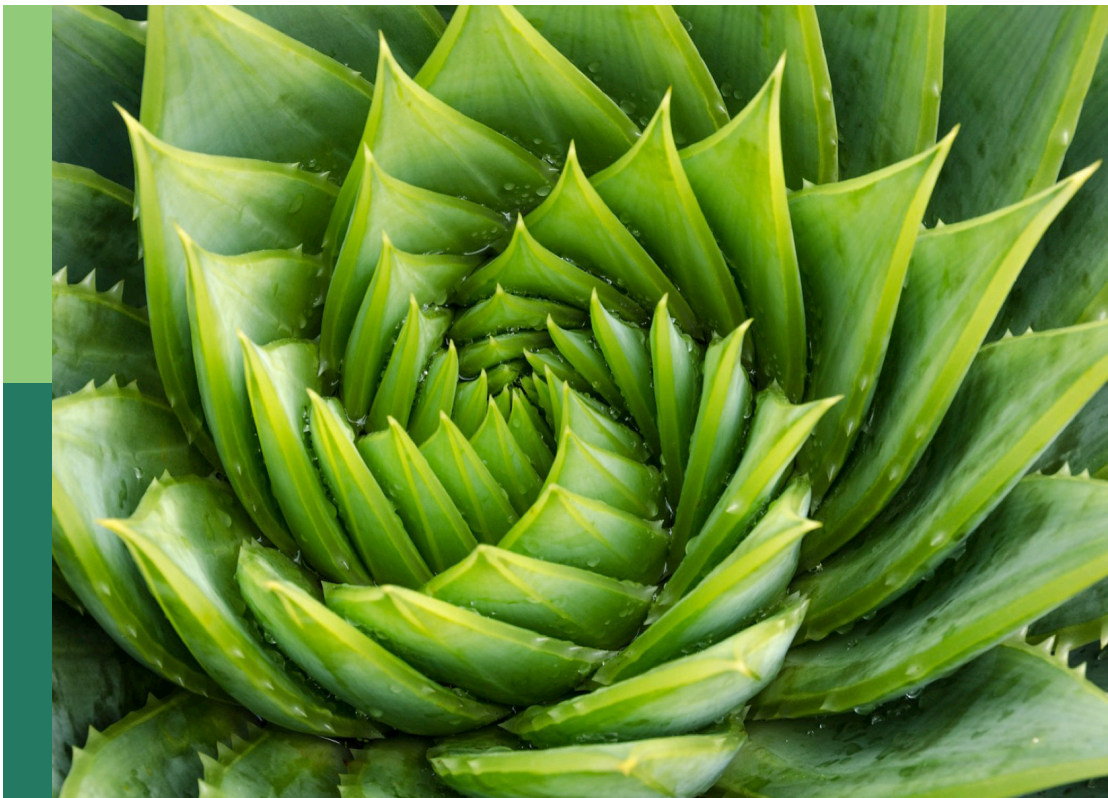
frontiers | Research Topics

# Metabolomics in crop research – current and emerging methodologies, volume II

**Edited by**
Marta Sousa Silva, Ute Roessner and Carlos Cordeiro

**Published in**
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3646-9
DOI 10.3389/978-2-8325-3646-9

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Metabolomics in crop research – current and emerging methodologies, volume II

**Topic editors**

Marta Sousa Silva — University of Lisbon, Portugal

Ute Roessner — The University of Melbourne, Australia

Carlos Cordeiro — University of Lisbon, Portugal

**Citation**

Silva, M. S., Roessner, U., Cordeiro, C., eds. (2023). *Metabolomics in crop research – current and emerging methodologies, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3646-9

# Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 22 September 2023
DOI 10.3389/fpls.2023.1292878



OPEN ACCESS

EDITED AND REVIEWED BY
Romina Beleggia,
Research Center for Cereal and Industrial Crops, Italy

*CORRESPONDENCE
Marta Sousa Silva
mfsilva@fc.ul.pt

RECEIVED 12 September 2023
ACCEPTED 18 September 2023
PUBLISHED 22 September 2023

CITATION
Sousa Silva M, Roessner U and Cordeiro C (2023) Editorial: Metabolomics in crop research – current and emerging methodologies, volume II.
*Front. Plant Sci.* 14:1292878.
doi: 10.3389/fpls.2023.1292878



# Editorial: Metabolomics in crop research – current and emerging methodologies, volume II

Marta Sousa Silva[1,2]*, Ute Roessner[3] and Carlos Cordeiro[1,2]

[1]Laboratório de FTICR e Espectrometria de Massa Estrutural, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal, [2]Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal, [3]School of BioSciences, The University of Melbourne, Melbourne, VIC, Australia

KEYWORDS

**metabolomics, agriculture, biomarkers, volatile organic compounds, pathogen resistance, species discrimination, new products discovery**

Editorial on the Research Topic
Metabolomics in crop research – current and emerging methodologies, volume II

Metabolomics' potential in plant and crop research is immense, thus giving an important contribution to improve crop yield and quality, to understand plant development and response to the environment, to pathogens, to assess food safety and maximize agriculture profits. This is the second edition of the Research Topic "*Metabolomics in Crop Research - Current and Emerging Methodologies*". It contains 10 articles from 81 authors that highlight the current methodologies and applications of metabolomics for the discrimination of crop species and cultivars, in the analysis of a crop's response to biotic and abiotic stresses, in the characterization of important compounds as well as the discovery of new compounds and metabolic pathways.

Crop domestication results from the selection and management of plant phenotypes over time, to achieve the desired characteristics. The domestication of *Capsicum annuum* (chili pepper), which started 6,000 years ago, caused considerable changes in the fruit and consequent alterations in metabolism. Cervantes-Hernández et al. analyzed the metabolic profiles during fruit development of wild and domesticated accessions of *C. annuum* L. Both groups were clearly separated based on their metabolic profiles and significant differences were observed in the fruit development between the wild and domesticated plants. This work highlights the effect of the domestication process in plant metabolism, contributing in the future to the development of plant breeding strategies to restore metabolic diversity. *Vitis vinifera* (grapevine) is another successful case of plant domestication, with the development of nearly 6,000 different varieties, grouped according to their geographic-genetic origin (GEN-GEO groups). Šikuten et al. were able to discriminate grapevine varieties belonging to different GEN-GEO groups based on their grapes' volatile metabolic profiles. This analysis allowed not only genetic discrimination, but also the identification of the country of origin, thus contributing to the promotion of grape quality and value linked to geographical origin. One of the main problems in grapevine cultivation is plant susceptibility to several diseases, particularly downy mildew.

In the work of Ciubotaru et al., the metabolic responses to the inoculation with *Plasmopara viticola* (downy mildew infectious agent) in different resistant genotypes were analyzed, following a post-infection time-course analysis. Several potential resistance biomarkers were identified, thus contributing to understanding grapevine resistance mechanisms and assisting future breeding programs for improving its resilience to infection.

The use of plants in traditional Chinese medicine is a millenary practice that is becoming widespread. Species discrimination is crucial since they can have different pharmacological effects. This is the case of *Polygonatum*, whose rhizomes are widely used in traditional Chinese medicine. Three *Polygonatum* species are included in the *Chinese Pharmacopoeia* (2020 Ed.), each one with different medicinal effects. Since their discrimination is often based on physiology and morphology, Wang et al. analysed the metabolic profile of these *Polygonatum* species and identified the discriminant compounds between them, thus contributing for a more accurate species identification and paving the way for a better understanding of its mechanism of action.

*Oryza sativa* (rice) is one of the most important food crops in the world. Hence, it is important to provide the best agricultural conditions for its growth. The use of volatile organic compounds produced by specific bacteria is a promising strategy to increase productivity in a more sustainable way. Almeida et al. analysed the metabolic composition and effects of bacterial volatilome on the growth and metabolism of rice and identified two promising bacterial isolates that improve rice growth, contributing to a more sustainable agriculture. Given the importance of rice, experiments have been carried out to understand the effects of deep space flight on rice seed transport and cultivation. Having in mind that spaceflight stress affected rice growth, leading to significant changes in metabolic pathways, Zeng et al. analysed the metabolome and proteome of F2 generation plants grown from seeds exposed to this stress. Rice F2 generation plants preserved the memory of spaceflight stress, thus bringing a new perspective for these studies in other crops and its potential for space flight and human colonization beyond Earth. A popular substitute to rice is quinoa (*Chenopodium quinoa* Willd.) given its high nutritional value. Although still considered a minor crop, the Food and Agriculture Organization (FAO) gave it the statute of a promising crop for the human diet that can contribute to food security in the 21$^{st}$ century. In the defense response against microbes, herbivores, and insects, quinoa plants produce the triterpenoid saponin. Zhao et al. studied this compound's biosynthesis pathway in quinoa using a combined transcriptomics-metabolomics strategy. These discoveries contributed to a thorough characterization of the metabolic mechanism of triterpenoid saponin in quinoa. Besides biotic stresses, quinoa is affected by drought stress. Huan et al. studied the drought tolerance response mechanisms in quinoa after drought and rewatering treatments, using a transcriptomics and targeted metabolomics approach. This work will impact future breeding programs of new drought-tolerance quinoa varieties.

Leaves from coffee (*Coffea* genus) plants, and others from the same family (*Rubiaceae*), produce highly important bioactive polyphenols, with applications in food and pharmaceutical industries. Leaves are a disregarded by-product of the coffee industry, which can be used for the extraction of these relevant compounds. Castro-Moretti et al. developed a single-step extraction methodology for the recovery of a vast diversity of phytochemicals from coffee and other *Rubiaceae* leaves and a method for quantification and detection of new polyphenols. This research will impact future breeding programs, boost phytochemical discovery and contribute to the circular economy, reducing waste and increasing value. The discovery of new compounds from the hemp seed (*Cannabis sativa* L.) is also quite relevant, given its applications in the pharmaceutical, nutraceutical and food industries. Padilla-González et al. analyzed the metabolic profile of several hemp seed accessions to identify new metabolites and associate specific biologically important molecules to particular plant accessions. Once again, these results can be used in future optimization of breeding programs for the development of new and improved varieties.

Metabolomics application to plant and crop research is constantly progressing. The resulting information can be used in the development of new improved crops, able to respond to climate change, environmental degradation, pathogen attacks, and loss of biodiversity. Metabolomics must be part of the agricultural research and innovation strategy towards a sustainable production. This is the Volume II of the Research Topic "*Metabolomics in Crop Research - Current and Emerging Methodologies*", and we are looking forward to future contributions in the next edition.

## Author contributions

MS: Conceptualization, Writing – original draft, Writing – review & editing. UR: Writing – review & editing. CC: Writing – original draft, Writing – review & editing.

## Funding

The authors acknowledge the support from the Research Infrastructure project European Network of Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry Centers (EU FT-ICR-MS) funded by the European Union's Horizon 2020 research and innovation programme, Grant Agreement No 731077, and also from the research center BioISI (ref. UIDB/04046/2020 and UIDP/04046/2020) and the Portuguese Mass Spectrometry Network (RNEM, (LISBOA-01-0145-FEDER-022125), both funded by the Fundação para a Ciência e a Tecnologia, Portugal.

## Acknowledgments

The Editors of this topic wish to thank all the authors and reviewers, and support staff who contributed to the success of the second volume of this Frontiers Research topic. We also acknowledge the Frontiers Editorial Office, the Chief Editors and the Frontiers Plant Science Production Office for their technical support.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

frontiers in Plant Science

ORIGINAL RESEARCH
published: 01 July 2021
doi: 10.3389/fpls.2021.693887

# Mono-Locus and Pyramided Resistant Grapevine Cultivars Reveal Early Putative Biomarkers Upon Artificial Inoculation With *Plasmopara viticola*

*Ramona Mihaela Ciubotaru[1,2], Pietro Franceschi[3], Luca Zulini[4], Marco Stefanini[4], Domen Škrab[1,2], Marcia Denise Rossarolla[5], Peter Robatscher[6], Michael Oberhuber[6], Urska Vrhovsek[2*†] and Giulia Chitarrini[2,6*†]*

[1] Department of Agri-Food, Environmental and Animal Sciences, University of Udine, Udine, Italy, [2] Food Quality and Nutrition Department, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy, [3] Unit of Computational Biology, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy, [4] Genomics and Biology of Fruit Crops Department, Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy, [5] Plant Science Department, Federal University of Santa Catarina, Florianópolis, Brazil, [6] Laimburg Research Centre, Auer, Italy

One of the most economically important grapevine diseases is Downy mildew (DM) caused by the oomycete *Plasmopara viticola*. A strategy to reduce the use of fungicides to compensate for the high susceptibility of *V. vinifera* is the selection of grapevine varieties showing pathogen-specific resistance. We applied a metabolomics approach to evaluate the metabolic modulation in mono-locus resistant genotypes carrying one locus associated with *P. viticola* resistance (*Rpv*) (BC4- *Rpv1*, Bianca- *Rpv3-1*, F12P160- *Rpv12*, Solaris- *Rpv10*), as well as in pyramided resistant genotypes carrying more than one *Rpv* (F12P60- *Rpv3-1; Rpv12* and F12P127- *Rpv3-1, Rpv3-3; Rpv10*) taking as a reference the susceptible genotype Pinot Noir. In order to understand if different sources of resistance are associated with different degrees of resistance and, implicitly, with different responses to the pathogen, we considered the most important classes of plant metabolite primary compounds, lipids, phenols and volatile organic compounds at 0, 12, 48, and 96 h post-artificial inoculation (hpi). We identified 264 modulated compounds; among these, 22 metabolites were found accumulated in significant quantities in the resistant cultivars compared to Pinot Noir. In mono-locus genotypes, the highest modulation of the metabolites was noticed at 48 and 96 hpi, except for Solaris, that showed a behavior similar to the pyramided genotypes in which the changes started to occur as early as 12 hpi. Bianca, Solaris and F12P60 showed the highest number of interesting compounds accumulated after the artificial infection and with a putative effect against the pathogen. In contrast, Pinot Noir showed a less effective defense response in containing DM growth.

**Keywords: downy mildew, metabolomics, mono-locus, pyramided, resistance**



***Edited by:***
*Ancheng Huang, Southern University of Science and Technology, China*

***Reviewed by:***
*Eva Maria Zyprian, Institut für Rebenzüchtung, Julius Kühn-Institut, Germany*
*Mariam Gaid, Independent Researcher, Braunschweig, Germany*
*Andreia Figueiredo, University of Lisbon, Portugal*

***Correspondence:**
*Urska Vrhovsek*
*urska.vrhovsek@fmach.it*
*Giulia Chitarrini*
*giulia.chitarrini@gmail.com;*
*giulia.chitarrini@fmach.it*

*†These authors share last authorship*

***Specialty section:***
*This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science*

***Received:*** *12 April 2021*
***Accepted:*** *04 June 2021*
***Published:*** *01 July 2021*

***Citation:***
*Ciubotaru RM, Franceschi P, Zulini L, Stefanini M, Škrab D, Rossarolla MD, Robatscher P, Oberhuber M, Vrhovsek U and Chitarrini G (2021) Mono-Locus and Pyramided Resistant Grapevine Cultivars Reveal Early Putative Biomarkers Upon Artificial Inoculation With Plasmopara viticola. Front. Plant Sci. 12:693887. doi: 10.3389/fpls.2021.693887*

## INTRODUCTION

Grapevine was among the first fruit species to be domesticated and today represents one of the most important crops in the world, with an essential role in the economy of many countries. Unfortunately, viticulture is threatened by numerous pathogens causing severe harvest losses. One of the most destructive diseases affecting grapevine is Downy mildew (DM), caused by the biotrophic pathogen *Plasmopara viticola*. DM affects the members of the family *Vitaceae* and in particular the cultivated species *Vitis vinifera* and it can attack all green parts of the vine (leaves, fruits, and shoots in particular) (Buonassisi et al., 2018; Vezzulli et al., 2018). DM infection leads to significant crop losses due to defoliation and to the production of low-quality, deformed or entirely damaged grapes (Yildirim et al., 2019; Nogueira Júnior et al., 2020). The most distinctive signs of infection are the sporangia formation apparent as whitish spots, commonly found on the abaxial surface of the first-formed leaves, which are accompanied by chlorotic spots (known also as oil spots) on the adaxial surface. The sporulation requires humidity > 93% and temperatures of 18–20°C and it can be observed on the abaxial side of the leaf and the surface of tendrils, inflorescence, and young berries (Buonassisi et al., 2018).

The application of large amounts of fungicides is the most diffused strategy to control DM, this practice, however, is not only expensive and in conflict with the requirements for sustainable and environment-friendly agriculture, but also promotes the emergence of fungicide-resistant strains (Buonassisi et al., 2018; Merdinoglu et al., 2018; Fröbel and Zyprian, 2019; Yildirim et al., 2019). A possible alternative to the use of fungicides is the valorization of the interspecific hybrids of *V. vinifera* with resistant genotypes from *Muscadinia*, several wild North American and Asian *Vitis* species which have been found resistant against *P. viticola* (Buonassisi et al., 2017; Merdinoglu et al., 2018; Vezzulli et al., 2018; Fröbel and Zyprian, 2019).

The resistance response to *P. viticola* is given by quantitative trait loci (QTLs) named *Rpv* (i.e., resistance to *P. viticola*). To date, 27 quantitative trait loci (QTL) have been identified in wild *Vitis* species and a descriptive list of them is available online (www.vivc.de/data on breeding and genetics/ Table of loci for Traits in Grapevine) (Bellin et al., 2009; Bove et al., 2019; Eisenmann et al., 2019; Vezzulli et al., 2019; Maul et al., 2020; Nogueira Júnior et al., 2020). The protection offered by these resistance genes (*R* genes) can be overcome by virulent strains of the pathogen, particularly in the genotypes carrying one *Rpv* gene (Peressotti et al., 2010; Merdinoglu et al., 2018; Fröbel et al., 2019). To avoid such resistance breakdowns a longer-lasting disease resistance is required. A possible strategy is pyramiding resistance, by accumulating several resistant genes in the same variety to create a durable disease resistance (Merdinoglu et al., 2018; Stam and McDonald, 2018). The study of varieties with different resistance genes can help us explore the mechanisms of resistance in *P. viticola*-grapevine interactions. Thus, this study initially screened four cultivars with mono-locus resistance (BC4, Bianca, F12P160 and Solaris) and subsequently two cultivars with pyramided resistance (F12P60 and F12P127).

Grapevine cultivar Bianca is a Bouvier and Villard Blanc hybrid, created in 1963 at the Kölyuktetö viticulture research facility in Hungary. Its resistance is given by the *Rpv3-1* locus located in chromosome 18 (Bellin et al., 2009). The cultivar Solaris was obtained by crossing the variety Merzling (*Rpv3-3*) with Gm6493 (*Rpv10*). It was created at the Geisenheim grape-breeding Institute (Germany) and is a carrier of resistance locus *Rpv10* that maps to chromosome 9 (Schwander et al., 2012). Both varieties are officially registered for use in wine production (http://www.vivc.de/). The resistance of the F12P160 genotype is explained by the *Rpv12* locus, located in chromosome 14 (Venuti et al., 2013). The cultivar BC4 was created in 2017 at INRA (France) as a cross between *Muscadinia* (*Rpv1*) X Regent (*Rpv3-1*). The *Rpv1* locus is responsible for its resistance and it maps to chromosome 12 (Merdinoglu et al., 2003). None of the two hybrids are officially registered for use in wine production. The latest cultivars, F12P60 and F12P127, are two pyramided hybrids created at Fondazione Edmund Mach (Italy). *Rpv3-1* and *Rpv12* are responsible for the resistance in cultivar F12P60 and they map to chromosomes 18 and 14, respectively (Bellin et al., 2009; Venuti et al., 2013). The resistance loci *Rpv3-1*, *Rpv3-3*, and *Rpv10* map to the chromosomes 18 and 9 and are engaged in the resistance of the cultivar F12P127 (Bellin et al., 2009; Di Gaspero et al., 2012; Schwander et al., 2012). Both varieties are not yet registered for cultivation.

Information about the different behavior of resistant and susceptible varieties coming from several cultivars is useful to understand the protection mechanisms involved in resistance to *P. viticola*. The plasticity of the plants in response to the pathogen is probably associated with the modulation of several classes of primary and secondary metabolites. For this reason, metabolomics is the most suitable approach in exploring the interaction between the grapevine and *P. viticola* and in extending the current knowledge about the perturbations of a wide range of molecules after biotic stress. To date, metabolomics studies have focused on several aspects: the differences between grapevine cultivars in berry composition in some cases (Mulas et al., 2011; Degu et al., 2014; Teixeira et al., 2014; Bavaresco et al., 2016), and the identification of metabolite changes in infected leaves in others (Ali et al., 2012). Some works focused on the metabolomic profiling of grapevine tissues infected with DM (Figueiredo et al., 2008; Ali et al., 2009; Buonassisi et al., 2017) and on metabolite changes due to the mono-locus resistance mechanism (Chitarrini et al., 2017, 2020). However, there is not yet a full description of which metabolites play a key role in resistance in the pyramiding resistance cultivars. This suggests the need to investigate further to identify the biomarkers of the defense response in resistant varieties.

In this study, we chose to examine first the reaction of primary and secondary metabolism of genotypes with mono-locus resistance against DM, and then we extended our investigation to the analysis of pyramided resistance genotypes. Among the hundreds of compounds identified, we decided to focus on those metabolites (not stilbenes and stilbenoids) that showed significant accumulation in resistant vs. susceptible genotypes over the course of the infection, and that can therefore be identified as putative markers of resistance. Within the class

of stilbenes and stilbenoids we decided to investigate not only the putative markers of resistance but also the markers of infection. The aim was to find previously unreported biomarkers of resistance, which are expected to pave the way for a better understanding of the different resistance mechanisms that underlie the hybrids-pathogen interaction affecting the *Vitis* species. All genotypes in the study were observed over 2 consecutive years and examined with a metabolomics approach for primary and secondary metabolism at 0, 12, 48, and 96 h post-inoculation. The assessment of the resistance level after artificial inoculation on leaves was carried out using the OIV-452 method (**Supplementary Table 1**).

# MATERIALS AND METHODS

## Plant Material and Artificial Inoculation

Grapevine plants with genotypes having different degrees of resistance to DM and one with a susceptible genotype were used in this study. The mono-locus resistance genotypes consisted of the varieties BC4, Bianca, F12P160 and Solaris whereas the pyramided resistance genotypes were F12P60 and F12P127 (**Table 1**). All the grapevine plants were grown in pots in controlled conditions in the Fondazione Edmund Mach grape germplasm collection located in San Michele all'Adige (Trento), Italy ($46^0$ 12′ 0″ N, $11^0$ 8′ 0″ E). The mono-locus resistance experiment was conducted in the 2 consecutive years 2016 and 2017; while the pyramided resistance experiment was conducted in the 2 consecutive years 2017 and 2018. For each experiment, the susceptible variety Pinot Noir was used as control genotype (**Table 1**).

During the experiment, the healthy plants ($n = 18$ per variety) were divided into two homogeneous groups (control and inoculated); the plants in the same group were further divided into three groups, each one representing one biological replicate (**Figure 1**). Plants were artificially infected with spores of the pathogen in the greenhouse. The inoculum was collected each year in late spring/early summer from naturally infected plants of the same untreated vineyard (grape cultivar: Pinot Noir) and was characterized by a mix of strains. Grapevine plants were inoculated by spraying the sporangial suspension at the rate of $1x10^6$ sporangia/mL on the lower surface of all leaves of plants, whereas the control plants were sprayed using milliQ water. Plants were kept in the greenhouse at a controlled temperature of 21°C and over 80% of relative humidity until the sampling. Leaves were sampled at four time points following a randomization scheme at 0, 12, 48, and 96 h post-inoculation/mock (**Figure 1**). Three biological replicates were sampled at each time point. Each sample was ground under liquid nitrogen and stored at −80°C until the extractions. The OIV-452 score was evaluated at 7 days post-inoculation on the first six fully expanded leaves (**Supplementary Table 1**) to assign a resistance score to *P. viticola* (leaves): 1 = very low 3 = low 5 = medium 7 = high 9 = very high or total. At the same time the Hypersensitive Response (HR) identified by the necrosis spots was evaluated.

## Extraction Procedures and Analysis of Compounds

### Primary Compounds

Primary compounds were extracted from 100 mg of fresh leaves and then subjected to derivatization using methoxamine hydrochloride in pyridine to inhibit the cyclization of reducing sugars and then with N-methyl-N-trimethylsilyl-trifluoroacetamide with 1% trimethylchlorosilane for trimethylsilylation following the Chitarrini et al. (2017) procedure. The derivatized extract was then injected for GC/MS analysis using a Trace GC Ultra with a fused silica RXI-5-Sil MS w/Integra Guard (30 m × 0.25 mm × 0.25 μm) column, combined with mass spectrometer TSQ Quantum GC (Thermo Electron Corporation) following the Chitarrini et al. (2017) parameters.

### Volatile Compounds

Volatile compounds were measured using a solid phase micro-extraction starting from 100 mg of fresh leaves and following the method of Chitarrini et al. (2017). Gas chromatography separation was done using a Trace GC Ultra gas chromatograph with a fused silica Stabilwax-DA column (30 m × 0.25 mm × 0.25 μm) (Restek Corporation) coupled to a Quantum XLS mass spectrometer (Thermo Electron Corporation) following the parameters of Matarese et al. (2014).

**TABLE 1 |** The genotypes used in this study, their source of resistance and their associated resistance-related loci (*Rpv*) with their references.

| Genotypes | | Resistance related loci (*Rpv*) | | | References |
|---|---|---|---|---|---|
| | | Downy mildew | Preliminary leaf resistance level | Source of resistance | |
| Mono-locus resistance | BC4 | *Rpv1* | Resistant | *M. rotundifolia* | Merdinoglu et al., 2003 |
| | Bianca | *Rpv3-1* | Resistant | *V. rupestris* | Bellin et al., 2009 |
| | F12P160 | *Rpv12* | Resistant | *V. amurensis* | Venuti et al., 2013 |
| | Solaris | *Rpv10* | Resistant | *V. amurensis* | Schwander et al., 2012 |
| Pyramided resistance | F12P60 | *Rpv3-1; Rpv12* | Resistant | *V. rupestris*<br>*V. amurensis* | Bellin et al., 2009; Venuti et al., 2013 |
| | F12P127 | *Rpv3-1, Rpv3-3; Rpv10* | Resistant | *V. rupestris*<br>*V. amurensis* | Bellin et al., 2009; Di Gaspero et al., 2012; Schwander et al., 2012 |
| Control | Pinot Noir | – | Susceptible | – | |

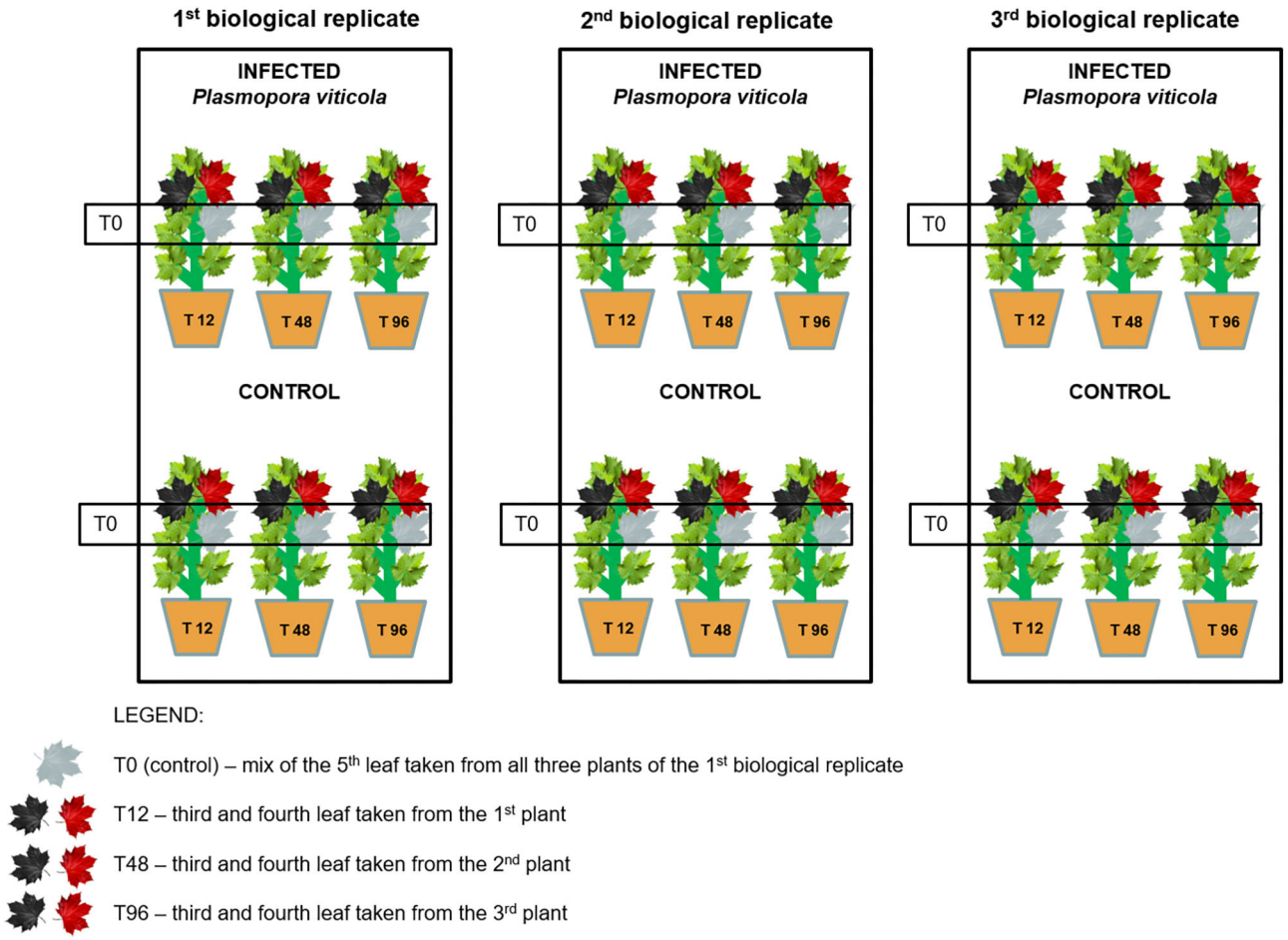



**FIGURE 1 |** Experimental design and randomization scheme.

### Lipidic Compounds

Lipid compounds analysis was done according to Della Corte et al. (2015) following the sample preparation described by Chitarrini et al. (2017). One hundred mg of fresh leaves were extracted using 0.3 mL of methanol; 0.6 mL of chloroform containing butylated hydroxyl toluene (500 mg/L); 0.25 mL water and then with 0.4 mL of chloroform containing butylated hydroxyl toluene (500 mg/L)/methanol/water 86:14:1 v/v/v; the combined lower lipid-rich layer was evaporated to dryness under $N_2$ and the samples were re-suspended in 300 µl of acetonitrile/isopropanol/water (65:30:5 v/v/v/). Samples were injected into a UHPLC Dionex 3000 (Thermo Fisher Scientific) with RP Ascentis column (15 cm × 2.1 mm; 2.7 µm C18) following a 30 min multi-step gradient coupled with an API 5500 triple-quadrupole mass spectrometer (Applied Biosystems/MDS Sciex) (Della Corte et al., 2015).

### Phenolic Compounds

The phenolic compounds were extracted from 100 mg of fresh leaves using 0.4 mL of chloroform and 0.6 mL of methanol:water (2:1); the extraction was repeated by adding 0.6 mL of methanol and water (2:1 v/v) and 0.2 mL of chloroform according to Vrhovsek et al. (2012) with some modifications, previously applied by Chitarrini et al. (2017). The aqueous-methanol phase of two extractions was collected, combined, and evaporated to dryness under $N_2$. Samples were re-suspended in 500 µl of methanol: water (1:1 v/v) and injected in a Waters Acquity UPLC system (Milford) with a Waters Acquity HSS T3 column (10 mm × 2.1 mm; 1.8 µm) coupled with a Xevo triple-quadrupole spectrometer (Waters) following Vrhovsek et al. (2012).

## Data Processing and Statistical Analysis

Data processing of primary and volatile compounds was performed using the software "Xcalibur" (version 4.0), whereas "Analyst" (version 1.7) and "MassLynx" (version 4.1) were used for processing lipids and phenols, respectively.

Lipid, phenols and primary compounds were identified using reference standards, retention time, quantifier and qualifier ion, and quantified using their standard calibration curves as mg/kg of fresh leaves. Volatile organic compounds were identified in the mass spectral database NIST MS Search 2.3 and results were semi

quantified as the equivalent of the internal standard (1-heptanol) and expressed as μg/kg of fresh leaves.

Statistical analysis and visualization were performed with R (R Core Team, 2020) relying on the following packages: tydiverse (Grolemund et al., 2019; Wickham et al., 2019) and egg (Baptiste, 2019) for data handling, manipulation and visualization; emmeans packages (Russell, 2020) for marginal means estimations; effsize for the effect size calculation (Sawilowsky, 2009; Torchiano, 2020). Logarithmic transformation was used to correct for the expected non-normality of metabolomics data. The average effect of each year was subtracted for each metabolite/genotype, to compensate for the expected year-to-year variability in the overall metabolic response. A linear modeling approach was used for each metabolite/genotype to assess the effects of time and artificial inoculationt (inoculated and non-inoculated). Cohen's d was used to estimate the size of the metabolic modulation induced by the pathogen inoculation for each time point. A metabolite was considered significantly perturbed if its concentration in the inoculated samples was significantly different from the control plants at least at one time point (uncorrected $p < 0.05$).

# RESULTS

## Dynamics of Metabolic Perturbations in Plant Defense Mechanism

In the 2 years considered, 264 compounds were identified in leaf samples under investigation. Among these, we quantified 175 compounds belonging to several classes: organic acids (29), amino acids (17), amines and others (12), sugars (25), benzoic acids derivatives (6), coumarins (3), dihydrochalcones (1), flavan-3-ols (11), flavanones (2), flavones (4), flavonols (15), phenylpropanoids (5), stilbenes and stilbenoids (13), fatty acids (15), glycerolipids (4), glycerophospholipids (2), prenols (1), sphingolipids (1), sterols (2) and other phenols (7). We semi-quantified 89 volatile organic compounds: volatile acids (5), alcohols (14), aldehydes (13), benzenoids (6), esters (3), hydrocarbons (1), other volatiles (6), fatty acids (2), benzofurans (1), terpenoids (10), terpenes (10), ketones (4) and unknown volatiles (14). In **Supplementary Table 2** the concentrations of VOCs (sheet 1) lipids (sheet 2) and polyphenols (sheet 3) identified as putative markers of resistance following the criteria described in section Putative Biomarkers of Resistance to *Plasmopara viticola* and The Effect of Pathogen Inoculation have been reported together with stilbenes and stilbenoids involved in the response to the infection and fight against the pathogen (sheet 3; see section Putative Biomarkers of Resistance to *Plasmopara viticola* and Stilbenes and Stilbenoids as Markers) for each genotype and for each year (**Supplementary Table 2**).

## Putative Biomarkers of Resistance to *Plasmopara viticola*

A global view of the metabolites that showed a significant effect after inoculation ($p < 0.05$ in at least one time point) is presented for the mono-locus resistant genotypes (BC4, Bianca, F12P160, Solaris) in **Figure 2** and for the pyramided resistant genotypes (F12P60, F12P127) in **Figure 3**.

In the plots, the dots indicate in which genotype(s) each metabolite was showing a significant difference in the inoculated vs. non-inoculated samples at least at one time point. This global visualization highlights that the resistant varieties Bianca, Solaris and F12P60 are showing a higher number of modulated metabolites. On the other hand, BC4 and F12P127 seem to show a more limited response to inoculation. The OIV-452 score was evaluated in the experiments (**Supplementary Table 1**) showing a very high degree of resistance for Bianca, F12P160, F12P60 and F12P127 (OIV-452 = 9); a high level for Solaris (OIV-452 = 7), medium for BC4 (OIV-452 = 5) and very low for Pinot Noir (OIV-452 = 1). At the same time, the hypersensitive response (HR) taking into account the necrosis was evaluated following the OIV-452 score. HR response was absent in Pinot Noir, medium in F12P160 and Bianca and high in BC4 and Solaris. For the pyramided genotypes, F12P127 was characterized by an high level of HR response, whereas the HR response was absent in F12P60 (**Supplementary Table 1**).

**Figures 2**, **3** clearly show that the interaction with the pathogen profoundly alters the plant metabolism, and some of the metabolites appear modulated after the artificial inoculation in both resistant genotypes and the susceptible Pinot Noir.

In order to pinpoint the most promising compounds, we defined the following for potential resistance biomarkers:

for metabolites excluding stilbenes and stilbenoids:

1. the metabolite was showing a significant modulation only in the resistant genotypes and, in addition, it was showing a large positive modulation (effect size d > 1) at the last two time points.

for stilbenes and stilbenoids:

1. the metabolite was showing a significant modulation only in the resistant genotypes and, in addition, it was showing a large positive modulation (effect size d > 1) at the last two time points (see section Stilbenes and Stilbenoids as Markers);
2. if modulated also in Pinot Noir, the metabolite was showing an effect size with a delta d > 1 compared with Pinot Noir (see section Stilbenes and Stilbenoids as Markers).

In the case of non-stilbenoids, we acknowledge that the magnitude and the timing of the accumulation of a compound could be important in characterizing the response of the plant to the pathogen attack (Pezet et al., 2004; Chitarrini et al., 2017), but the presence of a significant modulation also in Pinot Noir suggests that this metabolite is actually associated with infection. The second part of the first criterion (d > 1 in the last two time points), instead, stemmed from the hypothesis that the presence of the pathogen in the inoculated leaves was the main cause for the accumulation of the metabolites over time. In the case of stilbenoids, a more liberal criterion was applied since this class of compounds is known to hold a prominent role in the response of *V. vinifera* to pathogen infection; for these reasons we considered also those compounds with an effect size in the inoculated conditions with a delta d > 1 compared with Pinot Noir.

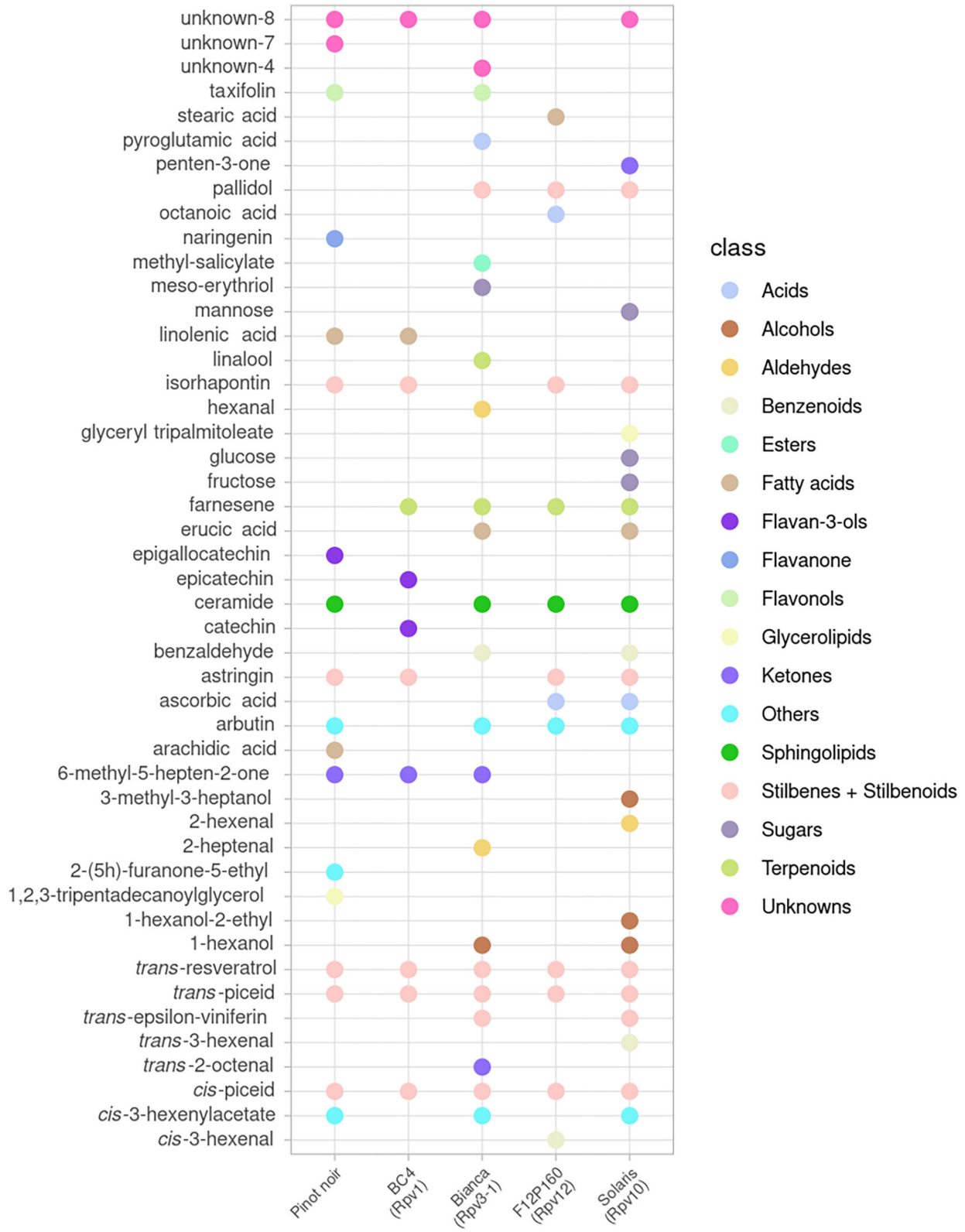



**FIGURE 2 |** Metabolites significantly modulated by the infection in at least one-time point for mono-locus resistant genotypes (BC4, Bianca, F12P160, Solaris) and for the susceptible Pinot Noir. All time points were considered in the 2 years of data analysis (2016–2017) and the color of each metabolite identifies the different chemical classes.


**FIGURE 3 |** Metabolites significantly modulated by the infection in at least one time point for the pyramided resistant genotypes (F12P60, F12P127) and for the susceptible Pinot Noir. All time points were considered in the 2 years of data analysis (2017–2018) and the color of each metabolite identifies the different chemical classes.

## The Effect of Pathogen Inoculation

The previous criteria led to the identification of 20 compounds, excluding the stilbenes and stilbenoids class (discussed in section Stilbenes and Stilbenoids as Markers), as putative biomarkers of resistance belonging to the plant primary metabolism: fatty acids (4) and secondary metabolism: flavan-3-ols (1), alcohols (4), aldehydes (2), benzenoids (1), benzoic acid esters (1), terpenoids (4), esters (1), and unknown volatiles (2) (**Table 2**). The concentrations of these compounds of interest are reported in **Supplementary Table 2**.

In order to discuss the strength of the modulation induced by the pathogen, the effect size (Cohen d) was calculated for each putative biomarker and for each time point (0, 12, 48, 96 hpi). According to the study of Sawilowsky (2009), the "d" values are associated with an effect size which can vary from a very small (d = 0.01) to a huge effect (d = 2.0). The "d" values of the identified putative biomarkers and their associated effect size are being presented in the **Supplementary Table 3**.

### BC4

In the resistant genotype BC4 we identified two compounds as putative biomarkers; one phenol, epicatechin, and one volatile, farnesene. Catechin and epicatechin have been recently identified as discriminatory factors, with a significantly higher amount in resistant/partial resistant plants (Maia et al., 2020). In our experiment, the effect size of epicatechin strongly grew at 48 and 96 hpi (1.99 and 1.64). Farnesene, instead, showed a higher and rapid accumulation after 12 hpi with high d values at 48 and 96 hpi (5.15 and 2.78) (**Figure 4**).

### Bianca

In the resistant genotype Bianca, six VOCs have been identified as potential biomarkers: 1-hexanol, erucic acid, benzaldehyde, farnesene, linalool, methyl-salicylate (**Figure 5**). In five of them we found an accumulation with a positive effect at both 48 and 96 hpi. The effect size of inoculation for 1-hexanol increased at 48 and 96 hpi, where it reached a positive effect (1.77 and 1.53, respectively). Linalool started increasing at 48 hpi (1.60), reaching a positive effect (3.05) at 96 hpi. Farnesene increased showing an effect size of 1.56 at 48 hpi and reaching a huge effect size of 4.23 at 96 hpi. The last two significant compounds of this resistant genotype, benzaldehyde, and methyl salicylate kept a positive effect immediately after the inoculation reaching an effect size at 48 and 96 hpi (benzaldehyde 1.66 and 3.16 at 48 and 96 hpi; methyl salicylate 2.96 and 2.61 at 48 and 96 hpi). Benzaldehyde was present also in F12P60 and Pinot Noir for 2017–2018, whereas methyl salicylate was detected in Pinot Noir for 2017–2018. Since in these cases the effect of the inoculation was much smaller, they remain putative biomarkers of resistance as initially assumed. Erucic acid reaches a peak with positive effect at 96 hpi (3.32).

**TABLE 2 |** Potential biomarkers among all metabolite classes except stilbenes and stilbenoids as identified by the selection criterion—modulation only in the resistant genotypes (d > 1).

| Class of the compounds | Compounds | GENOTYPES | | | | | |
|---|---|---|---|---|---|---|---|
| | | Mono-locus resistance | | | | Pyramided resistance | |
| | | BC4 (*Rpv1*) | Bianca (*Rpv3-1*) | F12P160 (*Rpv12*) | Solaris (*Rpv10*) | F12P60 (*Rpv3-1; Rpv12*) | F12P127 (*Rpv3-1, Rpv3-3; Rpv10*) |
| Fatty acids | erucic acid | | • | | | | |
| | oleic acid + *cis*-vaccenic acid | | | | | • | |
| | palmitic acid | | | | | • | |
| | stearic acid | | | • | | | |
| Flavan-3-ols | Epicatechin | • | | | | | |
| Alcohols | 1-hexanol | | • | | • | • | |
| | 1-hexanol-2 ethyl | | | | • | | |
| | (*E*)-2 hexenol | | | | | • | |
| | 1-octen-3-ol | | | | | • | |
| Aldehydes | 2-hexenal | | | | • | | |
| | nonanal | | | | | | • |
| Benzenoids | benzaldehyde | | • | | • | • | |
| Benzoic acid esters | methyl salicylate | | • | | | | |
| Terpenoids | farnesene | • | • | • | | | • |
| | linalool | | • | | | | |
| | (*E*)-nerolidol | | | | | • | |
| | neral | | | | | • | |
| Esters | *cis*-3-hexenyl benzoate | | | | | • | |
| Unknowns VOCs | unknown 4 | | | | | • | |
| | unknown 13 | | | | | • | |

### F12P160

For the resistant genotype F12P160, we identified farnesene and stearic acid in our inclusion criteria list. **Figure 6** highlights the interesting accumulation of farnesene with an increase of the effect size at 48 and 96 hpi (2.12 and 2.03, respectively).

### Solaris

In the resistant genotype Solaris, we identified four compounds: 1 hexanol, 1-hexanol-2-ethyl, 2-hexenal and benzaldehyde (**Figure 7**). All the four metabolites are accumulated at 48 hpi with a peak at 96 hpi and an effect size of 2.33, 1.61, 1.97 and 3.65.

### F12P127

The pyramided genotype F12P127 revealed two compounds in the inclusion criteria list, farnesene and nonanal (**Figure 8**). Farnesene was accumulated at 48 and 96 hpi with an effect size of 3.28 and 2.88, while nonanal showed an unclear trend among the time with an effect size of 1.48 and 1.10 at 48 and 96 hpi.

### F12P60

In F12P60 pyramided genotype, we identified eleven potential biomarkers (**Table 2**) in total. Benzaldehyde, as for F12P127 and Bianca genotypes, increased reaching an effect size of 4.92 and 4.76 at 48 and 96 hpi. Similar trends were found for (*E*)-2-hexenol (2.97 at 96 hpi) and 1-hexanol (2.74 at 96 hpi). The two terpenoids (*E*)-nerolidol and neral are accumulated after 24 hpi, with a peak at 48 hpi for (*E*)-nerolidol (2.17) and at 96 hpi for neral (2.27), respectively. Finally, we found a lipid compounds accumulation: oleic acid+*cis*-vaccenic and palmitic acid have an accumulation trend over time with an effect size of 7.01 at 48 hpi for palmitic aid and 4.44 and 4.5 for oleic acid+*cis*-vaccenic at 48 and 96 hpi (**Figure 9**).

## Stilbenes and Stilbenoids as Markers

Following the described criteria (see section Putative Biomarkers of Resistance to *Plasmopara viticola*), we found six significant compounds (**Table 3**); among them, pallidol and *trans*-epsilon-viniferin were the only compounds not modulated in Pinot Noir (for the concentrations seen **Supplementary Table 2** sheet 3). Pallidol reached the first criteria for stilbenes and stilbenoids in Bianca (effect size of 3.45 at 96 hpi), F12P160 (1.30 at 48 hpi and 3.07 at 96 hpi), and Solaris (2.02 at 48 hpi and 6.47 at 96 hpi); looking at the trend figures we found a comparable reaction in BC4 without a significant effect size (**Supplementary Figure 1**). The same situation is reported for *trans*-epsilon-viniferin, that reached the selected criteria in Bianca (3.17 at 96 hpi) and Solaris (1.19 at 48 hpi and 6.61 at 96 hpi) and reacted with a similar trend in F12P160 and BC4 but not with a significant effect size (**Supplementary Figure 1**).

The monomer *trans*-resveratrol was identified as significant in mono-locus resistant genotypes and in Pinot Noir comparing inoculated vs. not inoculated samples (**Supplementary Table 2**); anyhow, in mono-locusresistant genotypes the effect size was higher with a delta d > 1 compared with Pinot Noir. In the mono-locus genotypes the effect size had d > 1 already at 48 hpi with a peak at 96 hpi (Bianca 2.82; F12P160 3.41; Solaris 4.07; BC4 2.02); instead, in Pinot Noir we found an effect size of 1.5 at 96 hpi. As previously reported, *trans*-resveratrol has been identified as a monomer and precursor of active compounds in biotic stress plant defense (Langcake and Pryce, 1977; Jeandet et al., 2002). *trans*-Piceid, and *cis*-piceid were

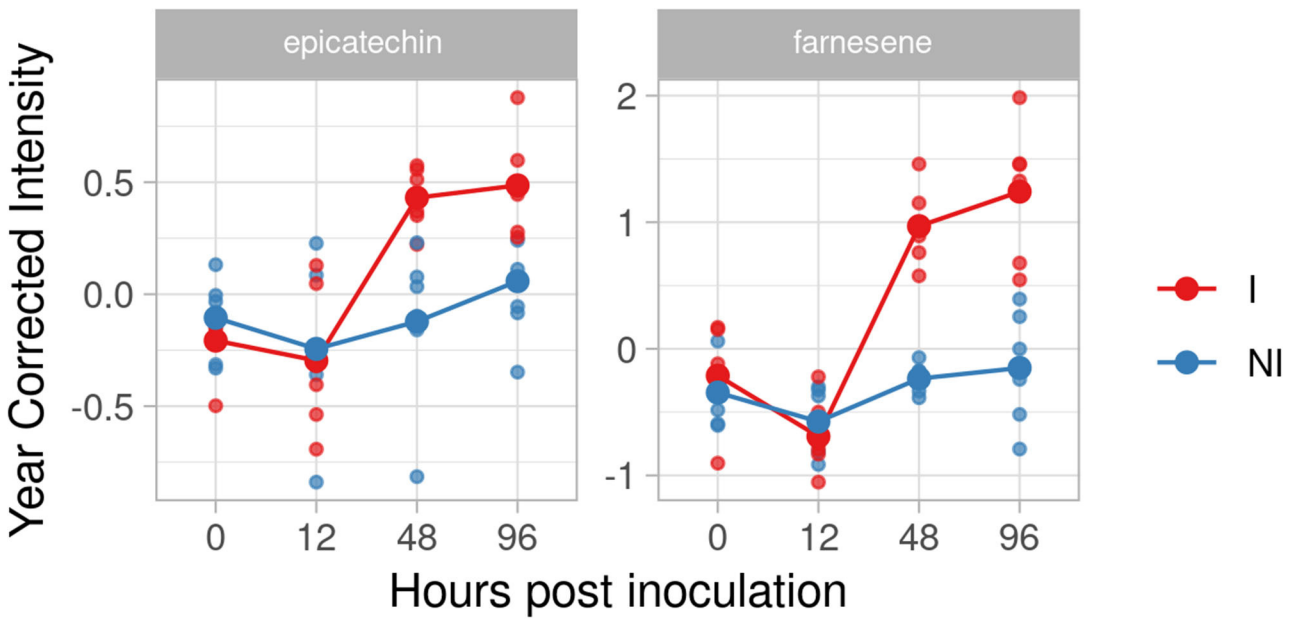



**FIGURE 4 |** Graphs for specific putative biomarkers of resistance to *Plasmopara viticola* in inoculated (Red) and not inoculated (Blue) BC4 genotype. Values of the 2 years are reported after subtracting the year's effect. I, inoculated; NI, not inoculated.

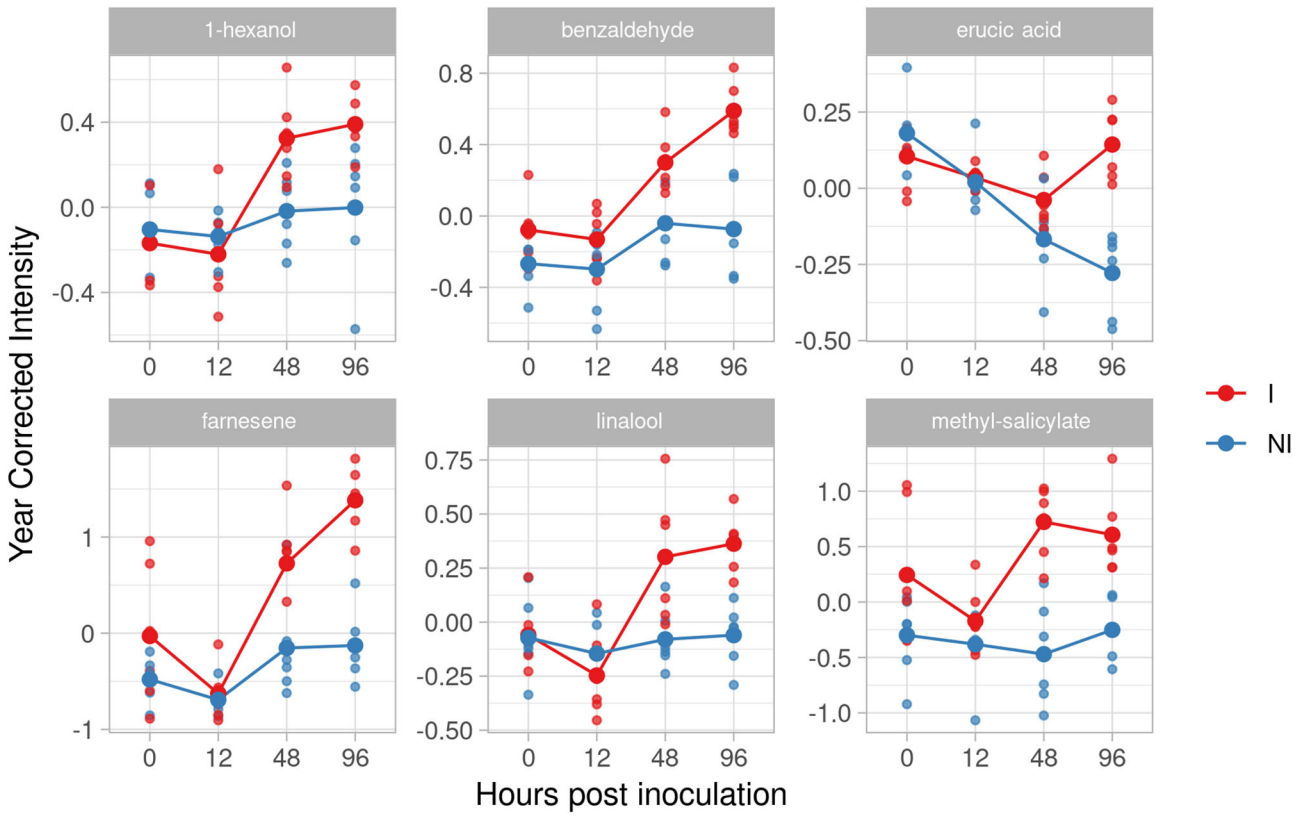



**FIGURE 5 |** Graphs for specific putative biomarkers of resistance to *Plasmopara viticola* in inoculated (Red) and not inoculated (Blue) Bianca genotype. Values of the 2 years are reported subtracting the year's effect. I, inoculated; NI, not inoculated.

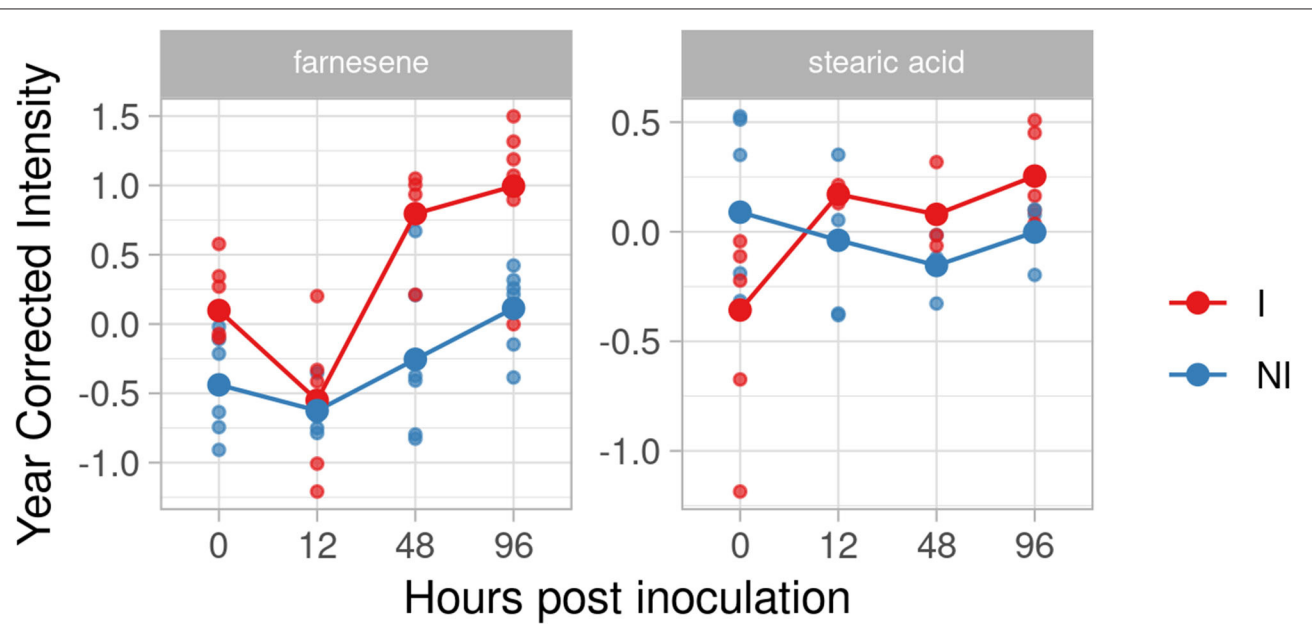



**FIGURE 6 |** Trend graph over time of putative biomarkers of resistance to *Plasmopara viticola* in F12P160 genotype inoculated (Red) and not inoculated (Blue). Values of the 2 years are reported subtracting the year's effect.

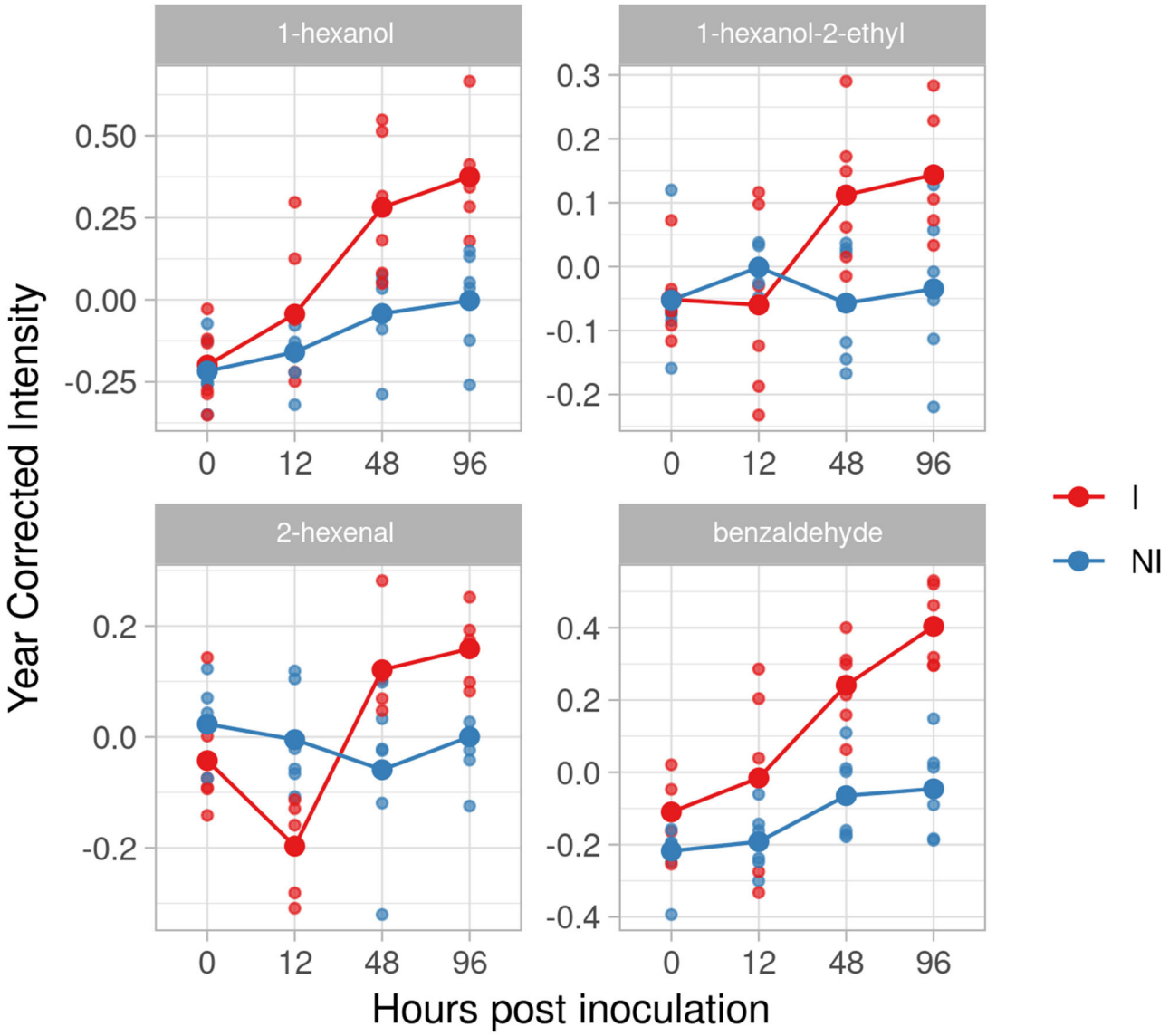



**FIGURE 7 |** Graphs for specific putative biomarkers of resistance to *Plasmopara viticola* in inoculated (Red) and not inoculated (Blue) Solaris genotype. Values of the 2 years are reported subtracting the year's effect.

identified as highly significant both in the mono-locus resistant varieties and in Pinot Noir and in the pyramided genotype F12P127 (**Figures 2**, **3**; **Supplementary Figures 1**, **2**). The effect size values showed an accumulation ($d > 1$) of these two compounds at 48 hpi and 96 hpi in the resistant genotypes while in Pinot Noir the accumulation has appeared only at 96 hpi (**Supplementary Table 3**). Astringin was significantly modulated in F12P160 (1.95 at 48 hpi and 1.72 at 96 hpi) and Solaris (1.50 at 48 hpi and 2.85 at 96 hpi) genotypes together with Pinot Noir (1.68 at 96 hpi). The trend of these compounds suggests a role in the response to biotic stress, supported by an early accumulation in the resistant genotypes compared to the susceptible one, but they are probably not directly involved in the defense against the pathogen. We are hypothesizing their modulation confirms that the artificial inoculation of the pathogen was successful. All the identified compounds increased with time after pathogen inoculation and their peak concentration was measured at 48 and 96 hpi (**Supplementary Figures 1**, **2**). The different behavior noticed for the phytoalexins agrees with the reports of Ali et al. (2010, 2012) who found that grapevine-specific phytoalexins can also be produced by the susceptible cultivars upon infection if we consider that at the beginning of the inoculation process the metabolic differences might be acting as the first inducible line of defense. Interesting accumulation was found for pallidol, and *trans*-epsilon-viniferin in all mono-locus genotype, with a significant effect size of these active compounds especially in Solaris at 48 and 96 hpi; these results confirm the importance of dimers biosynthesis and their accumulation in resistance

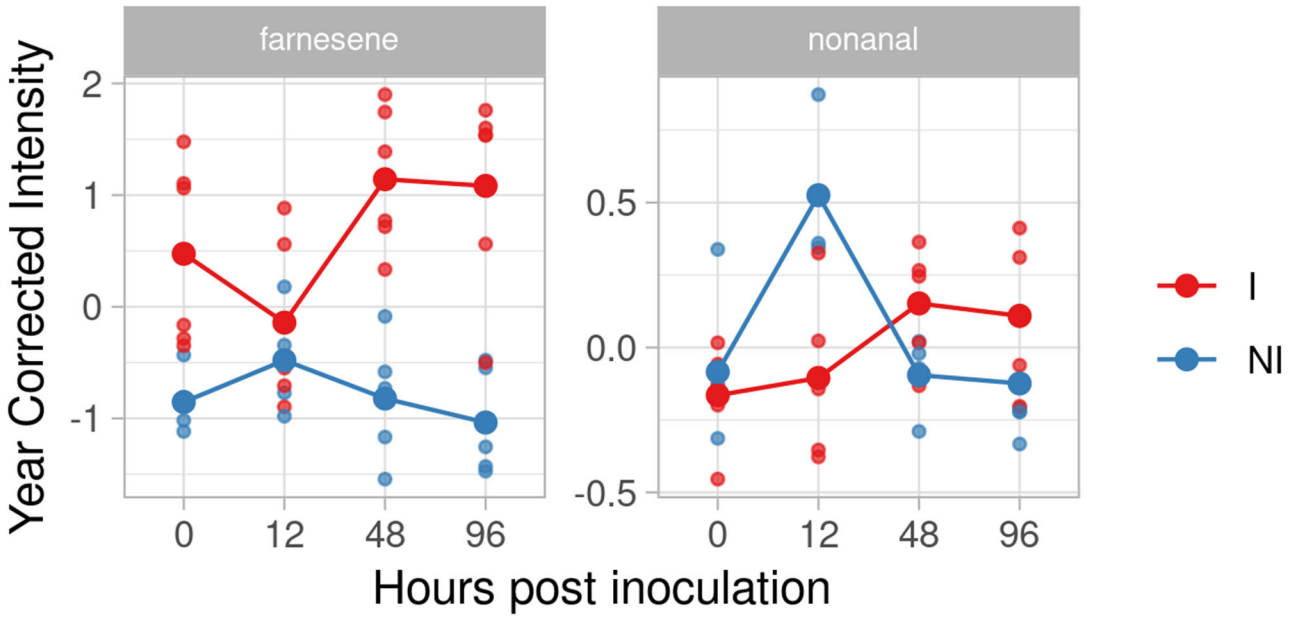



**FIGURE 8 |** Graphs for specific putative biomarkers of resistance to *Plasmopara viticola* in inoculated (Red) and not inoculated (Blue) F12P127 genotype. Values of the 2 years are reported subtracting the year's effect.

process (Malacarne et al., 2011; Bavaresco et al., 2012; Fröbel et al., 2019).

## DISCUSSION

In time, plants have developed different mechanisms of defense against abiotic and biotic stress. Among these mechanisms, the one between grapevine and *P. viticola* still raises questions concerning the interaction between the pathogen and the metabolism of the plant. It is already known that secondary metabolism has a defensive role against predators, parasites and diseases (Ali et al., 2010), but we shouldn't overlook the role of primary metabolism which, besides controlling the growth, development, and reproduction of plant species, also contributes to the plant defense. It can act as a source of energy, and it can signal molecules to directly or indirectly trigger defense response. This study showed findings of putative biomarkers in primary and secondary metabolism within resistant genotypes, as a defense response to *P. viticola.*

In the present 2-year study, we were able to use four analytical methods to identify and quantify or semi-quantify a large number of metabolites covering the most important compound classes. Among the extensive amount of obtained data, we arbitrarily choose to focus our investigation on the metabolites showing the most significant differences between inoculated vs. not inoculated samples, considering the time points with the criterion described in section Putative Biomarkers of Resistance to *Plasmopara viticola*.

An interesting aspect was observed in the alterations of the metabolism of most of the varieties, but mainly in mono-locus resistant genotypes. Several compounds identified as resistance putative biomarkers had their concentration reduced until 12 h after inoculation, followed by an increase at later time points. A similar reaction to the inoculation with DM was described by Ali et al. (2012) for quercetin-3-O-glucoside, glutamic acid and succinic acid in the resistant genotype Regent (*Rpv3-1*). Although we do not have substantial evidence to explain this behavior, we hypothesize that the pathogen might use these compounds to leak the necessary nutrients from the host cells, right before the activation of the plant defense.

During the infection, the pathogen disturbed the plant metabolism to different degrees. In F12P160 and Solaris, a decrease of the sugars was noticed at 12 h after inoculation, possibly because the pathogen was using them as a source of energy for its proliferation. Although sugars are mainly known in plants as a primary substrate to provide energy during the defense responses, they may also act as signal molecules interacting with the hormonal signaling network to regulate the plant immune system. In their role as plant resistance enhancers, sugars also stimulate the synthesis of flavonoids known as defense-related metabolites (Morkunas and Ratajczak, 2014).

In mono-locus resistant genotypes, the modulation of the metabolites was mainly noticed at 48 hpi and 96 hpi; this clearly indicates that 48 h after inoculation the plant defense mechanisms were active, just like Chitarrini et al. (2017) had noticed in a previous study. Solaris was an exception among the mono-locus resistant genotypes, as it reacted like the pyramided F12P60 genotype, where the modulation of 1-hexanol and benzaldehyde started earlier, between 0 and 12 hpi and reached its peak at 48 or 96 hpi. At the time of the experiments, one *Rpv* resistance gene was described in Solaris (**Table 1**); the latest report of Possamai et al. (2020) and Vezzulli et al.

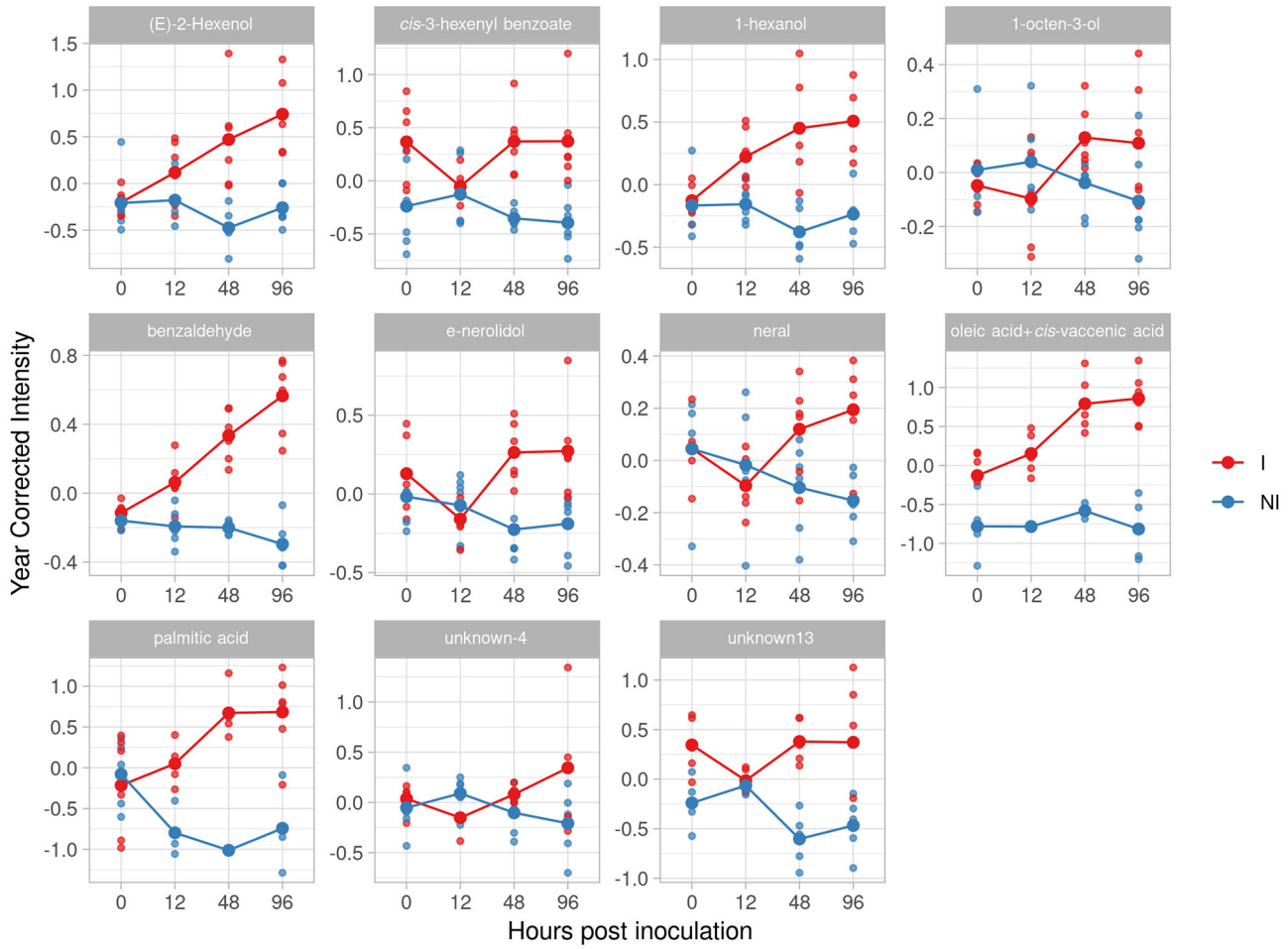



**FIGURE 9 |** Graphs for specific putative biomarkers of resistance to *Plasmopara viticola* in inoculated (Red) and not inoculated (Blue) F12P60 genotype. Values of the 2 years are reported subtracting the year's effect.

**TABLE 3 |** Potential biomarkers among stilbenes and stilbenoids as identified by the selection criterion.

| Compounds | Genotypes | | | | | | |
|---|---|---|---|---|---|---|---|
| | Susceptible | Mono-locus resistance | | | | Pyramided resistance | |
| | Pinot Noir | BC4 (*Rpv1*) | Bianca (*Rpv3-1*) | F12P160 (*Rpv12*) | Solaris (*Rpv10*) | F12P60 (*Rpv3-1; Rpv12*) | F12P127 (*Rpv3-1, Rpv3-3; Rpv10*) |
| *cis*-piceid | • | • | • | • | • | | • |
| *trans*-piceid | • | • | • | • | • | | • |
| *trans*-resveratrol | • | • | • | • | • | | |
| pallidol | | | • | • | • | | |
| *trans*-epsilon-viniferin | | | • | | • | | |
| astringin | • | | | • | • | | |

(2019) reveal the presence of two resistance sources in Solaris (*Rpv3-3* and *Rpv10*), explaining our results and supporting our conclusions. However, additional considerations at the genetic and metabolomics level should be made to fully support that the metabolic changes in Solaris are due to both *Rpv3-3+Rpv10*. The earlier activation of the defense response in the pyramided genotypes could be linked to the fact that the pathogen might take around 12 h to germinate and penetrate the leaf, inducing the first metabolic changes due to its colonization (Chitarrini et al., 2017). Another assumption is the presence of two or more resistance

sources for *P. viticola* within these genotypes. Besides ensuring a higher degree of resistance and a more stable and durable trait (Merdinoglu et al., 2018; Possamai et al., 2020) it could possibly also trigger a faster reaction against the pathogen.

In plants, lipids are energy storage and signaling compounds. In the defense against environmental factors and pathogens, they function as the structural components of cell membranes, which serve as permeable barriers to the external environment of cells. The accumulation of fatty acids (i.e., stearic acid, erucic acid, palmitic acid, oleic acid+*cis*-vaccenic acids) in plant metabolome after pathogen inoculation indicates their action in the adjustment of membrane fluidity mediated by desaturases and in the intracellular signaling processes (Nishida and Murata, 1996; Laureano et al., 2018) and their profile can be also involved in the protection of photosynthetic machinery in the early stages after the inoculation (Laureano et al., 2018). Thus, due to their role in activating the plant defense response, they are proposed as putative biomarkers. In plants, fatty acids have already been reported as important signaling molecules influencing genes involved in plant-microbe and plant-insect interaction (Savchenko et al., 2010; Walley et al., 2013). In previous experiments we found a decrease in oleic acid+*cis*-vaccenic acid together with other unsaturated fatty acids (16:1, 18:2, and 18:3) at the stage of 24 hpi in *Rpv3* and *Rpv12*-mediated resistance genotypes (Chitarrini et al., 2020). Previous studies report that the deactivation of the desaturase which converts stearic acid to oleic acids leads to an upregulation of salicylic acid (SA)-mediated responses and PR genes, with an inhibition of jasmonic acid (JA)-inducible defenses (Kachroo et al., 2008; Mandal et al., 2012). In our experiment we found an increase of palmitic acid and oleic acid+*cis*-vaccenic acid in F12P60; this situation, which is the opposite of what occurs with the mono-locus genotypes Bianca and F12P160, can be related to a different resistance response of the pyramided genotype.

A large variety of volatile compounds was emitted by the plants after the physiological stress induced by the *P. viticola* (green leaf volatiles, benzenoids, terpenoids, and some unknown compounds). This suggests that the secondary metabolism of the plant was seriously affected to a much higher degree by the pathogen. Green leaf volatiles (GLV) produced by the plant are volatile organic compounds that are released when plants suffer stress at the tissue level. Although the plants release GLVs constantly, they do so to a higher extent under conditions of stress (Hammerbacher and Coutinho, 2019). After pathogen inoculation, we identified two classes of GLVs that were released by plant leaves: alcohols and aldehydes. At physiologically relevant concentrations, a defense role of GLVs is suggested by this study based on their antifungal properties (Fallik et al., 1998). Plants are known to release *trans*-3-hexenal within minutes after they experience pathogen stress, and that such release can last for hours, after which it decreases in concentration as it undergoes enzymatic conversion to 2-hexenal (accumulated in our experiment in Solaris at 48 and 96 hpi) and unsaturated alcohols and esters (Davis et al., 2007). Chitarrini et al. (2017, 2020) had already suggested benzaldehyde as a putative biomarker of resistance, thanks to his role as a promoter of salicylic acid (SA)-mediated defense and its significant accumulation in the plant metabolome at 48 and 96 hpi, with an earlier accumulation in *Rpv12*-mediated resistance compared to the *Rpv3*-mediated one. This confirms our findings, and supports benzaldehyde being a biomarker also in the genotypes Solaris and F12P60, where it was found in significantly increased concentrations. Salicylic acid is the phytohormone precursor of the volatile methyl salicylate found in high concentration in the Bianca resistant genotype; in some plants, it is derived directly from the shikimate pathway in the plastids. Methyl salicylate is known for inducing systemic resistance after the attack of biotrophic organisms, like *P. viticola* (Hammerbacher and Coutinho, 2019).

The resistant grapevine genotypes in our study emitted significantly higher concentrations of terpenoids, both monoterpenes (linalool, neral) and sesquiterpenes (farnesene, (*E*)-nerolidol) than the susceptible genotype Pinot Noir. Hammerbacher and Coutinho (2019) found a positive correlation between an increased plant volatile emission and resistance to *P. viticola*. Algarra Alarcon et al. (2015) found a higher emission of sesquiterpenes and monoterpenes in grapevine genotypes resistant to *P. viticola*. Confirming their role in the fight against the pathogen, the antifungal activity of farnesene, and nerolidol together with ocimene and valencene have been recently tested by Ricciardi et al. (2021) showing a positive effect against the pathogen. In our experiment, farnesene was expressed in high concentrations in three mono-locus resistant genotypes (BC4, Bianca, F12P160) and included in the inclusion criteria for F12P127; linalool was significant only in Bianca genotype and (*E*)-nerolidol and neral were significant in the pyramided genotype F12P60.

The molecules of "unknown4" and "unknown13", have emerged in the pyramided genotype F12P60. Unfortunately, we do not have enough information about the chemical structure of these compounds; the likelihood of these molecules having a role in plant response to *P. viticola* infection is mentioned in the study by Lazazzara et al. (2018), who described an increase in the abundance of the unknown compounds in resistant genotypes compared to Pinot Noir. Nevertheless, further studies are required to identify the chemical structure and potential roles of these molecules.

Among the flavonoids, epicatechin has been identified in BC4 and, as per the studies of Ali et al. (2012) and Chitarrini et al. (2017); it plays a role in the resistance against pathogens, likely due to its antimicrobial properties.

The stilbenes and stilbenoids identified in mono-locus genotypes and F12P127 are produced through the phenylalanine/polymalonate pathway, and they can have a direct effect on fungal growth and sporulation by slowing down the growth of the pathogen and increasing plant resistance. Fröbel et al. (2019) found a significant induction of phenylalanine ammonium lyase (PAL) and stilbene synthase (STS) genes in *Rpv10* homozygous genotype stating the importance of the quantitative stilbenes produced to stop the pathogen. A recent study by Eisenmann et al. (2019) found that *Rpv3-1*-mediated resistance induces the production of toxic stilbenes and triggers programmed cell death, reducing, but not suppressing, the pathogen growth and development. The accumulation of

monomers (*trans*-resveratrol and *cis*- and *trans*-piceid) at the infection site is mainly related to the response to the pathogen inoculation, also found in the susceptible Pinot Noir. Instead, dimers biosynthesis and accumulation, significantly found only in resistant genotypes, can be related to the activity of these compounds against the pathogen (*trans*-epsilon viniferin and pallidol). These dimers have already been identified as markers of resistance representing key defense molecules because they are produced in response to biotic stress (Viret et al., 2018). Moreover, several studies (Del Rio et al., 2004; Atak et al., 2017) found a positive correlation between increased host resistance and an expression of a high content of phenolic compounds; indeed, according to Pezet et al. (2004) our observations demonstrate that stilbenes have significant inhibitory effects on the mobility of *P. viticola* zoospores and on subsequent disease development.

**Tables 2**, **3** give us a clear identification of the founded markers for each locus.

## CONCLUSIONS

This study describes different metabolic responses to the inoculation with *Plasmopara viticola* at various time points post-infection depending on the loci for resistance present in the genotypes.

