

Chemovar Typing of *Cannabis* Strains with MarkerView™ Software and the SCIEX X500R QTOF System

KC Hyland

¹SCIEX, Redwood City, CA USA

Consumers of *Cannabis* have long asserted that there are differences between cultivars, or strains, in terms of user experience. The historical classification scheme of *indica*, *sativa*, and hybrid has prevailed as a popular and simplified way to identify *Cannabis* cultivars of perceived properties and ancestral origin. Consumers often select desired products based on this class-based assumption of its qualities. However, modern approaches to analytical testing of *Cannabis*, spurred largely by a rapidly-expanding legal market and new regulations on the industry, have shown that these historical classifications no longer explain the differences between, or properties of, the hundreds of commercially available *Cannabis* strains¹. More relevant descriptions and explanations of strain-specific chemical profiles are increasingly being discussed, with significant implications for patient treatment, intellectual property development, metabolomic profiling, and economically motivated adulteration. The concept of chemovars – a chemistry-based, phenotypical fingerprint rather than a horticultural cultivar – has been proposed and has gained popularity in the scientific community.



A nontargeted approach is described which does not presume to know the differences in chemical signatures between strains, but instead uses PCA statistics and suspect screening to identify those differences. A targeted method, like those employed historically, operates on the assumption that differences can be identified using a set list of cannabinoids and terpenes.

Key Advantages of Chemovar Classification Approach

- The nontargeted approach does not try to target a “short-list” of cannabinoids or terpenes, and is therefore more able to identify novel biochemical markers which might explain the differences between sample groups.
- Generic mass spectrometry data acquisition in the form of SWATH® Acquisition for collection of HRAM MS and MS/MS information means that minimal method development is required to run this workflow.
- Atmospheric pressure chemical ionization (APCI) allows for robust ionization of endogenous compounds found in the *Cannabis* plant with minimal ion suppression.
- Ability to use MarkerView Software statistical tools to build up a “Peaks of Interest” list means that the workflow does not depend on processing data for potentially thousands of detected features resulting from a simple screening workflow.
- SCIEX OS Software 1.4 allows easy interfacing of acquired data with tools for compound ID, such as FormulaFinder and the ChemSpider database.

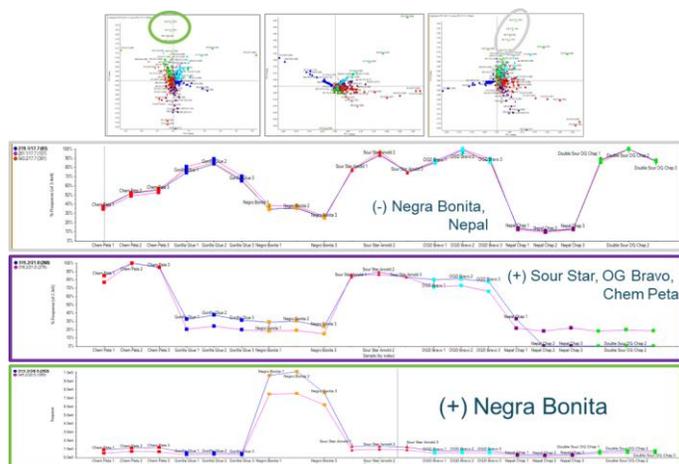


Figure 1. Principle Component Analysis of Seven *Cannabis* Strains using MarkerView software. PCA shows distinct clustering of the sample groups (strains) in several combinations of Principle Component (PC) combinations as well as with and without data normalization. Extracting features most distinctive along the PCA groups reveals the profile of those features for the different sample groups, and the Peaks of Interest feature can be classified as being upregulated (+) or downregulated (-) in different strains.

