CSIRO PUBLISHING

Reproduction, Fertility and Development, 2010, **22**, 138–144 www.publish.csiro.au/journals/rfd

Brief introduction to whole-genome selection in cattle using single nucleotide polymorphisms

G. E. Seidel, Jr

Animal Reproduction and Biotechnology Laboratory, Colorado State University, Fort Collins, CO 80523-1683, USA. Email: gseidel@colostate.edu

Abstract. Genomic selection using single nucleotide polymorphisms (SNPs) is a powerful new tool for genetic selection. In cattle, SNP profiles for individual animals are generated using a small plastic chip that is diagnostic for up to 50 000 SNPs spaced throughout the genome. Phenotypes, usually averaged over offspring of bulls, are matched with SNP profiles of bulls mathematically so that animals can be ranked for siring desirable phenotypes via their SNP profiles. For many traits in dairy cattle, the rate of genetic improvement can be nearly doubled when SNP information is used in addition to current methods of genetic evaluation. Separate SNP analyses need to be developed for different populations (e.g. the system for Holsteins is not useful for Jerseys). In addition, the value of these systems is very dependent on the number of accurate phenotypes matched with SNP profiles; for example, increasing the number of North American Holstein bulls evaluated from 1151 to 3576 quadrupled the additional genetic gain in net merit from this approach. Thus, the available information will be insufficient to exploit this technology fully for most populations. However, once a valid SNP evaluation system is developed, any animal in that population, including embryos, can be evaluated with similar accuracy. Biopsying embryos and screening them via SNP analysis will greatly enhance the value of this technology by minimising generation intervals.

Additional keywords: breeding, embryo, genetics, genomics, SNP.

Brief summary of some genetic principles

Phenotype and genotype

Phenotype, what organisms look like and how they perform, is determined entirely by genetic make-up for some traits (e.g. sex, hair colour) and mostly or entirely by environment for other traits (e.g. death due to lightning, becoming infected with certain viruses); however, for most traits, a combination of genetics and environment is involved, often with an interaction. An example of an interaction is dairy cows selected genetically for high milk production fed optimally or suboptimally. In addition to genetics and environment, epigenetics is also important (Bromfield *et al*. 2008), but will not be covered in this review. Although cattle examples will be used throughout this paper, the broad principles apply to most mammals.

The focus of animal breeding is to manipulate genetics, mostly by selective breeding, to obtain desired phenotypes. The genotype of an animal is fixed at fertilisation, when a haploid spermatozoon, containing approximately 2.8 billion base pairs (bp) DNA in the case of cattle (approximately 4% less for a Y spermatozoon than an X spermatozoon), fertilises a haploid oocyte, which, after extrusion of the second polar body, also contains approximately 2.8 billion bp DNA (Elsik *et al*. 2009). The resulting zygote duplicates this DNA and divides to produce a two-cell embryo, so each blastomere contains approximately 5.6 billion bp DNA. As the cells of the embryo continue to duplicate DNA and divide, the resulting adult animal will have well over 1 trillion somatic cells, each (with a few exceptions) containing the same 5.6 billion bp DNA (11.2 billion if duplicated in preparation for cell division) that were present in the zygote.

Genes and alleles

From classical genetic principles, genes are units or lengths of DNA that contain two kinds of information: (1) specification of the amino acid make-up of proteins via making mRNA; and (2) regulation of when, where and how much of that specific RNA is made. For example, the gene specifying RNA for the amino acid sequence of the milk protein casein has a regulatory part that causes that RNA to be made only in the mammary gland and only when lactation is physiologically appropriate. It turns out that there are approximately 22 000 such genes in cattle, specifying proteins ranging from haemoglobin to FSH (The Bovine Genome Sequence and Analysis Consortium 2009). There also are thousands of genes that produce RNA that is not translated into proteins. Several of these are structural RNAs (e.g. for ribosomes), but most are regulatory RNAs that interact with the regulatory regions of protein-specifying genes so that they are turned on to produce the right amount of RNA at the right time in the right tissues. Thus, skin cells do not make FSH and pituitary cells do not make skin, partly due to regulatory RNAs.

