Genetic Markers: Importance, uses and applications

Shahid Raza[1], Muhammad Waseem Shoaib[2] and Hira Mubeen[1]

[1]University of South Asia, Lahore, Pakistan
[2]District Head Quarter (DHQ) Hospital, Faisalabad, Pakistan

Abstract- Genetic markers are useful in identification of various genetic variations. The development of DNA-based genetic markers has had a revolutionary impact on genetic studies. With DNA markers, it is theoretically possible to observe and exploit genetic variation in the entire genome. These markers can be used to study the evolutionary relationships among individuals. Popular genetic markers include allozymes, mitochondrial DNA, RFLP, RAPD, AFLP, microsatellite, SNP, and EST markers. The application of DNA markers has allowed rapid progress in investigations of genetic variability and inbreeding, parentage assignments, species and strain identification, and the construction of high-resolution genetic linkage maps for aquaculture species. The advent of next-generation sequencing (NGS) has revolutionized genomic and transcriptomic approaches to biology. The new sequencing tools are also valuable for the discovery, validation and assessment of genetic markers in populations. This review focuses on importance and uses of genetic markers with advent of modern technologies.

Index Terms- allozymes, NGS, EST, genetic linkage map

I. Introduction

Genetic markers are variants in the DNA that are associated with a specific disease phenotype revealing variations. DNA Marker technology has revolutionized the world of genetic research. These markers can be used to detect polymorphism between different genotypes or alleles of a gene for a particular sequence of DNA in a gene pool. Markers whose presence confers a high level of probability of disease would be most useful as diagnostic tools. A marker may have functional consequences, such as altering the expression or function of a gene that directly contributes to development of disease. Alternatively, a marker may have no functional consequences but may be located near a functional variant such that both the marker and variant tend to be inherited together in the population at large.

DNA variations are classified as “neutral” when they cause no change in metabolic or phenotypic traits and not subjected to positive, negative, or balancing selection. In other case, they are referred to as “functional”. Mutations in key nucleotides of a coding sequence may change the amino acid composition of a protein and lead to new functional variants. Such variants may have an increased or decreased metabolic efficiency compared to the original “wild type”may lose their functionality completely, or even gain a novel function.

II. Methods of detection

There are two basic methods to detect the polymorphism: Southern blotting, a nuclear acid hybridization technique and PCR, a polymerase chain reaction technique. Using these methods can help to identify the variation in DNA samples or polymorphism for a specific region of DNA Sequence. Studies shown that in case of complex diseases, it is more difficult to identify genetic markers. Because complex diseases are polygenic i.e caused by defect in multiple genes.

III. Techniques for genetic polymorphisms

The most common methods used in various applications are the restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeat (SSR), and single nucleotide polymorphism (SNP). The genetic markers can also be classified into SNPs (due to sequence variation, e.g. RFLP) and non-SNPs (due to length variation, e.g. SSR) (Gupta et al., 2001). The marker techniques help in selection of multiple desired characters simultaneously using F2 and back-cross populations, near isogenic lines, doubled haploids and recombinant inbred lines. Single nucleotide polymorphisms (SNPs) have received much attention as potential genetic markers. They have the advantage of a high frequency in the human genome (1 occurs every 1000 nucleotides, on average) and are relatively easy to genotype using current technologies. A significant portion of current research is devoted to pool of gene variants that can be used to identify genes. The level of expression of certain genes may help to know the disease state. If these genes are overexpressed or suppressed they can be studied with biomarkers.

Type I Genetic Markers

Genetic markers are classified in two different categories. Type I markers are associated with genes of known function, while type II markers are associated with anonymous genomic segments (O’Brien, 1991). Under this classification, most RFLP markers are type I markers because they were identified during analysis of known genes. Similarly, allozyme markers are type I markers because the protein they encode has known function. EST markers are also type I markers because they represent transcripts of genes. Type I markers have utility in studies of comparative genomics, genome evolution, candidate gene identification, and enhanced communication among laboratories. Due to evolutionary constraints on the genome, many genes and their organization are conserved among species. Type I markers serve as a bridge for comparison and transfer of genomic

information from a map rich species into a relatively map-poor species.

Type II Genetic Markers

RAPD markers are type II markers because RAPD bands are amplified from anonymous genomic regions via the polymerase chain reaction (PCR). AFLP markers are type II because they are also amplified from anonymous genomic regions. Microsatellite markers are type II markers unless they are associated with genes of known function. SNP markers are mostly type II markers unless they are developed from expressed sequences (eSNP or cSNP). Indels are becoming more widely used as markers since they often are discovered during genomic or transcriptomic sequencing projects. They can be either type I or type II markers depending on whether they are located in genes. In general, type II markers such as RAPDs, microsatellites, and AFLPs are considered to be non-coding. Such markers have found widespread use in population genetic studies. (Brown and Epifanio, 2003).

Microsatellites

Currently, microsatellites are the most popular markers in livestock genetic characterization studies (Sunnucks, 2001). Their high mutation rate and codominant nature permit the estimation of within and between breed genetic diversity, and genetic admixture among breeds even if they are closely related.

AFLPs

AFLPs are dominant biallelic markers (Vos et al., 1995). Variations at many loci can be arrayed simultaneously to detect single nucleotide variations of unknown genomic regions, in which a given mutation may be frequently present in undetermined functional genes.

