HOW THE NEXT GENERATION OF GENETIC TECHNOLOGIES WILL IMPACT BEEF CATTLE SELECTION

Megan M. Rolf, Stephanie D. McKay, Matthew C. McClure, Jared E. Decker, Tasia M. Taxis, Richard H. Chapple, Dan A. Vasco, Sarah J. Gregg, Jae Woo Kim, Robert D. Schnabel and Jeremy F. Taylor

Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA

Abstract

Recent advancements in sequencing and genotyping technologies have enabled a rapid evolution in methods for beef cattle selection. The past three decades has seen the advancement from restriction fragment length polymorphism (RFLP) markers that were low-throughput, time-consuming and difficult to score to the newest high-density single nucleotide polymorphism (SNP) assays where marker genotypes are easily and inexpensively generated. The cattle genome sequence was published in 2009 and sequencing technologies have now advanced to the point that a complete genome can be resequenced to a relatively deep coverage for ~$30,000 on several different next generation sequencing platforms. While a reference genome to align the reads is currently required for this process, with read lengths increasing with each software or chemistry update, *de novo* sequence assemblies will become routine in the very near future. Once the analytical methodologies are developed and become widely available, animal scientists will begin to use them to develop cost-effective diagnostics for use in beef cattle production systems. As a result of this rapid expansion of technology, new tools will become available for beef producers to implement in the endeavor to efficiently produce high quality beef for today's consumer. Tools such as high-density genotyping assays and next generation sequencing instruments will help to shorten the generation interval, aid in the identification of causal mutations, increase the accuracy of EPDs on young sires and dams, provide information on gene expression and enhance our understanding of epigenetic and gut microbiome effects on cattle phenotypes.

Introduction

There will be many changes in methodologies for the genetic evaluation of beef animals in the near future due to rapid technological advances. These advances provide the momentum for change in the industry and will enhance our ability to produce beef efficiently in today's marketplace. The ability for beef producers to accurately select for genetically superior animals began over four decades ago when mixed model methods were first published by Henderson (1975). The first national cattle evaluation (NCE) was performed in 1974 (Willham, 1993), and since then, models have evolved from single-trait sire models to the multi-trait animal models used today. The next large step looming on the horizon will be the genomic revolution.

Current technologies are beginning to shape the next generation of genetic evaluation. One of the most useful advances has been the public availability of a published genome sequence for beef cattle. Baylor College of Medicine was the first to sequence the bovine genome and used a combination of bacterial artificial clone (BAC) methods as well as whole genome shotgun sequencing (The Bovine Genome Sequencing and Analysis Consortium, 2009). The University of Maryland quickly released their genome assembly based upon the sequences produced by the Baylor College of Medicine and this assembly has also been annotated (Zimin et al., 2009). These two assemblies differ slightly according to the methods used to assemble the sequence reads and the availability of both provides an invaluable resource for genomic studies in beef cattle.

Microarrays which are used to study the expression profiles of genes within specific tissues are available in two different forms: long oligonucleotide arrays (typically those spotted on glass slides that can be bought privately from researchers) and short oligonucleotide arrays (typically those commercialized by companies such as Affymetrix). Microarray technologies can allow the simultaneous profiling of very large numbers of genes and can be used to identify the pathways that are up- or down-regulated in different tissue types or disease states. This allows the identification of the key genes that regulate the behavior of entire pathways and possibly even phenotypes. These genes then become the targets of pharmocological intervention (drug targets) or even genetic manipulation. The greatest disadvantage of microarrays is that they can only query the genes for which probes are designed onto the array. Thus, we have to know the full complement of "genes" within a species genome to be able to design a comprehensive microarray, and unfortunately this is not the case even for humans, which have the most extensively studied genome. Microarrays also suffer from a loss of information in that, often, only probes are generated for one region of a gene and the gene may actually produce more than one type of transcript or protein. Finally, microarrays require quite a lot of technical skill and large numbers of replicates, normalization and dye swaps must be used to filter the true signal from the biological and technical noise.

