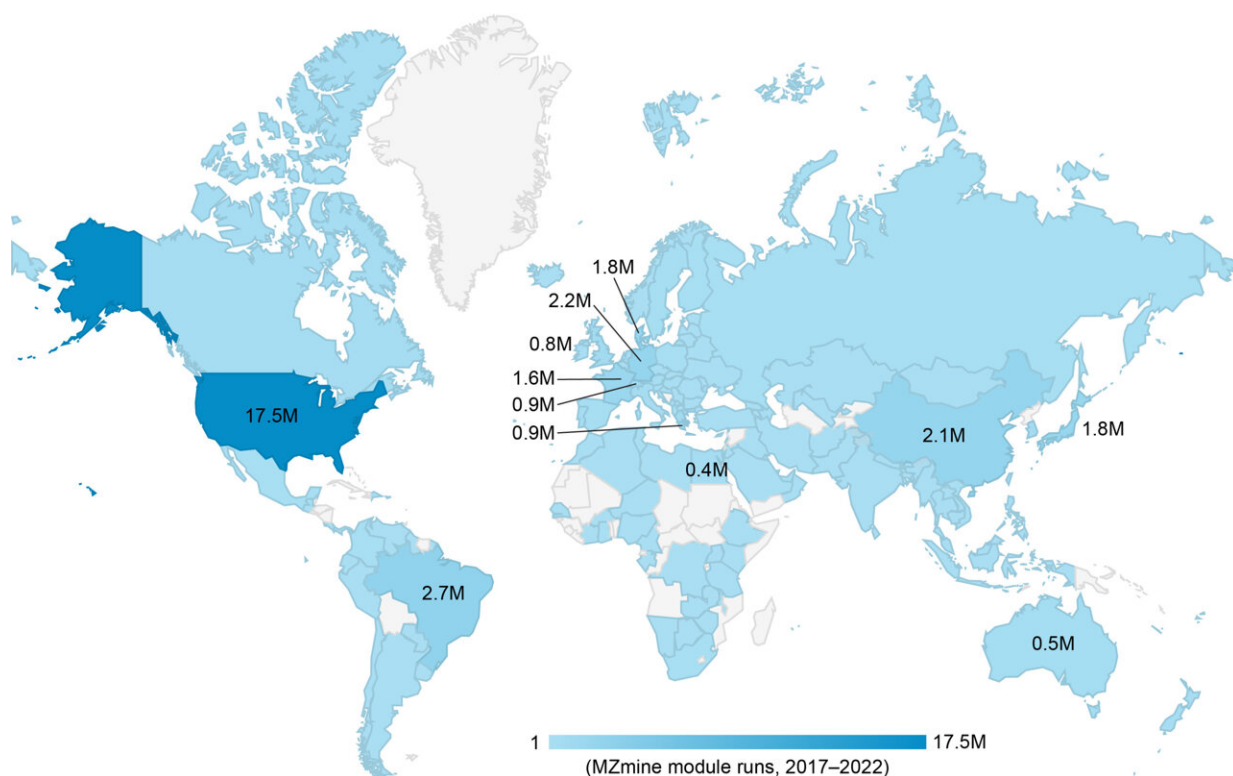# A breakthrough in big data processing helps trace chemicals in complex mixtures

March 8 2023

MZmine module runs. Credit: Robin Schmid / IOCB Prague

An international team of scientists led by Tomáš Pluskal from the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences (IOCB Prague) has introduced a new generation of software that enables scientists to analyze large volumes of data from mass

spectrometry, a technique that separates chemicals by their weights.

The open-source project MZmine provides a new window into the chemical space that surrounds us and lives within. The latest advances in MZmine 3 are now published in a *Nature Biotechnology* paper.

Analytical chemists of the world, unite! This paraphrase could characterize the joint efforts of scientists across the globe, who, using the methods of mass spectrometry, strive to decipher and analyze the chemical composition of complex samples from various origins, especially in biological and clinical studies. Each individual sample can contain hundreds of thousands of different chemical compounds that scientists need to trace, quantify, and identify to understand their impact on human health or their ecological role.

Even relatively small studies result in gigabytes of "raw" data to be processed and interpreted. It is the processing, analysis, and comparison of a multitude of molecular data that constitutes some of the most challenging steps in biochemical analysis today. This is also a major bottleneck that limits the ability of scientists to expand knowledge and come up with exciting new discoveries.

Community-driven development

For this reason, a group of international scientists started in 2005 to develop the open-source software MZmine to aid the analysis of mass spectrometry data. The community developing this software has been co-established by Czech scientist Tomáš Pluskal, who has been coordinating the project almost since its inception and is currently a group leader at IOCB Prague.

