

Hallando genes y explorando la página del gen (Ejercicio 1)

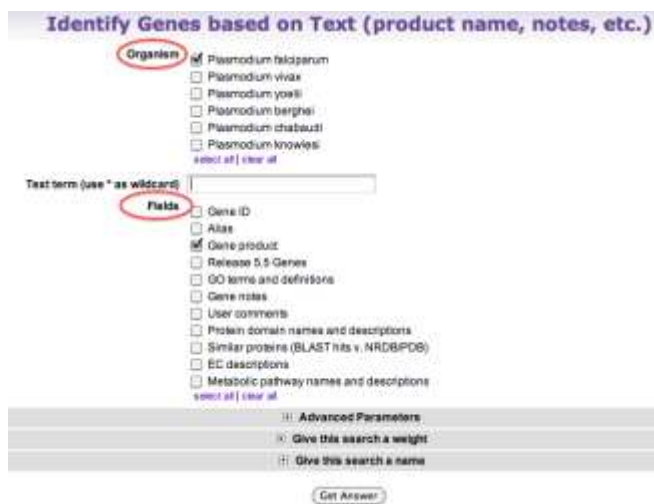
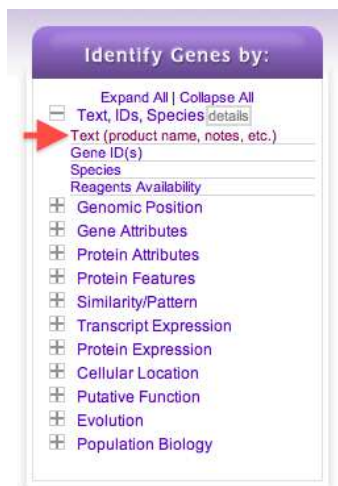
1.1 Hallando un gen usando la búsqueda de texto.

Nota: Para este ejercicio use <http://www.plasmodb.org>

- a. Hallar todas las posibles cinasas (kinases) en *Plasmodium*. (pista: use la palabra clave “kinase” en la caja “Gene Text Search”).



- ¿Cuántos genes obtuvo?
 - ¿Cuántos de estos genes son de *P. falciparum*? ¿Cómo hizo para descubrir eso?
 - ¿Qué sucede si busca usando la palabra “kinases”? ¿Cuántos resultados tienes de vuelta?
- b. ¿Cómo puedes incrementar el número de cinasas en tu resultado? (pista: La búsqueda hecha en ‘a’ no encontrara cosas como “6-phosphofructokinase” or “kinases” por tanto se necesita usar un comodín (“wildcard”) – intenta “kinase*”, “*kinase” y “*kinase*”
- ¿Obtuviste mas resultados?
 - ¿Cuál de las anteriores combinaciones de comodines dio la mayor cantidad de resultados?
- c. Hallar únicamente cinasas que específicamente tengan la palabra “kinase” en el nombre del producto del gen. (pista: para hacer esto tu necesitas ir la página de búsqueda por texto – hay muchas maneras de llegar ahí, ¿cómo llegaste ahí?)



- ¿Cuantas cinasas tienen la palabra kinase en nombres de producto? (pista: ¿Recuerdas el uso de comodín?)

1.2 Combinando los resultados de búsqueda de texto con los resultados provenientes de otras búsquedas.

- a. En el ejercicio 1.1 usted identificó genes que tienen la palabra “kinase” en alguna parte en el nombre de su producto. ¿Puede ahora hallar cuantas de estas cinasas son secretadas? (pista: aumenta tu estrategia agregando un paso. Escoge una búsqueda que identifique genes con un probable (péptido señal) preguntar Silvia de secreción). ¿Cómo combinaste los resultados de las búsquedas?

