**Descripción de script para la identificación de la proteína E, correspondiente al genoma del viral de la cepa   
SARS-CoV-2 (*severe acute respiratory syndrome coronavirus)***

## A partir del script adnframes.py ubicado en el repositorio GitHub (<https://github.com/pabeli-bero/Tareas.python/blob/main/adnframes.py> ) se procedió a identificar la proteína E.

## Líneas 5-19: Se realizaron las variables las bases nitrogenadas para cada organismo para formar el ADN complementario a RNAm, con las indicaciones adenina = adenina "A":"A", citosina=citosina "C":"C", timina=uracilo "T":"U”, y guanina = guanina "G":"G"

## Líneas 21-86: Se usaron los codones de inicio y paro para la codificación de aminoácidos, siendo en total 61 codificaciones para aminoácidos y 3 de paro. (UAA, UAG, UGA) estos codones se forman a partir de la combinación de tres bases nitrogenadas (A, C, G y U) para ARNm.

## Líneas 88-153: Se agregó la estructura primaria de la proteína de acuerdo a la nomenclatura de aminoácidos utilizada en genética molecular y en bases de datos, la cual consta de una sola letra, esto permite la representación de la estructura primaria de una proteína disponiendo a la izquierda el aminoácido N-terminal y a la derecha el aminoácido C-terminal.

## Líneas 155-175: Se introdujo la secuencia de la proteína E en el sentido 5’ 3’ dentro del archivo covid19.txt en la variable f, metiendo los resultados en u renglón nuevo dentro de la variable x.

## Líneas 178-210: Se buscó la secuencia complementaria en la variable X, obteniéndose la cadena anti sentido y la parte codificante. También se realizó la síntesis de ARNm – y +, para crear posteriormente identificar los codones.

## Líneas 212-249: Se identificaron los marcos de transcripción y paro a través de la construcción de codones de inicio/paro utiizando longitudes de palabras no mayores a 3 ciclor for y condicionantes if, else.

## Líneas 252-266: Se identificaron las cmbionaciones de adicionales mediante las condicionantes (if y else), codificantes para un aminoácido y sus codones, usando ciclos for dentro de las variables i y j.

## Líneas 270-282: Se obtuvieron los marcos y longitud de lectura abiertos. Se imprimieron los fragmentos, aminoácidos correspondientes en la nomenclatura de genética molecular, y finalmente se obtuvo la secuencia de la proteína E. Esta fue corroborada mediante blast en NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).