

Cazadores de genes: traducción de codones en secuencias de ADN mediante Python para la enseñanza en básica secundaria

*Gene Hunters: Codons Translation into DNA sequences
Through Python for Secondary School Education*

DARÍO ALFONSO CUELLO MEJÍA

Grupo CIDEINNOVA
Línea Grupos y Semilleros de Investigación
Centro Industrial y de Desarrollo Empresarial
Tecnoacademia, Semillero de Materiales,
Procesos de Manufactura y Automatización (MAPRA)
SENA, Soacha-Colombia
dacuellom@unal.edu.co

YICELY KATHERINE HERNÁNDEZ GÓMEZ

Grupo CIDEINNOVA
Línea Grupos y Semilleros de Investigación
Centro Industrial y de Desarrollo Empresarial
Tecnoacademia, Semillero de Materiales
Procesos de Manufactura y Automatización (MAPRA)
SENA, Soacha-Colombia
ykhernandez@sena.edu.co

CARLOS HERNANDO NIÑO RIVEROS

Grupo CIDEINNOVA
Línea Grupos y Semilleros de Investigación
Centro Industrial y de Desarrollo Empresarial,
Tecnoacademia, Semillero de Bioprocesos y
Biotecnología Aplicada (BIBA)
SENA, Soacha-Colombia
chninor@unal.edu.co

Fecha de recepción: 02-11-2018
Fecha de evaluación: 09-11-2018
Fecha de aceptación: 13-11-2018





Cazadores de genes: traducción de codones en secuencias de ADN mediante Python para la enseñanza en básica secundaria

Resumen

El presente estudio presenta un primer código en Python cuyo algoritmo identifica, ubica y traduce codones en secuencias de nucleótidos de ADN. La Biología Molecular es un campo donde el costo de los reactivos limita su enseñanza; mientras tanto, la incorporación de las tecnologías de la información (TIC) en el aula dista de ser una realidad en Colombia. El programa de Tecnoacademias es un escenario para desarrollar la innovación en comunidades vulnerables. En este programa los aprendices buscan nuevas formas de implementar las TIC en la difusión de la biología molecular y la bioquímica.

Este proyecto comprende una herramienta docente destinada a poblaciones de la educación media colombiana. El mismo muestra un inicio en la elaboración de herramientas interactivas relacionadas con la enseñanza de las disciplinas biotecnológicas. Se espera desarrollar interfaces amigables y evaluar su impacto sobre los estudiantes.

Palabras clave: aminoácido, biología molecular, codón, Python, residuo, traducción de RNA.

Abstract

The present study presents a first PYTHON code whose algorithm identifies, locates and translates DNA codons in nucleotide sequences. Molecular Biology is a field where reagent costs limits its teaching in public schools; meanwhile, information technologies (ICT) use in the classroom is far from a reality in Colombia. Techno-academies are a scenario for innovation development in vulnerable communities. In this program, apprentices seek new ways to implement ICT in the spread of molecular biology and biochemistry.

This project encompasses a teaching tool aimed at the population of Colombia's secondary education. It shows a start in the development of interactive tools related to the teaching of biotechnological disciplines. It is expected to develop friendly interfaces and evaluate the program's impact on the students.

Keywords: Aminoacid, Codon, Molecular biology, Residue, RNA Translation, Python.

Introducción

La Bioquímica y la Biología Molecular son dos campos de la ciencia con un alto crecimiento y desarrollo tecnológico en los últimos años. En consecuencia ejercen gran influencia en la vida cotidiana, abarcando ámbitos como la salud, la biotecnología y la agroindustria, entre otros (Moreno, 2015). Normalmente, los reactivos empleados en la ejecución de prácticas experimentales en estos ámbitos son costosos limitando su acceso. Por otra parte, la incorporación de las tecnologías de la información (TIC) a la práctica docente habitual está lejos de ser una realidad en Colombia (Morcillo-Ortega & López-García, 2007). En consecuencia es eficiente incentivar el uso de las TIC mientras se llenan nichos en la enseñanza de las ciencias básicas que no pueden suplirse en contextos de escasos recursos económicos.

