

# Chemometrics as a tool to answer to the challenges of the climate change and political situations on the consistency of cocoa flavor quality

Bagnolo E.<sup>1</sup>, Bortolini C.<sup>2</sup>, Liberto E.<sup>1</sup>

<sup>1</sup> Dept. of Drug Science and Technology, University of Turin, Via Pietro Giuria 9 (TO), Italy

<sup>2</sup> Soremartec Italia S.r.l. (Ferrero group), P.le P. Ferrero 1, 12051 Alba (CN), Italy

FERRERO

DSTF



## Aim and Scope

Cocoa is one of the most economically important commodities in the world and currently the main producers are West and East Africa (three-quarters of world production) and Latin America and Asia [1-2]. Cocoa production is strongly influenced by socio-political balances in farm areas and especially by climate change which affects both the amount of cocoa production and the flavour quality of the final products [3-6]. Since the flavour of cocoa products is the main property that directs the consumer's choice, as well as acting as an indicator of the quality of the product itself, it is necessary to ensure a high level of raw material quality standards [5]. Therefore, tools to track sustainable cocoa production are of paramount importance, especially due to climate change and the introduction of foreign varieties to support the high global demand for cocoa-derived products, which heavily influence cocoa production, quality and purity. With this in mind, the study we report is part of a broader project aimed at analysing the flavour profiles of cocoa from different origins and defining their chemical-sensory identity card aimed at a constant quality standard supply by also considering cocoa from different origins to cope with climate change and political conflicts.

## Materials and Methods

The volatile profile of one hundred sixty cocoa samples was analyzed by HS-SPME-GC-MS in combination with chemometrics for origin 'identification' [7-9]. The samples came from 4 origins: Colombia (COL), Ecuador (ECU), West Africa (WA) and Cameroon (CAM) and were harvested in 2015 and 2016 from different local regional farmers. Samples were of commercial grade (beans size "standard" based on counting test - federation of cocoa commerce) and were provided by Soremartec Italia s.r.l. (Alba, Italy). Cocoa samples were ground in liquid nitrogen to obtain a homogeneous powder and then stored at -80°C until the analysis. To achieve the above-mentioned objectives, fingerprinting (non-targeted) and profiling (targeted) approaches were used, which were able to decipher the information contained in the complex volatile dataset of both beans and liquor from different origins and validate the results, allowing the origins studied to be discriminated with successful classification models for their informative, discriminative and classification capacity.

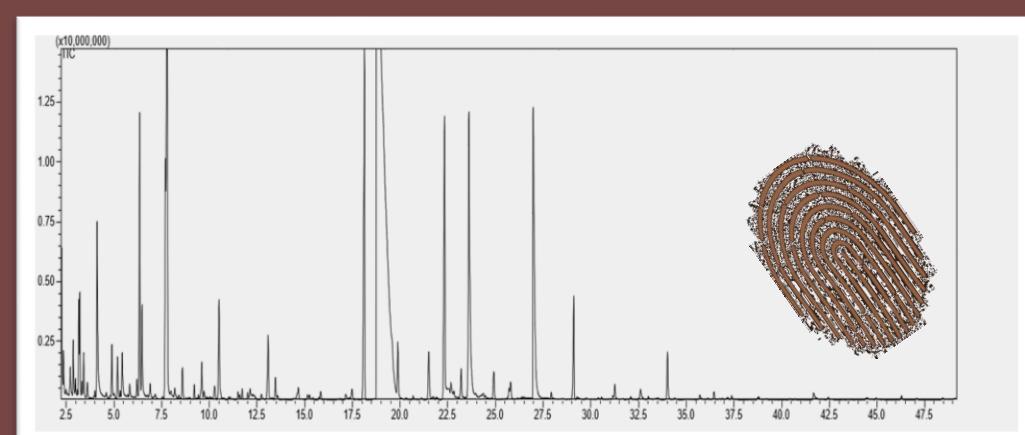


### HS-SPME-GS-MS: data acquisition

Automated Headspace Solid Phase Microextraction (auto-HS-SPME) was performed using a Combi-PAL AOC 5000 (Shimadzu, Milan, Italy) online integrated with a Shimadzu QP2010 GC-MS system provided with Shimadzu GC-MS Solution 2.51 software (Shimadzu, Milan, Italy). SPME fibre: DVB/CAR/PDMS df 50/30 µm - 2 cm length from Millipore (Bellefonte, PA, USA). The standard-in-fibre procedure was adopted to pre-load the ISTD (α-thujone) onto the fibre before sampling. 5.0 µL of ISTD solution were placed in a 20 mL glass vial and submitted to HS-SPME at 50°C for 20 min, stirring speed 350 rpm. Cocoa powder (1.00 g) was weighed in the headspace glass vials (20 mL) and submitted to automated HS-SPME sampling. After ISTD loading, the SPME device was exposed to the headspace of cocoa for 40 min at 50°C at a shaking speed of 350 rpm.

