

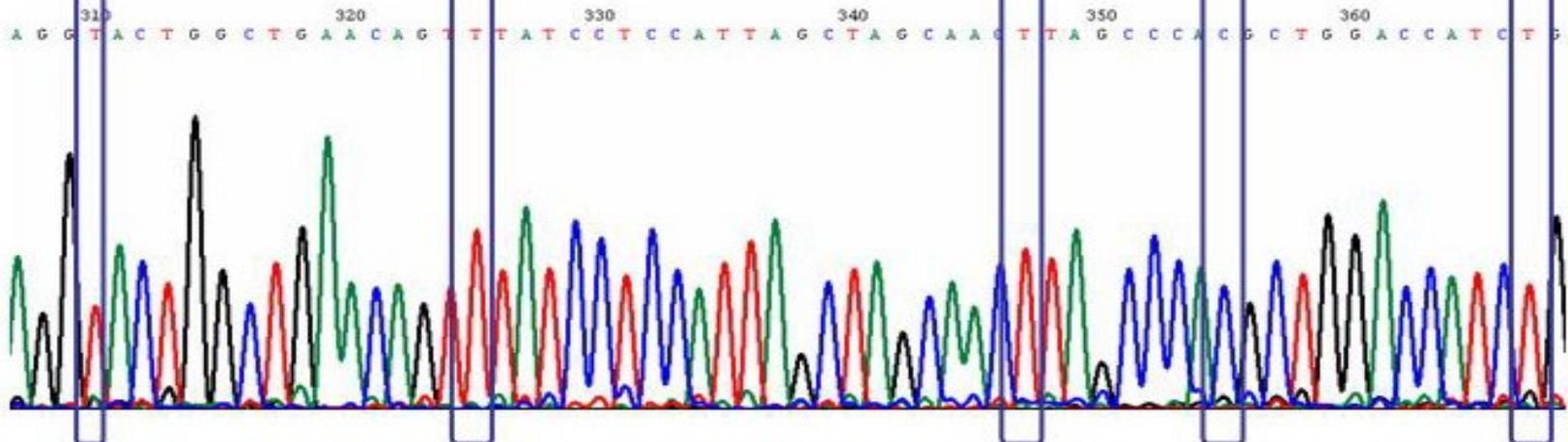
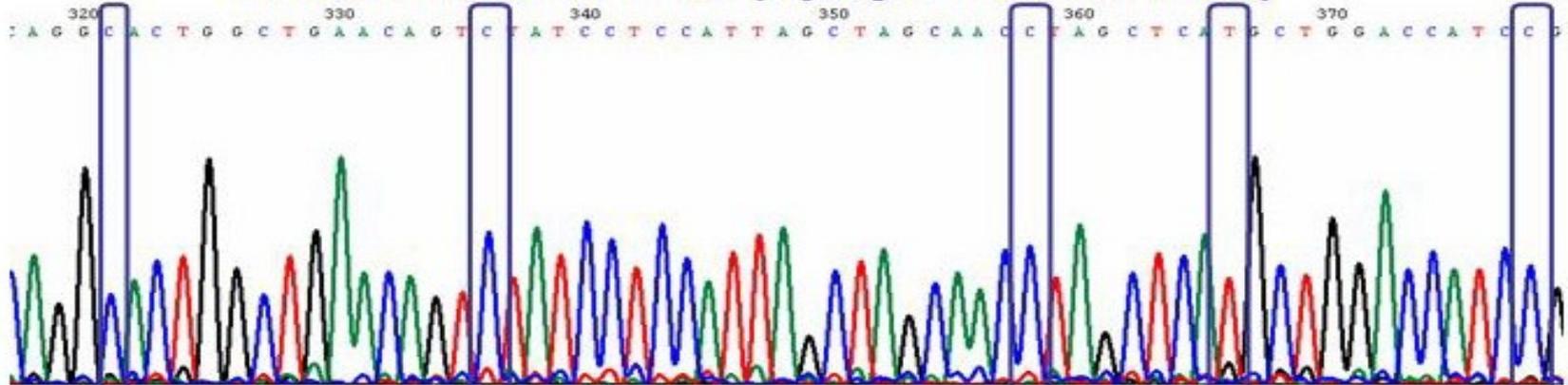
# GENETIC VARIATION & DNA MARKERS

- All organisms are subject to mutations as a result of normal cellular operations or interactions with the environment, leading to *genetic variation (polymorphism)*
- In conjunction with selection there arises genetic variation *within and among individuals, species, and higher order taxonomic groups.*
- To the geneticists, the variation must be *(1) heritable and (2) discernable to the researcher*, either as a recognizable phenotypic variation or as a genetic mutation distinguishable through molecular techniques.

- **Types of genetic variation:** Base substitutions – known as single nucleotide polymorphisms (SNPs); insertions or deletions of nucleotide sequences (indels) within a locus; inversion of a segment of DNA within a locus; and rearrangement of DNA segments around a locus of interest.
- Through long evolutionary accumulation, - **the number and degree of the various types of mutations define the genetic variation within a species.**
- DNA marker technology can be applied to reveal these mutations.
- **Molecular / genetic marker:** A DNA fragment, an allele or cytogenetic marker or gene product (allozymes/proteins) used as an experimental probe to keep track of a species, a population, an individual, a tissue, a cell, a nucleus, a chromosome, or a gene. **Molecular markers are used to identify different genetic varieties of microbes, animals & plants, or to trace genetic predisposition to certain cancers.**

# Single Nucleotide Polymorphisms (SNPs)

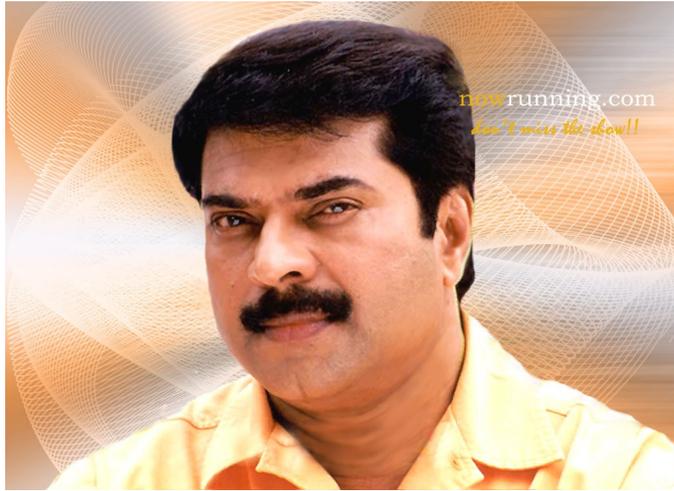
## Great hammerhead (*Sphyrna mokarran*)



## Scalloped hammerhead (*Sphyrna lewini*)

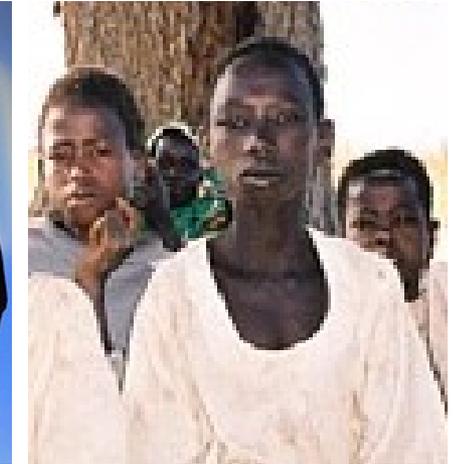
# Genetic Variation:

## 1) at individual level



# Genetic Variation:

## 2) at intra-specific level



*Canis familiaris*



# Genetic Variation:

## 2) at intra-specific level



# Genetic Variation:

## 3) at inter-specific/ inter-generic levels

*(Homo sapiens & Pan troglodytes)*



## Genetic Variation:

### 3) at inter-specific/ inter-generic levels



## **Types of markers:**

- **DNA markers (nuclear & mitochondrial)**
- **Protein Markers (allozymes; general proteins; salt soluble Myofibrillar proteins such as myosin and tropomyosin)**
- **Type I and Type II: Type I are markers associated with genes of known function, (eg., Allozymes, RFLP, ESTs); Type II markers are associated with anonymous genomic segments (RAPD, AFLP, microsatellites).**

# Properties of Molecular Markers

- 1) **Heritable in Mendelian fashion**
- 2) **Genetic variations - discernable**
- 3) **Neutral**
- 4) **Reproducible**
- 5) **Monomorphic (fixed) / Polymorphic**

**Very important to choose appropriate marker for a given objective.**

# Requirements for molecular markers studies

- **Tissue samples - accurately identified; properly preserved in ethanol**
- **Good quality DNA**
- **PCR Lab facilities**
- **Prior information of the genome (except in the case of RAPD) - 'primers'**
- **Electrophoresis apparatus & staining protocols**
- **Sequencing facility**
- **Data analysis**

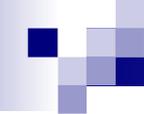


**1. RANDOM AMPLIFIED POLYMORPHIC  
DNA (RAPD) (Williams *et al.*,1990) or  
Arbitrarily Primed PCR (AP - PCR) (Welsh  
& McClelland, 1990)**

**2. MICROSATELLITES**

**3. Mitochondrial DNA & DNA Barcoding**

**4. ITS (Internal Transcribed Spacers)**



**RANDOM AMPLIFIED  
POLYMORPHIC DNA (RAPD)**

*(Williams et al.,1990)*

**or**

**Arbitrarily Primed PCR (AP - PCR)**

*(Welsh & McClelland, 1990)*

# RAPD

- **Single short Primer of Random Arbitrary Sequence (10 bp)**
- **No sequence data is required; hence can be initiated with minimal infrastructural facilities.**
- **Primer will bind to many sites at low PCR stringency conditions**
  - ❑ **Low annealing temp (36 – 40°C)**
  - ❑ **High MgCl<sub>2</sub> concentration**
- **Dominant Marker**

