

Uncovering patterns in skin and gut microbiota of rainbow trout (*Oncorhynchus mykiss*): insights from machine learning and feeding trials with sustainable aquafeeds based on yellow mealworm (*Tenebrio molitor*)

Aquaculture International

Silvio Rizzi¹, Giulio Saroglia², Violeta Kalemi¹, Simona Rimoldi¹, Genciana Terova^{1*}

¹ Department of Biotechnology and Life Sciences, University of Insubria, Via Jean Henry Dunant 3, 21100 Varese, Italy

² Medical Devices Area, Institute of Digital Technologies for Personalized Healthcare (MeDiTech), University of Applied Sciences and Arts of Southern Switzerland, Via la Santa 1, 6962 Lugano, Switzerland

* Corresponding author: genciana.terova@uninsubria.it

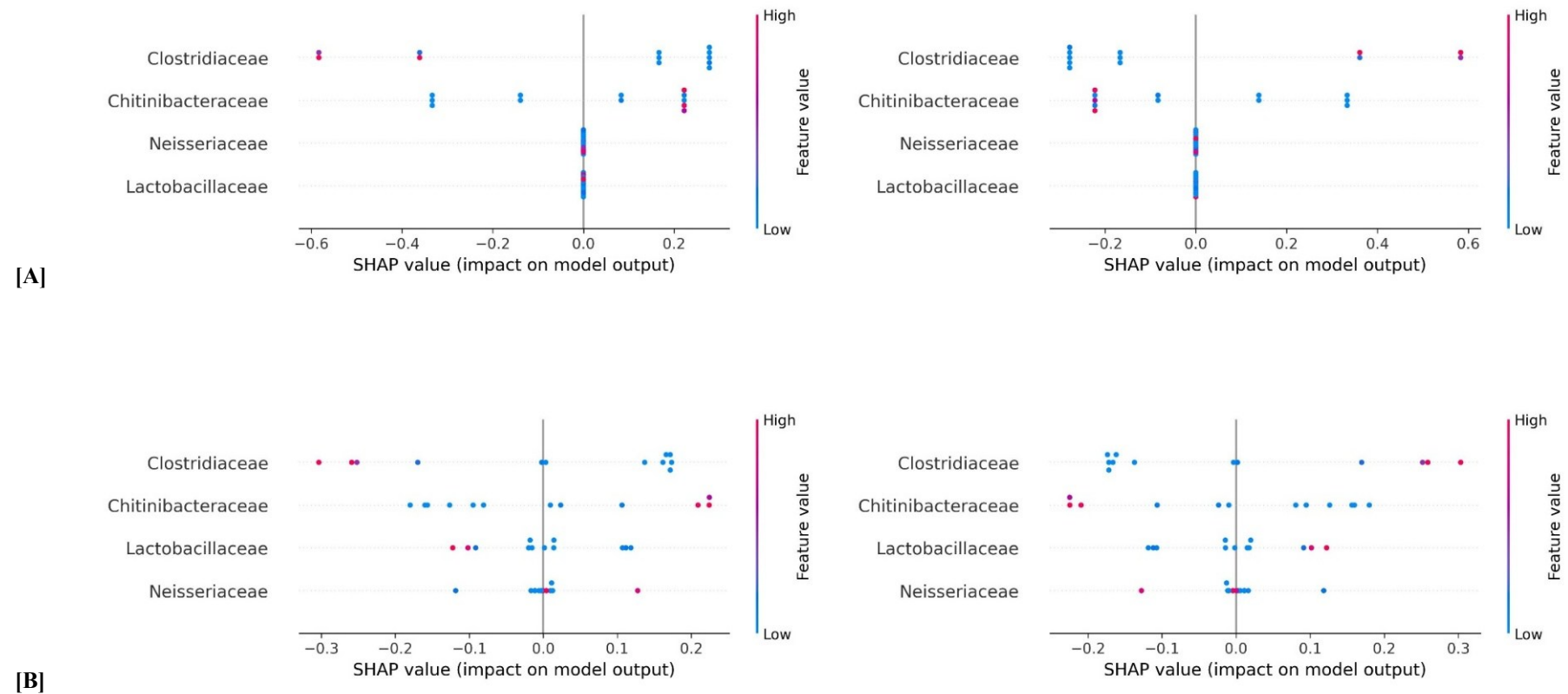
Understanding beeswarm plots in SHAP: a guide to interpretation

SHAP beeswarm plots provide a global visualization of feature influence on model prediction at an individual-sample level. On the y -axis, features are ranked in descending order of importance, with the most influential ones at the top and the least influential ones at the bottom. On the x -axis, each point represents an individual sample, encoding two key pieces of information: *(i)* color, representing the feature value; and *(ii)* position, corresponding to the SHAP value.

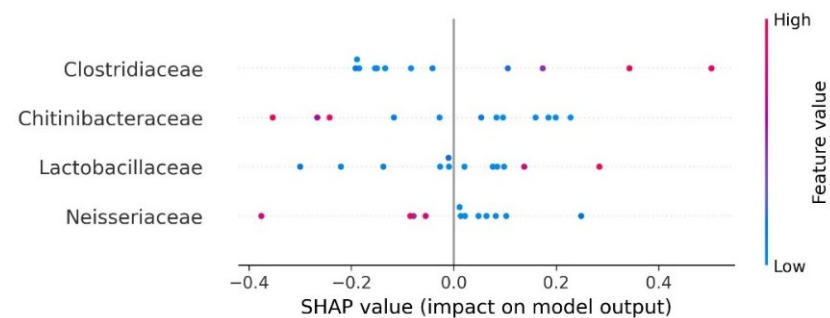
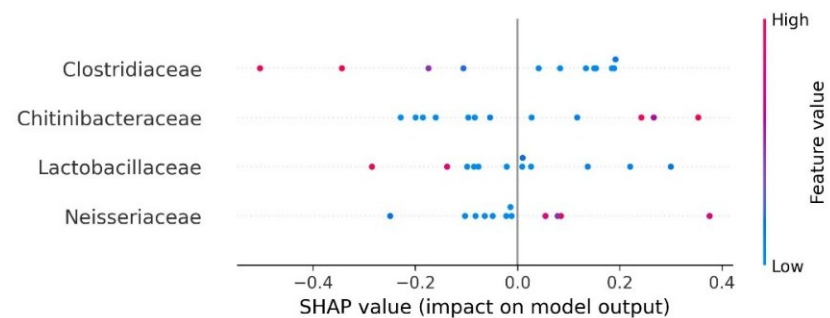
In binary classification, clusters of same-color points predominantly appearing at positive SHAP values suggest a strong contribution toward the class being examined, whereas clusters at negative SHAP values indicate influence toward the opposite class. While binary classification beeswarm plots often exhibit visual symmetry, this pattern depends on dataset balance and feature distribution, rather than being a strict rule.

Supplementary Figure 1

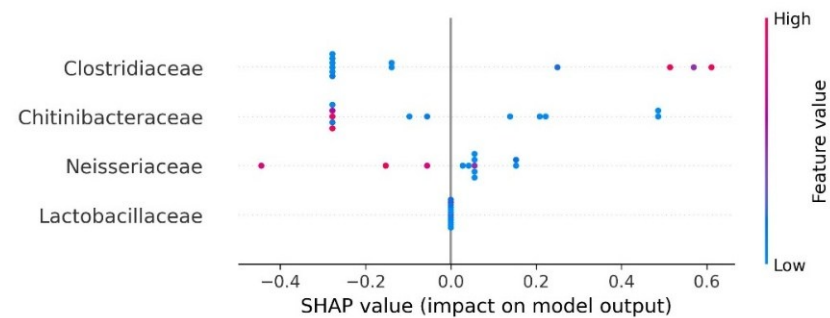
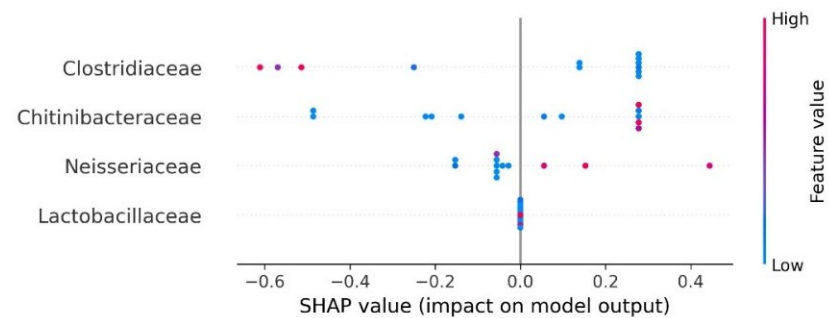
SHAP beeswarm plots of classification models for rainbow trout gut microbiota at the family level.