A final elementary concept is that of alleles, which are alternative forms of a gene. Genes inherited from one parent often differ in small but important ways from those inherited via the

 10.1071/RD09220 1031-3613/10/010138

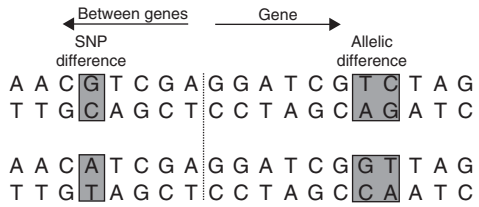

Fig. 1. Illustration of how a single nucleotide polymorphism (SNP) marks an allelic difference between two chromosomes, which could be considered homologous (one from each parent within an individual or chromosomes from two individuals). Note that the bp sequence is identical for the top and bottom chromosomes, except for the SNP marker and allele. This is simplified in various ways (e.g. the directionality of the DNA is not specified).

other parent (Fig. 1). These differences are the basis of genetic variation and are termed alleles. Familiar examples are coat colour, horned or polled, etc., with sex being a special case. For most of the 22 000 genes that specify proteins and their variants due to alternate splicing, these differences are less dramatic. For example, hundreds of genes affect growth; one of these is growth hormone (GH), which comes in different forms due to alleles, primarily because of differences in the regulatory parts of the gene. Thus, some animals produce more GH than others, affecting not only growth, but also traits like milk production.

Genetic selection for phenotypic traits is nothing more than choosing different combinations of alleles. For example, an individual animal will have three possibilities for the allelic composition of a regulatory region of the GH gene: (1) high from father and high from mother; (2) low from father and low from mother; or (3) high from one parent and low from the other. For many genes, there are more than two alleles present in the population of animals but, in an individual animal, only two alleles are possible.

Gene sequences and maps

Sequencing the bovine or human or rice genome is nothing more than determining the linear order of four bases in the genome: adenosine (A), thymidine (T), guanine (G) and cytosine (C). Each A is paired with a T, and each G is paired with a C. In the case of cattle, the DNA is arranged on 29 pairs of linear autosomal chromosomes (one of each pair from each parent) and the sex chromosomes. A special case is the circular mitochondrial genome of roughly 16 000 bp with approximately 35 genes; this is inherited maternally via mitochondria in the oocyte. Because sperm mitochondria degenerate after fertilisation, spermatozoa generally do not contribute mitochondrial genes to embryos.

A huge problem in sequencing 2.8 billion bases of the haploid bovine genome and using that information is knowing where you are, even though the genome is divided up into the 29 autosomes and two sex chromosomes; the process is called mapping and there are several kinds of maps. The same kinds of problems and solutions occur when designing and using maps for transportation. One needs to be able to match up information along the road, such as road signs and mile markers, with the map for it to be of any value. Particular DNA sequences serve as markers in the same way; for example, there are hundreds of ways to order the four DNA bases in sequences of four, such as ATGC, ATCG, AGTC, TACG, AATT, etc. With a sequence of 20 bases, there are billions of possible combinations and, with a few exceptions, such as 20 As in a row, 20 bp define a unique map marker. Large numbers of markers make it possible to know where you are in the genome with reasonable precision.

Another problem is that there really are two maps for each animal, one for the genome inherited from the mother and the other the genome from the father. As noted above, these maps differ in the alleles of genes. Another way they differ even more is in the DNA sequences between the genes, which comprise over 90% of the genome. Any difference at a particular point, termed a locus, whether in a gene or between genes, is called a polymorphism. Often these are a 1 bp change, frequently without the adjacent base pairs being different. These are called single base pair (nucleotide) polymorphisms or SNPs (pronounced ‘snips’).

