Molecular Marker

Prepared by: Pranab Borah, Dept. of Herbal Science & Technology

Molecular marker:

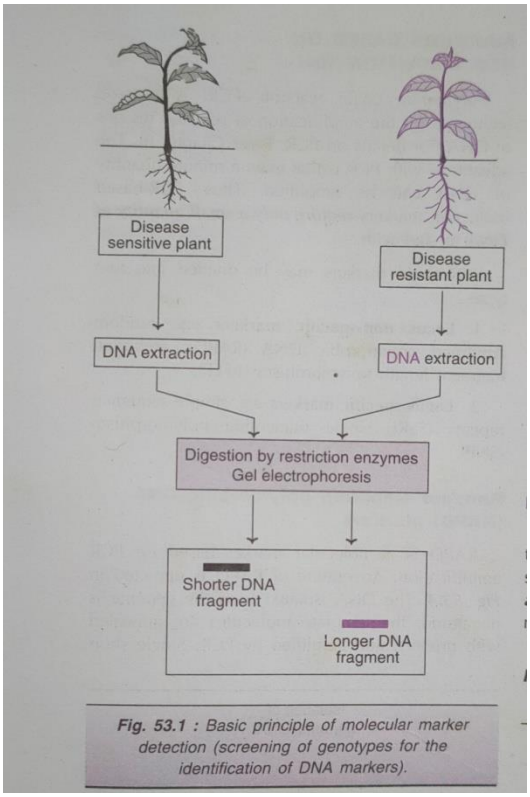

Fig. 53.1 : Basic principle of molecular marker detection (screening of genotypes for the identification of DNA markers).

Molecular marker is a DNA sequence used for chromosome mapping as it can be located at the specific site in a chromosome. A molecular marker is a heterozygocity for same type silent DNA variation, not associated with any measurable phenotypic variation.

- Molecular marker is identified as genetic marker.
- Molecular marker is a DNA or gene sequence within a recognized location on a chromosome which is used as identification tool.

- In the pool of unknown DNA or in a whole chromosome, these molecular markers help in identification of particular sequence of DNA at particular location.

Applications:

- It plays a crucial role in gene mapping by identifying the position of linked genes in the chromosome which inherited together
- It also detects any alteration in a sequence of DNA or any genetic oddity. It ascertains genes involved in genetic disorders.
- It is used to determine different characters in a DNA sequence which is used to distinguish between individuals, populations or species.

Molecular marker possess unique genetic properties (i.e. they are heritable DNA sequences and phenotypically neutral) and identified by techniques such as southern hybridization and PCR .

- **PCR based genetic markers**: **RAPD** (Random Amplified Polymorphic DNA), **AFLP** (Amplified Fragment Length Polymorphism), **SSR** (Simple Sequence Repeat), **STR** (Single Tendem Repeats), **VNTR** (Variable Number Tendem Repeat), **STS** (Sequence Tag Size), **SNP** (Single Nucleotide Polymorphism), **EST** (Expressed Sequence Tagged)

- **Hybridization based genetic markers**: RFLP (Restriction Fragment Length Polymorphism)

Molecular markers also can be classified as-

- **1st generation markers** (at the time of 1980-1990): RFLP, RAPD
- **2nd generation markers** (1990-2000): AFLP, SSR, STR, VNTR, STS
- **3rd generation markers** (After 2000): SNP, EST

Quality for a good genetic marker:

Molecular marker must possess the following desirable properties-

- It must be polymorphic so that diversity must be measured.
- It should be evenly distributed throughout the genome.
- It should be easily and fast detected.
- It must distinguish the homozygotes and heterozygotes.

1. Random Amplified Polymorphic DNA (RAPD):

- RAPD was developed by Welsh and McClelland along with Williams in 1990.
- It is pronounced as 'rapid'.
- It is based on PCR assay and it doesn't need require any prior sequencing of DNA.
- This procedure uses short arbitrary primer of 8-12 bp that randomly amplifies the region of DNA.
- This primer serves as both forward and reverse primer.
- This reaction proceeds when a single primer anneals to the genomic DNA at two distinct sites on the complementary strand of DNA template.
- The amplification of segment of DNA depends on the positions complimentary to the primers' sequence.
- The fragments obtained from RAPD are between 0.2 to 5.0kb and can be viewed by using agarose gel electrophoresis stained by ethidium bromide or with the help of polyacrylamide gel electrophoresis.
- If any mutation occurs in the primer binding region then no any PCR product will be produced, yielding a distinct pattern of amplified DNA segments on the gel.

