- **Funciones de creación de secuencias y base de datos (create_sequence y create_database):**
  create_sequence: Genera una secuencia aleatoria de bases nucleotídicas (A, C, G, T) con una longitud aleatoria entre 10 y 20 bases.
  create_database: Crea una base de datos ficticia de secuencias genéticas. Genera 50,000 secuencias utilizando la función create_sequence.
- **Cálculo de la Entropía de Shannon (shannon_entropy):**

  La entropía de Shannon se utiliza para medir la incertidumbre en una distribución de probabilidad. En este caso, se calcula la entropía para cada secuencia genética, lo que nos da una idea de su variabilidad o complejidad. Cuanto mayor sea la entropía, mayor será la diversidad en la secuencia.

  La función shannon_entropy calcula la entropía de Shannon para una secuencia dada. Utiliza la fórmula de la entropía de Shannon:

  $H(X)=-\sum_{i=1}^{n} P(x_i)\cdot \log_2(P(x_i))$, donde $P(x_i)$ es la probabilidad de ocurrencia del carácter x_i en la secuencia.

- **Filtrado de secuencias (filter_sequences):**
  Esta función filtra las secuencias basadas en un umbral de entropía. Aquellas secuencias con una entropía por encima del umbral se consideran suficientemente diversas y se mantienen en la base de datos filtrada.
  El umbral de entropía se establece en 1.5 como un valor arbitrario. Se puede ajustar según los requisitos específicos del análisis y la naturaleza de los datos.
- **Obtención de combinaciones y conteo de motivos (get_combinations y count_motif):**

  get_combinations: Genera todas las posibles combinaciones de bases nucleotídicas para un tamaño de motivo dado.

  count_motif: Cuenta el número de veces que un motivo específico aparece en todas las secuencias de la base de datos.

- **Obtención de motivos (get_motif):**

  Utiliza las funciones anteriores para encontrar el motivo más frecuente en las secuencias filtradas para tamaños de motivo dados.

  Itera sobre todas las combinaciones posibles de bases nucleotídicas y cuenta la frecuencia de cada una en las secuencias filtradas. Luego, devuelve el motivo más frecuente junto con su recuento.

- **Función principal (main):**

  Crea la base de datos de secuencias genéticas.

  Filtra las secuencias utilizando la Entropía de Shannon.

  Obtiene los motivos de tamaño 6 y 8 de las secuencias filtradas.

  Imprime los resultados.

En cuanto a la elección del factor de filtración (umbral de entropía), el valor de 1.5 fue seleccionado de manera arbitraria para este ejemplo. La elección del umbral puede variar según la naturaleza de los datos y los requisitos del análisis. Un valor más alto filtrará más secuencias, manteniendo solo aquellas con mayor variabilidad, mientras que un valor más bajo podría permitir más secuencias en la base de datos filtrada. Se recomienda ajustar este valor según el conocimiento del dominio y los objetivos del análisis específico.