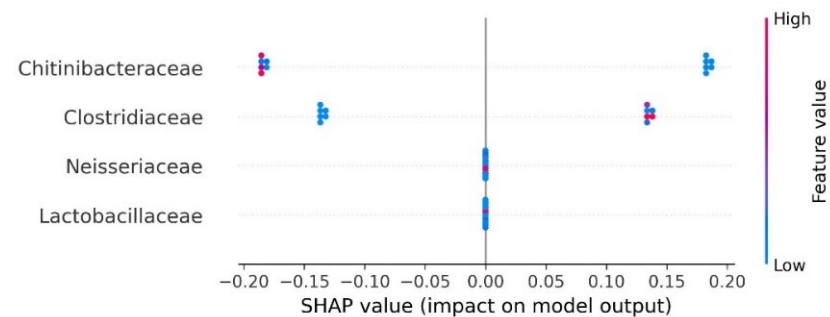
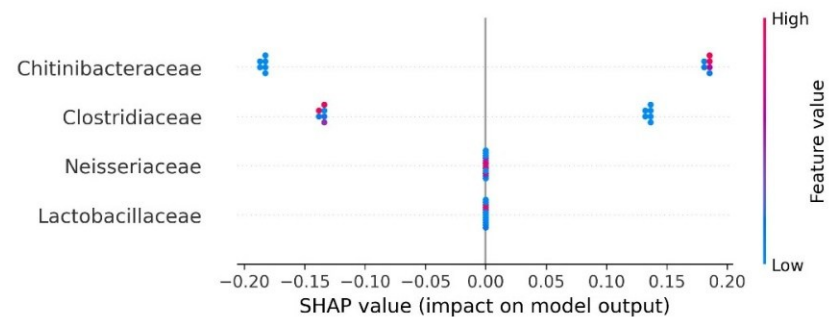
[C]



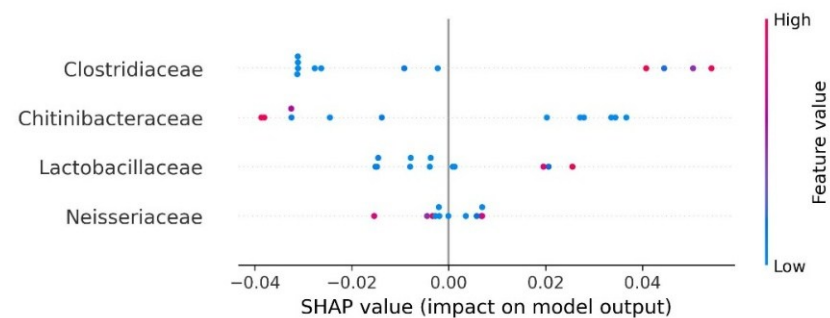
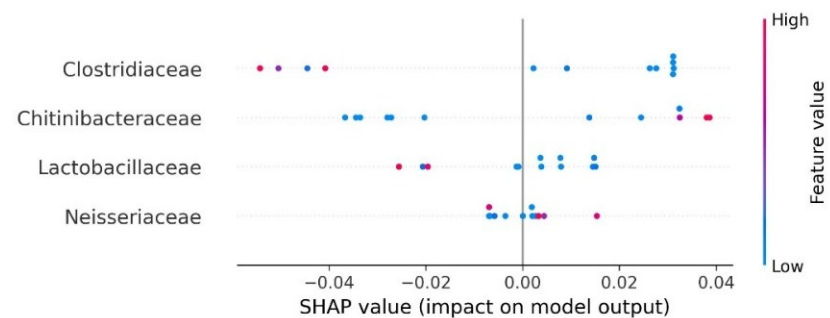
[D]



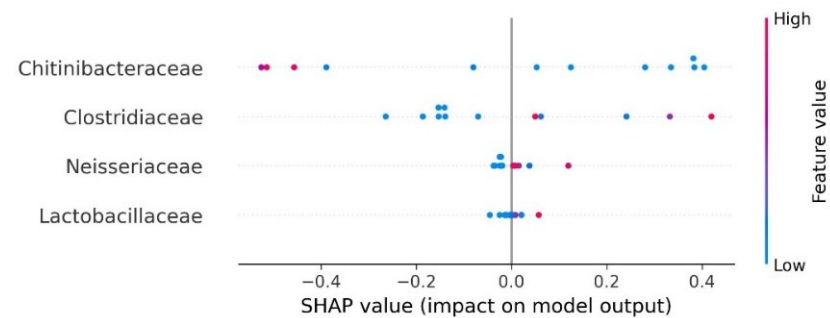
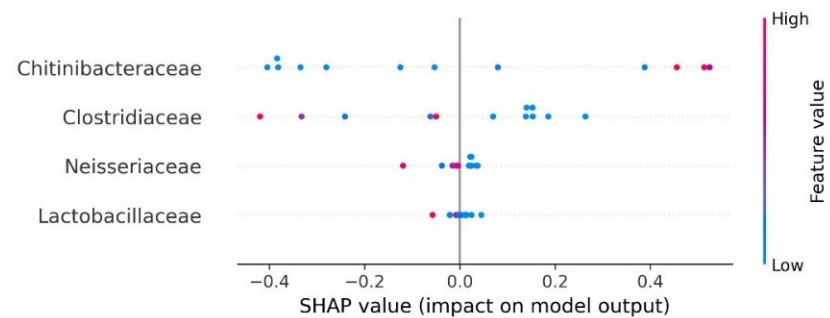
[E]



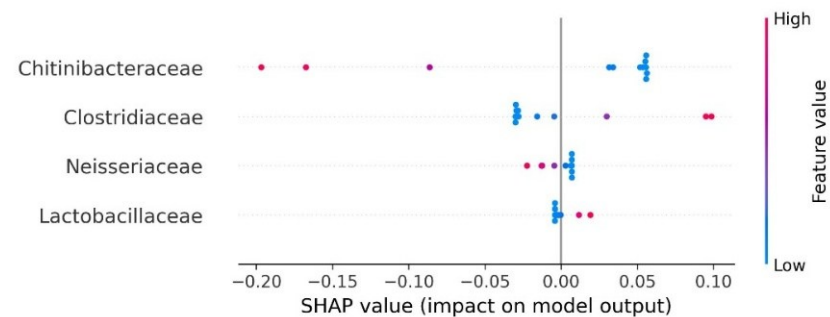
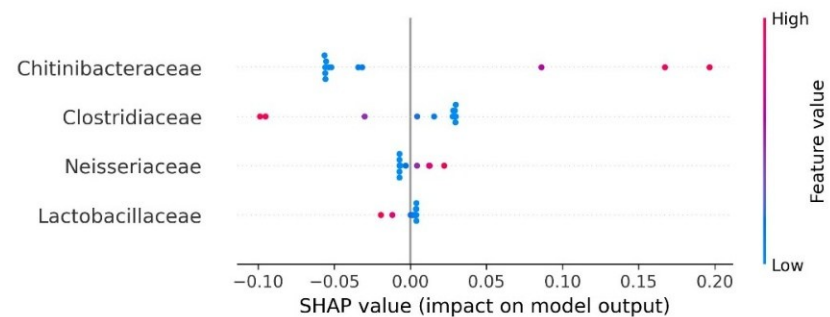
[F]



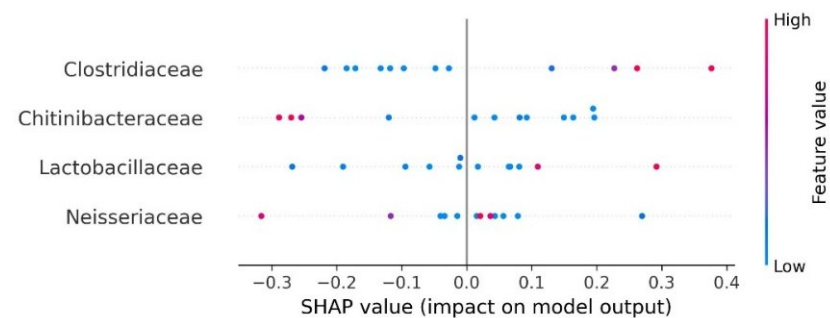
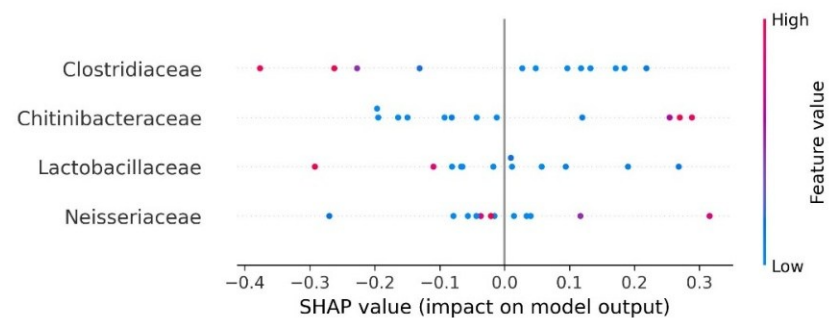
[G]



