REVIEW

Current analytical technologies and bioinformatic resources for plant metabolomics data

Chigateri M. Vinay¹ · Sanjay Kannath Udayamanoharan¹ · Navya Prabhu Basrur¹ · Bobby Paul² · Padmalatha S. Rai¹

Received: 18 June 2021 / Revised: 26 July 2021 / Accepted: 15 August 2021 / Published online: 25 August 2021
© The Author(s) 2021

Abstract

Plant metabolome as the downstream product in the biological information of flow starting from genomics is highly complex, and dynamically produces a wide range of primary and secondary metabolites, including ionic inorganic compounds, hydrophilic carbohydrates, amino acids, organic compounds, and compounds associated with hydrophobic lipids. The complex metabolites present in biological samples bring challenges to analytical tools for separating and characterization of the metabolites. Analytical tools such as nuclear magnetic resonance (NMR) and mass spectrometry have recently facilitated the separation, characterization, and quantification of diverse chemical structures. The massive amount of data generated from these analytical tools need to be handled using fast and accurate bioinformatics tools and databases. In this review, we focused on plant metabolomics data acquisition using various analytical tools and freely available workflows from raw data to meaningful biological data to help biologists and chemists to move at the same pace as computational biologists.

Keywords Metabolomics · Plant metabolome · Databases · Analytical tools · Computational tools · Mass spectrometry

Abbreviations

NMR	Nuclear magnetic resonance
LC–MS	Liquid Chromatography–mass spectrometry
GC–MS	Gas chromatography–mass spectrometry
CE–MS	Capillary electrophoresis–mass spectrometry
FT-ICR–MS	Fourier transform ion cyclotron resonance–mass spectrometry
MS/MS	Tandem mass spectrometry
PCA	Principal component analysis
HRMS	High-resolution tandem mass spectrometry
ESI	Electrospray ionization
APCI	Atmospheric pressure chemical ionization
HILIC	Hydrophilic interaction chromatography

Introduction

Over the past decades, metabolomics has been described as a popular omics approach along with other approaches like genomics, transcriptomics and, proteomics to study the metabolites (compounds less than 1500 Da) in a living system (Fiehn et al. 2000; Fiehn 2001, 2002; Bino et al. 2004; Saito and Matsuda 2010). Recently, metabolomics is a growing field and applied to study in-depth information on plant cellular metabolism, a group of the small molecular mass of compounds present in the plant extracts (Oliver et al. 1998). Usually, plants produce numerous metabolites for their developmental growth and environmental responses. Around 200,000 metabolites, including, primary and secondary metabolites are present in the plant kingdom (Wink 2010). These primary metabolites are liable for developmental growth and these are structurally conserved while, secondary metabolites are specialized, utilized during stress conditions, and vary across plant species (Scossa et al. 2016). Plant metabolome produces a wide selection of metabolites including, ionic inorganic compounds, hydrophilic carbohydrates, amino acids, organic compounds, and compounds associated with hydrophobic lipids. The complex metabolites present in plant samples bring challenges to analytical tools for the separation and

✉ Padmalatha S. Rai
padmalatha.raai@manipal.edu

¹ Department of Biotechnology, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, India

² Department of Bioinformatics, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, India

characterization of the metabolites. At present, the various analytical tools such as Nuclear magnetic resonance (NMR), Liquid chromatography–mass spectrometry (LC–MS), Gas chromatography–mass spectrometry (GC–MS), Capillary electrophoresis mass spectrometry (CE–MS), and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR–MS) (Okazaki and Saito 2012; Khakimov et al. 2014) are employed in the metabolomic analysis. Among these, LC–MS and GC–MS, and NMR are being used routinely in plant metabolomics. However, there is an intrinsic limitation for each analytical technology that can cover the whole plant metabolome, the combination of analytical techniques being used in plant metabolomics analysis.

The large amount of plant metabolome data produced by each analytical tool requires distinct handling, and thus, bioinformatics resources like software for data processing and analysis, databases for annotation, and workflows are required. There is a continuous increase in the advancement of analytical platforms and high-throughput data, newer tools are being developed for handling the metabolomic datasets. Recently, Spicer et al. provided a list of open-source software tools for metabolomic analysis. Most of the software tools are platform-specific and written in different scripting languages such as Python, R, Java, and MATLAB (Novikova et al. 2020). The automated software or workflows for plant metabolomics data analysis are not freely available. The metabolite identification process is manual and semi-automated to access the features of biological interest (Dunn 2008). To fill the gap, several workflows are scripted to perform integrated pre-processing, annotation, statistical analysis, and biological interpretation of metabolomic data. These are powerful, easy-to-use, modularized, and a good choice for plant metabolomics data analysis. In the present review, we will describe the typical plant metabolomics analysis workflow including analytical tools for data acquisition, software tools and databases for processing, annotation, and data analysis for GC–MS, LC–MS, NMR, GC–MS/MS, and LC–MS/MS data. This review also describes the multifunctional tools such as workflows for plant metabolomics analysis and limitations of current tools and future perspectives towards plant metabolomic analysis.

Workflow for plant metabolomic analysis

The typical workflow for plant metabolomic analysis comprised four steps namely (1) sample preparation (2) data acquisition (3) data mining/data analysis (4) data integration resulted in the interpretation of meaningful biochemical information (Fig. 1).

Sample preparation

This is often the critical step where practical considerations should be taken from the beginning to avoid the introduction of variability in metabolomics analysis. Sample preparation involves mainly three steps: harvesting, drying, and extraction methods. Plant selection depends on the biological issue to be explored (Kim and Verpoorte 2010; Álvarez-Sánchez et al. 2010). Harvesting the plant material is crucial to avoid metabolite changes caused by enzymatic reactions and contamination (Fiehn et al. 2000; Verpoorte et al. 2008; Xu et al. 2010). After harvested the plant material, drying can be performed using air, oven, freeze, and trap drying methods (Harbourne et al. 2009). Drying should be performed prior to the extraction to maintain a uniform water level within the plant sample. Extraction methods used for the extraction of metabolites depend on physicochemical properties, solvent properties, and biochemical system composition. The various extraction methods such as solvent extraction, ultra-sonication, microwave-assisted extraction, and supercritical fluid extraction are commonly and routinely used. However, no single extraction technique is been developed to extract all the metabolites till date. Thus, the combination of extraction methods usage is imperative for the extraction of various classes of metabolites. (Kim et al. 2011; Verpoorte et al. 2008; Hall, 2006; Dunn and Ellis 2005; Kim and Verpoorte 2010).

