



Genetic diversity studies using molecular genetic markers

Nuri Choi, Dongwon Seo, Prabuddha Manjula and Jun Heon Lee*

Division of Animal and Dairy Science, Chungnam National University, Daejeon, 34134, Korea

ABSTRACT

The phenotype traits of livestock are determined by the genetic variation with environmental variation of individual animals. In the past, the classification of breed was depending on morphological characteristics to identify breed. However, in recent years, various genetic markers can be analyzed to identify different breeds and individual animals. These molecular genetic markers can be used to assess the genetic diversity as well as tracking the species origin and identification. Recently, a genome sequencing technology can be rapidly and accurately confirmed the various type of genetic markers in the entire genome area. Of these, developed SNP markers and the rapidly evolving SNP array technology are increasingly replacing genetic diversity analysis using microsatellite markers. Therefore, this review discussed the basic information for developing various molecular genetic markers and their use in genetic diversity studies of livestock animals.

Keywords: mtDNA, SNP, marker, genetic diversity

Introduction

In general, many livestock breed present today have pass through the centuries of natural and human selection. Thus, different breeds were adopted to different environment condition and production criteria. As the results, livestock animals have different phenotypic characteristics that distinguish them from other subgroups in the same species; however, there are difficulties and limitations in distinguishing between different individuals and groups using their basic morphology or biological samples such as blood, tissue, and secretions. This problem has resolved by the genetic classification of individuals of different phenotypes by discovering their DNA profile and genomic level variation (Syvanen, 2001). Particularly, in mitochondrial DNA (mtDNA) and nuclear DNA. In fact, these patterns of genetic variation have various forms such as single nucleotide polymorphism (SNP), insertion and deletion (indel), simple tandem repeat (STR), copy number variation (CNV).

Genomic studies provide basic information to distinguish individuals and varieties at the DNA level (Singh et al., 2014). The discovery of these genetic variations that make the diverse population allows academic, industry and animal breeders to develop a new strain which survived in prevailing climatic changes and reaction changes in consumer demand. During past few decades, genotyping techniques have developed and their costs have sharply decreased, so the scope of application for the genetic markers is continuously expanding. In the animal industry, genetic markers can be a useful identification tool for establishment of traceability system, genetic improvement by recognition of chromosomal regions where harbor the loci that significantly affect economically important traits in livestock production (Fontanesi, 2009). The majority of molecular

*Corresponding author: Jun Heon Lee, PhD, Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, Korea
Tel: +82-42-821-5779, Fax: +82-42-825-9754, E-mail: junheon@cnu.ac.kr

Received: 14 March, 2018, Revised: 14 March, 2018, Accepted: 27 March, 2018



genetic markers such as SNP and microsatellite markers used nowadays are from high-throughput systems. Moreover, identification of maternal and paternal lineages, within diversity and phylogenetic studies using specific mtDNA and Y chromosome markers are also common in use.

In this purpose, genetic diversity analysis approach using genetic variation is discussed, because it can be used as basic data for developing various application studies in diverse areas. Further, the detail used of molecular marker and high thought put technology for evaluation of diversity and mapping quantitative traits loci (QTL) and candidate gene studies are presented.

Mitochondrial DNA for diversity study in livestock

Mitochondrion is an organelle with intracellular nuclear, consist of circular double helix DNA separate from the cell nucleus. In addition, it has characteristics of maternal origin, since the mitochondria in the oocyte that are inherited to descendants during the reproductive process (Cummins, 2000). Furthermore, unlike genomic DNA sequences, mtDNA has no recombination and lack of rapid nucleotide substitutions, making it suitable for genetic diversity studies and evolutionary studies (Hoque et al., 2013). The mtDNA consist of totally 37 genes, all of which are essential genes for normal function. Within the mitochondria, the D-loop control region, which does not have functions and cytochrome oxidase I (COI) gene, which is associated with oxygen metabolism (Hebert et al., 2003), are widely used for genetic diversity studies because they contain more genetic variations than other part of the mitochondria DNA (Jin et al., 2009). In addition, there are seven NADH dehydrogenase subunit genes called ND that catalyzed the oxidation of Nicotinamide Adenine Dinucleotide Hydrogen (NADH) in the mitochondria. Among them, ND4 and ND5 have reported in previous studies, which were contained many frameshift mutations in 41 amino acid complexes I. These two genes were used to confirming the diversity of breeds (Bourges et al., 2004). In particular, it was used as ideal markers for classifying *Bos Indicus* and *Bos Taurus* breeds (Yoon et al., 2008). Therefore, it is possible to distinguish the varieties among breeds by exploring the mutation of the mtDNA and can be used to identify the origin of breeds and the deciphering the background of evolution (Seo and Lee, 2016).

