Applications of Molecular Markers in Animal Breeding: A review

Firas R. Al-Samarai[1], Abdulkareem A. Al-Kazaz[2]

[1]Dep. Veterinary Public Health, College of Veterinary Medicine, University of Baghdad, Baghdad, Iraq
[2]Dep. Biotechnology, College of Science, University of Baghdad, Baghdad, Iraq

Email address:

firas_rashad@yahoo.com (F. R. Al-Samarai), kazaz_59@yahoo.com (A. A. Al-Kazaz)

Abstract: The dramatic development of molecular genetics has laid the groundwork for genomics. The applications of the new generations of molecular markers represent amazing tools for the genetic improvement of farm animals. These markers provide more accurate genetic information and better knowledge of the animal genetic resources. Scientists, who are unfamiliar with the different molecular techniques, need to know more about these techniques concerning applications, types, advantages and disadvantages. This review represents an attempt to highlight the different types of molecular markers by introducing a brief summary on the development of genetic markers including both the classical genetic markers and more advanced DNA-based molecular markers.This review could be helpful to better understand the characteristics of different genetic markers and its role in specifying the genetic diversity of animal genetic resources.

Keywords: Microsatellite, Molecular Markers, Genome, Polymorphism

1. Introduction

A genetic marker is a gene or DNA sequence with a known location on a chromosome and associated with a particular gene or trait. It can be described as a variation, which may arise due to mutation or alteration in the genomic loci that can be observed. A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, SNP), or a long one, like mini & micro satellites.

Recent years have witnessed a great interest of molecular markers, revealing polymorphism at the DNA level, as they play an important role in animal genetics studies. Sometimes the term "Smart Breeding" is used to describe marker supported breeding strategies.

The main aim of breeder is to select animal with superior genetic potential as a parents for the next generation. The first attempt to improve animals used the phenotyp of animal for a specific trait as a tool for selection. This application used external animal characteristics as a marker that called morphological markers (i.e. udder shape, coat color, body shape, skin structure, and anatomical characteristics) (Van Wezel and Rodgers, 1996). These markers depend on visual observation and measurement to identify, classify, and characterize the genetic evolution of different species or populations. The conclusions reached through applying morphological markers are often not completely accurate when they used for the evaluation of farm animal genetic because these markers based on subjective, judgments, and descriptions. Another type of markers represent by using of cytological markers that were included several criteria such as chromosome karyotypes, bandings, repeats, translocations, deletions, and inversions to investigate the genetic resources of animals (Yang et al., 2013). The chromosome mutations lead to genetic variation (Bitgood and Shoffner, 1990) . These mutations were used as markers to identify a certain location of the gene on a specific chromosome. In the domestic animals, cytological markers allow to investigate their genetic diversity by comparing chromosome number and structure between domesticated animals and their wild ancestors (Beack et al., 1973). Cytological markers still widely used in elucidating the origin and classification of species (Jonker et al., 1982) because of its good properties; rapid, economic, and straightforward technique.

The third type of markers is biochemical markers such as blood type and isozymes. These markers represent biochemical traits that could be analyzed by protein electrophoresis. The differences in the amino acid composition of isozymes and soluble proteins were used to investigate the genetic variation within species and

phylogenetic relationships between species (Buvanendran and Finney, 1967). The application of these markers was limited because the proteins and isozymes are not genetic materials. They are products of gene expression, so they could affect by environmental factors (Drinkwater and Hetzel, 1991). The molecular markers are based on the nucleotide sequence mutations within the individual's genome; they are the most reliable markers available (Yang et al., 2013).

2. Marker Assisted Selection (MAS)

Selection is one of the most important tools to improve the performance of animals. It can accomplish based on two types of data – pedigrees and phenotypes to estimate Best Linear Unbiased Prediction (BLUP) that combines these to generate estimated breeding values (EBVs). A third type of data is based on DNA markers to get a new approach named Marker assisted selection (MAS). MAS can be based on DNA in linkage equilibrium with a quantitative trait locus (QTL) (LE-MAS) –LE refers to genotype frequencies at one locus are independent of genotype frequencies at the second locus - , molecular markers in linkage disequilibrium with a QTL (LD-MAS) - LD refers to the non-random association of alleles between two loci-, or based on selection of the actual mutation causing the QTL effect (Gene-MAS). All three types of MAS are being used in the livestock industries (Dekkers, 2004).

