

DNA based Molecular markers

A molecular marker (identified as genetic marker) is a fragment of DNA that is associated with a certain location within the genome. Molecular markers are used in molecular biology and biotechnology to identify a particular sequence of DNA in a pool of unknown DNA.

What is an ideal DNA marker?

- 1) Highly polymorphic nature: It must be polymorphic as it is polymorphism that is measured for genetic diversity studies.
- 2) Codominant inheritance: determination of homozygous and heterozygous states of diploid organisms.
- 3) Frequent occurrence in genome: A marker should be evenly and frequently distributed throughout the genome.
- 4) Easy access (availability): It should be easy, fast and cheap to detect.
- 5) Easy and fast assay
- 6) High reproducibility
- 7) Easy exchange of data between laboratories
- 8) Gene based molecular markers

Non- PCR or Hybridization Based Markers

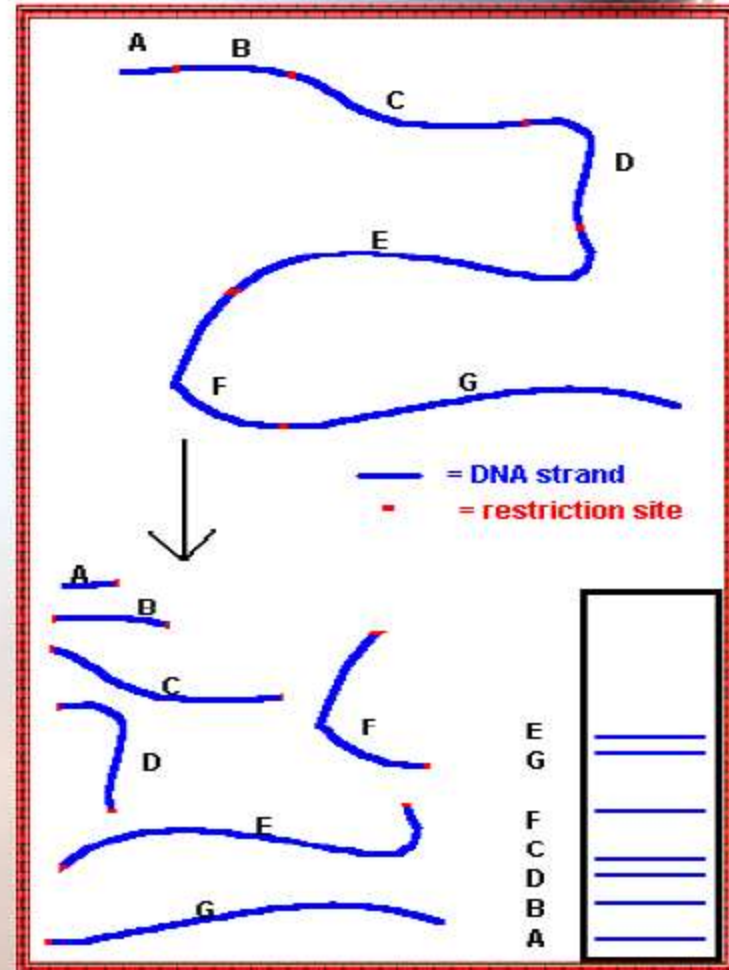
“The variation in the length of DNA fragments produced by specific Restriction Endonucleases from genomic DNA s of two or more individuals of a species is called hybridization and markers produced by this technique are called hybridization based molecular markers.”

This type of hybridization marker includes-

RFLP

RFLPs :

Genetic markers resulting from the variation or change in the length of defined DNA fragments produced by digestion of the DNA sample with restriction endonucleases



RFLPs

(restriction fragment length polymorphisms)

Electrophoretic comparison of the size of defined restriction fragments derived from genomic DNA

- 1. Isolate high quality DNA**
- 2. Digest with a combination of restriction enzymes**
- 3. Fractionate digested samples by electrophoresis**
- 4. Transfer fragments to membrane**
- 5. Hybridize with radioactively labeled DNA probe(s); detect by autoradiography. Can also use non-radioactive labeling systems**

Advantages of RFLP-

They are co- dominant.

- Measure variation at the level of DNA sequence, not protein sequence.
- RFLP loci are very large so even very small segments of chromosomes can be mapped and also study phylogenetic relationship.
- Very reliable for linkage analysis and for detecting coupling phase of DNA molecules.

