MOLECULAR BIOLOGY AND ANIMAL IMPROVEMENT

ALAN ROBERTSON*, SCOTLAND

In applying quantitative genetics theory to problems of animal improvement, the aim is to maximise the rate of improvement per unit of time. The theory is developed in terms of the variance of the characters under selection, the basic concept being the heritability, the proportion of the total phenotypic variance due to additive gene effects. From this central concept and the known rules of Mendelian segregation, we can derive the expected correlation between, for instance, the performance of an animal and its breeding value or between the performance of related animals. For most of the time, we can stay at this level and deal with individual genes vary rarely. Much of our efforts are then devoted to the optimum use of the available information.

Over the last thirty years, much information has accumulated about the segregation of identifiable genes in domestic animals, because of the development of more and more sophisticated biochemical techniques. I need hardly explain how the DNA sequence of a gene is first transcribed into a messenger sequence of RNA which is then translated, three bases at a time, into the amino acid sequence of a protein, which may be an enzyme like cytochrome c, a structural protein like myosin, concerned with transport like transferrin or haemoglobin or with the regulation of growth and development through a hormonal function like insulin. Note that there are $4 \times 4 \times 4 = 64$ possible triplet combinations of four bases but only 20 amino-acids are coded for. There is thus "redundancy" in the sense that many amino-acids are coded for by several triplets. Such sets of "synonymous" triplets are usually identical in the first two places. But I would stress that only a small proportion of the DNA in the genome codes for amino acid sequences. Some is concerned with the general purpose machinery of the cell, like the ribosomal RNA, some with the detailed regulation of specific functions in the cell and some may have no function at all.

New variation in the DNA may arise in several ways, of which the simplest is the substitution of one base for another. It may also involve loss or gain of a consecutive series of bases (deletions or insertions) or the multiplication of such a series. If the multiplication takes place at the level of whole genes, we speak of an alteration of "copy number" and of a gene family which may be "tandem" or "dispersed". Current evidence in humans would suggest that the majority of mutational changes are at the single base level but this may not be true for all species. Such new variants may remain in the population for many generations and give rise to a "polymorphism".

A new mutant remains of course in association with the preexisting variations on the chromosome on which it occurred and this may last for many generations. In humans, 1% crossing over per generation will occur between sites a million bases apart so that the half-life of such an association would be 70 generations.

The first such genes to be discovered were the red cell antigens, the "blood group genes". The development of gel electrophoresis, detecting differences between protein variants in migration rate on, for instance, a starch gel and due primarily to differences in charge, has shown up much

*Genetics, Edinburgh University EH9 3JN, Scotland

variation, depending on the ingenuity of the biochemist in inventing techniques for the visualisation of the protein concerned. This made it possible to measure the amount of the genetic variation in populations in an objective way - to answer the question, "what proportion of loci coding for proteins are showing genetic variation?", by modifying it to "what proportion of protein chains show variation on a gel?" Almost simultaneously, Lewontin and Hubby (1966) working with Drosophila and Harris (1966) with humans showed that the proportion was very high indeed. About one-third of all loci coding for proteins detectable by the method showed variation within populations. This figure is even more remarkable when one considers that the technique must be conservative in that only about 30% of amino acid substitutions produced by a single base change will show a difference in electric charge. Modifications of the methods used have in fact uncovered further genetic variation in some cases.

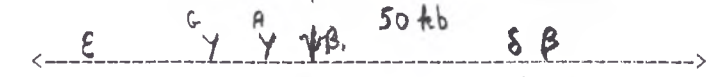
Analysing in finer detail we may examine the base sequence of structural genes . All coding sequences have features in common. Not only do they have specific triplets at the beginning and end of the coding part of the message, but also general features extending several hundred bases before its start, necessary for the processing of the message during translation and a further necessary terminal sequence, rather more variable between different proteins. Most surprising, however was the discovery that the message contains inserted sequences of whose function we have little idea, the introns, and which results in some cases in the functional message being read in up to forty pieces so that the DNA sequence is many times longer than that required to code for the corresponding amino acid sequence. We now possess very rapid methods of determining DNA sequences, but, since the relevant sequences are sometimes several thousand bases long, we have very little population information at this level. Indeed in that much studied organism, Drosophila melanogaster, our only published information on variation within populations is an analysis of eleven copies of the gene for alcohol dehydrogenase (Kreitman, 1983).