  Application:
  - Distinct pattern of amplification is seen in different samples. This is why RAPD can be used for studying polymorphism.
  - RAPD is applicable for the mapping of genome, analyzing linkage, and individual specific genotyping.

- RAPD markers are dominant in nature so it has restrictions for mapping purpose.
- RAPD is strictly laboratory dependent so it requires sensitivity.

- **Demerits:**
  - It has demerits as poor reproducibility, yields faint products, problems occur in band scoring.

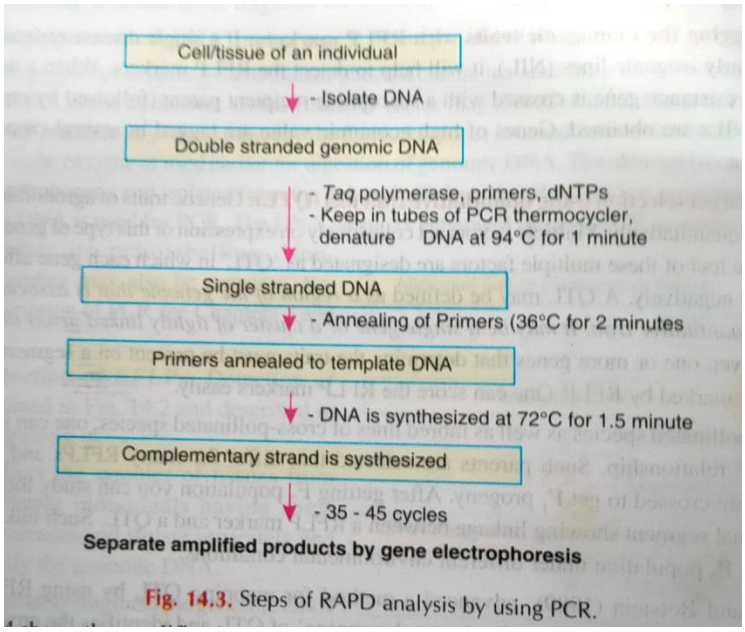

Fig. 14.3. Steps of RAPD analysis by using PCR.

2. Restriction Fragment Length Polymorphism (RFLP):

- It was one of the first methods used for the analysis of DNA in various fields such as forensic science.
- It is a hybridization based technique.
- It was invented by **Alec Jeffreys,** an English scientist in 1984 during his research in genetic diseases.
- RFLP uses particular restriction endonuclease enzymes that cut at its specific site yielding fragments of various lengths along with the fragment of interest.
- The length of the distinct fragments is determined by using blotting, now replaced with sequencing.
- RFLP markers are largely locus-specific and are co-dominant in nature due to the nature of restriction endonuclease used.
- **Applications:**
  - RFLP was one of the first techniques applied for genetic fingerprinting/profiling.
  - It is used for identification of inherited diseases, carrier of that diseases, genetic mapping, and heterozygous detection.
  - The molecular basis of the RFLP is that any point mutations as such deletions, substitutions and insertions or alterations like duplications, inversions within the genome can eliminate or form new restriction sites. These alterations in genome can

be detected by analyzing fragments of variable length, digested with restriction endonuclease enzyme

- **Demerits:**
  - requires relatively large DNA sample
  - laborious and tedious process
  - sensitivity and more precautions for contamination required