The first high-density and high-throughput genotyping assay was the 10K single nucleotide polymorphism (SNP) chip commercialized by Affymetrix (The Bovine HapMap Consortium, 2009). However, the density of SNPs in this panel was insufficient for many genomic studies (including genomic selection (GS) and genome-wide association analyses (GWAS)) which led to the need for a higher density chip. The Illumina BovineSNP50 chip was developed by a consortium of animal scientists using SNP discovery populations in Holstein, Angus and mixed breeds of beef cattle (Van Tassell et al., 2008) and provided much higher density (~50,000 SNPs per animal) than previous high-throughput genotyping assays (Matukumalli et al., 2009). This assay has become the international standard for GS and GWAS in cattle and has even been applied to other species to resolve the evolutionary relationship among the horned ruminants (Decker et al., 2009; MacEachern et al., 2009), testing the number of SNPs needed to form a genomic relationship matrix (Rolf et al., 2010) and investigating the amount of introgression of cattle DNA into bison populations (Schnabel et al., unpublished data). While the Illumina BovineSNP50 assay has proven to be extremely useful for many different types of genomic studies, our current data suggest that even higher density assays will be needed to build models for GS with utility across breeds.

New Technologies

High Density SNP Genotyping Chips from Affymetrix and Illumina

Two new high-density SNP genotyping chips will be introduced in 2010. The first is an assay from Illumina that will utilize the same bead technology and single base extension chemistry that is used for the current BovineSNP50 50K chip. The Illumina assay will genotype approximately 800K SNPs per animal and should be available by the time of this BIF meeting. The second high density SNP chip will be marketed by Affymetrix and will also genotype approximately 800K SNPs. This chip uses a different chemistry to the Illumina chip, but the ligation-based assay should result in almost the same call rate (% of genotypes called per sample) and the produced genotypes should be very high quality (low intrinsic error rate). Best of all, the two companies will compete for business and the cost of these assays may end up as low as we are currently paying for the Illumina 50K assay! With 50K SNPs available per animal, why do we need 800K? There are several important applications and advancements that will be made possible with the addition of more SNPs. The first is that SNPs will be distributed much more closely together in the genome. With 50K SNPs in a genome of approximately 3 billion base pairs, we would expect 1 SNP about every 60 Kb, but with 800K SNPs, we would expect 1 SNP approximately every 3.8 Kb. This inter-marker distance provides much finer resolution for mapping the causal mutations that underlie variation within a trait and also allows a much greater opportunity for identifying SNPs that can predict genotype at these causal mutations when scored in animals of different or even mixed breed content for use in GS.

SNP discovery for the Illumina BovineSNP50 assay was performed using pools of DNA samples from Angus, Holsteins and a group of bulls sampled from the next most important US beef breeds. As a result, there is a bias inherent in the assay towards SNPs that have high minor allele frequency in Angus and Holsteins and the assay performs slightly better for GWAS and GS in these breeds. However, the SNP discovery for the design of the Illumina and Affymetrix 800K panels was performed by sequencing a large number of animals from many different breeds (including both *Bos taurus* and *Bos indicus*) to minimize the ascertainment bias in SNP informativeness across breeds. The end result should be a panel of SNPs that will have high average allele frequencies in almost all cattle breeds but will also contain many loci with low allele frequencies. This is especially important for performing GWAS, since common SNPs cannot be strongly associated with rare or low frequency variants within a population. The larger, more variable panel will contain some SNPs which are at low frequency in the population of interest to facilitate the detection of rare variants within that population.

