

Low density genomic data for animal breeding: critical analysis and perspectives of the GoldenGate BeadXpress technology

Utilização de dados genômicos de baixa densidade no melhoramento animal: análise crítica e perspectivas da Tecnologia *GoldenGate BeadXpress*

Ronyere Olegário de Araújo¹ Alexandre Rodrigues Caetano^{II}

— REVIEW —

ABSTRACT

The increasing development of DNA sequencing and genotyping technologies has made possible to analyze the genomes of several species. Genomic studies of production animals have greatly increased the understanding of mechanisms that control the interactions of genetic and environmental factors involved in the expression of traits of economic importance. Several technologies have been presented by different companies for the genotyping of low-density SNP panels, which may be used in different applications with different goals, such as paternity testing, diagnosis of genetic diseases, and identification of genetically superior animals based on polymorphisms characterized in candidate genes. The present review critically analyzes the GoldenGate BeadXpress technology and puts its use in these applications into perspective.

Key words: *SNP genotyping, GoldenGate technology, haplotypic blocks, custom SNP panels, association studies.*

RESUMO

A crescente evolução das tecnologias de sequenciamento e genotipagem de DNA tornaram possível analisar o genoma de várias espécies, compreender suas funções dentro dos sistemas biológicos e, sobretudo, começar a entender os mecanismos que controlam as interações entre os genótipos e os efeitos ambientais que estão envolvidos com a expressão de características de interesse econômico. Várias tecnologias foram apresentadas por diferentes empresas para a genotipagem de painéis de SNP de baixa densidade, os quais podem ser utilizados em diferentes aplicações com objetivos variados, desde testes de paternidade e diagnósticos de doenças genéticas até a identificação de animais geneticamente superiores, com base em polimorfismos caracterizados em genes candidatos. Essa revisão analisa a tecnologia GoldenGate BeadXpress e coloca em perspectiva seu uso nessas aplicações.

Palavras-chave: *genotipagem de SNP, tecnologia GoldenGate, blocos haplotípicos, chips personalizados, estudos de associação.*

INTRODUCTION

Use of molecular markers in linkage disequilibrium, genetic mapping and association studies, as well as in diagnostic assays to detect genetic diseases and/or polymorphisms associated with production traits, have been long limited by technological constraints. However, this scenario has drastically changed by recent technological advancements observed in the past decade, which have generated new high-performance methodologies of greater accessibility and low cost for genotyping SNP (Single Nucleotide Polymorphisms) markers. These new methodologies revolutionized the way molecular data is used, especially in studies focused on prospecting genes of economic interest in production animals (CAETANO, 2009).

SNPs identified in candidate genes some biological functions associated with traits of interest, which have been used by different groups as a tool to identify and select genetically superior animals. A large number of candidate genes with tested effects on milk production traits was identified in cattle, including genes related to the resistance to pathogens

¹Faculdade Presidente Antônio Carlos (FAPAC), Instituto Tocantinense Presidente Antônio Carlos (ITPAC), 77500-000, Porto Nacional, TO, Brasil. E-mail: ronyereo@yahoo.com.br. Corresponding author.

^{II}Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil.

and parasites, and to milk production and composition (KHATKAR et al., 2008). Association of candidate genes with phenotypic traits may be performed based on a single polymorphic site or based on haplotypes reported in more than one SNP observed in the genes in question.

The aim of the current review was to provide a brief critical evaluation of the GoldenGate genotyping technology for generating low-density genomic data for application in association studies with polymorphisms in candidate genes associated with traits of interest for milk production in cattle.

DEVELOPMENT

GoldenGate genotyping technology

SNP genotyping technologies have advanced rapidly in recent years, followed by an increased interest in the use of these genetic markers to map complex diseases and to perform association studies with production traits. These combination interests resulted in a wide variety of methods which have been developed and described to characterize and genotype these genetic markers. The GoldenGate assay allows for a high degree of parallelization or multiplexing assays. A total between 96 and 1536 SNPs can be genotyped in a single sample, since the crucial stages of the process may be performed in parallel in the same tube. This parallelization results in great savings in time, required reagents and initial DNA samples. In addition, samples are processed in plate format, and allowed simultaneously genotyping a total of 96 samples.

