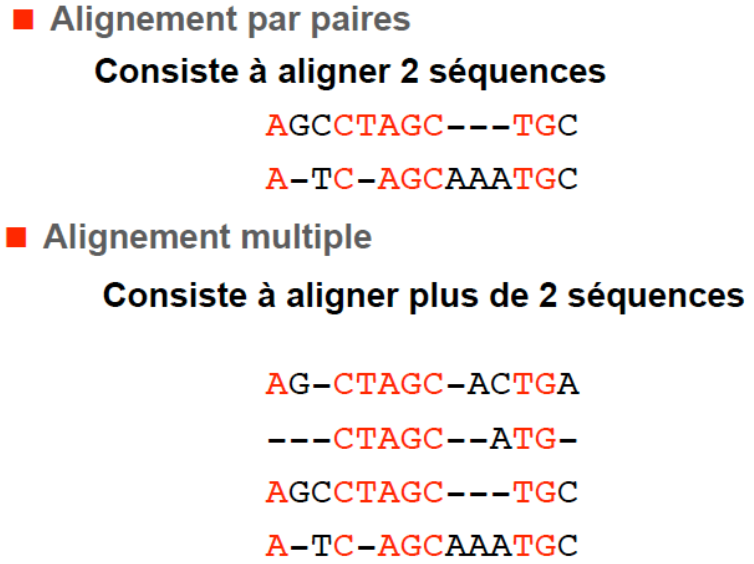
**CHAPITRE VIII : EXEMPLES D'ALGORITHMES ET DE PROGRAMMES EN BIOINFORMATIQUE**

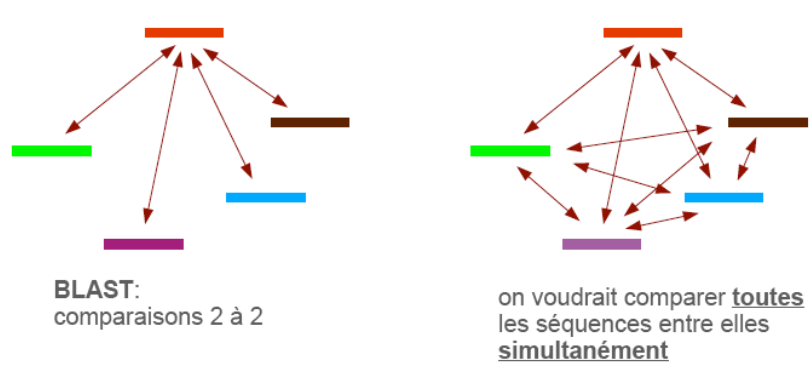
I. **Alignement de séquences**

Un ensemble de séquences similaires peuvent être soumis à un alignement multiple.



**I.1. Alignement multiple de séquences**

* Consiste à comparer toutes les séquences entre elles simultanément.
* Par l’alignement multiple on retrouve plus d’information que dans l’alignement entre pairs.
* L’alignement multiple permet d’identifier les acides aminés/ nucleotides conservés dans des séquences homologues



