Original article

Estimation of heritability in the base population when only records from later generations are available

L Gomez-Raya, LR Schaeffer, EB Burnside

University of Guelph, Centre for Genetic Improvement of Livestock Animal and Poultry Science , Guelph, Ontario, Canada N1G 2W1

Summary – The genetic variance and heritability of a quantitative trait decrease under directional selection due to the generation of linkage (gametic phase) disequilibrium. After a few cycles of directional selection in a population of infinite sire a steady-state equilibrium is approached. At this point there is no further reduction in these parameters since the disequilibrium generated by selection is offset by free recombination. In many situations records available to estimate genetic parameters come from populations at the steady-state equilibrium. A simple method to obtain estimates of genetic variance and heritability in the base population using estimates of these parameters at the equilibrium is described. The method makes use of knowledge of the effect of repeated cycles of selection on genetic variance and heritability to infer the base population parameters.

genetic variance / heritability / estimation of genetic parameters / linkage disequilibrium

Résumé – **Estimation de l'héritabilité dans la population initiale en utilisant seulement les données des générations subséquentes.** *Lorsqu'il y a sélection directionnelle sur un caractère quantitatif, la variance génétique et l'héritabilité sont réduites à la suite de la formation d'un déséquilibre de liaison (phase gamétique). Après quelques cycles de sélection directionnelle dans une population de taille infinie, un équilibre stable est atteint. À partir de ce moment, il n'y a plus aucune réduction de ces paramètres puisque le déséquilibre créé par la sélection est compensé par la recombinaison. Dans plusieurs situations, les données disponibles pour estimer les paramètres génétiques proviennent de populations en équilibre stable. Une méthode simple d'estimation de la variance génétique et de l'héritabilité dans la population initiale est présentée. Cette méthode tient compte de l'effet d'une succession de cycles de sélection sur la variance génétique et l'héritabilité pour inférer la valeur de ces paramètres dans la population initiale.*

variance génétique / héritabilité / estimation des paramètres génétiques / déséquilibre de liaison

INTRODUCTION

The estimation of genetic variances and heritabilities of quantitative traits in populations under artificial or natural selection is a common objective in animal breeding and evolutionary biology of natural populations. Standard methods to estimate genetic variances and heritabilities when information is available on the parents and offspring are the correlation among sib and the regression of offspring on parents (Falconer, 1989). Analysis of variance of half-sib yields biased estimates of heritability if the parents are a selected sample from the population (Robertson, 1977; Ponzoni and James, 1978). Unbiased estimates of heritability by half-sib correlation can be obtained after correcting for the bias induced by selection of sires (Gomez-Raya *et al*, 1991). Regression of offspring on parents is not altered by selection of animals to be parents (Pearson, 1903) and therefore estimates of heritability by regression are unbiased (Robertson, 1977). In both, half-sib and regression analyses, unbiased estimates of heritability are obtained after one cycle of selection. Regression estimates of heritability are not unbiased for the accumulated reduction in genetic variance after repeated cycles of selection (Fimland, 1979). The changes in genetic variance under selection were described by Lush (1945) using genetical arguments and a numerical example. Bulmer (1971) formally established the theory to explain the changes in the genetic variance under continued cycles selection. Under the assumption of an infinitesimal gene effect model the genetic variance and heritability are reduced due to the build-up of linkage (gametic phase) disequilibrium in a population of infinite size and with discrete generations. After only a few cycles of directional or stabilizing selection a limiting or steady-state equilibrium value for these parameters is approached. At this point the new disequilibrium generated by the selection of parents is offset by free recombination. Most animal populations are probably in the steady-state or close to it since the equilibrium is approached very quickly. The use of standard methods to estimate genetic variance and heritability yields estimates of these parameters in the limit situation. However, in many cases, interest is on the parameters in the non-selected base population. Sorensen and Kennedy (1984) have shown that mixed model methodology may be used to estimate the genetic variance and heritability in the base population. They carried out a simulation experiment for several cycles of mass selection and then proceeded to estimate genetic variances using a minimum variance quadratic unbiased estimator (MIVQUE) under the correct model. They found close agreement between observed and simulated parameters. The requirement of using the correct model implies making use of the relationship matrix with complete pedigree information back to the base population. Natural populations are currently under selection and pedigree information is not known. In livestock species, such as dairy cattle, pedigree information is only recorded from the later years. In general, mixed model methodology requires the genetic variance of the base generation as determined by the available data and corresponding pedigree information. Therefore, if the available data and pedigrees are only for animals at the point of selection equilibrium, then the genetic variance at selection equilibrium is needed to evaluate animals by mixed model methods. However, knowledge of genetic variance in the base population (prior to starting selection) is necessary to predict response to alternative breeding programmes in which selection intensity

