

CLASES PRÁCTICA DE ANALISIS DE VARIABILIDAD NUCLEOTÍDICA.  
**DOMESTICACIÓN EN CERDO: EL COLOR DEL PELAJE.**

Sebastián E. Ramos Onsins  
Centre of Research in Agricultural Genomics (CRAG)

### **Introducción**

El color del pelaje, así como la forma y el tamaño de los animales son características que se han tenido en cuenta en la cría de animales domésticos. En concreto, el color del pelaje varía en mayor grado en los animales domésticos que en los animales salvajes, posiblemente porque en los animales salvajes es una característica importante para mantener el mimetismo con el ambiente, y por tanto su supervivencia.

En el caso del cerdo (*Sus scrofa*), los jabalíes mantienen un color marrón oscuro que les permite mimetizarse en su ambiente, mientras las especies de cerdo domesticadas tienen colores variados: blanco, negro, rojo, diferentes tonos así como diferentes moteados o rayados.

El color del pelaje es un carácter determinado por diversos factores. Uno de ellos es el receptor *MC1R* (*melanocortin receptor 1*), un receptor de proteína-G que se expresa en melanocitos y que juega un papel fundamental en la determinación del color. En concreto, la presencia de este receptor induce la síntesis de eumelanina, asociado a un color oscuro, mientras que su ausencia solo permite expresar feomelanina, asociado a colores rojo/amarillos. La supresión o sobre-expresión de este receptor induce cambios en el color del animal.

En esta práctica se estudiará la variabilidad del DNA en el locus *MC1R* y su asociación con la variabilidad en el color de cerdos domésticos y salvajes. Para ello estudiaremos la variabilidad en cada grupo por separado y las

diferencias entre grupos. Después deduciremos la historia de los grupos y estudiaremos la asociación del fenotipo de coloración con la variabilidad en el locus *MC1R*. Las tres primeras prácticas están basadas en el artículo publicado en PLoS Genetics, Jan2009, Vol 5, e1000341.

### **PRÁCTICA 1-3.**

Se han recogido de GenBank las secuencias codificadoras del locus *MC1R* en 83 individuos, más una secuencia referencia con la anotación (***MC1R\_Pigs\_single.db.txt***). Estas secuencias se han alineado utilizando el software *muscle* y se han grabado en un fichero en formato FASTA con el nombre ***MC1R\_Pigs\_aligned.fas***. Los datos de razas y fenotipos (color) se encuentra en el fichero ***Pigmentation\_and\_breeds.pdf***.

En esta práctica, el OBJETIVO es calcular la variabilidad a nivel de DNA de cada una de las poblaciones, y determinar las diferencias entre poblaciones. Para ello estudiaremos las mutaciones que determinan cambios aminoacídicos (mutaciones no-sinónimas o de reemplazamiento) y las mutaciones que no producen cambios aminoacídicos (mutaciones sinónimas).

### **PASOS O EJERCICIOS A REALIZAR EN LA PRÁCTICA 1:**

1. Localizar el fichero ***MC1R\_Pigs\_aligned.fas*** y abrirlo con un editor de texto. Reconocer el formato FASTA y observar los nombre de las secuencias de los individuos.

Este fichero contiene 83x2 (este locus se encuentra en un autosoma, y por tanto hay dos copias en este organismo diploide) + 1 (secuencia de referencia) = 167 secuencias en total de los genotipos de todos los individuos muestreados. Las secuencias contienen nombres que definen diferentes grupos: EUROPA LOCAL, EUROPA COMERCIAL, EUROPA WILD BOAR, ASIA LOCAL y ASIA WILD BOAR. La razón de la separación entre Europa y

Asia es que estas poblaciones han divergido hace aproximadamente 500.000 años, y por tanto contienen historias diferentes (recordar que la domesticación ocurrió en el Neolítico, no más de hace 15.000 años).

2. Abriremos el programa **DnaSP** (<http://www.ub.edu/dnasp>). Este programa contiene un interfaz por ventanas y permite hacer un gran número de análisis de variabilidad en poblaciones. Aprenderemos a utilizar este software en algunas de sus opciones.

3. Observar el interfaz. Botones de acción (Abrir fichero, Cerrar, Grabar, Imprimir, Gráficos, Información, Mostrar Datos, Incluir/Excluir secuencias, Definir Grupos de secuencias, Simulaciones de Coalescencia, Calculadora Evolutiva, Ayuda). Entender los menús desplegados en la parte superior (File, Data, Analysis, Overview, Generate, Tools, Window, Help). Es de gran interés el menú de Ayuda (*Help*).

4. Abrir el fichero y visualizar los datos mediante *Display:View Data*. Buscar las posiciones variables mediante una opción en la ventana.

