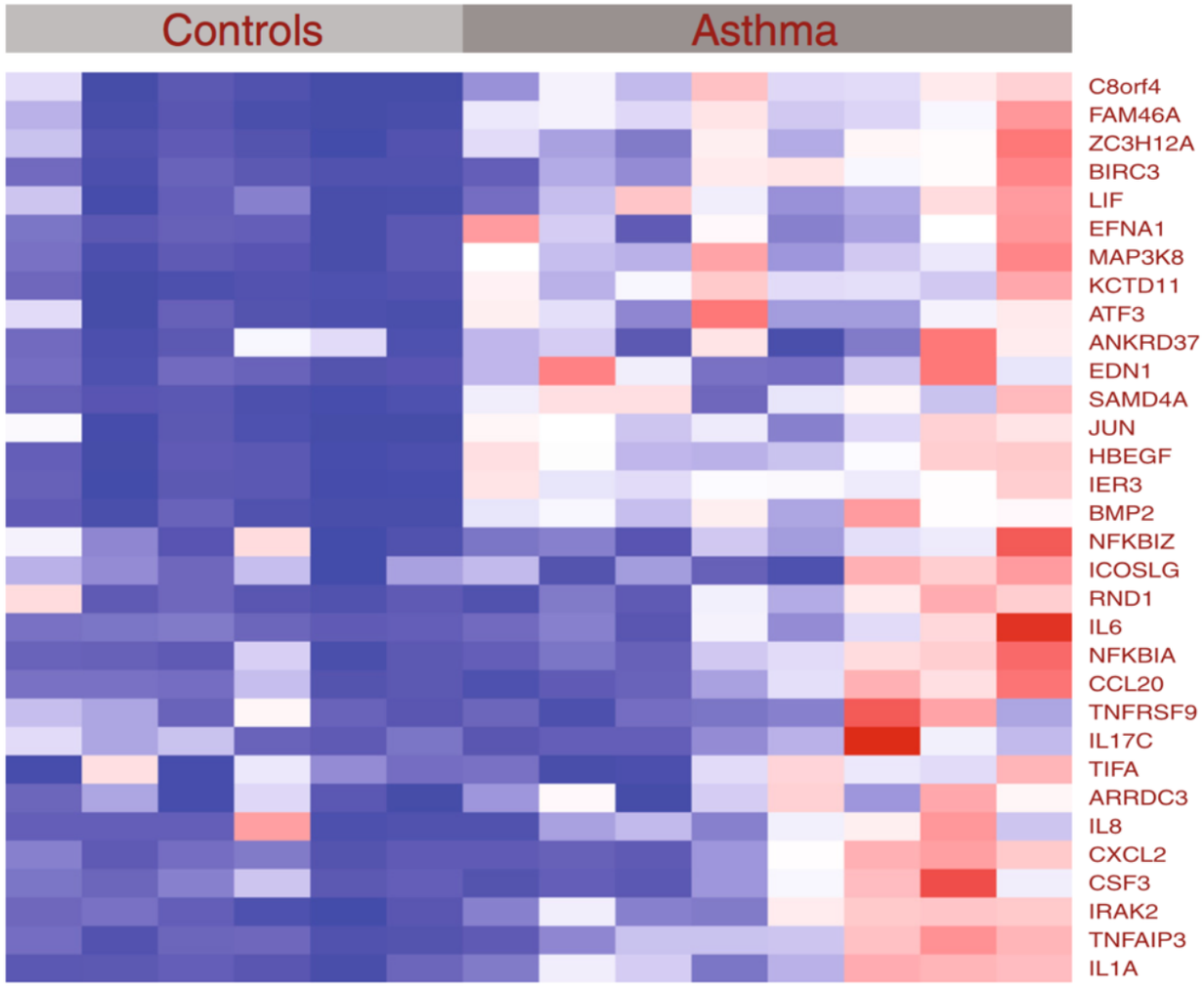
Mapas de calor, diagramas de Venn, filogenias

Visualización Científica
17 de octubre de 2016

Eduardo Castro-Nallar, PhD
Center for Bioinformatics and Integrative Biology
www.cbib.cl
www.castrolab.org

Mapas de calor o heatmaps

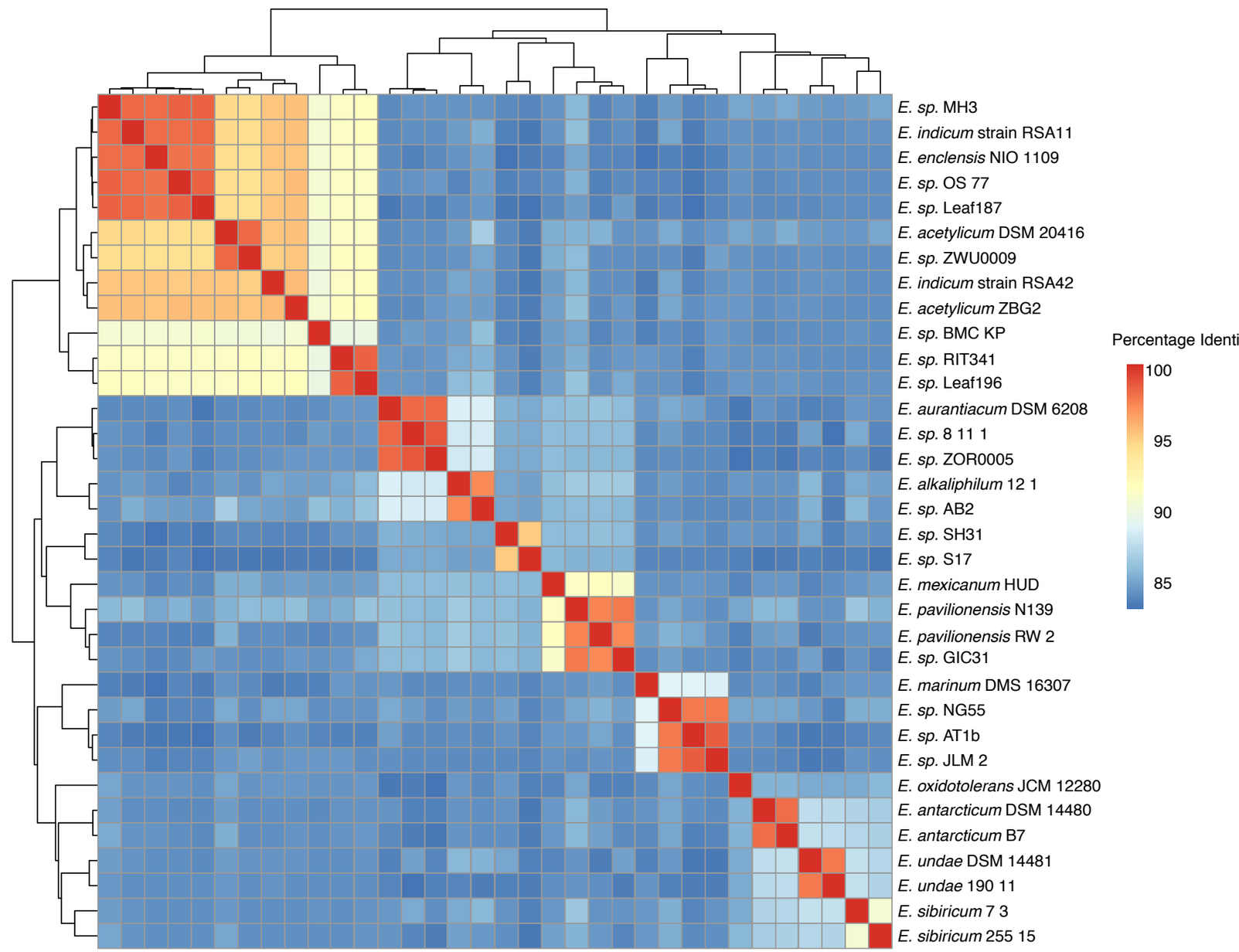
- Permiten resumir una gran cantidad de datos (data points) -> eficientes
- Comparación múltiple -> datos multivariados
- Muy útiles en genómica y metagenómica



Castro-Nallar, E., Shen, Y., Freishtat, R.J., Pérez-Losada, M., Manimaran, S., Liu, G., Johnson, W.E. and Crandall, K.A., 2015. Integrating microbial and host transcriptomics to characterize asthma-associated microbial communities. *BMC medical genomics*, *8*(1), p.1.

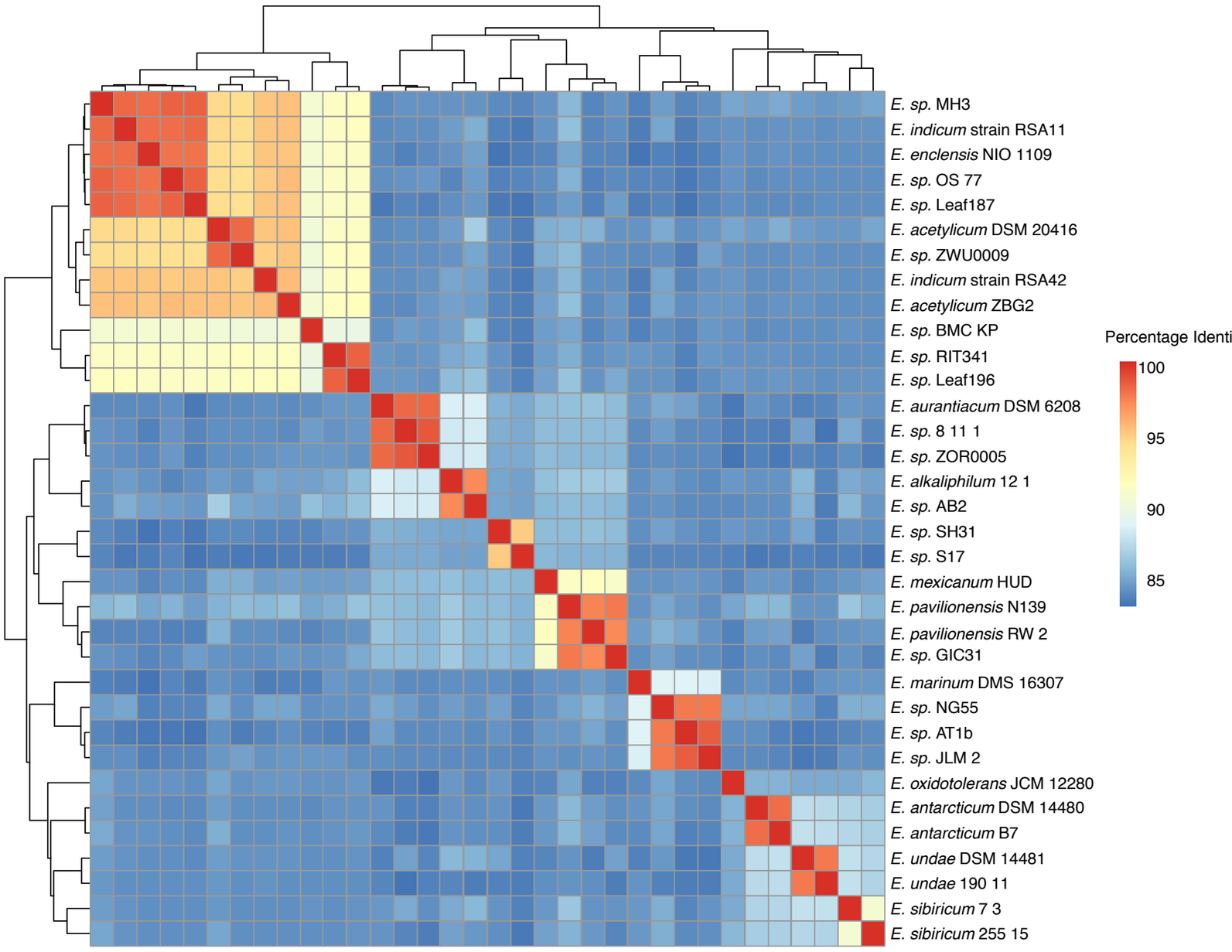
Heat maps representan tablas bidimensionales

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



Aspectos más relevantes

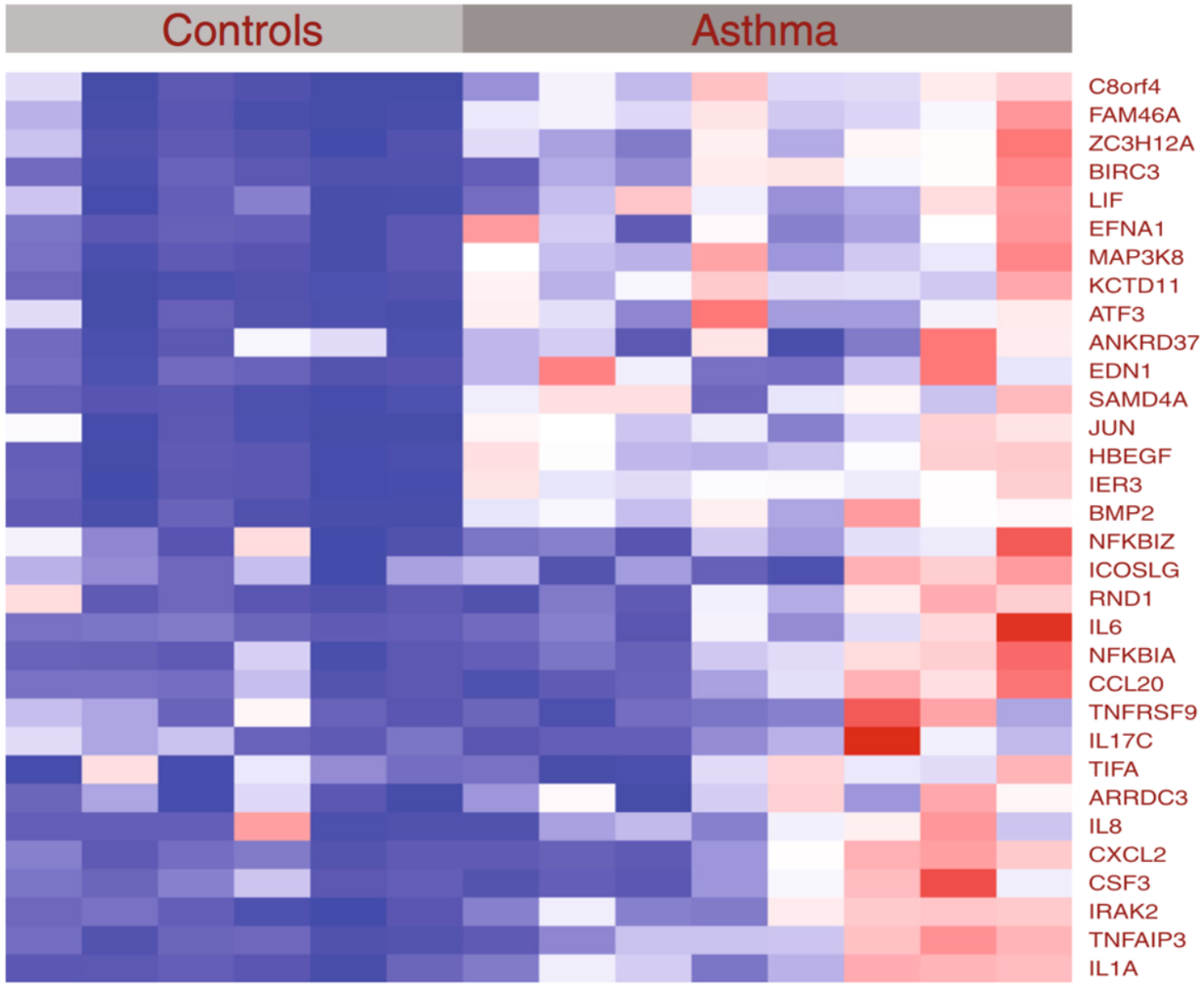
- El esquema de colores
- El orden de las filas/ columnas
- La magnitud que representan los colores