Data acquisition using analytical techniques

Once the sample is prepared, various analytical techniques are used to separate, quantify, and identify the various metabolite groups (Sumner et al. 2003; Allwood et al. 2008). Each analytical technique has its pros and cons as listed in Table 1. Selecting the analytical method relies on the type of approach (targeted or untargeted) and also based on the physical and chemical characteristics of the sample.

Nuclear magnetic resonance spectroscopy

NMR spectroscopy is a non-destructive and, a non-biased method that is easy to quantify, needs little to no preparation of samples, no analyte or chromatographic separation, and no sample derivatization prior to the analysis. This analytical tool is straightforward and depends on the external magnetic field with radiofrequency (RF) radiations applied to the nuclei of molecules. When samples present in NMR solvent are applied with the external magnetic field and radiofrequency pulses, the atomic nuclei transfer energy from

Fig. 1 Flowchart of a typical plant metabolomics study

lower states to higher states, and subsequent energy is emitted when spins back to lower energy states at an equivalent frequency. This generates signals and is reported as chemical shifts in NMR. Usually, the NMR chemical shifts are mentioned relative to an internal reference. The resultant spectrum is that all the peaks are distributed at different locations and intensities to obtain a unique pattern for each compound. In NMR, with the use of specific atomic nuclei like hydrogen (^1H), carbon (^{13}C), and phosphorous (^{31}P), information on specific metabolite types is obtained (Reo 2002). Hence, NMR analysis provides the global view of all metabolites in plant samples (Kim et al. 2011). It can also unveil structural information which helps to identify the unknown metabolite. Although NMR has the greatest advantages, it exhibits inherent low sensitivity and resolution which hampers the quality of analysis for complex samples (Welje et al. 2006). With the implementation of cryogenic probes, disadvantages are often addressed (Kovacs et al. 2005). Hence there is a shift within the metabolomics community towards the utilization of MS-based techniques.

Gas chromatography–mass spectrometry

GC–MS is the analytical tool often used in plant metabolomics. It is largely used to quantify volatile compounds at higher temperatures and derivatized polar metabolites (Dunn 2008). As separation in GC occurs at a higher temperature, derivatization is important for the samples to form them volatile, prior to the analysis. The electron ionization technique helps in the compound identification by generating informative and characteristic mass spectra with high degree fragmentation. The spectral data are often compared with the NIST database to identify unknown metabolites. However, molecular ions are often undetected which largely reduces the elemental composition, and hence, this technique is often used for the selective study of known primary metabolites. Recently, alternative techniques such as positive chemical ionization and negative ion chemical ionization are employed to enhance the separation and sensitivity of metabolites in complex mixtures (Raina and Hall 2008).

Table 1 Analytical techniques in plant metabolomics and its advantages and disadvantages

Techniques	Advantages	Disadvantages
Nuclear magnetic resonance (NMR)	High reproducibility Several analysis can be carried out with same sample Minimal sample preparation Exact quantification possible Relatively high throughput	Sensitivity is low Selectivity is low Not commonly used for targeted analysis
Gas chromatography mass spectrometry (GC–MS)	High sensitivity Availability of universal databases for identification Relatively high throughput Determination of molecular weight and fragmentation pattern of compound	Sample cannot be recovered Only volatile compounds are detected Requires extensive sample preparation Low ionization discrimination
Liquid chromatography–mass spectrometry (LC–MS)	High sensitivity Relatively high throughput Minimal sample preparation Superior technique for targeted analysis	High ionization discrimination Sample cannot be recovered Retention time depends on exact chromatographic condition Lack of extensive metabolite databases
Capillary electrophoresis–mass spectrometry (CE–MS)	Sensitivity is high Best suitable for polar analysis in aqueous samples Cost required for analysis low	Samples cannot be recovered Relatively low throughput
Fourier transform-ion cyclotron resonance–mass spectrometry (FT-ICR–MS)	High mass accuracy to resolve interferences High sensitivity	Use of superconducting magnets are expensive Computationally expensive for large amount of data generated

Liquid chromatography–mass spectrometry

LC–MS is another analytical technique utilized in plant metabolomics. In comparison with the GC–MS technique, it is more suitable for non-volatile and polar metabolites. Plant metabolomics is used for untargeted secondary metabolites analysis and complex lipids using chromatographic separation with reverse phased C18 columns (De Vos et al. 2007). These columns provide a strong separation between nonpolar and weak polar compounds. To analyze polar compounds in plant extracts, hydrophilic interaction chromatography columns are often used (Tolstikov and Fiehn 2002). In liquid chromatography mass spectrometry, the ionization techniques include electrospray ionization and atmospheric chemical ionization coupled with LC–ESI–MS (Codrea et al. 2007; Dunn 2008; Grandori et al. 2009). LC–ESI–MS produces molecular ions namely a few like $[M + H]^+$, $[M + 2H]^+$, $[M + NH_4]^+$ ions in positive mode of ionization and $[M - H]^-$, $[M - 2H]^-$, $[M - NH_4]^-$ ions in negative mode, for the identification of the proposed compounds. The advantage of this ionization source would be either volatile or non-volatile and derivatization is not necessary as that of GC–MS. The disadvantage of this ionization source is that the identification of metabolite is more time-intensive and lacks a mass spectral library for metabolite identification. However, with the utilization of tandem mass spectrometry (MS/MS) the metabolites are often identified (Lenz et al. 2004). Hence, more effort is required to identify

the metabolites through the development of comprehensive databases (Kind and Fiehn, 2006; Moco et al. 2006; Shinbo et al. 2006; Böcker and Rasche, 2008; Horai et al. 2010).

Capillary electrophoresis–mass spectrometry

This technique is employed to analyze polar and charged metabolites. The ionization method used for CE–MS is API, almost like LC–MS. In capillary electrophoresis, the migration of ions under the presence of an electrical field produces electro osmotic flow through the capillary action (Okazaki and Saito 2012). Ionic compounds are separated into capillaries based on the charge and size of ions. The samples in CE–MS are separated into the cationic and anionic analysis. In Soga's method for cationic analysis, formic acid used as the electrolyte provides good reproducibility of metabolites (Soga et al. 2006). While in the anionic analysis, the use of platinum electrospray needles has improved significantly in analyzing anionic metabolites (Soga et al. 2009). In the CE–MS technique derivatization is not required for the detection of compounds like GC–MS (Okazaki and Saito 2012). The metabolites identified using the CE–MS technique are physiologically important and similar in organisms. Hence, this technique is employed to profile metabolites of all organisms (Urano et al. 2009; Ishikawa et al. 2010).

Fourier transform ion cyclotron resonance mass spectrometry

This technique analyses the mass to charge ratio (m/z) of ions within the fixed magnetic field supported cyclotron frequency. This technique has higher importance in metabolomics because similar molecular mass compounds are often separately detected, and chromatographic separation is not required for the rapid detection of metabolites. Another advantage is that chemical structure can be obtained from the peak analysis, which helps to identify unknown compounds. Coupling tandem mass spectrometry with FT-ICR-MS with high precision and resolution can detect fragmented metabolite ions. This technique produces a large amount of data with hardware handling difficulties. Compared with other MS techniques, only a few reports are available with FT-ICR-MS (Kujawinski 2002).

Data mining/analysis

The analytical instruments produce a large amount of data that is used for metabolome analysis. To facilitate this, automated software is required for the identification of peaks from the raw MS or NMR data to identify and quantify the metabolites. Hence, informatics and statistics are essential for metabolomics data analysis (Weljie et al. 2006; Boccard et al. 2007; Liland 2011). Data analysis includes data pre-processing, annotation of metabolites, and statistical analysis.

Data pre-processing

During this step, peak detection and chromatogram alignment algorithm are used for the raw data signals (chromatograms, NMR) to obtain the metabolite signals. Different software tools or packages were developed to aid compound identification, including XCMS (Smith et al. 2006), mzMine2 (Katajamaa et al. 2006), Metabolomic Analysis and Visualization Engine, MAVEN (Clasquin et al. 2012), Mass Spectrometry-Data Independent AnaLysis software, MS-DIAL (Tsugawa et al. 2015), OpenMS (Sturm et al. 2008), automated data analysis pipeline, ADAP (Jiang et al. 2010), adaptive processing of liquid chromatography mass spectrometry, apLCMS (Yu et al. 2009). All the major software programs are compatible with the open format such as mzML for feature extraction and quantification (Kessner et al. 2008; Martens et al. 2011). The annotation of metabolites is required to interpret the result from pre-processed data. Several databases and tools are available for metabolite identification as listed in Table 2.

Databases and tools for annotation of metabolites

Annotation of metabolites is the connection between raw analytical data and biological knowledge through targeted or untargeted profiling approaches. But the identification or annotation of metabolites is usually a laborious process. Usually, identification of metabolites being administered with the fragmentation data obtained MS/MS targeted against online tandem MS databases when standard compounds are unavailable. For GC-MS data, NIST spectral library is employed for automated deconvolution and identification system of the mass spectra (Ausloos et al. 1999), and the Golm Metabolite database, GMD (Kopka et al. 2005) is another open-source software for spectral search. For LC-MS metabolite annotation and spectral search are often performed using spectral databases like METLIN (Smith et al. 2005), MoNA (<http://mona.fiehnlab.ucdavis.edu/>), mzCloud (<https://www.mzcloud.org/>), MassBank (Horai et al. 2010) and Global Natural Products Social Molecular Networking database, GNPS (Wang et al. 2016). For mass spectral and NMR data, the Human Metabolome Database, HMDB (Wishart et al. 2009) and Madison Metabolomics Consortium Database, MMCDDB (Markley et al. 2007; Cui et al. 2008) are useful for annotation of metabolites in both humans and plants.

Several software tools including xMSannotator (Uppal et al. 2017), RAMclust (Broeckling et al. 2014), CAMERA (Kuhl et al. 2012), and MetAssign (Daly et al. 2014) are available for metabolite annotation. Numerous compound repositories for metabolite annotation such as PubChem (Wheeler et al. 2008), Chemical Entities of Biological Interest, ChEBI (Degtyarenko et al. 2008), KNApSACk (Shinbo et al. 2006), LipidBank, and LIPID MAPS (Fahy et al. 2007), Kyoto Encyclopedia of Genes and Genomes, KEGG (Kanehisa et al. 2008) and PlantCyc, from Plant metabolite network (Seaver et al. 2012) are also accessible which are obliging for metabolite annotation. Many specialized databases of plant species are used to focus on mass spectra or metabolite annotation. These include Metabolome Tomato Database, MotoDB (Moco et al. 2006), and Kazusa Metabolomics Database, KOMICS (Iijima et al. 2008) contain information on metabolites present in the tomato and Plant Reactome database (Naithani et al. 2020) for curated pathway information related to 84 species of Gramineae family.

Statistical analysis

Statistical analysis is applied to compare the metabolite properties with the dependent variable. Depending on the experimental objective, appropriate statistical methods are

Table 2 List of tools and databases for metabolomics interpretation

Name	Description	References	Web link
XCMS	Bioconductor package in R for data pre-processing	Smith et al. 2006, Benton et al. 2008	https://www.bioconductor.org/packages/release/bioc/html/xcms.html
mzMine2	Open source software for data pre-processing	Katajamaa et al. 2006	http://mzmine.github.io/
MAVEN (metabolomic analysis and visualization engine)	Open source software for data pre-processing	Clasquin et al. 2012	http://genomics-pubs.princeton.edu/mzroll/index.php
MS-DIAL (mass spectrometry-Data Independent Analysis software)	Open source software for data pre-processing	Tsugawa et al. 2015	http://prime.psc.riken.jp/compms/msdial/main.html
OpenMS	Open source software for data pre-processing	Sturm et al. 2008	https://www.openms.de/
ADAP (automated data analysis pipeline)	Workflow for pre-processing GC-MS data	Jiang et al. 2010	http://www.du-lab.org/software.html
apLCMS (adaptive processing of Liquid Chromatography Mass Spectrometry)	R package for pre-processing LC-MS data	Yu et al. 2009	http://web1.sph.emory.edu/apLCMS/
mzMatch	Open source software for data pre-processing	Scheltema et al. 2011	http://mzmatch.sourceforge.net/
NIST (National Institute of Standards and Technology)	Comprehensive reference library for GC-MS data	Ausloos et al. 1999	https://www.nist.gov/
GMD (Golm Metabolome database)	Mass spectral database of GC-MS data	Kopka et al. 2005	http://gmd.mpimgolm.mpg.de/
METLIN	Mass spectral search and annotation	Smith et al. 2005	https://metlin.scripps.edu/
MassBank	Mass spectral search and annotation	Horai et al. 2010	http://www.massbank.jp/
MoNA (MassBank of North America)	Mass spectral search and annotation	–	http://mona.fiehnlab.ucdavis.edu/
mzCloud	Mass spectral search and annotation	–	https://www.mzcloud.org/
GNPS (Global Natural Products Social Molecular Networking)	Mass spectral search and annotation	–	http://gnps.ucsd.edu
HMDB (Human Metabolome database)	Collection of human and plant metabolite data from NMR and MS Spectra	Wishart et al. 2009	https://hmdb.ca/
MMCDDB (Madison Metabolomics Consortium Database)	Collection of human and plant metabolite data from NMR and MS Spectra	Markley et al. 2007, Cui et al. 2008	http://mmcd.nmrfam.wisc.edu/
CAMERA	Bioconductor package for compound spectra extraction and annotation of LC-MS data	Kuhl et al. 2012	–
xMSannotator	R package for annotation of LC-MS data	Uppal et al. 2017	https://sourceforge.net/projects/xmsannotator/
RAMClust	R package for annotation of MS data	Broekling et al. 2014	https://github.com/cbroeckl/RAMClustR
MetAssign	Annotation of metabolites from LC-MS data	Daly et al. 2014	http://mzmatch.sourceforge.net/
KEGG (Kyoto Encyclopedia of Genes and Genomes)	Compound databases for metabolite annotation	Kanehisa et al. 2008	https://www.genome.jp/kegg/compound/
PubChem	Compound databases for metabolite annotation	Wheeler et al. 2008	https://pubchem.ncbi.nlm.nih.gov/
ChEBI (Chemical Entities of Biological Interest)	Compound databases for metabolite annotation	Degtyarenko et al. 2007	https://www.ebi.ac.uk/chebi
KNAPSack	Comprehensive species-metabolite relationship database	Shinbo et al. 2006	https://www.genome.jp/db/pcidb
LipidBank	Open source database for lipid annotation	Yasugi and Watanabe 2002	http://lipidbank.jp/
Lipid Maps	Open source database for lipid annotation	Faluy et al. 2007	https://www.lipidmaps.org/

Table 2 (continued)

Name	Description	References	Web link
PMN (plant metabolite network)	Metabolic pathway database and compound search for annotation	–	http://www.plantcyc.org
MotoDB (Metabolome Tomato Database)	Metabolite annotation in tomato by LC–MS data	Moco et al. 2006	http://appliedbioinformatics.wur.nl/moto
KOMICS (Kazusa Metabolomics Database)	Web based database for metabolite annotation in tomato	Iijima et al. 2008	http://www.kazusa.or.jp/komics/
SIMCA-P (Soft Independent Modelling by Class Analogy)	Commercial software for statistical analysis	–	http://umetrics.com/products/simca
SMART (Statistical Metabolomics Analysis-An R Tool)	Statistical analysis tool	Liang et al. 2016	http://www.stat.sinica.edu.tw/hsinchou/metabolomics/SMART.htm
Specmine	Statistical analysis tool	Costa et al. 2016	https://cran.r-project.org/web/packages/specmine/
Metap-server	Pathway analysis tool	Suhre et al. 2011	http://metabolomics.helmholtz-muenchen.de/metap2/
MetPA (metabolomics pathway analysis)	Pathway analysis tool	Xia et al. 2011	http://metpa.metabolomics.ca
MetExplore	Pathway analysis tool	Cottret et al. 2018	http://metexplore.toulouse.inra.fr
MSEA (Metabolite Set Enrichment Analysis)	Pathway analysis tool	Xia and Wishart 2010	https://www.metaboanalyst.ca/
KEGG (Kyoto Encyclopedia of Genes and Genomes)	Pathway database	Ogata et al. 1998	https://www.genome.jp/kegg/pathway.html
PlantCyc	Pathway database	–	http://www.plantcyc.org/
MapMan	Pathway database	Thimm et al. 2004	https://mapman.gabipd.org/
KaPPA-View	Pathway database	Tokimatsu et al. 2005, Sakurai et al. 2011	http://kpv.kazusa.or.jp/kpv4/
Arabidopsis reactome	Curated pathway database of Arabidopsis thaliana	Tsesmetzis et al. 2008	www.arabidopsisreactome.org
Plant reactome	Curated pathway database of Gramineae family	Naithani et al. 2020	https://plantreactome.gramene.org

often applied to metabolomics data. Metabolomics studies from the analytical methods incorporate various biases into the data that make interpretation complicated and demanding, hence, multivariate analysis and chemometric methods are used for metabolomics data. The multivariate analysis consists of supervised and unsupervised methods. One of the unsupervised methods in metabolomics studies is principal component analysis. PCA describes the systematic variations in each sample's metabolites by projecting and clustering the samples in a data table. Many studies have reported the utilization of PCA for statistical analysis of the metabolomics data (Catchpole et al. 2005). Other statistical tools like hierarchical clustering (Murtagh 1983), *K* means clustering (Murtagh 1983), partial least square discriminant analysis, PLS-DA (Barker and Rayens 2003), and Orthogonal PLS, OPLS (Trygg and Wold 2002) are used in metabolomics data analysis. Commercial tools like SIMCA-P (<http://umetris.com/products/simca>), PLS-Toolbox (http://www.eigenvector.com/software/pls_toolbox.htm) are employed for statistical analysis. Also, some software packages written in R programs like Statistical Metabolomics Analysis-An R Tool, SMART (Liang et al. 2016), and specmine (Costa et al. 2016) are being used.

Data interpretation

After statistical analysis, the chosen metabolites from the metabolomics data got to be linked with biochemical pathways. There are several ad-hoc software tools includes metaP-server (Suhre et al. 2011), metabolic pathway analysis, metPA (Xia et al. 2011), MetExplore (Cottret et al. 2018), Metabolite Set Enrichment Analysis, MSEA (Xia

and Wishart 2010) for the enrichment and pathway analysis which map to the biochemical pathways available publicly databases like Kyoto Encyclopedia of Genes and Genomes, KEGG (Ogata et al. 1998). In addition, specific plant metabolomics pathway databases like PlantCyc (<http://www.plantcyc.org/>), KaPPA-View (Tokimatsu et al. 2005; Sakurai et al. 2011), MapMan (Thimm et al. 2004), Arabidopsis Reactome (Tsesmetzis et al. 2008), and Plant Reactome (Naithani et al. 2020) are available. Although, these biochemical pathways provide meaningful information, in the case of plants most of the metabolic pathways are highly interconnected and redundant to understand the biological interpretation from Metabolome data. The analysis tools used to identify perturbations in pathways depend on prior metabolite identification. Hence, analysis using untargeted metabolomics workflow has increased prominence in recent years (Heuberger et al. 2014).

Multifunctional tools as workflows for plant metabolomic analysis

The developed software tools should facilitate the data processing, annotation, statistical analysis, and interpretation of metabolite data. The workflows in the Web app and other programming languages like R, python are easy to use and allowing the researchers to perform analysis in a single tool. These workflows are primarily designed for the analysis of LC–MS data. In this section, the workflows for plant metabolomics are listed in Table 3.

XCMS Online (Tautenhahn et al. 2012), Scripps center for metabolomics and mass spectrometry, is an integrated freely accessible cloud-based platform to facilitate raw MS

Table 3 List of tools for metabolomics data analysis

Name	Description	References	Link
XCMS online	Data pre-processing and downstream analysis of metabolomics data	Tautenhahn et al. 2012	https://xcmsonline.scripps.edu/
MetaboAnalyst 5.0	Statistical analysis and visualization of metabolomics data	Pang et al. 2021	https://www.metaboanalyst.ca/
Workflow4 Metabolomics	Analysis workflow available in galaxy environment	Giacomini et al. 2015	https://workflow4metabolomics.usegalaxy.fr/
Metabox	Analysis workflow in R	Wanichthanarak et al., 2017	https://github.com/kwanjeeraw/metabox
WebSpecmine	Web based application based on R package specmine	Cardoso et al., 2019	https://webspecmine.bio.di.uminho.pt/
Galaxy-M	Analysis workflow available in galaxy environment	Davidson et al. 2016	https://github.com/Viant-Metabolomics/Galaxy-M
MZmine2	Analysis of targeted and untargeted LC–MS data	Pluskal et al. 2010	https://mzmine.sourceforge.net/
MS-DIAL 4.0	Analysis of tandem mass spectrometry data	Tsugawa et al. 2020	https://prime.psc.riken.jp/
IP4M	Analysis workflow for untargeted LC–MS and GC–MS data	Liang et al. 2020	https://github.com/IP4M

spectra processing and visualization of untargeted LC–MS data. XCMS Online retains the same features as the original XCMS software (Smith et al. 2006) with more flexibility in web browsing and also provides a direct link to the METLIN (Smith et al. 2005) database for metabolite identification. It also provides interactive metabolomic cloud plots, retention time correction plots, univariate statistical analysis and, pathway information.

Metaboanalyst 5.0 (Pang et al. 2021) is an integrated freely accessible web platform for metabolomic analysis, visualization, and interpretation of MS and NMR data. It contains several new modules like MS spectral processing, functional analysis, meta-analysis, joint pathway analysis, enrichment analysis, and network explorer to the existing modules from Metaboanalyst 4.0 (Chong et al. 2018).

Workflow4metabolomics (Giacomoni et al. 2015) is developed by the French Bioinformatics Institute for data processing, statistical analysis, and interpretation of LC–HRMS data. It is built on a Galaxy environment with unique computational modules like data normalization, multivariate analysis, and annotation. Galaxy-M (Davidson et al. 2016) is developed for LC–MS and DIMS metabolomic data analysis. It provides processing of raw data to preparation of statistical analysis and annotation in the galaxy-based platform. Both Workflow4metabolomics and Galaxy-M require implementations of XCMS (Smith et al. 2006) and CAM-ERA (Kuhl et al. 2012) packages for the analysis.

MZmine 2 (Pluskal et al. 2010) is developed for both targeted and untargeted LC–MS data. It is an open-source data processing toolbox obtained under the license of GNU GPL. Metabox (Wanichthanarak et al. 2017), an R-based software framework for metabolomics data processing, statistical analysis, and functional analysis. WebSpecmine (Cardoso et al. 2019) is an R-based web-based framework for metabolomics and spectral data processing. MS-DIAL 4.0 (Tsugawa et al. 2020) is developed for tandem mass fragmentation data in data-independent acquisition mode along with comprehensive identification and quantification of small molecules including lipids by mass spectral deconvolution. Integrated mass spectrometry-based untargeted data mining, IP4M (Liang et al. 2020) covers all eight modules include data preprocessing, peak annotation, statistical analysis, pathway and enrichment analysis, classification and biomarker detection, correlation analysis, regression analysis, and sample size and power analysis for untargeted LC–MS and GC–MS metabolomics data.

Limitations of software tools and workflows

The metabolomic tools written in different coding languages like C, C++, R, Python, and Java have their pros and cons. Most of the tools have different pre-knowledge and

dependencies as a result, not all tools are compatible with various operating systems. Tools with differential input–output behavior, lack of reliable metabolomics databases, sensitivity, cost, standardization of workflows are some of the issues in the metabolomics research (Perez-Riverol et al. 2014). Despite its challenges in the metabolomics community towards metabolite identification, archiving and biological interpretation, computational biologists and informaticians have laid enormous efforts towards the development of software tools, common interfaces, integrated tools, and workflows. This needs to be validated by the analytical chemists before accepted by the metabolomics community to gain popularity.

Conclusion and future perspectives

Metabolomics is rapidly evolving in the field of plant biotechnology with the development of new analytical tools for metabolomic data acquisition to cover the whole plant metabolome. This review provides an overview of the most widely used analytical tools in plant metabolomics experiments along with the typical workflow of plant metabolomics experiment starting from the sample preparation to data interpretation with supporting software tools, databases, and workflows. These computational tools and databases are expanding exponentially for functional characterization and biological interpretation of plant metabolomic profiles. The computational metabolomics community needs to develop workflows with the advancement of modules. These workflows are integrated multifunctional tools which are freely available, makes the beginner analyze the metabolomic data easily. Finally, there is an urgent need for the integration of metabolomics with other omics technologies and a combination of various instrumentation capabilities helps to understand complex biological systems in plants.

Acknowledgements The authors are thankful to the Manipal Academy of Higher Education for providing us with the infrastructure and facilities at Manipal School of Life Sciences, Department of Biotechnology, Government of India, and Technology Information Forecasting and Assessment Council-Centre of Relevance and Excellence (TIFAC-CORE) in Pharmacogenomics.

Author contributions PSR and CMV designed the review manuscript. CMV and NPB prepared the tables and figures. CMV and SKU wrote the manuscript. PSR and BP read and approved the final manuscript.

Funding Open access funding provided by Manipal Academy of Higher Education, Manipal. This review received no financial assistance from any funding body.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval This review article does not contain any studies with human or animal participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allwood JW, Ellis DI, Goodacre R (2008) Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiol Plant* 132:117–135. <https://doi.org/10.1111/j.1399-3054.2007.01001.x>
- Álvarez-Sánchez B, Priego-Capote F, de Castro MDL (2010) Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC Trends Anal Chem* 29:120–127. <https://doi.org/10.1016/j.trac.2009.12.004>
- Ausloos P, Clifton CL, Lias SG et al (1999) The critical evaluation of a comprehensive mass spectral library. *J Am Soc Mass Spectrom* 10:287–299. [https://doi.org/10.1016/S1044-0305\(98\)00159-7](https://doi.org/10.1016/S1044-0305(98)00159-7)
- Barker M, Rayens W (2003) Partial least squares for discrimination. *J Chemom* 17:166–173. <https://doi.org/10.1002/cem.785>
- Benton HP, Wong DM, Trauger SA, Siuzdak G (2008) XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. *Anal Chem* 80:6382–6389. <https://doi.org/10.1021/ac800795f>
- Bino RJ, Hall RD, Fiehn O et al (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 9:418–425. <https://doi.org/10.1016/j.tplants.2004.07.004>
- Boccard J, Grata E, Thiocone A et al (2007) Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding. *Chemom Intell Lab Syst* 86:189–197. <https://doi.org/10.1016/j.chemolab.2006.06.004>
- Böcker S, Rasche F (2008) Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btn270/>
- Broeckling CD, Afsar FA, Neumann S et al (2014) RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem* 86:6812–6817. <https://doi.org/10.1021/ac501530d>
- Cardoso S, Afonso T, Maraschin M, Rocha M (2019) WebSpecmine: a website for metabolomics data analysis and mining. *Metabolites* 9:237. <https://doi.org/10.3390/metabo9100237>
- Catchpole GS, Beckmann M, Enot DP et al (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc Natl Acad Sci* 102:14458–14462. <https://doi.org/10.1073/pnas.0503955102>
- Chong J, Soufan O, Li C et al (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* 46:W486–W494. <https://doi.org/10.1093/nar/gky310>
- Clasquin MF, Melamud E, Rabinowitz JD (2012) LC-MS Data Processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinformatics*. <https://doi.org/10.1002/0471250953.bi1411s37>
- Codrea MC, Jiménez CR, Heringa J, Marchiori E (2007) Tools for computational processing of LC-MS datasets: a user's perspective. *Comput Methods Programs Biomed* 86:281–290. <https://doi.org/10.1016/j.cmpb.2007.03.001>
- Costa C, Maraschin M, Rocha M (2016) An R package for the integrated analysis of metabolomics and spectral data. *Comput Methods Programs Biomed* 129:117–124. <https://doi.org/10.1016/j.cmpb.2016.01.008>
- Cottret L, Frainay C, Chazalviel M et al (2018) MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gky301>
- Cui Q, Lewis IA, Hegeman AD et al (2008) Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotechnol* 26:162–164. <https://doi.org/10.1038/nbt0208-162>
- Daly R, Rogers S, Wandy J et al (2014) MetAssign: probabilistic annotation of metabolites from LC-MS data using a Bayesian clustering approach. *Bioinformatics* 30:2764–2771. <https://doi.org/10.1093/bioinformatics/btu370>
- Davidson RL, Weber RJM, Liu H et al (2016) Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. *Gigascience* 5:10. <https://doi.org/10.1186/s13742-016-0115-8>
- De Vos RCH, Moco S, Lommen A et al (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2:778–791. <https://doi.org/10.1038/nprot.2007.95>
- Degtyarenko K, De matos P, Ennis M, et al (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 36:D344–D350. <https://doi.org/10.1093/nar/gkm791>
- Dunn WB (2008) Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Biol Phys*. <https://doi.org/10.1088/1478-3975/5/1/011001>
- Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. *TrAC Trends Anal Chem* 24:285–294. <https://doi.org/10.1016/j.trac.2004.11.021>
- Fahy E, Sud M, Cotter D, Subramaniam S (2007) LIPID MAPS online tools for lipid research. *Nucleic Acids Res* 35:W606–W612. <https://doi.org/10.1093/nar/gkm324>
- Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* 2:155–168
- Fiehn O (2002) Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171. <https://doi.org/10.1023/A:1013713905833>
- Fiehn O, Kopka J, Dörmann P et al (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18:1157–1161. <https://doi.org/10.1038/81137>
- Giacomini F, Le Corguille G, Monsoor M et al (2015) Workflow-4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* 31:1493–1495. <https://doi.org/10.1093/bioinformatics/btu813>
- Grandori R, Santambrogio C, Brocca S et al (2009) Electrospray-ionization mass spectrometry as a tool for fast screening of protein structural properties. *Biotechnol J* 4:73–87. <https://doi.org/10.1002/biot.200800250>

- Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol* 169:453–468. <https://doi.org/10.1111/j.1469-8137.2005.01632.x>
- Harbourne N, Marete E, Jacquier JC, O’Riordan D (2009) Effect of drying methods on the phenolic constituents of meadowsweet (*Filipendula ulmaria*) and willow (*Salix alba*). *LWT Food Sci Technol* 42:1468–1473. <https://doi.org/10.1016/j.lwt.2009.05.005>
- Heuberger AL, Robison FM, Lyons SMA et al (2014) Evaluating plant immunity using mass spectrometry-based metabolomics workflows. *Front Plant Sci* 5:291. <https://doi.org/10.3389/fpls.2014.00291>
- Horai H, Arita M, Kanaya S et al (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 45:703–714. <https://doi.org/10.1002/jms.1777>
- Iijima Y, Nakamura Y, Ogata Y et al (2008) Metabolite annotations based on the integration of mass spectral information. *Plant J* 54:949–962. <https://doi.org/10.1111/j.1365-313X.2008.03434.x>
- Ishikawa T, Takahara K, Hirabayashi T et al (2010) Metabolome analysis of response to oxidative stress in rice suspension cells overexpressing cell death suppressor bax inhibitor-1. *Plant Cell Physiol* 51:9–20. <https://doi.org/10.1093/pcp/pcp162>
- Jiang W, Qiu Y, Ni Y et al (2010) An automated data analysis pipeline for GC-TOF-MS metabonomics studies. *J Proteome Res* 9:5974–5981. <https://doi.org/10.1021/pr1007703>
- Kanehisa M, Araki M, Goto S et al (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:D480–D484. <https://doi.org/10.1093/nar/gkm882>
- Katajamaa M, Miettinen J, Orešič M (2006) MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22:634–636. <https://doi.org/10.1093/bioinformatics/btk039>
- Kessner D, Chambers M, Burke R et al (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* 24:2534–2536. <https://doi.org/10.1093/bioinformatics/btn323>
- Khakimov B, Bak S, Engelsens SB (2014) High-throughput cereal metabolomics: current analytical technologies, challenges and perspectives. *J Cereal Sci* 59:393–418
- Kim HK, Choi YH, Verpoorte R (2011) NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol* 29:267–275
- Kim HK, Verpoorte R (2010) Sample preparation for plant metabolomics. *Phytochem Anal* 21:4–13
- Kind T, Fiehn O (2006) Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* 7:234. <https://doi.org/10.1186/1471-2105-7-234>
- Kopka J, Schauer N, Krueger S et al (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21:1635–1638. <https://doi.org/10.1093/bioinformatics/bti236>
- Kovacs H, Moskau D, Spraul M (2005) Cryogenically cooled probes—a leap in NMR technology. *Prog Nucl Magn Reson Spectrosc* 46:131–155. <https://doi.org/10.1016/j.pnmrs.2005.03.001>
- Kuhl C, Tautenhahn R, Böttcher C et al (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 84:283–289. <https://doi.org/10.1021/ac202450g>
- Kujawinski EB (2002) Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS): characterization of complex environmental mixtures. *Environ Forensics* 3:207–216. <https://doi.org/10.1006/enfo.2002.0109>
- Lenz EM, Bright J, Knight R et al (2004) Cyclosporin A-induced changes in endogenous metabolites in rat urine: a metabolomic investigation using high field 1H NMR spectroscopy, HPLC-TOF/MS and chemometrics. *J Pharm Biomed Anal* 35:599–608. <https://doi.org/10.1016/j.jpba.2004.02.013>
- Liang YJ, Lin YT, Chen CW et al (2016) SMART: statistical metabolomics analysis—an R Tool. *Anal Chem* 88:6334–6341. <https://doi.org/10.1021/acs.analchem.6b00603>
- Liang D, Liu Q, Zhou K et al (2020) IP4M: an integrated platform for mass spectrometry-based metabolomics data mining. *BMC Bioinforma* 21:1–16. <https://doi.org/10.1186/S12859-020-03786-X3>
- Liland KH (2011) Multivariate methods in metabolomics—from pre-processing to dimension reduction and statistical analysis. *TrAC Trends Anal Chem* 30:827–841. <https://doi.org/10.1016/j.trac.2011.02.007>
- Markley JL, Anderson ME, Cui Q et al (2007) New bioinformatics resources for metabolomics. *Pacific Symp Biocomput* 12:157–168. https://doi.org/10.1142/9789812772435_0016
- Martens L, Chambers M, Sturm M et al (2011) mzML—a community standard for mass spectrometry data. *Cell. Proteomics Mol.* <https://doi.org/10.1074/mcp.R110.000133>
- Moco S, Bino RJ, Vorst O et al (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* 141:1205–1218. <https://doi.org/10.1104/pp.106.078428>
- Murtagh F (1983) A survey of recent advances in hierarchical clustering algorithms. *Comput J* 26:354–359. <https://doi.org/10.1093/comjnl/26.4.354>
- Naithani S, Gupta P, Preece J et al (2020) Plant Reactome: a knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res* 48:D1093–D1103. <https://doi.org/10.1093/nar/gkz996>
- Novikova DD, Cherenkov PA, Sizentsova YG, Mironova VV (2020) metaRE R package for meta-analysis of transcriptome data to identify the cis-regulatory code behind the transcriptional reprogramming. *Genes* 11:634. <https://doi.org/10.3390/genes11060634>
- Ogata H, Goto S, Fujibuchi W, Kanehisa M (1998) Computation with the KEGG pathway database. *BioSystems* 47:119–128. [https://doi.org/10.1016/S0303-2647\(98\)00017-3](https://doi.org/10.1016/S0303-2647(98)00017-3)
- Okazaki Y, Saito K (2012) Recent advances of metabolomics in plant biotechnology. *Plant Biotechnol Rep.* <https://doi.org/10.1007/s11816-011-0191-2>
- Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16:373–378. [https://doi.org/10.1016/S0167-7799\(98\)01214-1](https://doi.org/10.1016/S0167-7799(98)01214-1)
- Pang Z, Chong J, Zhou G et al (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkab382>
- Perez-Riverol Y, Wang R, Hermjakob H et al (2014) Open source libraries and frameworks for mass spectrometry based proteomics: a developer’s perspective. *Biochim Biophys Acta Proteins Proteomics* 1844:63–76. <https://doi.org/10.1016/j.bbapap.2013.02.032>
- Pluskal T, Castillo S, Villar-Briones A, Orešič M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395. <https://doi.org/10.1186/1471-2105-11-395>
- Raina R, Hall P (2008) Comparison of gas chromatography-mass spectrometry and gas chromatography-tandem mass spectrometry with electron ionization and negative-ion chemical ionization for analyses of pesticides at trace levels in atmospheric samples. *Anal Chem Insights* 3:111–125. <https://doi.org/10.4137/aci.s1005>
- Reo NV (2002) NMR-based metabolomics drug and chemical toxicology. Taylor Francis. <https://doi.org/10.1081/dct-120014789>
- Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* 61:463–489. <https://doi.org/10.1146/annurev.arplant.043008.092035>
- Sakurai N, Ara T, Ogata Y et al (2011) KaPPA-View4: a metabolic pathway database for representation and analysis of correlation

- networks of gene co-expression and metabolite co-accumulation and omics data. *Nucleic Acids Res* 39:D677–D684. <https://doi.org/10.1093/nar/gkq989>
- Scheltema RA, Jankevics A, Jansen RC et al (2011) PeakML/mzMatch: a File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis. *Anal Chem* 83:2786–2793. <https://doi.org/10.1021/AC2000994>
- Scossa F, Brotman Y, de Abreu e Lima F, et al (2016) Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. *Plant Sci* 242:47–64. <https://doi.org/10.3390/ijms17060767>
- Seaver SMD, Henry CS, Hanson AD (2012) Frontiers in metabolic reconstruction and modeling of plant genomes. *J Exp Bot* 63:2247–2258. <https://doi.org/10.1093/jxb/err371>
- Shinbo Y, Nakamura Y, Altaf-Ul-Amin M et al (2006) KNApSACk: A comprehensive species-metabolite relationship database. *Bio-technology in Agriculture and Forestry*. Springer
- Smith CA, Maille GO, Want EJ et al (2005) METLIN. *Ther Drug Monit* 27:747–751. <https://doi.org/10.1097/01.ftd.0000179845.53213.39>
- Smith CA, Want EJ, O'Maille G et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78:779–787. <https://doi.org/10.1021/ac051437y>
- Soga T, Baran R, Suematsu M et al (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. *J Biol Chem* 281:16768–16776. <https://doi.org/10.1074/jbc.M601876200>
- Soga T, Igarashi K, Ito C et al (2009) Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. *Anal Chem* 81:6165–6174. <https://doi.org/10.1021/ac900675k>
- Sturm M, Bertsch A, Gröpl C et al (2008) OpenMS—an open-source software framework for mass spectrometry. *BMC Bioinformatics* 9:1–11. <https://doi.org/10.1186/1471-2105-9-163>
- Suhre K, Kastenmüller G, Römisch-Margl W et al (2011) Meta P-server: a web-based metabolomics data analysis tool. *J Biomed Biotechnol*. <https://doi.org/10.1155/2011/839862>
- Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62:817–836. [https://doi.org/10.1016/s0031-9422\(02\)00708-2](https://doi.org/10.1016/s0031-9422(02)00708-2)
- Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 84:5035–5039. <https://doi.org/10.1021/ac300698c>
- Thimm O, Bläsing O, Gibon Y et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37:914–939. <https://doi.org/10.1111/j.1365-3113X.2004.02016.x>
- Tokimatsu T, Sakurai N, Suzuki H et al (2005) KaPPA-View: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol* 138:1289–1300. <https://doi.org/10.1104/pp.105.060525>
- Tolstikov VV, Fiehn O (2002) Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal Biochem* 301:298–307. <https://doi.org/10.1006/abio.2001.5513>
- Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). *J Chemom* 16:119–128. <https://doi.org/10.1002/cem.695>
- Tsesmetzis N, Couchman M, Higgins J et al (2008) Arabidopsis reactome: a foundation knowledgebase for plant systems biology. *Plant Cell* 20:1426–1436. <https://doi.org/10.1105/tpc.108.057976>
- Tsugawa H, Cajka T, Kind T et al (2015) MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* 12:523–526. <https://doi.org/10.1038/nmeth.3393>
- Tsugawa H, Ikeda K, Takahashi M et al (2020) A lipidome atlas in MS-DIAL 4. *Nat Biotechnol* 38:1159–1163. <https://doi.org/10.1038/s41587-020-0531-2>
- Uppal K, Walker DI, Jones DP (2017) xMSannotator: an R package for network-based annotation of high-resolution metabolomics data. *Anal Chem* 89:1063–1067. <https://doi.org/10.1021/acs.analchem.6b01214>
- Urano K, Maruyama K, Ogata Y et al (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. *Plant J* 57:1065–1078. <https://doi.org/10.1111/j.1365-3113X.2008.03748.x>
- Verpoorte R, Choi YH, Mustafa NR, Kim HK (2008) Metabolomics: back to basics. *Phytochem Rev*. <https://doi.org/10.1007/s11101-008-9091-7>
- Wang M, Carver JJ, Phelan VV et al (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 34:828–837
- Wanichthanarak K, Fan S, Grapov D et al (2017) Metabox: a toolbox for metabolomic data analysis, interpretation and integrative exploration. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0171046>
- Weljie AM, Newton J, Mercier P et al (2006) Targeted profiling: quantitative analysis of ¹H NMR metabolomics data. *Anal Chem* 78:4430–4442. <https://doi.org/10.1021/ac060209g>
- Wheeler DL, Barrett T, Benson DA et al (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36:13–21. <https://doi.org/10.1093/nar/gkm1000>
- Wink M (2010) Introduction biochemistry physiology and ecological functions of secondary metabolites. In *Biochemistry of Plant Secondary Metabolism*. Wiley, Oxford
- Wishart DS, Knox C, Guo AC et al (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 37:D603–610. <https://doi.org/10.1093/nar/gkn810>
- Xia J, Wishart DS (2010) MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* 38:W71–W77. <https://doi.org/10.1093/nar/gkq329>
- Xia J, Wishart DS, Valencia A (2011) MetPA: A web-based metabolomics tool for pathway analysis and visualization. In *Bioinformatics*. Oxford University Press
- Xu F, Zou L, Ong CN, Zou L (2010) Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatization-based GC-MS metabolomics. *TrAC Trends Anal Chem* 29:269–280. <https://doi.org/10.1016/j.trac.2009.12.007>
- Yasugi E, Watanabe K (2002) LIPIDBANK for Web, the newly developed lipid database. *Tanpakushitsu Kakusan Koso. Protein, nucleic acid, enzyme* 47:837–841
- Yu T, Park Y, Johnson JM, Jones DP (2009) apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* 25:1930–1936. <https://doi.org/10.1093/bioinformatics/btp291>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.