and/or accuracy of evaluation differ from those in the current breeding programme. Any changes in those parameters alter the amount of disequilibrium maintained in the population. After a few cycles of selection a new equilibrium will be approached which can be predicted with knowledge of new selection intensity, new accuracy of evaluation, and the genetic variance in the base population.

The objective of this paper is to describe a method to estimate base population genetic variance and heritability from data available at the steady-state equilibrium. Use is made of effect of repeated cycles of selection on genetic variance and heritability. Assuming the population is at the equilibrium, the base population parameters are obtained by reversing Bulmer's arguments.

THEORY

Consider an additive infinitesimal gene effect model. The trait under selection is determined by a very large number of loci with recombination rates of 1/2. Assume that selection intensity is constant across discrete generations and that each individual belonging to the same sex is evaluated with the same accuracy. Population size is infinite. Selection is by truncation. Assume that there are no departures from normality after selection (Bulmer, 1980).

The basic theory to explain the changes in genetic variance in populations undergoing selection was first given by Bulmer (1971). The breeding value of an individual i in a given generation is:

$$a_i = 1/2a_s + 1/2a_D + e_i$$

where a_s and a_D are the breeding values of the sire and dam respectively and e_i is the mendelian sampling effect in individual i. e_i is distributed normally with variance $((1/2)\ \sigma^2_{A_0})$ in a population of infinite size where $\sigma^2_{A_0}$ is the genetic variance in the base population. The genetic variance in the selected group of parents is reduced by kr^2 (Pearson, 1903), where r is the accuracy of selection and $k = (\phi(x)/p)((\phi(x)/p) - x)$ for selection of the top ranking individuals (directional selection) and $k = 2x(\phi(x)/p)$ for selection of the middle ranking individuals (stabilizing selection); x = standard normal deviate; $\phi(x)$ = ordinate at cutoff points for p = proportion selected. The genetic variance in the offspring can be partitioned into between and within family components. The within-family variance is not affected by selection of parents and has value $((1/2)\sigma^2_{A_0})$. This is true on the assumptions of a very large number of loci and infinite population size, ie no change in the gene frequencies of the segregating loci for the trait. The between-family variance has a value of $(1 - kr^2_{t-1})(1/2)\sigma^2_{A_{t-1}}$, where $\sigma^2_{A_{t-1}}$ is the genotypic variance in the previous generation. Therefore, the genetic variance in a given generation, t, assuming different selection intensities and accuracy of selection in the 2 sexes is:

$$\sigma^2_{A_t} = (1/2)\sigma^2_{A_{t-1}}[1 - 0.5k_s r^2_{s_{t-1}} - 0.5k_D r^2_{D_{t-1}}] + (1/2)\sigma^2_{A_0}$$

where $r_{s_{t-1}}$ = accuracy of selection of sires in generation $t-1$; $r_{D_{t-1}}$ = accuracy of selection of dams in generation $t-1$; k_s and k_D are the values of parameter k for sires and dams, respectively.

