



**FACULTAD DE FARMACIA  
UNIVERSIDAD COMPLUTENSE**

**TRABAJO FIN DE GRADO**

**TÍTULO:**

**Técnicas de secuenciación de nueva generación  
para el estudio del microbioma humano**

Autor: Andrea Vázquez Escribano

Tutor: Concepción Gil García

Convocatoria: Junio 2016

## Índice:

1. Resumen .....	3
2. Introducción y antecedentes .....	3
3. Objetivos .....	6
4. Metodología .....	6
5. Resultados y discusión .....	6
5.1. Métodos de secuenciación de nueva generación .....	6
5.1.1. Método de 454 life sciences (Roche®) .....	7
5.1.2. Método solexa de Illumina .....	9
5.2. Aplicaciones de la secuenciación de nueva generación .....	11
5.2.1. Secuenciación del genoma completo o secuenciación “shotgun” .....	11
5.2.2. Secuenciación dirigida del gen 16s rRNA .....	13
5.2.3. Análisis de los datos de la secuenciación del genoma completo y de la secuenciación dirigida .....	15
6. Conclusiones .....	16
7. Bibliografía.....	17

## 1. RESUMEN

El microbioma humano es el conjunto de microorganismos presentes en el cuerpo humano. Se suelen encontrar organizados en distintas comunidades, que tienen una función útil para el organismo. Conocer las características de los microorganismos y de las comunidades que forman es importante a la hora de manipular el microbioma con el fin de tratar enfermedades. Actualmente, para caracterizar el microbioma, la tecnología más utilizada es la secuenciación de nueva generación, que permite aumentar la eficacia y disminuir el gasto económico con respecto a las técnicas utilizadas anteriormente. Para secuenciar el microbioma humano, se utilizan principalmente dos métodos de secuenciación de nueva generación: la secuenciación del genoma completo o “shotgun”, y la secuenciación dirigida del gen 16S rRNA. Este trabajo pretende analizar estos dos métodos para discernir cuál sería más útil para determinadas aplicaciones. La principal ventaja de la secuenciación del genoma completo (“shotgun”) es que permite conocer la información funcional y genética, identificar las diferencias entre microorganismos y detectar microorganismos poco abundantes. Por otro lado, la secuenciación dirigida del gen 16S rRNA permite distinguir entre distintos microorganismos de una forma mucho más sensible, ya que dicho gen es único de cada especie, aunque no sería posible diferenciar entre cepas.

**Palabras clave:** microbioma, secuenciación de nueva generación, shotgun, 16S rRNA.

## 2. INTRODUCCIÓN Y ANTECEDENTES

El microbioma humano es el conjunto de todos los microorganismos (procariotas, eucariotas, y virus) que habitan el cuerpo humano. Es una parte importante del organismo, hasta el punto de que las bacterias se encuentran en un número diez veces mayor que las células del cuerpo humano, y el genoma bacteriano está formado por 1.000 genes más que el genoma humano (1). De hecho, el genoma humano coincide en un 99,99% con el de otra persona, mientras que el microbioma humano coincide en un 70-90% (2). Estos microorganismos generalmente no son perjudiciales para nosotros, siendo algunos esenciales para mantener la salud. Un número cada vez mayor de estudios han demostrado que los cambios en la composición de nuestro microbioma se relacionan con numerosos estados de enfermedad, aumentando la posibilidad de que la manipulación de las comunidades microbianas pueda ser utilizada para tratar ciertas enfermedades (1).

El Proyecto del Microbioma Humano (HMP) se inició en 2008, con el fin de caracterizar el microbioma humano y el análisis de su papel en la salud humana y la enfermedad. El HMP utiliza la metagenómica, que se encarga de analizar las comunidades microbianas, en lugar de estudiar cada cepa bacteriana por separado. Cada comunidad microbiana debe ser estudiada en su conjunto, ya que muchos organismos nunca han sido cultivados de manera independiente. El análisis de una comunidad es importante, porque el crecimiento de una especie está influenciado por el ambiente que la rodea (1).

El objetivo fundamental de la investigación del microbioma humano es medir la estructura y dinámica de las comunidades microbianas, las relaciones entre sus miembros, que sustancias son producidas y consumidas, la interacción con el huésped, y las diferencias entre personas sanas y enfermas (3). Por otro lado, la complejidad de las comunidades microbianas hace que su caracterización sea complicada debido a que muchos organismos pueden requerir condiciones especiales de crecimiento y es necesario un muestreo en profundidad para estudiar las especies menos abundantes. Cada comunidad del microbioma tiene sus propias funciones. Por ejemplo, la microbiota intestinal realiza actividades metabólicas que evitan pérdidas de energía y ayudan a absorber nutrientes, además de proteger al huésped frente a microorganismos patógenos. Por otro lado, la microbiota intestinal es un factor esencial en ciertos trastornos patológicos como el cáncer de colon y enfermedades inflamatorias del intestino (4). Para entender por qué diferentes sitios tienen diferentes propiedades y los mecanismos que conducen a la alteración de los ecosistemas y de la enfermedad se requiere el conocimiento de la estructura y el comportamiento de las comunidades microbianas (1). Esta información será útil a la hora de manipular el microbioma con el fin de tratar enfermedades. Un ejemplo de esto es el trasplante de microbiota fecal (desde una persona sana a una persona enferma), que es útil para tratar enfermedades infecciosas de microorganismos oportunistas resistentes a antibióticos, como *Clostridium difficile*, de las que todavía se desconoce que microorganismos son lo que ayudan a tratar la enfermedad (5).

