* MOLECULAR TAXONOMY

Dr . G. B. CHAND

Associate Professor

Aquatic Toxicology Laboratory

Department of Zoology

Patna University, Patna

Email : gbchand @rediffmail.com

INTRODUCTION

- The classification of organism on the basis of distribution and composition of molecular substances in them is called as Molecular taxonomy.
- Molecular systematics is the use of molecular genetics to study the evolution of relationship among individuals and species.
- Molecular phylo-genetics is the study of evolutionary relationship among biological entities(individual, population, species or higher taxa)by a combination of molecular data like DNA, protein sequence , presence or absence of transposable elements, gene order data and statistical techniques..
- Molecular phylo-genetics is one aspect of the molecular systematics a broad term that includes the use of molecular data in taxonomy and biogeography.
- It is the branch of phylogeny that analyses hereditary molecular differences mainly in DNA sequence to gain information of an organism evolutionary relationship. The result of molecular phylo-genetics is expressed as Molecular tree.
- In molecular taxonomy proteins and genes are used to determine evolutionary relationship. Among gene pool of a species the percentage of DNA , which is similar in species are taken into consideration.
- the rRNA molecule of an organism is known as ‘Signature sequence’. It is used to determine evolutionary relationship in prokaryotes.

INTRODUCTION

Historical background

- Application of Molecular systematics were pioneered by Charls G. Sibley(birds), Herbest C. Dessauer (reptile), Morris Goodman (primates) followed by Allan C. Wilson, Robert K. Sulender and John C. Ause work with protein electrophoresis, which began around 1956.
- Walter M. Fitch and Emanual Margoliash made first phylogenetic tree based on molecular data.

Aims Of Molecular taxonomy

- The aim of molecular systematics is to provide insight into the history of group of organism and evolutionary processes that create diversity among species.

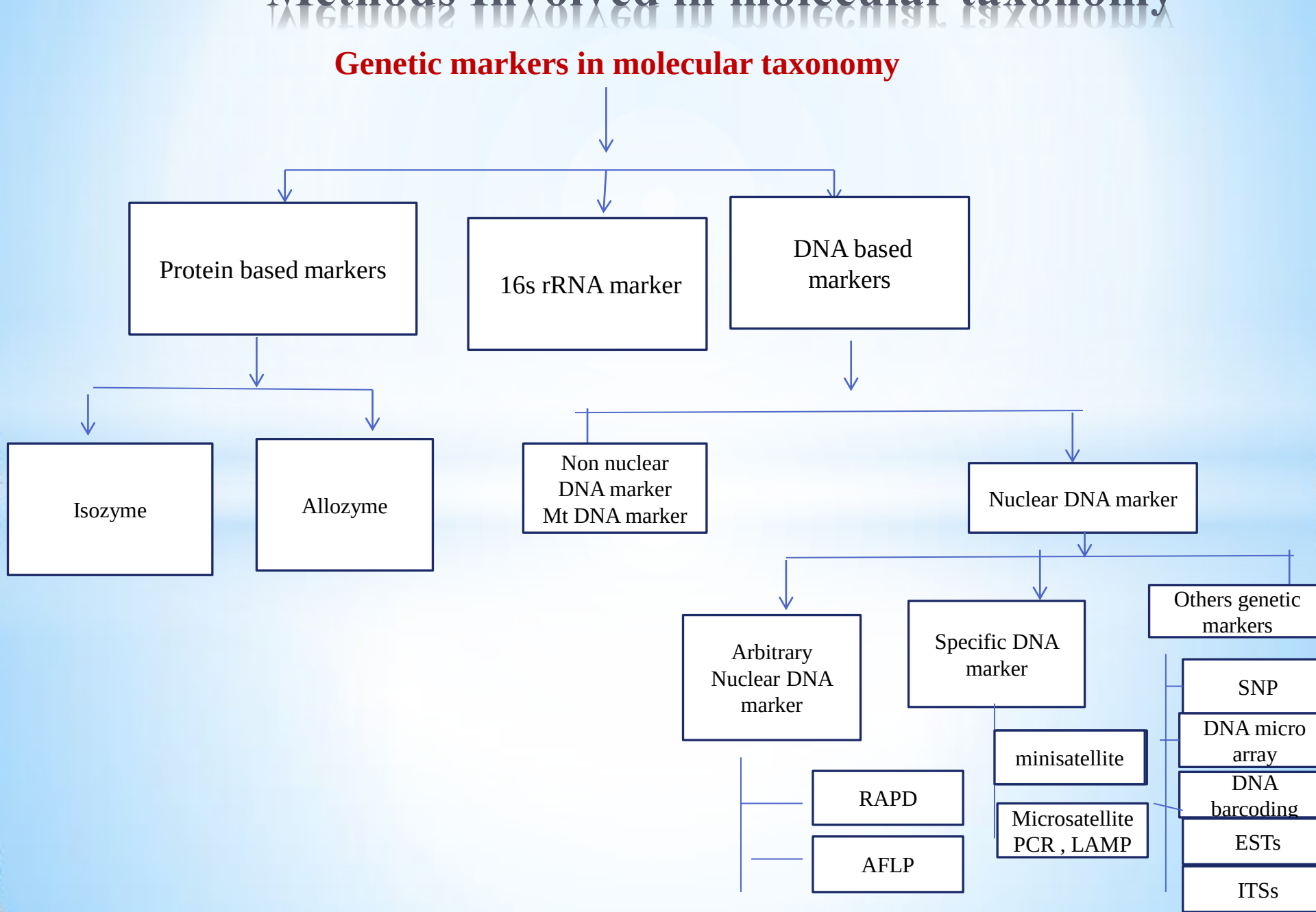
Objectives of Molecular taxonomy

- 1. To reconstruct the correct geneological tree among biological entities.
- 2. To establish the time of divergence between biological entities.
- 3. To chronicle the sequence of events along evolutionary lineage.

* Molecular markers used in Molecular taxonomy

- The molecules based on which molecular taxonomy is done are called **Molecular Markers**.
- Molecular markers can also be defined as signs especially along the DNA that pin-point the location of desirable genetic traits or specific genetic differences.
- A particular fragment of DNA can be used as a marker when differences can be detected in that fragment's DNA sequence among multiple plants or plant lines. These sequence variations, called polymorphisms, can be associated with different forms (alleles) of nearby genes involved with particular traits. The polymorphism, or difference, is the clue to find the gene of interest.
- Molecular markers are versatile tools in various fields other than taxonomy like physiology, embryology, genetic fingerprinting etc.
- Molecular phylogenetics and systematics have been found to be greatly promising in recent years, due to the development of new and diverse method.
- Molecular taxonomic approaches permit an exact and rapid method of distinguishing specimens based on their interspecific variations. These methods allow estimation of the genetic variability of the biota carrying to a super-estimation on the global biodiversity besides the relationships among taxa.

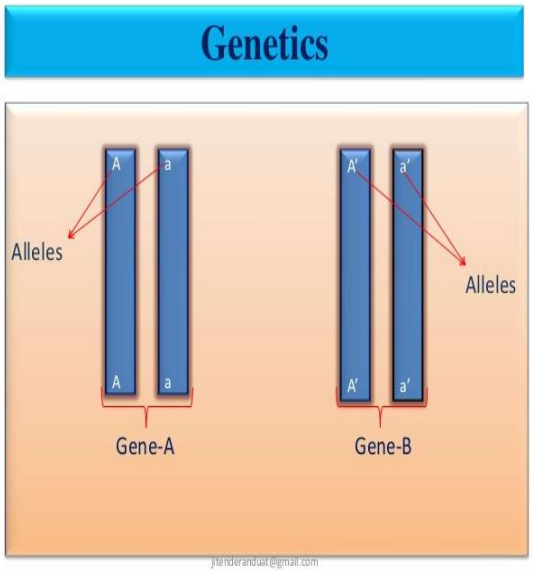
Methods Involved in molecular taxonomy

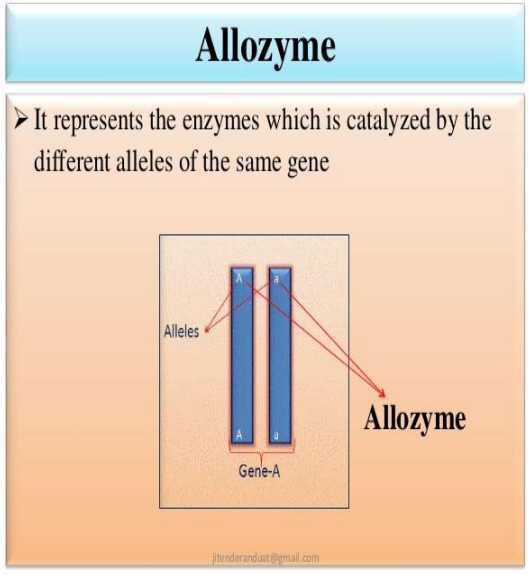


* Molecular markers used in Molecular taxonomy

Protein based markers

- Iso -enzyme s and allozymes were first discovered by R L Hunter and Clement Markert in 1957.
- Both these two variables are now used as interchangeable eg. Lactate dehydrogenase(LDH) chr.-12 and chr-15, malate dehydrogenase, glucose phosphate and glucokinase etc.
- Alloenzymes are common biological enzymes that exhibit high levels of functional evolutionary conservation throughout specific phyla and kingdoms. They are used by phylo-geneticists as molecular markers to gauge evolutionary histories and relationships between different species.
- A Allozyme electrophoresis is a method which can identify genetic variation at the level of enzymes that are directly encoded by DNA protein variants and they will differ slightly in electric charge.
- Allozyme provides us a data of single locus genetic variations which can answer many questions.





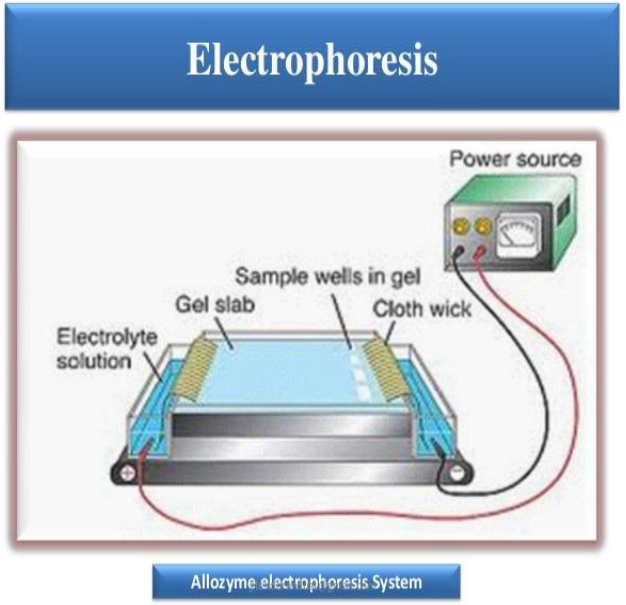
Steps in Allozyme analysis

1. Extract allozyme from tissues using standard specific protocol.
2. Then the variation is detected through electrophoresis in an acrylamide or in cellulose acetate gel.
3. Individuals that are homozygous show a single band where as heterozygous individuals show two bands.
4. it is a codominant Mendelian character

* Molecular markers used in Molecular taxonomy

Steps in Allozyme analysis

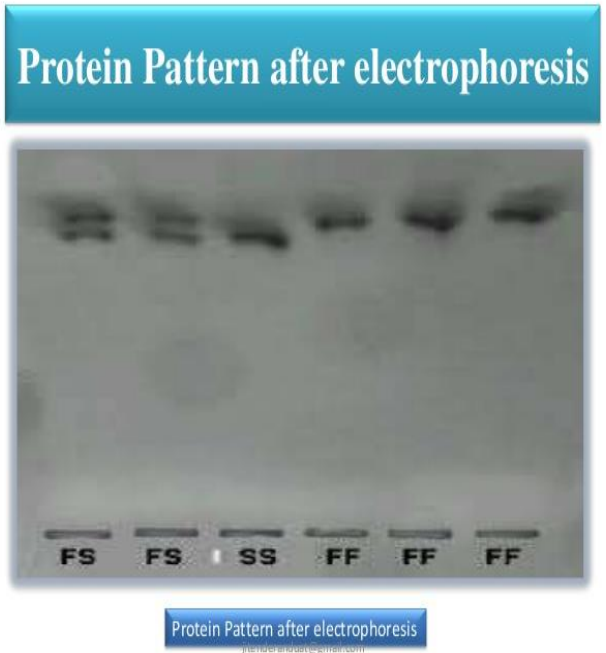
- Extract allozyme from tissues using standard specific protocol.
- Then the variation is detected through electrophoresis in an acrylamide, starch gel or in cellulose acetate gel.
- The protein bands obtained are observed carefully. Individuals that are homozygous show a single band where as heterozygous individuals show two bands.40-50 individuals can be analyzed per gel.
- It is a co-dominant Mendelian character.



- Molecular markers are versatile tools in various fields other than taxonomy like physiology, embryology, genetic fingerprinting etc.

- Molecular phylo-genetics and systematics have been found to be greatly promising in recent years, due to the development of new and diverse method.

- Molecular taxonomic approaches permit an exact and rapid method of distinguishing specimens based on their interspecific variations. These methods allow estimation of the genetic variability of the biota carrying to a super-estimation on the global biodiversity besides the relationships among



* Molecular markers used in Molecular taxonomy

Advantages of allozyme

- Due to their simplicity, low cost and requirement of little specialized equipment, it is most widely studied form of molecular evolution.
- It is co dominant type of marker and easily adjustable from species to species.
- The allozyme analysis has general applicability. It has wide spread use in fisheries mainly in Ichthyo- taxonomy, individual identification, establishing the phylogenetic relationship of the different species of fishes, identification and arrangement of the species and stock on the basis of their genetic affinity, identification of genetic diversity within stock and cultured species, study of population structure, genome mapping, study of the genetic behaviour of smolt stage, genetic characterization of the species and stock, to study the genetic drift of the species and stock ,Mixed Stock Fishery Analysis(MSFA), forensic analysis and study of conservation genetics.

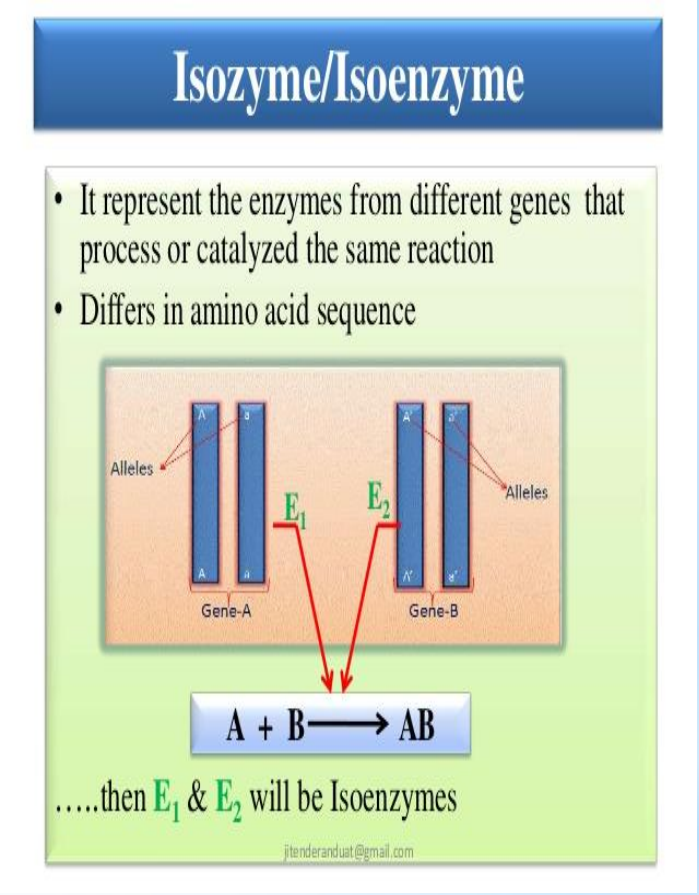
Limitations of Allozyme

- The technique requires a large amount of tissues and cannot be applied when the organism are very small i.e. larvae form.
- A tissue sampling method is invasive and the fish needs to be sacrificed and the tissues needs to be stored cyto-genetically. A point mutation in nucleotide sequence may not necessarily results in change in amino acid and hence could not be detected by protein electrophoresis.
- presence of limited number of allozyme loci precludes their use in large scale genome mapping.
- Heterozygote deficiencies due to null alleles.

* Molecular markers used in Molecular taxonomy

Isozyme

- Isozymes are protein markers.
- Isozyme staining is a kind of activity staining – a molecular marker technique based on the principle that allelic variation exists among many proteins which perform same enzymatic function but the electrophoretic mobility of the proteins may differ (depending on their respective molecular weights); therefore they migrate to different extents in a starch or polyacrylamide gel.
- In this technique, a crude protein extract is made from some tissue sources and separated by electrophoresis in a solution containing reagents required for the activity of the enzyme being monitored. The solution contains a dye that the enzyme can catalyse into a colour reagent that stains the protein. The allelic variants of the protein can be visualized in the gel.
- Isozyme markers are much used in molecular taxonomy as the technique is easy to perform, is cost-effective, less time-consuming, and gives a vivid representation of gel electrophoresis band patterns, as far as studies on taxonomic diversity.



