

Chapitre 4. Phylogénie moléculaire : retracer l'évolution à partir des séquences

Introduction à la bioinformatique (UE SSV3U15)
2025-2026

Jacques van Helden
Aix-Marseille Université
orcid.org/0000-0002-8799-8584

Contenu de ce chapitre

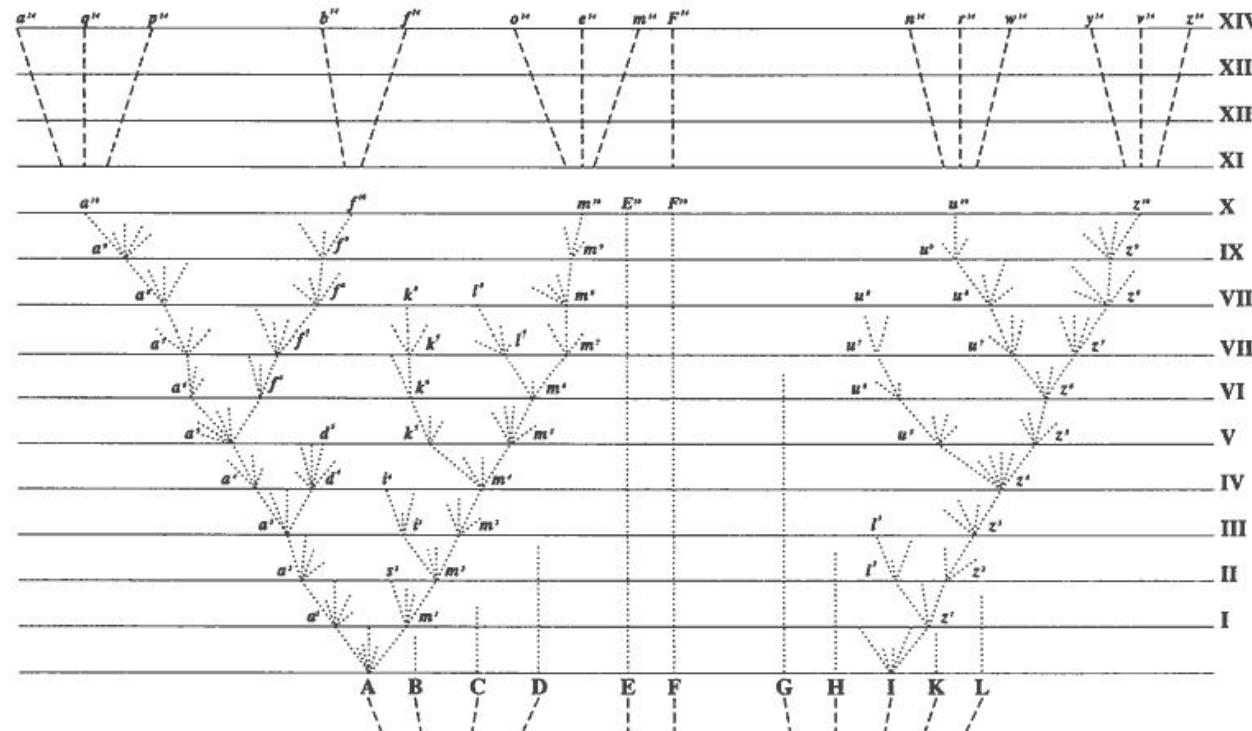
1. Représentations arborescentes de l'évolution
2. Concepts: homologie, analogie, paralogie, orthologie
3. Les duplications à l'origine de l'innovation
4. Phylogénomique : retracer l'évolution des espèces à partir des séquences génomiques
5. Retracer l'origine de SARS-CoV-2 dans les génomes des coronavirus
6. Pseudogènes ("gènes fossiles")
7. Quand les branches de l'arbre du vivant s'entrecroisent

Représentations arborescentes de l'évolution

La divergence des caractères

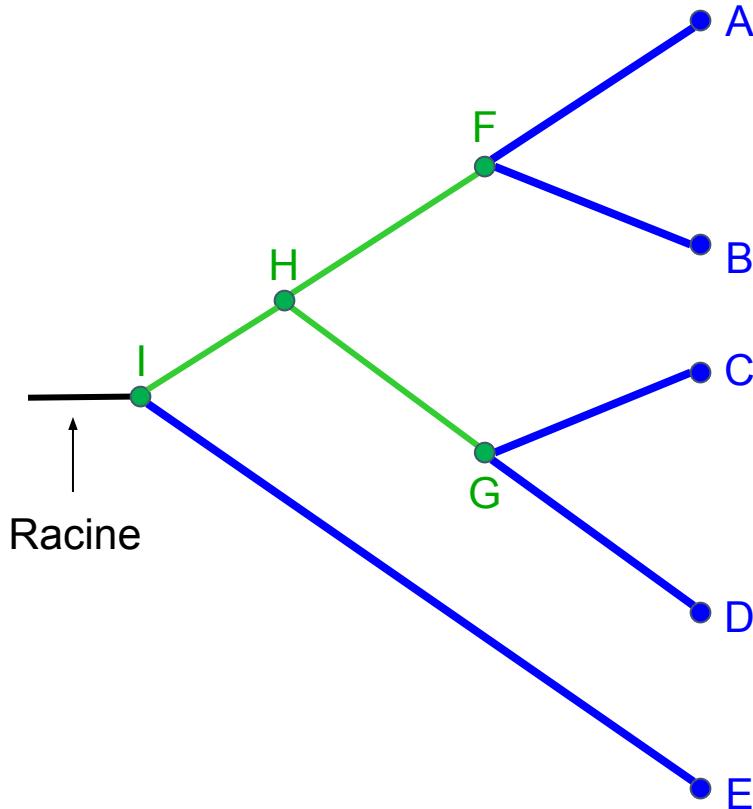
La seule figure de l'Origine des Espèces (C.Darwin, 1859)

- Arbre de la vie
- Synchrone : hauteur = temps
- Evolution graduelle
- Branches abortives = espèces éteintes



Unités taxonomiques opérationnelles (OTU) et hypothétiques (HTU)

- **Topologie** = ensemble des branchements de l'arbre.
- **Noeuds** = unités taxonomiques
 - **OTU** = Unités Taxonomiques Opérationnelles
 - = Noeuds externes = feuille
 - espèces actuelles ou éteintes pour lesquelles on dispose de données
 - **HTU** = Unités taxonomiques Hypothétiques = noeuds internes
 - on ne dispose pas de données
 - correspondent aux dernier ancêtre commun entre plusieurs OTU.
- Branches = relations de parenté(ancêtre/descendants) entre unités taxinomiques
 - Branches internes
 - Branches externes
-



Représentations arborescentes des histoires évolutives

• Cladogramme

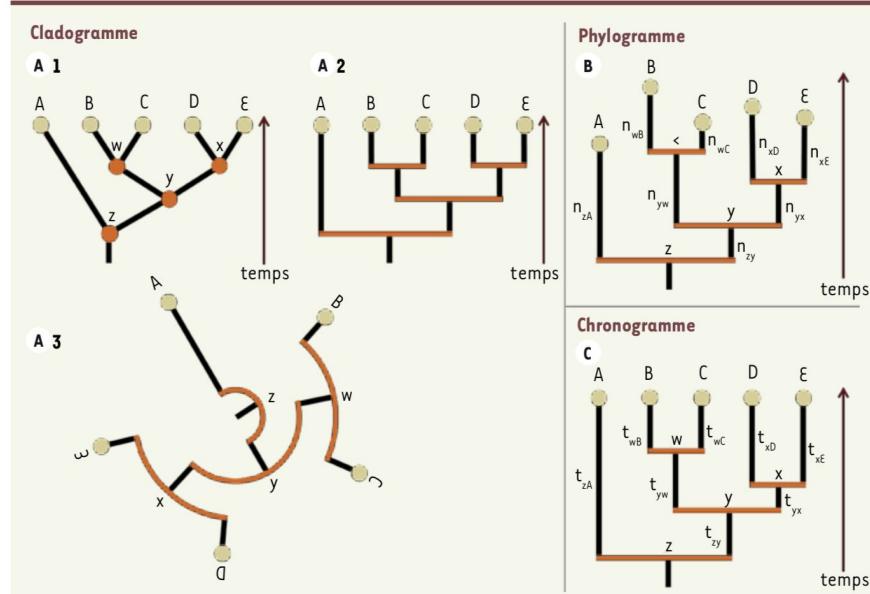
- Représente uniquement les branchements successifs
- Longueurs des branches non-informatives

• Phylogramme

- Longueurs des branches: **distance évolutive** (différences génétiques ou morphologiques)
- Certaines branches peuvent évoluer plus rapidement que d'autres → feuilles de l'arbre (OTU) pas forcément alignées

• Chronogramme

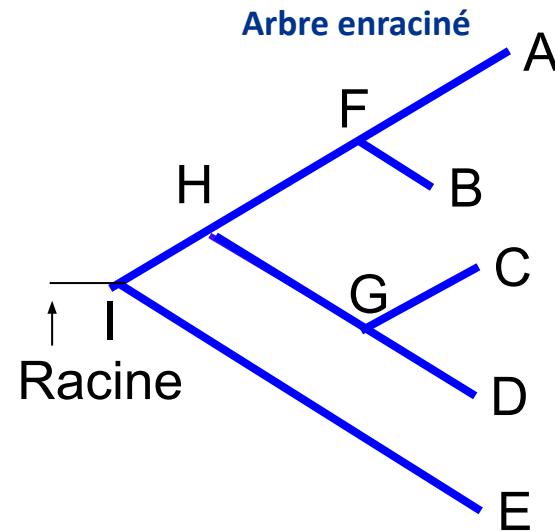
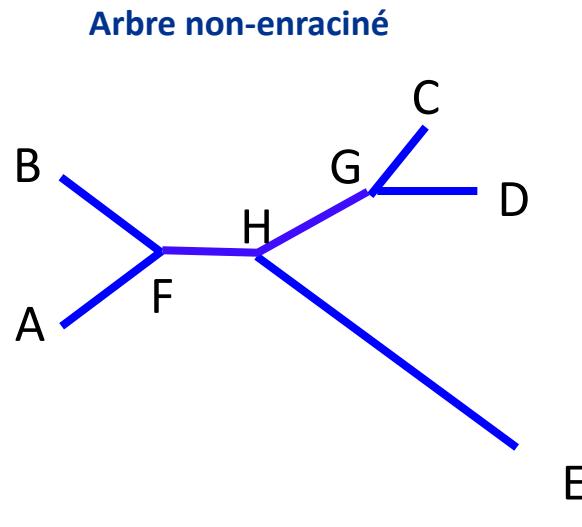
- Longueur des branches: **temps de divergence**.
- Si OTU contemporaines, les feuilles sont alignées



Casane, D. & Laurenti, P. Penser la biologie dans un cadre phylogénétique: L'exemple de l'évolution des vertébrés. Med Sci (Paris) 28, 1121–1127 (2012).
doi.org/10.1051/medsci/20122812024

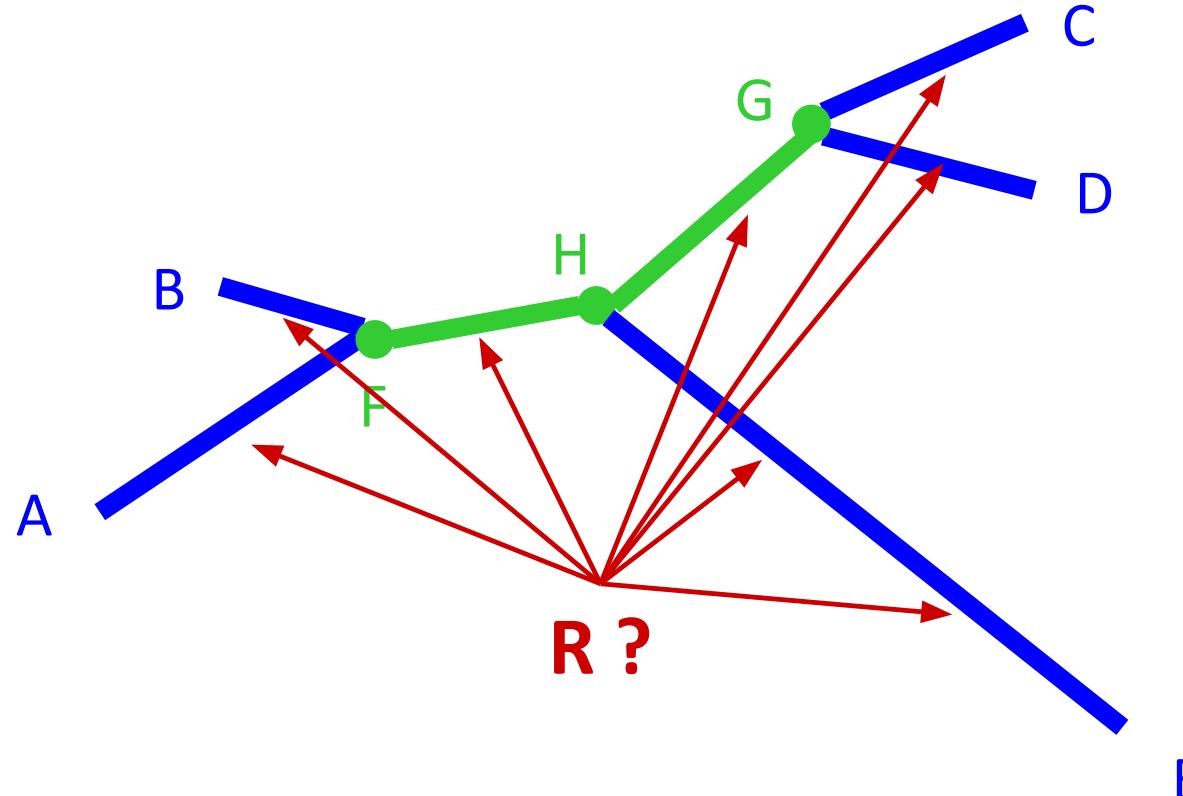
Arbres enracinés ou non enracinés

- **Arbre non-enraciné**
 - pas de direction temporelle
 - → indique les **distances**, mais pas les **relations de parenté** entre les noeuds.
- **Arbre racinés**
 - La **racine** symbolise le **dernier ancêtre commun** (l'ancêtre commun plus récent) de toutes les OTU.



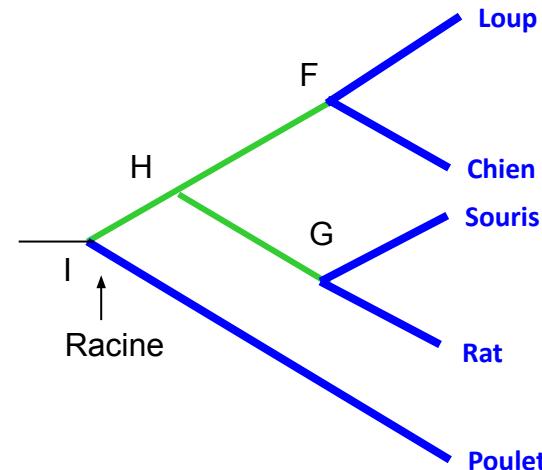
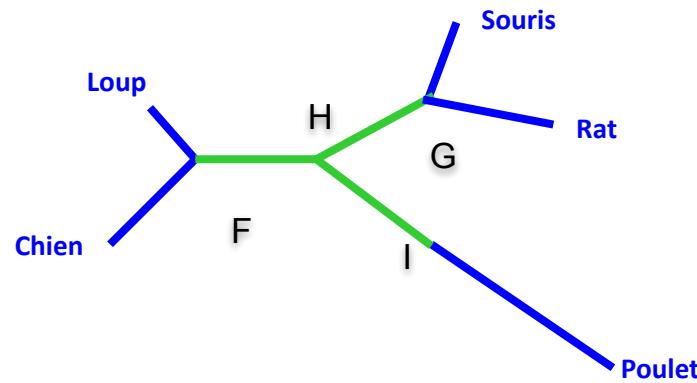
Comment enracer un arbre phylogénétique ?

- A priori, la racine pourrait se situer sur à n'importe quelle position de n'importe quelle branche de l'arbre.
- Cependant, une seule de ces positions correspond à la véritable histoire évolutive.



