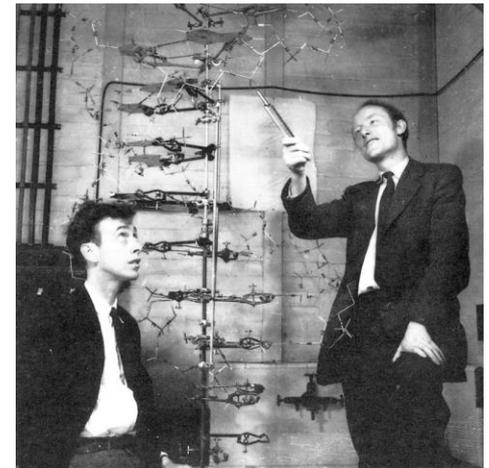
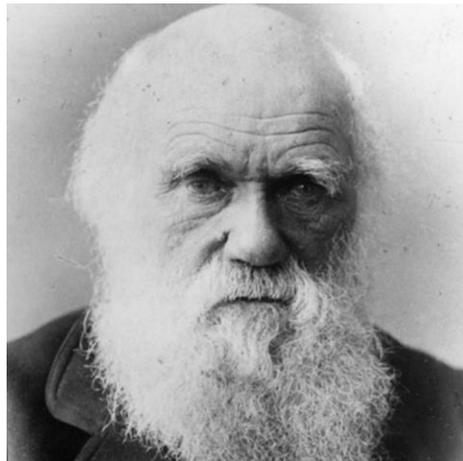


ADN
Analyses Génétiques
&
Phylogenie moleculaire

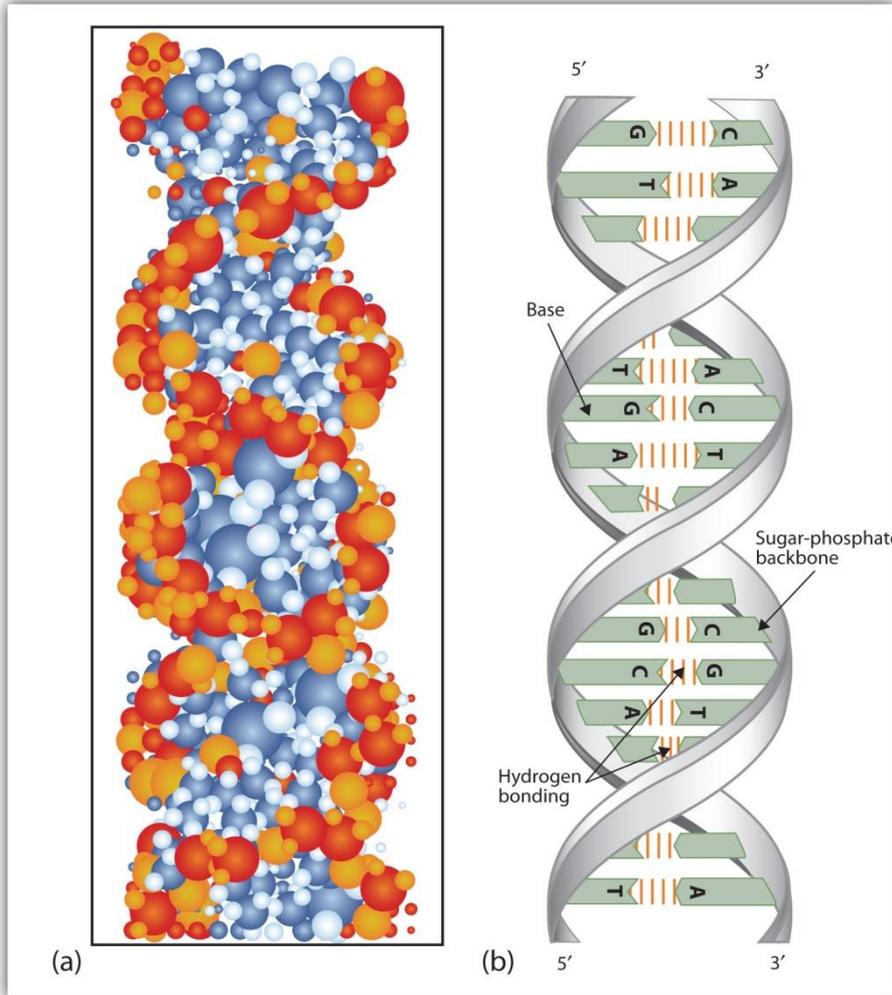
Formation fishbase
2018

Introduction

- Les lois de Mendel (1822-1884) → hérédité des caractères; transmission de certains 'facteurs' au fil des générations, il ne parlait pas encore des 'gènes'
- Darwin « l'origine des espèces (1859) »
- Watson & Crick (1953) ont proposé un modèle pour la structure de l'ADN : **la double hélice**



L'ADN



- La double hélice est composée de deux brins complémentaires qui sont formés par un assemblage de sucre (désoxyribose) unis par des groupements phosphates
- Les bases (A, T, G, C) sont liées aux groupements phosphate et forment les barreaux d'une échelle imaginaire dans laquelle A est toujours en face de T et G en face de C
- Les deux brins sont unis grâce à des liaisons hydrogène

L'ADN

- L'ADN est composé de 4 bases azotées :

- Adénine | Purines

- Guanine

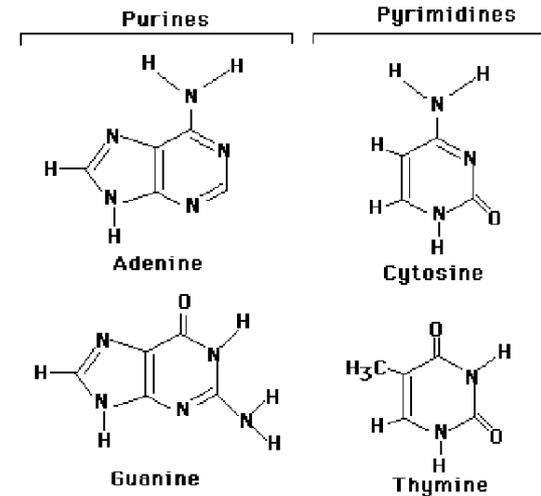
- Thymine | Pyrimidines

- Cytosine

- Les bases azotées sont complémentaires deux à deux, une base purique s'associant toujours à une base pyrimidique

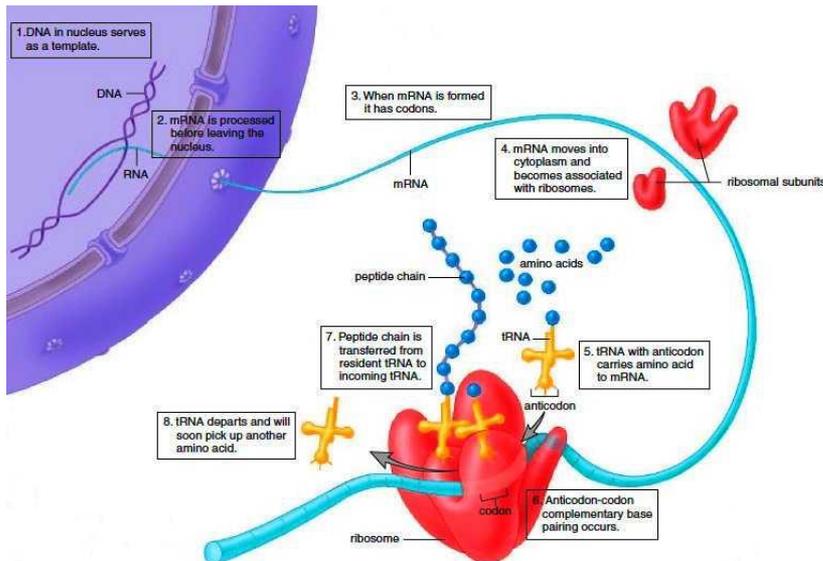
- L'information génétique va résider dans la succession de ces bases

➔ On pourrait comparer l'ADN à un texte écrit dans un alphabet de 4 lettres



L'ADN: code génétique

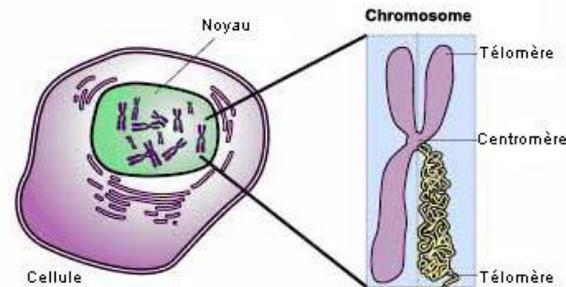
- Région codante pourrait être traduit en protéines
- Pluspart d'ADN non-codante!



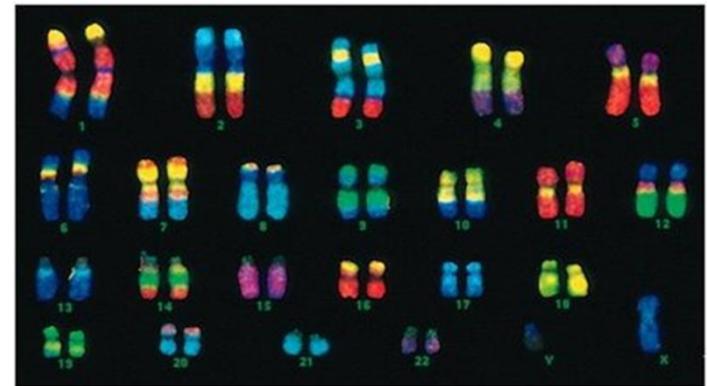
		Second Letter				
		T	C	A	G	
First Letter	T	TTT } Phe TTC } TTA } Leu TTG }	TCT } TCC } Ser TCA } TCG }	TAT } Tyr TAC } TAA } Stop TAG } Stop	TGT } Cys TGC } TGA } Stop TGG } Trp	T C A G
	C	CTT } CTC } Leu CTA } CTG }	CCT } CCC } Pro CCA } CCG }	CAT } His CAC } CAA } Gln CAG }	CGT } Arg CGC } CGA } CGG }	T C A G
	A	ATT } ATC } Ile ATA } ATG } Met	ACT } ACC } Thr ACA } ACG }	AAT } Asn AAC } AAA } Lys AAG }	AGT } Ser AGC } AGA } Arg AGG }	T C A G
	G	GTT } GTC } Val GTA } GTG }	GCT } GCC } Ala GCA } GCG }	GAT } Asp GAC } GAA } Glu GAG }	GGT } Gly GGC } GGA } GGG }	T C A G
						Third Letter

L'ADN

- Dans des cellules: ADN sur des 'chromosomes' → chaque chromosome contient une chaîne extrêmement longue d'acide désoxyribonucléique (ADN) qui est fortement empaquetée.