"The greatest strength of the MZmine project is the international community of experts that has formed around the project. At

conferences, presentations on MZmine are always well received," says Tomáš Pluskal about the project.
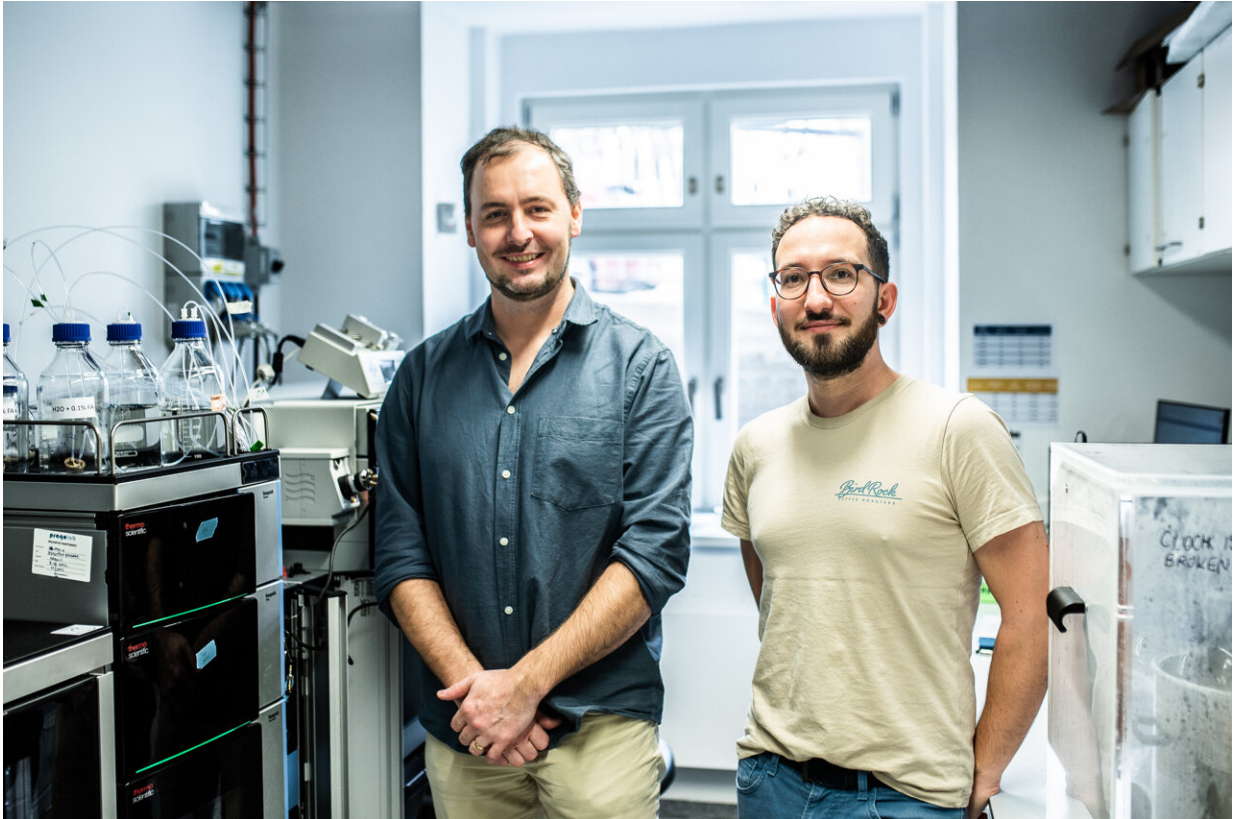
Tomáš Pluskal (IOCB Prague) & Robin Schmid (IOCB Prague & UC San Diego). Credit: IOCB Prague / Tomáš Belloň

Robin Schmid from IOCB Prague and UC San Diego (CA, U.S.), one of the first authors of the paper, adds, "It's fantastic when we meet researchers from other countries for the first time and they tell us that MZmine and our support has saved their Ph.D. thesis or projects. That's the best appreciation one can hope for."

The first version of MZmine has enabled scientists to automate the processing of datasets generated by analytical devices at an unprecedented scale. The second generation of MZmine, released in 2010, made the project more widely known and led to the formation of a worldwide community of researchers using the software and continuing to expand its functions with additional modules and applications. The publication introducing the second generation of MZmine has since collected more than 2,200 citations in scientific articles and the tool itself has been used to process millions of different measurements.

Third generation

The newest MZmine 3 brings several major improvements. Whereas the previous version allowed scientists to analyze hundreds of samples in a matter of days, the new generation makes it possible to process thousands of samples per hour. Besides vastly accelerating data processing, the new version of the software can also be used, for the first time, to link different data types, especially time-resolved and imaging data.

This opens up opportunities for researchers to more easily analyze and interpret complex biological samples. MZmine is a tool to investigate the causes and mechanisms of diseases, detecting useful clinical biomarkers for diagnostics and identifying chemicals in the environment. This includes previously unknown chemical structures, which might prove valuable for the discovery and development of new drugs for medical applications.

The third generation of MZmine was announced in a paper prepared, besides Tomáš Pluskal as the corresponding author, by the first authors Robin Schmid (IOCB Prague and UC San Diego), and Steffen Heuckeroth and Ansgar Korf (both from University of Münster, Germany), joined by over three dozen other contributors from around

the world.

"MZmine has established itself as a trusted tool for mass spectrometry researchers over the past decade. Its modular framework has fostered community participation in the development of the MZmine code, leading to significant advancements featured in the newly released MZmine 3," says Ansgar Korf of University of Münster.

More information: Robin Schmid et al, Integrative analysis of multimodal mass spectrometry data in MZmine 3, *Nature Biotechnology* (2023). DOI: 10.1038/s41587-023-01690-2

Provided by Institute of Organic Chemistry and Biochemistry of the CAS