GoldenGate assays may be adapted to genotype SNP marker panels using VeraCode™ microbeads or rods, which are individually barcoded. In combination, both methods fulfill the fundamental prerequisite for any type of DNA array, according to CHAUDHURI (2005): the existence of a “unique address” for each SNP in the panel. According to the protocol recommended by the manufacturer (FAN et al., 2003), the initial step in the GoldenGate assay corresponds to the activation of genomic DNA (gDNA) through incorporation of biotin, which allows for the subsequent ligation to streptavidin-conjugated paramagnetic particles (Figure 1 - step 1). Three oligonucleotides (oligos) are necessary to genotype each activated SNP (Figure 1 - step 2). Two oligos are specific to each of the alleles previously observed at the SNP site (3'-terminus), which are called allele-specific oligos (ASO). The third oligo (locus-specific oligo - LSO) is complementary to 50 bases starting from the SNP position (5'-terminus). When there is perfect

alignment between the specific oligo and the activated gDNA, the process of extension/ligation up to the locus-specific oligo is completed (Figure 1 - steps 3 and 4). With respect to the subsequent process step (Figure 1 - step 5), allele-specific oligos marked with two different fluorochromes - Cyanine-3 (Cy3 - green) and Cyanine-5 (Cy5 - red) - are used to perform an amplification reaction by PCR (Polymerase Chain reaction).

After the amplification step (Figure 1 - step 6), labeled PCR products are hybridized to VeraCode® silica rods (Figure 1 - Step 7), which contain complementary oligos and a bar code specific to each SNP evaluated in the assay (JALURIA et al., 2007). After the hybridization, the VeraCode® silica rods are read in the BeadXpress® scanner, which uses a laser to stimulate the fluorochromes, and analyze/detect the fluorescence signal emitted by each VeraCode® rod of each tested sample (Figure 1 - step 8). The device simultaneously captures two different wave lengths emitted by the used fluorochromes (Cy3 and Cy5) and uses the bar code recorded on the rod to determine the specific SNP marker to be evaluated (Figure 1 - step 8). Homozygous individuals emit only one fluorescence signal (Cy3 or Cy5) for a particular SNP marker, whereas heterozygous individuals simultaneously emit signal at both wavelengths (Figure 1 - step 9).

Linkage disequilibrium estimates and identification of haplotype blocks

The term Linkage Disequilibrium was first used in the study by LEWONTIN & KOJIMA (1960). Linkage disequilibrium (LD) is a measure of correlation between alleles at different loci and it represents the lack of independence between the alleles arranged in a haplotype. A haplotype is formed by a combination of alleles at adjacent loci. These alleles are part of the same chromosome and they are transmitted in a non-independent way. A haplotype may be formed by two or more adjacent markers.

LD and physical linkage are terms commonly confused with one another and that are often presented as synonymous. This confusion usually arises because closely linked loci may be at high LD, as it was pointed out by FLINT-GARCIA et al. (2003). However, the fact that any two loci are at LD does not mean that they are physically linked. Similarly, the fact that two loci are physically linked on a particular chromosome or haplotype does not necessarily imply that they are at LD (SLATKIN, 2008).

