Metabolomics Software Libraries and Tools using Python

Sai Shiva Bhaskar Emani[1], Ikuko Kato[2], Xiang Zhang[3], and Seongho Kim[2,4]

[1]Data Science and Business Analytics, Mike Ilitch School of Business, Wayne State University, Detroit, MI

[2]Department of Oncology, School of Medicine, Wayne State University, Detroit, MI

[3]Department of Chemistry, University of Louisville, Louisville, KY

[4]Biostatistics and Bioinformatics Core, Karmanos Cancer Institute, Wayne State University

WAYNE STATE
School of Medicine

Background

Metabolomics, as a distinct field of scientific research, is relatively young compared to other "omics" disciplines like genomics and proteomics. Its development has been driven by advancements in analytical techniques, bioinformatics, and the need for a comprehensive understanding of metabolic processes. The term "metabolomics" itself was coined in the late 1990s as a natural progression from genomics and proteomics.

Metabolomics has found applications in diverse fields, including medicine, nutrition, agriculture, environmental science, and drug development.

The field of Metabolomics is continuously growing at a rapid pace, as it continues to evolve along with the development of new software tools. Python with its extensive libraries has risen as one of the top programming languages to utilize its versatility, easy to read, learn and write while being highly efficient for data analysis.

In this study, we explore and review around 100 libraries currently available in Python for metabolomics and classify them into three distinct categories as Annotation, Pre-processing and Downstream Analysis.

This review further describes the strengths and limitations of each library within these categories which results in making informed choices in selecting the most suitable tools for specific metabolomics study or project.

Methods

In this review, we utilized the Python repository, The Python Package Index (PyPI; https://pypi.org), to identify the Python libraries for metabolomics.

We searched using the following keywords “Metabolomics and Metabolites” Using these keywords, the initial search yielded 117 Python libraries. After excluding libraries not relevant to MS and NMR instruments, a total of 96 libraries were remained.

These were then categorized into three distinct categories based on their primary functions as **Annotation, Pre-processing and Downstream Analysis.**

- **Annotation:** Identifying and assigning chemical identities to detected metabolites, encompassing tasks like metabolite identification, compound database matching, and structure elucidation.
- **Pre-processing:** Addresses the key tasks such as data filtering, normalization, scaling, and the imputation of missing values.
- **Downstream Analysis:** Enables the extraction of insights from data through statistical analysis, pathway analysis, and visualization methods like heatmaps and scatter plots, facilitating the identification of patterns and biomarkers.

Results

Annotation

Library Name	Link	Platform Dependency	Published Date	Latest Version	Latest Version Update Date	Author Name
khipu	https://pypi.org/project/khipu-metabolomics/	LC-MS	4/5/2023	0.7.0	5/13/2023	I.Blazenovic et al.
asari	https://pypi.org/project/asari-metabolomics/	LC-MS	7/11/2023	1.11.4	5/13/2023	S. Barnes
pyMolNetEnhancer	https://pypi.org/project/pyMolNetEnhancer/		5/31/2019	0.2.0	8/7/2023	M.Ernst et al.
beamspy	https://pypi.org/project/beamspy/	untargeted ultra (high) performance liquid chromatography-mass spectrometry		1.2.0	5/16/2023	
cometa	https://pypi.org/project/cometa/			0.1.69	2/8/2022	
mzOS	https://pypi.org/project/mzOS/	LC-MS		0.1.3	1/8/2016	
dimedbpy	https://pypi.org/project/dimedbpy/			0.1.2	8/29/2020	
isoenum-webgui	https://pypi.org/project/isoenum-webgui/	NMR		0.2.0	8/14/2019	
pyisopach	https://pypi.org/project/pyisopach/			0.1.2	8/10/2019	
MetaboDistTrees	https://pypi.org/project/MetaboDistTrees/		8/9/2016	0.1.3	3/30/2020	M.Wang et al.
DecoID	https://pypi.org/project/DecoID/	MS/MS	3/4/2022	0.3.3	3/4/2022	Ethan Stancliffe
idms-pairfinder-2	https://pypi.org/project/idms-pairfinder-2/	IDMS and LC-MS		0.0.7	7/17/2019	
hiconet	https://pypi.org/project/hiconet/			0.5.4	2/13/2023	
AMON-bio	https://pypi.org/project/AMON-bio/			1.0.0	12/1/2019	
django-mogi django-misa django-galaxy django-mbrowse django-gfiles	https://pypi.org/project/django-mogi/	LC-MS(/MS)		0.0.2	7/30/2019	
metaMS	https://pypi.org/project/metaMS/	GC-MS		2.1.3	6/24/2021	
pypath-omnipath	https://pypi.org/project/pypath-omnipath/			0.15.4	6/12/2023	

