

Ab⁴ims



Ifremer

23/06/2014

Introduction aux méthodes de phylogénie



CNRS UPMC

Station Biologique
Roscoff

UPMC
SORBONNE UNIVERSITÉS



INTRODUCTION

- Préambule historique et lexical

PARTIE I

- La phylogénie moléculaire
 - Les caractères
 - Le jeu de données
 - Les banques
 - Les alignements
 - Les arbres

PARTIE II

- Les distances évolutives

PARTIE III

- Les méthodes d'analyse phylogénétiques
 - Méthodes de distances
 - Méthodes de maximum de parcimonie
 - Méthodes de maximum de vraisemblance
 - Méthodes Bayésiennes

PARTIE IV

- Attraction des longues branches
- Robustesse des arbres phylogénétiques
- Analyse multigénique et superarbres
- Mesure de la pression évolutive

PARTIE V

- Les logiciels

PARTIE VI

- Les travaux pratiques



Laure Quintric



Fanny Marquer



Gildas Le Corguillé



Erwan Corre

Cellule bio-informatique
IDM/RIC – IFREMER
BREST



Plateforme bio-informatique ABiMS
FR2424-CNRS UPMC
ROSCOFF



CNRS UPMC INSU

Station Biologique
Roscoff



INTRODUCTION



- Préambule historique et lexical

PARTIE I



- La phylogénie moléculaire
 - Les caractères
 - Le jeu de données
 - Les banques
 - Les alignements
 - Les arbres



PARTIE II

- Les distances évolutives



PARTIE III

- Les méthodes d'analyse phylogénétiques
 - Méthodes de distances
 - Méthodes de maximum de parcimonie
 - Méthodes de maximum de vraisemblance
 - Méthodes Bayésiennes



PARTIE IV



- Attraction des longues branches
- Robustesse des arbres phylogénétiques
- Analyse multigénique et superarbres
- Mesure de la pression évolutive



PARTIE V



- Les logiciels

PARTIE VI

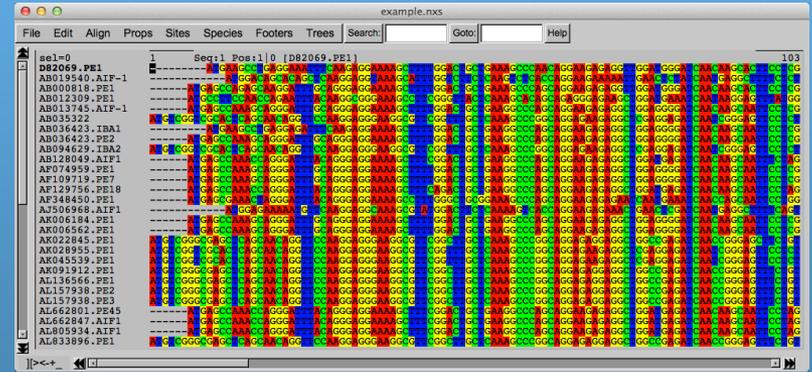
- Les travaux pratiques



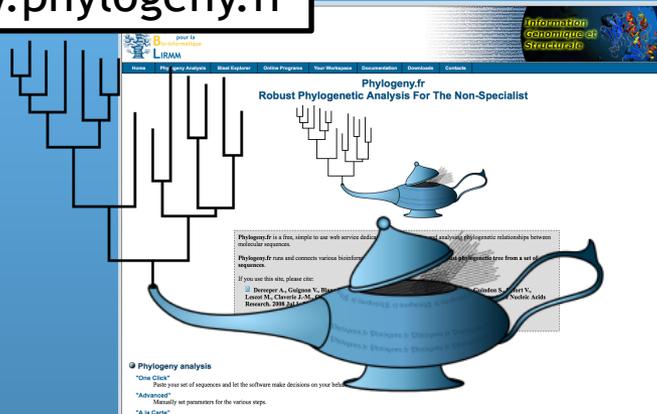
Galaxy



Seaview 4.5



www.phylogeny.fr



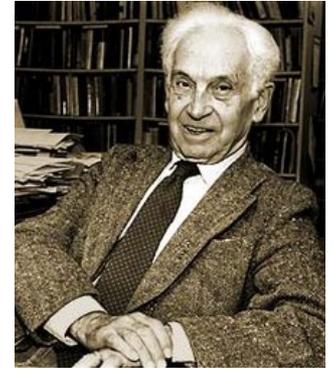
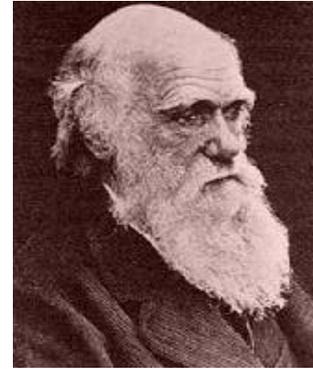
Cerveau 1.0



Support de cours
Lien ABlMS

- Lundi 9h30 ->(12h-13h30) ->17h
- Mardi 9h30 ->(~12h-13h30) -> ?? (17h max)

- Repas Gulf Stream ou Friterie

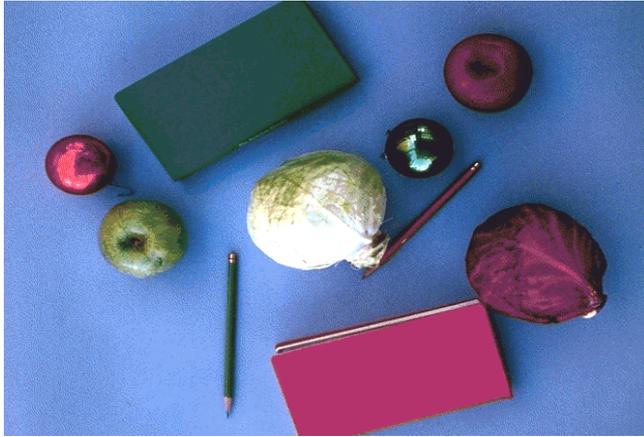


Introduction

PRÉAMBULE HISTORIQUE ET LEXICAL

Préambule : Taxonomie

L'étude des relations entre des groupes d'organismes est appelé Taxonomie, c'est l'art de classer des entités dans des groupes.



