

Analyse des séquences biologiques

Nous allons utiliser les libraires “Biostrings” et “DECIPHER” pour aligner les séquences biologiques et analyser les motifs séquentiels.

Travail 1 : Alignement des séquences biologiques (ADN et protéines)

1. Installer et charger le package “Biostrings”
2. Créer deux séquences ADN :

```
seq1 <- "ACGTCGATCGGCTA"
seq2 <- "ACGTCGATGGTCTA"
```

3. Aligner les deux séquences ADN en utilisant l'algorithme d'alignement global de Needleman-Wunsch. Utiliser pour cela la fonction `pairwiseAlignment()`.
4. Quel est le score d'alignement entre ces deux séquences ?
5. Créer et aligner les deux séquences protéiques en utilisant l'algorithme de Needleman-Wunsch et la matrice de substitution BLOSUM62.

```
prot1 <- "MPRCLCQRY"
prot2 <- "PYRCKCQR"
```

6. En utilisant la matrice BLOSUM62, comment les substitutions entre acides aminés sont-elles pénalisées ? Comment cela influence-t-il le score final par rapport à l'alignement ADN ?

Travail 2 : Analyse de motifs séquentiels dans des séquences biologiques

1. Créer un ensemble de séquences ADN représentant des motifs évolutifs potentiels.

```
sequences <- DNASTringSet(c("ACGTCGATCGGCTA",
"ACGTCGATGGTCTA", "ACGTCGCTCGTCTA", "ACGTCGATCGTCTA"))
```

2. Utilisez la fonction `AlignSeqs()` du package DECIPHER pour détecter les motifs communs dans ces séquences.
3. Quelle est la longueur des motifs alignés et les régions les plus conservées entre les séquences ? Est-il possible d'identifier des régions qui pourraient indiquer des ancêtres communs ?