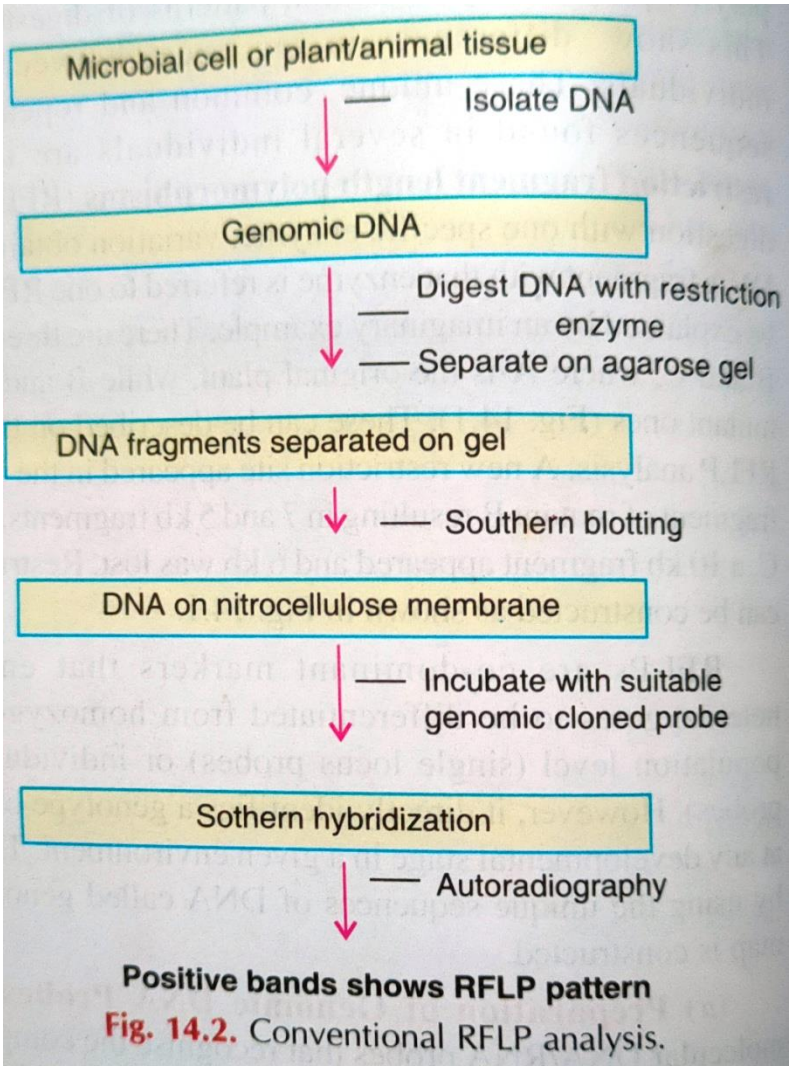

Fig. 14.2. Conventional RFLP analysis.

3. Amplified Fragment Length Polymorphism (AFLP):

- Zabeau and Vos invented the AFLP technique in 1993.
- AFLP was originally developed by the KeyGene in 1990.
- It is a PCR based technique for fingerprinting. It includes both PCR and RFLP.
- The basis of AFLP is the amplification of selected fragments followed by restriction digestion of whole genomic DNA of specific organism.
- The steps for the AFLP are as follows:

- DNA extraction and its restriction digestion followed by ligation with the short adaptor sequences.
- Amplification of restricted fragments by PCR
- Analysis of results in gel electrophoresis or PAGE followed by autoradiography.
- **Applications:**
  - AFLP has its ability for rapid generation of marker fragments for any organism without prior sequencing of DNA is required.
  - Also, it needs only small fragments of starting template DNA relative to RAPD and ISSR (inter-simple sequence repeats) and has much higher reproducibility.
  - AFLP is largely used for crop improvement programs, parentage and genomic interpretation of various crop species.
- **Demerits:**
  - AFLP require large DNA samples and require purification

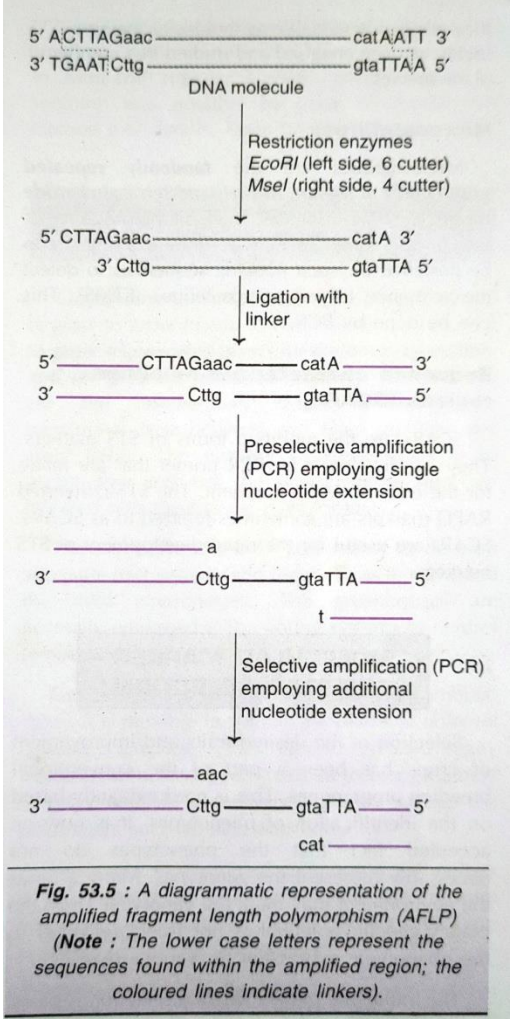

Fig. 53.5 : *A diagrammatic representation of the amplified fragment length polymorphism (AFLP)* (***Note :*** *The lower case letters represent the sequences found within the amplified region; the coloured lines indicate linkers).*

4. Inter Simple Sequence Repeat (ISSR) markers:

- Inter simple sequence repeat (ISSR) technique is PCR based method.

