Molecular identification of species, individuals and sex

Microsatellites

Pavel Munclinger

Species identification

- **DNA barcoding**

  – taxon identification
  using a standardized DNA region

Genebank (NCBI)
http://www.ncbi.nlm.nih.gov/genbank/

BOLDSYSTEM
http://v3.boldsystems.org/

BLAST or special programs

DNA barcoding

Hebert et al. 2003

The use of limited (approximately 600 bp) mitochondrial DNA sequence data as an inexpensive, easy way to "scan" and **identify all of life**.

SEQUENCE: COI-5P [Funding Source: N/A]

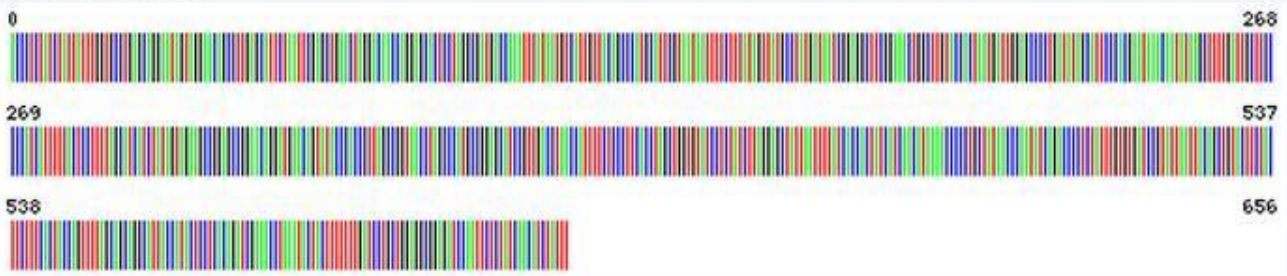
Sequence ID:	ABCMA692-07.COI-5P	GenBank Accession:	JF445285
Last Updated:	2013-02-12	Genome:	Mitochondrial
Locus:	Cytochrome Oxidase Subunit 1 5' Region		
Nucleotides:	657 bp		

```
ACCCTCTATCTATTATTTGGTGCCTGAGCAGGAATAGTAGGAACAGCCTTGAGCATTCTAATTCGAGCTGAACTA
GGACAACCAGGAGCACTCCTAGGCGATGACCAAATTTATAATGTCATTGTTACAGCCCATGCATTCGTAATAATT
TTCTTTATAGTTATGCCTATGATAATCGGAGGCTTCGGAAACTGGCTTGTACCACTAATGATTGGAGCCCCTGAT
ATAGCATTCCCACGAATAAACAATATAAGCTTTTGATTGCTTCCCCCATCATTTTTACTCCTTTTAGCATCATCT
ATAGTAGAAGCCGGAGCCGGAACAGGATGAACAGTATACCCACCCTTAGCCGGTAACCTAGCCCATGCCGGAGCA
TCCGTTGACCTAACCATTTTCTCCCTTCACCTAGCTGGTGTATCCTCTATCTTAGGAGCTATTAATTTTATCACC
ACTATCATCAACATAAAACCCCCTGCTATAACCCAATATCAGACCCCTCTATTTGTGTGATCCGTATTAATTACA
GCTGTACTTCTACTTCTTTCACTACCAGTTTTAGCAGCAGGCATTACCATACTCCTCACAGATCGAAACCTAAAT
ACTACTTTTTTTGATCCTGCTGGAGGCGGAGATCCAATTCTCTATCAACATCTATTT
```

Amino Acids:

```
TLYLLFGAWAGMVGTALSILIRAELGQPGALLGDDQIYNVIVTAHAFVMIFFMVMPMMIGGFGNWLVPLMIGAPD
MAFPRMNNMSFWLLPPSFLLLLASSMVEAGAGTGWTVYPPLAGNLAHAGASVDLTIFSLHLAGVSSILGAINFIT
TIINMKPPAMTQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLF
```

Illustrative Barcode:



. . . enabling the rapid and inexpensive identification of the estimated 10 million species on Earth. (*Savolainen et al. 2005*)

Database needed!

CONSORTIUM FOR THE BARCODE OF LIFE

Canadian Centre for DNA Barcoding

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

www.barcodeoflife.org

All Leps

Lepidoptera: The All Leps campaign is assembling barcodes for 25,000 species of Lepidoptera, focusing on the faunas of Australia, Canada, Costa Rica and the United States.
7512 species barcoded | **View**

FISH-BOL

Fishes: The FISH-BOL campaign is gathering barcodes for all species of fishes (approximately 30,000) with an emphasis on the 15,000 marine species.
2538 species barcoded | **View**

Canadian Fauna: The Canadian Barcode of Life Network, incorporates an on-going initiative to barcode all Canadian species.
2133 species barcoded | **View**

Birds: The All-Birds Barcoding Initiative (ABBI) is assembling DNA barcodes for all 10,000 bird species, and aims to complete this task within 5 years.
1233 species barcoded | **View**

COI

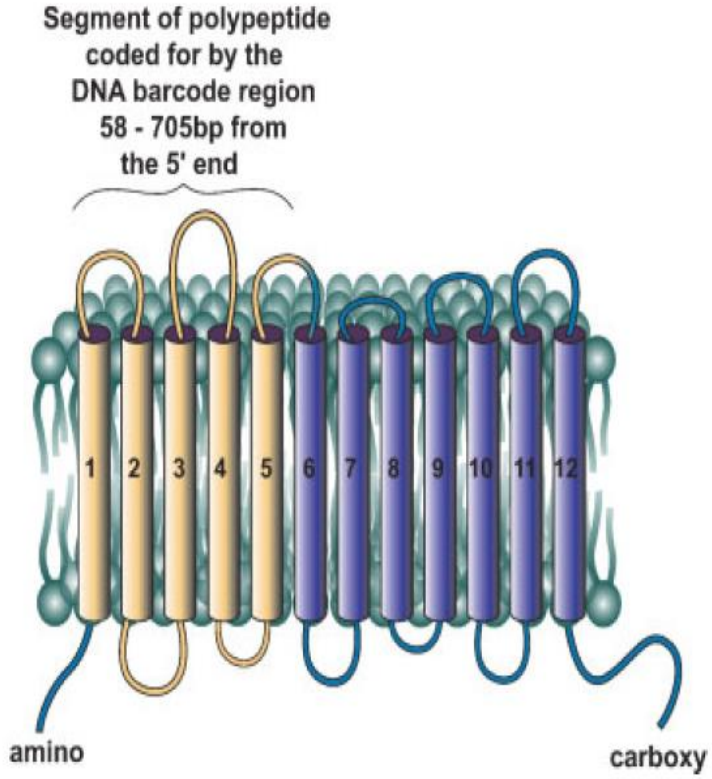


Figure 2. The predicted transmembrane structure for cytochrome *c* oxidase subunit 1 (COI). The area highlighted in yellow that includes five of the twelve transmembrane regions is coded for by the sequence designated by CBOL as the DNA barcode region.

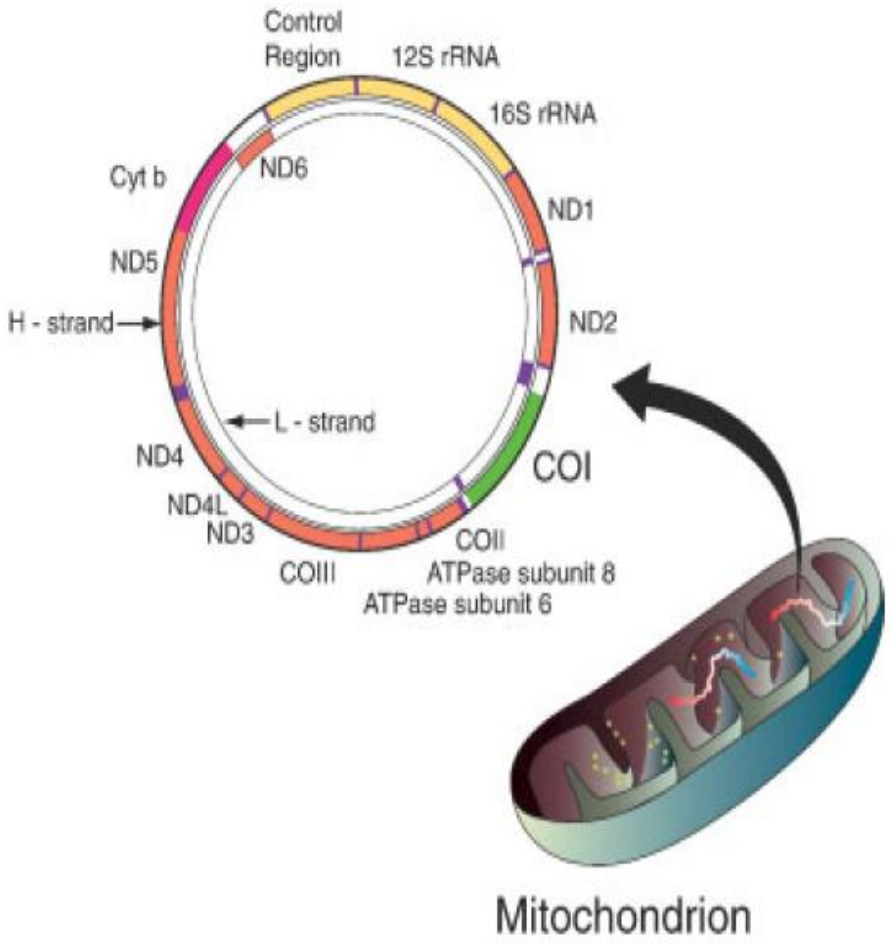
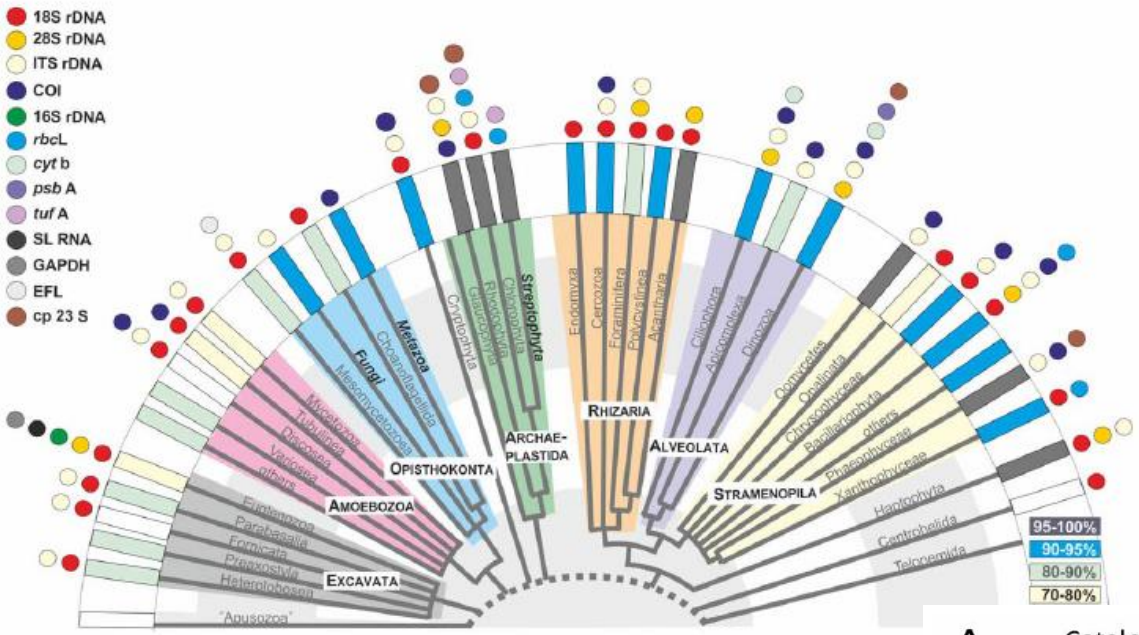
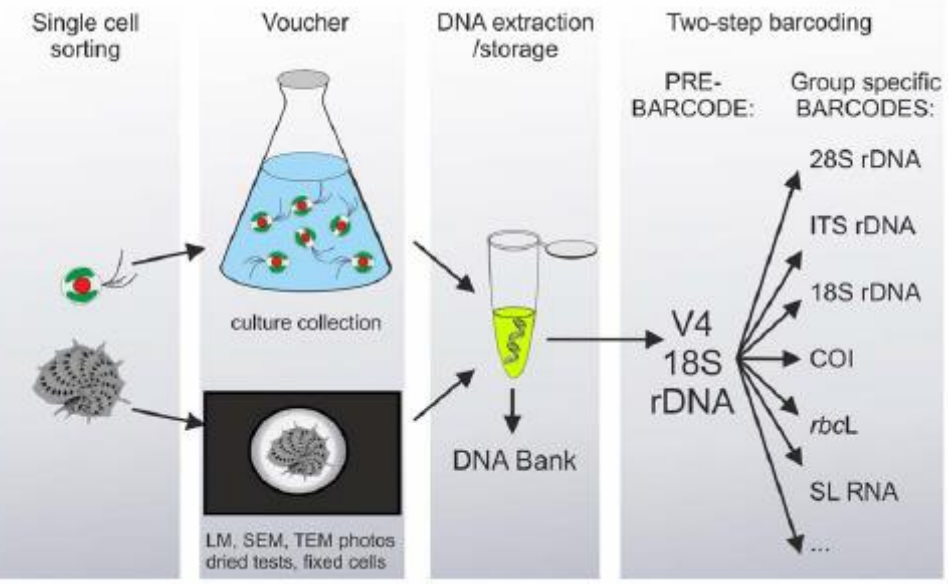
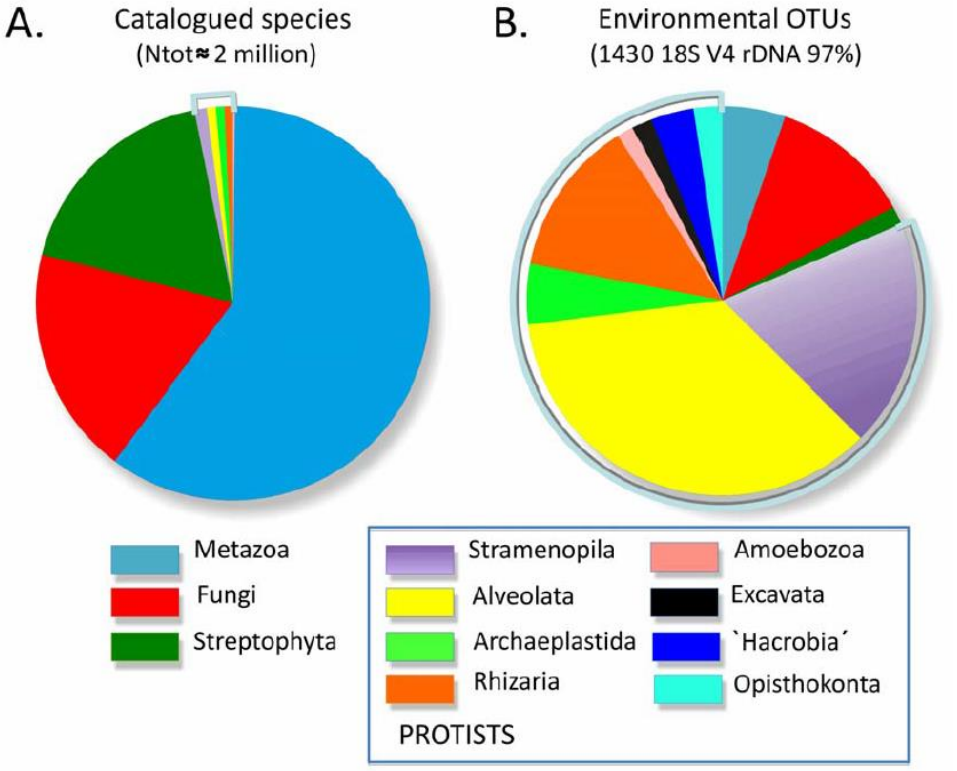


Figure 1. A diagrammatic representation of a mitochondrial genome based on complete mitochondrial DNA sequences from a variety of bird species.


18S rDNA
28S rDNA
ITS rDNA
COI
16S rDNA
rbcL
cyt b
psb A
tuf A
SL RNA
GAPDH
EFL
cp 23 S
Metazoa
Fungi
Mesomycetozoea
Choanomonadida
Choanoflagellatea
Opisthokonta
Amoebozoa
Mycetozoa
Tubulinea
Discosea
Variosea
others
Excavata
Euglenozoa
Parabasalia
Fornicata
Preaxostyla
Heterolobosea
"Apusozoa"
Archae-plastida
Cryptophyta
Rhodophyta
Glaucocystophyta
Chlorophyta
Streptophyta
Rhizaria
Foraminifera
Endomyxa
Cercozoa
Radiolaria
Acantharia
Alveolata
Ciliophora
Apicomplexa
Dinozoa
Stramenopila
Opalozoa
Oomycetes
Chrysophyceae
Bacillariophyta
others
Phaeophyceae
Xanthophyceae
Haptophyta
Centrohelida
Telonemida
95-100%
90-95%
80-90%
70-80%


Single cell sorting
Voucher
DNA extraction /storage
Two-step barcoding
culture collection
LM, SEM, TEM photos
dried tests, fixed cells
DNA Bank
PRE-BARCODE:
Group specific BARCODES:
V4 18S rDNA
28S rDNA
ITS rDNA
18S rDNA
COI
rbcL
SL RNA
...

