

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 ?

Substitutions, mais aussi insertions/délétions

ACCTCTAATCTATTTCGTACTGCTATT

ACCTCTGAATCCATTCGTCTGCTATT

10 différences

Substitutions, mais aussi insertions/délétions

ACCTCTAATCTATTTCGTACTGCTATT

ACCTCTGAATCCATTCGTCTGCTATT

10 différences

ACCTCT-AATCTATTTCGTACTGCTATT

ACCTCTGAATCCATTCGT-CTGCTATT

2 insertions/
délétions

1 substitution

Les séquences peuvent ne pas avoir la même longueur

```
ACCTCTAATCTATTTCGTACTGCTATT  
TGAATCCATTCGTCT
```

Les séquences peuvent ne pas avoir la même longueur

ACCTCTAATCTATTTCGTACTGCTATT
TGAATCCATTCGTCT

ACCTCT-AATCTATTTCGTACTGCTATT
-----TGAATCCATTCGT-CT-----

Aligner pour comparer

- Pour aligner les séquences, des caractères « - » doivent être insérés
- De très/trop nombreux alignements d'une même paire sont possibles
- Calculer un score pour chaque alignement
- Retenir le ou les alignements dont le score est optimal

L'alignement de séquences à des fins de comparaison est un problème d'optimisation