



Análisis de coabundancia Virus-Bacteria mediante un enfoque metagenómico

O'Bryan Cárdenas Andaur
obryan.cardenas.12@sansano.usm.cl

Centro de Biotecnología Dr. Daniel Alkalay Lowitt
UTFSM

14 de septiembre de 2017

CONTENIDO

- 1 Introducción
- 2 Objetivo
- 3 Algunas Definiciones
- 4 Índice de Jaccard
- 5 Validación
- 6 Resultados
- 7 ¿Y ahora qué?

INTRODUCCIÓN

La virología en ambientes acuáticos tiene 70 años. Este campo ha resurgido gracias a estudios metagenómicos, los cuales han resaltado que estos ambientes son excepcionalmente diversos: más de 5000 genotipos virales en 100 litros de agua marina. La mayoría de especies virales se mantienen desconocidas.

Importancia ecológica: Los virus pueden matar al 50 % de la población de bacterias generadas cada día.

OBJETIVO

Implementar el índice de Jaccard, basado en estudios metagenómicos, como una herramienta que permita identificar y/o predecir la relación fago-hospedero.

HIPÓTESIS

Coexistencia de genes de virus y bacterias en distintos medioambientes permite identificar la relación fago-hospedero y discriminar estrategias ecológicas de los virus (especificidad).

GLOSARIO

Hábitat: Colección de secuencias de DNA (metagenoma) extraídas desde un medioambiente.

Coexistencia: Es la presencia de un par de especies, virus:bacteria, en un hábitat (ambiente) común.

Coabundancia: Es la relación de la cantidad de individuos entre 2 especies, virus:bacteria, en un ambiente común.

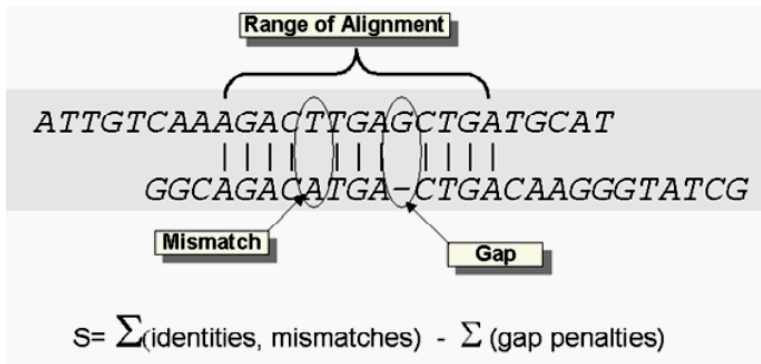
GLOSARIO

Presencia de especie viral: Es la identificación de al menos el 99 % de las proteínas que conforman su genoma en un hábitat.

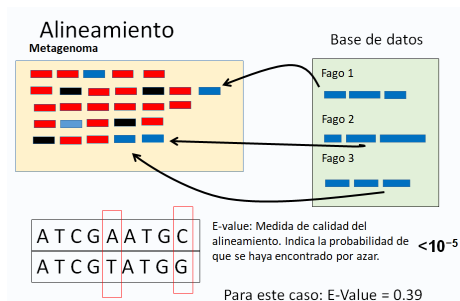
Presencia de especie bacteriana: Es la identificación de al menos una lectura (trozo de DNA de la bacteria) en un hábitat.

GLOSARIO

Homología existe cuando 2 secuencias comparten un origen evolutivo común, por ende las dos han divergido desde ahí.



GLOSARIO



La homología se puede medir a través de:

Identidad: es el numero de caracteres (bases) iguales que hay entre 2 secuencias.

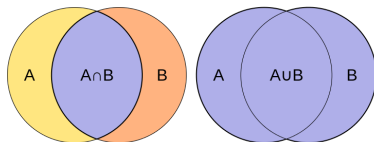
E-value: es la probabilidad de encontrar un hit de similaridad debido al azar. (A mayor E-value, menor similaridad).

DEFINICIÓN MATEMÁTICA

Sean A, B dos conjuntos, se define el índice de Jaccard entre A y B como:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

donde $|A|$ denota el número de elementos de A .