To our knowledge, this work is the first study to investigate biomarkers present in mono-locus and pyramided-resistant cultivars. We first screen the genotypes with one *Rpv* resistant gene, afterwards we look for genotypes with pyramided resistance to find potential biomarkers associated with different types of resistance to *P. viticola*.

We identified several classes of compounds responsible for the diversification of the resistant cultivars from the susceptible one. We found an interesting modulation on stilbenes and stilbenoids, already known as biomarkers of resistance (dimers active compounds) in the *Vitaceae* and we confirmed the implication of benzaldehyde as a valid biomarker. We found an increase of terpenes emitted by the resistant genotypes confirming their role against the pathogen. Our findings suggest the possibility to test the pathogen inhibition by these VOCs compounds on receiving tissues and the future perspective to use it as a formulation. Interesting accumulations of fatty acids and volatile organic compounds were observed in the pyramided genotype F12P60 which is the variety with the greatest accumulation of potentially active compounds. The high accumulation of the remaining identified metabolites in the resistant genotypes, as compared to the susceptible Pinot Noir, suggests their possible involvement as biomarkers of resistance in a successful defense against *P. viticola*. Further experiments are required to test the putative compounds investigating their effect on infected tissues.

Overall, the results indicate that the way the cultivars responded to pathogen attacks can be linked to genotype and/or to resistant gene differences; however, resistance is not exclusively related to the *Rpv* genes. In our experiment we did not find a strict relation between mono-locus and pyramided response genotypes, even if they have the same *Rpv* genes. We found a higher accumulation of potential resistance biomarkers in Bianca Solaris and F12P60 genotypes. As expected, in the resistance genotypes we identified an Hypersensitive Response (HR) with cell death and necrosis. The pyramided F12P60 genotype that showed interesting metabolites modulation, did not provide any phenotypic evidence of the HR response. Finally, this study provides novel insights into the resistance mechanisms underlying the hybrids-pathogen interaction that could be valuable for the genetic improvement of grapevines.

## DATA AVAILABILITY STATEMENT

Metabolomics raw data are available from MetaboLights (Study Identifier: MTBLS2876, https://www.ebi.ac.uk/metabolights/MTBLS2876)

## AUTHOR CONTRIBUTIONS

RC, GC, LZ, MS, and UV designed the experiment. MS provided the plant material. RC, GC, and LZ performed the experiment. RC, GC, DŠ, and MR did the extractions and analytical analysis. PF, RC, and GC conducted the data treatment and statistical analysis. RC, GC, and UV prepared the manuscript. UV, GC, MO, and PR supervised the project. All authors discussed the results and implications and commented on the manuscript at all stages.

## FUNDING

This research was supported by Laimburg Research Centre (Vadena, Italy) and Fondazione Edmund Mach (San Michele all'Adige), Italy in collaboration with Università degli studi di Udine.

## ACKNOWLEDGMENTS

Cesare Lotti is acknowledged for his assistance, support, and guidance in GC-MS analysis. MR acknowledges the scholarship supported by the International Cooperation Program CAPES/PDSE Financed by CAPES—Brazilian Federal Agency for Support and Evaluation of Graduate Education.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.693887/full#supplementary-material

**Supplementary Figure 1 |** Stilbenes and Stilbenoids meeting the described criteria in mono-locus genotypes; inoculated (Red) and not inoculated (Blue).

**Supplementary Figure 2 |** Stilbenes and Stilbenoids meeting the described criteria in pyramided genotypes; inoculated (Red) and not inoculated (Blue).

**Supplementary Table 1 |** Degree of resistance to *Plasmopara viticola* (OIV-452-leaves) evaluated at 7 days post-inoculation on the first six fully expanded leaves; 1, very low; 3, low; 5, medium; 7, high; 9, very high or total; HR, Hypersensitive Response (necrosis).

**Supplementary Table 2 |** Concentrations of the 22 compounds identified as putative markers of resistance (see section The Effect of Pathogen Inoculation) (VOCs in sheet 1; Lipids in sheet 2 and Polyphenols in sheet 3) and the four stilbenes and stilbenoids involved in the response to the infection (see section Stilbenes and Stilbenoids as Markers) (sheet 3) reported for each genotype and for each year.

**Supplementary Table 3 |** The "d" values of the identified putative biomarkers for the mono-locus and pyramided varieties.

## REFERENCES


Algarra Alarcon, A., Lazazzara, V., Cappellin, L., Bianchedi, P. G., Schuhmacher, R., Wohlfahrt, G., et al. (2015). Emission of volatile sesquiterpenes and monoterpenes in grapevine genotypes following *Plasmopara viticola* inoculation *in vitro*. *J. Mass Spectrom.* 50, 1013–1022. doi: 10.1002/jms.3615

Ali, K., Maltese, F., Choi, Y. H., and Verpoorte, R. (2010). Metabolic constituents of grapevine and grape-derived products. *Phytochem. Rev.* 9, 357–378. doi: 10.1007/s11101-009-9158-0

Ali, K., Maltese, F., Figueiredo, A., Rex, M., Fortes, A. M., Zyprian, E., et al. (2012). Alterations in grapevine leaf metabolism upon inoculation with *Plasmopara viticola* in different time-points. *Plant Sci.* 191–192, 100–107. doi: 10.1016/j.plantsci.2012.04.014

Ali, K., Maltese, F., Zyprian, E., Rex, M., Choi, Y. H., and Verpoorte, R. (2009). NMR metabolic fingerprinting based identification of grapevine metabolites associated with downy mildew resistance. *J. Agric. Food Chem.* 57, 9599–9606. doi: 10.1021/jf902069f

Atak, A., Göksel, Z., and ÇelIk, H. (2017). Relations between downy/powdery mildew diseases and some phenolic compounds in *Vitis spp*. *Turkish J. Agric. For.* 41, 69–81. doi: 10.3906/tar-1610-61

Baptiste, A. (2019). *egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size. R package version 0.4.5*. Available online at: https://CRAN.R-project.org/package=egg

Bavaresco, L., Mattivi, F., De Rosso, M., and Flamini, R. (2012). Effects of elicitors, viticultural factors, and enological practices on resveratrol and stilbenes in grapevine and wine. *Mini Rev. Med. Chem.* 12, 1366–1381. doi: 10.2174/13895575112091366

Bavaresco, L., Rosso, M. De, Gardiman, M., Morreale, G., and Flamini, R. (2016). Polyphenol metabolomics of twenty Italian red grape varieties. *BIO Web Conf.* 7, 1–3. doi: 10.1051/bioconf/20160701022

Bellin, D., Peressotti, E., Merdinoglu, D., Wiedemann-Merdinoglu, S., Adam-Blondon, A. F., Cipriani, G., et al. (2009). Resistance to *Plasmopara viticola* in grapevine "Bianca" is controlled by a major dominant gene causing localised necrosis at the infection site. *Theor. Appl. Genet.* 120, 163–176. doi: 10.1007/s00122-009-1167-2

Bove, F., Bavaresco, L., Cffi, T., and Rossi, V. (2019). Assessment of resistance components for improved phenotyping of grapevine varieties resistant to downy mildew. *Front. Plant Sci.* 10:1559. doi: 10.3389/fpls.2019.01559

Buonassisi, D., Cappellin, L., Dolzani, C., Velasco, R., Peressotti, E., and Vezzulli, S. (2018). Development of a novel phenotyping method to assess downy mildew symptoms on grapevine inflorescences. *Sci. Hortic.* 236, 79–89. doi: 10.1016/j.scienta.2018.03.023

Buonassisi, D., Colombo, M., Migliaro, D., Dolzani, C., Peressotti, E., Mizzotti, C., et al. (2017). Breeding for grapevine downy mildew resistance: a review of "omics" approaches. *Euphytica* 213:103. doi: 10.1007/s10681-017-1882-8

Chitarrini, G., Riccadonna, S., Zulini, L., and Vecchione, A. (2020). Two - omics data revealed commonalities and differences between Rpv12 - and Rpv3 - mediated resistance in grapevine. *Sci. Rep.* 10:12193. doi: 10.1038/s41598-020-69051-6

Chitarrini, G., Soini, E., Riccadonna, S., Franceschi, P., Zulini, L., Masuero, D., et al. (2017). Identification of biomarkers for defense response to *Plasmopara viticola* in a resistant grape variety. *Front. Plant Sci.* 8:1524. doi: 10.3389/fpls.2017.01524

Davis, M. E., Gilles, M. K., Ravishankara, A. R., and Burkholder, J. B. (2007). Rate coefficients for the reaction of OH with (E)−2-pentenal, (E)−2-hexenal, and (E)−2-heptenal. *Phys. Chem. Chem. Phys.* 9, 2240–2248. doi: 10.1039/b700235a

Degu, A., Hochberg, U., Sikron, N., Venturini, L., Buson, G., Ghan, R., et al. (2014). Metabolite and transcript profiling of berry skin during fruit development elucidates differential regulation between Cabernet Sauvignon and Shiraz cultivars at branching points in the polyphenol pathway. *BMC Plant Biol.* 14:188. doi: 10.1186/s12870-014-0188-4

Del Rio, J. A., Gomez, P., Baidez, A., Fuster, M. D., Ortuno, A., and Frias, V. (2004). Phenolic compounds have a role in the defense mechanism protecting grapevine against the fungi involved in Petri disease. *Phytopathol. Mediterr.* 43, 87–94. doi: 10.14601/Phytopathol_Mediterr-1736

Della Corte, A., Chitarrini, G., Di Gangi, I. M., Masuero, D., Soini, E., Mattivi, F., et al. (2015). A rapid LC-MS/MS method for quantitative profiling of fatty acids, sterols, glycerolipids, glycerophospholipids and sphingolipids in grapes. *Talanta* 140, 52–61. doi: 10.1016/j.talanta.2015.03.003

Di Gaspero, G., Copetti, D., Coleman, C., Castellarin, S. D., Eibach, R., Kozma, P., et al. (2012). Selective sweep at the *Rpv3* locus during grapevine breeding for downy mildew resistance. *Theor. Appl. Genet.* 124, 277–286. doi: 10.1007/s00122-011-1703-8

Eisenmann, B., Czemmel, S., Ziegler, T., Buchholz, G., Kortekamp, A., Trapp, O., et al. (2019). *Rpv3 – 1* mediated resistance to grapevine downy mildew is associated with specific host transcriptional responses and the accumulation of stilbenes. *BMC Plant Biol.* 19:343. doi: 10.1186/s12870-019-1935-3

Fallik, E., Archbold, D., Hamilton-Kemp, T., Clements, A., Collins, R., and Barth, M. (1998). (E)−2-hexenal can stimulate botrytis cinerea growth *in vitro* and on strawberries *in vivo* during storage. *J. Am. Soc. Hort. Sci.* 123, 875–881. doi: 10.21273/JASHS.123.5.875

Figueiredo, A., Fortes, A. M., Ferreira, S., Sebastiana, M., Choi, Y. H., Sousa, L., et al. (2008). Transcriptional and metabolic profiling of grape (Vitis vinifera L.) leaves unravel possible innate resistance against pathogenic fungi. *J. Exp. Bot.* 59, 3371–3381. doi: 10.1093/jxb/ern187

Fröbel, S., Dudenhöffer, J., Töpfer, R., and Zyprian, E. (2019). Transcriptome analysis of early downy mildew (Plasmopara viticola) defense in grapevines carrying the Asian resistance locus Rpv10. *Euphytica* 215:28. doi: 10.1007/s10681-019-2355-z

Fröbel, S., and Zyprian, E. (2019). Colonization of different grapevine tissues by *Plasmopara viticola* — a histological study. *Front. Plant Sci.* 10:951. doi: 10.3389/fpls.2019.00951

Grolemund, G., Hayes, A., Henry, L., and Yutani, H. (2019). Welcome to the tidyverse. *J. Open Sour. Softw.* 4:1686.

Hammerbacher, A., and Coutinho, T. A. (2019). Roles of plant volatiles in defense against microbial pathogens and microbial exploitation of volatiles. *Plant Cell Environ.* 42, 2827–2843. doi: 10.1111/pce.13602

Jeandet, P., Douillet-Breuil, A. C., Bessis, R., Debord, S., Sbaghi, M., and Adrian, M. (2002). Phytoalexins from the *Vitaceae*: biosynthesis, phytoalexin gene expression in transgenic plants, antifungal activity, and metabolism. *J. Agric. Food Chem.* 50, 2731–2741. doi: 10.1021/jf011429s

Kachroo, A., Fu, D. Q., Havens, W., Navarre, D., Kachroo, P., and Ghabrial, S. A. (2008). An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. *MPMI* 21, 564–575. doi: 10.1094/MPMI-21-5-0564

Langcake, P., and Pryce, R. J. (1977). A new class of phytoalexins from grapevines. *Experientia* 33, 151–152. doi: 10.1007/BF02124034

Laureano, G., Figueiredo, J., Cavaco, A. R., Duarte, B., Caçador, I., Malhó, R., et al. (2018). The interplay between membrane lipids and phospholipase A family members in grapevine resistance against *Plasmopara viticola*. *Sci. Rep.* 8:14538. doi: 10.1038/s41598-018-32559-z

Lazazzara, V., Buesc, C., Parich, A., Pertot, I., Schuhmacher, R., and Perazzolli, M. (2018). Downy mildew symptoms on grapevines can be reduced by volatile organic compounds of resistant genotypes. *Sci. Rep.* 8:1618. doi: 10.1038/s41598-018-19776-2

Maia, M., Ferreira, A. E. N., Nascimento, R., Monteiro, F., Traquete, F., Marques A. P., et al. (2020). Integrating metabolomics and targeted gene expression to uncover potential biomarkers of fungal/oomycetes-associated disease susceptibility in grapevine. *Sci. Rep.* 10:15688. doi: 10.1038/s41598-020-72781-2

Malacarne, G., Vrhovsek, U., Zulini, L., Cestaro, A., Stefanini, M., Mattivi, F., et al. (2011). Resistance to *Plasmopara viticola* in a grapevine segregating population is associated with stilbenoid accumulation and with specific host transcriptional responses. *BMC Plant Biol.* 11:114. doi: 10.1186/1471-2229-11-114

Mandal, M. K., Chandra-Shekara, A. C., Jeong, R. D., Yu, K., Zhu, S., Chanda, B., et al. (2012). Oleic acid-dependent modulation of nitric oxide associated 1 protein levels regulates nitric oxide-mediated defense signaling in *Arabidopsis*. *Plant Cell* 24, 1654–1674. doi: 10.1105/tpc.112.096768

Matarese, F., Cuzzola, A., Scalabrelli, G., and D'Onofrio, C. (2014). Expression of terpene synthase genes associated with the formation of volatiles in different organs of *Vitis vinifera*. *Phytochemistry* 105, 12–24. doi: 10.1016/j.phytochem.2014.06.007

Maul, E., Töpfer, R., and Eibach, R. (2020). *Vitis International Variety Catalogue*. Available online at: www.vivc.de

Merdinoglu, D., Schneider, C., Prado, E., Wiedemann-Merdinoglu, S., and Mestre, P. (2018). Breeding for durable resistance to downy and powdery mildew in grapevine. *Oeno One* 52, 189–195. doi: 10.20870/oeno-one.2018.52.3.2116

Merdinoglu, D., Wiedemann-Merdinoglu, S., Coste, P., Dumas, V., Haetty, S., Butterlin, G., et al. (2003). Genetic analysis of downy mildew resistance derived from *Muscadinia rotundifolia*. *Acta Hortic* 603, 451–456. doi: 10.17660/ActaHortic.2003.603.57

Morkunas, I., and Ratajczak, L. (2014). The role of sugar signaling in plant defense responses against fungal pathogens. *Acta Physiol. Plant* 36, 1607–1619. doi: 10.1007/s11738-014-1559-z

Mulas, G., Grazia, G.M., Pretti, L., and Nieddu, G. (2011). NMR analysis of seven selections of vermentino grape berry : metabolites composition and development. *J. Agric. Food Chem.* 59, 793–802. doi: 10.1021/jf103285f

Nishida, I., and Murata, N. (1996). Chilling sensitivity in plants and cyanobacteria: the crucial contribution of membrane lipids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 541–568. doi: 10.1146/annurev.arplant.47.1.541

Nogueira Júnior, A. F., Tränkner, M., Ribeiro, R. V., von Tiedemann, A., and Amorim, L. (2020). Photosynthetic cost associated with induced defense to *Plasmopara viticola* in grapevine. *Front. Plant Sci.* 11:235. doi: 10.3389/fpls.2020.00235

Peressotti, E., Wiedemann-Merdinoglu, S., Delmotte, F., Bellin, D., Di Gaspero, G., Testolin, R., et al. (2010). Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. *BMC Plant Biol.* 10:147. doi: 10.1186/1471-2229-10-147

Pezet, R., Gindro, K., Viret, O., and Spring, J. L. (2004). Glycosylation and oxidative dimerization of resveratrol are respectively associated to sensitivity and resistance of grapevine cultivars to downy mildew. *Physiol. Mol. Plant Pathol.* 65, 297–303. doi: 10.1016/j.pmpp.2005.03.002

Possamai, T., Migliaro, D., Gardiman, M., Velasco, R., and De Nardi, B. (2020). Rpv mediated defense responses in grapevine offspring resistant to *Plasmopara Viticola*. *Plants* 9:781. doi: 10.3390/plants9060781

R Core Team (2020). *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna. Available online at: https://www.R-project.org/

Ricciardi, V., Marcianò, D., Sargolzaei, M., Maddalena, G., Maghradze, D., Tirelli, A., et al. (2021). From plant resistance response to the discovery of antimicrobial compounds: the role of volatile organic compounds (VOCs) in grapevine downy mildew infection. *Plant Physiol. Biochem.* 160, 294–305. doi: 10.1016/j.plaphy.2021.01.035

Russell, L. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.5.0*. Available online at: https://cran.r-project.org/web/packages/emmeans/index.html

Savchenko, T., Walley, J. W., Chehab, E. W., Xiao, Y., Kaspi, R., Pye, M. F., et al. (2010). Arachidonic acid: an evolutionary conserved signaling molecule modulates plant stress signaling networks. *Plant Cell* 22, 3193–3205. doi: 10.1105/tpc.110.073858

Sawilowsky, S. S. (2009). New effect size rules of thumb. *J. Modern Appl. Stat. Methods* 8:26. doi: 10.22237/jmasm/1257035100

Schwander, F., Eibach, R., Fechter, I., Hausmann, L., Zyprian, E., and Töpfer, R. (2012). *Rpv10*: a new locus from the Asian *Vitis* gene pool for pyramiding downy mildew resistance loci in grapevine. *Theor. Appl. Genet.* 124, 163–176. doi: 10.1007/s00122-011-1695-4

Stam, R., and McDonald, B.A. (2018). When resistance gene pyramids are not durable—the role of pathogen diversity. *Mol. Plant Pathol.* 19, 521–524. doi: 10.1111/mpp.12636

Teixeira, A., Martins, V., Noronha, H., and Eiras-dias, J. (2014). The first insight into the metabolite profiling of grapes from three *Vitis vinifera* L. cultivars of two controlled appellation (DOC) regions. *Int. J. Mol. Sci.* 15, 4237–4254. doi: 10.3390/ijms15034237

Torchiano, M. (2020). *Effsize - A Package for Efficient Effect Size Computation. R Package Version 0.8.1*. Available online at: https://CRAN.R-project.org/package=effsize

Venuti, S., Copetti, D., Foria, S., Falginella, L., Hoffmann, S., Bellin, D., et al. (2013). Historical introgression of the downy mildew resistance gene *Rpv12* from the Asian species *Vitis amurensis* into grapevine varieties. *PLoS ONE* 8:e61228. doi: 10.1371/journal.pone.0061228

Vezzulli, S., Malacarne, G., Masuero, D., Vecchione, A., Dolzani, C., Goremykin, V., et al. (2019). The *Rpv3-3* haplotype and stilbenoid induction mediate downy mildew resistance in a grapevine interspecific population. *Front. Plant Sci.* 10:234. doi: 10.3389/fpls.2019.00234

Vezzulli, S., Vecchione, A., Stefanini, M., and Zulini, L. (2018). Downy mildew resistance evaluation in 28 grapevine hybrids promising for breeding programs in Trentino region (Italy). *Eur. J. Plant Pathol.* 150, 485–495. doi: 10.1007/s10658-017-1298-2

Viret, O., Spring, J. L., and Gindro, K. (2018). Stilbenes: biomarkers of grapevine resistance to fungal diseases. *Oeno One* 52, 235–240. doi: 10.20870/oeno-one.2018.52.3.2033

Vrhovsek, U., Masuero, D., Gasperotti, M., Franceschi, P., Caputi, L., Viola, R., et al. (2012). A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in fruits and beverages. *J. Agric. Food Chem.* 60, 8831–8840. doi: 10.1021/jf2051569

Walley, J. W., Kliebenstein, D. J., Bostock, R. M., Dehesh, K. (2013). Fatty acids and early detection of pathogens. *Curr. Opin. Plant Biol.* 16, 520–526. doi: 10.1016/j.pbi.2013.06.011

Wickham, H., Averick, M., Bryan, J., Chang, W., D'Agostino McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Yildirim, Z., Atak, A., and Akkurt, M. (2019). Determination of downy and powdery mildew resistance of some *Vitis spp*. *Ciencia Tec. Vitivinic.* 34, 15–24. doi: 10.1051/ctv/20193401015

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 11 March 2022
doi: 10.3389/fpls.2022.826902

# Metabolomics Study of Different Germplasm Resources for Three *Polygonatum* Species Using UPLC-Q-TOF-MS/MS

*Shiqiang Wang[1†], Wenna Li[1†], Xinfei Zhang[1], Gang Li[1], Xiao dong Li[2], Hui Chang[3], Junfeng Niu[1*] and Zhezhi Wang[1*]*

[1] National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, The Key Laboratory of Medicinal Resources and Natural Pharmaceutical Chemistry, The Ministry of Education, College of Life Sciences, Shaanxi Normal University, Xi'an, China, [2] Lueyang Chinese Herbal Medicine Industry Development Service Center, Hanzhong, China, [3] Shaanxi Buchang Pharmaceuticals Limited Company, Xi'an, China

OPEN ACCESS

**Edited by:**
*Marta Sousa Silva, University of Lisbon, Portugal*

**Reviewed by:**
*Kamalrul Azlan Azizan, National University of Malaysia, Malaysia*
*Mostafa Abdelwahed Abdelrahman, Faculty of Science, Aswan University, Egypt*

***Correspondence:**
*Junfeng Niu*
*niujunfeng@snnu.edu.cn*
*Zhezhi Wang*
*zzwang@snnu.edu.cn*

*[†] These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science*

**Received:** *01 December 2021*
**Accepted:** *21 February 2022*
**Published:** *11 March 2022*

**Citation:**
*Wang S, Li W, Zhang X, Li G, Li Xd, Chang H, Niu J and Wang Z (2022) Metabolomics Study of Different Germplasm Resources for Three Polygonatum Species Using UPLC-Q-TOF-MS/MS. Front. Plant Sci. 13:826902. doi: 10.3389/fpls.2022.826902*

Rhizomes of the *Polygonatum* species are well-known in traditional Chinese medicine. The 2020 edition of *Chinese Pharmacopoeia* includes three different species that possess different pharmacological effects. Due to the lack of standardized discriminant compounds there has often been inadvertently incorrect prescriptions given for these medicines, resulting in serious consequences. Therefore, it is critical to accurately distinguish these herbal *Polygonatum* species. For this study, UPLC-Q-TOF-MS/MS based metabolomics was employed for the first time to discriminate between three *Polygonatum* species. Partial least squares discriminant analysis (PLS-DA) models were utilized to select the potential candidate discriminant compounds, after which MS/MS fragmentation patterns were used to identify them. Meanwhile, metabolic correlations were identified using the R language package corrplot, and the distribution of various metabolites was analyzed by box plot and the Z-score graph. As a result, we found that adenosine, sucrose, and pyroglutamic acid were suitable for the identification of different *Polygonatum* species. In conclusion, this study articulates how various herbal *Polygonatum* species might be more accurately and efficiently distinguished.

**Keywords: *Polygonatum*, species discrimination, metabolomics, UPLC-Q-TOF-MS/MS, iconic metabolites**

## INTRODUCTION

Rhizomes of the *Polygonatum* species are well-known traditional Chinese medicines that have been extensively applied for the treatment of many diseases over hundreds of years in China, Korea, Japan, and other East Asian countries (Zhao and Li, 2015). Approximately, 60 species have been identified worldwide, 31 of which are found in China. However, only three species have been selected for *Chinese Pharmacopoeia*. These include *Polygonatum sibiricum* Delar. ex Redoute, *Polygonatum cyrtonema* Hua, and *Polygonatum kingianum* Coll. et Hemsl.

The herb Huangjing. *P. sibiricum* can assist with promoting the proliferation and enhancing the viability of Bone Mesenchymal Stem Cells (BMSCs) (Zong et al., 2015), prevent Alzheimer's disease (Zhang et al., 2015), osteoporosis (Zeng et al., 2011), hypolipidemia and atherosclerosis (Yang et al., 2015), and improve immunologic functions (Wu et al., 2014). *P. cyrtonema* can induce apoptosis

and autophagy in human lung adenocarcinoma A549 cells (Liu et al., 2016), whereas *P. kingianum* may be applied to prevent T2D by its regulatory role in gut microbiota (Yan et al., 2017).

Furthermore, other *Polygonatum* species have a history of medicinal use, with various species having different medicinal effects. *P. cirrhifolium* can reduce oxidative stress for the protection of neurons (Xu et al., 2017); *P. odoratum* can be used to inhibit the proliferation of cancer cells (Tai et al., 2016); and *P. stenophyllum* can suppress menopausal obesity (Lee et al., 2016).

Since standardized discriminant compounds are not yet available for the *Polygonatum* species, they are distinguished only by their physiological appearance and morphological features. However, due to the very similar physical features and distribution areas of some *Polygonatum* species, this has often led to prescription confusion, with serious consequences (Zhao et al., 2006). Although some researchers have focused on elucidating the discriminant compounds of *Polygonatum*, they have not yet been successfully employed to accurately differentiate *Polygonatum* species. Thus, it is imperative to distinguish between closely related herbal medicines. Currently, most studies on *Polygonatum* have focused on the isolation of active components and their functions. According to the *Chinese Pharmacopoeia*, polysaccharide is the main medicinal component of *Polygonatum*, so we measured the polysaccharide content of different *Polygonatum*. However, the determination of polysaccharide content alone could not distinguish the different *Polygonatum* species. So we want to find out discriminant compounds to distinguish *Polygonatum* species more accurately. With the development of biometric techniques, the identification of medicinal plants based on their signature components is becoming increasingly popular. Ultra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry (UPLC/Q-TOF MS) offers higher resolution and sensitivity; thus, UPLC-Q-TOF-MS/MSbased metabolomics has been used to determinate the raw materials of various plants, including *Salvia officinalis* (Sarrou et al., 2017), *Cimicifugae Rhizoma* (Fan et al., 2017), *Carthamus tinctorius* (Yao et al., 2017), and *Plantago* (Yao et al., 2017).

For this study, UPLC-Q-TOF-MS/MS-based metabolomics was employed for the first time to accurately distinguish between herbal *Polygonatum* species (we chose the five *Polygonatum* with the significant difference of polysaccharide content, three of them are the same species but different regions), with analytical results revealing the most suitable metabolites that could reliably discriminate between them.

# MATERIALS AND METHODS

## Chemicals and Herbal Medicines

HPLC grade methanol was obtained from Merck (Germany), HPLC grade chloroform, formic acid, and acetonitrile were purchased from Thermo Fisher Scientific (United States), and ultrapure water generated by a Milli-Q Progard TS2 system (France) was used in all experiments.

Seventeen germplasm sources of *Polygonatum* were collected across China in 2015 (**Supplementary Table 1**), which were authenticated by Prof. Yaping Xiao (Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, Shaanxi Normal University). Voucher specimens were deposited at the National Engineering Laboratory for Resource Development of Endangered Chinese Crude Drugs in Northwest China, Xi'an, China.

## Isolation and Detection of Polysaccharides

The rhizome of *Polygonatum* were dried at 60°C for 48 h and ground to a powder, 100 g of which were mixed with petroleum ether (300 mL) at 60°C for 12 h to remove the lipids, which were then extracted via distilled water at 80℃ for 3 h at a 1:30 ratio (w/v) and repeated twice. The water extract was collected and concentrated using a rotary evaporator. After concentrating the liquid, 95% ethanol was added to a fourfold volume, which was maintained at 4℃ overnight to precipitate the polysaccharide and then percolated. The proteinaceous sediment was removed by repeated freezing/thawing, dialyzing, and lyophilizing to obtain the crude polysaccharide. The content of the polysaccharide was then calculated for *Polygonatum* (PSP).

PSP (%) content = W/M × 100%.

W, weight of crude polysaccharide.

M, weight of dried *Polygonatum* rhizome powder.

## Sample Preparation

According to the PSP content and four key indexes including leaf width, impeller number, total pedicle length and total flower number were taken into consideration, and the specific proportion weighting method was adopted to select the *Polygonatum sibiricum* Delar. ex Redoute from Lueyang County (Hanzhong, Shaanxi Province. Longitude: 33°17′56″; latitude: 105°51′12″; elevation: 1192 m), Foping County (Hanzhong, Shaanxi Province. Longitude: 33°52′86″; latitude: 108°01′71″; elevation: 1494 m), Luanchuan County (Luoyang, Henan Province. Longitude: 34°39′41″; latitude: 112°24′7″; elevation: 1193 m), *Polygonatum cyrtonema* Hua from Meishan City (Sichuan Province. Longitude: 30°06′02″; latitude: 103°86′13″; elevation: 1278 m) and *Polygonatum kingianum* Coll. et Hemsl from Lincang (Yunnan Province. Longitude: 23°87′58″; latitude: 100°06′94″; elevation: 1527 m) as the metabolomic samples, they are abbreviated to PSLY, PS (PSFP), PSLC, PC, and PK, and repeated the processing of each sample six times. The rhizome samples were placed in liquid nitrogen and then ground into fine powder, after which 80 mg of each sample was placed into a 5 mL glass centrifuge tube. Subsequently, 1 mL of 100% precooled (−20℃) methanol was added and the mixture underwent vortex oscillation for 30 s, which was then placed into the ultrasound unit at room temperature for 30 min. Chloroform (750 μL) was then added with $ddH_2O$ (4℃) 800 μL and underwent vortex oscillation for 1 min and centrifugation for 10 min at 10,000 rpm. The upper methanol/water mixture layer (1,000 μL) was transferred to a new 1.5 ml centrifuge tube, wherein the sample was concentrated using a vacuum centrifuge.

## Liquid Chromatography-Mass Spectrometry Analysis

Liquid chromatography analysis was performed using a SHIMADZU LC-30A Ultra High-Performance Liquid Chromatography system (SHIMADZU, Japan), equipped with a C18 column (1.7 μm, 2.1 × 100 mm) at 40℃, with a 4 μL injection volume. The mobile phase consisted of 0.1% formic acid water (A) and acetonitrile (B), using 2% B at 0–0.5 min, 2–50% B at 0.5–9 min, 50–98% B at 9–12 min, 98% B at 12–13 min, 98-2% B at 13 – 14 min, and 2% B at 14–15 min. The sample was dissolved in the mobile phase. The next sample was collected after 1 min of equilibration in the column. The flow rate was 0.3 mL/min and the autosampler temperature was maintained at 4℃.

The MS analysis was performed using an ABSCIEXTripleTOF$^{TM}$5600 LC/MS/MS system (United States) with a Duospray source. The Duospray source parameters included (ion voltage = 4,000 V and 5,500 V in negative and positive mode, declustering potential voltage was 80 V, source temperature was 600℃, curtain gas was 35 psi, Gas1 (nebulizer gas) was 60psi; Gas2 (heater gas) was 65psi; and the mass analyzer was scanned over m/z 50–1000, and the scan mode was MS/MS, IDA [each cycle scan response to the highest 8 ion scatter secondary scan)]. The dynamic background was excluded, and the instrument was recalibrated every five samples.

## Data Pretreatment and Multivariate Statistical Analysis

The raw LC/MS data of the 30 test samples were processed using XCMS[1] for peak deconvolution and alignment (**Supplementary Table 2**). The parameters of XCMS are set as follows: (a) xset = xcmsSet [snthresh = 6, method = "centWave," ppm = 15, mzdiff = 0.01, peakwidth = c(10,120)]; (b) group [xset,bw = 5, minfrac = 0.5, mzwid = 0.15]; (c) retcor [xset1,method = "obiwarp," plottype = c ("deviation")]; (d) group (xset2, bw = 5, minfrac = 0.5, mzwid = 0.15). 11,442 variables (positive ion mode) and 16,637 variables (negative ion mode) were obtained. The aligned data were exported to an Excel table and the peak area was normalized by the sum[2]. Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed using SIMCA-P version 13.0 (Umetrics, Sweden) after Pareto scaling. Only variables with VIP $>$ 1.0 and $P < 0.05$ were selected as discriminant compounds. Next, the MS/MS fragmentation patterns were employed as potential discriminant compounds in the Human Metabolome Database (HMDB)[3] (Wishart et al., 2009), Metlin[4] (Smith et al., 2005), and massbank[5] (Horai et al., 2010) and LipidMaps[6]. The R language[7] package Pheatmap was used to classify the identified potential discriminant compounds, after which the metabolic correlations were analyzed using the R language (see text footnote 7) package corrplot. For an intuitive view of the distribution of the different metabolites between groups, the box plot and the Z-score were used for analysis. Z-score is based on the relative conversion of metabolites and used to measure the relative level of metabolites. $Z = (x - \mu)/\sigma$. Where x is a specific fraction, μ is the mean, and σ is the standard deviation.

[1]www.bioconductor.org
[2]www.metaboanalyst.ca
[3]http://www.hmdb.ca
[4]http://metlin.scripps.edu/
[5]http://www.massbank.jp/
[6]http://www.lipidmaps.org
[7]www.r-project.org

# RESULTS

## *Polygonatum* Content of Different Germplasms

In this study, the crude water-soluble polysaccharide was obtained from the dried rhizomes of different germplasms. The specific location data is depicted in **Supplementary Table 1**. As shown (**Table 1**), the PSP content of PSLY was highest (13.33%), while the PSLC was lowest (6.37%). The PSP content of the PC was 10.83%, while for the PK was 11.76%. All differences in the data above were significant. In order to better evaluate the differences in metabolites of the three *Polygonatum* species, according to the yield of polysaccharides and other biological traits, we selected the germplasms of PSLY, PSLC, PS, PC, and PK as the metabolomics samples, among them, PSLY, PSLC, and PS are the same species but different regions.

## Identification of Compounds Using UHPLC-Q-TOF-MS

An UHPLC-QTOF-MS technique was established and successfully applied to characterize the constituents of the methanol extract of rhizome of *Polygonatum*. Total ion flow chromatogram of *Polygonatum*, including positive and negative ion modes (**Figure 1**). As a result, 7,064 variables (positive ion mode) and 13,780 variables (negative ion mode) were obtained (**Supplementary Table 2**).

## Multivariate Statistical Analysis

After several pretreatment procedures the data were applied to PCA to visualize the grouping trends. In the automatic simulation process, positive ion obtained five principal components, R2X = 0.444, Q2 = 0.572; The negative ion obtained five principal components, R2X = 0.472, Q2 = 0.529. The samples are basically in the 95% confidence interval (Hotelling $T^2$ Ellipse). As shown in **Figures 2A,B**, the PCA score plot showed a clear distinction between the PS, PC, and PK samples, while the separation between the PS, PSLY, and PSLC was not clear. To select the chemical markers responsible for such separation, the data set was applied to PLS-DA. PLS-DA for supervised analysis model of multivariate statistical analysis, compared with PCA can distinguish the differences between different metabolic category, image more clear can argue, and further to test the quality of the model using cross validation method, using the R2X (model can explain the variable) and Q2 (predictable degree) of the model for effectiveness evaluation model. R2X = 0.44, Q2 = 0.971 (positive ion mode), R2X = 0.468, Q2 = 0.956

**TABLE 1 |** Polysaccharide content of *P. sibiricum* in different germplasm samples.

| No | Origin | Species | Abbreviation | PSP Content % |
|---|---|---|---|---|
| 1 | Xiakou Yi, ShaanXi, China | *P. sibiricum* | PSXKY | 11.91 ± 0.13 deCD |
| 2 | Shangluo, ShaanXi, China | *P. sibiricum* | PSSL | 11.63 ± 0.12 eDE |
| 3 | ZhenAn, ShaanXi, China | *P. sibiricum* | PSZA | 12.11 ± 0.11 cdCD |
| 4 | Danfeng, ShaanXi, China | *P. sibiricum* | PSDF | 11.2 ± 0.36 fEF |
| 5 | Liuba, ShaanXi, China | *P. sibiricum* | PSLB | 8.85 ± 0.14 jI |
| 6 | Lueyang, ShaanXi, China | *P. sibiricum* | PSLY | 13.33 ± 0.1 aA |
| 7 | Foping, ShaanXi, China | *P. sibiricum* | PS | 12.65 ± 0.12 bB |
| 8 | Luoyang, HeNan, China | *P. sibiricum* | PSLC | 6.37 ± 0.11 kJ |
| 9 | Anyang, HeNan, China | *P. sibiricum* | PSAY | 10.39 ± 0.09 hG |
| 10 | Yingshan, HuBei, China | *P. sibiricum* | PSYS | 10.85 ± 0.08 fgFG |
| 11 | Shiyan, HuBei, China | *P. sibiricum* | PSSY | 8.62 ± 0.1 jI |
| 12 | Tiantai, ZheJiang, China | *P. sibiricum* | PSTT | 12.4 ± 0.08 bcBC |
| 13 | Jiangxian, ShanXi, China | *P. sibiricum* | PSJX | 10.65 ± 0.1 ghFG |
| 14 | Meishan, Sichuan, China | *P. cyrtonema* | PC | 10.83 ± 0.08 fgFG |
| 15 | Guiling, Guangxi, China | *P. cyrtonema* | PCGL | 9.48 ± 0.09 iH |
| 16 | Lincang, Yunnan, China | *P. kingianum* | PK | 11.76 ± 0.07 deD |
| 17 | Wenshan, Yunnan, China | *P. kingianum* | PKWS | 11.11 ± 0.14 fEF |

*PS, Polygonatum sibiricum; PC, Polygonatum cyrtonema Hua; PK, Polygonatum kingianum Coll. et Hemsl.*

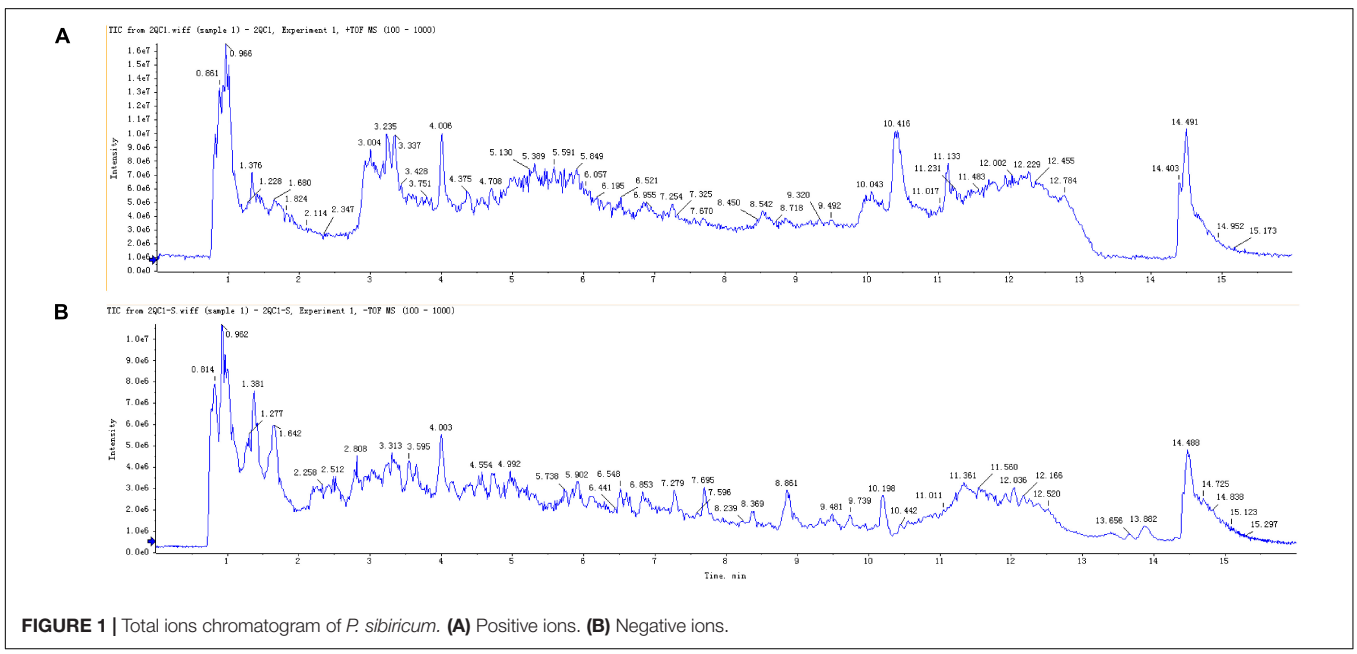



**FIGURE 1 |** Total ions chromatogram of *P. sibiricum*. **(A)** Positive ions. **(B)** Negative ions.

(negative ion mode) (**Figures 2C,E**). The abscissa is the first principal component score, represented by PC1; The ordinate is the second principal component score, represented by PC2, and the parentheses are the percentage of the explained model. R2Y (interpretation rate of supervision model) indicates that PLS-DA model can well explain the differences between each group of samples. After that, permutation test was performed on the PLS-DA model, and 100 times of positive ion mode verification results showed that the intercept of R2 on the *Y*-axis was 0.338, and that of Q2 on the *Y*-axis was −0.628 (**Figure 2D**). The 100 times verification results of negative ion mode show that the intercept of R2 on the *Y*-axis is 0.322, and that of Q2 on the *Y*-axis is −0.569 (**Figure 2F**). R2X is the explanatory degree of model X variable, R2Y is the explanatory degree of model Y variable, and Q2 is the predictability of the model. Theoretically, the closer R2X and Q2 are to 1, the better the model will be. The significance of variables using PLS-DA first principal component (VIP > 1), and metabolites with significant differences were screened with $P < 0.05$. In **Figures 3A,B**, each square of s-plot represents one compound, and compounds closer to the lower left and upper right corner represent greater contributions to each group of classification. Finally, 43 potential biomarkers were screened out (**Supplementary Table 3**). The scattered loading

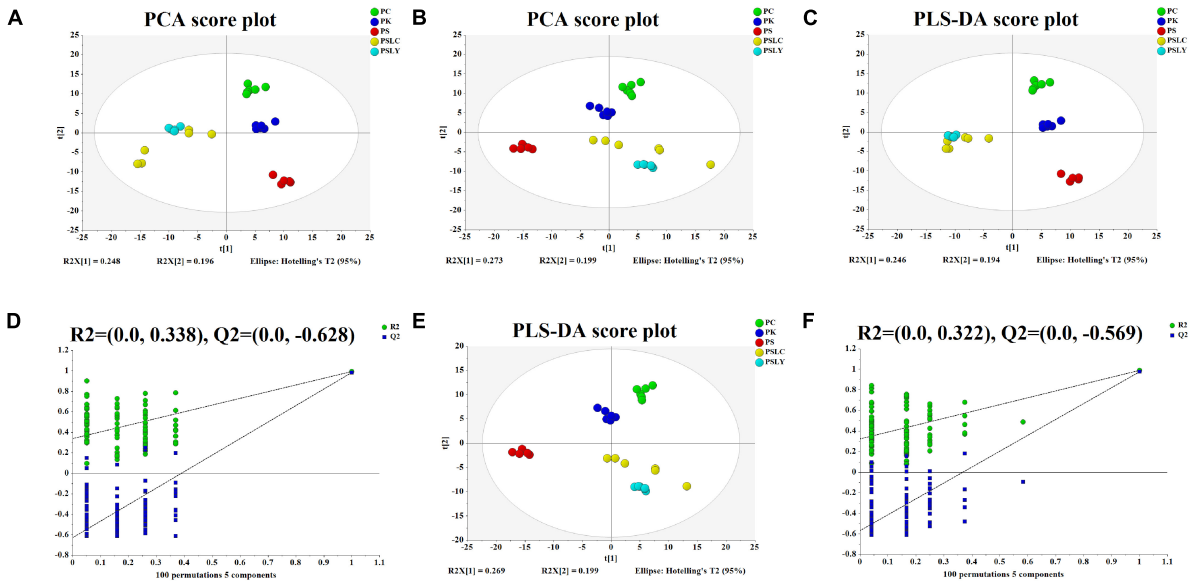



**FIGURE 2 |** PCA and PLS-DA score plots and the PLS-DA permutation test chart. **(A)** Positive ion mode of PCA score plot ($R^2X = 0.735$, $Q^2 = 0.572$). **(B)** Negative ion mode of PCA score plot ($R^2X = 0.736$, $Q^2 = 0.529$). **(C)** Positive ion mode of PLS-DA score plot ($R^2X = 0.734$, $R^2Y = 0.988$, $Q^2 = 0.971$). **(D)** Positive ion mode of PLS-DA permutation test. **(E)** Negative ion mode of PLS-DA score plot ($R^2X = 0.734$, $R^2Y = 0.983$, $Q^2 = 0.956$). **(F)** Negative ion mode of PLS-DA permutation test.

plot is shown in **Figure 4A** (positive ion mode) and **Figure 4B** (negative ion mode).

## Structural Annotation and Quantification of Potential Chemical Markers

According to the MS/MS fragmentation models in the HMDB, Metlin, massbank, and LipidMaps databases, the potential biomarkers (**Supplementary Table 3**) were identified using the exact molecular weights. As shown in **Supplementary Table 4** and **Supplementary Figure 1**, 25 key constituents were identified.

As shown in **Figure 5A**, a heat map of the 25 key constituents was used to visualize the relative level trends for all test samples. The different colors represent the relative levels, with red hues being higher and green hues being lower. From **Figure 5A**, there were significant differences in the contents of the five types of *Polygonatum* (PS, PC, PK, PSLY, and PSLC). These included adenosine, citric acid, guanine, L-pipecolic acid, L-tryptophan, pyroglutamic acid, and sucrose. The screening results, comparing each *Polygonatum* group are shown in **Table 2**. The results revealed that the iconic differential metabolites suitable for the identification of different *Polygonatum* species were adenosine, sucrose, and pyroglutamic acid. The iconic differential metabolites that were suitable for the identification of *P. sibiricum* of different regions included sucrose, citric acid, and pyroglutamic acid. Further, adenosine, sucrose, and pyroglutamic acid were appropriate for all five types of *Polygonatum* to distinguish.

For the 25 key constituents, we calculated the Pearson correlation coefficient and statistic test P value using the R language software (see text footnote 7) package corrplot. The highest correlation coefficient was −1 (red), which represented a completely positive correlation, and the lowest correlation coefficient was −1 (blue), which represented a completely negative correlation. The blank portion of the figure is the correlation statistic test p value $> 0.05$, whereas the colored portion is $p < 0.05$.

As shown in **Figure 5B**, the correlation analysis of the 25 key constituents intuitively showed the association between the different metabolites. Here, we focused on the iconic differential metabolites that were suitable for identifying all five *Polygonatum* species. These included adenosine, sucrose, and pyroglutamic acid, where sucrose was negatively correlated with adenosine and pyroglutamic acid.

For an intuitive view of the distribution of the different metabolites between groups, the box plot (**Figure 6**) was used for analysis. The box plot in **Figure 6A** shows the relative level of the three compounds in the *Polygonatum* extracts. The relative level of sucrose was highest in the *P. cyrtonema* extracts and about twofold higher than the *P. sibiricum* extracts, which had the lowest level between all *Polygonatum* species. However, the relative level of pyroglutamic acid and adenosine were highest in the *P. sibiricum* extracts. Finally, it was evident that the relative level of adenosine, sucrose, and pyroglutamic acid were significantly different for the five types of *Polygonatum*, which indicated that they could be employed for the identification of different *Polygonatum* species.

## DISCUSSION

The metabolites of identical plants can vary when grown in different environments, for different organs, different producing areas, or through various processing methods

(Souza et al., 2019). Consequently, the quality of the medicinal components of the *Polygonatum* species (particularly in terms of their chemical compositions) is uneven due to the mixed germplasm, transregional introduction, and diverse planting and processing procedures (Skubel et al., 2018). Moreover, traditional classification methods such as character (Xu et al., 2018), microscopic (Boiteau et al., 2018), and optical identification (Ullah et al., 2017), cannot directly reflect the nuances in their chemical compositions.

Metabolomics is a new discipline for the qualitative and quantitative analyses of small molecule metabolites, which has the advantages of high sensitivity, a low detection limit,

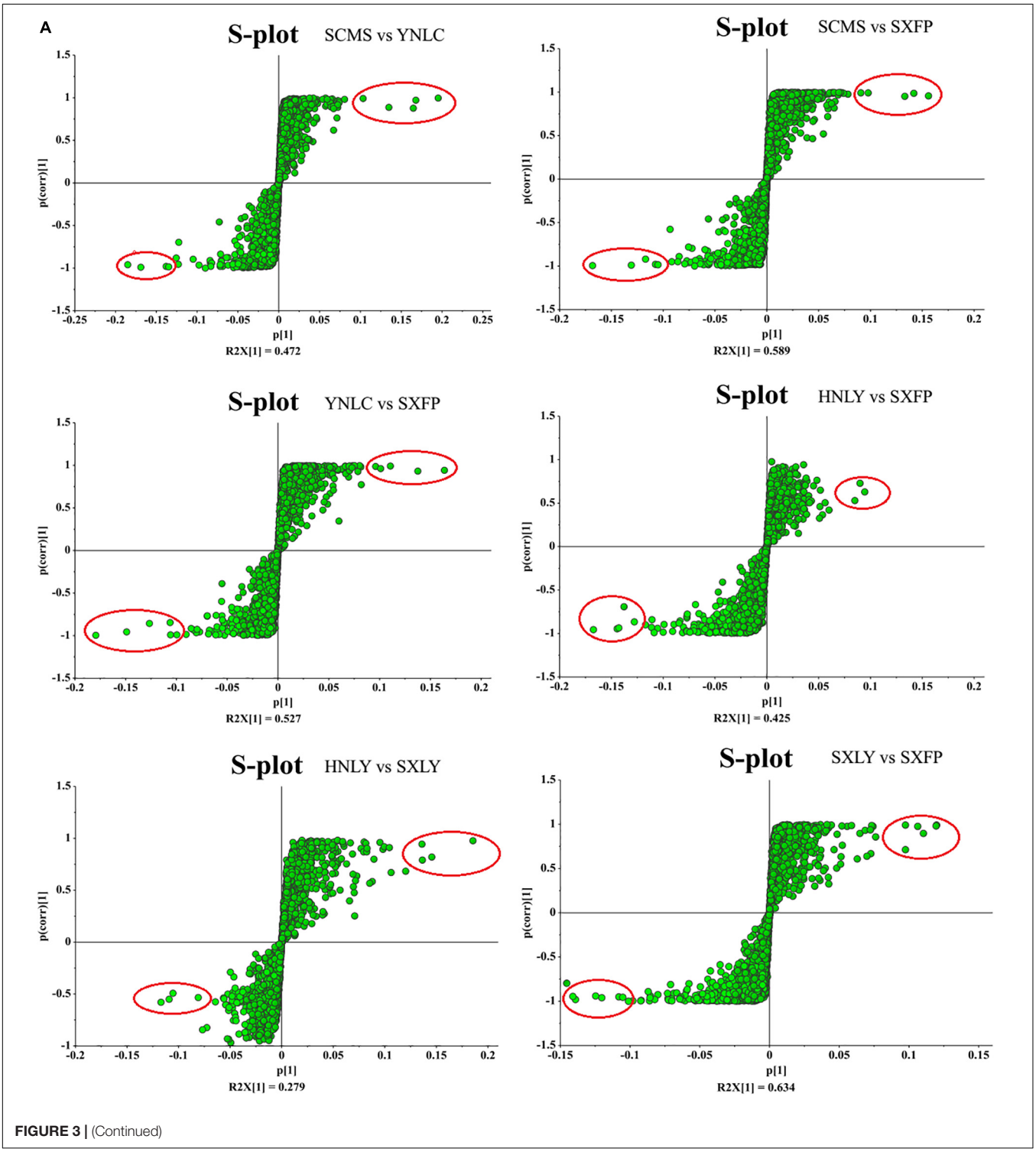



**FIGURE 3 |** (Continued)

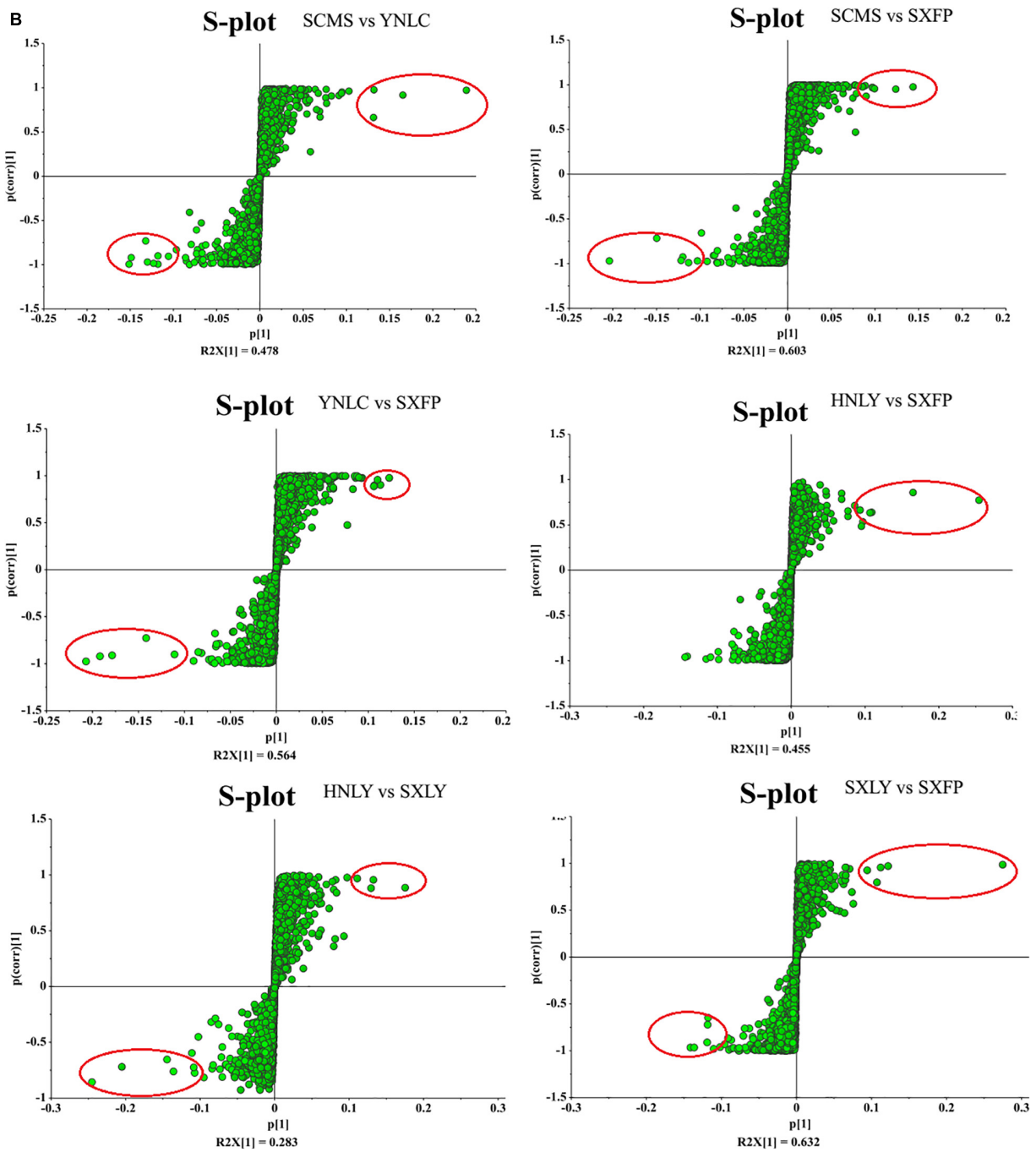



**FIGURE 3 | (A)** S-plot analysis of different origins of *P. sibiricum* (positive ion mode). **(B)** S-plot analysis of different origins of *P. sibiricum* (negative ion mode).

and small sample requirements, and is mostly used for the detection and analysis of plant signature components (Tian et al., 2014; Li et al., 2015; Puzanskiy et al., 2017). Therefore, we applied it for the identification of different types of Chinese medicinal *Polygonatum*, to facilitate the more rapid and accurate identification of Chinese medicinal ingredients. It is of great significance to analyze the metabolites of different germplasms. According to the *Chinese Pharmacopoeia*, polysaccharide is the main medicinal component of *Polygonatum*, so we measured the polysaccharide content of different *Polygonatum*. However, the determination of polysaccharide content alone could not distinguish the different *Polygonatum* species. Therefore, we combine the differences in polysaccharide content, as well four key indexes including leaf width, impeller number, total pedicle length and total flower number were taken into consideration, and the specific proportion weighting method was adopted to

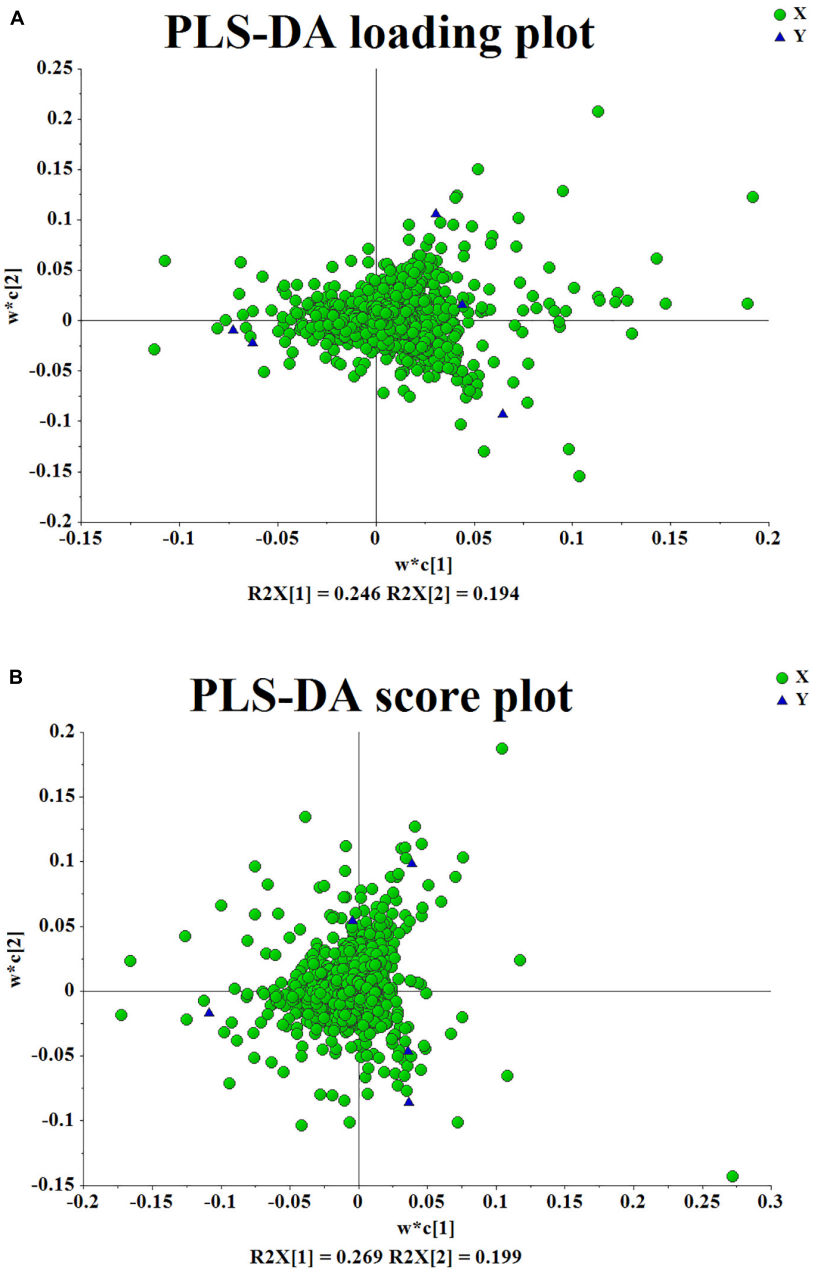



**FIGURE 4 | (A,B)** The scattered loading plot.

select. Finally, we selected PSLY, PS (PSFP), PSLC, PC, and PK for metabonomics analysis. Potential markers were screened by partial least-squares discriminant analysis (PLS-DA) and S-plot analysis. It was found that the models of each group were good, and had a high explanatory and predictive degree for X and Y variables.

According to heat map analysis, adenosine, sucrose and pyroglutamic acid were the most suitable metabolites for identification the different *Polygonatum*; Sucrose, citric acid and pyroglutamic acid were the characteristic metabolites suitable for identification of *Polygonatum* germplasm from different regions. Adenosine, sucrose and pyroglutamate acid were the metabolites that could distinguish the above five kinds of *Polygonatum* samples. Correlation analysis was conducted to focus on the metabolites that distinguish the five *Polygonatum*, namely adenosine, sucrose and pyroglutamate acid. Adenosine was negatively correlated with sucrose and positively correlated with pyroglutamate acid. Sucrose has a significant negative correlation with adenosine and pyroglutamic acid. After that, through the metabolite box diagram analysis, it was found that adenosine, sucrose and pyroglutamic acid were the signature metabolites that could distinguish the different species and the

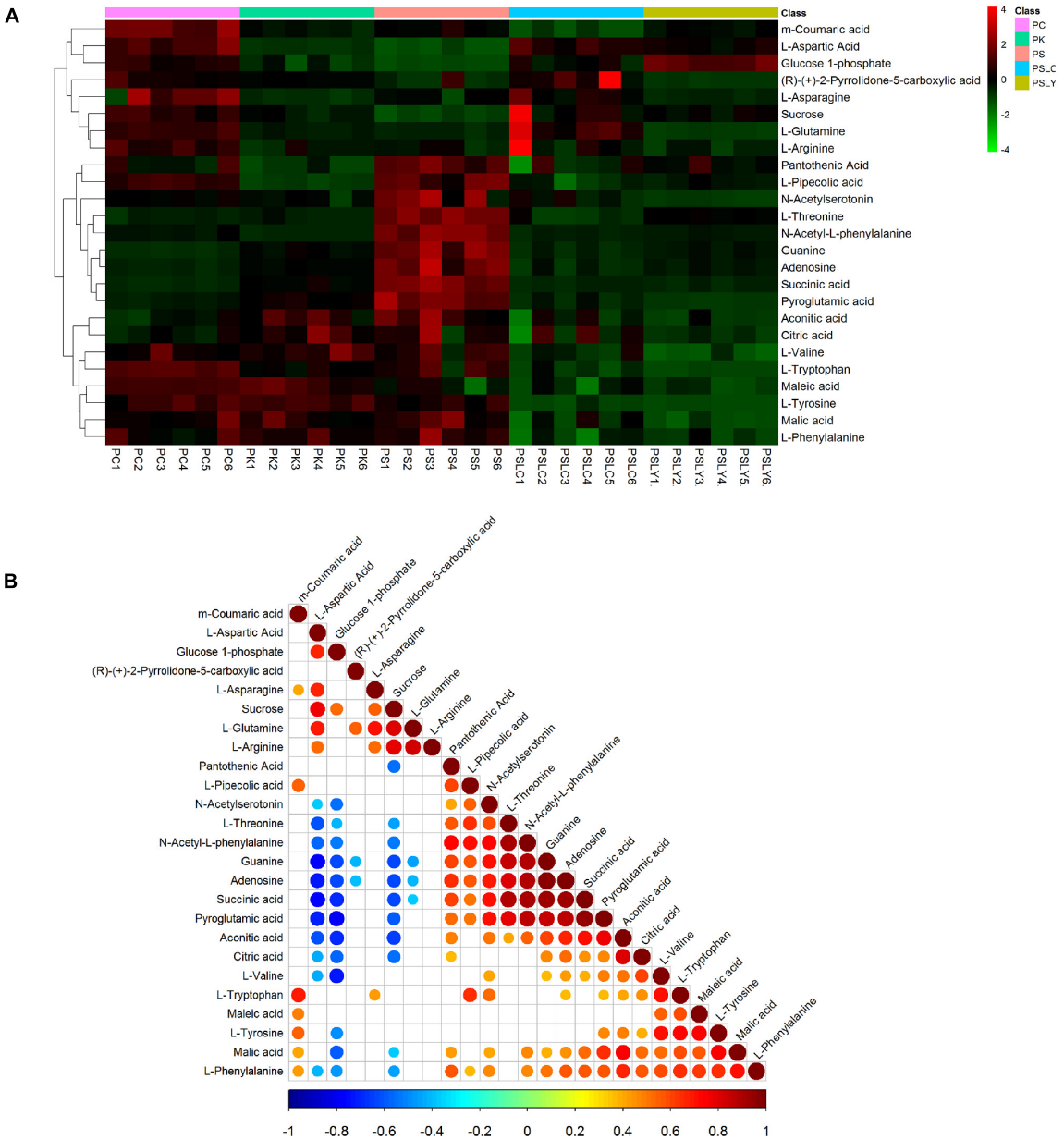



**FIGURE 5 |** Heat map and correlation analysis chart of differential metabolite. **(A)** Differential metabolite heat map. **(B)** Correlation analysis chart.

**TABLE 2 |** Discrepant metabolites.

| No. | mz | Rt (s) | exact_mass | mass_error | Metabolite | PC vs PK.vip | PC vs PS.vip | PK vs PS.vip | PSLC vs PS.vip | PSLC vs PSLY.vip | PSLY vs PS.vip |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 128.04 | 95.93 | 129.04 | 5 | Pyroglutamic acid | 5.20 | 7.71 | 7.06 | 8.75 | 4.07 | 8.80 |
| 2 | 130.09 | 83.12 | 129.08 | 1 | L-Pipecolic acid | 2.79 | 1.25 | 3.04 | 2.65 | 1.16 | 2.21 |
| 3 | 152.06 | 203.54 | 151.05 | 3 | Guanine | 1.39 | 2.41 | 2.43 | 2.54 | 1.04 | 2.18 |
| 4 | 191.02 | 85.06 | 192.03 | 3 | Citric acid | 8.17 | 4.90 | 1.14 | 4.05 | 5.26 | 7.36 |
| 5 | 205.10 | 253.37 | 204.09 | 3 | L-Tryptophan | 6.68 | 2.53 | 4.45 | 6.33 | 2.87 | 6.52 |
| 6 | 268.10 | 200.03 | 267.10 | 2 | Adenosine | 3.94 | 7.22 | 7.48 | 7.84 | 2.59 | 6.79 |
| 7 | 341.11 | 58.83 | 342.12 | 1 | Sucrose | 7.27 | 7.57 | 5.09 | 6.93 | 4.66 | 5.86 |

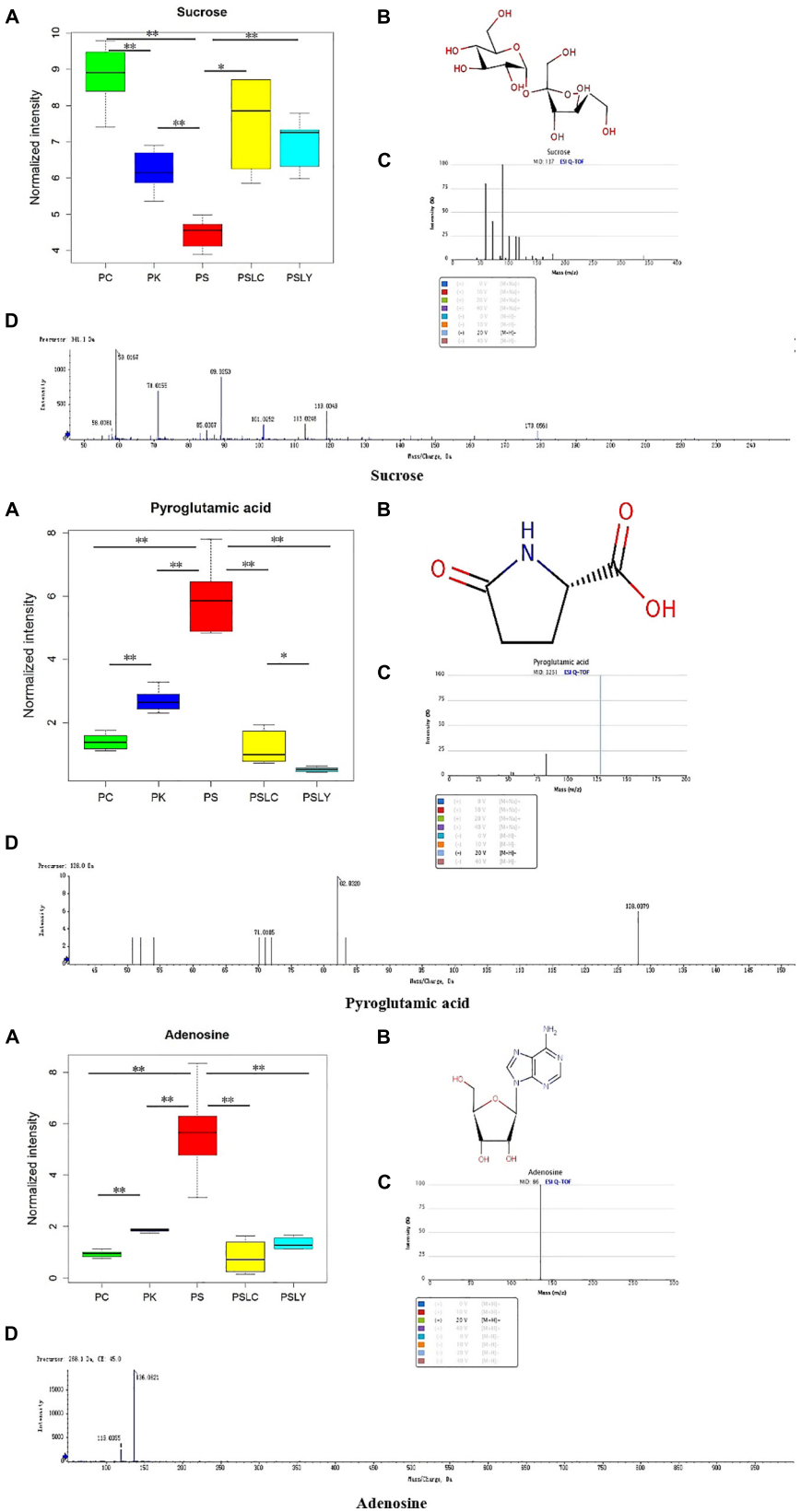



**FIGURE 6 |** Box diagram, spectra, and structural formulas of novel markers. **(A)** Box diagram. **(B)** Structural formulas. **(C)** Standard spectra of novel markers. **(D)** Detection spectra of novel markers. *represents significant differences, **represents extremely significant differences.

different germplasm. Sucrose is the precursor of polysaccharide synthesis pathway, which provides a new direction for germplasm differentiation and related metabolic pathway research of *Polygonatum*.

Sucrose is the main product of photosynthesis and is widely distributed in plants. Sucrose is broken down by digestive juices in the body's messaging system into fructose and glucose, which are absorbed through the small intestine (Szűts and Szabó-Révész, 2012). The sugars in sucrose are the building blocks of cells, such as their walls (Fewkes et al., 1971). Sugar is the main energy substance in living things. The sugars in sucrose are the precursors of many substances in the body, such as amino acids, nucleotides, fats, coenzymes, etc., all come from the intermediate products of glucose metabolism (Dissemond et al., 2020). Adenosine is an important intermediate for the synthesis of adenosine triphosphate (ATP), adenine, adenosine and adenosine (Novotný, 2015). Adenosine, as an endogenous nucleoside, can be phosphorylated to generate adenosine and participate in energy metabolism (O'Rangers, 1990). Pyroglutamate acid can stabilize intestinal structure, reduce intestinal inflammation and improve the structure and abundance of intestinal flora in mice fed a high-salt diet (Gunn et al., 2010). Furthermore, the development of plant metabonomics is bound to encounter a variety of difficulties and obstacles. Firstly, of prime importance is the limitation of the analytical methods (Kulkarni et al., 2018; Mai et al., 2020). Neither GC-MS or LC-MS can lengthy experiments (Fang and Luo, 2019). Secondly, data analysis methods are not efficient or convenient. Each metabonomics experiment generates a massive amount of data, and the task of data processing is heavy and complex (Li et al., 2015; Gargallo-Garriga et al., 2016). Consequently, more convenient data analysis methods are urgently needed to reduce the heavy workloads of researchers.

## CONCLUSION

This study provides a comprehensive profiling and putative identification of metabolites from *Polygonatum* through the UPLC-MS approach. For the first time, three different *Polygonatum* species were investigated simultaneously. After data screening and multivariate statistical analysis, the results revealed that the adenosine, sucrose, and pyroglutamic acid were suitable for identification of all five samples, which may be used for the further, more accurate identification of different *Polygonatum* species with potential as a chemical marker.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

JN and SW designed the research. WL, XZ and GL performed the experiments and data analysis. SW and WL analyzed the data and wrote the manuscript. ZW and JN advised on the result and discussions. All authors discussed the results and implications and commented on the manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Number 31670299); National Key Technologies R&D Program for Modernization of Traditional Chinese Medicine (Grant Numbers 2019YFC1712600, 2017YFC1701300, and 2017YFC1700706); Fundamental Research Funds for the Central Universities (Grant Numbers GK202103065 and GK201806006); Shaanxi Provincial Key R&D Program (2021SF-383, 2021SF-389, 2020LSFP2-21, and 2018FP2-26); Xi'an Science and Technology Project (20NYYF0057); and Research Project on Postgraduate Education and Teaching Reform of Shaanxi Normal University (GERP-20-41).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.826902/full#supplementary-material

**Supplementary Table 1 |** List of the specific location data of different germplasms.

**Supplementary Table 2 |** Raw LC/MS data of 30 test samples.

**Supplementary Table 3 |** Potential biomarkers for different species of *Polygonatum* (PS/PC/PK/PSLY/PSLC).

**Supplementary Table 4 |** Differences metabolites identified by MS/MS.

**Supplementary Figure 1 |** Key constituents identified by MS/MS spectrometry.

**Supplementary Figure 2 |** Z-score of differential metabolites.

## REFERENCES

Boiteau, R. M., Hoyt, D. W., Nicora, C. D., Kinmonth-Schultz, H. A., Ward, J. K., and Bingol, K. (2018). Structure Elucidation of Unknown Metabolites in Metabolomics by Combined NMR and MS/MS Prediction. *Metabolites* 8:8.

Dissemond, J., Augustin, M., Dietlein, M., Keuthage, W., Läuchli, S., Lobmann, R., et al. (2020). Sucrose octasulfate-evidence in the treatment of chronic wounds. *Hautarzt* 71, 791–801. doi: 10.1007/s00105-020-04637-9

Fan, M., Qin, K., Ding, F., Huang, Y., Wang, X., and Cai, B. (2017). Identification and differentiation of major components in three different "Sheng-ma" crude drug species by UPLC/Q-TOF-MS. *Acta Pharm. Sin. B* 7, 185–192. doi: 10.1016/j.apsb.2016.11.002

Fang, C., and Luo, J. (2019). Metabolic GWAS —based dissection of genetic bases underlying the diversity of plant metabolism. *Plant J.* 97, 91–100.

Fewkes, D. W., Parker, K. J., and Vlitos, A. J. (1971). Sucrose. *Sci. Prog.* 59, 25–39.

Gargallo-Garriga, A., Sardans, J., Pérez-Trujillo, M., Guenther, A., Llusià, J., Rico, L., et al. (2016). Shifts in plant foliar and floral metabolomes in response to the suppression of the associated microbiota. *BMC Plant Biol.* 16:78. doi: 10.1186/s12870-016-0767-7

Gunn, A. P., Masters, C. L., and Cherny, R. A. (2010). Pyroglutamate-Aβ: role in the natural history of Alzheimer's disease. *Int. J. Biochem. Cell Biol.* 42, 1915–1918. doi: 10.1016/j.biocel.2010.08.015

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–714. doi: 10.1002/jms.1777

Kulkarni, P., Wilschut, R. A., Verhoeven, K. J. F., van der Putten, W. H., and Garbeva, P. (2018). LAESI mass spectrometry imaging as a tool to differentiate the root metabolome of native and range-expanding plant species. *Planta* 248, 1515–1523. doi: 10.1007/s00425-018-2989-4

Lee, J. E., Kim, E. J., Kim, M. H., Hong, J., and Yang, W. M. (2016). Polygonatum stenophyllum improves menopausal obesity via regulation of lipolysis-related enzymes. *J. Nat. Med.* 70, 789–796. doi: 10.1007/s11418-016-1018-9

Li, L., Lu, X., Zhao, J., Zhang, J., Zhao, Y., Zhao, C., et al. (2015). Lipidome and metabolome analysis of fresh tobacco leaves in different geographical regions using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* 407, 5009–5020. doi: 10.1007/s00216-015-8522-8

Liu, T., Wu, L., Wang, D., Wang, H., Chen, J., Yang, C., et al. (2016). Role of reactive oxygen species-mediated MAPK and NF-κB activation in polygonatum cyrtonema lectin-induced apoptosis and autophagy in human lung adenocarcinoma A549 cells. *J. Biochem.* 160, 315–324. doi: 10.1093/jb/mvw040

Mai, Y., Yang, Z., Ji, X., An, W., Huang, Y., Liu, S., et al. (2020). Comparative analysis of transcriptome and metabolome uncovers the metabolic differences between Dendrobium officinale protocorms and mature stems. *All Life* 13, 346–359.

Novotný, J. (2015). Adenosine and its role in physiology. *Cesk Fysiol* 64, 35–44.

O'Rangers, E. A. (1990). Adenosine. *Conn. Med.* 54, 635–636.

Puzanskiy, R. K., Yemelyanov, V. V., Gavrilenko, T. A., and Shishova, M. F. (2017). The perspectives of metabolomic studies of potato plants. *Russian J. Genet.* 7, 744–756.

Sarrou, E., Ganopoulos, I., Xanthopoulou, A., Masuero, D., Martens, S., Madesis, P., et al. (2017). Genetic diversity and metabolic profile of Salvia officinalis populations: implications for advanced breeding strategies. *Planta* 246, 201–215. doi: 10.1007/s00425-017-2666-z

Skubel, S. A., Su, X., Poulev, A., Foxcroft, L. C., Dushenkov, V., and Raskin, I. (2018). Rapid, field-deployable method for collecting and preserving plant metabolome for biochemical and functional characterization. *PLoS One* 13:e0203569. doi: 10.1371/journal.pone.0203569

Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., et al. (2005). METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* 27, 747–751. doi: 10.1097/01.ftd.0000179845.53213.39

Souza, L. P. D., Alseekh, S., Naake, T., and Fernie, A. (2019). Mass Spectrometry—Based Untargeted Plant Metabolomics. *Curr. Protoc. Plant Biol.* 4:e20100.

Szűts, A., and Szabó-Révész, P. (2012). Sucrose esters as natural surfactants in drug delivery systems–a mini-review. *Int. J. Pharm.* 433, 1–9. doi: 10.1016/j.ijpharm.2012.04.076

Tai, Y., Sun, Y. M., Zou, X., Pan, Q., Lan, Y. D., Huo, Q., et al. (2016). Effect of Polygonatum odoratum extract on human breast cancer MDA-MB-231 cell proliferation and apoptosis. *Exp. Ther. Med.* 12, 2681–2687. doi: 10.3892/etm.2016.3630

Tian, Y., Amand, S., Buisson, D., Kunz, C., Hachette, F., Dupont, J., et al. (2014). The fungal leaf endophyte Paraconiothyrium variabile specifically metabolizes the host-plant metabolome for its own benefit. *Phytochemistry* 108, 95–101.

Ullah, N., Yüce, M., Neslihan Öztürk Gökçe, Z., and Budak, H. (2017). Comparative metabolite profiling of drought stress in roots and leaves of seven Triticeae species. *BMC Genom.* 18:696. doi: 10.1186/s12864-017-4321-2

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., et al. (2009). HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 37, D603–D610. doi: 10.1093/nar/gkn810

Wu, L. H., Lv, G. Y., Li, B., Zhang, Y. L., Su, J., and Chen, S. H. (2014). Study on effect of Polygonatum sibiricum on Yin deficiency model rats induced by long-term overload swimming. *Zhongguo Zhong Yao Za Zhi* 39, 1886–1891.

Xu, Q., He, Y., Yan, X., Zhao, S., Zhu, J., and Wei, C. (2018). Unraveling a crosstalk regulatory network of temporal aroma accumulation in tea plant (Camellia sinensis) leaves by integration of metabolomics and transcriptomics. *Environ. Exp. Bot.* 149, 81–94.

Xu, Y., Lin, W., Ye, S., Wang, H., Wang, T., Su, Y., et al. (2017). Protective Effects of an Ancient Chinese Kidney-Tonifying Formula against H(2)O(2)-Induced Oxidative Damage to MES23.5 Cells. *Parkinsons Dis.* 2017:2879495. doi: 10.1155/2017/2879495

Yan, H., Lu, J., Wang, Y., Gu, W., Yang, X., and Yu, J. (2017). Intake of total saponins and polysaccharides from Polygonatum kingianum affects the gut microbiota in diabetic rats. *Phytomedicine* 26, 45–54. doi: 10.1016/j.phymed.2017.01.007

Yang, J. X., Wu, S., Huang, X. L., Hu, X. Q., and Zhang, Y. (2015). Hypolipidemic Activity and Antiatherosclerotic Effect of Polysaccharide of Polygonatum sibiricum in Rabbit Model and Related Cellular Mechanisms. *Evid. Based Comp. Alter. Med.* 2015:391065. doi: 10.1155/2015/391065

Yao, C. L., Yang, W. Z., Si, W., Shen, Y., Zhang, N. X., Chen, H. L., et al. (2017). An enhanced targeted identification strategy for the selective identification of flavonoid O-glycosides from Carthamus tinctorius by integrating offline two-dimensional liquid chromatography/linear ion-trap-Orbitrap mass spectrometry, high-resolution diagnostic product ions/neutral loss filtering and liquid chromatography-solid phase extraction-nuclear magnetic resonance. *J. Chromatogr. A* 1491, 87–97. doi: 10.1016/j.chroma.2017.02.041

Zeng, G. F., Zhang, Z. Y., Lu, L., Xiao, D. Q., Xiong, C. X., Zhao, Y. X., et al. (2011). Protective effects of Polygonatum sibiricum polysaccharide on ovariectomy-induced bone loss in rats. *J. Ethnopharmacol.* 136, 224–229.