There is a short-cut method which allows us to recognise some of the variation in DNA sequences and, with some extrapolation, to estimate the proportion of bases at which there is segregation. Many species of bacteria carry enzymes capable of splitting DNA at rather short signal sequences (of 4 to 6 bases) - presumably serving as a mechanism against invasion by foreign DNA. The exposure of total DNA of, say, Drosophila to such a "restriction enzyme" allows us to measure the lengths of fragments of the DNA containing a previously marked sequence or probe. A base change in the signal sequence will then alter the pattern of fragments produced and indicate whether or not the individual carries the signal sequence. As many restriction enzymes are available, each recognising a different sequence, it is possible to recognise many of the polymorphisms in the base sequence as a polymorphism in fragment length.

Finally, I need to mention the many repetitive DNA sequences in the genome. There are families of "transposons", capable of moving from one site to another and in prospect very important in genetic engineering, and the clusters of tandemly arranged functional sequences such as those coding for ribosomal RNA, histones, globins and keratins. We may take, as an example of a cluster of functional genes, the β-globin cluster in man.

The major hemoglobin fraction in our blood is a molecule which involves

four amino-acid chains, two α and two β. These two are so similar in amino-acid sequence that they must have descended from the same ancestral molecule, some 500 million years ago. However the two genes are now on different chromosomes, each in a cluster of related genes, deriving from a process of duplication and divergence. The β-cluster in man is illustrated in the figure.



polymorphic restriction sites

*	*	*	*	*	*	*	
+	−	−	−	−	+	+	47
−	+	−	+	+	+	+	17
+	−	−	−	−	+	−	12
−	+	−	+	+	+	−	8
Others							16
							Frequency

Note that

(i) the cluster includes sequences coding for six different globin molecules.

(ii) though recognisable as coding for a globin, the fourth gene would not now function as such. It is a "pseudogene", which at some time has been switched off and which is now gradually decaying under the action of continued mutation. Such sequences are valuable in presenting a standard for the rate of change of sequences under mutation but with no natural selection. It has no introns.

(iii) all other globin genes in the cluster have two introns inserted at the same places.

(iv) the two fetal γ globins differ from one another by a single amino-acid.

So much for the amino-acid sequences. But we have further evidence at the DNA level from the analysis of restriction fragment polymorphisms and from some sequencing. Variation can be seen at seven restriction sites in the region and the figure shows the most frequent combination of sites found in chromosomes in Mediterranean populations. It will be seen that though there are 7 variable sites, giving 128 different possible combinations, the most frequent four combinations account for 84% of all the chromosomes analysed. In other words, there is an association between the variants present on the same chromosome. In the cluster, there are several mutants affecting haemoglobin function. These have been analysed by DNA sequencing of the β gene region. It has been found that

(i) each mutant is associated closely with a particular combination of restriction variants.

(ii) the mutants are almost always single base changes.

(iii) the changes do not all occur in the coding part of the DNA sequences but also at other sites affecting function, such as that part of the intervening sequences which controls their own removal before the gene can function. They are "regulatory" mutants.

(iv) the same combinations of restriction variants may be found in different human races.

(v) mutants affecting function are usually restricted to a single race.

The process leading to this pattern of variation can be envisaged as follows

(i) the mutational events producing the restriction variants are old - perhaps of the order of a million years.

(ii) the mutants producing variants affecting function are single events which have occurred since the pattern of restriction variants was established.

How much genetic variation is there in animal populations?