3. Molecular and Quantitative Genetics

The most economically important traits in livestock are quantitative, that they show continuous distributions. Two models have been proposed to explain the genetic variation among such traits, the infinitesimal model (the basis of quantitative genetics) and the finite loci model (the basis of molecular genetics). The infinitesimal model assumes that traits are determined by an infinite number of unlinked and additive loci, each with an infinitesimally small effect (Fischer, 1918). This model has been exceptionally valuable for animal breeding, and forms the basis for breeding value estimation theory (Henderson, 1984). The finite loci model assumes that the existence of a finite amount of genetically inherited material (the genome). There are a total of around 20000 genes or loci in the genome (Ewing and Green, 2000). Many evidences confirmed that the distribution of the effect of these loci on quantitative traits could be classified to a few genes with large effect and a many of small effect (Shrimpton and Robertson, 1988). The search for these loci, particularly those of moderate to large effect, and the use of this information to increase the accuracy of selecting genetically superior animals, has been subjected to intensive research studies in the last two decades.

Although this approach has achieved some success – for example a mutation that discovered in the estrogen receptor locus (ESR) which results in increased litter size in pigs (Rothschild et al., 1991), but two problems have faced this approach: Firstly, candidate genes affecting a trait usually have a large number, so many genes must be sequenced and a large sample of animals is needed. Thus, the likelihood that the mutation may occur in non-coding DNA further increases the amount of sequencing required. Secondly, the mutation that associated with the phenotypic variation in a certain trait could occur in another gene that considered a non-candidate gene.

Up to now, many types of molecular markers have been utilized to detect the variation among individual and population. These markers can be classified to three groups; protein variants (allozymes), DNA sequence polymorphism, DNA repeat variation. Therefore, the delving into the details of this subject undoubtedly involves a great importance.

4. Allozyme Markers

Allozymes are enzyme variants due to allelic differences and can be visualized through protein electrophoresis. This technique was developed to quantify the genetic and geographic variation in wildlife populations, and it remains a cost-effective and straightforward method (Avise, 1994). Genetic variations caused by mutations are expressed as amino acid replacements due to changes in protein compositions, and are resolved as bands (alleles) on electrophoretic gels (DeYoung and Honeycutt, 2005).

These markers provided a valuable tool for population genetic studies in natural populations of woody plants (Adams, 1983). They usually exhibit simple Mendelian inheritance and codominant expression, making genetic interpretations easy.

5. Mitochondrial DNA (mtDNA)

mtDNA is an extra-chromosomal genome in the cell mitochondria that resides outside of the nucleus, and is inherited from mother with no paternal contribution (Emadi et al., 2010). Due to higher evolutionary rates of mtDNA relative to the nuclear genome (Adams, 1983), this marker is preferred in constructing phylogenies and inferring evolutionary history, and is therefore, ideal for within- and between-species comparisons (Emadi et al., 2010).

6. Restriction Fragment Length Polymorphism (RFLP)

RFLP is a technique that is not widely used now but it was one of the first techniques used for DNA analysis in forensic science and several other fields. RFLP is defined by the existence of alternative alleles associated with restriction fragments that differ in size from each other.

The molecular basis of RFLP is that nucleotide base substitutions, insertions, deletions, duplications, and inversions within the whole genome can remove or create new restriction sites (Yang et al., 2013).

Despite the fact that it is less widely used now, there have

been numerous benefits to RFLP analysis. It plays an important role in allowing scientists to map the human genome as well as provide information on genetic diseases (Emadi et al., 2010). RFLP analysis is useful to find where a specific gene for a disease lies on a chromosome. RFLP was also one of the first methods used for genetic typing - also known as genetic fingerprinting, profiling or testing.

Despite that RFLP have many benefits but it is still a slow and more tedious process compared to some of the newer DNA analysis techniques. Ii is also requires substantially larger sample sizes than other forms of analysis.

7. Random Amplification of Polymorphic DNA (RAPD)

In the last decade, the RAPD technique based on the polymerase chain reaction (PCR) has been one of the most commonly used molecular techniques to develop DNA markers (Kumar and Gurusubramanian, 2011).

RAPD technology provides a quick and efficient screen for DNA sequence based polymorphism at a very large number of loci. The major advantage of RAPD includes that, it does not require pre-sequencing of DNA (Nandani and Thakur, 2014).