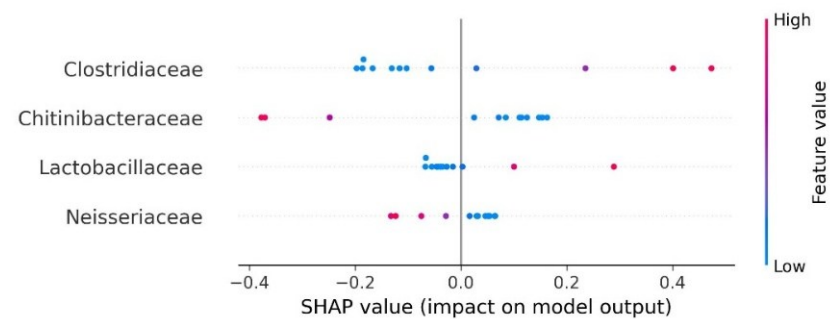
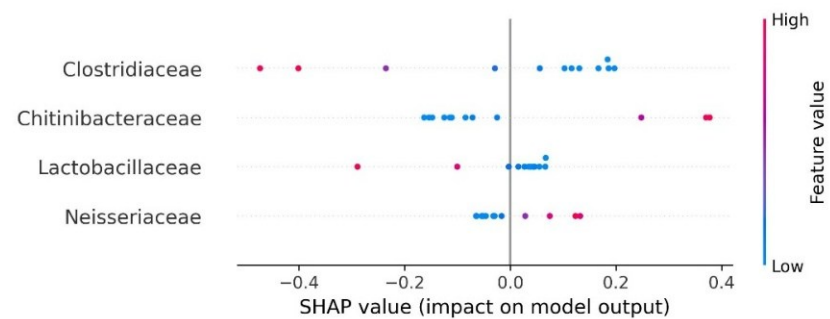
[H]



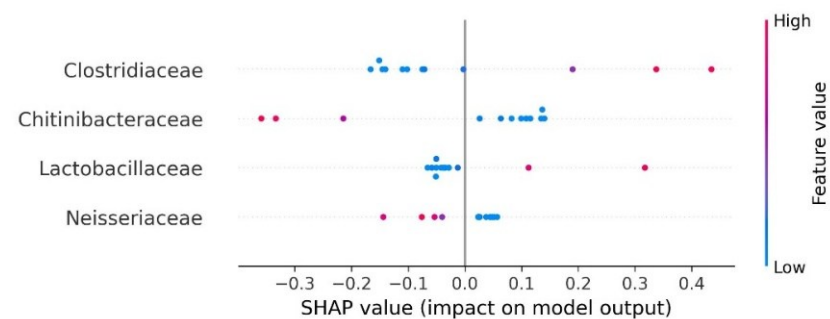
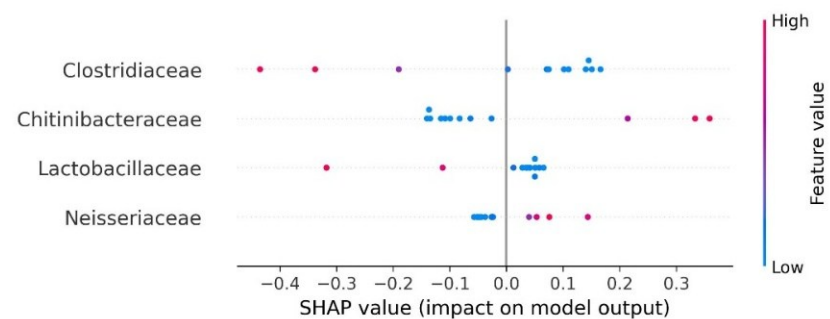
[I]



[L]



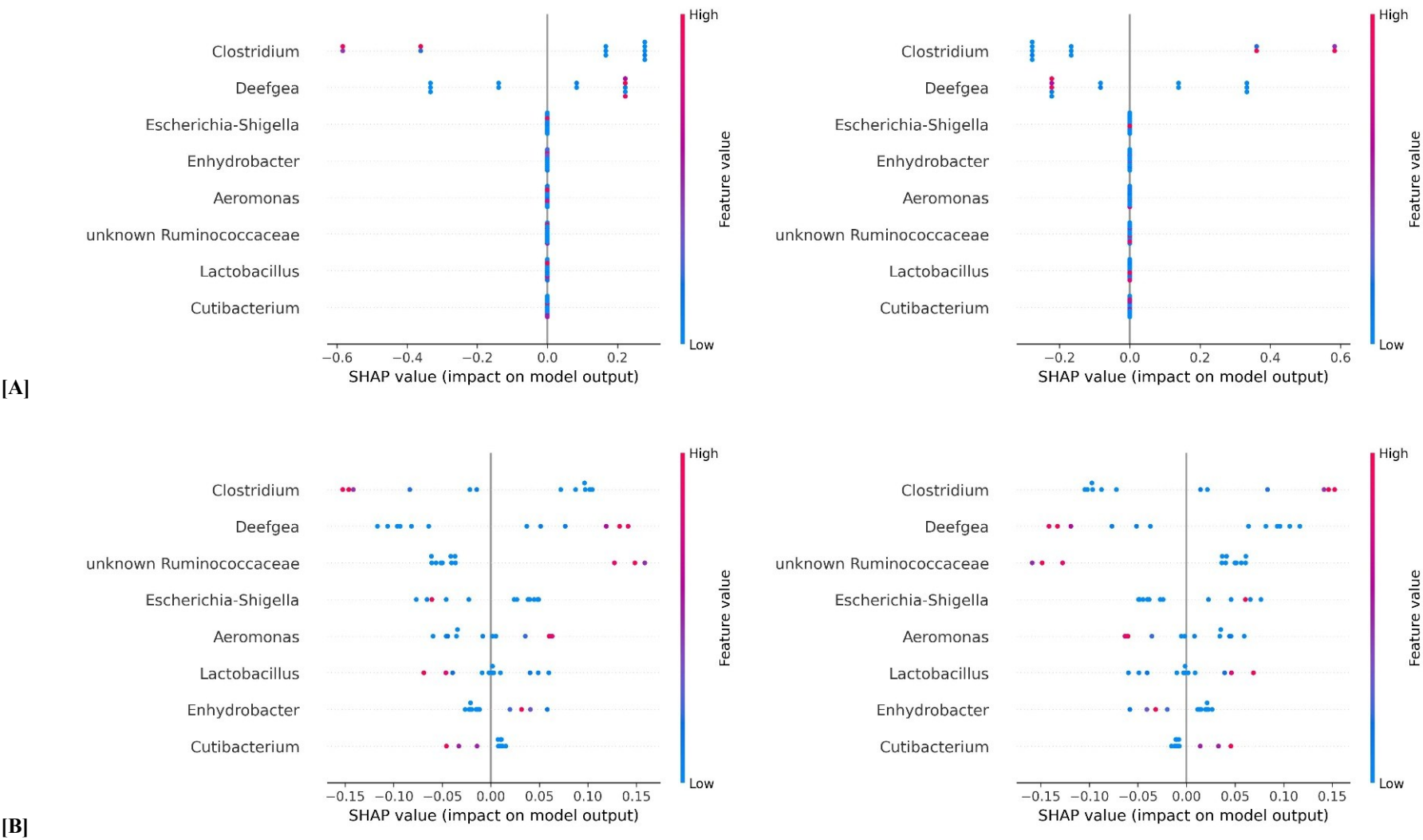
[M]