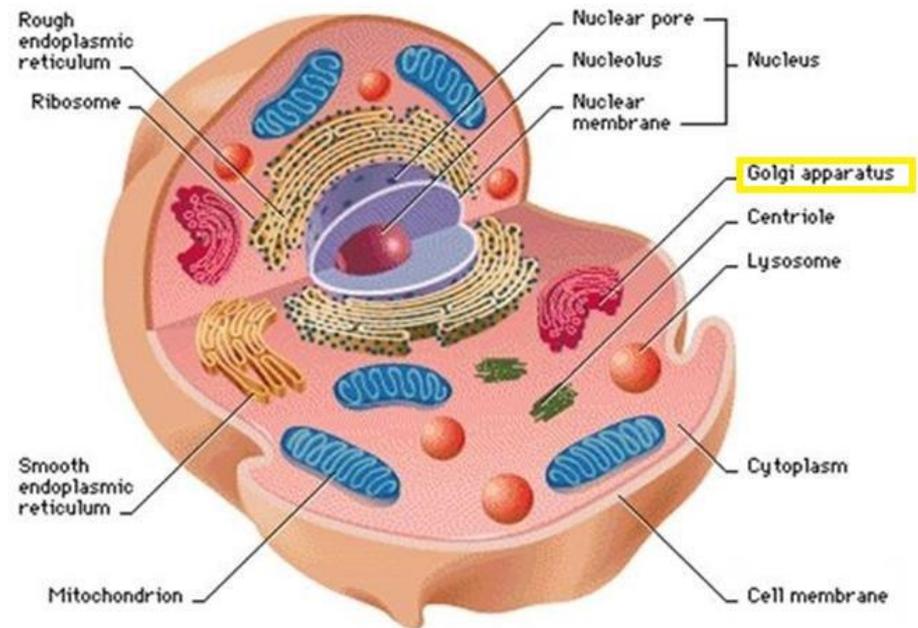


- Chez l'humain, il y a 2 copies de chaque chromosome (1 qui provient de la mère et 1 qui provient du père) et il y a 23 paires de chromosomes



L'ADN

- ADN dans différentes parties de la cellule
 - Nuclear (99%)
 - Mitochondrial
 - Chloroplastes (plantes)



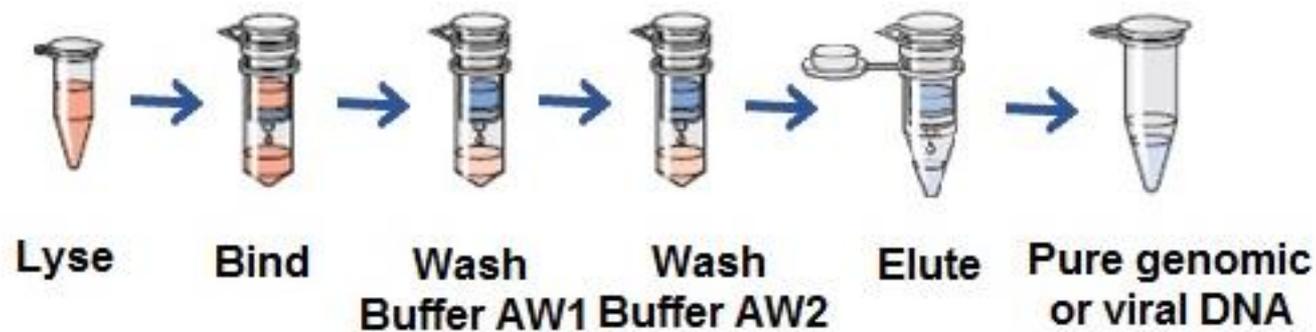
Lire l'ADN:
Extraction, PCR & Séquençage

Extraction, PCR & Séquençage

- Idée: On veut lire la sequence des bases (A,T,C,G) dans un gène
- Extraction → isoler l'ADN de la cellule
- PCR → augmenter la concentration de ce gène
- Séquençage → *lire* les nucleoides de ce gène

Extraction

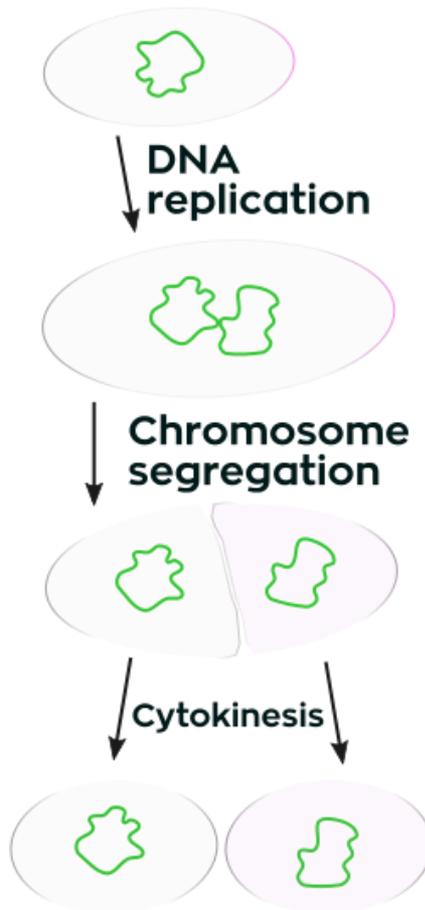
- Commencer avec une petite partie de tissu
 - Pour les poissons, souvent une boucle de nagoire
 - (séché ou en éthanol)



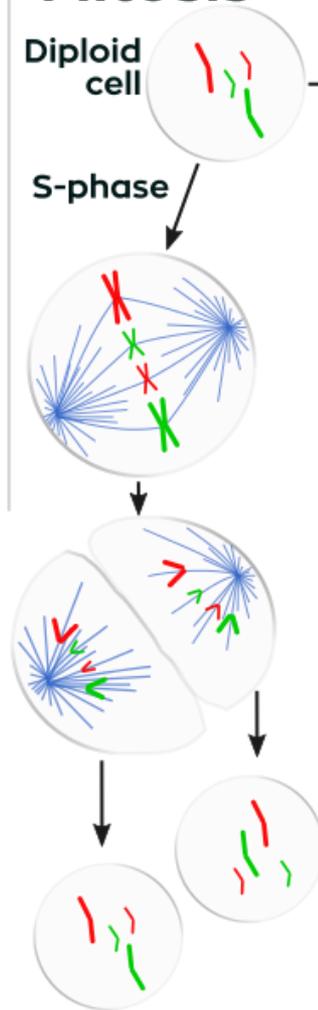
La PCR: Idée

Basé sur la division des cellules

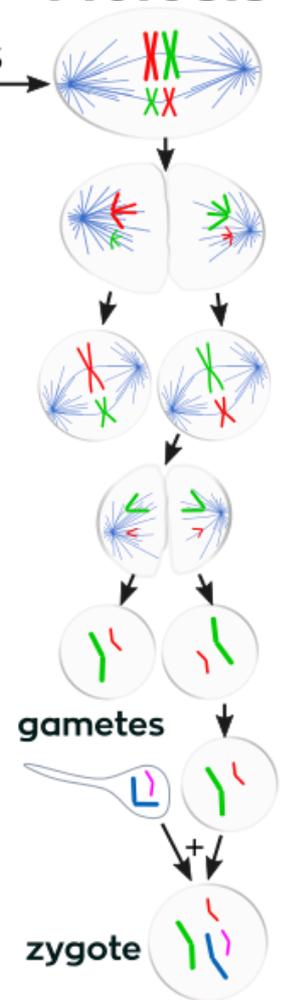
Binary fission



Mitosis

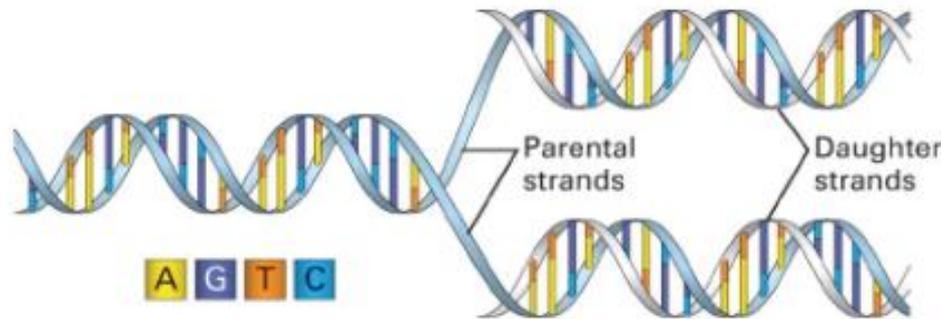


Meiosis



La PCR: La réplication de l'ADN

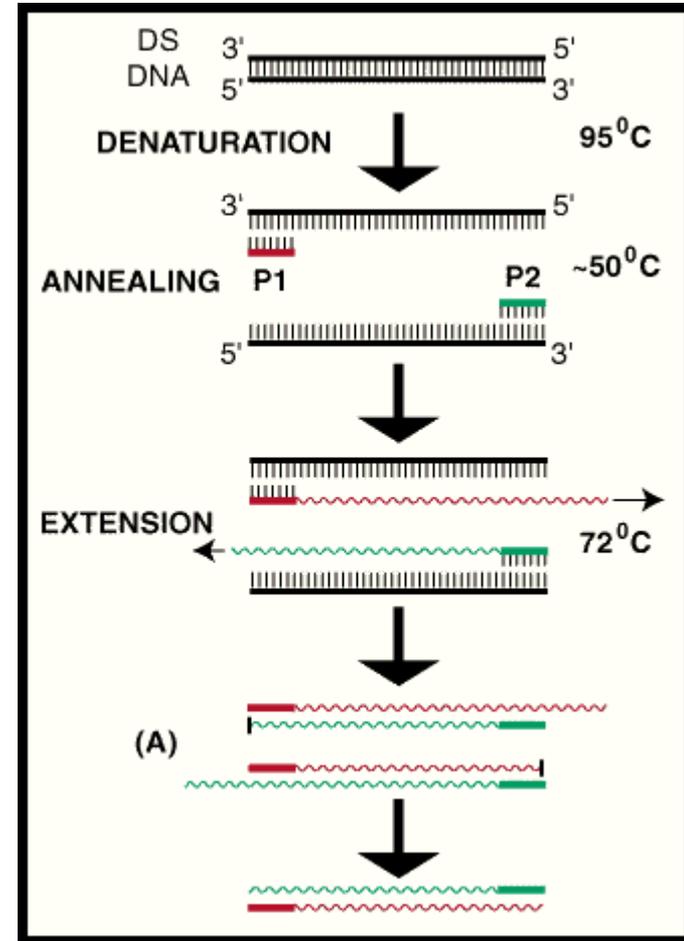
- Durant la réplication, les brins d'ADN se séparent et chaque brin sert de modèle pour la synthèse d'un brin complémentaire



