

SECUENCIACIÓN AUTOMÁTICA DEL GENOMA HUMANO

1. OBJETIVO DEL EXPERIMENTO

En esta práctica, los estudiantes leerán secuencias de ADN obtenidos a partir de técnicas de secuenciación automáticas de ADN. Los datos se analizaron utilizando las bases de datos disponibles al público para identificar los genes y productos genéticos. El impacto de la genómica se discutirá en el contexto de la sociedad actual.

2. COMPONENTES para 10 grupos de estudiantes

Esta práctica contiene un total de doce fragmentos de secuencias de ADN obtenidas automáticamente. Los estudiantes pueden usar cualquier base de datos de secuencias del Genoma Humano para llevar a cabo la práctica. En este protocolo se ha utilizado la base de datos ofrecida por el **Centro Nacional de Información Biotecnológica, NCBI (National Center for Biotechnology Information)**.

COMPONENTES	
Fragmentos de secuencias de ADN	12

2.1 Material requerido y no suministrado

- Ordenador con acceso a internet.

3. INTRODUCCIÓN

El genoma humano haploide comprende aproximadamente tres mil millones de pares de bases de ADN que se organizan en 23 cromosomas. El orden de estos nucleótidos crea genes, que son las unidades de la información genética que contienen las instrucciones para construir y mantener un organismo. La secuenciación de ADN es el proceso de determinar el orden preciso de estos nucleótidos. En 2001, se publicaron las secuencias del primer borrador del Genoma Humano, una por una coalición internacional conocida como el **Proyecto Genoma Humano** y otra por una empresa llamada **Celera**. Aunque estos primeros estudios necesitaron muchos años y una gran cantidad de personas para realizarse, los avances en la tecnología de secuenciación han permitido completar totalmente los datos del genoma de forma rápida y sencilla. En 2015 se puede conseguir una secuencia completa en un día por alrededor de mil dólares. Dado que gran parte de esta secuencia de datos está disponible gratuitamente en línea, el nuevo reto en la genética implica encontrar formas creativas y eficientes de analizar y gestionar las grandes cantidades de datos que se

generan. Esto ha dado lugar un floreciente campo de la bioinformática, una disciplina que combina la informática, la biología y la tecnología de la información.

La información del genoma humano puede ayudarnos a entender mejor nuestra fisiología y las bases biológicas de las enfermedades hereditarias. Por ejemplo, la secuenciación del ADN de los pacientes individuales, conocidos como medicina personalizada, está cambiando el papel de la genética en la medicina. La medicina personalizada utiliza el perfil genético de un individuo para guiar las decisiones relativas a la prevención, diagnóstico y tratamiento de la enfermedad. La **Tabla 1** destaca cinco áreas prometedoras de la medicina personalizada. Estas pruebas y servicios no solo ofrecen increíbles posibilidades terapéuticas, sino que también generan información para el paciente que no tiene precedentes en su detalle. Esto plantea nuevas cuestiones legales y logísticas sobre la protección de la confidencialidad médico-paciente. Por ejemplo, ¿tienen los padres derecho a solicitar exámenes genéticos para sus hijos menores de edad?, o ¿las compañías de seguros pueden aumentar las tasas o denegar el servicio debido a un potencial problema genético? La Ley de No Discriminación por Información Genética (**Genetic Information Nondiscrimination Act** o **GINA**) de 2008 (legislación de EEUU), aborda algunas de estas preocupaciones mediante la prohibición del uso de la información genética en las decisiones de empleo y seguros de salud. En el fondo de la discusión está la necesidad de equilibrar los avances para mejorar la salud humana, con sus consecuencias éticas.

Área de Aplicación	Objetivo
Caracterización molecular de las enfermedades raras.	Proporcionar un diagnóstico definitivo y sugerir nuevas opciones de tratamiento.
Tratamientos individualizados contra el cáncer.	Predecir la respuesta de los pacientes a una terapia dirigida específica.
Farmacogenómica.	Optimizar la terapia con medicamentos para garantizar la máxima eficacia con efectos adversos mínimos.
Cribado de la población de riesgo de enfermedades.	Aumentar conocimiento de la salud individual y fomentar los cambios de salud proactivos.
Evaluación preconcepción y prenatal.	Informar a los padres sobre el riesgo de enfermedades de sus hijos.

Tabla 1: Áreas prometedoras en la medicina personalizada

Además de un papel en el cuidado de la salud, la secuenciación del genoma humano tiene, y seguirá teniendo, un gran impacto en nuestra comprensión de la historia humana, la evolución y biología general. En el campo de la filogeografía, los científicos examinan los patrones geográficos actuales y antiguos de la variación genética para obtener más información sobre la expansión global del ser humano y su adaptación. Estudios realizados a escala genómica han demostrado extensos cruzamiento entre poblaciones separadas, mucho más que las estimaciones previas basadas en los estudios de loci individuales. Otra área prometedora es la comparación genómica entre los seres humanos y otros organismos que ayudan a conectar la biología de

organismos modelo a la fisiología humana. Los estudios de genómica comparada también están ayudando a descifrar las funciones de los genes de codificación de proteínas, ARN no codificantes y secuencias reguladoras en la evolución. Del mismo modo mediante el estudio de la estructura y la actividad del propio genoma humano podemos plantearnos preguntas acerca de la función del ADN a nivel de genes, transcritos de ARN, proteínas y productos. Sin embargo, para dar respuesta a estas preguntas, los científicos deben hacer frente a los retos computacionales del análisis de todos los datos genómicos, su integración y visualización.