-Observar el nombre de las líneas. Cada individuo tiene dos secuencias, la segunda con un "2" indicando la segunda secuencia.

5. Desplegar el Menú *Data* y estudiar las diferentes opciones. Vamos a definir la región codificadora, el código genético, los diferentes grupos de individuos, y sus características (autosoma, DNA, diploide).

5.1 Escoger las opciones apropiadas en *Data:Format*

5.2 Definir las regiones codificadoras: **ATENCIÓN. Se han de definir las regiones codificadoras teniendo en cuenta la secuencia REFERENCIA (la inserción-delección en las posiciones 62-63 no se encuentran en la secuencia referencia, y no deben ser anotadas como exones! [es necesario para un análisis comparativo]).** El

**codón STOP tampoco debe ser anotado como codificador (posiciones 963-965).**

5.3 Definir el código genético (Universal).

5.4. Anotación de los grupos de poblaciones mediante *Data:Define Sequence Sets*: Definir

EUROPALOCAL

EUROPACOMMERCIAL

EUROPAWILDBOAR

ASIADOMESTIC

ASIAWILDBOAR

y también

EUROPADOMESTIC

WILDBOARALL

DOMESTICALL

6. Grabar nuevo fichero con formato nexus para conservar todas las características incluidas. Grabarlo como ***MC1R\_Pigs\_aligned.nex***. Abrirlo en un editor de texto y observar la estructura del formato NEXUS.

7. Obtener ficheros con la variabilidad nucleotídica para el total de posiciones y para las posiciones sinónimas y no-sinónimas, y los haplotipos existentes mediante las opciones *Generate:Polymorphic/Variable sites Data File* y la opción *Generate:Haplotype Data File*. Observar las mutaciones en los diferentes grupos.

- Observar los ficheros en un editor de textos.

**- Fijaros que las secuencia EUROPALOCALCR0112 contiene un haplotipo solo presente en la poblacion de ASIA LOCAL, y que la secuencia ASIAWILDBOARDW12 contiene un haplotipo solo presente en Europa. Estas dos secuencias son muy probablemente producto de reciente introgresión (de Asia al cerdo criollo francés en un caso, y de cerdo doméstico en jabalí asiático en el otro). Por tanto eliminaremos**

estas dos secuencias del análisis (mediante *Data:Include/Exclude Sequences*).

## PASOS O EJERCICIOS A REALIZAR EN LA PRÁCTICA 2:

Una vez tenemos las poblaciones definidas, estimaremos la variabilidad nucleotídica. Existen varias maneras de calcular la variabilidad: se puede calcular a partir de una estimación de la **heterozigosidad** esperada mediante la comparación de todos los linajes entre sí, hacer el promedio y dividirlo por el número total de posiciones. También se puede calcular a partir de supuestos bajo el modelo estacionario de Wright-Fisher, que asume no migración, no selección, panmixia, igual tamaño poblacional a lo largo del tiempo pero permite mutación y deriva genética. Bajo estas condiciones se calcula la variabilidad mediante el estadístico de Watterson *theta* ( $\theta$ , que es un valor proporcional al número observado de mutaciones y ponderado por el número de muestras estudiadas).

1. Abrir el programa **DnaSP**. Recordatorio de los menús principales.
2. Abrir el fichero **MC1R\_Pigs\_aligned.nex**. Hacer un *Display: View Data* y *Display:Data Info* para comprobar que el fichero contiene las características definidas en la práctica anterior.
- 2a. HACER EL PUNTO 7 de la PRÁCTICA 1.
3. Estudiar la variabilidad. Primera aproximación con una visión general: hacer *Overview:Polymorphism Data* y escoger todas las secuencias. Entender el fichero de resultados. (*Number of Variable Sites, Number of Haplotypes, Haplotype diversity, Nucleotide diversity, Theta*)

4. Identificar las posiciones polimórficas: *Analysis:Polymorphic Sites*. Utilizar cada una de las poblaciones EUROPADOMESTIC, EURWILDBOAR, ASIADOMESTIC y ASIAWILDBOAR por separado.

- Apuntar el número de variantes totales, sinónimas y no-sinónimas.

Contrastar alguno de los resultados mediante *Display:View Data:Select Sites: Syn/NonSyn*.

5. Calcular la variabilidad por nucleótido en posiciones sinónimas y no-sinónimas para cada una de las poblaciones por separado. Para ello, utiliza *Analysis:Polymorphism and Divergence: Syn/NonSyn*.

- Mira el nivel variabilidad en cada grupo. ¿Cuál es más alta, la variabilidad sinónima o la no-sinónima?