My Strategies: New Opened (1) All (1) Basket Examples Help

(Genes)

Text* Rename Copy Save As Share Delete

Text 702 Genes Step 1 Add Step

Run a new Search for Transform by Orthology Add contents of Basket Add existing Strategy Filter by assigned Weight

Genes Genomic Segments (DNA) Motif SNPs GRFs SAGE Tags

Text IDs, Species Genomic Position Gene Attributes Protein Attributes Protein Features Similarity/Pattern Transcript Expression Protein Expression Cellular Location Putative Function Evolution Population Biology

Predicted Signal Peptide Transmembrane Domain Count P.F. Subcellular Localization Exported Protein

Questions or comments

Plasmod 7.2 May 3, 2011 2011 The EUPeptide Project

Add Step

Add Step 2 : Predicted Signal Peptide

Organism Plasmodium falciparum Plasmodium vivax Plasmodium yoelii Plasmodium berghei Plasmodium chabaudi Plasmodium knowlesi select all | clear all

Advanced Parameters

Minimum SignalP-NN Conclusion Score 3

Minimum SignalP-NN D-Score 0.5

Minimum SignalP-HMM Signal Probability 0.5

Matches any or all advanced parameters any

Give this search a weight

Give this search a name

Combine Genes in Step 1 with Genes in Step 2:

1 Intersect 2 1 Minus 2

1 Union 2 2 Minus 1

1 Relative to 2, using genomic locations

Which operation did you choose?

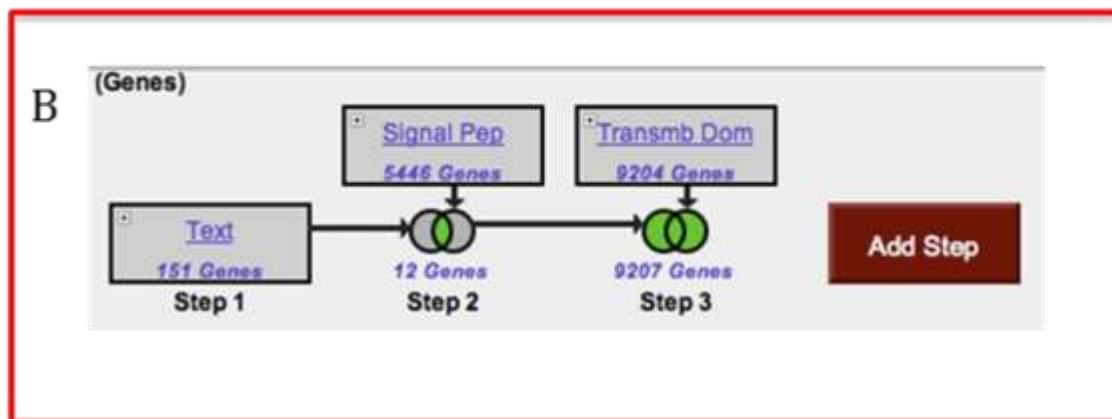
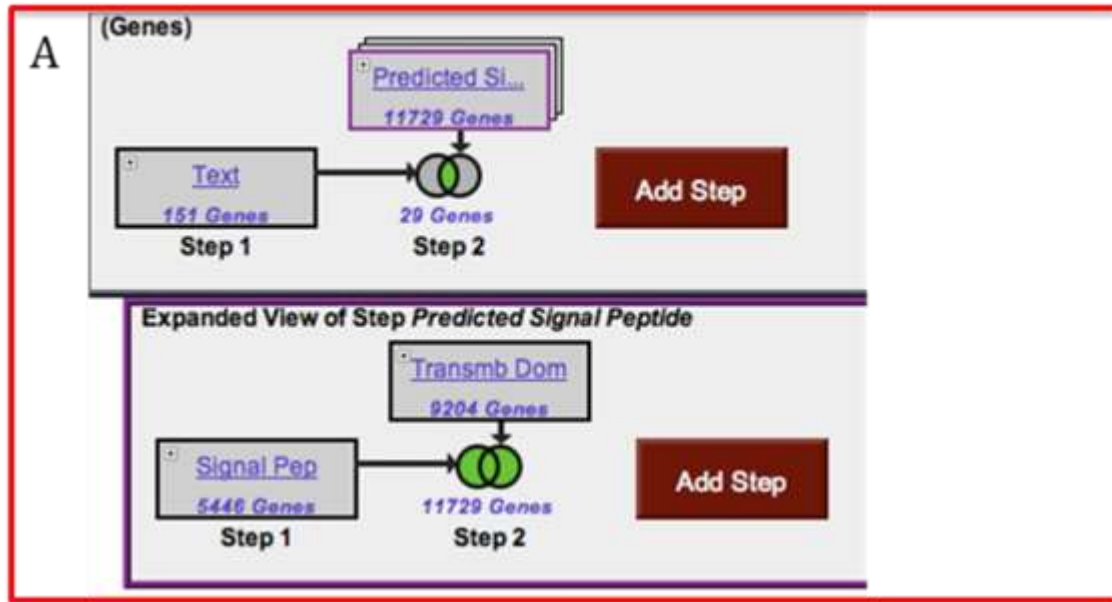
Run Step

