OPEN ACCESS

Next generation sequencing in livestock species- A Review

Aditi Sharma[1], Jong-Eun Park[1], Han-Ha Chai[1], Gul-Won Jang[1], Seung-Hwan Lee[2] and Dajeong Lim[1*]

[1]Animal Genomics & Bioinformatics Division, National Institute of Animal science, RDA, Jeollabuk-do, Republic of Korea
[2]Division of animal and dairy science, Chungnam National University, Daejeon 34134, Republic of Korea

ABSTRACT

Technological advances in molecular biology during the last decade have opened up possibilities to rapidly and accurately generate large-scale sequencing data from non-model organisms at an affordable cost. Next generation sequencing (NGS) had led us to a better understanding of genome organization, structure, function and evolution in livestock animals. NGS provides a high resolution view of the DNA/RNA sequence which is a distinct advantage over other methods. NGS is a first step toward understanding the genetic mechanisms of an animal's functions and its interaction with the environment. It has been utilized widely now to study complex traits in different species. NGS is expected to bring down the overall cost of the animal production, increase the yield, improve the quality of meat and milk, provide better disease resistance and improve reproductive health of livestock. In this paper we review the applications of NGS in livestock animals.

Keywords: Next generation sequencing, livestock, genomics, transcriptomics

Background

Next generation sequencing (NGS) is a term used for massively parallel sequencing technology that was developed after the Sanger and the Maxam and Gilbert chemical degradation sequencing method. In NGS technology the desired molecule to be sequenced is broken into small pieces, which are then ligated to adapters for random reading. Since the template is broken into many smaller pieces its read length is generally smaller than the Sanger sequencing method. However some latest methods like single molecule real time (SMRT) and Oxford Nanopore have overcome that drawback. These technologies have longer reads, and thus are more accurate to generate the consensus sequence. Due to its affordability NGS has now become a common tool of use in several fields of biological sciences. NGS generates large amounts of genomic data that can be used to detect genetic variants related to functional alterations. Single Nucleotide polymorphisms (SNPs) are the most abundant type of molecular markers and their high density facilitates interrogation by different genetic approaches. These include large-scale genome association analyses, genetic analysis of simple and complex disease states, genomic predictions and population genetic studies. The use of NGS has enabled to identify SNPs across genomes and allowed the development of pre-designed SNP chips for widespread testing of SNP associations with specific phenotypes of interest.

NGS has led to characterization and quantification of a whole range of "*omics*" like genomics, transcriptomics and Epigenomics. *Omics* are essential to understand the mechanisms and functions of different molecules. Different type of NGS

