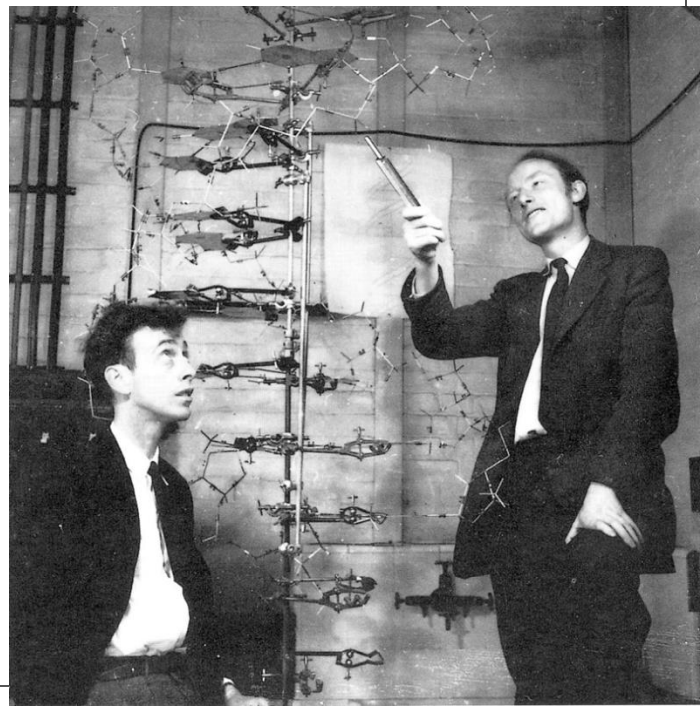
DNA
Genetic analyses
molecular phylogeny

Fishbase internship

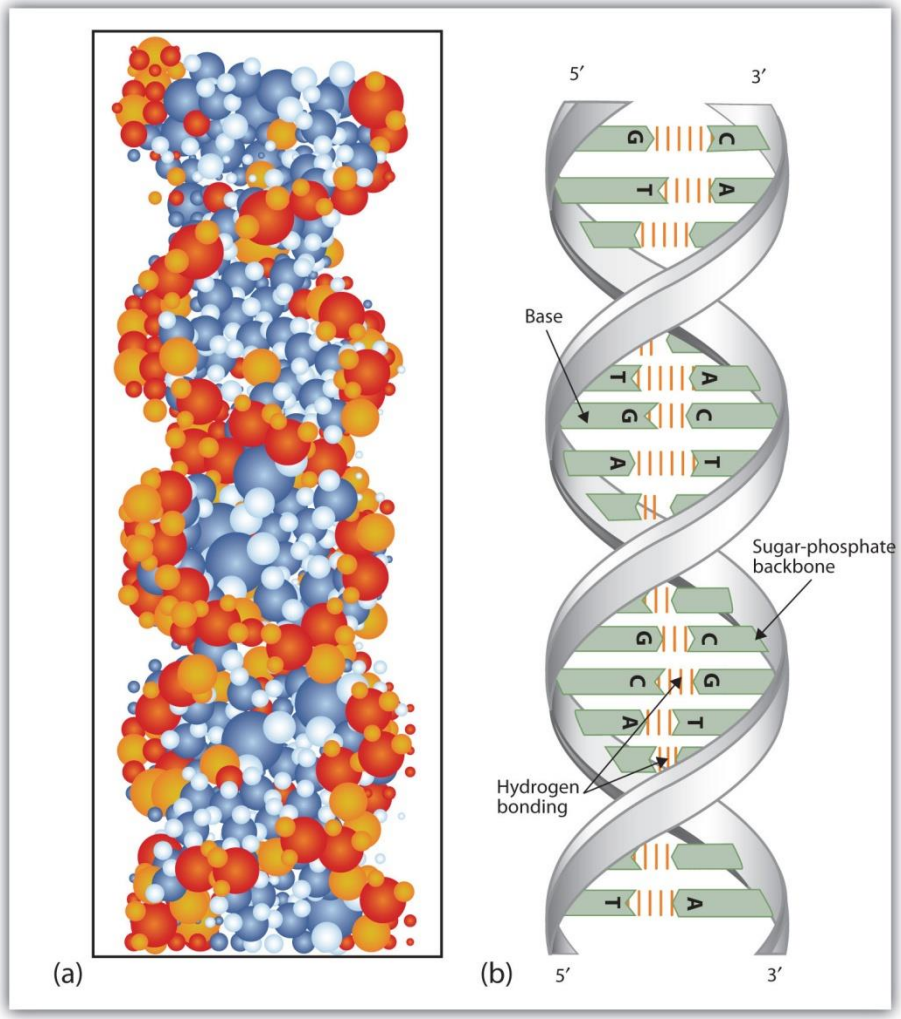
2017

DNA

- laws of Mendel (1822-1884) → heredity of characters; Transmission of certain 'factors' over generations, he did not yet speak of 'genes'
- Darwin "The origin of species (1859)"
- WatsoWatson & Crick (1953): proposed a model for the structure of DNA: the double helix

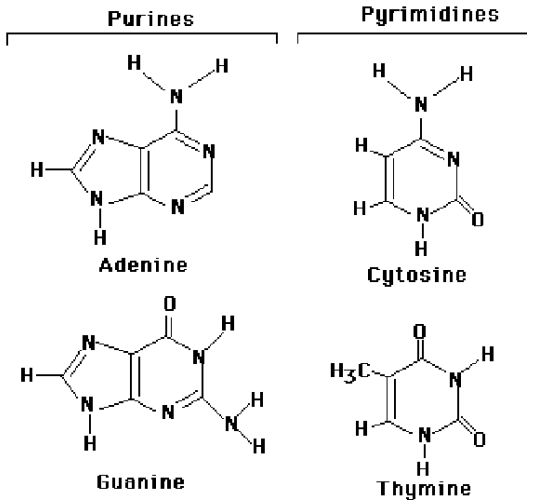


DNA



- The double helix is composed of two complementary strands, formed by sugars (deoxyribose) joined by phosphate groups
- The bases (A, T, G, C) are linked to the phosphate groups as the bars of a ladder in which A is paired with T and G with C
- The two strands are united through hydrogen bonds

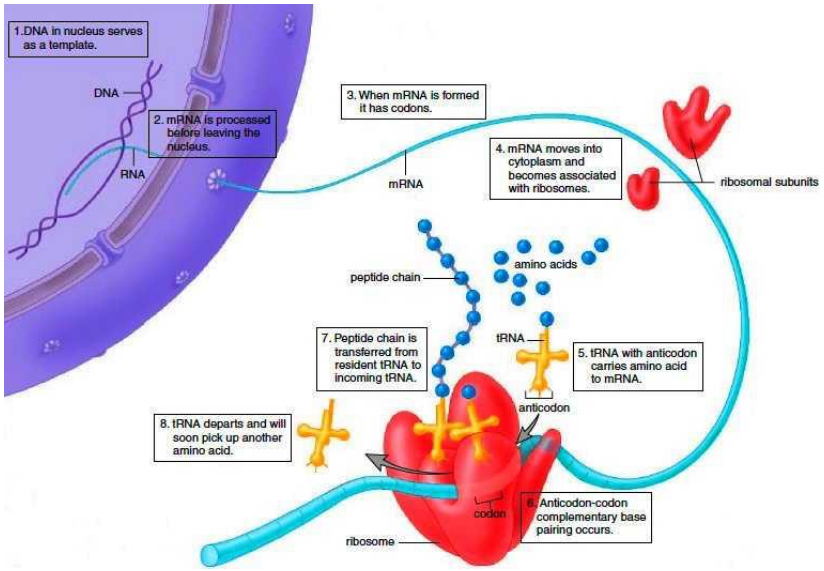
DNA

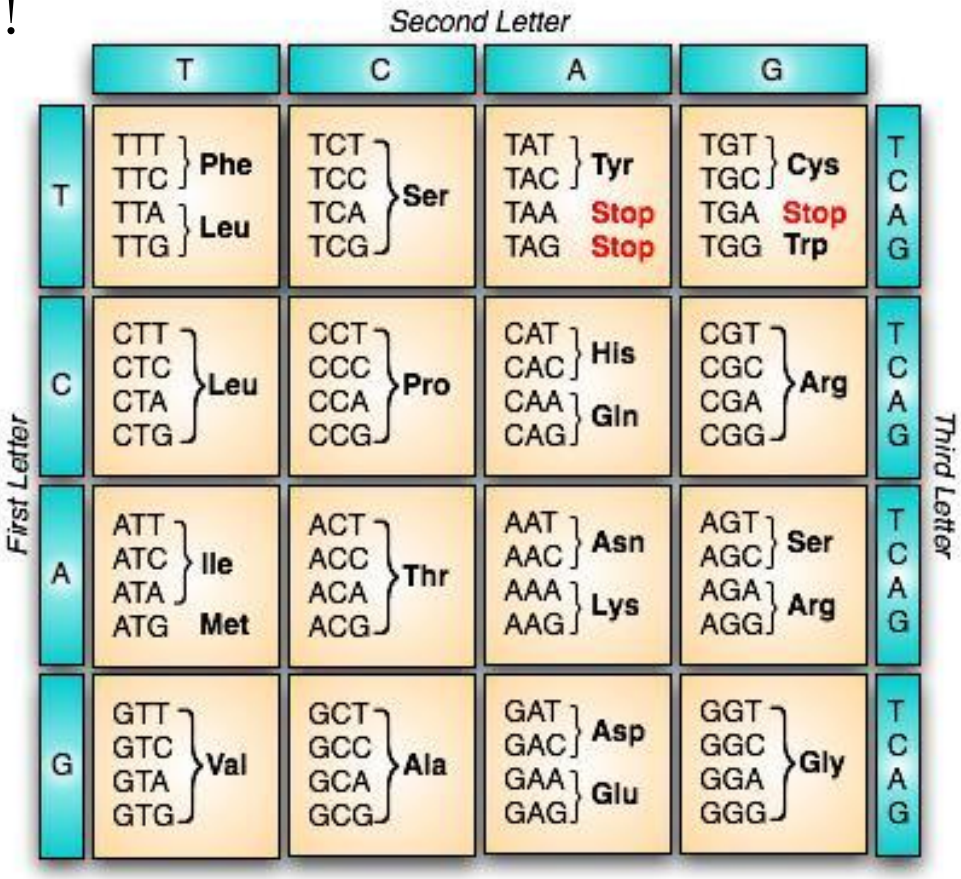


- DNA is composed of 4 bases :
  - Adenine — Purines
  - Guanine
  - Thymine — Pyrimidines
  - Cytosine
- Complementary, a purine associaties with a pyrimidine: A-T / G-C
- Genetic information lies in the succession of these bases → DNA could be compared to a text written in a 4 letter alphabet

