Single Nucleotide Polymorphism (SNP)

[1]Mallar, K. N. and [2]Prasenjit, D.

[1] M.Sc. (Agri.), Dept. of Biotechnology, AAU, Jorhat, Assam

[2] M.Sc. (Agri.), Dept. of Biotechnology, UAS, Dharwad, Karnataka

Correspondence mail id: mallar2015@gmail.com

➢ What is SNP?

If there is a difference of only one nucleotide (either Adenine / Guanine / Thiamine / Cytosine) in between two DNA sequences the difference is called Single Nucleotide Polymorphism (SNP), often pronounced as "*snip"* or *"snips"* in plural.

Eg: Sequence 1**. 5′-TCGAAGCCTA-3′**

Sequence **2. 5′-TCGAAGCTTA-3′**

These DNA sequences from two different individuals are heaving one single nucleotide difference in between them and there are two alleles (C/T). However, literally for a particular locus in the chromosome there could be four alleles representing one of the four nitrogen bases.

SNP can be of two categories: Transition type and Transversion type (Fig. 1). C/T transition is the most common type of SNP found in human and higher plants. A variation can be considered as SNP if this occurs in at least 1% individuals of the population, otherwise this will be treated as a mutation.

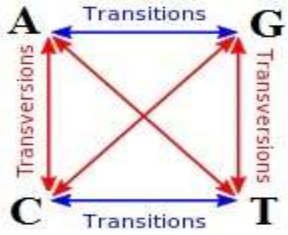

Fig 1.

➢ Where do SNP occur?

SNPs are widely distributed throughout the genome and can be present both in the coding region and in the non-coding region of the genome. However, SNPs are more frequently found in the non-coding regions. Diploid organisms, the two homologous gene copies (alleles) can also be

differentiated with SNPs. Alignment of the genome sequence data of *Japonica* and *Indica* rice indicated the occurrence of 1 SNP/ 268bp in rice genome approx. SNP frequency varies from 3 SNPs/ kb in coding regions and 27.6 SNPs/ kb in transposable element. In other crop plants (Barley, Soybean, Sugar beet, Maize, Potato) the frequency varies from 1 SNP/100-300 bp genome. Human genome contains 3-17 million SNPs. The frequency of SNP varies within each genome. Genomic regions with low recombination rate generally have low number of SNPs and the region maintaining two or more haplotypes (disease resistant genes) have greater SNP diversity.

- **Types of SNP:** SNP present in the non-coding region of genome: **ncSNP**; SNP found in exons: **cSNP** or coding SNP; **cDNA-SNP**: SNP discovered from cDNA; **pSNP**: SNP found in the promoter region; **rsSNP**: reference SNP serves as a reference point for locating other SNPs; **synSNP**: synonymous SNP changes the amino acid sequence; **nsSNP**: non synonymous SNP does not change the amino acid sequence.

How to find SNPs?

One of the most common methods of SNP discovery is the analysis of expressed sequence tag (EST) data of different individuals of a concerned species available in data bases. However, in broad sense methods of SNP discovery comes under two categories.

1. *In silico* Methods: *In silico* methods are the cheaper and more efficient methods of identifying novel SNPs. With developing of high throughout sequencing technique large number of data from various genome sequencing projects are being available. Alignment of these genome sequence data facilitates the discovery of SNPs but the true polymorphism should be differentiated from the sequencing errors (software programs like PolyByes and PolyPhred are most useful in distinguishing true SNPs out of sequencing errors).

EST sequence data has provided plenty of information about gene expression level even in tissue specific manner. EST dada can be the richest source of biologically active SNPs. One fact associated with the EST data is that multiple ESTs can be resemble to a single gene also, due to splicing mechanism and that is why EST data are full of redundancy. But in case of SNP discovery this redundancy nature of EST data is very much helpful in order to identify diversity.

2. *In vitro* Approaches: *In vitro* methods are generally more appropriate.