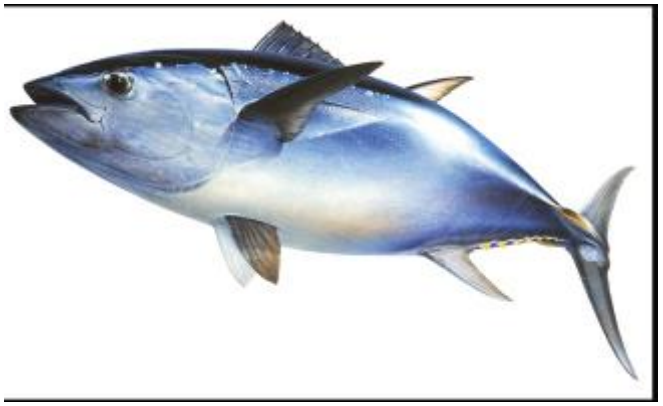

A.
Catalogued species
(Ntot≈2 million)
B.
Environmental OTUs
(1430 18S V4 rDNA 97%)
Metazoa
Fungi
Streptophyta
Stramenopila
Alveolata
Archaeplastida
Rhizaria
Amoebozoa
Excavata
`Hacrobia´
Opisthokonta
PROTISTS

Hesperiidae Costa Rica (Skippers)

The 28 last instar caterpillars

Lowenstein et al. 2009
Sushi

albacore
Thunnus alalunga



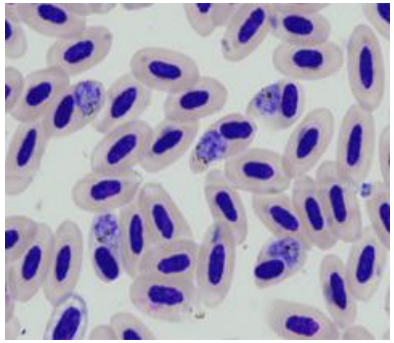
bluefin tuna
Thunnus maccoyii
critically endangered

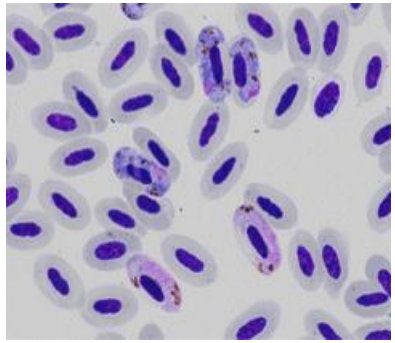
escolar
Lepidocybium flavobrunneum
health concerns

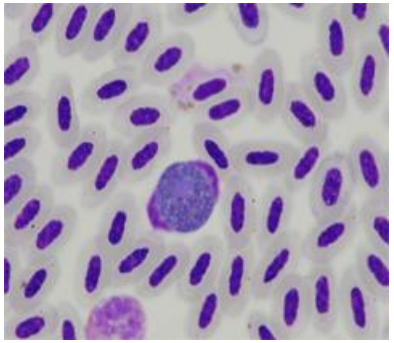
DNA barcoding – parasites

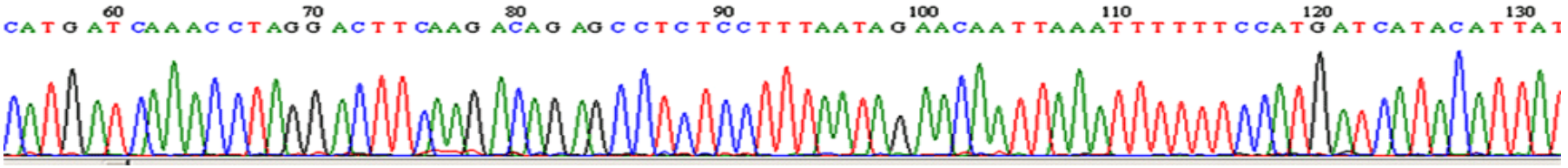
Avian malaria
Haemoproteus, Plasmodium, Leucocytozoon

DNA from bird blood, parasite-specific primers
→ parasite lineages



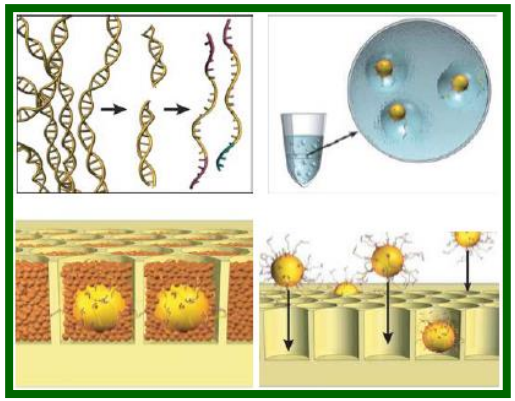






Alcaide et al. 2009
Arthropod bloodmeal

Culex pipiens

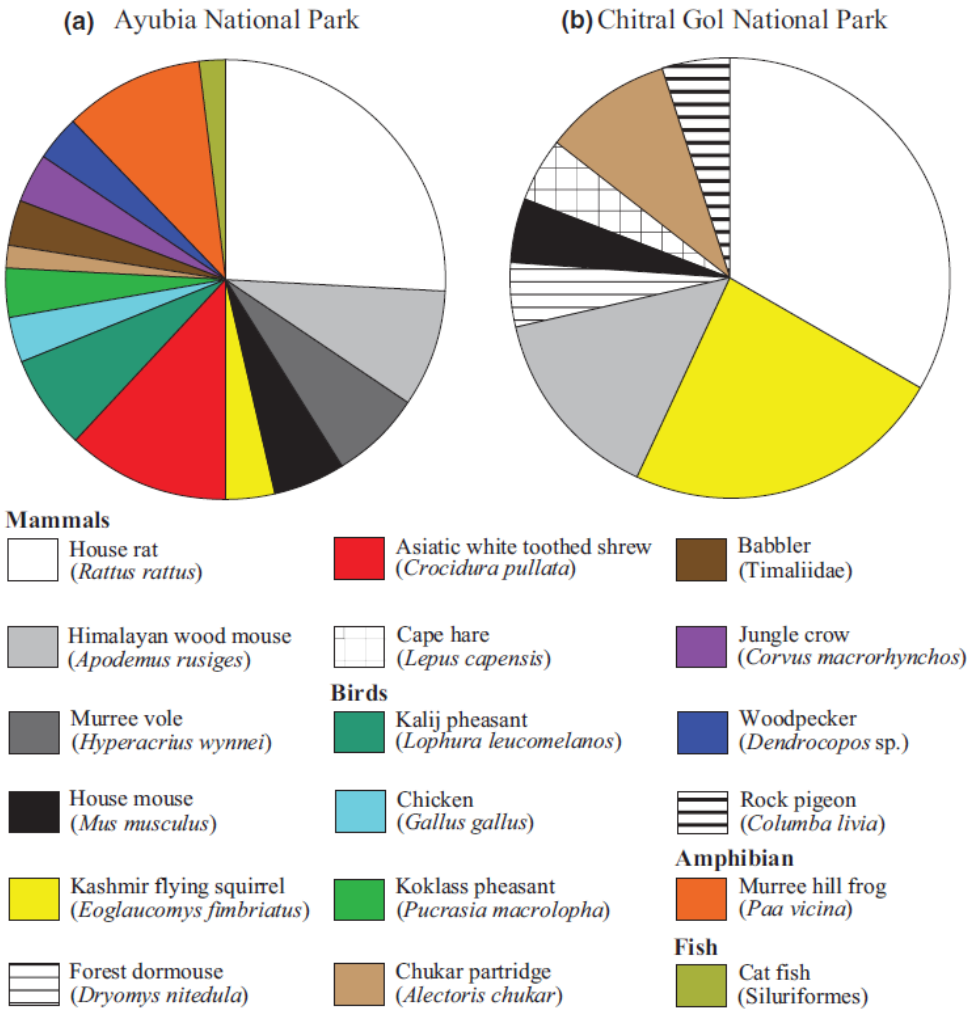


NGS

Illumina

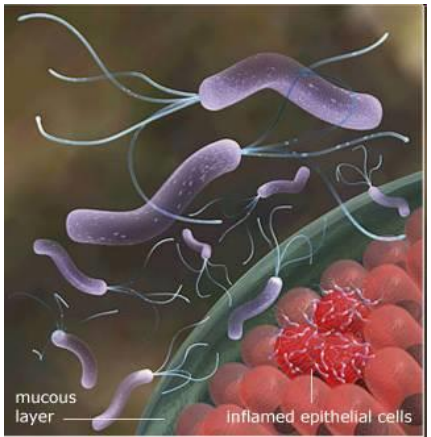
Leopard cat in Pakistan
Prionailurus bengalensis

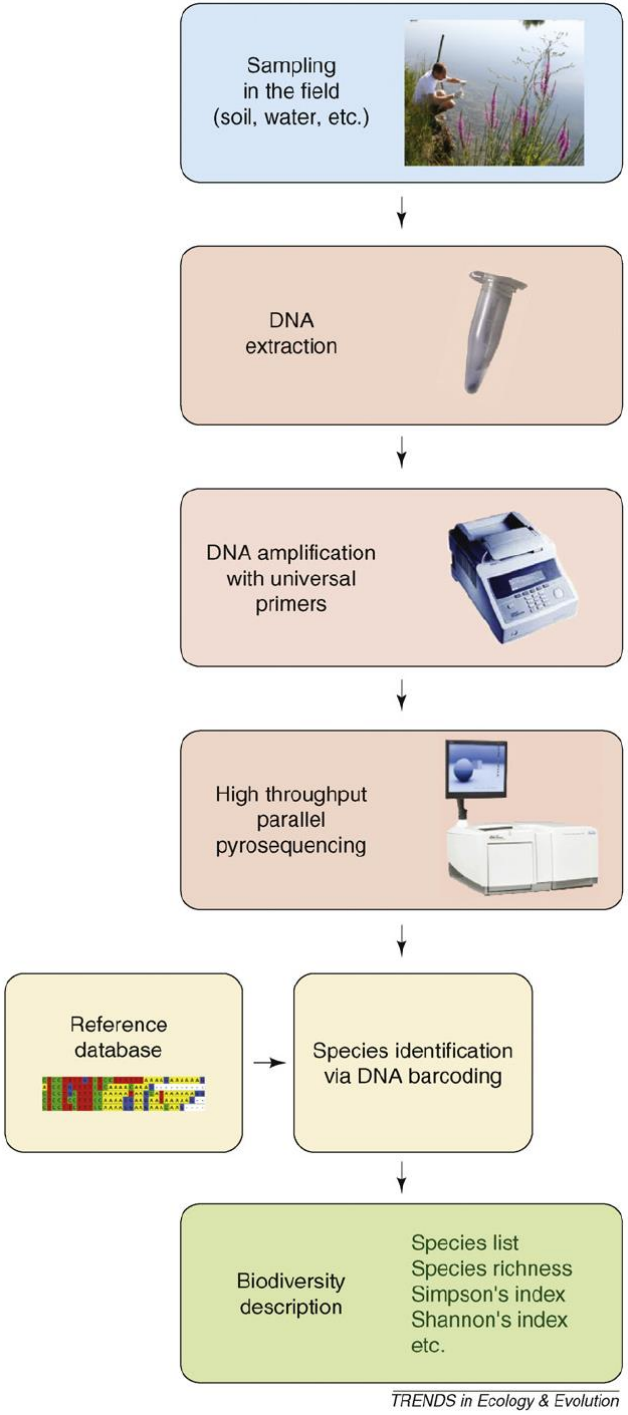
- Shezhad et al. 2012
- 12S rRNA
- Illumina



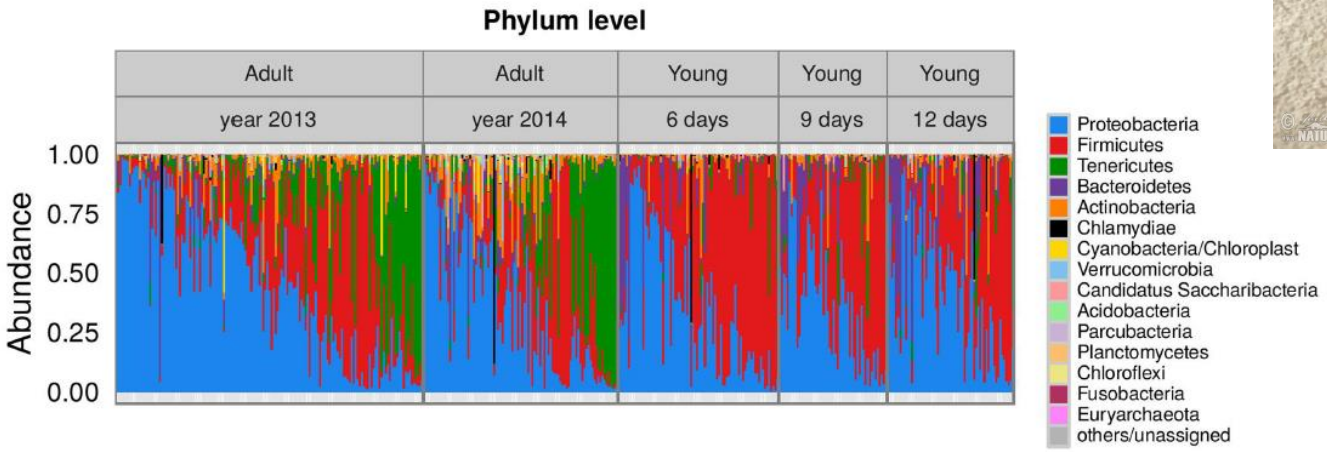
Metagenomics

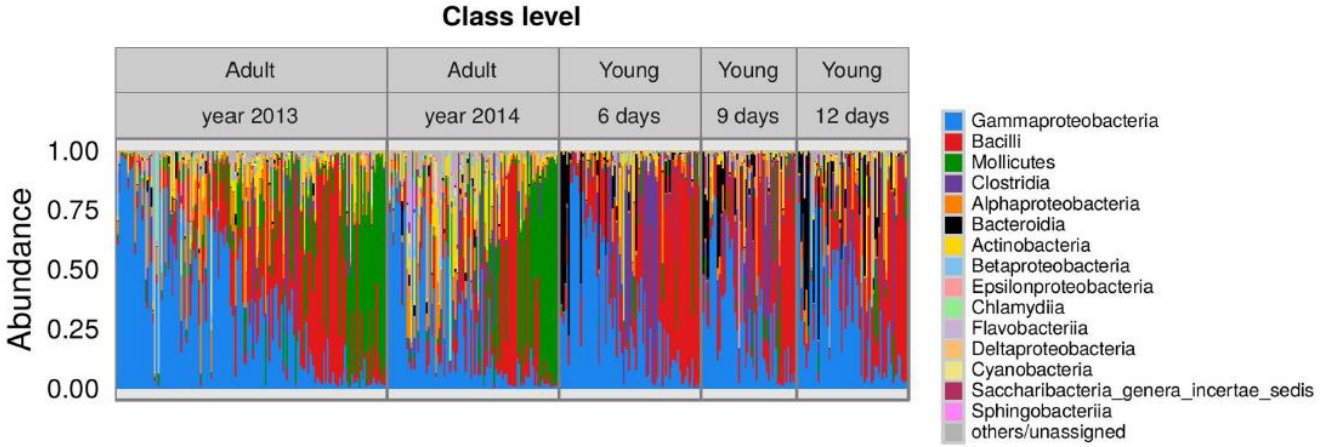
- Metagenomics:
  DNA sequencing of environmental samples
- water (sea, pool), clouds…
- Reconstruction of ancient vegetation (permafrost)
- Earthworm extracelular DNA from soil…
- Gut microbiome





Microbiota





Kreisinger et al. 2017

intraspecific divergence
<<
Interspecific divergence

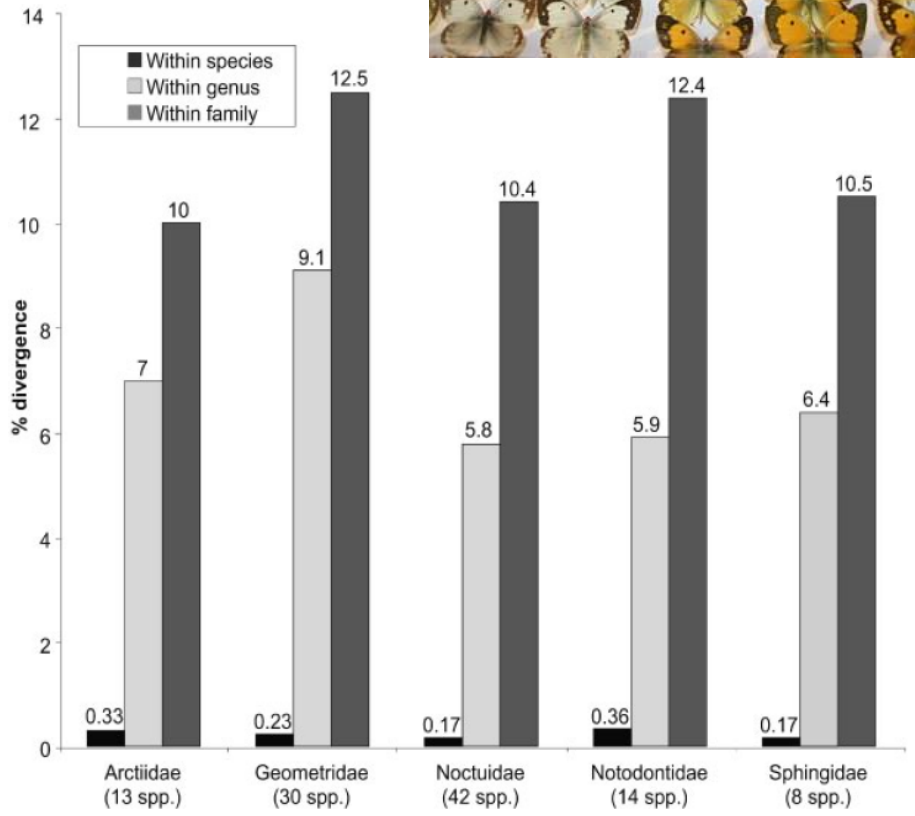


Figure 3. Nucleotide divergence in a 617 bp segment of the COI gene in five lepidopteron families at species, genus and family level. Data from Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a Proc R Soc Lond B 270:313–322