How much genetic variation is there at different levels? We saw above that perhaps 40% of peptide chains appear polymorphic on electrophoresis. What proportion of amino acids in such sequences show variation within the population? We may take a rough value of 200 as the average number of amino acids in proteins. In the great majority of cases (though by no means all) the difference between alternative AA sequences within populations is due to a single amino acid change, giving a value of 2×10^{-3} for the proportion of amino acids showing segregation. This is usually due to a single base change, occurring at the first or second positions in the codon, giving a value of 1×10^{-3} for the proportion of first and second positions of coding sequences which vary. Finally, it is possible, from the amount of polymorphism discovered by restriction enzymes, to give an estimate of the proportion of sites in the genome as a whole which are polymorphic and this, very crudely, comes to a value of 1×10^{-2} - indicating a vast amount of genetic variation within human populations at the DNA level (Jeffreys, 1979). I must emphasise the crudity of these figures - I am concerned to indicate their probable order of magnitude. But note that there are 3×10^{9} bases in the human genome and therefore 3×10^{7} polymorphisms at base level.

What use can be made of this information in animal improvement?

There are several different ways in which we can use this knowledge.

(i) the information on gene structure and regulation is of value for its own sake. In fact, the experiments on mice which I will discuss later have in the main been undertaken to gain new basic information on gene function.

(ii) increasing the accuracy of the selection process.

There are now many loci in the genome at which the genotype can be directly inferred from the phenotype, e.g. those controlling blood groups, enzymes, restriction sites, and so on. Some of these "marker" loci may prove to have a direct effect on "production" itself. Searches for such effects over the last 30 years have almost invariably failed. This is in fact not surprising. There may be in the genome several thousand marker loci segregating. There will by definition only be a few important "production" loci controlling the variation in any metric character. If we have, say ten "marker" loci segregating, the chance that any of the latter will also affect production must be very small.

If, however, we can screen a great many marker loci at the same time, we may markedly improve the chances of detecting those which also affect production. Bulfield (1984) has emphasised the potential value of 2-D (two dimensional) electrophoresis in identifying genes with direct effects on the character. It is possible to identify on such gels upwards of a thousand separate proteins in an individual. Comparison of lines which have been selected in opposite directions for several generations should show up spots present in one and absent in the other or perhaps differing greatly in intensity.

(iii) Can we use linked marker loci which have no direct effect themselves on production but which may be linked with such a locus? In the long term one would not expect association in a random breeding population between a production locus and a linked marker since it would decay gradually over time because of continued genetic recombination. There are two situations in which associations might be found. First, in early generations after a cross, associations in the parental strains might be expected to persist for some time in the crossbred population. Secondly, if there is little recombination between the two loci, it is possible that a particular mutant at the functional locus, which was at a selective advantage within the population, might spread through it and carry along genetic variants at the marker locus. A possible example might be the spread of sickle cell disease in humans in West Africa. Analyses of black populations in the United States show that a large proportion of chromosomes carrying the gene for sickle cell globin lack a particular restriction enzyme site (Kan and Dozy, 1979). This has provided a method of detecting S/S homozygote fetuses at a very early stage from their DNA restriction pattern, allowing an early abortion and reducing the risks to the mother. The association between the restriction site and the mutant form of the protein would imply that the majority of present S genes are identical copies of a single mutation on a chromosome lacking the restriction site concerned.

Further work (Antonarakis et al., 1984), considering these restriction sites in West African populations suggest that more than one event is involved. Of 170 HbS genes from American and Jamaican blacks, 113 had the most frequent pattern of restriction sites near the β globin gene and 52 had the second. Only the first group lacked the site used by Kan and Dozy. Almost all the 5 genes came from 2 mutational events.

Even though there may be linkage equilibrium between the marker locus and the functional locus in the population as a whole, it can be shown that there will be an association between the two within some families although its sign will vary from family to family. This happens in a family in which one of the parents is a double heterozygote at the marker locus and a locus controlling a recessive disease. Depending on the gene frequencies in the population, it is possible in a useful proportion of families in which the genotypes of both parents and an affected child are known, to predict with certainty whether a subsequent foetus is homozygous for the deficiency, from its genotype at the marker locus. In man, this method is now being used to reduce the number of individuals born who are homozygous for the recessive gene producing phenylketonuria, which has an incidence of 1 in 10,000 in Western populations (Woo et al., 1983). Three restriction site polymorphisms have been discovered close to the locus coding for the relevant enzyme.

