

# Bacteriófagos: analizando su diversidad

Leonardo Collado Torres  
Sur Herrera Paredes

Licenciatura en Ciencias  
Genómicas, UNAM

Cuernavaca, México

4 de Diciembre de 2008

# Índice

- Introducción
  - Diversidad, problema, interés.
- Uso de Codones
- Análisis de clusters
- Análisis de pares de fagos cercanos
  - Sensibilidad, sensitividad, red visual.
- Lo que sigue...

# Diversidad

- Son los organismos más **abundantes** en la Tierra.
- Metagenomas virales: más huérfanos (ORFans) que en cualquier otro conjunto de secuencias.
- Genomas **mosaicos**.
- Su clasificación taxonómica es por morfología. Tsk tsk.

# Problema

- No hay un elemento **común** entre fagos. ¡No hay equivalente al 16s ARNr!
- ¿Y los genes de las **cápsides**?
  - Cápsides morfológicamente parecidas no son homólogas a nivel de amino ácidos.
- ¿Tienen los fagos un **origen monofilético**?
  - Para hacer una clasificación filogenética la respuesta tendría que ser que sí.
  - Creemos que no es cierto.

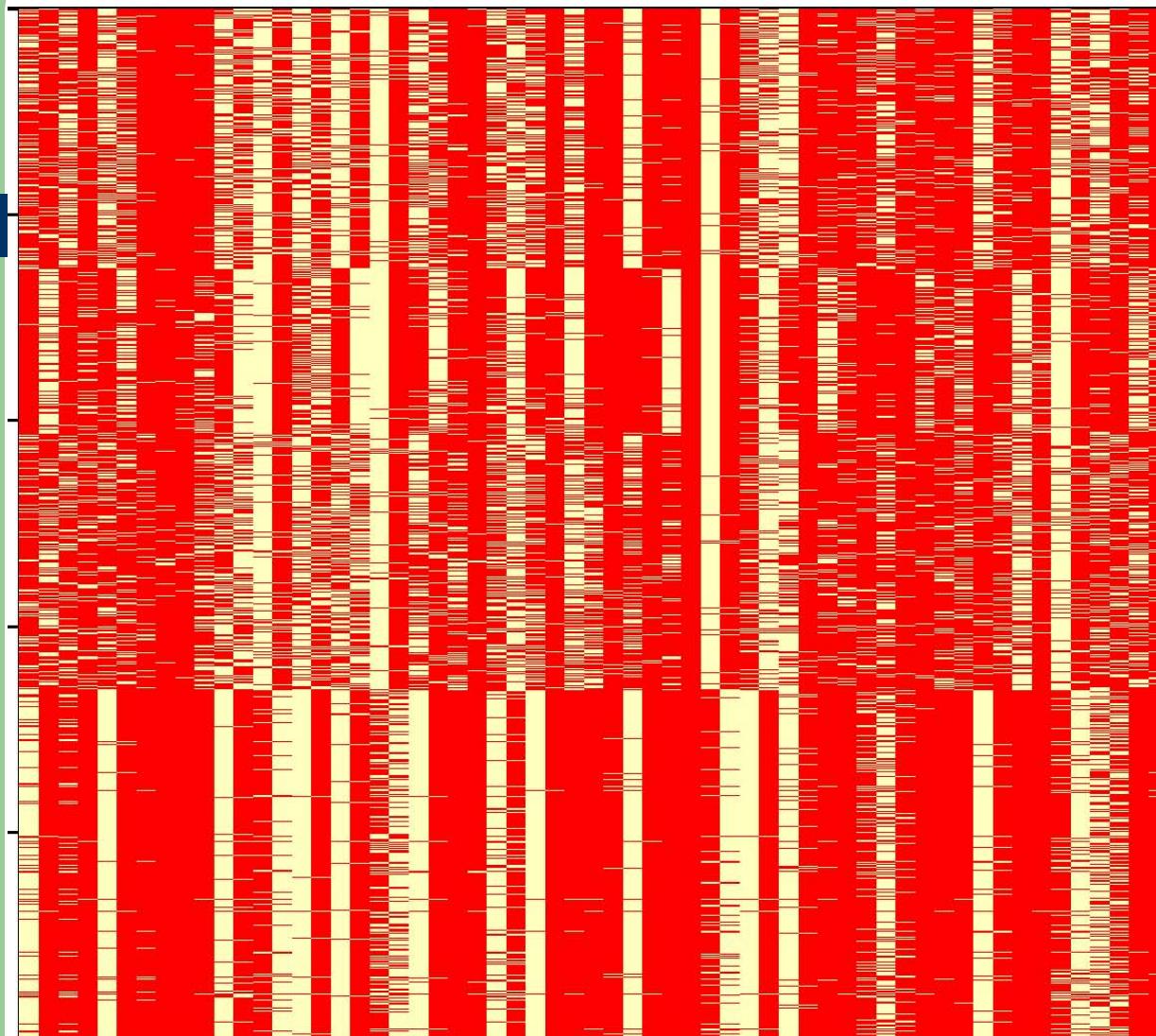
# Interés

- Hacer una clasificación **ecológica** de los fagos.
- Representar sus interacciones con los hospederos.
- Previamente: **contenido proteico**
- Este semestre nos enfocamos en frecuencias y uso de **codones**.