DNA: genetic code

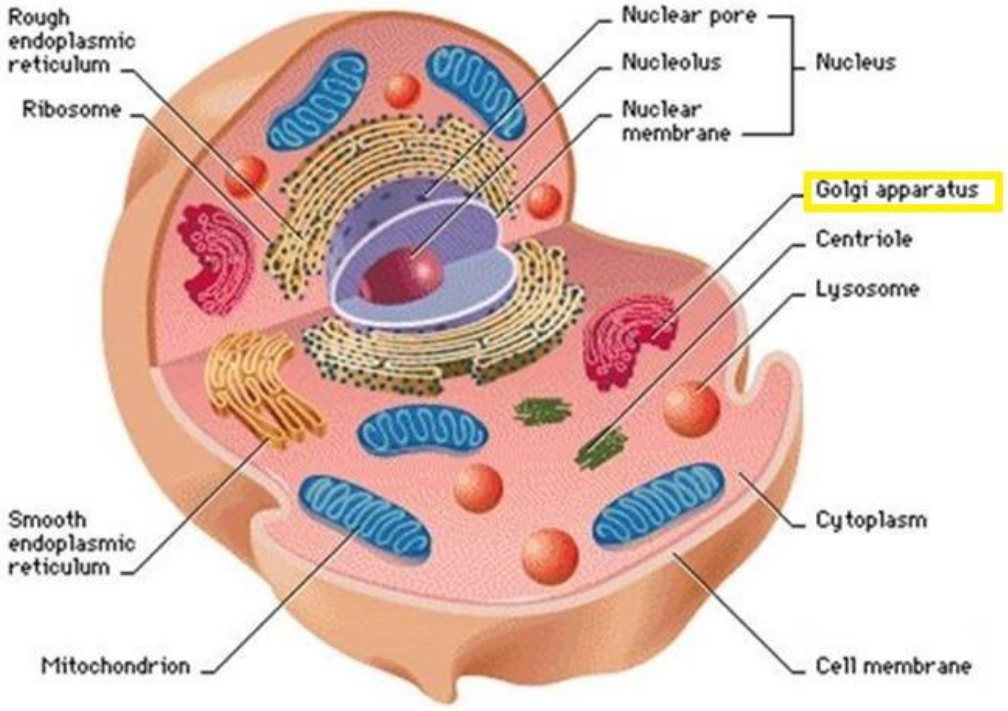
- Coding regions can be translated to proteins
- Most DNA non-coding!





DNA

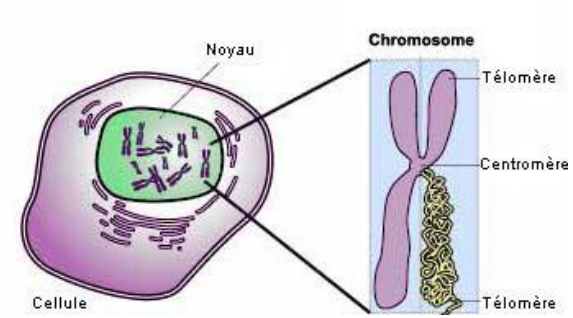
- DNA in different parts of the (eucariote) cell
  - Nucleus (Nuclear DNA >99%)
  - Mitochondria
  - Chloroplastes (plants)

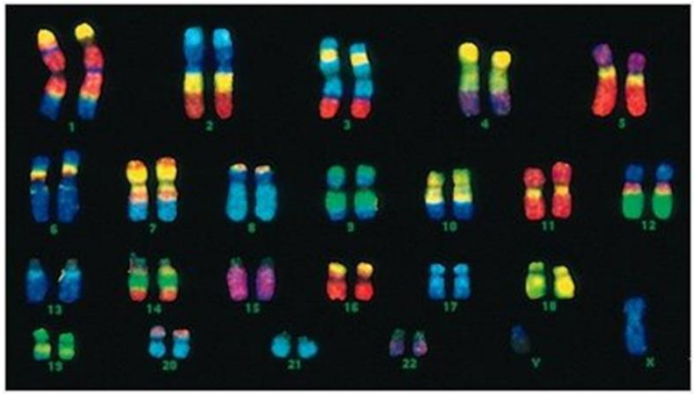


DNA

- In cells: most DNA on chromosomes each chromosome contains an extremely long chain of deoxyribonucleic acid (DNA) that is tightly packed.

- humans: there are 2 copies of each chromosome (1 from the mother and 1 from the father) and there are 23 pairs of chromosomes

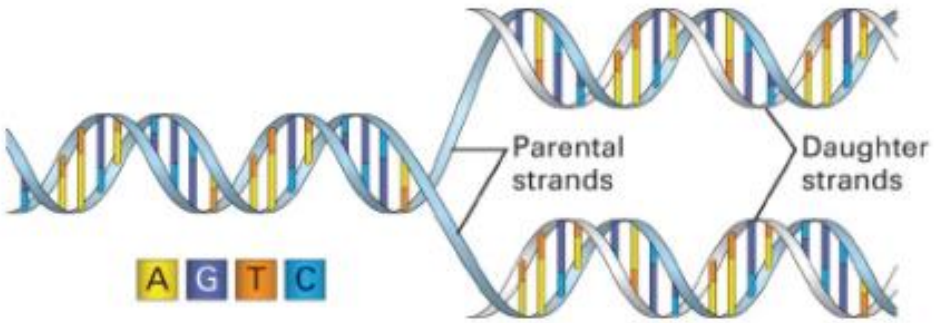




‘reading’ DNA strands

Replication of DNA

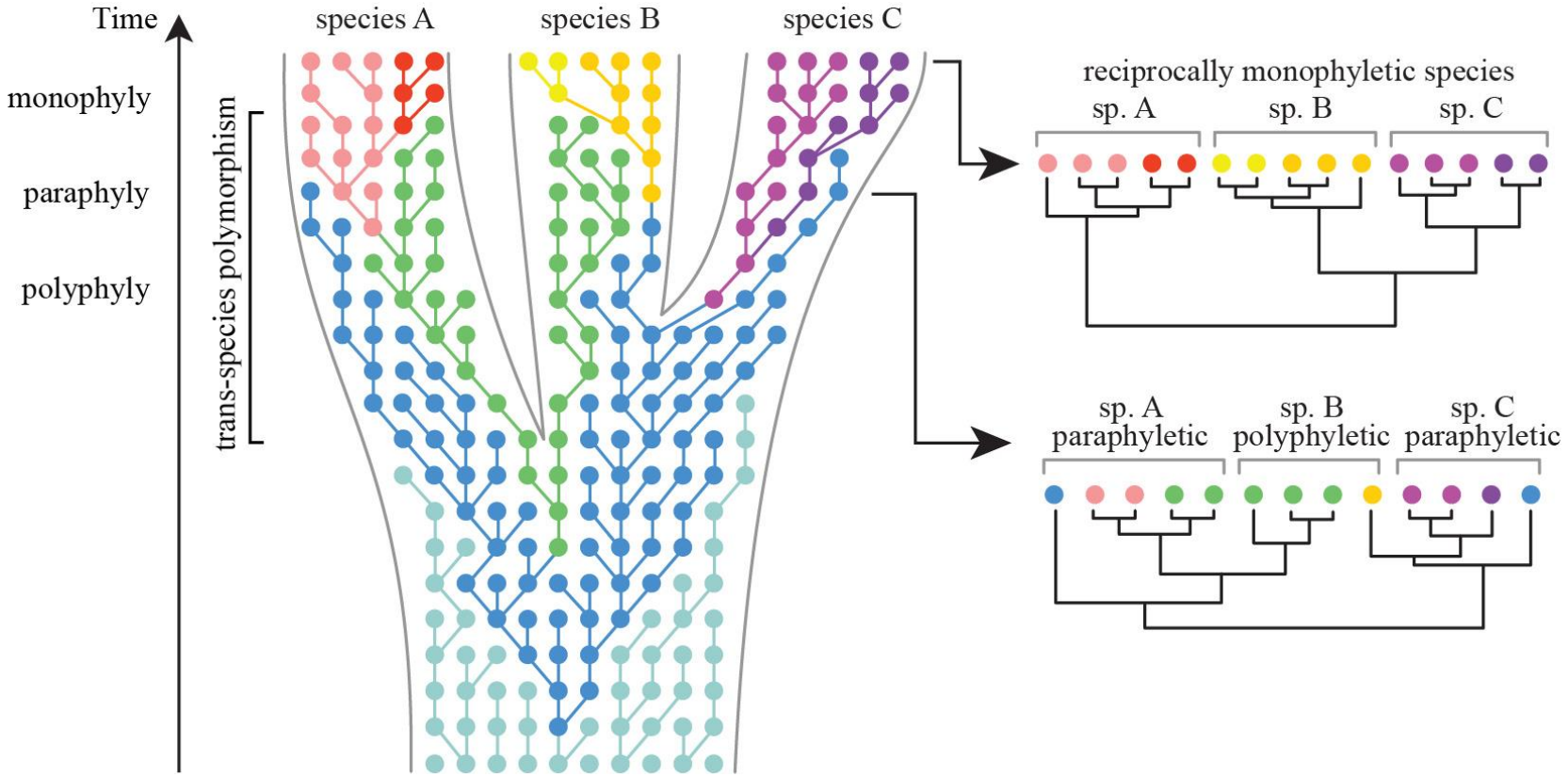
- During replication, DNA strands separate and each strand serves as a model for the synthesis of a complementary strand



- Replication thus ensures the transfer of genetic information from a cell to its descendants.
- During this process mistakes can occur → mutations!
  - 1 gene can have many alleles
  - Substrate for evolution
- Artificially: Polymerase Chain Reaction (PCR)

