

Écart 1 au modèle de Hardy-Weinberg : Sélection

Hardy-Weinberg : Modèle expliquant ce qui se passe quand aucune force évolutive agit (les fréquences alléliques ne varient pas au sein de la population) ;

Écart 2 au modèle de Hardy-Weinberg : Population de taille finie

Voir séance d'exercices d'hier

Un étrange phénomène sur les côtes bretonnes (ceci est une fiction)

On construit sur les côtes bretonnes une base navale, isolant une grande crique dans laquelle on trouve une population d'oursins. Dans un premier temps, les pêcheurs d'oursins découvrent 3 phénotypes différents : Pourpre, orangé et blanc.

La seule conséquence écologique qu'on peut aujourd'hui imputer à la construction de la base navale est la fragmentation de l'ancienne grande crique principale en isolats indépendants. Dans certaines de ces populations, on constate une disparition de deux des trois phénotypes présents au départ.

Classe de taille des populations / crique	3 phénotypes présents	1 seul phénotype présent	Nombre total de criques / classe
0 – 25 individus	2	18	20
25 – 75 individus	6	4	10
75 – 125 individus	8	2	10
+ de 125 individus	12	2	14

Comment peut-on expliquer que dans certaines populations on ne trouve plus qu'un seul phénotype ?
Hypothèse → Les premières sélections aléatoires lors de la création de l'isolat font qu'un des allèles peut avoir une fréquence relativement élevée.

Autre élément important qu'on constate : Les 3 phénotypes sont présents dans les populations les plus grandes. Les populations ne concentrant plus qu'un seul phénotype sont petites... Pourquoi ? (rappel : dans les pop° où y a qu'un seul phénotype, cela signifie que qu'il n'y a qu'un seul allèle présent)

→ La perte de diversité génétique est beaucoup plus importante dans les petites populations que dans les grandes. Les fluctuations alléliques sont plus importantes, la fixation ou la disparition d'un allèle en est donc facilitée.

Imaginons une population sans sélection : Cela veut-il dire que chaque individu aura le même succès reproducteur ? Non : Au départ, en se basant sur le génotype d'un individu, on ne peut prédire si son SR sera plus ou moins important que celui de son voisin. En d'autres termes, la probabilité de contribuer à la génération future sera la même pour tous les individus de la population.

Si les individus ont une probabilité de survie (ou de se reproduire) identique, ça ne veut pas dire que leur contribution dans la réalité sera identique.

Pour suivre l'évolution des fréquences alléliques dans la population, on utilise un modèle (=représentation mathématique de la réalité qui se concentre sur certaines caractéristiques importantes, ici l'évolution des fréquences alléliques).

Soit une population haploïde (reproduction asexuée) de dix individus. On a au départ 2 allèles, avec une présence de 50% chacun. Le modèle consistera à créer la génération suivante. Disons que la taille de la population est fixe. Comment peut-on simuler la création de la nouvelle génération à partir de la génération ancestrale ? Sachant qu'au départ, chaque individu a la même probabilité de survie que de se reproduire (même si, comme on l'a dit précédemment le succès reproducteur réel ne sera pas identique chez tout le monde).

Modèle du dé (aléatoire) : À la génération suivante, même si chaque copie a une probabilité égale de contribuer à la génération suivante, dans les faits, certaines sont plus représentées que d'autres. (Certains se font attaquer par un prédateur, d'autres par un virus, dans le cas de pop° avec reproduction sexuée p.e certains ne rencontrent pas de partenaires.... Tous ces éléments font que, même si au départ chaque individu avait un SR identique, au final seul une des copies du gène sera favorisée aléatoirement).

Un programme, *AlleleA1*, permet de modéliser ce phénomène. Pour une population infinie, sur 500 allèles, la fréquence allélique n'évolue (varie) pas. Avec une population finie, ça fluctue, et au plus la population est petite, au plus les fluctuations sont importantes.

Autre constatation : La **dérive génique** est une force génétique qui compte, puisqu'elle fait évoluer la fréquence allélique dans la population. Contrairement à la sélection qui va dans une direction précise, la dérive génique n'a pas une voie prédéterminée : Elle peut très bien fixer ou faire disparaître un allèle, aléatoirement.

Avec la dérive génique, on va vers une perte de diversité car, en attendant suffisamment longtemps, un des allèles va nécessairement disparaître. Un type de sélection permet d'augmenter la diversité génétique, contrecarrer l'action de la dérive génique : La sélection en faveur de l'hétérozygote.

Imaginons que le génotype hétérozygote soit favorisé par la sélection : Malgré une dérive génique importante (taille de la population petite), après 50 générations on a une fréquence allélique proche de la valeur initiale.

Sélection en faveur de l'hétérozygote et dérive génique sont deux forces qui s'opposent.

La dérive génique :

- Fluctuation des fréquences alléliques dans une population, due à la variation aléatoire du succès reproducteur des individus
- est d'autant plus forte que la population est petite
- par définition, n'oriente pas l'évolution dans une direction déterminée

Sur la slide des deux graphiques :

Phénomène connu dans les populations naturelles : Goulot d'étranglement ou effet fondateur. Une catastrophe écologique quelconque fait que la taille de la population diminue brusquement. Petit à petit, la taille retrouve sa valeur originelle. La fluctuation redevient égale à celle que l'on avait au départ.

Pour une population petite, la fluctuation est tellement forte qu'au final on a 4 populations très différentes du point de vue des fréquences alléliques. Un seul événement peut provoquer une évolution très importante de la population qui, une fois enregistrée, est définitive.

Imaginons une île qui apparaît dans l'océan. Si quelques individus provenant de l'océan colonisent l'île, on aura au départ une fluctuation importante des fréquences alléliques puis, une stabilisation grandissante en même temps que la population.

Exercice

Un allèle a une fréquence de 0,01 dans une population. Si la dérive génique est la seule force évolutive considérée, quelle est la probabilité que cet allèle finira par disparaître de la population, si la taille de cette population est de 50 individus, et qu'on fait l'hypothèse d'un accouplement aléatoire ?

On ne peut pas prédire le sens de la fluctuation puisque au départ, chaque allèle a une probabilité égale de se produire. La probabilité que l'allèle finisse par disparaître est uniquement dépend de sa fréquence de départ. En l'occurrence, elle est de 99% (car la fréquence initiale est de 0,01 soit 1%).

Exercice bis

Même question mais cette fois-ci avec une population de 50,000,000 individus. Ça ne change rien : Le raisonnement de tout à l'heure est juste basé sur la fréquence allélique initiale : La probabilité que l'allèle disparaît est toujours de 99%. Ce qui change c'est la vitesse à laquelle il disparaît... Dans le cas présent, ça ira plus lentement qu'avec une population de 50 individus : La dérive génique est plus faible dans les grandes populations.

Exercice 2

Dans les études qui analysent la variation génétique au sein des populations naturelles, on commence souvent par vérifier que les différents locus pour lesquels on a des données présentent des fréquences génotypiques à l'équilibre de H-W. Ceci pour vérifier si un ou plusieurs locus analysé(s) est sous sélection.

Pourtant, les populations naturelles étudiées présentent par définition des tailles finies (et non pas infinies). Une taille de population infinie étant une hypothèse du modèle de H-W, est-il légitime de s'attendre à observer des fréquences génotypiques à l'équilibre de H-W, en l'absence de sélection ?

Il suffit que les accouplements soient aléatoires. Si c'est le cas, les fréquences génotypiques correspondront à celles attendues sous HW.

Question d'examen

Chez la drosophile, le gène brown est impliqué dans la coloration de l'oeil. En 1956, Buri forme 107 populations composées chacune de 8 mâles et 8 femelles, tous hétérozygotes bw/bw⁷⁵. Dans chaque population, à chaque génération pendant 19 générations, Buri prélève au hasard 8 mâles et 8 femelles pour former la génération suivante, et compte le nombre d'allèles bw⁷⁵ qu'ils contiennent (les 3 génotypes possibles donnent 3 phénotypes différents). Il obtient les résultats suivants :

Voir graphique TP2

- a) Sur le graphique, que représente chacune des barres d'histogramme ?
- b) Que constatez-vous au cours des générations successives ?
- c) Comment expliquez-vous ces résultats ? Justifiez votre réponse
- d) Quel(s) résultat(s) pouvez-vous prédire si l'expérience est poursuivie bien au-delà de 19 générations ?

- a) Le nombre de populations qui possède un nombre d'allèles donné.
- b) Au cours des générations, le nombre de populations qui avait soit aucun allèle bw⁷⁵, soit 32 allèles bw⁷⁵ augmente. Au cours du temps, dans plus en plus de populations on observe la fixation de 1 des 2 allèles. La diversité génétique diminue au cours du temps : Au début, la majorité des populations ont une diversité maximale. Au fur et à mesure, on tend vers plus de populations qui ont soit aucun allèle, soit 32 allèles.

- c) Explication : Dérive génique. On s'attend à l'observer car il s'agit de petites populations. Le prélèvement au hasard sous-entend qu'il n'y a pas de sélection dirigée dans un sens. Autres caractéristiques de la dérive génique : La fluctuation forte des fréquences alléliques, appauvrissement de la diversité génétique.
- d) Toutes les populations seront soit dans un cas de fixation de l'allèle, soit dans un cas de disparition des allèles (si on attend suffisamment longtemps)

15 octobre 2015

DESEQUILIBRE DE LA LIAISON

On va continuer notre exploration de l'évolution des fréquences alléliques au sein des populations.

On va passer à un **3^e écart au modèle H-W : passage de un à deux locus.**

On part d'un exemple : concerne une espèce de papillon d'Amérique du Sud. *Heliconus numata* (?)

- Présente une grande variation de motifs sur leurs ailes. Chacun des dessins que l'on retrouve chez *Heliconus* se retrouve chez une autre espèce : *Melanarnia* (?)
- Comment expliquer cette information ?
→ Mimétisme : cette espèce génère des formes qui tentent d'imiter les dessins qu'on observe chez une autre espèce. Ces différentes espèces de *Melanarnia* produisent des toxines pour faire fuir les prédateurs. A chaque fois qu'un oiseau mange un individu de ces espèces, il va recracher l'insecte car le goût est mauvais, et va apprendre petit à petit à mémoriser le dessin pour éviter d'en manger la prochaine fois. *Heliconus* va donc imiter l'une des formes "toxiques" par les oiseaux de sorte qu'elle ne se fasse pas manger non plus, sorte de protection.
- Au niveau génétique, quel est le mécanisme qui permet au sein de l'espèce l'observation, l'apparition de ces différentes formes proches de celles de *Melanarnia* ?
*À souligner : il n'y a pas de formes intermédiaires, que les différentes formes représentées sur la slide.

On peut imaginer que par sélection, seules ses formes là survivent au détriment des autres formes intermédiaires. Mais on n'observe jamais d'intermédiaires (même pas en labo !).

Les premiers bios qui ont observé ce phénomène pensent que ça doit être nécessairement influencé par différents gènes, car pour expliquer ces différences importantes, un seul gène ne pourrait être le seul impliqué dans le mécanisme. On parle alors de **supergène**, qui n'est autre qu'un gène présentant un ensemble d'autres gènes.

- Une étude plus récente montre que ces patterns étaient déterminés par 18 gènes différents qui influencent la forme de ces papillons. Mais comment expliquer que les hybrides ne présentent pas de dessin intermédiaire ?

Imaginons qu'on ait ces 18 gènes, localisés sur le même chromosome, et imaginons que pour la première forme on ait une série d'allèle A et dans le deuxième individu on ait une série d'allèle B. Qu'est-ce qui se passerait si on avait un hybride entre les deux formes ? On aurait un hétérozygote pour tous les différents locus : AB, AB, AB... Alors pourquoi on obtiendrait uniquement une des formes de départ ? C'est dû à la dominance.

- Y a-t-il un moyen de formation d'un hybride malgré la dominance d'un allèle ?
Si les gènes étaient localisés sur des chromosomes différents, il n'y a plus d'explication du crossing-over. À la deuxième génération, soit les hétérozygotes se reproduisent entre eux, soit les hétérozygotes se reproduisent avec les formes parentales homozygote A ou B.
→ Si tous les gènes se retrouvent sur le même chromosome, des processus de recombinaison qui jouent. Et si c'est le cas, il y aura production d'un allèle qui combine à la fois des a et des b. on devrait alors s'attendre à la deuxième génération à des hybrides. Pkoi cela ne se présente pas ?
- En fait on s'est rendu compte que la région chromosomique concernée présente les 18 gènes

présentait une inversion. A partir du moment où il y a inversion, le phénomène du crossing-over ne peut plus se faire durant la méiose. Permet de maintenir la cohérence et ces différentes formes au sein de la population naturelle. (sur la slide H23 et H40 en bas. Alors que dans le segment pour cet allèle se sera inversé)

Par PCR on a obtenu un séquençage et amplification du chromosome et donc détermination du gène concerné, protégé de la recombinaison.

Autre exemple de papillon : *Papilio polites*, espèce asiatique. En haut sur la slide, forme normale.

- Les femelles présentent elles aussi une forme de mimétisme envers une autre espèce. Mais au contraire de l'exemple précédent, toute cette variation observée est déterminée par un seul gène, au lieu de 18 comme dans le cas précédent. Ce gène est extrêmement variable. Des variations nucléotidiques permettent l'apparition de ces différentes formes.
- Les formes mimétiques sont protégées elles aussi de la recombinaison par inversion chromosomique.

Si deux locus sont présents sur un même chromosome, ils vont être liés. Cette liaison peut être séparée par des événements de recombinaison ou pas, selon les cas.

Exercice : tableau qui représente les fréquences des 9 génotypes possibles d'un locus.

- ! La pop° est-elle à l'équilibre de Hardy Weinberg pour chacun des deux locus, a et b ?
 $F(AA) = 0,25$ $F(Aa) = 0,50$ $F(aa) = 0,25$
 $F(A) = 0,25 + (0,50/2) = 0,50$
 F attend sous HW (AA) = $p^2 = 0,5^2 = 0,25$ donc on est bien à l'équilibre de HW.

Et pour le deuxième locus : même raisonnement, on est bien à l'équilibre HW.

- ! Maintenant la question qui se pose est la suivante : la présence de l'allèle a ou A au locus 1 chez un individu dépend-elle de la présence de l'allèle B ou b au locus 2 ? En d'autres termes, la transmission du locus 1 à la descendance est-elle dépendante de la transmission du locus 2 (sont-ils 'liés' ou indépendants) ?
 Dans le cas d'une transmission indépendante, ils seraient sur des chromosomes différents. Hors si on observe une association entre les deux allèles, une hypothèse possible serait qu'ils soient sur le même chromosome.
- ! Comment tester que deux allèles sont sur le même chromosome ? (on doit réfléchir via les fréquences attendues si les deux locus étaient non associés, ensuite les comparer aux fréquences sur la slide).
 - S'ils sont indépendants, on aurait une $F(AA/BB)$ attendue = $0,25 \times 0,64 = 0,16$
 Hors F observée = $0,20$.
 - $F(Aa/BB) = 0,50 \times 0,64 = 0,32$
 Hors dans le tableau F observée = $0,40$.
 - On en déduit qu'il y a donc association entre les deux locus, leur transmission n'est donc pas totalement indépendante.
- ! Voyons maintenant comment les fréquences gamétiques peuvent évoluer dans la population à pd différents cas. On va de nouveau partir du cas de deux locus avec deux allèles.
 - Les gamètes possibles : AB Ab aB ab
 - Fréquences des génotypes : g(AB) g(Ab) g(aB) g(ab)
 - À pd ces différents gamètes il y aura formation de zygotes, plus précisément il y aura formation de 10 types de zygotes : AB/AB, AB/Ab, AB/aB, AB/ab, Ab/Ab, Ab/aB, Ab/ab, aB/aB, aB/ab, ab/ab.

- Fréquences des zygotes étant donné les fréquences gamétiques de départ, sachant les accouplements aléatoires :
Pour AB/AB : $g^2(AB)$
Pour AB/Ab : $2 [g(AB) \times g(Ab)]$ etc

À la génération suivant on va reformer les gamètes de départ :

AB Ab aB ab

Quelles sont les différentes possibilités pour former des gamètes AB à partir des zygotes ?

- Tous les gamètes produits par le zygote AB/AB seront de type AB. Fréquence = 1
- Les gamètes produits par le zygote AB/Ab seront 50% de type AB et 50% de type Ab. Fréquence = 0,5
- Les gamètes produits par le zygote AB/aB seront 50% de type AB et 50% de type aB.
- Etc...

On va considérer 3 cas différents :

- 1) Les deux locus sont sur deux chromosomes différents. Quels sont donc les autres zygotes contribuant à la formation de AB ? Tous ceux ayant un A et un B dans leur génotype, car les chromosomes sont différents, alors leur transmission se fera de manière indépendante.

AB/ab, étant double hétérozygote (pour chaque allèle), va former les gamètes $\frac{1}{4}$ AB, $\frac{1}{4}$ ab, $\frac{1}{4}$ Ab, $\frac{1}{4}$ aB.

Et donc à la génération suivante :

$$G'(AB) = g^2(AB) + g(AB)g(Ab) + \frac{1}{2} g(AB)g(aB) + \frac{1}{2} g(AB)g(ab) + \frac{1}{2} g(Ab)g(aB).$$

- 2) Les deux locus sont sur le même chromosome, mais il n'y a pas de recombinaison. Ils seront donc transmis avec moins de probabilité.

AB/ab formera alors des AB et des ab, mais pas d'intermédiaires car pas de recombinaison.

Et donc :

$$G'(AB) = g^2(AB) + g(AB)g(Ab) + g(AB)g(aB) + \mathbf{g(AB)g(ab)}.$$

La contribution des génotypes formant AB est donc réduite par rapport au premier cas.

- 3) Deux locus sur le même chromosome, recombinaison possible.

AB/ab, comme dans le premier cas va former les gamètes AB, ab, Ab, aB. Mais leur c'est au niveau de leur fréquence que ça va changer.

Tout dépend du **taux de recombinaison**, si les deux allèles sont éloignés sur le chromosome, la possibilité de recombinaison sera plus grande que s'ils étaient proches.

$$G'(AB) = g^2(AB) + g(AB)g(Ab) + g(AB)g(aB) + \mathbf{\frac{1}{2}(1-r) 2g(AB)g(ab)} + \mathbf{\frac{1}{2} r 2g(Ab)g(aB)}$$

r étant le taux ou la probabilité de recombinaison

Imaginons maintenant le cas suivant : un locus avec deux allèles (A,a) dont les fréquences sont représentées sur la slide (30% et 70%) et un autre locus avec deux allèles (B,b) dont les fréquences sont de 40% et 60%.

Quelle sera la fréquence attendue du gamète AB en l'absence d'association entre les deux locus.

(Locus sur chromosomes différents, aucun événement particulier qui provoquerait la recombinaison des allèles)

- ➔ Il suffit de faire le produit de la fréquence de chacun des allèles. Donc, en l'occurrence, $f(A) \times f(B)$, ce qui nous donne une fréquence attendue du gamète AB de 42%.

Imaginons que la fréquence observée soit de 50%, on peut comparer ces deux fréquences et la différence entre les deux c'est ce que l'on appelle le **DESEQUILIBRE DE LIAISON**.

Cette indice de déséquilibre de liaison **D** sera égale à 0 s'il n'y a pas d'association entre les deux locus (= s'ils sont transmis de façon indépendante).

Si au contraire il y a une association entre les deux locus, la valeur sera différente de 0.

QST : quelle sera la valeur maximale possible pour **D** ?

➔ Dans le cas de $g(AB) = 0,50$

Alors $f(A) = 0,50$ et $f(B) = 0,50$. Ce qui nous donne un $D = 0,25$. Ceci montre une association maximum.

16 octobre 2015 – DESEQUILIBRE DE LIAISON (2) - ACCOUPLEMENT ALEATOIRE

→ D peut-il évoluer au cours du temps dans une population ?

S'il y a un phénomène de crossing-over, oui. La recombinaison permet de casser l'association. Ainsi, D peut évoluer au cours du temps. Si les deux locus sont sur le même chromosome, au plus ils sont proches, au plus la probabilité de recombinaison sera faible, et au plus la vitesse d'évolution de D sera lente.

Sur la slide : Évolution du D au cours du temps en fonction du taux de recombinaison. Si r est faible (1 pour 1000 ici), le déséquilibre de liaison évolue lentement. S'il est maximum, c'est l'inverse.

$D_t = D_0 (1 - r)^t$ décrit l'évolution de D au cours du temps.

Cette propriété est utilisée pour cartographier les gènes : Afin de déterminer la position d'un locus par rapport à un autre sur un chromosome, on regarde leur degré d'association. Au plus ce degré d'association est important, au plus les deux locus sont proches et vice-versa. Cette technique fonctionne si la distance relative entre deux locus est relativement élevée... Si elle est faible, l'essentiel du déséquilibre de liaison est déterminé par le déséquilibre initial.

Si les locus sont proches au point que la probabilité de recombinaison est nulle, leur distance relative n'est plus proportionnelle au déséquilibre de liaison qui les sépare mais proportionnelle à d'autres paramètres.

Exercice

Considérons une population qui présente les fréquences gamétiques suivantes à une paire de locus : AB 0.4 ; Ab 0.3 ; aB 0.2 ; ab 0.1.

(a) Calculez le déséquilibre de liaison D

(b) Supposons que le taux de recombinaison entre les deux locus est de 0.2. Quel sera le D à la génération suivante si l'accouplement est aléatoire ? Et à la deuxième génération ?

(c) Refaire le même calcul que dans (b), mais avec un taux de recombinaison de 0.01.

(a) Calculer les fréquences alléliques

$f(A) = \text{Somme des fréquences des gamètes qui comprennent un A} = 0,4 + 0,3 = 0,7 = 70\%$

$f(B) = 0,4 + 0,2 = 0,6 = 60\%$

Ensuite, on calcule $D = g_{AB} - [f(A) \times f(B)] = 0,4 - [0,7 \times 0,6] = -0,02$

g_{AB} = Fréquence gamétique. Cette formule : Voir cours du 15 octobre.

(b) $D_t = D_0 (1 - r)^t = -0,02 - (1 - 0,02)^1 = -0,016$

* $t = 1$ car c'est la génération suivante.

À la génération encore suivante : remplacer D_0 par celui qu'on vient de calculer, donc -0,016.

(c) Il suffit de remplacer r par 0,01

Exercice 2

Considérons deux locus à deux allèles chacun (A, a ; B, b), caractérisés par un taux de recombinaison de 0.3. La population 1 a une fréquence de A de 0.5 et une fréquence de B de 0.1, et ne présente aucun déséquilibre de liaison. La population 2 a une fréquence de A de 0.9 et une fréquence de B de 0.9 et ne présente également aucun déséquilibre de liaison. Ces deux populations sont mélangées, avec la nouvelle population constituée de 60% d'individus de la population 1 et de 40% d'individus de la population 2.

Quel est le déséquilibre de liaison dans la population mélangée initiale ?

Pour calculer D, on a besoin des fréquences gamétiques et alléliques correspondantes.

$$D = g_{AB} - [f(A) \times f(B)]$$

On a besoin de ces 3 valeurs dans la population finale.

On sait que 60% d'individus de la population 1 et 40% d'individus de la population 2 constituent la population finale. Dans la population 1, on sait qu'il y a 50% de A et que, dans la population 2 il y a 90% de A.

Pour calculer la fréquence de A dans la population finale, on fait :

$$f(A) = [0,5 \times 0,60] + [0,90 \times 0,40] = 0,66$$
$$f(B) = 0,42$$

On a f(A), f(B) et D (= 0). On peut trouver g_{AB} !

Pour la population 1

$$g_{AB} = D + [f(A) \times f(B)] = 0 + [0,5 \times 0,1] = 0,05$$

Pour la population 2

$$g_{AB} = 0,81$$

Avec les fréquences gamétiques dans les populations initiales, on peut calculer la fréquence gamétique dans la population mélangée, qui vaut :

$$g_{AB} = [0,60 \times 0,05] + [0,40 \times 0,81] = 0,35$$

Maintenant, on peut calculer D dans la population finale :

$$D = 0,077$$

Constatation : Au départ, 2 populations étaient à l'équilibre de liaison. Une fois mélangées, avant que les accouplements aient commencées, on obtient un déséquilibre de liaison qui est assez élevé, sachant qu'il y a un maximum de 0,25 ! Comment l'expliquer ?

➔ Association entre les deux locus à cause du mélange. Avec le mélange, les fréquences gamétiques ne correspondent plus aux fréquences alléliques.

Application : Études d'association en Amérique du Nord. Population mélangée : Populations

d'origine européenne et d'origine africaine réunies sur le même endroit. Les accouplements se font préférentiellement au sein de la même communauté. En mélangeant les deux populations, et en réalisant une étude d'association entre les locus, on observera des déséquilibres de liaison : C'est dû à l'hétérogénéité de la population.

Écart 4 au modèle de Hardy-Weinberg : Accouplements non-aléatoires

1. Accouplements préférentiels entre individus apparentés ou accouplements consanguins

Cas extrême de consanguinité : Autofécondation.

$$f(AA) = 0,25 ; f(Aa) = 0,50 ; f(aa) = 0,25$$

Comment vont évoluer les fréquences génotypiques au fil des générations ? Si deux hétérozygotes s'accouplent : 1 AA ; 2 Aa ; 1 aa. 50% d'hétérozygotes, 50% d'homozygotes. Bref : On réduit la proportion d'hétérozygotes de moitié à la génération suivante.

Les autres accouplements consanguins (AA avec AA; aa avec aa) ne produisent que des homozygotes.

Comment peut-on définir un coefficient de consanguinité ? Dans une population où les accouplements sont aléatoires, la fréquence des hétérozygotes est celle que l'on obtient à l'équilibre d'Hardy-Weinberg.

On compare la fréquence des hétérozygotes observée dans la population avec la fréquence des hétérozygotes attendue sous Hardy-Weinberg.

$$f = 1 - (H_{\text{obs}}/H_{\text{H-W}}) = \text{Coefficient de consanguinité.}$$

Ex : Une communauté protestante vivant dans les Alpes a colonisé l'Amérique du Nord à la fin du 19ème siècle. Elle est restée relativement isolée. Au départ donc, on observe des accouplements consanguins **mais** les règles religieuses leur empêchent de se marier entre cousins proches.

Résultat :

- $f = -0,0615$ négatif (obtenu via la formule ci-dessus).
- $F = 0,0255$ **très élevé**

Deux indices, chacun mesurant la consanguinité d'une manière différente. Comment c'est possible d'avoir un f négatif et si oui, dans quelles conditions ? Quand on s'accouple avec un proche, on a une perte de l'hétérozygotie, donc le Hobs diminue par rapport au H attendu sous H-W. Mais ici y a la règle religieuse qui fait que l'accouplement est plus qu'aléatoire, car on ne peut s'accoupler avec des parents proches, hors dans la vraie vie il arrive parfois que, sans le savoir, on s'accouple avec des individus "relativement" proches. C'est ce qui explique le Hobs plus grand que H sous Hardy Weinberg et que donc f est négatif.

Pour calculer F , on se base sur un pedigree qui montre les relations généalogiques entre les individus. Dans ce type de graphes, les carrés représentent les hommes et les ronds représentent les femmes.

Voir graphe

Dans la 1ère génération du pedigree, on considère que tous les allèles sont différents : Bb ; Aa ; Cc. On calcule la probabilité que l'individu en F2 soit homozygote pour un locus, autrement dit qu'il y ait deux fois le même allèle chez cet individu.

La probabilité doit être considérée pour la femelle, mère des deux individus qui s'accoupleront en F1. Car pour les deux pères, la probabilité que l'individu en F2 ait deux fois un de leur allèle est nul (logiquement). Quelle est cette probabilité ? Du côté du 1er père, 1 chance sur 2. Pareil du côté du 2ème père.

