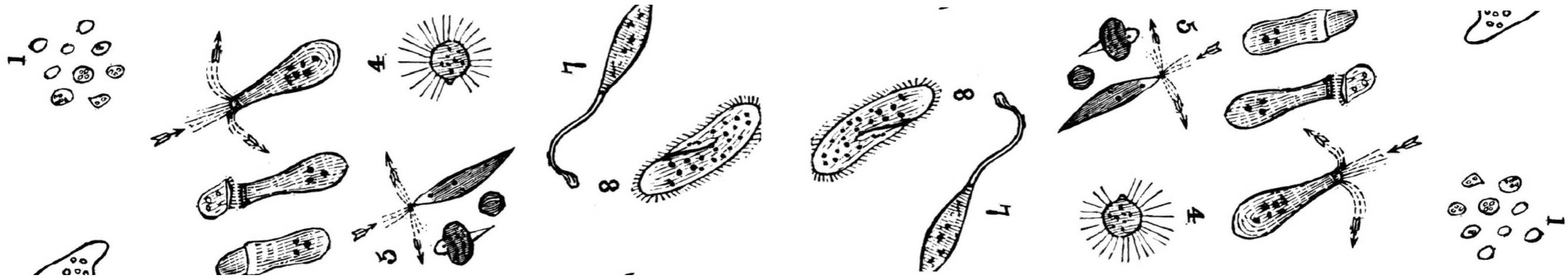




Mapas de calor, diagramas de Venn, filogenias

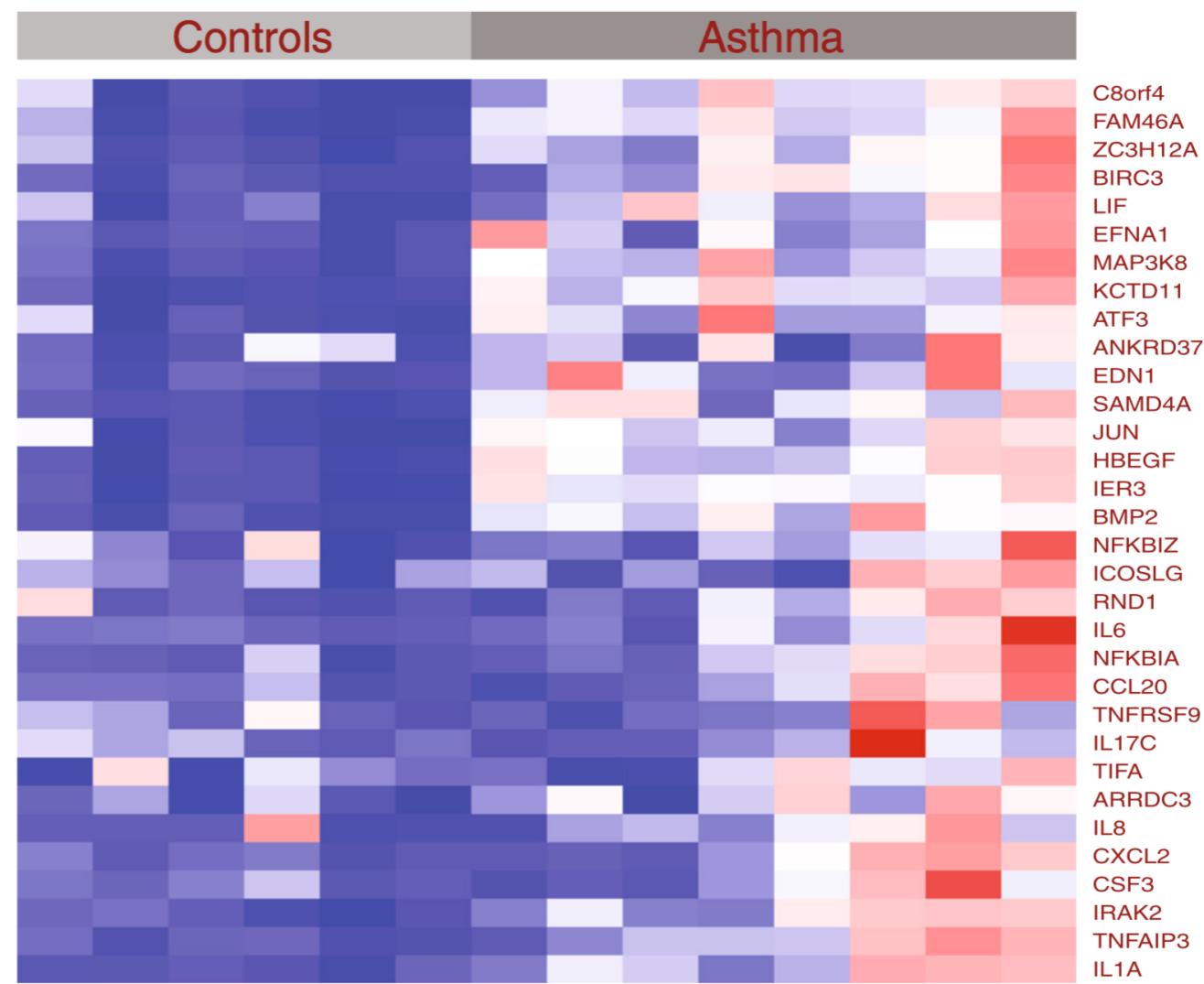
www.castrolab.org
www.cbib.cl

Eduardo Castro, PhD
Universidad Andrés Bello
9 de abril de 2018



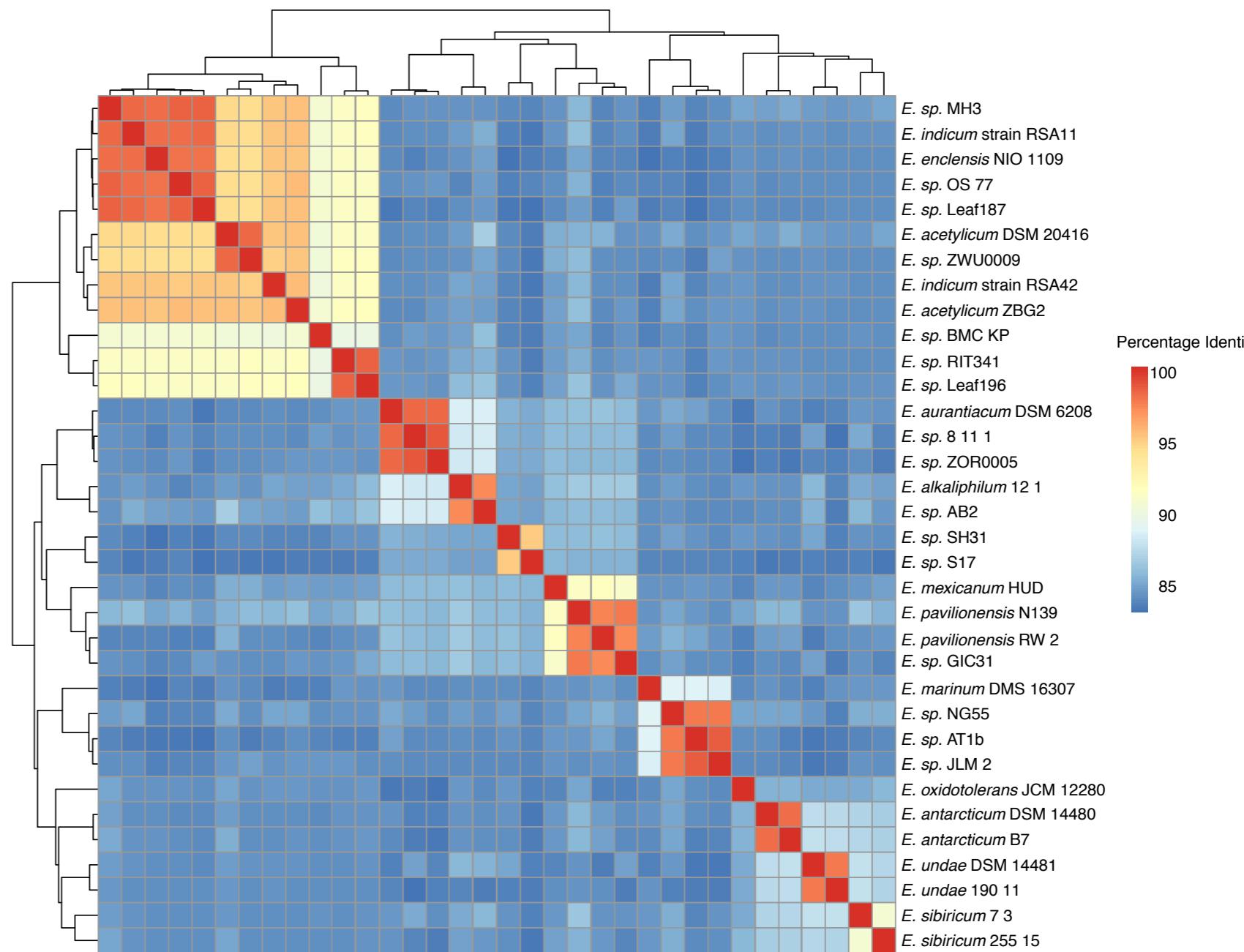
Mapas de calor o heatmaps

- Permiten resumir una gran cantidad de datos (data points) -> eficientes
- Comparación múltiple -> datos multivariados
- Muy útiles en genómica y metagenómica



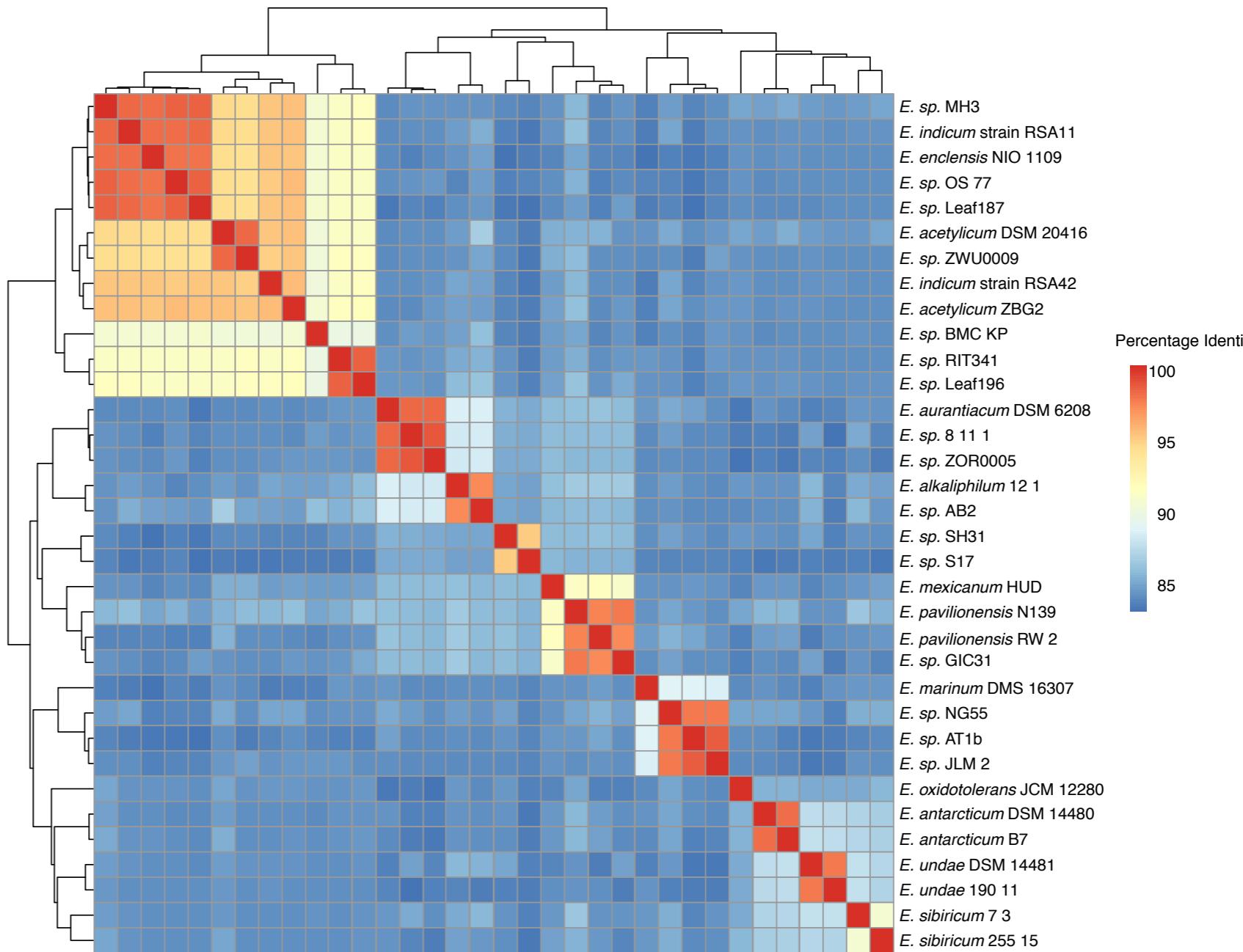
Heat maps representan tablas bidimensionales

- Dataset de genomas bacterianos
- Comparación de todos contra todos a partir de una matriz de distancia