DEKKERS (2004) suggested that, due to the LD extension seen in the entire genome in animal populations of zootechnical interest, informative

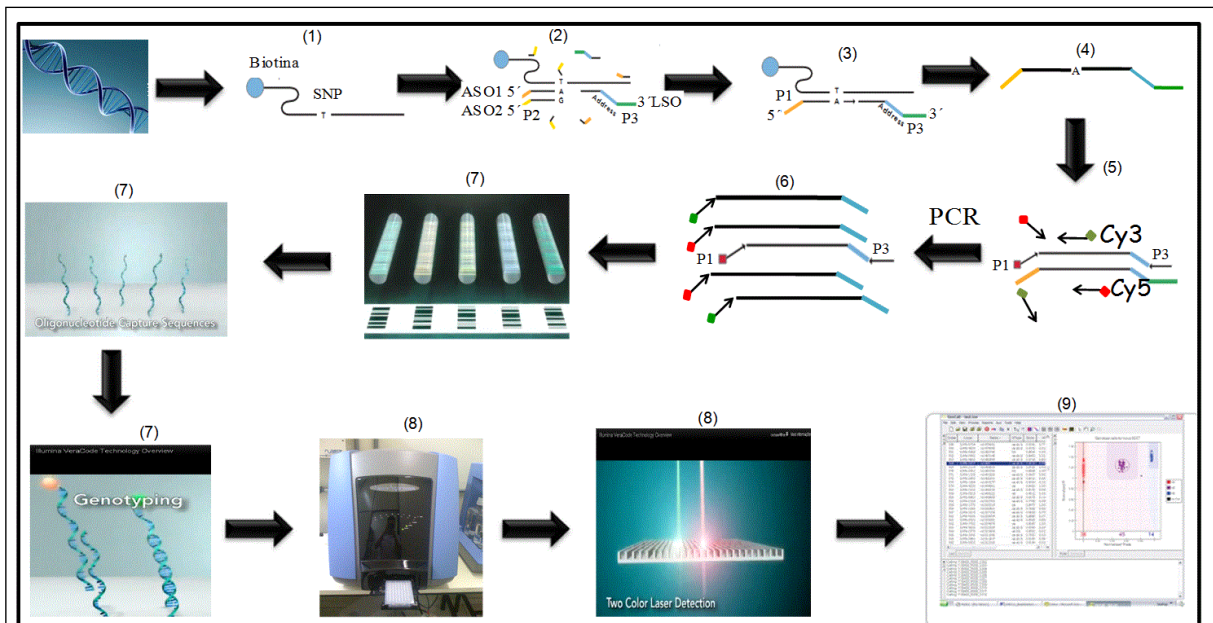


Figure 1 - Flowchart illustrating the oligonucleotide bonding technique used to detect alleles at SNP sites - VeraCode method (source: <www.illumina.com>).

markers every 1 or 2 centimorgans (cM) may be sufficient to detect most of the QTLs affecting production traits. A contrasting situation is observed in humans, since LD levels observed in most studied populations is very high due to high N_e . GABRIEL et al. (2002) showed that the density of one marker every 7.8Kb was necessary to identify the structure of haplotype blocks throughout the human genome.

MCKAY et al. (2007) evaluated the LD levels in Holstein and Angus cattle using approximately 3,000 SNPs in the genome and r^2 (defined in the next section) as a measure of LD. r^2 has been the most robust among all parameters used to estimate LD, especially in small populations, because it is less influenced by allele frequencies (DU et al., 2007). The results obtained by MCKAY et al. (2007) showed that at least 30,000 markers are required to capture the genetic variation distributed throughout the bovine genome in association analyses.

Linkage disequilibrium measures

Since SNP markers are normally biallelic, if one considers two polymorphic loci (A/T and C/G) in any N generation of a given population, four different gametes or haplotypes may be identified (AC, AG, TC, TG). If these loci are at linkage equilibrium (DL = 0 and LE = 1), i.e., in a situation of random association between their alleles,

the observed haplotype frequencies should equal the expected frequencies, which correspond to the product of the respective allele frequencies. Thus, key parameters used to estimate LD levels within a population are the expected deviations (D') and the coefficient of determination (r^2), as suggested by HILL & ROBERTSON (1968).