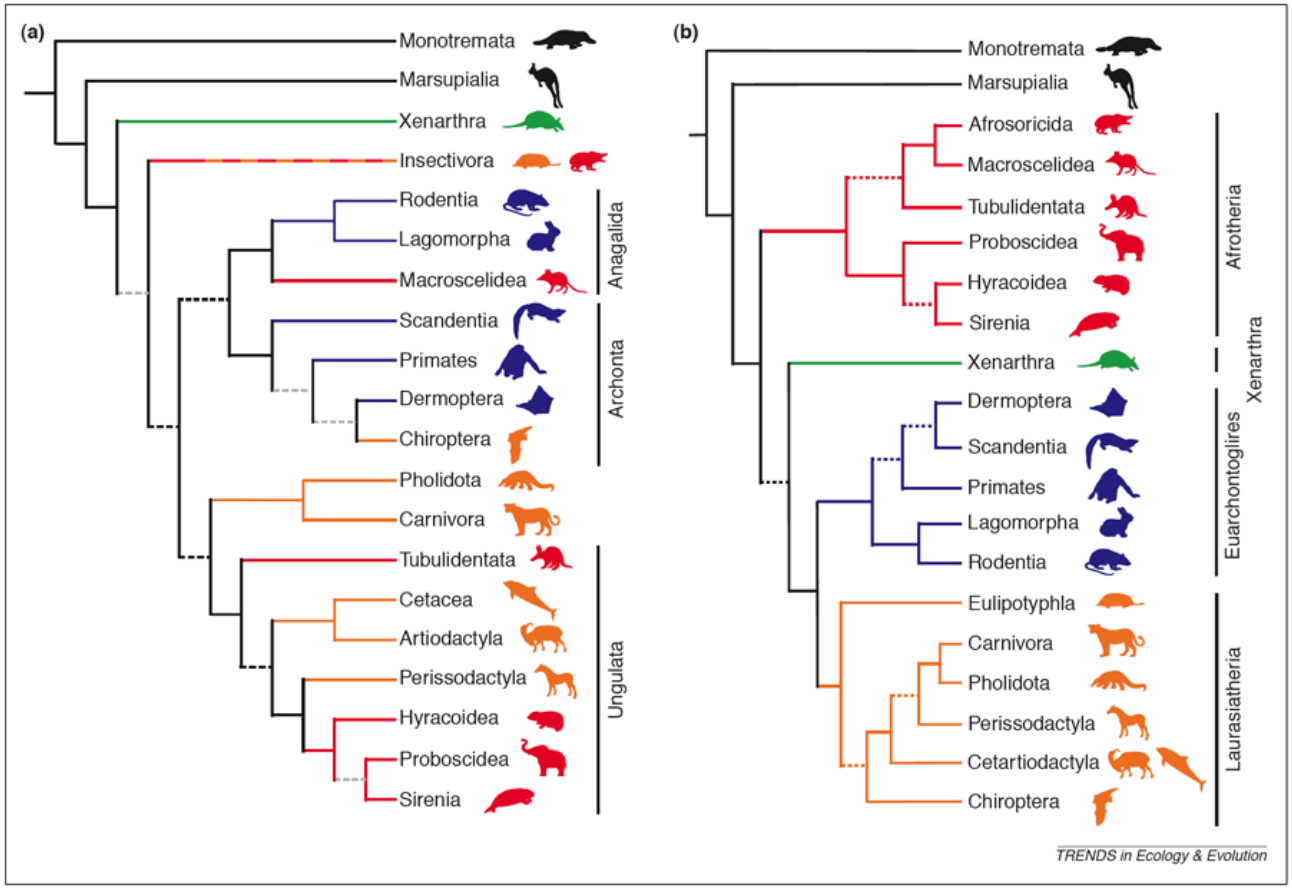
Evolutionary history

- Reconstruct species tree from gene trees



Evolutionary history

Relationships: classification based on a large amount of information

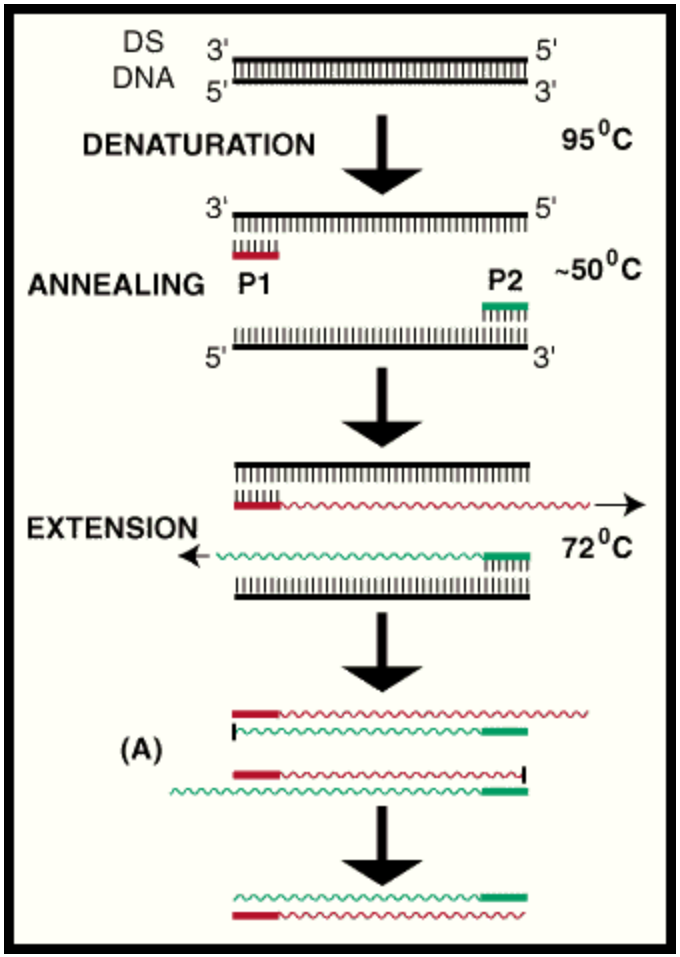


PCR

To amplify a gene of interest

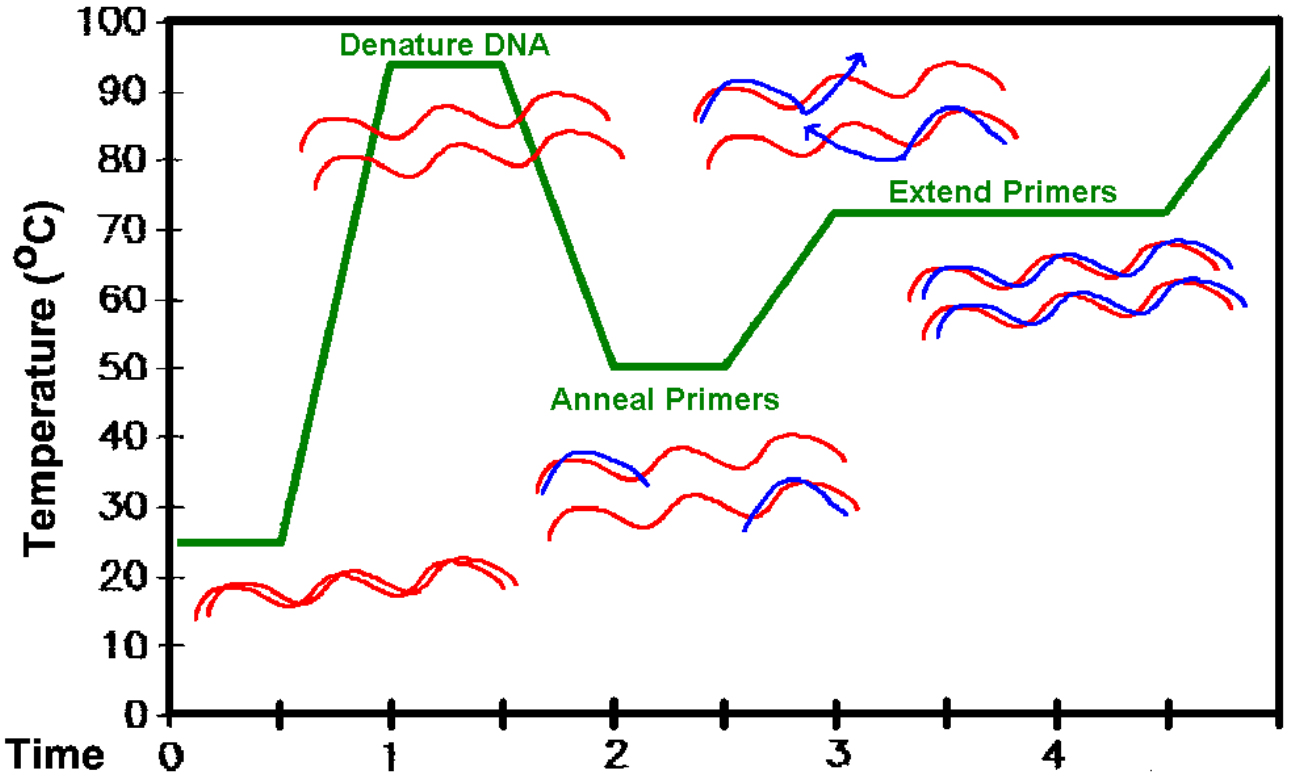
The PCR is carried out in 3 phases:

- Denaturation: separation of two DNA strands by rising the temperature
- Pairing primers: following a decrease in temperature, the specific primers **complementary to the DNA to be amplified** will hybridize on the DNA strand
- Elongation: synthesis of the complementary strand by Taq polymerase, which will add nucleotides (dNTPs) at the primers

