Future trends in Animal Breeding due to new genetic technologies

M. A. Toro

Departamento de Producción Animal, ETS Ingenieros Agrónomos, Universidad Politécnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

The Darwin theory of evolution by natural selection is based on three principles: (a) variation; (b) inheritance; and (c) natural selection. Here, I take these principles as an excuse to review some topics related to the future research prospects in Animal Breeding. With respect to the first principle I describe two forms of variation different from mutation that are becoming increasingly important: variation in copy number and microRNAs. With respect to the second principle I comment on the possible relevance of non-mendelian inheritance, the so-called epigenetic effects, of which the genomic imprinting is the best characterized in domestic species. Regarding selection principle I emphasize the importance of selection for social traits and how this could contribute to both productivity and animal welfare. Finally, I analyse the impact of molecular biology in Animal Breeding, the achievements and limitations of quantitative trait locus and classical marker-assisted selection and the future of genomic selection.

Keywords: animal breeding, Darwin, genetic technologies, imprinting

Introduction

The Origin of Species remains today the most significant text in the history of biology. Moreover, unlike what happens in other branches of science, where the references to the past are little more than a product of the historian’s erudition, very different aspects of Darwin’s original thinking are still under investigation at present. The relationship between Animal Breeding and evolution has been so close that theoretical developments of the former have contributed to those of the latter and vice versa. Notwithstanding, quantitative genetics, upon which Animal Breeding is based, was founded by the originators of the modern synthesis.

Darwin’s theory of evolution by natural selection is based on three principles: (a) variation (individuals differ from each other for particular characteristics); (b) inheritance (related individuals resemble more than unrelated individuals); and (c) natural selection (individuals with different characteristics differentially survive and leave offspring to the next generation). Here, I take these principles as an excuse to review some topics related to the future research prospects in Animal Breeding. This review is a bit idiosyncratic, without full coverage of all of the possible topics involved.

The principle of variation

Mutation (new variation) and recombination (shuffling of existing variation) provide the raw material for continued evolution. On the evolutionary scale, other forms of increasing genetic information have been important as, for example, symbiosis in the origin of eukaryotic cells.

Recently, however, other sources of variation different from mutation are becoming increasingly important. I will describe two of them: (a) variation in copy number (CNV); and (b) microRNAs (miRNAs). CNV refers to a segment of DNA in which copy number differences have been found by comparison between two or more genomes. The segment may range from 1 kb to several Mb in size. In humans, it is believed that approximately 0.4% of the genomes of unrelated people differ with respect to copy number and this has been associated with lung cancer, susceptibility to HIV, autism and schizophrenia. As a complement to the Bovine HapMap Consortium project, Liu et al. (2008) initiated a systematic study of the CNV and demonstrated that significant amounts of germline and fewer somatic CNVs exist in cattle, that many CNVs are common both across diverse cattle breeds and among individuals within a breed. Some of the CNV regions seem to be related to immunity and metabolism. Fadista et al. (2008) report a first account of CNVs in the pig genome based on a family material comprising 14 boars, 700 sows and about 12 000 offspring and covering part of the chromosomes 4, 7, 14 and 17 already sequenced and assembled. Using a strict CNV detection method, 37 copy number variable regions across all four chromosomes were identified. One of them, the mast/stem cell growth factor receptor (KIT) mutation is responsible for coat colour phenotypes in domestic pigs (white belt in Hampshire and white colour on Pietrain, Landrace and Large White).

MicroRNAs are single-stranded RNA molecules of 21 to 23 nucleotides in length, which regulate gene expression.
miRNAs are encoded by genes whose DNA is transcribed but not translated into protein, and mature miRNA molecules are complementary to one segment of the mRNA and their function is to downregulate gene expression. Although the first miRNA was identified years ago, it is only recently that we have begun to understand the function and diversity of these regulatory molecules. Lee et al. (1993) discovered that lin-4 in the worm Caenorhabditis elegans did not code for a protein, but instead produced a pair of short RNA transcripts each regulating the timing of larval development by trans-
lation repression of lin-14, which encodes for a nuclear protein. It was 7 years later that the second miRNA, let-7, was discovered in the same species, but thousands of miRNAs have been identified since then in organisms as diverse as viruses. An miRBase, hosted by the Faculty of Life Sciences, University of Manchester, is publicly available (http://www.mirbase.org/) and has more than 10,000 entries, out of which 1271, 77, 479, 615 and 4 corresponds to humans, pigs, chicken, cattle and sheep, respectively, and it seems that many of them are conserved among species. Several miRNAs have been found to have links with some types of cancer and to heart development and heart diseases. MicroRNAs also seem to regulate host-pathogen interactions in different directions (virus→virus, virus→host, host→virus) and stages of the viral life cycle (infectious, latent) and therefore pathways (replication, apoptosis, infection).

Because of the increasing evidence suggesting that miRNAs participate in muscle development in mice and humans, recently Xie et al. (2009) experimentally detected and identified 57 distinct miRNAs conserved and unique miRNAs from porcine skeletal muscle. Another prime example is the Belgian Texel sheep, famous for their hyper-developed muscles. A QTL study of the phenotype (Clop et al., 2006) uncovered a single-nucleotide polymorphism (SNP) in the the 3′ UTR of the myostatin gene (gdf8), which is involved in limiting the growth of muscle tissue. The G→A SNP creates target sites for mir-1 and mir-206, which result in downregulation of myostatin and thereby of higher muscle growth.

MicroRNA research has profoundly changed the perceptions of the role of RNAs from rather uninteresting carriers of coding information to key players in cellular regulation, implying an additional source for genome plasticity. Question on how novel RNAs contribute to an increase in genome complexity and how they lead to the emergence of novel traits remain largely unanswered.