Description: Find genes that are predicted to encode a signal peptide. The predictions are made with the SignalP program.

- b. Ahora que tienes las posibles cinasas secretadas, ¿cómo podrías expandir aún mas tu estrategia? (pista: No hay una respuesta equivocada aquí.... Desde el punto de vista biológico ¿qué otra cosa desearías conocer de estas cinasas? Si quiere, puede agregar mas búsquedas para aumentar esta estrategia.)
- Por ejemplo, ¿Cuántas de estas cinasas tienen residuos transmembranales (check with Silvia)?
- c. En el ejemplo anterior, ¿Cómo puedes definir cinasas que tienen tanto un péptido señal **O** dominio(s)? (pista: Para hacer esto de manera apropiada usted necesita generar una “Nested Strategy” (Estrategia anidada). ¿Por qué?)

The image displays two screenshots of a bioinformatics strategy editor interface. The top screenshot shows a strategy with two steps: 'Text' (114 Genes) and 'Predicted Signal Peptide' (5448 Genes). The 'Predicted Signal Peptide' step is expanded, showing parameters like Organism, Minimum SignalP-NN Conclusion Score, and Minimum SignalP-NN D-Score. A red circle highlights the 'Make Nested Strategy' button. The bottom screenshot shows the same strategy with an 'Expanded View of Step Predicted Signal Peptide' section below it, which contains a 'Text(3)*' step and an 'Add Step' button. A red arrow points from the 'Make Nested Strategy' button in the top screenshot to the 'Expanded View' section in the bottom screenshot.

**Note los diferentes resultados obtenidos en A (con anidamiento) y B (sin anidamiento) abajo:*



1.3 Visitando una página de un gen específico.

Nota: Para este ejercicio use <http://www.plasmodb.org>

- Hallar el gen bifuncional dihidrofolato timidilato reductasa sintetasa (reductase-thymidylate synthase) (DHFR-TS, por sus siglas en inglés) (o, si usted prefiere, el antígeno 1 apical de membrana; AMA1) en *P. falciparum*.
 - ¿Cómo navegó hasta este gen? ¿Qué otras vías pudieras usar para llegar ahí? (pista: ¿qué tal usando el gene ID? PFD0830w)
 - ¿Cuántos exones hay en este gen? ¿Cuántos nucleótidos tiene la secuencia codificante (Silvia)?
- ¿Cuales genes están localizados aguas arriba y aguas abajo (Silvia) de DHFR-TS (AMA1) en *P. falciparum*?
 - Se mantiene la sintenia (organización de los cromosomas) en esta región, en otras especies?

- ¿Cuan completo es el ensamblado del genoma para las otras especies? (pista: quizás le ayude ver el “genome browser”). Se puede acceder al “genome browser” desde la página de un gen haciendo click en el link “view in genome browser” (ver imagen abajo). Las secciones (“tracks”) del ”genome browser” pueden ser seleccionadas y luego haga click en el botón “update image”. ¿Que secciones selecciono usted?