Restriction enzyme based techniques: Use restriction enzyme was the first method to be used for detecting polymorphism in DNA in early days. This method was able to detect point mutation at the restriction sites, followed by restriction fragment polymorphism (RFLP). If one allele contains the restriction site while other one does not, digestion of those alleles will generate fragments of different length. Another method is cleaved amplified polymorphic sequence (CAPS), in which polymerase chain reaction (PCR) amplified fragments are digested using restriction enzymes. The main drawback of SNP-RFLP and CAPS technique is that the polymorphism must occur within a restriction site.

DNA conformation based techniques: Single stranded conformational polymorphism

(SSCP) is a method of distinguishing in between similar sized DNA fragments under poly acrylamide gel electrophoresis (PAGE). In heterozygous individuals, hetero-duplex is formed during continuous heating and cooling process which can be distinguished by PAGE. Because hetero-duplex migrates slowly due to the presence of mismatch base pairing than homo-duplex. Ultimately the mismatch contains SNP.

TILLING: Target induced local lesion in genome (TILLING) is technique in which mismatch-specific endonuclease from **CEL 1** family is employed. The CEL 1 restriction enzyme cleaves double stranded DNA at mismatch site. The mismatch can be created by denaturation/ cooling process of DNA pool or DNA of a heterozygous line. TILLING does not require the knowledge about the position mismatch, so this technique can be applied to any SNP.

Sequencing of PCR amplicon: In this method PCR amplicons are generated using locus specific primers (LAPs) and then the segment can be directly sequenced or cloner into plasmid for sequencing.

However, after discovery the true SNP should be validated with re-sequencing (this is done by 454 technology and Pyro-sequencing) and the following points should be considered.

- ✓ SNP is not the product of sequencing error.
- ✓ Alleles of SNP should represent homologous genomic region not the paralogous.
- ✓ Alleles of the SNP should follow Mendelian segregation pattern.
- ✓

➢ SNP as molecular marker:

SNP locus is a specific position in the genome at which different individuals of the same species have different nucleotide. SNPs are allele specific and thus represents greater diversity within the population (acts as co-dominant marker). SNPs are linked with important traits (economic traits in plants, disease resistant in human, etc.). Some SNPs present within the coding region may produce phenotypic effect by altering the amino acid sequence of the coded protein or by altering the splicing pattern. SNPs may also effect the promoter region and thus generates greater phenotypic effect. SNPs are abundant and presents throughout the genome indicating very high level of polymorphism. SNP system is low time consuming but has high reproducibility.

The SNP marker system can be used for almost all purposes like Diversity analysis, Genetics and Breeding purposes. Because of abundance a high resolution genetic map can be prepared with SNPs. Thus SNPs serve as the excellent molecular marker.

The limitations of SNP marker system is that it demands high technical knowledge for operating and the cost of initial setup is high.

➢ How to do SNP genotyping?

- **Allele Specific PCR**: This type of PCR reaction allows the amplification of a particular allele (containing the SNP) but at the same time prevents the amplification of another

allele (due to mismatch). Here, one primer is designed based on the conserved sequence and the other primer is designed based on the specific SNP locus at the 3′ end. Thus the SNP containing allele can be identified in gel electrophoresis (Fig. 2).