Las herramientas principales utilizadas por el HMP para la caracterización del microbioma humano se basan en la secuenciación del genoma completo y del gen 16S rRNA (1). La secuenciación de todo el genoma permite caracterizar las comunidades microbiomas, mientras que la secuenciación del gen 16S rRNA permite identificar las distintas especies. El gen 16S rRNA es importante porque es un gen distribuido universalmente que permite analizar las relaciones filogenéticas entre taxones distantes. Este gen solo se modifica por transferencia horizontal de genes, debido a que forma parte de los genes del cromosoma bacteriano. A

pesar de esto, el gen tiene regiones variables, lo que le proporciona la suficiente diversidad. Las regiones conservadas son útiles al diseñar los cebadores de PCR adecuados para su amplificación (6).

Además, también se pueden realizar análisis de RNA para describir el transcriptoma e identificar los virus de RNA. Los análisis no genómicos incluyen estudios proteómicos y metabolómicos.

El método tradicional de secuenciación más utilizado es el método de Sanger. Se basa en una DNA polimerasa que va incorporando bases a una cadena sencilla de DNA. La síntesis de la nueva cadena se interrumpe al incorporar un dideoxinucleótido, ya que debido a la ausencia de un grupo hidroxilo se impide que se añada otra base. De esta forma, se generan múltiples fragmentos de DNA de diferentes tamaños obteniéndose un patrón de bandas que permite leer la secuencia de bases. Un avance de este método es la utilización de terminadores fluorescentes que permiten que la lectura de la secuencia sea más sencilla ya que cada nucleótido emite una señal con una intensidad diferente. Gracias a este método, se consiguió completar la secuenciación del genoma humano en el año 2001. El principal problema del método de Sanger es que solo permite leer 96 secuencias de una vez.

Las tecnologías de secuenciación del DNA han ido avanzando rápidamente y en 2005 se publicó el primer método de secuenciación de nueva generación por la empresa Roche, basado en la pirosecuenciación. En 2006 la empresa Illumina publicó un nuevo método de secuenciación de nueva generación basado en un tipo de amplificación conocida como PCR de puente. Ambos métodos son más eficientes que la secuenciación de Sanger ya que la secuenciación se realiza simultáneamente en millones de fragmentos de DNA, aumentando el rendimiento. Además, esto permite reducir el precio por cada cadena de DNA secuenciada. La principal diferencia entre estos métodos es que el método de Roche es más utilizado para secuenciar fragmentos de mayor longitud, mientras que el método de Illumina se utiliza más para secuenciar fragmentos en mayores cantidades. Por esta razón, el Proyecto del Microbioma Humano utiliza el método de Roche para la secuenciación del gen 16S rRNA y el método de Illumina para la secuenciación del genoma completo (“shotgun”) (1).

### **3. OBJETIVOS**

El objetivo general de este trabajo es conocer las principales técnicas de secuenciación de nueva generación, así como analizar los distintos trabajos en los que se han utilizado las diferentes estrategias y la importancia de cada aplicación en el estudio del microbioma humano.

### **4. METODOLOGÍA**

Se ha realizado una revisión bibliográfica de distintos artículos y revistas de consulta on-line encontrados en PubMed y google académico, utilizando palabras clave como “microbiome”, “sequencing”, “next generation”, “Illumina”, “Roche”, “454 Life Sciences”, “16S rRNA”, “shotgun”. La selección inicial se llevo a cabo basándose en la fecha de publicación del artículo. Además, también ha sido utilizada información recogida de las páginas web oficiales de las compañías Illumina® y Roche®, así como de la página web oficial del Proyecto del Microbioma Humano (HMP).

### **5. RESULTADOS Y DISCUSIÓN**

#### **5.1. Métodos de secuenciación de nueva generación**

Todas las técnicas tienen en común la fragmentación del DNA, y su principal diferencia es el tipo de PCR utilizada. Los procedimientos de secuenciación de nueva generación más importantes son:

- Método de 454 Life Sciences-Roche® (pirosecuenciación): Fue el primer método de secuenciación de nueva generación. La amplificación se realiza mediante PCR de emulsión utilizando microesferas. Los productos de la secuenciación se analizan según la intensidad de señal luminosa obtenida.
- Método Solexa de Illumina®: La amplificación se realiza mediante PCR de puente. Para la secuenciación se utilizan cuatro nucleótidos con terminadores reversibles y cada ciclo tiene lugar con los cuatro nucleótidos simultáneamente.

### 5.1.1. Método de 454 Life Sciences (Roche®)

Se realiza en un sistema conocido como Genome Sequencer FLX y tienen lugar los siguientes pasos (8):