Recherche de la classification naturelle

C'est une discipline très ancienne et très noble de la biologie développée par Carolus Linnaeus (1707-1778).

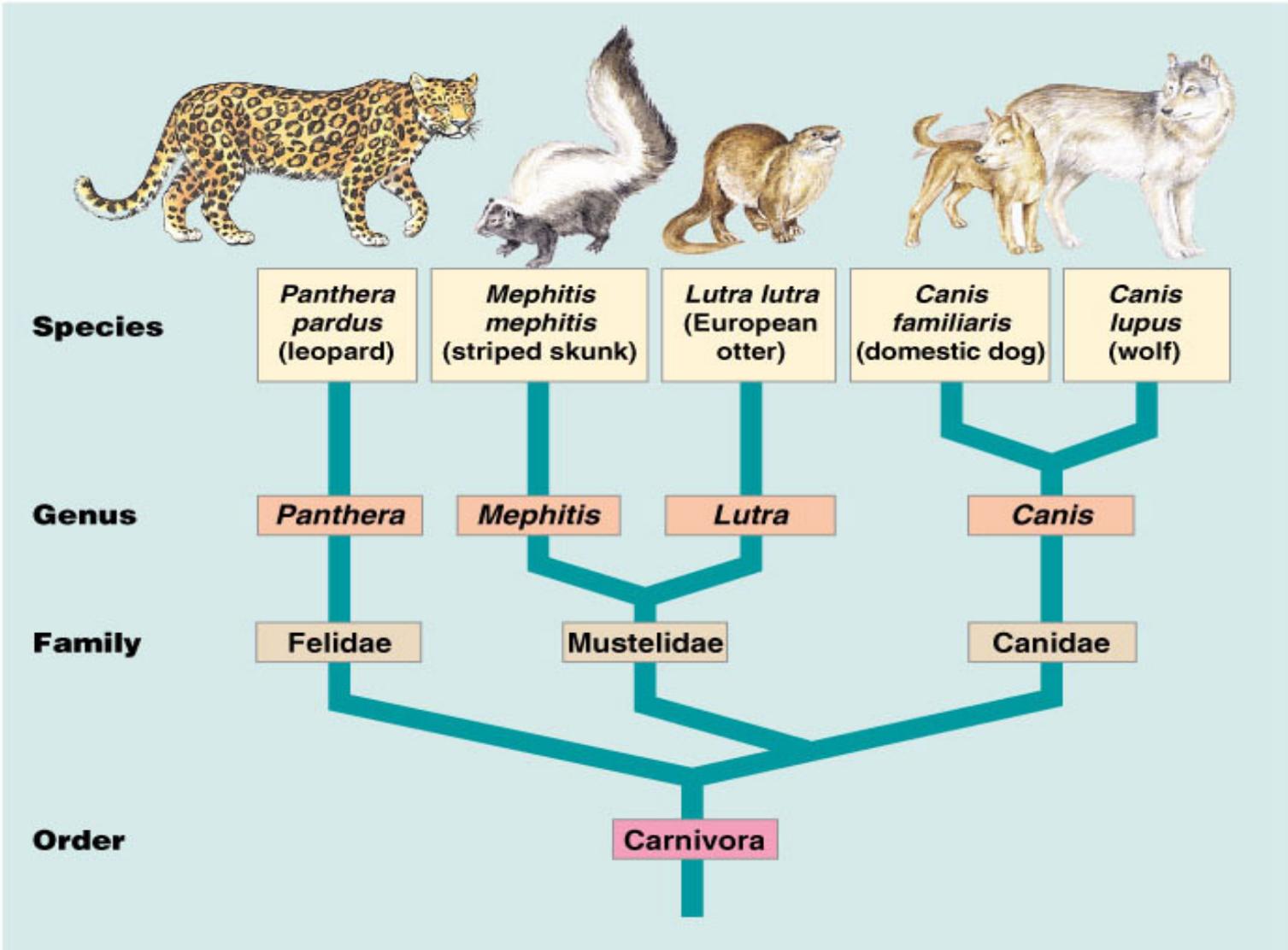
Distingue la classification populaire de la **classification naturelle**.