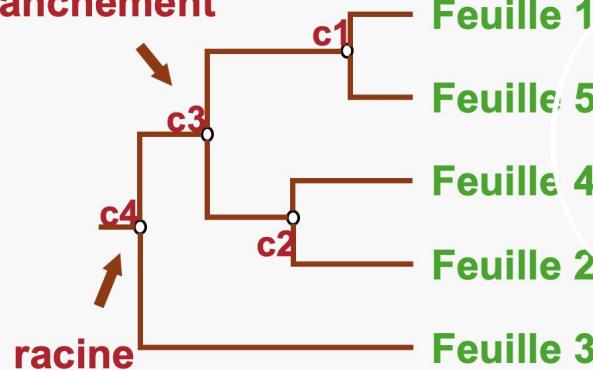
Approches pour enracer un arbre phylogénétique

- **Enracinement au poids moyen**
 - Aucune garantie de pertinence
- Inclusion d'un **groupe externe (outgroup)** dans les données
 - Enracinement sur la branche qui sépare ce groupe de tous les autres



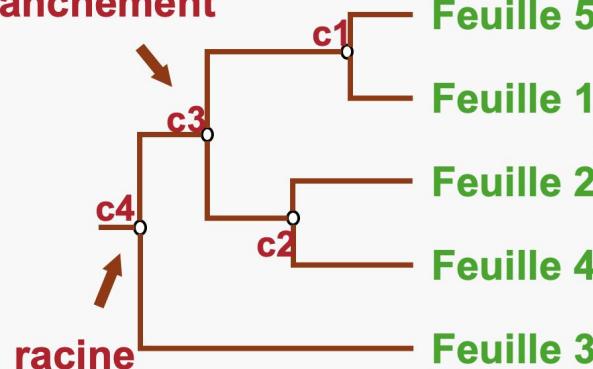
Isomorphisme des arbres phylogénétiques

branchement



- Arbres *isomorphiques*, considérés équivalents
- L'inversion verticale des branches d'une bifurcation ne modifie pas l'information.

branchement



Concepts: homologie, analogie, paralogie, orthologie

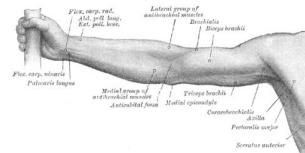
Homologie et analogie

• Homologie

- ressemblances héritées d'un **ancêtre commun**
- évolution **divergente**: accumulation de différences depuis la séparation des deux lignées

Homologie

Humain (mammifère)



Chimpanzé (mammifère)



• Analogie

- ressemblances résultent de trajectoires **indépendantes**
- **évolution convergente** : similarité augmente avec le temps
- peut éventuellement manifester l'effet d'une pression évolutive similaire

Analogie

Poule (oiseau)



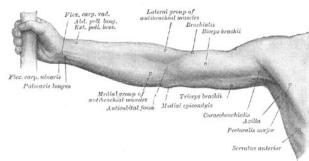
Mouche (insecte)



Homologie et analogie : un exemple non trivial

- Les **membres antérieurs** des chauves-souris sont homologues à ceux des oiseaux.
 - caractère ancestral** d'un groupe taxonomique = hérité du dernier ancêtre commun à ce groupe
 - commun à tous les tétrapodes
 - hérité à partir des premiers vertébrés terrestres.
- Cependant, les **ailes** des chauves-souris sont analogues aux ailes des oiseaux
 - caractère dérivé**: apparu ultérieurement à la séparation des lignées depuis leur dernier ancêtre commun

Humain (mammifère)



Chimpanzé (mammifère)



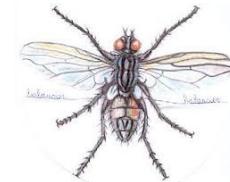
Poule (oiseau)



Pigeon (oiseau)



Mouche (insecte)



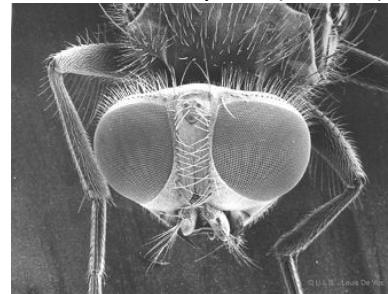
Chauve-souris (mammifère)



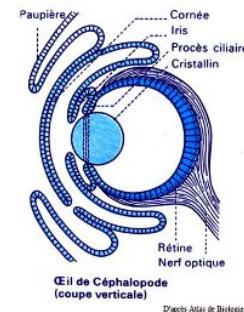
Structures analogues et convergence évolutive

- La vision des différents groupes d'animaux repose sur des yeux de structures très diverses.
- L'oeil à facettes des insectes est très différent de l'oeil des vertébrés
- L'oeil de pieuvre présente de fortes similarités de structures avec l'oeil humain, mais quelques différences notoires
 - **Similarités:** oeil sphérique, cornée, iris, cristallin, ...
 - **Défauts:** orientation des cellules rétinienques: les axones partent vers l'intérieur chez les vertébrés, vers l'extérieur chez les céphalopodes
- En dépit de leur ressemblance anatomique, l'oeil de pieuvre et l'oeil humain résultent de **voies évolutives indépendantes**. Leur ressemblance est due à une **convergence évolutive** plutôt qu'à une origine commune. Il s'agit d'un cas spectaculaire de ressemblance par **analogie**.

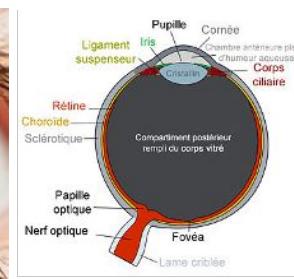
Oeil de drosophile (insecte)



Oeil de pieuvre (mollusque céphalopode)



Oeil humain (mammifère)

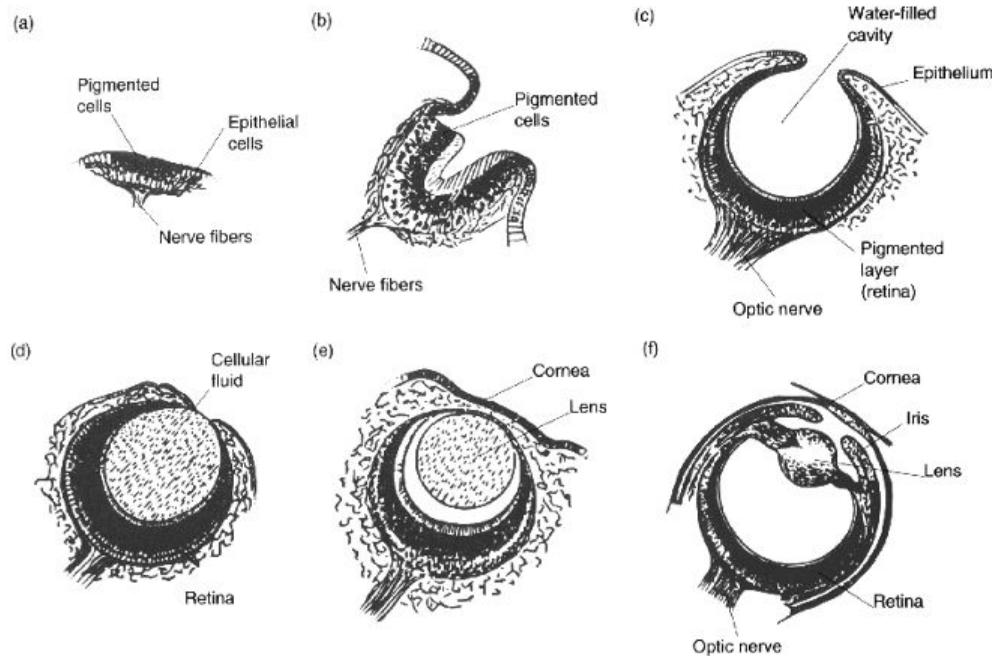


Figures:

<https://www.futura-sciences.com/sante/dossiers/medecine-oeil-vision-dela-vision-667/page/2/>

- Structures d'yeux de mollusques
- Complexité croissante
→ une évolution graduelle est capable de produire des structures extrêmement complexes

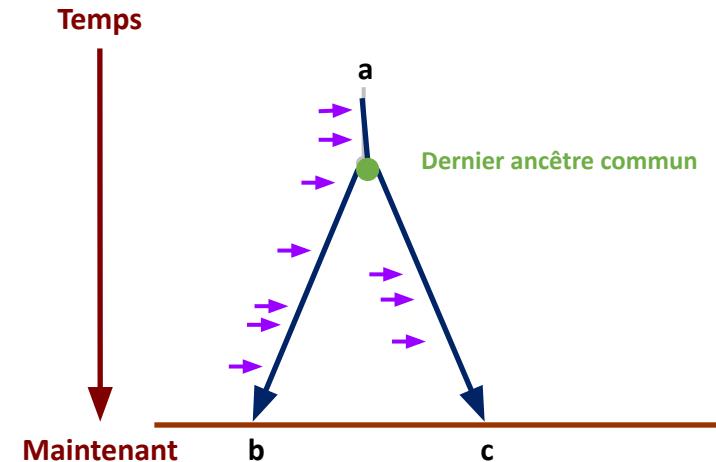
Figure 13.2 Stages in the evolution of the eye, illustrated by species of molluscs. (a) A simple spot of pigmented cells. (b) Folded region of pigmented cells, which increases the number of sensitive cells per unit area. (c) Pin-hole camera eye, as is found in *Nautilus*. (d) Eye cavity filled with cellular fluid rather than water. (e) The eye is protected by adding a transparent cover of skin, and part of the cellular fluid has differentiated into a lens. (f) Full, complex eye, as found in octopus and squid. Reprinted, by permission of the publisher, from Strickberger (1990).



Séquences homologues

■ Homologie

- ressemblances héritées d'une **séquence ancestrale commune**
- évolution **divergente**: accumulation de mutations à partir de la séquence ancestrale commune



Les séquences b et c sont homologues

- Nous ne disposons que des séquences actuelles → on ne peut pas prouver l'homologie
- On peut par contre *inférer* l'homologie sur base d'une estimation de la vraisemblance

Alignement de séquences

- calcul des taux d'identités / similarité
- significativité statistique

E-valeur élevée (>1) ?

- similarités vraisemblablement dues au hasard

E-valeur très faible (ex : 1e-5, 1e-100, ...) ?

- très significatif
- très improbable que la similarité résulte du hasard
- hypothèse alternative : la ressemblance entre les séquences résulte d'un héritage commun (à partir d'une séquence ancestrale)

L'homologie est une relation logique (soit vraie, soit fausse)

- Homologie = tout ou rien
 - Formulations inappropriées
 - “Ces séquences ont 50% d'homologie”
 - “Niveau d'homologie”



Formulation correcte

- Ces séquences sont vraisemblablement homologues car elles ont taux de similarité trop significatif pour s'expliquer par une coïncidence ou par une convergence évolutive.

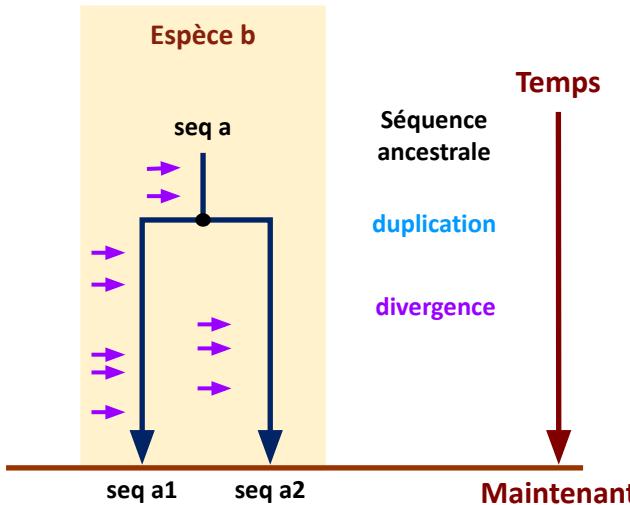
Evénements évolutifs générant des séquences homologues

- **Duplication**
 - Dédoublement d'une partie de l'ADN génomique.
 - Taille variable : quelques nucléotides, un ou plusieurs gènes, un chromosome entier, le génome entier
 - Crée de la redondance → plusieurs suites possibles
 - Délétion d'une copie
 - Mutations → on trouve dans le génome des traces d'un gène non-fonctionnel ("gène fossile", pseudo-gène)
 - Mutations → divergence, les deux copies restent fonctionnelles et se spécialisent
- **Spéciation**
 - Formation d'espèces distinctes à partir d'une espèce unique.

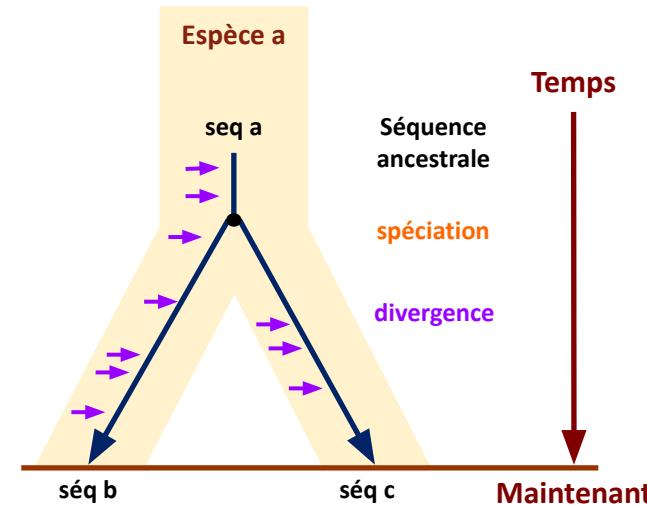
Scénarios évolutifs

- Séquences communes
- Deux scénarios alternatifs
 - **duplication** (création de deux copies du gène dans le même génome)
 - **spéciation** (formation d'espèces séparées à partir d'une espèce unique).
- Accumulation de **mutations** → diversification des séquences, des structures et des fonctions

Duplication



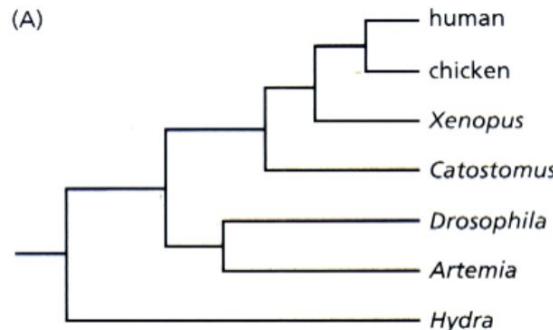
Spéciation



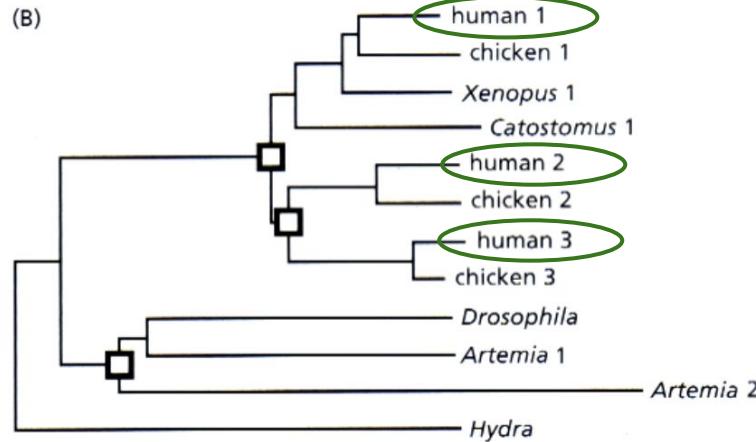
Arbre des espèces et arbre des molécules

- Arbre des espèces → chaque noeud interne représente une spéciation
- Arbre des molécules → chaque noeud interne représente soit une duplication, soit une spéciation

L'arbre des espèces



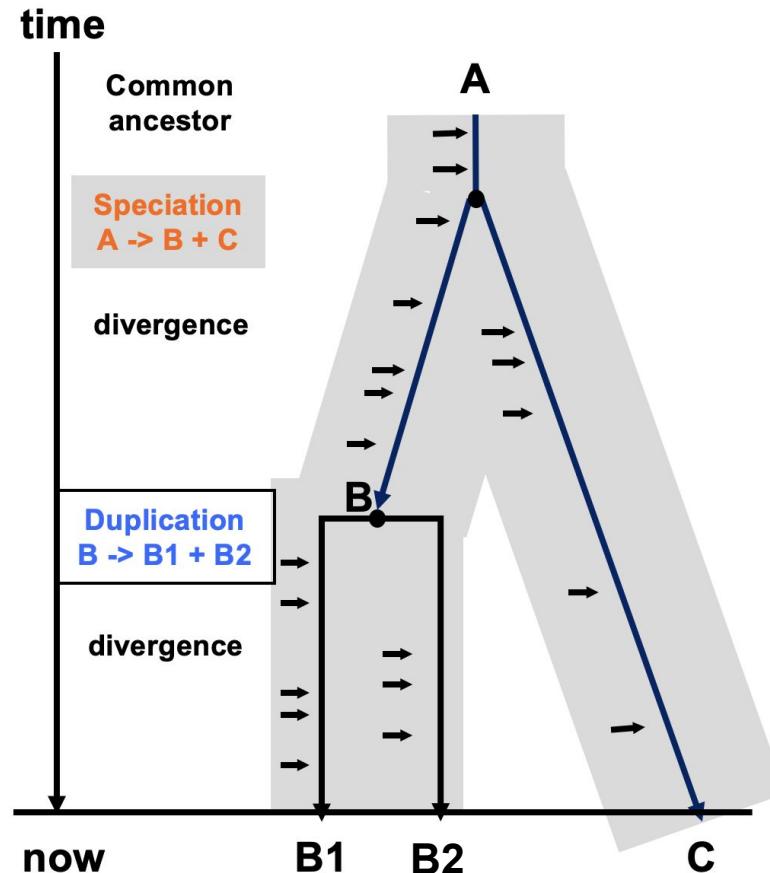
Arbre des molécules



Orthologie versus paralogie

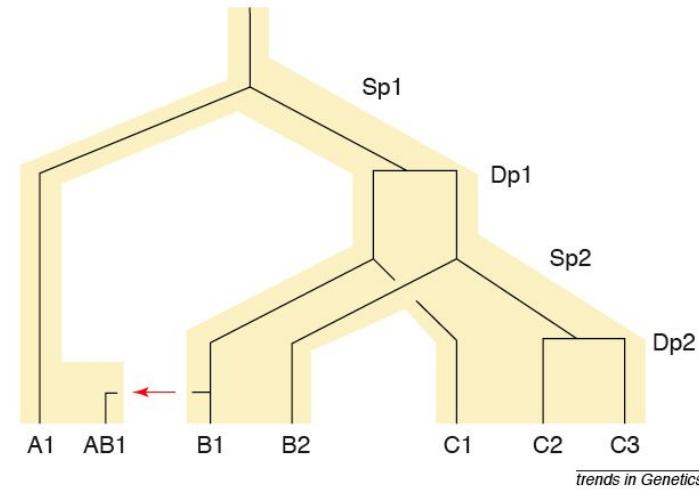
Définitions de Zvelebil (claires et opérationnelles)

- **Orthologues:** séquences dont le dernier ancêtre commun précède immédiatement un événement de spéciation.
- **Paralogues** séquences dont le dernier ancêtre commun précède immédiatement un événement de duplication
- Exemples:
 - B et C sont *orthologues*
 - B1 et B2 sont *paralogues*



Représentation détaillée des événements de spéciation / duplication

- **Arbre des molécules** : lignes noires
- **Arbre des espèces** : fond jaune
- **Spéciations (Sp)** : branchements triangulaires
- **Duplications (Dp)** branchements rectangulaires



The idealized evolution of a gene (lines) is shown from a common ancestor in an ancestral population (the gray background), descending to three populations labelled A, B and C. There are two speciation events (Sp1 and Sp2), each occurring at the junctions shown as an upside down Y. There are also two gene-duplication events (Dp1 and Dp2), depicted by a horizontal bar. Two genes whose common ancestor resides at a Y junction (speciation) are orthologous. Two genes whose common ancestor resides at a horizontal bar junction (gene duplications) are paralogous. Thus, C2 and C3 are paralogous to each other but are orthologous to B2. Both are paralogous to B1 but orthologous to A1. The red arrow denotes the transfer of the B1 gene from species B to species A. As a result, the AB1 gene is xenologous to all six other genes. All three subtype relationships are reflexive, that is, $A1 \Rightarrow B1$ implies $B1 \Rightarrow A1$ where \Rightarrow should be read, for example, as 'is orthologous to.' However, the relationships are not transitive. Thus, $C2 \Rightarrow A1 \Rightarrow C3$ might be true, but it is not necessarily therefore true that $C2 \Rightarrow C3$, as indeed it is not in the figure if \Rightarrow is read as 'is orthologous to.' A different non-transitivity occurs for 'is paralogous to' with $B2 \Rightarrow C1 \Rightarrow C2$.

