**Bioinformática 2014 - Lic. en Microbiología - UAQ**

**Examen Parcial III**

**Nombre:** Citlalli Limpens

**Correo electrónico:** citlimpens@yahoo.com.mx

1. ¿Qué técnicas de laboratorio se pueden emplear para determinar los diferentes niveles de estructuras de proteínas? Defina brevemente cada una de ellas.

* Estructura primaria
  + Se puede secuenciar la cadena de aminoácidos, mediante espectrometría de masas. Esto se realiza cortando la secuencia en pedazos pequeños, y separarlos de acuerdo a su masa. El espectrómetro de masas dará como resultado el peso de cada aminoácido obtenido. Dicho peso se puede comparer en la base de datos de los aminoácidos existentes, y, al ser diferente peso para cada uno, se podrá identificar cuáles se encuentran en la cadena problema.
  + También se puede deducir la secuencia teniendo la información genetica de la que provino la proteina problema. Esto se hace comparando cada 3 nucleótidos con el uso de codones del organismo (que, generalmente, es universal). Cada codon da como resultado un aminoácido, por lo que se puede saber cómo sera la proteína que provenga de esa secuencia.
* Estructura secundaria
  + Esta estructura está definida por las características fisicoquímicos de cada aminoácido que compone a la cadena, por lo que teninedo la estructura primaria de la proteína, la secundaria se puede calcular mediante la estimación de la energía libre de cada uno de los componentes, o, en caso de que se tengan las estructuras ya realizadas en proteinnas similares de organismos similares, se pueden comparer en una base de datos. Es difícil de definir en el laboratorio, generalmente se realiza mediante programas bioinformáticos (ver pregunta 2)
* Estructura terciaria
  + Se puede cristalizar la proteína de interés, y bombardearla con rayos X. Dichos rayos se verán difractados, mostrando un patrón específico en un detector. Dicho patrón mostrará la estructura de la proteína.
  + También se puede definir por NMR. Este proceso consiste en mantener la proteína de interés en una solución acuosa, y someterla a un campo magnético con ondas de radio. La proteína absorberá cierta energía gracias a sus puentes de hidrógeno. Dicha absorción es registrada, y gracias a esos resultados se puede definir la estructura terciaria.

1. ¿Qué métodos computacionales se pueden emplear para predecir los diferentes niveles de estructuras de proteínas? Defina brevemente cada una de ellas.

* Una vez obtenidos los resultados de la espectrometría de masas, los resultados para cada aminoácido se cotejan con una base de datos, para buscar la identidad de cada uno de ellos.
* Para la identificación de la estructura secundaria, lo más sencillo es utiilizar un servicio de internet que utiliza redes neurales (modelos matemáticos), modelos ocultos de Markov (modelo estadístico que se basa en el reconocimiento de patrones), o máquinas vectoriales. Los tres modelos se basan en la probabilidad y algunas de éstas páginas son:
  + PSIPRED: predice la estructura utilizando la base de datos PsiBLAST, y usa redes neurales como modelo.
  + SAM: también utiliza la base de datos PsiBLAST, pero usa modelos ocultos de Markov como modelo matemático.
  + PredictProtein: utiliza una mezcla de varios modelos
  + SABLE: usa redes neurales
  + Jpred también hace uso de redes neurales
* Todos estos programas son accesibles para la comunidad científica, son fáciles de utilizar, y, en general, son gratuitos.
* Estructura terciaria
  + El método más común de predicción, es pormedio de homología, ya que simplemente se basa en la búsqueda de secuencias similares en una base de datos. Es el más sencillo, ya que el otro método informático, llamado *ab initio*, es muy complejo y no se ha logrado definer un algoritmo que funcione para todas posibilidades y clases de plegamiento de la proteína. Los programas que utiizan esta técnica, toman únicamente fragmentos pequeños de la proteína problema. Algunos programas accesibles por medio de internet son:
    - HHpred, que utiliza bases de datos para la predicción por homología
    - I-TASSER, que utiliza varios algoritmos diferentes, así como comparaciones por homología
    - RaptorX, que utiliza varios protocolos de predicción
    - QUARK que usa fragmentos para realizar la técnica de *ab initio*
    - Folding@home, que utiliza la fuerza computadoras personales en descanso para realizar el doblamiento
    - Foldit, que usa a personas reales, volviendo el proceso de predicción en un juego.