Perhaps the greatest immediate value of the 800K chips will be the potential for implementing across-breed GS in the beef industry. The real advantage of GS is its ability to simultaneously select for desirable combinations at all loci responsible for genetic variation in a trait using panels of closely linked markers. Figure 1 provides a representation of the difference between traditional marker assisted selection (MAS) – the "single marker" tests that have been used in the industry to this point - and GS. MAS typically involves selecting for desirable genotypes at a small number of loci, which are usually of large effect, as these loci are usually the easiest to identify in association or linkage analyses. In contrast, GS allows the simultaneous selection for desirable genotypes genome-wide. The 50K SNP chip has been shown to be effective for GS within breeds of cattle such as for Net Merit in Holsteins (VanRaden et al., 2009). However, the computation of molecular breeding values with high accuracies requires that a large numbers of animals with high accuracy EPDs be genotyped and the lack of a centralized DNA repository (such as are utilized by the dairy breeds) has limited the numbers of animals available for genotyping within each of the beef breeds. Because of the shortage of DNA samples on animals with high accuracy EPDs, individuals from different breeds will need to be genotyped and the analysis performed across breeds. The assumption here is that the causal variants that create variation in traits are the same set of loci across breeds but they differ in frequency which leads to breed differences in the mean of these traits across breeds. However, SNP allele frequencies also differ across breeds and these differences in marker and causal variant frequencies mean that different SNPs are going to be more or less strongly associated with trait variation in different breeds. The density of SNPs on the 50K assay is not sufficient in an across breed analysis to arrive at a model for the prediction of molecular breeding values that will be highly accurate across breeds and the 800K chips will be vitally important for this application. With a SNP every 3.8 Kb, there will be a sufficient SNP density to surmount these issues and obtain accurate molecular estimates of genetic merit across breeds by identifying the markers that are very close to the causal mutations and that have the same SNP allele on the chromosomes harboring the trait enhancing allele at the causal mutation across all breeds. For example, in the Carcass Merit Project (CMP) 50K data for Warner-Bratzler shear force (WBSF) the correlations between SNP effects (Table 1) and molecular estimates of breeding value estimated from SNP effects produced within each of the breeds (Table 2) were low. However, in the region from 44,000,728-44,208,978 nucleotides on chromosome 29 which harbors the μ-calpain gene, we scored 37 additional SNPs to the 6 SNPs present on the BovineSNP50 assay to produce a mean marker spacing of 4.8 Kb. In the analysis of these data, we found one SNP was consistently associated with WBSF across breeds and that the same allele was predictive of increased tenderness across all 5 analyzed breeds.

Next Generation Sequencing

The ability to quickly, accurately, and inexpensively sequence the genomes of individual animals has the potential to revolutionize selection in beef cattle. Recent technological advancements have made great improvements in the affordability and accessibility of genomic sequence data. Two currently marketed platforms are the Illumina Genome Analyzer (or HiSeq 2000) and ABI SOLiD. Initially, read lengths for the Illumina and SOLiD were in the range 35-36 base pairs (bp) with a cost per million bases of sequence of approximately $2 (Shendure and Ji, 2008). However these platforms have been rapidly developed with improvements in chemistry and software allowing the Genome Analyzer to achieve reads of 125 bp and both technologies currently support paired-end reads in which each end of a 300 bp fragments are sequenced to a depth of 85 bp. More importantly, these instruments are now capable of producing up to 95 Gb of sequence in a single run of the instrument. After quality control processing of the data and mapping fragments to a genome assembly, this results in as much as a 15X coverage of animal genome. Two such runs at a cost of less than $30,000 will produce sufficient sequence data to allow a *de novo* assembly of an animal's genome sequence.

The data obtained from next generation sequencing has many applications. One application is the identification of the actual expression level of all of the genes that are expressed in essentially any tissue or animal. RNA sequencing (RNA-Seq) allows the novel assembly of a transcriptome (the set of expressed genes) for any tissue and provides quantitative data to identify differences in gene expression between two samples. The approach also identifies if alternative exons of a gene are used to create different forms of a protein in different tissues or animals and also produces the DNA sequence of each transcript. Thus any sequence differences (SNPs within coding regions) that result in amino acid changes could produce phenotypic variation within a trait. RNA-Seq produces estimates of the actual number of transcripts of a particular mRNA from the counts of the number of reads that map to each gene.

The top panel in Figure 2 shows the number of sequence reads observed for the *PRP* gene's messenger RNA in the brain of a dog. The figure shows that there is a large number of sequence reads observed for *PRP* (location indicated by the box) and another mRNA shown to the right of *PRP*.

Both RNA-Seq and genomic DNA sequencing data provide insight into novel and causal polymorphisms within an individual. The bottom panel in Figure 2 shows a C/G SNP polymorphism identified in the *PRP* gene from the RNA-Seq data, which results in the change of the amino acid at this position from aspartic acid (D) to glutamic acid (E). The discovery of mutations which actually cause variation within traits will become increasingly important and their knowledge will allow testing across breeds which will drastically reduce the number of loci that need to be tested to explain variation within a trait. If we know the causal mutations, we only have to test for those mutations, rather than using 800K SNPs to estimate the effects of these variants. This will result in the development of more affordable, accurate panels of SNPs that work across breeds. It will also suggest the genes that should be screened across populations in the endeavor to understand all existing naturally occurring variation which may have important phenotypic effects. Information will also be gained that will suggest drug targets or targets for genetic modification, if this technology is deemed acceptable for use in animals by consumers.