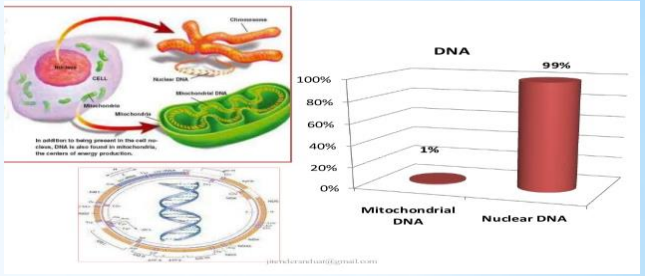
Limitations of Allozyme

- The number of isozyme loci that can be scored is limited. To date, only 40 – 50 reagent systems have been developed that permit staining of a particular protein in starch. Moreover, not all of these reagents work efficiently with all plant species. Hence, for many species only 15 – 20 loci can be mapped.
- Tissue variability causes some isozymes to be better expressed in certain tissues. Therefore, several samplings of the segregating population are necessary to score all the available isozymes.

* Molecular markers used in Molecular taxonomy

Mitochondrial DNA marker

- Mitochondrial DNA is non nuclear, remain present within mitochondria.
- Mt DNA is maternally inherited with haploid genome
- The entire genome undergoes transcription as one single unit. They are not subjected to any recombination and hence they are homologous marker .
- They are selectively neutral, occurring in multiple copies in each cell.
- Mt DNA is physically separated from the rest of the cells DNA and so it is relatively easier to isolate from any tissues or blood samples.
- Due to maternal inheritance of Mt. DNA, the effective population size is smaller than nuclear DNA and so Mt DNA variation is more sensitive to population bottle neck and hybridization.
- The difference in the nucleotide sequence of the DNA molecule in mitochondria can be determined directly or indirectly by several methods like-RFLP.
- The newly emerged sequencing technologies have enabled direct sequencing of Mt DNA and several sets of universal primer have been developed from conserved sequence region. Slow evolved gene regions are constantly being used for interspecies comparison while fast evolving gene region are used for population comparison eg. D-loops
- The only non coding region of Mt DNA is D-loop region which is fast evolving gene region and hence mostly used for population comparison. Besides the Cyt.b and ND-1and ND-5/6gene regions are also being used.
- Mit cytochrome c oxidase I gene(CO I gene) has been identified as universal barcode for species level identification due to its conserved nature across a wide range of taxa.



Mitochondrial DNA

mtDNA is small genome and **double stranded** circular DNA molecule

Haploid in nature

mt DNA contain **37 genes** all of essential for normal mitochondrial function

13 genes making enzymes involved in oxidative phosphorylation (**ATP production**)

process involved use of **oxygen** and simple **sugars** to form **ATP**

The Mitochondrial Genome

- **16,569** base pairs (**bp**) in length (16-18 kbp)
- encodes **37 genes**, **13 proteins**, **22 tRNAs**, and **2 rRNAs**

two general regions:

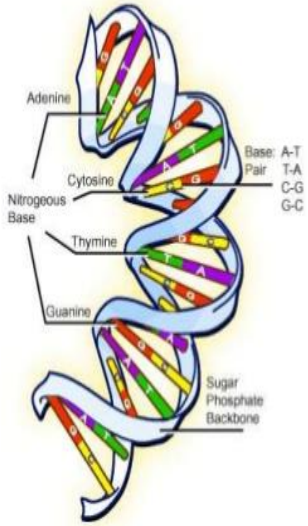
- **coding region**: responsible for the production of various biological molecules involved in **"cellular respiration"**
- **control region**: responsible for the **regulation of the mtDNA molecule**

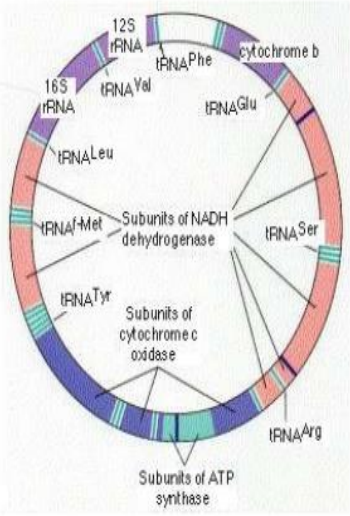
Mitochondrial DNA : At A Glance

Nuclear DNA vs. Mitochondrial DNA

• Nuclear DNA	• Mitochondrial DNA
– found in **nucleus** of the cell	– found in **mitochondria** of the cell
– 2 sets of **23 chromosomes**	– each mitochondria may have **several copies** of the single mtDNA molecule
– **maternal and paternal**	– **maternal only**
– can "**discriminate** between individuals of the same maternal lineage"	– **cannot "discriminate** between individuals of the same maternal lineage"
– **double helix**	– **Circular**
– **bounded** by a nuclear envelope	– **free** of a nuclear envelope
– DNA **packed** into chromatin	– DNA is **not** packed into chromatin

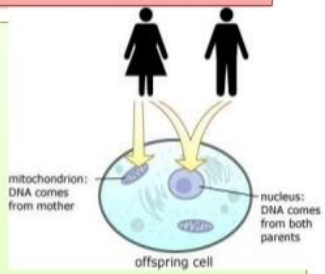
Nuclear DNA vs. Mitochondrial DNA





Maternal Inheritance of mtDNA

During **fertilization**, the sperm only contributes its nucleus (**23 chromosomes**)

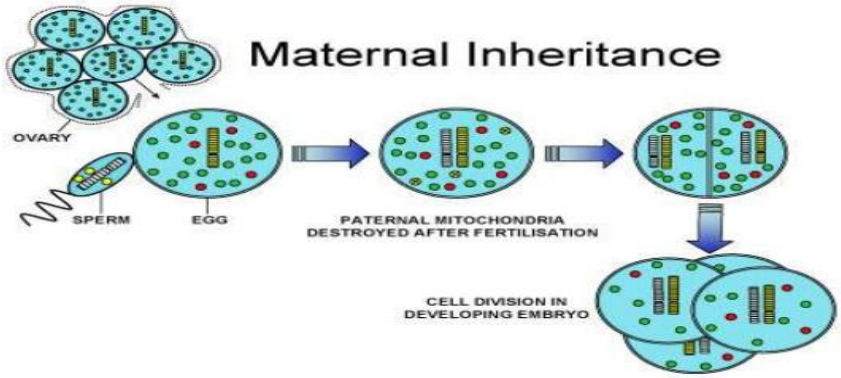


Mitochondria of the **sperm cell** are located at the mitochondrial sheath which is **destroyed** upon fertilization

Only available mitochondria (**mtDNA**) is that of the mother's; this is why mtDNA is of maternal origin

jitenderanduat@gmail.com

Maternal Inheritance of mtDNA



jitenderanduat@gmail.com

Methods for mtDNA analysis

DNA Extraction

- cellular homogenate is "**exposed** to a mixture of organic chemicals that separate the DNA from other biological molecules, such as proteins"
- mixture is **spun in a centrifuge**

- **DNA settles**
- top layer is **filtered** and concentrated
- DNA sample is now **purified**

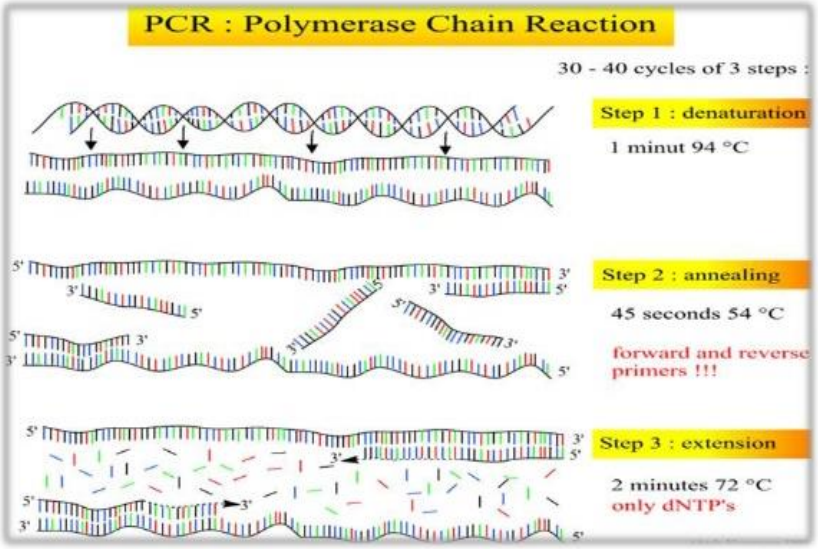
1

Polymerase Chain Reaction (PCR) Amplification

- PCR is a "**procedure that makes many copies of a small amount of DNA**."
- DNA is heated at **94 C** to separate the two strands of the DNA double helix in the sample
- **new DNA strands** are then made from the template (initially separated strands) of DNA by using DNA polymerase, **primers**, and free nucleotides
- the process is **repeated multiple times**, doubling the amount of DNA after each cycle

2

(PCR) Amplification



3

Post amplification Purification and Quantification

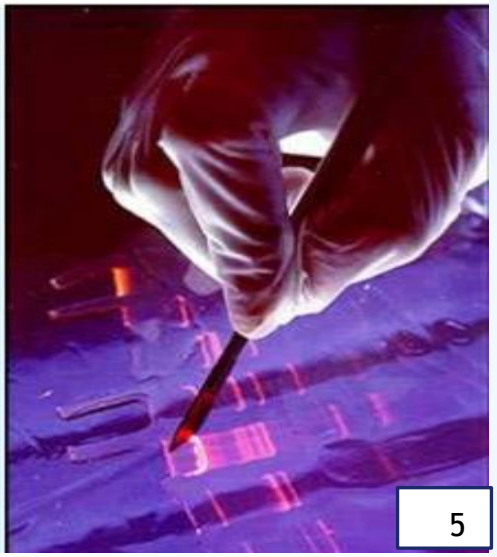
- Purification is performed "**using filtration** devices that remove the excess reagents used in the **PCR from the sample**."
- Quantification is performed "**using capillary electrophoresis** (CE)," which compares the amount of DNA in the PCR product to "a **known DNA standard to determine the concentration** of the DNA in the PCR-amplified sample."

4

Methods of Mitochondrial DNA marker contd.

DNA Sequencing

- **Gel Electrophoresis:**
  - DNA products are separated by **length of bp**
  - pore size of the gel influence how far the DNA fragments will travel when placed in an **electric field**
  - **smaller fragments** will **travel faster** and appear further from the wells in the gel
  - **larger fragments** will **travel slower** and appear closer to the wells in the gel
  - fluorescent detector "**records the emitted wavelength** of the fluorescent **dyes** on each base as the fragments travel past the detection area of the instrument"
  - a **chromatogram is generated, showing the colors** of the labeled fragments
  - "the **sequence** of the **mtDNA is determined** from a **series** of cycle **sequencing reactions**"



5

Data Analysis

mtDNA sequences are generated by a **computer** and edited by a **DNA examiner** to obtain the **final sequence**

Difference(s) is/are recorded by **comparing** the finalized sequence to the Anderson reference sequence

If sequence **concordance** ("the presence of the **same base** or a common base at every position analyzed") is observed,

then both mtDNA samples could be considered as originating from the **same source**

6

jitenderanduat@gmail.com

Application of Mitochondrial DNA Marker in molecular taxonomy

Molecular markers also show significant promise for aquaculture applications

(i) In comparison of hatchery and wild stocks;

(ii) Genetic identification and discrimination of hatchery stocks;

(iii)Monitoring inbreeding or other changes in the genetic variation;

(iv) Assignment of progeny to parents through genetic tags;

(v) Identification of quantitative trait loci (QTL) and use of these markers in selection programmes;

(vi)Assessment of successful implementation of genetic manipulations such as polyploidy and gynogenesis.

Source: (Magoulas, 1998 ; Davis and Hetzel, 2000; Fjalestad *et al., 2003; Subasinghe et al., 2003)*

DNA Barcodes

- DNA Barcodes are segments of approximately 600 b p of Mit. CO I gene, which is a fast , efficient and inexpensive technique helpful in cataloguing the biodiversity.
- With the increased resolution and maximum information obtained Mt DNA analysis is highly productive in molecular phylo-genetics.

Functions of mtDNA in Ichthyotaxonomy

➢Individual identification

➢Mixed Stock Fishery Analysis(MSFA)

➢To identify the phylogenetic relationship b/w the different species of fishes

➢To identify and arrangement of the species and stock on the basis of their genetic affinity

➢To identify the genetic diversity with in the stock & cultured species

Uses for mtDNA in Forensics

•mtDNA will be used when "biological evidence may be degraded [i.e. charred remains] or in small quantity“

•Cases in which evidence consists only of:
- –hairs
- –bones
- –Teeth

•Missing Persons Cases (use of skeletal remains)

•Establishing Individuals as suspects (hair evidence)

* Molecular Chronometer used in Molecular taxonomy

Signature sequence and its importance in tracing phylogeny among prokaryotes

- The rRNA molecule has been described as ultimate 'Molecular chronometer' enabling classification across all major taxonomic Classes.

Why rRNA molecule is considered as molecular chronometer or tool for molecular taxonomy?

1. Ubiquitous occurrence in all living beings
2. Functionally constant i.e. conserved over time(evolution of 16SrRNA is very slow).
3. Relatively smaller in size(50 helical stalk).
4. Having carrier regions which mutate at different rates.
5. Can easily be multiplied/amplified/sequenced and compared.

Why bacterial 16SrRNA used as a molecular clock in prokaryotes ?

- The bacterial 16S rRNA has a length of 1650 nucleotides, found on smaller ribosomal subunits (30S) of prokaryotic ribosomes. Since both mitochondria and chloroplast have their own ribosomes. So it is found in all three kingdoms.
- The sequence of 16S rRNA has been divided into distinct areas according to their degree of variability among different taxonomic groups.
- Actually the different areas of on rRNA molecule have a different mutation rate. Highly conserved region therefore be used to group bacteria into higher taxonomic order ,where as more variable regions allow classification at lower taxonomic level such as genes or species level.
- In recent times two organism that differ only by a few bases have diverged more recently in evolutionary time than organism that differ by more bases.

* Molecular Chronometer used in Molecular taxonomy Contd.

Example

- The following example show that how an evolutionary tree can be constructed for four hypothetical organism whose DNA sequence in the homologous region is known like-

  AACGTCGAAA(organism A).

  - **AACCTCGAAA(organism B).**
  - **AGGCTAGAAA(organism C).**
  - **AGGCTAGTAA(organism D).**

(1).'A' and 'B' differ by one base substitution at place 4, 'G' instead of "C".
(2). 'C' and 'D' differ only by one base substitution at 8th position 'T' instead of 'A'.
(3). 'A' and 'C' differ by three base substitution
(4). 'B' and 'C' differ by three base substitution.
(5). 'A' and 'D' differ by four base substitution.
(6). 'B' and 'D' duffer by four base substitution

Conclusion.

In the evolutionary history A and B appear to be very similar as do C and D, while A-B and C-D are more distantly related.

* Application of Bioinformatics in Molecular taxonomy

Role of Computer software

- Using DNA sequencer analyser, a large stretch of DNA can be sequenced and stored in computer data base for further analysis.
- During computational analysis once a nucleotide has been deduced to search an existing data base for similar homologous sequences and for generic genes or protein coding sequences , it becomes quite easier to gather information to construct phylogenetic tree.
- Other important sequences that may be used to define coding sequence are- ribosome binding sites, splice site junction, Poly A polymerase sequence, promoter sequences etc.
- Computer gathers such data and creates phylogenetic tree that accurately illustrates the divergence between different organism with linear distance being proportional to the number of accumulated errors.
- For this sequence analysis soft ware are used to search a new sequence for identity within a chosen data base.
- Software programmes like BLAST and FASTA provides the means to align the sequence.
- This homology searching provides important clues to the potential structure and function of a given DNA sequence.