Fondateur de la nomenclature binaire (*Systema Naturae*, 1758) : remplaçant le nom vernaculaire



Domaine	Eucaryotes
Règne	Métazoaires
Embranchement	Vertébrés
Classe	Mammifères
Ordre	Carnivores
Famille	Félidés
Genre	<i>Felis</i>
Espèce	<i>concolor</i>

Préambule : Taxonomie



Créationnisme et **fixisme** (chaque espèce serait apparue telle quelle au cours des temps géologique) prévalent pendant 2000 ans



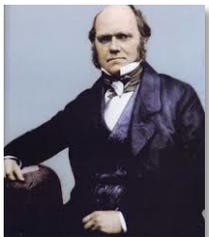
Le **transformisme** notion apparue au XVIII^{ème} siècle

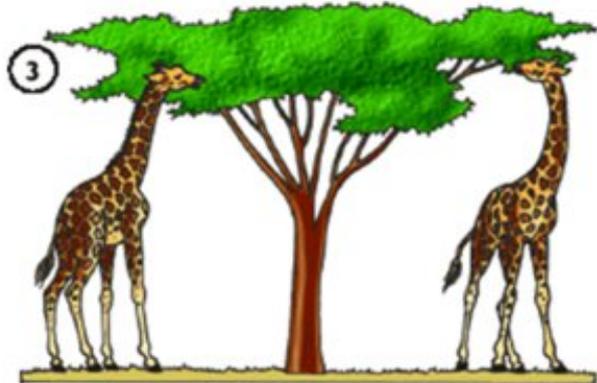
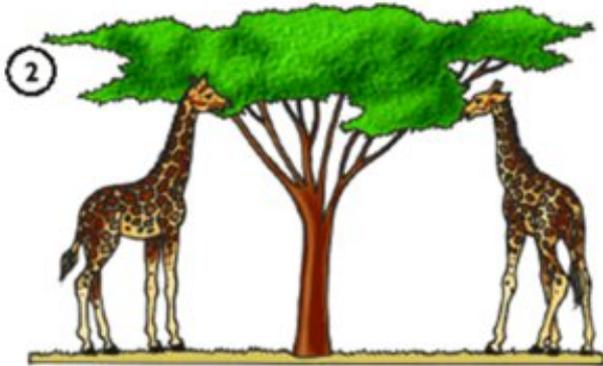
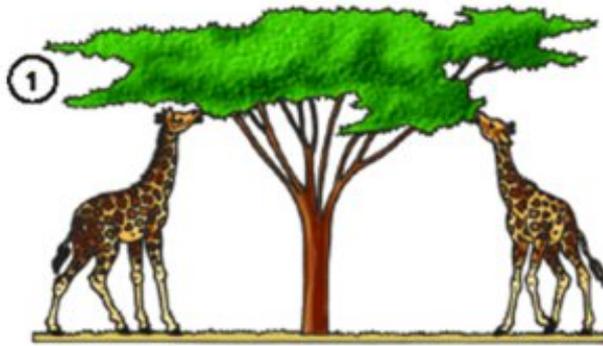
- Travaux de Buffon : transformation des espèces vivantes selon 3 facteurs : le climat, la nourriture et la domesticité
- Travaux de Lamarck : Les espèces se transforment au cours du temps en s'adaptant aux variations de leur milieu (transmissibilité des caractères acquis)**
- -> Cuvier lui oppose le **catastrophisme** : une création divine remplace les espèces disparues après chaque catastrophe anéantissant toute vie sur Terre.



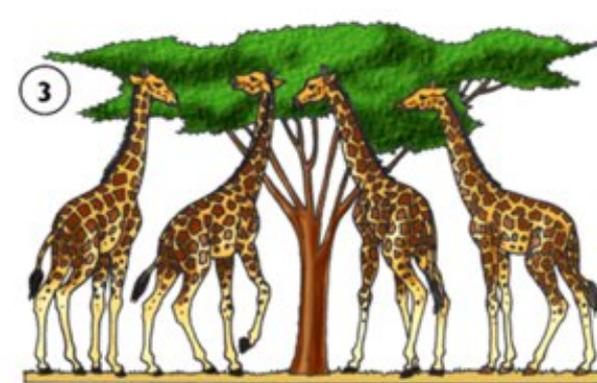
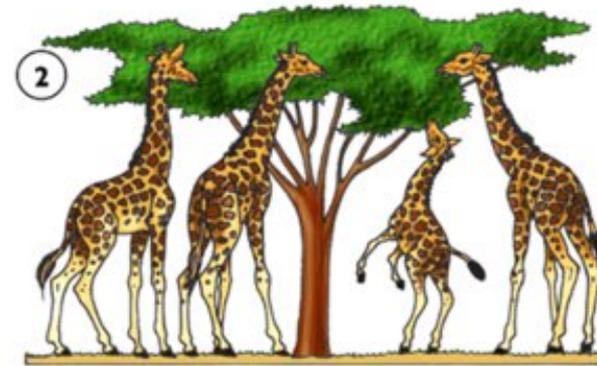
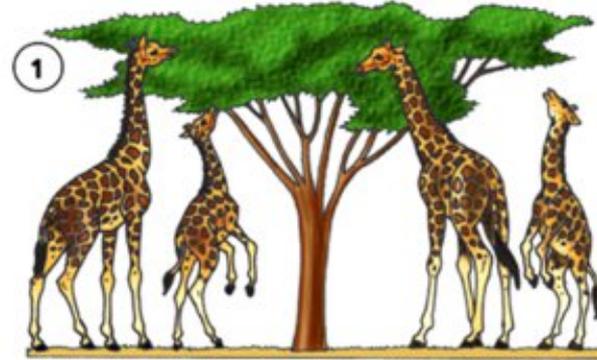
Le Darwinisme : Darwin, A. Wallace

Dans une population possédant une certaine variabilité de caractères sont sélectionnés les individus ayant les caractères les plus avantageux.