sequencing can be used depending upon objective of the project (Figure1). Several livestock genomes have been sequenced recently using NGS. Among livestock species *Bos taurus* has been the most highly sequenced species followed by *Sus scrofa* (Table 1). Fast and accurate acquisition of the genome sequence has led to genome-wide identification of causal common and rare variants. Identification of these candidate mutations has enabled the researchers to address phenotypic diversity among livestock species and breeds. Taking advantage of NGS data Sharma et al. (2017) identified 18 mutations involved in Mendelian diseases in Hanwoo cattle of Korea. This information could further be used in a customized SNP chip for this breed. It has also allowed the researchers to study copy number variations (CNVs), genes involved in different pathways and metagenomics. RNA-seq is another popular approach to quantify the expression of genes involved metabolic pathways (Salleh et al., 2017). It has also been used to identify splice variants accurately by mapping sequence fragments onto a reference genome (Suarez-Vega, et al., 2017).
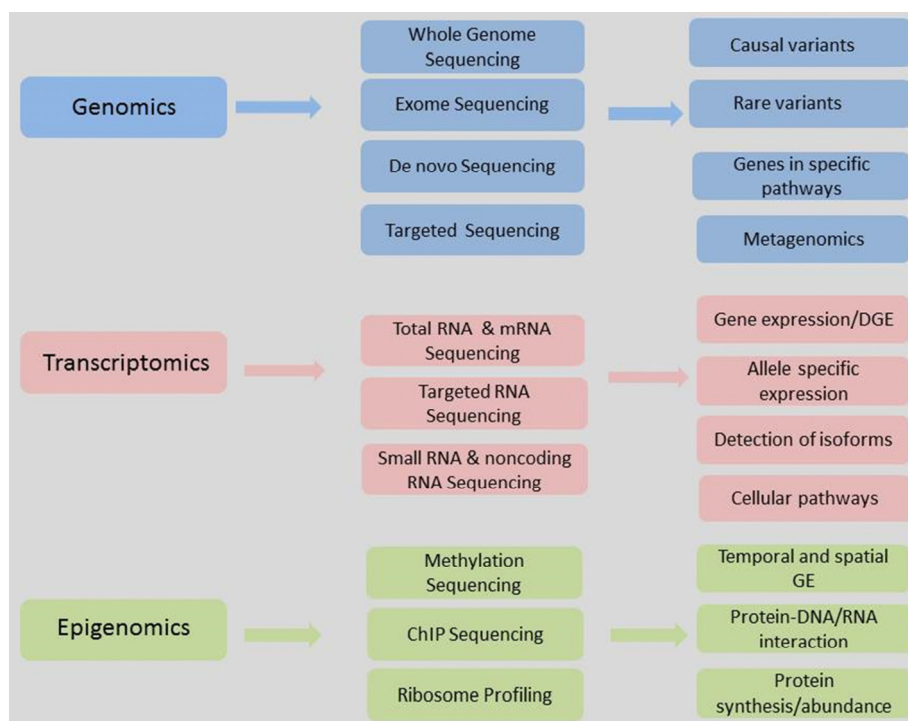
Figure 1. Different types of sequencing and their use in genomics, transcriptomics and epigenomics.

Table 1. Status of next generation sequencing in livestock (www.ncbi.nlm.nih.gov)

Species	SRA Experiments[#]
Dog (Canis familiaris)	3751
Chicken (Gallus gallus)	3637
Horse (Equus caballus)	3495
Cattle (Bos taurus)	12589
Pig (Sus scrofa)	6270
Rabbit (Oryctolagus cuniculus)	741
Sheep (Ovis aries)	2417
Duck (Anas platyrhynchos)	570

[#]The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms in NCBI

NGS has also made it possible to study genome-wide epigenetic modifications. Epigenetics may provide information about heritability of complex traits and diseases, imprinting and silencing of transposons which could be of much help in animal breeding (Triantaphyllopoulos et al., 2016). Epigenetic modifications are mediated by small RNAs (sRNAs). Studying methylome allows us to study relationship between sRNA and DNA methylation. NGS based methylome analysis provides a better understanding of methylation patterns across the genome. A better understanding of DNA methylation and other epigenetic modifications will help us establish a relationship among cellular, molecular, physiological and immune responses that play a role in disease resistance.

Studies have demonstrated the value of NGS technologies for molecular characterization, ranging from metagenomic characterization of unknown pathogens or microbial communities to molecular epidemiology and evolution of viral quasispecies (Jose et al., 2017, Yang et al., 2016). Moreover, high-throughput technologies now allow detailed studies of host-pathogen interactions at the level of their genomes (genomics), transcriptomes (transcriptomics), or proteomes (proteomics). The application of high-throughput NGS platforms and their typical low-cost per information content has revolutionized the resolution with which these processes can now be studied. In this paper we review the applications and impact of NGS on livestock species.

