

# 4. Comparaison de séquences

- Comment prédire les fonctions des gènes/protéines ?
- Évolution et similarité de séquences
- Quantifier la similarité de deux séquences
- L'alignement de séquences devient un problème d'optimisation
- Un alignement de séquences vu comme un chemin dans une grille
- Si un chemin est optimal, tous ses chemins partiels sont optimaux
- Coûts et alignement
- Un algorithme récursif
- Éviter la récursivité : une version itérative
- **Cet algorithme est-il efficace ?**

# Le nombre de comparaisons est quadratique

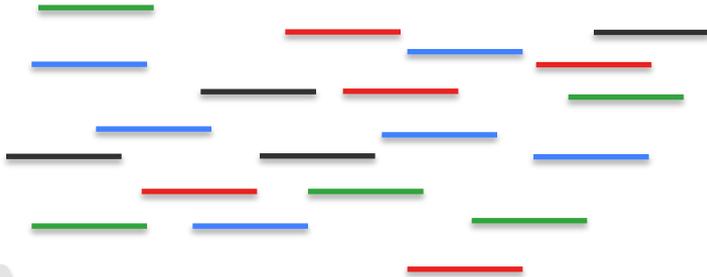
- Algorithme de Needleman et Wunsch (1970)
- Pour deux séquences de longueurs  $N$  et  $M$ , le nombre de comparaisons est  $O(N \times M)$
- La **complexité algorithmique** est dite **quadratique**,  $O(N^2)$
- Si la longueur des séquences double, le nombre de comparaisons (et donc le coût du calcul) quadruple
  - $N \times M \rightarrow (2 \times N) \times (2 \times M) = 4 \times (N \times M)$
- Sur 3 séquences :  $O(N^3)$
- Nécessité de disposer d'algorithmes dédiés  
à l'alignement multiple de séquences

# Blast

- La **méthode la plus fréquemment utilisée** en bioinformatique
- Ne fait **pas d'alignement exact sur toute la longueur** des séquences
- Largement utilisée pour la **recherche de séquences similaires** dans les bases de séquences
  
- **Analyses métagénomiques**
  - Séquençage de tout l'ADN présent dans un échantillon

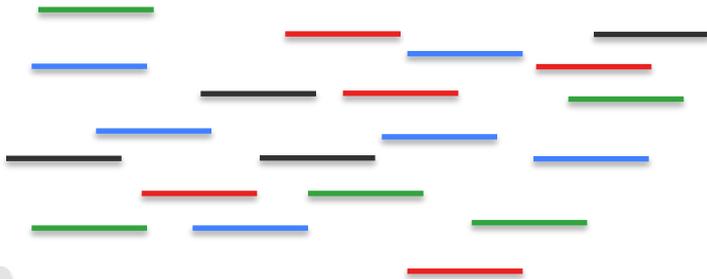
# Blast

- La **méthode la plus fréquemment utilisée** en bioinformatique
- Ne fait **pas d'alignement exact sur toute la longueur** des séquences
- Largement utilisée pour la **recherche de séquences similaires** dans les bases de séquences
- **Analyses métagénomiques**
  - Séquençage de tout l'ADN présent dans un échantillon



# Blast

- La **méthode la plus fréquemment utilisée** en bioinformatique
- Ne fait **pas d'alignement exact sur toute la longueur** des séquences
- Largement utilisée pour la **recherche de séquences similaires** dans les bases de séquences
  
- **Analyses métagénomiques**
  - Séquençage de tout l'ADN présent dans un échantillon
  - « Blaster » chaque read pour identifier l'organisme



# Blast

- La **méthode la plus fréquemment utilisée** en bioinformatique
- Ne fait **pas d'alignement exact sur toute la longueur** des séquences
- Largement utilisée pour la **recherche de séquences similaires** dans les bases de séquences
  
- **Analyses métagénomiques**
  - Séquençage de tout l'ADN présent dans un échantillon
  - « Blaster » chaque read pour identifier l'organisme