- La réplication assure donc le transfert de l'information génétique d'une cellule à ses descendantes. → polymerase
- Pendant ce procès des erreurs peuvent être faites → mutations!
- Artificiel → PCR (Polyperase Chain Reaction)

La PCR

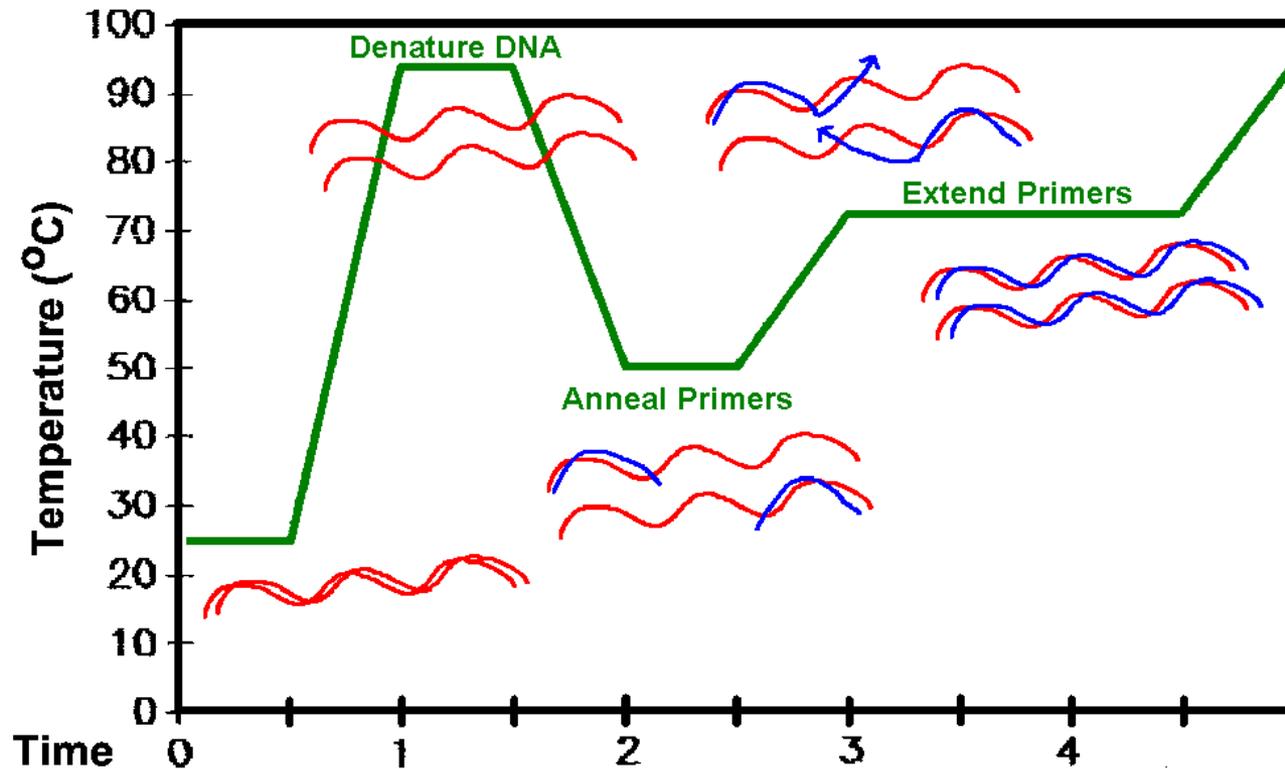
- La PCR se déroule en 3 phases :
 - **Dénaturation** : séparation des deux brins d'ADN grâce à l'élévation de la température
 - **Appariement des amorces** : suite à la diminution de la température, les amorces spécifiques complémentaires de l'ADN à amplifier vont s'hybrider sur le brin d'ADN
 - **Elongation** : synthèse du brin complémentaire grâce à la Taq polymérase qui ajoute les nucléotides (dNTPs)



ADN, nucléotides (dNPT), amorces, Taq polymérase

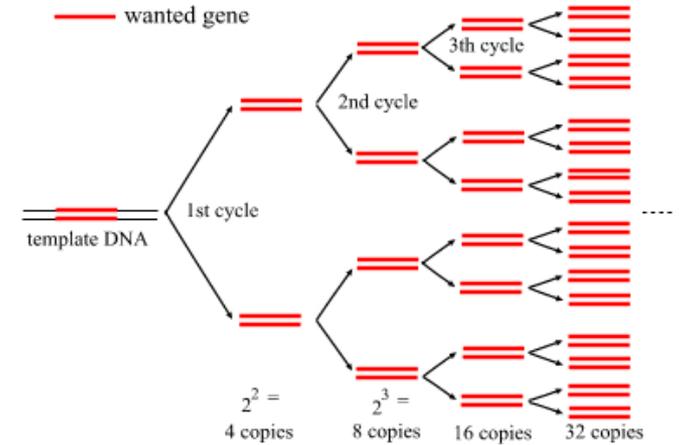
La PCR

- Les différentes étapes de la PCR se déroulent à des températures différentes



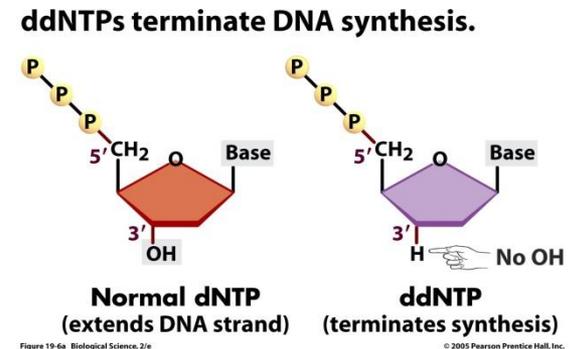
La PCR

- La PCR est une réaction exponentielle qui utilise les produits de chaque étape comme matrice des étapes suivantes. Ce processus va générer des milliers de copies
- En pratique, la réaction se réalise dans des tubes spéciaux qu'on va placer dans un appareil qui va permettre de régler les températures pour chaque étape: un **thermocycleur**



Séquençage de l'ADN

- Cette technique va permettre de connaître la succession des nucléotides qui composent une partie d'une molécule d'ADN
- La technique du **séquençage de Sanger** est la plus utilisée :
 - Semblable au processus de PCR, mais
 - Seulement **un des deux** amorces
 - Dans la pratique: deux réactions
 - l'incorporation de **ddNTP** (didésoxyribonucléotides) et **dNTP**
 - Après ddNTP la polymérase n'est plus capable d'ajouter le moindre nucléotide et la synthèse du brin d'ADN s'arrête
 - Les **ddNTP** sont fluorescents

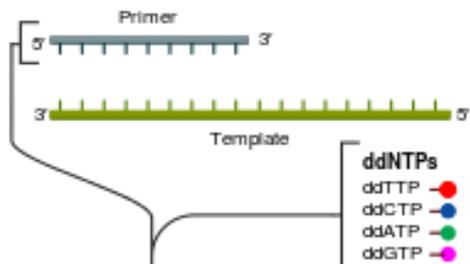


- On obtient un ensemble de brins d'ADN de tailles variées

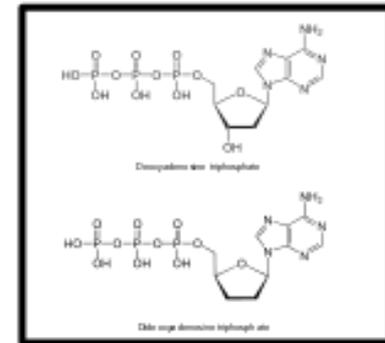
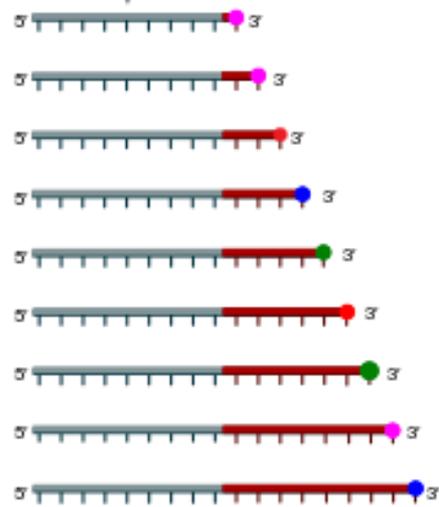
Séquençage de l'ADN

① Reaction mixture

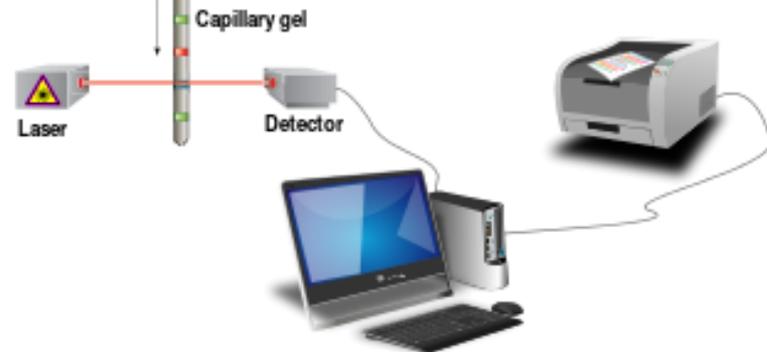
- ▶ Primer and DNA template
- ▶ DNA polymerase
- ▶ ddNTPs with flouochromes
- ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



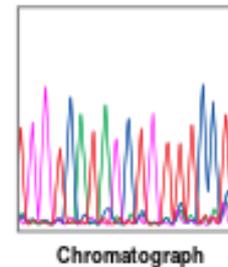
② Primer elongation and chain termination



③ Capillary gel electrophoresis separation of DNA fragments



④ Laser detection of flouochromes and computational sequence analysis



Utiliser l'information génétique
DNA barcoding, phylogénie moléculaire,...