Experimental

Sample Collection and Extraction: Seven different strains were sampled to test the workflow and the concept of statistical classification of chemovars based on chemical signatures. Triplicate samples of each strain were extracted into acetonitrile according to the modified vMethod™ protocol, followed by a 200x dilution in methanol.^{2,3}

HPLC Conditions: Analytes from all compound classes were separated on a Phenomenex Kinetex 2.6 μm Biphenyl LC Column (150 x 4.6 mm) held at 30° C using a SCIEX ExionLC™ AD System, with mobile phases consisting of A) Water + 5 mM ammonium acetate + 0.1% formic acid and B) Methanol:Water (98:2) + 5 mM ammonium acetate. The gradient and separation conditions had been previously optimized for the separation of terpene and cannabinoid isomers, and as such were deemed the most appropriate method for this application.³ The flow rate was 0.8 mL/min and the gradient program can be seen in Figure 2.

Mass Spectrometry Conditions: SCIEX X500R QTOF System with Turbo V™ Ion Source outfitted with a probe for atmospheric pressure chemical ionization (APCI) was operated in positive ionization mode. A TOF MS scan was conducted from 50-1000 m/z. Variable window SWATH Acquisition was employed to collect all MS/MS information for potential marker compounds in the highly complex *Cannabis* matrix. The variable window precursor ion widths were designated for maximum MS/MS specificity by using the Variable Window Optimization Tool.⁵ APCI utilized the following source settings: NC = 1 μA, TEM = 625 °C, CUR = 35 psi, CAD = 11, GS1 = 37 psi.

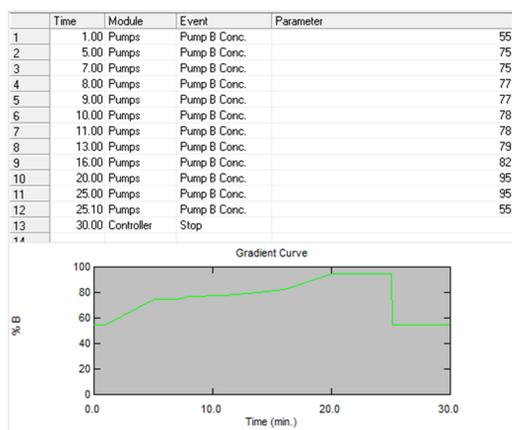


Figure 2. LC Gradient Time Program. The total gradient run time was 30 minutes and flow rate 0.8 mL/min.

A Nontargeted Approach

Workflow

This workflow is considered “nontargeted” for two main reasons. First, the experimental data collection is a SWATH Acquisition method with no defined target analytes. Second, the statistical data processing in MarkerView Software does not designate a target list of predetermined features on which to build the PCA model. It should be noted that certain method parameters (such as sample preparation conditions, LC separation conditions, ionization source conditions, and defined variable SWATH Acquisition windows) cannot be considered truly “nontargeted” techniques, as these were optimized for the detection of the endogenous *Cannabis* terpene and cannabinoid content. Since the goal of this work is to evaluate the chemical differences between *Cannabis* strains, such parameter selection is deemed appropriate.

The workflow used for processing the acquired data is outlined in Figure 3. Following the loading of the data for all samples and sample groups into MarkerView Software, the PCA modeling of the features and t-test comparison of sample groups was first tested. Using these functions is a rapid way to pinpoint important chemical features before performing a compound identification. Searching MS/MS databases represents the “first pass” at identifying the Peaks of Interest, and the All-in-One with NIST library is a key tool for maximizing coverage of natural products and other potential marker compounds. If the library search produces no spectral match for a Peak of Interest, the second portion of the workflow is triggered. In this procedure, the accurate mass and isotope ratio data from precursor ions allows the FormulaFinder algorithm to provide an empirical formula for the target feature. There might be more than one proposed formula, in which case the best match or most likely formula should be selected. The formula can be searched against the ChemSpider database within the SCIEX OS Software to achieve a list of structures in the database corresponding to the proposed formula. This list of structures can be narrowed down to a selection of potential candidate structures by comparing the experimental MS/MS data to in-silico predicted fragmentation of the database structures and selecting the structures demonstrating the closest match.

