

Ejercicios usando FungiDB

1.1. Encontrar un gen usando la búsqueda por texto.

a. Encontrar todas las posible kinasas en FungiDB. Pista: usar la palabra clave “kinases” (sin comillas) en el cuadro “Gene Text Search”.



- Cuantos genes encontraron?
- Cuantos corresponden a *Cryptococcus neoformans* (H99)? Donde encontraron esta información?
- Que pasa si buscan “kinases” con comillas? Cuantos resultados obtuvieron?

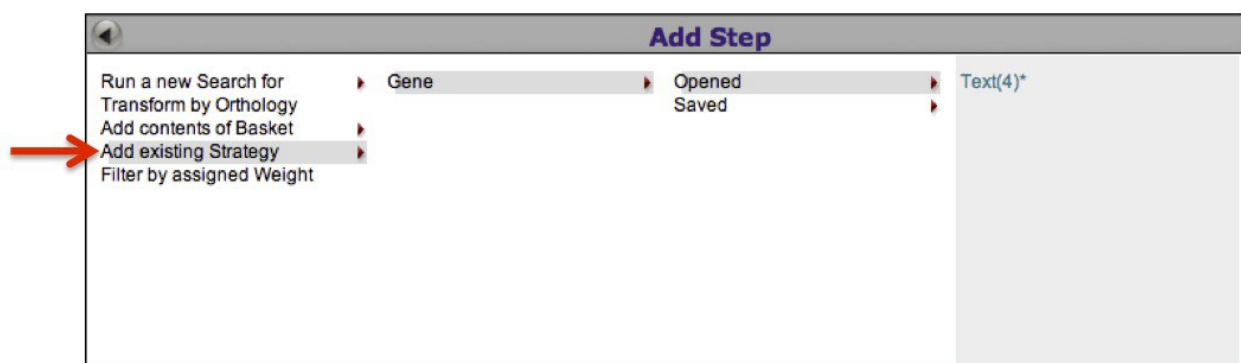
b. Como se puede aumentar el número de posibles kinasas en tu búsqueda?

Pista: La búsqueda que hicimos en ‘a’ no va a encontrar por ejemplo “diphosphokinase” o “kinases” entonces tenes que usar la wild card (no encontré traducción adecuada) que es el asterisco. Prueben buscar *kinase, kinase* y *kinase*.

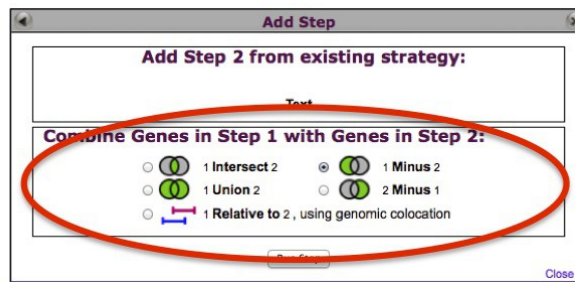
- Encontraron más resultados?
- Que combinación de wild card dio el número más grande de kinasas?

Ahora, como podemos examinar rápidamente los genes que se identificaron usando la palabra clave *kinase* pero no con la palabra clave kinase?

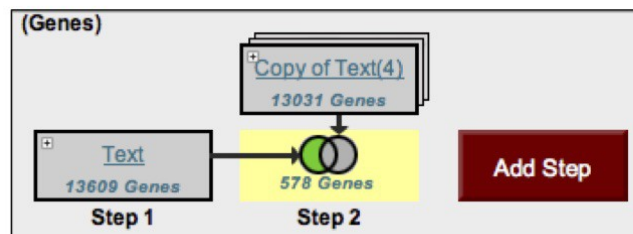
Pista: Se puede hacer combinando estrategias de búsqueda. Clickear “Add Step” y después seleccionar “existing strategy”.



Seleccionar la estrategia utilizada previamente de la lista. Luego elegir la operación que consideren correcta para combinar las estrategias:



Which operation did you choose?