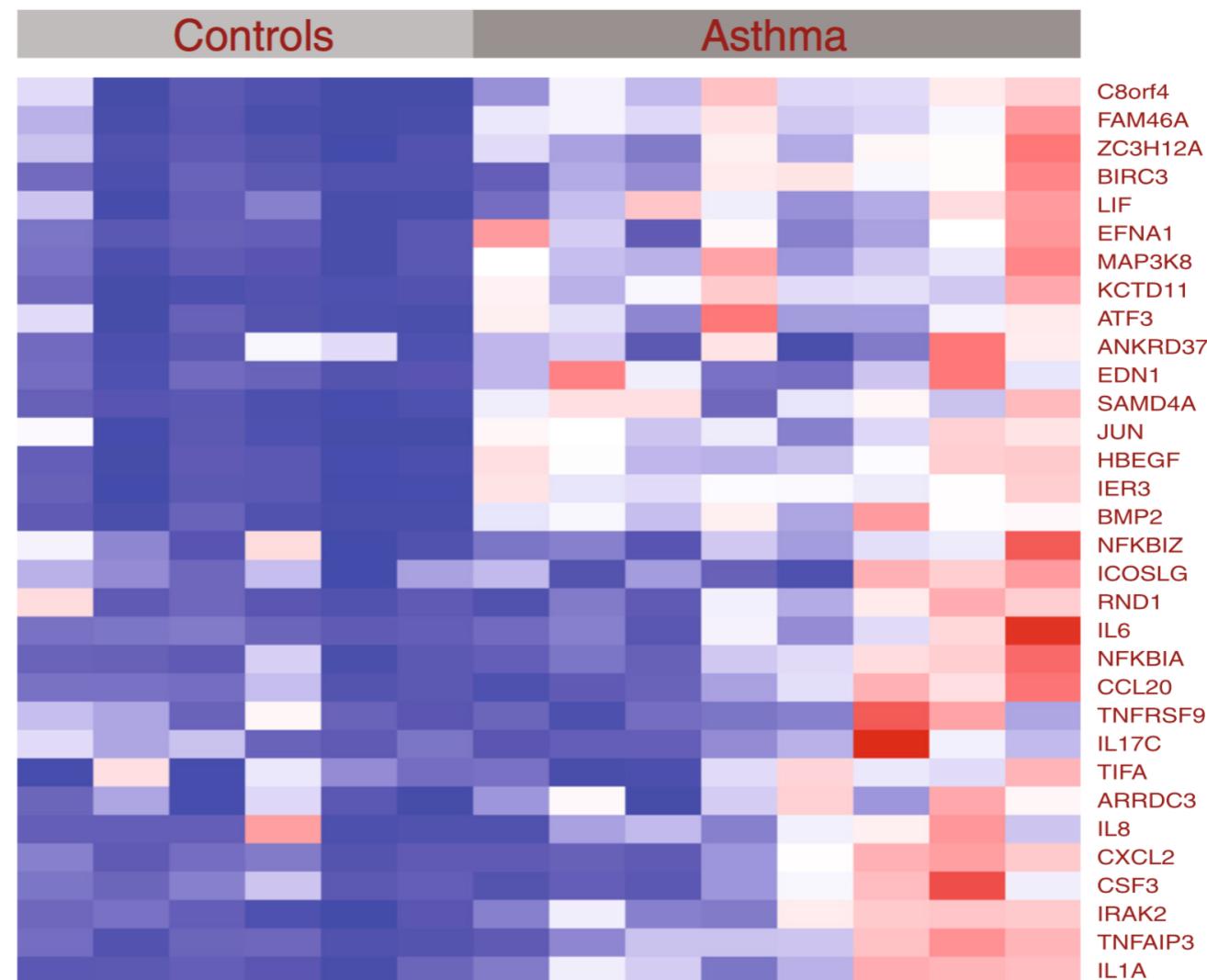
Aspectos más relevantes

- El esquema de colores
- El orden de las filas/ columnas
- La magnitud que representan los colores



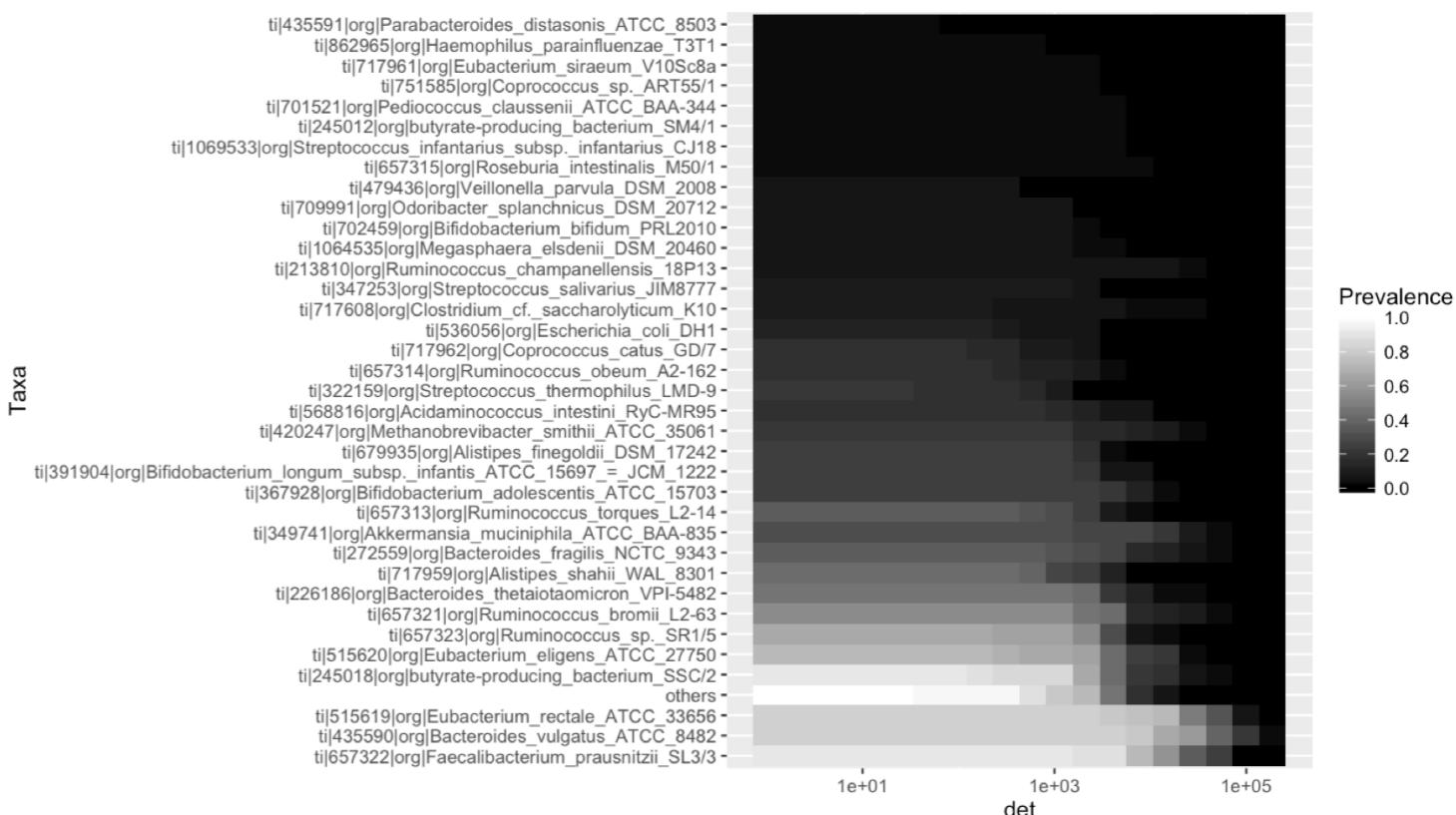
Esquema de color

- Para destacar diferencias, se prefiere un mapa de colores divergente
- De azul a blanco a rojo



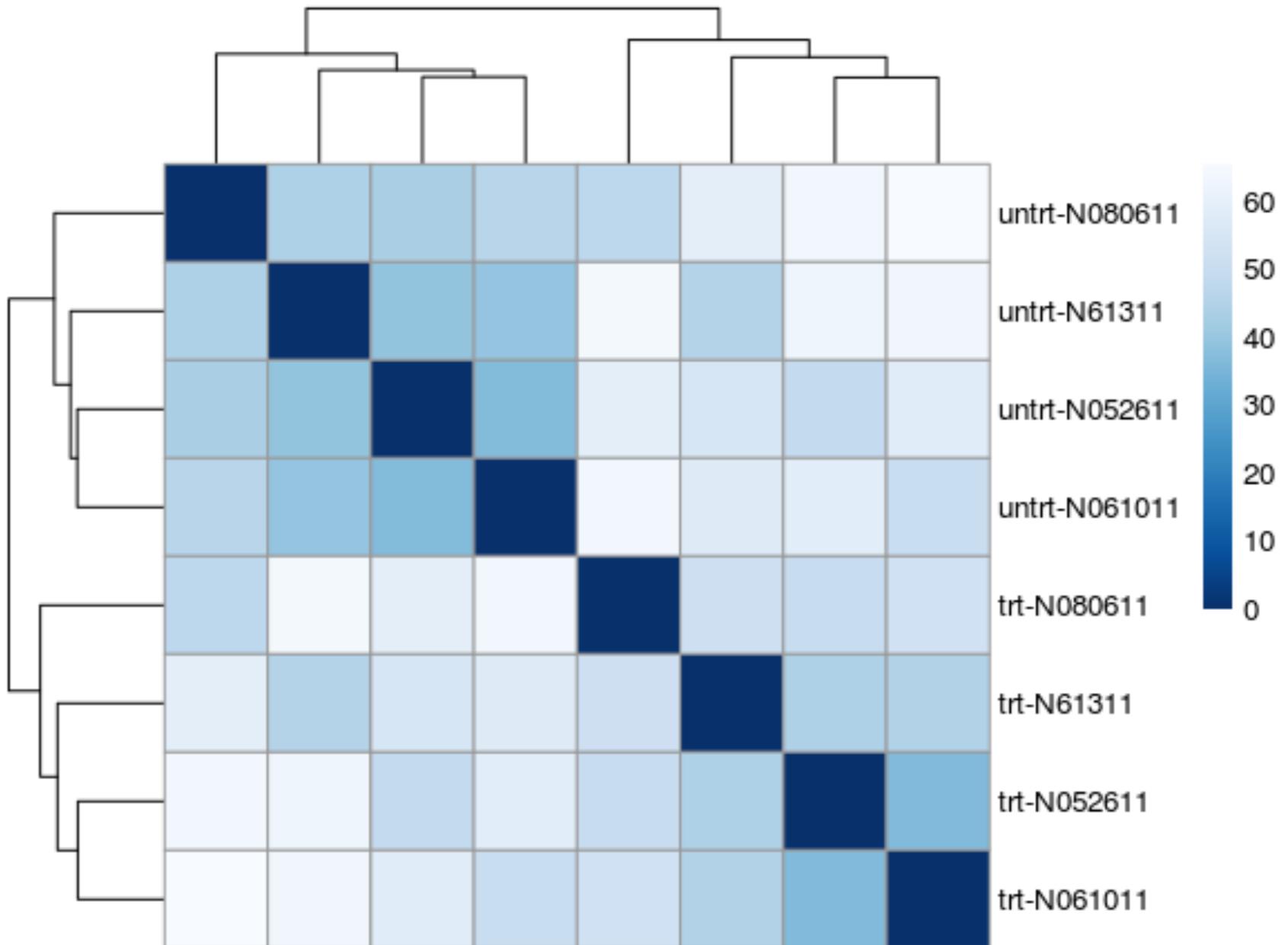
Esquema de color

- En cambio para mostrar intensidad es mejor usar tonalidades de un mismo color
- Ejemplo clásico es la escala de grises



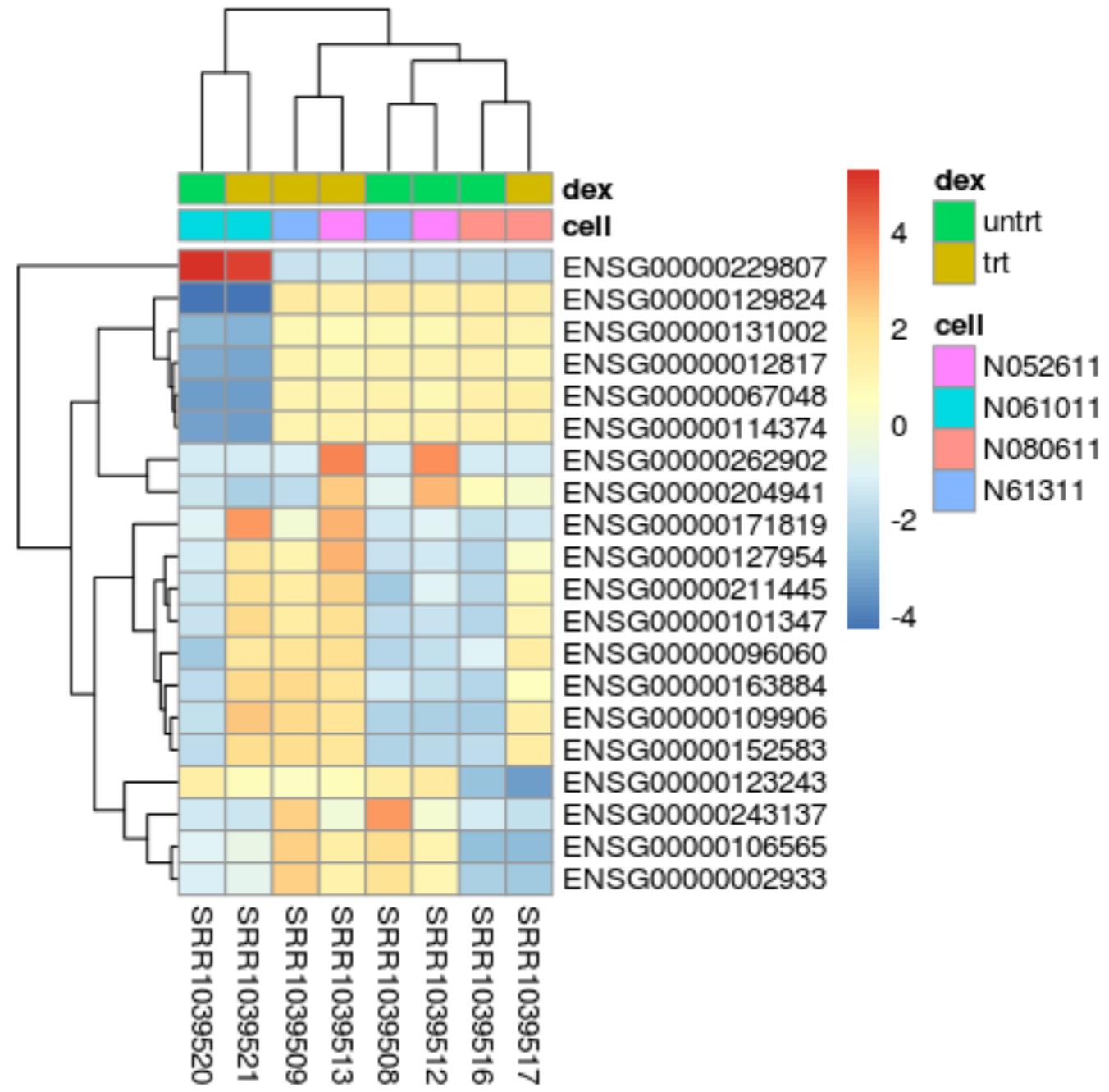
Heat maps: muestra a muestra

- Experimentos de transcriptómica
- ¿Qué muestras son más similares entre sí?
- Expectativa: muestras con tratamientos similares tienen expresión génica similar



