MAPPING QUANTITATIVE TRAITS BY MEANS OF GENETIC MARKERS

H. Geldermann
Institute of Animal Breeding and Genetics,
Hannover Veterinary School,
Buenteweg 17p, 3000 Hannover 71, FRG

INTRODUCTION

In recent years several traits were improved in different animal species. The development was forced by traditional breeding methods based on phenotypic assessment i.e. production, reproduction and viability. The estimates of such complex phenotypes are the result of a combined action of numerous genes and environmental factors. Since animal breeders were not able to visualize the real genotypes of individuals they selected on the basis of phenotypic values using complex biometrical procedures and applied them to populations. However, the network of effects masks the genotype, thus the estimated parameters are inadequate to analyze genes responsible for the expression of quantitative traits. Even where family data showed effects of major genes, e.g. the halothane susceptibility in pigs, the double muscling in cattle or the dwarfism in chicken, the primary gene products and their mechanisms of regulation remain largely unknown.

However technical advances during the last two decades of animal biotechnology and with it analysis of animal genomes had a large impact. Particular impulses on gene mapping are expected from methods related to molecular genetics.

STATE OF GENE MAPPING

During the last decade, new methods of cell genetics and recombinant DNA technology did improve the efficiency of gene mapping. With the help of somatic cell hybrids and banding of chromosomes, single chromosomes were identified and genes could be assigned to chromosomes or chromosomal regions. The number of assigned genes grew rapidly after DNA sequences were used as probes for *in-situ*-hybridization of chromosomes (Fries et al., 1989). Moreover, a mapping with RFLPs started, based on Restriction Site Variants (RSVs) or Variable Number of Tandem Repeats (VNTRs) (Skolnik and White, 1982; Jeffreys and Flavell, 1977; Bufton et al., 1986; Jeffreys et al., 1985, 1986;

Nakamura et al., 1987). Combining molecular biological techniques with analysis of recombination, DNA sequences were used as probes to detect polymorphic loci and to pursue the segregation for homologous regions of the chromosomes from parents to the offspring generation (Botstein et al., 1980). Linkage relationships were tested in pedigrees by established methods (Morton, 1956; Cerget-Darpoux and Baur, 1990), and the DNA marker loci could be arranged into linkage groups. Since many different DNA sequences can be tested, genetic linkage maps can be constructed which contain a very large number of markers at close intervals. For the purpose of mapping, DNA-variants offer significant advantages: they are screened at the level of DNA and thus normally behave codominantly, have a high number of allelic variation and are more or less free of epistatic effects.

The progress of gene mapping in the human grew exponentially and contains more than 3500 mapped genes, most of them identified by DNA probes (Human Gene Mapping, 1987, 1988). For farm animal species only a small number of genes is mapped when compared to species for which close gene maps are available (Lalley et al., 1988; Fries et al., 1989). However, extensive data, presented in human and mouse, are helpful for further activities, because most methods for marker gene analysis are transferable to several species, and the marker genes show homologies for the genomic arrangement in different mammalian species (O`Brien, 1986; Lalley et al., 1988).

One prerequisite for further detailed linkage studies is a large number of genes which are already mapped and cover most of the genome. The number of families required for such gene mapping is a function of the family size, the information content of the marker genes, and the recombination distance between the loci in question. Generally, large families yield more information than smaller. For gene mapping, DNA markers can be selected according to their value of information. Botstein et al. (1980) defined the Polymorphism Information Content (PIC) which depend on number and frequency of the alleles considered. Distances between the mapped genes desireable for an appropriate probability estimate of other genes by linkage analysis can be calculated with the LOD-scores (Morton, 1956). As seen from Thompson et al. (1978), distances of up to about 0.2 Morgan are feasible with reasonable numbers of families. Thus, newly detected genes, even those affecting quantitative traits, can be mapped easier if a detailed gene map for the species is already at hand or if a large number of DNA loci is included at once.