At the limit there are no further changes in the genetic variance since the new disequilibrium generated in that generation is compensated for by free recombination. Then, genetic variance becomes:

$$\sigma^2_{A_L} = (1/2)\sigma^2_{A_L}[1 - 0.5k_sr^2_{s_L} - 0.5k_Dr^2_{D_L}] + (1/2)\sigma^2_{A_0}$$

where $\sigma^2_{A_t} = \sigma^2_{A_{t-1}} = \sigma^2_{A_L}$.

After some algebraic manipulation this reduces to:

$$\sigma^2_{A_L} = \frac{\sigma^2_{A_0}}{[1 + 0.5k_sr^2_{s_L} + 0.5k_Dr^2_{D_L}]} \qquad [1]$$

Assuming constant environmental variance across generations and substituting expression [1] in the standard formula of heritability, the heritability at the equilibrium limit is:

$$h^2_L = \frac{h^2_0}{1 + (1 - h^2_0)[0.5k_sr^2_{s_L} + 0.5k_Dr^2_{D_L}]} \qquad [2]$$

If the population is at the steady-state equilibrium and records are available to estimate genetic variance, then estimates of base population parameters can be found by solving expressions [1] and [2] for $\sigma^2_{A_0}$ and h^2_0, respectively, and by substituting true values by their estimates. Thus, genetic variance and heritability in the base population can be obtained by:

$$\widehat{\sigma}^2_{A_0} = \widehat{\sigma}^2_{A_L}[1 + 0.5(k_s\widehat{r}^2_{s_L} + k_D\widehat{r}^2_{D_L})] \qquad [3]$$

$$\widehat{h}^2_0 = \frac{[1 + 0.5(k_s\widehat{r}^2_{s_L} + k_D\widehat{r}^2_{D_L})]\widehat{h}^2_L}{1 + [0.5(k_s\widehat{r}^2_{s_L} + k_D\widehat{r}^2_{D_L})\widehat{h}^2_L]} \qquad [4]$$

where "$\widehat{}$" denotes estimate.

If selection criterion is the individual phenotype then $\widehat{r}^2_{s_L} = \widehat{r}^2_{D_L} = \widehat{h}^2_L$ and expressions [3] and [4] reduce to:

$$\widehat{\sigma}^2_{A_0} = \widehat{\sigma}^2_{A_L}[1 + k\widehat{h}^2_L], \qquad [5]$$

and

$$\widehat{h}^2_0 = \frac{[1 + k\widehat{h}^2_L]\widehat{h}^2_L}{[1 + k\widehat{h}^4_L]}, \qquad [6]$$

respectively. In these expressions $k = 0.5k_s + 0.5k_D$. The required estimates of the genetic variance and heritability at equilibrium can be obtained by either regression or maximum likelihood methods. It is generally accepted that maximum likelihood estimates of genetic variances are unbiased by selection of parents if all the pedigree

information is included in the analysis. If REML (restricted maximum likelihood) account for selection, say, in generations 0 to 6, then it will also account for selection in generations 6 to 10 when only data from these generations are available. In the former case, the component of variance estimates the genetic variance in the base population, and in the latter the genetic variance in generation 6 which it is assumed to be the equilibrium genetic variance.

The approximate sampling variance of the estimate of heritability in the base population can be obtained by differentiating expression [6] with respect to $\widehat{h}_L^2$,

$$
\begin{aligned}
\mathrm{Var}\,(\widehat{h}_0^2) &\approx \left[\frac{d(\widehat{h}_0^2)}{d\widehat{h}_L^2}\right]^2 \quad \mathrm{Var}\,(\widehat{h}_L^2) \\
&\approx \left[\frac{(1+k\widehat{h}_L^4)(1+2k\widehat{h}_L^2)-(2k\widehat{h}_L^4)(1+k\widehat{h}_L^2)}{(1+k\widehat{h}_L^4)^2}\right]^2 \quad \mathrm{Var}\,(\widehat{h}_L^2) \\
&\approx f\,\mathrm{Var}(\widehat{h}_L^2)
\end{aligned}
$$