Zhang, H., Cao, Y., Chen, L., Wang, J., Tian, Q., Wang, N., et al. (2015). A polysaccharide from Polygonatum sibiricum attenuates amyloid-β-induced neurotoxicity in PC12 cells. *Carbohydr. Polym.* 117, 879–886. doi: 10.1016/j.carbpol.2014.10.034

Zhao, X., and Li, J. (2015). Chemical constituents of the genus Polygonatum and their role in medicinal treatment. *Nat. Prod. Commun.* 10, 683–688.

Zhao, Z., Hu, Y., Liang, Z., Yuen, J. P., Jiang, Z., and Leung, K. S. (2006). Authentication is fundamental for standardization of Chinese medicines. *Planta Med.* 72, 865–874. doi: 10.1055/s-2006-947209

Zong, S., Zeng, G., Zou, B., Li, K., Fang, Y., Lu, L., et al. (2015). Effects of Polygonatum sibiricum polysaccharide on the osteogenic differentiation of bone mesenchymal stem cells in mice. *Int. J. Clin. Exp. Pathol.* 8, 6169–6180.

**Conflict of Interest:** HC was employed by Shaanxi Buchang Pharmaceuticals Limited Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 13 June 2022
doi: 10.3389/fpls.2022.893055

# Metabolomic Analysis Identifies Differences Between Wild and Domesticated Chili Pepper Fruits During Development (*Capsicum annuum* L.)

*Felipe Cervantes-Hernández[1], Neftalí Ochoa-Alejo[2], Octavio Martínez[1] and José Juan Ordaz-Ortiz[1]**

[1] Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad de Genómica Avanzada, Irapuato, Mexico, [2] Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, Irapuato, Mexico

OPEN ACCESS

**Edited by:**
*Marta Sousa Silva, University of Lisbon, Portugal*

**Reviewed by:**
*Iris F. Kappers, Wageningen University and Research, Netherlands*
*Jose M. Manuel Palma, Spanish National Research Council (CSIC), Spain*

***Correspondence:**
*José Juan Ordaz-Ortiz*
*jose.ordaz.ortiz@cinvestav.mx*

**Specialty section:**
*This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science*

**Received:** *09 March 2022*
**Accepted:** *16 May 2022*
**Published:** *13 June 2022*

**Citation:**
*Cervantes-Hernández F, Ochoa-Alejo N, Martínez O and Ordaz-Ortiz JJ (2022) Metabolomic Analysis Identifies Differences Between Wild and Domesticated Chili Pepper Fruits During Development (Capsicum annuum L.). Front. Plant Sci. 13:893055. doi: 10.3389/fpls.2022.893055*

*Capsicum* spp. members are a rich source of specialized compounds due to their secondary metabolism. Some metabolic pathways have suffered modifications during the domestication process and improvement of agricultural traits. Here, we compared non-targeted LC–MS profiles from several areas: wild accessions (*C. annuum* L. var. *glabriusculum*), domesticated cultivars (*C. annuum* L.), and the F1 progeny of a domesticated, and a wild accession cross (in both directions) throughout seven stages of fruit development of chili pepper fruits. The main detected differences were in glycerophospholipid metabolism, flavone and flavonol biosynthesis, sphingolipid metabolism, and cutin biosynthesis. The domesticated group exhibited a higher abundance in 12′-apo-β-carotenal, among others capsorubin, and β-tocopherol. Palmitic acid and derivates, terpenoids, and quercitrin were prevalent in the wild accessions. F1 progeny showed a higher abundance of capsaicin, glycol stearate, and soyacerebroside I. This work supports evidence of the side-affectation of trait selection over the metabolism of chili pepper fruit development. Furthermore, it was also observed that there was a possible heterosis effect over the secondary metabolism in the F1 progeny.

**Keywords:** ***Capsicum annuum*, global profiling, metabolomics, domestication process, fruit development**

## INTRODUCTION

During the plant domestication process, humans have selected and managed phenotypes of plants that presented desired characteristics for crop yield and agricultural conditions. This conscious or unconscious process, has resulted in changes in the breeding phenotypes and genotypes, making them more useful to humankind (Doebley et al., 2006; Clement et al., 2010; Luna-Ruiz et al., 2018).

*Capsicum annuum* is one of the most important horticultural crops worldwide (Qin et al., 2014; Guevara et al., 2021; Martínez-Ispizua et al., 2021). Belonging to the Solanaceae family, originated and putatively domesticated from *Capsicum annuum* L. var. *glabriusculum* in México (Pickersgill, 1997; Hernández-Verdugo et al., 2001; Martínez-Ispizua et al., 2021), this species is a rich source of phytochemicals and nutrients, such as flavonoids, carotenoids, and vitamins that are important for

the human diet (Morales-Soto et al., 2013). In addition, the unique flavor and pungency are due to the existence of capsaicinoids, a group of alkaloids exclusively synthesized in the fruit of this plant species (Luna-Ruiz et al., 2018; Martínez-Ispizua et al., 2021).

The chili pepper domestication date is still under debate, but it has been estimated to be approximately 6,000 years BP based on the analysis of starch microfossils of domesticated accessions of chili pepper from seven Neotropical regions of America (Perry et al., 2007).

The domestication process resulted in a wide diversity of morphological, and phytochemical traits among the genus *Capsicum* (Hernández-Verdugo et al., 2001; Pickersgill, 2007; Qin et al., 2014). Some of the most appreciable changes in cultivars are the color, size, and form of the fruits (Hernández-Verdugo et al., 2001; Pickersgill, 2007; Qin et al., 2014; Solomon et al., 2021). Specialized metabolites from secondary metabolisms, such as capsaicinoids, carotenoids, and flavonoids, pass through alterations during domestication (Luna-Ruiz et al., 2018). All the latter are extensively reported to present shifts in their abundance across cultivars (Paran and van der Knaap, 2007; Wahyuni et al., 2011; Qin et al., 2014).

Wild accessions conserve high levels of genetic variation (Hernández-Verdugo et al., 2001; Hayano-Kanashiro et al., 2016) and thus it is reflected in their ability to adapt to different environments or diseases. Hernández-Verdugo et al. (2001), reported a wild accession (*C. annuum* var. *glabriusculum*) from the Norwest of México that was resistant to the *Pepper huasteco virus* (PHV). In contrast, most of the domesticated cultivars are poorly adapted to stressful environments; however, they are resistant to parasites and diseases. The domesticated cultivar "Criollo de Morelos landrace CM334" is a successful case of a genetic resistant inbred to *Phytophthora capsici* and nematodes (Pegard et al., 2005; Luna-Ruiz et al., 2018).

To date, shifts in phytochemical profiles of *C. annuum* have been minorly explored from a holistic perspective. Most of the metabolites in *C. annuum* remain unknown (Luna-Ruiz et al., 2018). Some metabolites have been reported for *C. annuum* L. species, for example, Pascale et al. (2020), outlined sixteen quercetin glycoconjugates in peperoni di Senise peppers; among them, five which were tentatively identified as quercetin-(galloyl-rhamnoside)-hexoside, quercetin-(sinapoyl-hexoside)-rhamnoside, quercetin-(galloyl-caffeoyl-hexoside)-rhamnoside, quercetin-(feruloyl-hexoside)-rhamnoside, and quercetin-(succinyl-rhamnoside)-rhamnoside, were reported for the first time. In another study, Morales-Soto et al. (2013) reported a total of 45 compounds tentatively identified mainly as glycoside derivatives of flavonoids such as quercetin 3,7-di-*O*-α-L-rhamnopyranoside and narigenin-7-*O*-β-D-(3″-*p*-coumaroyl)-glucopyranoside. Likewise, other polar metabolites such as organic acids, nucleosides, and amino acids were also detected. More recently, Guevara et al. (2021), reported the identification of compounds with potential therapeutic uses in fruits from sweet pepper (*C. annuum* L.) and their modulation by nitric oxide; 12 differential bioactive compounds were identified including quercetin and its derivatives, L-tryptophan, phytosphingosin, FAD, gingerglycolipid A, tetradydropentoxylin, blumenol C glucoside, colnelenic acid and capsoside A.

Fratianni et al. (2020), measured the amounts of bioactive compounds from different Campania native sweet pepper varieties. Polyphenols (caffeic, ferulic, *p*-coumaric, gallic, and chlorogenic acids, catechin, epicatechin, quercetin, rutin, naringenin, and apigenin) ranged between 1.37 nmol $g^{-1}$ and 3.42 mmol $g^{-1}$; β-carotene was abundant in a red variety, while yellow and red varieties showed a content of ascorbic acid not inferior to 0.82 mg $^{g-1}$, and in some varieties, the content of ascorbic acid was almost inconsistent. In the same manner, Ribes-Moya et al. (2020) reported the variation in flavonoids in a collection of chili peppers (*Capsicum* sp.) under organic and conventional cultivation practices. Their main results explained that luteolin and quercetin showed the highest contribution to the total phenolic content at both, unripe and fully ripe states evaluated, while myricetin, apigenin, and kaempferol showed lower contributions.

This study focused on exploring the metabolite diversity, and differences between domesticated and wild accessions of *Capsicum annuum* L. during fruit development. Furthermore, an F1 progeny between a domesticated and a wild accession was considered. This investigation showed a differential clustering and presence of metabolites related to terpenoid and fatty acid classifications in domesticated and wild accessions of chili pepper suggesting a possible impact of domestication.

# MATERIALS AND METHODS

## Chemicals and Reagents

All the chemicals and reagents were purchased from AccesoLab (AccesoLab S.A. de C.V., Mexico, Mexico). Formic acid, methanol, and acetonitrile were MS grade and purchased from Sigma-Aldrich (Merck S.A. de. CV., Mexico, Mexico).

## Plant Material and Fruit Harvesting

Seeds from each chili pepper accession (*C. annuum* L.; **Table 1**) were treated with 70% ethanol solution before treatment with 10% hypochlorite solution for 10 s, followed by six rinses with deionized water. Wild accession seeds were treated first with 50% of the sulfuric acid solution for scarification. All accession seeds were germinated using a mixture of peat moss, perlite, sludge, forest soil, and vermiculite in a plastic tray and incubated in a chamber maintained at 23–25°C. Three-week-old plants were individually transplanted to 5 L pots at greenhouse facilities in optimum conditions (30–32°C) and fertilized with Long Ashton solution every two weeks. Flowers were labeled and fruits from 10 different plants of each accession were collected from 0 to 60 days after anthesis (DAA) at intervals of 10 days during development. At least, 10 fruits and flowers were collected from each plant. All the collected fruits were immediately frozen in liquid nitrogen and stored at −80°C.

## Metabolite Extraction

The flowers and whole fruits were grounded manually in a mortar with a pestle using liquid nitrogen. Following this process, two

**TABLE 1 |** List of chili pepper (*C. annuum*) accessions used.

| Accession | Group | Key |
|---|---|---|
| *C. annuum* L. cv. Tampiqueño 74 | Domesticated | ST |
| *C. annuum* L. cv. Criollo de Morelos 334 | Domesticated | CM |
| *C. annuum* L. cv. California Wonder | Domesticated | CW |
| *C. annuum* L. cv. Zunla-1 | Domesticated | ZU |
| *C. annuum* L. var. *glabriusculum* (Querétaro) | Wild | QU |
| *C. annuum* L. var. *glabriusculum* (Coahuila) | Wild | CO |
| F1 *C. annuum* L. CM♂ X QU♀ | Cross | F1-CQ |
| F1 *C. annuum* L. QU ♂ X CM ♀ | Cross | F1-QC |

biological replicates were made for each developmental stage. A biological replicate comprised of at least two fruits harvested from 2 to 6 individual plants, depending upon the developmental stage and cultivar.

For the metabolite extraction, a modified Matyash method was employed as outlined previously (Matyash et al., 2008; Cervantes-Hernández et al., 2019): 1.5 ml of methanol was added to 100 mg of frozen sample in a test tube. After 1 min vortexed, 5 ml of diethyl ether were added. Samples were incubated at room temperature with gentle stirring for 1 h and then 1.5 ml of ultra-pure water (18 Ω, milli-Q system) were added and vigorously mixed for 1 min. The mixture was settled at room temperature for 10 min, and then centrifuged at 1,000 × *g* for 10 min. The organic phase was collected for this study. Samples were vacuum dried (miVac, Genevac) at 30°C for 30 min and stored at −80°C until further analysis.

## LC–MS Analysis

In total, five artificial quality control (QC) samples were prepared to verify the instrument drift and calibration along with the analysis in UHPLC-QTOF-HRMS. Each QC sample contained an equative portion of all the extracted samples in the experiment.

For UHPLC–MS analysis, all the samples were resuspended in 2 ml of acetonitrile/ultra-pure water 50:50 (v/v) and filtered using individual pre-packed filters (0.2 μm, PTFE, Agilent Technologies, Santa Clara, United States). Samples were injected by a randomized list order on a UPLC® (Acquity class I, Waters, Milford, CA, United States) coupled with an orthogonal QTOF mass spectrometer (SYNAPT G1 HDMS, Waters, Milford, CA, United States). Chromatographic separation was achieved on a reversed-phase CSH C18 column (2.1 mm × 150 nm, 1.7 μm, Waters, Milford, CA, United States) at 35°C during chromatographic separation. Mobile phase A was composed by ultra-pure water with formic acid (0.1%; v/v); Mobile phase B was acetonitrile with formic acid (0.1%, v/v) with a flow rate of 0.3 ml/min following a gradient method: from 0 to 0.5 min, 35% B; 0.5 to 10 min, 35–65% B; 10 to 30 min, 65–98% B; 30 to 31 min, 98% B; 31 to 31.5 min, 98–99% B; 31.5 to 33 min, 99% B; 33 to 33.1 min, 99–35% B; 33.1 to 36 min, 35% B. The mass spectrometer range was set from 50 to 1,500 Da. Both ionization modes were injected separately. For positive electrospray ionization (ESI) mode, the conditions were set as follows: capillary voltage 3 kV; cone voltage 40 V; source temperature 120°C; desolvation temperature 350°C; desolvation gas flow 500 L/h. For the negative ESI mode: capillary voltage 2.5 kV; cone voltage 40 V; source temperature 120°C; desolvation temperature 300°C; desolvation gas flow 500 L/h. Leucine–enkephalin (2 ng/ml) was infused as LockSpray reference internal mass calibrant at a flow rate of 5 μl/min and its signal was monitored every 10 s. Data were collected in a continuum mode with an MS scan time of 1.5 s. In both ionization modes, data were acquired in $MS^e$ experiments using Ar as the collision gas with collision energy in the trap region of 10 eV (Function 1, low energy) and ranging from 20–50 eV (Function 2, high voltage).

## Data Analysis

Positive and negative electrospray ionization data were independently analyzed using Progenesis QI for small molecules software (Non-Linear Dynamics, Waters, Milford, MA, United States). Alignment, normalization, and deconvolution were set at standard parameters. Pre-identification was performed using Chemspider Databases (PlantCyc, Plant Metabolic Network, KEGG, HMDB, and ChEBI) and with an in-house database with a minimum match of 80% for precursor ions, MS/MS data, retention times values, and isotope distribution were included for increasing match score values. Statistics and graphics were performed using R software (3.3.3v, Vienna, Austria). A *t*-test was used for significance analysis ($p$-value $< 0.05$). Clustering analysis was performed using the "pvclust" package in R, using Euclidean distance and Ward's linkage method for clustering. Values of bootstrap probability (bp) and approximately unbiased $p$-value (au) are parameters related to the grouping quality (Suzuki and Shimodaira, 2006). Compounds were classified according to their compound classes using the ClassyFire tool (Djoumbou Feunang et al., 2016). MetaboAnalyst (Pang et al., 2021) was used for principal component analysis, pathway analysis, and heatmap visualization. The data for principal component analysis (PCA) was first scaled to the Pareto method and submitted subsequently to PCA. The biosynthetic pathway analysis was performed using the *Arabidopsis thaliana* database of KEGG and pre-annotated compound names were used for the query. A heatmap was elaborated using normalized values of all the pre-annotated compounds during the development. A complete heatmap of the dataset is shown in **Supplementary Figure 1**. OPLS–DA and S-plots were generated with scaling data to Pareto using Ezinfo 3.0.3.0 software (Umetrics).

# RESULTS

## Metabolite Pre-identifications During Fruit Development of *C. annuum*

In total, 122 samples of pepper fruit extract from different accessions at different developmental stages were analyzed in both the positive and negative ionization modes. Chromatographic data were processed, aligned, and normalized using Progenesis QI for small molecules. A total of 456 and 292 features or m/z values were detected in positive and negative modes, respectively. **Supplementary Figure 2** shows some

examples of pre-identified metabolite distribution in QC samples per ionization mode.

According to the current levels of confidence for metabolite annotation (Blaženović et al., 2018; Reisdorph et al., 2019), around 160 putative compounds level 2 were annotated with a match score of over 80% which considered the exact mass of precursor ions m/z, precursor ion isotopic pattern, and fragmentation spectra matching. A complete dataset of putative compounds and unknown features is presented in **Supplementary Table 1**, containing metabolite ID, description, compound class, acquisition mode, and averaged abundance per developmental stage.

## Pathways Analysis Highlights Glycerophospholipids Metabolism During the Development of Chili Pepper Fruits

Using all 158 pre-annotated compound identifications (**Figure 1**), a biosynthesis pathway analysis was performed. In total, 19 pathway classifications were found in the dataset, where glycerophospholipids metabolism was significantly represented among the other categories according to the impact and *p*-value (**Figure 1**). Compounds in this category were PC [18:3(9Z,12Z,15Z)/18:3(6Z,9Z,12Z)], PE [18:2 (9Z, 12Z)/16:0]; LysoPC (15:0), and LysoPA (0:0/16:0). Other metabolic pathways, such as flavone and flavonol biosynthesis, linoleic acid metabolism, α-linoleic acid metabolism, sphingolipid metabolism, glycerolipids metabolism, and cutin, suberin, and wax biosynthesis were also highlighted based on the impact value, but they did not pass the standard significant threshold ($p$-value $< 0.05$). Arachidonic acid metabolism, glycosylphosphatidylinositol (GPI)-anchor biosynthesis, glycerolipid metabolism, biosynthesis of unsaturated fatty acids, fatty acid elongation, phosphatidylinositol signaling system, fatty acid degradation, ubiquinone, and another terpenoid–quinone biosynthesis, phenylpropanoid biosynthesis, porphyrin, and chlorophyll metabolism, fatty acid biosynthesis showed only one compound per category, that resulted in a low impact and non-significant *p*-value for the analysis.

## Clustering Analysis Segregates Chili Pepper Fruits Depending on Their Domestication Category

Bootstrapped hierarchical analysis of metabolic features during the whole development of chili pepper fruits showed clustering of the different accessions based on their domestication level. **Figure 2** shows the values of bootstrap probability (green) and approximately unbiased *p*-values (red) in each bifurcation. Domesticated cultivars clustered exclusively with other domesticated cultivars. The other big clade was constituted by the wild accessions and the F1 progenies from crosses, but even these exhibited a separation based on their similarities. F1-QC and F1-CQ were the most distant groups from the full dataset of the experiment.

Predominantly based on the domestication level, a 2D principal component analysis (**Figure 3**) shows consistent clustering of domestication level along the developmental stages, which is also coherent with the clustering analysis presented in **Figure 2**. Only metabolomic profiles from 00 DAA were more dispersed from other stages of development, suggesting that flower metabolites were different compared with the fruit for each accession. A total of 42.2% of the variance was explained using the two components.

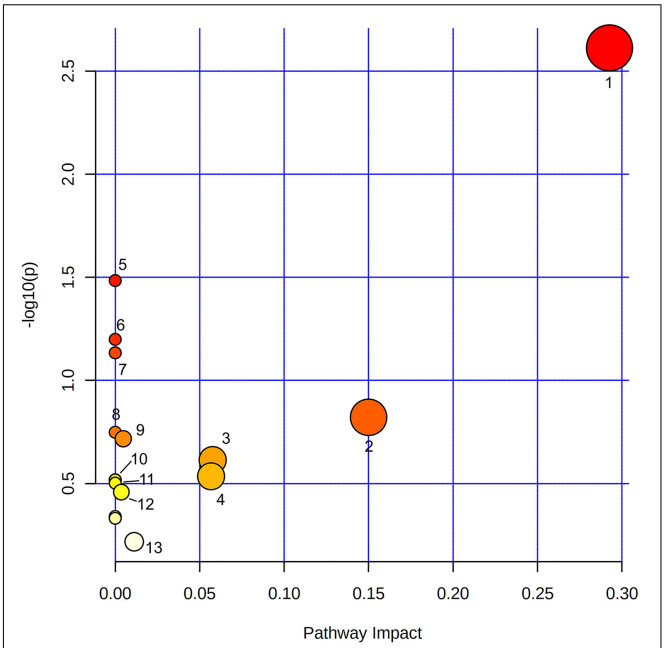



**FIGURE 1 |** Pathway analysis of pre-annotated compounds in fruits at different developmental stages from domesticated and wild accessions of chili pepper using LC–MS. 1. Glycerophospholipid metabolism; 2. Flavone and flavonol biosynthesis; 3. Sphingolipid metabolism; 4. Glycerolipid metabolism; 5. Cutin, suberin, and wax biosynthesis; 6. Linoleic acid metabolism; 7. α-Linoleic acid metabolism; 8. Arachidonic acid metabolism; 9. Glycosylphosphatidylinositol (GPI)-anchor biosynthesis; 10. Biosynthesis of unsaturated fatty acid; 11. Fatty acid elongation; 12. Phosphatidylinositol signaling system; and 13. Fatty acid biosynthesis. Color represents –log10(*p*) value, size correlates with pathway impact value.

## Significant Metabolite Differences During Fruit Development of Wild and Domesticated *C. annuum*

**Figure 4** displays the heatmap of the 60 top discriminant pre-annotated compounds with abundance scaled and grouped by accession. Dendrograms at the top and left side of **Figure 4** clustered the accessions and features, respectively. The accession dendrogram was consistent with the whole dataset dendrogram. It presented two big clades: one was exclusive for domesticated accessions and the other clade was constituted by the wild and F1 cross accessions.

The upper clade of the pre-annotated compound dendrogram represents the more representative abundance in domesticated *C. annuum* fruits. It was composed of jasmonic acid, capsorubin, 12′-apo-β-carotenal, β-tocopherol, murolic acid, LPA (0:0/16:0),

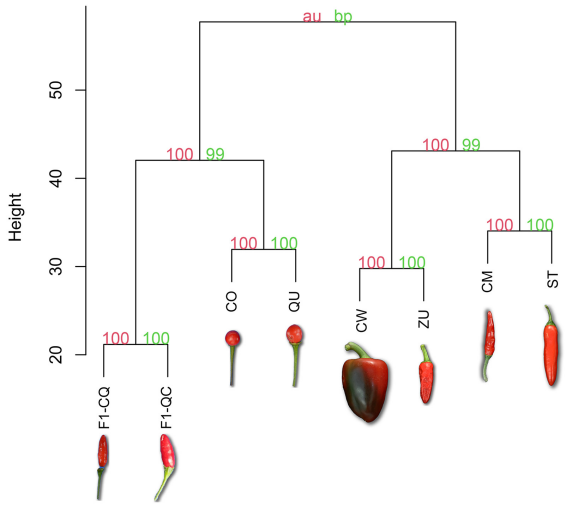



**FIGURE 2 |** Hierarchical clustering of mean values from 748 features during development of wild, domesticated, and cross F1 accession of *C. annuum*. Bootstrap probability (bp) appears in green; approximately unbiased *p*-value (au) appears in red. The orientation of fruits in the photo represents if they are erect or pendent in the plant. The distance was measured using Euclidean and Ward's linkage method for clustering.

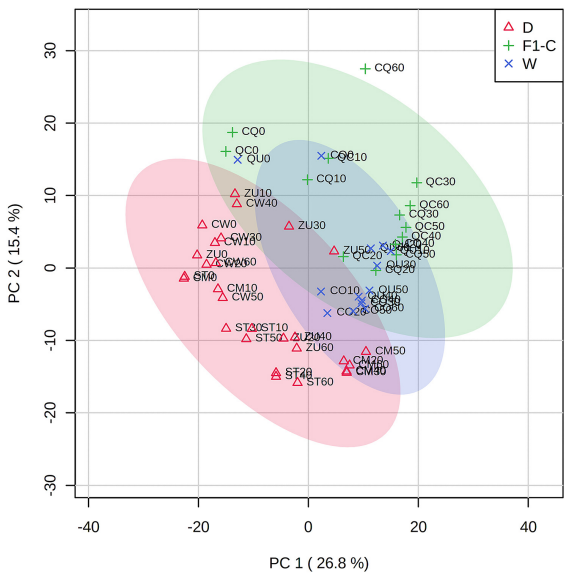



**FIGURE 3 |** Principal component analysis of fruits at different developmental stages from domesticated or wild accessions of *C. annuum*. Component 1 = 26.8%; component 2 = 15.4%. Color meaning is indicated in the legend. The developmental stage is indicated as two digits after the accession key, for example, ST0 for Tampiqueño74 stage 0 (flowers); ST20 Tampiqueño74 at 20 DAA.

4-hydroxy-6-methyl-3-(1-oxobutyl)-2H-pyran-2-one, dihydro-5-(2-octenyl)-2(3H)-furanone, parsonsine, C20 sphinganine 1-phosphate, C20 phytosphingosine, and thiamine, among others.

Capsaicin and some other capsaicinoids were more abundant in wild and F1 cross accessions (bottom clade). In addition, we found sphinganine, quercitrin, castasterone 23-*O*-α-D-glucoside, palmitic acid and derivates, N-(icosanoyl) ethanolamine, tuberoside, sphinganine 1-phosphate, prenyl arabinosyl-(1- > 6)-glucoside, soyacerebroside I, glycol stearate, and hexadecasphinganine as richer compounds in these groups.

In the F1 progenies cluster, (18R)-18-hydroxynonadecanoic acid, gypsogenin 3-*O*-β-D-glucuronide, pollinastanol, theasapogenol E were more abundant in contrast with the other two groups, domesticated and wild accession.

S-plots (**Figure 5**) were used to pair-comparison between the different accessions based on their domestication category. This plot is a powerful visualization method for comparing two conditions. S-plot considered, for the *y*-axis, the values of the correlation coefficient of the first component and the concentration of any feature; the *x*-axis is referred to the covariance of the first principal component and concentration of each feature (Pan et al., 2021). The red dots in **Figure 5** represent the significant features ($p$-value $< 0.05$ and correlation $> 60\%$) in each group. 12′-apo-β-carotenal, and 4-hydroxy-6-(2-oxononadecyl) pyran-2-one were significantly pre-annotated compounds in the domesticated group. Besides, the wild group presented significance only in features unfortunately no pre-identification. F1 cross accessions exhibited capsaicin as a significant compound compared with the domesticated cultivars but not against the wild group.

## DISCUSSION

Metabolic global profiles of chili pepper fruit (*C. annuum*) have been reported previously by different authors (Wahyuni et al., 2011, 2013b; Fratianni et al., 2020), including tissue-specific (Cervantes-Hernández et al., 2019) and fruits at different developmental stages (Jang et al., 2015). However, to the best of our knowledge, this is the first study that aimed at exploring comparatively the metabolomic profiles during fruit development in domesticated and wild accessions of chili pepper fruits.

Hierarchical cluster analysis (**Figure 2**) of features averaged intensities across the seven developmental stages showed clustering of the accessions based on their domestication level. The high values of bp and au demonstrated high reliability of the grouping clustering (Suzuki and Shimodaira, 2006). Domesticated accessions formed a big clade, suggesting alterations in the phytochemical profiles probably because of the domestication process and selection of different traits. The other main clade was outlined by the wild and F1 cross-group, which phenotypically was more similar than that of the domesticated accessions. Also, F1-QC and F1-CQ showed segregation from the two wild accessions in the superior clade.

The separation of the first filial progeny could suggest a potential heterosis event. Compounds related to the secondary metabolism, specifically capsaicinoids, were found more abundant in F1 crosses. This shift in plant defense mechanism could represent an advantage over the progenitors in the wild. Also, gypsogenin 3-*O*-β-D-glucuronide and theasapogenol E, are two saponins that were more abundant in the F1 group.

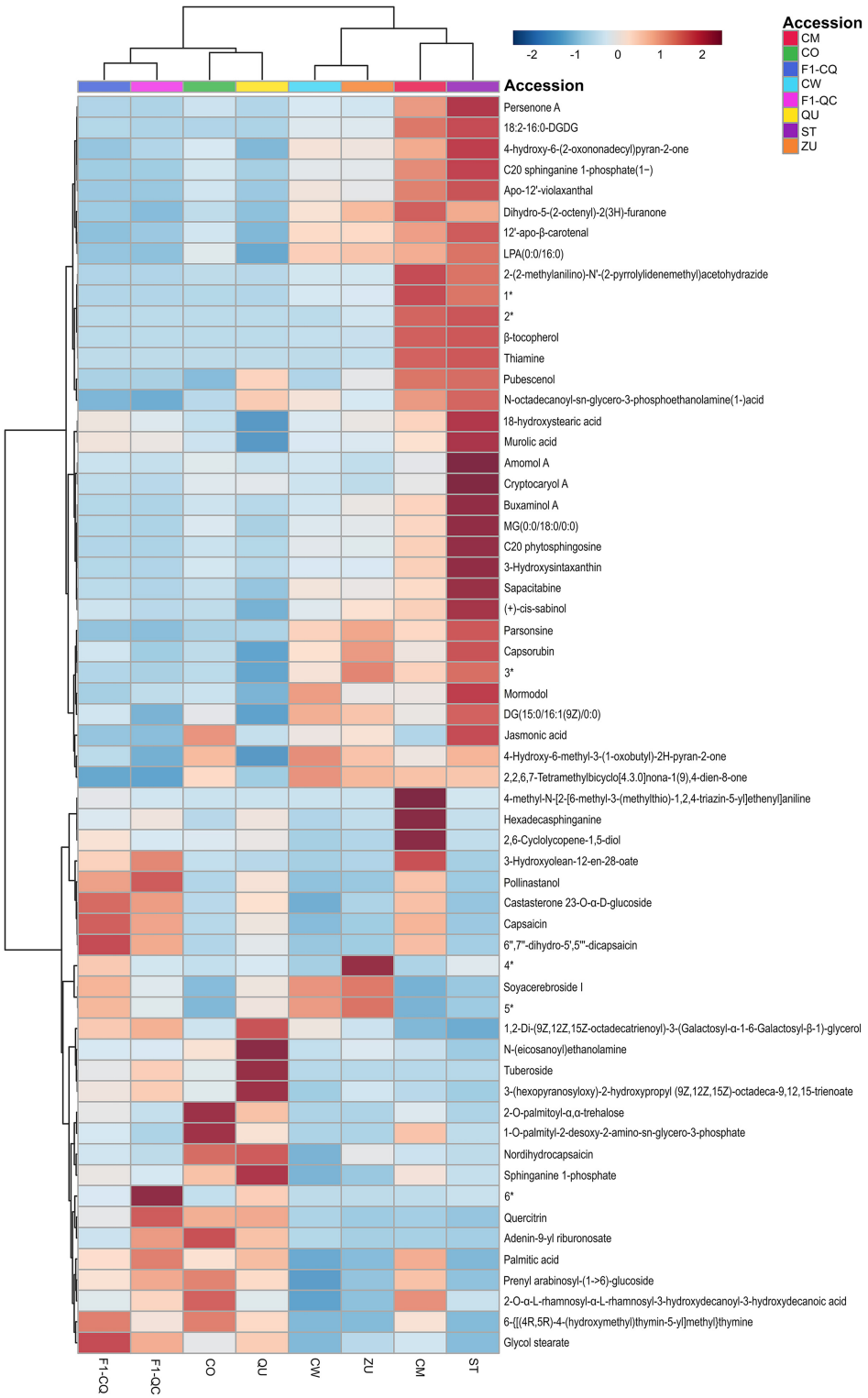



**FIGURE 4 |** Heatmap of pre-annotated compounds in *Capsicum annuum* during fruit development. 1* Glycerol 2-(9Z,12Z-octadecadienoate)1-hexadecanoate3-*O*-[α-D-galactopyranosyl-(1- > 6)-β-D-galactopyranoside]; 2* 2,5-epoxy-2β-hydroxy-8α-(2-methylbut-2-enoyloxy)-4(15),10(14),11(13)-germacratrien-12,6α-olide; 3* N-(4-{[(2-amino-4-oxo-1,4,5,6,7,8-hexahydro-6-pteridinyl)methyl]amino}benzoyl)-γ-glutamyl-γ-glutamylglutamic acid; 4* 3-{[(2,3-dihydroxypropoxy)(hydroxy)phosphoryl]oxy}-2-(palmitoyloxy) propyl (9E,12E)-9,12-octadecadienoate; 5* (2S)-3-(β-D-galactopyranosyloxy)-2-[(9Z,12Z)-9,12-octadecadienoyloxy]propyl (9Z,12Z,15Z)-9,12,15-octadecatrienoate; 6* 2,6,6-trimethyl-4-[[3,4,5-trihydroxy-6-(hydroxymethyl)-2-oxanyl]oxy]-1-cyclohexenecarboxaldehyde. Euclidean distance and Ward's linkage method was used for clustering.

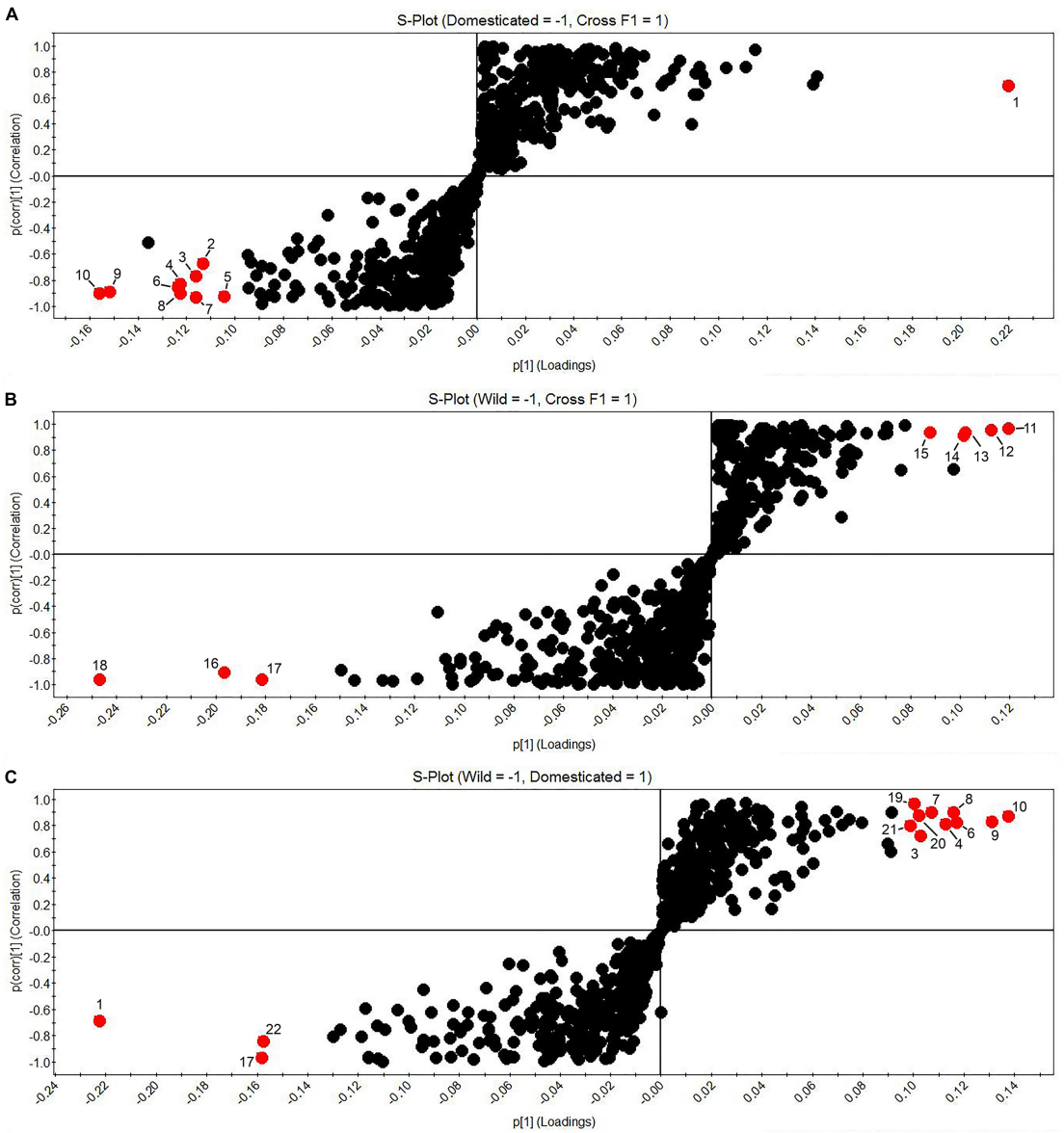



**FIGURE 5 |** S-plots. **(A)** Domesticated (left) vs. F1 progenies (right); **(B)** Wild (left) vs. F1 progenies (right); **(C)** Wild (left) vs. Domesticated (right). 1. UNK19.30_269.2689n; 2. UNK27.39_311.1710m/z; 3. 4-hydroxy-6-(2-oxononadecyl)pyran-2-one; 4. UNK26.15_337.3352n; 5. UNK24.68_309.3061n; 6. UNK26.98_395.3129m/z; 7. UNK23.37_283.2895n; 8. UNK27.29_363.3516n; 9. 12′-apo-β-carotenal, 10. dihydro-5-(2-octenyl)-2(3H)-furanone; 11. UNK33.34_638.4883n; 12. UNK23.32_414.3364n; 13. Capsaicin; 14. UNK2.77_583.2669n, 15. UNK2.90_612.2727m/z; 16. UNK33.96_859.6947n; 17. UNK1.41_678.5108n; 18. UNK30.56_774.5371n; 19. UNK27.83_311.3200n; 20. UNK19.55_405.2257m/z; 21. UNK15.76_299.2018m/z; and 22. UNK7.59_137.0603m/z.

It is also known that saponin compounds have a role in plant defense. Other authors have demonstrated the behavior of the genetic heritability in the F1 progeny using different plant species. Superior traits such as fruit yield per plant, fruit number, and fewer days for fruit ripening have been observed in the F1 progeny between wild and domesticated accessions (Geleta et al., 2004; Marame et al., 2009; Rodrigues et al., 2012).

Our results, using a biosynthetic pathway analysis, demonstrated a richer abundance of lipids and hydrophobic compounds in all the analyzed samples. Lipids in plants, as in other organisms, are essential for cell membrane generation, a source of energy, and act as signal molecules (Janda et al., 2013; Aid, 2019). Lipids also have a role as a precursor of important metabolic pathways. Fatty acids are precursors in the biosynthesis of secondary metabolites, such as isoprenoids and capsaicinoids (Arce-Rodríguez and Ochoa-Alejo, 2019; Schmid, 2021). Glycerophospholipids metabolism was the category with the highest impact and lower *p*-value among the most representative metabolic pathways. Our domesticated group

exhibited a significant abundance of these compound classes compared with the wild and cross F1 groups. This could probably be related to the modifications in the cell membrane structure of fruit pericarp. In *Capsicum annuum* L. cv. Bell pepper, Sutliff et al. (2021) reported the presence of 57 glycerophospholipids and glycerolipids compounds in the mature fruits. The selection of fruit color traits could be associated with the modification of nutritional characteristics in bell pepper (Sutliff et al., 2021). Some authors reported that this compound group is involved in different biological processes, such as plant defense, adaptation, and fruit water loss resistance, and probably resulted as an effect of the domestication process (Parsons et al., 2013; Villa-Ruano et al., 2018).

Previous studies have shown that fatty acids were possibly affected as a selected trait during the domestication process. Cuticle lipid composition seems to be unknowingly modified by this artificial selection by breeders with the intention of improving the appearance of the fruit and reducing fruit water loss (Parsons et al., 2013). Dormancy of fruit seeds is a trait of interest for producers, and they may have altered the profile of fatty acids. It has been reported that seeds of the domesticated chili pepper do not exhibit dormancy (Luna-Ruiz et al., 2018). Besides, wild chili pepper seeds have thick lignified test as to preserve water during drying (Luna-Ruiz et al., 2018). In addition, in different species of legumes, it seems that the domestication process has led to a decrease in the content of linoleic acid, another common lipid in plant tissues (Fernández-Marín et al., 2014). Guevara et al. (2021) reported the presence of higher levels of phytosphingosine, mainly in red/mature, than those in the green chili pepper fruit, and their implications in the ripening process.

Flavone and flavonol compounds belong to flavonoids, a large group of molecules in plants (Howard et al., 2000; Wahyuni et al., 2013a). Flavones present a 2-phenylchromen-4-one backbone and, flavonols bear a hydroxy group at the 3-position (Wahyuni et al., 2013a). Here, we pre-annotated some of these compound classes: such as quercitrin, neoflavan, and Kaempferol 3-xylosylglucoside. Different authors have reported this pathway as one of the most affected during the domestication and selection of chili pepper (Luna-Ruiz et al., 2018). It has been reported that almost all the flavonoid levels vary among cultivars and the fruit maturation stage (Morales-Soto et al., 2013). Wahyuni et al. (2014) reported an epistatic effect on semi-polar metabolites, including flavonol *O*-glycosides, in F2 segregation of crossing *C. annuum* AC1979 (female) and *C. chinense* (male) (Wahyuni et al., 2014).

There are several published reports that focus on pungent compounds and their related genes in chili pepper fruit. Without hesitation, this is one of the most attractive properties of the fruit. Therefore, the domestication process has had a relevant repercussion on the selection of traits that increase or decrease the content of capsaicinoids (Luna-Ruiz et al., 2018). In the wild, these metabolites function as a repellent against seed predators, such as small mammalians.

In the S-plots (**Figure 5**), 12′-apo-β-carotenal, a hydrophobic terpene of the group of apocarotenoids, is shown to be significant in the domesticated group. Zoccali et al. (2021), published the profiles of apocarotenoids in different accessions of *Capsicum*. They identified 19 free apocarotenoids and 8 apocarotenoids fatty acid esters, including 12′-apo-β-carotenal. Terpenoids in plants have different defensive functions as cytotoxic, phototoxic, and phytoalexins against microbials and insects (Cheng et al., 2007; Bakkali et al., 2008; Pandey et al., 2017). They are also attributed to attracting pollinators and seed dispersers (Langenheim, 1994). However, little is known about the terpenoids involvement in affectation during the domestication process in fruits.

Different authors agree that the domestication process has driven a decrease in metabolic diversity, mainly related to adaptation and pathogen resistance. Meyer et al. (2012), published a review about more than 200 different crops where wild progenitors showed a bigger abundancy in specialized metabolites, mainly related to flavor, pigments, and toxicity (Meyer et al., 2012).

The domestication of *C. annuum* plants changed the form, size, pungency, and color of the fruit based on the specific desired traits. Unfortunately, this artificial selection has also caused modifications in non-specific metabolism. While our results showed an increase in the abundance of lipid metabolism in domesticated cultivars, they also demonstrated a decrease of capsaicinoids, terpenoids, flavones, and flavonol compounds when compared with their wild ancestors. The resulting F1 progenies between the wild (QU) and the domesticated (CM) chili peppers (crosses in both directions) displayed a possible case of heterosis in genes related to the defensive strategies in the plants. It is our view that increasing the metabolomic analysis in a bigger number of accessions and cultivars will improve the resolution of the present results. Furthermore, the generation and comparison of additional metabolomic data, combined with molecular data, during the fruit development, could increase our understanding of the effects that occur due to the domestication process. This would enable the development of new genetic strategies to bring back metabolic diversity.

## DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

FC-H performed the experiments and wrote the manuscript. JO-O and OM designed and supervised the experiment. FC-H and OM analyzed the data. FC-H, JO-O, OM, and NO-A interpreted the results. JO-O, OM, and NO-A revised and edited the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was funded by the Conacyt's (Consejo Nacional de Ciencia y Tecnología) projects FC1570 and FOINS1776, as well as scholarship assigned to CVU 707294.

## ACKNOWLEDGMENTS

We like to thank Fernando Hernández-Godinez for the seed, grown, and collected plant material used in this experiment.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.893055/full#supplementary-material

## REFERENCES

Aid, F. (2019). "Plant lipid metabolism," in *Advances in Lipid Metabolism*, eds R. V. Baez (London: IntechOpen). 1–17 doi: 10.5772/intechopen.81355

Arce-Rodríguez, M. L., and Ochoa-Alejo, N. (2019). Biochemistry and molecular biology of capsaicinoid biosynthesis: recent advances and perspectives. *Plant Cell Rep.* 38, 1017–1030. doi: 10.1007/s00299-019-02406-0

Bakkali, F., Averbeck, S., Averbeck, D., and Idaomar, M. (2008). Biological effects of essential oils - A review. *Food Chem. Toxicol.* 46, 446–475. doi: 10.1016/J.FCT.2007.09.106

Blaženović, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* 8:31. doi: 10.3390/metabo8020031

Cervantes-Hernández, F., Alcalá-González, P., Martínez, O., and Ordaz-Ortiz, J. J. (2019). Placenta, pericarp, and seeds of tabasco chili pepper fruits show a contrasting diversity of bioactive metabolites. *Metabolites* 9:206. doi: 10.3390/metabo9100206

Cheng, A. X., Lou, Y. G., Mao, Y. B., Lu, S., Wang, L. J., and Chen, X. Y. (2007). Plant terpenoids: biosynthesis and ecological functions. *J. Integr. Plant Biol.* 49, 179–186. doi: 10.1111/J.1744-7909.2007.00395.X

Clement, C. R., de Cristo-Araújo, M., D'Eeckenbrugge, G. C., Pereira, A. A., and Picanço-Rodrigues, D. (2010). Origin and domestication of native Amazonian crops. *Diversity* 2, 72–106. doi: 10.3390/d2010072

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* 8, 1–20. doi: 10.1186/s13321-016-0174-y

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006

Fernández-Marín, B., Milla, R., Martín-Robles, N., Arc, E., Kranner, I., Becerril, J. M., et al. (2014). Side-effects of domestication: cultivated legume seeds contain similar tocopherols and fatty acids but less carotenoids than their wild counterparts. *BMC Plant Biol.* 14, 1–11. doi: 10.1186/s12870-014-0385-1

Fratianni, F., D'acierno, A., Cozzolino, A., Spigno, P., Riccardi, R., Raimo, F., et al. (2020). Biochemical characterization of traditional varieties of sweet pepper (*Capsicum annuum* L.) of the Campania region, southern italy. *Antioxidants* 9, 1–16. doi: 10.3390/ANTIOX9060556

Geleta, L. F., Labuschagne, M. T., and Viljoen, C. D. (2004). Relationship between heterosis and genetic distance based on morphological traits and AFLP markers in pepper. *Plant Breed.* 123, 467–473. doi: 10.1111/J.1439-0523.2004.01017.X

Guevara, L., Domínguez-Anaya, M. Á, Ortigosa, A., González-Gordo, S., Díaz, C., Vicente, F., et al. (2021). Identification of compounds with potential therapeutic uses from sweet pepper (*Capsicum annuum* L.) fruits and their modulation by nitric oxide (NO). *Int. J. Mol. Sci.* 22:4476. doi: 10.3390/ijms22094476

Hayano-Kanashiro, C., Gámez-Meza, N., and Medina-Juárez, L. Á (2016). Wild pepper L. var. : taxonomy, plant morphology, distribution, genetic diversity, genome sequencing, and phytochemical compounds. *Crop Sci.* 56:1. doi: 10.2135/cropsci2014.11.0789

Hernández-Verdugo, S., Luna-Reyes, R., and Oyama, K. (2001). Genetic structure and differentiation of wild and domesticated populations of *Capsicum annuum* (Solanaceae) from Mexico. *Plant Syst. Evol.* 226, 129–142. doi: 10.1007/s006060170061

Howard, L. R., Talcott, S. T., Brenes, C. H., and Villalon, B. (2000). Changes in phytochemical and antioxidant activity of selected pepper cultivars (*Capsicum* species) as influenced by maturity. *J. Agric. Food Chem.* 48, 1713–1720. doi: 10.1021/jf990916t

Janda, M., Planchais, S., Djafi, N., Martinec, J., Burketova, L., Valentova, O., et al. (2013). Phosphoglycerolipids are master players in plant hormone signal transduction. *Plant Cell Rep.* 32, 839–851. doi: 10.1007/s00299-013-1399-0

Jang, Y. K., Jung, E. S., Lee, H. A., Choi, D., and Lee, C. H. (2015). Metabolomic characterization of hot pepper (*Capsicum annuum* "CM334") during fruit development. *J. Agric. Food Chem.* 63, 9452–9460. doi: 10.1021/acs.jafc.5b03873

Langenheim, J. H. (1994). Higher plant terpenoids: a phytocentric overview of their ecological roles. *J. Chem. Ecol.* 20, 1223–1280. doi: 10.1007/BF02059809

Luna-Ruiz, J., de, J., Nabhan, G. P., and Aguilar-Meléndez, A. (2018). Shifts in plant chemical defenses of chile pepper (*Capsicum annuum* L.) due to domestication in Mesoamerica. *Front. Ecol. Evol.* 6:48. doi: 10.3389/fevo.2018.00048

Marame, F., Dessalegne, L., Fininsa, C., and Sigvald, R. (2009). Heterosis and heritability in crosses among Asian and Ethiopian parents of hot pepper genotypes. *Euphytica* 168, 235–247. doi: 10.1007/S10681-009-9912-9

Martínez-Ispizua, E., Martínez-Cuenca, M. R., Marsal, J. I., Díez, M. J., Soler, S., Valcárcel, J. V., et al. (2021). Bioactive compounds and antioxidant capacity of Valencian pepper landraces. *Molecules* 26, 1–23. doi: 10.3390/molecules26041031

Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A., and Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* 49, 1137–1146. doi: 10.1194/jlr.D700041-JLR200

Meyer, R. S., Duval, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

Morales-Soto, A., Gómez-Caravaca, A. M., García-Salas, P., Segura-Carretero, A., and Fernández-Gutiérrez, A. (2013). High-performance liquid chromatography coupled to diode array and electrospray time-of-flight mass spectrometry detectors for a comprehensive characterization of phenolic and other polar compounds in three pepper (*Capsicum annuum* L.) samples. *Food Res. Int.* 51, 977–984. doi: 10.1016/j.foodres.2013.02.022

Pan, Y.-Y., Chen, Y.-C., Chang, W. C.-W., Ma, M.-C., and Liao, P.-C. (2021). Visualization of statistically processed LC-MS-based metabolomics data for identifying significant features in a multiple-group comparison. *Chemom. Intell. Lab. Syst.* 210:104271. doi: 10.1016/j.chemolab.2021.104271

Pandey, A. K., Kumar, P., Singh, P., Tripathi, N. N., and Bajpai, V. K. (2017). Essential oils: sources of antimicrobials and food preservatives. *Front. Microbiol.* 7, 1–14. doi: 10.3389/fmicb.2016.02161

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., et al. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 49, W388–W396. doi: 10.1093/NAR/GKAB382

Paran, I., and van der Knaap, E. (2007). Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J. Exp. Bot.* 58, 3841–3852. doi: 10.1093/jxb/erm257

Parsons, E. P., Popopvsky, S., Lohrey, G. T., Alkalai-Tuvia, S., Perzelan, Y., Bosland, P., et al. (2013). Fruit cuticle lipid composition and water loss in a diverse collection of pepper (*Capsicum*). *Physiol. Plant.* 149, 160–174. doi: 10.1111/ppl.12035

Pascale, R., Acquavia, M. A., Cataldi, T. R. I., Onzo, A., Coviello, D., Bufo, S. A., et al. (2020). Profiling of quercetin glycosides and acyl glycosides in sun-dried peperoni di Senise peppers (*Capsicum annuum L.*) by a combination of LC-ESI(-)-MS/MS and polarity prediction in reversed-phase separations. *Anal. Bioanal. Chem.* 412, 3005–3015. doi: 10.1007/s00216-020-02547-2.

Pegard, A., Brizzard, G., Fazari, A., Soucaze, O., Abad, P., and Djian-Caporalino, C. (2005). Histological characterization of resistance to different root-knot nematode species related to phenolics accumulation in *Capsicum annuum*. *Phytopathology* 95, 158–165. doi: 10.1094/PHYTO-95-0158

Perry, L., Dickau, R., Zarrillo, S., Holst, I., Pearsall, D. M., Piperno, D. R., et al. (2007). Starch fossils and the domestication and dispersal of chili peppers

(*Capsicum* spp. L.) *in the Americas*. *Science* 315, 986–988. doi: 10.1126/science.1136914

Pickersgill, B. (1997). Genetic resources and breeding of *Capsicum* spp. *Euphytica* 96, 129–133. doi: 10.1023/A:1002913228101

Pickersgill, B. (2007). Domestication of plants in the Americas: insights from Mendelian and molecular genetics. *Ann. Bot.* 100, 925–940. doi: 10.1093/aob/mcm193

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Reisdorph, N. A., Walmsley, S., and Reisdorph, R. (2019). A perspective and framework for developing sample type specific databases for LC/MS-based clinical metabolomics. *Metabolites* 10:8. doi: 10.3390/metabo10010008

Ribes-Moya, A. M., Adalid, A. M., Raigón, M. D., Hellín, P., Fita, A., and Rodríguez-Burruezo, A. (2020). Variation in flavonoids in a collection of peppers (*Capsicum* sp.) under organic and conventional cultivation: effect of the genotype, ripening stage, and growing system. *J. Sci. Food Agric.* 100, 2208–2223. doi: 10.1002/jsfa.10245

Rodrigues, R., Gonçalves, L. S. A., dos Bento, C. S., Sudré, C. P., Robaina, R. R., and do Amaral, A. T. (2012). Capacidade combinatória e heterose para características agronômicas em pimenta. *Hortic. Bras.* 30, 226–233. doi: 10.1590/S0102-05362012000200008

Schmid, K. M. (2021). "Lipid Metabolism in Plants," in *in Biochemistry of Lipids, Lipoproteins and Membranes*, (Netherlands: Elsevier), 121–159. doi: 10.1016/B978-0-12-824048-9.00011-0

Solomon, A. M., Kim, T.-G., Han, K., Lee, H.-Y., Patil, A., Siddique, M. I., et al. (2021). Fine mapping and candidate gene identification for the *CapUp* locus controlling fruit orientation in pepper (*Capsicum* spp.). *Front. Plant Sci.* 12:1293. doi: 10.3389/fpls.2021.675474

Sutliff, A. K., Saint-Cyr, M., Hendricks, A. E., Chen, S. S., Doenges, K. A., Quinn, K., et al. (2021). Lipidomics-based comparison of molecular compositions of green, yellow, and red bell peppers. *Metabolites* 11, 1–16. doi: 10.3390/metabo11040241

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22, 1540–1542. doi: 10.1093/bioinformatics/btl117

Villa-Ruano, N., Velásquez-Valle, R., Zepeda-Vallejo, L. G., Pérez-Hernández, N., Velázquez-Ponce, M., Arcos-Adame, V. M., et al. (2018). 1H NMR-based metabolomic profiling for identification of metabolites in *Capsicum annuum* cv. mirasol infected by beet mild curly top virus (BMCTV). *Food Res. Int.* 106, 870–877. doi: 10.1016/j.foodres.2018.01.065

Wahyuni, Y., Ballester, A.-R., Sudarmonowati, E., Bino, R. J., and Bovy, A. G. (2011). Metabolite biodiversity in pepper (*Capsicum*) fruits of thirty-two diverse accessions: variation in health-related compounds and implications for breeding. *Phytochemistry* 72, 1358–1370. doi: 10.1016/j.phytochem.2011.03.016

Wahyuni, Y., Ballester, A.-R., Sudarmonowati, E., Bino, R. J., and Bovy, A. G. (2013a). Secondary metabolites of *Capsicum* species and their importance in the human diet. *J. Nat. Prod.* 76, 783–793. doi: 10.1021/np300898z

Wahyuni, Y., Ballester, A. R., Tikunov, Y., de Vos, R. C. H., Pelgrom, K. T. B., Maharijaya, A., et al. (2013b). Metabolomics and molecular marker analysis to explore pepper (*Capsicum* sp.) biodiversity. *Metabolomics* 9, 130–144. doi: 10.1007/s11306-012-0432-6

Wahyuni, Y., Stahl-Hermes, V., Ballester, A. R., de Vos, R. C. H., Voorrips, R. E., Maharijaya, A., et al. (2014). Genetic mapping of semi-polar metabolites in pepper fruits (*Capsicum* sp.): owards unravelling the molecular regulation of flavonoid quantitative trait loci. *Mol. Breed.* 33, 503–518.

Zoccali, M., Giuffrida, D., Salafia, F., Rigano, F., Dugo, P., Casale, M., et al. (2021). Apocarotenoids profiling in different *Capsicum* species. *Food Chem.* 334:127595. doi: 10.1016/j.foodchem.2020.127595

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/fpls.2022.900143

# Combining Proteomics and Metabolomics to Analyze the Effects of Spaceflight on Rice Progeny

*Deyong Zeng[1,2,3], Jie Cui[1,2], Yishu Yin[1,2], Cuihong Dai[1,2], Haitian Zhao[1,3], Chen Song[1,2], Shuanghong Guan[1,2], Dayou Cheng[1,2], Yeqing Sun[4] and Weihong Lu[1,2,3]**

[1] Department of Food Science and Engineering, School of Chemistry and Chemical Engineering, Harbin Institute of Technology, Harbin, China, [2] National and Local Joint Engineering Laboratory for Synthesis, Transformation and Separation of Extreme Environmental Nutrients, Harbin Institute of Technology, Harbin, China, [3] The Intelligent Equipment Research Center for the Exploitation of Characteristic Food & Medicine Resources, Chongqing Research Institute, Harbin Institute of Technology, Chongqing, China, [4] Institute of Environmental Systems Biology, Dalian Maritime University, Dalian, China

Spaceflight is a special abiotic stress, the biological effect mechanism of which on contemporary rice has been clarified, However, its effect on offspring rice was still unclear. In order to understand the response mechanism of F2 generation plants to space flight, this study used SJ-10 recoverable satellite to carry DN423 rice seeds for 12.5 days in orbit flight. After returning to the ground, the plants were then planted to F2 generation to explore the biological effect mechanism. Our research showed that in the F2 generation of TLS, the rice plant height of the space flight group increased by 33.8%, the ear length and thousand-grain weight decreased by 9.7 and 4.6%, respectively, and the grain number per panicle increased by 6.5%. Moreover, related proteins that control changes in agronomic traits have been identified. The changes of MDA, $H_2O_2$, soluble sugar, electron leakage and antioxidant enzyme activity confirmed the stress response in F2 generation plants. ITRAQ and LC-MS technology were used to reveal the change pattern of protein levels and metabolite levels in F2 generation plants, 389 and 405 proteins were identified as differentially abundant proteins in TLS and TS, respectively. In addition, there were 124 and 125 metabolites that changed during these two periods. The proteome and metabolome result further confirmed that the F2 generation plants still retained the memory of space flight stress, and retained the memory of space flight stress through genome instability. Oxidative stress signals activated sugar signals to rebuild metabolic networks to adapt to space flight stress. The reconstruction of energy metabolism, amino acid metabolism, phenylalanine metabolism, and flavonoid metabolism played an important role in the process of adapting to space flight stress. The results of this study broaden the perspective of space biological effects and provide a basis for studying the effects of abiotic stress on plant progeny.

**Keywords: rice, iTRAQ, metabolomics, space flight, SJ-10 returning satellite**

OPEN ACCESS

***Edited by:***
*Mirza Hasanuzzaman, Sher-e-Bangla Agricultural University, Bangladesh*

***Reviewed by:***
*Syarul Nataqain Baharum, National University of Malaysia, Malaysia*
*Klára Kosová, Crop Research Institute (CRI), Czechia*

****Correspondence:***
*Weihong Lu*
*lwh@hit.edu.cn*

***Specialty section:***
*This article was submitted to Plant Abiotic Stress, a section of the journal Frontiers in Plant Science*

***Received:*** *20 March 2022*
***Accepted:*** *10 May 2022*
***Published:*** *21 June 2022*

***Citation:***
*Zeng D, Cui J, Yin Y, Dai C, Zhao H, Song C, Guan S, Cheng D, Sun Y and Lu W (2022) Combining Proteomics and Metabolomics to Analyze the Effects of Spaceflight on Rice Progeny. Front. Plant Sci. 13:900143. doi: 10.3389/fpls.2022.900143*

**Abbreviations:** APX, Ascorbate peroxidase; CAT, Catalase; CK, F2 generation control group; DAPs, Differentially Abundant Proteins; DEMs, Differential metabolites; DN423, Dongnong423; EL, Electrolyte leakage rate; CK, F2 generation control group; EL, Electrolyte leakage rate; iTRAQ, Isobaric tags for relative and absolute quantification; PCA, Principal component analysis; POD, Peroxidase; qRT-PCR, Real-time quantitative PCR; ROS, Reactive oxygen species; SOD, Superoxide dismutase; SJ-10, ShiJian-10 retractable satellite; SSC, Soluble sugar content; TCA, Tricarboxylic acid cycle; TLS, Three-Leaf Stage; TS, Tillering Stage.

## HIGHLIGHTS

- Spaceflight changed the agronomic traits of F2 generation rice.
- The F2 generation rice retained the memory of stress to space flight, which was activated by ROS.
- The assembly process of mitochondrial complex III was blocked, which was the main reason for the increase of ROS in F2 generation rice.
- The response of F2 rice to spaceflight stress caused changes in the processes of sugar metabolism, amino acid metabolism, and energy metabolism.

## INTRODUCTION

Space environment has the characteristics of strong radiation, high vacuum, microgravity, changing magnetic field and so on. Different from the earth environment, space environment has a strong mutagenic effect on organisms. In the past few decades, scientists have tried to use various returnable satellites, the International Space Station, spacecrafts and other aircrafts to carry plants for research in space life sciences. From the change of plant agronomic traits to the research of molecular level, the influence of spaceflight on plants has been revealed gradually. In recent years, omics techniques have been used to explore the mechanism of plant mutations caused by spaceflight. However, these studies have rarely reported the simultaneous use of two or more omics to analyze the effects of spaceflight on plants.

It has been proved by Vaulina et al. (1984) that spaceflight can cause chromosome aberration, gene deletion and recombination of crops, thus causing genetic material variation. Moreover, crops adapted to the effects of spaceflight by adjusting their own metabolic network (Paul et al., 2017), which was induced by oxidative stress and heat stress (Manian et al., 2021). The current research conclusions showed that spaceflight has caused changes in plant lipid peroxidation, antioxidant enzymes, energy metabolism, signal transduction, protein synthesis, cell wall biosynthesis and other processes (Matía et al., 2010; Ferl et al., 2015; Choi et al., 2019). At the same time, studies have found that spaceflight increased the content of soluble sugar, glucose, fructose, sucrose and total starch (Mortley et al., 2008). Our previous study confirmed that spaceflight stress affected different growth periods in contemporary rice. After spaceflight, contemporary plants have undergone significant changes in metabolic pathways such as energy metabolism, amino acid metabolism, sugar metabolism, and vitamin B6 metabolism. It is believed that these changes were caused by the disruption of ROS balance in rice after spaceflight (Deyong et al., 2020; Zeng et al., 2020, 2021). Recent studies have shown that abiotic stress has genetic effects across generations (Kumar et al., 2015). As a special abiotic stress, spaceflight also has intergenerational genetic effects, and how the offspring rice adapts to the intergenerational genetic effects by adjusting its own metabolic network needs to be reported. Over the past 20 years, our team has carried 50 different varieties of rice seeds by using China's recoverable satellites and spacecraft. After returning to the ground for planting, we analyzed the biological effects and confirmed that short-term low orbit flight can also cause the biological effects on rice seeds (Sun et al., 2019). Our team planted rice seeds to M2 generation after spaceflight, and found mutation sites in the DNA of M2 generation plants, whose rate was between 0.05 and 0.52% (Yu et al., 2007). The plant height, heading date, leaf color, leaf shape, flag leaf angle, awn, panicle length, panicle shape, maturity and other agronomic traits of plants in the M2 generation also changed significantly (Wei et al., 2006; Yu et al., 2007). Our team's early results suggested that the impact of spaceflight on rice seeds were heritable, but how can M2 plants respond to the impact of spaceflight by regulating their own metabolic network? Our team used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) to analyze the proteome of the M2 generation plants. Proteins involved in amino acid and derivative metabolism, pentose phosphate shunting, stress response, seed maturation, protein folding, glycolysis, lipid biosynthesis, glycogen biosynthesis and TCA cycle were differentially expressed in M2 generation plants (Sun et al., 2019). However, the ability of 2D-PAGE to identify differentially abundant proteins was limited, so it may greatly limit our understanding of how M2 generation plants respond to spaceflight. At the same time, there was no report on the changes in the metabolic map of the M2 generation plants after spaceflight.

Clarifying the response mode of progeny plants to spaceflight is helpful to reveal the mechanism of spaceflight on crop genetic effects, which is crucial to the study of the mechanism of biological effects caused by spaceflight. In this study, DN423 seeds were placed in biological irradiation box A (BRB-A) for space flight. The surface material of the BRB is aluminum with an average thickness of 2.5 mm. The space radiation measurement module and the model organism module are included in BRB-A. Rice seeds were immobilized in the model organism module of BRB-A, and there was no mechanical collision between rice seeds and BRB-A. BRB-A was fixed on the -Y axis of the SJ-10 satellite for space flight (Sun et al., 2019). The SJ-10 returnable satellite was launched at 01:38 on April 6, 2016, with an orbital altitude of 252 kilometers, an inclination of 42°, and an orbital flight of 12.5 days (Sun et al., 2019). The g-profile during launch was as follows: the first-level maximum static overload was 4.8 g in flight for 150 s; the second-level maximum static overload was 6.0 g in flight for 180 s (Xu et al., 2018). After entering orbit, the radiation dose rate measured by TLD700 was 0.075 $\pm$ 0.005 mGy/d, and the total LET radiation dose measured by CR-39 was 0.970 $\pm$ 0.055 mGy (Zhou et al., 2018), the load temperature was 20.6–23.6°C, and the gravity level was $10^{-4}$∼$10^{-6}$ g. During landing, the temperature of the recovery cabin was controlled at 22 $\pm$ 2°C, and the landing speed was 12.5 m/s (Zhao et al., 2019). First, the agronomic traits of the F2 generation plants were analyzed, and then the changes in the redox state in the body were discussed. iTRAQ proteomics and non-targeted metabolomics were used to explain the response mode of F2 generation plants to spaceflight, and finally using RT-qPCR to verify the results of the proteomics and metabolomics. This study broadens our understanding of spatial biological effects, and also provides a basis for subsequent research on how

F2 generation plants retain memories of abiotic stress. At TLS and TS, we assessed the plant height of rice. At TS, we assessed tiller number. At maturity, we assessed the rice's, setting percentage, grain number per panicle, thousand seed weight. In addition, we also evaluated physiological indicators such as $H_2O_2$, MDA, soluble sugar contents, electrolyte leakage rate, APX activity, SOD activity, CAT activity, and POD activity of rice during TLS and TS, respectively.

# MATERIALS AND METHODS

## Plant Materials

The cultivation of rice material was used as described by Yu et al. (2007) and Deyong et al. (2020). Briefly, the DN423 rice seeds were carried by the SJ-10 returnable satellite for 12.5 days of spaceflight, and then planted after 400 rice seeds returning to the ground. When sampling at the tillering stage of the F1 generation, we randomly selected 30 rice plants for sampling and labeled them. After the F1 generation matures, the 30 rice seeds were harvested, and 400 F1 generation seeds were randomly selected for planting to obtain the F2 generation. The control group was sampled and planted in the same way. The F2 generation rice was planted in Wuchang City, Heilongjiang Province in April 2017. For the F2 generation plants, we harvested the leaves at the three-leaf stage (TLS) and the tillering stage (TS), and stored them at −80°C for further analysis. In F2 generation plants, five plants were randomly selected and their leaves were mixed to obtain a single biological replicate. In the metabolomics experiment, we performed six biological replicates, while we performed three biological replicates in others. Use SP2 to represent the F2 generation plant of the rice seed after spaceflight, and CK to represent the F2 generation plant of the rice seed without spaceflight. Seeds of DN423 (*Oryza sativa L.*) were provided and certified by the Agricultural College of Northeast Agricultural University.

## Detection of $H_2O_2$ and MDA Formation in Leaves

The $H_2O_2$ was measured according to the method described by Velikova et al. (2000). Simply, take about 200 mg of rice leaves were taken and homogenized in 2 mL of 0.1% (w/v) trichloroacetic acid. The homogenate was centrifuged at 4°C (12,000 × g for 15 min). Subsequently, take 0.25 mL of the homogenate and add 0.75 mL buffer solution to it, consisting of 10 mM potassium phosphate (pH 7.0) and 1 mL 1 M potassium iodide (KI), with a final volume of 2.0 mL in each tube. The absorbance value was measured at 390 nm and calculate the $H_2O_2$ content.

The determination of Lipid peroxidation was carried out according to the method described by Heath and Packer (1968). About 200 mg of rice leaves were homogenized in 2 mL of 0.1% (w/v) trichloroacetic acid, and then centrifuged at 4°C (12,000 × g for 15 min). Take 0.5 mL of the supernatant and add it to 1.5 mL of thiobarbituric acid, and then incubated at 90°C for 20 min. The reaction solution was put on ice to terminate the reaction, and the MDA content was calculated by reading the absorbance at 535 nm and 600 nm.

## Detection of Electrolyte Leakage Rate

The electrolyte leakage rate (EL) was measured according to the method described by Lutts et al. (1996). Simply, 0.5 g of rice leaves were placed in 25 mL of deionized water, at 25°C for 3 h, and the measured conductivity was recorded as H1. Then boiled for 10 min, cooled to 25°C and measured the conductivity, which was recorded as H2, and calculated EL according to the following formula:

$$EL = H1/H2 \times 100\%$$

## Detection of Soluble Sugar Content

The soluble sugar content was determined according to the method described in Bailey (1958). Simply, 100 mg of rice leaves were weighed, 5 mL of 80% ethanol was added, and extracted at 80°C for 30 min. Then the sample was centrifuged to collect the supernatant. The extraction was repeated three times. Next, anthrone reagent was added to the supernatant and incubated at 95°C for 20 min for color reaction. After completion, the reaction was stopped on ice and the absorbance was measured at 620 nm.

## Determination of Antioxidant Enzyme Activity

About 0.5 g rice leaves were placed in a homogenization tube, 5 mL phosphate buffer (100 mM, pH 7.4) was added, ground thoroughly on ice, and then centrifuged at 11,000 *g* for 15 min at 4°C to collect. The supernatant was used as a crude enzyme extract for subsequent enzyme activity analysis. The peroxidase (POD) activity was measured according to the method described by Castillo et al. (1984). The catalase (CAT) activity was measured according to the method described in Aebi (1984). The activity of SOD and APX was measured according to the method described by Foreman et al. (2003).

## Leaves Protein Extraction

The protein in the leaves was extracted using the method described previously (Deyong et al., 2020; Zeng et al., 2020). The rice leaves were grounded into powder in liquid nitrogen, mixed with a trichloroacetic acid/ethanol mixture (1:9), centrifuged and the supernatant was removed. The pellet was air-dried after centrifugation. Then, STD buffer (4% SDS, 1 mM DTT, 150 mM Tris–HCl pH 8.0) was added to the precipitate. After sonication, it was placed in a boiling water bath and kept for 5 min. Next, the supernatant was centrifuged and the total protein content was calculated with bicinchoninic acid (BCA). Then, 200 μg of protein was taken and 200 μL of buffer (8 M Urea, 150 mM Tris–HCl pH 8.0) was added for protein digestion. The absorbance was measured at 280 nm to quantify the peptide.

## iTRAQ Labeling and Peptides Analysis

ITRAQ labeling and peptides analysis was performed according to the method we described earlier (Deyong et al., 2020; Zeng et al., 2020). Take 80 μg peptides from each group for iTRAQ

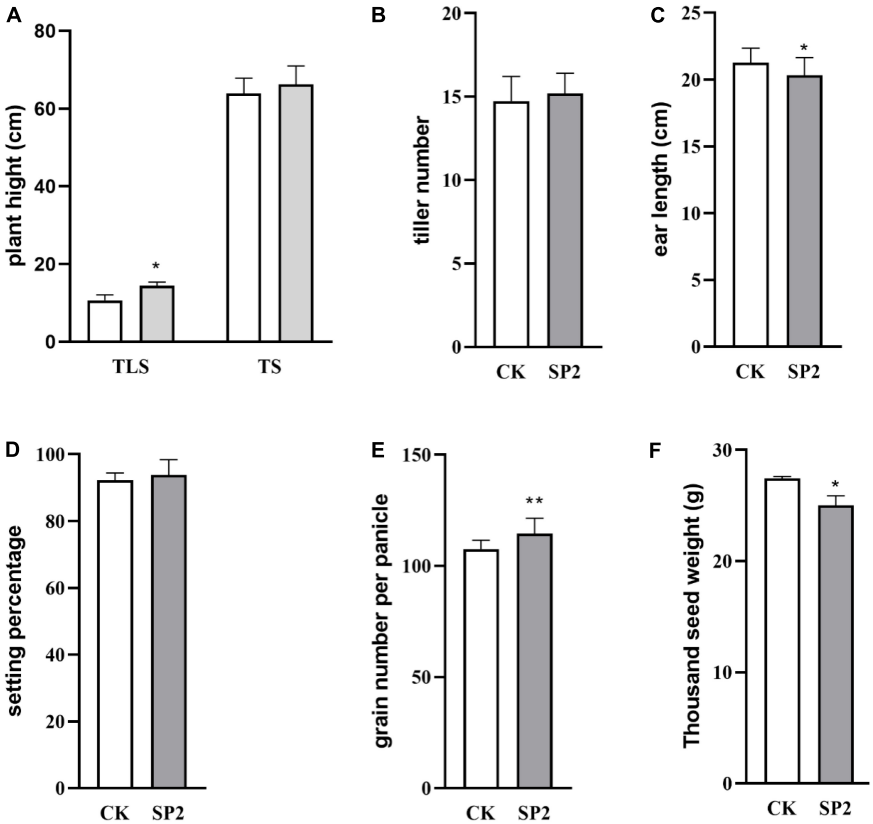



**FIGURE 1 |** Change of plant height **(A)**, tiller number **(B)**, ear length **(C)**, setting percentage **(D)**, grain number per panicle **(E)**, thousand seed weight **(F)** in rice different stages of growth and development. (TLS, three-leaf stage; TS, tillering stage; White column represents the control group; Gray column represents treatment group). Data are mean ± SD, $n = 30$, * and ** indicate significant difference at $p < 0.05$ and $p < 0.01$ by student *t*-test, respectively.

labeling with iTRAQ Reagent-8plex Multiplex Kit (AB SCIEX, United States) according to the instructions. Subsequently, EASY-nLC 1,000 liquid chromatograph (Thermo Finnigan, United States) was used for separation, and Q-Exactive mass spectrometer (Thermo Finnigan) was used for identification. Then the raw data of the mass spectrometer was analyzed by the method described in Cui et al. (2019). The UniProt database ([1]iTRAQ labeling and peptides analysisaccessed on 16 August 2020) was used for protein identification. When accessed, the database contained 148,104 rice protein sequences [Oryza sativa subsp. japonica (Rice, 39947)]. False discovery rate (FDR) < 0.01. The peptide mass tolerance was ± 10 ppm and the fragment mass tolerance were 0.2 Da. In order to determine the relative difference in protein abundance, we defined proteins with a fold change rate ≥ 1.2 or ≤ 0.83 and $p < 0.05$ as differentially abundant proteins (DAPs).

## Extraction of Metabolites and Metabolomics Analysis

Same as previous research (Zeng et al., 2021). About 0.1 g of sample was grounded into powder in liquid nitrogen, methanol: acetonitrile: water (1 mL, 2:2:1, v/v/v) was added and mixed well. Ultrasonic extraction (100 W, 60 min, 4°C) was performed kept standing for 60 min at −20°C, and centrifuged for 20 min (4°C, 14,000 *g*). The centrifuged supernatant was then vacuumed dry. Then 100 μL acetonitrile: aqueous solution (1:1, v/v) was added to the dried sample, and the supernatant was collected by centrifugation for analysis. The sample was separated by ultra-high-performance liquid chromatography (Agilent 1290, United States). The column used for separation was a HILIC column with a column temperature of 25°C, a flow rate of 0.3 mL/min, and an injection volume of 2 μL. The mobile phase consisted of A: water + 25 mM ammonium acetate + 25 mM ammonia, B: acetonitrile. The separated samples enter the Triple TOF 6600 mass spectrometer (AB SCIEX) to be identified. Data acquisition was conducted in full-scan mode in combination with information-dependent acquisition mode. The parameters were set as follows: ion spray voltage, 5,500 V (+) and 5,500 V (−); ion source temperature, 600°C (+) and 600°C (−); collision energy, 35 ± 15 eV(−); curtain gas of 30 PSI; The original data was converted into. mzML format by ProteoWizard, and then the XCMS program was used for peak alignment, retention time correction and peak area extraction. XCMS software parameter settings were as follows: For peak picking, cent

[1]https://www.uniprot.org/

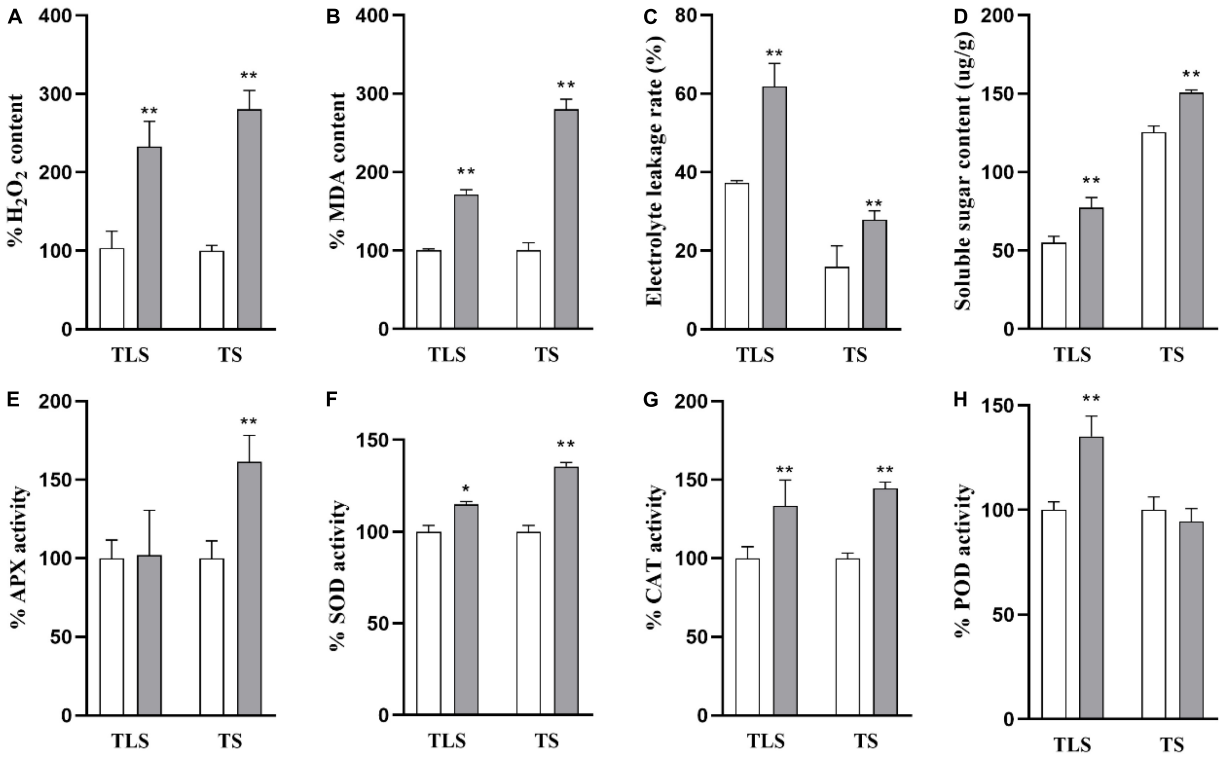



**FIGURE 2 |** The physiological changes of rice induced by spaceflight. **(A)** Percent values of $H_2O_2$. **(B)** Percent values of MDA. **(C)** The soluble sugar contents. **(D)** The electrolyte leakage rate change. **(E)** Percent values of APX activity. **(F)** Percent values of SOD activity. **(G)** Percent values of CAT activity. **(H)** Percent values of POD activity (TLS, three-leaf stage; TS, tillering stage; White column represents the control group; Gray column represents treatment group. The data (mean ± SD) are the means of three replicates with standard errors shown by vertical bars, $n = 3$, * and ** indicate significant difference at $p < 0.05$ and <0.01 by student *t*-test, respectively.

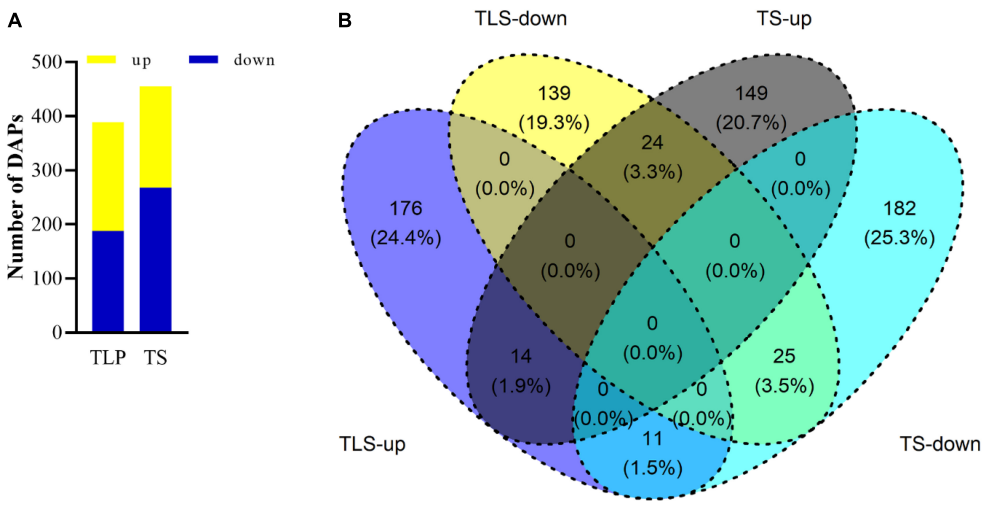



**FIGURE 3 |** The number of up-regulated and down-regulated DAPs in two different growth stages **(A)**. Venn diagram showed the overlaps of the DAPs in in different growth stages **(B)**. TLS-up represented the three-leaf stage up-regulated proteins; TLS- down represented the three-leaf stage down-regulated proteins; TS-up represented tillering stage up-regulated proteins; TS-down represented tillering stage down-regulated proteins.