REFERENCE FAMILIES

Because as many genetic markers as possible should be pursued within a distinct material of individuals or cells, families are important in which alleles of various loci are segregating. For this reason so-called reference families were established by human genetists (White and Lalouel, 1988). A reference material is a collection of samples from members of informative families. Informative individuals have, concerning a target trait or gene, a maximal heterogeneity of trait values and/or related genotypes. The samples (DNA, tissues, fluids, bone marrow, cell lines etc.) are stored and available for several research projects. This type of informative material has been described to study marker gene effects on the values of quantitative traits in cross-bred generations of mice (Kluge and Geldermann, 1982). Analogous cross-bred generations were proposed for farm animal species as well. Hetzel (1988) emphasized the importance of a collaborative effort between laboratories to maintain informative families for cattle. Moreover, Maijala (1989) proposed reference families by collecting data, which do not only include marker genes but physiological and immunological traits and results from *in-situ*-hybridization of chromosomes and somatic cell techniques. To this end, individuals of extremely different breeds can be crossed, followed by a crossing of the F1 generation or a back-crossing on both parental breeds (Figure 1). Then the generated individuals are maximally heterogenic both, for their genotypes and for their phenotypes that are responsible for differences between the original breeds or strains.

MAP-BASED GENE CLONING

Methods are well-established for isolation and cloning of a gene on the basis of its product. However for many important genes the products are unknown in farm animals. Even for genes already identified by classical genetics, the mechanisms by which they act are largely unknown. One approach to clone genes without knowledge of their products is the transposon tagging (Baker et al., 1986). But, a large number of transposon-mutagenized individuals must be screened which limits the value for farm animals. The other method for a product independent gene isolation is offered by a map-based cloning. Often referred to as reverse genetics (Orkin, 1986), this approach is based on physical linkage to mapped RFLP markers. Once tightly-linked RFLP markers are known, the flanking gene regions are identified using e.g. libraries with overlapping clones. A move or "walk" along the chromosome from the cloned gene to the gene of interest has already

been used to clone genes involved in hereditary diseases of human (Bender et al., 1979). For reasonable chromosomal distances large DNA segments were cloned in cosmid or yeast chromosome vectors (Poustka et al., 1984; Cooke, 1987). Poustka and Lehrach (1986) and Poustka et al. (1987) described linking and jumping libraries containing linked terminal DNA sequences from large DNA fragments. The approach of reverse genetics, however, needs a method to identify the DNA segment that actually contains the gene(s) of interest and is only feasible for genes physically tightly linked with the marker (Tanksley et al., 1989).

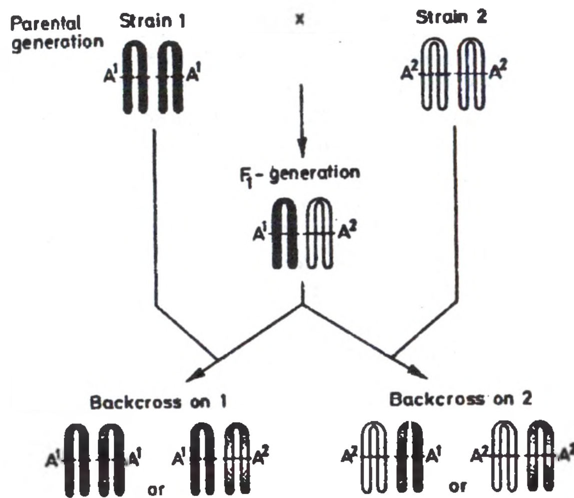


Figure 1 Outline for a backcross experiment to produce reference families (according to Kluge and Geldermann, 1982).
The linkage situation is given for a gene locus with the genotypes A^1A^1, A^1A^2 and A^2A^2. For the backcross generation the average arrangements of genes within the chromosome are shown for many individuals.
Examples for parental strains 1 and 2: More general mapping - in cattle, a European and a Zebu breed; in pigs, wild boar and a domesticated breed; in sheep, a merino breed and caraculs. Analysis of special genes - Zebu and N'Dama crosses established in Africa to study the inheritance of trypanosomiasis (Soller and Beckmann, 1987); homozygous halothane positive and negative individuals of Landrace pigs; Boroola merinos and European merino breeds.

ANALYSIS OF RELATIONS BETWEEN KNOWN MARKER GENES AND HYPOTHETICAL TARGET GENES FOR QUANTITATIVE TRAITS BY INTERVAL GENE MAPPING (IGM)

As mentioned, for most of the genes relevant in animal breeding no primary products or functions are known. To study such genes, an Interval Gene Mapping (IGM) may serve as a first crucial step. For this purpose, informative family groups can be used where marker genes segregate and also genes influencing the variance of the production criteria in question. Thereby, the association of variation in production criteria with single marker genes allows an indirect identification of closely linked but unknown target genes. This implies, in principle, a resolving of genetic components of a more or less complex trait.