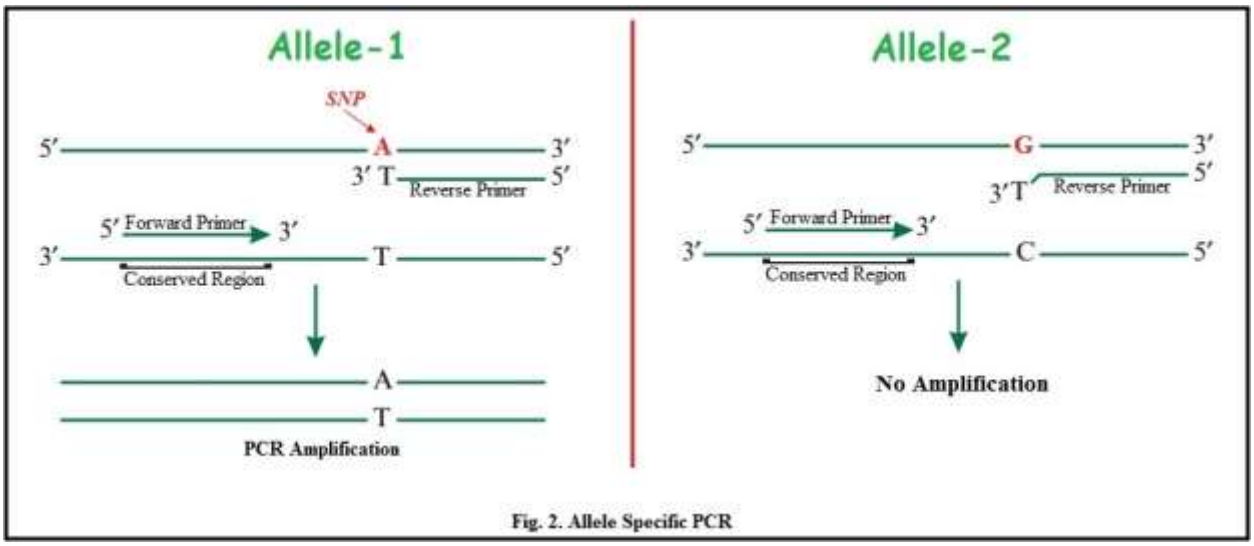

Fig. 2. Allele Specific PCR

- **TaqMan Assay:** In this technique real-time monitoring (qPCR) is done with the uses of one specially designed TaqMan™ probe which is complementary to the SNP (sequence containing the SNP). Here two primers are used to amplify the segment containing the SNP and one fluorescent labelled TaqMan probe is used. Speciality of this special probe is that, the attached Reporter die cannot give fluorescence until attached to the Quencer molecule (suppresses the fluorescence of Reporter die). At first the TaqMan probe is attached to the SNP containing region but gradually primer elongation starts and the Taq polymerase Enzyme cleaves the 5′ end of the TaqMan probe releasing the Reporter dye (due to 5′ nuclease activity). Thus fluorescence is generated. Taq polymerase will not able to cleave the TaqMan probe until it binds to the complementary strand and finally the presence of SNP is confirmed by a florescent detector (Fig. 3).

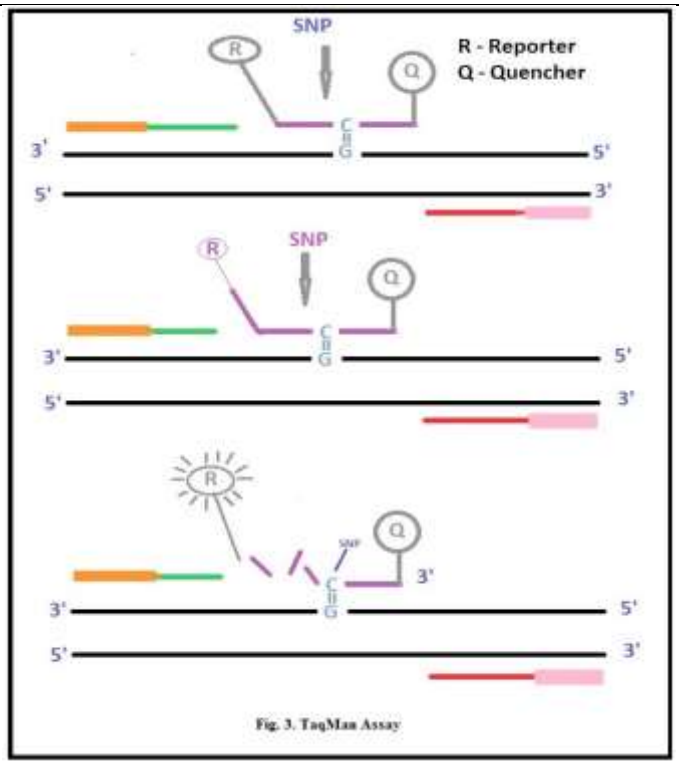

Fig. 3. TaqMan Assay

- **Molecular Beacon:** Molecular beacons are specially designed oligo nucleotides with a fluorophore attached to the 5′ region and a quencher at the 3′ region. The middle region of these probe is complementary to the sequences flanking the SNP (including the SNP). While remaining freely the oligo nucleotide forms a hair pin structure due to complementary base pairing within the molecule and do not producing fluorescence. But after base pairing with the complementary SNP region (of SNP allele) it becomes linear. In the linear molecule, fluorophore and quencher become separated thus generates fluorescence. In case of mismatch the molecular bacon remains in the hair-pin structure and fluorescence is not generated. Molecular Beacon is mixed with denatured PCR product and allowed for annealing (Fig.4).