Au final :

- $P(AA) = (1/2)^4 = 1/16$
- $P(aa) = (1/16)$

La probabilité que l'individu en F2 soit homozygote pour un locus est $1/16 * 1/16 = 1/8$

F est défini par rapport à un individu, pas par rapport à une population. Ceci dit, si on calcule F pour un échantillon d'individus de la population et qu'on fait la moyenne, on peut obtenir F_{moyen} qui ici vaut 0,0255.

On comprend mieux : F est très grand puisque la population est issue d'un petit nombre de migrants. En conséquence, il y a peu d'individus dans la 1ère génération du pedigree (tous proches les uns des autres). En conséquence, la probabilité au départ d'avoir des individus homozygotes est relativement élevée.

f est négatif car il mesure l'effet d'un accouplement aléatoire par rapport à l'effet d'un accouplement qui ne serait pas aléatoire.

Exercice : Quel est la probabilité que l'individu D soit homozygote, pour un locus localisé sur le chromosome X ?

Voir photos

Exercice : Si on fait l'hypothèse que toutes les relations généalogiques importantes sont présentées dans le pedigree ci-dessus, quel est le coefficient d'inbreeding F pour l'individu femelle colorié en noir ?

2. accouplements sélectifs – préférentiels entre phénotypes semblables

Au sein de certaines populations naturelles, on observe une attirance vers des individus de phénotypes semblables. Dans ce cas, l'accouplement est observé préférentiellement entre individus de phénotypes qui se ressemblent... C'est une préférence, mais ce n'est pas une nécessité.

Exemple : *Enchenopa sp.* Ils possèdent une 3ème paire d'ailes qui s'est transformée en « bouclier ». Chez certaines espèces de ce genre, on trouve une préférence pour différents types de plantes. Des individus se développent sur une plante-hôte, d'autres sur une autre plante-hôte, ce qui a un impact sur le cycle et la vitesse de développement de ces insectes.

Autrement dit, les insectes qui se développent sur la plante-hôte A arrivent à maturité à un moment différent de ceux qui se développent sur une plante-hôte B. Il y a donc un accouplement préférentiel entre les individus qui ont atteint la maturité sexuelle au même moment. Il n'y a pas une attirance particulière, c'est la concordance des cycles qui amène ce phénomène.

Quel impact sur les fréquences génotypiques de la population ? Diminution des hétérozygotes :

- AA avec AA donne des AA
- aa avec aa donne des aa
- Aa avec Aa donne 1/2 Aa, 1/4 AA et 1/4 aa

Comment évoluent les fréquences alléliques ? Elles ne changent pas... Quand on est dans un cas à 1 locus. Pour un cas à 2 locus, c'est différent.

22 octobre 2015

Éléments de microévolution

Voir slides

Objectifs susceptibles d'être évalués à l'examen

Voir slides

/!\ Objectifs les plus importants /!

- Prédire, y compris dans un cas concret défini, l'influence de chaque facteur exploré dans le cours (mutation, sélection, dérive génique, déséquilibre de liaison, différents types d'accouplements non aléatoires), séparément ou de concert (plusieurs facteurs combinés), sur l'évolution des fréquences alléliques, génotypiques et gamétiques (cas de >1 locus) d'une population
- À partir de fréquences alléliques/gamétiques observées, dans un contexte défini, identifier les hypothèses évolutives compatibles/incompatibles (en d'autres termes, quels sont les facteurs d'évolution qui peuvent expliquer les données)

2. accouplements sélectifs – préférentiels entre phénotypes semblables

[Rappel] Impact sur les fréquences génotypiques : réduction de la fréquence des hétérozygotes dans la population, cependant les fréquences alléliques ne changent pas au cours du temps comme dans l'accouplement consanguin.

Exercice

Locus 1 (A,a) Locus 2 (B,b)

Contribution de A et B au phénotype : +1 pour chacun

Contribution de a et b au phénotype : 0 (n'apportent rien au phénotype)

AB/AB aura un phénotype de +4.

Ab/ab = 0

Dans le cas d'un accouplement 100% sélectif, imaginons un individu AB/AB. Avec quel autre type d'individu peut-il alors s'accoupler pour porter le même phénotype au final ?

- Avec AB/AB uniquement. Tous les croisements entre génotypes de type 4, vont produire les mêmes génotypes.
- Même choses pour ab/ab, ils devront s'accoupler avec des ab/ab, et les générations suivantes auront le même génotype que leurs parents.

Si on dit maintenant que :

Ab/Ab = 2

Ab/Ab x Ab/Ab --> génotypes fils possibles pour cet accouplement?

➔ Ab/Ab

Avec ce type de croisement, les générations ne peuvent être que génétiquement identiques aux parents.

Ab/Ab, aB/aB, Ab/aB, AB/ab on des génotypes +2 et peuvent alors s'accoupler entre-eux.

Voyons par exemple : AB/ab x AB/ab (individu 100% hétérozygote, donc hétérozygote pour les deux locus), quels seront les génotypes produits, dans le cas où les locus sont sur des chromosomes sont différents ?

- $[1/2 \times (1-r)]^2$ de AB/ab (au carré, car deux possibilités de recombinaison : AB avec ab et ab avec AB). Il aura un phénotype **2**
- $1/4 \times (1-r)^2$ de AB/AB de phénotype **4**
- si recombinaison -> product° de Ab/aB avec une proportion de $(1/4 + 1/4)r^2$, de phénotype **2**

- $Ab/Ab : \frac{1}{4} r^2$ de phénotype 2
- $Ab/Ab : \frac{1}{2} (1-r)r$ de phénotype 3

On peut voir qu'un croisement entre deux génotypes de phénotype +2 peut produire des génotypes qui correspondent à des classes phénotypiques différentes.

Par contre, un génotype AB/AB ou ab/ab ne peut se croiser qu'avec un même génotype, produisant des homozygotes qui ne pourront se croiser qu'avec des génotypes homozygotes (=une fois qu'on a des homozygotes pour les deux locus, on ne produit que des homozygotes).

Hors, si on croise des hétérozygotes (p.e $AB/ab \times AB/ab$), on produit un mélange d'homozygotes et hétérozygotes. Si on continue ce croisement pendant un grand nombre d'année, la proportion d'hétérozygotes va diminuer de façon progressive et celle des homozygotes va augmenter, cela va dépendre de la force de l'accouplement sélectif.

En d'autres termes, si on attend suffisamment longtemps on va avoir que des AB/AB , ab/ab et éventuellement mais pas nécessairement des aB/aB ou Ab/Ab . Pas les deux en même temps car l'un peut s'accoupler avec l'autre et produire des génotypes différents des génotypes parents, se "neutraliser" et disparaître au final.

Impact de ce type d'accouplement dans le déséquilibre de liaison entre les deux locus ?

- La liaison sera importante
- la fréquence des gamètes AB et ab sera maximum dans la pop° : un allèle A sera associé qu'avec B , a avec b . Ceci que si les locus sont sur les deux chromosomes différents.
- Que ce soit sur des chromosomes différents ou pas, dans un cas d'accouplement préférentiel entre phénotypes semblables (accouplement 100% sélectif), le déséquilibre de liaison sera maximal.

Exercice

Considérons maintenant deux locus à deux allèles chacun ($A, a ; B, b$), caractérisés par un taux de recombinaison de 0.3. La population 1 a une fréquence de A de 0.5 et une fréquence de B de 0.1, et ne présente aucun déséquilibre de liaison. La population 2 a une fréquence de A de 0.9 et une fréquence de B de 0.9 et ne présente également aucun déséquilibre de liaison. Ces deux populations sont mélangées, avec la nouvelle population constituée de 60% d'individus de la population 1 et de 40% d'individus de la population 2.

- 1) Quel est le déséquilibre de liaison dans la population mélangée initiale ? (-0.02, voir avant)
- 2) Supposons que la population 1 parle une langue différente de celle de la population 2, et qu'il y a accouplement uniquement entre individus parlant la même langue (100% accouplement sélectif). Quel est le déséquilibre de liaison de la population mélangée après 1 génération d'accouplement sélectif ?
- 3) Supposons au contraire que les accouplements se produisent de façon aléatoire dans cette population mélangée. Quel est le déséquilibre de liaison après une génération d'accouplement aléatoire ?

2) L'accouplement sélectif augmente le déséquilibre de liaison entre deux locus... Mais celui caractérisant des locus définissant le phénotype sur lequel les individus se basent pour choisir leur partenaire sexuel. /\

On a vu une formule pour calculer l'évolution du déséquilibre de liaison au cours du temps... **Dans le cas d'un accouplement aléatoire**, ce qui n'est pas le cas ici ! On ne peut donc pas l'utiliser.

En faisant l'hypothèse que l'évolution des fréquences alléliques change très peu en une génération, le déséquilibre n'évolue pas car continue à refléter le mélange des populations isolées l'une de

l'autre. La réponse est donc 0,077.

Si on a ici un déséquilibre de liaison c'est parce que les fréquences alléliques des gamètes sont différentes.

3) On considère ici que la langue ne joue aucun rôle. On utilise la formule qui décrit l'évolution du déséquilibre de liaison : $D_t = D_0 (1-r)^t = 0,077 (1-0,3) = 0,054$

Exemple concret : Phénotype de la surdité profonde. Ce phénotype est influencé par leur génome : Plusieurs locus jouent. En général, ce sont des allèles récessifs qui sont responsables de cette maladie. Ces derniers ont une fréquence faible dans la population... Sauf 1 qui a une fréquence plus élevée : Le gène GJB2 avec une fréquence de 0,01.

Dans le cas d'un accouplement aléatoire, quelle serait la fréquence attendue des individus sourds dans la population ? On s'attend à une fréquence de $P^2 = (0,01)^2 = 10^{-4}$... Or dans les populations européennes et nord-américaines, ces valeurs sont 5 à 100 fois plus élevées ! **On l'explique par des accouplements préférentiels entre individus de phénotype semblable.**

En effet, les individus sourds communiquent via le langage des signes. Les sourds auront plus facile à communiquer entre-eux qu'avec des personnes « normales » qui ne connaissent pas le langage des signes. De plus, on les place dans des écoles et instituts qui leur donnent un suivi adapté à leur handicap, où elles rentrent en contact avec d'autres personnes qui sont aussi sourdes.

Comparaison accouplement sélectif – accouplement consanguins ?

L'accouplement sélectif augmente le déséquilibre de liaison uniquement entre les locus qui jouent un rôle dans la reconnaissance du partenaire (au moins deux locus doivent être impliqués). Dans le cas de l'accouplement consanguin, ça n'influe pas le déséquilibre de liaison car le choix du partenaire ne se fait pas par la reconnaissance d'un phénotype particulier. Simplement, il y a un accouplement préférentiel entre les individus plus apparentés... Mais on ne sélectionne pas certains allèles par rapport à d'autres.

Dans le cas d'un accouplement sélectif, on a une diminution des hétérozygotes uniquement pour les locus jouant un rôle dans la définition du phénotype. Pour tous les autres locus dans le génome, il n'y a aucun impact !

Par contre dans le cas d'un accouplement consanguin, tous les locus du génome sont concernés.

!/ Différence fondamentale : Pour l'accouplement sélectif, diminution de la proportion d'hétérozygote pour les locus influençant le phénotype d'intérêt. Pour l'accouplement consanguin, puisqu'on s'accouple avec une personne apparentée, on a une diminution du taux d'hétérozygote à travers tout le génome !/

Voir photos

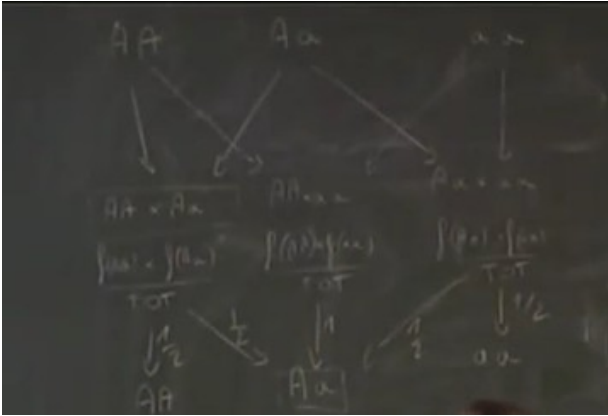
3. accouplements sélectifs – préférentiels entre phénotypes différents

On va se limiter à un locus, à deux allèles : A et a

AA Aa aa

Les individus de même génotype et donc de même phénotype ne s'accoupleront pas entre-eux. On aura donc que 3 accouplements possibles (voir notes du prof). La fréquence des descendants sera le produit des génotypes concernés.

(à compléter avec notes du prof)



On constate une augmentation des hétérozygotes dans la population !

La fréquence des allèles va évoluer au cours du temps, car on va augmenter les hétérozygotes.

Exemple : Complexe majeur d'histocompatibilité MH6. C'est un complexe d'une centaine de gènes présents sur le même chromosome. Notamment chez la souris on observe des variations génétiques associées à des odeurs différentes. On a démontré que les souris mâles étaient attirées par des souris femelles, ayant une odeur différente de la leur, associé à une variation génétique différente : variation forte au niveau des gènes MH6.

Il a été démontré qu'un système semblable existait chez l'Homme. C'est pour ça qu'on observe chez nous un polymorphisme au niveau des gènes de ce complexe plus marqué.

Dans la population humaine, la diversité génétique est très élevée pour ce locus et a été maintenue pendant une période de temps très longue à cause de ce phénomène d'accouplement sélectif. C'est pour ça qu'il n'y a pas eu de dérive génique éliminant plusieurs lignées et érigeant une seule lignée, comme ça a été le cas chez les primates par exemple.

Exercice

Considérons un dème génotypé pour 3 locus choisis de façon aléatoire. Les fréquences des génotypes sont reprises dans le tableau suivant :

Locus i	$A_i A_i$	$A_i a_i$	$A_i a_i$
1	0.09	0.42	0.49
2	0.36	0.48	0.16
3	111	378	0.511

Voir slides

/!\ erreur fréquente : Tant qu'on ne sait pas si on est dans un cas d'Hardy-Weinberg, on ne peut calculer la fréquence allélique par les formules connues.

23 octobre 2015

Suite de l'exercice de la fois dernière

Hypothèse d'absence de sélection ou absence de dérive génique.

Par vérification, le premier locus est à l'équilibre de H-W. Le troisième ne l'est pas, car la fréquence attendue est différente de celle calculée et la fréquence attendue des hétérozygotes est de 0,42, hors la fréquence observée est plus faible -> on conclut qu'il y a un déficit en hétérozygotes. Comment expliquer ça sachant qu'il n'y a pas de sélection ni de dérive génique ?

- Hypothèse de la consanguinité, qui a pour effet de diminuer la fréquence des hétérozygotes. Mais on devrait observer cet effet non seulement sur un locus mais sur la plupart du génome. Hors ce n'est pas le cas ici.
- Le 3ème locus est soumis à un accouplement préférentiel qui s'effectue entre individus de phénotypes semblables. Pour que l'observation soit compatible avec cette hypothèse il faudrait vérifier si le locus en question contribue au phénotype et influence le choix préférentiel des individus dans la pop°.

Est-ce que la sélection aurait pu expliquer ce phénomène ? (voir cours De Biseau)

Si l'hétérozygote a une probabilité de survie moindre des homozygotes dans le contexte environnemental, on pourrait être dans un cas de sélection contre l'hétérozygote qui va faire diminuer sa fréquence par rapport à celle attendue sous H-W.

Si le locus 4 est très proche du locus 3 sur le même chromosome, l'influence de la sélection va se voir sur les deux locus même si un des deux n'influence pas le phénotype.

-récapitulatif des forces évolutives- (slide)

* mutations : aléatoires, contribuent à la diversité génétique

* dérive génique : aléatoire, diminue la diversité génétique en faisant fluctuer les fréquences des allèles (un allèle pourrait arriver à une fréquence=0 et disparaître)

*sélection : évolution dirigée, diminution de la diversité génétique dans des pop° restreintes.

Dans la nature, ces forces sont combinées entre-elles et provoquent des changements différents sur la diversité génétique et fréquence allélique.

Évolution au sein des populations

Que se passe-t-il dans le cas d'une sélection en faveur d'un hétérozygote ? On tend vers une fréquence de 50%. Avec deux allèles dans une population, la diversité est plus grande s'ils ont chacun une fréquence de 50% que s'ils ont une fréquence inégale.

Dépend de la définition de la fréquence génétique que l'on prend.

Dans ce cas, la sélection peut diminuer la diversité génétique en favorisant les hétérozygotes. Dans certains cas de sélection en faveur des hétérozygotes elle égalise les fréquences alléliques des différents allèles au sein de la population et donc augmente la diversité génétique.

Accouplement consanguin et accouplement *spécifique* (vérifier) : Est-ce que les accouplements consanguins vont faire fluctuer les fréquences alléliques ? Non, donc pas d'altération de la diversité génétique car les fréquences alléliques ne bougent pas.

Accouplement préférentiel entre phénotypes différents ou dissemblables : Augmentation de la diversité génétique dans les populations. En effet : Augmentation de la fréquence des hétérozygotes, les fréquences alléliques tendent à aller vers 50%

Dérive génique et déséquilibre de liaison

Avec une fréquence allélique de 80% dans une population, pour une taille déterminée, la fréquence à la génération suivante fluctue. En prenant un autre allèle qui a une fréquence de 2%, on observera aussi une fluctuation d'ici la génération suivante mais relativement plus importante.

L'impact de la dérive génique est relatif à la fréquence de l'allèle de départ.

Soient deux locus :

- L₁ A(0,30) a(0,70)
- L₂ B(0,20) b(0,80)
- L₃ C(0,50) c(0,50)

Pour un gamète ABC, sa fréquence sera la multiplication de la fréquence des différents allèles : $f(ABC) = 0,3 \times 0,3 \times 0,5 = 0,03$

La dérive génique fait aussi fluctuer les valeurs des fréquences gamétiques, qui sont beaucoup plus faibles que les fréquences alléliques. L'impact est d'autant plus grand.

La dérive génique peut engendrer un déséquilibre de liaison entre deux locus sans que ces locus ne soient liés physiquement sur un même chromosome.

Effet fondateur : Événement de migration et de colonisation, un petit groupe d'individus qui proviennent d'une autre pop^o arrive et colonise un milieu. La population prend ensuite de l'importance au cours du temps. On pense qu'un tel effet est à l'origine de la population en Sardaigne et une autre région d'Italie (voir slide).

Nous observons un déséquilibre de liaison très important entre deux locus, localisés sur le même chromosome. Simplement, ce déséquilibre observé dans ces deux populations est beaucoup plus élevé que celui observé dans le reste de la population humaine... Serait-ce à cause de l'effet fondateur ? Vu que le groupe qui fonde est très petit, la force de la dérive génique est très forte.

S'en suit un déséquilibre de liaison important entre deux gènes :

- Le gène G6PD : Code pour la glucose phosphate déshydrogénase. La déficience pour cette activité de l'enzyme provoque des maladies plus ou moins importantes.
- Un gène dont la déficience de certains allèles provoque le daltonisme (= difficulté de distinction de couleurs, surtout au niveau des mâles de la pop^o).

Quasiment aucun homme ayant une déficience en G6PD est daltonien... Dans la population considérée c'est l'inverse : Tous les hommes ayant une déficience pour cette enzyme sont daltoniens. C'est l'allèle provoquant le daltonisme qui est lié à l'allèle déficient de G6PD.

Deux locus caractérisés par un déséquilibre de liaison différent de 0 sont-ils nécessairement localisés sur un même chromosome ?

Quand deux locus sont proches sur un chromosome, ils ont une probabilité élevée de présenter un déséquilibre de liaison... Dans certains cas, même quand ils ne sont pas physiquement liés, on peut observer un déséquilibre. Quels sont ces cas ? Comment un déséquilibre peut-il apparaître dans une population entre deux locus qui sont sur des chromosomes différents ?

Dans le cas d'une population « unique » constituée de deux populations isolées. Autre cas : Un accouplement sélectif engendre un déséquilibre de liaison entre deux locus, en particulier entre individus de phénotypes semblables. Ce type d'accouplement, lorsqu'on considère uniquement un locus il y a peu d'effet sur les fréquences alléliques. Ce n'est qu'à partir de deux locus qu'un

déséquilibre est envisageable.

La dérive génique. On peut imaginer, dans un cas extrême, que cette dérive peut faire apparaître un déséquilibre de liaison entre deux locus sur des chromosomes différents.

Mutations : imaginons qu'on ait 50% de A et 50% de a. imaginons un autre locus sur le même chromosome un peu plus loin, avec un allèle D. Une mutation apparaît en un exemplaire, qui va être relié à un C ou c. Par exemple : c – d, avec faible séparation entre les deux, va donner un c qui va être automatiquement lié à un d, et donc apparition d'un déséquilibre de liaison par mutation. On peut calculer le temps pour casser ce déséquilibre via la formule suivante : $D_t = D_o (1-r)^t$

Sélection naturelle : Soient deux locus avec deux allèles. Si l'allèle A donne un avantage aux porteurs dans un type d'habitat. Dans un autre type d'habitat, c'est à qui donne l'avantage. Dans le 1er habitat, sélection en faveur de A, dans la 2ème sélection en faveur de a. Un déséquilibre peut s'installer même si les deux locus ne sont pas sur le même chromosome.

Exercice

Un crustacé parasite d'origine tropicale est introduit accidentellement dans une rivière en Espagne, il y a plus ou moins 1000 générations. Environ 500 générations après son introduction, une usine est installée en bordure de cette rivière. Elle y rejette certains déchets, mais réchauffe également l'eau de la rivière, car elle l'utilise pour un système de refroidissement. Depuis le début de son introduction, une équipe de biologistes échantillonnent régulièrement cette population, et estime la fréquence de différents allèles du même gène, choisis de façon aléatoire dans le génome. Le graphe ci-dessous indique l'évolution de la fréquence d'un allèle de ce gène au cours du temps.

Voir graphique slide

- 1) Que peut-on constater à partir de ce graphique (= décrire les données observées) ?
- 2) Comment peut-on interpréter ces observations ? Expliquer

1) Quand on s'approche de 600 générations, la fréquence de l'allèle devient plus stable au cours du temps.

2) Cette stabilisation s'observe juste après l'installation de l'usine. Les conditions deviennent plus bénéfiques pour la population : La température de l'eau augmente. Pour une espèce tropicale, c'est avantageux ! La taille de la population augmente. Ça stabilise les fréquences car la force de la dérive génique est plus faible quand la taille de la population est plus grande.

Exercice

On échantillonne 50 individus d'une population de dauphin dans l'océan atlantique. Le génotypage de ces individus pour 2 locus donne les observations suivantes (nombres d'individus de chaque génotype) :

- Locus 1 : 2 aa, 9 bb, 18 cc, 6 ab, 5 bc, 10 ac
- Locus 2 : 10 rr, 36 rs, 4 ss

a) Comment peut-on interpréter ces données ?

b) Si la transmission de ces deux locus à la génération suivante se fait de manière indépendante, à quelles fréquences des gamètes ar et cs doit-on s'attendre dans la population ? Détachez vos calculs, et justifiez vos réponses

a)

Locus 2 : Excès d'hétérozygotes

Locus 1 : Déficit d'homozygotes

Quels sont les facteurs d'évolution ayant abouti à ces fréquences ?

1ère hypothèse : On peut imaginer un double accouplement non-aléatoire, avec un accouplement préférentiel entre phénotypes semblables pour le locus 1, et un accouplement sélectif pour les phénotypes différents par rapport au locus 2. Dans ce cas-là, chacun des deux locus influence le phénotype permettant l'accouplement sélectif.

2ème hypothèse : Sélection différente qui s'applique sur les deux locus.

(3ème hypothèse : Liaison ? On doit connaître les fréquences génotypiques combinées, ou les fréquences gamétiques pour calculer D ... Ni l'une ni l'autre ne sont données dans l'énoncé)

4ème hypothèse : Mix des hypothèses 1 et 2

b)

Fréquence du gamète ar sans association entre les deux marqueurs : $0,56 \times 0,20 = 0,42$

Exercice

Sur une île de l'océan pacifique, une large population de doryphores (insecte coléoptère diploïde) provoque des ravages dans les cultures de pommes de terre. Un biologiste récolte un échantillon représentatif de cette population et génotype les individus échantillonnés au niveau de 10 locus, sélectionnés au hasard dans le génome. Il observe entre 4 et 8 allèles différents par locus, d'une fréquence supérieure à 5%.

L'année suivante, les agriculteurs répandent de larges quantités d'insecticides sur leurs cultures, plusieurs années de suite, pour tenter de solutionner le problème. La population de doryphores est fortement réduite. Dix ans plus tard, le biologiste récolte à nouveau un échantillon de la population de doryphores, de taille équivalente à l'échantillon initial, pour génotyper les mêmes locus. Pour 9 locus sur 10, le nombre d'allèles observés par locus (d'une fréquence supérieure à 5%) n'est plus que de 1 ou 2. Par contre, pour 1 locus, le nombre d'allèles observés est toujours égal à 8.

Comment peut-on expliquer ces observations ? Expliquer

La dérive génique devient forte et fait diminuer la diversité génétique : Des allèles disparaissent, expliquant la perte d'allèles à la plupart des locus. Conservation des 8 allèles au dernier locus ? Effet dû à un accouplement sélectif.

Deux hypothèses :

- Tous les locus sont sous l'effet de la sélection (sauf 1)
- Tous les locus sont sous l'effet de la dérive génique

Quelle hypothèse est la plus probable ? Celle de la dérive génique (la première est trop peu probable). On peut toutefois imaginer qu'un des locus soit sous sélection... Mais pour expliquer la perte de diversité génétique à 9 locus sur 10, la dérive génique est plus convaincante.

6 novembre 2015

Exercice

On échantillonne une population de fourmis dans un désert d'Afrique du Nord. Les individus récoltés sont génotypés pour 3 gènes différents, dont on soupçonne qu'ils sont impliqués dans l'adaptation de cette population aux conditions désertiques. Les fréquences génotypiques suivantes sont observées :

- Gène 1 : Déficit en hétérozygotes
- Gène 2 : Déficit en hétérozygotes
- Gène 3 : Déficit en hétérozygotes (ici c'est donné, mais à l'examen faudra montrer qu'on est ou pas dans un équilibre de H-W).

1) Comment peut-on interpréter ces observations ?

Suite à une année particulièrement chaude, la taille de la population est divisée par 100.

- 2) Comment va évoluer la fréquence de l'allèle A (gène 1) dans les prochaines générations ?
- 3) Comment va évoluer la diversité génétique au sein de cette population ?
- 4) Risque-t-il d'y avoir un impact sur le déséquilibre de liaison observé entre les 3 gènes ?

1)

>> Accouplement sélectif (vu la diminution de la proportion en hétérozygotes...)? Plausible, pas vraisemblable dans notre contexte. Ici on n'a pas une réduction des hétérozygotes sur tous le génome. Les 3 gènes pourraient influencer un génotype, génotype qui influence l'accouplement entre individus de phénotypes semblables. Ce sont des gènes impliqués à l'adaptation à la chaleur, ils pourraient alors influencer la couleur de l'insecte : si l'insecte est plus clair il absorbera moins les rayons du soleil et donc être mieux protégé contre la chaleur (plus précisément contre les rayons UV). Et il pourrait y avoir comme ça, un mécanisme qui est apparu au sein de cette espèce, ou les partenaires sexuels ont tendance à choisir des individus plutôt clairs. Mais c'est plutôt une sélection dans ce cas et non un accouplement sélectif.

>> Sélection contre les hétérozygotes ? il faut favoriser nécessairement les homozygotes contre les hétérozygotes dans ce cas ? si on avait eu un déficit aussi au niveau des homozygotes, on l'aurait précisé dans l'énoncé, donc oui, cette sélection contre les hétérozygotes est plausible.

