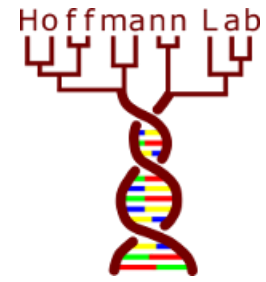




FACULTAD DE
CIENCIAS

UDELAR | fcien.edu.uy



Evolución de familias multigénicas 2020 Guía Proyecto Fin de Curso

Algunas consideraciones generales. El objetivo del trabajo de fin de curso no es el de hacer un trabajo de nivel de publicación, sino el de hacer un ejercicio donde se utiliza lo aprendido en el curso. Entonces, lo mejor es restringir el foco a familias multigénicas pequeñas de vertebrados. Está es el área donde los docentes nos sentimos más cómodos y podemos aportar más. Dicho esto, si tienen un interés claro por otro grupo de organismos, adelante. En este sentido conviene recordar que las duplicaciones genómicas son extremadamente frecuentes en plantas, y que las bacterias y arqueas son extremadamente promiscuas a la hora de pasarse genes entre organismos bastante distantes. Estas cosas son mucho menos frecuentes en los animales. En todo caso, siempre hay que usar los genomas más sencillos y de mayor calidad posible para empezar, y utilizar el menor número de ellos posibles para cubrir un grupo. En el caso de los vertebrados mandibulados, con comparar peces cartilagosos, el gar (pez lagarto, *Lepitosteus oculatus*), humano y gallina ya tenemos un muestreo razonable como para empezar.

Ensembl tiene bases de datos para un gran número de grupos, de manera que siempre es el lugar para empezar a explorar la variación en las familias multigénicas. Por ejemplo, si elegimos el gen BRCA2 de humanos, que está asociado al cáncer de mama, vemos que tiene ortólogos, pero no parálogos, de manera que no es un buen candidato.

Location: 13:32,315,086-32,400,266 Gene: BRCA2

Gene-based displays

- [-] Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- [-] Sequence
 - Secondary Structure
- [-] Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
 - Ensembl protein families

Gene: BRCA2

Description

Gene Synonyms

Location

About this gene

Transcripts

Tengan en cuenta que las anotaciones de Ensembl no tienen un alto nivel de curación, por lo que tienen que ser tomadas con pinzas. Aquí tienen un ejemplo de un gen con parálogos. El gen vecino a [BRCA2](#) se llama [ZARL1](#) y tiene un parólogo.

Location: 13:32,555,952-32,641,134 Gene: ZAR1L

Gene-based displays

- [-] Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- [-] Sequence
 - Secondary Structure
- [-] Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues**
 - Paralogues
 - Ensembl protein families

Gene: ZAR

Description

Gene Synonyms


Location

About this gene

Transcripts

Ensembl anota a su parólogo como un derivado de una duplicación en los gnatostomados.

