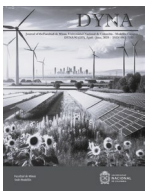
DYNA
http://dyna.medellin.unal.edu.co/



UNIVERSIDAD NACIONAL DE COLOMBIA

MetS: an interactive application to identify metabolites in liquid chromatography-mass spectrometry experiments with data independent acquisition

Daniel Estevan Garcia-Niño [a], Liliana Lopez-Kleine [b] & Federico Roda-Fornaguera [c]

[a] Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Departamento de Matemáticas, Bogotá, Colombia. dgarcian@unal.edu.co
[b] Universidad Nacional de Colombia, Sede Bogotá Departamento de Estadística, Facultad de Ciencias, Bogotá, Colombia. llopezk@unal.edu.co
[c] Universidad Nacional de Colombia, Sede Bogotá, Max Planck Tandem Group Leader, Bogotá, Colombia. frodaf@unal.edu.co

Received: May 17th, 2024. Received in revised form: October 2nd, 2024. Accepted: October 22nd, 2024.

Abstract
Liquid chromatography-tandem mass spectrometry (LC-MS/MS) with data-independent acquisition (DIA) enables the detection of metabolites in biological samples. However, identifying metabolites from DIA data remains challenging due to the complexity of the data. This work presents "Metabolomic Search" (MetS), a software application developed to facilitate metabolite identification in DIA experiments. The application supports filtering, correlation analysis, and similarity scoring algorithms to match DIA data to user-provided metabolite mass-to-charge ratios and fragmentation patterns. The graphical user interface is straightforward and intuitive, allowing easy data uploading, parameter configuration, and results exploration. Tests on different Solanaceae samples demonstrated successful identification of target metabolites such as scopolamine. By enabling rapid compound screening, MetS can support metabolomic based research in pharmaceutical, biotechnological, and clinical domains. The availability of this open-source tool could help address the pressing need for metabolite annotation in increasingly prevalent DIA experiments.

Keywords: application; metabolomics; metabolite identification; LC-MS/MS; data-independent acquisition.

MetS: una aplicación interactiva para identificar metabolitos en experimentos de cromatografía líquida-espectrometría de masas con adquisición independiente de datos

Resumen
La cromatografía líquida-espectrometría de masas en tándem (LC-MS/MS) con adquisición independiente de datos (DIA) permite la detección de metabolitos en muestras biológicas. Sin embargo, identificar metabolitos a partir de datos DIA sigue siendo un desafío debido a la complejidad de los datos. Este trabajo presenta "Metabolomic Search" (MetS), una aplicación de software desarrollada para facilitar la identificación de metabolitos en experimentos DIA. La aplicación admite algoritmos de filtrado, análisis de correlación y puntuación de similitud para hacer coincidir los datos de DIA con las relaciones masa-carga de metabolitos y los patrones de fragmentación proporcionados por el usuario. La interfaz gráfica de usuario es sencilla e intuitiva y permite cargar datos, configurar parámetros y explorar resultados fácilmente. Las pruebas en diferentes muestras de Solanaceae demostraron una identificación exitosa de metabolitos objetivo como la escopolamina. Al permitir la detección rápida de compuestos, MetS puede respaldar la investigación basada en metabolómica en los dominios farmacéutico, biotecnológico y clínico. La disponibilidad de esta herramienta de código abierto podría ayudar a abordar la necesidad apremiante de anotación de metabolitos en experimentos DIA cada vez más frecuentes.

Palabras clave: aplicación; metabolómica; identificación de metabolitos; LC-MS/MS; adquisición independiente de datos.

1. Introduction

At present, technological advancements allow the acquisition of high throughput molecular biology information from hundreds of samples, leading to the creation of the so-called omic sciences. Metabolomics is part of these emerging sciences and has enabled further research into the metabolism of living organisms. More specifically, metabolomics is the

How to cite: Garcia-Niño, D.E., Lopez-Kleine, L., and Roda-Fornaguera, F., MetS: an interactive application to identify metabolites in Liquid chromatography-mass spectrometry experiments with data independent acquisition. DYNA, 92(237), pp. 9-15, April - June, 2025.

systematic identification and quantification of all metabolites in a biological sample [1]. These metabolites are molecules that participate in or result from biochemical reactions in cells. Metabolomics is applied to a variety of research fields, such as microbiology, botany, pharmacology, toxicology, functional genomics, and medicine [1]. Furthermore, its study can help discover biomarkers, elucidate physiological mechanisms, evaluate drug efficacy and toxicity, and understand the production of medicinal compounds [1].

