Theoretical and Applied Genetics 45, 268–273 (1975)

Experimental Evidence on Limitations of the Heritability Parameter[1]

H. H. MEYER and **F. D. ENFIELD**

University of Minnesota, St. Paul, Minnesota (USA)

Summary. Two-way selection was practiced at three selection intensities for 21 day pupa weight in *Tribolium castaneum*. Each of the 19 single-generation replicates of the project consisted of three populations, one each being selected upward and downward at 10, 30 and 50% selection intensities. Realized heritability for divergence was significantly lower for the least intense (50%) level than for the 10% and 30% selection intensities. Both expected heritabilities, calculated from the regression of offspring on parents, and realized heritabilities were much larger for downward than for upward selection. Expected and realized estimates of heritability were consistent within direction of selection, but estimates of heritability for divergence were of little value in predicting directional weight change. The variation observed for heritability estimates was similar to that expected from published prediction equations. The asymmetry of heritability estimates and the differential effects of selection intensity on the magnitude of realized heritability are not consistent with the usual assumption of a linear relationship between genotype and phenotype.

Introduction

The extent of agreement between expected (estimates obtained from statistical parameters) and realized heritability estimates from a given experiment is often less than desired. Since one of the main uses of heritability estimates lies in their ability to predict response to selection in plant and animal improvement programs, it is useful to have reliable estimates prior to the onset of a selection experiment for optimizing the effects of selection. Realized estimates accurately measure response to selection but are available only 'after the fact'. On the other hand, expected estimates, usually based on resemblance between relatives, can be obtained in advance of a selection program but are of practical value only if they accurately reflect realized heritabilities. On theoretical grounds, expected and realized heritabilities measure the same thing and should thus be in close agreement. An implicit assumption, of course, is the population concept of heritability, that is, that heritability of a trait within a population is a constant at any point in time and is uniform over the phenotypic range of the population. Accordingly, the expected heritability based on resemblance among relatives should be the same regardless of the portion of the distribution that is observed. Also, the relative response to selection applied in a single generation, i.e. realized heritability, is expected to be the same regardless of the direction of selection, particularly in early generations of selection in populations initially at intermediate gene frequencies, as was pointed out by Falconer (1955). With these assumptions in mind, Hill (1970, 1971, 1972) has derived the expected variances of heritability estimates derived by mid-parent-offspring regression and of realized heritabilities measured in various ways. He has also suggested experimental designs to minimize these variances.

There has been some experimental evidence indicating that, in reality, the magnitude of heritability estimates is not constant over the phenotypic range of a population. Clayton, Morris and Robertson (1957) observed higher realized heritabilities at their more intense levels of selection than at lower selection intensities. The results of Frankham, Jones and Barker (1968) indicated that the effect of selection intensity on realized heritability may be dependent on population size. Meanwhile, several experiments have shown asymmetry of response to upward and downward selection (Falconer, 1953, 1955; Gall, 1971; and others).

The experiment described herein was designed to examine the effect of selection intensity on the magnitude of realized heritability estimates. The high degree of asymmetry found in both expected and realized heritabilities for response to upward and downward selection will be discussed. Finally, the observed variances of heritability estimates obtained from the multi-replicated experiment will be compared with predicted variances based on theoretical considerations.

Materials and Methods

The experiment was based on a system of multiple replicates of single generation, two-way selections for pupa weight in populations of *Tribolium castaneum*. The use of replicated single generation selections has been used previously by Wilson (1973) and avoids confounding genetic response with inbreeding and changes in gene fre-

---

[1] Supported by NSF Grants G-1238 and GB-5987, NIH Grant GM-16074 and USPHS Training Grant GM-01156.

quency over time. By using divergent selection, the effects of environmental variation can be removed by comparing upward with downward selected lines while, at the same time, maximizing the selection differential.

The genetic foundation consisted of two highly inbred lines (CSI-5 and CSI-10) which had been produced by 38 generations of full-sib matings followed by a number of generation of mass matings prior to the start of this study. These lines had earlier served as the base for a long-term selection project currently underway. Their background and the standard laboratory cultural practices used have been described by Enfield, Comstock and Braskerud (1966).

The two inbred lines were crossed reciprocally and the resulting F_1's were likewise reciprocally crossed to produce F_2 populations. Thirty F_2 adults of each sex were mass mated to produce each of the F_3 parental populations in which selection was to be practiced. A three day egg lay period was used and 60 F_3 pupa of each sex were weighed at random from each population 21 days following the middle day of egg lay.