Arbitrary Nuclear DNA marker in Molecular taxonomy

- Any genetic trait that can be identified with confidence and relative case and can be followed in a mapping population is known as genetic marker.
- Genetic marker is a specific location on a chromosome that is defined by a naked eye polymorphism as differences in electrophoretic mobility of specific protein or as difference in specific DNA sequence. Arbitrary nuclear DNA marker are used when a sequence of DNA of unknown function is targeted.
- The widely used method for amplifying unknown regions are-

Random Amplified Polymorphic DNA (RAPD)

- Random amplified Polymorphic DNA(RAPD) markers are DNA fragments from PCR amplification of random segments of genomic DNA with single primer of arbitrary nucleotide sequence.
- RAPD analysis is a PCR based molecular marker techni1que. Single short oligonucleotide primer is arbitrarily selected to amplify a set of DNA segments distributed randomly through out the genome (anonymous loci).
- RAPD uses random primer to generate multiple PCR products resulting in a fingerprint for a particular species. It is very fast, cheap and show very high amount of polymorphism and the marker does not require the prior

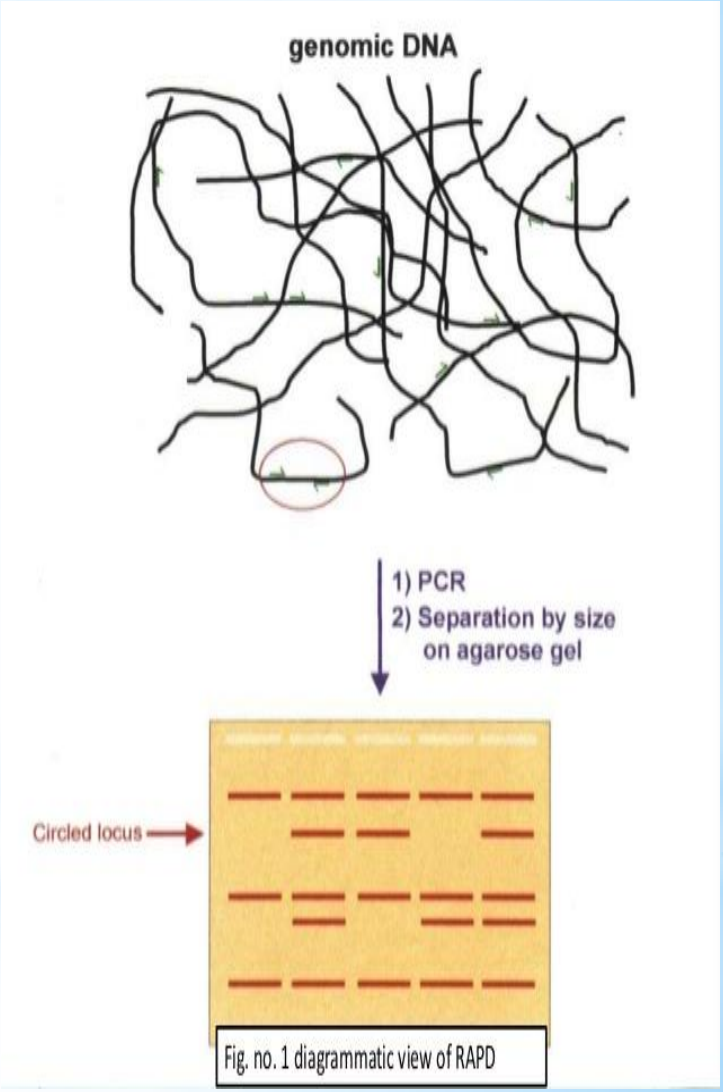


Fig. no. 1 diagrammatic view of RAPD

Steps in Random amplified Polymorphic DNA analysis

RAPD & ITS APPLICATION

PROCEDURE

❑RAPD involves following steps:-

1.The DNA of a selected species is isolated.

2. An excess of selected decaoligonucleotide added.

3. This mixture is kept in a PCR equipment and is subjected to repeated cycles of DNA denaturation-renaturation-DNAreplication.

4. During this process, the decaoligonucleotide will pair with the homologous sequence present at different locations in the DNA.

RAPD & ITS APPLICATION

PROTOCOL

Isolation of DNA

Keep the tubes in PCR thermocycler

Denature the DNA (94°C,1 min

DNA strands separated

Decaoligonucleotide enzyme, primer, Taq DNA polymerase,

Annealing of primer (36°C,2 min

Primer annealed to template DNA strands

DNA synthesis (72°C, 1.5 min

RAPD & ITS APPLICATION

PROCEDURE

❑RAPD involves following steps:-

5.DNA replication extend the decaoligonucleotide and copy the sequence continuous with the sequence with which the selected oligonucleotide has paired.

6.The repeated cycles of denaturation - renaturation-DNA replication will amplify this sequence of DNA.

7. Amplification will takes place only of those regions of the genome that has the sequence complementary to the decaoligonucleotide at their both ends.

8. After several cycles of amplification the DNA is subjected to gel electrophoresis.

9. The amplified DNA will form a distinct band. it is detected by ethidium bromide staining and visible fluorescence's under U.V. light

RAPD & ITS APPLICATION

PROTOCOL

Complementary strand synthesis

35 to 45 cycles

Amplified products separated by gel electrophoresis

Bands detected by Ethidium bromide staining

Applications , advantages and disadvantages of RAPD

Applications OF RAPD

- Molecular genetic marker have been developed into powerful tools to analyse genetic relationship and genetic diversity.
- As an extension to the variety of existing technique using polymorphic DNA markers, the RAPD may be used in molecular ecology to determine taxonomic identity, assess kinship relationship, analyse mixed genome sample sand create specific probes.

Advantages of RAPD

- It simultaneously target s multiple gene loci and is therefore more useful for discriminating closely related or cryptic species.
- DNA fingerprint is generated in a single reaction
- Data may be used for phylogenetic reconstruction in some instances.
- It is suitable for work on anonymous genome.
- It is quick and efficient screening for DNA sequence based polymorphism at many loci and low expanses technique. It does not involve any radio active assays.

Disadvantages of RAPD

- The major drawback of RAPD is the reproducibility and repeatability. At the same time a large number of products generated
- RAPD is a dominant marker and homozygous and heterozygous state can not be differentiated.
- It cannot be used on mixed samples.
- It is only useful as diagnostic, if RAPD fingerprint of unknown specimen has already been resolved for comparison.
- Co dominant RAPD markers observed as different sized DNA segments amplified from the same locus are detected only rarely.

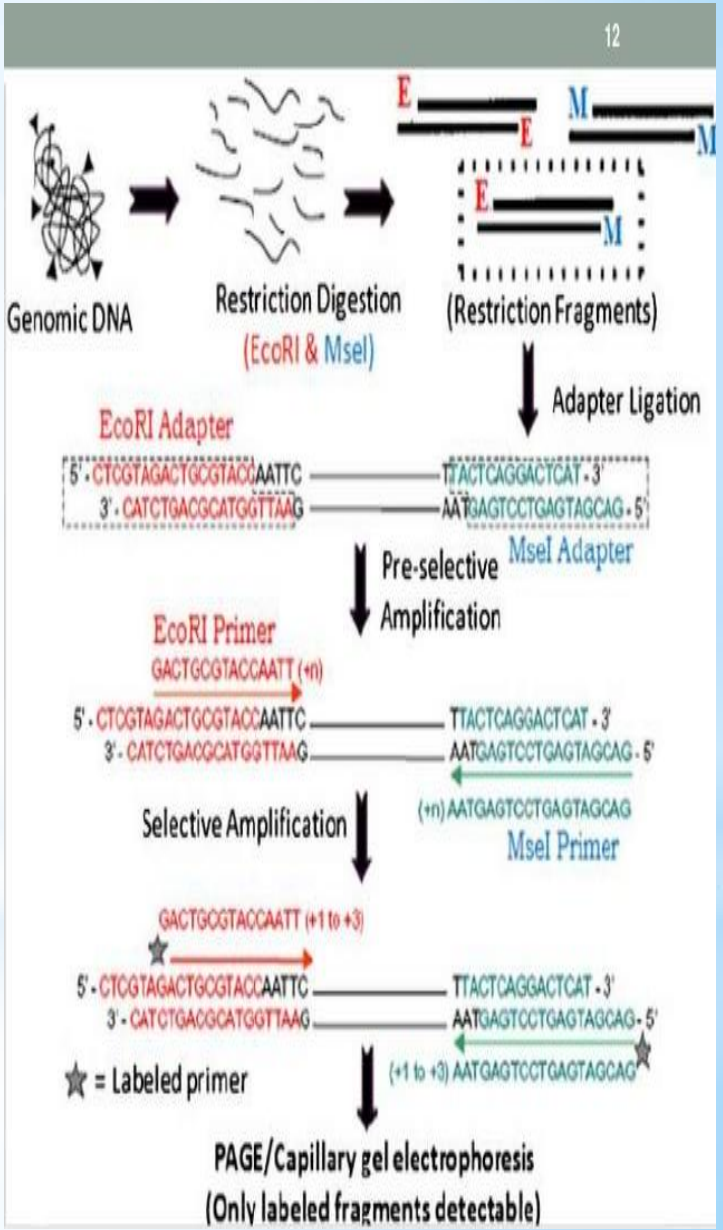
Amplified Fragment Length Polymorphism (AFLP) DNA Markers

- Amplified fragment length polymorphisms (AFLPs) provide an effective means of genotyping, particularly when little is known about the genome or genetics of an organism. It involves ligation of adaptors to digested DNA followed by PCR amplification using primer, that are primarily adaptor and partially gene specific.

- They may combine the benefit of both RAPD and RFLP. The total genomic DNA is digested using two restriction enzymes.

- Restriction enzymes cut the DNA and double stranded nucleotide adaptors are ligated to the ends of the fragments to serve as primer biding site for PCR amplification.

- Fragments are then amplified using PCR and the presence or absence of their varying lengths can then be visualized on polyacrylamide gel or capillary-based platform. Thus genetic polymorphism is studied.



Steps in AFLP Marker Test :At A Glance

7

Steps

1. Digestion
2. Adaptor ligation
3. Amplification
4. Electrophoresis

8

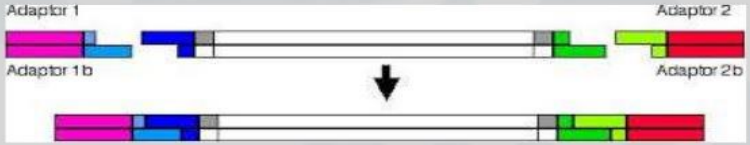
Digestion

- Different restriction endonucleases are used in digestion.
- One is four base cutter, *MseI*
- The other one is six base cutter, EcoR*I*

MseI	5'TTAA3'
EcoRI	5'GAATTC3'

Adaptor ligation

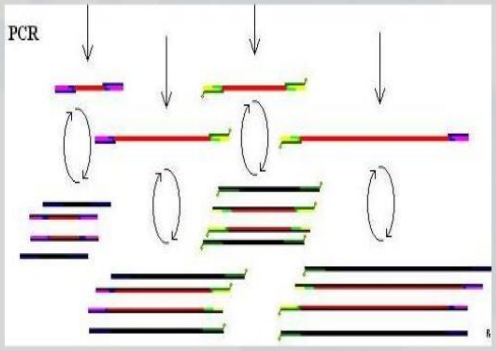
- Two different adaptors (short double stranded DNA with sticky ends) are ligated to the digested fragments.
- One adaptor will complement to *MseI* cut end and the other will complement to the *EcoRI* cut end.



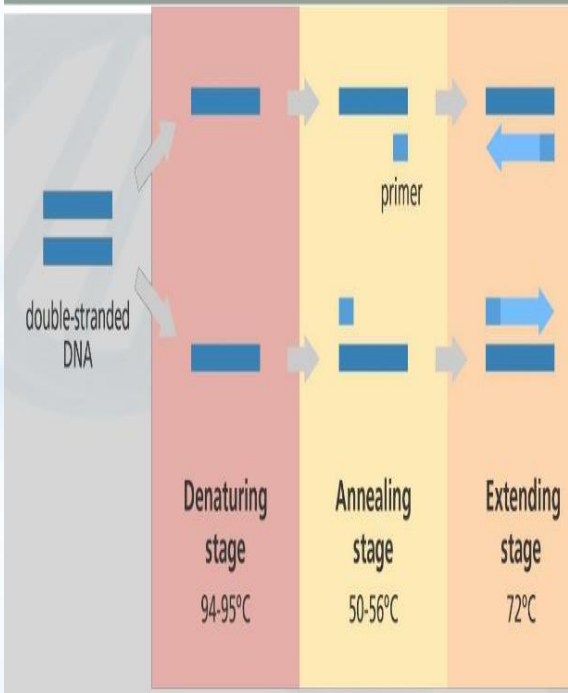
9

Amplification

- Selected fragments are amplified and separated by polyacrylamide gel electrophoresis.
- By repeating this second amplification with other primer pairs a different subset of the genome is amplified.



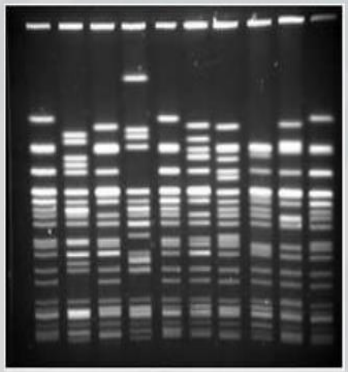
10



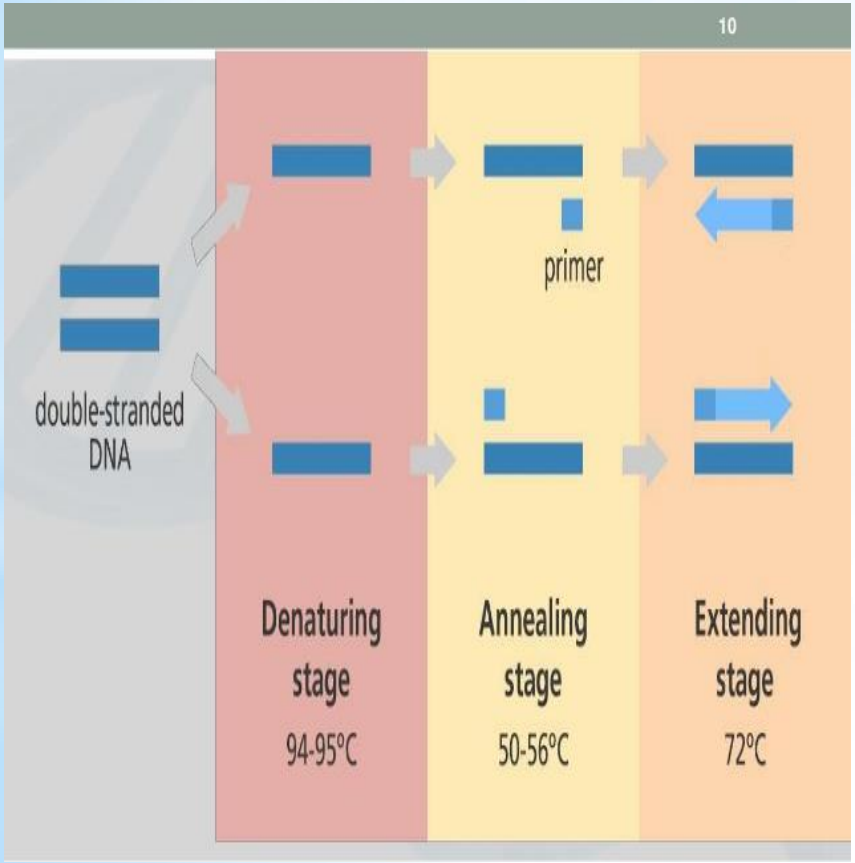
11

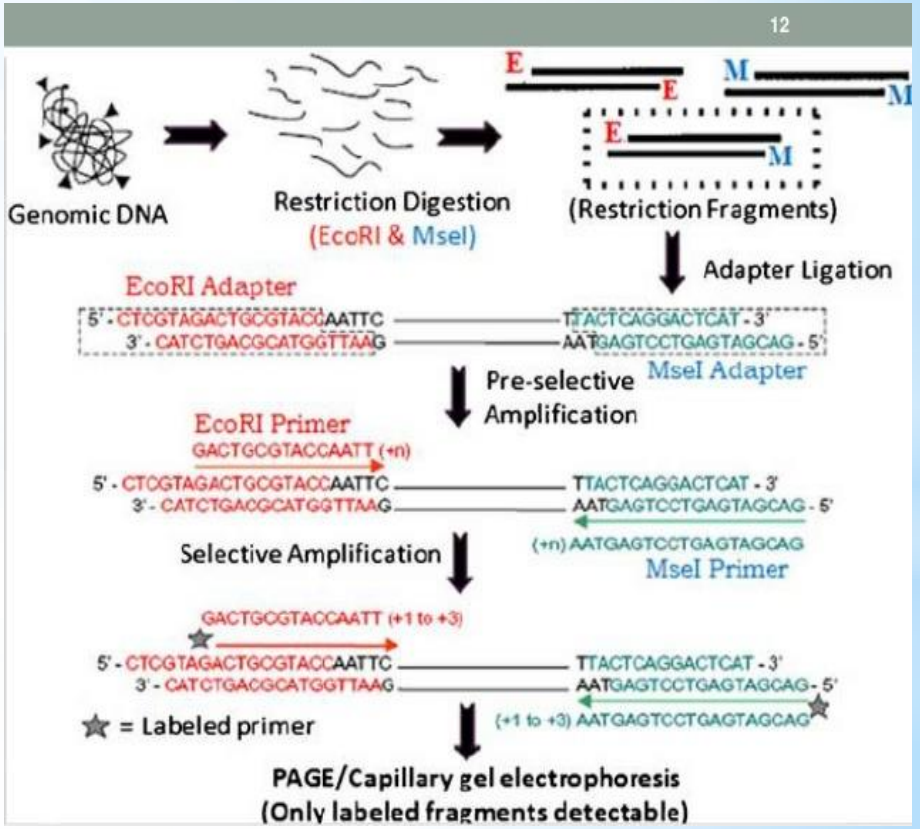
Electrophoresis

- Polyacrylamide gel is used for separating DNA bands.
- Normally, 30-100 DNA bands can be detected by AFLP on polyacrylamide gel.