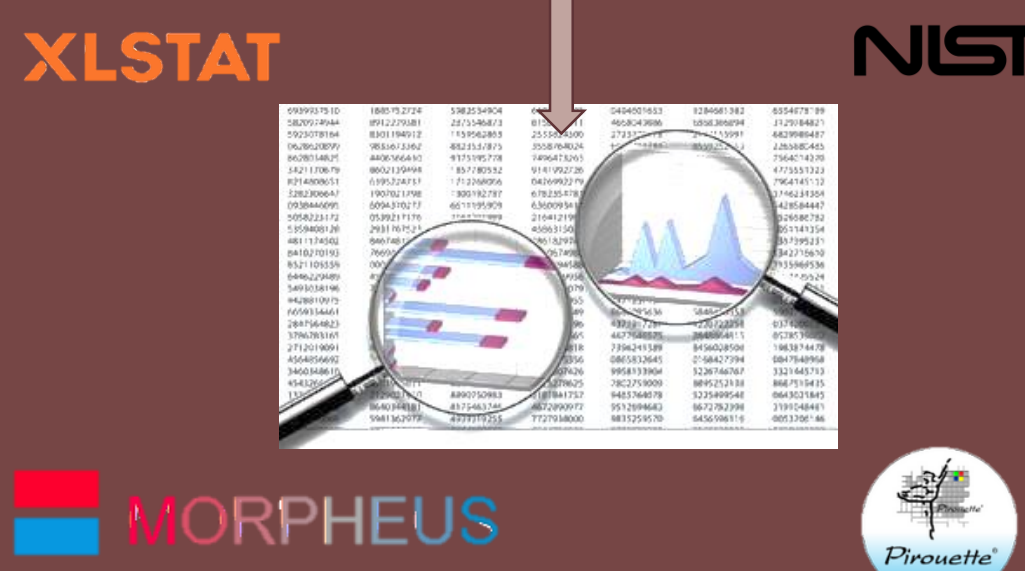


GC-MS analysis - Chromatographic conditions: analyses were run on a Shimadzu QP2010 GC-MS system, controlled by Shimadzu GC-MS Solution 2.5SU1 software (Shimadzu, Milan, Italy) injector temperature: 240°C, injection mode: splitless; carrier gas: helium, flow rate: 1 mL/min. Sampled analytes were recovered by thermal desorption into the split/splitless (S/SL) injection port of the GC system at 240°C for 5 min. GC column: SolGelwax (100% polyethylene glycol) 30 m x 0.25 mm dc x 0.25 µm df Trajan Analytical Science (Ringwood, Australia). Temperature program, from 40°C (2 min) to 200°C at 3.5°C/min, then to 240°C (5 min) at 10°C/min. MSD conditions: ionization mode: EI (70 eV); temperatures: ion source: 200°C; quadrupole: 150°C; transfer line: 260°C; scan range: 35-350 amu. Each sample was analyzed in triplicate.



### Data elaboration

Untargeted fingerprinting data elaboration, Principal Component Analysis (PCA), Partial Least Square Discriminant Analysis (PLS-DA) and regression analysis, was carried out with Pirouette® (Comprehensive Chemometrics Modelling Software, version 4.5-2014) (Infometrix, Inc. Bothell, WA). Heat-map was implemented in Morpheus (https://software.broadinstitute.org/morpheus/). Targeted analysis was focused on 55 compounds identified by matching their EI-MS fragmentation patterns (NIST MS Search algorithm, version 2.0, National Institute of Standards and Technology, Gaithersburg, MD, USA, with Direct Matching threshold 900) with those stored in commercial (NIST2014 and Wiley 7n) and in-house databases, and several standards available. Linear retention indices (LRI) were taken as a further parameter to support identification, and experimental values were compared to tabulated units. Profiling data elaboration and the Kruskal-Wallis test were performed with XLSTAT version 2021.4.1 statistical and data analysis solution. (Addinsoft (2022), New York, USA. https://www.xlstat.com/en)