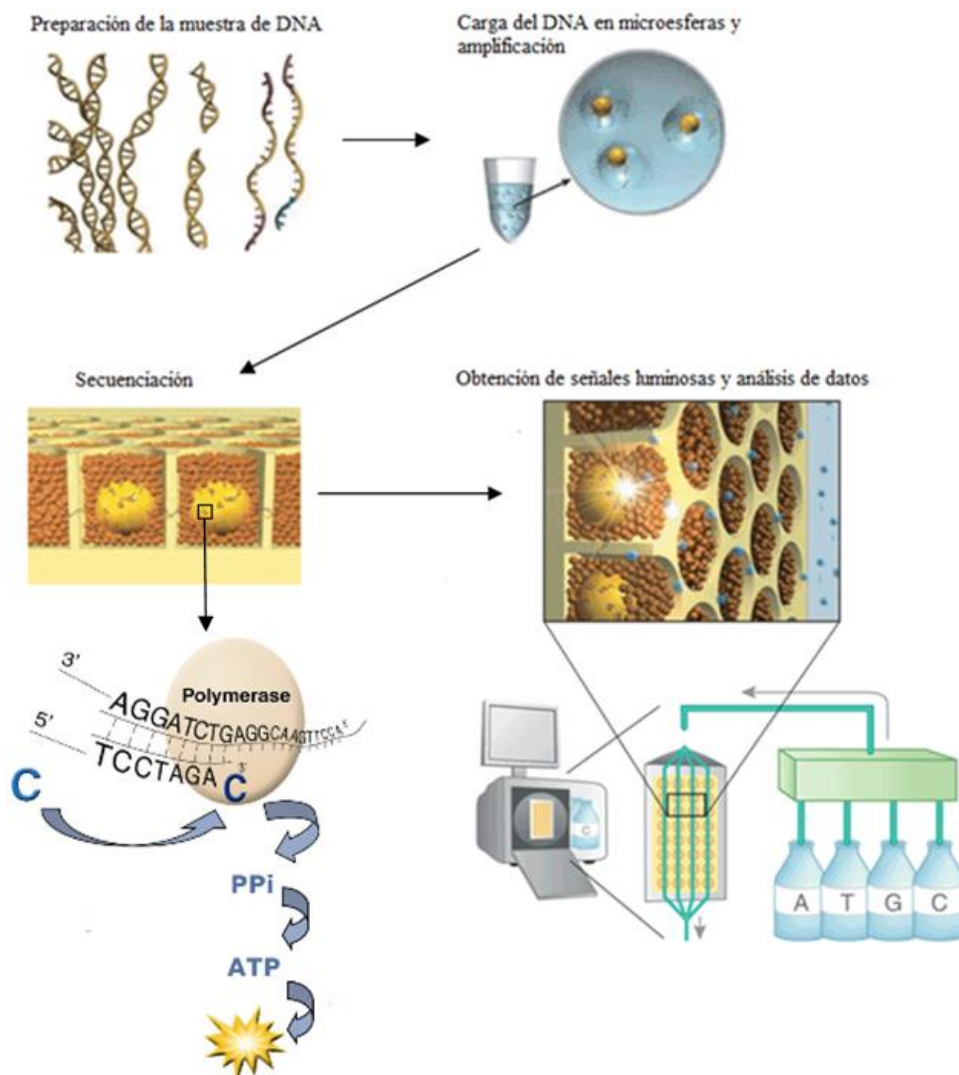
1. **Preparación de la muestra de DNA:** primero se rompe la doble hélice de DNA en fragmentos de doble cadena de aproximadamente 400 a 600 pares de bases. En el siguiente paso se conectan los cebadores a los fragmentos de DNA para la amplificación posterior. Por último, los fragmentos de cadena doble se separan en cadenas simples que serán las hebras moldes para la amplificación por PCR en emulsión.
2. **Carga de la muestra de DNA en microesferas:** Los fragmentos de DNA de la biblioteca se ponen sobre microesferas a través de un proceso de amplificación en emulsión (emPCR). Esto hace que las señales producidas tras la amplificación sean fácilmente detectables. La amplificación por el método de Roche tarda aproximadamente ocho horas, mientras que la amplificación por el método de Sanger dura aproximadamente tres semanas.

En la fase inicial del proceso de amplificación, se ponen en agua los fragmentos de la biblioteca de DNA, las perlas de captura y los reactivos de la enzima, que se inyectan en pequeños recipientes cilíndricos que tienen un aceite sintético. De este modo, se forman gotas de aceite en el agua, conteniendo cada una un fragmento de DNA. La mezcla de agua tiene una enzima que produce la amplificación de cada cadena sencilla de DNA presente en las gotas de aceite, teniendo lugar la PCR de emulsión. Los fragmentos amplificados se quedarán inmovilizados sobre las perlas de captura. Las microesferas con más de un tipo de DNA se filtran durante el procesamiento de las señales de secuenciación.

3. **Secuenciación del DNA:** En la secuenciación por síntesis de Roche, un fragmento de DNA de una sola cadena se va complementando con los nucleótidos mediante una enzima hasta formarse un fragmento de doble cadena. Cuando estos nucleótidos se incorporan en las cadenas de DNA, las enzimas convierten los productos químicos generados durante la incorporación de nucleótidos en una señal químico-luminiscente que es registrada por una cámara. La intensidad de la señal es proporcional al número de nucleótidos incorporados y varía con el número consecutivo de nucleótidos complementarios que se analizan en el fragmento de DNA. Por ejemplo, si hay tres

adeninas en el fragmento de cadena sencilla, la cantidad de luz generada sería tres veces mayor que la de una sola adenina. Por último, las señales creadas en el proceso de secuenciación se analizan por el software del Sistema de Secuenciación de Roche para generar millones de bases secuenciadas por hora a partir de una única prueba.

4. **Análisis de datos:** Los datos generados por la secuenciación de Roche tienen la ventaja de tener un alto rendimiento y una longitud de lectura más larga para crear una imagen más completa del genoma, lo cual es la principal diferencia con otros métodos de secuenciación. Esta tecnología permite acelerar el tiempo, la calidad y la profundidad de los resultados de secuenciación en cada procesamiento, permitiendo secuenciar todo el genoma de los individuos. Los resultados de cada secuenciación son recogidos y se comparan con los genomas de referencia para detectar regiones de coincidencia y diferencias.

