

# COMPARACIÓN DE DOS HERRAMIENTAS BIOINFORMÁTICAS PARA ANÁLISIS DE AMPLICONES 16S EN RUMEN DE VACUNO LECHERO

López-García, A.<sup>1</sup>; García-Rodríguez, A.; Pineda-Quiroga, C.; Atxaerandio, R.; González-Recio, Ó.

<sup>1</sup>INIA, Ctra. La Coruña km 7.5, 28040, Madrid, España; adrian.lopez@inia.es

## INTRODUCCIÓN

Actualmente la necesidad de incrementar la productividad ganadera disminuyendo su impacto ambiental ha supuesto el planteamiento de nuevas alternativas de mejora de caracteres fenotípicos, complementarios a los tradicionales. En este contexto, la metagenómica ha resultado ser una línea de investigación alternativa, dado que la composición del sistema microbiano digestivo y su expresión génica han resultado tener una fuerte relación con muchos caracteres fenotípicos de vital importancia en ganadería. En vacuno de leche, el microbioma ruminal está muy relacionado con caracteres como las emisiones de metano o la eficiencia alimentaria (Zilber-Rosenberg & Rosenberg, 2008).

Con la aparición de nuevas técnicas de secuenciación (NGS) la obtención de información genética de poblaciones de microorganismos, tanto metagenomas completos como amplicones de rRNA, se ha vuelto más rápida, precisa y económicamente rentable. Actualmente la secuenciación de amplicones de rRNA 16 y 18S es la opción preferida para analizar los sistemas microbianos en los animales, dado su menor coste y aceptable precisión.

En última instancia son las herramientas bioinformáticas las que van a determinar la fiabilidad de la identificación de los genes microbianos. Por ello los flujos de trabajo para el procesamiento y análisis de secuencias deben ser seleccionados y adaptados cuidadosamente. Algunos autores han contrastado la efectividad de varias herramientas de análisis de secuencias 16S (Lozupone *et al.*, 2005; Nilakanta *et al.*, 2014; Oulas *et al.*, 2015), y de entre todas ellas, Mothur y Qiime son de las más utilizadas. Aunque ambas han sido revisadas y se ha probado su efectividad, las comparaciones entre ellas usando datos reales son escasas, y no existen estudios previos que analicen su rendimiento en secuencias de microbiota ruminal vacuno.

Por ello se ha planteado un estudio para determinar si la elección del software o de la base de datos de asignación taxonómica para el procesamiento de secuencias de rRNA 16S, tienen influencia sobre los resultados obtenidos de composición del microbioma ruminal en vacuno lechero.

## MATERIAL Y MÉTODOS

Se recogieron muestras de líquido ruminal de 18 vacas lecheras de raza Holstein (10) y Brown Swiss (8) procedentes de una granja de la Escuela Agraria de Fraisoro (Zizurkil, Gipuzkoa, España). Para la amplificación se escogió la región hipervariable V4 del gen rRNA 16S. Para las librerías se usó el kit Nextera (Illumina) y las reacciones de secuenciación fueron llevadas a cabo en plataforma MiSeq (Illumina, San Diego, CA, USA). Los resultados de la secuenciación, fueron sometidos a un filtrado de calidad a través de la herramienta Trimmomatic (v 0.36), descartando secuencias de menos de 220 pb.

Las secuencias se procesaron usando los dos softwares mencionados previamente: Qiime (v 1.9.1) (Caporaso *et al.*, 2010) y Mothur (v 1.39.5) (Kozich *et al.*, 2013), utilizando los pipelines estándar recomendados por los desarrolladores para el análisis de amplicones rRNA, ligeramente modificados. Para la asignación de taxonomía se han utilizado las versiones compatibles con cada software de la base de datos Silva (rel. 132) y GreenGenes (2013). El procesado de secuencias se puso en marcha a través de los recursos informáticos del Centro de Supercomputación de Galicia (CESGA).

Se obtuvieron archivos con el número de OTUs clusterizados y de asignación de taxa a nivel de género para cada muestra. Las tablas de taxonomía se procesaron para calcular las abundancias relativas por muestra y se realizó un test de correlación de Pearson para los taxa comunes entre ambos pipelines. Posteriormente se eliminaron taxa espurios en base a

su abundancia relativa (< 0.1%). El procesamiento de los archivos de taxonomía y el cálculo de parámetros estadísticos y poblacionales se realizó en entorno R (v 4.3.2).

## RESULTADOS Y DISCUSIÓN

Una vez obtenida la salida, se realizó una comprobación del proceso de filtrado llevado a cabo por ambos pipelines, para ver si los resultados eran comparables. Para ello se utilizó el número de lecturas por muestra obtenido después de la eliminación de quimeras. Al usar GreenGenes como base de datos, el número de lecturas por muestra en ambas herramientas fue muy similar (53796 en Mothur y 52347 en Qiime,  $p$ -value = 0.6), de igual forma que con Silva (53784 lecturas por muestra en Mothur y 56697 en Qiime,  $p$ -value = 0.32), revelando que el filtrado de secuencias en ambos pipelines era equivalente.

