

2.3 Genomas de bacteriófagos

Diapositiva 2

Existe una enorme variedad de fagos en general. Además, esta diversidad también está representada a nivel del genoma. Debido a su pequeño tamaño y a la facilidad con que se aíslan, el genoma de los bacteriófagos fue el primero en ser secuenciado. Hoy en día, más de 2000 genomas de bacteriófagos han sido secuenciados.

Diapositiva 3

La gran mayoría de bacteriófagos son virus con ADN de doble cadena (dsDNA), aunque también existen de otros tipos. El genoma de los fagos es típicamente lineal cuando está empaquetado en la cápside. Algunos de los fagos que tienen ADN tienen extremos definidos. Esto significa que cuando observamos el ADN en distintas cápsides, siempre tendrán el mismo extremo tanto a la izquierda como a la derecha. Para otros fagos esto no es así, en estos virus, la composición general del genoma es la misma en cada cápside, pero los extremos no son siempre los mismos, el ADN parece haberse linearizado en distintos puntos a partir de círculos idénticos. Esto es lo que se denomina permutación circular. Además, el comienzo y el final, normalmente son secuencias idénticas y esto se denomina redundancia terminal.

Diapositiva 4

El tamaño del genoma de los fagos varía enormemente. Los más pequeños miden un poco más de 3000 nucleótidos, como por ejemplo los virus de ARN de cadena simple (ssRNA) que infectan a *Escherichia coli*, mientras que los mayores, como el fago G que infecta a *Bacillus megaterium* alcanza 500kpb. En general, el ADN de los fagos está densamente empaquetado dentro de la cápside con una densidad más o menos similar, por tanto, el tamaño de la cápside varía en función del tamaño del genoma. Muchos virus, especialmente los fagos dsDNA, contienen capas de ADN empaquetado, mientras que otros, fundamentalmente los filamentosos, son virus que empaquetan su genoma como hélices rodeadas de proteínas. Como no hay mucho espacio dentro de la cápside, el número de genes en un genoma fágico es muy alto, lo que significa que hay muy pocas regiones que no codifiquen ningún gen. Las regiones reguladoras están compactadas y ocasionalmente las regiones codificantes se solapan. El genoma de los bacteriófagos también contiene la mayor cantidad de genes nuevos en el mundo biológico, porque la mayoría de los genes que codifican (incluso más del 80%) no están relacionados con proteínas conocidas, sino con proteínas que tienen todavía una función desconocida.

Diapositiva 5

La clasificación taxonómica de virus de arqueas y de bacterias es muy importante para desarrollar un modelo predictivo para entender la microbiología de los ecosistemas, pero esto está constituyendo un gran reto. Recientemente, un número enorme de genomas víricos y de fragmentos de genomas están siendo identificados, y esto exige alejarse un poco de los métodos tradicionales de clasificación, orientándose hacia la utilización de los genomas para el estudio de la taxonomía. La taxonomía de organismos celulares, basada en el análisis del alineamiento de secuencias de genes homólogos, universalmente conservados, por ejemplo el gen ARN ribosomal (rRNA), es posible ya que todos los organismos comparten estos genes homólogos. Sin embargo, no hay ni un solo gen que esté conservado en todos los bacteriófagos, ni en ningún virus en general. Esto hace imposible determinar cualquier relación evolutiva basándonos en el alineamiento de secuencias de genes únicos entre todos los bacteriófagos.

Rohwer y Edwards propusieron una manera de clasificar taxonómicamente a los bacteriófagos basándose en los proteomas, mediante la construcción de “árboles proteómicos de fagos”. Estos son dendrogramas que revelan una similitud genómica global entre cientos o miles de virus basándose en un algoritmo que usa el genoma para agrupar a los fagos de acuerdo a los que están más próximos respecto al resto de fagos. Este método no requiere visualización directa de los viriones ni conocer su estilo de vida, y puede dividir a los fagos en grupos taxonómicos que predicen varios aspectos de su biología. Además, los árboles proteómicos son eficaces para investigar genomas de nuevos virus secuenciados y para asignar lecturas fragmentadas a los genomas correctos. Sin embargo, cuando se estudia un gran número de fagos, este método resulta complejo y bastante lento.

Diapositiva 6

El siguiente método omite la comparación de secuencias de proteínas y sólo considera la presencia o ausencia de genes. Este enfoque basado en la “red” se ha utilizado para organizar la secuencia del genoma de virus de doble cadena y funciona a través de la predicción de los genes virales en todos los genomas que luego se traducen en proteínas. A su vez, estos últimos se organizan en grupos de proteínas. La evaluación del número de grupos de proteínas compartidas a lo largo del conjunto de datos, sirve para establecer un perfil de proteínas. La representación de esta información como un gráfico ponderado, con nodos que representan los genomas virales y la similitud de su contenido de proteína compartida generarán una “red” de intercambio de genes.

Diapositiva 7

A pesar de que sus genomas son varios órdenes de magnitud más pequeños en tamaño comparados con los de las bacterias u otros organismos, la secuenciación de bacteriófagos conlleva una serie de desafíos: (1) la obtención de material genético purificado, ya que los fagos no se replican por sí solos, por lo tanto, el aislamiento de material genético de los fagos conlleva varias etapas de purificación; (2) la amplificación por PCR puede ser subjetiva

ya que el genoma de algunos fagos puede tener extremos con un alto contenido en GC. Estos extremos pueden suponer un problema tanto para la PCR como para la secuenciación; (3) el genoma de los fagos puede contener una estructura genómica compleja como repeticiones extremadamente largas o invertidas y redundancias terminales que son muy problemáticas a la hora de ensamblar genomas completos; (4) en genómica bacteriana y humana, asignar las lecturas a un genoma de referencia puede ser una herramienta muy útil, sin embargo, en genómica fágica, esto no es posible debido a la falta de genomas de referencia para ningún fago. Una posible solución es un enfoque combinado de una tecnología de larga lectura, junto con un gran número de lecturas cortas obtenidas a partir de una segunda tecnología para la secuenciación eficiente del ADN de los genomas de bacteriófagos.