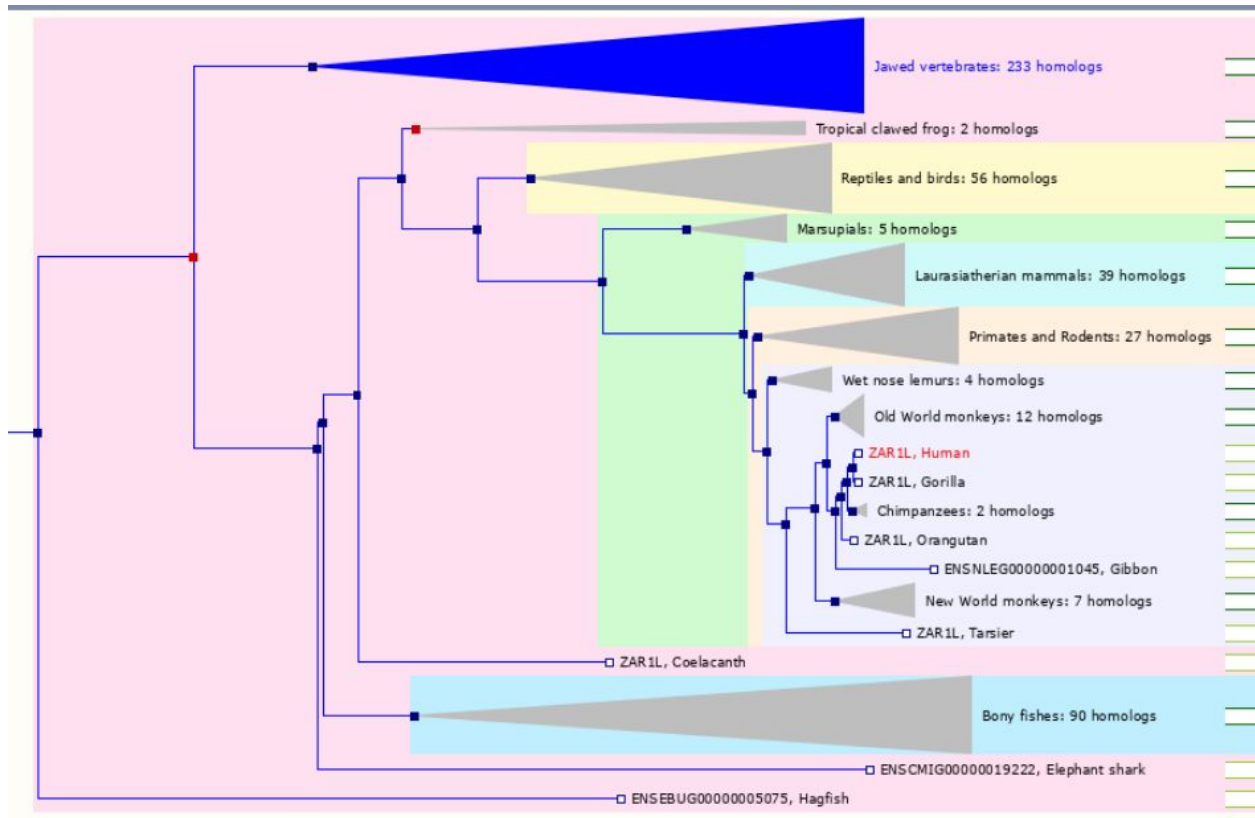
Paralogues

 Download paralogues

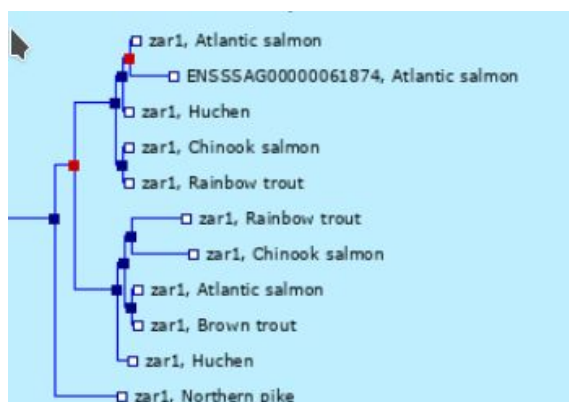
Show/hide columns

Type	Ancestral taxonomy	Ensembl identifier & gene name
Paralogues	Jawed vertebrates (Gnathostomata)	ENSG00000182223 ZAR1 zygote arrest 1 [Source:HGNC Symbol;Acc:HGNC:20436]

Si van al árbol de genes de [ZARL1 y ZARL1](#) correspondiente pueden ver que la filogenia sugiere que esto es correcto. Este árbol sugiere la presencia de parálogos adicionales en el Tropical clawed frog (*Xenopus tropicalis*)



Si abrimos el árbol del ZAR1, en azul, vemos que hay una duplicación el ZAR1 de los salmones, probablemente debida a la duplicación genómica en ese grupo.



Esta predicción se puede testear explorando la sintenia, dado que las duplicaciones genómicas afectan a todo el genoma, y entonces debería haber otros genes que se

han co-duplicado. Las duplicaciones sencillas afectan a un solo gen, de manera que no debería haber otros genes duplicados con ellos.

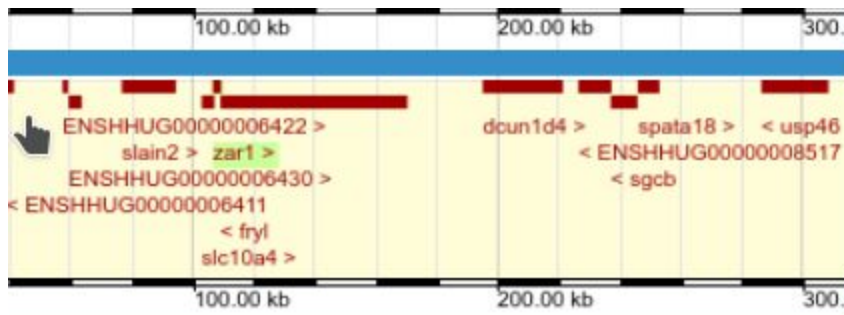
Para hacer esto, lo primero es explorar el contexto genómico de estos genes. Veamos a estos dos parálogos de ZAR1 en el genoma del [huchen](#) (un salmónido). Los dos genes están anotados como ZAR1. Como eso lleva a confusión, le pondremos a un ZAR1a ([ENSHHUG00000038706](#)) y al segundo ZAR1b ([ENSHHUG00000006535](#))

Si comparamos sus contextos genómicos vemos que los genes fryl, slain2, y slc10a4 se han co-duplicado junto a ZAR1a y ZAR1b, de manera que los datos sugieren que estos parálogos derivan de la duplicación genómica de los salmones.