Statistical Analyses in MarkerView Software

Figure 4 highlights the clustering of the seven strains analyzed in PCA space. The unsupervised statistical differentiation clustered together the replicate extractions within the same *Cannabis* strain giving confidence to the chemical differences observed between strains. The proximity of the strains in this PCA space can give insight into the chemical similarity between the strains,

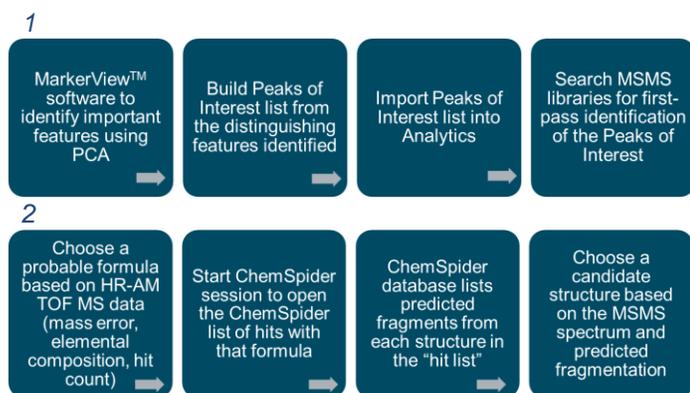


Figure 3: Detailed Data Processing Workflow for Nontargeted Chemotyping. 1.) Load acquired data into MarkerView Software and utilize available statistical tools to build a list of potential marker compounds which describe or distinguish the chemovar differences between the sample groups. Use SCIEX OS Analytics module to screen this target list against MS/MS libraries for candidate identification. 2.) If the library search produces no spectral match for a Peak of Interest, this portion of the workflow is triggered. Empirical formula was achieved using HRAM MS data and searched against the ChemSpider to produce structures corresponding to the formula. This list was narrowed down to a selection of candidates by matching the experimental MS/MS data to predicted fragmentation of the database structures.

while the PCA loadings plot gives insight into the specific chemical features causing the most differentiation; features existing further away from the center of the PCA loadings plot are responsible for the most differentiation, whereas features towards the center of the loadings plot are the most similar between the samples.

T-test analysis also provides statistical information when comparing one strain to another, or one strain to the rest of the samples. This can be further used to identify features that are uniquely upregulated or downregulated in a sample. One example t-test is shown in Figure 5, demonstrating the functionality of the MarkerView Software to take the t-test results and construct the volcano plot, in the example, for the Negro Bonita strain versus all other strains. This plot displays the statistical significance against the features causing differentiation between strains. Here, features at the extremes of the x-axis show the most difference between chosen samples, whereas those lowest on the y-axis exhibit the most statistical certainty of that difference, hence features in the bottom left and right corners are the most distinguishing features and make reasonable sense to first probe for structural identification. Features can be highlighted in this plot, and their area distribution across the samples and replicates can be automatically displayed as a profile (Figure 5, bottom).