Metabolomics employs advanced analytical techniques such as liquid chromatography (LC) in conjunction with tandem mass spectrometry (MS/MS) to separate, detect, and characterize thousands of metabolites. Liquid chromatography is a technique that allows for the separation and purification of the chemical components of a mixture (considering their physicochemical properties) through the use of a chromatographic column and two liquid phases [2]. Tandem mass spectrometry enables the determination of molecular structure, physical and chemical properties, and reaction dynamics of compounds by ionizing, fragmenting, and separating them by their mass-to-charge ratio (m/z) in two or more mass analyzers [3].

From mass spectrometry, there are two types of acquisition modes, 'Data-dependent acquisition' (DDA) and 'Data-independent acquisition' (DIA). In DDA, the most intense precursor fragments are selected to isolate and fragment them again, generating simple MS2 spectra; it has high sensitivity and selectivity but is prone to sampling bias and variability between experiments. In DIA, precursors are isolated in predefined mass windows and all precursors in each window are fragmented, generating complex MS2 spectra; it has good reproducibility and precision but requires a spectral library or advanced bioinformatics tools to identify and quantify compounds [4].

Recent reviews have highlighted the challenges in processing and interpreting data from untargeted metabolomics experiments based on mass spectrometry [5]. These reviews emphasize the need for specialized computational methods, particularly for data-independent acquisition (DIA) approaches. A comprehensive review of metabolomics-focused data mining techniques [6] discusses various analytical tools, including those designed for DIA data.

One such method, SWATH (Sequential Window Acquisition of all Theoretical Mass Spectra), has garnered attention for its DIA capabilities. Several tools have been developed to analyze SWATH data, such as OpenSWATH [7], initially created for proteomics but adaptable to metabolomics. Other software packages like XCMS [8], MS-DIAL [9], MetFrag [10], and MS-FINDER [11] have been employed for processing untargeted metabolomics data, metabolite annotation, and identification of unknown metabolites in mass spectrometry data.

A review of annotation tools for untargeted metabolomics [12] introduces MetDNA, a tool capable of processing DIA data. However, MetDNA is limited to annotation rather than identification in samples and requires previously annotated metabolites to create similarity networks.

Specific software tools have been developed to address various aspects of metabolomics data analysis. MetPathwayMap [13] organizes mass spectrometry peaks from untargeted metabolomics experiments and maps them to MetaCyc metabolic pathways. While this tool provides valuable pathway information, it relies on external software and is not explicitly designed for DIA data.

Another tool, MetMiner [14], is specifically designed for large-scale metabolomic data analysis in plants, utilizing LC-MS/MS experimental data. Although MetMiner focuses on processing, annotation, and statistical analysis of metabolomic data using plant-specific databases to improve annotation accuracy, it is not explicitly mentioned to be compatible with DIA data.

Despite these advancements, significant challenges remain in the field of metabolomics data analysis. Many of these tools are not specifically designed for DIA data, while others are restricted to specific data acquisition methods like SWATH. Some tools focus solely on annotation without providing metabolite identification in samples. These limitations significantly hinder our ability to identify and study novel compounds, highlighting the need for more versatile and comprehensive analytical tools in metabolomics research.

Within the Max Planck Tandem Group in Evolutionary Genomics of Specialized Metabolism (GEME) research group at the Universidad Nacional de Colombia in Bogotá, information was collected from LC-MS/MS experiments with DIA to identify specific classes of specialized secondary metabolites in a large set of species from the Solanaceae plant family. Currently, the collaborators of the Max Planck Institute of Molecular Plant Physiology (MPIMPP) use mainly manual inspection for identifying metabolites in DIA data. There is interest in standardizing, facilitating and optimizing this search. Especially increasing the precision and efficiency of identifications of novel compounds.

This purpose is accomplished by MetS, a computer tool that improves the process of identifying metabolites in LC-MS/MS experiments with DIA, so that it can be used by any researcher (thanks to an intuitive user interface) and is computationally efficient. The rationale that allowed the final tool to be successful is based on three main multidisciplinary aspects: 1) Comprehension of the biological problem and experiments, 2) statistical exploration to find methods to solve the problem, and finally, computer science thinking that could depict the conceptual steps in a computational tool. Given the complexity of the topic, the work was limited only to tabulated data over samples (like obtained with RefinerMS software, etc).