DownStream Analysis

Library Name	Link	Platform Dependency	Published Date	Latest Version	Latest Version Update Date	Author Name
pymantra	https://pypi.org/project/pymantra/	Mass Spectometry (MS)	5/15/2023	1.0.1	5/18/2023	
MetEvolSim	https://pypi.org/project/MetEvolSim/	Liquid Chromatography-Mass Spectrometry (LC-MS)	8/15/2022	0.6.3	1/9/2023	
Xconnector	https://pypi.org/project/Xconnector/	Database	8/15/2021	1.0.5	7/9/2020	
mummichog	https://pypi.org/project/mummichog/	Liquid Chromatography-Mass Spectrometry (LC-MS)	7/4/2013	2.6.1	2/7/2022	
catbridge	https://pypi.org/project/catbridge/	multi-omics		0.5	6/29/2023	
msp2db	https://pypi.org/project/msp2db/	Mass Spectrometry		0.0.9	11/4/2019	
sspa	https://pypi.org/project/sspa/		8/26/2023	0.2.4	7/4/2023	C.Wieder and M.Brochut
vimms	https://pypi.org/project/vimms/	Liquid Chromatography-Mass Spectrometry (LC-MS)	10/9/2019	2.1.1	3/11/2023	J.Wandy et al.
ezcalour	https://pypi.org/project/ezcalour/			1.12.4	12/1/2020	
HMKG	https://pypi.org/project/HMKG/	Human Metabolome Database (HMDB)		0.0.1	2/20/2023	
bmxp	https://pypi.org/project/bmxp/	Liquid Chromatography-Mass Spectrometry (LC-MS)		0.0.15	8/14/2023	
secimtools	https://pypi.org/project/secimtools/			22.3.23	3/24/2022	
nmrml2isa	https://pypi.org/project/nmrml2isa/		4/1/2020	0.3.1	12/19/2016	N.L.Thomas et al.
cheMLearning	https://pypi.org/project/cheMLearning/			0.0.1	11/9/2022	
dynamet	https://pypi.org/project/dynamet/	Liquid Chromatography-Mass Spectrometry (LC-MS)		2.8.6	6/7/2016	
isosolve	https://pypi.org/project/isosolve/	NMR and MS		1.0.1	3/2/2021	
dimet	https://pypi.org/project/dimet/	targeted isotope-labeled metabolomics data		0.1.0	7/22/2023	Galvis et al.
meta2vec	https://pypi.org/project/meta2vec/			0.1.0.4	2/19/2023	
LOCATE-mode	https://pypi.org/project/LOCATE-mode/			0.0.3	11/23/2022	
mummichog1	https://pypi.org/project/mummichog1/	Liquid Chromatography-Mass Spectrometry (LC-MS)	7/4/2013	1.0.11	3/6/2018	Li et al.
framed	https://pypi.org/project/framed/		1/12/2017	0.5.2	5/22/2019	D.Machado
cimcb-litecimcb	https://pypi.org/project/cimcb-lite/ https://pypi.org/project/cimcb/			1.0.2	10/28/2019	
pyopls	https://pypi.org/project/pyopls/		1/18/2002	20.3.post1	3/4/2020	J.Trygg and S.Wold
omicscope	https://pypi.org/project/omicscope/			1.3.3	8/16/2023	
bio2bel-wikipathways	https://pypi.org/project/bio2bel-wikipathways/		11/10/2017	0.2.3	3/13/2019	Slenter. D.N et al.
ASCA	https://pypi.org/project/ASCA/		5/12/2005	1	12/6/2022	A.K. Smilde et al.
GEMAP	https://pypi.org/project/GEMAP/			0.0.3	8/16/2023	
meta-matching-tool	https://pypi.org/project/meta-matching-tool/	Liquid Chromatography-Mass Spectrometry (LC-MS)		1.8.0	5/31/2023	
PALS-pathway	https://pypi.org/project/PALS-pathway/		6/8/2020	1.5.2	3/17/2021	Mcluskey. K. et al.
nplinker	https://pypi.org/project/nplinker/		8/12/2022	1.3.2	8/12/2022	Geng et al.
SIMSIToolBoxSIMSIToolBo	https://pypi.org/project/SIMSIToolBox/			0.1.4	3/16/2023	
gpml2svg	https://pypi.org/project/gpml2svg/			0.3.0	8/19/2014	
dFBApy	https://pypi.org/project/dFBApy/			0.0.4	5/10/2022	
MAGINE	https://pypi.org/project/MAGINE/	multi-omics		0.1.5	2/25/2021	
MetaCerberus	https://pypi.org/project/MetaCerberus/	metagenomics/metatranscriptomic	8/12/2023	1.1	7/17/2023	Figueroa JL et al.
diffupath	https://pypi.org/project/diffupath/		5/13/2020	0.0.3	5/13/2020	D.D. Fernandez et al.
metDataModel	https://pypi.org/project/metDataModel/	LC-MS		0.4.14	3/2/2022	
Pathomx	https://pypi.org/project/Pathomx/	NMR		3.0.2	11/28/2014	
pynmranalysis	https://pypi.org/project/pynmranalysis/			1.1.3	8/19/2021	

