QUANTITATIVE APPROACHES TO ANIMAL IMPROVEMENT

E. P. Cunningham
Dept. of Genetics, Trinity College,
Dublin 2, Ireland

SUMMARY

The amount and kind of genetic variation for quantitative traits revealed by allozyme polymorphism and nucleotide diversity is reviewed. At the nucleotide level, practically every locus is polymorphic, while only 15% of polymorphisms show up at the biochemical level. The increasingly complex picture of the genetic material revealed by molecular studies is discussed. Despite the fact that it does not accommodate these complexities, the standard quantitative genetics model has been a satisfactory basis for successful breeding programmes. Developments in the last two decades which reinforce this success are the statistical developments which lead to optimal breeding value estimation, and the gene flow theory which accommodates complexities in population structure.

With the sharpening division of the world agricultural economy between developed and developing countries, the major challenge for quantitative genetics and animal breeding in the decades ahead is clearly in the latter group. Additional challenges are presented by the need to exploit new molecular and reproductive technologies, and by unresolved questions concerning the effects of heterozygosity, epistasis and non-nuclear inheritance.

THE QUANTITATIVE MODEL

At the first International Conference on Quantitative Genetics, held in Iowa State in 1977, Richard Lewontin offered the following challenge, "Quantitative genetics, upon which modern plant and animal breeding is based, is an attempt to produce knowledge by a systematisation of ignorance". He was being provocative, of course. His point was that the very simple models of gene action which underlie our operational models of genotype and phenotype do not take account of the real complexity of genetic organisation. In particular, they take insufficient account of the detailed structure of the genome, and of the way in which the genetic message is translated into phenotype.

It must be acknowledged that his basic complaint is a valid one. With gathering speed, the genetic research of the past twenty years has uncovered new layers of complexity in gene structure, function and organisation.

Among the complexities that we now know of, one could include the distinction between coding and non-coding DNA segments, and the distinction between nuclear genes inherited in a Mendelian way, and mitochondrial genes transmitted in a maternal and thus a clonal fashion. There is considerable evidence against the independent functioning of genes (Kacser, 1989). There is evidence of parental imprinting, in which genes derived from one parent are activated while those from the other are not. Coding segments have associated promoter regions. Some genetic

elements are transposable within the genome, whereas most are relatively fixed. Some are highly conserved, being relatively stable across species and evolving slowly, while others are much less constrained, and evolve at perhaps ten times that rate. Furthermore, the genetic material is highly differentiated in function: in some regions, active coding segments are packed tight, while in other regions they are interspersed with inactive introns.

Many DNA sequences, including protein-coding genes and non-functional sequences are repetitive. The size, number and distribution of repeats varies among sequences: there are repeated arrays of two to hundreds (or more) copies of a simple dinucleotide sequence (such as CA repeats), or of a complete coding array (such as histone and ribosomal RNA genes), as well as gene families widely dispersed around the chromosomes. Evolution of these sequences may involve slippage in DNA replication, unequal crossing-over, gene conversion and transposition. These processes can lead to a great amount of intrapopulation polymorphism. This is not only a valuable new source of variation for the study of genetic similarities and differences, but it may also give rise to variation in quantitative traits in a way quite different from simple point mutations.

Lewontin is right to the extent that the standard quantitative model takes no account of all this richness of variation in the genome. However, he is incorrect to infer that the standard model is therefore a less than adequate basis for the analysis and genetic management of quantitative traits in animal populations. The gene effects specified in the model are statistical abstractions representing average effects of single genetic units, intralocus gene pair effects, or gene interaction effects involving two or more loci. The model has parallels in physics and in economics, where aggregates of many small effects behave in highly consistent and predictable ways. So in genetics, the simple additive model of gene action has proved an adequate basis for the prediction and interpretation of selection responses, at least in the short term. Likewise, the inclusion of dominance in the model gives good prediction and satisfactory interpretation in most crossbreeding situations. This is not to say that these simple models are universally adequate, but simply that those cases where they fall short are the exception rather than the rule.

THE AMOUNT AND KIND OF GENETIC VARIATION

When Fisher (1918) wrote the first quantitative genetic model representing the Mendelian basis of inheritance, the biochemical nature and structure of the genetic material, and indeed its extent, were unknown. That largely remained the situation for 50 years.