# ¿Por qué uso de codones?

- Un fago usa la **maquinaria** del hospedero para traducirse.
- Complementación de **tRNAs** con el hospedero.
- Hay cierta **correlación** entre:
  - nivel de expresión y uso de codones (bact).
  - nivel de expresión y genes esenciales (bact).
- En bacterias, hay una relación entre la **filogenia** de estas y el uso de codones.

## Secuencias x Codones – Binario



Bacillus  
Subtilis

Pseudomonas  
Aeruginosa

Staphylococcus  
\_aureus\_N315  
(2 replicones)

Escherichia\_col  
i\_K12\_substr\_  
MG1655

# ¿Cómo medir el uso de codones?

- Métodos anteriores:
  - Una ji cuadrada dividida por la longitud del organismo.
  - CAI: depende de un grupo de secuencias nuclear.
  - SCCI (Carbone).
  - SCUO.
- Desventajas
  - Te dan un valor de 0 a 1 para todo el organismo.
  - No puedes agrupar fagos con ellos.
- Creamos y exploramos unos nuevos
  - Eliminamos errores en archivos Genome\_Reviews.

# Nubes de frecuencias de codones

- Por cada fago, **relativizamos** el promedio de las frecuencias de los 64 codones.

	C1	C2	C3	...	Suma Linea
Gene1	3	4	5		12
Gene2	7	12	23		42
Gene3	20	4	3		27
...					
Total	30	20	31		

Relativo	C1	C2	C3	Suma Linea
Gene1	0.25	0.333333	0.416667	1
Gene2	0.166666667	0.285714	0.547619	1
Gene3	0.740740741	0.148148	0.111111	1

0.385802469 0.255732 0.358466 Promedio  
0.096221994 0.009248 0.050175 Desv. Est.

0.051881635 Máxima Distancia Base

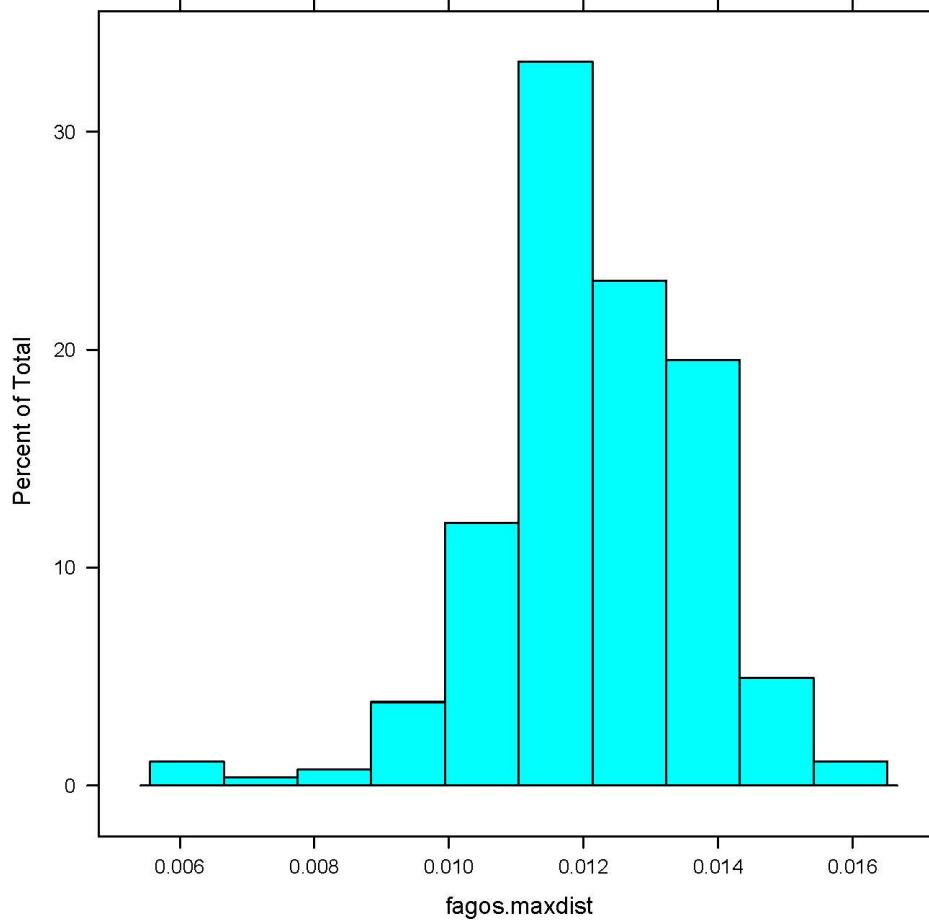
...

- Definimos “**nubes**” por fago
  - El centro es la media de cada codón: punto en 64 D.
  - El grosor de la nube es el promedio de las 64 desviaciones estándar.
- Analizamos los **pares** de fagos que cayeron adentro de sus nubes.
- Hicimos un **clustering** y analizamos los clusters resultantes.