1. ¿Cuáles son las principales características que distinguen a los equipos de secuenciación de primera, segunda y tercera generación? Comparelas y dé ejemplos de equipos.

|  |  |  |
| --- | --- | --- |
| **Primera generación** | **Segunda generación** | **Tercera generación** |
| Utilizaban PCR como base para la secuenciación | Generan miles de copias de la secuencia por medio de métodos de amplificación (tipo PCR), y las inmoviliza en una superficie sólida | Secuencia una molécula única de DNA |
| Alto costo de secuenciación por base, procesos muy tardados. | Bajó el precio de secuenciación por base, pero los equipos son muy caros. | Muy alta velocidad, pero han tenido menor nivel de confianza |
| Sólo es capaz de secuencias secuencias cortas | Tiene problemas como repeticiones en tandem de algunos sitios o puede tener lecturas cortas. Las muestras son difíciles de procesar | Poca confianza, no han sido útiles para secuenciación *de novo* |
| EJEMPLOS: secuenciación de Sanger | EJEMPLOS: SOLiD, HiSeq2000 | EJEMPLOS: IonTorrent, PacBio |

1. ¿Cómo puede determinar si los datos producidos por un secuenciador son viables para analizarlos?

* Se pueden utilizar herramientas como FastQC, que analizan la calidad de la secuencia obtenida por el secuenciador base por base. Gracias a esto, se pueden detector errores en la secuenciación. El programa da como resultado una interfaz gráfica de fácil interpretación que ayuda a definir qué partes de la secuencia son utilizables y cuáles contienen una mayor cantidad de errores.

1. Dado un genoma bacteriano de tamaño estimado de 5 Mb, escriba su estrategía para secuenciarlo, indicando tecnología a usar, longitud y candidad de fragmentos, protócolos de laboratorio y de cómputo a usar, etc.

* Si se trata de una secuenciación *de novo*, utilizaría un secuenciador HiSeq2000, ya que tiene una precision del 98%, y a pesar de que tardaría entre 3 y 10 días en tener mi secuencia, el precio por base es muy económico para lo que estoy buscando. Si, en cambio, no cuento con el equipo, podría usar IonTorrent, ya que es una tecnología más barata, y tardaría unas dos horas de obtener mi genoma. IonTorrent sí funciona para secuenciar bacterias, pero no es muy recomendable si dicha secuenciación es *de novo*.
* En caso de usar IonTorrent, los fragmentos obtenidos serían de 200 pares de bases, obteniendo como resultado aproximadamente 25,000 de ellos.
* Para realizar la secuenciación, simplemente tendrila que extraer el genoma y purificarlo, posteriormente entraría en el chip del IonTorrent para ser fragmentado y amplificado en pequeñas esferas. La máquina leerá los hidrógenos que soltará cada base agregada, por lo que dará como resultado la secuencia de cada una de mis esferas.
* Para recibir mis datos, usaría IonTorrent suite, y, al ser un genoma pequeño, podría guardarlo en cualquier computadora.
* Una vez teniendo mi secuencia, analizaría su calidad utilizando FastQC para verificar que mi secuencia sea la adecuada. No tomaría en cuenta los fragmentos de mala calidad, realizando un trimming utilizando Perl o Python.
* Para el ensamblaje, utilizaría el programa Velvet, ya que es ideal para el ensamblaje de genomas pequeños, como de virus y bacterias. Una vez obtenido mi resultado, lo compararía con otros genomas de bacterias similares utilizando blastx; para verificar que mi ensamblaje no sea erróneo.

1. ¿Cómo puede evaluar si un ensamble de un genoma es correcto?

* Se puede comparer el genoma con el de una especie que ya se encuentre secuenciada
* Se pueden alinear por medio de blastx con genes ya secuenciados y conocidos
* Se pueden predecir genes con la secuencia obtenida y comparer dichos genes predichos contra una base de datos.