6. Calcular las diferencias entre las poblaciones. Es de interés observar las diferencias entre domesticadas y salvajes en cada uno de las regiones. Es decir, comparar jabalés versus Domesticados en Asia y también en Europa (EUROPADOMESTIC vs EUROPAWILDBOAR, ASIADOMESTIC vs ASIAWILDBOAR, EUROPADOMESTIC vs ASIADOMESTIC y EUROPAWILDBOAR vs ASIAWILDBOAR). Realizar este análisis mediante la opción *Analysis:DNA Divergence Between Populations*.

- Calcular el número de mutaciones fijadas, exclusivas y compartidas.

- Calcular la variabilidad dentro y entre poblaciones.

7. Interpretación

### **PASOS O EJERCICIOS A REALIZAR EN LA PRÁCTICA 3:**

En esta práctica vamos a aprender a construir una red de haplotipos para una región de DNA (o gen). Esta red de haplotipos tiene una mejor interpretación para regiones en las que no exista recombinación.

En este caso utilizaremos, como en las anteriores prácticas, el locus *MC1R* y reconstruiremos todos los haplotipos y las relaciones que hay entre ellos. Después, asignaremos el origen (Asia o Europa), el tipo de población (doméstico o salvaje) y el color de la capa e intentaremos ver si existe algún patrón.

1. La reconstrucción de haplotipos la realizaremos con el software **Network** (<http://www.fluxus-engineering.com/sharenet.htm>). Este programa calcula la relación entre haplotipos mediante diferentes algoritmos basados en Máxima Parsimonia. En el caso que no esté instalado este software bajar el programa e instalarlo. Clicar dos veces sobre el icono Network para abrir el programa.
2. En el manual accesible desde su página web se encuentran las instrucciones de uso de este software. Nosotros vamos a realizar un análisis sencillo. En primer lugar necesitamos el fichero de entrada de nuestros datos (extensión *rdf*). Este fichero se puede construir a partir del fichero de formato *nexus*. El programa podría ser capaz de importar ficheros *nexus*, pero en muy particulares opciones. En la práctica hay que construir el fichero *.rdf* a partir de un fichero *nexus* que contenga solo los polimorfismos (ejemplo: a partir del fichero *MC1R\_PigsOutg\_aligned\_pol.nex*).
3. El fichero *MC1R\_PigsOutg\_aligned\_pol.rdf* contiene los linajes a estudiar y también una secuencia de una especie cercana (*Potamochoerus porcus*). Este fichero tiene un formato particular. Abrir el fichero con un editor de texto para verlo.
4. Una vez construido el fichero, es posible añadir datos sobre la geografía, fenotipo y otros datos de la muestra. En el fichero *MC1R\_PigsOutg\_aligned\_nointrg\_pol.rdf* se ha añadido la información desde el programa **Network**:
  - a. Abrir Network
  - b. Data Entry: Import rdf file
  - c. Seleccionar fichero

- d. Visualizar datos
  - e. Clicar *Switch Input Window*
  - f. Salir mediante Exit
5. Dos algoritmos de reconstrucciones se han incluido en el programa:
- a. Median Joining (MJ). Datos multiestado o binarios. Une los estados observados mediante nodos intermedios con los estados “ancestrales”. La red será mas completa cuanto mayor sea el valor del parámetro *epsilon*, pero también más difícil de interpretar. Empezar por defecto y aumentar el valor es la sugerencia de los autores.
  - b. Reduced Median (RM). Para datos binarios. Red reducida para clarificar los resultados. Al igual que MJ, la red será mas completa cuanto mayor sea el valor del parámetro *r*, pero también más difícil de interpretar. Empezar por defecto y aumentar el valor es la sugerencia de los autores.
6. Calcular la red con los dos algoritmos:
- a. *Calculate Network: Network Calculations: Reduced Median*
    - i. *File open: MC1R\_PigsOutg\_aligned\_nointrg\_pol.rdf*
    - ii. *Calculation:* Nombre del fichero:
      - 1. *MC1R\_PigsOutg\_aligned\_nointrg\_pol\_RM.rmf*
      - 2. *MC1R\_PigsOutg\_aligned\_nointrg\_pol\_RM.out*
  - b. *Calculate Network: Network Calculations: Median Joining*
    - i. *File open: MC1R\_PigsOutg\_aligned\_nointrg\_pol.rdf*
    - ii. *Calculation:* Nombre del fichero:
      - 1. *MC1R\_PigsOutg\_aligned\_nointrg\_pol\_MJ.out*
7. Una vez calculada la red de relaciones entre haplotipos, procedemos a visualizarla. Clicar en *Draw Network*.
- a. *File: Open:* Buscar los ficheros *.out*
  - b. Abrir el generado por MJ primero.
  - c. Observar el resultado.
8. Introducir colores para visualizar las asociaciones. Mediante clic sobre los nodos se pueden observar los linajes de los individuos. Mediante clic en el botón de la derecha se puede cambiar el color de cada grupo que nosotros queramos escoger (domestico, salvaje, asia, europa, colores....)

