

Darwin Day 2020

L'Arbre du Vivant



Prof. Denis BAURAIN
InBioS / Sciences de la Vie

Plan

1. L'Arbre du Vivant

- Quels sont les grands groupes d'êtres vivants ?
- Les 5 règnes du Vivant
- Les 3 domaines du Vivant

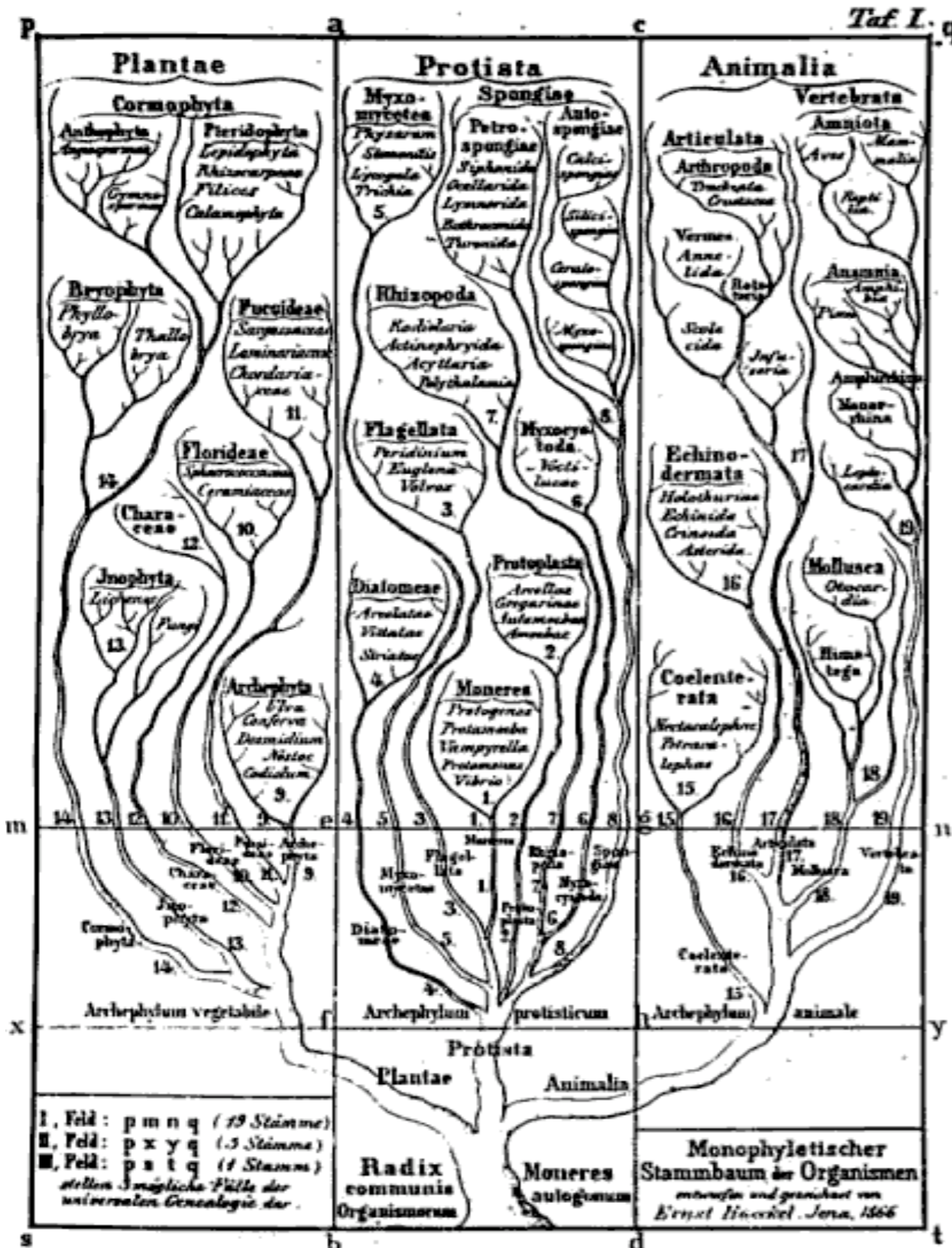
2. Phylogénie moléculaire

- D'où vient le signal ?
- Exemple : le cytochrome c

3. Phylogénomique

- Une analogie linguistique
- Exemple : l'arbre des Vertébrés

L'Arbre du Vivant

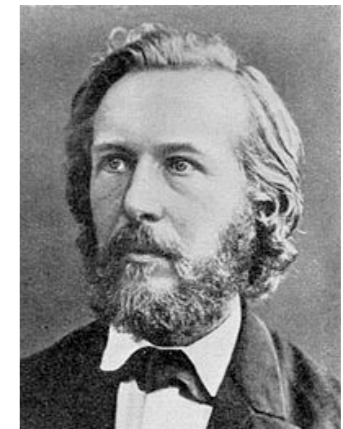


L'arbre du Vivant est une **classification** des êtres vivants tenant compte de leur histoire évolutive.

1859 1866



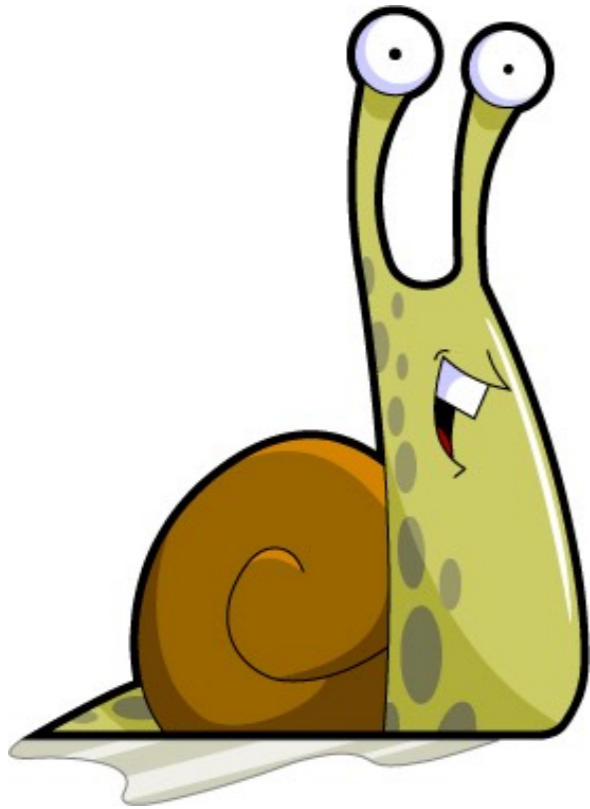
Ch. Darwin



E. Haeckel

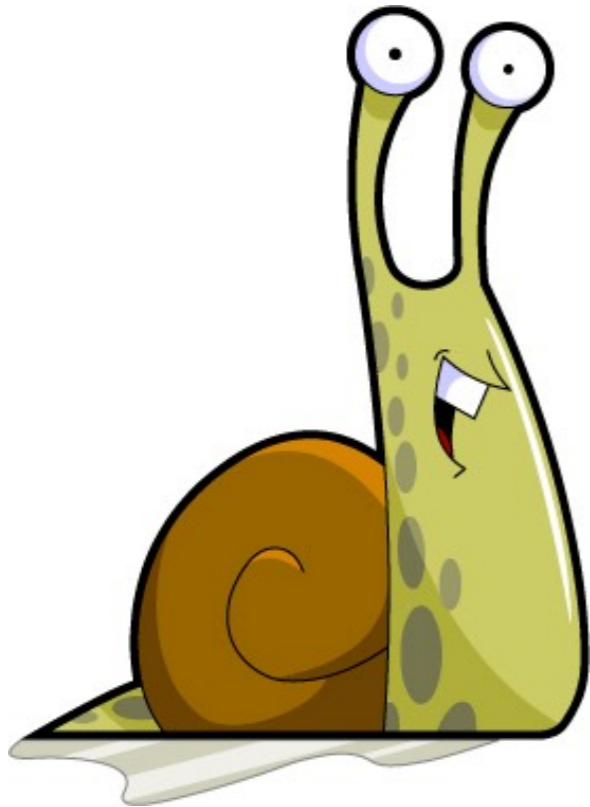
*Quels sont les grands
groupes d'êtres vivants ?*

Groupes d'êtres vivants



animaux

Groupes d'êtres vivants



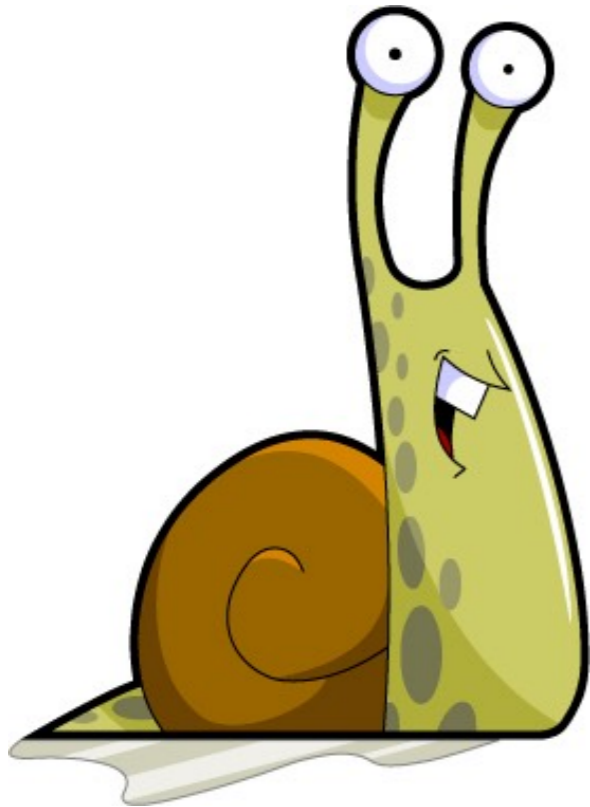
animaux



www.shutterstock.com · 48126004

plantes

Groupes d'êtres vivants



animaux

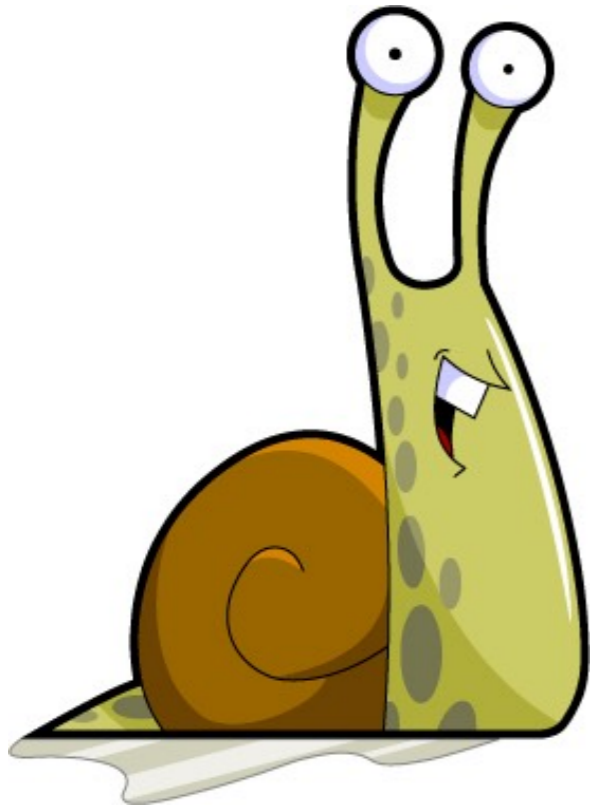


www.shutterstock.com · 48126004

plantes



champignons



animaux

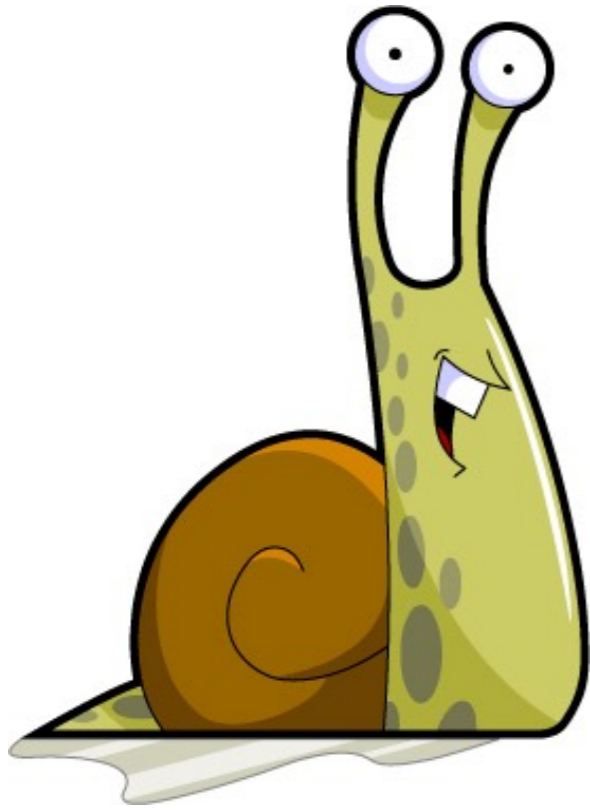


www.shutterstock.com · 48126004

plantes



champignons



animaux

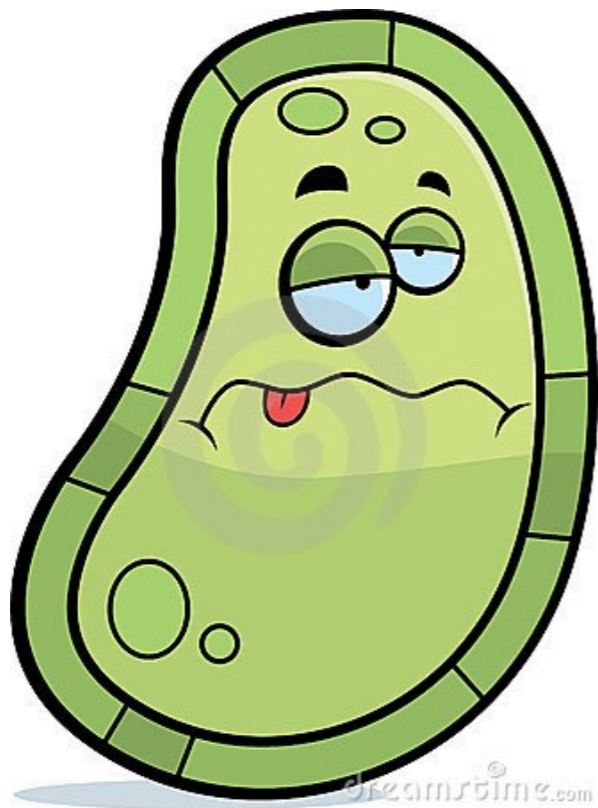


www.shutterstock.com · 48126004

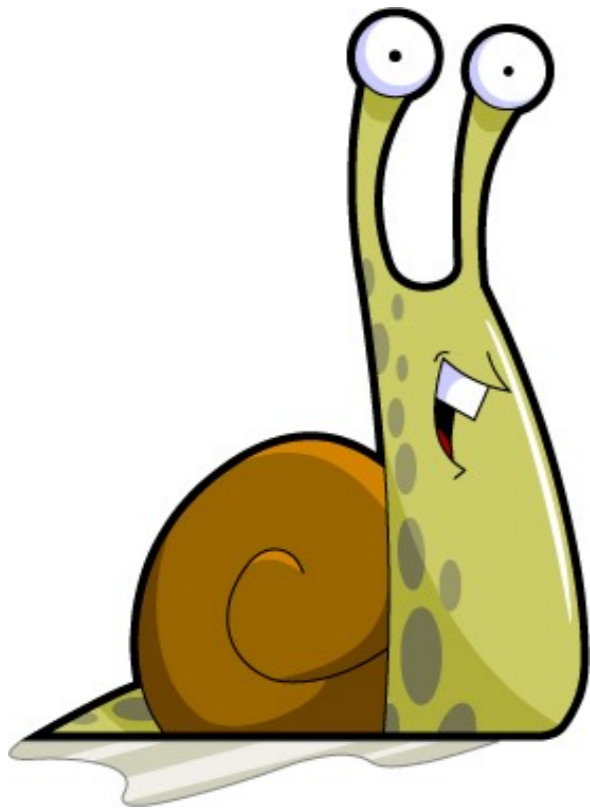
plantes



champignons



microbes



animaux

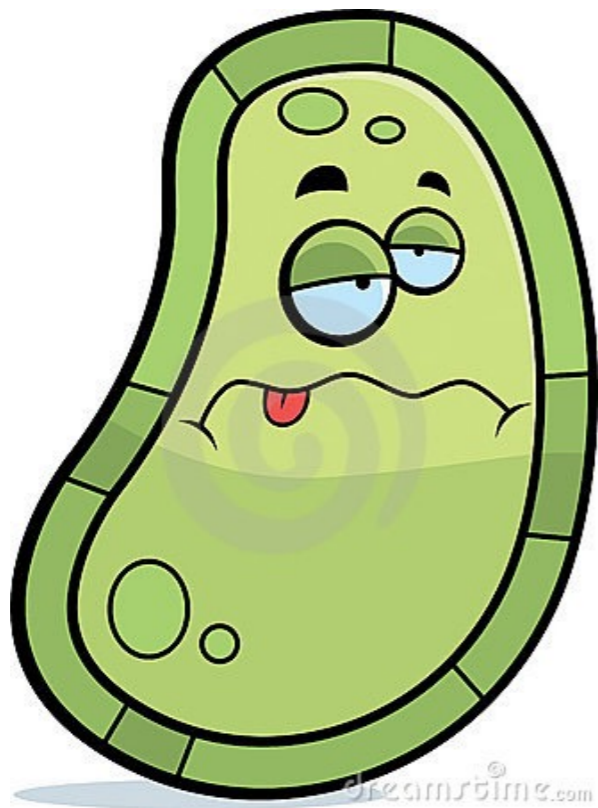


www.shutterstock.com · 48126004

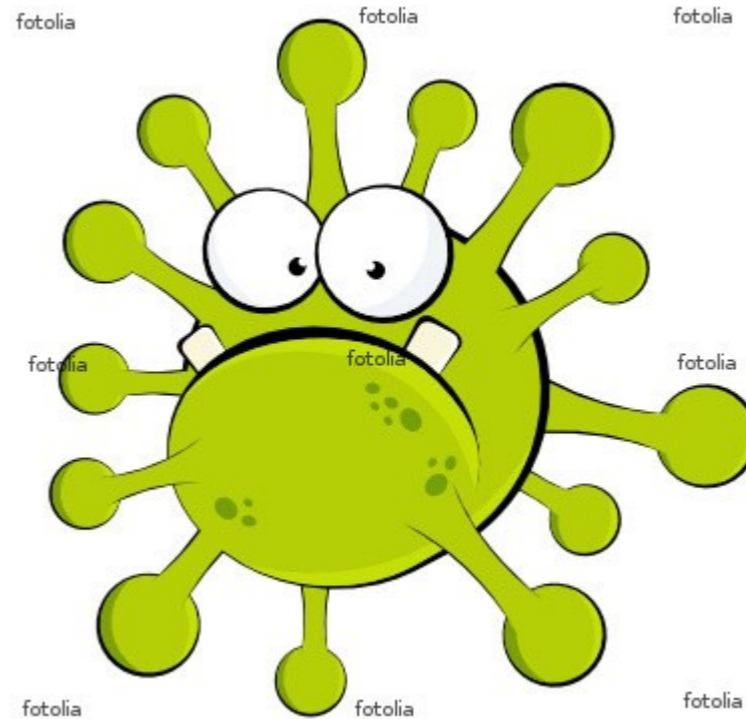
plantes



champignons

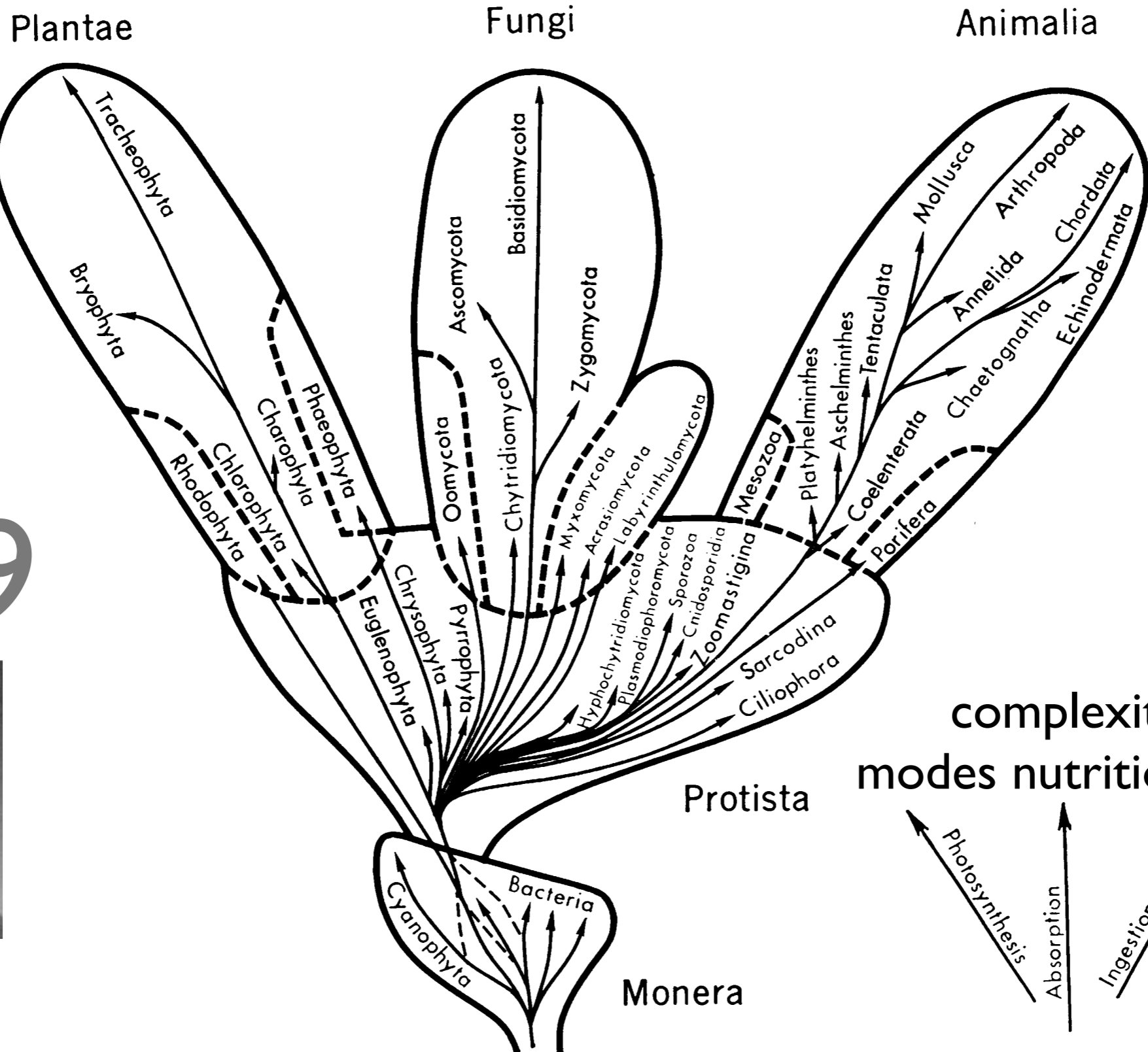


bactéries



virus

Les 5 règnes du Vivant

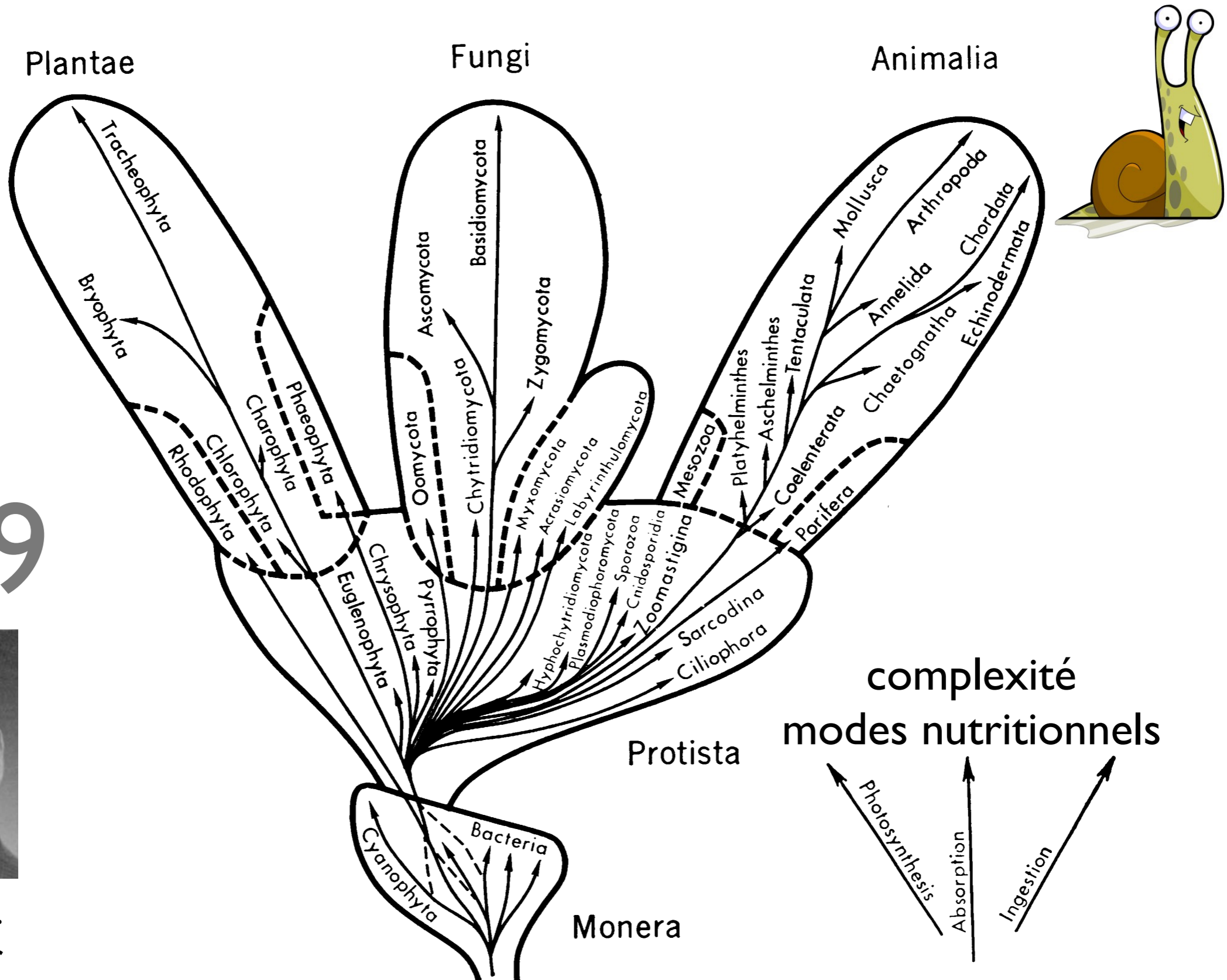


1969



Robert H. Whittaker

Les 5 règnes du Vivant



1969



Robert H. Whittaker

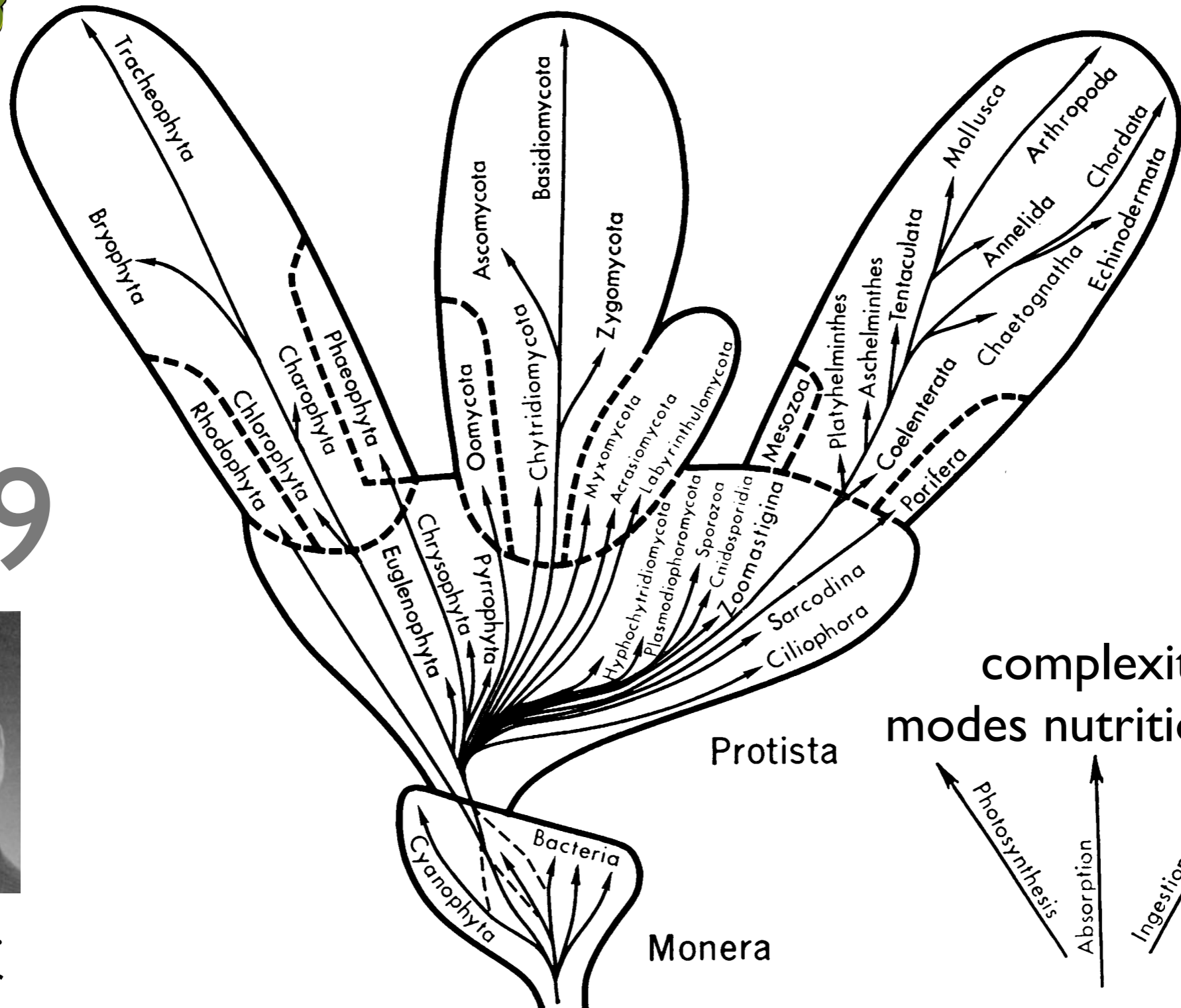
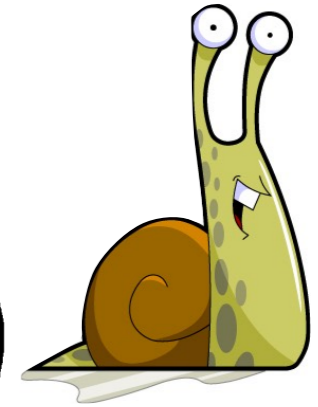
Les 5 règnes du Vivant