Juan Castro-Severyn, Francisco Remonsellez, Sandro Luis Valenzuela, Jonathan Fortt, Pablo Aguilar, Coral Pardo-Esté, Cristina Dorador, Raquel Quatrini, Franck Molina, **Eduardo Castro-Nallar** and Claudia P Saavedra. 2016. Allegiance of a new *Exiguobacterium* sp. strain onto a divided genus: a comparative genomic study. Frontiers in Microbiology. In review

Esquema de color

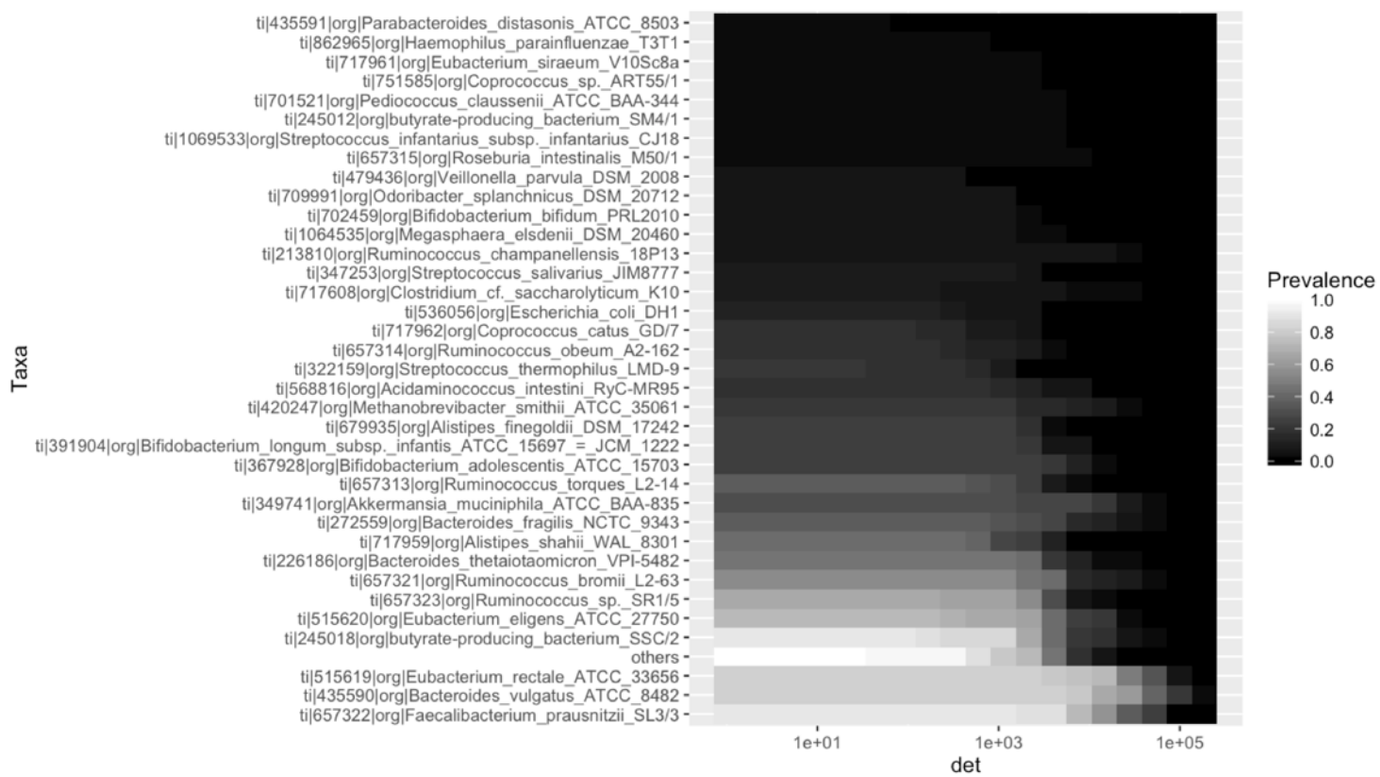
- Para destacar diferencias, se prefiere un mapa de colores divergente
- De azul a blanco a rojo



Castro-Nallar, E., Shen, Y., Freishtat, R.J., Pérez-Losada, M., Manimaran, S., Liu, G., Johnson, W.E. and Crandall, K.A., 2015. Integrating microbial and host transcriptomics to characterize asthma-associated microbial communities. *BMC medical genomics*, *8*(1), p.1.

Esquema de color

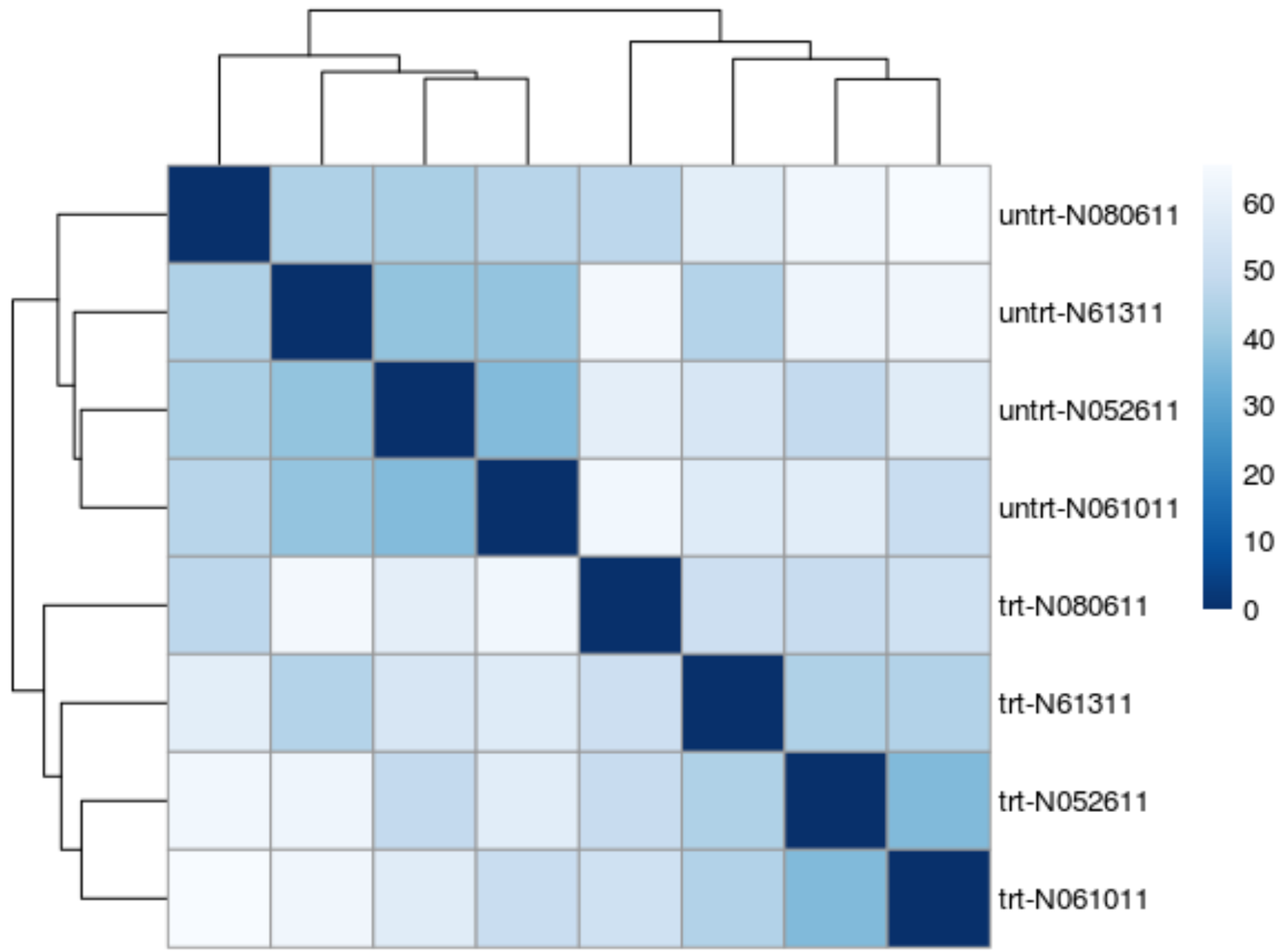
- En cambio para mostrar intensidad es mejor usar tonalidades de un mismo color
- Ejemplo clásico es la escala de grises



Castro-Nallar et al. PathoStat Bioconductor package

Heat maps: muestra a muestra

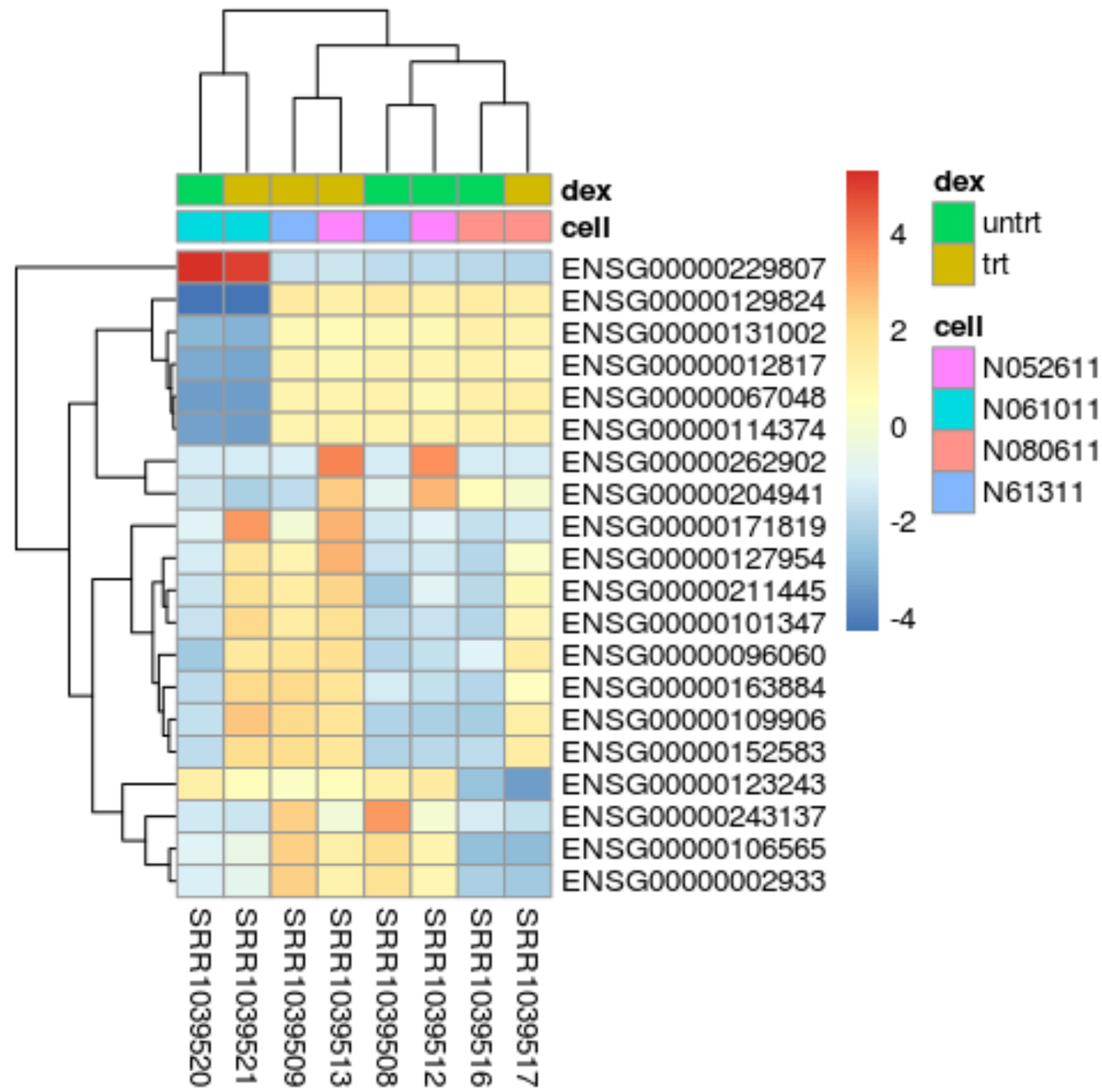
- Experimentos de transcriptómica
- ¿Qué muestras son más similares entre sí?
- Expectativa: muestras con tratamientos similares tienen expresión génica similar