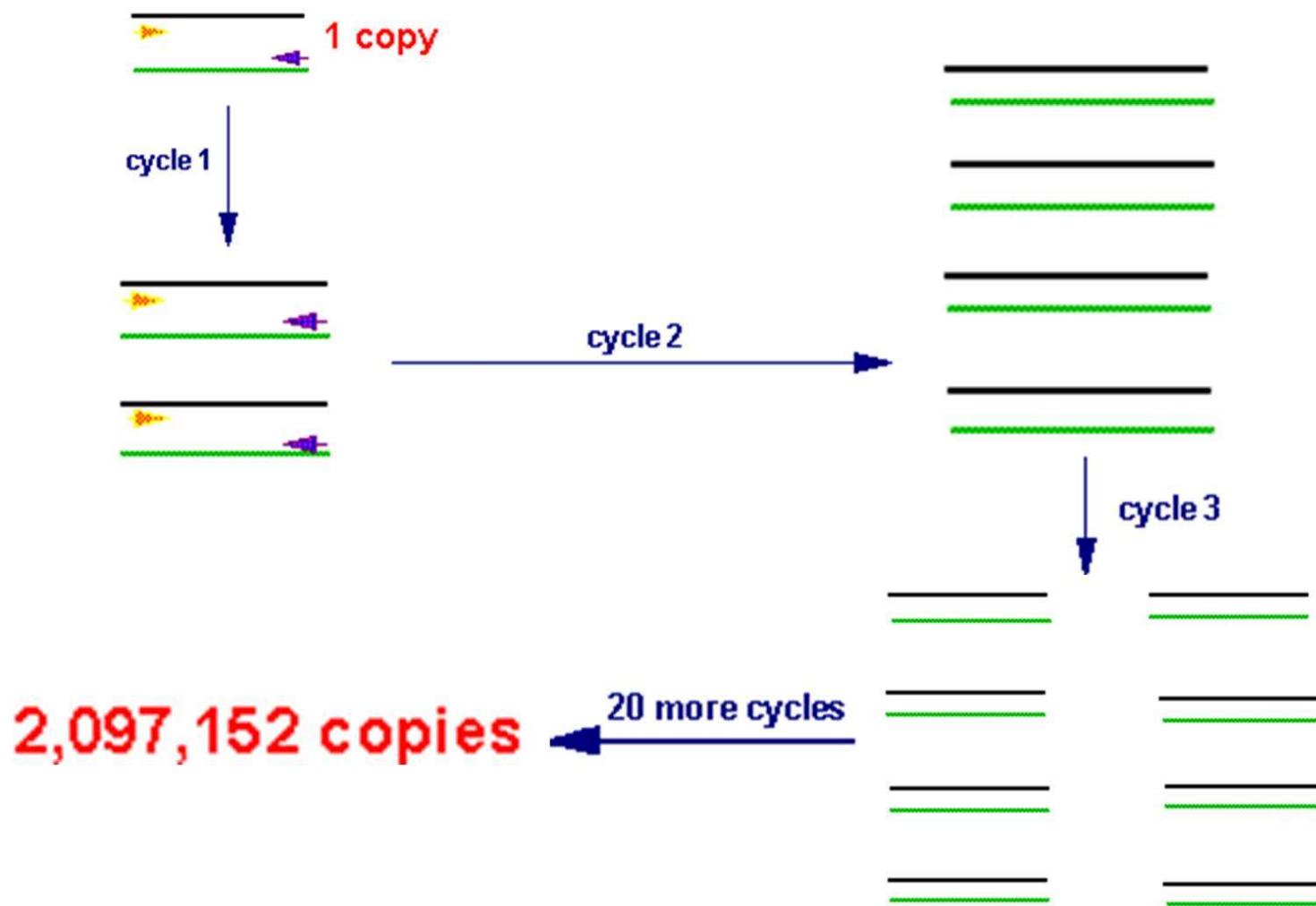
Disadvantages of RFLP-

- Requires relatively very large amount of DNA.
- Requirement of radioactive probe makes the analysis expensive and hazardous.
- They are not useful for detecting single base change or point mutations.
- It is time consuming, laborious, and expensive.
- The level of polymorphism is low.