Map markers

SNPs as map markers

SNPs occur approximately every 700 bp in *Bos taurus* and every 300 bp in *Bos indicus* cattle (The Bovine HapMap Consortium 2009), which means there is more genetic variation in *B. indicus* cattle. Thus, there are approximately 4 million SNPs in the *B. taurus* genome. If time and money were not limiting, one could sequence the genome of every animal and thus know their exact genetic make-up. This would be less useful for selecting animals for breeding purposes than one may surmise, mostly because we simply do not know what most genes do, nor do we know most of the genes that affect a given trait, such as milk production, growth rates, feed efficiency, disease resistance, etc., or how alleles of genes differ in their effects on phenotypes. However, the relationship between alleles and phenotype can be determined in several indirect ways and animal breeders have exploited such relationships for centuries. For example, large animals tend to have large offspring and thus usually have alleles for large size.

These approaches have become quite sophisticated with expected progeny differences (EPDs) in beef cattle and the estimated breeding values (EBVs) of dairy cattle (the term used in North America is ‘predicted transmitting ability’ or PTA). With rare exceptions, these very effective tools for animal breeding do not use any DNA sequence information, only pedigrees and phenotypes. However, applying these techniques does alter the allelic structure of populations. Obvious examples are the different breeds of cattle and changes within breeds, such as size changes in beef breeds and increased milk production in dairy breeds. It turns out that genomic selection using SNPs is just another method of matching unknown alleles with phenotypes, but now DNA sequence information is used in the process.

Direct selection for alleles

Hundreds of specific genes and their alleles that affect phenotypes are known, even though most are unknown. Most of these known genes belong to one of two categories: (1) all or none dominance effects; and (2) additive gene action effects. Familiar examples of the former are coat colour, polled/horned and certain genetic abnormalities, such as mulefoot, bovine leucocyte adhesion deficiency (BLAD) and curly calf syndrome (arthrogryposis multiplex). These dominant effect genes have the same

phenotype if the dominant allele is inherited from both parents or only one parent (e.g. one or two copies of the polled allele give polled offspring; one copy of black and one copy of the recessive red allele or two copies of black give black offspring; two copies of the recessive red allele result in red colour). With genetic diseases, one similarly needs two copies (one from each parent) of the defective (recessive) allele to get the disease. Note that the exact DNA sequence is known for most of these genes/alleles and that information forms the basis of the currently available genetic tests.

The second category for which alleles are known and can be selected for directly is illustrated by meat tenderness genes, such as calpain. In this case, one copy of the desirable allele (heterozygous) is intermediate in tenderness to two copies (homozygous). One submits blood or other body tissues to companies that analyse the DNA and report on the alleles for tenderness. There are hundreds of genes that affect meat tenderness in addition to those that are known and selected for. Although the presence of the tenderness alleles just described does not ensure tender meat, it increases the chances that the meat will be tender. Clearly, environment greatly affects meat tenderness (e.g. how animals are fed and how the meat is aged and cooked). However, some genetically inferior meat will not be tender, even if the environment is optimal.

Concept of marker-assisted selection

As mentioned earlier, many genes and/or alleles affect most production traits, but exactly which genes and/or alleles are responsible is largely unknown. However, there are animals in the population with different phenotypes, such as producing milk with high or low percentage protein. DNA from such cows can be evaluated to determine how it differs. There are a variety of ways of doing this. In most cases, the DNA differences found that correlate to the different phenotypes are not the different alleles themselves (see Fig. 1). These are called markers and, for that animal and its close relatives, for breeding purposes knowing the marker is just as useful as knowing the allele; if the correct marker is selected for, the desirable allele (e.g. top *v.* bottom in Fig. 1) is also selected. The SNPs are examples of such markers and, in sufficient and appropriately spaced numbers, can serve as markers for essentially all alleles of all genes in an animal.