Role of NGS in Livestock Diseases and Other Complex Traits

Next generation sequencing of livestock species had allowed a better understanding of their genome, transcriptome and epigenome. Among the livestock species dog was the first livestock animal to be sequenced in the year 2005 (Table 2). Dog has been a loyal companion to humans for thousands of years now. Due to human influence dogs have evolved into several different breeds ranging from difference in size, shape, color and behavior. This human selection had also led to various health issues in these animals. With the use of NGS data researchers have been able to identify key mutations involved in several dog diseases like Lundehund syndrome (LS), a severe gastro-enteropathic disease in the Lundehund dog. NGS pointed towards the association signal on CFA 34 for the LS disease (Metzger et al., 2016). In Golden Retrievers GWAS-guided fine mapping by targeted-NGS has identified novel mutation associated with Generalized progressive retinal atrophy (Downs et al., 2014) and in Tibetian Spaniels/Terriers Downs & Mellersh (2014) identified a short interspersed nuclear element insertion that was associated with Progressive retinal atrophy (PRA). Exome-sequencing has identified the CNGB1 mutation associated with PRA in Papillon and Phalene dog (Ahonen et al., 2013).

Table 2. Details of the first livestock whole genome sequence assemblies deposited in NCBI

Organism	Breed	Sex	Year	Total sequence length	Accession
Dog	Boxer	Female	2005	2,52,84,30,226	GCF_000002285.1
Chicken	Red jungle Fowl	Female	2006	1,09,87,54,166	GCF_000002315.1
Horse	thoroughbred	Female	2007	2,47,49,29,062	GCA_000002305.1
Cattle	Hereford	Pooled male and female	2009	2,87,49,56,119	GCA_000003055.1
Pig	Mixed	-	2009	2,26,25,79,801	GCA_000003025.1
Rabbit	Thorbecke inbred	Female	2009	2,73,74,62,810	GCA_000003625.1
Sheep	Texel	Female	2012	2,61,90,54,388	GCF_000298735.1
Duck	Pekin duck	Female	2013	1,10,50,52,351	GCF_000355885.1

Close physical contact of the dogs with the humans puts humans at risk of certain diseases as well. Studying molecular mechanism of the zoonotic transmission from domestic animals to their humans will help address such public health concerns

(Oh et al., 2015; Meinel et al., 2014). Comparison of oral microbiomes of dogs and their humans will provide the much needed information about the transmission of any microorganisms which might lead to human diseases.

The data obtained from next generation sequencing has many applications. One amongst them is the identification of the actual expression level of all of the genes that are expressed in essentially any tissue. RNA sequencing (RNA-Seq) allows the quantification of the gene expression for any tissue between two samples. In Horse, Illumina Next Generation Sequencing (NGS) technology was used to identify and characterize the global miRNA expression profile in normal tissues. MiRNAs are important as they provide an insight into various physiological and pathological conditions. Kim et al. (2014) identified a total of 292 known and 329 novel miRNAs in normal horse tissues including skeletal muscle, colon and liver. NGS has also been useful in studying Horse chromosome rearrangement and karyotype evolution (Huang et al., 2014). Diseases like Equine grass sickness and amoebic placentitis were also studied using NGS. NGS provides important opportunities to tackle problems associated with pathogenic illnesses. Based on NGS the identity of the etiological agent for amoebic placentitis in a mare from eastern Australia was confirmed as *Acanthamoeba hatchetti* (Begg et al., 2014).

