

# Recherche d'un Gène Défaillant : La Quête de la précision grâce à l'Algorithme de Boyer-Moore

## Introduction :

La recherche et la caractérisation des gènes défaillants sont essentielles pour comprendre les maladies génétiques et développer des traitements ciblés. Ce projet a pour but de montrer comment l'algorithme de Boyer-Moore est une technique de recherche de motifs efficace dans la quête spécifique de trouver rapidement et efficacement les gènes défectueux au sein d'un génome.

## Recherche naïve :

La recherche naïve consiste à parcourir le texte et à renvoyer la position de la première lettre du mot.

## Algorithme de Boyer-Moore-Horspool :

L'algorithme de Boyer-Moore-Horspool est un algorithme de recherche textuelle permettant de chercher un paterne dans une chaîne de caractère plus rapidement qu'un algorithme naïf. Il fonctionne au début comme un algorithme naïf, mais vérifie les caractères de droite à gauche, si un caractère ne correspond pas, alors on va suivre la règle du mauvais caractère.

Le but de la règle du mauvais caractère est de sauter un nombre exact de caractère dans la boucle afin de continuer sans rater une éventuelle occurrence du paterne / mot recherché par exemple si nous cherchons le mot "test", et que dans l'emplacement dans le texte actuelle on ne trouve pas de "t" mais un "e" alors, on sautera le nombre de caractère nécessaire afin d'aligner le "e" du paterne et le "e" du texte. Cette règle permet de passer certain passage du texte et d'effectuer beaucoup moins de vérification que l'algorithme naïf.

## Objectif du projet :

L'objectif de ce projet est de démontrer comment l'algorithme de Boyer-Moore peut être adapté pour localiser avec précision des séquences génétiques anormales ou défectueuses au sein de vastes ensembles de données génomiques. Nous aborderons les principes fondamentaux de l'algorithme, son implémentation en Python, et son application spécifique à la détection de gènes altérés.

## **Points principaux**

1. **Adaptation pour la Recherche Génétique** : Comment adapter cet algorithme pour rechercher des séquences génétiques spécifiques associées à des gènes défaillants.  
On réduit tout d'abord l'alphabet à quatre lettres (G, T, A, C) pour effectuer la recherche efficacement.

## **Thèse :**

La recherche génétique est au cœur de la médecine moderne, offrant des perspectives sur la compréhension et le traitement des maladies génétiques. L'identification précise des gènes défaillants dans le génome humain est cruciale pour diagnostiquer et traiter plus de 1500 infections telles que la mucoviscidose, la trisomie, divers cancers héréditaires et bien d'autres.

L'utilisation de l'algorithme de Boyer-Moore en Python pour la recherche de gènes défaillants propose une approche efficace et précise, ouvrant ainsi la voie à des avancées significatives dans la compréhension et le traitement des maladies génétiques.

## **Sources :**

Genomes : [kaggle.com/datasets/zusmani/mygenome](https://kaggle.com/datasets/zusmani/mygenome)

Source d'information pour l'algorithme de Boyer-Moore-Horspool :

[https://fr.wikipedia.org/wiki/Algorithme\\_de\\_Boyer-Moore-Horspool](https://fr.wikipedia.org/wiki/Algorithme_de_Boyer-Moore-Horspool)

## Planning et répartition des tâches :



