ChromMine™ Application Note 25-001

Cutting Through Complex GC-MS Data German Beers

Abstract

In this application note, we present the analysis of 28 German beers using SPME–GC-MS, comprising four defined styles and three unlabelled challenge samples. Data were first processed with Agilent MassHunter™ Unknowns Analysis (UA) and then imported into ChromMine™ for automated filtering, alignment, and clustering.

ChromMine's import routines corrected repeat identifications by grouping compounds to a median retention time and removing siloxane contaminants. Interactive bubble plots highlighted occasional misassignments, particularly with similar hydrocarbons. Subsequent clustering based on chromatographic profile and compound concentration reproducibly separated beers into distinct style groups.

The three challenge samples were correctly assigned to their respective styles, demonstrating accurate classification despite natural variation. This workflow shows how ChromMine converts complex chromatograms into interpretable fingerprints, supporting rapid, objective differentiation of beer types.

Introduction

Beer is one of the most widely consumed beverages worldwide, valued for its diversity of flavours and styles. From lagers and pilsners to IPAs and stouts, each beer type reflects a distinctive chemistry shaped by raw materials and fermentation.

German beers, the subject of this study, are restricted by the *Reinheitsgebot* (Purity Law) to just four ingredients - water, malt, hops, and yeast. Yet despite this apparent simplicity, the finished product is an extraordinarily complex mixture. Hundreds of volatile organic compounds contribute to its aroma, flavour, and style, producing chromatograms with multiple peaks, many of which overlap.

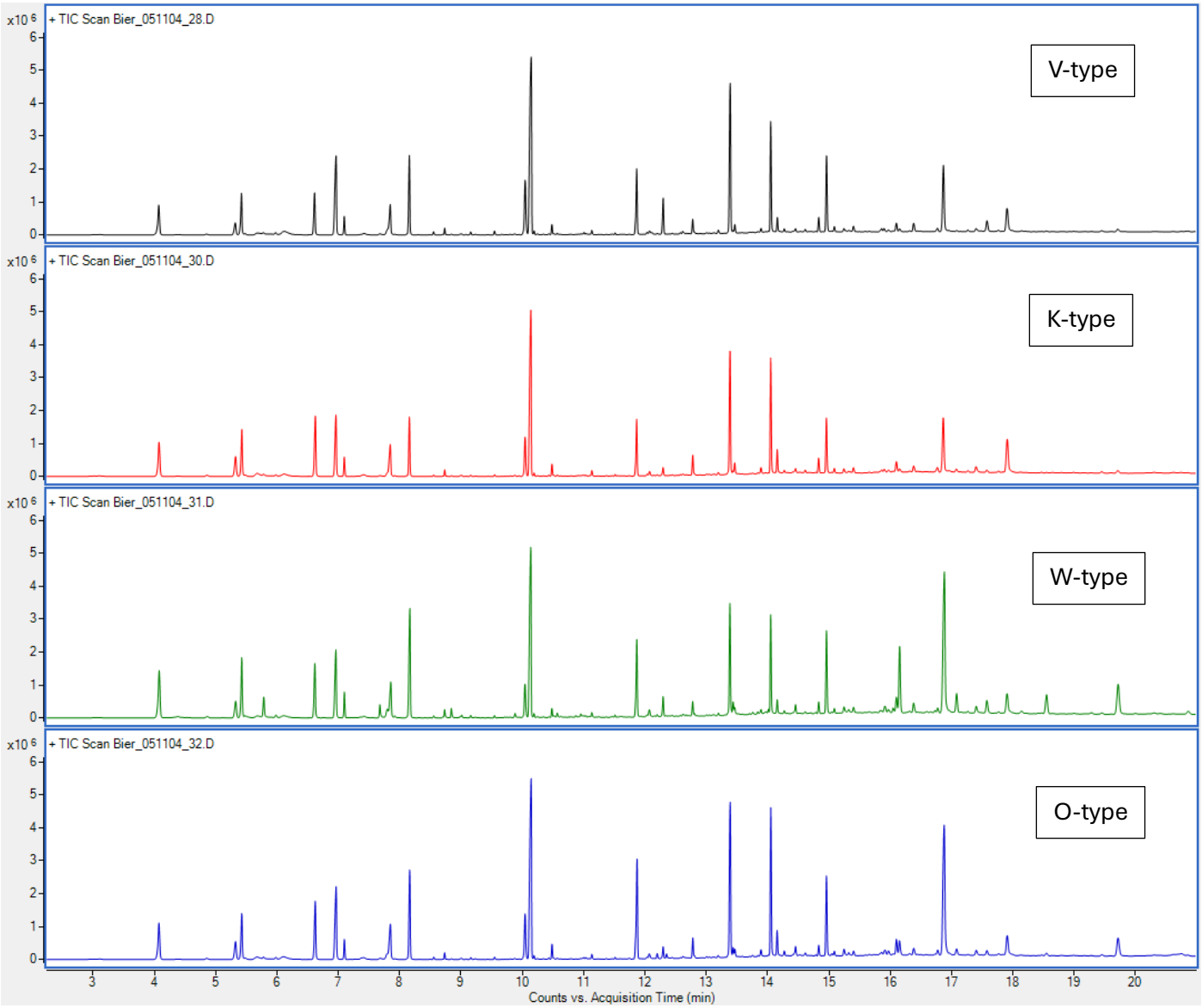
Analytically, this complexity presents several challenges:

- Chromatograms are dense and information-rich, often containing hundreds of peaks.
- Powerful deconvolution and library matching software, such as UA, are excellent at untangling this complexity, but can occasionally misidentify compounds. These library mismatches - where the same compound is assigned different identifications across samples - are especially prevalent with hydrocarbons.
- Siloxane and other background contaminants can obscure real signals.

These issues hinder objective comparisons and classification. Figure 1 shows sample chromatograms for each of the beer types studied. At first glance, the profiles appear broadly similar. However, closer inspection reveals subtle differences between beer styles. Traditionally, results might be exported to Microsoft Excel for comparison, but this process is both time-consuming and error-prone.

Reliable differentiation of beer types requires not only detection, but also systematic cleaning, retention time alignment, and clustering of compounds across replicate analyses. ChromMine provides this capability. By identifying and reclassifying repeat identifications, dynamically filtering contaminants, and clustering compounds based on retention time and library identity, ChromMine reduces data noise and reveals consistent sample groupings.

Figure 1: Example chromatograms from the four different beer types.



Experimental

Sampling and GC-MS analysis

Four different types of commercially available German beer (V, K, W and O types) were sampled using SPME and analysed using GC-MS. There were six replicates of each type, and three additional unlabelled "mystery" beers.

The chromatograms were first processed using Agilent MassHunter™ Quant-My-Way Unknowns Analysis (UA) and the resulting outputs were exported as a CSV file into ChromMine.

ChromMine import and processing

The raw data exported from UA consisted of a per-compound, per-sample CSV file with 11 columns (*including file name, sample name, beer type, retention time, compound name, component area, CAS number, molecular formula, and match factor*), and 1812 rows – amounting to nearly 20,000 data points. ChromMine™ ingested and parsed this dataset within seconds, automatically structuring the raw UA output into an analysis-ready table for downstream filtering and clustering.

During import, compounds identified at multiple retention times (within the same sample and/or across multiple samples) were automatically flagged and grouped. The user was then prompted to apply a median retention-time prefix (or a unique prefix) so that multiply identified compounds could be treated as distinct grouped entities.

Dynamic filters were then applied to remove all siloxanes – a common system/SPME contaminant not related to the sample – ensuring these did not influence subsequent clustering.

Bubble plot data review: The next step was to review the chromatograms using the bubble plot view (see Figure 2). Here retention time runs vertically (top to bottom) and samples are arranged as columns. Large markers represent larger peaks as does colour: red indicating high concentration and blue low. This facilitated rapid identification of potentially misidentified compounds (a common occurrence when multiple "best hits" are possible, particularly for hydrocarbons). Replicate samples made this process straightforward, as mis-assignments were immediately visible in the bubble plot representation (Figure 2).

