = 1/(1 + 4N_0u)$. Whether a locus is monomorphic or polymorphic (with $p = 0.5$) in a cross depends on the probability that genes are alike among the inbred lines which is $q_\tau = q_0\rho^\tau$ for $K \rightarrow \infty$ (Cockerham, 1984). The expected genetic variance in these base populations is then proportional to $0.5(1 - q_\tau)$, which equals to $(1 - q_0)$ when $\tau = 0.5 \ln(2 - 1/q_0) / \ln(1 - u) \approx 2N_0$. When $\tau < 2N_0$, $0.5(1 - q_\tau) < (1 - q_0)$. Note that $1 - q_\tau$ is also the proportion of loci segregating in the base populations to the total number of loci that *could* mutate. As $\tau \rightarrow \infty$, no genes will be alike among the inbred lines. Simulations for the inbred cross case were performed independently for the hybrid locus ($p = 0.5$) and the fixed locus ($p = 0$, or 1), and appropriate weights were placed on these cases to obtain the observed responses.

Some simulation results are presented in Table 2 for $N_0 = 250$ (with $S = 4N_0\sigma_a/\sigma$). For the inbred cross case, $\tau = 527$ so that the initial genetic variance is the same as the neutral equilibrium case with $N = 100$ and $u = 10^{-4}$. Together with other results not shown here, the simulations show that: (i) strong selection can significantly increase the ratio R_i^M/R_i^T even at early stages of selection; (ii) although r_i/r_1 in (1) is primarily a function of σ_a/σ (Hill, 1982b), R_i^M/R_i^T is primarily a function of $N\sigma_a/\sigma$ (Hill and Rasbash, 1986); (iii) given other parameters (e.g., N_0 , N , S , *et al.*), a change of mutation rate u has little effect on the ratio even for large S ; (iv) the effect of N on the ratio is primarily through changing S and, for given S , N has a relatively small, but not negligible, effect on the ratio; (v) compared with the neutral equilibrium case, the inbred cross case gives a rather smaller ratio when $S = 1, 5$, and 10 , a higher one when $S = 50$, which shows the dependency of the ratio on the structure of gene frequencies, not just the heritability. For very strong selection, about a quarter of the selection response can be attributed to mutation at $t = 25$, and half at $t = 50$. But the ratio is very sensitive to the change of parameters N_0 (or σ_g^2/σ_m^2) and S . The sampling variances are also very large in these examples (our replicates are ≥ 50000). The smallest coefficient of variation is 0.33. When $t \leq 50$ and/or $S \leq 20$, the coefficients of variation are hopelessly large.

In comparison with our mutation model, Hill's prediction of R^M from (1) is inflated. When selection is extremely weak, the difference between (6) with $Q_0 = 1$ and the equation (6) of Hill (1982b) is, however, very small when $t < 1/u$ and the significant difference develops only after $t > 1/u$. But for $S \geq 1$, our simulations show that significant differences

can exist even in early generations. In the inbred cross case, Hill's prediction inflates the ratio $(R_t^T - R_t^I)/R_t^T$ by roughly 1 to 5% for $20 \leq t \leq 100$ for a wide range of parameters. The inflation was greater than 10% for $t \geq 150$ in the case of $\tau = \infty$, $S = 20$, $N = 20$, $N_0 = 250$, and $u = 10^{-4}$. Nevertheless, Hill's conclusion of possible significance of mutation in artificial selection experiments is supported.

TABLE 2
 $R_t^M/R_t^T \times 100$ (\pm standard error) from simulations

<i>S</i>	<i>t</i> = 25	<i>t</i> = 50	<i>t</i> = 75	<i>t</i> = 100	<i>t</i> = 150	<i>t</i> = 200
The neutral equilibrium case: $N = 20$, $u = 10^{-5}$						
1	23.08 \pm 11.59	16.22 \pm 4.50	19.23 \pm 3.25	24.19 \pm 2.71	34.21 \pm 2.14	41.38 \pm 1.83
5	8.96 \pm 2.06	15.50 \pm 1.45	22.75 \pm 1.23	29.07 \pm 1.09	39.74 \pm 0.89	48.03 \pm 0.75
10	12.81 \pm 1.72	23.87 \pm 1.33	32.20 \pm 1.15	39.29 \pm 1.00	49.79 \pm 0.80	57.17 \pm 0.66
20	13.93 \pm 1.46	27.54 \pm 1.14	38.02 \pm 0.94	45.69 \pm 0.80	56.40 \pm 0.62	63.69 \pm 0.50
The neutral equilibrium case: $N = 100$, $u = 10^{-4}$						
1	21.18 \pm 11.86	15.47 \pm 6.33	20.28 \pm 4.42	24.13 \pm 3.53	25.61 \pm 2.56	27.99 \pm 2.09
5	19.43 \pm 4.70	22.35 \pm 1.45	23.73 \pm 1.57	26.05 \pm 1.26	31.78 \pm 0.96	37.70 \pm 0.80
20	19.06 \pm 1.66	26.68 \pm 1.04	35.58 \pm 0.81	42.77 \pm 0.67	52.78 \pm 0.52	59.41 \pm 0.43
50	25.68 \pm 0.98	43.87 \pm 0.63	54.24 \pm 0.48	60.72 \pm 0.40	68.33 \pm 0.31	72.72 \pm 0.26
The inbred cross case: $N = 100$, $u = 10^{-4}$						
1	0.67 \pm 2.40	2.21 \pm 1.74	6.13 \pm 1.44	8.91 \pm 1.27	15.22 \pm 1.05	20.97 \pm 0.92
5	3.00 \pm 0.92	6.78 \pm 0.71	11.67 \pm 0.62	16.73 \pm 0.59	27.58 \pm 0.53	36.90 \pm 0.47
20	6.57 \pm 0.72	23.82 \pm 0.70	39.46 \pm 0.60	50.15 \pm 0.49	63.17 \pm 0.34	70.19 \pm 0.26
50	26.03 \pm 0.69	53.15 \pm 0.46	65.22 \pm 0.31	71.80 \pm 0.24	78.69 \pm 0.16	82.39 \pm 0.12

This is paper no. 12552 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, N. C. 27695-7643. This investigation was supported in part by Research Grant GM 11546 from the National Institute of General Medical Sciences.

REFERENCES

- COCKERHAM, C. C. 1984. *Proc. Natl. Acad. Sci. USA* 81: 530-534.
 COCKERHAM, C. C. and TACHIDA, H. 1987. *Proc. Natl. Acad. Sci. USA* 84: 6205-6209.
 HILL, W. G. 1982a. *Proc. Natl. Acad. Sci. USA* 79: 142-145.
 HILL, W. G. 1982b. *Genet. Res.* 40: 255-278.
 HILL, W. G. and RASBASH, J. 1986. *Genet. Res.* 48: 125-131.
 KIMURA, M. and TAKAHATA, N. 1983. *Proc. Natl. Acad. Sci. USA* 80: 1048-1052.
 ZENG, Z.-B., TACHIDA, H. and COCKERHAM, C. C. 1989. *Genetics* 122: 977-984.
 ZENG, Z.-B. and COCKERHAM, C. C. 1990. *Theor. Pop. Biol.* (in press).