The principle of inheritance

There was a deep flaw in Darwin’s theory: it lacks a theory of heredity. Darwin, like many of his contemporaries, speculated that the characteristics of the parents were blended – like mixing paint – as they passed to the offspring. But if that was true, some of Darwin’s critics, notably the Scottish engineer Henry Fleming Jenkin, pointed out, how could a single fortunate mutation be spread through a species? It would be blended out, just as a single drop of white paint would be in a gallon of black one. Evolution does not work with blending inheritance: the new variants would rapidly be lost through mating to the common form.

Although Darwin knew that there should be a mechanism for parental characteristics to be transmitted to its progeny, he was completely unaware of this transmission mechanism. His close contact with the practices of farmers producing new varieties of ornamental plants or new breeds had not given him the key mechanism. The reason was simple: never came to fix his attention on a single property but on complex morphological traits. The first is exactly what Mendel did in his studies with the pea (Pisum sativum). He chose individual characteristics that differed from each other unequivocally and who also had the peculiarity of the trait being present or absent: plant height (tall or short), flower colour (white or purple), seed colour (yellow or green) and seed shape (smooth or wrinkled). Thus, he could apply a careful mathematical analysis that allowed him to analyse the results of the experiments in a rigorous way. The eminent population geneticist R. C. Lewontin has suggested that in full historical justice, if we were to personalize our modern explanation of evolution, we should call it neither ‘Darwinism’, nor even ‘Darwin–Wallaceism’, but ‘Darwin–Wallace–Mendelism’. The idea of inheritance of acquired characters is associated to the French biologist J. B. Lamarck, although it was held by most evolutionists, including Darwin. It was A. Weismann who set up the concept of the independence of germ line and soma. In most animals, cells that are going to give rise to germ cells are put aside very early in development. Further, as Weismann realized, a process cannot be imagined for the flow of information from the soma to the germ line. In his classical example, suppose that a blacksmith does develop big muscles. How could that alter the sperm he produces so that some of his sons resemble him? The sperm cell has no muscles, so the big muscles of the father should be translated into some code, and later on, this code will be translated back in the muscles of the sons. It is quite remarkable that Weismann was the first to understand that heredity is a matter of information. He, for example, remarks that to accept the inheritance of acquired characters would be ‘very like supposing that an English telegram to China is received there in the Chinese language’.

Non-Mendelian inheritance: epigenetic effects

Today, we would express Weismann’s argument in molecular terms as the central dogma of molecular biology that affirms that genetic information is passing from DNA to RNA and from RNA to protein, but not from protein to DNA. Although this assertion is essentially true, there are cases that are quoted as examples of Lamarckian inheritance (Maynard Smith, 1997). For example, the members of a clone of Daphnia can have different morphologies as a consequence of environmental changes; they develop spines in the presence of predators. The change is adaptive and it is transmitted through the egg. Another very classical example occurs in flax (Linum) that suffer morphological changes when the plants are treated with high levels of fertilizer and that persist for a number of sexual generations. Finally, on
the surface of ciliate protozoa there are complex patterns of cilia. If the pattern of an individual is changed, either by accident or by artificial manipulation, the new pattern may be transmitted through many binary fissions.

The term epigenetics was coined by C. H. Waddington in 1942 by joining the words genetics and epigénesis. Epigenesis was an old word to describe the differentiation of cells from their initial totipotent state in embryonic development. In modern terms it refers to heritable traits in the phenotype (or gene expression) that are not caused by changes in the underlying DNA sequence. It has been claimed, for example, that in humans the paternal granddaughters of women who experienced famine while in the womb lived shorter lives.

**Genomic imprinting**
The most well-known epigenetic effect in animal breeding is genomic imprinting. This is a genetic phenomenon by which certain genes are expressed in a parent-of-origin-specific manner: either only from the allele inherited from the mother (e.g. H19 or CDKN1C), or from that inherited from the father (e.g. IGF2). It involves methylation and histone modifications to achieve monoallelic gene expression without altering the genetic sequence that it is established in the germline and maintained throughout all somatic cells of an organism. In mammals, about 1% of genes are imprinted, depending upon its parental origin. In the pig, the IGF2 region, located in a telomeric position on chromosome 2, harbours a paternally expressed mutation that increases muscle growth and leanness. The causative mutation (intron.3-g.3072G.A) occurs in a CpG island of intron 3, which has a regulatory role; pigs receiving the A allele from their sire have a threefold increase of IGF2 mRNA in muscle (Van Laere et al., 2003). The mutation has a considerable effect, explaining about 10% to 30% of the total phenotypic variability for these traits and has been confirmed in several independent studies (Estelle et al., 2005). An allele of the ‘callipyge’ (from the Greek for ‘beautiful buttocks’) or CLPG gene in sheep produces large buttocks consisting of muscle with very little fat. The large-buttocked phenotype only occurs when the allele is present in the copy of chromosome 18 inherited from a sheep’s father and is not on the corresponding copy of inherited from the sheep’s mother (Georges et al., 2004).

From an evolutionary point of view a hypothesis widely accepted for the evolution of genomic imprinting is the ‘parental conflict hypothesis’ (Moore and Haig, 1991). This hypothesis states that the inequality between parental genomes due to imprinting is a result of the differing interests of each parent in terms of the evolutionary fitness of their genes. The father is often more ‘interested’ in the growth of his offspring, at the expense of the mother. The mother’s interest is often to conserve resources for her own survival while providing sufficient nourishment to current and subsequent litters. Accordingly, paternally expressed genes tend to be growth promoting, whereas maternally expressed genes tend to be growth limiting. In support of this hypothesis, many imprinted genes have a growth-related function; paternally expressed genes tend to increase growth, whereas maternally expressed genes tend to decrease growth. Furthermore, genomic imprinting has been found in all placentals mammals, where post-fertilization offspring resource consumption at the expense of the mother is high; it has not been found in oviparous birds or monotremes (a class of oviparous mammals), where there is relatively little post-fertilization resource transfer and therefore less parental conflict. From these ideas about imprinted genes, it is reasonable to speculate that part of the genetic variation for growth-related traits of economical importance in livestock might be due to polymorphisms in imprinted genes.

