

# TRABAJO

# PRÁCTICO N°3

NOMBRES:

Chaile Lucia.

Mansilla Milagros.

Menel Angelo.

Vargas Micaela.

1) **Biología computacional** : el estudio de la biología utilizando técnicas computacionales. El objeto <sup>2.1</sup> es aprender nueva biología, conocimiento sobre sistemas vivos. Se trata de la ciencia.

**Bioinformática** : la creación de herramientas (algoritmos, bases de datos) que resuelven problemas. El objetivo es crear herramientas útiles que funcionen con datos biológicos. Se trata de ingeniería

La **bioinformática** es una disciplina mucho más dirigida a los biólogos, de algún modo es menos interdisciplinaria, y si bien, usan ampliamente y sostenidamente la tecnología informática aplicada a contextos biológicos. Es una disciplina mucho más práctica y su propósito es el desarrollo de soluciones informáticas de diversas clases para biólogos.

La **Biología Computacional** es mucho más interdisciplinaria, o sea posee un profundo conocimiento de matemática y estadística, programación y aún de la problemática biológica. Es una disciplina más teórica y mucho más basada en el uso de Matemática y Estadísticas, y creo que entiende en grado mayor los escenarios y procesos biológicos. <sup>2.3</sup>

Si bien constituyen un campo interdisciplinar que desarrolla procedimientos computacionales, la computación, la bioinformática son diferentes; la bioinformática está enfocada a la extracción de información en la biología molecular, mientras que la biología computacional busca mejorar el trabajo de los biólogos mediante herramientas matemáticas.




2) Explique brevemente cómo se aplican las 4V del “Big Data” a las siguientes Bases de Datos Biológicas. Para ello, indague sobre el tipo de datos, estructura de datos, curado de los datos, etc.:

Las 4V del Big Data se refieren a: Volumen, Variedad, Velocidad y Veracidad


- Volumen: La cantidad de datos con la que se trabaja.
- Variedad: Los datos pueden proceder de distintas fuentes y pueden estar presentes en diferentes formatos.
- Velocidad: La rapidez necesaria para escribir, procesar y utilizar los datos en tiempo real.
- Veracidad: Se debe comprobar la validez de los datos y “limpiarlos” para eliminar sesgos y ruido. Hacemos esto para asegurarnos que los datos que obtenemos son de confianza.

Ahora aplicamos las 4V en las Bases de datos biológicas:


### **GenBank:**

- Volumen: Va a trabajar con un gran volumen de datos ya que recopila toda la información relacionada con las secuencias del ADN de la comunidad científica.
  - Variedad: Al ser GenBank parte de la INSDC, va a trabajar con miles de datos que provienen de distintas fuentes tales como DataBank de Japón, El Archivo de Nucleótido Europeo y GenBank en NCBI.
  - Velocidad: Los datos que se encuentran cargados en esta base de datos están al día con los avances que realiza la comunidad científica (apenas la información se ve publicada online es subida a la base de datos) y pueden ser accedidos en cualquier momento, aunque los datos nuevos pueden ser visualizados públicamente cada dos meses.
  - Veracidad: GenBank sube a la base de datos toda la información relacionada con las secuencias de ADN y que se encuentra publicada, por lo tanto no puede comprobar si los datos que tiene son correctos o si tienen errores.
- 

### **RNAcentral:**

- Volumen: Se encarga de procesar los datos de múltiples bases de datos que colaboran entre sí para aportar información sobre las secuencias de ARN.
  - Variedad: Al trabajar con muchas bases de datos distintas, los datos que existen son muy diferentes entre sí y varían tanto en contenido como en tipo.
  - Velocidad: Cuenta con un sistema de identificadores que separan la información según el tema al que corresponden, haciendo que la búsqueda de información sea mas rápida y sencilla.
  - Veracidad: Tiene un sistema de anotaciones (Rfam annotations) que controla todas las secuencias e información de RNAcentral e identifica a aquellos que tengan posibles problemas.
- 