Lamarckismo



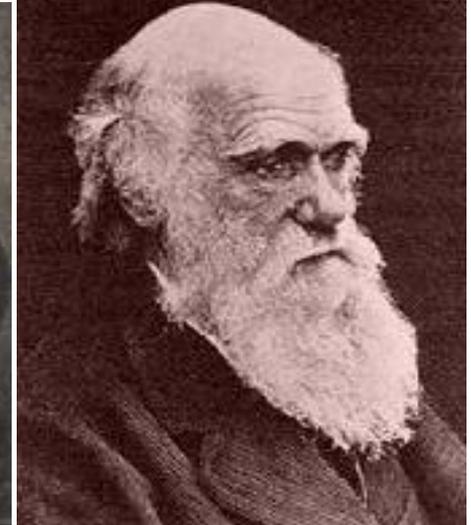
Darwinismo

La théorie de l' évolution est la base de toute la biologie moderne

De l' anatomie, aux études comportementales ou à la génomique, les méthodes scientifiques nécessitent une **appréciation des changements au sein des organismes au cours du temps.**



Alfred Russel Wallace



Charles Darwin

Le terme de phylogénie est inventé par Ernst Haeckel en 1866 pour définir l'enchaînement des espèces animales et végétales au cours du temps. Jusqu'alors le concept était exprimé par le terme de généalogie.

Ce terme va être repris par Darwin dans « l'origine des espèces ».

Phylogénie = Evolution + Taxonomie

Du grec · Phûlon = tribus · Genesis = origine

- Le réel développement de la phylogénie date du milieu du 20ème siècle.
- La cladistique a été développée par Willi Hennig, un entomologiste allemand en 1950.
- La systématique phylogénétique (cladistique) est une méthode de classification taxonomique des organismes basée sur leur histoire évolutive



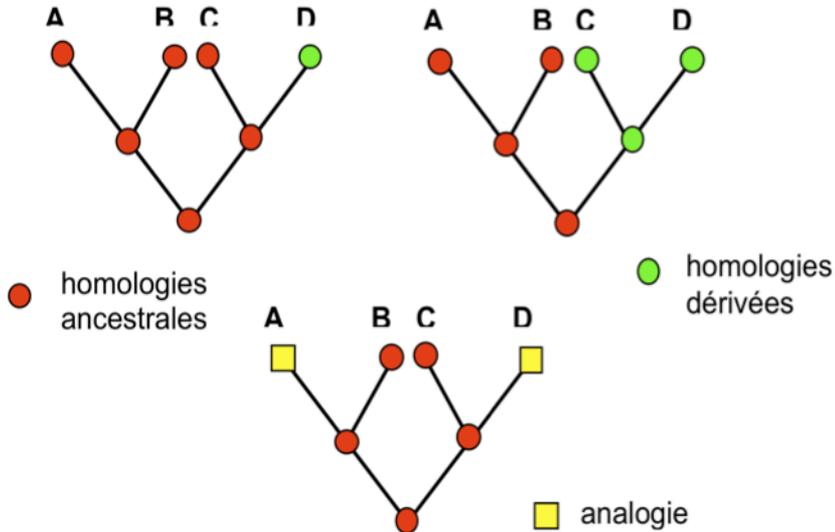
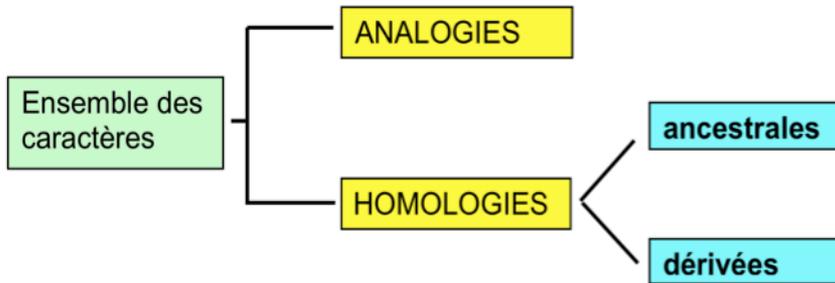
Le double constat de Hennig : les Synapomorphies

A. Tous les caractères n'évoluent pas à la même vitesse.

Exemple: les 5 doigts à la main et au pied chez l'homme, apparus déjà à l'époque des premiers amphibiens ; et la bipédie, apparue il y a seulement 5 millions d'années.

B. Seul le partage de caractères dont l'état s'est transformé est signe d'une parenté étroite (homologie phylogénétique ou **synapomorphie**).

A contrario, les caractères restés à l'état primitif (ou **symplesiomorphies**) ne témoignent pas d'une parenté étroite



Les homoplasies :

Analogies = ressemblances
= pas informatifs
= sources d'erreurs

Les homologies :

Synplésiomorphie
= caractère plésiomorphe
= ancestral
= pas informatif

Synapomorphie

= caractère apomorphe
= dérivé
= informatif

Ex. protéines

Homologue: même structure, fonction identique ou similaire, ascendance commune.

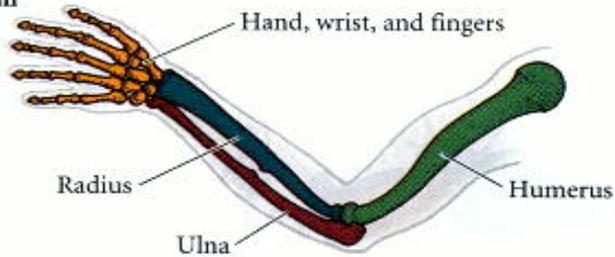
Analogie: même structure, fonction différente, origine ancestrale inconnu

Définitions de l'homologie

- Owen (1843). «Le même organe sous toutes les variétés de forme et de fonction ».
- Fitch (2000). Homologie est la relation de deux caractères qui sont descendus, généralement avec divergence, d'un caractère ancestral commun. (Fitch, W. M. (2000). Homology a personal view on some of the problems. Trends Genet 16, 227-31.)
- Un "caractère" peut être un trait phénotypique, ou d'un site à une position donnée d'une protéine, un gène ou ensemble, ...
- D'un point de vue moléculaire: deux gènes sont homologues si ils divergent à partir d'un gène ancestral commun.

