



Database tool

MMHub, a database for the mulberry metabolome

DongLi^{1,†}, Bi Ma^{1,†}, Xiaofei Xu^{2,†}, Guo Chen¹, Tian Li¹ and Ningjia He^{1,*}

¹State Key Laboratory of Silkworm Genome Biology, Southwest University, No. 2, Tiansheng Road, Beibei, Chongqing 400715, China and ²College of Computer and Information Science, Southwest University, No. 2, Tiansheng Road, Beibei, Chongqing 400715, China

*Corresponding author: Ningjia He, State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China. Tel: +86-023-68250797; Fax: +86-023-68251128; Email: hejia@swu.edu.cn

†These authors contributed equally to this work.

Citation details: Li,D., Ma,B., Xu,X. *et al.* MMHub, a database for the mulberry metabolome. *Database* (2020) Vol. 2020: article ID baaa011; doi:10.1093/database/baaa011

Abstract

Mulberry is an important economic crop plant and traditional medicine. It contains a huge array of bioactive metabolites such as flavonoids, amino acids, alkaloids and vitamins. Consequently, mulberry has received increasing attention in recent years. MMHub (version 1.0) is the first open public repository of mass spectra of small chemical compounds (<1000 Da) in mulberry leaves. The database contains 936 electrospray ionization tandem mass spectrometry (ESI-MS²) data and lists the specific distribution of compounds in 91 mulberry resources with two biological duplicates. ESI-MS² data were obtained under non-standardized and independent experimental conditions. In total, 124 metabolites were identified or tentatively annotated and details of 90 metabolites with associated chemical structures have been deposited in the database. Supporting information such as PubChem compound information, molecular formula and metabolite classification are also provided in the MS² spectral tag library. The MMHub provides important and comprehensive metabolome data for scientists working with mulberry. This information will be useful for the screening of quality resources and specific metabolites of mulberry.

Database URL: <https://biodb.swu.edu.cn/mmdb/>

Introduction

Mulberry (*Morus* spp.) is widely recognized as an important economic crop plant and is irreplaceable in the sericulture industry. Recently, increasing attention has been paid to mulberry bioactive metabolites, which have a variety of pharmacological benefits (1–4). These metabolites include

1-deoxynojirimycin (DNJ) (5), flavonoids (6, 7), polysaccharides (8) and anthocyanins (9). Any single plant species can produce a huge array of metabolites, with estimates ranging from 5000 to the tens of thousands (10, 11). Mulberry is also considered as a traditional medicine that is rich in bioactive metabolites.

Over the past 10 years, metabolomics based on compound separation and mass spectrometry has evolved from a little-known branch of analytical chemistry to a mainstream enterprise. It is now possible to identify and quantify hundreds of metabolites in a single analysis (12–15). A widely targeted metabolic profiling method (13) has been successfully applied to the studies of rice, maize, qingke and barley metabolomes, providing MS² spectral tag (MS2T) libraries with 840, 983 and 2059 distinct molecular features, respectively. However, these valuable mass spectral data and research products should be presented in a more visual and accessible way than simply listing supporting data with mass-to-charge ratio (m/z) and major peaks in a digital format. Tandem mass spectrum (MS/MS) spectral data are important experimental data that support life science research (16). Although thousands of metabolites with distinct chemical structures and mass spectra are listed in published mass spectral databases like MassBank (17), HMDB (12) and METLIN (18), it is impossible for researchers to screen for metabolites only detected in a particular species from these knowledgebases.

The draft genome sequence of *Morus notabilis* has been reported (19), and a supporting platform (MorusDB) is available for researchers to search and analyze the mulberry genomics and related data (20). However, at present, there is no comprehensive database of information about the composition, distribution and structures of mulberry metabolites. By combining a widely targeted metabolic profiling approach and high-throughput quantitative analysis, we created MMHub (version 1.0), a publicly available mulberry metabolome database of small chemical compounds (<1000 Da). This database is now available for other researchers to search and analyze the distinct mass spectra, distribution and structures of mulberry metabolites.

Materials and Methods

Construction and MMHub

MMHub (version 1.0) was constructed using LAMP (CentOS Linux release 7.4.1708, Apache 2.4.6, MariaDB 5.5.56 and PHP 5.4.16), which is an efficient platform for developing a comprehensive database. All metabolome data and information were stored in a MySQL relational table, which allowed for structural data, spectra data and other useful data about metabolites to be entered manually and indexed as quickly as possible.

Plant materials

Mulberry (*Morus* spp.) resources were cultivated at the Mulberry Germplasm Nursery in Southwest University, Chongqing, China. Leaves collected from the mulberry

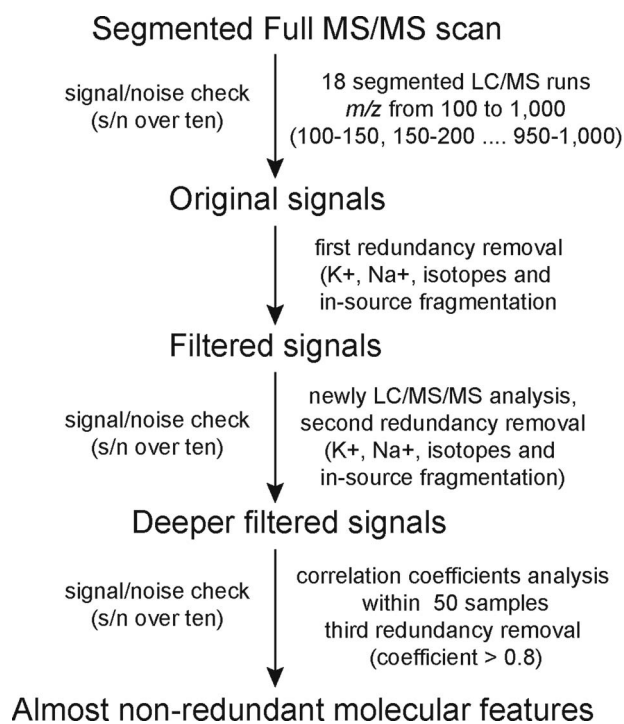


Figure 1. The workflow of MS2T library construction based on HPLC-MS/MS and its practical application in mulberry.

plants were used for MS2T library construction. A total of 91 resources with two replicates were sampled in June 2018. Two or three mature leaves were collected from the longest branch of each resource and frozen in liquid nitrogen. All 91 resources (Table S3) were classified into three species according to Zeng *et al.*'s study (21): 89 *Morus alba* L., one *Morus nigra* and one *M. notabilis*. Samples were then freeze-dried, ground and extracted (22). The sample extracts were analyzed using a liquid chromatography electrospray ionization tandem mass spectrometry system in positive ion mode (Ultra High-performance liquid chromatography, Thermo Scientific™ Dionex™ UltiMate™ 3000; MS, Q Exactive hybrid quadrupole-Orbitrap mass spectrometer; Thermo Fisher Scientific, Waltham, MA, USA).

MS2T library construction

A segmented full MS/MS scan strategy developed from the widely targeted metabolomics approach was used to construct the mulberry MS2T library (Figure 1). This strategy comprised 18 full MS/MS runs (segmented with 50 m/z) instead of thousands of multiple ion monitoring-enhanced product ions transitions in 113 runs (22). Original chromatographic peaks (signals) and related mass spectra were manually checked. Further filtering was conducted to remove potentially redundant signals including those of isotopes, in-source fragmentation products, K⁺, Na⁺ and NH₄⁺ adducts and dimerizations. The signal/noise

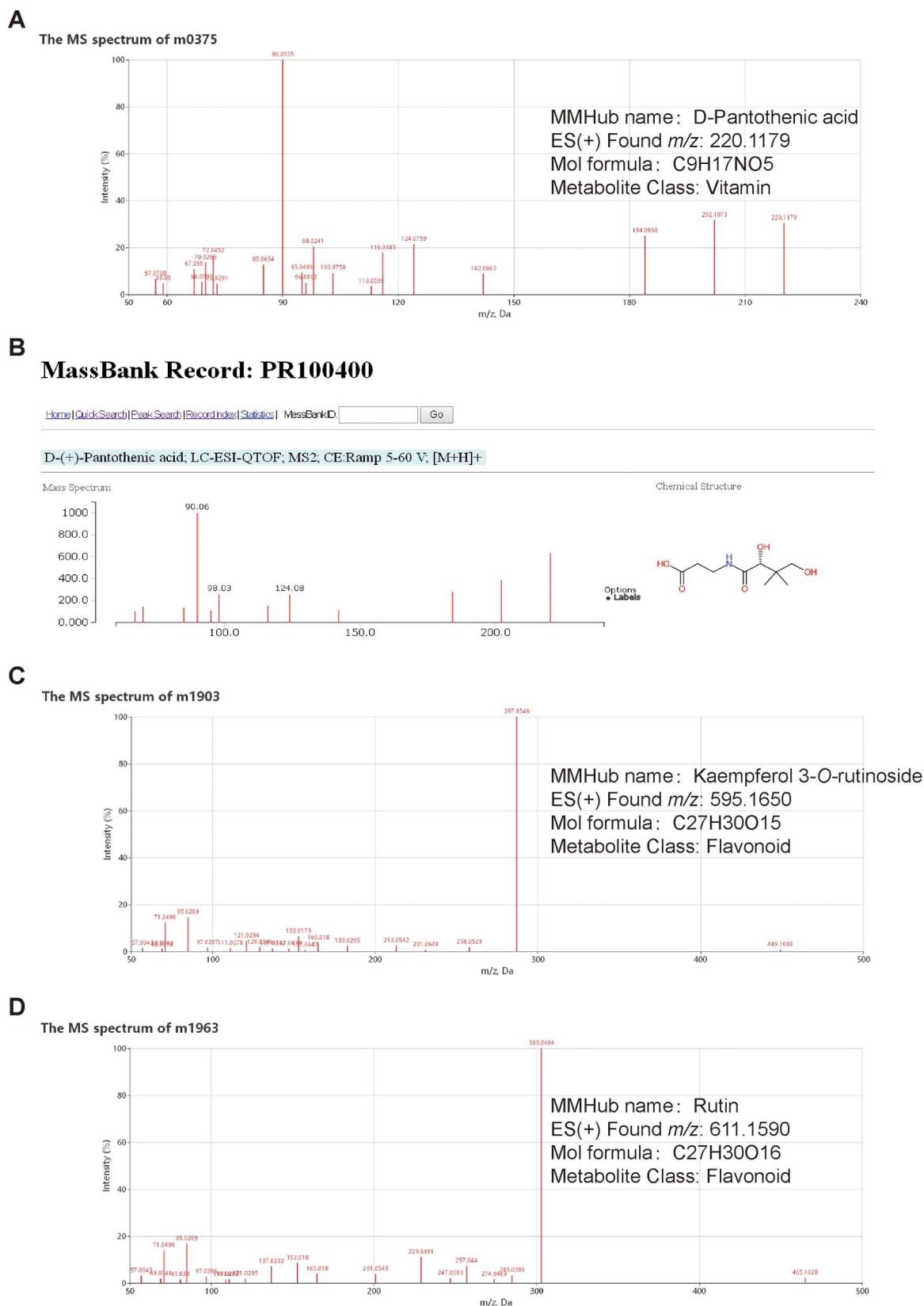


Figure 2. Examples of metabolite identification based on database comparisons and specific fragmentation patterns in MMHub. Molecular feature m0375 (A) was identified as D-pantothenic acid by comparison with MassBank (B). Two molecular features (m1903 and m1963) were annotated as kaempferol 3-*O*-rutinoside (C) and rutin (D) based on flavonoid-specific fragmentation patterns. Identities of three metabolites shown in this figure were confirmed by comparison with authentic standards.

(s/n) check and three rounds of redundancy removal were performed using the same standards with the software that we developed in-house. Supporting information such as PubChem compound information, molecular formula and main fragments were also added to the MS2T library according to Fernie's recommendation (23).

Metabolite identification and annotation

Metabolite identification/annotation was based on the accurate m/z , retention time (RT) and fragmentation patterns. The accuracy of metabolite identification (from high to low) was divided into four levels (A–D). Level A was the most accurate identification, indicating that those metabolites had the same RT (± 0.1 min) and mass spectra as those of authentic standards. Level B indicated those metabolites showed $>85\%$ match rate when their main fragments were searched against public databases (MassBank, KNApSAcK, HMDB and METLIN), or showed specific fragmentation patterns. Metabolites only with confident m/z ($|\text{error}| \leq 10$ ppm) by comparison with references or detected in other species were defined as level C and D (relatively low accuracy). The proportions of metabolites in categories C and D were relatively low.

Quantification of metabolites

To improve the sensitivity and accuracy of quantification, 936 molecular features were analyzed in two m/z ranges (from 100 to 400 and from 400 to 1000). The quantitative calculations were conducted using Thermo Scientific™ Xcalibur™ software v. 2.2. Each metabolite was quantified with accurate mass tolerance (units, 20 ppm) and precision (decimals, 0.0001). To verify data stability and reproducibility, a mixture of 50 randomly chosen extracts was repeatedly analyzed ($n = 6$) as the reference control through the analytical procedure. Each set of 20 samples and the relative content of each metabolite were normalized to the average in the reference control. The relative content of each metabolite in 91 mulberry resources has been deposited in the MMHub as its \log^2 value.

Results

Database description

We analyzed the metabolome of mulberry using a widely targeted metabolomics approach based on high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) (13). By applying this segmented full MS/MS scan strategy (Figure 1), metabolites with mass range from m/z 100 to 1000 were detected with high sensitivity in 18 LC-MS/MS runs. In total, 4924 chromatographic peaks (signals) with $s/n > 10$ were manually

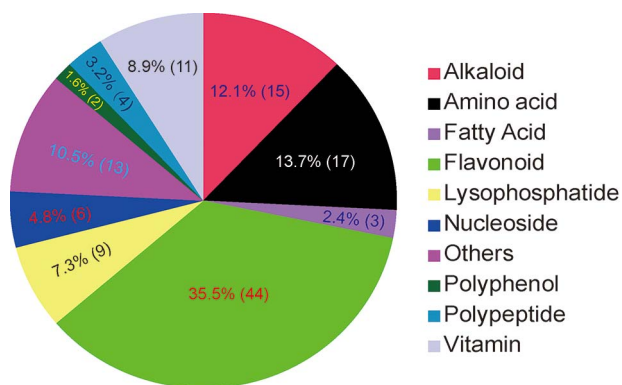


Figure 3. Classification of 124 identified/annotated metabolites in MMHub. The numbers of metabolites in each classification are shown in parentheses.

checked. To produce a matrix with fewer biased and redundant data, redundancies caused by signals from isotopes, in-source fragmentation products, K^+ , Na^+ and NH_4^+ adducts and dimerizations were removed using in-house software written in Perl (22). After the first redundancy removal, 2319 signals were obtained. To produce optimal mass spectra data and remove potential redundancies, a high-concentration mixture of compounds was subjected to LC-MS/MS analysis under the data-dependent MS^2 (dd- MS^2) mode. These new data were run through redundancy removal software (second redundancy removal) and generated a data matrix with 1577 signals.

We expected that transitions derived from the same metabolites would be strongly correlated among different samples. Thus, a pre-test quantitative analysis of 50 mulberry resources was conducted and correlation coefficients between transition pairs were calculated. We found that 641 metabolites with high correlation coefficients (>0.9 for metabolites eluted earlier than 2.5 min or >0.8 for metabolites eluted later than 2.5 min). The same retention time (± 0.1 min) were considered as another criterion of redundancy (third redundancy removal). Finally, an MS2T library containing 936 molecular features with almost no redundancy was obtained and is reported as recommended previously (Tables S1 and S2) (23).

Identification and annotation of metabolites in mulberry leaves


In MMHub (version 1.0), 37 commercially available standards were identified using the same profiling conditions as those used to analyze the extracts. Eighty-seven metabolites were putatively identified by querying MS/MS spectra data against the literature or databases (MassBank, KNApSAcK, HMDB and METLIN) (Figure 2A and B). A number of metabolites were identified by analyzing specific fragmentation patterns. For example, two flavonoids, m1903 (RT

MMHub Home Browse Submit About


Submit your data:

Essential information (marked with *) including peak no., compound name, molecular formula, main MS/MS fragments, and localization in mulberry tissues should be supplied.

Individual submission with essential information



Direct submission using excel



Submit your data:

Essential information (marked with *) including peak no., compound name, molecular formula, main MS/MS fragments, and localization in mulberry tissues should be supplied.

Field	Value
Peak no. *	<input type="text" value="Peak no."/>
Ret. Time (min) *	<input type="text" value="Ret. Time (min)"/>
m/z	<input type="text" value="m/z"/>
m/z (y)	<input type="text" value="m/z (y)"/>
PubChem Compound Identifiers (CID)	<input type="text" value="PubChem Compound Identifiers (CID)"/>
Database comparisons	<input type="text" value="Database comparisons"/>
Putative metabolite name *	<input type="text" value="Putative metabolite name"/>
Metabolite Class	<input type="text" value="Metabolite Class"/>
Mol formula	<input type="text" value="Mol formula"/>
ESI(+)/Threat m/z	<input type="text" value="ESI(+)/Threat m/z"/>
ESI(+)/Fused m/z *	<input type="text" value="ESI(+)/Fused m/z"/>
m/z error (ppm)	<input type="text" value="m/z error (ppm)"/>
MS/MS ESI(+)/fragments *	<input type="text" value="MS/MS ESI(+)/fragments"/>
MS/MS ESI(+)/CE (m/z)	<input type="text" value="MS/MS ESI(+)/CE (m/z)"/>
Species detected below*	<input type="text" value="Species detected below"/>
References	<input type="text" value="References"/>
Identification level (A-D)	<input type="text" value="Identification level (A-D)"/>
Vendor specificity Distribution file (if file exists)	<input type="button" value="Choose File"/> No file chosen
The MS spectrum file (if file exists) *	<input type="button" value="Choose File"/> No file chosen

Submit your data using excel:

Information

Step 1. Download template [Click here download the excel template](#)

Step 2. Up load file No file chosen

Figure 5. User guide for uploading and submission of mass spectra data into MMHub.

keys, PubChem compound identifiers, metabolite name, metabolite class, molecular formula and other information supported in the MS2T library (Tables S1 and S2). Clicking on the link ‘Peak no.’ in the browsing page takes the reader to a page with detailed information about that peak, including basic qualitative data (RT, m/z , main fragments and available compound information) (Figure 4D), quantitative data (specific distribution in 91 mulberry resources with two biological duplicates) (Figure 3E) and original mass spectrum (Figure 4F). A key feature of MMHub is that it provides links for known compounds to the PubChem database, where more features are listed (2D and 3D structures).

As an open metabolome database for mulberry, we expect to accommodate more metabolite data for other mulberry tissues, which can be added by researchers from any institution. By clicking on the ‘Submit’ hyperlink, any user can submit metabolite data to MMHub (Figure 5). Essential information (marked with *) including peak no., compound name, molecular formula, main MS/MS fragments and localization in mulberry tissues should be supplied. Users can choose to submit data manually one by one or directly upload data in file format. To ensure the authenticity and reliability of the data, all uploaded data will be temporarily stored in our local server before being

deposited in MMHub. New supporting data that have been contrasted and filtered against available data in MMHub will be updated regularly and displayed to all users.

Discussion

Mulberry has been identified as a potential functional nutraceutical food in recent years (24). Bioactive compounds in mulberry have been shown to prevent and treat hyperglycemia (25, 26), hyperlipidemia (3, 27), Alzheimer’s disease (28) and cancer (29). Mass spectral data are important experimental data for research on bioactive compounds. MMHub (version 1.0) is the first public repository of mass spectra of small chemical compounds (<1000 Da) in mulberry leaves. Although our ESI-MS² data were obtained under non-standardized and independent experimental conditions, Volna *et al.* found that the fragmentation patterns are almost identical for all tandem mass analyzers and that only the ratios of the product ions differ somewhat (16, 30). For instance, m0167 and m0375 are identified as DNJ and D-pantothenic acid, respectively, in MMHub and show similar fragmentation patterns and main fragments in the HMDB (HMDB ID: HMDB0035359) and MassBank (MassBank record: PR100400) databases (Figure 2B). Actually, 87.1% of 124

identified metabolites (108/124) in MMHub are also listed in other public databases (MassBank, METLIN, HMDB and KNApSAcK).

The widely targeted metabolic profiling method based on HPLC-MS/MS is not limited to a single mulberry species or particular tissue (Figure 1). This approach has been used to construct multiple MS2T libraries including 983 molecular features for maize kernels (22), 840 for rice leaves (31) and 2059 for qingke and barley (14). Primary and secondary metabolites like amino acids, nucleic acids, vitamins, flavonoids and vitamins can be efficiently detected using this high-throughput method. Qualitative and quantitative variations in metabolism are regarded as the ‘metabotype’ (32), which represents the bridge between the genotype and the phenotype of a plant (11). Recently, metabolomics combined with broad profiling approaches like genomics and transcriptomics has become an essential method to explore the diversity of plant metabolism and its underlying molecular mechanisms (22, 33, 34). This comprehensive metabolome database will provide a starting point for research on bioactive metabolites and will be useful for the identification of new candidate genes in mulberry.

The existing public metabolome databases are richly annotated and can meet many of the needs of biochemists, clinical chemists, physicians, medical geneticists, nutritionists and members of the metabolomics community (12). MMHub is a specialized database for mulberry research. Although many metabolites remain unidentified, MMHub provided a comprehensive collection of data for research on the mulberry-specific metabolome. The database also includes related quantitative data for 91 mulberry resources with two biological duplicates. This information may be useful for mulberry breeders and plant biochemists to screen for quality resources and specific metabolites.

MMHub was constructed to help researchers mine for metabolomics information easily and effectively. It has several features and advantages: (i) all metabolomics information, including metabolites, structure, mass spectra and validated metabolite concentrations, are strictly curated and verified; (ii) as an initial metabolomics data repository for mulberry, it can be used to provide basic data for other databases; and (iii) MMHub is a work in progress. Newly identified metabolomics data for mulberry will be integrated into MMHub as quickly as possible, and submission queries from all researchers will be encouraged. The comprehensive metabolomic studies of multiple mulberry tissue including fruit and bark are carrying out. It is particularly exciting given that the tissue-specific metabolites, such as anthocyanidins, will greatly enrich the diversity of mulberry metabolites in MMHub. In addition, the agronomic traits of mulberry resources including leaf type, fruit size

and fruit color will be integrated into the next version of MMHub.

Conclusion

In summary, MMHub is a user-friendly, freely available and comprehensive metabolomics database that brings together data for thousands of endogenous mulberry metabolites. We believe that MMHub is unique. Although MMHub is a work in progress, our aim is to make it a special metabolomics repository for mulberry species. We are committed to continuously updating this database with strictly curated data, including more metabolites and more tissues. We expect that MMHub will be useful for applications in metabolomics, traditional medicines, biomarker discovery and general education.

Availability

All metabolomics data deposited in this database are freely available to any researcher without restrictions.

Database name: MMHub.

Database URL: <https://biodb.swu.edu.cn/mmdb/>

Conflict of interest. The authors declare no competing financial interest.

Acknowledgements

The authors are grateful to all the laboratory members who provided advice during this work.

Funding

National Key Research and Development Program (2018YFD1000602); Chongqing Research Program of Basic Research and Frontier Technology (cstc2018jcyjAX0407); Fundamental Research Funds for the Central Universities (XDJK2017C076 and SWU118040); China Postdoctoral Science Foundation (2017M612884).

References

1. Park,E., Lee,S.M., Lee,J.E. *et al.* (2013) Anti-inflammatory activity of mulberry leaf extract through inhibition of NF-kappa B. *J. Funct. Foods*, 5, 178–186.
2. Jeong,J.H., Lee,N.K., Cho,S.H. *et al.* (2014) Enhancement of 1-deoxyojirimycin content and alpha-glucosidase inhibitory activity in mulberry leaf using various fermenting microorganisms isolated from Korean traditional fermented food. *Biotechnol Bioproc E*, 19, 1114–1118.
3. Hu,X.Q., Thakur,K., Chen,G.H. *et al.* (2017) Metabolic effect of 1-deoxyojirimycin from mulberry leaves on db/db diabetic mice using liquid chromatography-mass spectrometry based metabolomics. *J. Agric. Food Chem.*, 65, 4658–4667.
4. Hunyadi,A., Liktó-Busa,E., Marki,A. *et al.* (2013) Metabolic effects of mulberry leaves: exploring potential benefits in type

- 2 diabetes and hyperuricemia. *Evid-Based Compl Alt*, 2013, 948627.
5. Deng, M.J., Lin, X.D., Wen, C.W. *et al.* (2017) Metabolic changes in the midgut of Eri silkworm after oral administration of 1-deoxyojirimycin: a 1H-NMR-based metabolomic study. *PLoS One*, 12, e0173213.
 6. Zhang, P., Zhu, K.L., Zhang, J. *et al.* (2019) Purification of flavonoids from mulberry leaves via high-speed counter-current chromatography. *Processes*, 7, 91–101.
 7. Kwon, O.C., Ju, W.T., Kim, H.B. *et al.* (2019) UPLC-DAD-QTOF/MS analysis of flavonoids from 12 varieties of Korean mulberry fruit. *J. Food Qual.*, 2019, 1–8
 8. Chen, X., Sheng, Z., Qiu, S. *et al.* (2019) Purification, characterization and *in vitro* and *in vivo* immune enhancement of polysaccharides from mulberry leaves. *PLoS One*, 14, e0208611.
 9. Yan, F.J. and Zheng, X.D. (2017) Anthocyanin-rich mulberry fruit improves insulin resistance and protects hepatocytes against oxidative stress during hyperglycemia by regulating AMPK/ACC/mTOR pathway. *J. Funct. Foods*, 30, 270–281.
 10. Fernie, A.R., Trethewey, R.N., Krotzky, A.J. *et al.* (2004) Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.*, 5, 763–769.
 11. Fang, C.Y., Fernie, A.R. and Luo, J. (2019) Exploring the diversity of plant metabolism. *Trends Plant Sci.*, 24, 83–98.
 12. Wishart, D.S., Knox, C., Guo, A.C. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, 37, D603–D610.
 13. Chen, W., Gong, L., Guo, Z. *et al.* (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant*, 6, 1769–1780.
 14. Zeng, X., Yuan, H., Dong, X. *et al.* (2020) Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. *Mol. Plant*, 13, 112–127.
 15. Tohge, T., Wendenburg, R., Ishihara, H. *et al.* (2016) Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat. Commun.*, 7, 12399.
 16. 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.
 17. Taguchi, R., Nishijima, M. and Shimizu, T. (2007) Basic analytical systems for lipidomics by mass spectrometry in Japan. *Methods Enzymol.*, 432, 185–211.
 18. Smith, C.A., O'Maille, G., Want, E.J. *et al.* (2005) METLIN: a metabolite mass spectral database. *Ther. Drug Monit.*, 27, 747–751.
 19. He, N., Zhang, C., Qi, X. *et al.* (2013) Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat. Commun.*, 4, 2445.
 20. Li, T., Qi, X., Zeng, Q. *et al.* (2014) MorusDB: a resource for mulberry genomics and genome biology. *Database (Oxford)*, 2014, 1–8.
 21. Zeng, Q., Chen, H., Zhang, C. *et al.* (2015) Definition of eight mulberry species in the genus *Morus* by internal transcribed spacer-based phylogeny. *PLoS One*, 10, e0135411.
 22. Wen, W., Li, D., Li, X. *et al.* (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.*, 5, 3438.
 23. Fernie, A.R., Aharoni, A., Willmitzer, L. *et al.* (2011) Recommendations for reporting metabolite data. *Plant Cell*, 23, 2477–2482.
 24. Srivastava, S., Kapoor, R., Thathola, A. *et al.* (2003) Mulberry (*Morus alba*) leaves as human food: a new dimension of sericulture. *Int. J. Food Sci. Nutr.*, 54, 411–416.
 25. Kimura, T., Nakagawa, K., Kubota, H. *et al.* (2007) Food-grade mulberry powder enriched with 1-deoxyojirimycin suppresses the elevation of postprandial blood glucose in humans. *J. Agric. Food Chem.*, 55, 5869–5874.
 26. Katsube, T., Imawaka, N., Kawano, Y. *et al.* (2006) Antioxidant flavonol glycosides in mulberry (*Morus alba* L.) leaves isolated based on LDL antioxidant activity. *Food Chem.*, 97, 25–31.
 27. Chen, J.J. and Li, X.R. (2007) Hypolipidemic effect of flavonoids from mulberry leaves in triton WR-1339 induced hyperlipidemic mice. *Asia Pac. J. Clin. Nutr.*, 16, 290–294.
 28. Niidome, T., Takahashi, K., Goto, Y. *et al.* (2007) Mulberry leaf extract prevents amyloid beta-peptide fibril formation and neurotoxicity. *Neuroreport*, 18, 813–816.
 29. Chen, A.Y. and Chen, Y.C. (2013) A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem.*, 138, 2099–2107.
 30. Volna, K., Holcapek, M., Kolarova, L. *et al.* (2008) Comparison of negative ion electrospray mass spectra measured by seven tandem mass analyzers towards library formation. *Rapid Commun. Mass Spectrom.*, 22, 101–108.
 31. Chen, W., Gao, Y.Q., Xie, W.B. *et al.* (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.*, 46, 714–721.
 32. Suhre, K. and Gieger, C. (2012) Genetic variation in metabolic phenotypes: study designs and applications. *Nat. Rev. Genet.*, 13, 759–769.
 33. Fernie, A.R. and Tohge, T. (2017) The genetics of plant metabolism. *Annu. Rev. Genet.*, 51, 287–310.
 34. Polturak, G., Heinig, U., Grossman, N. *et al.* (2018) Transcriptome and metabolic profiling provides insights into Betalain biosynthesis and evolution in *Mirabilis jalapa*. *Mol. Plant*, 11, 189–204.