DIFFERENT COMPOSITION PATTERNS OF MICROBIOTA IN *C. difficile* COLONISED AND NON - COLONISED HUMANS AND ANIMALS

Škraban Jure[3], Džeroski Sašo [3,4,5], Ženko Bernard [3,4], Rupnik Maja[1,2,3]

[1]University of Maribor, Faculty of Medicine, Maribor, Slovenia; [2]Institute for Public Health Maribor, Centre for Microbiology, Maribor, Slovenia; [3]Centre of Excellence for integrated approaches in chemistry and biology of proteins, Ljubljana, Slovenia; [4]Jožef Stefan Institute, Ljubljana, Slovenia; [5]Jožef Stefan international postgraduate school, Ljubljana, Slovenia; (Corresponding author: maja.rupnik@zzv-mb.si)

INTRODUCTION

An unperturbed intestinal microbiota is important in conferring the colonisation resistance against *C. difficile* infection (CDI). Studies on the association of gut microbiota and *C. difficile* are still not numerous and all of them deal only with the bacterial microbiota. In this study, we have used a simple molecular method (**DHPLC** - denaturing high pressure liquid chromatography), combined with **machine learning analysis** methods, to identify different patterns in the composition of **bacterial, archaeal and fungal microbiota** in *C. difficile* colonised and non - colonised humans and poultry.

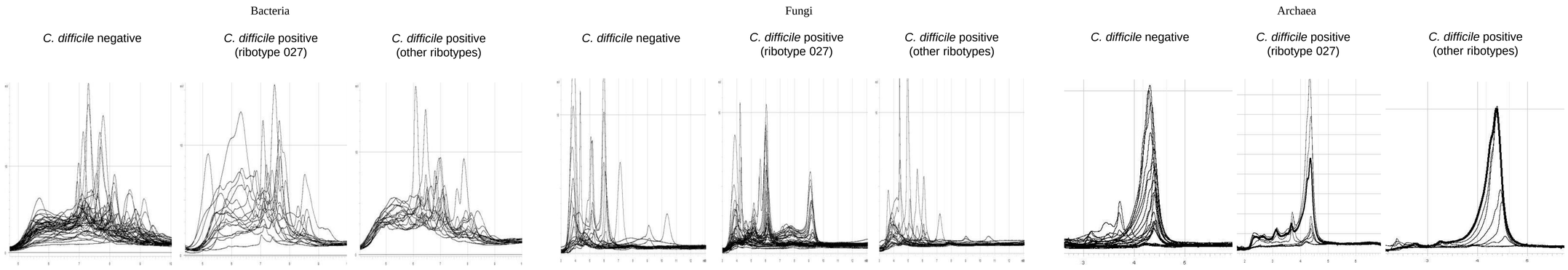

Figure 1. DHPLC profiles of dominant bacterial, fungal and archaeal faecal microbiota in humans. Microbial groups are represented by the peaks eluting at different retention times (x-axis).

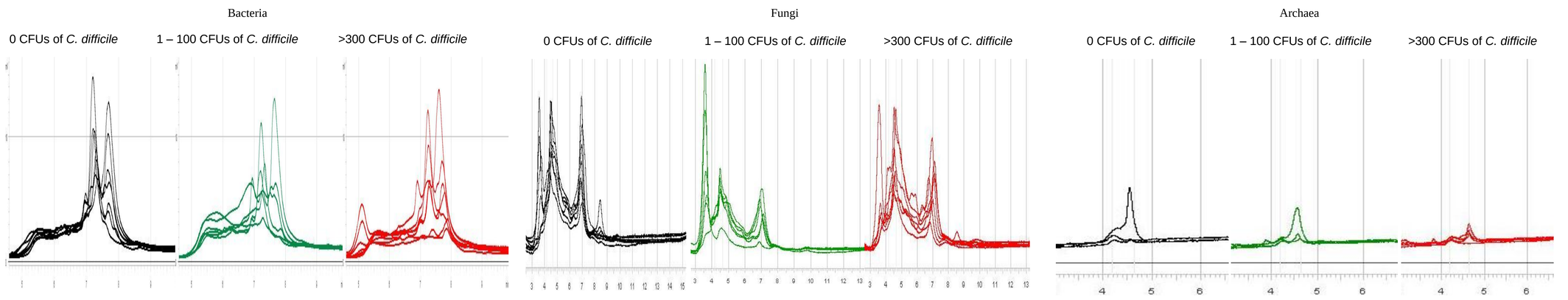

Figure 2. DHPLC profiles of dominant bacterial, fungal and archaeal faecal microbiota in 14 days old chickens. Microbial groups are represented by the peaks eluting at different retention times (x-axis). Sample groups: black / 0 CFUs of *C. difficile*, green / 1-100 CFUs of *C. difficile*, red / >300 CFUs of *C. difficile*.