Wave m/z = 25 ppm, peak width = c (10, 60), prefilter = c (10, 100). For peak grouping, bw = 5, mz wid = 0.025, minfrac = 0.5 were used. Metabolite structure identification was carried out by accurate mass matching (< 25 ppm) and secondary spectrum matching. Reference material databases built by Dalian Institute of Chemical Physics and Shanghai

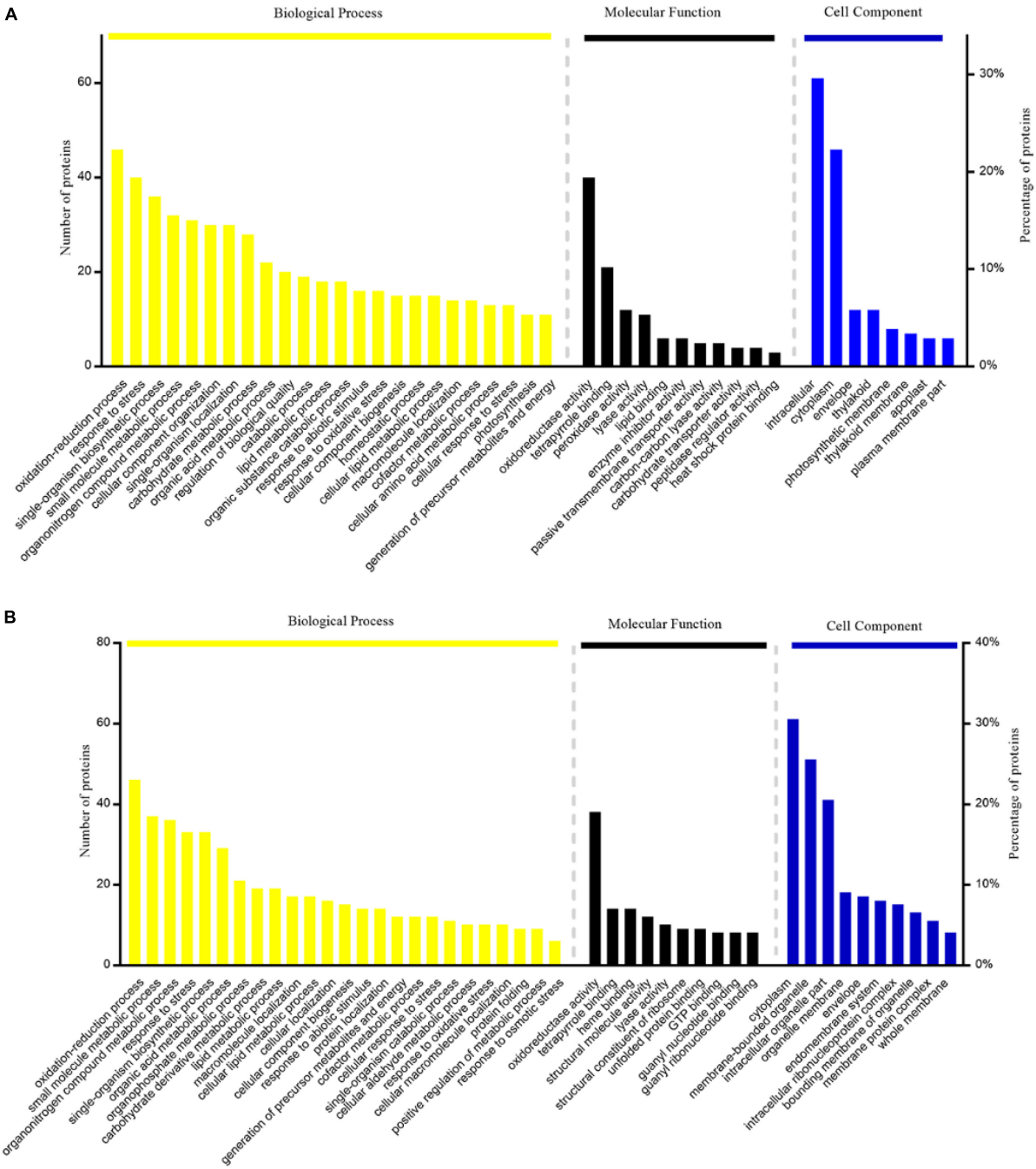



**FIGURE 4 |** GO enrichment analysis of proteins with differential abundance: **(A)** three-leaf stage SP2 vs CK; **(B)** tillering stage SP2vs CK.

Zhongke New Life Biotechnology Co., Ltd. The data was input into the software SIMCA-P 14.1 (Umetrics, Umea, Sweden) for pattern recognition, preprocessed by Pareto-scaling, and then subjected to multi-dimensional statistical analysis. Orthogonal PLS-DA (OPLS-DA) and metabolic pathway analysis both used MetaboAnalyst 4.0 software[2]. Principal component analysis (PCA) was performed by R software. The pheatmap package of R software was used for hierarchical cluster analysis.

## Bioinformatics Analysis

Gene Ontology and eggnog databases were used for functional annotation and classification of DAPs. Kyoto Encyclopedia of Genes and Genomes (KEGG)[3] was used to analyze the major

[2]http://www.metaboanalyst.ca/

[3]http://www.kegg.jp/

metabolic pathways of DAPs and DEMs. We believed that the GO terms and KEGG pathways were significantly enriched when $P$-value $\leq 0.05$ (Yang et al., 2018). In addition, we used https://www.ricedata.cn/gene/ search for DAPs related to rice agronomic traits.

## Total RNA Extraction and Real-Time PCR

As mentioned in the previous study (Deyong et al., 2020; Zeng et al., 2020), the rice leaves were fully ground in liquid nitrogen, and then we used TaKaRa kit (9767) to extract total RNA, and denaturing agarose gel electrophoresis and Micro Drop (BIO-DL Co., Ltd., Shanghai, China) were used to evaluate the quality and concentration of total RNA. Next, TaKaRa kit (RR037A) was used to reverse transcribe total RNA into cDNA. SYBR Premix Ex Taq II [TaKaRa kit (820A)] was used to detect gene expression. Relative expression levels of genes were determined using a relative quantitative method ($2^{-\Delta\Delta CT}$) (Cui et al., 2019). All primers were listed in **Supplementary Table 1**.

# RESULTS

## The Morphological Changes of Rice Progeny Induced by Spaceflight

In order to better understand the impact of spaceflight on rice progeny, we analyzed the agronomic traits of F2 plants at different growth and development stages (**Figure 1**). The plant height of the progeny after spaceflight increased significantly in TLS, while showed no significant difference in TS (**Figure 1A** and **Supplementary Figure 1**). Then we evaluated the impact of spaceflight on the tiller number ability of the F2 generation, which showed that the effective tiller number of the offspring did not change obviously after spaceflight (**Figure 1B**). Similarly, we evaluated the seed setting rate and ear length, and the results showed that there were rarely obvious changes in the F2 generation after spaceflight (**Figures 1C,D**). However, they showed significant differences in the number of grains without ears and thousand grain weight (**Figures 1E,F**). From the above results, we preliminarily concluded that the effect of spaceflight on rice still existed in the F2 generation of rice, which made the F2 generation plants of spaceflight show different agronomic traits from the control. This also showed that spaceflight had an impact on the physiological processes of F2 generation plants.

## The Physiological Changes of Rice Progeny Induced by Spaceflight

In order to further explore the changes in the physiological processes of rice F2 generation plants after spaceflight, we measured the levels of $H_2O_2$, MDA, EL, SSC and the activity of antioxidant enzymes (**Figure 2**). In the TLS and TS, $H_2O_2$, MDA, EL, SSC in SP2 were much higher than that of the control (**Figures 2A–D**). In detail, $H_2O_2$ levels increased by 132.9 and 180.6%, respectively (**Figure 2A**), MDA levels increased by 71.4 and 179.7%, respectively (**Figure 2B**), EL levels increased by 66.2 and 79.1%, respectively (**Figure 2C**), SSC levels increased by 41.1 and 20.2%, respectively (**Figure 2D**), in TLS and TS. The SOD, POD and CAT activities all significantly increased by 14.9, 35.1, and 33.4%, respectively, in the TLS (**Figures 2F–H**). However, there was no significant difference in the activity of APX in the SP2 group compared to the control in the TS (**Figure 2E**). Similarly, our results showed that the activity of antioxidant enzymes changes during TS (**Figures 2E,G,H**). Totally, these results suggested that the redox balance of SP2 was broken.

## Protein Profiles of Rice Leaves at Different Developmental Stages

We used iTRAQ proteomics technology to evaluate the changes in the proteome during the two growth stages. A total of 3,867 proteins were identified in three biological replicates, and

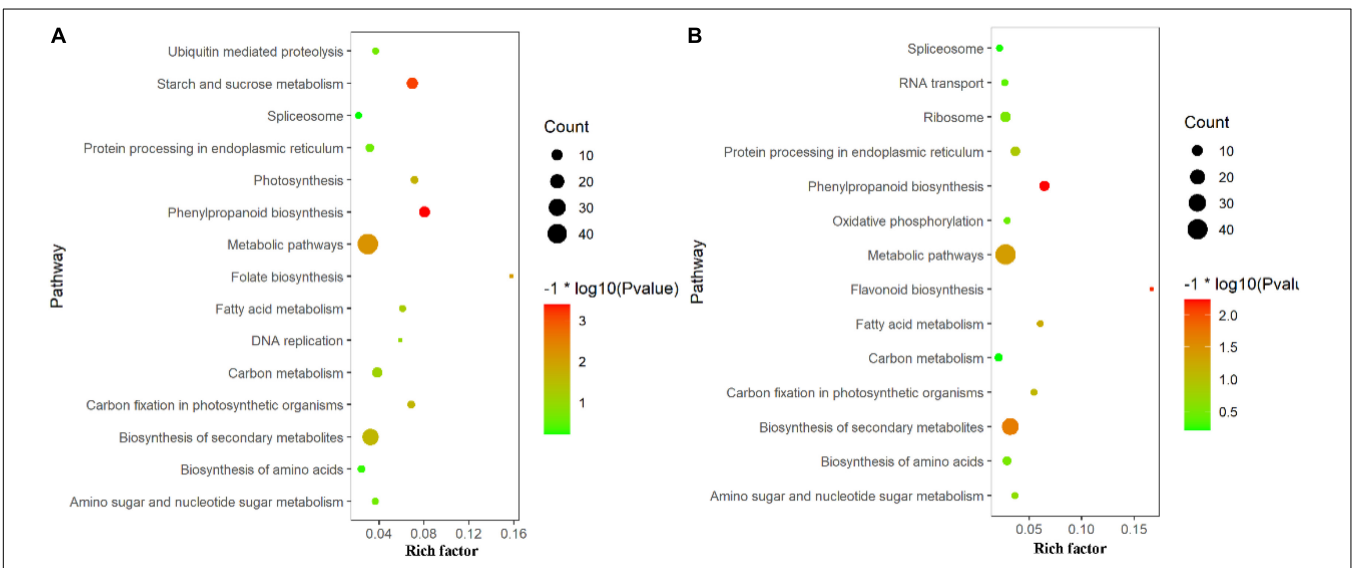



**FIGURE 5 |** KEGG pathway enrichment analysis of proteins with differential abundance; TOP 15 of pathway enrichment; **(A)** three-leaf stage; **(B)** tillering stage.

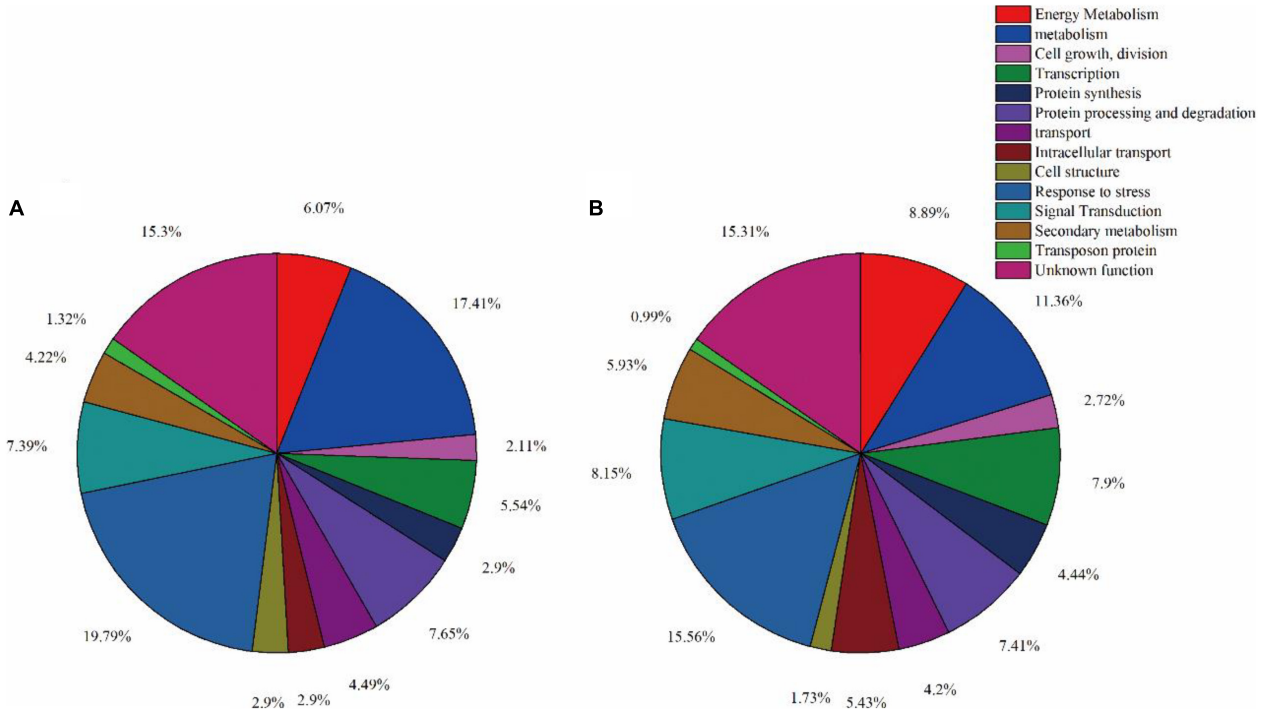



**FIGURE 6 |** Function classification of differentially expressed proteins in leaves at three-leaf stage **(A)** and tillering stage **(B)**.

proteins with a ratio $\geq$ 1.2 (treatment group/control group) and adjusted $p$ value were identified as differentially abundant proteins (DAPs). The abundances of 389 and 405 proteins were significantly changed in the TLS and the TS, respectively. Among them, 201 proteins were up-regulated while the other 188 proteins down-regulated in the TLS. Meanwhile, 187 proteins up-regulated and 268 proteins down-regulated during the TS (**Figure 3A**). The Venn diagram was used to characterize the overlap of proteins at different growth stages (**Figure 3B**). There were 315 and 331 unique proteins in TLS and TS, respectively. At the same time, there were 74 co-expressed proteins in the two growth stages. Among them, the expression abundance of 36 proteins showed opposite changes, and the expression trend of 38 proteins did not change (**Figure 3B**).

## Gene Ontology Annotation of DAPs

In this study, Gene Ontology (GO) term enrichment was used to reveal the functional characteristics of DAPs. We selected GO items with level 4 for analysis. The biological processes (BP) that were significantly enriched in TLS include oxidation-reduction process, response to stress, single-organism biosynthetic process, small molecule metabolic process, etc. (**Figure 4A**). For molecular functions (MF) that were significantly enriched, there were oxidoreductase activity, tetrapyrrole binding, peroxidase activity, lyase activity, etc. (**Figure 4A**). The intracellular, cytoplasm, envelope, and thylakoid (**Figure 4A**) were the most abundant groups under Cell component (CC). Similarly, in the TS (**Figure 4B**), the BP analysis showed that oxidation-reduction process was the most representative term, followed by small molecule metabolic process, organonitrogen compound metabolic process, response to stress and single-organism biosynthetic process, etc. As for CC, cytoplasm, membrane-bounded organelle, intracellular part, etc. In the MF, the annotated proteins were mainly involved in the oxidoreductase activity, tetrapyrrole binding and heme binding, etc. In addition, we used the KEGG database to map DAPs on metabolic pathways. We have enriched 68 metabolic pathways in the 389 DAPs of TLS. They mainly included Phenylpropanoid biosynthesis, Starch and sucrose metabolism, Metabolic pathways, DNA replication, Biosynthesis of amino acids, Biosynthesis of secondary metabolites, Photosynthesis, etc. (**Figure 5A**). The 405 DAPs of TS were mapped to 67 metabolic pathways, including Phenylpropanoid biosynthesis, Flavonoid biosynthesis, Biosynthesis of secondary metabolites, Metabolic pathways, Biosynthesis of amino acids, Oxidative phosphorylation, etc. (**Figure 5B**).

## Functional Classification of DAPs

In order to better understand the impact of spaceflight on rice progeny, we classified the functions of DAPs according to the method described by Bevan et al. (1998) and the results of GO annotation. We divided the function of protein into 14 major functional categories and several functional sub-categories including metabolism, energy metabolism, cell growth/division, etc. (**Supplementary Tables 2**, **3** and **Figure 2**). The 389 DAPs in the TLS leaves were related to energy metabolism (6.07%), metabolism (17.41%), cell growth and division (2.11%), transcription

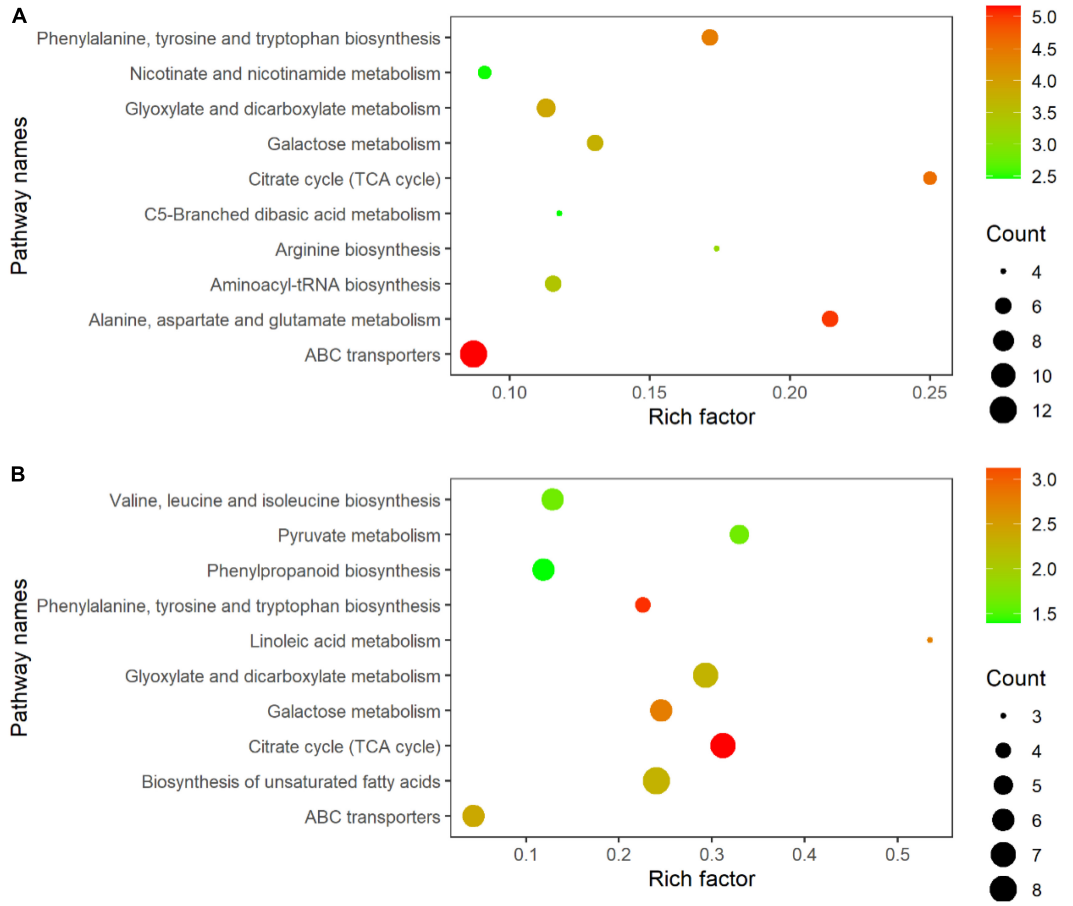



**FIGURE 7 |** KEGG pathway enrichment analysis of metabolites with differential abundance; **(A)** three-leaf stage; **(B)** tillering stage.

(5.54%), protein synthesis (2.11%), protein processing and degradation (7.65%), transport (4.49%), intracellular transport (2.90%), cell structure (2.90%), response to stress (19.79%), signal transduction (7.39%), secondary metabolism (4.22%), transposon protein (1.32%) and unknown (15.30%) (**Figure 6A**). The 405 differentially expressed proteins in the TS leaves were related to energy metabolism (8.89%), metabolism (11.36%), cell growth and division (2.72%), transcription (7.9%), protein synthesis (4.44%), protein processing and degradation (7.41%), transport (4.20%), intracellular transport (5.43%), cell structure (1.73%), response to stress (15.56%), signal transduction (8.15%), secondary metabolism (5.93%), transposon protein (0.99%) and unknown (15.31%) (**Figure 6B**).

## Metabolome Profiles of Rice Leaves at Different Developmental Stages

LC-MS was used to evaluate the changes in the metabolic profile of rice offspring leaves by spaceflight. PCA was used to determine the degree of separation of the metabolic profiles of the SP2 group and the CK group. At the three-leaf stage, the difference between CK and SP2 was 75.91% (**Supplementary Figure 2**), and at the TS, the difference between CK and SP2 was 79.3% (**Supplementary Figure 2**). In the two periods, the CK group was completely separated from the SP2 group, which indicated that the impact of spaceflight on rice continued to the offspring. In the **Supplementary Figure 2**, all the data points formed close clusters between different comparison groups, which also reflected the repeatability of metabolomics results.

In order to find metabolites with significant differences, we performed an orthogonal PLS-DA (OPLS-DA) analysis (**Supplementary Figure 3**). We defined metabolites with a ratio $\geq 1.3$, Variable of Importance to the Projection (VIP) $> 1$ and adjusted $p$ value $< 0.05$ as differential metabolites (DEMs). A total of 124 DEMs were identified during TLS, of which 70 increased and 54 decreased (**Supplementary Table 4**). 125 DEMs were identified in TS, 87 of which increased and 38 decreased (**Supplementary Table 5**).

Hierarchical cluster analysis showed that the metabolites of SP2 and CK groups were separated from each other in two different growth and development stages (**Supplementary Figure 4**), which indicated that the impact of spaceflight panicle rice still existed in the offspring, and the physiological response of rice was still affected

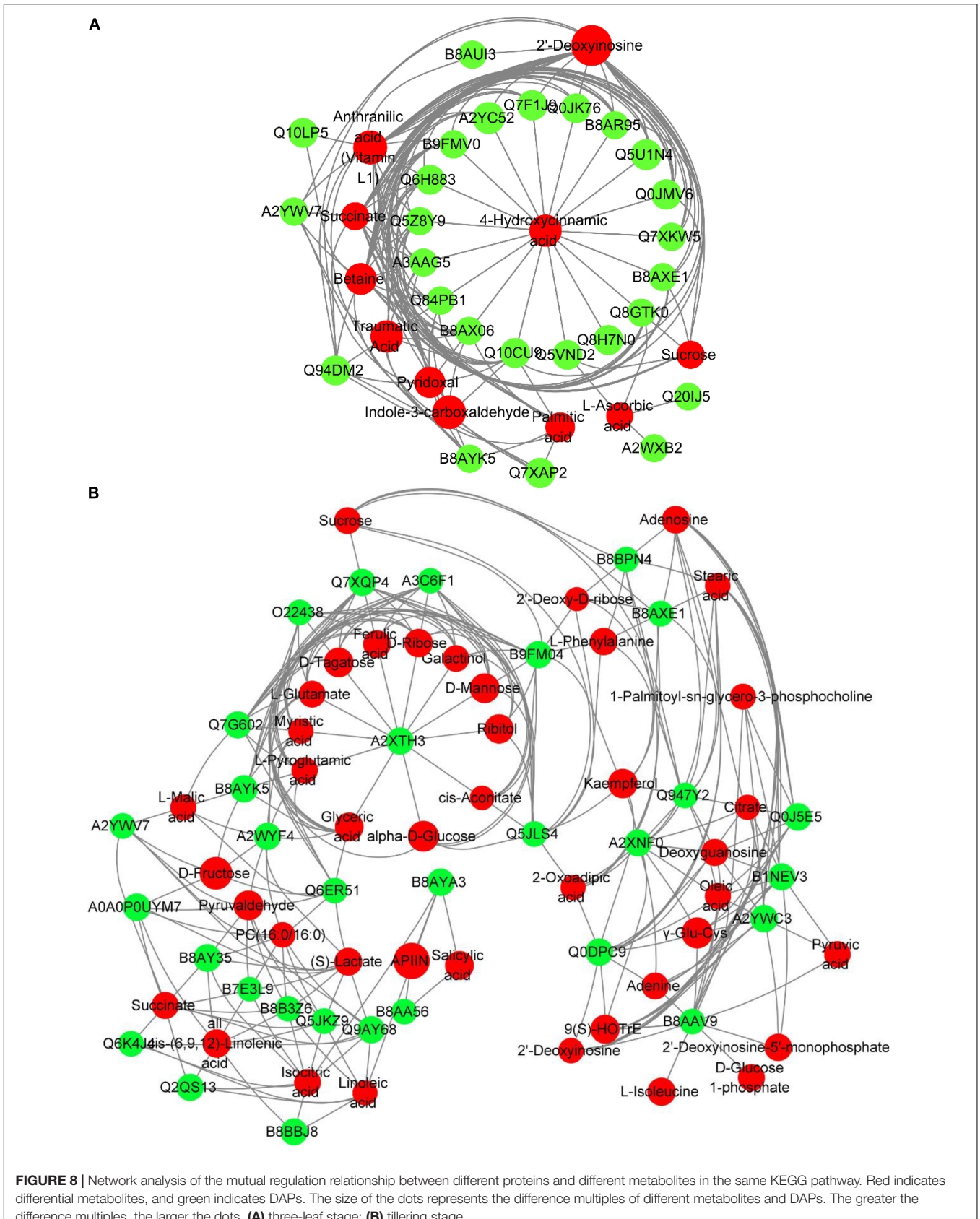



**FIGURE 8 |** Network analysis of the mutual regulation relationship between different proteins and different metabolites in the same KEGG pathway. Red indicates differential metabolites, and green indicates DAPs. The size of the dots represents the difference multiples of different metabolites and DAPs. The greater the difference multiples, the larger the dots. **(A)** three-leaf stage; **(B)** tillering stage.

(**Supplementary Figure 4**). In addition, it indicated that the two growth and development stages of the F2 generation rice had different response mechanisms to the spaceflight (**Supplementary Figure 4**).

## Analysis of Metabolic Pathways of Rice Leaves at Different Developmental Stages

All DEMs were submitted to the KEGG database for analysis of related pathways. The results showed that DEMs in TLS were mapped to 68 metabolic pathways (**Supplementary Table 6**), of which the top ten metabolic pathways mainly included ABC transporters; Alanine, aspartate and glutamate metabolism; Citrate cycle (TCA cycle); Phenylalanine, tyrosine and tryptophan biosynthesis; Glyoxylate and dicarboxylate metabolism; Galactose metabolism; Aminoacyl-tRNA biosynthesis; Arginine biosynthesis; Nicotinate and nicotinamide metabolism; C5-Branched dibasic acid metabolism (**Figure 7A**). Similarly, we also mapped the metabolic pathways of DEMs in TS, and a total of 66 pathways were enriched (**Supplementary Table 7**). The top ten pathways of these enriched metabolic pathways included citrate cycle (TCA cycle); Phenylalanine, tyrosine and tryptophan biosynthesis; Galactose metabolism; Linoleic acid metabolism; ABC transporters; Biosynthesis of unsaturated fatty acids; Glyoxylate and dicarboxylate metabolism; Valine, leucine and isoleucine biosynthesis; Pyruvate metabolism; Phenylpropanoid biosynthesis (**Figure 7B**). The changes in these metabolites and metabolic pathways provided important information on how the offspring of rice retained the memory of spaceflight stress.

## Comprehensive Analysis of Rice Leaves Metabolomics and Proteomics

In order to further understand the molecular mechanism of rice offspring in response to the spaceflight, we comprehensively analyzed the KEGG pathway of DAPs and DEMs. We found that the changes of DEMs and DAPs during TLS were related to 34 metabolic pathways

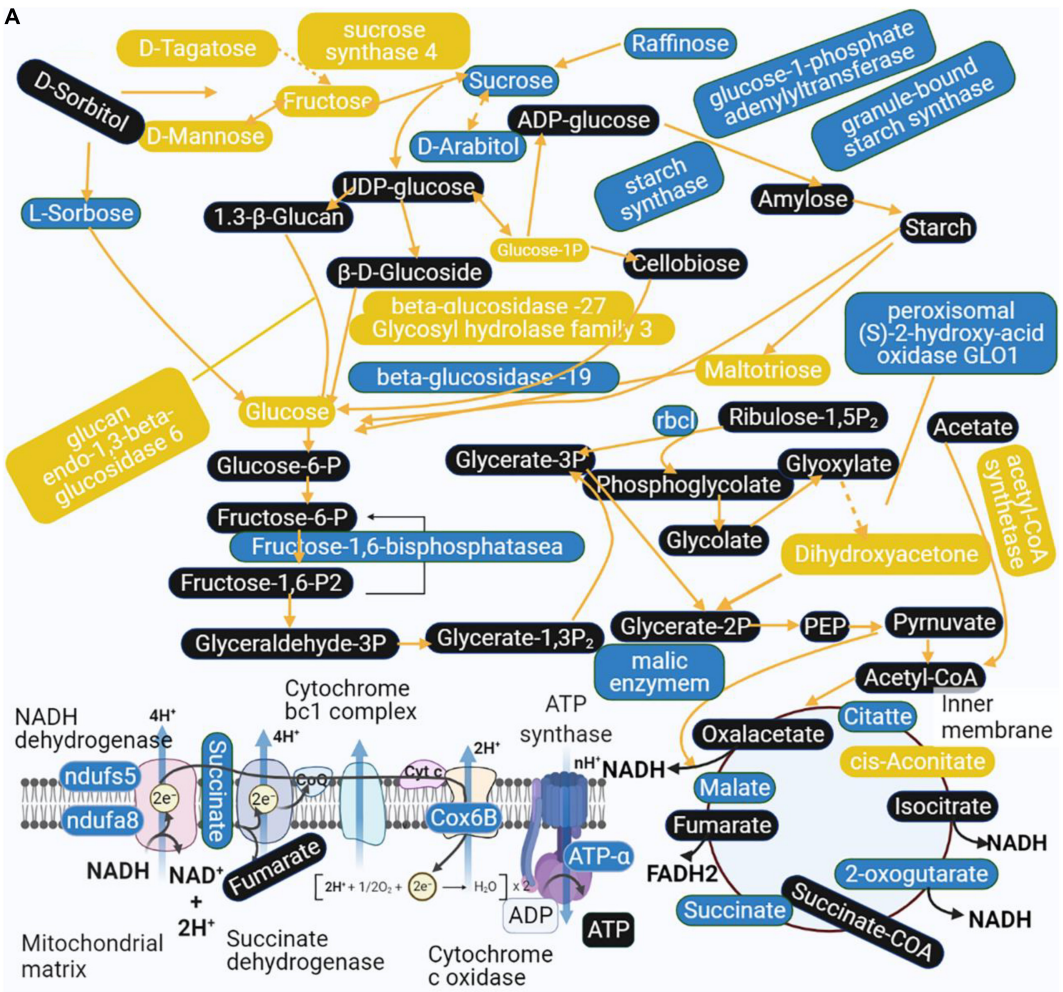



**FIGURE 9 |** (Continued)

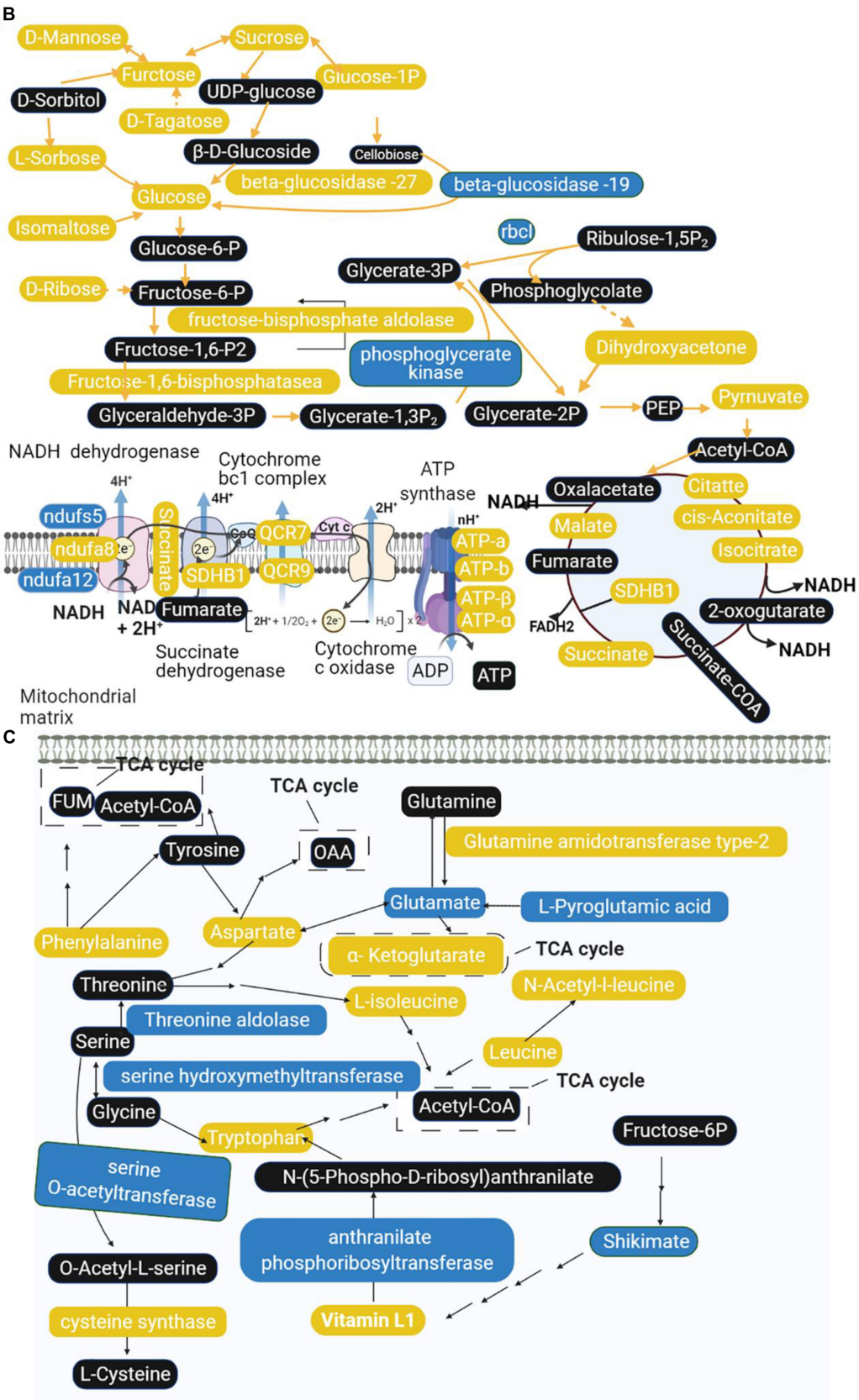



**FIGURE 9 |** (Continued)

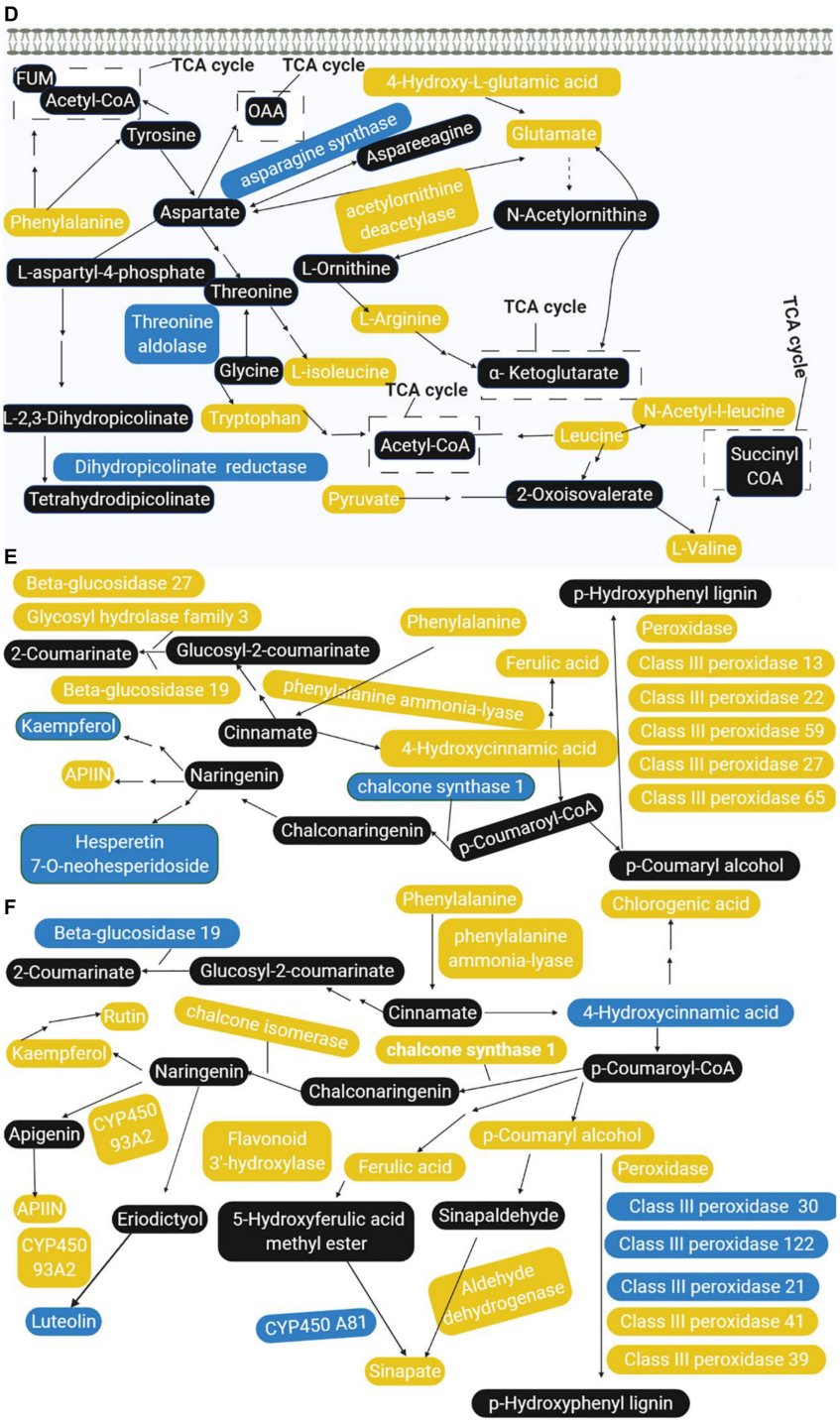



**FIGURE 9 |** Main biological pathway responses to spaceflight stress in rice progeny. **(A)** Carbohydrate metabolism and energy metabolism in TLS; **(B)** carbohydrate metabolism and energy metabolism in TS; **(C)** amino acid metabolism in TLS; **(D)** amino acid metabolism in TS; **(E)** phenylpropane biosynthesis and flavonoid metabolism in TLS; **(F)** phenylpropane biosynthesis and flavonoid metabolism in TS. The DEPs and metabolites are marked in box; yellow indicates upregulation; blue indicates downregulation; black indicates no significant change.

(**Table 1**). Moreover, we found that there were 36 pathways that were changed by DAPs and DEMs in TS (**Table 2**).

To further evaluate the interaction between DAPs and DEMs, we assessed the correlation of DAPs with DEMs by Pearson's test and visualized the correlation network using

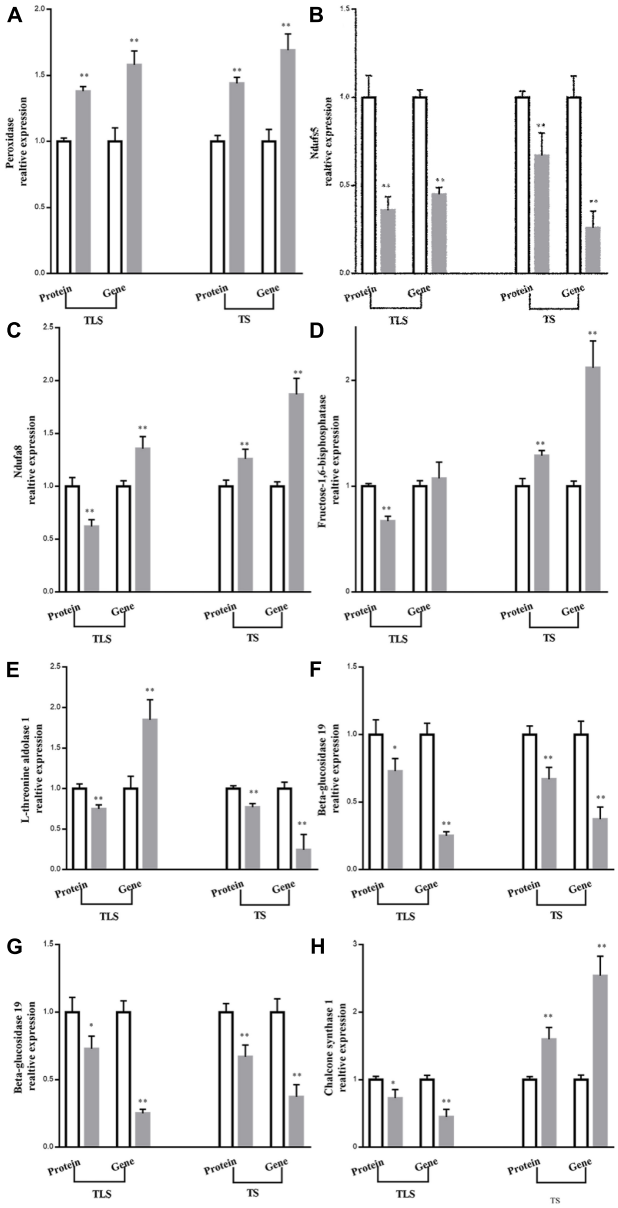



**FIGURE 10 |** Comparison between protein abundance and mRNA level. The protein abundance in ground treatment and after spaceflight treatment leaves at three-leaf stage (TLS) and tillering stage (TS) were assessed by ITRAQ based quantitative proteome analysis. Transcript abundance at two development stages (TLS, TS) was determined by quantitative RT-PCR and normalized against the *ACTIN* gene. The white column represents the control group; gray column represents treatment group. The means and standard error values from three independent samples are shown (means ± SE; $n = 3$). * and ** indicate significant difference at $p < 0.05$ and $<0.01$ by student *t*-test, respectively.

cytoscape (**Figure 8**). In TLS, 4-Hydroxycinnamic acid was at the center of the entire network, and the changes in the abundance of 11 types of DAPs [Q84PB1 (phosphoribosyl anthranilate transferase), Q7F1J9 (L-ascorbate peroxidase), Q7XKW5 (L-threonine aldolase 1), B9FMV0 (Malic enzyme), Q0JK76 (Glutathione-*S*-transferase), Q10CU9 (Glycosyl hydrolase family 3 N terminal domain protein), Q8H7N0 (alcohol dehydrogenase), Q5VND2 (Cysteine synthase), Q5Z8Y9 (Cysteine synthase), B8AXE1 (3-hydroxyacyl-CoA dehydrogenase), B8AR95 (COX6B), Q6H883 [H(+)-exporting diphosphatase), Q0JMV6 (Os01g0357100 protein), A2YC52 (Peroxidase), Q5U1N4 (Class III peroxidase 59), Q8GTK0 (Starch synthase), A3AAG5 (Methyltransf_11 domain-containing protein), B8AX06 (UbiA prenyltransferase family)] were related to changes in the content of 4-Hydroxycinnamic acid (**Figure 8A**). Interestingly, among these 11 proteins, there are proteins involved in energy metabolism (COX6B, Malic enzyme), redox balance (Class III peroxidase 59, Peroxidase, L-ascorbate peroxidase), and sugar metabolism (glutathione Stransferase, Glycosyl hydrolase family 3 N terminal domain protein, Starch synthase). This may suggest that changes in these pathways may be one of the reasons for the changes in 4-Hydroxycinnamic acid content in rice during spaceflight-induced TLS. Further, in TS, the protein A2XTH3 (Peroxidase) was located in the center of the network, and there were 12 metabolites (Ferulic acid, L-Glutamate, Ribitol, L-Pyroglutamic acid, Myristic acid, alpha-D-Glucose, D-Mannose, D-Ribose, Glyceric acid, D-Tagatose, Galactinol, *cis*-Aconitate) content changes directly related to their abundance changes (**Figure 8B**). Most of these 12 metabolites were organic acids and sugars. Notably, changes in peroxidase abundance in both TLS and TS appear to be related (directly or indirectly) to glucose metabolism. Furthermore, our results show a reciprocal regulatory relationship between changes in DAPs and DEMs.

## Analysis of the Comprehensive Systemic Metabolic Pathways Diagram

Combining the results of proteomics and metabolomics, we found that the amino acid metabolism in the offspring rice was abnormal, accompanied by related metabolites and protein metabolism disorders. Sugar metabolism has been affected, and energy metabolism has also been disturbed. It is worth noting that the amino acid metabolism has also been affected. In addition, phenylpropane biosynthesis and flavonoid metabolism showed significant changes. Based on this, we constructed a comprehensive systemic metabolic pathway map to reveal the response mechanism of rice progeny plants to the spaceflight (**Figure 9**). In TLS, the abundance of a total of seven proteins changed significantly during the process of sugar metabolism. The abundance of sucrose synthase 4, glucan endo-1,3-beta-glucosidase 6, beta-glucosidase −27, and Glycosyl hydrolase family three increased, which may account for the accumulation of fructose and glucose. The content of intermediate metabolites during the TCA cycle in TLS was reduced, and the abundance of related proteins in the electron transport chain was reduced either, which may lead to a reduction in energy metabolism flux at this time (**Figure 9A**). Unlike TLS, only one glycometabolizing protein abundance changed in TS, and D-Mannose, fructose, Sucrose, D-Tagatose, L-Sorbose, Isomaltose, Glucose, and D-Ribose were all accumulated. The changes in the abundance of related metabolites and proteins in the TCA cycle and the

electron transport chain were increased, which indicated that energy metabolism was activated compared to the control group (**Figure 9B**). There were six proteins involved in amino acid metabolism in TLS. In addition to the up-regulated expression of Glutamine amidotransferase type-2, the abundance of other proteins was down-regulated (**Figure 9C**). In TS, only four proteins were involved in amino acid metabolism at this time (**Figure 9D**). Obviously, phenylpropane biosynthesis and flavonoid metabolism in TS had a more complex regulatory network than in TLS (**Figures 9E,F**). Interestingly, chalcone synthase 1 may have an important influence on the content of Rutin, Kaempferol, APIIN and Hesperetin 7-*O*-neohesperidoside (**Figures 9E,F**).

## mRNA Expression Validations

qRT-PCR was used to further verify the validity of the results. This study selected eight proteins involved in amino acid metabolism, sugar metabolism, energy metabolism, and phenylpropane metabolism, and showed significant differences in TLS and TS for qRT-PCR (**Figure 10**). In TS, the expression abundance of the eight proteins was the same as the expression trend of the genes encoding them, which confirmed that the iTRAQ results was reliable (**Figure 10**). Moreover, in TLS, the protein expression trend of Ndfua8 and fructose-1,6-bisphosphatase was different from the gene expression trend (**Figures 10C,D**). **Figure 6** showed that DAPs involved in transcription modification and protein synthesis were enriched. The process of gene transcription to protein synthesis was affected, which may also explain why gene and protein expression trends were not the same, which further indicated that mRNA expression levels had limited effects on proteins. Therefore, it is more scientific to use multi-omics to reveal the mechanism of plants adapting to abiotic stress.

# DISCUSSION

Spaceflight is a special abiotic stress. We have reported that spaceflight causes ROS accumulation in contemporary rice plants to respond to the impact of spaceflight, thereby changing protein expression and metabolite changes to achieve metabolic rearrangement. Thus, it is necessary to further study whether the memory of spaceflight stress is retained in rice progeny plants after spaceflight, and how to adapt to the effects of spaceflight through metabolic rearrangement. In this study, we have used metabolomics and proteomics to systematically analyze how offspring rice responded to the effects of spaceflight stress.

## Effects of Spaceflight on Agronomic Characters of Offspring Rice

The results of this study showed that after space flight, the plant height of the F2 generation plants increased significantly at the TLS, but there was no significant change at the TS (**Figure 1A**), which has the same trend as the plant height of F1 generation rice plants after space flight (Deyong et al., 2020). Moreover, compared with the CK group, the SP2 group had significant changes in ear length, number of grains per ear, and 1,000-grain weight (**Figures 1C,E,F**), while the number of tillers and seed setting rate did not change significantly (**Figures 1B,D**). The agronomic traits of plants often change when they are subjected to abiotic stress (Bali and Sidhu, 2020). Under drought, salt stress, high temperature, and heavy metal stress, the plant height, ear length, grain and growth rate of crops would change (Saleem, 2003; Mahmood et al., 2007 Bakht et al., 2011; Abbas, 2012), and it is believed that the impact of abiotic stress on agronomic traits is related to the accumulation of ROS (Thounaojam et al., 2012; Bali and Sidhu, 2020). Therefore, the changes in agronomic traits of the F2 generation plants suggested that the F2 generation plants still retained the effects of spaceflight stress. In order to further explore the reasons for the changes in the phenotype of the F2 generation plants, we compared all DAPs with the proteins that control the phenotype in the https://www.ricedata.cn/gene/ database. We found that in TLS, two DAPs (Q5QM60, Q6YYV8) were associated with rice plant height, and their abundance both increased. Q5QM60 (photoperiod-sensitive dwarf 1) deficiency can lead to impaired cell division and elongation, and severely dwarf plants under long-day conditions, and neither gibberellin nor brassinosteroids can save plant height changes caused by this gene (Li et al., 2014). Q6YYV8 (BAHD acyltransferase-like protein gene; slender grain Dominant), maintained the steady state of brassinosteroids in the plant. Silencing this gene would result in smaller rice grains and dwarf plants, but over-expression of this gene would not change the existing phenotype (Feng et al., 2016). Therefore, in this study, the change in plant height during TLS was mainly due to the difference in the expression of photoperiod-sensitive dwarf 1. Although there was no significant change in plant height during TS, we also found two DAPs related to plant height. A0A0P0VS15 (basic transcription factor 3) was a basic transcription factor, and inhibiting its expression would result in rice plant dwarf and typical pollen abortion. In addition, basic transcription factor 3 was regulated by abiotic stresses such as salt, high temperature and exogenous plant hormones (Wang et al., 2012). Q6EUP4 (GA-insensitive dwarf 2), which was an F-box subunit of a SCF E3 complex, mediated GA signal transduction in rice and affects rice plant height. Rice mutants that silence GID2 exhibited severe dwarfing, broadened leaves, dark green in color, and sterility (Gomi et al., 2004). We found a significant change in the protein that controlled the number of grains per panicle (Q03200). Q03200 (Light-induced rice 1, LIR1) was a chloroplast protein, which regulated the adhesion of ferredoxin $NADP^+$ oxidoreductase (LFNR) to thylakoid membrane. LIR1 and LFNR form thylakoid protein complexed with TIC62 and TROL. LIR1 can increase the affinity of LFNR and TIC62. Light induced the rapid degradation of LIR1 and releases LFNR from Thylakoid Membrane. Silencing the protein showed growth retardation, and the seeds produced were about 75% of the control group (Yang et al., 2016).

## The F2 Generation Rice Still Retained the Memory of Spaceflight Stress

Among the various types of ROS, $H_2O_2$ has received the most attention (Xie et al., 2019). Several recent studies have shown

**TABLE 1 |** Proteins and metabolites involved in common pathways at TLS.

| Pathway name | Metabolomics | | Proteomics | |
|---|---|---|---|---|
| | Pathway ID | *P*-value | Pathway ID | *P*-value |
| Phenylalanine, tyrosine and tryptophan biosynthesis | map00400 | 3.6484E-05 | osa00400 | 6.1700E-01 |
| Glyoxylate and dicarboxylate metabolism | map00630 | 1.2823E-04 | osa00630 | 1.6800E-01 |
| Ascorbate and aldarate metabolism | map00053 | 1.2647E-02 | osa00053 | 5.7100E-01 |
| Glycine, serine and threonine metabolism | map00260 | 1.3556E-02 | osa00260 | 7.2900E-01 |
| Pyruvate metabolism | map00620 | 1.9406E-02 | osa00620 | 8.2900E-01 |
| Glutathione metabolism | map00480 | 3.3203E-02 | osa00480 | 2.2500E-01 |
| Cyanoamino acid metabolism | map00460 | 5.0944E-02 | osa00460 | 4.2800E-02 |
| Tyrosine metabolism | map00350 | 5.6997E-02 | osa00350 | 5.8000E-01 |
| Carbon fixation in photosynthetic organisms | map00710 | 6.7894E-02 | osa00710 | 2.3200E-02 |
| Fatty acid biosynthesis | map00061 | 9.3531E-02 | osa00061 | 2.2700E-01 |
| Purine metabolism | map00230 | 9.6256E-02 | osa00230 | 9.4900E-01 |
| Cysteine and methionine metabolism | map00270 | 1.1285E-01 | osa00270 | 3.6700E-01 |
| beta-Alanine metabolism | map00410 | 1.1973E-01 | osa00410 | 5.1800E-01 |
| Pentose phosphate pathway | map00030 | 1.3869E-01 | osa00030 | 3.1900E-01 |
| Biosynthesis of unsaturated fatty acids | map01040 | 1.6000E-01 | osa01040 | 1.7700E-01 |
| Monobactam biosynthesis | map00261 | 1.6489E-01 | osa00261 | 2.2200E-01 |
| Arginine and proline metabolism | map00330 | 1.7846E-01 | osa00330 | 5.9900E-01 |
| alpha-Linolenic acid metabolism | map00592 | 1.9873E-01 | osa00592 | 2.1800E-01 |
| Oxidative phosphorylation | map00190 | 2.6142E-01 | osa00190 | 6.0100E-01 |
| Fructose and mannose metabolism | map00051 | 2.6829E-01 | osa00051 | 7.3500E-01 |
| Linoleic acid metabolism | map00591 | 2.6991E-01 | osa00591 | 2.8900E-01 |
| Nitrogen metabolism | map00910 | 3.0232E-01 | osa00910 | 4.7200E-01 |
| Phenylpropanoid biosynthesis | map00940 | 3.6540E-01 | osa00940 | 3.9400E-04 |
| Vitamin B6 metabolism | map00750 | 4.1210E-01 | osa00750 | 1.6600E-01 |
| Glycolysis/Gluconeogenesis | map00010 | 4.4476E-01 | osa00010 | 7.7100E-01 |
| Thiamine metabolism | map00730 | 4.4476E-01 | osa00730 | 2.0400E-01 |
| Sulfur metabolism | map00920 | 4.6553E-01 | osa00920 | 4.2800E-02 |
| Starch and sucrose metabolism | map00500 | 5.0464E-01 | osa00500 | 7.0900E-04 |
| Amino sugar and nucleotide sugar metabolism | map00520 | 6.0660E-01 | osa00520 | 2.3400E-01 |
| Fatty acid degradation | map00071 | 6.1384E-01 | osa00071 | 2.3500E-01 |
| Pentose and glucuronate interconversions | map00040 | 6.4914E-01 | osa00040 | 6.0800E-01 |
| Biotin metabolism | map00780 | 7.5902E-01 | osa00780 | 7.3100E-02 |
| Ubiquinone and other terpenoid-quinone biosynthesis | map00130 | 8.2816E-01 | osa00130 | 5.5000E-01 |
| Porphyrin and chlorophyll metabolism | map00860 | 9.3529E-01 | osa00860 | 5.7100E-01 |

that $H_2O_2$ can activate multiple acclamatory responses that reinforce resistance to various abiotic stressors (Li et al., 2019; Hasanuzzaman et al., 2022). Excessive ROS can lead to oxidative stress (Hasanuzzaman et al., 2020). Oxidative stress is the driving force for evoking stress responses. MDA content is one of the cytotoxic chemicals that determine oxidative damage to cell membranes and the final product of lipid peroxidation (Hossen et al., 2022). Electrolyte leakage (EL) is a sign of stress response in plant cells, which is usually accompanied by the accumulation of reactive oxygen species (ROS) and often leads to programmed cell death (PCD) (Demidchik et al., 2014). Soluble sugar is considered to be an important molecule for sensing the concentration of ROS in cells, and involves in the response of plants to oxidative stress (Afzal et al., 2021). Studies have reported that $H_2O_2$, EL, and MDA increased in rice under salt stress. Salt stress increased EL in soybean by 69%, while $H_2O_2$ and MDA contents increased by 75 and 56%, respectively (Alharby et al., 2021). Similar results were observed in our study (**Figure 2**). In the SP2 group, these four oxidative stress markers all increased (**Figures 2A–D**), which indicated that the F2 generation plants still retained the stress response generated by space flight.

Antioxidative defense systems protect plants from oxidative damage under stress by detoxifying ROS and maintaining the balance of ROS production under abiotic stress (Xie et al., 2019). Antioxidative enzymes are an important part of plant antioxidant system. It was reported that the activities of CAT and APX in plants increased with the intensification of salt stress (Abulfaraj and Jalal, 2021). Meanwhile, salt stress can lead to the increase of SOD activity and APX activity in rice (Hossen et al., 2022). The activities of SOD, CAT, and APX were increased in soybean plants treated with 7.46 dS $m^{-1}$ NaCl (Soliman et al., 2020). Soybean plants treated with 100 mM NaCl for 25 days increased SOD,

**TABLE 2 |** Proteins and metabolites involved in common pathways at TS.

| Pathway name | Metabolomics | | Proteomics | |
|---|---|---|---|---|
| | Pathway ID | *P*-value | Pathway ID | *P*-value |
| Citrate cycle (TCA cycle) | map00020 | 4.69E-04 | osa00020 | 2.47E-01 |
| Galactose metabolism | map00052 | 1.62E-03 | osa00052 | 6.99E-01 |
| Linoleic acid metabolism | map00591 | 1.72E-03 | osa00591 | 4.25E-02 |
| Biosynthesis of unsaturated fatty acids | map01040 | 5.11E-03 | osa01040 | 1.74E-01 |
| Glyoxylate and dicarboxylate metabolism | map00630 | 5.24E-03 | osa00630 | 7.60E-01 |
| Phenylpropanoid biosynthesis | map00940 | 4.03E-02 | osa00940 | 5.81E-03 |
| Glycolysis/Gluconeogenesis | map00010 | 4.55E-02 | osa00010 | 5.19E-01 |
| Pentose phosphate pathway | map00030 | 5.85E-02 | osa00030 | 3.14E-01 |
| Flavone and flavonol biosynthesis | map00944 | 5.99E-02 | osa00944 | 6.52E-02 |
| Cyanoamino acid metabolism | map00460 | 9.08E-02 | osa00460 | 1.82E-01 |
| Glycine, serine and threonine metabolism | map00260 | 1.09E-01 | osa00260 | 3.63E-01 |
| Glycerophospholipid metabolism | map00564 | 1.16E-01 | osa00564 | 8.00E-01 |
| Fructose and mannose metabolism | map00051 | 1.23E-01 | osa00051 | 7.31E-01 |
| Plant hormone signal transduction | map04075 | 1.28E-01 | osa04075 | 9.86E-01 |
| Phenylalanine metabolism | map00360 | 1.46E-01 | osa00360 | 5.46E-01 |
| Glutathione metabolism | map00480 | 1.52E-01 | osa00480 | 2.20E-01 |
| Flavonoid biosynthesis | map00941 | 2.03E-01 | osa00941 | 6.80E-03 |
| Riboflavin metabolism | map00740 | 2.04E-01 | osa00740 | 1.83E-01 |
| Sphingolipid metabolism | map00600 | 2.48E-01 | osa00600 | 3.91E-01 |
| Purine metabolism | map00230 | 2.91E-01 | osa00230 | 9.48E-01 |
| Lysine biosynthesis | map00300 | 3.29E-01 | osa00300 | 2.87E-01 |
| Starch and sucrose metabolism | map00500 | 3.44E-01 | osa00500 | 6.90E-01 |
| Amino sugar and nucleotide sugar metabolism | map00520 | 3.46E-01 | osa00520 | 2.28E-01 |
| alpha-Linolenic acid metabolism | map00592 | 3.51E-01 | osa00592 | 5.46E-02 |
| Ascorbate and aldarate metabolism | map00053 | 4.14E-01 | osa00053 | 1.98E-01 |
| Pentose and glucuronate interconversions | map00040 | 4.67E-01 | osa00040 | 6.04E-01 |
| Monobactam biosynthesis | map00261 | 4.77E-01 | osa00261 | 2.19E-01 |
| Fatty acid biosynthesis | map00061 | 4.85E-01 | osa00061 | 2.23E-01 |
| Carbon fixation in photosynthetic organisms | map00710 | 4.95E-01 | osa00710 | 7.79E-02 |
| Arachidonic acid metabolism | map00590 | 5.77E-01 | osa00590 | 2.02E-01 |
| Nitrogen metabolism | map00910 | 6.23E-01 | osa00910 | 1.27E-01 |
| Sulfur metabolism | map00920 | 7.05E-01 | osa00920 | 5.46E-01 |
| Porphyrin and chlorophyll metabolism | map00860 | 8.93E-01 | osa00860 | 5.66E-01 |
| Fatty acid degradation | map00071 | 9.52E-01 | osa00071 | 6.04E-01 |
| Cysteine and methionine metabolism | map00270 | 9.77E-01 | osa00270 | 8.88E-01 |

CAT, and APX activities by 31, 16, and 20, respectively (Taha et al., 2020). In this study, we evaluated changes in the activities of four antioxidant enzymes. The results showed that the activities of the four antioxidant enzymes in the SP2 group changed during TLS and TS, but there were differences in the changes in enzyme activities during the two periods (**Figures 2E–H**), which indicated that there may be differences in their ways to eliminate ROS. This also shows that oxidative stress exists in SP2 group again.

Searched for DAPs related to ROS clearance (**Table 3**). In TLS, 13 proteins related to ROS clearance were differentially expressed, and most of them were Peroxidase family proteins. It is worth noting that the expression abundance of these proteins increased, indicating that the redox state of the SP2 group plants was severely damaged at this time. In TS, there were nine proteins that were related to ROS clearance. Two of these nine proteins were Glutathione S-transferase family proteins, and the rest were Peroxidase family proteins. We noticed that the enzyme activities of APX and CAT have changed, while the protein abundance has not changed significantly, which indicated that the changes in the activity of these two enzymes may not be affected by the protein abundance. The changes in the abundance of these proteins further indicated that the stress response caused by space flight was still retained in the F2 generation of plants, and continued to the F2 generation of TS. There were also some flavonoids accumulated in this study, which would be analyzed in the follow-up discussion.

What's interesting is why the F2 generation plants can retain the stress response caused by space flight? It is well known that the most serious damage to the body caused by space flight was DNA damage (Moreno-Villanueva et al.,

**TABLE 3 |** Changes in protein abundance associated with ROS scavenging in the three-leaf stage (TLS) and the tillering stage (TS) as identified by iTRAQ.

| Uniprot ID | Description | Fold change | | Adjusted *P* value | |
|---|---|---|---|---|---|
| | | TLS | TS | TLS | TS |
| A2YC52 | Peroxidase | 2.31 | | 2.2903E-03 | |
| Q5U1T0 | Class III peroxidase 13 | 1.88 | | 9.2736E-04 | |
| Q5U1N4 | Class III peroxidase 59 | 1.88 | | 3.2026E-04 | |
| Q7F1J9 | L-ascorbate peroxidase | 1.6 | | 3.5152E-03 | |
| A2WNR8 | Peroxidase | 1.51 | | 6.1473E-05 | |
| A0A0P0XR31 | Peroxidase | 1.46 | | 2.8026E-02 | |
| Q94DM2 | Class III peroxidase 22 | 1.42 | | 1.0709E-02 | |
| Q6AVZ8 | Class III peroxidase 65 | 1.41 | | 3.6169E-02 | |
| A2XTH3 | Peroxidase | 1.38 | 1.44 | 4.0609E-02 | 2.2156E-03 |
| Q9ST82 | Peroxidase | 1.35 | | 3.9481E-02 | |
| Q6EUS1 | Class III peroxidase 27 | 1.32 | | 4.0609E-02 | |
| Q0JK76 | Glutathione Transferase | 1.32 | | 4.0609E-02 | |
| A2XGP6 | Superoxide13 dismutase (Cu-Zn) 1 | 1.27 | | 7.6207E-04 | |
| B1NEV3 | Peroxidase | | 1.3 | | 1.0556E-03 |
| Q5U1Q2 | Class III peroxidase 41 | | 1.26 | | 1.0564E-02 |
| Q5U1Q4 | Class III peroxidase 39 | | 1.25 | | 4.1361E-03 |
| Q6K4J4 | Class III peroxidase 122 | | 0.76 | | 6.3815E-03 |
| O22438 | Class III peroxidase 21 | | 0.73 | | 1.0191E-02 |
| Q93WY5 | Glutathione S-transferase | | 0.72 | | 9.7683E-03 |
| Q6ER51 | Class III peroxidase 30 | | 0.68 | | 4.5544E-03 |
| Q6QN17 | Glutathione S-transferase GSTU35 | | 0.34 | | 7.1210E-06 |

*Blue represents down-regulated.*
*Yellow represents up-regulated.*

**TABLE 4 |** Differentially expressed proteins and metabolites related to DNA damage.

| Uniprot ID | Description | Fold change | | Adjusted *P* value | |
|---|---|---|---|---|---|
| | | TLS | TS | TLS | TS |
| **Protein** | | | | | |
| B8BHL4 | DNA ligase | 0.76 | | 4.0332E-02 | |
| Q852K3 | Replication factor C subunit 5 | 0.71 | | 2.2429E-02 | |
| B8AKX8 | Cullin 4 | 0.44 | | 2.5086E-04 | |
| Q9FTU2 | Replication factor A1 | | 0.77 | | 1.0564E-02 |
| Q0J8Y6 | DNA mismatch repair protein MutS2 | | 0.75 | | 1.9912E-04 |
| **Metabolites** | | | | | |
| | Deoxyguanosine | | 1.32 | | 1.1766E-02 |

*Blue represents down-regulated.*
*Yellow represents up-regulated.*

2017), the genomic instability caused by which can lead to heritable environmental stress (Boyko and Kovalchuk, 2011). Therefore, we speculated that the stress response of F2 generation plants to the space environment was caused by DNA damage and genome instability. To confirm this conjecture, we explored the changes in proteins and metabolites related to DNA damage and repair (**Table 4**). In TLS, there were three different proteins involved in the DNA repair process. Q852K3 (Replication factor C subunit 5) was a subunit of Replication factor C complex, which was involved in DNA repair, DNA replication and checkpoint control in cell cycle progression. The biological function of ScRFC5 in rice DNA repair process was still unclear, but it was only necessary for embryo development and mitosis at the cell stage (Chen et al., 2018). However, in yeast cells, ScRFC5 was necessary for DNA damage checkpoint control (Naiki et al., 2000). The lack of B8BHL4 (DNA ligase) increased the sensitivity of plants to ionizing radiation, and it participated in T-DNA integration and regulation mechanisms in DNA repair (Friesner and Britt, 2003). B8AKX8 (Cullin 4) was an important member of the Cullin family. It acted as a scaffold in the CUL4-DDB1-based ubiquitin ligase and regulated cell proliferation, DNA repair and genome integrity through key regulatory factors of ubiquitination (Lee and Zhou, 2007). Interestingly,

**TABLE 5 |** Differentially expressed proteins and metabolites related to saccharide metabolism.

| Uniprot ID | Description | Fold change | | Adjusted *P* value | |
|---|---|---|---|---|---|
| | | TLS | TS | TLS | TS |
| **Protein** | | | | | |
| Q10CU9 | Glycosyl hydrolase family 3 protein | 1.41 | | 1.2836E-03 | |
| Q10LP5 | Sucrose synthase 4 | 1.33 | | 3.4614E-02 | |
| Q84YK7 | Beta-glucosidase 27 | 1.3 | | 3.0213E-02 | |
| A2YSI0 | Glucan endo-1,3-beta-D-glucosidase | 1.3 | | 4.0609E-02 | |
| D0TZC9 | Glucose-1-phosphate adenylyltransferase | 0.74 | | 3.3343E-02 | |
| Q8GTK0 | Starch synthase | 0.63 | | 3.4245E-02 | |
| A2Y7W1 | Glucose-1-phosphate adenylyltransferase | 0.58 | | 2.2508E-03 | |
| D0U0Q5 | Soluble starch synthase II-3 | 0.5 | | 7.6898E-04 | |
| B8AXU3 | Beta-glucosidase 19 | 0.73 | 0.67 | 2.6617E-02 | 1.0564E-02 |
| **Metabolites** | | | | | |
| | D-Arabitol | 4.34 | | 1.6250E-05 | |
| | D-Fructose | 3.06 | 3.7 | 6.4945E-06 | 1.6293E-05 |
| | Alpha-D-Glucose | 2.27 | 2.93 | 1.6199E-06 | 7.6749E-04 |
| | D-Tagatose | 2.16 | 2.41 | 3.4048E-07 | 2.5078E-03 |
| | D-Mannose | 2.1 | 2.47 | 1.9112E-06 | 1.1283E-05 |
| | Maltotriose | 2.05 | | 6.1085E-04 | |
| | Isomaltose | | 1.5 | | 5.5054E-05 |
| | D-Ribose | | 1.9 | | 1.4766E-03 |
| | D-Glucose 1-phosphate | 1.711 | 1.25 | 1.0986E-03 | 2.4599E-04 |
| | Sucrose | 0.69 | 1.31 | 1.4117E-06 | 4.4672E-06 |
| | L-Sorbose | 0.69 | | 1.8522E-07 | |
| | Raffinose | 0.39 | | 4.0648E-05 | |

*Blue represents down-regulated.*
*Yellow represents up-regulated.*

the expression abundance of these three proteins related to DNA damage repair was all down-regulated in the TLS in the SP2 group, indicating that the DNA damage response existed in the plant and may affect the development of the plant simultaneously. In TS, we found two proteins and one metabolite related to DNA damage repair in the SP2 group. Q9FTU2 (replication factor A1) is an important protein for double-strand break repair in the process of meiotic homologous recombination (Liu et al., 2013), and help mediate genome stability and transcriptional gene silencing (Liu et al., 2010). Q0J8Y6 (DNA mismatch repair protein MutS2) is a member of the MutS family of proteins. This family of proteins participate in the process of DNA mismatch repair and is a natural candidate for maintaining a low mutation rate in the plant cell genome (Wu et al., 2020). Generally, deoxyguanosine can effectively inhibit the division of plant cells (Brulfert et al., 1974), which can be transformed into a DNA damage marker 8-oxo-2′-deoxyguanosine *in vivo*. Therefore, the increase in the content of the compound can reflect the damage of DNA. We noticed that the abundance of proteins involved in the DNA repair process in the SP2 group of plants decreased during TLS and TS, which was unfavorable for maintaining the genomic stability of the dimensional plant, indicating that the space flight reduced the genomic stability of the F2 generation plant and retained the stress response brought about by the space flight.

## Saccharide Metabolism Was Involved in the Response of F2 Generation Plants to the Spaceflight

Sugar plays a central role in maintaining plant cell structure and metabolism, and participates in the response of plant cells to abiotic stress (Kumari and Parida, 2018). Under different abiotic stresses, there are almost no consensus on the changes of specific carbohydrates in different species, which implies that there are different metabolic rearrangements in different species (Kumari and Parida, 2018). In addition, sugar can also act as a signal molecule for nutrients and metabolites in plant cells, and activate or interact with specific plant hormone transmission pathways, leading to changes in gene expression and protein abundance (Couée et al., 2006).

Monosaccharides are the basic structural unit of carbohydrate molecules and an important part of soluble sugars in plant cells. Compared with CK, the content of D-glucose in the SP2 group increased approximately 2.27 times and 2.93 times during the TLS and TS stages, respectively (**Table 5**). The accumulation of D-glucose can reduce the damage of salt stress to wheat seedlings and improve its photosynthetic capacity (Wang et al., 2019). D-Fructose was accumulated 3.06 and 3.70 times in the TLS and TS stages, respectively (**Table 5**). Fructose participated in the antioxidant protection of plants with a high ability to remove ROS, which was twice that of glucose (Bogdanović et al., 2008).

**TABLE 6 |** Differentially expressed proteins and metabolites related to energy metabolism.

| Uniprot ID | Description | Fold change | | Adjusted *P*-value | |
|---|---|---|---|---|---|
| | | TLS | TS | TLS | TS |
| **Protein** | | | | | |
| Q7 × 8V5 | Acetyl-coenzyme A synthetase | 1.27 | | 3.9493E-02 | |
| P0C2Z4 | ATP synthase subunit alpha | 0.77 | | 4.0609E-02 | |
| B8AR95 | Cytochrome c oxidase subunit 6b | 0.65 | | 1.6451E-02 | |
| B8AY35 | Fructose-bisphosphate aldolase | | 1.59 | | 3.6623E-03 |
| B9FW35 | QCR7 | | 1.51 | | 1.2327E-03 |
| Q5ZAS4 | ATPase beta subunit | | 1.43 | | 3.3091E-03 |
| Q0DIC9 | QCR9 | | 1.31 | | 3.4060E-03 |
| A2WXB2 | Fructose-1,6-bisphosphatase | 0.67 | 1.29 | 2.0885E-02 | 3.2994E-03 |
| A6N0U5 | Ndufa8 | 0.62 | 1.26 | 7.3520E-03 | 1.0564E-02 |
| Q75HQ5 | ATP synthase subunit beta | | 1.26 | | 6.3815E-03 |
| A2XF65 | Atpb | | 1.26 | | 9.8779E-03 |
| P0C2Y5 | ATP synthase subunit a | | 1.25 | | 1.0564E-02 |
| Q9S827 | SDHB1 | | 0.79 | | 1.0564E-02 |
| Q7XBT1 | NDUFA12 | | 0.69 | | 2.0238E-04 |
| B8AAV9 | Phosphoglycerate kinase | | 0.68 | | 5.3781E-03 |
| Q6ZJ19 | Ndufs5 | 0.36 | 0.67 | 2.7798E-04 | 4.1503E-04 |
| B7EBN1 | Mitochondrial carrier protein | | 0.49 | | 9.7310E-04 |
| Q9FP98 | TOM7-1 | | 0.11 | | 2.2796E-07 |
| **Metabolites** | | | | | |
| | *cis*-Aconitate | 1.71 | 5.35 | 3.2585E-04 | 4.5697E-10 |
| | Citrate | | 1.76 | | 2.1810E-06 |
| | Pyruvic acid | | 1.55 | | 6.1532E-07 |
| | Isocitrate acid | | 1.23 | | 4.8333E-02 |
| | Succinate | 0.58 | 1.3 | 1.0859E-07 | 4.3228E-02 |
| | Alpha-ketoglutarate | 0.44 | | 4.7704E-07 | |
| | L-Malic acid | 0.15 | 2.09 | 7.5190E-10 | 6.7041E-04 |

*Blue represents down-regulated.*
*Yellow represents up-regulated.*

At the same time, studies have shown that the accumulation of fructose under abiotic stress was also related to the synthesis of erythrose-4-P, which was a substrate for the synthesis of lignin and phenolic compounds (Hilal et al., 2007). Moreover, the D-Tagatose and D-Mannose were also accumulated in the TLS and TS stages (**Table 5**). Studies have shown that D-Tagatose can inhibit the metabolism of D-Mannose (Mochizuki et al., 2020), and both sugars responded to abiotic stresses of plants (Shahbazy et al., 2020, (Nishizawa et al., 2008; Li et al., 2017; Kumari and Parida, 2018). D-Mannose reduced the damage under abiotic stress by removing ROS in plants (Nishizawa et al., 2008). This study has also found changes in other monosaccharides, such as L-Sorbose and D-Ribose (**Table 5**), which only changed during the TLS and TS stages, respectively.

In addition to monosaccharides, soluble sugars also included disaccharides (sucrose, trehalose), raffinose family oligosaccharides (RFO) and fructan, which were mainly involved in stress responses in plants (Keunen et al., 2013). Sucrose is the main product of plant photosynthesis and the basic form of sugar storage in plants (Hilal et al., 2007; Rosa et al., 2009), which is composed of glucose and fructose. Studies have shown that sucrose was closely related to plant growth, development, signal transmission and adversity adaptation (Hilal et al., 2007; Chen et al., 2019). The sucrose content in the SP2 group decreased by 31% during TLS, while increased by 1.31 times during TS, which may be due to the different hydrolysis rates of sucrose in different growth and development stages. Certainly, this also implied that there were differences in rice energy metabolism between the two growth and development stages. The results of this study also showed that the content of Raffinose and Maltotriose changed significantly during the TLS phase (**Table 5**). Mannose and raffinose can protect plant cells from oxidative damage caused by various stress conditions, which has been determined that they have the ability to eliminate ROS (Nishizawa et al., 2008).

As a major polysaccharide, starch is the most important supplier of various carbohydrates in plants. It regulates plant growth by producing energy through respiratory metabolism, and can be hydrolyzed into soluble sugars by amylase (Chen et al., 2019). Previous studies have shown that space flight affects plant starch metabolism (Guisinger and Kiss, 1999). In this study, three types (Soluble starch synthase II-3, Starch synthase, and Glucose-1-phosphate adenylyl transferase) were involved in the reduction of protein expression in starch synthesis (**Table 5**), which indicated the process of starch synthesis in TLS Be suppressed. In addition, Soluble starch synthase II-3 determines the structure type of starch in the process of starch synthesis. The

**TABLE 7 |** Differentially expressed proteins and metabolites related to amino acid metabolism.

| Uniprot ID | Description | Fold change | | Adjusted *P*-value | |
|---|---|---|---|---|---|
| | | TLS | TS | TLS | TS |
| **Protein** | | | | | |
| Q8H7N0 | Alcohol dehydrogenase | 1.44 | | 1.0709E-02 | |
| Q6K7D6 | Lysine-ketoglutarate reductase | 1.37 | | 6.2624E-03 | |
| Q01IY5 | Amine oxidase | 1.32 | | 1.6010E-02 | |
| Q5VND2 | Cysteine synthase | 1.32 | | 4.0609E-02 | |
| H9N066 | Betaine aldehyde dehydrogenase | 1.3 | | 3.0213E-02 | |
| A0A0P0YAL3 | Phenylalanine ammonia-lyase | 1.3 | | 4.0609E-02 | |
| A3C0A7 | Sarcosine oxidase | 1.29 | | 4.0609E-02 | |
| B9FLJ1 | Glutamine amido transferase type-2 | 1.26 | | 4.0609E-02 | |
| Q0INQ6 | Serine hydroxymethyl transferase | 0.78 | | 1.4128E-02 | |
| B9FFK4 | Anthranilate synthase component | 0.78 | | 4.0609E-02 | |
| B9FHA5 | Serine *O*-acetyltransferase | 0.77 | | 4.0609E-02 | |
| Q84PB1 | Phosphoribosyl anthranilate transferase | 0.77 | | 9.2369E-05 | |
| Q6K6Q1 | Phenylalanine ammonia-lyase | | 1.35 | | 5.9680E-03 |
| Q0JFF8 | Hydroxy pyruvate reductase HPR3 | | 1.27 | | 4.8507E-03 |
| B8AGS8 | Acetylornithine deacetylase | | 1.27 | | 3.2391E-03 |
| A0A0P0W7B4 | Amidase At4g34880 | | 1.26 | | 9.8779E-03 |
| B9G099 | Indole-3-glycerol phosphate synthase | | 0.79 | | 3.3091E-03 |
| Q7XKW5 | L-threonine aldolase 1 | 0.75 | 0.77 | 1.1910E-02 | 5.6268E-03 |
| A0A0P0WFZ1 | Asparaginase | | 0.76 | | 6.8962E-03 |
| B7E3L9 | Dihydropicolinate reductase | | 0.28 | | 2.3522E-05 |
| **Metabolites** | | | | | |
| | Leucine | 3.3 | | 3.7182E-06 | |
| | Tryptophan | 2.65 | | 9.9516E-08 | |
| | L-isoleucine | 2.3 | 1.78 | 1.0986E-03 | 7.7095E-05 |
| | Phenylalanine | 2.08 | 1.49 | 2.0535E-05 | 2.0198E-04 |
| | *N*-Acetyl-L-leucine | 1.24 | 3.2 | 6.7834E-04 | 7.2787E-05 |
| | L-Valine | | 2.14 | | 4.7120E-02 |
| | L-Arginine | | 1.95 | | 4.1766E-02 |
| | 4-Hydroxy-L-glutamic acid | | 1.86 | | 2.0251E-02 |
| | Glutamate | 0.37 | 1.29 | 1.0079E-07 | 1.0862E-02 |
| | L-Pyroglutamic acid | 0.69 | 0.63 | 5.6325E-06 | 1.5291E-04 |
| | Aspartate | 0.36 | | 3.25E-06 | |

*Blue represents down-regulated.*
*Yellow represents up-regulated.*

absence of Soluble starch synthase II-3 would make the starch granules smaller and rounder. Therefore, our results suggested that in F2 generation plants, there may be difference in starch structure between the spaceflight group and the control group. The above results also suggested that in the SP2 group, plants may adjust sugar metabolism to eliminate excess ROS, and may induce the reconstruction of other metabolic pathways through sugar metabolism to adapt to the effects of spaceflight.