Secuenciación de ADN y búsquedas en las bases de datos

El primer paso en un proyecto de secuenciación es la obtención de los datos en bruto, el orden preciso de los cuatro nucleótidos: **A, G, C** y **T**. Existen varios enfoques para generar la información de la secuencia y nuevos métodos están apareciendo cada año. Dos métodos populares son la **Secuenciación de terminación de la cadena** y la **Secuenciación por síntesis**. La **Secuenciación de terminación de la cadena**, a menudo llamado **Secuenciación de Sanger**, permite a los investigadores generar una cadena larga de ADN a partir de una secuencia diana, también conocido como plantilla o molde ("template"). La plantilla de ADN se combina con un cebador de ADN ("primer"), el enzima ADN polimerasa I (ADN Pol I), y una mezcla de dos tipos de nucleótidos libres, desoxinucleótidos (dNTP) y didesoxinucleótidos (ddNTPs) (**Figura 1A**). Durante la reacción de secuenciación, el ADN Pol I utiliza la plantilla de ADN y añade dNTPs al cebador para formar una cadena complementaria de ADN. De vez en cuando, la ADN Pol I añadirá un ddNTP en la cadena de ADN, terminando la reacción (**Figura 1B**). Esta terminación final es debido a la falta de un grupo 3' hidroxilo en los ddNTPs (**Figura 1A**) por lo que es imposible para la polimerasa para añadir otro nucleótido en el extremo de la hebra. Por lo tanto, el resultado de la reacción es una serie de fragmentos de ADN de diferentes tamaños que se pueden separar por electroforesis capilar (**Figuras 1B, 1C**). Es importante destacar que cada ddNTP esta unido a un marcador fluorescente diferente, permitiendo que la fluorescencia de cada cadena de ADN individual pueda ser "leída" por un láser (**Figura 1C**). A continuación, los cuatro colores fluorescentes diferentes de los ddNTP se detectan automáticamente y la intensidad de fluorescencia se traducen en un "pico" de datos que representa el orden de los nucleótidos en la plantilla de ADN (**Figura 1D**). La **Secuenciación de Sanger** se introdujo en 1977 y fue el principal método utilizado para crear la primera secuencia del genoma. Todavía se utiliza hoy en día debido a su capacidad para generar secuencias largas para su lectura (500-800 pb).