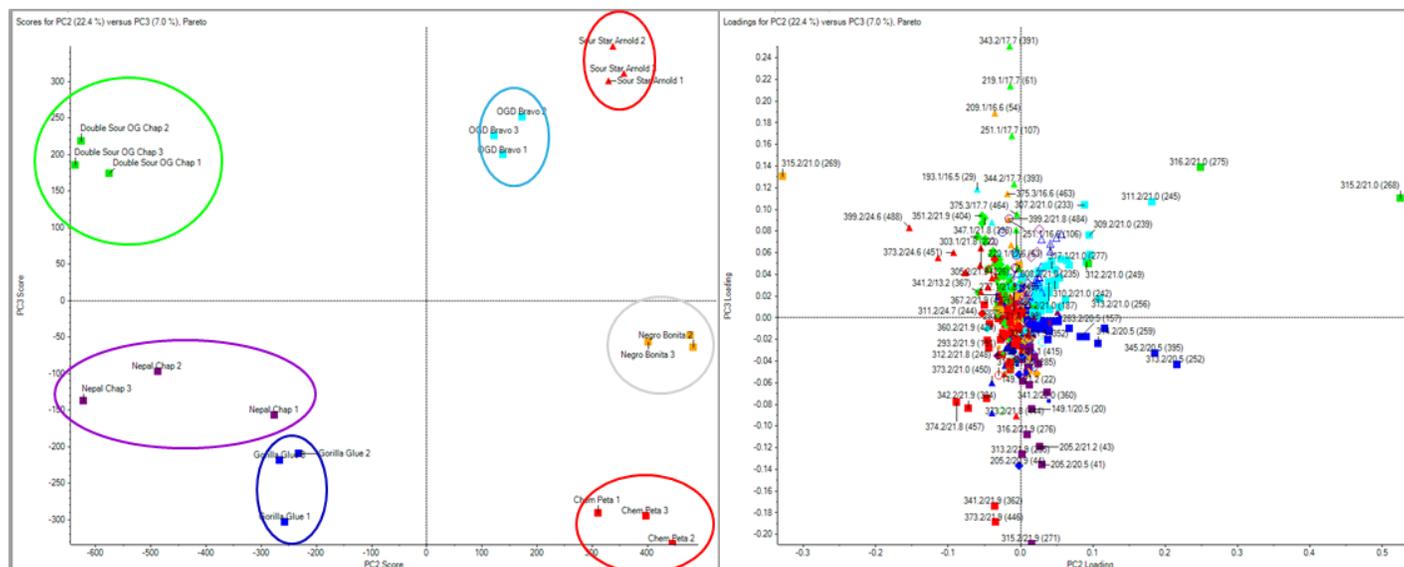
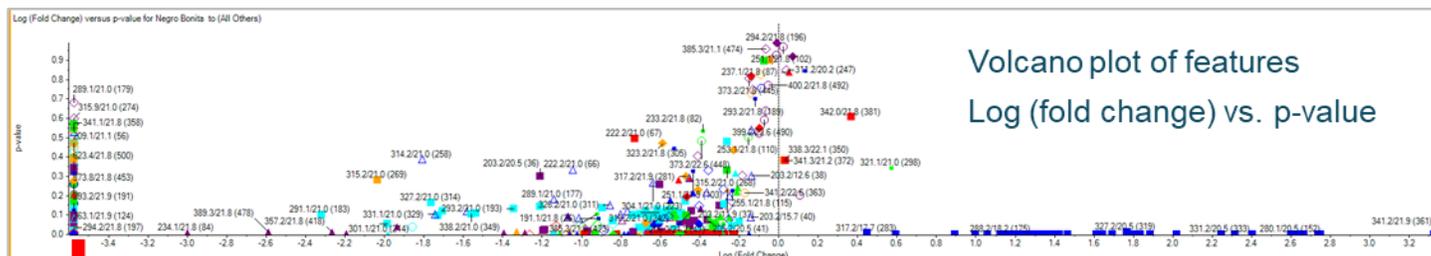


Figure 4: (Left) PCA Score Distribution and (Right) PCA Loadings Plot for the Seven Cannabis Strains Analyzed. This data visualization tool is designed for scientists who wish to interpret their data in terms of sample groupings and apply statistics in order to gain valuable insight into any trends within their mass spectral data. MarkerView Software is unique in that SCIEX users can explore statistical correlations with direct connections back to the raw data. This allows users to find meaningful relationships quickly.