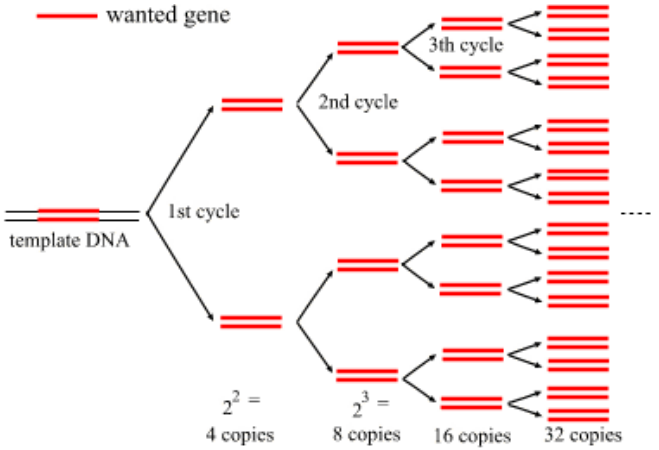


PCR

- Different stages of PCR occur at different temperatures

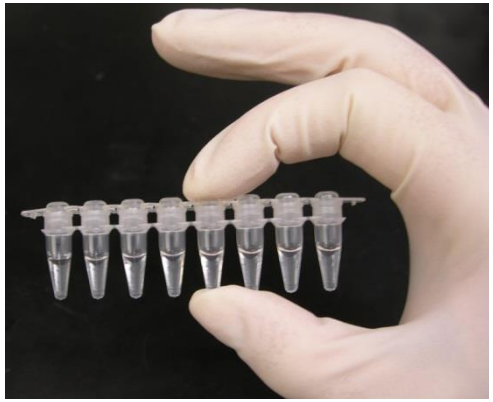


PCR



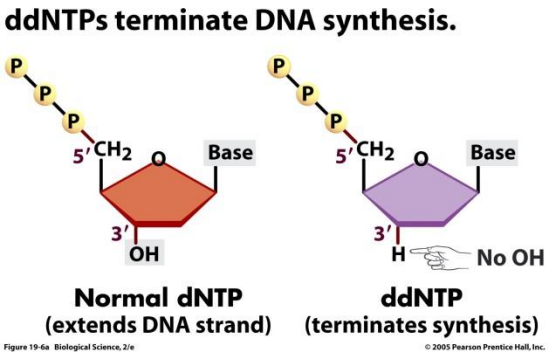
- Exponential reaction that uses the products of each step as a matrix of the next steps.
- This process will generate thousands of copies

The reaction is carried out in special tubes, placed in an apparatus that makes it possible to adjust the temperatures for each stage: a **PCR block**



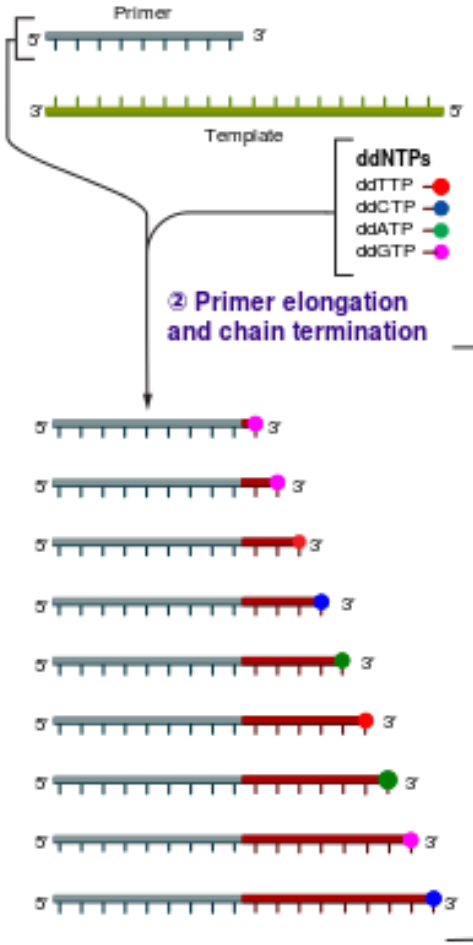
DNA sequencing

- This technique makes it possible to know the succession of the nucleotides that compose a part of a DNA molecule
- **Sanger sequencing** is the most used technique
  - Similar to PCR, the difference is the incorporation of coloured ddNTP (dideoxyribonucleotides): the polymerase is stops adding nucleotides to the sequence
  - A set of DNA strands of varying sizes is obtained, depending on where a ddNTP is inserted
  - Fluorescent signal is 'read' in a sequencer

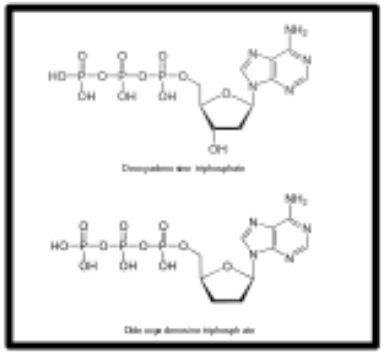


DNA Sequencing

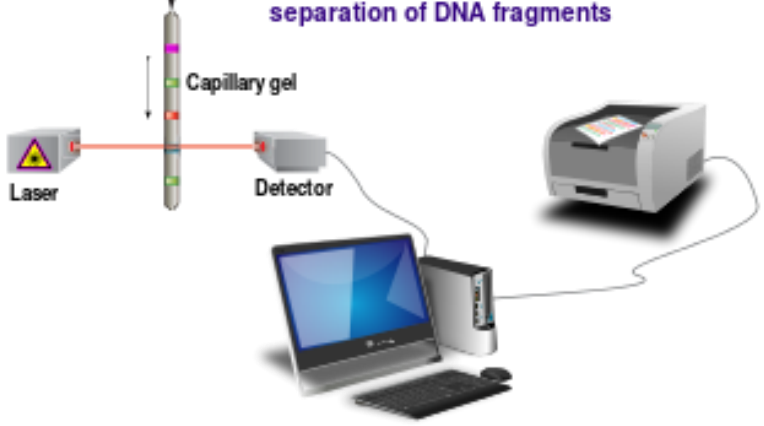
① Reaction mixture
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)

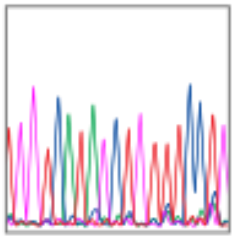


② Primer elongation and chain termination



③ Capillary gel electrophoresis separation of DNA fragments



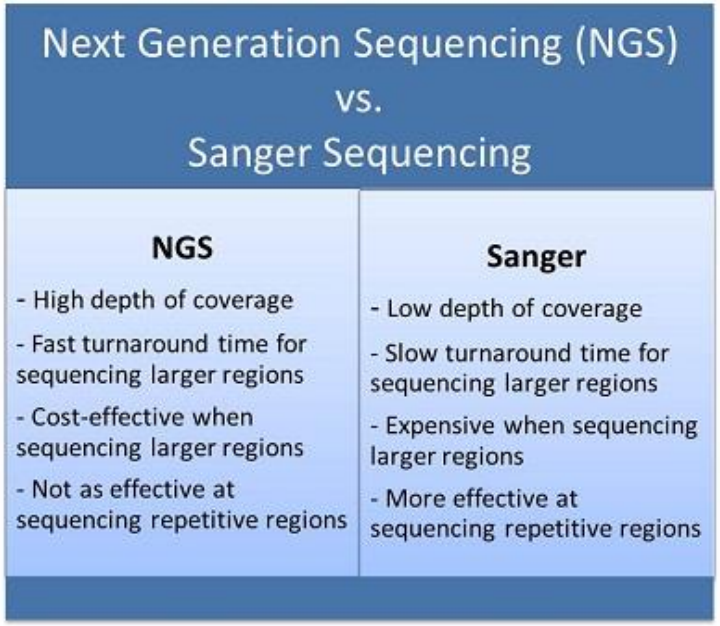


Chromatograph

④ Laser detection of flourochromes and computational sequence analysis

DNA Sequencing

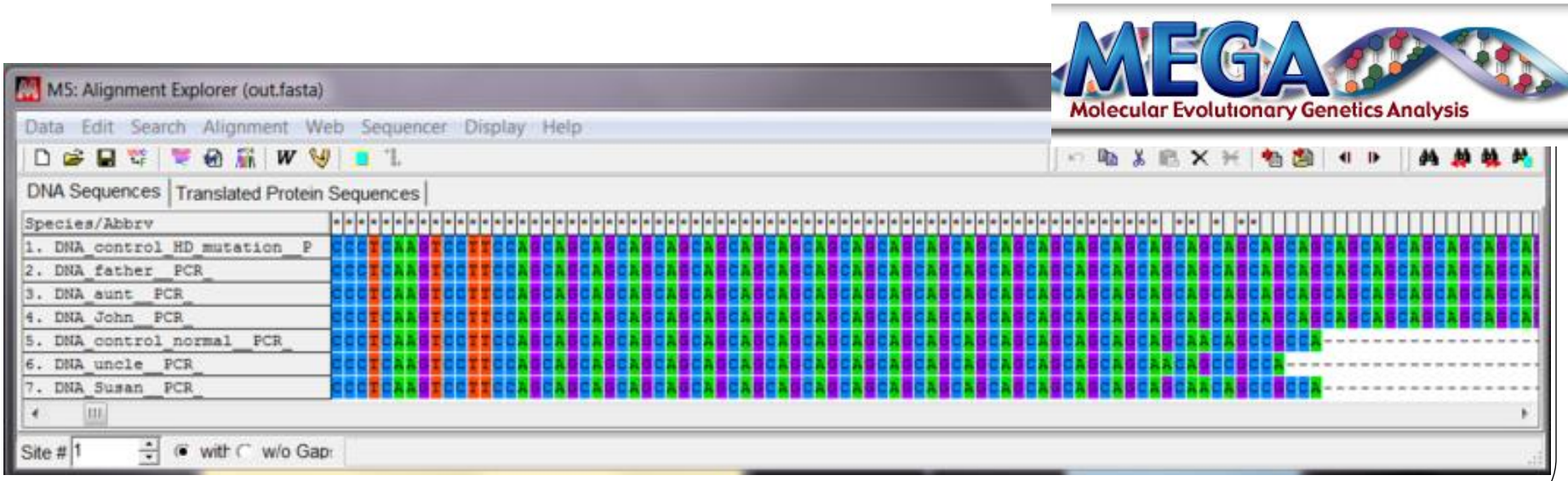
- This technique is sometimes replaced by "**Next Generation Sequencing**" (NG) that allows to obtain a complete genome but:
  - Higher cost
  - Much more data to analyze
  - Interesting only in the framework of big project, not in routine