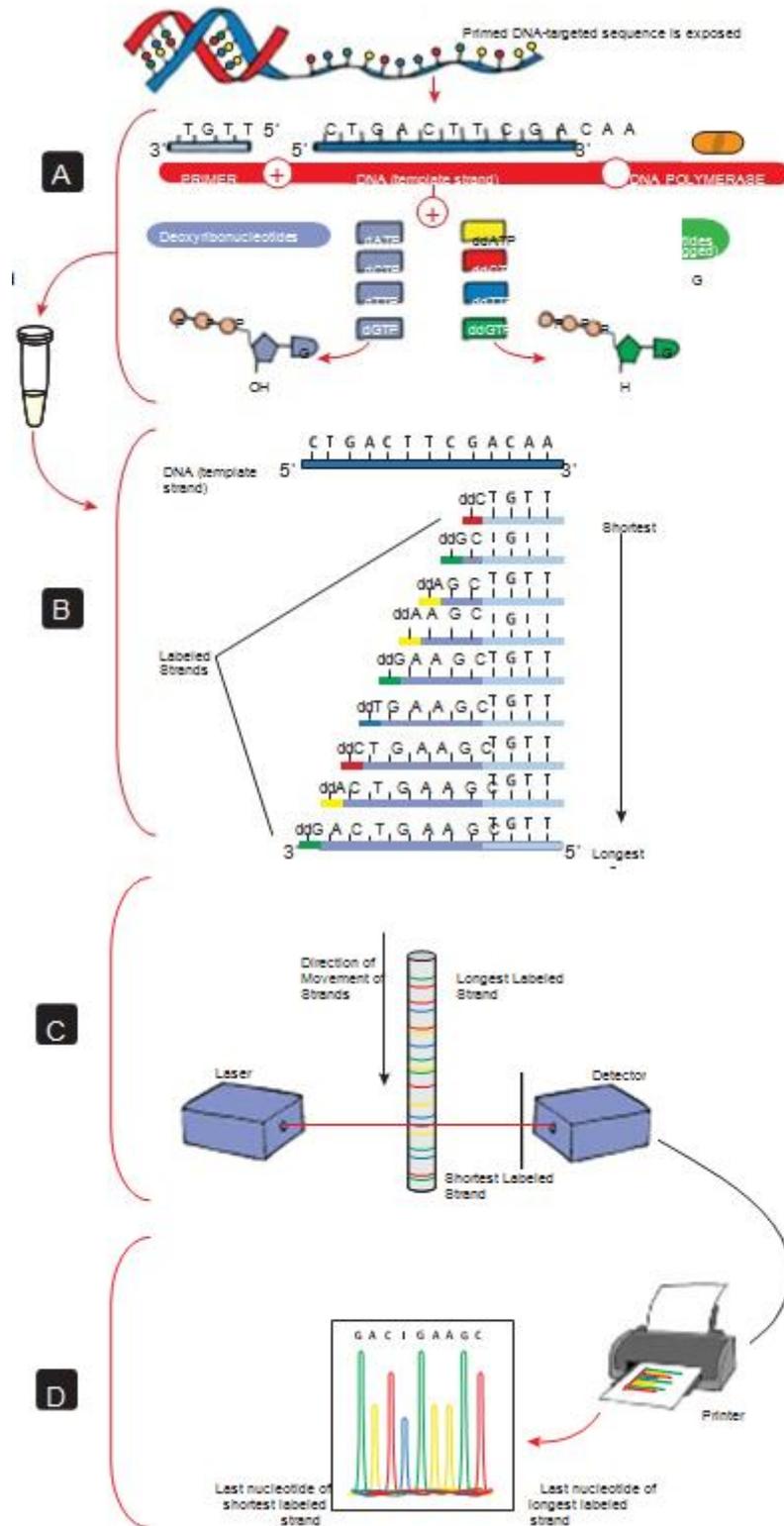


Figura 1: Secuenciación de ADN de Sanger.

(A) Creación de la reacción de secuenciación.

(B) Incorporación de los ddNTPs creando fragmentos de ADN de diferentes tamaños.

(C) La mezcla marcada se secuencia utilizando electroforesis en gel capilar. Un láser detecta la marca fluorescente de cada uno de los ddNTPs.

(D) La información se analizó utilizando un ordenador.

En la **Secuenciación por síntesis**, una hebra individual de ADN que complementa la hebra de interés se construye nucleótido a nucleótido, con cada nuevo nucleótido añadido se libera una señal específica. Antes de la secuenciación, la muestra de ADN debe prepararse e inmovilizarse. Esto implica la fragmentación al azar el ADN de interés, el anclaje del fragmento de ADN a una superficie sólida, y la eliminación de una de las hebras de la doble cadena. Este procedimiento produce zonas cortas detectables y distintas de moléculas idénticas de ADN. Estas moléculas de ADN reconstruyen ellas mismas la doble cadena de ADN. A medida que se añade cada nuevo nucleótido a la cadena complementaria se libera una señal única para cada nucleótido. Estas señales se registran en un ordenador y se traducen como la información de la secuencia. La ventaja de este método es que muchas cadenas de ADN diferentes se pueden examinar en una sola prueba. Esto significa que la **Secuenciación por síntesis** tiene un rendimiento mucho más alto y un coste menor que la **Secuenciación de Sanger**. Una desventaja de esta técnica es que produce más fragmentos de lectura cortos (50-150 pb), lo que significa que se deben realizar muchas lecturas para secuenciar un solo gen.

Interpretar la información de secuencia de ADN

Después de obtener los datos de secuenciación de ADN, los biólogos moleculares suelen buscar bases de datos públicas de secuencias similares. Esta búsqueda puede revelar una investigación ya realizada con el gen secuenciado, incluyendo la estructura tridimensional del producto del gen, enfermedades asociadas con la secuencia, y en que tejidos está activo el gen. En los casos en que la región no haya sido estudiada, encontrar secuencias similares pueden proporcionar pistas sobre la función de la secuencia y su relación evolutiva con otros genes humanos.

Uno de las mayores y más influyentes bases de datos es la GenBank. Esta base de datos de código abierto contiene más de un billón de bases de datos de secuencias de nucleótidos disponibles públicamente. Cada entrada en GenBank contiene una secuencia y un número de acceso, así como notas bibliográficas y biológicas tales como referencias de autor y los datos taxonómicos. El NCBI supervisa y mantiene la base de datos en su conjunto, pero cada entrada la presenta directamente cada laboratorio. La presentación directa de cada entrada ha permitido que la base de datos mantenga un ritmo rápido de crecimiento de los datos de secuencias. Sin embargo, esto también significa que existe heterogeneidad en la calidad de las entradas, especialmente en la certeza de la identidad de cada nucleótido y en la medida de notas adjuntas. Estas pueden variar dependiendo de los objetivos del estudio, las propiedades físicas de la región(es) de ADN, y el método de secuenciación elegido. Para solucionar esto, GenBank clasifica la información de la secuencia en base a la estrategia de secuenciación utilizada para obtener los datos.

