



Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals

Bin Fan^{1,2}, Zhi-Qiang Du¹, Danielle M. Gorbach¹ and Max F. Rothschild^{1,*}

¹ Department of Animal Science and Center for Integrated Animal Genomics,
Iowa State University, Ames, IA 50011, USA

ABSTRACT : In the past decade, there have been many advances in whole-genome sequencing in domestic animals, as well as the development of “next-generation” sequencing technologies and high-throughput genotyping platforms. Consequently, these advances have led to the creation of the high-density SNP array as a state-of-the-art tool for genetics and genomics analyses of domestic animals. The emergence and utilization of SNP arrays will have significant impacts not only on the scale, speed, and expense of SNP genotyping, but also on theoretical and applied studies of quantitative genetics, population genetics and molecular evolution. The most promising applications in agriculture could be genome-wide association studies (GWAS) and genomic selection for the improvement of economically important traits. However, some challenges still face these applications, such as incorporating linkage disequilibrium (LD) information from HapMap projects, data storage, and especially appropriate statistical analyses on the high-dimensional, structured genomics data. More efforts are still needed to make better use of the high-density SNP arrays in both academic studies and industrial applications. (**Key Words :** Whole-genome Sequencing, HapMap, SNP Array, Genomics, Domestic Animals)

INTRODUCTION

The first decade of the 21st century has been a golden time for the advancement of genomics, driven by the completion of the Human Genome Project (HGP). Various methodologies and technologies have been developed during and after the process of building the human genetic blueprint that have been directly transferred into the studies of domestic animal genomics (Andersson, 2009; Goddard and Hayes, 2009; Rothschild et al., 2010). The search for genetic underpinnings of human diseases perplexed researchers for many years. Only recently did the genetic factors underlying various human diseases begin to be revealed, especially with the help of genome-wide association studies (GWAS) using SNP arrays.

Single nucleotide polymorphisms (SNPs) are bi-allelic genetic markers, and they are easy to evaluate and interpret and are widely distributed within genomes. With proper

coverage and density over the whole-genome, SNPs could capture the linkage disequilibrium (LD) information embedded in the genome, which could be used to pinpoint genes underlying human diseases. For domestic animals, these tools can contribute to i) better understanding of species' evolution, domestication and breed formation, and developing new theories of population genetics; ii) dissecting the genetic mechanisms of complex agricultural traits; and iii) improving selection methods for genetic improvement of animal production. High-density SNP arrays were built for important farm animals, first for those with reference genomes and then recently also for others without reference genomes with the advent and application of the massive parallel sequencing technologies. The preparation and utilization of SNP arrays are having considerable impacts on the theory and practice of animal breeding and genetics, which will play important roles in the years to come.

In this review, the whole-genome sequencing and HapMap studies of several important domestic animals are briefly summarized. Then, details about the development of SNP arrays and their applications in various genetics and genomics research projects are also reviewed. Lastly, lessons learned from the reported studies and prospects for

* Corresponding Author: Max F. Rothschild. Tel: +1-515-294-6202, Fax: +1-515-294-2401, E-mail: mfrothsc@iastate.edu

² Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education & College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, P.R. China.

future work are discussed.

ANIMAL GENOME SEQUENCING AND HapMap PROJECTS

The whole-genome sequencing strategies for most domestic animals were taken directly from human genome sequencing, *i.e.*, combining both whole-genome shotgun

(WGS) and BAC-to-BAC sequencing (Green, 2001). Based on their significance in agriculture and as biomedical models, chickens, dogs, cattle, horses and pigs have had their genomes sequenced, as well as some other important animals (Table 1). Due to the rapid development of “next-generation” sequencing technologies and the availability of reference genomes, these strategies have been modified for different species. For those well-studied species in which

Table 1. A summary of the sequenced whole genomes of important domestic animals

| Species (Latin name) | Sequenced animal | Sequencing strategy (Fold coverage) | Genome length (Assembly) | No. of coding genes* | Sequencing organization | Release year | URL | Reference |
|---|---|--|--------------------------|----------------------|--|--------------|---|--|
| Chicken (<i>Gallus gallus</i>) | A female inbred red jungle fowl | Whole-genome shotgun/BAC and other clones (6.6×) | 1.05 Gb (WASHUC2) | 16,450 | Washington University Genome Sequencing Center | 2004 | http://genome.wustl.edu/genomes/view/gallus_gallushttp://www.ensembl.org/Gallus_gallus/Info/Index/ | International Chicken Genome Sequencing Consortium. 2004 |
| Dog (<i>Canis familiaris</i>) | A male poodle | Whole-genome shotgun (1.5×) | ~2.3-2.47 Gb | ~18,473-24,567 | The Institute for Genomic Research/ The Center for Advancement of Genomics | 2003 | http://www.ncbi.nlm.nih.gov/nuccore/36796739?report=genbank | Kirkness et al. 2003 |
| Dog (<i>Canis familiaris</i>) | A female boxer | Whole-genome shotgun/BAC and other clones (7.5×) | 2.38 Gb (CanFam2.0) | 15,900 | Broad Institute/MIT Center for Genome Research | 2005 | http://www.broadinstitute.org/mammals/dog http://www.ensembl.org/Canis_familiaris/Info/Index/ | Lindblad-Toh et al. 2005 |
| Bovine (<i>Bos taurus</i>) | A Hereford cow | Whole-genome shotgun/BAC and other clones (7.1×) | 2.91 Gb (Btau4.0) | 20,684 | Baylor HGSC | 2009 | http://genomes.arc.georgetown.edu/drupal/bovine http://www.hgsc.bcm.tmc.edu/project-species-m-Bovine.hgsc?pageLocation=Bovine http://www.ensembl.org/Bos_taurus/Info/Index/ | The Bovine Genome Sequencing and Analysis Consortium. 2009 |
| Horse (<i>Equus caballus</i>) | A female thoroughbred | Whole-genome shotgun/BAC and other clones (6.8×) | 2.47 Gb (EquCab 2) | 17,254 | Broad Institute/MIT Center for Genome Research | 2009 | http://www.broadinstitute.org/mammals/horse http://www.ensembl.org/Equus_caballus/Info/Index/ | Wade et al. 2009 |
| Pig (<i>Sus scrofa</i>) | Blood samples from five breeds** | Whole-genome shotgun (0.66×) | ~2.1 Gb | - | The Sino-Danish pig genome sequencing project | 2005 | http://www.piggenome.dk/ | Wernersson et al. 2005 |
| Pig (<i>Sus scrofa</i>) | A single Duroc sow | Minimal tile-path BAC by BAC (6×) | 2.26 Gb (Sscrofa9) | 12,678 | Wellcome Trust Sanger Institute | 2009 | http://www.piggenome.org/ http://www.sanger.ac.uk/Projects/S_scrofa/ http://www.ensembl.org/Sus_scrofa/Info/Index/ | - |
| Sheep (<i>Ovis aries</i>) | Blood samples from six breeds*** | Whole-genome shotgun (3×) | 2.78 Gb (OAR1.0) | - | AgResearch/ Baylor HGSC/CSIRO/ University of Otago | 2008 | http://www.sheepmap.org/ http://www.livestockgenomics.csiro.au/sheep/ https://isgdata.agresearch.co.nz/ | - |
| Cat (<i>Felis catus</i>) | A female Abyssinian cat | Whole-genome shotgun (1.87×) | 1.64 Gb (CAT) | 13,271 | Agencourt Bioscience/ Broad Institute | 2006 | http://www.ensembl.org/Felis_catus/Info/Index/ | Pontius et al. 2007 |
| Rabbit (<i>Oryctolagus cuniculus</i>) | A female Thorbecke New Zealand White rabbit | Whole-genome shotgun (7×) | 2.67 Gb (OryCun2) | 14,346 | Broad Institute | 2009 | http://www.ensembl.org/Oryctolagus_cuniculus/Info/Index/ http://www.broadinstitute.org/science/projects/mammals-models/rabbit/rabbit-genome-sequencing-project | - |
| Turkey (<i>Meleagris gallopavo</i>) | A female unknown Turkey | BAC/other large clone shotgun (-) | 1.08 Gb (UMD2) | 11,145 | Virginia Bioinformatics Institute/ USDA Beltsville/ University of Maryland | 2009 | http://www.ensembl.org/Meleagris_gallopavo/Info/Index/ | - |

* Including known protein-coding genes and projected protein-coding genes (<http://www.ensembl.org/info/about/species.html>, up to May 1, 2010).