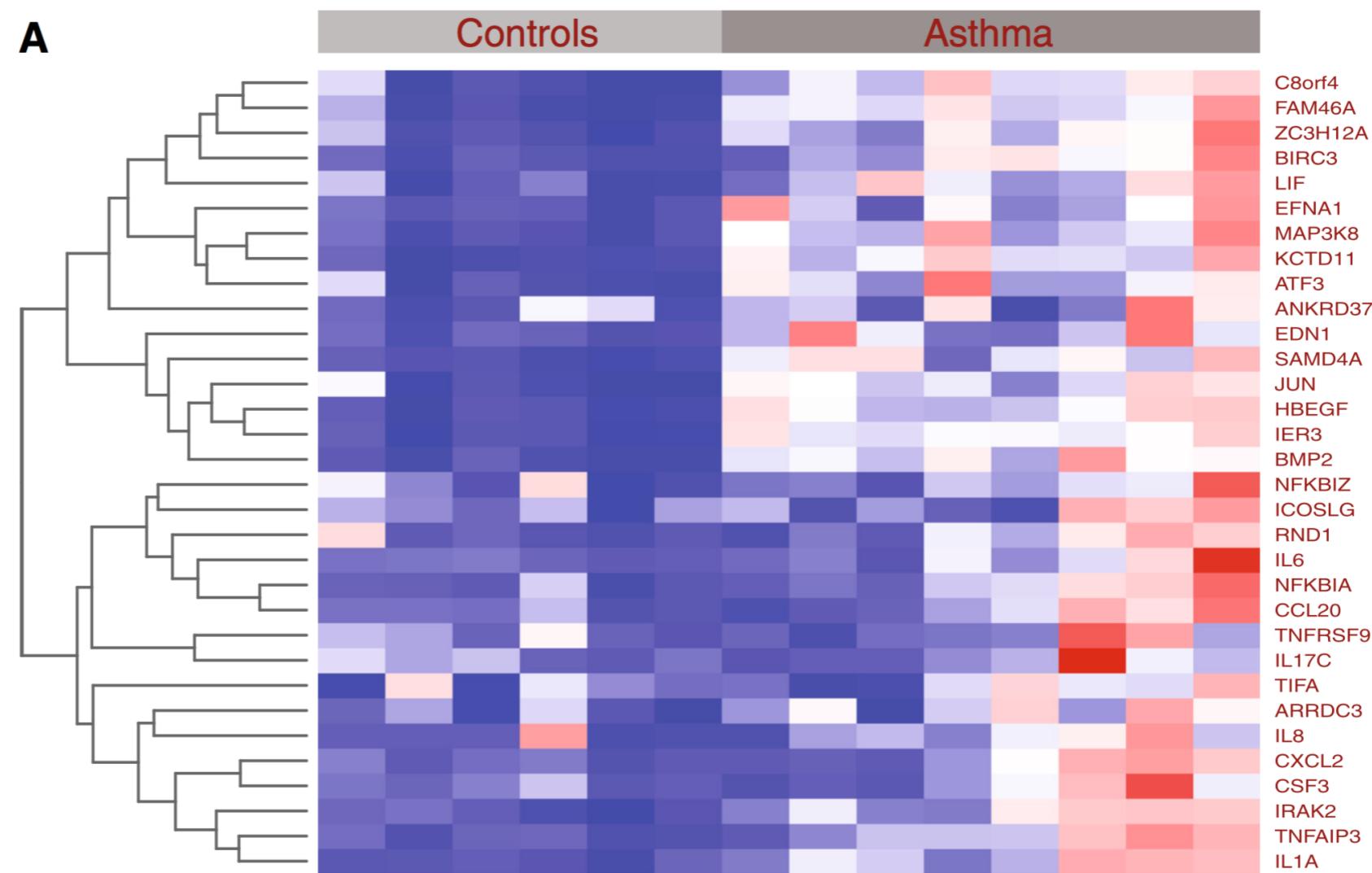
Heat maps: muestra a gen

- Experimentos de transcriptómica
- ¿Qué genes cambian su expresión de manera similar entre muestras?
- Expectativa: muestras cuyos genes cambian su expresión de manera similar están relacionadas biológicamente

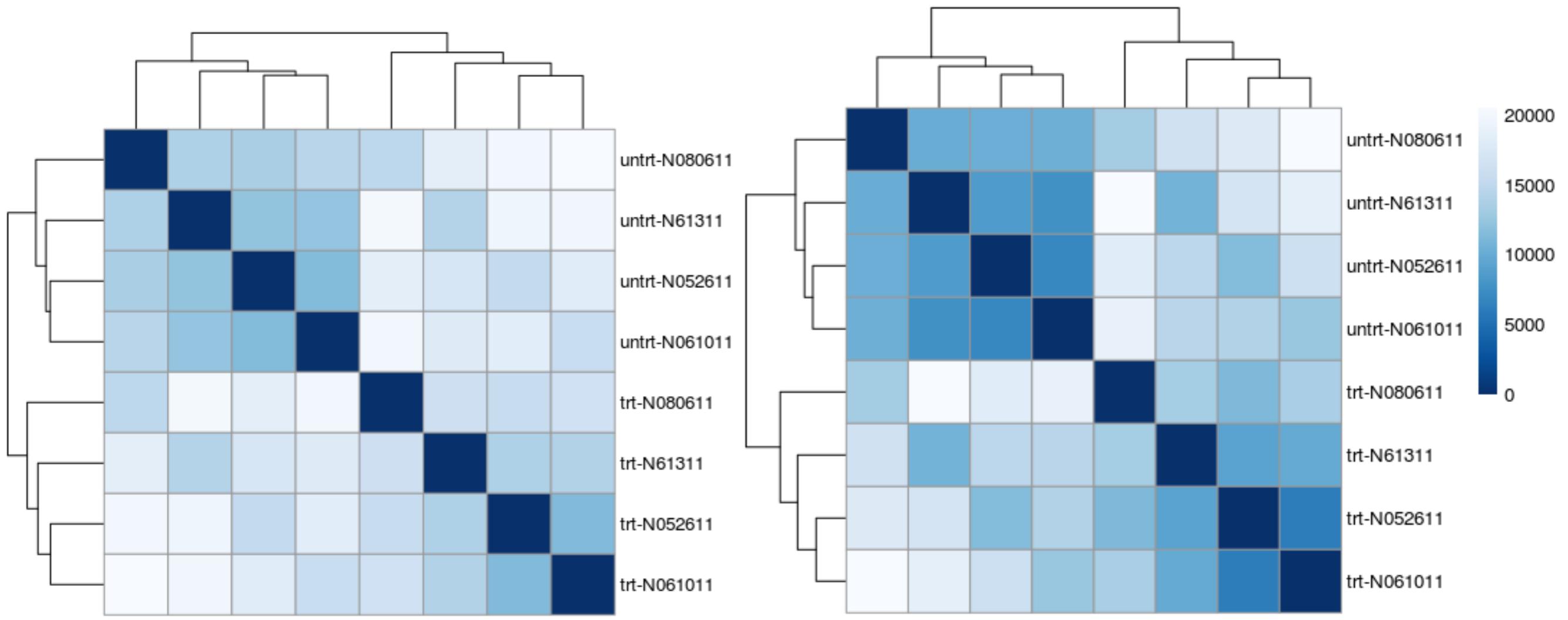


¿Cómo funciona el clustering?

- 1. Calcular distancia “todos contra todos”. Distancia es qué tan diferentes dos muestras son
- 2. Juntar puntos menos distantes entre sí
- 3. Juntar grupos de puntos
- 4. Dendrograma es la representación gráfica del clustering



Método de distancia

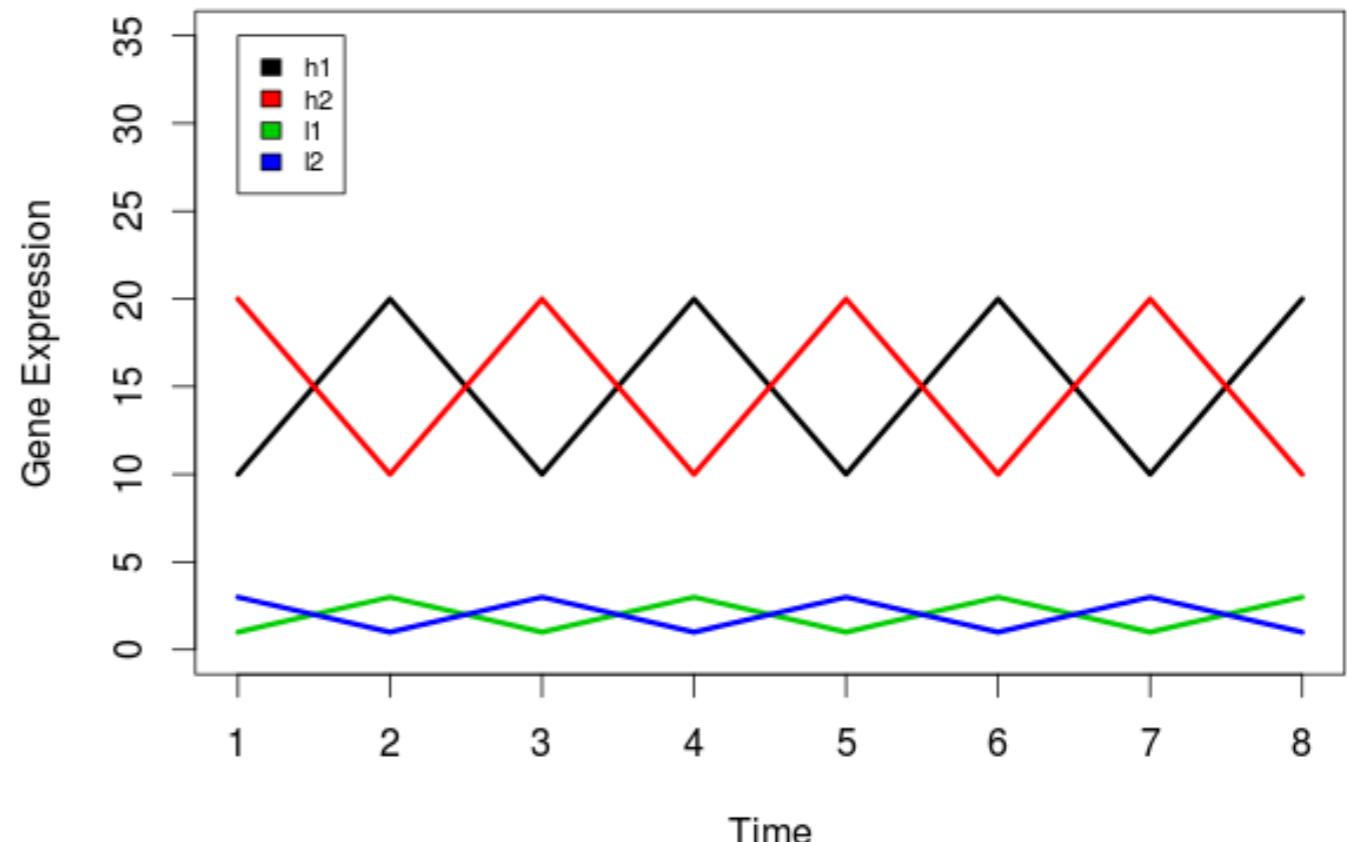


Distancia Euclidiana

Distancia de Poisson

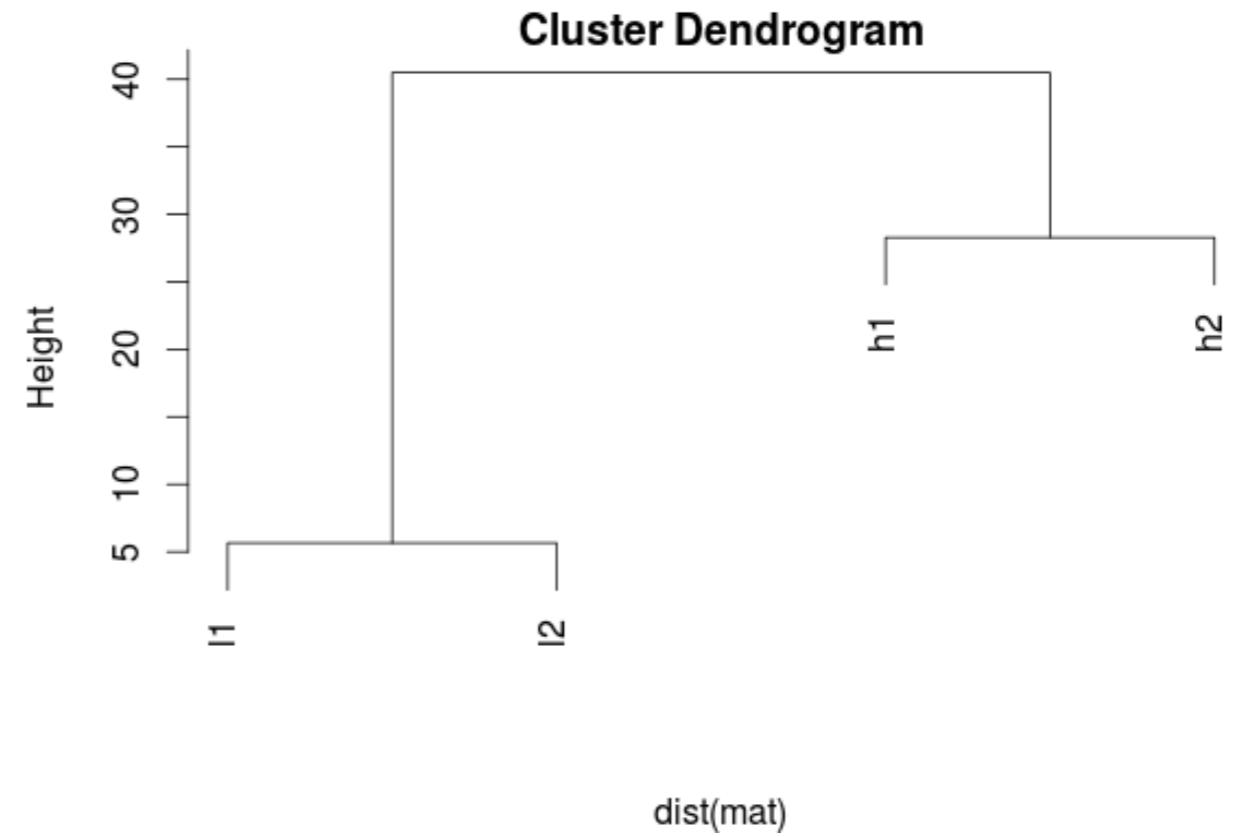
En un heat map queremos representar “la forma” de los datos