Steps in AFLP Marker Test :At A Glance





* Applications , advantages and Disadvantages of AFLP

Applications

- The AFLP has the capability to detect various polymorphism in different regions simultaneously.
- AFLP has been widely used for identification of genetic variations in strains or closely related species of plants , fungi animals and bacteria.
- The AFLP technique has been used extensively in criminal and paternity test , also to determine slight variation within the population and linkage studies to generate maps for quantitative Trait Locus(QTL) analysis.

Advantages

- It simultaneously target multiple gene loci and therefore useful for discriminating closely related or cryptic species.
- It is very sensitive and more robust than RAPD.
- Data may be used for phylogenetic reconstruction in some instances.
- The whole genome analysis is possible without any knowledge of known genome sequence by changing the selective nucleotide.

Disadvantages

- It is a complex procedure and requires manipulations in addition to PCR.
- It cannot be used on mixed samples.
- It is only useful as diagnostic if AFLP fragment of the unknown specimen has already been resolved for comparison.
- AFLP is very sensitive for detecting genetic polymorphisms but requires relatively large amounts of high-quality DNA and has difficulty with mixture analysis. Thus, AFLP is not an ideal candidate for genotyping forensic rich samples but has been used for population genetics of plants.

Restriction Fragment Length Polymorphism (RFLP) DNA Markers

The University of Edinburgh

What is RFLP ?

- Restriction Fragment Length Polymorphism is a variation in the length of a DNA fragment produced by a specific restriction enzyme acting on a DNA from different individuals that usually results from a genetic mutation.
- If two organisms differ in the distance between site of cleavage of a particular restriction endonuclease, the length of the fragments produced will be different when the DNA is digested with a restriction enzyme.
- RFLP analysis is the detection of the change in the length of the restriction fragments.

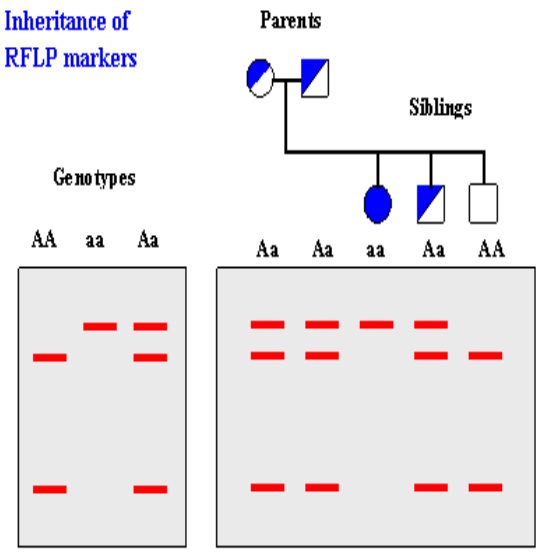
The University of Edinburgh

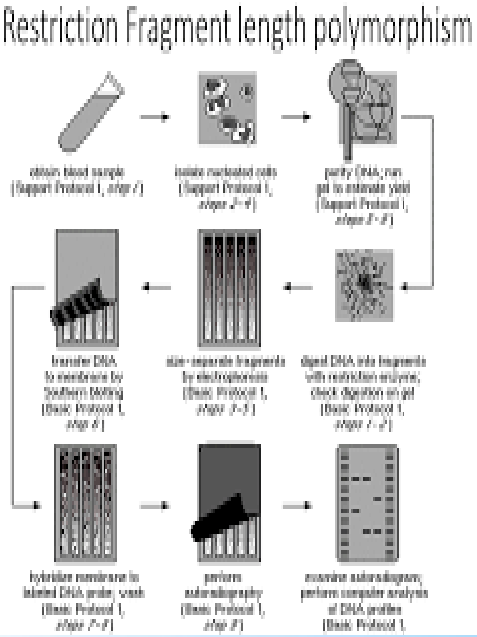
Continuous

- A restriction enzyme cuts the DNA molecule at every occurrence of a particular sequence, called restriction site.
- For Example, HindIII enzyme cuts at AAGCTT.
- If we apply a restriction enzyme on DNA, it is cut every occurrence of the restriction site into a million restriction fragments each a few thousands nucleotides long.

Continuous

- Any mutation or polymorphism of a single nucleotide may destroy (AAGCTT for HindIII) and change the length of the fragment.
- The term polymorphism refers to the slight differences between individuals, in base pair sequences of genes
  or
- A polymorphism is a clinically harmless DNA variation that does not affect the phenotype.
- So RFLP analysis is a technique which is used to detect the change in the length of the restriction fragments.





In molecular biology, restriction fragment length polymorphism (RFLP) is a technique that exploits variations in homologous DNA sequences, known as polymorphisms, in order to distinguish individuals, populations, or species or to pinpoint the locations of genes within a population

Techniques in RFLP: At A Glance

The University of Edinburgh

RFLP Technology

- RFLP detection depends on the possibility of comparing band profiles generated after restriction enzyme digestion of target DNA. These differences in fragment lengths can be seen after gel electrophoresis, hybridization and visualization. The basic steps involved are as follows:

1. Isolation of DNA
2. Restriction Digestion & Gel Electrophoresis
3. DNA transfer by Southern blotting
4. DNA hybridization

The University of Edinburgh

(1) Isolation of DNA

- Isolating DNA is the first step for many DNA-based technologies. DNA is found either in nuclear chromosomes or in organelles (mitochondria and chloroplast).
- To extract DNA from its location, several laboratory procedures are needed to break the cell wall and nuclear membrane, and so appropriately separate the DNA from other cell components.
- When doing so, care must be taken to ensure the process does not damage the DNA molecule and that it is recovered in the form of a long thread.

The University of Edinburgh

(2) Restriction Digestion & Gel Electrophoresis

- Extracted DNA is digested with specific, carefully chosen, restriction enzymes.
- Each restriction enzyme, under suitable conditions, will recognize and cut DNA resulting in a restriction fragments of different lengths.
- The thousands of restriction fragments produced are commonly separated by Gel Electrophoresis on agarose gels.

The University of Edinburgh

Continuous

- For visualization of DNA bands on the gel, it is stained with ethidium bromide. But staining alone cannot detect polymorphisms.
- Hybridization must therefore be used to detect specific fragments.

Techniques in RFLP: At A Glance

The University of Edinburgh

(3) DNA transfer by Southern blotting

- DNA transfer is called "Southern blotting", after E.M. Southern (1975), who invented the technique.
- In this method, the gel is first denatured in a basic solution of NaOH and placed in a tray. A porous Nylon or nitrocellulose membrane is laid over the gel.
- All the DNA restriction fragments in the gel transfer as single strands by capillary action to the membrane. All fragments retain the same pattern on the membrane as on the gel.

The University of Edinburgh

(4) DNA Hybridization

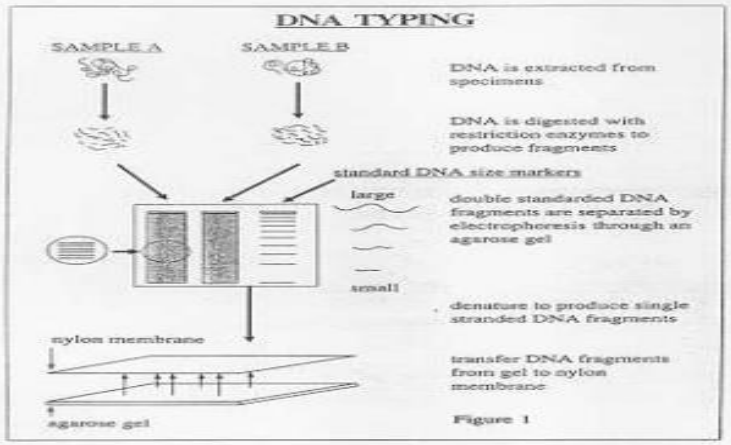
- The membrane with the target DNA is exposed to the DNA probe (radioactively labelled). On the basis of availability and complementarity, hybridization will occur.
- The DNA probe is a singe-stranded molecule, conveniently labelled, using any standard method (e.g. a radioisotope), and hybridize with the target DNA, which is stuck to the membrane.

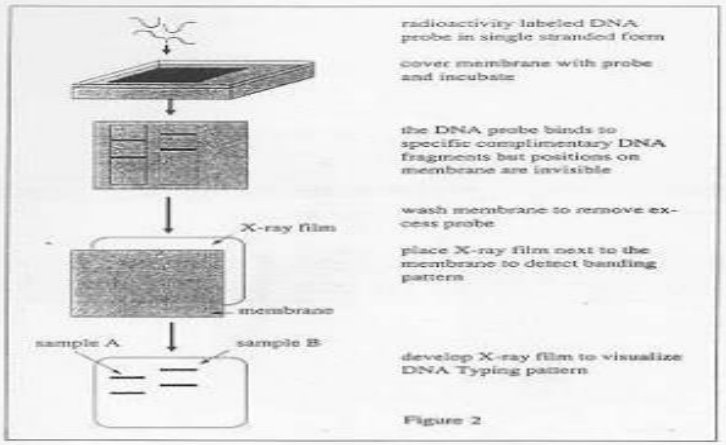
The University of Edinburgh

Continuous

- Thus the DNA probe binds with the sequences that are complementary to it among the thousands or millions of undetected fragments that migrate through the gel.
- Desired fragments maybe detected after simultaneous exposure of the hybridized probe to X-rays, which on exposure appear as black spots on the X-ray film.

Figure 3





Applications, Advantage and Disadvantages of RFLP

Applications

- RFLP can be used for paternity cases or criminal cases to determine the source of a DNA sample (i. e. It has forensic applications).
- RFLP can be used to detect mutations.

Advantages

- It is simple and accurate process and does not require sequence information.

Disadvantages

- It is an expensive process.
- It is very slow, labour intensive and time consuming process.
- It requires a very large amount of DNA
- It is difficult to automate.

Specific DNA Marker in Molecular Taxonomy

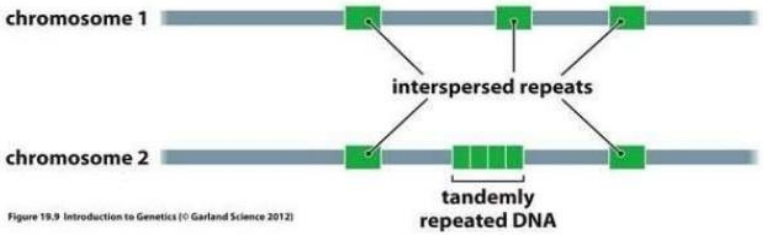
Repetitive DNA Pattern: At A Glance

✓ Repetitive DNA -two patterns of distribution in the chromosomes.

The two major classes of repetitive elements are

✓ Interspersed elements and

✓ Tandem arrays.



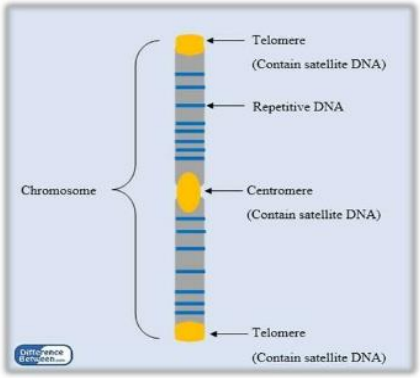
1. Interspersed repeated elements are usually present as single copies and distributed widely throughout the genome.

2. Tandemly-repeated DNA may be further classified according to the length and copy number of the basic repeat units as well as its genomic localization;

3 main types

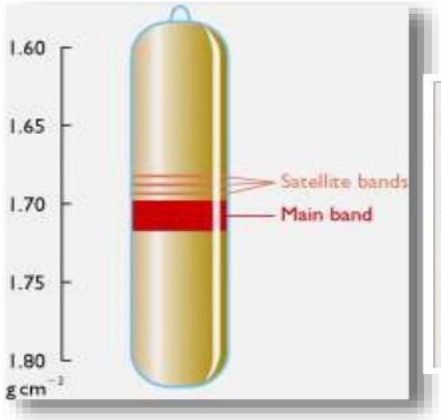
- **Satellite**
- **Minisatellite**
- **Microsatellite**

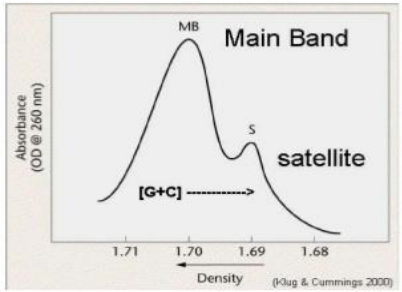
- A **satellite** is a highly repetitive DNA sequence with each repeated sequence ranging from a thousand to several thousand base pairs.
- The entire satellite can be up to 100 million base pairs long,



- Occur in regions of heterochromatin.
- Satellite DNA - clustered in discrete areas such as centromere.

- Differ in density from the remaining chromosomal DNA and are separable as minor band in **CsCl Density Gradient Centrifugation**, called as **SATELLITE DNA.**





Specific DNA markers in Molecular taxonomy

- There are two main classes of repetitive and highly polymorphic DNA- Mini satellite DNA (VNTRs) and Microsatellite DNA(STRs).

Mini-satellite DNA or Variable number of tandem repeats (VNTR)

- Variable number of tandem repeats are a sequence of DNA i.e. repeated tens or hundreds to thousands of time in nuclear genome.
- They repeat in tandem, vary in number in different loci and differently in individuals.
- Minisatellite DNA referring to genetic loci with repeats of length 9-65 bp, are not restricted to Nuclear DNA, rather they are also present in MtDNA or cpDNA .
- VNTR have been thought to arisen due to slippage during DNA replication or as a consequences of unequal crossing over. They may be of two types- Multi-locus mini-satellite and single locus minisatellite.

Multi locus minisatellite

- They are composed of tandem repeats of 9-65bp and have a total length ranging from 0.1to7 kb

Disadvantages

1. Multilocus minisatellite is are generally used in percentage analysis.

They are less useful for population genetics analysis unless we use large sample size.

2. The complexity of mutation process undergone by mini satellite loci is also a limitations.

Single locus minisatellite

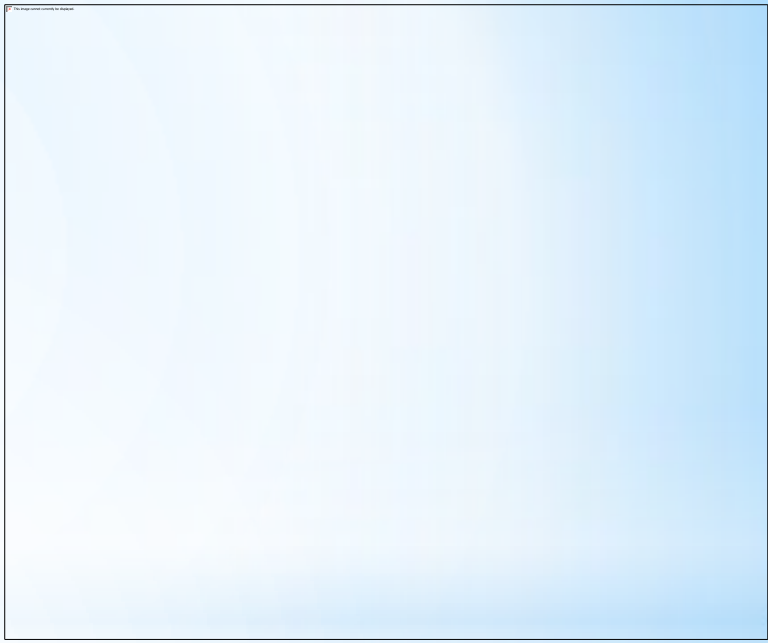
Advantages

1. Single locus minisatellite have been very useful and successful detecting in genetic variation within and between population.

2. It has also been used in fisheries for forensic, parentage, genetic identity, estimating mating success and conforming gynogenesis.