http://www.bioconductor.org/help/workflows/rnaseqGene/#gene-clustering

Heat maps: muestra a gen

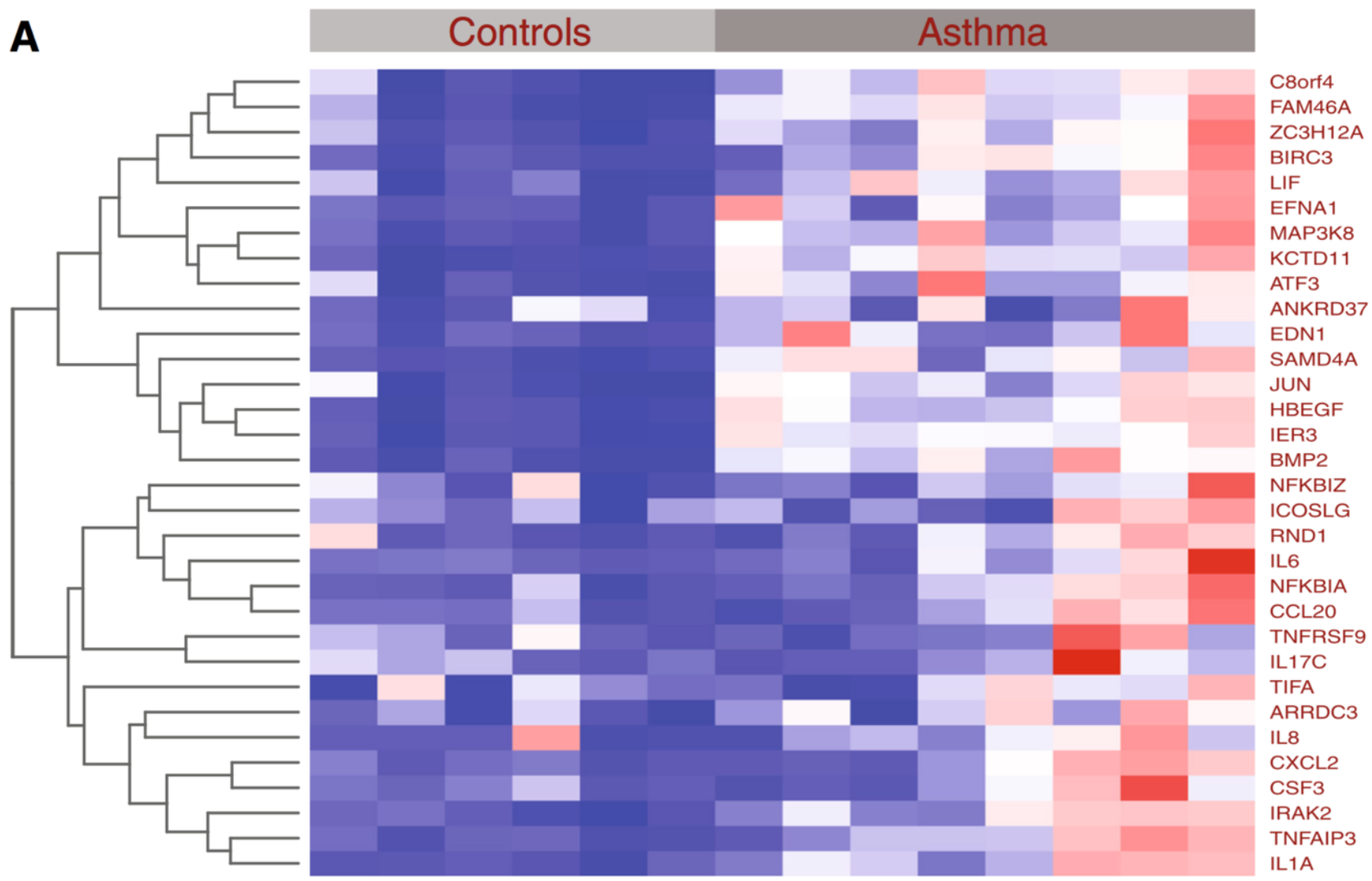
- Experimentos de transcriptómica
- ¿Qué genes cambian su expresión de manera similar entre muestras?
- Expectativa: muestras cuyos genes cambian su expresión de manera similar están relacionadas biológicamente



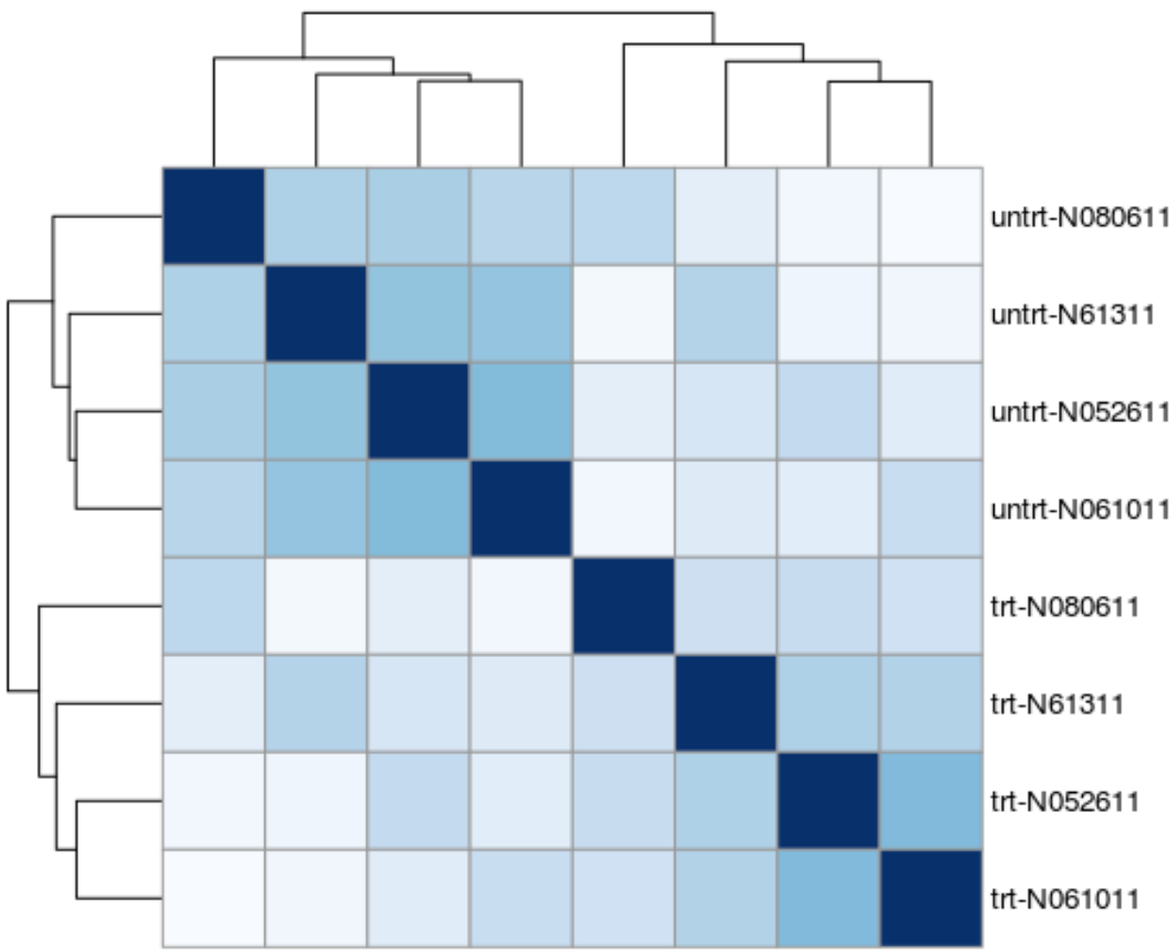
http://www.bioconductor.org/help/workflows/rnaseqGene/#gene-clustering

¿Cómo funciona el clustering?

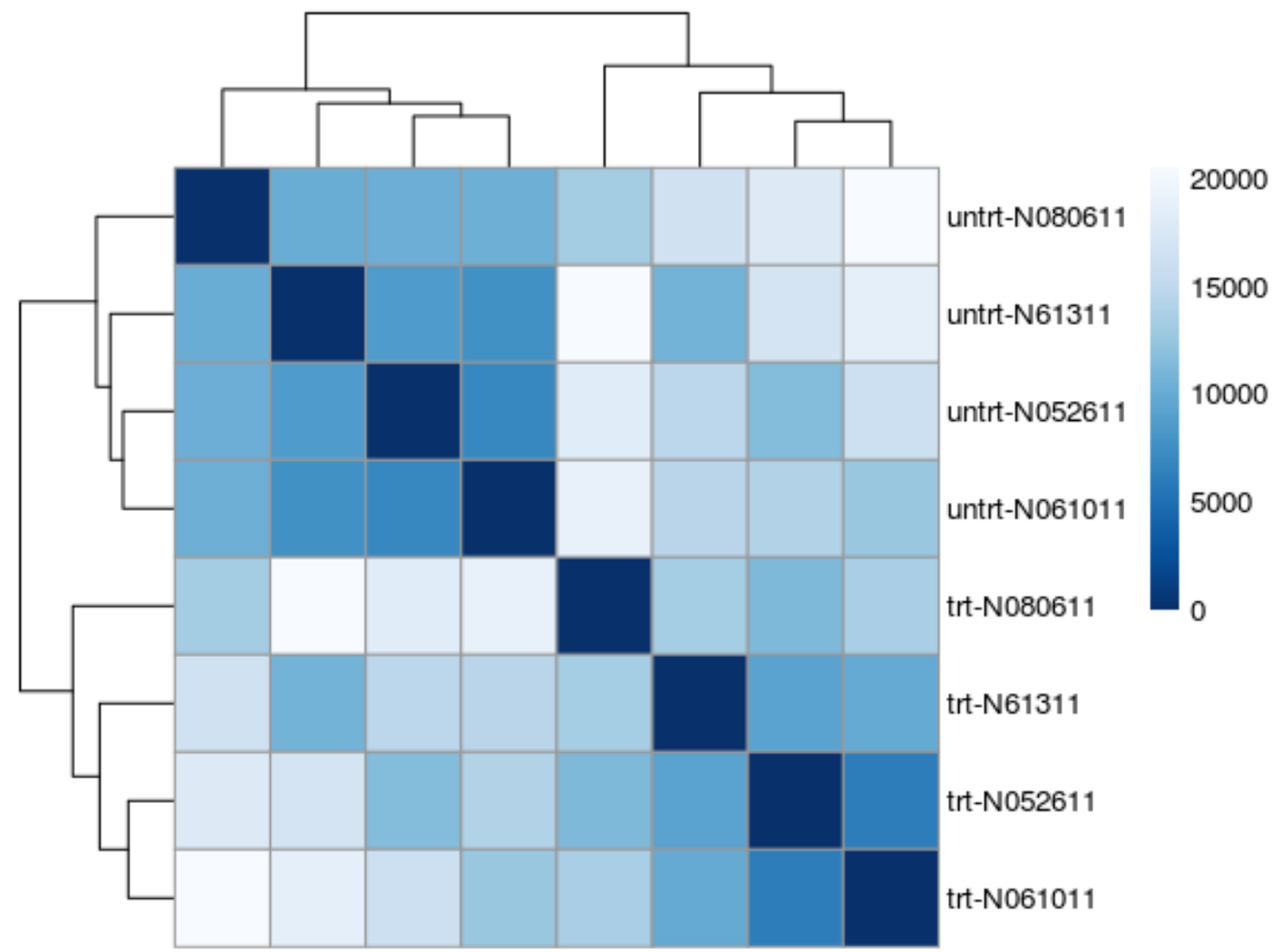
- 1. Calcular distancia “todos contra todos”. Distancia es qué tan diferentes dos muestras son
- 2. Juntar puntos menos distantes entre sí
- 3. Juntar grupos de puntos
- 4. Dendrograma es la representación gráfica del clustering