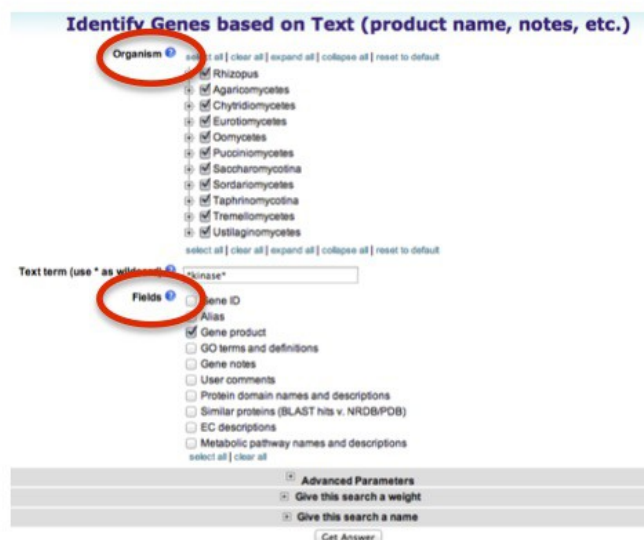
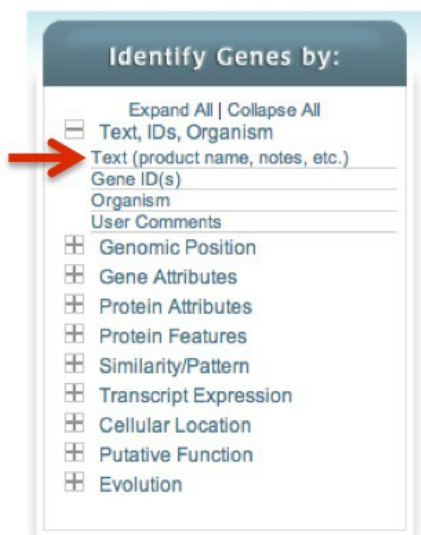


- Esta estrategia es coherente? Todos los resultados contienen la palabra kinase?

c- Encontrar solo las kinasas que tienen específicamente la palabra “kinase” in su nombre (Gene Product Name).

Pista: Usar la pagina de “Text Search” donde pueden definir los campos. Hay muchas formas de utilizar esta página.

- Como llegaron hasta esa página?
- Cuantas kinasas tienen la palabra “kinase” en su nombre?
- Se acordaron de usar la wild card?



1.2. Combinar búsqueda por texto con resultados de otras búsquedas

a. En el ejercicio 1.1 identificaron genes que tenían la palabra “kinase” en algún lugar de su “Product Name”. Ahora, pueden buscar cuantas de estas kinasas son supuestamente secretadas?

Pista: expandir la estrategia de búsqueda agregando un paso. Elijan una búsqueda que identifique genes que tengan peptidos señal secretorios putativos. Como podemos combinar los resultados de ambas búsquedas?

My Strategies: New Opened (1) All (47) Basket Examples Help

(Genes)

Text(7)*
Rename
Duplicate
Save As
Share
Delete

Text
5608 Genes
Step 1

Add Step

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
ORFs

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

Predicted Signal Peptide
Transmembrane Domain Count

Close nes

Add Step 2 : Predicted Signal Peptide

Organism select all | clear all | expand all | collapse all | reset to default

☒ Rhizopus
☒ Agaricomycetes
☒ Chytridiomycetes
☒ Eurotiomycetes
☒ Oomycetes
☒ Pucciniomycetes
☒ Saccharomycotina
☒ Sordariomycetes
☒ Taphrinomycotina
☒ Tremellomycetes
☒ Ustilaginomycetes

select all | clear all | expand all | collapse all | reset to default

☐ Advanced Parameters
☐ 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 colocation

Run Step

Which operation did you choose?

b. Ahora que tienen una lista de posibles kinasas secretadas, como podríamos expandir aún más la búsqueda?

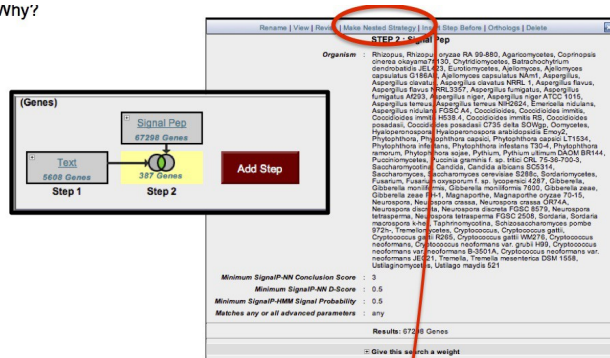
Pista: es un ejercicio libre para practicar...

- Desde un punto de vista biológico, que otra cosa sería interesante saber sobre estas kinasas? Agreguen mas búsquedas para que la estrategia crezca.