** ErHuaLian, Duroc, Landrace, Yorkshire and Hampshire.

*** Romney, Texel, Scottish Blackface, Merino, Poll Dorset and Awassi.

high levels of genomics knowledge and sequence coverage are required, such as chicken, dog, cattle, horse and pig, the dual approaches of WGS and BAC sequencing were applied.

These sequences provide comprehensive information for comparative genomics studies on the evolution and function of important genes and genomic regions (International Chicken Genome Sequencing Consortium, 2004; Lindblad-Toh et al., 2005; Dalrymple et al., 2007; The Bovine Genome Sequencing and Analysis Consortium, 2009; Wade et al., 2009; Groenen et al., 2010a). The comparative studies among the genomes of human and domestic animals have also demonstrated a high level of conservation and orthology for protein coding genes. However, huge differences were found in non-coding regions, especially intergenic repetitive regions, which may be one of the major forces driving evolution. The HapMap studies also revealed abundant genetic variability within and between domestic breeds. The majority of the variation was discovered by large-scale genotyping of SNPs and insertions or deletions of DNA fragments with variable sizes, such as copy number variation (CNV), which could in part contribute to the phenotypic diversity of domestic animals.

The additional information derived from linkage mapping, radiation hybridization (RH) mapping, fluorescent *in situ* hybridization (FISH) mapping and expressed sequence tags (ESTs) were used to assist in the genome assembly and annotation. For those with 'light' coverage genomes, such as dog (1.5×) and cat (2×), WGS with "next-generation" sequencing technologies were utilized, and the sequences were assembled using human and other closely-related species as references. More recently, due to the reduced cost of sequencing, both deep-sequencing and individual genome sequencing have been attempted in the cow and chicken (Eck et al., 2009; Rubin et al., 2010), along with the 1,000 Genomes Project in human.

With the completion of whole-genome sequencing of domestic animals, HapMap projects were developed. Since domestic animals have rich sources of phenotypic diversity, which can be interrogated by SNPs across the genome, HapMap studies can be helpful to characterize the complexity of a genome and disentangle the genetic bases of complex traits (International Chicken Polymorphism Map Consortium, 2004; Lindblad-Toh et al., 2005; The Bovine HapMap Consortium, 2009; Wade et al., 2009; Groenen et al., 2010a).

The advantages of the HapMap studies include i) production of a large number of SNPs for design and preparation of high-density SNP arrays; ii) clarification of the genetic relationships among diverse breeds and the phylogenetic relationships between domestic animals and their wild ancestors; iii) prediction of the potentially significant historical events that occurred during

domestication and breed formation, such as bottleneck effects and selective sweeps; and iv) identification of the potentially important candidate genomic regions associated with distinct morphology, disease and other quantitative traits. The implications of HapMap studies will be demonstrated in the later sections of this review.

DESIGN AND DEVELOPMENT OF HIGH-DENSITY SNP ARRAYS

SNP discovery

A very large number of SNPs are essential for the design and construction of arrays. Different methods and resources can be used for SNP discovery, including analysis of predicted SNPs generated from genome sequencing and HapMap studies, completing reduced representation library (RRL) sequencing, downloading SNP information from dbSNP of NCBI or collections of SNPs from individual research institutes or lab groups.

The completion of whole-genome sequencing and HapMap projects uncovered a large number of genetic variants across the genomes of domestic animals, most of which were SNPs. In the chicken, ~2.8 million SNPs were identified (International Chicken Polymorphism Map Consortium, 2004). There were more than 2.5 million potential SNPs in the dog genome, with one SNP per 0.9 kb between breeds and one SNP per 1.5 kb within breeds (Lindblad-Toh et al., 2005). In cattle, ~2.2 million draft SNPs were detected with one SNP per kb (The Bovine HapMap Consortium, 2009). In the horse genome, ~1.1 million draft SNPs were discovered with one SNP per 2 kb (Wade et al., 2009).

Although many SNPs were predicted during genome sequencing projects, SNP prediction software could confuse sequencing errors with true SNPs, meaning further validation is needed. The candidate SNPs for array design should be validated and have high minor allele frequency (MAF) in the testing populations. Matukumalli et al. (2009) found an uneven distribution of SNPs across the genome based on an analysis of cattle draft SNPs. Additionally, it was determined that the nucleotide conversion rate of SNPs was usually low and MAF was not estimated accurately because of the limited sample size of animals used in the HapMap studies. In the commercially released BovineSNP50 array, around three-fifths of SNPs were from the filtered draft SNPs from genome sequencing (Matukumalli et al., 2009). Kerstens et al. (2009) used a similar pipeline to obtain 104,525 SNPs from 1.2 Gb of draft swine genome sequence and verified the polymorphisms of 134 from 163 filtered SNPs in several tested pig populations.

Another effective approach to identify large numbers of candidate SNPs is RRL sequencing, which was first

introduced for creating a human SNP map (Altshuler et al., 2000). This approach could reduce the complexity of the genome by several orders of magnitude, help discover SNPs that are extensively dispersed across the genome, and can even be performed without *a priori* knowledge of the genome sequence. The RRL sequencing procedure is briefly described in Figure 1.

Due to each species' unique genome sequence, the most suitable restriction enzymes for RRL sequencing are variable. In the human genome, *Bgl*III cut sites are commonly distributed (Altshuler et al., 2000). Van Tassell et al. (2008) constructed several RRLs that were generated from DNA of eight commercial dairy and beef breeds that were digested by the *Hae*III restriction enzyme, and identified 62,042 putative SNPs by deep-sequencing and filtering procedures. Around two-fifths of the SNPs on the BovineSNP50 array were from the RRL sequencing approach (Matukumalli et al., 2009). In pigs, Wiedmann et al. (2008) identified 115,572 putative SNPs by sequencing of RRLs that were built from seven predominant commercial pig breeds and were digested by *Hae*III. Amaral et al. (2009) also detected 17,489 pig SNPs using RRLs sequencing of pools of DNA from five Large White×Pietrain crossbred boars digested by the *Dra*I enzyme. Ramos et al. (2009) prepared 19 RRLs derived from four popular commercial pig breeds and a wild boar and three restriction enzymes (*Alu*I, *Hae*III and *Msp*I), and eventually obtained 372,886 high-confidence SNPs. In total, the SNPs obtained from the RRLs sequencing comprised about 94% of the 64,232 SNPs used in the commercially released PorcineSNP60 array (Ramos et al., 2009).

The dbSNP database of NCBI (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) is also a SNP resource for array design. However, the unknown certainty level of each SNP polymorphism and limited genomic distribution of the SNPs might reduce their usefulness and the probability of being selected for a SNP array. On the PorcineSNP60 array, around 5,100 were from dbSNP and other private collections (Ramos et al., 2009).

Illumina's iSelect technology

Illumina's BeadArray based on single-base extension or allele-specific primer extension (<http://www.illumina.com>) and Affymetrix's GeneChip based on molecular inversion probe hybridization (www.affymetrix.com) are the two biggest and most competitive SNP chip genotyping platforms. The approaches of both arrays are different, but they have the capability to perform high-throughput genotyping for large scale samples. In comparison to the GeneChip, the BeadArray is cheaper and more flexible on probe designs (Perkel, 2008). Currently, the majority of the commercially released SNP arrays for domestic animals are constructed using the BeadArray platform with Illumina's