Método de distancia



Distancia Euclidiana



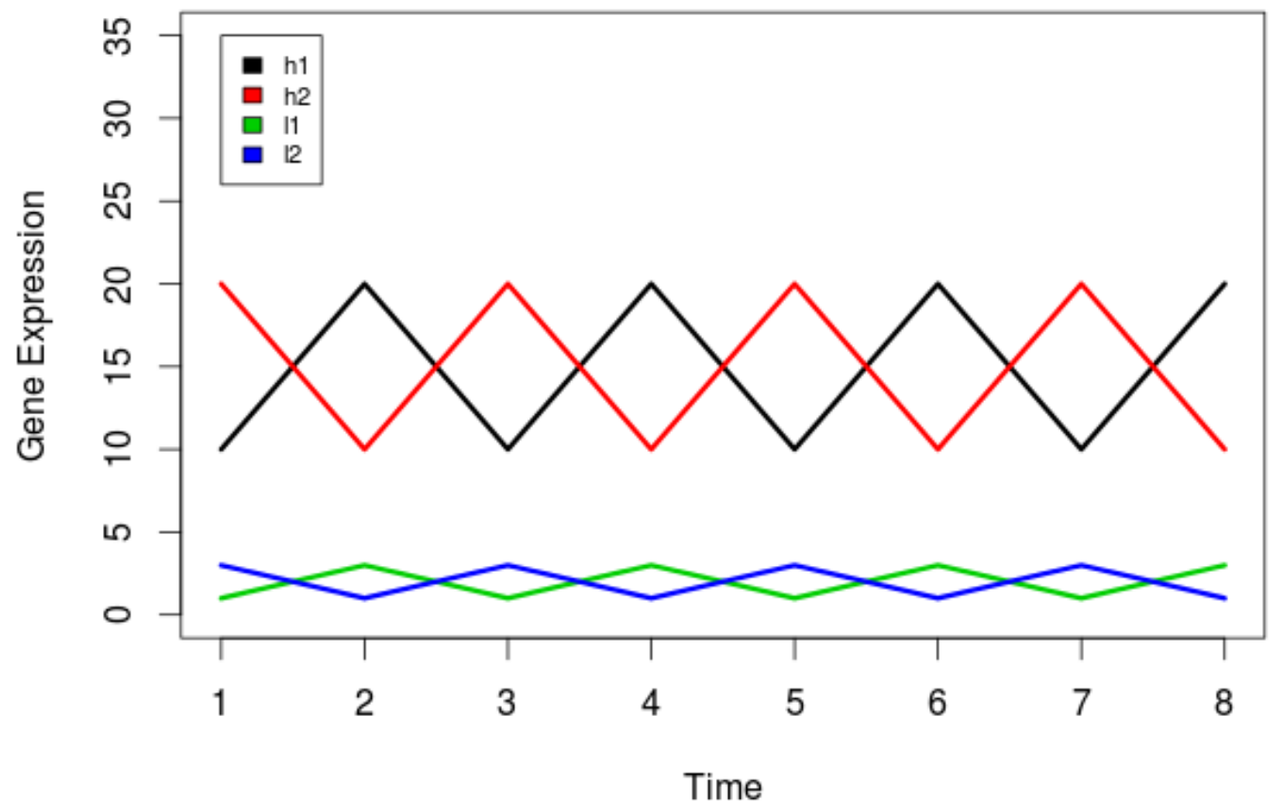
Distancia de Poisson

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

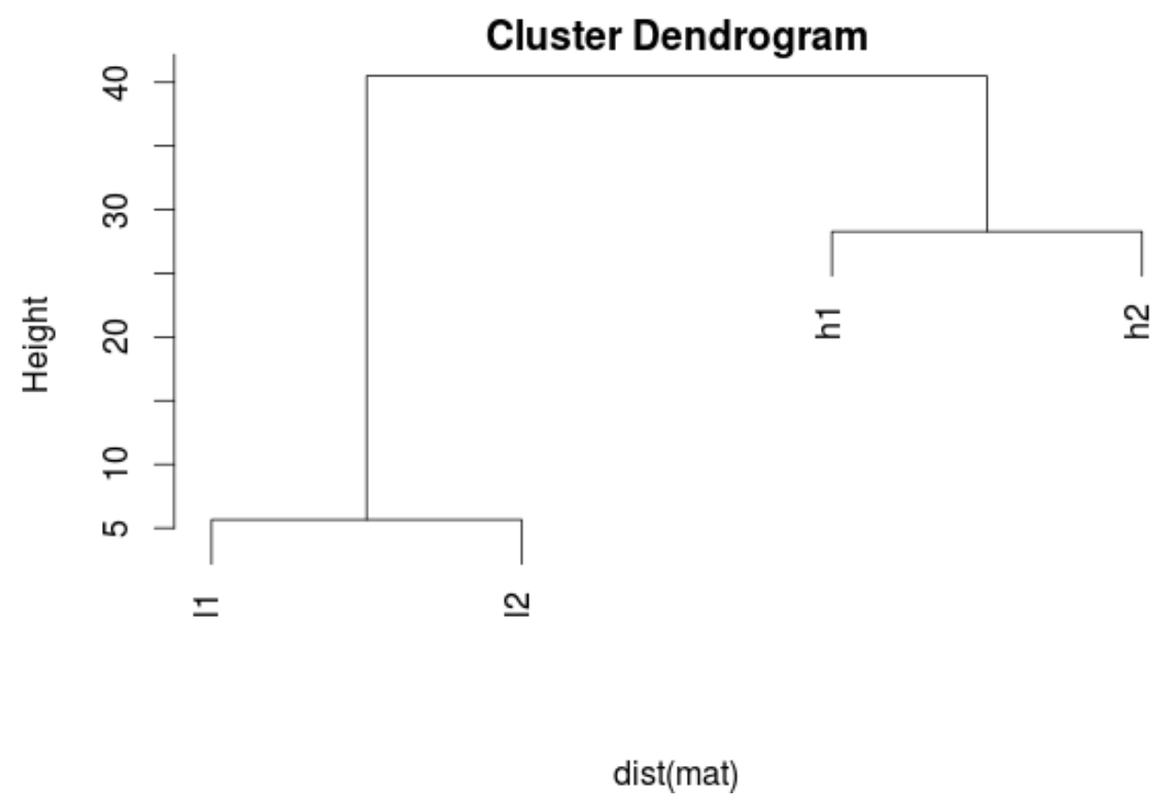


http://www.opiniomics.org/you-probably-dont-understand-heatmaps/

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

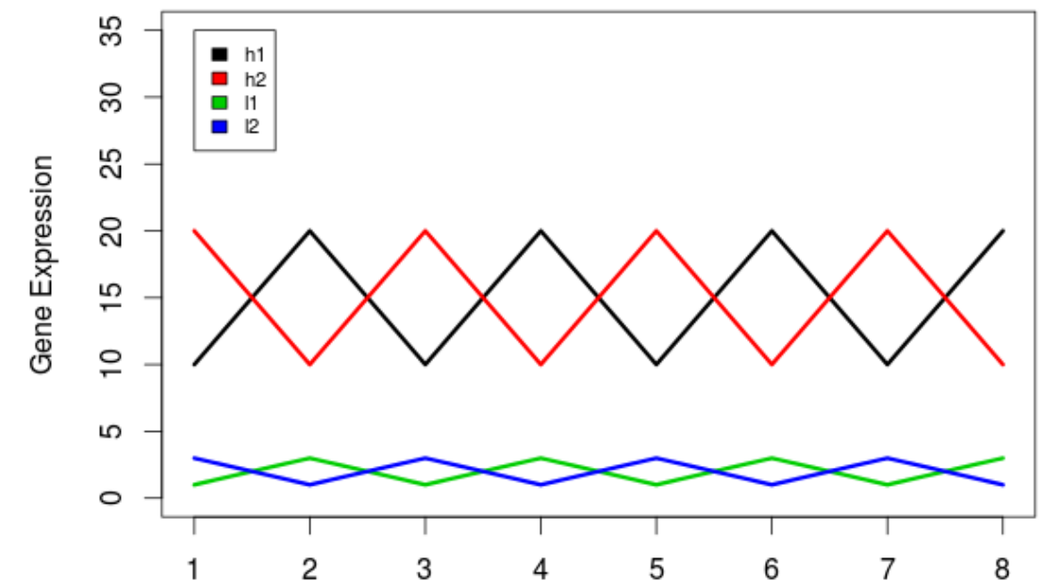
- Dos genes de expresión alta y dos de baja

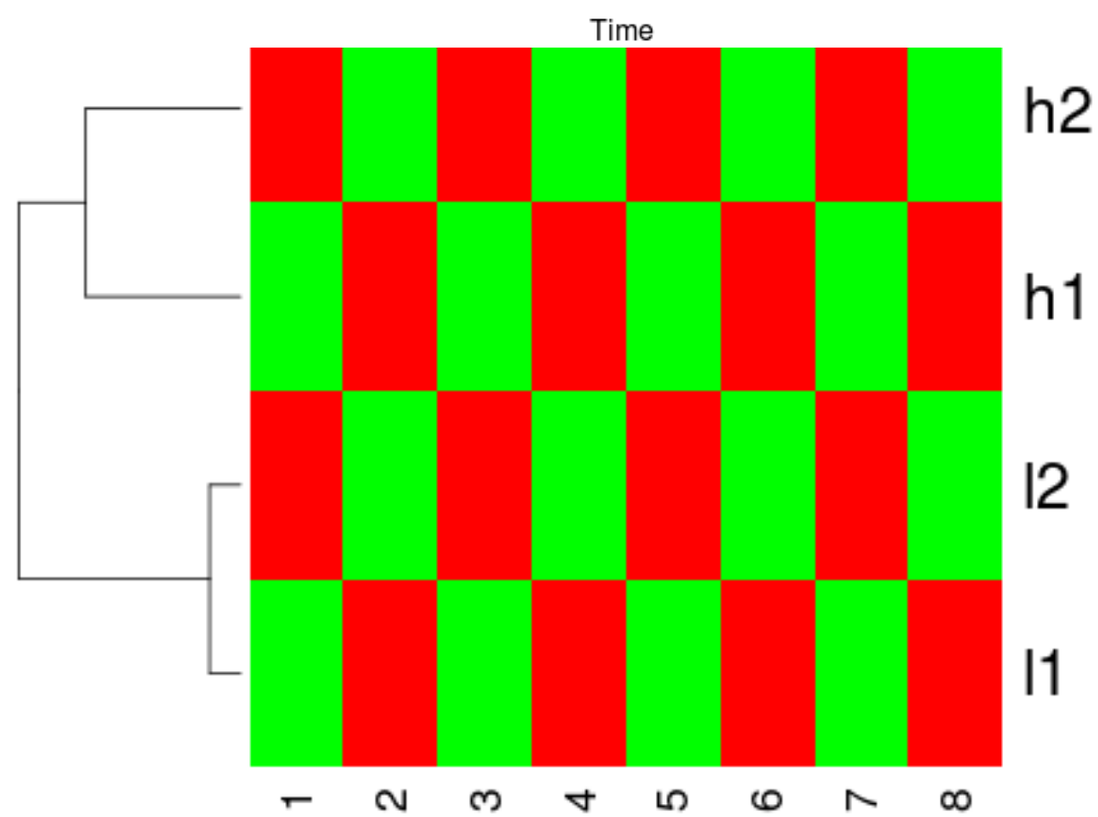


http://www.opiniomics.org/you-probably-dont-understand-heatmaps/

Distancia: queremos representar la forma de los datos

- El problema con este heat map es que tradujo los números literalmente en colores
- l1 y l2 forman clusters pero sus colores son opuestos
- l1 y h1 tienen los mismos colores, sin embargo su expresión es diferente

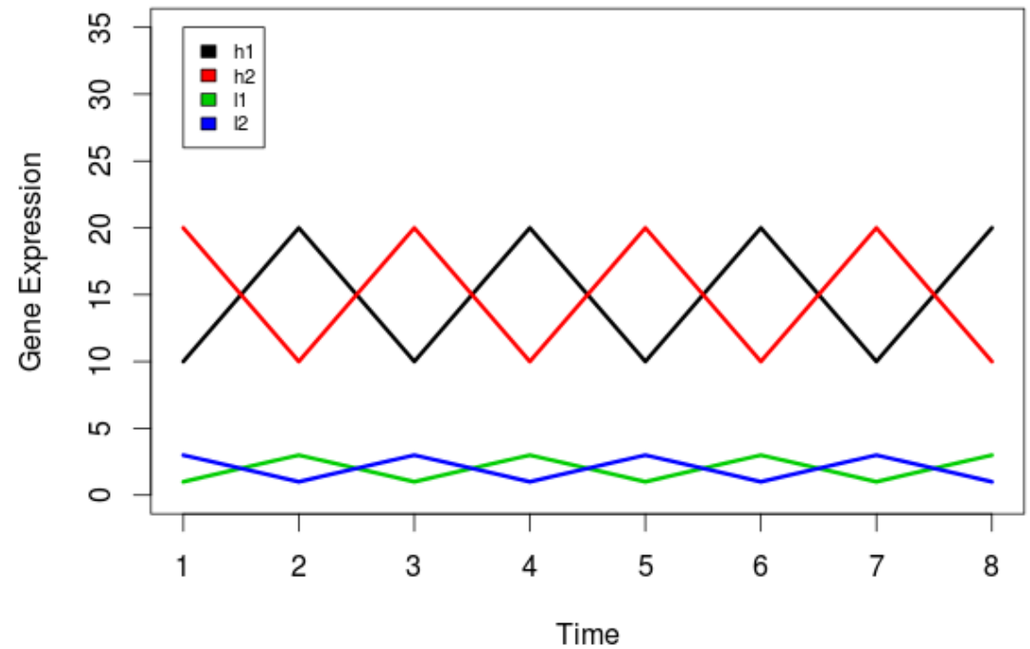


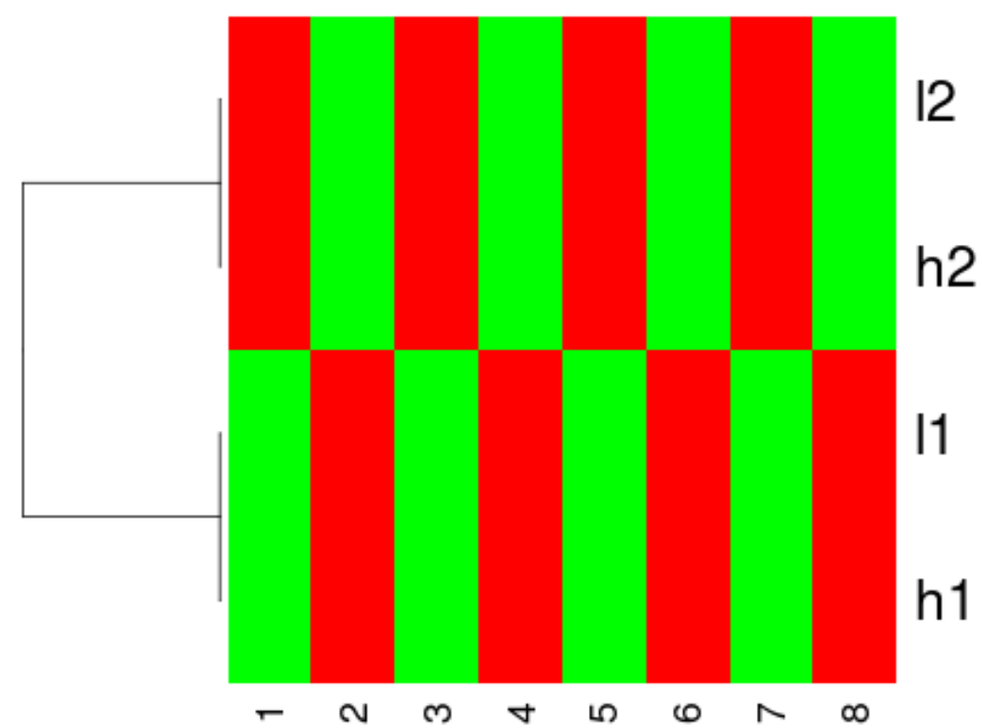


http://www.opiniomics.org/you-probably-dont-understand-heatmaps/

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”

