Can we predict feed efficiency using fecal microbiota information?

Le Sciellour Mathilde[1]*, S. Dejean[2], D. Renaudeau[1], O. Zemb[3]

[1]PEGASE, INRA, Agrocampus-Ouest, 35042, RENNES, France

[2]Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS, UPS, Toulouse Cedex 9, France, F-31062

[3]GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, 31320, Castanet Tolosan, France

*Corresponding author: mathilde.lesciellour@inra.fr

Abstract

The present study aimed at predicting feed efficiency (FE) based on fecal microbiota, using partial least square regression (PLSR), sparse PLSR, and random forest regression (RF). Fecal samples from 147 Pietrain x (Large White x Landrace) pigs reared in two consecutive batches were collected at 99 days of age. Daily live body weight and feed intake were individually measured in pigs fed ad libitum with a corn soybean diet. The relative abundances of operational taxonomic units (OTU) resulting from fecal 16S rRNA sequencing were used to build the prediction models of FE between 99 and 113 days. From these data, neither PLSR nor RF models have been validated on external datasets. An important over-fitting has been observed in PLSR. With this aim to test the ability of the methods to retrieve information, synthetic OTU were created to fit an artificial Pearson correlation with FE (r^2 = 0 to 0.9) and were added among the predictors in the dataset. Artificial OTU correlated above 0.37 with FE improved the prediction in sparse PLSR and RF, and reduced the over-fitting. The best predictions were achieved by sparse PLSR. The present study emphasized the ability of sparse PLSR and RF to build valid prediction models of a quantitative phenotype, based on fecal microbiota composition. Since no OTU was correlated above 0.30 with FE in the real dataset, the power of the prediction methods was not enough to extract useful information from the fecal microbiota. The functional redundancy of the microbiota could explain the lack of relevant information in the real dataset to predict pigs' quantitative phenotype. These results suggest that the best strategy is to run sparse PLSR only if a correlation higher than 0.37 is observed.

This study is part of the Feed-a-Gene Project funded from the European Union's H2020 Program (grant 633531).