Interpreting Genetic information by molecular phylogeny

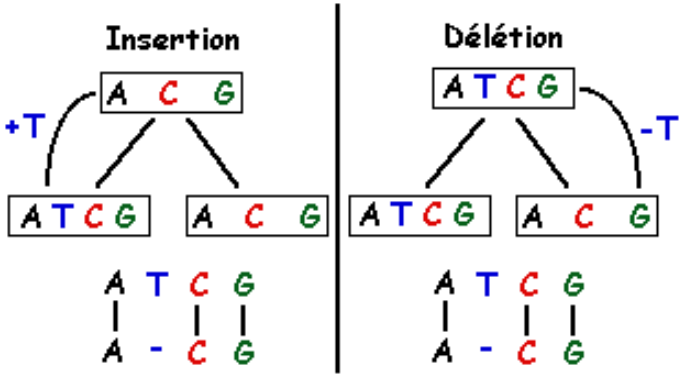
Molecular Phylogeny

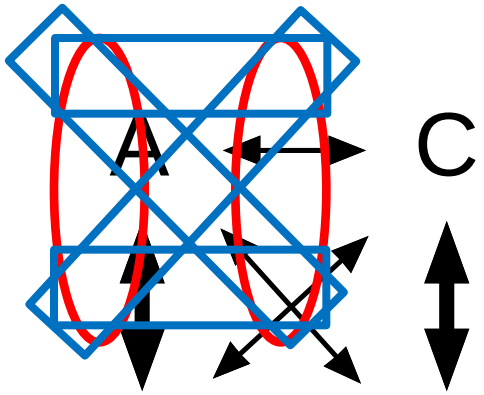
- Molecular phylogeny uses software packages to trace the history of the mutations that have appeared during the evolution of a given gene, by comparing the sequences of different species
- If we want to compare the sequences, we must align them
- **MEGA** is a free software package for many aspects of molecular phylogenetics



Molecular phylogeny

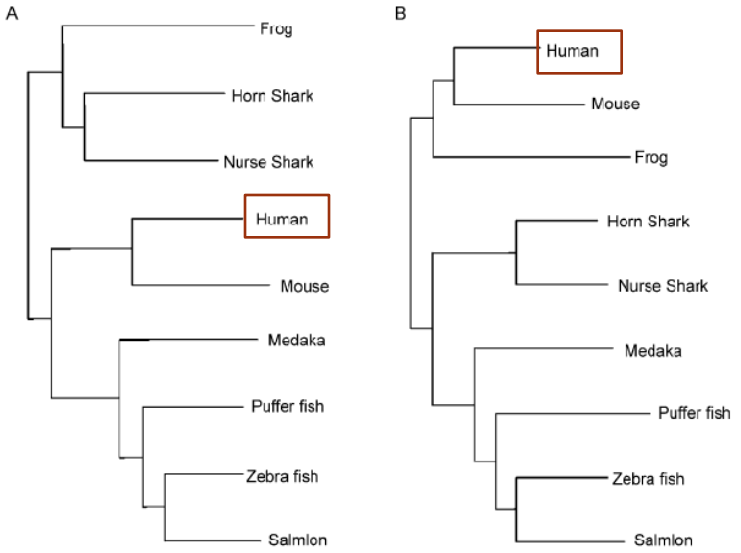
- After alignment, we can see which mutations have taken place:
  - Transitions between purines or between pyrimidines
  - Transversions of a purine to a pyrimidine
  - Deletions or insertions
- A model is chosen to weigh the importance of these differences





Molecular phylogeny

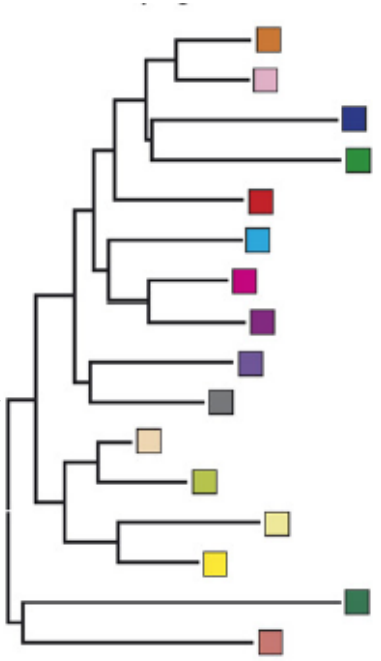
- There are different models of evolution which will analyze the rates of substitutions, the frequency of the bases, the number of transitions or transversions, the possibilities of mutations with different parameters
- Different models will give different results, so we must choose the most suitable!



Two different phylogenetic trees, based on the same data but using two different models of molecular evolution!

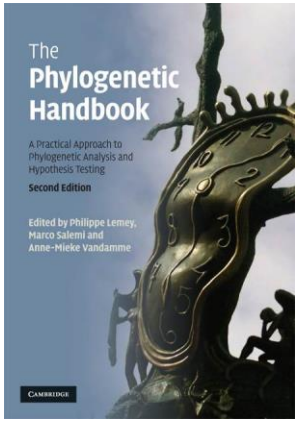
Molecular phylogeny

- A phylogenetic tree can then be constructed by grouping first the most similar sequences and then gradually those which are the most different
  - Neighbour Joining
  - This can be based on different measures of distance
- To construct the phylogenetic trees, we need a model of evolution of the nucleotide sequences which will evaluate
- In case of doubt on the method to choose, it is always better to make phylogenetic trees with all the methods and comparing the results
- To have a first impression we can already build some trees in MEGA



Molecular phylogeny

- If we have the correct model, there are still different possibilities
- Different methods to build a tree
- We need different programs to use different methods (Neighbor joining, bayesian analysis, maximum likelihood, maximum parsimony)
- All methods will use different ways to build phylogenetic trees
- For more information: 'phylogenetic handbook'

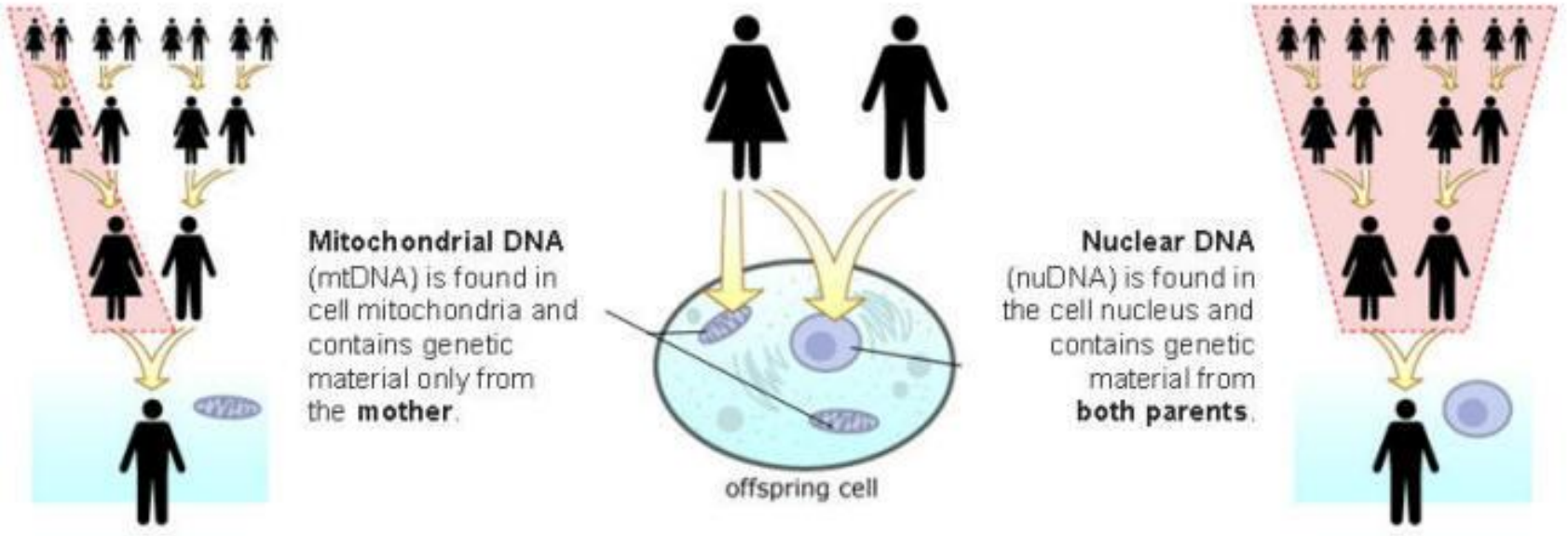


Choice of genes

- Mutation rate
- Coding or non-coding
- Mitochondrial or Nuclear?
- Tree of a gene = tree of species?

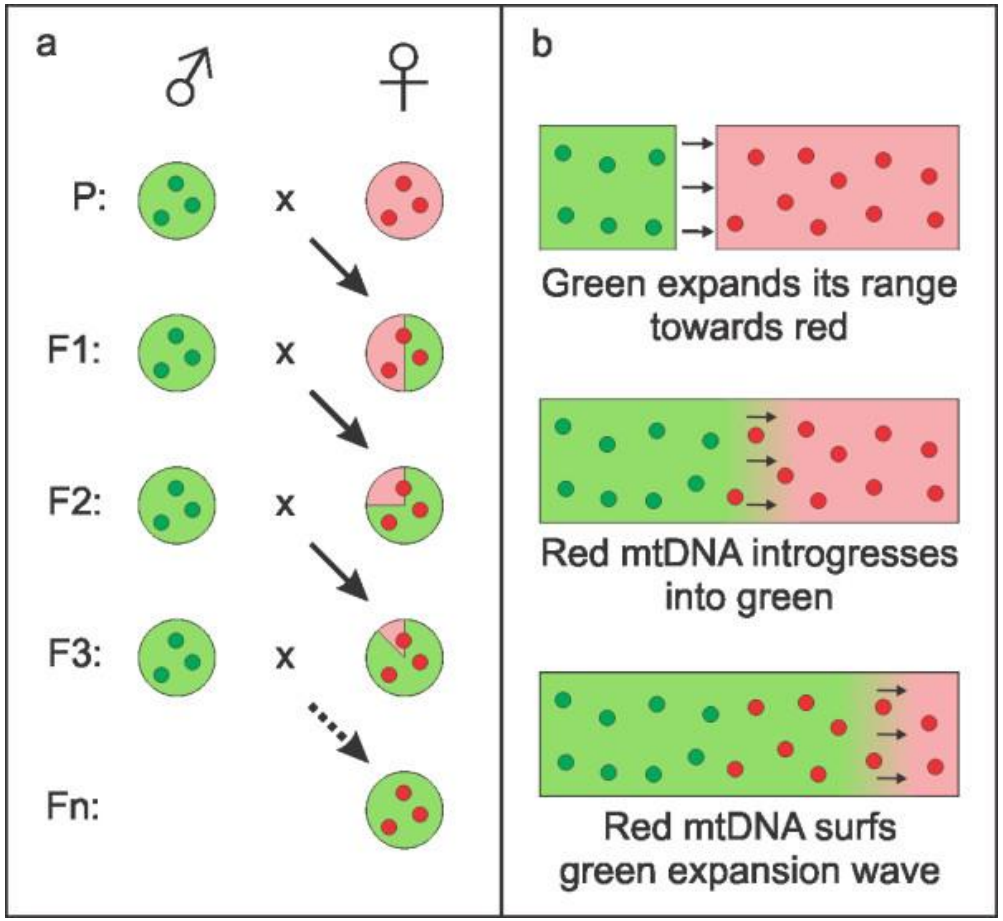
Choice of genes

- Mitochondrial or nuclear?