## Energy Metabolism Was Involved in the Response of F2 Generation Plants to the Spaceflight

Energy metabolism is one of the most important regulators for plants to adapt to abiotic stress (Gharechahi et al., 2015). We previously reported changes in the expression abundance of proteins involved in energy metabolism in contemporary rice plants after space flight (Deyong et al., 2020). Here we examined the changes in related proteins and metabolites during glycolysis, TCA cycle, and oxidative phosphorylation (**Table 6**). It has been proposed that glycolysis played an important role in the response of plants to abiotic stress. Three proteins involved in glycolysis were differentially abundant in the SP2 group. The expression abundance of Fructose-1,6-bisphosphatase changed significantly in both periods. Studies have shown that Fructose-1,6-bisphosphatase can regulate plant tolerance to abiotic stress (Chee Hark et al., 1997). Inhibiting the expression of this protein would result in reduced plant growth, dwarf phenotype and delayed flowering, as well as changes in important metabolites such as amino acids, sugars and organic acids (Rojas-González et al., 2015). Therefore, the changes in amino acid metabolism and sugar metabolism in this study may be related to the

**TABLE 8 |** Differentially expressed proteins and metabolites related to phenylpropanoid biosynthesis pathway and flavonoid synthesis metabolism.

| Uniprot ID | Description | Fold change | | Adjusted *P*-value | |
|---|---|---|---|---|---|
| | | TLS | TS | TLS | TS |
| Q10CU9 | Glycosyl hydrolase family 3 | 1.41 | | 1.2836E-03 | |
| A0A0P0YAL3 | Phenylalanine ammonia-lyase | 1.3 | 1.35 | 4.0609E-02 | 4.5000E-02 |
| Q84YK7 | Beta-glucosidase 27 | 1.3 | | 3.0213E-02 | |
| A2ZEX7 | Chalcone synthase 1 | 0.77 | 1.6 | 1.8876E-02 | 6.4225E-04 |
| A2WS12 | Aldehyde dehydrogenase | | 1.24 | | 1.0564E-02 |
| A2XPI4 | CYP 450 93A2 | | 1.44 | | 2.5489E-03 |
| Q7G602 | Flavonoid 3′-hydroxylase | | 1.31 | | 2.6334E-03 |
| A2XNF0 | Chalcone-flavonone isomerase | | 1.3 | | 6.3815E-03 |
| A3C6F1 | CYP 450 A81 | | 0.68 | | 6.9292E-03 |
| B8AXU3 | Beta-glucosidase 19 | 0.73 | 0.67 | 2.6617E-02 | 1.0564E-02 |
| **Metabolites** | | | | | |
| | 4-Hydroxycinnamic acid | 2.4 | 0.43 | 3.1061E-09 | 1.1828E-03 |
| | Ferulic acid | 2.16 | 1.59 | 2.3178E-08 | 3.2950E-04 |
| | APIIN | 1.57 | 5.2 | 6.1736E-07 | 1.1729E-04 |
| | Malvidin 3-*O*-glucoside cation | 1.22 | 1.5 | 7.4074E-04 | 2.0241E-02 |
| | Peonidin 3-galactoside cation | 1.2 | 1.5 | 2.1901E-04 | 2.4146E-02 |
| | Chlorogenic acid | | 4.4 | | 6.6984E-06 |
| | p-Coumaryl alcohol | | 2.91 | | 1.1109E-04 |
| | Oenin | | 2.61 | | 1.1157E-02 |
| | Kaempferol | 0.25 | 2.34 | 7.1541E-09 | 3.6814E-02 |
| | Hesperetin 7-*O*-neohesperidoside | 0.19 | | 1.2644E-09 | |
| | Rutin | | 2.01 | | 4.6557E-02 |
| | Orobol | | 1.84 | | 1.5832E-03 |
| | Sinapate | | 1.62 | | 4.1766E-02 |
| | Kuromanin | | 1.34 | | 1.1766E-02 |
| | Formononetin | | 1.33 | | 1.4771E-03 |

*Blue represents down-regulated.*
*Yellow represents up-regulated.*

differential expression of the protein. Fructose-bisphosphate aldolase was involved in the production of D-glyceraldehyde 3-phosphate during glycolysis. And the expression of this protein was induced by salt, drought, heat and other abiotic stresses, and affects plant soluble sugar content, stem straightness, dry weight and seed size (Cai et al., 2018). In plants, Phosphoglycerate kinase converts 1,3-bisphosphoglycerate to 3-phosphoglycerate during glycolysis, and the reduction in the expression of this enzyme would also lead to reduced plant growth, photosynthetic capacity and starch content (Rosa-Téllez et al., 2018). In fact, these three enzymes participated not only in the glycolysis process, but also in the photosynthetic Calvin-Benson cycle reaction, thus having an impact on photosynthesis (Rojas-González et al., 2015; Cai et al., 2018; Rosa-Téllez et al., 2018).

Recently, the TCA cycle is considered as a pressure sensor for plants, and changing the TCA cycle has been an inevitable response of plants under abiotic stress (Das et al., 2019). Under abiotic stress conditions, the TCA cycle was an important protective system (Fernie et al., 2004), and the increase in the TCA cycle helped to improve plant tolerance to abiotic stress (Zhong et al., 2016). In TLS, the SP2 group detected significant changes in four metabolites involved in the TCA cycle. Except for *cis*-Aconitate, the content of the other three metabolites decreased, indicating that the TCA cycle was inhibited at this time. However, during TS, there were six metabolites involved in the TCA cycle in the SP2 group, and the content of these six metabolites all increased, which showed that the TCA cycle was activated. These results suggested that the F2 generation plants have different responses to space flight stress at different growth stages. Many abiotic stresses led to the destruction of the mitochondrial electron transfer chain, which led to the reduction of ATP and the production of ROS.

It is well known that the reducing equivalent produced by the activity of the TCA cycle was used by the mitochondrial electron transport chain to promote the synthesis of ATP (Fernie et al., 2004). This study revealed that some proteins participated in the ETC (**Table 6**), NDUFS5, NDUFA12, NDFUA8, Cytochrome b-c1 complex subunit 9 (QCR9), Cytochrome b-c1 complex subunit 7 (QCR7), ATP synthase subunit beta, ATP synthase subunit alpha and ATP b. Our previous research has reported that NDUFS5 and NDUFA12 changed significantly in the F1 generation after spaceflight (Deyong et al., 2020). NDUFS5 and NDUFS12 have no contribution to the activity of complex I, but were related to their electron transfer efficiency (Loeffen et al., 1999). Obviously, the spaceflight caused a decrease in the electron

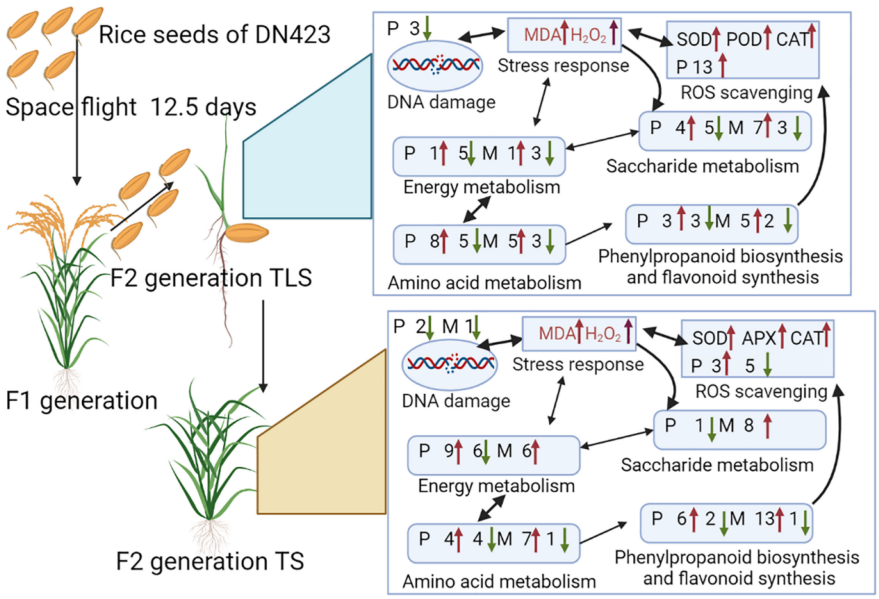



**FIGURE 11 |** Mechanisms of memory in response to space flight stress in F2 generation rice. *P* represent proteins; M represent metabolites; numbers indicate the amount of proteins and metabolites; red and green arrows represent increase and decrease, respectively.

transfer efficiency in rice mitochondrial complex I, which this effect continued from the F1 generation to the F2 generation.

QCR9, QCR7, TOM7-1, Mitochondrial carrier protein were necessary during the assembly process of mitochondrial complex III. The decrease in protein abundance of TOM7-1, Mitochondrial carrier protein would cause the assembly engineering of mitochondrial complex III to be hindered (Deyong et al., 2020). The changes of QCR9, QCR7, TOM7-1, and Mitochondrial carrier protein were also detected in the F1 generation plants after spaceflight (Deyong et al., 2020). This indicated that plant mitochondrial complex III may be a more sensitive part of space flight, and complex III was also one of the main sources of ROS. Therefore, the increase in rice ROS caused by space flight may be due to the dysfunction of complex III. Furthermore, we found that the expression abundance of ATP synthase changes. The abundance of mitochondrial electron transport chain-related proteins in the SP2 group decreased in TLS while increased in TS, which indicated that the rate of energy metabolism decreased in TLS and the respiratory rate increased in TS.

## Amino Acid Metabolism Was Involved in the Response of F2 Generation Plants to the Spaceflight

Previous studies have shown that the abundance of proteins related to amino acid metabolism in F2 generation plants changed after spaceflight (Ma et al., 2007; Wang et al., 2008). In higher plants, amino acids accumulated in response to various stresses and had multiple functions in plant growth (Less and Galili, 2008). Moreover, amino acids were also necessary for protein synthesis and provided necessary intermediate products for many metabolic reactions (Pratelli and Pilot, 2014).

In TLS, five types of amino acids were accumulated and 3 type of amino acids were reduced in SP2 group, and some proteins involved in amino acid metabolism also changed significantly (**Table 7**). In TS, seven types of amino acids were accumulated and one type of amino acids was reduced in SP2 group. Moreover, seven proteins involved in amino acid metabolism were altered in abundance during TS (**Table 7**). What's interesting was that our results showed that the changes in amino acid metabolism-related proteins were inconsistent with the changes in amino acid content. Therefore, it can be speculated that the accumulation of amino acids in our study was caused protein hydrolysis instead of biosynthesis (Fan et al., 2018). In addition, plants can also activate the accumulation of amino acids in response to abiotic stress through sucrose signals (Jia et al., 2019). L-aspartic acid, leucine, isoleucine, valine, glutamic acid, and phenylalanine can generate TCA cycle intermediate products through oxidation. At the same time, the electrons generated during the oxidation process are directly sent to ETC. This would help the production of ATP under abiotic stress (Galili, 2011; Hildebrandt et al., 2015). In SP2, the changes in the content of these amino acids may be related to their participation in energy metabolism (**Table 7**). Moreover, these amino acids may also act as potential signal molecules themselves, or act as precursors for the synthesis of other secondary metabolites and plant hormones to trigger abiotic stress signals. For example, phenylalanine (**Figure 9**) can be used as a precursor of plant synthetic alkaloids and flavonoids (ROS scavengers) (Kumari and Parida, 2018). Therefore, the accumulation of phenylalanine can ensure the production of antioxidants in SP2 plants to improve their adaptation to

spaceflight. Glutamate can be converted into alpha-ketoglutarate. In F2 plants, alpha-ketoglutarate was increased in TLS, while Glutamate content was reduced, which indicated that the increase in alpha-ketoglutarate content in SP2 plants was caused by the degradation of glutamate.

### Phenylpropanoid Biosynthesis Pathway and Flavonoid Synthesis Metabolism Was Involved in the Response of F2 Generation Plants to the Spaceflight

It is well known that phenylpropane metabolism is the upstream reaction of flavonoid and lignin biosynthesis, and phenylalanine is an important precursor of phenylpropane metabolism (Cramer et al., 2013). In this study, phenylalanine was significantly accumulated, which indicated that the phenylpropane metabolic pathway in SP2 plants was affected. Current studies have shown that the phenylpropane biosynthetic pathway was activated under various abiotic stress conditions (drought, heavy metals, salinity, high/low temperature and ultraviolet radiation), thereby inducing the accumulation of various flavonoids (Sharma et al., 2019). In this study, some metabolites and proteins involved in the metabolism of phenylpropane during the TLS and TS stages of SP2 plants has changed significantly (**Table 8** and **Figures 9E,F**). This confirmed that the impact of space flight on rice seeds lasted until the F2 generation. In addition, a large number of flavonoids were accumulated in this study, including APIIN, Oenin, Rutin, Kuromanin, Orobol, Formononetin, Malvidin 3-*O*-glucoside cation and Peonidin 3-galactoside cation. The accumulation of flavonoids in the cytoplasm can effectively decompose ROS generated by abiotic stress, and after the oxidation of flavonoids ends, ascorbic acid-mediated flavonoids were reconverted into primary metabolites and enter cell metabolism (Hernández et al., 2009). Our research results suggested that after rice seeds fly through space, there was still an oxidative stress effect in the F2 generation plants, and the F2 generation plants may also synthesize flavonoids to achieve the balance of ROS in the body.

## CONCLUSION

Overall, the F2 generation plants of rice still retained the stress of space flight to seeds, and this memory was caused by the instability of the rice genome after spaceflight. The memory of spaceflight stress induced the accumulation of ROS in the F2 generation plants, and the F2 generation plants-maintained ROS homeostasis by rebuilding their own metabolic pathways to adapt to the effects of spaceflight. As a signal molecule, ROS interacted with sugar signal pathways to mediate changes in amino acid metabolism, energy metabolism, and phenylpropane metabolism (**Figure 11**). These pathways helped maintain the homeostasis of ROS in plants. Moreover, the metabolic process of phenylpropane induced the biosynthesis of flavonoids, which were important for regulating ROS homeostasis (**Figure 11**). This study was the first one to combine metabolomics and proteomics methods to confirm that the effects of space flight on rice seeds lasted until the F2 generation, and the ability of F2 generation plants to adapt to space flight stress by rebuilding their own metabolic network. This research provides a new perspective for the study of spatial biological effects.

## DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

WL and JC designed the study. DZ conducted the experiments and analyzed the data. DZ and YY wrote the manuscript. HZ, CS, and CD revised the manuscript. SG conducted the field planting of experimental materials. DC provided the experimental site. YS guided the experiment. All authors read and approved the manuscript in its final form.

## FUNDING

This research was supported by The National Key Research and Development Program of China (2017YFC1601900), Heilongjiang Touyan Team (HITTY-20190034) and Planning Project for Space Application (01-1-08).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.900143/full#supplementary-material

## REFERENCES

Abbas, S. M. (2012). Effects of low temperature and selenium application on growth and the physiological changes in sorghum seedlings. *J. Stress Physiol. Biochem.* 8, 268–286.

Abulfaraj, A. A., and Jalal, R. S. (2021). Use of plant growth-promoting bacteria to enhance salinity stress in soybean (Glycine max L.) plants. *Saudi J. Biol. Sci.* 28, 3823–3834. doi: 10.1016/J.SJBS.2021.03.053

Aebi, H. (1984). Catalase in Vitro. *Methods Enzymol.* 105, 121–126. doi: 10.1016/S0076-6879(84)05016-3

Afzal, S., Chaudhary, N., and Singh, N. K. (2021). "Role of Soluble Sugars in Metabolism and Sensing Under Abiotic Stress, "in *Plant Growth Regulators*. eds T. Aftab, K. R. Hakeem. (Cham: Springer). 305–334. doi: 10.1007/978-3-030-61153-8_14

Alharby, H. F., Nahar, K., Al-Zahrani, H. S., Hakeem, K. R., and Hasanuzzaman, M. (2021). Enhancing Salt Tolerance in Soybean by Exogenous Boron: intrinsic Study of the Ascorbate-Glutathione and Glyoxalase Pathways. *Plants* 10:2085. doi: 10.3390/PLANTS10102085

Bailey, R. W. (1958). The reaction of pentoses with anthrone. *Biochem. J.* 68, 669–672. doi: 10.1042/bj0680669

Bakht, J., Shafi, M., Jamal, Y., and Sher, H. (2011). Response of maize (Zea mays L.) to seed priming with NaCl and salinity stress. *Span. J. Agric. Res.* 9, 252–261. doi: 10.5424/sjar/20110901-113-10

Bali, A. S., and Sidhu, G. P. S. (2020). "Growth and Morphological Changes of Agronomic Crops Under Abiotic Stress," in *Agronomic Crops*. Ed. M. Hasanuzzaman. (Singapore: Springer), 1–11. doi: 10.1007/978-981-15-0025-1_1

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. *Nature* 391, 485–488. doi: 10.1038/35140

Bogdanović, J., Mojović, M., Milosavić, N., Mitrović, A., Vučinić, Ž, and Spasojević, I. (2008). Role of fructose in the adaptation of plants to cold-induced oxidative stress. *Eur. Biophys. J.* 37, 1241–1246. doi: 10.1007/s00249-008-0260-9

Boyko, A., and Kovalchuk, I. (2011). Genome instability and epigenetic modification-heritable responses to environmental stress? *Curr. Opin. Plant Biol.* 14, 260–266. doi: 10.1016/j.pbi.2011.03.003

Brulfert, A., Clain, E., and Deysson, G. (1974). Deoxyguanosine, a potent cytokinesis inhibitor in plant cells. *Experientia* 30, 1010–1011. doi: 10.1007/BF01938977

Cai, B., Li, Q., Liu, F., Bi, H., and Ai, X. (2018). Decreasing fructose-1,6-bisphosphate aldolase activity reduces plant growth and tolerance to chilling stress in tomato seedlings. *Physiol. Plant.* 163, 247–258. doi: 10.1111/ppl.12682

Castillo, F. J., Penel, C., and Greppin, H. (1984). Peroxidase release induced by ozone in Sedum album leaves. Involvement of Ca2+. *Plant Physiol.* 74, 846–851. doi: 10.1104/pp.74.4.846

Chee Hark, H., Dale, J., and Liu, J. R. (1997). Regulation of cytosolic fructose-1,6-bisphosphatase under water-stressed leaves of sugar beet: protein modification is not a mechanism for coarse control. *J. Plant Biol.* 40, 261–266. doi: 10.1007/bf03030458

Chen, J., Le, X. C., and Zhu, L. (2019). Metabolomics and transcriptomics reveal defense mechanism of rice (Oryza sativa) grains under stress of 2,2",4,4"-tetrabromodiphenyl ether. *Environ. Int.* 133:105154. doi: 10.1016/j.envint.2019.105154

Chen, Y., Qian, J., You, L., Zhang, X., Jiao, J., Liu, Y., et al. (2018). Subunit interaction differences between the replication factor C complexes in arabidopsis and rice. *Front. Plant Sci.* 9:779. doi: 10.3389/fpls.2018.00779

Choi, W.-G., Barker, R. J., Kim, S.-H., Swanson, S. J., and Gilroy, S. (2019). Variation in the transcriptome of different ecotypes of *Arabidopsis thaliana* reveals signatures of oxidative stress in plant responses to spaceflight. *Am. J. Bot.* 106, 123–136. doi: 10.1002/ajb2.1223

Couée, I., Sulmon, C., Gouesbet, G., and El Amrani, A. (2006). Involvement of soluble sugars in reactive oxygen species balance and responses to oxidative stress in plants. *J. Exp. Bot.* 57, 449–459. doi: 10.1093/jxb/erj027

Cramer, G. R., Van Sluyter, S. C., Hopper, D. W., Pascovici, D., Keighley, T., and Haynes, P. A. (2013). Proteomic analysis indicates massive changes in metabolism prior to the inhibition of growth and photosynthesis of grapevine (Vitis vinifera L.) in response to water deficit. *BMC Plant Biol.* 13:49. doi: 10.1186/1471-2229-13-49

Cui, J., Xia, W., Wei, S., Zhang, M., Wang, W., Zeng, D., et al. (2019). Photosynthetic Performance of Rice Seedlings Originated from Seeds Exposed to Spaceflight Conditions. *Photochem. Photobiol.* 95, 1205–1212. doi: 10.1111/php.13097

Das, P., Manna, I., Sil, P., Bandyopadhyay, M., and Biswas, A. K. (2019). Exogenous silicon alters organic acid production and enzymatic activity of TCA cycle in two NaCl stressed indica rice cultivars. *Plant Physiol. Biochem.* 136, 76–91. doi: 10.1016/j.plaphy.2018.12.026

Demidchik, V., Straltsova, D., Medvedev, S. S., Pozhvanov, G. A., Sokolik, A., and Yurin, V. (2014). Stress-induced electrolyte leakage: the role of K+-permeable channels and involvement in programmed cell death and metabolic adjustment. *J. Exp. Bot.* 65, 1259–1270. doi: 10.1093/jxb/eru004

Deyong, Z., Jie, C., Yishu, Y., Meng, Z., Shan, S., Xin, G., et al. (2020). Effects of Space Flight on Expression of Key Proteins in Rice Leaves. *Rice Sci.* 27, 423–433. doi: 10.1016/j.rsci.2019.12.011

Fan, W., Ge, G., Liu, Y., Wang, W., Liu, L., and Jia, Y. (2018). Proteomics integrated with metabolomics: analysis of the internal causes of nutrient changes in alfalfa at different growth stages. *BMC Plant Biol.* 18:78. doi: 10.1186/s12870-018-1291-8

Feng, Z., Wu, C., Wang, C., Roh, J., Zhang, L., Chen, J., et al. (2016). SLG controls grain size and leaf angle by modulating brassinosteroid homeostasis in rice. *J. Exp. Bot.* 67, 4241–4253. doi: 10.1093/jxb/erw204

Ferl, R. J., Koh, J., Denison, F., and Paul, A. L. (2015). Spaceflight induces specific alterations in the proteomes of arabidopsis. *Astrobiology* 15, 32–56. doi: 10.1089/ast.2014.1210

Fernie, A. R., Carrari, F., and Sweetlove, L. J. (2004). Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr. Opin. Plant Biol.* 7, 254–261. doi: 10.1016/j.pbi.2004.03.007

Foreman, J., Demidchik, V., Bothwell, J. H. F., Mylona, P., Miedema, H., Torres, M. A., et al. (2003). Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* 422, 442–446. doi: 10.1038/nature01485

Friesner, J., and Britt, A. B. (2003). Ku80- and DNA ligase IV-deficient plants are sensitive to ionizing radiation and defective in T-DNA integration. *Plant J.* 34, 427–440. doi: 10.1046/j.1365-313X.2003.01738.x

Galili, G. (2011). The aspartate-family pathway of plants: linking production of essential amino acids with energy and stress regulation. *Plant Signal. Behav.* 6, 192–195. doi: 10.4161/psb.6.2.14425

Gharechahi, J., Hajirezaei, M. R., and Salekdeh, G. H. (2015). Comparative proteomic analysis of tobacco expressing cyanobacterial flavodoxin and its wild type under drought stress. *J. Plant Physiol.* 175, 48–58. doi: 10.1016/j.jplph.2014.11.001

Gomi, K., Sasaki, A., Itoh, H., Ueguchi-Tanaka, M., Ashikari, M., Kitano, H., et al. (2004). GID2, an F-box subunit of the SCF E3 complex, specifically interacts with phosphorylated SLR1 protein and regulates the gibberellin-dependent degradation of SLR1 in rice. *Plant J.* 37, 626–634. doi: 10.1111/j.1365-313X.2003.01990.x

Guisinger, M. M., and Kiss, J. Z. (1999). The influence of microgravity and spaceflight on columella cell ultrastructure in starch-deficient mutants of Arabidopsis. *Am. J. Bot.* 86, 1357–1366. doi: 10.2307/2656918

Hasanuzzaman, M., Bhuyan, M. H. M. B., Zulfiqar, F., Raza, A., Mohsin, S. M., Al Mahmud, J., et al. (2020). Reactive Oxygen Species and Antioxidant Defense in Plants under Abiotic Stress: revisiting the Crucial Role of a Universal Defense Regulator. *Antioxidants* 9:681. doi: 10.3390/ANTIOX9080681

Hasanuzzaman, M., Parvin, K., Anee, T. I., Awal, A., Masud, C., and Nowroz, F. (2022). "Salt Stress Responses and Tolerance in Soybean, "in *Plant Stress Physiology*. eds H. Mirza, N. Kamran. (London: Intech Open). doi: 10.5772/INTECHOPEN.102835

Heath, R. L., and Packer, L. (1968). Photoperoxidation in isolated chloroplasts. I. Kinetics and stoichiometry of fatty acid peroxidation. *Arch. Biochem. Biophys.* 125, 189–198. doi: 10.1016/0003-9861(68)90654-1

Hernández, I., Alegre, L., Van Breusegem, F., and Munné-Bosch, S. (2009). How relevant are flavonoids as antioxidants in plants? *Trends Plant Sci.* 14, 125–132. doi: 10.1016/j.tplants.2008.12.003

Hilal, M., Parrado, M. F., Rosa, M., Gallardo, M., Orce, L., Massa, E. M., et al. (2007). Epidermal Lignin Deposition in Quinoa Cotyledons in Response to UV-B Radiation. *Photochem. Photobiol.* 79, 205–210. doi: 10.1111/j.1751-1097.2004.tb00011.x

Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L., and Braun, H. P. (2015). Amino Acid Catabolism in Plants. *Mol. Plant* 8, 1563–1579. doi: 10.1016/j.molp.2015.09.005

Hossen, M. S., Karim, M. F., Fujita, M., Bhuyan, M. H. M. B., Nahar, K., Masud, A. A. C., et al. (2022). Comparative Physiology of Indica and Japonica Rice under Salinity and Drought Stress: an Intrinsic Study on Osmotic Adjustment, Oxidative Stress, Antioxidant Defense and Methylglyoxal Detoxification. *Stresses* 2, 156–178. doi: 10.3390/STRESSES2020012

Jia, X. M., Zhu, Y. F., Hu, Y., Zhang, R., Cheng, L., Zhu, Z. L., et al. (2019). Integrated physiologic, proteomic, and metabolomic analyses of Malus halliana adaptation to saline–alkali stress. *Hortic. Res.* 6:91. doi: 10.1038/s41438-019-0172-0

Keunen, E., Peshev, D., Vangronsveld, J., Van Den Ende, W., and Cuypers, A. (2013). Plant sugars are crucial players in the oxidative challenge during abiotic stress: extending the traditional concept. *Plant Cell Environ.* 36, 1242–1255. doi: 10.1111/pce.12061

Kumar, S., Kumari, R., and Sharma, V. (2015). Transgenerational Inheritance in Plants of Acquired Defence Against Biotic and Abiotic Stresses: implications

and Applications. *Agric. Res.* 4, 109–120. doi: 10.1007/S40003-015-0170-X/FIGURES/2

Kumari, A., and Parida, A. K. (2018). Metabolomics and network analysis reveal the potential metabolites and biological pathways involved in salinity tolerance of the halophyte Salvadora persica. *Environ. Exp. Bot.* 148, 85–99. doi: 10.1016/j.envexpbot.2017.12.021

Lee, J., and Zhou, P. (2007). DCAFs, the Missing Link of the CUL4-DDB1 Ubiquitin Ligase. *Mol. Cell* 26, 775–780. doi: 10.1016/j.molcel.2007.06.001

Less, H., and Galili, G. (2008). Principal transcriptional programs regulating plant amino acid metabolism in response to abiotic stresses. *Plant Physiol.* 147, 316–330. doi: 10.1104/pp.108.115733

Li, M., Guo, R., Jiao, Y., Jin, X., Zhang, H., and Shi, L. (2017). Comparison of salt tolerance in Soja based on metabolomics of seedling roots. *Front. Plant Sci.* 8:1101. doi: 10.3389/fpls.2017.01101

Li, R., Xia, J., Xu, Y., Zhao, X., Liu, Y. G., and Chen, Y. (2014). Characterization and genetic mapping of a Photoperiod-sensitive dwarf 1 locus in rice (Oryza sativa L.). *Theor. Appl. Genet.* 127, 241–250. doi: 10.1007/s00122-013-2213-7

Li, Y., Cao, X. L., Zhu, Y., Yang, X. M., Zhang, K. N., Xiao, Z. Y., et al. (2019). Osa-miR398b boosts H2O2 production and rice blast disease-resistance via multiple superoxide dismutases. *N. Phytol.* 222, 1507–1522. doi: 10.1111/NPH.15678

Liu, Q., Wang, J., Miki, D., Xia, R., Yu, W., He, J., et al. (2010). DNA replication factor C1 mediates genomic stability and transcriptional gene silencing in Arabidopsis. *Plant Cell* 22, 2336–2352. doi: 10.1105/tpc.110.076349

Liu, Y., Deng, Y., Li, G., and Zhao, J. (2013). Replication factor C1 (RFC1) is required for double-strand break repair during meiotic homologous recombination in arabidopsis. *Plant J.* 73, 154–165. doi: 10.1111/tpj.12024

Loeffen, J., Smeets, R., Smeitink, J., Triepels, R., Sengers, R., Trijbels, F., et al. (1999). The human NADH: ubiquinone oxidoreductase NDUFS5 (15 kDa) subunit: cDNA cloning, chromosomal localization, tissue distribution and the absence of mutations in isolated complex I-deficient patients. *J. Inherit. Metab. Dis.* 22, 19–28. doi: 10.1023/a:1005434912463

Lutts, S., Kinet, J. M., Bouharmont, J. (1996). NaCl-induced Senescence in Leaves of Rice (Oryza sativaL.) Cultivars Differing in Salinity Resistance. *Ann. Bot.* 78, 389–398. doi: 10.1006/anbo.1996.0134

Ma, Y., Cheng, Z., Wang, W., and Sun, Y. (2007). Proteomic analysis of high yield rice variety mutated from spaceflight. *Adv. Space Res.* 40, 535–539. doi: 10.1016/j.asr.2007.05.028

Mahmood, T., Islam, K. R., and Muhammad, S. (2007). Toxic effects of heavy metals on early growth and tolerance of cereal crops. *Pak. J. Bot.* 39, 451–462. doi: 10.1093/jxb/erw224

Manian, V., Gangapuram, H., Orozco, J., Janwa, H., and Agrinsoni, C. (2021). Network analysis of local gene regulators in Arabidopsis thaliana under spaceflight stress. *Computers* 10:18. doi: 10.3390/computers10020018

Matía, I., González-Camacho, F., Herranz, R., Kiss, J. Z., Gasset, G., van Loon, J. J. W. A., et al. (2010). Plant cell proliferation and growth are altered by microgravity conditions in spaceflight. *J. Plant Physiol.* 167, 184–193. doi: 10.1016/j.jplph.2009.08.012

Mochizuki, S., Fukumoto, T., Ohara, T., Ohtani, K., Yoshihara, A., Shigematsu, Y., et al. (2020). The rare sugar d-tagatose protects plants from downy mildews and is a safe fungicidal agrochemical. *Commun. Biol.* 3:423. doi: 10.1038/s42003-020-01133-7

Moreno-Villanueva, M., Wong, M., Lu, T., Zhang, Y., and Wu, H. (2017). Interplay of space radiation and microgravity in DNA damage and DNA damage response. *NPJ Microgravity* 3:14. doi: 10.1038/s41526-017-0019-7

Mortley, D. G., Bonsi, C. K., Hill, W. A., Morris, C. E., Williams, C. S., Davis, C. F., et al. (2008). Influence of microgravity environment on root growth, soluble sugars, and starch concentration of sweetpotato stem cuttings. *J. Am. Soc. Hortic. Sci.* 133, 327–332. doi: 10.21273/jashs.133.3.327

Naiki, T., Shimomura, T., Kondo, T., Matsumoto, K., and Sugimoto, K. (2000). Rfc5, in Cooperation with Rad24, Controls DNA Damage Checkpoints throughout the Cell Cycle inSaccharomyces cerevisiae. *Mol. Cell. Biol.* 20, 5888–5896. doi: 10.1128/mcb.20.16.5888-5896.2000

Nishizawa, A., Yabuta, Y., and Shigeoka, S. (2008). Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiol.* 147, 1251–1263. doi: 10.1104/pp.108.122465

Paul, A. L., Sng, N. J., Zupanska, A. K., Krishnamurthy, A., Schultz, E. R., and Ferl, R. J. (2017). Genetic dissection of the Arabidopsis spaceflight transcriptome: are some responses dispensable for the physiological adaptation of plants to spaceflight?. *PLoS One* 12:e0180186. doi: 10.1371/journal.pone.0180186

Pratelli, R., and Pilot, G. (2014). Regulation of amino acid metabolic enzymes and transporters in plants. *J. Exp. Bot.* 65, 5535–5556. doi: 10.1093/jxb/eru320

Rojas-González, J. A., Soto-Súarez, M., García-Díaz, Á, Romero-Puertas, M. C., Sandalio, L. M., Mérida, Á, et al. (2015). Disruption of both chloroplastic and cytosolic FBPase genes results in a dwarf phenotype and important starch and metabolite changes in Arabidopsis thaliana. *J. Exp. Bot.* 66, 2673–2689. doi: 10.1093/jxb/erv062

Rosa, M., Prado, C., Podazza, G., Interdonato, R., González, J. A., Hilal, M., et al. (2009). Soluble sugars-metabolism, sensing and abiotic stress a complex network in the life of plants. *Plant Signal. Behav.* 4, 388–393. doi: 10.4161/psb.4.5.8294

Rosa-Téllez, S., Anoman, A. D., Flores-Tornero, M., Toujani, W., Alseek, S., Fernie, A. R., et al. (2018). Phosphoglycerate kinases are co-regulated to adjust metabolism and to optimize growth. *Plant Physiol.* 176, 1182–1198. doi: 10.1104/pp.17.01227

Saleem, M. (2003). Response of Durum and Bread wheat Genotypes to Drought Stress: biomass and Yield Components. *Asian J. Plant Sci.* 2, 290–293. doi: 10.3923/ajps.2003.290.293

Shahbazy, M., Moradi, P., Ertaylan, G., Zahraei, A., and Kompany-Zareh, M. (2020). FTICR mass spectrometry-based multivariate analysis to explore distinctive metabolites and metabolic pathways: A comprehensive bioanalytical strategy toward time-course metabolic profiling of Thymus vulgaris plants responding to drought stress. *Plant Sci.* 290:110257. doi: 10.1016/j.plantsci.2019.110257

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M., and Zheng, B. (2019). Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules* 24:2452. doi: 10.3390/molecules24132452

Soliman, M. H., Abdulmajeed, A. M., Alhaithloul, H., Alharbi, B. M., El-Esawi, M. A., Hasanuzzaman, M., et al. (2020). Saponin biopriming positively stimulates antioxidants defense, osmolytes metabolism and ionic status to confer salt stress tolerance in soybean. *Acta Physiol. Plant.* 42:114. doi: 10.1007/S11738-020-03098-W/FIGURES/6

Sun, Y., Wang, W., Zhang, M., Zhao, L., Mi, D., Zhang, B., et al. (2019). "Space Radiation Systems Biology Research in SJ-10 Satellite," in *Research for Development*. eds D. Enkui, L. Mian. (Berlin: Springer), 43–68. doi: 10.1007/978-981-13-6325-2_3

Taha, R. S., Seleiman, M. F., Alotaibi, M., Alhammad, B. A., Rady, M. M., and Mahdi, A. H. A. (2020). Exogenous Potassium Treatments Elevate Salt Tolerance and Performances of Glycine max L. by Boosting Antioxidant Defense System under Actual Saline Field Conditions. *Agronomy* 10:1741. doi: 10.3390/AGRONOMY10111741

Thounaojam, T. C., Panda, P., Mazumdar, P., Kumar, D., Sharma, G. D., Sahoo, L., et al. (2012). Excess copper induced oxidative stress and response of antioxidants in rice. *Plant Physiol. Biochem.* 53, 33–39. doi: 10.1016/j.plaphy.2012.01.006

Vaulina, E., Anikeeva, I., and Kostina, L. (1984). Radiosensibility of higher plant seeds after space flight. *Adv. Space Res.* 4, 103–107. doi: 10.1016/0273-1177(84)90231-X

Velikova, M., Bankova, V., Marcucci, M. C., Tsvetkova, I., and Kujumgiev, A. (2000). Chemical composition and biological activity of propolis from Brazilian Meliponinae. *Z. Nat. Forsch. A. J. Phys. Sci.* 55, 785–789. doi: 10.1515/znc-2000-9-1018

Wang, L. H., Li, G. L., Wei, S., Li, L. J., Zuo, S. Y., Liu, X., et al. (2019). Effects of exogenous glucose and sucrose on photosynthesis in triticale seedlings under salt stress. *Photosynthetica* 57, 286–294. doi: 10.32615/ps.2019.030

Wang, W., Gu, D. P., Zheng, Q., and Sun, Y. Q. (2008). Leaf proteomic analysis of three rice heritable mutants after seed space flight. *Adv. Space Res.* 42, 1066–1071. doi: 10.1016/j.asr.2008.02.004

Wang, Y., Zhang, X., Lu, S., Wang, M., Wang, L., Wang, W., et al. (2012). Inhibition of a basal transcription factor 3-like gene Osj10gBTF3 in rice results in significant plant miniaturization and typical pollen abortion. *Plant Cell Physiol.* 53, 2073–2089. doi: 10.1093/pcp/pcs146

Wei, L. J., Xu, J. L., Wang, J. M., Yang, Q., Luo, R. T., Zhang, M. X., et al. (2006). A Comparative Study on Mutagenic Effects of Space Flight and Irradiation of

γ-rays on Rice. *Agric. Sci. CHIN.* 5, 812–819. doi: 10.1016/S1671-2927(06)60129-6

Wu, Z., Waneka, G., Broz, A. K., King, C. R., and Sloan, D. B. (2020). MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. *Proc. Natl. Acad. Sci. U.S.A.* 117, 16448–16455. doi: 10.1073/pnas.2001998117

Xie, X., He, Z., Chen, N., Tang, Z., Wang, Q., and Cai, Y. (2019). The roles of environmental factors in regulation of oxidative stress in plant. *BioMed. Res. Int.* 2019:9732325. doi: 10.1155/2019/9732325

Xu, P., Chen, H., Jin, J., and Cai, W. (2018). Single-base resolution methylome analysis shows epigenetic changes in Arabidopsis seedlings exposed to microgravity spaceflight conditions on board the SJ-10 recoverable satellite. *NPJ Microgravity* 4:12. doi: 10.1038/s41526-018-0046-z

Yang, C., Hu, H., Ren, H., Kong, Y., Lin, H., Guo, J., et al. (2016). LIGHT-INDUCED RICE1 regulates light-dependent attachment of LEAF-TYPE FERREDOXIN-NADP+ OXIDOREDUCTASE to the thylakoid membrane in rice and arabidopsis. *Plant Cell* 28, 712–728. doi: 10.1105/tpc.15.01027

Yang, Y., Ma, L., Zeng, H., Chen, L. Y., Zheng, Y., Li, C. X., et al. (2018). iTRAQ-based proteomics screen for potential regulators of wheat (Triticum aestivum L.) root cell wall component response to Al stress. *Gene* 675, 301–311. doi: 10.1016/j.gene.2018.07.008

Yu, X., Wu, H., Wei, L. J., Cheng, Z. L., Xin, P., Huang, C. L., et al. (2007). Characteristics of phenotype and genetic mutations in rice after spaceflight. *Adv. Space Res.* 40, 528–534. doi: 10.1016/j.asr.2007.06.022

Zeng, D., Cui, J., Yin, Y., Xiong, Y., Liu, M., Guan, S., et al. (2021). Metabolomics analysis in different development stages on SP0 generation of rice seeds after space flight. *Front. Plant Sci.* 12:1235. doi: 10.3389/FPLS.2021.700267

Zeng, D., Cui, J., Yin, Y., Zhang, M., Shan, S., Liu, M. Y., et al. (2020). Proteomic analysis in different development stages on SP0 generation of rice seeds after space flight. *Life Sci. Space Res.* 26, 34–45. doi: 10.1016/j.lssr.2020.02.001

Zhao, H., Qiu, J., Wang, Y., Zhao, H., Qiu, J., and Wang, Y. (2019). "System Design and Flight Results of China SJ-10 Recoverable Microgravity Experimental Satellite, "in *Research for Development*. eds D. Enkui, L. Mian. (Singapore: Springer). 9–42. doi: 10.1007/978-981-13-6325-2_2

Zhong, M., Yuan, Y., Shu, S., Sun, J., Guo, S., Yuan, R., et al. (2016). Effects of exogenous putrescine on glycolysis and Krebs cycle metabolism in cucumber leaves subjected to salt stress. *Plant Growth Regul.* 79, 319–330. doi: 10.1007/s10725-015-0136-9

Zhou, D., Sun, Y., Zhang, B., Zhang, S., Sun, Y., Liang, J., et al. (2018). Radiation Measured for Chinese Satellite SJ-10 Space Mission. *J. Geophys. Res. Space Phys.* 123, 1690–1700. doi: 10.1002/2017JA024697

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 October 2022
DOI 10.3389/fpls.2022.964558



OPEN ACCESS

EDITED BY
Kalenahalli Yogendra,
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

REVIEWED BY
Miguel García-Parra,
University of Cauca, Colombia
Niranjan Hegde,
McGill University, Canada

*CORRESPONDENCE
Wei Lv
lvwei91@126.com
Guojun Mu
mgj99999@126.com

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 08 June 2022
ACCEPTED 20 September 2022
PUBLISHED 19 October 2022

CITATION
Zhao Y, Ma Y, Li J, Liu B, Liu X, Zhang J, Zhang M, Wang C, Zhang L, Lv W and Mu G (2022) Transcriptomics–metabolomics joint analysis: New highlight into the triterpenoid saponin biosynthesis in quinoa (*Chenopodium quinoa* Willd.). *Front. Plant Sci.* 13:964558. doi: 10.3389/fpls.2022.964558



# Transcriptomics–metabolomics joint analysis: New highlight into the triterpenoid saponin biosynthesis in quinoa (*Chenopodium quinoa* Willd.)

Yulu Zhao[1], Yucong Ma[1], Jiawei Li[1], Bin Liu[1], Xiaoqing Liu[1], Jianheng Zhang[1], Min Zhang[1], Chunmei Wang[1], Liping Zhang[1], Wei Lv[2]* and Guojun Mu[1]*

[1]North China Key Laboratory for Crop Germplasm Resources of Education Ministry, Laboratory of Hebei Provincial Crop Germplasm Resources, Hebei Agricultural University, Baoding, China, [2]National Semi-arid Agricultural Engineering Technology Research Center, Shijiazhuang, China

Quinoa (*Chenopodium quinoa* Willd.) contains various physiologically active substances, including vitamins, polyphenols, flavonoids, phytosterols, and saponins. Research showed that saponins were the protective substances in the outer layer of quinoa seeds to defend against microbes, herbivores, and insects. Because the aglycones of quinoa saponins are triterpenoids, they are called triterpenoid saponins (TSs). In addition, the presence of TS imparted bitterness in quinoa and resulted in anticancer and anti-inflammatory effects. In this study, the seeds of low-saponin quinoa, NT376-2 (N), and high-saponin quinoa, B-12071(B), at 30 and 60 days after flowering (DAF) were used to measure the TS content and evaluated for their transcriptomic and metabolomic profiles. The amounts of TS were found to significantly differ between all possible comparisons: N and B at 30 DAF (N1_*vs*_B1), N and B at 60 DAF (N2_*vs*_B2), N at 30 DAF and 60 DAF (N1_*vs*_N2), and B at 30 DAF and 60 DAF (B1_*vs*_B2). RNA sequencing (RNA-seq) was used to screen differentially expressed genes (DEGs) and revealed 14,703 upregulated DEGs and 26,267 downregulated DEGs in the four comparison groups. The 311 overlapping DEGs found in the four comparisons were used for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses to screen for DEGs related to TS biosynthesis in quinoa. Metabolomics analysis identified acetyl-CoA, 1-hydroxy-2-methyl-2-butenyl-4-diphosphate, farnesal, and (S)-2,3-epoxysqualene as the key differentially accumulated metabolites (DAMs). Transcriptomics–metabolomics joint analysis showed that triterpenoid biosynthesis and terpenoid backbone biosynthesis were the enriched pathways of TS biosynthesis; farnesal were the key DAMs shared in the four comparison groups and associated with 10 key candidate DEGs related to TS biosynthesis in

quinoa. These results provided important references for in-depth research on the metabolic mechanism of TS in quinoa.

KEYWORDS

quinoa, triterpenoid saponins (TSs), triterpenoid biosynthesis, transcriptomics-metabolomics joint analysis, qRT-PCR

## Introduction

Quinoa (*Chenopodium quinoa* Willd., *2n* = *4x* = 36), an annual dicotyledonous plant belonging to the Amaranthaceae family, is known as "nutritional gold" and "superfood." Currently, quinoa is used to cure obesity, cardiovascular diseases, diabetes, and other chronic diseases (Lim et al., 2020). Furthermore, research had shown that quinoa is affluent in all amino acids essential for human beings and contained various physiologically active substances, including vitamins, polyphenols, flavonoids, phytosterols, and saponins. Saponins were the major anti-nutritional factor in the grain and presented in the outer layer of quinoa seeds and were used to defend against microbes, herbivores, and insects in plants. El Hazzam et al. (2020) demonstrated that the quinoa saponin was the pentacyclic triterpenoid saponin (TS) derived from β-amyrin and was an important secondary metabolite of the triterpenoid biosynthesis. Research showed that the bitterness of quinoa was closely related to the presence of TS (Gómez-Caravaca et al., 2014). Nowadays, because TSs have diverse pharmacological activities such as hemolytic, anticancer, and anti-inflammatory, TS biosynthesis has been widely studied (Yao et al., 2020).

The TS biosynthetic pathway was mainly divided into three stages: the initial stage, the terpenoid framework construction stage, and the modification stage (Zhao et al., 2017). In the initial stage, 3-hydroxy-3-methylglutaryl-coenzyme A synthase (HMGS) catalyzes the condensation of acetoacetyl-CoA and acetyl-CoA to produce S-3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). HMG-CoA and mevalonate kinase (MVK) *via* the mevalonic acid pathway (MVA) produce the upstream precursor isopentenyl diphosphate (IPP). The MVA pathway exists in the cytoplasm and endoplasmic reticulum and is the original pathway for triterpene and sterol biosynthesis (Huang et al., 2014). Meanwhile, the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway is also presented in the initial stage. Herein, IPP is synthesized by pyruvate through 1-deoxy-D-xylulose-5-phosphate synthase (DXS) to produce the isomer of IPP and dimethylallyl pyrophosphate (DMAPP). The MEP pathway, located in plastids, is mainly involved in the biosynthesis of diterpenes, monoterpenes, carotenoids, and isoprenes (Luo et al., 2016). In the terpenoid framework construction stage, IPP condenses with DMAPP to form farnesyl pyrophosphate (FPP) through geranyl pyrophosphate synthase (GGPS) and farnesyl-pyrophosphate synthase (FPS1). Subsequently, FPP forms the (S)-2,3-oxidosqualene under the action of squalene synthase (SQS1) and squalene epoxidase (SQE). (S)-2,3-oxidosqualene is an important precursor for the formation of the triterpenoid framework (Haralampidis et al., 2002). In the modification stage, the triterpenoid framework is hydroxylated by cytochrome P450 (CYP450) and is glycosylated by glycosyltransferases (UTG), which eventually forms TS.

At present, the research on TS is mainly presented in medicinal plants, such as *Panax notoginseng* and *American ginseng* (Niu, 2013). However, studies of TS in quinoa are mainly limited in its extraction methods. Ultrasonic extraction is the most common method for extracting TS and is beneficial to save the extraction times compared with the traditional extraction methods (Espinoza et al., 2021). Meanwhile, most research is also focused on the expression analysis of genes about TS in quinoa. Fiallos-Jurado et al. (2016) showed that candidate genes related to quinoa saponin biosynthesis, *Cqβ-AS*, *CqCYP716A78*, and *CqCYP716A79*, were induced by MeJA, suggesting that this phytohormone was used to modulate saponin biosynthesis in quinoa. So far, there were a few reports on the pathway of TS biosynthesis in quinoa. In addition, research revealed that TS shares a common precursor synthesis pathway with other triterpenoid compounds in the initial stage and the terpenoid framework construction stage; however, the modification stage is highly specific and diverse (Zhao and Li, 2018). Therefore, the TS biosynthesis pathway needs to be further explored. In this study, the interactions between the differentially expressed genes (DEGs) and differentially accumulated metabolites (DAMs) of TS biosynthesis in quinoa were investigated using transcriptomics–metabolomics joint analysis, and the key metabolic pathways of TS biosynthesis were explored. Results will provide an important reference for the in-depth study of biosynthetic mechanisms of TS in quinoa.

## Materials and methods

### Sample preparation

In this study, 140 quinoa germplasm materials from the National Semi-Arid Agricultural Engineering Science and Technology Research Center, Hebei Province, were used for

the determination of saponin criteria (Supplementary Table S1). The saponin content between 4.7 and 11.3 mg/g was identified as a high-saponin quinoa variety, and the saponin content<0.2–0.4 mg/g was identified as a low-saponin quinoa variety (Ruiz et al., 2017). According to the classification criteria of saponin content in quinoa, the low-saponin quinoa variety, NT376-2 (N) (0.253 mg/g), and the high-saponin quinoa variety, B-12071 (B) (10.514 mg/g), were screened and were used for the next study. Significant differences of saponin content existed between the two quinoa varieties ($p< 0.0001$) (Supplementary Figure S1). N and B were planted separately at individual spacings of 18 cm and row spacings of 45 cm in Guyuan County, Hebei Province, P. R. China (41°31'21.59″N, 115°52'34.37″E, 1,422 m a.s.l.), in May 2020 and May 2021. The field was routinely managed, and the plants were labeled with tags after flowering from August to September 2021. Quinoa starts to flower at 40–55 days after sowing. Five grams of quinoa seeds for N and B at 30 DAF, 40 DAF, 50 DAF, and 60 DAF were collected and were used for measuring the content of TS with three biological replicates (Supplementary Table S2). Because the TS content of quinoa seeds at 30 DAF and 60 DAF was significantly different ($p< 0.001$), the 3–5 g of seeds of N at 30 DAF (N1), N at 60 DAF (N2), B at 30 DAF (B1), and B at 60 DAF (B2) were taken and frozen in liquid nitrogen and were used for the transcriptomics and metabolomics analyses.

## Extraction and measurement of triterpenoid saponins in quinoa

The TS content was measured using the vanillin-glacial acetic acid-perchloric acid colorimetric method with three biological replicates (Huang et al., 2019). In this study, 1-g seeds were grounded to powder at 40 Hz for 20 min by a high-speed grinder (FW100, China), mixed with ethanol (1:30), sonicated at 45°C for 20 min by Ultrasonic Cleaner (SB-5200DTDN, China) to collect the supernatant by centrifuge (SIGMA3-18K, Germany) at 3,000 rpm for 15 min. Subsequently, 0.2 ml sample solution was dried at 70°C for 15 min and heated with 0.2 ml 5% vanillin-glacial acetic acid solution and 0.8 ml perchloric acid solution at 70°C for 15 min. After the chromogenic reaction, the sample solution was cooled for 10 min and finally mixed with 4 ml glacial acetic acid for measurement of UV absorbance. The absorbance was measured at 545 nm using UV-Vis spectrophotometer (UV-5800, Shanghai, China) and compared with methanol as the control. The total content of TS in the sample solution was calculated as follows:

$$m_{total}(mg/g) = mV/Mv$$

Note: $m$ is the mass of TS in the sample solution (mg) according to the calibration curve, $V$ is the total volume of the sample solution (ml), M is the total mass of quinoa (g), and the $v$ is the volume of the sample solution (ml). The two-tailed t-test was used to assess the significance of differences ($p\_value$). In this study, "*," "**," and "***" mean $p< 0.05$, $p< 0.01$, and $p< 0.001$, respectively.

## Transcriptomics analyses

### RNA sequencing and quality control

Total RNA was extracted from 12 samples of seeds (three biological replicates for N1, B1, N2, and B2) using TRIZOl Kit (Invitrogen, Carlsbad, CA, USA) (Saccenti et al., 2014). RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) according to the manufacturer's protocol, and RNA purity was tested using 1% agarose gel electrophoresis by NanoPhotometer spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). After total RNA was extracted, eukaryotic mRNA was enriched by Oligo(dT) beads, and the enriched mRNA was fragmented into short seqences using fragmentation buffer and reversly transcribed into cDNA using NEBNext Ultra RNA Library Prep Kit (NEB, USA). The cDNA libraries were sequenced using Illumina Novaseq6000 platform by Gene Denovo Biotechnology Co. (Guangzhou, China) and generated 150-bp paired-end reads. Raw reads containing 10% unknown nucleotides (N) or 50% low-quality bases ($Q \leq 20$) were filtered with fastp default parameters (Chen et al., 2018) to get high-quality clean reads. Clean reads were aligned with the tetraploid quinoa reference genome (https://www.ncbi.nlm.nih.gov/genome/?term=txid63459[orgn]) by HISAT2 system (Kim et al., 2015) according to the bowtie2 method (Freeberg and Kim, 2016). ASM168347v1 was used as the proper assembly version.

## Identification of differentially expressed genes

The genes were counted, and the expression level of genes, Fragments Per Kilobase of exon model per Million mapped fragments (FPKM), was calculated by RSEM (Li and Dewey, 2011). Principal component analysis (PCA) was performed based on hypergeometric distribution, and the DEGs between different comparison groups (N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2) were identified using the DESeq2 in the R package (Love et al., 2014). The $|\log2FC| \geq 1$ and false discovery rate (FDR)<0.05 were the criteria used to identify the significant DEGs in the transcriptomics, of which fold change (FC) denotes the difference in the relative expression levels of genes (Ruiying et al., 2020). FDR was obtained by correcting for the significance of differences ($p\_value$) using the Benjamini–Hochberg method (Pertea et al., 2015). Venny 2.1 online web tool (http://bioinfogp.cnb.csic.es/tools/venny/) was used to create a Venn diagram to identify the intersectional genes among the above comparison groups.

## Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed based on the DAVID database (david.abcc.ncifcrf.gov/). GO enrichment analyses annotated DEGs from cellular components (CC), biological processes (BP), and molecular functions (MF) using GoSeq (Young et al., 2010). All DEGs were mapped with GO terms in the GO database (http://www.geneontology.org/), and gene numbers for each term were calculated. The *q_value* (Benjamini–Hochberg-corrected *p_value*)<0.05 was used as the threshold to screen the significantly enriched GO terms. KEGG enrichment analysis was applied to identify the significantly enriched pathways of DEGs, and the *q_value*<0.05 was used as the threshold (Kanehisa et al., 2004).

## Metabolomics analysis

### Qualitative and quantitative analyses of metabolites

Liquid chromatography–tandem mass spectrometry (LC-MS/MS) was used for the qualitative and quantitative analyses of metabolites. Twelve freeze-dried seed samples (three biological replicates for N1, B1, N2, and B2) were extracted with 70% methanol at 4°C for 24 h, and the homogenate was centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant was taken into a 2-ml centrifuge tube and was filtered through a 0.22-μm membrane to obtain the prepared sample for LC-MS. In this study, 20 μl was taken from each sample as the quality control (QC), and the remaining samples were used for LC-MS detection. The sample extracts were analyzed using ultraperformance liquid chromatography-electrosspray ionization mode-tandem mass spectrometry (UPLC-ESI-MS/MS) system. The UPLC conditions based on Thermo Ultimate 3000 system were set as f ollows: ACQUITY UPLC® HSS T3 (150 mm × 2.1 mm, 1.8 μm, Waters) column, 40°C, 5 mM ammonium formate in water (A) and acetonitrile (B) at a flow rate of 0.25 ml/min. The injection volume of each sample was 2 μl after equilibration. The ESI-MS experiments were executed on the Thermo Q Exactive mass spectrometer with 3.8 kV and -2.5 kV spray voltage of positive and negative ion modes, respectively. The capillary temperature was 325°C (Zelena et al., 2009).

The format of raw data files was converted into mzXML format using Proteowizard (v3.0.8789) to differentiate the MS/MS data. The R (v3.3.2) package XCMS (Smith et al., 2006) was used to perform peak identification, peak filtration, and peak alignment for each metabolite to obtain the mass-to-charge ratio (m/z), retention time and intensity, and positive and negative precursor molecule. The identification of metabolites is based on the exact molecular formula (molecular formula error<20 ppm), and the peaks were matched with the mass spectrometry public databases, including HMDB (http://www.hmdb.ca), Massbank (http://www.massbank.jp/), LipidMaps (http://www.lipidmaps.org), and mzClound (https://www.mzcloud.org) to confirm the annotations for the metabolites.

## Identification of differentially accumulated metabolites

All detected metabolites were annotated on the MetWare database, and multivariate statistical processing of metabolomic data was performed by MetaboAnalyst version 5.0 software. Quality control was carried out by evaluating the repeatability in the process of sample extraction with three biological replicates. For a preliminary visualization of differences between different sample groups, the unsupervised dimensionality reduction method and the PCA were applied using R package models (http://www.r-project.org/). The orthogonal projection to latent structures-discriminant analysis (OPLS-DA) model was further validated by cross-validation and permutation test for the metabolomics data, and a score chart and permutation chart of each comparison group were drawn to visualize differences among the comparison groups. The variable importance in projection (VIP) score of the OPLS model was applied to rank the metabolites that were best distinguished among the four comparison groups (Saccenti et al., 2014). The VIP ≥1 and |log2 (fold change)| ≥ 2< 0.05 were used as the criteria to identify the DAMs. The identified DAMs were mapped with KEGG database to enrich the significant pathways based on the *q_value*<0.05.

## Transcriptomics and metabolomics joint analyses

To investigate the relationship of DEGs and DAMs annotated with KEGG pathways, the transcriptomics and metabolomics joint analyses were carried out using the Pearson correlation coefficient (PCC), and the PCC was calculated by the COR program in R language. The PCC ≥0.6 and ≥0.8 mean the normal or significant correlation between the DEGs and DAMs, respectively. The network diagram was visualized by Cytoscape software (Shannon et al., 2003). The DEGs and DAMs were annotated to the KEGG database to obtain their common KEGG pathway.

## Verification of differentially expressed genes by quantitative real-time PCR

qRT-PCR was used to verify the expression levels of the candidate DEGs. The total RNA was extracted from the 12 samples (three biological replicates of N1, B1, N2, and B2) and reverse-transcribed into cDNA for qRT-PCR. The $ddH_2O$ was

added in the qRT-PCR reaction mixture, including 2 µl cDNA, 2 µl BlazeTaq™ SYBR ® Green qPCR mix2.0 (Applied Biosystems, Carlsbad, CA, USA), 2 µl qPCR Primer (2 µM), and 2 µl cDNA Template, to make up to 20 µl. The qRT-PCR analysis was performed by Bio-Rad CFX96 qRT-PCR system, and the reaction conditions were set as follows: 95°C for 15 min, PCR cycle step at 94°C for 20 s, annealing, and extension step at 60°C for 34 s. The specific primers were designed using Premier 5.0 software based on the Coding sequence (CDS) sequence of the target gene (Supplementary Table S3). The *ACT7* in quinoa was used as the internal reference, and the calculation of relative expression of DEGs by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) and t-test was applied to analyze the significant differences.

# Results

## Transcriptomics analyses

### RNA sequencing and quality control analyses

Total RNA was extracted, and 12 cDNA libraries were constructed for high-throughput sequencing. After removing 10% of unknown nucleotides and 50% of low-quality data (Q20<20%), a total of 87.59 Gb of clean data were obtained. The Q20 and Q30 scores were 97.26% and 92.72%, respectively. The Guanine Cytosine (GC) content was maintained at 43.60%–44.83%. The N50 of each sample was more than 14,851 bp, indicating the high sequencing quality of each sample. The sequence comparison between the clean reads and the reference genome showed that more than 95.51% of clean reads were mapped to the reference genome, 12.49%–14.23% of clean reads were mapped to multiple locations in the reference genome, and 4.06%–4.59% of clean reads were unmapped to the reference genome, indicating that a high proportion of reads from each sample was compared with the reference genome (Supplementary Table S4). In addition, the PCA showed that the first and second principal components (PC1 and PC2) were 71.8% and 17.4% of the variation, and the samples of the same variety were clustered together and those of different varieties were apart. The correlation heatmap analysis showed that the Pearson correlation coefficient (PCC) ranged from 0.375 to 1, and most of these samples met with PCC >0.8, indicating that these samples were significantly different and have good biological repeatability (Supplementary Figure S2).

## Identification of differentially expressed genes and enrichment analyses

'Base on |log2FC| ≥1 and FDR<0.05, a total of 40,970 DEGs were identified in the N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2 comparison groups. Among these DEGs, 6,307 DEGs were detected in the N1_*vs*_B1, including 3,544 upregulated genes and 2,763 downregulated genes; 2,130 DEGs were detected in the N2_*vs*_B2, including 618 upregulated genes and 1,512 downregulated genes; 13,560 DEGs were detected in the N1_*vs*_N2, including 4,693 upregulated genes and 8,873 downregulated genes; and 18,973 DEGs were identified in the B1_*vs*_B2, including 5,848 upregulated genes and 13,125 downregulated genes (Figure 1A; Supplementary Table S5). Subsequently, Venn diagram analysis revealed that 311 overlapping DEGs were shared among the four comparison groups, whereas 796, 183, 1,530, and 5,512 were specific genes in each comparison group, respectively (Figure 1B). These results indicated that the numbers of DEGs in N1_*vs*_N2 and B1_*vs*_B2 were more than those in N1_*vs*_B1 and N2_*vs*_B2, and DEGs were mainly downregulated (except for N1_*vs*_B1).

To further investigate the functions of the identified DEGs, GO and KEGG enrichment analyses were adopted. The GO enrichment analyses showed that 1,742, 1,390, 2,257, and 2,452 GO terms were enriched in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2, respectively, and were annotated with three categories, including CC, BP, and MF. In the CC, the DEGs were mainly enriched in the cell (GO:0005623), cell part (GO:0044464), organelle (GO:0043226), and membrane (GO:0016020). In the BP, the DEGs were mainly enriched in the metabolic process (GO:0008152) and cellular process (GO:0009987). In the MF, the DEGs were mainly enriched in the catalytic activity (GO:0003824) and binding (GO:0005488) (Supplementary Figure S3; Supplementary Table S6). Moreover, the KEGG enrichment analyses indicated that 131, 109, 136, and 136 pathways were enriched in the above four comparison groups, respectively. Among these pathways, the DEGs were mainly enriched in metabolic pathways (ko01100) and biosynthesis of secondary metabolites (ko01110) (Supplementary Figure S4; Supplementary Table S7).

## Expression analysis of differentially expressed genes related to triterpenoid saponin biosynthesis

In this study, there were 23 candidate DEGs related to TS biosynthesis, including two *CqHMGR*, one *CqHMG3*, two *CqHMGS*, one *CqGGPS*, two *CqMVK*, three *CqAAT1*, one *CqISPD*, one *CqDXS*, one *CqPMK*, one *CqFPS1*, three *CqSQS1*, one *CqSQE3*, two *CqSQE1*, and one *CqCYP716A15*, and one *Cqβ-AS* by Nr functional annotation (Supplementary Table S8). Subsequently, in the GO enrichment analyses, the TS biosynthesis-related DEGs were mainly annotated to the nine GO terms, including triterpenoid biosynthetic process (GO:0016104), acetyl-CoA-acetyltransferase activity process (GO:0006084), phytosteroid biosynthetic process (GO:0016129), steroid biosynthetic process (GO:0006694), isoprenoid metabolic process (GO:0006720), terpenoid metabolic process (GO:0006721), acetyl-CoA-acyltransferase activity (GO:0003988), transferase activity (GO:0046912), and microbody (GO:0042579). Based on these

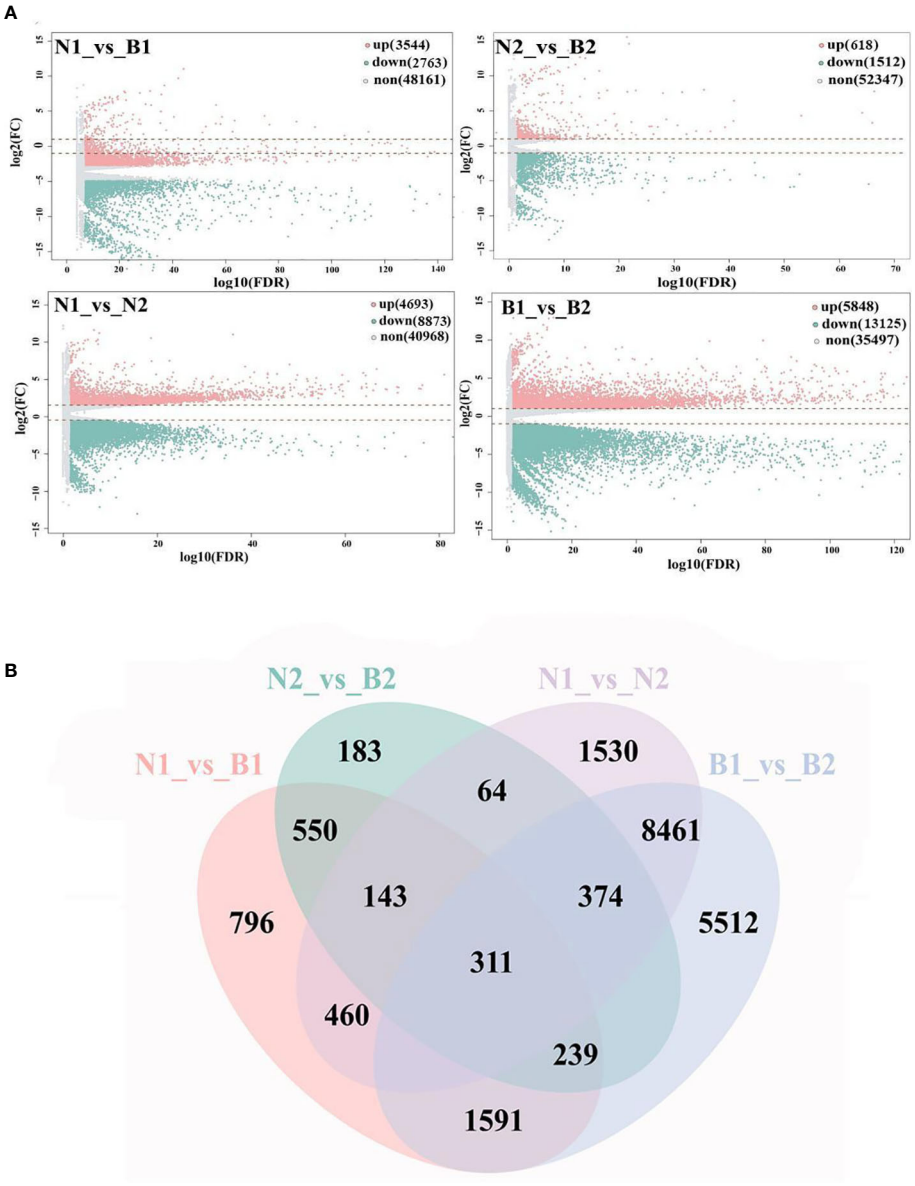



FIGURE 1
Screening and enrichment analyses of differentially expressed genes (DEGs). Volcano map **(A)** shows the DEGs identified from the N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2. Up, down, and none in panel **(A)** mean the upregulated genes, downregulated genes, and no-expressed genes and are indicated in pink, green, and gray, respectively. The values at the top right indicate the number of upregulated, downregulated, and non-differential expression genes. The abscissa shows the false discovery rate (FDR) (-log10), and the ordinate shows the log2(FC). The Venn diagram **(B)** shows the overlapping of DEGs in the above four comparison groups. The numbers on the overlapping non-colored and colored parts indicate the DEGs in a single comparison group and the DEGs shared among different comparison groups.

GO terms, the isoprenoid metabolic process was enriched in most DEGs, including 42, 15, 94, and 122 DEGs in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2, respectively (Figure 2A; Supplementary Table S9). In the KEGG pathway, the TS biosynthesis-related DEGs were mainly annotated with four pathways, consisting of biosynthesis of secondary metabolites (ko01110), triterpenoid biosynthesis (ko00909), terpenoid backbone biosynthesis (ko00900), and steroid biosynthesis (ko00100) (Figure 2B; Supplementary Table S10). The enrichment factors for the triterpenoid biosynthesis were 0.323, 0.077, 0.446, and 0.554, and the number of DEGs for terpenoid backbone biosynthesis was enriched with 18, 4, 36, and 52 DEGs in the four comparison groups, respectively. These results showed that isoprenoid metabolic process, triterpenoid biosynthesis, and terpenoid backbone biosynthesis were the key enrichment pathways related to TS biosynthesis in quinoa.

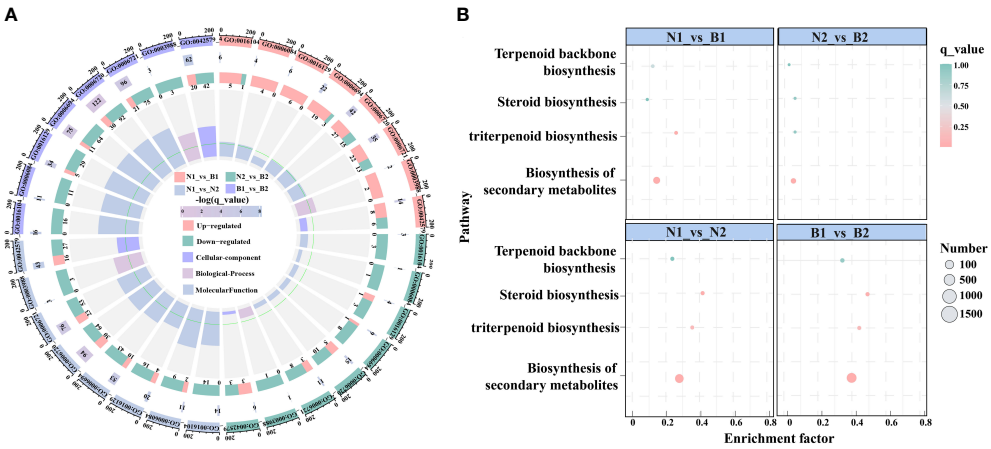



FIGURE 2
Enrichment analyses of the differentially expressed genes (DEGs) related to triterpenoid saponin (TS) biosynthesis. **(A)** shows Gene Ontology (GO) enrichment analyses of DEGs among the above four comparison groups. It is composed of four cycles. Cycle 1 (Cy1, the outermost cycle) means nine GO terms located in the above four comparison groups. Cycle 2 (Cy2, the secondary outer cycle) means the number of DEGs corresponding to GO terms. The different colors of the rectangles below the number are the *q_value* (-log10). Cycle 3 (Cy3, the sub-internal cycle) means the number of upregulated and downregulated genes corresponding to GO Classify2, which were labeled in pink and green, respectively. Cycle 4 (Cy4, the innermost cycle) means the enrichment range and the enrichment factors of GO terms, including cellular components, biological processes, and molecular functions. The green circular line indicates enrichment factor = 1. **(B)** shows the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2 comparison groups. The abscissa and ordinate denote the enrichment factors and pathways, respectively. The size and color of the circles represent the number and *q_value* of DEGs.

## Metabolomics analyses

### Qualitative and quantitative analyses of metabolites

The off-targeted metabolomics analyses were used to perform the qualitative and quantitative analyses of metabolites in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2. The PCA plot of the detected metabolites revealed that the PC1 and PC2 explained 27.1% and 12.3% of the variation, respectively, indicating that the metabolites of the four comparison groups were changed significantly (Supplementary Figure S5). In addition, the result of OPLS-DA showed that Q2Y values in the four comparison groups were 0.919, 0.914, 0.978, and 0.972, respectively. The Q2 value was greater than 0.9, indicating that the model was stable and reliable (Supplementary Figure S6).

## Identification and enrichment analyses of differentially accumulated metabolites

A total of 9,940 DAMs in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2 were identified. There were 6,899, 4,369, 9,214, and 10,697 DAMs in each of the four comparison groups, respectively, among which 3,376, 2,621, 4,052, and 2,763 were upregulated metabolites, and 3,523, 1,748, 5,162, and 5,068 were downregulated metabolites, respectively (Supplementary Table S11). KEGG enrichment analysis revealed that 103, 106, 104, and 107 pathways were enriched in the four comparison groups, of which the identified DAMs were significantly enriched in biosynthesis of antibiotic (ko01130), biosynthesis of secondary metabolites (ko01110), glycerophospholipid metabolism (ko00564), and carotenoid biosynthesis (ko00906) (Supplementary Figure S7).

## Analyses of differentially accumulated metabolites related to triterpenoid saponin biosynthesis

Among those identified DAMs, a total of eight DAMs related to TS biosynthesis were detected, including acetyl-CoA (POS_M810T467), 4-cytidine-5'-diphospho-2-C-methyl-D-erythritol (POS_M544T217), (R)-5-diphosphomevalonate (POS_M326T503), 1-hydroxy-2-methyl-2-butenyl-4-diphosphate (POS_M147T481), farnesal (POS_M221T594), (S)-2,3-epoxysqualene(POS_M409T832), 24-hydroxy-beta-amyrin (POS_M442T757), and ursolic acid (POS_M457T774) in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2 (Figure 3; Supplementary Table S12). The TS biosynthesis-related DAMs were involved in triterpenoid biosynthesis (ko00909) and terpenoid backbone biosynthesis (ko00900). These results indicated that the farnesal and acetyl-CoA were overlapped in all four comparison groups, and 1-hydroxy-2-methyl-2-butenyl-

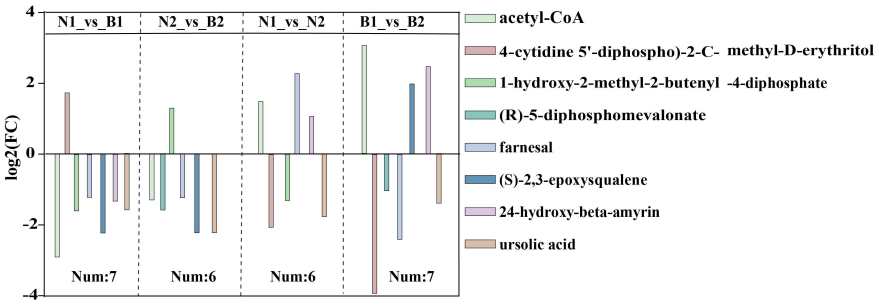



FIGURE 3
The differentially accumulated metabolites (DAMs) in the N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2 comparison groups. Different colors in the legend indicate different types of DAMs. The ordinate indicates the abs (log2FC), and the number indicates the number of DAMs in the above four comparison groups. Rectangular squares represent the different comparison groups.

4-diphosphate and (S)-2,3-epoxysqualene were overlapped in three of the four comparison groups. Those DAMs was identified as the key DAMs based on the transcriptomics–metabolomics joint analysis.

## Transcriptomics–metabolomics joint analysis on triterpenoid saponin biosynthesis

According to the DEGs and DAMs related to TS biosynthesis, the correlation analysis was conducted in the triterpenoid biosynthesis (ko00909) and terpenoid backbone biosynthesis (ko00900). In N1_*vs*_B1, seven DEGs were associated with farnesal and 1-hydroxy-2-methyl-2-butenyl-4-diphosphate, including two positively correlated DEGs and five negatively correlated DEGs. In N2_*vs*_B2, 13 DEGs were associated with farnesal, (S)-2, 3-epoxysqualene, and acteyl-CoA, including three positively correlated DEGs and 10 negatively correlated DEGs. In N1_*vs*_N2, 11 DEGs were associated with farnesal, including five positively correlated DEGs and six negatively correlated DEGs. In B1_*vs*_B2, 13 DEGs were negatively associated with farnesal and (S)-2,3-epoxysqualene. In summary, the result of transcriptomics–metabolomics joint analysis showed that farnesal was the DAM shared in the four comparison groups and correlated with 10 key candidate DEGs related to TS biosynthesis in quinoa, including *CqAAT1*, *CqHMGS*, *CqMVK*, *CqPMK*, *CqMVD2*, *CqDXS*, *CqGGPS*, *CqFPS1*, *CqSQS1*, and *CqSQE1* ($R^2 \geq 0.6$) (Figure 4).

## Verification of differentially expressed genes related to triterpenoid saponin by qRT-PCR

Ten key candidate DEGs related to TS biosynthesis in quinoa were screened by transcriptomics–metabolomics joint analysis and were validated by qRT-PCR detection. In N1_*vs*_B1, six DEGs were verified with upregulated genes, including *CqAAT1*, *CqMVD2*, *CqHMGS*, *CqMVK*, *CqSQS1*, and *CqSQE1*. In N2_*vs*_B2, two DEGs were verified with downregulated genes, including *CqMVK* and *CqFPS1*. In N1_*vs*_N2, 10 DEGs were verified with downregulated genes, including *CqAAT1*, *CqFPS1*, *CqMVD2*, *CqHMGS*, *CqGGPS*, *CqDXS*, *CqSQS1*, *CqSQE1*, and *CqMVK*. In B1_*vs*_B2, eight DEGs were verified with downregulated genes, including *CqAAT1*, *CqFPS1*, *CqHMGS*, *CqMVD2*, *CqMVK*, *CqDXS*, *CqSQS1*, and *CqSQE1*. These results show that *CqMVK* was validated in all of the four comparison groups, and that *CqAAT1*, *CqMVD2*, *CqHMGS*, *CqSQS1*, and *CqSQE1* were validated in three of the four comparison groups (Figure 5; Supplementary Table S13).

# Discussion

## Advantages of triterpenoid saponin content measurement and two quinoa varieties

TS imparts a bitter taste in quinoa, and their extracts can be processed into medicines and cosmetics. In this study, the TS content was measured by the vanilla-glacial acetic acid method because this method is available for the variety classification directly based on the color change, which plays an important role in the development of healthy products and TS drugs (Konishi et al., 2004). Previous studies have shown that the saponin content of high-saponin quinoa varieties is usually 4.7–11.3 mg/g and is called bitter quinoa, and the saponin content of low-saponin quinoa varieties is usually 0.2–0.4 mg/g, known as sweet quinoa (Ruiz et al., 2017). In this experiment, the saponin content classification of 140 quinoa varieties can provide a basis for future breeding work. Saponin was extracted from the

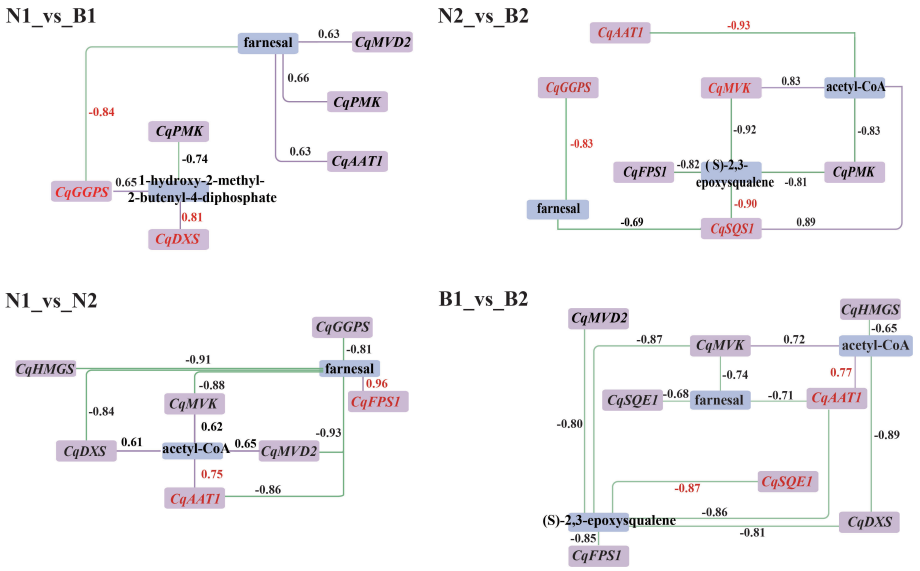



FIGURE 4
Correlation analysis between the candidate differentially expressed genes (DEGs) and differentially accumulated metabolites (DAMs). Purple and blue rectangles indicate the DEGs and DAMs associated with the biosynthesis of triterpenoid saponin (TS), respectively. The labels above the rectangle indicate the specific names of DEGs and DAMs. The purple solid lines and green dashed lines indicate positive correlation and negative correlation between DEGs and DAMs in the above four comparison. The numbers represent the correlation values, where the correlation coefficients and gene names are highlighted in red for the key DEGs and DAMs.

selected high-saponin quinoa varieties, and the extract is developed into saponin drugs, and saponin was removed from low-saponin quinoa varieties to produce edible quinoa. In addition, although the taste of TS in quinoa was unpalatable, the bitterness of quinoa is beneficial for the cultivation of the plants, as saponins prevent herbivory. The TS content of 60 DAF was significantly higher than that of 30 DAF, which was in accordance with previous findings that the TS content in quinoa was increased along with kernel development (Mastebroek et al., 2000). This phenomenon has also been observed in *Panax ginseng* (Zhang et al., 2012) and *Panax notoginseng* (Wu, 2018).

## Regulation of key candidate differentially expressed genes and differentially accumulated metabolites associated with triterpenoid saponin biosynthesis

*HMGS* and *AAT1* are important catalytic genes in the MVA pathway (Luskey and Stevens, 1985; Dyer et al., 2009). It is supposed that overexpression of *HMGS* may lead to enhanced production of *Centellosides yield* in transgenic plants (Afroz et al., 2022). In this study, upregulation of *CqHMGS* promoted the accumulation of acetyl-CoA, and *CqAAT1* coregulated the synthesis of the upstream precursor IPP in the MVA pathway. This result infers that *CqAAT1* and *CqHMGS* promoted the accumulation of acetyl-CoA, thereby accumulating upstream substrates, and finally promoted the formation of downstream products of TS biosynthesis.

*MVK* is an important rate-limiting and ATP-dependent enzyme gene involved in TS biosynthesis (Bohlmann and Keeling, 2008). The increased expression of *NtMVK* improved the metabolite content in the MVA pathway in *Nicotiana tabacum* (Champenoy et al., 1999). In this study, *CqMVK* was upregulated at 30 DAF, while the relative content of (R)-5-diphosphomevalonate also increased. Thus, the upregulated expression of *CqMVK* could lead to the accumulation of (R)-5-diphosphomevalonate during the TS biosynthesis and consequently increased the TS biosynthesis. In addition, *DXS* is the first important rate-limiting enzyme gene in the MEP pathway (Rodríguez-Concepción et al., 2001). In *Camellia sinensis*, *CsDXS1* had different expression levels at different developmental stages (Guo et al., 2018). In this study, *CqDXS* was downregulated in both N1_*vs*_N2 and B1_*vs*_B2, indicating that its differential expression is associated with kernel development in quinoa.