Exercice: types d'homologie

Sur base des définitions de **Zvelebil & Baum** (paralogie et orthologie), et de **Fitch** (xénologie), qualifiez la relation entre chaque paire de gènes dans le schéma de Fitch (ci-contre).

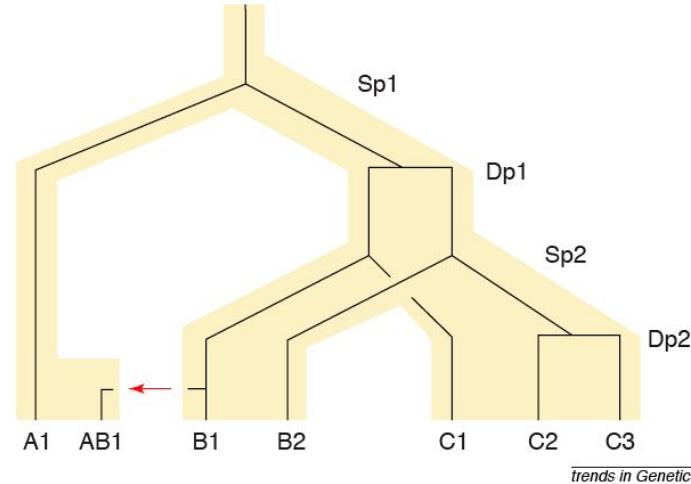
- P paralogie
- O orthologie
- X xenologie
- A analogie

Solutions dans un diaporama séparé

	A1	AB1	B1	B2	C1	C2	C3
A1							
AB1							
B1							
B2							
C1							
C2							
C3							

Figure: Fitch, W. M. Homology a personal view on some of the problems. Trends Genet 16, 227–231 (2000). [doi.org/10.1016/s0168-9525\(00\)02005-9](https://doi.org/10.1016/s0168-9525(00)02005-9)

Définitions : Zvelebil, M. J. & Baum, J. O. Understanding bioinformatics. (Garland Science, 2008).



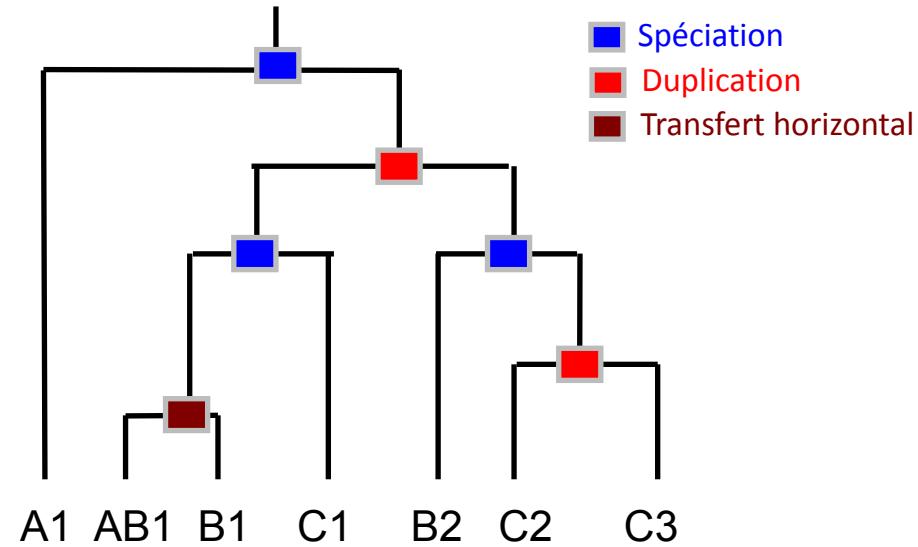
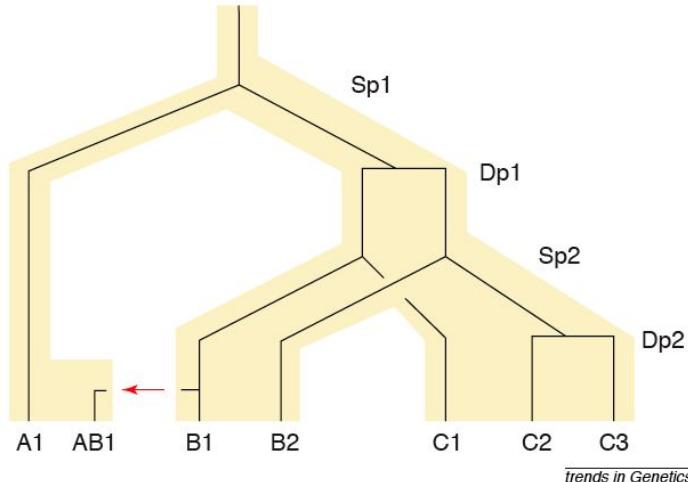
- Paire d'**orthologues**: paire de gènes dont le dernier ancêtre commun précède immédiatement un événement de spéciation (ex: a_1 and a_2). *Source: Zvelebil & Baum, 2000.*
- Paire de **paralogues**: paire de gènes dont le dernier ancêtre commun précède immédiatement une duplication génique (ex: b_2 and b_2'). *Source: Zvelebil & Baum, 2000.*
- **Xénologie**: relation entre deux caractères dont l'histoire, depuis leur dernier ancêtre commun, inclut un transfert entre espèces (horizontal) du matériel génétique pour au moins l'un de ces caractères. *Source: Fitch, 2000.*

Représentation classique des spéciations / duplications

Représentations de la même histoire évolutive.

- Gauche (Fitch, 2000): arbre moléculaire (lignes noires) **superposé** à l'arbre des espèces (fond jaune)
- Droite (du type [Ensembl genomes](#)) : arbre des molécules, avec coloration des noeuds internes selon le type d'événement évolutif

A, B, C représentent les espèces
2, 3, 3 les copies des gènes

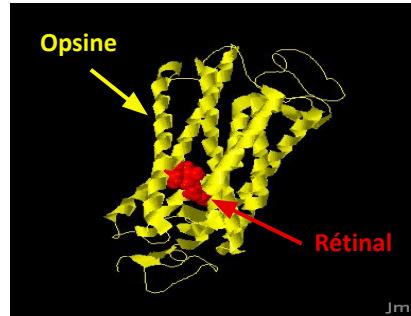


Les duplications à l'origine de l'innovation

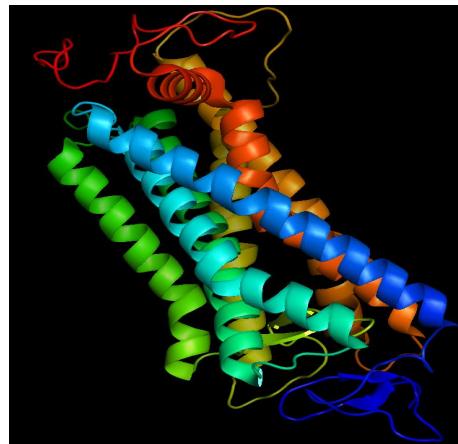
Structure d'une opsine

Modèle tridimensionnel du pigment des cônes bleus

(Structure PDB 1kpn affichée avec Jmol)



(Structure PDB 1kpn affichée avec MacPyMol)



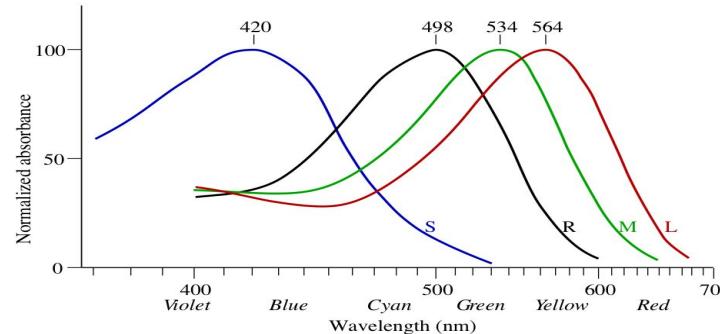
Perception de la lumière

- **Rétinal** : petite molécule, absorbe la lumière
- **Opsine** :
 - protéine transmembranaire
- Complexe Opsine-Rétinal → perception de la lumière
- La séquence d'une opsine détermine le spectre de sensibilité → mutations peuvent modifier la longueur d'onde de sensibilité maximale

La vision trichromatique chez les primates de l'ancien monde

Vision trichromatique

- Primates de l'ancien monde (Afrique + Asie + Europe)
- Trois pigments : protéines “opsines”
 - Bleu (short-waves opsin, SW)
 - Vert (medium-waves opsin, MW)
 - Rouge (long-waves opsin, LW)



[http://fr.wikipedia.org/wiki/Cône_\(biologie\)](http://fr.wikipedia.org/wiki/Cône_(biologie))

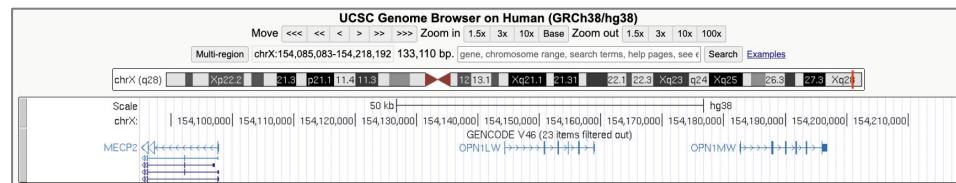


Vision dichromatique

- Autres mammifères
- 2 opsines
 - sensible au bleu
 - sensible aux ondes vert et rouge
- équivalent au daltonisme humain

Origine de la vision trichromatique

- duplication du gène de l'opsine rouge-verte
- → 2 gènes en tandem sur chromosome X
 - OPN1LW : “Long Waves” = opsite rouge
 - OPN1MW: “Medium Waves” = opsite verte
- OPN1SW (Short Waves) : gène de l'opsine verte, sur chromosome 7 chez l'humain



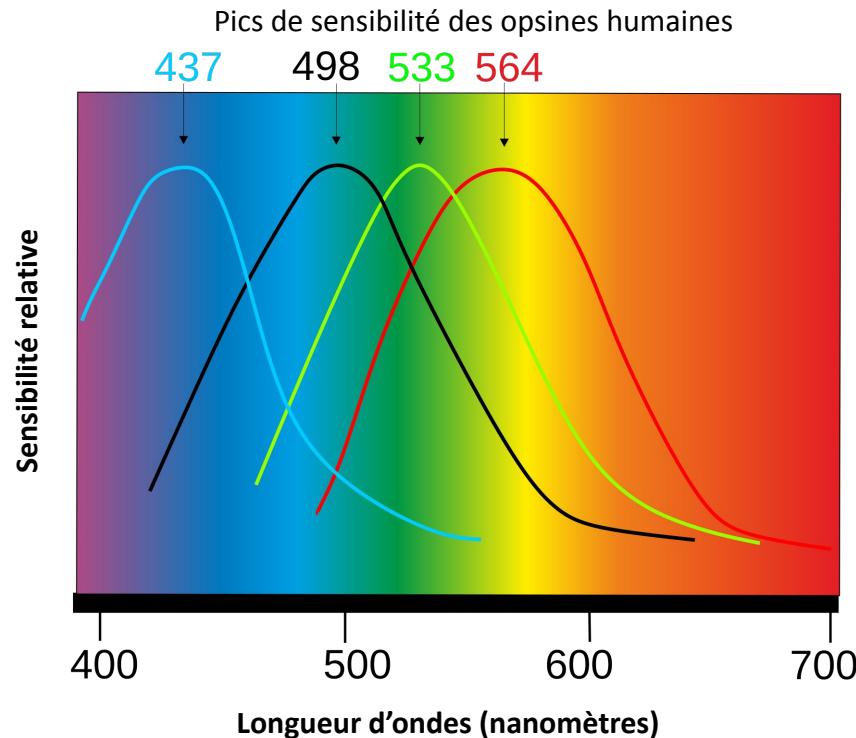
Le spectre visible

Pics de sensibilité

- Cônes sensibles au bleu (short-wave sensitive: SWS):
420 nm
- Bâtonnets: 489 nm
- Cônes sensibles au vert (medium-wave sensitive: MWS) :
534 nm
- Cônes sensibles au rouge (long-wave sensitive: LWS) :
564 nm

Distances entre pics

- Opsine verte versus rouge : $564 - 533 = 31$ nm
 - Fort recouvrement entre spectres de sensibilité
- Opsine bleue versus verte: $533 - 437 = 96$ nm



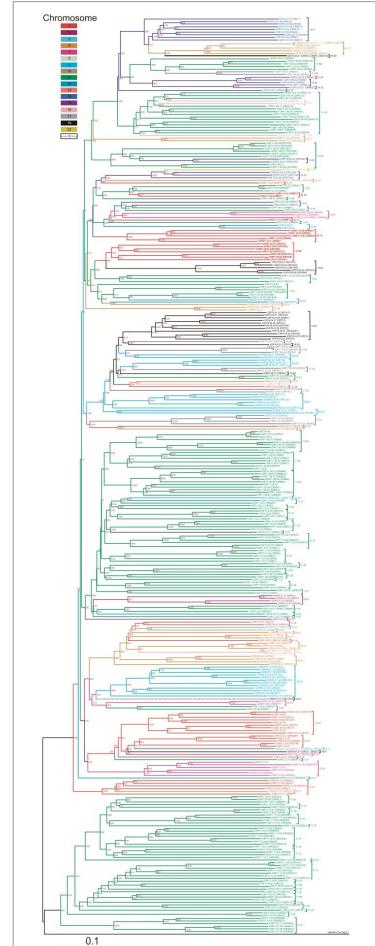
Une grande famille ... de gènes

Récepteurs olfactifs

- La plus grande famille de gènes chez les métazoaires
 - Souris : ~800 gènes
 - Humain : ~ 400 gènes
- Résultent de fréquentes duplications intra-chromosomiques.