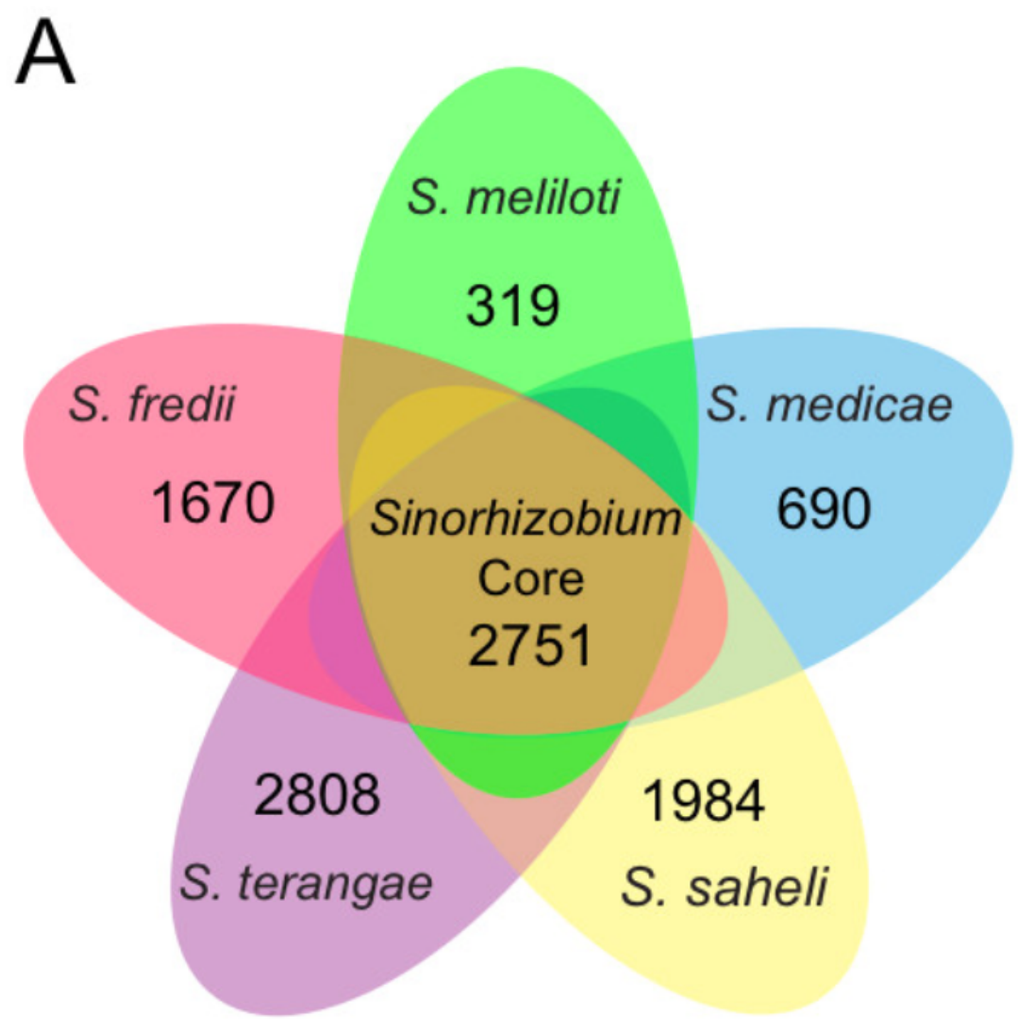




http://www.opiniomics.org/you-probably-dont-understand-heatmaps/

Diagramas de Venn

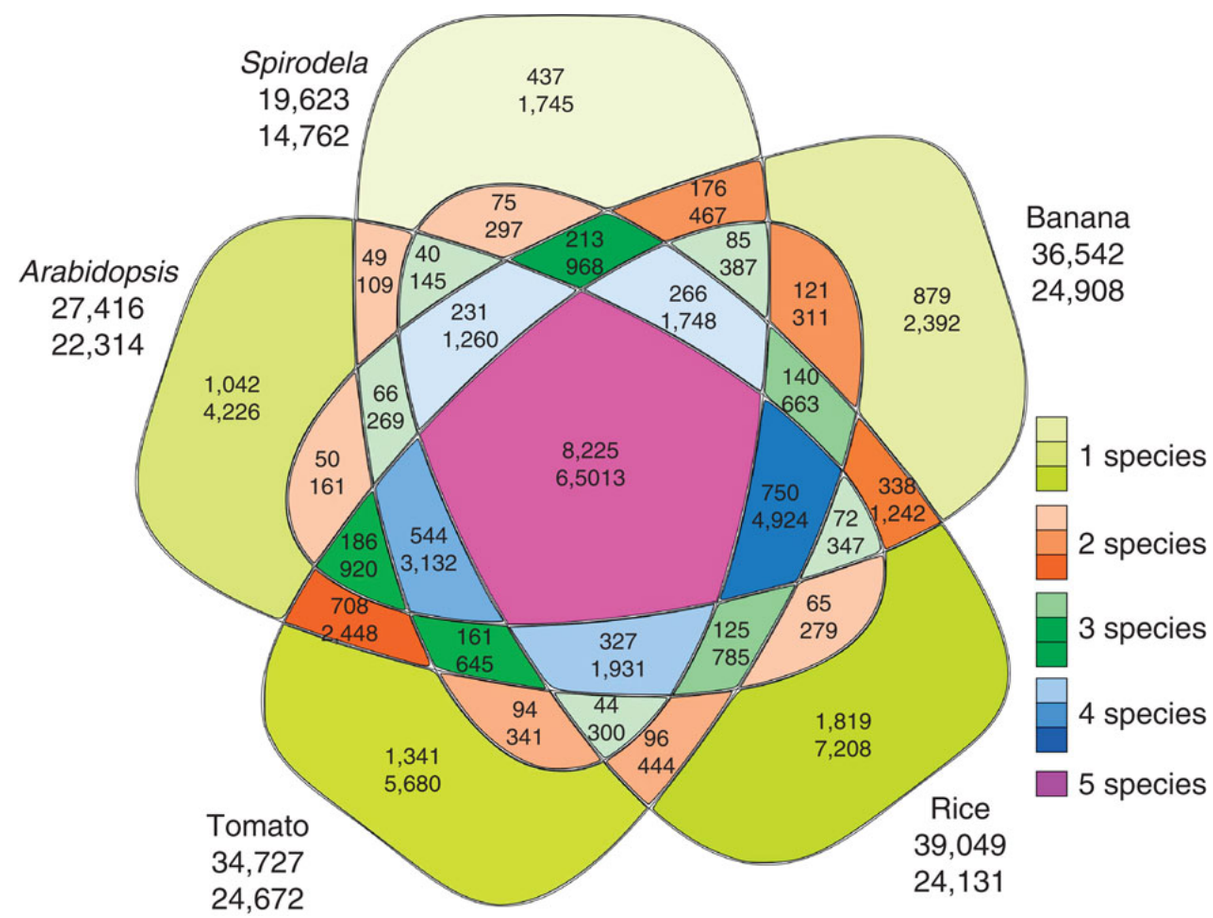
- Un diagrama que representa conjuntos
- Muestra todas las relaciones lógicas entre muestras
- Familias de genes comunes entre especies



https://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-2-r17#Sec15

Diagramas de Venn

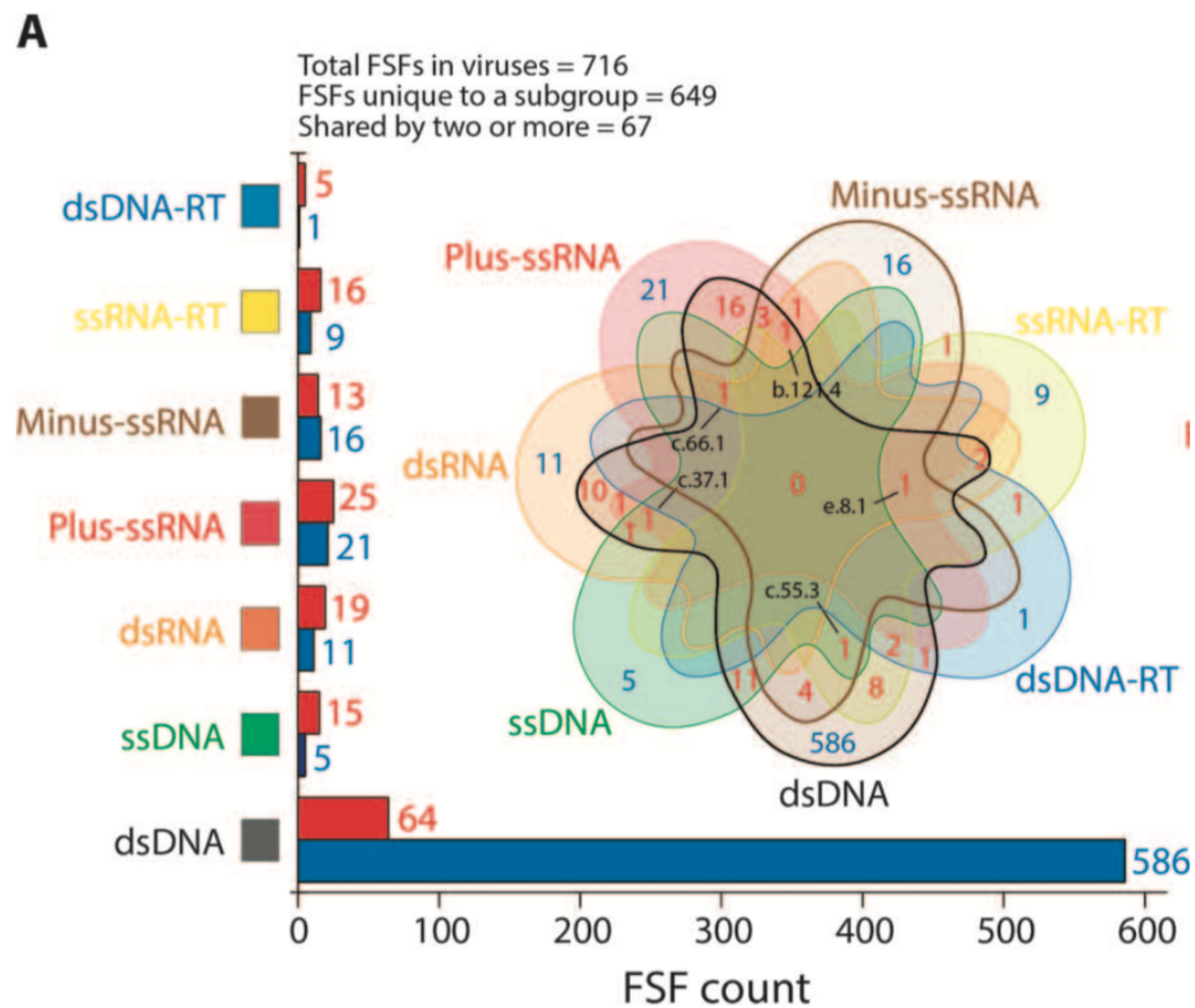
- Keep it simple
- Más de 5 conjuntos y la visualización comienza a jugar en contra de la ciencia



http://www.nature.com/articles/ncomms4311

Diagramas de Venn

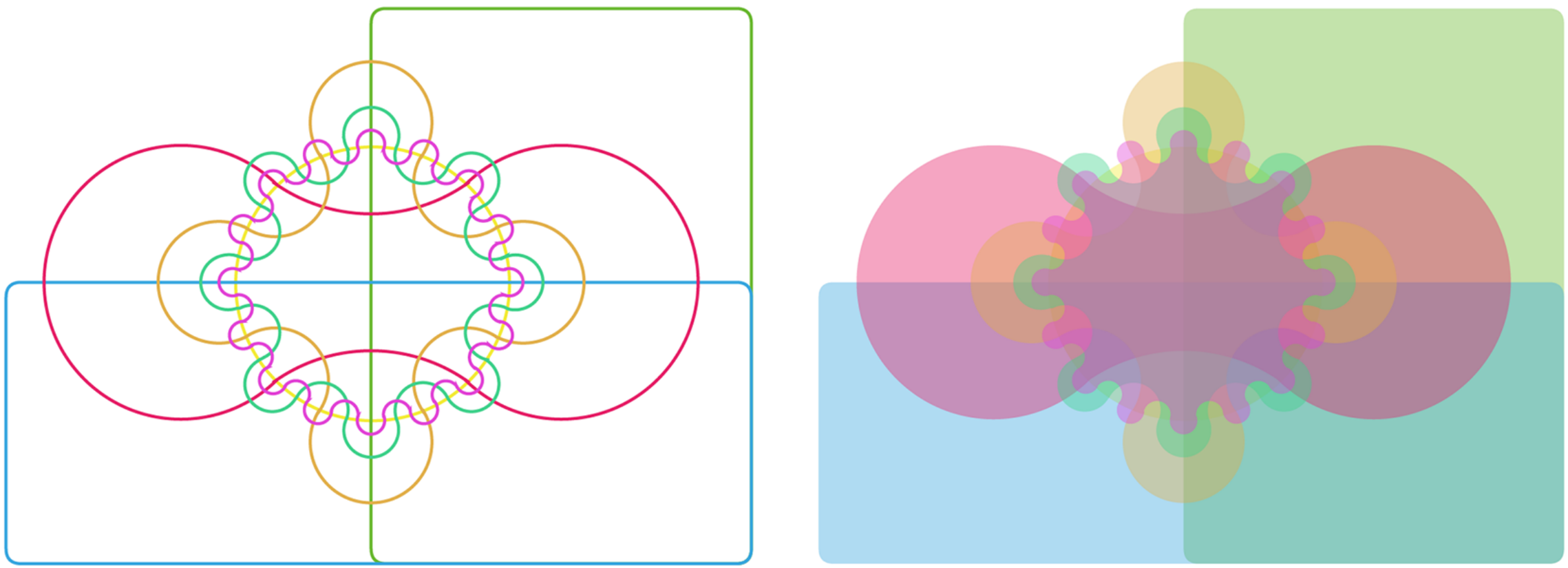
- Keep it simple
- Más de 5 conjuntos y la visualización comienza a jugar en contra de la ciencia