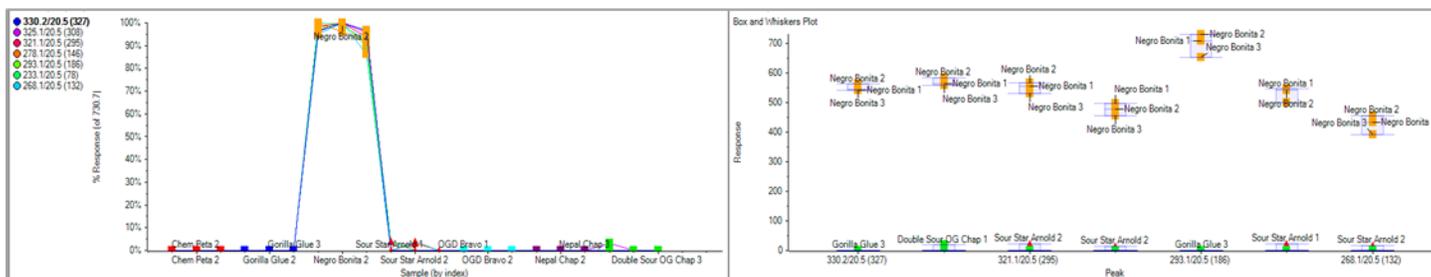


Volcano plot of features

Log (fold change) vs. p-value

Lowest p-value and greatest fold change

Features which are most distinguishing between the two groups compared in the t-Test



Example: Negro Bonita vs. All Others

Figure 5: T-test Analysis Comparing Features in Negro Bonita Strain Against All Other *Cannabis* Samples. (Top) Volcano plot highlighting features with the greatest fold change and statistical significance. (Bottom) Area distribution of selected features as a function of (left) sample and (right) feature. Here, it is easy to determine that the selected features are up-regulated in the Negro Bonita strain.

Peaks of Interest: Identifying the Unique Features

These tools can be used to create a target list of compounds to identify using the Analytics tools (examples in Table 1). This is an advantageous approach, as only the features that differ between samples are processed for identification, rather than attempting to process, search, and identify thousands of features, most of which do not represent differences or unique characteristics between the strains.

Utilizing the MS/MS spectral library and the Analytics tools in SCIEX OS software, some tentative structural identifications are proposed for some of the unique components of different *Cannabis* strains (Figure 6). Several factors can be assessed for the quality of the spectral match, including mass error of the MS1 data versus the accurate mass of the candidate structure and fit score of the empirical MS/MS spectrum to that in the database. Qualitative analysis rules can be set by the user in the SCIEX OS Software in order to automatically display the best potential matches.

In those instances where the MS/MS library does not produce a candidate match, the candidate empirical formula (produced by the FormulaFinder using the experimentally collected accurate

mass of the precursor ion) can be searched against the ChemSpider database, which will produce a list of candidate structures in the database which match that formula. *In silico* MS/MS spectra overlaid with the experimentally collected spectrum can suggest some potential candidate identities. These functionalities are all accessible within the Analytics module of SCIEX OS Software and represent the final stage of the nontargeted workflow. Figure 7 shows an example of an *m/z* feature with an experimentally collected MS/MS spectrum that did not produce any library hits; however, the likely empirical formula ($C_{22}H_{30}O_3$) has many structures in the ChemSpider database and one of them (Myrsinoic Acid) is a potential candidate based on the predicted MS/MS fragmentation pattern. It is important to note that without an analytical reference standard, it is impossible to confidently confirm any structural identification.

Summary

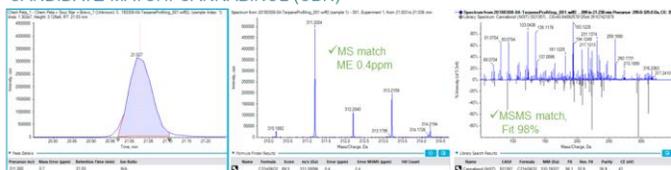
A nontargeted approach is described which utilizes advanced software and statistical data processing of nontargeted high resolution accurate mass spectrometric data in order to assess chemical signature differences between unknown *Cannabis*

strains. Some potential feature markers for different strains can be identified using high resolution mass spectra, MS/MS libraries, and SCIEX OS software tools.

Table 1. Some Example Peaks of Interest. Features (m/z and RT pairs) are identified as uniquely upregulated (+) or downregulated (-) in different *Cannabis* strains. This Peaks of Interest table can be built up by probing the statistical analyses from MarkerView Software, then imported to Analytics for searching against spectral libraries.