**The principle of natural selection**
Back from his trip on the Beagle, Darwin gestated his theory of evolution by natural selection and wrote the first draft of it in 1842. Without doubt a key factor in developing the concept of natural selection was the work of R. Malthus’ *Essay on the Principle of Population*, from which Darwin took the idea of struggle for existence. The thesis of Malthus argued that while populations grow according to a geometric progression, food grow as an arithmetic progression. This results in an intra-population struggle for resources. Historians have endlessly debated what exactly shocked Darwin when reading of Malthus. It seems that one important element was the change in perspective from giving priority to competition between individuals of a species against that of interspecific competition. When a population has abundant resources, it grows rapidly; but as it expands, the resources become limiting factors and the census of the population stabilizes. It thus reaches an equilibrium that arises as an unintended consequence of the conflict of interest between individual units.

What Darwin discovered reading Malthus is that this balance is more apparent than real. The equilibrium does not lead to an optimal situation, but continues the struggle, the evil and the suffering. Moral issues were not what impressed Darwin. He took notice of what might be called environmental aspects applicable to the population. He realized that in all phases of the growth curve of a population, including the balance, individuals do not use resources in a symmetrical manner. There is always competition that means that some individuals contribute more than others to the next generation. This idea, which today seems basic ecology, was not perceived at the time of Darwin. The prevalent idea in economics argued that a balance was achieved between resources and population census and then nothing changed, that is, competition did not lead to continuous change, but to a static situation. Malthus’s ideas about the importance of conflict and individual competence in the population enabled Darwin to find a completely different answer to explain the presence of design in living things: the struggle for survival, the Darwinian natural selection that could generate adaptation.

Both in *The Origin of Species by Means of Natural Selection* and in his later work *The Variation of Animals and Plants under Domestication*, Darwin showed that natural selection
is a process analogous to artificial selection that man applied to plants and domestic animals. He was aware that artificial selection could not have an effect at once, but by gradual accumulation of changes through selective mating. Although the similarity between the two types of selection is visible, there are important differences. Artificial selection is guided by a breeder that imposes the desired criteria, including the production of milk, fruit size, etc., and maintains it during the successive generations. In natural selection, there is no plan, only individuals that reproduce differentially: some leave more children than others depending on the environmental circumstances that may be changing with each generation.

Natural selection
Natural selection could be seen as survival and/or differential reproduction of units capable of replication, which is the same as to say that these units differ in their biological fitness. This notion is usually applied to individuals, and in loose words, is defined as the contribution of offspring to the next generation and, as indicated before, for the selection of the parents to produce a response in the offspring is necessary that the differences in replicative capacity show some hereditary component, that is, the similarity between the fitness of parents and children must be greater than between individuals belonging to two successive generations randomly sampled from the population.

Natural selection plays a role in the changes in the genetic structure of populations by promoting the adaptation of individuals to their environment. However, it is very important to note that not all evolutionary changes are explained by the action of natural selection. Some of them are due to random processes (genetic drift) that occur, especially in small populations, because some individuals leave more offspring than others by chance. These processes involve evolutionary changes, but no adaptation. In a sense, natural selection acts as an algorithmic process, that is, as a formal, blind, working process from a set of simple rules, but the presence of chance in the same prevent us from exactly predicting the end result.

The action of natural selection promotes the adaptation of organisms to their environment. Natural selection can explain the structural and functional complexity of organisms without having to resort to a creator. However, natural selection is not perfect but a mechanism of adaptation is accompanied, not infrequently, by opportunistic solutions or even the design of obvious errors, such as occurs in the arrangement of the nerve endings of the retina cells that causes a blind spot when they meet to form the optic nerve. In animal breeding we know that genetic improvement through selection for one trait may lead to a change (improvement or worsening). This was recognized by Darwin in The Origin of Species:

Hairless dogs have imperfect teeth; long-haired and course-haired animals are apt to have, as is asserted, long or many horns; pigeons with feathered feet have skin between their outer toes; pigeons with short beaks have small feet, and those with long beaks large feet. Hence if man goes on selecting, and thus augmenting any peculiarity, he will almost certainly modify unintentionally other parts of the structure, owing to the mysterious laws of correlation (quoted in Flint and Woolliams, 2008).

Natural and sexual selection
One of the forces that can lead to maladaptation is sexual selection. This is a special case of natural selection. Sexual selection acts on an organism’s ability to obtain or successfully copulate with a mate. The fantastic array of feathers of the peacock’s tail lead Darwin to consider that the presence of certain characters in males are not because they have survival value; if so, it is difficult to imagine why females do not have them, but because these ornaments attractive and therefore are traits that facilitate mating. The theory of sexual selection tries to explain the evolution of ornamental characters operating upon by two opposite forces. Larger and ornamental males would be more successful in mating. Minor exhibitionism facilitates survival against predators. The final outcome would depend on the balance between these two forces. Now the question is what is the origin of female preferences? Why do females prefer wings of exuberant plumage? It was R. A. Fisher, one of the main founders of theoretical neo-Darwinism, who offered the first solution. The reason why females prefer males with an extreme trait is simply that there are other females who also prefer them. Although this looks like a circular argument, it is not. Once, for any reason, even arbitrary, some females show a preference for a particular ornament, females not showing this preference do not have sons that show that ornament and, therefore, they are not appealing to the females. This establishes a co-evolution between the character exhibited by the male and the preference shown by the female, a selection process that has been termed runaway selection.