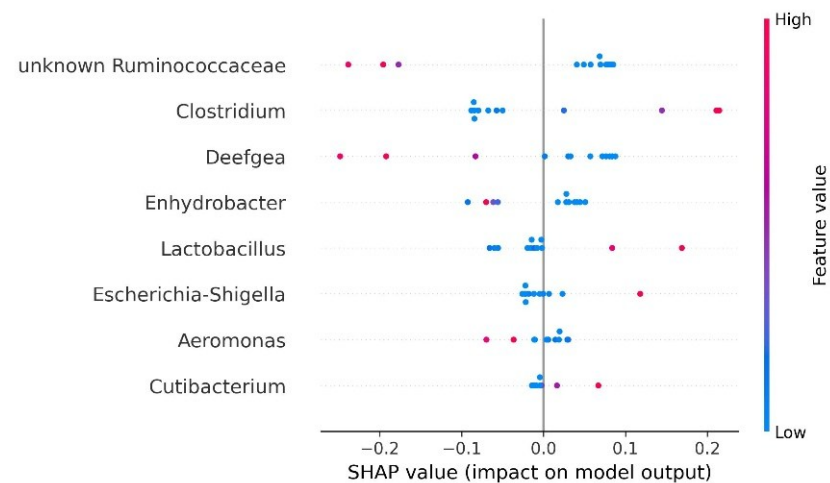
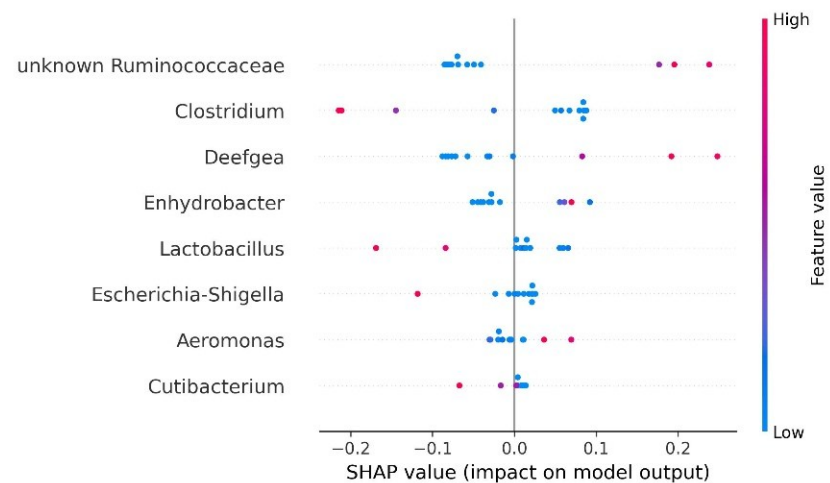
On the left, SHAP beeswarm plots referred to the TM0 diet. On the right, SHAP beeswarm plots referred to the TM100 diet. (**A**: decision tree classifier. **B**: random forest classifier. **C**: extra tree classifier. **D**: gradient boosting classifier. **E**: extreme gradient boosting classifier. **F**: categorical boosting classifier. **G**: multinomial Naïve Bayes. **H**: support vector machine classifier with linear kernel. **I**: k -nearest neighbor classifier. **L**: logistic regression. **M**: multilayer perceptron classifier.

Supplementary Figure 2

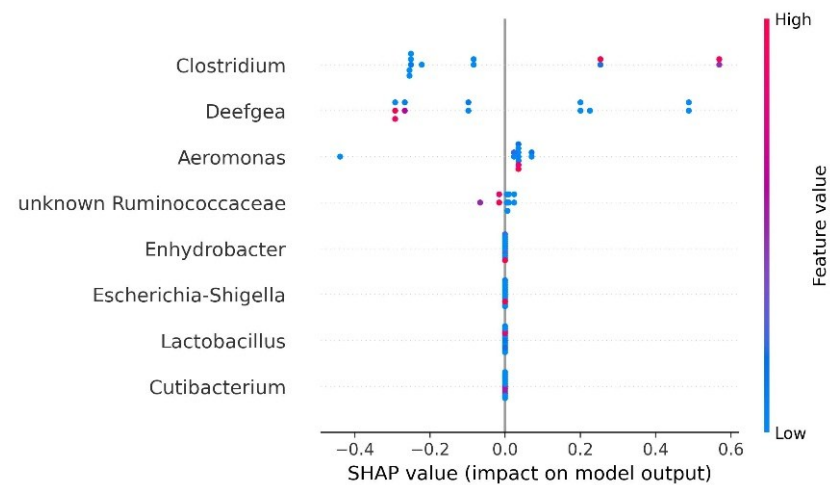
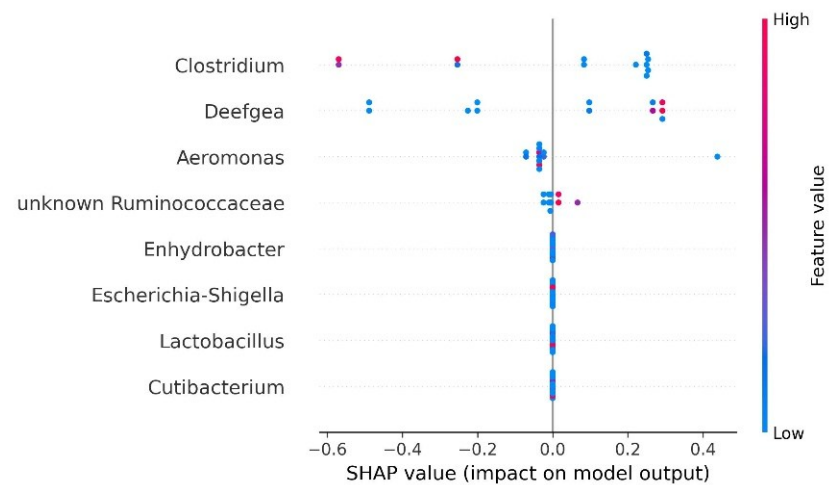
SHAP beeswarm plots of classification models for rainbow trout gut microbiota at the genus level.



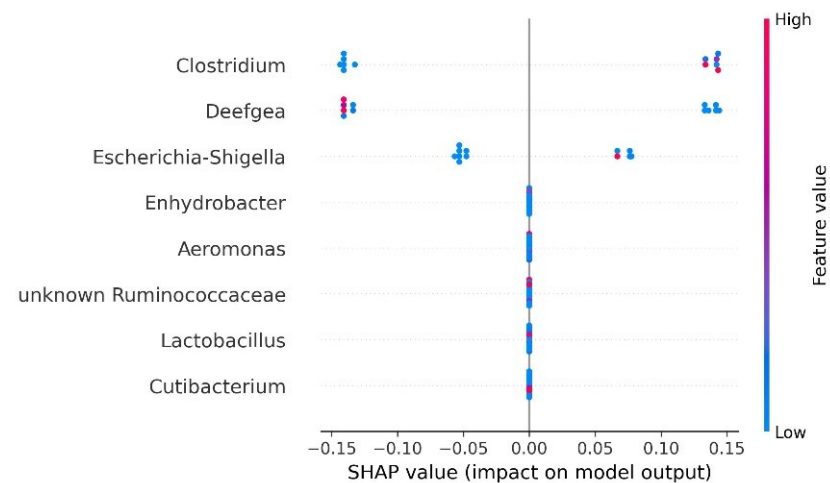
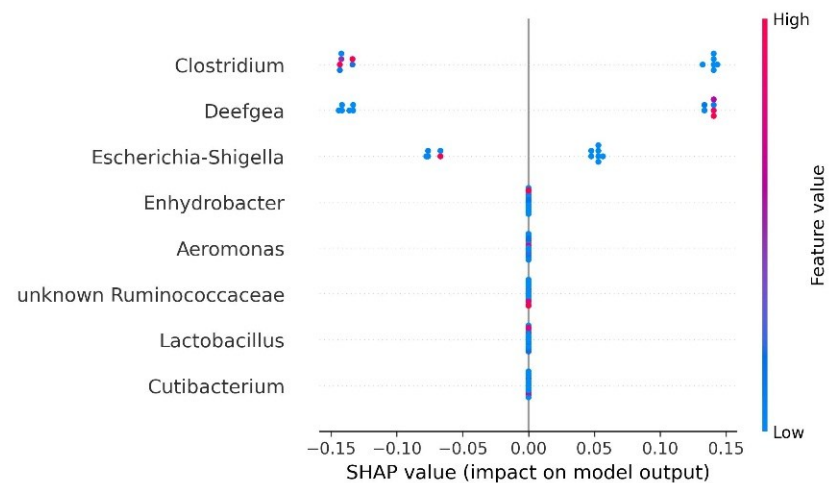
[C]



