

RESEARCH

Open Access



# Metabolic profile analysis based on GC-TOF/MS and HPLC reveals the negative correlation between catechins and fatty acids in the cottonseed of *Gossypium hirsutum*

MA Lei<sup>1,2,3\*</sup>, CHEN Yabing<sup>1</sup>, XU Shuangjiao<sup>1,2</sup>, DONG Ruidan<sup>1</sup>, WANG Yiwen<sup>1</sup>, FANG Dan<sup>1</sup>, PENG Jun<sup>1,2\*</sup> and TIAN Xinquan<sup>1,2\*</sup>

## Abstract

**Background:** The diversified and high value-added utilization of cotton by products can promote the sustainable development of modern agriculture. Differences in potential nutrients among varieties can be explained by variations in the composition and abundance of fatty acids, polyphenols, carbohydrates, amino acids, and organic acids. Therefore, the analysis of metabolite species and relationships in cottonseed is meaningful for the development of cotton byproducts.

**Results:** In this study, the metabolomes of three representative cotton cultivars of different species were compared using untargeted GC-TOF/MS analysis. A total of 263 metabolites were identified from 705 peaks, and their levels were compared across cultivars. Principal component analysis and OPLS-DA clearly distinguish these samples based on metabolites. There were significant differences in the contents of amino acids, carbohydrates, organic acids, flavonoids, and lipids in *G. hirsutum* TM-1 compared with *G. arboreum* Shixiya1 and *G. barbadense* Hai7124. Notably, the bioactive nutrient compound catechin obtained from the differential metabolites significantly accumulated in TM-1. Furthermore, a comprehensive analysis using catechin and oil-related traits was conducted in core collections of *Gossypium hirsutum*. The results revealed the reliability of the GC-TOF/MS analysis, as well as that catechin content has a negative association with myristic acid, palmitic acid, stearic acid, oleic acid, linoleic acid, arachidic acid, and total fatty acids.

**Conclusion:** These findings suggest that untargeted GC-TOF/MS analysis could provide a new method for investigating the underlying plant biochemistry of nutrient variation in cottonseed, and that catechin content has a negative association with oil-related traits in cottonseed. This study may pave the way to exploit the value of cotton byproducts.

**Keywords:** Cotton, Metabolic profile, GC-TOF/MS, Catechin, Fatty acid

## Introduction

Cotton is grown in over 75 countries mainly for its fiber or lint, the most important natural and renewable material for the textile industry. However, other parts of the cotton plant also have commercial potential, such as the seeds, stalks, hulls, and flowers (Egbuta et al. 2017;

\*Correspondence: malei@caas.cn; jun\_peng@126.com; tianxq0918@163.com

<sup>1</sup> Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Henan 450001, Zhengzhou, China

Full list of author information is available at the end of the article



© The Author(s) 2022. **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/>.

Xu et al. 2021; Yuan et al. 2016). These are commonly referred to as cotton byproducts, and are widely used to obtain edible oil and protein feed in cotton producing countries. The nutritional and functional quality of cottonseeds is far from meeting the requirements of the diversified development of agricultural products. Cottonseed is a non-fresh agricultural product with quality characteristics that often been ignored in the process of planting and primary processing. And the latest cutting-edge technology is rarely applied to cotton byproduct, resulting in its low level of utilization.

Plants can produce a variety of metabolites with structurally diverse, which play an essential role in growth and development (Saito 2009; Saito and Matsuda 2010; Wang et al. 2019). In addition, these metabolites provide necessary and sufficient resources for human and animal nutrition, bioenergy, and medicine (Jacobs et al. 2021; Sharma et al. 2021). Understanding plant biochemistry and phytochemistry is thus of fundamental importance for sustainable agriculture and resource conservation. Metabolomics technology can be used to analyze the types and contents of small molecule metabolites in samples, providing a new method for us to study metabolic diversity to evaluate the nutritional and functional quality of crops. The development of modern analytical instruments with high resolution and high sensitivity has elevated the rapid development of metabolomics. Gas chromatography/mass spectrometry (GC/MS), as the most mature method has been used to separate metabolites occurring in urine or tissue extracts, and was also widely used in plant metabolome analysis (Dalglish et al. 1966; Beale et al. 2018; Choudhury et al. 2022). Liquid chromatography/mass spectrometry (LC/MS) and nuclear magnetic resonance (NMR) have also been developed as the main analytical techniques for metabolomics in recent years (Rochfort 2005; Razzaq et al. 2019; Patel et al. 2021). Due to its high resolution, high sensitivity, large number of mass spectral libraries, good reproducibility, and relatively low cost, GC/MS is often used for the analysis of volatile and semivolatile metabolites with relative low molecular mass, low polarity, low boiling point or volatile compounds after derivatization. High-resolution time-of-flight (TOF) mass spectrometry can not only obtain the mass spectrum of compounds, but also accurately detect each fragment ion.

There are only four *Gossypium* species producing commercially spinnable fibers in the world, among which allotetraploid cultivars of *G. hirsutum* (upland cotton) accounts for approximately 90% of global cotton production, with most of the remaining production coming from *G. barbadense* (sea island cotton) (Mansoor and Paterson 2012; Zhang et al. 2015). The diploid cultivars of *G. arboreum* and *G. herbaceum* currently have

no agricultural output. At present, there are few literature reports on the use of metabolomics techniques to evaluate nutrients, bioactivity, and health benefits for cottonseed, and fewer comparative studies among cotton species. In this study, to better understand the metabolite variations among cotton species, untargeted GC-TOF/MS analysis was performed to identify and quantify metabolites including fatty acids, polyphenols, carbohydrates, organic acids, and amino acids in three representative cultivars. Then, classical chromatographic analysis was used to further verify the test results. Last, correlation analysis of micronutrient catechin with oil-related traits was conducted to explore the underlying relationship between nutritional quality in core collections of *Gossypium*. This study provides a reference for exploiting the value of cotton byproducts and the development of functional foods from cottonseed in the future.

## Materials and methods

### Plant material

Three cultivars Shixiya1, Hai7124, and TM-1, which belong to *Gossypium arboreum* (*G. arboreum*), *Gossypium barbadense* (*G. barbadense*) and *Gossypium hirsutum* (*G. hirsutum*), respectively, were used for gas chromatography combined with time-of-flight mass spectrometry (GC-TOF/MS) analysis. These representative cultivars have complete genome sequences and a clear genetic background; they have also been intensively studied. The cotton was planted at Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China, and all cultivars were cultivated under same conditions. The seeds of the three cotton cultivars were harvested in the same period. Then, the samples were freeze-dried and stored at  $-80^{\circ}\text{C}$  until GC-TOF/MS analysis.

To illustrate the content distribution of catechin across different species, 47 accessions of *G. arboreum*, 37 accessions of *G. barbadense*, and 144 accessions of *G. hirsutum* were chosen from core collections of *Gossypium* used for measuring catechin content by HPLC analysis. Moreover, to explore the correlation between catechin and fatty acids in *G. hirsutum*, 144 accessions were measured for fatty acids by GC analysis.

### Sample extract preparation and GC-TOF/MS analysis

Briefly, cottonseed sample powder (50 mg) was extracted with 0.48 mL 75% methanol containing 10  $\mu\text{L}$  adonitol ( $0.5\text{ mg}\cdot\text{mL}^{-1}$  stock in  $\text{dH}_2\text{O}$ ) as an internal standard. The resulting mixture was ultrasound treated for 5 min and then centrifuged ( $12\,000\text{ r}\cdot\text{min}^{-1}$ ) at  $4^{\circ}\text{C}$  for 15 min. The supernatant (0.4 mL) was transferred to a new 2 mL GC/MS glass bottle and dried completely in a vacuum concentrator. Then, the extracts were oximated using 80  $\mu\text{L}$

methoxyamine hydrochloride (20 mg·mL<sup>-1</sup> in pyridine) at 80 °C for 25 min. Subsequently, 100 µL BSTFA reagent (1% TMCS, v/v) was added to the samples and incubated at 70 °C for 1.5 h. GC-TOF/MS was performed as described by Deng et al (2020). Briefly, the GC-TOF/MS analysis was performed using an Agilent 7890 gas chromatograph (GC, Agilent Technologies, USA) system coupled with a Pegasus HT time-of-flight mass spectrometer (Leco, USA). The system utilized a DB-5MS capillary column (J&W Scientific, Folsom, CA, USA). A total of 1 µL samples were injected by the autosampler into GC. Helium was used as the carrier gas, the front inlet purge flow was 3 mL·min<sup>-1</sup>, and the gas flow rate through the column was 1 mL·min<sup>-1</sup>. The initial temperature was kept at 50 °C for 5 min, then raised to 210 °C at a rate of 3 °C·min<sup>-1</sup>, and held for 3 min. The injection, transfer line, and ion source temperatures were set at 280 °C, 280 °C, and 250 °C, respectively. The energy was 70 eV in electron impact mode.

#### Data preprocessing and annotation

Chroma TOF 4.3X software (LECO Corporation) and the LECO-Fiehn Rtx5 database were used to preprocess and annotate the GC-TOF/MS analysis data. In addition, the mass spectrum match and retention index match were also noted in metabolite identification (Kind et al. 2009).

#### Catechin extraction and analysis by high performance liquid chromatography (HPLC)

Briefly, approximately 100 mg cottonseed powder was extracted using 3 mL of 80% acidified methanol (1% hydrochloric acid). After 30 min of sonication, the homogenate was incubated for 12 h at 25 °C. Then samples were centrifuged at 10 000 r·min<sup>-1</sup> for 15 min. The supernatant was used by an Agilent 1100 series HPLC system (Agilent Technologies, USA) to determine the content of catechin (Peng et al. 2018). The elution contained A (0.2% formic acid/water) and B (methanol). The linear gradient program was as follows: 0–5 min, 5%~20% B; 5–18 min, 20%~25% B; 18–25 min, 25%~42%; and 25–32 min, 42% B. The flow rate was 1 mL·min<sup>-1</sup>. 10 µL of supernatant sample was injected. The column temperature was 30 °C and the detection wavelength was set at 278 nm.

#### Fatty acids extracted and analyzed by gas chromatography (GC)

The extraction of fatty acids from mature cottonseeds was performed as described previously. Briefly, approximately 50 mg of cottonseed powder was added to 1 mL of 0.3 mol·L<sup>-1</sup> potassium hydroxide methanol solution, and 1 mL of n-hexane containing 500 µg·mL<sup>-1</sup> C11:0 was used as an internal standard. After shaking for 30 s and

holding at 25 °C for 1.5 h, the homogenate was added 1.5 mL of 0.9% (w/v) sodium chloride solution, followed by centrifugation (10 000 r·min<sup>-1</sup>) at 25 °C for 5 min. The supernatant was separated by gas chromatography (Agilent 7890-FID) according to previously described procedures (Dowd et al. 2010). Briefly, the gas chromatography was fitted with a DB-23 capillary column (Agilent Technologies, USA). Injectors were operated in split mode with a split ratio of 1:100. The temperature of the injection port was set at 240 °C. Helium was used as the carrier gas, and the injection volumes were 1 µL. The column temperature was held at 170 °C for 5 min, and increased to 180 °C at 1 °C·min<sup>-1</sup>, and then increased to 240 °C with a rate of 4 °C·min<sup>-1</sup>, and held for 5 min.

#### Multivariate analysis and statistics

For better visualization and subsequent analysis, GC-TOF/MS data were used to perform Principal component analysis (PCA) and OPLS-DA analysis by SIMCA software (V14.1, MKS Data Analytics Solutions, Umea, Sweden). *Student's t-test* and one-way ANOVA were used to compare differences between two groups and multiple groups, respectively. Pearson's correlation coefficient analysis was performed by OriginPro 2021 (<https://www.originlab.com/>) to assess the correlation between catechin and fatty acid content. *P* values < 0.05 or < 0.01 or < 0.001 were considered statistically significant.

## Results and discussion

#### Characterization and analysis of GC-TOF/MS data

Metabolomics analysis was employed to elucidate the potential chemical basis of different species of the *Gossypium* genus. Three representative cultivars of cotton species, Shixiya1, Hai7124, and TM-1, were used for untargeted metabolome analysis by GC-TOF/MS, each with six biological replicates. A typical total ion chromatogram (TIC) from these cottonseed samples is shown in Additional file 1: Fig. S1, representing the summed intensity of all mass spectral peaks at every point in the analysis. From this step, obvious differences could be found in some TIC peaks across samples. A total of 705 peaks were extracted; these are listed in Additional file 2: Table S1. Based on the local metabolite database, 263 metabolites were identified and classified into 15 different categories (Additional file 3: Fig. S2, Additional file 4: Table S2), including 53 organic acids, 44 carbohydrates, 43 amino acid derivatives, 34 other metabolites, 25 alcohols and polyols, 14 lipids, 10 benzene and substituted derivatives, 9 nucleotide and its derivatives, 7 amines, 6 phenylpropanoids, 6 flavonoids, 3 sphingolipids, 2 quinates and their derivatives, 6 alkaloids, and 1 vitamin. The above results indicate that many types of metabolites

were obtained, consistent with the characteristics of metabolomics data based on GC-TOF/MS.

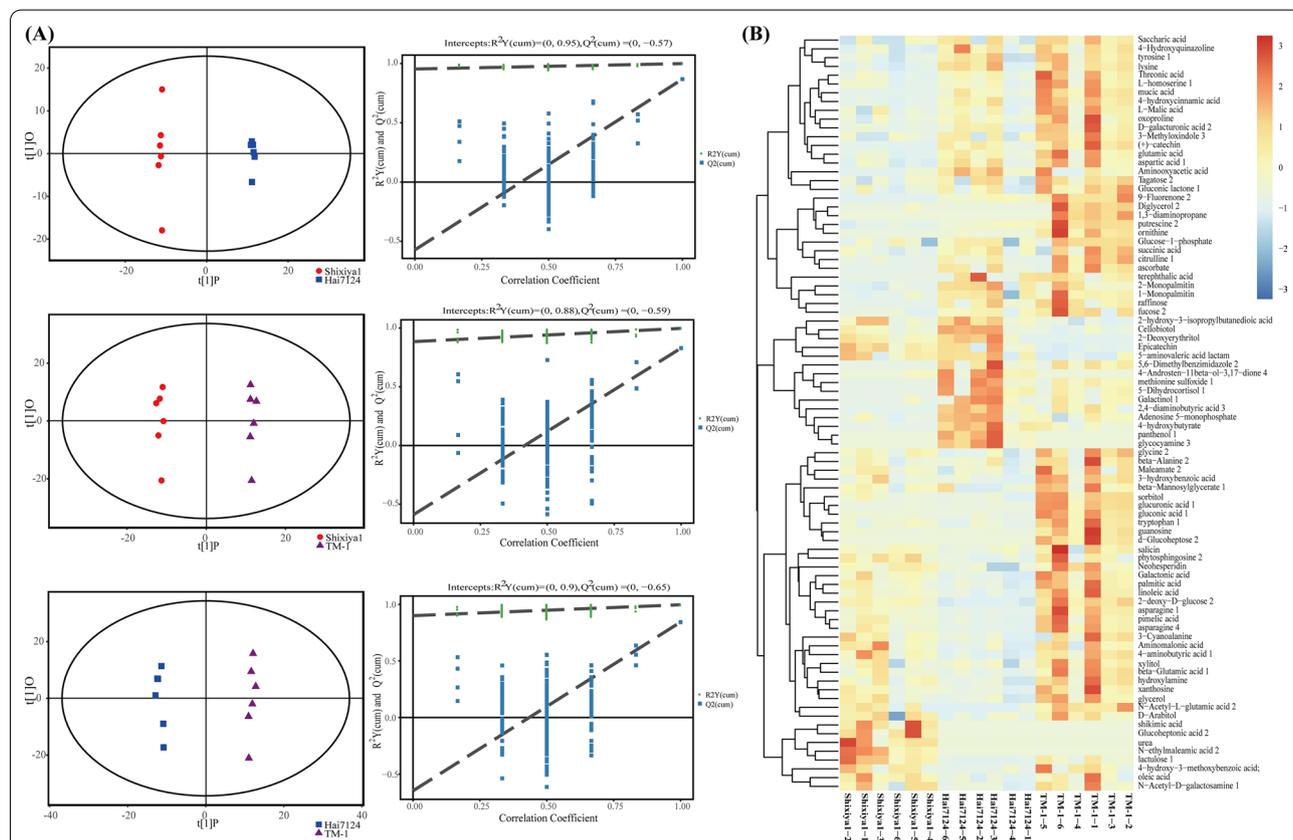
**Identification of differential metabolites from three cotton cultivars**

Principal component analysis (PCA) is the most common dimensionality reduction method. It converts a large number of variables into principal components (PCs) that still contain most of the information found in the large set. PCA is an unsupervised mode; thus, it shows the distribution of the origin data and is used to evaluate the difference for intergroup comparisons. As illustrated in Additional file 5: Fig. S3, the PCA score scatter plot indicated significant differences in the three cotton cultivars, and all the samples were within the 95% confidence intervals (Hotelling’s T-squared ellipse). PC1 and PC2 accounted for 29.1% and 14.7% of the total variation, respectively.

The orthogonal partial least square-discriminant analysis (OPLS-DA), as a supervised mode, was conducted to provide a precise level of group separation, as well as to

test the correlation of samples. It is apparent from Fig. 1A that, OPLS-DA was better for separating the differences in the metabolic phenotypes between three varieties of *Gossypium*. Furthermore, sevenfold cross validation and permutation tests were used to estimate the effectiveness of the model. The R2 value was close to 1 and the Q2 value was negative, indicating that the model was reliable and that the risk of overfitting was low. These results suggested that there were significant differences in the metabolites of each cultivar.

Subsequently, cutoffs for variable importance in the projection (VIP) ≥ 1 obtained from OPLS-DA and P value < 0.05 were used for screening the differential metabolites (DMs). A volcano plot was employed to visualize the differential metabolites (Additional file 6: Fig. S4). The results showed that there were 119, 131, and 128 DMs between the two cultivars, respectively. Specifically, the Shixiya1/Hai7124 comparison had 64 upregulated (red scatter points) and 55 downregulated (blue scatter points) DMs (Additional file 6: Fig. S4A, Additional file 7: Table S3). The Shixiya1/TM-1 comparison



**Fig. 1** OPLS-DA analysis and heatmap of hierarchical clustering analysis of differential metabolites for three cotton cultivars. **A** Score scatter plot of OPLS-DA between each two cultivars. **B** The differential metabolites from three cotton cultivars used to draw the heatmap of hierarchical clustering analysis (HCA). The horizontal axis and vertical axis represent the different cottonseed samples and different metabolites, respectively. Each relative metabolite content is listed in Additional file 10: Table S6

had 29 upregulated and 102 downregulated DMs (Additional file 6: Fig. S4B, Additional file 8: Table S4) and the Hai7124/TM-1 comparison had 26 upregulated and 102 downregulated DMs (Additional file 6: Fig. S4C, Additional file 9: Table S5). In addition, one-way ANOVA was also conducted for the comparative analysis of metabolic differences between multiple experimental groups, and hierarchical cluster analysis (HCA) of the selected differential metabolites ( $P$  value < 0.05) was used to classify metabolites with similar characteristics and further study intergroup variation (Fig. 1B). The relative contents of metabolites represented by colored segments at the corresponding locations are showed in Additional file 10: Table S6. We observed a significant enrichment of metabolites in TM-1 compared to Hai7124 or Shixiya1. Then, we classified these metabolites and found that they included carbohydrates, amino acids, lipids, organic acids and flavonoids. These results indicated that the seeds of *G. hirsutum* TM-1 may have high levels of these metabolites, and could be used as raw materials for processing and utilization in the future.

The seeds of higher plants can be described as valuable factories capable of converting photosynthetically derived sugars into a variety of storage compounds, such as oils, proteins, sugars, and secondary metabolites (Baud 2018). *G. hirsutum* is the most widely planted cotton cultivar and makes up the absolute majority of cotton fiber yield and cottonseed byproducts worldwide. Therefore, further nutritional evaluation research on the cottonseeds of TM-1 is needed. There have been many studies and evaluations of primary metabolites in cottonseed, and this study focuses on the evaluation of secondary metabolites proanthocyanidins. Notably, catechin and epicatechin were both detected, catechin significantly accumulated in TM-1, while epicatechin accumulated in Hai7124. Proanthocyanidins are a class of polyphenols produced by the flavonoid pathway, that have important biological functions. Numerous *in vivo* and *in vitro* experimental studies have demonstrated that proanthocyanidins are beneficial to human and animal health (Dixon et al. 2005; Gonzalo-Diago et al. 2013). Proanthocyanidins are compounds formed by the polymerization of flavan-3-ol structural unit. And common flavan-3-ol units include catechin and epicatechin. Therefore, we find that the main monomeric forms in TM-1 and Hai7124 are different, which will also lead to the differences in oligomers formed by varying amounts of catechin or epicatechin.

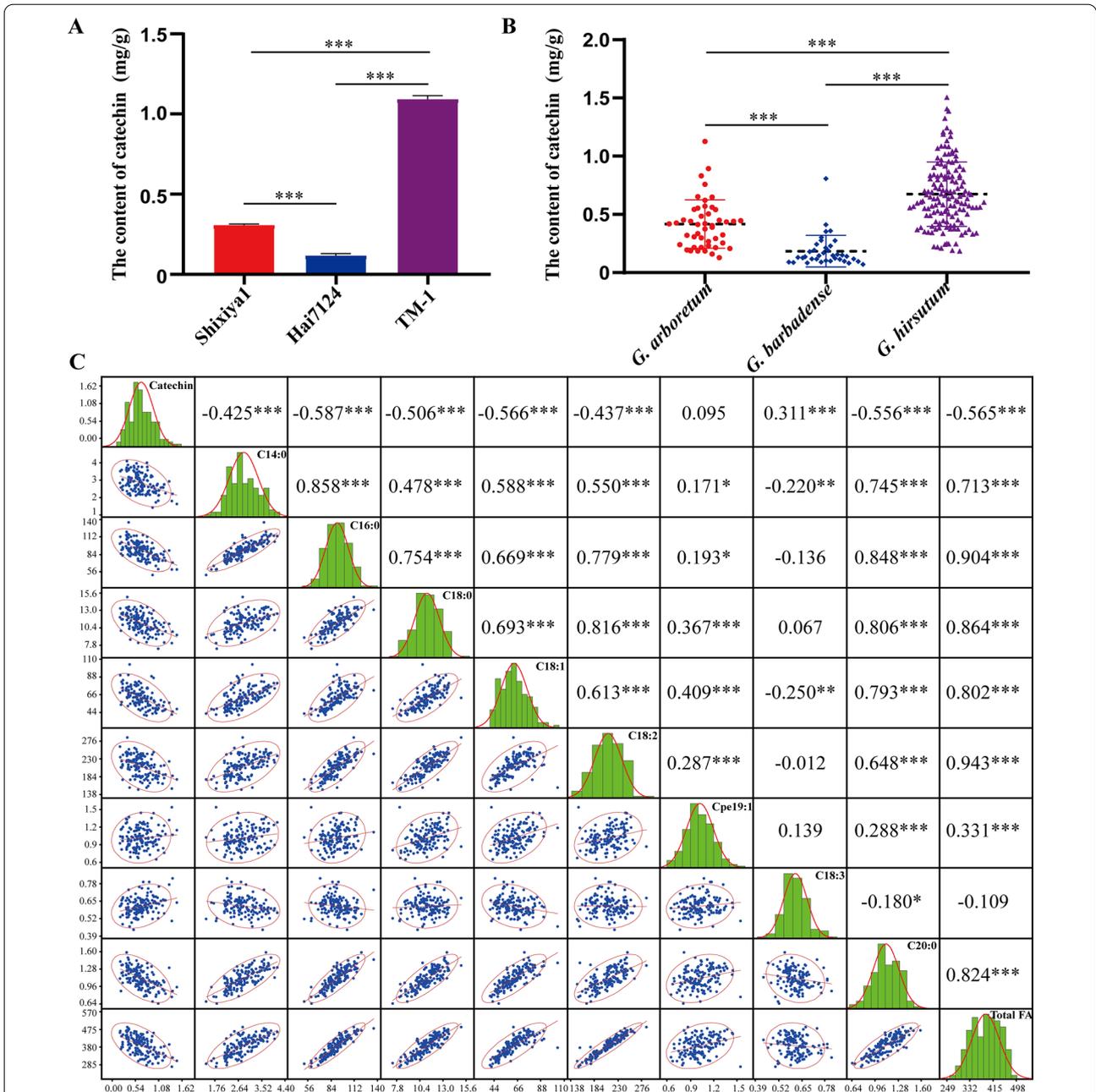
#### Correlation analysis of catechin and oil-related in core collections of *Gossypium*

To verify the reliability of the data obtained, we conducted confirmatory studies with different test methods

and an increased sample size. A classic chromatographic target HPLC method was used to evaluate the content of catechin, and the results indicated that the catechin content in TM-1 was significantly higher than that in other two cultivars, which is consistent with the above analysis (Fig. 2A). However, interestingly, we found that Shixiya1 had slightly higher values than Hai7124, which is different from the GC-TOF/MS results. Similar results have also been verified in core collections of *Gossypium* (47 accessions of *G. arboreum*, 37 accessions of *G. barbadense*, and 144 accessions of *G. hirsutum*) (Fig. 2B) (Du et al. 2018; He et al. 2021). Based on a comparative analysis of the detection methods, we thus believed that the acid used in the pretreatment method for HPLC could effectively release bound catechins into free catechins, which could be the potential reason behind this phenomenon. In Arabidopsis, the regulatory network of proanthocyanidins (PAs) in the seed coat is mainly controlled by the *TT2* gene. Previous studies have shown that *TT2* can suppress fatty acid (FA) biosynthesis, and that the *tt2* mutation results in significantly elevated seed oil content and an alteration in FA composition (Chen et al. 2012; Wang et al. 2014). Therefore, the correlation analysis was conducted between catechin and oil-related traits to evaluate their relationship in cottonseed. As indicated in Fig. 2C, catechin content has a negative association with myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), arachidic acid (C20:0) and total fatty acid content in the core collections of *G. hirsutum* (Pearson,  $P < 0.001$ ). In addition, the results also showed that there was no correlation between the content of catechin and cyclopropene fatty acid (C19:1), an anti-nutritional factor in cottonseed that is used for feed (Pearson,  $r = 0.095$ ,  $P > 0.05$ ). Taken together, this work advances our understanding of PAs functions in seed FA accumulation, and provides technical support for the development of special cotton varieties rich in catechins or high-quality fatty acids in the future. Lastly, further research should be undertaken to investigate the other nutrient components for the comprehensive evaluation of cottonseed.

#### Conclusions

In this work, the differential metabolites of three representative cultivars were identified and quantified using untargeted GC-TOF/MS to compare different cotton species. Catechin, as a bioactive component, was chosen to further verify the test results by classical HPLC analysis. The results revealed the reliability of the GC-TOF/MS analysis, as well as that catechin content has a negative association with that of myristic acid, palmitic acid, stearic acid, oleic acid, linoleic acid, arachidic



**Fig. 2** Assessment of differential catechins content and correlation analysis in core collections of *Gossypium*. **A** The contents of catechin in Shixiya1, Hai7124 and TM-1 measured by HPLC. **B** Distribution of the catechin content in three cotton species (*G. arboreum*, *G. barbadense*, and *G. hirsutum*) **C** Frequency distribution of phenotypic variation of catechin and oil-related traits, and correlation analysis in core collections of *G. hirsutum*. Myristic acid (C14:0); Palmitic acid (C16:0); Stearic acid (C18:0); Oleic acid (C18:1); Linoleic acid (C18:2); Cyclopropane fatty acid (C19:1); Linolenic acid (C18:3); Arachidic acid (C20:0) and Total fatty acids (Total FAs). Columns show the mean  $\pm$  SD with three replicates. \*, \*\*, \*\*\* represent  $P < 0.05$ ,  $P < 0.01$ , and  $P < 0.001$ , respectively

acid, and total fatty acids. The findings of the present study provide a theoretical basis for the rational utilization of cottonseeds in the future.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-022-00122-x>.

**Additional file 1: Fig. S1.** The typical total ion chromatogram (TIC) from all cottonseed samples by mass spectrometry detection.

**Additional file 2: Table S1.** The total peaks were extracted from cottonseeds by using GC-TOF/MS analysis.

**Additional file 3: Fig. S2.** The classification of all identified metabolites.

**Additional file 4: Table S2.** The information of 263 identified metabolites.

**Additional file 5: Fig. S3.** The PCA analysis of the metabolites based on GC-TOF/MS in Shixiya1, Hai7124 and TM-1. The X axis and Y axis represent the first principal component (PC1) and the second principal component (PC2), respectively.

**Additional file 6: Fig. S4.** Volcano plot of differential metabolites among three cultivars. **A** Shixiya1 vs. Hai7124. **B** Shixiya1 vs. TM-1. **C** Hai7124 vs. TM-1. Significantly up- and down-regulated metabolites are indicated in red and blue.

**Additional file 7: Table S3.** The differential metabolites between Shixiya1 and Hai7124 obtained from OPLS-DA. Up-regulated and down-regulated metabolites are shown in red and blue, respectively.

**Additional file 8: Table S4.** The differential metabolites between Shixiya1 and TM-1 obtained from OPLS-DA. Up-regulated and down-regulated metabolites are shown in red and blue, respectively.

**Additional file 9: Table S5.** The differential metabolites between Hai7124 and TM-1 obtained from OPLS-DA. Up-regulated and down-regulated metabolites are shown in red and blue, respectively.

**Additional file 10: Table S6.** The information of differential metabolites in the three cotton varieties obtained from one-way ANOVA.

## Acknowledgements

The authors thank Prof. DU Xiongming from Institute of Cotton Research, CAAS for supplying the cotton germplasm resources.

## Author contributions

Ma L and Tian XQ designed the experiments. Tian XQ, Chen YB and Xu SJ carried out the experiments with the help of Dong RD, and Wang YW. Fang D collected the samples and pretreatment; Ma L managed the main data analysis and wrote the manuscript. Ma L and Peng J conceived the project. All authors have read and approved the final manuscript.

## Funding

This work was supported by China Agriculture Research System (CARS-15-27) and Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Henan 450001, Zhengzhou, China. <sup>2</sup>State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China. <sup>3</sup>Laboratory of Quality and Safety Risk Assessment for Cotton Products, Ministry of Agriculture and Rural Affairs, Institute of Cotton Research of CAAS, Anyang, Henan 455000, China.

Received: 26 January 2022 Accepted: 8 April 2022

Published online: 21 May 2022

## References

- Baud S. Seeds as oil factories. *Plant Reprod.* 2018;31:213–35. <https://doi.org/10.1007/s00497-018-0325-6>.
- Beale DJ, Pinu FR, Kouremenos KA, et al. Review of recent developments in GC-MS approaches to metabolomics-based research. *Metabolomics.* 2018;14:152. <https://doi.org/10.1007/s11306-018-1449-2>.
- Chen MX, Wang Z, Zhu YN, et al. The effect of TRANSPARENT TESTA2 on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in Arabidopsis. *Plant Physiol.* 2012;160:1023–36. <https://doi.org/10.1104/pp.112.202945>.
- Choudhury FK, Pandey P, Meitei R, et al. GC-MS/MS profiling of plant metabolites. *Methods Mol Biol.* 2022;2396:101–15. [https://doi.org/10.1007/978-1-0716-1822-6\\_9](https://doi.org/10.1007/978-1-0716-1822-6_9).
- Dalgliesh CE, Horning E, Horning MG, et al. A gas-liquid-chromatographic procedure for separating a wide range of metabolites occurring in urine or tissue extracts. *Biochem J.* 1966;101:792–810. <https://doi.org/10.1042/bj1010792>.
- Deng H, Chen S, Zhou Z, et al. Transcriptome analysis reveals the effect of short-term sunlight on aroma metabolism in postharvest leaves of oolong tea (*Camellia sinensis*). *Food Res Int.* 2020;137: 109347. <https://doi.org/10.1016/j.foodres.2020.109347>.
- Dixon RA, Xie DY, Sharma SB. Proanthocyanidins—a final frontier in flavonoid research? *New Phytol.* 2005;165:9–28. <https://doi.org/10.1111/j.1469-8137.2004.01217.x>.
- Dowd MK, Boykin DL, Meredith WR, et al. Fatty acid profiles of cottonseed genotypes from the national cotton variety trials. *J Cotton Sci.* 2010;14:64–73.
- Du XM, Huang G, He SP, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet.* 2018;50:796–802. <https://doi.org/10.1038/s41588-018-0116-x>.
- Egbuta MA, McIntosh S, Waters DLE, et al. Biological importance of cotton by-products relative to chemical constituents of the cotton plant. *Molecules.* 2017;22:93. <https://doi.org/10.3390/molecules22010093>.
- Gonzalo-Diago A, Dizy M, Fernandez-Zurbano P. Taste and mouthfeel properties of red wines proanthocyanidins and their relation to the chemical composition. *J Agric Food Chem.* 2013;61:8861–70. <https://doi.org/10.1021/jf401041q>.
- He SP, Sun GF, Geng XL, et al. The genomic basis of geographic differentiation and fiber improvement in cultivated cotton. *Nat Genet.* 2021;53:916–24. <https://doi.org/10.1038/s41588-021-00844-9>.
- Jacobs DM, van den Berg MA, Hall RD. Towards superior plant-based foods using metabolomics. *Curr Opin Biotechnol.* 2021;70:23–8. <https://doi.org/10.1016/j.copbio.2020.08.010>.
- Kind T, Wohlgemuth G, Lee D, et al. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem.* 2009;81:10038–48. <https://doi.org/10.1021/ac9019522>.
- Mansoor S, Paterson AH. Genomes for jeans: cotton genomics for engineering superior fiber. *Trends Biotechnol.* 2012;30:521–7. <https://doi.org/10.1016/j.tibtech.2012.06.003>.
- Patel MK, Pandey S, Kumar M, et al. Plants metabolome study: emerging tools and techniques. *Plants.* 2021;10:2409. <https://doi.org/10.3390/plants10112409>.
- Peng YR, Meng QL, Zhou J, et al. Nanoemulsion delivery system of tea polyphenols enhanced the bioavailability of catechins in rats. *Food Chem.* 2018;242:527–32. <https://doi.org/10.1016/j.foodchem.2017.09.094>.

- Razzaq A, Sadia B, Raza A, et al. Metabolomics: a way forward for crop improvement. *Metabolites*. 2019;9:303. <https://doi.org/10.3390/metabo9120303>.
- Rochfort S. Metabolomics reviewed: a new "Omics" platform technology for systems biology and implications for natural products research. *J Nat Prod*. 2005;68:1813–20. <https://doi.org/10.1021/np050255w>.
- Saito K. Plant metabolomics: a basis for plant functional genomics and biotechnology. *New Biotechnol*. 2009;25:317–8. <https://doi.org/10.1016/j.nbt.2009.06.869>.
- Saito K, Matsuda F. Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol*. 2010;61:463–89. <https://doi.org/10.1146/annurev-arplant-043008.092035>.
- Sharma V, Gupta P, Kagolla P, et al. Metabolomics intervention towards better understanding of plant traits. *Cells*. 2021;10:346. <https://doi.org/10.3390/cells10020346>.
- Wang Z, Chen MX, Chen TL, et al. *TRANSPARENT TESTA2* regulates embryonic fatty acid biosynthesis by targeting *FUSCA3* during the early developmental stage of Arabidopsis seeds. *Plant J*. 2014;77:757–69. <https://doi.org/10.1111/tpj.12426>.
- Wang SC, Alseekh S, Fernie AR, et al. The structure and function of major plant metabolite modifications. *Mol Plant*. 2019;12:899–919. <https://doi.org/10.1016/j.molp.2019.06.001>.
- Xu SJ, Fang D, Tian XQ, et al. Subcritical water extraction of bioactive compounds from waste cotton (*Gossypium hirsutum* L.) flowers. *Ind Crop Prod*. 2021;164:113369. <https://doi.org/10.1016/j.indcrop.2021.113369>.
- Yuan XF, Ma L, Wen BT, et al. Enhancing anaerobic digestion of cotton stalk by pretreatment with a microbial consortium (MC1). *Bioresour Technol*. 2016;207:293–301. <https://doi.org/10.1016/j.biortech.2016.02.037>.
- Zhang TZ, Hu Y, Jiang WK, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015;33:531–52. <https://doi.org/10.1038/nbt.3207>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

