

Guía Práctico

Estudio de familias génicas

Curso: Genómica evolutiva

14 de junio, 2019

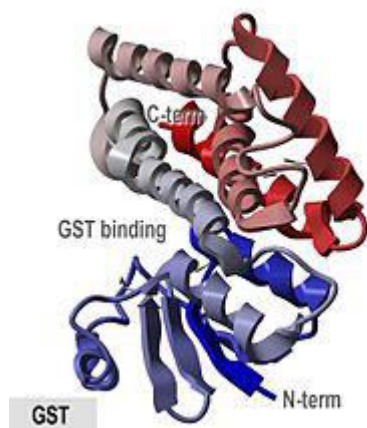
Introducción a conceptos:

Una familia génica está formada por dos o más loci con secuencias nucleotídicas similares que han derivado de una secuencia ancestral común. Una familia se forma como consecuencia de los procesos de especiación y duplicación de genes (por errores en la recombinación homóloga, retrotransposición y duplicación cromosómica).

Los genes homólogos, que forman estas familias, pueden ser ortólogos (derivados por procesos de especiación) o parálogos (derivados de procesos de duplicación). La duplicación es un fenómeno frecuente en todos los organismos vivos y es relevante en la formación de nuevos genes y funciones. Luego de una duplicación el destino de la nueva copia génica es en general la pérdida (o silenciamiento), aunque también puede retenerse, ya sea manteniendo su función “ancestral” (redundancia), divergiendo hasta adquirir una nueva función (neofuncionalización) o reajustando su función ancestral, ahora dividida en las dos nuevas funciones (subfuncionalización).

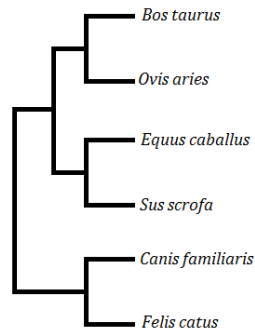
Existen varios modelos que intentan explicar cómo se da la evolución de las familias génicas, algunos modelos se enmarcan en una visión neutralista y otras en una visión seleccionista de la evolución molecular. Discutiendo cual es el aporte relativo de la selección positiva en el destino de estas nuevas copias.

En este práctico utilizaremos el análisis filogenético como herramienta principal para estudiar la evolución de la familia génica de las Glutación S-Transferasas (GSTs) en 6 especies de mamíferos domésticos. Tienen funciones muy variadas aunque son mejor conocidas como isozimas metabólicas de fase II, capases para catalizar la conjugación de la forma reducida de glutatión a sustratos xenobióticos con el propósito de desintoxicación. La (Super-) familia de las GSTs se divide en clases, siendo mu, omega, sigma, kappa, alpha, zeta, algunas de ellas. Generalmente forman homodímeros, aunque se han descrito casos de heterodímeros. No todas están presentes en todos los organismos, pero algunas se han conservado desde bacterias a humanos (e.j. omega). La identidad aminoacídica es relativamente baja entre clases pero la similitud estructural 2D y 3D considerable.



Aplicaremos una estrategia sencilla para el análisis filogenético, incluyendo traducción de secuencias, estimación de la diversidad aminoacídica, alineamiento de secuencias, que correremos en programas con funciones específicas en línea de comandos.

En principio trabajaremos con seis especies del superorden Laurasiatheria (*Bos taurus*, *Ovis aries*, *Canis familiaris*, *Felis catus*, *Equus caballus* y *Sus scrofa*).



ACTIVIDADES.

Antes de comenzar, conteste:

- ¿Por qué cree usted que el tema de familias multigénicas se entiende dentro del área de la genómica comparativa?

Dentro de **home**, cree un directorio con su nombre, ese será su directorio de trabajo.

- `$ mkdir nombre_apellido_familias`
- `$ cd nombre_apellido_familias`

Los miembros de las familias fueron identificados mediante blastp contra las proteínas anotadas en los genomas publicados en la base de datos del Ensembl (www.ensembl.org). Las secuencias fueron clasificadas en clases de acuerdo a secuencias consenso que definen las clases. Las especies tienen las referencias en las secuencias: Bt (*Bos taurus*), Ov (*Ovis aries*), Cf (*Canis familiaris*), Fc (*Felis catus*), Ec (*Equus caballus*) y Ss (*Sus scrofa*).

1. En caso que fuera necesario, como podría chequear que todas las secuencias obtenidas pertenecen a la familia GSTs
2. Copiar a su directorio de trabajo los 5 archivos multifasta, uno representando cada clase a estudiar: alfa, kappa, omega, eef1g y mu, ubicados en el directorio MATERIAL-flias.