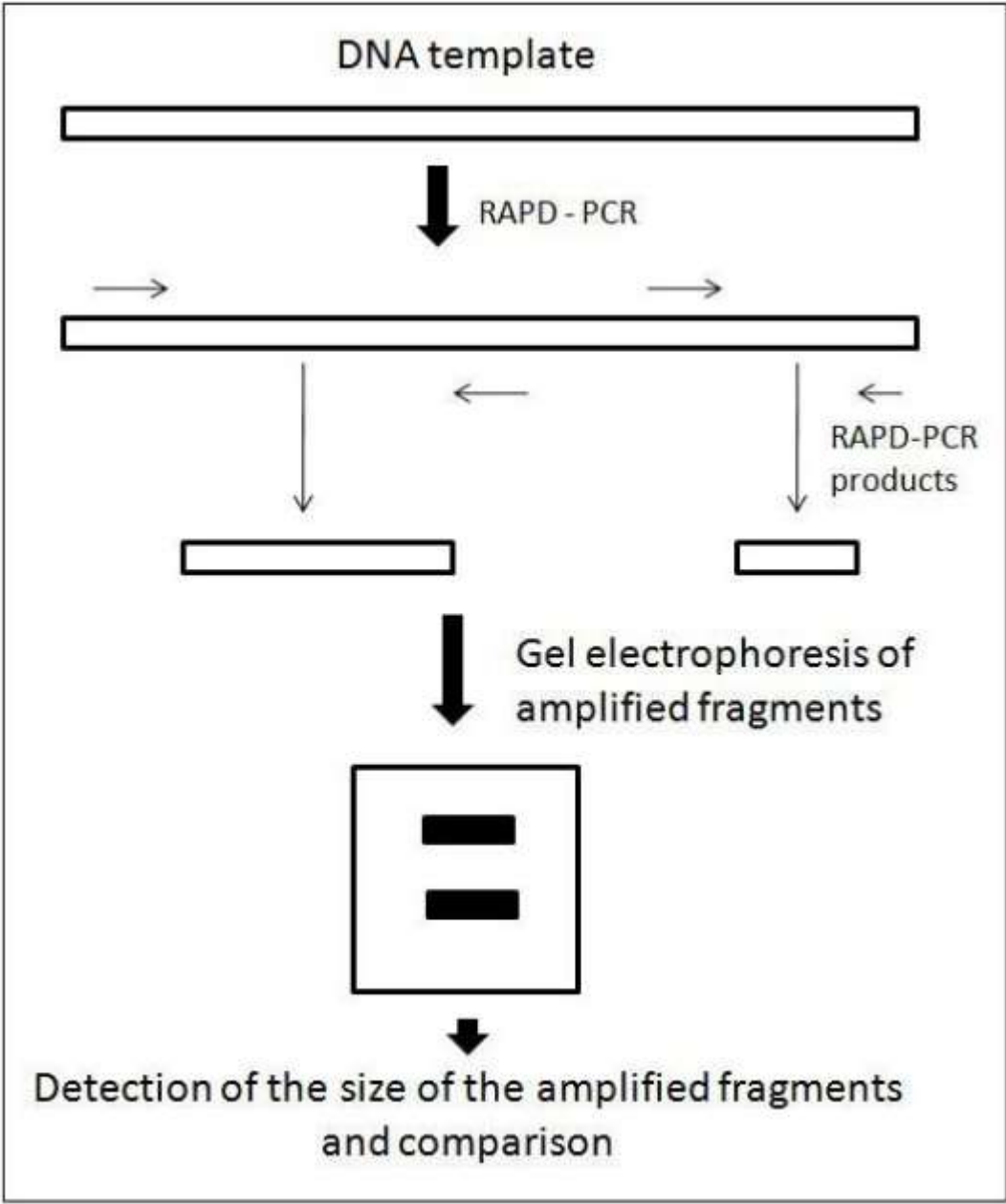
PCR based techniques

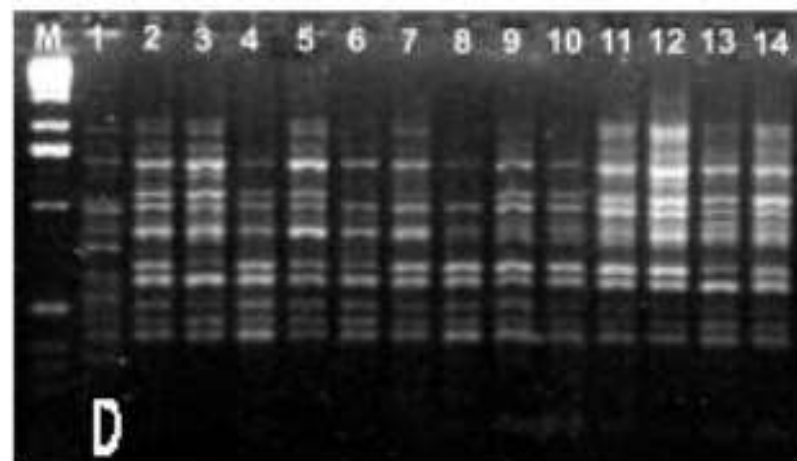
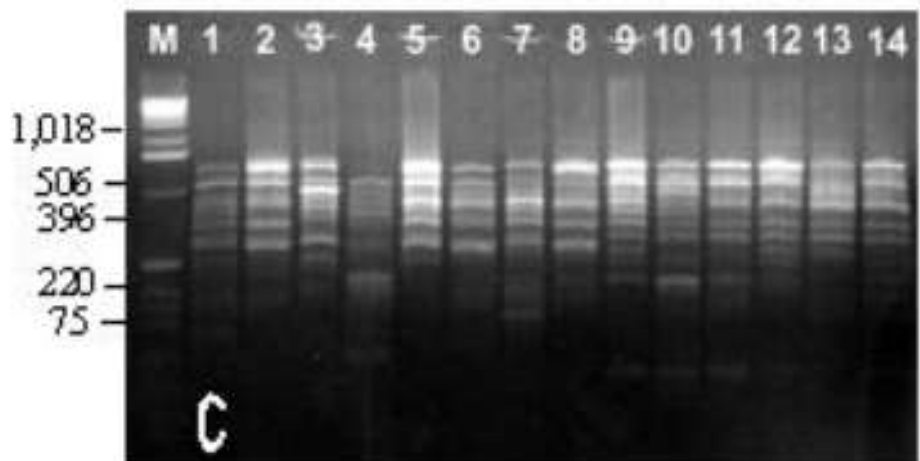
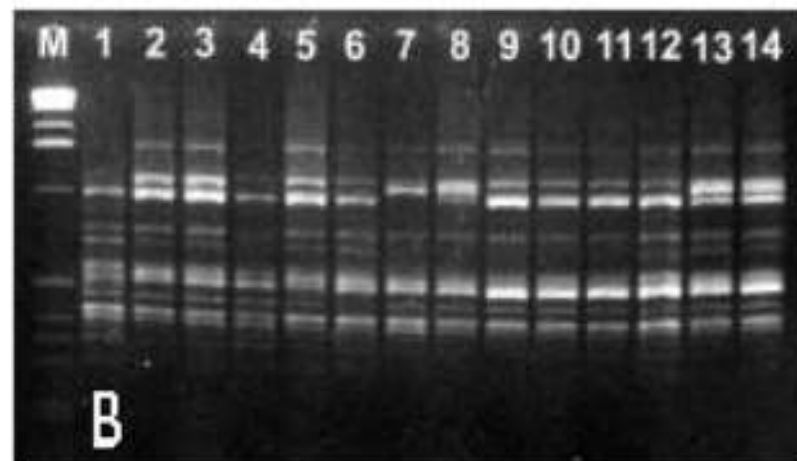
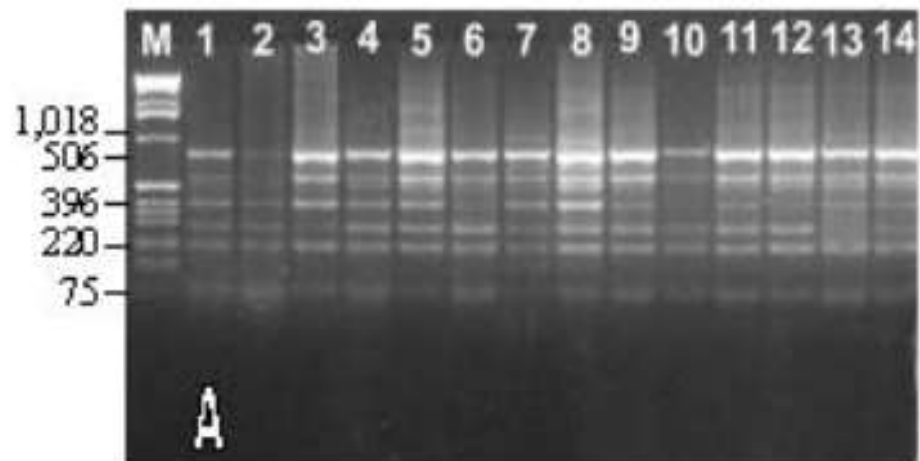
(RAPD, ISSR, SSR, AFLP, EST ,SCoT)



Random Amplified Polymorphic DNA (RAPD)

- ➔ **R**andomly Amplified Polymorphic DNA (RAPDs) are genetic markers resulting from PCR amplification of genomic DNA sequences recognized by **len-mer random** primers of arbitrary nucleotide sequence (Williams et al., 1990).
- ➔ **R**APDs are **dominant** markers that require no prior knowledge of the DNA sequence, which makes them very suitable for investigation of species that are not well known (Williams et al. 1993).





Advantages:

- They are quick and easy to assay.
- Because PCR is involved, only low quantities of template DNA are required.
- No sequence data for primer construction are needed.
- RAPDs have a very high genomic abundance and are randomly distributed throughout the genome.

Disadvantages:

- Low reproducibility.
- RAPD analyses generally require purified, high molecular weight DNA, and precautions are needed to avoid contamination of DNA samples because short random primers are used that are able to amplify DNA fragments in a variety of organisms.
- The inherent problems of reproducibility make RAPDs unsuitable markers for transference or comparison of results among research teams working in a similar species and subject.
- RAPD markers are not locus-specific, band profiles cannot be interpreted in terms of loci and alleles (dominance of markers), and similar sized fragments may not be homologous.

SSR (Simple sequence repeat)

Simple sequence repeat or Microsatellites, also known as Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STRs), are repeating sequences of 2-6 base pairs of DNA. It is a type of Variable Number Tandem Repeat (VNTR). Microsatellites are typically co-dominant.