Préambule : Cladiste->vocabulaire

Human arm



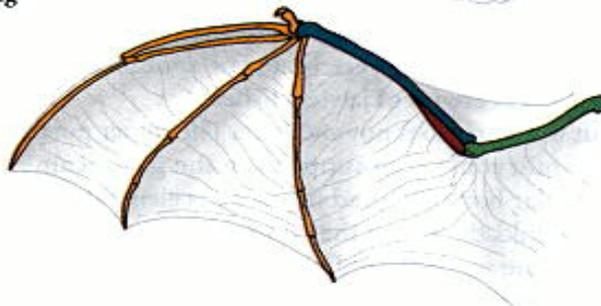
Seal limb



Bird wing



Bat wing



L'aile chez l'oiseau et le bras de l'homme sont **homologues**, en tant que **patte avant** !

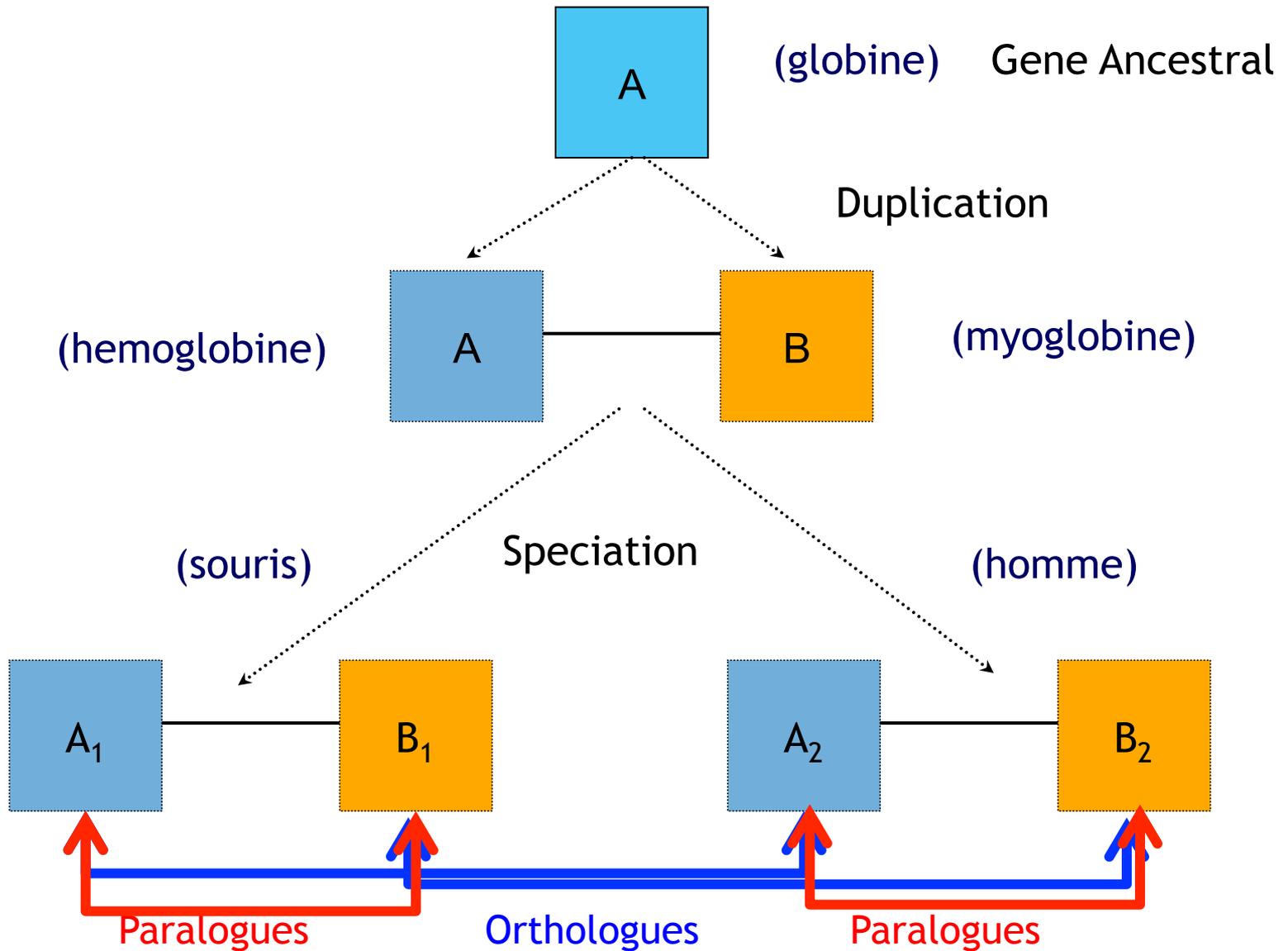
Par contre l'aile chez la chauve souris et l'aile chez l'oiseau **ne sont pas homologues en tant qu'aile**, car la fonction d'aile a été acquise séparément !

- Homologie est une relation booléenne (vrai ou faux):
- deux séquences sont homologues, ou ne le sont pas.
- Il est donc inexact de parler de "pourcentage d'homologie".
- La formulation correcte est :
 - Deux séquences sont homologues (avec un risque mesurable d'erreur) , parce qu'elles partagent un certain pourcentage de identité ou de similitude.

