

## FUNDAMENTOS EN BIOLOGIA COMPUTACIONAL CB0834

Prof. Javier Correa Álvarez y Alejandro Rodriguez

### PARCIAL 1

1. Los siguientes son secuencias de anticodones que participaron en la síntesis de una proteína: **UAC UAA CGC CCC AGA UCA**. Derive:

- La cadena de ARNm:
- La cadena de ADN que sirvió de molde:
- Teniendo en cuenta la siguiente tabla de codones, traduzca la secuencia de aminoácidos correspondiente:

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

2. ¿Cuál es la relación entre un alineamiento de secuencias y evolución?
3. Según el modelo estadístico de Karlin- Altschul (1990), el cual hace referencia a la estimación del puntaje más alto (HSP) en un alineamiento de dos secuencias. ¿Cómo es argumentado este modelo y qué aplicabilidad tiene en la búsqueda de homología de secuencias hoy en día?. Explique brevemente el modelo.
4. Consulte qué tipo de algoritmos de alineamientos utilizan los programas de FASTA, CAP3, Bowtie. Compare los tres en términos de velocidad, capacidad de análisis y sus principales limitaciones.
5. Represente como sería una matriz de alineamiento global para las siguientes dos proteínas. Asuma los valores estándar de puntuación que hay reportado en la literatura para amino ácidos.

A: SPAALKALAEAAAGS

B: SGAALKALAEAAAPS

6. Realizar los siguientes pasos.
  - A. Ingrese a este website: <http://www.ebi.ac.uk/Tools/sss/fasta/>
  - B. Ingrese una secuencia fasta de proteínas de su interés.
  - C. Extienda las opciones y explore los diferentes programas para hacer la búsqueda en las bases de datos (FASTA, FASTAX, FASTAY, SSERCH).
  - D. ¿Qué diferencias existen entre ellos? Explique
7. ¿Qué modificaciones haría al algoritmo de alineamiento global, para alinear dos secuencias A y B, si se quisiera exigir que B esté totalmente contenida en A. Ejemplo de aplicación: Alineamientos de reads contra un genoma referencia. Pista: un cambio sutil en la inicialización de la matriz y en la vuelta para reconstrucción del alineamiento.
8. Scripting
  - a. Escribir un script en Perl donde utilizando el módulo de BioPerl, se pueda bajar una secuencia FASTA de genbank: EU856392
  - b. Abra e imprima el archivo fasta
  - c. Traduzca la secuencia a proteína
  - d. Realice un blastp contra la base de datos de proteínas
  - e. Imprima el resultado del Blast
9. Suponga que usted conoce que genoma de una organismo *Hypoteticus organismos* tiene pocas repeticiones. Usted secuenció 100 millones de reads por medio de la metodología de shotgun, con 500 pares de bases en promedio cada uno. ¿Qué usted haría intentar estimar el tamaño del genoma de este organismo? Pista: La teoría de Lander-Waterman es capaz de modelar el desarrollo de un proyecto de secuenciamiento y prever varios parámetros a partir de L (tamaño del read), T (overlap mínimo), G (Tamaño del genoma) y N (número de reads).

10. Considere un organismo con tamaño de genoma estimado de 100Mpb, probablemente contiene muchos genes duplicados y sin genoma referencia conocido. Fueron sugeridas tres estrategias para el secuenciamiento del DNA para este organismo. Discútalas con relación al costo – beneficio, recursos computacionales, información con potencial de ser obtenidas y calidad probable del montaje.
- a. 300 Mpb de datos por 454 (single-ends de 400 pb) + 3 Gbp de datos solexa – illumina (paired-end de 75 pb y insertos de 500 pb)
  - b. 3 Gbp de datos solexa – illumina (paired-end de 75 pb y insertos de 250 pb) y 3 Gbp de datos solexa – illumina (paired-end de 75 pb y insertos de 5 Kpb).
  - c. 10 Gbp de datos solexa – illumina (paired-end de 75 pb y insertos de 250 pb)