SNP (Single Nucleotide Polymorphism): One of the important Nuclear DNA markers

Single nucleotide polymorphism (SNP) indicates the single base substitution in DNA sequence. In principle, one of the four nucleotide bases can be changes at the individual sequence position but it is appeared in diallelic (Nielsen, 2000). There are two main reasons for this. First, the frequency of single nucleotide substitution at the polymorphism site is very low, ranging from 1×10^{-9} to 5×10^{-9} per nucleotide and per year at the neutral positions in mammals. Second reason is the bias mechanism towards the mutation, resulted in two types of SNP. The mutation mechanism may be transition (purine-purine; A ↔ G or pyrimidine-pyrimidine; C ↔ T) or transversion (purine-pyrimidine; A ↔ C and A ↔ T or pyrimidine-purine; G ↔ C and G ↔ T) (Collins and Jukes, 1994; Vignal et al., 2002). The SNP markers that are found as differences in the DNA sequences can be used for genetic identification of varieties and as causative mutations that can affect economic phenotypes of animals (Seo and Lee, 2016). In this respect, the discovery of SNPs and their uses in a variety of genetic studies suggest that they can be efficiently used to investigate genetic diversity and improve the economic traits.

SNP markers generated by Sanger sequencing method

DNA sequencing technology has started by Fred Sanger in 1977 and been developing rapidly during last couple of decades. To identify SNP mutations in the DNA sequence, the techniques including RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), and AFLP (Amplified Fragment Length Polymorphism)

were used. RFLP was firstly used as a genetic analysis tool in 1974 (Grodzicker et al., 1974). This can identify specific base changes in the DNA sequence using restriction enzymes that recognize specific sequences and generate endonucleolytic cleavages, which are of defined length fragments discriminated using agarose gel electrophoresis techniques (Soller and Beckmann, 1983). This technique has been developed as a simple Mendelian codominant marker that allows easy identification of the individual genotype. In addition, it can be used as a genetic marker to identify whole or partial genetic characteristics in the pedigree relationship (Botstein et al., 1980). Usually, gel electrophoresis has a disadvantage of taking long time to confirm the results. To overcome these limitations, an automated electrophoresis technique was developed to identify various species such as cattle, sheep, chicken, turkey and fish using the Agilent 2100 Bioanalyzer chip (Dooley and Garrett, 2001; Fajardo et al., 2010). The RAPD technique was used for randomly amplified the anonymous segment of nuclear DNA using PCR in 1974 (Welsh and McClelland, 1990). This method is used short primer set of 8-10 bp and amplified with low annealing temperature (from 36 to 40°C) and confirmed the presence or absence of a band, indicating the polymorphisms (Welsh and McClelland, 1990; Williams et al., 1990; Liu and Cordes, 2004). Because, it generates multiple products representing the different locus, genomic variation can be easily and accurately identified without prior knowledge of the DNA sequence (Huang et al., 2003; Liu and Cordes 2004). The RAPD genotyping technique can determine the nucleotide sequence differences or the presence of indel(s) in the primer binding site (Liu and Cordes 2004). Therefore, RAPD is previously indicated as the useful tool for identification of species and was developed to the fingerprinting technique. The AFLP is usually identified with detection of restriction fragments in genome wide area by PCR amplification (Vos et al., 1995). The restriction enzyme cleaves the double strand of DNA and adapter is connected to the end of DNA fragment, and template DNA can be generated using complementary primer sets (Vos et al., 1995). This method is a fingerprinting technique for unknown genetic information site using the limited sets of generic primer pairs and is used as a genome mapping tool for genetic diversity study of species without a dense marker map (Mueller and Wolfenbarger, 1999).