- Por ejemplo, cuantas de estas kinasas secretadas tienen además dominios transmembrana?

c. En el ejemplo anterior, como se pueden definir kinasas que tengan un péptido señal de secreción y/o dominios transmembrana?

Why?

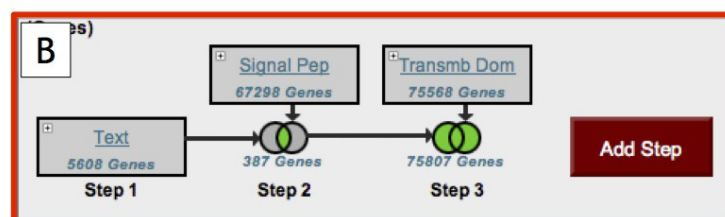
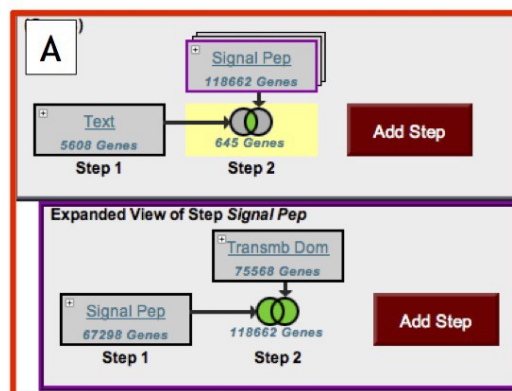


Pista: hay que usar una estrategia anidada (Nested Strategy). Por que?



Ahora les muestro los diferentes resultados obtenidos con anidado (A)

o sin anidado (B):



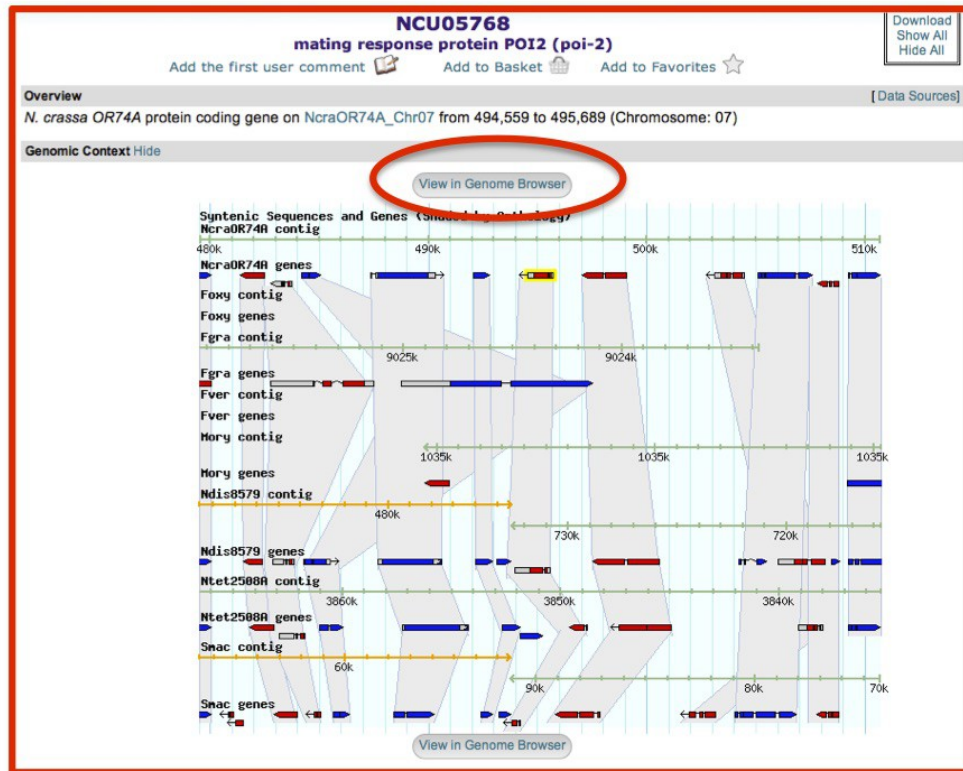
1.3. Visitando la página de un gen específico en FungiDB

a. Encontrar el gen poi-2 (maiting response protein POI2).

- Como llegamos a este gen? De que otras maneras? (Pista: el ID del gen es NCU05768).
- Cuantos exones tiene este gen?
- Cuantos nucleótidos tiene su secuencia codificante?

b. que genes se localizan corriente arriba y abajo de POI2 en *Neurospora crassa*?