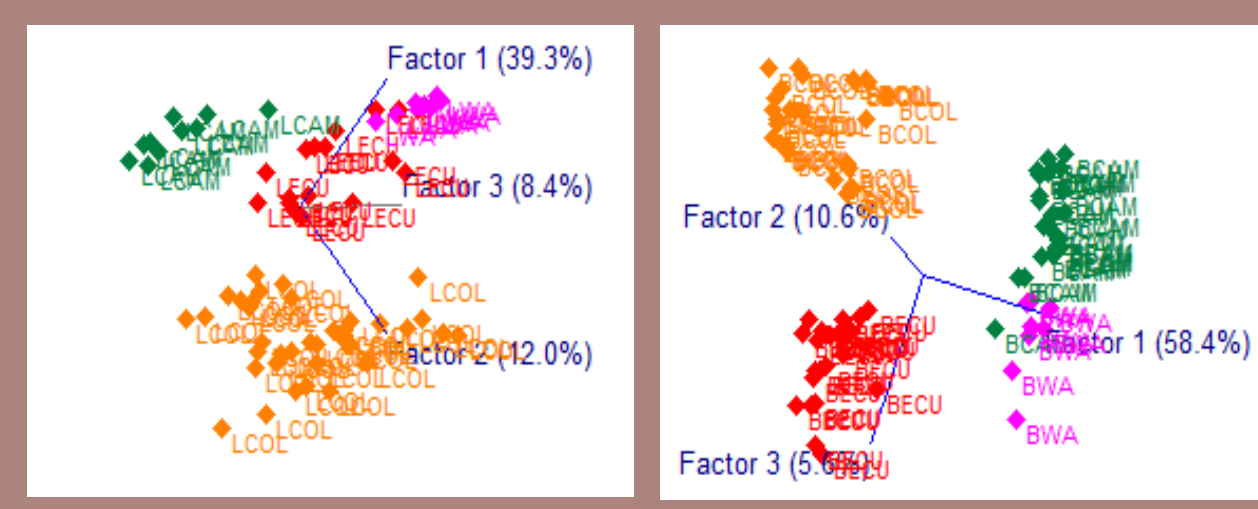


## RESULTS AND DISCUSSION

The results indicate a coherent, clear clustering of samples according to their origin with the two analytical strategies, both on raw beans and on cocoa liquors, albeit with differences at the molecular level. Predicting the classification of cocoa beans with the untargeted fingerprint on an external test set gave excellent results for beans with a classification rate of 100% and very good results for liquors (88%), despite the processing they underwent. Better results were obtained for targeted approaches with a classification rate above 92% (i.e. 92.86% for beans and 92.31% for liquors).

## Untargeted fingerprinting approach and origin identity

The increasing difficulties in procuring high quality raw materials due to the climate crisis have prompted the large-scale industry to look for new sources or new strategies to maintain or improve the standard quality of raw and finished products. Cocoa volatile profile is the final expression of the different "chemical and biochemical activities" occurring in the beans as a result of agronomic practices, post-harvest treatments and industrial processing [7]. With these perspectives, untargeted volatiliomics can be applied to define the fingerprint of a flavour quality standard and/or to benchmark with a quality reference e.g. linked to the origin. To define an origin identity based on the cocoa volatile profile it is necessary to have representative reference samples of each investigated origin; this operation is known as "identification" [8][9].



Exploratory data analysis on untargeted chromatographic fingerprints shows a better origin description with beans than with liquors, but in any case but in any case, with very good results. This agreement was not obvious since liquors undergo further processing steps that tend to homogenize origin information.

Supervised approach by PLS-DA for beans based on a training set of 66 samples internally cross-validated (CV 5) and an external test set of 18 samples shows excellent results with 100% of correct classification within origins. An OPLS-DA model cross-validated (CV 4) based on a liquor training set of 58 samples presents a total ability in origin classification of 88.24% when applied to the external test set (17 samples), with high specificity for all origins but a lower sensitivity (71%) for Colombia (COL) compared to other origins.