These ideas are relevant in a practical setting as been shown by the well-known animal breeder of Purdue University W. Muir together with R. Howard. Although production of transgenic organisms offers great agricultural potential, there is some concern that the introduction of genetically modified organisms into natural populations could result in ecological risk, such as species extinction. It is sometimes thought that this threat is not important because transgenic organisms are evolutionary novelties that would have reduced viability. However, transgenic organisms may also possess an advantage in some aspect of reproduction that may increase their success in nature. For example, a commonly desired characteristic in transgenic fish species (important in aquaculture and sport fishing) is accelerated growth rate and larger adult body size.

Muir and Howard (1999) conducted experiments with transgenic medaka – a fast-breeding Japanese fish. They found that normal medaka males, which are larger than average, are four times more successful in breeding. Transgenic medaka males, which are still larger, are up to seven times more successful in breeding. But they also found that the offspring of transgenic males had a low survival rate. Then they introduced the real parameter values into a
computer model to see what might happen if 60 transgenic medaka males were introduced into a population of 60,000 wild medaka. The wild medaka female naturally sees the larger transgenic male as the more attractive mate. But looks are deceiving, in this case, because even though the transgenic male is bigger and mates better, his offspring die sooner than those of his smaller, wild counterpart. In just 40 generations, the whole of the species is driven to extinction. They coined the term 'Trojan gene effect' because the transgenic medaka males hide inside an attractive package the gene that can destroy an entire population.

**Selection for social traits**

*Altruism and cooperation: group selection*

The existence of altruistic behaviour poses a challenge to the neo-Darwinian interpretation of the behaviour: how does natural selection encourage behaviour that is harmful, in terms of lower biological fitness, for the individuals who practice it? A first alternative, already proposed by Darwin, is called selection between groups. This is a type of Pareto's optimum. Its logic is simple. When studying a type of behaviour we should not only consider its implications for the individual who shows it, but also for the rest of the individuals of his group. If a behaviour benefits all of them, natural selection will favour it, while if it is harmful to all it will disappear. Now, if there is a negative impact on the individual but a positive one for the group, the answer will depend on the relationship between costs and benefits.

Until the sixties, it was thought that there were many characteristics of living things that had not emerged to promote the survival of the individual but, in the terminology of that time, for the good of the species. This was the position taken up by Darwin:

> There can be no doubt that a tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to give aid to each other and to sacrifice themselves for the common good, would be victorious over other tribes; and this would be natural selection (The Descent of Man, 1871).

It can be said that nowadays most evolutionary biologists doubt whether this process can be effective and think that natural selection acts by favouring some individuals over others and no one group over others. Furthermore, the maintenance of altruistic behaviour by selection between groups is essentially unstable, since an altruistic group can always be invaded by selfish individuals by mutation or migration and would be favoured by natural selection, since they would receive benefits without cost. To counter this effect would require a very high rate of extinction and formation of new groups, which seems not to be a common situation in most species.

Although there are theoretical reasons that the effectiveness of selection between groups is limited, there may be situations in which it may happen. For example, in the context of the interactions between parasites and hosts, it is expected that in many viruses natural selection favours intermediate virulence. An active virus will have an advantage in competition with others within the same host but, as a result of this increased virulence, the host will very soon die, which will limit the chances of transmitting the virus to other guests. It seems that this is what happened to the myxomatosis virus when introduced into Australia in the 50s with the aim of ending the plague of rabbits. Twenty years later, it was observed that deaths caused by the virus were much less than initially, and although part of this effect was because the rabbits had acquired some resistance, it was also shown that the virulence was reduced.

*Kin selection*

In a classic article published in 1964, the then young British biologist W. Hamilton offered an explanation of altruistic behaviour alternatives to group selection, which is known as kin selection. This author noted that if a particular gene induces an individual to sacrifice his life, saving those of various relatives, the number of copies of that gene in subsequent generations could grow faster than if the sacrifice had not been done, since relatives have a higher probability of being carriers of the same genes than other individuals of the population and that probability increases as the relationship is close. In short, the altruistic behaviour is a cost for the individual who performs it, but entails a benefit to those who interact with him and, if these individuals are relatives, this benefit will reverse indirectly in the selfless. R. Dawkins has popularized the mechanism with the expression 'selfish gene', which regards its actions neither as beneficial to the group, nor to the individual, but to the conditional gene itself. A similar idea to the selfish gene concept has some acceptance in the field of molecular biology. In most higher organisms, a considerable fraction of DNA contains a large number of copies of the same unit with no known function. It has been suggested that such sequences are selfish or parasitic institutions, which multiply by forming additional copies of themselves within the genome.

Hamilton established the relationship between the cost attributed to the author of altruistic behaviour (c) and the benefit to their receptors (b). This relationship is known as Hamilton's rule and is stated as follows: a trait will be favoured by natural selection if the product \( r \times b \) is greater than \( c \), where \( r \) is the proportion of genes shared by the author and the recipient (the additive relationship coefficient in animal breeders jargon). The coefficient \( r \) is equal to 0.5 if they are brothers, 0.25 if they are half siblings and 0.125 if they are cousins. In this sense it is noted that Haldane, one of the founders of the neo-Darwinian synthesis, said in the 30s while drinking beer in a pub, that he would not mind risking his life if that would save two brothers or eight cousins. The idea was also recalled by the famous animal breeder J.L. Lush in his book:

> The competition and selection among families thus introduced could make selection favor any genes which
tend to cause their possessor to sacrifice himself for his own welfare. This is, provided the sacrifice promotes the biological welfare of his relatives (some of whom will have some of the gene he has) enough to more than compensate for the genes lost in his own sacrifice (The Genetics of Populations, 1951).