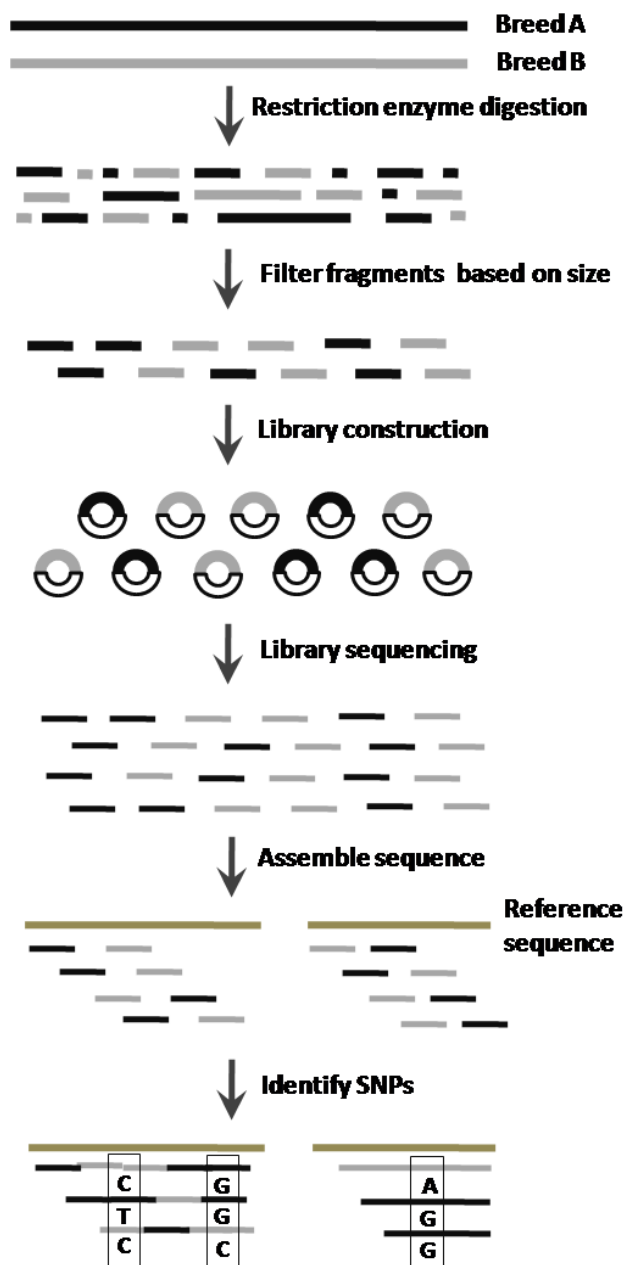


Figure 1. The strategy of reduced representation library (RRL) sequencing for SNP discovery. The details about each procedure can be obtained from Altshuler et al., 2000; Van Tassell et al., 2008; Wiedmann et al., 2008.

iSelect Infinium technology.

A bead chip is a micro-electro-mechanical system, in which wells attaching the beads are created by combining photolithography and plasma etching on silicon wafers. The beads are randomly dispersed and assembled into wells on a silicon wafer (Steemers and Gunderson, 2007). The location of each bead on the array can be identified through a decoding process that uses a 29 base tag sequence linked to the bead. Each bead has a number of 50-mer locus-specific primers following the tag sequence, which are used to

Table 2. Illumina's BeadChips developed for important domestic animals

| Species | BeadChip name | No. of SNPs (Approximate) | No. of mapped SNPs (Assembly) | Average interval between SNPs (kb) | Average MAF across tested populations | Release status |
|---------|-----------------|---------------------------|-------------------------------|------------------------------------|---------------------------------------|------------------------|
| Chicken | Multiple chips* | - | - | - | - | Open with restriction |
| Dog | CanineSNP20 | 22,362 | 22,000 (CanFam2.0) | 125 | 0.27 | Commercially available |
| Dog | CanineHD | 170,000 | 170,000 (CanFam2.0) | 14.3 | 0.23 | Commercially available |
| Cattle | BovineSNP50 | 54,001 | 52,255 (Btau4.0) | 51.5 | 0.25 | Commercially available |
| Cattle | BovineHD | >500,000 | - | - | - | Being developed |
| Cattle | Bovine3K** | 3,000 | 3,000 (Btau4.0) | - | - | Being developed |
| Horse | EquineSNP50 | 54,602 | 54,602 (EquCab2.0) | 43.2 | 0.21 | Commercially available |
| Pig | PorcineSNP60 | 64,232 | 55,446 (Sscrofa9) | 40.7 | 0.27 | Commercially available |
| Sheep | OvineSNP50 | 54,241 | - | 46 | ~0.3 | Commercially available |

* Multiple chips were produced, including a 60K SNP array.

** Selected from SNPs on the BovineSNP50, with the potential use for selecting breeding cattle prior to purchase in the dairy industry.

anneal the genomic sequences flanking the target SNPs (Figure 2). After direct hybridization of the genomic DNA to the SNP array probes, each SNP locus is scanned by an enzymatic-based extension assay using fluorescent labeled nucleotides. The labels are visualized by staining with an immunohistochemistry assay to increase the signal intensity (Steemers and Gunderson, 2007). The two different primer extension assays are allele-specific primer extension (ASPE) and single-base extension (SBE), which are called Infinium I and II assays, respectively. Infinium II can reduce the required number of synthesized beads by nearly half compared with Infinium I, and thus make this bead chip more economical. Therefore, the probes for the majority of SNPs in the chip follow the Infinium II design

(Figure 2).

Criteria for SNP selection

Whole-genome sequencing and HapMap projects provided each species with draft SNPs. High quality control (QC) criteria were then set up to filter these draft SNPs, which required i) each allele of the SNP is included in at least two sequence reads; ii) no repetitive elements surrounding the SNP (within 100 bp); iii) the SNP must be predicted by a minimum of six sequence reads; and iv) predicted SNPs cannot overlap with complex regions (e.g. duplicated sequences) (Matukumalli et al., 2009).

Candidate SNPs following this preliminary step were considered for placement on the SNP array. A set of

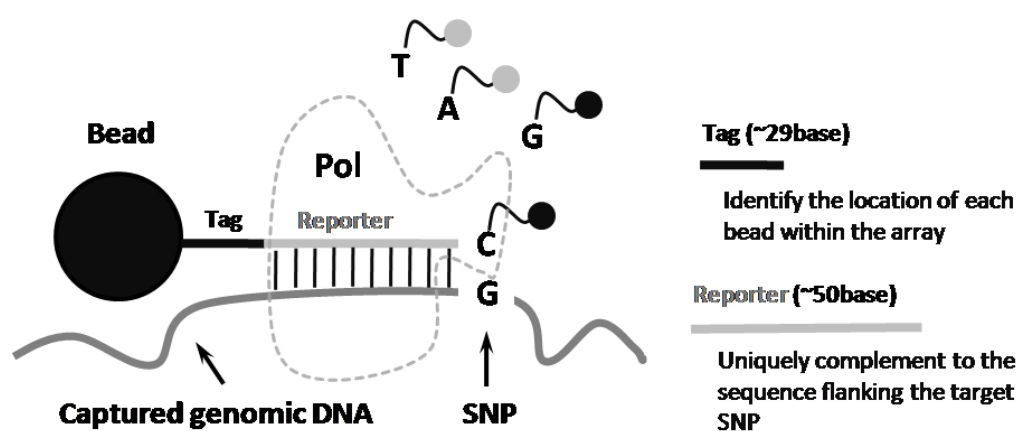


Figure 2. The illustration of Infinium II assay that was developed based on single-base extension (SBE) for SNP genotyping (Steemers and Gunderson, 2007, Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission). In this genotyping system, A and T nucleotides were labeled in one color, and C and G were in another. The polymorphisms A>T and G>C could not be detected.

additional criteria were set up, such as SNP distribution across the genome and each SNP's properties. The physical distribution of SNPs evenly across the genome and reasonable intervals between neighboring SNPs (except Y or W chromosomes in mammals and birds, respectively) were prioritized. As far as the properties of each SNP were concerned, high MAF determined from sequencing of representative samples, high quality score and good validation status were required as well as the design score of Illumina's assay design (Matukumalli et al., 2009; Ramos et al., 2009).

In addition, the bead density, the redundancy of beads per bead type and the final expense (cost-effectiveness) influence the number of SNPs being assembled into the commercially released SNP arrays (Stemers and Gunderson, 2007). For several important farm animals including horse, cattle and pig, the number of SNPs in the first generation chips was slightly more than 50K, ensuring that at least 50K SNPs would work. The first version of the canine SNP array contained 20K, and now a high-density chip with 170K has been released. A high-density 500K cattle SNP array is also being developed (<http://www.illumina.com>).