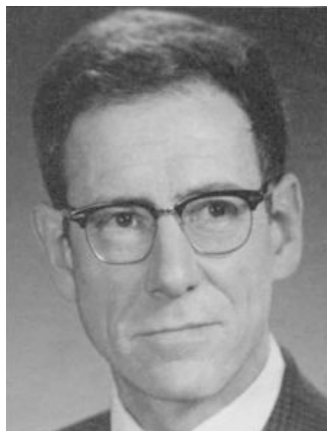
Plantae

Fungi

Animalia

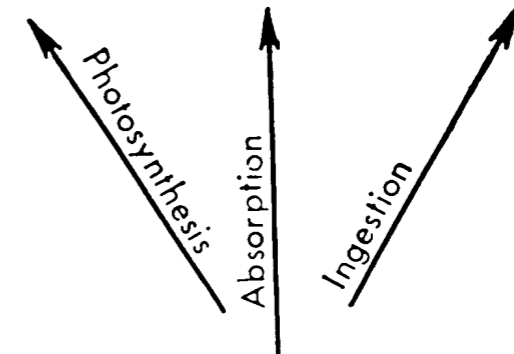


1969



Robert H. Whittaker

complexité
modes nutritionnels



Les 5 règnes du Vivant

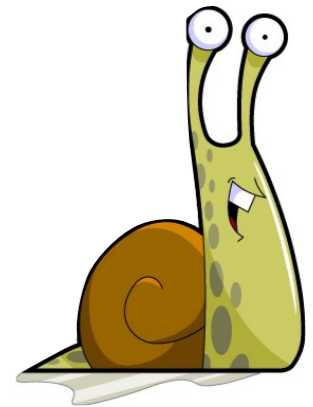


Plantae



Fungi

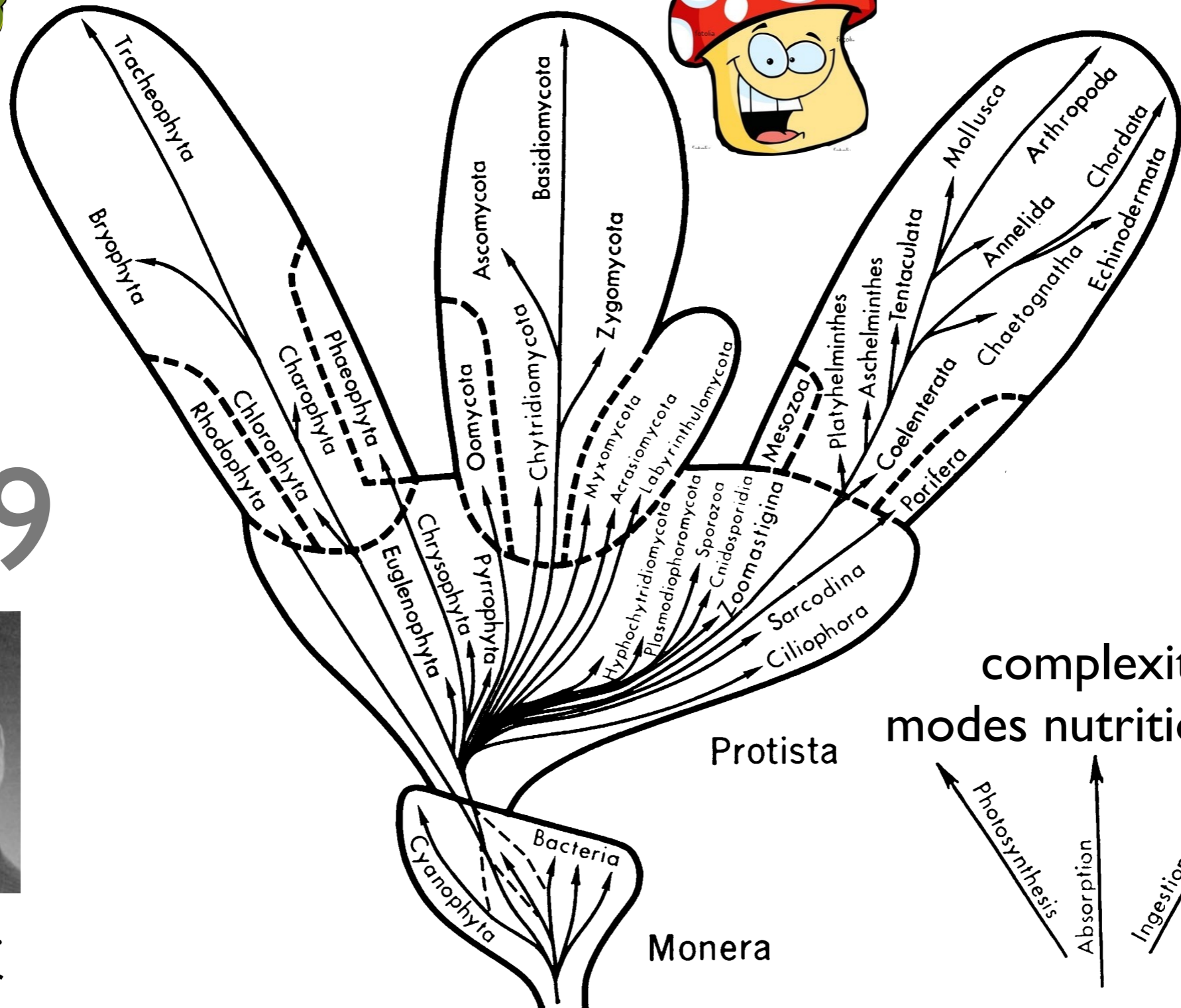
Animalia



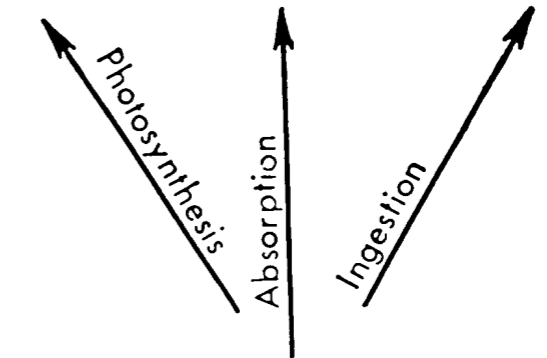
1969



Robert H. Whittaker



complexité
modes nutritionnels



Les 5 règnes du Vivant

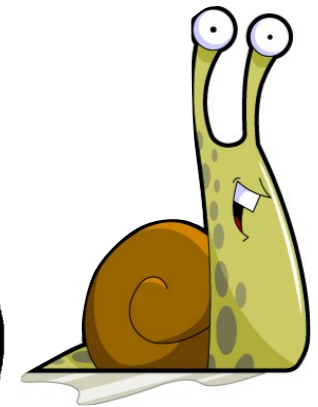


Plantae



Fungi

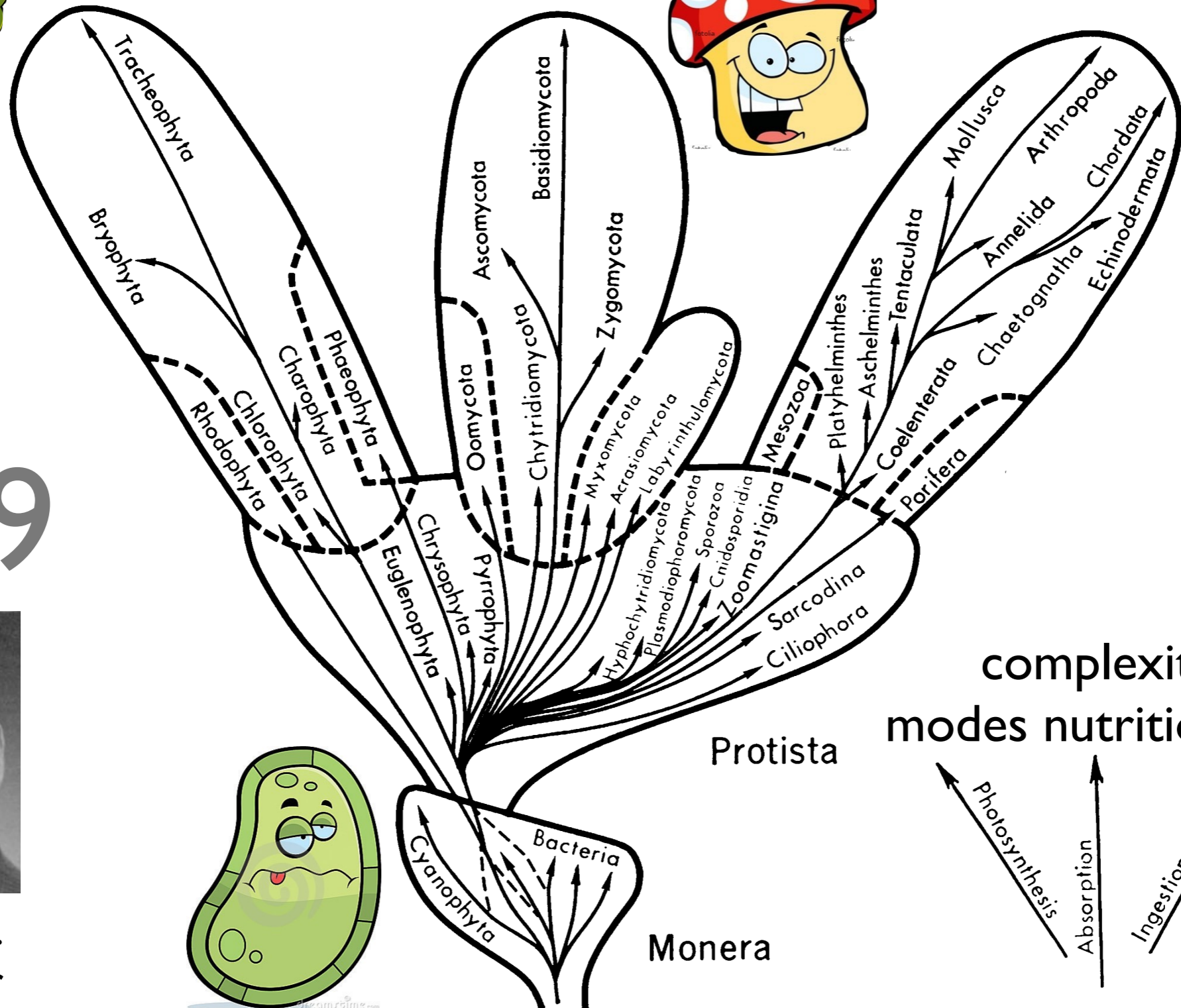
Animalia



1969



Robert H. Whittaker



complexité
modes nutritionnels

Les 5 règnes du Vivant

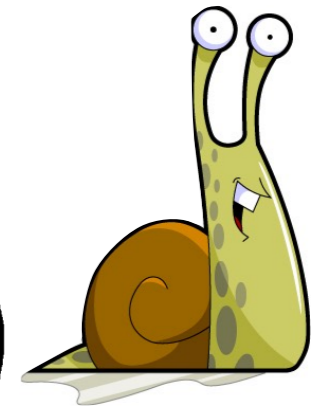


Plantae



Fungi

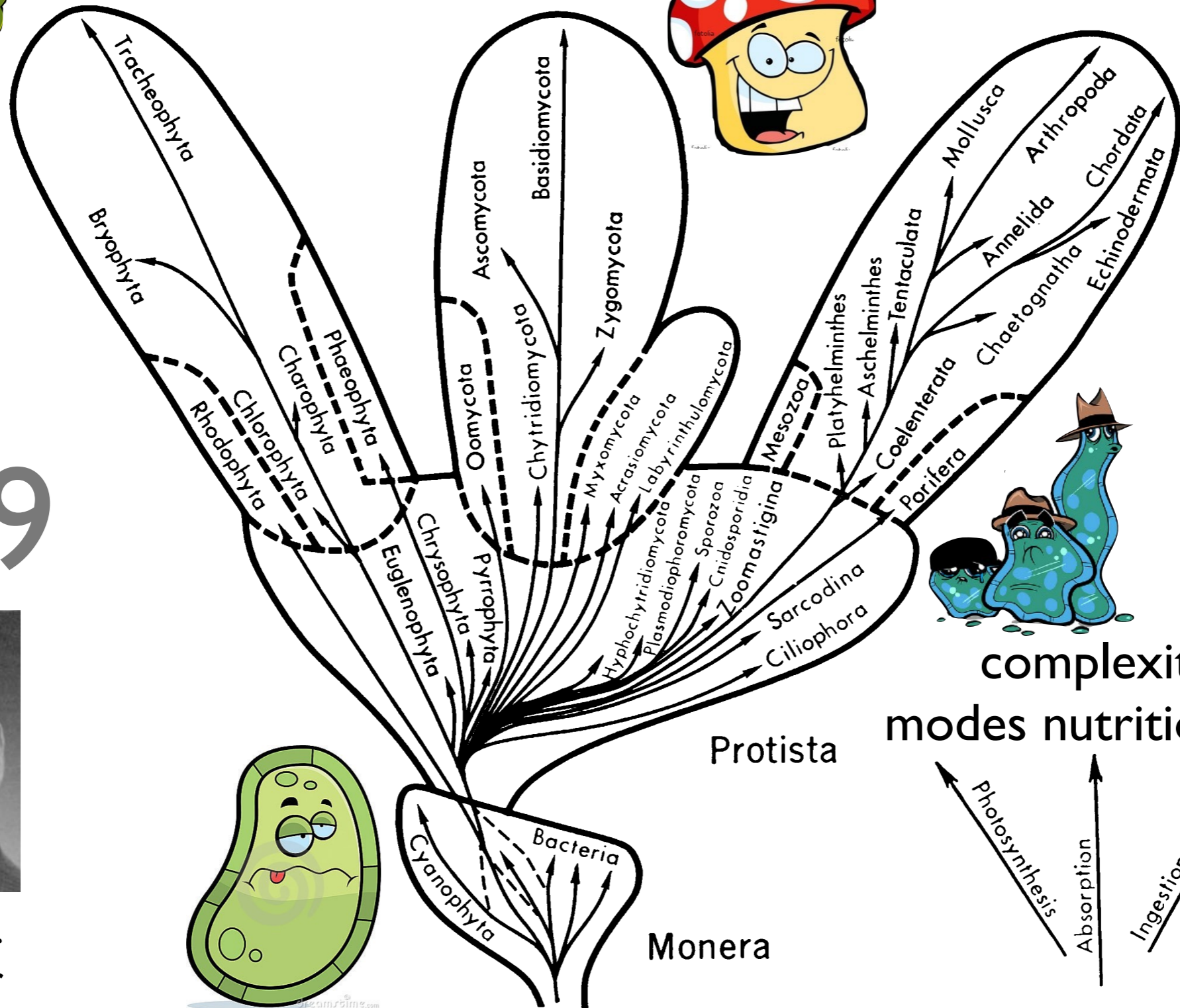
Animalia



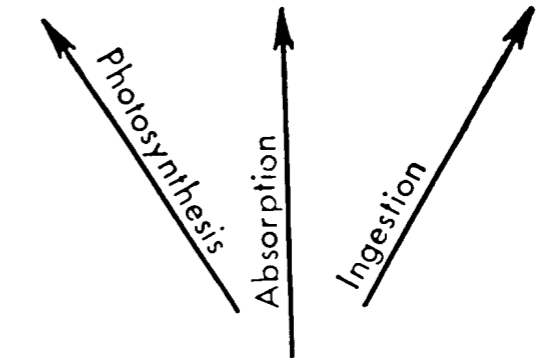
1969



Robert H. Whittaker



complexité
modes nutritionnels



Les 5 règnes du Vivant

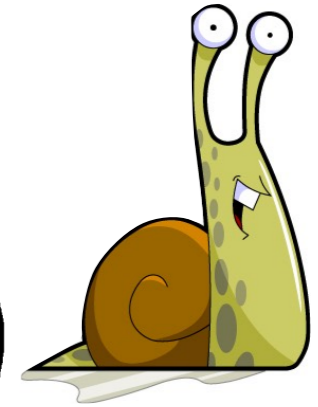


Plantae



Fungi

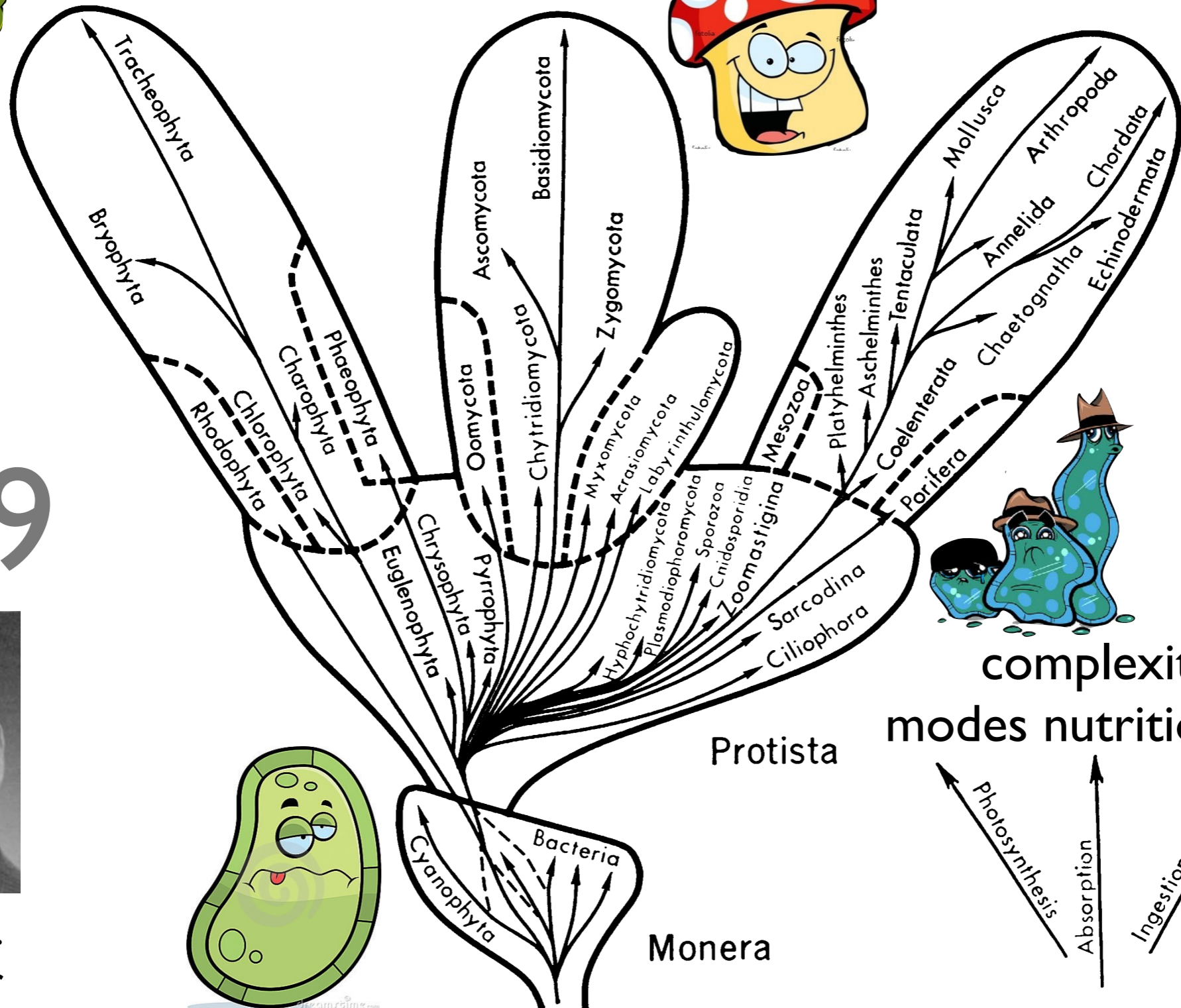
Animalia



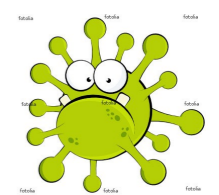
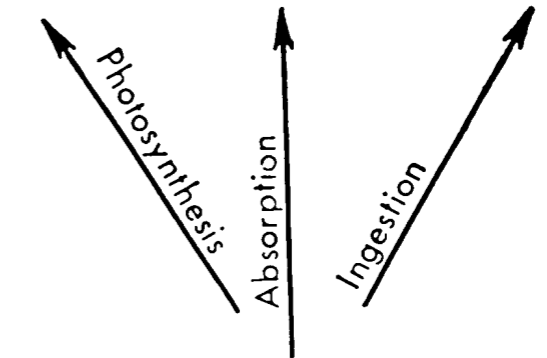
1969