A novel application of this technology that is becoming increasingly important in the human and mouse communities is the sequencing of gut microbiomes. Most work in humans and mice to this point has focused on profiling the *16S* ribosomal RNA (rRNA) gene to identify the microbes present in the gut using long-read and low throughput (traditional Sanger) sequencing methods. However, this type of research has recently expanded to utilize next generation sequencing technologies (Qin et al., 2010). The study of "gut microbiomes" and their interactions with the genotype of the host is important because previous studies have shown that there is substantial genetic diversity in the species present within the gut microbiome (Li et al., 2009; Turnbaugh et al., 2006, 2009) and that the gut microbes have a significant impact on energy harvest and obesity in humans and mice (Backhed et al., 2004). One study observed that when germ-free mice were inoculated with a gut microbiome from either a lean or an obese individual, the mice that received the gut flora from the obese mice gained more weight than their counterparts which received the gut flora from the lean mice (Turnbaugh et al., 2006). Because of the way that nutrients (especially those from forages) are harvested in ruminants, it is likely that gut microbiomes have an even greater impact on energy metabolism than in human or mouse. Furthermore, the composition of these gut populations may also be related to feed efficiency, methane/greenhouse gas emissions and manure production; thus, it is imperative that we explore whether the host genotype has an effect on the composition of the gut flora, and if so, select for favorable gut populations. Currently, these relationships are poorly understood, and the host interactions that may be under nuclear genetic control are either confounded with additive genetic effects (which would be desirable) or are being placed in the residual component of our genetic models, where they are not selectable.

Epigenetics

Epigenetics is a field of rising importance in genetics and genomics. Epigenetics involves DNA and histone modifications which can influence gene expression and thus the genetic variation in a trait, even if animals have identical genotypes and DNA sequence. Some examples of these modifications include imprinting, X-inactivation, gene silencing and embryonic reprogramming (Sellner et al., 2007). Epigenetic effects such as methylation involve the addition of methyl groups to cytosines and if these occur in the promoters of genes, transcription machinery can be blocked from binding to the DNA. DNA methylation is influenced by both the genetics and environment of the individual, but has been shown to be stably transmitted from parents to offspring for several generations. Once the bovine epigenome (the set oft nucleotides that are methylated in the DNA that is present across different tissues) has been characterized, there is the potential to select or perhaps even induce favorable effects and include this information into breeding programs. Most of the new high-throughput sequencing instruments can elucidate whether nucleotides are methylated (however the new sequencer from Pacific Biosystems can detect methylation as a by-product of sequencing by measuring the time it takes to incorporate a new base while reading genomic sequence), which will allow rapid advances into the understanding of these effects and their influence on phenotypes in beef cattle.

Finally, technologies such as ChIP-Seq (which allows determination of which proteins, such as transcription factors, interact with DNA to influence gene expression and also allows examination of epigenetic chromatin modifications) and Bis-Seq (massively parallel sequencing of bisulfate-treated DNA, which converts unmethylated cytosines to uracils) are powerful new tools for providing insight into the nature and extent of epigenetic modifications within the genome.

Conclusions

The fantastic pace at which new technologies are being developed to study the genome make it an exciting time in the beef industry for producers and scientists alike. High-density genotyping assays will soon revolutionize the way we conduct genetic prediction and whole-genome sequencing of animals and their gut populations along with epigenetic profiling will lead to new tools to ethically and efficiently provide high quality beef that meets consumer demands in an increasingly competitive marketplace.

Table 1: Correlation coefficients between SNP effects estimated for Warner-Bratzler shear force (WBSF) between five different breeds of animals involved in the NCBA sponsored Carcass Merit Project. The number of animals used in the analysis are shown on the diagonal.

WBSF SNP Effects	ANGUS	CHAROLAIS	HEREFORD	LIMOUSIN	SIMMENTAL
ANGUS	651	0.0267	0.0351	0.0134	0.0260
CHAROLAIS		695	0.0135	0.0019	0.0081
HEREFORD			1095	-0.0196	0.0251
LIMOUSIN				283	-0.0047
SIMMENTAL					516

Table 2: Correlation coefficients between molecular estimates of breeding value (MBVs) estimated from SNP allele substitution effects for Warner-Bratzler shear force (WBSF) in five breeds of animals involved in the NCBA sponsored Carcass Merit Project. Elements in each row represent correlations between MBVs computed using the SNP effects for the breed in that row with MBVs computed for the breed in that row using SNP allele substitution effects for the breed in each column.