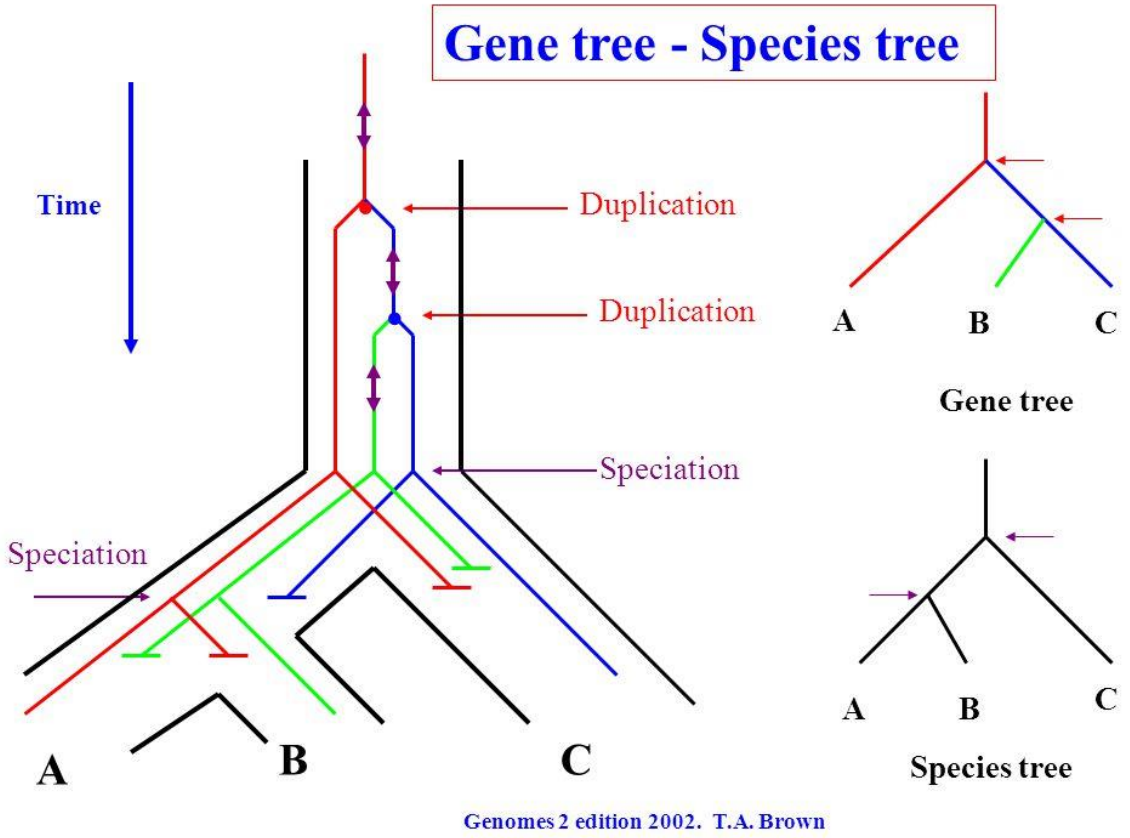
Choice of genes

- Mitochondrial or nuclar?: introgression



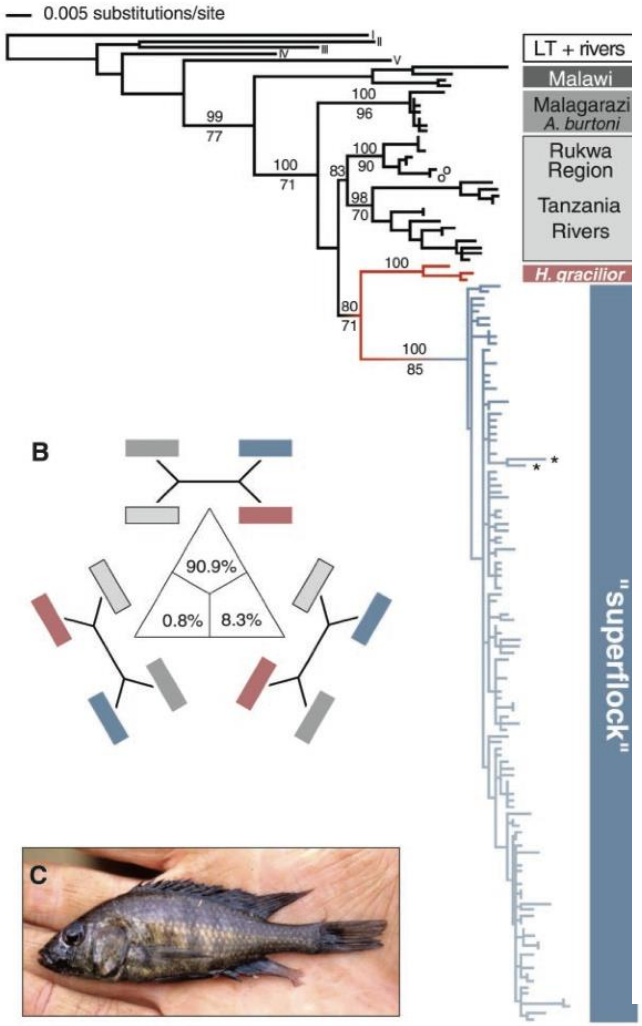
Choice of genes

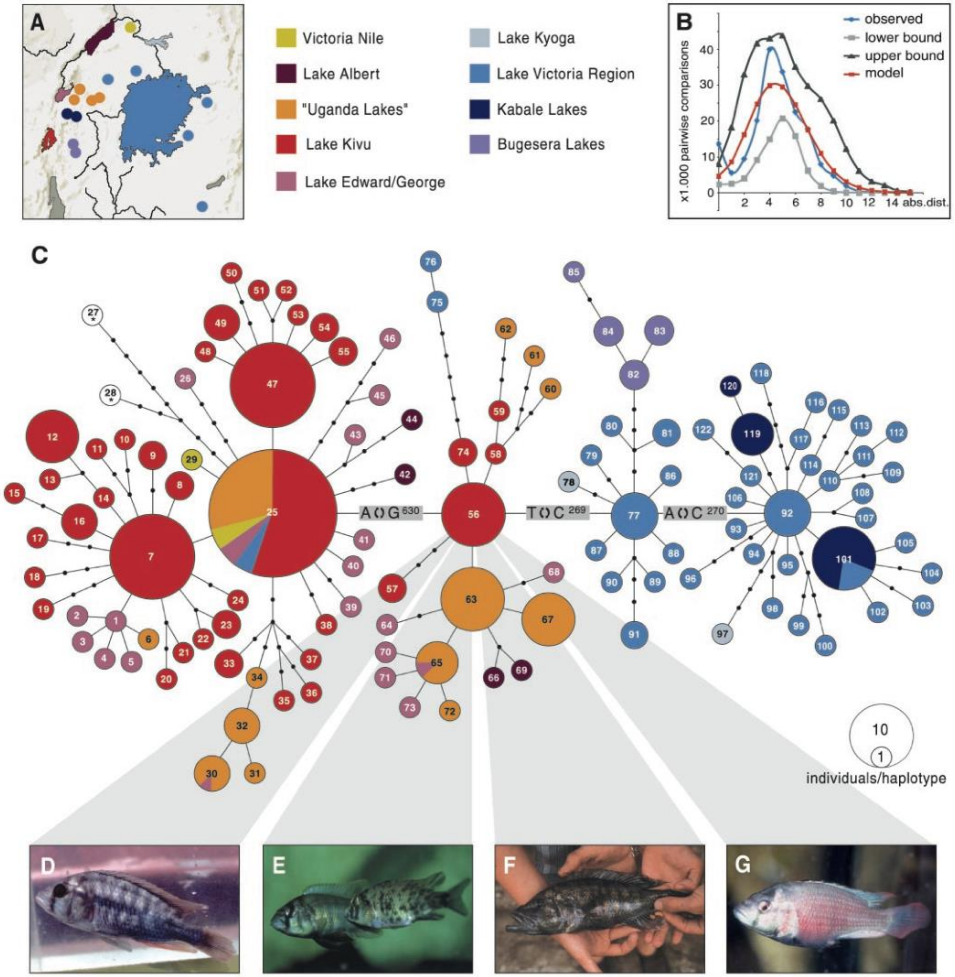
- Gene tree = species tree?