Development and application of Next Generation Sequencing (NGS) using High-Density SNP data

Thirty years after the development of Sanger sequencing, Roche 454 of Life Science, Solexa of Illumina and SoLiD system of Applied Biosystems was developed since 2005. This sequence genotyping technology called Next Generation Sequencing (NGS) (Schuster, 2007). These three sequencing technology platforms were compared and analyzed for the sequencing accuracy, variant accuracy, coverage rate, false positive rate, false negative rate and variant discrepancy rate (Harismendy et al., 2009). Many of genome sequencing data accumulated in the field of animal genomics including humans have confirmed nucleotide sequence variations throughout the genome, leading to the development of SNP chips using the array method. For an instant, in cattle, 50K SNP BeadChip array was used for bovine genome-wide association study (GWAS) not only that but also sequence variations have been detected and provide a platform for bovine disease gene study and QTL mapping (Matukumalli et al., 2009). Also, large scale porcine genomes were investigated using 60K BeadChip arrays in pigs (Ramos et al., 2009). In case of sheep, a copy number variation (CNV) map of ovine was generated using a 50K SNP BeadChip array (Liu et al., 2013). For the chicken, draft genome sequencing was first released in 2004 (Hillier et al., 2004) and, subsequently, high density 600K SNP chip was released for SNP genotyping in chicken with wide range of implication (Kranis et al., 2013).

Thereby, SNP array is a new popular technology in diverse animal species and can be used for association mapping, genetic diversity, and phylogenetic study. In addition, mass SNP chip data can be analyzed rapidly and accurately for better understanding of LD (Linkage Disequilibrium), Tandem Repeat Elements, Indel (Indels and deletion), and CNV (Copy Number Variation).

Understanding of LD structure in a population is necessary for developing phenotype related genetic markers and effective application of MAS (Marker Assisted Selection) (Abasht et al., 2009). In general, LD structure change is caused by factors such as migration, selection, and genetic drift in the population. This LD size is determined by the recombination rate in the meiosis stage. The size of LD (r^2) can be confirmed by calculating the degree of linkage of the adjacent two loci using allele frequency (Hill and Robertson, 1968). In addition, effective population size can be estimated by calculated LD (r^2) values (Sved, 1971). Effective population size is the minimum number of individuals in a population that remains unchanged allele frequency in a repeated generation (Wright, 1940). Since LD can be measured the association of a pair of loci (Devlin and Risch, 1995; Hayes et al., 2003), a chromosome segment homozygosity (CSH) measurement was attempted (Hayes et al., 2003). LD evaluation of the population is essential to evaluate the genetic characteristics of the population and establish an efficient selection strategy (Barrett and Cardon, 2006). In addition, accurate LD evaluation of a population can provide a basis for tracking signature of selection for a phenotype of population and can help to detect causal mutations and genetic markers, which are associated with a phenotype through the genome-wide association study (GWAS). This can be a basic study to maximize the amount of genetic improvement by increasing the selection effect. These LD evaluation methods are also developing together with the recent progress of NGS technology and various types of high-density SNP chips. In addition, with the help of recent imputation method, low-density SNP marker genotyping has been widely used for LD evaluation and GWAS (Hayes and Goddard, 2001).

The VNTR (Variable Number Tandem Repeat) technique is based on the differences in the tandem repeat of eukaryotic genome nucleotide sequences (Takezaki and Nei, 2008). It is a sequence of 1 to 6 bp are called microsatellite (STR; Short Tandem Repeats) (Litt and Luty, 1989) and a sequence of 10 to 60 bp are called minisatellite (VNTR), according to the number of repeat motifs (Jeffreys et al., 1985). Microsatellite is widely used for genetic diversity analysis and breeding program studies in developing countries due to its relatively easy genetic analysis and relatively low cost (Rege et al., 2011). This marker has reported to have an ability to identify polymorphisms by comparing the sizes of repeated simple sequences. It has reported that individuals and breeds can be identified if the expected heterozygosity (H_{exp}) is higher than 0.5 and polymorphism information contents (PIC) values is higher than 0.6, respectively (Botstein et al., 1980; Vignal et al., 2002). Though, high polymorphic status, quantitative trait loci (QTL) studies were conducted to find positional candidate genes related to economic traits in the early 2000s, recently, they have been replaced by GWAS studies using HD SNPs.

