



Annexe 1- Mesure du taux d'homologie entre deux séquences

Notions de score d'alignement et de pourcentages d'identité

Soit les deux séquences A et B dont l'alignement est représenté ci-dessous :

SEQUENCE A : ATGTTGCT---GTAAATGC-ATATGGCCTAAA
!!!!...----.-!!!-!!-!!!!.!!-!!!!
SEQUENCE B : ATGTGCT-AACC-AAACGCTATATCGC-TAAA

Trois types de correspondances peuvent être faites entre les nucléotides de la séquence A et ceux de la séquence B :

- les **identités** : les deux nucléotides sont identiques sur les deux séquences. Elles sont représentées par un « ! »
- les **substitutions** : une base est remplacée par une autre. Elles sont représentées par un « . »
- les **insertions et délétions**: l'ajout d'un nucléotide sur une séquence entraîne une brèche sur l'autre séquence (et vice versa). Elles sont représentées par un « - »

Il existe deux moyens simples de mesurer le taux d'homologie, c'est à dire le degré de parenté entre deux séquences :

- Le calcul du **pourcentage d'identité** : on place les deux séquences au mieux l'une de l'autre pour obtenir un maximum de « ! » et on calcule le % de « ! » par rapport au nombre total de correspondances.
Ce calcul est simple mais il n'est pas forcément le plus pertinent car il donne le même poids aux substitutions qu'aux insertions-délétions.
- Le calcul du **score d'alignement** est plus intéressant : il consiste à attribuer à chaque correspondance un score élémentaire ;
par exemple : identité +1, mésappariement 0, brèche -1.
Pour l'alignement parfait de deux séquences de 32 nucléotides, le score serait de 32.

IMPORTANT :

. Pour deux séquences données, le meilleur alignement est celui qui donne le score le plus élevé avec si possible le % d'identité le plus élevé.

. Il est possible de faire des alignements avec toutes les molécules possédant une séquence : ADN, ARN, Protéines.

. Les gènes de très nombreuses espèces ont été entièrement séquencés et sont stockés dans des banques de données en réseau. Elles sont toutes libres d'accès sur internet et accessibles par le portail américain NCBI (National Center for Biotechnology Information).