The principle of RAPD is that, a single, short oligonucleotide primer, which binds to many different loci, is used to amplify random sequences from a complex DNA template. This means that the amplified fragment generated by PCR depends on the length and size of both the primer and the target genome (Nandani and Thakur, 2014). Since the advantages of RAPDs are the technical simplicity and the independence of any prior DNA sequence information, (Weising et al., 2005) it is viewed as having several advantages compared to RFLP and fingerprint (Lynch and Milligan, 1994).

A disadvantage of RAPD markers is the fact that the polymorphisms are detected only as the presence or absence of a band of a certain molecular weight, with no information on heterozygosity besides being dominantly inherited, and also show some problems with reproducibility of data (Brumlop and Finckh, 2010).

8. Amplified Fragment Length Polymorphism (AFLP)

AFLP markers have found the widest application in analyses of genetic variation below the species level, particularly in investigations of population structure and differentiation (Hedrick, 1992).

AFLP methods rapidly generate hundreds of highly replicable markers from DNA; thus, they allow high-resolution genotyping of fingerprinting quality. The time and cost efficiency, reproducibility and resolution of AFLPs are superior or equal to those of other markers (RAPD, RFLP and microsatellites) (Brumlop and Finckh, 2010).

The AFLP method is an ideal molecular approach for population genetics and genome typing, it is consequently widely applied to detect genetic polymorphisms, evaluate, and characterize animal genetic resources (Ajmone-Marsan et al., 2002).

9. Microsatellites

Microsatellites or simple sequence repeat (SSR) loci, which have been referred to in the literature as variable number of tandem repeats (VNTRs) and simple sequence length polymorphisms (SSLPs), are found throughout the nuclear genomes of most eukaryotes and to a lesser extent in prokaryotes (Varshney et al., 2005).

Microsatellites range from one to six nucleotides in length (Van Oppen et al., 2000) and are classified as mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats. The sequences of di-, tri- and tetranucleotide repeats are the most common choices for molecular genetic studies (Selkoe and Toonen, 2006). They are repeated (usually 5-20 times) in the genome with a minimum repeat length of 12 base-pairs (Goodfellow, 1992).

This approach assumes that a certain quantitative trait was affected by many unknown genes. So, this approach is looking for associations between the variation of allele and quantitative traits at the neutral DNA markers. The DNA marker is located on a chromosome and its inheritance can be monitored (Hyperdictionary, 2003).

10. Single-Nucleotide Polymorphism (SNP)

In 1996, Lander proposed a new molecular marker technology named SNP. When a single nucleotide (A, T, C, or G) in the genome sequence is altered this will represent the SNP. In other words, it refers to a sequence polymorphism caused by a single nucleotide mutation at a specific locus in the DNA sequence (Yang et al., 2013).

This sort of polymorphism includes single base transitions, transversions, insertions and deletions (Goodfellow, 1992), and the least frequent allele should have a frequency of 1% or greater (Lander, 1996). Transitions are the most common (approx.2/3) among all the SNP mutation types (Zhao and Boerwinkle, 2002). SNP markers are one of the popular approach, despite they can be considered as a step backwards (simple bi-allelic co-dominant markers) when compared to the highly informative multi-allelic microsatellites. The more recent SNP concept has basically arisen from the recent need for very high densities of genetic markers for the studies of multifactorial diseases (Vignal et al., 2002).

The fundamental principle of SNPs is to hybridize detected DNA fragments with high-density DNA probe arrays (also called SNP chips); the SNP allele is then named according to the hybridization results (Yang et al., 2013).

SNPs are third generation molecular marker technology coming after RFLPs and SSRs (Peter, 2001); it was successfully performed to investigate genetic variation

among different species and breeds (The Bovine HapMap Consortium.2009).

The role of SNPs in farm animals was very important concerning the population structure, genetic differentiation, origin, and evolution research (Yang et al., 2013). On the other hand, the most important disadvantage of SNPs is the low level information obtained as compared with that of a highly polymorphic microsatellite but this can be solved by using a higher numbers of markers (SNP chips) and whole-genome sequencing (Werner et al., 2002).

11. DNA Barcoding Markers

A DNA barcode is a short DNA sequence from a standardized region of the genome used for identifying species. The essential aim of DNA barcoding is to use a large-scale screening of one or more reference genes in order to assign unknown individuals to species, and to enhance discovery of new species (Hebert et al., 2003).