However, heritable multifactorial characters, often referred to as quantitative, are influenced by a combined action of several genes. For expression of such characters little is known on the number, chromosomal position and effects of genes. Therefore the problem is that only high numbers of DNA markers make it feasible to measure and map the effects of genes underlying quantitative traits (Quantitative Trait Loci, QTL; Geldermann, 1975). To detect QTLs two individuals may be crossed which are genetically different for quantitative characters of interests. The progeny (F_2 or backcross generation) is then obtainable, segregating for these traits as well as for DNA markers. An outline of an analysis of effects on a quantitative trait assessable from gene markers in a F2-generation of breeding strains is given in Figure 2. Associations between a segregating marker gene and the values of quantitative characters should then be due to linkage of the marker gene to one or several QTLs. The ability to detect QTLs by marker genes is a function of the magnitude of QTL effects, the number of offspring studied, and the recombination frequency between marker gene and QTL(s). Likelihood maps can be established to show chromosomal intervals and effects for within located QTLs, as described for the genome of tomato (Tanksley et al., 1989). Linked markers, bounding an interval which may contain QTLs, will reduce the likelihood, that genotypes at the marker genes differ from genotypes at within linked QTLs, to the square of the recombination frequency between the flanking markers. For a demonstration of effects and interactions of QTL(s) each interval between marker genes can be studied as a discrete entity, and at last a biophysical basis of complex traits will be available (Tanksley et al., 1989).

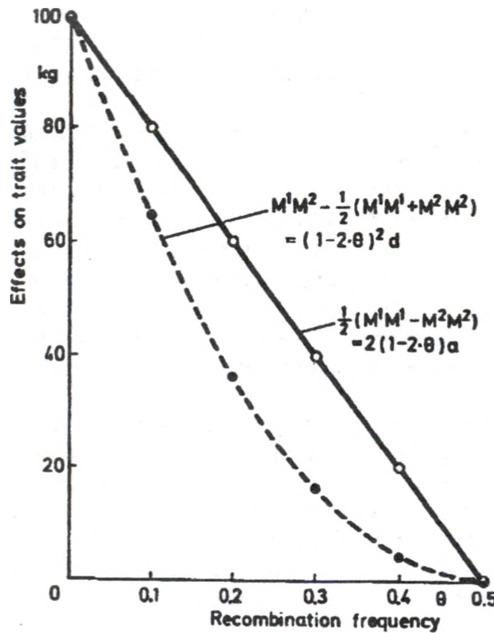
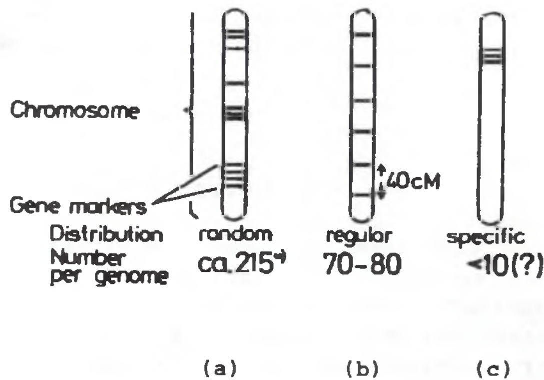


Figure 2 Estimate of genetic effects on a quantitative trait by a segregation analysis in a F2-generation using the example of milk yield in cattle.
Assumptions: Parental sources with different alleles at a marker gene M and a target gene A ($M^1M^1A^1A^1$ and $M^2M^2A^2A^2$); effects of the gene locus A on milk yield: a = d = 100 kg.
For a target gene with the genotypes A^1A^1, A^1A^2 and A^2A^2, effects on the quantitative trait are designed as +a, d resp. -a (Falconer, 1981). Then, from a marker gene, with a recombination frequency θ to the target gene and with the genotypes M^1M^1, M^1M^2, M^2M^2, the following effects on the quantitative trait are measured in the F_2-generation: $(1-2\cdot\theta)a+2\cdot\theta(1-\theta)d$; $((1-\theta)^2+\theta^2)d$; $-(1-2\cdot\theta)a+2\cdot\theta(1-\theta)d$ resp.. The mean trait values are compared for individuals with different marker genotypes: $\frac{1}{2}(M^1M^1 - M^2M^2)$ and $M^1M^2 - \frac{1}{2}(M^1M^1+M^2M^2)$.