DNA Barcoding

« identification moléculaire »

DNA Barcoding

- Le **DNA Barcoding** ou la méthode du “code-barres ADN” est une technique qui permet une identification moléculaire en se basant sur la comparaison de courtes séquences d’ADN mitochondrial
- Un bon code-barres ADN est :
 - Une séquence variable entre espèces mais très conservée au sein d’une même espèce, ce qui lui confère un fort pouvoir discriminant
 - Une séquence assez courte pour pouvoir séquencer mais suffisamment longue pour avoir assez d’information



En général, on utilise le fragment d’un gène codant pour la première sous-unité de la **cytochrome oxydase** (COI) qui a une longueur de 650 bp

DNA Barcoding

- Le DNA barcoding est utilisé pour différentes raisons :
 - Identification des **espèces cryptiques** dont la morphologie est quasi identique ce qui rend leur distinction difficile
 - Identification de **spécimens incomplets**, afin d'identifier l'espèce à laquelle une touffe de poils ou une patte d'insecte appartient
 - Identification d'espèces à large échelle pour étudier la **biodiversité** et les risques environnementaux
 - Identification des **espèces invasives**
 - Découverte de **nouvelles espèces**
 - **Association de mâles et femelles** d'une même espèce (lorsque les mâles et les femelles sont dimorphiques)
 - Association des **stades de développement** chez une même espèce
 - ...

DNA Barcoding



specimen : MNHN-IC-2009-1050

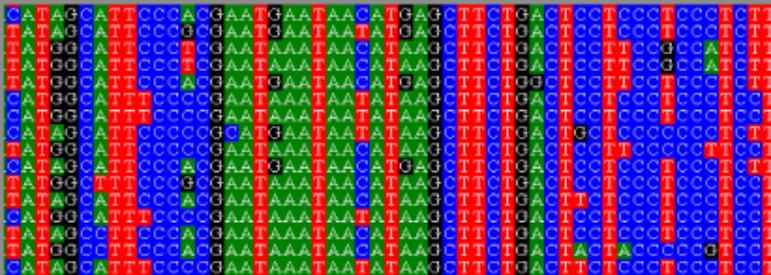
Spécimen à identifier

Extraction d'ADN, amplification du COI, séquençage

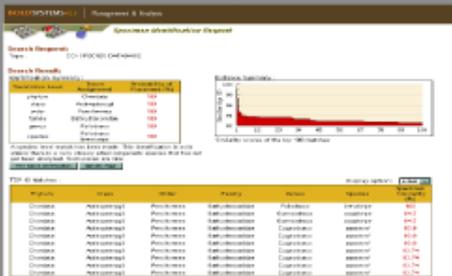


TATAGCATTCCCACGAAATAAATAACATAAGCTTCGACTTCTCCCTCCCTCCT

Comparaison avec le jeu de données de référence dans la base de données BOL



← Séquence identique ou très similaire

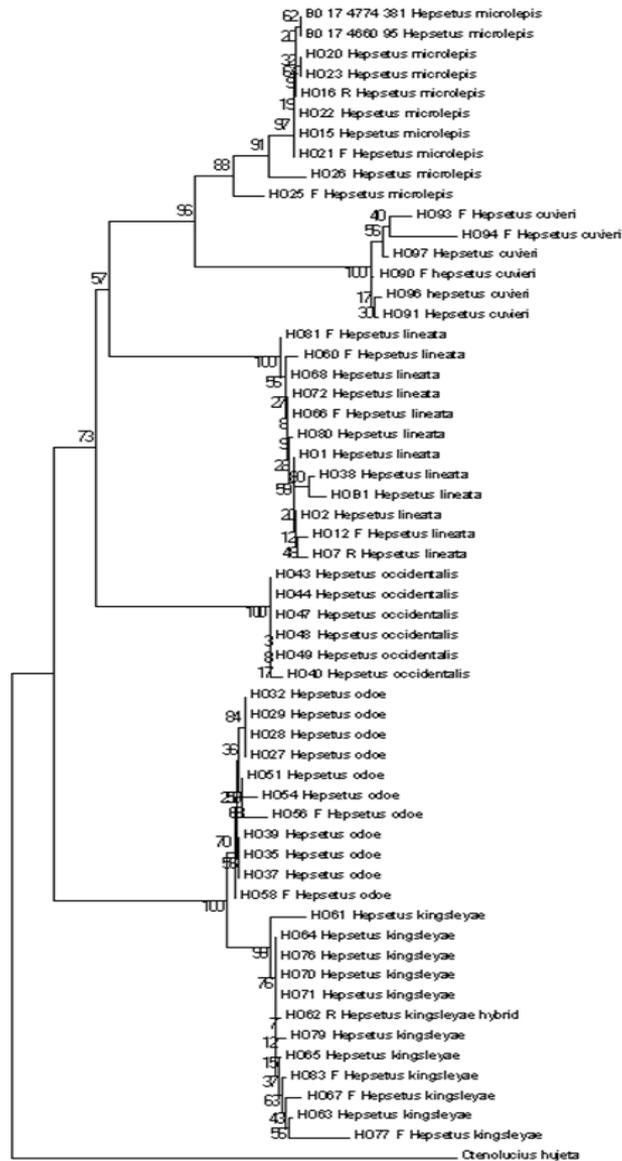


Si une séquence proche est disponible dans la base de données, l'identification est possible

➔ *Psilodraco breviceps*
Espèce identifiée



Confirmation of morphological studies e.g. revision of *Hepsetus*



H. microlepis

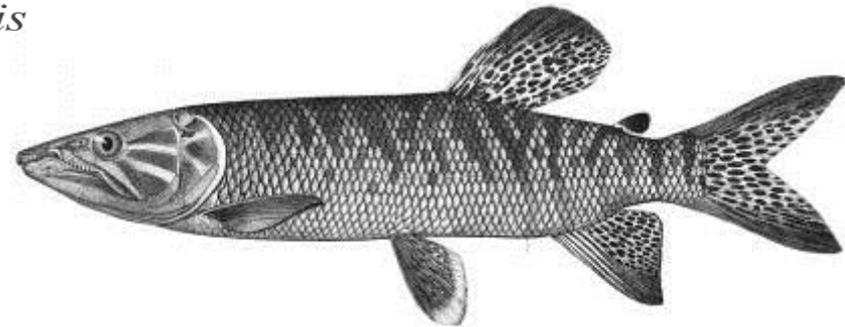
H. cuvieri

H. lineata

H. occidentalis

H. odoe

H. kingsleyae



DNA Barcoding

- **MAIS** on ne peut pas remplacer les analyses taxonomiques par les techniques moléculaire ! Le DNA barcoding peut aider et faciliter le processus d'indentification et permettre la découverte de nouvelles espèces ou répondre à d'autres questions biologiques mais ne peut **en aucun cas** remplacer la taxonomie classique

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

Mehrdad Hajibabaei¹, Gregory A.C. Singer², Paul D.N. Hebert¹ and Donal A. Hickey³

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

- Le barcoding moléculaire est une campagne qui est largement financée à travers le monde, comme le suggère ces différentes organisations :
- « 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 Centre for DNA Barcoding » (CCDB) : <http://www.ccdb.ca/>
- « The Fish Barcode of Life Initiative » (Fish-BOL) : <http://www.fishbol.org/>
-

Genbank

- Database des séquences publiés
- www.genbank.org

NCBI GenBank Overview

PubMed Entrez BLAST OMIM Books Taxonomy Structure

Search Entrez for **eat-4 elegans** Go

NCBI
SITE MAP

Submit to GenBank
Bankit
Sequin

► **What is GenBank?**

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research 2004 Jan 1;32\(1\):23-6](#)). There are approximately 37,893,844,733 bases in 32,549,400 sequence records as of February 2004 (see [GenBank growth statistics](#)). As an example, you may view the [record](#) for a *Saccharomyces cerevisiae* gene. The complete [release notes](#) for the current version of GenBank are available. A new release is made every