We now know a good deal about DNA. We know, for instance, that the mammalian genome (to take the human case as an example) extends to more than 3 billion base pairs of DNA. Perhaps 1% of this DNA complement is believed to be directly involved in coding for and activating functional genes. How much variation is detectable within this active part of the genome? A great deal of effort over many years has been invested in mapping this variation at the level of the gene product, i.e. the protein. Studies in 242 separate species were summarised by Nevo (1978). He documented the extent of polymorphism for a range of enzymes which could be

identified electrophoretically. This is expressed as the percent of loci examined which show polymorphism. However, a locus with two alleles at frequencies of 0.99 and 0.01 is contributing very little variation. These studies also, therefore, noted the percent of loci which were heterozygous, as an additional measure of genetic variation. These results are summarised in Table 1.

Table 1: Allozyme polymorphism and heterozygosity (Nevo, 1978).

Group	No. of species	% of loci poly-	% of loci heterozygous per individual
Plants	15	26	7
Invertebrates	93	40	11
Vertebrates	135	17	5
Birds	7	15	5
Mammals	46	15	4
Total	242	26	7

For birds and mammals, which are generally the species of interest to us in our present context, about 15% of loci were detectable as polymorphic, and in any one individual about one-third of these were heterozygous. Lewontin (1977) however concluded that the levels of variability were somewhat higher, and that about one-third of structural gene loci in a species were polymorphic, with heterozygosity per individual at about 10%.

These allozyme data are, of course, describing what is measurable by electrophoresis. Since they measure the end product of the gene, these data by definition relate to functional genes. The variation in the less active parts of the genome is likely to be a great deal higher. On the other hand, it is also possible that relatively few controlling genes may determine the expression of much of this variation, with considerable uniformity in the underlying structural genes. My general impression is that, given the well demonstrated malleability of most traits in most populations, these electrophoretic studies show a surprisingly small amount of variability at the gene level.

Restriction site analysis brings a new level of precision to the measurement of genetic variability. The recognition sites for the 50 or so restriction enzymes now in general use represent very small segments of DNA - most generally 4 or 6 bases in length. Restriction enzyme studies therefore produce large volumes of binary type data - presence or absence of a particular site. These sites may bear no relation to gene function, and therefore should give a more representative sampling of the genome as a whole. Not surprisingly, therefore, the amount of restriction site polymorphism and heterozygosity are considerably higher than for the whole-gene polymorphism displayed by allozyme studies.

Nei (1987) summarised estimates of nucleotide diversity from both restriction site studies and sequence comparisons. Selected data from his summary are given in Table 2. The important statistic is the proportion of nucleotides which, on average, are different between any two individuals within the population. It can be seen that this proportion is reasonably consistent across species, varying between .002 and .013. A more recent study (McConnell et al., 1990) compared these results with sequence comparisons of four protein-coding genes in mammals, and produced results within the same range. These are also included in Table 2.

Table 2: Estimates of nucleotide diversity. (Nei, 1987; McConnell et al., 1990)

DNA segment or gene	Species	Nucleotide diversity
	Restriction site analysis	
Mitochondrial	Man	.004
Mitochondrial	Chimpanzee	.013
Mitochondrial	Drosophila m.	.008
β - globin	Man	.002
Growth hormone	Man	.002
Adh gene region	Drosophila m.	.006
	Sequence comparisons	
Adh gene coding region	Drosophila m.	.006
Prochymosin	Cow	.004
Apolipoprotein B	Man	.002
Prolactin	Cow	.013
Growth hormone	Pig	.007

If we assume 1,000 bases per coding region, and also that restriction site differences reflect single base differences, then these figures translate into effectively 100% expected polymorphism at the gene level. Taking the pig growth hormone figure, for example, six separate sequences were compared, and nucleotide diversity was found to be .007. For any 1,000 base sequence, corresponding to a normal gene, we would therefore expect two individuals to differ for 7 bases in the sequence. Even allowing that about 25% of possible nucleotide changes within coding sequences are silent, in the sense of not affecting the resulting amino acid, this points to a very substantial amount of variability below the level of which has been picked up by the allozyme studies.