One of the biggest successes of Hamilton ideas is the interpretation of the evolution of social insects where this behaviour has emerged 12 times, of which 11 belong to the order Hymenoptera. In this group, there is an haplodiploid system of sex determination: the males develop from eggs without fertilization and, consequently, have a single chromosome that is inherited from his mother, while females come from fertilized eggs and have inherited, therefore, a chromosome from the father and another one from the mother. The most striking result of this particular mechanism is that the female share three quarters of their genes with their sisters but only half with their daughters, that is, the sisters are in some sense super-relatives. Remembering that their sisters but only half with their daughters, that is, the sisters are in some sense super-relatives. Remembering that each colony is produced from a single queen and the proportion of females is higher than that of males, one of the predictions of kin selection model is that altruistic behaviours will be exhibited by females, since their relationship with the rest of the colony is increased. Indeed, the workers are female members of the colony that made the most useful work, such as regulating temperature, brood care, foraging and defense of the group. The males, rightly called drones, hardly contribute anything to the welfare of the colony.

Reciprocal altruism
A third form of how cooperative altruism can evolve is if there is reciprocity, the benefits and costs of altruistic behaviour are balanced over time between pairs of interacting individuals. If individuals take turns as authors and recipients of altruistic actions the benefits of altruism in the long term can outweigh costs. This theory was first proposed by Trivers (1971), but corresponded again by the biologist Hamilton and the economist R. Axelrod through the mathematical formulation in the framework of the prisoner’s dilemma (Axelrod and Hamilton, 1981). This theory received a lot of attention but it will not be reviewed here.

Group selection and kin selection in animal breeding
Starting just shortly after the publication of Hamilton’s paper, Griffling (1967, 1968a and 1968b, 1969, 1976a and 1976b, 1981a and 1981b) provided a rigorous approach to the quantitative genetics of the interactions between individuals including the theory of selection response. However, his work had limited impact because it lacks an approach to the problem of identifying heritable components and an implementation in the framework of mixed model equations. The last has been achieved mainly by B. Muir and P. Bijma in several very elegant papers (Bijma et al., 2007a and 2007b; Ellen et al., 2007; Bijma and Wade, 2008).

Griffing shows that the phenotype $P_i$ of a particular individual $i$ included in a group of $n$ other interacting individuals can be decomposed into a direct effect $P_{D,i}$ from individual $i$ plus the sum of all the associate effects $P_{S,j}$ of others in its group.

$$P_i = P_{D,i} + \sum_{j \neq i} P_{S,j}$$

Thus, each individual has two unobserved effects: a direct effect expressed in its own phenotype and an associative effect expressed in the phenotypes of its associates.

For example, in commercial egg production, individuals are frequently beak trimmed to reduce mortality due to pecking behaviour. In canibalistic pecking behaviour in chickens, the direct effect of an individual indicates its ability to survive by avoiding being pecked, whereas the associate effects refer to the effect on its own survival of the pecking behaviour of its cage members. Notice that a maternal effects model, in which association is between mother and offspring, is a well-known special case of the above equation.

Phenotypic direct and associative effects in equation (1) are both decomposed into breeding values ($A$), and environmental effects ($B$), in which $A_{D,i}$ is the direct breeding value (DBV) of individual $i$, and $A_{S,j}$ the associative breeding value (SBV) of associate $j$. The DBV and SBV represent the heritable components of the direct and associative effects, and both DBV and SBV may respond to selection. Furthermore, Griffing showed that the response to individual selection of intensity $i$ for a trait with phenotypic standard deviation is:

$$R = (i/\sigma)(\sigma^2_A + \sigma^2_{SA})$$

where $\sigma^2_A$ is the additive variance of the direct effects and $\sigma^2_{SA}$ is the additive covariance between direct and associated effects. This covariance may even be negative, as expected when those individuals phenotypically superior to the selected trait are, at the same time, the most aggressive one. This problem may perhaps be avoided in a breeding nucleus by eliminating competition, although it will be present, however, in commercial farms. Therefore, part of the genetic progress achieved in the nucleus cannot be transmitted to the industry.

If the group is defined as the unit of selection:

$$R = (i/\sigma)(\sigma^2_A + 2 \sigma^2_{SA} + \sigma^2_{AA})$$

where $\sigma^2_{SA}$ is the additive variance for the associate effects. Thus, transferring the selection pressure from the individual to the group ensures that the population mean will not decrease. Griffing (1976a and 1976b) also pointed out that the efficiency of the process can be improved with the use of non-random groups, that is, families. In this situation, group selection would be operatively equivalent to family selection. However, the rationale is different: a family selection scheme can be of use, not because the selection criterion (family means) allows the elimination of random environmental differences between families, but because it detects
useful indirect genetic effects shared by all members of a family. In essence, this is the Hamilton mechanism that favors altruistic behaviour through natural selection.

The simplest experimental way of detecting social effects was proposed by Pérez-Tome and Toro (1982) and it relies in a simple comparison of means. First, a set of families of sibs are produced. Then, two treatments are compared. In the first, the individuals will live in groups of sibs and in the second the individuals (coming from the same families) will live in groups of unrelated individual. If the productivity in the first treatment is less (greater) than the productivity in the second treatment, it will indicate that relatives compete more (or less) than unrelated individuals. Experimental evidence supporting Griffling’s theory was obtained using the above experimental design in Drosophila melanogaster (Pérez-Tome and Toro, 1982; Martin et al., 1988; López-Suárez et al., 1993) and Tribolium castaneum (Garcia and Toro, 1992 and 1993). More sophisticated designs and analyses involving the estimation of variance components using mixed-model methodologies have been developed by Muir (2005), Van Vleck and Cassady (2005), Van Vleck and Cassady (2006), Cantet and Cappa (2008) and also implemented in Drosophila (Wolf, 2003), trees (Cappa and Cantet, 2006), mussels (Brichette et al., 2001), poultry (Craig and Muir, 1989), cattle (Van Vleck et al., 2007) and pigs (Arango et al., 2005).

