Introduction to Molecular Markers

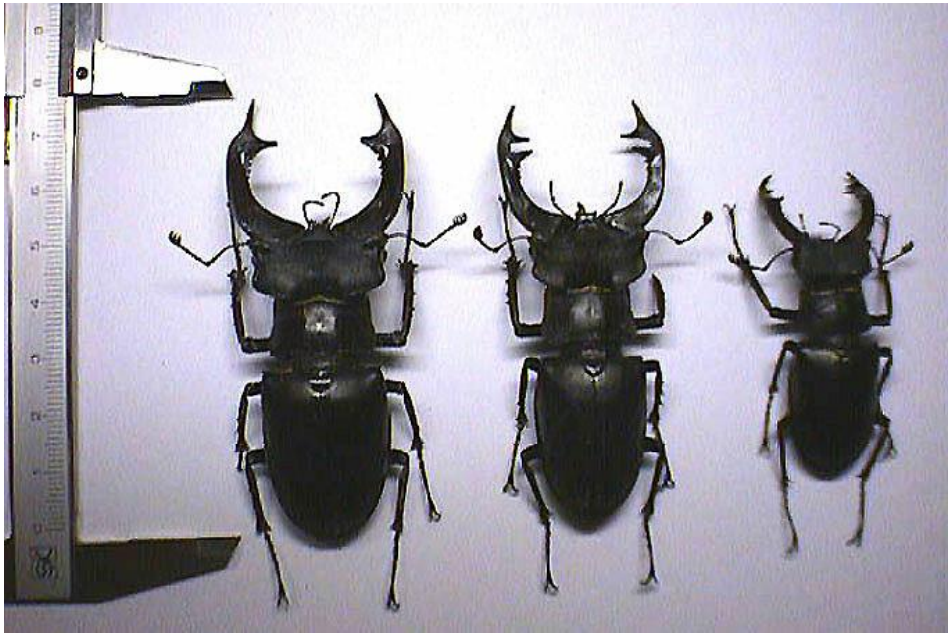
CLASSES OF MARKERS

Class	Level of Analysis
Morphological	Phenotype
Biochemical	Gene Product
Molecular	DNA

Within-population variation: Polymorphism

Happy-face spiders

Within population sex-linked visible polymorphism



Stag beetle

Using visible polymorphisms

Advantages:

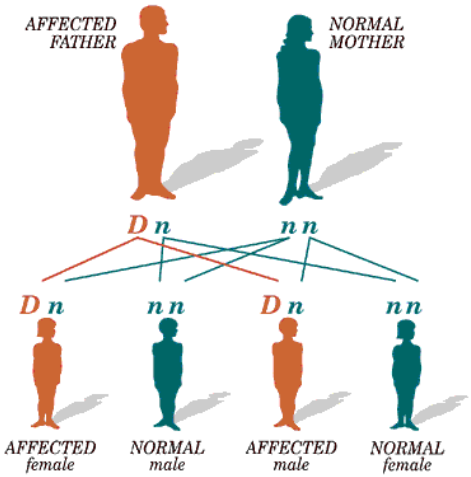
- inexpensive to score,
- amenable to experiments in natural populations

Disadvantages:

- Visible polymorphisms relatively rare.
- Most genetic variation not so easily observed.
- Genetic basis of variation can be complex, and is not necessarily easy to determine.

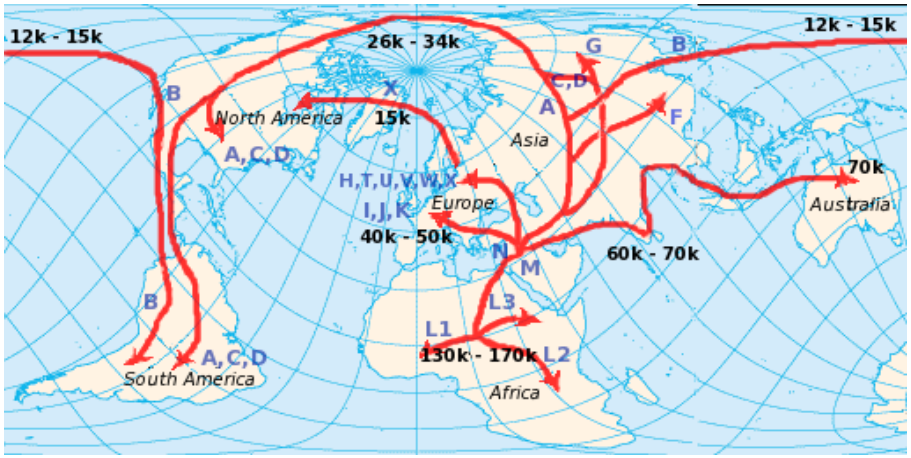
Why do we care about variations?

underlie phenotypic differences



cause inherited diseases

allow tracking human history (ancient and modern)



Biochemical Polymorphism

Protein variation

Molecular Polymorphism

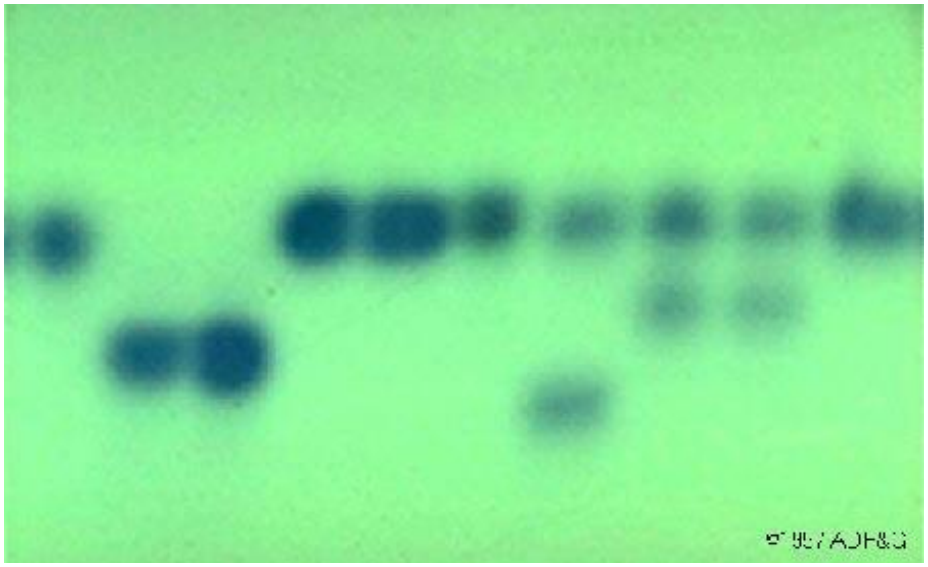
DNA sequence variation

Biochemical Markers

Protein allozymes:

Electrophoretic variants of proteins produced by different alleles at protein-colding genes.