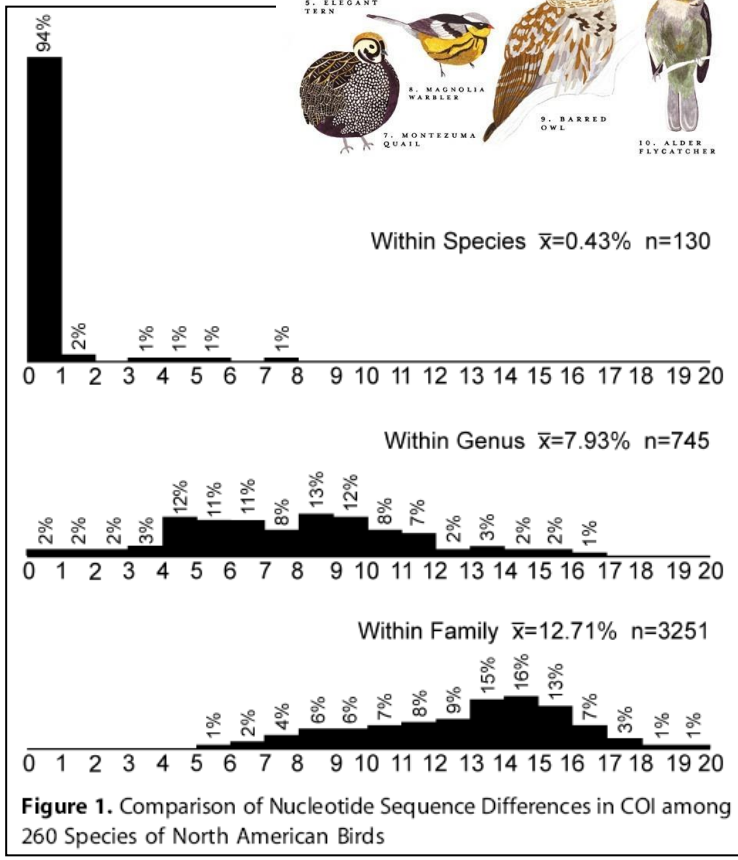
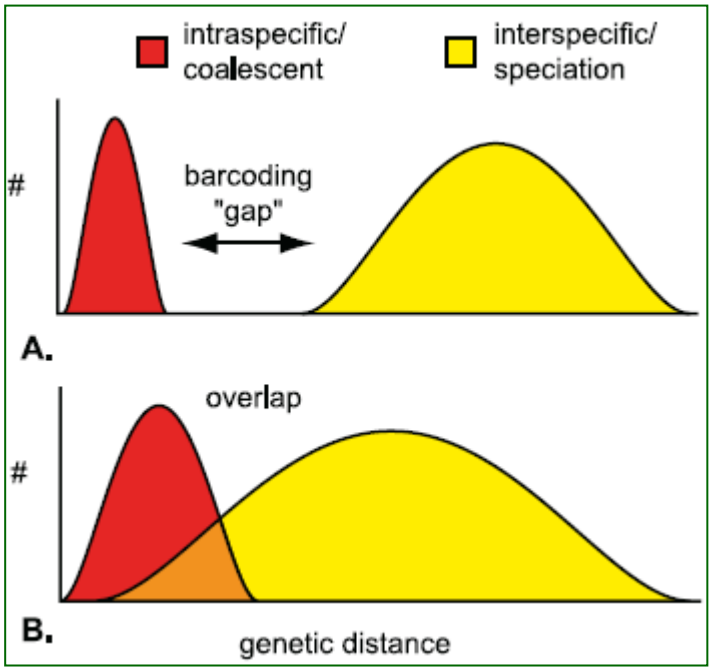


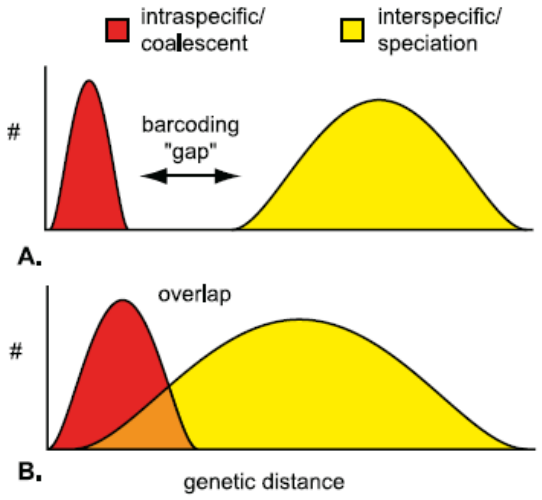
Figure 1. Comparison of Nucleotide Sequence Differences in COI among 260 Species of North American Birds

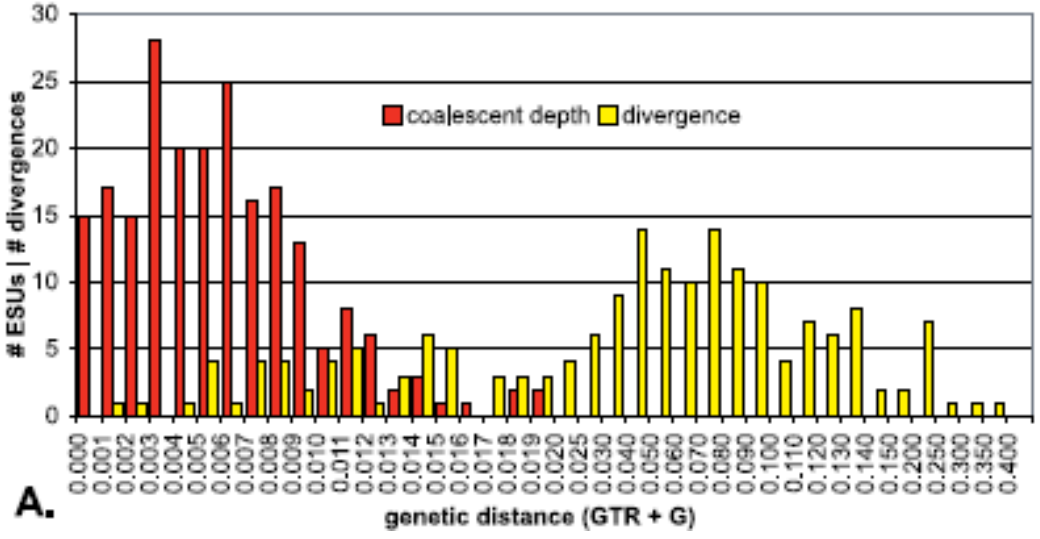
Barcoding gap



Cypraeidae

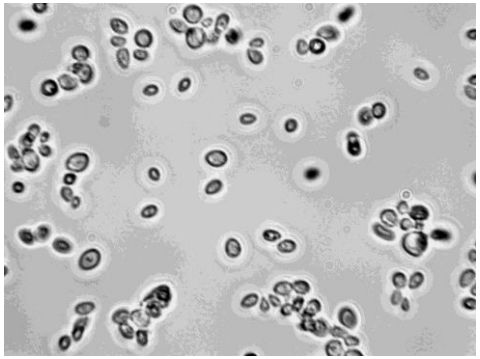
Meyer & Paulay 2005

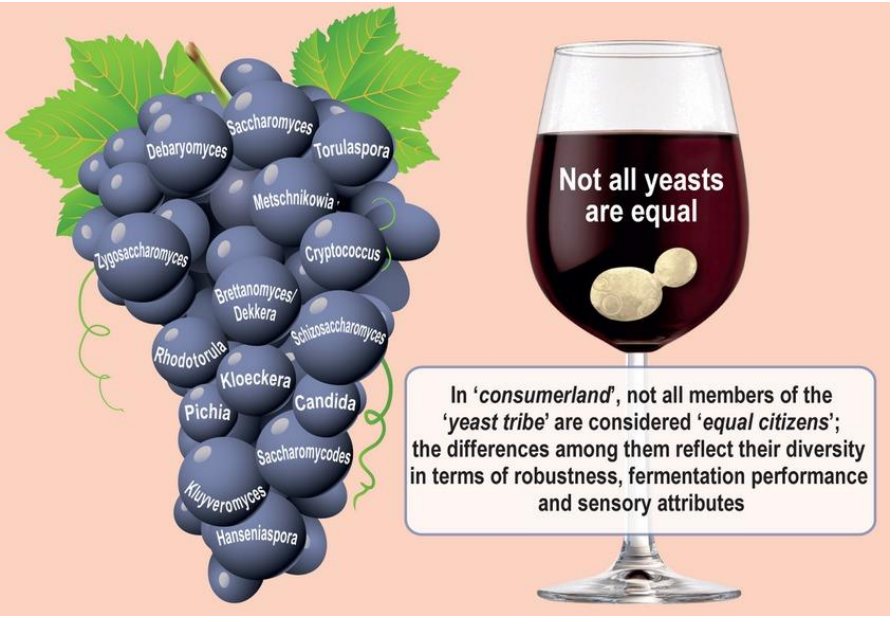


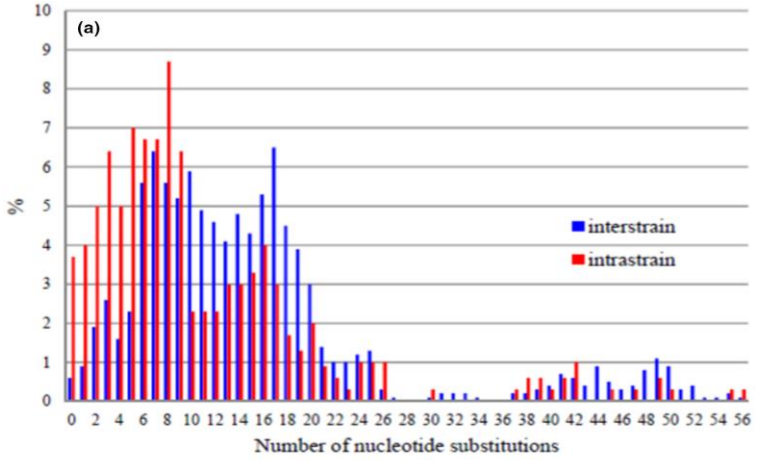


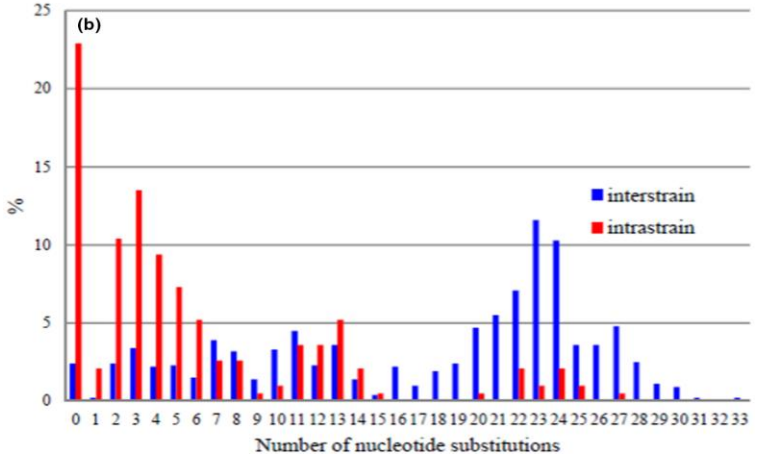
When barcoding fails: Genome chimerization (admixing) and reticulation obscure phylogenetic and taxonomic relationships

Metschnikowia





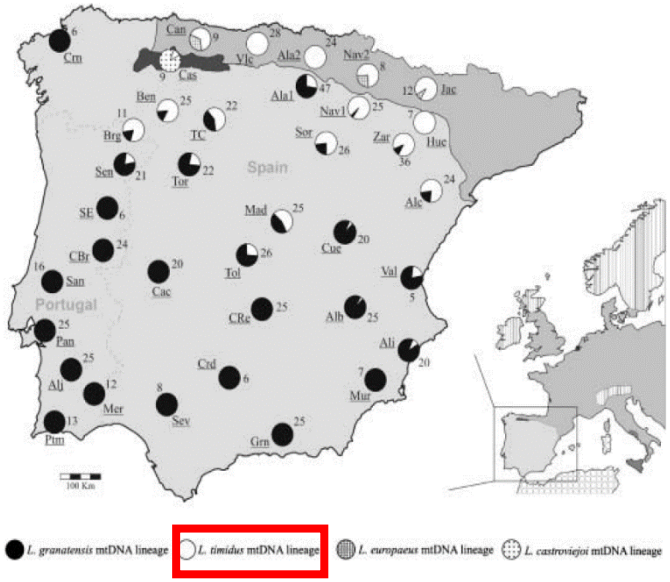




(a) D1/D2 domain sequences. (b) ITS1–5.8S–ITS2 sequences

Molecular Ecology Resources, First published: 20 January 2022, DOI: (10.1111/1755-0998.13586)

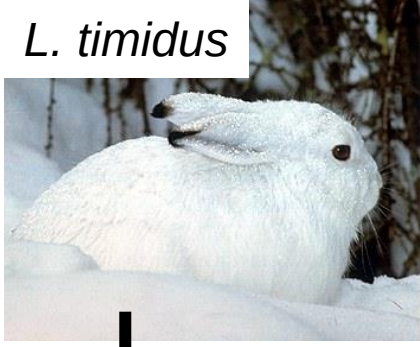
Hares in Spain and Portugal



L. timidus mtDNA

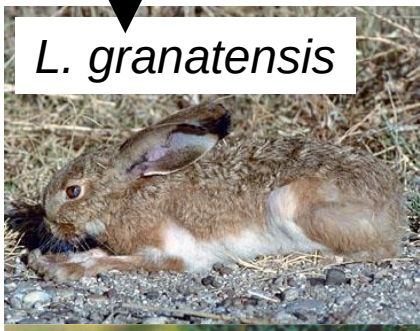
- *L. timidus* mtDNA in *Lepus granatensis, L. castroviejoi,* and *L. europaeus*
- *L. timidus* retreated from this region at the end of the last ice age
- Similar situation in bats, newts, fish...

L. timidus

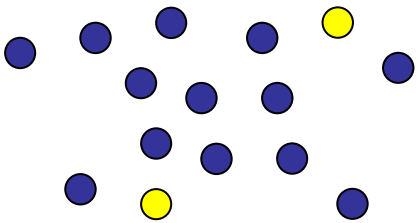


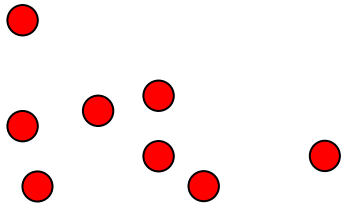
mtDNA

L. granatensis

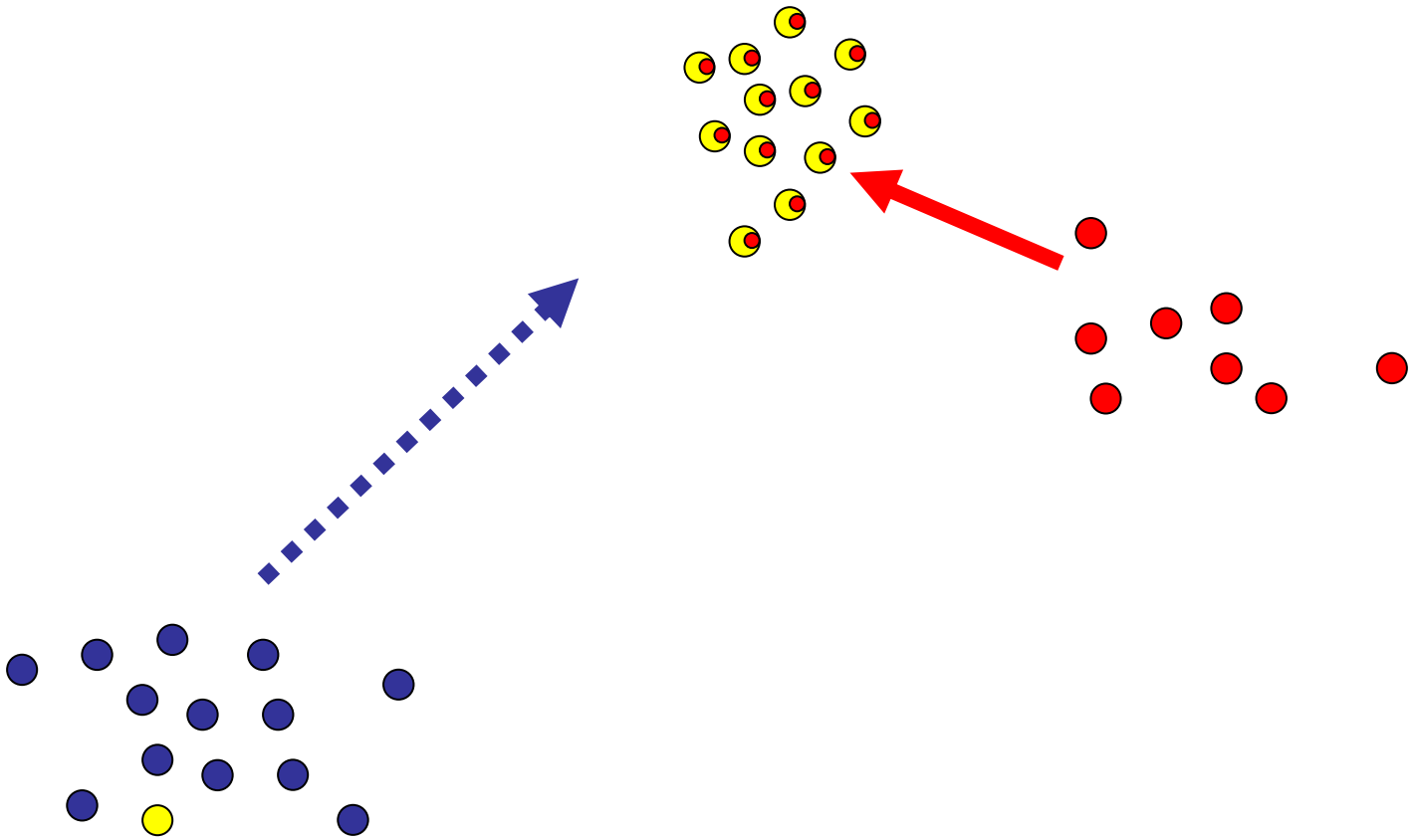


L. europaeus





Introgression from local into the invading species



SCIENCE • VOL. 266 • 18 NOVEMBER 1994

DNA Sequence from Cretaceous Period Bone Fragments

Scott R. Woodward,* Nathan J. Weyand, Mark Bunnell

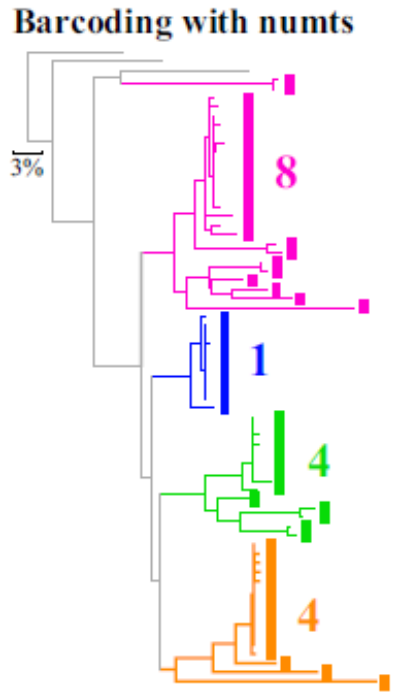
DNA was extracted from 80-million-year-old bone fragments found in strata of the Upper Cretaceous Blackhawk Formation in the roof of an underground coal mine in eastern Utah. This DNA was used as the template in a polymerase chain reaction that amplified and sequenced a portion of the gene encoding mitochondrial cytochrome b. These sequences differ from all other cytochrome b sequences investigated, including those in the GenBank and European Molecular Biology Laboratory databases. DNA isolated from these bone fragments and the resulting gene sequences demonstrate that small fragments of DNA may survive in bone for millions of years.

