***Exercice 1 :***

Écrire un programme MapReduce pour calculer la distance de Hamming (le nombre de positions où les nucléotides sont différents) entre deux séquences ADN de même longueur.

Exemple :

Séquence 1 : AGCTTAGCTA

Séquence 2 : AGCTCAGTTA

Distance de Hamming : 3

***Exercice 2 :***

Écrire un programme MapReduce pour analyser les codons (groupes de trois nucléotides) dans une séquence ADN et compter le nombre d'occurrences de chaque codon.

Exemple :

AGCTTAGCTA

CGTACGTA

AGC 1

TTA 1

GCT 1

CGT 1

ACG 1

GTA 1

***Exercice 3:***

Écrire un programme MapReduce pour identifier les séquences répétées inversées (palindromes) dans une séquence ADN.

AGCTTAGCTA

CGTACGTA

AGCT 1

CTTA 1

CGTA 1

***Exercice 4 :***

Écrire un programme MapReduce pour trouver les régions riches en GC dans une séquence ADN (régions avec un pourcentage de GC supérieur à un seuil spécifié).

AGCTTAGCTA

CGTACGTA

GCTTA 60%

CGTAC 60%

***Exercice 5 :***

Écrire un programme MapReduce pour calculer la fréquence des dinucléotides (paires de nucléotides) dans une séquence ADN.

AGCTTAGCTA

CGTACGTA

AG 2

GC 2

CT 2

TT 1

TA 2

CG 1

GT 1

***Exercice 6 :***

Écrire un programme MapReduce pour trouver les motifs répétés (séquences de longueur fixe) dans une séquence ADN.

***Exercice 7 :***

Écrire un programme MapReduce pour identifier les mutations (différences) entre deux séquences ADN.

***Exercice 8 :***

Écrire un programme MapReduce pour trouver toutes les séquences palindromiques d'une certaine longueur dans une séquence ADN

***Exercice 9 :***

Écrire un programme MapReduce pour identifier les paires de k-mers adjacentes (où k est la longueur du k-mer) dans une séquence ADN.

***Exercice 10 :***

Écrire un programme MapReduce qui trouve les k-mers (sous-séquences de longueur k) les plus fréquents dans une séquence ADN.