One of the genetic structural mutations, copy number variation (CNV) can contribute to genetic diversity and evolution. The beginning of CNV study, it was defined as a DNA segment larger than 1 kb in size compared to the reference genome, but recently it contains a short segment of 50 bp (Feuk et al., 2006; Alkan et al., 2011). Recently, NGS data have been used to estimate CNV more precisely for the genome-wide studies, for an example, Korean native chicken as well as comparison of CNV in Hanwoo, Angus and Holstein population (Cho et al., 2014; Seo et al., 2015).

SNP markers *versus* MS markers

Despite the fact that the MS markers are high polymorphic, the MS typing technique is becoming limited use. Especially, because of the technical limit, the labor intensive and the cost of the experiment. One of the technical problems with MS markers is the inconsistency of allele sizes, which makes it difficult to contrast the results generated from various laboratories (Vignal et al., 2002). The reason for the different size of alleles is that allele sizes are measured differently depending on the condition of PCR reaction. For an example; *Taq* polymerase has synthesized different adenine number at the 3' end of the PCR fragment, depending on the PCR amplification conditions (Brownstein et al., 1996). These additional nucleotide sequences lead to incorrect PCR amplification sizes (Ginot et al., 1996). The error of the fragment size results in the generation of an artificial new allele, and these false positive alleles can be corrected in the pedigree population analysis but in the random

population analysis it can cause different results (Vignal et al., 2002). In addition, MS markers require professional analysis to minimize genotyping errors compared with recently developed SNP analysis methods, and it is also relatively expensive and time-consuming to perform PCR and data analysis. On the other hand, SNP markers can be simply confirmed the existence or not. SNP errors are not easy to generate in one of the two alleles which genotyped heterozygote to homozygote or homozygote to heterozygote. In comparison, the SNP analysis method has many advantages in terms of cost, accuracy of analyzing, time-consuming. Also, array-based full automated analysis system is used to identify allele genotypes (Vignal et al., 2002). In general, the MS markers were widely used for identification of individuals and paternity discrimination, but SNP markers were relatively fewer alleles than MS markers. Previously, comparing the efficiency power between 13 MS markers and 37 SNP markers for pig identification and parentage test was obtained and the results indicated that they had similar estimates of a probability of identity (PI) (Lee et al., 2012). Therefore, if a small number of SNP markers was customized, fast and accurate results can be obtained. In particular, the number of SNPs annotated in the genome-wide region through the NGS sequencing analysis developed in the 2000s became the basis for the development of low-density and high-density SNP platforms, and the developed array tools which makes easier and cheaper for the genomic researches.

Acknowledgements

Agriculture, Food and Rural Affairs (MAFRA) (213010-05-2-SB250), and “Cooperative Research Program for Agriculture Science and Technology Development (Project No.PJ012820052018) Rural Development Administration, Republic of Korea was supported for this work.

References

1. Abasht B, Sandford E, Arango J, Settar P, Fulton JE, O'Sullivan NP, Hassen A, Habier D, Fernando RL, Dekkers JC (2009) Extent and consistency of linkage disequilibrium and identification of DNA markers for production and egg quality traits in commercial layer chicken populations. *BMC genomics* 10:S2.
2. Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363-376.
3. Barrett JC, Cardon LR (2006) Evaluating coverage of genome-wide association studies. *Nat Genet* 38:659-662.
4. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331.
5. Bourges I, Ramus C, Mousson de Camaret B, Beugnot R, Remacle C, Cardol P, Hofhaus G, Issartel JP (2004) Structural organization of mitochondrial human complex I: role of the ND4 and ND5 mitochondria-encoded subunits and interaction with prohibitin. *Biochem J* 383:491-499.
6. Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20:1004-1006, 1008-1010.
7. Cho ES, Chung WH, Choi JW, Jang HJ, Park MN, Kim N, Kim TH, Lee KT (2014) Genome-wide copy number variation in a Korean native chicken breed. *Korean J Poult Sci* 41:305-311.
8. Collins DW, Jukes TH (1994) Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics* 20:386-396.
9. Cummins JM (2000) Fertilization and elimination of the paternal mitochondrial genome. *Human Reproduction* 15:92-101.
10. Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29:311-322.