Another source of genetic variation which is receiving increasing attention is

that which is carried in the mitochondria. While the total mammalian genome is perhaps 3 million base pairs in extent, only something in the order of 1% of it is believed to be functional in the sense of coding for necessary protein sequences. By comparison, the amount of DNA in a mitochondrion is tiny: about 16,000 base pairs. However, almost all of this is in coding regions. Furthermore, a typical mammalian cell might have a thousand such mitochondria, and therefore carry 16 million base pairs of potentially functional mitochondrial DNA. When this is compared with the perhaps 30 million base pairs of functional DNA in the single nuclear set, it puts the mitochondrial and nuclear complements on a much more comparable basis. That being said, we have no evidence that mitochondrial DNA is half as important as nuclear DNA in determining the functional genome.

SUCCESS IN LAB AND FIELD

The real test of the robustness of the theory behind applied breeding programmes is its success in application. This success has been documented by Smith (1984). His tables are worth revisiting.

Table 3: Reported annual rates of genetic gain (as % of the mean) from experiments and breeding programmes (condensed from Smith 1984).

		Percent Gain Per Year	
Species	Trait	Experiments	Breeding Programmes
Poultry	Weight gain	4.1	6.5, 5.6
	Egg production		0.9, 1.7
Pigs	Fat depth	2.1	
	Litter size	0.0	1.5
	Index (6 traits)	1.8	
Sheep	Weaning weight	1.5	
	Litter size	1.2	2.9
	Index (7 traits)		1.2
Beef Cattle	Weaning weight	0.7	
	Yearling weight	1.1, 0.6	0.3
Dairy Cattle	Milk yield	2.2, 2.0	1.0
	Fat yield		0.7

What they show is that in almost every case where selection has been applied consistently, and on a reasonable scale for a long enough period, substantial genetic

change in a desired direction has been achieved. Many selection experiments (reviewed in the Proceedings of the Harrogate meeting of 1979), have demonstrated the short term predictability of genetic response in laboratory populations. It was not, however, until within the last decade, that any number of comparable experimental results in domestic livestock have become available. The reason for this is the quite simple one that to achieve even four generations of selection in sheep or cattle takes the best part of two decades to accomplish. To carry out such an experiment on an adequate scale, and with adequate controls, is therefore an expensive undertaking, requiring a large commitment of resources, and substantial institutional stamina to see it through.

The historical origins of genetic variability are one question. Its maintenance in the course of a selection programme is another, and a more immediate one, because it concerns the sustainability of current breeding programmes.

Selection limits undoubtedly do exist, though they have not yet become evident in domestic animal selection programmes, with the exception of egg-laying in poultry, though even here, some reports indicate continued responses (Flock, 1979). Enfield (1979) discussed long-term selection responses in laboratory populations, and produced the reassuring result that the plateaux encountered were at least 20, and usually 30 generations into the programme, and that even in these cases considerable genetic variance for the traits under selection was still present.

These results apply to experiments in which a single consistent selection objective was sustained over many generations. Practical breeding programmes are relatively protected from the problem of a genetic limit because it apparently takes many generations to develop, and also because over such long periods of time (say ten generations or more) shifts in the breeding objective will normally take place. The breeding goal is normally a multiple one in any case. Furthermore, it is unusual for a commercial population not to import genes over such a long period, and this will tend to restore genetic variance.

Apart from long-term depletion of genetic variance, the other major disturbing factor is expected to be the development of unfavourable correlations with fitness traits. This was investigated by Nicholas & Robertson (1980) and more recently by Hill & Keightley (1988), who concluded that the effect is likely to be greater on limiting response than on reducing variance. The problem has already required corrective action in two notable cases: because of a decline in cow fertility, it has been found necessary to include that trait in progeny testing and selecting dairy bulls in Sweden and Norway; in some European beef breeds, selection for muscularity has led to increased incidence of calving difficulty and calf losses, to the point where much of the selection exercised is now devoted to countering these consequences.