SNP evaluation systems

SNP chips

As mentioned earlier, there is a SNP approximately every 700 bp in the *B. taurus* genome and, because the genome is approximately 2.8 billion bp in length, there are approximately 4 million SNPs. For reasons of economy and practicality, smaller samples of SNPs are used. Specific SNPs can be identified in a variety of ways, but the current, most practical approach is the SNP chip, which is a small piece of plastic or glass with dozens to hundreds of thousands of small dots on it that bind DNA. Each dot corresponds to a specific SNP and, for a given animal, the SNP can be present in zero, one or two copies, corresponding to having been inherited from neither, one or both parents. The most common SNP chip used for cattle is from the company, Illumina (Illumina Inc., San Diego, CA, USA). This SNP chip has approximately 50 000 SNPs and is thus called a 50K SNP chip; only approximately 40 000 of these SNPs are reasonably useful for a variety of reasons (e.g. some SNPs provide redundant or ambiguous information). An attempt was made to scatter these SNPs throughout the 2.8 billion bp genome so, if evenly spaced (and they are not), the 40 000 SNPs would provide a marker at approximately 70 000-bp intervals. Although far from perfect, this number of SNPs turns out to be very useful for selection purposes in some populations of dairy cattle. Subsets of 10 000 SNPs are almost as useful as the 40 000 set (VanRaden *et al*. 2009). Much larger and more expensive SNP chips are used for studying the genetic basis of disease in human populations (Adelson 2008) and much smaller and cheaper SNP chips are being planned for cattle (e.g. choosing the 300 most useful SNPs from the Illumina 50K chip). The current cost to researchers for one 50K SNP chip plus analysis is approximately US$200; smaller chips could cost as little as US$20–50.

Definition of genomic selection

Whole-genome selection (or genomic selection) may be defined as using genotypes defined by a set of SNPs to select for optimal phenotypes. Considerable mathematics are involved in the process and some of the properties of SNPs are illustrated by the 27 possible configurations of three SNPs in Table 1. Each SNP can be in one of three configurations in the diploid genome, designated arbitrarily by using Letters A and B for SNP-1, and CD and EF for SNP-2 and -3, respectively. There are billions of combinations when there are thousands of SNPs as opposed to 27 for three SNPs. For this example, the optimal SNP configuration for percentage milk protein is BB DD EE or BB DD EF. When SNP-3 is in the EE or EF configuration, the more Bs and Ds, the higher the milk protein; when SNP-3 is in the FF configuration, SNP-1 and -2 have no effect on milk protein. Consider a second trait, namely productive herd life. This is not affected by SNP-3, but is negatively correlated with the numbers of Bs from SNP-1 and the number of Ds from SNP-2. Another way of illustrating the same point is the positive correlation with the number of As and Cs in SNP-1 and -2 (i.e. selecting for certain SNP configurations is equivalent to selecting against others).

Things get further complicated when selecting for more than one trait, as is true for conventional animal breeding and often dealt with by using a selection index approach or using conglomerate variables such as ‘net merit’. In Table 1, productive herd life is negatively correlated with percentage milk protein when SNP-3 is in the EE or EF configuration, but not the FF configuration. This implies that the configurations of SNP-1 and -2 drive productive herd life, not the percentage milk protein itself. However, these relationships are unrealistically simplified and meant only to illustrate principles. Although it is unclear whether SNP systems will help unravel relationships such as those illustrated, the potential could be considerable.

Putting SNP chips to use

The 50K SNP chip provides useful markers for most alleles of genes affecting phenotypes of cattle. Especially important is that essentially all phenotypes, from docility to the protein content of milk, can be evaluated (Lee *et al.* 2008). The problem then

Table 1. Illustration of the 27 combinations of three single nucleotide polymorphisms and how they may be associated with phenotypes of percentage milk protein and productive herd life
SNP, single nucleotide polymorphism