### **Uniprot:**

- Volumen: Va procesar datos de 3 bases de datos (UniProtKB, UniRef y UniParc).
  - Variedad: Recibe datos de diversas fuentes externas.
  - Velocidad: Cuenta con TrEMBL (Translated EMBL Nucleotide Sequence Data Library) que fue creado debido a que las secuencias de datos eran generadas a una velocidad que supera a la velocidad que tenía la base de datos para controlar y subir los datos, entonces TrEMBL sirve como un refuerzo a la base de datos Swiss-Prot, lo cual aumenta la velocidad a la hora de cargar datos y buscarlos.
  - Veracidad: Es utilizada y avalada por múltiples organizaciones gubernamentales e institutos científicos de todo el mundo, por lo tanto, los datos que se encuentran en esta base de datos biológica deben de ser correctos y revisados constantemente, es por esto que hay 100 personas que se ven involucradas en la tarea de revisar la base de datos así como también desarrollar el software y el soporte técnico, y los datos siempre son revisados antes de ser subidos.
- 

## KEGG:

- Volumen: Va a mantener una colección de bases de datos en línea de [genomas](#), rutas enzimáticas, y químicos biológicos. La base de datos PATHWAY registra las redes de interacciones moleculares dentro de las células, y variantes de ellas específicas a organismos particulares.
- Variedad: Cuenta con un KEGG Identifier, que identifica cada objeto KEGG para saber en cual base de datos va a guardar ese objeto. Por ejemplo: si entra un archivo relacionado a una enfermedad se le da el prefijo <H> y se guarda con el código [H00004](#) en la base de datos correspondiente.
- Velocidad: KEGG mantiene una gran variedad de bases de datos para poder tener todos los datos ordenados según el tema al que se refiere, esto facilita mucho la distribución y el acceso a los datos.
- Veracidad: Esta base de datos es usada mundialmente para acceder a información relacionada con el genoma secuencial y otras tecnologías experimentales, por lo tanto los datos que se encuentran en ella deben ser actualizados y revisados para asegurar que los datos no tengan errores.

## PubMed:

- Volumen: Los datos que recopila se obtienen de NLM (National Library of Medicine), la librería biomédica más grande del mundo.
- Variedad: Procesa más de 30 millones de distintos tipos de datos, desde artículos biomédicos hasta citas, también contiene links a textos.
- Veracidad: PubMed está interconectado con MedlinePlus, el cual es un proyecto que supervisa la selección de revistas aprobadas por la Biblioteca Nacional de Medicina de EE.UU.

## 3) Estudio de ácidos nucleicos:

### **ELECTROFORESIS**

La aplicación de la electroforesis en la separación de ácidos nucleicos permite hacer tareas simples, como verificar su síntesis o integridad, y complejas, como seguir los pasos enzimáticos de modificación durante la construcción de elaboradas colecciones de ADN. Su fundamento químico reside en que los ácidos nucleicos son polímeros de carga negativa unidos por enlaces covalentes fosfodiéster. De esta manera, el ADN y el ARN se moverán en un campo electroforético hacia el polo positivo.

### **PROCESO DE HIBRIDACIÓN**

Los ácidos nucleicos difieren entre sí principalmente en la secuencia de nucleótidos. Por este motivo, se han desarrollado técnicas capaces de separar las moléculas de ácidos nucleicos basándose en dicha secuencia de nucleótidos. Se basan en la capacidad que tienen las hebras de cadena simple de ADN para formar puentes de hidrógeno con sondas complementarias.

## **SECUENCIACION**

De todos los métodos históricos de secuenciación de ácidos nucleicos diseñados por Sanger, el método por terminadores dideoxi es el más exitoso (Sanger, 1988). Este método usa nucleótidos modificados con la ausencia del extremo 3'OH, grupo esencial para la polimerización por la ADN Polimerasa (figura 2A). De esta forma, cuando se incorpora un dideoxinucleótido, queda un trozo trunco de ADN que puede ser separado electroforéticamente (Sanger *et al.*, 1977). Si cada uno de los cuatro monómeros de ADN (los nucleótidos adenina, guanina, citosina y timina) se marca con una molécula fluorescente diferente, se puede determinar rápidamente el orden de los nucleótidos. La secuenciación con el método dideoxi se provee usualmente en centrales de secuenciación a muy bajo costo. La secuenciación dideoxi acoplada a métodos de reconstrucción computacional (Shotgun) ha sido fundamental en la secuenciación de los genomas modelo de humano, animales y plantas (Galperin y Koonin, 2010).