APPLICATIONS OF THE SNP ARRAY

Genomic selection

Genomic selection (also termed as genomic prediction or genomic evaluation) perhaps is the most fundamental change to breeding and genetics in agriculture as a direct result of the application of SNP arrays. It is different from human studies which have mainly concentrated on searching for disease genes or genealogy. Genomic selection is an advanced form of marker assisted selection (MAS) which concentrates on all markers across the whole genome (Meuwissen et al., 2001; Goddard and Hayes, 2009; Calus, 2010). The MAS strategy has been advocated for the past two decades, but its utilization is limited in practice because of lack of genetic markers linking to QTL with significant effects and the expense of genotyping. Meuwissen et al. (2001) proposed the original concept of genomic selection, *i.e.*, predicting breeding values of animals using information offered by thousands of SNPs across the genome (genomic estimated breeding value, GEBV), by assuming the availability of abundant SNPs scattered throughout the genome and LD relationships between SNPs and QTL. With the new SNP arrays, more SNP effects need to be predicted than there are phenotyped animals for use in predicting these effects. Consequently, Bayesian analysis methods were initially tested to address this problem. Two kinds of Bayesian approaches were developed to predict GEBV using dense SNPs in this pilot study. In both Bayes models, the effect of each SNP was

considered to be independent and random, and the variance of SNP effects were either assumed to be constant or locus specific, and then SNP effects were estimated by a Bayesian procedure with a prior distribution for this variance. These Bayesian methods had higher prediction accuracy compared to those of least squares (LS) and conventional best linear unbiased prediction (BLUP) based on simulation data. In recent years, different statistical approaches for genomic selection have been developed, derived from either non-parametric Bayesian models or parametric methods including genomic best linear unbiased prediction (GBLUP) and mixed regression models (Gianola et al., 2006; Aulchenko et al., 2007; Verbyla et al., 2009; Calus, 2010).

Based on most of the studies using simulated data (Goddard, 2009; Hayes et al., 2009b), several major factors which influence the accuracy of genomic selection were recognized: i) the LD extent between SNPs and the QTL; ii) the size of the training population (the individuals both phenotyped and genotyped for building statistical models and predicting SNP effects); iii) the heritability or genetic basis of the analyzed trait and iv) the distribution of QTL effects. Meuwissen et al. (2001) found that the prediction accuracy of genomic selection could reach 85% when r^2 between the adjacent SNPs was greater than 0.2, which required that SNPs have high-density and even distribution across the genome. For those traits with low heritability and composed of many QTL with small effects, a larger training population is necessary (Goddard and Hayes, 2009; Calus, 2010). Given the heritability of the targeted traits and the prediction accuracy in genomic selection programs, the numbers of animals required in the training population can be estimated (Figure 3).

Additionally, other methods have been proposed to improve the accuracy of genomic selection, such as estimation of missing genotypes, distinguishing the actual SNPs in LD with QTL from those only tracing relationships between animals, and developing novel approaches considering dominance and epistatic effects. Due to consistency of LD across populations, the prediction accuracy will likely remain at a high level whenever the training populations are at least partially related to the validation populations (animals for which GEBVs are being predicted without phenotype data), suggesting that SNP effects obtained from crossbred populations are suitable for genomic selection in pure breeds (Ibáñez-Escriche et al., 2009; Toosi et al., 2010).

Another challenge is to carry out genomic selection in livestock across both national and global regions. Genomic selection has been adopted for genetic evaluation of dairy cattle in the United States of America (VanRaden et al., 2009) and is being considered in the International Cattle Genetic Evaluation Project (<http://www-interbull.slu.se>).

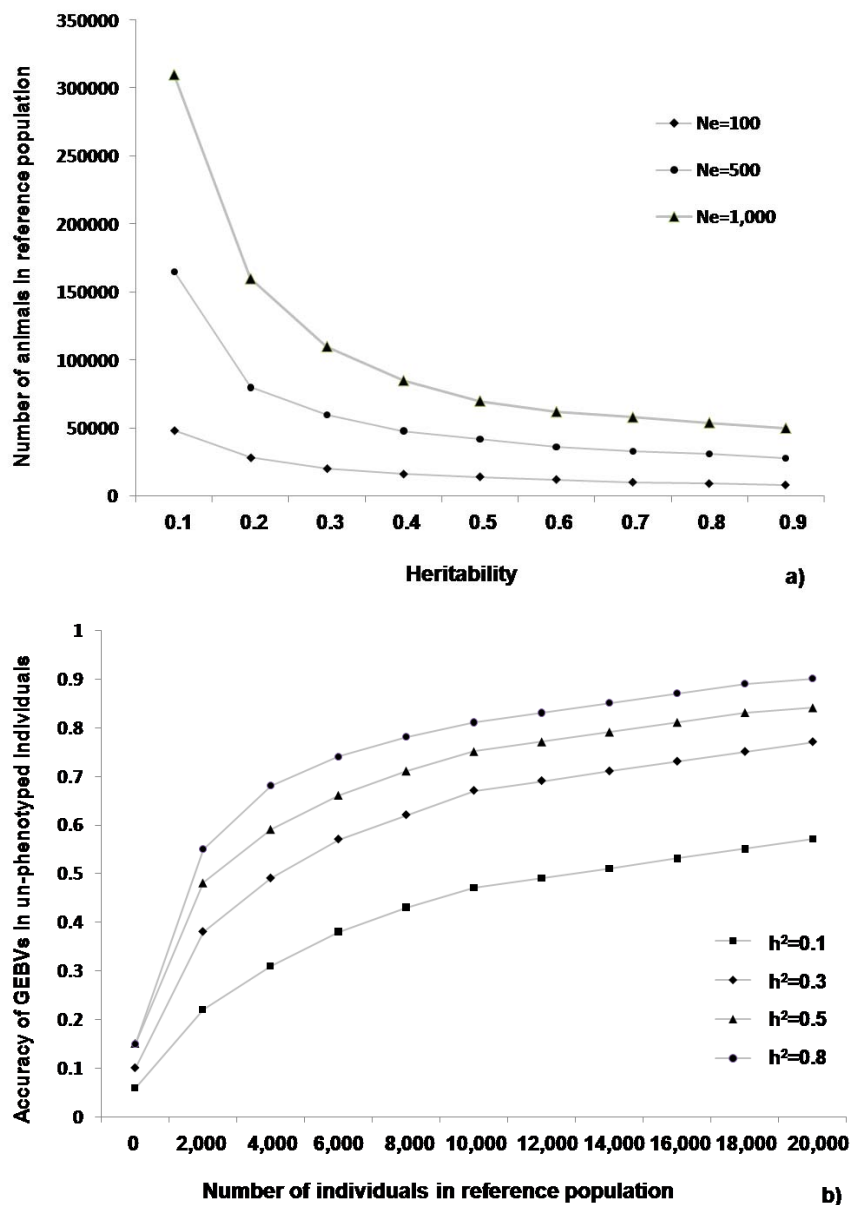


Figure 3. Estimation of number of individuals required in a reference population (Goddard and Hayes, 2009. Copyright Nature Publishing Group. Reproduced with permission). a) Number of individuals required in a reference population to obtain an accuracy of 0.7 for genomic estimated breeding values (GEBVs). The number of animals is negatively correlated to heritability of trait. b) Accuracy of the predicted GEBVs for individuals without genotypes in a validation population, assuming $N_e = 100$. The prediction accuracy is positively correlated to population size of reference individuals.

With the availability of higher density SNP chips which can help find more common haplotypes between breeds, the improvement of advanced statistical approaches and computer programs, and joint sharing of phenotypes and SNP genotypes among research groups, breeding companies will be able to apply genomic selection for livestock across the globe (VanRaden and Sullivan, 2010).

Genome-wide association studies

Both candidate gene and QTL mapping strategies have been extensively utilized in domestic animals for the

discovery of genetic markers suitable for MAS. However, the limitations of these approaches are becoming apparent. The biological mechanisms of quantitative traits and diseases are complicated, and they are still being explored. The determination of candidate genes according to their putative physiological roles is often difficult, and the candidate gene approach may miss the identification of novel genes and pathways associated with some traits. The regions with identified QTL are generally large and further fine mapping is necessary, and often consistency of results from QTL mapping is limited among different resource

families (Rothschild et al., 2007). GWAS (also termed as whole-genome association studies, WGAS) is one of the most promising approaches to overcome these limitations.

Although GWAS have been carried out in domestic animals using the commercially available SNP arrays, most of them were on disease related traits because case-control study strategies could be easily utilized for association analyses (Karlsson et al., 2007; Andersson, 2009; Feugang et al., 2009; Snelling et al., 2009; Wood et al., 2009; Wilbe et al., 2010). For quantitative traits such as growth rate, lean meat percentage, intramuscular fat content and milk production, some researchers tried single marker mixed model or mixed regression models for association analyses (Abasht et al., 2009; Settlles et al., 2009). Other researchers have used the posterior probability that is derived from a Bayesian approach originally designed for genomic selection (Fan et al., 2009; Gorbach et al., 2009; Onteru et al., 2009), where the SNPs having highest posterior probability (*i.e.*, the frequency of a SNP included in the model for GEBV prediction) are most likely to be linked to the QTL (Meuwissen et al., 2001; Fernando and Garrick, 2009; Verbyla et al., 2009). A number of GWAS in several important domestic animals have been completed to date with significant results (Table 3).