**Figure 31: Les deux stratégies d’alignement de séquences**



**Figure 32: Exemple d’alignement multiple de séquences peptidiques**

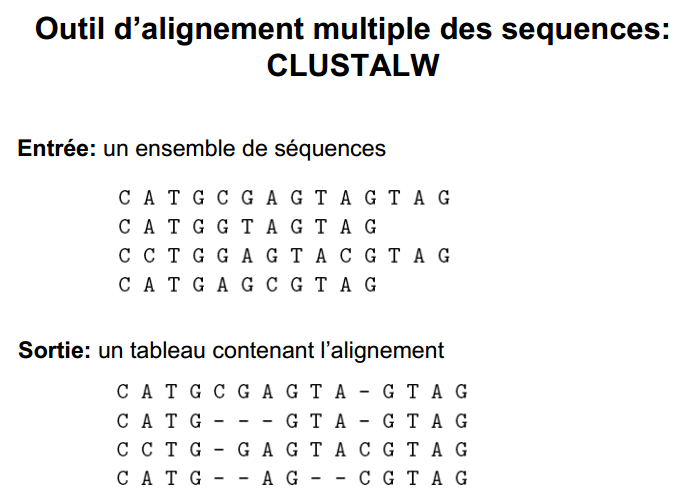
**I.2. But de l’alignement multiple**

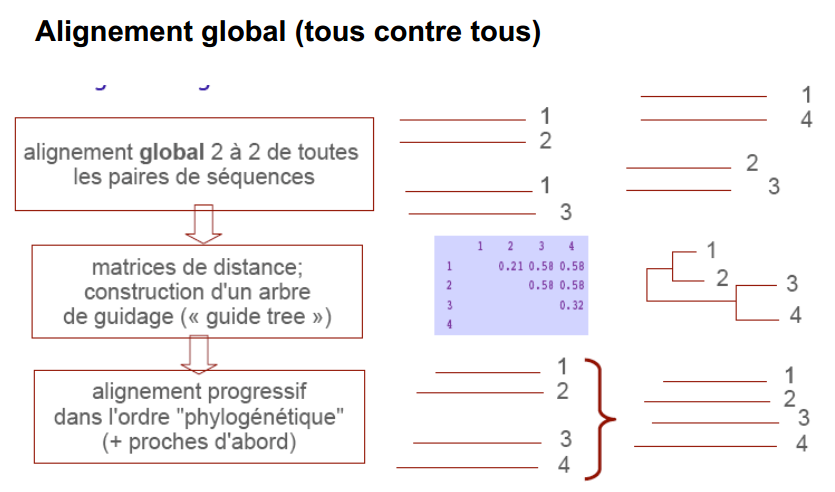
Un alignement d’un ensemble de séquences révèle des aspects que l’on ne pouvait pas visualiser en comparant 2 séquences:

1. Trouver des caractéristiques communes à une famille de protéines.
2. L'identification rapide des sites fonctionnels importants dont la conservation est requise pour la fonction.
3. Prédiction de structure 3D des protéines.
4. Recherche d'amorces de PCR et de sondes spécifiques (l'identification des résidus conservés est nécessaire pour dessiner des amorces de PCR pour amplifier le gène dans une nouvelle espèce).
5. Construire l’arbre phylogénétique des séquences homologues considérées.

**I.3. Outil d’alignement multiple de sequences: ClustalW** [**http://www2.ebi.ac.uk/clustalw/**](http://www2.ebi.ac.uk/clustalw/)

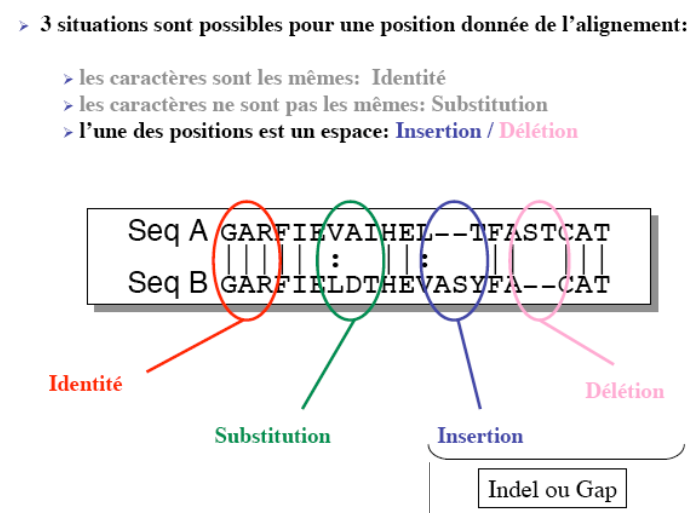
* Le programme ClustalW permet d'établir l'alignement multiple de sequences nucléiques ou protéiques (personnelles ou de banques).
* Il détermine le meilleur alignement de l'ensemble des séquences et les disposent de manière à distinguer les identités, similitudes et différences.
* Il propose en outre de générer l'arbre phylogénétique à partir d'un alignement donné





**I.4. Interpretation des résultats de l’alignement**

* Mise en correspondance des lettre d’une séquence par rapport a une autre  
  - **Match** ou appariement = lettres identiques  
  Conservation des lettres au cours de l’évolution  
  - **Mismatch** ou mésappariement = lettres différentes  
  Changement d’une lettre par une autre = substitution
* L’ordre des lettres est conservé mais décalage possible d’une partie d’une séquence par rapport a l’autre (insertion de blancs entre les lettres)



**I.5.Comment évaluer un alignement de séquences? « Les matrices de substitution »**

Pour comparer deux sequences, on a besoin de comparer chaque position dans les deux sequences.