This approach is very similar to others applied for origin discrimination but in addition shows that beans and liquors provide similar chemical information [10]. Despite these excellent results in origin identification, the untargeted strategies do not provide detailed information about the aroma chemicals profile that, on the contrary, enable an in-depth knowledge of the aroma components that characterise the quality standards for cocoa flavour.

### References

- [1] Somarriva, E.; Pignero, F.; Cerda, R.; Orozco-Aguilar, L.; López-Sampson, A.; Leandro-Muñoz, M. E.; Jagoret, P.; Sinclair, F. L. (2021) Rehabilitation and Renovation of Cocoa (Theobroma Cacao L.) Agroforestry Systems. *A Review*. *Agron. Sustain. Dev.* 41 (5), 1-19.
- [2] ICCO Statistics - International Cocoa Organization https://www.icco.org/ (accessed Sep 5, 2022).
- [3] Boeckx, P.; Bultman, M.; Dewettinck, K. (2020) Poverty and Climate Change Challenges for Sustainable Intensification of Cocoa Systems. *Curr. Opin. Environ. Sustain.* 47, 106-111.
- [4] Agostosi, A. C.; Villal, L.; Milon, A. (2016) Flavor Chemistry of Cocoa and Cocoa Products: An Overview. *Compr. Rev. Food Sci. Food Saf.* 15 (1), 73-91.
- [5] Cacao Beans (2015) Cocoa Beans: Chocolate & Cocoa Industry Quality Requirements.
- [6] Perotti, P.; Cordero, C.; Bortolini, C.; Rubiolo, P.; Bicchi, C.; Liberto, E. (2020) Cocoa Smoky Off-Flavor: Chemical Characterization and Objective Evaluation for Quality Control. *Food Chem.* 309, 125561.
- [7] Lytton, A. E.; Panagou, E. Z.; Nychas, G. J. E. (2019) Volatiliomics for food quality and authentication. *Current Opin. Food Science*, 28, 88-95. https://doi.org/10.1016/j.cofs.2019.10.003
- [8] Cuadros-Rodríguez, L.; Ruiz-Sambalás, C.; Valverde-Som, L.; Pérez-Castaño, E.; & Casado, A. G. (2016). Chromatographic fingerprinting: An innovative approach for food 'identification' and food authentication - A tutorial. *Analytica Chimica Acta*, 909, 9-23. https://doi.org/10.1016/j.aca.2015.12.042
- [9] Sileo, F.; Jiménez-Canelo, A. M.; Liberto, E.; Bicchi, C.; Reschenbach, S. E.; Cuadros-Rodríguez, L.; & Cordero, C. (2021). Chromatographic Fingerprinting Enables Efficient Discrimination and Identification of High-Quality Italian Extra-Virgin Olive Oils. *Cite This: J. Agric. Food Chem.* 69, 8889. https://doi.org/10.1021/acs.jafc.1c02981
- [10] Medina, S.; Perestrello, R.; Silva, P.; Pereira, J. A. M.; & Câmara, J. S. (2019). Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends in Food Science & Technology*, 85, 163-176. https://doi.org/10.1016/j.tifs.2019.01.017
- [11] Belitz, H. D.; Grosch, W.; Schieberle, P. (2009). *Food Chemistry*. Berlin Heidelberg: Springer.

## Acknowledgments

The authors acknowledge the financial support of PhD program of the University of Turin and Soremartec Italia Srl (Ferrero group). The attribution is based on the PNRR, Mission 4, component 2 "From Research to Business" - Investment 3.3 "Introduction of innovative doctorates that respond to the innovation needs of companies and promote the recruitment of researchers from companies" DM 352/2022.