Quelques concepts de phylogénie moléculaire

The Phylogenetic Handbook

A Practical Approach to
Phylogenetic Analysis and
Hypothesis Testing

Second Edition

Edited by Philippe Lemey,
Marco Salemi and
Anne-Mieke Vandamme

CAMBRIDGE

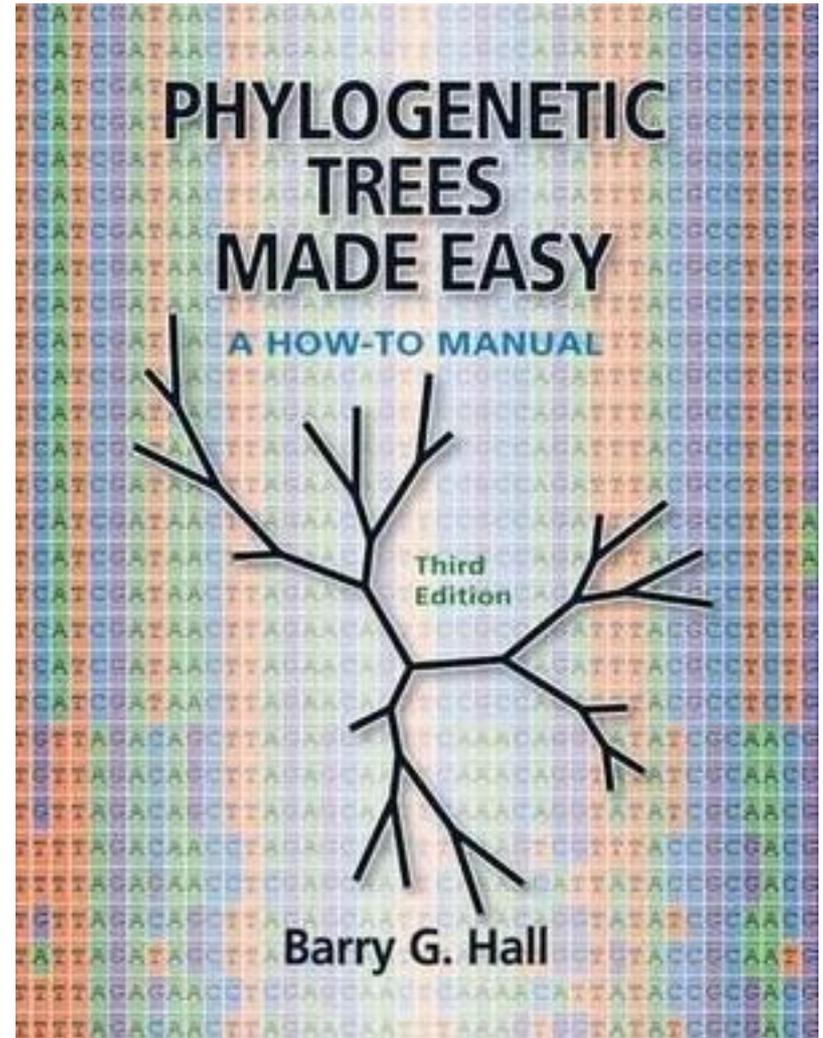


PHYLOGENETIC TREES MADE EASY

A HOW-TO MANUAL

Third
Edition

Barry G. Hall

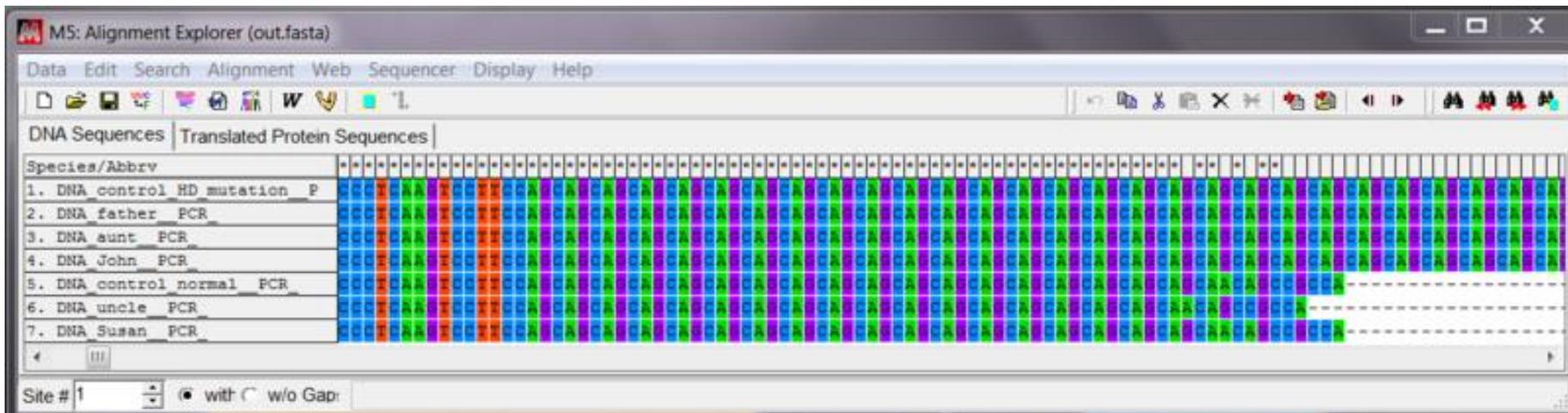


Phylogénie moléculaire

- La phylogénie moléculaire consiste à utiliser des méthodes informatiques pour retracer « l'histoire » des mutations apparues au cours de l'évolution d'un gène donné, en comparant les séquences de différentes espèces plus ou moins proches
- Si on veut comparer les séquences, on doit les aligner, on appelle ça un **alignement de séquences**.

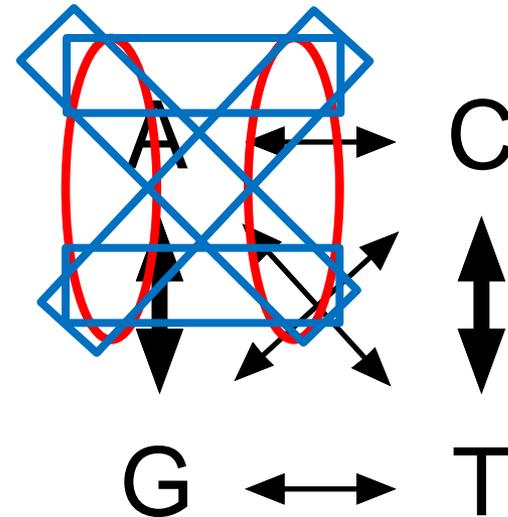
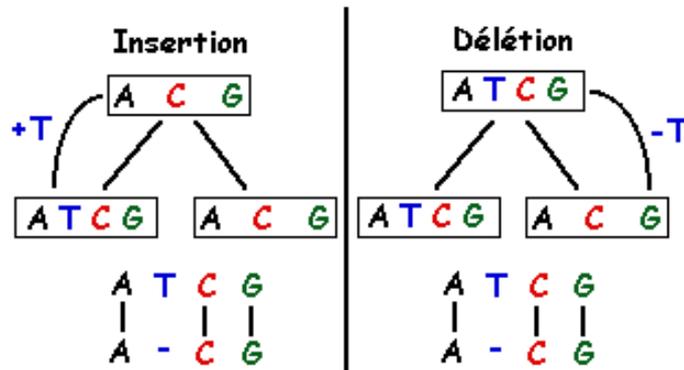
Phylogénie moléculaire

- Un programme gratuit permet de réaliser ces alignements : **MEGA**



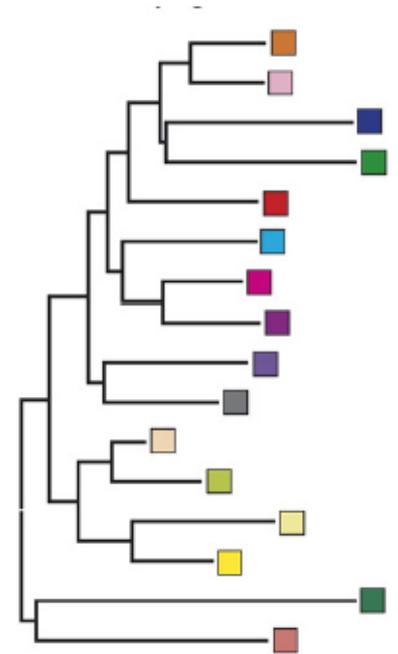
Phylogénie moléculaire

- Après l'alignement, on peut voir quelles mutations ont eu lieu :
 - Transitions entre purines ou entre les pyrimidines
 - Transversions d'un purine vers une pyrimidine
 - Délétions ou insertions



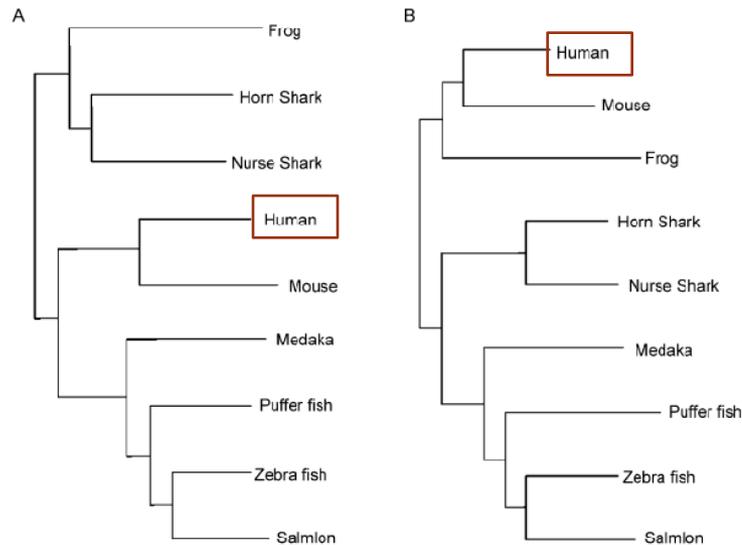
Phylogénie moléculaire

- On peut ensuite construire un **arbre phylogénétique**, en regroupant d'abord les séquences les plus similaires, puis progressivement celles qui sont les plus différentes
- Pour construire les arbres phylogénétiques, on a besoin d'un **modèle d'évolution** des séquences nucléotidiques qui va évaluer :



Phylogénie moléculaire

- Il y a différents modèles d'évolution qui vont analyser les taux de substitutions, la fréquence des bases, le nombre de transitions ou de transversions, les possibilités de mutations avec différents paramètres
- On a des résultats différents avec des modèles d'évolution différents donc il faut choisir le plus adapté !



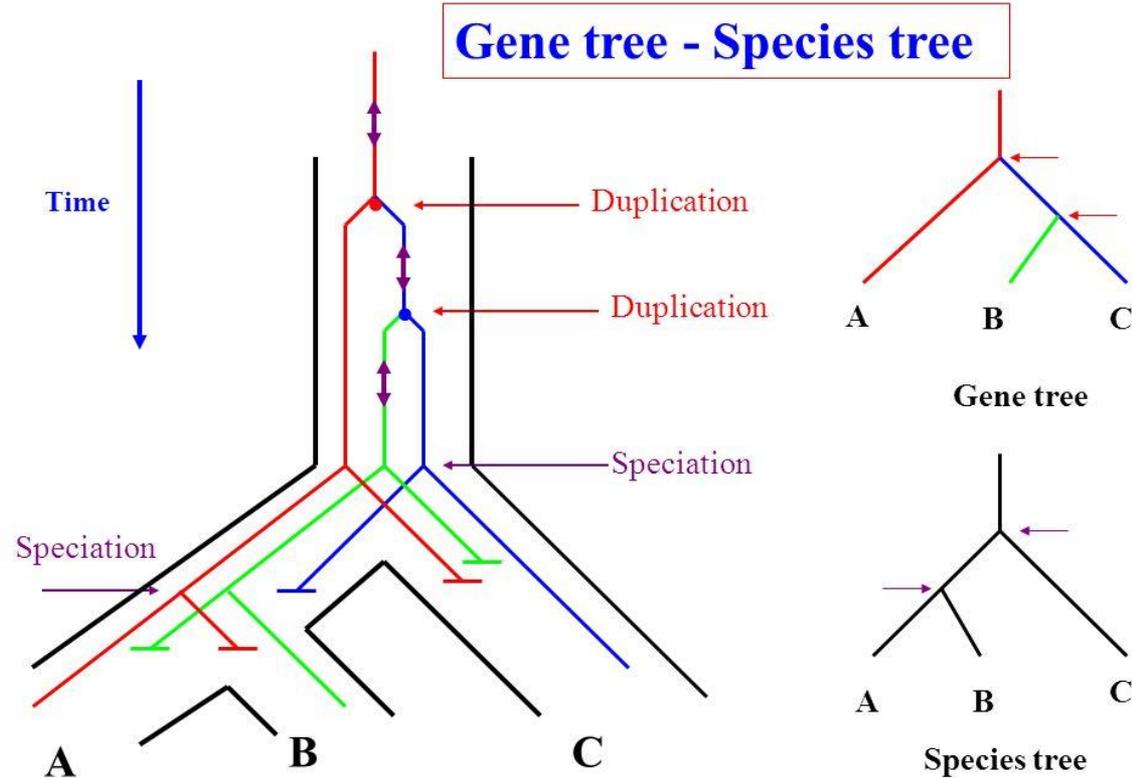
Deux arbres phylogénétiques différents, sur base des mêmes données mais en utilisant deux modèles d'évolution différents !