Disadvantages

1. Single locus minisatellite analysis requires reasonable quantities of high quality DNA.

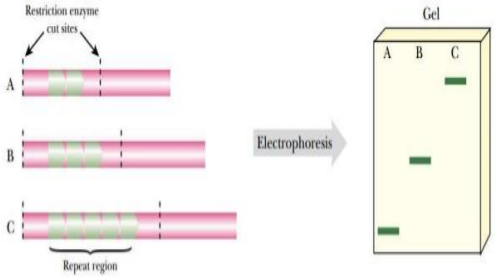


* VNTR (mini satellite)DNA Marker : Characteristics

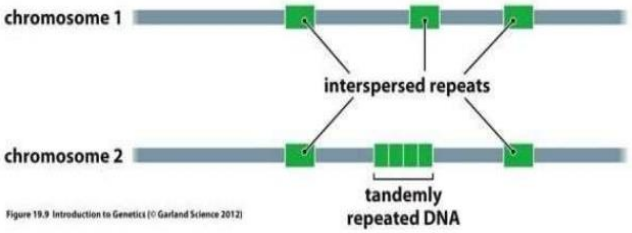
Variable Number Tandem Repeats

- Tandem repeats occur in DNA when a pattern of one or more nucleotides is repeated and the repetitions are directly adjacent to each other.
- An example would be:
  AATTTTCCGGCCCCAAAATTCCAATTTTCCGGCCCCAAAATTCCAATTTTCCGGCCCCAAAATTCC
- These can be found on many chromosomes, and often show variations in length between individuals.
- The number of elements in a given region may vary, hence they are known as **variable number tandem repeats.**
- Each variant acts as an inherited allele, allowing them to be used for personal or parental identification.
- Scientists use polymorphic loci that are known to contain VNTRs/STRs in order to differentiate people based on their DNA.

- Because different alleles consist of different numbers of repeats, VNTR alleles can be identified by their lengths.
- If DNA fragments of different lengths are placed on a semisolid medium (gel) in an electric field, they migrate at different rates;
- Different-sized fragments can therefore be identified by the distance they travel between electrodes in such a gel



- Individual repeats can be removed from (or added to) the VNTR via recombination or replication errors.
- Different individuals within a population may have different numbers of repeats.
- For example, one person might carry 3 tandem repeats in a particular gene, while another person might bear 15 repeats



- **In standard genetics, we often think of an allele as a variation of gene that would result in a difference in a physical manifestation of that gene.**
- **In the case of VNTRs, these alleles are simply a difference in number of repeats.**
- The length of DNA within this locus is either longer or shorter and gives rise to many different alleles.
- A locus in a population is polymorphic if it has more than one allele.
- This is often used in forensic science or in maternity/paternity cases.
- Any variation of a locus is referred to as an **allele**.
- VNTR regions are not genes, and our interest in them is solely related to their use for identifying individuals. We therefore refer to them as *markers*.

- First human minisatellite was discovered in 1980 by A.R. Wyman and R. White
- Alec Jeffreys coined the term 'minisatellite'
- 14-100 bp sequences which repeats 20-50 times
- **Hypervariable minisatellites**, highly polymorphic, contain 100-1000 repeats
- HVMS DNA sequences; Not transcribed (ex; MUC1 locus)
- The variability is not just limited to the number of repeat but also the sequence of repeats can vary in different members of an array
- Of the 300 minisatellite families, which have been typed less than 10 are hypervariable

- Due to the repetitive nature of the sequence have allowed further increases/decreases in the number of repeats resulting in the polymorphisms observed
- These can be isolated from an individual's DNA and therefore relatively easy to map.

VNTRs have a disadvantage as genetic markers:

- They tend to bunch together at the ends of chromosomes, leaving the interiors of the chromosomes relatively devoid of markers.

Specific DNA markers in Molecular taxonomy : Microsatellite(STRs)

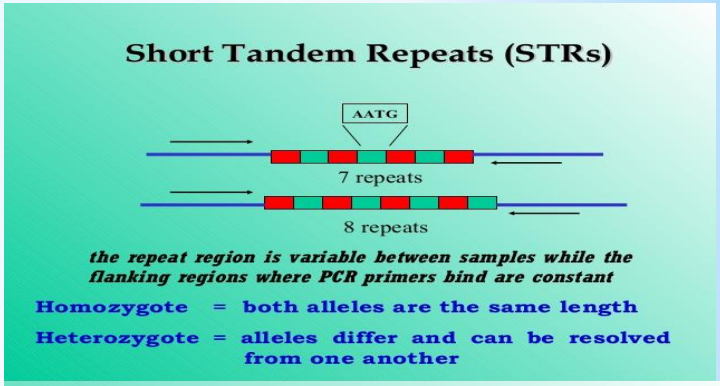
1. **Microsatellite** is a simple DNA sequence which is repeated several times across various points in the DNA of an organism.

2. They are also known as short tandem repeats (STRs), consisting of tracts of repeats of 2-8 bp (1-6) long .They are highly variable and these loci can be used as markers.

3. Microsatellite occurs once in every 10kbp, while minisatellites occur in every1500kbp in fishes. So they are more useful in genome mapping and population genetics studies.
4. They are highly variable, noncoding and selectively neutral.

Principle

Microsatellite analysis involves DNA extraction, amplification of microsatellite nuclei using specific primer in a PCR machine (thermocycler) and examination of band under SDS PAGE .The basic assumption while using microsatellite loci is that the predicted amount of sequence divergence between units of interest is directly related to the length of time since separation.
Cross amplification with primers developed in closely related species is also possible which minmizes the cost associated with detecting microsatellite sequence in different species.
The recent introduction of automated genotyping machine has lead to the use of a large number of samples and loci, which has increased precision and speed with micro satellite analysis.



Microsatellites – Types

Based on repeat pattern

1. Perfect – CACACACACACACACACACA
2. Imperfect – CACACACACA CACACACA CACACACA
3. Compound – CACACACACACA CATACATACATA CATACATACATA
4. Complex – CACACACACACACACA AATAATAATAATAATAATAAT

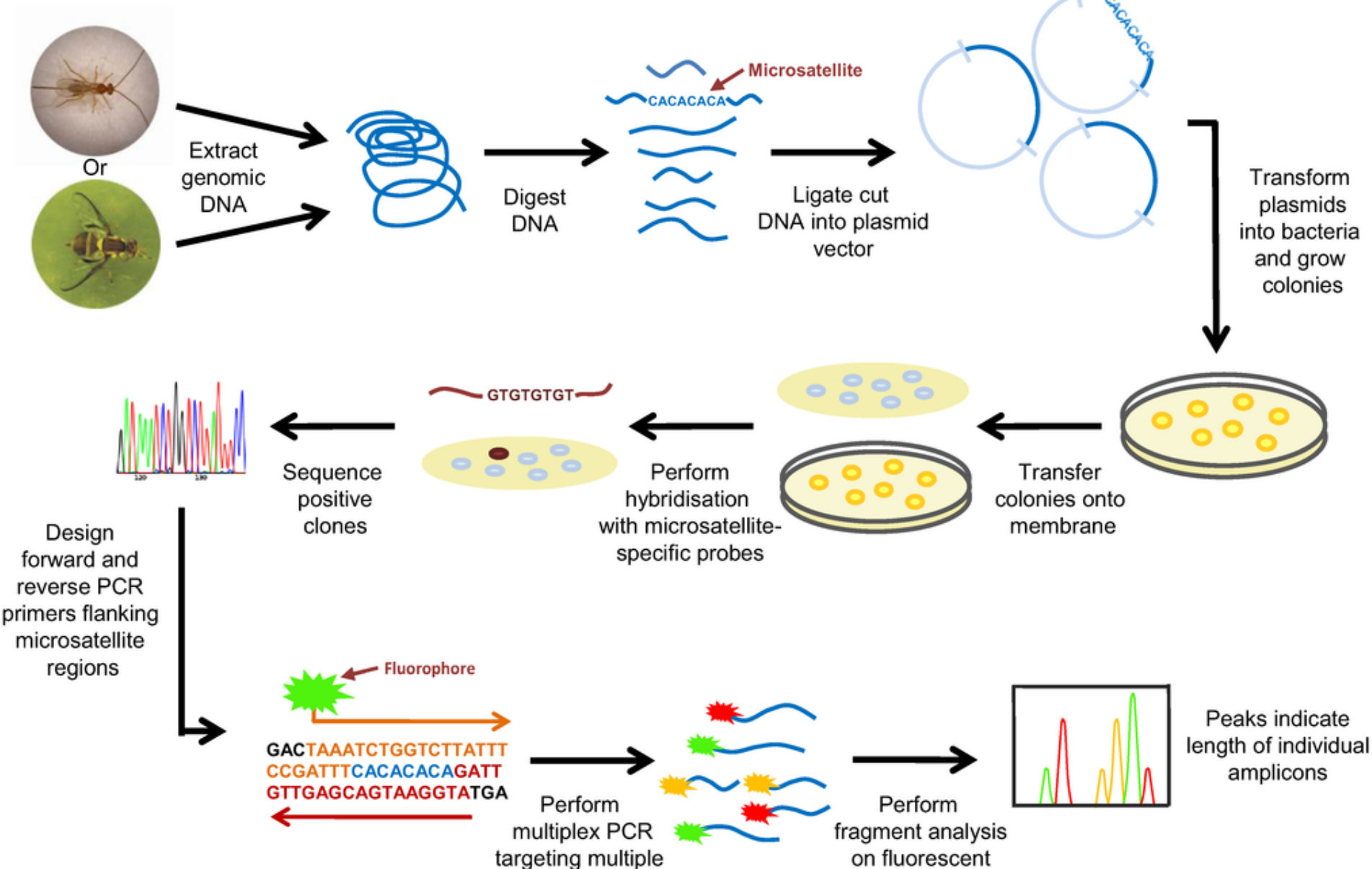
Based on number of base pairs

1) Mono (e.g. CCCCCCCC or AAAAAA)
2) Di (e.g. CACACACACA)
3) Tri (e.g. CCA CCA CCA CCA)
4) Tetra (e.g. GATA GATA GATA GATA GATA GATA GATA)

Minisatellites: - (9 – 65 base pairs repeated from 2 to several hundred times)

CGCCATTGTAGCCAATCCGGGTGCGATTGCAT CGCCATTGT AGCCAATCCGGGTGCGATTGCAT CGCCATTGTAGCCAATCCGGG TGCGATTGCAT CGCCATTGTAGCCAATCCGGGTGCGATTGCAT CGCCATTGTAGCCAATCCGGGTGCGATTGCAT

Development and Use of Microsatellite Marker for Species Identification



Specific DNA markers in Molecular taxonomy :Microsatellite(STRs)

Advantages

1. Micro satellite analysis simultaneously targets multiple gene loci and therefore more useful for discriminating closely related or cryptic species.
2. When fluorescent primers are used fragment analysis is readily automated .
3. Assays can be multiplexed during PCR and detection(fragment analysis) phase.
4. Some microsatellite assays can be applied across a number of different species.

Microsatellite analysis have been extensively used in fisheries and aquaculture for phylo-genetics and phylo-geographical studies, population genetics, structure, biodiversity, conservation, stocking impact and hybridization. Besides, it is extensively used for forensic identification, genome mapping, determination of kinship and behavioural pattern.

Disadvantages

1. Assay development is tissue consuming initially and cannot be used in mixed samples.
2. The presence of null alleles and shutter bands are major constraints of microsatellite markers.
3. Null allele occurs when mutation occur at primary binding sites of microsatellite locus it reduces the accuracy in terms of relatedness analysis and assignment test.
4. Shutter bands occur when a ladder of bands differing in 1-2 bp is seen and these occur due to slipped strand impairing PCR or incomplete denaturation of amplification products.

Specific DNA markers in Molecular taxonomy :Polymerase chain reaction(PCR)

- **Polymerase chain reaction (PCR)** is a method used widely in molecular biology to make millions to billions of copies of a specific DNA sample rapidly, allowing scientists to take a very small sample of DNA and amplify it to a large enough amount to study in detail.
- PCR was invented in 1983 by the American biochemist Kary Mullis. It is fundamental to much of genetic testing including analysis of ancient samples of DNA and identification of infectious agents. Using PCR, copies of very small amounts of DNA sequences are exponentially amplified in a series or cycles of temperature changes.
- Applications of the technique include DNA cloning for sequencing, gene cloning and manipulation, gene mutagenesis**; construction of DNA-based phylogenies**, or functional analysis of genes; diagnosis and monitoring of hereditary diseases; amplification of ancient DNA;] analysis of genetic fingerprints for DNA profiling (for example, in forensic science and parentage testing); and detection of pathogens in nucleic acid tests for the diagnosis of infectious diseases.
- A basic PCR set-up requires several components and reagents, including:a DNA template that contains the DNA target region to amplify, a DNA polymerase(heat-resistant Taq polymerase),two DNA primers that are complementary to the 3' (three prime) ends of each of the sense and anti-sense strands of the DNA target ,deoxynucleoside triphosphates, or dNTPs (sometimes called "deoxynucleotide triphosphates"; buffer solution providing a suitable chemical environment for optimum activity and stability of the DNA polymerase and bivalent cations, typically magnesium (Mg) or manganese (Mn) ions; Mg2+

* Specific DNA markers in Molecular taxonomy :Polymerase chain reaction(PCR)

PCR consists of a series of 20–40 repeated temperature changes, called thermal cycles, with each cycle commonly consisting of two or three discrete temperature steps. The individual steps common to most PCR methods are as follows:

Initialization: This step is only required for DNA polymerases that require heat activation by hot-start PCR. It consists of heating the reaction chamber to a temperature of 94–96 °C (201–205 °F), or 98 °C (208 °F) if extremely thermostable polymerases are used, which is then held for 1–10 minutes.

Denaturation: This step is the first regular cycling event and consists of heating the reaction chamber to 94–98 °C (201–208 °F) for 20–30 seconds. This causes DNA melting, or denaturation, of the double-stranded DNA template by breaking the hydrogen bonds between complementary bases, yielding two single-stranded DNA molecules.

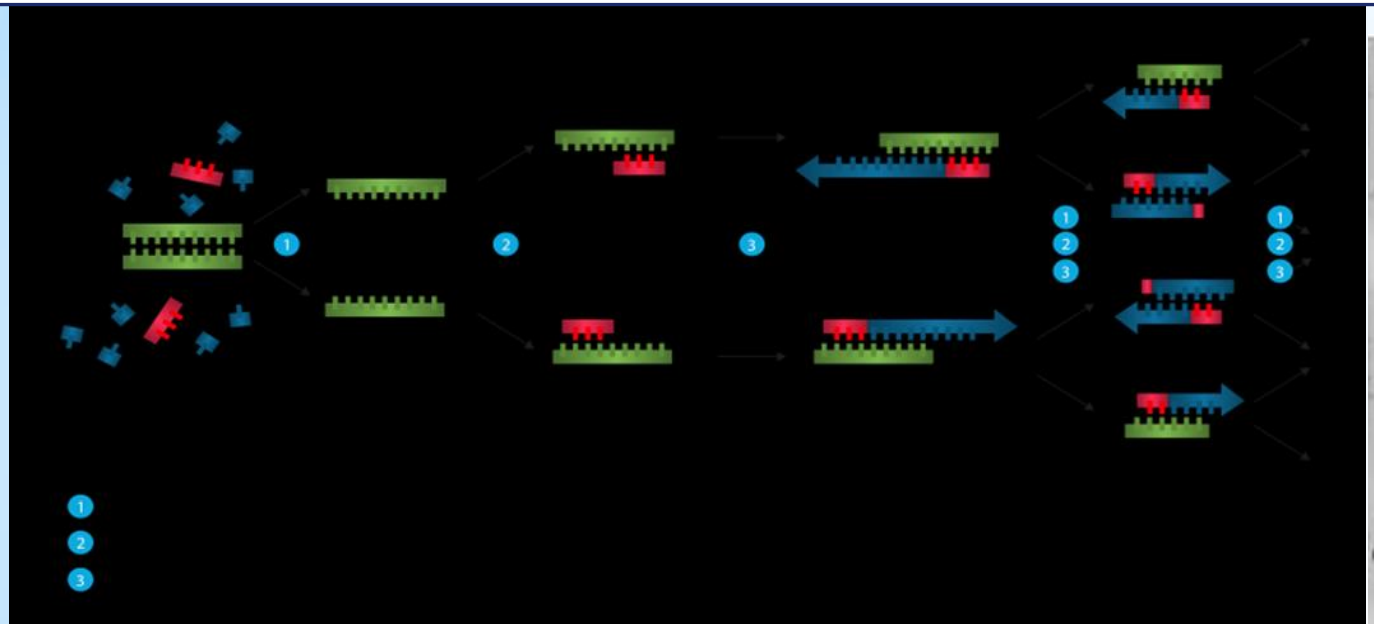
Annealing: In the next step, the reaction temperature is lowered to 50–65 °C (122–149 °F) for 20–40 seconds, allowing annealing of the primers to each of the single-stranded DNA templates. Two different primers are typically included in the reaction mixture: one for each of the two single-stranded complements containing the target region. The primers are single-stranded sequences themselves, but are much shorter than the length of the target region, complementing only very short sequences at the 3' end of each strand.