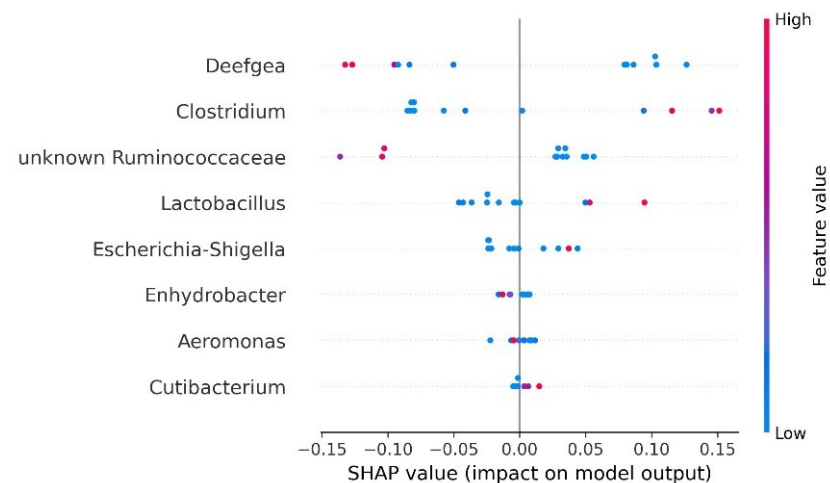
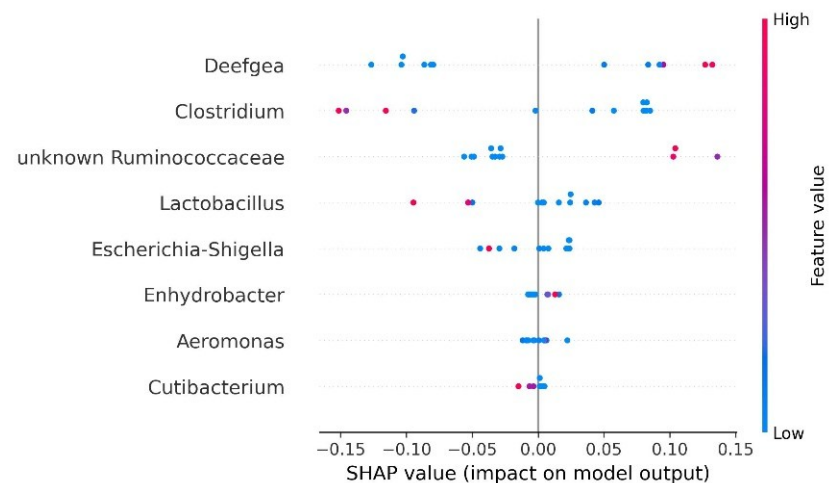
[D]



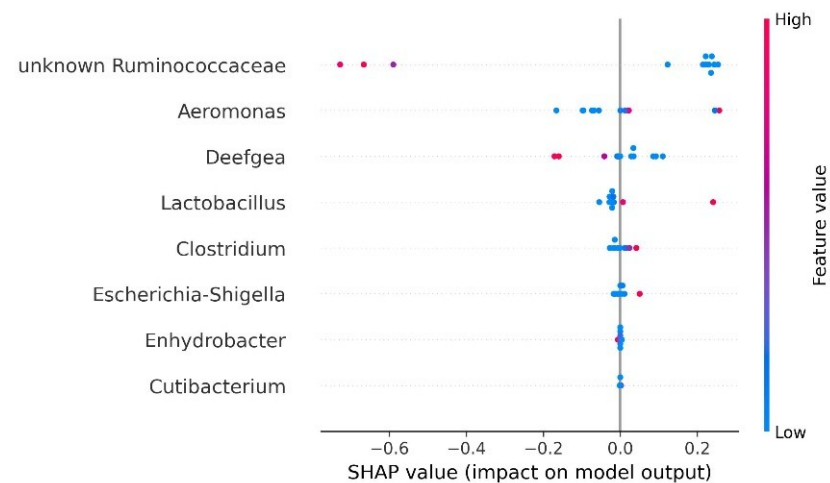
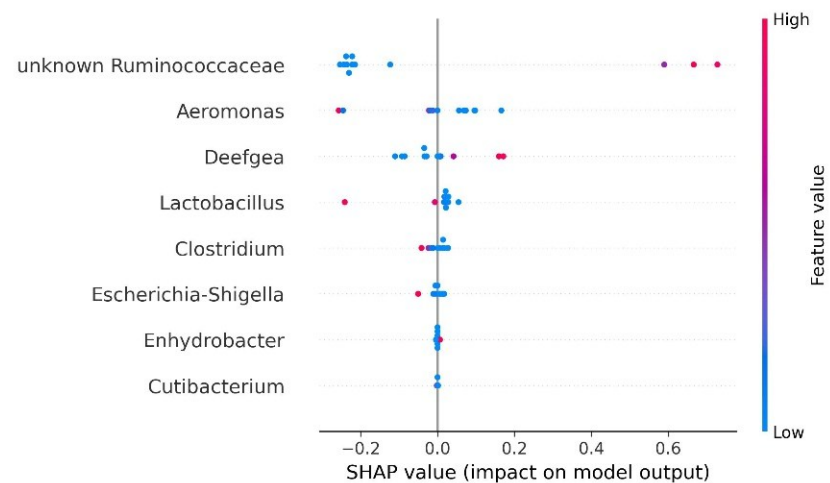
[E]



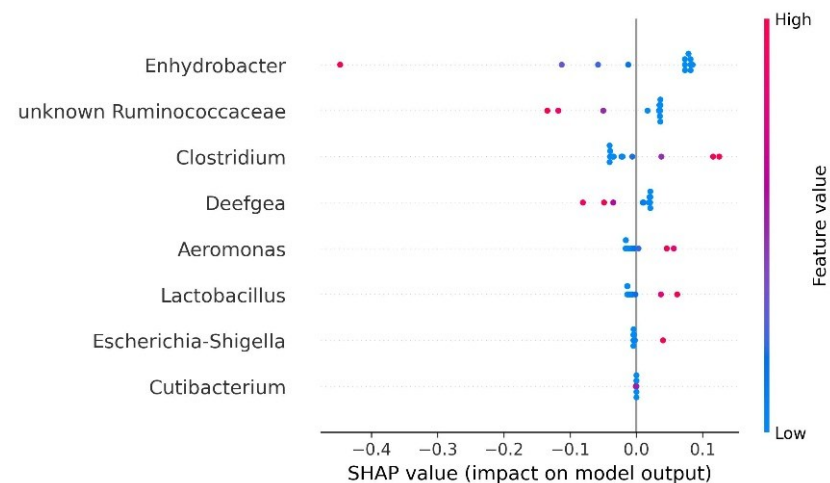
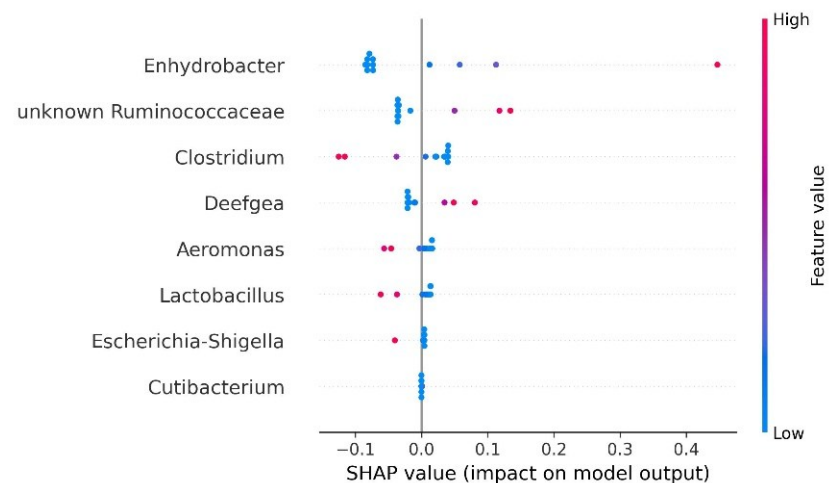
[F]



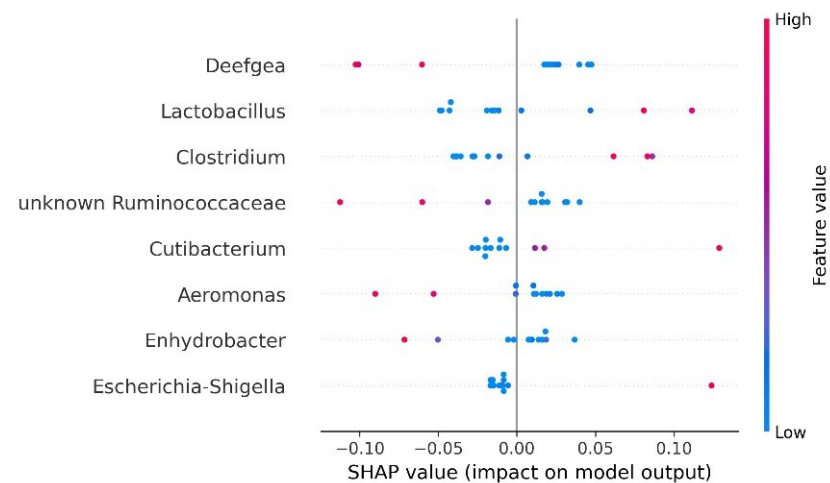
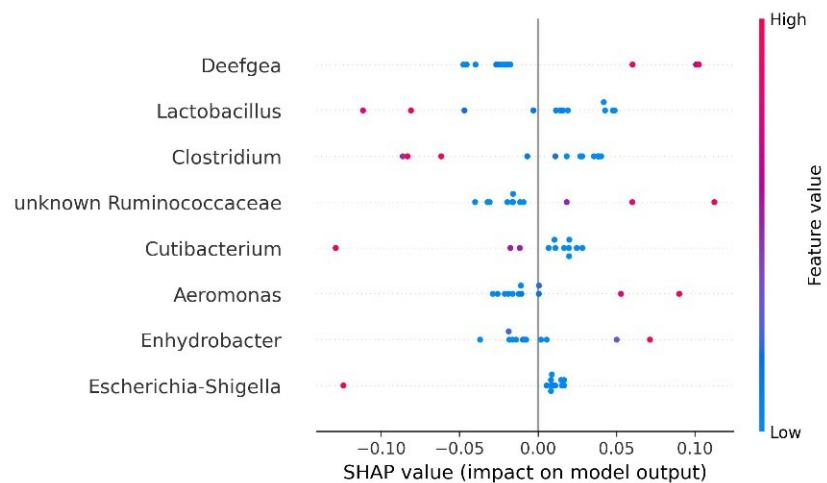
[G]