>> Accouplement consanguin ? Dans ce cas, la grande majorité des gènes au sein du génome présentent un déficit en hétérozygotes. Il n'est donc pas étonnant que quand on prend 3 individus au hasard, on y trouve également un déficit en hétérozygotes. (Si par contre on en prend 3 dont 1 ne présente pas ce déficit, on aurait dû rejeter cette hypothèse non compatible avec les données)

Aucune de ces 3 hypothèses peut être rejeté par rapport aux observations.

2)

La fréquence va fluctuer plus intensément par rapport à la population initiale. Vu que la taille est plus petite (population divisée par 100), la force de la dérive génique est plus importante et, avec elle, la fluctuation. La probabilité qu'un des allèles atteigne une fréquence de 0 à un moment donné va donc augmenter également. Via cette perte d'allèle on va diminuer la diversité génétique de la population dans des populations de petite taille.

Revenons en arrière : s'il y a consanguinité, cela ne va pas changer la réponse 2, la fréquence de l'allèle va fluctuer également.

En cas de sélection de défaveur contre les hétérozygotes, la fréquence de l'allèle dominant (par

exemple) va diminuer vers 0. En cas de population de petite taille, la probabilité de perte de la diversité génétique va être importante si la sélection en plus va tendre à éliminer un des allèles, ces deux forces vont se combiner puisqu'elles vont dans la même direction. Alors dans ce cas la diversité génétique peut diminuer plus rapidement, mais uniquement si la sélection favorise un des allèles par rapport à l'autre. Par contre si la sélection favorise les hétérozygotes, on aura deux effets opposés.

Quand on compare sélection / dérive génique, la sélection n'agit que sur un ou quelques gènes, alors que la dérive génique agit sur tout le génome (son effet est plus global).

Imaginons une population avec une fréquence de 5% pour un allèle X et 95% pour un allèle Y. Au cours de la génération suivante, si la taille de la pop° est réduite fortement, ces fréquences peuvent fluctuer fortement en passant de 5% à 50% et ainsi contribuer à augmenter la diversité génétique. Si on attend suffisamment longtemps, la probabilité de disparition un allèle est plus élevée (= perte de diversité génétique).

3)

Diversité génétique se mesure de 2 façons :

- En regardant le nombre d'allèles : Là, elle ne peut qu'au mieux se stabiliser, au pire diminuer car un allèle peut disparaître sous l'effet de la dérive génique
- En regardant la fréquence des allèles : Là, elle peut diminuer ou augmenter ou rester stable

En regardant l'ensemble du génome, au cours du temps on tendra obligatoirement vers une perte de la diversité génétique (on ne sait pas à quelle vitesse). Autre paramètre : Les mutations. Si la force de la mutation est égale ou plus élevée que la force de la dérive génique, alors là la diversité génique peut augmenter au sein de la population.

Si la population est très petite, ça va fluctuer très fort et la probabilité qu'il y ait une mutation qui viennent contrecarrer la dérive génique sera faible. A terme on ira vers une perte de diversité génétique. → Il y a un équilibre entre mutation et dérive génique.

4)

La dérive génique peut-elle modifier un déséquilibre de liaison entre 2 allèles ? Oui, car elle change les fréquences alléliques et gamétiques. Or, pour être à l'équilibre de liaison, la fréquence des gamètes doit être égale au produit des fréquences alléliques qui les composent.

La dérive génique a un impact plus élevé sur les fréquences alléliques que sur les fréquences gamétiques. La petite taille de la population donc, a une influence sur le déséquilibre de liaison (puisque si la population est petite, la dérive génique est grande).

Évolution de la tolérance au lactose chez l'Homme

Les entérocytes sont des cellules localisées au niveau de l'intestin, responsables de la transformation et de l'absorption des nutriments. Chez les jeunes mammifères, ces entérocytes contiennent des niveaux élevés de lactase, une enzyme qui permet la digestion du lactose. Logique : À ce moment de leur vie, le lait est leur source principale de nourriture.

ROLE = La lactase va transformer le lactose et le diviser en deux sucres :

- Le glucose
- Le galactose

Ces deux molécules sont utilisables par l'organisme et les cellules pour libérer de l'énergie. La lactase est une protéine transmembranaire, capable de capter les molécules de lactose dans l'intestin et de les transformer en sucres, sucres qui sont ensuite transférés dans la circulation sanguine et acheminés vers les différentes cellules de l'organisme qui les utilisent pour produire de l'énergie.

Chez presque tous les mammifères, on observe une diminution de la synthèse de lactase dans les années qui suivent le sevrage (chez 65% des individus de la population humaine). La régulation de la synthèse de lactase avec le sevrage est le facteur principal séparant les individus intolérants au lactose et ceux qui y sont tolérants. Chez 35% des êtres humains, qui sont tolérants au lactose, cette synthèse de lactase est élevée.

Pourquoi les individus adultes ne peuvent-ils pas digérer le lactose (=taux de lactase diminue) ? Le mécanisme de réduction de la lactase permet de conserver l'énergie cellulaire. La production de lactase requiert de l'énergie, or les mammifères ne consomment presque plus de lait une fois qu'ils sont sevrés : la lactase devient inutile. La régulation est alors un mécanisme visant à conserver l'énergie cellulaire et éviter toute perte énergétique inutile.

Chez les Hommes qui sont intolérants au lactose, s'ils continuent à boire du lait, le lactose une fois dans le grand intestin déclenche des symptômes d'intolérance :

- Augmentation de lactose dans l'intestin car il n'est pas transformé par l'organisme, ce qui crée un gradient osmotique attirant l'eau dans l'intestin, ce qui cause des crampes et de la diarrhée
- Les bactéries de l'intestin se nourrissent de lactose, et libèrent des déchets cellulaires : Dégagement gazeux (CH_4 , H_2 ...)

Quelle base génétique pour la tolérance au lactose ?

La protéine correspondant à la lactase contient 1927 acides aminés. Synthèse de lactase : Transcription (contrôlée par des facteurs), traduction à partir de la séquence ARN, modifications post-traductionnelles puis expression au niveau de la membrane cellulaire (car c'est une protéine transmembranaire).

La tolérance au lactose est provoquée par une mutation ponctuelle, concernant 1 nucléotide situé à 14,000 bases en amont du gène lactase. Ce C transformé en T permet la tolérance : Sur cette région se fixe le facteur de transcription Oct1, permettant la transcription du gène lactase. La mutation augmente la fréquence ou l'intensité de la fixation.

Chez les individus mutants (état évolutif dérivé) donc, plus de facteurs de transcription sont attirés ce qui empêche la diminution de la transcription du gène lactase → le niveau de lactase dans les entérocytes de ces personnes reste élevé.

Comment l'expliquer dans l'Histoire de la population humaine ?

Il faut remonter au Néolithique : Révolution, l'Homme passe du stade « chasseur-cueilleur » à un mode de vie sédentaire grâce à l'agriculture et l'élevage. De nombreuses innovations ont été développées par l'Homme, changeant son rapport à l'environnement (outils pour planter, réserves pour stocker la nourriture, ...) notamment le **pastoralisme**, c'est-à-dire l'élevage d'animaux qui produisent du lait, adopté par différentes cultures à cette période.

Le pastoralisme et la tolérance au lactose sont apparus au même moment. Ces deux changements se sont inter-influencés : Coévolution.

Comment la dissémination dans la population humaine s'est-elle faite ?

→ Voir slide : carte actuelle.

Dans nos régions, la tolérance au lactose est élevée : on atteint presque 90%. Dans d'autres régions du monde (ex : Indonésie, Asie) ce n'est pas le cas. En Afrique : Continent contrasté (régions de tolérance, d'autres non).

Deux histoires d'apparition de la tolérance au lactose différentes : Elles se sont répandues à travers l'Europe et l'Afrique de manières différentes. On a ici un phénomène d'évolution convergente.

En Europe, les premiers indices de pastoralisme ont été découverts au Moyen-Orient (découverte d'ossements de vache vieux de 10,500 ans). Les premières vaches domestiques européennes sont apparues il y a 8,000 ans. Ces vaches domestiques sont plus proches de celles du Moyen-Orient que des espèces sauvages présentes en Europe à l'époque : Migration du pastoralisme du Moyen-Orient vers l'Europe.

Comme ces individus au mode de vie sédentaire basé sur l'agriculture et l'élevage avaient une technologie avancée, ils se sont répandus en Europe et sont arrivés dans nos contrées.

Paradoxe ? Des indices anthropologiques suggèrent que le pastoralisme est apparu il y a 10,500 à 6,500 ans... Hors la tolérance est arrivée il y a entre 7,000 et 5,000 ans. Les Hommes récoltaient du lait, alors qu'ils ne pouvaient pas le digérer ! Comment l'expliquer ? Le **fromage**. On pense qu'au début, le lait n'était pas consommé directement mais intervenait dans la fabrication du fromage.

Le fromage en effet, a un contenu en lactose beaucoup plus réduit que le lait. La faible concentration de lactose lui permettait de le digérer, et a permis petit à petit de faire développer la tolérance au lactose. Preuve : Un reste de poterie trouvé en Europe intervenant dans la filtration du lait et dans la fabrication du fromage.

En Afrique, les mutations à la base de la tolérance sont différentes (== l'apparition de cette tolérance est apparue en Afrique de manière indépendante de l'Europe et le Moyen Orient). En comparant les séquences au niveau du site Oct1 :

- En Europe, le C se transforme en T chez les tolérants en lactose
- En Afrique, d'autres mutations (sur des G) mais elles ont le même effet

On pense qu'il y a eu deux apparitions indépendantes (une dans la région du Kenya, l'autre du Soudan) ayant le même effet d'augmenter la transcription de la lactase : Cas d'**évolution convergente**.

Évolution de E. Coli

Avantage de la bactérie : facile à élever, le temps de génération est très court (20 minutes dans les conditions idéales). Idée de Richard Lenski : Faire une expérience simple sur un temps très long. Son expérience a débuté en 1988 et se poursuit encore à l'heure actuelle. Ça représente un temps évolutif extrêmement considérable.

L'expérience s'est déjà déroulée actuellement sur plus de 60,000 générations. Pour atteindre ce nombre avec la population humaine, il faudrait attendre 1 million d'années.

Description de l'expérience

Lenski a au départ débuté avec 12 cultures. Dispositifs : erlenmeyers avec des milieux de cultures dans lesquels les bactéries qui se développent. Ces populations sont isolées les unes des autres. Lenski prend 100 µl d'une population qu'il transfère dans un nouvel erlenmeyer (avec un nouveau milieu de culture) mis à 37°C, ceci chaque 24h pour les 12 populations.

Après 16 ans, une observation inattendue concernant la culture 9 (voir slide) : pour une des populations, le milieu de culture est beaucoup plus trouble que pour les autres ! Comment

l'expliquer ? La population bactérienne y est plus importante.

Le milieu de croissance utilisé contient du glucose mais également du citrate. Normalement, la bactérie ne s'en nourrit pas en présence d'oxygène. Les bactéries de la culture 9 ont-elles acquis la capacité de croître à pd citrate ? On imagine l'apparition d'une mutation au niveau des acides nucléiques, la sélection naturelle choisit ces individus dont la fréquence augmente dans la population.

Métabolisme du citrate : les cellules de cette culture trouble sont capables d'importer le citrate présent dans le milieu culture... Mais en conditions aérobies, cette bactérie ne peut normalement pas en importer ! Pourquoi est-ce intéressant pour la cellule de se nourrir de citrate ? Une fois rentré dans la cellule, le citrate peut être métabolisé via le Cycle de Krebs.

Pour rappel, le citrate est un intermédiaire du cycle de Krebs qui produit de l'énergie au niveau cellulaire chez les organismes qui vivent en présence d'oxygène.

Le milieu de culture contient beaucoup plus de citrate que de glucose.

La protéine de transport CitT transmembranaire permet de faire rentrer le citrate dans la cellule (voir slide). Il y rentre uniquement si la cellule est capable d'éjecter un succinate (qui fait lui aussi partie du Cycle de Krebs).

Dans l'expérience de Lenski, les cellules de la population trouble (n°9) ont acquis la capacité de produire cette protéine de transport en présence d'oxygène. Cela permet l'importation d'un citrate et le succinate conjointement est expulsé en dehors de la cellule. Une fois rentré, le citrate est incorporé dans le Cycle de Krebs et transformé en succinate, une opération qui produit deux NADH et un ATP. → la cellule utilise le citrate pour générer de l'énergie !

La deuxième moitié du Cycle de Krebs est alors court-circuitée : Le succinate issu du citrate est sorti de la cellule et ne poursuit pas le cycle !

Intéressant : Produire deux NADH et un ATP donne plus d'énergie que la voie classique qui produit un FADH₂ et un NADH. Le citrate a donc un potentiel énergétique plus élevé que le succinate.

Au niveau moléculaire ? Le gène citT code pour la protéine transmembranaire réalisant l'antiport. Il est inclus dans l'opéron cit. Les gènes de l'opéron sont transcrits à partir d'un promoteur unique localisé au début de l'opéron. L'ARNm transcrit est ensuite transformé en protéine (système classique).

L'opéron subit un contrôle négatif en présence d'oxygène : Une protéine répresseur se lie au promoteur et bloque la transcription de tous les gènes en aval, donc de citT.

Dans la population trouble n°9, une mutation s'est produite : Un fragment d'ADN de la région du gène citT a été dupliqué et se trouve donc en deux exemplaires dans le génome de la bactérie. Résultat : L'opéron cit est toujours inhibé par le répresseur, les gènes ne sont pas transcrits... Mais les promoteurs en aval de l'opéron ne sont pas inhibés en présence d'oxygène ! La deuxième copie du gène peut donc être transcrite, car non-soumise à l'inhibition.

→ Ceci explique la croissance rapide de la culture car les bactéries n°9 peuvent utiliser le citrate et ainsi générer plus d'E que les autres cultures.

Dans cet erlenmeyer, les cellules pouvant utiliser le citrate sont nommées Cit+. Celles ne le pouvant pas Cit-. Comment se fait-il qu'il reste des Cit- dans cet erlenmeyer ?

Hypothèse la plus probable : Dans les populations normales, tous les individus se nourrissent de glucose. Dans l'ermenmeyer considéré, les individus se nourrissent de glucose et de citrate. Il y a **deux niches différentes** au sein de cette culture : Les bactéries qui se nourrissent de citrate ont moins de facilité à se nourrir de glucose. Celles se nourrissant de glucose n'ont pas accès au citrate.

On a 3 ressources disponibles désormais dans l'environnement : Le glucose, le citrate et le succinate. Les Cit+ se nourrissent des 3, les Cit- uniquement du glucose et du succinate... Mais elles sont plus efficaces que les Cit+ pour se nourrir du glucose.

On a alors réalisé un arbre phylogénétique. En suivant l'évolution de la population au cours du temps en gardant des échantillons à différents moments, on peut récolter les individus de ces différentes périodes et les séquencer. Ce faisant, on a mis en évidence les moments où sont apparus les mutations permettant l'apparition de la souche Cit+ dans la culture 9. **Des mutations préalables étaient nécessaires pour qu'apparaisse la mutation débouchant sur Cit+.**

Arbres phylogénétiques

Référence suggérée : *Tree thinking – an introduction to phylogenetic biology* (Baum D and Smith SD, 2013) – Roberts & Company, Greenwood Village, CO, USA

La **microévolution** est l'étude de l'Évolution au cours des générations jusqu'à une centaine de générations). La **macroévolution** c'est la même chose sur une échelle de temps plus grande, permettant d'observer l'émergence de motifs évolutifs intéressants. Outil principal dans notre étude de la macroévolution : L'arbre phylogénétique.

#slide : montre les liens de parenté entre individus et l'arbre phylogénétique des espèces.

Une ligne représente une génération. Chaque individu est connecté à la génération précédente par 2 lignes et peut avoir des descendants ou non. Comment expliquer une telle différence de succès reproducteur ?

>>Au hasard : on ne peut pas déterminer le SR de l'individu en fonction de son génotype. Tous les individus ont une probabilité équivalente, mais au final ils n'auront pas le même nombre de descendants.

>>Par sélection : on peut prédire le succès reproductif

Quels sont les facteurs expliquant le succès reproducteur ?

- Les ressources
- Les prédateurs
- Les couleurs (camouflage, ...) → la sélection intervient !
- Le hasard (= aucune prédiction ne peut être faite dans ce cas).

→ Dans toute population, on a une variation du SR, prédictive (sélection) ou non (hasard).

Différence entre **cladogramme** et **phylogramme** : Dans le premier, on n'utilise pas l'information sur la longueur des branches. La longueur entre l'ancêtre commun à toutes les espèces et chacune des espèces est représentée de manière identique. Dans un phylogramme, la longueur des branches varie.

Dans un arbre enraciné, on a déterminé la position de l'ancêtre commun. Dans un arbre déraciné, on ne sait pas où l'ancêtre commun se trouve.

Polytomie : Plus que deux branches partent d'un nœud. Cela traduit un manque d'informations. Et ça peut aussi signifier que la spéciation a eu quasiment lieu en même temps. En d'autres termes, une espèce ancestrale se divise en 3.

Groupe **monophylétique** : Un ancêtre commun et l'ensemble des descendants

Groupe **paraphylétique** : Un ancêtre commun et une partie des descendants

Groupe **polyphylétique** : N'inclut pas l'ancêtre commun, cas de convergence évolutive. On utilise ce terme pour grouper des espèces qui se ressemblent à cause d'une convergence évolutive, pas à cause d'un ancêtre commun.

12 novembre 2015

#slide : on prélève l'ADN de différents organismes (différents individus de la même espèce, différentes espèces du même genre...), on les séquence et ainsi on reconstruit l'arbre phylogénétique pour voir la relation entre ces différents organismes.

Intérêt d'un arbre phylogénétique :

A) Retracer l'histoire évolutive d'une espèce ou de gènes (voir plus loin)

B) Classification, qui se base sur l'évolution et l'arbre phylogénétique.

Exemple : l'homme, qui est inclus au sein des primates. Controverse : place de l'homme par rapport aux chimpanzés et aux gorilles → on a hésité pendant longtemps à placer l'homme/chimpanzés dans un même groupe à l'exclusion du gorille ; l'autre hypothèse était de regrouper chimpanzés/gorilles, en supposant qu'ils n'avaient pas d'ancêtre commun avec l'homme. Pourquoi ces hypothèses ? Parce que la branche qui sépare l'ancêtre commun entre les gorilles et d'autres primates et l'ancêtre commun entre l'homme et les chimpanzés/bonobos était trop courte. Cela signifie que peu de mutations se sont produites le long de cette branche. En fonction du type de séquence d'acides nucléiques analysées, parfois on obtenait un groupe monophylétique entre chimpanzés et gorilles, parfois autre chose. Comme les techniques d'analyse ont évolué, on a pu analyser un grand nombre de gènes différents, et on a pu ainsi créer l'arbre phylogénétique actuel.

Exemple : Placement des cétacés parmi les différentes espèces de mammifères. Leur placement est difficile car ils ont colonisé un habitat totalement différent des autres mammifères : le milieu aquatique. Suite à ça, développement d'adaptations morphologiques. En comparant les squelettes ou différents caractères morphologiques, il est difficile de regrouper les cétacés avec d'autres mammifères... Pendant longtemps on a considéré qu'il s'agissait d'un groupe à part.

En analysant et comparant les séquences ADN, l'inclusion dans la classe des mammifères a été avérée : ce groupe forme un groupe monophylétique avec les hippopotamidae. Une des conclusions que l'on peut tirer de l'arbre (#slide) : les cétacés sont inclus dans les artiodactyles (nombre de doigts paire $><$ périssodactyle : nombre de doigts impaire). On peut en conclure qu'un cétacé est plus proche d'une vache que d'un cheval, ce qui paraît peu intuitif sur base de caractères morphologiques.

Autre exemple : Tenrecidae, vit en Afrique et ressemble très fortement aux hérissons. Ils sont plus proches phylogénétiquement des éléphants que des hérissons. La ressemblance entre les Tenrecidae et les hérissons : convergence évolutive → des caractères morphologiques semblables se sont développés de façon indépendante chez différents organismes suite à l'adaptation à l'environnement.

Exemple : Strepsiptera : Insecte parasite. Les femelles ne se développent pas, elles restent au stade larvaire (elles sont dites « néoténiques »). Pour s'adapter à l'hôte, la femelle a subi des adaptations morphologiques importantes. #slide : photographie à microscopie électronique → on constate que le mâle possède une paire d'ailes, l'autre est transformée en altères ce qui fait penser aux diptères. Mais, chez ces derniers, ce sont les ailes arrières qui sont modifiées. Ici, il s'agit de la première paire d'ailes.

Plusieurs hypothèses de classement des Strepsiptera aux seins des insectes. Récemment, en comparant les séquences ADN pour différents gènes de ce groupe, on constate que les Strepsiptera forment un groupe monophylétique avec les diptères. Il y a donc probablement une origine commune à la transformation d'une des deux paires d'ailes en altères. Le problème, c'est que cette

hypothèse a été contestée par la suite,

Des auteurs ont montré que le regroupement de ces deux types d'insecte est dû à un artefact de l'analyse phylogénétique : dans le phylogramme (ici on tient compte de la longueur des branches), la branche qui mène au Strepsiptera, de la même manière que la branche qui mène au diptère, est très longue. Alternance de branches courtes et longues peut aboutir à un artefact = les longues branches ont tendance à s'attirer.

En utilisant des méthodes plus sophistiquées (modèles de substitution nucléotidique), on observe un arbre différent où les Strepsiptera se collent avec les coléoptères.

Exemple : Myzostomida, organisme parasite d'échinodermes, qui paralyse des organismes marins. En séquençant les gènes et comparant l'ADN avec d'autres organismes, on a pu le classer proche des plathelminthes.

C) Épidémiologie

Épidémiologie (=histoire de transmission) du virus HIV : Taux de mutation x1,000,000 plus élevé que ce qu'on connaît chez l'Homme... Grâce à ce taux de mutations élevé, les outils permettant d'étudier l'Évolution des espèces peuvent être utilisés pour étudier l'évolution du virus HIV.

Deux types de virus :

- SIV : homologue de HIV qui infecte les populations de singes, de chimpanzés
- HIV : infecte les hommes
 - HIV1 : assez distant de HIV2, mais plus proche de certaines souches de SIV
 - HIV2 : plus proche de certaines souches infectant les singes verts d'Afrique de l'ouest que de HIV1

Hypothèse : Deux histoires de contamination du virus HIV à partir de différentes populations de singes. → **Plusieurs événements de transmission du virus HIV à la population humaine.**

A) Médecine légale

➔ Pour voir quelle personne peut avoir transféré le virus HIV à une autre.

Sur le slide : Un dentiste soupçonné de transmettre HIV à ses patients. Autre cas célèbre : Un médecin soupçonné d'avoir transmis HIV par injection à son ex petite amie. Grâce à un arbre phylogénétique, on a établi que le HIV qui lui a été inoculé l'a bien été par ce médecin (via le prélèvement du virus HIV d'un de ses patients).

II. Évolution des organismes et de certaines de leurs caractéristiques

II.1. Évolution des sociétés d'insectes

Eusocialité : Constitution de sociétés d'insectes, d'hyménoptères très socialisées. Deux groupes d'insectes ont établi des sociétés grandes, très hiérarchisées :

- Abeilles
- Méléipones (insectes tropicaux)
- Bourdons : Niveau de complexité plus faible, les colonies de bourdons d'ailleurs ne survivent qu'un an
- Euglossines : Abeilles solitaires

Hypothèse : Apparition de la sociabilité de manière progressive, basé sur des séquences morphologiques. Après analyse des séquences ADN : Les abeilles forment un groupe monophylétique avec les euglossines (qui elles sont solitaires).

Probablement : Apparition des sociétés d'insectes complexes 2 fois de manière indépendante (une fois chez les abeilles-genre *Apis*, une autre fois chez les mélipones).

II.2. Adaptations convergentes

Grenouilles de Madagascar/Asie : on constate différents écomorphes (espèce adaptée à un habitat particulier et y développe des adaptations morphologiques ou autres particulières) sont les mêmes dans les deux environnements :

- Grenouilles fouisseuses, adaptation des pattes pour mieux creuser le sol
- Grenouilles arboricoles, développement larvaire entièrement dans l'œuf (au lieu de se faire dans l'eau comme la plupart des grenouilles).
- Grenouilles vivant sur des rochers, présentent des pattes très musclées adaptées au saut

➔ Point important sur l'évolution : si chaque fois qu'une grenouille arrive dans un habitat et s'adapte à celui-ci de la même manière chez plusieurs espèces, cela signifie que l'évolution est contrainte (ne peut pas aller dans n'importe quelle direction), en tout cas certainement au niveau morphologique.

Autre exemple : Caraïbes. De nouveau, des écomorphes sont semblables dans les Grandes Antilles (Haïti, Cuba, Jamaïque, Porto Rico). Chaque écomorphe se trouve dans plusieurs îles. On utilise leurs caractéristiques morphologiques typiques pour les regrouper ensemble. En établissant un arbre phylogénétique, on observe que les écomorphes sont complètement mélangés. On peut avant tout penser qu'un ancêtre commun se soit diversifié sur une île, et que ses descendants aient colonisé ensuite les autres îles. En réalité, ce qui s'est passé c'est que chaque écomorphe s'est développé de façon différentes sur chaque île.

II.3. Coévolution parasites-hôtes

Si un parasite est fortement associé à son hôte, il va suivre un même chemin évolutif. Si l'espèce-hôte est divisée en deux parties par une spéciation, le parasite fait de même. Le parasite ancestral va également se différencier en deux espèces. L'arbre phylogénétique de l'hôte serait donc exactement identique à l'arbre phylogénétique du parasite ? On peut tester cette hypothèse par comparaison des arbres phylogénétiques du parasite et de l'hôte via les séquences ADN.

Dans la réalité, l'hypothèse peut être réalisée ou non. Comment l'expliquer si les arbres ne concordent pas ? Deux explications :

- 1) Transfert du parasite d'une des branches de la spéciation à l'hôte. Le parasite reste sur les individus de la 1ère branche (puis ensuite disparaît), mais certains peuvent s'adapter secondairement aux individus de la 2ème branche. On imagine, une fois que P2 a colonisé H2, s'il est plus adapté que P4, il a pu faire disparaître P4.
- 2) Pas de transmission du parasite d'une des branches à l'autre. On imagine alors un événement de spéciation chez l'ancêtre commun : L'espèce ancestrale possède les deux parasites, qui coexistent durant un temps. Chez les différents hôtes (H1, H2 et H3) qui apparaîtront par la suite (spéciation), un des parasites de départ disparaît. Mais ce n'est pas tout le temps la même lignée qui disparaît. Ceci explique pourquoi l'arbre phylogénétique du parasite est différent de celui de l'hôte.

On peut séparer ces deux hypothèses en se basant sur le phylogramme du parasite :

- 1) Les deux espèces (P1 et P2) sont proches, les branches sont relativement courtes depuis la différenciation à partir de l'ancêtre commun.
- 2) L'ancêtre commun entre les deux espèces est beaucoup plus ancien dans le temps.

#slide – cas concret de coévolution : pucerons, et bactéries endosymbiotiques qui vivent dedans. Ici,

on constate que la phylogénie du puceron et des bactéries est semblable, confirmant que l'histoire évolutive a eu lieu en parallèle. On a pu dater les événements de spéciation car des fossiles existent chez les pucerons.

III.

III.1. Évolution d'une protéine

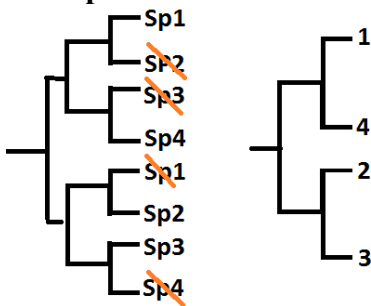
- On se focalise ici sur les gènes interloqués : gènes qui sont impliqués dans la régulation de l'infection chez les mammifères et notamment chez l'homme.