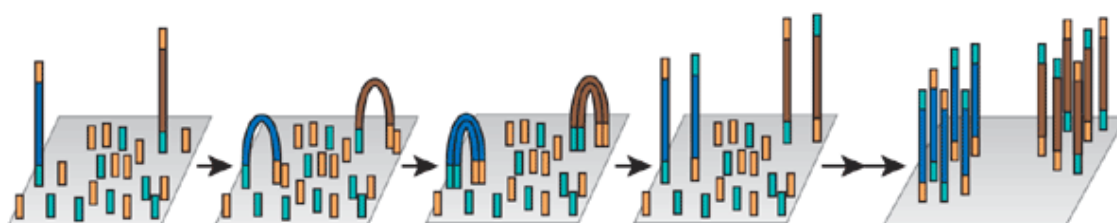


**Figura 1.** Esquema de la secuenciación de nueva generación por el método de Roche® (modificado de 8, 9)

### 5.1.2. Método Solexa de Illumina®

La secuenciación por síntesis (sequencing by synthesis, SBS) de Illumina se divide en cuatro etapas:

1. **Preparación de la genoteca:** son muestras de fragmentos aleatorios de DNA o de cDNA, seguidos de adaptadores (linkers) 5' y 3'. Al mismo tiempo, el marcaje combina reacciones de fragmentación y unión en un solo paso que aumenta la eficiencia del proceso de preparación de la genoteca. Después, los fragmentos unidos con los adaptadores son amplificados y purificados (10).
2. **Generación de cluster:** Los moldes para la secuenciación son inmovilizados en una superficie de celdillas que presenta el DNA accesible para las enzimas mientras asegura una alta estabilidad de los moldes ligados a una superficie y uniones no específicas de nucleótidos etiquetados con fluorocromos. Cada fragmento es amplificado y la amplificación en fase sólida crea hasta 1.000 copias idénticas de cada molde de moléculas. La amplificación se realiza por PCR de puente (los fragmentos se amplifican a partir de unos cebadores unidos a una superficie sólida) (11). Para la secuenciación de las cadenas sencillas se añade un cebador, mientras que para el DNA de doble cadena se elimina la cadena original, dejando la cadena complementaria como molde para la segunda secuenciación (12).

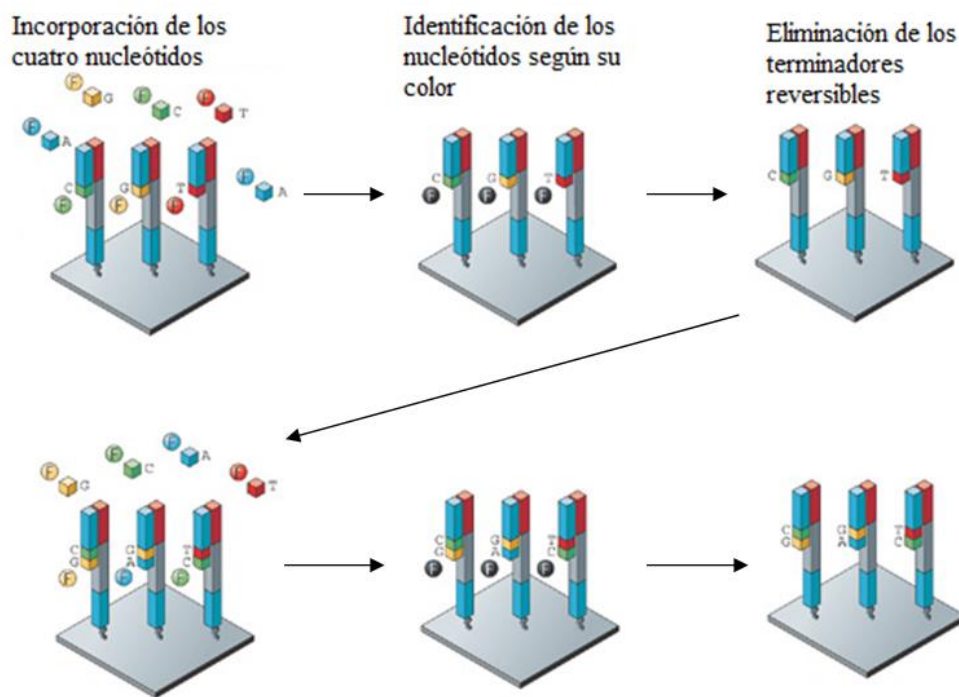


**Figura 2.** Amplificación en puente según el método de Illumina® (obtenido de 13)

3. **Secuenciación por síntesis (SBS):** se secuencian cadenas molde de DNA por ciclos repetidos en los que se va añadiendo cada base de forma individual. La tecnología SBS de Illumina utiliza cuatro terminadores reversibles, cada uno marcado con un fluoróforo diferente, de forma que se detecta cada base cuando son incorporadas a las cadenas moldes de DNA. Los cuatro terminadores reversibles se encuentran ligados a los nucleótidos y bloquean la polimerización, por lo que la enzima polimerasa solo puede añadir una base a cada cadena de DNA que se está sintetizando. Al añadirse los nucleótidos simultáneamente y no secuencialmente, se produce una competencia

natural por lo que se minimizan los errores de incorporación. Como las bases individuales se añaden a todos los moldes de una manera uniforme, el proceso de secuenciación produce unas secuencias de DNA de longitud uniforme. El resultado es una secuenciación base a base muy precisa que elimina virtualmente los errores específicos de la secuencia, permitiendo asignaciones de bases fuertes entre el genoma, incluyendo regiones de secuencias repetidas y homopolímeros. De esta forma, se consiguen secuenciar los millones de clusters de la superficie de la célula al mismo tiempo.

