STATE OF THE ART OF GENOMIC SEQUENCING IN LIVESTOCK

ESTADO DA ARTE DO SEQUENCIAMENTO GENÔMICO NA PECUÁRIA

G. L. PEREIRA[1*], K. O. ROSA[1], R. A. CURI[2], L. C. A. REGITANO[3], M. D. S. MOTA[2]

SUMMARY

Genomics has advanced greatly since 1975 when the possibility of determining the nucleotide sequence of the genome was first described. The human genome sequencing started in the 90s and it was greatly favored by advances in computer technologies. In the last ten years the development of next generation sequencing technologies allowed the sequencing of millions of bases at low cost and in a shorter time period compared to the previous technologies. After the conclusion of the human genome project, several initiatives to sequence the genome of domestic animals species were undertaken, which resulted in a large amount of data and are redirecting the goals of genetic studies in domestic animals. This review describes the state of the art genomic sequencing of the main domestic animals species of economic interest, as well as the most important tools available to access this information.

KEY-WORDS: BAC. Genome. Next Generation Sequencing. SNP. WGS.

RESUMO

Desde que a possibilidade de determinar a sequência nucleotídica de genomas surgiu em 1975, muitos foram os avanços da genômica. Na década de noventa teve início o sequenciamento do genoma humano, viabilizado em grande parte pelos avanços nos métodos computacionais. Nos últimos dez anos, o advento de tecnologias de sequenciamento de nova geração permitiu o sequenciamento de milhões de bases a baixo custo e em curto espaço de tempo quando comparadas às técnicas de sequenciamento anteriores. Após a conclusão do projeto genoma humano, muitas iniciativas foram tomadas para a realização do sequenciamento de diversas espécies domésticas, gerando grande volume de dados, e redirecionando os estudos genéticos. Esta revisão teve como objetivo descrever o estado atual do sequenciamento das principais espécies de animais domésticos de interesse econômico, bem como de expor as ferramentas mais utilizadas no acesso às informações genômicas.

PALAVRAS-CHAVE: BAC. Genoma. Next Generation Sequencing. SNP. WGS.

[1] Universidade Estadual Paulista "Julio de Mesquita Filho" - Faculdade de Ciências Agrárias e Veterinárias. Jaboticabal, SP.
*Corresponding author: guipicoia@hotmail.com
[2] Universidade Estadual Paulista "Julio de Mesquita Filho" - Faculdade de Ciências Agrárias e Veterinárias. Botucatu, SP.
[3] EMBRAPA - Pecuária Sudeste. São Carlos, SP.

INTRODUCTION

Genome sequencing began in the mid-70s, first with the development of the methodology described by Frederich Sanger and colleagues in 1975, which became known as Sanger's method, dideoxy or chain termination. Independently, Maxam and Gilbert developed in 1976, the chemical sequencing technique, which has been overlooked/ignored due to greater technical difficulty for automation (MAXAM & GILBERT, 1977; SANGER & COULSON, 1975). The Sanger method is based on the principle of controlled polymerization termination by using Dideoxynucleotides, thus enabling to determine the sequential order of nucleotides that make up the fragments of the DNA molecule. The original version was not the appropriate tool for the sequencing of complex genomes (FURLAN *et al, 2007;* SANGER & COULSON, 1975). However, in the late 80s, after the Sanger method had been modified, the sequencing was automated and integrated with computerized reading systems (VENTER *et al.,* 2001).

In the middle of the last decade, platforms based on Next-Generation Sequence (NGS) technologies, started being commercialized and have been widely employed to replace the Sanger method. The use of the Sanger's technique to sequence large genomes such as the human genome is less efficient than NGS based methods, presenting, in addition to high costs, great time requirements.

The genome of many domestic animals of economic interest has already been sequenced, for the most part. Some projects that started before the NGS, later adopted such technologies to complete or improve the quality of generated data. Goats *(Capra hircus),* one of the last species to have their genome decoded, was completely sequenced by the NGS technology (WADE *et al., 2009;* DOAN *et al., 2012;* DONG *et al., 2012).* These technologies have been widely used for sequencing of animals of different breeds of species already sequenced, aiming to discover the copy number variants (CNVs), insertions and deletions (InDels) and new single nucleotide polymorphisms (SNPs), among others. These studies have contributed greatly to improve the quantity and quality of genomic data for these species (DOAN *et al., 2012;* BARRELS *et al., 2012).*

Given the above, the objective of this review was to describe the current state of genomic sequencing of the major livestock species and to describe computational tools for accessing the genomic data and technologies developed from this information.

Strategies for sequencing and next-generation sequencing

Before the sequencing itself, either by Sanger or NGS method, it is necessary to adopt strategies to obtain DNA fragments of the genome to be sequenced since it is not possible to sequence large DNA segments.

The two main strategies used for this purpose are the whole genome shotgun (WGS) and another called sequencing of ordered clones. Both strategies aim to generate genomic libraries of clones obtained in the fragmentation of the genome in thousands of small random segments, followed by the reading of each sequence and their overlaps, by computational means (Griffiths, 2008).

The basic difference between the strategies is prior knowledge of the location of the sequenced genome fragments. Strategies that use clones of known sequences, such as ordered clone sequencing, can obtain the sequences from bacterial artificial chromosomes (BAC), which are isolated and mapped to the genome, then individually sequenced. The WGS genome sequences random fragments obtained by non-mapped clones, so that after overlapping the read sequences and forming the *contigs,* they are mapped to the genome (VENTER et al., 2001; Griffiths, 2008).