Alternative methods

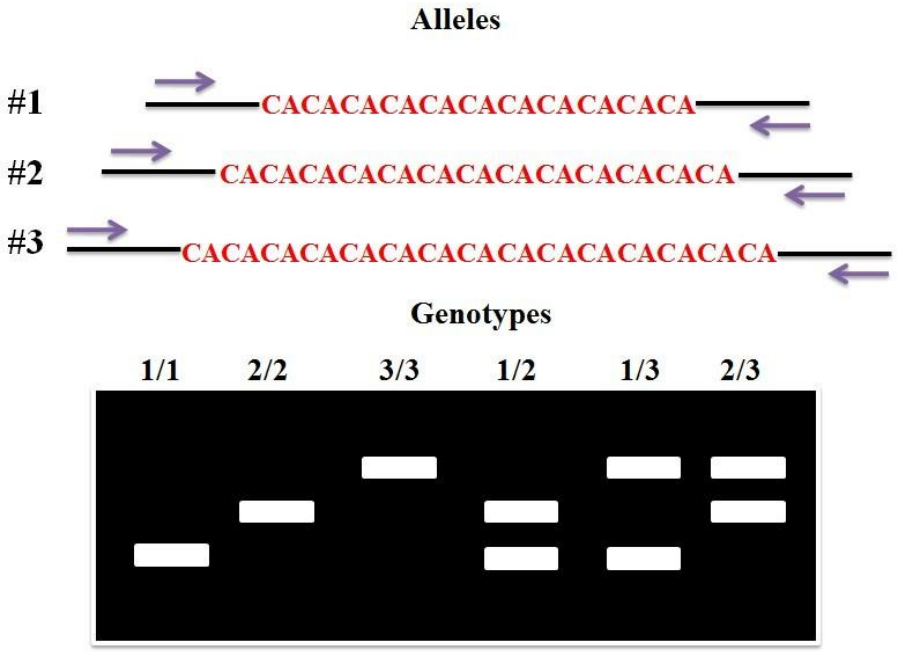
Visualisation: Trees or networks





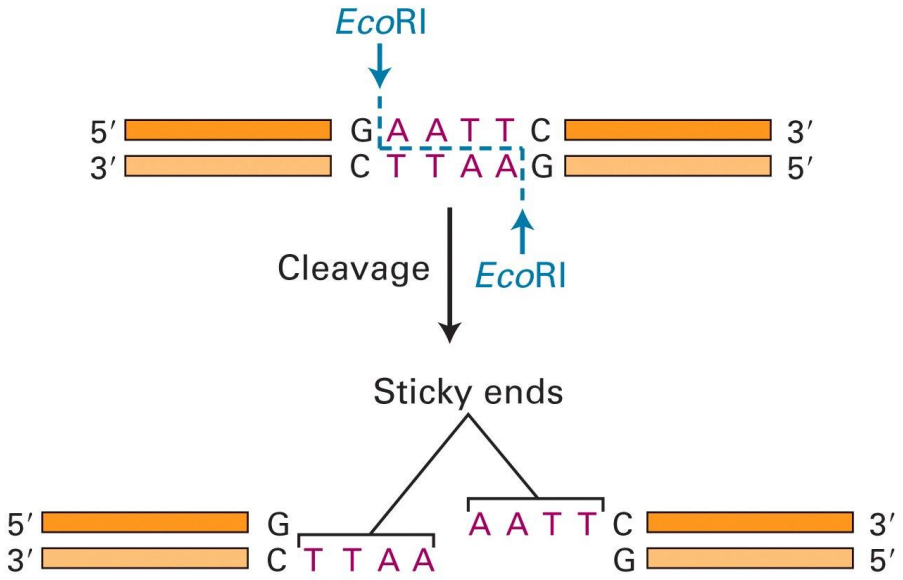
Intra-specific variation

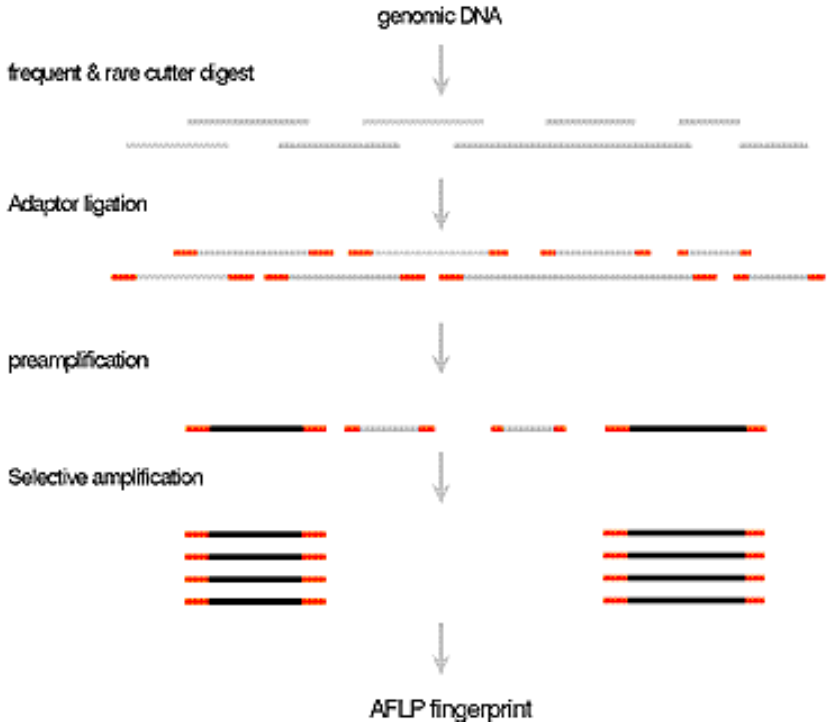
- Microsatellites
  - Population genetics
  - Paternity
  - Legal Investigations
  - Medicinal



Complete genome

- Restriction enzymes
  - AFLP
  - Next generation sequencing





DNA Barcoding

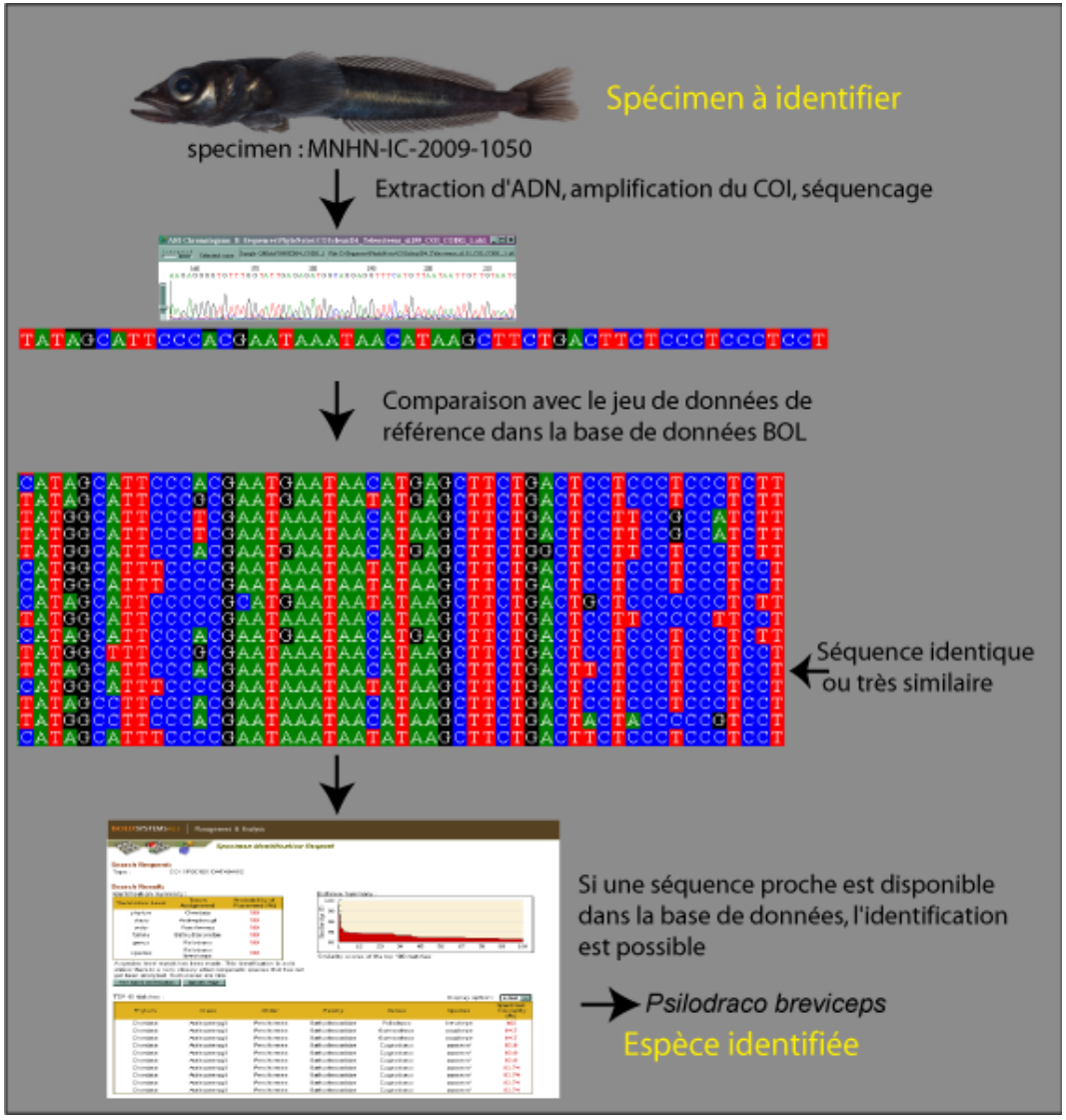
DNA Barcoding

- DNA Barcoding is a technique that allows molecular identification based on the comparison of short (mitochondrial) DNA sequences
- A good barcode gene is:
  - A variable between species but very conserved within a species, giving it a strong discriminating power
  - A sequence short enough to be able to sequence easily but long enough to have enough information
- In many taxa (including fish), the first subunit of cytochrome oxidase (COI) gene is used (length of 650 bp)

DNA Barcoding

- DNA barcoding is used for various reasons:
  - Quickly assessing the diversity of a region/group
  - Identification of cryptic species whose morphology is almost identical, making their distinction difficult
  - Identification of incomplete specimens, in order to identify the species to which a tuft of hair or an insect's leg belongs
  - large-scale studies to study biodiversity and environmental hazards
  - Identification of invasive species
  - Discovery of new species
  - Association of males and females of the same species (when males and females are dimorphic)
  - Association of stages of development in the same species

