



Annexe 2- Les alignements de séquences sont réalisés par des programmes puissants accessibles en ligne

Les calculs présentés dans l'annexe 1 sont simples à réaliser mais ils deviennent extrêmement compliqués dès que les séquences comportent plusieurs dizaines d'éléments. Il est alors indispensable d'avoir recours à un programme informatique. Ces programmes sont basés sur des algorithmes mathématiques complexes qui permettent de faire des alignements de séquence à très grande vitesse en repérant les parties identiques au sein de plusieurs millions de séquences disponibles dans les banques de données bioinformatiques.

Il en existe plusieurs accessibles en ligne. Les plus utilisés par la communauté scientifique sont **BLAST** (Activité 1) ou **ClustalW** (activité 2).

BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

Pour les alignements de séquences de petite taille, BLAST (Basic Local Alignment Search Tool) est l'outil le plus utilisé actuellement. Sa rapidité et son efficacité en font un outil idéal pour rechercher dans des bases de données les séquences homologues à une séquence d'intérêt, appelée séquence « requête ».

La réponse donnée par le logiciel fournit de nombreuses informations :

- les séquences les plus homologues de la séquence requête classées par ordre de scores décroissants (en indiquant le % d'homologie).
- pour chaque séquence : le nom de l'espèce à laquelle elle appartient (lorsque celle-ci est référencée).
- quand il s'agit d'une séquence codante, un lien est proposé vers la fiche NCBI du gène fournissant toutes les informations disponibles sur celui-ci.

Pour une utilisation simple de BLAST avec des séquences de nucléotide (ADN ou ARN), seuls les paramètres suivants sont nécessaires :

- ENTER QUERY SEQUENCE : entrer la séquence requête sur laquelle vous interrogez BLAST. Par copier-coller de la séquence ou par saisie de son code de référence;
- CHOOSE SEARCH SET : choisir l'étendue de la recherche. En général on utilise la banque de données (=database) par défaut : « nucléotide collection », elle regroupe tous les ADN et ARN connus.
- PROGRAM SELECTION : choix de l'algorithme de recherche. On conserve en général le programme par défaut « highly similar sequence (Megablast) »

Une fois le programme paramétré, il est possible de lancer l'alignement de séquence en cliquant sur l'icône « **BLAST** » en bas à gauche. La recherche peut prendre quelques minutes et rend le résultat sous trois formes : graphe d'alignement, liste des séquences, alignements détaillés.



ClustalW

<http://www.ebi.ac.uk/Tools/msa/clustalw2/>

C'est un programme spécialisé dans l'alignement de séquences multiples, c'est à dire que ClustalW est capable de mesurer le degré de parenté de différentes séquences correspondant au même gène dans des espèces différentes.

A partir des résultats d'alignements multiples, ClustalW2 peut réaliser des arbres phylogénétiques.

ClustalW2 peut analyser des séquences d'ADN, d'ARN ou de protéines. Son utilisation est assez simple quand on utilise les paramètres par défaut.

La première étape consiste à saisir les unes au dessous des autres les séquences à analyser. La requête est déclenchée en cliquant sur l'icône « **SUBMIT** » en bas à gauche.