Processus évolutif : duplication - divergence

Duplication
→ mutations
→ divergence de séquences
→ différences fonctionnelles (spécificité olfactive)



Alignment multiple

Matrice de pourcentages d'identité (opsines de mammifère, export de clustalx)

Mammalian opsins - percent identify matrix generated by clustalx

Groupe extérieur: 2 monotrèmes
(ornithorynque et echidné)

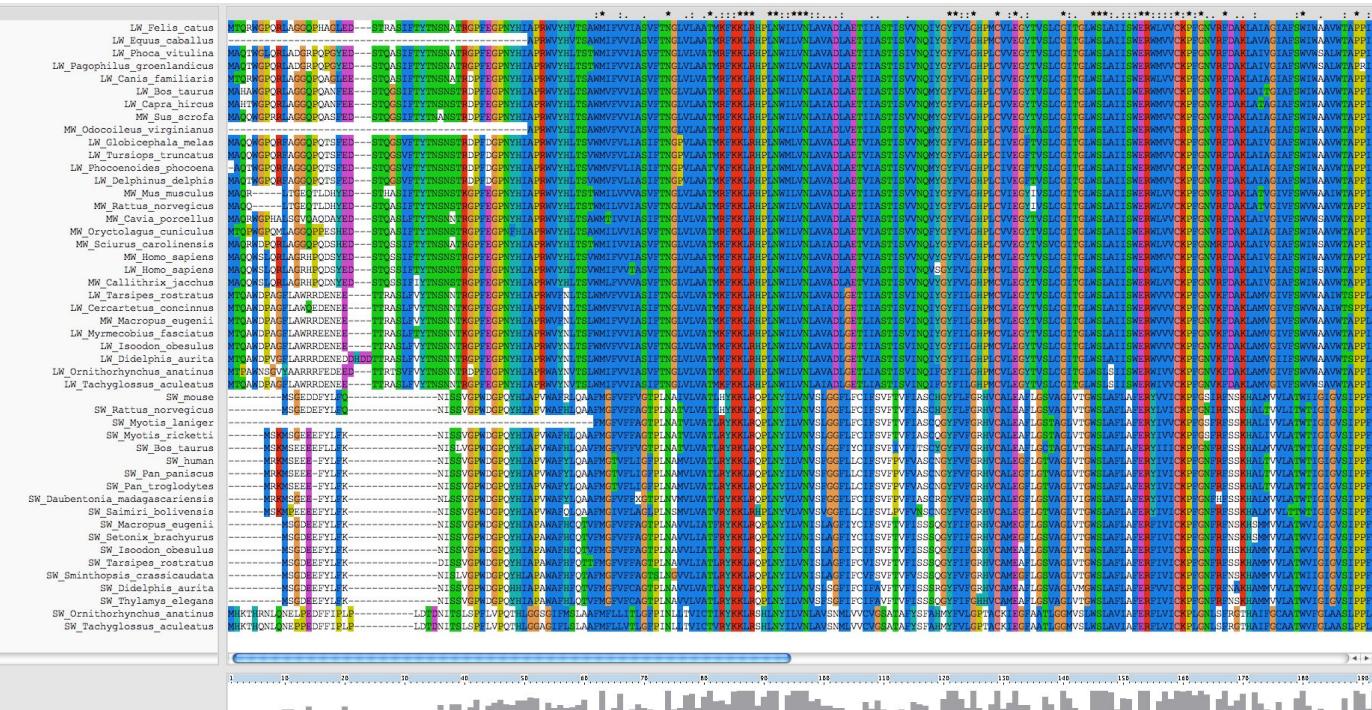
Alignement multiple des opsines de mammifère

Figure

- 50 opsines de différents mammifères (début d'alignement)

2 groupes évidents

- Haut : opsines vertes (MWS) et rouges (LWS)
- Bas : opsines bleues (SWS)

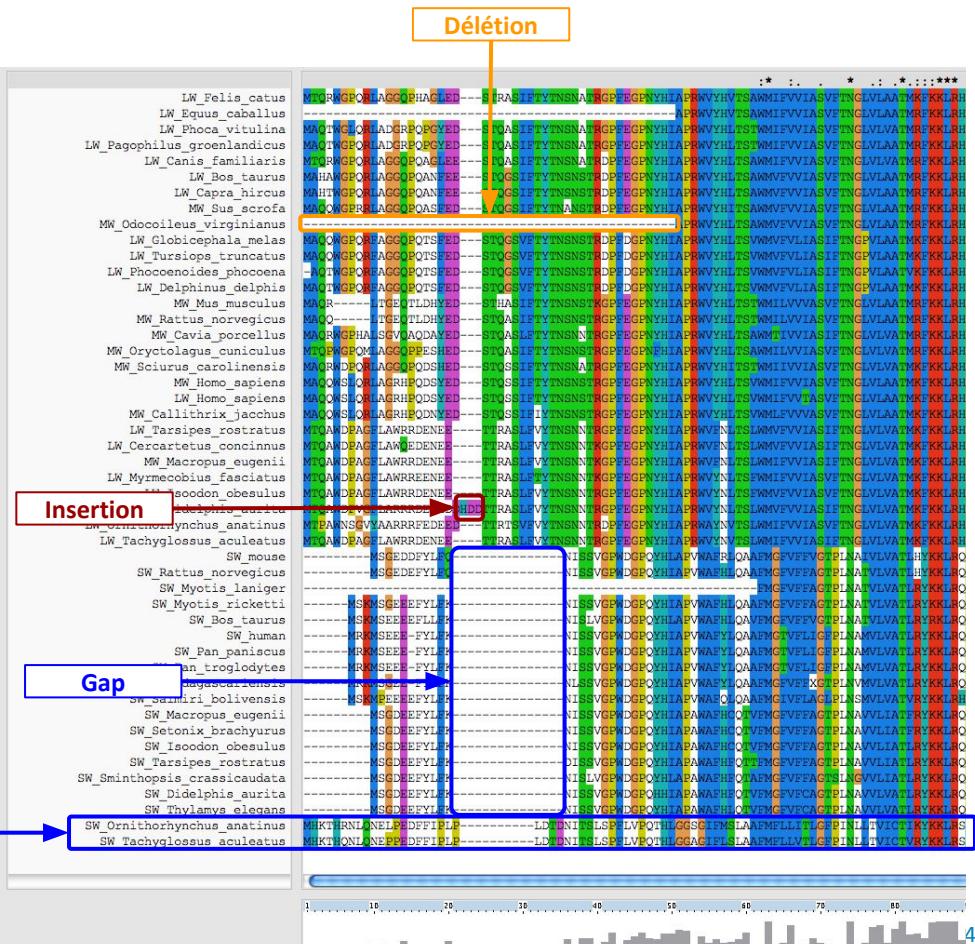


Lecture d'un alignement multiple

- Une ligne par séquence
- **Groupe extérieur** : 2 monotrèmes (*Ornithorhynchus* et *Tachyglossus*)
- Une colonne par position de l'alignement multiple
- **Blocs conservés**
 - sur l'ensemble de l'alignement
 - sur un sous-ensemble des séquences
- **Gaps**
 - délétions : fragments de séquence absents d'une ou de quelques séquences mais présents ailleurs
 - insertions : fragment de séquence présents dans une ou quelques séquences mais absents ailleurs

Pour départager insertion ou délétion il faut généralement comparer l'alignement multiple à l'arbre des molécules (inférence phylogénétique)

Groupe extérieur (outgroup)



Inférence d'un arbre des molécules à partir de l'alignement multiple

Inférence d'un arbre des molécules à partir de l'alignement multiple

Méthodes alternatives

- Algorithmes non présentés au cours
- **A connaître : avantages et inconvénients (à droite)**
 - comment choisir l'algorithme approprié en fonction des jeux de données ?

Puissent donner des résultats différents

- topologie de l'arbre (branchements)
- longueurs des branches
 - comment estimer la fiabilité d'un arbre ?

Maximum de vraisemblance (Maximum likelihood)

- Considérée comme la plus fiable des méthodes
- Permet d'estimer la longueur des branches (produit un phylogramme)
- Coûteuse en temps
- Ne permet pas de traiter des grandes familles de séquences

Neighbour Joining

- Rapide
- Permet de traiter un grand nombre de séquences
- Permet d'estimer la longueur des branches (produit un phylogramme)
- Résultats moins fiables que par la méthode du maximum de vraisemblance

Parcimonie

- Permet d'inférer les caractères ancestraux (pour les séquences: émettre une hypothèse concernant le résidu chaque noeud ancestral à chaque position de l'alignement)
- Très coûteux en temps
- Limité à un petit nombre de séquences
- Ne permet pas d'estimer la longueur des branches (produit un cladogramme)

Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

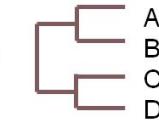
- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Arbre inféré



Méthode de bootstrap pour estimer la robustesse des arbres

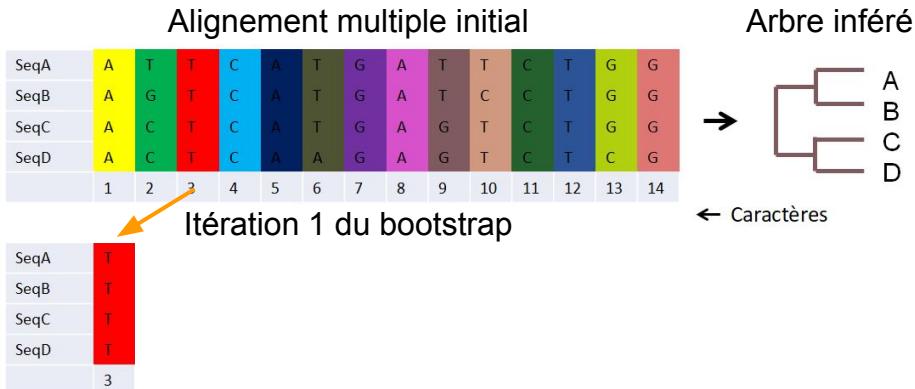
Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.



Méthode de bootstrap pour estimer la robustesse des arbres

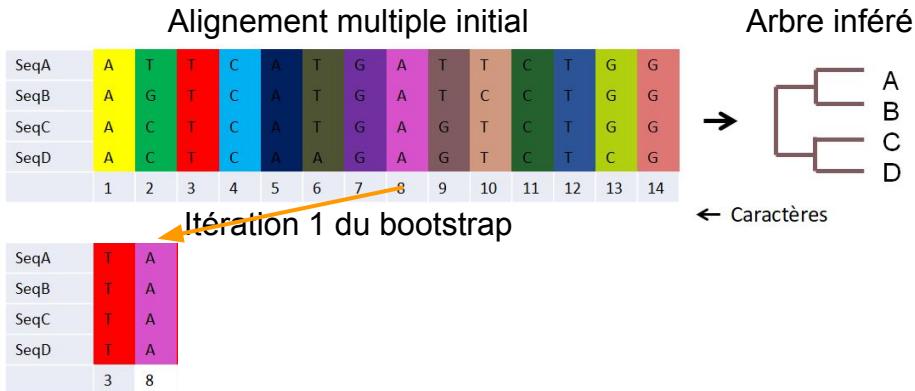
Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.



Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

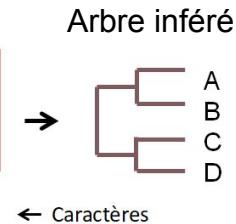
Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Itération 1 du bootstrap



SeqA	T	A	T
SeqB	T	A	T
SeqC	T	A	T
SeqD	T	A	T
	3	8	3

Méthode de bootstrap pour estimer la robustesse des arbres

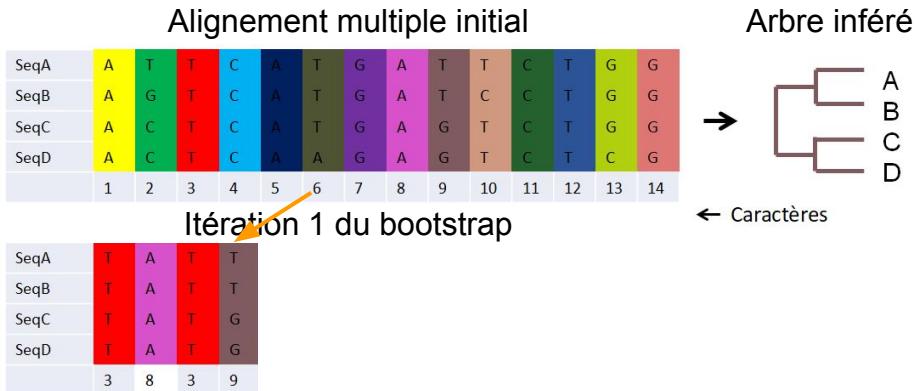
Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.



Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.

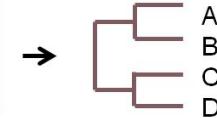
Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Itération 1 du bootstrap

SeqA	T	A	T	T	A
SeqB	T	A	T	T	A
SeqC	T	A	T	G	A
SeqD	T	A	T	G	A
	3	8	3	9	1

Arbre inféré



← Caractères

Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

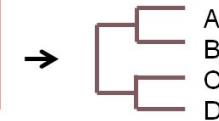
Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Arbre inféré



← Caractères

Itération 1 du bootstrap

SeqA	T	A	T	T	A	C	T	G	C	A	T	T	G	A
SeqB	T	A	T	T	A	C	C	G	C	A	T	T	G	A
SeqC	T	A	T	G	A	C	T	G	C	A	G	T	G	A
SeqD	T	A	T	G	A	C	T	G	C	A	G	T	C	A
	3	8	3	9	1	4	10	14	4	8	9	3	13	1

Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.
- Inférence d'un arbre avec ces colonnes ré-échantillonnées.

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

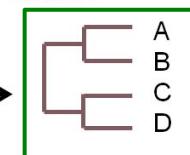
Itération 1 du bootstrap

SeqA	T	A	T	T	A	C	T	G	C	A	T	T	G	A
SeqB	T	A	T	T	A	C	C	G	C	A	T	T	G	A
SeqC	T	A	T	G	A	C	T	G	C	A	G	T	G	A
SeqD	T	A	T	G	A	C	T	G	C	A	G	T	C	A
	3	8	3	9	1	4	10	14	4	8	9	3	13	1

Arbre inféré



← Caractères



Identique à
l'arbre original

Méthode de bootstrap pour estimer la robustesse des arbres

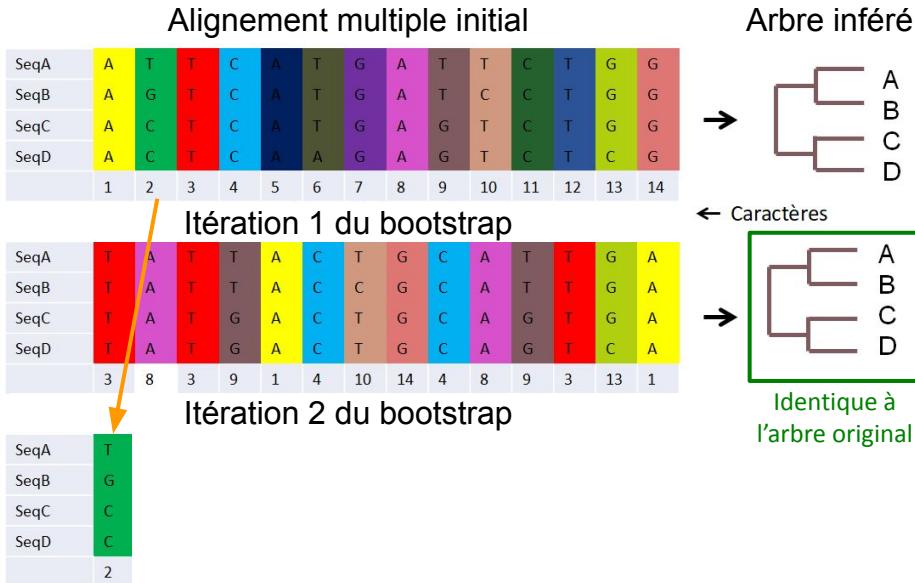
Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- N iterations (ex: N=100 ou N=1000)
 - Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.
 - Inférence d'un arbre avec ces colonnes ré-échantillonées.



Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- N iterations (ex: N=100 ou N=1000)
 - Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.
 - Inférence d'un arbre avec ces colonnes ré-échantillonées.

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

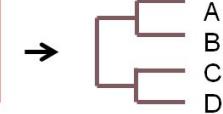
Itération 1 du bootstrap

SeqA	T	A	T	T	A	C	T	G	C	A	T	T	G	A
SeqB	T	A	T	T	A	C	C	G	C	A	T	T	G	A
SeqC	T	A	T	G	A	C	T	G	C	A	G	T	G	A
SeqD	T	A	T	G	A	C	T	G	C	A	G	T	C	A
	3	8	3	9	1	4	10	14	4	8	9	3	13	1

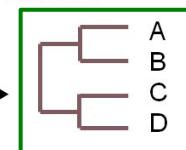
Itération 2 du bootstrap

SeqA	T	T
SeqB	G	T
SeqC	C	G
SeqD	C	G
	2	9

Arbre inféré



← Caractères



Identique à
l'arbre original

Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- N iterations (ex: N=100 ou N=1000)
 - Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.
 - Inférence d'un arbre avec ces colonnes ré-échantillonées.

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Itération 1 du bootstrap

SeqA	T	A	T	T	A	T	T	G	C	A	T	T	G	A
SeqB	T	A	T	T	A	C	C	G	C	A	T	T	G	A
SeqC	T	A	T	G	A	C	T	G	C	A	G	T	G	A
SeqD	T	A	T	G	A	C	T	G	C	A	G	T	C	A
	3	8	3	9	1	4	10	14	4	8	9	3	13	1

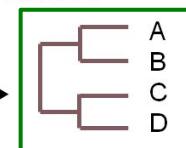
Itération 2 du bootstrap

SeqA	T	T	G
SeqB	G	T	G
SeqC	C	G	G
SeqD	C	G	G
	2	9	7

Arbre inféré



← Caractères



Identique à
l'arbre original

Méthode de bootstrap pour estimer la robustesse des arbres

Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- N iterations (ex: N=100 ou N=1000)
 - Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.
 - Inférence d'un arbre avec ces colonnes ré-échantillonées.