Observación: $0 \leq J(A, B) \leq 1$

DEFINICIÓN UTILIZADA

Considere un conjunto de Metagenomas, una especie de bacteria B y un virus V . Para medir la coexistencia entre la bacteria y el virus definimos:

$$J(V, B) = \frac{\# \text{metagenomas donde hay presencia de } V \text{ y } B}{\# \text{metagenomas donde hay presencia de } V \text{ o } B}.$$

PROBLEMA

Representatividad: un virus presente en M_1 y una bacteria presente en M_1 tienen el mismo índice de Jaccard que un virus presente en $\{M_1, \dots, M_{100}\}$ y una bacteria presente en los mismos metagenomas.

Solución: Índice de Jaccard Relativo

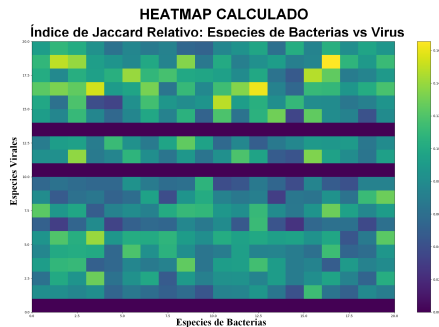
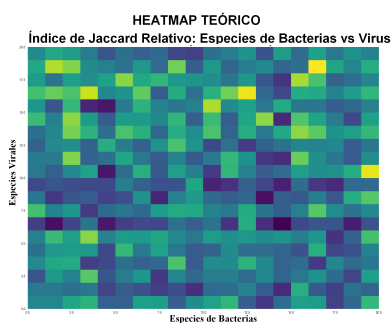
$$J_r(V, B) = J(V, B) \times \frac{\# \text{metagenomas donde hay presencia de V y B}}{\# \text{metagenomas analizados}}.$$

VALIDACIÓN DE HERRAMIENTA IMPLEMENTADA

Construcción de metagenomas sintéticos vía **randomreads**, considerando 20 especies virales y 20 bacterianas.

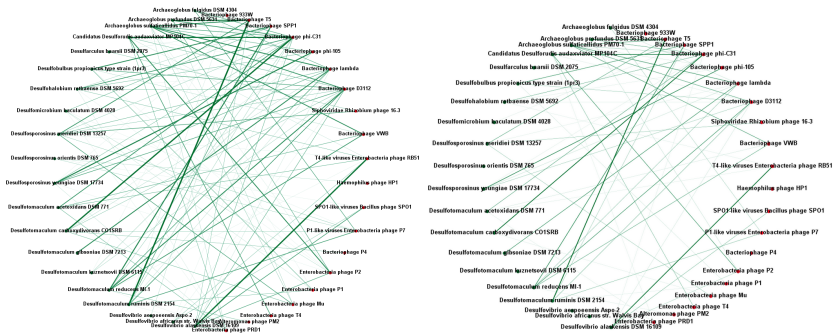
Cada metegenoma construido consta de 10 virus y 10 bacterias seleccionados de forma aleatoria.

HEATMAP TEÓRICO VS HEATMAP OBTENIDO



La herramienta, al considerar al menos un 95 % del genoma completo de fagos, utilizando metagenomas sintéticos y el top2 de homologías, comete un error de 12 % al agregar falsos positivos y una pérdida del 5 % al no identificar 3 especies.

GRAFO TEÓRICO VS OBTENIDO



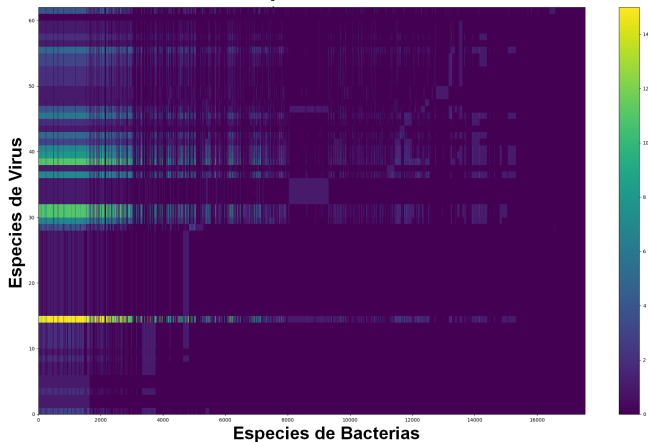
La distribución de las interacciones fagos:bacterias (links en verde) dada por la herramienta implementada (derecha) se mantiene similar a la teórica (izquierda).