Sequence

Primer

ACTGTCG**ACACACACACACAC**GCTAGCT (AC)₇
TGACAGC**TGTGTGTGTGTGT**GCGATCGA

ACTGTCG**ACACACACACACACAC**GCTAGCT (AC)₈
TGACAGC**TGTGTGTGTGTGTGT**GCGATCGA

ACTGTCG**ACACACACACACACACAC**GCTAGCT (AC)₁₀
TGACAGC**TGTGTGTGTGTGTGTGT**GCGATCGA

ACTGTCG**ACACACACACACACACACAC**GCTAGCT (AC)₁₂
TGACAGC**TGTGTGTGTGTGTGTGTGT**GCGATCGA

Table 2 Classification of microsatellites

(A) Based on the number of nucleotides per repeat

Mononucleotide (A)_n

Dinucleotide (CA)_n

Trinucleotide (CGT)_n

Tetranucleotide (CAGA)_n

Pentanucleotide (AAATT)_n

Hexanucleotide (CTTTAA)_n (n = number of variables)

(B) Based on the arrangement of nucleotides in the repeat motifs (Weber 1990; Jarne and Lagoda 1996; Wang et al. 2009a)

Pure or perfect or simple perfect (CA)_n

Simple imperfect (AAC)_n ACT (AAC)_n + 1

Compound or simple compound (CA)_n (GA)_n

Interrupted or imperfect or compound imperfect
(CCA)_n TT (CGA)_n + 1

(C) Based on location of SSRs in the genome

Nuclear (nuSSRs)

Chloroplastic (cpSSRs)

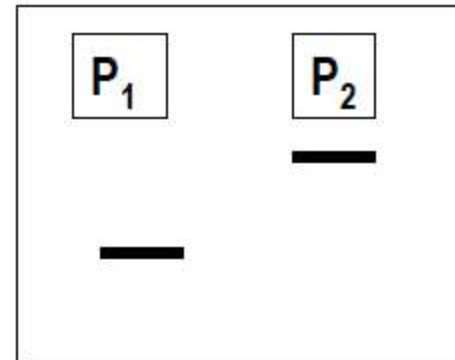
Mitochondrial (mtSSRs)

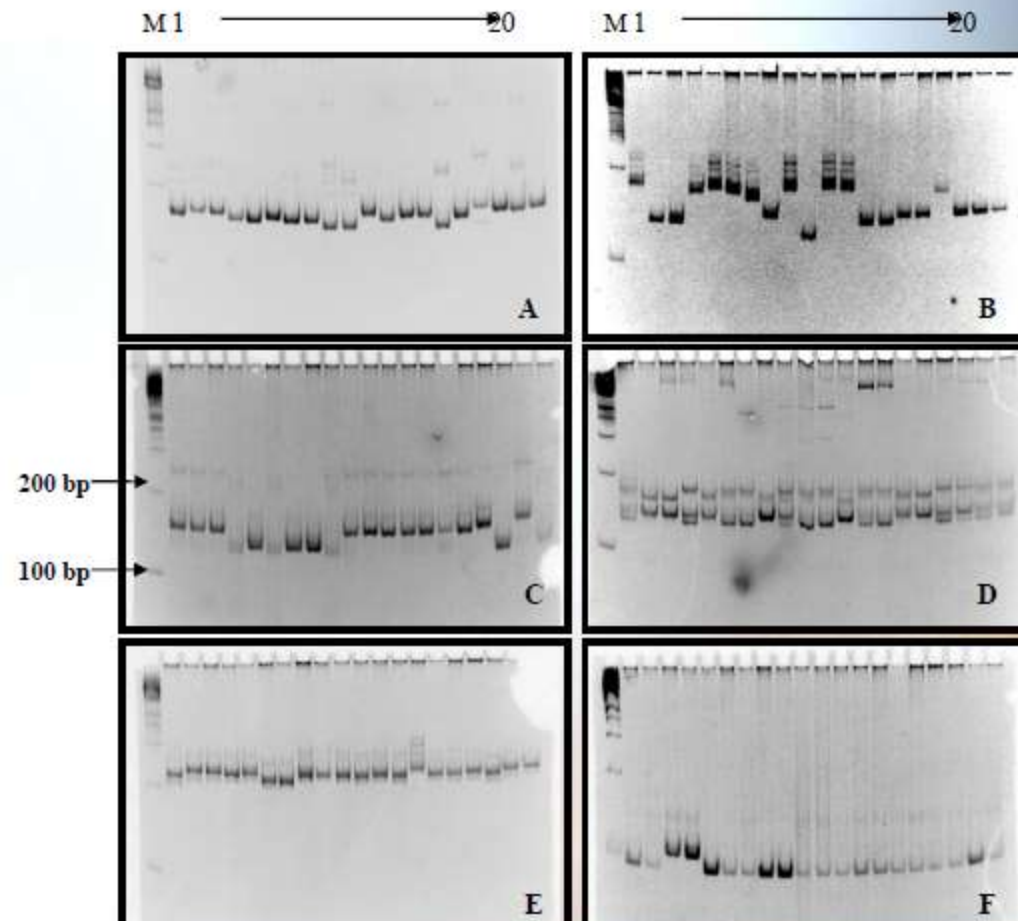
SSR polymorphisms

P₁ AATCCGGACTAG **CTTCTTCTTCTTCTTCTT** TAGCGAATTAGG

P₂ AAGGTTATTT **CTTCTTCTTCTTCTTCTTCTTCTT** TAGGCTAGGCG

Gel configuration





Photograph of EtBr stained polyacrylamide gels of polymorphic **SSR** products from 20 maize inbred lines as detected by SSR primers (A: M28, B: M27, C: M25, D: M20, E: M18 and F: M22). M (100bp DNA ladder)

Advantage-

- SSR markers tend to be **highly polymorphic**.
- The genotyping throughput is high.
- This is a **simple PCR assay**. Many SSR markers are multi-allelic and highly polymorphic.
- Most SSRs are **co-dominant** and locus specific.
- **No special equipment** is needed for performing SSRs assays; however, special equipment is needed for some assay methods,
- Start-up costs are low for manual assay methods (once the markers are developed). SSR assays can be performed using very small DNA samples (~100 ng per individual).
- SSR markers are easily **shared between laboratories**.