Pre-processing

Library Name	Link	Platform Dependency	Published Date	Latest Version	Latest Version Update Date	Author Name
asari	https://pypi.org/project/asari-metabolomics/	LC-MS	7/11/2023	1.11.4	5/13/2023	S. Barnes
mwtab	https://pypi.org/project/mwtab/	MS and NMR	3/12/2021	1.2.5	3/18/2022	Powell Christian D. and H. NB. Moseley
matchms	https://pypi.org/project/matchms/	MS/MS	8/31/2020	0.22.0	8/18/2023	S. F. Huber et al.
amimspy	https://pypi.org/project/amimspy/	Acoustic mist ionization mass spectrometry		0.1.0	4/15/2021	
dimspy	https://pypi.org/project/dimspy/	Direct-infusion mass spectrometry	4/27/2020	2.0.0	4/26/2020	Ralf J. M. Weber and J. Zhou
tidyms	https://pypi.org/project/tidyms/	LC-MS	10/16/2020	0.7.0	4/24/2023	G. Riquelme et al.
metabolinks	https://pypi.org/project/metabolinks/	MS		0.75	12/22/2022	
ms-mint-app ms-mint	https://pypi.org/project/ms-mint-app/ https://pypi.org/project/ms-mint/	LC-MS		0.3.1.2	8/15/2023	
pymean	https://pypi.org/project/pymean/			0.3	7/3/2019	
jms-metabolite-services	https://pypi.org/project/jms-metabolite-services/	LC-MS		0.5.7	5/13/2023	
imzml2isa-qt mzml2isa-qt	https://pypi.org/project/imzml2isa-qt/			0.3.4	1/12/2017	
nmrml2isa-qt	https://pypi.org/project/nmrml2isa-qt/					
mzml2isa	https://pypi.org/project/mzml2isa/		4/1/2020	1.1.1	10/16/2022	T. N. Lawson et al.
breathXplorer	https://pypi.org/project/breathXplorer/	MS/MS spectra MGF		0.1.7	6/2/2023	
metabo-adni	https://pypi.org/project/metabo-adni/	NMR and p180		0.5.2	7/29/2022	
metabolabpytools metabolabpy qtmetabolabpy	https://pypi.org/project/metabolabpytools/ https://pypi.org/project/metabolabpy/ https://pypi.org/project/qtmetabolabpy/	NMR		0.9.8	8/11/2023	
DIMEpy	https://pypi.org/project/DIMEpy/			1.0.1	9/3/2020	
BioDendro	https://pypi.org/project/BioDendro/	MS/MS		0.0.3	2/25/2021	
messes	https://pypi.org/project/messes/	MS and NMR		1.0.1	5/31/2023	
spectrum-utils	https://pypi.org/project/spectrum-utils/	MS	12/6/2019	0.4.2	4/26/2023	W. Bittremieux
iptkl	https://pypi.org/project/iptkl/		8/17/2021	0.6.24	8/8/2022	H. ElAbd et al.
pcpfm	https://pypi.org/project/pcpfm/	LC-MS		0.0.31	8/11/2023	
PeakDetective	https://pypi.org/project/PeakDetective/	LC-MS		0.1.8	7/31/2023	
jonwestern icoshift-py3 icoshift icoshift3	https://pypi.org/project/jonwestern/ https://pypi.org/project/icoshift-py3/ https://pypi.org/project/icoshift/ https://pypi.org/project/icoshift3/	NMR	11/18/2009	0.0.1	2/10/2021	F. Savorani et al.
metabolights-utils	https://pypi.org/project/metabolights-utils/	ISA Files		0.9.6	7/20/2023	
srm-helper	https://pypi.org/project/srm-helper/	HRMS MS/MS		0.0.2	7/13/2021	
ms-autoqc	https://pypi.org/project/ms-autoqc/	LC-MS		1.0.1	2/26/2023	
DBDIpy	https://pypi.org/project/DBDIpy/	FT-ICR-MS	10/18/2022	1.2.2	7/3/2023	L. Weidner
breathpy	https://pypi.org/project/breathpy/	GC-MS and LC-MS	8/19/2020	0.9.6	2/7/2022	W. Philipp
memo-ms	https://pypi.org/project/memo-ms/		12/24/2021	0.1.4	2/18/2022	A. Gaudry et al.
pathme	https://pypi.org/project/pathme/		5/15/2019	0.1.13	3/26/2020	D. D. Fernandez et al.
MolNotator	https://pypi.org/project/MolNotator/	LC-MS/MS	12/22/2021	0.1.2	2/5/2023	D. O. Jimenez et al.
qurro	https://pypi.org/project/qurro/		10/20/2022	0.8.0	10/20/2022	M. Fedarko wt al.
empress	https://pypi.org/project/empress/		3/16/2021	1.2.0	7/20/2021	K. Cantrell et al.
metDataModel	https://pypi.org/project/metDataModel/	LC-MS		0.4.14	3/2/2022	
Pathomx	https://pypi.org/project/Pathomx/	NMR		3.0.2	11/28/2014	
pynmranalysis	https://pypi.org/project/pynmranalysis/			1.1.3	8/19/2021	
django-mogi django-misa django-galaxy django-mbrowse django-gfiles	https://pypi.org/project/django-mogi/	LC-MS(/MS)		0.0.2	7/30/2019	
metaMS	https://pypi.org/project/metaMS/	GC-MS		2.1.3	6/24/2021	
pypath-omnipath	https://pypi.org/project/pypath-omnipath/			0.15.4	6/12/2023	