2. Design

The R programming language was used due to its advantages in the analysis of biological data. In addition, a user interface was created with the Shiny library that allows the application to run in any web browser, regardless of the operating system. To make the application available, a GitHub repository containing installation instructions was created: https://github.com/EstevanGN/Metabolomic-search.

The development of the application required two distinct programming paradigms. The functional paradigm was integral to the design of the Metabolomic Search (MetS)

methodology, particularly in the filtering and transformation of datasets. By implementing these operations as functions, the code became more compact, readable, and self-referential [15], allowing for flexibility in modifying different datasets without needing to restart the program. In contrast, the object-oriented paradigm was employed to take advantage of language-specific methods and to structure the user interface [16]. This combination of paradigms enhances the maintainability of the codebase and ensures that both the methodological core and the interface are easily extensible.

The structure of the application was divided into three sections: Data upload, Metabolomic search and More tools.

3. Import files

To achieve the objectives of this project, a methodology based only on tabulated data for all samples is implemented in MetS. This data set has the following configuration:

- Retention Time (RT): In chromatography, it is the time required for an analyte to migrate from the column to the detector after injection. In a sample, compounds with different physicochemical properties will arrive at different retention times. This time is generally measured in seconds or minutes.
- Intensity: In chromatography, it refers to the electrical signal generated when the ions of a sample are detected after their separation in the chromatographic column. It represents the intensity of the ion signal relative to the RT.
- m/z: In mass spectrometry, it is a measure that represents the ratio between the mass of an ion and its charge (positive or negative). The m/z value is generally presented as a decimal number, and it is used to generate a mass spectrum, which is a graphical representation of the intensity of the detected ions against their respective m/z ratios. These m/z values generate the so-called fragmentation pattern of the compounds.
- Sample: It is the identification of the organic sample that was run in the LC-MS/MS experiment with DIA in metabolomics.

The file is read in csv format and can be configured with the application parameters for correct operation and a list of 'warnings' to achieve correct loading. It is also possible to upload an additional csv file to do an automatic search on a list of compounds additional to individual searches.

4. Processing data

To identify metabolites in this type of data set, we rely on two fundamental aspects: the user knows the type of metabolite to search for, and the metabolome is similar among similar samples. That is:

- The m/z of each metabolite to be searched is known. Alternatively, the chemical formula of the metabolite can be used thanks to the added tool that calculates the m/z. The user also knows the fragmentation pattern of the molecule, when possible.
- If a metabolite exists in one sample, it will also be present in other similar samples, maintaining the fragmentation pattern given by mass spectrometry, creating a footprint in the data.

Therefore, the methodology for finding metabolites is reduced to the following steps (see Fig. 1):

1. Input:
   a. Provide an m/z.
   b. Provide a fragmentation pattern (optional). This is a theoretical fragmentation pattern found in the literature for the molecule in question. Given in m/z values.
   c. Provide an RT (optional). The RT may vary depending on the chromatogram configuration and its use represents a strong restriction on MetS. It is recommended to use it with caution only as long as the LC-MS/MS data and the metabolite input come from a practically identical experimental design.

2. Metabolomic search:
   a. With the input m/z, set a search interval, allowing variability across samples.
   b. The previous m/z interval has associated certain RT values. Generate a list with these possible RT values.
   c. For each RT in the above list, set a search interval again, allowing for variability in samples.
   d. In the RT interval, find the input m/z (previous steps ensure this). Together, the RT interval and the input m/z allow the generation of a possible fragmentation pattern for all m/z values that correlate highly (Pearson linear correlation is used given the large number of samples that are usually analyzed) across intensities in samples with the main m/z.
   e. The aforementioned possible fragmentation pattern is nothing more than a subtable of the data, presenting a similar RT and several m/z values.
   f. For each of these subtables, generate an approximation score to the input fragmentation pattern (if any), or an approximation score to the input m/z in another way.
   g. Upon completion, there will be a subtable and an approximation score for each RT.

3. Output:
   a. Best result: the subtable with the lowest approximation score.
   b. All results: a table of all possible subtables for each RT.
   c. Intensity graph: it is possible to interactively select the rows of the subtables to see the intensity of the m/z on each sample.