Apart from the prospect of long-term depletion of genetic variance, there is a more immediate effect of selection on genetic variance, first developed by Bulmer (1971, 1989). He predicted a reduction in the genetic variance because of the development of negative associations between loci induced by selection. The phenotypic variance is not reduced in proportion, with the consequence that the heritability declines. Most of the reduction takes place in the first few generations after the first round of selection, and is

typically of the order of 20-30%. The predicted reduction in variance and response has been observed in selection experiments (e.g. Atkins & Thompson, 1986). Much of the planning of breeding programmes carried out over the last few decades, and involving predictions of genetic change, has ignored this effect, with the consequence that the expectations of response have been generally too high.

OPTIMAL ESTIMATION

The development of optimal estimation procedures for breeding values owes much to a particular conjunction of topic, time and temperament. The topic was the problem of genetic evaluation of bulls destined for widespread use in artificial insemination. As AI became widespread in the 1950's, this emerged as a real and immediate challenge. Since individual selected bulls could have an enormous impact on the population, accurate discrimination between them was a high priority. The challenge was all the sharper, since, particularly in the United States at that time, the breeding objective was clearly volume of milk per cow, complicated only by some considerations of type. The time brought not just the challenge consequent on AI, but the steady expansion in the power of computers, which gave purpose and expression to the developments in statistical theory. The temperament belonged to C.R. Henderson, who focused a lifetime of intellectual energy on the single theme of optimal estimation.

I have used the term "optimal estimation" from the title of a book (Liebelt, 1967) which covered the range of retrospective, current and prospective statistical estimation. Henderson took the view that, in the genetic context, it is always concerned with the genetic values of future generations, and insisted on "prediction", hence Best Linear Unbiased Prediction (BLUP). This is now the procedure of choice wherever it is feasible to use it. Its origins have been recounted many times, and recently by Henderson (1988) and Dempfle (1989). It is essentially general linear model theory, usually involving both fixed and random effects. In certain cases, it is equivalent to selection index. BLUP therefore subsumes all of the work of the last 50 years on contemporary comparisons, combined selection, and much of selection index into a broad framework which guarantees optimal selection provided its assumptions are met.

Is there unfinished business in this area? Are there negative aspects to be corrected? One of the arguments that was used to defend Contemporary Comparisons was that the principles could be understood by any user. That is certainly no longer true with BLUP. One danger therefore is that, as large volumes of data are processed to produce BLUP estimates, unrecognised biases may be included. The largest application is in dairy sire evaluation. Before the animal model BLUP was introduced, evaluations were solely on progeny data. Dams and sisters can now contribute substantially. With the intense concentration on the outstanding females which are used as dams, is there a danger that their records may be manipulated, and that the resulting biases will feed through into the published proofs? In the same vein, is there a danger that confidence in the theoretical power of the method will lead to reduced concern about good structure in the data used?

A further factor in the efficiency of selection with multiple sources of information and multiple traits in the selection objective is the way in which the importance of the

traits and sources of information is evaluated. Much confusion exists in this area. Many of the statements of relative importance that are made are scale dependent. Thus, to speak of the relative economic weights, even per standard deviation, can be misleading. Likewise, the weighting factors applied to the sources of information in calculating an index tell nothing about the relative importance of the individual sources. There are, however, two statistics, which serve the purpose admirably. For the sources of variation, it is possible to calculate a statistic which specifies the percent reduction in the accuracy of the index, and hence in the genetic gain achieved, if that variate is omitted. This statistic is entirely analogous to the reduction in sum of squares corresponding to a single degree of freedom in a regression model. On the response side, the total gain from selection is an aggregate of gains (and perhaps losses) for the different traits. This aggregate is stated in economic units. A second statistic expresses the percentage of this total gain which is accounted for by gain in each of the target traits. The use of these two statistics permits very rational exploration of selection programmes.

POPULATION STRUCTURE

Under this broad heading, some very significant developments have taken place in animal breeding theory and practice in the last two decades. The evolution of these ideas is briefly reviewed by James (1989). The key element is the development in the mid-70's of the gene flow matrix theory, which gives a general methodology for tracking the consequences of selection through a population, no matter how complex its structure. To complete this armoury of methods for predicting breeding value and measuring selection responses, we still need a good general set of methods for the continuous monitoring of genetic relationships within the population and consequent developments in inbreeding.

