

# ¿Microbiomas con sobredispersión, o solo variabilidad en la dispersión?

J. Casellas<sup>1\*</sup>, C. Casto-Rebollo<sup>2</sup>, L. Azougagh<sup>2</sup>, P. Núñez<sup>2</sup>, L. Varona<sup>3</sup>, N. Ibáñez-Escriche<sup>2</sup>

<sup>1</sup>Universitat Autònoma de Barcelona, 08193, Bellaterra, Barcelona

<sup>2</sup>Universitat Politècnica de València, 46022, València

<sup>3</sup>Universidad de Zaragoza, 50018, Zaragoza.

\* Corresponding autor: joaquim.casellas@uab.cat

## Resumen

Aunque el concepto teórico de microbioma se desarrolló durante la segunda mitad del siglo XX (*i.e.*, comunidad característica de microorganismos que ocupan un hábitat razonablemente caracterizado con propiedades físico-químicas diferenciadas), su aplicación práctica en múltiples campos de la ciencia no se generalizó hasta la introducción de las técnicas de secuenciación masiva. Dentro el ámbito concreto de la ganadería, el microbioma juega un papel principal en múltiples procesos fisiológicos, con un gran potencial tanto para impulsar la productividad de los distintos sistemas ganaderos, como para mejorar la salud animal y reducir los niveles de contaminación ambiental. Los microbiomas se caracterizan a partir de los recuentos de los distintos microorganismos presentes en cada muestra y obtenidos mediante secuenciación, lo que da lugar a abundancias absolutas que pueden, posteriormente, transformarse en abundancias relativas (*i.e.*, datos composicionales) y someterse a normalización si el procedimiento analítico así lo requiere. En cualquier caso, una asunción típica dentro del análisis de datos de microbioma es la sobredispersión de los mismos. Al tratarse de recuentos, su análisis se podría desarrollar bajo distribuciones paramétricas clásicas como la binomial, Poisson o binomial negativa, aunque en todos estos casos la variabilidad está directamente relacionada con la media y no puede ajustarse de forma independiente en cada base de datos (varianza menor, igual y superior a la media, respectivamente). Es por esto que se acostumbra a elegir el modelo binomial negativo para garantizar una varianza superior a la media, con el objetivo de capturar la sobredispersión que se asume por defecto en los datos.

En el presente trabajo se generaliza el modelo binomial (*i.e.*, regresión logística) con el objetivo de introducir un parámetro adicional de dispersión ( $\pi$ ) que permita ajustar tanto sobre como subdispersión en los datos de microbioma. Tomando  $s_i$  como el número de recuentos de un determinado microorganismo de un total de  $n_i$  recuentos totales en la muestra  $i$ , podemos definir

$$E(s_i) = n_i p_i$$

y

$$V(s_i) = \pi n_i p_i (1 - p_i),$$

dónde

$$p_i = \exp(\mathbf{x}_i \boldsymbol{\beta}) / [1 + \exp(\mathbf{x}_i \boldsymbol{\beta})]$$

y  $\mathbf{x}_i \boldsymbol{\beta}$  es el término que integra las variables independientes de la regresión. El modelo resultante se resuelve mediante una aproximación de quasi-verosimilitud, y permite estimar una varianza independiente de la media de los datos, dentro de un espacio paramétrico teórico entre 0 y más infinito. A modo de ejemplo, se analizaron microbiomas fecales a 180 días de edad de cerdos ibéricos de de las estirpes Entrepelado ( $E$ ;  $n = 67$ ) y Retinto ( $R$ ;  $n = 54$ ), así como sus cruces recíprocos  $E \times R$  ( $n = 47$ ) y  $R \times E$  ( $n = 43$ ). El modelo desarrollado se aplicó individualmente sobre 57 taxones a nivel de género mediante inferencia evidencial, detectándose valores de  $\pi$  entre 0,02 y 145,7. La relevancia estadística de los parámetros de dispersión se testó mediante un cociente de verosimilitudes (CV) entre el modelo binomial desarrollado y un modelo binomial estándar sin dispersión adicional. La evidencia estadística a favor de modelo binomial con dispersión era superior a  $CV = 32$  en los valores más extremos de  $\pi$ . Estos resultados sugieren la presencia no solo de sobredispersión sino también de subdispersión, variable dependiendo de los taxones analizados, lo cual debería ser tenido en cuenta en los análisis de microbioma.

*Keywords: binomial, microbiome, likelihood, variability*