## Advantages:

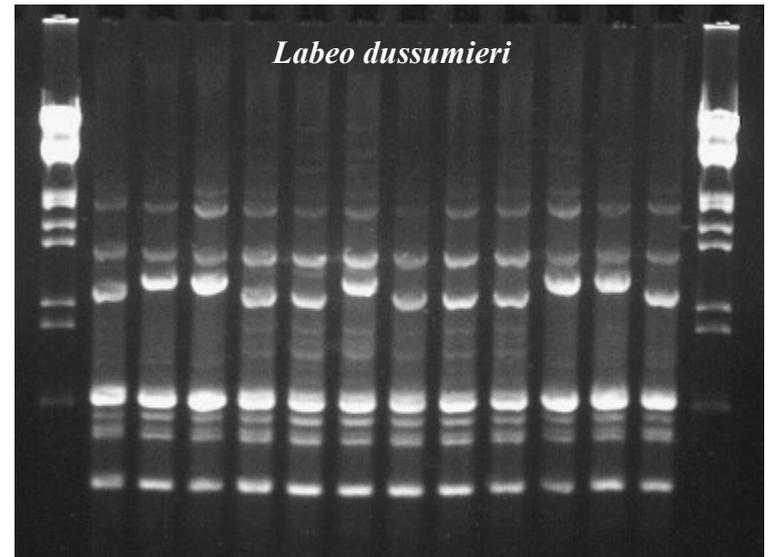
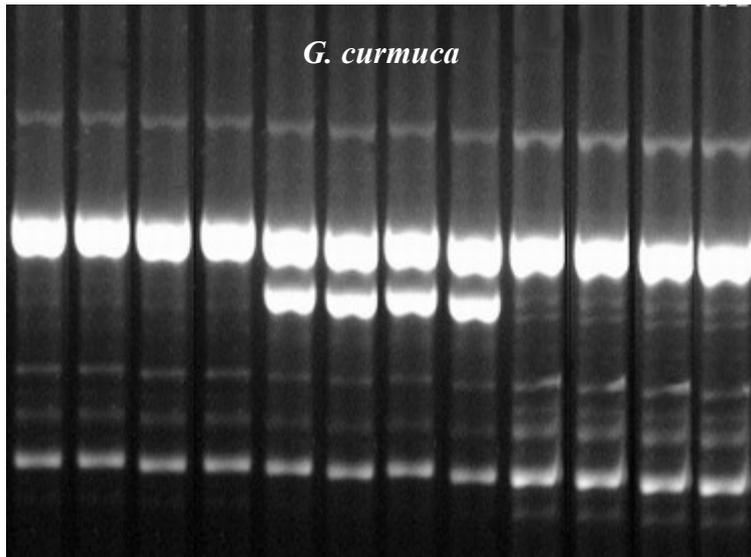
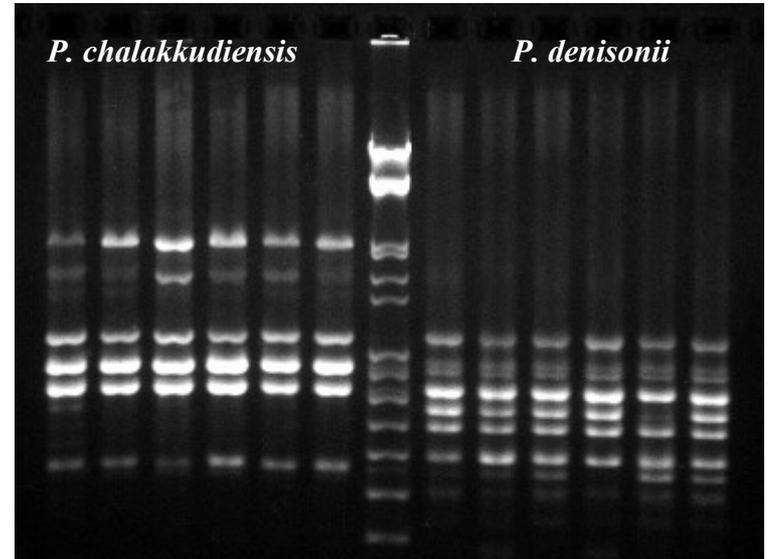
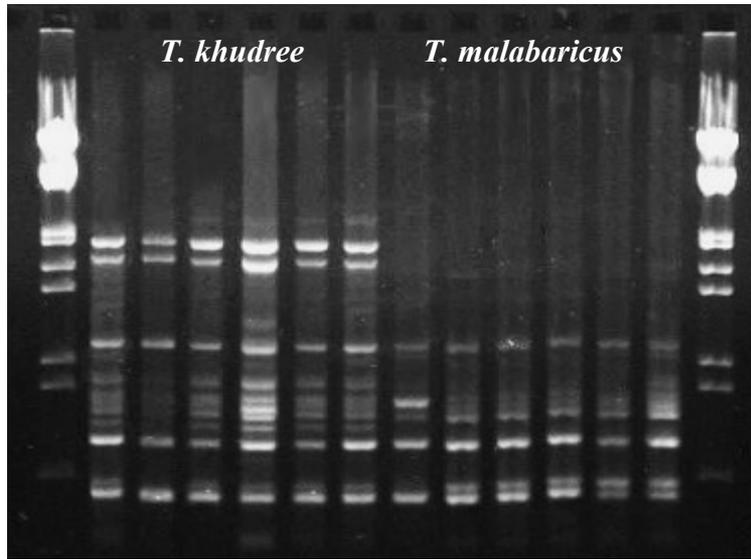
- **Sequence data is not required.**
- **RAPD studies in a new species can be initiated with the minimal infrastructure facilities.**

## Disadvantages:

- **Short primer (of generally 10mer) with random sequence will bind to many sites within the template DNA at low PCR stringency conditions. Due to this property, a slight change in the parameters of PCR would result in change in RAPD-PCR product patterns. Hence, for obtaining the reproducibility of the results, strict standardization of the PCR are required.**

# Points to be taken care of

- 🔑 With every PCR, set up a PCR reaction without genomic DNA (called negative PCR reaction) to monitor the DNA contamination during PCR set up.
- 🔑 **RAPD primer to be used is first tested for amplification of clear and repeated bands.**
- 🔑 **Utmost care has to be taken during setting up of RAPD-PCR, to avoid contamination of PCR products**
- 🔑 **For repeatability of results, to strictly follow:**
  - # Proper lab techniques.
  - # Strict standardisation of all reagents and protocols



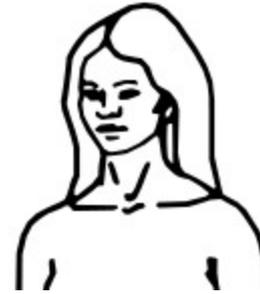
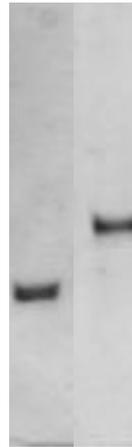
# SCAR

- ♣ **RAPD-PCR technique can also generate species-specific, sex-specific and population specific fragments. These fragments are useful in developing specific “Sequence Characterized Amplified Region (SCAR) Markers”**
- ♣ **Cut out species specific band from the gel → purify → sequence → synthesize specific primer → amplify → only a single species specific band in the target species.**



# **MICROSATELLITES**

- **Powerful DNA markers for quantifying genetic variations within & between populations of a species**
- **Also called as STR, SSR, VNTR**
- **Tandemly repeated DNA sequences with the repeat/size of 1 – 6 bases repeated several times**
- **Highly polymorphic; can be analysed with the help of PCR**
- **Individual alleles at a locus differ in number of tandem repeats of unit sequence owing to gain of loss of one or more repeats and they can be differentiated by electrophoresis according to their size**



Co-dominant marker



# Microsatellites – Types

## Based on repeat pattern

1. Perfect – CACACACACACACACACACA

2. Imperfect – CACACACACA CACACACA  
CACACACA

3. Compound –  
CACACACACACA CATACTACATA CATACTACATA

4. Complex – CACACACACACACA  
AATAATAATAATAATAAT

## 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)**

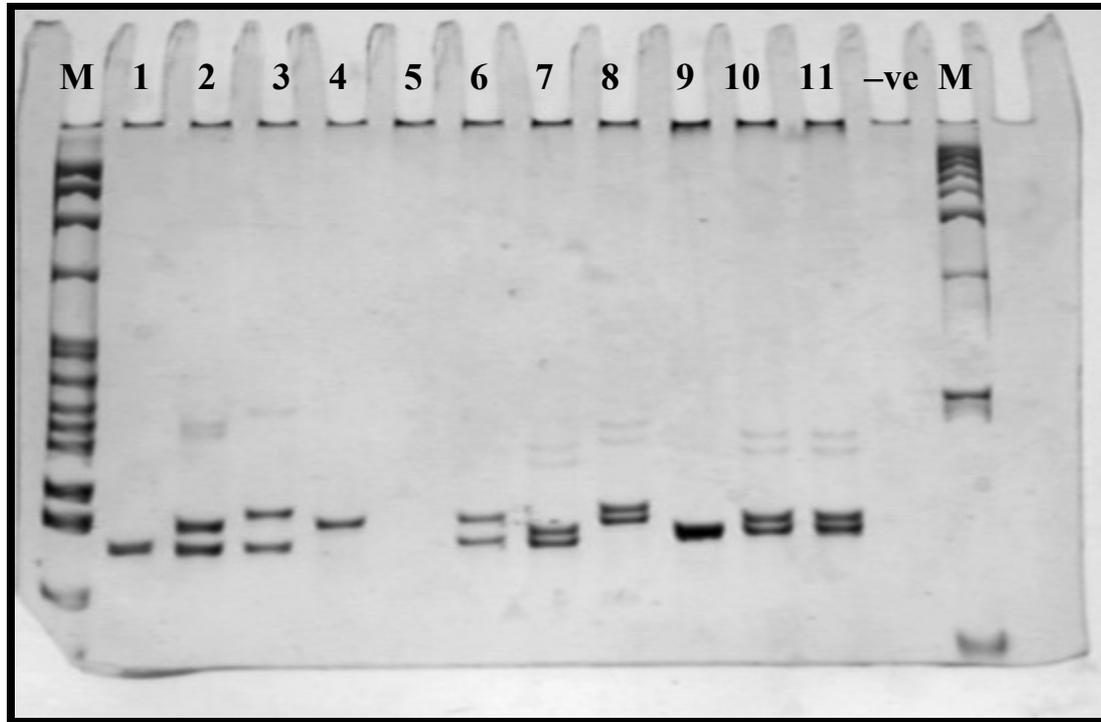
———— CGCCATTGTAGCCAATCCGGGTGCGATTGCAT CGCCATTGT  
AGCCAATCCGGGTGCGATTGCAT CGCCATTGTAGCCAATCCGGG  
TGCATTGCAT CGCCATTGTAGCCAATCCGGGTGCGATTGCAT  
CGCCATTGTAGCCAATCCGGGTGCGATTGCAT ———

## **Microsatellites – Properties**

- ✓ **Co-dominant**
- ✓ **Inherit in Mendelian Fashion**
- ✓ **Polymorphic loci with allele number as high as 14 – 15 per locus**
- ✓ **Mostly reported from non-coding region, hence can be independent of selection**
- ✓ **Flanking region is highly conserved in related species**
- ✓ **Can be obtained from small amounts of tissues**
- ✓ **PAGE separation; silver staining/automated sequencing**
- ✓ **Abundant in the genome**