Protein Electrophoresis Gel



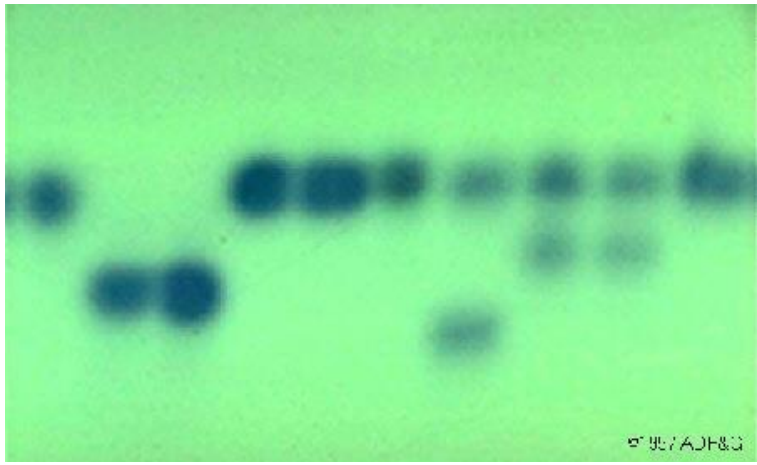
Using protein polymorphism

Advantages:

- inexpensive;
- markers are co-dominant.

Disadvantages:

- Only reveals small proportion of DNA variation.
- Many DNA variants do not result in changes in amino acid sequence (e.g., synonymous substitutions).
- Some changes in amino acid sequence do not result in changes in mobility on the gel.



Molecular Markers

- Molecular markers are based on naturally occurring polymorphisms in DNA sequences (i.e. base pair deletions, substitutions, additions or patterns)
- There are various methods to detect and amplify these polymorphisms
  - RFLP
  - RAPD
  - AFLP etc

There are 5 conditions that characterize a suitable molecular marker:

- Must be polymorphic
- Co-dominant inheritance
- Randomly and frequently distributed throughout the genome
- Easy and cheap to detect
- Reproducible

- Molecular markers can be used for several different applications including:
  - Germplasm characterization,
  - Genetic diagnostics,
  - Characterization of transformants,
  - Study of genome
  - Organization and phylogenic analysis.

TECHNIQUES USED FOR ANALYSIS OF MOLECULAR MARKERS

- Restriction Digestion
- Gel Electrophoresis
- PCR

Molecular Marker Techniques

1) Restriction Fragment Length Polymorphism (RFLP)

- The technique centers around the digestion of genomic DNA digested with restriction enzymes.
- These enzymes are isolated from bacteria and consistently cut DNA at specific base pair sequences which are called recognition sites.
- These recognition sites are not associated with any type of gene and are distributed randomly throughout the genome.
- When genomic DNA is digested with one of these restriction enzymes, (of which there are thousands, each cutting at a specific sequence), a series of fragment are produced of varying length.

- These fragments are separated using agarose or polyacrylamide gel electrophoresis (PAGE) and yield a characteristic pattern.

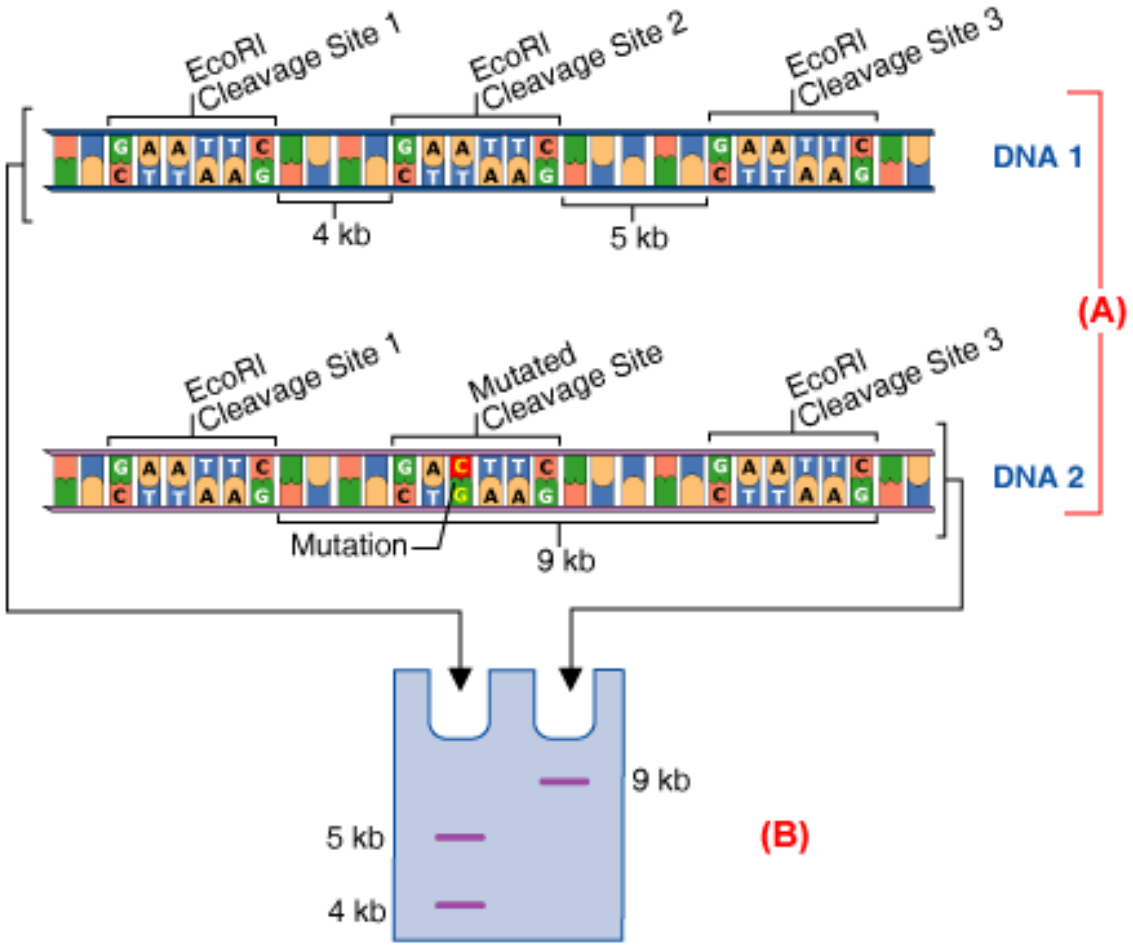
- Variations in the characteristic pattern of a RFLP digest can be caused by
  - base pair deletions,
  - mutations,
  - inversions,
  - translocations and transpositions
- which result in the loss or gain of a recognition site resulting in a fragment of different length and polymorphism.