SNP-1	SNP-2	SNP-3	% Milk protein	Productive herd life (months)
AA	CC	EE	3.0	+5.0
AB	CC	EE	3.1	+4.5
BB	CC	EE	3.2	+4.0
AA	CD	EE	3.3	+3.5
AB	CD	EE	3.4	+3.0
BB	CD	EE	3.5	+2.5
AA	DD	EE	3.6	+2.0
AB	DD	EE	3.7	+1.5
BB	DD	EE	3.8	+1.0
AA	CC	EF	3.0	+5.0
AB	CC	EF	3.1	+4.5
BB	CC	EF	3.2	+4.0
AA	CD	EF	3.3	+3.5
AB	CD	EF	3.4	+3.0
BB	CD	EF	3.5	+2.5
AA	DD	EF	3.6	+2.0
AB	DD	EF	3.7	+1.5
BB	DD	EF	3.8	+1.0
AA	CC	FF	3.4	+5.0
AB	CC	FF	3.4	+4.5
BB	CC	FF	3.4	+4.0
AA	CD	FF	3.4	+3.5
AB	CD	FF	3.4	+3.0
BB	CD	FF	3.4	+2.5
AA	DD	FF	3.4	+2.0
AB	DD	FF	3.4	+1.5
BB	DD	FF	3.4	+1.0

becomes obtaining accurate phenotypes from thousands of animals from which one also can obtain DNA for SNP analysis. This is difficult to do accurately because phenotypes are greatly influenced by environment and thus can be misleading when matched to the SNP profile for an individual animal. Fortunately, with certain populations of dairy and beef cattle, phenotypic information has been accumulated in the form of sire proofs derived from hundreds to thousands of phenotypes of the respective sire's offspring. Thus, the SNP profile of a bull can be evaluated and correlated with the phenotypic characteristics of his progeny, such as birthweight, weaning weight, milk production, somatic cell count in milk, etc. This results in a reasonably accurate phenotype averaged over many progeny (suitably adjusted for various factors, such as overall herd performance, age, etc.).

The next step is to take the information from thousands of bulls and to determine which SNP profiles correspond to which phenotypes (undesirable phenotypes are just as valuable because, as indicated earlier, one selects for desirable and against undesirable). This process involves using thousands of simultaneous equations; what is remarkable about the process is that although phenotypes are matched with SNPs that match desirable alleles, it does not require knowledge of which alleles of which genes are actually involved (Lee *et al*. 2008)!

Probably the best-characterised system of selection with SNPs is Holstein dairy cattle in North America (VanRaden *et al*. 2009). This information has been provided to the public in the form of enhanced dairy bull proofs. To develop the system, information from over 5000 bulls and a few females, with millions of progeny, was used. The rate of genetic improvement can nearly double using this technology (Hayes *et al*. 2009) because genetically valuable animals can be identified more accurately and at younger ages. A measure of the power of this approach is that rather than progeny testing 1000 bulls per year, the same genetic progress can be made by progeny testing approximately 500 bulls that have been screened from a larger population with a SNP analysis. Another measure is that the additional information a SNP analysis provides to the pedigree analysis is equivalent to having phenotypic information from an additional 10–20 daughters per bull for most traits in dairy cattle (VanRaden *et al*. 2009), and these numbers increase as more bulls are genotyped.

Genomic selection and genetic progress

There are three practical ways of increasing genetic progress: (1) increase the accuracy of selection; (2) increase selection intensity; and (3) decrease the generation interval, which also results in more selection steps per unit time. Whole-genome selection is one of the few tools that can affect all three of these components affecting genetic progress. Accuracy is clearly increased with genomic selection using SNPs. Generation interval can be lowered easily because SNP evaluations of embryos are equally valid as evaluations for young or old animals. Because the technology can be applied broadly at relatively low cost (screening hundreds of embryos or calves), practical opportunities are provided for increasing selection intensity as well. The combination of these advantages, when added to pedigree and phenotypic information on each respective individual, becomes the most powerful, practical approach available for making genetic changes.