Although group selection has been tested in Tribolium (Wade, 1976 and 1977; Garcia and Toro, 1990; Goodnight and Stevens, 1997), there are few applications of group selection in domesticated species. In the classic experiment of Muir (1985 and 1996), layers were housed in half-sib groups and selected as a group for either group livability or egg mass over 1 year of housing. After five generations of selection, results were dramatic and rapidly achieved, that is, eggs per hen housed increased and mortalities declined in group cages until they were similar to those in single-bird cages. The rapid rate of response is attributed to genetic variability previously unused in breeding programmes and perhaps to the presence of major genes for the trait selected. After the seventh generation of selection, a commercial layer line was included in the comparisons. All lines were housed by line, at random, in 12-bird cages. Mortalities were greatest among the commercial birds, followed by the control, and least among the group of selected birds ages. The results clearly demonstrate that, in relation to the control, group selection has succeeded. Further, other observations also led observers to the overall conclusion that, for physiological measures of stress, the wellbeing of the bird had been improved.

One of the best documented fish selection experiments is that carried out by Moav and Wohlfarth (1976) for growth rate in carp. No response to upward mass selection was achieved during the initial five generation periods. However, a clear response to downward selection was observed in that period, showing the existence of genetic variation for the selected trait in the base population. Subsequently, family selection was practised in the upward line, resulting in considerable progress. These results were interpreted in terms of direct and associated effects on growth rate, the late being ascribed to within-family competition.

The methodology presented above has been recently applied to analyse mortality in a commercial population of layer chickens (Bijma et al., 2007b; Ellen et al., 2008). The data were provided by Hendrix Genetics and consisted of observations on survival days of a single generation of 3800 hens bred from 36 sires and 287 dams, which had been mated at random. Each sire had been mated to approximately eight dams, and each dam contributed on average 13.2 females. At an age of 20 weeks, individuals were allocated randomly to 950 standard commercial battery cages, four individuals per cage. Due to chance, some cages contained full or half sibs, but most cages contained unrelated individuals only. For each individual, survival days were defined as age at death in days. Mean survival time was 454 days, with a standard deviation of 122 days. Inspection of dead hens showed that the vast majority of chickens had died due to being pecked. Data were analysed with the previous model, using restricted maximum likelihood. When using a conventional model without associative effects, the estimated heritability for survival days in the chicken line was 6.7%. However, the total heritable variance estimated from the full model was 20%. Two-thirds of the heritable variation, therefore, is due to social interactions among individuals and is hidden in classical analyses. From these data, classical theory suggests a response of only 7.8 days of survival. Predictions accounting for heritable interactions, however, yield substantially higher responses. Selection among individuals applied to a population composed of groups of unrelated individuals yields an expected response of approximately 11 days of survival. Mild multilevel selection applied to a population composed of groups of full sibs more than double the predicted response. The maximum response that can be obtained equals 23 days, which is nearly threefold greater than that suggested by classical theory.

In summary, animal breeders concerned for interaction among individuals should be aware that there are statistical tools and experimental designs that will permit them to estimate genetic variance for associate breeding values and evaluate its importance. Furthermore, from the estimate of these parameters optimal breeding programmes can be implemented that maximize total genetic merit. This will contribute to sustained improvement of both productivity and animal welfare (Muir and Craig, 1998).

**Molecular biology and Animal Breeding**

Only 30 years after the discovery of the DNA structure, the advent of the new techniques of molecular genetics marks the beginning of the new field of genomics: the scientific discipline of mapping, sequencing and analysing genomic level of DNA information. Taking advantage of polymorphic markers called microsatellites, spread throughout the genome, researchers were able to build genetic maps of domestic species and to search for regions of the genome harbouring genes
affecting the performance for economically important traits. With the new century, a number of genome projects have finished, first the human one and later cattle, chicken, dog and horse (and very soon swine). We are now seeing the large panels of thousand of SNPs made available that will provide new tools for individual, family and breed traceability, to diagnose genetic defects and to select on a whole-genome basis. Here we review some of these achievements. There have been parallel advances in what could be called animal biotechnology (cloning, transgenesis, xenotransplantation, etc.) that will not be dealt with here (for a more complete review, see Flint and Woolliams, 2008).

QTL detection in domestic species

In the '90s the QTL detection experiment started. Initially, two basic designs were used. In the first, we utilize the linkage disequilibrium between markers and QTL generated by crosses. Typically, F1 animals are generated by crossing breeds that are highly divergent for the traits of interest (e.g. European wild boar and domestic Large White or junglefowl and domestic White Leghorn chicken) and the F1 animals are then either intercrossed (F2) or backcrossed (BC) to one of the parental lines. The second design is to mainly utilize the within-family linkage disequilibrium. This design is especially well suited for commercial populations as dairy cattle where large half-sib families are available. Most QTL studies have been conducted with panels of 100 to 300 microsatellite markers covering the genome, corresponding to an average density of markers covering the genome, corresponding to an average distance between markers of about 5 to 20 cM, and this activity has been very successful. In the database (http://www.animalgenome.org/QTLdb/), the number of reported QTLs are 4928 affecting 499 traits (pigs), 2344 affecting 185 traits (cattle), 1290 for 164 traits (chicken) and 84 for 30 traits (sheep).