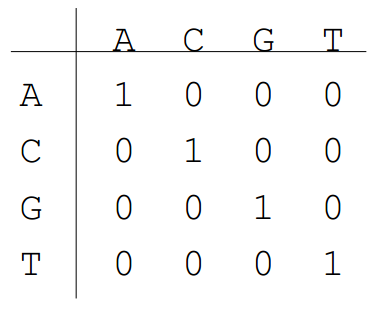
On a besoin de connaitre combien ça “coute” pour substituer une base par une autre ou un acide amine par un autre.

* **Exemple:**  
   - Alanine en Isoleucine  
   - Tryptophan en Glycine

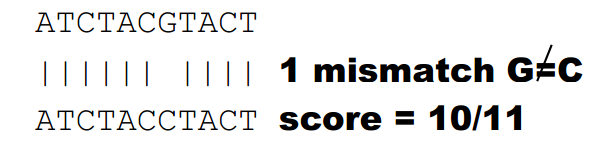
La table qui contient les valeures de toutes les substitutions possible s’appele matrice de substitution (substitution matrix).

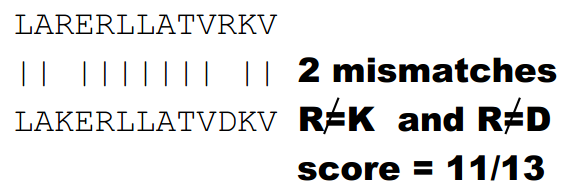
Les matrices de substitution sont utilises pour calculer le score d’un alignement d’une base avec une base (DNA) ou acide amine avec acide amine (proteine).

