Reconstituer l'histoire évolutive d'une famille multigénique

Comment reconstituer l'histoire évolutive d'une famille multigénique à partir d'une matrice de comparaison de séquences ?

1. Identifier le type de matrice

Le titre de la matrice indique son type, mais vérifiez quand même sa diagonale:

- Diagonale à 100 % : matrice de similitudes. Plus la valeur est élevée, plus les séquences se ressemblent.
- Diagonale à 0 % : matrice de distances. Le raisonnement s'inverse dans toute la suite.

2. Montrer l'origine commune

- Relever la similitude minimale de la matrice et la longueur des séquences.
- La comparer à la valeur due au hasard : 25 % pour des nucléotides, 20 % pour des acides aminés.

> « La plus faible similitude de la matrice (… %) porte sur des séquences de … nucléotides. Elle ne peut pas être due au hasard. Les gènes sont issus d'un gène ancestral commun : ils forment une famille multigénique. »

3. Expliquer les différences

- Les similitudes restent inférieures à 100 %.
- Chaque duplicata accumule ses propres mutations depuis la duplication : c'est la divergence.
- Horloge moléculaire : plus la similitude est élevée, plus la divergence a été courte.

> « Les différences correspondent aux mutations accumulées indépendamment par chaque duplicata depuis la duplication (divergence). Le taux de mutation est supposé constant (horloge moléculaire). Plus la similitude est élevée, plus la divergence a été courte, et plus la duplication est récente. »

4. Compter les duplications

- n gènes : n - 1 duplications.

> « La famille compte … gènes. Elle résulte donc de … duplications. »

5. Repérer la duplication la plus récente

- Chercher le couple de gènes le plus semblable : sa divergence est la plus courte.
- Ces deux gènes sont issus de la duplication d'un gène ancestral commun qui leur est exclusif.

> « Les gènes … et … présentent la similitude la plus élevée (… %). Leur divergence a donc été la plus courte. Ils sont issus de la duplication d'un gène ancestral commun qui leur est exclusif. Ce gène ancestral est plus récent que celui qu'ils partagent avec le gène … »

6. Reconstituer l'histoire

Cas d'une famille de trois gènes

- Les gènes A et B sont les plus semblables. Ils sont issus de la duplication la plus récente (étape 5).
- Le gène C présente des similitudes voisines avec A et avec B. Elles sont plus faibles que celle de A et B.
- Les divergences de C avec A et avec B ont donc été de même durée. Elles ont été plus longues que celle de A et B.
- Une première duplication du gène ancestral commun aux trois gènes a produit le gène ancestral de C et le gène ancestral commun exclusif à A et B.
- Une seconde duplication plus récente de ce gène ancestral commun exclusif a produit A et B.

> « Le gène … présente des similitudes voisines avec … et … (… % et … %). Elles sont plus faibles que celle de … et … . Les divergences de … avec … et avec … ont donc été de même durée, et plus longues que celle de … et … . Une première duplication du gène ancestral commun aux trois gènes a produit le gène ancestral de … et le gène ancestral commun exclusif à … et … . Une seconde duplication plus récente de ce gène ancestral commun exclusif a produit … et … . »

Cas d'une famille de plus de trois gènes

- Partir des deux gènes les plus semblables, issus d'un gène ancestral commun qui leur est exclusif. Ils forment un premier groupe.
- Comparer un groupe déjà formé comme un seul gène : ses gènes descendent d'un gène ancestral commun qui leur est exclusif. Un gène extérieur au groupe présente des similitudes voisines avec chacun des gènes du groupe.

- Prendre la similitude immédiatement inférieure. Elle rattache un gène à un groupe, réunit deux autres gènes en un nouveau groupe, ou réunit deux groupes.
- Recommencer par similitude décroissante, jusqu'à ce que tous les gènes soient rattachés. Chaque rattachement situe une duplication plus ancienne que le précédent.
- Le dernier rattachement correspond à la divergence la plus longue, donc à la duplication la plus ancienne. C'est la duplication du gène ancestral commun à toute la famille.
- Vérifier que la reconstitution compte n - 1 duplications (étape 4).

> « Le gène … présente des similitudes voisines avec chacun des gènes du groupe … (… % en moyenne). Elles sont plus faibles que celles des gènes du groupe entre eux. Les divergences de … avec chacun de ces gènes ont donc été de même durée. Elles ont débuté à partir de la duplication du gène ancestral commun à ce gène et au groupe. Cette duplication est plus ancienne que celles qui ont produit les gènes du groupe. »

7. Identifier le mécanisme de chaque duplication

- Gènes sur le même chromosome : crossing-over inégal lors de la méiose.
- Gènes sur des paires différentes : translocation entre chromosomes non homologues lors de la méiose.

> « Les gènes … et … sont situés sur le même chromosome. Leur duplication résulte d'un crossing-over inégal. Le gène … est situé sur une autre paire. La duplication qui l'a séparé de … et … résulte d'une translocation. »

Limites: Les seuils de 25 % et 20 % ne valent que pour des séquences longues. L'horloge moléculaire est une hypothèse. La méthode établit l'ordre des duplications, pas leur date.