Biological taxonomists apply this principle to species classification. The first application of using the DNA sequences in systematic biological taxonomy (also called DNA taxonomy) was conducted by Tautz et al., (2002) and then , Hebert et al., (2003) proposed the concept of DNA Barcoding and suggested its use for a single mtDNA gene, mitochondrial cytochrome c oxidase I (COI), as a common sequence in animal DNA barcoding studies.

DNA Barcoding has a high accuracy of 97.9% (Goodfellow, 1992), and provides a new, quick, and convenient identification strategy for animal genetic diversity (Morin et al., 2004). This approach like previous mentioned markers have some disadvantages represents by the genome fragments are very difficult to obtain and are relatively conservative and have no enough variations. Some organisms cannot be identified with COI because of the low evolution rates of COI sequences in some species. Moreover, COI is an mtDNA sequence of maternal origin, which could bias species diversity (Hajibabaei et al., 2006).

12. Conclusions

The accurate genetic evaluation of animals is the primary target for their conservation and utilization. Different methods have been developed and tested at the DNA sequence level. These methods provide a large number of markers and opening up new opportunities for evaluating diversity in farm animal genetic resources. Among all these methods, microsatellites (SSR) remained the marker of choice for the past 15 years (Morin et al., 2004).

Wang et al., (2009) reported that SNP markers will replace microsatellites for some applications as SNP markers have good genome coverage. However, the results of recent studies (de Bakker *et al.* 2006 ; Vezzulli *et al.* 2008) revealed that SNP markers can only be transferred to different mapping populations within the same species, but not across species. This will limit the applications of SNP markers on related minor species. In contrast, due to multiple alleles,

cost-effectiveness, and transferability, SSR markers will continue to play an important role in different genetic studies in the future.

References

[1] Adams W. T. (1983). Application of isozymes in tree breeding. *In* S. D. Tanksley and T. J. Orton (eds.), Isozymes in plant genetics and breeding, part A, 381–400. Elsevier Science, Amsterdam, The Netherlands.

[2] Ajmone-Marsan P, Negrini R, Milanesi E, Bozzi R, Nijman IJ, Buntjer JB, Valentini A, Lenstra JA. (2002). Genetic distances within and across cattle breeds as indicated by biallelic AFLP markers. Anim Genetics, 33:280–286.

[3] Avise JC (1994). Molecular markers, natural history and evolution. Chapman and Hall, New York.

[4] Becak ML, Becak W, Roberts FL.(1973). Fish, amphibians, reptiles and birds. Berlin, Heidelberg, New York: Springer-Verlag.

[5] Bitgood JJ, Shoffner RN.(1990). Cytology and cytogenetics. Poult breeding Genet, 22:401–427.

[6] Brumlop S., and Finckh, M.R.(2010). Applications and potentials of marker assisted selection (MAS) in plant breeding. Final report of the F+E project "Applications and Potentials of Smart Breeding" (FKZ 350 889 0020) On behalf of the Federal Agency for Nature Conservation December 2010.

[7] Buvanendran V, Finney DJ.(1967). Linkage relationships of egg albumen loci in the domestic fowl. Br Poult Sci., 8:9–13.

[8] Dekkers JC. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci. 82 E-Suppl: E313-328.

[9] DeYoung RW, Honeycutt RL (2005). The molecular toolbox: genetic techniques in wildlife ecology and management. J Wildl Manage 69:1362–1384.

[10] Drinkwater RD, Hetzel DJS.(1991). Application of molecular biology to understanding genotype-environment interactions in livestock production. In Proc. of an International Symposium on Nuclear Techniques in Animal Production and Health. Vienna: IAEA, FAO; 437–452. 15–19 April.

[11] Emadi, A.; Crim, MT.; Brotman, DJ. et al. (2010). Analytic validity of genetic tests to identify factor V Leiden and prothrombin G20210A. Am J Hematol, Vol.85, No.4, (April 2010), pp. 264-270, ISSN 0361-8609.

[12] Ewing B, Green P. (2000). Analysis of expressed sequence tags indicates 35,000 human genes. Nat Genet. 25:232-4.

[13] Fischer, R. A. (1918). The correlation between relatives: the supposition of mendelain inheritance. Transactions of the royal society of Edinburgh. 52:399.

[14] Goodfellow PN. (1992). Variation is now the theme. Nature 359, 777-778.

