*Jour. Ind. Soc. Ag. Statistics*
*51(2&3), 1998 : 147-156*

# Quantitative Trait Loci: Polygenes Revisited and the Future of Animal Breeding

O. Mayo and I. R. Franklin
*CSIRO, Australia*

SUMMARY

The number of genes that influence a quantitative trait has been important for the strategy of animal improvement only in certain specific cases: incorporation of genetic material from one strain or breed into another, assessing the ultimate selection response; assessing the significance of gene interaction. With the accurate mapping of the whole genome for major species a goal now well within reach, the assumptions underlying quantitative genetics should be critically re-examined.

*Keywords:* Marker assisted selection (MAS), Pure infinitesimal model, Epistatic interaction, Genetic load, Genomic scanning.

## *1. Introduction*

The idea that a major gene can influence the magnitude of a quantitative trait is as old as the science of genetics. Mendel [22] found the first quantitative trait locus (QTL), a major gene that affected the height of garden peas but it was not a polygene, in the sense considered over the last half century. Polygenes are the many genes of individually small effect that are frequently assumed to influence a quantitative trait.

Fisher identified the possibility of a very large number of polygenes in 1918 but also showed that the model did not need an indefinitely large number of genes. Castle and Wright provided the first means of estimating the minimum number of genes that might affect a quantitative trait (see Mayo and Hopkins [19] for references and discussion). "Student" [31] provided the first estimation from a selection experiment of the number of genes affecting a trait and found that there were very many. Panse [25] provided the next major advance, in showing not only how the minimum number of genes (the effective number of factors) might be more precisely estimated but also in showing how such estimators depended on the magnitude of gene effects and the motivation of the genes. It is worthwhile quoting his analysis at length; his data came from two crosses between cotton lines selected for different production traits:

If the $F_2$ variance is $(a + b + \ldots)$, the mean variance with $F_3$ progenies is $1/2(a + b + \ldots) = V_3$, and the variance of this variance is $1/4\ (a^2 + b^2 + \ldots) = V(V_3)$.

Assuming the segregation in $F_2$ to be due to n factors all with equal variance a, then,

$$V_3 = \frac{nx}{2} \text{ and } V(V_3) = \frac{nx^2}{4}$$

$$\therefore \qquad V_3^2/V(V_3) = \frac{n^2}{4} x^2 / \frac{n}{4} x^2 = n$$

where n the number of factors, hypothetically with equal variance and without linkage, can be termed the "effective" number of factors. This number can thus be calculated if the mean genotypic variance within $F_3$ progenies is known. These quantities are shown below for the staple-length data assuming the genotypic variance to be the same as genetic.

| Cross | $V_3$ (= 1/2 genotypic variance in $F_2$) | $V(V_3)$ | $V_3^2/V(V_3) = n$ |
|---|---|---|---|
| C 520 × Bani | 0.772 | 0.363 | 1.64 |
| C 529 × Malvi | 0.788 | 0.224 | 2.77 |

The ratio $V_3^2/V(V_3)$ calculated from the table indicates on the same scale as in the whole $F_3$ population, the "effective" number of factors operating in the selected portion of the population. Its calculated values, which are given below, show the greatest reduction in heterozygosity when dominance is absent, and in its presence a great reduction with three factors than with an infinite number.

| | System of factors | $V_3^2/V(V_3)$ |
|---|---|---|
| I | No dominance, geometric series | 2.210 |
| II | No dominance, three equal factors | 2.265 |
| III | Balanced dominance, geometric series | 2.835 |
| IV | Balanced dominance, three factors with equal variances | 2.586 |
| V | Balanced dominance, three equal factors | 2.611 |

At that time, the number of genes that might affect a trait was thought to be of considerable importance, but there was no way of going past the statistical estimate of the minimum number of genes to determine how this related to the actual number of genes affecting a trait. Panse used as one of his cases a geometric model of magnitude of gene effect because it was mathematically tractable (see Mayo, Eckert and Nugroho [21] for discussion). There is, as yet, no theoretical basis for adopting any particular model for gene effects, nor indeed a basis for each model.