El presente estudio plantea el desarrollo de un programa capaz de identificar codones en una secuencia de nucleótidos de ADN y traducir el respectivo codón en el correspondiente residuo de aminoácido que resultaría tras la traducción de la secuencia. Este programa se constituye como una herramienta de docencia en el aula de clase.

El programa de Tecnoacademias se plantea como un escenario de aprendizaje, dotado de tecnologías emergentes para desarrollar competencias orientadas a la innovación mediante la formación por proyectos en comunidades vulnerables. Su población objetivo son los estudiantes de la educación básica y media constituyéndose como un complemento de su educación formal (Tecnoacademias, 2017). Dentro de este programa los aprendices vinculados a los semilleros de investigación Bioprocesos y Biotecnología aplicada (BIBA) y el semillero de investigación en materiales, procesos de manufactura y automatización (MAPRA) han buscado nuevas formas de implementación de las TIC para la difusión del conocimiento en áreas como la biología molecular y la bioquímica desde un enfoque

práctico y destinado a alumnos de los grados octavo y noveno, a los cuales estos pertenecen.

El desarrollo de un programa capaz de identificación de codones y residuos en una secuencia de nucleótidos de ADN representa una herramienta práctica para la docencia de las ciencias básicas, mediante un enfoque práctico y centrado en la socialización de la biología molecular en los grados correspondientes a la educación básica secundaria colombiana. En este proyecto se desarrolló un primer algoritmo para la identificación, ubicación y traducción de codones en secuencias de nucleótidos de ADN; enfocado como una herramienta docente en la ejecución de talleres prácticos destinados a poblaciones de los grados octavo y noveno, de acuerdo con el sistema educativo colombiano.

Metodología

Lenguaje de programación

Para el desarrollo del programa y la implementación del algoritmo se utilizó el lenguaje de programación Python. Python es un lenguaje interpretado de código abierto cuya principal característica es la simplicidad en la sintaxis (Python Software Foundation [US], 2017). Al ser un lenguaje tan sencillo de leer y escribir, es una herramienta ideal para la enseñanza con la que se pueden implementar diferentes tipos de proyectos tanto a nivel educativo como a nivel profesional (Tobis, Papert, Siegel & Nevins, 2015). En este caso, Python se utiliza para implementar el algoritmo de identificación de codones y residuos en una secuencia de nucleótidos de ADN, enfocado a la enseñanza de los conceptos relacionados a través del programa.

Desarrollo del algoritmo

Para el desarrollo del algoritmo se buscó emular la traducción a residuos de aminoácidos directamente desde tripletas de codones de ADN embebidos en la secuencia; omitiendo la transcripción a secuencias de ARN como es establece en el dogma central de la biología molecular (Crick, 1970).

En la Figura 1 se muestra el diagrama de bloques del programa. La ejecución de cada uno de los bloques se realiza de forma secuencial, utilizando los resultados obtenidos en los bloques anteriores. De igual forma, cada uno de los bloques se implementa en una *función*, permitiendo la re-utilización de cada bloque de código para la organización y el procesamiento de la información.