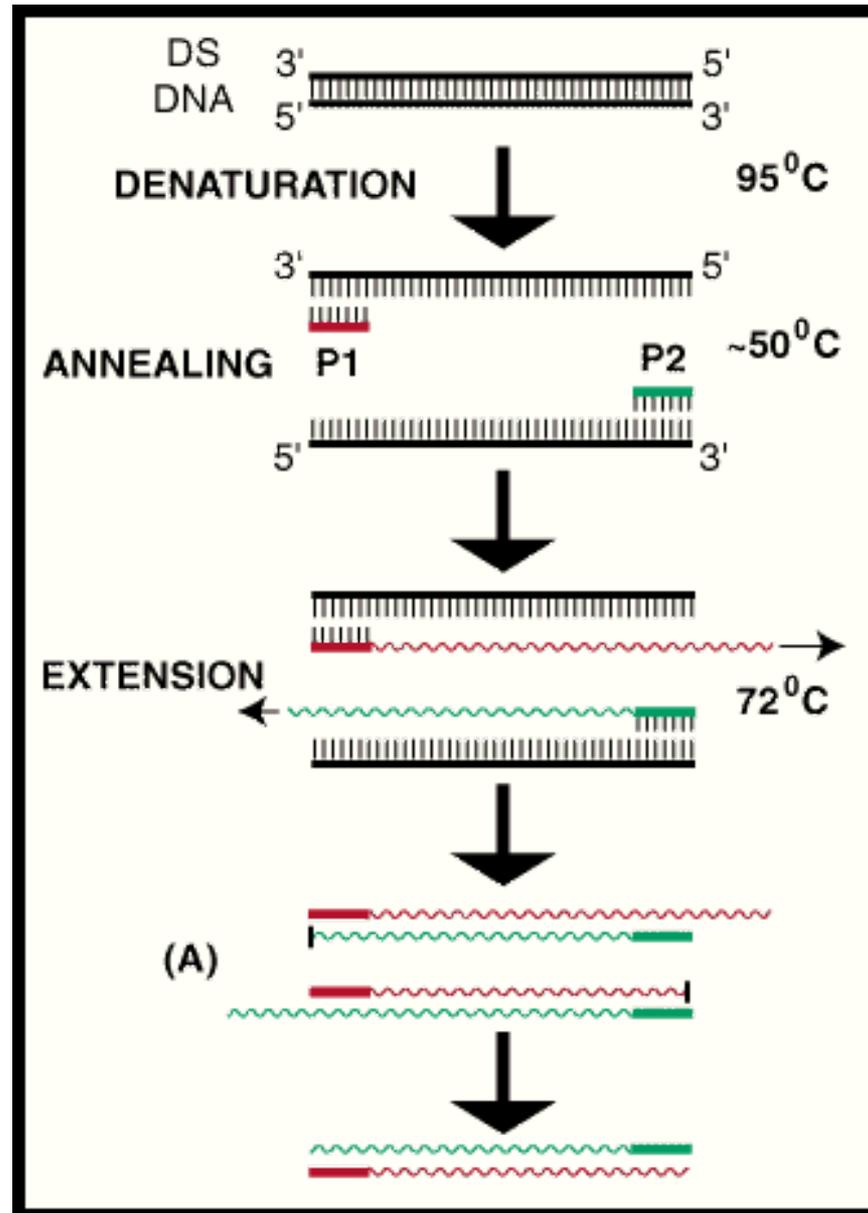
## **Problems:**

- ♣ **Prior information of the genome is essential as the primers are highly specific**
- ♣ **Stutter bands**
- ♣ **High sample size**
- ♣ **Mutation in primer binding sites and the region between primers and repeats can create problems**



**Fig. Single Locus Microsatellites. (There are five alleles at this locus.)M: Standard molecular weight marker, 1 -11:Different individuals -ve : Negative control**

# Microsatellite PCR





# Microsatellite multiplexing

4275

3895

3520

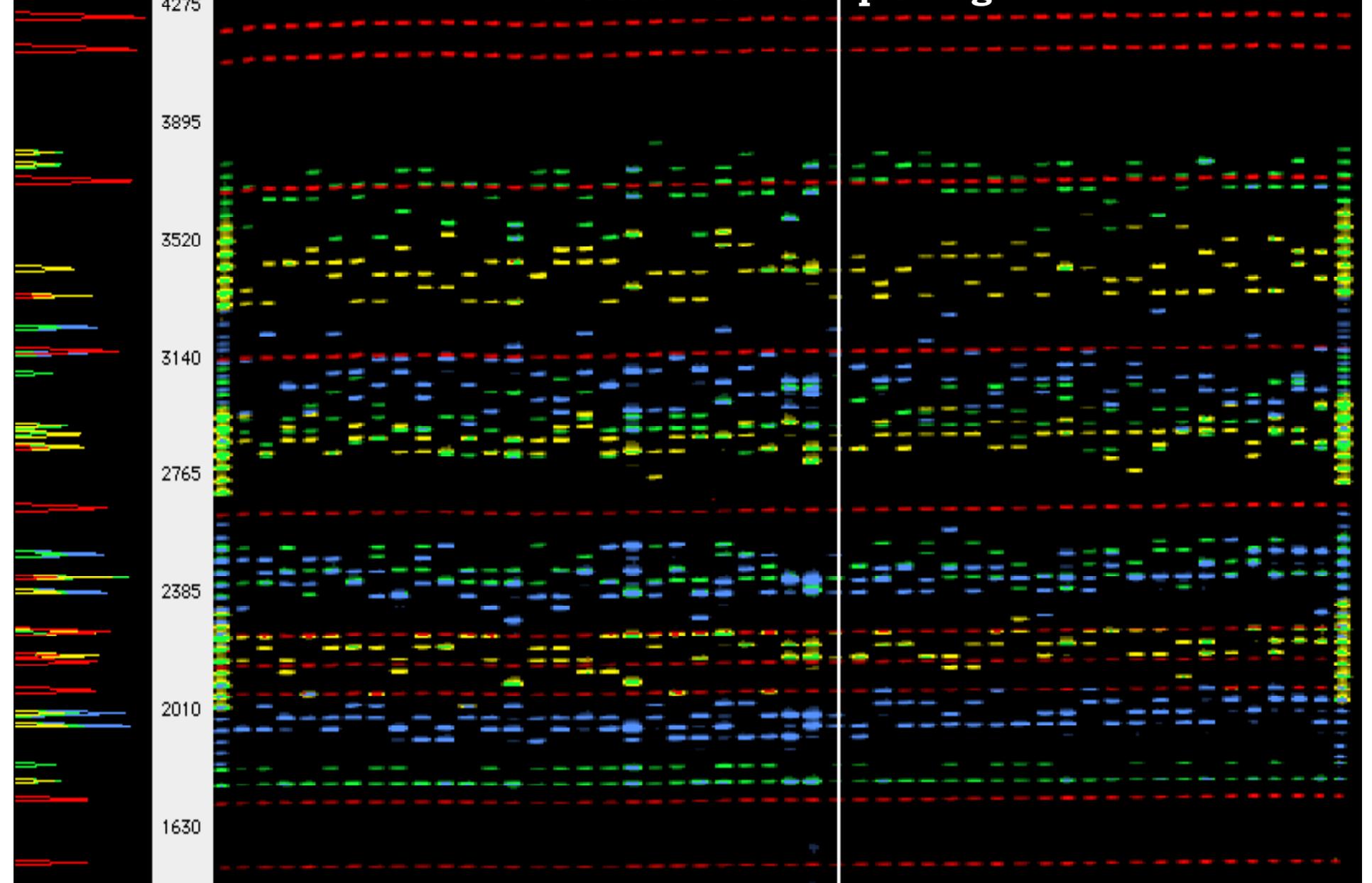
3140

2765

2385

2010

1630



# Genotyping

Individuals with single bands constitute homozygotes while heterozygotes have double bands. On this basis, each individual is genotyped as homozygote/heterozygote with allele designation. The alleles are designated on the basis of their molecular weights.

**Example:** An individual showing (i) one band of 100 bp is genotyped as 100100, and (ii) two bands of molecular weight 100 and 120 bp is genotyped as 100120.

# Genetic differentiation/Heterogeneity

- o **Co-efficient of genetic differentiation ( $F_{ST}$ )**

( $F_{ST}$  is the index of genetic differentiation that describes how much variation in allele frequencies is present between the local populations. It is a measure of population differentiation and ranges from 0 where all the population have the same allele frequencies at all the loci to 1.0, where all the populations are fixed for different alleles at all loci)