On constate 4 copies différentes du gène chez l'homme. En regardant l'arbre du gène, il ne correspond pas vraiment à l'arbre des espèces. On peut expliquer sa part des événements de duplication et d'extinction.

Imaginons qu'on ait 4 espèces Sp1, Sp2, Sp3, Sp4 et qu'au départ avant la spéciation on ait un phénomène de duplication, ce qui implique 2 copies du gène dans le génome de ces 4 espèces. Si on prend les différentes séquences de la première copie, on va trouver l'arbre phylogénétique des espèces. Si on regarde la même chose pour la deuxième copie, on devrait trouver le même arbre phylogénétique.

Maintenant, à partir du moment où duplication, une des deux copies n'est plus indispensable : soit elle peut prendre une autre fonction, soit elle évolue dans un cas où la contrainte sélective disparaît, alors elle se différencie rapidement et tellement fort de l'autre copie qu'on ne la reconnaît plus dans le génome. On a alors une extinction d'une des deux copies.

Imaginons que ce soit Sp2 et Sp3 qui disparaissent chez deux espèces et Sp1 et Sp4 qui disparaissent dans deux autres espèces (voir schéma ci-dessous) : un biologiste séquence ce gène qui se retrouve à une copie dans le génome, il va alors établir les relations de parenté en construisant un arbre phylogénétique **du gène** qui ne va plus correspondre à l'arbre phylogénétique **des espèces**.



Si on veut étudier les événements de duplication, on va essayer de concilier l'arbre du gène et des espèces de la façon suivante : (sur la slide) en gris on a les copies que l'on n'a pas trouvées mais qui sont nécessaires pour réconcilier les deux arbres, en sombre on a les copies qui étaient présentes au départ.

Quand on a une copie en gris : soit on n'a pas trouvé le gène, soit il s'est tellement différencié de la première copie et donc on ne peut l'identifier.

Comme chez l'homme il y a 4 copies du gène, cela implique au moins 3 événements de duplication, mais pour pouvoir concilier les arbres, il faudra prendre en considération 4 événements de duplication.

#slide suivante : évolution de lambda max des protéines opsin sensibles aux UV chez les vertébrés. C'est une étude réalisée par des chercheurs américains.

Protéine opsin : impliquée dans le mécanisme de la vision, notamment chez les vertébrés mais aussi chez les insectes. Ces opsines ont été développées de façon indépendante chez plusieurs espèces.

L'opsine est un gène qui code pour une protéine associée à un chromophore. Ce couple prot-chromophore est capable de capter la lumière dans une longueur d'onde plus ou moins étroite. La séquence d'acides aminés de cette protéine va déterminer la gamme de longueur d'onde à laquelle ce chromophore devient sensible. Donc, en fonction de l'évolution de la séquence de cette protéine, plusieurs longueurs d'onde différentes peuvent être captés par le chromophore.

Les chercheurs ont séquencé l'opsine dans divers groupes de vertébrés, et ont pu déterminer par expériences en labo à quelle gamme de longueur d'onde ce couple prot-chromophore était sensible.

λ_{\max} : longueur d'onde à laquelle la protéine capte le maximum de la lumière, ici 359, 363 430... etc.

La longueur d'onde aux alentours de 360 correspond aux UV, une longueur d'onde captée par des protéines primitives. Si on va plus haut (grenouille, poulet) la longueur d'onde est plus élevée, et donc la protéine ne capte plus des UV mais d'autres.

Les chercheurs ont utilisé l'arbre des vertébrés pour inférer les séquences des protéines ancestrales (en d'autres termes, en utilisant l'arbre et la séquence de la protéine, on estime la séquence ancestrale de la protéine). Ils ont alors synthétisé cette protéine ancestrale estimée en labo pour in fine, mesurer la longueur d'onde pour laquelle cette protéine était sensible.

On a pu constater que cette sensibilité aux UV a disparu chez plusieurs groupes de vertébrés au cours de l'évolution : dans le groupe des oiseaux par exemple, on part de 360 à 393 → l'ancêtre des oiseaux a perdu cette sensibilité aux UV, sauf que chez certains elle est réapparue (zebrafish, canari...).

Pseudogène (voir définition google) : exemple du cœlacanthe, qui a perdu la capacité de détecter la lumière UV mais en plus a perdu la capacité de produire l'opsine (même chose pour le dauphin).

III.2. Transferts horizontaux

Les éléments transposables (ou éléments P) se trouvent entre autre chez les drosophiles. Une fois qu'un élément P est présent dans un génome, il peut se dupliquer et s'y insérer à un autre endroit. En comparant les séquences des éléments P de différentes espèces de drosophiles, on obtient l'arbre phylogénétique présent sur la slide.

Melanogaster forme un groupe monophylétique avec *willistoni*, leurs séquences sont très proches. *Mélanogaster* ne présente qu'une seule mutation par rapport à un des allèles de *willistoni*.

Mais en comparant ceci à l'arbre des espèces : *melanogaster* est très éloigné de *willistoni*. La plupart des espèces du groupe de *melanogaster* ne possèdent pas d'éléments P... Étonnant puisque sur le 1er arbre les séquences de *willistoni* et de *melanogaster* sont très proches. Ça s'explique par un transfert horizontal, d'un élément P du génome de *willistoni* (où la variation génétique est très importante) vers le génome *melanogaster* (où il n'y a pas de variation génétique).

Comment ce transfert est-il possible ? Plusieurs hypothèses. La plus vraisemblable : Les deux espèces sont parasitées par le même tique. Ce dernier, se nourrissant du sang de l'un, passe sur l'autre espèce et en se nourrissant, transfère des éléments génétiques. Mais reste à comprendre comment l'élément P soit passé du sang d'un individu au génome de l'autre...

En comparant la phylogénie de l'espèce à celle de l'élément P : nécessairement au moins trois éléments de transfert d'une espèce à l'autre.

Arbre sur la slide : en noir = arbre des espèces, en blanc = phylogénie des éléments P. Comme ces deux arbres ne correspondent pas, on imagine 3 éléments de transferts apparus au cours de l'évolution des éléments P entre les différentes espèces de Drosophiles.

- Le transfert est tjrs à pd *willistoni* (variation intra-espèce, où on a plusieurs allèles) à *melanogaster* (présence d'un seul allèle—pas de variation génétique)

- Ces 4 hypothèses sont considérées es plus vraisemblables.

Exercices

1)

a. Faux.

b. Faux.

c. **!\ Question d'examen /** Les 3 espèces affichées sont toutes présentes aujourd'hui. Aucune de ces 3 -là descendent de l'autre. Elles descendent chacune d'ancêtres communs présents dans le passé. « A is the most ancient species » ne veut rien dire... A est aussi ancien que B ou que C car toutes les 3 ont à un moment donné partagé un ancêtre commun. Ils ont donc tous les 3 eu un temps d'évolution équivalent.

Important car on entend souvent ça : « Cette espèce-là est basale, donc elle est plus ancienne que les autres » **c'est absolument faux.**

d. Faux pour les mêmes raisons

e. Vrai.

2)

a. Faux.

b. Faux.

c. Vrai.

d. Vrai.

e. Faux.

3) Viridiplantae inclus tous les descendants de l'ancêtre de Volvox et des plantes terrestres.

a. Vrai.

b. Faux

c. Faux

d. Faux

e. Faux

4)

c. Les 3 autres sont identiques entre-eux. Dans le cas de a. et d. : on intervertit les branches mais on garde la même relation phylogénétique.

5)

a. Les 3 autres sont identiques entre-eux, et présentent BED comme groupe monophylétique.

6)

a. Faux.

b. Faux.

c. Vrai.

d. Faux.

7)

- a. Faux
- b. Faux
- c. Vrai
- d. Faux
- e. Faux

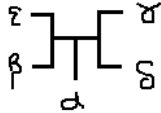
19.11.15 - ANALYSE PHYLOGENETIQUE

I.

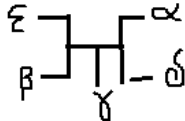
Analyse phylogénétique : méthode qui permettent d'estimer des arbres phylogénétiques.

Ex concret (slide): imaginons 5 espèces (alpha, beta, gamma, delta, epsilon) et 6 caractères (1 → 6) dont chacun a un état 0 ou un état 1. D'après les données de la slide, quel arbre phylogénétique pourrait-on construire pour ces espèces ?

Proposition : → Problème : Le deuxième caractère n'est pas compatible dans ce cas.



On change alors l'arbre : → compatible à 100% avec les données observées.



Dans la réalité, on n'a pas toujours un arbre compatible à 100% avec les données. Alors, en changeant la matrice de caractères (slide suivant), l'arbre va changer aussi.

Les caractères 1 et 3 suggèrent un regroupement de bêta et epsilon. Les caractères 4,5,6 suggèrent un regroupement de bêta avec gamma. → Contradiction !

Pkoi a-t-on des caractères en conflit ?

- La convergence évolutive
- Mutations
- ➔ Font apparaître des caractères de façon indépendante chez différentes espèces.

La convergence évolutive est un phénomène rare mais survient quand même souvent, et important dans l'analyse génétique. Si on analyse le génome, il y a 4 caractères possibles : A, T, C, G. La probabilité que deux espèces éloignées aient un AA par convergence est très faible, mais cette probabilité reste toujours non nulle. Cela va entraîner des **conflits de caractères**.

Quand l'arbre n'est pas compatible à 100% avec les données, il faudra s'organiser différemment.

➤ METHODE DE PARCIMONIE

On utilise un critère de parcimonie pour comparer différents arbres et identifier le meilleur par rapport à un set de données (ici, séquences d'ADN).

- Principe de parcimonie : la meilleure explication est l'explication la plus simple.

Exemple : regroupement entre cheval/vache/lombric. A première vue, selon des caractères morphologiques communs (poils, sabots, 4 pattes, mammifères...) on aurait tendance à regrouper le cheval avec la vache. Hypothèse alternative : si le cheval et la vache se ressemblent c'est un phénomène de convergence évolutive. Cette hypothèse a une probabilité beaucoup plus faible, car pour expliquer la parenté entre un lombric et une vache, il faudrait

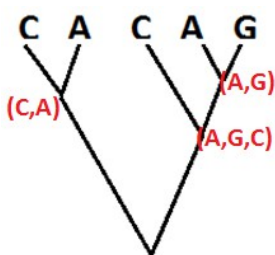
un grand nombre de phénomènes de convergences évolutives → trop complexe, peu probable.

En pratique, on recherche l'arbre qui explique les données observées par le plus petit nombre de convergences évolutives (arbre avec le plus petit nombre de mutations !).

Pour calculer la longueur d'un arbre, on calcule le nombre de mutations sur cet arbre pour chaque site nucléotidique pris séparément (hypothèse= l'évolution se fait indépendamment d'un site à l'autre)., puis on additionne ces nombres de mutations sur l'ensemble des sites.

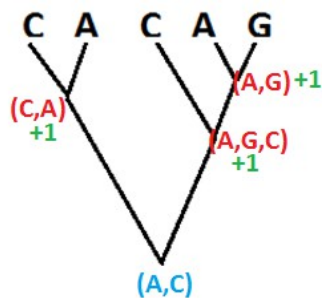
L'analyse de parcimonie consiste donc à comparer la longueur de tous les arbres possibles et de retenir celui ou ceux de longueur minimum.

Exercice. Voici l'arbre qu'on veut évaluer :



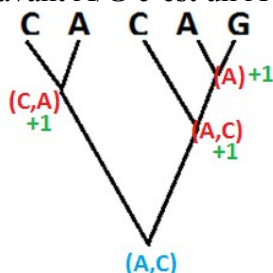
Quand l'intersection est nulle, on y ajoute une mutation (+1) : la longueur totale de l'arbre est alors de 3.

Quand l'intersection est non nulle, on y ajoute A/C : on considère alors que l'état ancestral de l'arbre c'est soit un C soit un A. Les deux sont aussi parcimonieux l'un que l'autre.

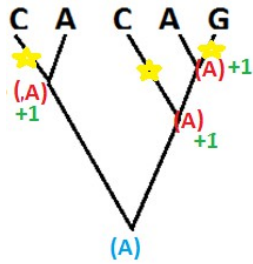


On va alors refaire la même chose cette fois-ci du bas de l'arbre vers le haut :

- L'intersection entre (A,C) et (A,G,C) est non nulle, alors on supprime le G (arder C ou A est plus parcimonieux).
 - L'intersection entre le nouveau (A,C) et (A,G) est non nulle, on va encore supprimer le G.
- Cela veut dire que l'état ancestral le plus parcimonieux en bas cest soit C soit A, et juste avant A/G c'est un A (voir schéma).



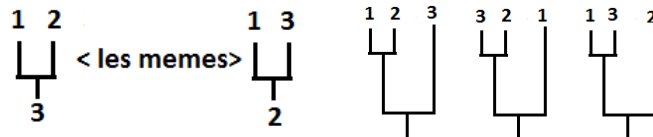
Si on choisit A comme caractère ancestral, on ajoute alors les mutations (étoile jaune) sur l'arbre :



Alors la longueur minimale de l'arbre c'est 3. En d'autres termes, la longueur pour cet arbre et pour ce site nucléotidique, c'est 3 (nombre de mutations minimum pour un set de données).

Maintenant la longueur totale de l'arbre sera la longueur de ce site nucléotidique + la longueur du 2^e site nucléotidique + la longueur du 3^e site nucléotidique + Pour tout le génome.

Problème avec cette méthode : il faut passer en revue tous les arbres possibles. Pour 3 espèces, il y a un arbre non enraciné possible et 3 arbres enracinés possibles.



Pour 5 espèces : 15 arbres non enracinés possibles. Pour 10 espèces : plus de 24 millions.

Le nombre d'arbres possibles augmentent de façon très importante pour chaque espèce en plus étudiée. Comme il y a plusieurs millions d'espèces dans le monde, ce serait impossible de passer en revue tous les arbres pour calculer la longueur minimale, même avec l'aide des ordinateurs (ce serait trop long).

➤ ANALYSE « BRANCH and BOUND »

Cette méthode fonctionne comme suit : on part d'un arbre où l'on connecte 3 espèces. On va alors construire un nouvel arbre à partir des espèces étudiées, en additionnant séparément chaque espèce et en l'additionnant sur l'arbre de façon à l'améliorer.

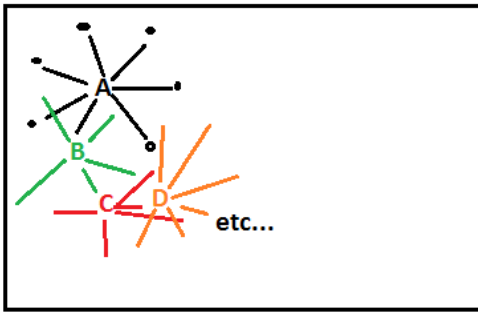
On part de 3 espèces (A,B,C) et on ajoute une nouvelle espèce D que l'on peut placer à différents endroits en gardant l'arbre résultant qui suit le principe de parcimonie. On ajoute l'espèce E, on regarde les différentes possibilités et on regarde la longueur minimale. Ainsi, on élimine à chaque fois des arbres inutiles à comparer. → Cette méthode permet d'éliminer une partie de l'espace à explorer et rendre la tâche plus facile, ainsi on a une garantie de trouver le meilleur arbre.

Problème de limitation : cette méthode arrive jusqu'à une vingtaine d'espèce, mais ne va pas plus loin.

➤ METHODES HEURISTIQUES

Explorent l'espace des possibilités, mais problème de garantie : ne garantissent pas de trouver le meilleur arbre.

Permettent d'analyser un nombre d'espèces/séquences plus ample (jusqu'à une cinquantaine).



Rectangle : espace de tous les arbres possibles.

Une recherche heuristique va démarrer à un endroit de l'espace (A), on va alors calculer la longueur de l'arbre avec le critère utilisé (par exemple celui de parcimonie), on compare cette longueur à la longueur des arbres voisins.

Soit les arbres voisins sont plus longs, alors on ne s'en préoccupe plus. Soit il y en a un parmi les voisins dont la longueur est plus petite que la longueur du premier arbre étudié (A), on va l'inclure dans l'analyse. On explore alors les arbres voisins de B etc... L'analyse continue jusqu'à trouver un arbre qui n'a plus de voisins meilleurs. On a trouvé un arbre optimum.

Problème : il pourrait exister un ou des arbres encore meilleur(s) que celui qu'on a trouvé, qui se trouve(nt) dans un espace du rectangle qui n'a pas encore été exploré.

On doit alors lancer plusieurs analyses, en partant d'un point de départ chaque fois différent.

Comment choisir l'arbre de départ ?

- Possibilité : méthode de « branch and bound » → on part de 3 premières séquences, on y ajoute une autre, ensuite une autre etc... et on obtient un arbre qui contient toutes les espèces. Le seul hic c'est que l'analyse redémarre alors à chaque fois sur le même point de départ, ce que l'arbre meilleur par la méthode « branch and bound » reste tjrs le même. On risque alors de se retrouver sur un optimum local.
- Autre possibilité : on change aléatoirement l'ordre d'ajout des séquences via toujours la méthode « branch and bound ».
- Possibilité radicale : on choisit aléatoirement toujours un arbre différent à partir de l'ensemble des arbres de l'espace.

Sur la slide : chaque plateau représente un critère, et au plus le plateau est haut, au plus le critère est meilleur et ainsi l'arbre est meilleur.

Comment l'espace d'analyse est défini ?

Comment définir les arbres voisins ?

Plusieurs définitions d'arbres voisins donnent des espaces différents.

- Méthode SPR (à ne pas connaître par cœur pour l'examen). On déconnecte une partie de l'arbre et on la reconnecte ailleurs sur la même branche.
- Méthode TBR, une des plus utilisées, qui donne plus d'arbres voisins car elle est moins contraignante.
- Nearest-neighbor interchange process

II. DISTANCES GENETIQUES

== Autre classe de méthode d'analyse phylogénétique, basées sur les distances génétiques.

Distance génétique = distance mesurée entre deux séquences.

1^{ère} étape de toutes les méthodes à pd distances génétiques : reconstruction d'une matrice de

distances génétiques à partir de la matrice de séquence.

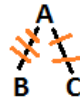
Une première mesure possible de la distance consiste à mesurer la proportion des sites entre deux séquences, qui sont différents.

Si on a deux séquences de 1000 nucléotides, et 5 nucléotides qui sont différents (mutation), la distance génétique sera égale à 5 pour mille (5/1000).

Problème : s'il y a beaucoup de mutations qui séparent deux séquences.

→ **Distance génétique** = nombre de mutations qui se sont accumulées entre deux séquences depuis qu'elles divergent de leur ancêtre commun.

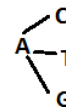
On a donc au départ un ancêtre commun A, qui se différencie en 2 espèces. La distance génétique entre B et C est alors mesurée par le nombre de mutations qui se sont accumulées dans les lignées évolutives depuis que les branches se sont séparées.



Si le nombre de mutations devient trop important, la probabilité qu'on ait deux mutations qui se produisent au niveau du même site nucléotidique est devenue non négligeable.

Dans ce cas si on veut estimer le nb de mutations accumulées, on va sous-estimer le nb de mutation totale qui se sont accumulées. → On utilise alors des méthodes de **substitutions nucléotidiques**, qui va permettre de dériver une équation afin d'estimer le nb réel de mutations qui séparent deux séquences à partir du nb observé entre les deux séquences.

Le taux de mutation de A vers C ou de A vers T ou de A vers G est la même :



Ce modèle est simple. On peut calculer la distance génétique réelle: $D = \frac{-3}{4} \ln \left(1 - \frac{4}{3} D_s \right)$

D_s : le nb de différences observées entre les deux séquences.

Par exemple une séquence de 1000 nucléotides, on observe une mutation au 100ème nucléotide, alors D_s vaut 1/10.

#Slide/

En abscisse : le taux de mutation x le temps.

Au cours du temps, le nombre de sites qui vont différer entre les séquences va augmenter. Cependant, on observe un plateau (un maximum) autour de 75%. Pourquoi ? Car il y a que 4 états possibles (A, C, G, T), si on prend deux séquences générées au hasard (qui n'ont pas une origine commune), une fois sur 4 on aura deux nucléotides identiques par hasard. Si les séquences évoluent depuis longtemps avec un taux de mutation important qui les sépare, les séquences deviennent aléatoires → on aura par hasard des nucléotides qui seront semblables, et non pas parce qu'ils ont une origine commune mais parce qu'il y aura mutation vers la même base azotée. C'est pour cette raison qu'il existe un plateau à 75%.

→ D'où le besoin de calculer le nb de mutation **réel** qui sépare deux séquences.

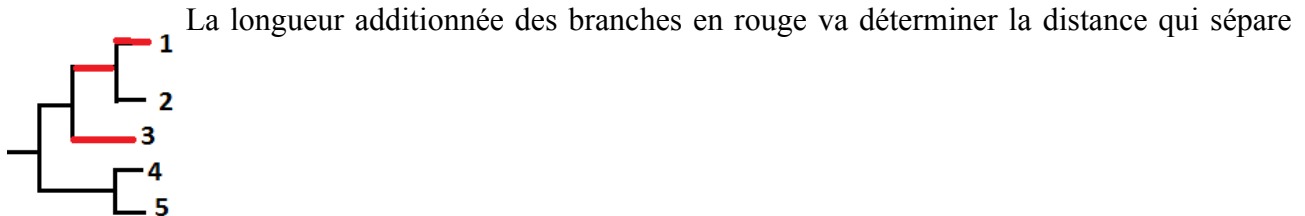
Comment fonctionnent ces méthodes basées sur les distances génétiques ?

Matrice de distances : estime le nb de mutations qui sépare deux espèces

Sp	1	2	3	4
----	---	---	---	---

1	0	1/2	1/3	...
2	2/1	0	2/3	...
3	3/1	3/2	0	...
4	0

Etc.



On compare cette distance avec la distance estimée du tableau àpd séquences ADN (1/3).

On va définir un critère Q : (voir notes du prof)

$$Q = \sum_{i=1}^n \sum_{j=1}^n (d_{ij} - d_{ij}^*)^2$$

Q représente l'écart entre les distances calculées sur l'arbre et celles du tableau.

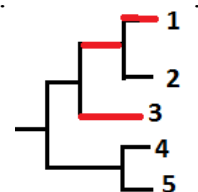
Méthode d'évolution minimum

Se base sur les distances génétiques.

Caractéristique commune : pour un arbre donné, elle va déterminer la longueur optimale des branches en utilisant la méthode des moindres carrés (Q, page précédente).

Pour un cladogramme, cette fois, on définit le meilleur phylogramme, en faisant varier la longueur des différentes branches pour minimiser la valeur de Q.

La longueur totale de l'arbre est toujours donnée par l'addition des branches :



Pour trouver le meilleur arbre, on va ensuite comparer ces longueurs totales de chaque arbre et voir quelle est la plus petite. L'arbre de longueur minimum = meilleur arbre.

Méthode d'UPGMA

On va partir d'une matrice de distance, pour former UN SEUL arbre final. On n'explore donc plus tout l'espace de possibilités !

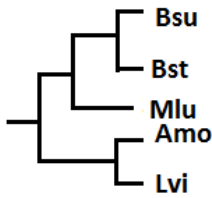
A pour but de regrouper ensemble des espèces ou objets les plus proches possibles, ce n'est donc pas une bonne méthode d'analyse phylogénétique mais va servir pour une autre méthode (voir plus loin).

On identifie les deux séquences/espèces les plus proches (sur la slide : Bsu/Bst).

On va regrouper Bsu et Bst ensemble. Une fois le regroupement fait, on recalcule la matrice, en considérant Bsu/Bst comme une seule espèce.

Par exemple, pour calculer la distance entre Bsu-Bst et Lvi, on va prendre les distances Bsu/Lvi et Bst/Lvi. On en fait une moyenne, qu'on place dans la nouvelle matrice.

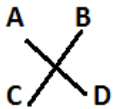
On crée un nouveau groupe : Bsu-Bst-Mlu et on refait la même chose jusqu'à trouver une valeur minimale. On génère alors un arbre.



Problème : si on suit la matrice des 4 espèces A,B,C,D on aurait tendance à les regrouper de manière différente car ne tient pas compte de la longueur des branches de l'arbre de départ.

Méthode de Neighbor-joining

Diffère de l'UPGMA car on démarre d'un arbre non résolu en étoile. On détermine ensuite quels sont les deux premiers taxa qu'on va regrouper en un clade.



On va corriger la matrice de distances de départ et en former une nouvelle, en tenant compte de la divergence d'une espèce par rapport aux autres taxa.

Pour chaque nœud terminal, ici ABCD, on va calculer une **divergence nette** :

$$n_i = \sum_{j=1}^n d_{ij}$$

(dans l'exemple $n=4$)

On calcule la distance entre A et B, entre A et C, entre A et D etc...

On établit alors une matrice de distance corrigée, grâce aux valeurs de divergence nette.

La distance corrigée sera donc :

$$M_{ij} = d_{ij} - \frac{n_i + n_j}{n-2}$$

On compare enfin les distances, en regroupant les espèces présentant la distance la plus petite.

Avantage : tient compte de la longueur des branches, ne donne pas d'estimations biaisées lorsque les distances entre les branches sont différentes. Il n'y a pas non plus de comparaison entre plusieurs arbres, c'est une méthode directe.

Désavantage : on ne peut comparer l'arbre résultant avec d'autres, on ne sait donc pas si cet arbre est le meilleur ou pas au final, mais il s'y rapproche beaucoup.

Cette méthode peut être cependant utilisée comme point de départ d'une méthode heuristique, pour explorer tout l'espace.

< faire les exos postés sur l'UV >

26.12.15

CORRECTION DES EXOS SUR L'UV

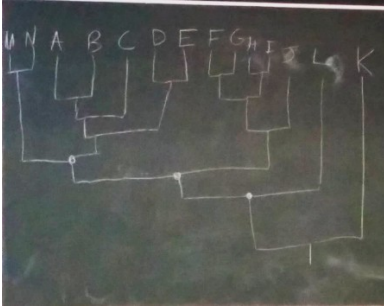
1) Arbres phylogénétiques

a. Quels sont les groupes monophylétiques :

- A et B
- FGHI
- KL
- FGHIJ

b. ancêtre de : ADKGMN --> aucun des propositions car toutes les espèces ici sont des descendants
L'ancêtre de tous se trouve ici (flèche rouge):

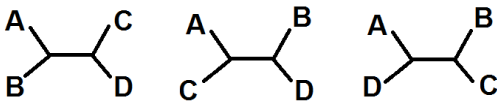
c. On refait l'arbre en partant de K.



2) Critère de parcimonie

a. *Dessinez tous les arbres non enracinés différents pouvant être construits à partir des 4 taxa.*

Pour 4 espèces, nous possédons 3 arbres non-enracinés et 15 arbres enracinés :



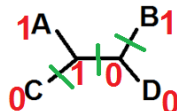
b.

1^{er} caractère : 0 pour les 3 arbres. Ce caractère ne permet pas de différencier un des arbres en se basant sur le critère de parcimonie. Quelle est la longueur de cet arbre (= combien de mutations pourraient expliquer cet arbre) ? 0 mutations, car on peut considérer qu'entre les nœuds il y ait aussi 0.

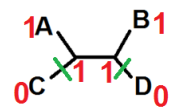
2^{ème} caractère : Dans les 3 cas, on a une longueur minimum qui est égale à 1 (une mutation). Pas de différence entre les 3 arbres.

3^{ème} caractère : $L_{C3} = 1$; $L_{C3} = 2$; $L_{C3} = 2$ -> pour le caractère 3, l'arbre 1 est le plus parcimonieux.

