**información Básica del PIA**

**¿Qué debo saber antes de hacer el PIA?**

Cada **nucleótido** puede ser considerado como una letra de un alfabeto sencillo de 4 letras **( ABCD , RGFB , FRTYL, etc …)** con el que se escriben los mensajes biológicos. Hay segmentos del **ADN** con una determinada cantidad de **nucleótidos** que contienen la información necesaria para sintetizar una secuencia de **aminoácidos**

Interfaz de usuario gráfica

Descripción generada automáticamente

**¿Cuál es el objetivo de este proyecto?**

Se busca determinar cuál de estos algoritmos (que se presentan a continuación) es más eficiente en la **búsqueda de patrones en secuencias de ADN y aminoácidos en secuencias biológicas**.

Es decir:

Encontrar en donde coinciden el **aminoácido o ADN** en la **secuencia biológica** gigante

Interfaz de usuario gráfica

Descripción generada automáticamente

En este caso no se ve que coincidan ni el **Aminoácido** ni el **ADN,** por lo que se tiene que usar un **algoritmo** (como los que veremos a continuación)para encontrar patrones donde **coincida** y cuantas **veces coincide**

**¿Qué debemos investigar?**

Algoritmos para encontrar patrones en las secuencias Biológicas:

Nota: Los números al lado de los nombres de los métodos son de la bibliografía del documento

* Estos son los algoritmos que menciona mas frecuentemente y usa como base para las pruebas:

"New" de Lecroq [13]

SSABS [16]

TVSBS [19]

SBNDM2 [6]

Boyer-Moore [3]

FAOSO [4]

SBNDM [14,15]

Our algorithm [18]

( El algoritmo de ellos )

-Modificacion del algoritmo Boyer–Moore–Horspool [7]

BMHq

-Algoritmo sin desarrollar en la pagina 4

BMH4

BMH4b

Lee una palabra en 32 bits

BMH4c

Lee dos medias palabras consecutivas

BMH2

Basada en BMHq el rango de letras en código ASCII incremento de 4 a 20 (lee el alfabeto de aminoácidos en lugar del de ADN de la imagen)

BMH2c

Lee 2-gram como 16-bit media palabra

* Estos algoritmos que usa mas que nada para comparación, no debemos investigarlo, pero podemos tenerlos como referencia que se pueden usar:

Boyer–Moore–Horspool [7]

Kim y Shawe-Taylor [10]

bucle de salto rápido [8]

Wu-Manber [20]

Berry-Ravindran [2]

Zhu-Takaoka

QS de Sunday [17]

**¿Cuáles son los alfabetos de entrada para comparar con la secuencia biológica?**

Vamos a tener 2 alfabetos para comparar con los algoritmos que vamos a crear

**Alfabeto de ADN**

Composición: El alfabeto del ADN está compuesto por 4 caracteres, que representan las bases nitrogenadas: **A, C, G, T** (Adenina, Citosina, Guanina y Timina).

**Alfabeto de aminoácidos**

Composición: El alfabeto de aminoácidos está compuesto por 20 caracteres, que representan las diferentes cadenas laterales de los aminoácidos: **A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y.**

Cada letra corresponde a un aminoácido, por ejemplo:

A: Alanina

C: Cisteína

D: Ácido aspártico

E: Ácido glutámico

F: Fenilalanina

G: Glicina

H: Histidina

K: Lisina

Y así sucesivamente.