Mitochondrial DNA markers

MtDNA markers may also provide a rapid way of detecting hybridization between livestock species or subspecies (e.g. Nijman et al., 2003). The polymorphisms in the sequence of the hypervariable region of the D-loop or control region of mtDNA have contributed greatly to the identification of the wild progenitors of domestic species, the establishment of geographic patterns of genetic diversity, and the understanding of livestock domestication (Bruford et al., 2003).

USES

The usefulness of molecular markers can be measured based on their polymorphic information content (Botstein et al., 1980). PIC refers to the value of a marker for detecting polymorphism in a gene pool or complete population. PIC depends on the number of detectable alleles and the distribution of their frequencies.

Traditionally, the development of markers such as microsatellites, RFLPs and AFLPs was a costly, iterative process that involved time-consuming cloning and primer design steps that could not easily be parallelized[5-7]. Scoring of marker panels across target populations was also expensive and laborious. The advent of high- throughput SNP arrays removed this bottleneck from the genotyping process, but not from the discovery process: the production of a high-quality array requires a substantial investment of resources.

Restriction enzymes have been a core tool for marker discovery and genotyping for decades, ever since the development and use of RFLPs to link many genes to human diseases and to construct the first complete linkage map of the human genome[8-10]. Restriction enzymes remain central to the genome-wide NGS methods discussed here, but rather than length polymorphisms, the developed markers are sequenced SNPs or structural variants. The diversity of restriction enzymes are available which makes them an extremely versatile assay tool.

IV. Discussion

Genetic markers can be considered as heritable polymorphisms that can be measured in one or more populations of individuals. “The ideal molecular approach for population genomics should uncover hundreds of polymorphic markers that cover the entire genome in a single, simple and reliable experiment (2003, Luikart et al. 2003). Previously, there was no such approach but now with emergence of next genome sequencing, it has been easier to discover, sequence and genotype thousands of genetic markers in a single step[2]. Many of these NGS methods depend on restriction enzymes to produce a reduced representation of a genome. The use of restriction enzymes combined with NGS for genome wide marker discovery in new technologies such as reduced-representation sequencing, restriction-site-associated DNA sequencing (RAD-seq) and multiplexed shotgun genotyping (MSG), and we make recommendations for the use of these technologies in future studies.RNA can be reverse transcribed into cDNA and cut with restriction enzymes, producing a small set of markers from the transcriptome that can be used to assay gene expression without the burden of transcriptome assembly. However, we expect the largest gains to come from improved analysis of the data produced by these methods. A better understanding of the variation in the data will enable more robust inference of marker identity and genotypes. Rapidly increasing throughput will allow more individuals to be sequenced in a population, more markers to be sequenced per individual and each marker to be genotyped at greater depth and so with greater accuracy.

V. Conclusion

DNA markers are useful in many aspects of studying genetic polymorphisms in in huamans. The development and application of DNA marker technologies already underway in other areas such as molecular systematics, population genetics, evolutionary biology and conservation genetics. Advances in genomics are also likely to affect other areas utilizing molecular markers as well. Researchers claimed that it will be possible to sequence tens of thousands of markers in thousands of individuals in the near future. This will be far in excess of what is required for many studies in which a small number of markers are quite sufficient and will be accessible using the various methods.

REFERENCES

[1] O'Brien, S.J., 1991. Molecular genome mapping: lessons and prospects. Curr. Opin. Genet. Dev. 1, 105– 111.

[2] Brown, B., Epifanio, J., 2003. Nuclear DNA. In: Hallermann, E.M. (Ed.), Population Genetics: Principles and Applications for Fisheries Scientists. American Fisheries Society, Bethesda, MD. 458 pp.

[3] Sunnucks, P. 2001. Efficient genetic markers for population biology. Tree, 15: 199–203.

[4] Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J. & Kuiper, M. 1995. AFLP: a new technique for DNA fingerprinting.Nucleic Acids Research, 23: 4407–1444.

[5] Nijman, I.J., Otsen, M., Verkaar, E.L., de Ruijter, C. & Hanekamp, E.2003. Hybridization of banteng (Bos javanicus) and zebu (Bos indicus) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. Heredity, 90: 10–16

[6] Bruford, M.W., Bradley, D.G. & Luikart, G.2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics, 4: 900–910.

[7] Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using

[8] restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314–331.

[9] Vos, P. et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res.23, 4407–4414 (1995).

[10] Jarne, P. & Lagoda, P. J. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11, 424–429 (1996).

[11] Gusella, J. F. et al.A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306, 234–238 (1983).

[12] Riordan, J. et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245, 1066–1073 (1989).

[13] Donis-Keller, H. et al. A genetic linkage map of the human genome. Cell 51, 319–337 (1987).

[14] Luikart G, England P R, Tallmon D, Jordan S and Taberlet P The power and promise of population genomics: from genotyping to genome typing. Nature Rev Genet 4 (2003) 981-994.

[15] Stapley, J. et al. Adaptation genomics: the next generation. Trends Ecol. Evol. 25 , 705–712 (2010).

AUTHORS

First Author – Shahid Raza, University of South Asia, Lahore, Pakistan
Second Author – Muhammad Waseem Shoaib, District Head Quarter (DHQ) Hospital, Faisalabad, Pakistan
Third Author – Hira Mubeen, University of South Asia, Lahore, Pakistan

Correspondence Author – mianrs@yahoo.com