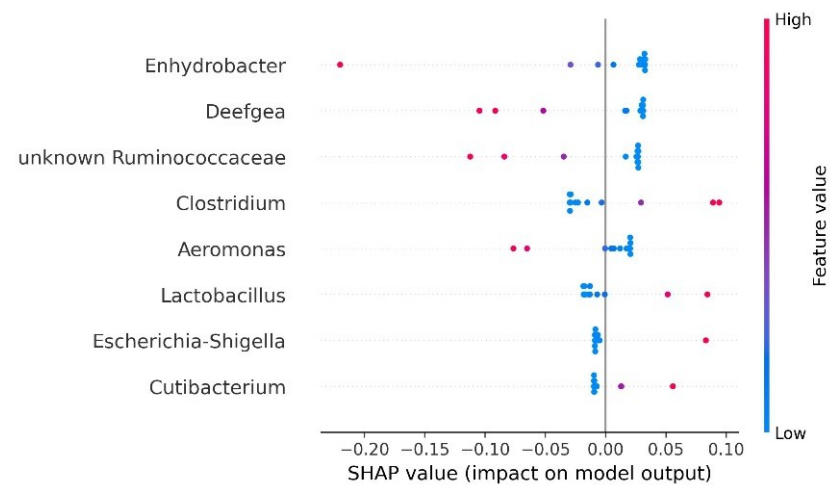
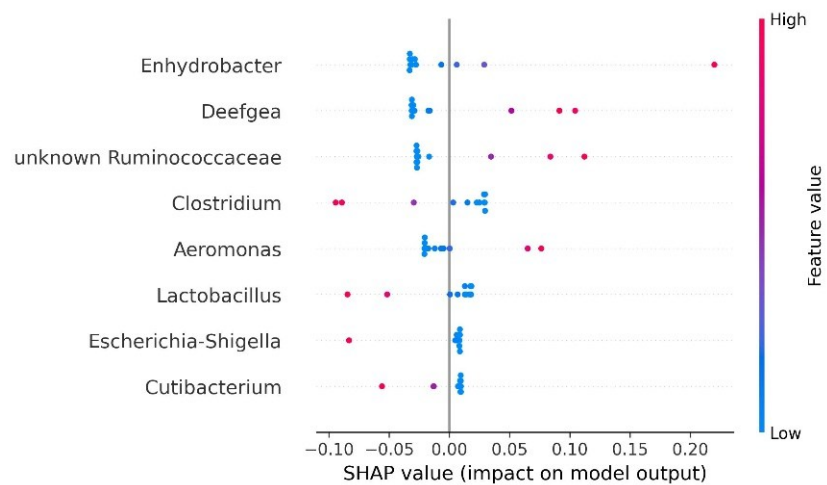
[H]



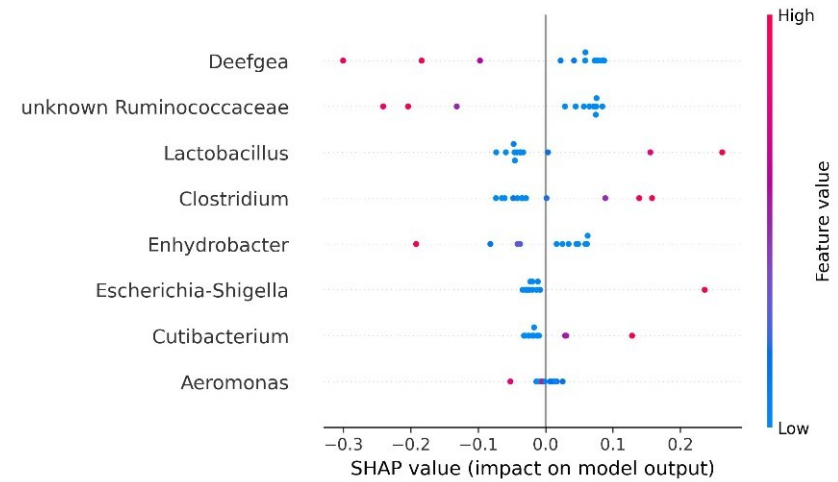
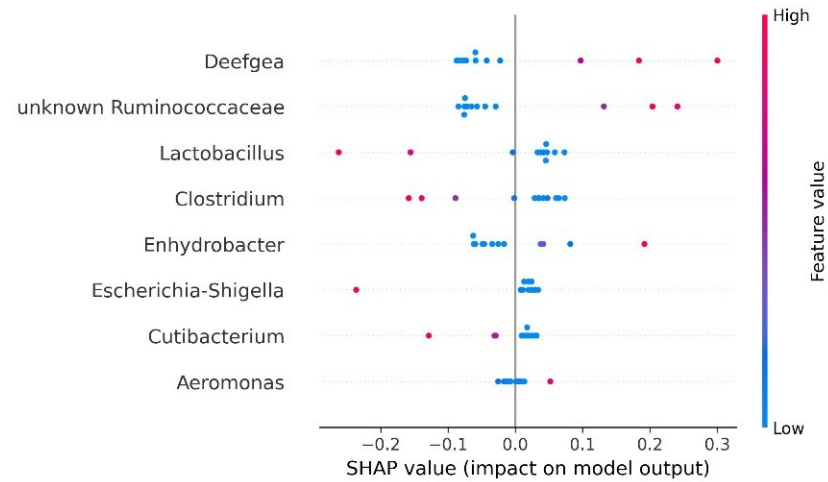
[I]



[L]



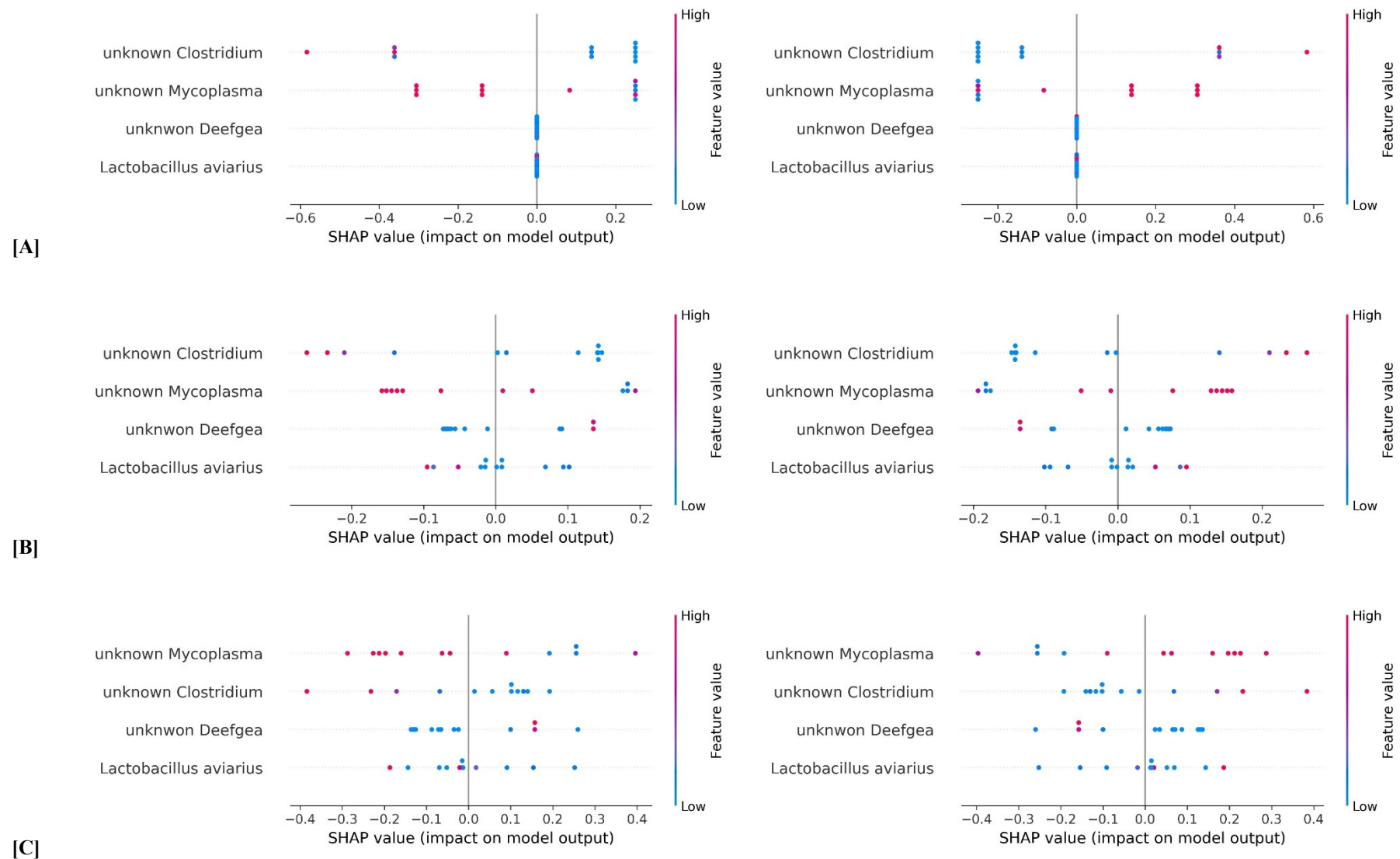
[M]