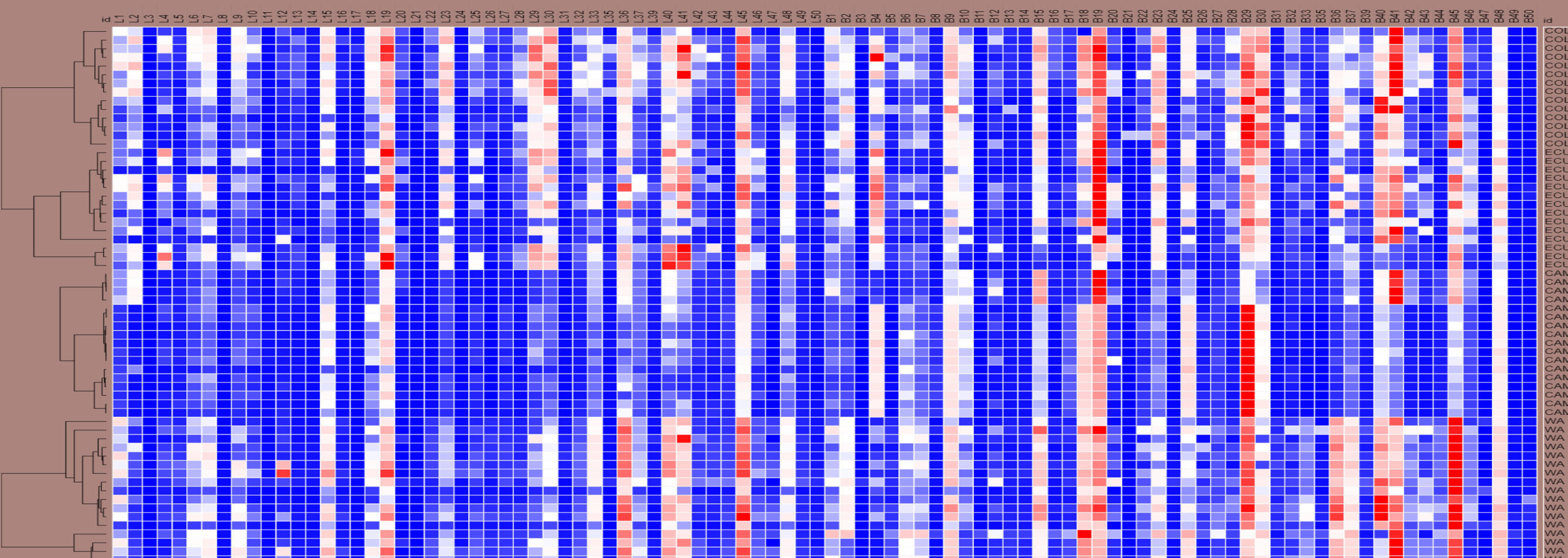


## Profiling and origin identity: chemical information provided from beans and liquors

Flavour is an essential criteria of quality for manufacturers of cocoa products. It represents an overall sensation that emerges from the interaction of taste, odour and textural feeling and it results from three classes of compounds, i.e. those responsible for taste (generally non-volatile compounds), for odours (aroma volatile substances) or for both of them [11]. Several chemical compounds are involved in the flavour composition (aldehydes, ketones, esters, alcohols, pyrazines, quinolines, furans, pyrones, lactones, pyrroles, and diketopiperazines) and their concentration strictly depends on the components formed during processing steps of the cocoa supply chain (fermentation, drying, roasting etc.). Therefore, the content of these components is related to the processing methods applied both in the countries of origin and by chocolate manufacturer. Gas chromatographic analysis of cocoa volatile profile benefits from detailed profiling, which can then be used for correlations with sensory data and characterisation of the chemical odour code.