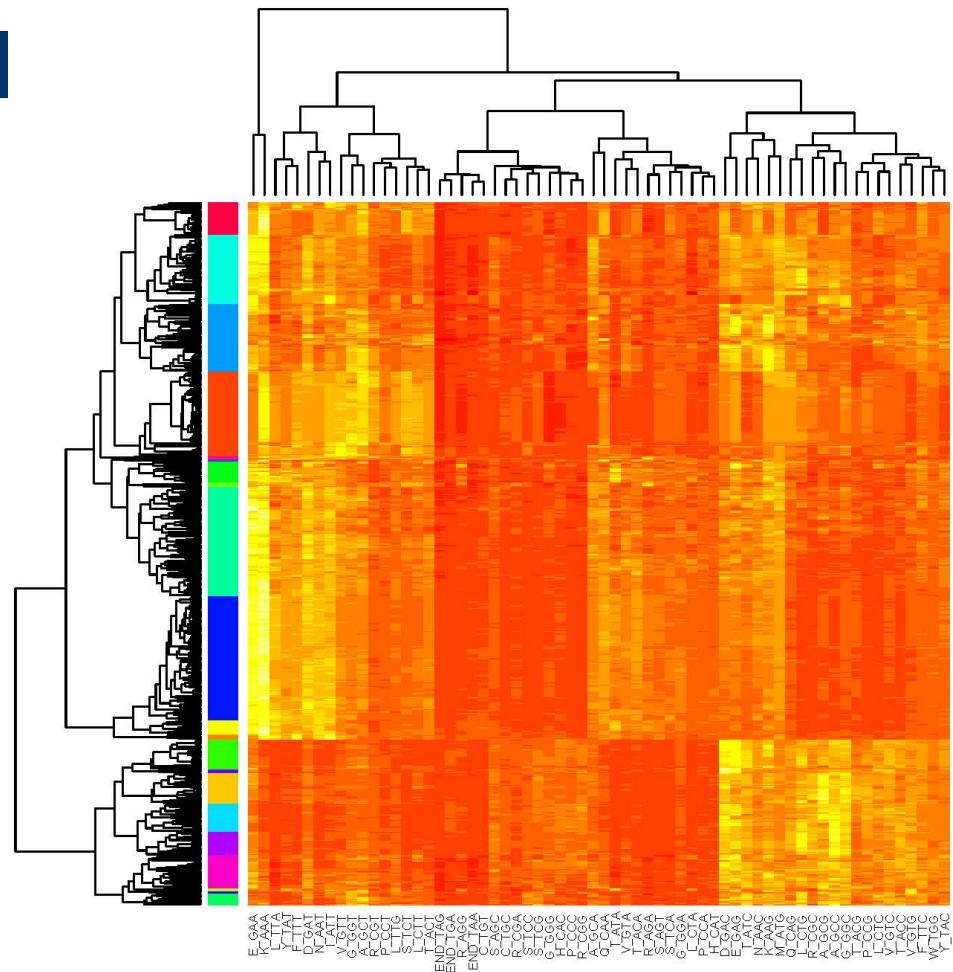
## Uso de codones: binario

- Eliminamos los codones de paro, W y M.
- Por fago, relativizamos los valores por amino ácido.
- Al máximo(s) por aa le dimos el valor de 1, al resto 0.
- No pudimos definir un grosor de la nube por fago, pero usamos uno general.