Asociadas con esta base de datos hay varias herramientas bioinformáticas útiles, incluyendo la **Herramienta de Búsqueda de Alineaciones Locales Básicas** o **BLAST (Basic Local Alignment Search Tool)**. El programa **BLAST** encuentra regiones de similitud local entre la secuencia de ADN de un usuario y las secuencias de la base de datos de GenBank. Tal similitud sugiere homología, la existencia de ascendencia compartida entre los genes. En la terminología de BLAST la secuencia entrada por el usuario se conoce como la **secuencia de consulta**, las secuencias de la base de datos se conocen como **secuencias diana**, y secuencias con similitudes con la secuencia de entrada son "**hits**". El usuario puede extraer conclusiones sobre la posible función molecular de la secuencia de consulta al mirar los "hits". El programa **BLAST** también se utiliza para identificar las especies desconocidas, localizar dominios conocidos de proteínas, y encontrar posibles localizaciones cromosómicas.

El programa **BLAST** tiene un enfoque heurístico para el problema de la búsqueda en la gigantesca base de datos de secuencias diana. Esto significa que toma atajos con el fin de hallar las secuencias que coinciden en un tiempo razonable. Estos accesos

directos se basan en el supuesto de que las secuencias biológicamente similares contendrán tramos cortos con una coincidencia muy alta. El programa **BLAST** intenta hallar estos segmentos de elevada coincidencia mediante la eliminación de las regiones de baja complejidad, dividiendo la secuencia en fragmentos mucho más cortos, y luego escaneando la base de datos para encontrar las coincidencias. Una vez que se ha generado una lista de fragmentos con elevadas coincidencias, el programa **BLAST** amplía el campo de observación de los fragmentos para ver si las alineaciones que coinciden son más largas. Mediante este tipo de búsqueda en la base de datos de GenBank, el programa **BLAST** puede obtener resultados muy rápidamente a pesar de sacrificar algo de exactitud y precisión.

El programa **BLAST** es popular no sólo debido a su velocidad, sino también porque se calcula la significación estadística de las soluciones. Además del número de acceso, la descripción y enlace al genoma, el programa **BLAST** proporciona una **puntuación**, la **puntuación de bits**, y el **valor E**. La **puntuación**, **S** (del inglés "score"), es una medida de referencia de la calidad de la alineación entre la consulta y la respuesta positiva. Las variables elegidas por el usuario que incorporan conceptos moleculares y bioquímicos tienen una gran influencia en este valor. La **puntuación en bit** es la puntuación inicial ajustada en función del tamaño de la base de datos y la longitud de secuencia. El **valor E** se traduce en la probabilidad, debido a la casualidad, de que haya otra alineación con una puntuación superior a la puntuación **S** dada. **Puntuación**, la **puntuación de bits**, y el **valor E** son buenos indicadores iniciales de la similitud entre secuencias, sin embargo la alineación en sí también debe ser examinado para asegurar la exactitud.

Esta práctica introduce a los estudiantes en la genómica y la bioinformática. Con el fin de adquirir experiencia en la búsqueda en bases de datos, los estudiantes usarán el servicio gratuito ofrecido por el NCBI. En la actualidad, GenBank comprende varias bases de datos, incluyendo las secuencias de nucleótidos GenBank y EMBL, los CDS de GenBank no redundante (secuencias de proteínas), traducciones y la base de datos de Etiquetas de Secuencias Expresadas, EST (Expressed Sequence Tag). Para esta práctica, se recomienda utilizar la base de datos que ofrece el NCBI. Estos ejercicios implicar el uso de **BLASTN**, mediante el cual una secuencia de nucleótidos se compara con otras secuencias en la base de datos de nucleótidos. Para cada una de las tres secuencias, los estudiantes deben identificar una enfermedad humana potencial y discutir temas bioéticos relacionados.

4. DESCRIPCIÓN DEL EXPERIMENTO

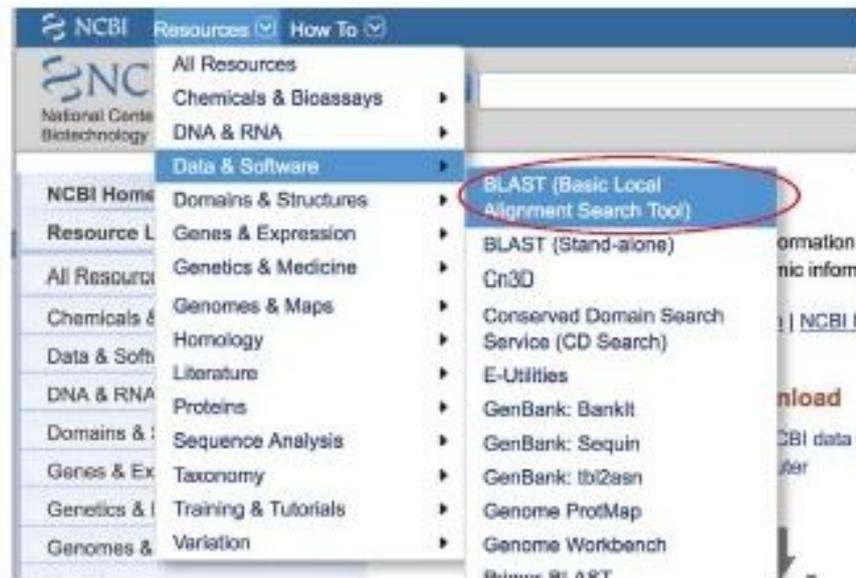
En esta práctica, los estudiantes leerán secuencias de ADN obtenidos a partir de técnicas de secuenciación automáticas de ADN. Los datos se analizaron utilizando las bases de datos disponibles al público para identificar los genes y productos genéticos. El impacto de la genómica se discutirá en el contexto de la sociedad actual.