- Se mantiene la sintenia de esta región en otras especies?



c. POI2 tiene ortologos en otras especies fúngicas?

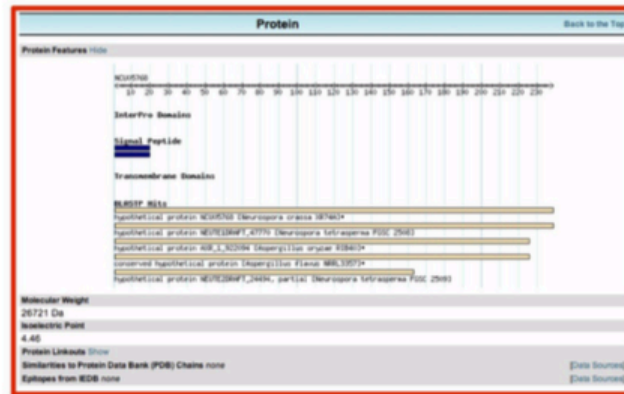
Volver a la página del gen. Mirar en la sección de anotaciones (Annotation) la tabla en “Orthologs and Paralogos in FungiDB”.

Annotation					Back to the Top
Add a comment on NCU05768					
User Comments none					[Data Sources]
External Links Show					[Data Sources]
Names, Previous Identifiers, and Aliases Show					[Data Sources]
GO Terms none					[Data Sources]
EC Number none					[Data Sources]
Orthologs and Paralogos within FungiDB Hide					[Data Sources]
Gene	Organism	Product	is syntenic	has comments	
AFLA_094980	Aspergillus flavus NRRL3357	conserved hypothetical protein	yes	no	
FOXG_17469	Fusarium oxysporum f. sp. lycopersici 4287	unspecified product	no	no	
Ndis8579_168059	Neurospora discreta FGSC 8579	unspecified product	yes	no	
NEUTE1DRAFT_47770	Neurospora tetrasperma FGSC 2508	hypothetical protein	yes	no	
SMAC_09010	Sordaria macrospora k-hell	unspecified product	yes	no	

Find the group containing NCU05768 in the OrthoMCL database

d. POI2 se secreta?

Vayan a la sección “Protein”.



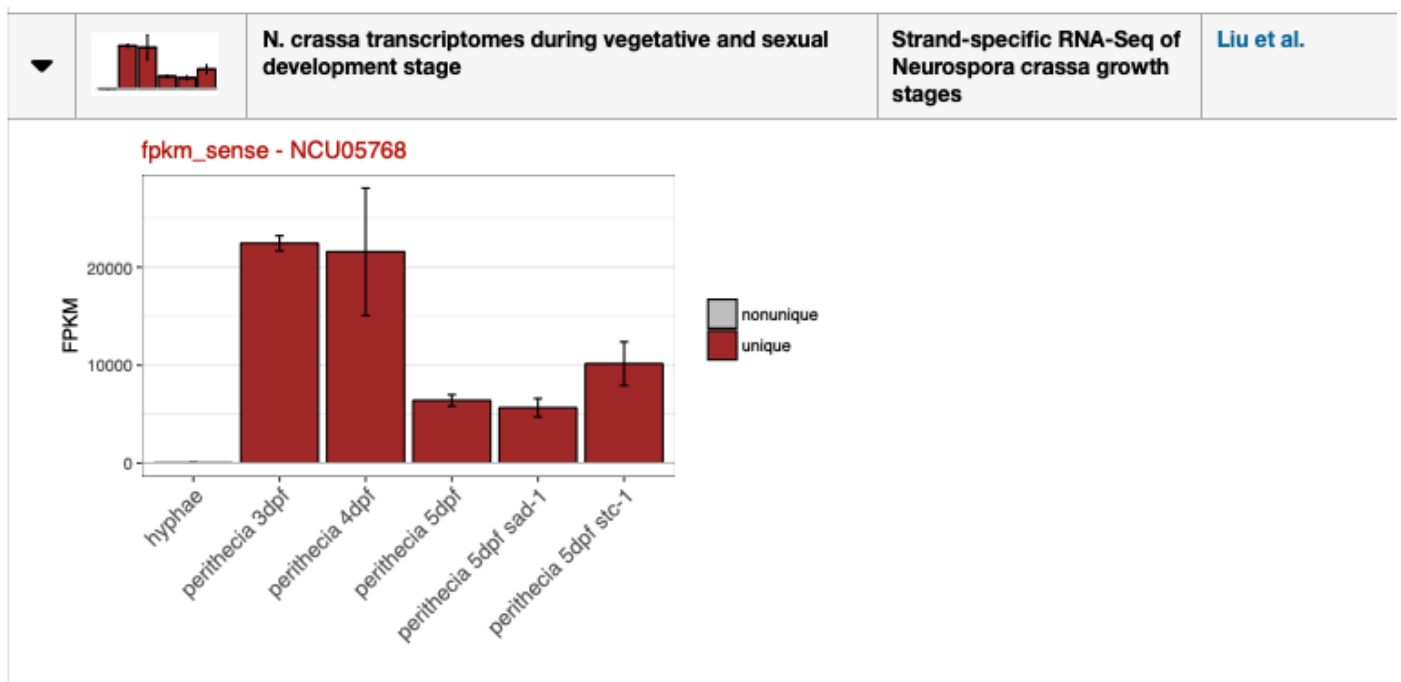
e. Se expresa el gen POI2?