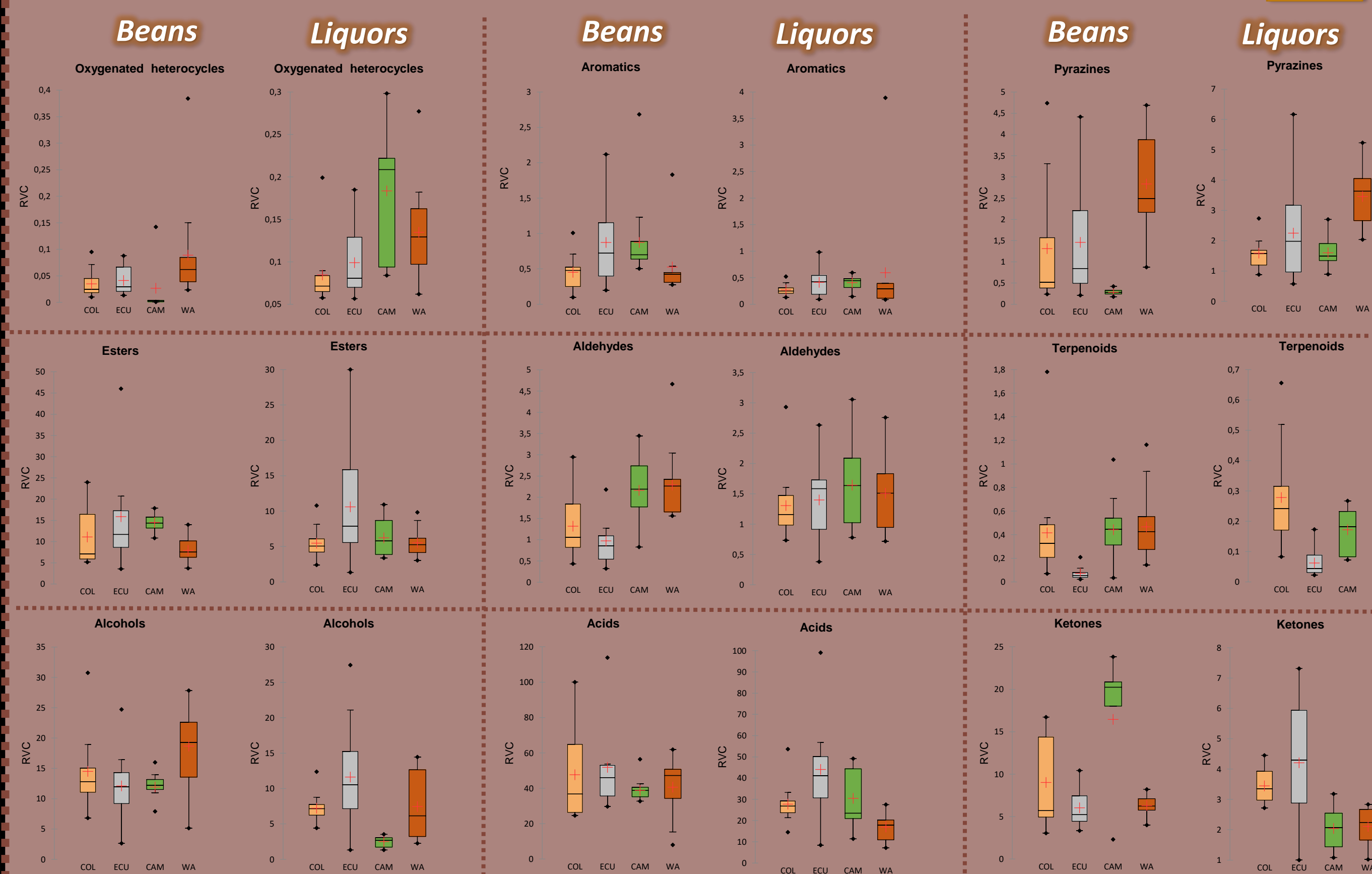
Figure 1 shows the chromatographic profiles in plan view, where L and B indicate liquors and beans, and the volatile distribution along the origins. The clustering of the origins is based on the agglomerative average linkage method and the one-minus Pearson correlation distance on the zeta score-adjusted data using Morpheus. The colour scale ranges from red (volatile substances in larger quantities) to blue (volatiles in smaller quantities).

Figure 1



From figure 2 we can observe that beans from different countries of origin have different ratios between volatiles, which mainly reflect the origin and primary processing in the country of origin (mainly fermentation and drying), as we can see from the boxplots of the different chemical classes. The content of volatiles of cocoa beans from Colombia was somewhat lower than in the other countries of origins. The relative content of volatiles was highest in cocoa liquors from Ecuador. The processing of the beans leads to a quasi-quantitative "flattening" of some volatiles within the origins, e.g. aromatics and aldehydes, while it increases the differences in esters, alcohols, terpenes, ketones and acids although with a different trend depending on the origins. For example, acids decrease more in COL and WA, but less on ECU and CAM liquors, and they changes their distribution with the origins. Acids are generated in the fermentation process, and in particular, acetic acid whose content is in strict relationship with the fermentation conditions. Esters are an important class of volatiles formed from amino acids that impart fruity and floral aromas whose content is negatively influenced from roasting. Liquors from Ecuador contains higher amounts of esters compared to other origins.