Orthologues vs. Paralogues

- Lorsque l'on compare des séquences de gènes, il est important de distinguer dans différents organismes les gènes identiques de ceux seulement similaires.
- **Orthologues** sont des gènes homologues chez différentes espèces ayant des fonctions analogues.
- **Paralogues** sont des gènes similaires qui sont le résultat d'une duplication génique.
- Une phylogénie qui comprend des orthologues et des paralogues est susceptible d'être inexacte.
- Parfois, l'analyse phylogénétique est la meilleure façon de déterminer si un nouveau gène est un orthologue ou paralogue à d'autres gènes connus.

Préambule : Orthologues vs Paralogues



Cenancestor: le plus récent ancêtre commun des taxons étudiés.

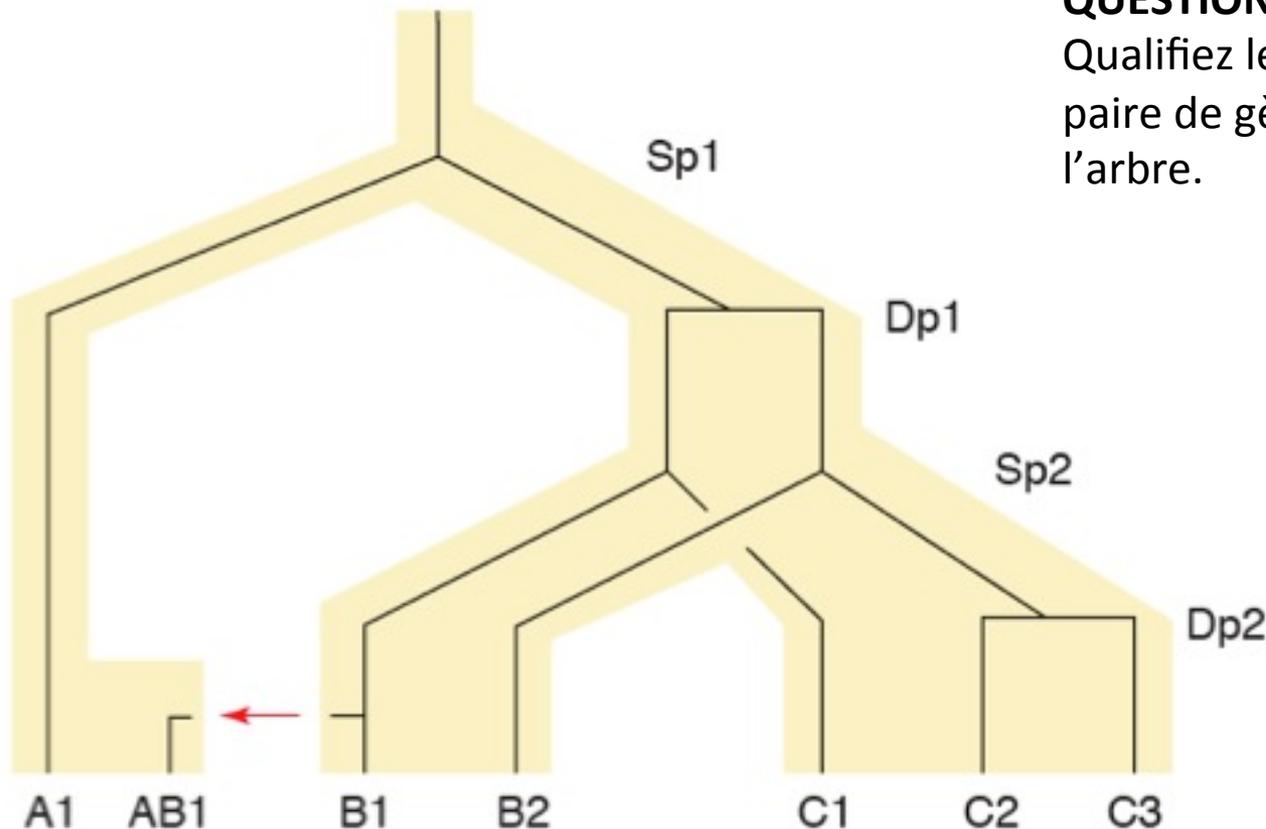
Xenology: relation de deux caractères dont l'histoire, depuis leur ancêtre commun, implique le transfert interspécifique (horizontal) de matériel génétique pour au moins un de ces caractères.

DONNEES:

- 3 taxa : A,B,C
- 2 évènements duplication (Dp1 Dp2)
- 2 évènements spéciation (Sp1 Sp2)

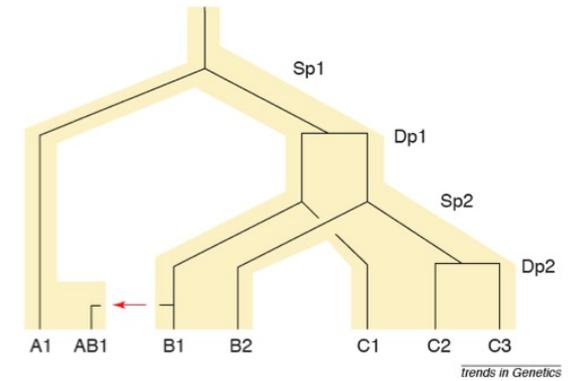
QUESTION :

Qualifiez les relations entre chaque paire de gènes représentés dans l'arbre.