- o **Genetic Distance & Similarity Index**

- o **Dendrogram (UPGMA)**

- o **Softwares used: GENEPOP, GENETIX, POPGENE, WINBOOT**

- o **Interpretation of Results**



# **DNA Barcoding**

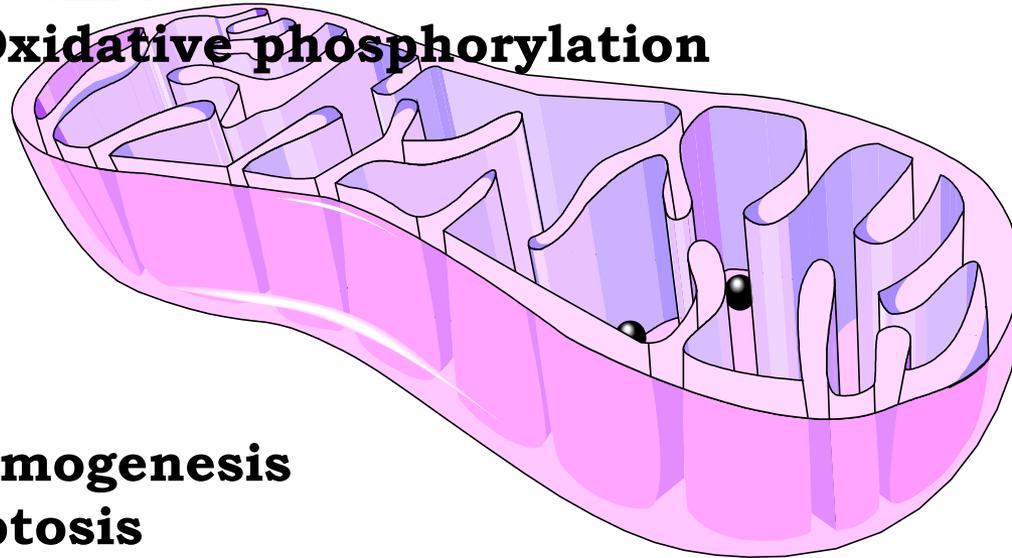
# Mitochondria

- **Production of energy [ATP] for the cell:**

- **Krebs cycle**

- ∇  **$\beta$ -oxidation**

- **Oxidative phosphorylation**

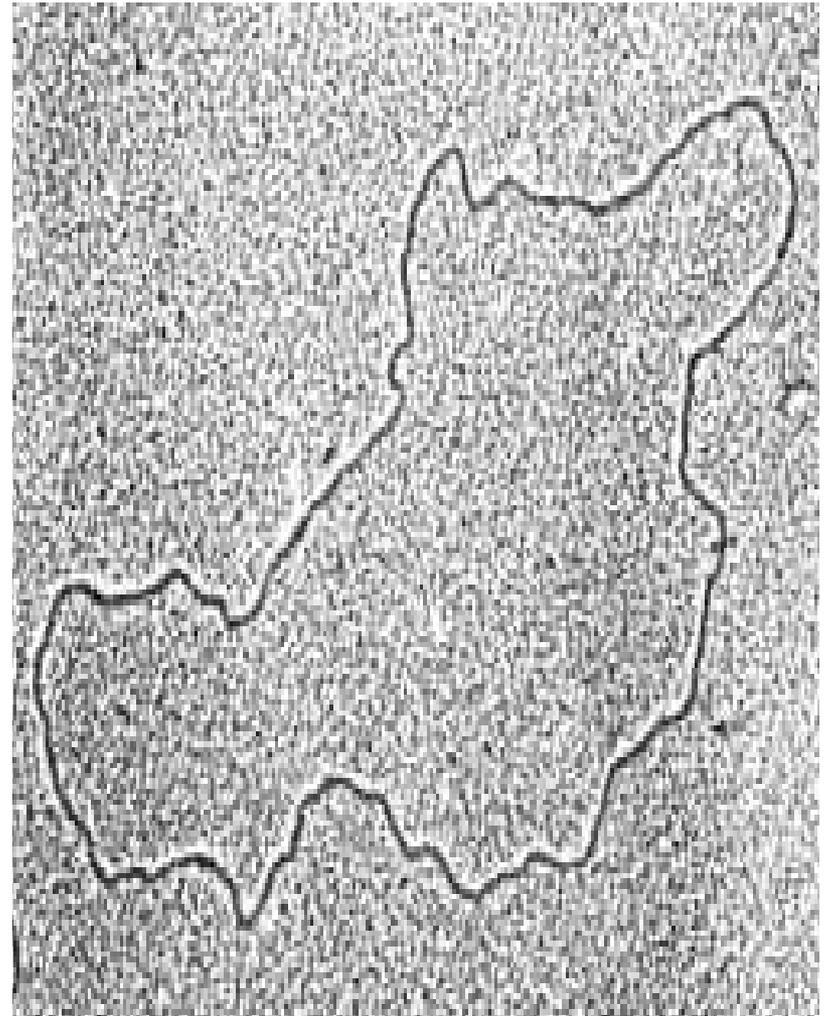


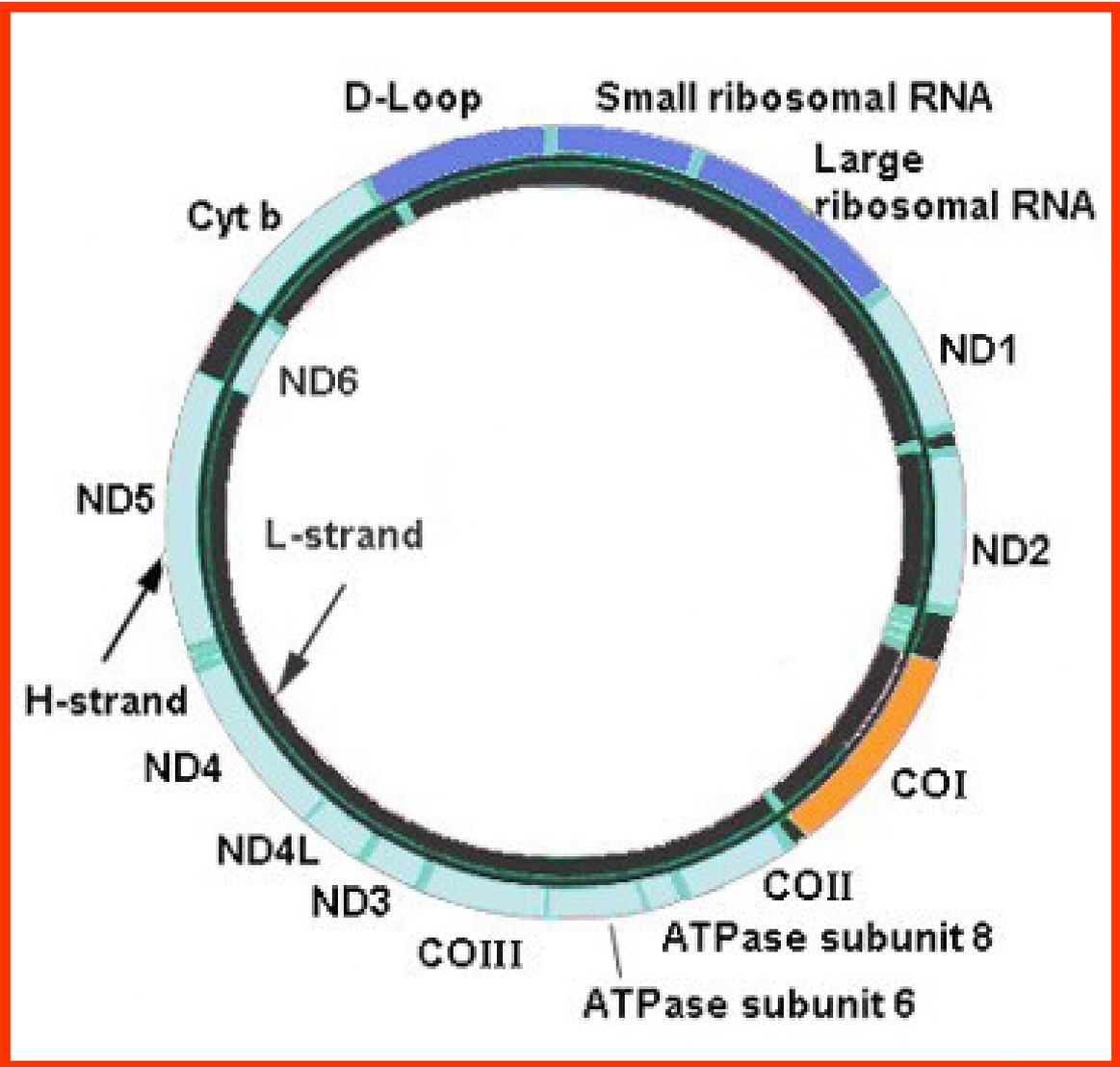
- **Thermogenesis**

- **Apoptosis**

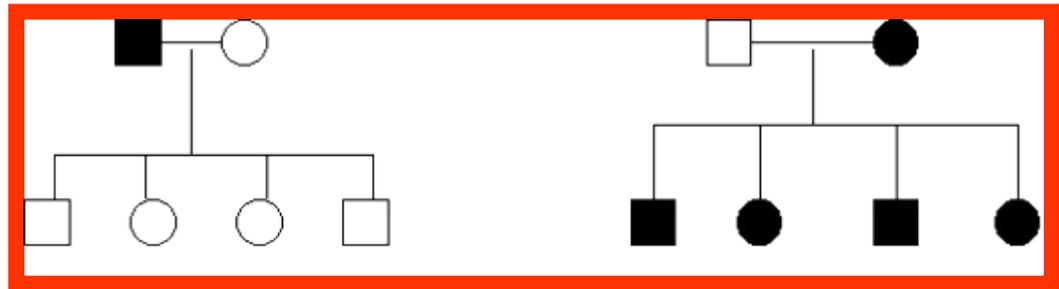
# Mitochondrial DNA

- ♫ **Mitochondria contain their own DNA (mtDNA).**
- ♫ **Not a linear chromosome; Circular haploid structure about 16.5 kb in size.**
- ♫ **Humans: 16,000 bases.**
- ♫ **Similar to bacterial DNA.**
- ♫ **37 Genes: rRNA (2), tRNA (22), 13 proteins in the electron transport chain that produces ATP; no introns; Inter-generic sequences (10 bp long) are small or absent; Repetitive DNA is rarely present.**
- ♫ **Both nDNA and mtDNA are required for normal function of the mitochondria.**

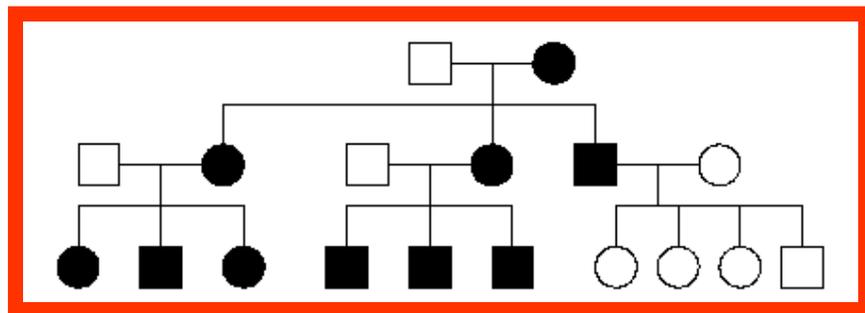




# Mitochondrial inheritance



**Maternal inheritance**



## **Haplotype:**

**A combination of alleles of different genes which are located closely together on the same chromosome and which tend to be inherited together**

## **Heteroplasmy:**

**Presence of more than one type of mtDNA in an individual**

## **Paternal leakage**

# Properties of mtDNA

- ✦ **Exclusive maternal inheritance except in marine mollusc *Mytilus spp.***
- ✦ **No recombination of mtDNA genome. In plants – it appears to be mosaic *i.e.* a combination of chloroplast, mitochondrial & nuclear genes**
- ✦ **Evolution is 5-10 times faster than single copy nDNA due to :**
  - (i) Error prone system of DNA replication;**
  - (ii) Absence or deficiency in DNA repair function;**
  - (iii) Incapability of removing mildly deleterious mutation in the absence of recombination.**