Después de cada ciclo de incorporación, se eliminan los terminadores y se determina la identidad de la base insertada por la excitación inducida por láser de los fluoróforos y la consiguiente formación de imágenes que se registran. La asignación de bases está hecha directamente de las medidas de intensidad de señal durante cada ciclo, lo cual reduce las tasas de error comparado con otras tecnologías (12).



**Figura 3.** Secuenciación por síntesis mediante terminadores reversibles según el método de Illumina® (obtenido de 14)

- Análisis de los datos:** El enfoque de la secuenciación Illumina se basa en la lectura simultánea de grandes cantidades de secuencias. El muestreo en profundidad y una cobertura uniforme asegura una alta precisión en la determinación de diferencias genéticas. Cada lectura de base tiene una puntuación asignada según la calidad por lo que el software de Illumina puede hacer diferencias y generar puntuaciones seguras.

Se pondera la contribución de cada base a la secuencia y se detecta las variantes de las secuencias (11). Hay muchas variantes del análisis posibles como el polimorfismo de nucleótido único (SNP), la identificación de inserciones o deleciones, los recuentos de lecturas para los métodos RNA, los análisis filogenéticos o metagenómicos, y más. El software es capaz de recoger los datos, procesarlos y analizarlos para que haya una mínima intervención del usuario (11).

Estos métodos de secuenciación de nueva generación permiten una amplia variedad de técnicas, resolviendo cualquier pregunta relacionada con el genoma, transcriptoma, o epigenoma de cualquier organismo. Los métodos de secuenciación difieren principalmente por como son obtenidas las muestras de DNA o RNA y por las opciones de los análisis de los datos utilizados. Después de que las genotecas de secuenciación sean preparadas, la etapa de secuenciación real sigue siendo la misma, independientemente del método.

## **5.2. Aplicaciones de la secuenciación de nueva generación**

### **5.2.1. Secuenciación del genoma completo o secuenciación “shotgun”**

Se suele asociar con secuenciación de genomas humanos, pero también puede ser útil para la secuenciación de genomas microbianos. La metodología actual se basa principalmente en la secuenciación “shotgun” de Illumina (secuenciación de fragmentos aleatorios) y una variedad de métodos para ensamblar las lecturas en un genoma (no es un verdadero genoma completo, pero cubre casi todo el genoma). El principal método de ensamblaje es conocido como *de novo* y consiste en una secuenciación en la que no hay secuencia de referencia disponible para complementar. Las lecturas de las secuencias se ensamblan como “contigs” y la calidad de la cobertura de los datos de la secuencia *de novo* depende del tamaño y de la continuidad de los “contigs” (es decir, el número de espacios en los datos). Otro factor importante en la generación de secuencias *de novo* de alta calidad es la diversidad de tamaños incluidos en la genoteca. La combinación de insertos cortos emparejados al final e insertos largos de secuencias es el enfoque más eficaz para la máxima cobertura en todo el genoma, ya que dicha combinación permite la detección de la más amplia gama de tipos de variantes estructurales y es esencial para identificar con precisión los reordenamientos más complejos. Los avances tecnológicos permiten realizar uniones *de novo* de alta calidad utilizando lecturas NGS. Por ejemplo, mediante la secuenciación *de novo*, el genoma de *E. coli* se puede montar en tan sólo 15 minutos utilizando un ordenador de escritorio Windows de 32 bits con 32 GB de RAM (11).

La secuenciación del genoma completo permite identificar las diferencias entre microorganismos aislados y conocer la información funcional y genética. Además, gracias a la secuenciación de todo el genoma también se pueden identificar las mutaciones genéticas que provocan el aumento de la virulencia. Existe una considerable variación en los genomas entre cepas de la misma especie. Un ejemplo de la utilidad de este método fue en 2011, durante el brote de *E. coli* en Europa (15).

Además, la secuenciación del genoma completo permite hacer un censo basado en los genes, que puede ser realizado mediante secuenciación aleatoria; pero en una comunidad microbiana con muchas especies diferentes, es necesaria una secuenciación más profunda para conocer componentes menores que pueden ser importantes. La concentración de bacterias en el intestino es de  $10^{11}$  células/ml, así que para que un microorganismo que está presente a una concentración de  $1 \times 10^6$  hay  $10^5$  células/ml, que es suficiente para que los productos del microorganismo, como metabolitos y toxinas, tengan un efecto sobre la comunidad y el huésped (3).

La secuenciación “shotgun” también es importante para la metagenómica ya que muestran el contenido de genes de la comunidad, que es útil para definir las capacidades de la comunidad e identificar los microorganismos aislados. Mediante las lecturas de la secuencia los datos obtenidos pueden ser comparados con las bases de datos para identificar qué genes están presentes. Además de esto, también se puede conocer el contenido de los genes. Estos métodos hacen que los datos de las secuencias de metagenómica se conviertan en catálogos de genes (3).

Los datos de la secuenciación de “shotgun” proporcionan información sobre los organismos que conforman las comunidades. También permiten relacionar las secuencias de “shotgun” a genomas de referencia bacterianos, lo que ayuda a determinar la abundancia relativa de especies sobre la base del número de lecturas que se relacionan con cada genoma de referencia (también útil para estudios comparativos). Esto ha sido utilizado, por ejemplo, en el proyecto MetaHIT (investigación del microbioma del tracto gastrointestinal) para clasificar a los individuos en diferentes grupos sobre la base de la estructura de la comunidad en muestras fecales (16); y en el Proyecto del microbioma humano (17).