Both *GGPS* and *FPS1* control the accumulation of TS and triterpenoid compounds together (Zhou et al., 2013; Zu et al., 2020). Previous research revealed that *SmGGPS* was significantly differentially expressed at different developmental stages in

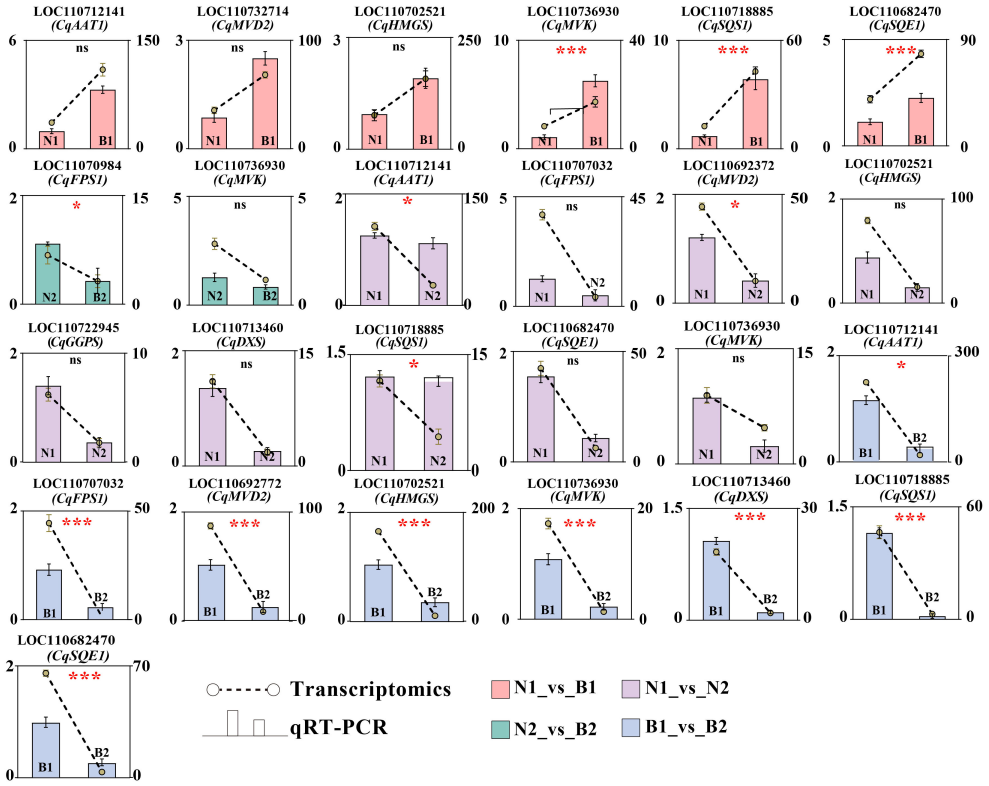



FIGURE 5
Verification of candidate differentially expressed genes (DEGs) by qRT-PCR. The Y-axis as the ordinate denotes the Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) values. The ncbi serial number and name of the DEGs in N1_*vs*_B1, N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2 were marked in pink, green, purple, and blue, respectively. The error bars for qRT-PCR and transcriptomics were marked in black and yellow, respectively. In this figure, "ns," "*," and "***" mean $p > 0.05$, $p < 0.05$, and $p < 0.001$, respectively.

*Salvia miltiorrhiza* (Hua et al., 2014). In this study, *CqGGPS* and *CqFPS1* were upregulated at 30 DAF and downregulated at 60 DAF (Figure 6), providing the changes in gene expression levels along with the development of the kernels. Moreover, a unique mutual relationship between *TaGGPS* and *TaFPS1* was observed on identification and functional analyses of genes and metabolites in *Triticum aestivum* (Zhang, 2012). The result in this study showed that *CqGGPS* and *CqFPS1* were negatively correlated with farnesal, suggesting that these two genes play a negatively regulatory role in the TS biosynthesis in quinoa.

*SQS1* is a key enzyme gene that catalyzes the transformation of farnesal into TS and phytosterols (Shimada et al., 1998). *SQE3* is the regulatory enzyme gene involved in forming the TS framework (Laranjeira et al., 2015). In this study, *CqSQS1* and *CqSQE3* were significantly downregulated in N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2, which promoted the increase in the content of (S)-2, 3-epoxysqualene. It was speculated that *CqSQS1* and *CqSQE3* promoted the accumulation of TS by inhibiting its competing substrates, the phytosterols. This result is mutually verified with the research on *Saccharomyces cerevisiae* (Niu et al., 2012) and *Ashwagandha* (Singh et al., 2015). Meanwhile, *CqCYP716A15* is a key catalytic enzyme gene involved in the formation of TS, which oxidizes inert methyl groups of the triterpenoid framework (Nelson et al., 2008). In this study, the expression of *CqCYP716A15* was upregulated in both N and B at 30 DAF, but it was downregulated at 60 DAF, indicating that the expression of *CqCYP716A15* was regulated along with kernel development.

## Discoveries in the biosynthesis of triterpenoid saponin in quinoa

In this study, most DEGs were downregulated during the TS biosynthesis in N2_*vs*_B2, N1_*vs*_N2, and B1_*vs*_B2. Meanwhile, the expression levels of genes and their related metabolites were also significantly different in the developmental stages. Thus, we speculated that the DEGs in the TS biosynthesis mainly exist in the early developmental periods in quinoa kernels. *AsFPS1* was positively correlated with saponins in *Acanthopanax senticosus* (Zhou et al., 2013), and the upregulated expression of *FPS1* increased the content of saponin

10.3389/fpls.2022.964558

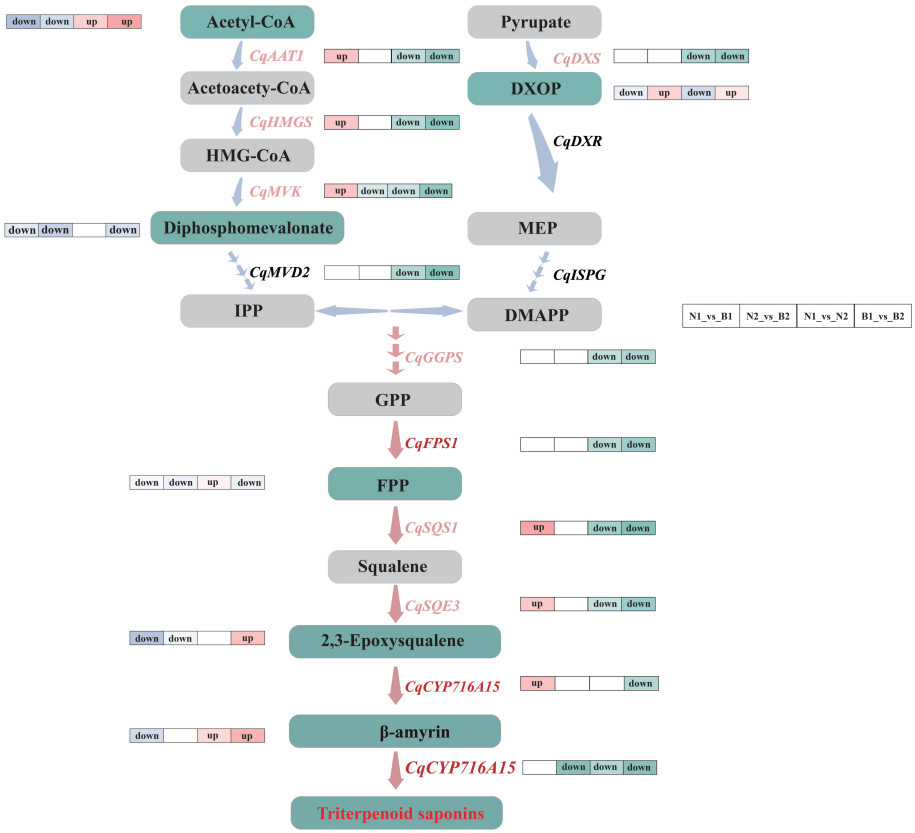



FIGURE 6
Flow-processing chart of the triterpenoid saponin (TS) biosynthesis in quinoa. The light blue and pink arrows between the metabolites indicate the upstream and downstream metabolic pathways. The labels next to the blue arrows are differentially expressed genes (DEGs), in which the key DEGs are labeled with pink. The gray and green rectangles represent the normal and key differentially accumulated metabolites (DAMs). The label "up" next to the genes and metabolites represents the upregulated key candidate DEGs and key DAMs in N1_vs_B1, N2_vs_B2, N1_vs_N2, and B1_vs_B2 comparison groups, and the label "down" next to the genes and metabolites represents the downregulated key candidate DEGs and key DAMs. The pathways marked by pink arrows are the key parts during the TS biosynthesis in quinoa.

in *ginseng* (Kim et al., 2014). Interestingly, *CqFPS1* did not increase the content of farnesal during the biosynthesis of TS in quinoa in this study. On the contrary, a significant negative correlation was observed between *CqFPS1* and farnesal, implying that *CqFPS1* promoted TS by regulating other downstream genes during the TS biosynthesis. Furthermore, *CqCYP71615*, which hydroxylates the TS framework, was not annotated with any relevant metabolic pathway, although the Nr annotation provided that *CqCYP71615* was the key gene in β-amyrin synthesis and that the cytochrome P450 enzyme gene regulates the conversion from β-amyrin into TS. These data infer multiple regulation roles of *CqCYP71615* in the TS biosynthesis in quinoa (Figure 6).

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA836289.

## Author contributions

GM, the corresponding author. GM and WL are responsible for the project planning and the material selection. YZ wrote the manuscript. YZ and YM performed most of the data integration, analysis, and prepared Figures 1, 2 and 3. XL and JZ participated in most of experiment. BL, MZ, CW and LZ prepared Figures 4, 5 and 6. JL revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by the Key Project of Technologies for Industrialization of Quinoa in Bashang

Region (19227527D), the Key Project of High-eficiency Technology Integration and Demonstration Project of Water-saving for Special Crops in Bashang Region of Hebei Province (21327005D), the Key Project of Modern Seed Industry Science and Technology Special Project of the Key Research and Development Program of Hebei Province (19226363D).

## Acknowledgments

We are specially appreciative to all members of National Semi-arid Agricultural Engineering Technology Research Center, Hebei Shijiazhuang, P.R.China for their efforts in breeding of varieties and selection of experimental materials. Thanks for the members of Quinoa Breeding Laboratory of Hebei Agricultural University, Baoding, P.R.China in helping to measure TS content and is grateful to omicsmart platform for the efforts in transcriptomics and Metabolomics sequencing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.964558/full#supplementary-material

## References

Afroz, S., Warsi, Z. I., Khatoon, K., Sangwan, N. S., Khan, F., and Rahman, L. U. (2022). Molecular cloning and characterization of triterpenoid biosynthetic pathway gene HMGS in centella asiatica (Linn.). *Mol. Biol. Rep.* 49 (6), 4555–4563. doi: 10.1007/s11033-022-07300-9

Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313X.2008.03449.x

Champenoy, S., Vauzelle, C., and Tourte, M. (1999). Activity of the yeast mevalonate kinase promoter in transgenic tobacco. *Plant Sci.* 147, 25–35. doi: 10.1016/S0168-9452(99)00091-6

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinf. (Oxford England)* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Dyer, J. H., Maina, A., Gomez, I. D., Cadet, M., Oeljeklaus, S., and Schiedel, A. C. (2009). Cloning, expression and purification of an acetoacetyl CoA thiolase from sunflower cotyledon. *Int. J. Biol. Sci.* 5, 736–744. doi: 10.7150/ijbs.5.736

El Hazzam, K., Hafsa, J., Sobeh, M., Mhada, M., Taourirte, M., El Kacimi, K., et al. (2020). An insight into saponins from quinoa (*Chenopodium quinoa* willd): A review. *Molecules (Basel Switzerland)* 25, 5–22. doi: 10.3390/molecules25051059

Espinoza, C. R., Ruiz, C., Ramos, O. P. F., Solano, M., Quiñonez, G. H., and Mallma, N. E. S. (2021). Optimization of the ultrasoud-assisted extraction of saponins from quinoa (Chenopodium quinoa wild) using response surface methodology. *Acta Sci. Pol. Technol. Aliment.* 20 (1), 17–23. doi: 10.17306/j.Afs.0859g

Fiallos-Jurado, J., Pollier, J., Moses, T., Arendt, P., Barriga-Medina, N., Morillo, E., et al. (2016). Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in chenopodium quinoa leaves. *Plant Sci.* 250, 188–197. doi: 10.1016/j.plantsci.2016.05.015

Freeberg, M. A., and Kim, J. K. (2016). Mapping the transcriptome-wide landscape of RBP binding sites using gPAR-CLIP-seq: Bioinformatic analysis. *Methods Mol. Biol.* 1361, 91–104. doi: 10.1007/978-1-4939-3079-1_6

Gómez-Caravaca, A. M., Iafelice, G., Verardo, V., Marconi, E., Barriga-Medina, N., and Caboni, M. F. (2014). Influence of pearling process on phenolic and saponin content in quinoa (Chenopodium quinoa Willd). *Food Chem*, 157, 174–178. doi: 10.1016/j.foodchem.2014.02.023

Guo, Y. F., Wang, J. Y., Guo, F., and Ni, D. J. (2018). Cloning and expression analysis of *CsDXS1* gene encoding 1-deoxy- d-Xylulose-5-Phosphate synthase in *Camellia sinensis*. *Biotechnol. Bull.* 34, 144–152. doi: 10.13560/j.cnki.biotech.bull.1985.2017-0652

Haralampidis, K., Trojanowska, M., and Osbourn, A. E. (2002). Biosynthesis of triterpenoid saponins in plants. *Adv. Biochem. Eng. Biotechnol.* 75, 31–49. doi: 10.1007/3-540-44604-4_2

Huang, J., Guo, X., Xu, T., Fan, L., Zhou, X., and Wu, S. (2019). Ionic deep eutectic solvents for the extraction and separation of natural products. *J. Chromatogr. A.*, 1598:1–19. doi: 10.1016/j.chroma.2019.03.046

Huang, Z. J., Liu, D. Q., Ge, F., and Chen, C. Y. (2014). Advance in studies on key post-modification enzymes in triterpenoid saponins in biosynthesis. *Acta Botanica Boreali-Occidentalia Sinica* 34, 2137–2144. doi: 10.7606/j.issn.1000.2014.10.2137

Hua, W. P., Song, S. H., Zhi, Y., and Wang, Z. Z. (2014). Cloning and expression analysis of the *SmGGPPS3* gene from *Salvia miltiorrhiza*. *Plant Sci. J.* 32, 50–57. doi: 10.3724SP.J.1142.2014.10050

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32 (Database issue), D277–D280. doi: 10.1093/nar/gkh063

Kim, Y. K., Kim, Y. B., Uddin, M. R., Lee, S., Kim, S. U., and Park, S. U. (2014). Enhanced triterpene accumulation in panax ginseng hairy roots overexpressing mevalonate-5-pyrophosphate decarboxylase and farnesyl pyrophosphate synthase. *ACS Synth. Biol.* 3, 773–779. doi: 10.1021/sb400194g

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Konishi, Y., Hirano, S., Tsuboi, H., and Wada, M. (2004). Distribution of minerals in quinoa (*Chenopodium quinoa* willd.) seeds. *Biosci. Biotechnol. Biochem.* 68, 231–234. doi: 10.1271/bbb.68.231

Laranjeira, S., Amorim-Silva, V., Esteban, A., Arró, M., Ferrer, A., Tavares, R. M., et al. (2015). Arabidopsis squalene epoxidase 3 (SQE3) complements SQE1 and is important for embryo development and bulk squalene epoxidase activity. *Mol. Plant* 8 (7), 1090–1102. doi: 10.1016/j.molp.2015.02.007

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323

Lim, J. G., Park, H. M., and Yoon, K. S. (2020). Analysis of saponin composition and comparison of the antioxidant activity of various parts of the quinoa plant (Chenopodium quinoa willd.). *Food Sci. Nutr.* 8, 694–702. doi: 10.1002/fsn3.1358

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *ELSEVIER* 25, 402–408. doi: 10.1006/meth.2001.1262

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Luo, Z. L., Zhang, K. L., Ma, X. J., and Guo, Y. H. (2016). Research progress in synthetic biology of triterpen saponins. *Chin. Tradit. Herbal Drugs* 47, 1806–1814. doi: 10.7501/j.issn.0253-2670.2016.10.029

Luskey, K. L., and Stevens, B. (1985). Human 3-hydroxy-3-methylglutaryl coenzyme a reductase. conserved domains responsible for catalytic activity and sterol-regulated degradation. *J. Biol. Chem.* 260, 10271–10277. doi: 10.1016/S0021-9258(17)39242-6

Mastebroek, H. D., Limburg, H., Gilles, T., and Marvin, H. J. P. (2000). Occurrence of sapogenins in leaves and seeds of quinoa ( *Chenopodium quinoa* willd). *J. Sci. Food Agriculture* 80, 152–156. doi: 10.1002/(SICI)1097-0010(20000101)80:13.0.CO;2-P

Nelson, D. R., Ming, R., Alam, M., and Schuler, M. A. (2008). Comparison of cytochrome P450 genes from six plant genomes. *Trop. Plant Biol.* 1, 216–235. doi: 10.1007/s12042-008-9022-1

Niu, Y. Y. (2013). *Cloning and expression analysis of the key genes involved in triterpene saponin biosynthesis in panax notoginseng and panax quinquefolium. master's thesis* (China: Peking Union Medical University).

Niu, Y. Y., Luo, H. M., Huang, L. F., Chen, S. L., and He, S. Z. (2012). Research progress of cytochrome P450 in ginsenoside biosynthesis pathway. *World Sci. Technology/Modernization Tradit. Chin. Med. Mater. Medica* 14, 1177–1183. doi: 10.3969/j.issn.1674-3849.2012.01.007

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Rodríguez-Concepción, M., Ahumada, I., Diez-Juez, E., Sauret-Güeto, S., Lois, L. M., Gallego, F., et al. (2001). 1-Deoxy-D-xylulose 5-phosphate reductoisomerase and plastid isoprenoid biosynthesis during tomato fruit ripening. *Plant J.* 27, 213–222. doi: 10.1046/j.1365-313x.2001.01089.x

Ruiying, C., Zeyun, L., Yongliang, Y., Zijia, Z., Ji, Z., Xin, T., et al. (2020). A comprehensive analysis of metabolomics and transcriptomics in non-small cell lung cancer. *PloS One* 15 (5), e0232272. doi: 10.1371/journal.pone.0232272

Ruiz, K. B., Khakimov, B., Engelsen, S. B., Bak, S., Biondi, S., and Jacobsen, S.-E. (2017). Quinoa seed coats as an expanding and sustainable source of bioactive compounds: An investigation of genotypic diversity in saponin profiles. *Ind. Crops Products* 104, 156–163. doi: 10.1016/j.indcrop.2017.04.007

Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A., and Hendriks, M. M. W. B. (2014). Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 10, 361–374. doi: 10.1007/s11306-013-0598-6

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Shimada, H., Kondo, K., Fraser, P. D., Miura, Y., Saito, T., and Misawa, N. (1998). Increased carotenoid production by the food yeast candida utilis through metabolic engineering of the isoprenoid pathway. *Appl. Environ. Microbiol.* 64, 2676–2680. doi: 10.1128/aem.64.7.2676-2680.1998

Singh, A. K., Dwivedi, V., Rai, A., Pal, S., Reddy, S. G., Rao, D. K., et al. (2015). Virus-induced gene silencing of withania somnifera squalene synthase negatively regulates sterol and defence-related genes resulting in reduced withanolides and biotic stress tolerance. *Plant Biotechnol. J.* 13, 1287–1299. doi: 10.1111/pbi.12347

Smith, C. A., Want, E. J., O'maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78 (3), 779–787. doi: 10.1021/ac051437y

Wu, Y. M. (2018). *Study on the accumulation of saponins in different growth stages of panax notoginseng. master's thesis* (China: Yunnan University of Chinese Medicine).

Yao, L., Lu, J., Wang, J., and Gao, W. Y. (2020). Advances in biosynthesis of triterpenoid saponins in medicinal plants. *Chin. J. Nat. Med.* 18 (6), 417–424. doi: 10.1016/s1875-5364(20)30049-2

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11 (2), R14. doi: 10.1186/gb-2010-11-2-r14

Zelena, E., Dunn, W. B., Broadhurst, D., Francis-Mcintyre, S., Carroll, K. M., Begley, P., et al. (2009). Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal. Chem.* 81 (4), 1357–1364. doi: 10.1021/ac8019366

Zhang, Y. (2012). *Molecular and functional analysis of common wheat genes encoding farnesyl pyrophosphate Synthase(Tafps) in sesquiterpene synthetic pathway* (Beijing: Master's thesis. Chinese Academy of Agricultural Sciences).

Zhang, Y. D., Wu, J., Li, Z., Xu, Y. H., Zhang, X. L., and Zhang, L. X. (2012). The effect of the different development stages photosynthetic leaf area on the contents of ginsenosides. *Chin. Agric. Sci. Bull.* 28, 233–239.

Zhao, Y. N., Dang, B., Yang, X. J., Liu, Y., Yao, Y. H., and Chi, D. Z. (2017). Ultrasonic extraction technology of saponins and its antioxidant activity in quinoa from qinghai. *J Agric Food Chem* 38, 12155–12165. doi: 10.13386/j.issn1002-0306.2017.19.009

Zhao, Y. J., and Li, C. (2018). Biosynthesis of plant triterpenoid saponins in microbial cell factories. *J. Agric. Food Chem.* 66, 12155–12165. doi: 10.1021/acs.jafc.8b04657

Zhou, L., Hua, W. P., Yang, Y., and Li, C. Q. (2013). Expression pattern of farnesyI diphosphate synthase (*SmFPPS1*) from salvia miltiorrhiza bunge. *J. Shaaixi Normal Univ. (Natural Sci. Edition)* 41, 70–75. doi: 10.15983/j.cnki.jsnu.2013.02.010

Zu, Y., Prather, K. L., and Stephanopoulos, G. (2020). Metabolic engineering strategies to overcome precursor limitations in isoprenoid biosynthesis. *Curr. Opin. Biotechnol.* 66, 171–178. doi: 10.1016/j.copbio.2020.07.005

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpls.2022.942148



OPEN ACCESS

EDITED BY
Jens Rohloff,
Norwegian University of Science and Technology, Norway

REVIEWED BY
José S. Câmara,
Universidade da Madeira, Portugal
Maurício Bonatto Machado de Castilhos,
Minas Gerais State University, Brazil

*CORRESPONDENCE
Iva Šikuten
isikuten@agr.hr

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 12 May 2022
ACCEPTED 25 August 2022
PUBLISHED 20 October 2022

CITATION
Šikuten I, Štambuk P, Tomaz I, Marchal C, Kontić JK, Lacombe T, Maletić E and Preiner D (2022) Discrimination of genetic and geographical groups of grape varieties (*Vitis vinifera* L.) based on their volatile organic compounds.
*Front. Plant Sci.* 13:942148.
doi: 10.3389/fpls.2022.942148



# Discrimination of genetic and geographical groups of grape varieties (*Vitis vinifera* L.) based on their volatile organic compounds

Iva Šikuten[1,2]*, Petra Štambuk[1,2], Ivana Tomaz[1,2], Cecile Marchal[3], Jasminka Karoglan Kontić[1,2], Thierry Lacombe[4], Edi Maletić[1,2] and Darko Preiner[1,2]

[1]Department of Viticulture and Enology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia, [2]Centre of Excellence for Biodiversity and Molecular Plant Breeding, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia, [3]Grapevine Biological Resources Center, INRAE, Unité Experimental Domaine de Vassal, University of Montpellier, Marseillan, France, [4]AGAP, University of Montpellier CIRAD, Institut Agro, Montpellier, France

Grape volatile organic compounds (VOCs) play an important role in the winemaking industry due to their contribution to wine sensory characteristics. Another important role in the winemaking industry have the grapevine varieties used in specific regions or countries for wine production. Due to the high variability of grapevine germplasm, grapevine varieties are as classified based on their genetic and geographical origin into genetic-geographic groups (GEN-GEO). The aim of this research was to investigate VOCs in 50 red grapevine varieties belonging to different GEN-GEO groups. The study included varieties from groups C2 (Italy and France), C7 (Croatia), and C8 (Spain and Portugal). The analysis of VOCs was performed by SPME-Arrow-GC/MS directly from grape skins. The analyzed VOCs included aldehydes, ketones, acids, alcohols, monoterpenes, and sesquiterpenes. The most abundant VOCs were aldehydes and alcohols, while the most numerous were sesquiterpenes. The most abundant compounds, aldehydes and alcohols, were found to be (*E*)-2-hexenal, hexenal, (*E*)-2-hexen-1-ol, and 1-hexanol. Using discriminant analysis, the GEN-GEO groups were separated based on their volatile profile. Some of the individual compounds contributing to the discrimination were found in relatively small amounts, such as benzoic acid, (*E,E*)-2,4-hexadienal, 4-pentenal, and nonanoic acid. The groups were also discriminated by their overall volatile profile: group C2 was characterized by a higher content of aldehydes and alcohols, and group C8 was characterized by a higher content of sesquiterpenes and acids. Group C7 was characterized by all low amount of all classes of VOCs.

KEYWORDS

**volatile organic compounds, grapevine varieties, GEN-GEO groups, discrimination, volatile profiles**

## Introduction

According to OIV (2017), there are 6,000 grapevine varieties in the world, which were developed during the long domestication history of grapevine by the combined actions of selection, breeding, admixture, and migration (Bacilieri et al., 2013). Traditionally, grapevine varieties are classified based on their usage in wine and table grapevine varieties. However, the development of biochemical and molecular markers enabled the assessment of the genetic diversity and structure of grapevine varieties. Using different molecular markers and structure analysis, many authors have classified grapevine varieties into different genetic groups (Aradhya et al., 2003; Arroyo-Garcia et al., 2006; Bacilieri et al., 2013; Laucou et al., 2018), which overlap with certain geographic areas and confirm the accepted classification of Negrul et al. (1946). The GEN-GEO groups of varieties included in this paper are based on the work of Laucou et al. (2018). The authors divided a large sample of 783 grapevine varieties into eight GEN-GEO groups using single nucleotide polymorphism (SNP) markers and structure analysis. The groups are C1 (cultivars from Western and Central Europe, and Iberian Peninsula), C2 (similar to group C1 with the addition of wine cultivars from the Italian peninsula), C3 (wine and table cultivars from the Iberian Peninsula), C4 (table cultivars from Western Europe), C5 (table cultivars from Eastern regions), C6 (wine cultivars from Eastern Mediterranean and Caucasus regions), C7 (cultivars from the Balkan region), C8 (mainly Iberian cultivars, as well as cultivars from Western Europe, the Balkan region, and the Italian peninsula). This classification was used in our previous work (Šikuten et al., 2021b), where we investigated polyphenolic profiles of GEN-GEO groups and used these profiles in discriminant analysis.

In the last 40 years, almost 1,000 volatile compounds have been identified in wine, with a content ranging from μg/L up to mg/L (Pons et al., 2017). VOCs are secondary metabolites responsible for grape and wine sensory properties. They can come from several sources: directly from grape berry, from alcoholic fermentation through yeast and bacterial metabolism, or from aging (Francis and Newton, 2005). VOCs that come directly from grapes are the product of the grapes' own metabolism, therefore influenced by grape variety, climate conditions, and viticultural practices (Bretón et al., 2020). The main groups of volatile compounds in grapes are terpenoids (monoterpenes, sesquiterpenes), $C_{13}$-norisoprenoids, alcohols, carbonyls, and methoxypyrazines. Furthermore, all these compounds can be found in free form, as volatile molecules, or in glycosidically bound form, as non-volatile molecules.

Monoterpenes are a class of compounds that give rise to Muscat's characteristic floral aroma. In grape berries, they can be found both in skins and pulp with different distributions, depending on the compound (Luan and Wust, 2002), and can be found as free volatile compounds or as non-volatile glycosidically bound compounds (Ilc et al., 2016). Most of the wine monoterpenes contribute toward floral and citrusy notes (Ilc et al., 2016). The main representatives and most odoriferous are monoterpene alcohols, notably linalool, citronellol, $\alpha$-terpineol, nerol, geraniol, and hotrienol (Gonzalez-Barreiro et al., 2015). Sesquiterpenes in grapes and wines have received less attention due to their lower volatility and higher detection threshold (Lin et al., 2019). In a recent review by Li et al. (2020), the authors extensively summarized the presence and impact of 97 sesquiterpenes in grapes and wines. Despite the numerous compounds present in grapes, the most significant sesquiterpene is rotundone, a compound responsible for the peppery aroma of Australian shiraz wines (Wood et al., 2008). Norisoprenoids are a diverse group of widespread compounds derived from the oxidative breakdown of carotenoids (Rienth et al., 2021). Similar to monoterpenes, their aroma is mostly described as floral or fruity, and the majority of compounds can be found as non-volatile glycosides (Ilc et al., 2016; Lin et al., 2019). The most important norisoprenoids are $\beta$-damascenone, $\beta$-ionone, vitispirane, and TDN (1,1,6-dimethyl-1,2-dihydronaftalene; Gonzalez-Barreiro et al., 2015). Methoxypyrazines are a class of nitrogenated heterocyclic compounds found in many plants, contributing to aromas described as herbaceous, green, vegetal, and earthy (Ilc et al., 2016; Rienth et al., 2021). These compounds are characterized by an extremely low odor detection threshold, and their excessive levels may result in unacceptable green and unripe aromas that negatively affect wine quality (Lei et al., 2018). $C_6$ and $C_9$ alcohols and aldehydes are products of the lipoxygenase pathway and have the characteristic green aroma that can be a negative contributor to wine aroma (Ilc et al., 2016; Lin et al., 2019). However, the levels of these compounds in wines are mainly modulated by the winemaking process. This happens when the more odorous $C_6$ aldehydes are reduced to less odorous $C_6$ alcohols by yeast activity during alcoholic fermentation (Rienth et al., 2021).

There is a lot of research regarding the aromatic profiles of grape varieties, the influence of climate conditions on the accumulation of VOCs, and the effect of winemaking practices on the sensory properties of wine. However, only recently has research started to elucidate the genetic mechanism behind the biosynthesis and metabolism of grape VOCs (Lin et al., 2019). The grape VOCs and their precursors originate from multiple biosynthetic pathways and can undergo enzyme-catalyzed modifications and spontaneous chemical transformations (Dunlevy et al., 2009; Lin et al., 2019), thus making the research on the biosynthesis of VOCs extremely difficult and complex. Hence, the volatile compounds are not usually used for chemotaxonomic purposes. One study that explores VOCs on a germplasm level is by Yang et al. (2009), who evaluated the composition and concentration of volatiles in berries of 42 grape cultivars belonging to seven genotypic groups. Based on the aromatic profile and PCA analysis, the authors divided cultivars into three groups: 1. *V. labrusca* and its hybrids with *V. amurensis* or *V. vinifera*; 2. *V. vinifera* with muscat aroma; 3. others, including *V. vinifera* without muscat

aroma plus *V. amurensis*, and hybrids between *V. vinifera* and *V. thunberghii* or *V. amurensis*. The authors also observed that quantitative variations of VOCs were influenced by the growing season, but the qualitative volatile composition of the cultivars was consistent.

Since research of VOCs on *V. vinifera*'s germplasm level is rare, the aim of this study is to analyze volatile profiles of red varieties from different GEN-GEO groups and to determine the differences between these groups based on the analyzed volatile profiles. Furthermore, using discriminant analysis, we wanted to determine the individual compounds contributing to the differences between GEN-GEO groups. This research included 50 grapevine varieties with different genetic and geographic origins. As mentioned, the GEN-GEO groups are based on the work of Laucou et al. (2018).

# Materials and methods

## Grape samples

Grape samples were collected during the 2019 growing season at proximity of full ripeness. During the season, maturity controls were performed by measuring the sugar content and visually by checking the seed color. Samples were collected when the sugar content stopped increasing and the seeds were brown in color. The samples were collected from 21 August to 30 August. Only true-to-type accessions were chosen to represent three GEN-GEO groups and five countries of origin, and were collected in a single collection in the INRAE Grape Germplasm Repository 'Domaine de Vassal'. The grapevine varieties are grown on sandy soil on their own roots.

For each grape variety, five clusters were randomly chosen from three vines. Berries with attached pedicels were removed from the clusters using small scissors. One hundred and fifty berries were randomly chosen and divided into three batches of 50 berries each. Each batch was considered as one replication, resulting in three replications for each grape variety. The same replications were used for the analysis of free VOCs. Until analysis, the samples were stored at −20°C. The remainder of berries (~300) were removed from the clusters, divided into three uniform batches, and manually crushed to obtain juice for analysis of basic parameters such as total soluble solids, titratable acidity, and pH value. The basic parameters were measured according to the methods of the International Organization of Vine and Wine (OIV, 2019). The results of the analysis of basic parameters have been published by Šikuten et al. (2021b) as Supplementary material.

The selection of varieties was based on their genetic and geographic origin. The groups are as follows: C2 (10 grape varieties from France, 10 grape varieties from Italy), C7 (10 grape varieties from Croatia), and C8 (10 grape varieties from Spain, 10 grape varieties from Portugal). The country of origin was confirmed by the *Vitis* International Variety Catalogue (VIVC). The list of varieties, their country of origin, and GEN-GEO groups are presented in Supplementary Table 1.

## Analysis of volatile organic compounds

### SPME-arrow extraction of free VOCs

For the analysis of VOCs, three batches of 50 berries (representing three replications) for each grape variety were used. The grape skins were manually removed from the frozen berries and freeze-dried. To obtain powder, the skins were ground using a MiniG Mill (SPEX Sample Prep, Meutchen, United States) and were stored at −20°C until analysis.

SPME-Arrow extraction was carried out based on the method described by Šikuten et al. (2021a). Briefly, SPME-Arrow extraction was conducted using an RSH Triplus autosampler (Thermo Fisher Scientific Inc., Brookfield, United States). Each sample weight of 100 mg was placed in a 20 ml headspace screw-top vial with a cap consisting of a PTFE/silicone septum.

The sorption conditions were as follows: the sample was incubated at 60°C for 20 min, and then the SPME-Arrow fiber DVB/CWR/PDMS (120 µm × 20 mm; Thermo Fisher Scientific Inc., Brookfield, United States) was exposed for 49 min. Then, the fiber was inserted into a GC injector port operating in splitless mode and desorbed at 250°C for 10 min.

### GC–MS analysis

Separation and detection of the analytes was carried out by TRACE™ 1300 Gas Cromatographer coupled with ISQ 7000 TriPlus quadrupole mass spectrometer (Thermo Fisher Scientific Inc., Bartlesville, OK, United States) equipped with TG-WAXMS A capillary column (60 m × 0.25 mm × 0.25 µm film thickness; Thermo Fisher Scientific Inc., Bartlesville, OK, United States). The volatile compounds injected into the inlet were delivered to the column in splitless mode, and helium was used as a carrier gas at a constant flow rate of 1 ml/min. The oven temperature program was as follows: the initial temperature of 40°C was maintained for 5 min, increased by 2°C every minute until the temperature reached 210°C, and held for 10 min. The MS spectra was recorded in the electron impact ionization mode (EI) at an ionization energy of 70 eV. The mass spectrometer performed in full scan mode in the range of 30–300 *m/z*. The obtained data was processed using Chromeleon™ Data System (Thermo Fisher Scientific Inc., Bartlesville, OK, United States). Identification of volatile compounds was achieved by comparing the recorded mass spectrum with the data available in the Wiley Registry 12th Edition/NIST Spectral Library. The Retention Index (RI) was calculated using alkane standards $C_8$–$C_{20}$ (Sigma Aldrich, St. Louis, United States) according to the equation in Song et al. (2019) and compared with results previously reported in the literature (Babushok and Zenkevich, 2009; Babushok et al., 2011). The results are presented in Supplementary Table 2. All the results are expressed as absolute peak areas (×$10^6$). Figure 1 represents a typical chromatogram.

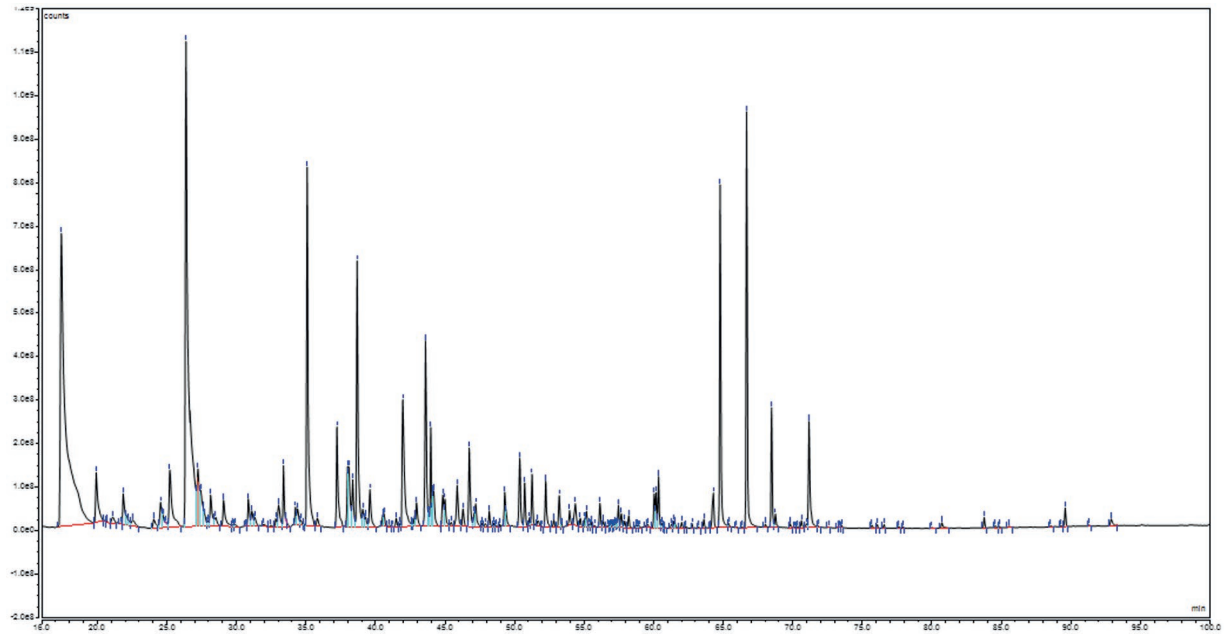



FIGURE 1
Typical GC–MS chromatogram of analyzed samples.

## Statistical analysis

The individual VOCs were analyzed using one-way ANOVA, and the differences between given means for countries and GEN-GEO groups were evaluated by Duncan's multiple range test at a confidence level of 95% ($p < 0.05$). The data reported in all the tables are the average triplicate observations. Discrimination among five groups of varieties based on country of origin and among three different GEN-GEO groups was performed by discriminant analysis (DA) stepwise forward model using the average grape skin volatile profiles of varieties to define multivariate difference among these groups, as well as to define the contribution of VOCs in discrimination. Statistical analysis was carried out using XLSTAT (Addinsoft, 2021, New York, United States).

# Results

## Volatile profile

In Supplementary Tables 3, 4 are given mean values of analyzed volatile organic compounds (VOCs) for all varieties, countries, and GEN-GEO groups. In total we analyzed 119 volatile compounds, among which are 50 sesquiterpenes, 28 alcohols, 16 aldehydes, 8 acids, 8 monoterpenes, 3 ketones, 3 lactones, 2 esters, and 2 $C_{13}$-norisoprenoids.

## Carbonyls, alcohols, acids, and esters

Carbonyl compounds (aldehydes and ketones) represent the most abundant group of VOCs, representing almost 50% of total VOCs. The most abundant compounds were (*E*)-2-hexenal, representing 60.83% of total carbonyls, and hexanal, representing 27.08% of total carbonyls. The compounds nonanal, benzaldehyde, and 2,4-hexadienal were found in higher abundance, while other compounds were found in low abundance or were detected in a small number of varieties (for example (*E,Z*)-2,6-nonadienal and phenylacetaldehyde). Varieties with the highest abundance of total carbonyls were Dobričić, Cahors, and Trepat. These varieties also had the highest abundance of (*E*)-2-hexenal and hexanal. On the other hand, the lowest abundance of total carbonyls were found in varieties Vranac, Icod do Vinao (Listan negro), and Rudežuša. In addition, these varieties had the lowest abundance of (*E*)-2-hexenal and hexanal. The $C_9$ compound 2-nonenal was found in all varieties, except in Mencía and Rudežuša. In the case of (*E,Z*)-2,6-nonadienal, it was identified in seven varieties, namely Ancellotta, Servanin, Petit Verdot, Tinto Cão, Cabernet franc, Montepulciano, and Barbera. The GEN-GEO groups a had similar abundance of total carbonyl compounds. Similar to the varieties, the most abundant compounds were (*E*)-2-hexenal and hexanal, but the GEN-GEO groups did not differ significantly in the content of these compounds. However, they did differ in the abundance of compounds found in small amounts, such as 4-pentenal, decanal, benzaldehyde, and 6-methyl-5-heptene-2-one. Similar to the GEN-GEO groups, there were no significant differences in the abundance of total carbonyl compounds, (*E*)-2-hexenal, and hexanal based on country of origin.

Alcohols were the second most abundant class of compounds after carbonyls. The most abundant alcohol compounds were (*E*)-2-hexen-1-ol, representing 28.87% of total alcohols, followed by 1-hexanol, representing 24.6% of total alcohols, and benzyl alcohol, representing 15.62% of total alcohols. Other alcohol compounds found in higher abundance were (*Z*)-3-hexen-1-ol, isoamyl alcohol, and phenyl ethanol. Alcohol compounds represented in small abundance or detected in a small number of varieties included 1-nonanol and 2-nonanol. The varieties with the highest abundance of total alcohols were Petit Verdot, Tannat, and Touriga nacional. (*E*)-2-hexen-1-ol, as the most abundant alcohol, had the highest abundance in varieties Petit Verdot and Tinto Cão, while Vranac had a considerably lower abundance than the other varieties, followed by Garnacha. A considerably higher abundance of 1-hexanol was found in the variety Alvarelhão. The Tannat variety had the next highest abundance of 1-hexanol, at one times lower than Alvarelhão. The results for the GEN-GEO groups are similar to those of varieties, with (*E*)-2-hexen-1-ol and 1-hexanol as the most abundant compounds. The content of the above-mentioned compounds was similar for groups C2 and C8, while group C7 had a significantly lower abundance. The same results were found for total abundance of alcohols. When looking at countries of origin, the smallest abundance of total alcohols was found in Croatian varieties, followed by a similar content in Spanish and Italian varieties. The highest content of total alcohols was found in French and Portuguese varieties. It is a similar situation with the most abundant compounds, (*E*)-2-hexen-1-ol and 1-hexanol.

Only two compounds belonging to esters were found, ethyl hexanoate and ethyl octanoate. Ethyl hexanoate was only found in two Spanish varieties, Sumoll Tinto and Carignan. Ethyl octanoate was identified in all varieties, except Barbera, while Mancens had the highest abundance.

Acids are another class of volatile compounds identified in the analyzed grapevine varieties. This class of compounds represented only 4% of total VOCs, with eight compounds identified. The most abundant acid was hexanoic acid, representing 62.49% of total acids, followed by (*E*)-2-hexenoic acid, representing 22.34% of total acids. Other acids found in higher abundance were nonanoic and benzoic acids. The variety with a significantly higher abundance of total acids was Trepat, followed by Mourisco tinto and Barbera. These varieties also had the highest abundance of hexanoic and (*E*)-2-hexenoic acid. The smallest abundance of total acids was found in varieties Lasina and Ninčuša, which also had the smallest abundance of the most abundant acids. The results for the GEN-GEO groups followed the results of the individual varieties. Group C7 had the lowest abundance of total acids, hexanoic and (*E*)-2-hexanoic acid, while groups C2 and C8 did not differ in the abundance of mentioned parameters. In the context of countries of origin, Croatian varieties had the lowest abundance of all identified acids and total acids, while other countries had similar abundance.

## Terpenoids

In our samples, the representatives of the terpenoid family are monoterpenes, sesquiterpenes, and $C_{13}$-norisoprenoids. Only eight monoterpene compounds were identified in analyzed samples and in relatively small quantities. The variety with a significantly higher abundance of monoterpenes was Dolcetto, followed by Terrano and Baga, while the lowest abundance was found in varieties Crljenak kaštelanski, Sušćan, and Ninčuša. The most abundant monoterpene was *β*-pinene, followed by 2-pinen-4-one and *β*-ocimene. Among monoterpene alcohols, only linalool was identified but in small quantities. Regarding the GEN-GEO groups, the C7 group had the lowest abundance of total monoterpenes, while the C2 and C8 groups had similar abundance. Similar results were shown for countries of origin. Croatian varieties mostly differed from the others by having a low abundance of analyzed monoterpenes. Varieties from other countries differed slightly in the abundance of monoterpenes.

In analyzed varieties, sesquiterpenes represented the third most abundant group of VOCs, comprising 21.6% of total VOCs. The most abundant compound was ylangene, followed by *β*-copaene and *β*-burbonene. Some of the compounds were found in only a few varieties, like isospathulenol, (*Z*)-*β*-farnesene, or *α*-farnesene. The varieties with the highest abundance of sesquiterpenes were Dolcetto and Baga, while the lowest abundance was recorded for Vranac, Mancin, and Tinto Cão. The highest abundance of ylangene was recorded in Trepat, Baga, and Terrano, while the lowest abundance was recorded in Montepulciano, Manseng, and Ninčuša. The GEN-GEO groups did not differ significantly in the abundance of total sesquiterpenes. The content of ylangene, the most abundant compound, was significantly higher in group C8, while the abundance in groups C2 and C7 was similar. When looking at the countries of origin, the abundance of total sesquiterpenes did differ significantly. The highest abundance was recorded for Italian and Portuguese varieties, followed by Spanish varieties. Croatian and French varieties had the lowest abundance of sesquiterpenes, which is in accordance with the results presented for varieties and GEN-GEO groups. The content of ylangene also follows the results presented for GEN-GEO groups. Hence, the Croatian and French varieties showed the lowest abundance, while the highest abundance was recorded for Spanish and Portuguese varieties, followed by Italian varieties.

In analyzed varieties, only two $C_{13}$-norisoprenoids were detected, (*E*)-β-ionone and TDN. The varieties with a considerably higher abundance of $C_{13}$-norisoprenoids included Dolcetto and Tinto Cão, while the lowest abundance was recorded for Mancens and Soić. Among the GEN-GEO groups, group C7 had the lowest abundance of analyzed norisoprenoids. Groups C2 and C8 had a similar content of analyzed norisoprenoids. The results based on countries of origin followed those reported for GEN-GEO groups. Hence, the Croatian varieties had the lowest abundance of analyzed norisoprenoids, and varieties from other countries did not differ in the content of norisoprenoids.

## Discriminant analysis

The discriminant analysis (DA) of varieties based on their country of origin is presented in Figure 2, which shows the distribution of the varieties in the space defined with the first two canonical factors. The factors explained 96.59% of variability (F1 90.03%, F2 6.56%). In Supplementary Table 5, the correlations of variables/factors are shown, which represent the compounds contributing the most to the discrimination. As presented in Figure 1, the countries are distinctly separated based on their volatile profile. Croatian and Spanish varieties are distinctly separated and are located in the first and third quadrant, respectively. It can be seen that Croatian varieties are discriminated by the compounds benzoic acid, isoamyl alcohol, heptanal, hexanal, and (*S,R*)-2,3-butanediol. Spanish varieties are discriminated by *exo*-2-hydroxycineole, pentanoic acid, $\gamma$-undelactone, and isoamyl alcohol. French and Italian varieties are located in the second quadrant, but they are clearly separated. These varieties are differentiated mostly by alcohols, chiefly 3-methoxy-1-butanol, (*Z*)-2-penten-1-ol, (*E*)-2-hexen-1-ol, and 2-nonanol. Furthermore, two other compounds had a high correlation with certain varieties, namely 4-pentenal in French varieties, and nonanoic acid in Italian varieties. Portuguese varieties were located in the fourth quadrant and were differentiated by the presence of linalool, (*E*)-2-hexenoic acid, ylangene, and 1-butoxy-2-propanol.

Figure 3 represents the results of DA for GEN-GEO groups, that is, the distribution of varieties in the space defined with the first two canonical factors, explaining 100% of variability (F1 63.71%, F2 36.29%). In Supplementary Table 3 the variable/factor correlations are presented. Similar to the groups based on countries of origin, the GEN-GEO groups were also distinctly separated. Group C7 located in the first quadrant, containing Croatian varieties, was differentiated by the presence of compounds benzoic acid, heptanal, (*E*)-$\alpha$-bergamontene, and both (*S,R*)- and (*R,R*)-2,3-butanediol. The second quadrant contains group C2, discriminated by the presence of compounds (*E*)-2-hexen-1-ol, 2,4-dimethyl-3-pentanol, (*E,E*)-2,4-hexadienal, and (*E*)-2-hexenal. Group C2 consists of French and Italian varieties, with high correlations of 4-pentenal and nonanoic acid. Group C8 is located in third and fourth quadrant, near the y axis, and was discriminated by the presence of compounds *exo*-2-hydroxycineole, isoamyl alcohol, (*E*)-$\beta$-ionone, linalool, (*E*)-2-hexenoic acid, and phenyl ethanol.

## Discussion

Fifty grapevine varieties, representing different GEN-GEO groups and countries of origin, varied greatly in their volatile profile. VOCs in grapes are represented by different groups of compounds including aldehydes, alcohols, esters, and acids. Although these compounds are mostly produced during fermentation by yeast metabolism, many compounds originate directly from grapes. In our samples, the most abundant are $C_6$ aldehydes; however, some of the most powerful aroma compounds have nine carbon atoms, such as 2-nonenal or (*E,Z*)-2,6-nonadienal, contributing, like all aldehydes, to green and herbaceous aromas (Ferreira and Lopez, 2019). Furthermore, their content during ripening can be extremely low (Zhu et al., 2012), which can also be seen in our results. Alcohols, like aldehydes, contribute to green and herbaceous aromas. Besides being important aroma contributors, alcohols could also be used as varietal markers. In a study on $C_6$ alcohols, Oliveira J. M. et al. (2006) showed that 1-hexanol, (*E*)-3-hexanol, and (*Z*)-3-hexanol could be used in

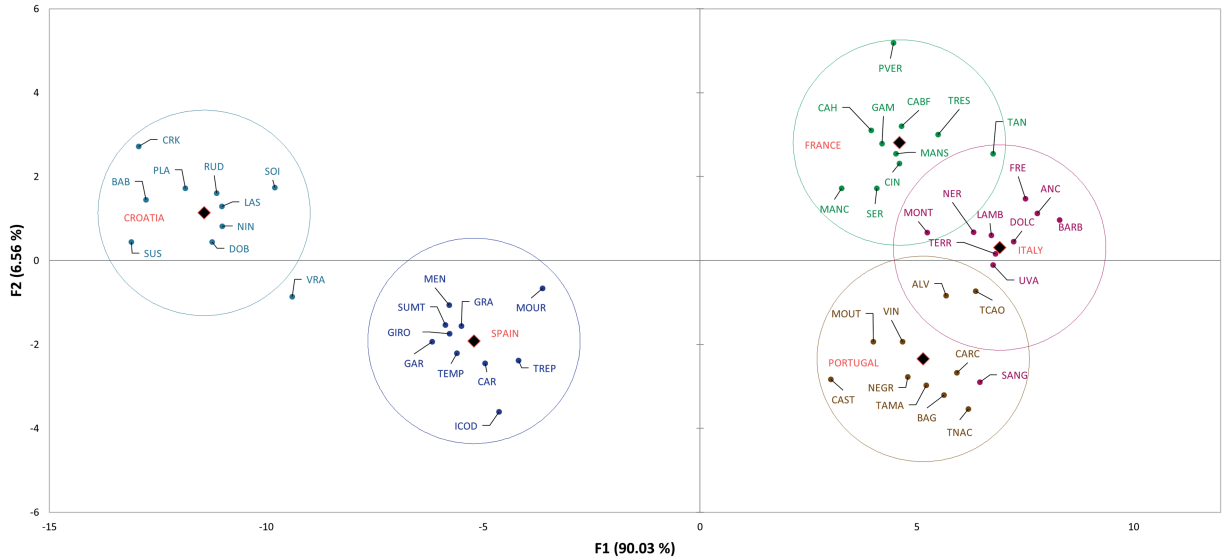



FIGURE 2
The scatter plot of discriminant analysis representing cultivars from different countries.

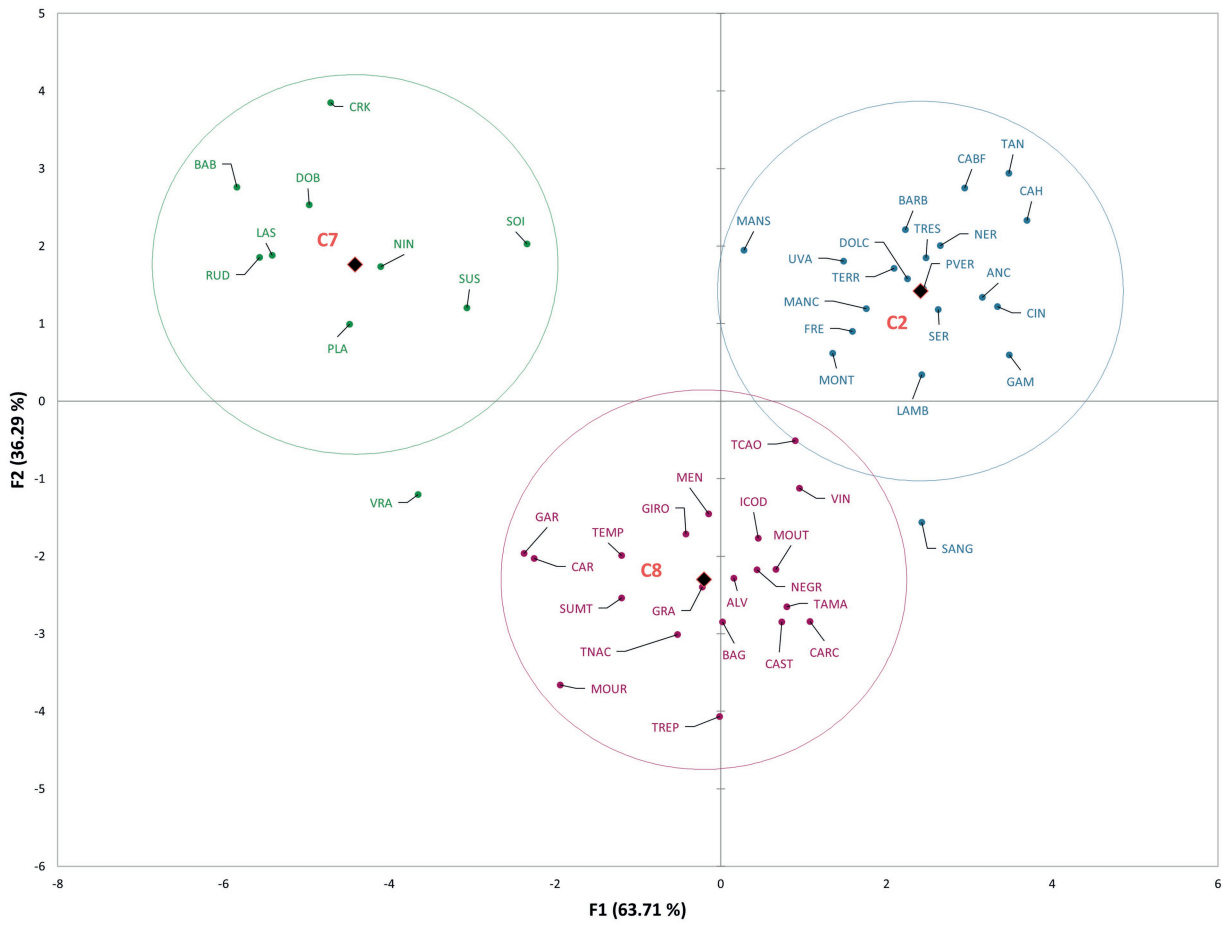



FIGURE 3
The scatter plot of discriminant analysis representing cultivars from different GEN-GEO groups.

discrimination of varieties. Regarding esters, they are powerful odorants, with a very low odor threshold (Pineau et al., 2009). In the analyzed samples, only two esters were identified, belonging to the group of ethyl esters: ethyl hexanoate contributed to the apple-like and aniseed aromas, and ethyl octanoate contributed to the sour apple aroma (Saerens et al., 2010). Most of the esters are produced during fermentation by yeast metabolism, and the production is dependent on the presence of precursors in the must (Saerens et al., 2010; Dennis et al., 2012; Boss et al., 2015). Furthermore, in a study on Cabernet Sauvignon, the esters were identified from early developmental stages of grape berries. However, their concentrations significantly dropped at véraison (Kalua and Boss, 2009). Thus, this could be an explanation as to why only two compounds were identified. All these classes of compounds are mutually connected through biosynthetic pathways and can be transformed into each other. In grapes, $C_6$ and $C_9$ alcohols and aldehydes are products of lipoxygenase pathway (Lin et al., 2019). In this pathway, the fatty acids are oxidized by lipoxygenases (LOX) and modified by hyperoxide lyase to form aldehydes (Dunlevy et al., 2009). The most abundant acids in our samples were (*E*)-2-hexenoic and hexanoic acids, which through the above-mentioned modifications can yield aldehydes (*E*)-2-hexenal and hexanal, the most abundant aldehydes. The produced aldehydes can be further metabolized by alcohol dehydrogenase (ADH) to form corresponding alcohols (Schwab et al., 2008), (*E*)-2-hexen-1-ol and 1-hexanol. The production of ethyl esters is dependent on the concentration of fatty acid precursors (Saerens et al., 2010), which in our samples were hexanoic and octanoic acids. Although hexanoic acid was the most abundant acid, it did not yield esters, except in two varieties. On the other hand, octanoic acid was not identified in grape samples, but the corresponding ester, ethyl octanoate, was identified. Thus, it could be hypothesized that the small quantities produced were transformed into esters.

Terpenoids are the most extensively studied group of VOCs in *Vitis vinifera* grapes and are an extremely diverse and abundant group (Yu and Utsumi, 2009; Gonzalez-Barreiro et al., 2015). The major representatives are monoterpenes, sesquiterpenes and norisoprenoids, which are all responsible for fruity (citric) and floral aromas of grapes and wines (Gonzalez-Barreiro et al., 2015; Lin et al., 2019). As presented, there were only 8 monoterpene compounds identified in analyzed samples. This is not surprising since red varieties are not characterized by high levels of terpenes (Hernandez-Orte et al., 2015; Yuan and Qian, 2016; Luo et al., 2019). Furthermore,

the levels of bound monoterpenes, which were not included in this research, are usually much higher than levels of free monoterpenes (Li et al., 2017; Yue et al., 2020). Sesquiterpenes in grapes and wines have received less attention due to their lower volatility and detection threshold (Lin et al., 2019). In total we identified 50 sesquiterpene compounds, which makes sesquiterpenes the most numerous class of VOCs. The most known compound is rotundone, which was not identified in our samples. Regarding other sesquiterpenes, May and Wust (2012) showed that grapes emit numerous sesquiterpene hydrocarbons and the sesquiterpene profile depends on grape variety and developmental stage. Norisoprenoids are another class of terpenoid compounds found in grapes and wines. Among these carotenoid-derived compounds, $C_{13}$-norisoprenoids are the most widespread (Winterhalter and Rouseff, 2002). In analyzed samples only two compounds belonging to norisoprenoids were identified and in small quantities. However, other research on red varieties reported the presence of other norisoprenoids, such as $\beta$-damascenone or $\alpha$-ionone (Oliveira C. et al., 2006; Bindon et al., 2007; Yuan and Qian, 2016). What we did find in accordance with these researches are the low abundances of identified compounds. Since the majority of norisoprenoids, like monoterpenes, are found in non-volatile bound form (Lin et al., 2019; Mele et al., 2021), this could be potential explanation for detecting only two compounds in relatively small abundance.

Research on germplasm level that explore volatiles in grapevine varieties are scarce and do not include genetic and geographic origin. In our work when talking about geographic origin, we are referring to it in the sense of where these varieties are considered to be native, not in the sense of terroir or winegrowing regions. Furthermore, the samples were collected from a single location with the aim of minimalizing the effect of different environmental factors and rootstocks (Olarte Mantilla et al., 2018; Carrasco-Quiroz et al., 2020).

The discriminant analysis clearly separated the groups based on the country of origin and GEN-GEO groups. The mean values of compounds, above-mentioned to discriminate the groups, are the highest for the country of origin that they discriminate (Supplementary Tables 3, 4). However, when looking at the whole volatile profile, these compounds are found in relatively small abundance or in just few varieties. For example, $\gamma$-undelactone was identified in only four varieties, Rudezusa (Croatia), Uva rara (Italy), Graciano and Trepat (Spain). Nevertheless, they contribute to the discrimination and show that these small quantities identified are important varietal characteristics, that is characteristics defined by geographical origin. Alongside these compounds, the overall volatile profile also contributed to discrimination, especially to the French, Italian and Portuguese varieties, which are closely located on the scatter plot. These groups of varieties, as can be seen from correlations and mean values, are characterized by high content of alcohols, carbonyls, and sesquiterpenes. Similar results were obtained for GEN-GEO groups, and figures show similar position on scatter plots. GEO groups, like countries of origin, were discriminated by the compounds found in high quantities for the group that they discriminate. However, in the overall volatile profile, these compounds represent a small proportion. Furthermore, GEO groups are also discriminated by their overall volatile profile. Again, based on the correlations and mean values, it can be seen that C2 group contains higher abundance of carbonyl compounds and alcohols, while C8 group contains higher abundance of sesquiterpenes and acids. Group C7 is not characterized by high quantities of VOCs, except the compounds that discriminate it. Regarding the C8 group and its position on the scatter plot, it is interesting that the varieties separated within group, with Spanish varieties near y axis in the third quadrant, and Portuguese varieties near y axis in the fourth quadrant. That clear separation within group was not visible in group C2, containing French and Italian varieties. In the context of winemaking, even the small changes or differences in volatile profiles can have an impact on the sensory properties of wine (Ilc et al., 2016; Ferreira and Lopez, 2019).

On both figures representing DA results, varieties Vranac and Sangiovese singled out from their groups. In Figure 2 Sangiovese was located in the group containing Portuguese varieties and within the group near the varieties Carcajolo and Touriga nacional. Comparing the mean values of these three varieties, as well as the mean values of all Portuguese varieties, Sangiovese indeed has similar profile to Portuguese varieties. Although Sangiovese was not included by DA in group C8, on scatter plot is located more closely to group C8 and Portuguese varieties, than to group C2 containing all other Italian varieties. Vranac on the other hand was not placed near any group. The reason is probably its poor volatile profile, compared even to the Croatian varieties, which in general had low quantities of VOCs.

The volatile profiles of grape varieties are complex and include a large number of compounds. This research gives insight into the volatile profiles of red grape varieties with different genetic and geographic background. The most abundant compounds were carbonyls, while sesquiterpenes were the most numerous. Discriminant analysis clearly separated both countries of origin and GEN-GEO groups based on their volatile profile, with all classes of compounds contributing to the discrimination.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

IŠ: conceptualization, investigation, and writing–original draft. PŠ: formal analysis and investigation. IT: methodology. CM: resources. JK: writing–review and editing. TL: resources, writing–review and editing. EM: funding acquisition. DP: conceptualization, formal analysis, and writing–review and

editing, supervision. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by Centre of Excellence for Biodiversity and Molecular Plant Breeding (CoE Crop-BioDiv KK.01.1.1.01.005).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.942148/full#supplementary-material

## References

Aradhya, M. K., Dangl, G. S., Prins, B. H., Boursiquot, J. M., Walker, M. A., Meredith, C. P., et al. (2003). Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genet. Res.* 81, 179–192. doi: 10.1017/S0016672303006177

Arroyo-Garcia, R., Ruiz-Garcia, L., Bolling, L., Ocete, R., Lopez, M. A., Arnold, C., et al. (2006). Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp sativa) based on chloroplast DNA polymorphisms. *Mol. Ecol.* 15, 3707–3714. doi: 10.1111/j.1365-294X.2006.03049.x

Babushok, V. I., Linstrom, P. J., and Zenkevich, I. G. (2011). Retention indices for frequently reported compounds of plant essential oils. *J. Phys. Chem. Ref. Data* 40:043101. doi: 10.1063/1.3653552

Babushok, V. I., and Zenkevich, I. G. (2009). Retention indices for Most frequently reported essential oil compounds in GC. *Chromatographia* 69, 257–269. doi: 10.1365/s10337-008-0872-3

Bacilieri, R., Lacombe, T., Le Cunff, L., Di Vecchi-Staraz, M., Laucou, V., Genna, B., et al. (2013). Genetic structure in cultivated grapevines is linked to geography and human selection. *BMC Plant Biol.* 13:14. doi: 10.1186/1471-2229-13-25

Bindon, K. A., Dry, P. R., and Loveys, B. R. (2007). Influence of plant water status on the production of $C^{-13}$-norisoprenoid precursors in *Vitis vinifera* L. cv. Cabernet sauvignon grape berries. *J. Agric. Food Chem.* 55, 4493–4500. doi: 10.1021/jf063331p

Boss, P. K., Pearce, A. D., Zhao, Y., Nicholson, E. L., Dennis, E. G., and Jeffery, D. W. (2015). Potential grape-derived contributions to volatile Ester concentrations in wine. *Molecules* 20, 7845–7873. doi: 10.3390/molecules20057845

Bretón, M. P. R., Salinas, M. R., Nevares, I., Pérez-Álvarez, E. P., Álamo-Sanza, M. D., Román, S. M. S., et al. (2020). *Recent Advances in the Study of Grape and Wine Volatile Composition: Varietal, Fermentative and Ageing Aroma Compounds*. London: Taylor & Francis.

Carrasco-Quiroz, M., Martínez-Gil, A. M., Gutiérrez-Gamboa, G., and Moreno-Simunovic, Y. (2020). Effect of rootstocks on volatile composition of merlot wines. *J. Sci. Food Agric.* 100, 3517–3524. doi: 10.1002/jsfa.10395

Dennis, E. G., Keyzers, R. A., Kalua, C. M., Maffei, S. M., Nicholson, E. L., and Boss, P. K. (2012). Grape contribution to wine aroma: production of hexyl acetate, Octyl acetate, and benzyl acetate during yeast fermentation is dependent upon precursors in the must. *J. Agric. Food Chem.* 60, 2638–2646. doi: 10.1021/jf2042517

Dunlevy, J. D., Kalua, C. M., Keyzers, R. A., and Boss, P. K. (2009). "The production of flavour & aroma compounds in grape berries," in *Grapevine Molecular Physiology & Biotechnology*. ed. K. A. Roubelakis-Angelakis (Dordrecht: Springer Netherlands)

Ferreira, V., and Lopez, R. (2019). The actual and potential aroma of winemaking grapes. *Biomol. Ther.* 9:818. doi: 10.3390/biom9120818

Francis, I. L., and Newton, J. L. (2005). Determining wine aroma from compositional data. *Aust. J. Grape Wine Res.* 11, 114–126. doi: 10.1111/j.1755-0238.2005.tb00283.x

Gonzalez-Barreiro, C., Rial-Otero, R., Cancho-Grande, B., and Simal-Gandara, J. (2015). Wine aroma compounds in grapes: a critical review. *Crit. Rev. Food Sci. Nutr.* 55, 202–218. doi: 10.1080/10408398.2011.650336

Hernandez-Orte, P., Concejero, B., Astrain, J., Lacau, B., Cacho, J., and Ferreira, V. (2015). Influence of viticulture practices on grape aroma precursors and their relation with wine aroma. *J. Sci. Food Agric.* 95, 688–701. doi: 10.1002/jsfa.6748

Ilc, T., Werck-Reichhart, D., and Navrot, N. (2016). Meta-analysis of the Core aroma components of grape and wine aroma. *Front. Plant Sci.* 7:15. doi: 10.3389/fpls.2016.01472

Kalua, C. M., and Boss, P. K. (2009). Evolution of volatile compounds during the development of cabernet sauvignon grapes (*Vitis vinifera* L.). *J. Agric. Food Chem.* 57, 3818–3830. doi: 10.1021/jf803471n

Laucou, V., Launay, A., Bacilieri, R., Lacombe, T., Adam-Blondon, A. F., Berard, A., et al. (2018). Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS One* 13:27. doi: 10.1371/journal.pone.0192540

Lei, Y. J., Xie, S., Guan, X. Q., Song, C. Z., Zhang, Z. W., and Meng, J. F. (2018). Methoxypyrazines biosynthesis and metabolism in grape: a review. *Food Chem.* 245, 1141–1147. doi: 10.1016/j.foodchem.2017.11.056

Li, Z. Z., Howell, K., Fang, Z. X., and Zhang, P. Z. (2020). Sesquiterpenes in grapes and wines: occurrence, biosynthesis, functionality, and influence of winemaking processes. *Compr. Rev. Food Sci. Food Saf.* 19, 247–281. doi: 10.1111/1541-4337.12516

Li, X. Y., Wen, Y. Q., Meng, N., Qian, X., and Pan, Q. H. (2017). Monoterpenyl glycosyltransferases differentially contribute to production of Monoterpenyl glycosides in two aromatic *Vitis vinifera* varieties. *Front. Plant Sci.* 8:13. doi: 10.3389/fpls.2017.01226

Lin, J., Massonnet, M., and Cantu, D. (2019). The genetic basis of grape and wine aroma. *Horticulture Res.* 6:24. doi: 10.1038/s41438-019-0163-1

Luan, F., and Wust, M. (2002). Differential incorporation of 1-deoxy-D-xylulose into (3S)-linalool and geraniol in grape berry exocarp and mesocarp. *Phytochemistry* 60, 451–459. doi: 10.1016/S0031-9422(02)00147-4

Luo, J. Q., Brotchie, J., Pang, M., Marriott, P. J., Howell, K., and Zhang, P. Z. (2019). Free terpene evolution during the berry maturation of five *Vitis vinifera* L. cultivars. *Food Chem.* 299:13. doi: 10.1016/j.foodchem.2019.125101

May, B., and Wust, M. (2012). Temporal development of sesquiterpene hydrocarbon profiles of different grape varieties during ripening. *Flavour Fragr. J.* 27, 280–285. doi: 10.1002/ffj.3096

Mele, M. A., Kang, H.-M., Lee, Y.-T., and Islam, M. Z. (2021). Grape terpenoids: flavor importance, genetic regulation, and future potential. *Crit. Rev. Food Sci. Nutr.* 61, 1429–1447. doi: 10.1080/10408398.2020.1760203

Negrul, A., Baranov, A., Kai, Y., Lazarevski, M., Palibin, T., and Prosmoserdov, N. (1946). "Origin and classification of cultured grape," in *The Ampelography of the USSR. Vol. 1*. 159–216. Moscow: Pischepromizdat.

OIV (2017). *Distribution of the World;s Grapevine Varieties*. Paris: International Organization of Vine and Wine.

OIV (2019). *Compendium of International Methods of Wine and must Analysis France*. Paris: International Organization of Vine and Wine.

Olarte Mantilla, S. M., Collins, C., Iland, P. G., Kidman, C. M., Ristic, R., Boss, P. K., et al. (2018). Shiraz (*Vitis vinifera* L.) berry and wine sensory profiles and composition are modulated by rootstocks. *Am. J. Enol. Vitic.* 69, 32–44. doi: 10.5344/ajev.2017.17017

Oliveira, C., Barbosa, A., Ferreira, A. C. S., Guerra, J., and De Pinho, P. G. (2006). Carotenoid profile in grapes related to aromatic compounds in wines from Douro region. *J. Food Sci.* 71, S1–S7. doi: 10.1111/j.1365-2621.2006.tb12398.x

Oliveira, J. M., Faria, M., Sá, F., Barros, F., and Araújo, I. M. (2006). C6-alcohols as varietal markers for assessment of wine origin. *Anal. Chim. Acta* 563, 300–309. doi: 10.1016/j.aca.2005.12.029

Pineau, B., Barbe, J. C., Van Leeuwen, C., and Dubourdieu, D. (2009). Examples of perceptive interactions involved in specific "red-" and "black-berry" aromas in red wines. *J. Agric. Food Chem.* 57, 3702–3708. doi: 10.1021/jf803325v

Pons, A., Allamy, L., Schuttler, A., Rauhut, D., Thibon, C., and Darriet, P. (2017). What is the expected impact of climate change on wine aroma compounds and their precursors in grape? *Oeno One* 51, 141–146. doi: 10.20870/oeno-one.2017.51.2.1868

Rienth, M., Vigneron, N., Darriet, P., Sweetman, C., Burbidge, C., Bonghi, C., et al. (2021). Grape berry secondary metabolites and their modulation by abiotic factors in a climate change scenario–a review. *Front. Plant Sci.* 12:643258. doi: 10.3389/fpls.2021.643258

Saerens, S. M. G., Delvaux, F. R., Verstrepen, K. J., and Thevelein, J. M. (2010). Production and biological function of volatile esters in Saccharomyces cerevisiae. *Microb. Biotechnol.* 3, 165–177. doi: 10.1111/j.1751-7915.2009.00106.x

Schwab, W., Davidovich-Rikanati, R., and Lewinsohn, E. (2008). Biosynthesis of plant-derived flavor compounds. *Plant J.* 54, 712–732. doi: 10.1111/j.1365-313X.2008.03446.x

Šikuten, I., Štambuk, P., Karoglan Kontić, J., Maletić, E., Tomaz, I., and Preiner, D. (2021a). Optimization of SPME-arrow-GC/MS method for determination of free and bound volatile organic compounds from grape skins. *Molecules* 26:7409. doi: 10.3390/molecules26237409

Šikuten, I., Štambuk, P., Tomaz, I., Marchal, C., Kontić, J. K., Lacombe, T., et al. (2021b). Discrimination of genetic and geographical groups of grape varieties (*Vitis vinifera* L.) based on their polyphenolic profiles. *J. Food Compos. Anal.* 102:104062. doi: 10.1016/j.jfca.2021.104062

Song, N. E., Lee, J. Y., Lee, Y. Y., Park, J. D., and Jang, H. W. (2019). Comparison of headspace-SPME and SPME-arrow-GC-MS methods for the determination of volatile compounds in Korean salt-fermented fish sauce. *Appl. Biol. Chem.* 62:8. doi: 10.1186/s13765-019-0424-6

Winterhalter, P., and Rouseff, R. (2002). "Carotenoid-derived aroma compounds: an introduction," in *Carotenoid-Derived Aroma Compounds*. eds. P. Winterhalter and R. Rouseff (Washington: American Chemical Society).

Wood, C., Siebert, T. E., Parker, M., Capone, D. L., Elsey, G. M., Pollnitz, A. P., et al. (2008). From wine to pepper: Rotundone, an obscure sesquiterpene, is a potent spicy aroma compound. *J. Agric. Food Chem.* 56, 3738–3744. doi: 10.1021/jf800183k

Yang, C. X., Wang, Y. J., Liang, Z. C., Fan, P. G., Wu, B. H., Yang, L., et al. (2009). Volatiles of grape berries evaluated at the germplasm level by headspace-SPME with GC-MS. *Food Chem.* 114, 1106–1114. doi: 10.1016/j.foodchem.2008.10.061

Yu, F. N. A., and Utsumi, R. (2009). Diversity, regulation, and genetic manipulation of plant mono-and sesquiterpenoid biosynthesis. *Cell. Mol. Life Sci.* 66, 3043–3052. doi: 10.1007/s00018-009-0066-7

Yuan, F., and Qian, M. C. (2016). Development of $C^{-13}$-norisoprenoids, carotenoids and other volatile compounds in *Vitis vinifera* L. Cv. Pinot noir grapes. *Food Chem.* 192, 633–641. doi: 10.1016/j.foodchem.2015.07.050

Yue, X. F., Ren, R. H., Ma, X., Fang, Y. L., Zhang, Z. W., and Ju, Y. L. (2020). Dynamic changes in monoterpene accumulation and biosynthesis during grape ripening in three *Vitis vinifera* L. cultivars. *Food Res. Int.* 137:109736. doi: 10.1016/j.foodres.2020.109736

Zhu, B. Q., Xu, X. Q., Wu, Y. W., Duan, C. Q., and Pan, Q. H. (2012). Isolation and characterization of two hydroperoxide lyase genes from grape berries. *Mol. Biol. Rep.* 39, 7443–7455. doi: 10.1007/s11033-012-1577-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 October 2022
DOI 10.3389/fpls.2022.988861



OPEN ACCESS

EDITED BY
Ravi Gupta,
Kookmin University, South Korea

REVIEWED BY
Wang Pu-chang,
Guizhou Normal University, China
Dinesh Kumar Saini,
South Dakota State University,
United States

*CORRESPONDENCE
Peng Qin
qinpeng77@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Crop and Product Physiology, a section of the journal Frontiers in Plant Science

RECEIVED 07 July 2022
ACCEPTED 10 October 2022
PUBLISHED 26 October 2022

CITATION
Huan X, Li L, Liu Y, Kong Z, Liu Y, Wang Q, Liu J, Zhang P, Guo Y and Qin P (2022) Integrating transcriptomics and metabolomics to analyze quinoa (*Chenopodium quinoa Willd.*) responses to drought stress and rewatering.
*Front. Plant Sci.* 13:988861.
doi: 10.3389/fpls.2022.988861



# Integrating transcriptomics and metabolomics to analyze quinoa (*Chenopodium quinoa Willd.*) responses to drought stress and rewatering

Xiuju Huan[1†], Li Li[1†], Yongjiang Liu[1†], Zhiyou Kong[2], Yeju Liu[3], Qianchao Wang[1], Junna Liu[1], Ping Zhang[1], Yirui Guo[1] and Peng Qin[1*]

[1]College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China, [2]College of Resources and Environment, Baoshan College, Baoshan, China, [3]Graduate Office, Yunnan Agricultural University, Kunming, China

The crop production of quinoa (*Chenopodium quinoa Willd.*), the only plant meeting basic human nutritional requirements, is affected by drought stress. To better understand the drought tolerance mechanism of quinoa, we screened the drought-tolerant quinoa genotype "Dianli 129" and studied the seedling leaves of the drought-tolerant quinoa genotype after drought and rewatering treatments using transcriptomics and targeted metabolomics. Drought-treatment, drought control, rewatering-treated, and rewatered control were named as DR, DC, RW, and RC, respectively. Among four comparison groups, DC vs. DR, RC vs. RW, RW vs. DR, and RC vs. DC, we identified 10,292, 2,307, 12,368, and 3 differentially expressed genes (DEGs), and 215, 192, 132, and 19 differentially expressed metabolites (DEMs), respectively. A total of 38,670 genes and 142 pathways were annotated. The results of transcriptome and metabolome association analysis showed that gene-*LOC110713661* and gene-*LOC110738152* may be the key genes for drought tolerance in quinoa. Some metabolites accumulated in quinoa leaves in response to drought stress, and the plants recovered after rewatering. DEGs and DEMs participate in starch and sucrose metabolism and flavonoid biosynthesis, which are vital for improving drought tolerance in quinoa. Drought tolerance of quinoa was correlated with gene expression differences, metabolite accumulation and good recovery after rewatering. These findings improve our understanding of drought and rewatering responses in quinoa and have implications for the breeding of new drought-tolerance varieties while providing a theoretical basis for drought-tolerance varieties identification.

KEYWORDS

**quinoa, drought stress, transcriptomics, metabolomics, flavonoid biosynthesis**

## Introduction

Quinoa (*Chenopodium quinoa* Willd.) is native to the Andean region of South America, with a long history of cultivation and has received considerable attention as a functional food in recent years. Quinoa grains are small and round with white, black, and red colors (Repo-Carrasco et al., 2003). Their outstanding physicochemical, nutritional, and functional properties result in their use as a staple food for humans (Melini and Melini, 2021). With its high nutritional potential and genetic diversity, the Food and Agriculture Organization (FAO) classified quinoa as a promising crop for humans that can contribute to food security in the 21st century (FAO Regional Office for Latin America and the Caribbean & Proinpa, 2011). Therefore, the selection and exploitation of quinoa varieties is crucial.

Quinoa contains high quality protein with a perfect balance of essential amino acids and a suitable fatty acid composition. In addition, it is rich in bioactive compounds such as polyphenols, flavonoids, and minerals (Nowak et al., 2016); the high content of phenolics makes it an important antioxidant active. The polysaccharide fraction of quinoa also has potential as a natural antioxidant, antidiabetic, and immunomodulatory food (Tan et al., 2021). Both the seeds and leaves of quinoa have some nutritional value, and current research on quinoa has focused on the nutritional composition of the seeds as functional foods. However, quinoa leaves also have some nutritional potential and may prevent cancer and other diseases related to oxidative stress (Gawlik-Dziki et al., 2013). Quinoa grains and leaves both contain a high content of phenolics, although the leaves have a higher protein content than the grains. The consumption of nutrient-rich green leaf quinoa can prevent nutritional deficiencies caused by iron and zinc, and there is higher saponin content in the grains compared with the leaves (Pathan et al., 2019; Villacrés et al., 2022). Quinoa grains also contain sufficient micronutrients such as calcium, phosphorus, potassium, copper, iron, and zinc (Ayasan, 2020). In addition, quinoa has some anti-nutritional factors such as saponins, phytic acid, tannins, and protease inhibitors, among which the saponins can resist the adverse conditions of the external environment (Filho et al., 2017). Quinoa has a wide genetic diversity and can adapt to various harsh environments including biotic and abiotic stresses, and can be grown on plateaus from sea level to 4,500 m (Zhang et al., 2021). However, the nutritional value of quinoa differs by variety and environment (Nowak et al., 2016), making it particularly important to study quinoa breeding in adverse conditions.

Drought conditions threaten crop production and food security (Yang et al., 2020). Drought stress affects morphological and physiological changes in plants, leading to severe crop yield deficits; agricultural drought affects global food production and is among the most serious challenges facing sustainable agriculture (Fadiji et al., 2022). Water stress not only affects metabolic activities such as plant respiration, sugar metabolism, and photosynthesis (Dos Reis et al., 2012), but also reduces the cellular water potential, affecting growth and cell elongation. The occurrence of drought stress during the reproductive period may also lead to interruption of flowering and yield loss (Kaur and Asthir, 2017). Plants under drought stress release reactive oxygen species (ROS) and free radicals, triggering an increase in ethylene content (Nair et al., 2008; Narayanasamy et al., 2020). Drought stress not only causes physiological responses in plants, but also impacts mineral nutrition, where a reduction in iron uptake occurs through a specific response to drought, leading to a reduction in zinc and manganese uptake, which is associated with differential expression of transport-related genes (D'Oria et al., 2022). Drought tolerance in plants refers to the ability of plants to tolerate drought and rapidly resume growth after rehydration. Drought severely affects plant growth and development; however, there is some compensation through rehydration (e.g., photosynthesis) (Chaves and Oliveira, 2004). In conclusion, increasing future food production under drought stress will be challenging. While many crops have been extensively studied under drought stress, research on the unique mechanisms of quinoa, an important gluten-free crop, to cope with different degrees of drought have been limited. Among the mechanisms of action for drought adaptation in quinoa, bio-promoters can lead to an increase in total soluble sugars (SS), proteins, and antioxidant enzyme activities in quinoa leaves and roots; however, drought decreases biomass, leaf water potential, and stomatal conductance, and increases malondialdehyde and hydrogen peroxide content (Benaffari et al., 2022). Physiological analysis of drought tolerance mechanisms in quinoa demonstrated an increase in $H_2O_2$ and malondialdehyde (MDA) content in drought-treated quinoa, with differences in the physiological response to different varieties of quinoa. This suggests that different varieties of quinoa have different drought tolerance mechanisms (Lin and Chao, 2021). Further, physiological characteristics of quinoa under rehydrated conditions after drought stress have not been reported. Quinoa genotypes grown in coastal lowlands have always exhibited better yields and larger seeds with reduced

**Abbreviations:** ABA, abscisic acid; CHS, chalcone synthase; DC, drought control; DEG, differentially expressed gene; DEM, differentially expressed metabolite; DFR, dihydroflavonol 4-reductase; DR, drought treatment; EC, Enzyme Commission; F3H, flavanone 3-hydrogenase; FC, fold-change; FPKM, fragments per kilobase of transcript, per million mapped reads; GO, Gene Ontology; JA, jasmonates; KEGG, Kyoto Encyclopedia of Genes and Genomes; KOG, Eukaryotic Orthologous Group; MDA, malondialdehyde; PCA, principal component analysis; PCC, Pearson's correlation coefficient; POD, peroxidase; Pro, proline; RC, rewatered control; RW, rewatering treatment; SA, salicylic acid; SOD, superoxide dismutase; SS, soluble sugar; TF, transcription factor TIC, total ion current.

irrigation relative to commercial varieties (Dumschott et al., 2022). Among them, *CqZF-HD14* further enhanced the drought tolerance of quinoa seedlings in synergy with *CqNAC79* or *CqHIPP34* and may be a key gene in the drought tolerance regulatory network of quinoa (Sun et al., 2022). Varieties and methods that can withstand drought stress, among others, are being extensively researched worldwide to alleviate water stress (Philippot et al., 2013). Thus, it is imperative to investigate the mechanisms of drought tolerance in quinoa in order to successfully breed drought-tolerant varieties.