WBSF	Angus SNP effects	Charolais SNP effects	Hereford SNP effects	Limousin SNP effects	Simmental SNP effects
Angus MBVs	1.0000	0.2229	0.2500	-0.0625	0.0661
Charolais MBVs	0.0442	1.0000	0.0407	0.0035	0.0715
Hereford MBVs	0.2997	0.1100	1.0000	-0.3259	-0.0068
Limousin MBVs	-0.0220	-0.0391	-0.1794	1.0000	-0.1875
Simmental MBVs	0.1502	0.1624	0.1160	-0.0255	1.0000

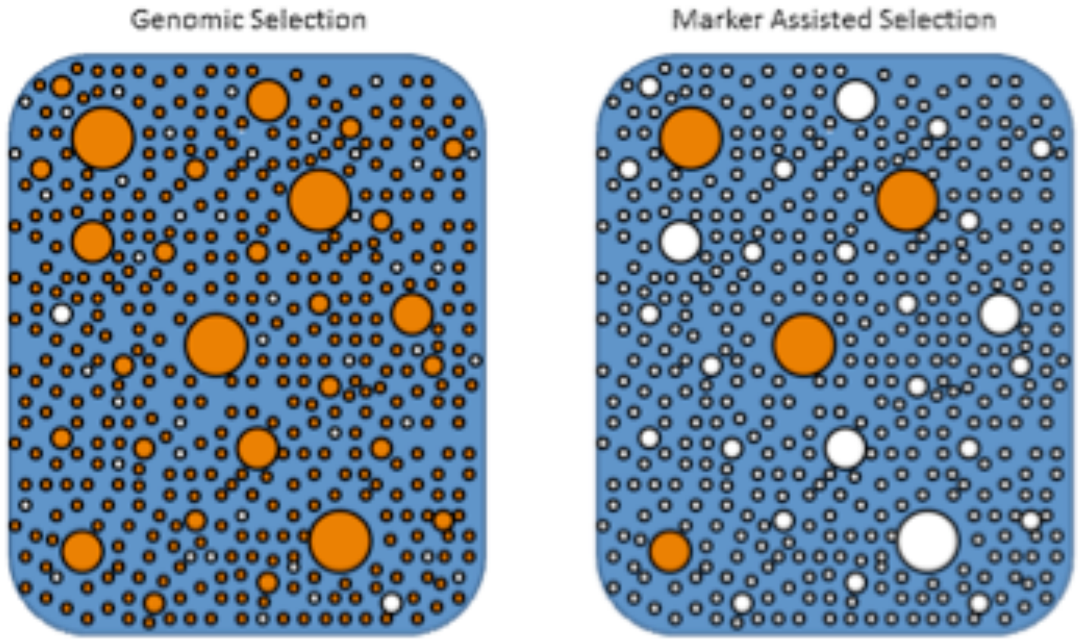
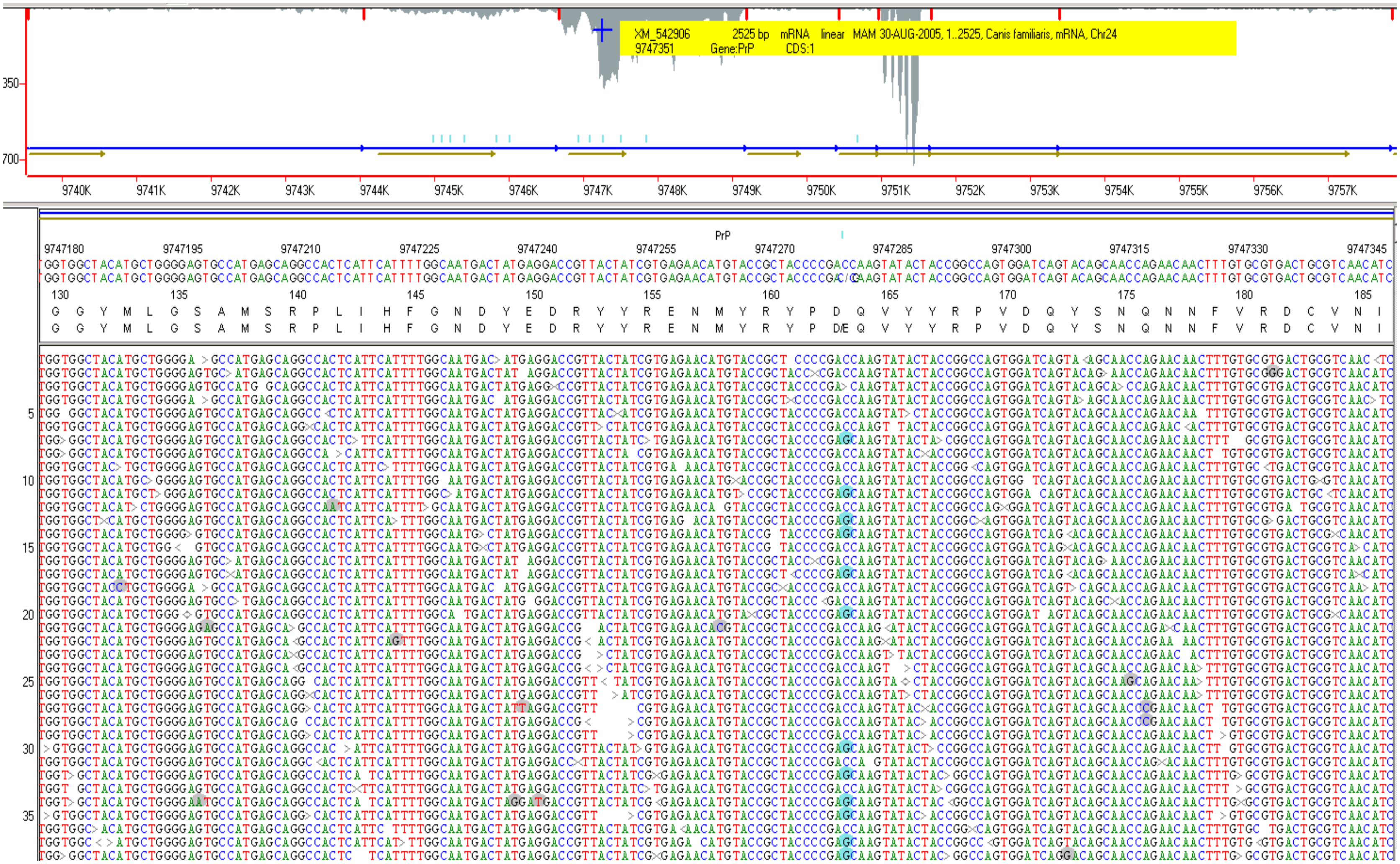