- **Mutation rate differs in different regions:**
  - D loop/control region – fastest evolution;**  
 ATPsynthase – 6, ND3, ND4, etc., are fast evolving;  
 COI, 16S rRNA, 12S rRNA, Cyt b are slow evolving.
- **Fast evolving genes are useful in studying intra-specific genetic diversity & slow evolving (highly conserved) ones are useful in studying evolutionary relationships.**
- **Length heteroplasmy:- due to insertion/deletion of tandem repeats in D-loop region.**
- **Site polymorphism – due to more common transition (A↔G, C↔T) & rare transversions (A↔C; A↔T; G↔C; G↔T)**

- ✓ **Two strands – Light (L) & Heavy (H) chains due to more purine bases (A&G) in heavy chains and the behaviour in CsCl gradients.**
- ✓ **All genes are coded by H strand except ND6 and 8 tRNAs by L strand**
- ✓ **Genetic Code: UGA codon – stop codon in nuclear genes, but codes for tryptophan in mitochondria. Mt code requires 22 tRNAs rather than 24 tRNAs of the nucleus.**

**All these factors helps to reduce the effective population size for mtDNA to  $\frac{1}{4}$  to that nuclear genes of the same organism. This means that genetic drift can cause allele frequency differences more rapidly in mtDNA than nDNA; less sample size required.**

### **Analysis**

- **Earlier days – RFLP**
- **Now sequence information**
- **Using specific primers, mtDNA genes can be amplified & sequenced.**



## **Problems of mtDNA Analysis**

- **Crossbreds, Hybrids & recently diverged species cannot be distinguished**
- **Mitochondrial pseudogenes in nDNA**

# DNA Barcoding

- **DNA sequence analysis of a uniform target gene like the mitochondrial cytochrome oxidase subunit I (COI) to enable species identification has been referred to as “DNA Barcoding”, by analogy with the Universal Product Code (UPC) system barcodes used to identify manufactured goods.**
- **DNA barcoding has the potential to be a practical method for identification of the estimated 10 million species of eukaryotic life on earth.**
- **Expected to work at all stages of life, resolve taxonomic ambiguities, unmask look-alikes and opens the way for an electronic handheld field guide, ‘the life barcoder’.**

- **An appropriate target gene like COI is conserved enough to be amplified with broad range of primers and divergent enough to allow discrimination of closely allied species and it shows comparatively less sequence divergence within a species compared to the fast evolving genes of mtDNA.**
- **Alternate target genes are needed for cnidarians (sea anemones, corals and some jelly fishes) and plants as these groups show too little mitochondrial sequence diversity.**
- **mtDNA of fungi contains introns that can complicate DNA amplification.**
- **Global effort, recently initiated to compile DNA barcodes of all known eukaryotes along with collection of voucher specimens with authoritative taxonomic identification.**

# Biological specimens come in many forms



Identifiable  
adults

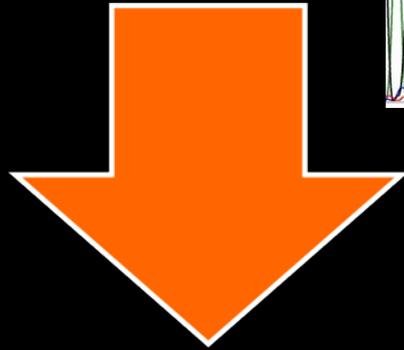
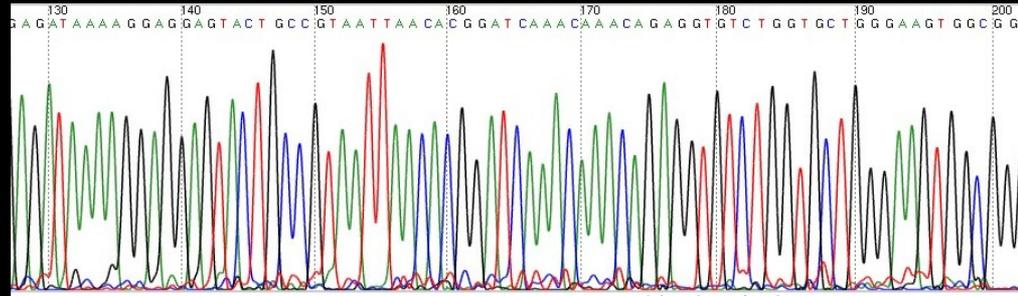


Juvenile stages



Processed products

# Barcoding: Integrating Best Practices



**Genomics**



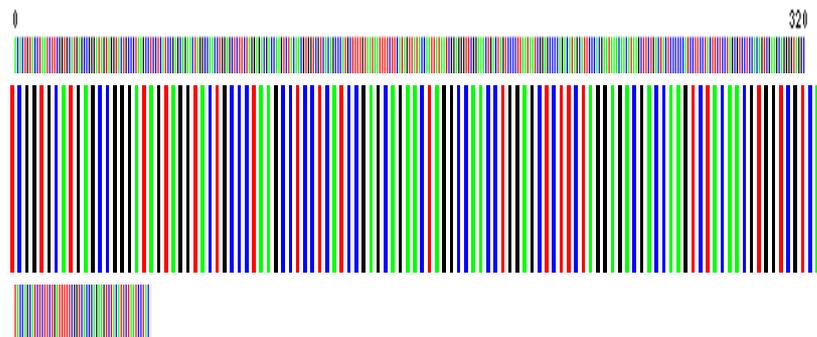
**Classical  
Taxonomy**



## DNA-based identification system:

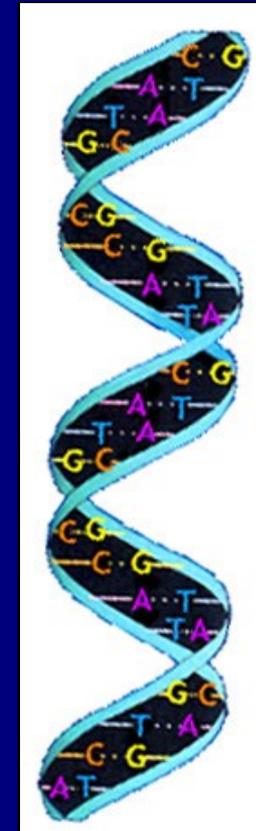


- Universal Product Code
- 10 state
- 11 positions

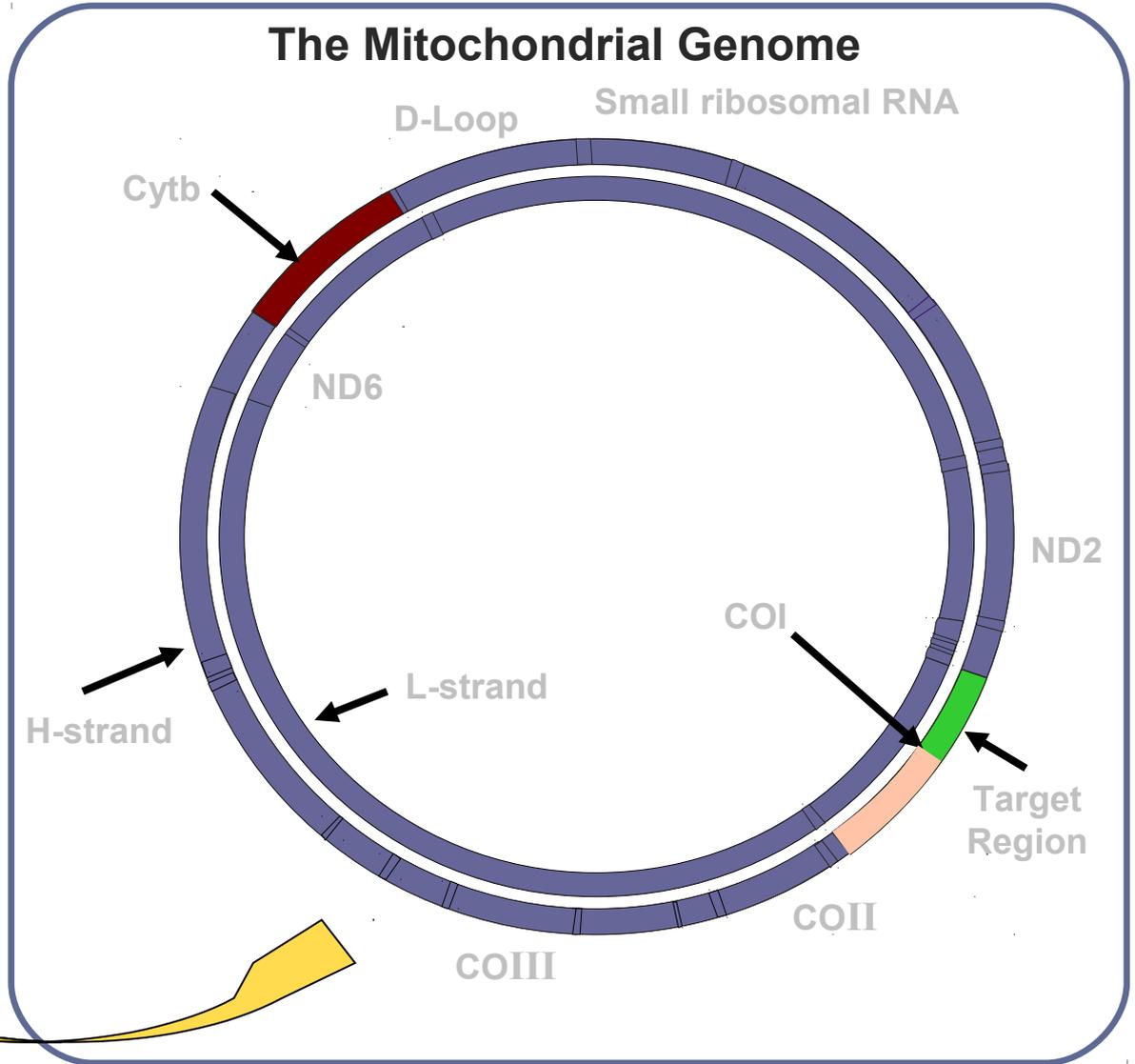
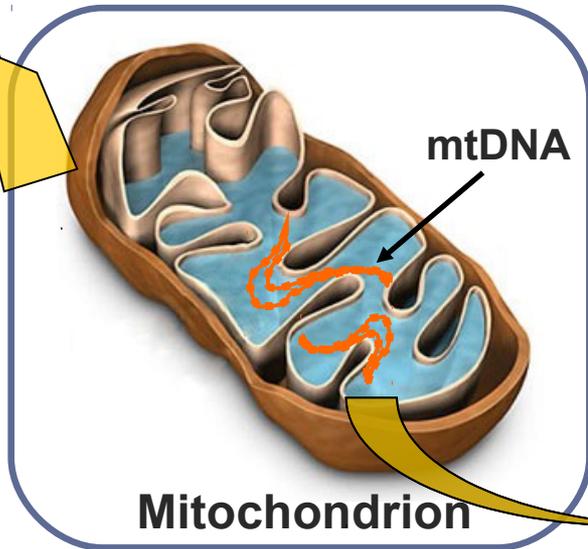
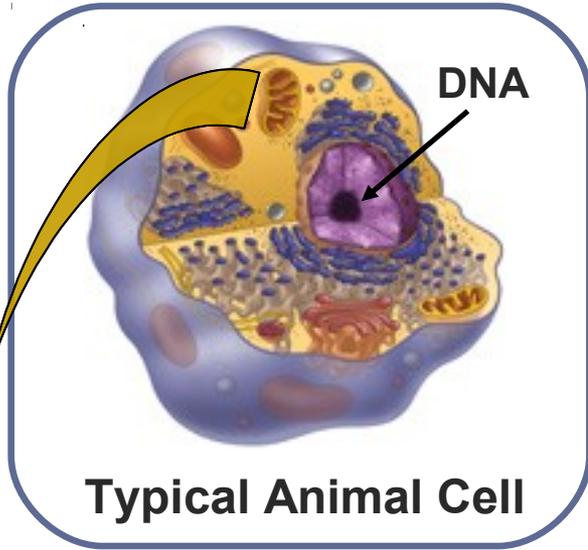