Gracias a la secuenciación “shotgun”, también se han podido secuenciar virus humanos (18). Esto ha llevado a su identificación, así como a la detección de virus conocidos en los sujetos sanos y enfermedades de etiología desconocida (19).

### 5.2.2. Secuenciación dirigida del gen 16S rRNA

Con la secuenciación dirigida, un subconjunto de genes o regiones del genoma son aislados y secuenciados. Permite centrarse en los análisis de datos de áreas específicas de interés y una mayor cobertura durante la secuenciación, lo que facilitaría a los investigadores identificar variantes raras y mutaciones. Este tipo de secuenciación es útil para examinar los genes en las vías específicas, o para estudios de seguimiento de la secuenciación del genoma completo. Actualmente, hay dos métodos diferentes de enriquecimiento para la secuenciación (11):

- Enriquecimiento dirigido: regiones específicas de interés son capturadas por hibridación con sondas con biotina y aisladas magnéticamente.
- Enriquecimiento por amplificación: se realiza mediante PCR con oligos multiplexados. Esto permite descubrir, validar, o detectar variantes genéticas. Algunas de sus aplicaciones son: descubrir mutaciones somáticas raras en muestras complejas, y secuenciación del gen bacteriano 16S rRNA, un método ampliamente utilizado para estudios de filogenia y taxonomía (11).

La secuenciación de los genes 16S rRNA es el principal método para hacer un censo de una comunidad porque permite la identificación de microorganismos poco abundantes (20).

El gen 16S rRNA es diferente para cada especie bacteriana. Una especie bacteriana es difícil de definir, pero las secuencias de 16S rRNA indican al menos el 97% de identidad. Una secuencia del gen 16S rRNA tiene nueve regiones hipervariables cortas que diferencian los distintos taxones bacterianos, por lo que la secuenciación de una o más de estas regiones están dirigidos a hacer un censo de la comunidad (3).

Antes de la introducción de los métodos de nueva generación, la secuenciación de genes 16S rRNA se llevaba a cabo por el método de Sanger, que limitaba el estudio de los microorganismos menos abundantes por su menor precisión. La secuenciación de nueva generación para el análisis de genes 16S rRNA ha conseguido mejoras en el precio y la profundidad de muestreo. El método de Roche es uno de los más utilizados en la secuenciación de los genes 16S rRNA, siendo utilizado, por ejemplo, para la secuenciación de microorganismos poco abundantes en el mar (21). La secuenciación por el método de Roche lee aproximadamente 400 bases, por lo que solo una parte del gen 16S rRNA será secuenciado. Esto provoca que haya una menor sensibilidad, aunque el hecho de que sea un muestreo más profundo y los costes sean menores supera la pérdida de sensibilidad.

El Proyecto del Microbioma Humano (HMP) ha secuenciado organismos gracias al método Roche de secuenciación del gen 16S rRNA, identificando la estructura de la comunidad microbiana de cada sitio. Aunque para un estudio más profundo ha sido necesario el uso de la secuenciación del genoma completo (“shotgun”) por el método de Illumina (22). La secuenciación del gen 16S rRNA también ha sido utilizada en el estudio de los efectos de los antibióticos. Por ejemplo, se realizó un estudio sobre los efectos del ciprofloxacino y gracias a la pirosecuenciación (Roche) del gen 16S rRNA se pudo saber que después del tratamiento, la diversidad y la uniformidad de la microbiota intestinal había disminuido. Además, también se pudo observar una diferente respuesta interindividual al tratamiento con ciprofloxacino, pese a que se comprobó que antes del tratamiento la microbiota de dos personas no relacionadas compartía una gran similitud (23).

Cuando se utiliza la secuenciación de genes 16S rRNA para comparar individuos no es necesario saber qué organismos están presentes, solamente si los espectros de secuencias de genes 16S rRNA son similares y el grado de diferencia entre las muestras. Muchos proyectos que comparan individuos sanos y enfermos para determinar si hay una diferencia en el microbioma debido a efectos de la dieta, el tratamiento con antibióticos o factores ambientales sobre el microbioma; se centran en la detección de diferencias en las comunidades, en lugar de identificar los taxones reales (3). Las técnicas de secuenciación de nueva generación de genes de 16S rRNA permiten conocer las diferencias entre un gran número de individuos gracias a un muestreo profundo. El método de secuenciación de Illumina se ha utilizado en el análisis del metagenoma, pero debido a que esta plataforma de secuenciación lee de 100 a 150 bases, sólo puede ser secuenciada una única región hipervariable (3). Gracias a la profundidad de esta tecnología se han hecho estudios en los que se identificaban taxones de la microbiota bucal que todavía no habían sido descubiertos (24). También se han realizado estudios en los que se han tratado de identificar géneros mediante la secuenciación de Illumina. Hay estudios que afirman que las lecturas realizadas por la tecnología de Illumina no permiten clasificar a los organismos en géneros debido a que la longitud de secuencias usadas no es lo suficiente larga (25). Por otro lado, otro estudio afirma que las lecturas cortas realizadas por Illumina son suficientes para asignar a los organismos el nivel de género o de especie en la mayoría de los casos (26). En cualquier caso, el método Roche, al leer un mayor número de bases, sí que sería útil para diferenciar con seguridad entre especies.