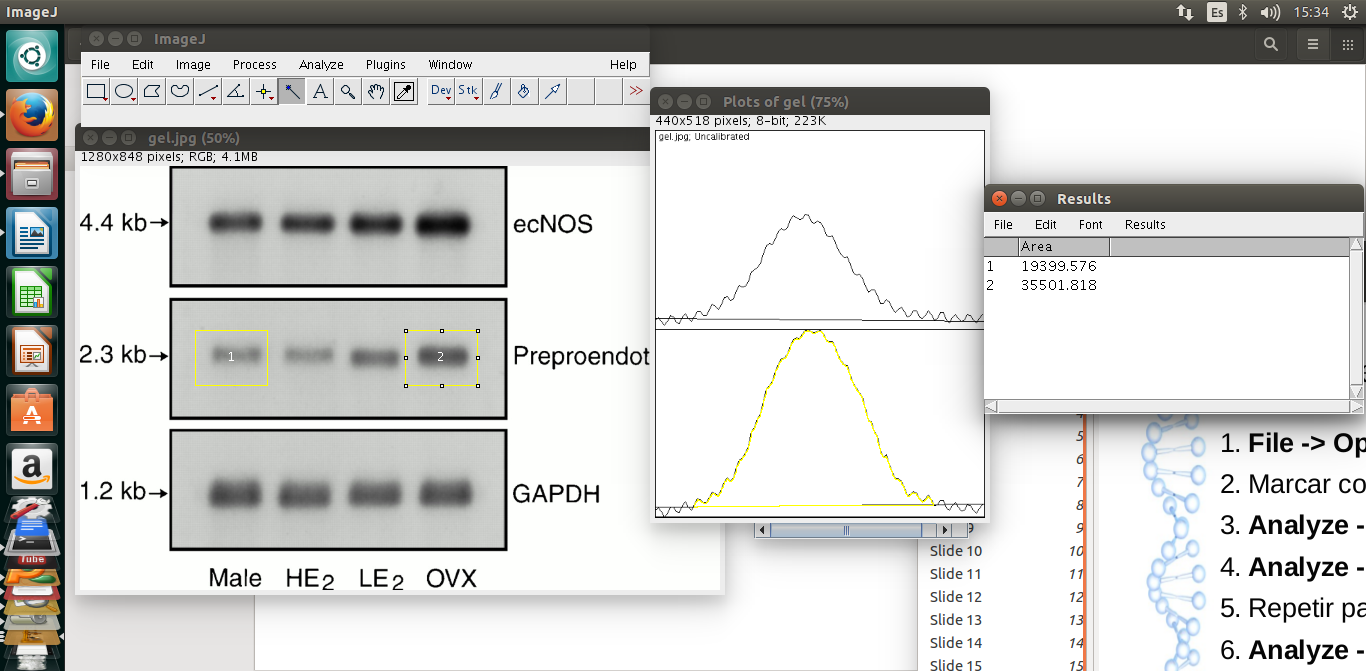
1. ¿Qué describe un pixel en una imágen digital?