After detecting a QTL, the next task is to locate the gene responsible (causal mutation). In QTL detection studies, we can locate one QTL in a chromosome as a region of about 5 to 20 cM (probably harbouring 200 to 400 genes), which makes it difficult to identify the underlying gene responsible, as has been recently emphasized by Georges (2007). To refine the position several actions can be taken. The first is to increase the number of F2 or BC individuals, but about 5000 individuals are required to bring down the resolution to about 5 cM. The second is to work with advanced intercross, that is, F3, F4, ..., Fn generations, where the confidence interval for the QTL is reduced by a factor of about 2/n when compared with an F2 population of the same size, where n is the number of generations of intercrossing. Both strategies are expensive and time consuming.

The third is fine mapping through linkage disequilibrium. These fine-mapping studies were mostly based on addition of new sire families, additional markers notably using high-throughput SNP analysis and SNP microarrays, and statistical methods combining linkage analysis (LA) and linkage-disequilibrium (LD) analysis (Meuwissen and Goddard, 2004). LD can be used for fine mapping of QTL because LD decays quickly as the distance between marker and QTL increases. The pattern of LD observed in a population depends on the effective population size along the history of the species. For instance, in Bos taurus cattle Ne was large before domestication (>50 000), declined to 1000 to 2000 after domestication and, in many breeds, declined to ~100 after breed formation. This causes some LD to exist at long distances (>1 cM), but not to increase markedly until very short distances are reached. This is in sharp contrast with the human species, where Ne was only ~3000, but increased enormously in the last 10 000 years. Consequently, humans have similar LD to cattle at short distances but almost no LD at long distances (Goddard and Hayes, 2009). Therefore, in domestic species, there is no need for a panel of SNP as dense as in human (probably 50 to 60 000 SNPs instead of 300 to 500 000). On the other hand, SNPs that are located several cM from the QTL can show an association to the trait, making precise mapping more difficult. This problem can be overcome by using multiple breeds: markers that show a consistent pattern of LD with a QTL across breeds must be close to that QTL (Goddard and Hayes, 2009). Fine-mapping strategies are now being implemented in domestic species with successful results being reported mainly for monogenic traits. For complex traits, as expected, the results suggest the existence of many mutations, each with small effect (see Goddard and Hayes, 2009 for a review).

The fourth approach to detect QTLs is the 'candidate gene approach': to look for known genes that for physiological reasons could be the gene responsible. For some genes, prior knowledge about function suggests that they might contribute to the genetic variation observed for a given trait. It is therefore logical to look for correlations between phenotypic variation and DNA sequence variation in these candidate genes. This has led to some well established and replicated associations as, for example, that between the porcine melanocortin-4 receptor (MC4R) and fatness, growth, and feed intake (Kim et al., 2000). Unfortunately, the candidate gene literature has too many non-reproducible results that made the field a bit problematic and it has been suggested that the standard used in these type of studies should be raised (Georges, 2007). At present, we have tools (microarrays) that allow us to quantify the level of expression of virtually all genes in a large number of livestock species. We can compare the expression levels between two populations that differ for the trait of interest. The genes expressed differentially could be considered as candidate genes responsible for the phenotypic differences between populations. In the strategy called genomics genetics (Jansen and Nap, 2001) we combine both tools, those of genotyping and functional genomics to identify polymorphisms responsible for the variability in the levels of expression. The rationale for this approach is that the expression levels of each gene are phenotypes, but more 'simple' than classic characters like the growth or disease resistance due to its 'proximity' to genotype. Therefore, it should be easier to identify polymorphisms causal in this case, in a second stage, correlate the genes whose expression varies with the QTLs classically obtained. However, genetical genomics is not free from problems and pitfalls (Haley and de Koning, 2006).
The final approach, more in the Darwin's spirit, is to look for signatures or footprints of selection. The rationale behind this strategy is that selection affects nucleotide variability, that is, if we find a gene with a pattern of nucleotide diversity incompatible with a drift model, it is possible that this region contains genes subject to selection and, thus, genes with an economic interest potential. Note that no phenotypic measures are required, as this is an indirect approach in which sequencing is needed only to identify SNPs and genotyping the population studied. The problem is that we can detect the presence of selection, but not knowing with certainty which character is under selection, and what affects the gene in question. These ideas have also been applied to the domestication process (see Ajmone Marsan, in this issue).

Unfortunately, although easy to find QTLs to locate the responsible gene is a formidable task. Georges (2007) describes three successful stories: DGAT1 and ABCG2 that affect milk composition in cattle and IGF2 and MSTN influencing muscle mass in pigs and sheep respectively. The DGAT1 gene, which is located on the BTA14 chromosomal site and encodes Acyl-coenzyme A: diacylglycerol acyltransferase, is known to catalyse the last step in triglyceride synthesis and influences milk yield and composition. The intron 3-3072(G-A) mutation in the imprinted porcine IGF2 gene, described before, leads to increased muscle mass and decreased fat deposition. The Texel MSTN g+6723(G-A) mutation creates an illegitimate target site for microRNAs miR-1 and miR-206, and as result, mutant MSTN miRNAs are translationally repressed, causing decreased levels of the MSTN protein and increased muscle mass.

The difficulties for finding the causal mutations can be illustrated by some examples. In pigs from almost 5000 QTLs reported, less than a dozen of causative mutations have been firmly established. Interestingly, the first QTL reported in livestock was FAT1 QTL located in swine chromosome 4 (Andersson et al., 1994). However, its causal mutation is still unknown. Rothschild et al. (1996) show that variation in the ESR in pigs was associated with litter size but ten year later, it was still controversial as to whether the ESR is the responsible gene (Alfonso, 2005). Similarly, Georges et al. (1995) found a QTL on chromosome 6 with effect on milk production but, however, de Koning in 2006 discussed which of the two possible genes OPN (Osteopontina) or ABCG2 protein is the causal mutation.