http://advances.sciencemag.org/content/advances/1/8/e1500527.full.pdf

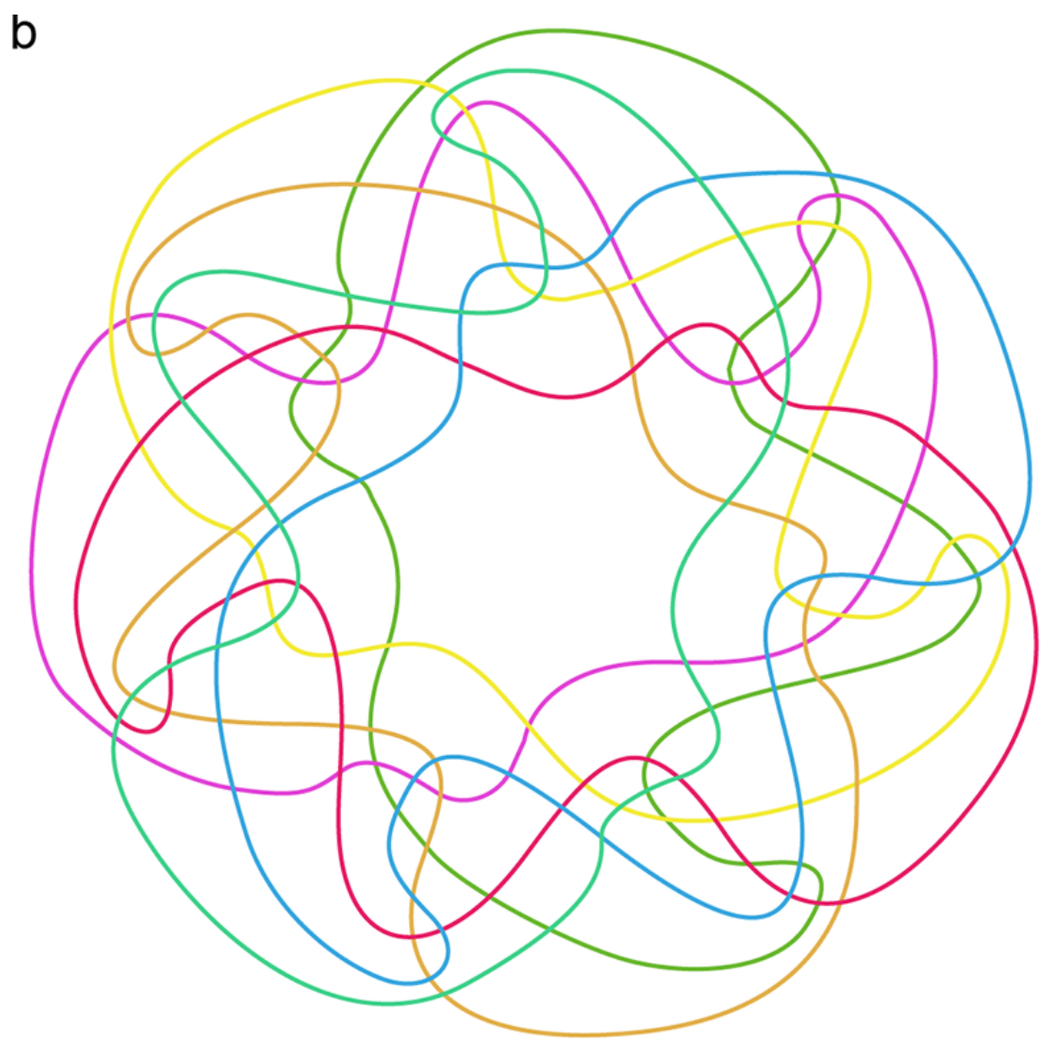
Variaciones a diagramas de Venn

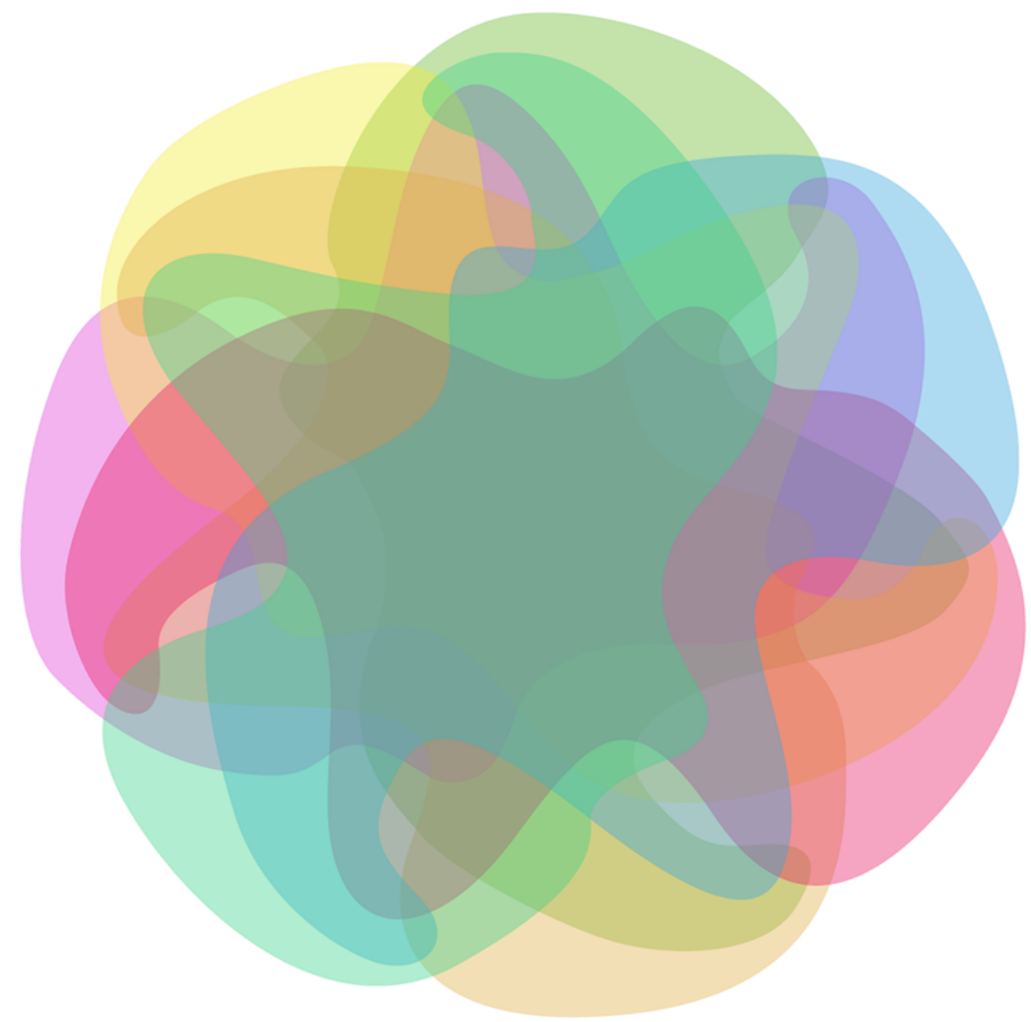
a



- Diagrama de Venn, tipo Edwards

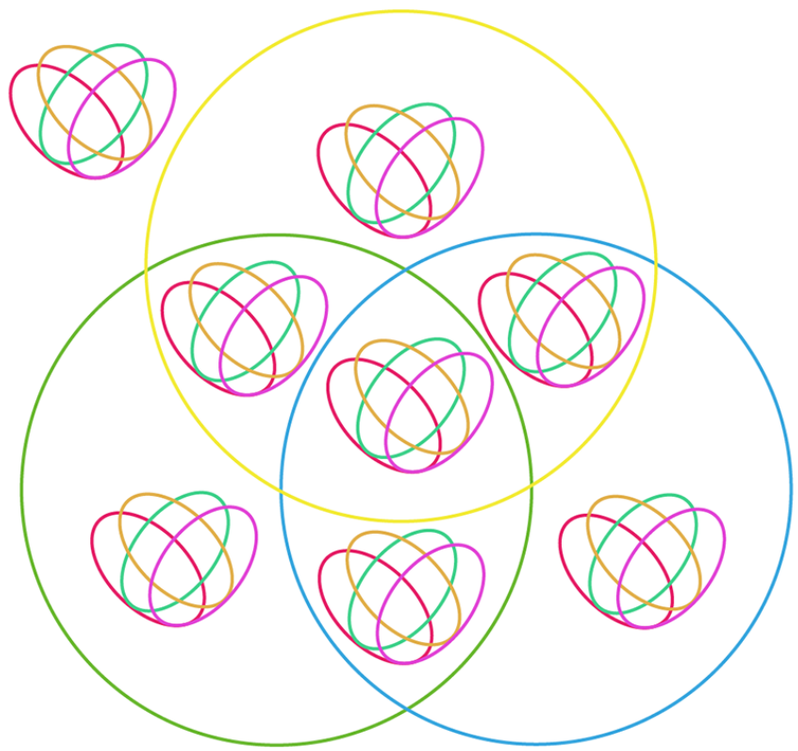
Variaciones a diagramas de Venn

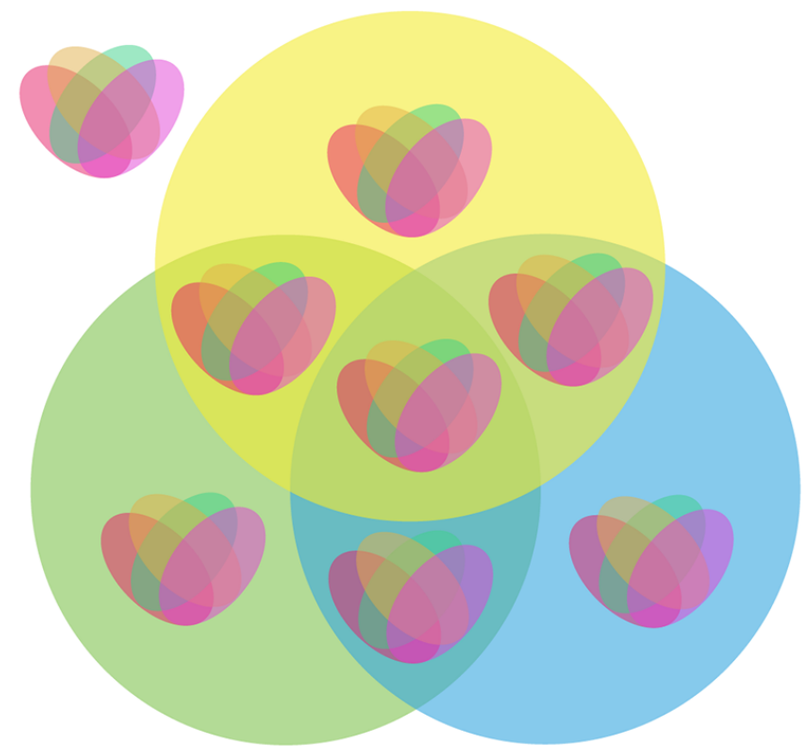




- Diagrama de Venn de curvas irregulares

Variaciones a diagramas de Venn



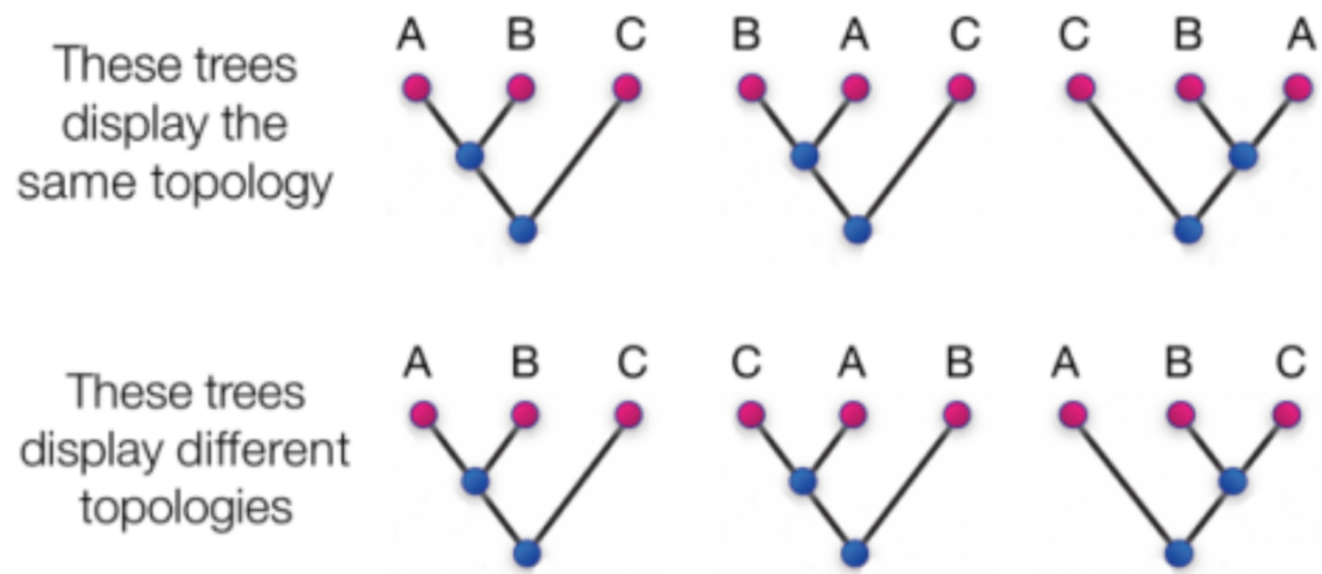


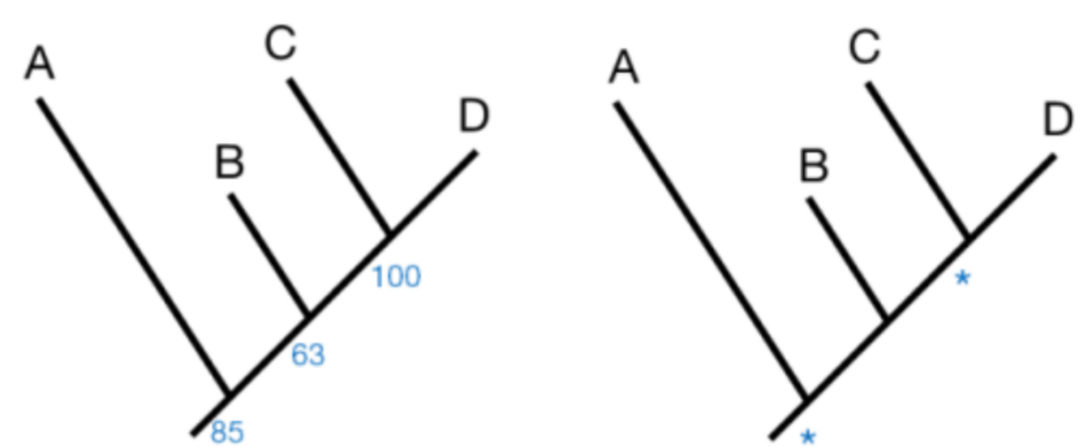
- Diagrama de Venn anidado
- Implementado en el software VennPainter

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154315

Características de las filogenias

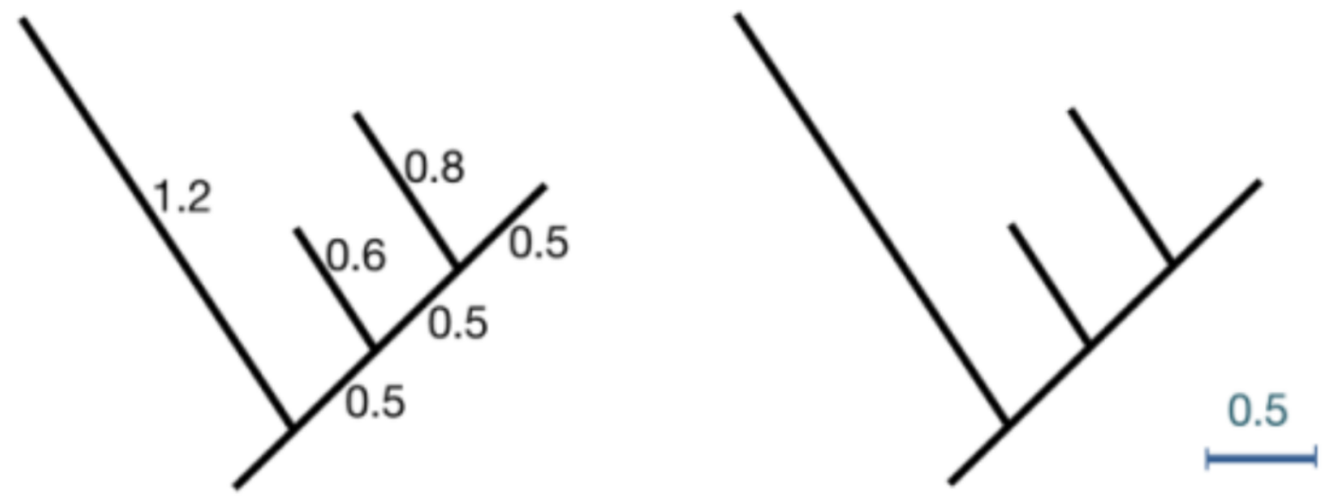
- Topología
- Ramas - branches (edges)
- Nodos - nodes (vertices)
    - Raíz - root
    - Nodos internos
    - Extremos - tips
- Confianza en nodos





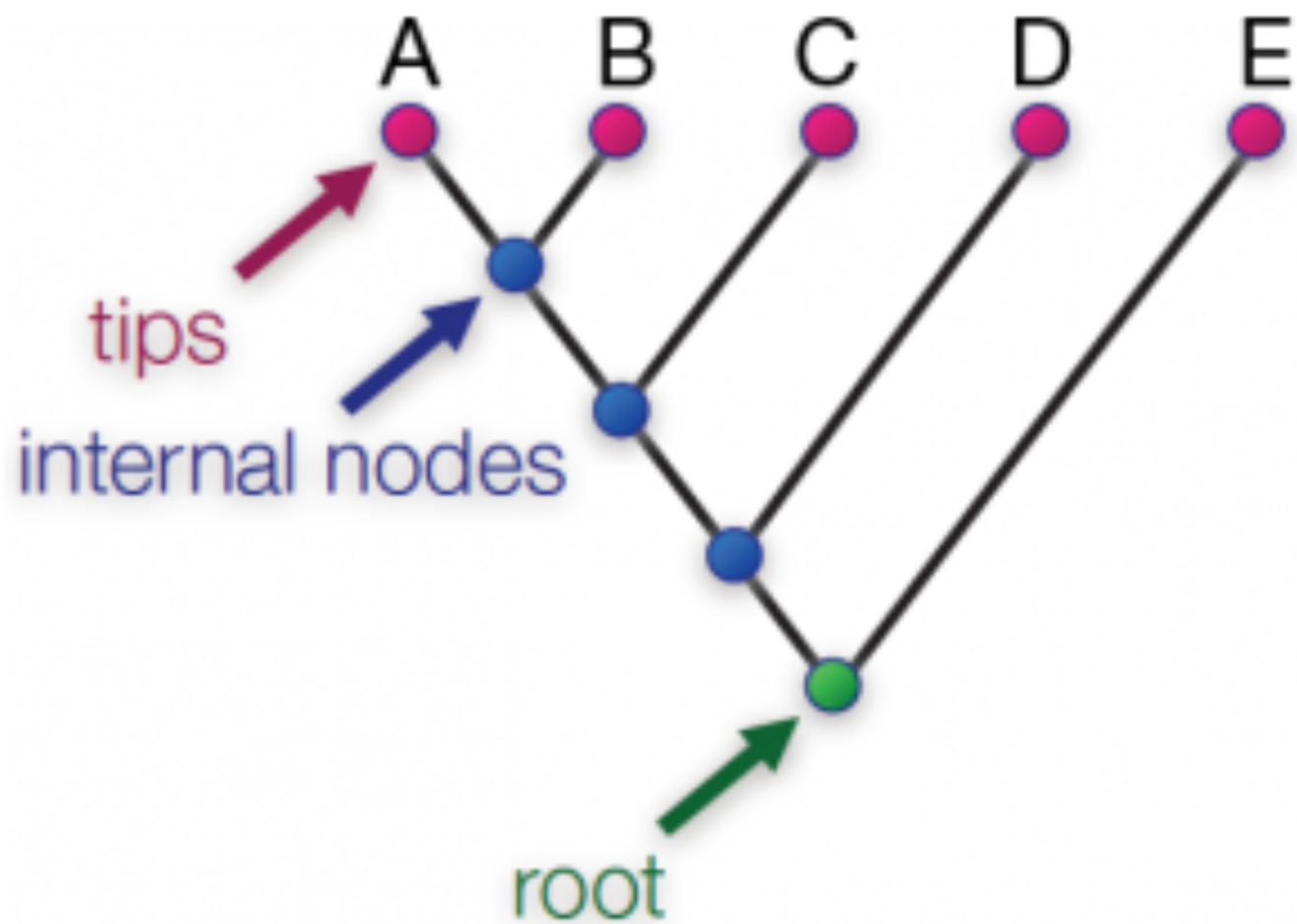
Características de las filogenias

- 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)