Therefore, the sampling variance of the estimate of heritability in the base population depends on a factor f, which is a function of h_0^2 and k because under phenotypic selection h_L^2 depends only on h_0^2 and k, and on the sampling variance of $\widehat{h}_L^2$. Values of the f factor for different $\widehat{h}_0^2$ are represented in figure 1 for varying selected percentages (p) 50%, 20%, 10%, and 1%. The value of h_L^2 was obtained by solving expression [6] as a function of known h_0^2:

$$
h_L^2 = \frac{-1+\sqrt{1+4h_0^2 k(1-h_0^2)}}{2k(1-h_0^2)}
$$

as described by Gomez-Raya and Burnside (1990). For traits with heritability values less than 0.70, f is larger than 1 and therefore the sampling variance of $\widehat{h}_0^2$ will be increased with respect to the sampling variance of the estimates at the limit (Var $(\widehat{h}_L^2)$). Selection intensity appears to have small effect on f.

In practical animal breeding, the performance of relatives can be used to maximize response by the use of selection indices. For example, consider a population where sires are selected on the average of records of d daughters each with one record and dams are selected on the average of n records each. Estimates of heritability in the base population can be obtained by substituting in expression [4] the appropriate equilibrium values of accuracy for sires $\widehat{r}_{s_L} = [d\widehat{h}_L^2/(4+(d-1)\widehat{h}_L^2)]^{1/2}$ and dams $\widehat{r}_{D_L} = [n\widehat{h}_L^2/(1+(n-1)r\widehat{e}p_L)]^{1/2}$, where $r\widehat{e}p_L = [(\widehat{\sigma}_{A_L}^2+\widehat{\sigma}_{PE}^2)/(\widehat{\sigma}_{A_L}^2+\widehat{\sigma}_{PE}^2+\widehat{\sigma}_{TE}^2)]$, $\widehat{\sigma}_{PE}^2$ = estimated permanent environmental variance and $\widehat{\sigma}_{TE}^2$ = estimated temporary environmental variance.
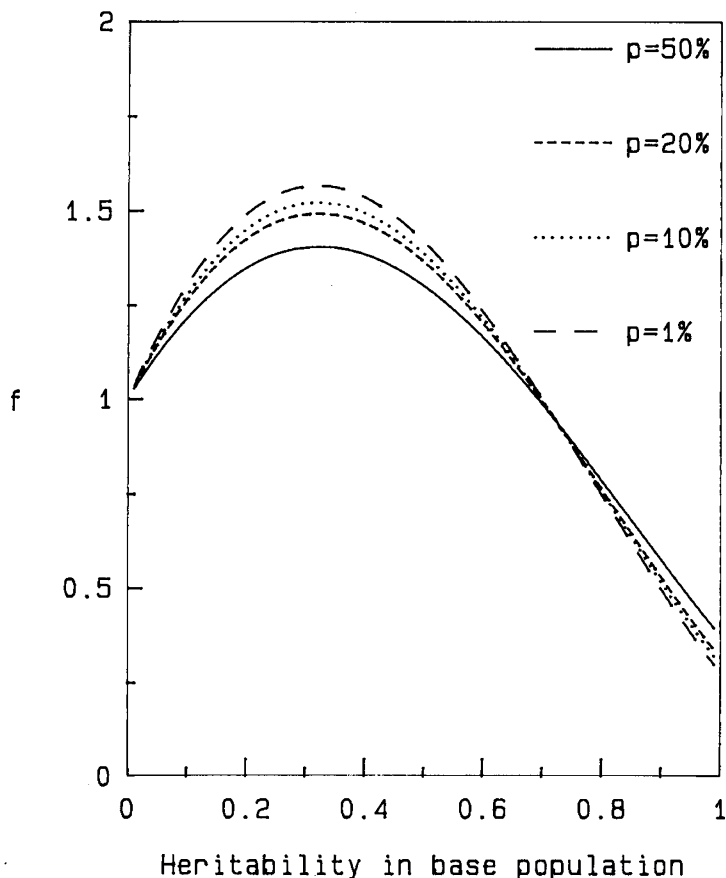
Fig 1. Changes in the value of f with heritability in the base population for varying selection intensities corresponding to selected percentages (p) 50%, 20%, 10% and 1%.