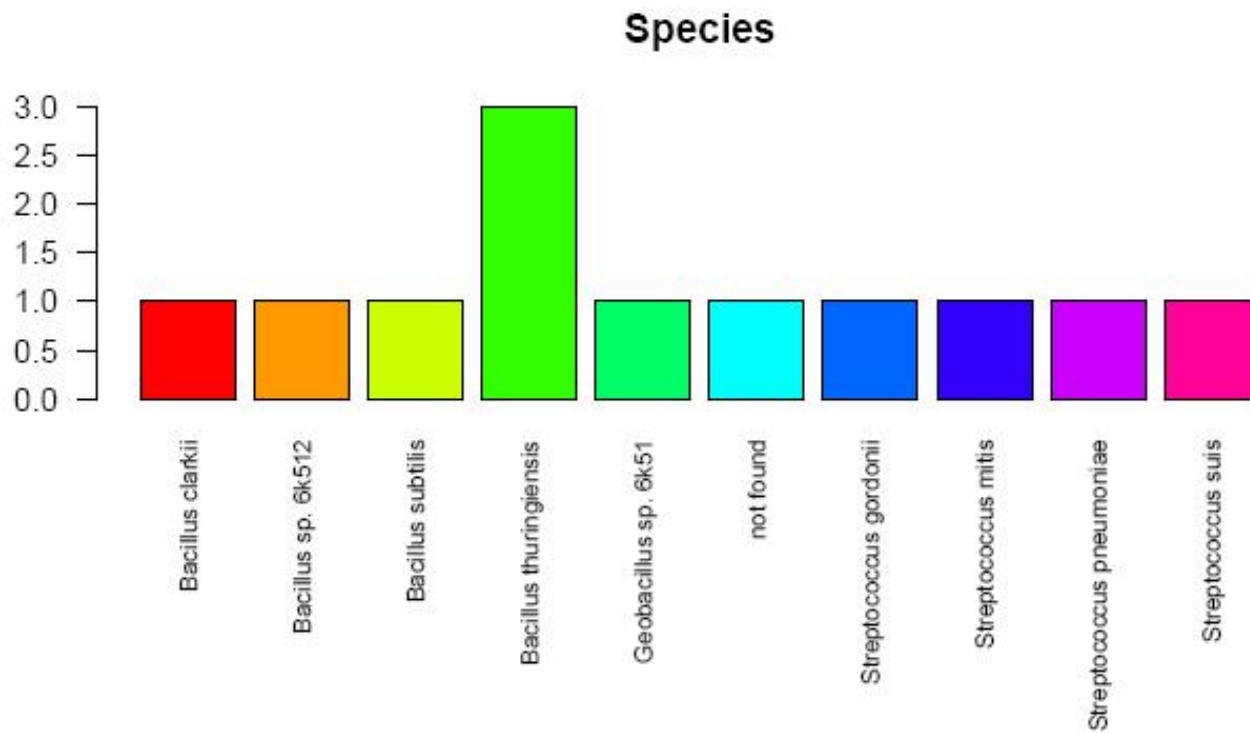
# Grosor de las nubes



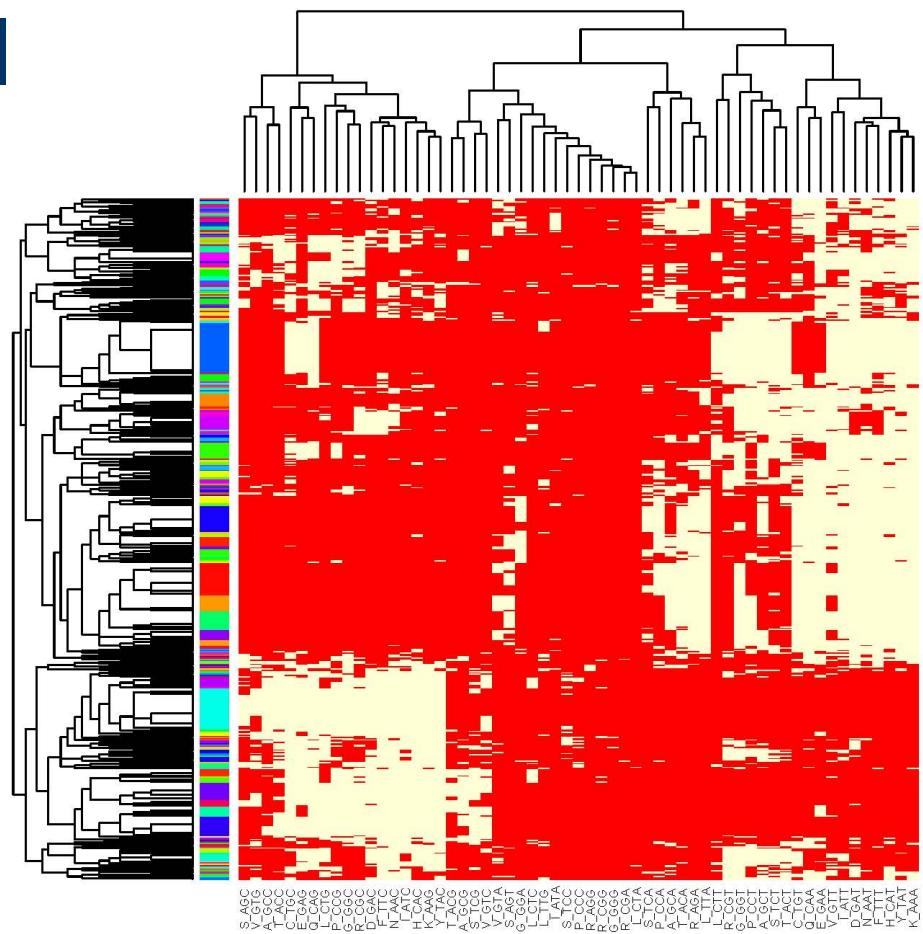
# Frecuencia de codones



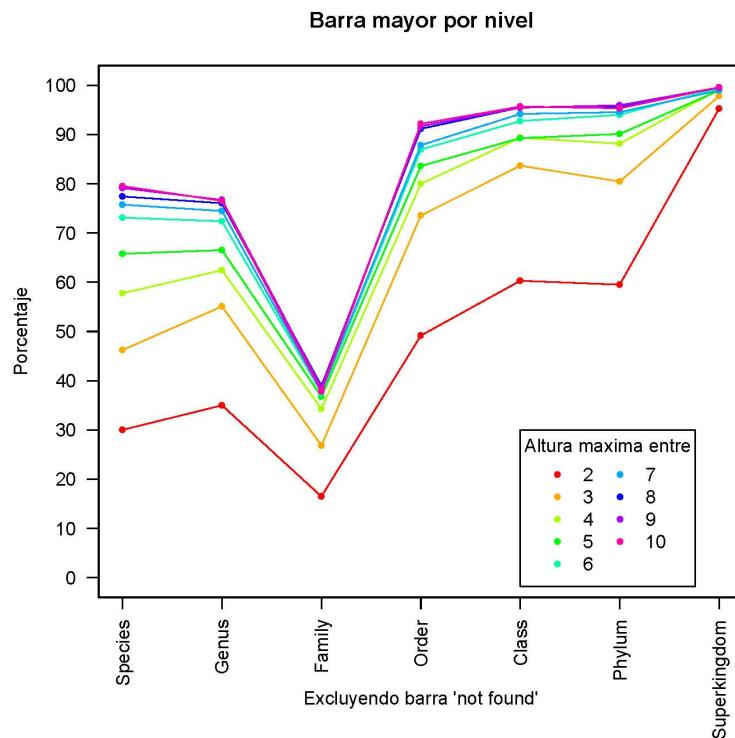