Robert H. Whittaker



complexité
modes nutritionnels



Echelle \neq Arbre du Vivant

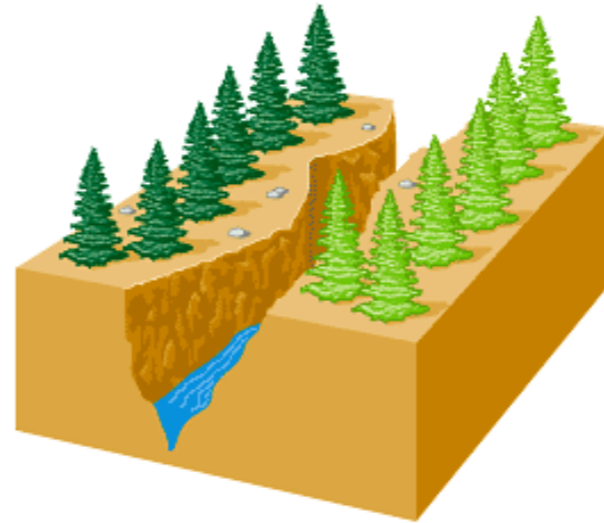
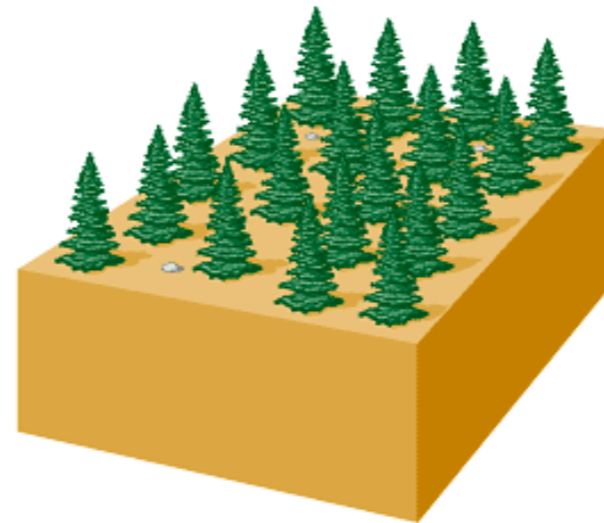
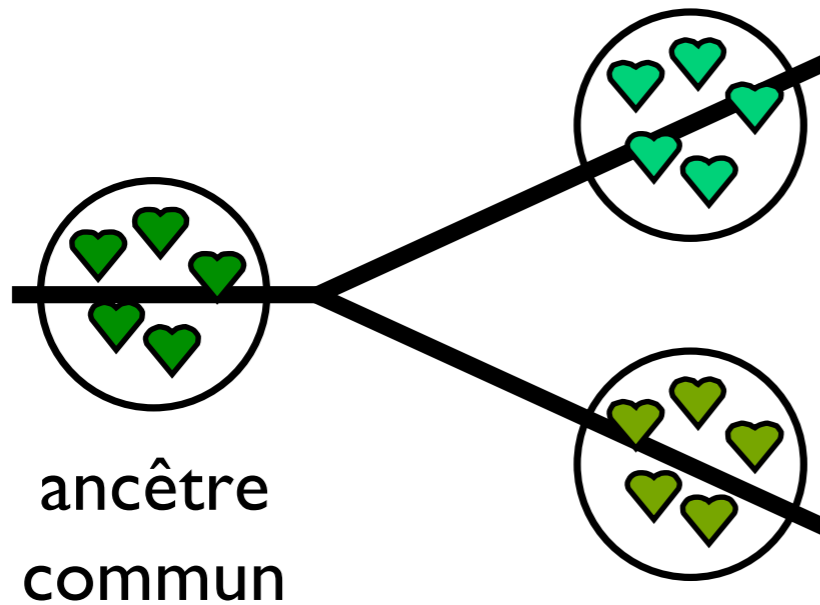


wrong



right

Phylogénie = Spéciations



spéciation allopatrique

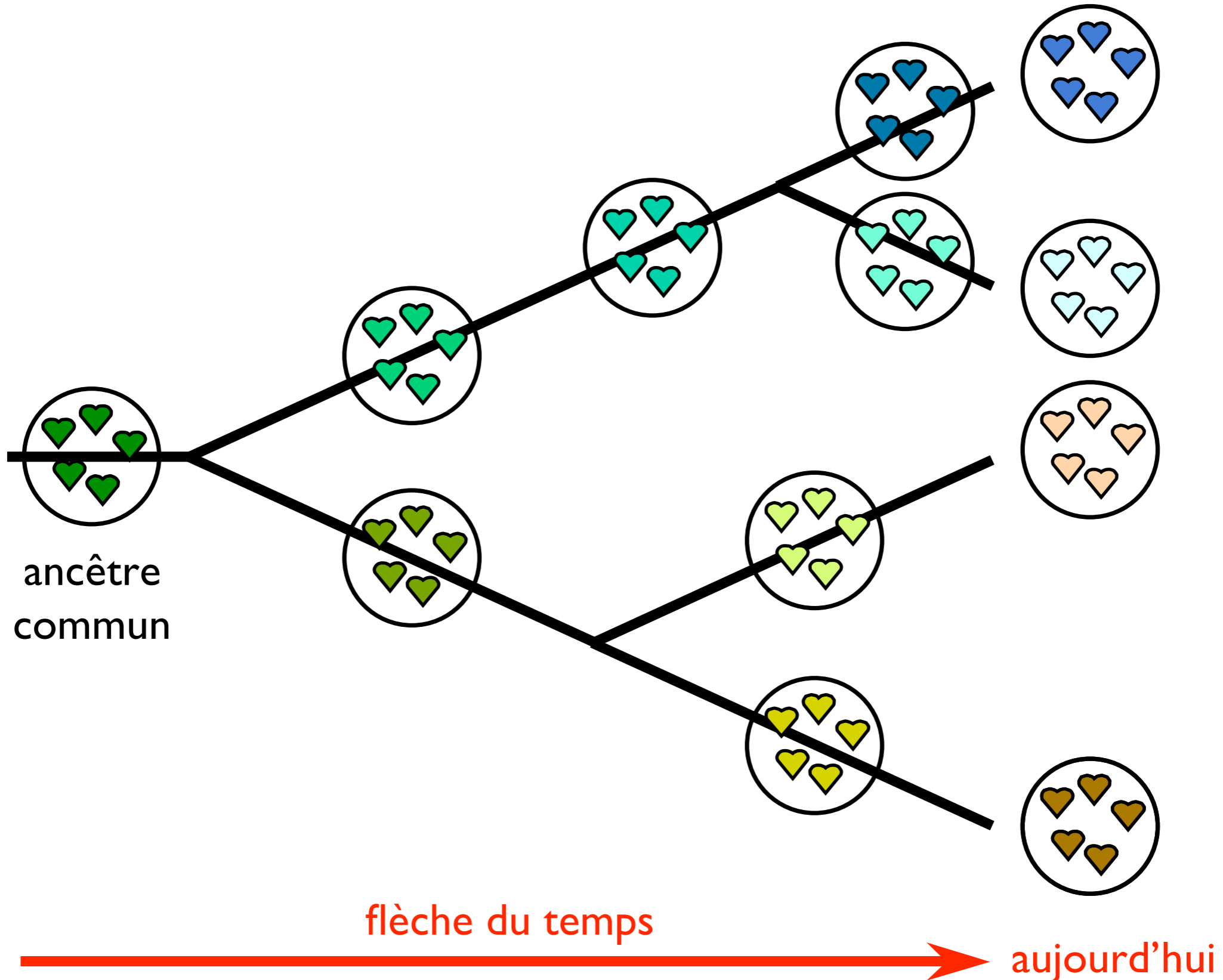
spéciation sympatrique

flèche du temps

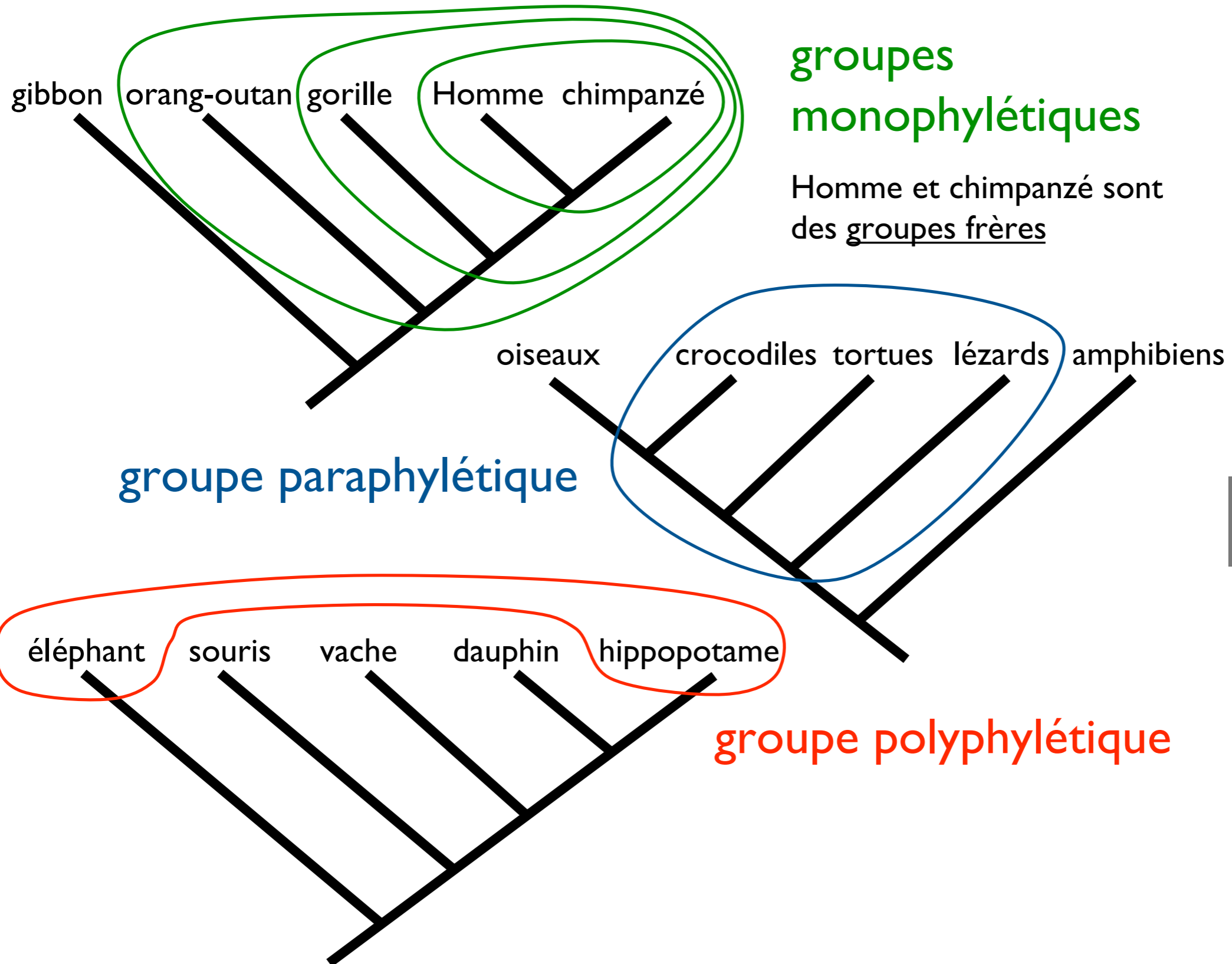


aujourd'hui

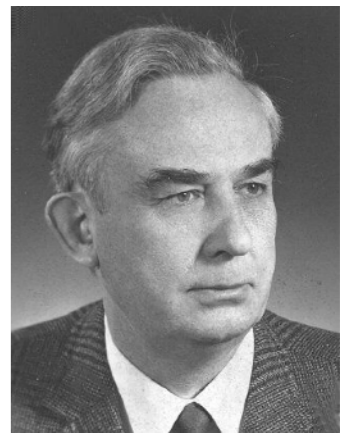
Phylogénie = Spéciations



Systematique phylogénétique

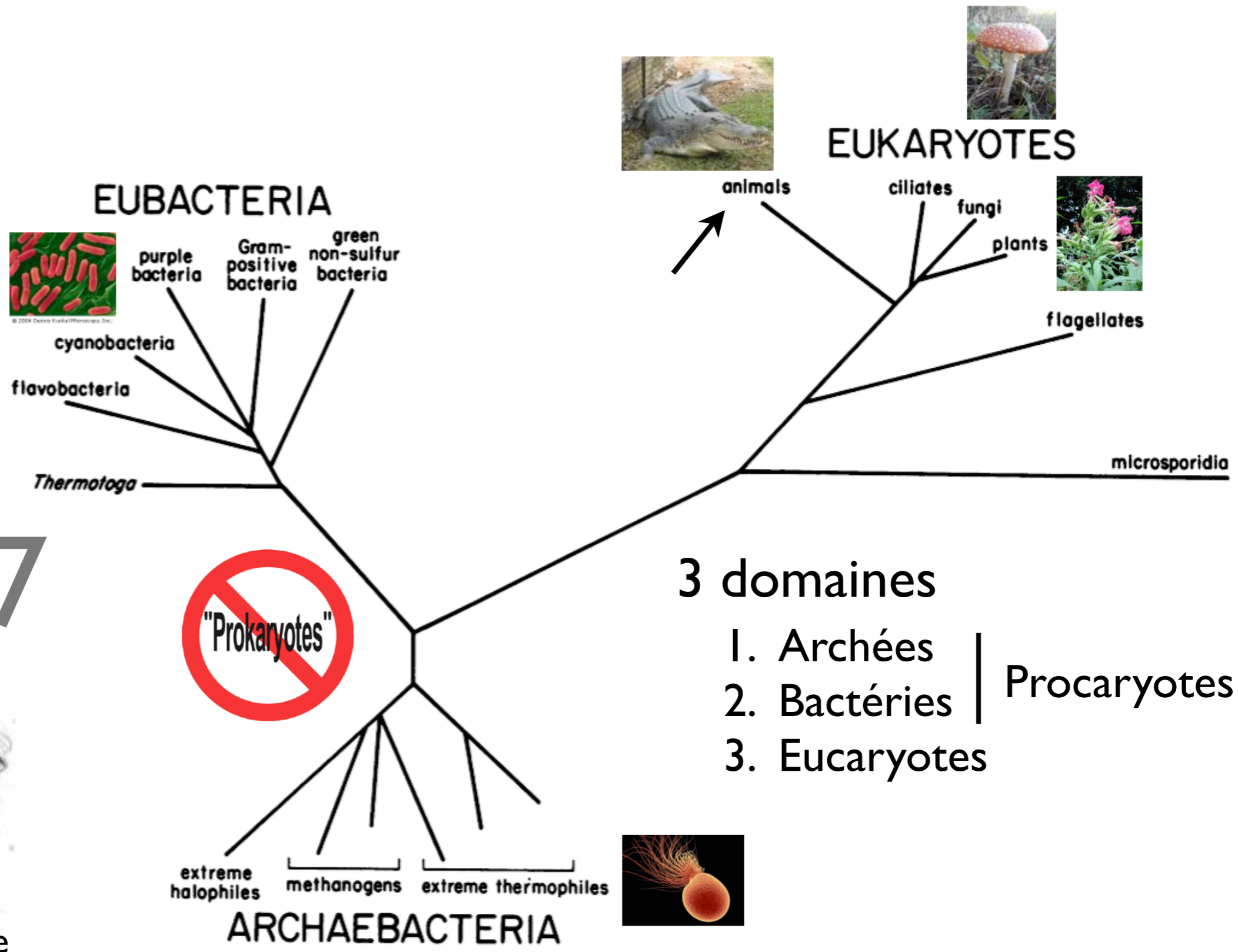


1966

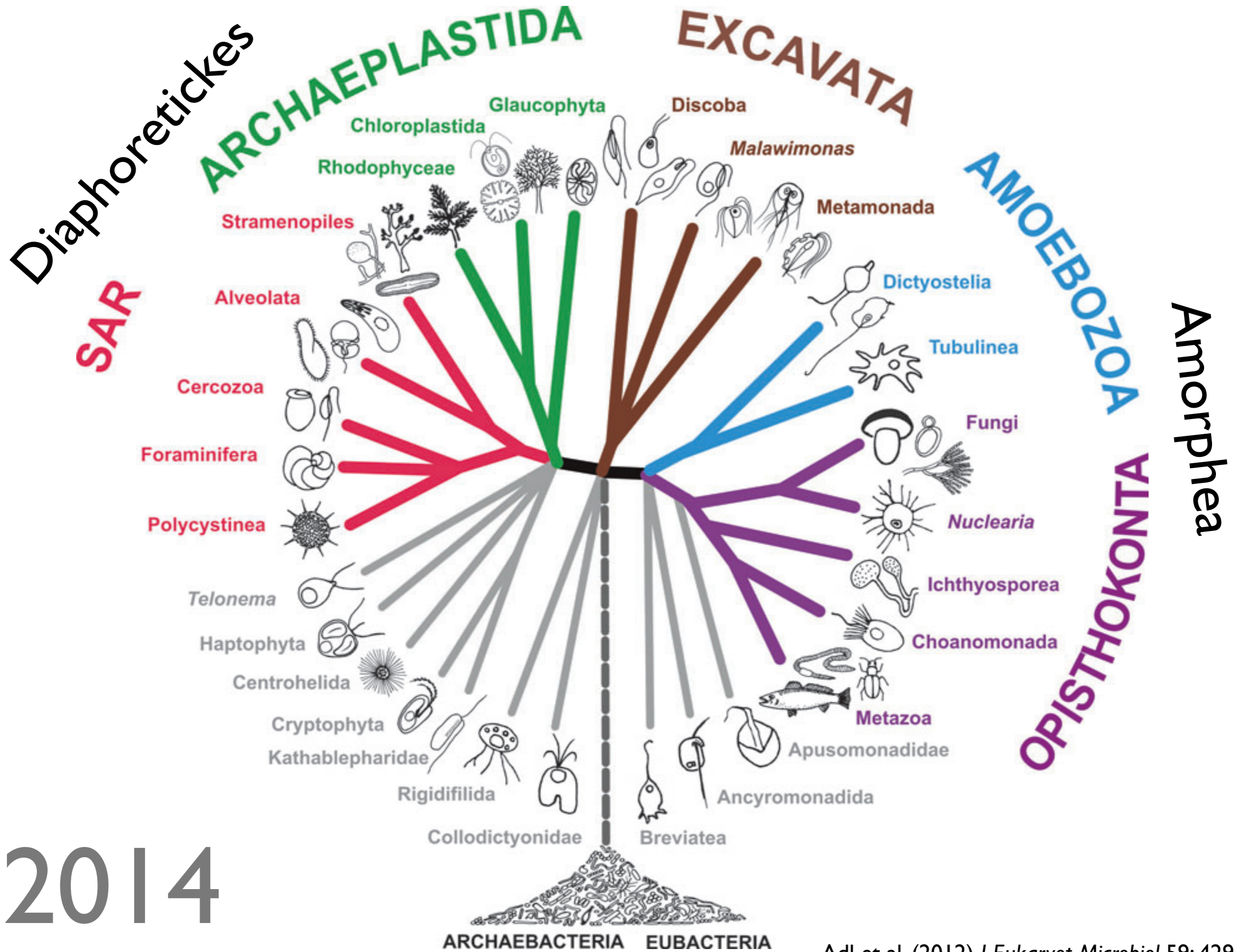


Willi Hennig

Les 3 domaines du Vivant



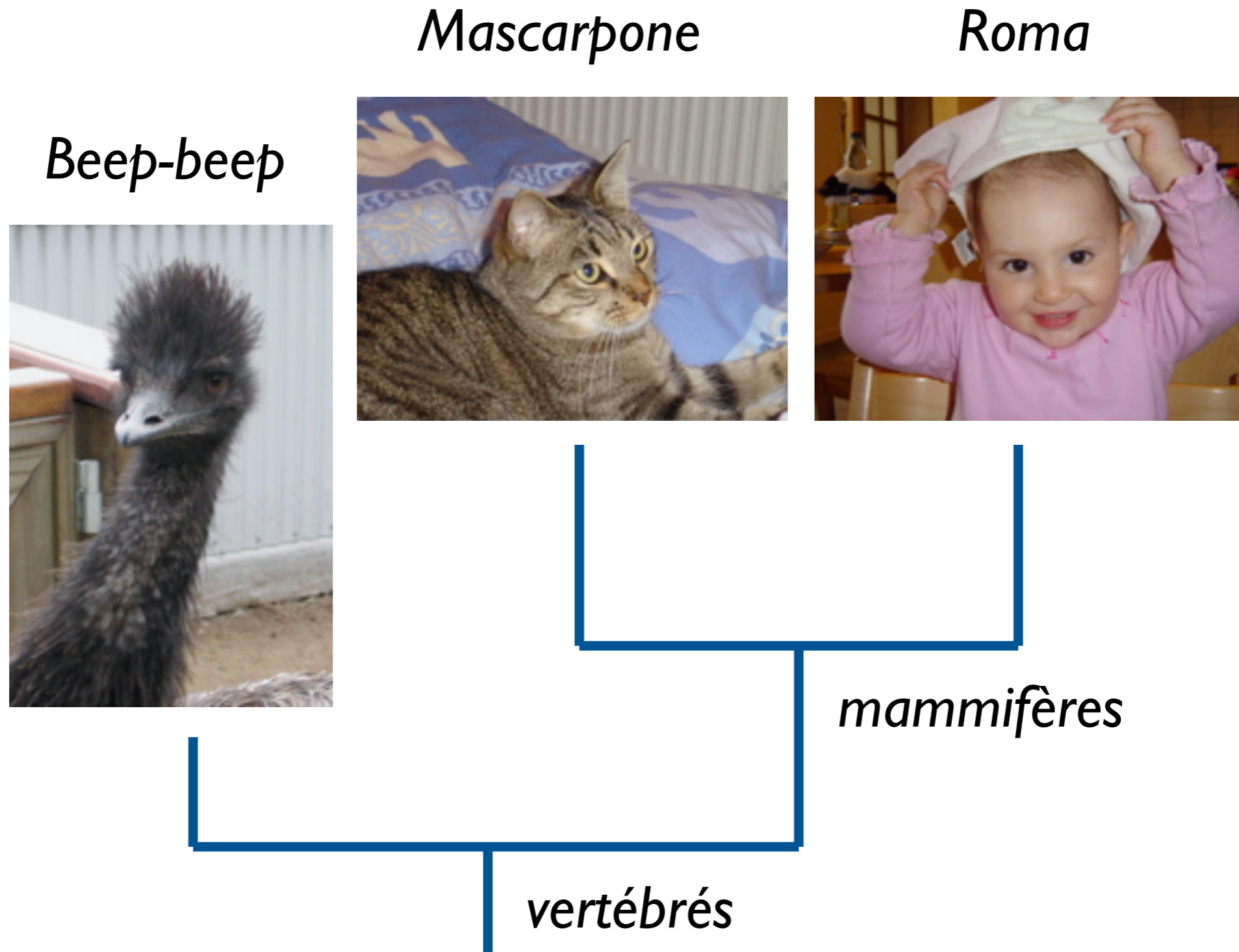
Carl Woese



2014

Phylogénie morphologique

On compare les caractères qu'on peut observer à l'oeil.



Phylogénie moléculaire

On compare les génomes sous-tendant les phénotypes.



Mónica & Penélope : *Homo sapiens*



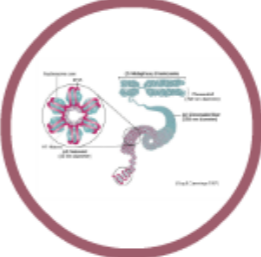
César : *Pan troglodytes*



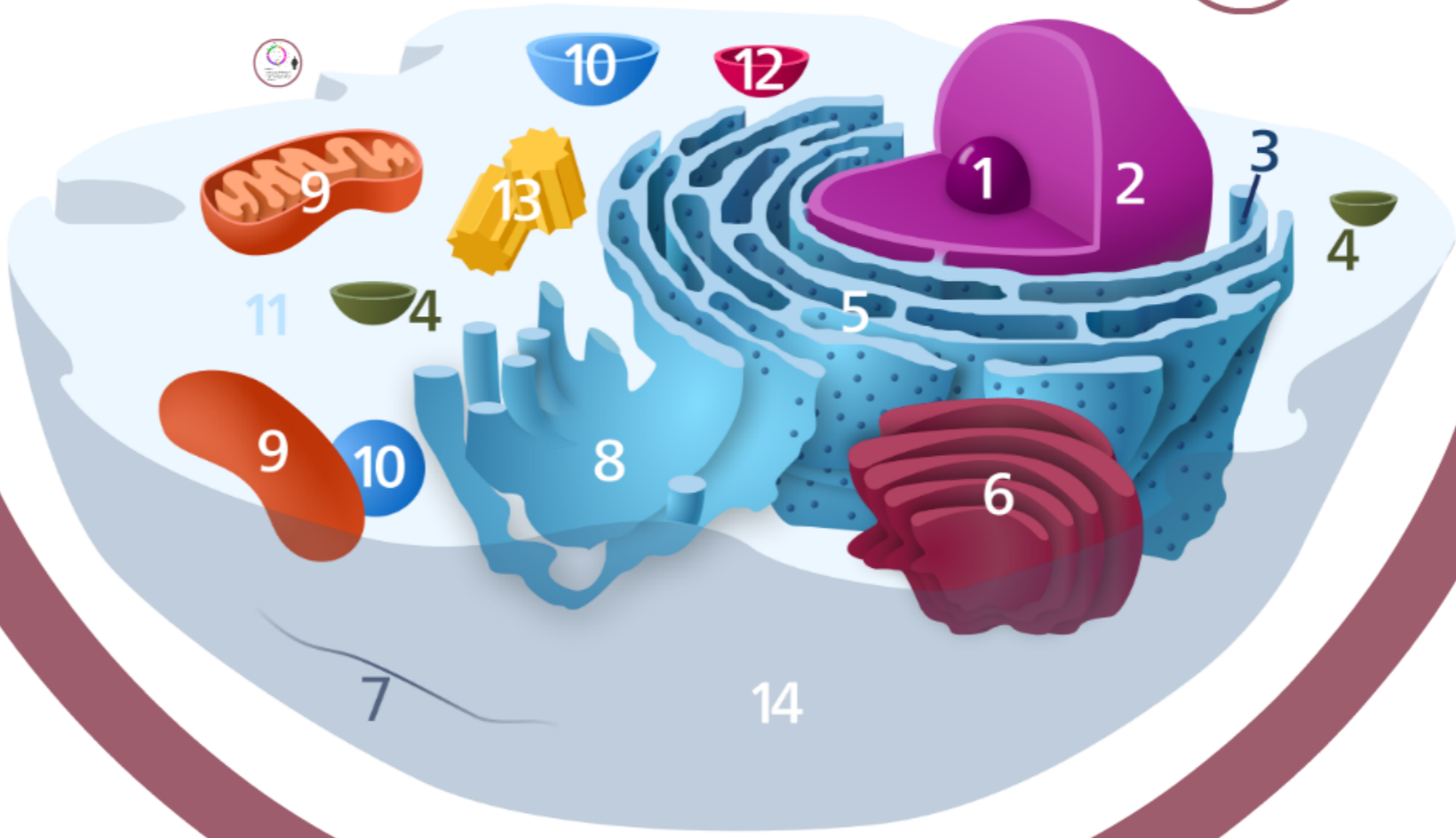
Andy : *Homo sapiens*

D'où vient le signal ?

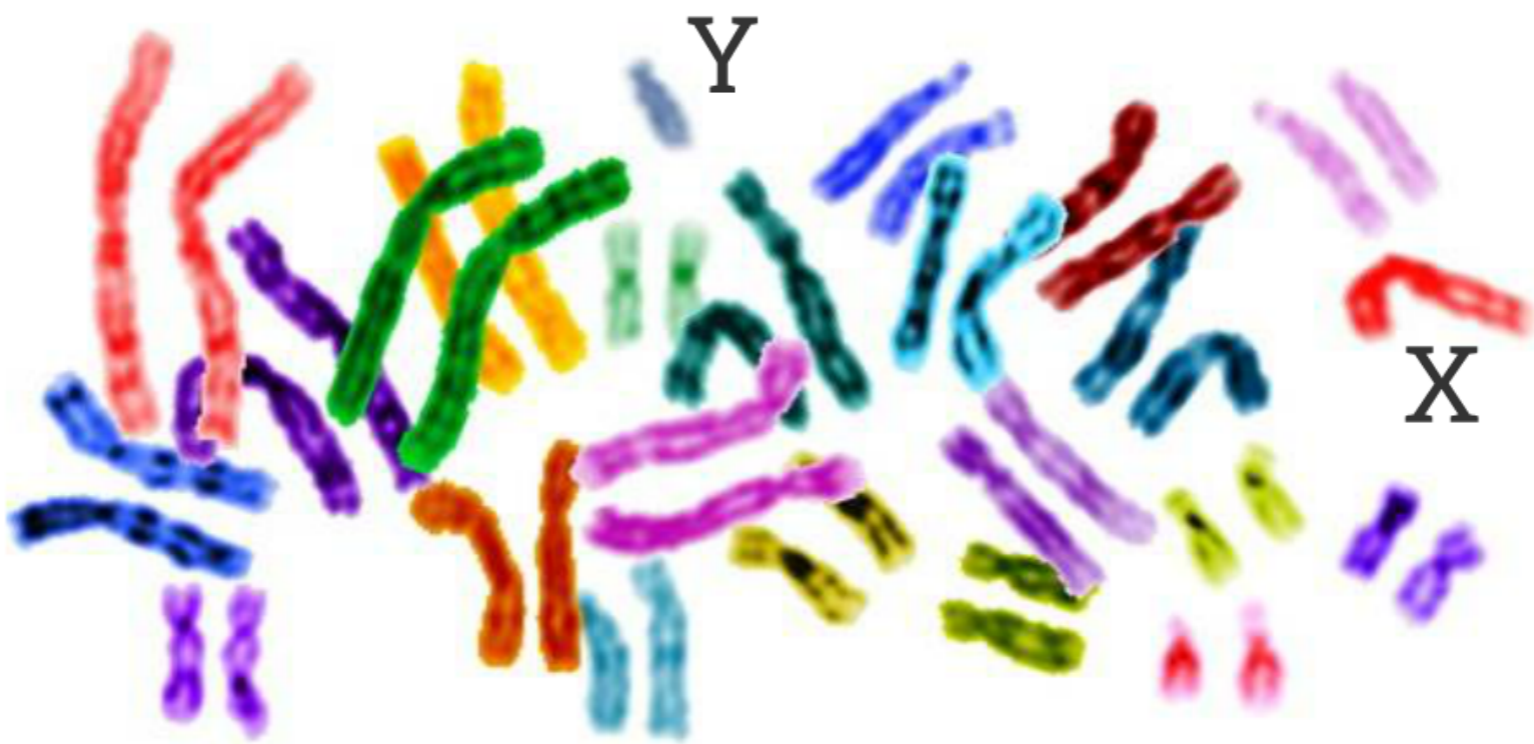
```
ACGACTCA
GATAGCATACATACAT
AACGACTATACATACGACT
CAACCTCAGACCCATACTCAGCAT
ACGACATAOGAACGACTACAGCAGCA
GACATCAGCAGTACGACATCAGCATCGA
CACGCTCAGACTCTCCATACTCAGCATAAG
ATACGACATACGACTACAGCAGCAGCAGCAT
GCAGACTACGACATCAGCATOGACAGCATA
GACAGCATAOGGACTACAAGCTACAGCA
ACCTACATCGACTACATCATCAGACATCA
ATCATAOGGACAGCATAGACATACTAGCT
TCTATATCACAGACAATCTACTAAGC
GTACACTACAATCGAGACAGGCTC
TCTCCATACCGCAAGGAGCG
TGAGACTTACCGAAC
TATAGCT
```



D'où vient l'information ?