- ¿Cuántos polimorfismos de un solo nucleótido (SNPs, por sus siglas en inglés) puedes identificar dentro del gen de *P. falciparum* DHFR-TS (o AMA1)? (Nota: regrese a la página del gen - use la flecha de regreso en su navegador o haga click en el gen resaltado en amarillo).
 - ¿Cuántos de estos SNPs están en la secuencia codificante? ¿Cuánto impactan la secuencia de la proteína predicha? ¿Cuántos alelos hay por cada SNPs por cepa (Silvia)? ¿Cuál es el máximo número de SNPs por cepa? (¿Es probable definir el espectro total de variación de secuencia en esta cepa en particular? ¿Qué hay acerca de los campos aislados? Rewrite) (pista: coloca el ratón sobre los rombitos para obtener mas información acerca de los SNPs).

- ¿Cómo compara estos resultados con la distribución de SNP en otros genes?
- d. ¿Está el gen DHFR-TS (o AMA1) expresado? (pista: busque en las secciones de la página del gen que se titulan “Protein” y “Expression” - quizás debes hacer click en “Show” para ver la información).
- ¿Qué tipo de datos en PlasmoDB proporcionan evidencia de expresión? ¿En cuál estadio del ciclo de vida se encuentra DHFR-TS (AMA1) más abundante? ¿Tiene esto algún sentido?
 - ¿Cómo comparar los diferentes perfiles de expresión en microarreglos de cada estadio con cada uno de los otros? ¿Los resultados son similares? ¿Qué hay acerca de los datos de RNA-sequence?, ¿están estos de acuerdo con los datos de microarreglos?
 - ¿Cuán abundante es la proteína DHFR-TS (AMA1)? ¿Qué tan seguro estás tú de este análisis?

1.4 Hallando genes a través de BLAST.

Nota: Para este ejercicio use: <http://www.toxodb.org>

Adicionalmente, usted necesitará abrir otra ventana en su buscador para salvar la secuencia que usted usará en nuestra búsqueda de BLAST, ir a:

<http://tinyurl.com/ycz2hwh>

- Imagine que usted generó un mutante por inserción en *Toxoplasma* este le provee algunos de los resultados mas interesantes de su carrera! Usted secuenció la región flanqueante y usted solo puede obtener la secuencia a partir de la inserción (la secuencia en el enlace anterior). Usted inmediatamente va a ToxoDB para hallar cualquier información acerca de esta secuencia. ¿Qué hace usted?
- Trate de ejecutar un BLAST con esta secuencia (pista: Usted puede llegar a la herramienta BLAST haciendo click en el enlace BLAST s que se encuentra debajo de “Tools” en la página principal).

The screenshot shows the ToxoDB website with the 'Tools' menu open. A red arrow points to the 'BLAST' option in the menu. Another red arrow points to the 'Tools' section on the right side of the page, which also lists 'BLAST' and other tools.

- ¿Cuál programa de blast debería usar? (pista: Intnte diferentes combinaciones, justo manteniendo en mente que usted tiene una secuencia de nucleótidos, entonces tiene que usar un arpoiado programa de BLAST).

BLAST

Target Data Type

Transcripts
 Proteins
 Genome
 ORF
 EST
 Isolates

BLAST Program blastn blastp blastx tblastn tblastx

Target Organism

Neospora caninum
 Toxoplasma gondii GT1
 Toxoplasma gondii ME49
 Toxoplasma gondii VEG
 Toxoplasma gondii apicoplast
 select all | clear all

Nucleotide Sequence

Expectation value

Maximum descriptions (V)

Maximum alignments (B)

Low complexity filter

1. Choose what you want to BLAST against
2. Choose which BLAST program to use
3. Choose your target organism

- ¿Obtuvo usted algún resultado de blastx? ¿tblastn? ¿Qué hay de blastn?
- ¿Qué es tu gen? (pista: después de ejecutar un blastn contra la secuencia genómica de *Toxoplasma*, haga click sobre “link to the genome browser”. En el buscador del genoma aumente la escala para ver que’ gen está en el área.)