ChromMine sample clustering and challenge sample attribution

Figure 2 shows how well ChromMine's median RT grouping and renaming works – with, in this case, only one obvious outlier shown at 6.616 minutes where an unsaturated C10 hydrocarbon has been identified as a branched-chain octadiene, whereas in all the other repeat samples it was identified as a branched-chain heptadiene. In the case of these beer samples there were very few misattributed compounds, so clustering was continued without any realignment.

Clustering samples

Principal Component Analysis (PCA) has its place – but is a black box – making it hard to interpret, compare clusters and give meaning to the separations. ChromMine's clustering works differently – it looks at profile similarity, and compound concentrations across the whole chromatogram. Then rather than projecting onto a principal component 1 (PC1) versus principal component 2 (PC2) feature space, ChromMine projects the samples onto volatile concentration similarity versus profile/sample similarity cluster plot.

This novel style of clustering is beneficial to chemists as proximity or separation horizontally means that the samples are similar or dissimilar from a chromatographic profile perspective, and proximity or separation vertically tells the same story but with regards to concentration.

Figures 3 through 5 show how the beer samples project into this space and how step by step the clustering is refined through threshold adjustments.

Figure 2: Bubble plot excerpt for O-type beer replicates.
Horizontally aligned compounds (green lines) with similar concentration are likely correctly identified (or at least reproducibly so). Compounds with similar concentrations that are not horizontally aligned (red line) likely contain mass-spectrometric misidentifications and are best rectified if clustering is unreliable. The red arrow indicates where a compound has the wrong ID.

Figure 3: Beer samples with initial clustering with default settings.

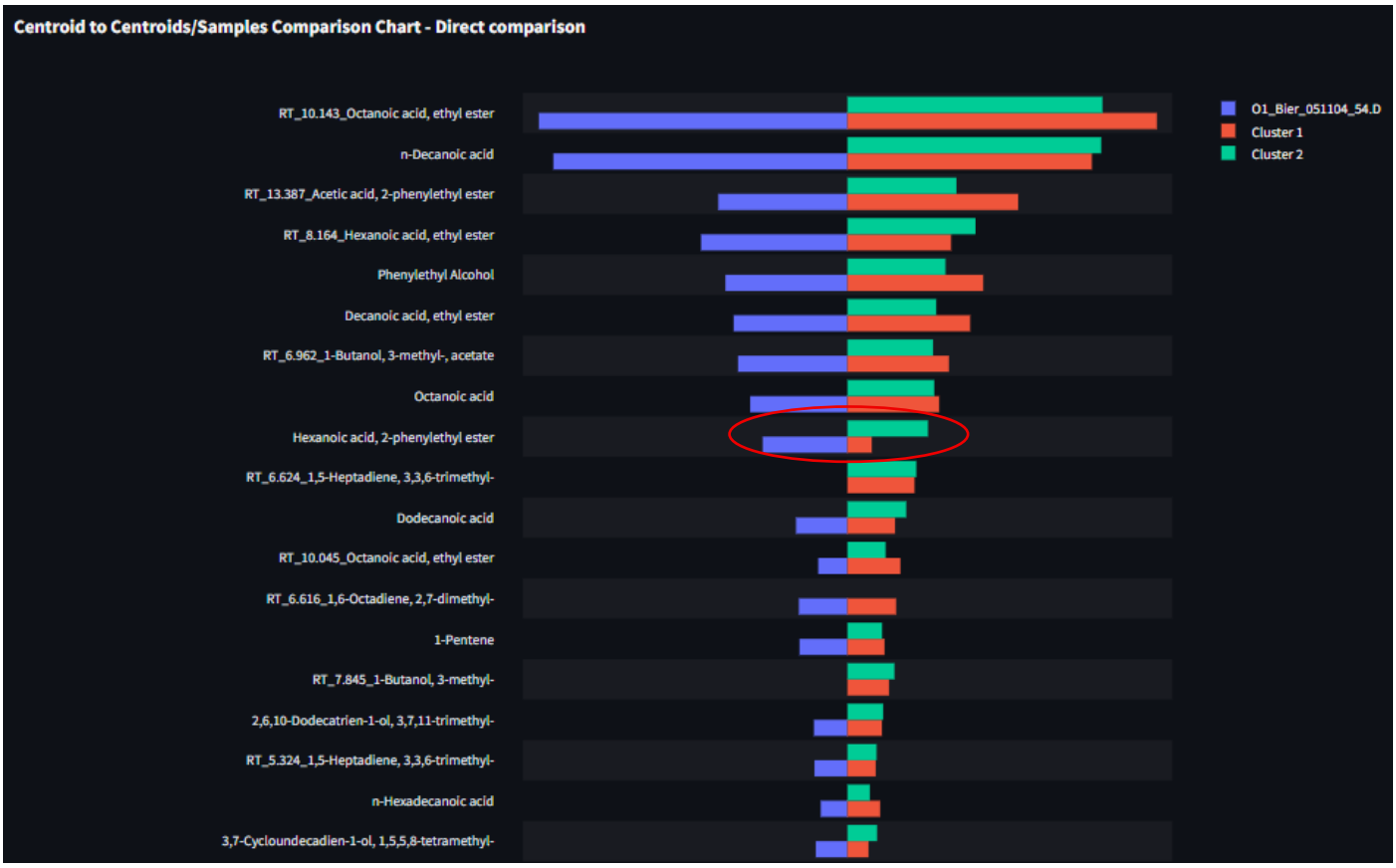
Figure 3 shows that the initial import groups all samples as one cluster – however closer inspection shows that the O beers, K beers, V beers and W beers cluster as discrete groups. Note the larger "glowing" dot in the middle represents the geometric mean profile for that cluster. Changing the sample and concentration similarity thresholds produces figure 4.