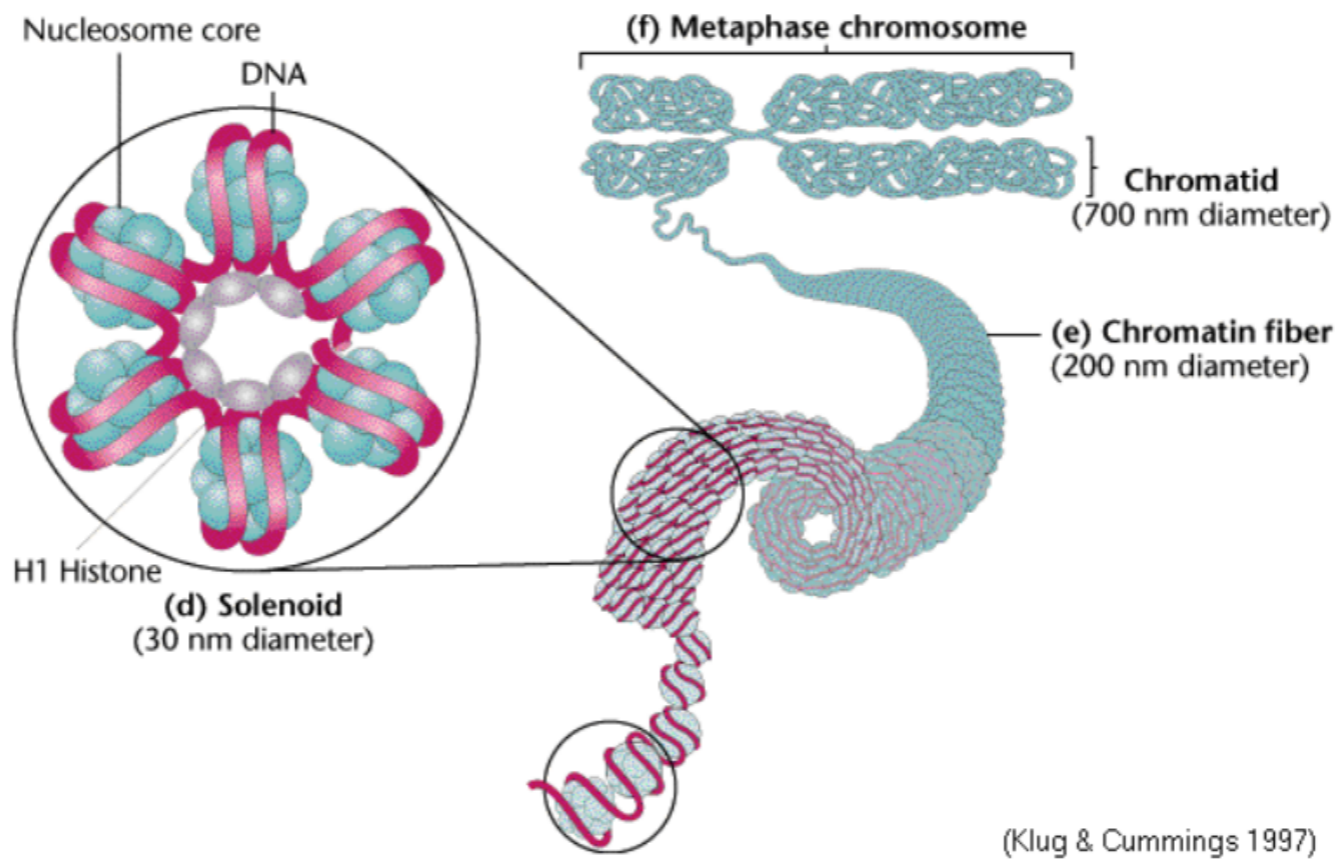
D'où vient le signal ?



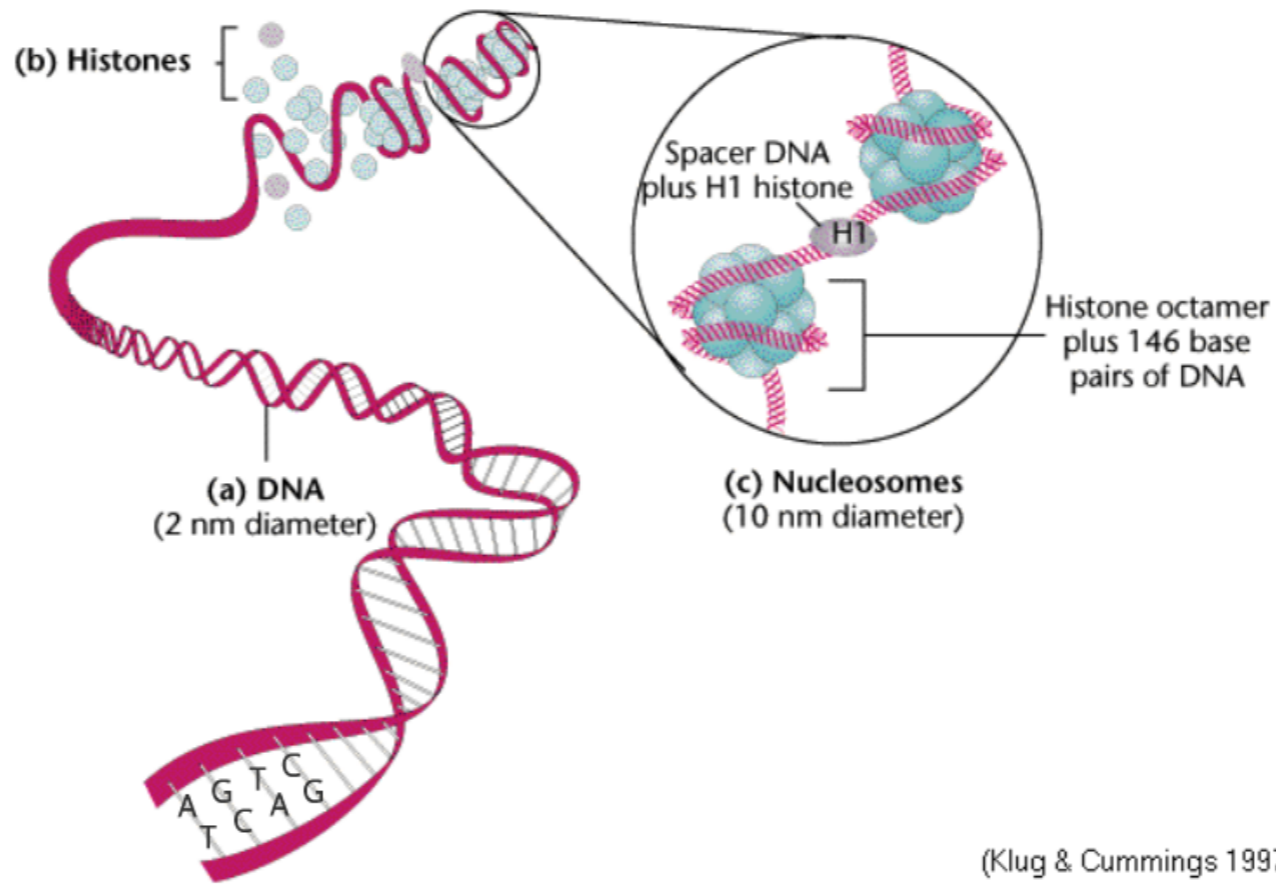
Caryotype humain :

- 22 paires d'autosomes
- 2 chromosomes sexuels

D'où vient le signal ?



D'où vient le signal ?



Génome humain :

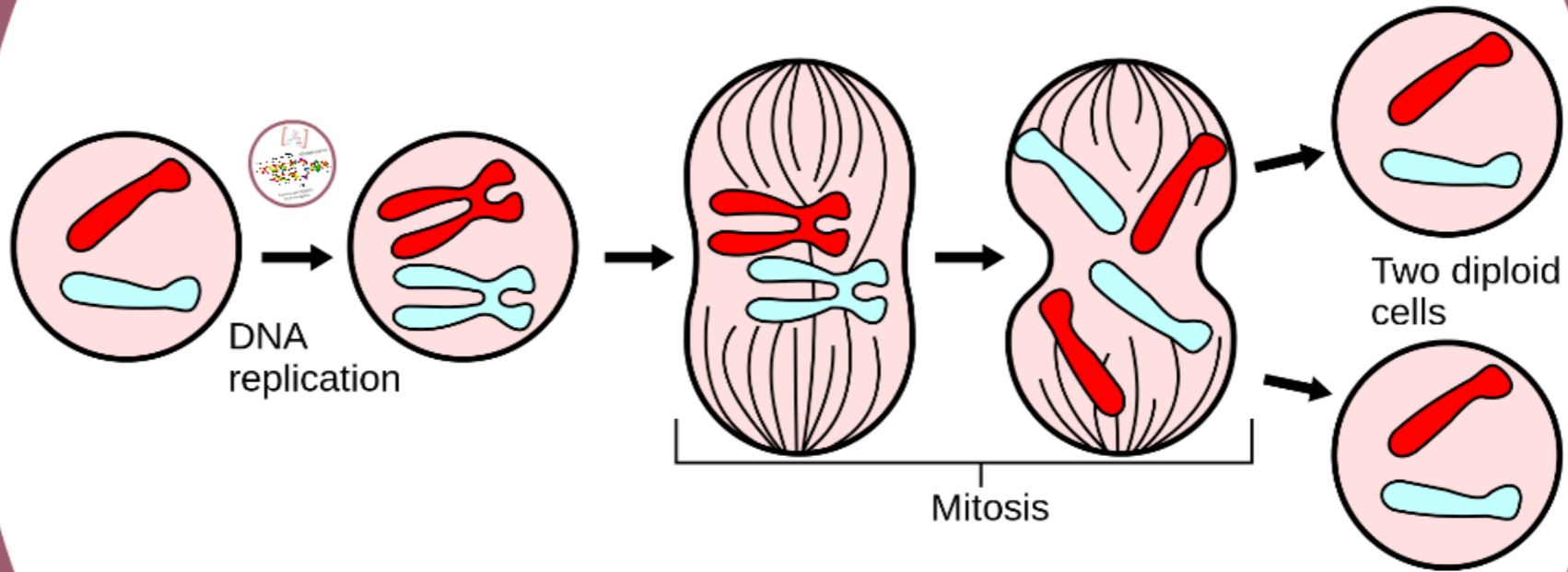
- 3 milliards de pb (x2)

D'où vient le signal ?

ACGACTCA
GATAGCATAACAT
AACGACTATACATACGACT
CAACCTCAGACCCATACTCAGCAT
ACGACATACGAACGACTACAGCAGCA
GACATCAGCAGTACGACATCAGCATCGA
CACGCTCAGACTCTCCATACTCAGCATAACG
ATACGACATACGACTACAGCAGCACGACAT
GCAGACTACGACATCAGCATCGACAGCATA
GACAGCATAACGAGACTACAAGCTACAGCA
ACCTACATCGACTACATCATCAGACATCA
ATCATACGACAGCATAGACATACTAGCT
TCTATATCACAGACAATCTACTAAGC
GTACACTACAATCGAGACACGCTC
TCTCCATAACCGCAAGGAGCG
TGAGACTTACCGAAC
TATAGCT

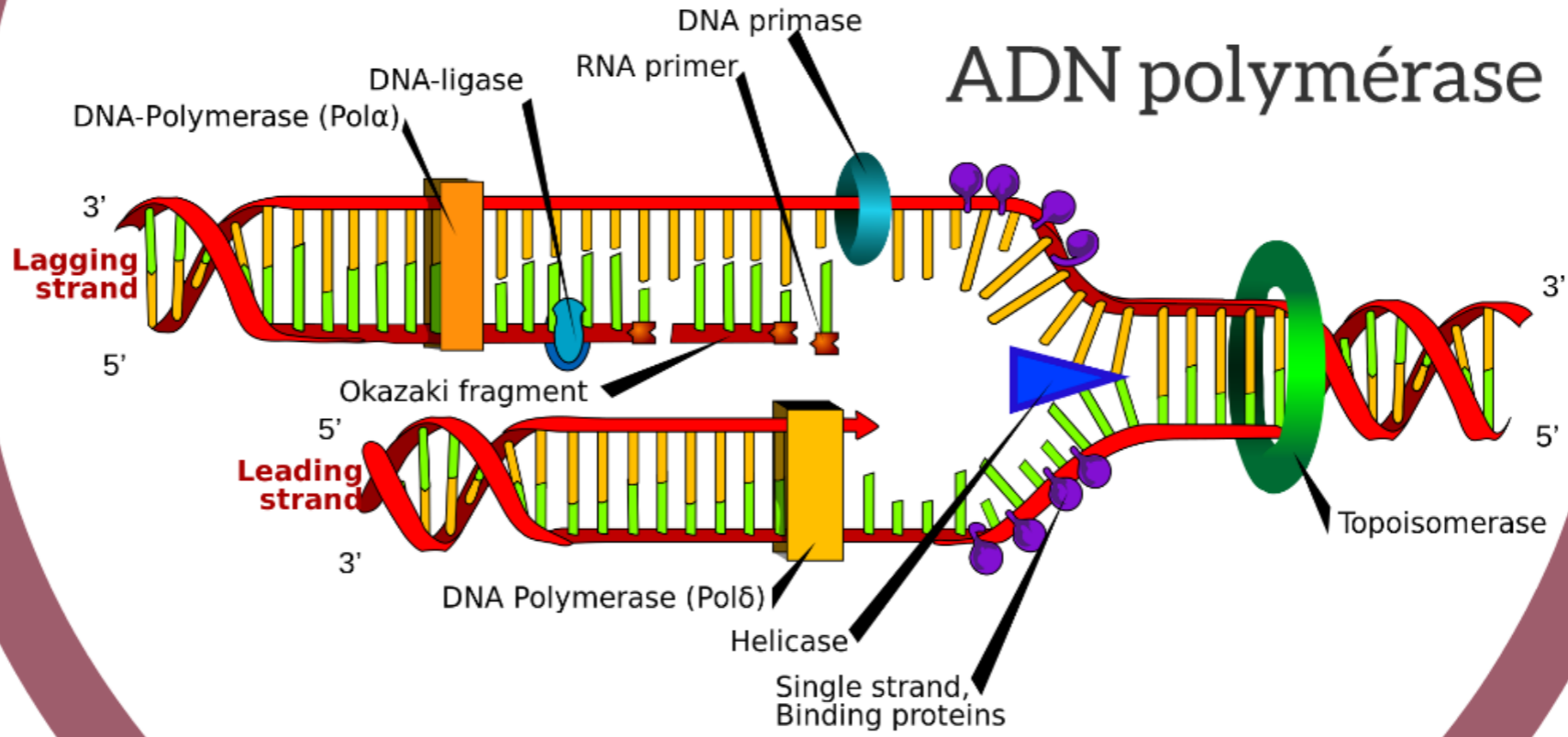
D'où vient le signal ?

Le matériel génétique est transmis à la descendance. C'est la base de l'hérédité.



L'ADN est répliqué en phase S du cycle cellulaire.

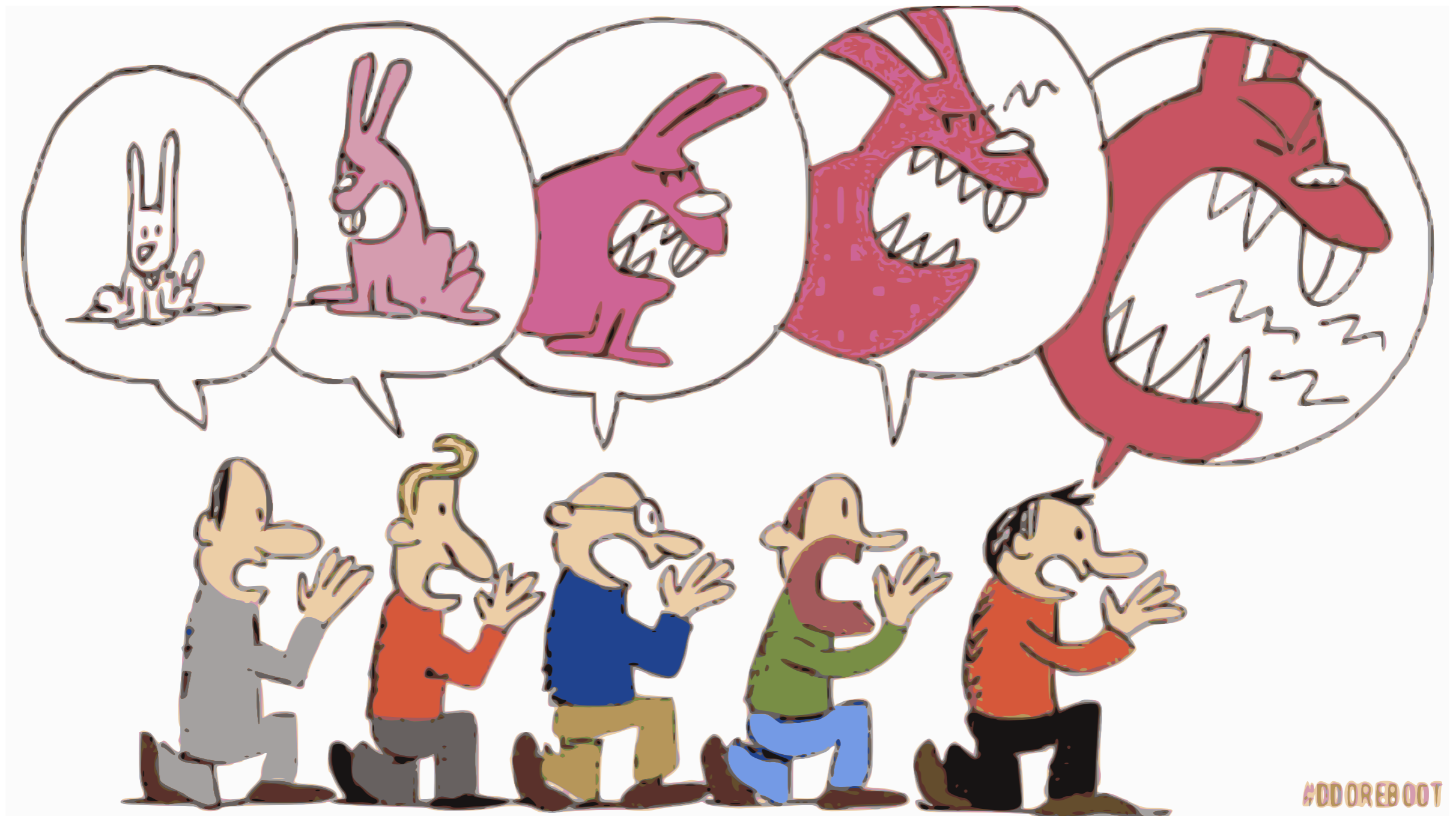
D'où vient le signal ?



1 erreur par milliard
de pb copiées

Phylogénie moléculaire

L'accumulation des erreurs forme le signal phylogénétique.

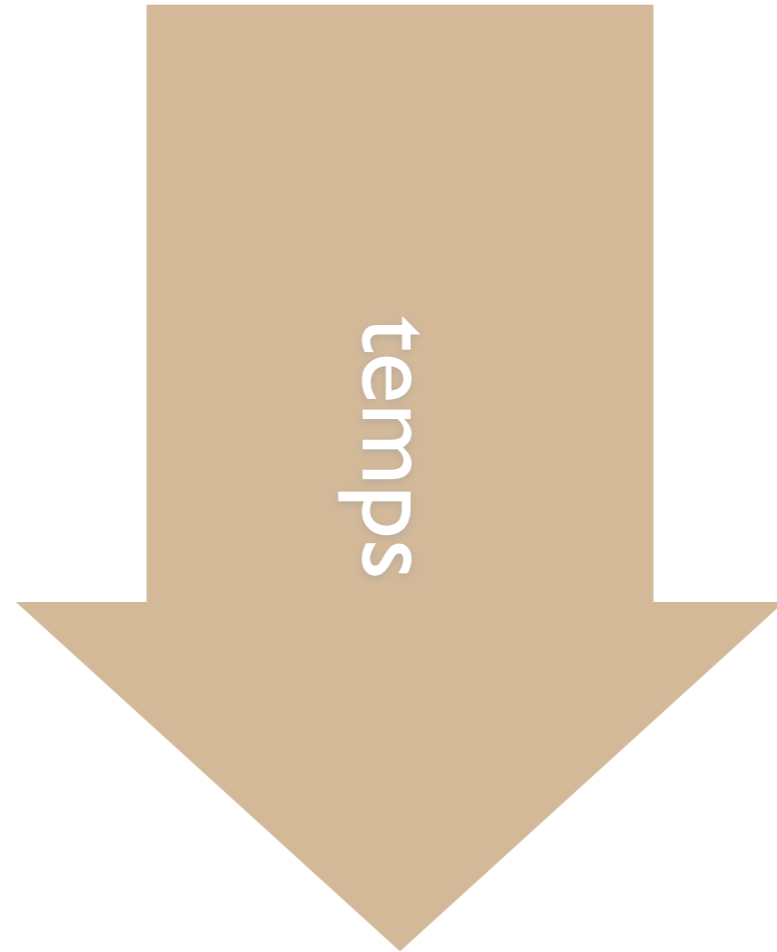


Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.

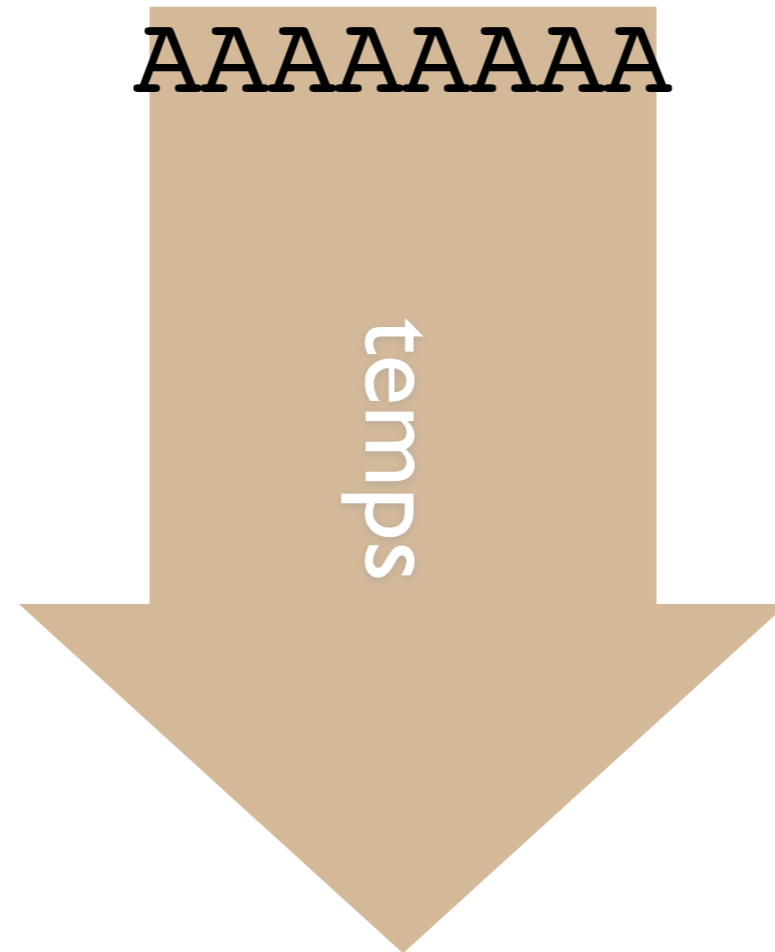
Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.



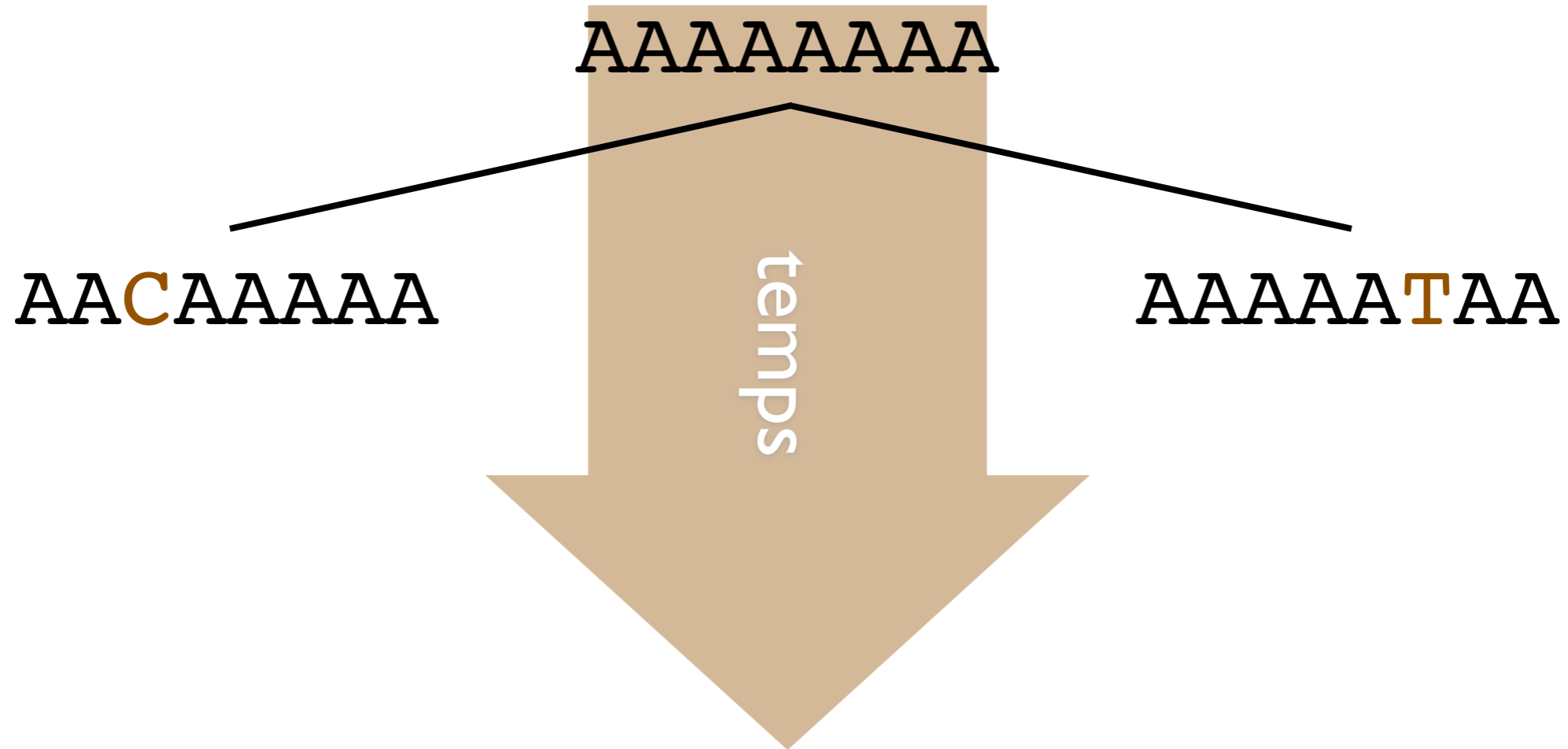
Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.



