**CHAPITRE 7 : LA GENOMIQUE COMPARATIVE**

1. **Définition**

 La génomique comparative est une approche permettant d’étudier la comparaison et la structure des génomes. On peut la considérer comme une branche de la bio-informatique vouée à l’étude des similarités et des différences dans l’organisation et la structure des génomes. Son objectif principal est de trouver les relations généralisables entre les génomes étudiés tout en appréciant l'histoire individuelle de chacun. Les relations qui peuvent exister entre gènes et entre génomes sont diverses (orthologie, paralogie, synténie, duplication, gain ou perte de gènes et de fonctions). Les protéines engendrées par les gènes ont aussi des relations, dynamiques, qui décrivent leurs manières d'interagir et de coopérer pour réaliser les comportements cellulaires.

1. **Le principe de la génomique comparative**

A l’échelle d’un génome entier (pour les procaryotes) ou d’une partie de génome (pour les eucaryotes), nous disposons de solutions intégrées permettant de mettre en évidence la présence, l’absence ou la duplication de gènes (ou de groupes de gènes) dans le cadre d’études comparatives d’un génome par rapport à une référence connue. Cette application passe par l’extraction du génome étudié, son marquage à l’aide de fluorophores et enfin par une hybridation compétitive sur « microarray » en présence du génome de référence. Donc elle passe par la comparaison du ou des génomes étudiés à un génome de référence en utilisant la technologie des « microarrays » ou ce qui on appele puce à ADN.

1. **Homologie et identité de séquence**

La comparaison de séquences avec des logiciels de comparaison de séquences type BLAST conduit à mesurer un pourcentage d’identité entre les séquences nucléotidiques ou d’acides aminés. Pour les séquences d’acides aminés, on mesure également un pourcentage de similarité, qui tient compte, outre l’identité des acides aminés, du fait que les acides aminés appartiennent à une même famille selon que leur chaine latérale est acide, polaire neutre, non polaire. Cette identité et cette similarité sont qualifiables.

Si deux séquences présentent des identités ou similarités de séquence, cela peut traduire :

* Qu’elles ont une même origine phylogénétique. On dit alors que les séquences sont homologues, c'est-à-dire qu’elles sont issues d’une séquence ancestrale commune. Contrairement à la notion d’identité ou de similarité (deux séquences peuvent être similaire à 55%), la notion d’homologie n’est pas quantifiable (il existe ou il n’existe pas de séquence ancestrales communes) ;
* Qu’elles ont convergé bien qu’elles n’aient pas d’ancêtre commun. Ceci peut être du au fait du hasard (dans des régions de faible complexifié, c'est-à-dire avec des nucléotides ou des acides aminés très répétés) ou d’une pression de sélection commune.

On considère en général qu’il y a homologie entre deux séquences, si leur similarité est supérieure à 30%.

1. **Homologie et origine ancestrale**

Au sein des gènes homologues, selon la nature de l’événement évolutif qui les séparés, on décrit trois sous-classes :

* **Les gènes orthologues** : ce sont des gènes ayant évolué à partir d’un même gène ancestral suit à des événements de spéciation (séparation d’espèces). Souvent, la fonction des gènes orthologues est conservée au cours de l’évolution :
* **Les gènes paralogues**: ce sont des gènes issus d’événements de duplication au sein d’un génome. Un ou plusieurs paralogues peut subir des modifications de séquence, telles que sa fonction en deviennent différentes. Il peut également perdre sa fonction et devenir un pseudo-gène
* **Les gènes xénologues**: ce sont des gènes ayant été acquis par transfert génétique horizontal (transfert de gènes entre bactéries). Souvent, les gènes xénologues confèrent à la bactérie receveuse une nouvelle fonction qu’elle ne possédait pas initialement.

On parle également de protéines orthologues, paralogues ou xénologues selon qu’elles sont codées par des gènes orthologues, paralogues ou xénologues.

1. **Relation de synténie entre les génomes**

**5-1 définitions de la synténie**

Dans l’hypothèse ou deux espèces comparées sont issues d’un même ancêtre commun, il est raisonnable de penser que leur organisation génomique à garder des traces de ce passé commun comme, par exemple, des fragments génomiques dans lesquels l’ordre des gènes est conservé. On appelle synténie la conservation de l’ordre des gènes sur le chromosome au sein de lignées différentes. La **synténie** se caractérise par 2 types de relations :

• Relation de "**co-localisation**" (intra génome).