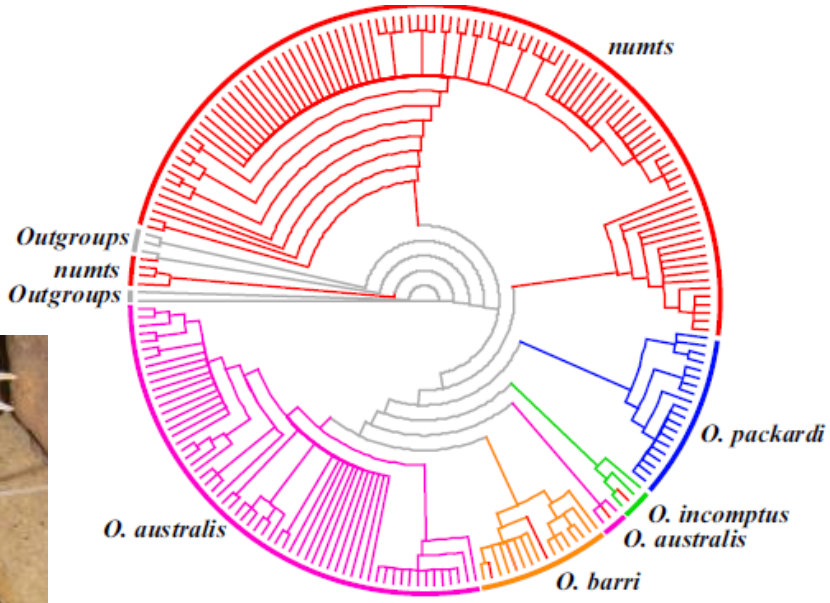
```
          15,627
Consensus CC CTT CTA TTA TCC ATT CTC ATT CTA TTC GTT ATT CCT GTA CTC CAC ACA TCC (C) AAA
2-37                   ..A ... ... ... ..C ... C.. ..C ... T.. ... ... .A. ...  /  ...
3-37      .. ... G.. ... ... ... T.. ... ... ... ... ... ... ... ..T ... ..G ..T  /  ...
4-37      .. ... .G. ... ... ... ... ... ... ... ... ... G.. ... .CC ... ... ... ..G  .  ...
31-44                          ... ... ... ..T ... ... ... ... ... ... ... ... ...  /  .G.
2-61      .. ... ... C.. .T. ..C ... ... ... ... ... .CA ... ... A.. ... ... .T. ...  /  ...
2-18                     ... ... ... ... ... ... T.. ... ... ... ... ... .G. ... ...  /  ...
20-61                          ..T ... ... ... ... ... ... ... ... ... ... .T. ...  /  ...
5-37      .. ... ... C.. T.. .A. T.. ..C ... ... ..A ... ..C A.. ..T ... ... .GT  .  ...
6-37                ... ..T ... ..T ..C ..G ... ..A T.. ... .C. ..T ... ... .GT  /  ...

Consensus CAA CAA AGC ATA ATA TTC CAC CCA TTG AGT CAA TTC CTA TCC TGA TTC TTA GTC CCC GAA
2-37      ... .G. .C. ... .C. ... ... ... ... T.. ... ... ... .A. ... ..T ... .
3-37      ... ... .C. ... ... ... ..A ... ... ... ..G .G. ... .T. ... A.. ... A.G G.. ...
4-37      ... ... ... ..C ... ... .G. ... ... ... ... ..T .A. ... ... A.. ... ... A.. ...
31-44     ..G ... ... ... ... ... ... ... ... ... C.. ... ... ... A.. ... ... C.. ... ..A ..C
2-61      ... ... ... ... ... ..T ... ... ... ... ... ... ... ... .T. ... ... ... ..G ...
2-18      ... ..T ... ... ... ... ... ... ... ..C ... ... ... ... .T. ... A.. ... A.. ...
20-61     T.. ... ... ..T ... ..T ... ... ... ... ... ... ... ... ... C.. ... ... ... ...
5-37      ... ... G.. ... ... ... GGT ... C.. G.C ... ... ..T ..T ..G C.. C... .
6-37      ... ... ... ... ... ... GG. ... C.A G.. ... ...

Consensus CCT TTT ACA CTC ACA TG
2-37
3-37      .TA ... ... ... ... ..
4-37      .G.
31-44     ... ... .
2-61      ... ... ... ..A ..C
2-18      .T. ... ... ... ...
20-61     ... ..
```

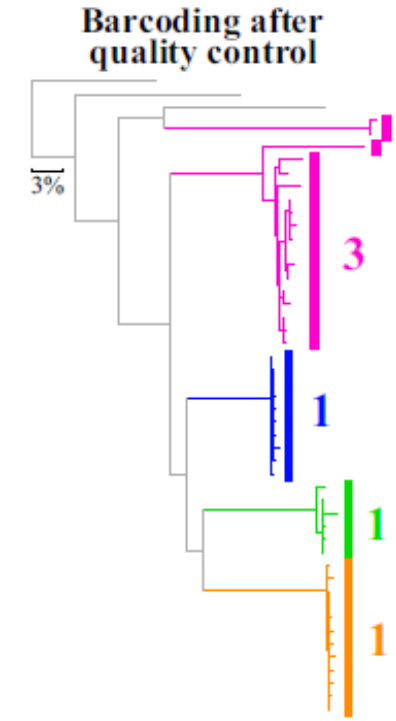
→ human numt

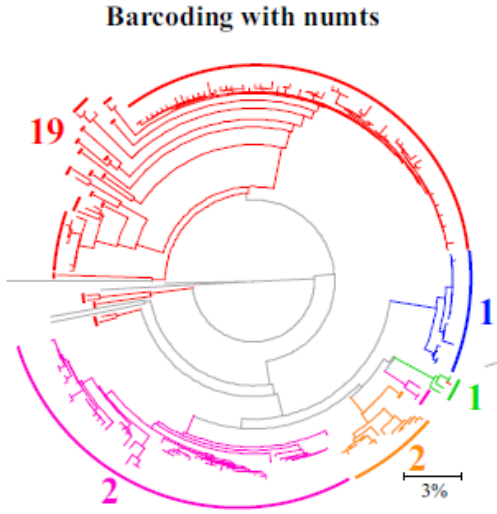
numts and DNA barcoding

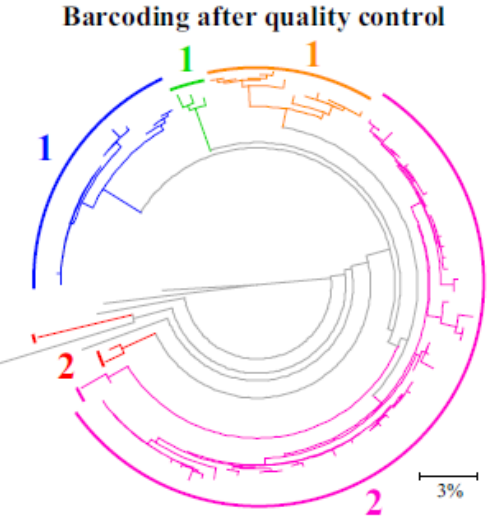
Song et al. 2008





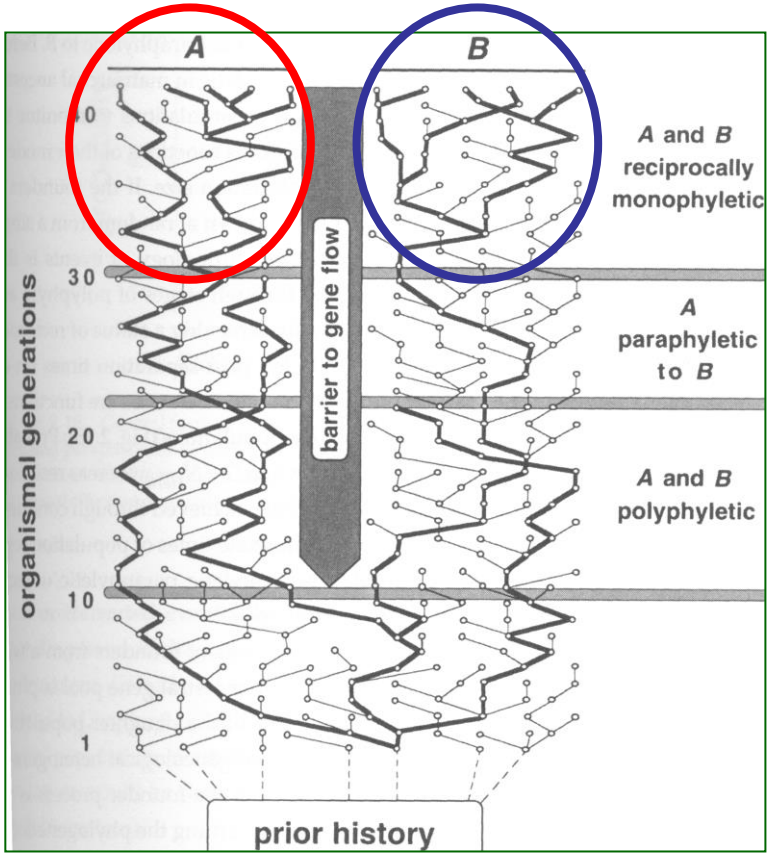






Retension of ancestral polymorfism

Lineage sorting
→ reciprocal monophyly



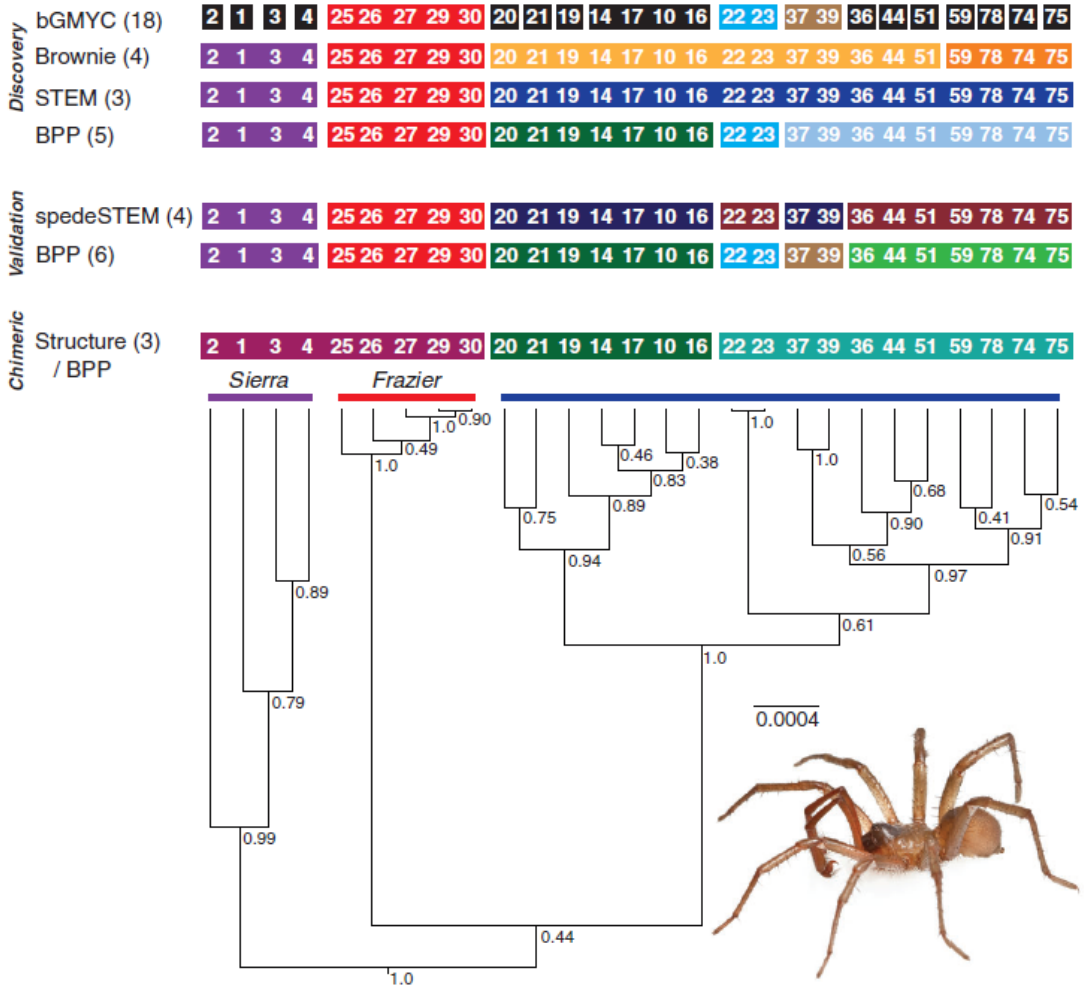
… more and more problems

Table 1 Outline of potential problems, consequences and solutions for the 'seven deadly sins' of DNA barcoding, as presented here

Problem	Consequence	Solution
Failure to test clear hypotheses	Choice of inappropriate or suboptimal analytical method due to confusion as to the objectives of the study	Explicitly state each hypothesis, and for each distinct aspect of the study present separate headings in methods and results sections
Inadequate a priori identification of specimens	Conflicting identifications made by different labs can compromise the effectiveness of reference libraries that are ultimately used as a resource for scientific or regulatory purposes	Present a bibliography of references, as well as the distinguishing morphological characters used in the identification process. Follow recommendations outlined by Steinke & Hanner (2011)
The use of the term 'species identification'	Confusion between identification of individuals, and delimitation/discovery of species	To clarify objectives, use the term 'specimen identification' or 'species discovery' where appropriate
Inappropriate use of neighbour-joining trees	(a) Relying on strict monophyly for identification can reduce the apparent effectiveness of DNA barcoding as an identification tool. This can be due to either mtDNA paraphyly or misidentification of specimens. (b) For biodiversity assessment and species discovery, NJ trees cannot estimate the number of species independently with respect to the taxonomic names	(a) Alternative criteria such as 'best close match' are readily available, and have higher rates of identification success. This method can be implemented using the free software packages TaxonDNA (Meier *et al.* 2006) or Spider (Brown *et al.* 2012). (b) Estimate species richness using ABGD (Puillandre *et al.* 2012), GMYC (Monaghan *et al.* 2009) or BOLD's BIN system (http://v3.boldsystems.org)
Inappropriate use of bootstrap resampling	For specimen identification purposes, bootstrap resampling can further reduce the already low identification success rates associated with NJ trees	Only use bootstrapping where appropriate: e.g. as part of a species delimitation process on preestimated groups
Inappropriate use of fixed distance thresholds	For specimen identification purposes, a generic threshold which is set too low or high can reduce or bias identification error rates	Thresholds can now be optimized for specific data sets using the method of Virgilio *et al.* (2012), or with software such as ABGD (Puillandre *et al.* 2012) and Spider (Brown *et al.* 2012)
Incorrectly interpreting the barcoding gap	Overlapping distributions of intra-/interspecific distances do not necessarily mean that barcodes perform poorly for identification	For specimen identification studies, dotplots of intra-/interspecific distances are a better way to illustrate the barcoding gap (e.g. Robinson *et al.* 2009)

Species delimitation

- **Distances**
  - ABGD
  - (Automatic Barcode Gap Discovery)
- **Phylogenies (trees)**
  - „minimal phylogenetic units“ (OTUs)
  - BPP (Bayesian Phylogenetics and Phylogeography)
  - GMYC, PTP…



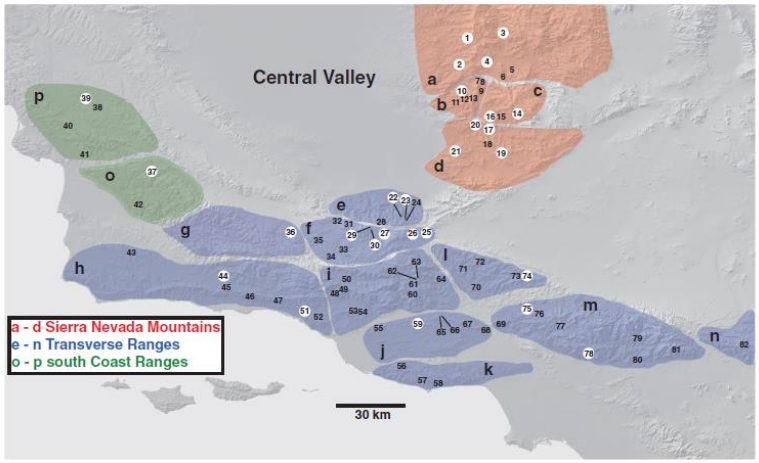


FIGURE 6. Summarized results from all species delimitation analyses, represented on *BEAST tree resulting from analysis of all sampling locales. Colors correspond to recovered groupings, with each partition represented by a unique color. Colored bars above phylogeny represent hypothesized species groupings based on multiple analyses. Insert image of an adult male *Aliatypus starretti*, sp. nov. (Kern Co., Poso Flat Road).