Each result table can be copied or downloaded in csv, xlsx, and pdf formats for further filters. Finally, this process can be visually understood in Fig. 2.

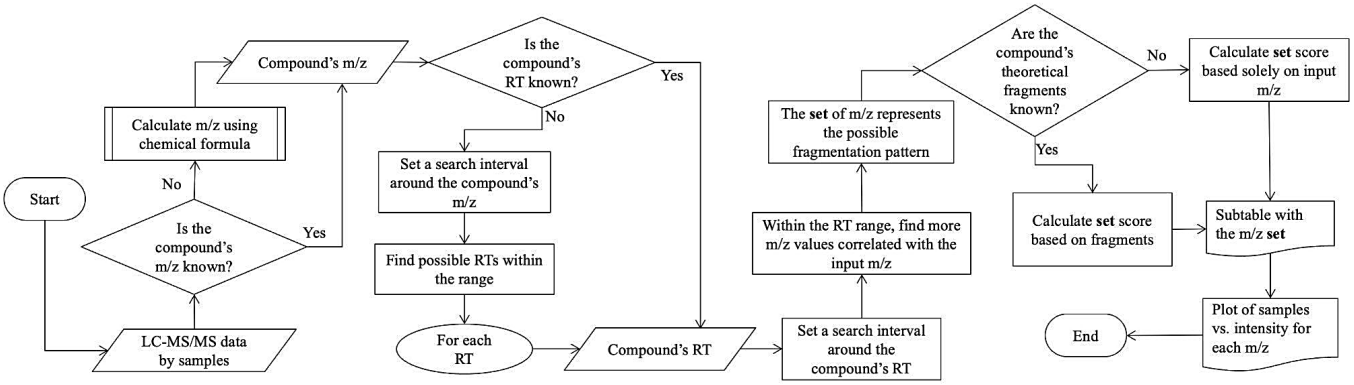


Figure 1. Flowchart for the MetS algorithm. The exact compound's RT may not be known for various reasons (LC-MS/MS experimental design was different at some stage) and a general search for the compound may be desired. For each possible RT found in the algorithm, 'Subtable with the m/z set' and 'Plot of samples vs intensity for each m/z' will be returned.
Source: Own elaboration.

5. Results

Thanks to the collective work carried out in the GEME group, MetS could be tested with LC-MS/MS data in Solanaceae samples and metabolites already known from this plant family. In Fig. 3, we see the result of finding the chemical compound 'scopolamine' in leaf samples from multiple species from the Solanaceae family. The configuration was as follows:

- Search only with m/z value.
- m/z of scopolamine: 304.15.
- m/z interval of search: 0.01.
- Fragments: 138.09 and 156.10.
- Correlation: 0.95.
- RT interval of search: 0.01.

The best result of the search is a subtable with seven m/z values at an approximate RT of 4.97. Within this subtable is the fragmentation pattern for scopolamine in rows 2, 3, and 6. Again in Fig. 3, the intensity graph for each sample in the selected fragmentation pattern is shown (rows 2, 3, and 6). Therefore, the method allows identifying metabolites with a certain level of error given by the variability of the samples, the intervals of the parameters, and the chosen correlation level.

6. Discussion

With the MetS application we have narrowed down the search due to the search mechanism and its options in comparison to other pipelines to analyze DIA data. Moreover, it is important to note that other methodologies proposed earlier [17-19] use complementary information, which is not exclusive to the LC-MS/MS experiment with DIA in metabolomics.

In [17], the authors developed a workflow named DaDIA, which combines DIA analysis of biological samples with DDA analysis of quality control samples. The DIA analysis provides high coverage of metabolic features and MS/MS spectra, and the DDA analysis generates high-quality MS/MS spectra to enhance the confidence of metabolite annotation. The authors also created an R package, DaDIA.R, to automate data processing and metabolite annotation from DaDIA data. The DaDIA workflow was applied to a study that compared metabolic alteration in the plasma of leukemia patients before and after chemotherapy. The results demonstrated that the DaDIA workflow can detect and annotate approximately four times more significantly altered metabolites than the conventional DDA workflow.