# Análisis de clusters: frec. codones



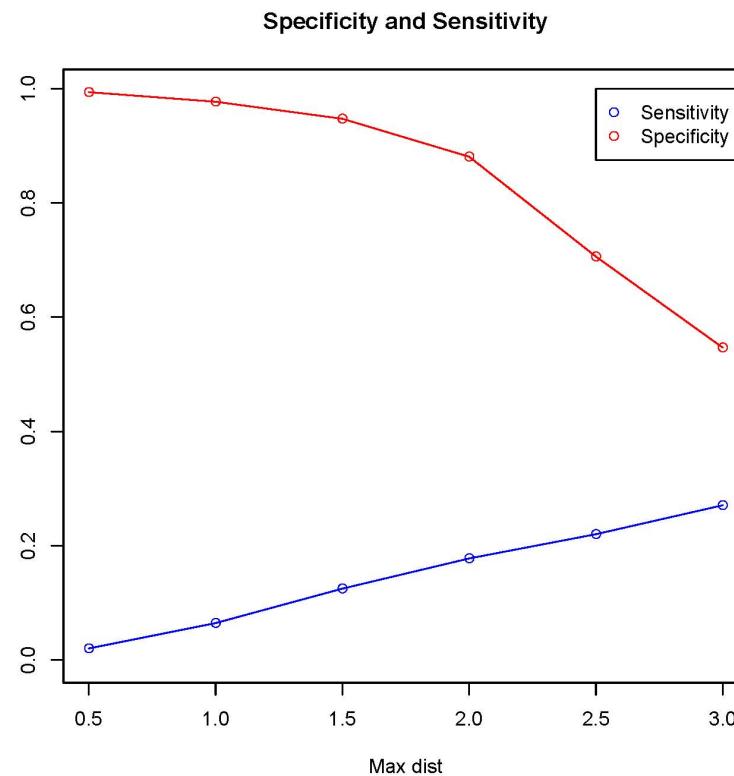
# Método binario



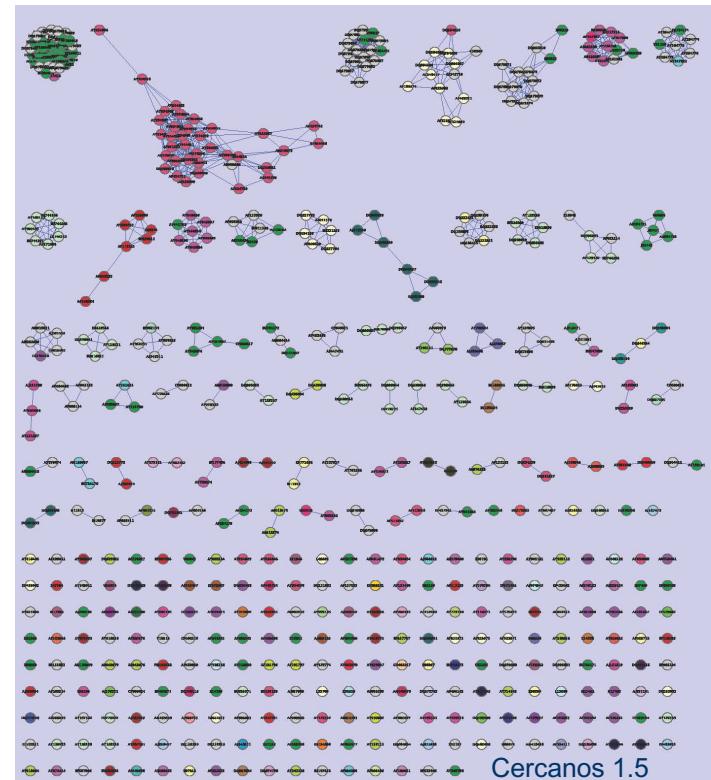
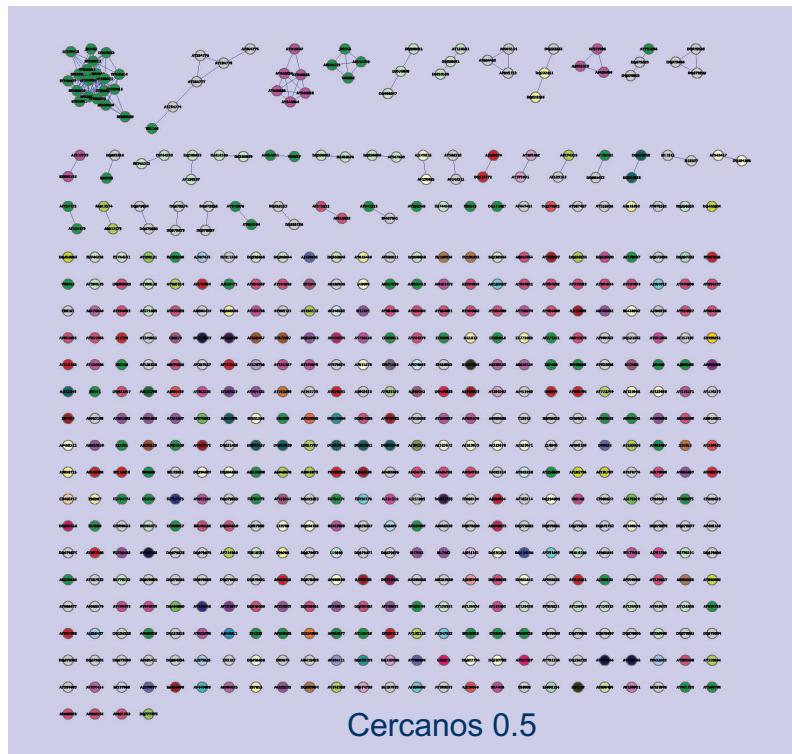
# Análisis de clusters: método binario