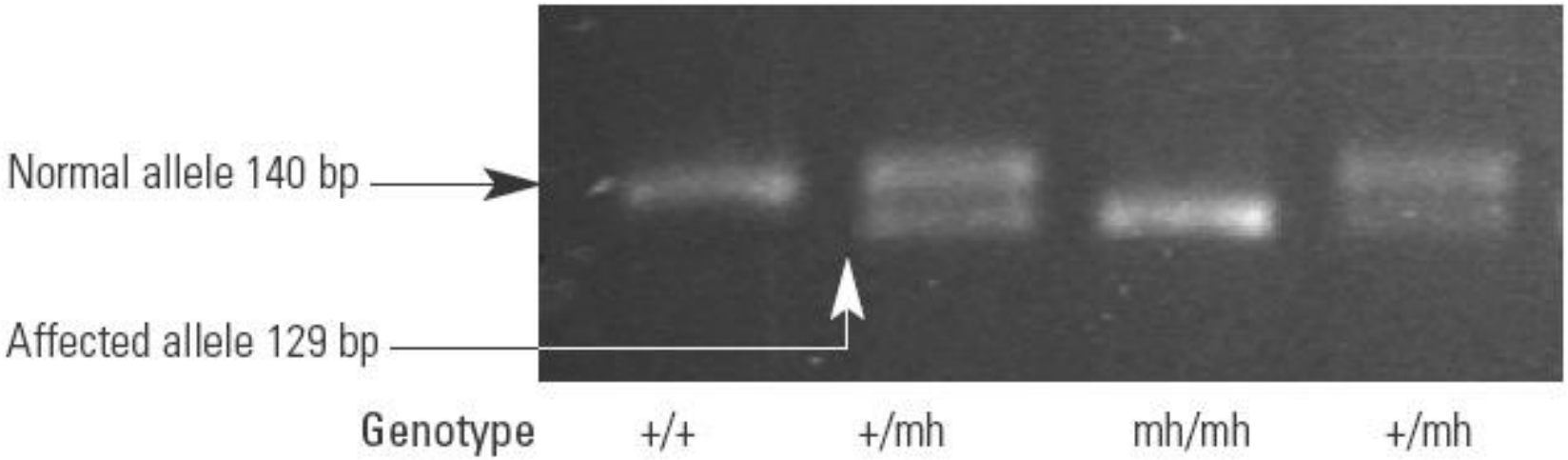
RFLP

Enzymes cut DNA at specific sequences
Restriction sites are often palindromes:

6-cutter GAATTC / CTTAAG

4-cutter TCGA / AGCT


EcoRI Cleavage Site 1
EcoRI Cleavage Site 2
EcoRI Cleavage Site 3
DNA 1
4 kb
5 kb
(A)
EcoRI Cleavage Site 1
Mutated Cleavage Site
EcoRI Cleavage Site 3
DNA 2
Mutation
9 kb
9 kb
5 kb
4 kb
(B)


Normal allele 140 bp
Affected allele 129 bp
Genotype
+/+
+/mh
mh/mh
+/mh

Using RFLP polymorphism

Advantages:

- variants are co-dominant;
- measures variation at the level of DNA sequence, not protein sequence.

Disadvantages:

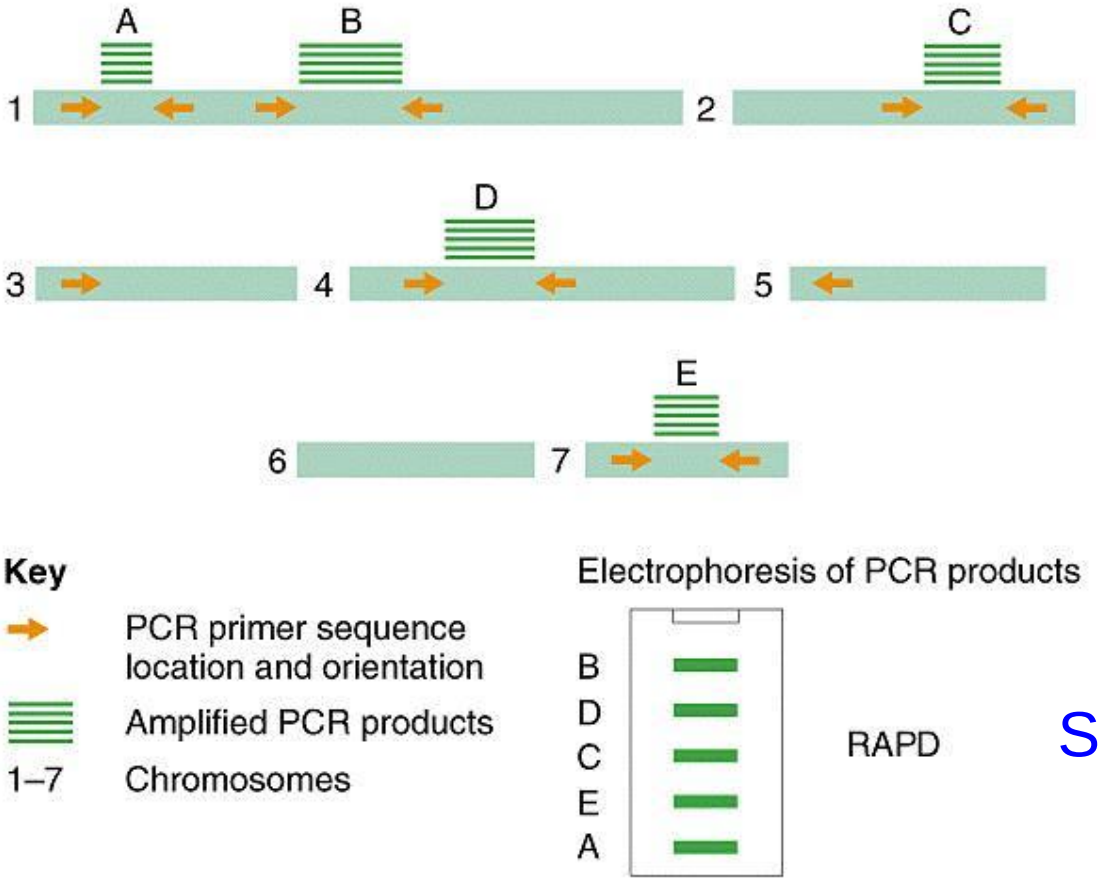
- labor intensive;
- requires relatively large amounts of DNA

PCR Based Molecular Markers

2. Randomly amplified polymorphic DNA Markers (RAPD)

- RAPD was the first PCR based molecular marker technique developed and it is by far the simplest.
- Short PCR primers (approximately 10 bases) are randomly and arbitrarily selected to amplify random DNA segments throughout the genome.
- The resulting amplification product is generated at the region flanking a part of the 10 bp priming sites in the appropriate orientation.
- RAPD often shows a dominant relationship due to primer being unable to bind on recessive alleles.
- RAPD products are usually visualized on agarose gels stained with ethidium bromide.

RAPD: Randomly amplified polymorphic DNA



Size sorted

RAPDs

Advantages:

- fast,
- relatively inexpensive,
- highly variable.