- DNA “Barcode”
- 4 states
- 655 positions

## DNA-based identification system:



# An Internal ID System for All Animals



## Costs and Time



**Tissue Sampling**

**Fresh/Frozen**

**\$0.41**

**Time**

**10**



**DNA Extraction**

**\$0.34**

**10**



**PCR Amplification**

**\$0.24**

**20**



**PCR Product Check**

**\$0.35**

**5**



**Cycle Sequencing**

**\$1.04**

**30**



**Sequencing Cleanup**

**\$0.32**

**5**



**Sequence**

**\$0.40**

**35**

**Total:**

**\$3.10**

**115 min**

# Producing Barcode Data: 2006



- **Hundreds of samples per day**
- **costing several dollars per sample**

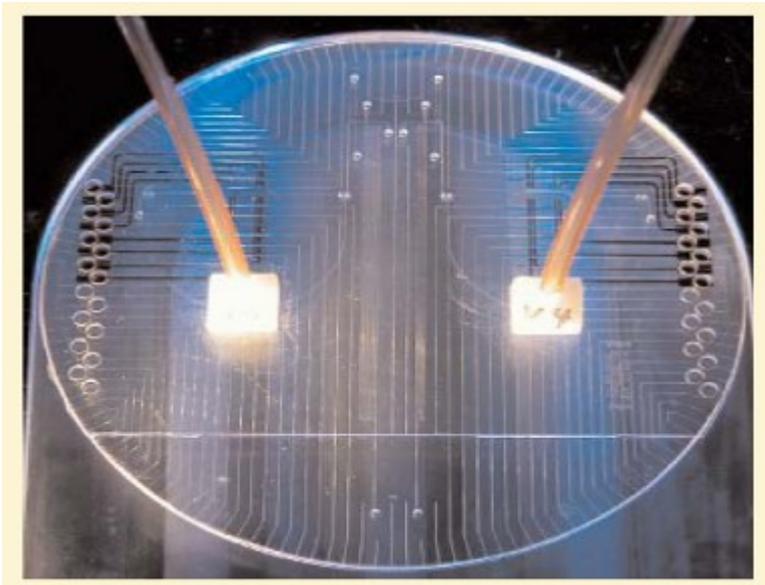


**ABI 3100 capillary**

**automated sequencer**

# Producing Barcode Data: 2008

**Faster, more portable: Hundreds of samples per hour**



**Integrated DNA microchips**



**Table-top microfluidic systems**

# Producing Barcode Data: 2010?

Barcode data anywhere, instantly



- Data in seconds to minutes
- Pennies per sample
- Link to reference database
- A taxonomic GPS
- Usable by non-specialists

# A Field Guide for the Third Millennium



# FISH-BOL

## Fish Barcode of Life Initiative

[Home](#)[Vision](#)[People](#)[Progress](#)[Checklists](#)[News & Updates](#)[Links](#)

## People

[Home](#) > [People](#)

Listed below are the people involved in running FISH-BOL, both the administrative heads and the members of the various working groups. Please click on the members names to retrieve their contact information

### Governance of FISH-BOL

The global administration of the Fish Barcode of Life Initiative is led by;

- Co-Chair: [Paul Hebert](#)
- Co-Chair: [Bob Ward](#)
- Campaign Coordinator: [Robert Hanner](#)

### Regional Working Groups

FISH-BOL has been broken into 10 regional working groups, the contacts for each group are listed below;

#### [India](#)

[Top](#)

The administration of the Indian Regional Working Group is led by;

- Chair: [Wazir Lakra](#)

The members are;

- [Achamveettil Gopalakrishnan](#)
- [Pallipuram Jayasankar](#)
- [Wazir Lakra](#)
- [Mahender Verma](#)

#### [North America](#)

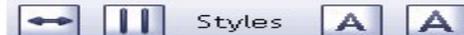
The administration of the North American Regional Working Group is led by;

- Chair: [Phil Hastings](#)
- Vice Chair: [Richard Mayden](#)

The members are;

- [Jesse Ausubel](#)
- [Andy Bentley](#)
- [Louis Bernatchez](#)
- [Vadim Birstein](#)
- [Jim Ruffillier](#)

 Search



### People Sections

- [Global Administration](#)
- [Africa](#)
- [Australia](#)
- [Central America](#)
- [Europe](#)
- [India](#)
- [North America](#)
- [North East Asia](#)
- [Oceania/Antarctica](#)

### Achamveettil Gopalakrishnan



**Title:** Senior Scientist  
**Address:** NBFGR Cochin Unit  
 Central Marine Fisheries Research Institute Campus  
 P.B. No. 1603  
 Cochin-682018, Kerala  
 India  
**Tel:** +91 984 711 5515  
**Fax:** +91 484 239 5570  
**Email:** [agopalkochi@rediffmail.com](mailto:agopalkochi@rediffmail.com)



# FISH-BOL Team at NBFGR





# First International Training on “DNA Barcoding of Marine Life” organized at NBFGR, Lucknow, India



## Molecular identification and phylogenetic relationships of seven Indian Sciaenids (Pisces: Perciformes, Sciaenidae) based on 16S rRNA and cytochrome c oxidase subunit I mitochondrial genes

Wazir S. Lakra · M. Goswami · A. Gopalakrishnan

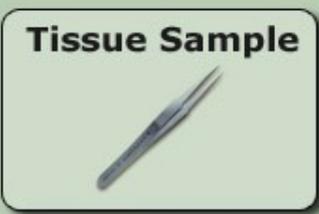
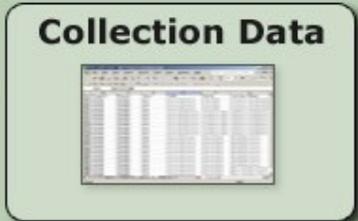
Received: 31 December 2007 / Accepted: 3 April 2008

© Springer Science+Business Media B.V. 2008

**Abstract** The partial sequences of 16S rRNA and cytochrome c oxidase subunit I (COI) mitochondrial genes were analyzed for species identification and phylogenetic relationships among the commercially important Indian sciaenids (*Otolithes cuvieri*, *Otolithes ruber*, *Johnius dussumieri*, *Johnius elongatus*, *Johnieops vogleri*, *Otolithoides biauritus* and *Protonibea diacanthus*). Sequence analysis of both genes revealed that the seven species fell into three distinct taxonomic groups, which are distributed

and South America, West Africa, South and South-East Asia [4]. The Indian sciaenids, including the present commercially important species (*Otolithes cuvieri*, *Otolithes ruber*, *Johnius dussumieri*, *Johnius elongatus*, *Johnieops vogleri*, *Otolithoides biauritus* and *Protonibea diacanthus*), contribute approximately 4.6% to the total Indian marine fish production.

Conventionally, sciaenid fishes are identified using



the animal kingdom  
**Barcodes of Life**

#### SPECIMEN DATA -Profile - Guelph Moths

Identification :	<i>Baleya ophthalmica</i>		
Specimen Accession :	moth784.01	Museum Accession :	moth 784.01
Specimen Label :	moth 784.01	Institution Holding :	University of Guelph
Sex :	Male	Collector :	Paul Hebert
Reproduction :	Sexual	Date Collected :	2001-06-20
Life Stage :	Adult	Identifier :	Paul Hebert
GPS Latitude :	43.537	Common Name :	Eyed <i>Baleya</i>
GPS Longitude :	-80.1253	Taxonomy :	phylum - Arthropoda
Elevation (meters) :	320	class - Insecta	order - Lepidoptera
Country :	Canada	family - Noctuidae	genus - <i>Baleya</i>
State/Province :	Ontario	species - <i>Baleya ophthalmica</i>	
Region :	Wellington County		
Sector :	Paslinch Township		
Site :	Concession 11		

Notes :

Collection Location (click on image to zoom)

Wing Span = 3.2 cm  
Click on image to enlarge

the animal kingdom  
**Barcodes of Life**

#### SEQUENCE RELATED DATA -Profile - Guelph Moths

Specimen Name :	<i>Baleya ophthalmica</i>	Translation Matrix :	Invertebrate Mitochondrial Code
Specimen Accession :	moth784.01	Electropherogram :	
Sequence ID :	PH0093-05		
Primer Name :		A - Count :	100
5' Primer :		G - Count :	87
3' Primer :		T - Count :	247
		C - Count :	95
Sequence Length :	617	AT - Content :	70.5%
		GC - Content :	29.5%

**Nucleotide Sequence :**

```

GAACTCTTTAAGCTTATTAATTCGAGCAGAATTAGAAAACCCGTTCAATTAATGGAG
ATGATCAAAATTTAATACTATTGTTACAGCTCATGCTTTTATTATAATTTCTTTATAG
TTATACCAATATATTGGAAGATTTTGAAGATTAATTAATACCTCAATATTAATGAGCC
CTGATATAGCTTCCGCCAAATAAATAATTAAGTITTTGATTTATAGCCGCTCTTGA
CTCTACTAATTTCTAGTAGAATCTGAGAAATGGAGCAGGACAGGATGACAGCTTACC
GDCGACTTCATCAAAATATTCCTATGCGGAAATTCAGTAAATTAAGTATTTCTCAT
TACATTTAGCTGGAATTTGATCTATTTAGAGCTATTAAATTTTACAGCTATATTA
ATATAGCAATTAATAATTAATCATTTGATGAAATAGCTTATTTGATGAGGAGTAA
TTACAGCTTTTACTCTATTATCTCTCTGTTTAACTGAGCTGAGCTATTATAT
TAACAGATCGTAATTTAAATCTCAATTTTGTGCTCATGGAAGAGGAGATCTTATT
TAAATCAACATTTATT
  
```

**Amino Acid Sequence :**

```

YTSLLSLDAELGHPDGLIGDQQLYHTVTAHAFHFFNVDPIKGGFHWLSLKLGLGAPGMATSEM
NNHSMLEPPSLTLRRTSIVNSGATAMTVVLESLRINAGRREVDLAPSLHAGLISRLGATNF
TTEHPLNLNSLFDQMLFVWAVGTFAPLLLSLPLVLAGAITMLLDRNLTFRFPAGGQGPLNQH
LF
  
```

# BARCODE OF LIFE DATA SYSTEMS

Advancing species identification and discovery through the analysis of short, standardized gene regions

SEARCH

About BOLD Contact Us



[Published Projects](#) | [Taxonomy Browser](#) | [Request an Account](#) | [Identify Specimen](#) | [Introductory Tutorial](#) | [Documentation](#) | [Citation](#)

The Barcode of Life Data Systems (BOLD) is an online workbench that aids collection, management, analysis, and use of DNA barcodes. It consists of 3 components (MAS, IDS, and ECS) that each address the needs of various groups in the barcoding community.