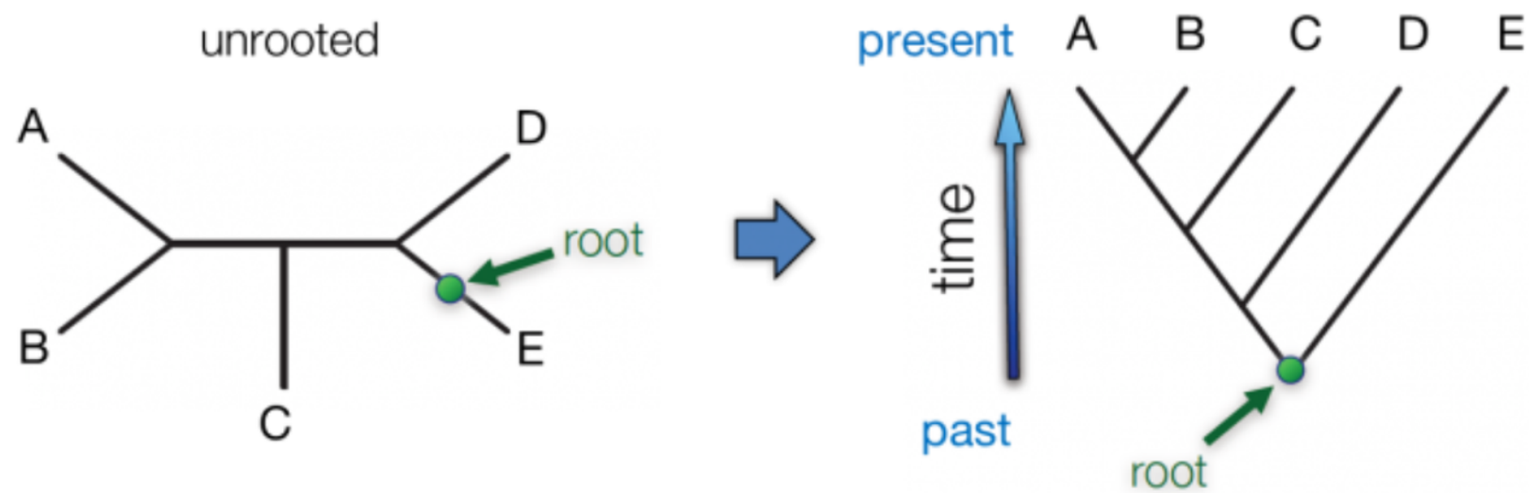
Características de las filogenias

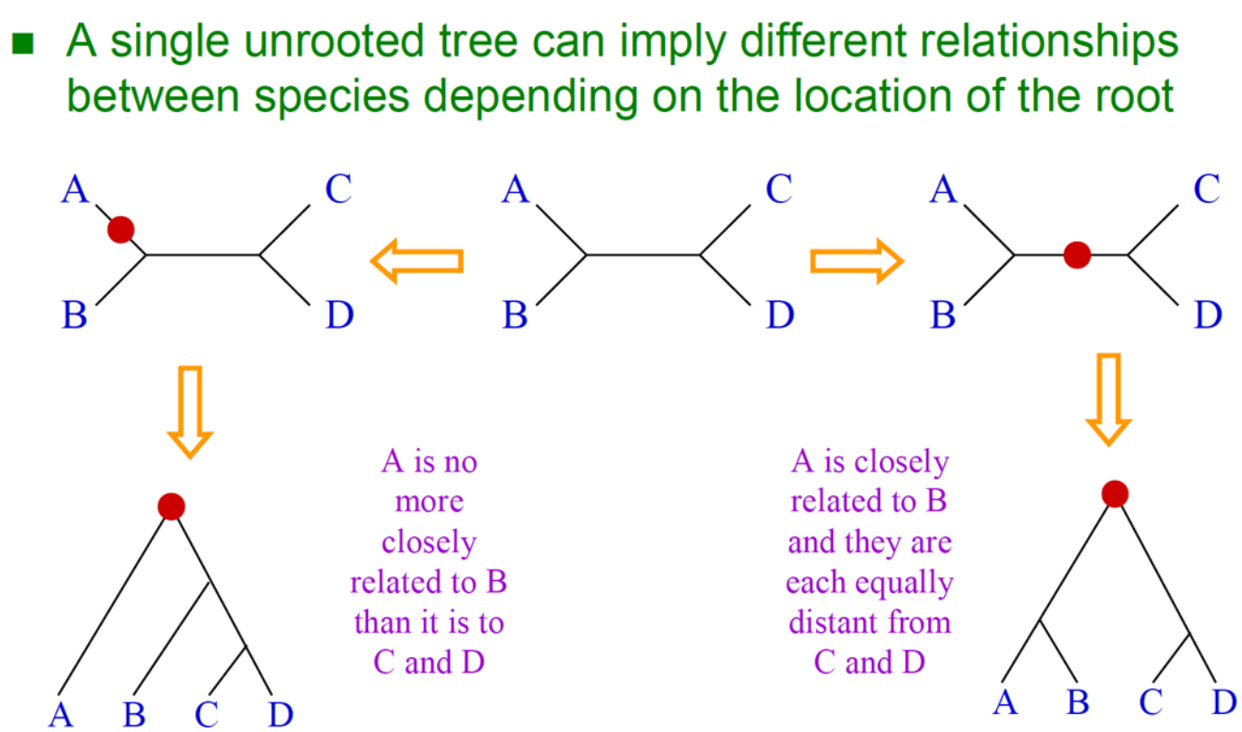
- Nodos
- Representan ancestros
- Mayoría de las veces no podemos observar a esos ancentros
- Excepciones?