Panse [26] further showed how selection in the $F_2$ would result in different outcomes that depended on the number of genes, their relative magnitude and their mode of action. Unfortunately this work has not been followed up.

It is noteworthy that the very effective technology of plant and animal breeding, although it uses major genes where possible, has been built largely on Fisher's original model of an indefinitely large number of genes. Major genes influencing quantitative traits are likely to be found in unselected populations (e.g. Mayo and Hancock [18]), but are expected to have been fixed in highly selected breeds or strains. How many genes do affect quantitative traits, whatever their magnitude, and for what cases is this an important consideration?

### 2. *Introgression*

The incorporation of a gene or genes from one strain (in animals) or even a different species in plants is usually carried out by repeated back-crossing, though other designs are possible (e.g. Fisher's, as discussed by Morton [23]). If the aim is to include a specific major gene and nothing else from one strain into another strain, then probabilistic theory tells us how many generations we need to be confident at any chosen level of probability that we have incorporated just the gene we want and no other genetic material. Where we wish to incorporate a chromosomal region or regions, the situation is more complex and breeding, even with the aid of many DNA markers, is slow and inefficient (Koudande, Thomson and van Arendonk [15]).

If many genes affect a trait and they are widely dispersed, then incorporation of new variation in the manner of Fisher mentioned above will be slow.

It is noteworthy that none of the QTL for prolificacy in sheep, from the Booroola gene (Piper *et al* [27]) to the Cambridge gene (Ap Dewi *et al* [2]), has yet been mapped to within less than 2 cm, let alone identified.

This has meant that development of a new, high prolificacy meant breed, the Booroola BL, has had to follow the traditional path while DNA markers have been under development.

### *3. QTL Detection and Measurement*

There are two general strategies for detecting QTL in commercially significant species. One is to search, in pedigreed populations, for associations between genetic variations at a marker locus and a phenotypic variation for a trait sampled from the proposed target population. This strategy is true to the utilitarian goals of marker use, but suffers from the disadvantage that alleles with potentially large effects on a trait may be rare, and hence not detected. The second is to cross lines, breeds or even species that differ substantially for the traits of interest, and then to identify chromosomal regions that contribute to the differences between the parents. This second strategy is commonly used, and is thought to maximise the chance of finding major QTL. Such crosses may well be useful for studies of the genetic architecture of a quantitative trait, but it does not follow that identifying loci that are associated with differences accumulated over a long selection history will yield major loci of use in a marker assisted selection (MAS) program. It also does not seem to have been widely recognised that all of the designs used (back-crosses, half-sibs, $F_3$, grand-daughter) allow only four genotypic affects to be estimated for two loci, whereas ten possible genotypes exist for two linked biallelic genes. This is a fundamental limitation.

We (Mayo and Franklin [17]) have examined the potential errors in QTL detection by choosing an extreme model of multiple QTL effects, the "Pure infinitesimal model" in which the parental differences are spread uniformly over the entire genome. In this case, the effective length of the chromosomal segment of a given parental origin is a function only of the length of that segment.

Franklin and Mayo [12] have shown that when the infinitesimal model holds large QTL effects may be detected despite the fact, that no large QTL are, in fact present (see also Wilcox, Richardson and Carson [33]). Our study highlights three important results for QTL detection. First, analytical methods based on the assumption that a segment-defined by flanking markers has no effect on the trait are questionable if the parental lines differ substantially. Secondly, the methods are not robust to assumptions of multiple QTL and users of the commonly available software packages should exercise considerable caution unless there is ancillary evidence that only few loci are involved.

Thirdly, the genetic model chosen as the null hypothesis affects the outcome of any application of the methodology.