Whole-genome LD patterns

Construction of high-resolution LD maps, calculation of the extent of LD at the population level, and characterization of haplotype block structures are crucial for fine mapping and genomic selection (Georges, 2007; Goddard and Hayes, 2009). In most cases, the extent of LD between loci varies between populations, including lines, breeds and even different populations within a breed (Amaral et al., 2008; Bovine HapMap Consortium, 2009; Wade et al., 2009; Megens et al., 2010), and this inconsistency between groups of animals may have a significant impact on fine mapping, genomic selection and GWAS.

The extent of LD in a population also plays an important role in helping a researcher to decide the SNP density needed for a particular study. Differences in population structure and evolutionary forces affect how much LD exists in a population. For populations with longer range LD, there is less value in moving to a higher density SNP array because most QTL may already be in LD with markers on a smaller array. If LD has a relatively short range, then not all QTL may be in LD with markers on a smaller array such that use of a larger SNP array may be worth the extra cost. The extent of LD in a given population can be easily calculated from any SNP array study to predict the best array size to use in future studies.

The findings from the cattle and horse HapMap projects have demonstrated that the decay of LD relationship

between SNPs slows beyond 100 kb, and haplotype blocks become smaller between breeds. It has been suggested that ~100K SNPs may be sufficient for association mapping within and across breeds (Amaral et al., 2008; The Bovine HapMap Consortium, 2009; Wade et al., 2009).

Additionally, effective population size could be derived from the extent of LD within a given interval length, $r^2 = 1/(4Nc+1)$ or $r^2 = 1/(4Nc+2)$ (when mutation is considered in the model). Where N is the effective population size $1/2c$ generations in the past, and c is the recombination rate based on the number of Morgans between the examined markers (Sved, 1971). Although values of N varied between populations, rapid decreases in N were observed in recent generations in the populations examined (de Roos et al., 2008; Kim and Kirkpatrick et al., 2009; Villa-Angulo et al., 2009; Qanbari et al., 2010b), which implied that domestic animals have undergone inbreeding and extensive selection in the past two centuries, both well known occurrences.

In general, the amount of LD between any two markers decreases as the physical distance between those markers increases. Forces such as selection, however, can cause markers that are far apart physically (or even on different chromosomes) to be in high LD with one another. Having high LD for long stretches or between unlinked markers complicates fine mapping. One feasible approach for discovering SNPs that are widely applicable for selection is to carry out LD mapping in multiple breeds, so the SNPs in high LD with QTL across populations can confirm the associations (Goddard and Hayes, 2009).

Population genetics

Selective sweeps : During domestication and breed formation of domestic animals, they have experienced both natural and artificial selection. These selection pressures have led to increased allele frequencies of some mutations in a few specific genomic regions because these mutations made the animals more adaptable or gave them favorable characteristics based on human demands. Over time other polymorphisms may have decreased in frequency or vanished, and a single haplotype containing multiple genes may have become the only one or the most prominent in the population. This has been termed as a selective sweep or positive selection (Andersson and Georges, 2004).

Several statistical methods were proposed for detecting selective sweeps (Sabeti et al., 2007). The integrated haplotype score (iHS) developed from integrated extended haplotype homozygosity (EHH) detects selective sweeps by identifying genomic regions with increased local LD. The fixation coefficient (F_{st}) can be used to predict selective sweeps by comparing the F_{st} values among populations. The composite likelihood ratio test (CLR) is based on the comparison of the maximum composite likelihoods under models with and without selective sweeps.

Table 3. Genome-wide association studies with reported candidate genes in domestic animals

| Species | Population (Size) | Phenotype | Analysis method (Software) | Candidate gene(s) | Replication | Reference |
|--------------------------|---|---|--|--|------------------------------------|--|
| Dog | Boxer (19) | White coat color | Case-control (Plink) | <i>MITF</i> | Yes | Karlesson et al., 2007 |
| | Rhodesian Ridgeback (21) | Ridgeless | Case-control (Plink) | <i>FGF3, FGF4, FGF19</i> | No | Karlesson et al., 2007 |
| | Pembroke Welsh Corgi (55) | Degenerative myelopathy | Case-control (Plink) | <i>SOD1</i> | Yes | Awano et al., 2009 |
| | Chinese Crested dogs (19) | Canine ectodermal dysplasia | Case-control (Plink) | <i>FOXI3</i> | Yes | Drogemuller et al., 2008 |
| | Golden Retriever (48) | Atopic dermatitis | Case-control (Plink) | <i>RAB3C, RAB7A, SORCS2 etc</i> | Yes | Wood et al., 2009 |
| | ~ 80 breeds (>1,000) | Coat phenotypes | Case-control (Plink) | <i>RSPO2, FGF5, KRT71 etc</i> | Fine mapping | Cadieu et al., 2009 |
| | 76 breeds (835) | Chondrodysplasia | Case-control (Plink) | <i>FGF4</i> | Fine mapping | Parker et al., 2009 |
| | Multiple breeds (51) | Brachycephalic head type | Case-control (Plink) | <i>THBS2, SMOC-2</i> | Fine mapping | Bannasch et al., 2009 |
| Unknown population (138) | Systemic lupus erythematosus | Case-control (Plink) | <i>PPP3CA, HOMER2, PTPN3 etc.</i> | Yes | Wilbe et al., 2010 | |
| Cattle | Multiple breeds (~1,500) | Efficient food conversion | Linear regression | <i>PLA2G5, ATP1A1, DAG1 etc</i> | Yes | Barendse et al., 2007 |
| | Belgian Blue (26) | Congenital muscular dystony 1 (CMD1) | Case-control (ASSHOM/ASSIST) | <i>ATPA2A1</i> | Fine mapping | Charlier et al., 2008 |
| | Belgian Blue (31) | Congenital muscular dystony 2 (CMD2) | Case-control (ASSHOM/ASSIST) | <i>SLC6A5</i> | Fine mapping | Charlier et al., 2008 |
| | Italian Chianina (12) | Ichthyosis fetalis | Case-control (ASSHOM/ASSIST) | <i>ABCA12</i> | Fine mapping | Charlier et al., 2008 |
| | Holstein cows (245) | Johne's disease (JD) | Case-control (Plink) | <i>EDN2, SOD1, PARP1 etc</i> | No | Settles et al., 2009 |
| | Black/Red Angus (76) | Black/Red coat color | Case-control (Plink) | <i>MC1R</i> | No | Matukumalli et al., 2009 |
| | Holstein bulls (5,360) | Dairy-related traits | Linear regression/ Bayesian methods | <i>DGAT1, ABCG2, Integrin β2 etc.</i> | No | Cole et al., 2009 |
| | Holstein bulls (20) | Bull fertility | Two stage phases/ Linear regression | <i>ITGB5</i> | Yes | Feugang et al., 2009 |
| | Holstein Friesian cows (798) | Sensitivity of milk production to environment | Bayesian method | <i>FGF4, G3PD etc</i> | Yes | Hayes et al., 2009a |
| | Multiple breeds/crossbreed (~3,000) | Body weight | Linear regression | <i>SPPI, NCAPG etc</i> | No | Snelling et al., 2010 |
| | Holstein cows (481/343) | Bovine spongiform encephalopathy (BSE) | Sib-TDT/Case-control (Plink) | <i>ANKS1B, B3GALT1, EXT1 etc</i> | No | Murdoch et al., 2010 |
| | Multiple breeds/crossbreed (464) | Feed intake and feed efficiency | Linear regression | <i>FND3C3B, RUFY3, RAI etc</i> | No | Sherman et al., 2010 |
| | Unknown population (232) | Johne's disease (JD) | Principal component regression | <i>TUBA3D, CCDC59, TMTC2 etc</i> | No | Pant et al., 2010 |
| Pig | Yorkshire (716) | Average daily gain | Bayesian method (GenSel) | <i>MC4R etc</i> | No | Gorbach et al., 2009 |
| | Large White/ Large White \times Landrace (815) | Back fat thickness | Bayesian method (GenSel) | <i>MC4R, CHCHD3, ATP6VIH etc</i> | No | Fan et al., 2009 |
| | Large White/ Large White \times Landrace (815) | Loin muscle area | Bayesian method (GenSel) | <i>BMP2, IGF2, FST etc</i> | No | Fan et al., 2009 |
| | Large White/ Large White \times Landrace (815) | Leg action | Bayesian method (GenSel) | <i>HOXA, TWIST1, SP4 etc</i> | No | Fan et al., 2009 |
| | Large White/ Large White \times Landrace (683) | Reproduction related traits | Bayesian method (GenSel) | <i>MEF2C, PTX3, ITG6 etc</i> | No | Onteru et al., 2009 |
| | Large White/ Large White \times Landrace (683) | Sow longevity related traits | Bayesian method (GenSel) | <i>ANXA6, ZIC3, ZIC5</i> | No | Onteru et al. (personal communication) |
| | Sheep | Texel (23) | Chondrodysplasia | Identity by descent (IBD) mapping | <i>CADPS2, SLC13A1, NDUFA5 etc</i> | No |
| | Texel (46) | Microphthalmia | Case-control (Plink) | <i>PITX3</i> | Fine mapping | Becker et al., 2010 |
| Horse | Arabian (36) | Lavender Foal Syndrome | Case-control (R) | <i>MYO5A</i> | Fine mapping | Brooks et al., 2010 |

