Traitement informatique des données

Présentation du module:

Cette matière contribue au développement des acquis de l'étudiant en ce
Qui concerne les nouvelles bases de la bioinformatique

Connaissances préalables recommandées
Génétique
Biologie moléculaire
Microbiologie
Biochimie

Cours 04: Concepts évolutifs en bioinformatique

Chargée de cours : ALIANE S

1. Similarité :

% Identité « identities »: Proportion de résidus (acide aminés ou nucléotides) identiques entre les séquences.

% Similarité « positives »: Proportion de résidus similaires entre les séquences

Identities = 14/26 (54%),

Positives = 15/26 (58%),

VCGMWLDGDIAAVDMFTHVEIGDDEV

V G W+ GDI MFTH DD EV

VAGIWVAGDIRGGPMFTHTAYDDFEV

Important

- ***% Identité ou similarité maximale: 100%***
- ***% Identité entre deux séquences nucléiques aléatoires ?***

4 bases => % Identité attendu entre séquences aléatoires: ¼= 25%

- ***% Identité entre deux séquences protéiques aléatoires ?***

20 aa => % Identité attendu entre séquences aléatoires: 1/20 = 5%

2. Similarité, homologie, analogie :

La similarité entre deux séquences peut s'interpréter par trois hypothèses alternatives:

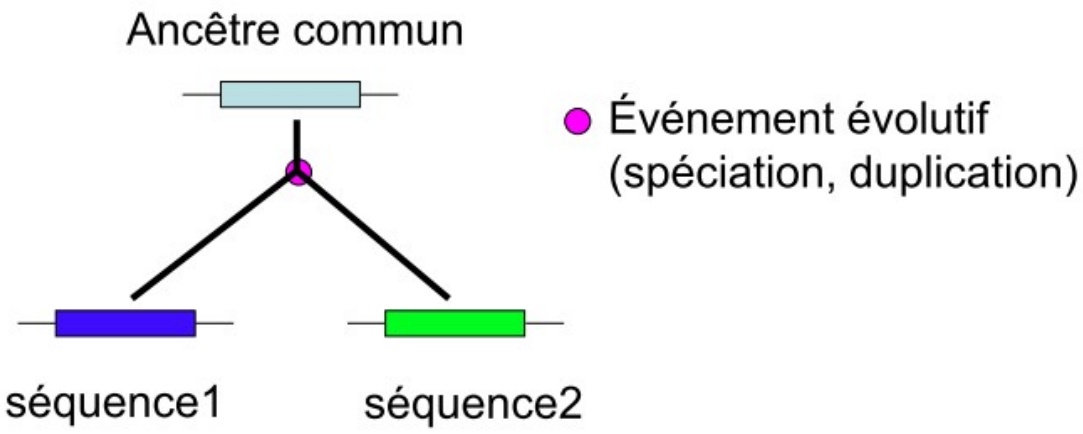
– **Homologie**: la ressemblance s'explique par le fait que les deux séquences divergent d'un ancêtre commun.

– **Evolution convergente (analogie):** les similarités sont apparues dans les deux séquences de façon indépendante, mais ont été sélectionnées pour la même raison.

– **Similarité due au hasard**

2 .1. Homologie • Deux séquences sont dites homologues si elles possèdent un ancêtre commun

• L'existence d'un ancêtre commun est inférée à partir de la similarité



➢ Homologie ≠ Similarité •

• L'homologie n'est pas quantifiable

– Deux séquences sont homologues (possèdent des caractères communs parce qu'elles dérivent d'un ancêtre commun) ou elles ne le sont pas.

– Raisonnement binaire (vrai ou faux)

• La similarité est quantifiable

– On peut dire de deux séquences qu'elles sont similaires à 50% ou 75%

• Généralement on infère l'homologie sur base de similarité élevée entre les séquences,

2.2. Duplication et spéciation

• Deux séquences qui descendent d'un ancêtre commun divergent.

• La divergence peut résulter d'une duplication ou d'une spéciation.