Benefits of using embryos

One of the most useful aspects of SNP evaluation systems is that they can be applied to evaluate any animal within the population used to develop a SNP system, even embryos or fetuses. Because genotype is fixed at the time of fertilisation, the information from a SNP analysis of an embryo biopsy is just as valid as that from an animal of any age. It also is equally valid for each sex. Because most Holstein bull progeny tested in North America are produced by embryo transfer, the screening process for selecting breeding bulls can be initiated between embryo recovery and embryo transfer. Because analysis of the SNP chip takes several days, including sending the sample to a laboratory, biopsied embryos would ideally be frozen and selected embryos thawed and transferred. Because only a few cells would be available from a biopsy, as opposed to millions from a blood sample, a DNA amplification step is needed for this approach, a step that is relatively easy and inexpensive (Le Bourhis *et al*. 2009). The value of starting the genetic selection process with the embryo cannot be overemphasised (Sonstegard and iBMC Consortium 2008). Not only does one not waste costs of embryo transfer, recipients and

Table 2. Relative rates of genetic improvement when selecting breeding males as embryos, prepubertally or as adults after a progeny test for traits not measurable before puberty

Note: rates depend on the heritability of the trait, the intensity of selection and whether the trait can be measured in both sexes. When the trait can be measured before puberty (e.g. birthweight), the phenotypic information may alter the relative rates in the table. SNPs, single nucleotide polymorphisms

Method	Stage of selection		
	Embryo	Prior to puberty	After progeny test (adult)
Progeny test systems	++	++	++
Progeny test and genomic selection using SNPs	+++++	++++	+++

post-birth rearing on inferior genotypes, but the generation interval is decreased, increasing the rate of genetic progress. Without biopsy and evaluation of genomic information, there is no way to distinguish differences in genetic value among embryos from a particular mating but, by combining technologies, the embryos that will turn into genetically inferior animals can be culled.

Dairy cattle selection has been emphasised because of the large amount of information available and because the benefits of genome selection using SNPs have been verified empirically for dairy cattle breeding (Hayes *et al*. 2009). Although the same principles apply, it remains to be seen how useful this approach will be for beef cattle. It appears that it will be quite useful but much less than for dairy cattle, at least for traits that can be measured in both sexes, like weaning weights. There may be special advantages of genomic selection for both beef and dairy cattle in breeding for difficult to measure traits, such as disease resistance (Hayes *et al*. 2009).

Table 2 provides an overview of the relative advantages of genomic selection of embryos using SNPs for selecting breeding males. The same relationships generally apply to selecting breeding females, but cows rarely have accurate progeny tests and, for some traits, their phenotypic information is an important part of the selection process. Genomic information will greatly increase the accuracy of genetic evaluations of cows compared with previous procedures. One factor not considered in Table 2 is cost (e.g. the costs of embryo transfer are spread over many more future animals via offspring of selected males compared with females). Because most dairy bulls used for AI in North America are the result of embryo transfer, imposing selection at the embryo stage would not affect costs greatly. Another important point is the additional value of genomic information for progeny tested males. The SNP information will be of considerably greater value in the early stages of accumulating progeny test data than later, when reliabilities exceed 90%. However, achieving high reliability of progeny testing usually takes years of accumulating data.

Some aspects of Table 2 will be considered controversial (e.g. that selecting prepubertally with progeny test systems can be as effective as selecting after a progeny test). Although selection before a progeny test is risky for an individual male, the average of a population of young bulls selected based on their pedigree (plus phenotype for some traits) can result in similar genetic gain relative to waiting for progeny test information. The risk of this approach in an individual herd is mitigated by using five or more bulls rather than fewer; the reason that using young, non-progeny tested bulls is effective is entirely due to taking advantage of shorter generation intervals.

A third line may have been added to Table 2: genomic testing with SNPs without progeny testing. However, it is difficult to envision how such a system would work, because accumulated progeny test information is the basis of using the SNP information, at least as currently performed with cattle. All these systems depend on collecting accurate phenotypic information that can be correlated with genomic information or pedigree information. High reliability progeny test information is a 'gold standard' measure of phenotype.