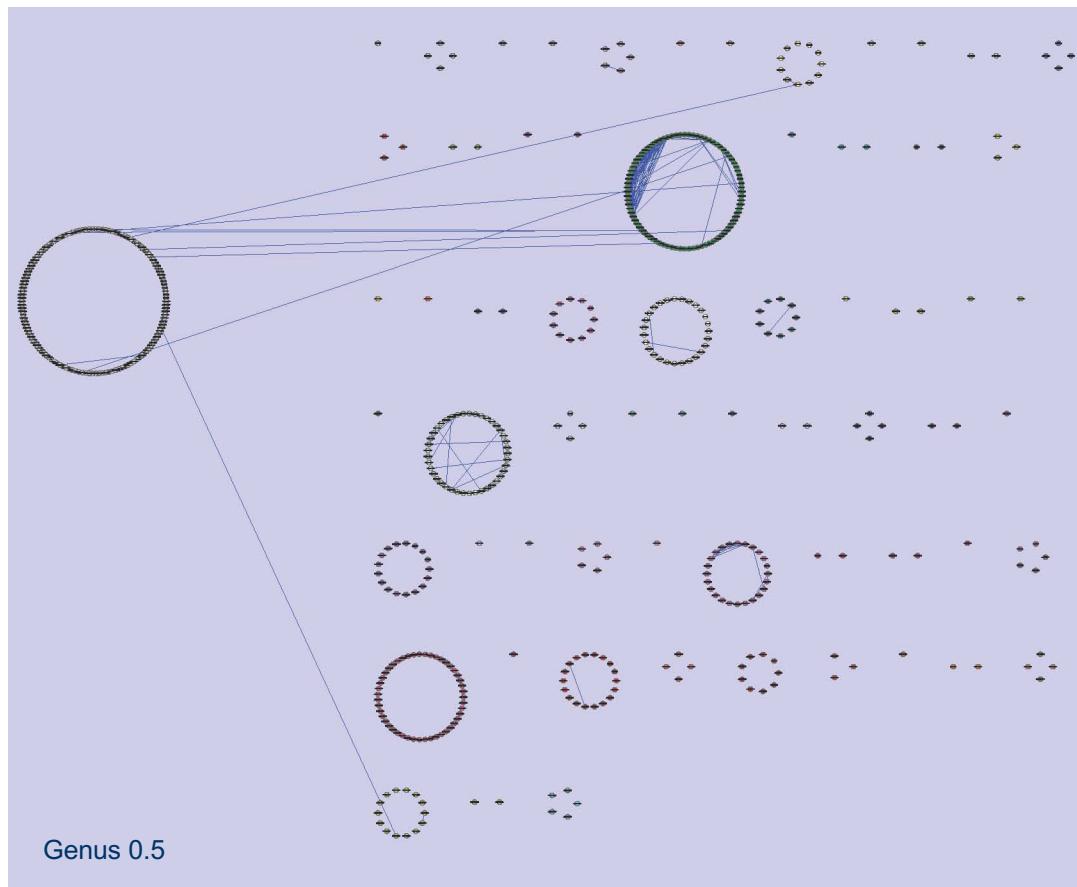
# Sensibilidad y sensitividad: frec. codones