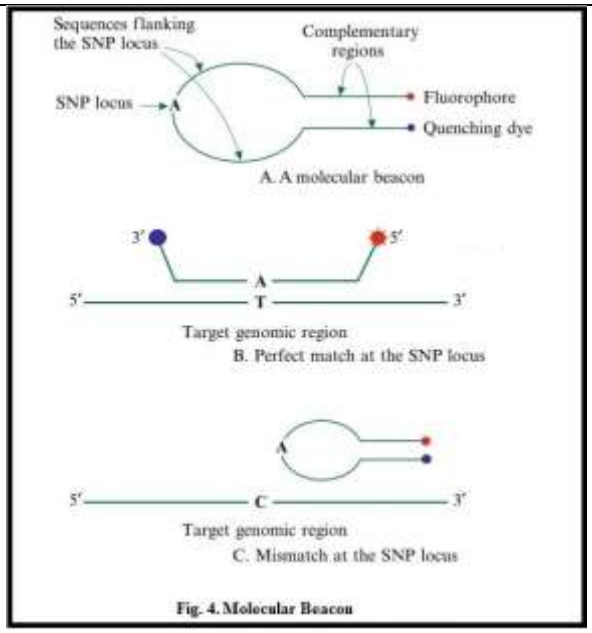

Fig. 4. Molecular Beacon

.

- **Microarray based technique:** In this technique several SNP loci can be genotyped together by developing *SNP-Microarray* or SNP-DNA chip. DNA-chip / *microarray* is a silicon plate on which a number of different ssDNA molecular are covalently linked from one end. In each spot of *microarray* chip each DNA molecule is of 25nt length and contains complementary region of same SNP (in different position). This for ensuring reliability of the system. Short genomic regions containing the SNP are amplified with PCR. All the PCR products are then fluorescently labelled and used for hybridization (under optimized condition) with the ssDNA attached onto the chip. All non-hybridized PCR products are washed away and a detector records the fluorescence. The fluorescence data is analyzed to find SNPs (Fig. 5).

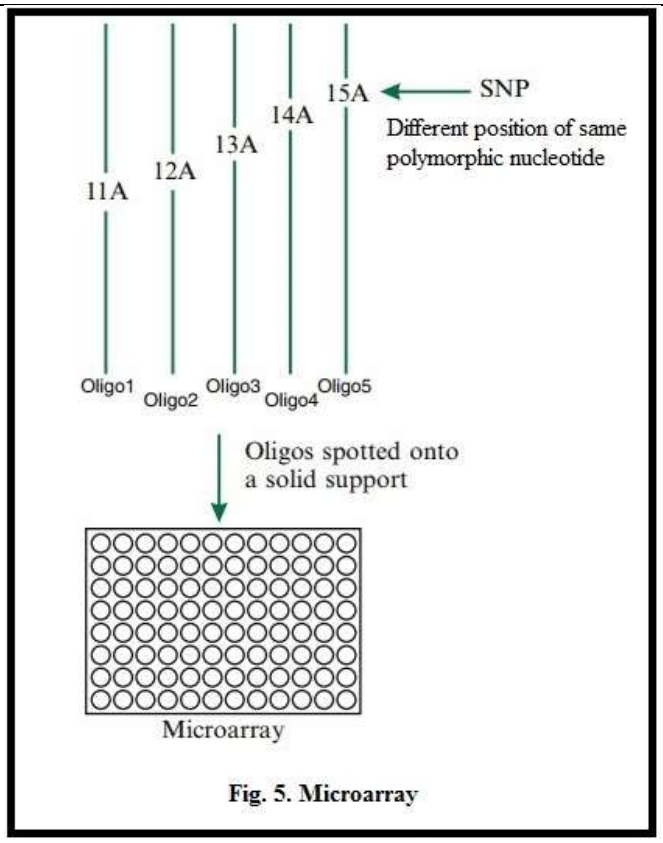

Fig. 5. Microarray

There are some similar technique in which molecular beats are used onto which oligo nucleotides are attached.