Pista: Miren en la sección “Transcriptomics”. Experimento: *N. crassa* transcriptomics during vegetative and sexual development stage.

Se expresa?

Cual es la evidencia que soporta esta conclusión?

Se expresa en todos los estadios de desarrollo del hongo?



1.5. Usar la ortología para definir genes presentes en hongos que generan hifas y están ausentes en hongos que no las generan.

Un ejemplo de un organismo que genera hifas es *Candida albicans* y uno que no las genera es *Batrachomyces dendrobatidis*. Como se les ocurre que podemos buscar genes que están presentes en *C. albicans* y no tienen ortólogos en *B. dendrobatidis*.

Primero identificar todos los genes presentes en *C. albicans* (Pista: buscar genes por organismo en “Identify genes by”).

Search for Genes

expand all | collapse all

Find a search...

Text

Gene models

Annotation, curation and identifiers

Genomic Location

Taxonomy

Organism

Orthology and synten

Phenotype

Genetic variation

Transcriptomics

Sequence analysis

Structure analysis

Protein features and properties

Protein targeting and localization

Function prediction

Pathways and interactions

Proteomics

Immunology

expand all | collapse all

Identify Genes based on Organism

Organism

1 selected, out of 138

Filter list below...

Oomycetes

Fungi

Agaricomycetes

Blastocladiomycetes

Chytridiomycetes

Dothideomycetes

Eurotiomycetes

Glomeromycetes

Leotiomycetes

Pneumocystidiomycetes

Pucciniomycetes

Saccharomycetes

Candida

Candida albicans

Candida albicans SC5314

Candida albicans SC5314_B

Candida albicans WO-1

Candida auris

Candida glabrata

Candida parapsilosis

Candida tropicalis

Clavospora

Saccharomyces

Yarrowia

Schizosaccharomycetes

Sordariomycetes

Tremellomycetes

Combinar estos resultados con una búsqueda similar para *B. dendrobatidis*. Que operación debemos usar para combinar los resultados?

Add Step 2 : Organism

Organism select all | clear all | expand all | collapse all | reset to default

- ☐ Rhizopus
- ☐ Agaricomycetes
- ☒ Chytridiomycetes
 - ☒ Batrachochytrium dendrobatidis JEL423
- ☐ Eurotiomycetes
- ☐ Oomycetes
- ☐ Pucciniomycetes
- ☐ Saccharomycotina
- ☐ Sordariomycetes
- ☐ Taphrinomycotina
- ☐ Tremellomycetes
- ☐ Ustilaginomycetes

select all | clear all | expand all | collapse all | reset to default

☐ 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
☐ 1 Relative to 2, using genomic colocation

- Cuantos genes obtuvieron? Son todos los genes de *C. albicans*? Es correcto? O cometimos algún error?

Pista: para comprar genomas de diferentes especies hay que convertir uno de los genomas en sus ortólogos en el otro genoma. Por ejemplo, transformar los genes de *B. dendrobatidis* en sus ortólogos en *C. albicans*.

Clickear en nombre del paso y seleccionar “Orthologs” y luego seleccionar *C. albicans*.

- Click on the name of the step and select “Orthologs”, then select *C. albicans*.