Transcriptome sequencing technologies and metabolome assays have been used to analyze plant tolerance mechanisms (Lenka et al., 2011; Nakabayashi et al., 2014; Yuan et al., 2018; Mu et al., 2021). Comparative transcriptome analysis of drought tolerance in two rice varieties showed that it was attributable to enhanced expression of several enzyme-coding genes, and drought sensitivity was attributed to significant down-regulation of regulatory components that confer drought tolerance (Lenka et al., 2011). Ramie plants exhibited differential expression of *AP2, MYB, NAC*, zinc finger proteins, and the *bZIP* transcription factor (TF) (An et al., 2015), suggesting an association with osmotic treatment. The maximum activity of superoxide dismutase (SOD) and peroxidase (POD), as well as the contents of MDA and proline (Pro), increased in wheat plants to differing degrees during a winter drought (Mu et al., 2021). Flavonoids can alleviate oxidation and drought stress in *Arabidopsis thaliana* (Nakabayashi et al., 2014). In the study of Tibetan hullless barley under salt stress the main compounds included amino acids and their derivatives, organic acids, nucleotides and their derivatives, and flavonoids (Wang et al., 2019). Transcriptomic and metabolomic studies on powdery mildew tolerance in Tibetan hullless barley showed a significant enrichment of genes related to pathways such as phenylalanine metabolism, terpene biosynthesis, zeatin biosynthesis, and isoflavonoid biosynthesis, which may be associated with downy mildew tolerance in fully tolerant varieties (Yuan et al., 2018).However, a multi-omics-based study of drought tolerance mechanisms in quinoa seedlings has not yet been reported.

Therefore, it is imperative to rapidly screen genes associated with drought stress and rehydration to provide excellent genetic resources for creating drought-tolerant quinoa germplasm. However, studies on the molecular regulation of quinoa adaptation to drought stress have been limited, and multi-omics-based studies on drought tolerance mechanisms in quinoa seedlings have not yet been reported. In this study, we analyzed drought-tolerant quinoa plants at the seedling stage (six-leaf stage) after drought and rewatering by transcriptome sequencing and metabolomics. Our results shed light on the mechanisms of drought tolerance in quinoa and direct research towards a comprehensive exploration of drought tolerance genes in quinoa. In addition, this topic is important to understand the drought tolerance mechanism of quinoa in order to breed optimal varieties.

# Materials and methods

## Material planting

Material was sourced from China and the United States, and quinoa genotypes were introduced through selective breeding (The natural variation of existing varieties during the breeding process is used as the original material for the selection work, which is then carried out continuously according to the requirements of high and stable yield and disease resistance, and the selected lines are planted and selected again until the traits are relatively stable). While quinoa grains are available in a variety of colors, four colors were initially selected for this study: red, yellow, white, and black, with five genotypes of each color for a total of 20 an advance generation genotypes. The 20 quinoa genotypes were planted in a greenhouse seedling tray. Fifty-cell seedling growth trays were used (50 mm × 50 mm × 90 mm for each point). Three seeds were sown at each point, the seedlings were thinned when the plants had grown to the two-leaf stage, and finally, one plant was left at each point. All genotypes were managed in the same way with water and fertilizer before the drought treatment. Humus soil was mixed with perlite at a 4:1 ratio as the cultivation substrates for the drought stress-treated and control groups.

## Drought treatment and drought-tolerance material screening

Quinoa seedlings were treated with natural drought stress (i.e., no watering during this period) for 5 ds at the 6-leaf-one stage, followed by 1 day of rewatering. The drought-treated control was watered normally during the drought treatment and the rewatered control was watered normally on both the 5 days of the drought treatment and the 1 day of rewatering. The average temperature of the greenhouse was 27.1°C and the average humidity was 59.6%. Based on the degree of wilting and survival rate of plants on the fifth day of drought treatment, two drought tolerant genotypes (Dianli 66 and Dianli 129) and two drought sensitive genotypes (Dianli 58 and Dianli 114) were initially screened out of 20 genotypes. Only a few plants in each genotype with slightly curled leaves were considered drought tolerant and most plants in each genotype with all curled leaves was considered drought sensitive, according to degree of wilting criteria used. For the survival rate criteria, seedlings of quinoa genotypes with more than 85% survival are considered drought tolerant genotypes and those with less than 60% survival are considered drought sensitive genotypes. The most drought tolerant genotype, Dianli 129, and the most drought sensitive genotype, Dianli 114, were further screened among the four genotypes according to the method above. The leaves of these two genotypes were sampled on day 5 of the drought treatment and on day 1 of the rewatering treatment, with DR on day 5 of

the drought treatment, drought treatment was rewatered for 1 day after 5 days as RW and normal watering during both the drought and rewatering treatments was the rewatering control RC. The physiological parameters related to drought tolerance were measured, and the drought tolerant genotypes and drought sensitive genotypes were analyzed for comparison (Lin and Chao, 2021). Finally, metabolome determination and transcriptome sequencing were performed on the most drought tolerant genotype, Dianli 129, to further investigate the drought tolerance mechanism of quinoa. Dianli 129 had four treatments, drought treatment (drought treatment for 5 days), drought control (normal watering), rewatering treatment (drought treatment for 5 days followed by rewatering for 1 day) and rewatering control (normal watering), each with three replicates for a total of 12 samples, one sample from each treatment was mixed with five biological replicates.

## Morphological parameters, physiological parameters measurement and statistical analysis

The most drought tolerant genotype, Dianli 129, and the drought sensitive genotype, Dianli 114, were screened for morphological and physiological parameters measurement. Single quinoa seedlings of uniform growth were selected to determine plant height, above- and below-ground part biomass, and leaf morphology and each parameter were replicated three times. The root length, average root diameter, volume and root surface area were determined by scanning with a Topper root scanner (MRS-9600TFU2L). The plants were placed in an oven, killed at 110°C for 15 min, dried at 80°C until a constant weight, and the dry weight of above- and below-ground parts was measured, and the root-to-crown ratio = root dry weight/above-ground part dry weight was calculated. Leaf color was measured by a Minolta colorimeter (CR-20). Among the physiological parameters determined, Chlorophyll content was determined using the ethanol acetone method. Soluble protein content was determined by the Komas blue colorimetric method. Pro content was determined by the acidic ninhydrin method. MDA content was determined by the thiobarbituric acid (TBA) method. Catalase (CAT) activity was determined by the hydrogen peroxide reduction method. POD activity was determined by the guaiacol method. SOD was determined by the nitrogen blue tetrazolium photochemical reduction method. Total antioxidant capacity (T-AOC) was determined by the iron ion reduction method. SS were determined by the anthrone colorimetric method. Relative conductivity was also measured. Each parameter was repeated three times (Benaffari et al., 2022). Microsoft Excel 2010 was employed for graphical analysis and statistical analysis was performed using SPSS25, DPS version 7.05 software.

## Metabolite extraction and detection

The samples were placed in a SCIENTZ-100F Lyophilization Dryer Laboratory LCD Display Freeze Dryer (SCIENTZ) for vacuum freeze-drying. The samples were ground to powder form using a MM 400 grinding machine (30 Hz, 1.5 min; Retsch). Subsequently, 100 mg of powder was dissolved in a 1.2 mL 70% methanol extract. The dissolved samples were refrigerated overnight at 4°C, while swirling six times during the period to improve extraction rate. Each sample was centrifuged at 12,000 rotations per min for 10 min, after which the upper liquid fraction was filtered through a 0.22-μm membrane filter (0.22-μm pore size). Each sample was saved in a bottle for analysis using ultra-performance liquid chromatography and tandem mass spectrometry.

Metabolome profiling was performed using a widely targeting metabolomics method, based on a database (MWDB) built by Wuhan Mettware Biotechnology (http://www.metware.cn/), and qualitative analysis was based on secondary spectrum information. Metabolites were quantitated using triple-level quadrupole mass spectra obtained in multiple-reaction monitoring mode. Before data analysis, quality-control analysis was performed to confirm the reliability of the data. Principal component analysis (PCA) was conducted to analyze variabilities between and within groups. DEMs were subjected to orthogonal partial least-squares discriminant analysis (OPLS-DA). Metabolites with a variable importance in projection (VIP) ≥ 1 and a fold-change (FC) of ≥ 2 (or ≤ 0.5) were defined as DEMs. See Supplementary Material 1 for the collection conditions and experimental methods of chromatography-mass spectrometry.

## Transcriptome sequencing and data analysis

Total RNA was extracted from 12 quinoa leaf samples using the TRIzol reagent according to the manufacturer's instructions (Beijing TransGen Biotech). After RNA extraction, an RNA-sequencing library was constructed, and then the quality of the library was determined. The Illumina HiSeq platform was used for sequencing after determining that the sequencing library met the requirements. To ensure the accuracy of subsequent analysis, the original data were filtered and screened, and low-quality reads, adapter sequences, and when the N content (proportion of reads with N bases) of any sequenced read exceeds 10% of the number of bases in that read were removed. The high-quality clean reads obtained by screening were compared with the reference genome. The fragments per kilobase of transcript (FPKM) per million mapped reads was used as an parameters to measure gene expression levels. The screening criteria for identifying DEGs were a |log2 FC| value of ≥ 1 and a false-

discovery rate of < 0.05, a positive value of |log2 FC| is an up-regulated gene, while a negative value is a down-regulated gene. Functional annotations of DEGs were performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Eukaryotic Orthologous Group (KOG), PfAM, Swiss-Prot, TrEMBL, and NR databases. See Supplementary Material 2 for the experimental procedure of transcriptome sequencing.

## Quantitative real-time PCR validation (qRT-PCR)

qRT-PCR was conducted in Step One in addition to a real-time fluorescence quantitative PCR instrument (Thermo Fisher, USA). TUB1 was used as the internal reference gene. The reaction procedure was as follows: 95°C for 30 s, followed by 40 cycles, 95°C for 5 s, and 60°C for 30 s. According to the kit instructions (Beijing TransGen Biotech), a 20 μL system was used for each reaction: 3 μL cDNA, 0.4 μL forward primer (10 μM) and 0.4 μL reverse primer (10 μM), 10 μL green qPCR SuperMix, 0.4 μL passive reference dye I, and 5.8 μL nuclease-free water. The experiment was repeated with three biological replicates on 96-well plates.

# Results

## Effects of drought stress and rewatering on the morphology and physiology of quinoa seedlings

The morphological parameters (total root length, total root surface area, total root volume, etc.) and physiological parameters (CAT, SOD, MDA, T-AOC, etc.) of the drought tolerant genotype Dianli-129 and drought-sensitive genotype Dianli 129 were compared. The difference of morphological indexes between two strains "dianli-114" and "dianli-129" was compared. The root shoot ratio was P > 0.05, so the difference was not significant. The total root length and other six indexes were P<0.01, so the difference was extremely significant; The comparison among the four treatments (DR, DC, RW, RC) showed that the root shoot ratio was P < 0.05, so the difference was significant. The leaf area and other six parameters were P < 0.01, so the difference was extremely significant (Table 1). Compared between the two strains, leaf width P < 0.05, the difference is not significant, and leaf perimeter and other seven parameters P <0.01, the difference is extremely significant; Compared among the four treatments, all indexes were P < 0.01, the difference was extremely significant (Table 2). The ANOVA found no significant differences in CAT, SOD, MDA and T-AOC between the two genotypes, however, significant differences in PDO and all other parameters were identified. In each treatment, the difference of all parameters was highly significant; Under the cross action of the two factors, the chlorophyll difference was significant, and all other parameters reached a highly significant level (Table 3).

## Metabolomics of quinoa leaves under drought and rewatered conditions

Four groups of samples (DR, DC, RW, and RC) were analyzed using a widely targeted metabolomics approach that enabled the detection of 701 metabolites divided into 12 categories including amino acids and derivatives, flavonoids, and phenolic acids. A total of 99 flavonoids related metabolites were detected, namely 53 flavonols, 36 flavonoids, 4 dihydroflavonols, 2 dihydroflavone, 2 isoflavones, and 2 chalcones. The contents of 11 flavonoids such as 6-hydroxyluteolin 5-glucoside were lower than the control during drought but increased to the control level or significantly higher after rehydration. The content of 15 flavonoids such as kaempferol-4'-o-glucoside increased during drought, and the content after rehydration exhibited little difference compared with the control or continued to increase after rehydration. The contents of 58 flavonoids such as naringenin (5,7,4'-trihydroxyflavanone) were significantly lower than the control during drought. After rehydration, their contents did not adjust to the control level; however, they all exhibited varying upward trends (Supplementary Table 1). The total ion current (TIC) diagram of total ion flow shows detection and analysis of essential spectra for different quality-control samples, which overlapped. The total ion flow metabolite curves showed a high degree of overlap, that is, the retention times and peak strengths were consistent. This indicates that the signal stability of mass spectrometry for detecting the same sample at different time points was good, that is, the technology of metabolite extraction

TABLE 1 Analysis of variance for morphological parameter determination 1 (*F* value).

| Source of variation | DF | Total root length (cm) | Total root Surface Area ($cm^2$) | Total root volume ($cm^3$) | Average root Diameter (mm) | Plant height (cm) | Leaf area ($mm^2$) | Root-shoot ratio |
|---|---|---|---|---|---|---|---|---|
| Materials | 1 | 150.409** | 374.858** | 559.578** | 29.356** | 757.412** | 11.689** | 1.850 |
| Treatments | 3 | 112.004** | 269.226** | 274.101** | 13.582** | 249.828** | 203.869** | 6.319* |
| Materials×Treatments | 3 | 102.450** | 231.390** | 407.149** | 8.466** | 255.951** | 20.927** | 19.002** |

* and ** indicate P<0.05 and P<0.01. Table 2.

10.3389/fpls.2022.988861

TABLE 2 Analysis of variance for morphological parameter determination 2 (*F* value).

| Source of variation | DF | Root dry weight | Shoot dry weight | Leaf length | Leaf width | Leaf perime-ter | Lightness L | Red-Green a | Yellow-Blue b |
|---|---|---|---|---|---|---|---|---|---|
| Materials | 1 | 16.0** | 15.4** | 51.5** | 2.5 | 14.5** | 21.9** | 2040.0** | 19.4** |
| Treatments | 3 | 31.2** | 11.8** | 45.1** | 172.0** | 41.6** | 74.2** | 874.7** | 1155.0** |
| Materials×Treatments | 3 | 50.8** | 15.6** | 22.2** | 10.7** | 14.7** | 0.9 | 500.1** | 119.6** |

* and ** indicate P<0.05 and P<0.01. Table 3.

and detection had good repeatability and high reliability (Figures 1A, B). Cluster-heat map analysis was performed on all samples. All samples grouped together after performing three replicates, indicating that the metabolome data had high reliability. Significant differences were found between the control groups (RC and DC) and the treatment groups (DR and RW), in terms of the metabolite levels. After rewatering treatment, the metabolites slowly returned to normal levels After rewatering treatment, the metabolites slowly returned to normal levels (Figure 1C). PCA of the samples revealed that there was an evident trend of separation between groups on PC1 and PC2, differences among the groups and good repeatability (Figure 1D).

## Identification of different metabolites in quinoa leaves

DEMs were identified between samples according to the criteria of a VIP of ≥1 and an FC of ≥ 2 or ≤ 0.5. To study trends in different samples, the relative contents of different metabolites were standardized, centralized, and analyzed by K-means clustering (Supplementary Figure 1; Supplementary Table 2). The different metabolites were divided into nine groups, In the 6th cluster, the levels of amino acids and their derivatives with higher metabolite levels under drought conditions, which returned to normal after rewatering. In the seventh cluster, metabolites such as flavonoids were lower under drought conditions and returned to normal after rewatering. Four differential metabolite-comparison groups were obtained through pairwise comparisons. Specifically, in the DC vs DR comparison group, 84 metabolites were up-regulated and 131 were down-regulated. In the RC vs RW comparison group, 84 metabolites were up-regulated and 108 were down-regulated. In the RW vs DR comparison group, 67 metabolites were up-regulated and 65 were down-regulated. In the RC vs DC comparison group, 15 metabolites were up-regulated and four were down-regulated (Supplementary Table 3). The Venn diagram in Figure 2 shows that the different groups had 40, 23, 0, and 47 DEMs, respectively, four of which were common among all four groups. By comparing the metabolite FCs in each group, we determined that among the four groups, up-regulated DEMs with the largest |log2 FC| values included N-feruloyltyramine, quercetin-7-O-rutinoside-4-O-glucoside,4-O-(6-O-glucosylferuloyl)-3,4-dihydroxybenzyl alcohol, and quercetin-7-O-rutinoside-4-O-glucoside. Down-regulated DEMs with the highest |log2FC| values included quercetin, 4-O-(6-O-glucosylferuloyl)-3,4-dihydroxybenzyl alcohol, quercetin, and LysoPE 15:1 (Supplementary Figure 2). In the four differential metabolite-comparison groups, the DEM-associated metabolite pathways with significant enrichment included cyanoamino acid metabolism, flavonoid biosynthesis, starch and sucrose metabolism, penicillin and cephalosporin biosynthesis, indole alkaloid biosynthesis, sulfur metabolism, propanoate metabolism, glycerolipid metabolism, glucosinolate biosynthesis, aminoacyl-tRNA biosynthesis, synthesis and degradation of ketone bodies, fatty acid metabolism, and lysine biosynthesis (Figure 3; Supplementary Figure 3).

## Transcriptomics of quinoa leaves under drought and rewatered conditions

Transcriptome sequencing analysis of 12 samples yielded 77.06 GB clean data, Among the high-quality clean reads, the percentage of the Q20 base was >98%, the percentage of the Q30 base was >94%, and the GC contents were >43.0%. These reference data indicated that the sequencing results were reliable and could be used for further analysis (Supplementary

TABLE 3 Analysis of variance for physiological parameter determination(*F* value).

| Source of variation | DF | CAT | POD | SOD | MDA | Proline | T-AOC | Soluble protein | Soluble sugar | Relative conductivity | Chlorophyll |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Materials | 1 | 1.8 | 6.6* | 1.14 | 0.22 | 94.2** | 0.74 | 238** | 29.7** | 46.2** | 23.183** |
| Treatments | 3 | 140** | 71** | 79.0** | 65** | 1885** | 26.5** | 133** | 199.9** | 43.6** | 8.154** |
| Materials×Treatments | 3 | 272** | 218** | 55.1** | 21** | 278** | 9.73** | 132** | 28.8** | 25.3** | 3.747* |

* and ** indicate P < 0.05 and P < 0.01, respectively. DF, degree of freedom; CAT, catalase; POD, Peroxidase; MDA, Malondialdehyde; T-AOC, Total antioxidant capacity.

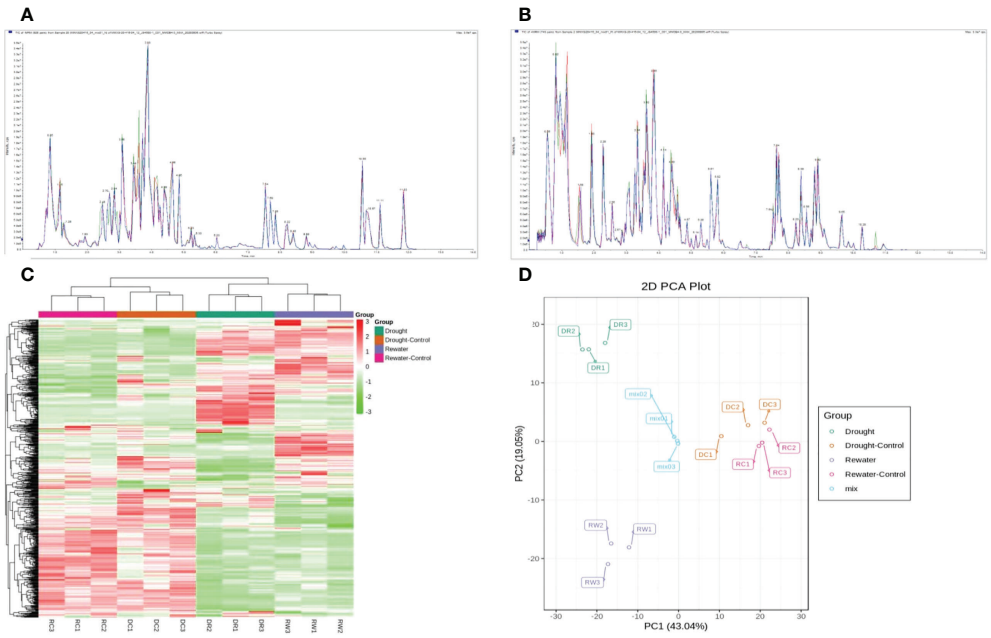



FIGURE 1
Analysis of metabolites in different comparison groups. **(A, B)** TIC overlap diagram of indicating the sample qualities, based on the observed spectra. **(C)** Overall clustering diagram for several samples. **(D)** PCA diagram. In **(A, B)**, **(A)** indicates the for negative-ion mode, while **(B)** indicates the positive-ion mode. In **(C)**, the sample names are shown horizontally, the metabolite information is shown vertically, and the values obtained after standardizing the relative contents are shown in different colors (red represents a high content, green represents a low content). In **(D)**, PC1 represents the first principal component, PC2 represents the second principal component, and the percentages represent the estimated contribution of the principal component to the data set. Each data point in the figure represents a sample. Samples in the same group are represented using the same color, and "MIX" is the quality-control sample containing a mixture of equal amounts of each sample.

Table 4). By adopting the PCA method of multivariate statistical analysis, the data for each group of triplicate samples showed that the method had good stability and quality. Significant separation was found between the treated and control samples, indicating that changes in metabolite accumulation were strictly controlled by differential gene expression (Figure 4A). Using FPKM as an parameter of gene expression levels, the density map showed that the gene-abundance trends in the samples changed with the expression levels, which clearly reflected the gene expression levels in the samples (Figure 4B), FPKM = $10^{-2}$ ~ $10^{4}$. Pearson's correlation coefficient (PCC, abbreviated as "r") was used as an parameter to evaluate correlations with biological replicates. The closer the R2 is to 1, the stronger the correlation between the two replicate samples. This study requires that the R2 between biological replicate samples be at least greater than 0.8 before further study of DEGs.

## DEGs in quinoa leaves

Using the following databases, we annotated of DEGs with KEGG (38,670; Supplementary Table 5), GO (38,191; Supplementary Table 6), KOG (45,935; Supplementary Table 7), PfAM (77,193; Supplementary Table 8), Swiss-Prot (31,837; Supplementary Table 9), TrEMBL (47,387; Supplementary Table 10), and NR (49,054; Supplementary Table 11). KEGG involves 142 pathways. By analyzing DEGs in quinoa leaves under drought stress and rewatering conditions, 14,883 differentially expressed genes were found. When comparing them in pairs, we generated four DEG comparison groups, among which 4,104 up-regulated genes and 6,188 down-regulated genes were found in the DC vs. DR comparison group. In the RC vs. RW comparison group, there were 1,439 up-regulated genes and 868 down-regulated genes. In the RW vs. DR comparison group, there were 5,374 up-regulated genes and 6,994 down-regulated genes. In the RC vs. DC comparison group, there were no up-regulated genes and three down-regulated genes (Supplementary Table 12). After drought treatment, the gene expression patterns changed significantly, and the genes tended to be stably expressed after rewatering (Figure 5A). The Venn diagram in Figure 5B shows 0 DEGs in common among all four groups, and 1,830, 3,343, 0, and 407 specific DEGs in the four comparison groups, respectively. KEGG analysis (Figure 6 and Supplementary Figure 4) showed that the significantly enriched pathways in the four comparison groups

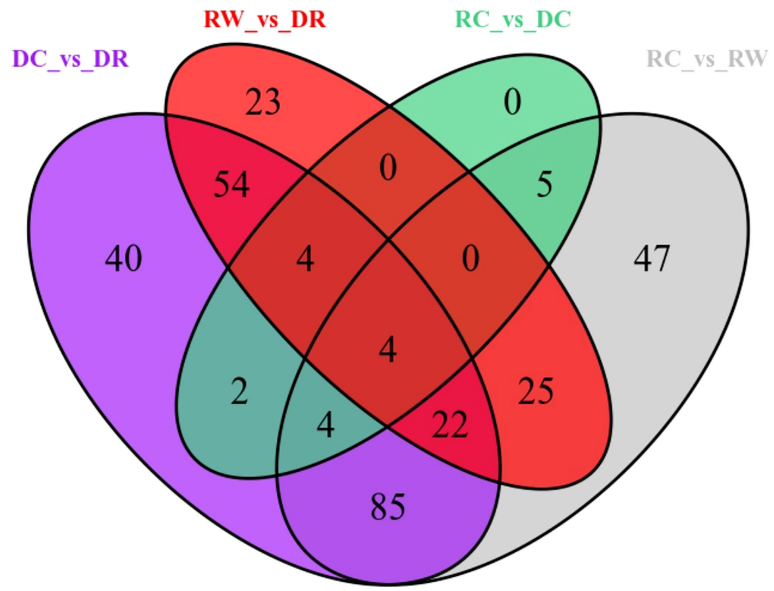



FIGURE 2
Venn diagram for differential metabolites. Each circle in the figure represents a comparison group. The numbers in the circles and overlapping regions represent the number of differential metabolites in common with the comparison group, whereas the numbers without overlaps represents the number of DEMs unique to the comparison group.

included zeatin biosynthesis, photosynthesis, photosynthesis-antenna proteins, ribosome, ribosome biogenesis in eukaryotes, biotin metabolism, alpha-linolenic acid metabolism, and linolenic acid metabolism. The DEGs in quinoa leaves were classified by GO enrichment (Supplementary Figure 5), to evaluate enrichment for DEGs in terms of molecular function, cellular component, and biological process. In terms of biological process, this included metabolic and cellular processes. Cellular components mainly included cells, cell parts, and organelles. Molecular functions mainly included binding and catalytic activities. The results indicate that cells and cell parts were most enriched for DEGs in quinoa leaves, indicating that cellular components played important roles in responding to drought stress. The 50 GO terms with the lowest q values in the enrichment analysis were selected, and the enrichment entries were plotted in a bar chart. The genes showing greater enrichment were related to several biological factors, such as fatty acid metabolic processes, ribosome biogenesis, apoplasts, ribosomal subunits, lyase activity, rRNA binding, and photosynthesis (Supplementary Figure 6).

## Analysis of TFs under drought and rewatering conditions

TFs play important roles in plant responses to drought stress and rewatering by regulating the expression levels of target genes. We analyzed TFs associated with the DEGs identified in quinoa leaves. In four comparison groups (Supplementary Table 13), 598, 116, 629, and 0 TFs were detected, indicating the key roles of TFs during drought and rewatering treatment. The 1,343 TFs were divided into 55 families. The main TFs in this study included the AP2 (130), MYB (126), bHLH (80), WRKY (79), NAC (64), and bZIP (41) families.

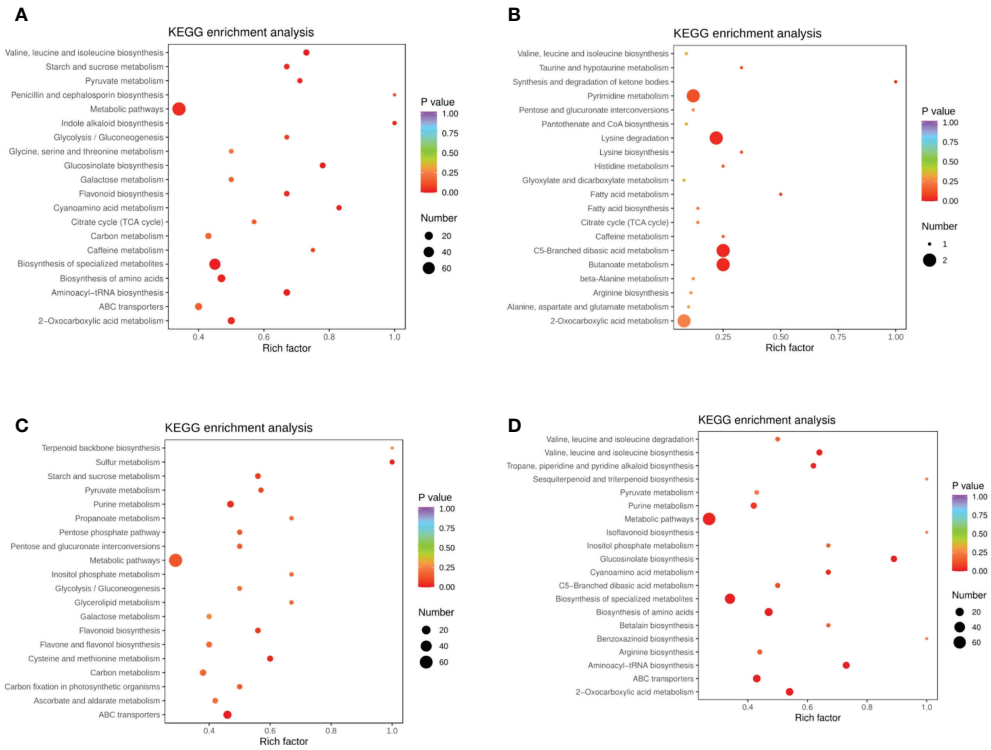



FIGURE 3
KEGG enrichment diagram of different metabolites in quinoa leaves. **(A)** DC vs DR. **(B)** RC vs DC. **(C)** RC vs RW. **(D)** RW vs DR. The horizontal coordinate represents the Rich factor corresponding to each pathway (i.e., the ratio between the number of metabolites in the corresponding pathway and the total number of metabolites detected and annotated in the pathway, where larger Rich factors correspond to greater enrichment). The vertical coordinate represents pathway name, the color of each data point represents the P value (where a deeper shade of red corresponds to more significant enrichment). The size of each data point represents the number of enriched DEMs.

## QRT-PCR validation results

QRT-PCR was used to validate the randomly selected genes. With primers designed by Beacon Designer 7.9 (Supplementary Table 14), 17 differentially expressed genes were analyzed using the $2^{-\Delta\Delta C_T}$ method. Compared with $\log_2$FC, the results show that the gene-*LOC110732446*(Beta-amylase), gene-*LOC110694254* (Beta-amylase), gene-*LOC110688573*(Beta-glucosidase), gene-*LOC110730263*(Beta-amylase), gene-*LOC110684791*(Glucose-1-phosphate adenylyltransferase), gene-*LOC110686667*(Trehalose 6-phosphate synthase), gene-*LOC110710941*(Beta-amylase), gene-*LOC110739236*(Beta-glucosidase), gene-*LOC110693889* (Maltase-glucoamylase), gene-*LOC110696059*(Beta-amylase) had the same up-regulated or down-regulated trend (Supplementary Table 15). Ten genes were randomly selected to calculate the relative expression levels of differentially expressed genes by $2^{-\Delta\Delta C_T}$. This data was compared with $\log_2$FC values from transcriptome sequencing, and these 10 genes had the same up- or down-regulation trend (Supplementary Figure 7). The results showed that the transcriptome sequencing was reliable.

## Analysis of quinoa drought-tolerance mechanisms using combined transcriptomics and metabolomics

Plants can produce a large number of specialized metabolites under drought stress; most of which are antioxidant substances. Among them, flavonoids can improve the drought tolerance of plants under drought stress (Nakabayashi et al., 2014). The biosynthesis of flavonoids originates from Cinnamoyl CoA, and p-coumaroyl-CoA is synthesized through the action of CYP73A. Then, the types and contents of metabolites formed under the action of different enzymes differ, resulting in differences in concentrations of quercetin, hesperetin 7-o-glucoside, kaempferol, and phlorizin, and finally the drought tolerance of quinoa (Figure 7). DEGs and DEMs with PCC values of >0.8 in the flavonoid-synthesis pathway were selected. Correlation analysis was performed on the differential genes and differential metabolites, and results with Pearson's correlation coefficient |PCC| greater than 0.8 were selected, with a positive PCC being a positive correlation and the opposite being a

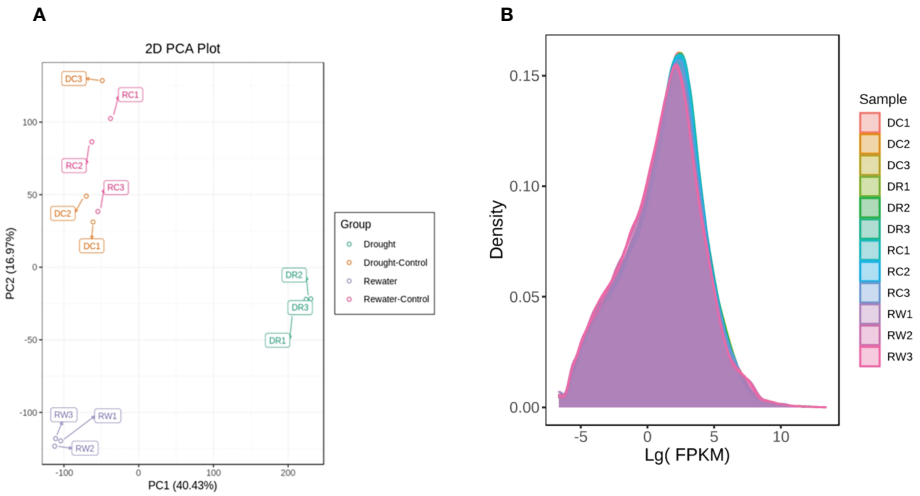



FIGURE 4
PCA of the different comparison groups analyzed in this study. **(A)** PCA diagram. **(B)** Diagram showing the expression-density distribution of PC1 in panel A representing the most obvious features that could be described in the multidimensional data matrix. The distribution for PC2 represents the most significant features that could be described in the data matrix, except for PC1. The curves are indicated with different colors to represent different samples. The abscissa represents the logarithm of FPKM of each sample and the ordinate represents the probability density.

negative correlation. Network diagram of synthetic pathways in flavonoids(Supplementary Figure 8), the *LOC110682233* gene was significantly negatively correlated with the metabolites phlorizin, hespertin-7-o-glucoside, kaempferol, and naringenin in DR vs. DC group. The *LOC110703828* gene was significantly negatively correlated with the metabolites phlorizin, hespertin-7-O-glucoside, and kaempferol. The *LOC110713661* gene was significantly positively correlated with kaempferol, phlorizin, and hespertin-7-O-glucoside, The expression of this gene was approximately 20 in different treatments and controls, and the order from high to low was DR, RW, RC and DC, therefore, this may be the key gene for drought tolerance in quinoa. The LOC110729560 gene was significantly negatively correlated with phlorizin (Supplementary Table 16). In the RW vs. RC group, the *LOC110722063* gene was significantly positively correlated with phlorizin and hespertin-7-O-glucoside

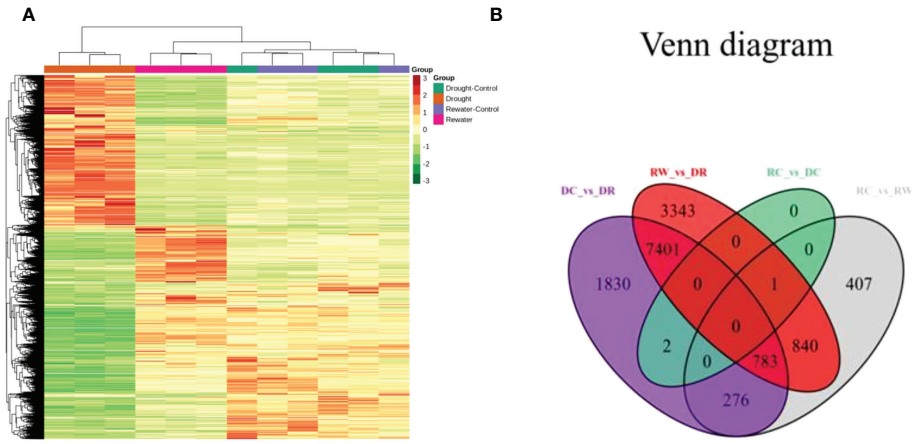



FIGURE 5
Cluster analysis of gene-expression data and a Venn diagram related to DEGs. **(A)** The abscissa represents the sample names and hierarchical clustering results, whereas the ordinate represents DEGs and hierarchical clustering results. Red shading indicates high expression, and green shading indicates low expression. **(B)** The non-overlapping regions represent specific DEGs for each group, and the overlapping areas represent DEGs common to the indicated subgroups.

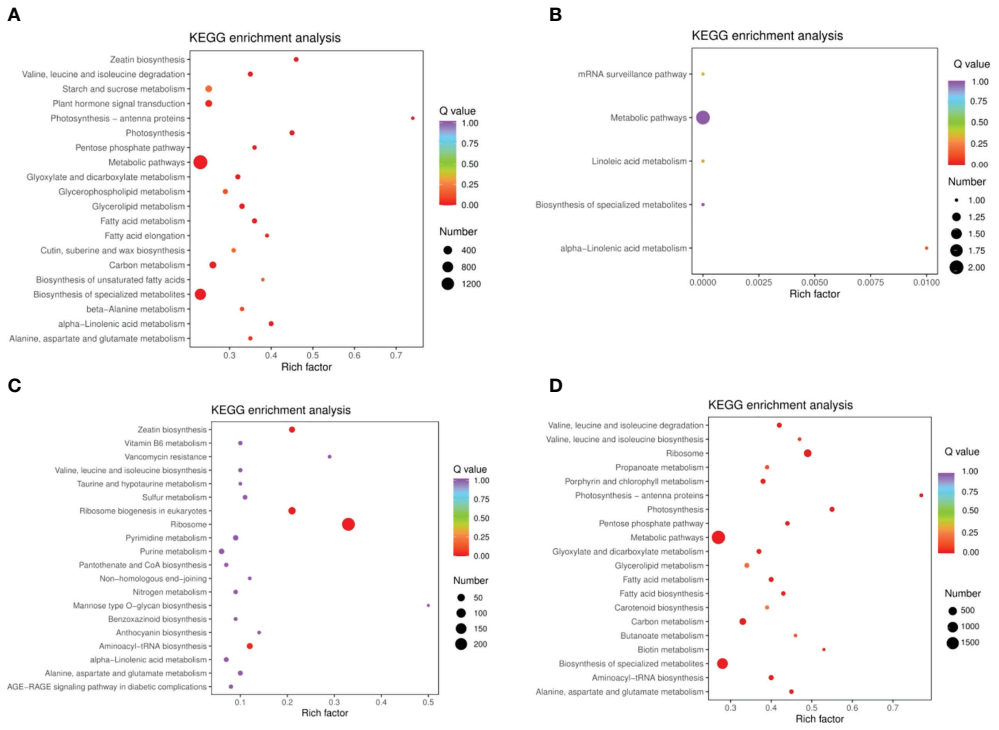



FIGURE 6
KEGG enrichment analysis of DEGs. **(A)** DC vs DR **(B)** RC vs DC **(C)** RC vs RW **(D)** RW vs DR. The ordinate represents the KEGG pathway. The abscissa represents the Rich factor the ratio of the number of different genes enriched in a pathway to the total number of annotated genes, where greater Rich factors correspond to greater enrichment. The larger the data point, the greater the number of DEGs enriched in the pathway. Deeper red shading indicates greater statistical significance in terms of the enrichment.

(Supplementary Table 17).Through metabolome analysis, we found that phlorizin, naringenin, hespertin-7-O-glucoside, dihydrokaempferol, kaempferol, and quercetin were down-regulated metabolites in the flavonoid-synthesis pathway (Supplementary Figure 9). Through transcriptomics analysis, 44, 7, and 51 differentially expressed genes related to flavonoid biosynthesis were found in the DC vs. DR, RC vs. RW, and RW vs. DR comparison groups, respectively (Supplementary Table 18). Among these, the *LOC110724467* gene (chalcone synthase, CHS, EC: 2.3.1.74) was down-regulated during drought but stabilized after rewatering. The *LOC110695126* gene (flavanone 3-hydrogenase, F3H, EC: 1.14.11.9) was up-regulated during drought and stabilized after rewatering. The *LOC110709209* gene (5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase, EC: 1.14.14.96) was stable during drought, but up-regulated after rewatering. The *LOC110736236* gene (anthocyanidin reductase, EC: 1.3.1.77) was stably expressed during drought, but down-regulated after rewatering. Eleven genes, including the *LOC110715013* (Shikimate O-hydroxycinnamoyltransferas, EC: 2.3.1.133), were up-regulated or down-regulated during drought, four genes were up-regulated after rewatering, and three genes, including the *LOC110682224* gene (DFR, EC: 1.1.1.219; 1.1.1.234) were up-regulated or down-regulated during drought, but stabilized after rewatering. Eight genes, including the *LOC110708783* gene (FLS, EC: 1.14.20.6), were up-regulated or down-regulated during drought, but stabilized after rewatering (Supplementary Figure 9, Supplementary Tables 19, 20). In the flavonoid biosynthesis pathway, the up-regulated expression of enzymes such as F3'H under drought stress, and the stable expression of enzymes such as CHS after rewatering, affected the synthesis of flavonoids, which helps achieve the goal of improving the drought tolerance of quinoa.

Under drought conditions, down-regulated expression of five genes (EC: 2.4.1.357, phlorizin synthase), including the *LOC110693894* gene, promoted the down-regulation of phlorizin. Up-regulated expression of the *LOC110690711* gene and down-regulated expression of four genes, including the *LOC110702757* gene (EC: 5.5.1.6, chalcone isomerase) contributed to naringenin down-regulation. Up-regulated expression of the *LOC110695126* gene (EC:1.14.11.9, naringenin 3-dioxygenase) contributed to dihydrokaempferol down-regulation. Up-regulated expression of three genes, such as the *LOC110698563* gene, and down-regulated expression of five genes, such as the *LOC110699285* gene (EC: 1.14.20.6, FLS) promoted kaempferol and quercetin down-regulation (Supplementary

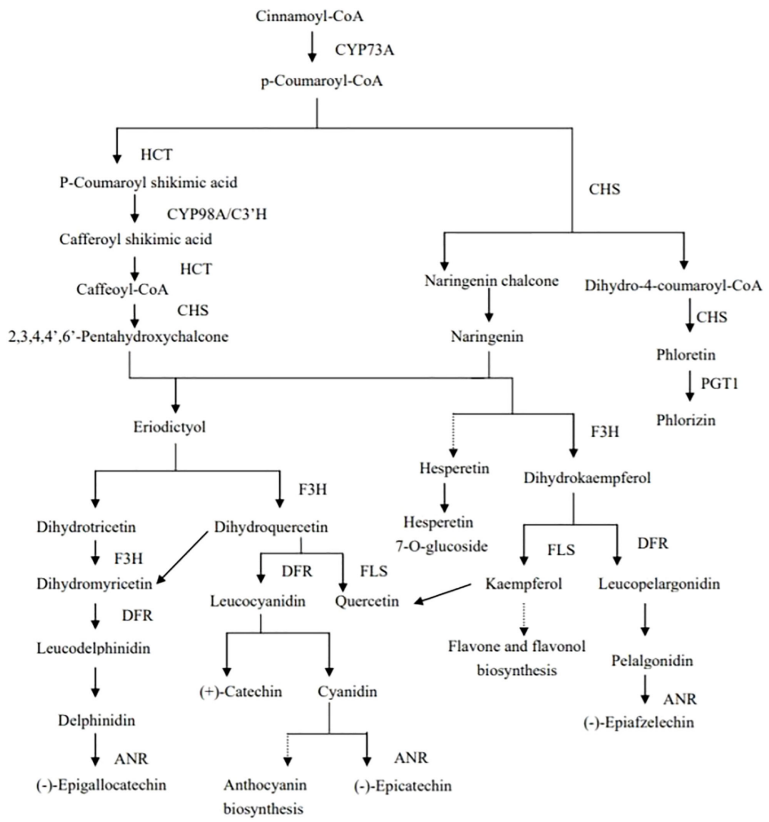



FIGURE 7
Mechanisms of flavonoid biosynthesis in quinoa.CYP73A is trans-cinnamate 4-monooxygenase, HCT is shikimate O-hydroxycinnamoyl transferase, CYP98A/C3'H is 5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase, CHS is chalcone synthase, PGT1 is phlorizin synthase,F3H is naringenin 3-dioxygenase, DFR is bifunctional dihydroflavonol 4-reductase/flavanone 4-reductase, ANR is anthocyanidin reductase, FLS is flavonol synthase.

Figure 9, Supplementary Table 19). The part of the genes shown in Table 4 were strongly correlated with metabolites in the flavonoid biosynthesis pathway, and the starch- and sucrose metabolism pathways. The *LOC110729560* gene (EC: 2.4.1.357, phlorizin synthase) was significantly negatively correlated with phlorizin, which was down-regulated during drought, but recovered after rewatering. Phlorizin contents were down-regulated during drought but changed minimally after rewatering. KEGG-based functional enrichment of DEGs (transcriptome data) and DEMs (metabolome data) showed that starch and sucrose metabolism were significantly enriched. The DEGs and metabolites with |PCC| >0.8 in the starch- and sucrose metabolism pathways were selected for further analysis. During drought treatment, 32 genes, including the *LOC110715744* gene, were significantly negatively correlated with D-fructose 6-phosphate, whereas the *LOC110712600* gene was significantly positively correlated with D-fructose 6-phosphate. The *LOC110722127* gene and other 12 genes were significantly negatively correlated with D-glucose. Fifteen genes, including the *LOC110695607* gene, were significantly negatively correlated with D-glucose 6-phosphate. The *LOC110738152* gene exhibited a significant negative correlation with the metabolite trehalose 6-phosphate, whereas the *LOC110715744* gene showed a significant positive correlation with trehalose 6-phosphate. Eighteen genes, including the *LOC110687336* gene, showed a significant negative correlation with the metabolite, glucose 1-phosphate (Supplementary Table 21). After rehydration, significant negative correlations were found between the *LOC110684752* gene and D-sucrose and D-trehalose. The *LOC110721696* gene was significantly positively correlated with D-glucose 6-phosphate (Supplementary Table 22).

Through metabolomics analysis, we determined that trehalose-6P, D-glucose-6P, D-fructose-6P, D-glucose, α-D-glucose-1P, trehalose, and sucrose were involved in the starch- and sucrose-metabolic pathways (Supplementary Figure 10). Through transcriptome analysis, 187, 30, and 205 differentially expressed genes were identified that were related to starch and sucrose metabolism among the three comparison groups (Supplementary Table 23). Three genes, including the *LOC110686362* gene (hexokinase, EC: 2.7.1.1), were down-

10.3389/fpls.2022.988861

TABLE 4 Correlation between differential genes and differential metabolites.

| Gene | Compounds | PCC | PCCP |
|---|---|---|---|
| *gene-LOC110682233* | Phloretin-2'-O-glucoside (Phlorizin) | -0.859 | 0.0003 |
| | Hesperetin-7-O-glucoside | -0.884 | 0.0001 |
| | Kaempferol (3,5,7,4'-Tetrahydroxyflavone) | -0.881 | 0.0002 |
| | Naringenin (5,7,4'-Trihydroxyflavanone) | -0.828 | 0.0009 |
| *gene-LOC110703828* | Phloretin-2'-O-glucoside (Phlorizin) | -0.817 | 0.0012 |
| | Hesperetin-7-O-glucoside | -0.838 | 0.0007 |
| | Kaempferol (3,5,7,4'-Tetrahydroxyflavone) | -0.865 | 0.0003 |
| *gene-LOC110713661* | Kaempferol (3,5,7,4'-Tetrahydroxyflavone) | 0.844 | 0.0006 |
| | Phloretin-2'-O-glucoside (Phlorizin) | 0.827 | 0.0009 |
| | Hesperetin-7-O-glucoside | 0.807 | 0.0015 |
| *gene-LOC110729560* | Phloretin-2'-O-glucoside (Phlorizin) | -0.804 | 0.0016 |
| *gene-LOC110722063* | Phloretin-2'-O-glucoside (Phlorizin) | 0.833 | 0.0008 |
| | Hesperetin-7-O-glucoside | 0.814 | 0.0013 |
| *gene-LOC110682423* | D-Fructose 6-phosphate | -0.824 | 0.0010 |
| *gene-LOC110686141* | D-Glucose 6-phosphate | -0.885 | 0.0001 |
| | D-Glucose | -0.84 | 0.0006 |
| | Glucose-1-phosphate | -0.892 | 0.0001 |
| | D-Fructose 6-phosphate | -0.919 | 0.0000 |
| *gene-LOC110686362* | D-Fructose 6-phosphate | -0.836 | 0.0007 |
| | Glucose-1-phosphate | -0.826 | 0.0009 |
| | D-Glucose 6-phosphate | -0.81 | 0.0014 |
| *gene-LOC110686667* | D-Fructose 6-phosphate | 0.827 | 0.0009 |
| *gene-LOC110687336* | Glucose-1-phosphate | -0.904 | 0.0001 |
| | D-Fructose 6-phosphate | -0.917 | 0.0000 |
| | D-Glucose 6-phosphate | -0.895 | 0.0001 |
| | D-Glucose | -0.816 | 0.0012 |
| *gene-LOC110688727* | Glucose-1-phosphate | -0.82 | 0.0011 |
| | D-Glucose | -0.812 | 0.0013 |
| | D-Glucose 6-phosphate | -0.819 | 0.0011 |
| | D-Fructose 6-phosphate | -0.88 | 0.0002 |
| *gene-LOC110690443* | D-Fructose 6-phosphate | -0.833 | 0.0008 |
| *gene-LOC110692291* | D-Glucose | -0.842 | 0.0006 |
| | D-Fructose 6-phosphate | -0.854 | 0.0004 |
| *gene-LOC110693540* | Glucose-1-phosphate | -0.804 | 0.0016 |
| | D-Fructose 6-phosphate | -0.854 | 0.0004 |
| *gene-LOC110693889* | D-Fructose 6-phosphate | -0.84 | 0.0006 |
| *gene-LOC110695607* | D-Glucose 6-phosphate | -0.808 | 0.0015 |
| | D-Fructose 6-phosphate | -0.862 | 0.0003 |
| | Glucose-1-phosphate | -0.822 | 0.0010 |
| | D-Glucose | -0.838 | 0.0007 |
| *gene-LOC110703195* | D-Fructose 6-phosphate | -0.807 | 0.0015 |
| *gene-LOC110703632* | D-Fructose 6-phosphate | -0.908 | 0.0000 |
| | Glucose-1-phosphate | -0.86 | 0.0003 |
| | D-Glucose 6-phosphate | -0.849 | 0.0005 |
| | D-Glucose | -0.816 | 0.0012 |
| Gene | Compounds | PCC | PCCP |
| *gene-LOC110704748* | D-Fructose 6-phosphate | -0.808 | 0.0015 |
| *gene-LOC110709538* | D-Fructose 6-phosphate | -0.856 | 0.0004 |

*(Continued)*

10.3389/fpls.2022.988861

TABLE 4 Continued

| Gene | Compounds | PCC | PCCP |
|---|---|---|---|
| | D-Glucose 6-phosphate | -0.823 | 0.0010 |
| | Glucose-1-phosphate | -0.835 | 0.0007 |
| *gene-LOC110710484* | D-Fructose 6-phosphate | -0.843 | 0.0006 |
| *gene-LOC110710504* | D-Fructose 6-phosphate | -0.826 | 0.0009 |
| | Glucose-1-phosphate | -0.803 | 0.0017 |
| | D-Glucose | -0.815 | 0.0012 |
| *gene-LOC110711798* | D-Fructose 6-phosphate | -0.834 | 0.0007 |
| *gene-LOC110712600* | D-Fructose 6-phosphate | 0.833 | 0.0008 |
| *gene-LOC110714985* | Glucose-1-phosphate | -0.834 | 0.0007 |
| | D-Fructose 6-phosphate | -0.87 | 0.0002 |
| | D-Glucose 6-phosphate | -0.835 | 0.0007 |
| *gene-LOC110715744* | D-Fructose 6-phosphate | -0.845 | 0.0005 |
| | Glucose-1-phosphate | -0.89 | 0.0001 |
| | Trehalose 6-phosphate | 0.804 | 0.0016 |
| | D-Glucose 6-phosphate | -0.894 | 0.0001 |
| *gene-LOC110717307* | D-Glucose 6-phosphate | -0.86 | 0.0003 |
| | Glucose-1-phosphate | -0.858 | 0.0004 |
| | D-Fructose 6-phosphate | -0.89 | 0.0001 |
| *gene-LOC110719674* | D-Fructose 6-phosphate | -0.804 | 0.0016 |
| *gene-LOC110720321* | D-Glucose 6-phosphate | -0.82 | 0.0011 |
| | Glucose-1-phosphate | -0.826 | 0.0009 |
| | D-Fructose 6-phosphate | -0.878 | 0.0002 |
| *gene-LOC110721974* | D-Glucose | -0.818 | 0.0012 |
| *gene-LOC110722127* | D-Glucose | -0.83 | 0.0008 |
| | D-Fructose 6-phosphate | -0.845 | 0.0005 |
| | Glucose-1-phosphate | -0.806 | 0.0015 |
| *gene-LOC110723159* | D-Glucose 6-phosphate | -0.827 | 0.0009 |
| | D-Glucose | -0.813 | 0.0013 |
| | D-Fructose 6-phosphate | -0.874 | 0.0002 |
| | Glucose-1-phosphate | -0.825 | 0.0009 |
| *gene-LOC110723897* | D-Fructose 6-phosphate | -0.836 | 0.0007 |
| *gene-LOC110725158* | D-Glucose 6-phosphate | -0.86 | 0.0003 |
| | Glucose-1-phosphate | -0.853 | 0.0004 |
| | D-Glucose | -0.813 | 0.0013 |
| | D-Fructose 6-phosphate | -0.906 | 0.0001 |
| *gene-LOC110726353* | D-Fructose 6-phosphate | -0.83 | 0.0008 |
| | D-Glucose | -0.819 | 0.0011 |
| *gene-LOC110727926* | D-Glucose 6-phosphate | -0.837 | 0.0007 |
| | Glucose-1-phosphate | -0.843 | 0.0006 |
| | D-Fructose 6-phosphate | -0.84 | 0.0006 |
| *gene-LOC110728048* | D-Fructose 6-phosphate | -0.812 | 0.0013 |
| *gene-LOC110733459* | D-Glucose 6-phosphate | -0.946 | 0.0000 |
| | D-Fructose 6-phosphate | -0.91 | 0.0000 |
| | Glucose-1-phosphate | -0.941 | 0.0000 |
| *gene-LOC110736307* | D-Fructose 6-phosphate | -0.804 | 0.0016 |
| *gene-LOC110738152* | Trehalose 6-phosphate | -0.837 | 0.0007 |
| *gene-LOC110684752* | D-Sucrose | -0.817 | 0.0012 |
| | D-Trehalose | -0.837 | 0.0007 |
| *gene-LOC110721696* | D-Glucose 6-phosphate | 0.82 | 0.0011 |

regulated during drought, but showed stable expression after rewatering. Eight genes, including *LOC110693889* gene (maltase-glucoamylase, EC: 3.2.1.20), were down-regulated during drought, but stably expressed after rehydrating. Four genes, including the *LOC110690627* gene (glucan 1,3-beta-glucosidase, EC: 3.2.1.58), were down-regulated during drought, but stable after rehydration. Three genes, including *LOC110689796* gene (sucrose synthase, EC: 2.4.1.13), were up-regulated during drought, but down-regulated after rehydration. The *LOC110729741* gene (phosphoglucomutase, EC: 5.4.2.2) was down-regulated during drought but stabilized after rehydration. The *LOC110719410* gene (ADP-sugar diphosphatase, EC: 3.6.1.21) was up-regulated during dry drought but stabilized after rehydration. The *LOC110683757* gene (starch synthase, EC: 2.4.1.21) was up-regulated during drought, but down-regulated after rehydration. The *LOC110738898* gene (granule-bound starch synthase, EC: 2.4.1.242) was up-regulated during drought but stabilized after rehydration. Four genes, including *LOC110703195* gene (4-alpha-glucanotransferase, EC: 2.4.1.25), were down-regulated during drought, but stably expressed after rehydrating (Supplementary Figures 10 and Supplementary Tables 24, 25). In the starch- and sucrose metabolism pathways, the expression levels of enzyme-related genes such as ADP-sugar diphosphatase were up/down-regulated under drought stress, and the expression levels of enzyme-related genes such as hexokinase were restored after rehydration, which enhanced the drought tolerance of quinoa. The *LOC110686362* gene and two other genes were down-regulated (EC: 2.7.1.1, hexokinase), which promoted D-fructose-6P and D-glucose-6P down-regulation. The down-regulated expression of eight genes (EC: 3.2.1.20, maltase-glucoamylase), including the *LOC110693889* gene, promoted the down-regulation of D-glucose, and four genes (EC: 3.2.1.58, glucan 1,3-beta-glucosidase) including the *LOC110690627* gene, promoted the down-regulation of D-glucose. The down-regulation of three genes (EC: 2.4.1.13, sucrose synthase) including the *LOC110689796* gene promoted the down-regulation of sucrose. The down-regulation of the *LOC110729741* gene (EC: 5.4.2.2, phosphoglucomutase) was associated with that of D-glucose-6P. Up-regulation of the *LOC110719410* gene (EC: 3.6.1.21, ADP sugar diphosphatase) promoted the down regulation of α-D-glucose-1P. Down-regulation of the *LOC110703195* gene and three other genes (EC: 2.4.1.25, 4-alpha-glucanotransferase) promoted the down-regulation of D-glucose. Up-regulation of the *LOC110737398* gene (EC: 3.1.3.24, sucrose-6-phosphatase) promoted the up-regulation of sucrose (Supplementary Figure 10, Supplementary Table 24).Which the *LOC110703632* gene (EC: 2.7.1.4, fructokinase) was significantly negatively correlated with the metabolism of D-fructose 6-phosphate, which was down-regulated during drought, but tended to be steadily expressed after rehydration, whereas the D-fructose 6-phosphate content was down-regulated during drought, but gradually recovered after rehydration. Gene-*LOC110709538* and gene-*LOC110693540* expressed 0 in DR, but low after rehydration. In this study, 42 differential genes (5 for flavonoid biosynthesis and 37 for starch and sucrose metabolism) and 11 differential metabolites (4 for flavonoid biosynthesis and 7 for starch and sucrose metabolism) with differential metabolites were identified as key factors of drought tolerance in quinoa leaves (Tables 4 – 6). Gene-*LOC110738152* was not expressed in DC, and was very low or not expressed in RC and RW. This gene was highly expressed in DR. Therefore, gene-*LOC110738152* may be the key gene to improve the drought tolerance of quinoa.

## Discussion

Water scarcity has devastating effects on the yield and quality of major crops, and water deficits caused by drought can lead to severe growth retardation and yield loss (Zhang et al, 2014). When plants are exposed to a water deficit, they undergo highly complex morphological, physiological, biochemical, and molecular changes (Bhargava and Sawant, 2013). Drought stress can reduce plant heights, ear lengths, chlorophyll contents, and root and stem biomass, thereby reducing grain yields (Abbas et al., 2018). In the present study, similar conclusions to previous studies were reached, whereby several parameters of the morphology of the drought tolerant genotype Dianli 129 showed increase or a small decrease under drought conditions. The ability to maintain key biological functions during drought and recover quickly after rewatering are important determinants of the maximum lifetime productivity and high drought tolerance (Abid et al., 2018). Dianli 129 showed better recovery ability after rewatering, and its drought tolerance mechanism may occur by reducing the above-ground part biomass and leaf area while maintaining a larger root-to-crown ratio to better maintain normal growth. In contrast, the drought-sensitive genotype Dianli 114 demonstrated a decline in total root length and total root volume under drought and did not recover well after rehydration, likely because drought stress inhibited root growth. Photosynthesis is enhanced under drought stress, which confers a high potential to withstand drought stress (Abid et al., 2016). The physiological changes of tea plants under drought and rehydration conditions have been studied (Liu et al., 2015). With the development of drought stress, the MDA, SS, and Pro contents, as well as the SOD and CAT activities increased significantly, however, decreased rapidly after rehydration. The ABA and SA levels peaked at an early stage of drought stress and then declined rapidly (Liu et al., 2015). Pro, total SS, ascorbic acid, and ABA levels increased in the drought-tolerant varieties, whereas hydrogen peroxide, superoxide anions, lipid peroxidation, and electrolyte leakage increased rapidly in the drought-sensitive varieties, indicating that the tolerant varieties showed higher antioxidant capacities and stronger protective mechanisms (Das et al., 2015). Proteins

10.3389/fpls.2022.988861

TABLE 5 The $log_2FC$ of differentially expressed genes.

| Gene | EC | Enzyme | $Log_2FC$ DC_vs _DR | $log_2FC$ RC_VS_ RW | $log_2FC$ RW_vs_DR |
|---|---|---|---|---|---|
| *gene-LOC110682233* | 1.1.1.219 | Dihydroflavonol-4-reductase | -1.5923 | -1.1454 | -0.0146 |
| *gene-LOC110703828* | 2.1.1.104 | Caffeoyl-CoA O-methyltransferase | -1.0410 | -0.7165 | 0.1076 |
| *gene-LOC110713661* | 2.3.1.133 | Shikimate O-hydroxycinnamoyltransferase | 1.4368 | 0.7523 | 0.2064 |
| *gene-LOC110722063* | 2.3.1.133 | Shikimate O-hydroxycinnamoyltransferase | 0.9084 | 1.7961 | -0.6467 |
| *gene-LOC110729560* | 2.4.1.357 | Phlorizin synthase | -2.3684 | -0.9085 | -1.1941 |
| *gene-LOC110686141* | 2.4.1.1 | Glycogen phosphorylase | -2.1955 | -0.8257 | -1.234 |
| *gene-LOC110703195* | 2.4.1.1 | Glycogen phosphorylase | -2.0669 | -0.2735 | -1.8612 |
| *gene-LOC110711798* | 2.4.1.1 | Glycogen phosphorylase | -2.4101 | -0.6955 | -1.823 |
| *gene-LOC110714985* | 2.4.1.1 | Glycogen phosphorylase | -2.547 | -1.1384 | -1.3998 |
| *gene-LOC110717307* | 2.4.1.1 | Glycogen phosphorylase | -1.4565 | -0.7021 | -0.6242 |
| *gene-LOC110722127* | 2.4.1.1 | Glycogen phosphorylase | -1.7839 | -0.3585 | -1.2523 |
| *gene-LOC110686667* | 2.4.1.15 | Trehalose 6-phosphate synthase | 1.3188 | 0.5304 | 0.8782 |
| *gene-LOC110686362* | 2.7.1.1 | Hexokinase | -1.2759 | -0.3831 | -0.799 |
| *gene-LOC110688727* | 2.7.1.1 | Hexokinase | -1.848 | -0.7427 | -1.2488 |
| *gene-LOC110703632* | 2.7.1.4 | Fructokinase | -1.1064 | -0.4192 | -0.7366 |
| *gene-LOC110720321* | 2.7.1.4 | Fructokinase | -1.5628 | -0.5718 | -1.0074 |
| *gene-LOC110721696* | 2.7.1.4 | Fructokinase | 0.5286 | 1.3556 | -0.7011 |
| *gene-LOC110695607* | 3.1.3.12 | Trehalose 6-phosphate phosphatase | -1.2305 | -0.2609 | -0.8932 |
| *gene-LOC110725158* | 3.1.3.12 | Trehalose 6-phosphate phosphatase | -1.9397 | -1.161 | -1.0502 |
| *gene-LOC110690443* | 3.2.1.1 | Alpha-amylase | -1.2586 | -0.8134 | -0.668 |
| *gene-LOC110728048* | 3.2.1.2 | Beta-amylase | -1.2861 | -0.3646 | -1.0604 |
| *gene-LOC110693889* | 3.2.1.20 | Maltase-glucoamylase | -2.0635 | -0.7661 | -1.3983 |
| *gene-LOC110687336* | 3.2.1.21 | Beta-glucosidase | -1.9779 | -0.7351 | -0.9874 |
| *gene-LOC110693540* | 3.2.1.21 | Beta-glucosidase | -6.080 | -2.6781 | / |
| *gene-LOC110738152* | 3.2.1.21 | Beta-glucosidase | 7.413 | 0.1203 | 3 |
| *gene-LOC110684752* | 3.2.1.26 | Beta-fructofuranosidase | -0.7047 | -1.5606 | 0.187 |
| *gene-LOC110710504* | 3.2.1.26 | Beta-fructofuranosidase | -2.2885 | -0.5805 | -1.5052 |
| *gene-LOC110723159* | 3.2.1.26 | Beta-fructofuranosidase | -2.0124 | -0.8969 | -1.1699 |
| *gene-LOC110682423* | 3.2.1.39 | Glucan endo-1,3-beta-D-glucosidase | -3.2789 | -0.6438 | -2.344 |
| *gene-LOC110704748* | 3.2.1.39 | Glucan endo-1,3-beta-D-glucosidase | -2.3616 | -0.3512 | -1.8591 |
| *gene-LOC110710484* | 3.2.1.39 | Glucan endo-1,3-beta-D-glucosidase | -2.2403 | -0.5962 | -1.6781 |
| *gene-LOC110715744* | 3.2.1.39 | Glucan endo-1,3-beta-D-glucosidase | -1.0465 | -0.7182 | -0.123 |
| *gene-LOC110721974* | 3.2.1.39 | Glucan endo-1,3-beta-D-glucosidase | -2.6315 | -0.5894 | -2.3957 |
| *gene-LOC110712600* | 3.2.1.4 | Endoglucanase | 2.1741 | 0.8014 | 1.6283 |
| Gene | EC | Enzyme | $Log_2FC$ DC_vs _DR | $log_2FC$ RC_VS_ RW | $log_2FC$ RW_vs_DR |
| *gene-LOC110719674* | 3.2.1.4 | Endoglucanase | -1.2416 | -0.1616 | -0.9423 |
| *gene-LOC110723897* | 3.2.1.4 | Endoglucanase | -3.1699 | -1.1673 | -1.8015 |
| *gene-LOC110726353* | 3.2.1.4 | Endoglucanase | -4.812 | -0.848 | / |
| *gene-LOC110727926* | 3.2.1.4 | Endoglucanase | -1.2437 | -0.4567 | -0.5519 |
| *gene-LOC110733459* | 3.2.1.4 | Endoglucanase | -3.0444 | -2.663 | -0.585 |
| *gene-LOC110736307* | 3.2.1.4 | Endoglucanase | -3.0686 | -0.5679 | -2.8164 |
| *gene-LOC110709538* | 3.2.1.58 | Glucan 1,3-beta-glucosidase | -6.174 | -1.2345 | / |
| *gene-LOC110692291* | 3.2.1.68 | Isoamylase | -2.7769 | -0.8795 | -2.0478 |

TABLE 6 The $log_2$FC of differential metabolites.

| Compounds | $Log_2$FC DC_vs _DR | $log_2$FC RC_VS_ RW | $log_2$FC RW_vs_DR |
|---|---|---|---|
| Phlorizin | -1.7576 | -2.0336 | 0.1234 |
| Kaempferol | -3.1101 | -2.0164 | -1.4189 |
| Hesperidin-7-O-glucoside | -1.5406 | -1.7914 | -0.1764 |
| Naringenin | -12.4423 | -3.1737 | -10.0105 |
| D-Glucose 6-phosphate | -1.7741 | -1.5406 | -0.0440 |
| D-Fructose 6-phosphate | -1.6879 | -1.4563 | 0.2797 |
| D-Glucose | -1.0217 | -0.7181 | 0.2220 |
| Glucose 1-phosphate | -1.8962 | -1.4918 | -0.1369 |
| Trehalose 6-phosphate | 2.0636 | 0.2035 | 1.4972 |
| D-Sucrose | -0.9879 | -2.9574 | 2.3223 |
| D-Trehalose | -1.0142 | -2.5478 | 1.8593 |

rich in glutathione, taurine, hypotaurine, methionine, cysteine, and other amino acids involved in sulfur-dependent metabolic pathways were significantly altered under drought stress (Wang et al., 2017). In this study, the drought tolerant genotype Dianli 129 maintained a high total antioxidant capacity under drought. Soluble protein content increased sharply under drought conditions, which may be closely related to the assay method, varietal differences, stress intensity and time. Relative conductivity increased less, chlorophyll content also decreased less and recovered better or showed supercompensation after rehydration. Under drought stress and rehydration, the drought tolerant genotype Dianli 129 showed higher enzyme activity and higher osmoregulatory substances, and therefore exhibited a greater ability to resist drought stress as well as recover after rehydration. This genotype could be used as an important drought tolerant quinoa germplasm resource.

In recent years there have been many histological studies demonstrating that plant resistance is associated with flavonoids. Transcriptome and metabolite analysis of grapes under drought stress showed that water scarcity regulated the expression of structural genes related to phenylpropane, flavonoids, carotenoids, and terpenoids, and these metabolic pathways underwent transcriptional regulation in grapes under water stress (Savoi et al., 2016). Comparative transcriptome analysis of *Ammopiptanthus mongolicus* under drought and cold stress revealed that flavonoid biosynthesis genes were enriched in DEG up-regulated by both stresses (Wu et al., 2014). The effects of fulvic acid on genes and metabolites of tea plants during different drought stress stages were studied by transcriptomics and metabolomics. The results showed that fulvic acid could enhance ascorbic acid metabolism, improve glutathione metabolism, and promote the biosynthesis of flavonoids (e.g., C4H, CHS, F3'5'h, F3H, kaempferol, and quercetin). Thus, fulvic acid could significantly improve the antioxidant-defense abilities of tea plants under drought stress, to enhance the drought tolerance of tea plants (Sun et al., 2020). Sugars (sucrose and trehalose) influence the regulation of cell osmotic pressure during the stress response of plants during drought (Shinozaki and Yamaguchi-Shinozaki, 2007). Similar to previous studies in the present study, DEGs were analyzed using the KEGG database, and the haircut showed a significant enrichment of KEGG metabolite pathways including flavonoid biosynthesis, suggesting that possibly quinoa drought tolerance is also associated with flavonoids. The gene gene-LOC110713661 was significantly positively correlated with flavonoids (Kaempferol, Phlorizin, Hesperetin-7-O-glucoside), which means that this gene promotes the synthesis of flavonoids to tolerate drought. However, gene-LOC110738152 was highly expressed in the drought treatment, but not in the drought control, and had very low or no expression in the rewatering treatment and the rewatering control. We therefore infer that gene-LOC110713661 and gene-LOC110738152 may be key genes for enhancing drought tolerance in quinoa. The tolerance to drought stress may be due to the elevated expression levels of these two genes, and the drought tolerance genes may still be lowly expressed or slowly not expressed after later rewatering due to environmental changes. This study involved six metabolites (phlorizin, naringenin, hespertin-7-O-glucoside, dihydrokaempferol, kaempferol, and quercetin), as well as CHS, naringenin 3-dioxygenase, anthocyanidin reductase, and FLS, among other enzymes. Previous data showed that DREB, ERF, NAC, and WRKY were jointly regulated by drought and cold stress (Wu et al., 2014). The main TFs examined in this study include AP2, MYB, BHLH, WRKY, NAC, and bZIP. The gene expression levels may be related to drought tolerance in quinoa. Trehalose-6p, D-glucose-6P, D-fructose-6P, D-glucose, α-D-glucose-1P, trehalose, and sucrose, as well as related genes such as hexokinase, fructokinase, malpase-glucoamylase, and glucan 1,3-beta-glucosidase, may be correlated with drought tolerance in quinoa. In this study, 42 DEGs and 11 DEMs were identified

as key factors of drought tolerance in quinoa leaves. We also discovered 3,259 genes potentially related to drought tolerance in quinoa (Supplementary Table 26).

## Conclusion

To study the mechanisms mediating responses to drought stress at the seedling stage in quinoa, we adopted a drought stress–rehydration method and conducted transcriptomics and metabolomics analyses on the drought-tolerant Dianli 129 quinoa genotype. The results showed that the gene-*LOC110713661* and gene-*LOC110738152* may be key genes for drought tolerance in quinoa, and our findings provide a theoretical basis for breeding drought-tolerant quinoa genotypes. In this study, we investigated the biosynthetic pathways of flavonoids in quinoa and the active roles of the starch- and sucrose metabolism pathways in quinoa under drought stress. The results of this study confirm that a strategy to protect quinoa from drought stress is to regulate the antioxidant systems and the accumulation of metabolites. The Dianli 129 quinoa genotype showed a high ability to resist drought stress and to recover after rehydration.

## Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number PRJNA857812.

## Author contributions

XH: Writing - Original Draft, Methodology. LL: Conceptualization, Writing- Review & Editing. YL: Formal analysis, Methodology. ZK: Data Curation, Visualization. YL: Data Curation, Investigation. QW, JL: Methodology, Visualization. PZ,YG: Formal analysis, Investigation. PQ: Supervision, Project administration, Funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This research was funded by Yunnan Academician Workstation (2019IC006), Central Government for Guiding Local Science and Technology Development (2020, Quinoa), and the Kunming Science and Technology Innovation Center (2019-1-N-25318000002317).

## Acknowledgments

We thank the staff of Wuhan Metware Biotechnology Co., Ltd.(Wuhan, China), for their support during the metabolite data analysis. We would like to thank Editage (www.editage.cn) for English language editing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.988861/full#supplementary-material

## References

Abbas, T., Rizwan, M., Ali, S., Adrees, M., Mahmood, A., Zia-Ur-Rehman, M., et al. (2018). Biochar application increased the growth and yield and reduced cadmium in drought stressed wheat grown in an aged contaminated soil. *Ecotoxicol. Environ. Safety* 148, 825–833. doi: 10.1016/j.ecoenv.2017.11.063

Abid, M., Ali, S., Qi, L. K., Zahoor, R., Tian, Z., Jiang, D., et al. (2018). Physiological and biochemical changes during drought and recovery periods at tillering and jointing stages in wheat (*Triticum aestivum l.*). *Sci. Rep.* 8, 4615. doi: 10.1038/s41598-018-21441-7

Abid, M., Tian, Z., Ata-Ul-Karim, S. T., Liu, Y., Cui, Y., Zahoor, R., et al. (2016). Improved tolerance to post-anthesis drought stress by pre-drought priming at vegetative stages in drought-tolerant and -sensitive wheat cultivars. *Plant Physiol. Biochem.* 106, 218–227. doi: 10.1016/j.plaphy.2016.05.003

An, X., Chen, J., Zhang, J., Liao, Y., Dai, L., Wang, B., et al. (2015). Transcriptome profiling and identification of transcription factors in ramie (*Boehmeria nivea l. gaud*) in response to PEG treatment, using illumina paired-end sequencing technology. *Int. J. Mol. Sci.* 16, 3493–3511. doi: 10.3390/ijms16023493

Ayasan, T. (2020). Determination of nutritional value of some quinoa varieties. *Turk J. Vet. Anim. Sci.* 44, 950–954. doi: 10.3906/vet-2003-53

Benaffari, W., Boutasknit, A., Anli, M., Ait-El-Mokhtar, M., Ait-Rahou, Y., Ben-Laouane, R., et al. (2022). The native arbuscular mycorrhizal fungi and vermicompost-based organic amendments enhance soil fertility, growth performance, and the drought stress tolerance of quinoa. *Plants (Basel)* 11, 393. doi: 10.3390/plants11030393

Bhargava, S., and Sawant, K. (2013). Drought stress adaptation: Metabolic adjustment and regulation of gene expression. *Plant Breed.* 132, 21–32. doi: 10.1111/pbr.12004

Chaves, M., and Oliveira, M. (2004). Mechanism's underlying plant resilience to water deficits: prospects for water-saving agriculture. *J. Exp. Botany* 55, 2365–2384. doi: 10.1093/jxb/erh269

D'Oria, A., Courbet, G., Billiot, B., Jing, L., Pluchon, S., Arkoun, M., et al. (2022). Drought specifically downregulates mineral nutrition: Plant ionomic content and associated gene expression. *Plant Direct.* 6, e402. doi: 10.1002/pld3.402

Das, A., Mukhopadhyay, M., Sarkar, B., Saha, D., and Mondal, T. K. (2015). Influence of drought stress on cellular ultrastructure and antioxidant system in tea cultivars with different drought sensitivities. *J. Environ. Biol.* 36, 875–882.

Dos Reis, S. P., Lima, A. M., and De Souza, C. R. B. (2012). Recent molecular advances on downstream plant responses to abiotic stress. *Int. J. Mol. Sci.* 13, 8628–8647. doi: 10.3390/ijms13078628

Dumschott, K., Wuyts, N., Alfaro, C., Castillo, D., Fiorani, F., and Zurita-Silva, A. (2022). Morphological and physiological traits associated with yield under reduced irrigation in chilean coastal lowland quinoa. *Plants (Basel)* 11, 323. doi: 10.3390/plants11030323

Fadiji, A., Santoyo, G., Yadav, A., and Babalola, O. (2022). Efforts towards overcoming drought stress in crops: Revisiting the mechanisms employed by plant growth-promoting bacteria. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.962427

FAO regional office for latin america and the caribbean, & proinpa (2011) *Quinoa: An ancient crop to contribute to world food security*. Available at: http://www.fao.org/alc/file/media/pubs/2011/cultivo_quinua_en.pdf.

Filho, A. M. M., Pirozi, M. R., Borges, J. T. D. S., Pinheiro Sant´Ana, H. M., Chaves, J. B. P., and Coimbra, J. S. D. R. (2017). Quinoa: Nutritional,functional, and antinutritional aspects. *Crit. reciews Food Sci. Nutr.*, 1618–1630. doi: 10.1080/10408398.2014.1001811

Gawlik-Dziki, U., Sułkowski, M., Dziki, D., Baraniak, B., and Czy˙z, J. (2013). Antioxidant and anticancer activities of chenopodium quinoa leaves extracts–*In vitro* study. *Food Chem. Toxicol.* 57, 154–160. doi: 10.1016/j.fct.2013.03.023

Kaur, G., and Asthir, B. (2017). Molecular responses to drought stress in plants. *Biol. plantarum* 61, 201–209. doi: 10.1007/s10535-016-0700-9

Lenka, S. K., Katiyar, A., Chinnusamy, V., and Bansal, K. C. (2011). Comparative analysis of drought-responsive transcriptome in indica rice genotypes with contrasting drought tolerance. *Plant Biotechnol. J.* 9, 315–327. doi: 10.1111/j.1467-7652.2010.00560.x

Lin, P., and Chao, Y. (2021). Chenopodium quinoa different drought-tolerant mechanisms in quinoa (*Willd.*) and djulis *(Koidz.*) based on physiological analysis. *Plants (Basel)* 10, 2279. doi: 10.3390/plants10112279

Liu, S., Yao, M., Ma, C., Jin, J., Ma, J., Li, C., et al. (2015). Physiological changes and differential gene expression of tea plant under dehydration and rehydration conditions. *Scientia Horticulturae* 184, 129–141. doi: 10.1016/j.scienta.2014.12.036

Melini, V., and Melini, F. (2021). Functional components and anti-nutritional factors in gluten-free grains: A focus on quinoa seeds. *Foods* 10, 351. doi: 10.3390/foods10020351

Mu, Q., Cai, H., Sun, S., Wen, S., Xu, J., Dong, M., et al. (2021). The physiological response of winter wheat under short-term drought conditions and the sensitivity of different indices to soil water changes. *Agric. Water Management* 243, 106475. doi: 10.1016/j.agwat.2020.106475

Nair, A. S., Abraham, T., and Jaya, D. (2008). Studies on the changes in lipid peroxidation and antioxidants in drought stress induced cowpea (*Vigna unguiculata l.*) varieties. *J. Environ. Biol.* 29, 689–691. doi: 10.2112/08A-0006.1

Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., et al. (2014). Enhancement of oxidative and drought tolerance in arabidopsis by overaccumulation of antioxidant flavonoids. *Plant J.* 77, 367–379. doi: 10.1111/tpj.12388

Narayanasamy, S., Thangappan, S., and Uthandi, S. (2020). Plant growth-promoting bacillus sp. cahoots moisture stress alleviation in rice genotypes by triggering antioxidant defense system. *Microbiol. Res.* 239, 126518. doi: 10.1016/j.micres.2020.126518

Nowak, V., Du, J., and Charrondiere, U. R. (2016). Assessment of the nutritional composition of quinoa (*Chenopodium quinoa* willd.). *Food Chem.* 193, 47–54. doi: 10.1016/j.foodchem.2015.02.111

Pathan, S., Eivazi, F., Valliyodan, B., Paul, K., Ndunguru, G., and Clark, K. (2019). Nutritional composition of the green leaves of quinoa (*Chenopodium quinoa* willd.). *J. Food Res.* 8, 55–65. doi: 10.5539/jfr.v8n6p55

Philippot, L., Raaijmakers, J. M., Lemanceau, P., and Van Der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* 11, 789–799. doi: 10.1038/nrmicro3109

Repo-Carrasco, R., Espinoza, C., and Jacobsen, S.-E. (2003). Nutritional value and use of the andean crops quinoa (*Chenopodium quinoa*) and kañiwa (*Chenopodium pallidicaule*). *Food Rev. Int.* 19, 179–189. doi: 10.1081/FRI-120018884

Savoi, S., Wong Darren, C. J., Arapitsas, P., Miculan, M., Bucchetti, B., Peterlunger, E., et al. (2016). Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (*Vitis vinifera l.*). *BMC Plant Biol.* 16, 67. doi: 10.1186/s12870-016-0760-1

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *J. Exp. Botany* 58, 221–227. doi: 10.1093/jxb/erl164

Sun, J., Qiu, C., Ding, Y., Wang, Y., Sun, L., Fan, K., et al. (2020). Fulvic acid ameliorates drought stress-induced damage in tea plants by regulating the ascorbate metabolism and flavonoids biosynthesis. *BMC Genomics* 21, 411. doi: 10.1186/s12864-020-06815-4

Sun, W., Wei, J., Wu, G., Xu, H., Chen, Y., Yao, M., et al. (2022). CqZF-HD14 enhances drought tolerance in quinoa seedlings through interaction with *CqHIPP34* and *CqNAC79*. *Plant Sci.* 323, 111406. doi: 10.1016/j.plantsci.2022.111406

Tan, M., Zhao, Q., and Zhao, B. (2021). Physicochemical properties, structural characterization and biological activities of polysaccharides from quinoa (*Chenopodium quinoa* willd.) seeds. *Int. J. Biol. Macromol.* 193, 1635–1644. doi: 10.1016/j.ijbiomac.2021.10.226

Verma, V., Ravindran, P., and Kumar, P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16, 86. doi: 10.1186/s12870-016-0771-y

Villacrés, E., Quelal, M., Galarza, S., Iza, D., and Silva, E. (2022). Chenopodium quinoa nutritional value and bioactive compounds of leaves and grains from quinoa ( *Chenopodium quinoa* willd.). *Plants (Basel)* 11, 213. doi: 10.3390/plants11020213

Wang, Y., Fan, K., Wang, J., Ding, Z., Wang, H., Bi, C., et al. (2017). Proteomic analysis of camellia sinensis (L.) reveals a synergistic network in the response to drought stress and recovery. *J. Plant Physiol.* 219, 91–99. doi: 10.1016/j.jplph.2017.10.001

Wang, Y., Zeng, X., Xu, Q., Mei, X., Yuan, H., Jiabu, D., et al. (2019). Metabolite profiling in two contrasting Tibetan hulless barley cultivars revealed the core salt-responsive metabolome and key salt-tolerance biomarkers. *AoB Plants* 11, plz021. doi: 10.1093/aobpla/plz021

Wu, Y., Wei, W., Pang, X., Wang, X., Zhang, H., Dong, B., et al. (2014). Comparative transcriptome profiling of a desert evergreen shrub, ammopiptanthus mongolicus, in response to drought and cold stresses. *BMC Genomics* 15, 671. doi: 10.1186/1471-2164-15-671

Yang, A., Akhtar, S., Li, L., Fu, Q., Li, Q., Naeem, M., et al. (2020). Biochar mitigates combined effects of drought and salinity stress in quinoa. *Agronomy* 10, 912. doi: 10.3390/agronomy10060912

Yuan, H., Zeng, X., Shi, J., Xu, Q., Wang, Y., Jabu, D., et al. (2018). Time-course comparative metabolite profiling under osmotic stress in tolerant and sensitive Tibetan hulless barley. *BioMed. Res. Int.* 2018, 9415409. doi: 10.1155/2018/9415409

Yuan, H., Zeng, X., Yang, Q., Xu, Q., Wang, Y., Jabu, D., et al. (2018). Gene coexpression network analysis combined with metabonomics reveals the resistance responses to powdery mildew in Tibetan hulless barley. *Sci. Rep.* 8, 14928. doi: 10.1038/s41598-018-33113-7

Zhang, X., Lu, G., Long, W., Zou, X., Li, F., and Nishio, T. (2014). Recent progress in drought and salt tolerance studies in brassica crops. *Breed. Sci.* 64, 60–73. doi: 10.1270/jsbbs.64.60

Zhang, D., Wei, X., Liu, Z., Wu, X., Bao, C., Sun, Y., et al. (2021). Transcriptome analysis reveals the molecular mechanism of GABA accumulation during quinoa (*Chenopodium quinoa* willd.) germination. *J. Agric. Food Chem.* 69, 12171–12186. doi: 10.1021/acs.jafc.1c02933

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 January 2023
DOI 10.3389/fpls.2022.1057645



OPEN ACCESS

EDITED BY
Marta Sousa Silva,
University of Lisbon, Portugal

REVIEWED BY
Emmanuelle Meudec,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France
Melissa Hamner Mageroy,
Norwegian Institute of Bioeconomy Research (NIBIO), Norway

*CORRESPONDENCE
Ana Paula Alonso
Anapaula.Alonso@unt.edu

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 29 September 2022
ACCEPTED 14 December 2022
PUBLISHED 06 January 2023

CITATION
Castro-Moretti FR, Cocuron J-C, Castillo-Gonzalez H, Escudero-Leyva E, Chaverri P, Guerreiro-Filho O, Slot JC and Alonso AP (2023) A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry. *Front. Plant Sci.* 13:1057645. doi: 10.3389/fpls.2022.1057645



# A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography mass spectrometry

Fernanda R. Castro-Moretti[1], Jean-Christophe Cocuron[2], Humberto Castillo-Gonzalez[3], Efrain Escudero-Leyva[4,5], Priscila Chaverri[3,4], Oliveiro Guerreiro-Filho[6], Jason C. Slot[7] and Ana Paula Alonso[1,2]*

[1]BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, TX, United States, [2]BioAnalytical Facility, University of North Texas, Denton, TX, United States, [3]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD, United States, [4]School of Biology and Natural Products Research Center Centro de Investigaciones en Productos Naturales (CIPRONA), University of Costa Rica, San Jose, Costa Rica, [5]Centro Nacional de Alta Technologia-Consejo Nacional de Rectores (CeNAT-CONARE), National Center for Biotechnological Innovations (CENIBiot), San Jose, Costa Rica, [6]Coffee Center Agronomic Institute, Campinas, Sao Paulo, Brazil, [7]Department of Plant Pathology, The Ohio State University, Columbus, OH, United States

**Introduction:** Products of plant secondary metabolism, such as phenolic compounds, flavonoids, alkaloids, and hormones, play an important role in plant growth, development, stress resistance. The plant family *Rubiaceae* is extremely diverse and abundant in Central America and contains several economically important genera, e.g. *Coffea* and other medicinal plants. These are known for the production of bioactive polyphenols (e.g. caffeine and quinine), which have had major impacts on human society. The overall goal of this study was to develop a high-throughput workflow to identify and quantify plant polyphenols.