Dans un des cas on pourrait avoir :



mais c'est moins parcimonieux que



L'arbre le plus parcimonieux est celui qui a la longueur minimum à travers tous les caractères (pour avoir la longueur totale, il faut additionner tous les caractères).

III. Maximum de vraisemblance

Il ne s'agit pas d'une méthode d'analyse génétique uniquement : C'est une méthode de statistique générale, mise en œuvre pour l'estimation phylogénétique. Elle n'est toutefois pas restreinte à ce domaine.

La **vraisemblance** d'un arbre T correspond à la probabilité d'observer les données (sur la slide, ce sont les séquences ADN) si l'arbre T est l'arbre phylogénétique correct qui a généré les données observées. Il s'agit donc d'une probabilité, qu'on calcule pour tous les arbres possibles. L'idée est de rechercher parmi eux celui qui a la probabilité maximum d'avoir généré les données: il est considéré comme l'arbre le plus vraisemblable.

$$L(T) = \text{Prob}(D/T)$$

Prenons un alignement de séquences pour 4 espèces (slide : matrice ADN). On possède n nucléotides dans chacune des séquences. Comme dans toute analyse phylogénétique, on considère que chaque cycle nucléotidique évolue indépendamment des autres. Les mutations dans un de ces cycles n'influencent pas les autres. On peut donc les traiter séparément.

Comme on l'a montré tout à l'heure, pour 4 espèces il y a 3 arbres possibles. On doit commencer par enraceriner l'arbre au niveau d'un de ces nœuds. En l'occurrence, on a enraciné au niveau du nœud où sont présentes les espèces 3 et 4. On se focalise dans un premier temps sur le site j. On calcule la vraisemblance, la probabilité d'observer les données CCAG pour cet arbre-ci. Pour se faire, on considère tous les cas de figure possibles d'états ancestraux : on a deux nœuds ancestraux, et pour chacun on peut s'imaginer toutes les combinaisons impossibles de combinaisons de nucléotides CCAG.

La vraisemblance totale est la somme des vraisemblances de chacun des cas de figure possible. Pour chacun d'entre-eux, on calcule la probabilité que l'arbre ait engendré les données observées.

Il est nécessaire de calculer la probabilité d'absence de mutations le long d'une branche, la probabilité de mutation de A vers C le long d'une branche, puis le long d'une autre branche, etc... On réduit le calcul au produit des probabilités de transition le long de chacune des branches.

Pour calculer les probabilités, on utilise le **modèle de Jukes-Cantor**. Dans ce modèle, toutes les substitutions nucléotidiques ont la même probabilité de se produire. En se contentant d'un modèle aussi simple, on peut calculer la probabilité de transition :

$$(E) \quad L = L_{(1)} \cdot L_{(2)} \cdot \dots \cdot L_{(N)} = \prod_{j=1}^N L_{(j)}$$

$$(F) \quad \ln L = \ln L_{(1)} + \ln L_{(2)} + \dots + \ln L_{(N)} = \sum_{j=1}^N \ln L_{(j)}$$

Contrairement à la méthode de parcimonie, ici la longueur des branches compte beaucoup : au plus la branche est longue au plus la probabilité qu'il y ait une/des mutation/s est élevée.

Pour calculer la probabilité de substitution, on a besoin d'un modèle de substitution nucléotidique. Par exemple : modèle de Jukes-Cantor.

(pas connaître formules par cœur pour l'examen)

Handwritten formulas on a chalkboard:

$$L(t) = \text{Prob}(D|t)$$
$$P_{ij}(t) = \begin{cases} \frac{1}{4} + \frac{3}{4}e^{-\mu t} & (i=j) \\ \frac{1}{4} - \frac{1}{4}e^{-\mu t} & (i \neq j) \end{cases}$$

Jukes-Cantor

μt

Quand $i = j$: pas de mutation // $i \neq j$: il y a eu substitution.

μ : taux de mutation

t : temps

La longueur d'une branche (nb de mutations attendues le long de cette branche) correspond à : μt

➔ Dans ce modèle on évalue un phylogramme.

Plusieurs modèles nucléotidiques existent :

- Le plus simple : **Jukes-Cantor**. Deux caractéristiques fondamentales : 25% de chacune des 4 bases (A, C, G et T) dans les séquences et tous les types de substitution ont la même probabilité de se produire.
- **Modèle GTR**. On considère que la fréquence des bases nucléotidiques n'est pas égale (contrainte : La somme des 4 fréquences doit être égale à 1). Autre chose : Tous les types de substitution peuvent être associés à une probabilité différente. Ce modèle est réversible, donc la probabilité de passer de A à T (par exemple) est la même que celle de passer de T à A. Si ce n'est pas le cas, ça veut dire que l'ADN va voir sa composition changer au cours du temps... Impossible.
- Modèles **intermédiaires** pour passer entre Jukes-Cantor et GTR :
 - 2 types de substitution (les transitions c-à-d T vers C et A vers G ; les traversions c-à-d les autres types de mutations) : TrN

Sur la slide : Figure qui explique la différence entre l'analyse de parcimonie et celle du maximum de vraisemblance. Dans l'un ou l'autre cas, quel est l'état le plus optimal pour les arbres montrés ?

- Parcimonie : A (pour le nœud 2) // A ou C ou G (pour le nœud 1).
- Vraisemblance : Pour le nœud 1, un A serait la meilleure option car la branche séparant le nœud 1 est plus petite que la branche séparant le nœud 1 à C.

IV. Méthodes bayésiennes

iv Methode bayesiennes

La proba se pom um sête $P(H|D) = \frac{P(H) \times P(D|H)}{P(D)}$

Donnée étant donnée P l'hypothèse = vraisemblance

$P(T_i|D) = \frac{P(t_i) \times P(D|t_i)}{\sum_{j=1}^T (P(D|t_j) \times P(t_j))}$

arbre i

grand T est astronomique donc ce terme doit être calculé pour tout les arbres, c'est énorme.

$\rightarrow \text{Ratio} = \frac{P(t_1|D)}{P(t_2|D)} = \frac{P(t_1) \times P(D|t_1)}{P(t_2) \times P(D|t_2)}$

↳ = fraction

\rightarrow donc calcul un rapport entre 2 possibilités

$P(D|H)$ = vraisemblance

$P(H)$: probabilité « à priori » = hypothèse avant d'observer les données (on tient compte de données préalables ou pas)

$P(H|D)$: probabilité « à postériori » (après avoir observé les données)

$P(D)$: probabilité des données, calculée sur l'ensemble des hypothèses possibles

T_i : arbre T_i .

T : nb total d'arbres possibles

$\Sigma \dots$: probabilité de toutes les données (somme de l'ensemble des arbres possibles)

$P(T_j)$: probabilité à priori de l'arbre T_j

→ L'arbre ayant la probabilité maximum sera le meilleur arbre

Le ratio de probabilités sert à calculer deux arbres entre-eux. Pour le ratio, les deux termes T (énormes) s'annulent. Magie hehe ☺

$P(D|T_1)$: probabilité des données étant donné que T_1 est correct (= vraisemblance de T_1)

Durant le cours précédent, on a utilisé la technique des arbres voisins : on représente un espace d'arbres possible, on choisit un endroit-début, on regarde ses voisins etc... On calcule alors le ratio pour comparer cet arbre-début à ses voisins !

Deux cas :

- Soit T_1 est meilleur que T_2 : le ratio des probabilités sera plus grand que 1. On garde alors T_1 .
- Soit T_2 est meilleur que T_1 : on continue la comparaison mais en prenant T_2 comme arbre-début.

(Problèmes audio avec le podcast, à compléter la fin avec les notes



du prof ☺)

Questions d'examen :

- Comparer différents arbres phylogénétiques et utiliser le critère de parcimonie pour les différencier
- Donner un critère d'un des modèles vus plus haut (on nous donner les formules)
- Comparer le principe de méthode bayésienne et du maximum de vraisemblance : Points communs ? Différences ? Dans tel contexte, quelle est la meilleure méthode à utiliser ?
- On ne nous demandera jamais à l'examen d'implémenter manuellement une méthode bayésienne par exemple... C'est impossible sans un algorithme. Mais on nous demande de comprendre les méthodes étudiées.

De manière générale, il n'y aura pas de questions de restitution.

27 novembre 2015

Correction de questions d'examen

Si on considère le cladogramme de gauche, qui représente les relations évolutives réelles, indiquez si les interprétations de cet arbre phylogénétique qui sont reprises ci-dessous sont correctes (V/F) et pourquoi :

- le genre *Peripatus* est l'ancêtre commun de tous les autres genres représentés. **FAUX**, il fait partie des descendants car c'est une des bifurcation de l'ancêtre commun.
- Le genre *Pechygmatha* est aussi distant du genre *Scolopendra* que du genre *Daphnia*. **VRAI**, car la distance n'est pas mesurée en terme de nb de nœuds mais on calcule les longueurs des branches (nb de mutations accumulées identiques entre chaque paires d'espèces)
- Le genre *Drosophila* est plus récent que *Scolopendra*. **FAUX**, car ce n'est pas parce on a une division que nécessairement le genre *Scolopendra* est apparu à ce moment.
- Les *Drosophila* et les *Daphnia* sont plus évolués que les *Peripatus*. **FAUX**, Il a autant évolué que les autres car ils partagent tous un ancêtre commun. Le temps d'évolution séparant chacune de ces espèces de l'ancêtre commun est identique.

4) Question d'examen 2011

Une résolution complète sera postée sur l'UV via un enregistrement podcast.

IV. Méthodes bayésiennes - récapitulation

Le but d'une telle méthode n'est pas de trouver le meilleur arbre, mais d'échantillonner l'espace des arbres possibles de telle manière que l'on se focalise sur les arbres les plus probables. Au plus un arbre a une forte probabilité au plus il aura la probabilité d'être sélectionné. Avec le paquet d'arbres obtenu (=échantillon), on réalise une comparaison et on identifie les clades qui se retrouvent dans la majorité d'entre-eux. On ajoute une valeur de probabilité à ces clades, qui correspond à la fréquence de présence de ces arbres dans l'ensemble de l'échantillon (si un clade est présent dans 90% d'un échantillon, il sera associé à une probabilité de 90%).

Recherches heuristiques : se basent sur un critère de parcimonie, etc... Pour les méthodes bayésiennes, c'est différent : on estime une distribution de probabilité.

V. Biais potentiels

L'arbre obtenu via les méthodes décrites précédemment est-il le bon ? Non, car il s'agit d'estimations, on n'a aucune garantie que le passé phylogénétique a été retracé correctement. Tout ce qu'on peut faire, c'est tenter d'associer une valeur de support aux différents clades, une valeur valable que si les hypothèses explicites ou implicites faites pour développer ces méthodes sont correctes. Les modèles utilisés pour construire des arbres phylogénétiques sont des approximations (plus simples que la réalité – par définition d'un modèle).

Certaines méthodes sont soumises à des biais, qui correspondent à des hypothèses de départ des méthodes qui peuvent aboutir si on n'y prend pas garde à des résultats erronés.

Une des premières études à ce sujet : faite par Felsenstein (1978) concernant la méthode de parcimonie. A cette époque les méthode de vraisemblance n'était pas encore utilisé en analyse phylogénétique, car on n'avait pas les ressources mathématiques pour le faire. Il a identifié un cas simple avec 4 séquences pour lequel la méthode de parcimonie va systématiquement se tromper et donner un résultat biaisé.

Arbre qu'il propose :

Un arbre non-enraciné regroupe 4 séquences et est caractérisé par 2 longueurs de branche différentes (p et q). Dans un cas pareil, puisque certaines branches sont petites, entre les nœuds

ancestraux et les espèces 2 et 4 il y aura très peu de mutations (voire aucune dans un cas extrême). Dans un cas où les nœuds ancestraux contiennent des A (nucléotide), le 3 sera donc un A et le 4 sera un A aussi, car la probabilité de mutations est très petite. Les autres branches plus longues (celles menant vers 1 et 3) présentent une probabilité de mutation importante.

Pour séparer ces 4 espèces, une séquence nucléotidique typique serait :

- 1 – C
- 2 – A
- 3 – T
- 4 – A

De temps en temps, du fait qu'il n'y a que 4 états nucléotidiques possibles (ATCG), le long d'une branche la mutation pourrait aboutir à un nucléotide similaire à celui d'une autre branche. On pourrait par exemple aboutir à :

- 1 – G
- 2 – A
- 3 – G
- 4 – A

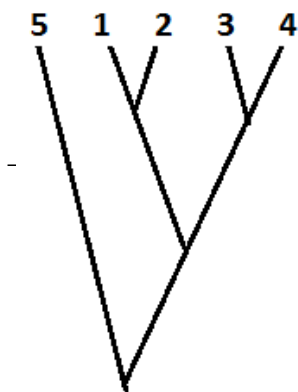
Dans le 1er cas, on ne favorise aucun arbre mais dans le 2ème, on favorise un arbre par rapport à l'autre, on y regroupera 2 et 4 d'une part et 1 et 3 de l'autre..

Même en ayant beaucoup de caractères, vu que ce phénomène de convergence évolutive se reproduit souvent, on aboutira à un grand nombre de caractères polymorphes qui vont ressembler au 3ème cas, et l'arbre favorisé sera différent de l'arbre réel (1^{er} cas).

Felsenstein a identifié la différence entre p et q nécessaire pour observer ce phénomène. Conséquence : **p^2 plus grand ou égal que $q(1-q)$** . Dès que cette condition est remplie, l'analyse par parcimonie nous mène à un résultat biaisé.

Conclusion de Felsenstein : ce qui pose problème, c'est le taux d'évolution le long des différentes branches de l'arbre, qui peut être différent. L'analyse par parcimonie fait l'hypothèse de l'horloge moléculaire, c'est-à-dire que chaque espèce évolue à la même vitesse. Si ce n'est pas le cas, la méthode par parcimonie peut amener à des résultats biaisés.

D'autres scientifiques ont ensuite montré que ce n'est pas que les espèces n'évoluent pas avec des vitesses différentes, donc le taux d'évolution n'entre pas en compte... Ils ont pris des exemples précis où, malgré une horloge moléculaire parfaite, l'analyse par parcimonie aboutit à quelque chose de faux : *Voir photo dans les notes du prof.*



Dans l'arbre proposé, la distance entre chaque séquence (5,1,2,3,4) et l'ancêtre commun est identique : L'horloge moléculaire est donc respectée. Les espèces ont évolué à la même vitesse.

Rappel : La longueur d'une branche = Temps x Taux de mutation ($\mu \times t$). La seule variable ici est le taux de mutation, mais sous l'hypothèse de l'horloge moléculaire, ce dernier reste constant à travers tout l'arbre.

Pourtant, une analyse de parcimonie va produire un arbre différent de celui-ci.

Dans l'étude faite par Hendy & Renny (1985), 12 arbres plus parcimonieux que l'arbre de départ ont été trouvés... Clairement, les résultats via la méthode de parcimonie sont biaisés. Ce qui pose problème à l'analyse phylogénétique n'est pas la différence de taux d'évolution entre les différentes branches, mais la juxtaposition de branches courtes et de branches longues.

Si la différence entre les branches courtes et longues est suffisamment grande, l'analyse est biaisée car les longues branches ont tendance à s'attirer l'une l'autre dans une analyse phylogénétique. (on parle d'artefact des longues branches dans la littérature)

Exemple de cet artefact : Insecte parasite qui ressemble à un diptère mais n'en est pas un (très petit, une paire d'ailes antérieure est transformée en altères et pas la paire d'ailes postérieure comme chez les diptères). Une analyse par parcimonie inclue cet insecte dans le clade des diptères... Mais ce n'est pas le cas. Les deux arbres en fait ont des branches longues qui s'attirent, ce qui fausse l'analyse. En utilisant une méthode de maximum de vraisemblance (qui tient compte de la longueur des branches), on a bien séparation des deux groupes.

Pendant longtemps, des simulations ont été faites pour tester différentes méthodes d'analyse phylogénétique. Sur la slide : Graphe, avec plusieurs types d'arbres pour seulement 4 taxa ou séquences. On simule des séquences le long de ces différents arbres, qu'on a testés avec différentes méthodes d'analyse phylogénétique (parcimonie, max de vraisemblance etc...).

Au final, on a regardé si dans les résultats on trouvait l'arbre initial. Sur la slide suivante : Les résultats pour la méthode par parcimonie. La courbe représente la limite au-delà de laquelle le nombre d'arbres corrects retrouvés après la méthode est inférieur à 95%. Plusieurs courbes car au fur et à mesure on teste des séquences (des sets de données) plus grandes : Au plus la taille de la séquence augmente, au plus l'analyse est fiable. Il y a une zone pour laquelle la méthode de parcimonie, même avec des séquences très longues, n'aboutit jamais à l'arbre correct.

Slide suivante (méthode de maximum de vraisemblance) : Cette zone, appelée **zone de Felsenstein**, ne pose pas problème pour l'analyse du maximum de vraisemblance. Par contre, pour la méthode du minimum d'évolution, basées sur des distances non corrigées, elle pose encore problème. En utilisant des modèles de substitution nucléotidique toutefois, on améliore la méthode. C'est possible car ces méthodes tiennent compte de la longueur des branches.

-

Autre type de biais : Lié au processus d'évolution. La majorité des méthodes d'analyse phylogénétique partent du principe que le processus d'évolution est identique à travers tout l'arbre... En comparant des espèces proches ça ne pose pas de problèmes. Mais en comparant des espèces éloignées, ça peut être problématique.

On note de grosses différences de fréquences de bases. En considérant (comme dans la plupart des modèles de substitution nucléotidique) que la probabilité de substitution dans un sens est équivalente à la proba de la même substitution mais dans l'autre sens, on ne peut expliquer ça.

En se basant uniquement sur les différences de fréquences des bases entre les espèces, on obtient le même arbre phylogénétique... Il y a de fortes chances que les analyses aient été influencées par le fait que les arbres ont des fréquences similaires. Des fréquences trop semblables mènent à plus de chances d'avoir une convergence évolutive qu'habituellement.

Une seule distance génétique tient compte de ces différences : **LogDet**. En l'utilisant, et en reconstruisant un arbre phylogénétique qui se base sur une matrice de distances corrigée par

LogDet, on obtient en effet un arbre correct.

Deux livres utiles :

- Tree thinking : An introduction to phylogenetic biology (David A. Baum & Stacey D. Smith) (2013)
- Molecular Evolution: A statistical approach (Ziheng Yang). Basé sur un traitement plus mathématique des analyses phylogénétiques (2014)

FIN DU CHAPITRE SUR L'ANALYSE PHYLOGÉNÉTIQUE

GÉNÉALOGIES DE GÈNE ET ÉVOLUTION MOLÉCULAIRE NEUTRE

Des notes concernant ce chapitre vont être postées sur l'UV.

➤ **ÉVOLUTION MOLECULAIRE NEUTRE**

Au début, sans le séquençage de gènes, les biologistes n'avaient pas accès à la variabilité génétique qui existait au niveau des séquences d'ADN et des séquences protéiques. La plupart des généticiens se basaient uniquement sur la variation morphologique, pour voir à travers elle la variation génétique au sein des espèces.

Ex : Drosophiles. Depuis l'extérieur, elles se ressemblent globalement, ce qui laissait penser que la variation génétique était faible. De temps en temps, une mutation apparaissait (cadre rouge) et soit elle était si défavorable qu'elle en était létale, soit elle était favorable et se répandait rapidement au sein de la population.

Vision de l'époque : Les populations sont peu variables génétiquement. De temps en temps une mutation apparaît et la sélection naturelle s'en occupe : Disparition si défavorable ou répartition dans la population si favorable jusqu'à fixation complète de cette caractéristique dans la population.

Deux scientifiques (Sturtevant et Dobzhansky) ont étudié les caractères moléculaires et la structure des chromosomes. Ils se sont aperçus qu'on pouvait colorer les chromosomes et faire apparaître un pattern de bandes, permettant de mettre en évidence des régions du chromosome qui ont subi des inversions. On constatait que ces inversions sont héréditaires et qu'on pouvait les suivre d'une génération à l'autre. On commence à comprendre qu'il y a plus de variation génétique que prévu...

Hypothèse des deux scientifiques pour expliquer le peu de variation visible et cette grande variation au niveau chromosomique : Cette variation chromosomique est maintenue au sein des populations par la sélection en faveur des hétérozygotes. Cette sélection favorise la diversité génétique, car elle garde deux allèles de façon identique dans une pop°.

Deux méthodes moléculaires sont ensuite apparues :

- Électrophorèse des protéines : Cette technique permet de comparer des protéines au sein d'un même individu ou entre des individus de la même population en les faisant migrer au sein d'un champ électrique à travers un gel. Au départ, on prend un morceau de tissu d'un individu et on l'écrase dans un tampon, ce qui nous donne un extrait enzymatique. On l'insère à l'intérieur d'un gel, que l'on soumet à un champ électrique. Les protéines migrent à travers ce champ en fonction de leur charge et leur masse.

Selon la composition en acides aminés (certains sont +, certains -, certains neutres), la protéine migre dans le gel à une vitesse plus ou moins importante. Ça permet de mettre en évidence des variations de séquences d'acides aminés. Au départ : Extrait protéinique conséquent. On le coupe en tranches. Chacune d'entre-elles est mise en contact avec un substrat particulier qui fait réagir un des enzymes de l'extrait. On peut alors faire apparaître la protéine à l'aide de colorant et la visualiser sur le gel.

Au final : Bandes dans le gel en fonction de la composition en acides aminés. On peut déduire le génotype des individus sur base du pattern observé sur le gel.

Le pattern de bandes peut différer d'une enzyme à l'autre : Pour certaines protéines, une seule sous-unité qui la compose. Dans d'autres cas : Deux sous-unités. Dans le cas des hétérozygotes, des sous-unités peuvent se mélanger et former des bandes intermédiaires.

Pourquoi est-ce une méthode révolutionnaire ? Avant, on se basait sur la variation morphologique et les patterns chromosomiques. Pour la première fois, une méthode identifie des locus et vérifie si il est variable ou pas. Pour certaines protéines, on peut imaginer un pattern où on a le même allèle pour toute la population, et d'autres protéines pour lesquelles on a un polymorphisme... On peut alors pour la première fois estimer la fréquence des protéines enzymatiques polymorphes par rapport à celles qui ne varient pas au sein de la population.

Comment interpréter les observations (fig 3.3 dans les slides) ? Les idées divergent... Débat : D'une part, Sturtevant et Dobzhansky disent que la variation génétique au sein des populations s'explique par la sélection en faveur des hétérozygotes. D'autre part, certains auteurs (surtout ceux qui pensaient initialement qu'il n'y avait pas de variation au sein des populations) disent que la variation génétique observée n'est pas soumise à sélection. On appelle ça de la **variation neutre**.

L'idée apparaît que même si au niveau morphologique il y a peu de variation, au niveau moléculaire il y en a beaucoup plus. Simplement, cette variation est **soumise à la dérive génique**.

- SLIDE : aa qui diffèrent entre les chaînes d'hémoglobine a de différents vertébrés-

Un deuxième type de données apparaît et accrédite la variation neutre: Le séquençage des protéines, pas au niveau des acides nucléiques mais des acides aminés. On a séquencé une protéine d'hémoglobine alpha pour différentes espèces et on a fait une comparaison. Sur la slide : Nombre de différences qui existent entre les séquences de différentes espèces.

Curieux : Entre la carpe et le requin, qui morphologiquement se ressemblent, on a 85 différences. Et entre la carpe et l'Homme, 68 différences... ?

Si on compare cela au temps d'évolution qui sépare de l'ancêtre commun (slide suivant), on se rend compte que les valeurs du tableau sont cohérentes : Au plus la distance est plus grande, au plus le nombre de différences est élevé. Il y a une proportionnalité facilement observable. Ici la distance qui sépare homme – requin est plus grande que la distance qui sépare requin – carpe.

Émergence de cette idée : L'évolution de ce type de caractères est neutre, car ces derniers accumulent un nombre de mutations proportionnel à la longueur des branches qui les sépare sur l'arbre phylogénétique. Apparaît sur base de ça l'**hypothèse de l'horloge moléculaire**, disant que le taux de mutations est plus ou moins constant à travers l'arbre. C'est cette hypothèse qui permet d'estimer les temps de divergence de ces espèces : en se basant sur le nb de mutation qui sépare le requin et l'homme, si on a une estimation du taux de mutations (dans la séquence d'acides aminés) on peut effectuer une estimation du temps auquel a existé l'ancêtre commun le plus récent entre le requin et l'homme.

Théorie de l'évolution moléculaire neutre (Kimura 1968)

Motoo Kimura (1924-1994) propose un modèle neutre : La plupart des allèles présents dans les séquences ADN/des acides aminés ne sont pas soumis à la sélection, et évoluent simplement par dérive génique. Ça ne veut pas dire que la sélection ne joue aucun rôle dans l'Évolution ! Mais, Kimura dit que l'essentiel de la variation génétique observée est imputable à la dérive génique pure. De temps en temps, une mutation et un allèle est soumis à la sélection, ce qui joue un rôle essentiel dans l'évolution de l'espèce... Mais ça ne concerne qu'une petite partie.

Comment expliquer l'existence d'une horloge moléculaire sachant que les espèces sont caractérisées par des abondances relatives différentes ? Pr qu'une mutation caractérise une espèce, il faut qu'elle s'y fixe. Cette probabilité de fixation est-elle plus petite ou plus grande dans une petite population ? Plus grande, car la dérive génique est plus forte. Dans une grande population au contraire, la probabilité de fixation diminue...

Donc à priori dans une espèce avec beaucoup d'individus, l'évolution serait plus lente car la fixation de l'allèle prendrait plus de temps ? Kimura dit que non : Quel que soit la taille de la population, l'évolution accumule les mutations de manière plus ou moins constante quelles que soit les espèces considérées. Comment peut-on l'expliquer ?

Dans les grandes populations, la force dérive génique est faible, ce qui fait que si une mutation apparaît elle ne va pas se fixer immédiatement. Hors dans une grande population la probabilité de mutation est bcp plus grande que dans une petite population, ce qui compense l'effet fixateur d'une mutation. Dans les petites populations, où la dérive génique est plus forte, l'effet inverse s'observe. Kimura a théorisé ça par un modèle mathématique, et montre que quel que soit la taille de la pop, le nombre de mutation qui va s'accumuler au sein des lignées génétiques (ou entre lignées) sera +/- le même.

→ en conclusion : le nombre de mutation est compensé par la force de la dérive génique selon la taille de la pop°.

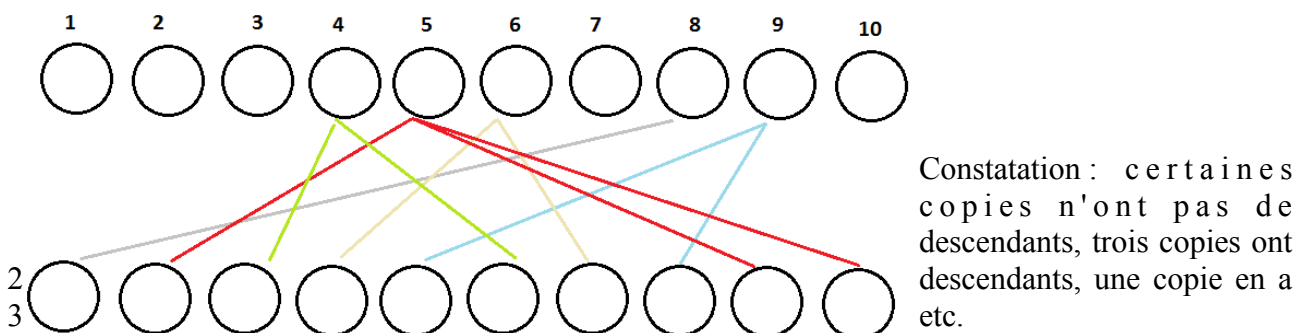
➤ GENEALOGIE DES GENES

Rappel : En voyant la dérive génique, on a utilisé un modèle simple pour faire évoluer des copies de gène de génération en génération et faire évoluer les fréquences alléliques au sein de la population. On va refaire la même chose mais cette fois en gardant l'information concernant la transmission d'une copie d'une génération à l'autre.