(Genes)

Step 1: *C. albicans* 6446 Genes

Step 2: *B. dendrobatidis* 8794 Genes

STEP 2 : B. dendrobatidis

Rename | View | Revise | Make Nested Strategy | Insert Step Before | Orthologs | Delete

Organism : Chytridiomycetes, Batrachochytrium dendrobatidis JEL423

Results: 8794 Genes

☐ Give this search a weight

Transform by Orthology

Organism select all | clear all | expand all | collapse all | reset to default

- ☐ Rhizopus
- ☐ Agaricomycetes
- ☐ Chytridiomycetes
- ☐ Eurotiomycetes
- ☐ Oomycetes
- ☐ Pucciniomycetes
- ☐ Saccharomycotina
- ☒ Candida
 - ☒ Candida albicans SC5314
- ☐ Sordariomycetes
- ☐ Taphrinomycotina
- ☐ Tremellomycetes
- ☐ Ustilaginomycetes

select all | clear all | expand all | collapse all | reset to default

Syntenic Orthologs Only? ☐ no

☐ Give this search a name

My Strategies: New | Opened (1) | All (53) | Basket | Examples | Help

(Genes)

Step 1: *C. albicans* 6446 Genes

Step 2: *Orthologs* 2966 Genes

Organism(4)*

- Rename
- Duplicate
- Save As
- Share
- Delete

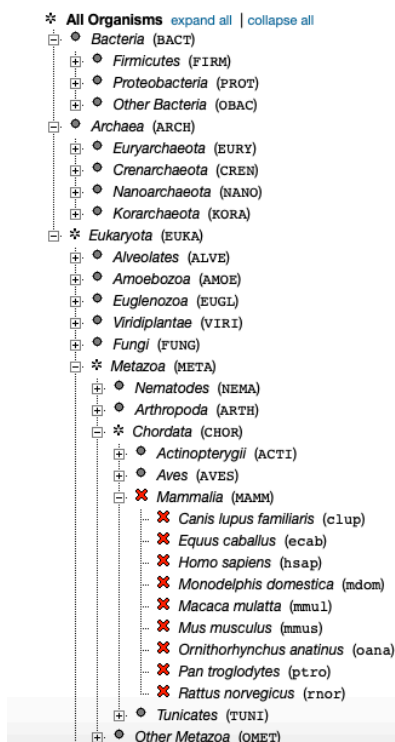
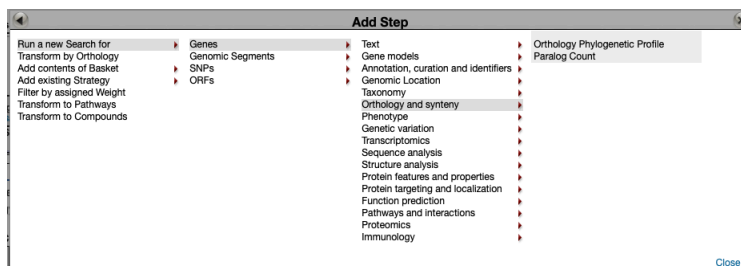
Análisis filogenético para definir genes de hongos que no tienen ortólogos en mamíferos

Vamos a buscar genes que sean posibles blancos para drogas. Una manera de resolver esto es buscando genes que no tienen ortólogos en mamíferos.

*Las **secuencias o genes ortólogos** son aquellas secuencias homólogas que se han separado por un evento de especiación. Es decir, cuando una especie diverge en dos especies separadas, se dice que las copias divergentes de un mismo gen en las especies resultantes son ortólogas. En otras palabras, las secuencias ortólogas son las secuencias que se encuentran en diferentes especies y que son altamente similares debido a que se han originado en un antepasado común.*

Usamos la opción de búsqueda: Orthology and synteny -> Orthology and phylogenetic Profile

Elegir genoma de *Candida albicans* SC5314 como en el ejercicio anterior y luego seleccionar en Perfil de ortología Mamíferos (doble click para que aparezca la cruz, con un click aparece una tilde que significa que queremos que los ortólogos de mamíferos estén incluidos).



-Cuántos genes encontraron?

-Alguno es una enzima? Vayan a Analyze results -> Metabolic Pathway enrichment

Encontrar un gen por similitud usando BLAST

En este caso vamos a usar la base de datos ToxoDB pero son las mismas herramientas que encontramos en el resto de las bases de datos compiladas en EuPathDB.