Phylogénie moléculaire

- Si on a le model correct il y a encore des différents possibilités
 - Les méthodes utilisent des autres techniques pour construire un arbre
 - On a besoin de différents programmes pour utiliser de différents méthodes (Neighbour joining, bayesian analysis, maximum likelihood, maximum parsimony)
- Tous les méthodes vont utiliser des différents manières pour construire des arbres phylogénétique
- Pour plus d'information 'phylogenetic handbook'

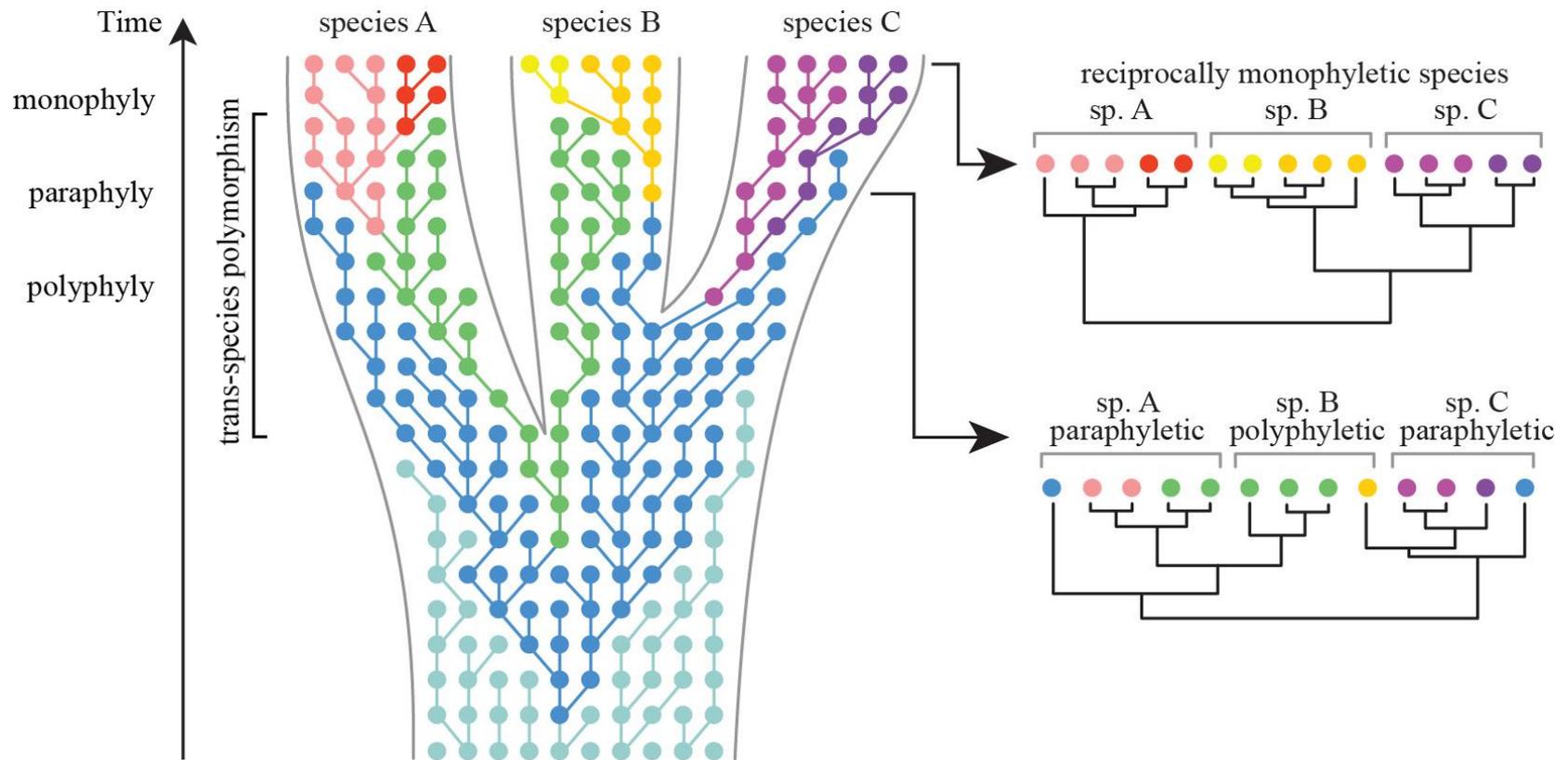
La choix des gènes

- Vitesse de mutation
- Arbre d'un gene = arbre d'espèces?



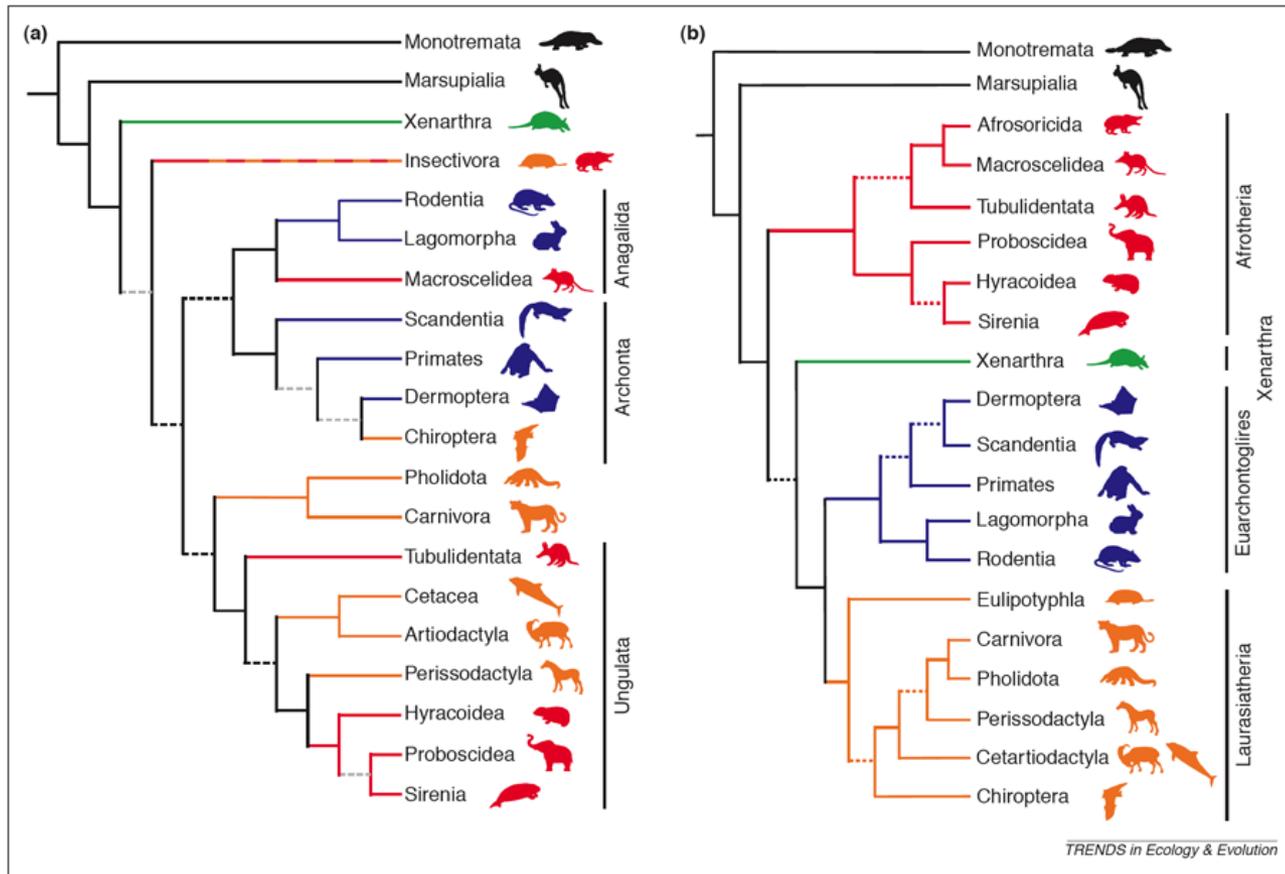
Evolutionary history

L'histoire d'un gène n'est pas nécessairement l'histoire de l'espece



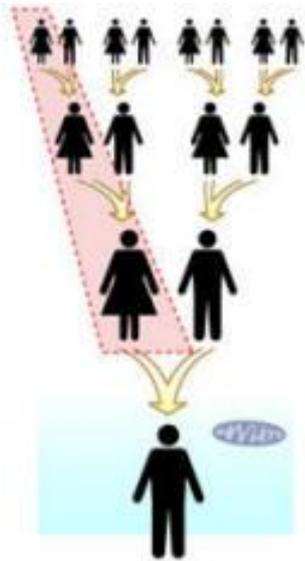
Histoire évolutif

Phylogenies: classifications basés sur multiple lignes d'information

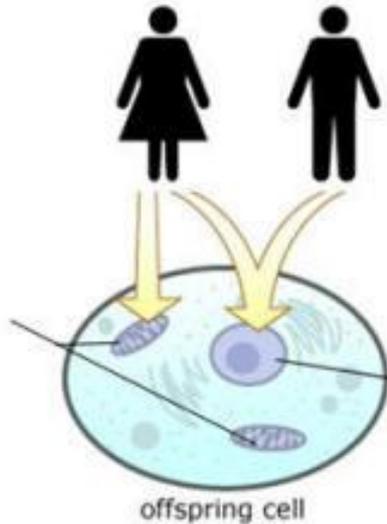


Choix de gènes

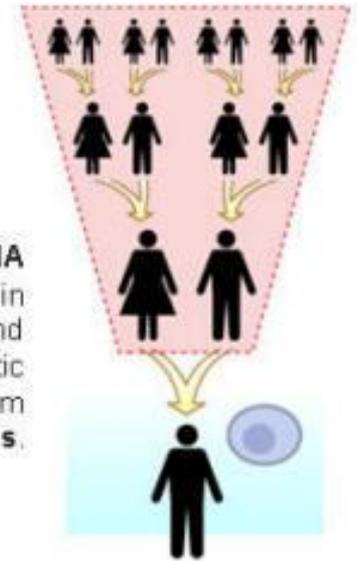
- Mitochondrial ou nuclear?