DATOS ANALIZADOS

Los datos analizados corresponden a 140 metagenomas de ambientes marinos, todos ellos extraídos de MG-RAST.

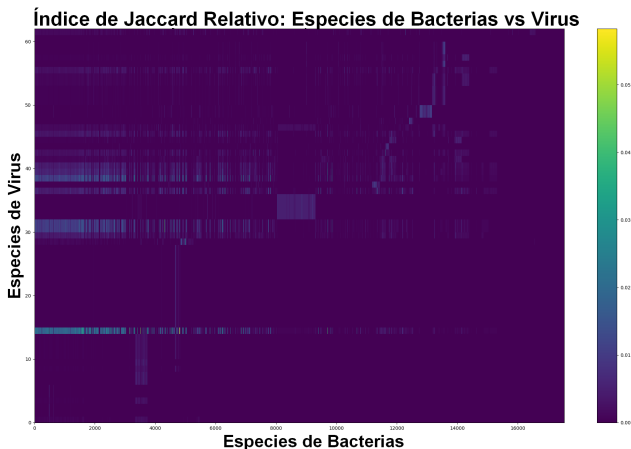
HEATMAP DE COEXISTENCIA DE DATOS ANALIZADOS

Coexistencia: Especies de Bacterias vs Virus



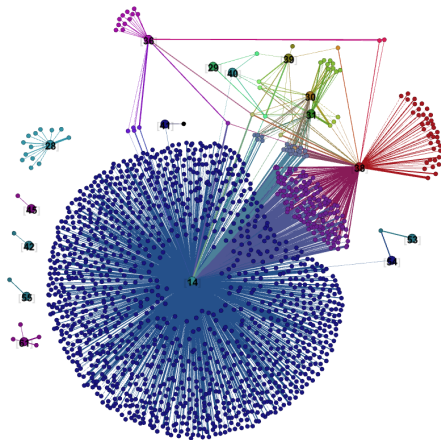
En el heatmap la coexistencia se entiende como la proporción de metagenomas en los que está presente el par virus-bacteria frente al total de metagenomas (140).

HEATMAP DE J_r DE DATOS ANALIZADOS



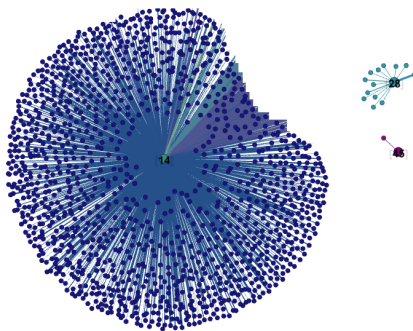
En el heatmap se grafica el índice de Jaccard Relativo tal y como se definió anteriormente.

RED DE COEXISTENCIA VÍA JACCARD RELATIVO



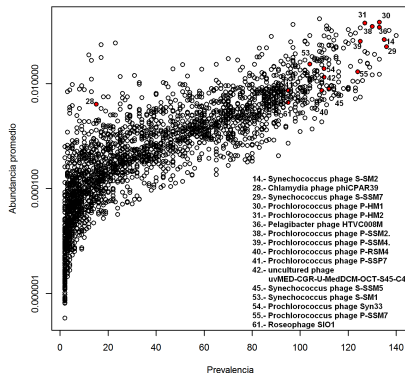
Cada nodo representa una especie de fago (nodos grandes) o bacteria (nodos pequeños). Cada conexión representa una relación de coexistencia.

ESTRATEGIAS VIRALES



La promiscuidad, puede verse reflejada como el número de conexiones que presenta un fago. En este caso se dice que el fago 14 (*Synechococcus* phage S-SM2) es mucho más promiscuo que por ejemplos los fagos 28 (*Chlamydia* phage phiCPAR39) y 45 (*Synechococcus* phage S-SSM5).

