

Patrones Ortólogos y Filéticos (Ejercicio 8)

8.1 Acceder a OrthoMCL desde las bases de datos de EuPathDB

Nota: Para este ejercicio utiliza <http://microsporidiadb.org/micro/>

- Ve a la página del gen de *Encephalitozoon cuniculi* con el ID: ECU07_0290
- ¿Qué es lo que hace este gen? ¿Está anotado como hipotético!
- Ve hacia abajo de la página a la tabla etiquetada “Orthologs and Paralogs within MicrosporidiaDB”. ¿Tiene este gen ortólogos en otras especies de *Encephalitozoon*? ¿Y en otros organismos? (sugerencia: haz click en el link debajo de la tabla para llegar a OrthoMCL).

Orthologs and Paralogs within MicrosporidiaDB [Hide](#)

Gene	Organism	Product	is syntenic	has comments
ECU10_1730	<i>Encephalitozoon cuniculi</i> GB-M1	hypothetical protein	no	no
Ein07_0230	<i>Encephalitozoon intestinalis</i>	hypothetical protein	yes	no
Ein10_1700	<i>Encephalitozoon intestinalis</i>	hypothetical protein	no	no
EBI_25559	<i>Enterocytozoon bienewisi</i> H348	hypothetical protein	yes	no
EBI_27325	<i>Enterocytozoon bienewisi</i> H348	hypothetical protein	no	no

[Find the group containing ECU07_0290 in the OrthoMCL database](#)

- ¿Tiene esta proteína ortólogos en otros organismos? ¿Tiene ortólogos en bacteria o archaea? (sugerencia: mueve el mouse sobre los cuadros de colores en las tablas para revelar el nombre completo de la especie y el filum – mira la imagen abajo).

OrthoMCL DB
Ortholog Groups of Protein Sequences

[Home](#) [About OrthoMCL](#) [Data](#) [Search](#) [Tools](#)

Group Summary (OG4_10720)

# Sequences	# Taxa	# Match Pairs (%)	Ave E-Value	Ave % Coverage	Ave % Identity
118	86	6306 (91.3%)	4.03e-08	88.8	32

no ortholog; group details
 one ortholog; phyletic patterns
 more than one ortholog; pattern labels & counts

[Links to OG4_10720](#)
[Prim Domain Architecture](#) (circled in red)
[BioLayout Graph](#)
 No Multiple Seq. Alignment (too big)
[Get Sequences \(FASTA\)](#)

Mouse over these boxes

List of Protein Sequences

- e. Observa la arquitectura de dominios de PFAM - ¿Tienen todas las proteínas en este grupo similar arquitectura?
- f. Basado en los ortólogos, ¿qué piensas que hace esta proteína? ¿Cómo llamarías a este gen si tuvieras que darle un nombre?

8.2 Utilizando la herramienta para patrones filéticos en OrthoMCL

Nota: Para este ejercicio utiliza <http://www.orthomcl.org>

- a. ¿Cuántos grupos de proteínas en OrthoMCL no tienen ortólogos en bacteria o archaea? (sugerencia: ve hacia la búsqueda “by phyletic pattern”, esta puede accederse desde diversos lugares en OrthoMCL, mira la imagen de abajo)

The screenshot shows the OrthoMCL DB website interface. At the top, there is a navigation bar with links for Home, About OrthoMCL, Data, Search, and Tools. Below the navigation bar, there is a 'NEWS' section on the left with three entries: Oct 9, 2009, Sep 22, 2009, and Feb 28, 2008. The main content area features a 'Welcome to' message and statistics for the current release (Version 4, 138 Genomes, 110550 Protein Sequences, 110550 Ortholog Groups). There are two main search sections: 'Search for Groups' and 'Search for Sequences'. The 'Search for Groups' section has a dropdown menu with options: 'by IDs / Keyword / PFam', 'by Phyletic Pattern', 'by Phyletic Pattern - Advanced', 'by Group Properties', and 'Query History - Groups'. The 'Search for Sequences' section has options: 'by IDs / Keyword / Taxon / PFam', 'by BLAST Search', and 'Query History - Sequences'. A 'Tools' section at the bottom has one option: 'Assign your proteins to OrthoMCL groups'. Red arrows in the image point to the 'by Phyletic Pattern' option in both the dropdown menu and the search for groups list.