Figure 4, with adjusted sample and concentration similarity thresholds, shows the beer types grouping well – aside from one O1 beer that has fallen into the K beer type (green Cluster 1) rather than with the O beer type (blue Cluster 2).

By using the cluster/sample comparison tool ChromMine easily allows you to identify why that one O-type beer replicate fell into the K-type beer cluster. Figure 5 shows this comparison and this one run of an O-type beer had a higher-than-normal concentration of hexanoic acid 2-phenyl-ethyl ether, and which was closer in concentration to that seen in the Cluster 2 beers (K beer type). A major time saving compared to manual data comparison.

Figure 4: Beer samples clustering with adjusted thresholds – O type beer outlier circled in red.

Figure 5: Comparing the O1 outlier (blue bars) to the geometric mean profiles for O type (red bars, cluster 1) and K type (green bars, cluster 2) beers. Assymetry, ie. compounds responsible for outliers, are much easy to spot.



Projecting the unknown challenge beers into the clustering feature space

The primary goals of this work were twofold:

1) to see if the very profile similar beers (see Figure 1) could be reliably clustered, and
2) to project the unknown challenge samples into the clustering feature space and see if they could be reliably attributed to specific beer types.

Figure 6 shows the unknown beers projected into the feature space previously shown in figure 4. The previously unknown samples are encircled in red. Each challenge beer was correctly identified: one was a W-type beer, two was a V- type beer and three was a K-type beer.

Conclusions

ChromMine transforms complex beer chromatograms into interpretable datasets. By removing siloxane contaminants and clustering profiles consistently across samples, it enables:

- reliable style differentiation despite natural variation,
- objective classification of unknown samples, and
- faster, cleaner insight into complex chemistry.

This approach offers broad potential in flavour, quality control, and authenticity testing — both for beer and for other complex natural products.

Acknowledgements

Thanks to the application team at GERSTEL for the beer data.

Figure 6: Unknown challenge samples projected into cluster feature space – the previously unknown samples are encircled in red.