

Universidad Catolica de Cordoba
Facultad de Ingeniería
Córdoba, Argentina

Trabajo Práctico N° 3

Integrantes:
Cabral, Camila
Caceres, Martin
Nanfara, Abril
Zuliani, Francina

Tema 3.1

Biología, Bioinformática y Biología Computacional

1. Existen controversias y similitudes entre las definiciones de Bioinformática y de Biología Computacional que puede ser evidenciada en la siguiente definición:

“La biología computacional es a veces definida como sinónimo de Bioinformática y a veces como una disciplina emparentada, pero distinta, de esta. El NIH define a ambas disciplinas como distintas aunque con cierto grado de solapamiento, según esta definición la bioinformática está más relacionada con el desarrollo de herramientas computacionales con el fin de analizar y procesar datos y la biología computacional con el estudio por medios computacionales de sistemas biológicos”

Lea cuidadosamente los textos de las siguientes páginas webs y discuta con sus compañeros si hay o no diferencia entre Bioinformática y Biología Computacional.

¿Deberían fusionarse los términos y por consiguiente su definición?

Webs:

- *. <http://www.bioinformaticos.com.ar/biologia-computacional-en-argentina/>
- *. <https://biology.meta.stackexchange.com/questions/168/merging-bioinformatics-and-computational-biology-tags> *.
- <https://respuestas.me/q/La-biolog-a-computacional-es-diferente-de-la-bioinform-tica-34039920924> *.
- <https://rubenyciencia.wordpress.com/tag/diferencias-entre-bioinformatica-y-biologia-computacional/> *.
- <http://www.euskonews.com/0334zbnk/gaia33402es.html>

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.

Desarrolle las 4V para:

- GeneBank: <https://www.ncbi.nlm.nih.gov/genbank/>
- RNACentral: <https://rnacentral.org/>
- Uniprot: <https://www.uniprot.org/>
- KEGG: <https://www.genome.jp/kegg/>
- PubMed: <https://www.ncbi.nlm.nih.gov/pubmed>

Tema 3.2

1. Busque y desarrolle brevemente 2 o 3 (dos o tres) metodologías y técnicas utilizadas en los laboratorios, tanto a escala simple como a gran escala, para el estudio de:

a. Ácidos Nucléicos.

b. Proteínas

c. Metabolitos

Respuestas:

Tema 3.1

- 1) Últimamente los términos “**Bioinformática**” y “**Biología Computacional**” se están usando como sinónimos entre sí. Ya bien ambas definiciones están compartiendo mucho de la misma rama, Biología, tanto la bioinformática como la biología computacional, esto **no** significa que deberían unirse bajo un mismo término. Primero en principal estas abordan su área de trabajo de formas muy distintas. La biología computacional es aquella que investiga y busca nueva vida, es decir, nueva ciencia. Por otra parte, la bioinformática crea herramientas que ayudan con los datos biológicos. Es decir, uno investiga y otro crea herramientas.
- 2) Cuando hablamos de Big Data nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (**volumen**), complejidad (**variabilidad**), velocidad de crecimiento (**velocidad**) y al sesgo de los datos (**veracidad**) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles.

*. GeneBank: <https://www.ncbi.nlm.nih.gov/genbank/>

- Volumen: cuenta con una gran cantidad de información. Contiene 6.25 Trillones de pares de bases de 1.6 billones de secuencias de nucleótidos de unas 450,000 especies.
- Variedad: sus datos se ven reflejados en imágenes, cuadros, gráficos, estadísticas, secuencias.
- Veracidad: Banco de información. Cada persona puede subir anónimamente las secuencias que desee.
- Velocidad: Sus datos están actualizados, (son datos de este año)



*. RNACentral: <https://rnacentral.org/>

- Volumen: cuenta con una gran cantidad de información principalmente con secuencias de RNA
- Variedad: sus datos se ven reflejados en imágenes, cuadros, gráficos, estadísticas estadísticas, secuencias, códigos, links.
- Veracidad: Organización con colaboración de Expert Databases. Su desarrollo está coordinado con el “Instituto Europeo de Bioinformática”.
- Velocidad: Sus datos están actualizados, (son datos de este año)

*. Uniprot: <https://www.uniprot.org/>

- Volumen: cuenta con una gran cantidad de información que abarca muchos temas.
- Variedad: sus datos se ven reflejados en archivos y grupos de secuencias, archivos de virus, bacterias, etc.

