

METAGENÓMICA DE RNA RIBOSOMAL 16S (16S rRNA) COMO HERRAMIENTA DIAGNÓSTICA EN MICROBIOLOGÍA: MÉTODOS Y APLICACIONES BASADOS EN *NEXT-GENERATION SEQUENCING*

Martínez José Gregorio^{1,2,3}

¹ Grupo de Investigación Biociencias, Facultad de Ciencias de la Salud, Institución Universitaria Colegio Mayor de Antioquia, Cra 78 N° 65 - 46, Código Postal 4-72, Medellín, Colombia.

² Grupo de pesquisa em Genética Molecular e Citogenética, Laboratório de Genômica e Proteômica, Escola Superior de Ciências da Saúde, Universidade do Estado do Amazonas, Avenida Carvalho Leal 1777 - Cachoeirinha, Manaus, Brasil.

³ Laboratório de Evolução e Genética Animal, Departamento de Genética, Instituto de Ciências Biológicas, Universidade Federal do Amazonas, Av. Gal. Rodrigo Octávio Jordão Ramos, 3000, Setor Sul, Coroadó I, Manaus, Brasil.

Abstract

Identification methods in traditional microbiology are of great relevance in medicine and bioprospecting, however, there are still limitations. It is estimated that only 50% of the bacteria are cultivable, and that between 10-20% of the isolates are new organisms for science, so that culture-dependent identification approaches can induce error in several cases. The development of DNA sequencing and molecular marker technologies has led to the leap into new and faster forms of bacterial identification. For example, metagenomic analyzes of the 16S rRNA gene sequence. The most classical method consists in the amplification of this molecular marker (via PCR) in the sample, cloning of the PCR fragments and SANGER sequencing of the species-specific fragment. However, this method either underestimates the number of actual taxonomic units in a sample or ignores the abundance of them within the community from which it was extracted. To solve this limitation, Next Generation Sequencing (NGS) is useful, which allows partial parallel sequencing of the 16S gene for all species harbored in the sample. Bioinformatics is the main ally of this latter method for the processing of sequences and identification. The 16S metagenomics have applications in the determination of the composition of intestinal, buccal, leather and soil microbiots, which have allowed to contribute for new information to the science in diagnostic and pathological association studies. Entering the philosophy of these techniques is fundamental, since metagenomics will revolutionize the research and development processes of modern microbiology, of which you are part.

Keywords: Bacterial community, molecular identification, metagenomic library, massive parallel sequencing

Resumen

Los métodos de identificación en la microbiología tradicional son de grande relevancia en la medicina y la bioprospección, sin embargo, aún persisten limitaciones. Se estima que sólo el 50% de las bacterias son cultivables, y que entre 10-20% de los aislados son organismos nuevos para la ciencia, por lo que los abordajes de identificación dependientes de cultivo pueden inducir al error en varios casos. El desarrollo de tecnologías de secuenciamiento de ADN y de marcadores moleculares, han propiciado el salto hacia nuevas y más rápidas formas de identificación bacterianas. Por ejemplo, los análisis metagenómicos de secuencia del gen 16S rRNA. El método más clásico consiste en la amplificación de este marcador molecular (vía PCR) en la muestra, clonación de los fragmentos de PCR y secuenciamiento SANGER del fragmento especie-específico. Sin embargo, este método llega a subestimar el número de unidades taxonómicas reales en una muestra o a ignorar la abundancia de las mismas dentro de la comunidad de la que se extrajo. Para solucionar esta limitación, es útil el secuenciamiento de nueva generación (*Next-Generation Sequencing*, NGS), el cual permite el secuenciamiento parcial en paralelo del gen 16S para todas las especies albergadas en la muestra. La bioinformática es la aliada principal de este último método para el procesamiento de las secuencias y la identificación. La metagenómica de 16S posee aplicaciones en la determinación de la composición de microbiotas intestinales, bucales, piel y suelo, los cuales han permitido aportar nueva información a la ciencia en estudios diagnósticos y de asociación con patologías. Adentrarnos en la filosofía de estas técnicas es fundamental, pues la metagenómica revolucionará los procesos de investigación y desarrollo de la microbiología moderna, de la cual usted hace parte.

Palabras clave: Comunidad bacteriana, identificación molecular, librerías metagenómicas, secuenciamiento paralelo en masa.