* **Exemple d’une simple matrice de score**



* **Calcul du score d’un alignement à partir d’une simple matrice de score**

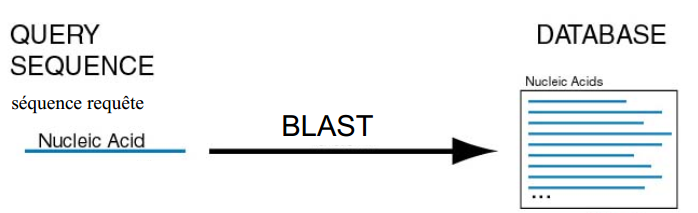


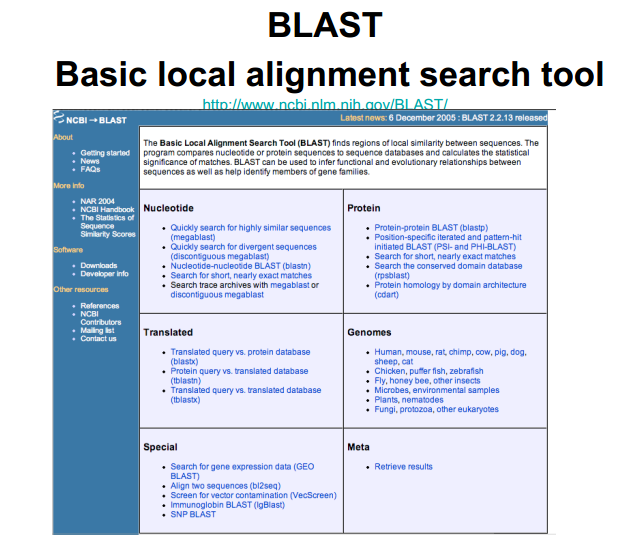


**II. BLAST: Basic local alignment search tool**

**II.1. Définition**

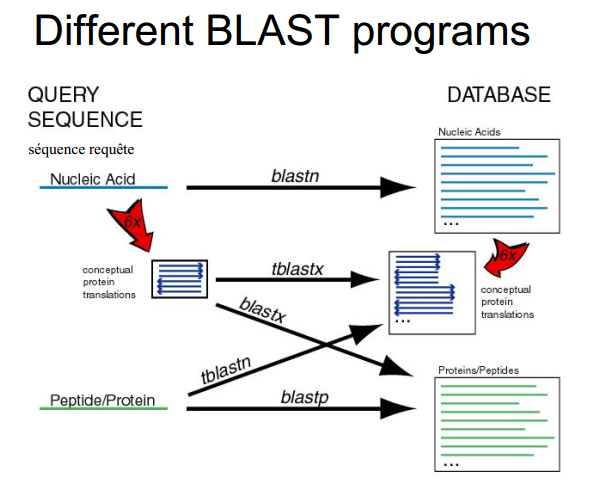
* BLAST est un algorithme de recherche de similarité entre une séquence et une banque de séquences.
* Le programme le plus utilise en bioinformatique (the workhoese of bioinformatics)
* BLAST: incontournable “pipette” de la bioinformatique





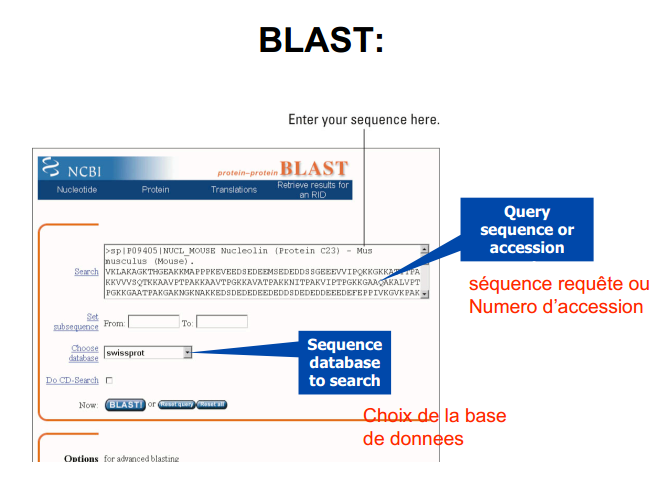
**II.2. Types de programs BLAST**

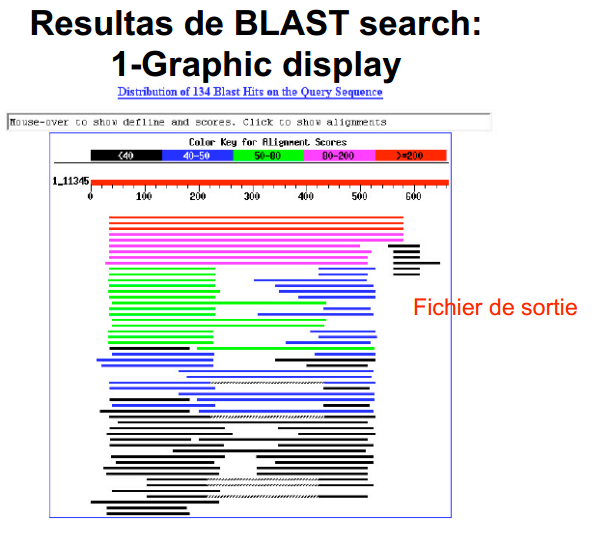
* DNA vs DNA (blastn)
* DNA translation vs Protein (blastx)
* Protein vs Protein (blastp)
* Protein vs DNA translation (tblastn)
* DNA translation vs DNA translation (tblastx)

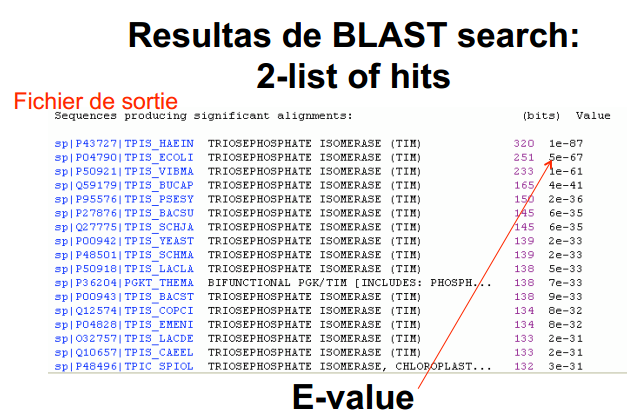


**III.3. La E-value**

La E-value (Expect value) est le nombre attendu d’alignement qui par chance aurait un score ≥S.  
Plus la E-value est proche de 0, plus la similarite est significative.







**Figure 33: Resultats de recherche de séquences par le programme BLAST**