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

Lei Ma (malei@caas.cn)
Chinese Academy of Agricultural Sciences Cotton Research Institute
Chen Yabing
Zhengzhou University
Xu Shuangjiao
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
Dong Ruidan
Zhengzhou University
Wang Yiwen
Zhengzhou University
Fang Dan
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
Peng Jun
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute
XinQuan Tian
CAAS CRI: Chinese Academy of Agricultural Sciences Cotton Research Institute

Research

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

Posted Date: February 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1297603/v1

Abstract

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

Results: In this study, the metabolomes of the three cotton cultivars of different species were compared using an untargeted GC-TOF/MS analysis. A total of 263 metabolites were identified from 705 peaks and their levels were compared across the cultivars. PCA and OPLS-DA of these metabolites could clearly distinguish. There were significant differences in the content of amino acids, carbohydrates, organic acids flavonoids, and lipids in *G. hirsutum* TM-1 compared with *G. arboretum* Shixiya1 and *G. barbadense* Hai7124. It is noteworthy that the bioactive nutrient compound catechin obtained from the differential metabolites is significantly accumulated in TM-1. Furthermore, a comprehensive analysis using catechin and oil-related traits was conducted in core collections of *Gossypium*. And the results revealed the feasibility 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, linolenic acid, arachidic acid and total fatty acids.

Conclusion: These fndings suggest that catechin content has a negative association with most fatty acids in cottonseed, which may provides a new solution into the underlying plant biochemistry of nutrient variation in cottonseed and pave the way to exploit the value of cotton by-products.

Introduction

Cotton is grown in over 75 countries mainly for its fibre 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; Xu et al. 2021; Yuan et al. 2016). These are commonly referred to cotton by-products, for example, 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 kind of non-fresh agricultural product, whose quality characteristics is easy to be ignored in the process of planting and primary processing. And the latest cutting-edge technology is rarely to be applied to cotton byproduct, resulting in its low utilization level.

Plants could produce a variety of metabolites with structurally diverse, which play essential roles 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, medicine and so on (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 crop. 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 was reported for separating metabolites occuring in urine or tissue extracts (Dalgliesh et al. 1966). Besides, liquid chromatography/mass spectrometry (LC/MS) and nuclear magnetic resonance (NMR) have also developed into main analytical techniques for metabolomics in recent years (Rochfort 2005). Due to high resolution, high sensitivity, large number of mass spectral library, good reproducibility, and relatively low cost, GC/MS is suitable for the analysis of volatile and semi-volatile metabolites with relative molecular mass, low polarity, low boiling point or volatile compounds after derivatization. High resolution 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 spinnable fibers in the world, among which allotetraploid cultivars *G. hirsutum* (upland cotton) accounts for about 90% of global cotton production, with the remaining production made up by, *G. barbadense* (sea island cotton) (Mansoor and Paterson 2012; Zhang et al. 2015). The diploids cultivars of *G. arboreum*, and *G. herbaceum* have no agricultural output today. At present, there are few literature reports on the use of metabolomics techniques to evaluate nutrients, bioactivity and health benefits for cottonseed, and there are fewer comparative studies among cotton species. In this study, to understand better the metabolite variations for nutrient among cotton species, untargeted GC-TOF/MS analysis was performed to identify and quantify metabolites including fatty acids, polyphenols, carbohydrates, organic acids, amino acids and so on in three representative cultivars. Then classical chromatographic analysis was used to further verify the test results. Lastly, correlation analysis of micronutrient catechin with oil-related traits was conducted for exploring the underlying relationship between nutritional quality in core collections of *Gossypium*. This study provides a reference for exploiting the value of cotton by-products 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 arboretum* (*G. arboretum*), *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. The cotton was planted in Chinese Academy of Agricultural Sciences, Anyang, China, and were cultivated under the 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℃ until GC-TOF/MS analysis.