9. También podemos cambiar la posición de las ramas de la red mediante clic sobre las ramas y después escoger el nodo a mover.
10. Grabar los resultados en formato *.fdi* (para posterior modificación) y en formato *.pdf* o *.bmp* para la visualización gráfica sobre un documento.
11. Ver los arboles con los dos algoritmos. Ver las diferencias y probar a cambiar los parámetros epsilon y r para ver el efecto.

#### **PASOS O EJERCICIOS A REALIZAR EN LA PRÁCTICA 4:**

Vamos a realizar tests de neutralidad (analizar estadísticamente si tenemos señales de procesos selectivos) para el locus *MC1R* en la especie porcina. Vamos a utilizar el programa **DnaSP** para esta tarea.

Los tests a realizar serán el **test MK** (MacDonald y Kreitman, 1992) y el **test HKA** (Hudson, Kreitman y Aguadé, 1987).

El test **MK** asume que el número de mutaciones que se acumulen en una especie en las posiciones sinónimas y no-sinónimas es proporcional al número de mutaciones sinónimas y no-sinónimas que se acumulen entre especies.

En la práctica, utilizaremos el fichero nexus que contiene toda la información con la que hemos trabajado anteriormente mas una especie de referencia (y así conseguir también los valores de divergencia). El fichero a utilizar es entonces: *MC1R\_PigsOutg\_aligned.nex*

1. Abrir *DnaSP*, abrir el fichero *MC1R\_PigsOutg\_aligned.nex*,
2. Excluir las líneas introgresadas **CR0112#EL** y **DW12#AWB**.
3. Realizar el Analisis *MacDonald and Kreitman test* desde el Menu:Analysis.  
Hacer el análisis para los diferentes grupos:
  - a. ASIADOMESTIC vs *Babyrousa bayrussa*
  - b. ASIAWILDBOAR vs *Babyrousa bayrussa*
  - c. EUROPADOMESTIC vs *Babyrousa bayrussa*

- d. EUROPAWILDBOAR vs *Babryrousa bayrussa*
  - e. WILDBOARALL vs *Babryrousa bayrussa*
  - f. DOMESTICALL vs *Babryrousa bayrussa*
4. Guardar la tabla de contingencia y el resultado (Fisher test y p-value).
  5. ¿Que pasa si utilizamos *Potamochoerus porcus* como especie referencia?
  - ¿Que pasa si incluimos las lineas introgresadas?

El test **HKA** compara el polimorfismo y divergencia entre dos diferentes regiones de DNA. El test asume que una de las regiones se comporta de manera neutra y bajo un modelo estacionario, es decir, una población de tamaño constante sin selección (y otros supuestos) y con una tasa de mutación que permanece constante a lo largo del tiempo. Para realizar este análisis se necesitan al menos dos regiones, una neutra (por ejemplo, el locus *Glucose phosphate Isomerase pseudogen GPIP*) y la otra nuestra región de interés, el locus *MC1R*.

Vamos a utilizar en este caso los datos de Li *et al.*, Heredity (2010) para las poblaciones salvajes y domesticadas localizadas en China. Existen dos ficheros nexus: uno para *MC1R* y otro para *GPIP*. El trabajo consiste en calcular los datos de polimorfismo y de divergencia para cada uno de los genes y compararlos mediante HKA test. El proceso a realizar es el siguiente:

1. Se habrá de hacer el cálculo por separado por cada gen para luego integrarlo en el análisis de HKA en la opción “manual”. Hay que calcular los siguientes valores para cada uno de los grupos CHINADOMESTIC y CHINAWILDBOAR: Número de nuestras, número de posiciones, número de mutaciones polimórficas y divergencia entre especies, tanto para el locus *GPIP* como para el locus *MC1R*:
  - a. Abrir el fichero nexus *GPIP\_aligned\_all.nex*. Menu:Analysis:DNA *Divergence between Pops*. Recoger (o apuntar en la libreta) los valores para el número de uestras y posiciones, polimorfismo y divergencia para cada poblacion por separado:
    - i. CHINADOMESTIC vs *Potamochoerus*

- ii. CHINAWILDBOAR vs *Potamochoerus*.
  - b. Hacer lo mismo con el fichero *MC1R\_COD\_aligned.nex*.
- 2. Ir al Menu: *Tools: HKA Test (Direct Mode)* y completar la Tabla. Calcular HKA.
- 3. Discutir el resultado.