D' may be calculated as follows:

$$|D'| = \frac{|D|}{D^{max}}$$

wherein: D' is the measure of LD, $D = f_{AB} - f_A f_B$, and D^{max} is the maximum possible value of D . Thus, if $D_{ab} < 0$, we have:

$$|D'| = \frac{(D_{ab})^2}{\min(f_A f_b, f_a f_B)}$$

However, when $D_{ab} > 0$, we have:

$$|D'| = \frac{(D_{ab})^2}{\min(f_A f_B, f_a f_b)}$$

wherein: f is the frequency of each allele (A or B) and min = is the lowest possible frequency.

The expected deviation (D') is a relative measure and it may vary between 0 and 1 ($0 \leq D' \leq 1$). If $D' = 1$, it can be said that at least one of the four

haplotypes is absent, thus featuring an LD situation (SLATKIN, 2008). Another LD measure is the coefficient of determination (r^2), which was suggested by HILL & ROBERTSON (1968). This coefficient represents the squared correlation between the presence and absence of alleles at different loci. This measure may be described as:

$$r^2 = \frac{(D_{ab})^2}{\min(fAfB, fafb)}$$

wherein: $D = f_{AB} - f_{AfB}$; $f_{A,fa}$, $f_{B,fb}$ and fb are the allele frequencies and f_{AB} , f_{Ab} , f_{aB} and f_{ab} are the haplotype frequencies.

Similarly to D' , the r^2 may vary between 0 and 1 ($0 \leq r^2 \leq 1$). Therefore, when r^2 tends to 1, it suggests that at least one haplotype is absent in the sample, thus featuring an LD situation (SLATKIN, 2008). The r^2 and D' statistical measures reflect different linkage disequilibrium aspects and show different behaviors under different conditions. There is only one value equal to 1 in r^2 , i.e., complete linkage disequilibrium, when the two loci in question have identical allele frequencies. This difference in linkage disequilibrium observed when r^2 or D' is adopted results from the evolutionary history that each of these statistics incorporates to measure the linkage disequilibrium, and each method has its advantages and disadvantages.

GABRIEL et al. (2002) determined that at high LD levels (little or no evidence of recombination in the genome) it can be observed estimated D' values between allele pairs with 95% confidence intervals, i.e., with upper limits above 0.98 and lower limits above 0.7. Conversely, pairs of loci for which D' values showed upper limits below 0.9 can be considered as "strong evidence of historical recombination". In parallel, the four-gamete rule is a variation that uses the four possible haplotype population frequencies of each pair of computed markers. If all four are observed with frequencies of at least 0.01, recombination is considered to have occurred. Blocks are formed by consecutive markers in which just three gametes are observed. Another way to define haplotype blocks is through the Solid Spine of Linkage Disequilibrium (BARRETT et al., 2005). All these methods are used to statistically determine a value for the association between alleles and, from this value, define the assembly forming the haplotype block.

Another advantage of using haplotype blocks to perform genetic characterization is that high LD regions showed limited diversity (ALTSHULER et al., 2005). Thus, a small number of distinct

haplotypes is predicted in each block for most of the chromosomes in the populations in comparison to the number predicted by the permutation of combinations among them, if the alleles of different loci join together at random, i.e., 2^n , wherein n is the number of markers. Therefore, haplotype LD blocks have been widely used in association studies and they have shown great importance in the association of candidate genes in complex diseases. The advancement of the simpler, faster and more accurate technologies allows projects to be carried out both in gene regions of small extent in indirect association studies and in association studies on full genome.

Practical application of VeraCode (Beadxpress) technology

Results presented by GONZÁLEZ-NEIRA (2013) point out to the GoldenGate methodology viability in genotyping procedures conducted in custom SNP panels. According to the author, the tool ensures the verification of minor-allele frequencies and allows obtaining higher probability of success in the experiment validation. In addition, since the filtering criteria exclude unreliable samples and markers from the experiment, one can establish a panel of low-density SNP markers to verify criteria such as reproducibility and genetic inheritance.