**Marker-assisted selection**

One of the main motivations for QTL detection in domestic animals is marker-assisted selection (MAS). The usual way of thinking of MAS is a three-step process. First, detect one or several QTLs. Second, find the gene responsible (causal mutation). Third, increase the frequency of the favourable allele either by selection or by introgression. There are some examples as the halothane gene in pigs or the booroola gene in sheep. This strategy should better be called gene-assisted selection. Another approach is to use markers that are in linkage disequilibrium with QTLs. Several companies are developing and releasing commercial DNA test toolboxes involving a small number of markers (Van Eenennaam et al., 2007) and some initiatives exist to combine pedigree, markers and phenotypic information (for a review, see Dekkers, 2004).

The impact of MAS in livestock breeding programmes has been modest because the QTL that exceed the chosen significance thresholds usually only account for a minor fraction of the trait variance. This can be illustrated in the following quotations:

**Despite the great enthusiasm for breeding companies to be involved there are very few applications of MAS in commercial poultry breeding. They are not convinced about the economic feasibility of MAS**’ (de Koning and Hocking, 2007).

**Although several useful genes (primarily gene-linked markers) have been identified in pigs, their application has been limited and their success inconsistent**’ (Spötter and Distl, 2006).

**The much anticipated benefits of DNA-based tools to routinely guide selection decisions in cattle have not been fully met since the origin of this premise**’ (Sonstegard and van Tassell, 2004).

**Although initial expectations for the use of marker assisted selection were high the current attitude is one of cautious optimism**’ (Dekkers, 2004).

**Unfortunately, as is too often the case, the immediate promise of genomics was clearly oversold, as it has since become clear that the identification of QTL was only a first baby step in the process to bringing these results to a practicable technology** (Green, 2009).

A more radical proposal: bypassing QTL detection by genomic selection

Meuwissen et al. (2001) proposed a different approach towards MAS, called genomic selection. It is rooted in two assumptions that have now been accomplished. The first is that panels with tens of thousands of markers will be available together with cost-effective genotyping procedures, and the second is that marker-density will be sufficient for all responsible genes of a trait to be in linkage disequilibrium with flanking markers. The conseqution of genomic projects in several domestic species has allowed that a large numbers of SNPs were discovered as a by-product of sequencing or in subsequent re-sequencing. Although we are still far from the latest human SNP chips with over 1 000 000 SNPs, commercial ‘SNP chips’ exist for cattle (60 000 SNPs, soon hundred of thousands), dogs (250 000 SNPs), sheep (50 000 SNPs), pigs (60 000 SNPs), horses (55 000 SNPs) and chickens (60 000 SNPs) that can be easily genotyped using the same well established technology that in human and as with a reasonable cost (€100 to €200).

In the simplest terms, genomic selection is a two-step process. First, estimate the effects of markers (>50 000 SNPs) in a reference (training) populations that has been phenotyped
and genotyped. Second, use this information to predict the breeding value of candidates to selection in a testing (evaluation) population that has been only genotyped for the previous markers. The main difference between genomic selection and MAS is that MAS concentrates on few QTLs with well-verified association, whereas genomic selection uses a genome-wide panel of dense markers so that all QTLs are in linkage disequilibrium with at least one marker.

Genomic selection has been met with a lot of enthusiasm and some breeding companies are re-designing their breeding programme. The idea is that with genomic selection we can potentially predict the breeding values for selection candidates at birth with an accuracy of 0.8. Consequently, we can select animals at an early age and it is therefore expected to double the rate of genetic improvement per year. For example, in dairy cattle an optimal breeding design with genomic selection will be more or less as follows: (a) genotype a large number of bull calves from the population; (b) calculate GEBVs (Genomic Extended Breeding Values) for these calves (accuracy = 0.8); (c) select team based on GEBV and sell semen from these bulls as soon as they can produce it. The generation interval will be reduced from ~4 to ~2 years and the rate of genetic gain will be doubled.

There are still several unknowns in the implementation of genomic selection and a lot of work need to be done in the next few years (Goddard and Hayes, 2009). First, design issues such as the gathering of an appropriate reference population of one or several breeds, the number of animals and the number of SNPs to be genotyped and how often will the marker effects have to be re-estimated and new markers discovered. Second, statistical models and approaches should be investigated. Although the Bayesian methods proposed by Meuwissen et al. (2001) have captured a lot of attention other alternatives are being developed as non-parametric methods (Gianola et al., 2006; Gianola and de los Campos, 2008; González-Recio et al., 2008). Third, more species (such as the goat and the rabbit) need to be sequenced and SNP chips made available, although in these species (such as the goat and the rabbit) need to be sequenced and SNP chips made available, although in these species the cost of genotyping may delay its implementation. Fourth, from a practical point of view, the most important limitation is the lack of measured phenotypes for difficult traits, such as resistance to disease and stress, fertility, longevity or nutrient utilization (the so-called phenomic gap). Finally, what ultimately we would like to do is to unravel the complexities of epistatic and genotype by environment interactions and how they affect phenotypic expression.

Animal Breeding fiction: synergistic use of genomic selection and germ-line manipulation

Meuwissen et al. (2001) also emphasized that the rate of genetic improvement can be increased still further by the combined use of genomic selection and embryo technologies. The last will allow very short generation intervals and therefore ability to increase the number of selection cycles per unit of time. Georges and Massey (1991) and Haley and Visscher (1998) took this idea to the extreme in what they call ‘velogenetics’ schemes for cattle, where oocytes were harvested from in utero calves, matured in vitro, fertilized, selected on the basis of their marker genotypes, and implanted in recipient cows (or cultured again), resulting in generation intervals of 6 months or less. The process can be repeated by harvesting oocytes from these second generation animals with the generation interval being reduced to around 3 to 6 months.

In the proposal of ‘whizzogenetics’: instead of being re-implanted the development of embryos continue, in vitro meiosis is induced, and embryos are formed (genomic selection in the laboratory).

References