RELACIÓN CON ABUNDANCIA Y PREVALENCIA

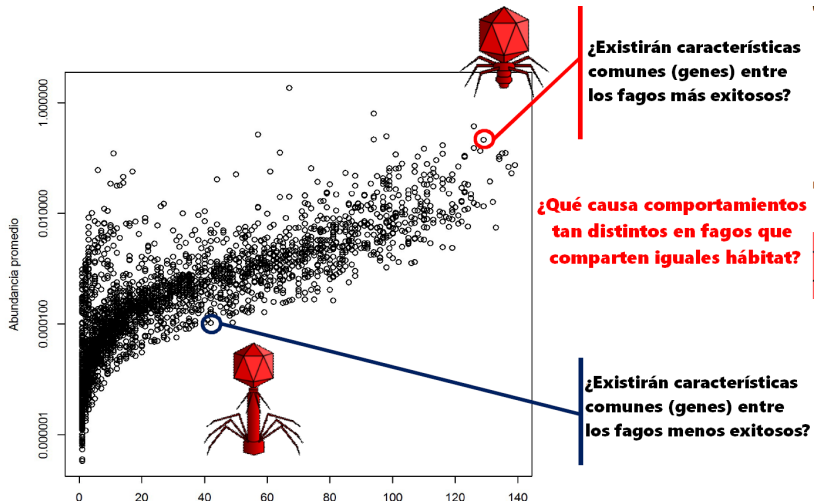


Aquellos fagos identificados con un mayor índice de Jaccard Relativo son aquellos con mayor prevalencia y abundancia en las muestras metagenómicas (140) de medioambientes marinos.

INTERPRETACIÓN

Los resultados sugieren que la promiscuidad, entendida como el número de especies bacterianas que puede habitar un virus es una herramienta fundamental para su éxito ecológico.

¿QUÉ HACE EXITOSOS A LOS FAGOS EXITOSOS?



EN EL GENOMA DE LOS FAGOS

PROBLEMA: Existe poca caracterización de los genes de los virus, hay muchos "genes misteriosos".

EN EL GENOMA DE LOS FAGOS

PROBLEMA: Existe poca caracterización de los genes de los virus, hay muchos "genes misteriosos".

SOLUCIÓN: Estudiar algunas proteínas mejor conservadas y de las que se tiene más información: Las Endolisinas.

ENDOLISINAS

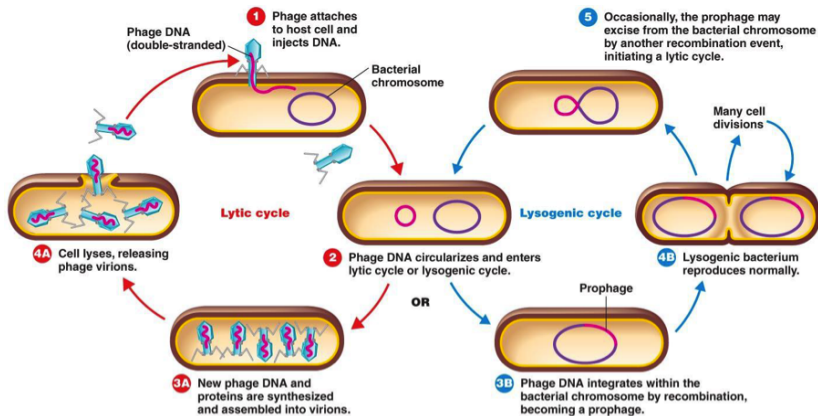
Las endolisinas son enzimas hidrolíticas producidas por bacteriófagos al final de la fase lítica para romper enlaces del péptidoglicano.

ENDOLISINAS

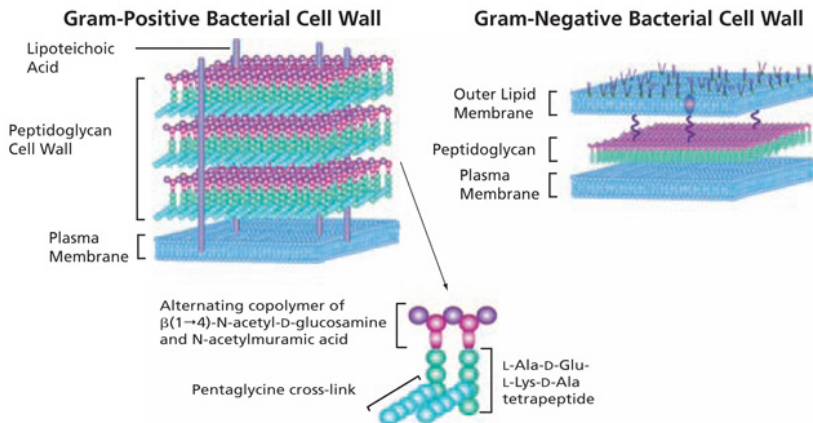
Las endolisinas son enzimas hidrolíticas producidas por bacteriófagos al final de la fase lítica para romper enlaces del péptidoglicano.

