

COMPARACIÓN DE MÉTODOS DE ANÁLISIS MICROBIANO: PLS-DA, RANDOM FOREST Y BAYES-C

Biada^{1*}, I., Tiezzi², F., Ibáñez-Escriche¹, N., García³, M.L., Argente³, M.J. y Santacreu¹, M.A.

¹Instituto de Ciencia y Tecnología Animal, Universitat Politècnica de València, València 46022, España; ²Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università degli Studi di Firenze, Firenze, Italia; ³Departamento de Tecnología Agroalimentaria, Universidad Miguel Hernández de Elche, Orihuela, España
*ibiada@posgrado.upv.es

INTRODUCCIÓN

Los datos del microbioma son de alta dimensionalidad, dispersos y composicionales (Gloor *et al.*, 2017), y su análisis requiere métodos analíticos capaces de abordar esta complejidad. Este estudio compara los métodos Random Forest (RF), Partial Least Square Discriminant Analysis (PLS-DA) y Bayes-C en la clasificación de dos líneas genéticas de conejo a partir de su microbiota intestinal. El objetivo es evaluar el rendimiento predictivo de estos tres métodos y su relevancia biológica en la identificación de taxones microbianos determinantes para la clasificación de dos líneas maternas con diferente longevidad.

MATERIAL Y MÉTODOS

El análisis microbiano se hizo a partir de muestras de heces blandas de 102 conejas nulíparas de dos líneas maternas (A y LP), 51 muestras por línea. La línea LP, fue fundada utilizando criterios de longevidad y productividad, mientras que la línea A es una línea materna comercial estándar (Sánchez *et al.*, 2008). Se compararon tres modelos para predecir la línea genética a partir del microbioma: Bayes-C, regresión bayesiana que asume que solo algunos taxones influyen en la clasificación; Random Forest, que construye árboles de decisión; y PLS-DA, que maximiza la covarianza entre la microbiota y la línea genética. Se siguió el procedimiento propuesto por Topçuoğlu *et al.* (2020) para la clasificación basada en el microbioma. Además, la selección de variables se realizó de manera iterativa, priorizando los taxones microbianos más relevantes en cada modelo.

RESULTADOS Y DISCUSIÓN

El método PLS-DA mostró la mayor precisión predictiva y menor variabilidad ($AUC = 0,95 \pm 0,03$), seguido de Bayes-C ($AUC = 0,92 \pm 0,05$) y Random Forest ($AUC = 0,87 \pm 0,06$). El método PLS-DA seleccionó más variables (86), mientras que BayesC identificó 24 unidades taxonómicas y Random Forest 10, aunque con un solapamiento de taxones alto entre los modelos, destacando taxones compartidos como Eubacteriaceae y Lachnospiraceae. El método PLS-DA optimizó la precisión y Random Forest priorizó taxones biológicamente relevantes.

CONCLUSIÓN

Los métodos PLS-DA y Bayes-C lograron mayor precisión predictiva al capturar más características, mientras que Random Forest seleccionó menos taxones, pero con mayor relevancia biológica. El método PLS-DA ofreció una visión amplia de diferencias de microbioma entre líneas, mientras que Random Forest identificó taxones clave, siendo más eficiente para la selección dirigida de variables.

REFERENCIAS BIBLIOGRÁFICAS

- Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J. 2017. *Frontiers in Microbiology* 8: 2224
 - Sánchez, J.P., Theilgaard, P., Mínguez, C., Baselga, M. 2008. *Journal of Animal Science* 86: 515-525.
 - Topçuoğlu, B.D., Lesniak, N.A., Ruffin, M.T., Wiens, J., Schloss, P.D. 2020. *mBio* 11: e00434-20.
- Agradecimientos:** Agradecemos a la Unión Europea y Generalitat Valenciana implicadas en AGROAL-NEXT/2022/037;PRTR-C17.I1), a la Generalitat Valenciana por la beca CIACIF/2021/005, y a Tamara Romero Rueda y Rafael Torres Lozano por su apoyo en la recolección de muestras y manejo de animales.