Peak of Interest	m/z	RT
(-) <i>Negro Bonita_1</i>	341.2107	21.86
(-) <i>Nepal_1</i>	219.1014	17.67
(+) <i>Chem Peta + Sour Star + Bravo_1</i>	311.2002	21
(+) <i>Negro Bonita_1</i>	313.1794	20.52
(+) <i>Sour Star + Double Sour_1</i>	375.2528	16.56

PEAK OF INTEREST: (+) IN CHEM PETA, SOUR STAR, OG BRAVO (M/Z 311.2003)
CANDIDATE MATCH: CANNABINOL (CBN)



PEAK OF INTEREST: (-) IN NEPAL (M/Z 219.1015)
CANDIDATE MATCH: XANTHORRIZOL

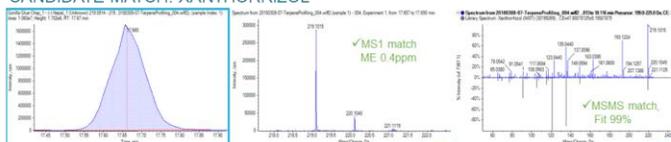


Figure 6. Candidate Structure Identifications for Some Peaks of Interest. Cannabinol, a minor, non-psychoactive cannabinoid, was tentatively identified as being upregulated in Chem Peta, Sour Star, and OG Bravo versus the other strains. Xanthorizol, a sesquiterpenoid, was tentatively identified as being a unique feature in the Nepal strain.

References

1. P. Henry. (2017) *Cannabis* chemovar classification: terpenes hyper-classes and targeted genetic markers for accurate discrimination of flavours and effects. *PeerJ Preprints*.
2. Quantitation of Oregon List of Pesticides and Cannabinoids in *Cannabis* Matrices by LC-MS/MS. SCIEX Technical Note RUO-MKT-02-6729-B.
3. Comprehensive *Cannabis* Analysis: Pesticides, Aflatoxins, Terpenes, and High Linear Dynamic Range Potency from One Extract Using One Column and One Solvent System. SCIEX Technical Note RUO-MKT-02-7218-A.
4. Elzinga S., Fishedick J., Podkolinski R. and J. Raber (2015) Cannabinoids and Terpenes as Chemotaxonomic Markers in *Cannabis*. *Natural Products Chemistry & Research* **3:4**, 181.
5. Improved Data Quality Using Variable Q1 Window Widths in SWATH® Acquisition. SCIEX Application Note RUO-MKT-02-2879-C.
6. A. Schreiber, N. Pace. (2010) Identifying Unexpected Environmental Contaminants with High-Resolution, Accurate Mass LC-MS-MS. *LCGC Chromatography Online*.

PEAK OF INTEREST: (-) IN NEPAL (M/Z 343.2266)

Index	Sample Name	Component Name	Precursor Mass	Area Ratio of...	Mass Error...	Library Confidence	Found At Mass	Mass Error (pp...)	Library Hit	Library Score	Formula Finder Results	Formula Finder Sc...
6	MeOH Blank	(-) Nepal_3	343.227	N/A			N/A	N/A		N/A		N/A
439	Chem Peta_1	(-) Nepal_3	343.227	N/A	✓	●	343.2268	0.5	No Match	0.0	C22H30O3	99.125
872	Gonilla Glue Chap_1	(-) Nepal_3	343.227	N/A	✓	●	343.2265	-0.3	No Match	0.0	C22H30O3	87.331
1305	Negro Bonita_1	(-) Nepal_3	343.227	N/A	✓	●	343.2264	-0.5	No Match	0.0	C22H30O3	90.203
1738	Sour Star Arnold_1	(-) Nepal_3	343.227	N/A	✓	●	343.2266	0.0	No Match	0.0	C22H30O3	95.001
2171	OGD Bravo_1	(-) Nepal_3	343.227	N/A	✓	●	343.2265	-0.3	No Match	0.0	C22H30O3	88.962
2604	Nepal Chap_1	(-) Nepal_3	343.227	N/A	✓	●	343.2267	0.4	No Match	0.0	C22H30O3	97.504
3037	Double Sour OG Chap_1	(-) Nepal_3	343.227	N/A	✓	●	343.2266	-0.1	No Match	0.0	C22H30O3	93.956