Rappel : toutes les copies de gène de tout individu ont des probabilités équivalentes de se reproduire, de générer une descendance. Mais cela ne veut pas dire que chaque individu a le même succès reproducteur. On a vu ça dans le cas de la dérive génique : on peut modéliser la transmission des gènes d'une génération à l'autre en échantillonnant au hasard une copie de gène de la génération précédente.

Ex : 10 copies de gènes. On tire au hasard : 8, 5, 4. On s'intéresse à "qui vient d'où ?".

On retire au hasard : 6, 9, 4. On voit déjà qu'un individu a deux descendants. Ensuite : 6, 9, 5, 5.



On pourrait de cette manière la transmission de copies de gènes de génération en génération, en gardant à chaque fois ces liens généalogiques en mémoire.

Slide : c'est ce qui se passerait si chaque individu avait un seul et même descendant à chaque génération, si le succès reproductif est identique entre tous les individus.

Slide suivant : expérience qu'on vient de faire, au hasard. En se concentrant sur la génération actuelle (ronds du bas), on remonte les générations jusqu'à trouver l'ancêtre commun le plus récent de toutes les copies présentes. Les liens généalogiques qui permettent cette connexion entre ancêtre commun et descendants actuels sont représentés en rouge. Ces liens représentent la généalogie des gènes.

On remplace ces liens par une représentation simplifiée (arbre rouge).

β ==> Deux remarques importantes :

- Ici, on prend un exemple simple car la taille de la population est petite. Donc, il ne faut pas remonter très haut pour avoir l'ancêtre commun à toutes les copies présentes... Avec une population plus grande, la probabilité que 2 copies partagent un ancêtre commun aurait été beaucoup plus faible. Faisons le calcul, pour une pop° de taille constante : 1/10. C'est la probabilité que deux copies du gène (p.e rond 9 et rond 10) partagent le même ancêtre commun (p.e rond 10).

Si maintenant on a 10.000 copies, la probabilité tombe à 1/10.000.

→ La probabilité d'avoir 2 lignées qui fusionnent est d'autant plus faible que la taille de la population est grande. Comme dans notre exemple la population est petite, ces événements de fusion apparaissent plus souvent. La hauteur de ces généalogies dans des pop° naturelle est beaucoup plus grande : comme la pop° est grande, il faudra remonter des milliers de générations avant de trouver le plus récent ancêtre commun.

- Du fait que la population soit petite, on se focalise sur tous les individus, toutes les copies de gène. Dans la réalité, vu que les populations sont plus grandes, on ne s'intéresse pas à tous les individus mais sur un échantillon.

Une **généalogie du gène** = c'est un graphe présentant les relations généalogiques entre différentes copies échantillonnées depuis la génération actuelle jusqu'à l'ancêtre commun le plus récent.

Imaginons des individus diploïdes : on a refait les mêmes exercices, mais cette fois avec des ronds qui représentent des individus, et les petits cercles représentent les 2 copies du gène (diploïdie).

Certains individus ont plusieurs descendants, mais chaque individu va donner à sa progéniture une seule copie du gène. Certains n'ont pas de descendants mais chaque individu a nécessairement deux parents.

La présence de ces deux copies dans chaque individu est transitoire qui dure le temps d'une génération : elles proviennent de même parent, mais vont se retrouver dans des individus différents.

Slide suivante : on élimine les individus, on garde juste les copies de gène.

Fig 1.18 : généalogies de gènes simulées à partir d'un même modèle simple. Ce qui est représenté est un échantillon de la population et non l'ensemble de la population. On observe que la généalogie de gène d'une simulation à l'autre varie de façon très importante, hors le modèle est identique (même taille de la pop°, etc...). Ceci dû au fait que la transmission de copies de gène d'une génération à l'autre est aléatoire.

C'est important car si on compare la généalogie de deux gènes qui se trouvent dans la même population chez les mêmes individus, alors l'échantillon d'une dizaine ou vingtaine d'individus, on séquence les deux gènes sur deux chromosome différents (chromosome 1 et chromosome 2). On s'attend à quoi ? Est-ce que leur histoire de transmission d'une génération à l'autre va être corrélée ?

- × Ces gènes viennent de parents différents.
- × L'histoire de transmission de gène varie fortement d'un chromosome à l'autre !

Par contre s'il y a recombinaison, l'histoire de transmission commune va être "cassée" même au sein d'un même chromosome. **Le génome d'un organisme est constitué d'une mosaïque de gène avec une généalogie différente.**

03 décembre 2015

résumé de la semaine passée. Fig. 1.18 : six arbres qui ont été simulés en utilisant le même modèle, les mêmes paramètres. Ces arbres présentant des grandes différences qui représentent le caractère stochastique de la transmission des copies de gène d'une génération à l'autre à travers l'histoire des populations.

Cette différence de généalogie pourrait en fait caractériser différents morceaux ou fragments de notre génome : quand on regarde le génome d'un organisme, l'ensemble du génome n'est pas caractérisé par une seule et même généalogie, ceci parce que différents morceaux 'un génome n'ont pas été transmis de la même façon au cours du temps.

Slide suivant : les rond sont les copies de gène et non plus l'individu. Chaque individu (ici tous diploïdes) à un succès reproducteur différent, il peut ou non se reproduire. Tout individu à la même probabilité de se reproduire.

Prenons maintenant en considération qu'une copie du gène, ici la copie en rouge : si on remonte le temps, on voit que pour chaque copie il y a deux possibilités (chaque individu a un parent, mais une copie peut venir soit du père soit de la mère). On finit par trouver la génération ayant la copie ancestrale de toute celles qui ont été échantillonnées. On peut retracer le chemin généalogique (lignes rouges) !

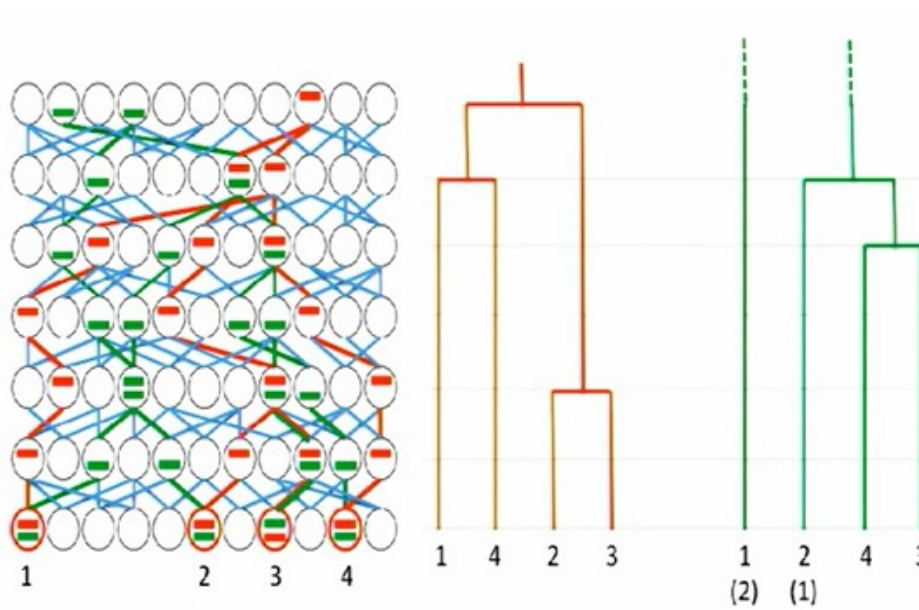
On refait la même chose mais avec un autre gène (en vert). De nouveau on choisit un des deux parents de façon aléatoire, jusqu'à trouver les liens généalogiques pour ce gène.

/!\ Quand on a deux copies qui se rejoignent dans un individu, cela ne veut pas dire qu'il y a un événement de coalescence (pas de fusion des lignées à ce niveau-là), tout simplement parce que les deux copies sont présentes dans l'individu.

L'association de deux copies de gènes qui se retrouvent dans un même individu est une association transitoire, car si on regarde la génération précédente ils ne sont pas ensemble et la génération suivant non plus.

On peut redessiner la généalogie du gène de façon schématique. Deux constatations :

- Les relations entre les individus (ou mieux entre les copies de gène de ces individus) sont différentes : Pour le gène rouge, 2 et 3 sont regroupés ensemble
- La génération de l'ancêtre commun le plus récent n'est pas la même. Dans le cas du 2ème gène, la génération n'est pas encore atteinte : Il faut remonter plus loin dans le temps.

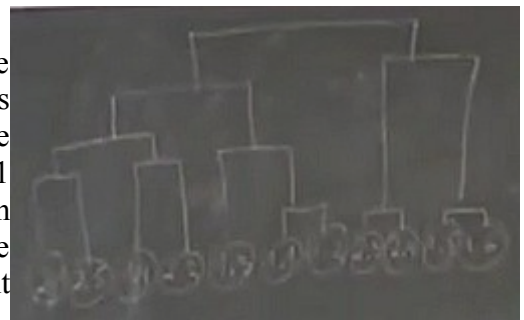


Quand on s'intéresse à l'histoire évolutive, on ne peut considérer qu'il y a une seule généalogie pour l'ensemble du génome : Chaque morceau génomique a subi une histoire différente et indépendante des autres... Sauf dans un cas, où deux gènes partagent exactement la même généalogie : S'ils se trouvent sur le même chromosome, sans qu'il n'y ait d'événement de recombinaison à un moment donné qui sépare les deux gènes (s'il y a recombinaison, leurs histoires divergent et leur généalogie va être totalement ou partiellement différente).

Si les généalogies de gène sont si différentes d'un gène à l'autre, cela a-t-il un sens de se baser uniquement sur **un** gène pour établir/estimer un arbre phylogénétique ? Il est important de faire le lien entre l'**arbre phylogénétique** et la **généalogie de gènes (ou arbre des gènes)**.

Sur la slide, nous avons représenté un événement de spéciation, une divergence qui commence au temps t_1 . Après division, les deux espèces se différencient et évoluent chacune de leur côté. Les individus sont séparés en deux groupes (du fait par exemple d'une barrière géographique). En se plaçant juste au moment de la mise en place de la barrière, on a une série de copies de gènes qui se retrouvent dans l'espèce 1, une autre série dans l'espèce 2.

Si on regarde en détails les relations reliant ces copies de gènes entre-elles, on constate qu'il n'y a pas différentes copies d'une espèce qui forment un groupe monophylétique (*voir dessin*)... On constate qu'il y a des copies de l'espèce 1 qui sont regroupées avec des copies de l'espèce 2. En d'autres termes, la copie 1.4 est plus proche d'une copie de l'espèce 2 que d'autres copies de l'espèce 1. Comment l'expliquer ?



Considérons que l'espèce est une grande population panmictique (distribution des individus aléatoires). La subdivision se fait aléatoirement : Aucune raison ne permet d'imaginer que toutes les copies qui se retrouvent pas hasard dans le territoire de l'espèce 1 sont plus proches les unes des autres, que les copies qui se retrouvent dans le territoire de l'espèce 2.

Après plusieurs années, ce sera pareil ? Est-ce que certaines copies de l'espèce 1 pourraient-elles être plus proches des copies de l'espèce 2 que des autres copies de l'espèce 1 ? Qu'est-ce qui peut faire évoluer cette situation... Les mutations ?

Les mutations font que la distance entre les différentes lignées s'agrandit... Mais les relations entre les lignées elles restent les mêmes.

Certaines des lignées, de par la dérive génique, vont disparaître au cours du temps (pour rappel, au plus la population est grande, au plus ce processus prend du temps). À un moment donné si on attend suffisamment longtemps, toutes les copies échantillonnées partagent un ancêtre commun. Uniquement à ce moment-là, toutes les copies de l'espèce 1/de l'espèce 2 forment un groupe monophylétique.

Ça nous donne un moyen intéressant pour évaluer depuis quand deux espèces se sont séparées. On voit qu'au temps t_3 , slide suivante, toutes les copies du gène de l'espèces 1 sont monophylétiques, même chose pour l'espèce 2.

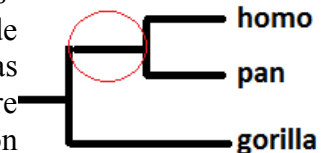
Le processus de formation de la généalogie peut être formulé de la même façon que le processus de ... pour peu que le gène soit neutre, c'est-à-dire que tous les nouveaux allèles qui apparaissent sont ni plus ou moins avantageux que ceux existant dans la population. Si c'est vrai, dans un premier temps

on peut simuler la généalogie (transmission des copies de gène d'une génération à l'autre) puis rajouter les mutations de façon aléatoire mais en tenant compte de la longueur des branches (plus une branche est longue, plus la probabilité qu'une mutation s'y insère est élevée).

Les mutations sont représentées par des croix oranges (slide suivant). On voit des mutations qui engendrent des nouveaux allèles dans la population, mais ces nouveaux allèles peuvent disparaître par la suite via la dérive génique. Par contre, les mutations engendrent un polymorphisme au sein de l'espèce, mais elles sont fixées dans la population.

Sur la slide : Phylogénie entre 3 espèces. Les espèces 1 et 2 partagent un ancêtre commun (1 et 2 forment un groupe monophylétique). Deux situations différentes par rapport à la généalogie de gène. Imaginons que ça soit deux gènes différents :

- À droite : Cas le plus simple. Les différentes copies de l'espèce 2/espèce forment un groupe monophylétique. En prenant n'importe quelle copie de gène de l'espèce 1/l'espèce 2/l'espèce 3 et qu'on établit un arbre phylogénétique, on retrouve un arbre de gènes identique à l'arbre des espèces.
- À gauche : La copie H5 est plus proche des différentes copies de l'espèce 1 que de celles de l'espèce 2. Les deux copies de gènes de l'espèce 2 sont plus proches des copies de l'espèce 3 que l'autre copie de l'espèce 2. Suivant la copie que l'on prend, on retrouve un arbre qui regroupe l'espèce 2 et 3 d'abord, puis par après l'espèce 1. Dans ce cas, l'arbre du gène ne correspond plus à l'arbre des espèces. Dans quelles situations on aura un arbre du gène différent de l'arbre des espèces (dans un cas d'un gène neutre) ?
 - Si la taille de la population est petite, le phénomène de dérive génique est accéléré et un conflit entre l'arbre du gène et l'arbre des espèces est moins probable
 - Si la branche est très longue, la probabilité que l'événement de coalescence des copies aille au-delà de ladite branche est fatalement plus faible. On connaît un exemple concret où ce phénomène peut se produire : les différentes relations entre les Primates avec les différents genre (Homo, Pan, Gorilla). La branche qui sous-tend Homo et Pan est très courte, donc tous les événements de coalescence entre les copies d'Homo et de Pan ne sont pas terminés et peuvent aller au-delà (en remontant vers l'arrière dans l'arbre). Ça explique pourquoi dans certains cas on regroupe et le chimpanzé et le gorille ensemble.



β → en résumé == condition pour qu'il y ait conflit entre les deux arbres : Les différentes copies du gène au sein des deux espèces ne coalescent pas avant le deuxième événement de spéciation. Paramètres : Si on a une branche très courte et si la population est grande, on a une probabilité plus élevée d'avoir un conflit entre l'arbre du gène et l'arbre des espèces.

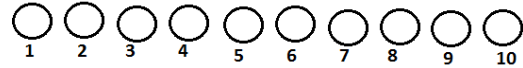
Les généalogies de gène sont également étudiées pour obtenir des infos sur l'évolution des populations naturelles. Sur la slide : Deux exemples schématiques montrant l'évolution d'une taille de population de taille constante au cours du temps dans le 1er cas et une population de taille qui varie brusquement et rapidement dans le 2ème cas. Ces deux types d'histoire évolutive influencent la forme de la généalogie.

Même chose sur la slide suivante – Fig 1.2 : Si une population de taille constante reste telle quelle (arbre de gauche) ou se divise en deux (arbre de droite), dans l'un ou l'autre cas la forme de la généalogie évoluera différemment. La séparation en 2 populations isolées nous dirige vers une situation où les différentes copies d'une population forment un groupe monophylétique par rapport aux copies de l'autre population.

Fig 4.2 et 4.3 : on a réalisé une simulation dans deux cas différents. Dans le 1^{er} cas la population a

une taille constante, dans le deuxième cas la population voit sa taille augmenter de façon exponentielle au cours du temps. Les différentes généalogies varient fortement entre-elles car le processus de transmission des gènes est un processus aléatoire/stochastique, mais par contre on voit quand même des caractéristiques communes. Différence fondamentale :

Dans la figure du haut, la fusion des lignées se fait essentiellement dans les temps les plus récents et la dernière fusion se fait beaucoup plus tard. Dans la figure du bas, l'événement de coalescence se fait au début de l'arbre.



Imaginons 10 copies de gène et une population de taille constante. Quelle sera la proba que deux de ces copies (1 et 2) partagent un ancêtre commun à la génération précédente ? $1/10$. La proba que 3 et 4 ait un ancêtre commun est aussi $1/10$ etc.

Au plus la taille de la population augmente, au plus la probabilité de coalescence (fusion des lignées) diminue. Car la probabilité de coalescence est $1/\text{taille de la population}$. C'est ce qui explique la différence entre les deux généalogies sur la slide.

Slide suivant : l'arbre suggère que l'origine de l'homme s'est faite en Afrique.

En 1987, Cann et al. ont comparé l'ADN mitochondrial de 147 hommes provenant de plusieurs régions du monde. La généalogie reconstruite à partir de ces séquences montre un ancêtre commun le plus récent (MRCA) dont l'âge est estimé à $\pm 200,000$ ans (140-290 ka). Cela suggère que :

[il faut savoir que le génome mitochondrial est transmis QUE par la femme]

- 1) L'ensemble des femmes actuelles descendent toutes d'une seule femme qui vivait il y a $\pm 200,000$ ans

Faux ! Ce qui est vrai pour l'ADN mitochondrial n'est pas vrai pour un autre gène du génome. En regardant ce qui se passe sur les chromosomes 1 et 2, l'ancêtre le plus ancien d'un entre-eux pouvait être présent à une génération il y a 400,000 ans, un autre à une génération il y a 500,000 ans... Chacun a suivi une histoire différente.

- 2) Il y a $\pm 200,000$ ans, la population humaine s'est réduite considérablement et ne comprenait qu'une seule femme.

Faux ! De nouveau on ne tient en compte que l'ADN mitochondrial ici, et pas les autres gènes du génome. On ne peut donc affirmer cela sans prendre en considération tout le génome.

- 3) L'ensemble des génomes mitochondriaux de la population humaine actuelle descendent tous d'une seule femme, qui vivait il y a $\pm 200,000$ ans

Vrai ! Si 147 copies c'est suffisant pour représenter la population, cette affirmation est vraie. Cette femme par contre ne possédait pas l'ancêtre commun de toutes les autres copies du génome ! Elle possédait juste l'ancêtre commun du génome mitochondrial.

- 4) Si on échantillonne 10 copies de n'importe quel gène dans la population humaine actuelle, l'ancêtre commun le plus récent partagé par ces 10 copies est âgé de moins de $\pm 200,000$ ans

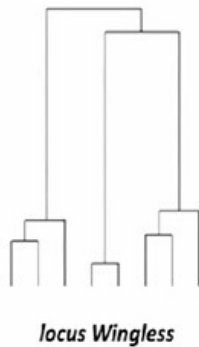
Faux ! Vrai que si les 10 copies échantillonnées ont été échantillonnées de façon biaisées.

- 5) Si on échantillonne 10 copies d'un locus mitochondrial parmi la population humaine mâle actuelle, l'ancêtre commun le plus récent partagé par ces 10 copies est âgé de moins de $\pm 200,000$ ans

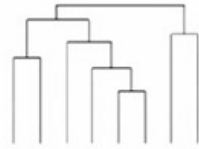
Ça dépend !

Les séquences de deux locus (gène Wingless et gène Adh) sont échantillonnées chez 4 individus

d'une population de *Drosophila heteroneura*, localisée sur une île de l'archipel Hawaï. Ces séquences sont reliées par les généalogies suivantes (une généalogie par locus ; même échelle de temps) :



locus Wingless



locus Adh

a) comment pourrait-on expliquer la différence de hauteur entre les deux généalogies ?

-Un effet de sélection, car elle agit sur un gène et pas un autre.

-Si les gènes se trouvent sur deux chromosomes différents ou sont sur le même chromosome mais séparés par une distance suffisamment grande pour qu'il y ait recombinaison entre eux, leur histoire généalogique sera différente.

b) comment pourrait-on expliquer les différences de relations généalogiques reliant les séquences, observées entre les deux généalogies ?

-même chose, si les gènes se trouvent sur deux chromosomes différents.

4 décembre 2015

EVOLUTION DES GENOMES

Table 13.3 : nombre de nucléotide dans les différents génomes d'être vivants. La taille des génomes varie très fort. L'espèce qui a le génome le plus grand est une amibe, l'homme a une taille de ~3 milliards de nucléotide. Pourquoi une telle variation et pourquoi il n'y a pas de relation entre la complexité d'un individu et la taille du génome ?

Figure C-value : représente l'étendue des génomes de différent organismes. On voit que les mammifères ont un génome d'un range important.

Tableau suivant : on a évalué le nombre de gènes codants, qui varie beaucoup moins que la taille du génome. → les variations de tailles du génome concernent peu le nombre de gènes codants mais plutôt tout le reste du génome.

Arbre génétique de différents groupes de plantes. On constate que l'augmentation de taille du génome s'est produite de façon indépendante dans les clades. On a également une diminution de la taille du génome dans quelques clades.

Globalement, le génome a tendance à devenir de plus en plus grand au sein des plantes.

Structure répétitive de l'ADN génomique

(Cette figure a été réalisée après avoir coupé l'ADN génomique en plusieurs morceaux, et après l'avoir chauffé. À plus de 90°C, l'ADN double-brin se dénature en simple brin. En mesurant le pourcentage des fragments réassociés, on réalise la courbe sur la slide. Les différents paliers observables sur le graphique correspondent à différents types d'ADN, différents types de fragments. On constate que la courbe n'est pas régulière mais comporte des paliers, qui correspondent à différents types d'ADN. On peut alors classer les types d'ADN en plusieurs groupes :

Foldback DNA : Séquences ADN palindromiques dont une partie correspond au reverse-complement de l'autre. Ces séquences simple brin peuvent donc se refermer sur elles-même pour former une structure en épingle à cheveux.

Highly repetitive DNA : Hautement répétitif. Ces petites séquences sont répétées plusieurs milliers à plusieurs millions de fois dans le génome.

Middle-repetitive DNA : Séquences plus longues de plusieurs milliers de nucléotides qui sont répétées de quelques dizaines à quelques milliers de fois dans le génome.

Single-copy DNA : Une seule copie présente dans le génome.

Pourquoi le Foldback se réassocie plus vite ? Les simples copies mettent plus de temps à trouver leur brin complémentaire dans le mélange. Le foldback DNA quant à lui n'a qu'à se replier sur lui-même.

On parle d'ADN satellite car, auparavant, en faisant une extraction d'ADN on réalisait une centrifugation dans un gradient de chlorure de césium. On obtenait ainsi la bande principale et les bandes satellites qui correspondaient à ces séquences répétées un certain nombre de fois.

Les microsatellites quant à eux sont des petits nombres de nucléotides répétés plusieurs fois.

Mécanismes responsables de l'évolution de la taille des génomes

Un des mécanismes les plus importants : Crossing-over inégal. Figure 1.14, exemple : 2 chromosomes homologues se réassocient au mauvais endroit. L'échange d'ADN entre les deux se

fait de telle manière qu'au final, deux copies de l'un se retrouvent sur l'autre. Cette erreur est rare sauf dans le cas d'ADN répétés, puisque les deux morceaux sont identiques... Donc le mauvais appariement se fait plus facilement sur ces deux copies identiques. Cette mutation fait apparaître un fragment supplémentaire sur un des brins et une délétion au niveau de l'autre brin.

Figure 1.15 : Cas de duplication double brin. On a un segment TATA (fragment répété). Dans le premier cas, un TA précède TATA. Dans le deuxième cas à droite on observe une délétion.

Slide suivant : Glissement entre les deux séquences, des enzymes de correction viennent corriger la boucle au niveau du double brin.

Mécanismes responsables de l'évolution de la taille des génomes

Le graphe représente la composition du génome humain. Les gènes qui codent pour les protéines représentent une petite fraction de ce génome... La plus grande est composée d'éléments transposables (presque 50% du génome), capables de se répliquer eux-mêmes au sein du génome : Ils peuvent donc être eux aussi responsables de l'évolution de la taille du génome.

Il existe deux grands types d'éléments transposables (représentés sur la slide) qui se différencient en fonction de leur mécanisme de transposition :

- Classe 1 : Passent par la formation d'un transcrit ARN. Ce dernier, par une enzyme (la transcriptase reverse) va être transformé en ADN, qui sera de nouveau intégré dans le génome. Ça engendre une duplication.

→ Les **rétrotransposons** (caractérisés par les **LTRs**)

→ Les **rétrotransposons LINEs** et **SINEs** (les premiers sont autonomes et peuvent se transposer par eux-mêmes, les seconds ont besoin de la machinerie enzymatique des LINEs pour se transposer)

→ Les **introns mobiles**, capables de se déplacer et aller se mettre dans d'autres gènes codants.

- Classe 2 : Ils sont transposés au cours de la réplication. On reste au niveau de l'ADN, pas de passage par un stade ARN. Au cours de la réplication, un des deux segments est transposé dans le segment en cours de réplication. Au final, on se retrouve sur l'un des deux brins avec la copie insérée mais pas la copie de départ. → **Les transposons**

L'origine évolutive des éléments transposables serait des copies qui nous viennent du génome bactérien. On pense que les rétrotransposons et les transposons ont une origine purement bactérienne, ce qui veut dire que ces éléments transposables de bactéries ont été transmis au génome eucaryote. Les rétrotransposons par contre ont une origine hybride et sont composés d'éléments transposable bactériens et de segments provenant de l'organisme qui les porte.

Slide arbre génétique : Transferts horizontaux == Des éléments transposables (ici éléments P) sont passés d'une espèce de drosophile à l'autre. Leur présence dans les deux espèces n'est pas due à une origine commune (sinon ça aurait été un transfert vertical).

On utilise les SINEs souvent pour construire des arbres phylogénétiques. Une fois intégrés dans un morceau de génome, les SINEs ne peuvent se retirer car ne sont pas autonomes comme les LINEs.

Intéressant car :

- Peu de convergence évolutive possible
- Une fois l'intégration faite, s'il est tout de même retiré (peu probable) ça ne sera pas fait « proprement » : Soit un morceau du SINE reste, soit un morceau de la région manquante v être retiré.

Une fois qu'on a isolé et identifié la présence d'un SINE dans le génome d'un organisme, en faisant une PCR (amplifier un fragment contenant le SINE), on peut utiliser le caractère « présence du SINE » pour regrouper deux espèces dans un même clade. On voit la slide que le SINEL1 est

présent dans l'espèce 1 et l'espèce 2 mais pas dans la 3 ni la 4, on regroupe alors 1 et 2 dans un même clade.

Duplication de gènes : Nous parlons de gènes codants. Une fois qu'ils sont dupliqués, une des deux copies n'est plus nécessaire et ne sera pas maintenue par sélection : Elle va évoluer et perdre sa fonction (à travers une mutation). On peut observer dans le génome plusieurs gènes qui ont été codants à un moment donné mais ne le sont plus, les pseudogènes. Notion de « junk DNA » : Une grande partie du génome a joué un rôle important à un moment donné, mais aujourd'hui (de façon transitoire), ce rôle fonctionnel n'est plus indispensable (n'est plus contraint par la sélection).