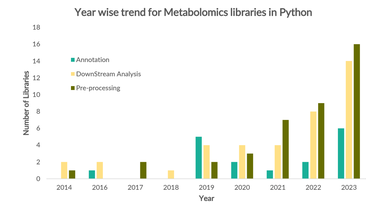

Figure 1 – Year wise trend of Python Libraries with version updates

Table 1 – Comparisons among ASARI, MZmine, and XCMS

Feature	ASARI	MZmine	XCMS
Maturity	Newer	Well-established	Well-established
Ease of use	Easy to use	Difficult to learn and use	Easy to use
Versatility	Very versatile	Very versatile	Limited versatility
Configurability	Highly configurable	Highly configurable	Less configurable
Speed and efficiency	Fast and efficient	Fast and efficient	Very fast and efficient
Accuracy and reproducibility	Accurate and reproducible	Accurate and reproducible	Very accurate and reproducible
Integration with other software packages	Well-integrated	Not as well-integrated	Not as well-integrated

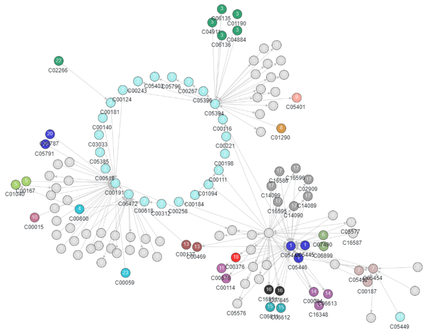

Figure 2 – Circular Spring Embedder (CiSE) for Pathway Analysis of Empirical Compounds in Negative data with 69 nodes and 67 edges

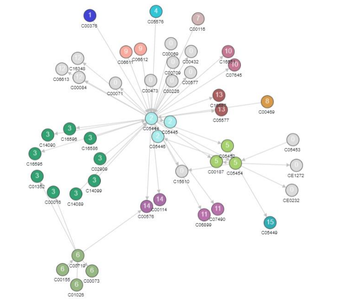

Figure 3 –Circular Spring Embedder (CiSE) for Pathway Analysis of Empirical Compounds in Positive data with 37 nodes and 39 edges

The results depicted in Fig. 2 and Fig. 3 are based on the outputs of Mummichog (https://mummichog.org/), a software tool developed in 2013. This tool was primarily designed for untargeted/global pathway analysis, where prior identification of metabolites is not required.

Conclusion

During the review of Metabolomics software libraries and tools using python, we noticed platforms like Mummichog and Asari have significantly advanced the field of metabolomics. Asari, a tool recently produced for pre-processing computationally outperforms XCMS and Mzmine.

Python libraries play a vital role in simplifying and enhancing metabolomics research, making it easier to analyze, interpret, and draw meaningful conclusions from complex metabolomics datasets due to its readability and simplicity, versatility, open source and free along with large community support and Cross platform compatibility.

References

Li, S., Siddiqa, A., Thapa, M. et al. Trackable and scalable LC-MS metabolomics data processing using asari. Nat Commun 14, 4113 (2023). https://doi.org/10.1038/s41467-023-39889-1

Li S, Park Y, Duraisingham S, Strobel FH, Khan N, et al. (2013) Predicting Network Activity from High Throughput Metabolomics. PLOS Computational Biology 9(7): e1003123. https://doi.org/10.1371/journal.pcbi.1003123

Acknowledgment

This research was partially funded by NIH grants (National Institutes of Health R21GM140352), (National Cancer Institute P30CA022453), (National Institutes of Health P20GM113226) and (National Institutes of Health P50AA024337).