Although QTL investigations are much simpler in plants, Kearsey and Farquhar [13] have raised concerns about the imprecision of the analytical tools that reinforce those expressed here.

### *4. Epistasis*

Evidence is now accumulating that epistatic interactions are not at all unusual in the determination of the magnitude of quantitative traits. For example, Eshed and Zamir [9] have concluded that epistasis leading to a reduction in additivity occurs in determination of some quantitative traits in tomato. They propose "that the diminishing additivity of QTL effects is amplified when more loci are involved; this mode of epistasis may be an important factor in phenotype analysation and in breeding." Blows and Hoffmann [3] have shown that there can be an association between non-additive genetic variation and extreme expression of a trait, such as comes about when strong selection has been applied as in plant and animal breeding. Routman and Cheverud [30] have confirmed this result in selected lines of mice. At a much simpler wholly phenotypic level, Akerlind and Emanuelson [1] have shown that in selection for high and low fat content in milk, the relative proportion of long and short chain volatile fatty acids changes with the two directions of selection. That is, the higher the fat level, the higher the proportion of short chain VFA and vis-versa. This might be expected on bio-energetic grounds, but it certainly suggests epistasis at the genetic level. Again, Brocks, Husegge and Markus [4] have shown in selection in pigs that the fibre composition in muscles changes as selection is for lean growth or for overall rapid growth.

Thus, if epistatic effects become more important as more genes affect a trait, and if epistasis has important implications for product quality or animal welfare, it is at a practical level important to know the number of genes. Study of QTL frequency differences among selected lines (e.g. Keightley *et al* [14]) strains of a breed or breeds, which will only become possible as mapping becomes more complete, may allow epistasis to used or overcome, as necessary.

### *5. Genetic Load*

Over time, as progress under selection is successful, strains of animal may be developed that raise welfare concerns. Extreme milk production, for example,

may be accompanied by a unacceptable incidence of mastitis. This may arise because extreme phenotypes may include genotypes homozygous for alleles with deleterious recessive side effects. Such populations may be said to have a high "genetic load".

This concept, introduced by Muller [24], is the burden "felt in terms of death, sterility, illness, pain and frustration" (Crow [6]) thought by its proponent to be caused by mutation. This view of genetical variation has recently been revived: see e.g. Lynch and Gabriel [16] and Crow [7]: "[s]ince most mutations, if they have any effect at all, are harmful, the overall impact of the mutation process must be deleterious". This view of populations as end-products rather than work in progress may be contrasted with that of Fisher [10] in his fundamental theorem of natural selection:

> Against the rate of progress in fitness [equal to the additive genetic variance in fitness] must be set off, if the organism is, properly speaking, highly adapted to its place in nature, deterioration due to undirected changes either in the organism, or in its environment. The former, typified by the pathological mutations observed by geneticists, annul their influence by calling into existence an equivalent amount of genetic variance. The latter, which are due to geological and climatological changes on the one hand, and to changes in the organic environment, including the improvement of enemies and competitions, on the other, may be in effect either greater or less than the improvement due to Natural Selection.

If there is indeed an ideal human, then the concept of load has validity; if not, not (Fraser and Mayo [11]). The ability to identify deleterious alleles, their number and their effect in homozygote and heterozygote will, subject to the availability of very large numbers of observations, allow largely fruitless controversy like the reality of load to be disposed of with facts. It may, for example, be possible to screen populations for existing and new recessive deleterious alleles, such as stop codons within structural genes (functional exons). In mice, using site-specific mutagenesis, it will be possible to repeat the bacterial experiments of Elena and Lenski [8] and assess the form of interaction of different deleterious and advantageous mutations. In highly selected strains, we may be able to estimate the extent to which particular new recessive phenotypes contribute to variation in both the selected traits and fitness (Mayo, Burger and Leach [20]).