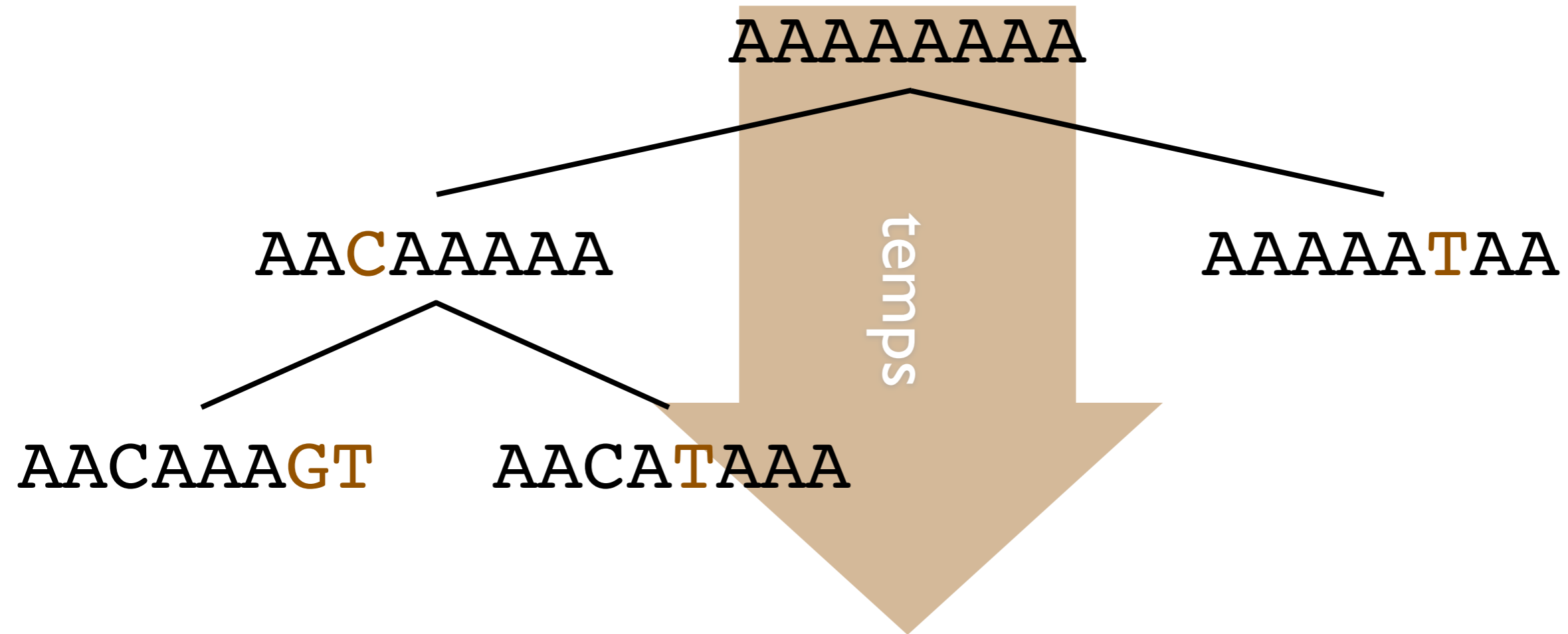
Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.



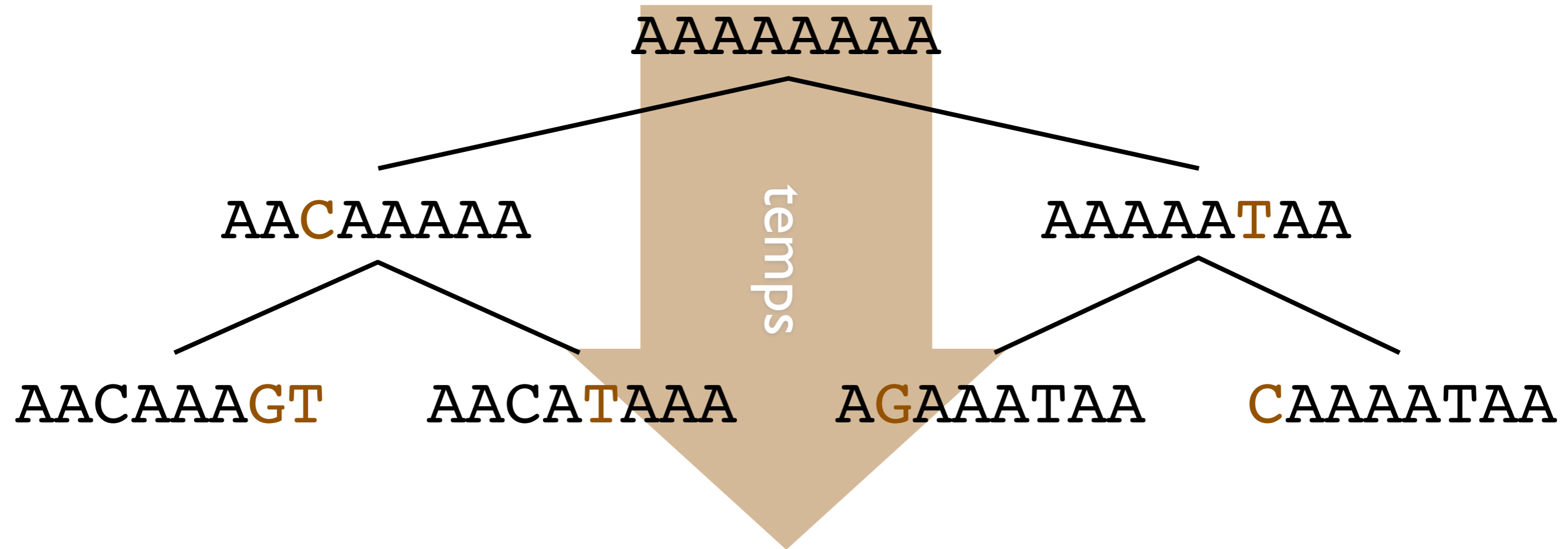
Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.



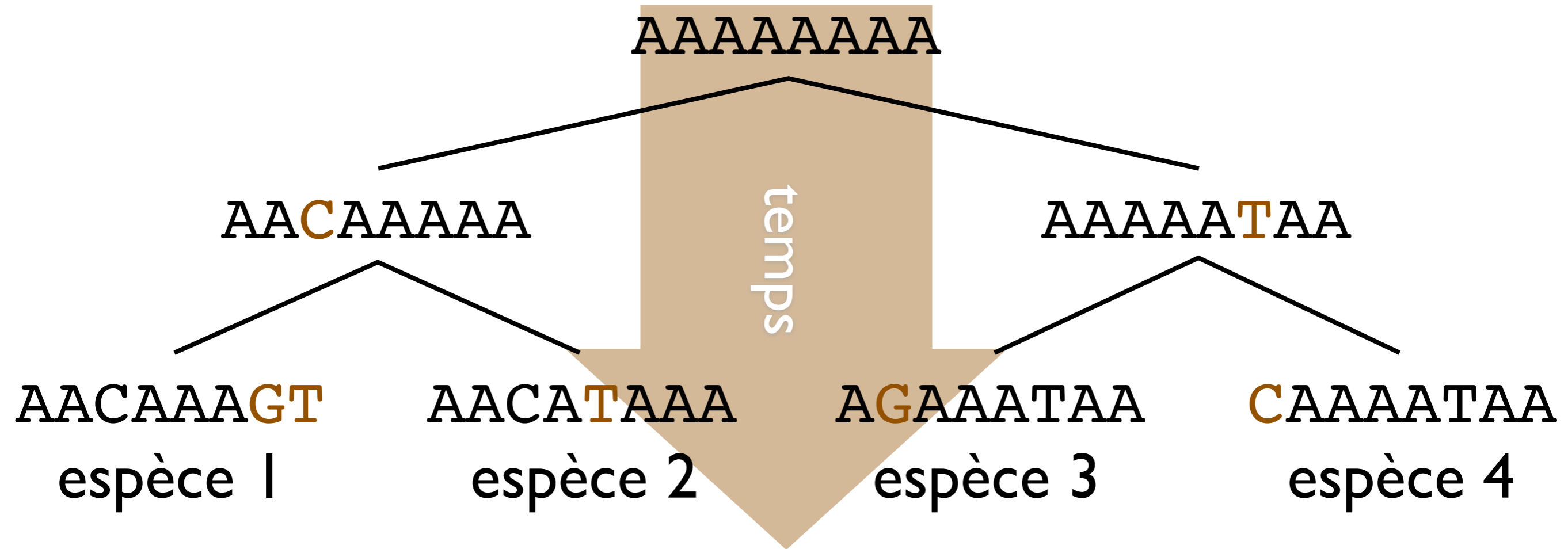
Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.



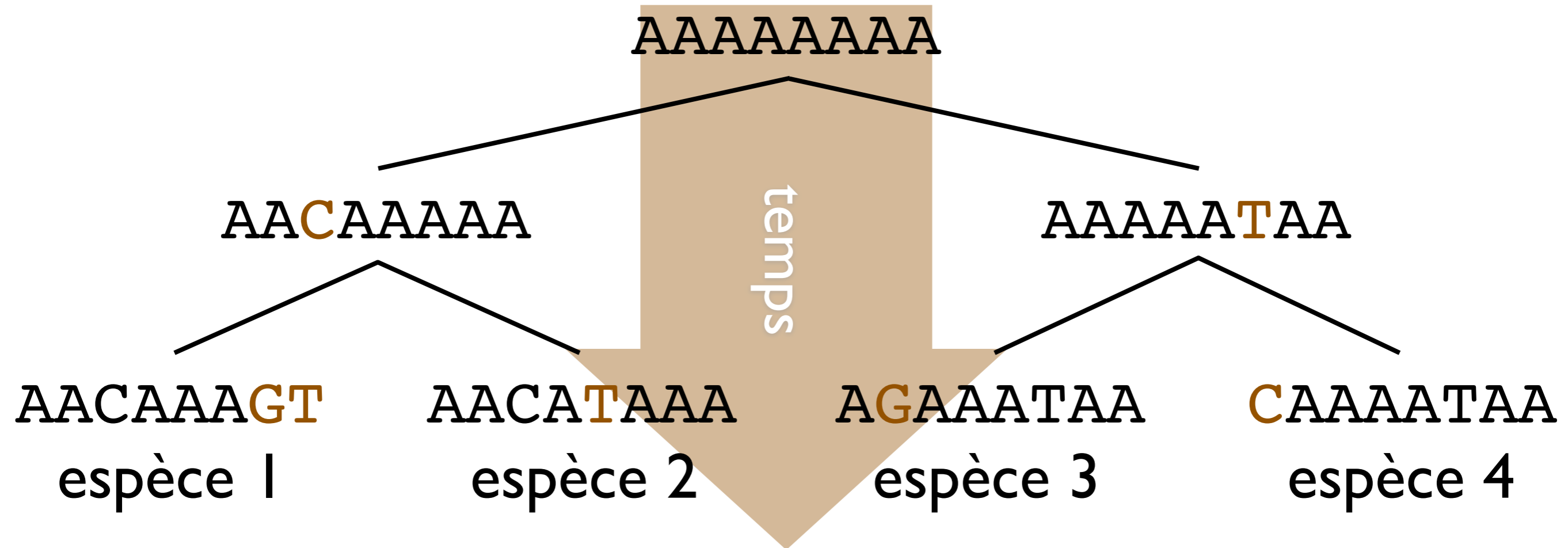
Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.



Phylogénie moléculaire

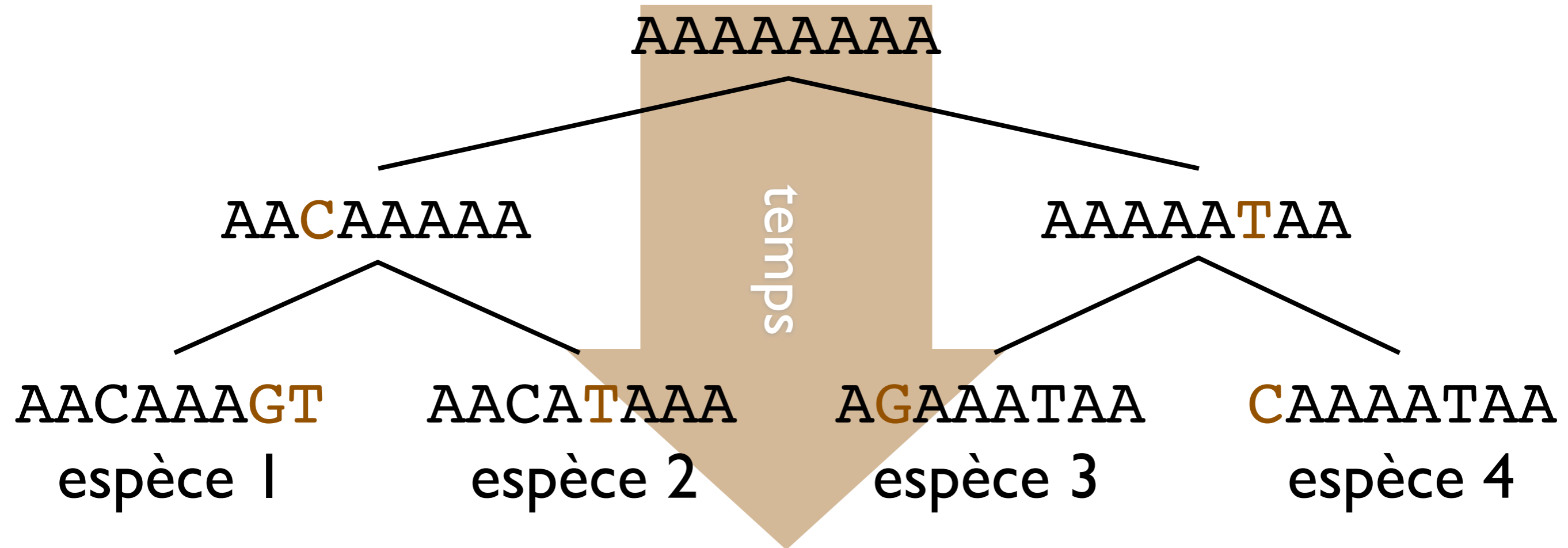
On peut reconstituer l'arbre en suivant la piste des erreurs.



- 1 AACAAAGT
- 2 AACATAAA
- 3 AGAAATAA
- 4 CAAAATAA

Phylogénie moléculaire

On peut reconstituer l'arbre en suivant la piste des erreurs.

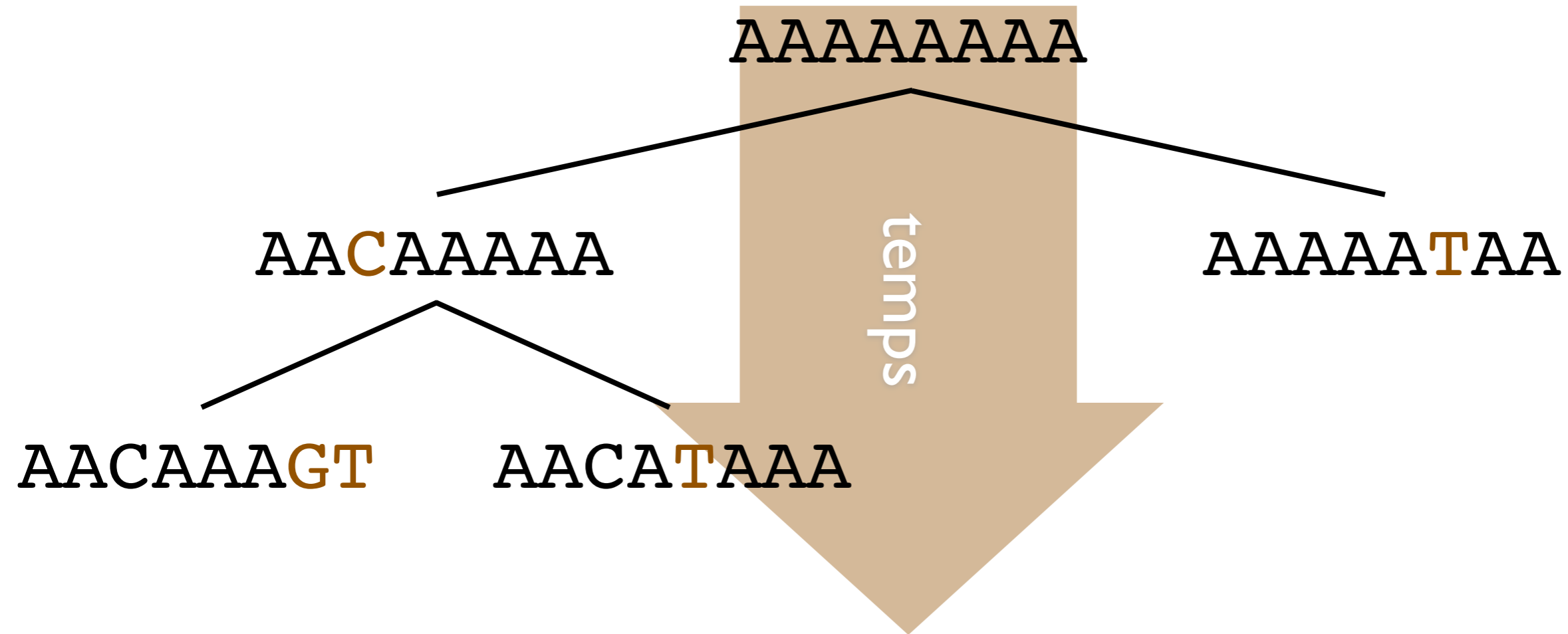


- 1 AACAAAAGT
- 2 AACATAAAA
- 3 AGAAAATAA
- 4 CAAAATAA

Les erreurs étant assez **rare**s, la majorité des différences de séquence entre espèces sont donc **héritées** des ancêtres communs.

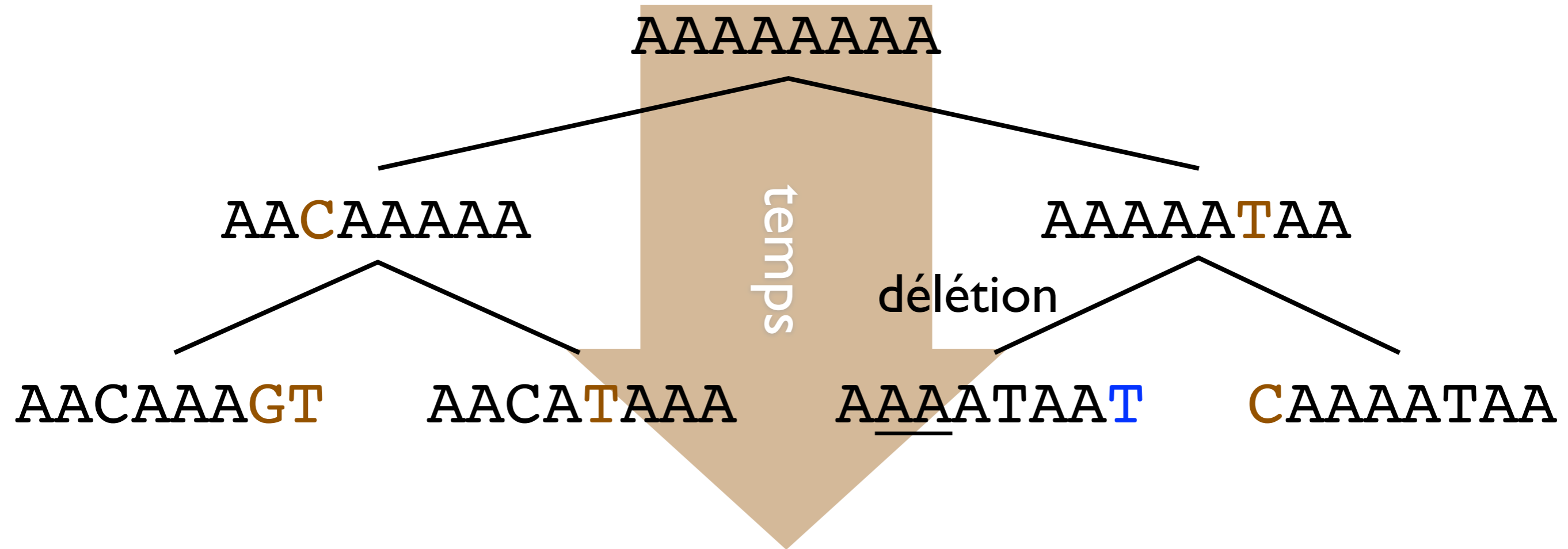
Phylogénie moléculaire

Il est impératif de comparer des positions homologues.



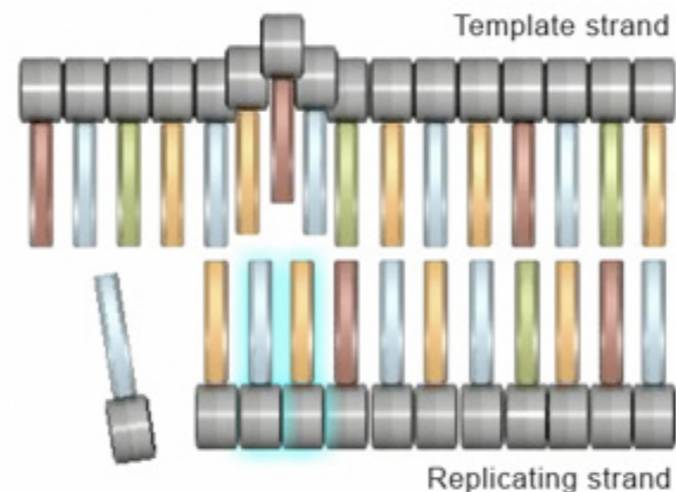
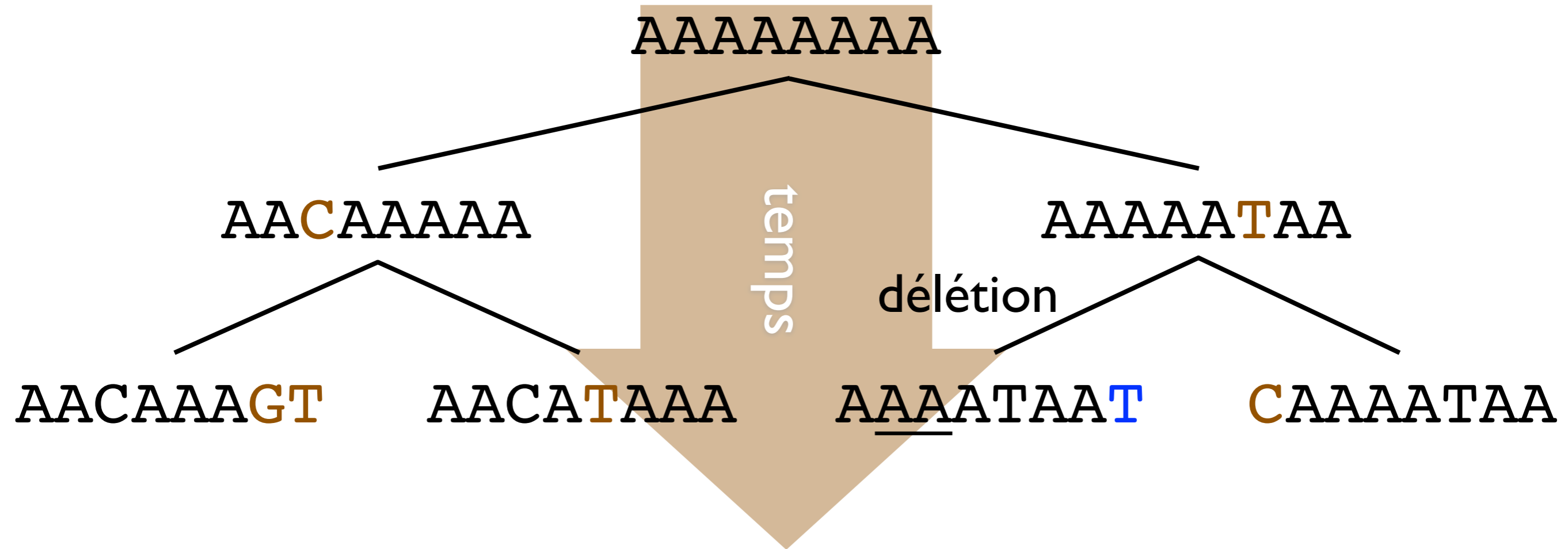
Phylogénie moléculaire

Il est impératif de comparer des positions homologues.