Disadvantages:

- The development of SSRs is **labor intensive**
- SSR marker **development costs are very high**

Application of SSRs

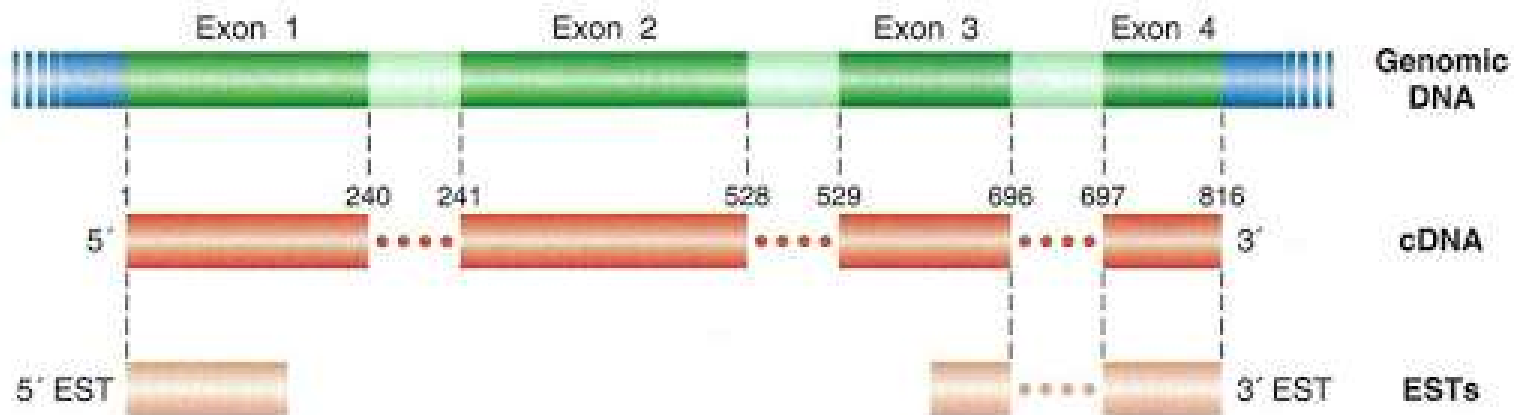
- Genetic diversity analysis (measurement of genetic similarity or differences among plant species)
- Cultivar identification
- Phylogenetic relationships
- Gene mapping and identification of quantitative trait loci (QTLs) which can lead to the identification of candidate genes for the trait of interest
- Marker-assisted selection (MAS)
- sex determination of dioecious plants

EST-SSR

An expressed sequence tag (EST) is a short sub sequence of a cDNA sequence

Expressed sequence tags (ESTs), obtained by partial random sequencing of cDNA libraries, are 300–500 nucleotide long mRNA sequences from many of the genes expressed in a sample from an organism

SSRs present in EST region is known as ESR-SSRs or genic microsatellites.



Inter-Simple Sequence Repeats (ISSR)

- ➔ The generation of ISSR markers involve PCR amplification of DNA using a single primer composed of a microsatellite repeated sequence and in some cases primer also contains 1-4 base anchor at either 3' or 5' or at both ends, which target a subset of 'simple sequence repeats' (SSRs) and amplify the region between two closely spaced and oppositely oriented SSRs (Fang et al., 1997; Fang and Roose, 1997; Moreno et al., 1998).
- ➔ ISSR technique permits the detection of polymorphisms in microsatellites and inter-microsatellites loci without previous knowledge of the DNA sequence (Moreno et al., 1998).

Start Codon Targeted Polymorphism (SCoT)

PCR based technique developed by Collard and Mackill in 2009.

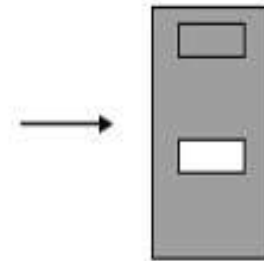
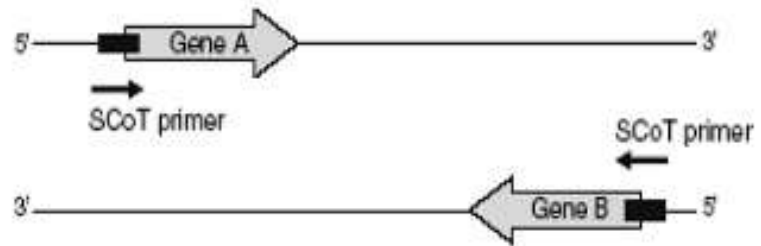
Simple and novel DNA marker technique, uses 18-mer single primer in PCR and an annealing of 50 °C.

PCR products are resolved using standard agarose gel electrophoresis.

SCoT based on the short conserved region flanking the start codon (ATG) in plant genes.

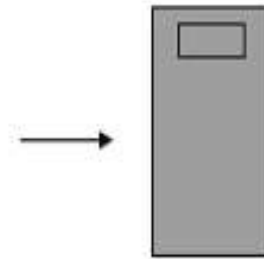
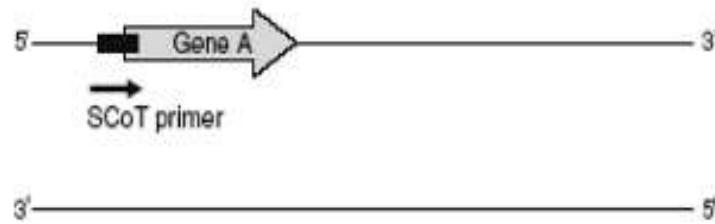
This technique has been demonstrated high polymorphic and efficient, and successfully utilized in a number of plant species for cultivar identification and genetic diversity analysis

GENOTYPE #1



PRESENT

GENOTYPE #2



ABSENT

Advantages and applications of SCoT

- Technically simple
- SCoT employs longer primers (18-mer) producing high polymorphism which is reproducible
- There is requires no prior sequence information and the polymorphism is correlated to functional genes and their corresponding traits
- It could be used directly in marker-assisted breeding programmes.
- Point mutations affecting the primer binding region generate polymorphism with SCoT primers
- It has been used to evaluate genetic relationships and genetic diversity analysis in plants.
- Since, it is a gene targeted marker, therefore it is useful in QTLidentification and gene mapping.