Molecular identification of individuals

- microsatellites

Repeats

- more than half of our genome
- Interspersed (transposony)
- Tandem (mini- a mikrosatellites)
- Minisatellites (longer motifs: 10 - 100 nucleotides)
- Mikrosatellites (1 - 6 nucleotides motifs)
- Human - 700 000 msats, 3 % of genome

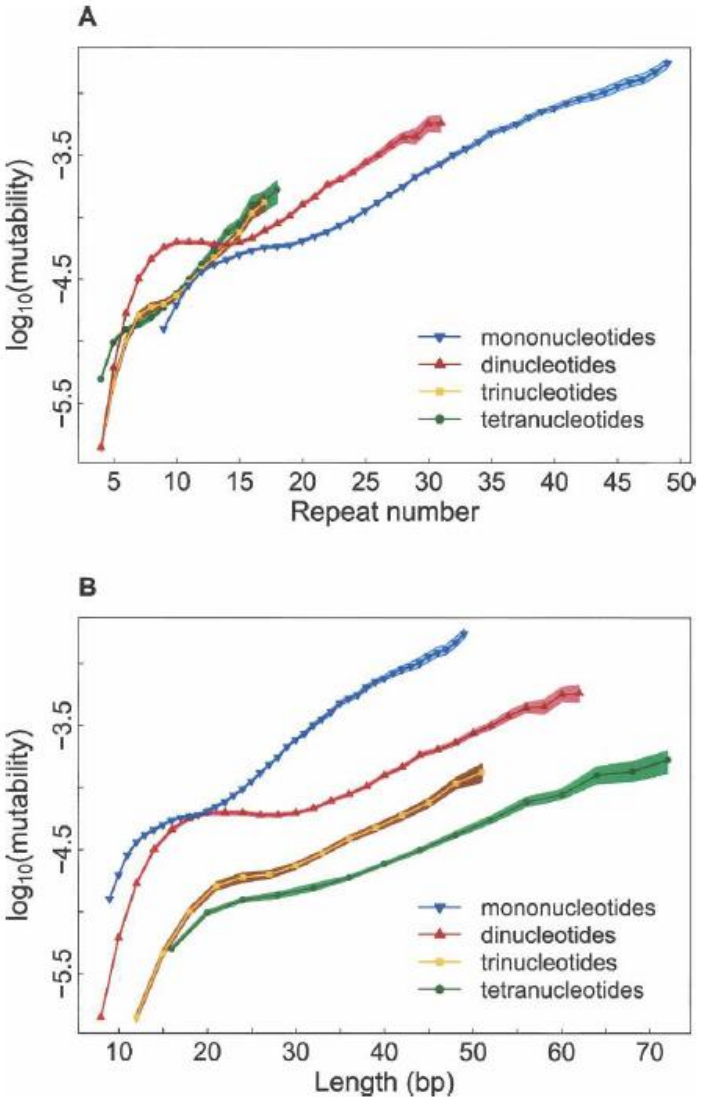
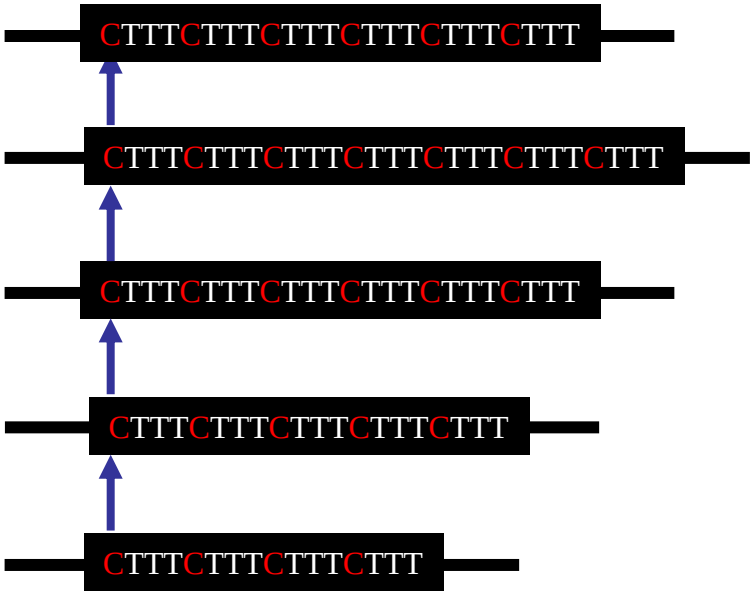
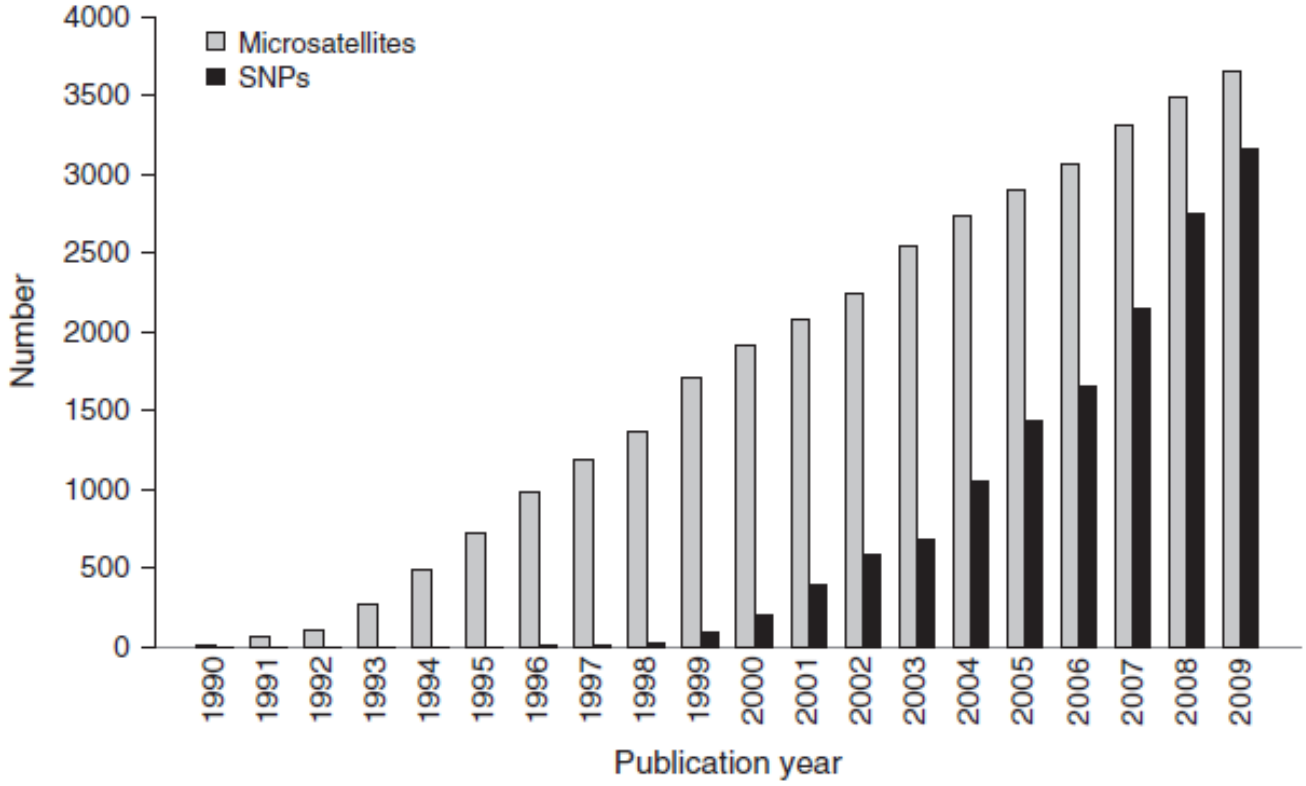


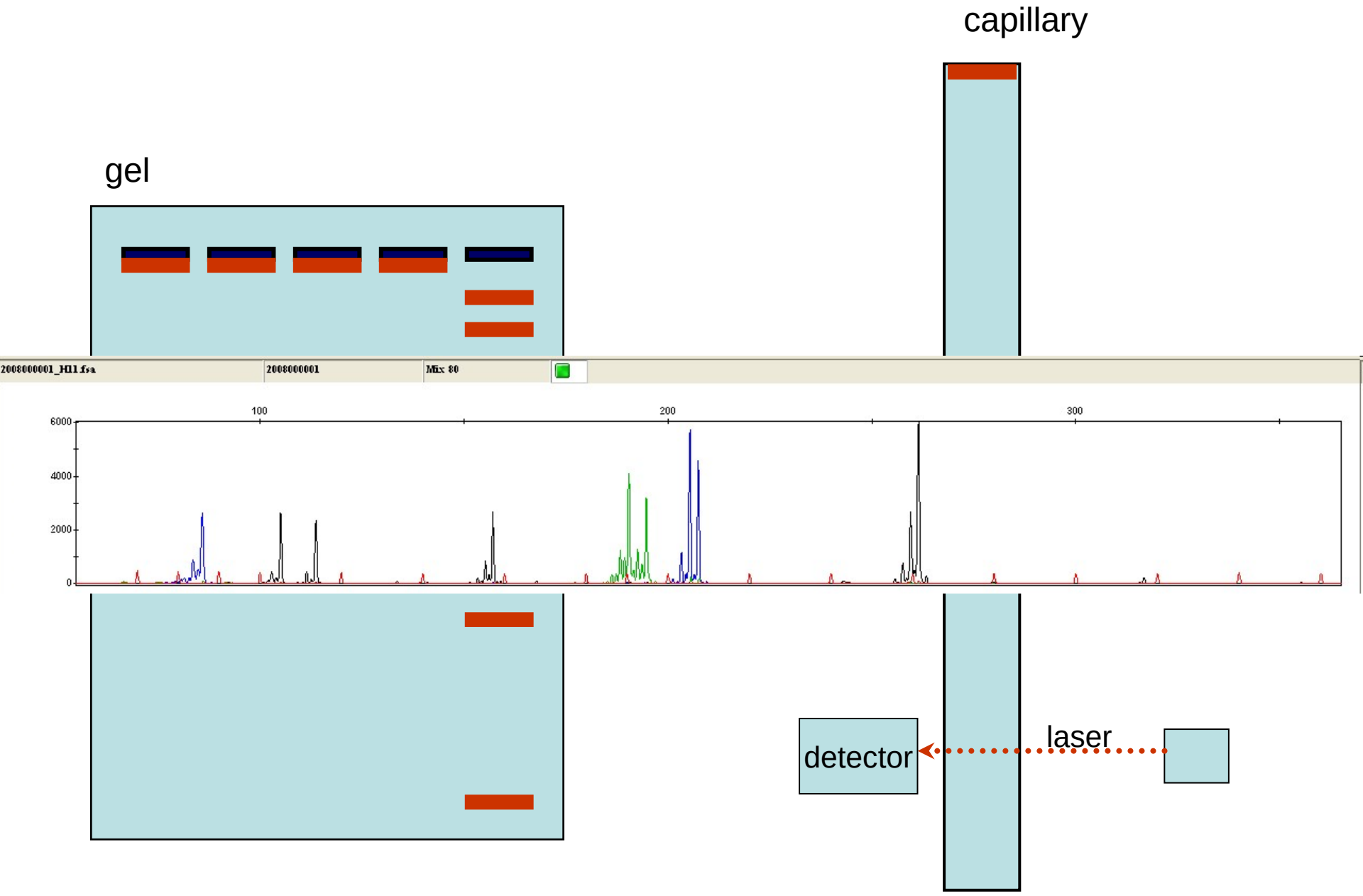
Figure 1. Dependence of microsatellite mutability on repeat number (*A*) and length (*B*). Mutability is per locus per generation. The bands around the curves indicate the 2.5th and 97.5th percentiles of empirical distributions obtained through a resampling procedure (see Methods). Only points with at least 30 microsatellites are plotted.

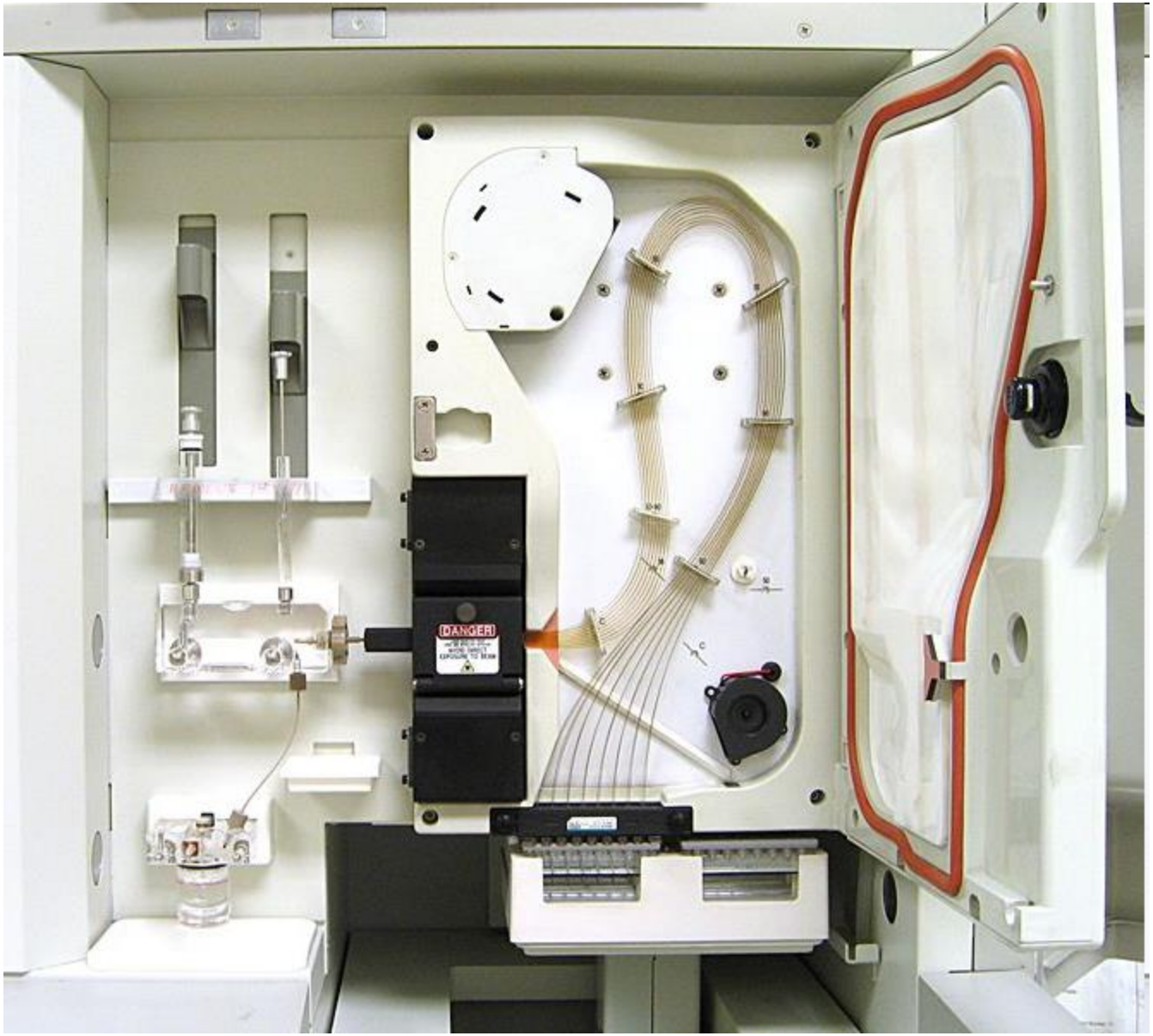
- Mutation rate $10^{-3(2)}$ 10^{-7}
- Male germline > female germline
- Slippage

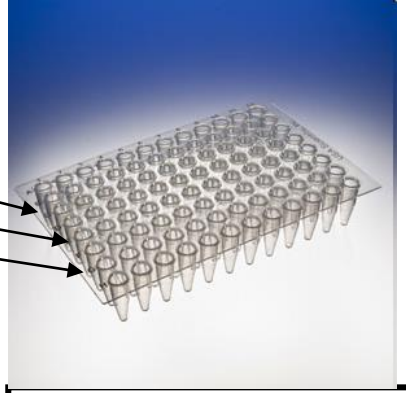


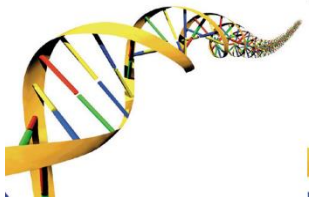
- Simple Mendelian inheritance
- Highly variable
- Paternity, population structure…


Microsatellites
SNPs
Number
4000
3500
3000
2500
2000
1500
1000
500
0
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
Publication year

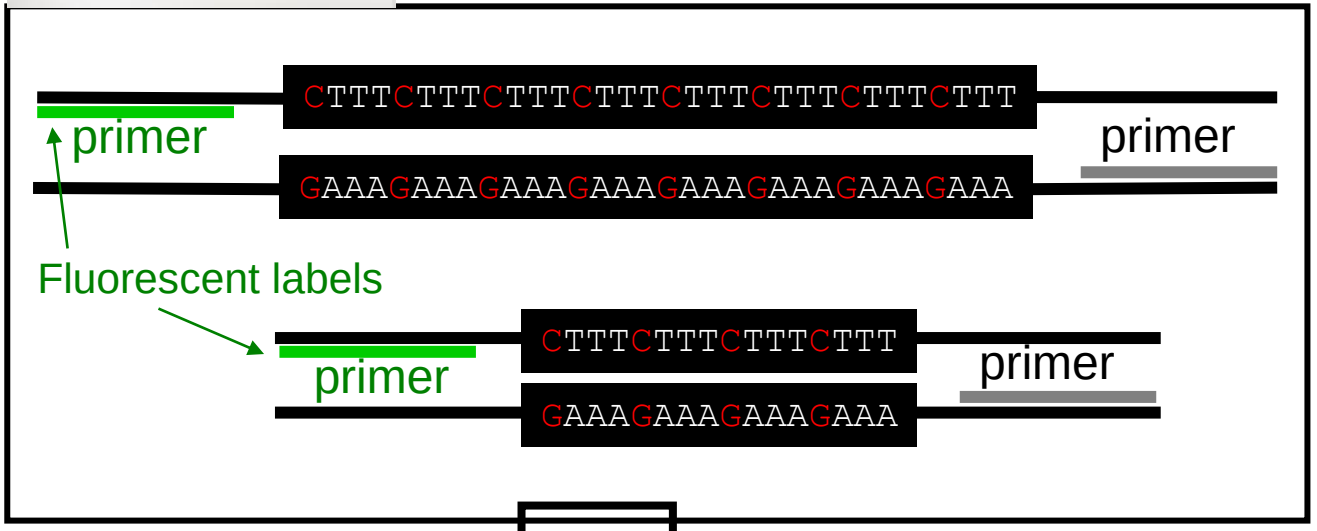

capillary
gel
2008000001_H11.fsa
2008000001
Mix 80
100
200
300
6000
4000
2000
0
detector
laser

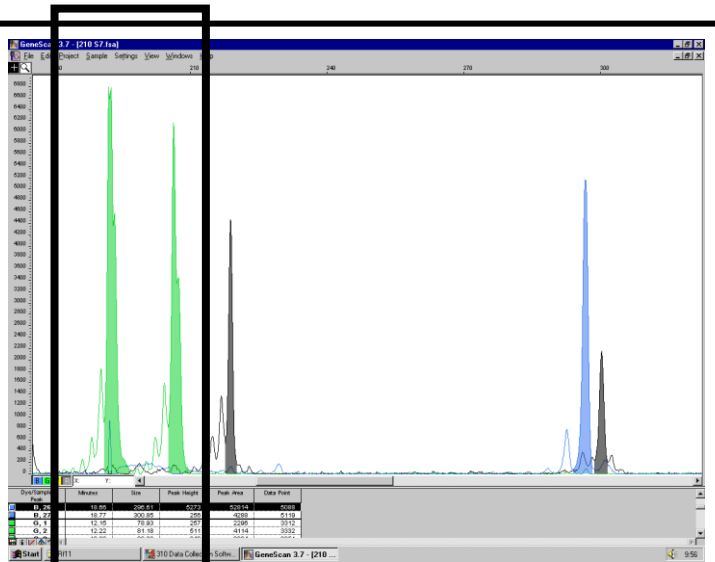

DANGER!