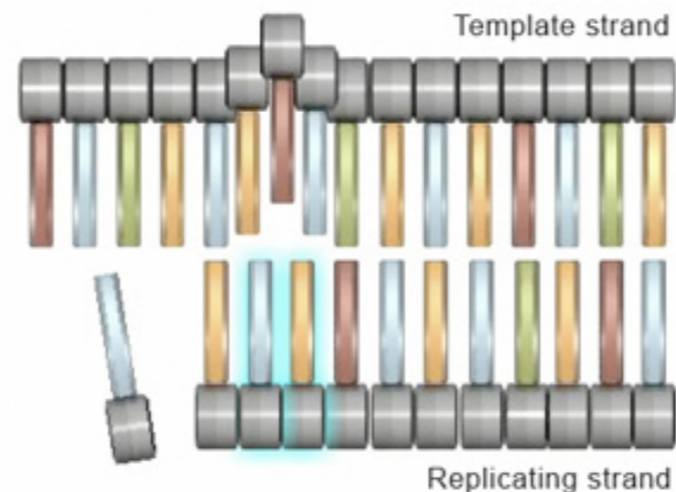
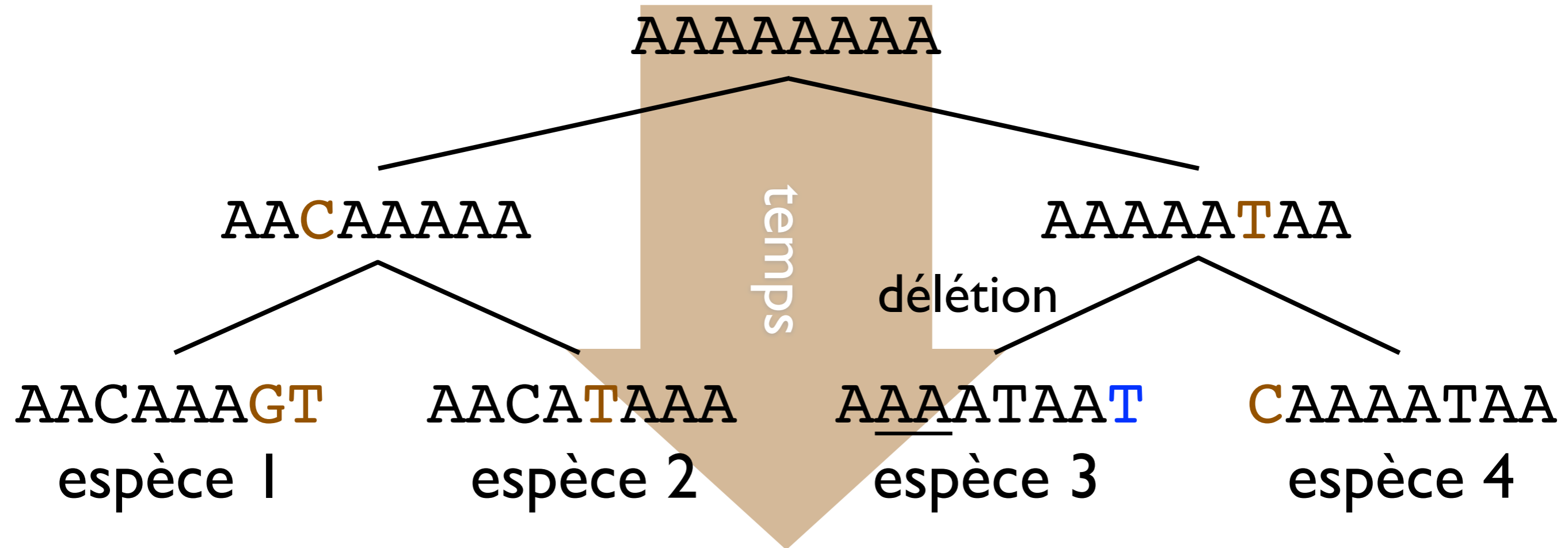
Phylogénie moléculaire

Il est impératif de comparer des positions homologues.



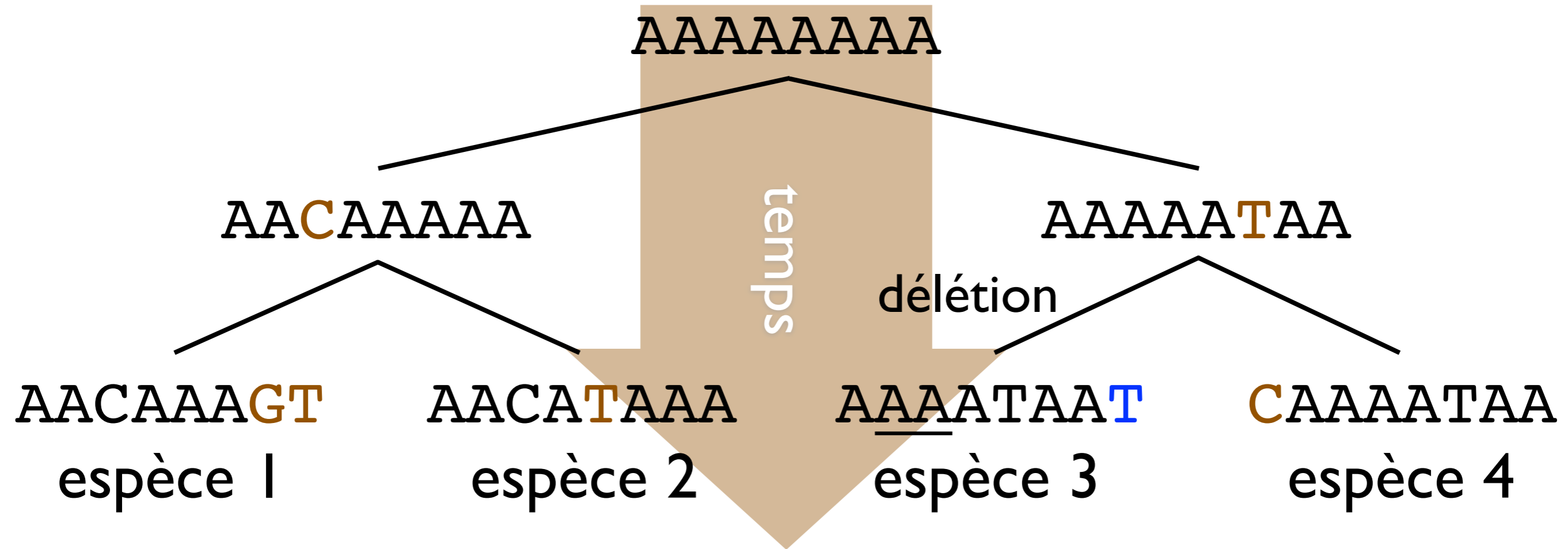
Phylogénie moléculaire

Il est impératif de comparer des positions homologues.

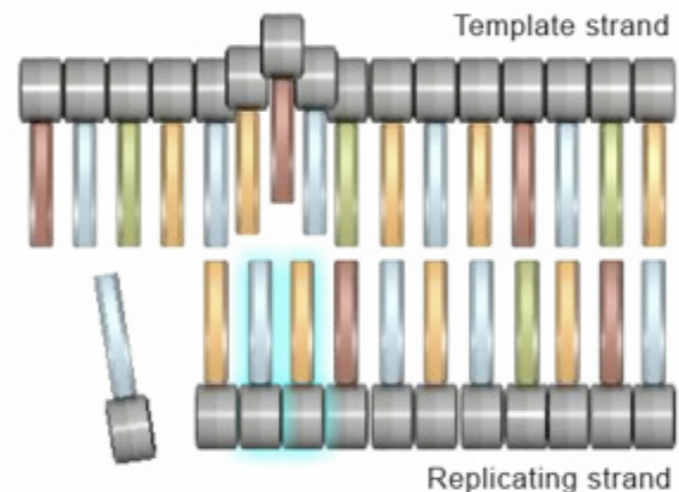


Phylogénie moléculaire

Il est impératif de comparer des positions homologues.

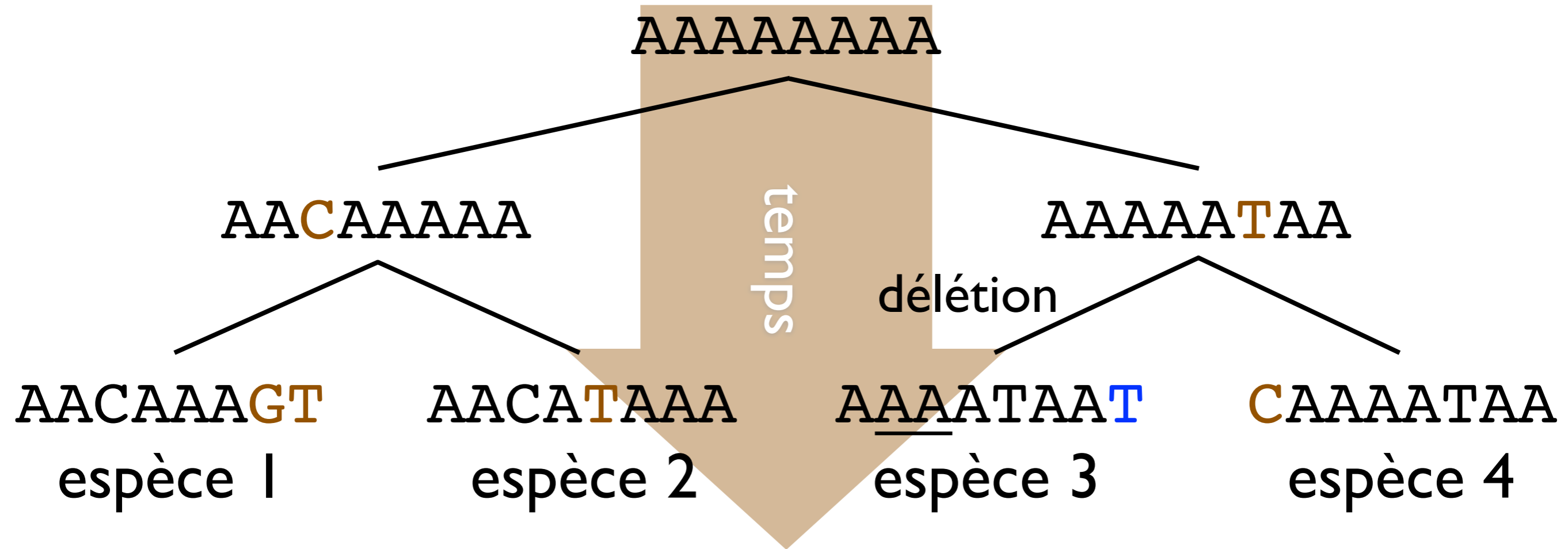


- 1 AACAAAGT
- 2 AACATAAA
- 3 AAAATAAT
- 4 CAAAATAA

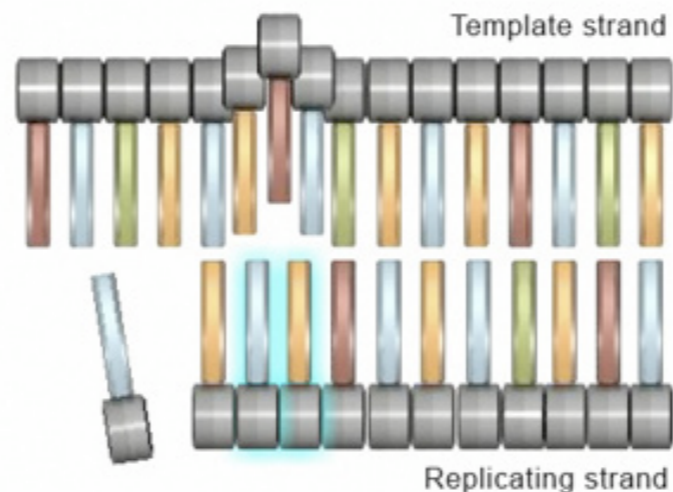


Phylogénie moléculaire

Il est impératif de comparer des positions homologues.



1 AAC**C**AAAG**T**
2 AAC**C**ATAAA
3 AAAA**T**AA**T**
4 CAAAA**T**AA



1 AAC**C**AAAG**T**
2 AAC**C**ATAAA
3 AA-**A**A**T**AA
4 CAAAA**T**AA

Exemple : cytochrome c

Exemple : cytochrome c

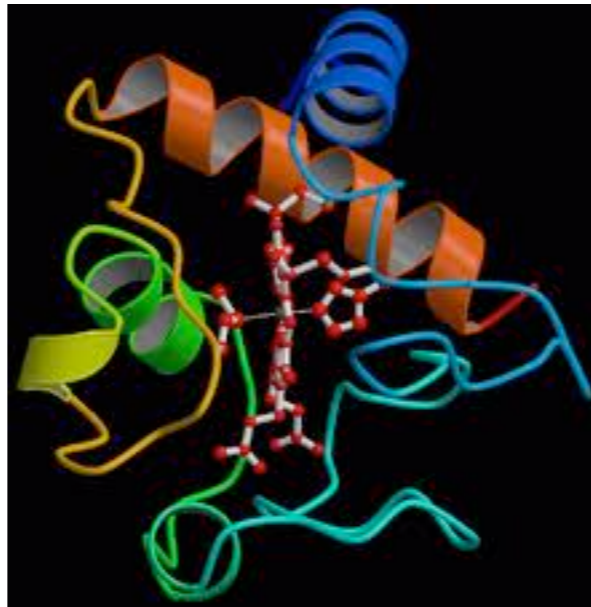
>cheval

MGDVEKGGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

Exemple : cytochrome c

>cheval

MGDVEKGGKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE



cytochrome c

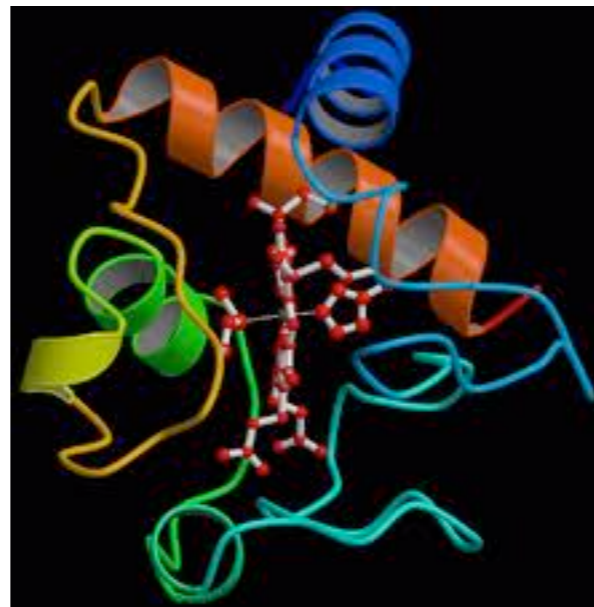
Exemple : cytochrome c

>cheval

```
MGDVEKGGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITW  
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE
```

>humain

```
MGDVEKGGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIW  
GEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE
```



cytochrome c

Exemple : cytochrome c

>cheval

MGDVEKGGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

>humain

MGDVEKGGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIW
GEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE

Exemple : cytochrome c

>cheval

MGDVEKGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKATNE

>humain

MGDVEKGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIW
GEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKATNE

>poisson

MGDVEKGKKVVFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGFSYTDANKSKGIVW
GEDTLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKSATS

Exemple : cytochrome c

>cheval

MGDVEKGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKATNE

>humain

MGDVEKGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIW
GEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKATNE

>poisson

MGDVEKGKKVVFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGFSYTDANKSKGIW
GEDTLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKSATS

>plante (arabette des dames)

MQVADISLQGDAAKKGANLFKTRCAQCHTLKAGEGNKIGPELHGLFGRKTGSVAGYSYTD
NKQKGI EWKDDTLFEYLENPKKYIPGTKMAFGGLKKPKDRNDLITFLEETK

Exemple : cytochrome c

>cheval

-----MGDVEKGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDA
NKNKGITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

>humain

-----MGDVEKGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAA
NKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE

>poisson

-----MGDVEKGKKVVFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGFSYTDA
NKSKGIVWGEDTLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKSATS

>plante (arabette des dames)

MQVADISLQGDAAKKGANLFFKTRCAQCHTLKAGEGNKIGPELHGLFGRKTGSVAGYSYTDA
NKQKGI EWKDDTLFEYLENPKKYIPGTKMAFGGLKKPKDRNDLITFLEEETK

Exemple : cytochrome c

>cheval

-----MGDVEKGGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDA
NKNKGITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

>humain

-----MGDVEKGGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAA
NKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE

>poisson

-----MGDVEKGGKKVVFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGFSYTDA
NKSKGIVWGEDTLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKSATS

>plante (arabette des dames)

MQVADISLQGDAAKKGANLFKTRCAQCHTLKAGEGNKIGPELHGLFGRKTGSVAGYSYTDA
NKQKGIEWKDDTLFEYLENPKKYIPGTKMAFGGLKKPKDRNDLITFLEEETK

>microbe (amibe)

MSDIIARGNVENGDKLFFKARCAQCHTTANGAPNKQGPPLYGLFFPKSRSFPGYAYSDPNK
NTGKFCIMWGEQTLFDYLENPKKYIPKTKMAFAGFKSEQDRADVVAYLEQSTK

Exemple : cytochrome c

>cheval

-----MGDVEKGKKIFVQKCAQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGFSYTDA
NKNKG---ITWKEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE

>humain

-----MGDVEKGKKIFIMKCSQCHTVEKGGKHKKTGPNLHGLFGRKTGQAPGYSYTAA
NKNKG---IIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKATNE

>poisson

-----MGDVEKGKKVVFVQKCAQCHTVENGGKHKVGPNLWGLFGRKTGQAEGFSYTDA
NKSkg---IVWGEDTLMEYLENPKKYIPGTKMIFAGIKKKGERADLIAYLKSATS

>plante (arabette des dames)

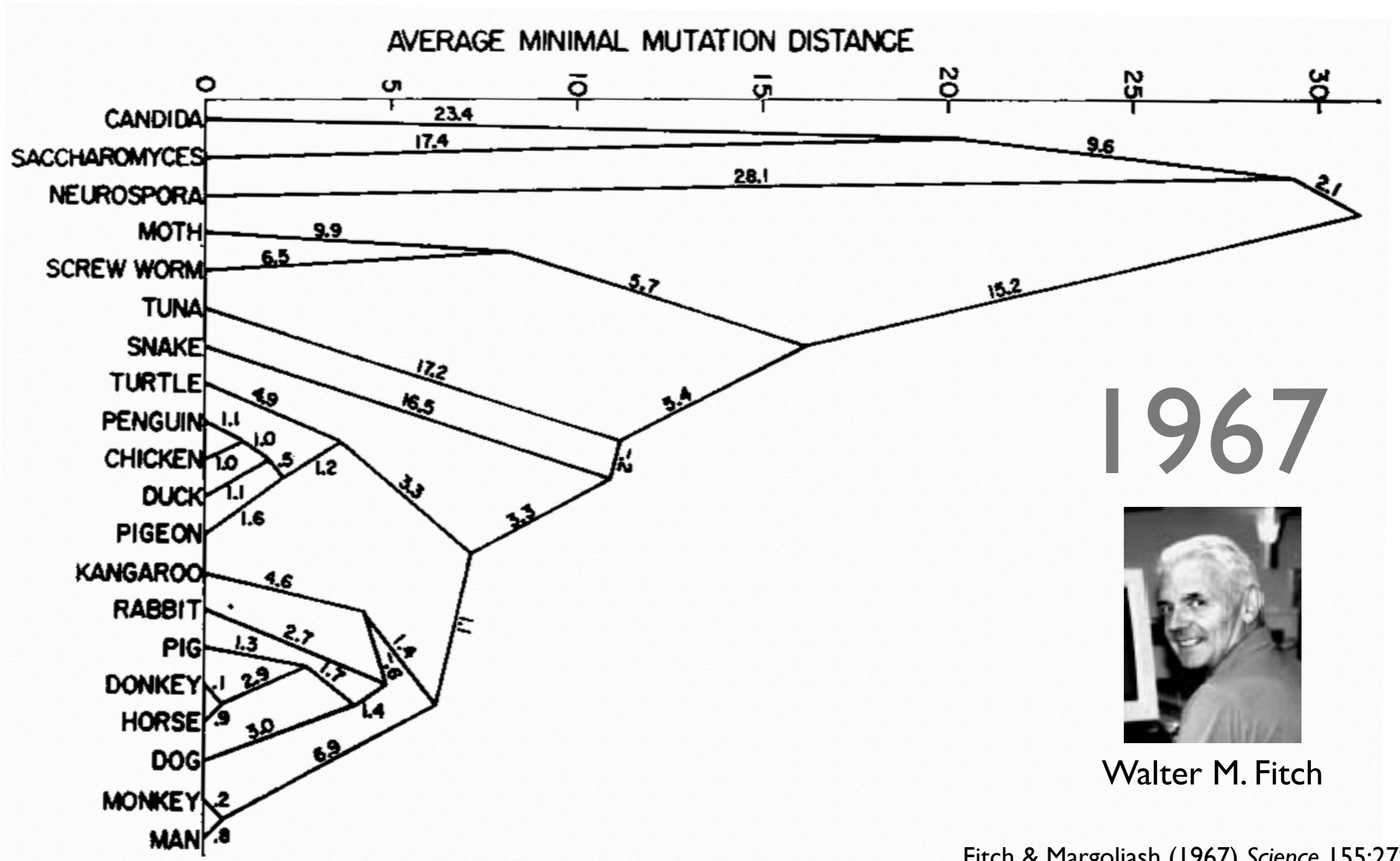
MQVADISLQGDAAKKGANLFFKTRCAQCHTLKAGEGNKIGPELHGLFGRKTGSVAGYSYTDA
NKQKG---IEWKDDTLFEYLENPKKYIPGTKMAFGGLKKPKDRNDLITFLEEETK

>microbe (amibe)

--MSDIIARGNVENGDKLFFKARCAQCHTTANGAPNKQGPPLYGLFFPKSRSFPGYAYSDF
NKNTGKFCIMWGEQTLFDYLENPKKYIPKTKMAFAGFKSEQDRADVWAYLEQSTK

Exemple : cytochrome c

La comparaison des séquences alignées produit un arbre.