Mitochondrial DNA (mtDNA) is found in cell mitochondria and contains genetic material only from the **mother**.

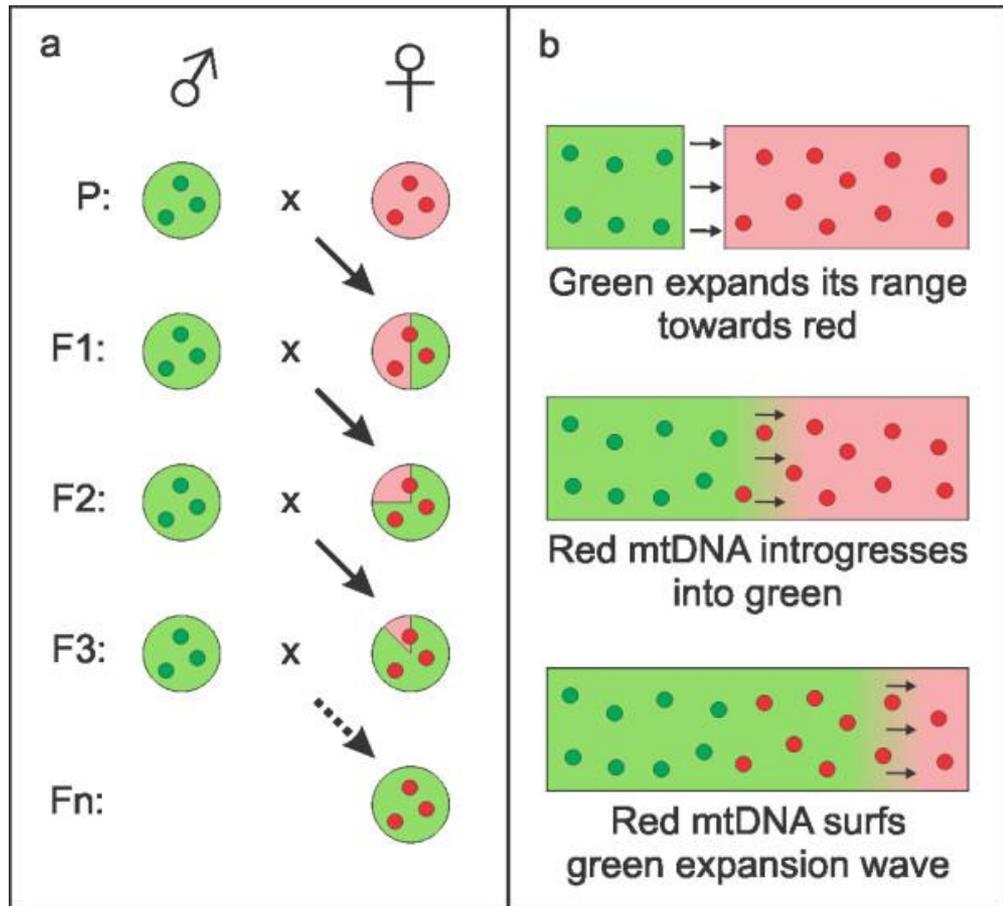
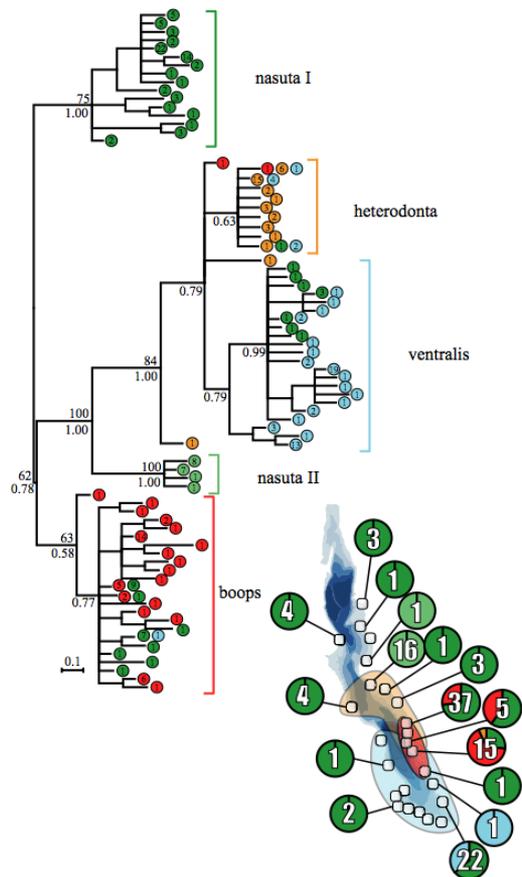


Nuclear DNA (nuDNA) is found in the cell nucleus and contains genetic material from **both parents**.



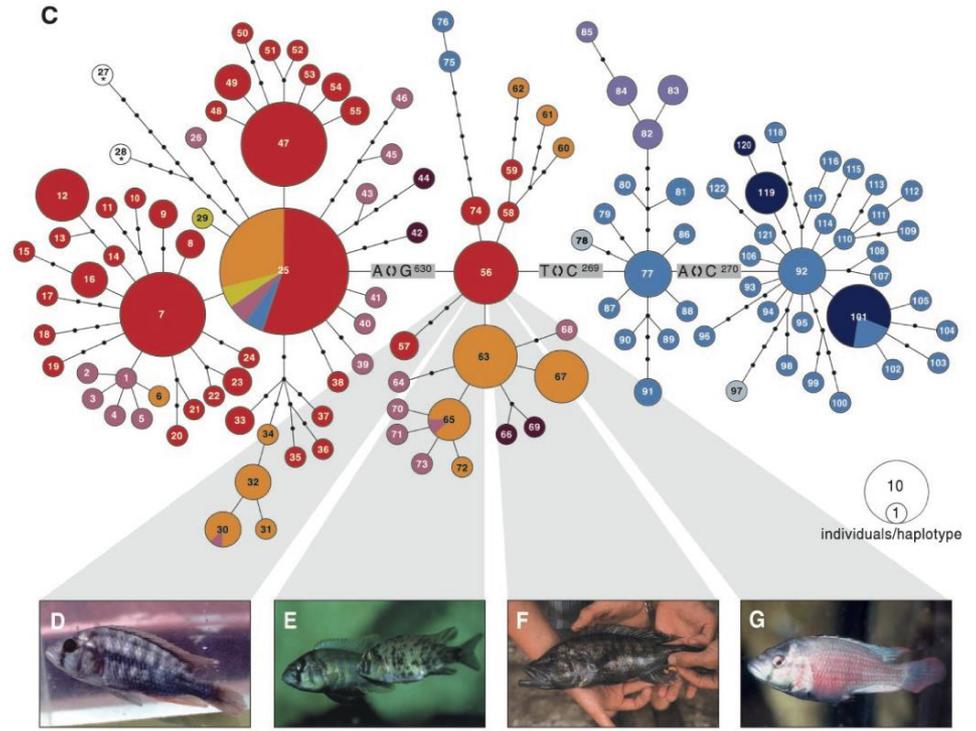
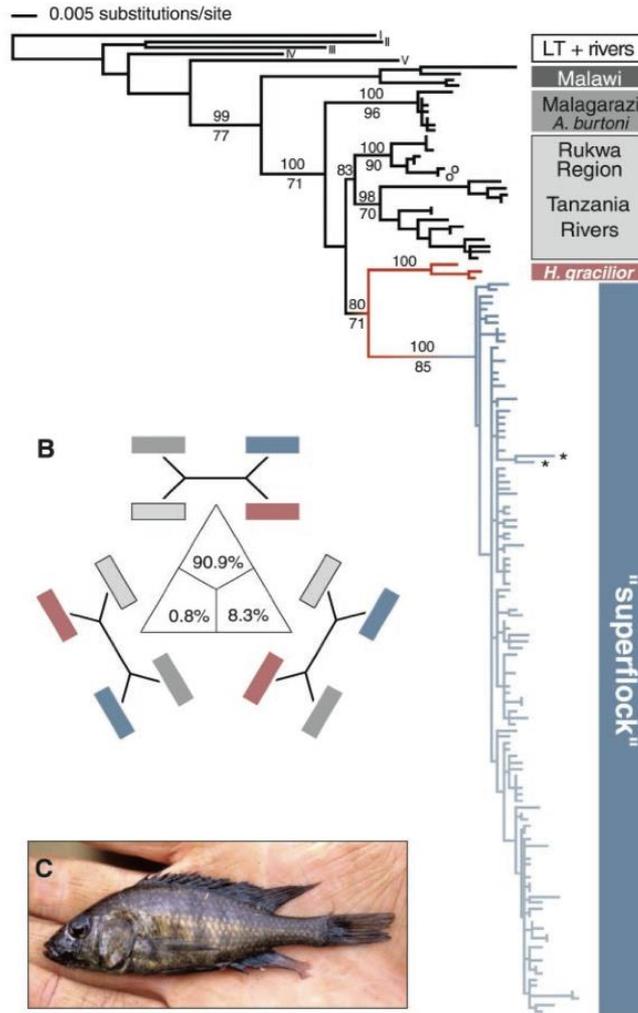
Choix de gènes

