

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 ?

Distance de Hamming

ACCTCTGTATCTATTTCGGCATCATCAT
ACCCCTGAATCTATTTCGGGATCATCAT

2 différences

ACCTCTGTATCTATTTCGGGATCATCAT
ACCTCTGAATCTATCCGGGATCATGAT

3 différences

Distance de Hamming

- $D(S1, S1) = 0$
- $D(S1, S2) = D(S2, S1)$
- $D(S1, S2) + D(S2, S3) \geq D(S1, S3)$

- C'est une distance au sens mathématique

Calculer la distance de Hamming entre 2 séquences

```
function HammingDistance (Sequence1, Sequence2 : character string
[1,*], Length: integer)
    return integer
    I, Distance: integer
    Distance ← 0
    for I from 1 to Length do
        if Sequence1[I] ≠ Sequence2[I] then
            Distance ← Distance + 1
    endfor
    return Distance
end HammingDistance
```