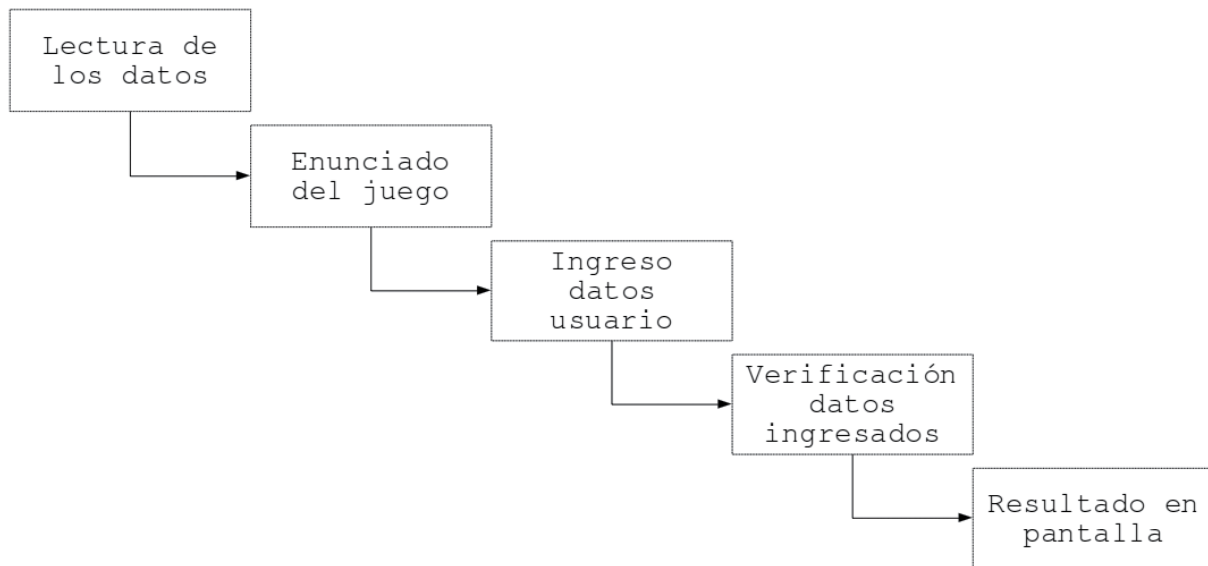


Figura 1. Diagrama de bloques del algoritmo. Fuente: Elaboración propia.

La primera etapa es la lectura de los datos por parte del programa. La secuencia de nucleótidos de ADN se encuentra guardada en un archivo de texto plano (.txt), el cual es leído por el programa y organizado en seis bloques de diez secuencias para que sea más legible para el usuario. Esa lectura y organización de

los datos se implementa con la función ***open_file()*** que tiene como parámetros de entrada la ubicación del archivo de texto y como parámetros de salida, una variable llamada ***adn_string*** en donde queda almacenada la secuencia. En la Figura 2 se muestra el resultado obtenido en pantalla.

La cadena de ADN es:

```

GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC

```

Figura 2. Resultado en pantalla de la función ***open_file()***. Fuente: Elaboración propia.

La segunda etapa es la representación en pantalla del enunciado del programa. En este bloque se imprime en consola la información necesaria para completar la actividad, brindando información relevante al usuario para comprender en qué consiste el juego y los conceptos generales de lo que es una secuencia

de nucleótidos de ADN. Esta parte del programa se implementa en una función llamada **statement()** que no cuenta con parámetros de entrada, siendo la cadena de texto del enunciado, el único parámetro de salida. En la Figura 3 se muestra un prototipo del enunciado que se imprime en pantalla.

Las cadenas de ADN se componen de 4 bases nitrogenadas:

+++++++ Adenina (A), Timina (T), Citosina (C) y Guanina (G) +++++++

A su vez, las combinaciones de a tres de estos componentes conforman 20 aminoácidos.

La actividad consiste en adivinar las combinaciones de BASES NITROGENADAS:

ATENCIÓN: Recuerde que son combinaciones de 3 bases nitrogenadas (A-T-C-G)

Figura 3. Enunciado del juego implementado en la función *statement()*. Fuente: Elaboración propia.

La tercera etapa es el ingreso de los datos requeridos por parte del usuario. En este bloque se solicita al usuario que ingrese una posible combinación de bases nitrogenadas teniendo en cuenta las condiciones del juego presentadas en el bloque anterior. Esta etapa se implementa en una función llamada

usrdata_input(), en donde solamente se tiene un parámetro de entrada, el cuál es la combinación ingresada por el usuario y ésta se almacena en la variable **user_input**. En la Figura 4 se muestra el enunciado que se imprime en pantalla.