AFLP

(Amplified Fragment Length Polymorphisms)

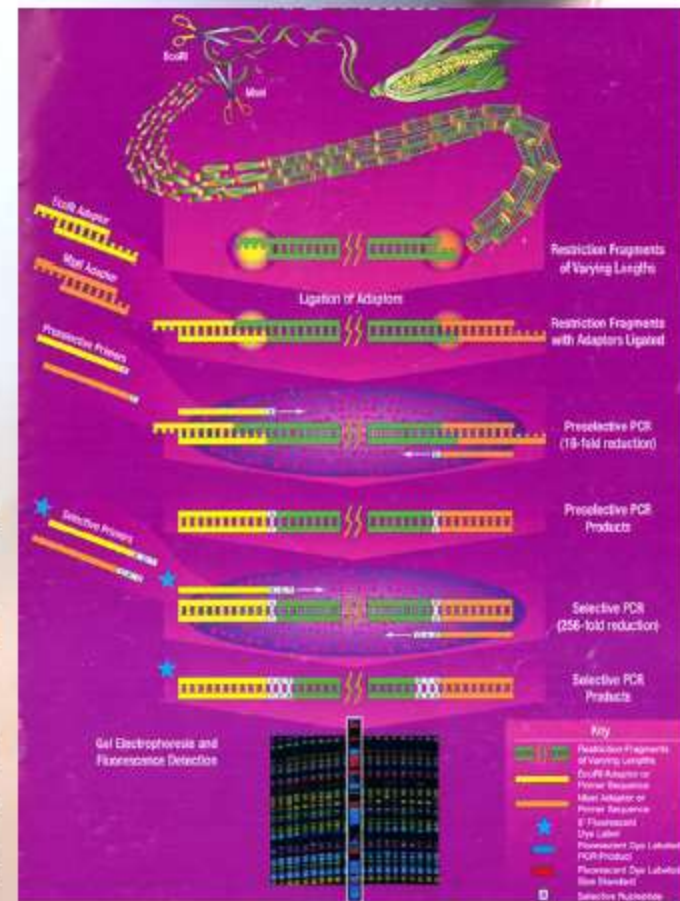
- **A combination of PCR and RFLP**
- **Informative fingerprints of amplified fragments**

Amplified Fragment Length Polymorphism (AFLP)

AFLP technology is a DNA fingerprinting technique that combines RFLP and PCR. It is based on the selective amplification of a subset of genomic restriction fragments using PCR.

AFLP process

1. **Digest** genomic DNA with restriction enzymes
2. **Ligate** commercial adaptors (defined sequences) to both ends of the fragments
3. **Carry out PCR** on the adaptor-ligated mixture, using primers that target the adaptor, but that vary in the base(s) at the 3' end of the primer.

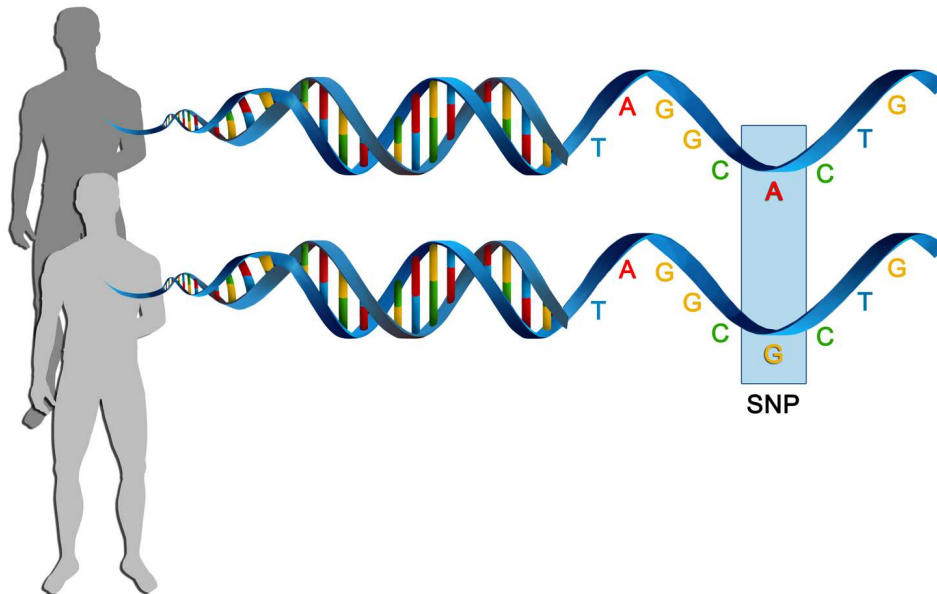


Single-nucleotide polymorphism (SNP)

A single-nucleotide polymorphism (SNP) is a substitution of a single nucleotide that occurs at a specific position in the genome

Single-nucleotide polymorphisms may fall within coding sequences of genes, non-coding regions of genes, or in the intergenic regions (regions between genes).

It determines genetic variation in an organism which is associated with a disease or trait.



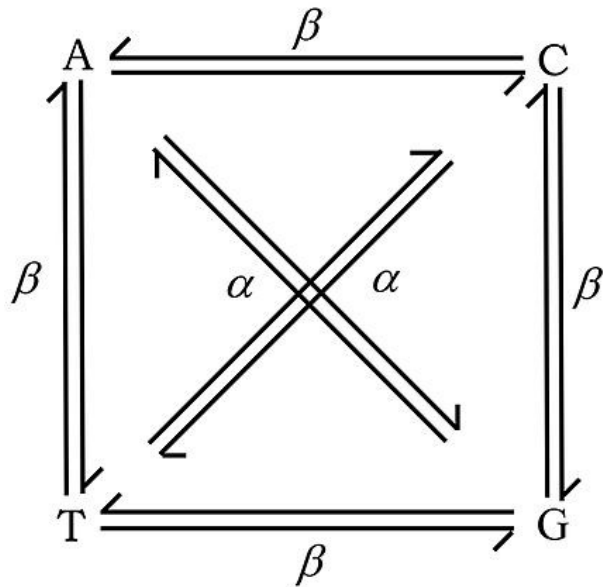
Point mutation: major cause of SNP

A point mutation or substitution is a genetic mutation where a single nucleotide base is changed, inserted or deleted from a sequence of DNA or RNA

Transitions are replacement of a purine base with another purine or replacement of a pyrimidine with another pyrimidine.

Transversions are replacement of a purine with a pyrimidine or vice versa.

Transition mutations are about ten times more common than transversions.



Transitions (Alpha) and transversions (Beta)

In human beings, 99.9 percent bases are same.