NGS has also allowed CNV detection which has opened new avenues for studying genes associated with complex traits in livestock. In Holstein bulls with extremely high and low estimated breeding values (EBVs) for milk protein percentage and fat percentage whole-genome resequencing data identified a total of 14,821 CNVs and 487 differential CNVRs. In addition, 10 genes (INS, IGF2, FOXO3, TH, SCD5, GALNT18, GALNT16, ART3, SNCA and WNT7A), were identified as candidate genes for milk protein and fat traits. In another such study, in Korean Hanwoo cattle a total of 6,811 deleted CNVs were identified using Hiseq 2000 (Illumina, Inc) sequencing data. 33 genes that had high deletion scores were identified to be involved in the domestication process. Their genetic functions were found to be related to nervous transmission, neuron motion and neurogenesis. These genes and the nervous system may be associated with the changes in behavior due to domestication (Shin et. al, 2014).CNVs are known to affect a wide range of phenotypic traits and CNVs in or near segmental duplication regions are difficult to track. However read depth approach based on next-generation sequencing had made it possible to detect such CNVs. Bickhart et al. (2012) used NGS to provide the first individualized cattle CNV and segmental duplication maps and genome-wide gene copy number estimates. A comparative analysis between taurine and indicine cattle breeds was made. It was found that the genes related to pathogen- and parasite-resistance, such as CATHL4 and ULBP17, were highly duplicated in the Nelore cattle relative to the taurine cattle, while genes involved in lipid transport and metabolism, including APOL3 and FABP2, were highly duplicated in the taurine breeds (Beef cattle). These CNV regions harbored genes like BPIFA2A (BSP30A) and WC1, suggesting that some CNVs may be associated with breed-specific differences in adaptation, health and production traits. A similar study was performed in Meishan pigs where segmental duplication (SD) map for pigs was constructed. Genome-wide CNV hotspots were found which were significantly enriched in SD regions, suggesting evolution of CNV hotspots to be affected by ancestral SDs. It was also found that the CNV-related and CNV-unrelated genes undergo a different selective constraint and CNVs may be associated with or affect pig health and production performance under recent selection (Jiang et al., 2014). Such information is of much help in the studies where pig is used as a biomedical model to study human diseases.

In dairy and meat type animals, increase in production and betterment of the quality of the produce is an active area of research. Transcriptomics data can facilitate the functional studies where high and low producing animals can be compared and differentially expressed genes could be identified. Further their metabolic pathways could be identified. All this information could be incorporated in breeding programs. Chen et.al (2015) sequenced and characterized divergent marbling levels in FLW beef cattle. RNA-seq data from the *Longissimus dorsi* muscle was used to identify the genes that were expressed in low and high marbling animals along with differentially expressed genes.

Recently Bovo et al. (2017) demonstrated the potential of NGS dataset mining for viral metagenomics analysis in livestock.
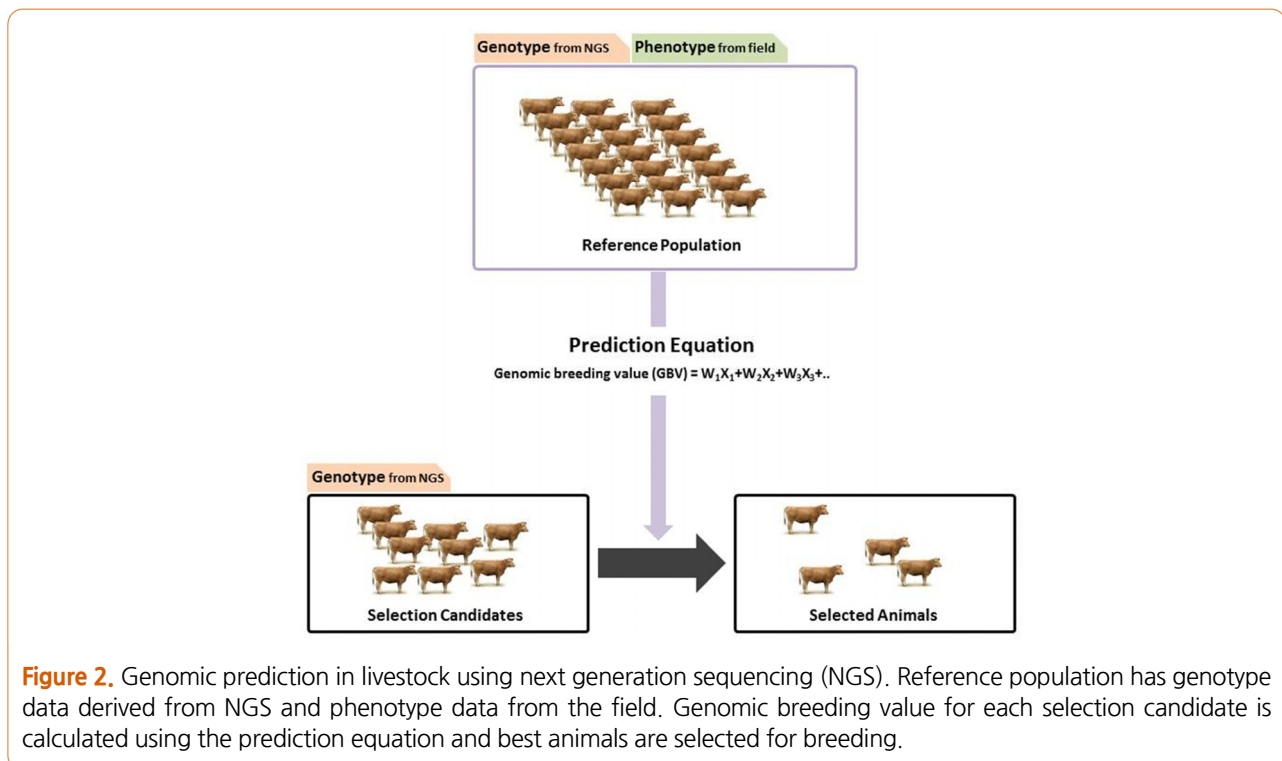
In usual practice the unmapped reads from the sequencing projects are discarded as by-products. But Bovo et al mined these reads in 100 performance tested Italian large white pigs. They assembled these reads for viral metagenomics analysis and were able to identify several viruses of the Parvoviridae family. It was found that the pigs were infected with parvovirus. This study validates the usefulness of NGS for viral metagenomics analysis in livestock. In a similar study Singh et al. (2016) sequenced mitogenome in Indian pig using NGS without designing mitogenome-specific primers.