Limitations of RAPD

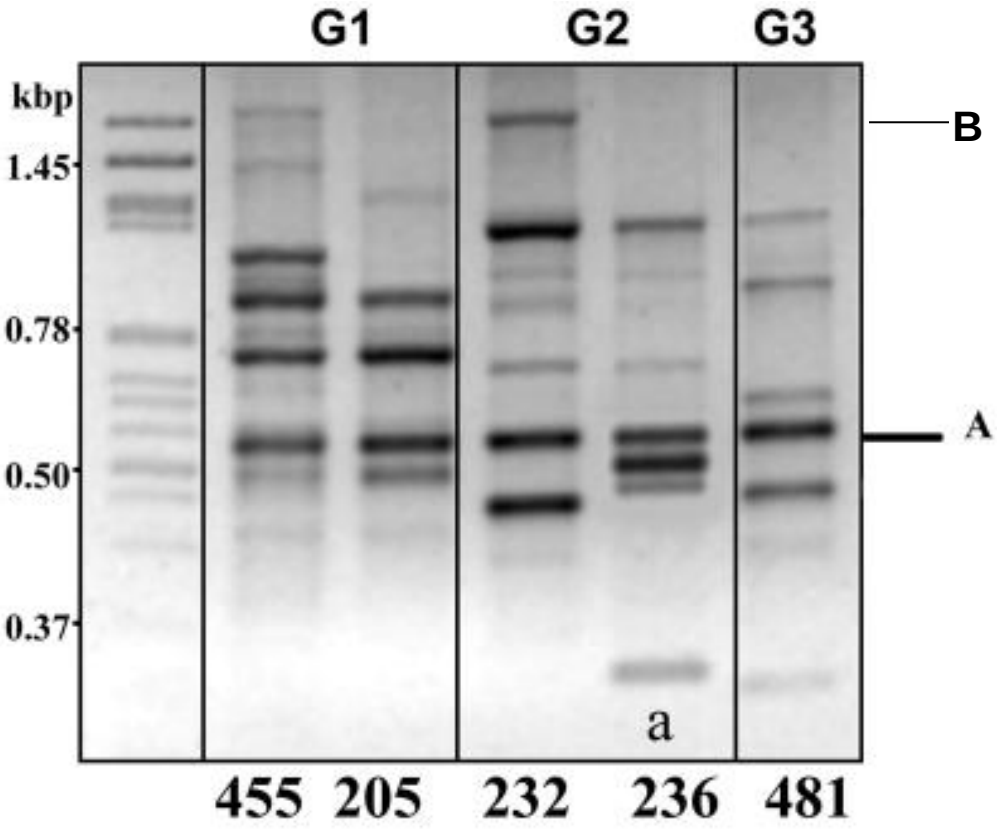
- Nearly all RAPD markers are dominant, i.e. it is not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies). Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are detected only rarely.
- PCR is an enzymatic reaction, therefore the quality and concentration of template DNA, concentrations of PCR components, and the PCR cycling conditions may greatly influence the outcome. Thus, the RAPD technique is notoriously laboratory dependent and needs carefully developed laboratory protocols to be reproducible.
- Mismatches between the primer and the template may result in the total absence of PCR product as well as in a merely decreased amount of the product. Thus, the RAPD results can be difficult to interpret.

.

- Presence of a band could mean the individual is either heterozygous or homozygous for the sequence--can't tell which.

- Data analysis more complicated.

RAPD Analysis



Questions:

1. Is the locus represented by band “B” polymorphic? Band A?
2. Is individual 232 a homozygote or heterozygote for alleles represented by band “B”? What about individual 236?
3. Does band “B” represent a longer or shorter DNA fragment than band “A”.

2. Simple Sequence Repeats (SSR)/Microsatellites

- Simple sequence repeats are present in the genomes of all eukaryotes and consists of several to over a hundred repeats of a 1-4 nucleotide motif.

Some common motifs are:

Mono: A, T

Di: AT, GA

Tri: AGG

Tetra: AAAC

- **Microsatellites** (sometimes referred to as a variable number of tandem repeats or VNTRs) are short segments of DNA that have a repeated sequence such as CACACACA, and they tend to occur in non-coding DNA. In some microsatellites, the repeated unit (e.g. CA) may occur four times, in others it may be seven, or two, or thirty. In diploid organisms such as elephants, each individual animal will have two copies of any particular microsatellite segment. For example, a father might have a genotype of 12 repeats and 19 repeats, a mother might have 18 repeats and 15 repeats while their first born might have repeats of 12 and 15.
- On rare occasions, microsatellites can cause the DNA polymerase to make an extra copy of CA similar to the way we find it difficult to say ³toy boat² several times in a row with consistent accuracy. If an individual¹s DNA polymerase adds to the repeated sequence, then this slightly larger version can be passed on to offspring who will usually replicate it accurately. Over time, as animals in a population breed, they will recombine their microsatellites during sexual reproduction and the population will maintain a variety of microsatellites that is characteristic for that population and distinct from other populations which do not interbreed.

- The most common way to detect microsatellites is to design PCR primers that are unique to one locus in the genome and that base pair on either side of the repeated portion (figure 1). Therefore, a single pair of PCR primers will work for every individual in the species and produce different sized products for each of the different length microsatellites.

- **Figure 1.** Detecting microsatellites from genomic DNA. Two PCR primers (forward and reverse gray arrows) are designed to flank the microsatellite region. If there were zero repeats, the PCR product would be 100 bp in length. Therefore, by determining the size of each PCR product (in this case 116 bp), you can calculate how many CA repeats are present in each microsatellite (8 CA repeats in this example).