Una aplicación de la secuenciación de nueva generación es su uso en un estudio de microbiomas vaginales en pacientes con el VIH, en el que se comparaban la vaginosis antes y después de la terapia con antibióticos gracias a la gran resolución filogenética proporcionada por el método de Illumina (27). Al secuenciar virus, se utiliza la tecnología de Illumina, ya que las secuencias son más cortas que las generadas en la pirosecuenciación por Roche y proporciona la sensibilidad necesaria para detectar secuencias de virus raros. Las longitudes cortas parecen ser suficientes para la detección de nuevos virus dentro de una muestra de una comunidad microbiana, aunque el montaje de secuencias de Illumina también se puede utilizar para lograr secuencias más largas contiguas (28).

Como resultado de los aumentos excepcionales en el número de lecturas y el bajo costo, cada vez se está utilizando más para la secuenciación de genes 16S rRNA (3).

### 5.2.3. Análisis de los datos de la secuenciación del genoma completo y de la secuenciación dirigida

Los datos metagenómicos son una fuente de información para los métodos de secuenciación y análisis. En el análisis de los datos se distinguen tres fases:

1. La primera fase consiste en el procesamiento de los datos. Para la secuenciación del gen 16S rRNA, la calidad del análisis es importante para que los organismos no estén mal clasificados. El procesamiento inicial es dirigido hacia la calidad de la secuencia, el quimerismo (una lectura formada a partir de diferentes genes 16S rRNA), la lectura de la longitud después de la eliminación de las bases de baja calidad y problemas relacionados. Para datos de secuencias “shotgun” deben hacerse, por ejemplo, lecturas duplicadas y eliminación de la contaminación a partir de secuencias humanas. En el procesamiento de lectura para el análisis del viroma también hay que eliminar secuencias bacterianas (3).
2. La segunda fase consiste en la generación de diversos conjuntos de datos. Para el análisis del gen 16S rRNA, se realizan tablas de taxones que se comparan con bases de datos de secuencias 16S rRNA o mediante el uso de softwares que agrupan las lecturas en OTUs (unidad taxonómica operacional, permite diferenciar los datos obtenidos en la secuenciación mediante la similitud para la clasificación de microorganismos). La comparación de lecturas de “shotgun” con bases de datos de genes mediante el uso de BLAST (Basic Local Alignment Search Tool, compara una secuencia con las secuencias de las bases de datos) da lugar a las listas de genes (3).

Si la secuencia no es encontrada en las bases de datos puede deberse a dos razones: que el organismo no tenga secuenciado el genoma o que las funciones de los genes no hayan sido descubiertas y no se conozca como puede repercutir esto sobre el resto de los genes reconocidos.

3. La tercera fase de análisis utiliza estos datos para producir representaciones de la similitud de las comunidades, abundancia, biodiversidad y otros descriptores de la estructura de la comunidad. Los resultados obtenidos mediante BLAST se utiliza para construir las vías metabólicas para conocer las funciones de la comunidad gracias a programas informáticos que analizando la secuencia de DNA son capaces de reconocer la función de la comunidad (29). El análisis informático también es útil para conocer el lugar y la estructura de las comunidades a lo largo del tiempo (30).

Las comparaciones con bases de datos de genes a nivel de proteínas son complicadas porque las secuencias de “shotgun” deben traducirse en polipéptidos en los seis marcos de lectura, y cada uno debe compararse con una base de datos de genes representados en el nivel de proteínas. El uso de los programas BLAST convencionales para comparar grandes conjuntos de datos, como en el HMP o el análisis de datos procedentes de virus (28), podría llevar décadas, por lo que deben utilizarse programas BLAST acelerados. En el HMP se han establecido protocolos para que el procesado de los datos se realice a varios niveles y en una única lectura, guiándose por los datos de las secuencias 16S rRNA y del genoma completo, y por la situación esperada de cada taxón (31).

## **6. CONCLUSIONES**

La secuenciación del microbioma es la mejor estrategia para conocer las características de los microorganismos y las comunidades microbianas que habitan el cuerpo humano. La tecnología más utilizada para secuenciar genomas es la secuenciación de nueva generación. Las dos técnicas principales son: la secuenciación del genoma completo y la secuenciación del gen 16S rRNA. Ambas técnicas son muy eficaces, siendo cada una útil en diferentes aplicaciones.

La secuenciación del genoma completo:

- Permite conocer la información funcional y genética de cada microorganismo.
- Permite diferenciar entre las distintas cepas de una misma especie, lo cual no se puede realizar mediante la secuenciación del gen 16S rRNA, ya que en las distintas cepas de una misma especie el gen es el mismo. Esto es útil para conocer las cepas causantes de brotes de enfermedades infecciosas.
- Permite conocer microorganismos poco abundantes en el microbioma humano ya que se realiza una secuenciación más profunda.

La secuenciación del gen 16S rRNA:

- Permite evaluar los microorganismos presentes en las comunidades microbianas, pero no aporta información funcional o genética.
- Es más sensible que la secuenciación del genoma completo debido a que la cantidad de material para secuenciar es menor.
- Tiende a producir sesgos que pueden ser debidos a la elección de la secuencia del gen 16S rRNA para ser secuenciada.

Muchos laboratorios están trabajando en el desarrollo de nuevos métodos de secuenciación o en la mejora de los ya existentes, ya que esto ayudaría a conocer mejor el microbioma humano pudiendo dar lugar a nuevas vías terapéuticas.