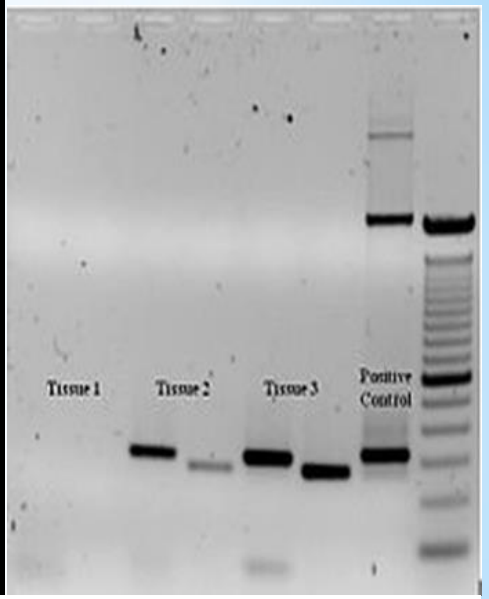
The processes of denaturation, annealing and elongation constitute a single cycle. Multiple cycles are required to amplify the DNA target to millions of copies. The formula used to calculate the number of DNA copies formed after a given number of cycles is 2n, where n is the number of cycles. Thus, a reaction set for 30 cycles results in 230, or 1,073,741,824, copies of the original double-stranded DNA target region.

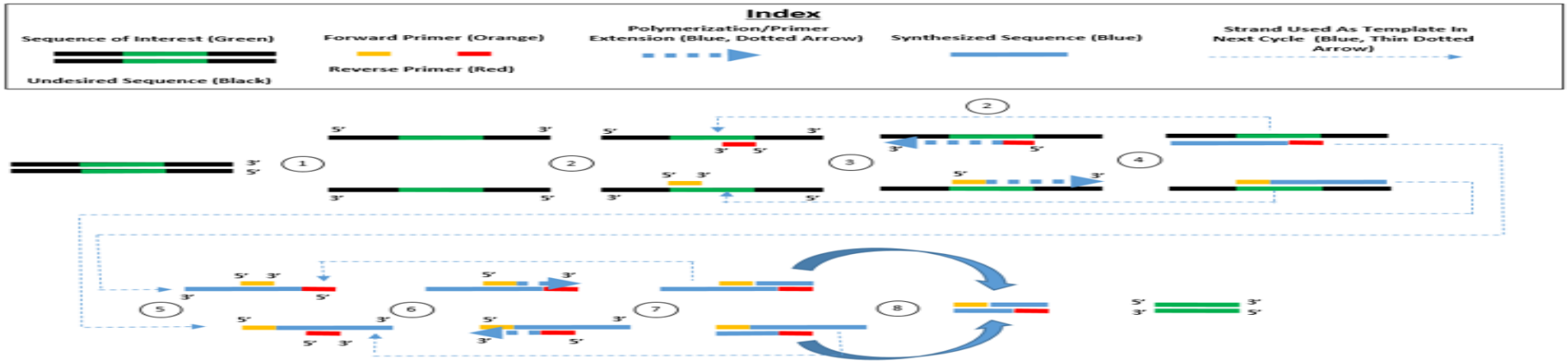
Final elongation: This single step is optional, but is performed at a temperature of 70–74 °C (158–165 °F) (the temperature range required for optimal activity of most polymerases used in PCR) for 5–15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully elongated.

Final hold: The final step cools the reaction chamber to 4–15 °C (39–59 °F) for an indefinite time, and may be employed for short-term storage of the PCR products.

Specific DNA markers in Molecular taxonomy :Polymerase chain reaction(PCR)







1. The DNA double helix is melted apart at T> 90°C and its strands separate.
2. The temperature is decreased to slightly below the T_m of both the primers being used. Both primers bind to the available strands. These primers are supplied in excess to insure that the strands do not only come back and reanneal to one another.
3. Polymerization (extension) occurs via DNA Polymerase in the 5' to 3' direction on each strand.
4. Incorporated additional nucleotides give rise to new strands that extend past the sequence of interest.
5. The previously polymerized strands act as template for the other primer (if forward primer bound first, reverse primer now binds and vice versa).
6. Polymerization occurs via DNA Polymerase in the 5' to 3' direction on each strand, this time ending at the end of the sequence of interest.
7. Incorporated additional nucleotides give rise to new strands that only encode the sequence of interest.
8. The synthesized strands encoding the sequence of interest anneal to one another to form the end product.

* **Specific DNA markers in Molecular taxonomy :Polymerase chain reaction (PCR)**

Specific PCR

Principle: It is a targeted assay giving a presence or absence result for a particular gene or species.

Advantages

1. It is useful diagnostically as it targets a specific region.
2. It can be used to target specific genes species or strain within a mixed sample.
3. No frequency of PCR product is required.

Disadvantages

1. It requires specific primer design, assay optimization, specificity testing prior to use as a diagnostics.

Size Differential PCR

Principle: It employs generic PCR primer but yields amplicon, that differ in length, usually targets the Intergenic transcribed spacer region (ITS).

Advantage

1. it can discriminate between a range of species simultaneously.
2. Differentiation is determined by electrophoresis.
3. It does not require screening of amplicon.

Disadvantages

1. size of amplicon needs to vary substantially to enable discrimination.
2. ITS regions contain repetitive region that can results in PCR products with multiple beds.

* Specific DNA markers in Molecular taxonomy :Polymerase chain reaction (PCR)

- **Multiplex PCR**

Principle : It combine multiple primer sets with different specification in a single assay.

Advantages

- It detects and differentiates multiple species in a single assay.
- It can be used as multiple genetic marker.
- Differentiation is conformed by electrophoretic size differentiation so no downstream processing of amplicon is required.
- It is useful for simultaneous detection of species in mixed samples (Detection of host and parasitoid DNA in one assay).

Disadvantages

- It is a difficult assay technique to optimize due to the presence of multiple primer set .
- Potentially cross hybridization of primers may interfere with reaction.

- **PCR-RFLP**

Principle: It involves discrimination of species based on restriction profile of amplicon.

Advantage

- It can discriminate range of species simultaneously.
- It can be used on a range of genetic marker(not restricted to size variable markers).
- It can provide an additional level of discrimination, if differentiation based on size fails.
- It may be able to detect some new types in some instances.

Disadvantages

- It requires downstream digestion of amplified DNA.
- Mutation may occasionally result in unidentified RFLP pattern.

* Specific DNA markers in Molecular taxonomy :Polymerase chain reaction (PCR)

Quantitative PCR

Principle : Short regions of DNA are PCR amplified and products are detected either by SYBR Green(double stained DNA dye)or via specific probe labelled with fluorescent dyes.

Advantages

- Amplification is monitored in real time against standard of known concentration allowing for quantification of target DNA.
- When using specific probe for amplicon detection, the reaction can be multiplexed for simultaneous detection up to 4 or 5 species and can be used on mixed samples.
- No electrophoresis is required. Detection is automated and involves detection of fluorescent entities. It allows for rapid and high throughput detection.

Disadvantages

- It need s specialized equipment.
- Multiplexing is limited by choice of limited dyes.

Loop mediated Isothermal Amplification(LAMP)

Principle: It employs three sets of specific primers used for amplification under isothermal condition.It yields a ladder of amplicons on electrophoresis. Here, amplicons can be detected using SYBR Green (double stained DNA dye).

Advantages

- It provides rapid and specific amplification under isothermal condition.
- The technique is potentially most suitable for field condition.
- It can be used with mixed samples due to primer specificity.

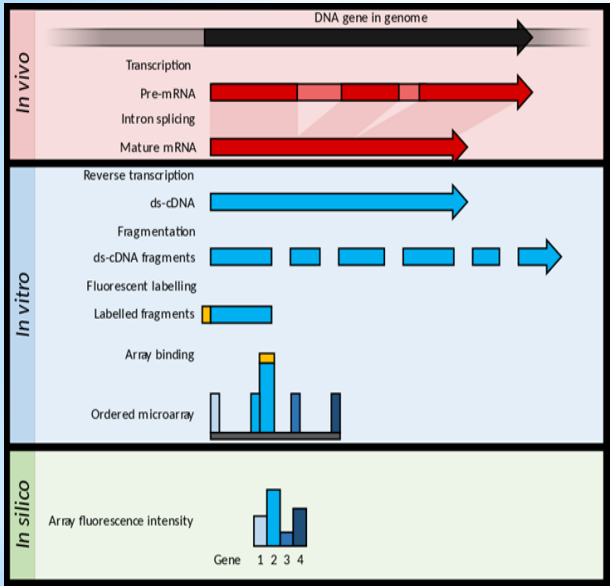
Disadvantages

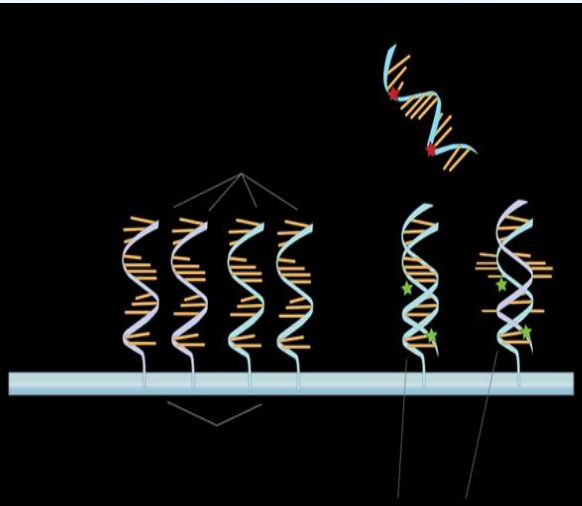
- The assay has a relatively complex design.
- It is only suitable for field condition, when paired with simple DNA extraction method.

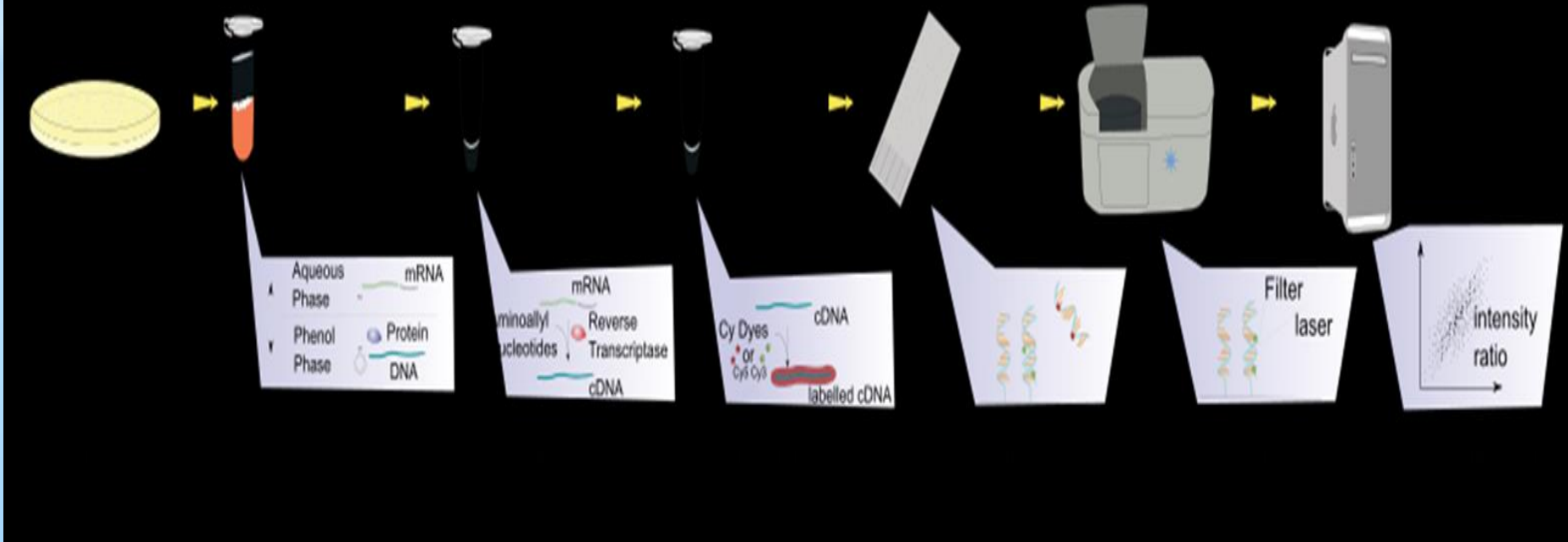
Other Useful DNA Markers in Molecular taxonomy :DNA Micro array

- **DNA Micro array/ DNA microchips** A DNA microarray (also commonly known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface. Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. A dDNA microarray is a test that allows for the comparisons of thousands of genes at a time.
- The principle behind microarrays is that complementary sequences will bind to each other. The unknown DNA molecules are cut into fragments by restriction endonucleases; fluorescent markers are attached to these DNA fragments. These are then allowed to react with probes of the DNA chip.
- In this technique cDNA or RNA (molecules of known sequences or probe, reporters, oligos, or antisense RNA) covalently attached on the solid surface. Each DNA spot contains pico moles (10−12 moles) of a specific DNA sequence. The experimental DNA or RNA (unknown sequences or test, target, or sample DNA or RNA) are tagged with fluorescent dye and poured over the probe area for hybridization. Probe-target hybridization is usually detected and quantified by techniques like autoradiography, laser scanning, fluorescence, and enzyme detection devices. Such techniques can be used to read the chip surface and hybridization pattern.
- Each DNA spot contains many thousands of copies of a specific DNA sequence, known as probes. These usually correspond to a short section of a gene – generally at the 3′ end. Each microarray includes one or a few probe sets for each interrogated gene. These are used to hybridize a cDNA sample (the target) under high-stringency conditions. Probe–target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine the relative abundance of transcripts in the target sample. Data on about 700 000 sample hybridizations performed on DNA microarrays are accessible through the databases Gene Expression Omnibus (GEO) at NCBI, and Array Express at EBI.
- They have diverse applications in gene expression analysis, DNA sequencing, characterization of

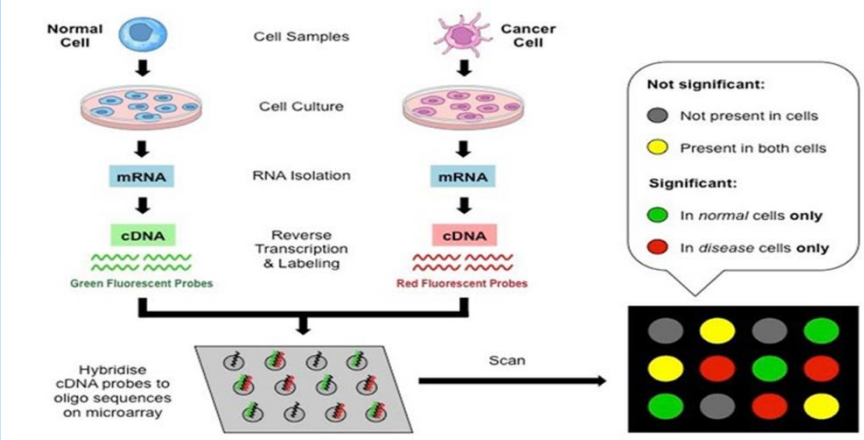
Steps in DNA Microarray Analysis: At A Glance