- Veracidad: Recurso integral para la secuencia de proteínas y de datos. Su información proviene de literatura científica, herramientas de análisis de secuencia y otras fuentes de datos científicas
- Velocidad: Sus datos están actualizados, (son datos de este año)

*. KEGG: <https://www.genome.jp/kegg/>

- Volumen: cuenta con una gran cantidad de información enfocada en genes y genomas.
- Variedad: sus datos se ven reflejados en imágenes, cuadros, gráficos, estadísticas estadísticas, secuencias, códigos, links.
- Veracidad: Enciclopedia de genes y genomas
- Velocidad: Sus datos están actualizados

*. PubMed: <https://www.ncbi.nlm.nih.gov/pubmed>

- Volumen: cuenta con una gran cantidad de información. Más de 30 millones de citas de literatura biomédica.
- Variedad: está compuesta por artículos ya que es un buscador de datos e información.
- Veracidad: Librería de medicina nacional. Este sitio es principalmente mantenido por el "Centro Nacional para la Información Biotecnológica" al igual que Genbank.
- Velocidad: Sus datos están actualizados hasta el 31/10/2020

Tema 3.2

Los ácidos nucleicos y las proteínas componen una red de biomacromoléculas que almacenan y transmiten la información que sustenta la vida celular. De entre los métodos disponibles para el estudio de estos mecanismos, cinco se destacan: electroforesis, secuenciación, clonación, hibridación y la reacción en cadena de la polimerasa.

- Electroforesis: es un método de separación basado en la movilidad de las biomoléculas en una fase líquida sometida a un campo eléctrico. Las moléculas que posean carga negativa migrarán hacia el polo positivo de un aparato electroforético y viceversa. La inclusión de una matriz sólida, además de la fase líquida, permitió agregar un nuevo punto de separación y versatilidad en la electroforesis. De esta manera, no solo las biomoléculas pueden ser separadas por su carga, sino también por su tamaño.
- ◆ En 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.
- ◆ En proteínas: son biomoléculas cuyos monómeros son aminoácidos polimerizados por medio de enlaces peptídicos. La carga de las proteínas es muy variable porque en ellas puede haber aminoácidos tanto negativos como positivos. Laemmli (1970) implementó un tratamiento para poder separar a las proteínas con base en su peso molecular; es útil para las aplicaciones rutinarias como el seguimiento de los niveles de acumulación de proteínas, así como para su purificación, separación, hibridación con anticuerpos y análisis de peso

molecular.

→ Secuenciación: su objetivo es conocer exactamente el orden de sus monómeros.

◆ En ácidos nucleicos: el método por terminadores dideoxi es el más exitoso. Este método usa nucleótidos modificados con la ausencia del extremo 3'OH, grupo esencial para la polimerización por la ADN Polimerasa. De esta forma, cuando se incorpora un dideoxinucleótido, queda un trozo trunco de ADN que puede ser separado electroforéticamente. Si cada uno de los cuatro monómeros de ADN 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 y ha sido fundamental en la secuenciación de genomas modelo.

◆ La secuenciación de proteínas mediante el método de Ed-man consiste en marcar el extremo N-terminal de las moléculas con el reactivo químico fenilisotiocianato (PITC). La identidad del PITC-aminoácido escindido de la proteína puede ser determinada mediante la observación de su presencia directamente en cromatografía de capa fina o de su ausencia por exclusión en un analizador de aminoácidos. Pronto el método de Edman fue automatizado y es actualmente el estándar de secuenciación de proteínas. En años recientes, este método químico ha coexistido con el método analítico maldi (matrix-associated laser desorption/ionization) acoplado a ms, que identifica los péptidos de acuerdo a su huella de fragmentación iónica.