NGS data in European pig identified three loci that were the elongation of the back and an increased number of vertebrae. The three loci were associated with the NR6A1, PLAG1, and LCORL genes. PLAG1 and LCORL are repeatedly associated with stature in other domestic animals and in humans (Rubin et al., 2012). Choi et al. (2015) carried out genome-resequencing analysis of five pig breeds including Korean native and wild pig and provided a comparative analysis of these breeds. Using NGS data they identified 25.5% novel SNPs and 35,458 non synonymous SNPs in 9904 genes which may contribute to traits of interest. They also identified two genes viz. CLDN1 and TWIST1 that could be associated with economically relevant traits.

Apart from giving insights into the diseases NGS is also used for elucidating breed specific SNPs which could further help in exploring the potential of the breed (Barris et al., 2012; Mengistie et al., 2017, Wang et al., 2017). Outcome of such studies could help design better breeding programs and have a practical benefit in the livestock industry.

Role of Next Generation Sequencing in Animal Breeding

Next generation sequencing has opened up new avenues to explore relationship between genetic and phenotypic diversity with high resolution. Many whole genome sequences of livestock from different breeds and species already exist in the public domain and many new sequencing projects are ongoing. This wealth of data allows us to identify genetic markers spanning the entire genome. In the last five years, large numbers of SNPs have been identified in livestock (particularly bovine,

Figure 2. Genomic prediction in livestock using next generation sequencing (NGS). Reference population has genotype data derived from NGS and phenotype data from the field. Genomic breeding value for each selection candidate is calculated using the prediction equation and best animals are selected for breeding.