ZAR1a



ZAR1b



Por razones que desconozco, coleccionar datos con BLAST es más sencillo en NCBI, mientras que explorar relaciones de ortología y paralogía es mucho más sencillo en Ensembl.

Aquí va una lista familias multigénicas de animales que merecen atención (Disclosure: hemos publicado en muchas de ellas).

1. Myb (mioglobinas)
2. Relaxinas y sus receptores
3. Gap Junction Alfas y Betas
4. RAFs de Animales
5. Nodal
6. Insulina y sus receptores (INSL)
7. PIWIs y Argonautas de vertebrados (estas dos van juntas)
8. Luc7L
9. HIF

Coviene recordar que los términos parólogo y ortólogo son relativos, ya que describen relaciones entre genes por lo que la elección de genomas tiene que ser hecha con criterio de cubrir los nodos de interés. Aquí tienen una discusión extensa sobre el tema: <https://pubmed.ncbi.nlm.nih.gov/16285863/>.

En vertebrados, tenemos que con los genomas de los ciclostomados (lampreas y peces bruja o hagfish en inglés), peces cartilaginosos (tiburones, rayas, chuchos, quimeras, peces guitarra y afines), más unos pocos representantes de los peces óseos u [osteíctios](#), donde están los actinopterigios y sarcopterigios, alcanzan para cubrir el grupo muy bien. La estructura de los genomas de vertebrados es conservada en peces óseos y cartilaginosos, por lo que comparaciones entre el pez lagarto y el tiburón son bien informativas y convenientes ya que estos genomas parecen no haber cambiado demasiado a nivel estructural. Los mamíferos somos más complicados. Si agregamos los tunicados y cefalocordados como grupos externos, ya tenemos un buen marco de análisis.

Como dijimos más arriba, conviene usar los genomas más sencillos y de mayor calidad posible para empezar. Aquí abajo les muestro la selección que combina estas dos variables para estudiar a una familia multigénica en vertebrados como para una publicación.

Espece	Clasificación	Notas
<i>Homo sapiens</i>	Mamifero placentario	
<i>Monodelphis domestica</i>	Marsupial	
<i>Ornithorhynchus anatinus</i>	Prototeria	
<i>Alligator mississippiensis</i>	Cocodrilo	Este es el mejor genoma, pero en ensembl está el cocodrilo australiano de agua salada
<i>Crocodylus porosus</i>	Cocodrilo	En ensembl. Pero de menor calidad.
<i>Gallus gallus</i>	Ave	
<i>Pelodiscus sinensis</i>	Tortuga	
<i>Anolis carolinensis</i>	Lagartija	
<i>Sphenodon punctatus</i>	Rincocefalio	
<i>Xenopus tropicalis</i>	Anfibio	
<i>Latimeria chalumnae</i>	Sarcopterigio	
<i>Lepisosteus oculatus</i>	Holósteo	
<i>Oryzias latipes</i>	Teleósteo	
<i>Callorhynchus milii</i>	Condrictio	Este genoma es razonable, pero no excelente.
<i>Rhincodon typus</i>	Condrictio	Tiburón ballena. Su genoma está bastante bien. No está en Ensembl. Animal muy carismático!
<i>Petromyzon marinus</i>	Ciclostomado	Lamprea
<i>Eptatretus burgeri</i>	Ciclostomado	Pez bruja o hagfish en inglés.
<i>Ciona savignyi</i>	Urocordado	Ascidio
<i>Ciona intestinalis</i>	Urocordado	Ascidio
<i>Branchiostoma lanceolatum</i>	Cefalocordado	Anfioxo

Claramente, conviene ajustar el muestreo según la pregunta a contestar. Si queremos trabajar con mamíferos, debemos incrementar el muestro de estos, y probablemente restringirnos a unos pocos linajes amniotas adicionales. En el caso de los peces teleósteos, hay que recordar que han pasado por una duplicación genómica extra relativa a su grupo hermano, que son los Holósteos, y que algunos teleósteos como los salmones han pasado por duplicaciones genómicas adicionales. A menos que tengan un fuerte interés por los teleósteos, puede ser conveniente evitar esos casos

complejos, que tienen interés en sí mismo pero complican si el objetivo es estudiar otros vertebrados.