To dissect complex traits into their genetic components many mapped marker loci are necessary. As early as 1980 Botstein et al. proposed to cover the whole genome of a species with polymorphic DNA markers. Each gene should then be judged by gene linkage. Assuming a genome size of 2500 cM and aiming at a probability of greater 95% for a linkage of 20 cM or less between marker genes and unknown genes,

about 215 randomly distributed marker genes (Beckmann and Soller, 1983) have to be used (Figure 3a). The number of marker genes is reduced to about a third if their loci are regularly distributed (Figure 3b). In some cases a few chromosome sections may contain the most important genes ("major genes") of the considered trait, e.g. the "halothane gene" region in pigs if regarding influences on meat quality and stress resistance. Then the number of genes necessary for breeding purposes is considerably reduced (Figure 3c) because markers of flanking regions of target genes will be sufficient.



(a) (b) (c)

Figure 3 Number of polymorphic marker genes applied for a genome with 2,500 cM.
*) According to Beckmann and Soller (1983) as explained in the text.

DETECTION OF GENOMIC SIMILARITIES AND DIFFERENCES IN GENE ARRANGEMENTS

Comparative gene mapping and the identification of conserved linkage groups (O'Brien, 1986; Lalley et al., 1988) stimulated new research activities. Using sets of clones for RFLP mapping the degree to which chromosome content and gene order have been conserved was determined. To this end, gene clones from one species were mapped in different species. The high degree of linkage conservation found between species suggests that substitutions of chromosomes or chromosomal segments from one species to the other might be definable. For example, all

members of the Bovidae family share the same basic number of chromosomal arms (29 autosomal arms per haploid set). If the gene contents and orders within arms were highly conserved among related species, single chromosomes or chromosomal segments might be substituted between species (e.g. by somatic cell hybridization; Rossant et al., 1982) to combine genes that are not available in the normal crossing range of a species.

On the other hand, different arrangements of genes have been described even within species (Mettler and Gregg, 1969). Investigations on gene arrangements in different farm animal populations can achieve information on evolution or special effects of gene clusters and provide new parameters for breeding programs. For example, genomic differences may be associated with fertility and viability. This may become evident when cross-bred generations are formed between populations with large genetic distance, e.g. Zebu and European cattle.

CONCLUSIONS

New approaches of biotechnology have a significant impact on the progress of gene mapping. Thus, activities of gene mapping are widespread in several countries and are supported by far-reaching experiments. Recombined DNA sequences, polymorphic DNA regions and gene maps deliver important information. Methods and DNA probes arise which are usable for analyzing complex traits for genetic components. In animal genetics, gene mapping does not only enhance basic knowledge but supports future practical breeding programs as well. Recent developments in molecular genetics suggest an approach by which breeders could reduce effects of environmental variation on selection response and sustain various goals, e.g. the increase of efficiency, the improvement of animal product quality and the development of alternative products. To gain these goals, combinations of conventional methods along with new technologies of genome research are required.