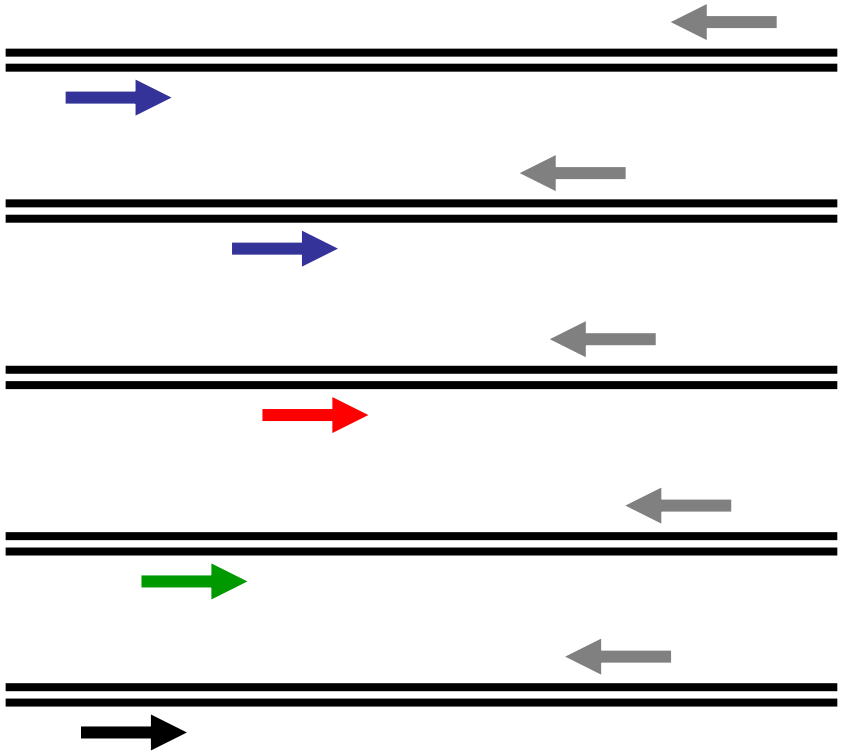

eppendorf
Mastercycler
epgradient S

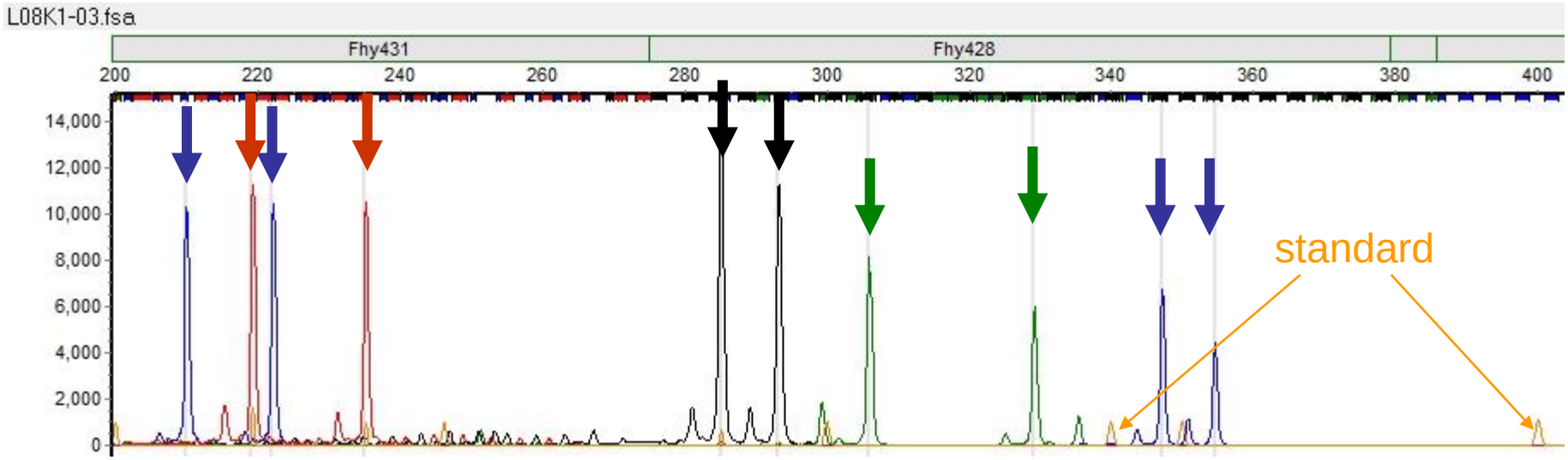

CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT
primer
primer
GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA
Fluorescent labels
CTTTCTTTCTTTCTTT
primer
primer
GAAAGAAAGAAAGAAA



Multiplex

More loci in one reaction





Primers for new loci: traditional approaches

genome library, hybridization with probe, sequencing

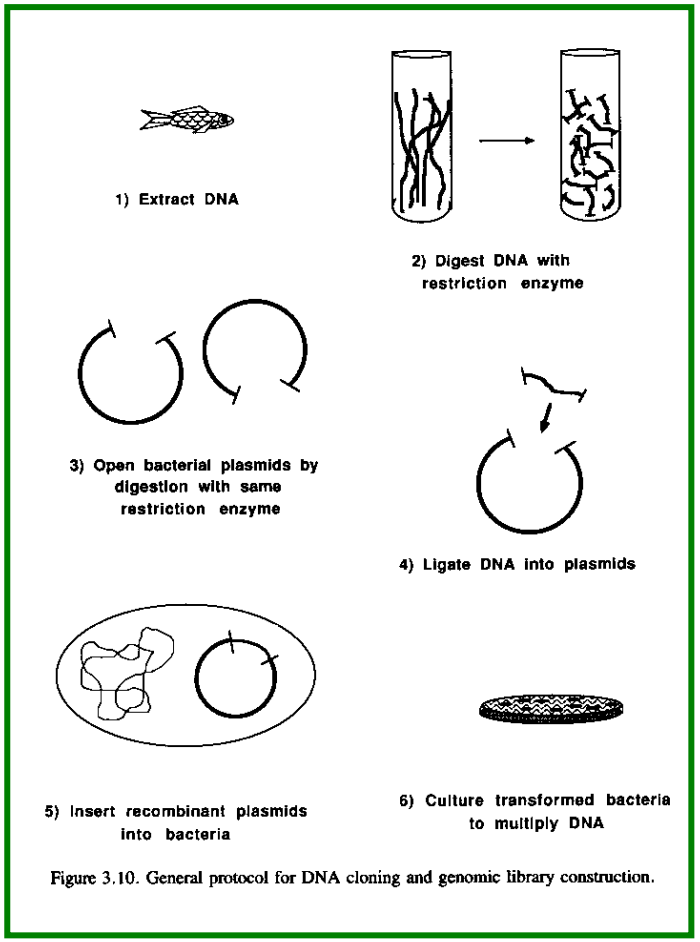


Figure 3.10. General protocol for DNA cloning and genomic library construction.

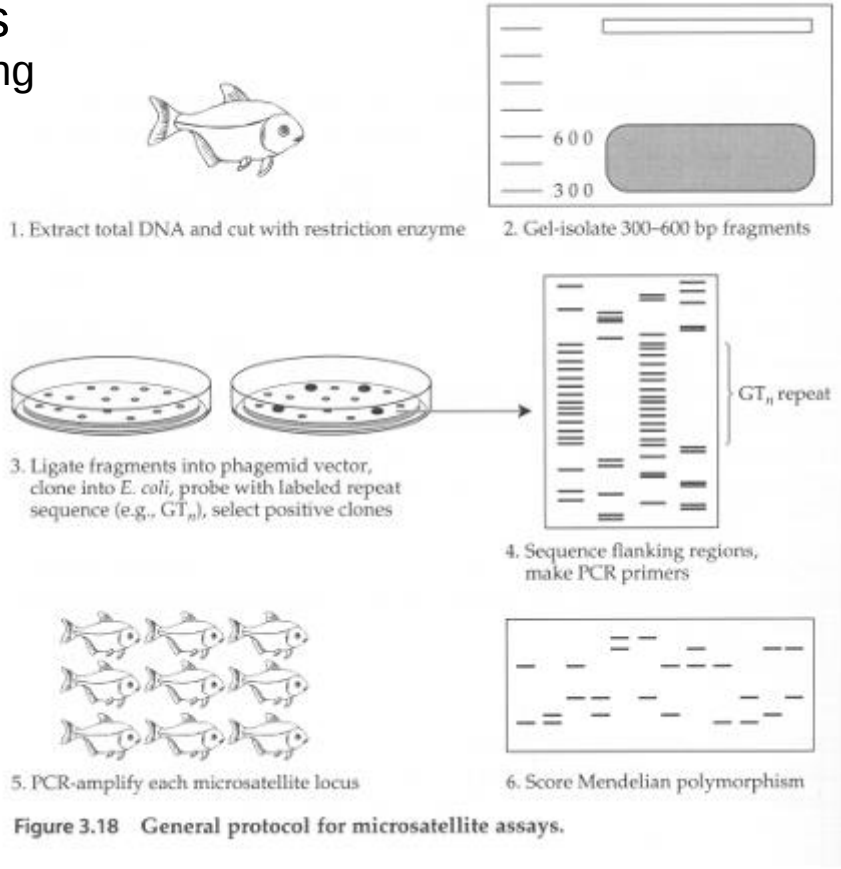
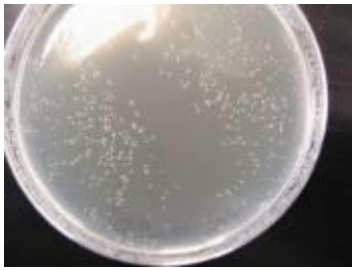
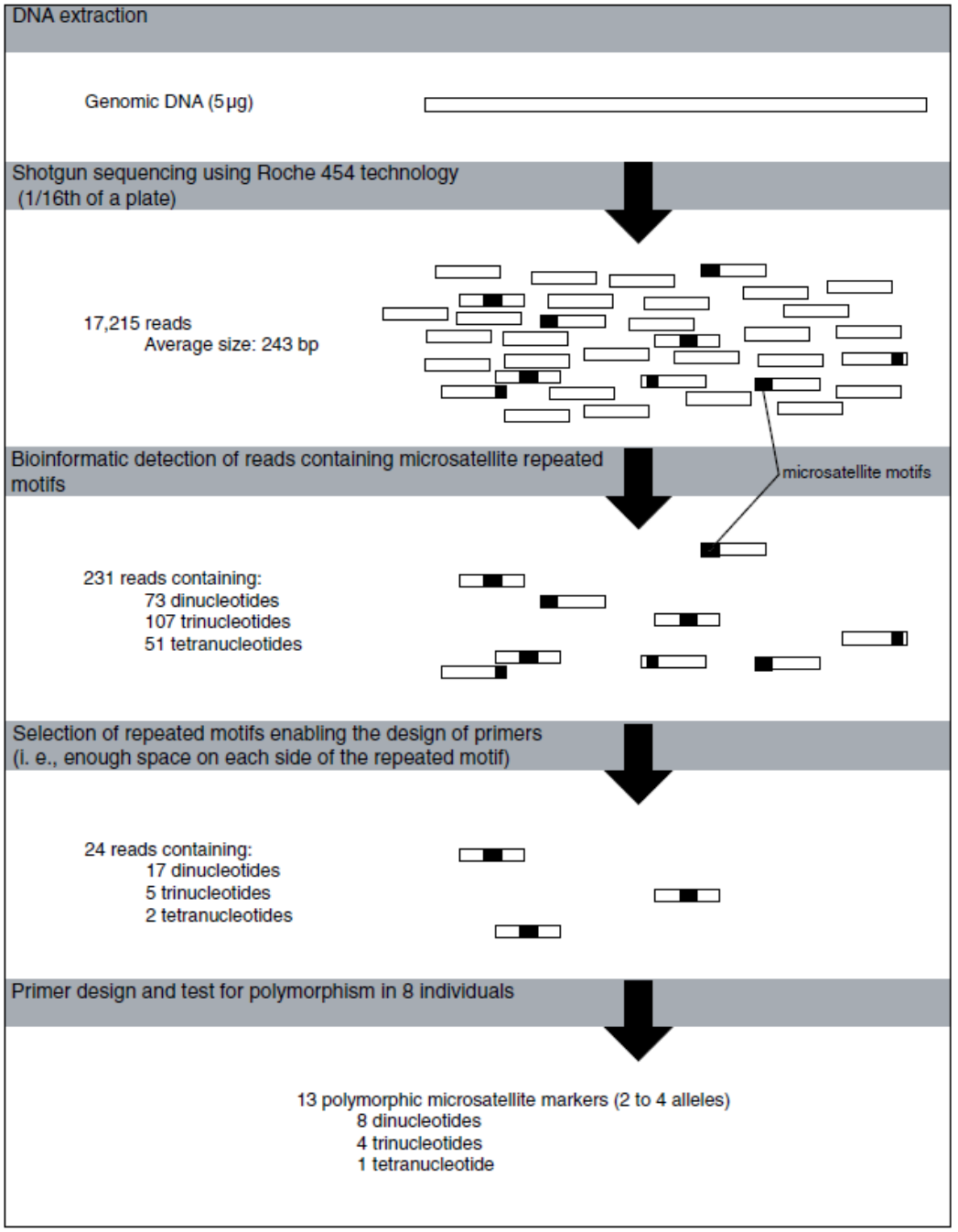


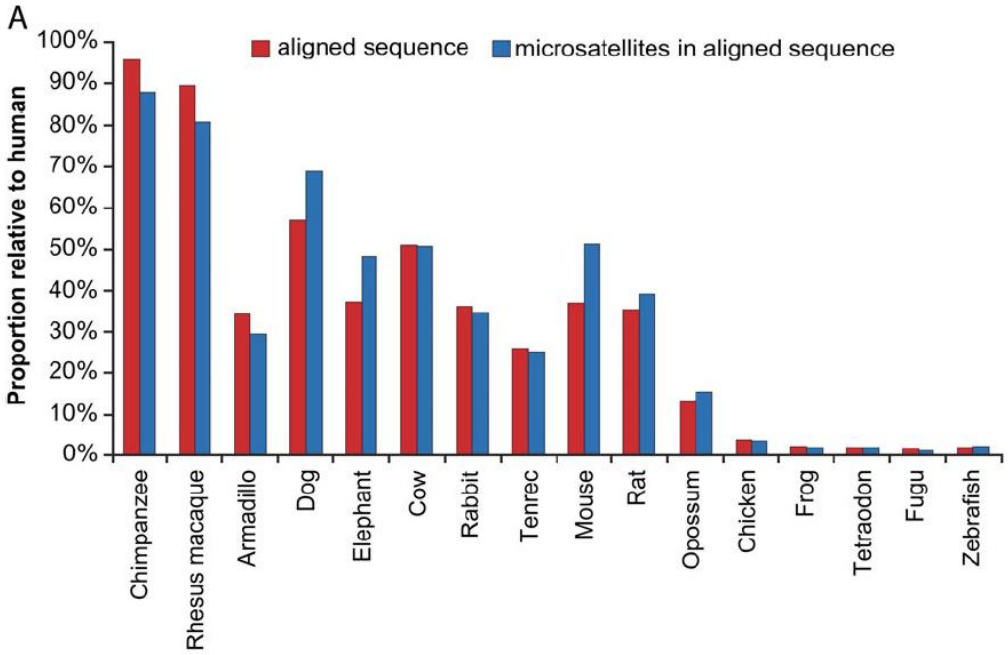
Figure 3.18 General protocol for microsatellite assays.