[15] Hajibabaei M, Janzen DH, Burns JM. (2006). DNA barcodes distinguish species of tropical Lepidoptera. Proc Nat Acad Sci USA, 103:968–971.

[16] Hebert PDN, Cywinska A, Ball SL, de Waard JR. (2003). Biological identifications through DNA barcodes. Proc R Soc Biol Sci., 270:313–321.

[17] Hedrick P. (1992). Shooting the RAPDs. Nature 355: 679-680.

[18] Henderson, C. R. (1984). Applications of linear models in animal breeding. Can. Catal. Publ. Data, Univ Guelph, Canada.

[19] Hyperdictionary, (2003). hyperdictionary. Available from: http://www.hyperdictionary.com.

[20] Jonker J, Meurs G, Balner H.(1982). Typing for RhLA-D in rhesus monkeys: II.genetics of the D antigens and their association with DR antigens in a population of unrelated animals. Tissue Antigens, 19:69–78.

[21] Kumar N.S., and Gurusubramanian G.(2011). Random amplified polymorphic DNA (RAPD) markers and its applications. Sci Vis 11 (3), 116-124.

[22] Lander ES. (1996). The new genomics: global views of biology. Science,274:536–539.

[23] Lynch, M., and Milligan, B.G. (1994). Analysis of population genetic structure with RAPD markers. Molecular Biology., 3:91-99.

[24] Morin P. A., Luikart G., Wayne R. K., and the SNP American Association of Blood Banks, Arlington, workshop group .(2004). SNPs in ecology, evolution VA, USA, pp. 277D280. and conservation. Trends Ecol. Evol. 19: 208-216.

[25] Nandani K. N. and Thakur,S.K.(2014). Randomly amplified polymorphic DNA- a brief review. American Journal of Animal and Veterinary Sciences 9 (1): 6-13, 2014.

[26] Payseur BA, Cutter AD. (2006). Integrating patterns of polymorphism at SNPs and STRs. Trends Genet. 22:424–429.

[27] Peter G., (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med, 114:204–210.

[28] Rothschild MF, Larson RG, Jacobson C, Pearson P. (1991). PvuII polymorphisms at the porcine oestrogen receptor locus (ESR). Anim Genet. 22(5):448.

[29] Selkoe KA, Toonen RJ (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615-629.

[30] Shrimpton, A. E., Robertson, A. (1988). The Isolation of Polygenic Factors Controlling Bristle Score in Drosophila melanogaster. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics 118: 445-459.

[31] Tautz D, Arctander P, Minelli A, Thomas RH: DNA points the way ahead in taxonomy. (2002). Nature, 418:479.

[32] The Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science, 324:528–532.

[33] Van Oppen MJ, Rico C, Turner GF, Hewitt GM (2000). Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. Molecular Biology and Evolution.17, 489-498.

[34] Van Wezel IL, Rodgers RJ. (1996). Morphological characterization of bovine primordial follicles and their environment in vivo. Biol Reprod 55:1003–1011.

[35] Varshney, R.K.; Graner, A. & Sorrels, M.E. (2005). Genetic microsatellite markers in plants: features and applications. Trends in Biotechnology, Vol.23, No.1, (January 2005), pp. 48-55.

[36] Vignal A., Milan D., SanCristobal M., and Eggen A. (2002). A review on SNP and other types of molecular their use in animal genetics. Genet. Sel. Evol. 34: 275-305.

[37] Wang, M.L., Barkley, N.A., Jenkins, T.M. (2009) Microsatellite Markers in Plants and Insects. Part I: Applications of Biotechnology. www.ent.uga.edu/pubs/jenkins_microsatellitei.pdf

[38] Weising K, Nybom H, Wolff K, Kahl G (2005) DNA fingerprinting in plants – principles, methods, and applications, 2nd edn. CRC Press, Boca Raton, FL.

[39] Werner M, Sych M, Herbon N, Illig T, Konig IR, Wjst M. (2002). Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. Hum Mutat, 20:57–64.

[40] Yang, W., Kang, X., Yang, Q., Lin, Y. and Fang, M.(2013). Review on the development of genotyping methods for assessing farm animal diversity. J Anim Sci Bio. 4, 2:1 – 6.

[41] Zhao ZM, and Boerwinkle E. (2002). Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. Genome Res, 12:1679–1686.