En cuanto al número de taxa asignados a nivel de género, al usar GreenGenes ambas herramientas detectaron un total de 40 taxa, 23 de ellos a nivel de género, de los cuales *Prevotella*, *Butyrivibrio*, *Succiniclasicum* y *Methanobrevibacter* fueron los más representativos (RA > 1%). Adicionalmente Mothur detectó otros 13 taxa y Qiime, 5. Con Silva como base de datos el número de taxa detectados aumentó considerablemente: ambos pipelines asignaron 63 taxa, 52 de ellos a nivel de género, y, de nuevo, Mothur detectó 13 taxa adicionales y Qiime 14. Los géneros más abundantes (RA > 0.5%) comunes a ambos pipelines al usar la base de datos Silva se representan en la Tabla 1. En cuanto a las diferencias a nivel estadístico, con GreenGenes el número de taxa por muestra identificado por Mothur fue significativamente superior respecto a Qiime ( $p$ -value = 7.1e-12), pero con Silva estas diferencias no se observaron ( $p$ -value = 0.28).

Al comparar las abundancias relativas a nivel de género se observó un elevado grado de correlación positiva ( $r = 0.996$ ,  $p$ -value < 2.2e-16) al usar tanto Silva como GreenGenes. Sin embargo, al considerar solamente los géneros con RA < 5%, esta correlación disminuyó considerablemente en la comparación de pipelines con GreenGenes ( $r = 0.750$ ,  $p$ -value < 2.2e-16), mientras que con Silva (Figura 1) el descenso fue más leve ( $r = 0.976$ ,  $p$ -value < 2.2e-16). Esto indica que GreenGenes produce diferencias en la asignación de taxa menos abundantes, pero usando Silva con ambas herramientas la asignación es más parecida.

Las diferencias observadas en número de taxa ya se han reportado previamente (Plummer & Twin, 2015), y podrían deberse básicamente al uso de la base de datos, que además se ha construido de forma diferente en cada uno de los softwares para permitir la compatibilidad, y al algoritmo de clasificación. En este sentido hay que destacar que ambos programas usan métodos de agrupación diferentes: Mothur usa un sistema clasificador Bayesiano “ingenuo” que utiliza un procedimiento de *pseudo-bootstrapping* para generar estimas de confianza (Wang *et al.*, 2007), y Qiime usa el algoritmo Usearch para encontrar el equivalente más próximo a cada secuencia en una base de datos de referencia (Edgar, 2010). Sin embargo, este estudio ha revelado que, si bien el uso de bases de datos obsoletas como GreenGenes hace patentes estas diferencias en la asignación taxonómica derivadas del método de actuación del propio software, las versiones más actuales de bases de datos las minimizan.

A la vista de estos resultados podemos concluir que la utilización de una base de datos apropiada parece tener gran importancia en la obtención de resultados sobre caracterización de la estructura poblacional de microorganismos. Por otra parte, utilizando una base de datos apropiada la elección del software no parece tener influencia en la asignación de taxonomía en muestras de microbioma ruminal, con lo que la elección de la herramienta dependerá en última instancia de las preferencias del investigador.

