



**CRC HBP**  
FOR HONEY BEE PRODUCTS

# Development of a Phytochemical Reference Profile for Honey Based on HPTLC and Machine Learning

Md Khairul Islam<sup>1,2</sup>, Kevin Vinsen<sup>1,3</sup>, Lee Yong Lim<sup>2</sup>, Katherine Hammer<sup>1,4</sup>, Cornelia Locher<sup>1,2</sup>

<sup>1</sup>Cooperative Research Centre for Honey Bee Products Limited (CRC HBP); <sup>2</sup>Division of Pharmacy, School of Allied Health, University of Western Australia, WA, Australia; <sup>3</sup>International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, WA, Australia; <sup>4</sup>School of Biomedical Sciences, University of Western Australia, WA, Australia



THE UNIVERSITY OF  
**WESTERN  
AUSTRALIA**

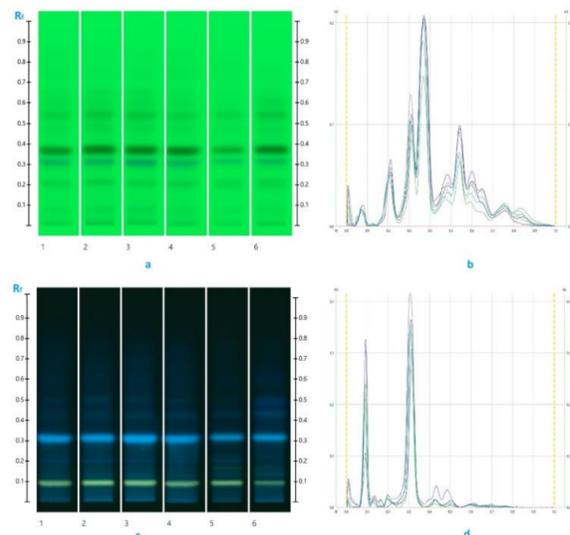
## Background

Honey is a sweet, viscous and highly concentrated natural product, produced by honey bees from flower nectar. The organoleptic and bioactivity profile of honey is directly related to its floral origin [1]. Monofloral honeys, which are predominately derived from a single botanical source, are highly sought after and priced accordingly [2]. However, as a natural product, honey is rarely completely monofloral, as honey bees forage opportunistically to collect nectar based on preference and nutritional needs. Geographical and environmental variations, differences in harvesting methods and post-harvest processing, as well as variations in storage conditions can also impact on the final phytochemical composition of honey. To obtain a reference phytochemical profile, which can be used in the authentication of monofloral honeys, sample pooling might be necessary to amplify typical and 'dilute out' unusual honey constituents [3]. We propose High-Performance Thin Layer Chromatography (HPTLC) coupled with machine learning (ML) as a tool to guide effective pooling for the preparation of phytochemical reference profiles for monofloral honeys.

## Methodology

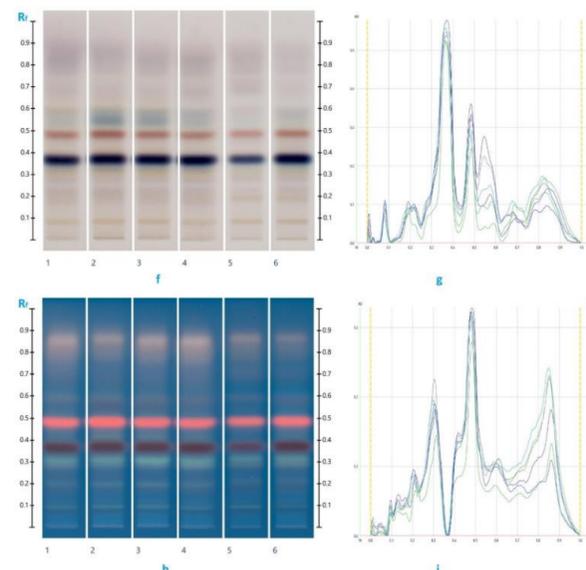
The HPTLC chromatographic profiles of various Honeys were obtained as described in [4,5]. Four sets of fingerprints for each sample (at 254 nm and 366 nm, and at white light and 366 nm after derivatisation with vanillin reagent) were converted into their respective chromatograms to derive values for Retention Factor (Rf) and corresponding intensities (AU) (Figure 1 and 2). The colours of the corresponding HPTLC bands were also recorded as RGB values. Machine learning was then performed using this HPTLC derived dataset to perform unsupervised (no ground truth labels) clustering of the data.