### *6. Genomic Scanning*

Though the development of the technique of expressed sequenced tags (EST), it is now possible to assay the metabolism in any tissue in a plant or animal to determine precisely which genes are activated. From our own work on the wool follicle, we know that at least 1300 genes are functional in the growing wool follicle (D Adelson personal communication). Which of these genes are important for wool growth is not yet known. By the use of the related technique of differential display (DD), we can determine which genes are active in the cells of the follicle as opposed to those in the surrounding skin, and this will allow us to identify genes specifically involved in wool growth. It is not necessarily true that these genes are going to be the most important for increased quantity or quality of wool, however. Transgenic research has shown how easily the follicle may be disrupted, not how its performance may be enhanced (Powell and Rogers [28]). It is fair to say that, despite all our knowledge of metabolism, we have at the moment no model from which to make predictions.

### *7. Conclusion*

Development of a follicle model has a very high priority for us, but a good model is not a lack only for the wool follicles, it applies also to body growth, especially partition between fat and other tissues (see e.g. Rance *et al* [29] and Veerkamp *et al* [32]).

While I have only referred to our own work on the follicle, evidence that thousands of genes are involved in the metabolism in any given tissue, and hence, in principle, involved in the determination of a quantitative trait, is now overwhelming. That Fisher's model remains robust and effective is a tribute not only to the profundity of Fisher's thought, but also a measure of how difficult it is to move beyond the model based on the undifferentiated effects of many small interacting causes, when we have no idea how to model these interactions.

ACKNOWLEDGEMENT

Wool-related research discussed here has been partly supported by the International Wool Secretariat. We thank Dr. R. R. Woolaston for helpful comments on the paper.

## REFERENCES

[1] Akerlind, M. and Emanuelson, M., 1997. Conjugated linoleic acid in milk fat from cows selected for high or low milk fat percentage. *Proceedings of European Association for Animal Production*, 74.

[2] Ap Dewi, I., Owen, J.B., Axford, R.F.E. and Beigi-Nassiri, M.T., 1997. Selection for high and low prolificacy in Cambridge sheep. *Animal Science*, **65**, 209-215.

[3] Blows, M.W. and Hoffmann, A.A., 1996. Evidence for an association between nonadditive genetic variation and extreme expression of a trait. *The American Naturalist*, **148**, 576-587.

[4] Brocks, L., Hulsegge, B. and Markus, G. 1997. Histochemical in relation to meat quality properties in fast and lean growing pigs. *Proceedings of European Association for Animal Production*, 42.

[5] Cheverud, J.M., Routman, E.J., Duarte, F.A.M., van Swinderen, B., Cothran, K. and Perel, C., 1996. Quantitative trait loci for murine growth. *Genetics*, **142**, 1305-1319.

[6] Crow, J.F., 1970 Genetic loads and the cost of natural selection. In: *Biomathematics*, Vol. 1, Mathematical Topics in Quantitative Genetics (ed.K. Kojima) Berlin, Springer.

[7] Crow, J.F., 1997. The high spontaneous mutation rate: is it a health risk? *Proceedings of National Academy of Science*, **94**, 8380-8386.

[8] Elena, S.F. and Lenski, R.E., 1997. Test of synergistic interactions among deleterious mutations in bacteria. *Nature*, **390**, 395-397.

[9] Eshed, Y. and Zamir, D., 1996. Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics*, **143**, 1807.

[10] Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*. Oxford University Press.

[11] Fraser, G.R. and Mayo, O., 1974. Genetical load in man. *Hummangenetik*, **23**, 83-110.

[12] Franklin, I.R. and Mayo O., 1998. The place of QTL in the basis quantitative genetics. II. Mapping QLT from a wide cross. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale*, **26**, 273-276.

[13] Kearsey, M.J. and Farquhar, A.G.L., 1998. QTL analysis in plants: where are we now? *Heredity*, **80**, 137-139.

[14] Keightley, P.D., Morris, K.H., Oliver, F., Bullfield, G., Falconer, V.M. and Ishikawa, A., 1998. Replicated QTL mapping experiments in mice. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale*, **26**, 205-208.