To illustrate the content distribution of catechin in different species, 47 accessions of *G. arboretum*, 37 accessions of *G. barbadense*, 144 accessions of *G. hirsutum* were chose from core collections of

Gossypium used for measure catechin content by HPLC analysis respectively. 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 μL adonitol (0.5 mg/mL stock in dH_2O) as internal standard. The resulting mixture was ultrasound treated for 5min, and then centrifuged (12000 rpm) at 4°C for 15 min. Transferred the supernatant (0.4 mL) to a new 2 mL GC/MS glass bottle and dried it completely in a vacuum concentrator. Then the extracts were oximated using 80 μL methoxyamine hydrochloride (20 mg/mL in pyridine) at 80℃ for 25 min. Subsequently, the samples were added 100 μL BSTFA regent (1% TMCS, v/v) and incubated at 70℃ for 1.5 h. The GC-TOF/MS was performed as described by Deng et al (Deng et al, 2020)

Data preprocessing and annotation

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

Catechin extraction and analysis by High Performance Liquid Chromatography (HPLC)

Briefly, about 100 mg cottonseed powder was extracted using 3 mL of 80% acidified methanol (1% hydrochloric acid). After 30 min sonication, the homogenate was incubated for 12 h at 25℃. Then samples were centrifuged at 10000 rpm 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).

Fatty acids extracted and analyzed by Gas chromatography (GC)

Extraction of fatty acids from mature cottonseeds was performed as described previously [16]. Briefly, about 50 mg of cottonseed powder was added 1 mL of 0.3 M potassium hydroxide methanol solution, 1 mL of n-hexane containing 500 μg/mL C11:0 as an internal standard. After shaking for 30 seconds and holding at 25℃ for 1.5 h, the homogenate was added 1.5 mL of 0.9% (w/v) sodium chloride solution, followed by centrifugation (10000 rpm) at 25℃ for 5 min. The supernatant was separated by gas chromatography (Agilent7890-FID) according to previously described procedures (Dowd et al. 2010).

Multivariate analysis and statistics

To better visualization and subsequent analysis, the data of GC-TOF/MS was used to perform PCA and OPLS-DA analysis by the SIMCA software (V14.1, MKS Data Analytics Solutions, Umea, Sweden). The *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 acids 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

The metabolomics analysis was employed to elucidate the potential chemical basis of different species of *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 was shown in 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 peaks of TIC between samples. A total of 705 peaks were extracted and listed in Table S1. Based on the local metabolite database, 263 metabolites were identified and classified into 15 different categories (Fig. S2, Table S2), including 53 organic acids, 44 carbohydrates, 43 amino acid derivatives, 34 others metabolites, 25 alcohols and polyols, 14 lipids, 10 benzene and substituted derivatives, 9 nucleotide and its derivates, 7 amines, 6 phenylpropanoids, 6 flavonoids, 3 sphingolipids, 2 quinate and its 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, which converts a large number of variables into principal components (PC) that still contains most of the information in the large set. PCA is an unsupervised mode, showing the distribution of origin data and used to evaluate the difference for inter-group comparison. As illustrated in Fig. S3, the PCA score scatter plot indicated significant differences in 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 the correlation of samples. It is apparent from Fig. 1A that, the OPLS-DA could be better to separate the difference in metabolic phenotypes between three varieties of *Gossypium* genus. Furthermore, 7-fold cross validation and permutation test 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 the risk of over-fitting was low. These results suggested that there were significant differences in the metabolites of each cultivar. Subsequently, the variable importance (VIP) value in the projection obtained by the OPLS-DA analysis was greater than 1, and the *P* value of the *Student's t-test* analysis was less than 0.05, which were used

for the screening of differential metabolites. The volcano plot was employed for visualize the differential metabolites (Fig. S4). The results showed that there were 64 up-regulated (the red scatter points) and 55 down-regulated (the blue scatter points) metabolites in the Shixiya1 compared with their counterparts in the Hai7124 (Fig. S4A, Table S3); 131 differentially abundant metabolites (29 up-regulated and 102 down-regulated) were found in the Shixiya1 compared with their counterparts in the TM-1 (Fig. S4B, Table S4); Moreover, comparing with TM-1, 26 up-regulated and 102 down-regulated metabolites were obtained in the Hai7124 (Fig. S4C, Table S5); The classification of metabolites shows that the content of some carbohydrates, amino acids, lipids, organic acids and flavonoids in TM-1 were significantly higher than that in Hai7124 or Shixiya1. These results indicated that the seeds of upland cotton TM-1 may contain more abundant metabolites, which is beneficial to the further processing of cottonseeds. One-way ANOVA was also conducted for 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 inter-group variation (Fig. 1B). The relative contents of metabolite represented by color segments at the corresponding locations were showed in Table S6.