DNA Barcoding



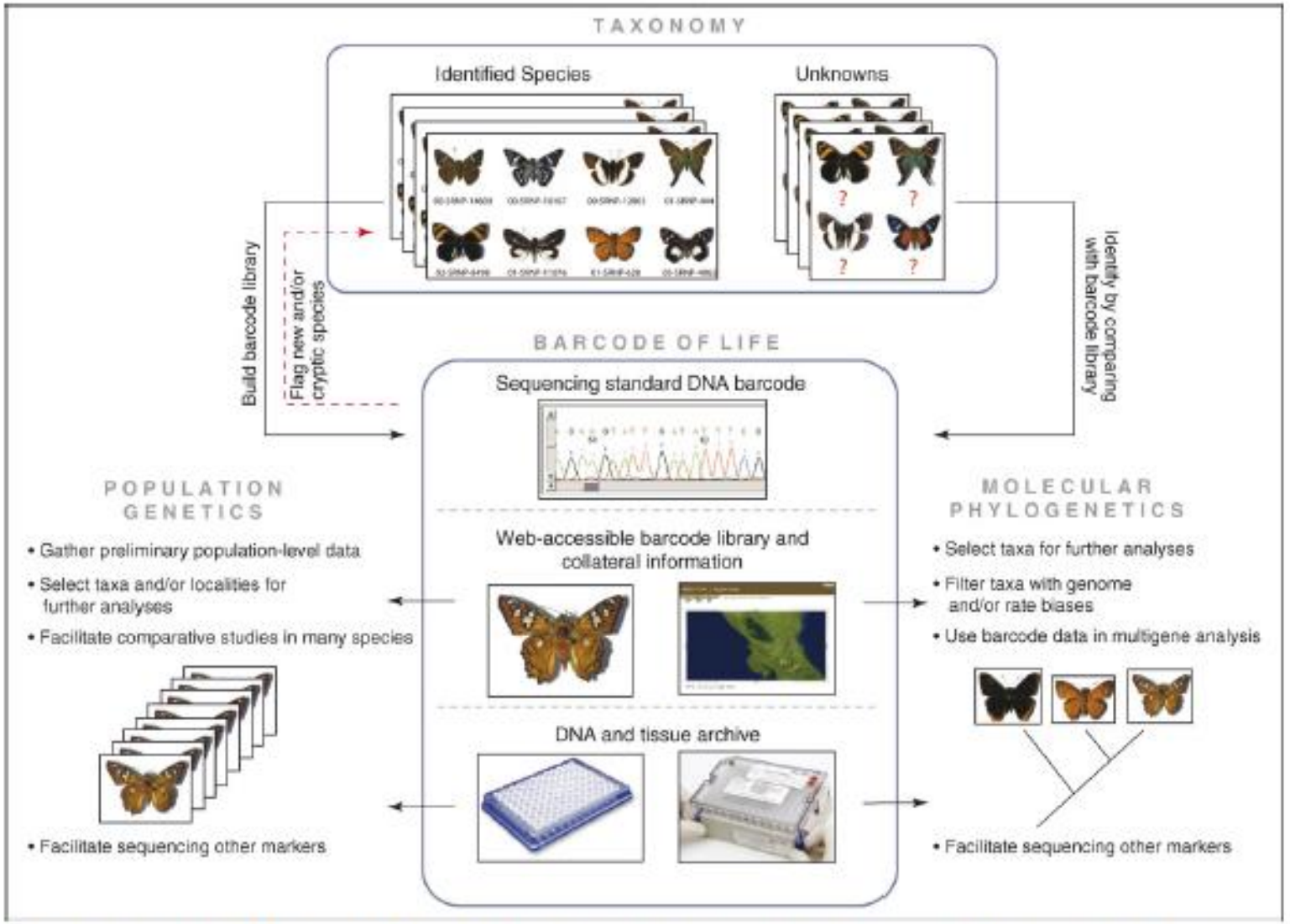
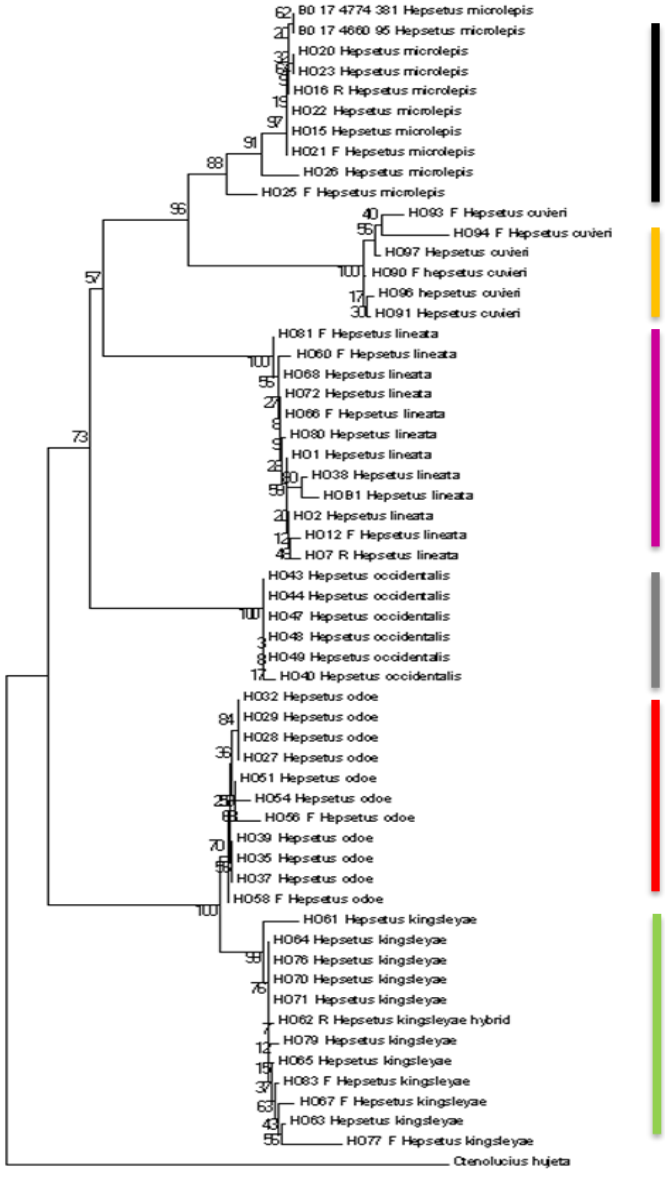


Figure 2. Major components of the Barcode of Life projects and their contribution to taxonomy, reconstruction of molecular phylogenies and population genetics investigations. This diagram shows how DNA barcoding libraries can support the conventional taxonomic workflow by high-throughput identification of unknown specimens and by helping to draw attention to new and cryptic species. Barcode sequences and collateral data for each specimen are accessible through a global online data base (e.g. BOLD: http://www.barcodinglife.org). This information can be useful in other contexts, such as phylogenetics (Tree of Life projects) and population-level studies. In addition, archival DNA and tissue specimens collected in barcoding projects provide an excellent resource for other investigations. Butterfly images are taken from the database of Daniel Janzen and Winnie Hallwachs (http://janzen.sas.upenn.edu/).

Confirmation of morphological studies

e.g. revision of *Hepsetus*



H. microlepis

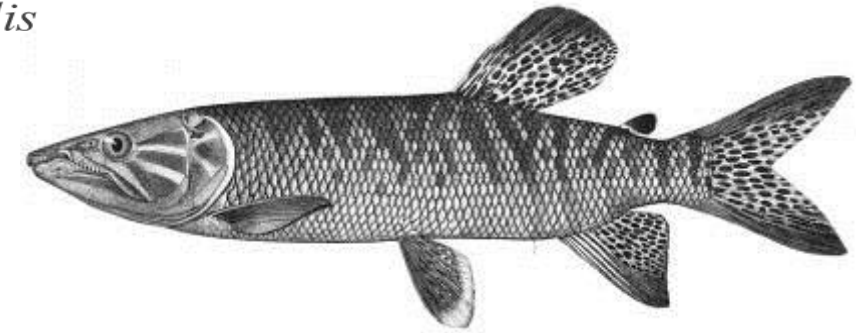
H. cuvieri

H. lineata

H. occidentalis

H. odoe

H. kingsleyae



DNA Barcoding

- **BUT** taxonomic analyzes can not be replaced by molecular techniques! DNA barcoding can help and facilitate the identification process and allow the discovery of new species or other biological questions but can in no way replace conventional taxonomy

DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics

Mehrdad Hajibabaei[1], Gregory A.C. Singer[2], Paul D.N. Hebert[1] and Donal A. Hickey[3]

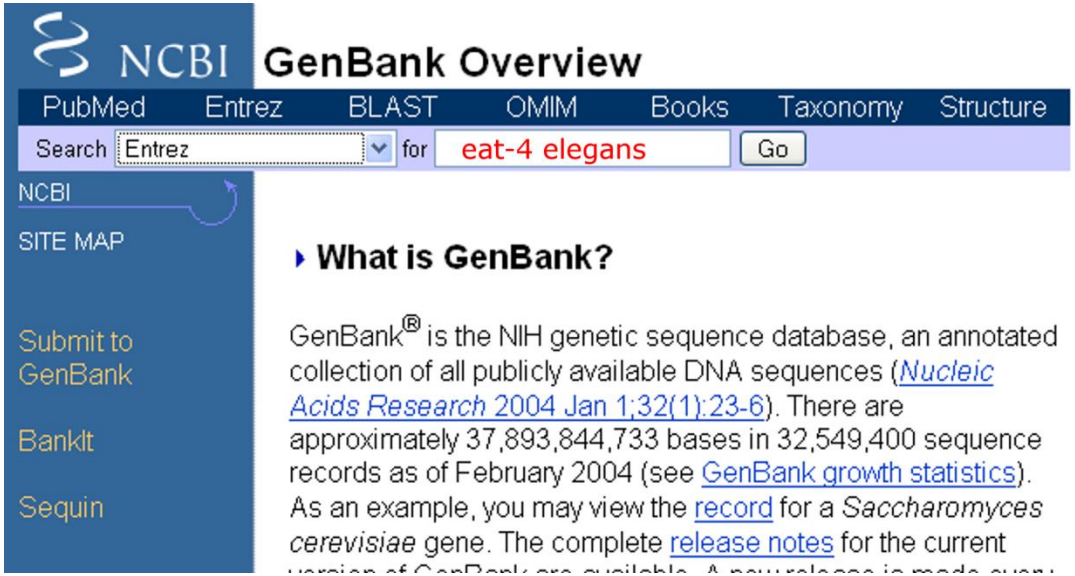
[1] Biodiversity Institute of Ontario, Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[2] Human Cancer Genetics Program, The Ohio State University, Columbus, OH 43210, USA
[3] Department of Biology, Concordia University, 7141 Sherbrooke Street, Montreal, Quebec H4B 1R6, Canada

DNA Barcoding

- Molecular barcoding is a campaign that is performed around the world, as shown these organizations:
  - "Barcode of Life Data Systems" (BOLD): www.boldsystems.org
  - "Consortium for Barcode of Life" (CBOL): http://www.barcodeoflife.org/
  - "European Consortium for the Barcode of Life": http://www.ecbol.org/
  - "International Barcode of Life" (iBOL): www.iBOL.org
  - "The Belgian Network for DNA Barcoding" (BeBOL): http://bebol.myspecies.info
  - Canadian Center for DNA Barcoding (CCDB): http://www.ccdb.ca
  - "The Fish Barcode of Life Initiative" (Fish-BOL): http://www.fishbol.org

Genbank

- Database for published sequences
- www.genbank.org



DNA Barcoding

- But everything is not so simple in practice …

BARCODING

Universal primer cocktails for fish DNA barcoding

NATALIA V. IVANOVA,* TYLER S. ZEMLAK, ROBERT H. HANNER and PAUL D. N. HEBERT
Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1

- The primers described in the literature are not always universal and often, many adjustments are necessary.

Questions?