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): ■

Figura 4. Solicitud de ingreso de datos por el usuario en la función *usrdata_input()*. Fuente: Elaboración propia.

La cuarta etapa es la verificación de los datos ingresados por el usuario. La primera parte de la verificación consta de dos fases. La primera de ellas es revisar que los datos ingresados sean correctos, que no tengan más de 3 bases hidrogenadas y que efectivamente la combinación sea de dichas bases. La segunda fase es la de revisar que tipo de codón es, verificando que tipo de aminoácido fue ingresado e imprimiendo en pantalla información relevante sobre el mismo. Esta verificación se implementa en la función **aminoacid()**, cuyo parámetro de entrada es la variable **user_input** en donde se encuentra almacenada la combinación ingresada por el usuario, y el parámetro de salida es la cadena de texto que brinda la información sobre el aminoácido encontrado.

La segunda parte de la verificación es la revisión de la cadena de ADN, identificando si la combinación ingresada por el usuario se encuentra en dicha cadena. En esta parte se recorre toda la cadena cargada del archivo de texto, en busca de identificar si lo que ingresó el usuario se encuentra dentro de la cadena. Este procedimiento se implementa en una función llamada **string_verification()**, en donde los parámetros de entrada son la variable **adn_string** en donde se encuentra la cadena de ADN y la variable **user_input** en donde está la combinación ingresada por el usuario; mientras que la variable de salida es una cadena de texto donde se muestra en que fracción de la cadena se encuentra la combinación del usuario o un texto que informa que dicha combinación no se encuentra en esa cadena.

Finalmente, la quinta etapa consiste en mostrar los datos y resultados en pantalla. En esta etapa se combinan los anteriores módulos y se imprime en la consola en la secuencia definida.

Se implementa en la función llamada **main()**, nombre asignado por convención para definir la función principal de un código, en donde se ejecutan las demás funciones.

Resultados

Las pruebas del programa se realizaron sobre el sistema operativo Ubuntu con la versión de Python

3.7.0. Sin embargo, el programa se puede ejecutar en Windows desde que se tenga instalada la versión de Python indicada. En la Figura 5 se muestra el resultado en consola del programa.

Como se describió anteriormente, el programa se ejecuta de forma secuencial de acuerdo al diagrama de flujo presentado en la Figura 1. La primera parte (1), muestra la cadena de ADN organizada para que sea más fácil identificar las secuencias. La segunda parte (2), brinda la base teórica para la actividad y las condiciones de operación del juego, que en este caso es ingresar una posible combinación de tres bases nitrogenadas. La tercera parte (3), muestra resultados arrojados por el

```
dario@XPS-12-9023:~/Documentos/Python/adn$ python3 adn_string.py

La cadena de ADN es:
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC
GAGTTTACTT CACCAGATAC GCTGTCTCTG GCACGTGGAT GGTTTAGAGG AATCACATCC

Las cadenas de ADN se componen de 4 bases nitrogenadas:
+++++ Adenina (A), Timina (T), Citosina (C) y Guanina (G) +++++
A su vez, las combinaciones de a tres de estos componentes conforman 20 aminoácidos.
La actividad consiste en adivinar las combinaciones de BASES NITROGENADAS:
ATENCIÓN: Recuerde que son combinaciones de 3 bases nitrogenadas (A-T-C-G)

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): ATG
El aminoácido encontrado es: Metionina
La sección de la cadena de ADN donde se encuentra es: ATGGTTTAGA

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): AAT
El aminoácido encontrado es: Asparagina
La sección de la cadena de ADN donde se encuentra es: AATCACATCC

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): CGT
El aminoácido encontrado es: Arginina
La sección de la cadena de ADN donde se encuentra es: CGTGGATGGT

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): GTG
El aminoácido encontrado es: Valina
La sección de la cadena de ADN donde se encuentra es: GTGGATGGTT

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): TTTT
El aminoácido encontrado es: ¡No se encontró esa combinación!

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G): DTG
¡Alguna de las letras ingresadas no hacen referencia a una base hidrogenada!
El aminoácido encontrado es: ¡No se encontró esa combinación!

Ingrese la posible combinación de bases hidrogenadas (A-T-C-G):
```