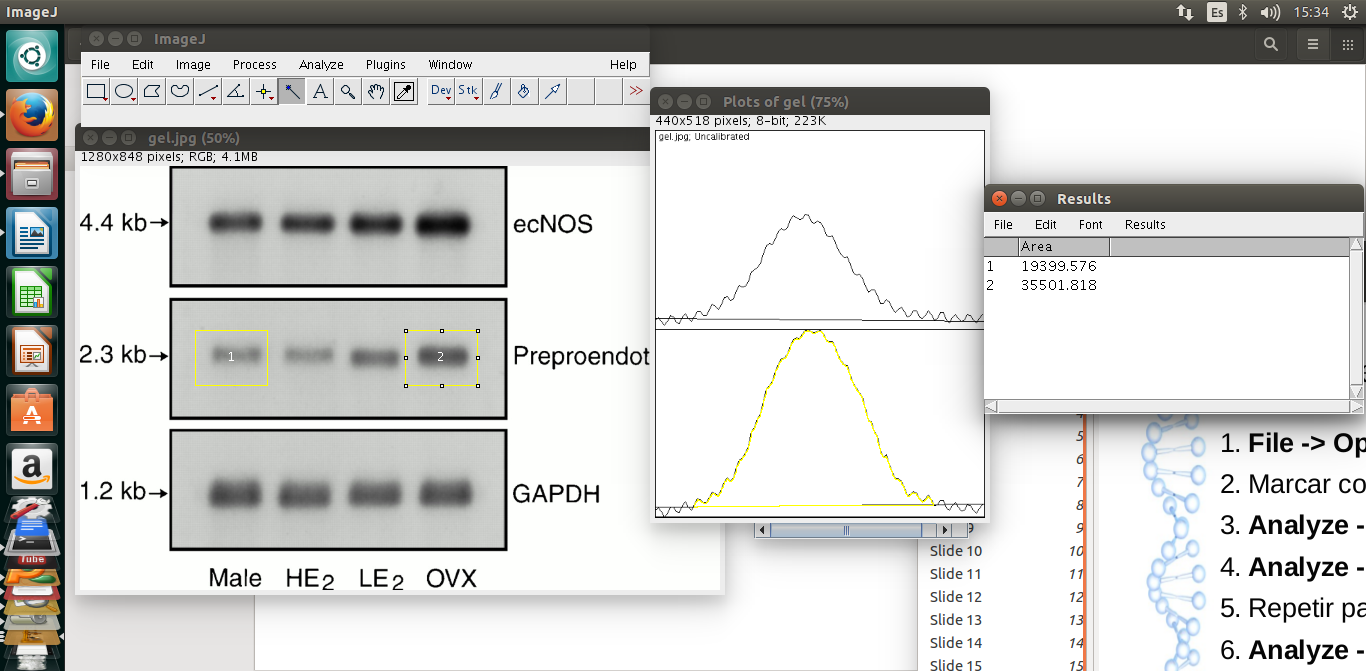
* Describe cada punto de la imagen con la intensidad determinada de su color. Este color puede ser definido como simplemente blanco y negro; como una escala de grises o con los colores básicos RGB (rojo, verde, azul) o CMYK (versiones claras del azul, rojo y Amarillo; complementadas con negro)

1. ¿Qué operaciones podemos hacer en una imágen digital?

* Podemos modificar colores, logrando que se marquen únicamente las secciones de determinada intensidad de color.
* Podemos poner una escala real a la imagen para determinar tamaños reales en una fotografía de microscopía
* Podemos contar los elementos que se requieran en una imagen
* Podemos analizar geles con diferentes operaciones, como por ejemplo, un dot blot que nos da el valor numerico que tiene cada banda de un gel gracias a su intensidad de color.

1. Análice la imagen “gel.jpg” con ImageJ e indique númericamente el cambio de expresión de la preproendotelina-1 en “OVX” comparado con “Male”.

*  Los resultados indican el cambio de expresión. La diferencia entre ambos es de 16,102.242, es decir, hay una expresión 54.64% mayor de preproendotelina-1 en OVX que en Male.



10. ¿Qué recomendaciones le parecen más importantes en el artículo *So you want to be a computational biologist?*, explique sus razones propiamente.

* “La clave para ser un buen trabajo de biología computacional es la selección y el uso del software apropiado”. Esta es la primer recomendación, y la que a mí me parece más importante. Esto, porque el software que se va a usar es la base del análisis bioinformático. Para poder realizar un buen trabajo se debe conocer el programa y saber qué esperar de él y qué no, así de qué se tiene que hacer para obtenerlo. Cada programa es diferente y aunque varios funcionen para el mismo propósito, elegir el ideal para tu investigación, estilo y acomodo es importante para realizar un buen trabajo y entender el trabajo que la computadora realiza por ti.
* Ser creativo y aventurero. Esto es muy importante, no solo en la biología computacional, sino en todos los aspectos de la ciencia y la vida diaria. Para poder lograr algo, lo primero que se debe pensar es que cualquier problema se puede resolver, nada más se debe inovar lo suficiente como para encontrar la forma de solucionarlo.
* Por ultimo, me pareció muy importante la recomendación de “Ser un detective”. Esto porque la computadora y el programa utilizado únicamente sabe entregar la información de maneras limitadas y es trabajo del investigador interpreter esa información de la forma adecuada. Esto puede llegar a ser muy difícil, por lo que hay que tener un ojo bien entrenado y saber identificar cualquier anormalidad en el resultado. En este punto, recalco la importancia del primero que mencioné, ya que la interpretación adecuada de los resultados que arroja un programa, entra en su uso adecuado.