## 7. BIBLIOGRAFÍA

1. Hmpdacc.org. Human Microbiome Project - About the HMP. 2016 [acceso 18 de abril de 2016]. Disponible en: <http://hmpdacc.org/overview/about.php>
2. Gutiérrez P. A. Caracterización de dos glicosiltransferasas productoras de oligosacáridos prebióticos de las levaduras *Phaffia rhodozyma* y *Rhodotorula dairenensis*. Universidad Autónoma de Madrid; 2013.
3. Weinstock G. M. Genomic approaches to studying the human microbiota. *Nature*. 2012; 489(7415): 250-256.
4. Guarner F., Malagelada J. R. Gut flora in health and disease. *The Lancet*. 2003; 361(9356), 512-519.

5. Langdon A., Crook N., Dantas G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Medicine*. 2016; 8(1), 1.
6. Vetrovsky T., Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*. 2013 [acceso 18 de abril de 2016]; 8(2), e57923. Disponible en: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0057923>
7. Ogiza A., Gamba Alessio., Pereda A., Pérez E., Lázaro N., Bettinelli P. Secuenciación de nueva generación. *Fundaciontelefonica.com* [Internet]. 2010 [acceso 13 de abril de 2016]. Disponible en: <https://biotecnologia.fundaciontelefonica.com/2010/04/19/secuenciacion-de-nueva-generacion/>
8. 454.com. How is genome sequencing done? [sede Web]. *454 Life Sciences*; 2016 [acceso 30 de Abril de 2016]. Disponible en: [http://www.454.com/downloads/news-events/how-genome-sequencing-is-done\\_FINAL.pdf](http://www.454.com/downloads/news-events/how-genome-sequencing-is-done_FINAL.pdf)
9. Biosonda S.A. Reacción de pirosecuenciación [imagen modificada de Internet]. 2016 [acceso 30 de abril de 2016]. Disponible en: <http://www.biosonda.cl/producto/pyromark>
10. Illumina.com . An introduction to Next Generation Sequencing technology (NGS) [sede Web]. *Illumina*. 2016 [acceso 16 de abril de 2016]. Disponible en: <http://www.illumina.com/technology/next-generation-sequencing.html>
11. Illumina.com. Technology spotlight: Illumina Sequencing [Sede Web]. *Illumina*. 2010 [acceso 16 de abril de 2016]. Disponible en: [http://www.illumina.com/documents/products/techspotlights/techspotlight\\_sequencing.pdf](http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf)
12. Bentley D.R., Balasubramanian S., Swedlow H.P., et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456(7218):53-59.
13. Shendure J., Ji H. Next-generation DNA sequencing. *Nature*. 2008; 26(19):1135-1145.
14. Metzker M.L. Sequencing technologies - the next generation. *Nature*. 2010; 11(1):31-46.

15. Grad T.H., Lipsitch M., Feldgarden M., et al. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe. PNAS. 2012 [acceso 16 de abril de 2016]. 109(8), 3065-3070.
16. Metahit.eu. MetaHIT: Sequencing [Sede Web]. MetaHIT. 2016 [acceso 16 de Abril de 2016]. Disponible en: <http://www.metahit.eu/index.php?id=233>
17. Qin J., Li R., Raes J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010; 464(7285):59–65.
18. Breitbart M., Rohwer F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. Biotechniques. 2005; 39(5):729–736.
19. Wylie K. M., Mihindukulasuriya K. A., Sodergren E., Weinstock G. M., Storch G. A. Sequence analysis of the human virome in febrile and afebrile children. PLoS ONE. 2012 [acceso 16 de abril de 2016]; 7(6): e27735. Disponible en: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0027735>
20. Bent S. J., Pierson J. D., Forney L. J. Measuring species richness based on microbial community fingerprints: the emperor has no clothes. Applied and Environmental Microbiology. 2007; 73(7): 2399–2401.
21. Sogin M. L., Morrison H. G., Huber J. A., Welch D. M., et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proceedings of the National Academy of Sciences. 2006; 103(32): 12115–12120.
22. The NIH HMP Working Group et al. The NIH Human Microbiome Project. Genome Research. 2009; 19(12): 2317–2323.
23. Dethlefsen L., Huse S., Sogin M. L., Relman D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol. 2008 [acceso 17 de abril de 2016] 6(11): e280. Disponible en: <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0060280>
24. Lazarevic V., Whiteson K., Huse S., et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. J. Microbiological Methods. 2009; 79(3), 266–271.

25. Claesson, M. J., Wang Q. et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 2010 [acceso 17 de abril de 2016] 38(22), e200. Disponible en:  
<https://nar.oxfordjournals.org/content/38/22/e200.full>
26. Gloor G. B., Hummelen R., et al. Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS ONE.* 2010 [acceso 17 de abril de 2016] 5(10): e15406. Disponible en:  
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0015406>
27. Hummelen R., Fernandes A. D., Macklaim J. M., et al. Deep sequencing of the vaginal microbiota of women with HIV. *PLoS ONE.* 2010 [acceso 17 de abril de 2016] 5(8), e12078. Disponible en:  
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012078>
28. Wylie K. M., Weinstock G. M., Storch G. A. Emerging view of the human virome. *Translational Res.* 2012; 160(4): 283-290.
29. Abubucker S., Segata N., Goll J., et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* 2012 [acceso 17 de abril de 2016] 8(6), e1002358. Disponible en:  
<http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002358>
30. Caporaso J. G., Lauber C. L., Costello E. K., et al. Moving pictures of the human microbiome. *Genome Biol.* 2011 [acceso 17 de abril de 2016] 12(5), R50. Disponible en: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-5-r50>
31. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature.* 2012; 486(7402), 215–221.