porcine, and ovine species) by performing whole-genome association studies (WGAS). These studies can detect statistical associations between economically important traits and SNP markers, leading to the development of custom marker arrays for genomic selection. Till now genomic prediction depended on SNP arrays but using NGS data provides a clear advantage over SNP arrays as it is not bound by the extent of linkage disequilibrium between SNP markers and the causal mutation as the causal mutation is in the data itself. Use of NGS data for genomic prediction is also believed to yield better results (Figure 2). Advantage of using sequencing data for genomic selection is increased with increase in effective population size and size of the reference population (Druet et al., 2014). Increase in prediction accuracy could be achieved if all the SNPs with causal genes are included in the model equation. Perez-Enciso et al. (2015) found that prediction accuracy increased by 40% when causal genes and SNPs were included in prediction equation. However in dearth of correct biological information the accuracy will drop dramatically which was seen in study reported by Perez-Enciso et al. (2015). Only a 4% increase in accuracy was seen with the whole genome sequencing data over the HD array. Whole genome sequencing (WGS) data can increase accuracy of genomic prediction for low to moderately heritable traits in small populations depending upon QTL density, the size of the reference population and the evaluation method used. The use of WGS data was especially beneficial for multibreed predictions can specifically benefit from the use of WGS data (Iheshiulor et al., 2016).

Use of NGS in breeding and its practical benefits are yet to be seen. It is an active field of research and its advantages and drawbacks are to be accessed in practical situations before it becomes a standard practice in livestock breeding.

Conclusions

Development in sequencing technologies have opened up plethora of opportunities for researchers to better understand the complex traits and use the information thus gained in livestock breeding programs. NGS has already been used in livestock animals to identify breed specific variants, signatures of selection, causal mutations etc. Current breeding programs mostly rely on marker assisted selection (MAS). And how NGS fairs in the livestock breeding sector still remains to be seen.

Acknowledgements

References

1. Ahonen SJ, Arumilli M, Lohi H (2013) A CNGB1 Frameshift mutation in papillon and phalène dogs with progressive retinal atrophy. PLoS ONE 8: e72122.

2. Barris W, Harrison BE, McWilliam S, Bunch RJ, Goddard ME, Barendse W (2012) Next generation sequencing of African and Indicine cattle to identify single nucleotide polymorphisms. Animal Production Science 52: 133.

3. Begg AP, Todhunter K, Donahoe SL, Krockenberger M, lapeta J (2014) Severe amoebic placentitis in a horse caused by an acanthamoeba hatchetti isolate identified using next-generation sequencing. J Clinical Microbiology 52: 3101-4.

4. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF, Garcia JF, Van Tassell CP, Sonstegard TS, Eichler EE, Liu GE (2012) Copy number variation of individual cattle genomes using next-generation sequencing. Genome Research 22: 778-90.

5. Bovo S, Mazzoni G, Ribani A, Utzeri VJ, Bertolini F, Schiavo G, Fontanesi L (2017) A viral metagenomic approach on a non-metagenomic experiment: Mining next generation sequencing datasets from pig DNA identified several

porcine parvoviruses for a retrospective evaluation of viral infections. PLOS ONE 12: e0179462.

6. Chen D, Li W, Du M, Wu M, Cao B (2014) Sequencing and Characterization of Divergent Marbling Levels in the Beef Cattle (*Longissimus dorsi* Muscle) Transcriptome. Asian-Australasian J Animal Sciences 28: 158-65.

7. Choi JW, Chung WH, Lee KT, Cho ES, Lee SW, Choi BH, Lee SH, Lim W, Lim D, Lee YG, Hong JK, Kim DW, Jeon HJ, Kim J, Kim N, Kim TH (2015) Whole-genome resequencing analyses of five pig breeds, including Korean wild and native, and three European origin breeds. DNA Research 22: 259-67.

8. Downs LM, Mellersh CS (2014) An intronic SINE insertion in FAM161A that causes exon-skipping is associated with progressive retinal atrophy in tibetan spaniels and tibetan terriers. PLoS ONE 9: e93990.

9. Downs LM, Wallin-Håkansson B, Bergström T, Mellersh CS (2014) A novel mutation in TTC8 is associated with progressive retinal atrophy in the golden retriever. Canine Genetics and Epidemiology 1: 4.

10. Druet T, Macleod IM, Hayes BJ (2014) Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. Heredity 112: 39-47.

11. Iheshiulor OOM, Woolliams JA, Yu X, Wellmann R, Meuwissen THE (2016) Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. Genetics Selection Evolution 48.