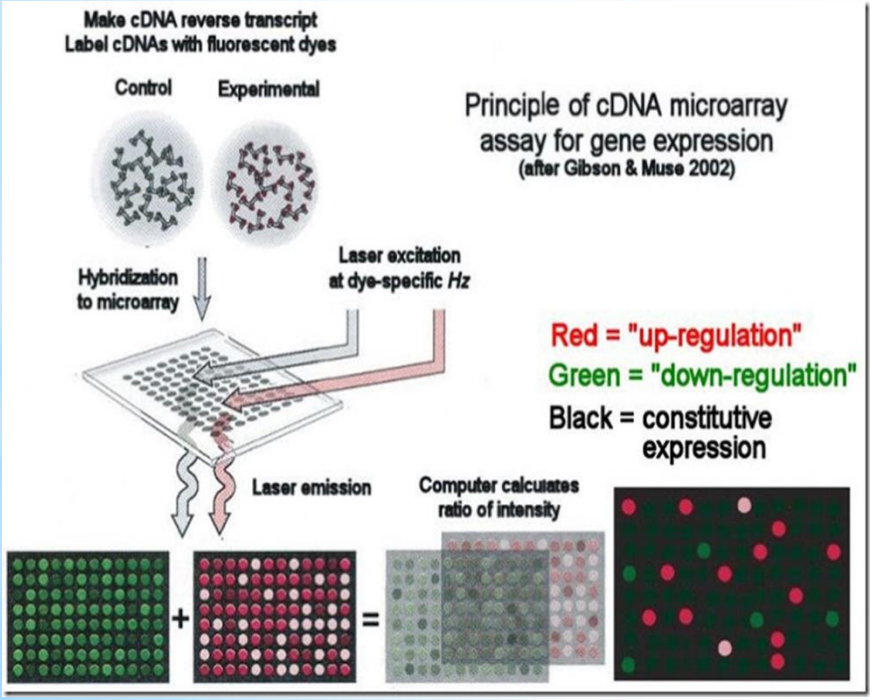


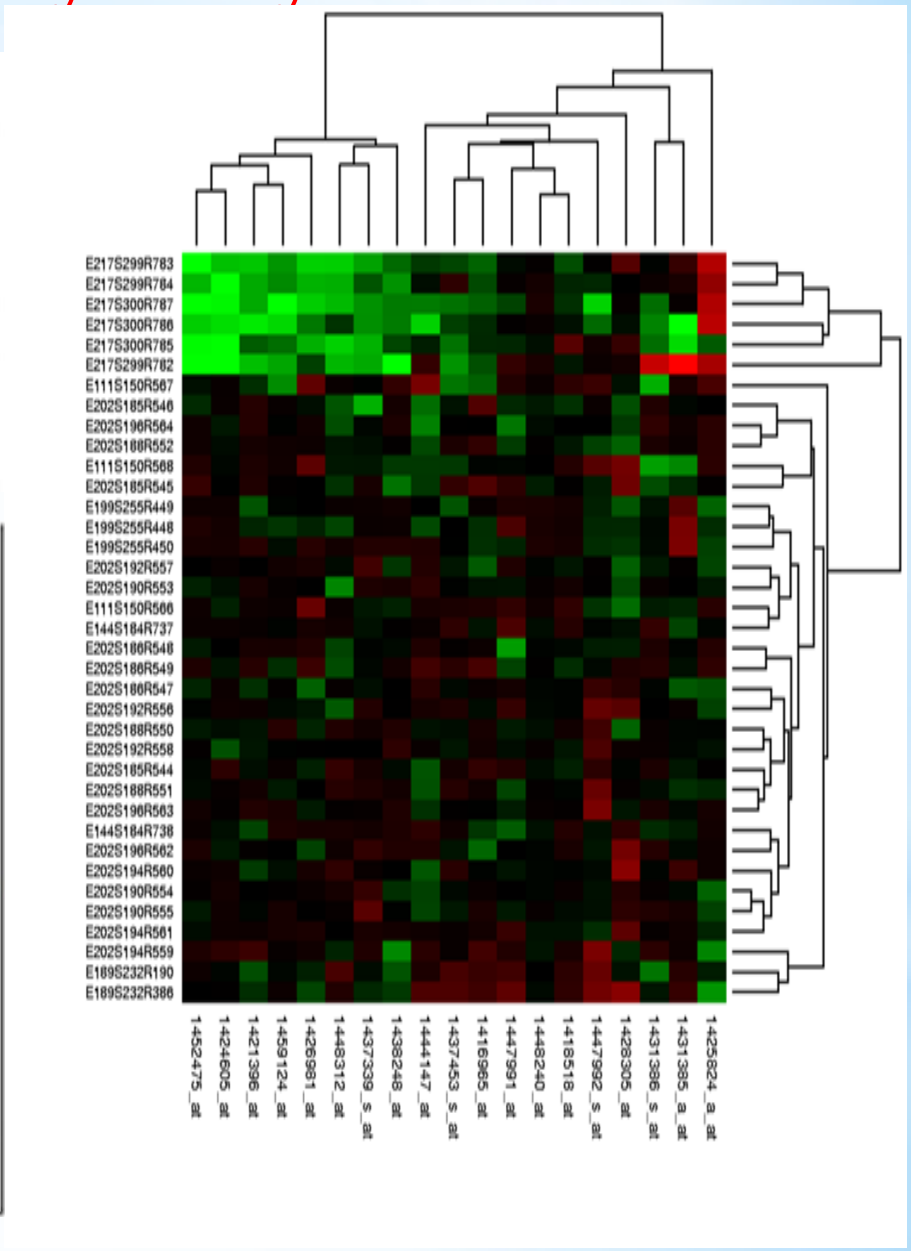




*Steps in DNA Microarray Analysis: At A Glance



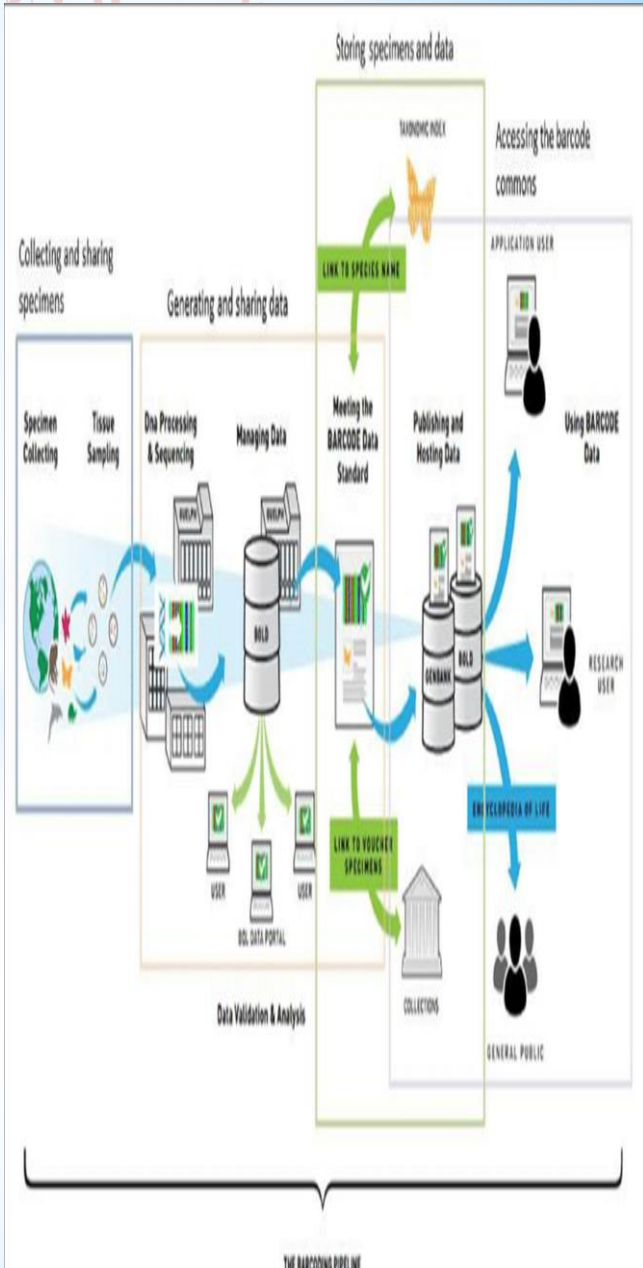




Specific DNA markers in Molecular taxonomy : DNA Barcoding

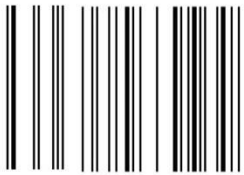
DNA barcoding is a method of identifying organisms based on a short, standardized fragment of genomic DNA and has been developed for use by taxonomists, ecologists, conservation biologists, regulatory agencies, and others.

- DNA barcoding involves the use of a single gene to identify a given species through the comparison of nucleotide sequences in the DNA to that of the same gene in other species. This is in marked contrast to the multiple STR approach used for identifying individuals within a species.
- A bar code (often seen as a single word, barcode) is the small image of lines (bars) and spaces that is affixed to retail store items, identification cards, and postal mail to identify a particular product number, person, or location. ... A barcode reader is used to read the code.
- A region of the mitochondrial gene COI (cytochrome c oxidase subunit I) is used for barcoding animals. COI is involved in the electron transport phase of respiration. Thus, many genes used for barcoding are involved in the key reactions of life: storing energy in carbohydrates and releasing it to form ATP.
- DNA barcodes can aid conservation and research by assisting field workers in identifying species, by helping taxonomists determine species groups needing more detailed analysis, and by facilitating

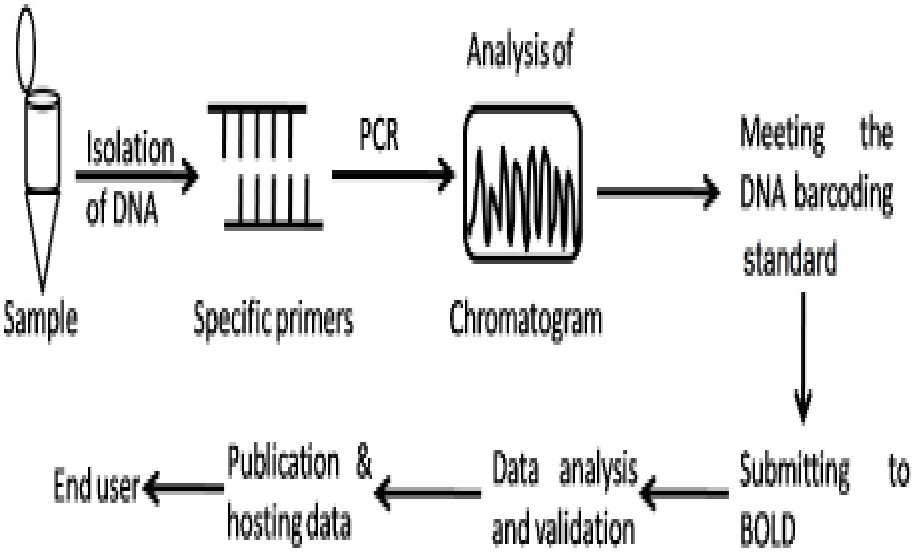


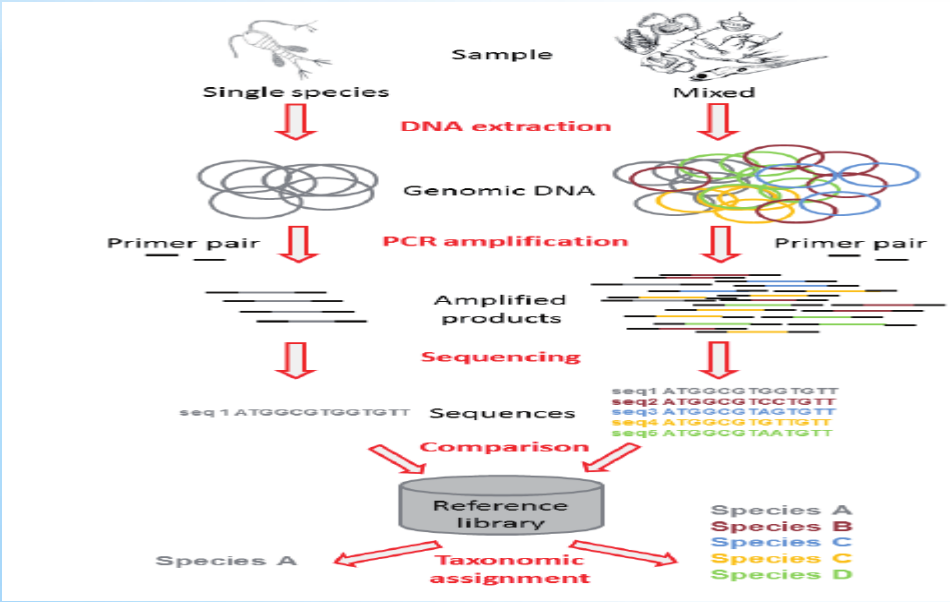
Specific DNA markers in Molecular taxonomy : DNA Barcoding

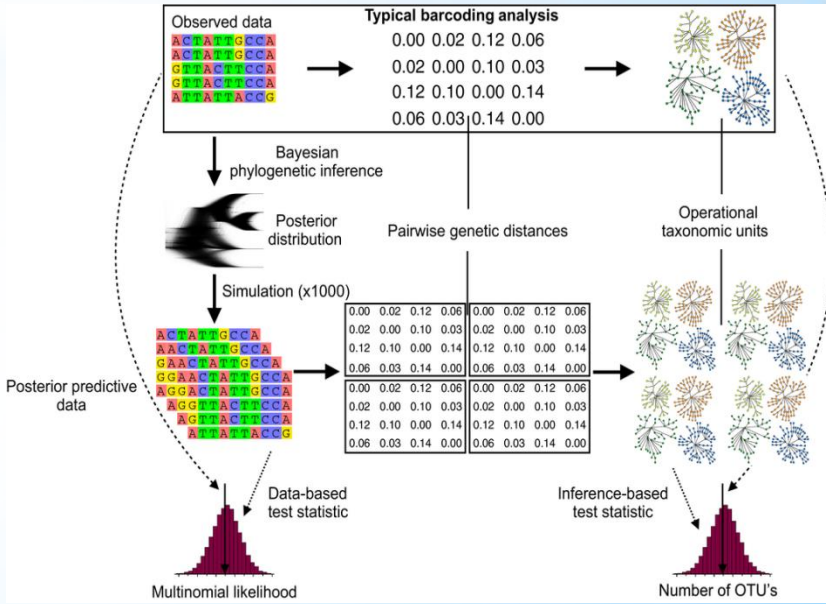
What is DNA Barcode?



- DNA barcode is a short genetic sequence that can be used for biological species identification (Hebert *et al.*, 2003).
- Named after the black stripes used to identify products in the supermarket: Universal Product Code (UPC).
- One of the potential plant DNA barcode is the coding *rbcL* plastid region (CBOL Plant Working Group 2009).







Specific DNA markers in Molecular taxonomy : DNA Barcoding

Advantages

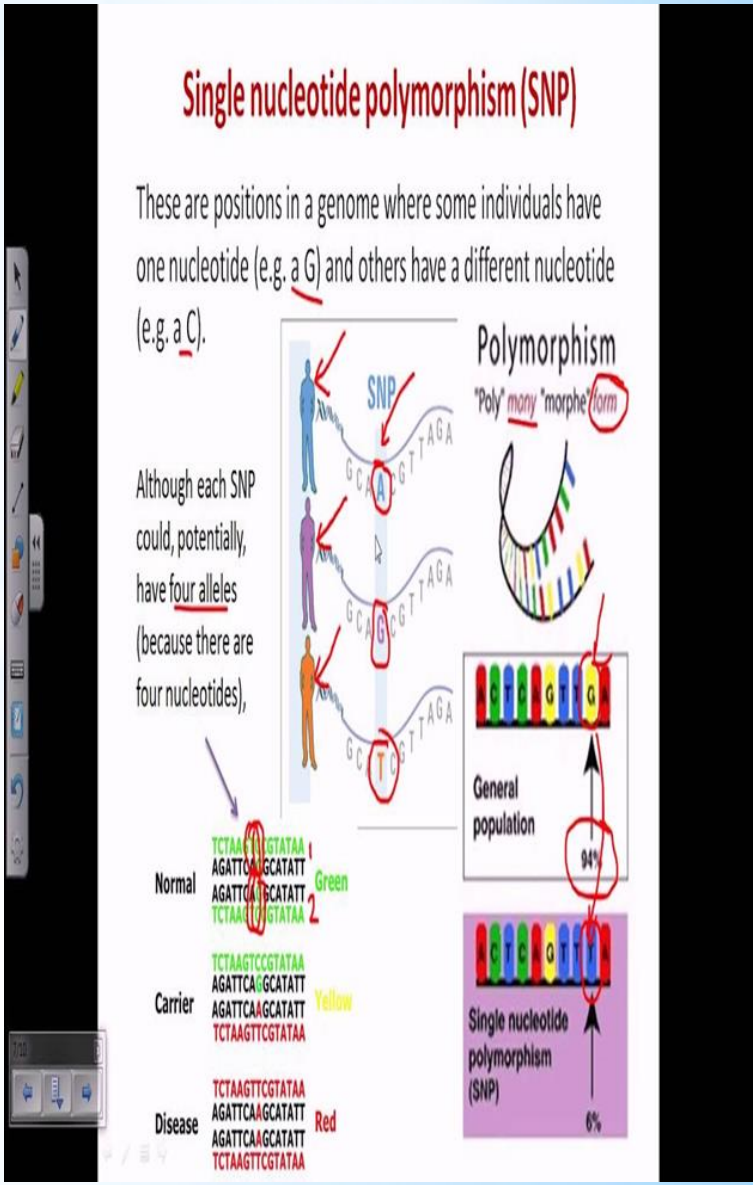
- It is widely used in Arthropod identification.
- Genes primers are available for COI barcode region. COI is widely used for distinguishing closely related and less closely related taxa.
- Alternate markers can be sequenced ,if COI barcode is not differentiating.
- It may be highly useful for taxonomic analysis.