Limitations of SNP selection

The main limitation of SNP selection is having good phenotypic information from a large number of animals in the population of interest (Hayes *et al*. 2009). For example, accuracy improved markedly when the number of progeny tested American Holstein bulls increased from 1151 to 3576 (VanRaden *et al*. 2009). There are very few populations of cattle that have thousands of accurately progeny tested bulls whose phenotypes can be matched with their SNP profiles. A related major limitation is that a separate system must be set up for each population, at least for best accuracy (e.g. a system developed for Holsteins is essentially useless for evaluating Jersey cattle; see Hayes *et al*. 2009). Ironically, most of the same alleles are desirable in both breeds, but most of the SNPs are different, at least in how they match up to alleles, due to meiotic events over the centuries since Holsteins and Jerseys diverged (see below). Thus, although still useful, the SNP system developed for American Jerseys is much less accurate and useful than that for Holsteins, almost entirely because there are not enough accurately progeny tested sires available. Note that information from dead animals can also be used as long as there is tissue available, such as frozen semen, for producing a SNP profile. Conversely, the more that populations overlap, the more valid is the application of a particular SNP system. For example, the system developed for American Holsteins will

likely be fairly accurate for European Holsteins because there is considerable overlap in their genetic make-up.

A theoretical limitation of SNPs is their bi-allelic nature. For many genes, there are more than two alleles in the population, so other markers, such as microsatellites, could be more informative than SNPs. However, it is unclear whether this is of much practical consequence, because multiple SNPs marking a particular gene may be fairly good at identifying multiple alleles of that gene, especially with the higher-density SNP chips.

Another problem with SNP analyses is that they degrade very slightly with each new generation in the population due to the crossing-over that occurs during meiosis. If a crossing-over event occurs between a marker and the allele it is marking, one not only gets the wrong information, but it is exactly opposite, which equates to selection for the undesirable allele. Such events are rare in a statistical sense and of little consequence in applying this technology over several generations, but eventually will cause problems. Fortunately, this will be a minor practical problem because SNP analysis systems will likely be rederived on a regular basis as information becomes available on new sets of progeny tested bulls. Although less accurate, females also can contribute SNP and phenotypic information to such systems. Because cross-over events differ in the different populations, SNP analysis must be done within populations. The same is true for marker-assisted selection, which is most valid within families and less valid as the family (or population) becomes more extensive. It is possible to use less homogeneous populations (e.g. beef cattle instead of Angus cattle) by using more SNPs, but a 250K SNP chip may be needed for equivalent accuracy to a 50K SNP chip for the Angus breed.

Intellectual property issues

A major limitation of SNP evaluation systems is that they are very expensive to set up, especially in developing and validating SNP chips. Once set up, the cost of making and analysing each additional SNP chip will often be less than US$100. However, millions of dollars are required to set up the system to make the chips and then to set up the analysis system. Although these costs will likely decline dramatically with time, high cost is a current reality. This also means that patents and other intellectual property issues arise, as those who spent the millions of dollars seek to recoup their investments. For example, applications of the Illumina 50K SNP chip using certain dairy cattle information in North America is proprietary and can only be used by the bull studs that paid for development to select bulls for progeny testing through the year 2013. The SNP chips developed in different countries also represent a huge duplication of an expensive procedure. For both logistical and intellectual property reasons, prospects for sharing such genomic information appear remote, at least for the next several years. This is in contrast with the widespread sharing of progeny test information. Almost all SNP technology has intellectual property constraints; in most cases, users pay indirectly by the costs of semen and other fees.

Refinements of SNP selection and future aspects

SNP technology will not replace current methods of selecting breeding cattle, but will be used in addition. We will use all the pedigree information (including the new SNP information on parents) in decision making, as well as the individual's phenotype (e.g. birthweight). The term 'genomic enhanced EPDs' has already been introduced to beef cattle breeders. In addition, specific allelic information will be used, such as sex (embryos), coat colour, freedom from recessive alleles for genetic diseases, desirable alleles (such as for meat tenderness or milk composition), etc.

The SNP marker information will also be used by researchers to locate and identify the alleles being marked; gene chips will eventually discriminate among those alleles, making the whole process of using SNP markers obsolete. This also gets around having to work within populations or families, because problems due to crossing-over between markers and alleles will disappear. This will take years to accomplish for a substantive number of genes and, because genes and/or alleles interact, SNP-type systems, including thousands of simultaneous equations, will likely be used for some time; sophisticated mathematics will be needed to find optimum combinations of alleles.