The above methods have been utilized for selective sweep detection in the Bovine HapMap Project (The Bovine HapMap Consortium, 2009). Based on the iHS method, specific haplotype frequencies in the genomic regions containing *MSTN* (relevant to muscle development), *ABCG2* (relevant to milk yield and composition) and

KHDRBS3 (relevant to intra-muscular fatness) might have resulted from selective sweeps. Genomic regions relevant to behavior, immune response and feed efficiency were discovered based on F_{st} estimates (The Bovine HapMap Consortium, 2009). Both iHS and CLR approaches have revealed that one region including *SPOK1* was subject to a selective sweep in beef and dairy cattle. A total of 12 putative selective sweep regions associated with residual feed efficiency, beef yield and intra-muscular fatness were discovered when additional data sets were included (Barendse et al., 2009). In addition, a set of genes including *GHR*, *MC1R*, *FABP3*, *CLPN3*, *SPERT*, *HTR2A5*, *ABCE1*, *BMP4* and *PTGER2* were possibly subject to selective sweeps (Flori et al., 2009; Qanbari et al., 2010a).

In the previously described chicken HapMap Project, most SNPs were thought to arise before domestication. However, Rubin et al. (2010) using massively parallel sequencing identified a possible selective sweep resulting from domestication and specialization of broiler and layer birds and found one putative region including *TSHR* that was associated with metabolic regulation and photoperiod control of reproduction in vertebrates. The *TSHR* selective sweep may represent a significant feature of domestic animals, *i.e.*, the restriction of seasonal reproduction that is now absent from domestic animals. In broilers, the selective sweep regions contained the genes *IGF1*, *PMCH1* and *TBC1D1*, which are related to growth, appetite and metabolic regulation.

In pigs, putative selective sweep regions on *SSC1* and *SSC3* have been observed (Groenen et al., 2010a). The regions containing the genes *IGF2*, *PRLR* and *GHR* also had undergone possible selection (Andersson and Georges, 2004; Iso-Touru et al., 2009).

Genetic diversity and genetic relationship analyses : Population genetics studies of domestic animals focus on genetic variability within breeds and genetic distances between breeds. Their purposes are to unravel the possible historic events during domestication and breed formation and assist in preserving the genetic diversity within endangered indigenous breeds. These studies will be helpful for scientific conservation and preservation measures, and for clarifying the population stratification for genomic selection and GWAS.

Genetic relationships among 19 cattle breeds with different geographical distributions were analyzed using 37,470 SNPs during the Bovine HapMap Project (The Bovine HapMap Consortium, 2009). When the population was divided into two groups ($K = 2$) using Bayesian approaches, the cattle from the taurine and indicine breeds could be distinguished and crossbred populations showed admixture characteristics. Assuming nine groups ($K = 9$), most of the analyzed cattle breeds could be classified into separate groups. Recently the phylogenetic relationships

among 372 animals from 48 cattle breeds were characterized using the BovineSNP50 array. The results were consistent with the biogeography of breeds but also clearly depicted the admixed nature of many populations and revealed pedigree relationships between individuals (Decker et al., 2009).

Kijas et al. (2009) analyzed the genetic relationships among 403 individuals from 23 sheep breeds and 210 individuals from two wild sheep species with 1,536 SNPs. The genetic variability within both African and Asian sheep breeds were lower than those of European breeds, and genetic distances between individuals from African and Asian breeds were smaller than those of European breeds. The genetic relationships among breeds were consistent with the geographical distribution and history of breed formation. Close phylogeographical structure, high genetic similarity and low differentiation were observed in sheep breeds, which was in agreement with the previous findings from other genetic markers.

vonHoldt et al. (2010) detected genetic relationships of 912 dogs from 85 breeds and 225 grey wolves using 48,000 SNPs. Both the neighbor-joining (NJ) clustering tree based on SNP genotypes of individuals and population clustering based on haplotype similarity showed single breeds could be distinguished from one another and grouped into Asian, Middle Eastern and Northern groups, which were consistent with the history of breed formation. In addition, domestic dogs had a higher proportion of multi-locus haplotypes unique to Middle Eastern grey wolves, suggesting that domestic dogs may originate from the Middle East instead of the Far-east as previously hypothesized.

Breed clustering is not always as successful as it was in the previous studies. Wade et al. (2009) analyzed the genetic relationships between 11 horse breeds using 1,007 SNPs and found that the relationships between the studied breeds could not be clarified. This result may be due to the close relationships among domestic horse breeds.

Muir et al. (2008) examined genetic variability of chickens representing commercial, experimental and standard breeds using 2,551 SNPs. Based on the proportion of missing alleles and inbreeding coefficients, commercial broiler and layer line birds were found to have lost a significant amount of genetic diversity (~50% or more) from ancestral breeds. It was suggested that genetic diversity could be recovered within lines by crossing multiple pure lines from chicken breeding companies.

The high-density SNP array has also been useful in understanding the phylogenetic relationships of domestic animals. The dog was determined to be most closely related to the grey wolf, followed by the coyote, the golden jackal and the Ethiopian wolf (Lindblad-Toh et al., 2005). For pecoran (higher ruminant) species, 17 novel relationships

were identified and another 16 previously proposed nodes within the infraorder were confirmed (Decker et al., 2009).

CNV detection

Copy number variation (CNV) refers to a DNA segment that is 1 kb or larger and has variable numbers of copies in comparison with a reference genome. CNVs generally occur in more than 1% of the population, and they have often been found to be associated with specific diseases in humans. The comparison of the fluorescent signal intensity ratios of alleles at each SNP across the genome based on the Illumina BeadChip platform is one approach for CNV identification (<http://www.illumina.com>). Matukumalli et al. (2009) predicted 79 CNVs in diverse cattle breeds using the BovineSNP50 array, and ten of them were verified by comparative genome hybridization (CGH) array genotyping results. Fan et al. (unpublished data) predicted 12 CNV regions in pigs with the PorcineSNP60 array and found two large CNV regions of interest on SSC14.

Other applications

High-density SNP arrays have been used for relationship and paternity testing and tracing the geographic origins of animal products (Fisher et al., 2009; Kijas et al., 2009; Weller et al., 2010). The arrays were also utilized for constructing high-resolution linkage maps, improving physical mapping orders, exploring the potential relationships between genomic sequence features and recombination rates, and carrying out linkage disequilibrium and linkage analysis (LDLA) mapping in particularly interesting regions (Arias et al., 2009; Groenen et al., 2010b).

FUTURE PROSPECTS

Even though SNP arrays have been widely applied in animal breeding and genomics, they still have some limitations with regard to the coverage and annotation of probes on the arrays. The coverage of the arrays for certain species is still low and uneven. Some genomic regions have very few SNPs. Further population information from HapMap studies could potentially solve this issue in combination with new SNP arrays. In addition, on some of the currently released commercial SNP arrays, there are still a number of unmapped SNPs. For example, ~1,800 SNPs on the BovineSNP50 array were unassigned based on Btau4.0, and ~8,000 SNPs on the PorcineSNP60 array were unmapped based on *Sus scrofa* build 9. WGAS have uncovered some associations with unmapped SNPs, and a few of them could be localized by LD estimates with the mapped SNPs. Additionally, the physical locations of some mapped SNPs have been corrected with the production of

new genome assemblies (Ramos et al., 2009). Therefore, it is necessary to continuously improve the genome assembly and assign these SNPs to the correct physical positions.