DISCUSSION

In this paper a method to estimate heritability in the base population from data at the steady-state equilibrium is presented. The method to obtain estimates of parameters at the equilibrium is assumed to be unbiased by selection of parents in that particular generation. Estimates of heritability by regression of offspring on parents is unbiased by selection in a given generation (Robertson, 1977). Estimation of heritability by half-sib correlation is biased by selection of sires (Robertson, 1977; Ponzoni and James, 1978), but estimates can be corrected (Gomez-Raya *et al*, 1991), and then final estimates free of selection bias can be obtained. Another alternative is to use the method given by Sorensen and Kennedy (1984). They proposed the estimation of genetic variance in later generations using the MIVQUE algorithm and assuming that individuals in the generation in question are unrelated. In the same

paper they carried out a simulation experiment to test the validity of this method. In generation 7 actual genetic variance had decreased from 10 to 8.41. The simulated environmental variance was 10, so heritability at the limit was 0.457, assuming that environmental variance was known without error. The percentage selected in males was 50% ($k_s = 0.637$) in each generation. Dams were not selected ($k_D = 0$). Using expression [6] after substituting estimated with true parameter values and corresponding values of h_L^2, k_s and k_D the heritability in the base population is expected to be 0.491, which is very close to the simulated heritability in the base population (0.50). On the other hand, Van der Werf (1990) carried out 2 different simulation experiments in which mass selection was practised on males at different selection intensities corresponding to percentage selected $p = 10\%$ and $p = 25\%$. He proceeded to estimate components of variance using REML (restricted maximum likelihood) and the data from generations 4 and 5 with pedigree information known back to generation 3. Treating sires as random in the model he obtained biased estimates (8.58 for $p = 10\%$ and 8.71 for $p = 25\%$) of the base population genetic variance (10). If we assume that the population is at the steady-state equilibrium in generation 3 then genetic variance in the base population can be estimated using [5] after substituting appropriates values of $k(k_s = 0.830$ for $p = 10\%$ and $k_s = 0.759$ for $p = 25\%$; $k_D = 0$) and $\widehat{h}_L^2$ (0.45 for $p = 10\%$ and 0.46 for $p = 25\%$). The values of $\widehat{h}_L^2$ can be obtained from the estimates of genetic (8.58 for $p = 10\%$ and 8.71 for $p = 25\%$) and residual variances (10.44 for $p = 10\%$ and 10.17 for $p = 25\%$) given by Van der Werf (1990) in table II. Proceeding in this way, estimates of the genetic variance in the base population are 10.18 ($p = 10\%$) and 10.23 ($p = 25\%$). These values are very close to the simulated genetic variance in the base population (10). The slight discrepancy, in these studies, occurs because the formulae derived in this paper have not taken into account the effect of inbreeding in the reduction of genetic variance. Throughout this paper, population size has been assumed infinite, and therefore, inbreeding effects on genetic variance were not considered. Both natural and livestock populations are finite. The reduction in genetic variance due to the build-up of linkage disequilibrium occurs rapidly in the first generations whereas inbreeding effect is small but accumulates gradually in later generations. After the steady-state equilibrium is achieved, the genetic variance reduces gradually due to inbreeding and so does the amount of linkage disequilibrium maintained in the population. Thus, correction for selection at this point would not yield estimates of the genetic variance and heritability in the original base population. Rather, the estimates of these parameters would be those obtained after relaxing selection for several generations, in other words, the genetic variance due to the gene frequencies segregating in the population at the generation in question. In most situations, these are the parameters of interest because they explain how much genetic variability could be used in selection programmes. Prediction of the joint effects of inbreeding and selection on genetic variance is rather difficult (Robertson, 1961; Verrier *et al*, 1990; Wray and Thompson, 1990).