Figure 1: Depiction of traditional marker assisted selection (MAS) versus genomic selection (GS). The box represents the genome of an animal and the circles represent variation within the genome. The size of the circles represents the amount of genetic variation explained at that locus. The white circles represent variation that is not being selected due to a lack of a suitable closely-linked marker and the filled circles represent the variation which is under selection using each of the approaches.



```
XM_542906        2525 bp   mRNA   linear   MAM 30-AUG-2005, 1..2525, Canis familiaris, mRNA, Chr24
9747351        Gene:PrP        CDS:1
```

```
9747180         9747195         9747210         9747225         9747240         9747255         9747270         9747285         9747300         9747315         9747330         9747345
TGGTGGCTACATGCTGGGGAGTGCCATGAGCAGGCCACTCATTCATTTTGGCAATGACTATGAGGACCGTTACTATCGTGAGAACATGTACCGCTACCCCGACCAAGTATACTACCGGCCAGTGGATCAGTACAGCAACCAGAACAACTTTGTGCGTGACTGCGTCAACATC
TGGTGGCTACATGCTGGGGAGTGCCATGAGCAGGCCACTCATTCATTTTGGCAATGACTATGAGGACCGTTACTATCGTGAGAACATGTACCGCTACCCCGAC/GAAGTATACTACCGGCCAGTGGATCAGTACAGCAACCAGAACAACTTTGTGCGTGACTGCGTCAACATC
130             135             140             145             150             155             160             165             170             175             180             185
G  G  Y  M  L  G  S  A  M  S  R  P  L  I  H  F  G  N  D  Y  E  D  R  Y  Y  R  E  N  M  Y  R  Y  P  D   Q  V  Y  Y  R  P  V  D  Q  Y  S  N  Q  N  N  F  V  R  D  C  V  N  I
G  G  Y  M  L  G  S  A  M  S  R  P  L  I  H  F  G  N  D  Y  E  D  R  Y  Y  R  E  N  M  Y  R  Y  P  D/E Q  V  Y  Y  R  P  V  D  Q  Y  S  N  Q  N  N  F  V  R  D  C  V  N  I
```

Figure 2: Dog RNA-Seq data in a NextGene viewer showing the PRP region. The top panel shows the number of copies on the Y axis and chromosomal position on the X axis. The center panel shows the reference sequence compared to the sample sequence assembly and any detected amino acid change. The bottom panel shows the tiled sequences. A SNP can be observed and is highlighted in the tiled sequence.