In selecting for a quantitative trait, successful use of such temporary associations would depend on a detailed preliminary analysis of the sign and magnitude of the association in the progeny of individuals heterozygous at a suitable marker locus. The predictions would be much improved if enough markers were available that important genes affecting the character could be bracketed by a marker gene on either side of it. There is considerable interest at the moment in the use of such techniques to obtain complete coverage of the human genome. Bishop et al. (1983) have calculated that 400 marker loci are required to give a probability of 90% that a locus chosen at random in the human genome would be within 20 cM of a marker, if the latter are scattered at random through the genome. Several groups are now involved in such efforts in man, and as much of the material used is likely to be useable across mammalian species, the best strategy for domestic animals might be to follow closely in the wake of human investigations. Bishop et al. suggest that sufficient work is now going on on the human genome that "the genome would be effectively covered by 1985". Recently Soller and Beckmann (1983) have studied the possibility of applying these techniques to improve the efficiency of selection in animals, primarily in the context of dairy cattle improvement.

Jeffreys et al. (1985) have used "mini satellites", loci at which chromosomes differ in the number of repeated copies of a small DNA sequence that they carry in tandem, to give parentage tests of a much higher accuracy than was previously available. The techniques have been used successfully in man but not so far in animals.

It is possible that some of the changes brought about by selection have been produced, not by selecting between alleles at the same locus, but by selecting those chromosomes with the largest number of copies of useful loci. Frankham has shown that response to selection for reduced abdominal bristles in Drosophila is probably due to a reduction in the number of copies of genes coding for the ribosomal RNA (Frankham, 1980). Other aspects peculiar to Drosophila selection programmes - the frequent occurrence of balanced states with a high level of phenotypic variability in which recessive lethals are being selected as heterozygotes, which have apparently occurred during selection and also the evidence that in several unrelated lines the same mutation has appeared - could be explained by an increase in the number of copies of a particular sequence. We shall see below, in discussing genetic transformation, that the easiest change to make is to insert several copies of a locus. Repeated elements have proved very valuable as carriers of foreign DNA sequences in such transformations.

Genetic Transformation

The last two years has seen some startling papers published concerning the "transformation" of mice and of Drosophila by the insertion into their genomes of small DNA fragments from foreign species. The problems of the genetic engineer may be described as, first to catch the useful genes, perhaps from a different species, and then insert them into the DNA so that they are transmitted stably from generation to generation in the usual way and are expressed in the usual tissues. The "recombinant DNA" is the product of a series of cuttings and rejoinings. I would emphasise that the first step may

be the most difficult one. In the majority of selected strains we have no clue whatsoever as to the basic genetic changes that we have brought about by selection, we don't know what kind of genes we are concerned with and we are not sure whether the modification is in the kind of alternative allele fixed or in the number of copies. There are at the moment several variant alleles segregating in animal populations which are of potential value, such as the "halothane" gene in the pig which produces lean carcasses and the Booroola gene in Merino sheep which markedly increases fecundity. At an experimental level, genes like obese in the mouse and fatty in the rat are valuable as models for obesity in the human. For all these, we remain ignorant of the basic genetic cause which produces the observed effects and, until we can "catch" the gene by isolating its primary protein product, genetic manipulation is difficult.

When the economic product is itself a protein, such as casein in milk, the experimental approach is clear. In wool, it is known that a complex family of genes code for the important protein, keratin. The use of genetic engineering techniques will be valuable in analysing the genetical and biochemical organisation of the family, apart from direct practical implications (Ward et al., 1982). Finally, the peptide hormones, already used so brilliantly by Parmiter et al. (1982) are an obvious target. It might also be possible to "engineer" strains of cattle or sheep capable of producing medically important peptides in their milk, such as interferrin or blood factors VIII or IX, in place of the usual casein or -lactoglobin.