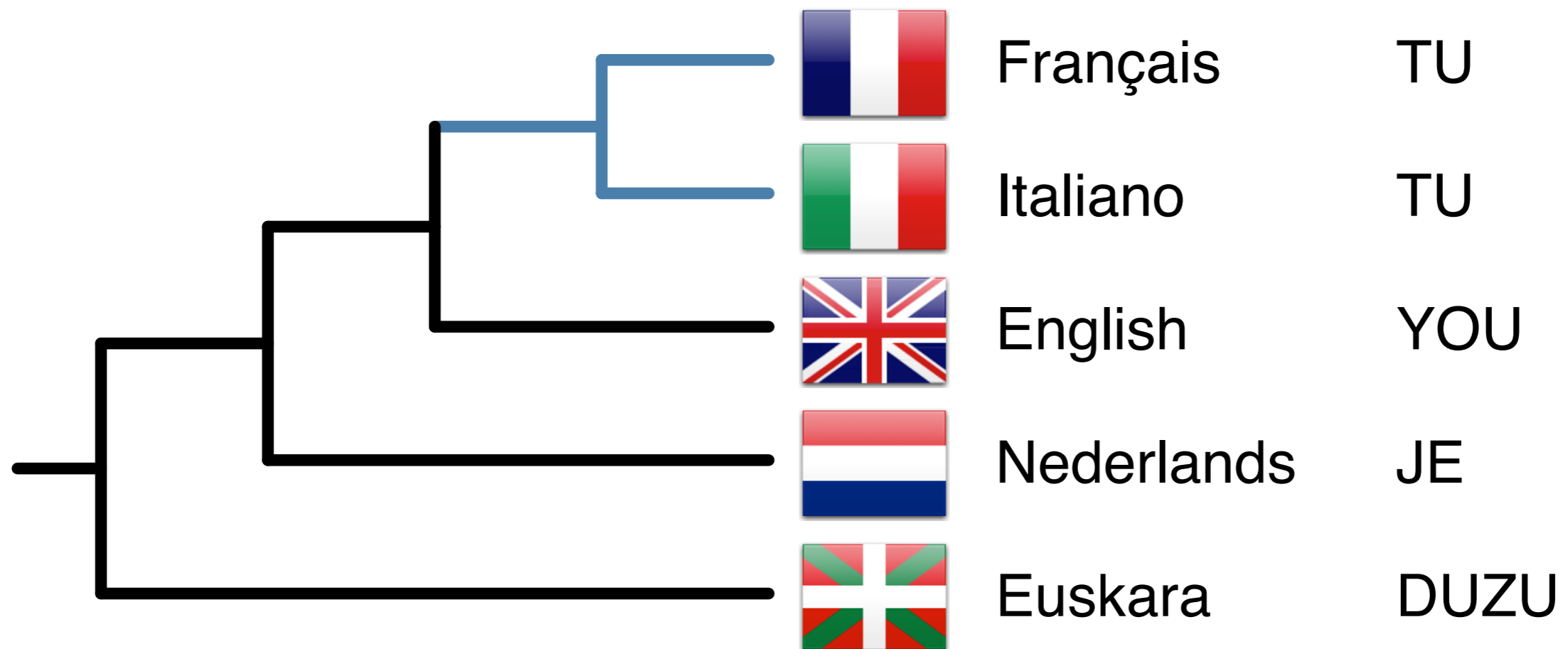
Walter M. Fitch

Phylogénomique

un exemple tiré de la linguistique

Potential relationships among European languages

based on the naive analysis of 1 word

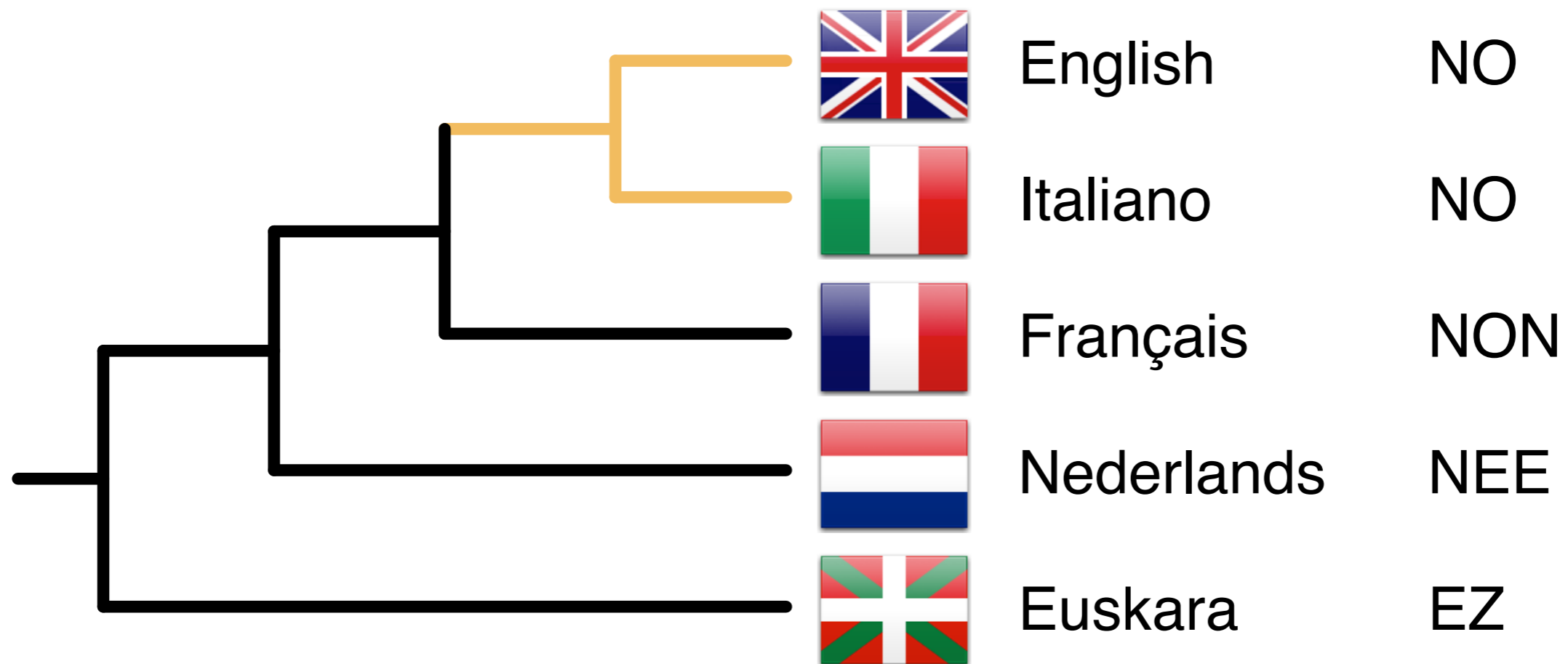


Phylogénomique

un exemple tiré de la linguistique

Potential relationships among European languages

based on the naive analysis of 1 word



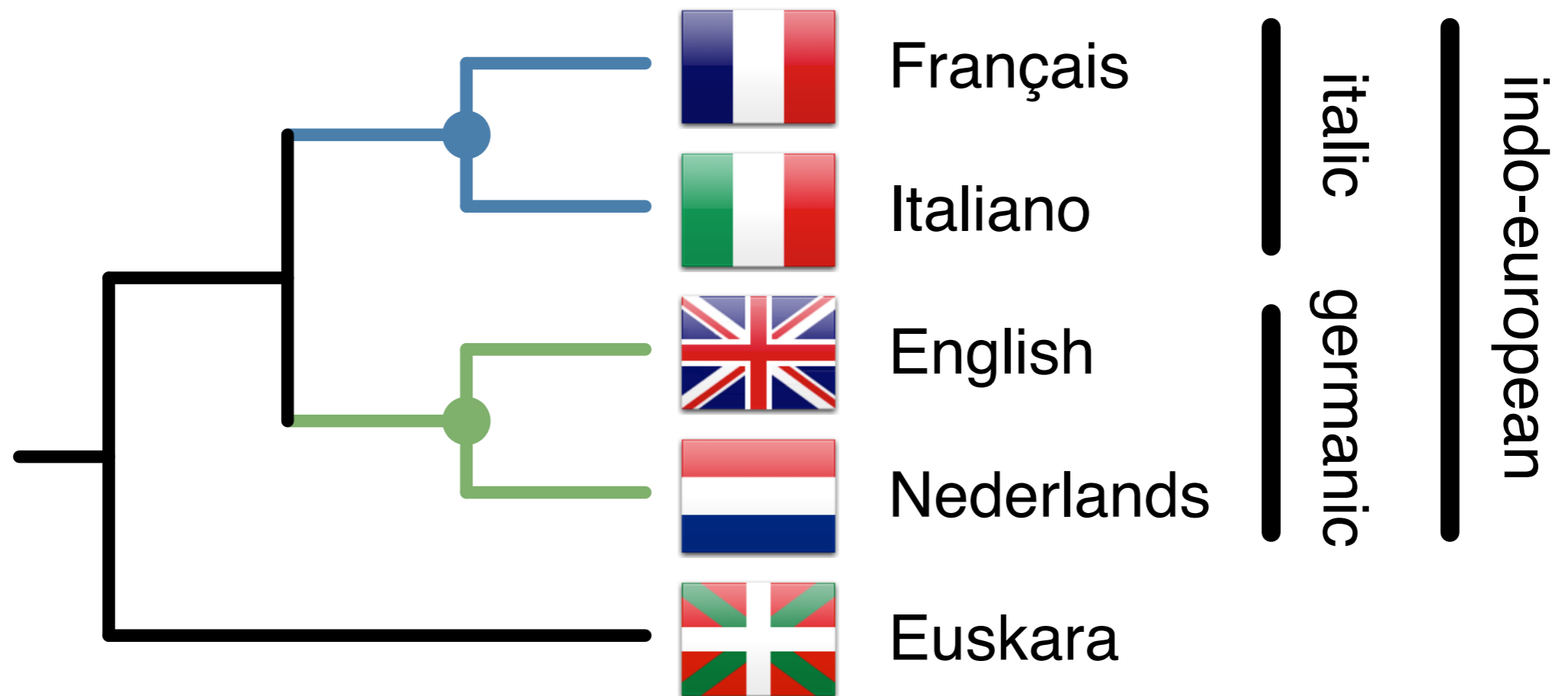
	français	italiano	english	nederlands	euskara
1	un	uno	one	een	bat
2	deux	due	two	twee	bi
3	trois	tre	three	drie	hiru
4	je	io	I	ik	ni
5	tu	tu	you	je	duzu
6	qui ?	chi?	who?	wie?	nor?
7	oui	si	yes	ja	bai
8	non	no	no	nee	ez
9	mère	madre	mother	moeder	ama
10	père	padre	father	vader	aita
11	dent	dente	tooth	tand	hortz
12	coeur	cuore	heart	hart	bihotza
13	pied	piede	foot	voet	oinez
14	souris	topolino	mouse	muis	saguaren

Phylogénomique

un exemple tiré de la linguistique

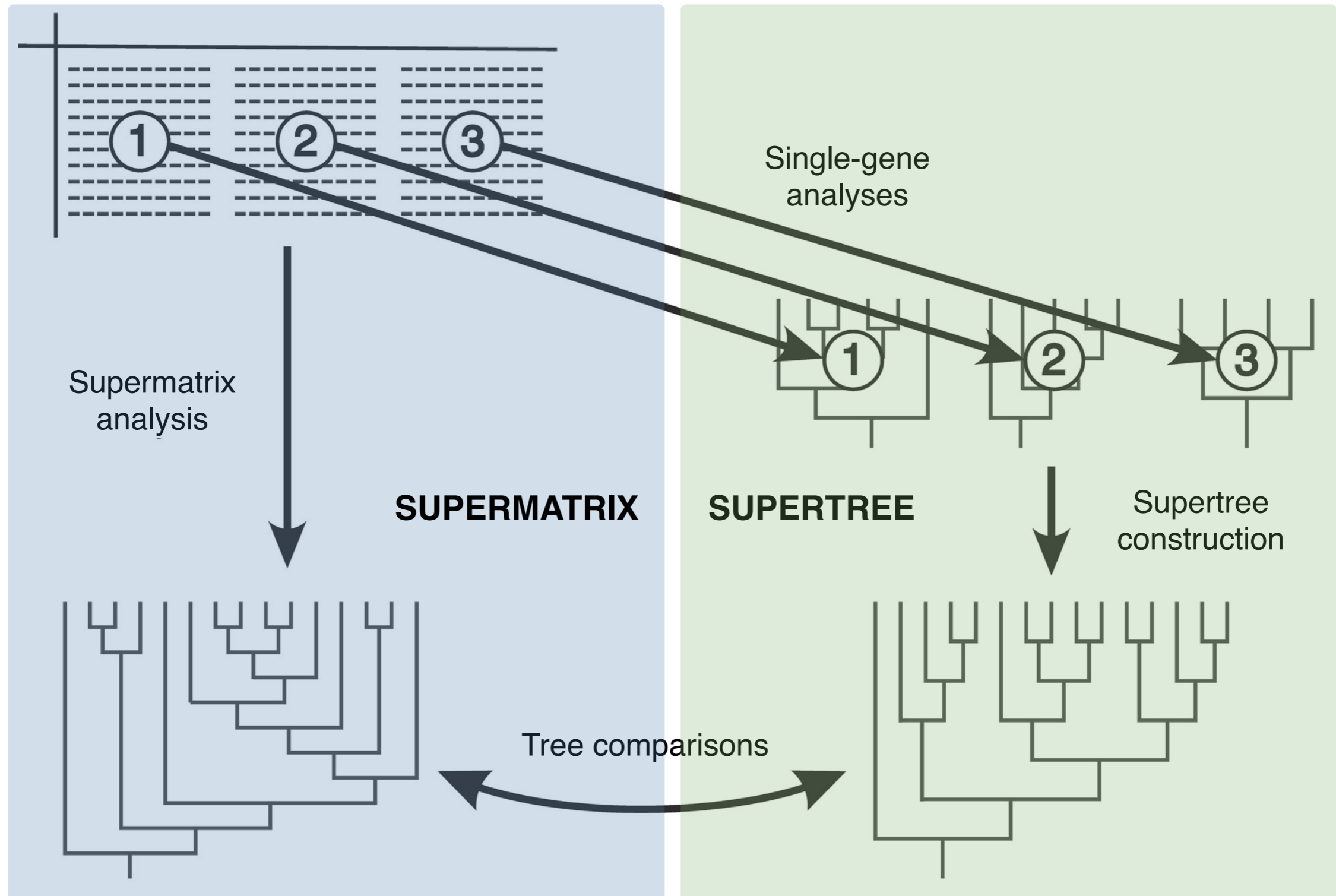
Known relationships among European languages

strongly supported by the naive analysis of 14 words



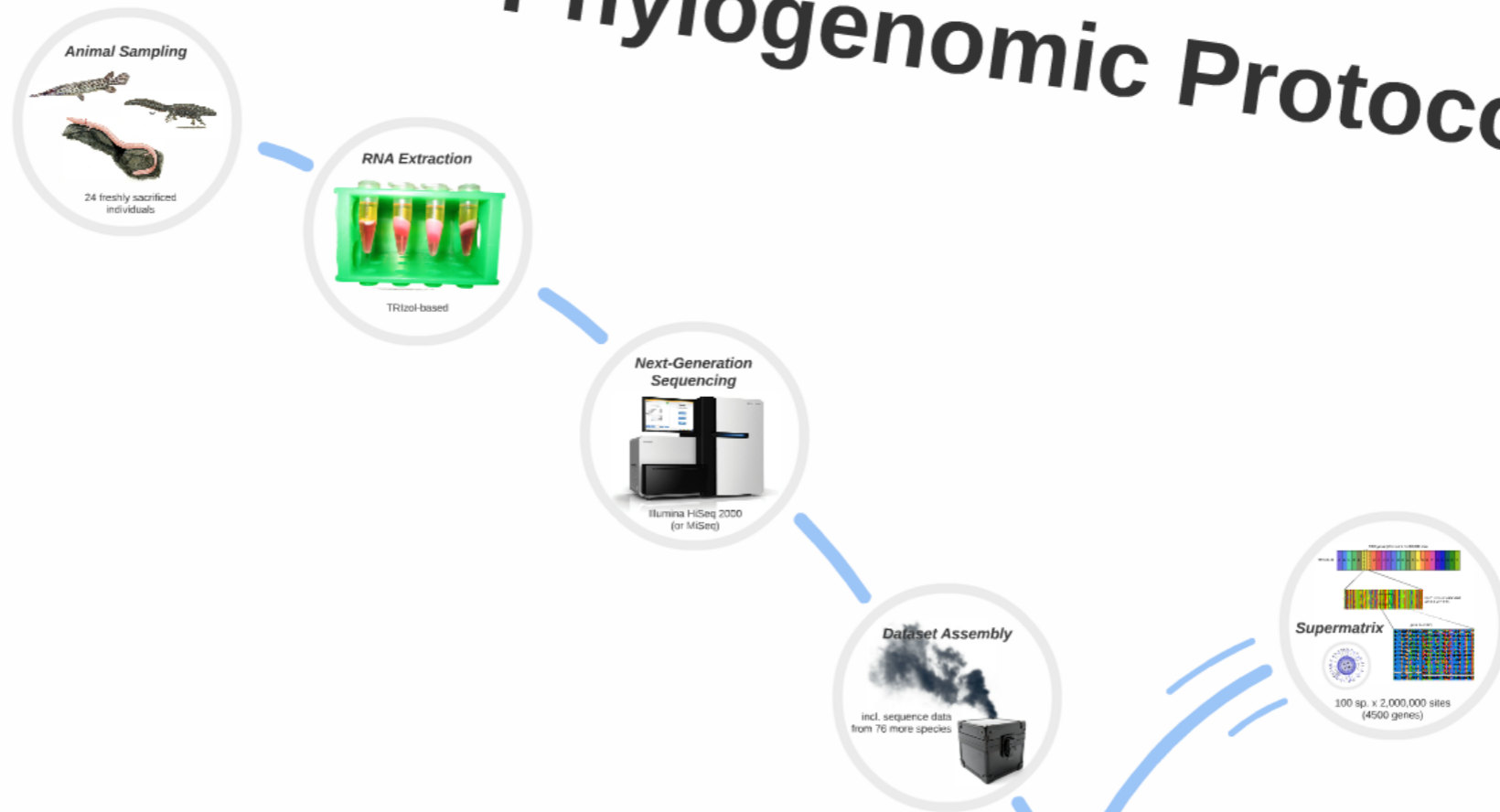
Phylogénomique

application aux gènes des génomes à comparer



Arbre des Vertébrés

Phylogenomic Protocol



Phylogenomics is the use of genome-scale data to reconstruct evolutionary relationships.

Arbre des Vertébrés

Animal Sampling

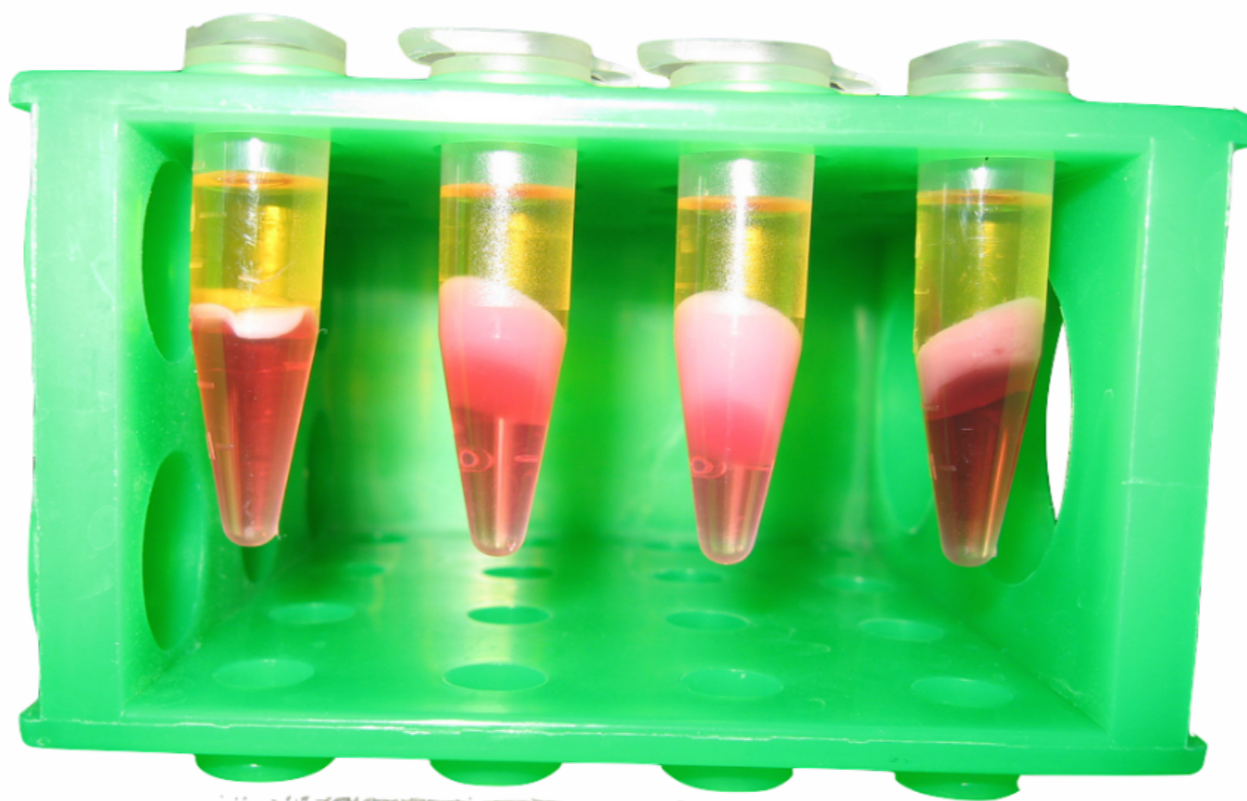


24 freshly sacrificed
individuals



Arbre des Vertébrés

RNA Extraction



TRIzol-based

Arbre des Vertébrés

Next-Generation Sequencing



Illumina HiSeq 2000
(or MiSeq)

Arbre des Vertébrés

Dataset Assembly

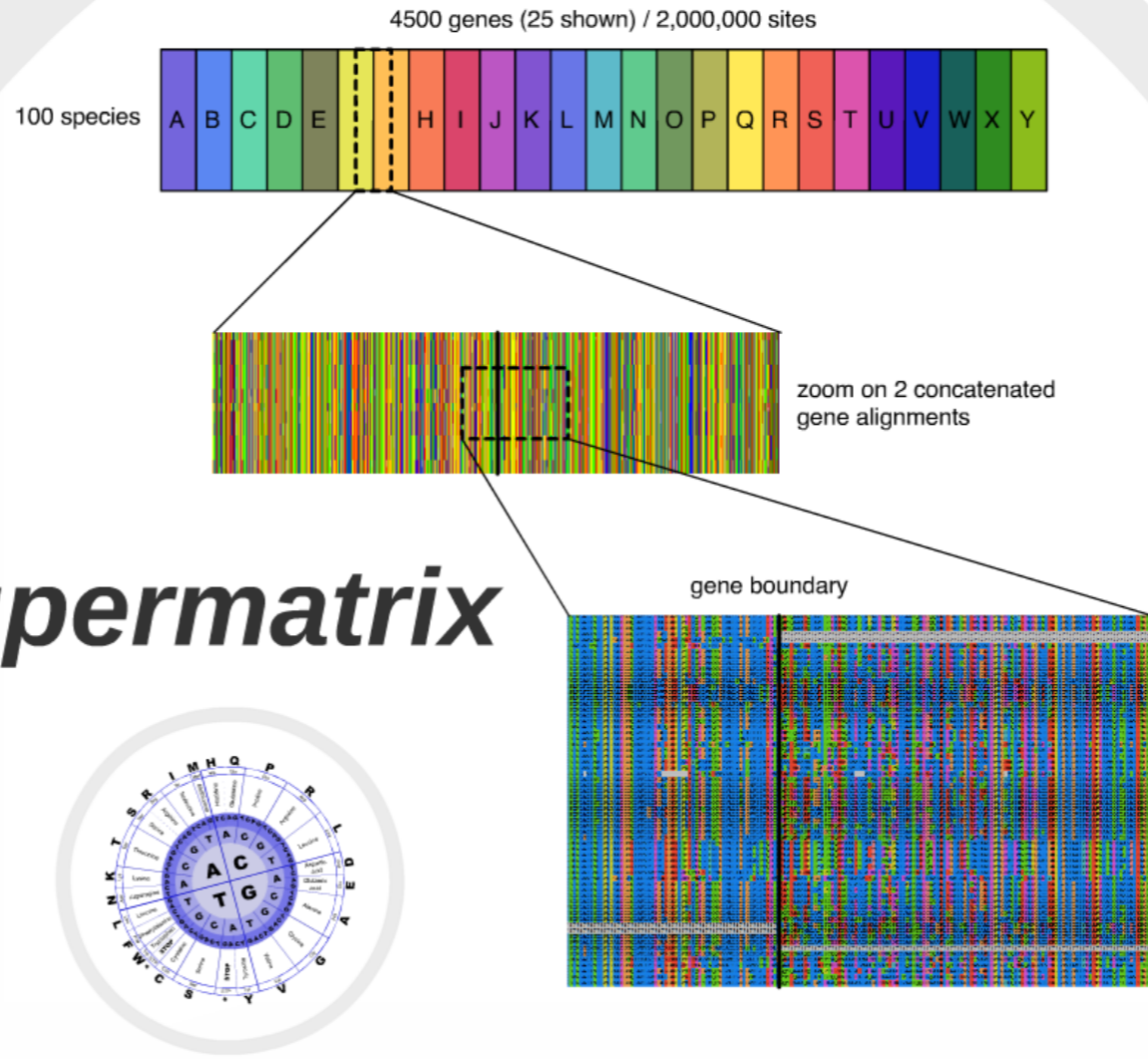
incl. sequence data
from 76 more species



Arbre des Vertébrés

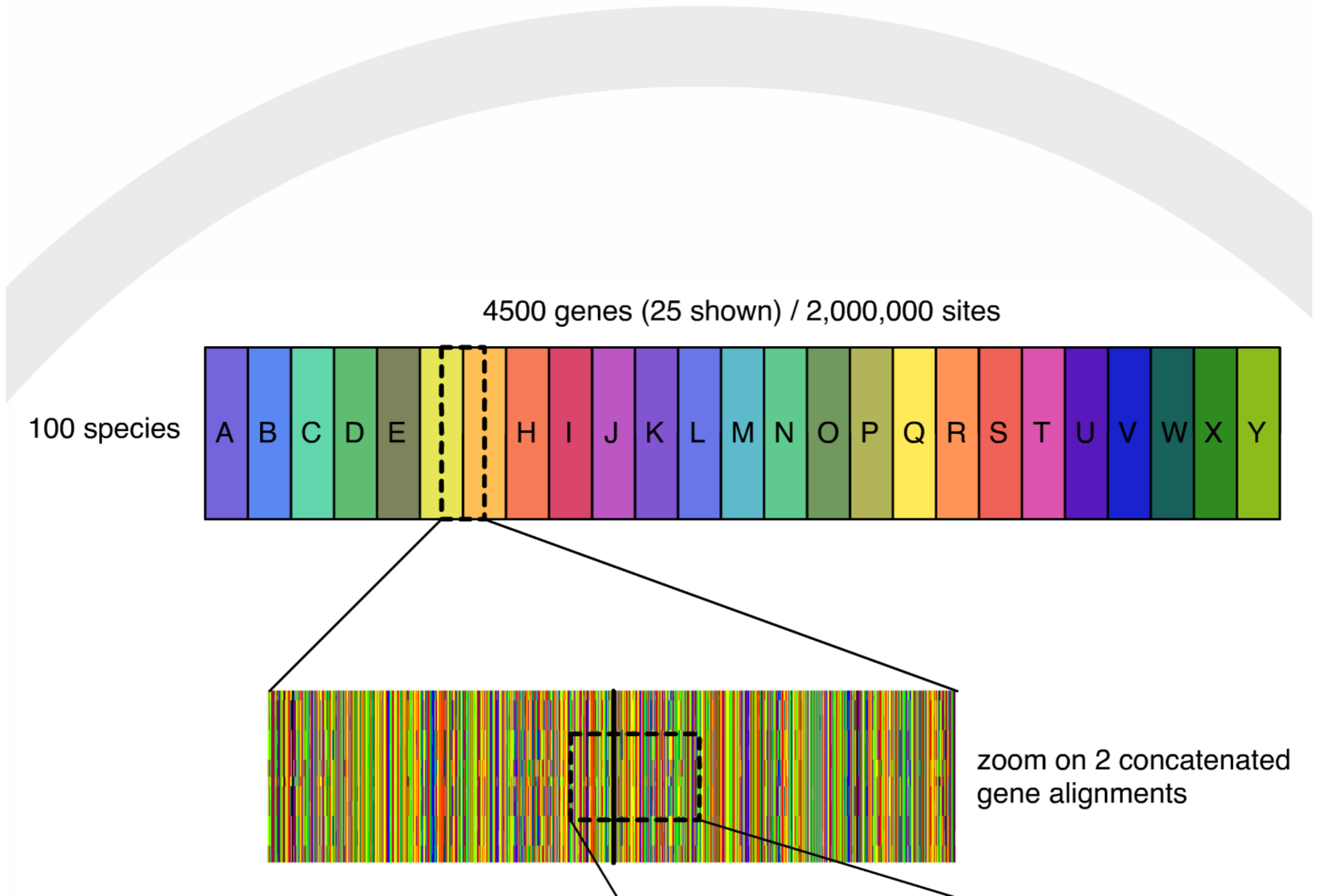


Supermatrix

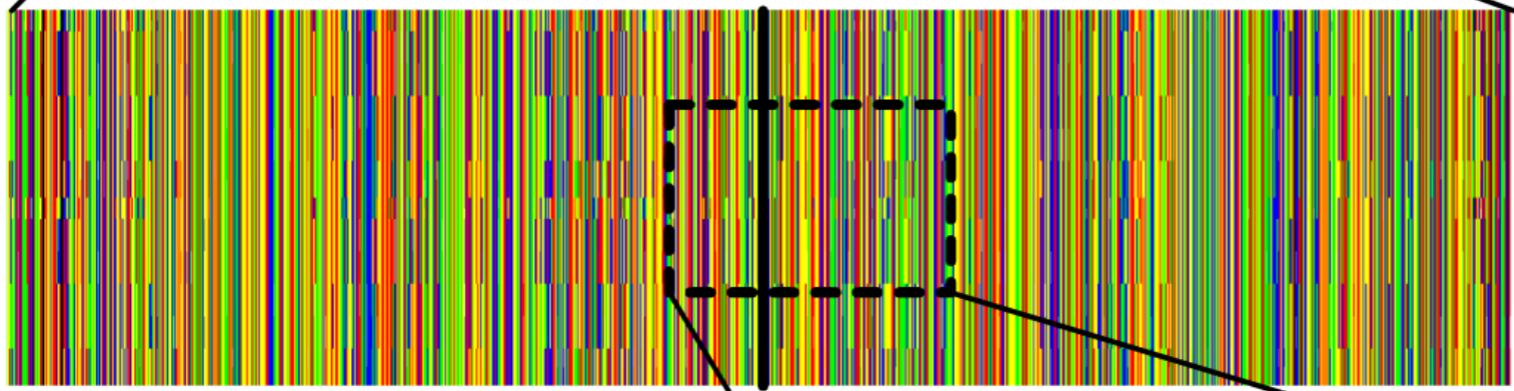
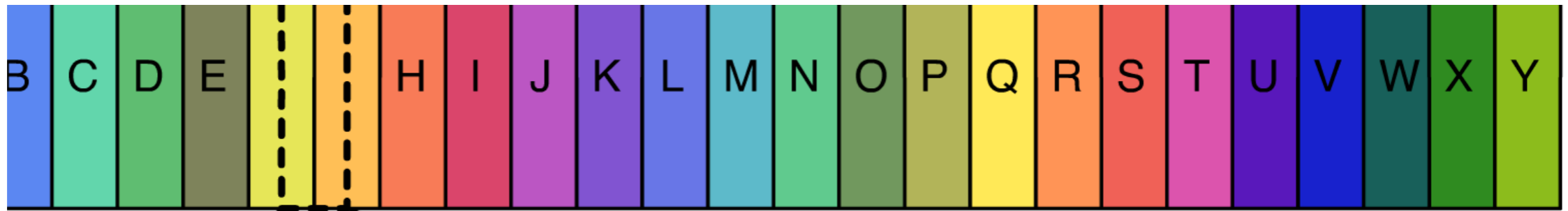


100 sp. x 2,000,000 sites
(4500 genes)