Twelve years after the human genome sequencing was published on February 15, 2001 in *Nature* (VENTER *et al.,* 2001), sequencing techniques and genome studies have advanced greatly. Next-generation sequencing methods have become reality due to the rapid technological advances of the last decade. NGS allows the sequencing and analysis of millions of base pairs (bp) in a single round of reading aided by bioinformatics tools. In the next 5 to 10 years, besides the sequencing speed, costs will also be greatly reduced, about 10,000-fold compared to the first complete sequencing of genomes (PAREEK *et al.,* 2011).

The first NGS generation began with the development of the 454 life science equipment in 2000 by Jonathan Rothberg (www.454.com). This new platform of massive DNA sequencing was validated by Margulies *et al.* (2005) early on, who describes in an article published in *Nature* a "highly parallel and scalable sequencing system with significantly more automation than standard capillary electrophoresis instruments", which allows to produce 25 million base pairs in a single reading with over 99% accuracy. On this occasion, the resequencing of 96% of the bacterium *Mycoplasma genitalium* genome was performed with high accuracy (99.96%) in a single reading. This procedure, unlike the technique that adopts capillaries (needed for electrophoresis), has real-time sequencing by synthesis approach, and relies on the detection of luminometric inorganic pyrophosphate (PPi) released with the incorporation of the triphosphate deoxyribonucleotides (RONAGHI *et al.*, 1998).

The second NGS generation is represented primarily by three platforms: 454 FLX (Roche, USA); Solexa (Illumina Inc., USA) and SOLiD (Applied Biosystems, USA), all of which have a high rate of sequencing, yielding 150 to 200 Gb per week and average fragment size of 100 bp (OAK & SILVA, 2010). The third NGS generation, unlike the second, does not require DNA amplification, whose goal is to strengthen the light signal for capture based on CCD cameras, which can generate distortion due to the abundance of fragments. The nano-scale miniaturization and minimal reagents use of the third

generation have made possible the sequencing of a single DNA molecule (SCHADT *et al.,* 2010). The first sequencer of this generation was the HeliScope™ (Pacific Bioscience, USA), followed by others as the SMRT™ (Pacific Bioscience, USA) (PAREEK e*t al.,* 2010). The third generation platforms produce sequences between 30 and 200 times longer than the second. This feature can be of great advantage when assembling genomes using bioinformatics (ROBERTS *et al.,* 2013).

Genomic sequencing in livestock

The last two decades have seen an explosion of genome sequencing projects. Thus, all species that somehow are of interest to humans, including the human genome itself, had or are having their genetic code deciphered. Among these species are those explored as livestock, such as poultry, cattle, swine and ovine, among others. The genomic sequencing state of species of economic interest in livestock is summarized in Table 1.

Poultry Genome

In 2002, prior to the complete sequencing of the genome, 339,314 *expressed sequence tags* (ESTs) were sequenced, from 64 libraries of complementary DNA (cDNA) originating from different tissues of domestic fowl, which were clustered and assembled into 85,486 contigs (contiguous sequences). At least 38% of orthologous contigs were observed in other organisms and 13,000 new genes have been documented (BOARDMAN, 2002).

Funded by the National Human Genome Research Institute (NHGERI, USA), the project of the genome sequencing of the domestic fowl was started in March 2003, spending about $13 million (NHGRI, 2012). In 2004, with 38 pairs of autosomal and one pair of sexual chromosomes, the domestic chicken *(Gallus gallus)* was the first livestock species to have its genome sequenced.

The species genome comprises approximately one billion base pairs, about 40% of the size of the human genome, and has approximately 23,000 genes in its structure. This draft sequence of the chicken genome was obtained from the DNA from an inbred female Red Junglefowl, to minimize heterozygosity and provide sequences of both sexual chromosomes, Z and W. The sequences were generated with 6.6x coverage by WGS, plasmids and BACs (HILLIER *et al.,* 2004) libraries.

In May 2006, the Gallus_Gallus 2.1 set of data produced by the Genome Sequencing Center at Washington University School of Medicine was submitted by the International Chicken Genome Sequencing Consortium. Approximately 95% of the sequences were anchored to autosomal chromosomes 1-24, 26-28, and 32, and sexual W and Z. Later on, in September 2011, the same Sequencing Center, aiming to correct gaps and duplication of version 2.1, presented a new set of data obtained from the RJF # 256 individual, already used in previous assemblies. This new version 4.0 with coverage of 12x, obtained from Roche® 454™ platform, represented a significant improvement of the information contained in the Gallus_Gallus 2.1 data set (UCSC, 2012).

Two thousand seven hundred thirty-six Quantitative Trait Loci (QTL) representing 257 different features have been described in 132 publications, which are included in the data base *chickenQTLdb* (NAGRP, 2012).

More than 2.8 million SNPs were identified by comparing the genome sequence of the ancestral *(Gallus gallus)* with genome sequences obtained from three domesticated lineages: a male broiler *(White Cornish;* a laying hen *(White Leghorn);* and, a female ornamental species *(Chinese Silkie)* (NCBI, 2012). The identification of SNPs from contigs derived from NGS resulted in increased coverage of the chicken genome, allowing the design of Illumina SNP BeadChip (Illumina Inc., USA) of moderate density with approximately 64,000 SNPs (60k) (GROENEN *et al .,* 2011).

Bovine genome

In December 2003, the International Bovine Genome Sequencing Project was launched as a result of a workshop held in June 2003, in Montreal, Canada. In December 2004, the first draft of the bovine genome sequence was published in free public databases. Among the many entities involved, Genome British Columbia, along with the National Human Genome Research (NHGRI) and the United States Department of Agriculture (USDA) were some of the most important partners in the financing and development of the project (Genome Canada, 2013).

The bovine genome sequencing employed the WGS approach combined with BAC. Many of the BACs were sequenced using the so-called Clone-Array Pooled Shotgun Strategy (CAPSS) to reduce costs (LIU *et al.,* 2009).