One consequence of these studies on population structure has been a renewed interest in the benefits of selection within a nucleus population. The main impetus for this came from the paper of Nicholas & Smith (1984) on MOET schemes. However, the focus has now shifted to the possibilities of using nucleus selection schemes to good effect in the populations of developing countries which, because they lack the documented infrastructure normal in the developed world, are not in a position to apply conventional population-wide selection programmes. The advantages are obvious: in most relevant situations there exist publicly funded research or development centres with livestock resources that could be used as a population nucleus. If relevant, it is feasible to use the most advanced technology in such nuclei: AI, ET, biochemical or DNA genotyping. Much more work needs to be done to develop these possibilities, at both the theoretical and practical level.

APPROPRIATE BREEDING GOALS

Breeding objectives in modern improvement programmes are never simple. Modern dairy cattle, pig and poultry programmes normally have three or four major traits and up to six minor ones in their formal goals. Throughout the developing world, while less formality in selection objectives exists, there is even greater complexity caused by the necessity to reconcile the needs of increasing productivity per animal with the needs of adaptation to highly demanding environments.

In the first of these cases, formal weightings are normally used to define breeding objectives. These have traditionally been specified as they were in the first selection index example by Hazel in 1942, i.e. as economic weightings reflecting expected producer returns from unit improvements in each of the traits. This notion has been refined somewhat over the years. Quadratic relationships of value to level of merit have been developed. Discounting has been introduced to allow for the dispersion in time of multiple expressions of the trait. Allowance has been made for the increased costs involved in expression of higher merit, to give net rather than gross returns.

However, these and other adjustments have all been made on the general assumption that the market is an open one, and that increments of productivity do not affect the net returns. In the short term, this is undoubtedly true. However, the arrival of milk quotas in Europe and elsewhere has brought many of these programmes sharply up against the reality that the benefits of genetic improvement, as indeed of any other technology, primarily belong to consumers, rather than to producers. In the long run, genetic improvement, leading to higher productivity, results in lower unit costs, and competition ensures that these cost reductions are passed on to the end user of the product.

Between this broad economic reality and the present position of many breeding programmes, of course, there may be many intermediate stages. In an important paper in 1985 (Brascamp *et al.*) this question is addressed. They introduced the notion of "normal profit" from economic theory to reformulate the method of deriving economic weights. Paradoxically, the appropriate value for normal profit is zero. This says that in the long term, and in a perfectly competitive situation with fixed total market, the average profit is simply that required to keep production going. It can therefore be considered as part of the cost of production, and in this sense is equivalent to the approach of Dickerson (1978) where cost minimisation rather than profit maximisation is the objective.

The conclusion of Brascamp *et al.* was that, in meat animals, evaluation on the basis of the individual animal, the reproducing female or the unit of product all gave the same selection objectives.

Throughout the developed world, where fast moving, sophisticated breeding programmes are generally in place, and where total markets for the products of the livestock industry are either full or expanding very slowly, this is the appropriate way to define breeding objectives.

The situation throughout the developing world is quite different. There, real demand is expanding steadily, and will continue to do so for several generations into the future. This can be illustrated with a few stark statistics. Population in the developing world will double by the year 2050. Consumption levels for milk and meat throughout the developing world are currently about one-fifth of what they are in the developed world. The driving forces of an expanding consumer base and rising consumption levels should therefore ensure that the "closed market" objectives should not apply there. What this means in practice is that the returns to investment in livestock improvement in the developing world should be greater, both in strict

financial terms, and more significantly in terms of service to real human needs, than they are in the developed world.

THE CHALLENGES AHEAD

The challenges facing quantitative genetics and animal breeding in the decades ahead are determined by the requirement to serve real human needs, the requirement to both develop and accommodate advances in relevant technology, and by the intellectual drive of the individual scientist to innovate, to discover, and to improve.

Serving real needs means a reorientation of effort towards the developing world. In that context, livestock improvement is a genuine public service, and will continue in that mode. Elsewhere, the application of science in this field is steadily moving into place as another technology more and more closely driven by competitive market forces. This is already wholly the case in poultry, and largely in pigs. It is becoming so in cattle and sheep. There are clear signs that, beyond certain limits, consumers in these markets are now interested more in quality than in decreasing real prices.