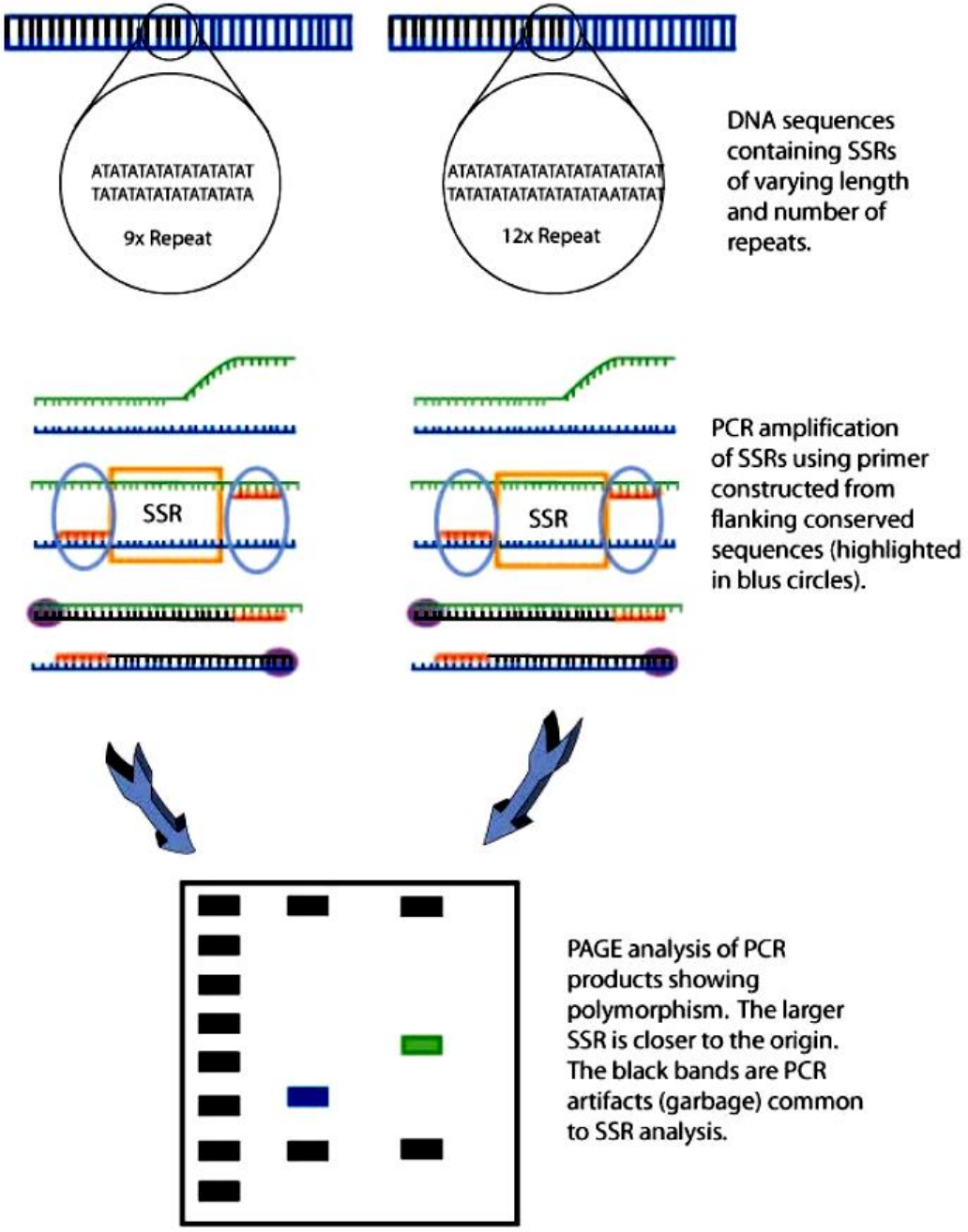
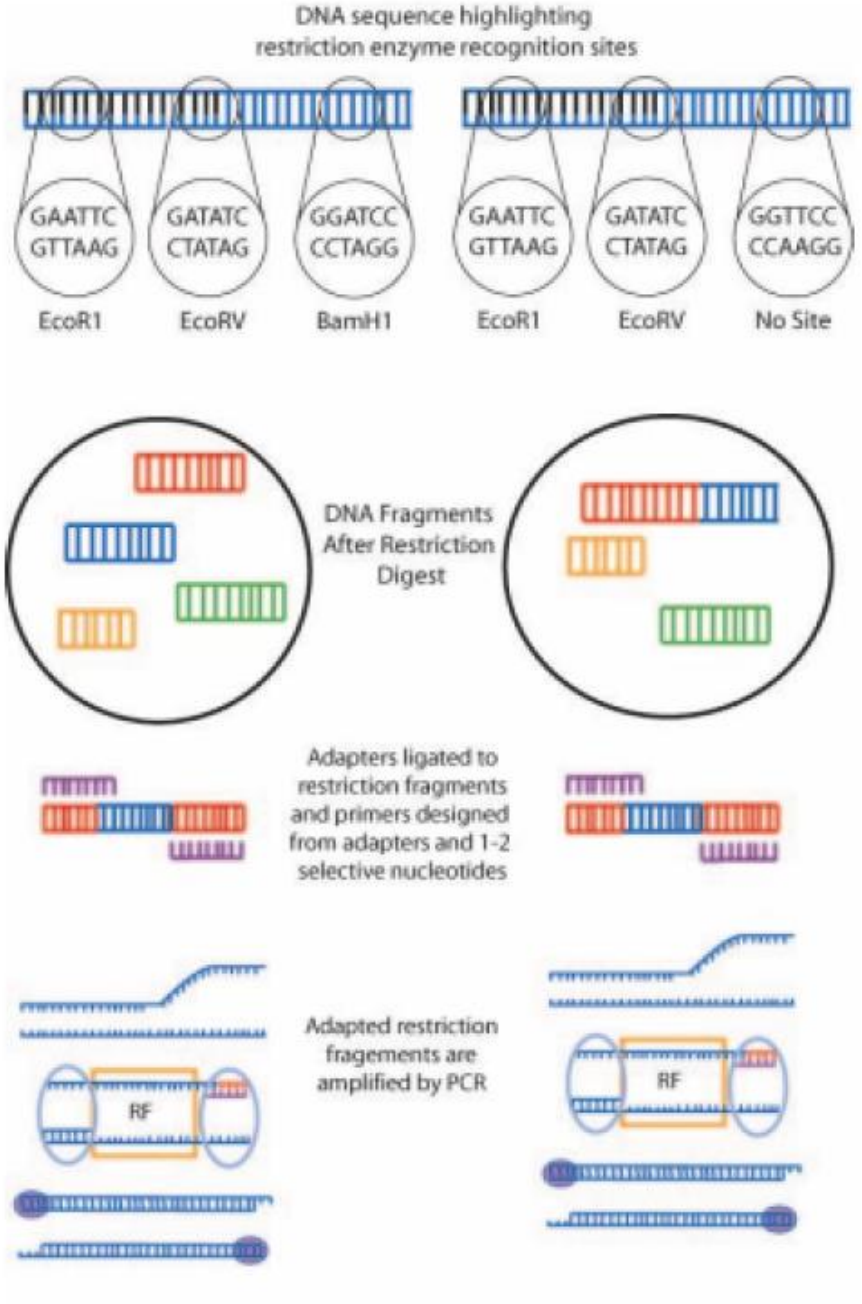


Fig 7: Diagram of major events in SSR analysis

4. Amplified Fragment Length Polymorphism (AFLP)

- AFLP is the latest form of marker assisted selection and is a highly sensitive method based on the combined concepts of RFLP and RAPD.
- This technique is applicable to all species giving very reproducible results.
- The basis of AFLP is the PCR amplification of restriction enzyme fragments of genomic DNA.



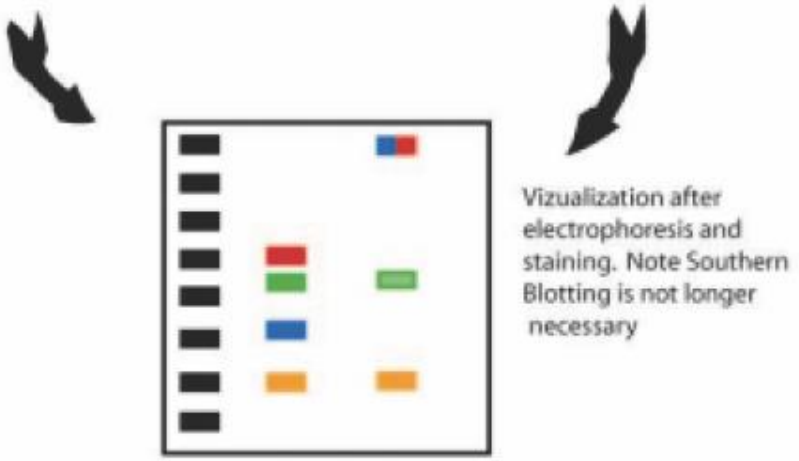
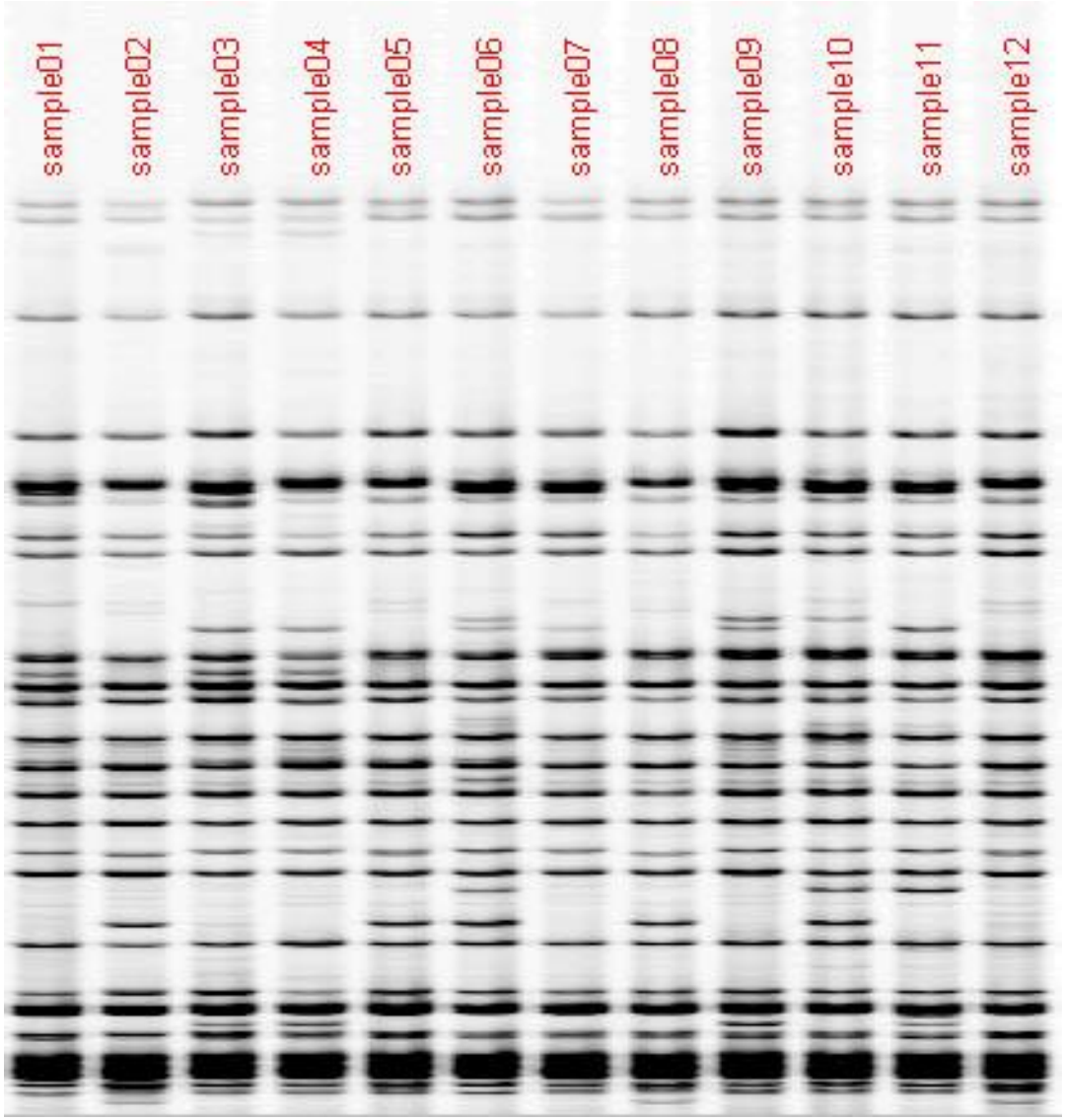