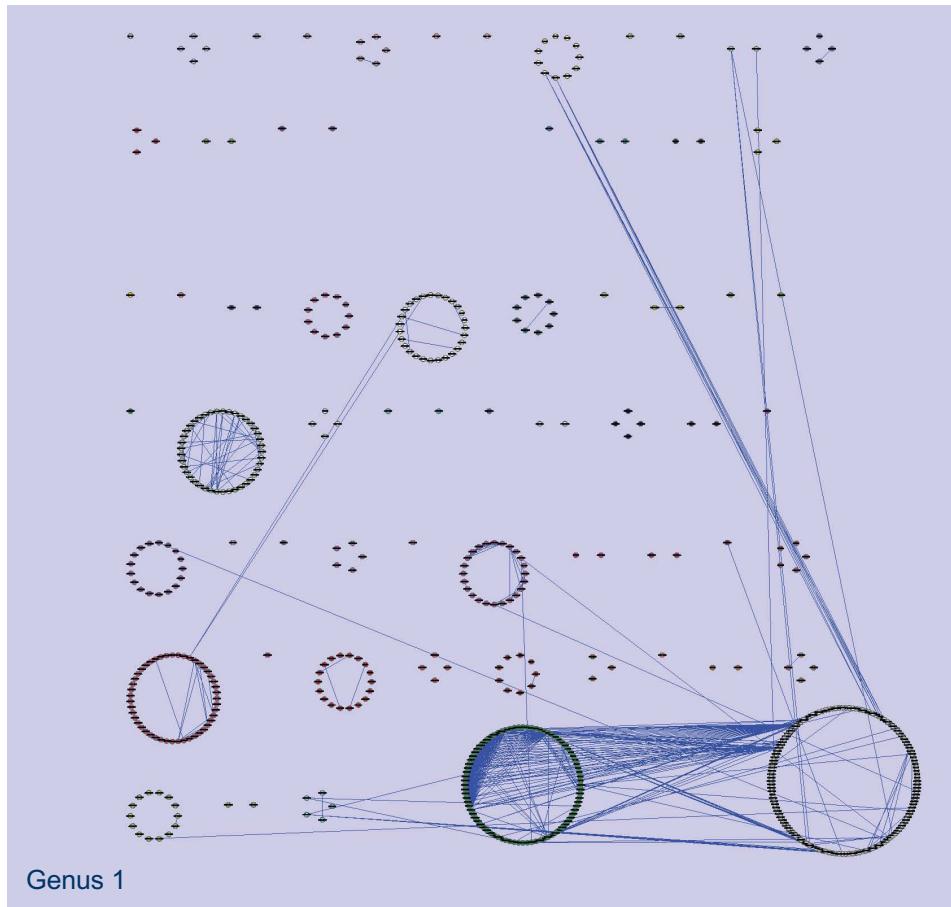
# Cytoscape: pares cercanos (freq)



# Agrupados por género (0.5\*MD)



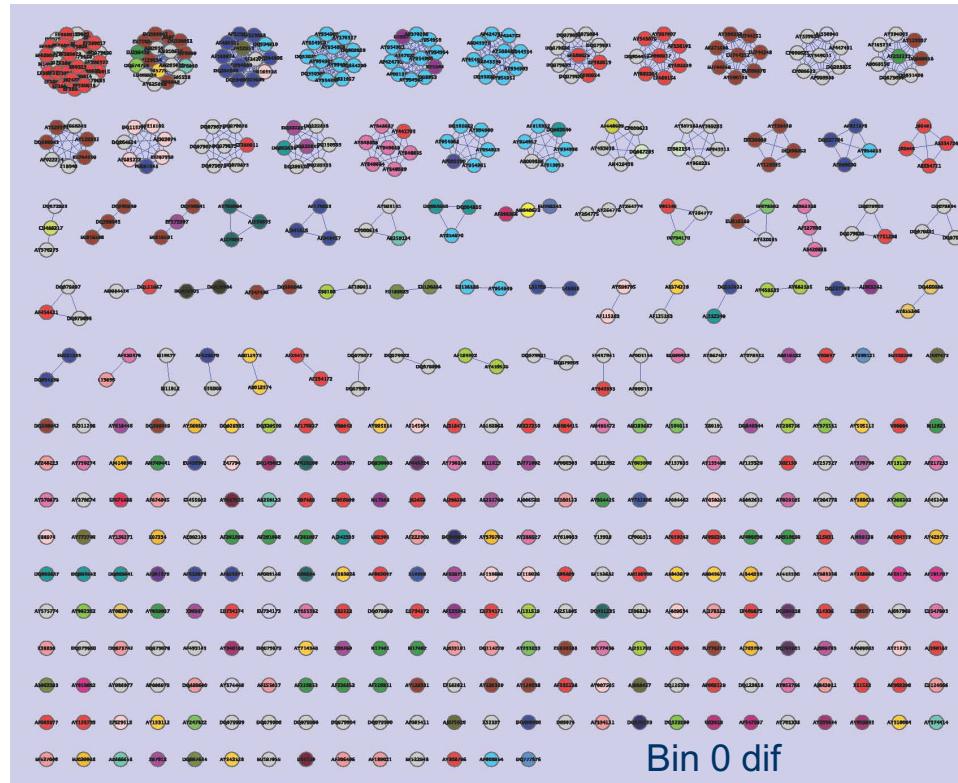
# Agrupados por género (1\*MD)