Implicit in this discussion is that the objective has been selecting the optimal animals for breeding purposes. This is somewhat different from choosing the best animals from a phenotypic standpoint. For example, for breeding purposes, homozygosity of alleles (resulting in breeding 'true') will be aimed for, whereas for production purposes heterozygosity will likely be more desirable for some genes. This gets further complicated in the context of optimal maternal *v.* terminal cross lines.

Intense selection for homozygous desirable alleles directly results in increased inbreeding, but also increases inbreeding indirectly because adjacent lengths of DNA are selected to homogeneity. However, SNP systems can be directed to minimise inbreeding, and they automatically select against allelic combinations that cause inbreeding depression, at least for the trait(s) being selected. Because genomic-enhanced selection procedures can be used either to ameliorate or exacerbate problems such as decreased fertility of high milk-producing dairy cows, thoughtful application is essential. One other obvious approach just alluded to is to have parental lines for breeding purposes that are crossed for production purposes, as is routine for poultry, pigs, sheep and hybrid plants. However, SNPs can be used to mark alleles causing inbreeding depression and then to ensure that those alleles result in a heterozygous configuration for particular matings.

It is only a small step to incorporate transgenic technology into such systems (e.g. polled alleles may be added to genotypes for high-producing dairy cows without compromising the allelic profile for profitable milk production). This particular example may not be considered transgenic, because it would simply be changing a bovine allele to result in the exact end-point that would occur by introgression via repeated back-crossing.

The value of this technology for basic research has only been alluded to briefly. As one example, this technology can be used for pinpointing the meiotic crossing-over events that occur in each chromosome in every gamete produced. The next few decades will be truly exciting as assisted reproduction technologies, such as sexed spermatozoa and embryo transfer, are combined with genetic technologies, such as genomic-enhanced selection using SNPs. Although this paper has emphasised cattle,

SNP technology is already being used on a large scale for plant breeding and will also find use with most domestic animals, including companion animals. There will also be applications for endangered species and, as hinted earlier, related technology will be of great value applied to our own species, particularly for treating diseases and possibly even for retarding aging, especially for conditions such as Parkinson's or Alzheimer's diseases.

Acknowledgements

Several people provided useful comments during the preparation of this manuscript. The references cited are excellent sources for those readers wanting additional information.

References

Adelson, D. L. (2008). Insights and applications from sequencing the bovine genome. *Reprod. Fertil. Dev.* **20**, 54–60. doi:10.1071/RD07157

Bromfield, J., Messare, W., and Albertini, D. F. (2008). Epigenetic regulation during oogenesis. *Reprod. Fertil. Dev.* **20**, 74–80. doi:10.1071/RD07181

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **92**, 433–443. doi:10.3168/JDS.2008-1646

Le Bourhis, D., Amigues, Y., Charreaux, F., Lacaze, S., Tissier, M., Guyader-Joly, C., Mervant, G., Moulin, B., Vignon, X., Gonzalez, C., and Humblot, P. (2009). Embryo genotyping from *in vivo* biopsied bovine embryos after whole genome amplification. *Reprod. Fertil. Dev.* **22**, 192. [Abstract] doi:10.1071/RDV21N1AB187

Lee, S. H., van der Werf, J. H. J., Hayes, B. J., Goddard, M. E., and Visscher, P. M. (2008). Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.* **4**, e1000231. doi:10.1371/JOURNAL.PGEN.1000231

Sonstegard, T. S., and iBMC Consortium (2008). Applied genomics in cattle SNPs to chips: changing the paradigm of selection of cattle. *Acta Sci. Vet.* **36** (Suppl. 2), s287–s298.

The Bovine Genome Sequencing and Analysis Consortium (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528. doi:10.1126/SCIENCE.1169588

The Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532. doi:10.1126/SCIENCE.1167936

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., and Schenkel, F. S. (2009). Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **92**, 16–24. doi:10.3168/JDS.2008-1514