4 hypothèses permettent d'expliquer le maintien d'une quantité d'ADN non génique au sein du génome :

- 1) L'ADN non génique réalise des fonctions essentielles encore mal comprises
- 2) L'ADN non génique est de l'ADN inutile (« junk DNA ») pour l'instant (il pourrait acquérir des fonctions nouvelles par la suite). Il est transformé de façon passive par le chromosome, car relié physiquement à des gènes fonctionnels. L'ADN en excès n'a pas d'effet sur l'aptitude (fitness) des individus qui les portent
- 3) L'ADN non génique est parasite, sans utilité pour l'organisme qui le porte (« ADN égoïste ») qui s'accumule et est maintenu par sélection intragénomique
- 4) L'ADN non génique a une fonction de structure c-à-d une fonction reliée à la détermination du volume nucléaire, mais non reliée au transport de l'information génétiquement

Ces hypothèses ne sont pas mutuellement exclusives.

Hypothèse du junk DNA : Une grande partie du génome n'assume pas une fonction utile. Cette hypothèse a fait l'objet d'un débat récemment suite à plusieurs études sur le séquençage du génome humain.

En 2001, première version d'un génome humain entièrement séquencé. Suite à ça, un consortium de laboratoires dans le monde a lancé le projet **Encode** dont l'idée est de déterminer les gènes fonctionnels dans l'ensemble du génome séquencé l'année précédente. Une des conclusions de ce consortium est que l'hypothèse du junk DNA est fautive.

Ils concluent que 80% du génome assume une fonction biochimique. L'hypothèse du junk DNA serait donc dépassée... Suite à ça, de nombreux articles dans la littérature scientifique de spécialistes de l'évolution moléculaire ont remis en question ces conclusions.

Le consortium Encode s'est rétracté par après sur le pourcentage d'ADN fonctionnel dans le génome : 80% ? 40% ? 20% ? Ils ne sont pas sûrs...

On a remis en cause leur définition de « fonction ». Rappel : TATABOX = fragment d'ADN sur lequel les facteurs de transcription viennent s'attacher, en amont des gènes transcrits. Les membres du consortium ont recherché les séquences ressemblant à une TATABOX dans le génome et, dès le moment où ils en trouvaient une, ils estimaient être en face d'un génome fonctionnel. Or, dans le génome, certaines séquences ressemblant à une TATABOX peuvent apparaître au hasard, et peuvent même accueillir un facteur de transcription, mais ce n'est pas pour autant que ça activera la transcription de quelque chose... Les procédures utilisées par le consortium semblent donc fautes.

Un segment d'ADN qui présente une propriété d'un élément fonctionnel (être transcrit, liaison à un facteur de transcription, ...), n'est pas nécessairement fonctionnel. Tous les chats ont 4 pattes, donc tous les organismes à 4 pattes sont des chats ? ... Il y a une erreur de logique.

Les contestataires du consortium disent que le seul moyen de démontrer qu'un élément est fonctionnel est de mettre en évidence que l'évolution de sa séquence est contrainte par sélection. Sans preuve d'une sélection qui empêche la nouvelle séquence d'être proche de la séquence initiale, impossible de dire que l'élément analysé est fonctionnel.

D'autres séquences peuvent être responsables de la différence de taille du génome :

- Apparition/Délétion d'introns dans les gènes codants.
- Les réarrangements chromosomiques : mutations à large échelle. En comparant le génome humain au génome des primates, le nombre de chromosomes chez ces derniers est de 48 (alors qu'il est moindre chez l'Homme) probablement du à la fusion de chromosomes, qui pourrait expliquer que le génome humain est plus petit que celui du chimpanzé.
- Polyploïdie : Mécanisme courant chez les plantes. C'est une « mutation » qui duplique le génome au sein de l'organisme (passage d'une diploïdie à une tétraploïdie par exemple). Dans un premier temps, l'individu n'a pas d'information génétique supplémentaire ! Cette information est juste présente un plus grand nombre de fois (d'ailleurs, la plupart du temps les tétraploïdes doivent consacrer plus d'énergie à la réplication de l'ADN pour peu de bénéfices, et sont donc moins « optimisés » que les diploïdes). Cette polyploïdie est appelée autopolyploïdie.

Allopolyploïdie : Hybridation entre deux espèces différentes, où l'on met en commun le patrimoine génétique de deux espèces différentes, mettant ainsi en évidence des complémentarités qui parfois sont intéressantes pour l'hybride.

Au cours de l'Évolution, on constate que certains états tétraploïdes ont été conservés, permettant l'augmentation de l'information génétique dans le génome des organismes. On a pu le mettre en évidence par observation dans le génome de grosses portions localisées en des endroits différentes mais très proches l'une de l'autre : Des régions qui provenaient d'une duplication très ancienne du génome complet.

Pourquoi ce phénomène est plus courant chez les plantes que chez les animaux ? La plante peut s'autoféconder, donc si un tétraploïde apparaît dans la population, il peut se reproduire avec lui-même. Au contraire chez les animaux, le tétraploïde ne pourra se reproduire avec personne, et n'aura aucune descendance.

Duplication de génomes complets

Il semble que les duplications de génome sont souvent associées à l'apparition de grandes innovations. Ce qui a permis l'apparition des vertébrés et toutes leurs spécificités sont les événements de duplication, et il semble que ça soit la même chose pour les plantes.

Slide suivant : uniquement pour les plantes. En comparant les événements de duplication de génome pour différentes espèces à travers un arbre phylogénétique, on constate que beaucoup de ces duplications ont eu lieu dans une période de temps assez courte (barrière entre Crétacé et Phérozoïque : une soixantaine de millions d'années). Pourquoi, à cette époque spécifiquement, a-t-on eu autant d'événements de duplication ?

*****Suite des exercices du cours précédent*****

L'arbre phylogénétique qui relie 6 espèces est estimé par l'analyse séparée des alignements de séquences d'ADN de 4 gènes différents. Le résultat de ces 4 analyses est présenté ci-dessous.

(a) D'après vous, quel est l'arbre phylogénétique qui relie ces 6 espèces ?

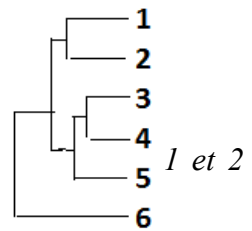
2

**les gènes 2 et 3 sont vraisemblablement des phylogénies d'espèces. Les autres sont des arbres des gènes car ils varient fortement.*

(b) Qu'est-ce qui peut expliquer les estimations contradictoires observées,

- 1) de façon générale
- 2) spécifiquement pour cette phylogénie

Le phénomène des lignées généalogiques : La fusion des lignées n'est pas nécessairement terminée au-delà de 2 événements de spéciation. Pour cette phylogénie, ça concerne la branche sous-tendant le clade contenant les espèces qui est particulièrement courte.



Trois gènes différents sont séquencés chez un bonobo et un chimpanzé, dans le but d'estimer la période à laquelle ces deux espèces ont commencé à diverger de leur espèce ancestrale commune. Sur base de l'estimation du nombre de mutations qui séparent les séquences des deux espèces, et d'un taux de mutation supposé connu pour ces trois gènes, on obtient les estimations suivantes pour les temps de divergence associés aux trois gènes : $0.8 \cdot 10^6$ années, 10^6 années, et $1.2 \cdot 10^6$ années.

(a) Si on fait l'hypothèse que les différences entre les valeurs observées ne peuvent pas être expliquées seulement par l'erreur liée au processus d'estimation, comment peut-on expliquer ces différences ?

-Via la sélection

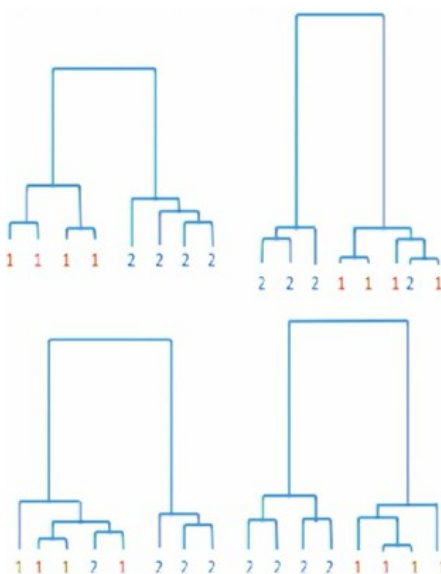
-On peut poser la question différemment : est-ce que l'ancêtre commun des deux espèces date exactement de quand les deux espèces se sont séparées ? Non cela dépend de l'histoire de transmission des copies de gène.

(b) Quelle(s) hypothèse(s) a-t-on dû faire pour calculer ces trois estimations ?

L'hypothèse du taux de mutation constant, en d'autres termes celle de l'horloge moléculaire. On considère ainsi que le gène est neutre. S'il était soumis à sélection, le temps de sélection serait allongé ou ralenti.

Deux populations de drosophiles sont isolées chacune sur une île différente de l'archipel Hawaï. Pour vérifier si les deux îles échangent des migrants, on séquence 4 copies de 4 gènes différents par île, et on estime leurs relations phylogénétiques. Les 4 arbres de gène sont représentés ci-dessous.

Que peut-on en conclure concernant la migration entre les deux îles ?



- La convergence évolutive pourrait être une explication : des séquences de la population 1 sont très proches des séquences de la population 2. Cela dépend des mutations qui ont eu lieu. Ici les branches sont assez longues, il y a donc plus de probabilité de mutations dans ces branches.

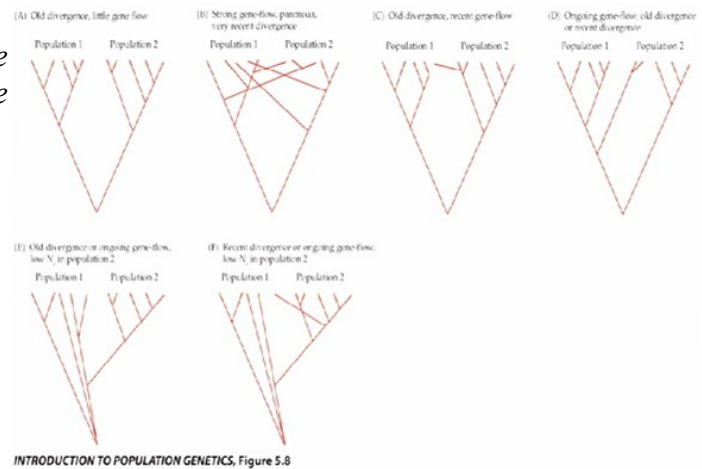
- la dérive génique a formé deux groupes monophylétiques : pour les gènes a et d, toutes les séquences de la population 1 forment un groupe monophylétique, et toutes les séquences de la population 2 forment un groupe monophylétique. Même que les séquences de ces 2 groupes sont séparées par de longues branches (bcp de mutations). Cela implique que les deux populations soient séparées par les deux îles depuis longtemps.

- En fait il y a d'abord eu un isolement, suffisamment longtemps pour qu'il y ait formation des deux groupes monophylétiques. Le phénomène de migration que l'on voit que pour deux arbres est précise : de l'île 1 vers l'île 2. C'est une migration récente, car si elle était ancienne on devrait voir le 2 qui se rattache à la branche avant (le 2 devrait être "plus ancien"). Comme il y a eu cet

isolement et la dérive génique, cette séquence aussi proche des séquences 1 soit présente dans l'île 2 s'explique plus facilement par un phénomène de migration récente. → l'une des séquences de l'île 1 se retrouve dans l'île 2.

Si dans le futur, aucune migration ne se produit et qu'on attend suffisamment longtemps, on finira par retrouver cette monophylie au sein des deux populations !

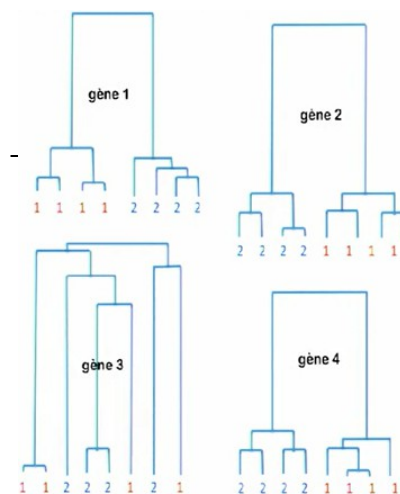
Pour mieux comprendre la différence entre migration récente et ancienne, voir slide précédente-->



Mus macedonicus (1) et Mus spicilegus (2) se sont séparées il y a environ deux millions d'années. Un biologiste séquence 4 gènes chez les deux espèces (4 séquences par espèce), et estime l'arbre phylogénétique qui les relie. Les 4 arbres obtenus sont présentés ci-dessous.

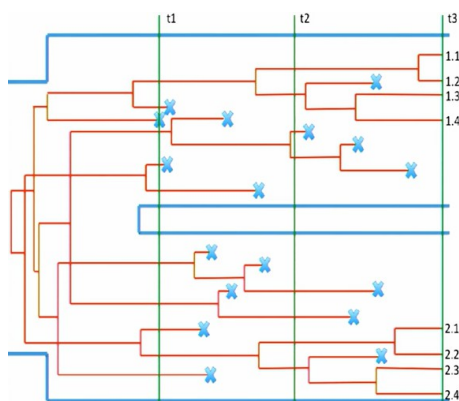
Comment peut-on interpréter ces observations?

-explication peu probable:



Un sélection adaptative qui favorise une mutation apportant une adaptation par rapport au milieu, accélère le processus de tri des lignée et donc accélère le processus qui se fait normalement par dérive génique (accélère la monophylie des espèces)

peu probable si les gènes sont choisis au hasard, peu de probabilité de choisir des gènes impliqués dans l'adaptation !



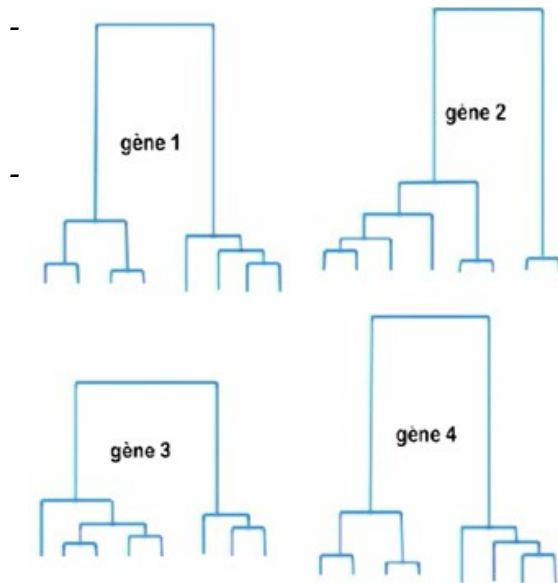
- Référence à la figure ci dessus : Quand les espèces divergent, les lignées sont mélangée. Au fur et à mesure que le temps passe, on finira par voir que toutes les copies de l'espèce 1 sont plus proches entre-elles qu'elles ne le sont des copies de l'espèce 2.

- Pour 3 gènes, on observe cette situation, par contre pour le gène 3 c'est différent. Le phénomène qui pourrait ralentir ou empêcher le phénomène de monophylie chez 1 gène sur 4 serait :

- la sélection en faveur des hétérozygote, qui maintient la diversité.
- Accouplement préférentiel pour individu de phénotype différent. (même histoire que les gènes MHC, qui

présentent une diversité génétique chez les hommes et les primates. Voir cours précédents).

Huit séquences de 4 gènes différents échantillonnées dans une population de dauphin de l'océan atlantique. Les 4 arbres de gènes estimés à partir de ces séquences sont présentés ci-dessous. Peut-on en déduire quelque chose par rapport à la position de ces 4 gènes dans le génome ?



Les gènes 1 et 4 ont des généalogies identiques= ils partagent donc la même histoire de transmission, et la seule explication possible est qu'ils soient sur le même chromosome et proches.

dans toute leur histoire de transmission depuis leur ancêtre commun, les gènes n'ont pas subi de recombinaison !

10 décembre 2015

Spéciation : ensemble des mécanismes qui vont permettre la création de nouvelles espèces (= mécanismes qui génèrent la diversité biologique)

La différenciation en espèces selon le critère morphologique n'est pas évidente : Fig22.2, représente les relations entre les espèces du genre *Partula* sur une île. (S = coquille lévogyre et D = coquille dextrogyre) Deux situations différentes sur l'île : d'une part dans quelques parties les deux espèces s'hybrident, dans d'autres elles restent séparées. D'autres encore sont reliées par un continuum morphologique.

Espèce : définition biologique de l'espèce (Dobzhansky, Mayr, Wright ~1940), basée sur :

! Critère d'isolement reproductif : deux individus appartiennent à deux espèces différentes s'ils ne peuvent plus se reproduire entre-eux dans les conditions naturelles.

Néanmoins, la définition biologique de l'espèce ne peut pas être utilisée dans par exemple chez les organismes se reproduisant par voie asexuée. Autres cas problématiques : définition des espèces chez les fossiles, cas de *Salix*.

! Critère : groupe monophylétique. L'espèce est définie via l'arbre. Ce critère est applicable pour les organismes asexués et pour les fossiles.

Désavantages : établissement de l'arbre pas évident pour tous les métazoaires + tendance à multiplier le nombre d'espèces.

Exemple d'espèces établies selon un critère phylogénétique : éléphant d'Afrique qui vit dans des forêts (*Loxodonta cyclotis*) et l'éléphant d'Afrique qui vit dans les savanes (*Loxodonta africana*).

**Elephas maximus* = éléphant d'Asie.

⇒ Il n'y a pas une définition universelle pour définir le concept d'espèce.

MECANISMES DE SPECIATION

L'isolement génétique est nécessaire, indispensable pour qu'il y ait différenciation de deux groupes de populations (par mutation, dérive génique et/ou sélection).

1) ISOLEMENT GENETIQUE

a. Via un isolement géographique : spéciation allopatrique

== empêche le flux de gènes entre deux populations

Spéciation allopatrique :

- Par dispersion : dans ce cas on a une population ancestrale unique. Quelques individus colonisent un nouvel habitat, plus aucun flux de gène, entraîne la différenciation qui aboutira à la formation de nouvelles espèces.

Ex : Archipel Hawaï (l'île plus récente). On peut déduire l'ordre d'apparition des îles en regardant l'ordre dans lequel elles sont localisées l'une par rapport à l'autre. Si deux espèces sont localisées sur des îles proches, elles auront une parenté génétique plus proche que si elles sont localisées sur des îles éloignées. On observe une concordance entre l'arbre phylogénétique et la séquence des îles représentées.

C'est compatible avec une hypothèse de spéciation par dispersion :

- Au départ, ancêtre commun localisé quelque part (Hawaï)
- Dispersion des individus vers les îles proches (Molokaï, Maui)

- Par vicariance : une barrière géographique apparaît subitement et scinde la population en deux ou plusieurs morceaux. Ex : Isthme de Panama qui coupe l'Océan Pacifique de l'Océan Atlantique. Il est apparu il y a 3,000,000 d'années environ. Étude : échantillonnage de différentes espèces de crevettes de part et d'autre de l'isthme pour voir si ces espèces sont connectées entre-elles.

Résultat : Pour la plupart des espèces que l'on retrouve dans le Pacifique, une espèce sœur est retrouvée dans l'Atlantique et inversement. Confirme l'hypothèse que suite à l'apparition de l'isthme, l'aire de distribution des espèces a été coupée en deux. Chaque groupe a pu se différencier et aboutir à de nouvelles espèces. Expériences en laboratoire pour voir si les espèces peuvent encore s'hybrider. Constatation et tests en labo : Dans la plupart des cas, les nouveaux mécanismes comportementaux des individus du Pacifique les empêchaient de se reproduire avec ceux de l'Atlantique et inversement.

L'ISOLEMENT GEOGRAPHIQUE EST-IL NECESSAIRE ?

Spéciation sans isolement géographique :

1. Spéciation parapatrique : Les deux groupes ne sont pas totalement isolés, quelques échanges de migrants sont encore possibles entre les deux. Mais malgré ce flux de gènes on a une différenciation entre les deux entités
2. Spéciation sympatrique : Recouvrement majeur, voir total des aires de distribution des deux espèces

Dans le premier cas, différentes situations sont imaginables :

- Les entités se différencient car les deux parties correspondent à un habitat différent (par sélection, chacun s'est adapté à son habitat)
- Il y a 20,000,000 d'années, la Terre a subi une période glaciaire avec un glacier qui recouvrait une grande partie du nord de l'Europe. Suite à ça, l'aire d'espèces non-adaptées à ce climat a été réduite et confinée dans certaines zones du sud de l'Europe : Des zones refuges (péninsule ibérique, péninsule italienne, Balkans, ...) isolées les unes des autres. Une fois le climat réchauffé, les espèces confinées dans ces zones refuges ont pu recoloniser l'Europe du Nord. On passe d'une séparation allopatrique à une spéciation parapatrique via la recolonisation.

Quel que soit le mécanisme, dans les deux cas on note une zone d'hybridation où peut se former les hybrides. Différents cas :

- 1) Aptitude des hybrides équivalente à celle des formes parentales. Dans ce cas, ces hybrides s'accouplent et se reproduisent, la zone d'hybridation s'étend et peut éventuellement recouvrir l'ensemble de l'aire de distribution de l'espèce ==> zone d'hybridation large.
- 2) Aptitude des hybrides moindre que celle des formes parentales. La différenciation entre les deux populations est trop forte, les hybrides constituent alors des individus qui ne sont adaptés à aucun des deux milieux. On peut aussi imaginer des incompatibilités entre gènes qui ont évolué chacun de leur côté faisant que les hybrides sont moins aptes à se reproduire. ==> zone d'hybridation plus étroite : mise en place possible d'un **mécanisme de renforcement**.

Soient deux espèces d'oiseaux : *Ficedula hypoleuca* et *Ficedula albicollis*, localisés en Europe sur des territoires différents mais avec une aire de recouvrement dans laquelle ils peuvent s'hybrider. Le comportement des femelles pour la sélection des mâles est différent suivant qu'ils proviennent de la zone d'hybridation ou de la zone d'allopatrie.

On a alors fait des tests d'accouplements : On a proposé à une femelle (provenant de la zone d'hybridation) deux mâles, de deux espèces différentes et provenant de la zone d'hybridation. La

femelle choisit quasiment tout le temps le mâle de la même espèce.

Si on lui propose deux mâles mais provenant des zones d'allopatrie :

- Première constatation : les mâles sont différents, celui localisé dans la zone d'hybridation a des couleurs plus ternes
- La femelle a beaucoup plus de mal à reconnaître le mâle de sa propre espèce. Elle s'accouple donc dans plus de cas avec le mâle de l'autre espèce.

D'autres tests de la même nature ont été réalisés. Que conclure de ces expériences ?

Il y a clairement une sélection sexuelle, qui favorise dans la zone de sympatrie la reconnaissance du mâle de sa propre espèce. Le comportement des femelles va évoluer, par rapport à celui que l'on trouve dans la zone d'allopatrie, de manière à ce que la femelle s'accouple préférentiellement avec le mâle de la même espèce. Pourquoi un tel comportement est-il favorisé ? Et pourquoi les individus dans la zone de sympatrie ont des couleurs plus ternes ?

Les hybrides entre espèces sont moins adaptés (dans ce cas-ci) que les individus non-hybrides. Dans la zone d'hybridation, des accouplements inter-espèces sont possibles mais, puisque les deux espèces sont trop éloignées, les hybrides sont moins viables et leur probabilité de se reproduire est plus faible. La sélection en conséquence favorise des mécanismes qui empêche la formation de ces hybrides. **C'est ça un mécanisme de renforcement** (car RENFORCE la différenciation).

Au départ : Deux entités bien différenciées, à tel point que les hybrides sont moins adaptés et ont une probabilité plus faible de se reproduire. Le mécanisme de renforcement est alors mis en place pour empêcher la formation d'hybrides.

b) Isolement reproductif pré- ou post-zygotique

La barrière prézygotique se met en place avant la formation de la descendance. Pour la postzygotique, la formation du zygote n'est pas empêchée mais ce dernier peut ne pas se développer, ou meurt prématurément, il peut être stérile, etc...

La sélection peut-elle favoriser des barrières prézygotiques de la même manière qu'elle favorise les barrières postzygotiques ? Ou ne peut-elle favoriser qu'un seul type de barrière ?

Dans le cas précédent des oiseaux, la sélection favorise la barrière prézygotique (la barrière a lieu avant la formation du zygote).

La sélection naturelle peut favoriser une barrière prézygotique, mais pas une barrière postzygotique (illogique par rapport au fonctionnement de la sélection naturelle) : imaginons un allèle provoquant la mort du zygote, cet allèle ne sera pas transmis à la descendance, donc la sélection ne favorisera pas par définition cet allèle. Par contre, un allèle qui empêche un accouplement qui forme un hybride non viable allèle peut être sélectionné car il favorisera la descendance fertile.

*Drosophila*

On a pu remarquer qu'il y avait une relation entre la distance génétique qui séparait ces populations, et leur état d'isolement prézygotique. Nous constatons que dans la grande majorité des paires sympatriques, on a un isolement prézygotique très important. Ce qui montre bien que pour qu'il y ait une différenciation quand deux espèces sont en sympatrie, il faut un mécanisme d'isolement (comportemental dans ce cas-ci).

3) Aptitude des hybrides plus grande que celle des formes parentales

Dans ce cas on aura une zone d'hybridation stable et éventuellement une nouvelle espèce. Exemple

des tournesols sur la slide. *H. anomalus* est le produit d'hybridation qui a des caractères particuliers non présents chez ses espèces parentales (cette espèce peut vivre en zone sablées !). Revenons-en à la question initiale... L'isolement géographique est-il nécessaire ? Un scientifique célèbre considérerait que seul l'isolement géographique pouvait aboutir à la formation de nouvelles espèces et que les spéciations parapatriques etc... N'étaient pas possibles. Depuis, d'autres scientifiques (un peu moins célèbres ceci dit) ont tenté de prouver le contraire mais ça n'est pas évident.

Des modèles mathématiques montrent que, théoriquement, on peut aboutir à la formation de nouvelles espèces en parapatrie et en sympatrie. Mais est-ce que ça se produit dans la réalité ? Là se situe la difficulté : Lorsque deux espèces sœurs bien différenciées sont en condition de sympatrie, ça ne signifie pas forcément qu'elles sont apparues en condition de sympatrie ! Elles ont pu apparaître en conditions d'allopatrie et ensuite les conditions du milieu ont évolué...

Conditions requises pour suspecter une spéciation sympatrique

- Distribution sympatrique des deux espèces sœurs les plus proches
- Données génétiques qui indiquent qu'elles sont isolées sur le plan reproductif (ne peuvent plus se reproduire entre-elles aujourd'hui)
- Monophylie des deux espèces en sympatrie
- Conditions écologiques telles que la différenciation allopatrique est peu probable (ex : île isolée, petits lacs, hôte d'une espèce parasite)

Un exemple de spéciation sympatrique ?

Étude faite sur les cyclidées (diversité morphologie très différente). On les retrouve en Amérique centrale, dans différents lacs volcaniques. On se concentre sur le lac Apoyo, apparu il y a 23.000 ans avec une profondeur de maximum 200m (prérequis : faible profondeur donc on ne peut pas définir plusieurs habitats différents, de plus le lac est récent du coup il est peu probable qu'il y ait eu des spéciations).

Réseau d'allèles : représentation graphique qui montre les relations évolutives entre les différents allèles d'un gène. Chaque cercle représente un allèle. Chaque branche reliant deux allèles représente une mutation : par exemple l'allèle 1 et l'allèle 11 sont séparés par une mutation etc.