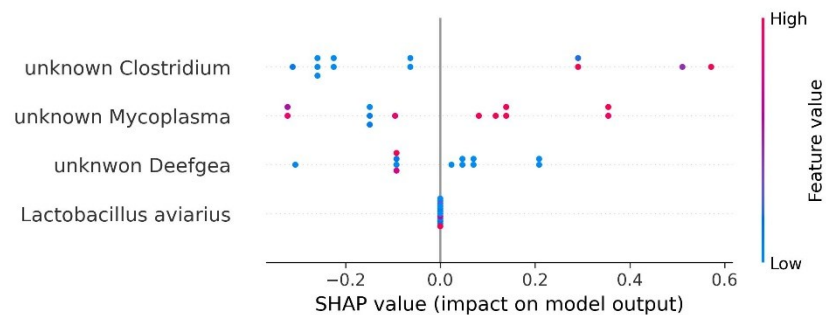
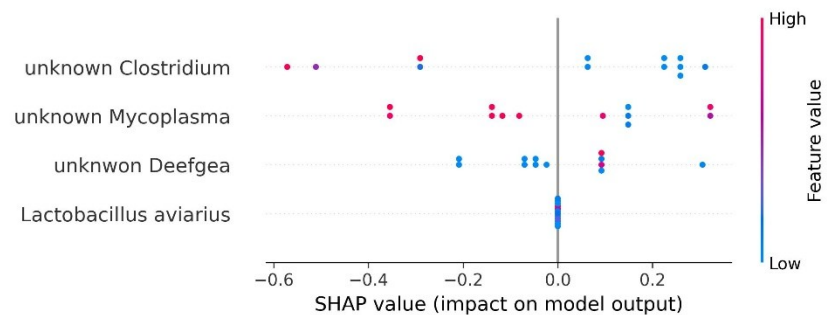
On the left, SHAP beeswarm plots referred to the TM0 diet. On the right, SHAP beeswarm plots referred to the TM100 diet. (**A**: decision tree classifier. **B**: random forest classifier. **C**: extra tree classifier. **D**: gradient boosting classifier. **E**: extreme gradient boosting classifier. **F**: categorical boosting classifier. **G**: multinomial Naïve Bayes. **H**: support vector machine classifier with linear kernel. **I**: k -nearest neighbor classifier. **L**: logistic regression. **M**: multilayer perceptron classifier.

Supplementary Figure 3

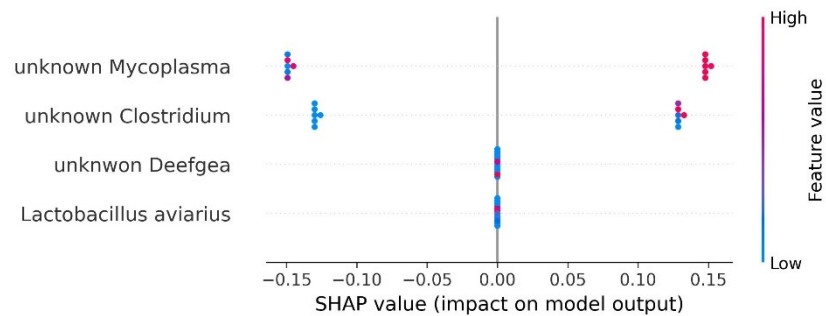
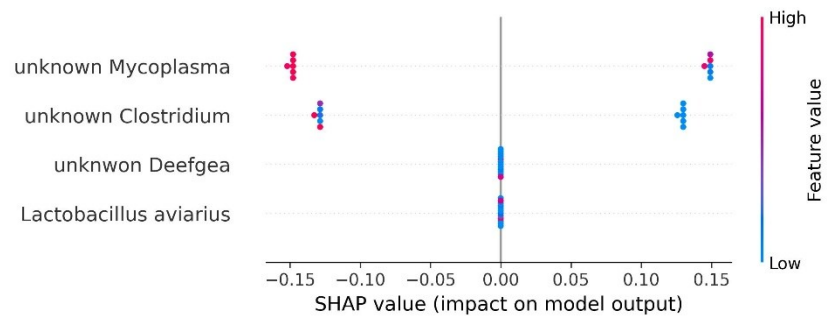
SHAP beeswarm plots of classification models for rainbow trout gut microbiota at the species level.



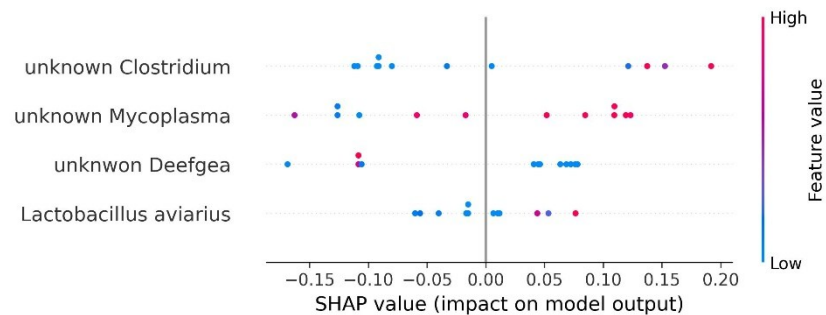
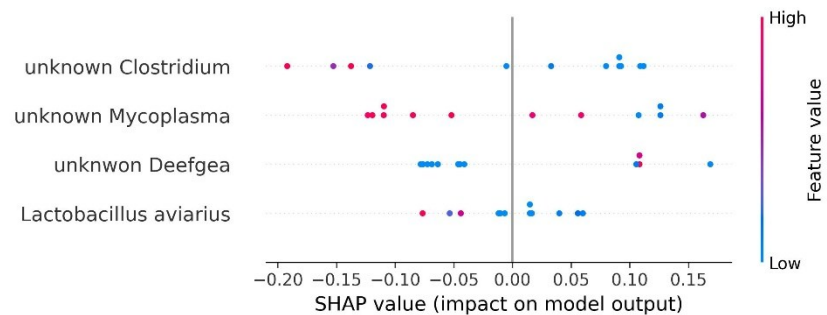
[D]