Fig 8: Diagram of major events in AFLP analysis

AFLPs


sample01
sample02
sample03
sample04
sample05
sample06
sample07
sample08
sample09
sample10
sample11
sample12

AFLPs

Advantages:

- fast,
- relatively inexpensive,
- highly variable.

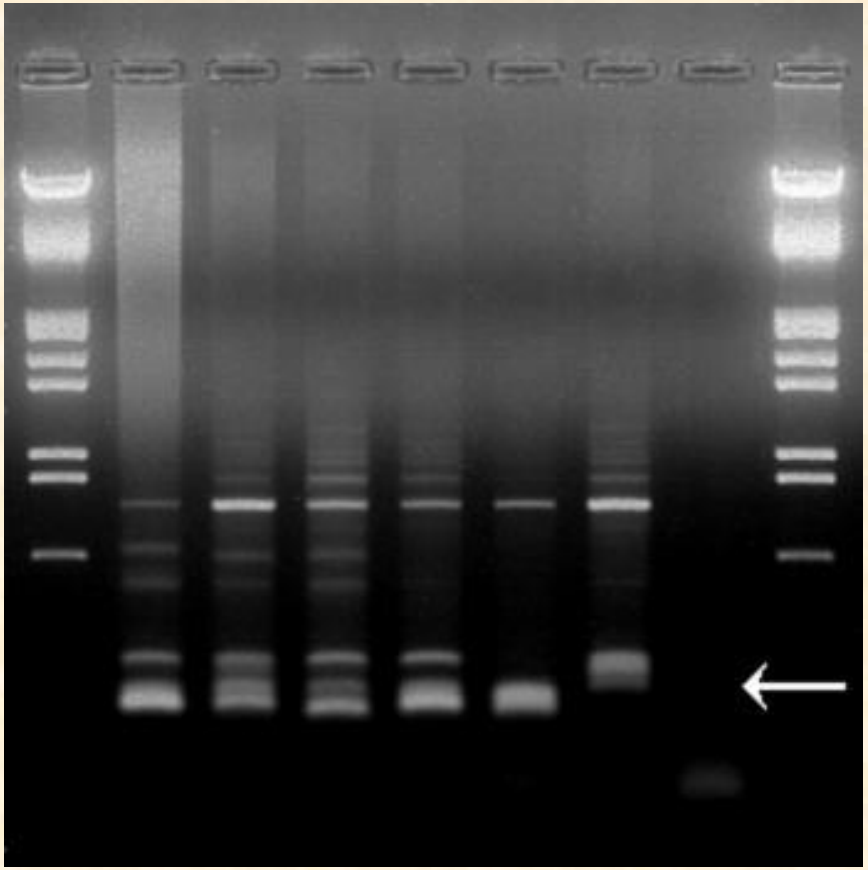
Disadvantages:

- markers are dominant.
- Presence of a band could mean the individual is either heterozygous or homozygous for the sequence--can't tell which.