→ Clonación: Las ERs (enzimas de restricción) cortan de forma precisa secciones de ADN y hacen compatibles los extremos de los productos escindidos con otras moléculas de ADN, como adaptadores, plásmidos, promotores y regiones codificantes, a los que se unen en una reacción in vitro catalizada por la enzima ligasa. Estas nuevas construcciones de ADN formadas por moléculas de orígenes genéticos distintos, se denominan quimeras. Las quimeras a su vez pueden ser escindidas para ligarse a otro fragmento de ADN, en un proceso llamado subclonación. Mediante programas de cómputo se puede simular la mejor estrategia para obtener quimeras en el laboratorio. Las quimeras son útiles para comprobar hipótesis sobre la función génica, así como para la producción en masa de proteínas recombinantes. En combinación con las hibridaciones Southern Blot, las ERs fueron esenciales para descubrir los primeros genes responsables de enfermedades genéticas humanas y para seleccionar características deseables en vegetales y animales. Debido a lo laborioso que es obtener huellas genéticas con ERs, se favorece la obtención de huellas por PCR o por secuenciación masiva del genoma completo con métodos post-Sanger.

→ Detección de biomoléculas por hibridación: Cuando los ácidos nucleicos o las proteínas en solución interactúan con superficies sólidas pueden llegar a adsorberse en ellas. Las superficies más utilizadas son las membranas de nitrocelulosa y de nylon.

Esta propiedad de interacción con fases sólidas, en conjunto con la capacidad de los ácidos nucleicos de formar puentes de hidrógeno con otras moléculas de ADN complementarias a su secuencia (hibridación), se han utilizado para detectar y cuantificar su abundancia en extractos celulares de las más variadas condiciones. La técnica original se conoce, por su nombre en inglés, como Southern Blot. Los Southern Blot consisten en adsorber una muestra de ADN (digerida por enzimas de restricción y separada previamente de manera electroforética) en una membrana, en presencia de un agente que rompa los puentes de hidrógeno intermoleculares (ej. NaOH), esto es, que

desnaturalice la estructura del ADN en la membrana. Las moléculas contenidas en la muestra adsorbida en la membrana, de las cuales se busca obtener información específica, se llaman "blanco". Posteriormente, otra molécula de ADN en solución, también desnaturalizada (por calor) o de cadena simple, se hace interactuar con el ADN en la membrana a una temperatura que permita un reconocimiento de bases lo más específico posible. Esta molécula que servirá para extraer información sobre el blanco, se denomina "sonda". Si la sonda ha sido marcada con radioactividad o grupos antigénicos (capaces de interactuar con anticuerpos), producirá manchas oscuras en la zona de hibridación si se le expone a una placa fotográfica (Southern, 1975). El patrón de manchas producidas indicará la presencia de su blanco, y dará información sobre su tamaño, frecuencia y abundancia en la muestra.

Si en este arreglo en lugar de colocar ADN en la membrana se coloca ARN desnaturalizado con formaldehído, se le denomina Northern Blot, esto debido a un juego de palabras con el Southern Blot. Si en la membrana se colocan proteínas desnaturalizadas con SDS y en la solución anticuerpos (proteínas del sistema inmune capaces de interactuar con otras proteínas), el método se conoce como Western Blot (continuando con el juego de los puntos cardinales).

El diseño de experimentos con estos "blots" involucra la comparación de las propiedades de las manchas que se obtengan (tamaño molecular, intensidad, número) de una muestra control o estándar contra las muestras donde se espera un comportamiento diferente, por ejemplo, las señales detectadas en ARN extraído de células sanas contra las detectadas en ARN extraído de células de un tumor cancerígeno. Ha permitido los más complejos análisis como la detección de los defectos moleculares donde reside la causa de una enfermedad humana, por ejemplo.

→ **Reacción en cadena de la polimerasa (PCR):** La PCR es un método que permite la amplificación o copiado masivo in vitro de un fragmento específico de ADN. La longitud del fragmento copiado está indicada por dos secuencias pequeñas de ADN (llamadas oligonucleótidos) que son adicionadas a la reacción y que flanquean al fragmento que se desea amplificar. Estas secuencias pequeñas indican además el sitio desde el cual debe iniciarse la reacción de amplificación. La reacción es llevada a cabo por la ADN polimerasa. Las variantes de polimerasas termoestables que existen en el mercado se pueden contar por decenas. Se pueden seleccionar por velocidad de síntesis, fidelidad del copiado, capacidad de sintetizar moléculas largas, tolerancia a inhibidores de reacción y por ornamentaciones adicionales que hacen a los productos.

Cada ciclo de PCR tiene tres pasos térmicos:

Desnaturalización: se calienta la reacción (90–94°C) para romper los puentes de hidrógeno y separar la doble cadena de ADN.

Alineamiento: se disminuye la temperatura (50–65°C) para alinear los oligonucleótidos con su molécula blanco.

Extensión: se eleva la temperatura (68–72°C) para que la polimerasa termoestable busque su sustrato de doble cadena (ADN molde + oligonucleótido) y sintetice la nueva hebra de ADN utilizando a los nucleótidos libres en solución.

Los metabolitos son moléculas pequeñas y son el resultado de la actividad metabólica de las células. Las principales técnicas utilizadas para su análisis en laboratorio son: Electroforesis capilar acoplada a espectrometría de masas (EC-EM), Cromatografía líquida acoplada a espectrometría de masas (CL-EM), Resonancia magnética nuclear (RMN), Cromatografía de gases acoplada a espectrometría de masas (CG-EM) e Infusión directa acoplada a

espectrometría de masas (ID-EM).

- CG-EM: ofrece una poderosa técnica analítica que combina la cromatografía (de líquidos o de gases) como técnica de separación, y la espectrometría de masas como técnica de detección, identificación y cuantificación para compuestos orgánicos/organometálicos. ✓

La Espectrometría de Masas es una potente técnica instrumental de análisis, de alta sensibilidad, basada en la ionización de las moléculas y en la separación y registro de los iones producidos según su relación masa/carga (m/z) en un sistema a vacío. Los iones que llegan al detector producen una señal eléctrica que es ampliada, procesada y registrada en un ordenador, dando lugar al correspondiente espectro de masas que no es más que una representación gráfica de la abundancia de los iones detectados en función de su relación m/z . ✓

- RMN: Esta técnica se basa en la separación de iones en función de su masa y carga. Estos se forman por la ionización de una molécula orgánica y en dependencia del tipo de molécula será el tipo de ionización utilizado. En este caso, en contraste con la resonancia magnética nuclear (RMN), la muestra si se destruye y no puede volver a utilizarse ✓

Referencias:

“Los métodos experimentales que permiten el estudio de las macromoléculas de la vida: historia, fundamentos y perspectivas.” Julián M. Peña–Castro,* Oscar Gregorio–Ramírez y Blanca E. Barrera–Figueroa.

“El futuro de la metabolómica en el diagnóstico clínico”. Ana Karenina Rocha Viggiano, Mariana Salgado Bustamante y Yamilé López Hernández.

“Cromatografía de Gases/Líquidos acoplado a espectrometría de masas de Alta Resolución (MS-AR)” Universidad de Burgos

Índice de comentarios

- 4.1 No es completamente correcto. Recordar las definiciones de Bioinformática que hemos estado trabajando desde la Unidad 2
- 4.2 ¿?
- 4.3 ¿?
- 4.4 ¿?
- 5.1 ¿?