Figura 5. Resultado del programa ejecutado con diferentes entradas de usuario.
Fuente: Elaboración propia.

programa cuando el usuario ingresa una combinación correcta. En este caso se imprime en pantalla el nombre del aminoácido encontrado y luego la sección de la cadena de ADN donde se encuentra ubicado, resaltando en negrita la cadena que ingresó el usuario.

Finalmente, en la cuarta parte (4), se muestran los resultados cuando el usuario ingresa combinaciones erradas. El primer caso se da cuando se ingresa más de 3 caracteres, en donde el programa indica que dicha combinación no se encontró ya que solamente se pueden ingresar 3 bases nitrogenadas. El otro caso se presenta cuando se ingresa un carácter que no hace referencia ninguna base nitrogenada. Al darse esta situación, el programa indica el error y recuerda al usuario que una de las letras ingresadas no hace referencia a una base nitrogenada.

Conclusiones

El artículo presenta un programa diseñado de forma modular para la identificación de codones en cadenas de ADN con la interacción del usuario. Cada uno de los módulos planteados se encarga de una función específica, dando una capacidad de re-utilización del código para otro tipo de implementaciones. El programa se desarrolló utilizando Python, un lenguaje de programación con una sintaxis simple, con la capacidad de ser utilizado en cualquier plataforma de forma gratuita y contando con una gran cantidad de documentación con la filosofía de software libre.

El programa se diseñó con la intención de que se utilice en ambientes de formación dedicados a la enseñanza de la bioquímica y la biología molecular, en donde una herramienta de este tipo permite acercar la ciencia y la tecnología a estudiantes de básica secundaria. De igual forma, como trabajos futuros, se plantea el desarrollo de una interfaz gráfica más atractiva e intuitiva para los usuarios y la complementación de programa existente con mayor información sobre la temática.

Agradecimientos

Agradecemos a la Tecnoacademia por el uso de espacios y equipos. A los aprendices de los semilleros BIBA y MAPRA por su participación activa en el estudio.

Referencias

- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227, 561. Retrieved from <http://dx.doi.org/10.1038/227561a0>
- Morcillo-Ortega, J. & López-García, M. (2007). Las TIC en la enseñanza de la Biología en la educación secundaria: los laboratorios virtuales. *Revista electrónica de enseñanza de las ciencias*, 6(3), 562-576. Recuperado de <https://www.researchgate.net/publication/28184291>
- Moreno, A. L. (2015). "Retos en la enseñanza de la biología molecular y la bioquímica en las carreras del área de la salud". *Boletín Virtual*, 4-9(2266-1536), 26-39.
- Python Software Foundation [US]. (2017). Welcome to Python.org. Retrieved from <https://www.python.org/>
- Tecnoacademias, L. (2017, noviembre 7). Tecnoacademia. Servicio Nacional de Aprendizaje. *Servicio Nacional de Aprendizaje*, pp. 3-5. Recuperado de <http://www.sena.edu.co/es-co/formacion/Paginas/tecnoacademia.aspx>
- Tobis, M., Papert, S., Siegel, A. & Nevins, D. (2015). Python in Education, 2(2), 9-16. Retrieved from <http://www.oreilly.com/programming/free/files/python-in-education.pdf>

Si desea consultar derechos de autor, declaración de privacidad, la lista de comprobación para la preparación de envíos y el aviso de derechos de autor/a, favor diríjase a la siguiente página web: http://revistas.sena.edu.co/index.php/inf_tec/about/submissions#authorGuidelines