¿¿TRADUCCIÓN??

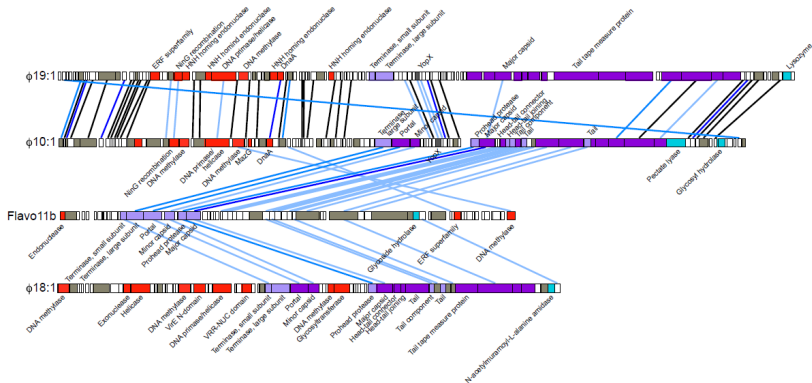
CICLO LÍTICO Y LISOGÉNICO



BACTERIAS GRAM + Y GRAM -



28 / 31



PERO POR AHORA:

Burkholderia phage Bcep22	MGNRARTLIGALAVSAAAFGTWVAS	E	GFAPKAEIPTKGDVPTIGHGSTRY
Cronobacter phage ES2	MQIPAKLRTALVAAAAGGASFIAGVLIQYQ	E	GVKYKPYLDPVGIPTVCAG
Caulobacter phage Cd1	MQNRAKAGIIGAVAAISAAGVAFIAGW	E	GKENAPYKDMVGWTVCYGSTG
Stx2-converting phage 1717	MSPSLRKAVAAAIGGGAVAIASVLITGPGGNDGLE	E	GVSYIPYKDIVGVWT
Stx2-converting phage 86	MNTKIKYGLSAAVLALIAIGAPAPDILDQFLDEKE	E	GNHTTAYRDGSGIWT
Stx2 converting phage II	MSRKLRYGLSAAVLALIAAGASAPEILDQFLDEKE	E	GNHTTAYRDGAGIWT
Enterobacteria phage Min27	MSRKLRYGLSAAVLALIAAGASAPEILDQFLDEKE	E	GNHTTAYRDGAGIWT
Escherichia phage TL-2011c	MSRKLRYGLSAAVLALIAAGASAPEILDQFLDEKE	E	GNHTTAYRDGAGIWT
Enterobacteria phage RTP	MKQKLVVGAATAAAIYIAAPLIELV	E	GVENKPYMDIAGIPTVCAGVTGPD
Enterobacteria phage P1	MKGKTAAGGGAICAIIVMITIVMGNGNVRTNQAGLELIGNA	E	GCRRDPYM
Enterobacteria phage 933W	MSRKLRYGLSAAVLALIAAGASAPEILDQFLDEKE	E	GNHTTAYRDGAGIWT
Enterobacteria phage cdtI	MSPSLRKAVAAAIGGGAVAIASVLITGPSGDDGLE	E	GVSYIPYKDIVGVWT
Stx1 converting phage	MSRKLRYGLSAAVLALIAAGASAPEILDQFLDEKE	E	GNHTTAYRDGAGIWT
Stx2 converting phage I	MSRKFRLYGLSAVVLALIAAGASAPEILDQFLDEKE	E	GNHTTAYRDGVGIWT

APLICACIÓN INDIRECTA

¿Qué utilidad tienen las endolisinas en particular?