Remaining 0.1 percent makes a person unique. – Different attributes / characteristics / traits

These variations can be

Harmless (change in phenotype)

(The severity of illness and the way the body responds to treatments are also manifestations of genetic variations. For example, a single-base mutation in the APOE (apolipoprotein E) gene is associated with a lower risk for Alzheimer's disease.)

Harmful

(diabetes, cancer, heart disease, Huntington's disease, and hemophilia)

(SNPs pinpoint differences in our susceptibility to a wide range of diseases (e.g. sickle-cell anemia, β -thalassemia and cystic fibrosis result from SNPs)

- SNPs occur with a very high frequency – about 1 in 1000 bases to 1 in 100 to 300 bases.
- SNPs close to particular gene acts as a marker for that gene.
- SNPs in coding regions may alter the protein structure made by that coding region.
- In many organisms most polymorphisms result from changes in a single nucleotide position (point mutations), has led to the development of techniques to study single nucleotide polymorphisms (SNPs).
- SNPs and flanking sequences can be found by sequencing or through the screening of readily available sequence databases.
- Once the location of SNPs is identified and appropriate primers designed for cultivar discrimination in crops where it is difficult to find polymorphisms.

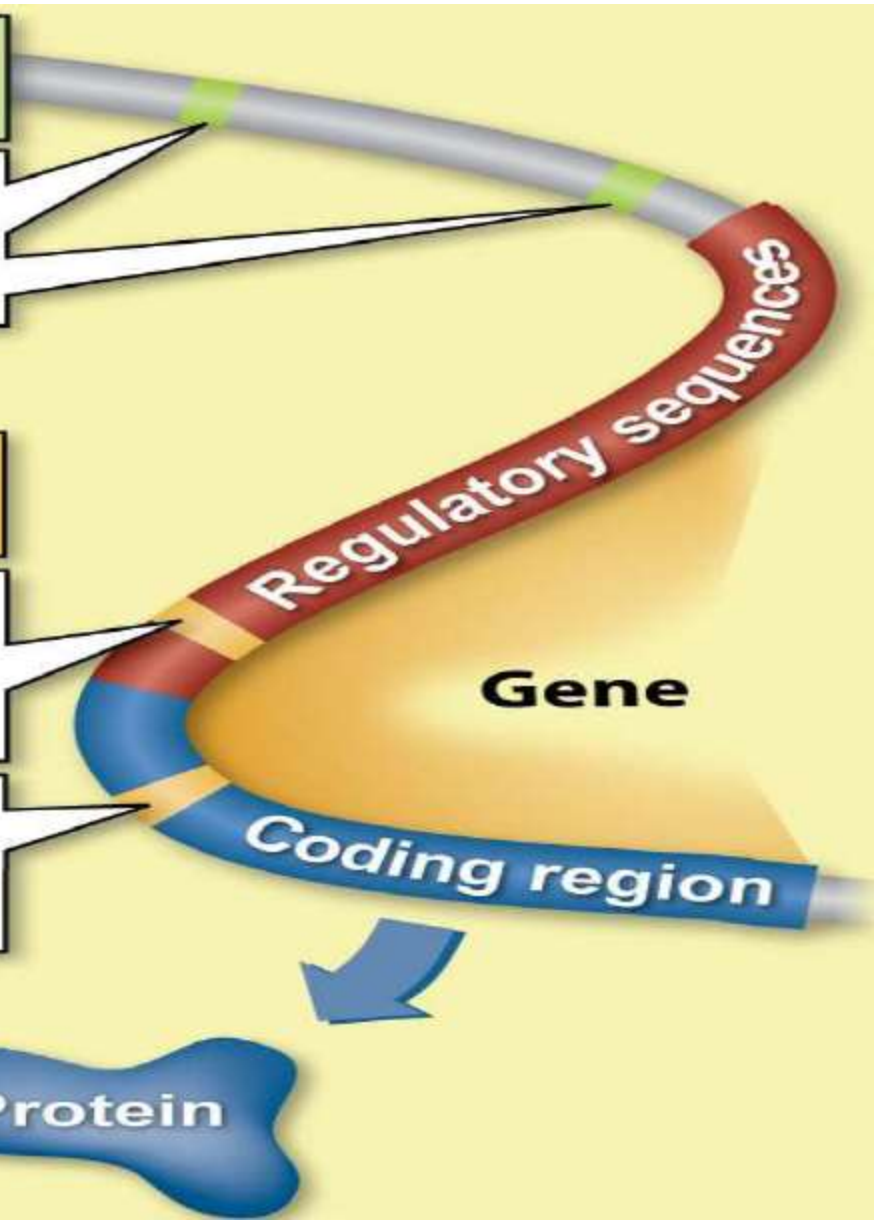
Linked SNPs
outside of gene

no effect on protein production or function

Causative SNPs
in gene

Non-coding SNP:
● changes amount of protein produced

Coding SNP:
● changes amino acid sequence

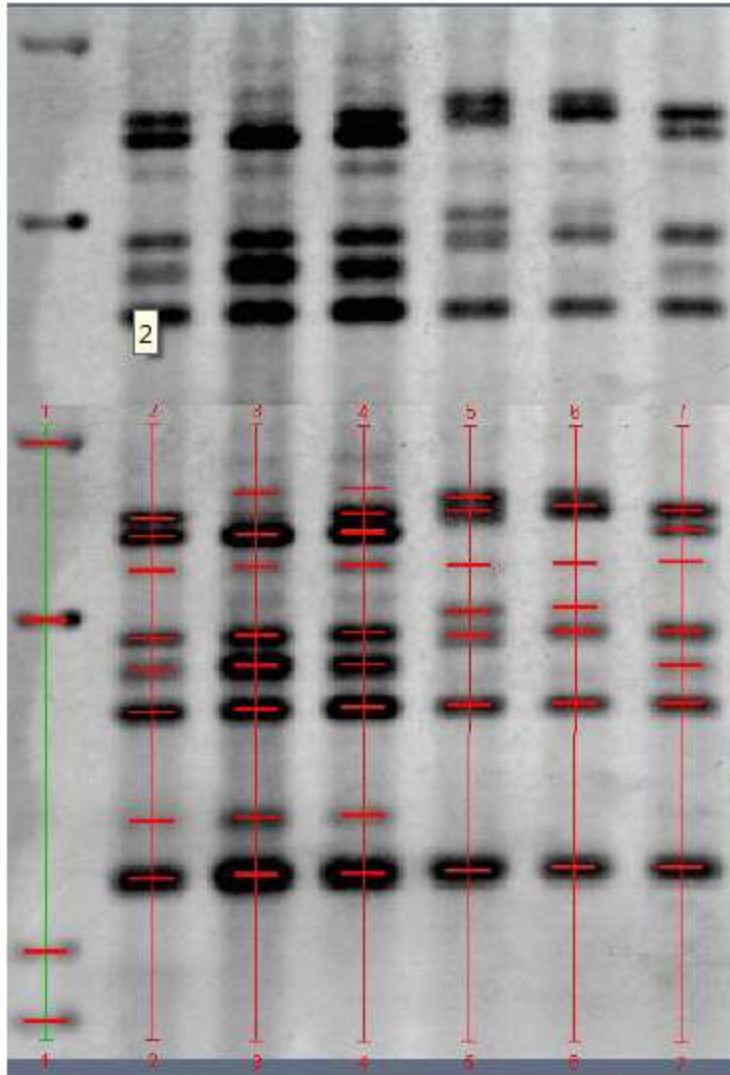


Advantages-

- **SNP markers are useful in gene mapping.**
- **SNPs help in detection of mutations at molecular level.**
- **SNP markers are useful in positional cloning of a mutant locus.**
- **SNP markers are useful in detection of disease causing genes.**

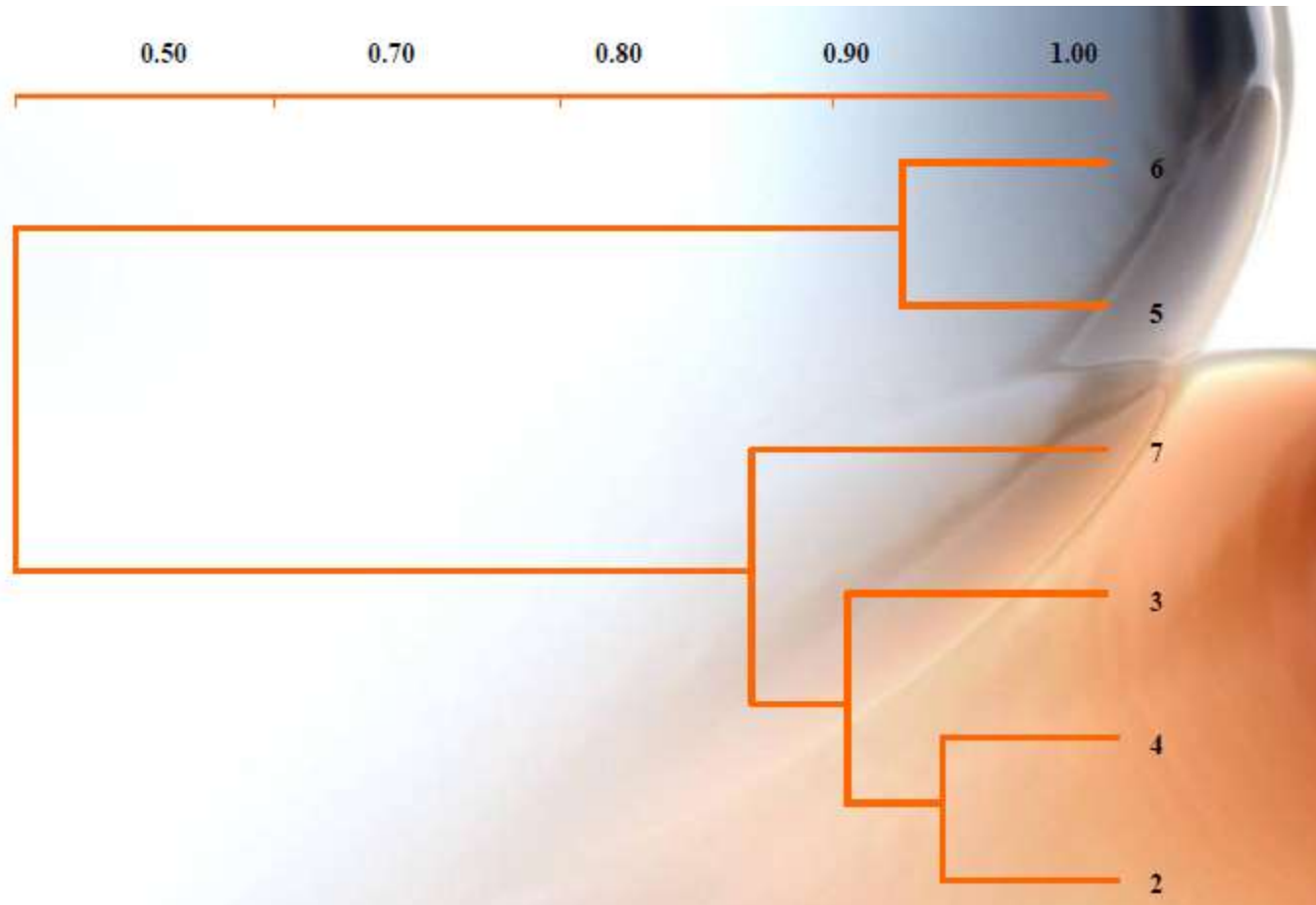
Table 1 Important features of different types of molecular markers

Features	Molecular markers			
	EST-SSRs	SSRs	RFLPs	RAPDs/AFLPs/ISSRs
Need for sequence data	Essential	Essential	Not required	Not required
Level of polymorphism	Low	High	Low	Low-moderate
Dominance	Co-dominant	Co-dominant	Co-dominant	Dominant
Interspecific transferability	High	Low-moderate	Moderate-high	Low-moderate
Utility in Marker assisted selection	High	High	Moderate	Low-moderate
Cost and labour involved in generation	Low	High	High	Low-moderate



2	3	4	5	6	7
0	1	1	0	0	0
0	0	0	1	1	0
1	0	1	1	1	1
1	1	1	0	0	1
1	1	1	1	1	1
0	0	0	1	1	0
1	1	1	1	1	1
1	1	1	0	0	1
1	1	1	1	1	1
1	1	1	0	0	0
1	1	1	1	1	1

Scoring of bands



Dendrogram constructed with UPGMA cluster analysis of marker data showing the genetic relationships among the different samples.