Estudio de proteínas:

### **TURBIDIMETRÍA Y NEFELOMETRÍA**


Cuando un haz de luz choca con una partícula en suspensión parte de la luz se dispersa, parte de la luz se refleja y parte de la luz se absorbe. La dispersión de la luz depende de: la longitud de onda de la luz (?), del tamaño de la partícula y del índice de refracción de la partícula en relación con el medio que la rodea. La dispersión de la luz se puede medir por turbidimetría o por nefelometría. En ambas técnicas, para dar como resultado una concentración de proteína concreta, se compara la cantidad de luz dispersada o la tasa de aumento de dispersión con los valores de dichos parámetros en estándares proteicos conocidos.

### **INMUNOELECTROFORESIS EN COHETE**

La inmunolectroforesis en cohete es una técnica cuantitativa equivalente a la inmunodifusión radial pero, a en la que la placa se expone a un campo eléctrico. Al igual que en la inmunodifusión radial, el antisuero (Ac) se incorpora al gel de manera que el Ac no pueda migrar. En el gel se realizan unos pocillos (generalmente en el cátodo) que se rellenarán con la muestra o con diluciones patrón. Una vez depositadas éstas se activa el campo eléctrico. El Ag se desplaza en la agarosa y precipita al encontrarse con el Ac. La precipitación va produciéndose a medida que el Ag avanza hacia el ánodo de manera que al ir disminuyendo la concentración de Ag los bordes laterales se van acercando hasta unirse. Así, se produce una precipitación triangular (en estela de cohete). La concentración de Ag de la muestra es directamente proporcional al área y altura del triángulo. Comparando los resultados de la muestra con los obtenidos en las diluciones patrón obtendremos un resultado cuantitativo.


### **INMUNOFIJACIÓN**

La inmunofijación consiste en la separación electroforética de las proteínas de una muestra seguida del contacto del gel con una tira de acetato de celulosa impregnada con antisuero. La unión Ag-Ac da lugar a la formación de bandas de precipitación visulalizadas tras lavar y teñir. Es una técnica muy rápida que se


utiliza especialmente en la detección de bandas oligoclonales en líquido cefalorraquídeo. 

estudio de metabolitos:

**La espectrometría de masas acoplada a la cromatografía de gases o líquidos:**

Permite detectar e identificar metabolitos polares y no polares (una molécula es polar cuando uno de sus extremos está cargado positivamente, y el otro de manera negativa. Cuando una molécula es polar, estas cargas no existen), azúcares, líquidos volátiles, metabolitos de gran tamaño y aminoácidos. La ventaja de esta técnica es que es posible recuperar la muestra después de su análisis para estudiarla por otros métodos. La gran desventaja es que se requiere de una cantidad mayor de muestra y la sensibilidad del método es relativamente baja, lo que significa que se detectarán los metabolitos de mayor concentración. 

**La espectrometría de masas (EM):**

Es utilizada para identificar y cuantificar metabolitos después de su separación por cromatografía de gases, HPLC, cromatografía líquida-espectrometría de masas (CL-EM), o EC. El método CG-EM es la combinación más 'natural' de ellos, y fue la primera técnica en ser desarrollada. Además, se han creado bases de datos de huellas de espectros de masa que permiten identificar un metabolito de acuerdo a su patrón de fragmentación. 

# Índice de comentarios

---

- 2.1      ¿De qué ciencia se trata?
- 2.2      ¿De qué modo? Recuerden las definiciones de bioinformática que hemos estado trabajando desde la Unidad 2
- 2.3      La redacción no es clara y no interpreto a qué se refieren. Y nuevamente se están olvidando de los conceptos ya vistos.
  
- 5.1      No encontré la Figura 1 ni la Figura 2 mencionada
- 5.2      Si colocan bibliografía, luego al final debe estar en el formato completo.