The gene must then be separated, purified and inserted into a vector in a bacterial cell in which it can be multiplied. It must be inserted into a recipient animal, perhaps at the stage of a just fertilised egg, in the hope that it will be taken into the genome of the recipient, transferred from parent to offspring in the normal way and expressed in the right tissue. There are several ways of carrying out such insertions and at least three different techniques have been used successfully. The most used method involves injection of a number of copies of the DNA sequence in a suitable carrier into the male pronucleus of fertilised eggs. In some experiments with mice, insertion of the genetic material has been satisfactorily achieved but expression of the gene has only been achieved at a low level and in the wrong tissue. The recipient genomes then prove to carry several copies (up to several hundred in some cases) of the donor sequences attached head to tail with one another in a tandem fashion. The most striking results have come from transformation of mice with growth hormone genes of rat and man (Parmiter et al., 1982, 1983). They and their colleagues have recently summarized their work with mice and domestic animals (Hammer et al., 1985; Brinster et al. 1985; Palmiter and Brinster, 1985). They mostly used a "construct" of a metallothionein promoter attached to a growth hormone gene from a different species and injected 700 or so copies of this into the male pronucleus of the mouse. The "integration frequency", the proportion of fetuses coming from injected eggs which retained injected DNA was of the order of 25-30% under optimal conditions. In experiments in which 1200 eggs were injected, 111 young mice were born and 23 of these expressed the growth hormone gene of the donor. The eggs used came from crosses between two standard laboratory inbred lines of mice and the success was 8 times greater than when eggs from the inbreds themselves were used. "When more than one copy integrates which is usual with microinjected DNA, the multiple copies are typically arranged in tandem head- to-tail containing up to several hundred copies". The level of expression usually does not correlate with gene copy number. When viruses

were used as a vehicle for the DNA, copies integrated singly at many sites.

A useful technique in Drosophila has been first to insert the donor sequence into a P element (a widespread transposable element present in several species) and to inject these into embryos. Of all the transformation experiments, these have been the most successful in achieving a high level of gene function in the right tissue. The donor sequences have been found to be inserted singly at several places in the genome. The size of the inserts have been around 10 kilobases suggesting that, for the genes concerned, all the necessary information for function is contained in a unit of that length (Rubin and Spradling, 1982). It will be interesting to see if such transformation techniques using transposons can be worked out in domestic animals. The RNA retroviruses are structurally very similar to the DNA in these Drosophila transposons and have been used as carriers to transfer genes into mouse bone marrow cells.

Hammer et al (1985) recently reported on the production of transgenic rabbits, sheep and pigs by direct microinjection. The techniques used so successfully with mice work well with rabbits. Visualisation of the pig and sheep micronuclei was difficult but was eventually possible for the former after a short period of centrifugation. They injected about 5000 ova in total of which about 500 led to neonates or fetuses. The frequency of integration was about 12% in pigs and rabbits and only 1% in sheep.

In the experiments on mice, the use of the metallothionein promoter with a growth hormone gene led to some of the mice being twice normal size, due to the usual feedback control mechanisms being bypassed. On the other hand, the growth rate of the transgenic pigs was apparently not increased by the level of growth hormone.

Because integration of DNA is at random in the recipient genome, the inserts might be expected in some cases to land in necessary part of the genome of the recipient and interfere with normal function. Present data would suggest that 10-20% of transgenic mice harbor recessive mutations of essential genes (insertional mutagenesis). Mackay (1984) has shown in Drosophila that such genetic variation arising from the movement of P elements can increase the selection response for abdominal bristles several fold .

I have dealt with new genetic variation of several kinds. Transformation by injection is of value in increasing the activity of a particular enzyme and may at the same time reduce the activity of others in a purely random way. One might however imagine situations in which a controlled reduction in the activity of an enzyme or perhaps of a peptide hormone might be agriculturally desirable. One such study involved making a sheep immunologically reactive to its own somatostatin. The consequent reduction of somatostatin levels produced an increase of growth hormone and of growth rate. One might imagine that a decrease of the enzyme producing lactose or a protease breaking down muscle proteins would be economically desirable.

The use of "anti-sense" RNA seems to give a method of accomplishing this. An anti-sense plasmid is made in vitro by inverting the protein coding sequence with respect to its promoter. Such a plasmid will therefore produce not the usual mRNA message but a complementary sequence which will hybridize with the usual message and inhibit production of the usual proteins. Such a procedure has recently been shown to work with Drosophila (Knipple et al.,

1985).