In [18], DDA and DIA data were compared in metabolomics. For the analysis of DIA-type data, the MS-DIAL software was used, which allows spectral deconvolution and similarity comparison with a reference spectral library. The quality of the MS/MS spectra obtained by the dot product method was evaluated, which measures the degree of match between the evaluated spectra and the reference ones. It was observed that the dot products obtained with the DIA method were generally lower than those obtained with the DDA method, due to the greater complexity and lower purity of the DIA spectra. However, the DIA method allowed obtaining structural information from all the detected compounds, which facilitates the identification of unknown metabolites.

Specialized software has been created to extract information from DIA in the same MS/MS experiment with different energy settings as in [19], where DIA-type data were analyzed using a workflow called MetaboMSDIA, which uses the R programming language and several open-source packages to extract, align, and annotate MS2 spectra in different acquisition channels with different collision energies. The workflow allows identifying metabolites by searching for their MS2 spectra in public or private databases, or by searching for characteristic fragmentation patterns of metabolite families.

There are approaches that do not use DDA data as support and instead base their search on statistical or machine learning techniques. Some examples are MetaboAnnotator, CANOPUS, NPClassifier, and MS2DeepScore [20]. GNPS is another tool that uses molecular networks to annotate compounds in both DIA and DDA [21]. The functionality of these tools is based, almost exclusively, on the structural comparison of fragmentation patterns with metabolite databases.

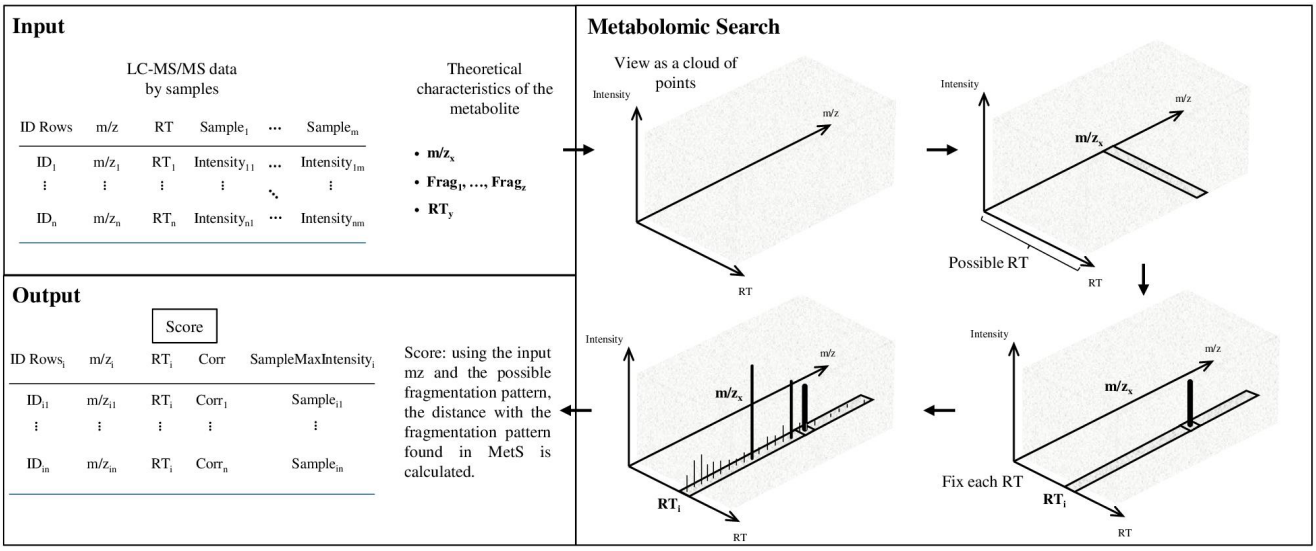


Figure 2. Internal methodology for the MetS algorithm. The input should contain the data from the scientific experiment and the characteristics of the metabolite to be found. LC-MS/MS data can be viewed as a three-dimensional point cloud. The m/z of the input is used to find possible RTs within a fixed interval. For each RT, the recommended fragmentation pattern is constructed, taking into account only the data that correlates with the m/z of the input. Each RT returns a table referring to the fragmentation pattern and a proximity score to the characteristics of the metabolite. This process can be extended to a list of compounds.
Source: Own elaboration.

Thus, all the software proposed to date use comparisons with complementary information, such as supporting DDA data or metabolite databases, to produce results. However, the complexity of the chemical structure of metabolites and the way they fragment [22] prevent these tools from having a real application in DIA type data. It is here, where MetS offers a novel solution by proposing possible fragmentation patterns, controlled by statistical theory, for each of the chemical compounds to be identified.