Selection within parental (F_3) populations was conducted at three intensities, 10, 30, and 50%, in upward and downward directions for 21 day pupa weight. The three intensities will be denoted as T-10, T-30 and T-50, respectively. Each replicate consisted of three F_3 populations randomly assigned to the three intensity levels. From the 60 pupa of each sex weighed from each parental population the appropriate proportion were selected in each direction and pair mated at random within the selected groups to form "high" and "low" lines. Sixty F_4 pupa of each sex were weighed from each line with equal numbers chosen at random from each full-sib family. Thus, each of the 19 replicates consisted of three F_3 populations, with differing proportions of parents selected from each population to produce high and low lines. Each replicate was comprised of a total of 360 F_3 and 720 F_4 individuals.

All data were analyzed on an intra-replicate, intra-treatment basis before being pooled over replicates within treatments. Selection differentials were calculated for high and low selection and for divergence of lines. Response to divergent selection was taken as the differences between F_4 progeny means of high and low lines, and directional responses were estimated for each line as the deviation of the F_4 progeny mean from the mean of the F_3 population from which the parents were selected. Realized heritabilities were then calculated as the ratios of responses to the appropriate selection differentials.

Expected or predicted heritabilities were calculated by parent-offspring regressions and pooled estimates were obtained by pooling appropriate sums of squares and cross products.

Results and Discussion

Parental populations

Means and phenotypic variances for pupa weight in F_3 populations are shown in Table 1. Average weight for both sexes was slightly lower than observed previously for the same cross (Enfield *et al.*, 1966), while variances reported here are slightly larger than earlier estimates. Mean weights within replicates differed little over the course of the experiment, indicating no substantial environmental change during that time.

Table 1. *Means and Variances for F_3 Pupa Weight*

Treatment	Males		Females	
	Mean (μg)	Phenotypic Variances[a]	Mean (μg)	Phenotypic Variances
T-10	2287	55 574	2462	76 409
T-30	2270	53 771	2448	73 923
T-50	2295	61 488	2473	78 309
Mean	2284	56 944	2461	76 214

[a] Variances pooled within populations.

Expected heritabilities

Expected heritabilities were estimated within replicates and selection lines from the regression of full-sib family progeny means on individual parent and midparent values. The results are shown in Table 2. Heritability estimates were consistently higher for downward than for upward selection, regardless of the particular parent-offspring regression used. Enfield *et al.* (1966) reported heritability estimates for upward selection somewhat larger than estimates for high lines in this experiment. They also reported that heritability estimates obtained from dam-offspring

Table 2. *Heritability Estimates Calculated from Parent-offspring Regression*

Treatment	Regression	Source of Parents for Regression					
		High selected		Low selected		All selected[a]	
		Mean	S.E.	Mean	S.E.	Mean	S. E.
T-10	Sire-son	−.44	± .24	.86	± .24		
	Dam-daughter	−.02	± .24	.26	± .30		
	Midparent-offspring	−.05	± .15	.47	± .17	.28	± .02
T-30	Sire-son	.14	± .10	.42	± .18		
	Dam-daughter	.34	± .12	.44	± .18		
	Midparent-offspring	.18	± .06	.57	± .11	.25	± .02
T-50	Sire-son	.16	± .08	.50	± .12		
	Dam-daughter	.10	± .08	.64	± .12		
	Midparent-offspring	.11	± .05	.51	± .07	.19	± .02

[a] Calculated within replicates using all (high and low) selected parents for each regression, then pooled over replicates.

regressions were significantly higher than those obtained from sire-offspring regression or sire component analysis, attributing the difference to a large nongenetic maternal effect. The data from this experiment provide no indication of such an effect, either in the parent-same sex-progeny regressions shown, or parent-opposite sex-progeny regressions (not shown).

In addition to regression analysis within high and low lines, the data were combined over lines within each replicate for a single regression using both high and low selected parents and their offspring in order to compare such results with those predicted from theoretical expectations. Such an analysis tends to minimize variance of the estimate due to the maximization of sums of squares for parents, the independent variable in regression. Heritabilities obtained in this manner were intermediate between estimates from high and low lines separately. These results will be discussed in more detail later in the paper.

Selection differentials, responses, and realized heritability

As expected, selection differentials and responses were progressively larger at more intense levels of selection. Within selection levels, however, the observed selection differentials were greater for high than for low lines in all treatments. Standardized selection differentials were obtained within replicates by dividing observed selection differentials by their respective intrapopulation standard deviations, then averaged over replicates to obtain a mean standardized selection differential for each selection direction within treatments. If one assumes a normal distribution of weights, the expected selection truncation points can be calculated and the proportion of the population selected can be estimated. The resulting estimates of percent selected were 11% high and 20% low for T-10, 30% high and 37% low for T-30, and 50% high and 53% low for T-50. The inequality of realized selection differentials suggested a high incidence of sterility for very small individuals, thus reducing the selection differential for downward selection. However, this was ruled out following a comparison of mean weights of fertile versus sterile adults in the low lines.

An examination of distributions of F_3 weights indicated the presence of a slight deviation from normality. This was detected only by virtue of the large number of populations of reasonable size available for examination. A simple indicator of skewness, the frequency of individuals more than two standard deviations above or below the mean, based on the means and standard deviations of the population from which they came, revealed a small but definite skewness in the upward direction. The frequency of extreme individuals is shown by treatment and sex in Table 3.

Table 3. *Number of Individuals more than Two Standard Deviation from the Mean*[a]

Treatment	Males		Females	
	Above	Below	Above	Below
T-10	40	14	30	7
T-30	47	14	29	12
T-50	34	9	25	6
Total	121 (3.5%)	37 (1.1%)	84 (2.5%)	25 (.7%)

[a] Based on intra-population means and variances; 1140 pupa of each sex weighed in each treatment.

The pooled estimates, over all 57 F_3 populations, were 3.5% of males above and 1.1% below the two standard deviation definition of extreme. Comparable figures for females were 2.5% high extremes with only .7% low extremes.

Two points should be made about the skewness observed here. On the one hand the degree of skewness is small, certainly not large enough to be noticeable in the majority of populations used in selection experiments and smaller than what most researchers can confidently say might be present in their populations. On the other hand, the effects of even this slight deviation from normality on realized selection differentials is obvious, particularly when selection is intense with only the most extreme individuals in each direction being selected. Given equal heritabilities in the two directions one would expect some degree of asymmetry of selection response due to the difference in selection differentials.

The observed selection differentials, responses and realized heritability estimates are shown by treatment and line in Table 4. The realized heritability for divergence was significantly lower ($P < .01$) for T-50, the least intense selection level, than for the other two treatments. For all three treatments the realized heritabilities for divergence were in excellent agreement with the midparent-offspring estimates obtained by using all parents selected, both high and low, within a replicate. Since the parent-offspring regressions had been found to differ so greatly between high and low lines (Table 2), we wished to compare those estimates with realized heritabilities for unidirectional selection. Because no unselected controls had been maintained, directional responses for high and low lines were calculated within replicates as the deviation of F_4 line means from the means of the F_3 populations from which their parents had been selected. This method fails to take into account environmental changes occurring over the one generation time span, however, replicate means of F_3 populations indicated no directional change over the course of the experiment. Unidirectional responses obtained in this way will show larger variances than divergent response estimates, but should, nevertheless, be un-

Table 4. *Mean Selection Differentials, Responses and Realized Heritabilities*

Treatment	Level	Selection differential		Response (μg)		Realized heritabilities	
		Observed (μg)	Standardized	Mean	S.E.	Mean	S.E.
T-10	High	427	1.69	70	± 21	.16	± .05
	Low	348	1.37	146	± 20	.42	± .06
	Total	775	3.06	216	± 17	.28	± .02
T-30	High	283	1.13	50	± 24	.20	± .10
	Low	250	1.00	92	± 24	.36	± .10
	Total	533	2.13	143	± 11	.28	± .02
T-50	High	207	.79	28	± 22	.14	± .11
	Low	195	.75	42	± 23	.21	± .07
	Total	402	1.54	69	± 11	.18	± .03

biased estimates, assuming variation in F_3 means was random. The resulting response estimates were consistently larger for low lines than high lines, despite larger selection differentials in the high lines. Corresponding realized heritability estimates for low lines were considerably larger than for high lines, the difference increasing over treatments as selection became more intense. When midparent-offspring and realized heritability estimates were compared, they were found in very good agreement for high selection. The two types of estimates were in poorer agreement for low selection, with realized being smaller than expected heritability, particularly at T-30 and T-50. The difference in heritability estimates between high and low lines was consistently larger for parent-offspring estimates than for realized heritabilities.