- Mitochondrial ou nucléaire?: introgression

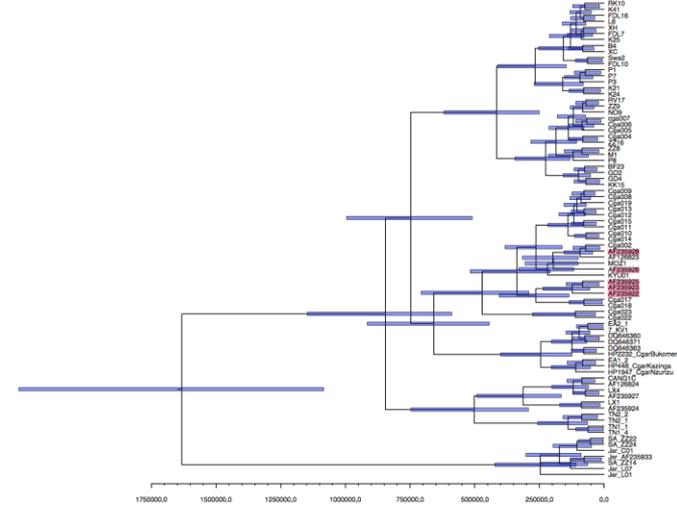
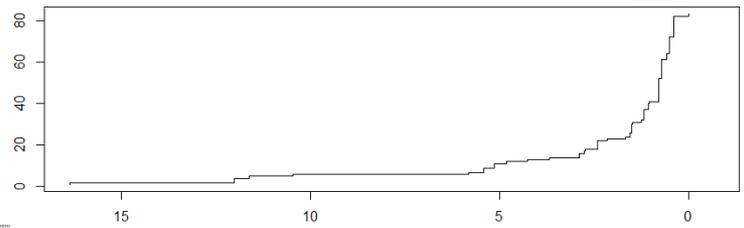
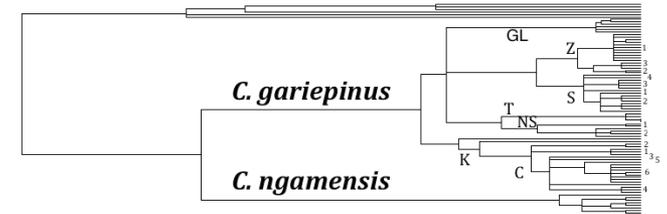
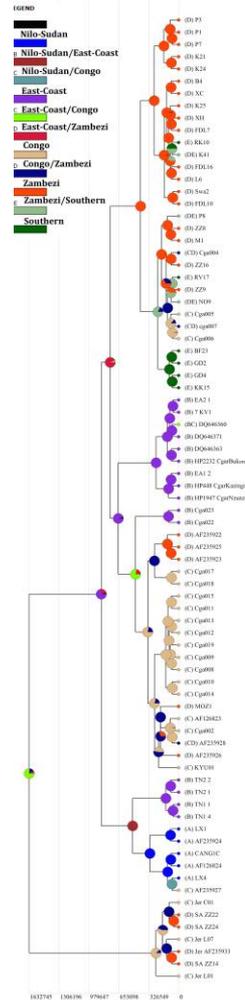
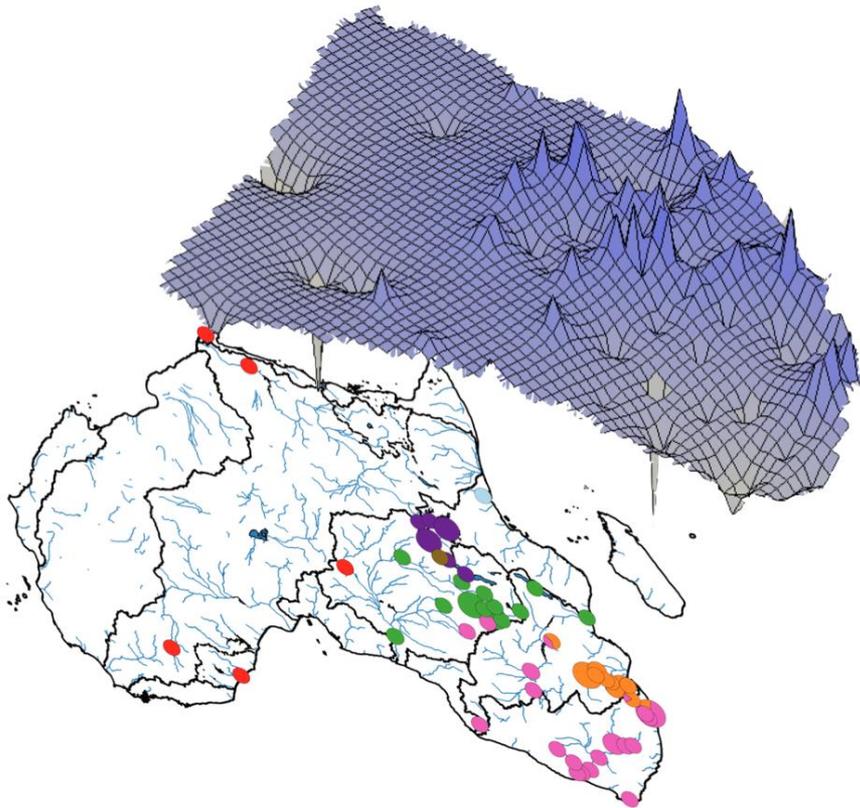


Méthodes alternatives

Visualisation: Trees or networks

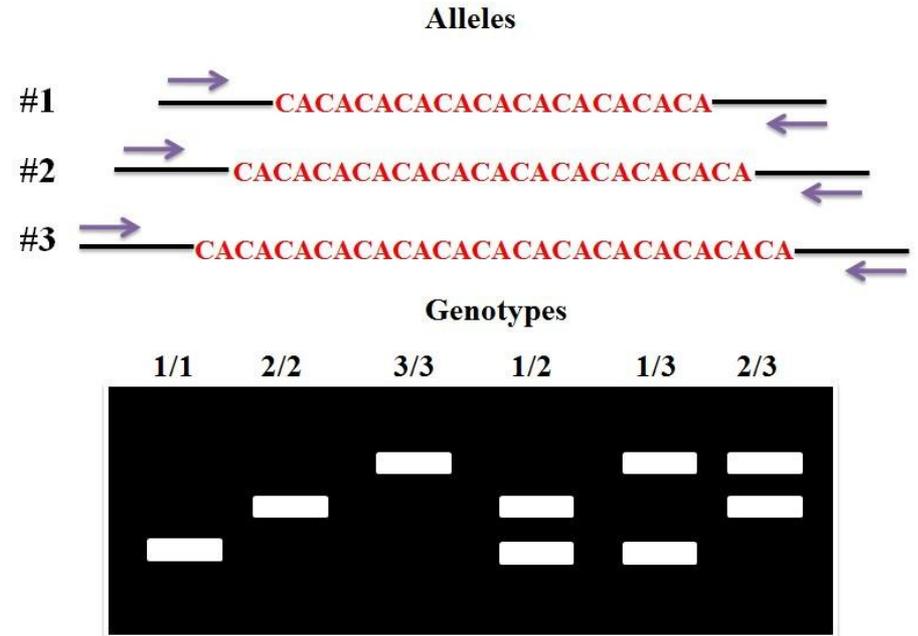


Biogéographie, horloge moléculaire



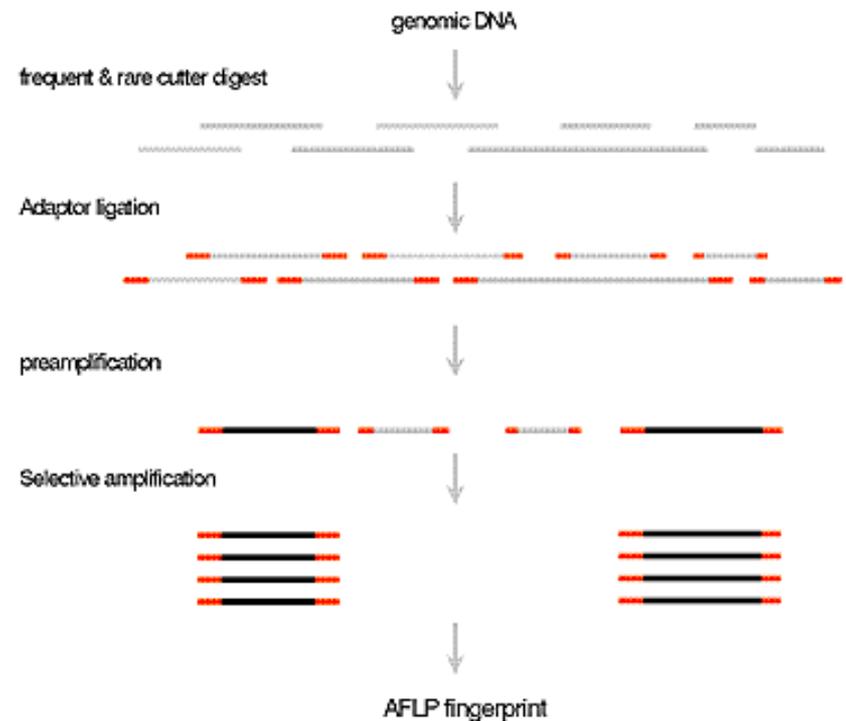
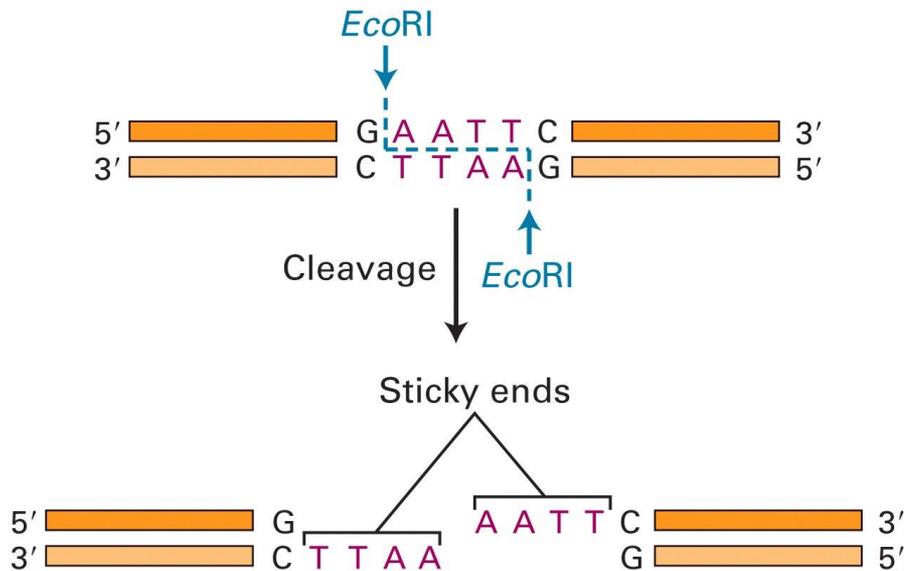
Intra-specific variation

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



Complete genome

- Restriction enzymes
 - AFLP
 - Next generation sequencing



Séquençage de l'ADN

- Cette technique est parfois remplacée par du « **Next Generation Sequencing** », qui permet d'obtenir un génome complet mais :
 - Coût plus élevé
 - Beaucoup plus de données à analyser
 - Intéressant uniquement dans le cadre de gros projet, pas en routine

Next Generation Sequencing (NGS) vs. Sanger Sequencing	
NGS	Sanger
<ul style="list-style-type: none">- High depth of coverage- Fast turnaround time for sequencing larger regions- Cost-effective when sequencing larger regions- Not as effective at sequencing repetitive regions	<ul style="list-style-type: none">- Low depth of coverage- Slow turnaround time for sequencing larger regions- Expensive when sequencing larger regions- More effective at sequencing repetitive regions

Questions?