4.1 Consideraciones previas

Los estudiantes pueden usar cualquier base de datos de secuencias del Genoma Humano para llevar a cabo la práctica. Para la realización de este protocolo se ha utilizado la base de datos ofrecida por el **Centro Nacional de Información Biotecnológica, NCBI (National Center for Biotechnology Information)**.

5. PRÁCTICA

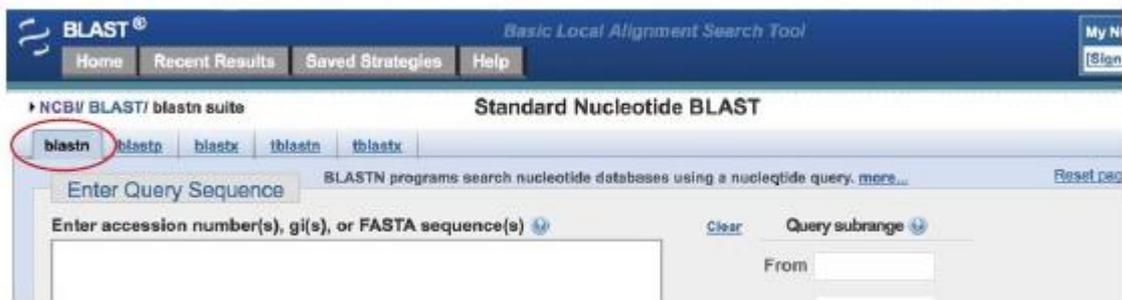
1. Escriba: www.ncbi.nlm.nih.gov para iniciar sesión en la página web del NCBI.
2. En la parte superior izquierda de la pantalla, haga clic en el apartado **"Resources"** (Recursos) del menú desplegable.
3. Haga clic en **"Data and software"** (Datos y software), a continuación, haga clic en **"BLAST (Basic Local Alignment Search Tool)"**.



4. En la nueva pantalla de inicio, seleccione **"Nucleotide blast"** (Nucleótido blast), que es la primera opción en la lista de **"BLAST Basic"**.

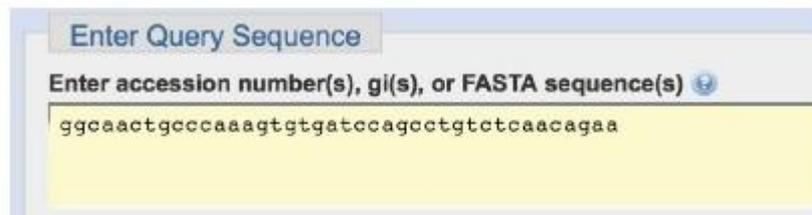


5. En la nueva pantalla asegúrese de que la pestaña seleccionada es "**BLASTN**".

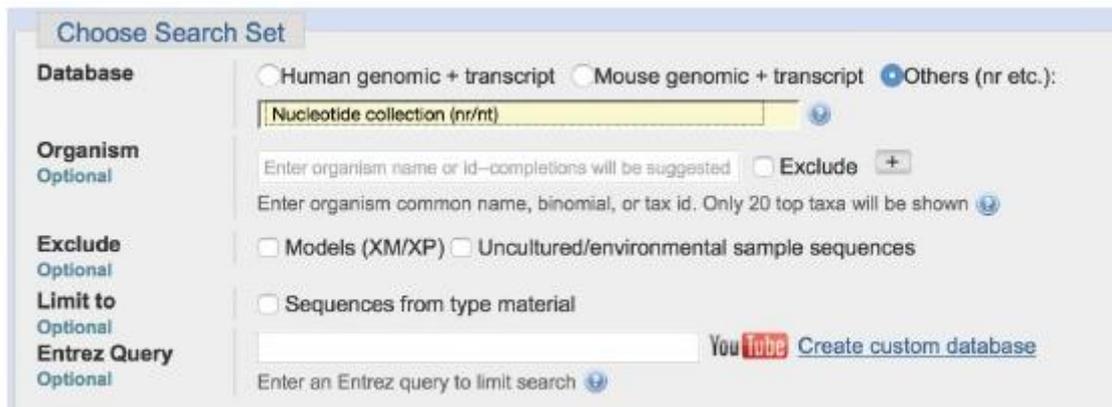


6. Introduzca la secuencia de nucleótidos en la caja grande en la sección "**Enter Query Sequence**" ("Entrar secuencia a consultar"); tenga cuidado al escribir la siguiente secuencia exactamente:

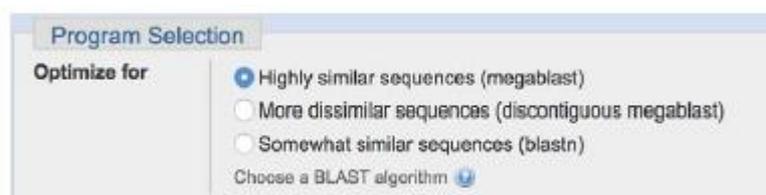
ggcaactgcccaaagtgtgatccagcctgtctcaacagaa



7. En "**Choose Search Set**" (Elija conjunto de búsqueda) asegúrese de que se ha seleccionado "**Others (nr etc)**" (Otros (nr etc.)) y que "**Nucleotide collection (nr/nt)**" (Colección de nucleótidos (nr/nt)) aparece en el menú desplegable. Las entradas restantes deben dejarse en blanco.



8. En la sección "**Program Selection**" (Selección del programa) seleccionar "**Highly similar sequence (megablast)**" (Secuencia de alta similitud (megablast)).



9. Haga clic en el cuadro de consulta azul **"BLAST"**



10. Una vez que se ha hecho clic en el cuadro de búsqueda **"BLAST"**, se le asignará un ID#. Anote este número para poder comprobar los resultados en un momento posterior.

11. Examine el informe de búsqueda **BLASTN**. El informe incluye:

a. **Informe Resumen de Búsqueda** muestra una visión general de los parámetros de la búsqueda **BLASTN**.

Your search parameters were adjusted to search for a short input sequence.

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [You](#)

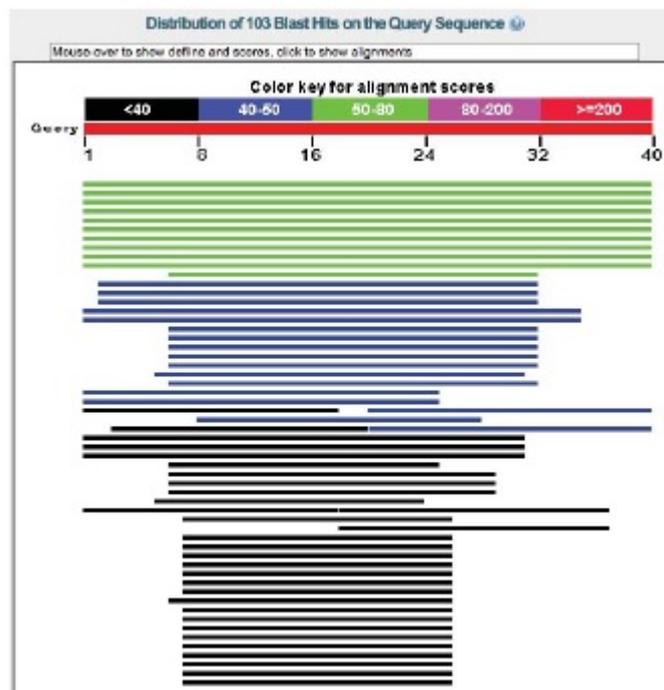
Nucleotide Sequence (40 letters)

RID S38347Y801R (Expires on 06-19 00:29 am)

Query ID lcl Query_44773	Database Name nr
Description None	Description Nucleotide collection (nr)
Molecule type nucleic acid	Program BLASTN 2.2.31+ Citation
Query Length 40	

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#)

b. Sección del **Gráfico Resumen** muestra la alineación de las coincidencias de la base de datos con la secuencia de consulta. El color de las cajas corresponde a la puntuación de la alineación, siendo el rojo el que representa las puntuaciones de alineación más altas.



clic en el nombre de la secuencia en la sección de **Descripción**, o desplácese hacia abajo a la sección de **Alineamiento**. A continuación, hacer clic en el identificador de secuencia. Esto nos lleva a obtener información adicional acerca de la secuencia sujeto, incluyendo el nombre del gen, el género y la especie de origen, y artículos escritos sobre el gen. Después de realizar esta búsqueda, uno de los "hits" debe ser **Bos taurus epidermal growth factor receptor (EGFR), mRNA** (Bos taurus factor de crecimiento epidérmico (EGFR), ARNm). Secuencia ID: **Ref |XM_002696890.4|**. Si la parte superior del hit no coincide, volver a entrar la secuencia. Asegurarse que los parámetros de la búsqueda del **BLASTN** son los correctos.

5.1 Ejercicio 1

Ahora que está familiarizado con el proceso de inscripción y presentación de **BLAST**, lea el análisis de la secuencia de ADN de la copia impresa del gel (cualquier carril) y encuentre el gen que identifica esta huella.

Para hacer esto:

- (1) Identificar la secuencia de nucleótidos (nucleótidos 100-200) a partir de la lectura de la secuencia de ADN.
- (2) Escriba al menos 70 bases en el cuadro de consulta del programa **BLAST** en la página web del NCBI. Las bases pueden ser de cualquier región de la secuencia, pero deben ser contiguas.
- (3) Examinar el informe de búsqueda **BLASTN**, identificar un gen probable, y examinar la identificación de genes para obtener información detallada.

Una vez que el gen ha sido identificado, responda a las siguientes preguntas:

- (A) ¿Cuál es el nombre de este gen?
- (B) En comparación con la entrada GenBank, ¿qué cadena has leído?
- (C) ¿Se puede encontrar algún artículo escrito sobre este gen? Anote el nombre de uno de los autores que han contribuido.
- (D) Al identificar una enfermedad causada por mutaciones en este gen. ¿Cuáles serían los motivos de un médico para realizar una búsqueda para esta enfermedad? ¿Y si quien realiza la búsqueda es una compañía de seguros?

Recuerde lo siguiente:

- La secuencia automatizada diferencia las bases de la siguiente manera: A es de color verde, C es azul, G es negro, y T es rojo.
- El ADN es de doble cadena y contiene una parte superior (5'→3') e inferior (3'→5') cadena (a veces esto corresponde a las hebras de codificación y no codificantes). Una secuencia de ADN se introduce siempre en la dirección 5'→3'.
- A veces es difícil de leer un pico de nucleótidos. Esto es particularmente cierto en el principio y al final de una secuencia de lectura donde los picos pueden superponerse. Estos lugares de difícil identificación a menudo son etiquetados con una N en lugar de con uno de los cuatro nucleótidos. Generalmente, es mejor saltar secciones con una gran cantidad de N's.
- Debido a que los investigadores pueden llamar a los mismos genes con diferentes nombres, pueden existir varias posibilidades para cada secuencia.
- Al hacer clic en el número de acceso de GenBank se puede acceder a información adicional, como la secuencia de proteína/aminoácidos, la descripción de la secuencia/gen, y los que contribuyen los nombres científicos.

- Algunas de las secuencias que tienen una puntuación elevada pueden predecir genes que no tienen ningún trabajo de investigación relacionados con ellos. Los estudiantes pueden tener que revisar varios emparejamientos antes de que hallar una secuencia con una referencia asociado.

5.2 Ejercicio 2

Intercambiar la copia de secuencia automática con otro grupo y enviar la secuencia de análisis **BLAST**. Anote el gen. Seleccione una secuencia asociada a un documento publicado, grabar el título y el primer autor del artículo.

1. ¿Qué es la bioinformática? ¿Cómo han avanzado en la tecnología de secuenciación en este campo?
2. Nombre dos métodos de secuenciación y describir el compromiso entre la velocidad de producción y la longitud de las secuencias producidas.
3. ¿Qué suposición hace **BLAST**? ¿Cuáles son las ventajas y desventajas de hacer esta suposición?

6. RESULTADOS Y PREGUNTAS DE ESTUDIO

La transducción de señales es un proceso importante por el que los receptores de las células actúan como amplificadores de la señal. La membrana de la superficie celular contiene sensores moleculares complejos que reciben, amplifican y reaccionan a señales extracelulares implicadas en el crecimiento y la diferenciación celular. La pérdida de capacidad de una célula para reaccionar a estas señales puede dar lugar a la transformación celular y la aparición de diversos tipos de cáncer.

El conjunto de tres secuencias corresponden a los nucleótidos de distintos genes que están funcionalmente relacionados y juegan un papel importante en la transducción de señales. La identidad de estos genes (y sus proteínas producto) puede determinarse utilizando la página web de NCBI como se ha descrito anteriormente en esta práctica. Tenga en cuenta que los detalles sobre identificación de genes, cadena, referencias e información sobre el autor, serán diferentes dependiendo del punto en que se centre el estudiante. Sin embargo, cada gen debe coincidir con la información básica que se presenta a continuación. Después de investigar las posibles enfermedades relacionadas con estos genes, los estudiantes pueden discutir los temas bioéticos relacionados.

Secuencia de ADN 1

Línea 1:

GNNNNNTGGNNNNNNNATANTTGC GGCCGCGGTTTTNTTTTTNTTNNNCNNGGAGCACAA
NCNAATGNANTGTGTTGTTGTGGCGARGGC-
GARGGCGCCGTTHTAAACTGTCTCCTGATATCCTACACAACAAACAAATTTTCAT

Línea 2:

CGGAGTATGTACCGACTGTTTTGACA ACTATGCAGTCACAGTTATGATTGGTGGAGAACCAT
ATACTCTTGGACTTTTTGATACTGCAGGGCAAGTTATGA-
CAGATTACGACCGCTCACTTATCCACAAACAG

Línea 3:

ATGTATTTCTAGTCTGTTTTTCAGTGGTCTCTCCATCTTCATTTGAAAACGTGAAAGAAAAGTG
GTGCCTGAGATAACTCACTGTCCAAGACNCCTTTCTT-
GCTTGTTGGGACTCAAATTGATCTCAACGAGATGACCC

Línea 4:

CTCTACTATTGAGAACTTGCCAAGAACAACAGAAAGCCTATCACTCCANAGACTGGGTGAAA
AGCTGGCCCGTGACCTGAANGCNGTCAAAGTATGTG-
GAGTGTCTGCACCTACACAGCAGANGTCTGAAAAATGTGTTNATGAAGC

Secuencia 1: Ciclo de división celular 42 (proteína de unión a GTP) o CDC42
La CDC42 es una molécula reguladora intracelular clave que está implicada en la forma de la célula. En células de mamíferos, la proteína CDC42 regula el citoesqueleto de actina (el andamiaje de una célula) para producir una estructura filamentosa llamada filopodio, que están implicadas en la detección de otras células y el movimiento celular. Las mutaciones en CDC42 y sus proteínas asociadas se correlacionan con ciertos tipos de cáncer.