[15] Koudande, O.D., Thomson, P.C. and van Arendonk, J.A.M., 1997. Experimental design for marker assisted Introgression. *Proceedings of European Association for Animal Production*, 74.

[16] Lynch, M.T. and Gabriel, W., 1990. Mutation load and the survival of small populations. *Evolution*, **44**, 1725-1737.

[17] Mayo, O. and Franklin, I.R., 1998. The place of QTL in the basis of quantitative genetics. I. General considerations. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale*, **26**, 77-84.

[18] Mayo, O. and Hancock, T.W., 1981. Fixation of genes having large or small effects on a trait with an intermediate optimum. *Hum Hered*, **31**, 286-290.

[19] Mayo, O. and Hopkins, A.M., 1985. Problems in estimating the minimum number of genes contributing to quantitative variation. *Biom. J.*, **27**, 181-187.

[20] Mayo, O., Burger, R. and Leach, C.R., 1990. The heritability of fitness: some single gene models. *Theor. Appl. Genet.*, **79**, 278-284.

[21] Mayo, O., Eckert, S.R. and Nugroho, W.H., 1982. Models of gene effects for a quantitative trait in man. *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Human Genetics and Adaptation*, 1.

[22] Mendel, G.J., 1965. Versuche uber Pflanzenhybriden. *Verhandlungen des Naturforschenden Vereins in Brun*, (translation in J H Bennett (editor) 1965 *Experiments in Plant Hybridisation*, Olive & Boyd, Edinburgh).

[23] Morton, J.R., 1879. Fisher's reticulate mating system for immigration into an improved stock. *Theoretical and Applied Genetics*, **55**, 279-284.

[24] Muller, H.G., 1950. Our load of mutations. *Amer. J. Hum. Genet.*, **2**, 111-176.

[25] Panse, V.G., 1940a. Application of genetics to plant breeding. II. The inheritance of quantitative characters and plant breeding. *Journal of Genetics*, **40**, 283-302.

[26] Panse, V.G., 1940b. A statistical study of quantitative inheritance. *Annals of Eugenics*, **10**, 76-105.

[27] Piper, L.R., Bindon, B.M. and Davies, R.H., 1985. The single gene inheritance of the higher litter size of the Booroola Merino. In: *Genetics of Reproduction in Sheep* (Ed R.B. Land and D.W. Robinson) Butterworths, London.

[28] Powell, B.C. and Rogers, G.E., 1997. The role of keratin proteins and their genes in the growth, structure and properties of hair. In: Jolles P, Zhan H, Hocker H (eds.). *Formation and Structure of Human Hair*. Birkhauser Verlag, Basel, Switzerland, 59-148.

[29] Rance, K.A., Hill, W.G. and Keightley, P.D., 1997. Mapping quantitative trait loci for body weight on the X chromosome in mice. I Analysis of a reciprocal $F_2$ population. *Genetical Research*, **70**, 117-124.

[30] Routman, E.J. and Cheverud, J.M., 1997. Gene effects on a quantitative trait: two locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution*, **51**, 1654-1662.

[31] "Student", 1934. A calculation of the minimum number of genes in winter's selection experiment. *Annals of Eugenics,* **6,** 77-82.

[32] Veerkamp, R.F., Haley, C.S., Knott, S.A. and Hastings, I.M., 1993. The genetic basis of response in mouse lines divergently selected for body weight or fat consent. II The contribution of genes with a large effect. *Genetical Research,* **62,** 177-182.

[33] Wilcox, P.L., Richardson, T.E. and Carson, S.D., 1997. Nature of quantitative trait variation in *Pinus radiata:* insights from QTL detection experiments. 304-312. *Proceedings of IUFRO 97 Genetics of Radiata Pine,* Rotorua NZ, 1-4 December *(F R Bulletin No. 203).*