Bioinformatics tools are increasingly important in modern biology, challenges related to usability, implementation, and user education remain [23]. Moreover, accessibility and usability of bioinformatics software tools is crucial for research [24]. Softwares like MetS overcome the principal challenges of installability and long-term availability assuring reproducibility of scientific findings due to its free and easy access and because it is the result of a collaborative and interdisciplinary effort.

7. Conclusions and recommendations

The application exclusively relies on information gathered through DIA and data analysis techniques, eliminating the necessity for supplementary experimental designs. This distinctive approach offers a novel utility and significance unparalleled by other methods addressing similar issues. MetS, due to its open-access code and construction paradigms allows ongoing development and enhancement. The user interface is intuitive for researchers and can handle user errors. The algorithm relies solely on filters, searches, and correlations, allowing for polynomial time computational complexity. This application works for all LC-MS/MS experiments with DIA type data and can be used by researchers from different areas of knowledge. Tests with metabolomic data from the Solanaceae plant family were successful, as the application found compounds characteristic of this family of plants, and demonstrated that MetS is robust against typographical errors.

This is a user friendly software that can be used by researchers from different areas. However, it is highly recommended to conduct research in collaboration with chemists or technicians specializing in LC-MS/MS machinery. This interdisciplinary approach can foster a more comprehensive understanding and facilitate the development of more robust solutions. The application could be further enhanced to perform not only the identification of metabolites but also their annotation, leveraging the power of artificial intelligence. This would provide a more detailed analysis and could potentially uncover new insights in the metabolomic data. In addition, incorporating an error calculation feature into the application could provide a measure of certainty regarding the results allowing researchers to gauge the reliability of the findings and make informed decisions about their research.