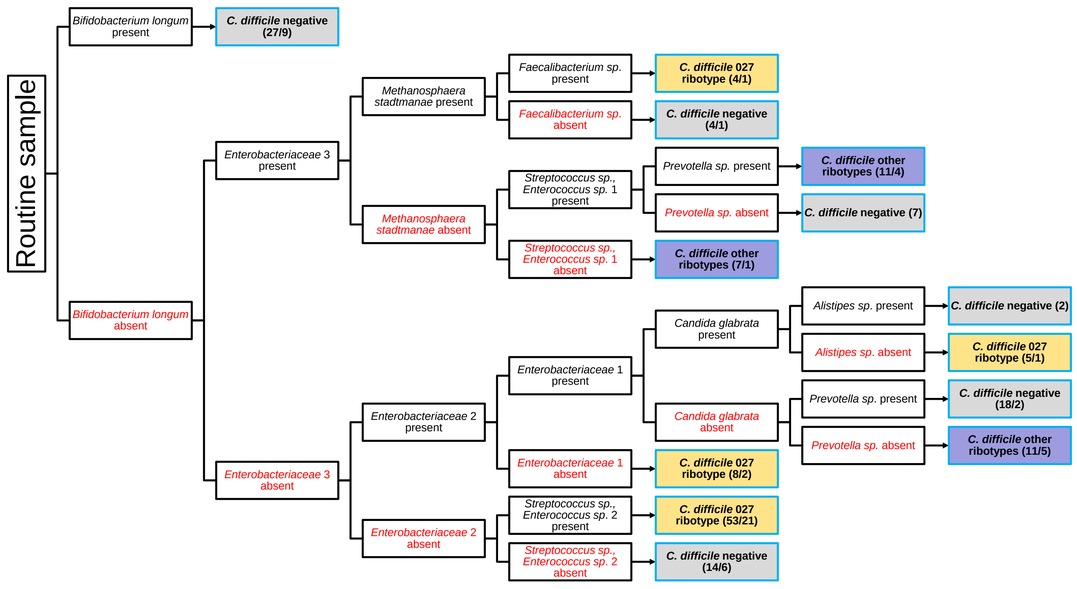

Figure 3. J48 decision tree describing type of colonisation with *C. difficile* in humans (*C. difficile* negative, *C. difficile* 027 ribotype, *C. difficile* other ribotype). The outcome is based on the presence and absence of microbial groups. In each leaf (colored prediction node), the numbers in parenthesis show the total number and the number of incorrectly cassified instances.

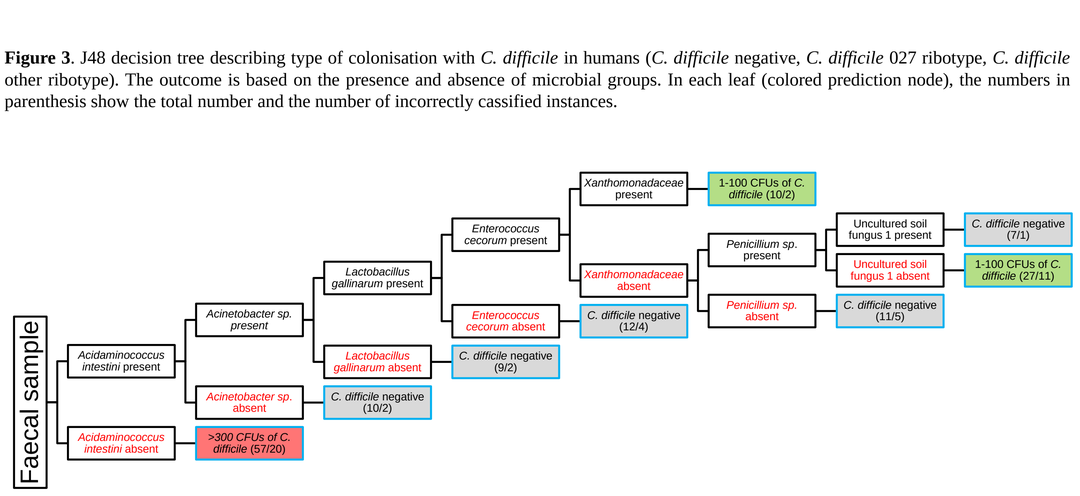

Figure 4. J48 decision tree describing the outcome regarding the number of CFUs of *C. difficile* in poultry (*C. difficile* negative, 0-100 CFUs of *C .difficile*, >300 CFUs of *C. difficile*). In each leaf (colored prediction node), the numbers in parentheses show the total number and the number of incorrectly cassified instances.

