Association rule mining to help detect plant phenolic compounds putatively involved in decreased ruminal methane production in vitro

Didier Macheboeuf, Sylvie Guillaume, Clara Leguay, Sylvain Kerros, Agnes Cornu

Association rule mining to help detect plant phenolic compounds putatively involved in decreased ruminal methane production *in vitro*

Take home Message In the virgin forest of unknown bioactive plant phenolic compounds, association rules work like bushcutters clearing pathways to help discover active compounds.

Introduction In the search for natural alternatives to synthetic chemicals able to mitigate methane emission by ruminants, bioactive plant secondary metabolites are valuable candidates. However, these phytochemicals come in myriad chemical structures, and any one plant may contain hundreds of them. Even in plant extracts containing tannins, saponins or essential oils, it is difficult to link the presence of a compound or combination to the plant's activity. Here we focus on low-molecular-weight (<1000 Da) phenolic compounds from hydroalcoholic plants extracts, using association rules mining as a tool to emerge compounds involved in the antimethanogenic activity of plants during ruminal fermentation.

Material and methods A set of 208 plants was collected of the French Massif Central area . Samples were frozen in liquid nitrogen, freeze-dried, ground, and kept out of light to protect phenolic compound integrity. Fermentation profiles (gas production and composition, and volatile fatty acids (VFA) produced) were determined *in vitro* by incubating 600-mg pure plant substrates in rumen mixed bacteria cultures (buffer solution:rumen fluid at 2:1, v/v) for 24h at 39°C. All incubations were repeated 3 times. Each run period included 4 incubations of perennial ryegrass (PRG*)* used as control for a total of 48 runs. Plant production of methane and VFA were normalized and expressed as a ratio of mean PRG values for each period to eliminate inter-period drift. A plant was declared antimethanogenic ($p<0.01$) when its methane production was lower than the value fitted to the methane=f(VFA) linear regression, minus 2.58 times the standard deviation (s.d.) of PRG.
Phenolic compound profiles were obtained using an Agilent 1100 HPLC system. After ethanol:water (80:20) extraction, chromatographic separation was performed with a water:methanol gradient for 90 min on an Agilent C18 column and recorded at 280 and 320 nm with a diode array detector. A mix of 21 pure phenolic compounds was systematically injected in each analytical sequence (comprising about 10 plant extracts), in which flavone was taken as reference to calculate relative retention times (RRT). Sequence alignment was carried out by repositioning the standard RRTs. The area values were binarized (absence/presence). The resulting pair of matrices, i.e. 280 and 320nm, comprised 208 plants × 1092 peaks.
Association rules were created for frequent antecedent-consequent patterns using *ARA* software (*Association Rules Analyzer*, Papon 2016). The consequent was tied to the antimethanogenic item and antecedents were itemsets from peak matrices. Support*(S)* and confidence*(C)* criteria were used to identify important relationships. *S* indicated how frequently the itemset appeared in the database. C is a measure of how often the rule was found to be true.

Results & Discussion The consequent item included 64 antimethanogenic plants (Macheboeuf *et al.* 2014), of which 15 had strong effect (outliers of the methane=f(VFA) linear regression). The analysis for these plants generated from 29 to 161 peaks, with an average of 100. The very low filling ratios of the two peak matrices (0.10 and 0.09) justified the use of association rules rather than multivariate analysis. The very frequent peaks (>0.5) were discarded before data mining to avoid false-positives. With the minimum thresholds of 5 for S and 0.5 for C, there were 205 candidate peaks. In a first strategy, results were filtered via the constraints of a) co-occurrence of the peak in the 280 and 320 nm matrices and b) $C > 0.65$, which narrowed the candidate peaks down to 28. In a second strategy, the constraints were that the peaks had to be major (i.e. more than 10 times the area of the median peak) and present in the plants that showed high antimethanogenic effect (outliers), which narrowed the candidate peaks down to 24. Combining the two strategies resulted in 7 candidate peaks. One peak was easily identified as gallic acid. Based on absorbance spectrum between 200 and 400 nm, three others were cinnamic acid derivatives and two were flavonols.

Conclusion Association rules mining was able to select a compact number of peaks making identification feasible. The effect of these pure compounds now has to be verified for proof of the concept. While the algorithm works with qualitative data, using strategy which selects among the major peaks of the profiles serves to integrate the quantitative aspect.

Acknowledgements We thank ethnobotanist G. Lalière and the *Conservatoire Botanique National du Massif Central* for plant collection.

References

Macheboeuf D, Coudert L, Bergeault R, Laliere G, Niderkorn V 2014. Animal, 8, 1797-1806.
Papon P A 2016. Thesis, https://tel.archives-ouvertes.fr/tel-01412054.