Arbre des Vertébrés



Arbre des Vertébrés



zoom on 2 concatenated gene alignments

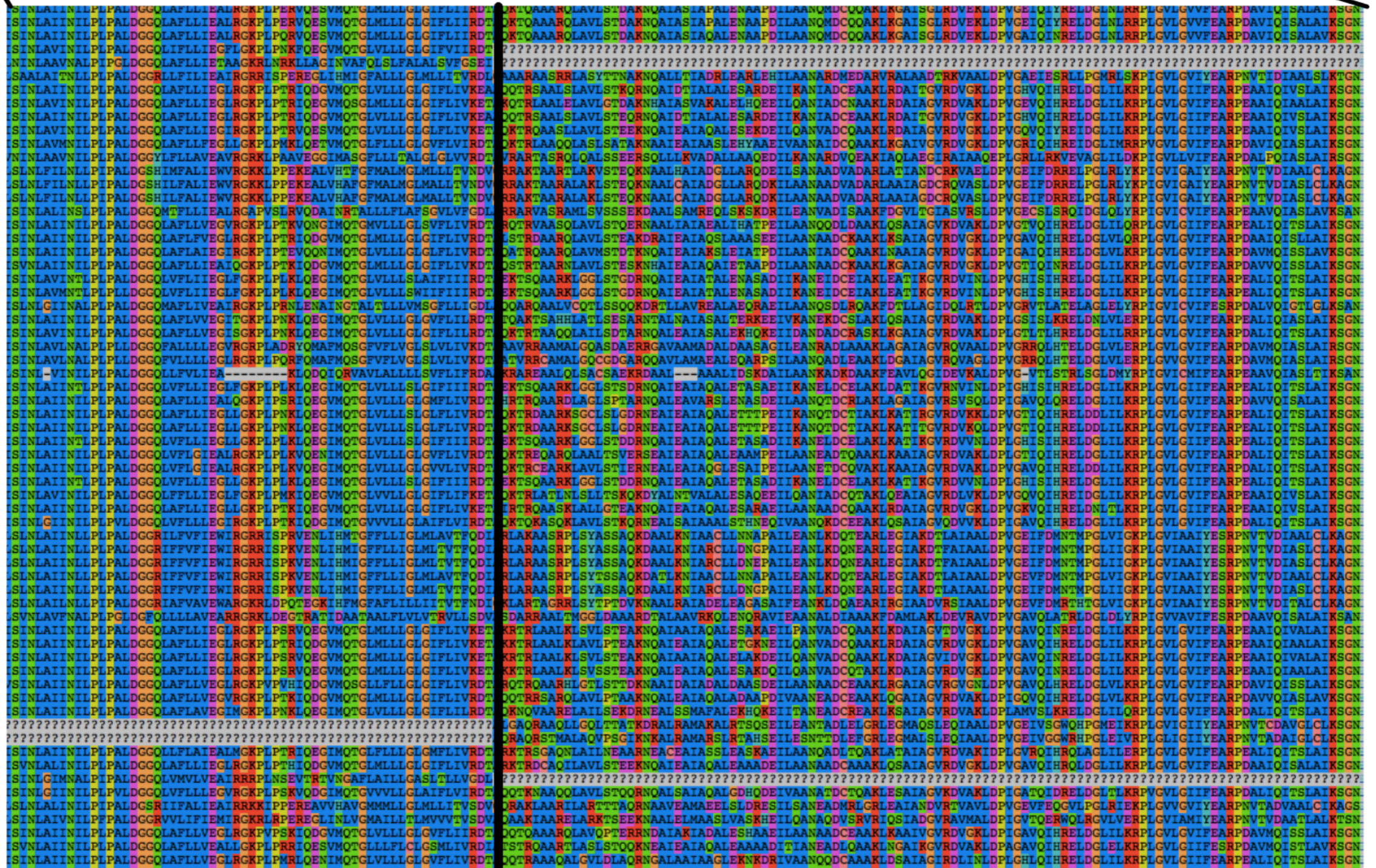
matrix

gene boundary

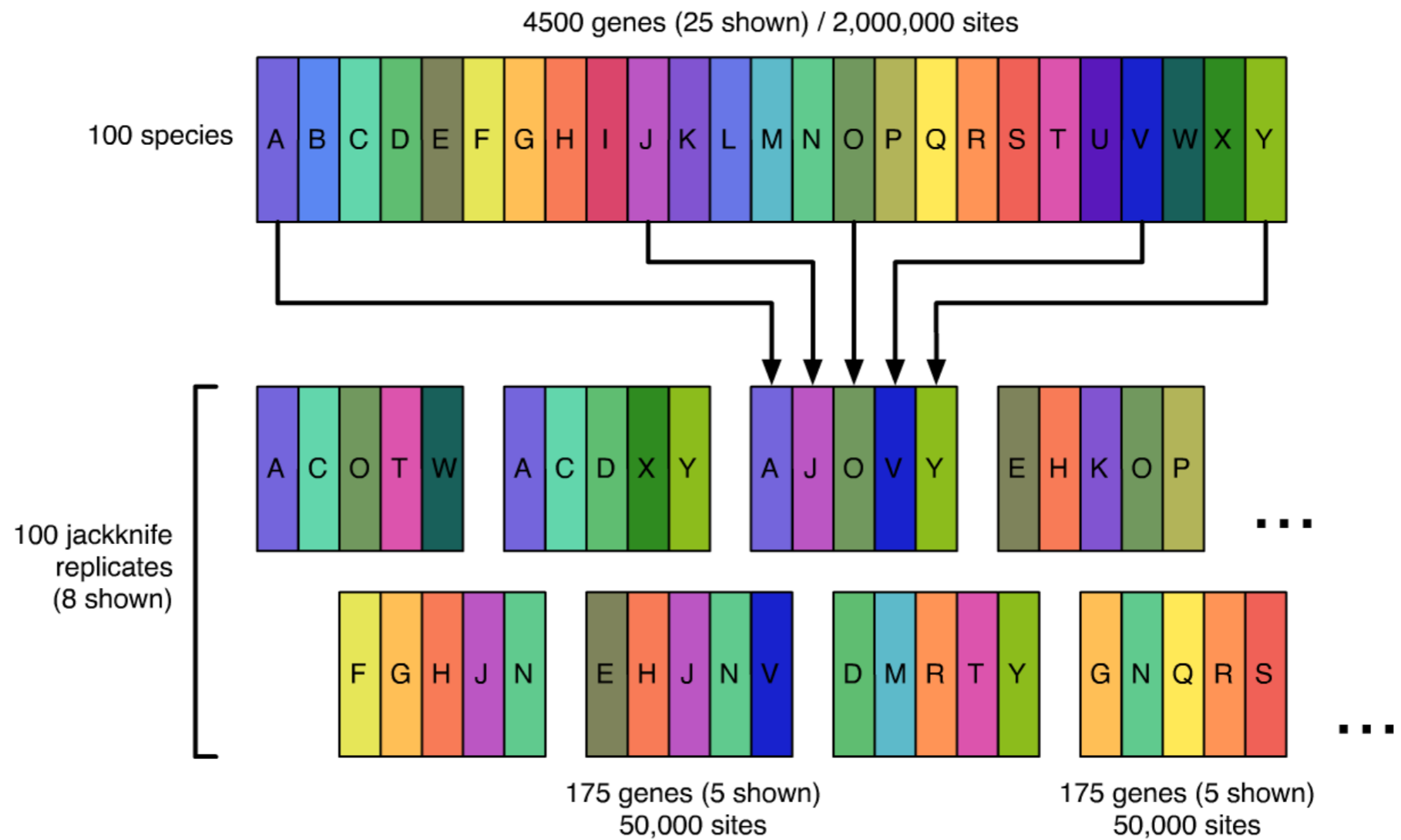


Arbre des Vertébrés

gene boundary




Arbre des Vertébrés



Our supermatrix was too large to be analysed "as is".
We thus used a gene resampling approach.

Arbre des Vertébrés



$$f(\theta|X) = \frac{f(\theta)f(X|\theta)}{f(X)} \text{ with } \theta = (\tau, v)$$

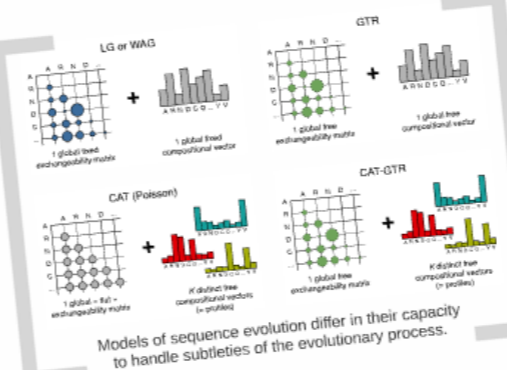
$$f(X) = \int f(\theta)f(X|\theta)d\theta$$

$$= \sum_{\tau} \int_{v} f(v)f(X|\tau, v)dv$$

Bayes' theorem applied to phylogenetics

The tree is the interesting part of the model while the model of sequence evolution is a necessary "nuisance".

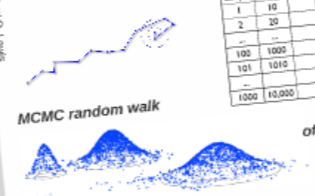
We want to compute the (posterior) probability distribution of the model (tree and sequence model) given the data (supermatrix).



Models of sequence evolution differ in their capacity to handle subtleties of the evolutionary process.



N. Lartillot



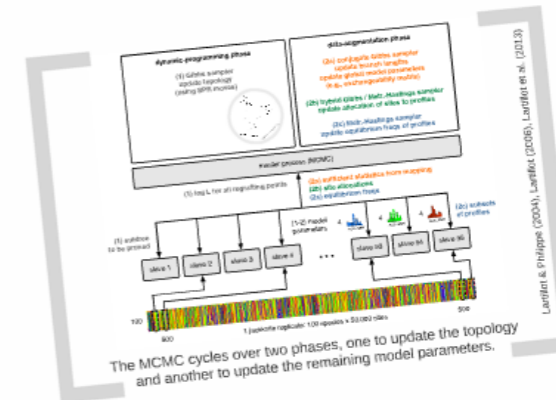
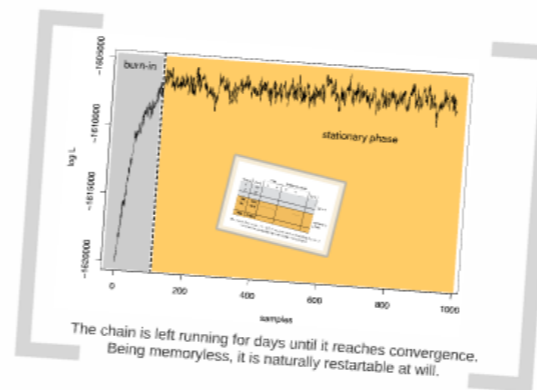
sample	cycle	tree					log L
		t	v	R	tt	-	
1	10						
2	20						
...	...						
100	1000						
101	1010						
...	...						
1000	10,000						

MCMC random walk

MCMC samples of the parameter space

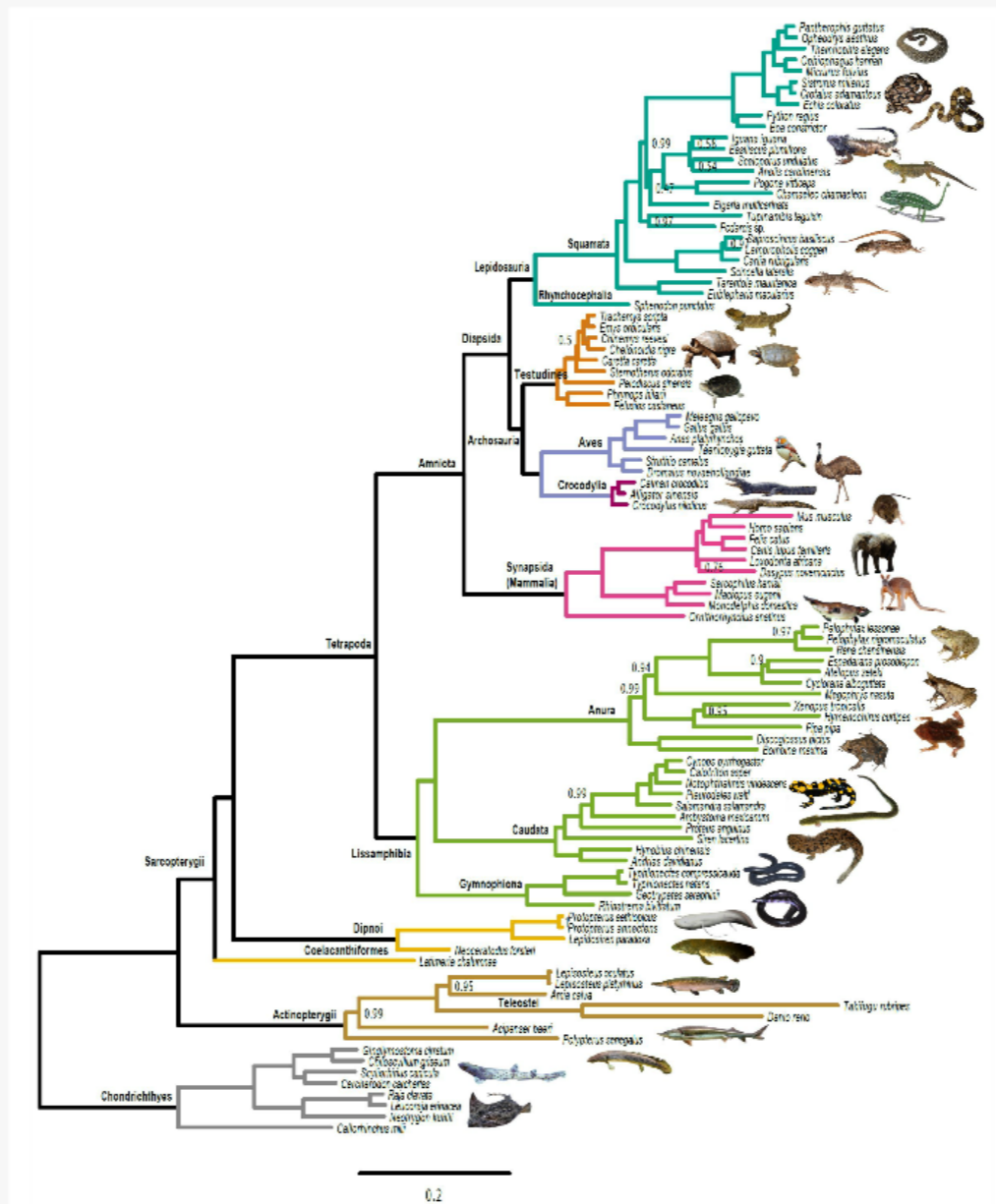
The best sequence models use complex mixtures that can only be implemented in a Bayesian framework.

The principle is to sample the posterior distribution using numerical simulation (Markov Chain Monte Carlo).



Each jackknife replicate was then analysed using a powerful Bayesian phylogenetic software: PhyloBayes MPI.

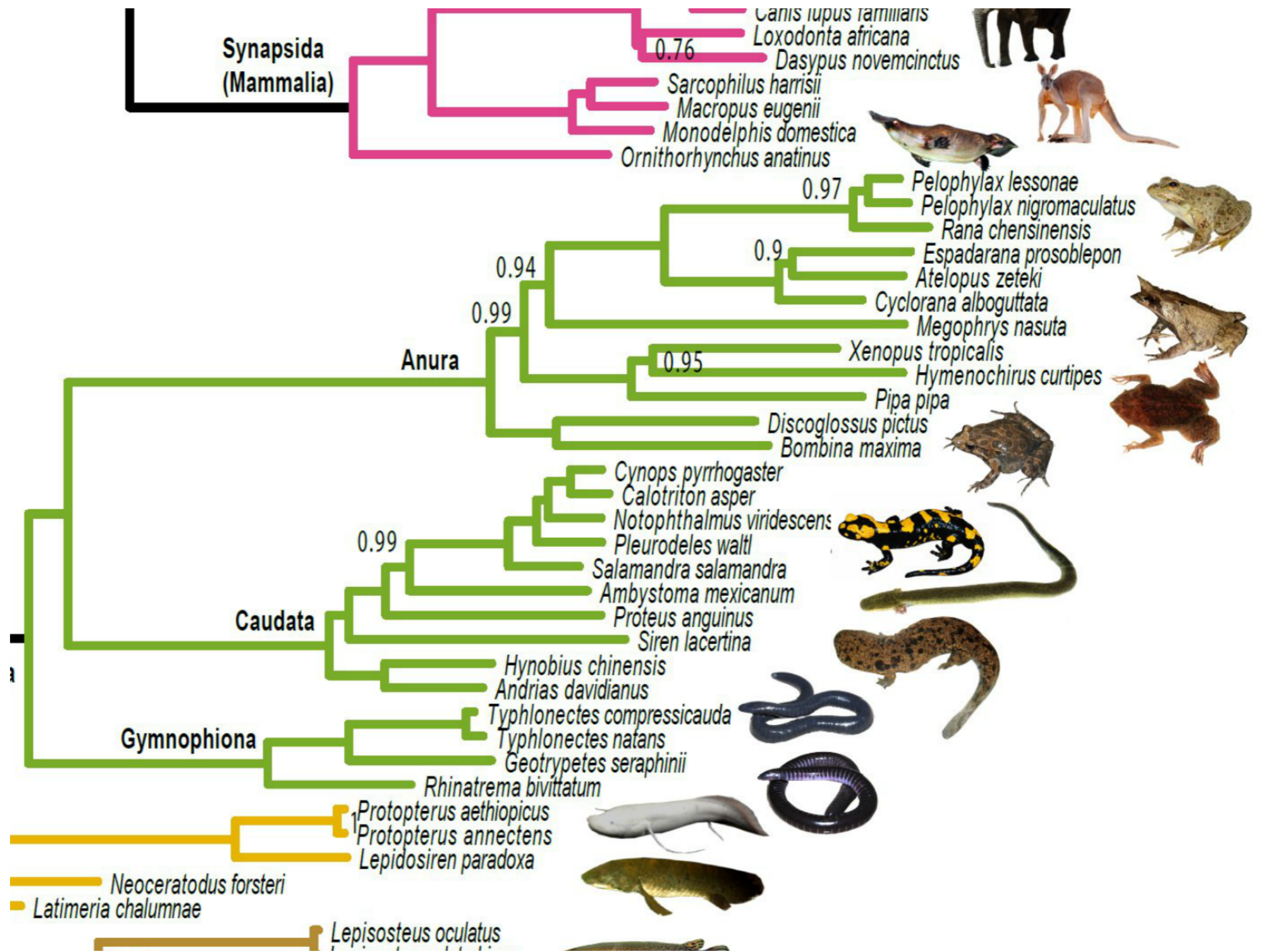
Arbre des Vertébrés



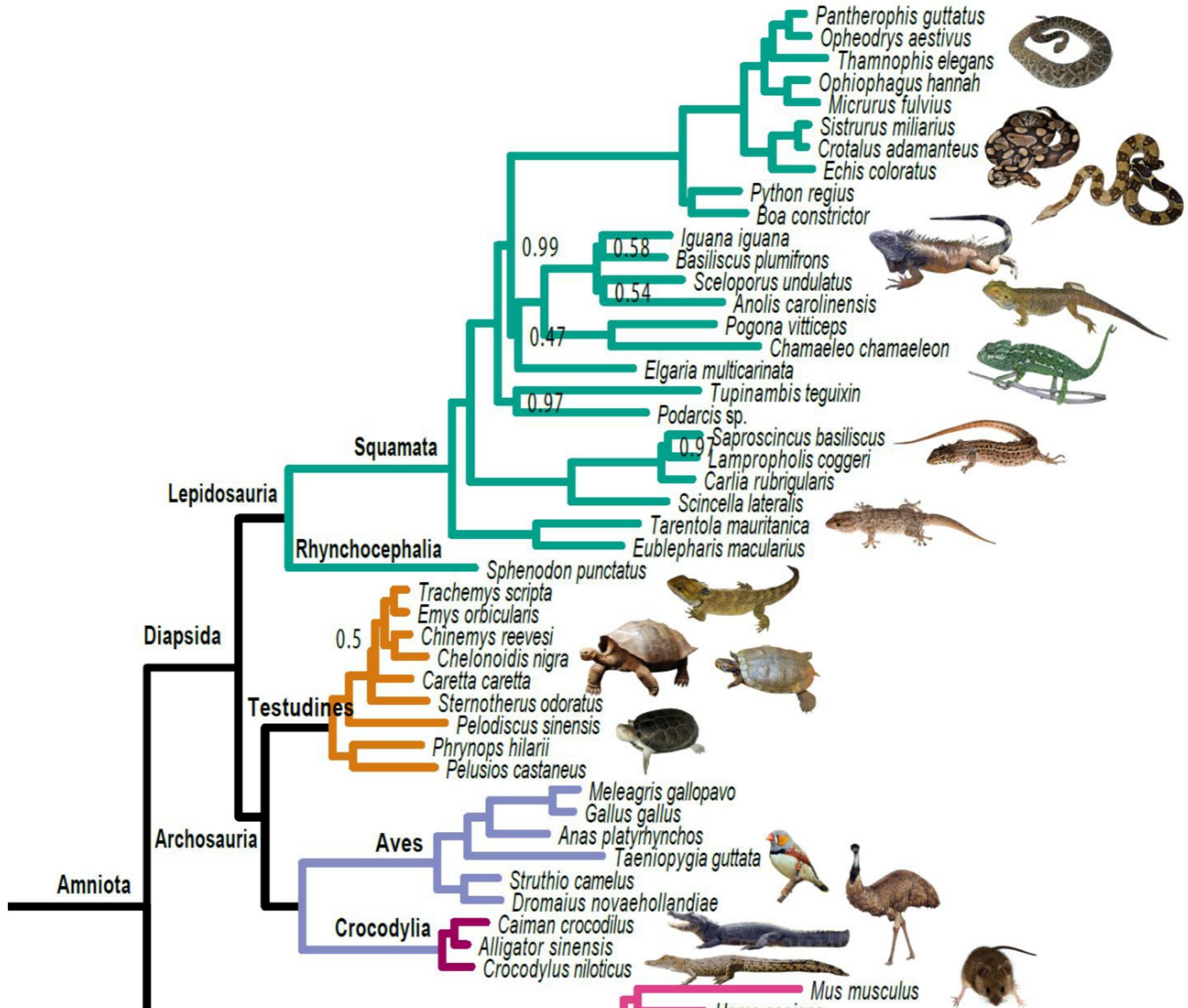
We obtained a new reference phylogenetic framework for the evolution vertebrates.

Irisarri, Baurain et al. (in prep.)

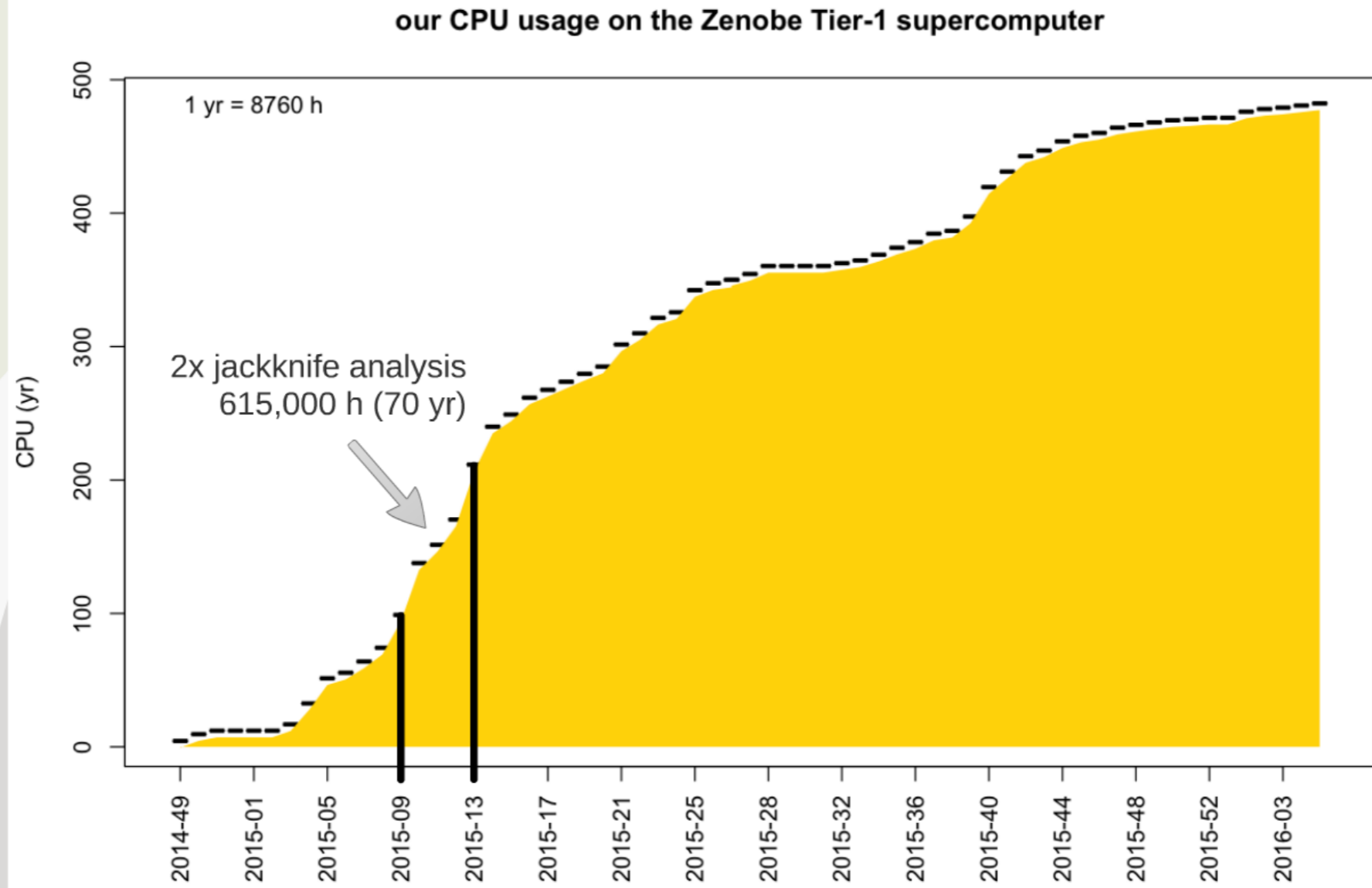
Arbre des Vertébrés



Arbre des Vertébrés



Arbre des Vertébrés



Even if we do not use many cores per job (96), we still need quite a lot of computing power (hence HTC).

Infrastructure / Supercalculateur zenobe

Supercalculateur zenobe

Le supercalculateur *zenobe* a été baptisé en l'honneur de Zenobe Gramme, l'électricien belge inventeur de la dynamo.

Il est constitué de deux parties combinées résultat d'un investissement de près de 5,5 millions d'euros, réalisé en trois temps :

- un calculateur acquis en 2011 dans le cadre du portefeuille de projets SINUS co-financé par le programme FEDER 2007-2013 constituant une base d'environ 3300 cœurs de calcul,
- son extension financée par la subvention PRACE Supercalculateur Tier-1 de la Région Wallonne, ajoutant fin 2013 environ 8200 cœurs de calcul à cette base, et
- le renouvellement, fin 2015, des nœuds de calcul acquis 4 ans plus tôt, dans le même cadre du portefeuille de projets SINUS, fournissant 5760 cœurs de calcul de dernière architecture

de sorte à ce qu'il constitue le supercalculateur Tier-1 de Wallonie avec près de 14000 cœurs de calcul. Le supercalculateur a été mis en exploitation, dans sa configuration Tier-1, pour 3 ans à compter du 1er juillet 2014.



Le supercalculateur a une capacité de calcul soutenue de plus de 330 TFLOPS (milliards d'opérations à la seconde). Sa meilleure position dans le classement mondial des 500 calculateurs les plus puissants a été enregistrée en novembre 2014 à la 300^e place. Il offre en outre une capacité de stockage haute performance de 350 To en parallèle des 120 To de stockage standard. L'infrastructure est connectée à haut débit (1 Gbps) au réseau de recherche belge, Belnet, et par là au réseau recherche européen GÉANT et à Internet. Les caractéristiques techniques détaillées du supercalculateur dans sa configuration actuelle sont reprises dans la [documentation](#).