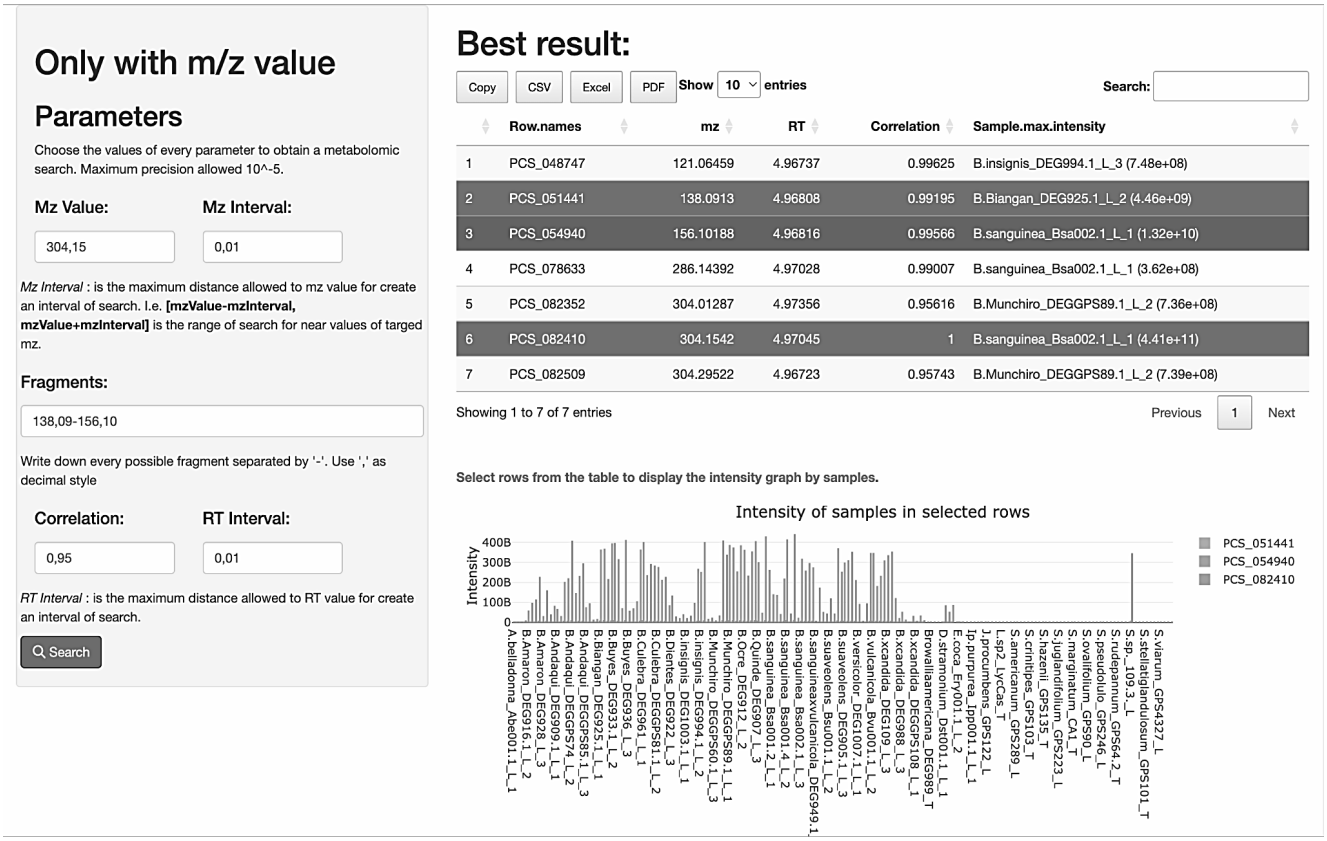


Figure 3. MetS results from LC-MS/MS experiment in Solanaceae samples searching for the metabolite scopolamine.
Source: Own elaboration.

8. Funding

Agreement 566 of 2014 between Universidad Nacional de Colombia (https://unal.edu.co/) and Colciencias (Now Minciencias https://minciencias.gov.co/).

References

[1] Yang, Q. et al., Metabolomics biotechnology, applications, and future trends: a systematic review. RSC Adv., 9(64), pp. 37245–37257, 2019. DOI: https://doi.org/10.1039/C9RA06697G

[2] Coskun, O., Separation Tecniques: Chromatography. North Clin Istanbul, 2016. DOI: https://doi.org/10.14744/nci.2016.32757

[3] Amorim-Madeira, P.J., and Helena, M., Applications of tandem mass spectrometry: from structural analysis to fundamental studies. Tandem Mass Spectrometry - Applications and Principles, J. Prasain, Ed., InTech, 2012. DOI: https://doi.org/10.5772/31736

[4] Li, J., Smith, L.S., and Zhu, H.-J., Data-independent acquisition (DIA): an emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. Drug Discovery Today: Technologies, 39, pp. 49–56, Dec. 2021. https://doi.org/10.1016/j.ddtec.2021.06.006

[5] Perez De Souza L. & Fernie, A. R., Computational methods for processing and interpreting mass spectrometry-based metabolomics. Essays in Biochemistry, 68(1), pp. 5–13, 2024. DOI: https://doi.org/10.1042/EBC20230019

[6] Shen, S., Zhan, C., Yang, C., Fernie, A.R., and Luo, J., Metabolomics-centered mining of plant metabolic diversity and function: past decade and future perspectives. Molecular Plant, 16(1), pp. 43–63, 2023. DOI: https://doi.org/10.1016/j.molp.2022.09.007

[7] Röst H.L. et al., OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol, 32(3), pp. 219–223, 2014. DOI: https://doi.org/10.1038/nbt.2841

[8] Mahieu, N.G., Genenbacher, J.L., and Patti, G.J., A roadmap for the XCMS family of software solutions in metabolomics. Current Opinion in Chemical Biology, 30, pp. 87–93, 2016. DOI: https://doi.org/10.1016/j.cbpa.2015.11.009

[9] Tsugawa, H. et al., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods, 12(6), pp. 523–526, 2015. DOI: https://doi.org/10.1038/nmeth.3393

[10] Ruttkies, C., Neumann, S., and Posch, S., Improving MetFrag with statistical learning of fragment annotations. BMC Bioinformatics, 20(1), art. 376, 2019. DOI: https://doi.org/10.1186/s12859-019-2954-7

[11] Lai, Z. et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. Nat Methods, 15(1), pp. 53–56, 2018. DOI: https://doi.org/10.1038/nmeth.4512

[12] Cai, Y., Zhou, Z., and Zhu, Z.-J., Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics. TrAC Trends in Analytical Chemistry, 158, art. 116903, 2023. DOI: https://doi.org/10.1016/j.trac.2022.116903