Also aiming to develop and validate a rapid method to genotype the main mutations of the bovine milk proteins, CHESSA et al. (2007) conducted experiments with microchips to simultaneously identify 22 distinct SNPs located in 8 different regions of milk-protein-related genes. The used platform allowed inserting and deleting the polymorphisms to be genotyped. Obtained results showed that the microchips allowed finding easy genotyping in different genomic regions, which could be investigated in phylogenetic and association studies; assessing the genetic distances between the cattle breeds; assessing the bovine species history; and identifying parentage.

ONTERU et al. (2011) analyzed the reproductive traits of female pigs using MAF thresholds >0.001 and genotyping efficiency $>80\%$ to check the EHW and to exclude SNPs with genotyping errors. The Gentrain Score used in this analysis was based on genotype-cluster analyses performed through the GenCall score developed by Illumina. After this quality control, 6,814 SNPs (11%) were removed from the genotypic database. These authors were less strict with the thresholds set in the quality criteria of animal samples/markers in comparison to those adopted in the herein presented chapter.

ANKUNOV et al. (2009) used the VeraCode technology to genotype SNPs in wheat-based experiments. The method effectively determined polyploid wheat samples, with average of 89% tetraploid and 84% hexaploid genotypes. After filtering the low-quality SNPs in this experiment, it was possible to observe genotyping accuracy at the average of 99% hexaploid and 100% tetraploid genotypes. According to the studies by DEULVOT et al. (2010), who also performed an experiment with vegetable culture (peas), the GoldenGate assay proved to be equally effective in validating the SNP panel. This panel was initially composed of 384 SNPs, and 356 (92.7%) of them remained after the filtering criteria was applied, thus keeping the GenTrain quality parameter level higher than 94%. Similarly to the research focus of the aforementioned study, the GoldenGate technology has been used in experiments performed by ANKUNOV et al. (2009) and DEULVOT et al. (2010), who considered it an appropriate tool to check genetic variations of loci associated with complex traits, with favorable results to this type of culture.

Critical evaluation of the VeraCode (BeadXPress) Technology

The GoldenGate assay associated with the VeraCode (BeadXPress) technology can be used to obtain 36,864 genotypes (384 SNPs X 96 samples) in two working days (FAN et al., 2003; Illumina - San Diego, USA). This scenario represents significant progress in the generation of molecular data in comparison to those obtained by microsatellites-based genotyping, which might generate 7,680 genotypes in the same working period under optimal conditions (standardized Multiplex, 5 primers, 96 samples, ABI 3730 sequencer with 16 plates).

Regarding the structure and procedures required to generate genomic data, it is worth mentioning some of the difficulties that can be encountered, especially in the laboratory perspective. The VeraCode (BeadXPress) technology has a complex laboratory routine, since the process used to obtain these results involves the DNA sample manipulation using 18 different reagents applied to 5 different plates. This complexity becomes even greater when one adds factors such as the “transport” of these reagents used. There are temperature specificities for each reagent and variations may affect the reagents’ chemical properties, which may eventually lead to biased results. Another transport-related limiting factor lies on the hybridization plates, in which pre-synthesized oligonucleotides are fixed on silica rods. These rods are glass cylinders with micrometric dimensions and

excessive vibrations on the plates may damage them and, consequently, make it difficult for the reader to capture the genotype information. The equipment used in the laboratory routine must be previously approved by the Illumina technicians and the end user must be qualified to initiate the activities (training provided by Illumina). In addition, annual calibrations - from the automatic pipettes to the BeadXPress scanner - are necessary so that the company can ensure the quality of generated data. As a consequence of all of these technological issues, the Illumina Company announced in early 2013 that it would no longer accept new purchase applications for the BeadXPress equipment, thus effectively discontinuing the sale of low-density commercial and customized assays based on this technology.

The current technological context shows positive perspectives, since new incremental advancements are often released to the already available technologies used in genotyping and sequencing with different levels of several species coverage. In addition, new technologies have been developed and improved to create possibilities for new cost reductions and to set new application limits, which were once limited by human inventiveness alone (Oxford Nanopore Technologies®, SCHNEIDER & DEKKER, 2012). In light of the foregoing, outsourced genotyping and sequencing services become more attractive and safe for the user/client.