## MANAGEMENT & ANALYSIS

**BOLD-MAS** provides a repository for barcode records coupled with analytical tools. It serves as an online workbench for the DNA barcode community.

Username

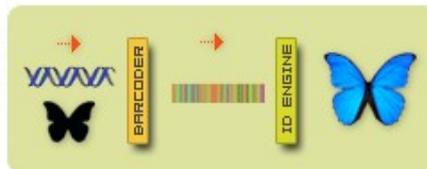
Password

[Request a new user account](#)



## IDENTIFICATION ENGINE

**BOLD-IDS** provides a species identification tool that accepts DNA sequences from the barcode region and returns a taxonomic assignment to the species level when possible.



## EXTERNAL CONNECTIVITY

**BOLD-ECS** provides web developers and bioinformaticians the ability to build tools and workflows that can be integrated with the BOLD framework. We welcome the addition of new analytical modules.



## BARCODE COUNTS

Formally Described Species With Barcodes	41,626
Total Barcode Records	426,327
Source	Breakdown
GenBank	61,689
Canadian Centre	333,006
Others	31,632

## SYSTEM UPDATES

### Nov-23--2007 - Primer Registration (MAS)

Primer submission is now possible through the Project Console. Primers that target multiple markers are accepted as a prelude to supporting multi-marker barcoding.



### Sept-30--2007 - ID Engine Integration (MAS)

The BOLD ID engine has now been integrated into the sequence page. Users can quickly test the identification of any record on BOLD by clicking on the ID buttons found directly below the nucleotide sequence on the page. Both full database and reference database searches are supported.



### Aug-29--2007 - Google Earth (MAS)

Google Earth extension has been added to the Distribution map function on BOLD. Users with Google Earth installed on their system can map sample collection points on ultra-high resolution maps with specimen images.



### Mar-2--2007 - Sequence Quality Measure (MAS)

BOLD now provides the number of ambiguous bases in a sequence wherever the sequence length is displayed. Sequences with greater than 1% ambiguous bases are highlighted for review. The Taxon ID tree function also provides a filter to restrict any sequences with greater than 1% ambiguous bases from the analysis.

### Feb-12-2007 - Taxonomy Browser

A Taxonomy Browser is now available off the front page of BOLD. Users can browse and search for taxon pages that provide an overview of the barcoding



Bats of Southeast Asia [BIM]

Specimen Identifiers

Edit Specimen

Sample ID:	ROM 101996	Museum ID:	101996
Isolate / Field Num:	F35806	Collection Code:	MAMM
Donated By:	Judith L. Eger	Deposited In:	Royal Ontario Museum

Taxonomy

Identifier:	Mark D. Engstrom
phylum:	Chordata
class:	Mammalia
order:	Chiroptera
family:	Pteropodidae
genus:	Macroglossus
species:	Macroglossus minimus

Specimen Details

Voucher Type:	Skin, Skull, Skeleton
Tissue Type:	Frozen Liver
Extra Info:	F35806 - E Kalimantan
Sex:	Male
Reproduction:	Sexual
Life Stage:	Adult

Collection Data

Collectors:	Mark D. Engstrom
Date Collected:	22-May-1993
Country:	Indonesia
State/Province:	Kalimantan Timur
Region/County:	East Kalimantan
Sector:	60
Exact Site:	
Latitude:	-0.8
Longitude:	117.1
Coord. Source:	
Elevation/Depth:	60



Photographs



Bats of Southeast Asia - first paper [BIM]

Barcode Identifiers

Barcode ID:	BM258-04	Sample ID:	ROM 101996
Gene:	COX1	GenBank Accession:	Record is unpublished
Last Updated:	2005-08-26	Translation Matrix:	Vertebrate Mitochondrial

Sequencing Runs

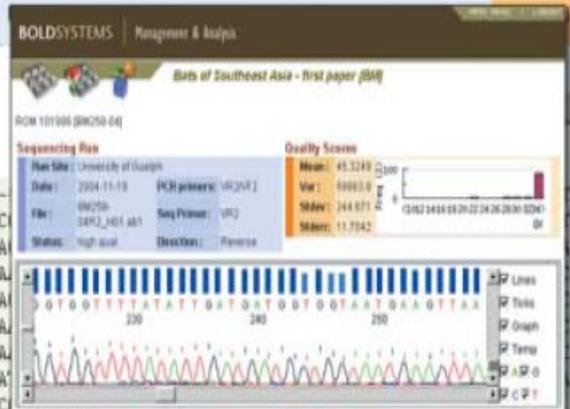
Run Date	Run Site	Direction	Trace File	PCR primers	Seq Primer	Status
2004-11-18 15:18:37	University of Guelph	Reverse	BM258-04R2_H01.ab1	VR2/VF2	VR2	high qual
2004-11-18 11:55:58	University of Guelph					

Nucleotide Sequence

Length: 619  
 Comp. A: 190  
 Comp. G: 95  
 Comp. C: 169  
 Comp. T: 165  
 Updated: 2005-08-26

```

-----XGT
GGACAACCAGGAGCC
TTTTCATAGTAATA
ATAGCATTCCCCGA
ACAGTAGAAGCTGGA
TCTGTAGATCTAGCA
ACCATCATCAATATA
GCAGTCTACTATTA
ACAACTTCTTCGAC
    
```

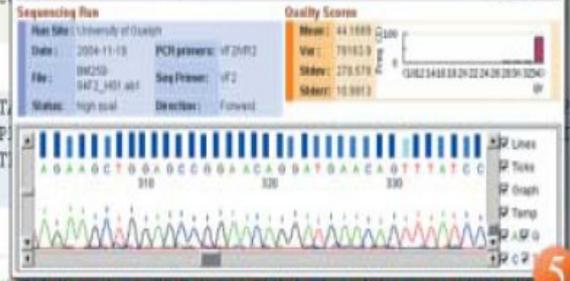


Amino Acid Sequence

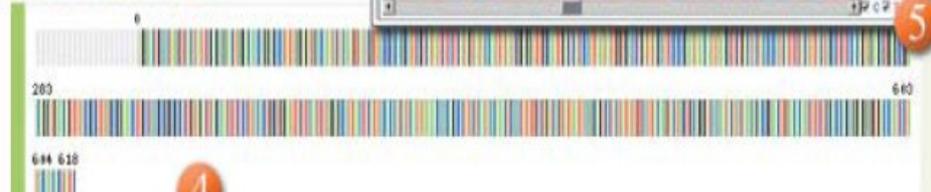
Length: 206

```

-----XGT
NAFFPMMNMSFWLLP
TIINRKPALSOYQT
    
```