Previous estimates of realized heritability for pupa weight in *Tribolium* have generally come from selection experiments for increased pupa weight. Enfield *et al.* (1966) and Goodwill (1974), using stocks from the same inbred lines as used in this experiment, reported estimates of .35 and .10, respectively, for upward selection compared to a mean estimate of .17 in this experiment. Gall (1971) reported mean realized heritabilities of .37 for upword change and .26 for downward selection. His larger realized heritability for upward than for downward change is in contrast to our findings and to the reports by Falconer (1953, 1955, 1960) for several traits in mice. It should be noted that his genetic stock was from a wild type foundation with gene frequencies and previous selection history unknown.

Predicted and observed variances of heritability estimates

Hill (1970, 1971, 1972) investigated the expected variances of both midparent-offspring and realized estimates of heritability and presented guidelines for obtaining minimum variance estimates of each type. The procedure recommended for parent-offspring estimation is the selection of extreme parents at both ends of the distribution to maximize the sums of squares of the independent variable. Minimum variance estimates for a given total number measured would be obtained by selecting approximately 10% at each end of the distribution and measuring about 10 progeny per mating. Assortative mating within selected groups reduces the variance of heritability estimates, but this effect diminishes as selection becomes more intense.

Using formula presented by Hill (1970), expected variances of midparent-offspring heritability estimates were calculated for each treatment using observed values of selection intensities, mean family size, etc. These predicted variances are shown in Table 5 for a range of heritabilities along with the observed variances and variances expected for the optimal design (percent of parents selected and number of offspring measured) in each treatment. Variance estimates increased from T-10 to T-50 as expected since the proportion of parents selected becomes progressively

Table 5. *Expected*[a] *and Observed Variance of Midparent-Offspring Regression Estimates of Heritability*

	Selection level								
	T-10			T-30			T-50		
$\hat{h}^2$	.2	.3	.4	.2	.3	.4	.2	.3	.4
Expected $\sqrt{\sigma^2_{\hat{h}^2}}$	.061	.066	.069	.067	.071	.072	.080	.082	.083
Observed h^2		.283			.263			.194	
Observed $\sqrt{\sigma^2_{h^2}}$		.076			.079			.089	

[a] Calculated by the methods of Hill (1970). For optimal design, expected $\sqrt{\sigma^2_{\hat{h}^2}}$ is .055 for $\hat{h}^2 = .2$ and .059 for $\hat{h}^2 = .4$.

farther removed from the optimal design as selection intensity decreases. For T-10, the design closest to optimum, the expected variance was 11% greater than for the optimum design.

Standard errors of the parent-offspring regression coefficients were available from each replicate within a treatment. These standard errors were averaged within treatments and the means found to be uniformly 11% larger than expected from the prediction equation. The prediction equation appears quite satisfactory for estimating variances of heritability estimates, at least for the heritability of divergence in a trait.

A similar comparison was made for expected and observed variances of realized heritability estimates (Hill, 1972 equation). Again the expected and observed standard deviations for heritability estimates were in close agreement (Table 6).

Table 6. *Expected*[a] *and Observed Variance of Realized Heritability Estimates*

	Selection level		
	T-10	T-30	T-50
Expected $\sqrt{\sigma^2_{\hat{h}^2_r}}$	.095	.090	.120
Observed $\sqrt{\sigma^2_{h^2_r}}$	.081	.105	.127

[a] Calculated by the methods of Hill (1972).

In general, it appears that the estimation equations give accurate estimates of the variation to be expected in heritability estimates when selection is practiced for genetic divergence. Likewise, the techniques recommended yield heritability estimates from parent-offspring regressions which agree very well with realized heritabilities for divergence. However, if the data from this experiment is typical of what might be found generally, i.e. different heritabilities in different parts of the distribution, it limits the practical value of the Hill approach. Heritability estimates for predicting rate of response for unidirectional changes would then best be obtained from parent-offspring regression using parents at the end of the distribution in which change is to be made. The variances of such estimates are larger than if parents at both ends are selected, but at least the parameter estimated would be relevant to the selection project to be carried out.