Another issue is the annotation of the findings from SNP arrays. According to the reported GWAS, most of the trait-associated SNPs (TASs) were located in genes without obvious biological significance on the analyzed phenotypes, or they were located in the intergenic regions or introns of certain genes. Similarly, in GWAS in humans, the TASs were not always in or near putative candidates relevant to the diseases (Manolio et al., 2009). Furthermore, a statistical summary indicated that TASs were intronic (45%), intergenic (43%), exonic and nonsynonymous (9%), exonic and synonymous (2%), or in a 5' or 3' untranslated regulatory regions (2%) (Hindorff et al., 2009). These unexpected results may be due to i) the TASs may be from genes that have not yet been annotated or may be unmapped SNPs demonstrating that further annotations of the current genome assemblies are necessary; ii) the sample size (especially for lowly heritable traits) and the genetic backgrounds of the studied populations may have effects on the association analyses, and multiple pure breeds and larger samples sizes may be of help; iii) the large (~40 Mb) average interval length between SNPs and uneven SNP distributions of the current arrays are major limitations for haplotype block analyses and fine mapping, so higher density SNP panels may be worth developing for improved analyses depending on the extent of LD in the analyzed populations; and iv) the robustness of statistical methods for GWAS could be improved.

As more and more studies using SNP arrays become available, effective storage of the original data and curation of results could be other important issues. With the emergence of the large WGAS and population genetics studies, it will be feasible to build databases related to genome variation and/or candidate genes as public repositories, facilitating the comparisons of data across studies. In humans, several genome variation and GWAS databases have been developed (<http://projects.tcag.ca/variation/>; <http://www.genome.gov/26525384>; https://gwas.lifesciencedb.jp/cgi-bin/gwasdb/gwas_top.cgi). Ogorevc et al. (2009) constructed a gene database on cattle milk production and mastitis traits, but the capabilities of this database are limited. The designed databases should be comprehensive toolkits, interactive with whole-genome sequence, QTL mapping results and as much other related information as possible (Hu Z-L, personal communication). Such comprehensive databases will contribute to better utilization by the community of researchers doing animal breeding and genetics research.

Lastly, statistical analysis of SNP array data still presents a challenge. The large volume of data generated by

SNP arrays is computationally demanding, which requires more sophisticated statistical models and efficient analytical methods. Statistical methods for genomic selection and GWAS are always being developed and improved. The approaches derived from different theories and algorithms will certainly impact the accuracy of the analyses. In addition, most early genomic selection studies were performed with simulated data, which are quite different from real data that often have limited sample sizes, and which may lack detailed pedigree information. Therefore, novel efforts are still needed in quantitative genetics, population genetics, and bioinformatics to develop advanced and efficient statistical approaches that will improve the applications of high-density SNP arrays in animal scientific research and production.

ACKNOWLEDGEMENTS

The authors thank Zhi-Liang Hu for valuable suggestions on manuscript preparation, and thank Drs. Hans Cheng, John McEwan, James Brian and Curt Van Tassell for providing references. The authors appreciate that Drs. Michael E. Goddard and Frank J Steemers allowed to reprint their published figures. Support for this work was provided by the College of Agriculture and Life Sciences, State of Iowa and Hatch and funding from the Ensminger Endowment. Support for Danielle M Gorbach is provided by USDA National Needs Graduate Fellowship Competitive Grant No. 2007-38420-17767 from the National Institute of Food and Agriculture.

REFERENCES

- Abasht, B., E. Sandford, J. Arango, P. Settar, J. E. Fulton et al. 2009. Extent and consistency of linkage disequilibrium and identification of DNA markers for production and egg quality traits in commercial layer chicken populations. *BMC Genomics* 10 (Suppl 2):S2.
- Altshuler, D., V. J. Pollara, C. R. Cowles, W. J. Van Ettern, J. Baldwin et al. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513-516.
- Amaral, A. J., H. J. Megens, H. H. D. Kerstens, H. C. M. Heuven, B. Dibbits et al. 2009. Application of massive parallel sequencing to whole genome SNP discovery in the porcine genome. *BMC Genomics* 10:374.
- Amaral, A. J., H. J. Megens, R. P. M. A. Crooijmans, H. C. M. Heuven and M. A. M. Groenen. 2008. Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179: 569-579.
- Andersson, L. and M. Georges. 2004. Domestic-animal genomics: Deciphering the genetics of complex traits. *Nat. Rev. Genet.* 5: 202-212.
- Andersson, L. 2009. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136:341-349.
- Arias, J. A., M. Keehan, P. Fisher, W. Coppieters and R. Spelman. 2009. A high density linkage map of the bovine genome *BMC Genet.* 10:18.
- Aulchenko, Y. S., D. J. de Koning and C. Haley. 2007. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177:577-585.
- Awano, T., G. S. Johnson, C. M. Wade, M. L. Katz, G. C. Johnson et al. 2009. Genome-wide association analysis reveals a SOD1 mutation in canine degenerative myelopathy that resembles amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. USA* 106:2794-2799.
- Bannasch, D., A. Young, J. Myers, K. Truvé, P. Dickinson et al. 2009. Localization of canine Brachycephaly using an across breed mapping approach. *PLoS ONE* 5:e9632.
- Barendse, W., B. E. Harrison, R. J. Bunch, M. B. Thomas and L. B. Turner. 2009. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* 10:178.
- Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris and M. B. Thomas. 2007. A validated whole-genome association study of efficient food conversion in cattle. *Genetics* 176:1893-1905.
- Becker, D., J. Tetens, A. Brunner, D. Bürstel, M. Ganter et al. 2010. Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. *PLoS ONE* 5:8689.
- Brooks, S. A., N. Gabreski, D. Miller, A. Brisbin, H. E. Brown et al. 2010. Whole-genome SNP association in the horse: Identification of a deletion in Myosin Va responsible for Lavender Foal Syndrome. *PLoS Genet.* 6:e1000909.
- Cadiou, E., M. W. Neff, P. Quignon, K. Walsh, K. Chase et al. 2009. Coat variation in the domestic dog is governed by variants in three genes. *Science* 326:150-153.
- Calus, M. P. L. 2009. Genomic breeding value prediction: methods and procedures. *Animal* 4:157-164.
- Charlier, C., W. Coppieters, H. Rollin, D. Desmecht, J. S. Agerholm et al. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.* 40:449-454.
- Dalrymple, B. P., E. F. Kirkness, M. Nefedov, S. McWilliam, A. Ratnakumar et al. 2007. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biol.* 8:R152.
- de Roos, A. P., B. J. Hayes, R. J. Spelman and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179: 1503-1512.
- Decker, J. E., J. C. Pires., G. C. Conant, S. D. McKay, M. P. Heaton et al. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl. Acad. Sci. USA* 106:18644-18649.
- Drogemuller, C., E. K. Karlsson, M. K. Hytonen, M. Perloski, G. Dolf et al. 2008. A mutation in hairless dogs implicates FOXI3 in ectodermal development. *Science* 321:1462.
- Eck, S. H., A. Benet-Pagès, K. Flisikowski, T. Meitinger, R. Fries and T. Storm. 2009. Whole genome sequencing of a single Bos