Los métodos de identificación en la microbiología tradicional basados en la caracterización fenotípica y bioquímica, han contribuido de forma significativa a diversos campos como la medicina moderna y a la bioprospección biotecnológica, sin embargo, estas aproximaciones han tenido limitaciones. Se estima que sólo el 50% de las bacterias son cultivables (Clemente et al. 2015), y que entre 10-20% de los aislados son organismos nuevos para la ciencia (Zhang and Versalovic 2007), por lo que los abordajes de identificación dependientes de cultivo pueden redundar en un desafío para la caracterización basada en el fenotipo/bioquímica, resultando en una tipificación errónea en varios casos y la subestimación severa del potencial de especies que realmente existen en una comunidad bacteriana examinada (Clarridge 2004).

Las técnicas moleculares como la PCR/secuenciamiento han sido particularmente útiles en microbiología para la identificación, debido a que no requieren del cultivo de la muestra inicial, existe menos consumo de tiempo en el retorno de la información y la misma puede ser almacenada en bancos de datos para posteriores análisis. El desarrollo acelerado de las tecnologías de secuenciamiento de ADN y de marcadores moleculares, han propiciado el salto hacia nuevas y más rápidas formas de identificación de unidades taxonómicas bacterianas. Por ejemplo, los análisis de secuencia del gen 16S rRNA, de aproximadamente 1500 pares de bases con 9 regiones variables (v1-v9) y 10 conservadas, se ha constituido en un método más exacto en la identificación de especies de micobacteria (y otros microorganismos de difícil identificación) suplementando tipificación fenotípica en muchos centros de diagnóstico (Hall et al. 2003).

El método más clásico consiste en la amplificación del marcador molecular 16S (vía PCR) para la muestra total conteniendo el genoma de toda la comunidad bacteriana o pool genómico, posterior individualización de los fragmentos de PCR por especie (vía clonación en bacterias recombinantes) y secuenciamiento SANGER del fragmento especie-específico (1400-1500 pb) contenido en las colonias originadas por las bacterias recombinantes. Este método puede lograr identificar hasta 500 especies o tal vez más (Contreras et al. 2010), dependiendo del esfuerzo, recursos y espacio del investigador para generar igual o mayor número de colonias recombinantes, cuya cantidad es quien determina a la postre el número máximo de especies o unidades taxonómicas que se identificarán. Aunque este método es importante, puede llegar a subestimar el número de unidades taxonómicas reales en una muestra o a ignorar la abundancia de las mismas dentro de la comunidad de la que se extrajo.

Dado que cubrir toda la diversidad bacteriana en una muestra, además de su abundancia, es prioridad para quien pretende hacer bioprospección, otras herramientas y tecnologías más robustas deberán ser usadas.

El secuenciamiento de nueva generación (*Next Generation Sequencing*, *NGS*), el cual permite el secuenciamiento en paralelo de genomas enteros o genes particulares (marcador/marcadores) para todas las especies albergadas en la muestra (metagenómica), es una de los métodos más avanzados en la identificación taxonómica de orden molecular y la determinación de abundancias relativas de estas unidades dentro de la muestra examinada (Obregon-Tito et al. 2015). Actualmente la identificación por este método comprende el secuenciamiento masivo y en paralelo de fragmentos de PCR del gen 16S de una o varias muestras de comunidades bacterianas, permitiendo dar cobertura total a las unidades taxonómicas que allí existan sin importar su grado de abundancia. Este método, a diferencia del anterior, no requiere pasos de clonación o crecimiento de bacterias recombinantes, sino el uso directo de la PCR en el

secuenciamiento, el cual incluye métodos internos de separación de fragmentos clonales y su secuenciamiento en paralelo para cada una de las especies o unidades taxonómicas que se encuentren representadas en la PCR por su gen 16S.

La capacidad de generación de lecturas totales para todas las secuencias, es del orden de los 5 millones (IonTorrent PGM) hasta más de 100 millones (Illumina), lo que garantiza que todas los genes 16S de todas las bacterias, por pequeña que sea su proporción, estarán secuenciadas al final. Los fragmentos originados van de 150 a 800 pb, dependiendo del secuenciador, siendo más común en la actualidad el uso de Illumina, que genera fragmentos del primer tamaño mencionado. En ese caso, para Illumina, las regiones que se amplifican para secuenciar son la v2 o la v4-v5 del gen 16S, en la mayoría de los casos (Contreras et al. 2010).