- It was reported by Ztetikiewicz et al. in 1994.
- The ISSR markers are developed by PCR amplification of DNA segments between 2 similar microsatellites repeat regions by use of single primer consisting of microsatellite core regions.
- The primers can be usually 16-25 bp long, and unattached or attached at 3′ or 5′ end.
- **Applications:**
  - It is simple, rapid and economical like the RAPD technique and has higher reproducibility compared to RAPD because of longer primer length.
  - ISSR doesn't require previous knowledge of genome for analysis and is highly polymorphic marker.
  - ISSR are used for identification of genetic diversity, phylogenetic analysis, to detect proximity of cultivars and to determine somaclonal variations in plants.
  - Due to the simple set up of ISSR, it is applicable for studying gene mapping, gene tagging, distinct strain identification, and parental recognition.
- **Demerits:**
  - ISSR has less reproducibility and non-homology of identical sized fragments due to multi locus feature.

5. Microsatellites or simple sequence length polymorphisms (SSLPs):

- Microsatellite was termed by Jeffery et al. in 1985.
- Microsatellites or simple sequence repeated (SSR) loci are PCR based markers which needs previous knowledge of gene sequence.
- In literature it is referred to as **variable number of tandem repeats (VNTRs)** or **simple sequence length polymorphisms (SSLPs**) or **sequence tagged microsatellites (STMS).**
- They are dispersed throughout the nuclear genomes in eukaryotes and to a few extents in prokaryotes.
- Microsatellite primers are short tandem repeats (STRs), or simple sequence repeats (SSRs), having 1-6 base pair long sequences repeated several times.
- Usually microsatellites are repeated less than 100 times.
- Microsatellites can be recognized by constructing a small-insert genomic library followed by screening of library and sequencing of positive clones.
- Microsatellites are used as markers for studying gene mapping, closeness among the species, and population genetics.
- The amplification of tandem arrays followed by visualization in gel helps to detect variation in DNA length.
- The main cause for the variation in DNA length is polymerase diminution during DNA replication, or slipped strand mispairing.
- **Applications:**
  - Microsatellite consists of co-dominance of alleles and requires low quantities of DNA templates.
  - It has high reproducibility and is economical in nature.

- The screening of microsatellite variation can be automated.

- **Demerits:**
  - Assay is costly if sufficient primer sequences for the species of interest are not available.
  - An error in genotype scoring occurs if alterations are seen in primer annealing sites.
  - chances of homoplasy (some characters are present in more than one species but not present in their common ancestor because of convergence evolution)

7. Expressed Sequence Tags (ESTs):

- ESTs are molecular markers synthesized by partial sequencing of random cDNA clones.
- Once cDNA is synthesized, then a few hundred nucleotides (500-800bp) from either end (3' or 5' end) is sequenced to create two different kinds of ESTs.
- Sequencing of the 5' end of cDNA produces a 5′ EST which usually codes for a protein.
- These regions tend to be conserved across species and do not change much within a gene family.
- Sequencing of 3' end of cDNA produces a 3′ EST which is non-coding, or untranslated regions (UTR) and it is less conserved among species.
- Therefore, ESTs are sub-sequence of cDNA which represents tag for the entire cDNA.
- **Applications:**
  - EST is used for the whole genome sequencing and studying gene of interest.
- In addition to it, it is used for cloning gene of interest and gene mapping.

9. Single nucleotide polymorphism (SNP):

- SNP was invented by Lander in 1996.
- SNP is formed when any alteration/mutation occurs in single nucleotide (A, T, C, or G).
- The point mutation as such substitutions, insertions or deletions in single nucleotide it represents SNP.
- SNPs are based on hybridization of detected DNA fragments with SNP chips (DNA probe arrays) and the SNP allele is named with respect to the hybridization results.
- **Applications:**
  - SNPs are widely used in biomedical research for comparing the case and control groups of disease.
  - It is also used in studying phylogenetics, genetic variation etc.
- **Demerits:**
  - The information obtained is low as compared to microsatellites and therefore large numbers of markers and complete genome sequencing is needed