Methods developed by (MEUWISSEN et al., 2001) termed Genomic Selection (GS) used high genome coverage with markers to partition a large part of the additive genetic variability of a trait throughout the genome, allowing to estimate the allele substitution effects in each of the loci involved in the trait. The procedure can be used to generate highly-accurate estimates of genetic values of any genotyped individual tested in a population or breed (termed Genomic Breeding Value - GBV), based on genetic effects estimated in a training dataset (MEUWISSEN et al., 2001).

Estimated GBVs may be used to prematurely select genetically superior individuals for breeding towards established improvement goals, thus, accelerating rates of genetic gain (RESENDE et al., 2008). Low-density tests (Infinium BovineLD - 7K) are already available for applications such as genomic selection, paternity testing and tracking. The users of this chip may compare its data to those generated in high-density platforms (charged with 98% reliability) and extend the genomic information application since the new product is complementary to Bovine SNP50 and to other similar tests.

CONCLUSION

As evidenced by the current review, the use of genomic tools to implement genetic-merit evaluation processes resulted in significant technological advancements, and it may lead to increased efficiency and higher profits due to increased accuracies and to reduced generation intervals. In addition, there are still great perspectives for launching new technological innovations for generating molecular data in the short and medium terms, with cost reduction trends even for the complete sequencing of genomes in several livestock species, thus increasing the limits provided by technologies.