En este tipo de sistemas moleculares basados en NGS, la bioinformática juega un papel fundamental en la identificación de las unidades taxonómicas, siendo el software más común Mothur (Schloss et al. 2009), el cual utiliza bases de datos de referencia donde previamente han sido depositadas secuencias de 16S de especies plenamente identificadas. Algoritmos de clusterización basados en distancia genética y similitud de secuencias, son la base de la identificación. Más de 3500 especies o Unidades Taxonómicas Operativas (OTUs) pueden ser identificadas (Contreras et al. 2010; Yatsunenکو et al. 2012), dependiendo de la cobertura de la corrida de secuenciamiento y de la riqueza de la muestra.

Así, no sólo la identificación taxonómica es posible por este programa a través de la presencia o ausencia, también la abundancia de cada unidad taxonómica, índices de diversidad filogenética y diversidad por riqueza de OTUs, además de otros índices de importancia para el estudio de diversidad en comunidades bacterianas en una o varias muestras. La capacidad de ver no solo datos cualitativos sino también cuantitativos, permite multiplicidad de comparaciones y análisis estadísticos posteriores como componentes principales, discriminantes, t-Student, entre otros, los cuales buscan diferenciar comunidades o microbiotas analizadas.

Dentro de las principales aplicaciones de la metagenómica de 16S rRNA, el estudio de composición de microbiotas ha tenido grande relevancia en áreas como las microbiotas intestinales (Yatsunenکو et al. 2012; Clemente et al. 2015; Obregon-Tito et al. 2015), bucales (Contreras et al. 2010), de piel (Clemente et al. 2015), suelo (Forsberg et al. 2012), entre otros, los cuales han permitido aportar nueva información, antes desconocida para la ciencia, en relación a la composición taxonómica de comunidades bacterianas ancestrales vivientes y no vivientes, de restos fósiles del ambiente y de museo, de personas con patologías y sin patologías o con hábitos alimenticios y estilos de vida contrarios.

Este método permite obtener resultados en un tiempo menor a un mes, con costos que pueden oscilar entre los 2000 a 6000 dólares por corrida de secuenciamiento, dependiendo el número de muestras analizadas y el estado de preparación de las mismas.

Finalmente, es posible concluir que los métodos moleculares de identificación en la era genómica incrementarán ostensiblemente su uso en microbiología en las siguientes décadas, principalmente permitiendo expandir las capacidades de los laboratorios para tipificar de forma exacta y rápida los microorganismos nuevos, no cultivables o fenotípicamente indistinguibles. Para ello, es necesario estar preparados y entender su filosofía, pues la metagenómica revolucionará los procesos de investigación y desarrollo de la microbiología moderna, de la cual usted hace parte.

Bibliografía

- Clarridge J (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 17:840–862.
- Clemente JC, Pehrsson EC, Blaser MJ, et al (2015) The microbiome of uncontacted Amerindians. *Sci Adv* 1:e1500183.
- Contreras M, Costello EK, Hidalgo G, et al (2010) The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology* 156:3282–3287. doi: 10.1099/mic.0.043174-0
- Forsberg KJ, Reyes A, Wang B, et al (2012) The shared antibiotic resistome of soil bacteria and human pathogens. *Science* (80-) 337:1107–11. doi: 10.1126/science.1220761
- Hall L, Doerr KA, Wohlfel SL, Roberts GD: Evaluation of the MicroSeq system for identification of Mycobacteria by 16S ribosomal DNA sequencing and its integration into a routine clinical mycobacteriology laboratory. *J Clin Microbiol* 2003, 41:1447–1453
- Obregon-Tito AJ, Tito RY, Metcalf J, et al (2015) Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun* 6:1–9. doi: 10.1038/ncomms7505
- Schloss PD, Westcott SL, Ryabin T, et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–41. doi: 10.1128/AEM.01541-09
- Yatsunencko T, Rey FE, Manary MJ, et al (2012) Human gut microbiome viewed across age and geography. *Nature* 486:222–7. doi: 10.1038/nature11053
- Zhang WW, Versalovic J (2007) Expanding the Diagnostic Capabilities of Molecular Microbiology by Genomic Methods. *J Mol Diagnostic* 9:572–573. doi: 10.2353/jmoldx.2007.070109