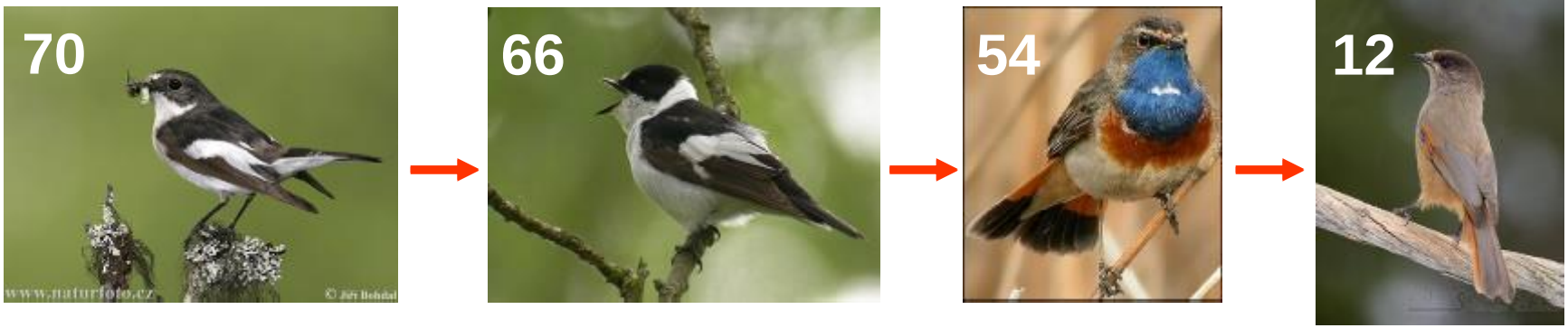
NGS


DNA extraction
Genomic DNA (5µg)
Shotgun sequencing using Roche 454 technology
(1/16th of a plate)
17,215 reads
Average size: 243 bp
Bioinformatic detection of reads containing microsatellite repeated motifs
microsatellite motifs
231 reads containing:
73 dinucleotides
107 trinucleotides
51 tetranucleotides
Selection of repeated motifs enabling the design of primers
(i. e., enough space on each side of the repeated motif)
24 reads containing:
17 dinucleotides
5 trinucleotides
2 tetranucleotides
Primer design and test for polymorphism in 8 individuals
13 polymorphic microsatellite markers (2 to 4 alleles)
8 dinucleotides
4 trinucleotides
1 tetranucleotide

Cross-species amplifications



Some msats are surprisingly conserved, but…



Pyrenean brown bears

Taberlet et al. 1997

- Faeces, hairs
- 24 msats
- 4 males a 1 female
- Otters, faeces
- →Molecular scatology

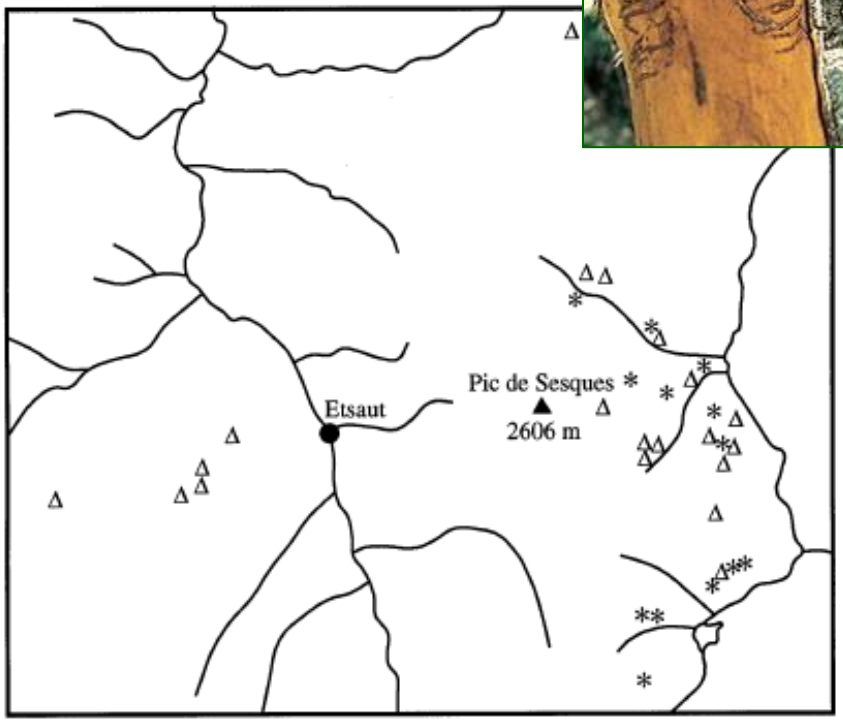
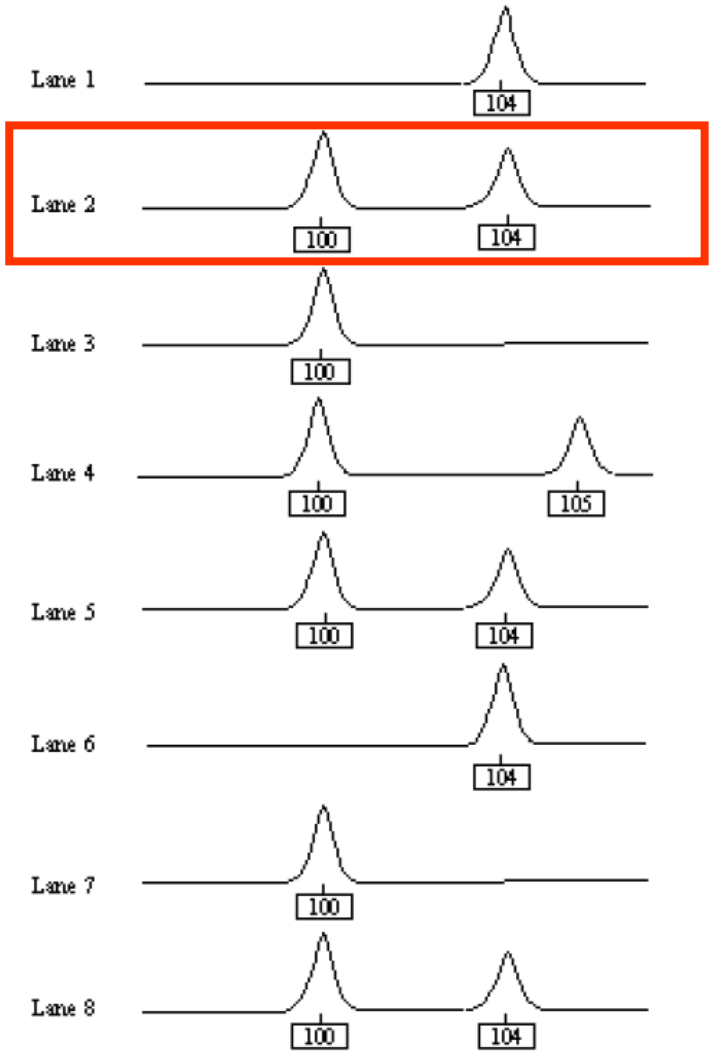


Fig. 3 Home range of two Pyrenean brown bears obtained by noninvasive genetic sampling and genotyping.



allelic dropout

allelic dropout

allelic dropout and false allele.

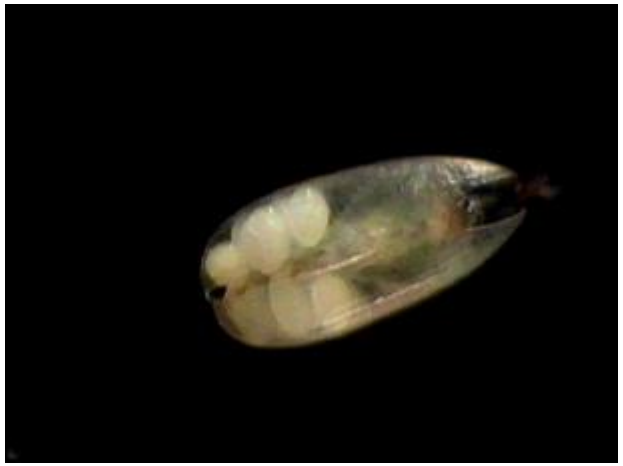
allelic dropout

allelic dropout

→ multiple tube approach

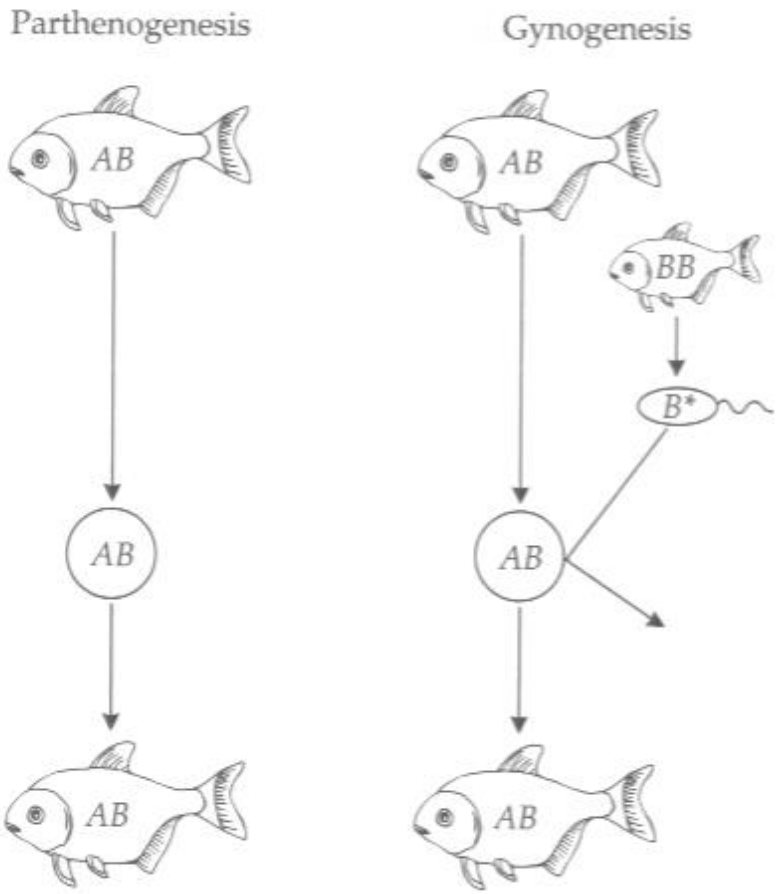
- Clonality
- Genetic elimination (genome loss)
- Genetic chimeras

- Rotifera – Bdelloidea
- Ostracoda (*Darwinula*)
- Partenogenetic clones

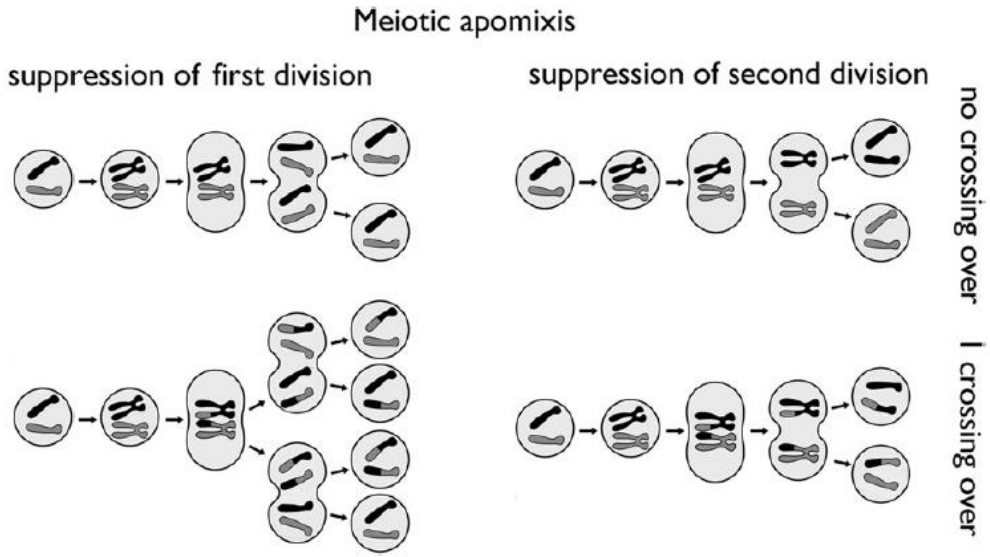


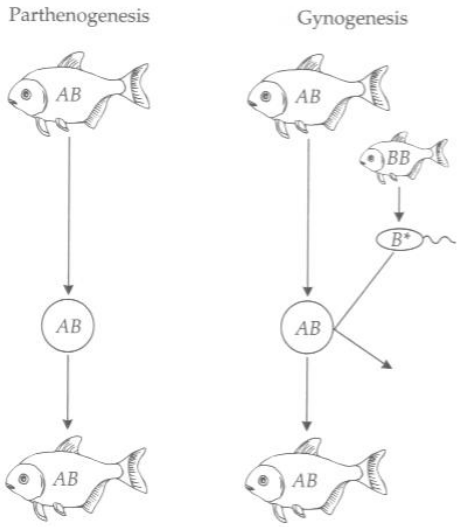
Darwinula stevensoni

Gynogenesis

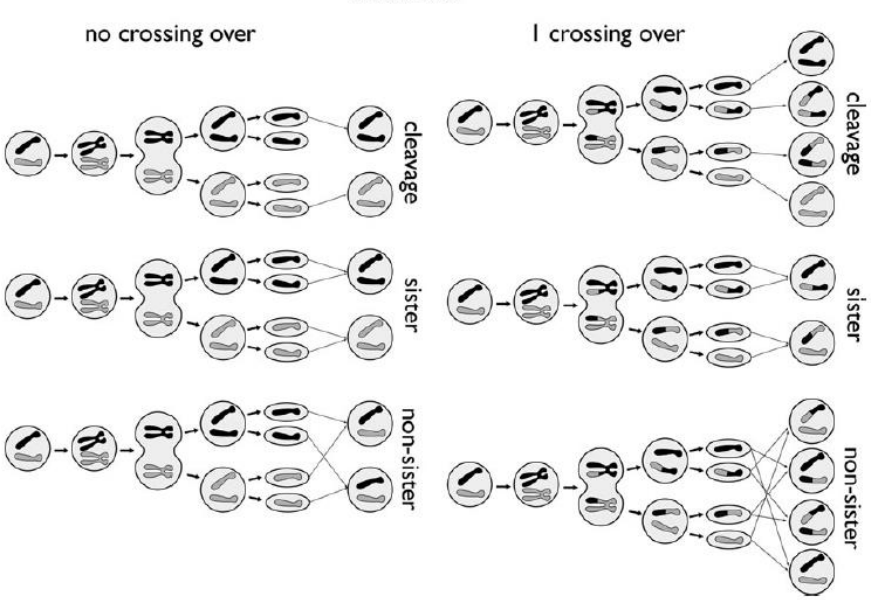


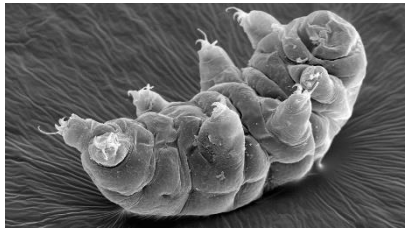
Ambystoma – genome of 4 species





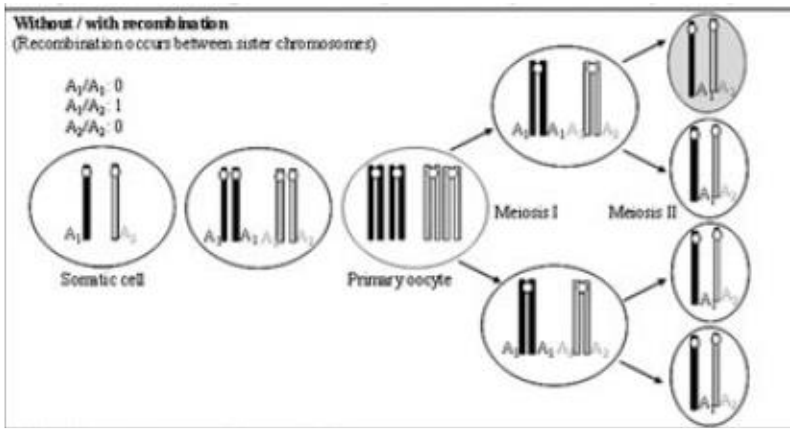
Meiotic apomixis with suppression of the first or second division, with or without recombination





Automixis with fusion of cleavage nuclei, sister nuclei, or nonsister nuclei, with or without recombination

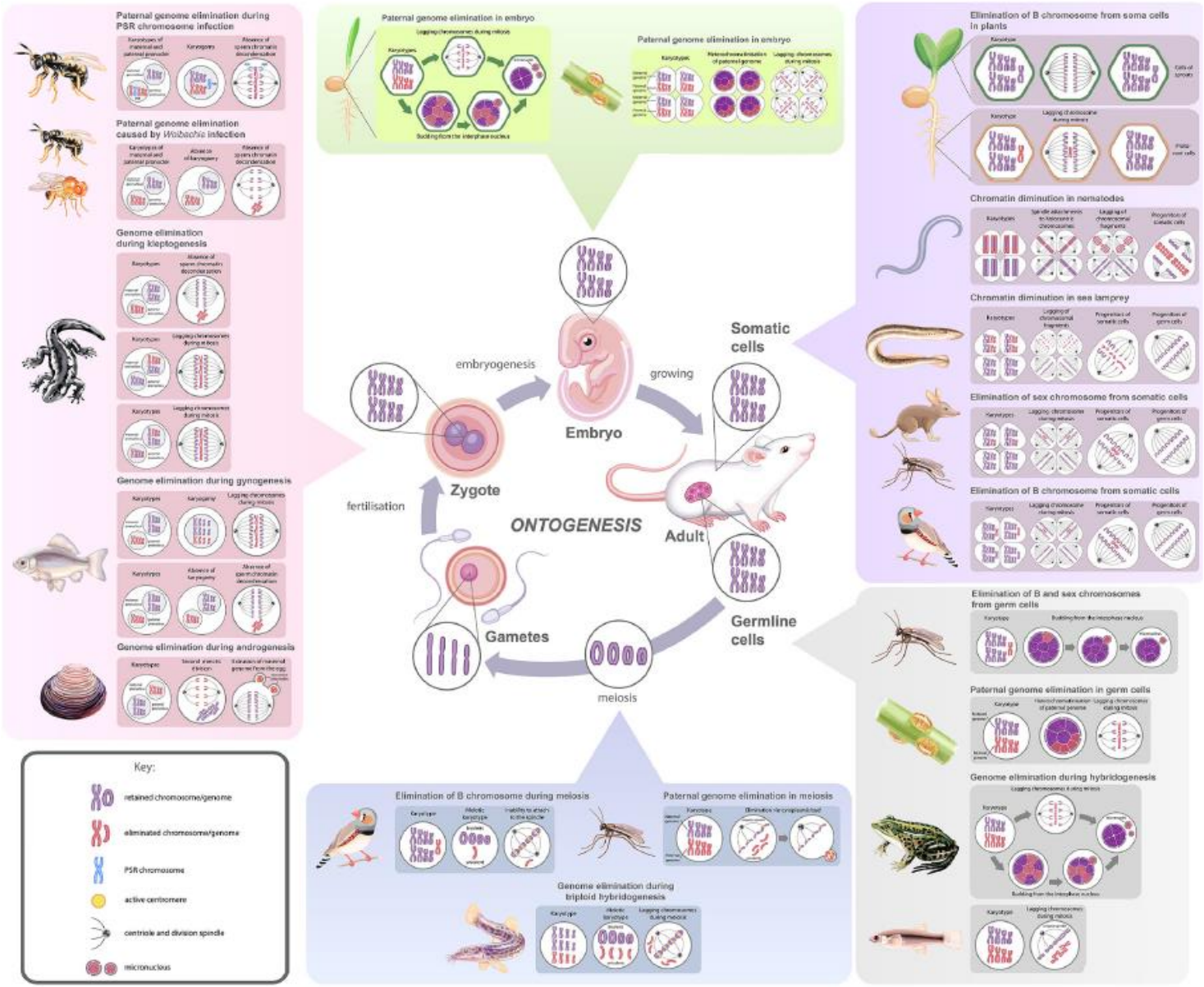
Premeiotic chromosome doubling



BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2022), **97**, pp. 195–216.
doi: 10.1111/brv.12796