Where do we go from here?

I found myself thinking about transformation very much as I would mutation - transformation is controlled mutagenesis. And it carries with it the disadvantages of mutagenesis. We may expect that each mutational event will produce its own syndrome of effects and that fairly rapidly we shall find ourselves with more genetic variation on our hands than we can easily deal with.

Transformation itself has proved surprisingly simple to carry out. It seems that we have two distinct methods available which have different genetic consequences. That used by Parmiter and Brinster appears to result in a single event in which many of the constructs are inserted head-to-tail in just one site. The alternative, not as yet completely worked out, results in many insertional events, each involving one copy of the donor sequence. We shall need to answer the following questions.

(i) how stable will be the inserts? If there is a single large insert, this may well break down under unequal crossing over. If the inserts go in singly, they will be subject to ordinary genetic recombination. Are they liable to excision?

(ii) what will be the effect on the main production characters?

(iii) what will be the "correlated responses" in characters like fertility? One would expect that the larger the number of insertions the greater the effect on fitness, due to insertional mutagenesis.

References

Antonarakis, S.E., Boehm, C.C., Serjeant, G.B., Theisen, C.E., Dover, G.A. and Kazazian, H.H. (1984). P.N.A.S., 81, 853-856.

Bishop, T.D., Cannings, C., Skolnick, M. and Williamson, J.A. (1983). in Statistical Analysis of DNA Sequence Data (ed. B.S. Weir), Dekker, New York, pp. 181-200.

Brinster, R.L., Chen, H.Y., Trumbauer, M.E., Yagle, M.K. and Palmiter, R.D. (1985). Proc.Nat.Acad.Sci. 82, 4438-42.

Bulfield, G. (1984). Proc. 18th Poultry Science Symposium.

Frankham, R.L. (1980). In "Selection Experiments in Laboratory and Domestic Animals" (ed. A. Robertson) pp.56-68. Commonwealth Agric. Bur. Slough.

Gavora, J.S. and Spencer, J.L. (1979). Compl.Immun.Microbiol.Infect.Dis. 2, 359-371.

Hammer, R.E., Pursel, V.G., Rexroad, C.E., Wall, R.J., Bolt, D.J., Ebert,K.M., Palmiter, R.D. and Brinster, R.L. (1985). Nature 315, 680-683.

Harris, H. (1966). Proc.Roy.Soc. B, 164, 298- 310.

Jeffreys, A.J. (1979). Cell 18, 1- 10.

Jeffreys, A.J., Brookfield, J.F.Y. and Semeonoff, H.R. (1985). Nature 317, 818-9

Kan, Y.W. and Dozy, A.M. (1979). P.N.A.S. 76, 2886- 2889.

Knipple, D.C. Seifert, E., Rosenberg, V.G., Preiss, A. and Jäckle, H. (1985). Nature 317, 40-43.

Kreitman, M. (1983). Nature 304, 412-417.

Lewontin, R.C. and Hubby, J.L. (1966). Genetics 54, 595-609.

Mackay, T.F.C. (1984). Genet.Res.(Camb.) 44, 231-237.

Orkin, S.H., Kazazian, H.H., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G. and Giardina, P.J.V. (1982). Nature 296, 627- 631.

Palmiter, R.D., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G., Birnberg, N.C. and Evans, R.M. (1982). Nature 300, 611- 615.

Palmiter, R.D., Norstedt, G., Gelinas, R.E., Hammer, R.E. and Brinster, R.L. (1983). Science, 222, 809-814.

Palmiter, R.D. and Brinster, R.L. (1985). Cell 41, 343-5.

Rubin, G.M. and Spradling, A.C. (1982). Science 218, 348-353.

Soller, M. and Beckmann, J. (1983). Theor.App.Genet. 67, 25-34.

Ward, K., Sleigh, M.J., Powell, B.C. and Rogers, G.E. (1982). In Proc. 2nd World Congr. Genetics Applied to Livestock Production VI, 146-153.

Woo, S.L.C., Lidsky, A.S., Guttler, F., Chandra, T. and Robson, K.J.H. (1983). Nature 306, 151-155.