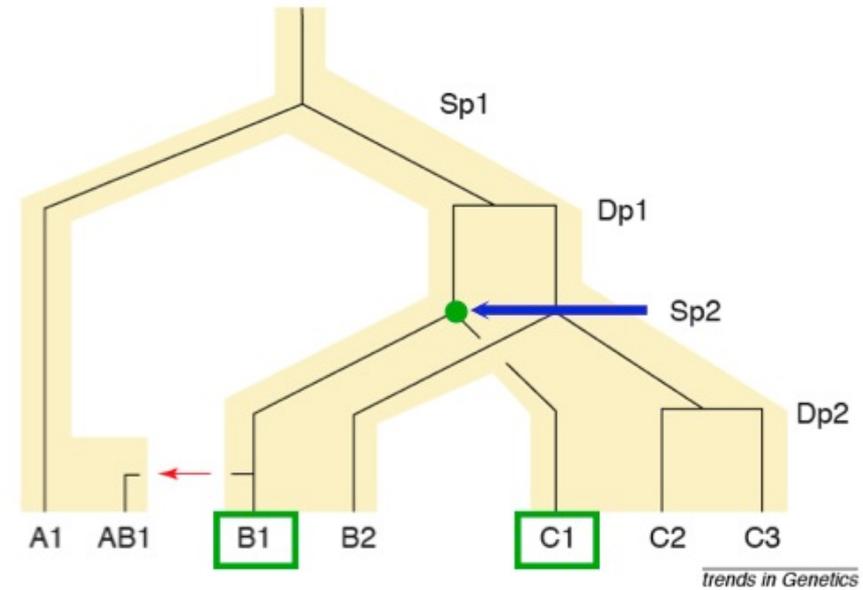


P : paralogue
O : orthologue
X : xenologue
A : analogue

P : paralogue
O : orthologue
X : xenologue
A : analogue



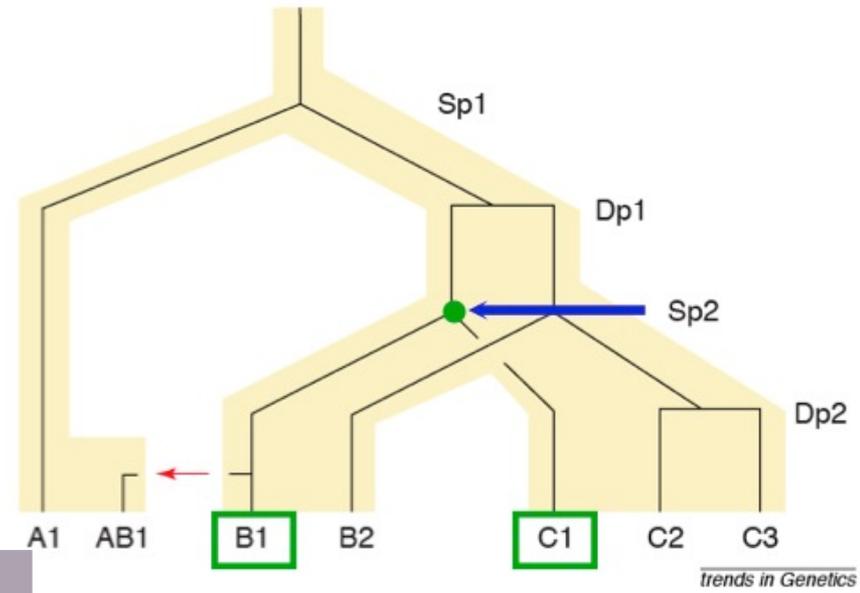
	A1	AB1	B1	B2	C1	C2	C3
A1							
AB1							
B1							
B2							
C1							
C2							
C3							



B1 - C1

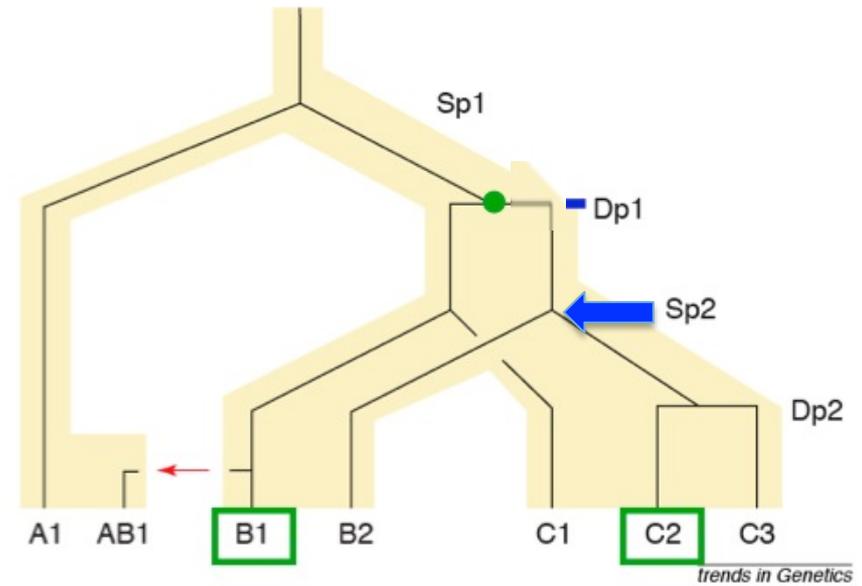
- Les deux séquences (B1 et C1) ont été obtenues à partir des taxons B et C.
- Le cenancestor (**flèche bleue**) est le taxon qui a précédé le second événement de spéciation (SP2).
- Le gène ancêtre commun (**point vert**) coïncide avec le cenancestor