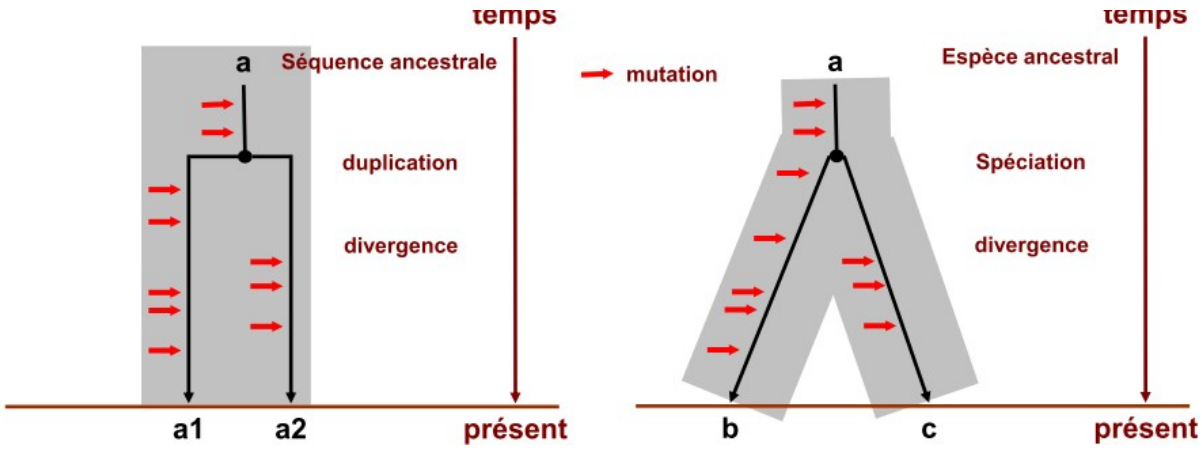
– ***Duplication*** : mutation génétique caractérisée par le dédoublement du matériel génétique sur un chromosome.

– ***Spéciation :*** Apparition d'une nouvelle espèce par différenciation entre deux populations

• ***Evénements de mutations*** :

– Substitutions

Délétions + Insertions = indel



• on observe un certain niveau de similarité entre deux séquences (% identité, % similarité).

• Sur cette base, on évalue des scénarios évolutifs:

cette similarité peut provenir

– d'une évolution convergente (analogie);

– d'une évolution divergente à partir d'un ancêtre commun (homologie);

– du hasard.

• Si la deuxième hypothèse est la plus vraisemblable, on infère que les séquences sont homologues.

❖ Homologie/analogie

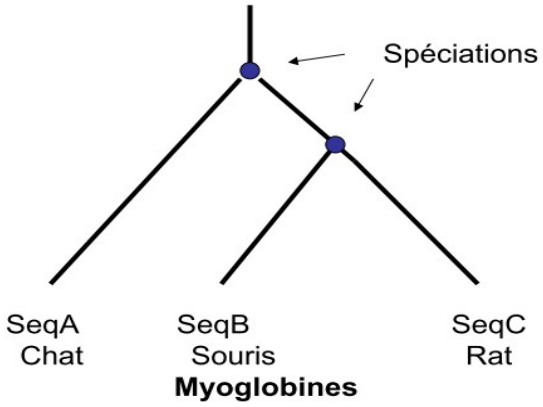
• ***Analogie:*** relation entre deux caractères qui se sont développés de manière convergente à partir d'ancêtres différents.

• ***Homologie:*** L'homologie est la relation entre les deux caractères qui sont descendus, le plus souvent avec une divergence, d'un caractère ancestral commun.

Les catégories d'homologies

1. Orthologie: Paire de gènes homologues dont le dernier ancêtre commun a eu lieu immédiatement avant un événement de spéciation

• A, B et C sont **orthologues**



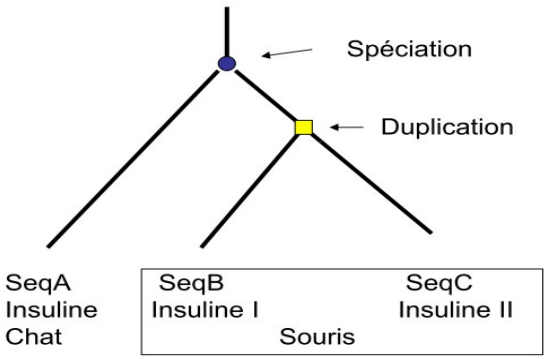
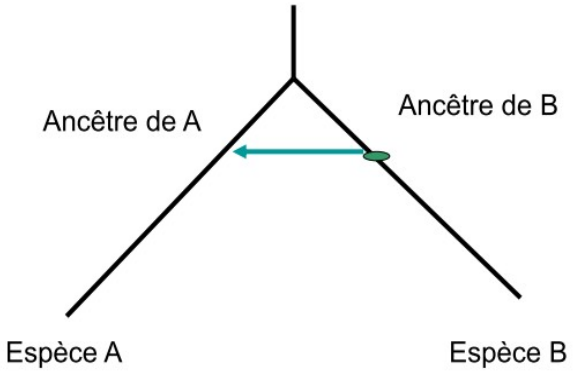
2. Paralogie:

Paire de gènes homologues dont le dernier ancêtre commun a eu lieu immédiatement avant un événement de duplication génique

• Les fonctions d'un ou de plusieurs paralogues peuvent changer au cours de l'évolution (spécialisation, nouvelle fonction)

• B et C sont paralogues

• A et C, A et B sont orthologues



3. Xénologie: relation entre les deux gènes homologues dont l'histoire, depuis leur dernier ancêtre commun, implique le transfert interspécifique (horizontal) du matériel génétique.