Disadvantages

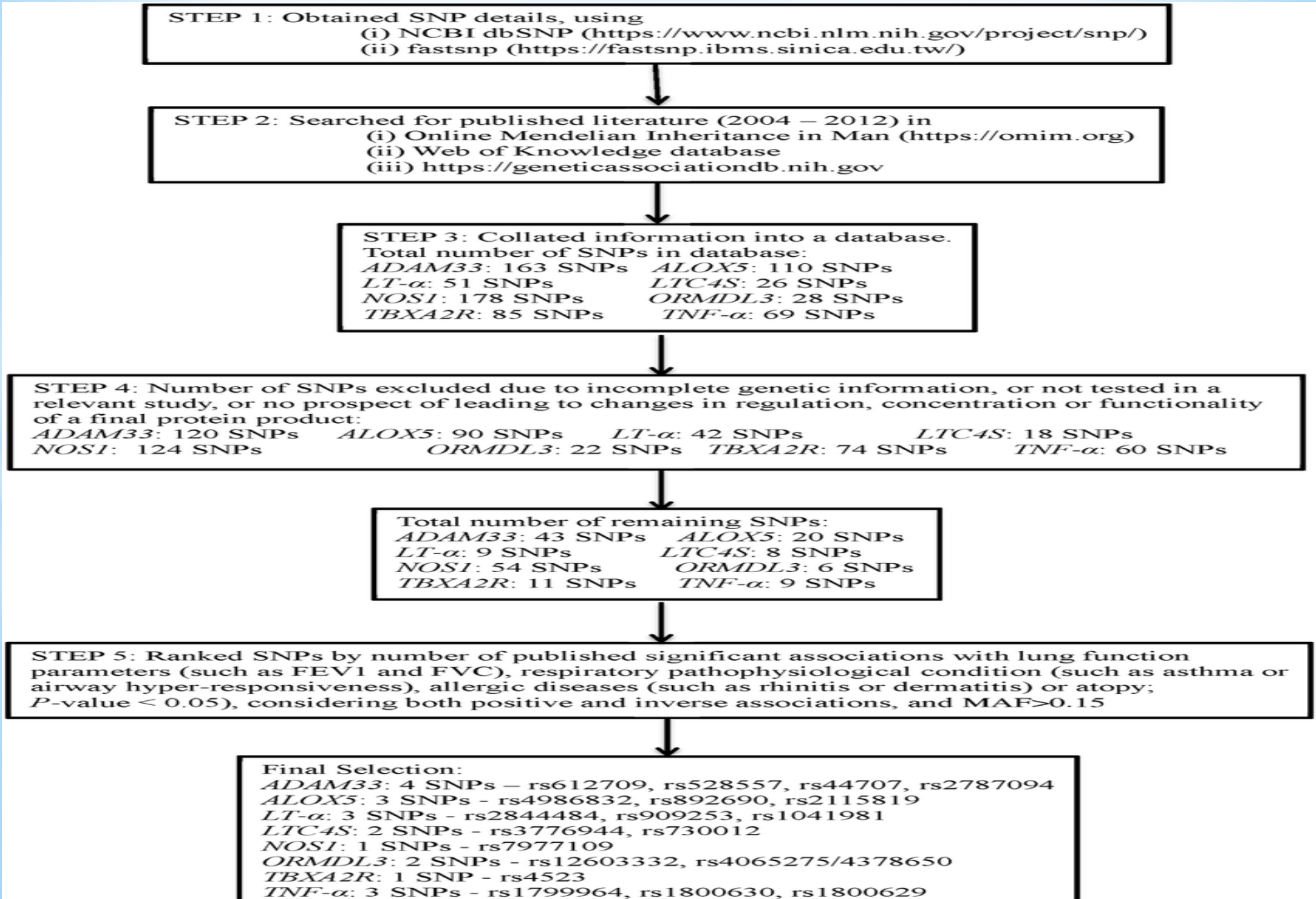
- It requires a large data base of sequences for comparison.
- Prior knowledge of barcoding region is required when applied diagnostically.
- Individual sequences may not provide sufficient discrimination, when studying cryptic species complexes.
- CO I and other mitochondrial genes are maternally inherited which may result in decreased barcode diversity and thereby slow phylogenetic inferences.

Specific DNA markers in Molecular taxonomy : Single Nucleotide Polymorphism (SNP)

- **A single-nucleotide polymorphism** (**SNP**, pronounced **snip**) is a DNA sequence variation occurring when a single nucleotide adenine (A), thymine (T), cytosine (C), or guanine (G]) in the genome (or other shared sequence) differs between members of a species or paired chromosomes in anindividual.
- A single-nucleotide polymorphism is a substitution of a single nucleotide that occurs at a specific position in the genome, where each variation is present at a level of 0.5% from person to person in thepopulation.
- SNP arise due to single nucleotide substitution (transition /trans-version) or single nucleotide (insertion or deletion).
- Here, point mutation gives rise to different alleles with a particular base at a particular mutational position.
- SNPs are the most abundant polymorphism in the genome (coding and noncoding region) of any organism.
- The SNP variables can be detected using PCR, DNA microarray and Fluorescent technology.
- They are considered as next generation markers in fisheries.



Specific DNA markers in Molecular taxonomy : Steps in Single Nucleotide Polymorphism (SNP)



Specific DNA markers in Molecular taxonomy : Expressed Sequence Tags (ESTs)

Expressed sequence tags (ESTs) are relatively short DNA sequences (usually 200–300 nucleotides) generally generated from the 3′ ends of cDNA clones from which PCR primers can be derived and used to detect the presence of the specific coding sequence in genomic DNA.

The identification of ESTs has proceeded rapidly, with approximately 74.2 million ESTs now available in public databases (e.g. GenBank 1 January 2013, all species).
ESTs may be used to identify gene transcripts and analyse their expression analysis, and are instrumental in gene discovery and in gene-sequence determination and linkage mapping.

An EST results from one-shot sequencing of a cloned cDNA. The cDNAs used for EST generation are typically individual clones from a cDNA library. The resulting sequence is a relatively low-quality fragment whose length is limited by current technology to approximately 500 to 800 nucleotides.
Because these clones consist of DNA that is complementary to mRNA, the ESTs represent portions of expressed genes. They may be represented in databases as either cDNA/mRNA sequence or as the reverse complement of the mRNA, the template strand.

One can map ESTs to specific chromosome locations using physical mapping techniques, such as radiation hybrid mapping, Happy mapping, or FISH. Alternatively, if the genome of the organism that originated the EST has been sequenced, one can align the EST sequence to that genome using a computer.

The current understanding of the human set of genes (as of 2006) includes the existence of thousands of genes based solely on EST evidence. In this respect, ESTs have become a tool to refine the predicted transcripts for those genes, which leads to the prediction of their protein products and ultimately of their function. Moreover, the situation in which those ESTs are obtained (tissue, organ, disease state - e.g. cancer) gives information on the conditions in which the corresponding gene is acting. ESTs contain enough information to permit the design of precise probes for DNA microarrays that then can be used to determine gene expression profiles.It is comperatively a fast and reliable method to analyse gene expression in particular tissue type under specific physiological condition differentially expresse3d genes could be identified using

Specific DNA markers in Molecular taxonomy : steps in Expressed Sequence Tags (ESTs)

Isolation of mRNA from the tissue

↓

mRNAs are reverse transcribed to cDNA using oligo-dT primers

↓

Generated cDNA are separated using electrophoresis

↓

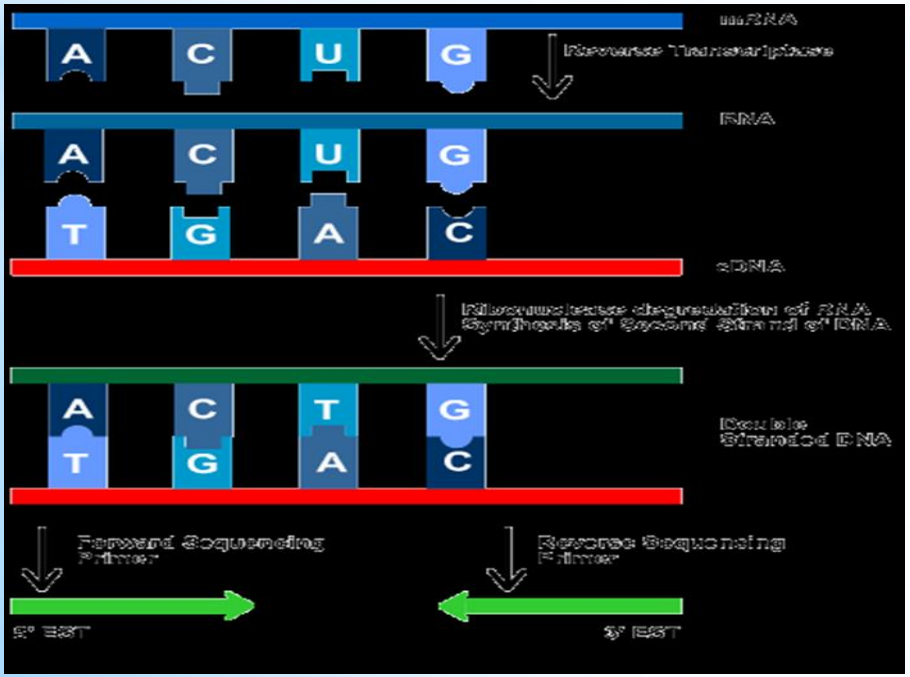
Separated cDNAs will be selected on the basis of size and eluted

↓

They are sequenced randomly from both the ends (5' and 3')

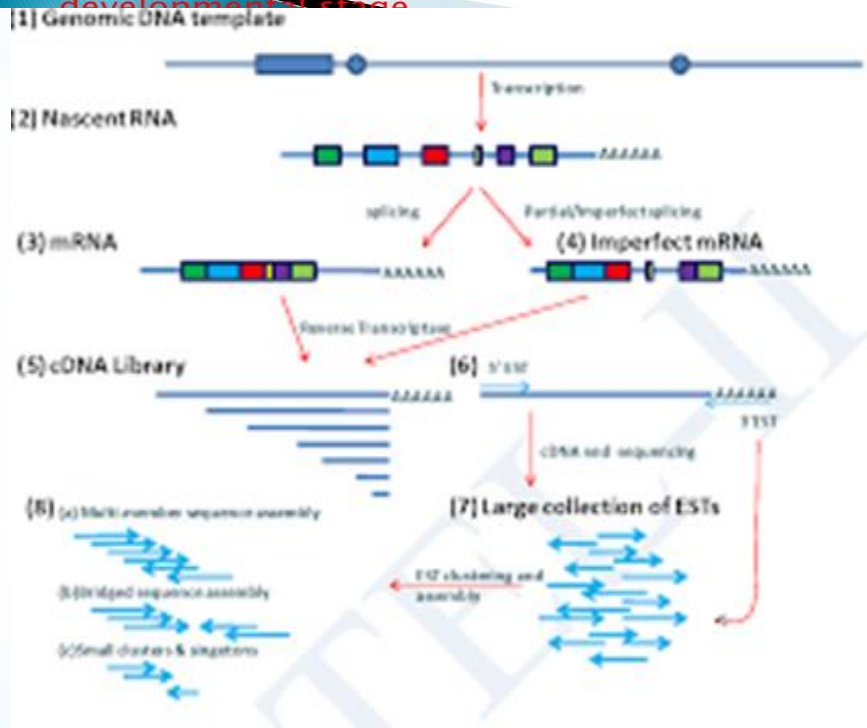
↓

ESTs are generated and their sequences are compared using different databases



Application of EST

- Identify unknown genes and to map their positions within a genome.
- ESTs provide researchers with a quick and inexpensive route for discovering new genes,
- For obtaining data on gene expression and regulation,
- constructing genome maps.
- Economical approach to identify and characterize expressed genes
- EST represent a snapshot of genes expressed in a given tissue and/or at a given



Conclusions

1. molecular-taxonomy is the most emerging branch of modern taxonomy. Molecular markers can be used as a supplementary marker system which will increase resolution in taxonomic research.

The molecular evolution among taxa is highly variable and the extent of divergence in DNA or gene can be taken as the basis for differentiation among species. Molecular markers are very useful when morphological identification is very difficult. Endangered and threatened whales, sharks and dolphins, when are dead and stranded can be identified using molecular markers.

Within species, genetic differences are more than between population so using molecular markers identification is possible even when the sample size is small(3-5). Molecular markers are used in sub species identification. Some of the high evolving loci will show more divergence within species, so these loci can be used for finding out intraspecific competition.

Molecular markers are also used in genetic tagging or marking, forensic investigations, (deliberate or accidental release in natural water), studying the trophic relationship (extracting DNA from naturally digested food) and analysis of ancient DNA.

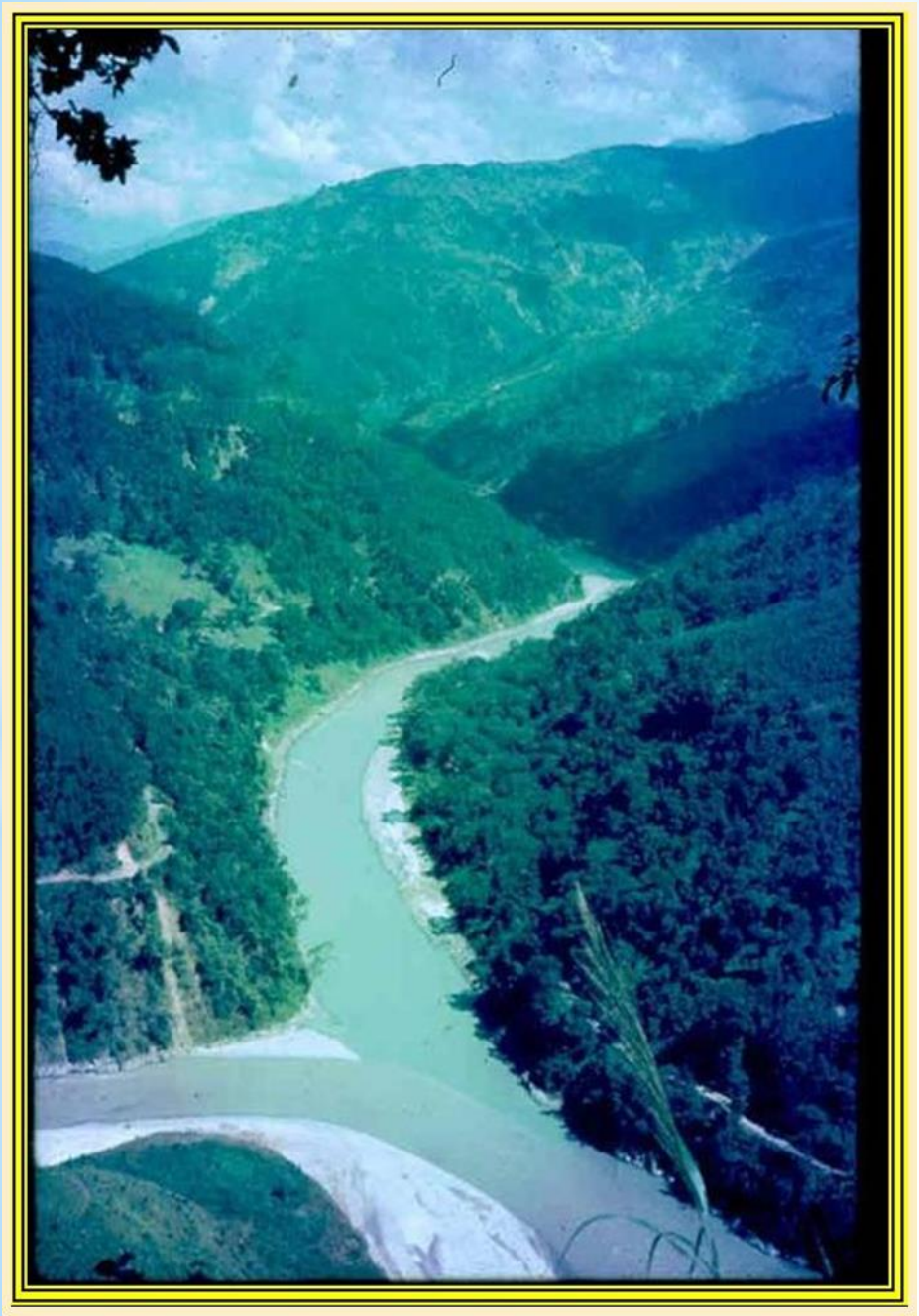
Conclusions (Continued)

Limitations:

1. It is an essentially a cladistic approach .It assumes that classification must correspond to phylogenetic, descent and that all valid taxa must be mono-phylectic.

2. The recent development of extensive horizontal gene transfer among organism provides a significant complications to molecular systematics, indicating that different genes within the same organism can have different phylogenies.

3. Molecular phylogeny are sensitive to assumption and model that go into making them. They face problem like long branch attraction, saturation and taxon sampling problem. This means that strikingly different results can be obtained applying different models to the same data.



THANK YOU