B1 et C1 sont orthologues



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

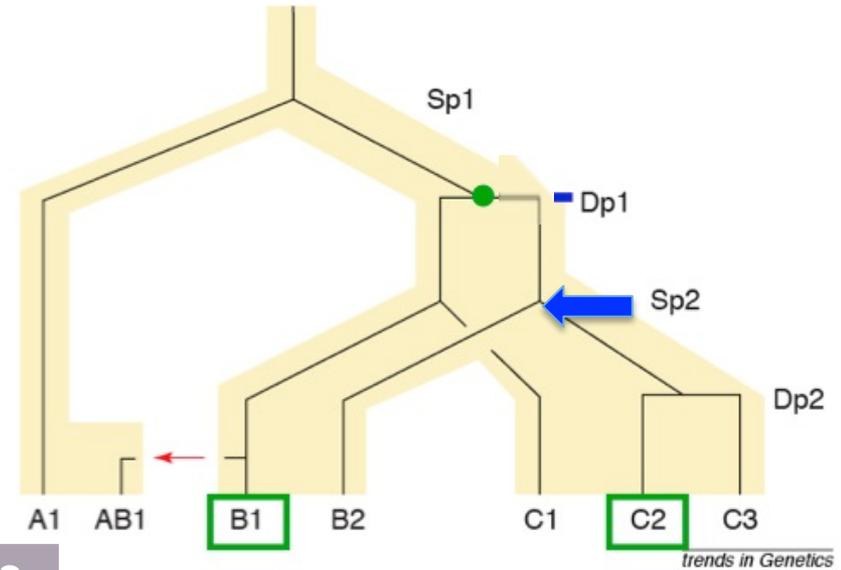
P : paralogue
 O : orthologue
 X : xenologue
 A : analogue



B1 - C2

- Les deux séquences (B1 et C2) ont été obtenues à partir des taxa B et C.
- Le gène ancêtre commun (**point vert**) est le gène qui précède l'événement de duplication Dp1.
- Cet ancêtre commun est très antérieur au cenancestor (**flèche bleue**).

B1 et C2 sont **paralogues**



	A1	AB1	B1	B2	C1	C2	C3
A1	I						
AB1		I					
B1			I				
B2			P	I			
C1			O		I		
C2						I	
C3							I

P : paralogue
 O : orthologue
 X : xenologue
 A : analogue

Les relations sont **réflexives**:

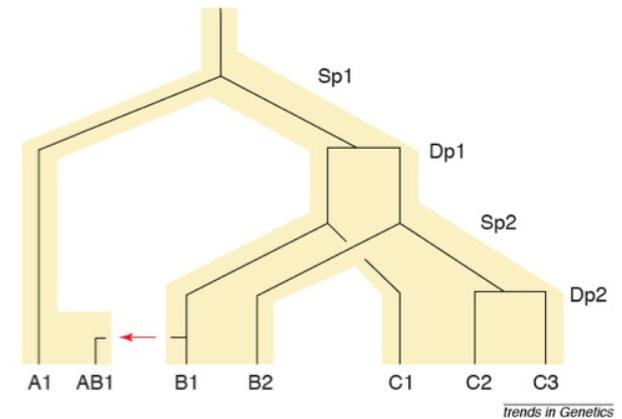
Ex : A1 est orthologue (\rightarrow) à B1

B1 est orthologue à A1

Elles ne sont **pas transitives**:

Ex: C2 \rightarrow A1 \rightarrow C3 est vrai

mais C2 est paralogue de C3



	A1	AB1	B1	B2	C1	C2	C3
A1	I						
AB1	X	I					
B1	O	X	I				
B2	O	X	P	I			
C1	O	X	O	P	I		
C2	O	X	P	O	P	I	
C3	O	X	P	O	P	P	I

P : paralogue
 O : orthologue
 X : xenologue
 A : analogue

Methodes

- Best hit (BH)
- Reciprocal (RBH) or bidirectional (BBH) best hit.

Problème 1: la non-réciprocité de la relation BH, qui peut résulter de divers effets

- protéines multidomaines -> non-transitivité de la relation d'homologie dans ce cas
détection: mais pas paralogie
- des paralogues dans un génome correspondent à un même orthologue dans un autre génome (fusion de gènes)
- non-symétrie de la suite BLAST

Problème 2: détection d'un RBH mais faux RBH

- Duplication suivie d'une suppression
- Deux paralogues peuvent être BBH, car les vrais orthologues ne sont plus présents dans le génome (en raison de la suppression).

Problème conceptuel: intrinsèquement incapable de traiter les relations multi-orthologie.

Ex figure: B2 est orthologue à la fois C2 et C3, mais un seul d'entre eux sera son meilleur Hit.

Conclusion: l'analyse de BBH est intrinsèquement incapable de révéler les véritables relations d'orthologie.

Les solutions aux problèmes de RBH

Analyse par domaines protéiques : analyser l'emplacement des hits dans les alignements

- Résout les problèmes de fusion de gènes (deux fragments différents d'une protéine dans le génome A correspondent aux deux protéines distinctes de génome B)

L'analyse de l'histoire de l'évolution: l'inférence phylogénétique complète + la réconciliation de l'arbre de la séquence et l'arbre des espèces

- Résout les cas de multiples relations d'orthologie (n à n)
- Ne résout pas les problèmes de délétions différentielles après duplications de régions

-> Résoudre le problème des duplications régionales suivie par des délétions

- Analyse de synténie: les relations de voisinage entre les gènes dans les génomes
- Analyse des pseudogènes: permet de déduire la présence d'un gène putatif à l'ancêtre commun.

Ces solutions nécessitent une analyse au cas par cas