An Open Software Development-based Ecosystem of R Packages for Metabolomics Data Analysis

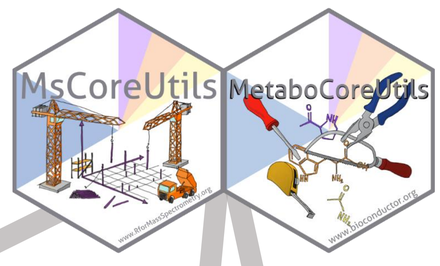
Andrea Vicini[1], Roger Gine[2], Michele Stravs[3,4], Josep M Badia[2], Carolin Huber[5], Liesa Salzer[6], Jan Stanstrup[7], Nir Shahaf[8], Thomas Naake[9], Helge Hecht[10], Steffen Neumann[11], Michael Witting[12], Sebastian Gibb[13], Laurent Gatto[1], Johannes Rainer[14]

[1]Computational Biology and Bioinformatics, de Duve Institute, UCLouvain, Belgium. [2]Department of Electronic Engineering & IISPV, Universitat Rovira i Virgili, Spain. [3]Department of Environmental Chemistry, Switzerland. [4]Institute of Molecular Systems Biology, EHT Zurich, Switzerland. [5]Department of Effect Directed Analysis, Helmholtz Center for Environmental Research, Germany. [6]Research Unit Analytical BioGeoChemistry, Helmholtz Munich, Germany. [7]Department of Nutrition, Exercise and Sports, University of Copenhagen, Denmark. [8]Department of Plant and Environmental Sciences, Weizmann Institute of Science, Israel. [9]Genome Biology Unit, EMBL, Germany. [10]RECETOX, Masaryk University, Czech Republic. [11]Computational Plant Biochemistry, MetaCom, Leibniz Institute of Plant Biochemistry, Germany. [12]Metabolomics and Proteomics Core, Helmholtz Munich, Germany. [13]Anesthesiology and Intensive Care Medicine, University Hospital Greifswald, Germany. [14]Institute for Biomedicine, Eurac Research, Italy.

In a nutshell

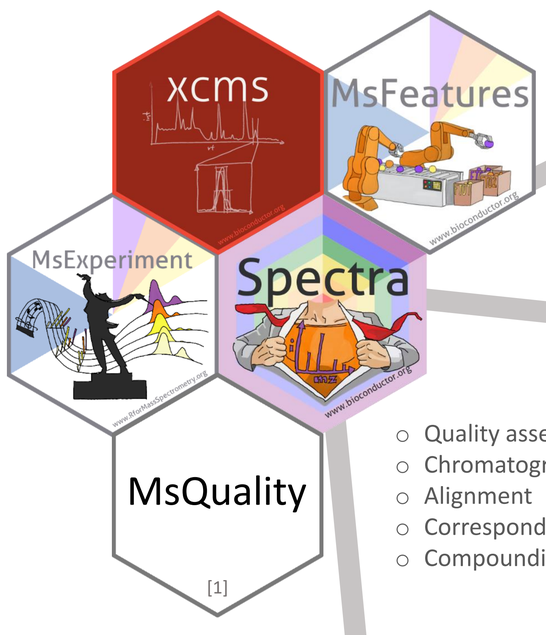
- Open collaborative development of software for MS data analysis:
  - Well documented and thoroughly tested
  - Long-term support and maintenance
- Modular package ecosystem to allow creation of custom analysis workflows
- Access to low-level functions enables *advanced users* and developers to integrate packages into their own projects
- Open for contributions from the community

Core *low-level* functions



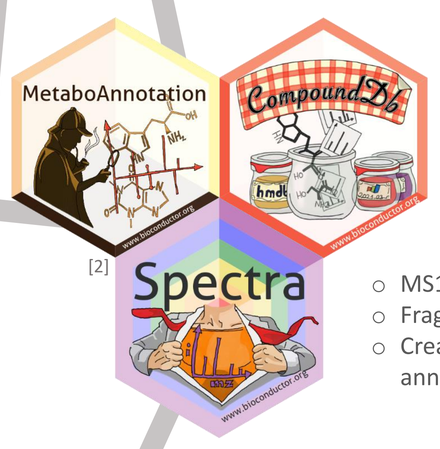
- Spectra similarity scores
- Peak alignment
- Handling chemical formulas
- Isotope peak prediction
- Adduct/exact mass calculations
- …