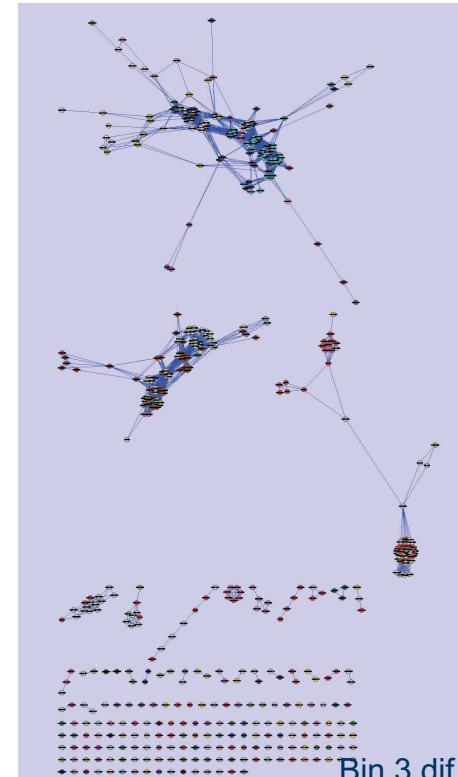
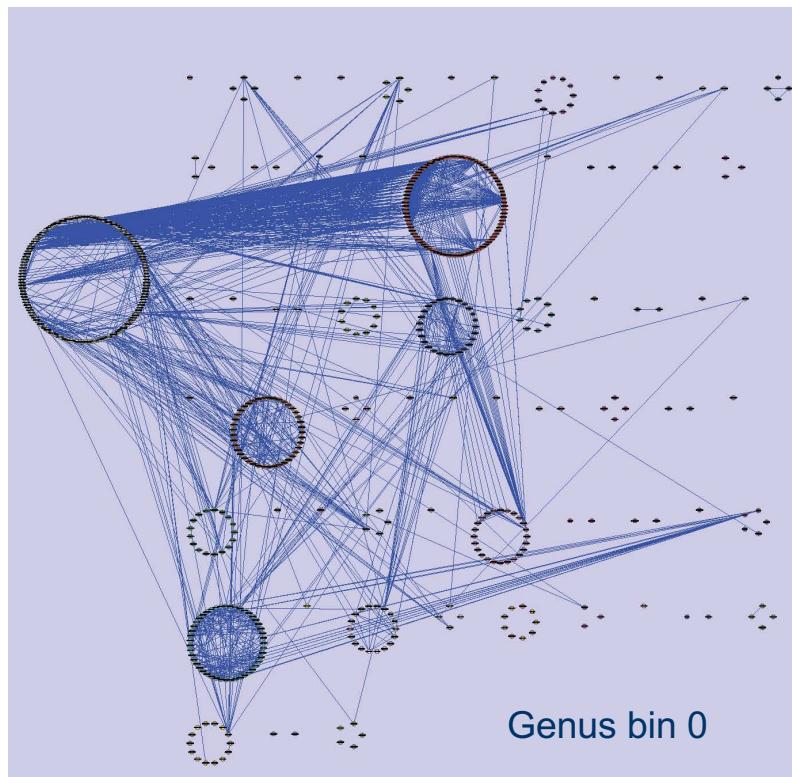
# Sensib. y sensit. con método binario

<i>DIST</i>	<i>sameHost</i>	<i>diffHost</i>	<i>NA</i>
0	700	172	438
1	700	172	438
2	1640	877	1434
3	1640	878	1434
4	2385	2743	2934
5	2397	2758	2959
6	3061	5845	4716
7	3063	5877	4752
8	3633	9779	6909
9	3637	9820	6941
10	3994	13949	9161
11	3995	13975	9169
12	4385	17978	12007
13	4386	18011	12016
14	5014	22440	15690
15	5014	22465	15699
16	5740	29018	20868
17	5742	29064	20887
18	6406	35773	26950
19	6410	35825	26981
20	6900	41688	32162

# Pares cercanos: 0-1 dif binario



# Género (0 dif) y pérdida de señal



# Lo que sigue...

- Usar los métodos a nivel de secuencias.
  - Hay que definir secuencias “no informativas” y filtrarlas.
- Ver si eliminamos codones “no informativos”.
- Checar la taxonomía\*\* de los fagos en los grupos.
- Hacer “nubes” irregulares en vez de las nubes esféricas.

# Otras cosas

- Explorar frecuencias de bi, tri, ... amino ácidos en fagos y demás organismos.
  - ¿Hay una firma para los fagos?
- Probar predicciones experimentalmente
  - Para aprender y probar rangos de infección de fagos (predicciones).