- taurus animal for single nucleotide polymorphism discovery. *Genome Biol.* 10:R82.
- Fan, B., S. K. Onteru, D. Garrick, K. J. Stalder and M. F. Rothschild. 2009. A genome-wide association study for pig production and feet and leg structure traits using the PorcineSNP60 BeadChip. Pig Genome III Conference, November 2-4, 2009, Hinxton, Cambridge, UK. Abstract No. 6.
- Fernando, R. L. and D. J. Garrick. 2009. GenSel - User manual for a portfolio of genomic selection related analyses. Animal Breeding and Genetics, Iowa State University, Ames.
- Feugang, J. M., A. Kaya, G. P. Page, L. Chen, T. Mehta et al. 2009. Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics* 10:176.
- Fisher, P. J., B. Malthus, M. C. Walker, G. Corbett and R. J. Spelman. 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. *J. Dairy Sci.* 92:369-374.
- Flori, L., S. Fritz, F. Jaffrézic, M. Boussaha, I. Gu et al. 2009. The genome response to artificial selection: A case study in dairy cattle. *PLoS ONE* 4:e6595.
- Georges, M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annu. Rev. Genomics Hum. Genet.* 8:131-162.
- Gianola, D., R. L. Fernando and A. Stella. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173:1761-1776.
- Goddard, M. E. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245-257.
- Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10:381-391.
- Gorbach, D. M., W. Cai, J. C. M. Dekkers, J. M. Young, D. J. Garrick et al. 2009. Whole-genome analyses for genes associated with residual feed intake and related traits utilizing the PorcineSNP60 BeadChip. Pig Genome III Conference. November 2-4. Hinxton, UK. Abstract No. 11.
- Green, E. D. 2001. Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* 2: 573-583.
- Groenen, M. A. M., A. Amaral, H. J. Megens, G. Larson, A. L. Archibald et al. 2010a. The Porcine HapMap Project: genome-wide assessment of nucleotide diversity, haplotype diversity and footprints of selection in the pig. The International Plant & Animal Genome XVIII Conference. January 9-13, San Diego, California. W609.
- Groenen, M. A. M., P. Wahlberg, M. Foglio, H. H. Cheng, H. J. Megens et al. 2010b. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res.* 19:510-519.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. van Tassell et al. 2009a. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS ONE* 4:e6676.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain and M. E. Goddard. 2009b. Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433-443.
- Hindorf, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins and T. A. Manolio. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106:9362-9367.
- Ibáñez-Escriche, N., R. L. Fernando, A. Toosi and J. C. M. Dekkers. 2009. Genomic selection of purebred for crossbred performance. *Genet. Sel. Evol.* 41:12.
- International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspective on vertebrate evolution. *Nature* 432:695-716.
- International Chicken Polymorphism Map Consortium. 2004. A Genetic variation map for chicken with 2.8 million single-nucleotide polymorphism. *Nature* 432:717-722.
- Iso-Touru, T., J. Kantanen, M. H. Li, Z. Gizejewski and J. Vilkki. 2009. Divergent evolution in the cytoplasmic domains of PRLR and GHR genes in Artiodactyla. *BMC Evol. Biol.* 9:172.
- Karlsson, E. K., I. Baranowska, C. M. Wade, N. S. Hillbertz, M. C. Zody et al. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* 39:1321-1328.
- Kerstens, H. H. D., S. Kollers, A. Kommadath, M. del Rosario, B. Dibbits et al. 2009. Mining for single nucleotide polymorphism in pig genome sequence data. *BMC Genomics* 10:4.
- Kijas, J. W., D. Townley, B. P. Dalrymple, M. P. Heaton, J. F. Maddox et al. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4:e4668.
- Kim, E. S. and B. W. Kirkpatrick. 2009. Linkage disequilibrium in the North American Holstein population. *Anim. Genet.* 40: 279-288.
- Kirkness, E. F., V. Bafna, A. L. Halpern, S. Levy, K. Remington et al. 2003. The dog genome: Survey sequencing and comparative analysis. *Science* 301:1898-1903.
- Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803-819.
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorf et al. 2009. Finding the missing heritability of complex disease. *Nature* 461:747-753.
- Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan et al. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4: e5350.
- Megens, H. J., R. P. M. A. Crooijmans, J. W. M. Bastiaansen, H. H. D. Kerstens, A. Coster et al. 2009. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. *BMC Genet.* 10:86.
- Meuwissen, T. H. E., B. J. Hayes and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Muir, W. M., G. K. S. Wong, Y. Zhang, J. Wang, M. A. M. Groenen et al. 2008. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc. Natl. Acad. Sci. USA* 105:17312-17317.
- Murdoch, B. M., M. L. Clawson, W. W. Laegreid, P. Stothard, M. Settles et al. 2010. A 2cM genome-wide scan of European

- Holstein cattle affected by classical BSE. *BMC Genet.* 11:20.
- Ogorevc, J., A. Razpet and P. Dovc. 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim. Genet.* 40:832-851.
- Onteru, S. K., B. Fan, D. Garrick, K. J. Stalder and M. F. Rothschild. 2009. Whole genome analyses for pig reproductive traits using the PorcineSNP60 BeadChip. Pig Genome III Conference. November 2-4, 2009, Hinxton, Cambridge, UK. Abstract No. 5.
- Pant, S. D., F. S. Schenkel, C. P. Verschoor, Q. You, D. F. Kelton et al.. 2010. A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle. *Genomics* 95:176-182.
- Parker, H. G., B. M. VonHoldt, P. Quignon, E. H. Margulies, S. Shao et al. 2009. An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325:995-998.
- Perkel, J. 2008. SNP genotyping: six technologies that keyed a revolution. *Nat. Methods* 5:447-453.
- Pontius, J. U., J. C. Mullikin, D. R. Smith, Agencourt Sequencing Team, K. Lindblad-Tohet et al. 2007. Initial sequence and comparative analysis of the cat genome. *Genome Res.* 17: 1675-1689.
- Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner et al. 2010a. A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* Published Online doi:10.1111/j.1365-2052.2009.02016.x
- Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner et al. 2010b. The pattern of linkage disequilibrium in German Holstein cattle. *Anim. Genet.* Published Online doi:10.1111/j.1365-2052.2009.02011.x
- Ramos, A. M., R. P. M. A. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald et al. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS ONE* 4: e6524.
- Rothschild, M. F., Z. L. Hu and Z. H. Jiang. 2007. Advances in QTL mapping in pigs. *Int. J. Biol. Sci.* 3:192-197.
- Rothschild, M. F., D. M. Gorbach, B. Fan, S. K. Onteru, Z. Q. Du et al. 2010. Applications of new porcine genomic tools to trait discovery and understanding genomic architecture. The International Plant & Animal Genome XVIII Conference. January 9-13, San Diego, California. W614.
- Rubin, C. J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464:587-591.
- Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913-918.
- Settles, M., R. Zanella, S. D. McKay, R. D. Schnabel, J. F. Taylor et al. 2009. A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *paratuberculosis* infection status in US Holstein cattle. *Anim. Genet.* 40:655-662.
- Sherman, E. L., J. D. Nkrumah and S. S. Moore. 2010. Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. *J. Anim. Sci.* 88:16-22.
- Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, T. McDaneld, T. P. L. Smith, T. S. Sonstegard, R. M. Thallman and G. L. Bennett. 2010. Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.* 88:837-848.
- Stemers, F. J. and K. L. Gunderson. 2007. Whole genome genotyping technologies on the BeadArray™ platform. *Biotechnol. J.* 2:41-49.
- Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* 2:125-141.
- 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.
- The Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324:528-532.
- Toosi, A., R. L. Fernando and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. *J. Anim. Sci.* 88:32-46.
- Van Tassell, C. P., T. P. L. Smith, L. K. Matukumalli, J. F. Taylor, R. D. Schnabel et al. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* 5:247-252.
- VanRaden, P. and P. G. Sullivan. 2010. International genomic evaluation methods for dairy cattle. *Genet. Sel. Evol.* 42:7.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel et al. 2009. Reliability of genomic prediction s for North American Holstein bulls. *J. Dairy Sci.* 92:16-24.
- Verbyla, K. L., B. J. Hayes, P. J. Bowman and M. E. Goddard. 2009. Accuracy of genomic selection using stochastic variable selection in Australian Holstein Friesian dairy cattle. *Genet. Res. Camb.* 91:307-311.
- Villa-Angulo, R., L. K. Matukumalli, C. A. Gill, J. Choi, C. P. Van Tassell and J. J. Grefenstette. 2009. High-resolution haplotype block structure in the cattle genome. *BMC Genet.* 10:19.
- vonHoldt, B. M., J. P. Pollinger, K. E. Lohmueller, E. Han, H. G. Parker et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898-902.
- Wade, C. M., E. Giuotto, S. Sigurdsson, M. Zoli, S. Gnerre et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326:865-867.
- Weller, J. I., G. Glick, Y. Zeron, E. Seroussi and M. Ron. 2010. Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. *Anim. Genet.* Published online doi:10.1111/j.1365-2052.2010.02035.x
- Wernersson, R., M. H. Schierup, F. G. Jørgensen, J. Gorodkin, F. Panitz et al. 2005. Pigs in sequence space: A 0.66× coverage pig genome survey based on shotgun sequencing. *BMC Genomics* 6:70.
- Wiedmann, R. T., T. P. L. Smith and D. J. Nonneman. 2008. SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet.* 9:81.
- Wilbe, M., P. Jokinen, K. Truvé, E. H. Seppala, E. K. Karlsson et al. 2010. Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nat. Genet.* 42: 250-254.

- Wood, S. H., X. Y. Ke, T. Nuttall, N. McEwan, W. E. Ollier et al. 2009. Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. *Immunogenetics* 61:765-772.
- Zhao, X., H. T. Blair, S. Onteru, S. A. Piripi, M. F. Rothschild et al. 2010. A genome wide association study and fine mapping for Chondrodysplasia of Texel sheep. The International Plant & Animal Genome XVIII Conference. January 9-13, San Diego, California. P579.