A second challenge is to balance conservation of genetic resources with development of productive potential. Over 95% of world milk and meat production derives from five species. The variation they encompass is a valuable resource. Its conservation and study is most urgent precisely where it is most difficult, in countries which have an even more urgent need to improve productivity. Resolving this dilemma is not an easy task.

Three areas which definitely need more research, at both theoretical and experimental levels, are the questions of inbreeding (and crossbreeding), epistasis, and non-nuclear inheritance. It is likely to be worthwhile to expend much greater effort to maximise heterozygosity, even where wide outcrossing is not practised. Mapping the benefits and developing the plans for achieving this, and for balancing it with the needs of selection programmes are unresolved questions. Epistasis is difficult to measure, not least because it is by definition an infinitely expandable category of gene action. However, there is evidence that something more than additive and dominance effects are involved in some important instances, including milk production in *Bos taurus x Bos indicus* crosses. I have no doubt that similar situations exist, as yet undocumented, in other crosses between genetically distant strains. While measuring epistasis is difficult, and exploiting it may be impossible, knowing its extent is important because it impinges strongly on the structure of crossbreeding programmes. Cytoplasmic inheritance is receiving increasing attention, partly because we now have better ways of studying it. The experimental possibilities offered by embryonic cloning, and by placing identical nuclear genotypes in varying cytoplasmic environments has intriguing possibilities.

A major challenge for quantitative genetics is how it should interact with new gene technology. Will marker assisted selection become part of routine programmes? Will gene insertion increase in precision and control to the point where it can become a part of improvement schemes? Will embryonic cloning, including sexing, by-pass much of the elaborate selection schemes in place today?

The body of theory which underpins today's successful breeding programmes has served us well. The two broad challenges of the coming decades are a reorientation towards the problems of livestock improvement in the developing world, and a fruitful amalgamation of quantitative and molecular genetics. These will require new thinking , new training, and a renewed commitment to one of the most worthwhile tasks that science can undertake.

REFERENCES

ATKINS, K. D. & THOMPSON, R. 1986. Anim. Prod. 43: 421-435.
BRASCAMP, F. W., SMITH, C. & GUY, D. R. 1984. Anim. Prod. 40: 175-180.
BULMER, M.G. 1971. The American Naturalist. 105: 201-211.
BULMER, M.G. 1989. Genome 31: 761-767.
COCKERHAM, C. C. 1954. Gen. 39: 859-882.
DEMPFLE, L. 1988. *Evolution & Animal Breeding,* 181-188. CAB.
DICKERSON, G. E. 1978. Anim. Prod. 27: 367-370.
ENFIELD, F .D. 1979. *Selection Experiments in Laboratory and Domestic animals,* 69-86. CAB.
FISHER, R. A. 1918. Trans. Roy. Soc. Edin. 52: 399-433.
FLOCK, D. 1979. *Selection Experiments in Laboratory and Domestic Animals,* 214-224. CAB.
HENDERSON, C. R. 1988. Proc. 2nd Int. Conf. Quant. Gen. 85-90. Sinauer.
JAMES, J. W. 1989. *Evolution and Animal Breeding.* 189-194. CAB.
KACSER, H. 1989. *Evolution and Animal Breeding.* 219-226. CAB.
LEWONTIN, R. C. 1977. Proc. Int. Conf. Quant. Gen. 55-62. Iowa State Univ. Press.
LIEBELT, B. 1967. *Introduction to optimal estimation.* Addison-Wesley, London.
McCONNELL, D. J., CHEN YUNG FU, QIN LU, D. HUGHES, P. M. SHARP, & D. P. CARROLL. 1990. In preparation.
NEI, M. 1987. *Molecular evolutionary genetics.* Colombia Univ. Press.
NEVO, E. 1978. Theor. Pop. Biol. 13: 121-177.
NICHOLAS, F. W. & ROBERTSON, A. 1980. Theor. Appl. Gen. 56: 57-64.
NICHOLAS, F. W. & SMITH, C. 1983. Anim Prod. 36: 341-353.
SMITH, C. 1984. Res. & Dev. in Agric. 1: 79-85.