Nota sobre los programas BLAST: *blastp* compara una secuencia de amino ácidos contra una base de datos de secuencias de proteínas; *blastn* compara una secuencia de nucleótidos contra una base de datos de secuencias nucleotídicas; *blastx* compara los seis marcos conceptuales de productos de traducción de una secuencia nucleotídica (ambas cadenas) contra una base de datos de secuencias de proteínas; *tblastn* compara una secuencia de proteína contra una base de datos de secuencias nucleotídicas dinámicamente traducidas en todos los seis marcos de lectura (ambas cadenas); *tblastx* compara los seis marcos de traducción de una secuencia nucleotídica contra los seis marcos de traducción de una base de datos de secuencias nucleotídicas.

1.5 Más BLASTing en EuPathDB (opcional).

Nota: para este ejercicio use <http://www.eupathdb.org>

- Lo primero que vamos a necesitar es una secuencia para usar en BLAST. Buscar por palabras claves "dihydrofolate". (Pista: bajo “Quick Search” en la página principal de EuPathDB).
- Usted debería obtener múltiples resultados. Halle la primera que está anotada como "dihydrofolate reductase-thymidylate synthase" (DHFR-TS). (Pista: busque en la columna “product description”).
- Una vez que usted ha hallado una, haga click sobre el gen y vaya a la página del gen. Haga click sobre el tercer resultado, el cual es de *Cryptosporidium parvum*. (pista: GenelD: cgd4_4460)
- Descienda hasta la parte inferior de la página del gen para la información de la secuencia.
- Nosotros vamos a comenzar con un BLAST para proteínas.
- Abre una nueva ventana y vaya a la página de BLAST de la página principal de EuPathDB. (pista: debajo de “Tools”).
- Copie la secuencia de amino ácidos de la parte inferior de la página del gen en la ventana del paso d en el paso l. (dont understand what is step l)
- Pegue la secuencia en la ventana de “input”.
- Seleccione el tipo de bases de datos (target data type) (Vamos a comenzar con "Proteins").
- Seleccione el programa BLAST. (Pista: BLASTP).
- Seleccione el organismo (target organism) (vamos a comenzar por seleccionar todos). Haga click en "Get Answer".

Basados en los resultados usted debiera haber identificado excelentes

resultados en casi todos los patógenos en EuPathDB pero, ¿puedes hallar unos buenos resultados en *Giardia*, *Entamoeba* or *Trichomonas*?

Vamos a intentar un método diferente de BLAST:

- a. Vuelva a la ventana de BLAST. Cambie el tipo de base de datos a "Genome".
- b. Seleccione el programa de BLAST. Fíjese que no puede seleccionar BLASTP. Intente las otras opciones. Note como su tipo de secuencia de entrada ha cambiado cuando usted selecciona un programa diferente. (Pista: TBLASTN Es el que usted necesita).
- c. Seleccione todos los organismos blancos. Haga click sobre "Get Answer".

Basados en sus resultados usted debería notar que usted ahora además tiene una secuencia de *Theileria*. Sin embargo, nosotros aún no tenemos la dihydrofolate de *Giardia* y *Trichomonas*.

Vamos a intentar un método diferente de BLAST.

- a. Ir a la ventana de la página del gen y copiar la secuencia de nucleótidos.
- b. Ir a la ventana de BLAST y pegar la secuencia de nucleótidos dentro de la ventana de entrada.
- c. Seleccione el tipo de base de datos (intenta con diferentes) y el programa de BLAST. Note que solamente puedes seleccionar TBLASTX o BLASTN cuando ingresa una secuencia de nucleótidos. (Pista: selecciona TBLASTX).
- d. Seleccione el organismo en las bases de datos (target organismo). Esta vez vamos a seleccionar específicamente solo *Giardia*, *Entamoeba* y *Trichomonas*. Haga click sobre "Get Answer".

¿Se encuentra frustrado? No obtener ningún resultado para *Giardia*, *Entamoeba* y *Trichomonas* en este caso es en realidad la respuesta correcta! Ambos no tienen actividad para dihydrofolato reductasa o timidilato sintetasa.