## REFERENCIAS BIBLIOGRÁFICAS

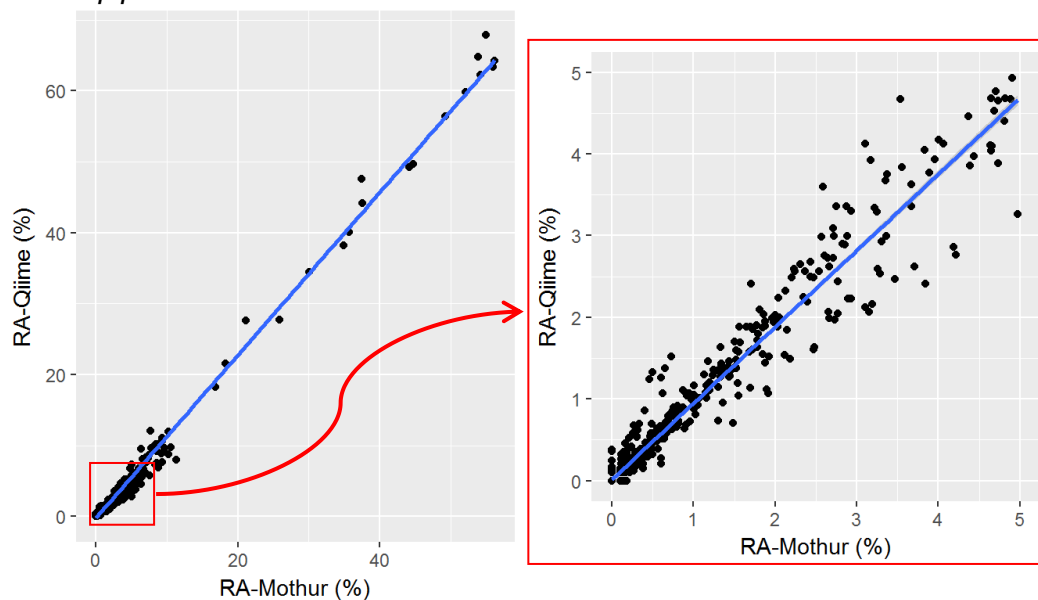
Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., et al. (2010). *Nat. Methods* 7, 335–6. • Edgar, R.C. (2010). *Bioinformatics* 26, 2460–2461. • Kozich, J. J., Westcott, S.L., Baxter, N.T., Highlander, S.K., & Schloss, P.D. (2013). *Appl. Environ. Microbiol.* 79, 5112–20. • Lozupone, C., Knight, R., Firrell, S., Foulkes, M.A., Jablonski, K.A., Collman, R., et al. (2005). *Appl. Environ. Microbiol.* 71, 8228–8235. • Nilakanta, H., Drews, K.L., Firrell, S., Foulkes, M.A., & Jablonski, K.A. (2014). *BMC Res. Notes* 7, 830. • Oulas, A., Pavloudi, C.,

Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N., Kotoulas, G., et al. (2015). *Bioinform. Biol. Insights* 9, 75–88. • Plummer, E., & Twin, J. (2015). *J. Proteomics Bioinform.* 8, 283–291. • Wang, Q., Garrity, G.M., Tiedje, J.M., & Cole, J.R. (2007). *Appl. Environ. Microbiol.* 73, 5261–5267. • Zilber-Rosenberg, I., & Rosenberg, E. (2008). *FEMS Microbiol. Rev.* 32, 723–735.

**Tabla 1:** Géneros con abundancia relativa (RA) > 0.5% detectados en ambos pipelines usando la base de datos de referencia Silva.

Taxa (genus)	Mothur	Qiime	Taxa (genus)	Mothur	Qiime
<i>Prevotella_1</i>	40,195	46,506	<i>Ruminococcaceae_UCG-014</i>	2,511	1,683
<i>Ruminococcaceae_NK4A214_group</i>	6,342	6,825	<i>Saccharofermentans</i>	1,756	1,878
<i>Butyrivibrio_2</i>	5,516	6,062	<i>Methanobrevibacter</i>	1,609	1,612
<i>Christensenellaceae_R-7_group</i>	5,726	4,625	<i>Prevotellaceae_UCG-001</i>	1,024	0,951
<i>Candidatus_Saccharimonas</i>	5,329	4,414	<i>p-1088-a5_gut_group</i>	0,957	0,934
<i>Succiniclaticum</i>	3,466	4,531	<i>Acetitomaculum</i>	0,702	0,665
<i>Rikenellaceae_RC9_gut_group</i>	4,101	3,890	<i>Lachnospiraceae_AC2044_group</i>	0,392	0,821
<i>Lachnospiraceae_XPB1014_group</i>	2,488	2,539	<i>CPla-4_termite_group</i>	0,567	0,619
<i>Lachnospiraceae_NK3A20_group</i>	2,175	2,393	<i>Schwartzia</i>	0,539	0,600

**Figura 1.** Abundancia relativa de los distintos géneros de microorganismos detectados en ambos pipelines usando la base de datos de referencia Silva.



## BIOINFORMATIC TOOL COMPARISON FOR 16S AMPLICON ANALYSIS IN DAIRY COW RUMEN MICROBIOME

**ABSTRACT:** With NGS techniques, microbiome studies have acquired a great importance in the improvement of cattle productivity and sustainability. In these studies, bioinformatics tools are determinant for the reliability of the results, so contrasts between different pipelines must be done. In this experiment a comparison between two of the most used tools, Qiime and Mothur, is carried out for 16S rRNA analysis from 18 dairy cattle rumen microbiome samples, using both GreenGenes and Silva reference databases. For each database, means for number of assigned taxa to genus level were calculated for each pipeline, as well as Pearson correlations for relative abundances between pipelines. In our results, Mothur detected a significantly higher number of taxa per sample than Qiime when using GreenGenes, whereas no differences were observed with Silva. The correlation between softwares for relative abundance among genera was very positive with both databases, although GreenGenes revealed some differences when assigning taxonomy to less abundant OTUs, meaning that the bioinformatics tool has not a relevant influence in the characterization of the core microbiome composition, while the chosen database is crucial for the results.

**Keywords:** microbiome, qiime, mothur, rumen