Alignement multiple initial

SeqA	A	T	T	C	A	T	G	A	T	T	C	T	G	G
SeqB	A	G	T	C	A	T	G	A	T	C	C	T	G	G
SeqC	A	C	T	C	A	T	G	A	G	T	C	T	G	G
SeqD	A	C	T	C	A	A	G	A	G	T	C	T	C	G
	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Itération 1 du bootstrap

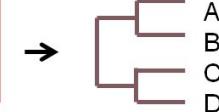
SeqA	T	A	T	T	A	C	T	G	C	A	T	T	G	A
SeqB	T	A	T	T	A	C	C	G	C	A	T	T	G	A
SeqC	T	A	T	G	A	C	T	G	C	A	G	T	G	A
SeqD	T	A	T	G	A	C	T	G	C	A	G	T	C	A
	3	8	3	9	1	4	10	14	4	8	9	3	13	1

Itération 2 du bootstrap

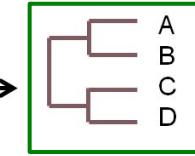
SeqA	T	T	G	C	A	A	T	T	T	A	A	T	A	A
SeqB	G	T	G	C	A	A	T	T	G	A	A	T	A	A
SeqC	C	G	G	C	A	A	G	T	C	A	A	G	A	A
SeqD	C	G	G	C	A	A	G	T	C	A	A	G	A	A
	2	9	7	4	1	8	9	3	2	1	8	9	5	1

Figure adapté du cours de Céline Brochier

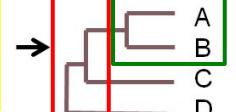
Arbre inféré



← Caractères



Identique



Branchements différents de l'arbre original

Méthode de bootstrap pour estimer la robustesse des arbres

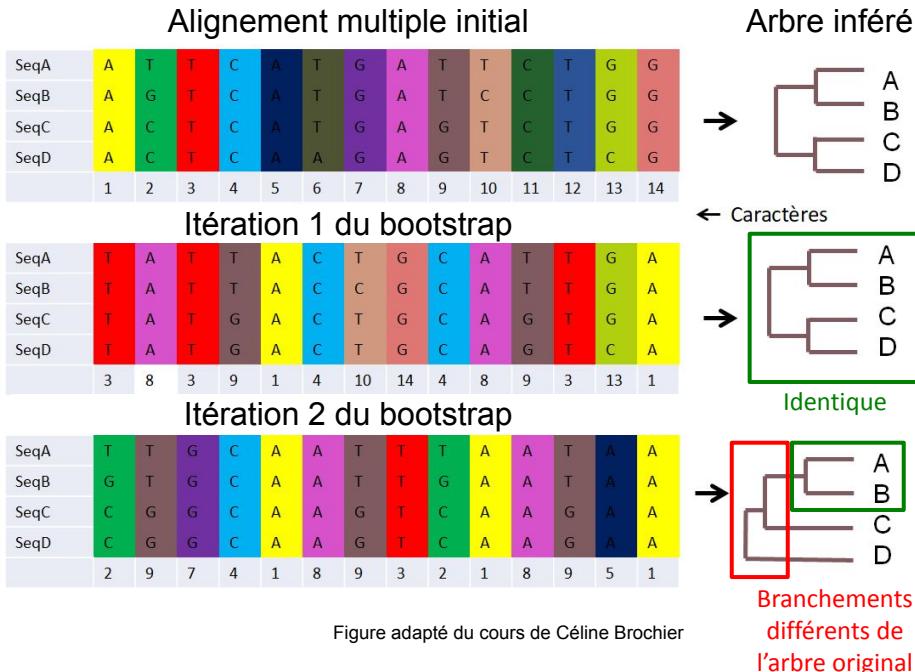
Alignement multiple complet

- M colonnes (positions d'alignement)
- S lignes (séquences)

→ inférence d'un l'arbre phylogénétique

Estimation de la robustesse : bootstrapping.

- N iterations (ex: N=100 ou N=1000)
 - Sélection aléatoire de M colonnes **avec remise**
→ Chaque colonne peut être tirée 0, 1 ou plusieurs fois.
 - Inférence d'un arbre avec ces colonnes ré-échantillonées.
- **Valeur de bootstrap** de chaque branchement
 - nombre de fois où ce branchement se retrouve à l'identique dans les N arbres produits.
 - → indique la robustesse de l'arbre par rapport aux fluctuations d'échantillonnage
 - **! Ne garantit pas que l'arbre reflète l'histoire évolutive** (les données peuvent être erronées ou biaisées)



Bootstrapping

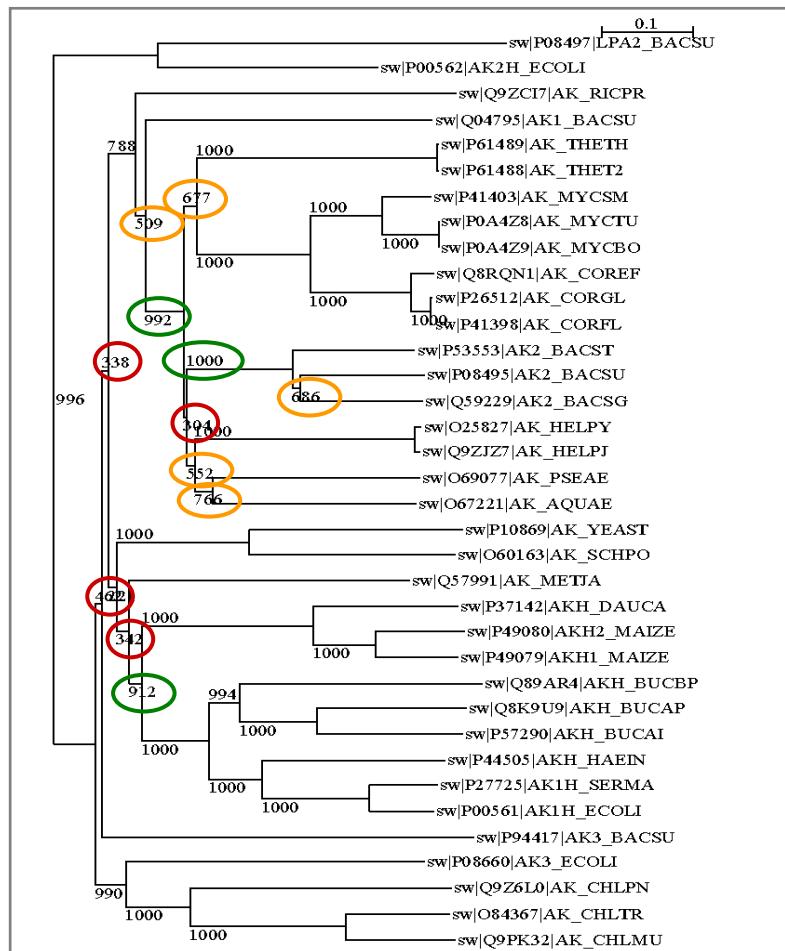
- **Valeur de bootstrap** assignée à chaque branchement
 - Indique le nombre de fois où ce branchement se retrouve à l'identique dans les N arbres de bootstrap
 - **Valeur élevée** → robuste aux fluctuations d'échantillonnage
 - **Valeur faible** → sensible.

Attention ! La valeur de bootstrap

- indique la robustesse par rapport aux données,
 - ne garantit pas la fiabilité par rapport à l'histoire évolutive réelle ("arbre vrai")

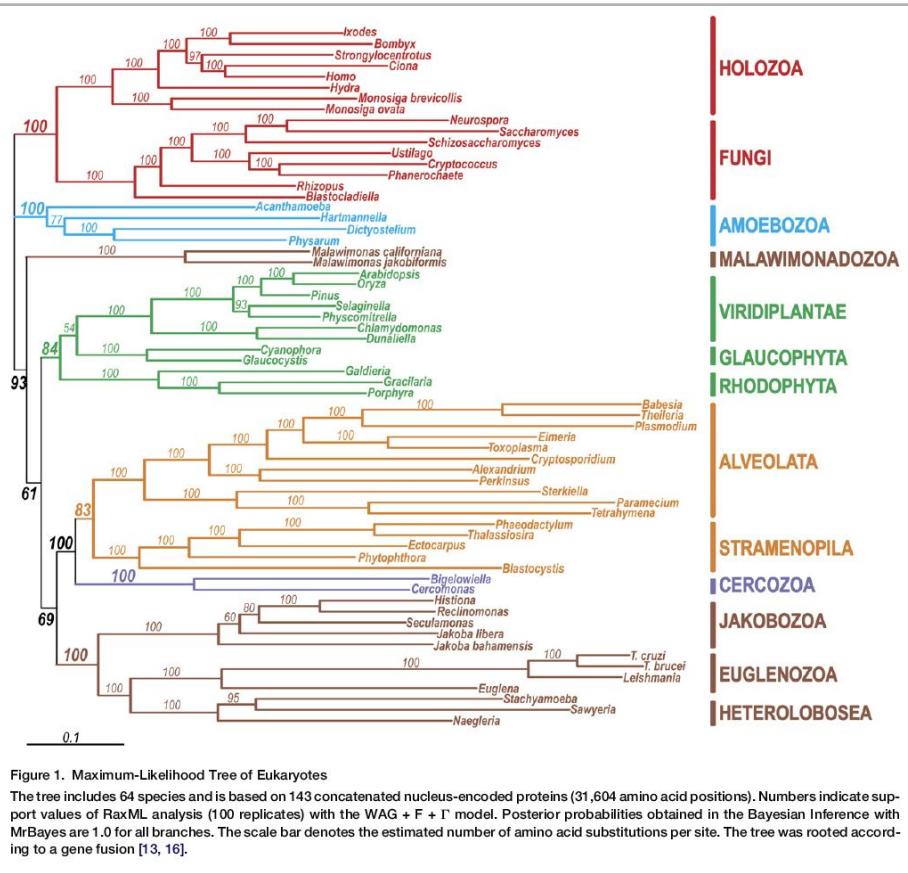
Problèmes potentiels

- Biais d'échantillonnage
 - Données erronées



Phylogénomique : retracer l'évolution des espèces à partir des séquences génomiques

Phylogénomique



- Phylogénie moléculaire “classique”: arbre d’une famille de molécule
- Phylogénomique
 - Inférence d’une arbre des espèces
 - Données: séquences de **plusieurs centaines de gènes ou protéines**
- → phylogénies entre organismes très éloignés (règnes différents)
- scénarios concernant les premières étapes de la diversification des êtres vivants.

Exemple (figure de gauche)

- Arbre basé sur 143 familles de protéines
- Grands groupes d'eucaryotes

Source: Rodríguez-Ezpeleta et al. Curr Biol (2007) vol. 17 (16) pp. 1420-5
Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans.

Suffit-il de manger des insectes pour être un Insectivore ?

Au 20^e siècle, les **tenrecs** étaient considérés comme une famille d'insectivores, présentant une ressemblance morphologique mais des différences anatomiques importantes par rapport aux hérissons.

Hérisson

classe	Mammifères
ordre	Insectivores
famille	Erinacéidés
genre et espèce	<i>Erinaceus europaeus</i> et autres

Tenrec

classe	Mammifères
ordre	Insectivores
famille	Tenrécidés
genres et espèces	<i>Centetes ecaudatus</i> (tenrec commun) <i>Hemicentetes semispinosus</i> (tenrec à bandes ou tenrec strié) <i>Limnogale mergulus</i> (tenrec à pieds palmés) <i>Microgale longicaudata</i> (tenrec à longue queue) <i>Oryzorictes hova</i> (tenrec des rizières) <i>Setifer setosus</i> (tenrec-hérisson) et autres



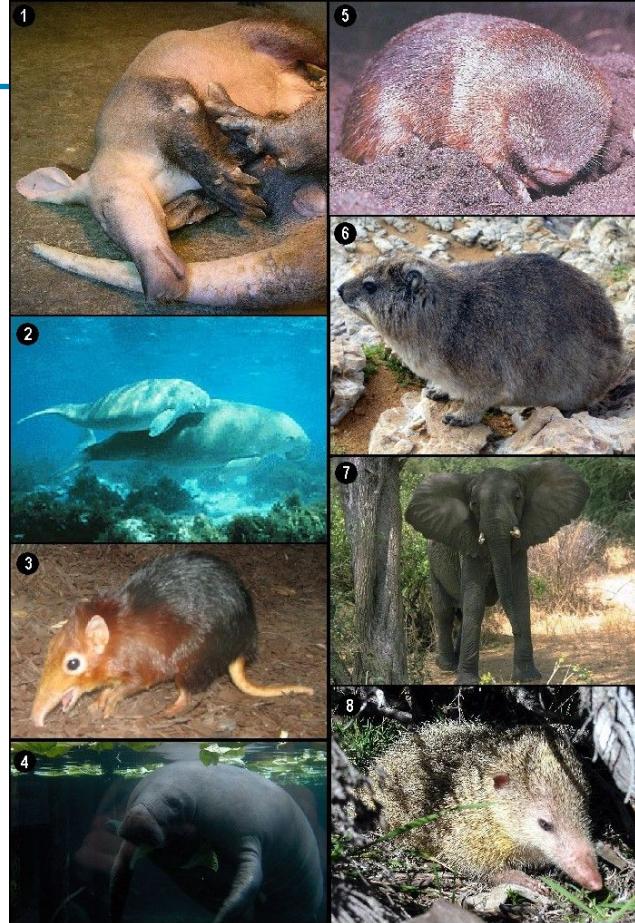
Les Afrotheria

Quels sont les plus proches cousins du tenrec ?

Inférence phylogénétique (fin des années 1990)

- regroupement d'espèces très différentes
- définition du **super-ordre des Afrotheria**

Autres Afrotheria: éléphant, lacentin, dugong, oryctérope, taupe dorée, daman, ...

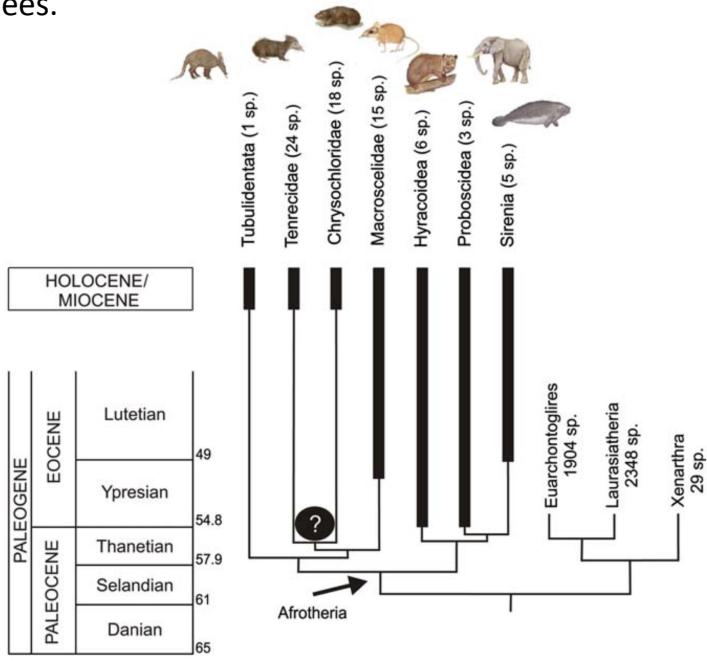


1. Oryctérope du Cap; 2. Dugongs; 3. Macroscéléide de Peters ; 4. Lamantin; 5. Taupe dorée; 6. Daman du Cap; 7. Éléphant de savane d'Afrique ; 8. Tangue ("tailless tenrec"); 8, 8b: Tenrec
<http://upload.wikimedia.org/wikipedia/commons/0/01/Kleiner-igeltanrek-a.jpg>

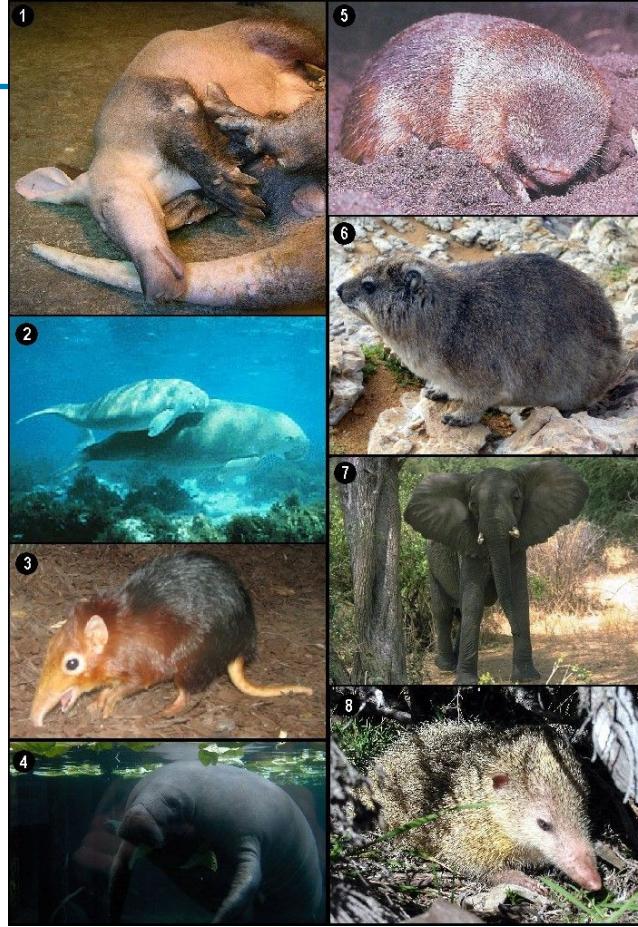
- Murphy, W. J. et al. Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. *Science* 294, 2348–2351 (2001). doi.org/10.1126/science.1067179
- Tabuce, R., Asher, R. J. & Lehmann, T. Afrotherian mammals: a review of current data. *mammalia* 72, (2008). doi.org/10.1515/MAMM.2008.004

Les Afrotheria

- Sur base d'analyse de l'ADN, on estime que la divergence entre Afrotheria remonte à 60-55 millions d'années.