11. Dooley J, Garrett S (2001) Development of meat speciation assays using the Agilent 2100 bioanalyser. Agilent Technologies Application. Note.
12. Feuk L, Marshall CR, Wintle RF, Scherer SW (2006) Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* 15:R57-66.
13. Fontanesi L (2009) Genetic authentication and traceability of food products of animal origin: new developments and perspectives. *Italian Journal of Animal Science* 8:9-18.
14. Ginot F, Bordelais I, Nguyen S, Gyapay G (1996) Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research* 24:540-541.
15. Grodzicker T, Williams J, Sharp P, Sambrook J (1974) Physical mapping of temperature-sensitive mutations of adenoviruses. In: *Cold Spring Harbor Symposia on Quantitative Biology*. 39:439-446.
16. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 10:R32.
17. Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
18. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13:635-643.
19. Hebert PD, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci* 270:S96-99.
20. Hill W, Robertson A (1968) Linkage disequilibrium in finite populations. *TAG Theoretical and Applied Genetics* 38:226-231.
21. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P., Burt DW, Groenen MA, Delany ME (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.
22. Hoque MR, Choi NR, Sultana H, Kang BS, Heo KN, Hong SK, Jo C, Lee JH (2013) Phylogenetic Analysis of a Privately-owned Korean Native Chicken Population Using mtDNA D-loop Variations. *Asian-Australas J Anim Sci* 26:157-162.
23. Huang MC, Horng YM, Huang HL, Sin YL, Chen MJ (2003) RAPD fingerprinting for the species identification of animals. *Asian Australasian Journal of Animal Sciences* 16:1406-1410.
24. Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. *Nature* 316:76.
25. Jin SD, Seo DW, Sim JM, Baek WK, Jung KC, Jang BK, Choi KD, Lee JH (2009) Single nucleotide polymorphism analysis of the COI gene in Korean native chicken. *Korean Journal of Poultry Science* 36:85-88.
26. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, Talbot R, Pirani A, Brew F, Kaiser P (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC genomics* 14:59.
27. Lee JB, Yoo CK, Jung EJ, Lee JG, Lim HT. (2012) A comparison of discriminating powers between 13 Microsatellite markers and 37 single nucleotide polymorphism markers for the use of port traceability and parentage test of pigs. *Jorunal of Agriculture and Life Science* 46:73-82.
28. Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44:397-401.
29. Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14:229.
30. Liu ZJ, Cordes J (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238:1-37.
31. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP,

- Sonstegard TS, Van Tassell CP (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 4:e5350.
32. Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Trends Ecol Evol* 14(10):389-394.
 33. Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154:931-942.
 34. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P (2009) Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PloS one* 4:e6524.
 35. Rege J, Marshall K, Notenbaert A, Ojango J, Okeyo A (2011) Pro-poor animal improvement and breeding—What can science do? *Livestock Science* 136:15-28.
 36. Schuster SC (2007) Next-generation sequencing transforms today's biology. *Nature methods* 5:16.
 37. Seo DW, Lee JH (2016) DNA markers for the genetic diversity in Korean native chicken breeds: A review. *Korean Journal of Poultry Science* 43:63-76.
 38. Singh U, Deb R, Alyethodi RR, Alex R, Kumar S, Chakraborty S, Dhama K, Sharma A (2014) Molecular markers and their applications in cattle genetic research: A review. *Biomarkers and Genomic medicine* 6:49-58.
 39. Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theor Appl Genet* 67:25-33.
 40. Sved J (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical population biology* 2:125-141.
 41. Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2:930-942.
 42. Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* 178:385-392.
 43. Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 34:275-305.
 44. Vos P, Hogers R, Bleeker M, Reijmans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407-4414.
 45. Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 18:7213-7218.
 46. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531-6535.
 47. Wright S (1940) Breeding structure of populations in relation to speciation. *The American Naturalist* 74:232-248.
 48. Yoon D, Kwon YS, Lee KY, Jung WY, Sasazaki S, Mannen H, Jeon JT, Lee JH (2008) Discrimination of Korean cattle (Hanwoo) using DNA markers derived from SNPs in bovine mitochondrial and SRY genes. *Asian-Aust J Anim Sci* 21:25-28.