- La clave es asegurarse que los colores representan exactamente lo que queremos → la forma de los datos
- Forma de los datos, suben o bajan juntos
- Dos genes de expresión alta y dos de baja



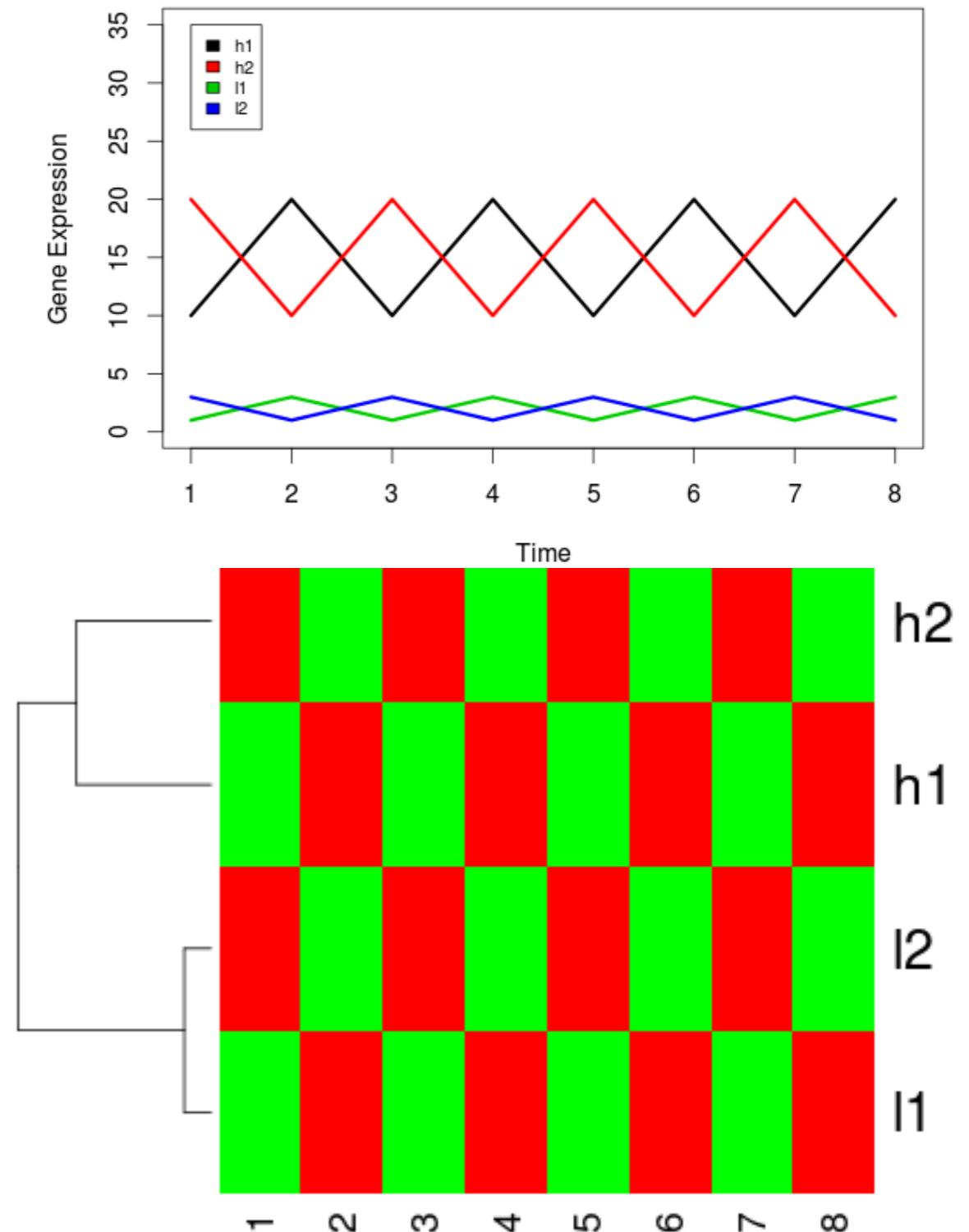
Distancia: queremos representar la forma de los datos

- La clave es asegurarse que los colores representan exactamente lo que queremos —> la forma de los datos
- Dos genes de expresión alta y dos de baja



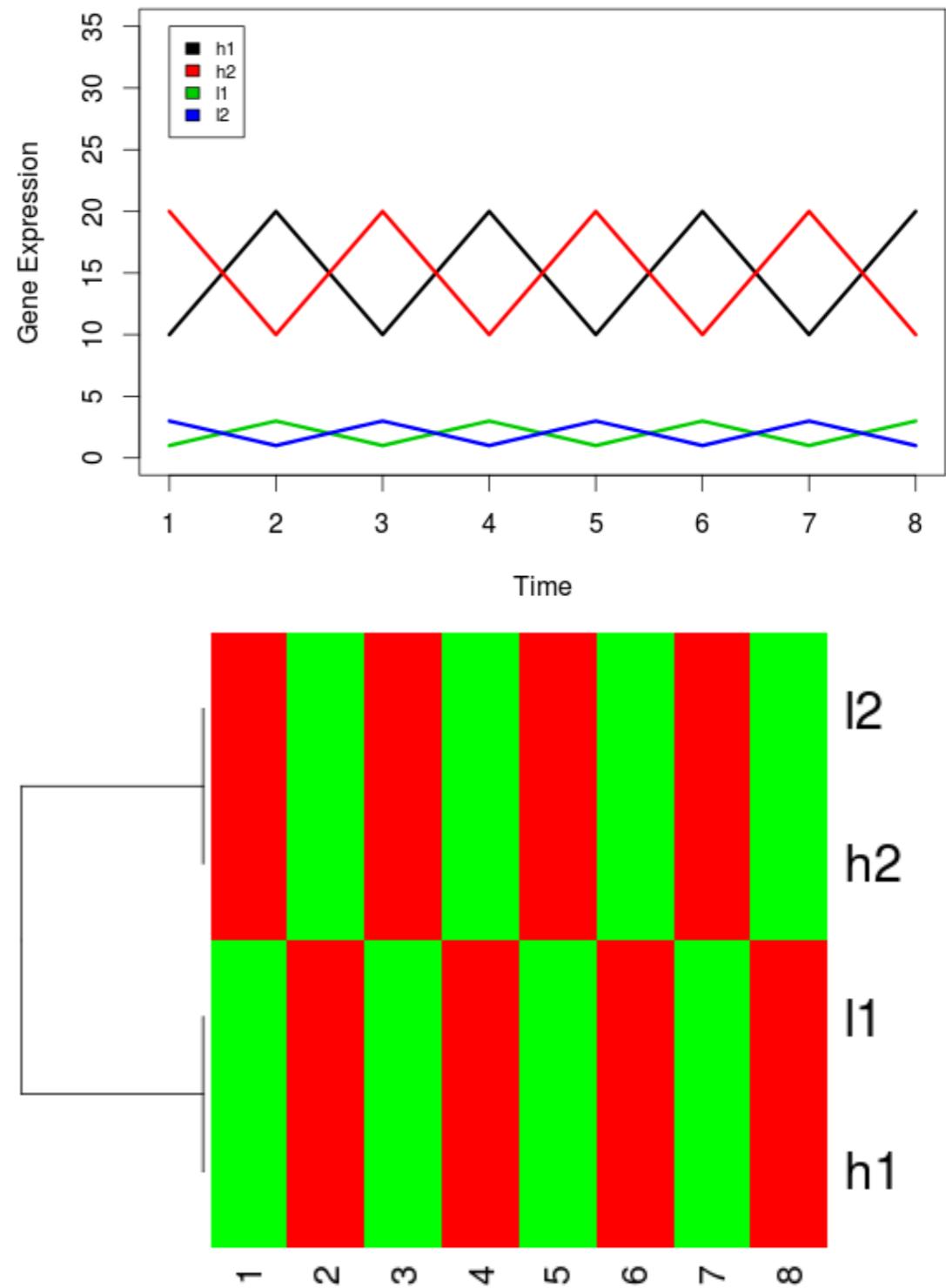
Distancia: queremos representar la forma de los datos

- El problema con este heat map es que tradujo los números literalmente en colores
- I1 y I2 forman clusters pero sus colores son opuestos
- I1 y h1 tienen los mismos colores, sin embargo su expresión es diferente



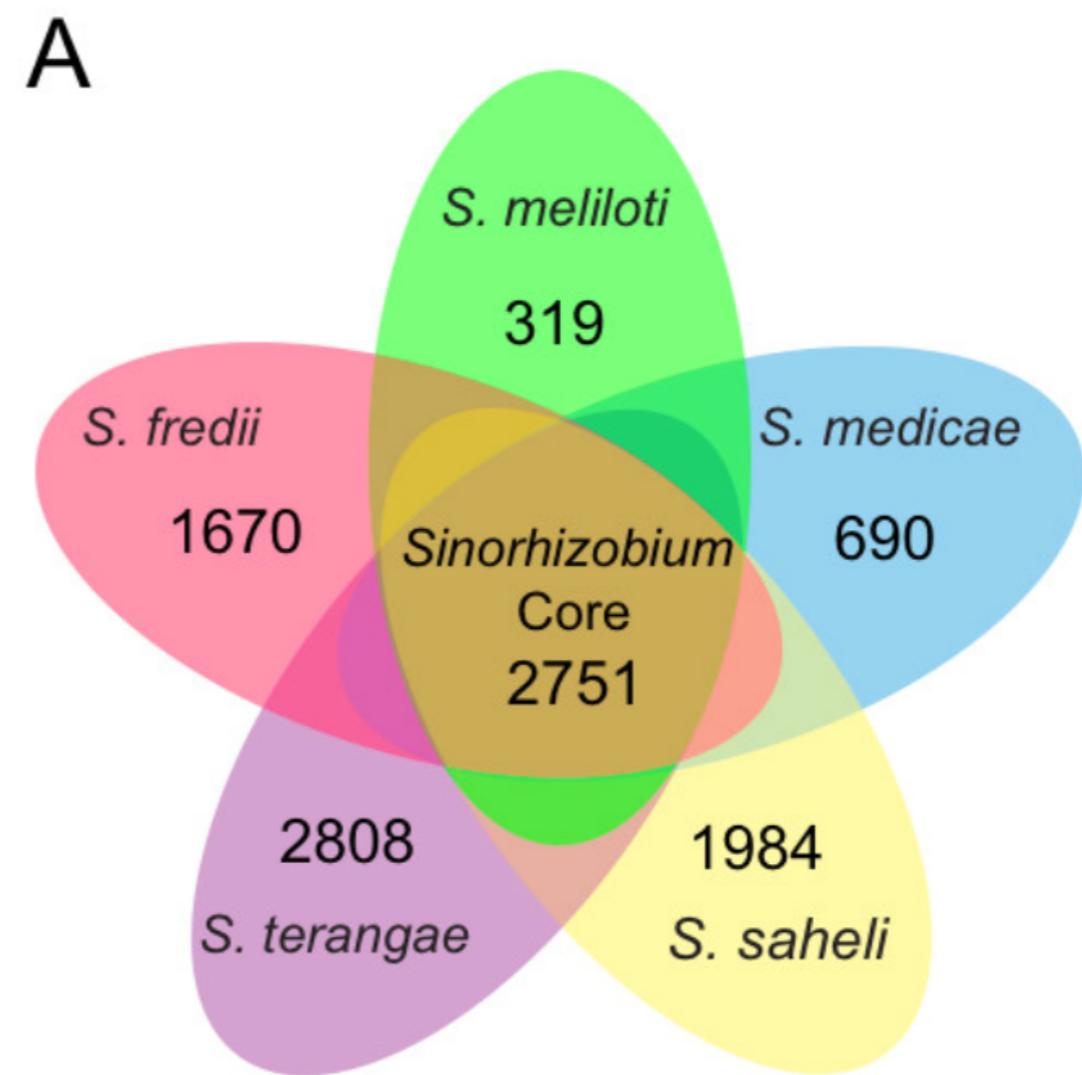
Distancia: queremos representar la forma de los datos

- En vez de una matriz de distancia, calculamos una de correlación
- Correlación de Pearson, de -1 a +1, completamente opuesto a idéntico, respectivamente
- La correlación captura “la forma de los datos”



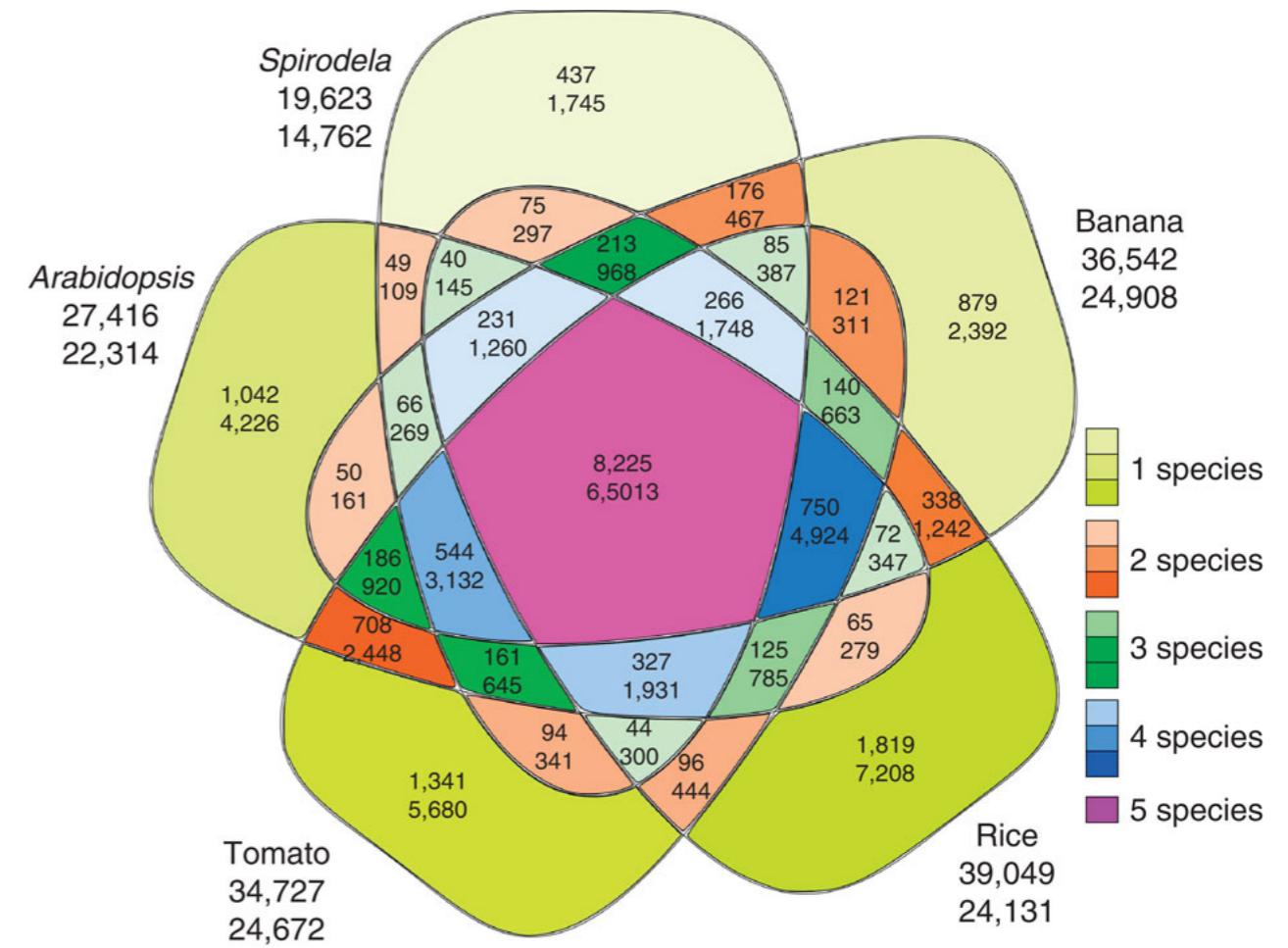
Diagramas de Venn

- Un diagrama que representa conjuntos
- Muestra todas las relaciones lógicas entre muestras
- Familias de genes comunes entre especies