Deux espèces présentes dans le lac :

- Allèles rouges : *Citrinellus*, qui est présente aussi dans un autre lac.
- Allèles jaunes : *Zallosus*

Le réseau d'allèles indique un seul événement de colonisation et ensuite une différenciation en espèces au sein du lac. Les deux espèces présentes dans le lac forment un groupe monophylétique.

Graphique avec trois couleurs : on constate qu'il n'y a pas d'existence d'hybrides entre *Zallosus* et *Citrinellus*.

➔ Tous les indices convergent fortement vers une hypothèse de spéciation sympatrique.

BASE GÉNÉTIQUE DE L'ISOLEMENT REPRODUCTIF

De temps en temps, une nouvelle espèce est créée sur base d'une seule modification génétique. On imagine l'hybridation entre deux espèces différentes. Parfois, la formation de l'hybride permet l'apparition d'une nouvelle espèce, qui peut se reproduire de façon asexuée. Imaginons par exemple que l'hybride est stérile (cas d'espèces trop différentes au départ que pour former un individu fertile), l'émergence d'une nouvelle espèce est quand même possible si les individus usent d'un mode de reproduction asexuée.

Chez les plantes particulièrement, la **polyploïdie** est possible : Suite à l'hybridation, on obtient un

polyploïde, un individu qui regroupe le génome des deux entités de départ et forme un nouvel individu qui peut se reproduire de manière sexuée et asexuée.

Exemple : *Galeopsis tetrahit* (4n)

Dans certains cas, la polyploïdie peut aboutir à la duplication de génomes complets, comme dans le cas de *Sedum suaveoleus* par exemple ($2n = 640$ chromosomes). On pense que la polyploïdisation s'est produite environ 80 fois chez cette espèce !

On peut retrouver la polyploïdisation chez les animaux, même si c'est plus rare (pour rappel, c'est plus courant chez les plantes car elles peuvent plus souvent faire de l'autofécondation).

Pourquoi deux espèces, deux populations qui se différencient aboutissent à un état où ils ne peuvent plus se reproduire entre-eux ? Si on attend suffisamment longtemps, des barrières post-zygotiques s'installent... Plusieurs études se concentrent là-dessus : Quels gènes sont concernés ? Combien sont nécessaires à l'établissement de la barrière ? Etc...

Règle de Haldane (1922) : « when in the offspring of two different animal races one sex is absent, rare or sterile, that sex is the heterozygous » Il ne propose pas un mécanisme mais réalise une observation. Chez la plupart des mammifères, le sexe est déterminé par les chromosomes sexuels (XY chez l'homme, XX chez la femme). Ce n'est pas le cas dans tout le monde vivant ! Chez certaines espèces, le sexe femelle est le sexe hétérogamétique, comme par exemple chez les oiseaux. C'est aussi le cas chez la plupart des papillons.

Observation : les entités se différencient. À un moment donné, la différenciation est telle que l'hybridation aboutit à des individus non-viables ou stériles. Lorsqu'un seul des deux sexes est stérile, c'est en général le sexe hétérogamétique (dans le cas des oiseaux c'est la femelle)

Certaines explications ont été proposées :

- Théorie de la dominance (Muller 1940) : Les mâles ne présentent qu'un des deux allèles pour les gènes liés au chromosome sexuel, du coup la dominance joue.
- Évolution plus rapide des mâles (Wu & Davis 1993) : Le processus de spermatogenèse est sensible. Des mutations présentes sur les gènes impliqués dans ce processus peuvent aboutir à l'hybridation des mâles. Les mâles évoluent plus rapidement car ils sont soumis à deux types de phénomènes différents :
 - Interactions entre mâles, bataille entre-eux pour avoir accès à la femelle
 - Interactions intersexes, nécessité de séduire les femelles et de présenter les caractères les plus attractifs
- Gènes qui biaisent le sex-ratio en leur faveur : Au sein du génome, certains gènes peuvent être favorisés par la sélection. Ces derniers vont biaiser le sex ratio pour être mieux transmis aux générations suivantes, ce qui a une incidence sur la présence d'individus stériles ou non-viables.

Gènes concernés ?

- Les gènes concernés sont déjà impliqués dans d'autres fonctions standards des organismes. Ce ne sont pas des gènes particuliers qui apparaissent avec comme fonction d'établir des barrières prézygotiques
- Les gènes responsables des barrières zygotiques ne peuvent être rassemblés dans une classe fonctionnelle particulière : On les trouve dans toutes les classes fonctionnelles qu'on pourrait établir dans le génome d'un organisme
- Les gènes responsables des barrières zygotiques ont évolué plus rapidement par sélection adaptative.

Souvenons-nous : En parlant des généalogies de gène, on a vu que la dérive génique permet la disparition de certaines lignées. En attendant suffisamment longtemps après la différenciation de deux espèces, on aboutit à deux groupes monophylétiques.

La sélection peut accélérer ce processus : Certaines mutations nouvelles qui arrivent dans l'une ou l'autre branche sont favorisées car permettant l'adaptation à l'habitat, à la survie, etc... Bref, la sélection accélère l'accumulation des mutations.

- Le nombre de gènes concerné dépend du stade de différenciation :
 - Début de la différenciation : Barrière présente à cause de 1 ou quelques gènes
 - Au fur et à mesure, le nombre de gènes s'accumule et devient de plus en plus nombreux

SPECIATION EN PRESENCE DE FLUX GENIQUE

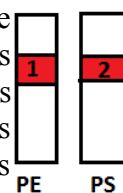
Deux espèces, qui coexistent par sympatrie, encore capables de s'hybrider. Malgré ça, deux espèces différentes se maintiennent avec leurs caractéristiques propres. En l'occurrence, il s'agit de deux espèces de drosophile.

- *D. Pseudoobscura* : Présent sur la côte pacifique de l'Amérique du Nord. Pareil pour *D. Persimilis*. Ces deux espèces peuvent alors former des hybrides qui sont fertiles.
- *D. pseudoobscura bogotana* localisé en Amérique du Sud
- *D. miranda* l'espèce de référence

Les espèces pseudoobscura et persimilis peuvent s'hybrider et restent fertiles : Échange de gènes interspécies mais les caractéristiques propres à chacune d'entre-elles se maintiennent. On aurait pu imaginer qu'à cause du flux de gène, un rassemblement en une seule même espèce aurait eu lieu... Ça n'est pas le cas.

→ Séquençage, analyse phylogénétique sur les différents individus de chacune de ces 3 espèces : tous les *pseudoobscura* forment un groupe monophylétique. Pareil pour *persimilis*. En regardant au niveau d'un autre gène, localisé sur le génome mitochondrial : tout est mélangé, ce qui signifie que *pseudoobscura* et *persimilis* sont mélangées sur une grande partie de l'arbre phylogénétique.

Au sein des différents chromosomes de l'espèce, on observe plusieurs inversions chromosomiques (= des portions des régions chromosomiques sont inversées). C'est important : Les régions inversées ne peuvent plus se recombiner lorsqu'elles sont associées à un chromosome sur lequel la portion n'est pas inversée. Sur les régions inversées, on a tous les gènes responsables de la différenciation entre les deux espèces. L'inversion empêche la recombinaison dans les hybrides formés par l'accouplement entre les deux espèces. Photo : Le crossing over n'est pas possible en 1



Si la portion A est inversée, elle ne pourra plus se recombiner avec la portion B. Par contre elle pourra toujours se recombiner avec les autres portions du chromosome.

→ Bref, ces inversions chromosomiques protègent le mélange des deux espèces car les régions ne peuvent se recombiner malgré le flux de gène.

EXEMPLES DE SPECIATION SYMPATRIQUE

La spéciation sympatrique est souvent proposée pour les insectes herbivores, car ils sont souvent spécialisés sur quelques plantes-hôtes. Dès lors, l'espèce se différencie et s'adapte aux plantes-hôtes différentes d'une manière différente, ce qui aboutit à de nouvelles espèces.

Rhagoletis pomonella : Espèce présente en Amérique du Nord qui se développe sur l'aubépine.

Les individus pondent leurs oeufs sur les fruits. Les accouplements se font sur les fruits des plantes. Il y a environ 300 ans, les pommiers ont été introduit en Amérique du Nord. *Rhagoletis pomonella* a alors investi les pommiers et cause des dégâts importants (récoltes moins importantes).

Des biologistes ont étudié les *Rhagoletis* des pommiers, et les *Rhagoletis* des aubépines. Ils constatent qu'ils forment des groupes différents : En utilisant des marqueurs génétiques, on peut différencier très clairement les populations à aubépine et les populations à pommier.

Proposition : Spéciation sympatrique en cours. On introduit le pommier il y a 3 siècles et, aujourd'hui, en comparant les populations à pommier et celles à aubépine on observe une différence génétique importante, à tel point que les deux ne peuvent plus se reproduire entre-elles.

Hypothèse : Aubépine et Pommier arrivent à maturité à une période différente de l'année. Du coup les accouplements ont lieu à des moments différents.

Problème : Aujourd'hui, on observe des accouplements interraciaux... Comment l'expliquer ? Alors que la différence génétique est si grande.

En génétique des populations, il suffit d'un flux de gènes faible pour que deux populations se mélangent. En analysant le génome et les chromosomes, on a pu mettre en évidence des inversions chromosomiques, expliquant que deux espèces en sympatrie continuant à échanger des gènes restent différenciées.

Tous les locus utilisés par les auteurs au départ étaient, par hasard, localisés dans des régions d'inversion chromosomique. La différenciation n'est pas si forte que ce que l'on pensait au départ, car une grande partie du génome est mélangée.

Autre élément apparu plus tard remettant en cause l'hypothèse de la spéciation sympatrique : Une autre population de cette espèce présente au Mexique est beaucoup plus proche de la race à pommiers, que de la race à aubépines.

Hypothèse actuelle : Dans un premier temps, séparation allopatrique. On avait une population en Amérique du Nord, l'autre au Mexique. Probablement, à partir de l'introduction des pommiers, colonisation des mexicains vers l'Amérique du Nord. Ces individus étaient préadaptés aux pommiers puisque ses fruits apparaissent plus tôt dans l'année. Ces insectes venant de régions chaudes, ils se reproduisaient plus tôt dans l'année également : concordance.

11 décembre 2015

CONFLITS INTRAGÉNOMIQUES ET UNITÉS DE SÉLECTION

- La biodiversité peut être divisée en différents niveaux d'organisation :
Nucléotide < gène < cellule < organe < individu < groupe < espèce < taxons supérieurs

Quel est le niveau d'organisation qui bénéficie d'une adaptation particulière ? Ex du lion qui capture une proie : l'adaptation bénéficie à l'individu qui va pouvoir survivre, bénéficie également au groupe de lions qui chassent ensemble et bénéficie pour l'espèce lion qui s'en porte mieux. Enfin, pour les différents Felidae, cette adaptation est également positive puisqu'au moins une espèce se portera mieux. Et la compétition entre espèces ? La proie capturée par le lion, c'est une proie en moins pour une autre espèce de Felidae... À regarder au cas par cas.

Au niveau inférieur : si l'individu survit, alors ses organes et ses cellules survivent. Les différents gènes ont une probabilité plus élevée d'être transmis à la génération suivante. Par contre, le nucléotide ou le niveau atomique ne sont pas affectés par la survie ou non dans le milieu.

➔ **Une adaptation constitue un bénéfice pour différents niveaux d'organisation.** On a :

- Des niveaux qui bénéficient de la sélection
- Le niveau d'organisation qui est la cause de l'apparition de cette adaptation, de la sélection en faveur d'une caractéristique : c'est ce qu'on appelle **unité de sélection**

Les adaptations ont des conséquences différentes pour les niveaux d'adaptation. En reprenant l'exemple du lion : Si une adaptation permet à un mâle de mieux combattre un concurrent (mâle aussi), cette dernière constitue un bénéfice pour l'individu mais pas pour le groupe.

Dans des cas extrêmes, certaines mutations sont bénéfiques pour tous les niveaux d'organisation : Par exemple, une mutation qui optimise la réplication de l'ADN.

Quelle est l'unité sur laquelle la sélection s'opère ? Quel niveau d'organisation est la cause de l'apparition de l'adaptation ? Le niveau sur lequel la sélection s'opère... On entend souvent dans la littérature de vulgarisation scientifique : « *Telle adaptation ou tel comportement a été sélectionné(e) pour le bien du groupe ou de l'espèce.* ».

Jadis, des biologistes ont proposé que la sélection puisse jouer pour faire apparaître une adaptation au niveau du groupe, bénéficiant à tous ses membres. D'autres biologistes s'inscrivent en faux, soutenant que l'adaptation ne peut être sélection qu'au niveau de l'individu.

Reprenons l'exemple du lion : Une adaptation le rend plus apte à capturer une proie. Cette adaptation bénéficie au lion, la sélection s'opère au niveau de l'individu. Autre vision : La sélection bénéficie au groupe. Peut-on pour autant dire que la sélection agit au niveau du groupe ? Non, car on peut expliquer cette solution simplement par un bénéfice au niveau de l'individu. Même si l'adaptation bénéficie au groupe, la sélection s'opère qu'au niveau de l'individu. Le groupe en bénéficie mais ce n'est pas lui l'unité de la sélection.

Pour mettre en évidence que le groupe est l'unité de sélection, il faut trouver une mutation qui fait apparaître une caractéristique, qui bénéficie au groupe sans bénéficier à l'individu (diminution pour lui de probabilité de survie ou de reproduction) : Comportements altruistes.

Sélection de la parentèle : Le comportement altruiste est dirigé vers des individus qui sont proches génétiquement. Donc là, c'est un cas particulier. Imaginons qu'un nouvel allèle apparaît dans la population et développe ce comportement altruiste : Ce dernier « se favorise lui-même » puisqu'il est destiné aux individus semblables génétiquement.

Dans un cas où le comportement altruiste se dirige vers tous les individus du groupe, qu'ils soient génétiquement proches ou pas, on pourrait penser à un cas de sélection de groupe.

Exemple proposé : Soit une population de souris placée dans un enclos fermé, auquel on donne dans un premier temps une grosse quantité de nourriture. Comme la nourriture est abondante, chaque couple de souris obtient un nombre de jeunes importants. Au fur et à mesure que la nourriture devient plus rare, les souris commencent à avoir des portées plus petites. Idée : Chaque reproducteur se sacrifie pour le groupe, en utilisant qu'une partie des ressources et donc en ayant moins de descendants. Cela permet la survie du groupe au détriment du succès reproducteur de l'individu.

Le problème est qu'on peut expliquer ça par une sélection au niveau de l'individu : Si la quantité de nourriture devient trop faible, la probabilité de survie pour les descendants devient plus faible pour une grande portée que pour une petite portée. L'individu a intérêt à avoir des portées réduites, car s'il doit nourrir tous ses descendants, du fait du manque de nourriture avec une grande portée, la majorité de ses enfants mourront.

Pour le moment, personne n'a pu démontrer un cas où une sélection de groupe est affirmée. C'est logique : Un comportement d'individu qui se sacrifie pour le groupe est défavorisé par la sélection. Même s'il apparaîtrait un jour, il disparaîtrait très rapidement.

Expérience sur *Tribolium* (Wade 1976)

Tentative de démontrer la sélection de groupe : On constitue 48 colonies de chacune 16 individus de coléoptères. On les laisse se reproduire pendant 37 jours. Ensuite, on sélectionne artificiellement les colonies qui produisent le moins d'individus pour qu'elles constituent la génération suivante. Au final, on se retrouve avec des colonies où les individus ont des taux de reproduction plus faibles qu'au départ. On a pu mettre en évidence que la sélection choisit des individus qui sacrifient leur succès reproducteur au groupe.

Certains pensent qu'on peut cependant aussi réduire ça à une sélection individuelle : Regrouper les individus en colonies ne change rien. On applique une sélection artificielle qui favorise les individus ayant une descendance faible. On applique une pression de sélection sur ces individus, qui au final sont choisis. Cette expérience montre juste que si le critère de sélection est « produire le moins d'enfants », les individus choisis seront ceux qui produisent le moins d'enfants.

CONFLITS INTRAGÉNOMIQUES

Certains allèles modifient en leur faveur le ratio 1:1 de ségrégation des allèles lors de la méiose. Lors d'une méiose normale, chaque allèle se retrouve avec une fréquence de 50% dans les différents gamètes produits.

Certains allèles sont capables de biaiser ce ratio de ségrégation en leur propre faveur.

Exemple : Chez les souris *Mus Musculus* ou *Mus Domesticus*. Un groupe de gènes, l'**haplotype t** existe sous plusieurs formes (3 allèles) : + (forme normale), t et T (formes Tailless, dont la queue est plus courte).

- Deux allèles ++ : phénotype normal
- Homozygotes TT ou tt : l'individu n'est pas viable
- Tt ou T+ : Tailless (viable mais avec phénotype particulier)
- t+ ou Tt : Chez le mâle de ces individus, production de gamètes qui comportent entre 90% et 100% de l'haplotype t (normalement on aurait 50% de t, hors ce n'est pas le cas).

Il existe en fait un niveau de l'haplotype au moins 1 gène qui a effet toxique sur le gamète, produisant une protéine poison qui, si elle se développe au niveau du gamète, le tue. Lorsque l'haplotype t est présent dans le gamète, les allèles contrecarrent l'effet toxique : Survie du gamète. Par contre, les gamètes t+ et Tt contiennent dans leur cytoplasme le poison produit par le zygote. Ne restent plus viables que les gamètes qui contiennent t.

Ce système est apparu au niveau du génome des souris, et qui favorise donc sa propre transmission via la production du poison.

Pourquoi n'atteint-on pas 100% d'allèles t dans la population ? Si tous les individus sont tt ... La population disparaîtrait, car ils ne sont pas viables !

Deux types de sélection qui agissent sur des niveaux différents se contrecarrent :

- Celle favorisant la transmission de t au niveau de la production des gamètes
- Celle empêchant les individus tt de se reproduire (ils meurent)

Ces deux sélections favorisent pour l'une et défavorisent pour l'autre t : on finira par avoir un équilibre, qui explique que le système se maintient. Normalement, avec la dérive génique on pourrait imaginer que soit t disparaît, soit qu'il atteigne une fréquence de 100%, mais à cause des deux sélections, l'équilibre est maintenu.

Un élément génétique qui ne se transmet que via un des deux sexes peut obtenir un avantage en biaisant le sex ratio en sa faveur. Ex : L'ADN mitochondrial qui ne se transmet que via le sexe femelle : en effet, tout le cytoplasme et les organites cellulaires qui vont constituer le zygote proviennent de l'ovule, car le spermatozoïde y introduit son génome et fournit son ADN et basta. Pour quelques espèces animales, il y a des transmissions des mitochondries mâles mais cela reste très rare.

Chez beaucoup d'Angiospermes, des individus portent à la fois des fleurs mâles et femelles. Mais chez 5% à 10% des individus d'espèces hermaphrodites, les fleurs mâles sont stériles. En effet : Des gènes au niveau de l'ADN mitochondrial sont transmis et stérilisent les fleurs mâles. Au niveau du génome nucléaire, d'autres gènes apparaissent et contrecarrent les gènes stérilisant mitochondriaux.

→ À nouveau, deux niveaux de sélection différents.

D'autres allèles peuvent apparaître qui contrecarrent l'effet de ceux qui biaisent le ratio de ségrégation. Ex : Une espèce de drosophile pour laquelle on a pu construire artificiellement un chromosome Y comportant un complexe de gènes qui biaise la ségrégation en sa faveur.

À ce niveau, les mâles produisent surtout des gamètes qui vont contenir ledit chromosome Y. Ceux qui contiennent le chromosome X seront décrits. Sur la slide : Évolution de la fréquence du chromosome Y, qui augmente dans un premier temps puis se stabilise un petit peu. On voit sur les seconds graphiques que dans un second temps le sex ratio est fortement biaisé en faveur des mâles.

Au bout de quelques jours, le sex ratio revient à une valeur proche de 50%, en raison de l'apparition d'autres allèles qui contrecarrent l'effet de cet allèle artificiel qui biaise le sex ratio en sa faveur. On a des conflits au sein du génome :

- D'une part, le chromosome Y qui biaise la ségrégation en sa faveur
- D'autre part, la sélection favorise d'autres allèles pour contrecarrer le chromosome Y et rétablir un sex ratio avoisinant les 50%

→ La distorsion du ratio de ségrégation peut être contrecarrée par des changements comportementaux.

Exemple des souris : Les femelles t+ ont développé un comportement qui favorise les

accouplements avec les mâles ++, par rapport aux mâles qui sont également t+.

Vu qu'ils ont une queue plus courte, les femelles peuvent les reconnaître aisément (hehe ☺). Ça génère une descendance en partie ++ qui contrecarre le biais de ségrégation en faveur des allèles t.

Autre exemple : Les *Diopsidae*, qui ont des yeux localisés au sommet de longs pédoncules (voir slide). Chez deux espèces, on met en évidence un allèle qui biaise la méiose en faveur de la transmission du X. Ce biais fait que la proportion de femelles dans la population est très importante.

Pour contrecarrer cet effet, d'autres gènes apparaissent sur le chromosome Y et rétablissent un sex ratio d'environ 50%. À nouveau, conflit intragénomique. On a pu mettre en évidence que certaines femelles ont développé un comportement pour sélectionner les mâles qui ont un pédoncule oculaire très grand. C'est un bel exemple de sélection sexuelle (*hypothèse* : si le mâle survit alors qu'il a des pédoncules très longs, il est très fort et survit, il va alors être choisi par les femelles).

Les gènes localisés sur Y pour contrecarrer sont très proches de ceux déterminant la longueur des pédoncules : déséquilibre de liaison entre les gènes qui contrecarrent et les gènes qui déterminent un long pédoncule. → En choisissant des mâles avec un long pédoncule, en fait les femelles choisissent les mâles leur permettant d'établir globalement un sex ratio de 50%.

Intérêt pour la femelle : Elle obtient une descendance avec plus de mâles, avantageux car cela augmente son succès reproducteur au final (+/- même nombre de mâles et de femelles : tout le monde trouve un partenaire pour se reproduire). De plus, les fils de la femelle posséderont les allèles qui contrecarrent, ces fils auront eux-mêmes un succès reproducteur plus élevé !

Les conflits intra-génomiques peuvent jouer un rôle dans le processus de spéciation

Le lien entre les conflits intra-génomiques et le processus de spéciation a été établi en 2009, présenté par un cas particulier qui concerne *D.pseudoobscura* (USA) et *D.bogotana* (Amérique du sud). Lorsqu'on croise des mâles de *pseudoobscura* avec des femelles de *bogotana* on constate que :

- Les mâles de la F1 sont stériles mais qui peuvent tout de même devenir fertiles quand ils sont âgés mais ne donnent naissance qu'à des filles. Cela pourrait s'expliquer par un allèle qui biaise le ratio de ségrégation en faveur du chromosome X.

Deux effets sont causés par le même gène :

- Effet du biais de ségrégation
- Effet de la stérilité

Le gène a été séquencé chez *bogotana*, chez *pseudoobscura* et chez certaines espèces-sœurs proche de *pseudoobscura*. On a constaté qu'un grand nombre de mutations (7 non-synonymes et 5 synonymes, soient 12 mutations au total) séparent les deux séquences *bogotana* et *pseudoobscura*. Par contre, en comparant *persimilis* et *m.randa*, la séquence des gènes est identique à celle qu'on avait trouvé chez *pseudoobscura*. *Bogotana* a fortement divergé (12 mutations différentes).

On peut reconstruire le fil des événements : Les deux sous-espèces de *pseudoobscura* vivent en allopatrie. Au sein de *bogotana* est apparu un allèle biaisant le ratio de ségrégation en faveur de X. Afin de contrecarrer ça, d'autres allèles sont apparus sur les chromosomes 2 et 3, ce qui fait que dans la population de *bogotana*, le ratio mâle/femelle est rétabli.

Pourquoi une séparation de 12 mutations pour *bogotana* ? Au sein de l'espèce, course évolutive entre d'une part les allèles qui biaisent le ratio de ségrégation, et d'autre part les allèles qui

contrecarrent le X → Sélection forte présente dans la population, ils évoluent alors plus rapidement par sélection adaptative.

Comme il n'y a pas d'effet de ségrégation chez *pseudoobscura* et les autres espèces, ce gène n'évolue pas (ou en tout cas beaucoup moins vite que pour *bogotana*).

/!\ Le conflit intragénomique au sein de *bogotana* accélère le processus de mutation, ce qui fait que la séquence de *bogotana* devient différente de celle de *pseudoobscura*. À un moment, cette incompatibilité est telle que si un hybride *bogotana/pseudoobscura* est produit par accouplement, il sera stérile.

L'unité sur laquelle la sélection agit ne serait-elle pas plutôt le « gène » ? **Proposition de Richard Dawkins en 1976.** Ici, par gène on entend fragment d'ADN transmis de génération en génération et qui ne subit pas de recombinaisons.

À l'origine, une molécule d'ADN se réplique et une erreur (mutation) apparaît : Si cette erreur lui confère une capacité de réplication plus grande, elle sera favorisée. C'est la sélection naturelle. À un moment donné, les molécules d'ADN se sont regroupées en gènes, qui se sont rassemblés dans des cellules, qui elles-mêmes ont construit des organismes, ... On peut considérer que le gène est une sorte de réplicateur qui permet aux individus de se multiplier.

Richard Dawkins dit que la sélection a lieu au niveau du gène, de la molécule d'ADN. Dans certaines circonstances, cette façon de penser plaçant le gène au centre de l'Évolution, est plus logique. Cette théorie ne fait pas consensus.

Exercice sur les mécanismes de spéciation (question d'examen) :

La plante *Argostis tenuis* est parvenue à coloniser des sols pollués en métaux lourds dans le nord du pays de Galle. À la frontière entre les sols pollués et les sols non pollués, on observe que les individus qui poussent sur des types de sol différents sont caractérisés par des phénologies florales incompatibles : les fleurs sur sols pollués n'arrivent pas à maturité au même moment que les fleurs présentes sur sols non pollués, ce qui empêche les croisements entre les deux. Cette différence de phénologie florale disparaît quand on s'écarte à la zone frontière.

- a) Comment et pourquoi la sélection naturelle a-t-elle pu favoriser cette différence dans les temps de floraison au niveau de la zone frontière ?

→ Il s'agit d'un phénomène de renforcement. La sélection est apparue car les hybrides entre les plantes sur les sols pollués et celles sur les sols non pollués sont désavantagés par rapport à leurs parents. Hypothèse : Les hybrides ne sont ni adaptés au 1er type de sol, ni adaptés au 2ème type de sol (ou en tout cas moins adaptés que leurs parents). Un mécanisme permettant d'éviter l'hybridation sera donc favorisé par la sélection.

Comment est-ce apparu ? Un mécanisme d'évitement, apparu par mutation, a été ensuite sélectionné.

- b) Quel pourrait-être l'impact à long terme de ce phénomène sur la différenciation entre la population d'*A. Tenuis* sur sol pollué et celle sur sol non pollué ? Expliquez.

→ La différenciation augmente. C'est la base du mécanisme de renforcement : Le mécanisme de sélection diminue les accouplements entre individus des deux populations, ce qui diminue le flux de gènes et augmente la différence entre les deux (pour rappel, le mécanisme de renforcement apparaît au niveau de la zone hybride).

- c) Ce même phénomène aurait-il pu favoriser (par sélection) un allèle qui rend les hybrides (issus des croisements entre plantes qui poussent sur sol pollué et plantes qui poussent sur

sol non pollué) stériles ? Expliquez.

Non. Si le résultat du croisement est une descendance stérile, les individus ne transféreront pas leurs gènes à la génération suivante. Donc la sélection désavantagera ça. Il n'y a que les barrières prézygotiques qui peuvent être favorisés par la sélection. Les barrières postzygotiques engendrent des individus stériles, et ne seront donc pas sélectionnées.