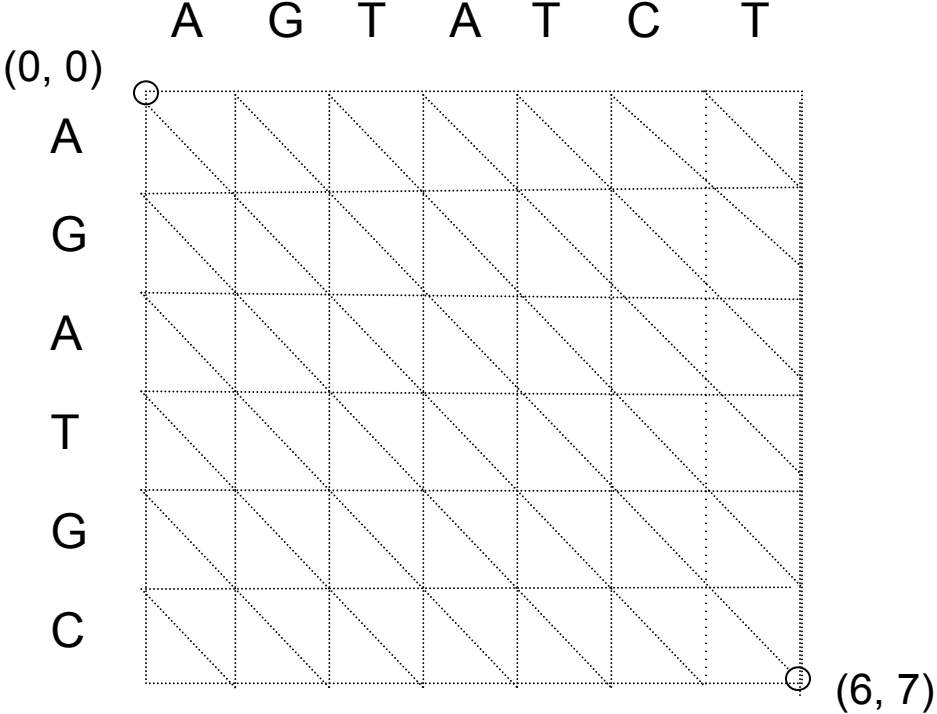


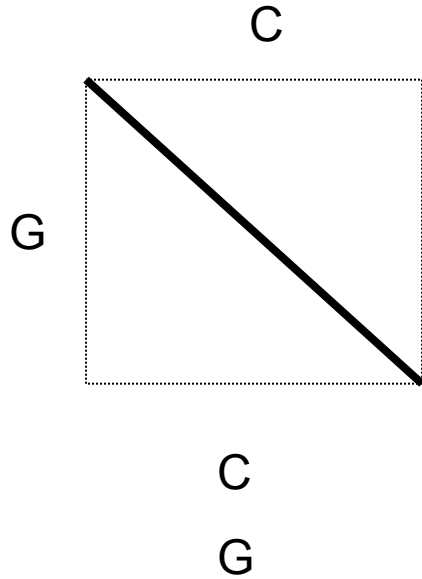
4. Comparaison de séquences

- 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 ?

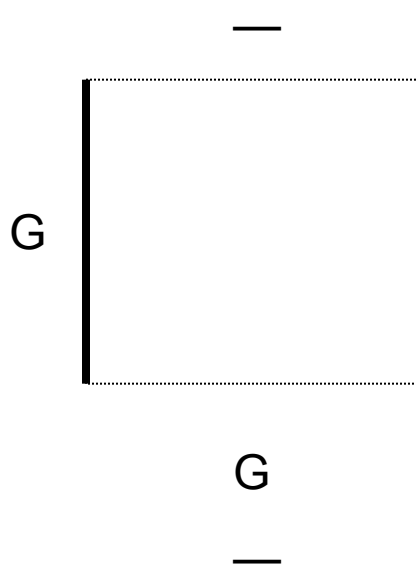
Seconde séquence



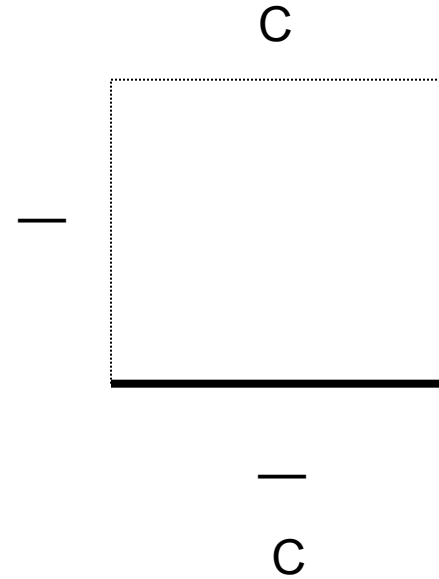
Première
séquence



Substitution

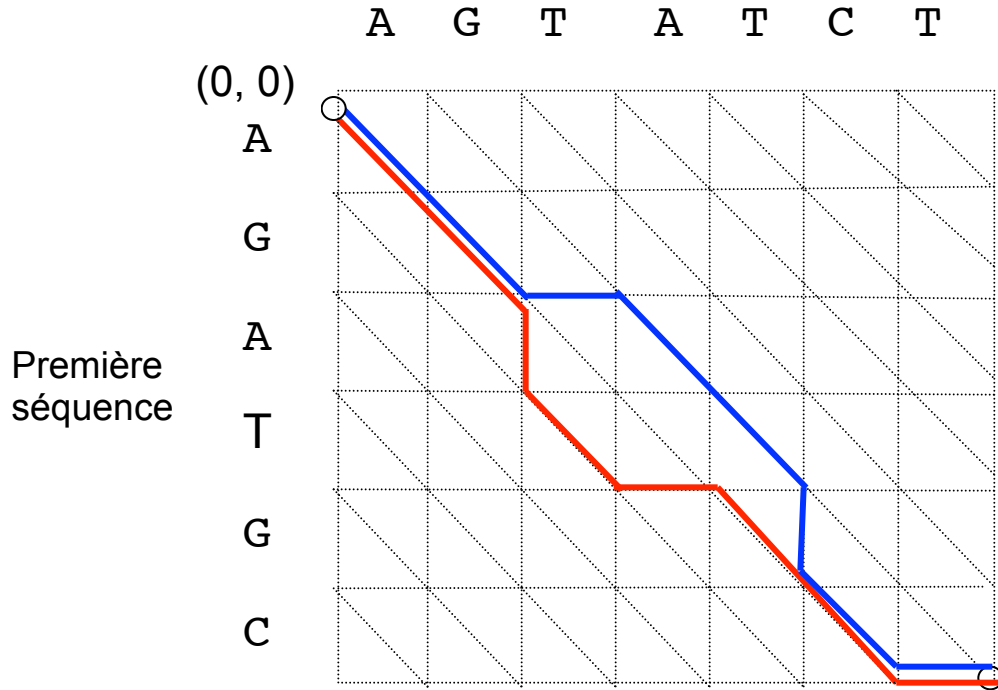


Insertion dans la première séquence (ou délétion dans la seconde)



Insertion dans la seconde séquence (ou délétion dans la première)

Seconde
séquence



A G - A T G C -
A G T A T - C T

A G A T - G C -
A G - T A T C T

(6, 7)

Alignement optimal

- Chaque substitution, délétion/insertion possède un coût
- Le coût d'un alignement est la somme des coûts obtenue en parcourant le chemin associé
- On recherche l'alignement dont le coût est le plus faible
- Pour une même paire de séquences, il peut exister plusieurs alignements optimaux, de même coût