`$ cp ../MATERIAL-flias/*.`

3. Visualizar los archivos.
 - Puede decir algo desde el punto de vista biológico respecto al número de secuencias por organismo en los distintos archivos?
 - ¿Cómo se estudiaría los grupo y ¿Cómo formularía una hipótesis de trabajo?

`$ less alpha.fas`

***Donde dice “alpha” deben figurar cada uno de los archivos de acuerdo a la clase que esta trabajando.**

4. Antes de comparar secuencias es necesario traducirlas y alinearlas, para esto utilizaremos el programa “transeq” del paquete Emboss y el programa “muscle”,

respectivamente. Luego estimaremos la distancia aminoacídica promedio utilizando el comando distmat del paquete Emboss.

```
$ transeq alpha.fas alpha.aa.fas
```

```
$ sed -i 's/_1$/g' alpha.aa.fas
```

```
$ muscle -in alpha.aa.fas -out alpha.aa.aln.fas
```

```
$ distmat alpha.aa.aln.fas -protmethod 2 alpha.aa.aln.mat
```

Copie los resultados y calcule las distancias promedio dentro de cada clase.

- ¿Qué puede interpretar de los resultados?
 - ¿Cómo es la variabilidad de una clase respecto a la otra?
 - Busque en NCBI mediante blastp información sobre la función de las proteínas “eef1g”. ¿Qué relación puede tener esto con su variabilidad, tanto en presencia como a nivel de secuencias?
5. Realizaremos una filogenia por cada clase siguiendo dos pasos sencillos: 1) estimar el modelo evolutivo más adecuado para cada clase utilizando el programa modelgenerator escrito en java y 2) construir la filogenia utilizando el método de Máxima Verosimilitud utilizando el programa Phym1.
- a. modelo evolutivo

```
$ cp ../MATERIAL-flías/modelgenerator.jar .
```

```
$ java -jar -Xmx2000m modelgenerator.jar alpha.aa.aln.fas 8
```

- ¿Cuál fue el modelo elegido en cada clase? Busque información sobre alguno de ellos.
- b. Muchos de los modelos estimados por el programa modelgenerator pueden ser incluidos en el programa. Consulte el manual para saber cómo pasar esta información al programa phym1. (Si resulta muy difícil y no le quiere dedicar más tiempo puede proseguir con el siguiente punto).

```
$ seqret alpha.aa.aln.fas -osformat phylip3 alpha.aa.aln.phy
```

- ¿Qué diferencia hay entre un formato fasta y un phylip?

```
$ phym1 -i alpha.aa.aln.phy -d aa --sequential --no_memory_check -b -4 -s BEST -v e -a e -c 8
```

- ¿Qué opción estamos usando para estimar el apoyo estadístico de los nodos? (ver manual online o “phym1 –help”)

6. El archivo con el árbol filogenético generado por phym1 se encuentra en la misma carpeta, está en formato newick y tiene el nombre del archivo de entrada más la extensión “_phym1_tree.txt”. En este caso alpha.aa.aln.phy_phym1_tree.txt.
 - a. Copie la información del archivo con el mouse y visualice el mismo en la página www.phylogeny.fr utilizando el programa TreeDyn.
 - b. Ajuste el formato utilizando el menú de opciones abajo.
 - c. Guarde la imagen generada en un documento para contestar las siguientes preguntas.
 - d. Envíe a airiardeo@gmail.com como documento adjunto sus árboles filogenéticos y conteste las siguientes preguntas. Los resultados serán discutidos la próxima clase.

Conteste las siguientes preguntas en relación a los resultados obtenidos:

1. ¿Cómo explica las diferencias en el número de miembros de las distintas clases entre estos organismos cercanamente relacionados? Discuta los resultados de las distintas clases.
2. ¿Qué le sugiere este resultado en cada clase en relación a la dinámica evolutiva?
3. ¿Por qué puede ser importante tener el árbol robusto (estadísticamente) de las especies que estamos trabajando?
4. Estudie los árboles generados e identifique los nodos que representen eventos de especiación y eventos de duplicación.
5. Identifique eventos de duplicación reciente, linaje-específicos y especie-específicos.
6. ¿Puede en algún caso identificar ortólogos?
7. ¿Cómo explica las observaciones en relación a los procesos evolutivos, deriva, mutación, selección?
8. Busque información en relación a las distintas clases estudiadas. ¿Qué interpretación biológica puede dar a la dinámica evolutiva observada en cada clase dentro de esta familia génica. Compare lo que observa en una clase con respecto a las otras.
9. ¿Qué ventaja adaptativa puede tener el hecho de aumentar o reducir el número de copias de una clase específica de GSTs?