[13] Moghe, G., and Strickler, S., metaPathwayMap: a tool to predict metabolic pathway neighborhoods from structural classes of untargeted metabolomics peaks. bioRxiv. Art. 48433, 2022. DOI: https://doi.org/10.1101/2022.03.15.484337

[14] Wang, X. et al. MetMiner: a user-friendly pipeline for large-scale plant metabolomics data analysis. JIPB, art. 13774, 2024. DOI: https://doi.org/10.1111/jipb.13774

[15] Abajirov, D., Does functional programming improve software quality?: an empirical analysis of open-source projects on GitHub. Universität Stuttgart. 2021, DOI: https://doi.org/10.18419/OPUS-11582

[16] Chambers, J.M., Object-Oriented Programming, Functional Programming and R. Statist. Sci., 29(2), 2014. DOI: https://doi.org/10.1214/13-STS452
[17] Guo, J., Yu, H., Xing, S., and Huan, T., Addressing big data challenges in mass spectrometry-based metabolomics. Chem. Commun., 58(72), pp. 9979–9990, 2022. DOI: https://doi.org/10.1039/D2CC03598G
[18] Barbier Saint Hilaire, P. et al. Comparative evaluation of data dependent and data independent acquisition workflows implemented on an orbitrap fusion for untargeted metabolomics. Metabolites, 10(4), art. 158, 2020. DOI: https://doi.org/10.3390/metabo10040158
[19] Ledesma-Escobar, C.A., Priego-Capote, F., and Calderón-Santiago, M., MetaboMSDIA: a tool for implementing data-independent acquisition in metabolomic-based mass spectrometry analysis. Analytica Chimica Acta, 1266, art. 341308, 2023. DOI: https://doi.org/10.1016/j.aca.2023.341308
[20] Ebbels, T.M.D. et al. Recent advances in mass spectrometry-based computational metabolomics. Current Opinion in Chemical Biology, 74, art. 102288, 2023. DOI: https://doi.org/10.1016/j.cbpa.2023.102288
[21] Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol, 34(8), pp. 828–837, 2016. https://doi.org/10.1038/nbt.3597
[22] Collins, S.L., Koo, I., Peters, J.M., Smith, P.B., and Patterson, A.D., Current challenges and recent developments in mass spectrometry–based metabolomics. Annual Rev. Anal. Chem., 14(1), pp. 467–487, 2021. DOI: https://doi.org/10.1146/annurev-anchem-091620-015205
[23] Ahmad. Software tools in Bioinformatics: a survey on the importance and Issues faced in implementation. Global Engineers and Technologists Review, 3, pp. 6–10, 2013
[24] Mangul, S. et al. Challenges and recommendations to improve the installability and archival stability of omics computational tools. PLoS Biol, 17(6), art. e3000333, 2019. DOI: https://doi.org/10.1371/journal.pbio.3000333

D.E. Garcia-Niño, received his BSc. in Statistics in 2022, and his BSc. in Computer Science in 2024, both degrees from the Universidad Nacional de Colombia, Bogotá Campus, Colombia. Currently, he is a PhD student in biotechnology at the same university. His research interests lie in statistical methods and computational algorithms focused on the biology and health sectors. He has worked at a multinational pharmaceutical company, research centers in health technology assessments, and research groups in omics.
ORCID: 0000-0001-6865-2468

L. Lopez-Kleine, is a BSc. in Biologist from the Universidad Nacional de Colombia in Bogotá Campus, with a MSc. in Evolution, Ecology and Biometrics of the University of Lyon I, and a PhD in Applied Statistics from the AgroParisTech in France. Her research concentrates in developing and applying methods for statistical genomics and biological network construction and analysis. She teaches biostatistics for undergraduate and postgraduate students at Universidad Nacional de Colombia where she has worked as associate professor since 2009.
ORCID: 0000-0001-9325-9529

F. Roda-Fornaguera, received the BSc. Biology in 2006, from the Universidad Nacional de Colombia, PhD in Evolutionary Genomics in 2014, from the University of Queensland, and he worked as a postdoc in Harvard University from 2014 to 2019. He is currently the leader of the Max Planck Tandem Group in Evolutionary Genomics of Specialized Metabolism (GEME), at the Universidad Nacional de Colombia, where his team investigates the genomic basis of plant metabolism.
ORCID: 0000-0002-7923-9713