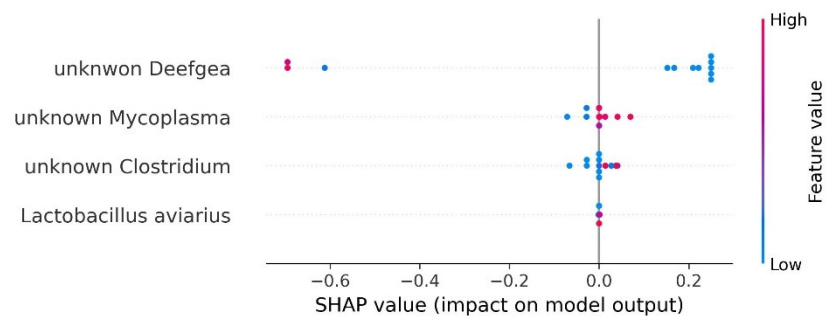
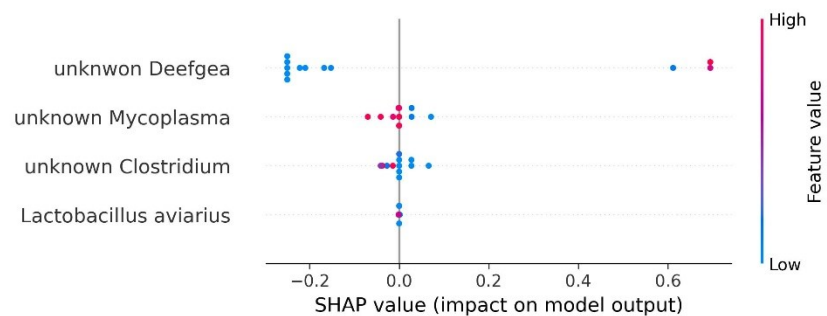
[E]



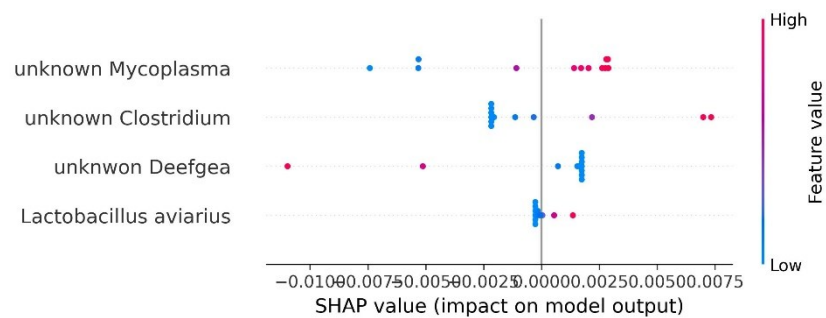
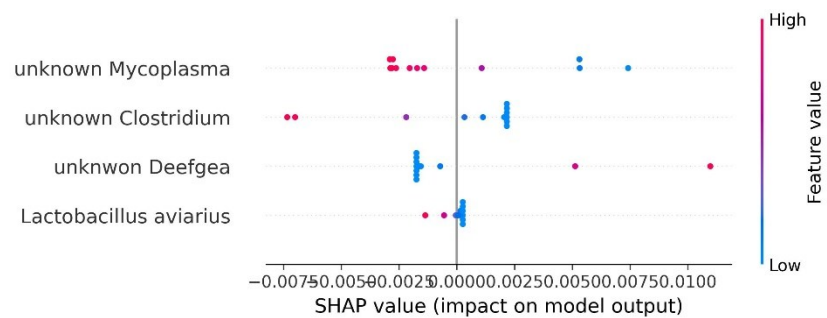
[F]



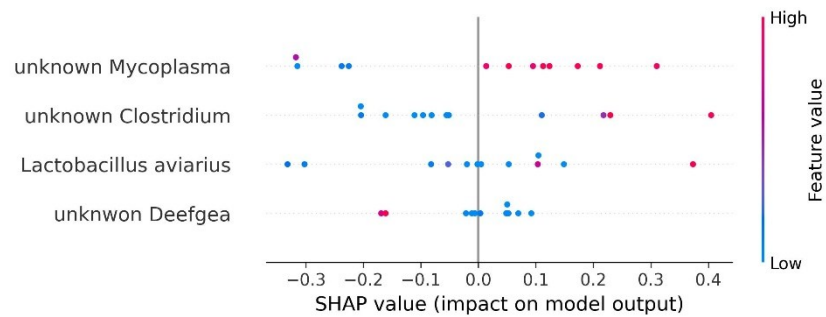
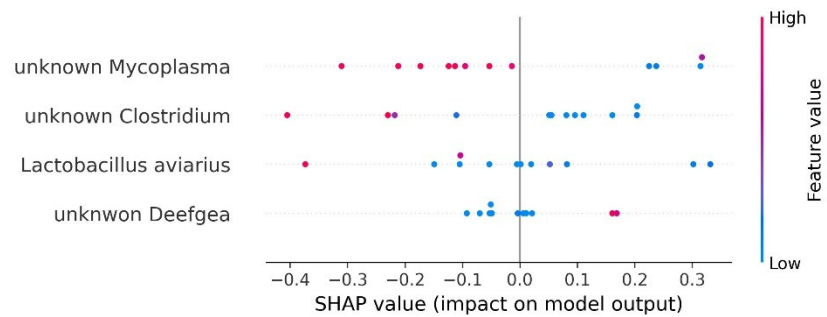
[G]



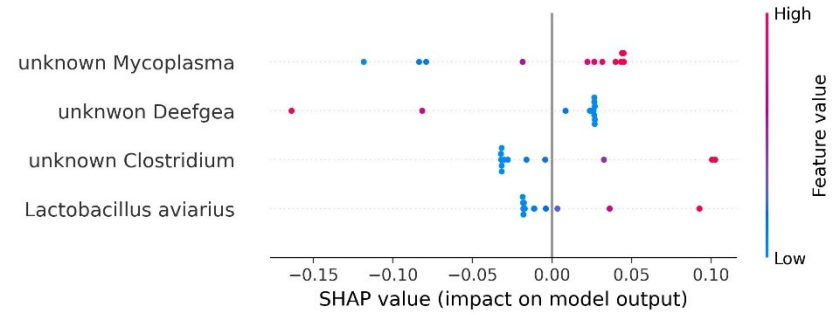
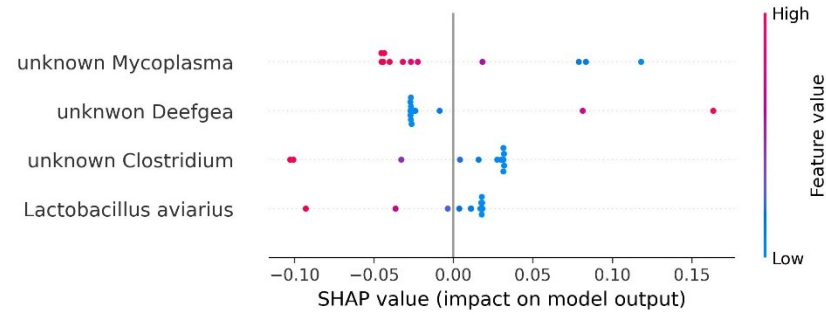
[H]



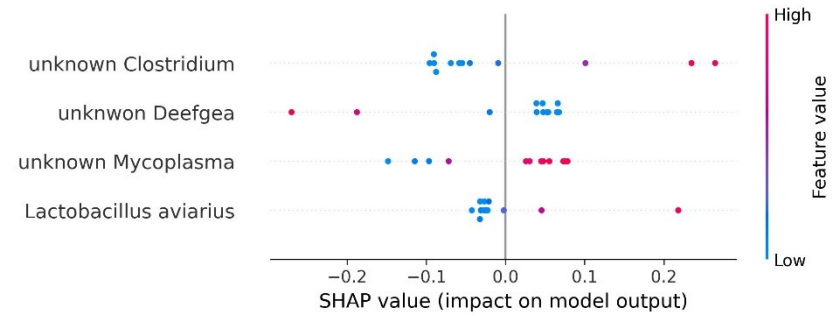
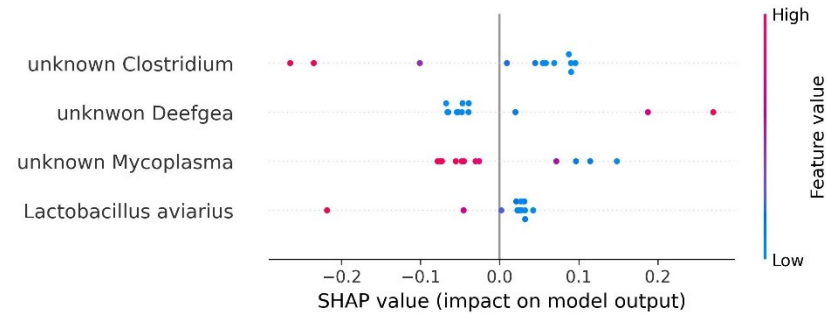
[I]



[L]



[M]



On the left, SHAP beeswarm plots referred to the TM0 diet. On the right, SHAP beeswarm plots referred to the TM100 diet. (A: decision tree classifier. B: random forest classifier. C: extra tree classifier. D: gradient boosting classifier. E: extreme gradient boosting classifier. F: categorical boosting classifier. G: multinomial Naïve Bayes. H: support vector machine classifier with linear kernel. I: k -nearest neighbor classifier. L: logistic regression. M: multilayer perceptron classifier.