A plethora of plant molecules provides a ple"flora" of data

January 18 2023

Researchers at the University of Geneva have established a searchable library of spectra and molecules found in a collection of 1,600 plant extracts. This collection was accessed through a collaboration with Pierre Fabre Laboratories. The resulting open resource, published in the journal *GigaScience*, shares both the obtained data and the employed methods. This will be useful for research ranging from drug discovery to the large-scale exploration of plants' chemical diversity.

Plant metabolites play a fundamental role in drug development, as they can display potent biological activities; many existing drugs are used either as pure natural products (for example the anti-malarial artemisinin and anti-cancer drug taxol), or derived from natural products (such as the anti-cancer drugs vinorelbine, brentuximab vedotin).

Pierre Fabre Laboratories have always been keen to share its pharmaceutical experience with academia. For instance, the anti-cancerous semi-synthetic natural product vinorelbine, which is derived from the Madagascar Periwinkle (Catharanthus roseus), was marketed in 1989 as a result of a collaboration with Professor Pierre Potier from the French National Center for Scientific Research (CNRS). Pierre Fabre Laboratories have made plant research a central aspect of its approach since its creation. A collection of plant samples was constituted over the 1998-2015 period with the main goal of finding novel anti-cancer drugs. This collection of botanical samples is one of the largest private plant libraries in the world numbering over 17,000 unique samples, including some rare species and covers a diverse range of botanical families from

all over the world. Since 2015, Pierre Fabre Laboratories have opened access to their private plant samples collection for interested partners.

This new study published in *GigaScience* reports the chemical characterization of circa 10% of the plant extracts in the Pierre Fabre Laboratories collection. This represents an important step towards making the chemical diversity of the full collection accessible to researchers around the world. The researchers at the University of Geneva used high-resolution mass spectrometry in combination with advanced computational pipelines to acquire over two million spectra and associated chemical information, providing valuable insights into the biochemical content of the plant extracts. The mass spectrometry profiles and associated metadata have been shared openly through the MassIVE repository (accession number MSV000087728), and this Data Note provides demonstrations on how to query this extensive curated resource. The Data Note shares both the resulting data and the employed methods, allowing for reproducibility, further exploration of the dataset and improvement of the proposed computational tools and methods. This is an exceptional resource for advancing the field of large-scale chemodiversity digitization. Such fruitful partnership between academia and industry illustrates that the fate of a historical and private collection of plant samples can be changed and that the richness of the associated chemical diversity can be made available to a wider public.

More information: Open and reusable annotated mass spectrometry dataset of a chemodiverse collection of 1,600 plant extracts, *GigaScience* (2023). DOI: 10.1093/gigascience/giac124

Allard P-M & Wolfender, J.-L. (2021). MassIVE MSV000087728—GNPS_PF_plant_extracts_library_dataset_01 [Data set]. MassIVE. doi.org/10.25345/C59J97

Allard Pierre-Marie et al, Supporting data for "Open and re-usable

annotated mass spectrometry dataset of a chemodiverse collection of 1,600 plant extracts.", *GigaScience Database* (2022). DOI: 10.5524/102323

Provided by GigaScience