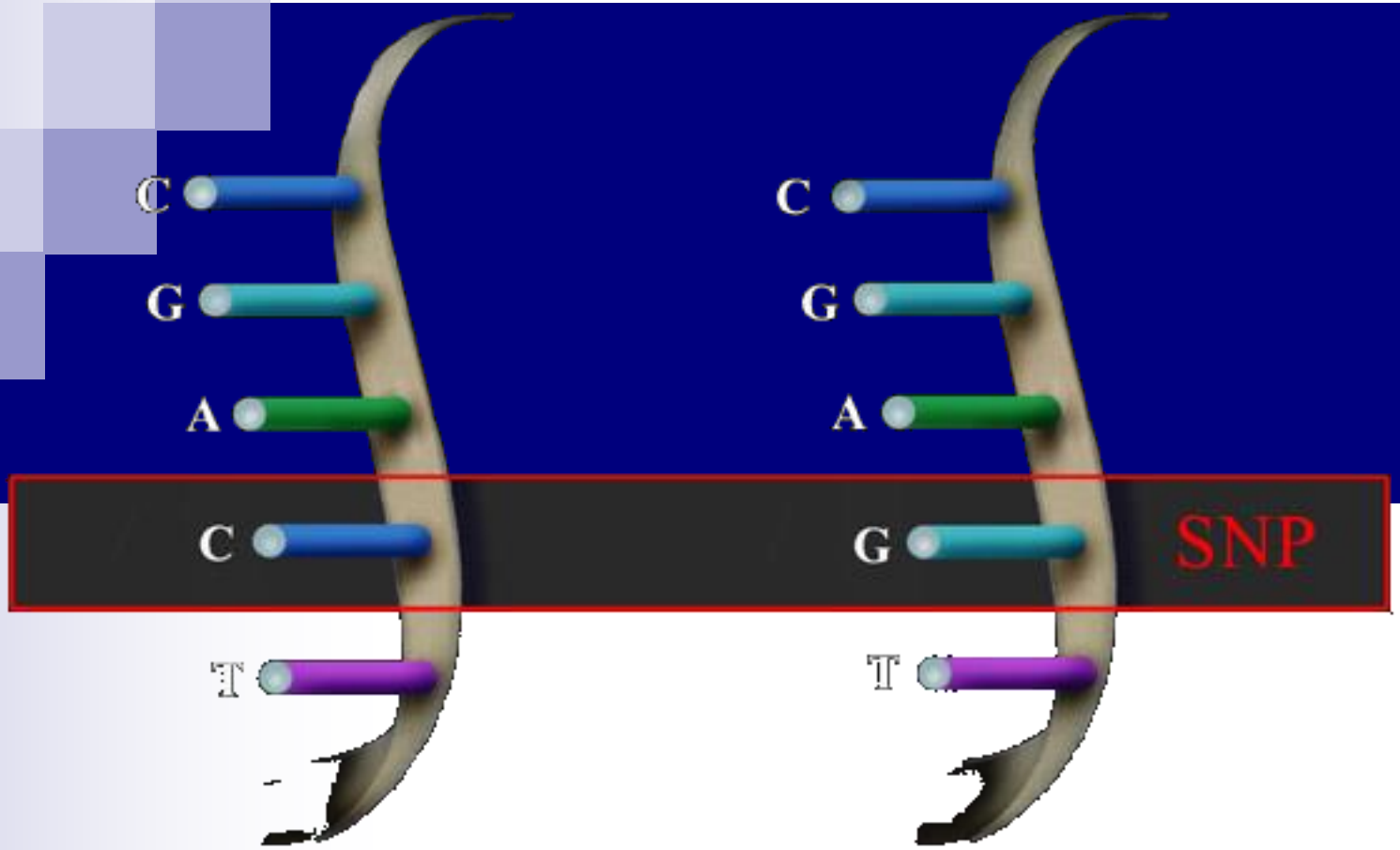
Microsatellites

Questions:

1. Is the locus represented by the bands at the arrow polymorphic?
2. If it is polymorphic, how many individuals are heterozygous?
3. How many individuals are homozygous for the “short” allele?



What is a SNP?



SNP

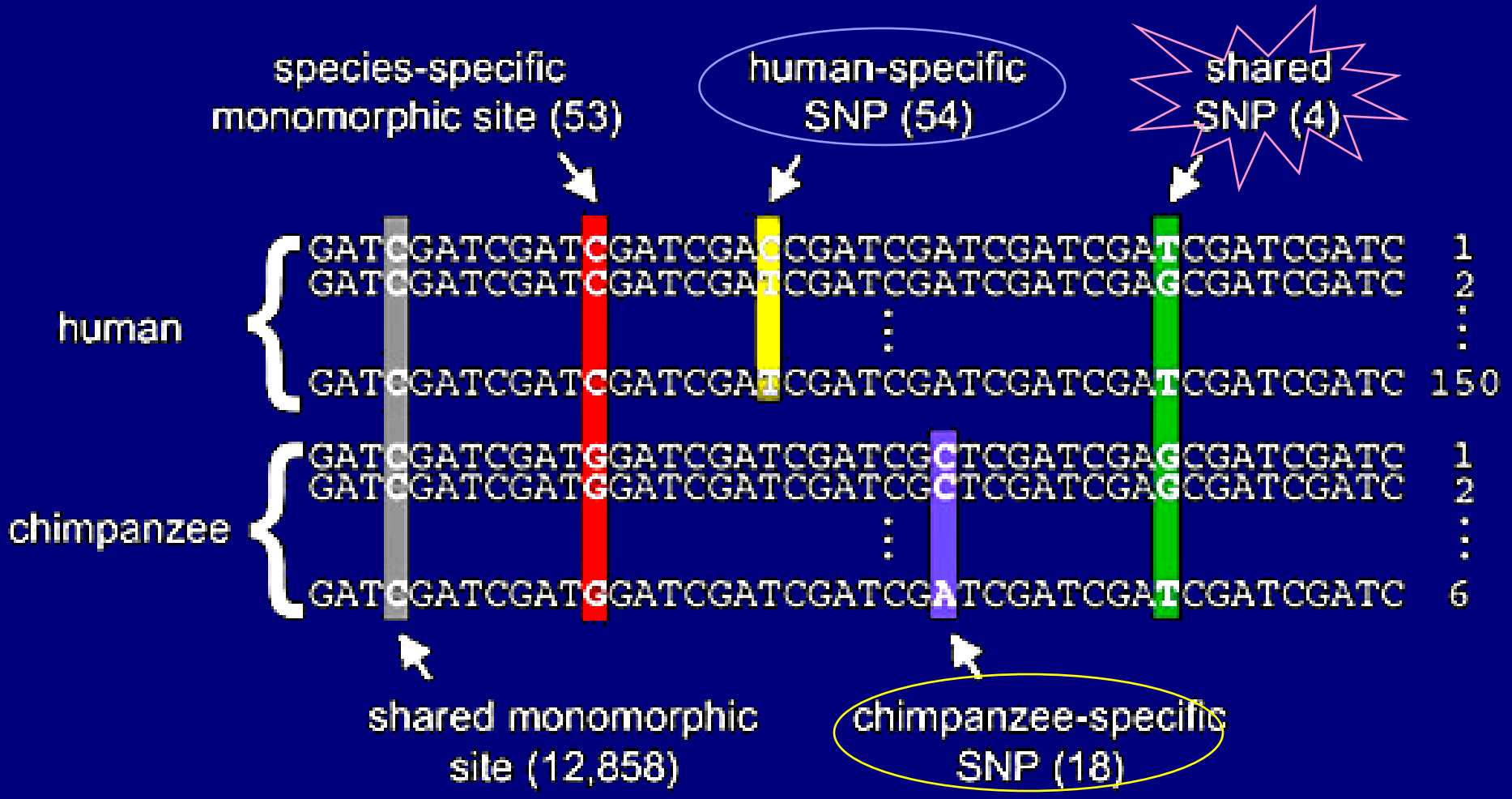
- The most common genetic polymorphism
- Distribute throughout genome with high density
- More stable and easy to assay
- Major cause of genetic diversity among different (normal) individuals, e.g. drug response, disease susceptibility.
- Facilitates large scale genetic association studies as genetic markers.

