

Predicción Taxonómica de Especies

Usando Frecuencias de Codones

Alumno: Esconjaureguy leonel

profesor: Germán Rodriguez

Tutor: Silvia Vilela

Comisión: #61190

Introducción:

Expresión Heteróloga de Genes

Definición: Expresión de genes de una especie en un organismo distinto (e.g., en bacterias)

Optimización de Codones

Función: Adapta la secuencia genética a los codones preferidos del organismo hospedador

Beneficios:

Aumenta la eficiencia con la que se producen proteínas, al ajustar el código genético (sesgo de uso de codones) a las preferencias del organismo

Aumenta la estabilidad de la proteína

Sesgo de Uso de Codones

- **Definición:** Preferencia de ciertos codones por diferentes especies para un mismo aminoácido
- **Factores Asociados:**
 - Eficiencia de Traducción
 - Disponibilidad de tRNAs específicos
- **Ejemplos de Preferencias:**
 - Codones más comunes en **eucariotas o procariotas**
- **Impacto en Transferencia Génica:**
 - Influencia en la **expresión de genes** al transferir entre especies

El objetivo final

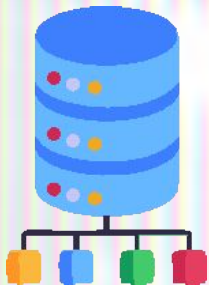
Desarrollar nuevos medicamentos, mejorar la calidad y rendimiento de los cultivos, y crear productos innovadores en biotecnología

Objetivo:



- Desarrollar un modelo de clasificación supervisado que aprende patrones distintivos en las frecuencias de codones para predecir con precisión el reino biológico de una especie (virus, bacterias, eucariotas).

Dataset:



El conjunto de datos examina las frecuencias de uso de codones en el ADN codificante de muestras de organismos de diferentes taxones.

Variables:

categorías:

- Kingdom
- SpeciesName

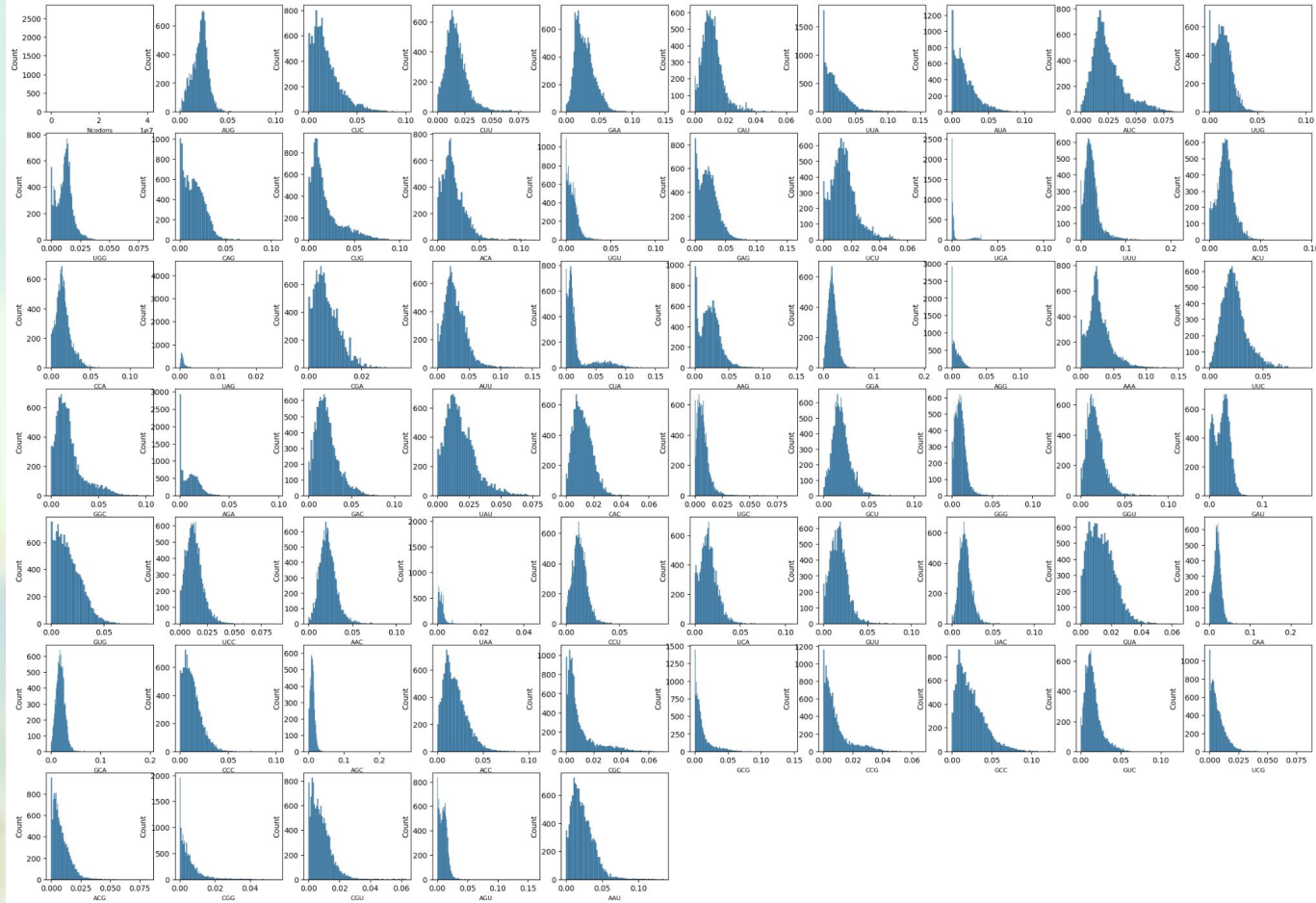
Dimensiones:

- columnas: 69
- filas: 13028

categorías:

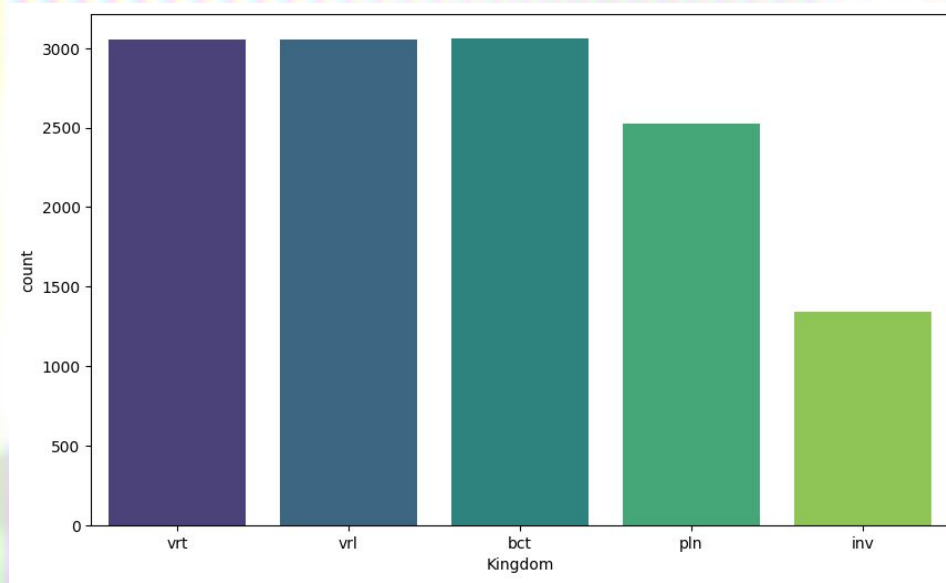
- Codon Frequencies:
- Ncodons
- DNAtype

Análisis univariado:



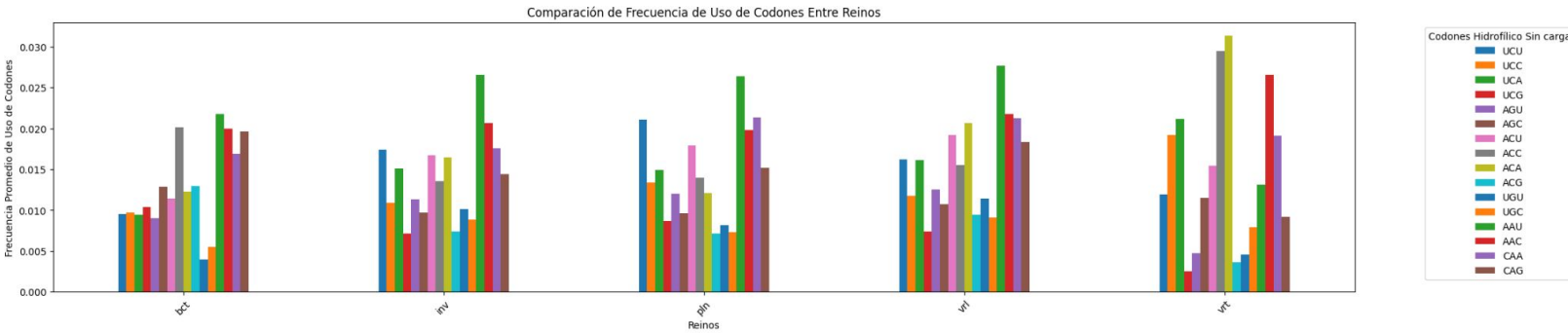
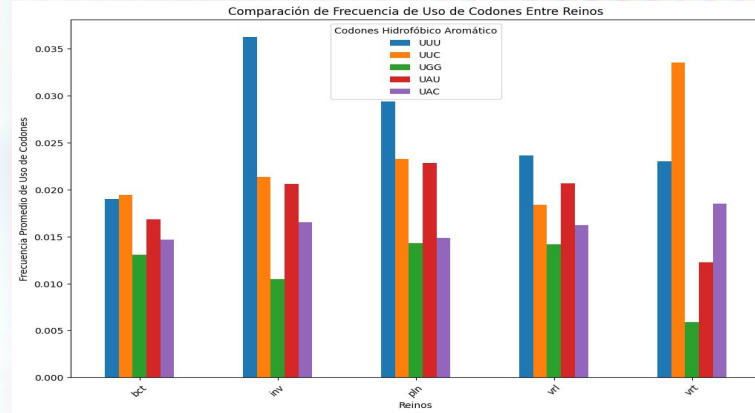
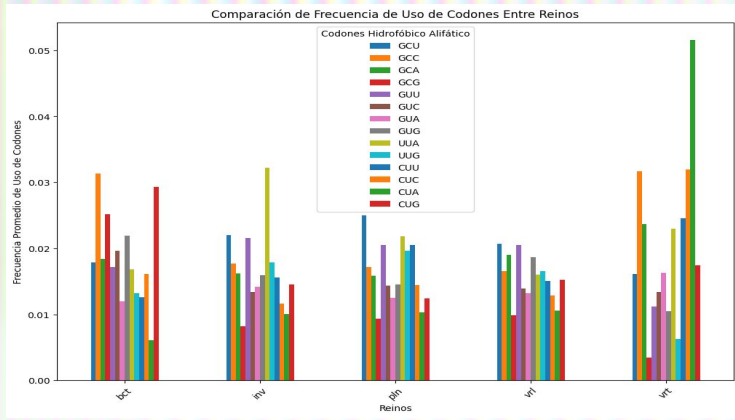
El gráfico muestra que ciertos codones se utilizan con más frecuencia que otros. Esto sugiere que podrían ser más eficientes para crear proteínas, lo que podría ayudar a los organismos a adaptarse mejor a su entorno.

Análisis univariado:



La variable que queremos predecir es el reino al que pertenece el gen en estudio. Conocer esta información es crucial, ya que nos orienta sobre los pasos a seguir en nuestra investigación y la implementación de herramientas adecuadas

Análisis bivariado:



Los análisis univariado y bivariado ayudan a identificar patrones y relaciones entre la frecuencia de codones y el reino biológico, lo cual es una buena base para inferir su potencial como predictor.