

Ejercicios en R: Introducción a la biología molecular

Introducción

Una molécula de ADN consta de dos hebras que se enrollan entorno a un eje imaginario común dando lugar a la famosa doble hélice, un modelo estructural propuesto en 1953 por James Watson y Francis Crick. Cada cadena se denomina también hebra (strand) y está formada por nucleótidos (nt) unidos entre sí mediante enlaces de tipo fosfodiéster que se forman entre el grupo fosfato en posición 5' de un nt y el grupo hidroxilo en posición 3' de otro. El orden en que se disponen cada una de las 4 bases posibles (A, C, T y G) determina su secuencia.

En los últimos años ha habido un avance exponencial en el campo de la secuenciación del ADN. Se han desarrollado métodos de alto rendimiento para acelerar el proyecto de secuenciación. Aun así, la secuenciación básica del ADN es una tecnología que lee cada nucleótido paso a paso por métodos químicos para descifrar el orden de las letras A, T, G y C que se encuentran en la secuencia de un gen en específico. Sin embargo, para este problema, le hemos dado un breve resultado de la secuencia de datos. En los resultados de secuenciación de laboratorio típicos, parte de la secuencia de ADN podría verse así: "CTGATT-GATGGTC-NAT", donde aparte de ATGC podemos ver algunos guiones (omitidos) y N (desconocido) nucleótido.

Instrucciones

1. Escribe una función que genere una secuencia aleatoria de DNA de tamaño "n".
2. Codifica una función que calcula el tamaño de una secuencia de DNA.
3. Crea una función que recibe una secuencia de DNA e imprime el porcentaje de cada base en la secuencia.
4. Para representar una molécula de ADN de doble hebra basta con escribir la secuencia de una de sus hebras. Consideremos, por ejemplo, la secuencia TGCGATAC. Como no se indica lo contrario, se asume que la secuencia está escrita en sentido 5'→3' y, por lo tanto, se trata de la hebra directa (forward strand):

Hebra directa: 5'-TGCGATAC-3'

Análogamente, si decido escribir esta misma secuencia empezando por el extremo 3' se obtiene la hebra inversa (reverse strand):

Hebra inversa: 3'-CATAGCGT-5'

Escribe una función que recibe una hebra directa y regresa la hebra inversa.

5. Normalmente se representa la molécula escribiendo primero la hebra directa y debajo la hebra complementaria (complementary strand). La hebra complementaria se escribe en sentido 3'→5' para que las bases de ambas hebras queden emparejadas:

Hebra directa: 5'-TGCGATAC-3'

Hebra complementaria: 3'-ACGCTATG-5'.

Escribe una función que recibe una hebra directa y obtiene la hebra complementaria.

6. Programa una función que transcribe DNA a RNA.
7. Crea una función que traduce una secuencia de RNA a una secuencia de aminoácidos (proteína). Usa la siguiente tabla del código genético:

		Second Letter					
		U	C	A	G		
1st letter	U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G	3rd letter
	C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G	
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G	
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G	

*Recuerda que las secuencias pueden contener caracteres especiales donde aparte de ATGC, en ADN, y AUGC, en ARN, podemos tener algunos guiones (omitidos) y N (desconocido) nucleótido.