12. Jiang J, Wang J, Wang H, Zhang Y, Kang H, Feng X, Wang J, Yin Z, Bao W, Zhang Q, Liu JF (2014) Global copy number analyses by next generation sequencing provide insight into pig genome variation. BMC Genomics 15: 593.

13. Jose VL, Appoothy T, More RP, Arun AS (2017) Metagenomic insights into the rumen microbial fibrolytic enzymes in Indian crossbred cattle fed finger millet straw. AMB Express 7.

14. Kim MC, Lee SW, Ryu DY, Cui FJ, Bhak J, Kim Y (2014) Identification and Characterization of MicroRNAs in Normal Equine Tissues by Next Generation Sequencing. Ray RB, editor. PLoS ONE 9: e93662.

15. Meinel DM, Margos G, Konrad R, Krebs S, Blum H, Sing A (2014) Next generation sequencing analysis of nine Corynebacterium ulcerans isolates reveals zoonotic transmission and a novel putative diphtheria toxin-encoding pathogenicity island. Genome Medicine, 6:113

16. Oh C, Lee K, Cheong Y, Lee SW, Park SY, Song CS, Choi IS, Lee JB (2015) Comparison of the oral microbiomes of canines and their owners using next-generation sequencing. white BA, editor. PLOS ONE 10: e0131468.

17. Pérez-Enciso M, Rincón JC, Legarra A (2015) Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. Genetics Selection Evolution 47.

18. Rubin CJ, Megens HJ, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Calborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen MAM, Andersson L (2012) Strong signatures of selection in the domestic pig genome. Proceedings of the National Academy of Sciences 109: 19529-36.

19. Salleh MS, Mazzoni G, Höglund JK, Olijhoek DW, Lund P, Løvendahl P, Kadarmideen HN (2017) RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics :18.

20. Sharma A, Cho Y, Choi BH, Chai HH, Park JE, Lim D (2017) Limited representation of OMIA causative mutations for cattle in SNP databases. Animal Genetics 48: 369-70.

21. Shin DH, Lee HJ, Cho S, Kim H, Hwang J, Lee CK, Jeong JY, Yoon DH, Kim HB (2014) Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. BMC Genomics 15: 240.

22. Singh AP, Jadav KK, Kumar D, Rajput N, Srivastav AB, Sarkhel BC (2016) Complete mitochondrial genome sequencing of central Indian domestic pig. Mitochondrial DNA Part B 1: 949-50.

23. Suárez-Vega A, Gutiérrez-Gil B, Klopp C, Tosser-Klopp G, Arranz JJ (2017)Variant discovery in the sheep milk transcriptome using RNA sequencing. BMC Genomics 18.

24. Taye M, Kim J, Yoon SH, Lee W, Hanotte O, Dessie T, Kemp S, Mwai OA, Caetano-Anolles K, Cho S, Oh SJ, Lee HK, Kim H (2017) Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. BMC Genetics 18.

Aditi Sharma, Jong-Eun Park, Han-Ha Chai, Gul-Won Jang, Seung-Hwan Lee, Dajeong Lim

25. Triantaphyllopoulos KA, Ikonomopoulos I, Bannister AJ (2016) Epigenetics and inheritance of phenotype variation in livestock. Epigenetics & Chromatin 9.

26. Wang Z, Chen Q, Liao R, Zhang Z, Zhang X, Liu X, M. Zhu, Zhang W, Xue M, Yang H, Zheng Y, Wang Q, Pan Y (2017) Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing. Animal Genetics 48: 38-47.

27. Yang X, Noyes NR, Doster E, Martin JN, Linke LM, Magnuson RJ, Yang H, Geornaras I, Woerner DR, Jones KL, Ruiz J, Boucher C, Morley PS, Belk KE (2016) Use of Metagenomic Shotgun Sequencing Technology To Detect Foodborne Pathogens within the Microbiome of the Beef Production Chain. Dozois CM, editor. Applied and Environmental Microbiology 82: 2433-43.