LC-MS(/MS) analysis



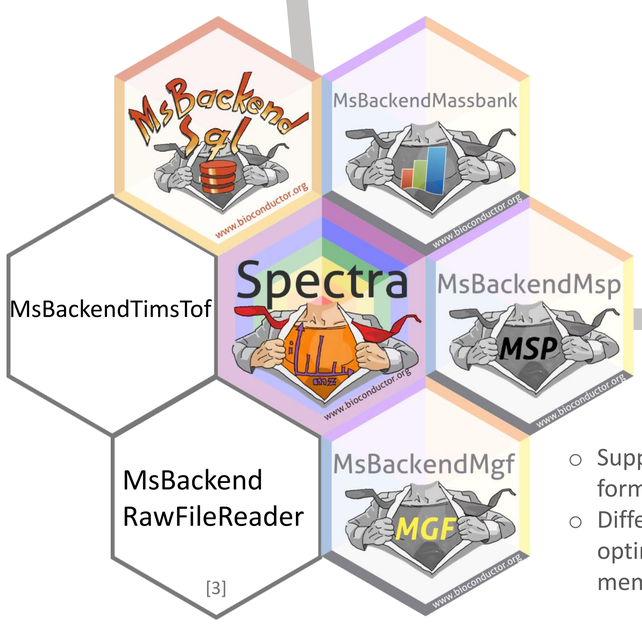
- Quality assessment
- Chromatographic peak detection
- Alignment
- Correspondence
- Compounding

Annotation



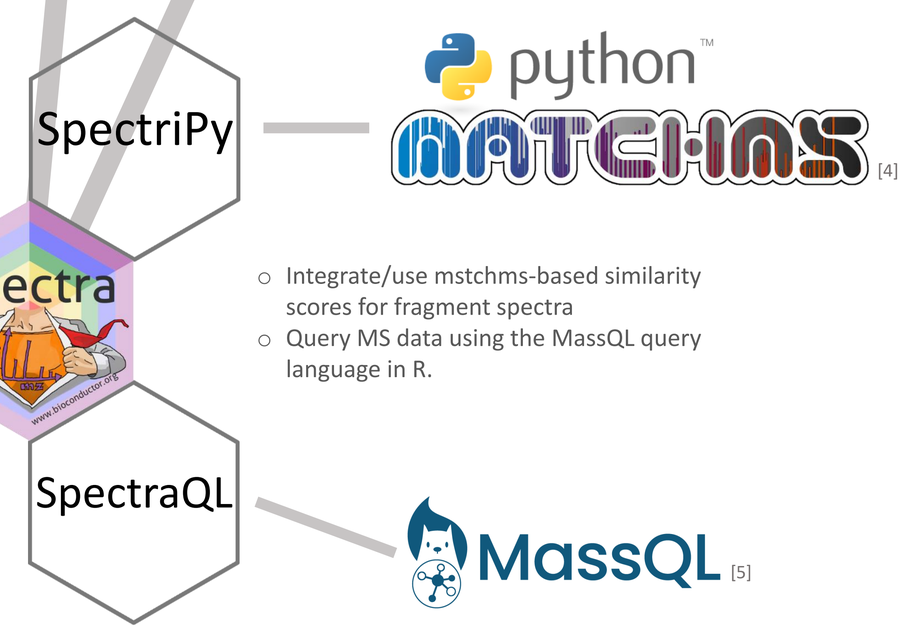
- MS1 and MS2-based annotation
- Fragment spectra matching
- Creating and maintaining annotation libraries

Supported data files/formats



- Support for additional MS data file formats
- Different representations of MS data in R optimized either for performance or low memory footprint.

Integration with external tools



- Integrate/use mstchms-based similarity scores for fragment spectra
- Query MS data using the MassQL query language in R.

Tutorials and documentation

xcms data analysis | MetaboAnnotation | Spectra

References

[1] MsQuality – an interoperable open-source package for the calculation of standardized quality metrics of mass spectrometry data. Naake T et al. bioRxiv 2023. https://doi.org/10.1101/2023.05.12.540477
[2] A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R. Rainer J et al. *Metabolites* 2022. https://doi.org/10.3390/metabo12020173
[3] The rawrr R Package: Direct Access to Orbitrap Data and Beyond. Kockmann T et al. Journal of Proteome Research 2021. https://doi.org/10.1021/acs.jproteome.0c00866
[4] matchms – processing and similarity evaluation of mass spectrometry data. Huber F et al. JOSS 2020. https://doi.org/10.21105/joss.02411
[5] A Universal Language for Finding Mass Spectrometry Data Patterns. Jarmusch AK et al. bioRxiv 2022. https://doi.org/10.1101/2022.08.06.503000

@jo_rainer
@jorainer@fosstodon.org

https://doi.org/10.5281/zenodo.7936787

eurac research