- Murphy, W. J. et al. Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics. *Science* 294, 2348–2351 (2001). doi.org/10.1126/science.1067179
- Tabuce, R., Asher, R. J. & Lehmann, T. Afrotherian mammals: a review of current data. *mammalia* 72, (2008). doi.org/10.1515/MAMM.2008.004



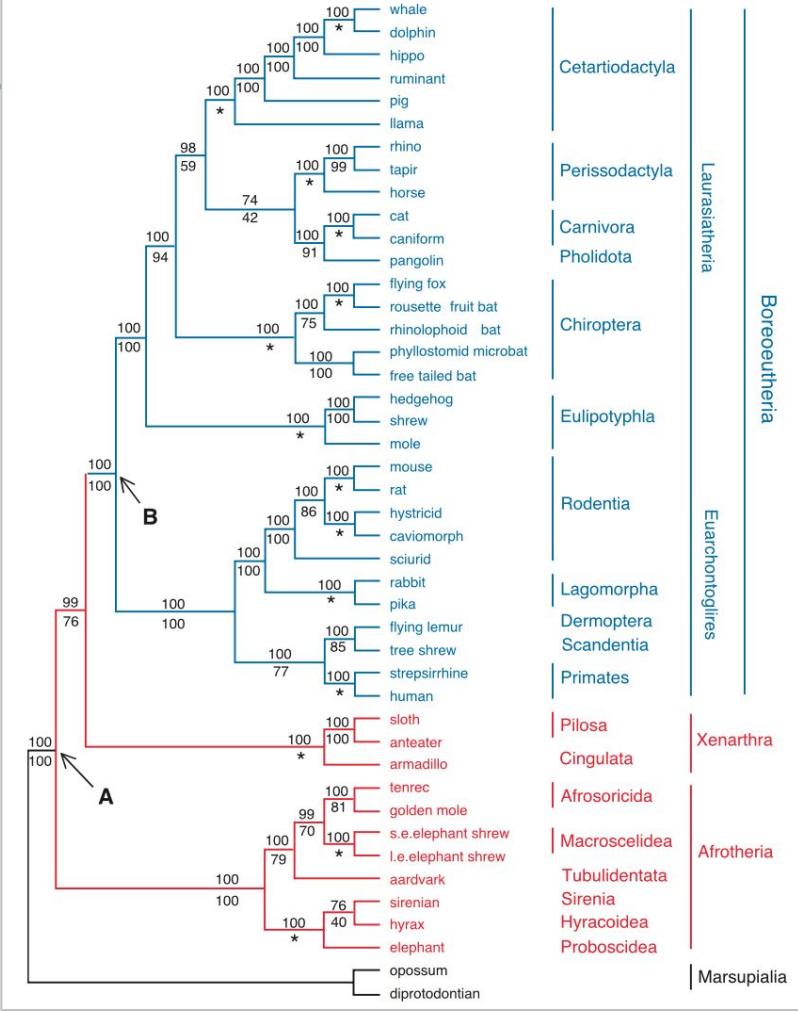
1. Oryctérope du Cap; 2. Dugongs; 3. Macroscéléide de Peters ; 4. Lamantin; 5. Taupe dorée; 6. Daman du Cap; 7. Éléphant de savane d'Afrique ; 8. Tangue ("tailless tenrec"); 8, 8b: Tenrec
<http://upload.wikimedia.org/wikipedia/commons/0/01/Kleiner-igeltanrek-a.jpg>

Resolution of the Early Placental Mammal Radiation Using Bayesian Phylogenetics

William J. Murphy,^{1,*} Eduardo Eizirik,^{1,2*} Stephen J. O'Brien,^{1†}
 Ole Madsen,³ Mark Scally,^{4,5} Christophe J. Douady,^{4,5}
 Emma Teeling,^{4,5} Oliver A. Ryder,⁶ Michael J. Stanhope,^{5,7}
 Wilfried W. de Jong,^{3,8} Mark S. Springer^{4†}

Molecular phylogenetic studies have resolved placental mammals into four major groups, but have not established the full hierarchy of interordinal relationships, including the position of the root. The latter is critical for understanding the early biogeographic history of placentals. We investigated placental phylogeny using Bayesian and maximum-likelihood methods and a 16.4-kilobase molecular data set. Interordinal relationships are almost entirely resolved. The basal split is between Afrotheria and other placentals, at about 103 million years, and may be accounted for by the separation of South America and Africa in the Cretaceous. Crown-group Eutheria may have their most recent common ancestry in the Southern Hemisphere (Gondwana).

Fig. 1. Phylogeny of living placental mammals reconstructed using a Bayesian phylogenetic approach. An identical topology was obtained with maximum likelihood [$-\ln L = 211110.54$; see (15) for methodological details]. The number above each branch refers to the Bayesian posterior probability (shown as percentages; i.e., 95 represents a posterior probability of 0.95) of the node derived from 26,250 MCMC sampled trees on the basis of the complete 16.4-kb data. Additional analyses with the full data set and with data sets that varied taxon sampling (i.e., jackknifing single outgroup taxa) and character sampling (nuclear only and nuclear coding loci only) produced similarly high posterior probabilities (15). Values below branches represent percent support in maximum likelihood ($\text{GTR} + \Gamma + I$) nonparametric bootstrap. An asterisk indicates nodes constrained in the ML nonparametric bootstrap analysis. (A) Bifurcation between Afrotheria and Xenarthra + Boreoeutheria at approximately 103 million years, which corresponds to the vicariant event that separated Africa and South America (Fig. 2B). (B) Branch where dispersal from South America to Laurasia is hypothesized to have occurred (15). Blue, monophyletic Northern Hemisphere group (i.e., Boreoeutheria); red, paraphyletic Southern Hemisphere group (i.e., Xenarthra + Afrotheria); black, outgroups.

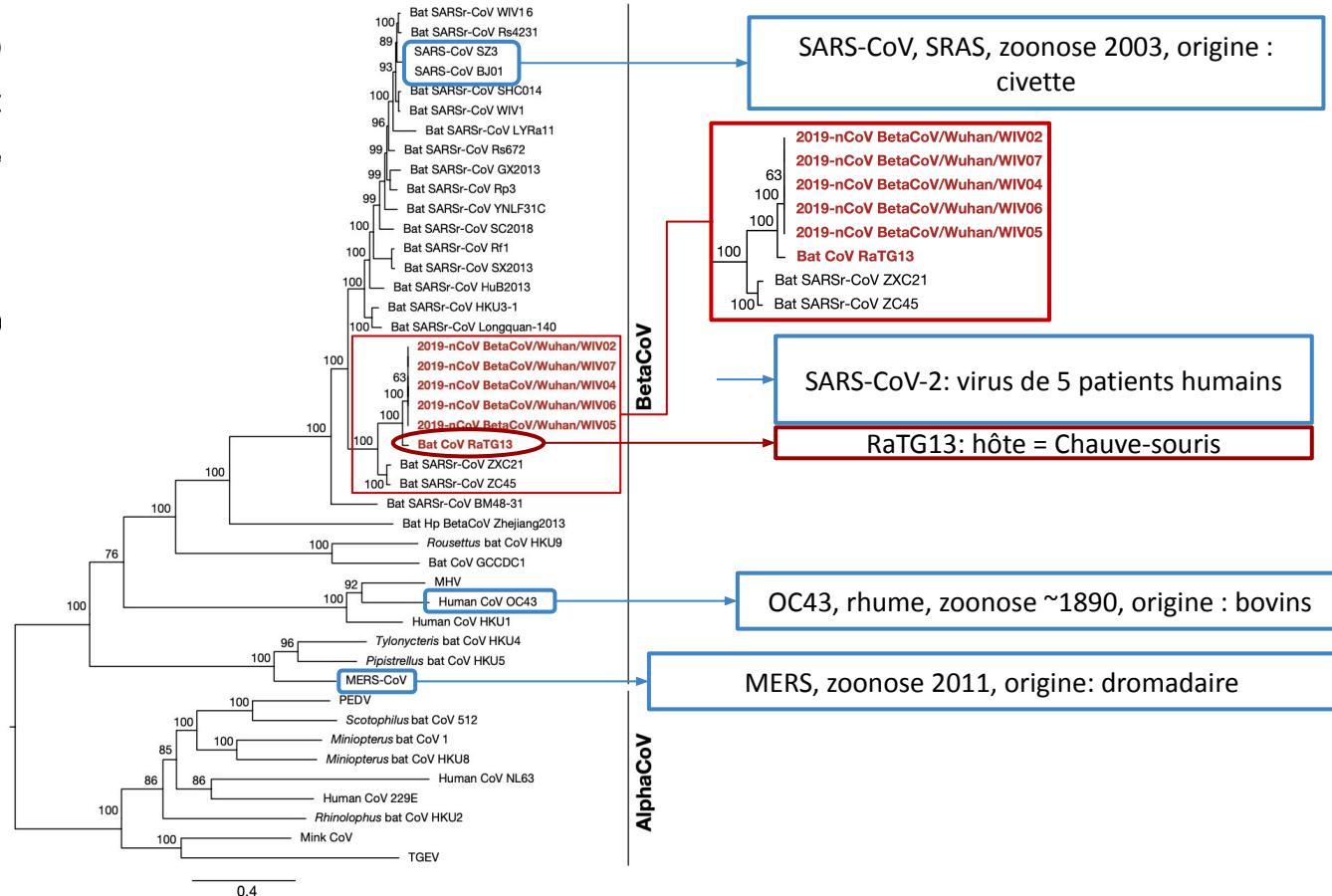


Retracer l'origine de SARS-CoV-2 dans les génomes des coronavirus

La publication du génome de SARS-CoV-2

3 février 2020 : génomes complets publiés

- **SARS-CoV-2 (COVID-19)**
- **RaTG13:** virus de chauve-souris, le plus proche de SARS-CoV-2
- 96.2% d'identité sur l'ensemble du génome
± 40 à 70 ans de divergence
→ cousin éloigné plutôt qu'ancêtre direct de SARS-CoV-2



Un virus synthétique avec des bouts de HIV ?

Le 17 avril 2020: Luc Montagnier (prix Nobel de médecine pour sa contribution à la découverte du HIV), affirme que SARS-CoV-2 résulte de manipulations : insertion de fragments de HIV dans un génome de coronavirus, et fuite accidentelle du labo.

Eléments cités par Mongagnier

- Publication de Jean-Claude Perez
- Préprint d'un groupe de chercheurs indiens "forcés à rétracter"

Selon Luc Montagnier, le virus covid19 est une manipulation humaine (17 avril 2020)

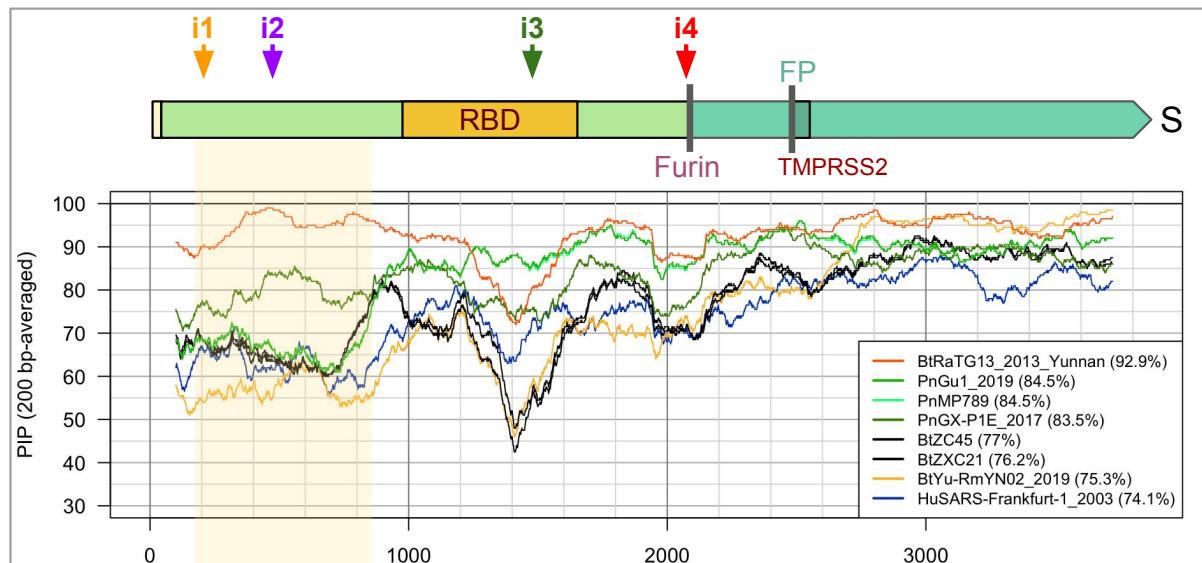
<https://www.youtube.com/watch?v=qSWCLHIOiMo>
(devenue inaccessible depuis lors)

"Je suis arrivé à la conclusion qu'il y avait eu une manipulation de ce virus. [...] Il y a un modèle qui est évidemment le virus classique, et là c'était un modèle venant de la chauve-souris, et là, à ce modèle on a par-dessus ajouté les séquences du VIH, du SIDA. ... Non, ce n'est pas naturel, c'était un travail de professionnel, de biologiste moléculaire, très minutieux, on peut dire d'horloger, au niveau des séquences. Dans quel but ce n'est pas clair. Mon travail c'est d'exposer les faits, c'est tout. Je n'accuse personne, je ne sais pas qui a fait ça et pourquoi. La possibilité c'est qu'on a voulu faire un vaccin contre le SIDA. Donc on a pris des petites séquences du virus [HIV] et on les a installées dans la séquence plus grande du coronavirus. [...] Il y a quand même une volonté d'étouffement, nous ne sommes pas les premiers. Un groupe de chercheurs indiens très renommés avaient publié la même chose, on les a forcés à rétracter. Si vous regardez leur publication vous voyez une grande bande "annulé". "

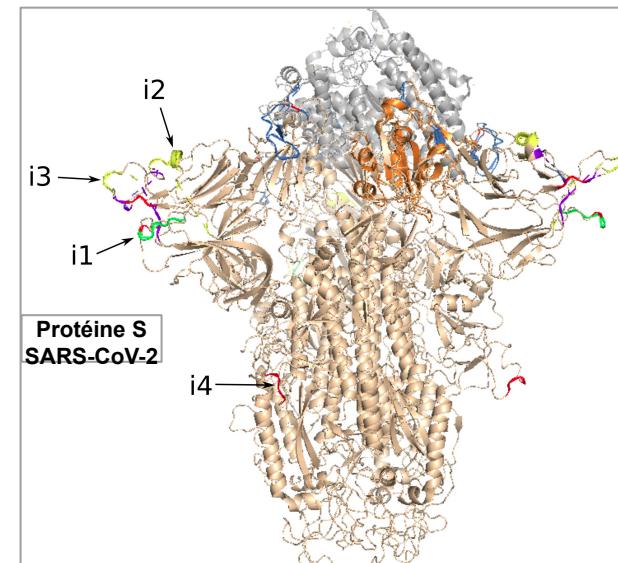
Quatre insertions dans le gène S de SARS-CoV-2

- Flèches: 4 insertions
- Note: i1, i2, i3 dans des régions “exposées” (boucles externes)

Structure et profil de conservation du gène S



Structure de la protéine spike



Alignement de séquences de SARS-CoV-2 sur le génome du HIV

Haut: alignement par paire:

- Query: gène S
 - Subject: génome complet de VIH (10kb)
 - **Expect = 7.5 → pas du tout significatif**

Bas: contrôle négatif

- Query : séquence aléatoire
 - Subject : génome complet de HIV
 - Expect = 2.1 → meilleur alignement que gène S , mais toujours pas significatif (comme attendu, la séquence étant aléatoire !)

Conclusion: les similarités ne permettent pas d'affirmer qu'on a inséré des séquences de HIV dans le génome de SARS-CoV-2.

HIV-1 isolate 19828.PPH11 from Netherlands envelope glycoprotein (env) gene, partial cds				
Sequence ID: HQ644953.1		Length: 1143	Number of Matches: 1	Range 1: 967 to 994
Score	Expect	Identities	Gaps	Strand
38.3 bits(41)	7.5	25/28(89%)	0/28(0%)	Plus/Plus
Query 86	AATGGTACTAAGAGGTTTGATAACCTG	113		
Sbjct 967	AATGGTACTAAAGGTTAGATAACACTG	994		

Des insertions bizarres?

Figure de Pradhan et al (2020), initialement sur bioRxiv, puis retirée par les auteurs suite à des critiques sur leur rigueur scientifique.