- **Primer Extension:** In this process specially designed primers are used for annealing with the targeted sequence containing the SNP and only one or few nucleotide is synthesized. The primer is so designed that the 3′ end of the primer base pairs up to the just proceeding nucleotide of the polymorphic nucleotide (the SNP), as a result the very first nucleotide added by the polymerase will be complementary to the polymorphic nucleotide (the SNP). Now the PCR mixture is prepared by adding one ddNTP (dideoxy nucleotide) and remaining three types of dNTP (deoxy nucleotide). In this way four separate PCR reaction mixtures are prepared.

  In case any particular ddNTP becomes complementary with the polymorphic nucleotide, there is no further extension happens. On the other hand if the dNTP becomes complementary, the primer extension continues. The PCR products of four separate reactions are analysed either by electrophoresis or by mass spectrometry.

- **Dynamic Allele Specific Hybridization (DASH):** In this technique specific probe is hybridized and the perfect match with the polymorphic nucleotide at specific locus is distinguished by the difference in the melting temperature of hybridized duplex.

- **Denaturing High Performance Liquid Chromatography (dHPLC):** The principle of this process is to separate the perfectly matched DNA homo duplex from the hetero duplex heaving two or more mismatch. Here PCR product of the test individual (containing genomic segment at a particular SNP locus) is mixed with the PCR product of a reference individual which contains a known polymorphic nucleotide (SNP) at that particular locus.

  The mixture than heated and cooled down to allow denaturation followed by renaturation. If homo duplex (with perfect match) is produced, it can be seen as a single peak otherwise two different peaks will be produced. In this way the SNP genotyping can be done with the help ok known reference locus.

SNP Database:

The SNP database of National Center for Biotechnology Information (NCBI) is known as **dbSNP**. Wikipedia based database for SNP is **SNPedia**. **OMIM:** This database is about the association in between human disease and polymorphism. **Rice SNP-Seek** is the data base of International Rice Research Institute (IRRI) for rice SNPs.

REFERENCES

1. Singh, B. D., & Singh, A. K. (2015). *Marker-assisted plant breeding: principles and practices*. Springer.

2. Xu, Y. (2010). *Molecular plant breeding*. Cabi.

3. Chagné, D., Batley, J., Edwards, D. and Forster, J.W. (2007) Single nucleotide polymorphisms genotyping in plants. In: Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N. (eds) *Association Mapping in Plants*. Springer.

Terms - Do not remove or change this section (It should be emailed back to us as is)

- This form is for genuine submissions related to biotechnology topics only.
- You should be the legal owner and author of this article and all its contents.
- If we find that your article is already present online or even containing sections of copied content then we treat as duplicate content - such submissions are quietly rejected.
- If your article is not published within 3-4 days of emailing, then we have not accepted your submission. Our decision is final therefore do not email us enquiring why your article was not published. We will not reply. We reserve all rights on this website.
- Your article will be published under our "Online Authors" account, but you will be clearly indicated as the original author inside the article. Your name and email address will be published. If we feel it is not feasible for us to publish your article in HTML format then we may publish it in PDF format.
- Do not violate copyright of others, you will be solely responsible if anyone raises a dispute regarding it.
- Similar to paper based magazines, we do not allow editing of articles once they are published. Therefore please revise and re-revise your article before sending it to us.
- Too short and too long articles are not accepted. Your article must be between 500 and 5000 words.
- We do not charge or pay for any submissions. We do not publish marketing only articles or inappropriate submissions.
- Full submission guidelines are located here: http://www.biotecharticles.com/submitguide.php
- Full Website terms of service are located here: http://www.biotecharticles.com/privacy.php

As I send my article to be published on BiotechArticles.com, I fully agree to all these terms and conditions.