The marked asymmetry of response to selection observed in this experiment was not expected and is certainly not easily explained. This is particularly true in light of the initial gene frequencies being near .5 for segregating genes, the large number of genes affecting this trait (Enfield, 1973), and the fact that the asymmetry appeared in the first generation of selection and thus does not reflect gene frequency changes due to prior selection or inbreeding. Falconer (1953, 1955, 1960) has repeatedly reported asymmetry of response to divergent selection for several traits in mice. He has suggested, as possible genetic explanations, directional dominance of genes affecting the trait, unequal frequencies of favorable and unfavorable alleles and inbreeding depression, citing the last of these as the most likely reason for asymmetry in his experiments. In view of the genetic conditions listed above, none of these suggested causes of asymmetry are satisfactory explanations for the results observed in this experiment. Neither is the skewness observed in parental populations an explanation, since it was in the upward direction while the greatest selection response was in the downward direction. Although the asymmetry of response was compatible with the unequal heritability estimates derived from parent-offspring regression within high and low lines, the reason for inequality of heritabilities in the two directions is no less a mystery.

The original intent of this experiment, to observe the effect of variation in selection intensity on realized heritability, has become somewhat obscured by the discovery of marked asymmetry of heritability estimates. This difficulty not being anticipated, we utilized a design which would be powerful in detecting small differences in realized heritability. This it clearly did, the standard errors of intra-treatment realized heritability estimates being very small and capable of detecting small differences between estimates. The realized heritability for divergence, which is essentially the mean of realized heritabilities for high and low selection when selection differentials are similar for the two directions, appears to decline at the 50% level because of lower estimates in both high and low directions. Realized heritability estimates in the low direction become progressively smaller as the proportion selected becomes larger. No trend is apparent over treatments in the high direction of selection. From this experiment it appears that progress relative to selection intensity is greater for downward than for upward selection, and that the more intense the selection in the downward direction, the larger the realized heritability.

Both the observed asymmetry of heritability estimates and the differential effect of selection intensity on realized heritability argue for limitations in the current quantitative genetic theory. Neither of these anomalies is expected or explainable on the basis of genetic models presupposed for quantitative traits. More work is obviously needed in comparing heritabilities at opposite ends of the phenotypic distribution to see if the asymmetry phenomenon observed here is widespread and, if so, the types of traits most likely to display substantially different heritabilities for upward and downward change. Additional work may also then be needed in optimizing designs for heritability estimation of single directional change in quantitative traits.

Literature

Clayton, G. A., Morris, J. A., Robertson, A.: An experimental check on quantitative genetic theory. I. Short-term responses to selection. J. Genet. **55**, 131–151 (1957).

Enfield, F. D.: Long term selection for pupa weight in *Tribolium*: gene action and gene number. Genetics **74** (Suppl.), S72–S73 (1973).

Enfield, F. D., Comstock, R. E., Braskerud, O.: Selection for pupa weight in *Tribolium castaneum* I. Parameters in base populations. Genetics **54**, 523–533 (1966).

Falconer, D. S.: Selection for large and small size in mice. J. Genet. **51**, 470–501 (1953).

Falconer, D. S.: Patterns of response in selection experiments with mice. Cold Spr. Harb. Symp. Quant. Biol. **20**, 178–196 (1955).

Falconer, D. S.: The genetics of litter size in mice. J. Cell. and Comp. Physiol. **56** (Suppl. 1), 153–167 (1960).

Frankham, R., Jones, L. P., Barker, J. S. F.: The effects of population size and selection intensity in selection for a quantitative character in *Drosophila*. I. Short-term response to selection. Genet. Res. **12**, 237–248 (1968).

Gall, G. A. E.: Replicated selection for 21-day pupa weight of *Tribolium castaneum*. Theoret. Appl. Genetics **41**, 164–173 (1971).

Goodwill, R.: Comparisons of three selection programs using *Tribolium castaneum*. J. Heredity **65**, 8–14 (1974).

Hill, W. G.: Design of experiments to estimate heritability by regression of offspring on selected parents. Biometrics **26**, 566–571 (1970).

Hill, W. G.: Design and efficiency of selection experiments for estimating genetic parameters. Biometrics **27**, 293–311 (1971).

Hill, W. G.: Estimation of realized heritabilities from selection experiments. I. Divergent selection. Biometrics **28**, 747–765 (1972).

Wilson, S. P.: Experimental comparisons of individual versus family selection systems. II. Single generation tests. 22nd National Breeders Roundtable, 45–60 (1973).

Received July 14, 1974
Communicated by H. Abplanalp

Dr. F. D. Enfield
H. H. Meyer
Department of Genetics and Cell Biology
University of Minnesota
St. Paul, Minn. 55108 (USA)