In 2004, map with 3x coverage was generated by the WGS strategy with small inserts. In March 2005, a second map with 6.2x coverage was generated using the WGS of small clones and BACs with sequenced ends. These were the preliminary BTAU 1.0 and 2.0 maps, respectively. In 2006, BTAU 3.1 with 7.1x coverage was obtained using WGS with small insertions and BACs sequenced individually or by CAPSS. In 2007, the BTAU 4.0 with 7.1x coverage was generated with the same techniques used to create the BTAU 3.1 map. Currently, the version 4.6.1 of the BTAU map was released in 2012, after the 4.2 and 4.5 versions published in 2009. In April 2009, the high-quality genome sequence of the bull *(Bos primugenius taurus)* was published in the journal *Nature,* with 7.1x coverage, by the Bovine Genome Sequencing and Analysis Consortium (BCM, 2012).

The bovine genome sequences UMD 2.0 and UMD 3.1 were mapped in Baltimore, Maryland, USA, by the Salzberg laboratory team using information from NCBI reinforced with modifications by powerful installation and mapping tools. UMD 2.0 was built with 24 million WGS sequences and 11,000 BAC clones. In December 2009, seven months after the publication of UMD 2.0, UMD 3.1 was released with coverage of 9.5x and 2.85 Gb sequenced, of which

more than 2.64 Gb were allocated on some chromosome, 30Mb more than UMD 2.0. UMD 3.1 corresponds to the same UMD 3.0 set, released in August 2009, with only minor modification to be incorporated into the GenBank, without changing the content (ZIMIN et al., 2009; BAI *et al.,* 2012).

In December 2011, the genome sequencing of zebu cattle *(Bos primogenius indicus)* using the SOLiD sequencing platform and 52x coverage was published. The DNA from a Nelore bull registered in the Brazilian Association of Zebu Breeders (ABCZ) was used for the sequencing. Ninety-seven percent of genes encoding proteins of bulls were covered (CANAVEZ *et al.,* 2011), and as expected, the bull/taurine and zebu genomes showed high nucleotide similarity in the autosomes and the X chromosome, which did not occur for Y, which is still incomplete for taurine.

In 2012, the genomes of three breeds well adapted to the tropical climate, Brahman, Africander and Tuli Bull were sequenced. The platform used for sequencing was Illumina GA-II (Illumina Inc., USA), where each genome was covered more than 6x, between 86.7% and 88.8% of the bases of each genome were covered by one or more readings, identifying a total of 3.56 million SNPs.

Also in 2012, the NGS technology was used to sequence seven cattle from four breeds: Angus, Holstein, Hereford and Nellore. However, the main objective of the study was the discovery of CNVs (BICKHART *et al.,* 2012).

The Illumina *BovineSNP50* V2 *BeadChip,* still commercially available, has 54,609 highly informative SNPs and evenly distributed throughout the bovine genome. In addition, Illumina provides the high density BovineHD chip (777,000 SNPS) and the low density BovineLD v1.1 (6,912 SNPs), the latter is a tool for imputation and genotyping of the whole herd (ILLUMINA, 2013). The company Neogen also offers low (10,000 SNPs) and medium (80,000 SNPs) density chips specific for zebu and taurine (NEOGEN, 2013).

Pig genome

In order to coordinate the sequencing of the pig genome, the Swine Genome Sequencing Consortium (GFCF) was formed in September 2003 by academic, government and industry representatives. In 2005 the strategy to be adopted was published. The rationale for sequencing the genome of this species lies in its outstanding importance for the production of animal protein and use as model organism, since its physiology is very close to human (SGSC, 2013).

In November 2009, a high quality draft sequence of the pig *(Sus scrofa domesticus)* genome was assembled with 2.7Gb approximate size (ARCHIBALD *et al.,* 2010; BAI *et al.,* 2012). BAC clones from the *Chori-242* library, prepared from the DNA of a single Duroc female (Duroc 2-14), were preferentially selected for sequencing. However, some BACs with genome fragments of other races were also used. The initial goal was to determine the sequence of the pig genome with 3x coverage. However, both ends of 768 subclones of each BAC, with 707 bp average length were sequenced to provide approximate 4x coverage. The WGS data were generated using both methods, sequencing by chain termination (or Sanger) in capillary system at the Korea Livestock Research Institute, and the Solexa/Illumina at the Beijing Genomics Institute and the Wellcome Trust Sanger Institute (SCHOOK *et al., 2005;* ARCHIBALD *et al.,* 2010).

In 2009, Ramos and colleagues (RAMOS *et al.,* 2009), used the Illumina's Genome Analyzer to sequence the DNA from four groups of pigs: the Duroc, Large White, Pietrain breeds and wild boar. From this sequencing, approximately 272,000 SNPs were found.

In September 2011, the Consortium of Swine Genome Sequencing published a mixed assembly of the pig genome based on BAC and WGS, the Sscrofa 10.2. This map includes the set of chromosomes from 1 to 18, X and Y. The NCBI *Sus scrofa build* 4.1 included the *Sscrofa* 10.2 assembly and the complete mitochondrial genome derived from a Landrace pig (WELLCOME TRUST SANGER INSTITUTE, 2012).

The database for the pig genome continues to receive significant updates and new QTL is still being added to the *PigQTLdatabase.* Currently, this database has a total 8,315 QTL, cited in 355 publications, for 622 different characteristics of interest (NAGRP, 2013). This tool enables to find in swine chromosomes the most likely location for the genes responsible for important quantitative traits for production and reproduction.

More than 549,000 SNPs were used to create the *PorcineSNP60* (Illumina Inc., USA), a commercial chip with 64,621 SNPs.

Equine genome

The Horse Genome Project to sequence the horse genome began in the mid-90s, specifically in October 1995, in Lexington, Kentucky, USA. In January 1997, the international community that participate in this project met as part of a national survey sponsored by the United States Department of Agriculture (USDA) and in 2005, a petition showing that the equine genome research would help to understand the mechanisms of the human genome was presented to the National Human Genome Research Institute (NHGRI), (HORSE GENOME PROJECT, 2011).

The sequencing and organizing of the equine genome were conducted by the Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard University. A high quality draft with 6.8x coverage was produced from the genomic DNA of a thoroughbred English mare named *Twilight,* owned by Cornell University in Ithaca, New York, USA (NIH, 2007).

Approximately 300,000 BACs with sequenced ends were generated in Germany by the University of Veterinary Medicine in Hanover and the Helmholtz Centre for Infection Research in Braunschweig. Between January 2006 and January 2007, BACs were sequenced and ordered generating an initial map, the *EquCab*1.0. In September 2007, with the high quality draft sequence finished, the virtual map *EquCab*2.0

Table 1 - Current Situation of Livestock Genome Sequencing

Animal	Chrs (n)	Sequenced Extension	Sequencing Strategy	Platform	Coverage	Newest versions available	SNPs Commercial panels	publication of the first high quality sequencing	Coordination of sequencing projects
Chicken	33	1.2 Gb	Bacterial artificial chromosome (BAC), fosmid and plasmids based on WGS	-	6.6x	Gallus gallus 4.0	60k	*Nature* 09/12/2004	*Washington University Genome Sequencing Center*
Swine	20	2.8 Gb	Based on clones	Sanger, Illumina/Solexa	-	Sscrofa 10.2	60k	*BMC genomics* 19/07/2010	*Swine Genome Sequencing Consortium*
Bovine	30	2.67 Gb	Combination of ordered clones and WGS sequencing	-	9.5 x	UMD 3.1	3 to 778k	*Genome Biology* 24/3/2009	*Baylor College of Medicine* in Houston, Texas The genome was reassembled by Salzberg in Baltimore, Maryland
	31	2.98 Gb	Combined assembling of WGS 7.15x and BAC sequences	-	7.15x	BTAU 4.6.1		*Science* 24/4/2009	*Bovine Genome Sequencing Project* led by *Baylor College of Medicine's Human Genome Sequencing Center* in Houston, Texas
Ovine	27/27	2.6Gb/2.86Gb	WGS	454-GS-FLX, Illumina GAII	-	OAR 3.1/Ovis Aries 1.0	50k	*Animal Genetics* 30/8/2010	*International Sheep Genomics Consortium*
Equine	32	2.47Gb	WGS	Sanger, Illumina GAII	6.79x	EquCab 2.0	70k	*Science* 06/11/2009	*Broad Institute* and *Horse Genome Project*
Caprine	30	2.66Gb	Fosmid, WGS	Illumina GA-II/ HiSeq 2000	> 5x	CHIR 1.0	-	*Nature Biotechnology* 23/12/2012	*International Goat Genome Consortium*

Sources: Bai *et al.,* 2012; JGI 2012, UCSC Genome Browser, 2012; Animal Genome 2012; NCBI, 2013.

"Chrs" = fully or partially sequenced chromosomes, including autosomes, sexual and mitochondrial chromosomes; "-" = information not available

was completed. At the approximate cost of 15 million dollars, close to 2.7 Gb were sequenced, showing that the horse genome is smaller than the human genome and slightly larger than the domestic dog (WADE *et al.,* 2009).

The map containing the location of just over 940,000 SNPs distributed throughout the equine genome, except chromosome Y, is available on www.broad.mit.edu/mammals/horse/snp. This content has been compiled by the Broad Institute's Equine Genome Sequencing Project, for which information on the genomes of the breeds Arabic, Andalusian, Akhal-Teke, Iceland, Standardbred, Thoroughbred and Quarter Horses were used.

In 2012, the genome sequence of a Quarter Horse mare was published. The Illumina sequencing platform generated 59.6 Gb of sequenced DNA, with 24.7x cover. The sequences were mapped in 97% of the reference genome of a Thoroughbred horse. In addition, 3.1 million SNPs, 193,000 InDels and 282 CNVs were identified (DOAN *et al.,* 2012). The information of the *EquCab2.0* set was used by Illumina initially to build the *Equine SNP50 BeadChip* with 54,602 SNPs. This chip was discontinued in early 2011 due to the commercial release of the denser *Equine SNP70 BeadChip,* containing more than 74,000 SNPs evenly distributed throughout the genome (NEOGEN, 2013)

Sheep genome

Unlike the bovine genome project that was funded by the USDA, NIH and others (Bovine Genome Sequencing and Analysis Consortium, 2009), the sequencing of the sheep genome had limited resources.

The NGS technologies seemed to be the solution due to the capability of obtaining large amounts of sequences at a low cost. However, because these approaches were unable to generate readings of large segments, they were not immediately adopted by the consortium members. But, in 2009, the panda genome was sequenced and mapped solely from NGS technology with sequencing of fragments ends. Thus, the consortium agreed to start the sequencing of a reference map in 2009, during the workshop of the International Sheep Genome Consortium (ISGC, 2010).

The reference data generation phase for the sheep genome project was initiated in two sequencing centers. The Kunming Institute of Zoology, Yunnan and the Beijing Genomics Institute (BGI), Shenzhen, both in China, ordered the WGS sequencing of a Texel sheep, using the Illumina platform. Libraries with clones ranging from 170pb to 40kb were built and used to generate approximately 220Gb of sequences with paired ends and 75x coverage. Simultaneously, the *ARK-Genomics Centre for Comparative and Functional Genomics* at the Roslin Institute (Midlothian, Scotland) used a Texel male to produce 140Gb (45x coverage) of sequences with paired ends from clone libraries ranging from 200 to 500bp. Additionally, sequences were generated from the library of clones (3-8kb) with both ends sequenced, from the same Texel male. A first *de-novo* assembling using sequences derived from the Texel sheep is performed by readings of small sequences generating *contigs* and *scaffolds.* Information on the Texel male was used to complete the spaces, increase the number of identified SNPs and facilitate the analysis of the Y chromosome (ARCHIBALD *et al.,* 2010).

Comparisons with genomic information of dogs, cattle and humans have been made to refine the hybrid radiation map (HR) and to produce an integrated physical map. These comparisons allowed evaluating the quality of *contigs* and *scaffolds* (DARYMPLE *et al.,* 2007; WU *et al.,* 2009). To aid the characterization of components of genome transcription (genes), the BGI (Shenzhen, China) generated approximately 15Gb of sequences derived from seven different body tissues. This information along with datasets from NGS of other tissues was used to predict the total number of genes, enabling to observe their structures and positions in the sheep genome (ARCHIBALD *et al.,* 2010).

Sequencing information of 60 animals of 15 different breeds led to the discovery of 18,000 new SNPs, which together with other 37,000 SNPs obtained from the draft sequence (OARv1.0) were used to make the *Ovine SNP50 BeadChip* commercialized by Illumina (ILLUMINA, 2010). The most recent virtual map, the OARv3 is intended to fill gaps; more correctly assign sequences to chromosomes and provide data of 5,000,000 identified SNPs. With the same objectives of the OARv3.1, the OARv4.0 is being built and is due for release in the second half of 2013 (LIVESTOCK GENOMICS, 2012).

Goat genome

The International Goat Genome Consortium (IGGC) was created in March 2010 in Shenzhen City, China, and is coordinated by Chinese and French researchers. The initial activity of the consortium has focused on three areas, namely: goat genome sequencing and de-novo assembling; mapping and developing a radiation hybrid panel, and producing a high density SNPs chip. The assembly of the goat genome was performed at the Beijing Genomics Institute (BGI) (Ruminant Genome Biology Consortium, 2010).

At the end of 2012, a 2.66 Gb draft sequence of goat genome was published, i.e., 91% of the estimated 2.92 Gb. The DNA of a three-year-old *Yunnan Black Goat* female was used for the sequencing, performed by Illumina® GAII platform (Illumina Inc., USA). High quality readings (191.5 Gb) with 65x coverage were generated from fragments corresponding to two libraries: one with short sequences (180, 350 and 800pb); and, another with long sequences (2, 5, 10, and 20kb). Fosmid libraries were also created with fragments of sizes between 400 and 700pb, sequenced like the short sequence libraries described above. Furthermore, the *RNAseq* was performed from ten different organic tissues of the *Yunnan Black Goat* and from the follicular tissue of a Cashmere goat from Mongolia, using the Illumina HiSeq 2000 (Illumina

Inc., USA) platform. *Scaffolds* were built by aligning the *contigs* while increased *Scaffolds* were generated from sequenced fosmid ends. Because there is no physical map for goats, a new optical mapping technology called whole-genome mapping (WGM), developed by the group *OpGen* (OpGen, USA) was used to facilitate uniting the *Super Scaffolds.* These *Super Scaffolds* were anchored on chromosomes based on information of syntenic regions between goats and cattle. Finally, 22,175 coding genes were annotated, of which the majority was recovered from transcriptome data of the ten tissues used for RNA sequencing (DONG *et al.,* 2012).

FINAL THOUGHTS

In recent years, various methods for DNA sequencing have been developed and their widespread use has revolutionized genetic and genomic research while generating large quantities of information for several different species of interest to man. These techniques have been applied to domestic livestock, providing important tools for phylogenetic studies and animal breeding, as well as for the discovery and understanding of several biological processes.

The low cost and short time required by the new methods of sequencing made it faster to discover DNA polymorphisms. Accordingly, platforms for the detection of thousands to hundreds of thousands of variations in DNA (high-density SNP chips) have been developed and are routinely used in conjunction with statistical methods and increasingly appropriate computational tools, drastically changing the strategies normally used for animal breeding programs. As an example, we can mention the use of these tools in herds of dairy cattle breeds of European and North American countries, which use genomic selection to generate early genomic breeding values (EGBV) with high accuracy by summing the effects of markers scattered over the whole genome of the genotyped animal. Thus, increasing the genetic gain by decreasing the generation interval reduces drastically the costs of these programs.

Genome-Wide Association Studies (GWAS), with the use of SNP chips, have enabled the discovery of genes with great effect on some traits of economic interest, allowing the use of information of these markers in the selection and understanding the genetic mechanisms that govern such traits. In order to understand these genetic mechanisms, much has been done in the area of functional genomics, mainly by RNA sequencing or assays with *Microarrays.* These studies have shown changes in levels of gene expression related to different phenotypes or treatments, and enable the identification of genetic variants that may be directly responsible for part of the variation of a trait.

Certainly there is still much more to advance, especially in relation to major transformations arising from sequencing technologies. With further decreasing costs and increasing data generation speed, in a near future, new technological tools will be available, bringing ever more promising possibilities for the genetic study of domestic animals that were unthinkable a few years ago.

REFERENCES

ARCHIBALD, A. L.; BOLUND, L.; CHURCHER, C.; FREDHOLM, M.; GROENEN, M. A. M.; HARLIZIUS, B.; LEE, K. T.; MILAN, D.; ROGERS, J.; ROTHSCHILD, M. F.; UENISHI, H.; WANG, J.; SCHOOK, L. B. Pig genome sequence - analysis and publication strategy. **BMC Genomics**, v.11, n.438, p.1-5, 2010.

BAI, Y.; SARTOR, M.; CAVALCOLI, J. Current. status and future perspectives for sequencing livestock genomes. **Journal of Animal Science and Biotechnology**, v.3, n.1, p.3-8, 2012.

BARRIS, W. HARRISON, B.E. MCWILLIAM, S. BUNCH, R. J. GODDARD M. E. BARENDSE, W. Next generation sequencing of African and Indicine cattle to identify single nucleotide polymorphisms. **Animal Production Science**, v.52, n.3, p.133–142, 2012.

BAYLOR COLLEGE OF MEDICINE (BCM). **Genome Project Bovine**. Disponível em: <http://www.hgsc.bcm.tmc.edu/project-species-mBovine.hgsc?pageLocation=Bovine> acesso em: 21/06/2012.

BICKHART, D. M.; HOU, Y.; SCHROEDER, S. G. *et al*. Copy number variation of individual cattle genomes using next-generation sequencing. **Genome Research**, v.22, n.4, p.778-790, 2012.

BOARDMAN, P. E.; SANZ-EZQUERRO, J.; OVERTON, I. M.; BURT, D. W.; BOSCH, E.; FONG, W. T.; TICKLE, C.; BROWN, W. R.; WILSON, S. A.; HUBBARD, S. J. A comprehensive collection of chicken cDNAs. **Current Biology**, v.12, n.22, p.1965-1969, 2002.

BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM *et al*. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, v.324, n.5926, p.522–528, 2009.

CANAVEZ, F. C.; LUCHE, D. D.; STOTHARD, P.; LEITE R. M. K.; SOUSA-CANAVEZ, J. M.; PLASTOW, G.; MEIDANIS, J.; SOUZA, M. A.; FEIJÃO, P.; MOORE, S. S.; CAMARA-LOPES, L. H. Genome Sequence and Assembly of Bos indicus. Journal of Heridity, doi: 10.1093/jhered/esr153, 2012. Disponível em: <http://jhered.oxfordjournals.org/content/early/2012/02/06/jhered.esr153.full> acesso em 13/01/2013.

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas

aplicações na genômica de plantas. **Ciencia Rural**, v.40, n.3, p.735-744, 2010.

DALRYMPLE, B. P.; KIRKNESS, E. F.; NEFEDOV, M.; MCWILLIAM,, S.; RATNAKUMAR, A.; BARRIS, W.; ZHAO, S.; SHETTY, J.; MADDOX, J. F.; O'GRADY, M.; NICHOLAS, F.; CROWFORD, A. M.; SMITH, T.; JONG, P. J.; McEVAN, J.; ODDY, V. H.; COCKETT, N. E. and INTERNATIONAL SHEEP GENOMICS CONSORTIUM. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. **Genome Biology**, v.8, n.7, p.152.1-152.20, 2007.

DOAN, R.; COHEN, N. D.; SAWYER, J.; GHAFFARI, N.; JOHNSON, C. D.; DINDOT, S. V. Whole-Genome Sequencing and Genetic Variant Analysis of a Quarter Horse Mare. **BMC Genomics**, v.13, n.78, p.1471-2164, 2012.

DONG, Y.; XIE, M.; JIANG, Y.; XIAO, N.; DU, X.; ZHANG, W.; TOSSER-KLOPP, G.; WANG, J. YANG, S.; LIANG, J.; CHEN, W.; CHEN, J.; ZENG, P.; HOU, Y.; BIAN, C.; PAN, S.; LI, Y.; LIU, X.; WANG, W.; SERVIN, B.; SAYRE, B.; ZHU, B.; SWEENEY, D.; MOORE, R.; NIE, W.; SHEN, Y.; ZHAO, R.; ZHANG, G.; LI, J.; FARAUT, T.; WOMACK, J.; ZHANG, Y.; KIJAS, J.; COCKETT, N.; XU, X.; ZHAO, S.; WANG, J.; WANG, W. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). **Nature Biotechnology**, v.31, p.135-143, 2013.

FURLAN, L. R.; FERRAZ, A. L.; BORTOLOSSI, J. C. A genômica funcional no âmbito da produção animal: estado e perspectivas. **Revista Brasileira de Zootecnia**, v.36, suppl.0 2007.

GENOME CANADA. Bovine Genomics. Disponível em: < http://www.genomecanada.ca/en/portfolio/research/bovine.aspx > acesso em: 07/04/2013.

GROENEN, M. A. M.; MEGENS, H. J.; ZARE, Y.; WARREN, W. C.; HILLIER, L. W.; CROOIJMANS, R. P. M. A.; VEREIJKEN, A.; OKIMOTO, R.; MUIR, W. M.; CHENG,H.H. The development and characterization of a 60K SNP chip for chicken. **BMC genomics**, v.12, p.1-9, 2011.

HILLIER, L. W. MILLER, BIRNEY W. E. *et al*. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, v.432, n.7018, p.695-716, 2004.

HORSE GENOME PROJECT (HGP). 2011. Disponível em: < http://www.uky.edu/Ag/Horsemap/abthgp.html > acesso em: 21/02/2013

ILLUMINA. BovineHD Genotyping BeadChip, 2013 Disponível em: <http://www.illumina.com/documents//products/datasheets/datasheet_bovineHD.pdf > acesso em 13/03/2013.

ILLUMINA. BovineLD v1.1 Genotyping BeadChip, 2013. Disponível em: <http://res.illumina.com/documents/products/datasheets/datasheet_bovineld.pdf> acesso em 29/08/2013.

ILLUMINA. Ovine SNP50 Genotyping BeadChip, 2010. Disponível em: <http://res.illumina.com/documents/products/datasheets/datasheet_ovinesnp50.pdf> acesso em: 10/03/2013.

INTERNATIONAL SHEEP GENOMICS CONSORTIUM; ARCHIBALD, A. L.; COCKETT, N. E.; DALRYMPLE, B. P.; FARAUT, T.; KIJAS, J. W.; MADDOX, J. F.; MCEWAN, J. C.; HUTTON ODDY, V.; RAADSMA, H. W.; WADE, C.; WANG, J.; WANG, W.; XUN, X. The sheep genome reference sequence: a work in progress. **Animal Genetics**, v.41, n.5, p.449–453, 2010.

INTERNATIONAL SHEEP GENOME CONSORTIUM (ISGC). 2010. Disponível em: < http://www.sheephapmap.org/ > acesso em: 10/12/2012.

LIU G. E. Applications and case studies of the next-generation sequencing technologies in food, nutrition and agriculture. **Recent Patents on Food, Nutrition & Agriculture**, v.1, n.1, p.75–79, 2009.

LIVE STOCK GENOMICS (LSG). **OAR v2.0**. 2010. Disponível em: < http://www.livestockgenomics.csiro.au/sheep/oar2.0.php > acesso em: 24/01/2013.

LIVE STOCK GENOMICS (LSG). **OAR v3.1**. 2012. Disponível em: < http://www.livestockgenomics.csiro.au/sheep/oar3.1.php > acesso em: 23/11/2012.

MARGULIES, M.; EGHOLM, M.; ALTMAN, W. E.; ATTIYA, S.; BADER, J. S.; BEMBEN, L. A.; BERKA, J.; BRAVERMAN, M. S.; CHEN, Y. J.; CHEN, Z.; DEWELL, S. B.; DU, L.; FIERRO, J. M.; GOMES, X. V.; GODWIN, B. C.; HE, W.; HELGESEN, S.; HO, C. H.; IRZYK, G. P.; JANDO, S. C.; ALENQUER, M. L. I.; JARVIE, T. P.; JIRAGE, K. B.; KIM, J. B.; KNIGHT, J. R.; LANZA, J. R.; LEAMON, J. H.; LEFKOWITZ, S. M.; LEI, M.; LI, J.; LOHMAN, K. L.; LU, H.; MAKHIJANI, V. B.; MCDADE, K. E.; MCKENNA, M. P.; MYERS, E. W.; NICKERSON, E.; NOBILE, J. R.; PLANT, R.; PUC, B. P.; RONAN, M. T.; ROTH, G. T.; SARKIS, G. J.; SIMONS, J. F.; SIMPSON, J. W.; SRINIVASAN, M.; TARTARO, K. R.; TOMASZ, A.; VOGT, K. A.; VOLKMER, G. A.; WANG, S. H.; WANG, Y.; WEINER, M. P.; YU, P.; BEGLEY, R. F.; ROTHBERG, J. M. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v.437, p.376-380, 2005.

MAXAM, A. M.; GILBERT, W. A new method for sequencing DNA. **Proceedings of the National Academy of Science of the United States of America**, v.74, n.2, p.560-564, 1977.

MCEWAN, J. C.; PAYNE, G. M.; BRAUNING, R.; MCCULLOCH, A. F.; STANTON, JO-A.; ODDY, V. H.; NICHOLAS, F. W.; DALRYMPLE, B. P.; INTERNATIONAL SHEEP GENOMICS CONSORTIUM. Using Roche 454 Flx Sequencing For SNP Discovery In Sheep. In: Plant & Animal Genomes XVI Conference. 12 a 16 de Janeiro de 2008.Town & Country Convention Center. San Diego, CA. Disponível em: <http://www.intl-pag.org/16/abstracts/PAG16_P05l_543.html> acesso em: 12/02/2013.

NATIONAL ANIMAL GENOMES RESEARCH PROGRAM (NAGRP). **QTLdb chicken**. Disponível em: < http://www.animalgenome.org/cgi-bin/QTLdb/GG/index > acesso em: 12/01/2013.

NATIONAL ANIMAL GENOMES RESEARCH PROGRAM (NAGRP). **QTLdb Sus Scrofa**. Disponível em: < http://www.genome.iastate.edu/cgi-bin/QTLdb/SS/index > acesso em: 13/02/2013.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI). **SNPdb**. Rockville Pike, Bethesda MD, USA. [2012]. Disponível em: < http://www.ncbi.nlm.nih.gov/projects/SNP/ > acesso em 25/06/2012.

NATIONAL INSTITUTE OF HEALTS (NHI). **NHI News**. 2007. Disponível em: <http://www.nih.gov/news/pr/feb2007/nhgri-07.htm> acesso em: 23/03/2013.

NATIONAL HUMAN GENOME RESEARCH INSTITUTE. **NHI news advisory: chicken genome assembled.** 2012. Disponível em: <http://www.genome.gov/11510730> acesso em 23/11/2012.

NEOGEN. Agrigenomic Solutions for Dairy Cattle, 2013. Disponível em: <http://www.neogen.com/Agrigenomics/Dairy.html> acesso em: 15/08/2013.

NEOGEN. Agrigenomic Solutions for Beef Cattle, 2013. Disponível em: <http://www.neogen.com/Agrigenomics/Beef.html> acesso em 15/08/2013.

NEOGEN, Equine SNP70 BeadChip Whole Genome SNP profiling, 2013. Disponível em: <http://www.neogen.com/Agrigenomics/pdf/Slicks/EquineGenotypingFlyer.pdf> acesso em 10/07/2013.

PAREEK, C. S.; SMOCZYNSKI, R.; TRETYN, A. Sequencing technologies and genome sequencing. Journal of Applied Genetics, v.52, n.4, p.413–435, 2011.

RAMOS, A. M.; CROOIJMANS, R. P. M. A.; AFFARA, N. A.; AMARAL, A. J.; ARCHIBALD, A. L.; BEEVER, J. E.; BENDIXEN, C.; CHURCHER, C.; CLARK, R.; DEHAIS, P.; HANSEN, M. S.; HEDEGAARD, HU, J. Z.; KERSTENS, H. H.; LAW, A. S.; MEGENS, H. J.; MILAN, D.; NONNEMAN, D. J.; ROHRER, G. A.; ROTHSCHILD, M. F.; SMITH, T. P. L.; SCHNABEL, R. D.; VAN TASSELL, C. P.; TAYLOR, J. F.; WIEDMANN, R. T.; SCHOOK, L. B.; GROENEN, M. A. M. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. **Plos one**, v.4, n.8, p.1-13, 2009.

ROBERTS, J. R.; CARNEIRO, M. O.; SCHATZ, M. C. The advantages of SMRT sequencing. Genome Biology, v.14, n.8, p1-4, 2013.

RONAGHI, M.; UHLEN, M.; NYREN, P. A sequencing method based on real-time pyrophosphate. **Science**, v.281, n.5375, p.363–365, 1998.

RUMINANT GENOME BIOLOGY CONSORTIUM. **The International Goat Genome Consortium.** 2010**.** Disponível em : <http://www.ruminants.org/index.php?option=com_content&view=article&id=4:the-goat-champion-is-&catid=5&Itemid=2> acesso em 17/04/2013.

SANGER, F.; COULSON, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. **Journal of Molecular Biology**, v.94, n.3, p.441-448, 1975.

SCHADT, E. E.; TURNER, S.; KASARSKIS, A. A window into third-generation sequencing. **Humane Molecular Genetics**, v.19, n.2, p.227–240, 2010.

SCHOOK, L. B.; BEEVER, J. E.; ROGERS, J.; HUMPHRAY, S.; ARCHIBALD, A.; CHARDON, P.; MILAN, D.; ROHRER, G.; EVERSOLE K. Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. **Comparative and Functional Genomics**, v.6, n.4, p.251-255, 2005.

SWINE GENOME SEQUENCING CONSORTIUM (SGSC). SGSC mission and goals. Disponível em: < http://piggenome.org/ > acesso em: 23/11/2012.

UNIVERSETY OF CALIFORNIA SANTA CRUZ (UCSC). ChickenSNPdb. Santa Cruz - CA 2012. Disponvel em: < http://genome.ucsc.edu/cgi-in/hgGateway?org=Chicken&db=galGal3&hgsid=250839081 > acesso em: 03/10/2012.

VENTER, J. C., ADAMS; M. D.; MYERS, .E.; LI, P. W.; MURAL, R. J.; SUTTON, G. G.; SMITH, H. O.; et al. The sequence of the human genome. **Science**, v.291, n.5507, p.1304–1351, 2001.

WADE, C. M.; GIULOTTO, E.; SIGURDSSON, S.; ZOLI, M.; GNERRE, S.; IMSLAND, F.; LEAR, T. L.; ADELSON, D. L.; BAILEY, E.; BELLONE, R. R.; BLÖCKER, H.; DIST, O.; EDGAR, R. C.; GARBER, M.; LEEB, T.; MAUCELI, E.; MACLEOD, J. N.; PENEDO, M. C. T.; RAISON, J. M.; SHARPE, T.; VOGEL, L.; ANDERSSON5, D. F.; ANTCZAK, T.; BIAGI, M. M.; BINNS, B. P.; CHOWDHARY, S. J.; COLEMAN, J.; DELLA VALLE, G.; FRYC, S.; GUÉRIN, G.; HASEGAWA, T.; HILL, E. W.; JURKA, J.; KIIALAINEN, A.; LINDGREN, G.; LIU, J.; MAGNANI, E.; MICKELSON, J. R.; MURRAY, J.; NERGADZE, S. G.; ONOFRIO, R.; PEDRONI, S.; PIRAS, M. F.; RAUDSEPP, T.; ROCCHI, M.; RØED, K. H.; RYDER, O. A.; SEARLE, S.; SKOW, L.; SWINBURNE, J. E.; SYVÄNEN, A. C.; TOZAKI, T.; VALBERG, S. J.; VAUDIN, M.; WHITE, J. R.; ZODY, M. C.; BROAD INSTITUTE GENOME SEQUENCING PLATFORM; BROAD INSTITUTE WHOLE GENOME ASSEMBLY TEAM;. LANDER, E. S.; LINDBLAD-TOH, K. Genome sequence, comparative analysis, and population genetics of the domestic horse. **Science**, v.326, n.5954, p.865-867, 2009.

WELLCOME TRUST SANGER INSTITUTE (WTSI). Swine Genome Sequencing Project. Disponível em: < http://www.sanger.ac.uk/resources/downloads/othervertebrates/pig.html > acesso em 12/12/2012.

WU, C. H.; JIN, W.; NOMURA, K.; GOLDAMMER, T.; HADFIELD, T.; DALRYMPLE, B. P.; MCWILLIAM, S.; MADDOX, J. F.; COCKETT, N. E. A radiation hybrid comparative map of ovine chromosome 1 aligned to the virtual sheep genome. **Animal Genetics**, v.40, p435–8, 2009.

ZIMIN, A. V.; DELCHER, A. L.; FLOREA, L.; KELLEY, D. R.; SCHATZ, M. C.; PUIU, D.; HANRAHAN, F.; PERTEA, G.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; MARCAIS, G.; ROBERTS M, SUBRAMANIAN, P.; YORKE, J. A.; SALZBERG, S. L. A whole-genome assembly of the domestic cow, *Bos taurus*. **Genome Biology**, v.10, R42, 2009.