SNP

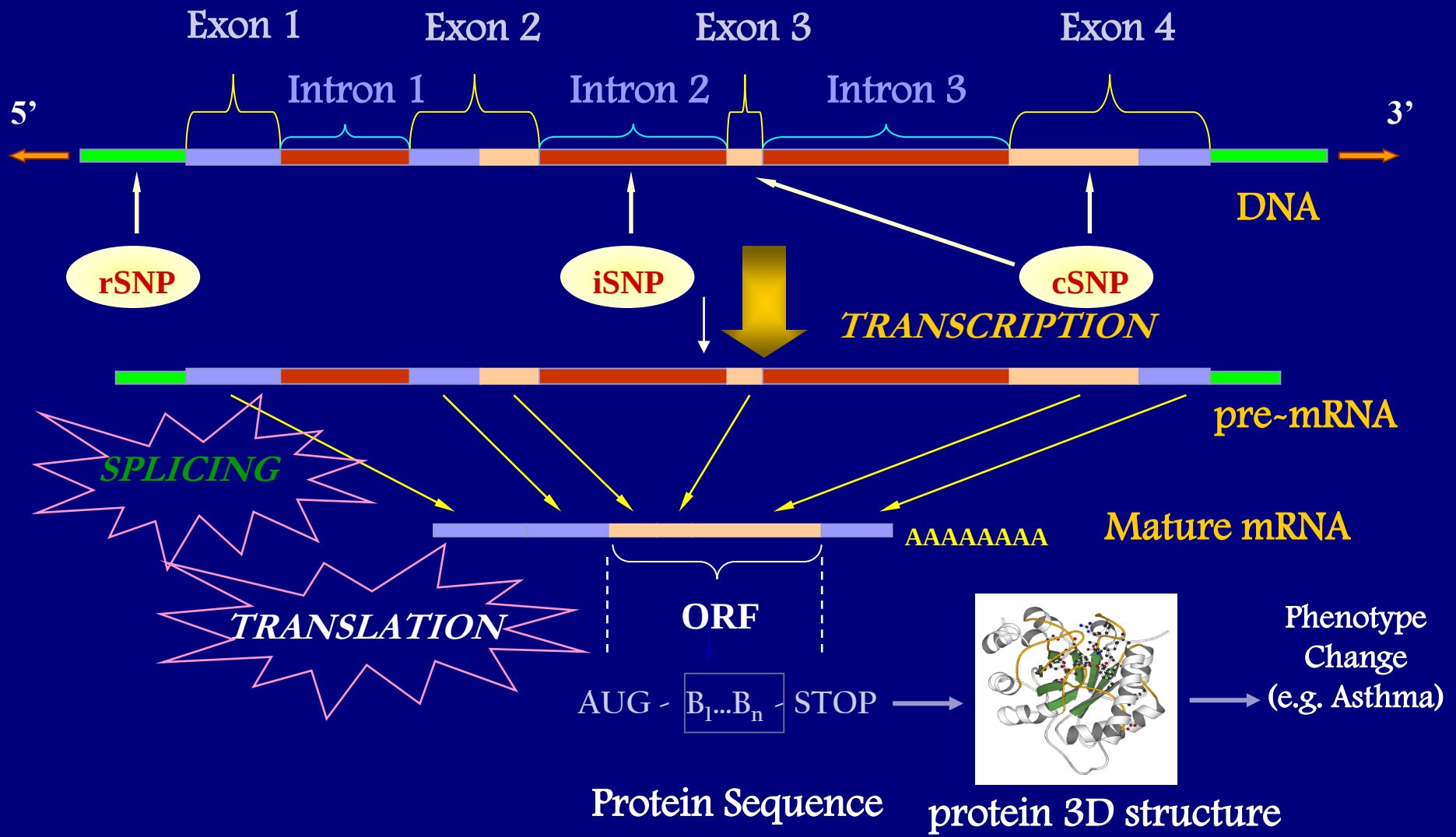
- Most of SNPs neither change protein synthesis nor cause disease directly. Rather, they serve as landmarks, since they may be physically close to the mutation site on the chromosome. Because of this proximity, SNPs may be shared among groups of people with common characteristics.

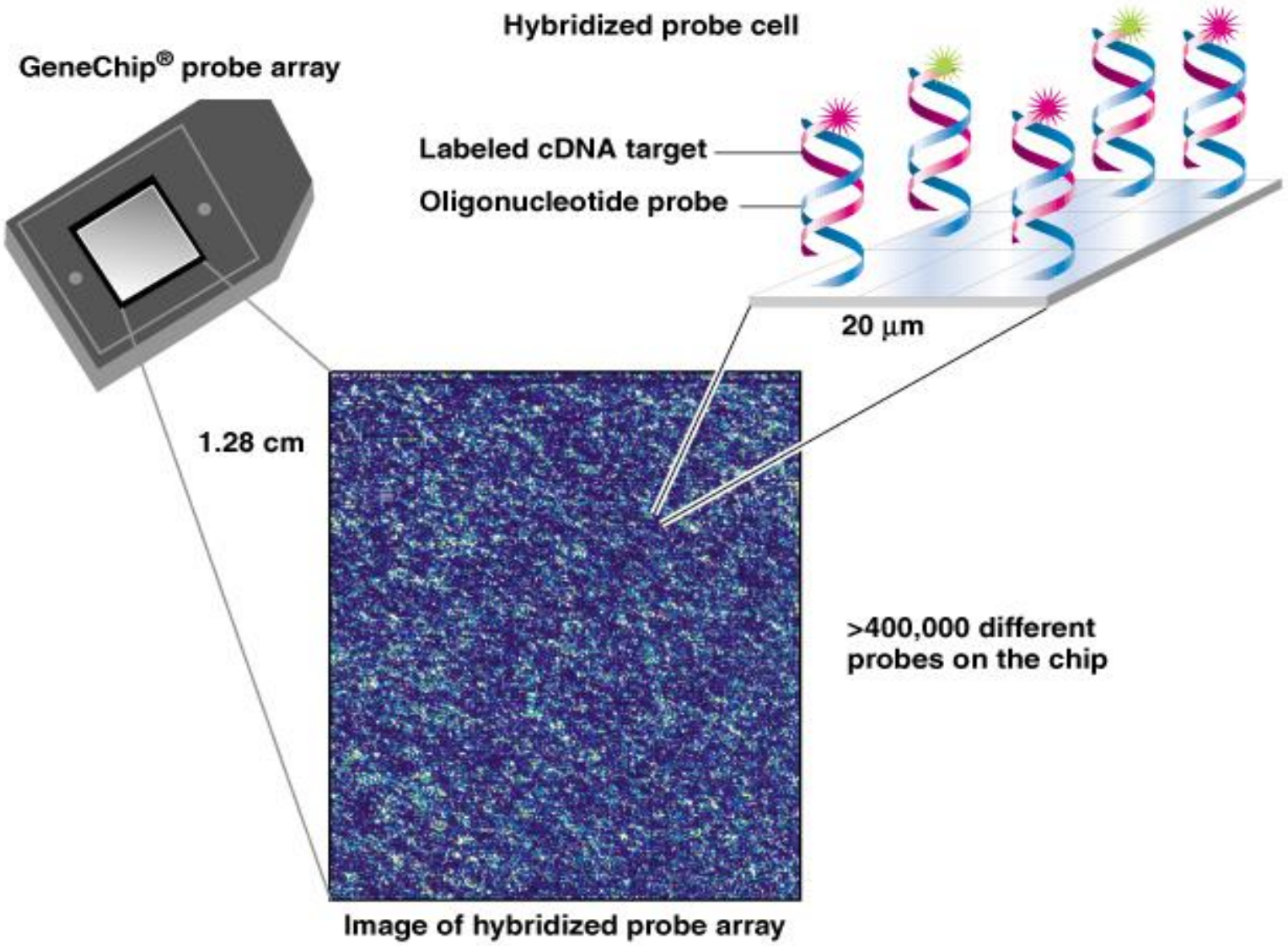
- Analyze SNP patterns among different groups of people may shed light on evolution of human race, understand ethnic groups and races.

SNP Types


species-specific
monomorphic site (53)
human-specific
SNP (54)
shared
SNP (4)
human
GATCGATCGATCGATCGACCGATCGATCGATCGATCGATCGATC 1
GATCGATCGATCGATCGATCGATCGATCGATCGAGCGATCGATC 2
GATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATC 150
chimpanzee
GATCGATCGATGGATCGATCGATCGCTCGATCGAGCGATCGATC 1
GATCGATCGATGGATCGATCGATCGCTCGATCGAGCGATCGATC 2
GATCGATCGATGGATCGATCGATCGATCGATCGATCGATCGATC 6
shared monomorphic
site (12,858)
chimpanzee-specific
SNP (18)

SNP Locations




Hybridized probe cell
GeneChip® probe array
Labeled cDNA target
Oligonucleotide probe
20 μm
1.28 cm
>400,000 different probes on the chip
Image of hybridized probe array

Thanks…