The method described in this paper is able to correct for the bias generated after repeated cycles of selection assuming equal information on each individual evaluated and constant selection intensity across generations. In practice, both assumptions may not hold. Methods to estimate breeding values such as best linear unbiased predictor (BLUP) are preferred to selection index in the improvement

of livestock. Each individual breeding value has a different accuracy in BLUP evaluations. If pedigree information is known back to the base population then mixed model methodology could be used (Sorensen and Kennedy, 1984). The effect of different accuracies among selection candidates on genetic variance is not known. Further work is needed to incorporate this kind of selection in the method presented in this paper. Changes in selection intensity across generations result in changes in the disequilibrium in the population parameters over time. In natural populations, selection intensity could oscillate due to changes in the pattern of interaction among species and/or environmental fluctuations. In livestock populations, selection intensity may fluctuate due to changes in production system or in market conditions. Therefore, the procedure described in this paper would give only approximate values of the base population heritability. However, oscillation in the selection intensity across generations has small effect on the estimation of heritability in the base population because the parameter k changes only very slightly with selection intensity. For example, if we use a wrong value of selection intensity corresponding to selection of the top 1% ($k_s = 0.903$; $k_D = 0$) in the simulation experiment of Sorensen and Kennedy (1984), then heritability in the base population after using expression [6] is 0.504. This value is again very close to the simulated heritability (0.50). Therefore, even though selection intensity is not constant across generations the method described in this paper could be used to estimate, in a very approximate manner, the value of heritability in the base population.

ACKNOWLEDGMENTS

REFERENCES

Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105, 201-211

Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford

Falconer DS (1989) *Introduction to Quantitative Genetics.* Longman Press, Essex, 3rd edn

Fimland E (1979) The effect of selection on additive genetic parameters. *Z Tierz Zuchtungsbiol* 96, 120-134

Gomez-Raya L, Burnside EB (1990) The effect of repeated cycles of selection on genetic variance, heritability, and response. *Theor Appl Genet* 79, 568-574

Gomez-Raya L, Schaeffer LR, Burnside EB (1991) Selection of sires to reduce sampling variance in the estimates of heritability by half-sib correlation. *Theor Appl Genet* 81,624-628

Lush JL (1945) *Animal Breeding Plans.* The Iowa State University Press, Ames, IA, 3rd edn

Pearson K (1903) Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil Trans R Soc Lond Ser A* 200, 1-66

Ponzoni RW, James JW (1978) Possible biases in heritability estimates from intraclass correlation. *Theor Appl Genet* 53, 25-27

Robertson A (1961) Inbreeding in artificial selection programmes. *Genet Res* 2, 189-194

Robertson A (1977) The effect of selection on the estimation of genetic parameters. *Z Tierz Zuchtungsbiol* 94, 131-135

Sorensen DA, Kennedy BW (1984) Estimation of genetic variances from unselected and selected populations. *J Anim Sci* 59, 1213-1223

Van der Werf JHJ (1990) Models to estimate genetic parameters in crossbred dairy cattle populations under selection. Doctoral thesis, Dept Anim Breeding, Agric Univ, Wageningen, The Netherlands, ch 5

Verrier E, Colleau JJ, Foulley JL (1990) Predicting cumulated response to directional selection in finite panmictic populations. *Theor Appl Genet* 79, 833-840

Wray NR, Thompson R (1990) Prediction of rates of inbreeding in selected populations. *Genet Res* 55, 41-54