Delete and survive: strategies of programmed genetic material elimination in eukaryotes

Dmitrij Dedukh and Alla Krasikova*



Pelophylax hybridogenesis

P. lessonae LL

P. esculentus LR

P. ridibundus RR

P. esculentus LR

Sympatry with P. lessonae LL

LR

↓

~~L~~R

↓

RR

↓

R

L → ↓

LR

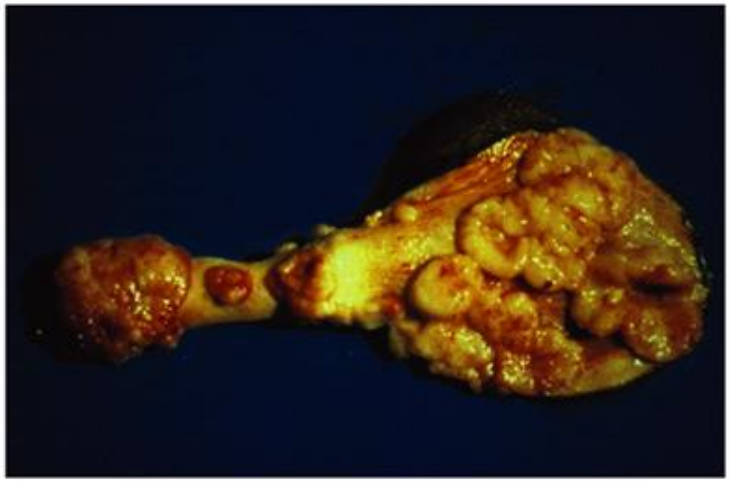
Genetic chimeras

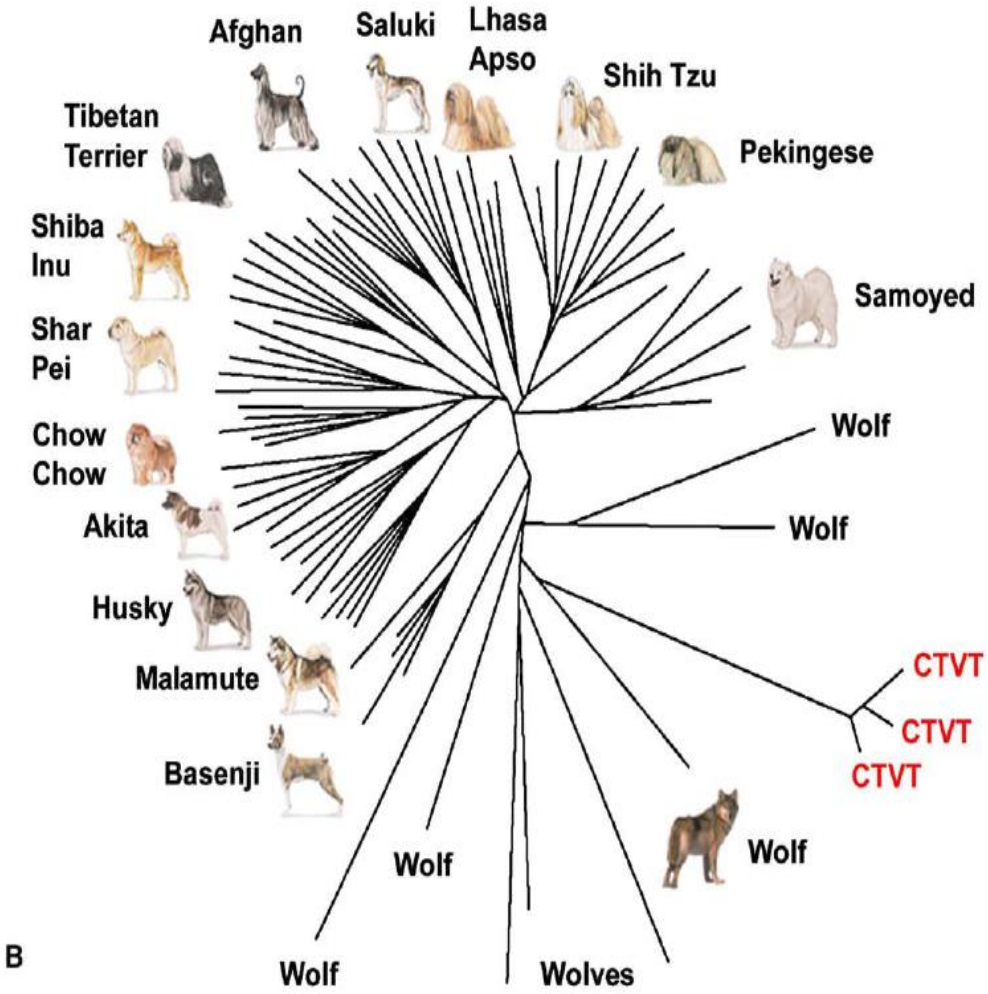
Deep sea fish

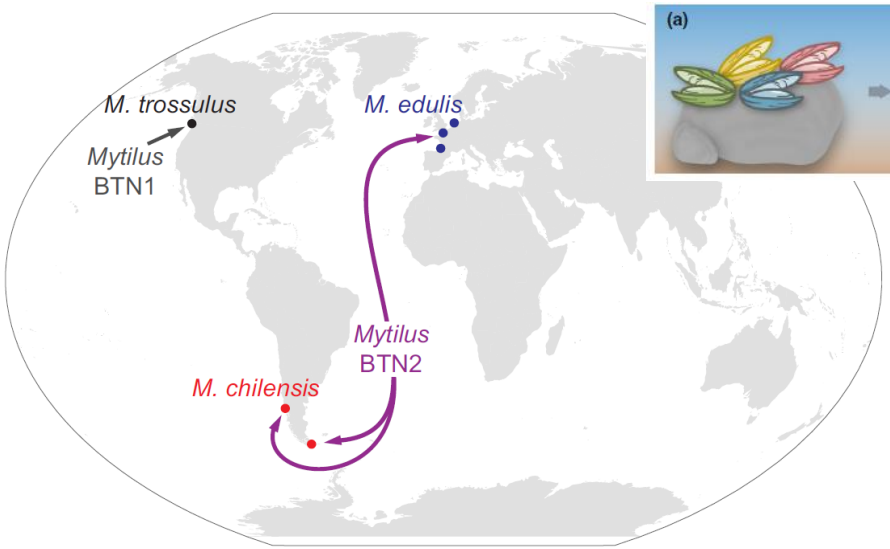
- marmosets and tamarins
- *Callithrix jacchus* (also *Saguinus*)
- Dizygotic twins
- Hematopoietic chimeras

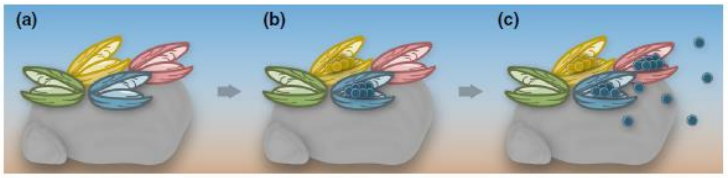
Canine transmissible venereal sarcoma (CTVS)

Sarcophilus harrisii









B

mtCOI alleles
Normal allele
Cancer allele

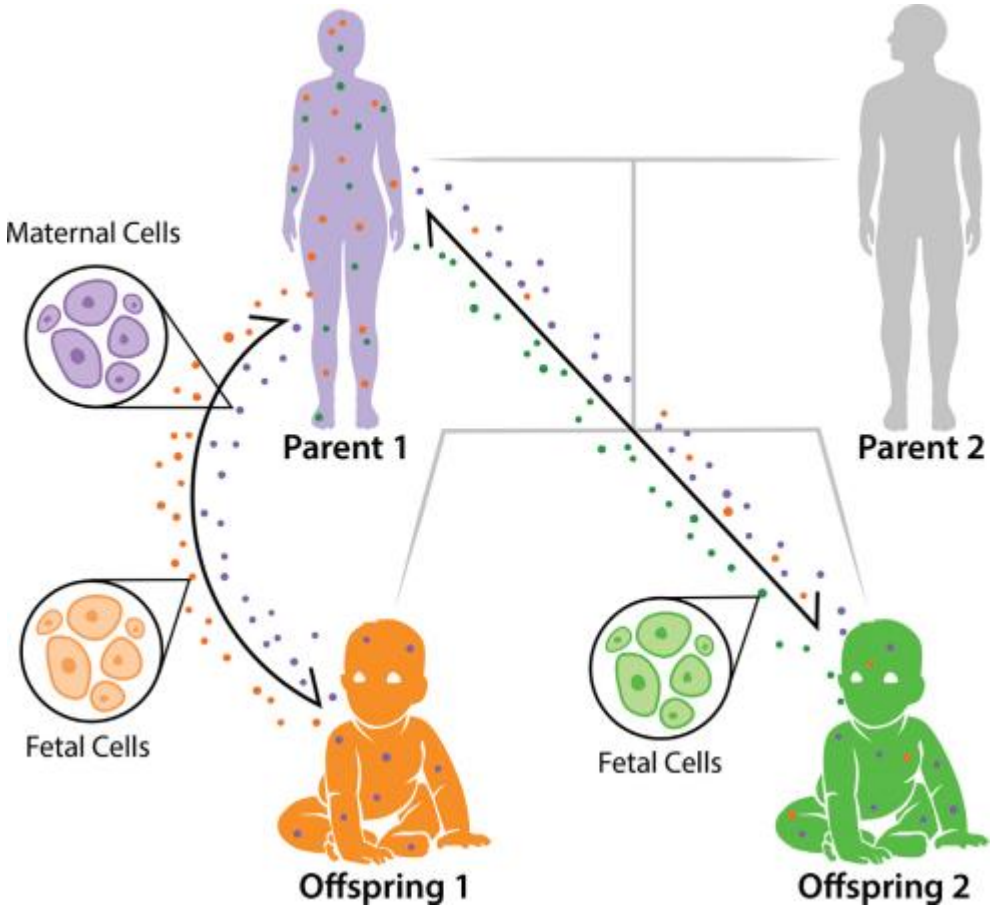
M. trossulus
M. edulis
M. chilensis

BA30-B'''
AR7-B''
HO8-B''
Mch41-B
Mch42-B'
Mch23-B
Mytilus BTN2 B allele
62mc10
MW59-Y
CB11-2B
CB22-2A
CB10-U
MW81-U
CB13-U
Mytilus BTN1 U allele
CB13-W
MW81-X
CB10-V
Castro52-K
Mch23-E
Mch3-A
Castro26-I
Mch41-G
Mch5-C
Castro5-C
Castro52-Q
Castro26-Q
Mytilus BTN2 Q allele
Mch7-D
Mch23-F
Mch42-F
Calbuco2-M
HO9-O
HO9-N
BA12-P
AR5-R
AR7-R
HO8-T
BA30-S

65 89 91 75 53 100 81 53 99 64 89 100 91 70 90 74 70

0.04

Microchimerism



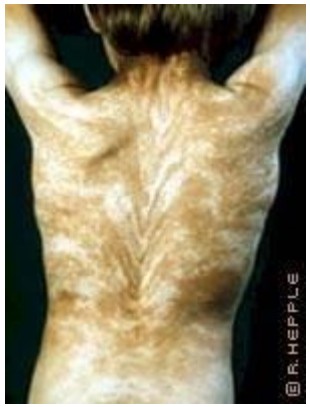
Lydia Fairchild

Nikita from Irkutska

Chang a Eng
Siam twins
born 1811

Fusion of embryos
(heteropaternal superfecundation 2.4%)

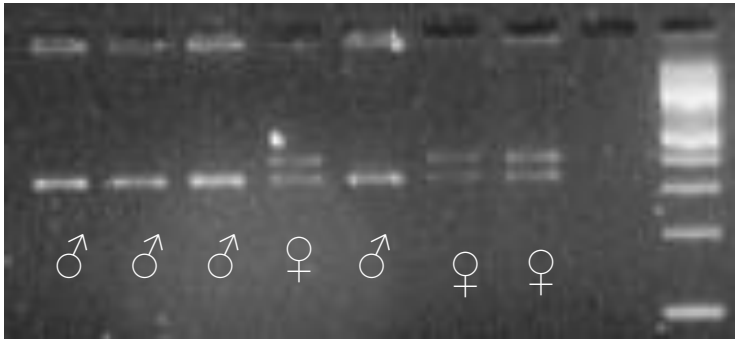
→ genotypes of ovaries and somatic tissues may differ

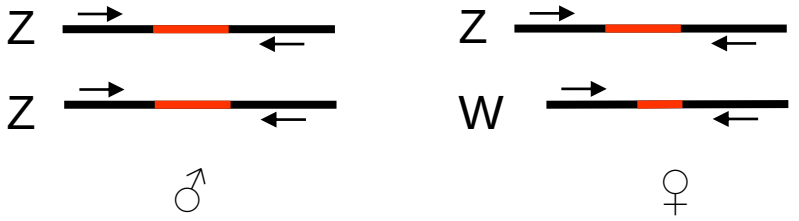


Blaschko's lines

Lydia Fairchild

Molecular sexing – birds

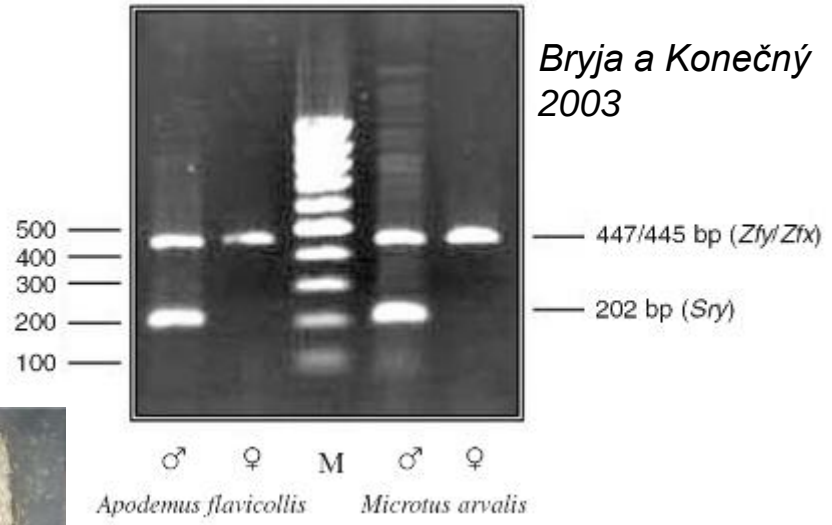




- CHD1W a CHD1Z (Griffith et al. 1998)
- ATP5A1Z a ATP5A1W (Bantock et al. 2008)
- Genes on sex chromosomes
- Primers amplify introns of both genes
- Introns may differ in length

Molecular sexing - mammals

- Y (*Sry*)
  (duplex PCR with an autosomal or X gene)
- *Sry* DNA-binding motif (HMG box)



Bryja a Konečný 2003

- *Microtus cabrerae*
  Sry on Chr X

  Ellobius, Tokudaia
  Sry is missing

  Dicrostonyx torquatus, Mus minutoides
  feminizing X* → X*Y females

- *Microtus oregoni*
  females XO somatic cells, XX germ cells, males XY somatic cells, 0Y germ cells

M. cabrerae

Ellobius

D. torquatus

Tokudaia osimensis

qPCR

Methods in Ecology and Evolution

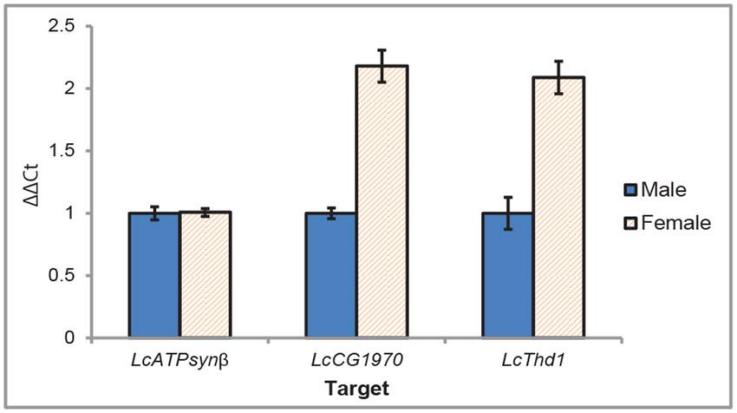
Methods in Ecology and Evolution 2017, **8**, 902–906 doi: 10.1111/2041-210X.12714

REVIEW

Molecular sexing applicable in 4000 species of lizards and snakes? From dream to real possibility

Michail Rovatsos and Lukáš Kratochvíl*

Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, 12844 Prague, Czech Republic



Gynandromorphs

Double fertilization of binucleate eggs

Loss of the W

ZW / ZZ