- alignement par paire, même si la légende indique “Multiple sequence alignment”)
- Insuffisant pour inférer si insertions ou délétion
- Insuffisant pour évaluer l'ancienneté de l'événement

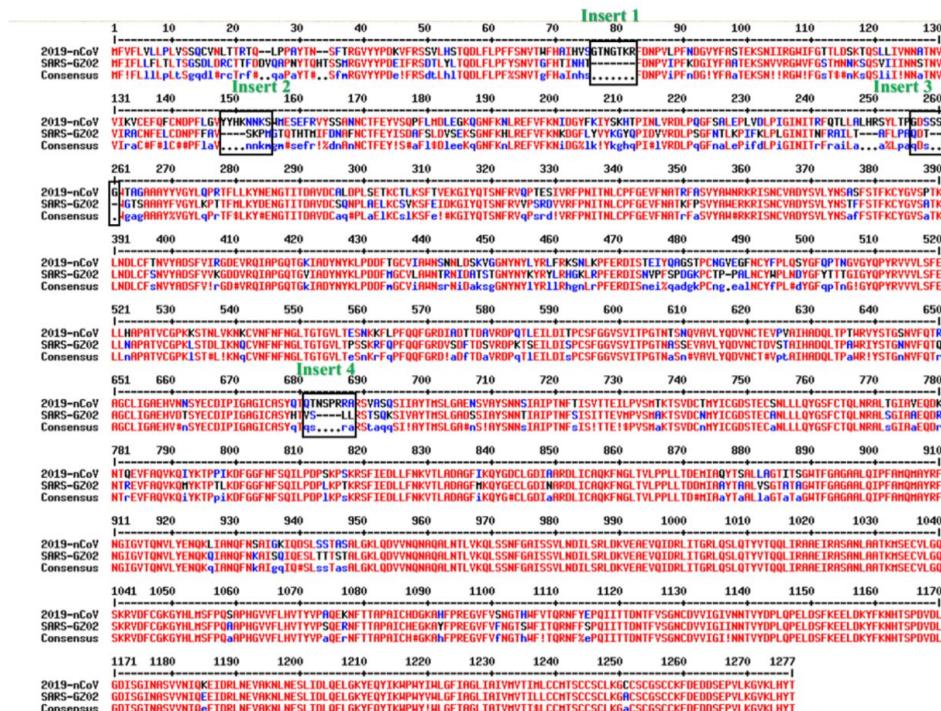
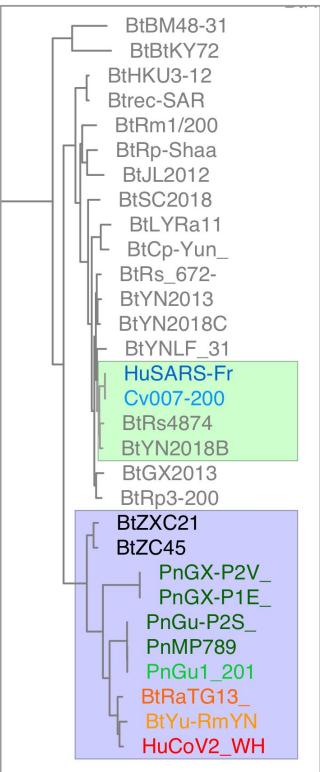


Figure 2: Multiple sequence alignment between spike proteins of 2019-nCoV and SARS. The sequences of spike proteins of 2019-nCoV (Wuhan-HU-1, Accession NC_045512) and of SARS-CoV (GZ02, Accession AY390556) were aligned using MultiAlin software. The sites of difference are highlighted in boxes.

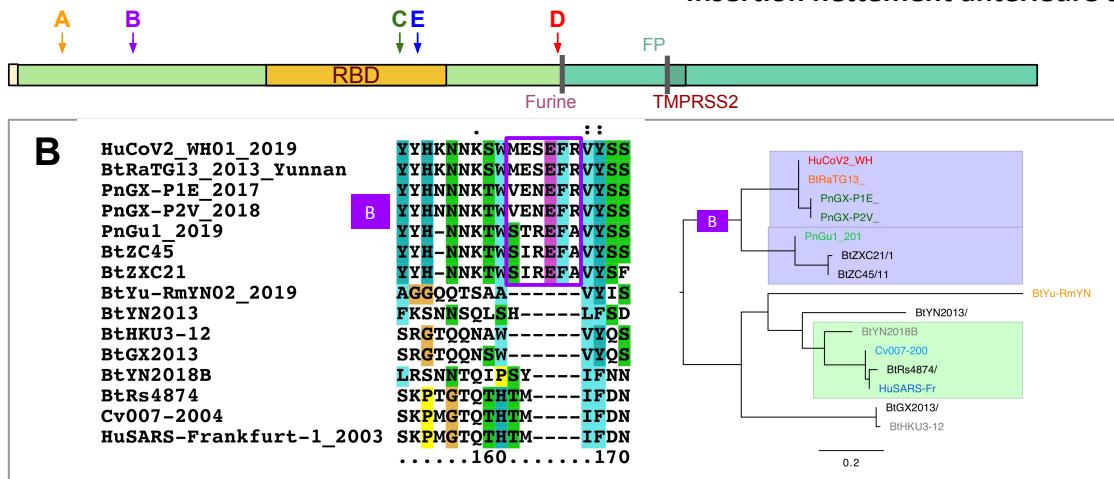
Insertion B : partagée entre plusieurs espèces de coronavirus

Arbre des génomes



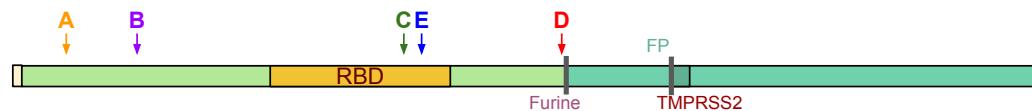
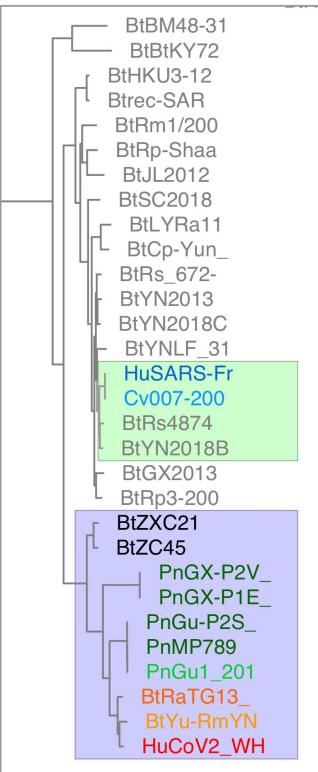
- B : insertion présente chez autres virus de chauve-souris + pangolin
- Séquences identiques entre SARS-CoV-2 et RaTG13, malgré 40 à 70 ans de divergence
- Séquences très proches entre virus de humain, de chauve-souris, et de pangolin
- Substitution secondaire, au sein de l'insertion :
 - MES dans 3 génomes
 - STR dans 3 autres

→ Insertion nettement antérieure à l'émergence de SARS-CoV-2

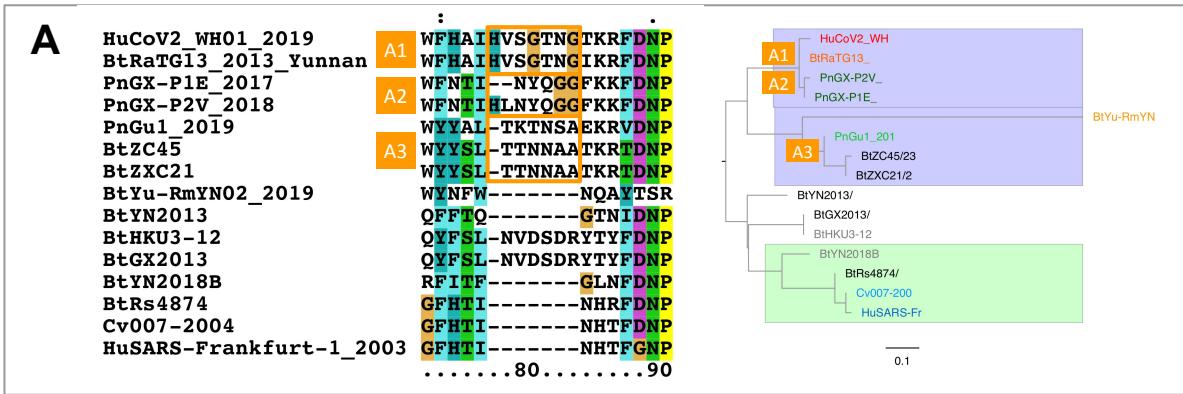


Insertion A : partagée entre SARS-CoV-2 et RaTG13

Arbre des génomes

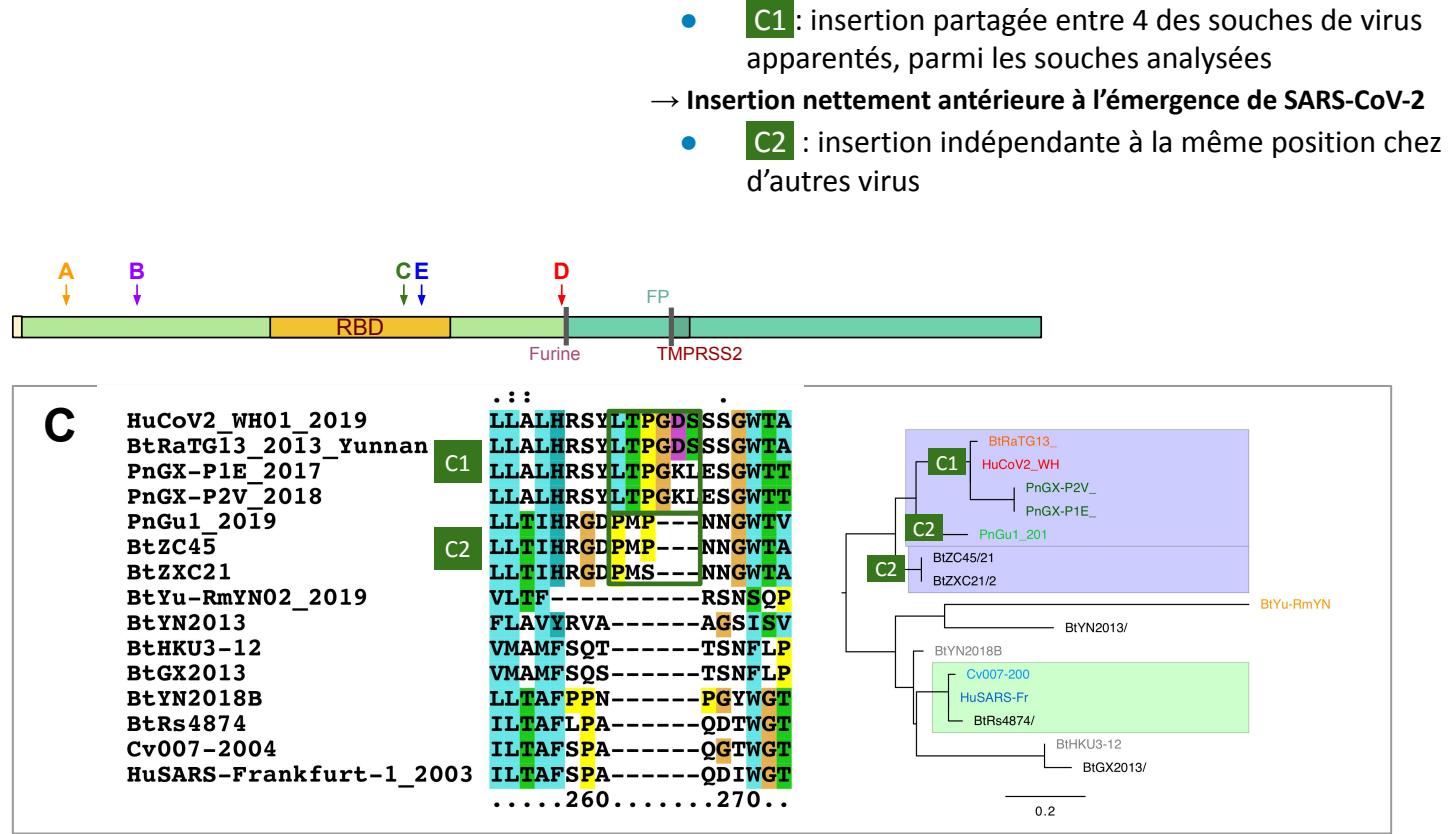
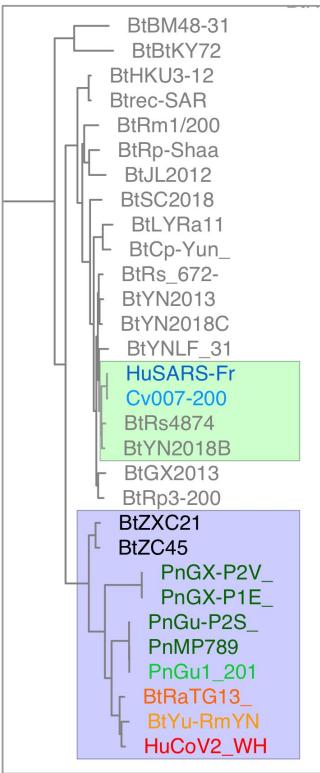


- A1 : Insertion identique entre SARS-CoV-2 et RaTG13, malgré 40 à 70 ans de divergence
→ **Insertion nettement antérieure à l'émergence de SARS-CoV-2**
- A2, A3 : Insertions au même site, dans des sous-groupes séparés → événements indépendants



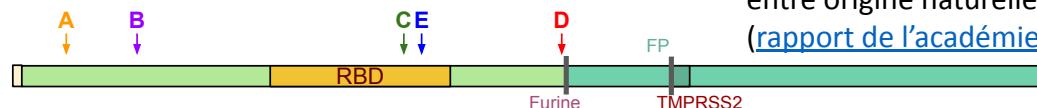
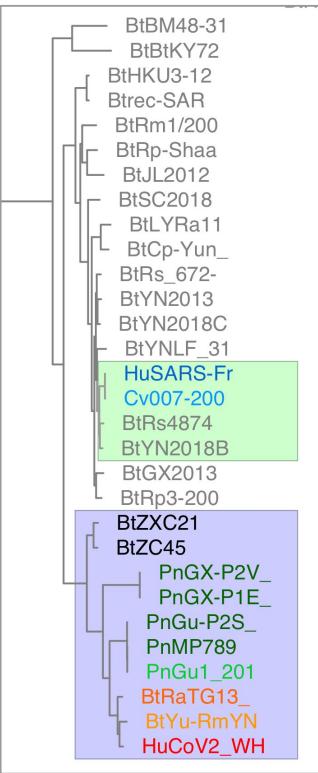
Insertion C

Arbre des génomes

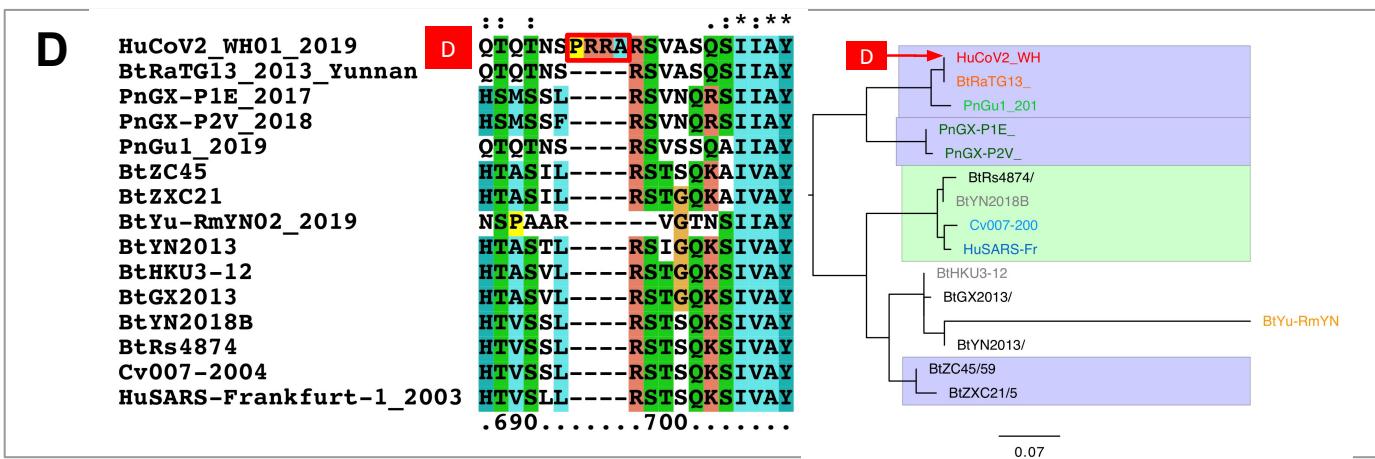


Insertion D

Arbre des génomes

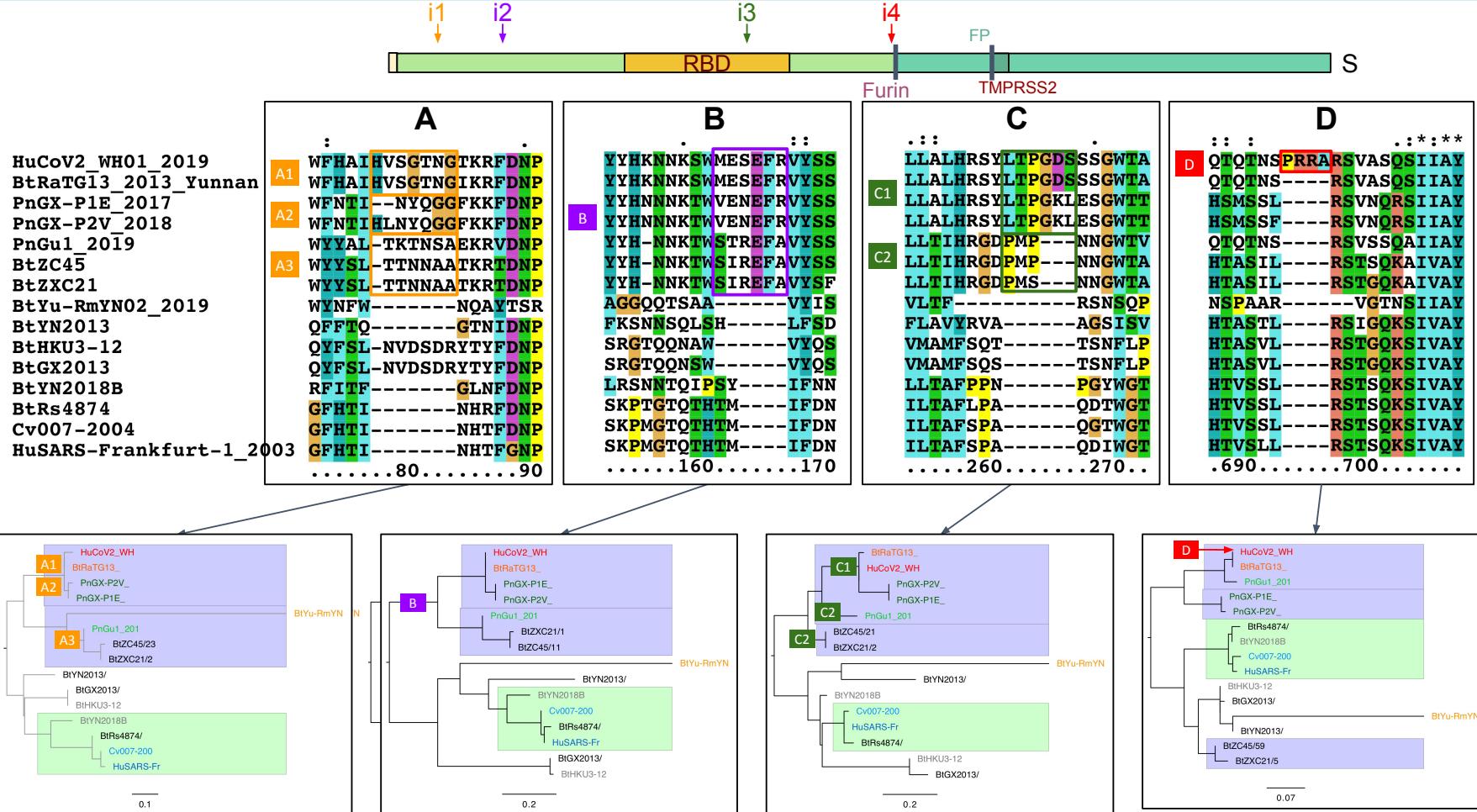


- **D** : insertion très courte, sans similarité significative avec HIV
- Insertion unique à SARS-CoV-2
- Crée un site furine, qui augmente la capacité de la protéine à se lier aux cellules humaines.
- Les données disponibles ne permettent pas de statuer entre origine naturelle ou artificielle de cette insertion ([rapport de l'académie nationale de médecine, 2025](#))