Figure 2



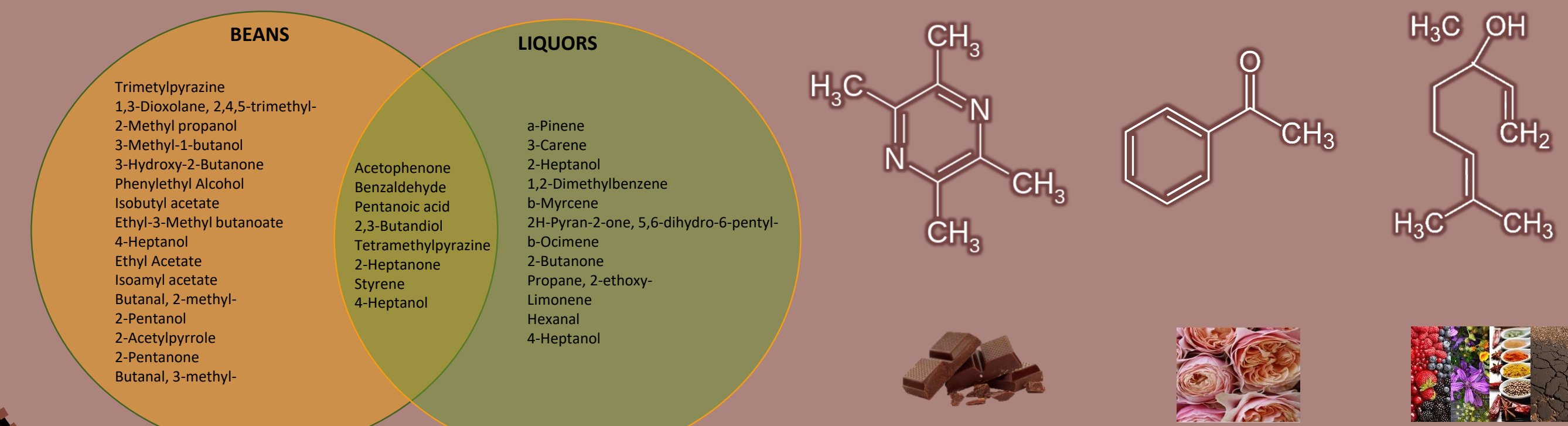
However, supervised methods are needed to classify and predict origin basing on chemical characteristics. The main objective of supervised machine learning is to model data to obtain qualitative prediction models, and thanks to dedicated algorithms to define target compounds that are characteristic of the origin. The Variable Importance in Projection (VIP), for example, is a well-known method to find significant variables in complex data sets. It enables to select variables from the PLS model by calculating the VIP scores for each chemical variable and to remove those with a VIP score below a predefined threshold (default=1).



Several VIP (>1) compounds are in common between beans and liquors: Acetophenone, Benzaldehyde, Pentanoic acid, 2,3-Butandiol, Tetramethylpyrazine, 2-Heptanone, Styrene, 4-Heptanol. These volatiles are important variables in the origin classification and share similar relative abundance in beans and liquors of the same sample.

**Tetramethylpyrazine** (odour quality; cocoa, chocolate - coffee) comes from Maillard reactions that can also occur in the beans during fermentation and drying due to the temperatures reached at this stage. However, roasting is the main factor contributing to its formation in liquors. **Acetophenone** is responsible for the floral and sweet notes among the ketones.

Besides these volatiles, a group of **monoterpenoids** contribute significantly in the definition of the origin of liquors. Although these volatiles are associated with odour description very far from cocoa flavour and have a high odour threshold, they do affect its overall aroma. These components derive from the monoterpene biosynthetic pathway and, during fermentation, they are present in higher amounts in beans than in pulp (Chetschik et al., 2018). Monoterpenoids have not yet been studied in depth in cocoa, but they can play a marked role in defining/monitoring the origin through the liquor volatile profile, the different fermentation processes and more in general, the post-harvest processing in the countries of origin.



## Conclusions and future perspectives

The volatile profile of cocoa beans and liquors is a very rich source of information to define the origin of incoming raw material. Both fingerprinting (untargeted) and profiling (targeted) strategies allow to decipher the information contained in the complex dataset resulting from HS-SPME-GC-MS analysis, and the cross-validation of the results enables to discriminate between the origins of both beans and liquors studied with suitable classification models. Volatile fingerprinting to be reliable and to be used as an authentication tool for origin requires a large number of representative samples, including the variability of cocoa products, to build-up a representative cocoa chemical identity card correlated to the sensory profile. The information provided by the untargeted approach can thereby be sufficient for a basic authentication of cocoa volatiles for routine control quality. On the other hand, the targeted approach is mandatory when a certification of origin is requested, as it requires an identity card detailing the chemical and sensory characteristics of the origin itself.