REFERENCES

BAKER, B., SCHELL, J., LÖRZ, H. and FEDERAFF, N. 1986. *Proc. Natl. Acad. Sci. U.S.A. 83:* 4844-4848.
BECKMANN, J.S. and SOLLER, M. 1983. *Theor. Appl. Genet. 67:* 35-43.
BENDER, W., SPIERER, P. and HOGNESS, D. 1979. *J. Supra Molec. Structure 10:* 32.
BOTSTEIN, D., WHITE, R.L., SKOLNICK, M. and DAVIS, R.W. 1980. *Am. J. Hum. Genet. 32:* 314-331.
BUFTON, L., MOHANDAS, T.K., MAGENIS, R.E., SHEELY, R., BESTWICK, R.K. and LITT, M. 1986. *Hum. Genet. 74:* 425-431.
CLERGET-DARPOUX, F. and BAUR, M.P. 1990. Linkage studies and mapping. In: *Genome analysis in domestic animals.* VCH, Weinheim.
COOKE, H. 1987. Trends in Genetics 3: 173-174.
FALCONER, D.S. 1981. *Quantitative Genetics.* 2nd Edition. Longman, London & New York.
FRIES, R., BECKMANN, J.S., GEORGES, M., SOLLER, M. and WOMACK, J. 1989. *Anim. Genet. 20:* 3-29.
GELDERMANN, H. 1975. Theor. *Appl. Genet. 46:* 319-330.
HETZEL, D.J.S. 1988. Genomic DNA banks. *Workshop on RFLP Markers, Gene Mapping and Linkage Studies in Domestic Animals.* Mt. Victoria: 6 pp.
Human Gene Mapping. 1987. Proc. 9th International Workshop Human Gene Mapping (Paris). *Cytogenet. Cell Genet. 46.*
Human Gene Mapping. 1988. Update of the 9th International Workshop Human Gene Mapping (Paris). *Cytogenet. Cell Genet. 49:* 1-258.
JEFFREYS, A.J., WILSON, V. and THEIN, S.L. 1985. *Nature 314:* 67-72.
JEFFREYS, A.J., WILSON, V., THEIN, S.L., WEATHERALL, D.J. and PONDER, B.A. 1986. *Am. J. Hum. Genet. 39:* 11-24.
JEFFREYS, A.J. and FLAVELL, R. 1977. *Cell 12:* 429-439.
KLUGE, R. and GELDERMANN, H. 1982. *Theor. Appl. Genet. 62:* 1-4.
LALLEY, P.A., DAVISSON, M.T., GRAVES, J.A.M., O'BRIEN, S.J., RODERICK, T.H., DOOLITTLE, D.P. and HILLYARD, A.L. 1988. *Cytogenet. Cell Genet. 49:* 228-235.
MAIJALA, K. 1989. Gene mapping in cattle. *4th International Symposium "Progress in Cattle Breeding",* Czechoslovakia.
METTLER, L.E. and GREGG, T.G. 1969. *Population genetics and evolution.* Prentice-Hall, Englewood Cliffs, New Jersey.
MORTON, N.E. 1956. *Am. J. Hum. Genet. 8:* 80-96.
NAKAMURA, Y., LEPPERT, M., O'CONNELL, P., WOLFF, R., HOLM, T., CULVER, M., MARTIN, C., FUJIMOTO, E., HOFF, M., KUMLIN, E. and WHITE, R. 1987. *Science 235:* 1616-1622.
O'BRIEN, B.J. 1986. Genetic analysis in mammals: past, present and future. *Genet. Engin. Anim.* (Eds. J. WARREN and A. HOLLAENDER) Plenum Press: pp 139-149.
ORKIN, S.H. 1986. *Cell 47:* 845-850.

POUSTKA, A. and LEHRACH, H. 1986. *Trends Genetics 2:* 174-179.
POUSTKA, A., RACKWITZ, H.-R., FRISCHAUF, A.-M., HOHN., B. and LEHRACH, H. 1984. *Proc. Natl. Acad. Sci. USA 81:* 4129-4133.
POUSTKA, A., POHL, T.M., BARLOW, D.P., FRISCHAUF, A.-M. and LEHRACH, H. 1987. *Nature 325:* 353-355.
ROSSANT, J., CROY, B.A., CHAPMAN, V.M., SIRACUSA, L. and CLARK, D.A. 1982. *J. Anim. Sci. 55:* 1241-1248.
SOLLER, M. and BECKMANN, S. 1987. Toward an understanding of the genetic basis for trypano-tolerance in the N'Dama cattle of West Africa. Consultation Report, FAO, Rome.
SOLNICK, M.H. and WHITE, R. 1982. *Cytogenet. Cell Genet. 32:* 58-67.
TANKSLEY, S.D., YOUNG, N.D., PATERSON, A.H. and BONIERBALE, M.W. 1989. *Bio/Technology 7:* 257-264.
THOMPSON, E.A., KRAVITZ, K., HILL, J. and SKOLNICK, M. 1978. Linkage and the power of a pedigree structure. In: *Genetic Epidemiology* (Eds. Morton, N.E., & C.S. Chung) Academic Press, New York: pp 247-253.
WHITE, R. and LALOUEL, J.M. 1988. *Sci. Amer. 258:* 20-28.