- Sallard, E., Halloy, J., Casane, D., van Helden, J., Decroly, E. Retrouver les origines du SARS-COV-2 dans les phylogénies de coronavirus. Médecine/Sciences 36, Sept-Août 2020. <https://doi.org/10.1051/medsci/2020123>
- [De l'Origine du SARS-CoV-2 aux risques de zoonoses et de manipulations dangereuses de virus](#), rapport de l'Académie Nationale de Médecine (2025)

Alignements multiples et arbres phylogénétiques des insertions de la protéine S



Points-clés : origine de SARS-CoV-2

- L'alignement multiple → blocs de séquences où un fragment est présent / absent.
- Comparaison avec arbre phylogénétique → inférence des événements évolutifs (insertion, délétion, substitution) et estimation de l'ancienneté
- L'hypothèse de Montagnier (insertion de séquences HIV dans un coronavirus) ne tient pas la route.
 - méconnaissance des méthodes bioinformatiques et des scores statistiques (E-valeur)
 - incompatible avec la présence de ces mêmes insertions dans plusieurs génomes
 - échantillons collectés à des dates et endroits indépendants
 - séquences publiées bien avant la pandémie.

Pseudo-gènes (“gènes fossiles”)

Les pseudogènes (“Gènes fossiles”)

- **Pseudogènes**
 - Gènes défectifs (ont perdu leur activité), mais dont on retrouve des traces dans les génomes.
 - Aussi appelés “gènes fossiles” mais risque de confusion avec les gènes fonctionnels dans les génomes fossiles (ex: *Homo neandertalis*)
- On trouve généralement de nombreux codons stops dans les séquences codantes.
- Succède à une relaxation de la pression sélective:
 - Mutations aléatoires sur toutes les parties du génome
 - Perturbent la fonction → éliminées (sélection naturelle « purificatrice »).
 - Occasionnellement, un gène devient dispensable → sélection relâchée → mutations s'accumulent
- Exemple : plusieurs espèces de levure (indépendamment)
 - perte de la capacité de digérer le galactose
 - Fossilisation de *chacun* des 7 gènes GAL (devenus dispensables)

La fossilisation des gènes de l'hémoglobine chez le poisson des glaces.

A



B



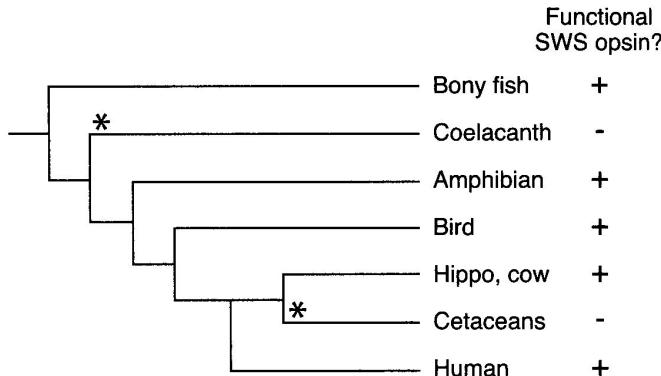
A Juvenile icefish. The transparent appearance is due to evolutionary loss of scales and red blood cells. (Photograph by Flip Micklin.)

B Adult mackerel icefish, *Champsocephalus gunnari*.

- Poissons des glaces (Channichthyidae)
- Océan Arctique, eaux de 4°C à -2°C (salée -> liquide).
- Protéines « antigels » dans le sang
- Absence de globules rouges !
- Échanges d'oxygène à travers la peau
- Ni hémoglobine ni myoglobine fonctionnelle
- Présence dans le génome de gènes fossilisés pour les deux chaînes de l'hémoglobine.

Carroll, S. B. (2006). The Making of the Fittest. DNA and the Ultimate Forensic Record of Evolution. Norton.

La perte de la perception des couleurs



* Gene fossilization

FIG. 5.2. **The same opsin gene has been fossilized twice.** The distribution of different mutations found in the coelacanth and cetacean SWS opsins and the evolutionary relationship of these species indicates that the SWS opsin was fossilized at least twice (asterisk). *Figure by Jamie Carroll.*

- Chez les coelacanthes et les cétacés, la perception des couleurs a été perdue.
- Perte indépendante chez les coelacanthes et chez les cétacés.
- Gènes des "c-opsines" (perception des couleurs) toujours présents, mais plus fonctionnels
- Grand nombre de mutations, y compris codons stops

La pseudogénisation des récepteurs olfactifs (OR) et des vomérorécepteurs (VR)

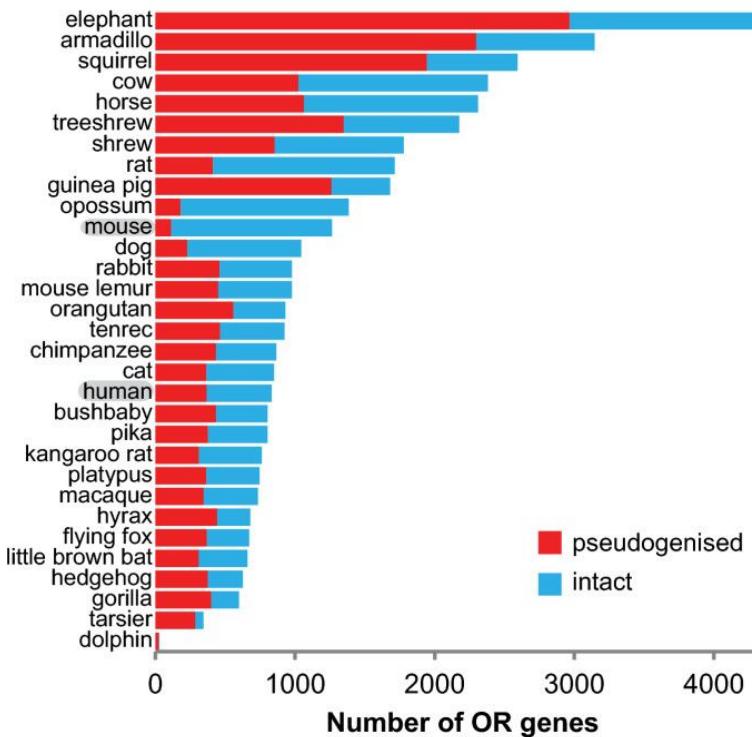
Gauche: récepteurs olfactifs (OR)

- Souris:
 - 25.000 gènes
 - 1.400 récepteurs olfactifs.
- Humain
 - 21.000 gènes
 - 400 récepteurs olfactifs
 - Perte massive de récepteurs olfactifs : fossilisés, non-fonctionnels.
- Idem chez les autres primates de l'ancien monde, qui ont acquis la vision trichromatique.

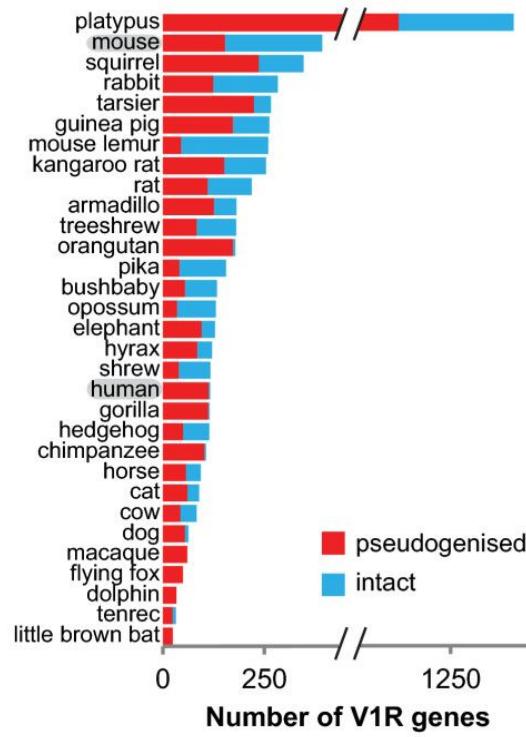
Droite: vomérorécepteurs (VR)

- Homologues des récepteurs olfactifs
- Perception des phéromones.
- Humain et gorille :
 - (quasiment) pas de communication par phéromones
 - génome comporte une centaine de vomérorécepteurs pseudogénisés.

(a)



(b)



L'évolution en marche: la fossilisation massive des gènes de *Mycobacterium leprae*

- *Mycobacterium tuberculosis*
 - Extracellulaire
 - 4.189 gènes codant pour des protéines
- *Mycobacterium leprae*
 - Intra-cellulaire
 - 1.605 gènes codant pour des protéines
 - A perdu (par délétions) ~1000 gènes présents chez *M.tuberculosis*
 - On trouve également plus de 1.000 gènes fossiles.

La fossilisation est une voie de non-retour

Dès que la pression sélective est relâchée, la probabilité de fossilisation d'un gène est très élevée. Par contre, la probabilité que de revenir à une copie fonctionnelle à partir d'un gène fossilisé est quasiment nulle.

Ces probabilités sont discutées par S. Carroll.

Quand les branches de l'arbre du vivant s'entrecroisent

Quand les branches de l'arbre du vivant s'entrecroisent

Fig. 1. Part of the only figure in the *Origin of Species*. Darwin first uses it to represent the divergence of variants within a species, showing successively more difference in a single lineage (a^1 through a^{10}) and splitting into multiple lineages (m , s , i , and so forth), some of which will become new species. Later, he expands the tree metaphor, explaining that "limbs divided into great branches ... were themselves once, when the tree was small, budding twigs; and this connection of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups" (3, p. 171).

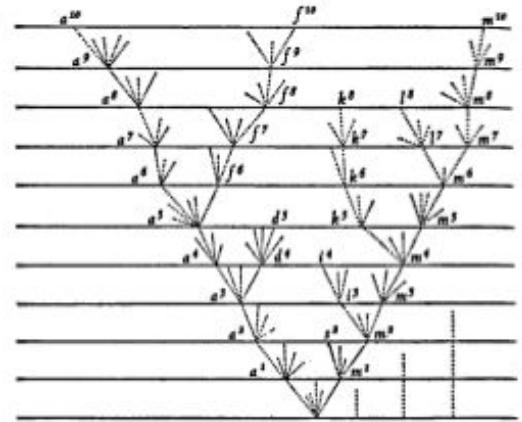


Fig. 2. The current consensus or standard model. Only a few of the "kingdoms" of the "domain" Bacteria are shown. Branching orders of several kingdoms within Bacteria and Eukarya remain in dispute. Mitochondrial and chloroplast endosymbioses are indicated by lower and upper diagonal arrows, respectively. Archaea, as a subkingdom composed of primitively amitochondriate protists, may be extinct. For SSU rRNA trees with much more detail, see (5).

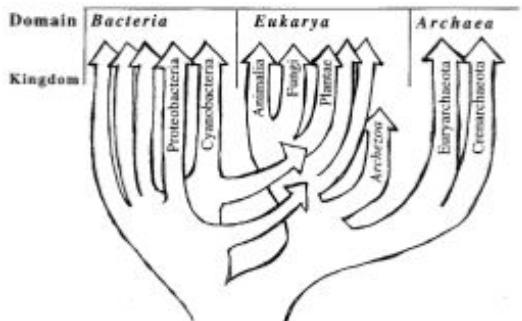
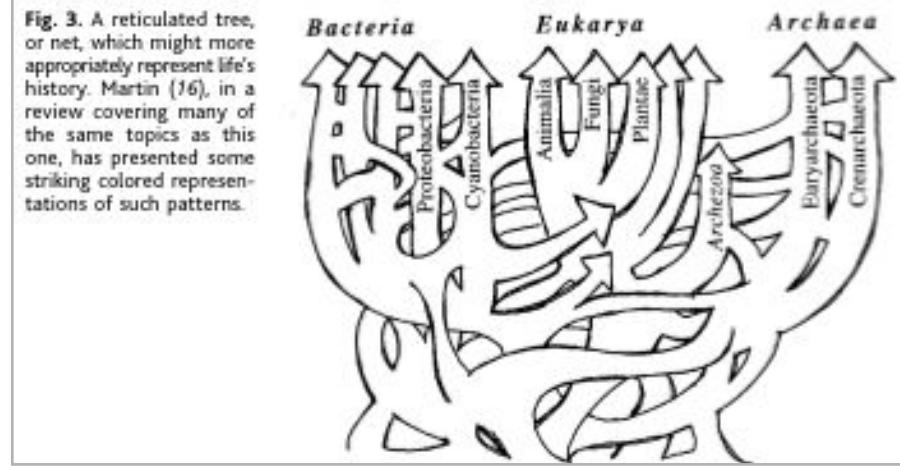


Fig 1: arbre de la vie de Darwin (Fig 1)

Fig 2-3: revisité par Doolittle (1999)

- Fig 2: événements d'**endosymbiose** liés à l'apparition des organelles des eucaryotes (mitochondrie et chloroplaste).
- Fig 3: **transferts horizontaux** entre génomes de procaryotes.

Fig. 3. A reticulated tree, or net, which might more appropriately represent life's history. Martin (16), in a review covering many of the same topics as this one, has presented some striking colored representations of such patterns.

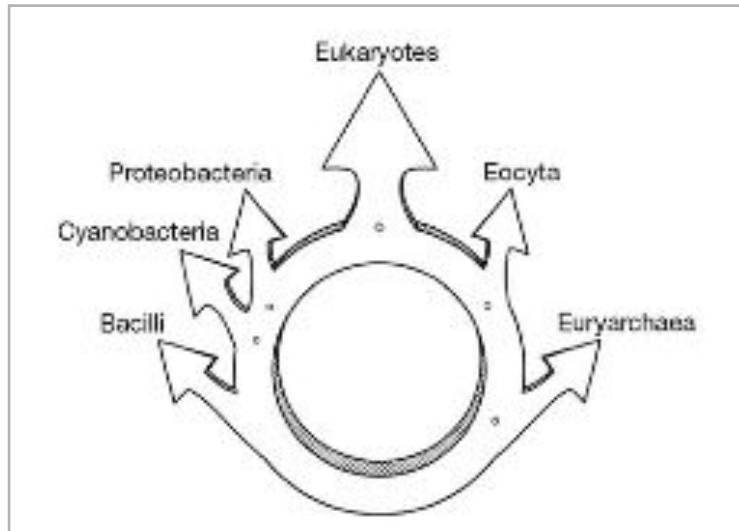


The ring of life provides evidence for a genome fusion origin of eukaryotes

Maria C. Rivera^{1,3,4} & James A. Lake^{1,3,4}

¹Molecular Biology Institute, MCD Biology, ²Human Genetics, ³IGPP, and ⁴Astrobiology Institute, University of California, Los Angeles 90095, USA

Genomes hold within them the record of the evolution of life on Earth. But genome fusions and horizontal gene transfer seem to have obscured sufficiently the gene sequence record such that it is difficult to reconstruct the phylogenetic tree of life. Here we determine the general outline of the tree using complete genome data from representative prokaryotes and eukaryotes and a new genome analysis method that makes it possible to reconstruct ancient genome fusions and phylogenetic trees. Our analyses indicate that the eukaryotic genome resulted from a fusion of two diverse prokaryotic genomes, and therefore at the deepest levels linking prokaryotes and eukaryotes, the tree of life is actually a ring of life. One fusion partner branches from deep within an ancient photosynthetic clade, and the other is related to the archaeal prokaryotes. The eubacterial organism is either a proteobacterium, or a member of a larger photosynthetic clade that includes the Cyanobacteria and the Proteobacteria.



- Relations entre tous les gènes d'eucaryotes, d'eubactéries, et d'archées.
- eucaryotes résulteraient d'une fusion entre bactérie et archée.
- Gènes hérités des archées : majoritairement maintien de la cellule (réplication, transcription, régulation).
- Gènes hérités des bactéries : majoritairement métabolisme.