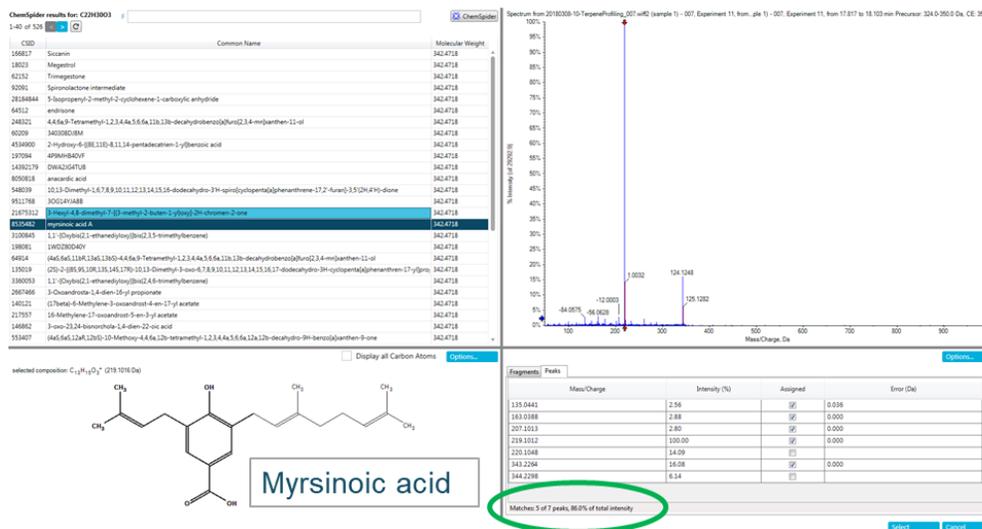
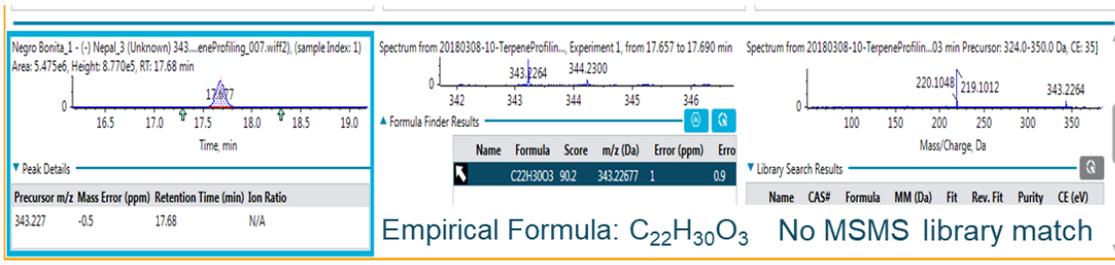


Figure 7. Using ChemSpider for Tentative Identification. One of the peaks of interest, when imported into the SCIEX OS Analytics module as a targeted component, did not have an MS/MS spectral match in the library. However, the m/z value did produce a candidate formula, which could be searched against the ChemSpider database from within the software interface. One structure in ChemSpider with this formula had a predicted fragmentation pattern which closely matched the acquired MS/MS spectrum. This potential candidate identification of Myrsinoic Acid remains unconfirmed without a reference standard for MS/MS fragmentation as well as retention time confirmation. This feature was uniquely downregulated in Nepal versus the other strains.

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to <https://sciex.com/diagnostics>. All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries.

© 2020 DH Tech. Dev. Pte. Ltd. AB SCIEX™ is being used under license. RUO-MKT-02-8085-A.