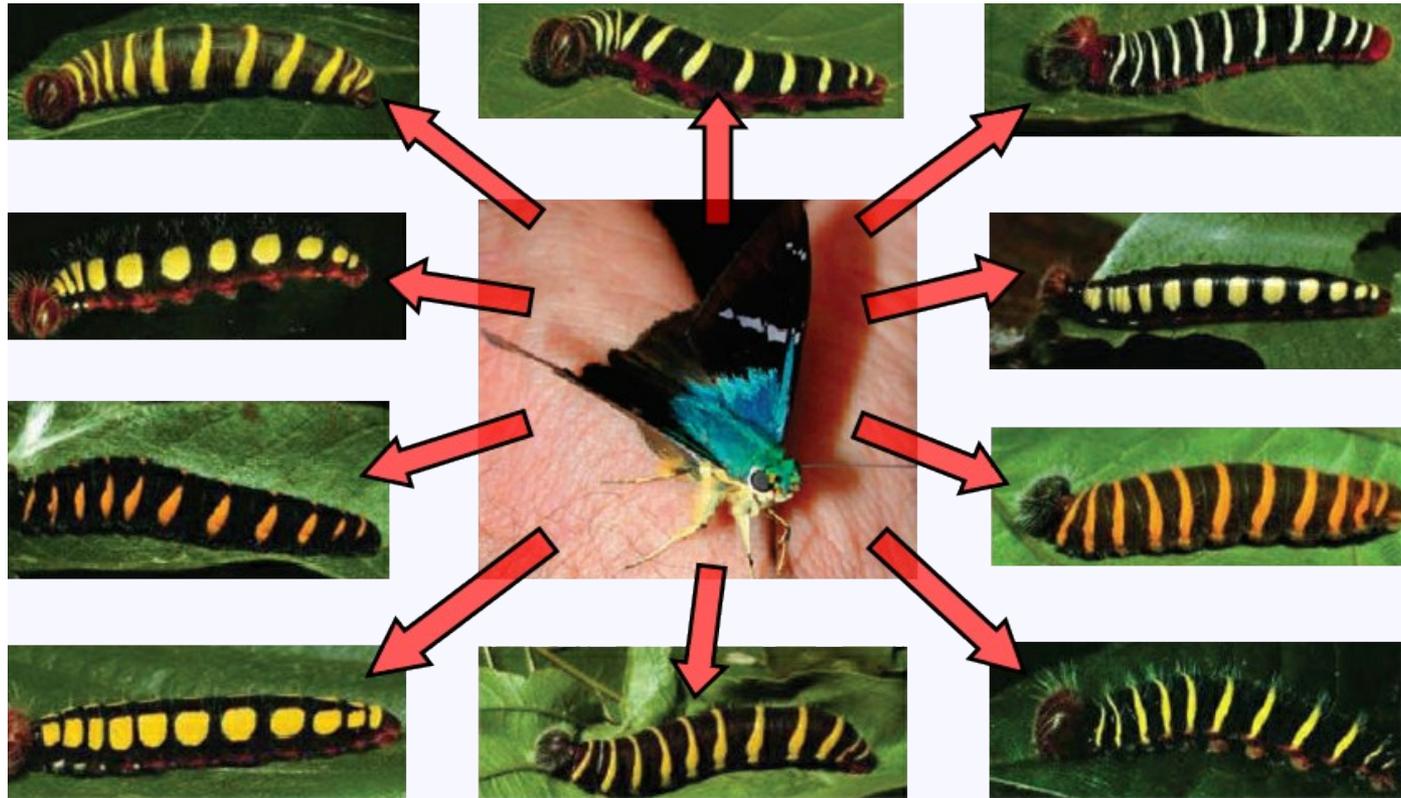
Illustrative Barcode



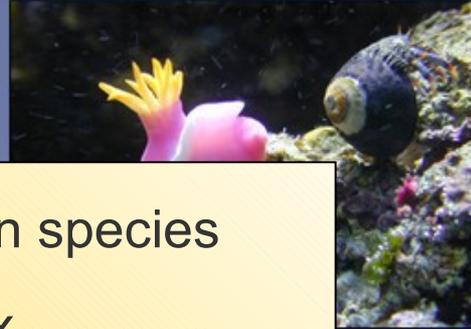
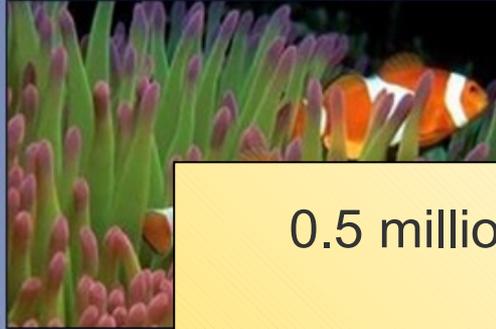
## Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*

Paul D. N. Hebert<sup>\*\*†</sup>, Erin H. Penton<sup>\*</sup>, John M. Burns<sup>‡</sup>, Daniel H. Janzen<sup>§</sup>, and Winnie Hallwachs<sup>§</sup>

<sup>\*</sup>Department of Zoology, University of Guelph, Guelph, ON, Canada N1G 2W1; <sup>†</sup>Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0127; and <sup>§</sup>Department of Biology, University of Pennsylvania, Philadelphia, PA 19104



# Gaining Closure for All Marine Life



0.5 million species

x

10 barcodes each

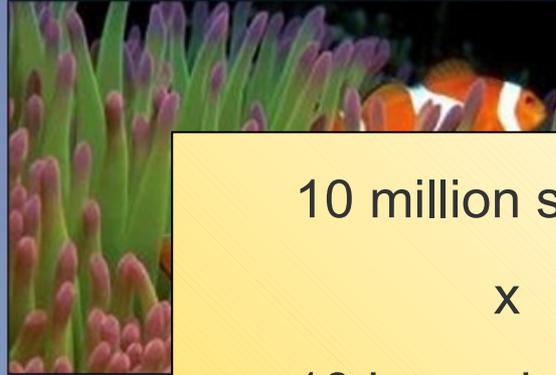
=

5 million barcodes

& DNA extracts



# Gaining Closure for All Animals



10 million species

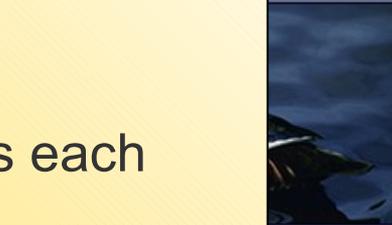
x

10 barcodes each

=

100 million barcodes

<\$500,000,000





# Forensic Applications of DNA Barcoding

## Identification of Endangered Whale Shark (*Rhincodon typus*).

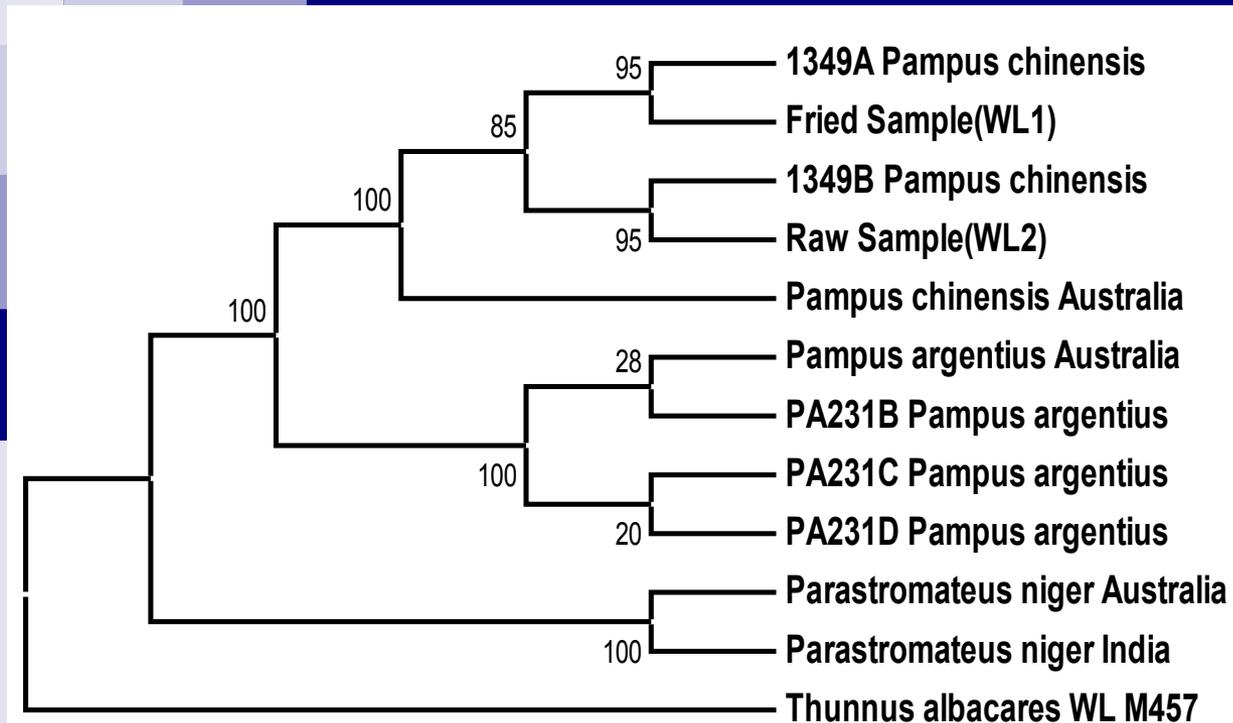
- Suspected Flesh of Banned Whale Shark (*Rhincodon typus*) seized by Forest Range Officer, Kannur, Kerala.
- The Judicial First Class Magistrate, Thalassery, Kannur, Kerala approached NBFGR Cochin Unit for Analysis and confirmation of species.
- Based on DNA Barcoding it was identified as of endangered Whale Shark (*Rhincodon typus*).
- Confirmed with DNA sequencing of 16S rRNA(525bp) and COI (600bp) Cyt b(541bp) genes.



# Forensic Applications of DNA Barcoding



Pomphret Identification from fried and raw fish.



**NJ Tree of Pomphret Species found in India with two samples(WL1 & WL2).**



# Issues and Challenges



- **DNA Barcoding vs. Traditional Taxonomists**
- **Networking of people and the funding agencies**
- **Lack of qualified Taxonomists**
- **Issue of voucher specimens located in India's vast network of Universities, Colleges and Research Institutes**
- **Plants, Hard corals & Fungi**

## Chloroplast DNA (cp DNA):

- **A double-stranded circular DNA - chloroplast chromosome or cpDNA.**
- Chloroplasts genomes - 3 types.

(Group I) - The Gymnosperms, Pinaceae and a group of legumes (peas and broad bean) have chloroplast chromosomes without an inverted repeat - IR

All other angiosperms have chloroplast genomes containing a large (6 - 76 kb) inverted repeat; these are Group II genomes.

The alga *Euglena* has three tandem repeats in its Group III chloroplast genome.

- **Uniparental mode of inheritance (usually maternal in angiosperms and paternal in gymnosperms)**
- **Chloroplast genomes contain between 120 - 140 genes and of ~160 kbp size (~10 times larger than mtDNA).**
- **Relatively stable genome with marked conservation of gene content.**

• Unlike the animal mitochondrial genome, cpDNA of several plants contain introns.

• The chloroplast genome shares many features with animal mtDNA and the two have been referred to as 'natural counterparts'.

• Regions such as *trH-psbA* spacer, *rbcL*, *matK*, *rpoC1*, *rpoB*, *accD* and *YCF5* are identified as the most promising regions in the cpDNA for DNA barcoding in plants and universal primer pairs for these regions have been developed ([barcoding@kew.org](mailto:barcoding@kew.org)).

• Occurrence of polymorphic mononucleotide microsatellites (cpSSR)

• Unlike, nuclear microsatellites, cpSSRs are uniparentally inherited, nonrecombinant and all loci are linked.

• Chloroplast microsatellites are fast evolving and typically consist of mononucleotide motifs that are repeated 8 to 15 times

• They represent potentially useful markers at the population level in plants.



# **ITS (Internal Transcribed Spacers)**

- **Eukaryotic ribosomal RNA genes (known as ribosomal DNA or rDNA) are found as parts of repeat units that are arranged in tandem arrays, located at the chromosomal sites known as nucleolar organizing regions (NORs).**
- **Each repeat unit consists of a transcribed region (having genes for 18S, 5.8S and 28S rRNAs and the external transcribed spacers i.e. ETS1 and ETS2) and a non-transcribed spacer (NTS) region.**
- **Internal transcribed spacers (ITS) are found on either side of 5.8S rRNA gene and these are described as ITS1 and ITS2.**
- **The length and sequences of ITS regions of rDNA repeats are believed to be fast evolving and therefore may vary**

- **This makes the ITS region an interesting subject for evolutionary/phylogenetic investigations as well as biogeographic investigations.**
- **Universal PCR primers designed from highly conserved regions flanking the ITS and its relatively small size (600-700 bp) enable easy amplification of ITS region due to high copy number (up to-30000 per cell) of rDNA repeats.**