Literature Cited

Backhed, F., H. Ding, T. Wang, L. V. Hooper, G. Y. Koh, A. Nagy, C. F. Semenkovich and J. I. Gordon. 2004. The gut microbiota as an environmental factor that regulates fat storage. Proc. Natl. Acad. Sci. 101(44):15718-23.

Decker, J. E., J. C. Pires, G. C. Conant, S. D. McKay, M. P. Heaton, K. Chen, A. Cooper, J. Vilkki, C. M. Seabury, A. R. Caetano, G. S. Johnson, R. A. Brenneman, O. Hanotte, L. S. Eggert, P. Wiener, J. J. Kim, K. S. Kim, T. S. Sonstegard, C. P. Van Tassell, H. L. Neibergs, J. C. McEwan, R. Brauning, L. L. Coutinho, M. E. Babar, G. A. Wilson, M. C. McClure, M. M. Rolf, J. Kim, R. D. Schnabel and J. F. Taylor. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. 106(44):18644-9.

Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. 1975. Biometrics 31:423-47.

Li, M., B. Wang, M. Zhang, M. Rantalainen, S. Wang, H. Zhou, Y. Zhang, J. Shen, X. Pang, M. Zhang, H. Wei, Y. Chen, H. Lu, J. Zuo, M. Su, Y. Qiu, W. Jia, C. Xiao, L. M. Smith, S. Yang, E. Holmes, H. Tang, G. Zhao, J. K. Nicholson, L. Li and L. Zhao. 2009. Symbiotic gut microbes modulate human metabolic phenotypes. Proc. Natl. Acad. Sci. 105(6):2117-22.

MacEachern, S., J. McEwan, A. McCulloch, A. Mather, K. Savin, and M. Goddard. 2009. Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in Domestic cattle? BMC Genom. 10:179.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4:e5350.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J. M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, MetaHIT Consortium, P. Bork, S. D. Ehrlich and J. Wang. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464(7285)59-65.

Rolf, M. M., J. F. Taylor, R. D. Schnabel, S. D. McKay, M. C. McClure, S. L. Northcutt, M. S. Kerley and R. L. Weaber. 2010. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. BMC Genet. 11:24.

Sellner, E. M., J. W. Kim, M. C. McClure, K. H. Taylor, R. D. Schnabel and J. F. Taylor. 2007. BOARD-INVITED REVIEW: Applications of genomic information in livestock. J. Anim. Sci. 85:3148-3158.

Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. Nat. Biotech. 26:1135-1145.

The Bovine Genome Sequencing and Analysis Consortium. 2009. The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324(5926):522-8.

The Bovine HapMap Consortium. 2009. Genome wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324(5926):528-532.

Turnbaugh, P.J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight and J. I. Gordon. 2009. A core gut microbiome in obese and lean twins. Nature 457:480-485.

Turnbaugh, P. J., R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis and J. I. Gordon. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444(7122):1027-31.

VanRaden P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor and F. S. Schenkel. 2009. Invited review: reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16-24.

Van Tassell C. P., T. P. L. Smith, L. K. Matukumalli, J. F. Taylor, R. D. Schnabel, C. T. Lawley, C. D. Haudenschild, S. S. Moore, W. C. Warren and T. S. Sonstegard. 2008. Simultaneous SNP discovery and allele frequency estimation by high throughput sequencing of reduced representation genomic libraries. Nat. Meth. 5:247-52.

Willham, R. L. 1993. Ideas into action: a celebration of the first 25 years of the Beef Improvement Federation. University Printing Services, Oklahoma State University, Stillwater, OK.

Zimin, A. V., A. L. Delcher, L. L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marcias, M. Roberts, P. Subramanian, J. A. Yorke and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos Taurus*. Gen. Bio. 10(4):R42.