As *G. hirsutum* is the most widely planted cotton cultivar, accounting for absolute majority of cotton fiber yield and cottonseed by-products in the world. It is interest for us to carry out the further nutritional evaluation research for TM-1. Notably, through visual analysis of different metabolites, we found that catechin and linolenic acid are significantly different in the three cotton varieties. Moreover, catechin, as a bioactive component of proanthocyanidin, was found to accumulated significantly higher in *G. hirsutum* TM-1 than in the other two cultivars. Therefore, since this metabolite have potential for the utilization of cotton by-product, we chose it to conduct evaluation research in the core collections of *Gossypium*.

Correlation analysis of catechins and fatty acids traits in core collections of Gossypium

A classic chromatographic target HPLC method was used to evaluate the content of catechin, and the results indicated that the catechin in TM-1 was significantly higher than that in other two cultivars, consistent with the above analysis (Fig. 2A). But interestingly, we found that Shixiya1 is slightly higher than Hai7124, which is different from GC-TOF/MS. Similar results have also been verified in core collections of *Gossypium* (47 accessions of *G. arboretum*, 37 accessions of *G. barbadense*, 144 accessions of *G. hirsutum*) (Fig. 2B) (Du et al. 2018; He et al. 2021). Through a comparative analysis of the detection methods, we thus believed that the acid used in in the pre-treatment method for HPLC could effectively release bound catechins into free catechins, which could be the potential reason behind this phenomenon. Next, to better evaluate the relationship between micronutrient catechin and the known major nutrients in cottonseed, the correlation analysis was conducted between catechin and oil-related traits. 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), linolenic acid (C18:3), arachidic acid (C20:0) and total fatty acids in the core collections of *G. hirsutum* (*P*-value < 0.001). In

addition, the results also showed that there was no correlation between catechins and cyclopropene fatty acid (C19:1), an anti-nutritional factor in cottonseed for feed (r=0.095, *P*-value > 0.05). In short, the above research provides technical support for development of special cotton varieties rich in catechins and high-quality fatty acids in future. Further research should be undertaken to investigate the other nutrient components for 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 chromatographic analysis HPLC. The results revealed the feasibility 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, linolenic acid, arachidic acid and total fatty acids. The findings of the present study provide a theoretical basis for the rational utilization of cottonseeds in the future.

Declarations

Acknowledgments

The author thanks Prof. Xiongming Du for supplying the cotton germplasm resources.

Authors' contributions

Tian Xinquan and Ma Lei designed the experiments. Tian Xinquan, Chen Yabing and Xu Shuangjiao carried out the experiments with the help of Dong Ruidan, and Wang Yiwen. Fang Dan collected the samples and pretreatment; Ma Lei managed the main data analysis and wrote the manuscript. Ma Lei and Peng Jun 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.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no confict of interest

Author details

[1] Research base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou, Henan 450001, China. [2] State Key Laboratory of Cotton Biology, Institute of Cotton Research of CAAS, Anyang 455000, China. [3]Laboratory of Quality & Safety Risk Assessment for Cotton Products, MARA, Institute of Cotton Research of CAAS, Anyang 455000, China.

References

1. Dalgliesh CE, Horning E, Horning MG, et al. A gas-liquid-chromatographic procedure for separating a wide range of metabolites occuring in urine or tissue extracts. Biochem J. 1966;101:792–810. https://doi.org/10.1042/bj1010792.
2. 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. doi. 10.1016/j.foodres.2020.109347.
3. 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.
4. 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.
5. 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.
6. 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.
7. Jacobs DM, van den Berg MA, Hall RD. Towards superior plant-based foods using metabolomics. Curr Opin Biotech. 2021;70:23–8. https://doi.org/10.1016/j.copbio.2020.08.010.
8. 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.
9. 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.
10. 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.

11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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: https://doi.org/10.1016/j.indcrop.2021.113369.
17. 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.
18. 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;33:531–252. https://doi.org/10.1038/nbt.3207.

Figures

Figure 1

The 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 of three cotton cultivars used to draw the heatmap of hierarchical clustering analysis (HCA).

Figure 2

Assessment of differential catechins content and correlation analysis in in core collections of *Gossypium*. **A** The contents of catechin in Shixiya1, Hai7124 and TM-1 measured by HPLC. **B** Distribution of catechin content in three cotton species (*G. arboreum*, *G. barbadense*, and *G. hirsutum*) **C** Correlation analysis of catechin content and fatty acids content in the seeds of *G. hirsutum*. Columns present mean ± SD with three replicates. *, **and *** represent $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tif
- FigureS2.tif
- FigureS3.tif
- FigureS4.tif
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx