

# Genome Wide Association Studies, Next Generation Sequencing and Their Application in Animal Breeding and Genetics: A Review

Review Article

H. Khanzadeh<sup>1</sup>, N. Ghavi Hossein-Zadeh<sup>1\*</sup> and S. Ghowvati<sup>1</sup><sup>1</sup> Department of Animal Science, Faculty of Agricultural Science, University of Guilan, Rasht, Iran

Received on: 31 May 2019

Revised on: 5 Aug 2019

Accepted on: 15 Aug 2019

Online Published on: Sep 2020

\*Correspondence E-mail: [nhosseinzadeh@guilan.ac.ir](mailto:nhosseinzadeh@guilan.ac.ir)

© 2010 Copyright by Islamic Azad University, Rasht Branch, Rasht, Iran

Online version is available on: [www.ijas.ir](http://www.ijas.ir)

## ABSTRACT

Recently genetic studies have been revolutionized by next generation sequencing (NGS) technology, and it is expected that the use of this technology will largely eliminate defects in the methods of association studies. The NGS technology is becoming the premier tool in genetics. However, at the moment the use of this method is limited especially in the livestock due to high cost and computational problems. But it is expected that the development of sequencing and computing technologies and reducing the cost will have significant impacts on the livestock health and production. This study reviews the literature on genetic association studies, NGS technologies and their application in animal breeding.

**KEY WORDS** genome wide association studies, genomics, livestock, next generation sequencing.

## INTRODUCTION

Linkage analysis and association studies made significant progress in understanding the genetic basis of common phenotypes and complex diseases (McCarthy *et al.* 2008; Chen, 2011). Next generation sequencing (NGS) technology has dramatically increased the human ability for DNA sequencing (Londin *et al.* 2013). Genomic studies in farm animals will increase our understanding of the genetic basis of traits and their results will be used in breeding programs and reduce the occurrence of diseases and improve products' quality and production efficiency.

Sequence of farm animal genome is expected to have a significant effect on sustainable production of animals (Andersson, 2001; Bisht and Panda, 2014). The present review aimed to summarize the current knowledge about genome wide association studies (GWAS), NGS and their application in animal breeding.

## Genome wide association study (GWAS)

The ability to predict genetic risk factors for human disease and important economic traits in animals, such as growth rate and production, requires understanding of responsible genetic loci for the phenotypic and genetic architecture of traits (Korte and Farlow, 2013). Genetic association study is a statistical method to identify genes or loci regulating complex traits that utilizes linkage disequilibrium (LD) to connect phenotypic trait with genetic polymorphisms. All mapping methods fall mainly into two categories: studies of candidate genes and whole genome studies (Jiang, 2013). The candidate genes study examines the relationship between known genes and traits (Liu *et al.* 2008; Bisht and Panda, 2014). Compared to candidate genes and linkage analysis, GWAS is studying the entire genome with a systemic method for detecting susceptible genetic variants for diseases and complex traits (Hirschhorn and Daly, 2005; Huang, 2015).

In general, the identification of genetic variants associated with complex traits requires a large number of variants and samples (Huang, 2015). Genome-wide association studies (GWAS) is the study of genotyped single nucleotide polymorphisms (SNPs) in the genome and their association with phenotype (Zeng *et al.* 2015). Literatures contain numerous examples of GWAS that explain the genetic background of traits. The missing genotypes, genetic heterogeneity, low LD, effect size, low allele frequencies and genetic architecture of complex traits are a challenge for GWAS (Korte and Farlow, 2013). Also, major technical and analytical challenges remain with the GWAS including multiple test corrections and missing loci or blocks, low power to identify sites with low effect, risk of stratification finding, overestimation of haplotype effects, poor model fitting, insufficient sample size, low-density SNP coverage, bringing out rare variants and unknown copy number variation (CNV) effects (Kadarmideen, 2014) and not justification of genetic variance for complex traits (Manolio *et al.* 2009; Clarke and Cooper, 2010; Gibson, 2010; Kadarmideen, 2014).

#### Linkage disequilibrium

The GWAS relies on LD between SNPs and causative genes (Schmid and Bennewitz, 2017). LD is non-random association of alleles among different loci within a population. LD would be affected by different factors such as selection, mutation, migration, population structure and recombination rate (Zhu *et al.* 2013). The efficiency of quantitative trait locus (QTL) mapping studies, e.g. GWAS and marker-assisted selection (MAS), depends on the LD in population (Sellner *et al.* 2007). LD is an ideal parameter for diagnosing genetic association between markers and genes or casual loci for complex traits with high accuracy (Jiang, 2013).

#### Missing heritability

Missing heritability refers to a part of the genetic variance that cannot be interpreted by all significant single-nucleotide polymorphisms (SNPs). A significant proportion of the heritability is not justified by common genetic variants in GWAS (Manolio *et al.* 2009). Missing heritability theory hypothesized that unknown and missing variants may be in GWAS with big effect on phenotype, but their frequency is much lower than that identified by SNP chips (Huang, 2015). Likely the use of NGS data for GWAS will help to fix this problem. Sequencing enables detection of low frequency and rare variants with medium to high effect and expected at least part of the missing heritability justified with this technology (Feng, 2015). The application of sequencing technology for a large number of samples with the appropriate phenotype provides a great opportunity to

uncover the missing heritability and genetic architecture of complex traits (Luo *et al.* 2011).

Many reasons have been suggested for the missing heritability including: the large number of unknown variants with small effects, rare variants which are less diagnosed by available genotyping arrays and probably have great effects, structural variation and inappropriate calculation of common environment among relatives (Manolio *et al.* 2009).

#### Copy number variation (CNV)

CNV is an important source of genetic diversity that provides structural information in genomics (Hou *et al.* 2011). CNV refers to a change in the number of copies of a region of the genome (between one kb to several mb) (Henrichsen *et al.* 2009). CNVs' size is defined differently in various sources. CNVs are the result of DNA deletion, duplication, insertion and rearrangement. Because most CNVs contain gene coding regions and regulatory factors, they play an important role in gene expression regulation (Conrad *et al.* 2010). It has been confirmed that CNVs have a higher mutation rate than SNPs (Zogopoulos *et al.* 2007). CNVs can be considered as a significant part of genetic variation for diseases or traits (McCarroll, 2008). It has been confirmed that an overlap exists between the CNVs and genes, and there is a correlation between CNVs and genes expression levels and between CNVs and some clinical phenotypes (Stranger *et al.* 2007). GWAS is a good tool for simultaneous survey of SNPs and CNVs (McCarroll, 2008). This study could help to explain genetic variability and heritability.

In the past few years, considerable progress has been occurred in identifying the CNVs in domestic animals. In the future, the development of accurate tools for the detection of CNVs and their application in combination with QTL and gene expression data will be necessary to identify the impact of structural variation on many phenotypes (Clop *et al.* 2012).

#### Application of GWAS in animal breeding and genetics

The advent of genome sequencing, including GWAS, whole genome prediction (WGP) and genomic selection has changed the pattern of animal breeding (Kadarmideen, 2014).

The combination of allelic and phenotypic information through GWAS facilitates the discovery of genetic loci associated with important traits (D'Agostino and Tripodi, 2017). Improving genomic selection through GWAS enhances biological knowledge about trait expression, provides information on genetic architecture of quantitative traits and makes gene mapping as a hot topic in the genetics of livestock (Goddard *et al.* 2016).

The use of GWAS in animal breeding and genetics has expanded since the genome sequence of domestic animals was identified and a large number of SNPs were discovered through sequencing. A variety of commercial SNP chips are available for cattle, sheep, poultry, horses, dogs and pigs. Despite toddler use of GWAS in domestic animals, desirable results have been reported, particularly in the analysis of the quantitative traits mechanism. Now, SNP chips are widely used in GWAS to identify QTL for traits in domesticated animals (Zhang *et al.* 2012). The use of SNP arrays considerably affected the theory and practice of animal breeding and genetics, which will play important roles in the future (Fan *et al.* 2010). Much progress has been made in GWAS in domestic animals and some genes have been identified for important traits (Zhang *et al.* 2012). Compared to SNP chips, sequencing can provide almost all information about variants including SNPs, CNV, insertions and deletions. By reducing the cost of sequencing, it is possible that everyone in the community is sequenced and GWAS done with this technique (Zhang *et al.* 2012). Some recent literature on the application of GWAS in animal breeding and genetics is presented in Table 1.

### Next generation sequencing and GWAS

NGS technology allows rare variants to be studied. Also, NGS technology enables us to identify many variants including SNP and structural variation and search for rare variants (Chen, 2011). Human understanding about the genetic basis of diseases is expanding due to increased use of NGS. Perhaps the biggest success of NGS is the discovery of variants for rare diseases with Mendelian inheritance (Londin *et al.* 2013).

While chip-based GWAS progresses, sequencing technology is developing rapidly and cost of sequencing is decreasing (Feng, 2015). With the advent of whole-genome sequencing (WGS) technology and increasing the capacity to rare variants detection, it is expected that GWAS using WGS will provide more opportunities to explore variants with larger size and causal effect (Huang, 2015). Unlike chip-based GWAS, sequencing supplies the direct analysis of causal genes and variants rather than considering their linkage disequilibrium (Feng, 2015).

NGS technology has a significant impact on our ability to find variants related with diseases and traits (Edwards *et al.* 2014). With the progress in implementation and invention for sequencing the entire genome, new valves have been opened for the recognition of DNA building (Feuk *et al.* 2006).

### First-generation sequencing technology

The first-generation sequencing was the sequencing of bacteriophage phiX174 that was done in 1977 by Frederick

Sanger (Sanger, 1977). Sanger sequencing was the basis for modern methods of sequencing that are already in use (Gabaldón and Alioto, 2016).

### Second-generation sequencing technology

General principles of NGS are similar to capillary electrophoresis sequencing (Sanger) in which sequencing occurs by the synthesis, but in NGS sequencing, millions of fragments are simultaneously sequenced instead of sequencing a single fragment of DNA (Gabaldón and Alioto, 2016). Five hundred millions to billions bases of raw sequence can be generated in a single run of the second-generation sequencing platforms (Pareek *et al.* 2011).

Illumina (sequencing by synthesis), SoLID (sequencing by ligation), Roche (pyrosequencing chemistry) and Ion Torrent (semiconductor detection of H<sup>+</sup>) are second-generation sequencing techniques. All second-generation sequencing techniques rely on polymerase chain reaction (PCR) to amplify DNA. The major challenges of second-generation techniques are short reads which can be complicated in genome assembly and alignment algorithms (Pareek *et al.* 2011).

### Third-generation sequencing technology

Third-generation sequencing technologies have several features including: 1) capability of detecting a single nucleotide change based on new visual and electrical single-molecule techniques, 2) these methods do not require amplification by PCR, thereby reduce the sequencing time and cost, 3) reading length in this method is long (1000 bp to 50 kb) (Steinbock and Radenovic, 2015). The third-generation sequencing techniques are explained in detail as follows:

#### I. PacBio technique

The first NGS tool is PacBio technique which is known as single-molecule real-time sequencing (SMRT) and has been used since 2011 (Steinbock and Radenovic, 2015). This technique is provided by Pacific Bio Sciences Company and has higher reading length than second-generation sequencing technology (SGS). Highly interconnected assemblies in *de novo* sequencing projects using PacBio technique have the ability to eliminate gaps in the current reference assemblies and identify structural variation (SV) in the personal genome.

#### II. Helicos technique

Helicos single molecule sequencing technique provides a particular vision of the genome biology through direct sequencing of nucleic acids. The sample preparation is simple and does not require any composition or amplification by PCR, and DNA and RNA are directly hybridized within the cell.

**Table 1** Recent literature on genome wide association study in domestic animals

Study	Used chip	Species	No. of subjects	Trait
Wolc <i>et al.</i> (2012)	A custom Illumina SNP 42k Infinium chip	Chickens	2900	Egg weight and egg uniformity
Garcia-Gamez <i>et al.</i> (2012)	Illumina OvineSNP50 BeadChip	Sheep	1681	Milk production traits
Zhang <i>et al.</i> (2013)	Illumina OvineSNP50 BeadChip	Sheep	329	Growth and meat production traits
Spehar <i>et al.</i> (2015)	Illumina Bovine SNP50 Bead Chip	Cattle	182	Milk, fat and protein yield
Yuan <i>et al.</i> (2015)	Chicken 600 K SNP genotyping array	Chickens	1534	Feed intake and efficiency
Ibeagha-Awemu <i>et al.</i> (2016)	Genotyping-by-sequencing (GBS)	Cattle	1246	Milk traits
Wang <i>et al.</i> (2015b)	Specific-locus amplified fragment sequencing technology	Chickens	400	Carcass traits
Lai <i>et al.</i> (2016)	QIAamp DNA Mini Kit	Goat	34 (14 cases, 20 controls)	Litter size
Yodklaew <i>et al.</i> (2017)	GeneSeek Genomic Profiler low-density (1255 cows) and high-density (55 cows) bead chips	Cattle	1305	Milk yield, persistency and age at first calving
Pétille <i>et al.</i> (2017)	Sequencing on Illumina platform	Chickens	464	Performance traits
Martin <i>et al.</i> (2017)	Illumina GoatSNP50 BeadChip	Goat	2209	Milk traits
Mucha <i>et al.</i> (2018)	Illumina Caprine 50K BeadChip	Goat	2381	Milk yield and conformation
Welderufael <i>et al.</i> (2018)	Illumina BovineSNP50_v2	Cattle	993	Mastitis
Abdoli <i>et al.</i> (2018)	Illumina Ovine SNP50	Sheep	124	Prolificacy
Abdoli <i>et al.</i> (2019)	Illumina Ovine SNP50	Sheep	124	Composite reproductive traits

This eliminates many intermediate stages which may cause distortion or loss of the sample (Milos, 2010).

Helicos sequencing technique is not dependent on the PCR (Schuster, 2007; Blow, 2008; Arif *et al.* 2010). This method does not need to convert RNA to cDNA for RNA sequencing and provides a new perspective to broad and unbiased understanding of the transcriptome (Ozsolak *et al.* 2009; Arif *et al.* 2010). The Helicos reading length is about 800-1000 bp (Ku and Roukos, 2013). In this method, millions of DNA single molecules trapped in two flow cells. These strings serve as sample for sequencing by synthesis. Then polymerase and a fluorescent-labeled nucleotide are added. Polymerase catalyzes specific binding of fluorescent nucleotide sequences into complementary strands in all samples. Then strings are washed and free nucleotides going out. Binding of nucleotides is made and position of banded nucleotides recorded. Fluorescent groups separated from strands but connected nucleotides are remained. The process repeated for other nucleotides (A/T/C/G) (Blow, 2008; Arif *et al.* 2010).

### III. Nanopore technique

Single-molecule techniques used in the nanopore method allow further studies such as DNA-proteins and protein-protein interactions (Feng *et al.* 2015). The idea of using nanopore for DNA sequencing was introduced in the 1990s (Deamer and Akeson, 2000). Recently this method has attracted considerable attention, due to its fast sequencing, low cost, long read length (5 kb) and no need for amplification of DNA or connection of enzymes or modified nucleotides (Steinbock and Radenovic, 2015).

The main advantages of nanopore technique are very long readings, high throughput and low requirements. These features simplify the use of these techniques (Feng *et al.* 2015). The entire genome sequenced in about 15 minutes and with very low cost.

Nanopore sequencing is based on the principle that single molecule DNA can be detected by passing through a very small channel (Ku and Roukos, 2013). The steps involved in nanopore sequencing technique are double stranded DNA conversion to single-stranded DNA using a polymerase. This will slow down the movement of ssDNA through the nanopore. Nanopore has the property of constriction around the channels that allow the read of ssDNA sequences. Sequences of ssDNA translated during the passage of nanopore and produce the signal. Each level of signal represents a nucleotide and sequence of DNA is decoded by detecting these levels (Steinbock and Radenovic, 2015).

### Restrictions of sequencing technologies

NGS is becoming the premier tool in genetic diagnostics. However, concerns are raised about the complexity and volume of data for genome full sequence that may lead to inefficiency of interpretation method for the relationship between genetic variants and diseases (Goldstein *et al.* 2013). NGS technology can generate millions of genetic diversity that densely distribute in the genome (Luo *et al.* 2012). Therefore, this sequencing method generated large amounts of data. However, the current computational methods are not able to harness the full potential of genome and epigenome data from NGS.

Therefore, there is a need for new and upgraded tools and systems (Chaitankar *et al.* 2016). The biggest constraint of NGS is bioinformatics methods for storing and analyzing the data (Trapnell and Salzberg, 2009; Blaby-Haas and de Crécy-Lagard, 2011; DePristo *et al.* 2011; Hinchcliffe and Webster, 2011; Nielsen *et al.* 2011; Londin *et al.* 2013). On the other hand, rare variants with high volume, sequencing errors and missing data are important challenges for association test of NGS data. These challenges are largely affected Type I error rate and power of test for phenotype-genotype correlation (Luo *et al.* 2011).

Unlike the highly accurate genotypes of GWAS, deep sequencing produces millions of DNA short fragments which this process requires precise and effective statistical algorithms for genotype calling and mapping (Chen, 2011). Hundreds samples collected and thousands to millions variants are genotyped in the genome for GWAS (Chen, 2011; Risch, 2000). Therefore, weak design of experiments and sample collection can cause challenges in the subsequent analysis (Gabaldón and Alioto, 2016).

### Categories of sequencing projects

Genome sequencing projects can generally be divided into two categories: 1) *de novo* sequencing where the goal is to obtain a high quality sequence of genome that can be used as reference for species and 2) resequencing where a reference genome is available and goal is to determine the sequence variation map for individuals. These variations may include all or some of single nucleotide polymorphism, rare variants, simple somatic mutations, deletion and insertion, copy number variations and other structural variation (Gabaldón and Alioto, 2016).

*De novo* genome sequencing is sequencing of a new genome for which there is not reference sequence for alignment. Quality of coated *de novo* sequencing data depends on contig size and continuity and variety of sizes included in the library. Researchers can made high quality *de novo* using NGS readings and available assembly tools.

The *de novo* sequencing required depth is determined by several factors including the sequencing method and strategy, reading length, assembly method and the complexity or repetitive regions of the genome (Chen, 2011). Studies have shown that the required depth of sequencing to detect more SNPs and indels are 15X and 33X for homozygous and heterozygous genotypes, respectively (Bentley *et al.* 2008; Ajay *et al.* 2011; Gabaldón and Alioto, 2016).

### Whole-genome and whole-exome sequencing

Whole-genome sequencing (WGS) and whole-exome sequencing (WES) technologies are purposeful, powerful and relatively unbiased methods for the discovery of genetic variations (Chaitankar *et al.* 2016).

Instead of sequencing the entire genome, targeted sequencing of coding regions such as exome sequencing produces valuable results to identify disease-related variants (Londin *et al.* 2013).

In the WES, protein coding regions of the genome are selected and sequenced. This method can efficiently identify variants for a large range of applications such as population genetics, genetic diseases and cancer studies.

WGS provides a special opportunity for surveying genetic and somatic variations but now large amounts of data and high computational requirement are limited the use of WGS in routine biological and genetic studies. But WES focuses on sequencing of protein-coding regions (exons) and therefore produces low data (Chaitankar *et al.* 2016). WES covers about 1.5 percent of the human genome (Lander *et al.* 2001; Huang, 2015) and has low cost. In the past few years, WES has conducted on a larger scale than WGS due to the economic performance, while WGS can discover more variants for complex traits (Huang, 2015). Despite the clear advantages of WES, this method has shortcomings, such as CNV detection (Londin *et al.* 2013). Sequencing of exons is based on the idea that mutations affecting the phenotype are in coding regions of the genome. However, we have very little information about the distribution of functional variants (Goldstein *et al.* 2013). Only relying on the sequencing of exons cannot be a good way and the entire genome of affected individuals must be sequenced to find all effective variants (Londin *et al.* 2013).

### Application of genome sequencing in animal science

Genome sequencing can transform food security and sustainable agriculture including food safety, public, animals and plants health, reducing the risk of diseases and increasing the development in agriculture through the breeding of animals and plants (FAO, 2016). Farm animals are valuable resources and often used as a model in studies of physiology and pathology, duo to very similar reproductive physiology and nutrition system in farm animals and humans. Thus, farm animals are unique resource for human researches (Bisht and Panda, 2014). However, production of farm animals is more important because of the provision of food for human society. Development of genomics in animal is the outcome of genomics development in human as a result of genome sequencing projects (Kadarmideen, 2014). In the recent years, application of genomic evaluation for mapping small effect QTLs using many markers in dairy cattle can greatly increase reliability, especially for young animals (Bisht and Panda, 2014).

Use of NGS enables us to detect SNP in the genome and the development of SNP chips for wide evaluation of SNPs with desired phenotypes (Kranis *et al.* 2013; Pértille *et al.* 2016).

**Table 2** Recent literature on genome sequencing in livestock and poultry

Study	Method	Species	No. of subjects	Major findings
Eck <i>et al.</i> (2009)	Paired-end sequencing	Cattle	96 (48 Fleckvieh and 48 Braunvieh bulls)	Generate 24 gigabases of sequence with 7.4-fold sequence depth (in average)
Cosart <i>et al.</i> (2011)	Exome sequencing	Cattle	3	Sequencing of exons in the <i>Bos taurus</i> genome
Kerstens <i>et al.</i> (2011)	Paired-end sequencing	Chickens	25	Identification of hundreds of shared and divergent SVs in four commercial chicken lines
Bickhart <i>et al.</i> (2012)	Single- and paired-end libraries	Cattle	6	Identification of 1265 CNV regions that 476 of which (~38%) have not been reported previously
Fouts <i>et al.</i> (2012)	Sanger and 454 sequences of small sub-unit rRNA loci	Cattle	12	Interrogate microbial diversity in the rumen of 12 cows consuming a forage diet
Dong <i>et al.</i> (2013)	High-throughput whole-genome mapping	Goat	1	Report of ~2.66-Gb genome sequence of a female Yunnan black goat
De Donato <i>et al.</i> (2013)	Illumina HiSeq 2000	Cattle	47	Adapting genotyping-by sequencing used in plants for genotyping 47 cattles
Yi <i>et al.</i> (2014)	Paired-end sequencing	Chickens	12	Identifying 8,840 CNV regions covering 98.2 Mb and representing 9.4% of the chicken genome
Shin <i>et al.</i> (2014)	Pair-end sequence	Cattle	32 (10 Holsteins and 22 Hanwoo)	Detecting 6,811 deleted CNVs across the analyzed individuals
Wang <i>et al.</i> (2015a)	Whole-genome and RNA-seq	Chickens	11	Genomic imprinting is absent in the brains of 1-day-old chickens
Sartelet <i>et al.</i> (2015)	BovineHD BeadChip	Cattle	19 (4 cases and 15 controls)	Identifying 31 candidate mutations by resequencing the genome
Pértille <i>et al.</i> (2016)	CornellGBS	Chickens	464	Optimize of the CornellGBS (genotyping by sequence approach developed in plants) for its application in chickens

The primary chips have limited coverage on the genome and not cover effective SNPs completely. NGS technology is powerful enough to detect casual polymorphisms but its use in animal breeding is impractical due to high cost (Elshire *et al.* 2011; Glaubitz *et al.* 2014; Pértille *et al.* 2016), although sequencing cost have decreased sharply in the past decade. Also, the huge expansion happened in the capacity and performance of information technology that made it possible to store and transfer large volume of information (FAO, 2016). Along with the development of sequencing methods and reducing costs, the widespread use of whole genome sequencing is likely in animal breeding.

NGS leads to better understanding of their genome, transcriptome and epigenome in livestock (Sharma *et al.* 2017). Genotyping is becoming a common tool in the development of poultry breeding (Pértille *et al.* 2016). Many scientists use genomic information to identify genes associated with diseases in cattle, sheep, goats and horses and created disease-resistant animals (Bisht and Panda, 2014). Some industrialized countries use WGS in the field of food and prevent and control of animal diseases. Genome sequencing also uses for inspection of food imports and exports (FAO, 2016). The genome sequence for a number of species, including poultry, cattle, horses, pigs and chimps have been completed that support many developments in animal breeding (Bisht and Panda, 2014). Sequencing was used for all pathogens in prevention and control of zoonotic diseases (FAO, 2016).

Also, NGS were used to identify breed specific variants, signatures of selection and mutations in livestock (Sharma *et al.* 2017). Recent advances in genotyping and sequencing technologies have created fast evolution in beef cattle evaluation methods. As a result, new tools provided for effective production of high quality meat (Rolf *et al.* 2010). RNA sequencing is also widely used in farm animal species such as chicken that leads to understanding of animal development mechanisms and use in functional genomics (Dunislawska *et al.* 2017). The development of new techniques and software in this field has made it possible to design effective strategies to improve livestock breeding and intended purpose using this technology and precise understanding of genomic structure and study of relationship between genotype and phenotype. Recent literature on genome sequencing in livestock and poultry are summarized in Table 2.

## CONCLUSION

Linkage analysis and GWAS studies have major role in understanding the genetic basis of traits and diseases and there has been much success using them. However, these methods have defects such as lack of full justification of genetic variance in genetic association studies. NGS technology can partly overcome on defects of GWAS studies. Many studies in field of NGS are related to human research, especially relationship between the genotype and the occur-

rence of various diseases. Human research has always been the first, and has been a model for undertaking genotype-phenotype studies on animals. Thus it is expected that in the next few years NGS will have a significant impact on livestock production and health. Currently the use of NGS especially in animals has been limited due to high cost of WGS and the huge amount of data produced by these methods and computational problem of data. WES is a cost effective way (rather than WGS), however, the sequencing of all exons gives much less information and knowledge than sequencing the entire genome. Thus, despite conducting a few works using whole-genome association studies and NGS for improving efficiency and quality of animal products, high cost and then huge volume of information and computational problems are the most important limiting factors for the use of NGS technology in farm animals. It is expected that development of sequencing methods and reducing the cost of sequencing with the progress of hardware and computational methods have significant impact on animal breeding and genetics.

## REFERENCES

- Abdoli R., Mirhoseini S.Z., Ghavi Hossein-Zadeh N., Zamani P., Ferdosi M.H. and Gondro C. (2019). Genome-wide association study of four composite reproductive traits in Iranian fat-tailed sheep. *Reprod. Fertil. Dev.* **31**, 1127-1133.
- Abdoli R., Mirhoseini S., Ghavi Hossein-Zadeh N., Zamani P. and Gondro C. (2018). Genome wide association study to identify genomic regions affecting prolificacy in Lori Bakhtiari sheep. *Anim. Genet.* **49**(5), 488-491.
- Ajay S.S., Parker S.C., Aabaan H.O., Fajardo K.V.F. and Margulies E.H. (2011). Accurate and comprehensive sequencing of personal genomes. *Genome Res.* **21**(9), 1498-1505.
- Andersson L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat. Rev. Genet.* **2**(2), 130-138.
- Arif I.A., Bakir M.A., Khan H.A., Al Farhan A.H., Al Homaidan A.A., Bahkali A.H., Al Sadoon M. and Shobrak M. (2010). A brief review of molecular techniques to assess plant diversity. *Int. J. Mol. Sci.* **11**(5), 2079-2096.
- Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L. and Bignell H.R. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* **456**(7218), 53-59.
- Bickhart D.M., Hou Y., Schroeder S.G., Alkan C., Cardone M.F., Matukumalli L.K., Song J., Schnabel R.D., Ventura M. and Taylor J.F. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* **22**(4), 778-790.
- Bisht S.S. and Panda A.K. (2014). DNA Sequencing: Methods and Applications Pp. 11-23 in *Advances in Biotechnology*. I.R. Mamta Baunthiya and J. Saxena, Eds. Springer-Verlag Berlin Heidelberg Publisher, Berlin, Germany.
- Blaby-Haas C.E. and de Crécy-Lagard V. (2011). Mining high-throughput experimental data to link gene and function. *Trends Biotechnol.* **29**(4), 174-182.
- Blow N. (2008). DNA sequencing: Generation next-next. Nature Publishing Group, Berlin, Germany.
- Chaitankar V., Karakulah G., Ratnapriya R., Giuste F.O., Brooks M.J. and Swaroop A. (2016). Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. *Prog. Retin. Eye. Res.* **55**, 1-31.
- Chen W. (2011). Statistical methods and analysis in genome wide association studies and next-generation sequencing. Ph D. Thesis. the University of Michigan, Michigan, United State.
- Clarke A.J. and Cooper D.N. (2010). GWAS: Heritability missing in action? *European J. Hum. Genet.* **18**(8), 859-861.
- Clop A., Vidal O. and Amills M. (2012). Copy number variation in the genomes of domestic animals. *Anim. Genet.* **43**(5), 503-517.
- Conrad D.F., Pinto D., Redon R., Feuk L., Gokcumen O., Zhang Y., Aerts J., Andrews T.D., Barnes C. and Campbell P. (2010). Origins and functional impact of copy number variation in the human genome. *Nature.* **464**(7289), 704-712.
- Cosart T., Beja-Pereira A., Chen S., Ng S.B., Shendure J. and Luikart G. (2011). Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genom.* **12**(1), 347-355.
- D'Agostino N. and Tripodi P. (2017). NGS-based genotyping, high-throughput phenotyping and genome-wide association studies laid the foundations for next-generation breeding in horticultural crops. *Diversity.* **9**(3), 38-42.
- De Donato M., Peters S.O., Mitchell S.E., Hussain T. and Imumorin I.G. (2013). Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS One.* **8**(5), e62137.
- Deamer D.W. and Akeson M. (2000). Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends Biotechnol.* **18**(4), 147-151.
- DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., Del Angel G., Rivas M.A. and Hanna M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**(5), 491-498.
- Dong Y., Xie M., Jiang Y., Xiao N., Du X., Zhang W., Tosser-Klopp G., Wang J., Yang S. and Liang J. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* **31**(2), 135-142.
- Dunisławska A., Lachmańska J., Sławińska A. and Siwek M. (2017). Next generation sequencing in animal science-a review. *Anim. Sci. Pap. Rep.* **35**(3), 205-224.
- Eck S.H., Benet-Pagès A., Flisikowski K., Meitinger T., Fries R. and Strom T.M. (2009). Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biol.* **10**(8), 82-90.
- Edwards J.S., Atlas S.R., Wilson S.M., Cooper C.F., Luo L. and Stidley C.A. (2014). Integrated statistical and pathway app-

- approach to next-generation sequencing analysis: A family-based study of hypertension. *BMC Proc.* **8(1)**, 104-111.
- Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S. and Mitchell S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* **6(5)**, e19379.
- Fan B., Du Z., Gorbach D.M. and Rothschild M.F. (2010). Development and Application of high-density SNP Arrays in genomic studies of domestic animals. *Asian-Australasian J. Anim. Sci.* **23(7)**, 833-847.
- FAO. (2016). Application of Genome Sequencing for Sustainable Agriculture and Food Security. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.
- Feng S. (2015). Design and association methods for next-generation sequencing studies for quantitative traits. Ph D. Thesis. the University of Michigan, Michigan, United State.
- Feng Y., Zhang Y., Ying C., Wang D. and Du C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genom. Proteom. Bioinf.* **13(1)**, 4-16.
- Feuk L., Carson A.R. and Scherer S.W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* **7(2)**, 85-97.
- Fouts D.E., Szpakowski S., Purushe J., Torralba M., Waterman R.C., MacNeil M.D., Alexander L.J. and Nelson K.E. (2012). Next generation sequencing to define prokaryotic and fungal diversity in the bovine rumen. *PLoS One.* **7(11)**, e48289.
- Gabaldón T. and Alioto T.S. (2016). Whole-Genome Sequencing Recommendations Pp. 13-41 in Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing. A.M. Aransay, T. Lavín and L. José, Eds. Springer-Verlag Berlin Heidelberg Publisher, Berlin, Germany.
- García-Gómez E., Gutiérrez-Gil B., Sahana G., Sánchez J.P., Bayón Y. and Arranz J.J. (2012). GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PLoS One.* **7(10)**, e47782.
- Gibson G. (2010). Hints of hidden heritability in GWAS. *Nat. Genet.* **42(7)**, 558-560.
- Glaubitz J.C., Casstevens T.M., Lu F., Harriman J., Elshire R.J., Sun Q. and Buckler E.S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One.* **9(2)**, e90346.
- Goddard M.E., Kemper K.E., MacLeod I.M., Chamberlain A.J. and Hayes, B.J. (2016). Genetics of complex traits: Prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. Biol. Sci.* **283**, 20160569.
- Goldstein D.B., Allen A., Keebler J., Margulies E.H., Petrou S., Petrovski S. and Sunyaev S. (2013). Sequencing studies in human genetics: design and interpretation. *Nat. Rev. Genet.* **14(7)**, 460-470.
- Henrichsen C.N., Chaignat E. and Reymond A. (2009). Copy number variants, diseases and gene expression. *Hum. Mol. Gen.* **18(1)**, 1-8.
- Hinchcliffé M. and Webster P. (2011). In silico analysis of the exome for gene discovery. *Meth. Mol. Biol.* **760**, 109-128.
- Hirschhorn J.N. and Daly M.J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6(2)**, 95-108.
- Hou Y., Liu G.E., Bickhart D.M., Cardone M.F., Wang K., Kim E.S., Matukumalli L.K., Ventura M., Song J. and VanRaden P.M. (2011). Genomic characteristics of cattle copy number variations. *BMC Genom.* **12**, 127-138.
- Huang J. (2015). Whole-genome sequencing-based association studies of cardiovascular biomarkers. Ph D. Thesis. University of Cambridge, Cambridge, United Kingdom.
- Ibeagha-Awemu E.M., Peters S.O., Akwanji K.A., Imumorin I.G. and Zhao X. (2016). High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci. Rep.* **6**, 31109.
- Jiang N. (2013). Linkage disequilibrium based eQTL analysis and comparative evolutionary epigenetic regulation of gene transcription. Ph D. Thesis. University of Birmingham, Birmingham, United Kingdom.
- Kadarmideen H.N. (2014). Genomics to systems biology in animal and veterinary sciences: Progress, lessons and opportunities. *Livest. Sci.* **166**, 232-248.
- Kerstens H.H., Crooijmans R.P., Dibbits B.W., Vereijken A., Okimoto R. and Groenen M.A. (2011). Structural variation in the chicken genome identified by paired-end next-generation DNA sequencing of reduced representation libraries. *BMC Genom.* **12(1)**, 94-110.
- Korte A. and Farlow A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods.* **9(29)**, 29-38.
- Kranis A., Gheyas A.A., Boschiero C., Turner F., Yu L., Smith S., Talbot R., Pirani A., Brew F. and Kaiser P. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genom.* **14(1)**, 59-72.
- Ku C.S. and Roukos D.H. (2013). From next-generation sequencing to nanopore sequencing technology: Paving the way to personalized genomic medicine. *Expert. Rev. Med. Devices.* **10(1)**, 1-6.
- Lai F.N., Zhai H.L., Cheng M., Ma J.Y., Cheng S.F., Ge W., Zhang G.L., Wang J.J., Zhang R.Q. and Wang X. (2016). Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (*Capra hircus*). *Sci. Rep.* **6**, 38096-38107.
- Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M. and FitzHugh W. (2001). Initial sequencing and analysis of the human genome. *Nature.* **409(6822)**, 860-921.
- Liu X., Zhang H., Li H., Li N., Zhang Y., Zhang Q., Wang S., Wang Q. and Wang H. (2008). Fine-mapping quantitative trait loci for body weight and abdominal fat traits: Effects of marker density and sample size. *Poult. Sci.* **87(7)**, 1314-1319.
- Londin E., Yadav P., Surrey S., Kricka L.J. and Fortina P. (2013). Use of linkage analysis, genome-wide association studies, and next-generation sequencing in the identification of disease-causing mutations. *Meth. Mol. Biol.* **1015**, 127-146.
- Luo L., Boerwinkle E. and Xiong M. (2011). Association studies for next-generation sequencing. *Genome Res.* **21(7)**, 1099-1108.
- Luo L., Zhu Y. and Xiong M. (2012). Quantitative trait locus analysis for next-generation sequencing with the functional



- linear models. *J. Med. Genet.* **49(8)**, 513-524.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R. and Chakravarti A. (2009). Finding the missing heritability of complex diseases. *Nature.* **461(7265)**, 747-753.
- Martin P., Pailhère I., Maroteau C., Bardou P., Canale-Tabet K., Sarry J., Woloszyn F., Bertrand-Michel J., Racke I. and Besir H. (2017). A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Sci. Rep.* **7(1)**, 1872-1880.
- McCarroll S.A. (2008). Extending genome-wide association studies to copy-number variation. *Hum. Mol. Gen.* **17(2)**, 35-42.
- McCarthy M.I., Abecasis G.R., Cardon L.R., Goldstein D.B., Little J., Ioannidis J.P. and Hirschhorn J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9(5)**, 356-369.
- Milos P.M. (2010). Helicos single molecule sequencing: Unique capabilities and importance for molecular diagnostics. *Genome Biol.* **11(1)**, 14.
- Mucha S., Mrode R., Coffey M., Kizilaslan M., Desire S. and Conington J. (2018). Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. *J. Dairy Sci.* **101(3)**, 2213-2225.
- Nielsen R., Paul J.S., Albrechtsen A. and Song Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **12(6)**, 443-451.
- Ozsolak F., Platt A.R., Jones D.R., Reifengerger J.G., Sass L.E., McInerney P., Thompson J.F., Bowers J., Jarosz M. and Milos P.M. (2009). Direct RNA sequencing. *Nature.* **461**, 814-814.
- Pareek C.S., Smoczynski R. and Tretyn A. (2011). Sequencing technologies and genome sequencing. *J. Appl. Genet.* **52**, 413-435.
- Pértille F., Guerrero-Bosagna C., Da Silva V.H., Boschiero C., da Silva Nunes J.R., Ledur M.C., Jensen P. and Coutinho L.L. (2016). High-throughput and cost-effective chicken genotyping using next-generation sequencing. *Sci. Rep.* **6**, 26929-26941.
- Pértille F., Moreira G.C.M., Zanella R., da Silva Nunes J.d.R., Boschiero C., Rovadoscki G.A., Mourão G.B., Ledur M.C. and Coutinho L.L. (2017). Genome-wide association study for performance traits in chickens using genotype by sequencing approach. *Sci. Rep.* **7**, 41748-41758.
- Risch N.J. (2000). Searching for genetic determinants in the new millennium. *Nature.* **405(6788)**, 847-856.
- Rolf M.M., McKay S.D., McClure M.C., Decker J.E., Taxis T.M., Chapple R.H., Vasco D.A., Gregg S.J., Kim J.W. and Schnabel R.D. (2010). How the next generation of genetic technologies will impact beef cattle selection. Pp. In Proc. 42<sup>nd</sup> Ann. Res. Symp. Ann. Meet., Columbia, United State.
- Sanger F. (1977). Nucleotide sequence of bacteriophage (D X174) DNA. *Nature.* **265(5596)**, 687-95.
- Sartelet A., Li W., Pailhoux E., Richard C., Tamma N., Karim L., Fasquelle C., Druet T., Coppieters W. and Georges M. (2015). Genome-wide next-generation DNA and RNA sequencing reveals a mutation that perturbs splicing of the phosphatidylinositol glycan anchor biosynthesis class H gene (PIGH) and causes arthrogryposis in Belgian Blue cattle. *BMC Genom.* **16(1)**, 316-324.
- Schmid M. and Bennewitz J. (2017). Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs. *Arch. Anim. Breed.* **60**, 335-346.
- Schuster S.C. (2007). Next-generation sequencing transforms today's biology. *Nat. Methods.* **5(1)**, 16-24.
- Sellner E., Kim J., McClure M., Taylor K., Schnabel R. and Taylor J. (2007). Board-invited review: Applications of genomic information in livestock. *J. Anim. Sci.* **85(12)**, 3148-3158.
- Sharma A., Park J.E., Chai H.H., Jang J.W., Lee S.H. and Lim D. (2017). Next generation sequencing in livestock species: A review. *J. Anim. Breed. Genom.* **1(1)**, 23-30.
- Shin D.H., Lee H.J., Cho S., Kim H.J., Hwang J.Y., Lee C.K., Jeong J., Yoon D. and Kim H. (2014). Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genom.* **15(1)**, 240-248.
- Špehar M., Mrak V., Smatko A., Potočnik K. and Gorjanc G. (2015). Genome-wide association study for dairy traits in Slovenian Brown swiss breed. *Slovenian Vet. Res.* **52(2)**, 49-55.
- Steinbock L. and Radenovic A. (2015). The emergence of nanopores in next-generation sequencing. *Nanotechnology.* **26(7)**, 074003.
- Stranger B.E., Forrest M.S., Dunning M., Ingle C.E., Beazley C., Thorne N., Redon R., Bird C.P., de Grassi A. and Lee C. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science.* **315(5813)**, 848-853.
- Trappnell C. and Salzberg S.L. (2009). How to map billions of short reads onto genomes. *Nat. Biotechnol.* **27(5)**, 455-457.
- Wang Q., Li K., Zhang D., Li J., Xu G., Zheng J., Yang N. and Qu L. (2015a). Next-generation sequencing techniques reveal that genomic imprinting is absent in day-old Gallus gallus domesticus brains. *PLoS One.* **10(7)**, e0132345.
- Wang W., Zhang T., Wang J., Zhang G., Wang Y., Zhang Y., Zhang J., Li G., Xue Q. and Han K. (2015b). Genome-wide association study of 8 carcass traits in Jinghai Yellow chickens using specific-locus amplified fragment sequencing technology. *Poult. Sci.* **95(3)**, 500-506.
- Welderufael B.G., Løvendahl P., De Koning D.J., Janss L. and Fikse F. (2018). Genome-wide association study for susceptibility to-and recoverability from mastitis in Danish Holstein cows. *Front. Genet.* **9**, 141-147.
- Wolc A., Arango J., Settler P., Fulton J., Osullivan N., Preisinger R., Habier D., Fernando R., Garrick D. and Hill W. (2012). Genome wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim. Genet.* **43**, 87-96.
- Yi G., Qu L., Liu J., Yan Y., Xu G. and Yang N. (2014). Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genom.* **15(1)**, 962-971.
- Yodklaew P., Koonawootrittriron S., Elzo M.A., Suwanasopee T.

- and Laodim T. (2017). Genome-wide association study for lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population. *Agric. Nat. Res.* **51(3)**, 223-230.
- Yuan J., Wang K., Yi G., Ma M., Dou T., Sun C., Qu L.J., Shen M., Qu L. and Yang N. (2015). Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. *Genet. Sel. Evol.* **47(1)**, 82-91.
- Zeng P., Zhao Y., Qian C., Zhang L., Zhang R., Gou J., Liu J., Liu L. and Chen F. (2015). Statistical analysis for genome-wide association study. *J. Biomed. Res.* **29(4)**, 285-292.
- Zhang H., Wang Z., Wang S. and Li H. (2012). Progress of genome wide association study in domestic animals. *J. Anim. Sci. Biotechnol.* **3(1)**, 26-32.
- Zhang L., Liu J., Zhao F., Ren H., Xu L., Lu J., Zhang S., Zhang X., Wei C. and Lu G. (2013). Genome-wide association studies for growth and meat production traits in sheep. *PLoS One.* **8(6)**, e66569.
- Zhu M., Zhu B., Wang Y., Wu Y., Xu L., Guo L., Yuan Z., Zhang L., Gao X. and Gao H. (2013). Linkage disequilibrium estimation of Chinese beef Simmental cattle using high-density SNP panels. *Asian-Australasian J. Anim. Sci.* **26(6)**, 772-779.
- Zogopoulos G., Ha K.C., Naqib F., Moore S., Kim H., Montpetit A., Robidoux F., Laflamme P., Cotterchio M. and Greenwood C. (2007). Germ-line DNA copy number variation frequencies in a large North American population. *Hum. Genet.* **122(3)**, 345-353.
-