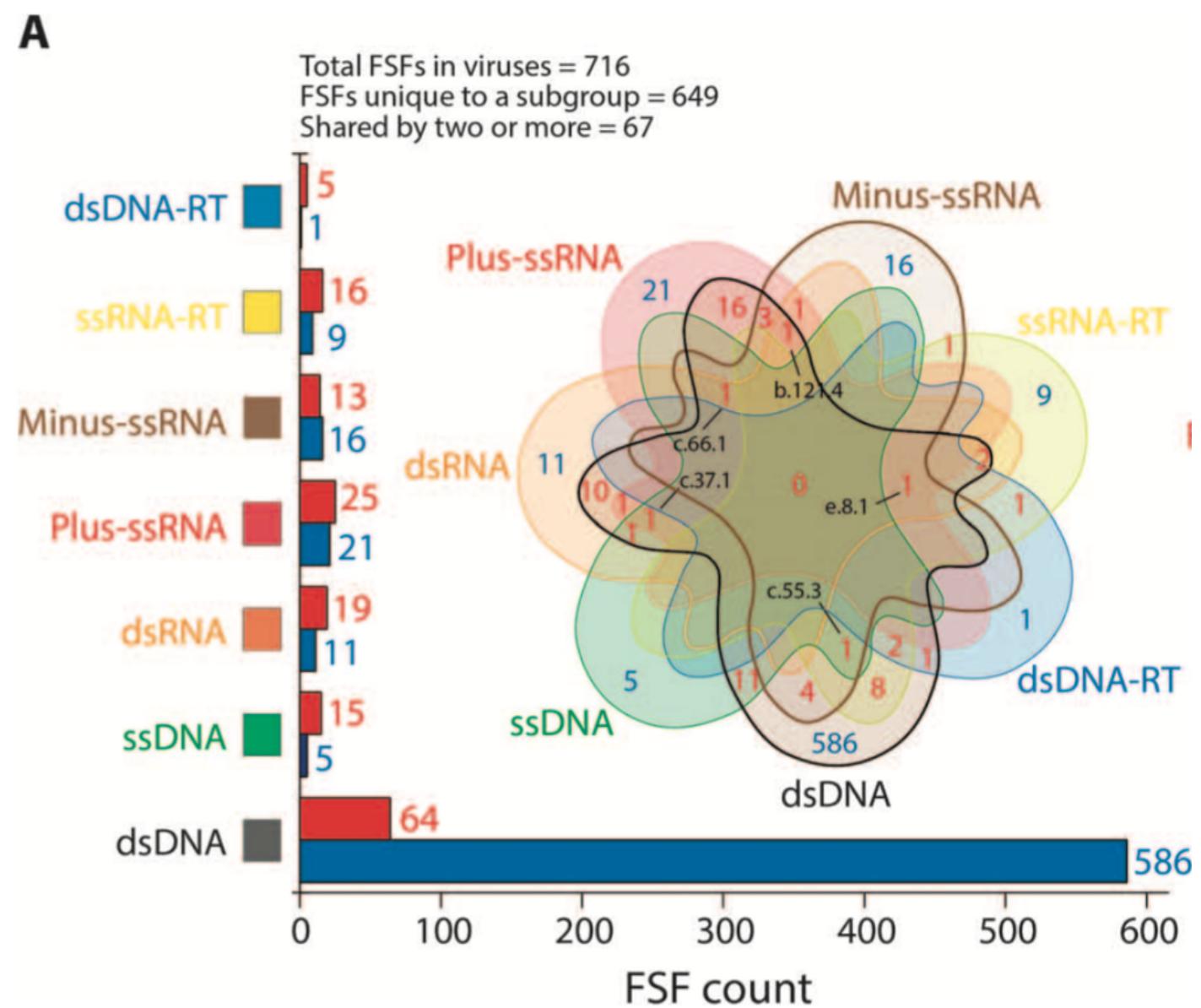
Diagramas de Venn

- Keep it simple
- Más de 5 conjuntos y la visualización comienza a jugar en contra de la ciencia

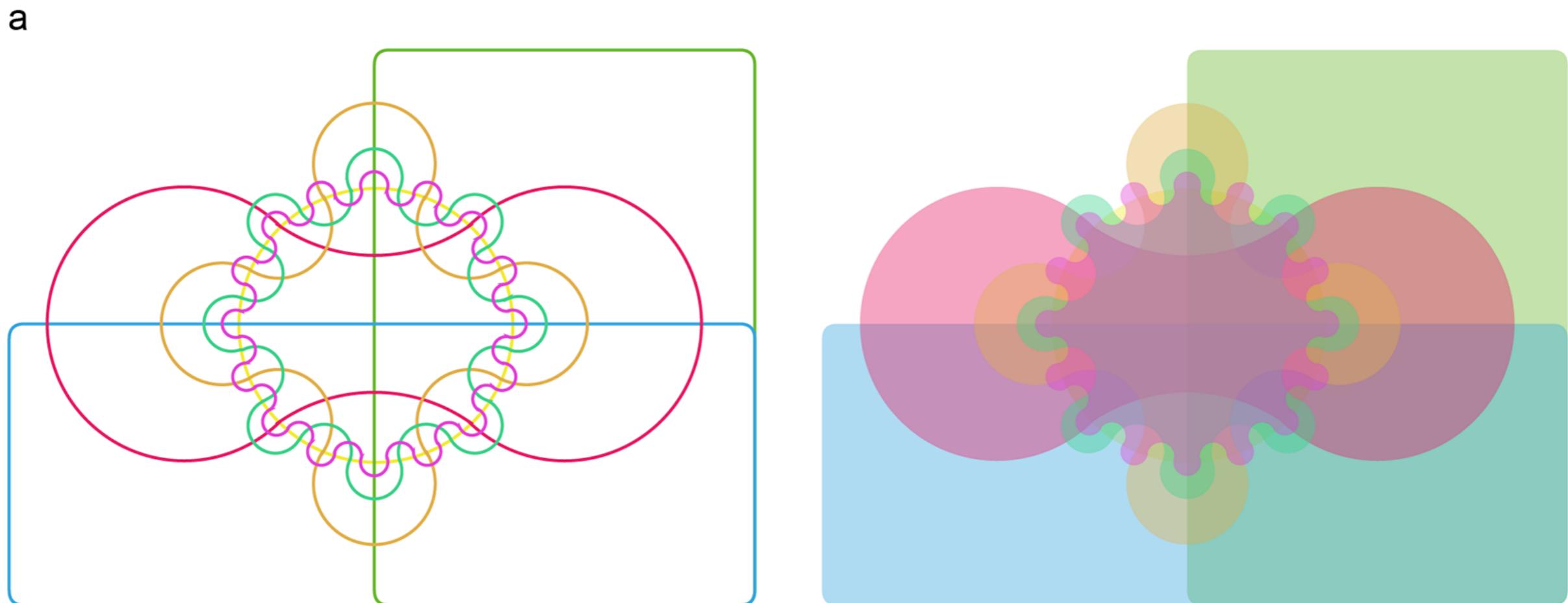


Diagramas de Venn

- Keep it simple
- Más de 5 conjuntos y la visualización comienza a jugar en contra de la ciencia

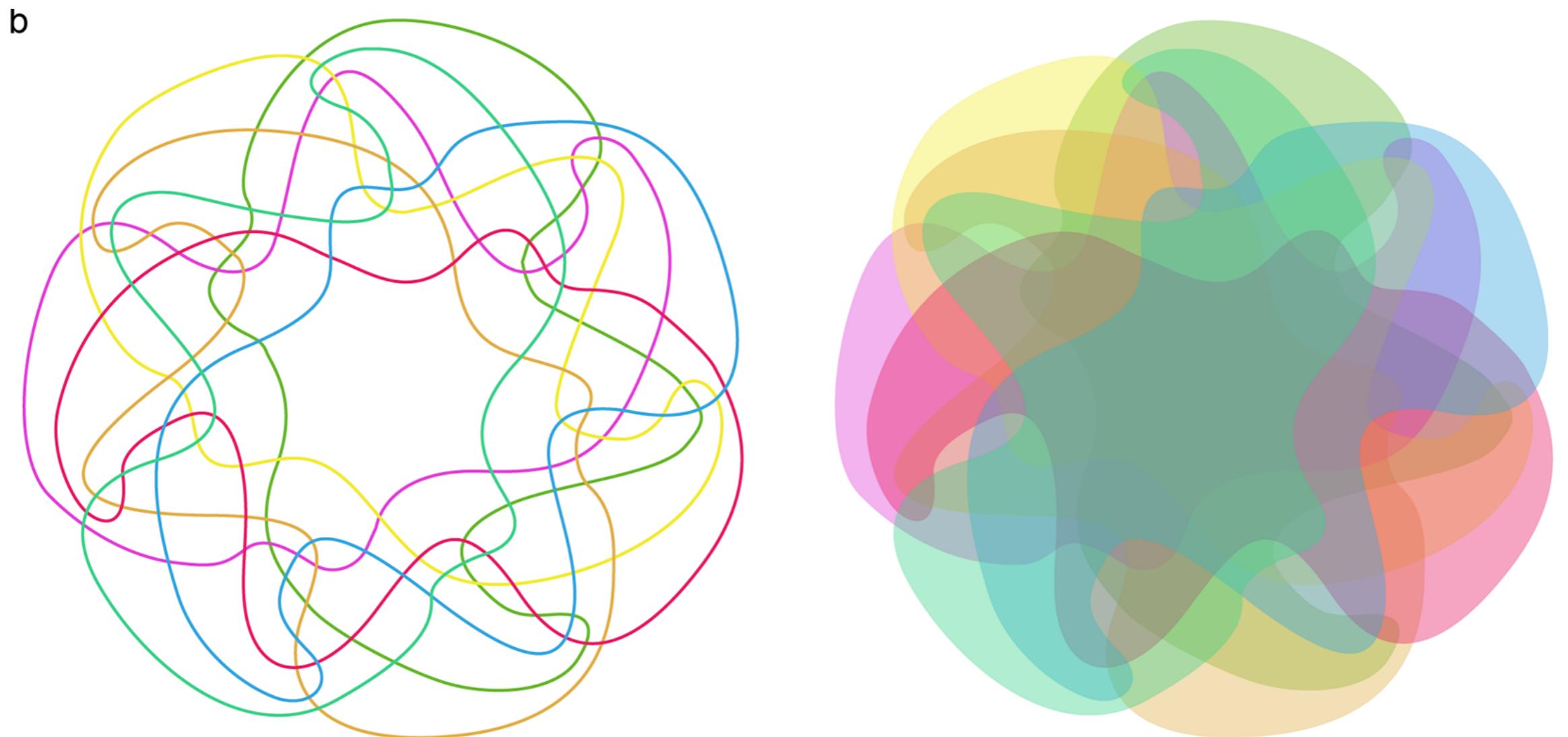


Variaciones a diagramas de Venn



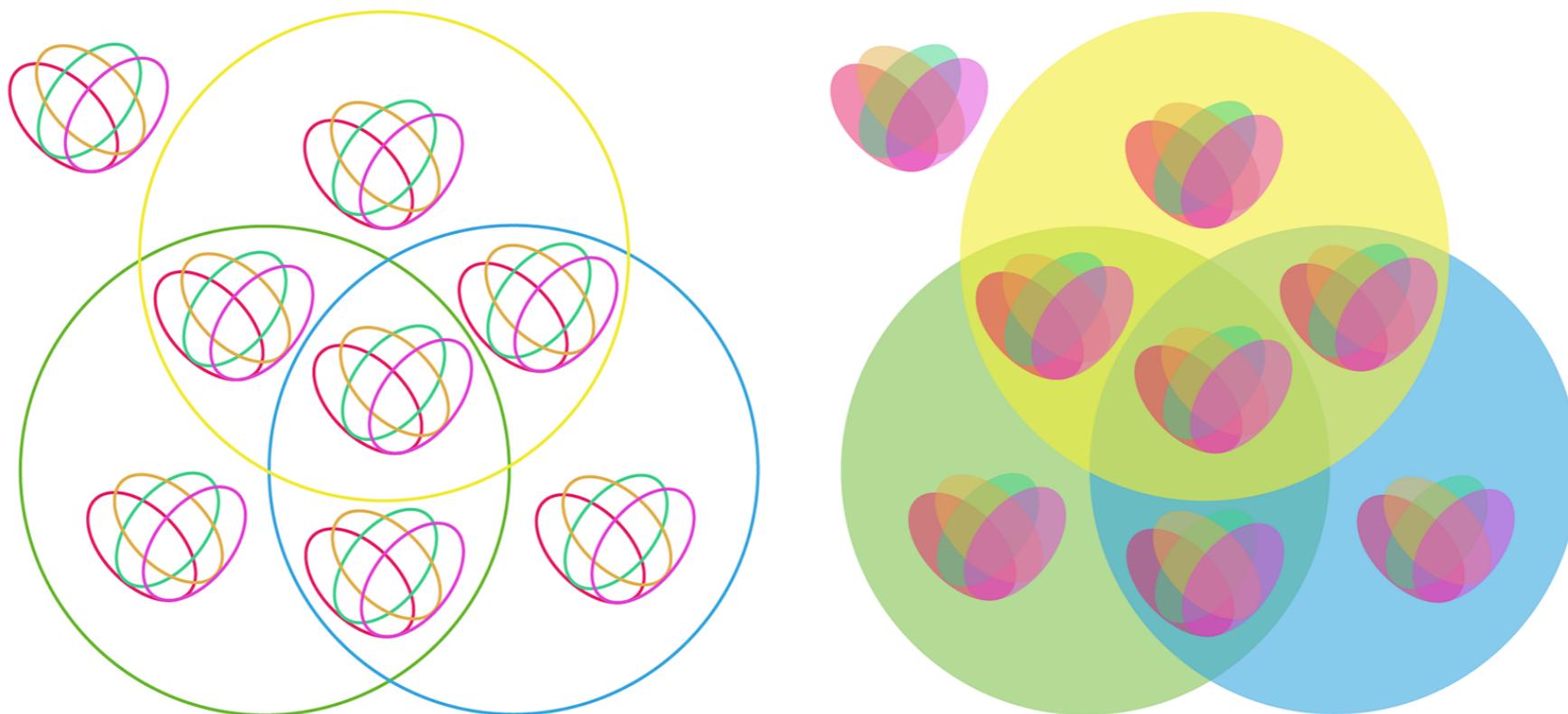
- Diagrama de Venn, tipo Edwards

Variaciones a diagramas de Venn



- Diagrama de Venn de curvas irregulares

Variaciones a diagramas de Venn

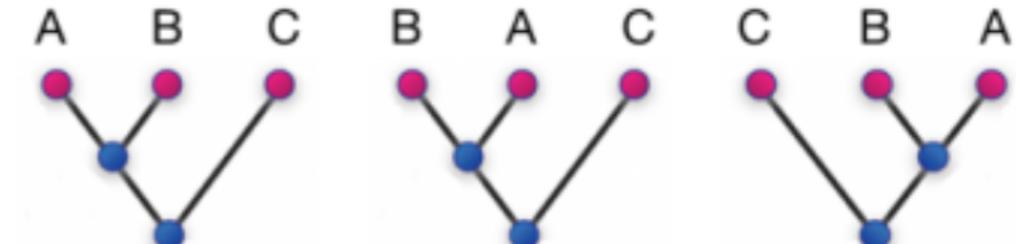


- Diagrama de Venn anidado
- Implementado en el software VennPainter

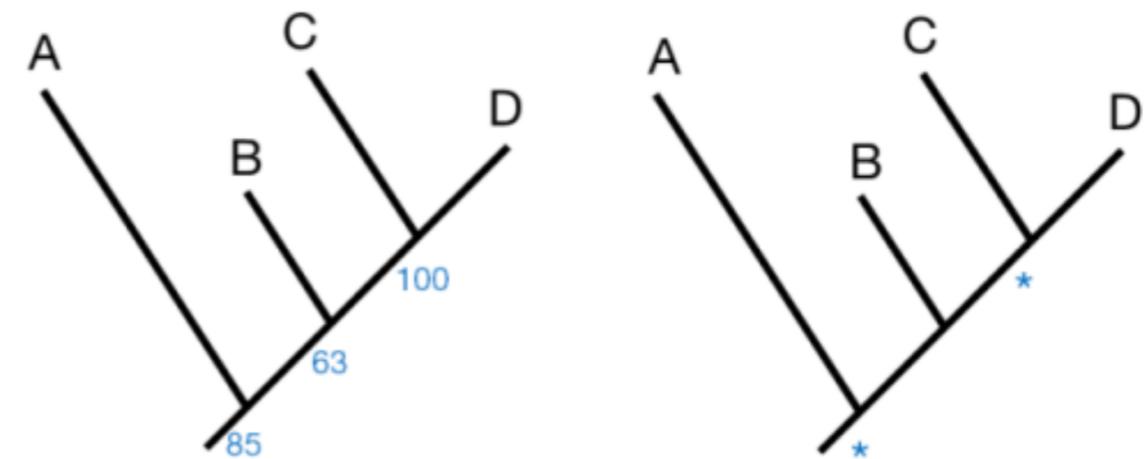
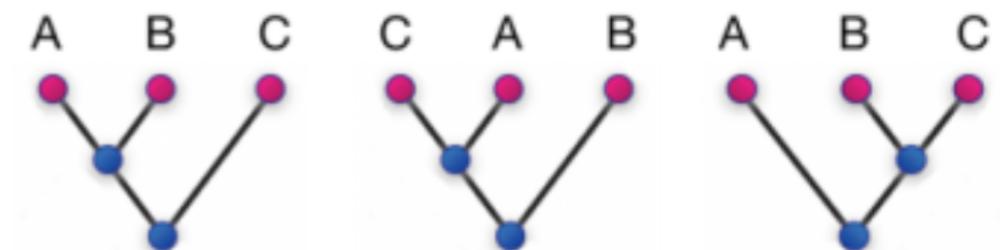
Características de las filogenias

- Topología
- Ramas - branches (edges)
- Nodos - nodes (vertices)
 - Raíz - root
 - Nodos internos
 - Extremos - tips
- Confianza en nodos

These trees display the same topology

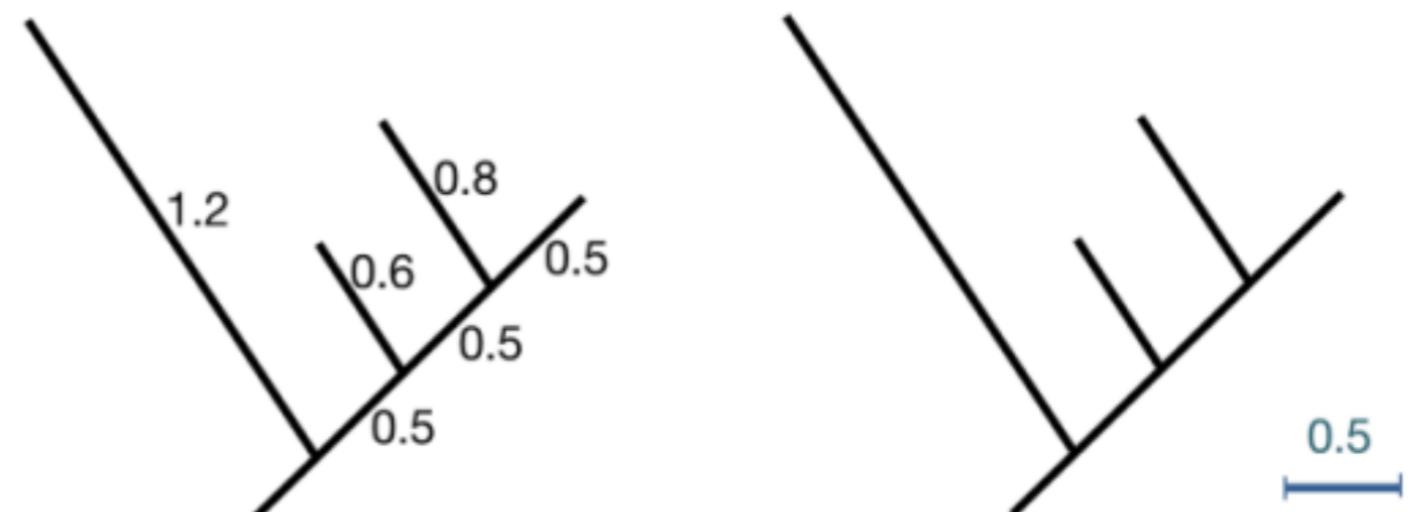


These trees display different topologies



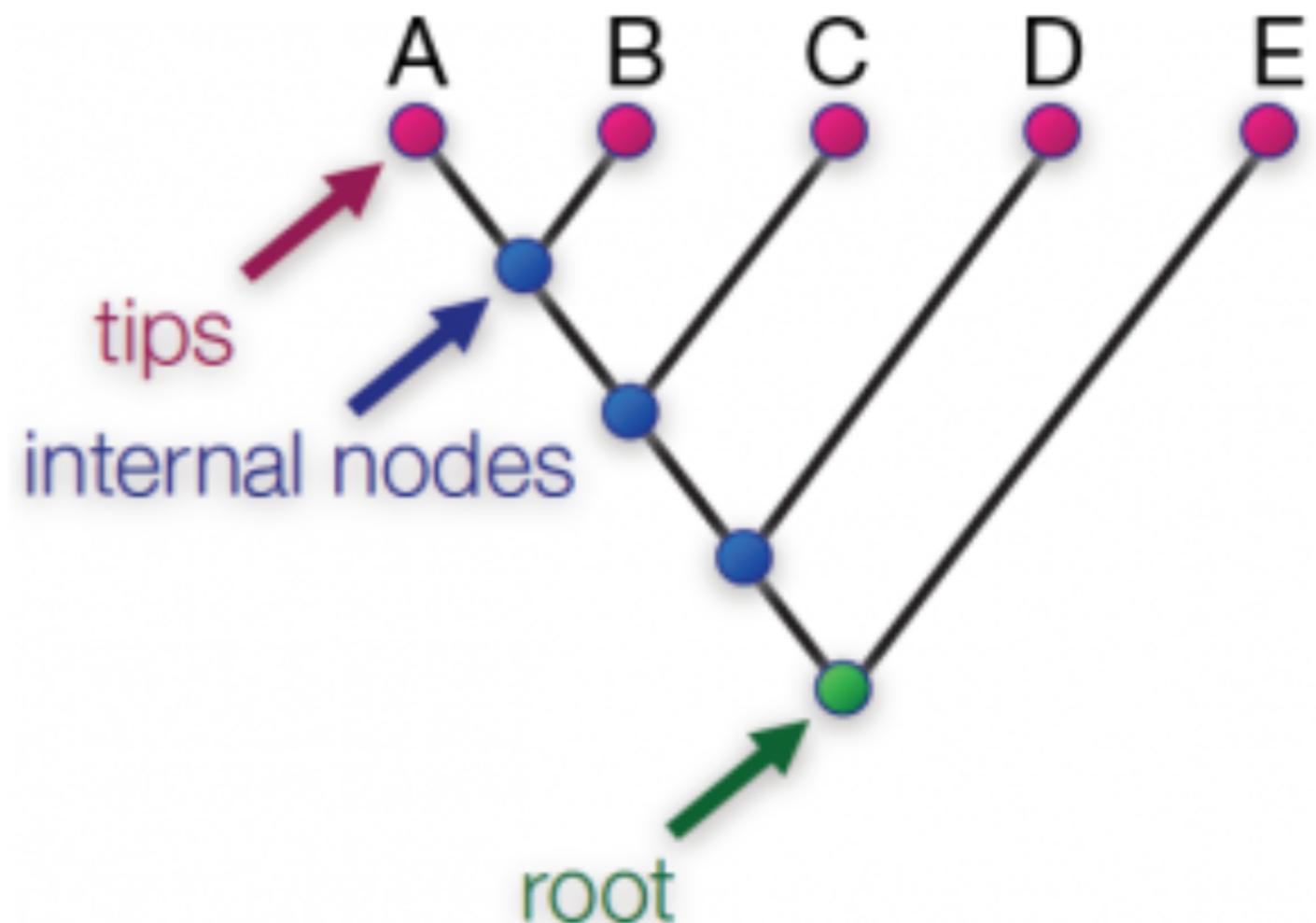
Características de las filogenias

- Ramas - branches (edges)
- Indican la cantidad de cambio
- Expresado en sustituciones por sitio, sustituciones por gen
- En algunos casos, el largo de las ramas está escalado por tiempo (relojes moleculares)



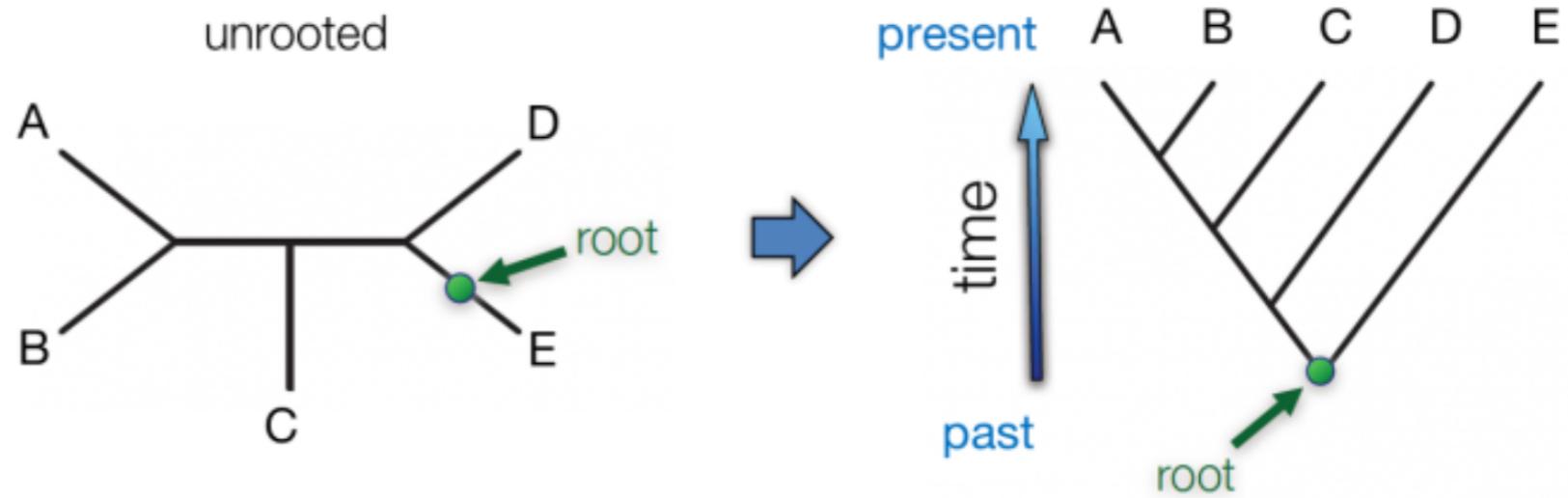
Características de las filogenias

- Nodos
- Representan ancestros
- Mayoría de las veces no podemos observar a esos ancetros
- Excepciones?

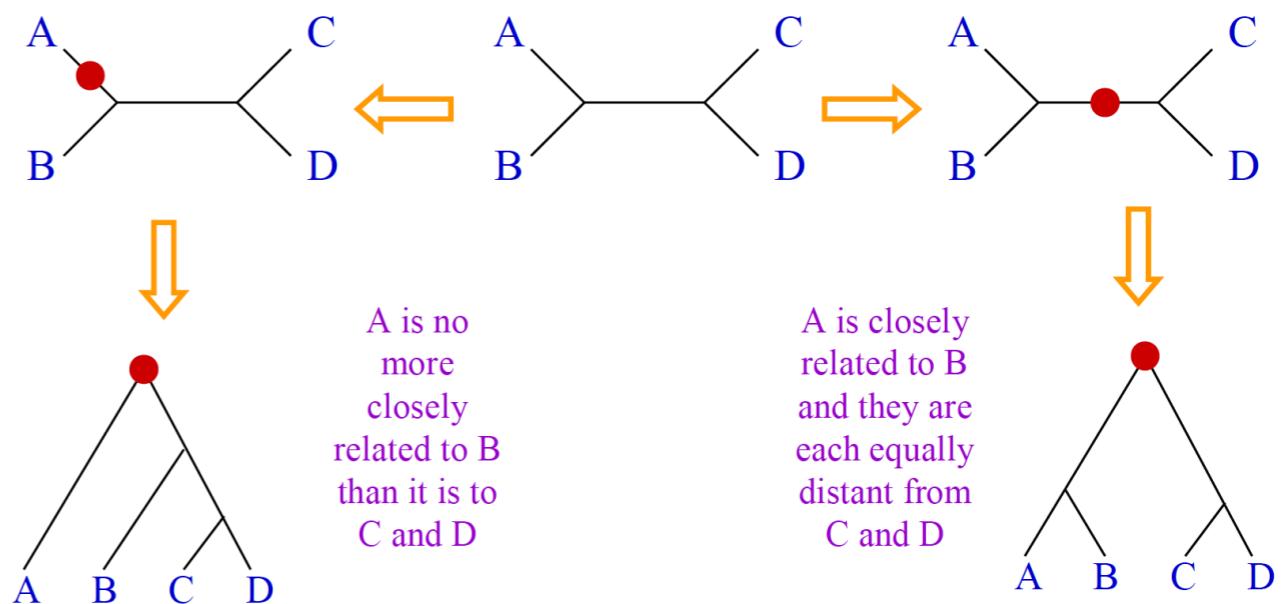


Características de las filogenias

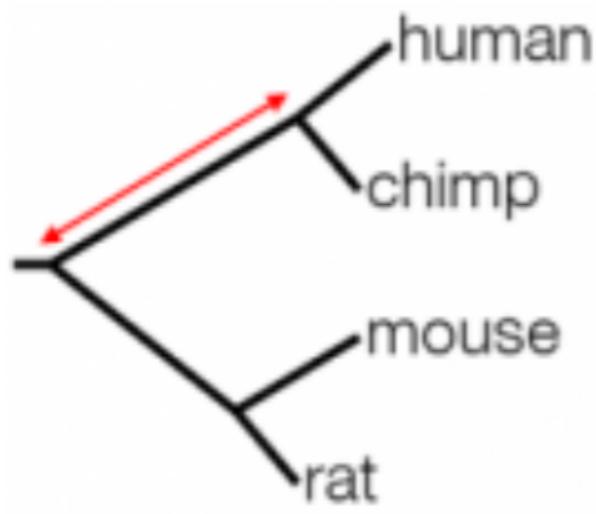
- Raíz
- Árboles enraizados y no enraizados
- Posibles árboles?
- Enraizados = $(2n-3)!!$
- No-enraizados = $(2n-5)!!$



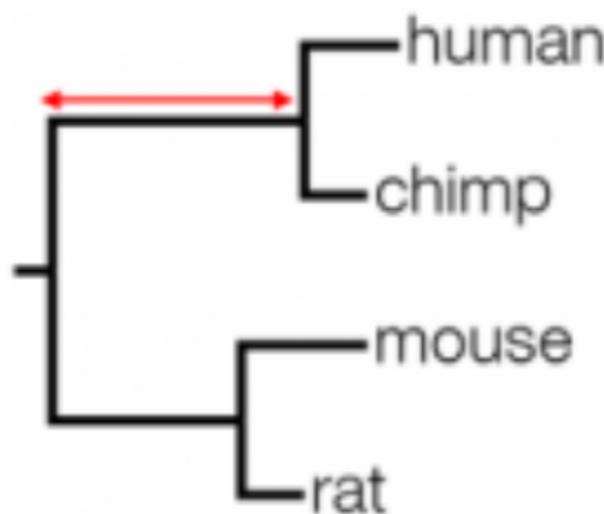
- A single unrooted tree can imply different relationships between species depending on the location of the root



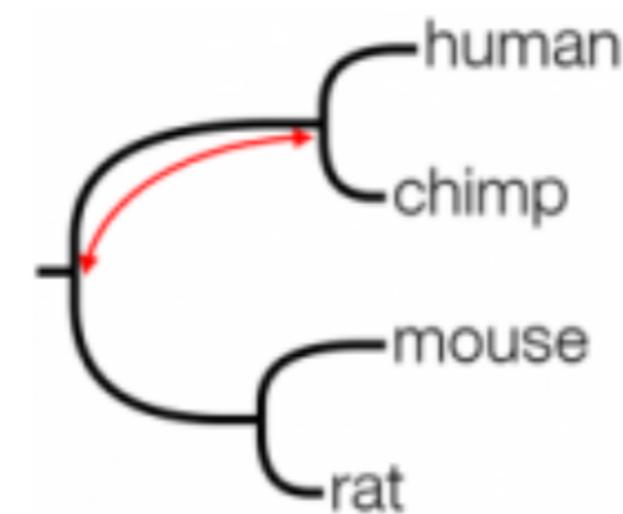
Características de las filogenias



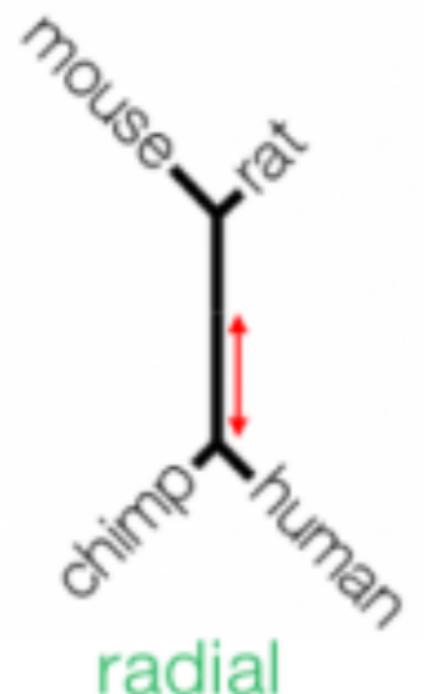
diagonal



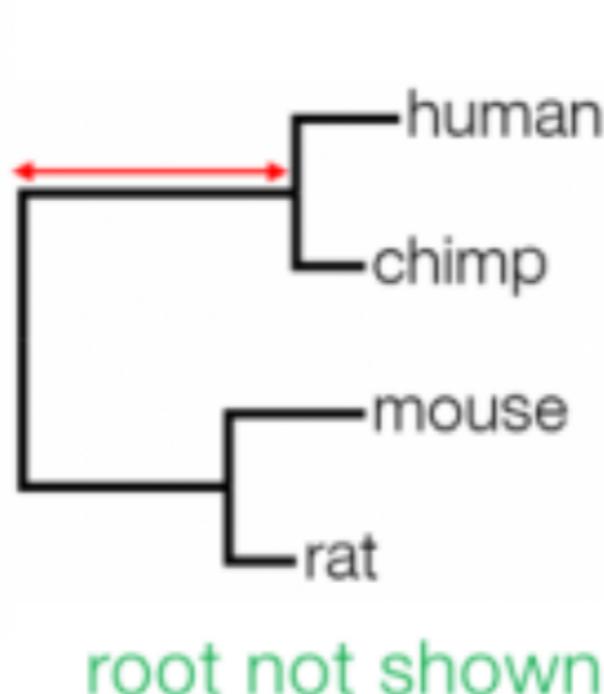
rectangular



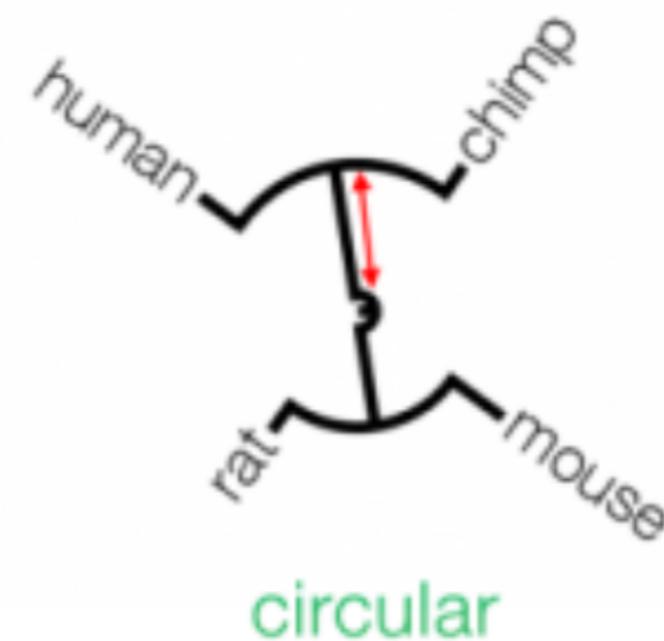
curved



radial



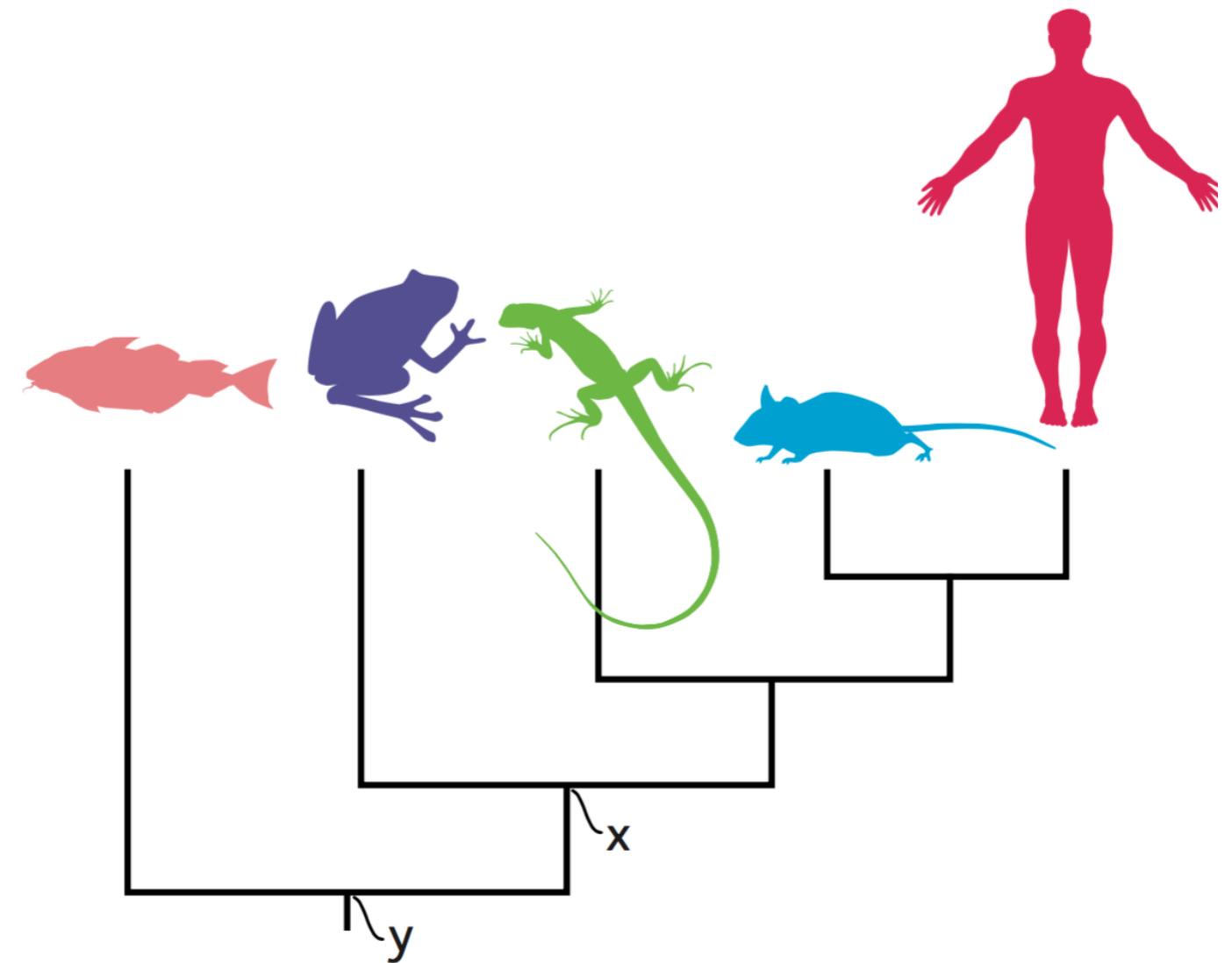
root not shown



circular

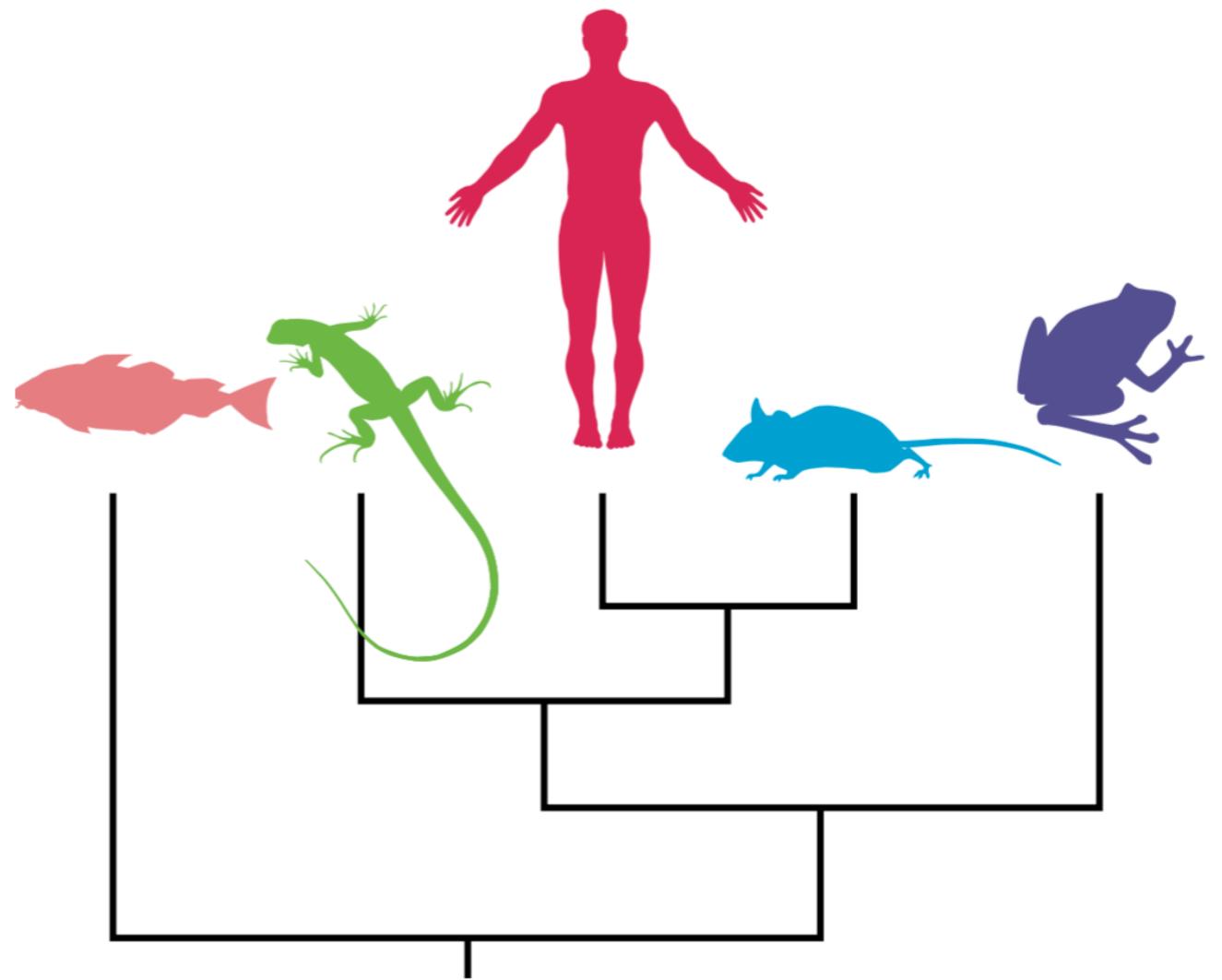
¿Cómo “leer” una filogenia?

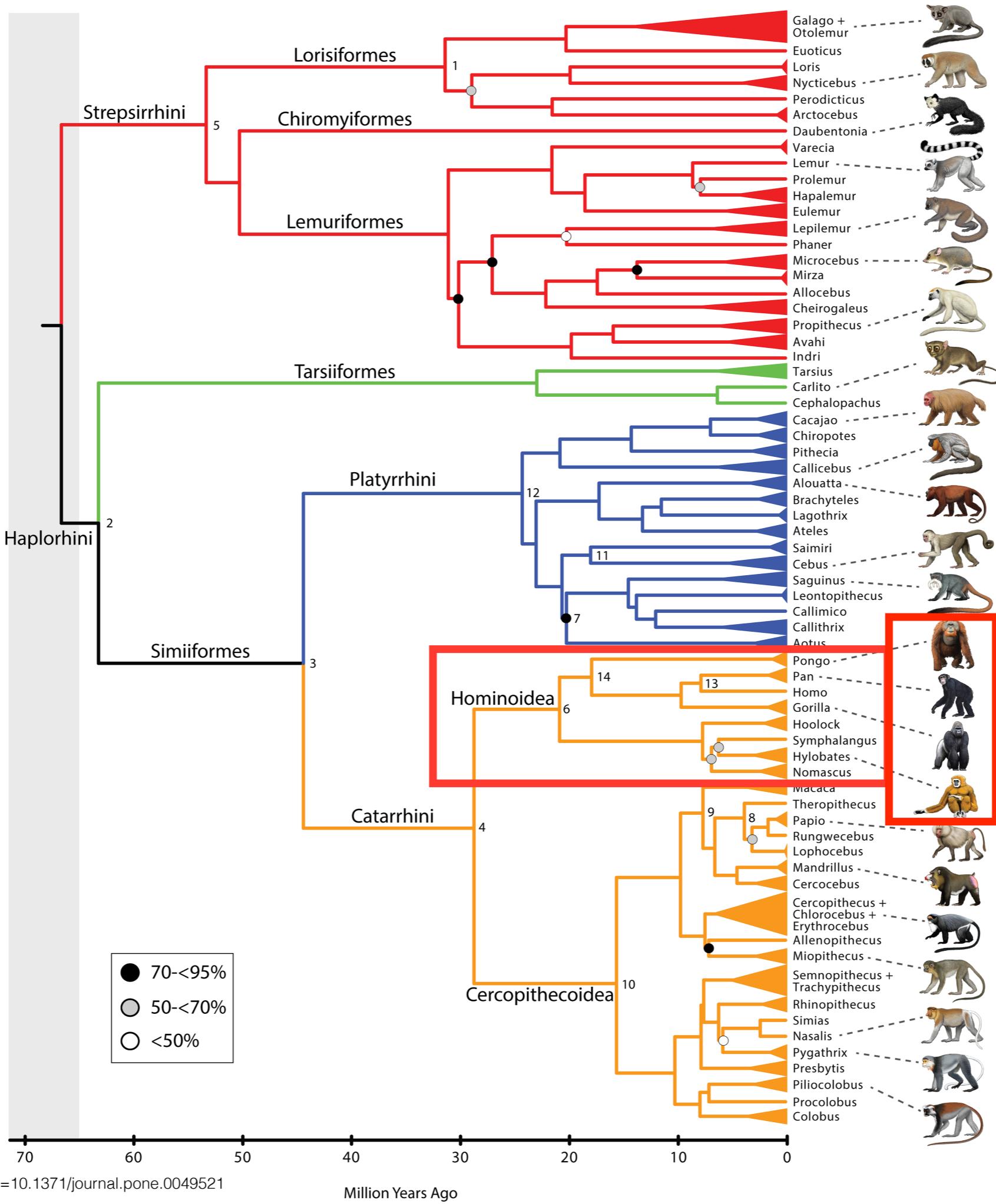
- Basándose en el árbol ¿la rana es más cercana al pez o al humano?



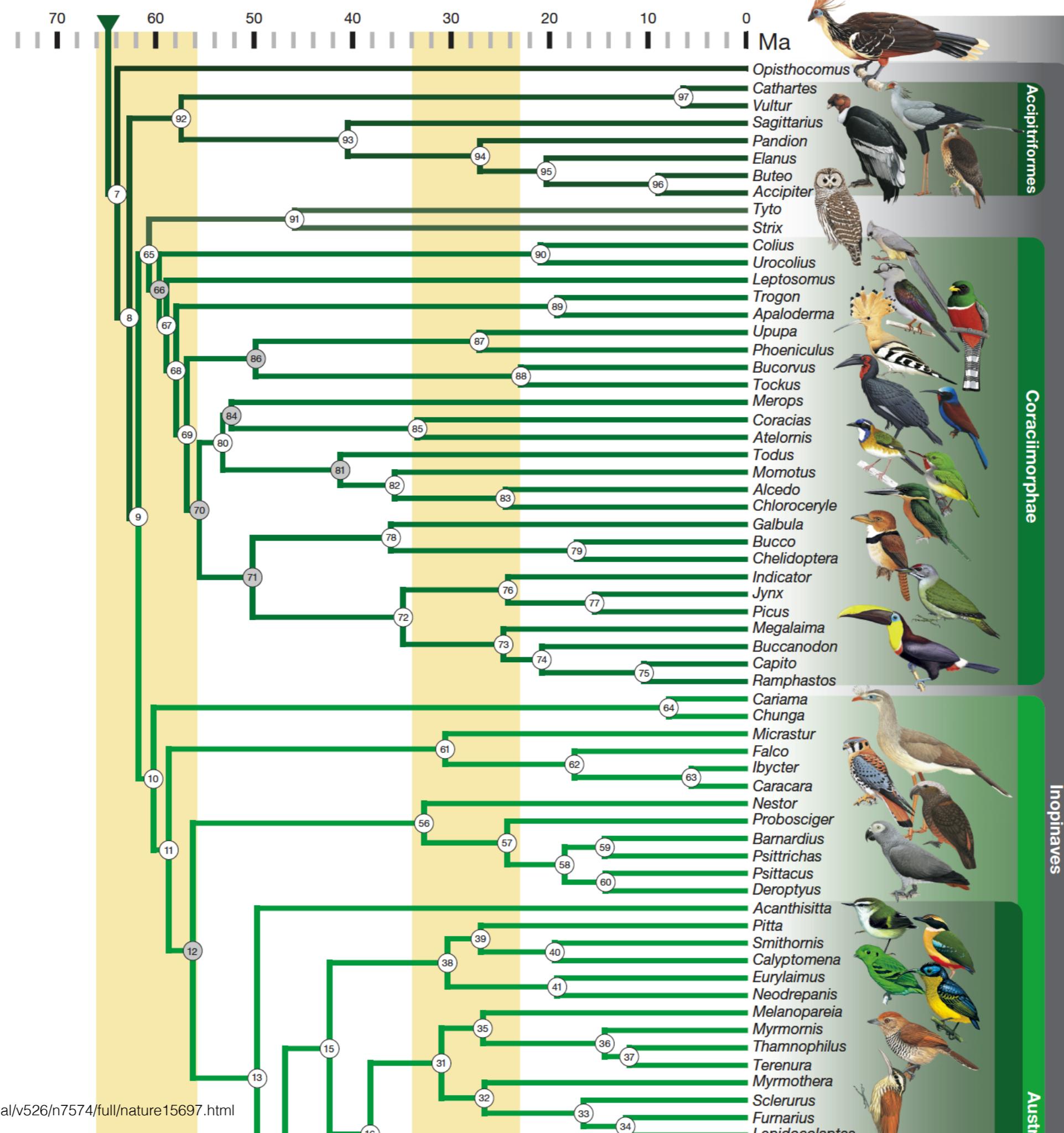
¿Cómo “leer” una filogenia?

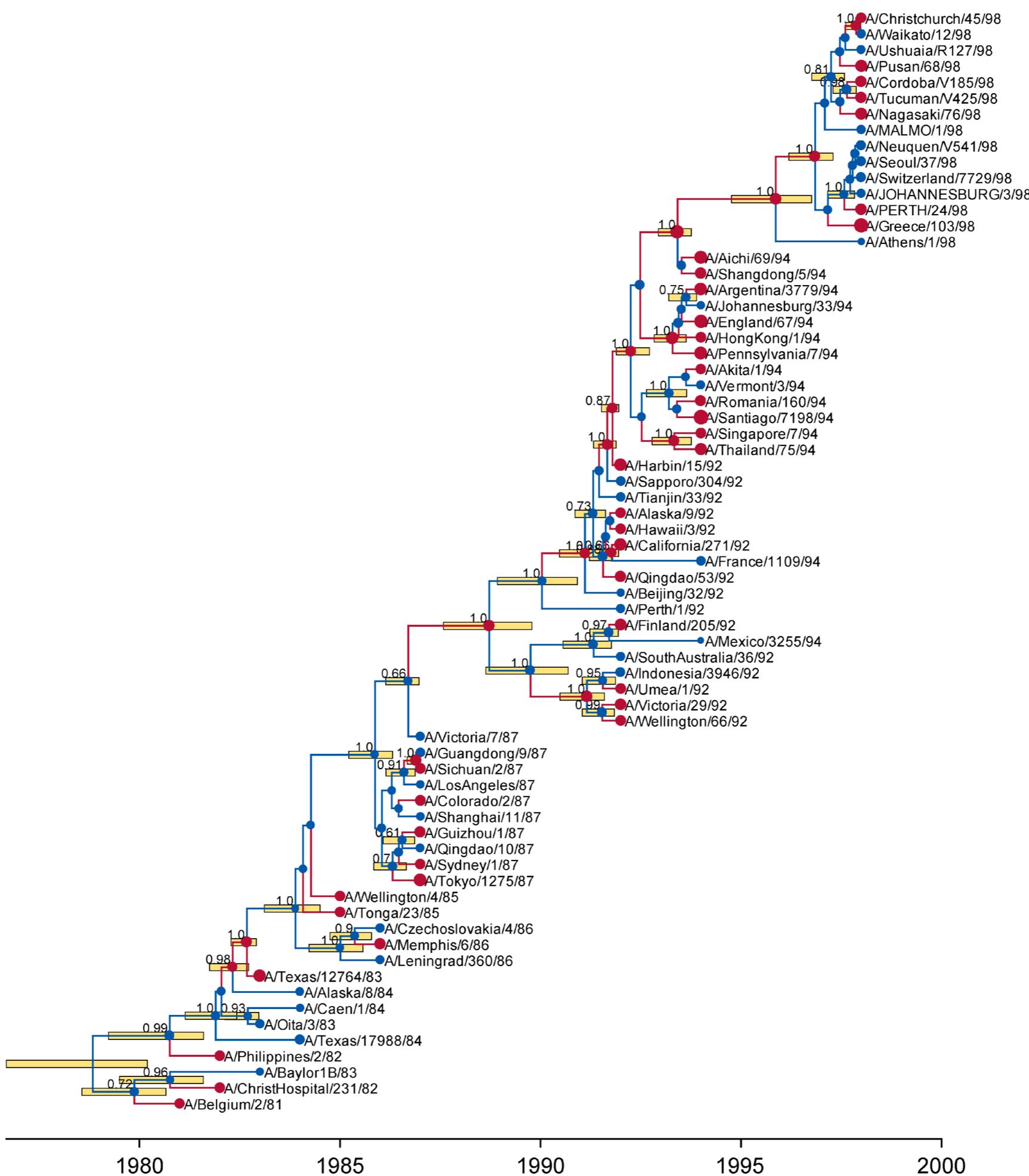
- ¿Y ahora?





Neoaves continued





```
install.packages(); biocLite()
```

- ggtree
- VennDiagram
- pheatmap