• Relation de "**correspondance**" (inter génomes).

**5-2 Plateformes informatiques d’étude de la synténie**

L’étude de la synténie passe par l’utilisation d’outils informatiques qui permettent

* Des comparaisons intra génomiques
* Des comparaisons inter génomiques (séquences de génomes proches)
* Une visualisation des alignements

**5-3 La synténie comme aide à l’annotation**

Les recherches en génétique (génétique inverse, mutagénétique aléatoire) et les tests physiologiques, biochimique, immunologiques, etc., menés depuis prés d’un siècle sur des animaux modèles comme la souris et le rat ont permis l’élucidation de la fonction d’un certain nombre de gène ces travaux ne sont pas transposables chez l’homme en recherchant des blocs de synténie entre les génomes des animaux modèles et de l’homme (par exemple, il y’a plusieurs dizaines de blocs de synténie conservés entre la souris et l’homme), on facilite l’identification de gènes orthologues entre les deux espèces et donc l’annotation fonctionnelle du génome humain.

**5-4 la synténie comme marqueur d’événements chromosomiques**

Le degré de synténie est inversement proportionnel au temps écoulé depuis la divergence à partir du génome ancestral. Les moteurs de cette divergence sont les remaniements génomiques de différentes natures :

* Fusions chromosomiques
* Cassures chromosomiques (échange de segments de séquences entre des chromosomes différents)
* Inversions chromosomiques
* Délétions (perte d’une région chromosomique)
* Duplication d’une région chromosomique
1. **La comparaison de l’organisation chromosomique:( identification des relations d'homologue entre génomes)**

Il existe plusieurs méthodes pour comprendre les relations d'homologue et la comparaison entre les génomes, et cela à l'aide d'un outil bioinformatique.

**1-La bioinformatique:** La bioinformatique est une composante essentielle de la génomique comparative à la fois pour le développement de bases de données, les développements méthodologiques (algorithmique et représentation des résultats) mais aussi pour la réalisation d'analyses fines. Dans ce domaine on utilise le serveur web, leur architecture est divisée en 3 niveaux: la base de données relationnelle contenant les informations génomiques, l’application chargée du calcul des groupes de synténie, et une série d’interfaces web dédiées à l’exploration des résultats.

**2- Quelques méthodes de comparaison**

**a. L’alignement réciproque de séquences**

La méthode la plus généralement utilisée consiste à réaliser des alignements de séquences réciproques entre deux génomes (méthode du “Best Bidirectional Hits” (BBH) ou “meilleurs alignements réciproques”). L’idée sous-jacente de la méthode implique que les séquences de gènes orthologues présentent plus de similitudes entre elles que de similitudes avec une séquence du génome initial.



**Fig:** Le principe du BLAST réciproque entre deux génomes de deux espèces SA et SB.

Les gènes orthologues réciproques identifiés par la méthode sont symbolisés par les triangles blancs.

Dans cette figure deux gènes A et B, appartenant respectivement aux espèces SA et SB, seront considérés orthologues si le meilleur alignement de la séquence du gène A sur le génome SB correspond au gène B et, réciproquement, si le meilleur alignement de la séquence du gène B sur le génome SA correspond au gène A

**b. La méthode de regroupement**

La méthode des “COGs” pour Cluster of Orthologous Groups” consiste à identifier, non seulement des paires de gènes orthologues entre deux génomes, mais plus généralement, des groupes de gènes orthologues entre plusieurs génomes selon les étapes suivantes :

* Toutes les séquences de gènes sont alignées entre les génomes considérés afin d’identifier l’ensemble des relations d’homologie.
* Puis, la détection des paralogues est réalisée par des alignements de la séquence d’un gène sur son propre génome.
* Enfin, la dernière étape consiste à rassembler les groupes de gènes (orthologues et paralogues) identifiés dans les deux premières étapes pour former des groupes de gènes homologues.
1. **Les limites de synténie:**

En pratique, la recherche de synténie nécessite préalablement d’identifier les familles de gènes orthologues entre tous les génomes comparés. Les principales limitations à la recherche de synténie sont en réalité liées à cette étape d’identification des orthologues. Les méthodes traditionnellement utilisées pour identifier les orthologues atteignent en effet leurs limites quand les séquences divergent. On peut considérer uniquement les orthologues qui limitent la synténie, car cette dernière peut être dupliquée au sein d’un même chromosome, donc la formation ce qu'on appelle les paralogues.