1- Imaginense que generaron un mutante insercional en *Toxoplasma* que les dió los resultados más interesantes de su carrera! Secuenciaron la región que flanquea el gen pero solo se pudo secuenciar un lado de la inserción (secuencia que se encuentra más abajo). Entonces van inmediatamente a ToxoDB para obtener información de esta secuencia. Que hacen?

```
aaaggagagaaagataaaaatatacaaaggtccccagagacacgatagtgttactgacaacatacagaatcaggtcgagcaatggaagaa  
ccaagcaccggcgccagagattgaactcgcttgattgccgtagcgtttatgagttgatagcttggtctaaaaaacaaggctgaaaaatgg  
aaaaaatgtctccaat
```

-Prueben usar la búsqueda por BLAST



- Que programa BLAST les conviene usar? Pueden probar diferentes combinaciones, pero recuerden que tienen una secuencia nucleotídica.

Machete programas BLAST:

- `blastp`: compara una secuencia aminoacídica contra la base de datos de secuencias de proteínas.

- **blastn**: compara una secuencia nucleotídica contra la base de datos de secuencias de nucleótidos.
- **blastx**: compara los productos traducidos teóricos de una secuencias nucleotídica en los seis marcos de lectura (en ambas hebras) contra la base de datos de secuencias de proteínas.
- **blastn**: compara una secuencia proteica contra la base de datos nucleotídica traducida dinamicamente en los seis marcos de lectura posibles (ambas hebras).
- **tblastx**: compara los productos traducidos teóricos de una secuencias nucleotídica en los seis marcos de lectura (en ambas hebras) contra la base de datos nucleotídica traducida en los seis marcos de lectura posibles (ambas hebras).

[illegible]

Pista: Luego de usar blastn contra las secuencias genómicas (Target Data Type) de *Toxoplasma* ME49 (organismo target) hagan click en “link to the genome browser”. En el Genome Browser hagan zoom hacia afuera para ver que gen esta en esa área.

a- Busquen la secuencia proteica de DHFR-TS y hagan BLAST contra todos los genomas:

- Primero hay que encontrar la secuencia, para esto busquen por palabra clave “dihydrofolato” (sin comillas) en **CryptoDB**.
- Van a obtener multiples resultados, encuentren la “dihydrofolato reductase-thymidylate synthase” (busquen en la columna Product Description).
- Una vez encuentran uno, hagan click en gene ID y vayan a la página del gen (mejor si lo abren en una ventana nueva).
- Vayan al final de la página donde esta la sección “Sequences”.

- Copien la secuencia aminoacídica y vuelvan a la página principal de **EuPathDB**. Vayan a la pagina BLAST (esta en la sección Tools).
- Usen los siguientes parámetros y copien la secuencia:

Target data Type: Proteins
 BLAST Program: blastp
 Targen Organism: Select All
 Input Sequence: Your pasted sequence
 Expectation value: 10
 Maximum description/alignments (V=B): 100
 Low complexity filter: no

-Click en “Get Answer”.

b- Basados en los resultados previos deben haber encontrado hits en casi todos los organismos patógenos de EuPathDB pero encontraron hits en *Giardia* o *Trichomonas*? Probemos otro método BLAST diferente:

- Vuelvan a la ventana BLAST. Cambien el “Target Data Type” a Genome.
- Seleccionen el Programa BLAST que corresponde, no pueden usar más blastp. Prueben con otras opciones. Tengan en cuenta que el tipo de “Input Sequence” debe cambiar cuando seleccionan un diferente programa BLAST.
- Seleccionen “All target organisms”- Click en “Get Answer”.

Fijense que aún no aparece la dihydrofolate de *Giardia* y *Trichomonas*. Prueben un método BLAST diferente.

- Vuelvan a la página del gen en **CryptoDB** y copien la secuencia nucleotídica.
- Vayan a la ventana de BLAST en **EupathDB** y peguen la secuencia nucleotídica en la ventana correspondiente.
- Seleccionen el “Target Data Type” (prueben diferentes) y el programa BLAST. Ojo que solo pueden seleccionar tblatx o tblastn si la secuencia input es nucleotídica.
- Seleccionen el organismo target. En este caso seleccionen específicamente *Giardia* o *Trichomonas*. Click en “Get Answer”.

Todavía no obtienen resultados? En el caso de *Giardia* es correcto! Este organismo no tiene dihydrofolate reductase o actividad thymidylate synthetase.