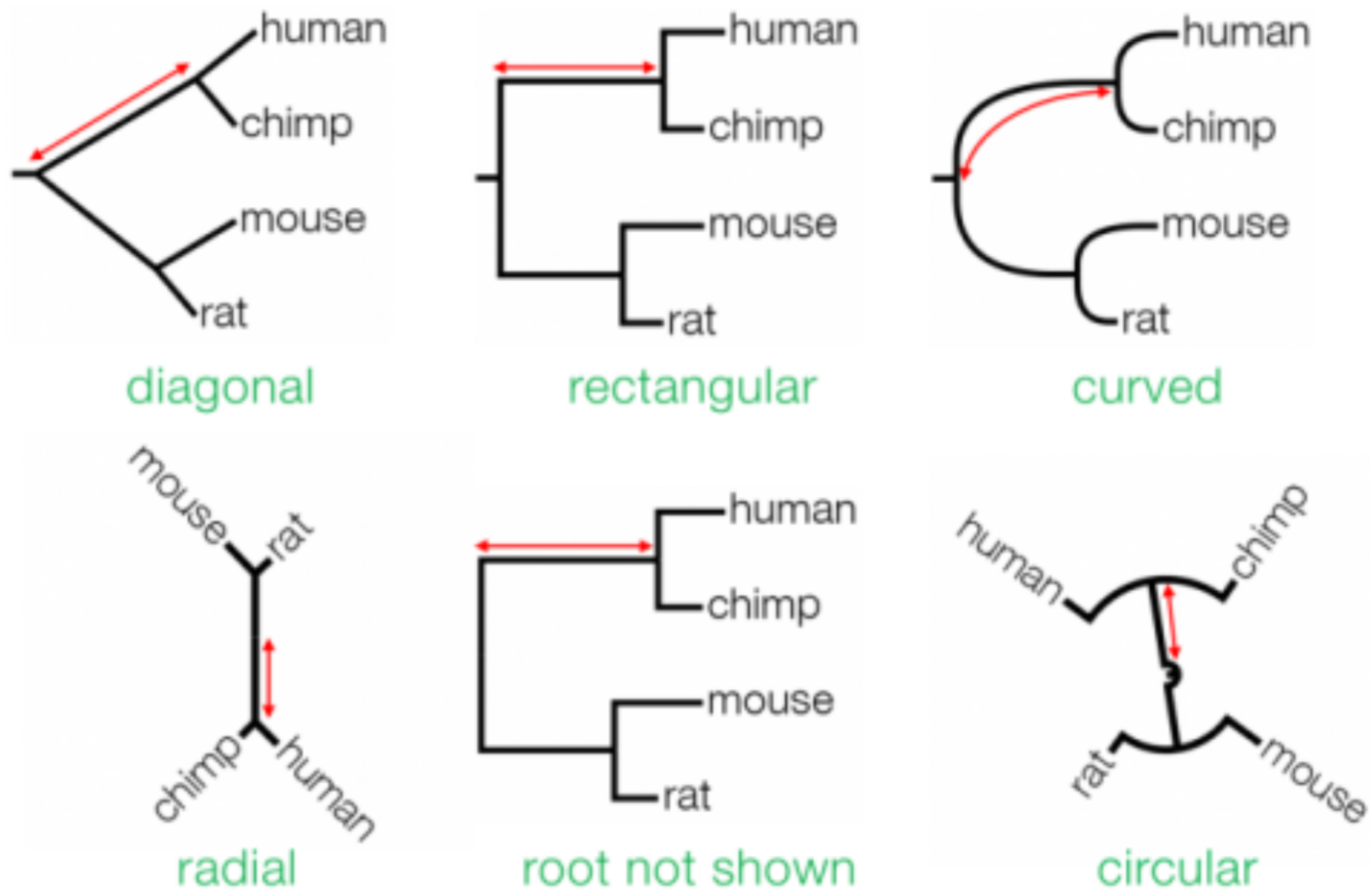
Características de las filogenias

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



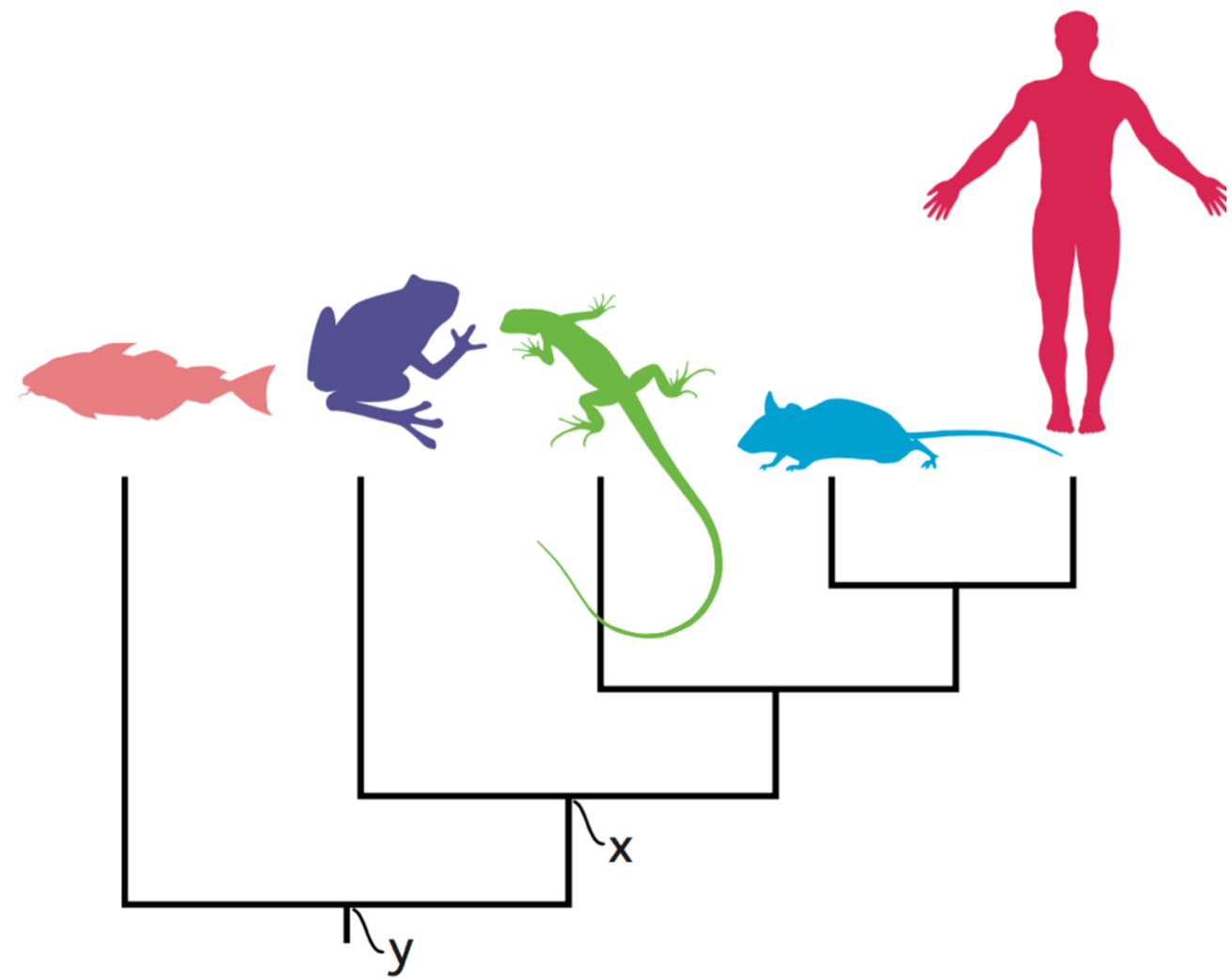


Características de las filogenias



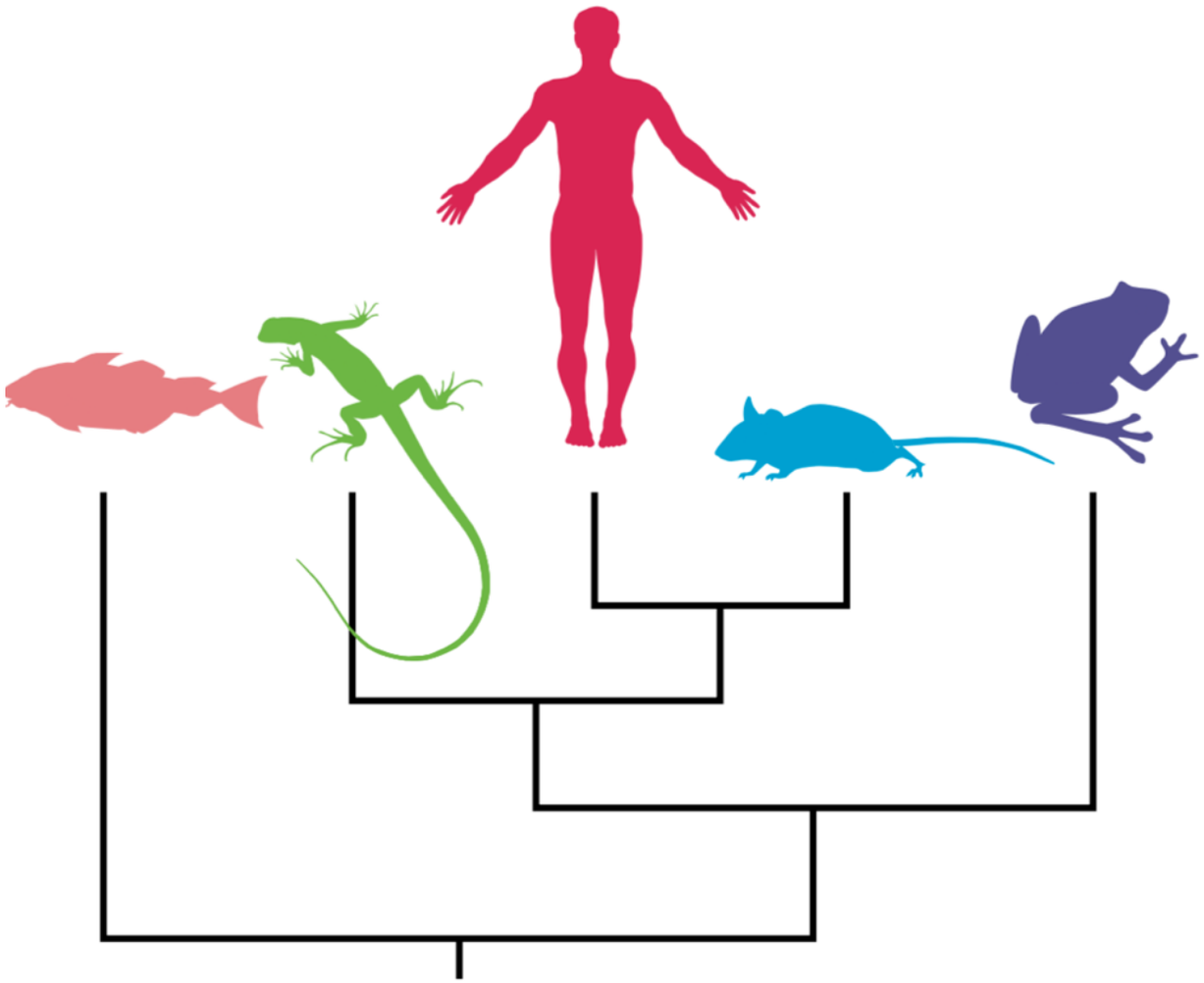
¿Cómo “leer” una filogenia?

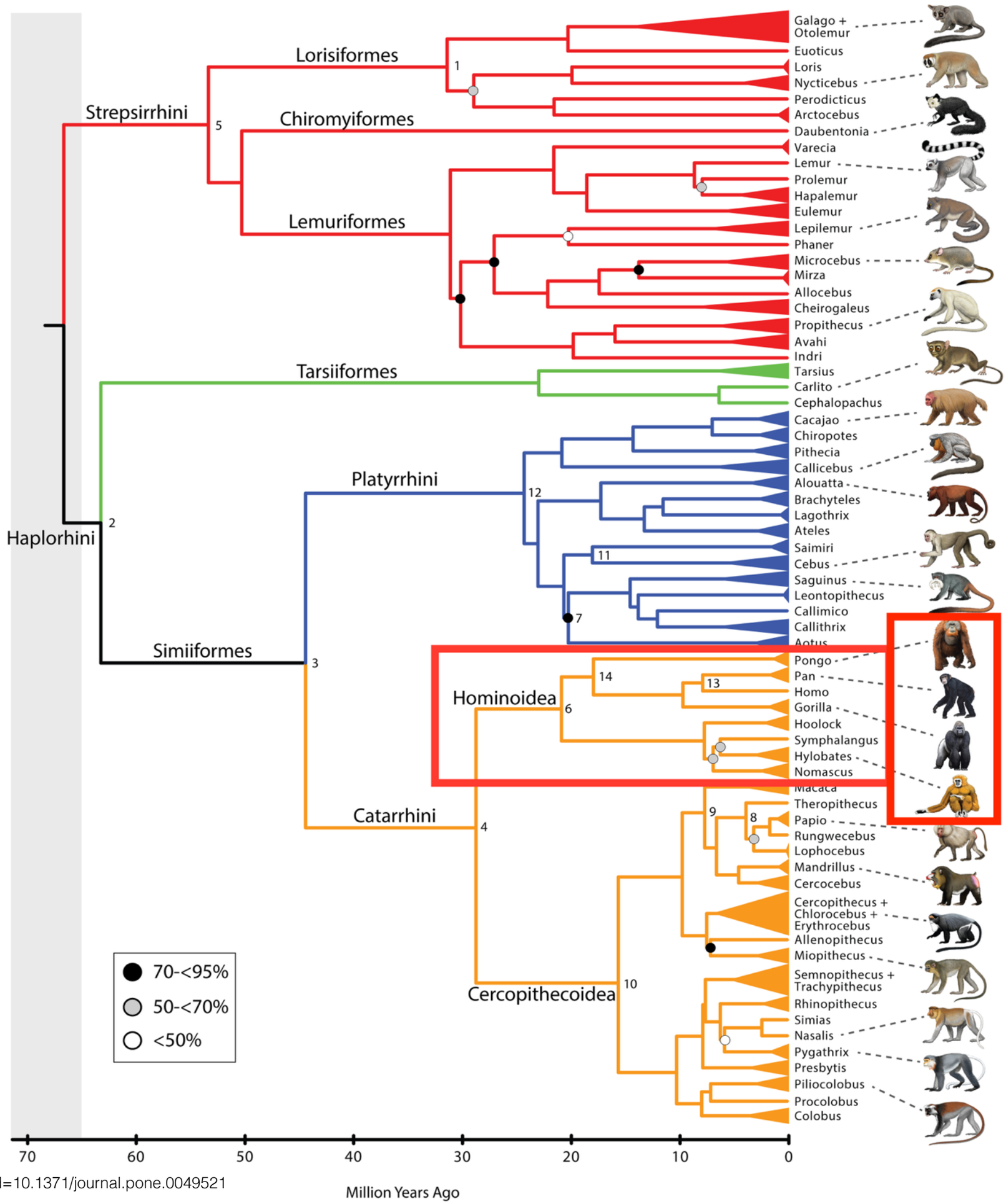
- Basándose en el árbol ¿la rana es más cercana al pez o al humano?



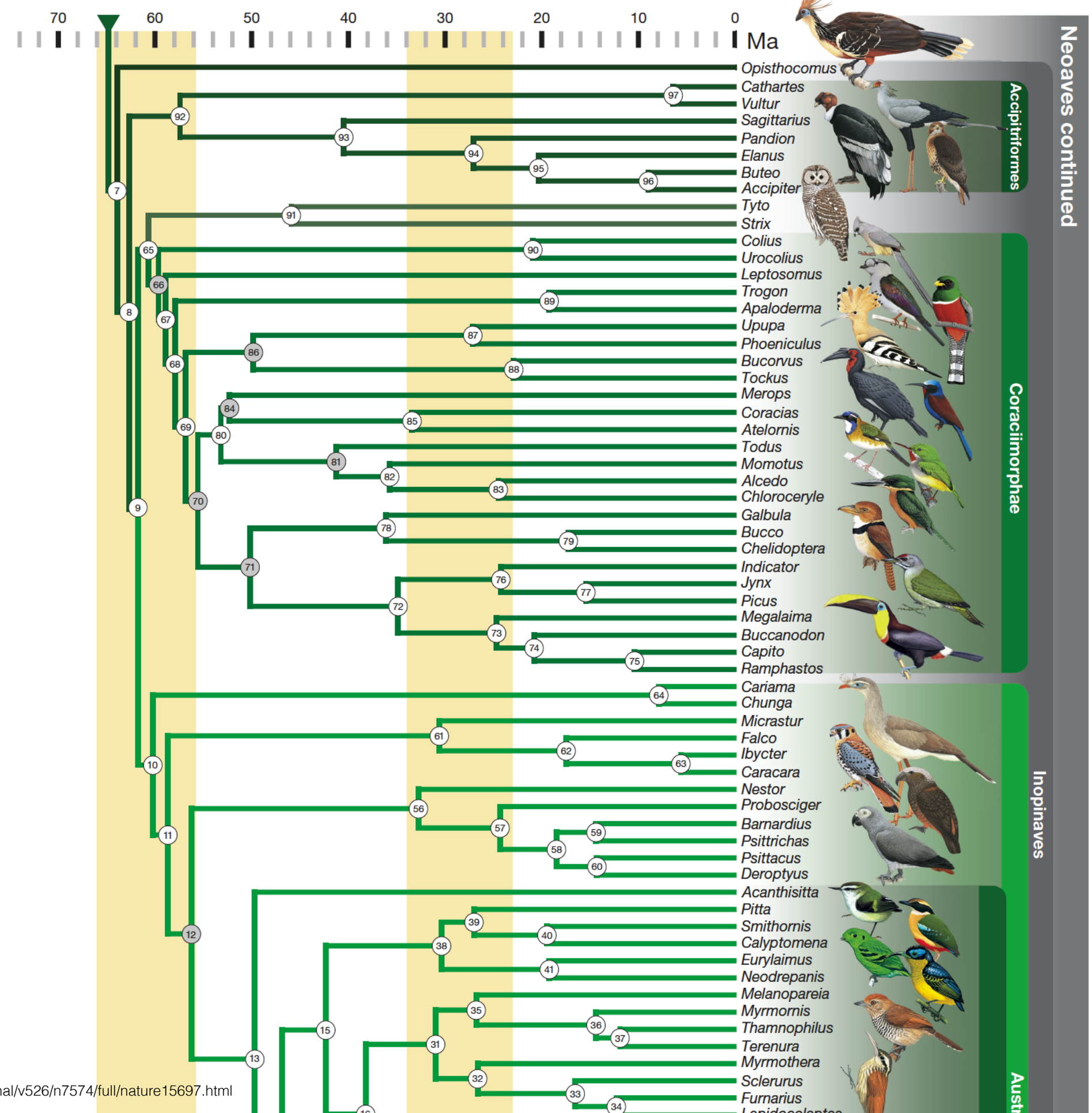
¿Cómo “leer” una filogenia?

- ¿Y ahora?

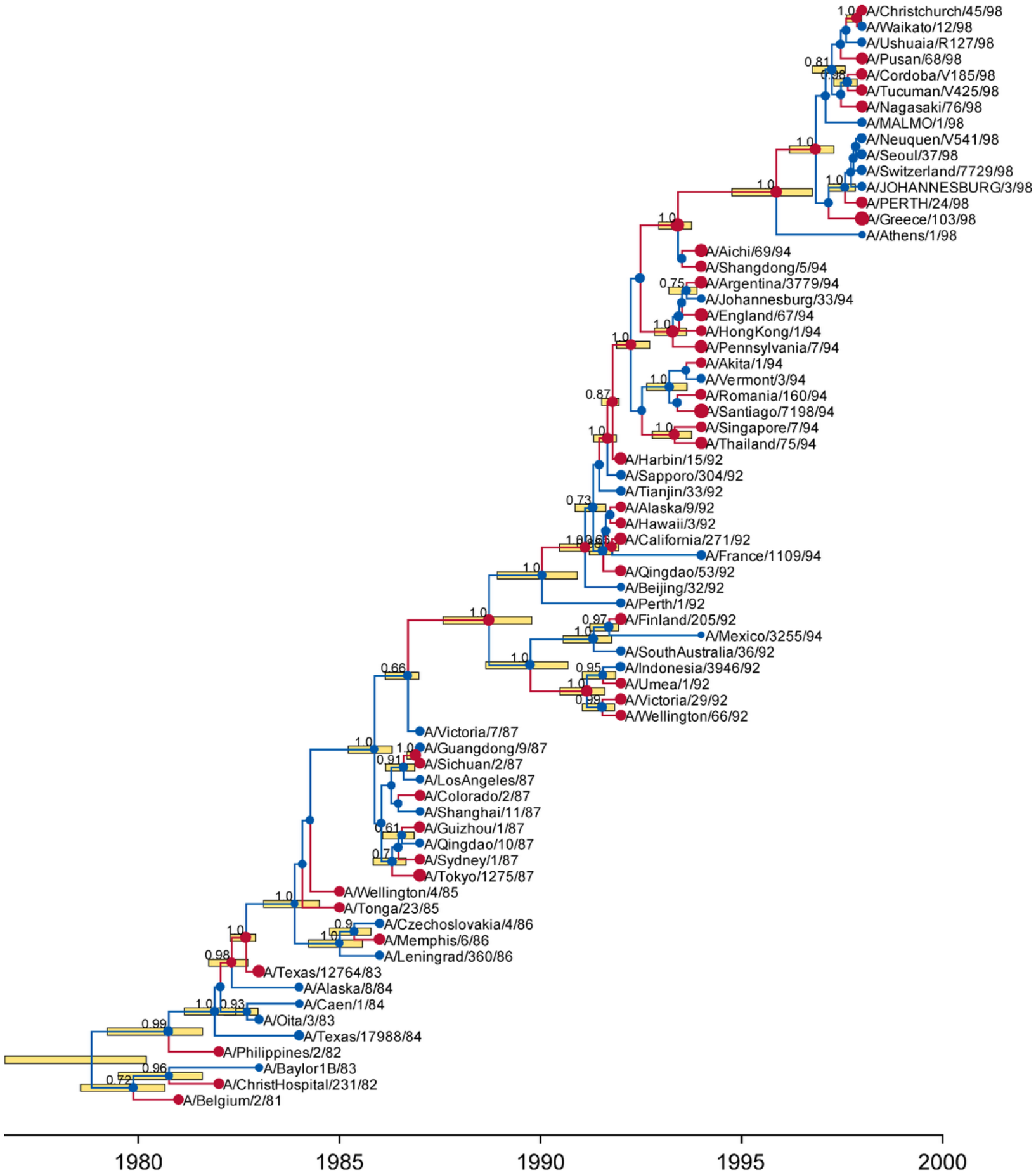




http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0049521



http://www.nature.com/nature/journal/v526/n7574/full/nature15697.html



http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040088