Secuencia de ADN 2

Línea 1:

TGCNNNNNTGGNTNNGGNNNNNATTGNNTCNCTNTACCATGCNNGNGCACAANGTTTTTTTT
TTTTTTTTTTGGGCAAAGCGTACAAAGGTTT-
CAAGGGACAGGACCAAGAACGAGGGGCTGAGACATTTACAACAGCAGGCATT

* NOTA: Para obtener los mejores resultados, se recomienda la eliminación de los primeros 73 nucleótidos (en rojo) en el análisis de este carril.

Línea 2:

TTTCTCTTCTTCTTCTTACGGGAGGCGGGCANAGGACTGCTCGGATCGCTTCGTCAAACACT
GTCTTGAGGCCTCNCTGTGTGAGCGCCGAGCACTCCAG-
GTATTTTACAGCACCAATCTCCTTANCCATGGCTANAC

Línea 3:

CCCTGCGGATAGGTGATGGGAGTCAGCTTCTTBCCTTCAGTTTCNCNATCGTGTCTTTATCAT
CCCTAAGATCAAGTTTAGTTCCCACTANGATGATGGGAGT-
GTTGGGACAGTGGTGCCGCACCTCAGGATACCACTT

Línea 4:

TGCACGGACATTTTCAAATGATGCAGGACTCACAAGGGAAAAGCAAATTAAGAACACATCTGT
TTGCGGATAGGATAGGGGGCGTATTCTGTGTCATA-
ATCTTCTTGTCCAGCTGTATCCCAAAAAGCCCAGATTCACCGGTTT

Secuencia 2: Ras relacionada con toxina botulínica C3 substrato 1 (familia rho, pequeña unión GTP de proteína Rac1) o RAC1

El gen RAC1 para una proteína reguladora que altera el citoesqueleto de actina para producir ondulaciones en la membrana, importantes para la motilidad celular. RAC1 es importante en un gran número de enfermedades, incluyendo el cáncer. Muchos tumores utilizarán la señalización RAC1 para conseguir un aumento de la motilidad y la capacidad invasiva, una de las primeras etapas requeridas para que el cáncer pueda propagarse por todo el cuerpo.

Secuencia de ADN 3

Línea 1:

TGNNNNNNTGNNNNNNNGNNANAAACGAAGTGCAGACTCAAAGTGCCATCTCCCTCCCGAC
CATTGGAGGATCCCAAGCTCTATGTTGCCCTTATTGT-
CACCAGTGACATTTAATTCAAACAGGAGTCTTTCGGGCCAGCAA

Línea 2:

GCTGCCAGGCTTAGCTGCGAGCCCGTCATGGAGGAAAAAAGCTCAGGAGAAAACAGTCTG
TTGGAGAATGGGACAGTCCACCAGGGAGACACCTC-
GTGGGGCTCCAGCGGTTCTGCATCTCAGTCCAGCCAAGGCAGA

Línea 3:

GACAGCCACTCCTCCAGCCTGTCCGAACAGTACCCCGACTGGGCCAGCCGAGGACATGTTTG
ACCATCCCACCCCATGCGAGCTCATCAAGGGGAAAGAC-
TAAGTCAGAGGAGTCCCTCTCTGACCTTACAGGTTCCCTCCCTCCTCTCC

Línea 4:

CCTGCAAGCTTGATCTTGGGCCCTCACTTTTGGATGANGTGCTGAATGTTATGGATAAAAATA
AGTAACTCGAGCATGCATCTAGAAAGGGCCTATTCTATA-
ATGTACCTAAATGCTAAACCTCGCTGATCAGCCTCGACTGTGCCNT

* NOTA: Para obtener los mejores resultados, se recomienda la eliminación de los primeros 70 nucleótidos (en rojo) en el análisis de este carril.

Secuencia 3: proteína efectora CDC42 (Rho GTPasa vinculante) 3 o CDC42EP3

La CDC42 (secuencia 1) regula la formación de estructuras que contienen F-actina, mediante su interacción con diferentes proteínas efectoras. La CDC42EP3, una de estas proteínas efectoras, está implicada en la unión a la proteína CDC42, provocando cambios en la forma de una célula. Una célula detecta señales del ambiente externo mediante la creación de redes de moléculas de transmisión que indican a la célula cómo responder a dichas señales externas. Los defectos en estas moléculas de transmisión están implicados en muchas enfermedades, incluyendo cáncer.

6.1 Preguntas de debate

Las siguientes preguntas son complejas y no siempre tendremos una respuesta correcta o incorrecta. Se plantean para estimular el debate sobre la aplicación de la genómica en el siglo XXI.

- 1. ¿Hay que hacer pruebas prenatales para las enfermedades que son actualmente incurables?**
- 2. ¿Hay que hacerles pruebas a nuestros hijos para enfermedades de aparición en la edad de adulto?**
- 3. ¿Qué papel debe desempeñar el gobierno en el establecimiento de directrices o leyes que regulen las pruebas genéticas humanas?**