CONCLUSIONS

In conclusion, this is the first study to apply machine learning methods to analyse microbial colonisation paterns of gut microbiota. We have shown that in addition to bacteria, fungal and archaeal microbiota could also be important in *C. difficile* colonisation and that the patterns in the composition of the gut microbiota (instead of a single microorganism) are predictive for *C. difficile* colonization in humans and poultry.

RESULTS AND DISCUSSION

J48 decision tree in Figure 3 shows, that different branches of the tree are connected with the same predicted outcome (colored prediction nodes in figures 3 and 4), which means that different combination of microbial groups are associated with *C. difficile* colonization.

J48 decision tree enabled us to recognise patterns in microbiota which are predictive for *C. difficile* colonization, e.g., the pattern *Bifidobacterium longum* negative / *Enterococcaceae* negative / *Streptococcus sp*. - *Enterococcus sp*. positive was highly associated (32 correct predictions out of 53) with the presence of the PCR ribotype 027 in humans (Figure 3).

The key predictor associated with *C. difficile* negative samples in humans was *Bifidobacterium longum*. In 18 out of 27 cases the model conected this bacterium to *C. difficile* negative samples (Figure 3).

In poultry the J48 model identified *Acidaminococcus intestini* as a strong predictor for good *C. difficile* growth where it formed a separate leaf in the decision tree (Figure 4).

The microbes identified by the J48 model as predictive for the *C. difficile* colonization were found to be different in humans and chickens (Figures 3 and 4).

The J48 model recognised not only bacteria, but also fungi and archaea as predictors in the outcome of *C. difficile* colonisation (Figures 3 and 4).

MATERIALS AND METHODS

Samples and isolation of DNA:
We have analysed 171 human faecal samples and 143 poultry faecal samples from a single farm. The samples were divided into three groups according to the *C. difficile* ribotype (none, other, 027) for human samples, or the number of CFUs of *C. difficile* (none, 1-100, >300) after enrichment for poultry samples. The total DNA was isolated from the faecal samples by a standard procedure (Stool isolation kit from Qiagen).

Amplification and Separation of 16S rRNA and ITS2 with DHPLC:
Amplified 16S rRNA (V6-V8 in bacteria, V3 in archaea) or ITS2 (fungi) sequences were separated by denaturing high performance liquid chromatography (DHPLC) (Wave system), (Doman *et al*. 2003, Goldenberg *et al*. 2005), which is a high-throughput system for DNA fragment analysis that separates DNA based on fragment size and sequence (Figures 1 and 2), and also enables collection of the peak fractions for subsequent identification of bacterial, fungal and archaeal groups by cloning and sequencing.

Machine learning:
We have analysed the differences between the sample groups in terms of presence or absence of microbial groups by using machine learning methods for induction of decision trees (Quinlan, 1993), more specifically classification trees (Breiman et al., 1984). In particular, we used the WEKA J48 implementation (Witten *et al*. 2011) of the C4.5 algorithm (Quinlan, 1993). The default parameter settings for J48 were applied. For the human samples, the dependent (class) variable was the ribotype of *C. difficile* detected in a sample (none, 027, other). The independent variables (attributes) were the indicators of presence/absence of each of the 34 microbial groups (Figure 3). For the poultry samples, the dependent (class) variable was the number of *C. difficile* CFUs after enrichment (none, 1-100 CFUs, >300 CFUs). The independent variables (attributes) were the indicators of presence/absence of each of the 26 microbial groups.

REFERENCES

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984, Classification and Regression Trees, Wadsworth, Belmont, CA.
Domann E., Hong G., Imirzalioglu C., Turschner S., Kuhle J., Watzel C., Hain T., Hossain H., Chakraborty T.. 2003. Culture – Independent Identification of Pathogenic Bacteria and Polymicrobial Infections in the Genitourinary Tract of Renal Transplant recipients. Journal of clinical Microbiology, Dec. 2003, p. 5500 – 5510.
Goldenberg, O., Herrmann, S., Adam, T., Marjoram, G., Hong, G., Gobel, U.B., Graf, B., 2005. Use of denaturing high-performance liquid chromatography for rapid detection and identification of seven Candida species. J Clin Microbiol 43, 5912-5915.
Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco, CA, USA.
Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, CA, USA.