REFERENCES

- ALTSHULER, D. et al. A haplotype map of the human genome. **Nature**, v.437, n.7063, p.1299-320, 2005. Available from: <<http://www.nature.com/nature/journal/v437/n7063/full/nature04226.html>>. Accessed: Jun. 04, 2016.
- ANKUNOV, E. et al. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. **Theoretical and Applied Genetics**, v.119, n.3, p.507-517, 2009. Available from: <<http://link.springer.com/article/10.1007%2Fs00122-009-1059-5>>. Accessed: Jun. 04, 2016.
- BARRETT, J.C. et al. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, v.21, n.2, p.263-265, 2005. Available from: <<http://bioinformatics.oxfordjournals.org/content/21/2/263.full>>. Accessed: Jun. 04, 2016.
- CAETANO, A.R. SNP markers: basic concepts, applications in animal breeding and management and perspectives for the future. **Revista Brasileira de Zootecnia**, v.38, n.spe, p.64-71, 2009. Available from: <http://www.scielo.br/scielo.php?script=sci_artext&pid=S1516-35982009001300008>. Accessed: Jun. 04, 2016.
- CHAUDHURI, J.D. Genes arrayed out for you: the amazing world of microarrays. **Medical Science Monitor**, v.11, n.2, p.52-62, 2005. Available from: <<http://www.medscimonit.com/download/index/idArt/14231>>. Accessed: Jun. 05, 2016.
- CHESSA, S. et al. Development of a single nucleotide polymorphism genotyping microarray platform for the identification of bovine milk protein genetic polymorphisms. **Journal of Dairy Science**, v.90, n.1, p.451-464, 2007. Available from: <[http://www.journalofdairyscience.org/article/S0022-0302\(07\)72647-4](http://www.journalofdairyscience.org/article/S0022-0302(07)72647-4)>. Accessed: Jun. 05, 2016.
- DEKKERS, J.C.M. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v.82, n.6, p.312-328, 2004. Available from: <http://www.psas-web.net/documents/ask_psas/Dekkers%20BLUP%20paper.pdf>. Accessed: Jun. 05, 2016.
- DEULVOT, C. et al. Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. **BMC Genomics**, v.11, n.468, p.1-10, 2010. Available from: <<http://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-11-468>>. Accessed: Jun. 05, 2016.
- DU, F.X. et al. Characterizing linkage disequilibrium in pig populations. **International Journal of Biological Sciences**, v.3, n.3, p.166-178, 2007. Available from: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1802018/>>. Accessed: Jun. 05, 2016.
- FAN, J.B. et al. Highly parallel SNP genotyping. **Cold Spring Harbor Symposia on Quantitative Biology**, v.68, p.69-78, 2003. Available from: <http://dnatech.genomecenter.ucdavis.edu/wp-content/uploads/2013/06/highly_parallel.pdf>. Accessed: Jun. 08, 2016.
- FLINT-GARCIA, S.A. et al. Structure of linkage disequilibrium in plants. **Annals Reviews in Plant Biology**, v.54, n.6, p.357-374, 2003. Available from: <<http://www.pnas.org/content/98/20/11479.full>>. Accessed: Jun. 08, 2016.
- GABRIEL, S.B. et al. The structure of haplotype blocks in the human genome. **Science**, v.296, n.5576, p.2225-2229, 2002. Available from: <<http://science.sciencemag.org/content/296/5576/2225.full>>. Accessed: Jun. 08, 2016.
- GONZÁLEZ-NEIRA, A. The GoldenGate genotyping assay: custom design, processing, and data analysis. **Methods in Molecular Biology**, v.1015, n.5, p.147-153, 2013. Available from: <http://link.springer.com/protocol/10.1007%2F978-1-62703-435-7_9>. Accessed: Jun. 09, 2016.
- HILL, W.G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, v.38, n.6, p.226-231, 1968. Available from: <<http://link.springer.com/article/10.1007/BF01245622>>. Accessed: Jun. 09, 2016.
- JALURIA, P. et al. A perspective on microarrays: current applications, pitfalls, and potential uses. **Microbial Cell Factories**, v.13, n.6, p.1-14, 2007. Available from: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1796898/>>. Accessed: Jun. 09, 2016.
- KHATKAR, M.S. et al. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics**, v.9, n.187, p.1-18, 2008. Available from: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2386485/>>. Accessed: Jun. 09, 2016.
- LEWONTIN, R.C.; KOJIMA, K. The evolutionary dynamics of complex polymorphisms. **Evolution**, v.14, n.4, p.458-472, 1960. Available from: <<http://ib.berkeley.edu/labs/slatkin/popgenclub/pdf/lewontin-kojima1960.pdf>>. Accessed: Jun. 09, 2016.
- MCKAY, S.D. et al. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, v.8, n.74, p.1-12, 2007. Available from: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2174945/>>. Accessed: Jun. 09, 2016.
- MEUWISSEN, T.H. et al. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v.157, n.4, p.1819-1829, 2001. Available from: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461589/pdf/11290733.pdf>>. Accessed: Jun. 09, 2016.
- ONTERU, S.K. et al. Whole-genome association analyses for lifetime reproductive traits in the pig. **Journal of Animal Science**, v.89, n.4, p.988-995, 2011. Available from: <<https://www.animalsciencepublications.org/publications/jas/articles/89/4/988>>. Accessed: Jun. 09, 2016.

RESENDE, M.D.V. et al. Genome wide selection (GWS) and maximization of the genetic improvement efficiency. **Pesquisa Florestal Brasileira**, v.1, n.56, p.63-77, 2008. Available from: <<http://pfb.cnpf.embrapa.br/pfb/index.php/pfb/article/view/63/65>>. Accessed: Jun. 09, 2016.

SCHNEIDER, G.F.; DEKKER, C. DNA sequencing with nanopores. **Nature Biotechnology**, v.30, n.4, p.326-328, 2012.

Available from: <<http://www.nature.com/nbt/journal/v30/n4/full/nbt.2181.html>>. Accessed: Jun. 09, 2016.

SLATKIN, M. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. **Nature Reviews Genetics**, v.9, n.6, p.477-485, 2008. Available from: <<http://cteg.berkeley.edu/papers/slatkin/SlatkinNatRevGenet2008.pdf>>. Accessed: Jun. 09, 2016.