**Figure 1:** HPTLC Fingerprints of six different Coastal Peppermint (*Agonis flexuosa*) honey extracts



Images taken at 254 nm (a), at 366 nm (c) after development and their corresponding chromatograms (b, d)

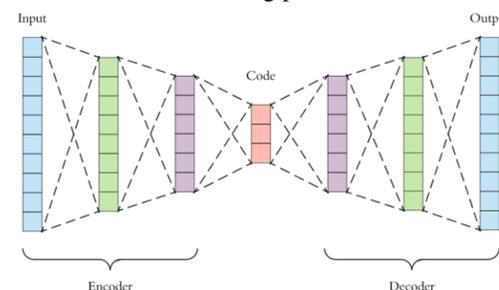
**Figure 2:** HPTLC Fingerprints of six different Coastal Peppermint (*Agonis flexuosa*) honey extracts



Images taken at white light (f), at 366 nm (h) after derivatization with vanillin reagent and their corresponding chromatograms (g, i)

The ML techniques employed were an autoencoder and a gaussian mixture model. An autoencoder is a type of neural network used to learn efficient coding of unlabelled data where the encoding or latent space is refined by attempting to regenerate the input from the encoding (Figure 3). Raw data generated by HPTLC analysis was segmented in the Rf plane, and reduced to an encoding of just 14 numbers. The autoencoder learns to ignore insignificant data ('noise') and shifting peaks in the AU and Rf domains. A gaussian mixture model (GMM) assumes that there are a certain number of Gaussian distributions, and each of these distributions represent a cluster. Using the encoder output from the autoencoder, the GMM groups the data points belonging to a particular set of distribution together. GMMs are probabilistic models and use a soft clustering approach for distributing the points in different clusters. Thus we can give the probability of a sample being in any cluster. The GMM output is shown in Figure 4 as a 3D cluster diagram.

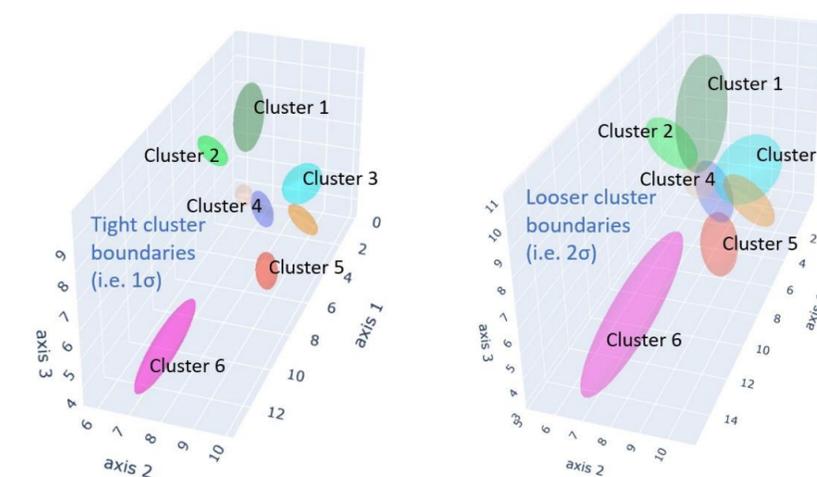
**Figure 3:** Graphical representation of autoencoding processes.



## Results

Honeys derived from the same predominant floral origin present as closely packed clusters in the 3D cluster diagram. Depending on specific industry needs, cluster boundaries can be defined more tightly or more loosely. Having the GMM allows us to assign probabilities to a particular honey sample falling within specific clusters, which will further support honey authentication by this method. For medicinal grade monofloral honeys, which require stricter quality control, cluster boundaries can be set more stringent compared to food grade honeys (Figure 4). These clusters can guide the preparation of pooled reference samples and from these phytochemical reference profiles can be derived that can be used in the authentication of monofloral honeys.

**Figure 4:** Graphical representation of cluster diagrams



## Conclusion

Rich HPTLC-derived data coupled with machine learning can be used as a convenient tool for the development of honey reference standards, which can be used in the authentication of monofloral honeys. The machine learning algorithm will improve the soft clustering as more and more data is added to the system.

## References

- [1] Soares, S., et al. (2017). Comprehensive Reviews in Food Science and Food Safety 16(5): p. 1072-1100; [2] da Silva, P. M., et al. (2016). Food Chemistry 196: p. 309-323; [3] Islam, M. K., et al. (2021). PloS One 16(7): e0254857; [4] Locher, C. et al., JPC - Journal of Planar Chromatography - Modern TLC, 2017. 30(1): p. 57-62; [5] Locher, C., et al., JPC - Journal of Planar Chromatography - Modern TLC, 2018. 31(3): p. 181-189.

## Funding

Cooperative Research Centre for Honey Bee Products Limited