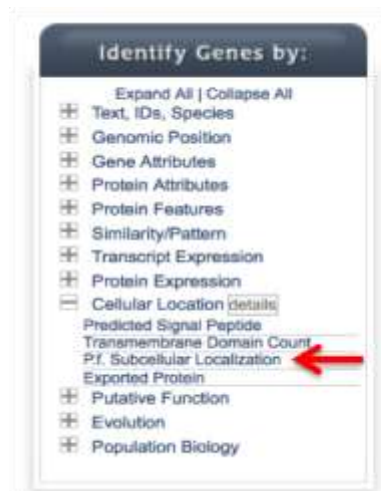
- b. ¿De los eucariotes, cuántos grupos de proteínas no contienen ortólogos? (sugerencia: Haz click en el ícono para especificar cuáles taxones o especies incluir o excluir en el perfil).

Key: ○ =no constraints | ✓ =must be in group | ✗ =must not be in group | ✓ =at least one subtaxon must be in group | ✳ =mixture of constraints

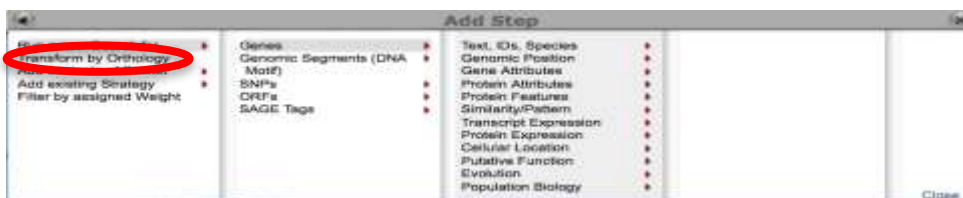
8.3 Utilizando la herramienta de transformación por ortología (“Transform by Orthology”) para identificar genes que codifican proteínas localizadas en el apicoplasto en *Toxoplasma* y *Neospora*.

Nota: Para este ejercicio utiliza <http://eupathdb.org>

- Comienza por encontrar los genes en *Plasmodium* que codifican proteínas cuya localización celular predicha es el apicoplasto (sugerencia: Haz click en “Cellular Location” y luego en “P.f. Subcellular Localization” – observa la imagen de abajo).



- Transforma los resultados de la búsqueda de arriba a los ortólogos de *Toxoplasma* (sugerencia: añade un paso, luego selecciona “Transform by Orthology”). En la búsqueda selecciona todas las cepas de *T. gondii* y *N. caninum*).



- Si bien *Cryptosporidium* es un parásito apicomplexan, ha perdido sus apicoplastos! ¿Puedes usar este hecho para refinar tus resultados con respecto a la búsqueda de arriba? (sugerencia: intenta abstraer cualquier

ortólogo presente en *Cryptosporidium* – necesitas utilizar una estrategia anidada).



8.4 (Opcional) Utilizando la herramienta de transformación por ortología para identificar el mayor número de genes de *P. falciparum* que contengan péptidos señal.

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

- ¿Cuántos genes en *P. falciparum* están anotados como péptidos señal (utiliza la configuración standard de la búsqueda en “Genes based on Predicted Signal Peptide”)?
- ¿Cuántos genes en *P. vivax* están anotados como péptidos señal (utiliza la configuración standard)?
- ¿Cuántos genes son comunes a estas dos listas? (sugerencia: utiliza la búsqueda de ortólogos para transformar entre organismos)
- ¿Cuántos ortólogos de *P. falciparum* en los genes de *vivax* con péptidos señal no contienen en ellos mismos péptidos señal? ¿Por qué ocurre esto? Observa un par de ellos utilizando el “synteny viewer” para generar algunas hipótesis.
- Utilizando PlasmoDB, genera la lista más amplia de genes de *P. falciparum* que puedan contener péptidos señal. ¿Cuántos encontraste?

8.5 (Opcional) Búsquedas integradas en OrthoMCL

Nota: Para este ejercicio utiliza <http://www.orthomcl.org>

Encuentra todas las proteínas de plantas que probablemente sean fosfatasa y que no tengan ortólogos fuera de las mismas plantas.

- Utiliza la búsqueda por palabras (keywords) para encontrar grupos que contengan la palabra “fosfatasa” (“phosphatase”)
- Ejecuta una búsqueda de patrón filético para grupos que contengan cualquier proteína en plantas pero que no contengan ningún otro organismo fuera de las plantas. (sugerencia: asegúrate de que todo tenga una x roja excepto las plantas, que deberán tener un círculo gris).

- Ahora, necesitas combinar las dos búsquedas de arriba con una intersección. (sugerencia: bajo el menú de búsqueda va hacia "Query History – Groups", selecciona los dos resultados y haz click en "intersect")
- ¿Cuántos grupos obtuviste? Si hay tiempo explora los alineamientos múltiples de secuencias de algunos de estos grupos.