**Methods:** First, a method was optimized to extract over 40 families of phytochemicals. Then, a high-throughput metabolomic platform has been developed to identify and quantify 184 polyphenols in 15 min.

**Results:** The current metabolomics study of secondary metabolites was conducted on leaves from one commercial coffee variety and two wild species that also belong to the *Rubiaceae* family. Global profiling was performed using liquid chromatography high-resolution time-of-flight mass spectrometry. Features whose abundance was significantly different between coffee species were discriminated using statistical analysis and annotated using spectral databases. The identified features were validated by commercially

available standards using our newly developed liquid chromatography tandem mass spectrometry method.

**Discussion:** Caffeine, trigonelline and theobromine were highly abundant in coffee leaves, as expected. Interestingly, wild *Rubiaceae* leaves had a higher diversity of phytochemicals in comparison to commercial coffee: defense-related molecules, such as phenylpropanoids (e.g., cinnamic acid), the terpenoid gibberellic acid, and the monolignol sinapaldehyde were found more abundantly in wild Rubiaceae leaves.

KEYWORDS

LC-MS/MS, phenolics, phytochemicals, *Rubiaceae*, secondary metabolism

## 1 Introduction

Plant secondary metabolites are byproducts of primary metabolism. They play important roles during plant development, reproduction and stress response (Patra et al., 2013; Ma et al., 2016; Böttger et al., 2018; Kessler and Kalske, 2018; Jain et al., 2019). Because plants are sessile organisms, they must endure environmental and biotic pressure. The production of phytochemicals is part of their response to these stresses. Besides their importance for plant adaptation, growth and development, plant secondary compounds are valuable resources for the food, pharmaceutical, and biofuel industries (Boudet, 2007; Korkina, 2007; Cragg and Newman, 2013; Chiocchio et al., 2021). Alkaloids, phenolics, and terpenoids are the three main families that comprise secondary metabolites produced by plants. They are synthesized through malonic acid, mevalonic acid, methylerythritol-phosphate, and shikimate pathways (Jain et al., 2019). Because of their vast structural/chemical diversity, low solubility, and small quantities in plant tissues, the recovery, identification and quantification of phytochemicals are particularly challenging.

Genetics studies in combination with metabolic profile are key for plant breeding and insertion of desired traits, such as specific polyphenols. Recovering lost attributes due to domestication using wild relatives in the breeding program is a promising strategy. However, it has been mostly applied to crops such as rice, wheat, barley and potatoes (McSorley and Phillips, 1992; Peleg et al., 2005; Feuillet et al., 2008; Spooner et al., 2014; Brar and Khush, 2018). Despite the limited number of studies on coffee leaf and other *Rubiaceae* metabolic content, the presence of phenolic compounds has been previously described (Souard et al., 2018; Cangeloni et al., 2022; Montis et al., 2022). Caffeine, chlorogenic acids, mangiferin and trigonelline are the main phytochemicals found in coffee leaves (Cangeloni et al., 2022). Indole alkaloids are the most common secondary metabolite class throughout *Rubiaceae* species, although other classes of alkaloids, terpenes and flavonoids have also been reported (Martins and Nunez, 2015). For example, akuamigine, vincoside, yohimbine and other indole alkaloids have been detected in the *Rubiaceae Uncaria* spp. (Laus and Teppner, 1996; Ndagijimana et al., 2013; Xie et al., 2013). Different species of *Gardenia* sp. produce iridoids, such as genipin and gardenoside, as well as flavonoids and triterpenes (Chen et al., 2009; Kunert et al., 2009; Yang et al., 2013; Wang et al., 2015). Many members the *Rubiaceae* family have been studied for their secondary metabolites with medicinal properties (Chen et al., 2009; Ahmad and Salim, 2015; Martins and Nunez, 2015). For instance, species that are used in traditional medicine from the genera *Borreria* and *Spermacoce* contain alkaloids, flavonoids, iridoids, and terpenoids (Conserva and Ferreira Júnior, 2012). Additionally, medicinal plants with known anti-inflammatory and antioxidant properties, such as species from the genera *Rytignia* and *Canthium multiflorum*, have bioactive compounds like tannins, saponins and flavonoids, coumarins and terpenoids (Chandra Kala, 2015). Therefore, it is important to develop efficient methodologies to monitor plant polyphenols, which will guide breeding programs and boost phytochemical discovery.

In order to fully grasp the diversity of phytochemicals present in leaves of coffee and other *Rubiaceae* species, i) a single extraction procedure allowing to recover the vast diversity phytochemical families is needed, ii) an untargeted metabolomics approach is required to detect unknown/new polyphenols, and iii) high-throughput targeted metabolomics method is necessary to quantify a maximum of secondary compounds within a single run. Developing a fast methodology that isolates most of the secondary metabolites present in leaves is challenging.

Metabolomics is the ideal technique for detecting small quantities of phytochemicals (Jorge et al., 2016; Cocuron et al., 2019; Castro-moretti et al., 2020). Nuclear magnetic resonance (NMR) and mass spectrometry (MS) are the most common analytical tools for performing plant metabolomics. MS coupled

with liquid or gas chromatography (LC and GC, respectively) is a preferred method due to its higher sensitivity and lesser amount of sample requirements (de Falco and Lanzotti, 2018; Liu et al., 2019; Perez de Souza et al., 2019). In this study, two MS instruments were used: a high-resolution quadrupole time-of-flight (HR-Q-TOF) for untargeted metabolomics, and a highly sensitive triple quadrupole for targeted quantification of known metabolites. On one hand, untargeted studies are designed to detect a broad range of molecules in a biological sample (Patti et al., 2012; Perez de Souza et al., 2019). It is common to use spectral libraries to attempt compound identification (Dunn et al., 2012; Perez de Souza et al., 2019; Jez et al., 2021). On the other hand, targeted metabolomics is used to quantify known metabolites using analytical standards (Patti et al., 2012; Roberts et al., 2012; Sawada and Yokota Hirai, 2013). Although complementary, these two approaches have been rarely combined (Montis et al., 2022).

Other approaches have attempted to isolate and quantify plant secondary metabolites; however, a limited number of families (one or two) and compounds (less than 40) were monitored (Sun et al., 2013; Orcic et al., 2014; Bataglion et al., 2015; Lin et al., 2015; Jaini et al., 2017; Cocuron et al., 2019; Gulcin et al., 2019; Marchetti et al., 2019; Quatrin et al., 2019). In this study, a single-extraction method was developed to recover 42 distinct families of phytochemicals. Untargeted and targeted metabolomics were combined to study the secondary metabolites present in coffee and wild *Rubiaceae* leaves. More specifically, a state-of-the-art targeted approach allowing the quantification of 184 phytochemicals was developed and was used to validate the identity of 74 compounds highlighted by the untargeted analysis. Combining both techniques and instruments along with an optimized extraction method resulted in a sensitive and thorough pipeline to detect, classify and quantify secondary metabolites in leaves of coffee and other *Rubiaceae* species. We anticipate that this thorough pipeline will boost the process of detection, classification and quantification of polyphenols in leaves of coffee and other *Rubiaceae*, and will be further applied to other plant organs and species.

# 2 Materials and methods

## 2.1 Chemicals

LC-MS-grade acetic acid, acetonitrile, methanol, DMSO, and water were ordered from Thermo Fisher Scientific (Hampton, NH). All non-labeled standards as well as trans-cinnamic acid-β,2,3,4,5,6-d6 were purchased from MilliporeSigma (Burlington, MA). N,N-dimethyltryptamine (N,N-DMT), bufotenin (5-OH-DMT) and psilocybin standards were obtained from Cayman Chemical (Ann Arbor, MI).

## 2.2 Preparation of standard stocks and working solutions

Stock solutions of dihydrokaempferol, dihydroquercetin, mitragynine, luteolin, luteolin-7-O-glucoside, naringenin, phloretin, piceid, prunetin, pterostilbene, orientin, quercetin, quercitrin, reserpine, rhamnazin, rhamnetin, schaftoside, spiraeoside, swertiajaponin, swertisin, tectochrysin, tricin, vicenin 2 and 3, vincosamide and yohimbine were prepared at 1,000 µM. Acacetin-7-O-rutinoside, afzelin, apigenin-7-glucuronide, calycosin, corynanthine, harmane, hordenine, ipriflavone, tomatidine, xanthohumol, sophoricoside, rauwolscine, idaein, keracyanin and neobavaisoflavone were prepared at 100 µM. All other stock solutions were prepared at 10,000 µM, using methanol or DMSO as solvents. Working solutions were prepared to final concentrations of 100, 50, 10 and 1 µM in methanol/water (40:60; v/v).

## 2.3 Leaf collection

Untargeted and targeted metabolomic analyses were performed using commercial coffee leaves (*Coffea arabica* cv. Obatã IAC 1669-20 - CC) that were collected in San José, Costa Rica, at the Coopetarrazú plantation. Three mature leaves were harvested from four trees and kept on ice during transportation to the laboratory, where they were flash-frozen in liquid nitrogen, and lyophilized until dryness. Wild *Rubiaceae* leaves were collected from the two species *Isertia hankeana* and *Simira maxonii* (WR1 and WR2, respectively) in a private rainforest of the Golfito (Puntarenas) region of Costa Rica. Three leaves were collected from each tree, one tree per species and kept in ice during their transportation to the laboratory. Then they were flash-frozen in liquid nitrogen and lyophilized (Labconco Freezone, South Kansas City, KS) until dryness. To optimize the extraction and LC-MS/MS method, mature leaves from the wild coffee species *Coffea liberica* var. *dewevrei* and *Coffea salvatrix* were used from plants grown in field conditions, collected at the Agronomic Institute in Campinas (São Paulo, Brazil). Four leaves were collected from each tree, two trees per species, kept in dry ice during harvest and transportation to the laboratory and then they were frozen in liquid nitrogen prior lyophilization until dryness. All collected leaves were lyophilized using a freeze dryer Labconco Freezone 12 plus (South Kansas City, KS). Dried leaves were ground into a fine powder using 15 mL plastic jars with four 20 mm metal beads in a tissue homogenizer Geno/Grinder 2010 from Spex (Metuchen, NJ) for two rounds of 30 sec at 1,750 rpm. Wild *Rubiaceae* leaves were collected from the same tree, grown in very different environmental conditions in comparison to CC; to mitigate variations due to leaf maturity level, light exposure, etc., and to focus on polyphenol differences across species, the ground

powder was pooled, homogenized and divided into five pseudoreplicates per species.

## 2.4 Intracellular secondary metabolite extraction

The extraction of leaf secondary metabolites was performed after grinding and weighting 10 mg of powdered leaf material. Metabolites were extracted by adding 10 µL of 1 mM trans-cinnamic acid-β,2,3,4,5,6-d6 and 490 µL of 100% methanol followed by grinding with one 5 mm metal bead at 30 Hz for 5 min using a mixer mill MM400 from Retsch (Haan, Germany). Then, the extracts were sonicated at 35-40°C for 20 min and centrifuged at 9,600 *g* for 5 min at room temperature. The supernatants were transferred to 1.5 mL microcentrifuge tubes. The remaining pellets were resuspended in 500 µL of methanol/water (30:70, v/v), sonicated for 20 min at 35-40 °C, and spun down under the same conditions as mentioned before. The supernatants were combined to the first ones, and then, 500 µL of extracts were filtered through 3 kDa Amicon filtering devices (MilliporeSigma, Burlington, MA) at 14,000 *g* for 60 min at room temperature. The resulting eluates were stored at -20°C until LC-MS/MS analysis.

## 2.5 Untargeted metabolomics

The analysis of the metabolites was carried out using an Exion ultra high-performance liquid chromatography system coupled with a high-resolution mass spectrometer TripleTOF6600+ from AB Sciex (Framingham, MA).

### 2.5.1 HPLC conditions

The compounds were separated using a C18 Symmetry column (75 x 4.6 mm, 3.5 µm) with a Symmetry C18 pre-column (20 x 3.9 mm; 5µm) from Waters (Milford, MA) as previously described (Cocuron et al., 2019). The temperatures of the column compartment and the autosampler were kept at 30 °C and 15 °C, respectively. The analytes were eluted using a gradient of 0.1% (v/v) acetic acid in acetonitrile (Solvent A) and 0.1% (v/v) acetic acid in water (Solvent B) under a flow rate of 0.8 mL/min. The following gradient was applied: 0-1.0 min, 98% B; 1.0-16.0 min, 98-42% B; 16.0-21.0 min, 42-20% B; 21.0-26.0 min, 20-10% B; 26.0-28.0 min, 10% B; 28.0-28.1, 10-98% B; 28.1-30.0 min, 98% B.

### 2.5.2 High-resolution discovery using triple TOF

#### 2.5.2.1 Data-dependent acquisition

The mass spectrometer was set to scan metabolites from m/z 100-1500 amu in negative or positive mode. For the negative polarity the ion spray voltage was 4,500 V, the accumulation time was 100 msec, the declustering potential and collision energy were 60 V and 10 V, respectively. MS/MS spectra were acquired over m/z 30-1500 amu with an accumulation time of 25 msec. Parameters such as declustering potential, collision energy, collision energy spread were set to 60 V, 45 V and 15 V, respectively. The parameters for the positive mode were very similar to the ones for the negative mode except for the ion spray voltage, and the declustering potential that were 5,000 V and 35 V, respectively. The total cycling time was 0.65 sec.

The parameters for the electrospray source ionization such as curtain gas (nitrogen), nebulizing gas, heating gas, and the temperature of the source were fixed at 40 psi, 70 psi, 70 psi, and 650 °C, respectively. The source conditions were the same for the negative and positive polarities. An atmospheric-pressure chemical ionization (APCI) negative or positive calibration solution was delivered by a calibrant delivery system every 5 samples to correct for any mass drift that may occur during the run. MS spectra were acquired using Analyst TF 1.8.1 software (AB Sciex, Framingham, MA). It is important to note that the data-dependent acquisition (DDA) was run in negative and positive modes, on a mixture composed of CC, WR1 and WR2 extracts. The precursor ions present in this mixture were then used for the sequential window acquisition of all theoretical mass spectra (SWATH-MS) scan survey.

#### 2.5.2.2 Data-independent acquisition using SWATH-MS

The precursor ion data obtained from the DDA (negative and positive ionizations) were used to generate short overlapping precursor ion windows which are the core of the SWATH-MS mode. Briefly, all precursors ions from the DDA mode (negative or positive polarity) as well as their intensities were exported to an excel file in which a total of 20 variable SWATH-MS windows were created with one amu overlapping mass. These SWATH-MS windows for the negative or positive polarity were saved as a ".txt" file, and uploaded to Analyst to build the mass spectrometry part of the LC-HR-MS/MS acquisition method. It is important to note that the source, MS scan and MS/MS scan parameters for the SWATH-MS mode were the same than the ones used for the DDA scan survey.

For the sequence injection, the total 15 biological samples were placed randomly in the autosampler. Two quality controls (QC) consisting of an equal mixture of each leaf extract, and two blanks containing the internal standard, trans-cinnamic acid-β,2,3,4,5,6-d6 in methanol: water (40:60, v/v), were included; one was injected at the beginning of the sequence, and the other at the end.

### 2.5.3 Data processing

A non-targeted screening MQ4 workflow was designed using the software SciexOS v.1.6.1 (Sciex, Framingham, MA) with the Smart Confirmation Search algorithm. Results were

sorted by purity, 0.02 Da as precursor mass tolerance, and 0.4 Da as fragment mass tolerance. The intensity threshold was set to 0.05 and minimal purity to 10%. All extract samples, blanks with internal standard, and QCs were considered for constructing the processing method. Quality of the processed data was assessed as follows: i) the difference in the area of the internal standard between the sample extracts and the blanks was less than 5%; and ii) a Principal Component Analysis (PCA) including the QCs was performed for the metabolites monitored in negative and positive modes and showed clustering of the two QCs (Supplemental Figure 1), which is indicative of a low variance. Then, data obtained from untargeted metabolomics were analyzed using the software XCMS (Tautenhahn et al., 2012) with the following parameters: method UPL/UHD Q-TOF matchedFilter; ppm error of 15; minimum peak width of 5; maximum peak width of 20; signal/noise threshold of 6; mzdiff of 0.01; integration method 1; prefilter peaks 3; noise filter 0; and retention time correction method obiwarp. Once peak picking, alignment and integration was performed, a table with mass to charge ratio, retention time, and normalized intensity of each feature (by the intensity of the internal standard, trans-cinnamic acid-β,2,3,4,5,6-d6) was generated. This table was then used for statistical analyses (see 2.7. Statistical Analyses). Further data curation was performed: features with signal intensity lower than 1,000,000.00 count per second were excluded, as well as the features not present on all replicates. Then, the feature with the highest intensity was selected for each peak group. Features from positive and negative ionization were merged and, when the same feature was present in both modes, the one with the highest intensity was selected. Metabolite identification was performed using the software SciexOS v.1.6.1 (Sciex, Framingham, MA) with spectral libraries from the National Institute of Standards and Technology (NIST, Gaithersburg, MA) and a homemade library.

## 2.6 Targeted metabolomics using LC-MS/MS scheduled multiple reaction monitoring

### 2.6.1 LC-MS/MS conditions

The detection and quantification of phytochemicals was performed as previously described by (Cocuron et al., 2019) with some slight modifications concerning the LC gradient and the use of scheduled multiple reaction monitoring (sMRM).

The 184 phytochemicals considered in this study were optimized one by one by direct infusion after having been diluted to 1 µM with acetonitrile/water solution (50:50; v/v) containing 0.1% of acetic acid as an additive. The flow for the direct infusion was set to 10 µL/min, and parameters such as declustering potential (DP), collision energy (CE), cell exit potential (CXP) were determined for the five most abundant product ions derived from each precursor ion (see Table 1). The compound optimization was done automatically for the negative and positive polarities using Analyst 1.7 software (AB Sciex, Framingham, MA). The source optimization for the electrospray ionization was conducted using different values for the curtain gas (25, 30, 35, 40 V), the nebulizer gas (GS1; 40, 45, 50, 55, 60, 65 V), the heating gas (GS2; 40, 45, 50, 55, 60, 65 V), the collision activated dissociation (CAD; low, medium, high), the temperature (300, 350, 400, 450, 500, 550, 600, 650°C), and the ionspray voltage (IS; 3000, 3500, 4000, 4500, 5000 V).

The compounds were detected and quantify using an Agilent 1290 Infinity II liquid chromatography system coupled to a hybrid Triple Quadrupole 6500+ from ABSciex (Framingham, MA). The extracts were kept at 10°C in an auto-sampler, and the phytochemicals were separated at 30°C using a reverse phase C18 Symmetry column (4.6 x 75 mm; 3.5 µm) coupled to a Symmetry C18 pre-column (3.9 x 20 mm; 5 µm) from Waters (Milford, MA). The liquid chromatography gradient was made of 0.1% (v/v) acetic acid in acetonitrile (A) and 0.1% (v/v) acetic acid in water (B). The total LC-MS/MS run was 15 min with a flow rate of 800 µL/min. The following gradient was applied to resolve the polyphenols: B= 0-1.0 minute 98%, 1.0-7.0 min 42%, 7.0-9.0 min 20%, 9.0-11.0 min 10%, 11.0-13.0 min 10%, 13.0-13.1 min 98%, and 13.1-15.0 min 98%. The injection needle was rinsed with 50% aqueous methanol. Five µL of external standard mixtures and 5 µL of biological sample were injected onto the column.

Electrospray ionization with polarity switch was applied to the extracts to acquire mass spectra of the different analytes. The settling time between each polarity was 15 msec. Phytochemicals were simultaneously detected as precursor ion/product ion pair using multiple reaction monitoring (MRM) at first to record the retention time for each of the polyphenols considered in this study (see Table 1). The retention times for the compounds were reported in the LC-MS/MS method to create scheduled MRM with MRM detection windows of 60 sec. The cycling time was set to 1.1 sec and the dwell time varied depending on the number of MRMs triggered at a specific point of time during the LC-MS/MS acquisition. The dwell time ranged from 3 to 250 msec. The source parameters for both modes were identical, and they were as followed: 4,500 V for the ionspray voltage, 40 V for the curtain gas, 550°C for the temperature, 50 psi for the nebulizer gas (GS1), 60 psi for the heating gas (GS2), and "Medium" for the collision activated dissociation (CAD).

### 2.6.2 Data acquisition and processing

Analyst 1.7 software (AB Sciex, Framingham, MA) was used to acquire the data whereas MultiQuant v3.0.3 (AB Sciex, Framingham, MA) was used to integrate the peaks corresponding to the different phytochemicals. Metabolite quantification was performed as previously explained (Arias et al., 2022). Briefly, the total amount of each analyte was calculated using the trans-cinnamic acid-β,2,3,4,5,6-d6 internal standard area, and the known concentration of its corresponding external standard run in parallel to the samples.

TABLE 1 Compound-dependent parameters for scheduled MRM scan survey, per metabolite and their chemical class (family): retention time (RT) in minutes, precursor mass (Q1) and product mass (Q3), declustering potential (DP), collision energy (CE) and collision cell exit potential (CXP) in volts.

| Family | Metabolite | RT | Q1 | Q3 | DP | CE | CXP |
|---|---|---|---|---|---|---|---|
| alkaloid | Corynanthine | 4.1 | 355.3 | 144.0 | 113 | 39 | 16 |
| alkaloid | Dihydrocinchonine | 4.2 | 297.0 | 279.0 | 40 | 31 | 14 |
| alkaloid | Harmane | 3.6 | 183.0 | 115.0 | 50 | 45 | 12 |
| alkaloid | Hordenine | 1.1 | 166.0 | 120.9 | 24 | 21 | 14 |
| alkaloid | Mitragynine | 4.9 | 399.0 | 174.0 | 75 | 41 | 20 |
| alkaloid | Seneciphylline | 3.4 | 334.0 | 120.0 | 105 | 35 | 14 |
| alkaloid | Tomatidine | 5.6 | 416.0 | 161.0 | 91 | 49 | 18 |
| amino acid derivative | 5-Hydroxy-tryptophan | 3.1 | 221.0 | 204.0 | 20 | 13 | 12 |
| amino acid derivative | Tyramine | 0.9 | 138.0 | 120.9 | 9 | 13 | 14 |
| anthocyanidin | Apigeninidin | 4.4 | 255.0 | 170.9 | 127 | 43 | 18 |
| anthocyanin | Idaein | 3.7 | 449.1 | 287.0 | 60 | 29 | 14 |
| anthocyanin glycoside | Keracyanin | 3.6 | 595.2 | 286.9 | 75 | 39 | 14 |
| benzodioxol | Piperonyloyl | 6.4 | 184.8 | 142.8 | 50 | 13 | 16 |
| chalcone | Xanthohumol | 9.6 | 355.0 | 299.0 | 27 | 15 | 16 |
| cinnamaldehyde | p-Coumaraldehyde | 5.9 | 147.1 | 119.0 | -50 | -24 | -13 |
| cinnamate ester | 3,4-Di-O-caffeoylquinic acid | 6.6 | 515.0 | 353.0 | -80 | -28 | -19 |
| cinnamate ester | 3-Caffeoylquinic acid | 4.7 | 353.0 | 134.0 | -25 | -62 | -15 |
| cinnamate ester | 4,5-Di-O-caffeoylquinic acid | 6.7 | 515.0 | 353.0 | -80 | -28 | -19 |
| cinnamate ester | 4-Caffeoylquinic acid | 5.3 | 353.0 | 173.0 | -45 | -20 | -11 |
| cinnamate ester | 5-Caffeoylquinic acid | 4.8 | 353.0 | 93.0 | -25 | -56 | -11 |
| coumarin | 6-Methylcoumarin | 7.7 | 161.0 | 105.0 | 80 | 29 | 12 |
| coumarin | 7,8-Dihydroxy-4-methylcoumarin | 5.4 | 191.0 | 119.0 | -65 | -26 | -11 |
| coumarin | Scopoletin | 5.5 | 192.9 | 133.0 | 100 | 29 | 14 |
| coumarin derivative | Esculetin | 4.7 | 178.9 | 123.0 | 80 | 31 | 14 |
| coumarin derivative | Mellein | 8.2 | 179.0 | 160.8 | 11 | 17 | 18 |
| cyclic ketone | Isophorone | 7.5 | 138.9 | 68.9 | 184 | 21 | 8 |
| cyclohexenecarboxylic acid | p-Coumaroyl-shikimate | 5.3 | 321.1 | 147.0 | 115 | 15 | 10 |
| dihydrochalcone | Phloretin | 6.9 | 274.9 | 106.9 | 50 | 21 | 12 |
| diterpenoid | Gibberellic acid | 5.4 | 345.0 | 239.0 | -70 | -20 | -13 |
| diterpenoid | Ginkgolide A | 6.6 | 409.1 | 345.0 | 80 | 27 | 16 |
| flavanol | Catechin | 4.3 | 291.0 | 139.0 | 50 | 20 | 16 |
| flavanol | Epicatechin | 4.6 | 291.0 | 139.0 | 50 | 20 | 16 |
| flavanol | Epigallocatechin | 4.1 | 307.0 | 138.9 | 13 | 19 | 16 |
| flavanol | Gallocatechin | 3.7 | 307.1 | 138.9 | 19 | 19 | 16 |
| flavanone | Eriodictyol | 6.5 | 287.0 | 151.0 | -30 | -20 | -7 |
| flavanone | Hesperetin/Homoeriodictyol | 7.3 | 303.0 | 177.0 | 85 | 25 | 10 |

*(Continued)*

TABLE 1 Continued

| Family | Metabolite | RT | Q1 | Q3 | DP | CE | CXP |
|---|---|---|---|---|---|---|---|
| flavanone | Isosakuranetin | 8.4 | 287.0 | 153.0 | 80 | 29 | 18 |
| flavanone | Naringenin | 7.1 | 272.9 | 153.0 | 50 | 31 | 16 |
| flavanone | Sakuranetin | 8.4 | 287.0 | 167.0 | 90 | 29 | 20 |
| flavanone-C-glycoside | Swertiajaponin | 4.6 | 463.1 | 445.1 | 60 | 17 | 22 |
| flavanone O-glycoside | Naringin | 5.3 | 579.1 | 271.0 | -150 | -44 | -13 |
| flavanone O-glycoside | Naringenin-7-O-glucoside | 5.5 | 435.1 | 273.0 | 70 | 19 | 14 |
| flavanonol | Dihydrokaempferol | 6.1 | 287.0 | 125.0 | -60 | -28 | -5 |
| flavanonol | Dihydroquercetin | 5.5 | 303.0 | 285.0 | -50 | -16 | -13 |
| flavone | 3-Deoxyrobinetin | 5.1 | 284.9 | 149.0 | -110 | -36 | -9 |
| flavone | Apigenin | 7.0 | 269.0 | 117.0 | -90 | -42 | -13 |
| flavone | Apigenin-7-glucuronide | 6.8 | 447.1 | 271.1 | 120 | 29 | 14 |
| flavone | Baicalein | 7.3 | 270.9 | 122.9 | 150 | 43 | 14 |
| flavone | Chrysin | 8.3 | 253.0 | 143.0 | -110 | -36 | -9 |
| flavone | Chrysoeriol | 7.0 | 301.0 | 286.0 | 80 | 37 | 32 |
| flavone | Diosmetin | 7.1 | 300.9 | 286.0 | 80 | 35 | 32 |
| flavone | Eupatorin | 7.9 | 344.9 | 283.9 | 100 | 41 | 30 |
| flavone | Flavopiridol | 4.8 | 402.0 | 341.0 | 75 | 33 | 16 |
| flavone | Genkwanin | 8.5 | 284.8 | 242.0 | 80 | 43 | 26 |
| flavone | Maysin | 5.5 | 577.0 | 431.0 | 50 | 19 | 24 |
| flavone | Myricetin | 5.8 | 317.0 | 151.0 | -70 | -32 | -15 |
| flavone | Scutellarein | 6.0 | 286.9 | 123.0 | 110 | 45 | 14 |
| flavone | Tectochrysin | 10.1 | 268.9 | 226.0 | 80 | 43 | 26 |
| flavone | Tricin | 7.0 | 331.0 | 315.0 | 100 | 41 | 34 |
| flavone C-glycoside | Isoorientin | 4.6 | 449.0 | 299.0 | 50 | 39 | 14 |
| flavone C-glycoside | Swertisin | 5.0 | 447.2 | 297.0 | 60 | 35 | 14 |
| flavone-C-glycoside | Isoschaftoside | 4.5 | 565.1 | 427.1 | 100 | 29 | 20 |
| flavone-C-glycoside | Isovitexin/Vitexin | 4.9 | 433.0 | 283.0 | 80 | 35 | 14 |
| flavone-C-glycoside | Orientin | 4.7 | 449.0 | 329.0 | 120 | 39 | 16 |
| flavone-C-glycoside | Rhamnosylisoorientin | 4.4 | 593.0 | 298.0 | -150 | -58 | -13 |
| flavone-C-glycoside | Vicenin 2 | 4.2 | 595.1 | 577.1 | 100 | 21 | 28 |
| flavone-C-glycoside | Vicenin 3/Schaftoside | 4.5 | 565.1 | 547.2 | 80 | 19 | 26 |
| flavone-O-glycoside | Acacetin-7-O-rutinoside | 5.8 | 593.1 | 447.1 | 60 | 25 | 22 |
| flavone-O-glycoside | Apigenin-7-O-glucoside | 5.3 | 432.9 | 271.0 | 60 | 25 | 14 |
| flavone-O-glycoside | Benzoic acid | 7.4 | 120.9 | 77.0 | -20 | -16 | -9 |
| flavone-O-glycoside | Luteolin-7-O-glucuronide | 6.5 | 463.0 | 287.0 | 130 | 29 | 14 |
| flavone-O-glycoside | Myricetin-3-O-Rhamnoside | 5.0 | 465.0 | 319.0 | 30 | 15 | 16 |
| flavone-O-glycoside | Neodiosmin | 5.3 | 607.0 | 299.0 | -150 | -40 | -15 |

*(Continued)*

TABLE 1 Continued

| Family | Metabolite | RT | Q1 | Q3 | DP | CE | CXP |
|---|---|---|---|---|---|---|---|
| flavonol | Fisetin | 5.9 | 286.9 | 137.0 | 120 | 43 | 16 |
| flavonol | Gossypetin | 5.8 | 319.0 | 169.1 | 160 | 43 | 18 |
| flavonol | Isorhamnetin | 7.2 | 315.0 | 300.0 | -80 | -28 | -15 |
| flavonol | Kaempferide | 8.5 | 300.9 | 229.0 | 120 | 53 | 24 |
| flavonol | Kaempferol | 7.1 | 286.9 | 153.0 | 120 | 43 | 18 |
| flavonol | Luteolin | 6.4 | 286.9 | 153.0 | 120 | 43 | 18 |
| flavonol | Morin/Tricetin | 6.3 | 302.9 | 152.9 | 140 | 39 | 18 |
| flavonol | Quercetin | 6.5 | 301.0 | 151.0 | -80 | -28 | -15 |
| flavonol | Quercitrin | 5.4 | 449.0 | 303.0 | 30 | 15 | 16 |
| flavonol | Rhamnazin | 8.6 | 331.0 | 316.0 | 140 | 35 | 36 |
| flavonol | Rhamnetin | 7.8 | 315.0 | 165.0 | -60 | -28 | -9 |
| flavonol | Tamarixetin | 7.2 | 316.9 | 302.0 | 120 | 35 | 14 |
| flavonol-O-glucuronide | Miquelianin | 6.7 | 479.0 | 303.0 | 60 | 21 | 16 |
| flavonol-O-glycoside | Isoquercetin | 5.1 | 465.1 | 303.0 | 30 | 17 | 16 |
| flavonol-O-glycoside | Isorhamnetin-3-O-glucoside | 5.3 | 479.1 | 317.0 | 30 | 17 | 16 |
| flavonol-O-glycoside | Kaempferitrin | 4.8 | 579.1 | 433.1 | 50 | 17 | 20 |
| flavonol-O-glycoside | Kaempferol-3-O-glucoside | 5.4 | 449.0 | 287.0 | 40 | 17 | 16 |
| flavonol-O-glycoside | Kaempferol-3-O-glucuronide | 6.4 | 463.0 | 287.0 | 50 | 21 | 14 |
| flavonol-O-glycoside | Kaempferol-3-O-rutinoside | 5.1 | 595.1 | 287.1 | 50 | 27 | 14 |
| flavonol-O-glycoside | Kaempferol-7-O-Neohesperidoside | 5.1 | 595.1 | 287.0 | 120 | 31 | 14 |
| flavonol-O-glycoside | Luteolin-7,3'-Di-O-glucoside/Kaempferol-3-O-sophoroside | 4.6 | 611.1 | 449.1 | 150 | 31 | 22 |
| flavonol-O-glycoside | Luteolin-7-O-glucoside/Kaempferol-7-O-glucoside | 5.4 | 449.0 | 287.0 | 91 | 27 | 14 |
| flavonol-O-glycoside | Quercetin-3,4'-O-diglucoside | 4.5 | 627.0 | 465.0 | 60 | 19 | 24 |
| flavonol-O-glycoside | Quercetin-3-O-galactoside | 5.0 | 465.0 | 303.0 | 40 | 17 | 16 |
| flavonol-O-glycoside | Rutin | 4.8 | 611.2 | 303.0 | 30 | 29 | 16 |
| flavonol-O-glycoside | Spiraeoside | 5.5 | 465.0 | 303.0 | 120 | 29 | 16 |
| flavonol-glycoside | Afzelin | 5.7 | 431.1 | 285.0 | -100 | -28 | -15 |
| glycosylated hydroquinone | Arbutin | 2.9 | 271.0 | 161.0 | -60 | -10 | -9 |
| hydroxycinnamic acid | Cynarin | 5.4 | 515.0 | 190.9 | -54 | -40 | -11 |
| hydroxycinnamic acid | Ferulic acid | 5.5 | 192.9 | 134.0 | -25 | -20 | -9 |
| hydroxycinnamic acid | p-coumaric acid | 5.4 | 162.9 | 119.0 | -25 | -18 | -15 |
| hydroxycinnamyl alcohol | p-Coumaryl alcohol | 5.1 | 149.1 | 131.0 | -25 | -14 | -13 |
| hydroxy monocarboxylic acid | Caffeoyl-shikimate | 4.8 | 337.0 | 163.0 | 120 | 15 | 10 |
| indole alkaloid | 5-Hydroxydimethyltryptamine | 1.2 | 205.0 | 160.0 | 40 | 19 | 10 |
| indole alkaloid | N,N-dimethyltryptamine | 3.3 | 189.0 | 144.0 | 25 | 23 | 16 |
| indole alkaloid | Psilocybin | 3.1 | 285.0 | 205.0 | 50 | 23 | 10 |
| indole alkaloid | Rauwolscine | 5.4 | 355.0 | 144.0 | 100 | 39 | 14 |

*(Continued)*

TABLE 1 Continued

| Family | Metabolite | RT | Q1 | Q3 | DP | CE | CXP |
|---|---|---|---|---|---|---|---|
| indole alkaloid | Reserpine | 5.5 | 609.0 | 195.0 | 50 | 47 | 10 |
| indole alkaloid | Strychnine | 3.5 | 335.0 | 184.0 | 120 | 49 | 20 |
| indole alkaloid | Theobromine | 3.5 | 181.0 | 138.0 | 85 | 23 | 14 |
| indole alkaloid | Theophylline | 3.9 | 181.0 | 124.0 | 60 | 25 | 14 |
| indole alkaloid | Tryptamine | 3.1 | 161.0 | 144.0 | 20 | 13 | 10 |
| indole alkaloid | Vincosamide | 6.2 | 499.0 | 337.0 | 60 | 23 | 18 |
| indole alkaloid | Yohimbine | 4.0 | 355.0 | 144.0 | 60 | 37 | 16 |
| indole alkaloid | Paynantheine | 4.9 | 397.0 | 174.0 | 50 | 37 | 20 |
| isoflavone | Calycosin | 6.5 | 285.0 | 269.9 | 77 | 31 | 28 |
| isoflavone | Glycitein | 6.3 | 284.9 | 270.1 | 100 | 35 | 30 |
| isoflavone | Ipriflavone | 10.2 | 280.8 | 239.0 | 114 | 27 | 12 |
| isoflavone | Neobavaisoflavone | 8.4 | 323.0 | 266.9 | 75 | 25 | 14 |
| isoflavone | Prunetin | 8.6 | 284.9 | 241.9 | 150 | 43 | 26 |
| isoflavone-glycosylated | Sophoricoside | 5.4 | 433.0 | 271.0 | 80 | 17 | 14 |
| isoflavone-O-glycoside | Genistein-7-O-glucuronide/Baicalin | 6.9 | 447.0 | 271.0 | 120 | 27 | 14 |
| isoflavone-O-glycoside | Glycitin | 4.7 | 447.0 | 224.8 | 42 | 59 | 24 |
| lactone | Caffeic acid | 6.9 | 178.9 | 135.0 | -35 | -20 | -15 |
| lignan | Arctigenin | 7.7 | 371.0 | 83.0 | -71 | -24 | -15 |
| lignan | Matairesinol | 7.1 | 357.2 | 82.9 | -76 | -26 | -9 |
| methylxanthine alkaloid | Caffeine | 4.3 | 195.0 | 138.0 | 150 | 27 | 14 |
| monocarboxilic acid | Cinnamic acid | 7.1 | 148.8 | 103.0 | 20 | 25 | 12 |
| monohydroxybenzoic acid | 4-Hydroxybenzoic acid | 4.6 | 136.9 | 93.0 | -15 | -20 | -13 |
| monohydroxybenzoic acid | Salicylic acid | 9.1 | 136.9 | 93.0 | -15 | -20 | -13 |
| monohydroxybenzoic acid | Vanillic acid | 4.8 | 166.9 | 108.0 | -50 | -30 | -47 |
| mycotoxin | Neosolaniol | 5.0 | 383.2 | 365.1 | 117 | 13 | 18 |
| mycotoxin | Roridin-L2 | 7.1 | 531.3 | 249.1 | 29 | 21 | 14 |
| octadecanoid | Oxo-phytodienoic acid | 9.6 | 293.0 | 275.0 | 35 | 15 | 16 |
| O-methylated isoflavone | Brefeldin A | 8.6 | 281.0 | 245.0 | 13 | 9 | 14 |
| oxopurine alkaloid | 1,3,7-Trimethyluric acid | 4.0 | 209.0 | 194.0 | -40 | -18 | -9 |
| oxylipin | Jasmonic acid | 7.3 | 209.0 | 59.0 | -80 | -16 | -27 |
| oxylipin | Methyl jasmonate | 8.9 | 225.0 | 151.0 | 20 | 17 | 10 |
| phenol | Gingerol | 8.4 | 292.8 | 99.1 | -60 | -16 | -11 |
| phenolic | Rosmarinic acid | 6.6 | 361.0 | 163.0 | 12 | 11 | 10 |
| phenolic acid | Sinapic acid | 5.4 | 225.0 | 175.0 | 20 | 19 | 10 |
| phenolic acid | Syringic acid | 4.8 | 199.0 | 155.0 | 20 | 13 | 10 |
| phenolic alcohol | Caffeyl alcohol | 4.5 | 165.0 | 147.0 | -20 | -16 | -7 |
| phenolic alcohol | Coniferyl alcohol | 5.2 | 163.0 | 131.0 | 45 | 13 | 12 |

*(Continued)*

10.3389/fpls.2022.1057645

TABLE 1 Continued

| Family | Metabolite | RT | Q1 | Q3 | DP | CE | CXP |
|---|---|---|---|---|---|---|---|
| phenolic alcohol | Sinapyl-alcohol | 5.1 | 209.0 | 194.0 | -40 | -18 | -9 |
| phenolic alcohol | Vanillyl alcohol | 4.2 | 137.0 | 122.0 | 20 | 23 | 14 |
| phenolic aldehyde | 2,5-Dihydroxybenzoic acid | 6.8 | 153.0 | 108.0 | -15 | -28 | -13 |
| phenolic aldehyde | 3,4-Dihydroxybenzaldehyde | 4.6 | 139.0 | 93.0 | 50 | 19 | 10 |
| phenolic aldehyde | 3,4-Dihydroxybenzoic acid | 6.2 | 153.0 | 108.0 | -15 | -28 | -13 |
| phenolic aldehyde | 3,4-Dimethoxycinnamic acid | 4.1 | 209.1 | 163.0 | 21 | 27 | 16 |
| phenolic aldehyde | 3,5-Dihydroxybenzaldehyde | 6.2 | 139.0 | 111.0 | 50 | 15 | 12 |
| phenolic aldehyde | 3,5-Dihydroxybenzoic acid | 4.1 | 153.0 | 109.0 | -15 | -18 | -13 |
| phenolic aldehyde | 3,5-Dimethoxybenzaldehyde | 7.9 | 167.0 | 139.0 | 20 | 15 | 14 |
| phenolic aldehyde | 3,5-Dimethoxybenzoic acid | 6.9 | 183.0 | 124.0 | 25 | 21 | 14 |
| phenolic aldehyde | 4-Hydroxybenzaldehyde | 5.3 | 121.0 | 92.0 | -20 | -28 | -11 |
| phenolic aldehyde | 5-Hydroxyconiferaldehyde | 5.2 | 193.0 | 178.0 | -30 | -20 | -9 |
| phenolic aldehyde | 5-Hydroxyconiferyl-alcohol | 5.5 | 195.0 | 180.0 | -135 | -18 | -9 |
| phenolic aldehyde | Caffeyl aldehyde | 5.3 | 163.0 | 135.0 | -40 | -24 | -13 |
| phenolic aldehyde | Coniferaldehyde | 6.1 | 179.0 | 147.0 | 30 | 17 | 10 |
| phenolic aldehyde | Sinapaldehyde | 5.1 | 209.0 | 177.0 | 30 | 15 | 10 |
| phenolic aldehyde | Vanillin | 5.5 | 153.0 | 92.9 | 50 | 19 | 10 |
| phenylpropanoid | 5-Hydroxy-ferulic acid | 4.8 | 209.0 | 150.0 | -40 | -24 | -15 |
| phenylpropanoid | Biochanin A | 6.3 | 283.0 | 267.9 | -120 | -30 | -13 |
| polyketide-derived mycotoxin | Citrinin | 12 | 251.0 | 233.0 | 101 | 23 | 12 |
| polyphenol | Ellagic acid | 5.1 | 301.1 | 284.0 | -150 | -40 | -13 |
| pryridine alkaloid | Trigonelline | 1.0 | 138.0 | 92.0 | 170 | 29 | 10 |
| pyridoisoquinoline | Emetine | 5.5 | 481.0 | 246.0 | 180 | 47 | 12 |
| pyrrolizine alkaloid | Erucifoline | 2.9 | 350.2 | 120.0 | 125 | 37 | 14 |
| sesquiterpene | Abscisic acid | 6.4 | 263.0 | 153.0 | -40 | -16 | -7 |
| sesquiterpene | Alpha-Cyperone | 11.1 | 219.0 | 111.0 | 109 | 29 | 12 |
| sesquiterpene | Artemisinin | 9.4 | 283.0 | 265.1 | 13 | 11 | 12 |
| sesquiterpene lactone | Heptelidic Acid | 7.6 | 279.0 | 205.0 | -33 | -12 | -11 |
| stilbenoid | Piceid | 5.1 | 389.0 | 227.0 | -90 | -20 | -11 |
| stilbenoid | Pinosylvin | 7.9 | 213.0 | 135.0 | 50 | 19 | 14 |
| stilbenoid | Pterostilbene | 8.9 | 255.0 | 240.0 | -80 | -26 | -13 |
| stilbenoid | Resveratrol | 6.3 | 229.0 | 135.0 | 50 | 19 | 14 |
| stilbenoid | t-trimethoxyresveratrol | 10.4 | 271.0 | 152.0 | 50 | 73 | 16 |
| stilbenol | 3-Hydroxystilbene | 9.2 | 197.0 | 119.0 | 45 | 17 | 14 |
| terpenoid indole alkaloid | 7-Hydroxymitragynine | 4.0 | 415.0 | 190.0 | 60 | 37 | 10 |
| trihydroxybenzoic acid | Gallic acid | 3.3 | 169.0 | 125.0 | -30 | -20 | -13 |
| triterpenoid | Enoxolone | 11.0 | 471.0 | 189.0 | 230 | 45 | 10 |

### 2.6.3 Standard curves: Limit of detection, limit of quantification, and linearity range

To determine the limits of detection (LOD) and quantification (LOQ), and the linearity range, standard curves were generated for each metabolite as previously described (Cocuron et al., 2014; Cocuron et al., 2019) with at least six points. Each standard curve was performed in five replicates.

### 2.6.4 Recovery efficiency, matrix effect, and accuracy intra- and inter- assay

Recovery efficiency (RE) and accuracy intra- and inter- assay were determined using five coffee leaf pseudoreplicates as previously described (Cocuron et al., 2019). To assess the matrix effect (ME), five coffee leaf pseudoreplicates were used according to the procedure previously published (Cocuron et al., 2017) and the following equation:

$$ME = \left( \frac{Analyte\ peak\ area_{Sample\ spiked\ after\ extraction} - Analyte\ peak\ area_{Sample}}{Average\ analyte\ peak\ area_{External\ standard}} \times 100\,\% \right) - 100\,\%$$

In these conditions, a negative value indicates an ion suppression whereas a positive value depicts an ion enhancement due to ME.

## 2.7 Statistical analyses

Principal component analysis (PCA), partial-least square discriminant analysis (PLS-DA), heatmap and ANOVA-simultaneous component analysis (ASCA), for both untargeted and targeted analyses, were performed after log-transformation and auto-scaling, using MetaboAnalyst 5.0 (Chong et al., 2018).

# 3 Results

## 3.1 Untargeted metabolomics

First, the extraction procedure was optimized using a standard mixture containing alkaloids, cinnamate esters, and flavonoids. For that purpose, different solvents/additive and multiple sonication time/temperature combinations were tested: i) defatting beforehand of after extraction with hexanes; ii) methyl ter-butyl ether, ethyl acetate, different percentages of methanol/water as solvents; iii) acetic acid (1%) as additive; iv) under sonication for 10 to 30 min, at temperatures varying from 25 to 40 °C (data not shown). The extracts were injected in the LC-HR-Q-TOF using the same column and solvents described in Cocuron et al. (2019) but a different gradient (see Materials and Methods section). The method which resulted in the best recovery for the most diverse set of metabolites was using two rounds of extraction: the first one 100% methanol, and the second one methanol/water (30:70, v/v) with sonication rounds of 20 min each at 35-40 °C. This procedure was adopted to extract phytochemicals from coffee and wild *Rubiaceae* leaves.

Leaves from three different *Rubiaceae* species were collected in Costa Rica: i) one commercial coffee (CC), *Coffea arabica* cv. Obatã IAC 1669-20 from a plantation, and ii) two wild *Rubiaceae*, *Isertia hankeana* (WR1) and *Simira maxonii* (WR2) from the rainforest. Leaves were freeze-dried and reduced to powder. Secondary metabolites were extracted from leaf powder using the optimized extraction procedure described above, adding a filtering step , and analyzed *via* LC-HR-Q-TOF in positive and negative modes.

Runs in the positive and negative mode resulted in 19,000 and 24,000 features, respectively. Principal component analysis (PCA) of unprocessed data (wiff files) from the positive mode, obtained by the untargeted analysis of the leaf extracts, resulted in three widely separated clusters, grouping the five pseudoreplicates from each plant species together (Figure 1). The principal component (PC 1) explained 53.7% of the variance, separating the CC from WR1 and WR2, whereas the PC 2 explained 38.8% of the variance, separating all species. This can be interpreted as those three plant species having significantly different metabolic profiles.

After chromatogram alignment, data curation was performed on peak intensities, excluding features with signal lower than one million counts per second and not present on all five pseudoreplicates. Features from positive and negative modes were merged, and when they were present on both modes, only the ones with the highest area were selected. This resulted in 324 features in total within the three species extracts and their pseudoreplicates (Supplemental Table 1). National Institute of Standards and Technology (NIST17, Gaithersburg, MD) and in-house library identified 69 and 81 metabolites in the negative mode and the positive mode, respectively. As previously mentioned, for phytochemicals detected in both modes, the highest area was reported (Table 2). Thirty-one families of phytochemicals were identified: flavone, flavone C-glycoside, flavonol, indole alkaloid, monolignol phenylpropanoid and triterpenoid were the most common families of compounds found in leaf extracts (Table 2). Interestingly, flavones and terpenes were among the families of phytochemicals that were not present in CC but were detected in the wild *Rubiaceae* leaves (Table 2). For instance, 6,2'-dihydroxyflavone, afzelin and apigenin were the flavones not detected in CC leaves extracts. Also, robinin, bisdemethoxycurcumin, and madecassic acid were only detected in the wild *Rubiaceae* leaves WR1 and WR2, while epicatechin, vicenin 2, isoquercetrin, 2,3-dehydrosilybin, theobromine, theophylline, trans-3-coumaric acid, and neomangiferin were only present in CC (Table 2).

To validate the findings from the metabolomic profiling, a targeted metabolomic approach was conducted using known quantities of commercially available phytochemical standards.

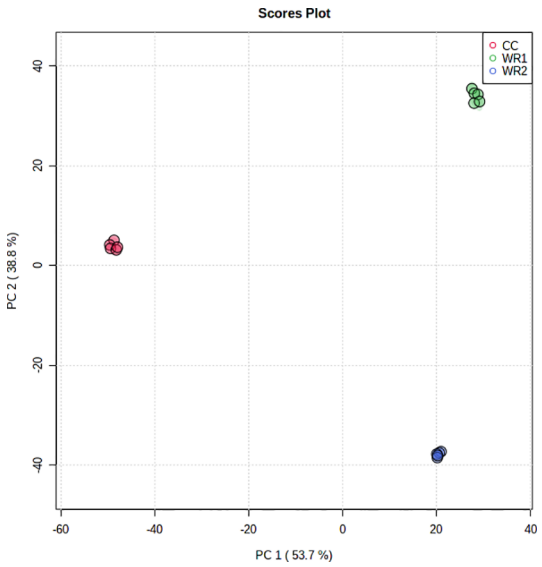



FIGURE 1
PCA analysis performed on features from untargeted metabolomic analysis in the positive mode of leaf extracts of commercial coffee (CC) and two wild *Rubiaceae* species (WR1 and WR2) collected in Costa Rica. The extractions and analyses were performed with five pseudoreplicates per species.

## 3.2 Targeted metabolomics

### 3.2.1 Method development

One hundred eighty-four phytochemical standards were individually infused in a triple quadruple to optimize the mass spectrometry parameters (declustering potential - DP, collision energy - CE and collision cell exit potential - CXP) associated with precursor/product ion (Q1/Q3) as shown in Table 1. A multiple reaction monitoring (MRM) scan survey was implemented using these parameters. For the liquid chromatography part, the same column, solvents, and additive than the untargeted metabolomics were used. Mixtures of standards were injected into the LC-MS/MS to obtain the retention time for each phytochemical (Table 1). It is important to note that standards with the same Q1/Q3 transition were separated by their retention times, except for hesperetin and homoeriodictyol, isovitexin and vitexin, vicenin 3 and schaftoside, morin and tricetin, luteolin-7,3'-Di-O-glucoside and kaempferol-3-O-sophoroside, luteolin-7-O-glucoside and Kaempferol-7-O-glucoside, genistein-7-O-glucuronide and baicalin. For the LC-MS/MS analysis of the 184 phytochemicals, polarity switching (between positive and negative modes) was used within the same run, and MRM windows (aka scheduled MRM) were set up for each compound according to its retention time.

A calibration curve was performed for each standard to determine the limits of quantification, detection, and the linearity range (Table 3). The coefficients of correlations were all above 0.98. Acacetin-7-O-rutinoside was the compound with the lowest limits of quantification and detection, at 0.11 and 0.03 fmol, respectively. On the other hand, 2-hydroxyconiferaldehyde was the compound with highest limits of quantification and detection, at 1724.63 and 517.39 fmol, respectively. Tricetin, 2,5-dihydroxybenzoic acid and 4,5-di-O-caffeoylquinic acid were the compounds with the widest range of quantification, from 32.77 to 1,250,000.00 fmol, 81.92 to 1,250,000.00 fmol and 163.84 to 2,500,000.00 fmol, respectively (Table 3).

Recovery efficiency (RE) and matrix effect (ME) were assessed for each metabolite. Ninety percent of the metabolites were recovered with an efficiency of at least 50%. Keracyanin and enoxolone were the compounds with the lowest REs of 23.3 and 28.9%, respectively (Table 4). On the other hand, 5-hydroxytryptophan and 3,4-di-O-caffeoylquinic acid had REs over 100 (Table 4). In parallel, ion suppression or enhancement was measured (i.e. ME > ± 30%) for less than 30% of the phytochemicals. For instance, there was a strong negative ME for 3,4-di-O-caffeoylquinic acid and naringin, underlying ion suppression from the sample matrix. Similarly, 4,5-dicaffeoylquinic acid and gossypetin were the compounds with the highest positive ME, indicating ion enhancement (Table 4).

To verify the method accuracy, at least one representative of each major phytochemical family was added to coffee leaf extracts at three concentrations: 0.25, 0.5, and 1μM. Intra-day and inter-day accuracy percentages were determined by re-injecting the samples on the same day and on three different

10.3389/fpls.2022.1057645

TABLE 2 Families of phytochemical detected by untargeted metabolomics.

| Family | Spectral library match | mz/RT (+/-) | CC | WR1 | WR2 |
|---|---|---|---|---|---|
| acyl glycine | Hippuric acid | 180.0596/2.1 (-) | 3.31E+05 ± 1.24E+04 | 2.27E+05 ± 2.45E+04 | 2.06E+05 ± 4.25E+03 |
| amine | Tyramine | 138.0913/0.9 (+) | 3.37E+06 ± 1.35E+05 | 5.53E+04 ± 3.27E+03 | u.d |
| benzopyrone | Coumarin | 147.0439/10.6 (+) | 2.15E+04 ± 3.32E+03 | 5.58E+05 ± 2.78E+04 | 5.12E+05 ± 2.93E+04 |
| cinnamate ester | Neochlorogenic acid | 353.0877/6.7 (-) | 1.08E+07 ± 2.04E+06 | 1.46E+06 ± 3.03E+05 | 3.58E+05 ± 7.17E+04 |
| dihyroflavonol | Dihydrokaempferol | 287.0563/9.7 (-) | 8.64E+04 ± 4.66E+03 | 1.19E+05 ± 4.98E+03 | 1.90E+04 ± 1.52E+03 |
| diterpenoid | Gibberellic acid | 345.1345/8.4 (-) | u.d | u.d | 1.30E+06 ± 9.65E+05 |
| flavan 3-ol | Catechin | 289.0715/6.7 (-) | 3.99E+05 ± 5.85E+04 | 1.85E+04 ± 3.20E+03 | 4.03E+05 ± 1.34E+04 |
| flavan 3-ols | Epicatechin | 289.0715/7.0 (-) | 1.36E+05 ± 1.57E+04 | u.d | u.d |
| flavanone | Eriodictyol | 287.0561/6.8 (-) | 9.14E+04 ± 7.19E+03 | u.d | 2.01E+05 ± 4.65E+03 |
| flavone | 6,2'-Dihydroxyflavanone | 257.0675/8.3 (-) | u.d | 2.16E+06 ± 1.15E+05 | 9.32E+03 ± 9.54E+02 |
| flavone | Afzelin | 431.0978/9.7 (-) | u.d | 6.86E+04 ± 7.93E+03 | 4.07E+04 ± 5.71E+03 |
| flavone | Apigenin | 269.0458/11.9 (-) | u.d | u.d | 2.69E+05 ± 1.64E+04 |
| flavone | Apigenin 7-glucoside | 431.0976/8.8 (-) | 2.21E+04 ± 3.10E+03 | 6.12E+04 ± 4.01E+03 | 5.06E+03 ± 1.76E+03 |
| flavone | Luteolin | 287.0550/8.4 (+) | 6.95E+04 ± 9.28E+03 | 7.96E+05 ± 4.60E+04 | 2.63E+06 ± 1.19E+05 |
| flavone | Naringenin | 273.0755/9.7 (+) | 9.78E+04 ± 4.34E+03 | 4.01E+04 ± 1.16E+04 | u.d |
| flavone C-glycoside | Kaempferol-3-O-rutinoside | 593.1525/8.0 (-) | 4.36E+04 ± 6.57E+03 | 3.33E+06 ± 1.38E+05 | 4.34E+06 ± 2.98E+05 |
| flavone C-glycoside | Kaempferol-7-O-Neohesperidoside | 595.1673/8.0 (+) | 3.70E+04 ± 5.42E+03 | 1.98E+06 ± 6.99E+04 | 3.19E+06 ± 2.32E+05 |
| flavone C-glycoside | Quercetin 3-glucoside | 463.0881/8.0 (-) | 4.48E+06 ± 3.81E+05 | 8.91E+04 ± 1.24E+04 | 1.41E+05 ± 1.16E+04 |
| flavone C-glycoside | Rutin | 611.1617/7.7 (+) | 1.94E+06 ± 1.45E+05 | 5.90E+04 ± 5.04E+03 | 1.54E+04 ± 1.53E+03 |
| flavone C-glycoside | Vicenin 2 | 595.1518/6.5 (-) | 1.64E+06 ± 1.33E+05 | u.d | u.d |
| flavonoid | Tiliroside | 593.1311/10.4 (-) | u.d | u.d | 6.30E+05 ± 4.19E+04 |
| flavonoid glycoside | Isoquercetrin | 465.1034/8.0 (+) | 2.31E+06 ± 1.65E+05 | u.d | u.d |
| flavonoid glycoside | Nepetin 7-glucoside | 479.1190/9.3 (+) | u.d | u.d | 2.77E+05 ± 1.64E+04 |
| flavonol | Kaempferol | 285.0401/12.3 (+) | 2.72E+04 ± 3.74E+03 | 1.64E+05 ± 1.01E+04 | 1.28E+05 ± 5.24E+03 |
| flavonol | Quercitrin | 447.0947/8.4 (-) | 1.40E+05 ± 4.23E+04 | 1.95E+06 ± 8.08E+04 | 5.65E+06 ± 2.06E+05 |
| flavonolignan | 2,3-Dehydrosilybin | 479.0823/8.0 (-) | 1.96E+05 ± 2.74E+04 | u.d | u.d |
| glycosyloxyflavone | Robinin | 739.1904/10.2 (-) | u.d | 1.18E+06 ± 8.03E+04 | 5.27E+03 ± 1.06E+03 |
| hydroquinone | 3,4-Dihydroxybenzoic acid | 153.0196/4.9 (-) | 1.05E+06 ± 1.21E+05 | 2.95E+06 ± 2.65E+05 | 1.91E+05 ± 7.31E+03 |
| indole alkaloid | Theobromine | 181.0715/4.4 (+) | 5.73E+06 ± 2.75E+05 | u.d | u.d |
| indole alkaloid | Theophylline | 181.0715/5.1 (+) | 5.16E+05 ± 2.53E+04 | u.d | u.d |
| indole alkaloid | Yohimbine/Rauwolscine | 355.2026/7.1 (+) | 1.01E+04 ± 3.91E+03 | 1.46E+06 ± 6.27E+04 | 4.89E+04 ± 1.93E+04 |
| indolizine | Isorhyncophylline | 385.2123/7.0 (+) | 6.02E+04 ± 1.73E+04 | 3.06E+06 ± 1.15E+05 | 9.46E+04 ± 3.38E+03 |
| methylxanthine alkaloid | Caffeine | 195.0891/6.1 (+) | 5.26E+08 ± 1.56E+07 | 1.50E+05 ± 4.41E+04 | 1.69E+05 ± 1.89E+04 |
| monohydroxybenzoic acid | Gentisic acid | 153.0191/6.0 (-) | 1.53E+05 ± 1.10E+04 | 1.14E+05 ± 6.57E+03 | 1.03E+05 ± 3.87E+04 |
| monohydroxybenzoic acid | Vanillic acid | 167.0344/8.2 (-) | 7.58E+04 ± 6.60E+03 | 2.80E+04 ± 2.22E+03 | 4.14E+04 ± 4.94E+03 |
| monolignol | Coumaric acid* | 163.0403/6.1 (-) | 4.51E+06 ± 5.51E+05 | 1.26E+05 ± 3.99E+03 | 8.31E+04 ± 2.90E+03 |
| monolignol | Scopoletin | 191.0352/8.4 (-) | 4.12E+04 ± 4.89E+03 | 2.81E+05 ± 1.29E+04 | 2.30E+04 ± 1.40E+03 |

*(Continued)*

TABLE 2 Continued

| Family | Spectral library match | mz/RT (+/-) | CC | WR1 | WR2 |
|---|---|---|---|---|---|
| monolignol | trans-3-Coumaric acid* | 147.0440/7.8 (+) | 1.07E+06 ± 1.77E+05 | u.d | u.d |
| oxopurine alkaloid | 1,3,7-Trimethyluric acid | 209.0684/5.3 (-) | 5.81E+05 ± 7.30E+04 | 3.13E+03 ± 7.31E+02 | 2.72E+03 ± 2.84E+02 |
| oxylipin | Methyl salicylate | 153.0544/7.2 (+) | 5.84E+05 ± 2.62E+04 | 5.59E+04 ± 1.96E+04 | 1.01E+04 ± 9.88E+02 |
| phenolic aldehyde | 2,5-Dihydroxybenzaldehyde | 137.0247/6.2 (-) | 4.28E+06 ± 1.86E+05 | 5.52E+05 ± 3.22E+04 | 1.01E+05 ± 1.23E+04 |
| phenylpropanoid | Benzoic acid | 121.0309/7.4 (-) | 1.87E+06 ± 8.04E+04 | 5.18E+06 ± 2.20E+05 | 5.11E+05 ± 2.90E+04 |
| phenylpropanoid | Esculetin | 177.0198/6.6 (-) | 1.20E+06 ± 4.69E+05 | 1.31E+05 ± 1.72E+04 | 7.18E+04 ± 5.87E+03 |
| phenylpropanoid | Esculin | 339.0719/6.4 (-) | 3.49E+04 ± 1.28E+04 | 9.44E+04 ± 2.03E+04 | 5.19E+04 ± 3.85E+03 |
| phenylpropanoid | p-Methoxycinnamic acid | 207.1011/8.1 (+) | 7.30E+04 ± 6.47E+03 | 2.94E+04 ± 1.03E+04 | u.d |
| phenylpropanoid | trans-Cinnamic acid | 181.0861/10.7 (+) | u.d | 6.00E+06 ± 3.18E+05 | u.d |
| polyphenol | Bisdemethoxycurcumin | 309.0964/8.9 (+) | u.d | 5.08E+05 ± 3.52E+04 | 1.08E+04 ± 1.64E+03 |
| polyphenol | Caffeic acid | 179.0349/6.7 (-) | 4.58E+05 ± 2.24E+04 | 2.08E+05 ± 1.01E+04 | 1.25E+05 ± 1.69E+04 |
| pyridine alkaloid | Trigonelline | 275.1030/1.0 (+) | 2.77E+07 ± 8.40E+05 | 6.01E+04 ± 9.90E+03 | 7.92E+03 ± 1.06E+03 |
| sesquiterpene | Abscisic acid | 263.1286/10.7 (-) | 7.83E+04 ± 4.48E+03 | 2.96E+04 ± 8.94E+02 | 1.43E+04 ± 5.80E+02 |
| trihydroxyanthraquinone | Chrysophanol | 255.0649/17.6 (+) | u.d | u.d | 1.66E+05 ± 1.28E+04 |
| trihydroxybenzoic acid | Gallic acid | 169.0140/3.5 (-) | 2.23E+05 ± 1.88E+04 | 4.58E+04 ± 9.22E+03 | u.d |
| triterpene | Sumaresinolic acid | 473.3626/11.9 (+) | u.d | u.d | 4.91E+05 ± 4.77E+04 |
| triterpenoid | Madecassic acid | 503.3371/15.2 (-) | u.d | 1.42E+05 ± 2.03E+04 | 1.39E+06 ± 2.18E+05 |
| triterpenoid | Maslinic acid | 471.3469/19.8 (+) | 5.32E+04 ± 9.97E+03 | 4.85E+05 ± 4.70E+04 | 8.16E+05 ± 4.40E+04 |
| triterpenoid | Soyasaponin I | 941.4949/15.3 (-) | u.d | 1.01E+05 ± 1.39E+04 | u.d |
| xanthone C-glycoside | Neomangiferin | 583.1305/5.3 (-) | 4.15E+05 ± 3.03E+04 | u.d | u.d |

The mass to charge ratio value represents the $[M+H]^+$ and $[M-H]^-$ for the positive and negative modes, respectively, except compounds marked by an asterisk (*) which denotes a loss of water in the source. The areas of secondary metabolites extracted from leaves of commercial coffee (CC) and two wild Rubiaceae species (WR1 and WR2) ± standard deviation (n = 5 pseudoreplicates) are presented. When metabolites were detected in both polarities, only the one with the highest area was reported. Areas labeled as u.d. were under the limit of detection.
For each compound identified using NIST and homemade spectral libraries with > 70% probability, the mass to charge ratio, retention time in min (mz/RT), and mode of detection (positive +, or negative - polarity) are specified.

days, respectively (Table 5). Out of the 25 phytochemicals tested, 18 and 22 had intra-day and inter-day accuracies above 75%, respectively.

### 3.2.2 Application of the analytical method to quantify phytochemicals in leaf extracts from commercial coffee and wild *Rubiaceae*

Validation of the method was performed using the same biological samples and extracts as the untargeted metabolomics (Supplemental Table 2). From the 184 phytochemicals monitored by LC-MS/MS, 74 were quantifiable in at least one of the *Rubiaceae* species. Interestingly, all the 74 phytochemicals were significantly different in at least one comparison by ANOVA ($p<0.05$). The PCA was consistent with the one from the untargeted metabolomic analysis (Figure 1), resulting in three separate clusters (Figure 2). The principal component 1 (PC 1) explained 70.1% of the variance, separating the CC from the wild *Rubiacaea*, whereas the PC 2 explained 28.9% of the variance, separating all species.

Targeted metabolomic analysis revealed compounds that were highly concentrated in commercial coffee leaf extracts when compared to wild *Rubiaceae* extracts (Figure 3; Supplemental Table 2). Caffeoylquinic acids, like 3, 4 and 5-caffeoylquinic acids, caffeine, trigonelline, vincenin 2, theobromine and others were more abundant in CC (Figure 3). Also, similarly to what was found in the untargeted analysis, some compounds were more abundant and even sometimes found exclusively in wild *Rubiaceae* leaf extracts. For instance, the levels of benzoic acid, 3,4-dimethylcinnamic acid and kaempferol-7-O-neohesperidoside, were the highest in WR1 in comparison to CC and WR2 (Figure 3). Cinnamic acid, yohimbine, corynanthine/rauwolscine, syringic acid, kaempferol were abundant in WR1 but absent in CC (Figure 3). Similarly, the levels of keracyanin, luteolin-7-O-glucoside, idaein,

10.3389/fpls.2022.1057645

TABLE 3 Range, linearity coefficient ($R^2$), limit of quantification (LOQ) and limit of detection (LOD) of secondary metabolites analyzed in this study.

| Metabolite | Range (fmol) | $R^2$ | LOQ (fmol) | LOD (fmol) |
|---|---|---|---|---|
| 1,3,7-Trimethyluric acid | 32.77 – 125,000.00 | 0.9985 | 4.54 | 1.36 |
| 2,5-Dihydroxybenzoic acid | 81.92 – 1250,000.00 | 0.9995 | 52.85 | 15.86 |
| 3,4-Dihydroxybenzaldehyde | 131.07 – 32,000.00 | 0.9941 | 46.81 | 14.04 |
| 3,4-Dihydroxybenzoic acid | 327.68 – 32,000.00 | 0.9935 | 62.30 | 18.69 |
| 3,4-Dimethoxycinnamic acid | 65.54 – 16,000.00 | 0.9989 | 22.91 | 6.87 |
| 3,4-di-O-Caffeoylquinic acid | 819.20 – 80,000.00 | 0.9951 | 364.09 | 109.23 |
| 3,5-Dihydroxybenzaldehyde | 327.68 – 1,250,000.00 | 0.9983 | 117.87 | 35.36 |
| 3,5-Dimethoxybenzaldehyde | 32.77 – 20,000.00 | 0.9977 | 9.55 | 2.87 |
| 3,5-Dimethoxybenzoic acid | 13.11 – 8,000.00 | 0.9923 | 4.68 | 1.40 |
| 3-Caffeoylquinic acid | 163.84 – 40,000.00 | 0.9965 | 81.92 | 24.58 |
| 3-Deoxyrobinetin | 65.54 – 100,000.00 | 0.9945 | 26.43 | 7.93 |
| 3-Hydroxystilbene | 327.68 – 200,000.00 | 0.9852 | 106.74 | 32.02 |
| 4,5-di-O-Caffeoylquinic acid | 163.84 – 2,500,000.00 | 0.9937 | 148.95 | 44.68 |
| 4-Ccaffeoylquinic acid | 163.84 – 100,000.00 | 0.9988 | 68.99 | 20.70 |
| 4-Hydroxybenzaldehyde | 16.38 – 10,000.00 | 0.9939 | 5.43 | 1.63 |
| 4-Hydroxybenzoic acid | 32.77 – 8,000.00 | 0.9994 | 10.05 | 3.02 |
| 5-Caffeoylquinic acid | 65.54 – 250,000.00 | 0.9928 | 40.96 | 12.29 |
| 5-Hydroxyconiferaldehyde | 4,096.00 – 2,500,000.00 | 1.0000 | 1,724.63 | 517.39 |
| 5-Hydroxydimetyltryptamine | 131.07 – 5,120.00 | 0.9920 | 46.32 | 13.89 |
| 5-Hydroxytryptophan | 16.36 – 62,425.00 | 0.9936 | 5.20 | 1.56 |
| 6-Methylcoumarin | 16.38 – 4,000.00 | 0.9990 | 7.99 | 2.40 |
| 7,8-Dihydroxy-4-methylcoumarin | 327.68 – 500,000.00 | 0.9948 | 114.98 | 34.49 |
| 7-Hydroxymitragynine | 163.84 – 100,000.00 | 0.9970 | 59.15 | 17.74 |
| Abscisic acid | 16.38 – 10,000.00 | 0.9978 | 3.60 | 1.08 |
| Acacetin-7-O-rutinoside | 0.33 – 500.00 | 0.9957 | 0.11 | 0.03 |
| Afzelin | 1.31 – 2,000.00 | 0.9994 | 0.31 | 0.09 |
| alpha-Cyperone | 16.38 – 4,000.00 | 0.9979 | 2.42 | 0.72 |
| Anthranilic acid | 131.07 – 200,000.00 | 0.9967 | 80.41 | 24.12 |
| Apigenin | 32.77 – 8,000.00 | 0.9955 | 20.61 | 6.18 |
| Apigenin-7-glucuronide | 1.31 – 5,000.00 | 0.9997 | 0.45 | 0.13 |
| Apigenin-7-O-glucoside | 3.17 – 1,936.00 | 0.9971 | 1.21 | 0.36 |
| Apigeninidin | 64.93 – 39,628.00 | 0.9984 | 13.09 | 3.93 |
| Arbutin | 327.00 – 80,000.00 | 0.9957 | 52.10 | 15.63 |
| Arctigenin | 327.29 – 79,905.60 | 0.9928 | 87.51 | 26.25 |
| Artemisinin | 13.11 – 8,000.00 | 0.9966 | 7.62 | 2.29 |
| Baicalein | 3.28 – 5,000.00 | 0.9993 | 2.05 | 0.61 |

*(Continued)*

TABLE 3 Continued

| Metabolite | Range (fmol) | $R^2$ | LOQ (fmol) | LOD (fmol) |
|---|---|---|---|---|
| Baicalin | 32.77 – 20,000.00 | 0.9971 | 14.83 | 4.45 |
| Biochanin A | 32.77 – 20,000.00 | 0.9957 | 20.10 | 6.03 |
| Brefeldin A | 131.07 – 80,000.00 | 0.9901 | 42.42 | 12.73 |
| Caffeic acid | 131.07 – 500,000.00 | 0.9983 | 48.55 | 14.56 |
| Caffeine | 3.28 – 800.00 | 0.9984 | 1.12 | 0.34 |
| Caffeoyl-shikimate | 163.84 – 250,000.00 | 0.9983 | 85.11 | 25.53 |
| Caffeyl alcohol | 16.38 – 25,000.00 | 1.0000 | 10.40 | 3.12 |
| Caffeyl aldehyde | 6.55 – 25,000.00 | 0.9961 | 6.43 | 1.93 |
| Calycosin | 6.55 – 4,000.00 | 0.9936 | 1.59 | 0.48 |
| Catechin | 16.38 – 10,000.00 | 0.9928 | 2.43 | 0.73 |
| Chrysin | 16.38 – 10,000.00 | 0.9976 | 5.67 | 1.70 |
| Chrysoeriol | 3.28 – 800.00 | 0.9968 | 1.34 | 0.40 |
| Cinnamic acid | 32.77 – 8,000.00 | 0.9992 | 14.63 | 4.39 |
| Coniferaldehyde | 16.38 – 10,000.00 | 0.9902 | 4.88 | 1.46 |
| Coniferyl alcohol | 65.54 – 16,000.00 | 0.9975 | 24.82 | 7.45 |
| Corynanthine | 1.31 – 320.00 | 0.9971 | 0.33 | 0.10 |
| Coumaric acid | 16.38 – 10,000.00 | 0.9984 | 7.95 | 2.39 |
| Cynarin | 1,310.72 – 2,000,000.00 | 0.9910 | 468.11 | 140.43 |
| Dihydrokaempferol | 13.11 – 3,200.00 | 0.9902 | 3.03 | 0.91 |
| Dihydroquercetin | 32.77 – 125,000.00 | 0.9935 | 26.21 | 7.86 |
| Diosmetin | 1.31 – 800.00 | 0.9935 | 0.56 | 0.17 |
| Enoxolone | 65.54 – 16,000.00 | 0.9945 | 10.39 | 3.12 |
| Epicatechin | 3.28 – 5,000.00 | 0.9952 | 1.38 | 0.41 |
| Epigallocatechin | 65.54 – 40,000.00 | 0.9955 | 11.13 | 3.34 |
| Eriodictyol | 6.55 – 4,000.00 | 0.9886 | 1.20 | 0.36 |
| Erucifoline | 16.38 – 25,000.00 | 0.9970 | 7.99 | 2.40 |
| Eupatorin | 3.28 – 800.00 | 0.9930 | 0.99 | 0.30 |
| Ferulic acid | 65.54 – 16,000.00 | 0.9914 | 13.43 | 4.03 |
| Fisetin | 65.54 – 250,000.00 | 0.9982 | 20.23 | 6.07 |
| Flavopiridol | 1.31 – 2,000.00 | 0.9964 | 0.58 | 0.17 |
| Gallic acid | 163.84 – 625,000.00 | 0.9944 | 43.57 | 13.07 |
| Gallocatechin | 65.54 – 40,000.00 | 0.9957 | 12.85 | 3.86 |
| Genistein-7-O-glucuronide | 3.28 – 5,000.00 | 0.9936 | 1.30 | 0.39 |
| Genkwanin | 6.55 – 4,000.00 | 0.9933 | 2.72 | 0.82 |
| Gibberellic acid | 32.77 – 20,000.00 | 0.9960 | 11.22 | 3.37 |
| Gingerol | 1,310.72 – 32,000.00 | 0.9934 | 379.92 | 113.98 |
| Ginkgolide A | 16.38 – 4,000.00 | 0.9978 | 3.19 | 0.96 |

*(Continued)*

TABLE 3 Continued

| Metabolite | Range (fmol) | $R^2$ | LOQ (fmol) | LOD (fmol) |
|---|---|---|---|---|
| Glycitein | 6.55 – 640.00 | 0.9953 | 2.32 | 0.70 |
| Glycitin | 64.49 – 15,744.00 | 0.9869 | 15.10 | 4.53 |
| Harmane | 1.31 – 800.00 | 0.9928 | 0.44 | 0.13 |
| Hesperetin | 3.28 – 2,000.00 | 0.9975 | 1.22 | 0.37 |
| Homoeriodictyol | 3.28 – 800.00 | 0.9998 | 1.21 | 0.36 |
| Hordenine | 3.28 – 2,000.00 | 0.9995 | 0.83 | 0.25 |
| Idaein | 163.84 – 250,000.00 | 0.9999 | 159.84 | 47.95 |
| Ipriflavone | 1.31 – 128.00 | 0.9989 | 0.29 | 0.09 |
| Isoorientin | 32.77 – 20,000.00 | 0.9991 | 12.90 | 3.87 |
| Isophorone | 6.47 – 632.32 | 0.9812 | 1.32 | 0.39 |
| Isoquercetrin | 6.55 – 4,000.00 | 0.9979 | 2.16 | 0.65 |
| Isorhamnetin | 6.55 – 1,600.00 | 0.9969 | 1.50 | 0.45 |
| Isorhamnetin-3-O-glucoside | 1.13 – 5,000.00 | 0.9964 | 0.81 | 0.24 |
| Isosakuranetin | 3.28 – 2,000.00 | 0.9923 | 1.21 | 0.36 |
| Isoschaftoside | 1.31 – 5,000.00 | 0.9921 | 0.52 | 0.16 |
| Isovitexin | 1.26 – 4,800.00 | 0.9969 | 0.92 | 0.28 |
| Jasmonic acid | 32.77 – 20,000.00 | 0.9898 | 6.39 | 1.92 |
| Kaempferide | 6.55 – 10,000.00 | 0.9924 | 2.31 | 0.69 |
| Kaempferitrin | 6.55 – 1,600.00 | 0.9989 | 3.58 | 1.07 |
| Kaempferol | 32.77 – 8,000.00 | 0.9986 | 13.43 | 4.03 |
| Kaempferol-3-O-glucoside | 6.55 – 640.00 | 0.9967 | 3.01 | 0.90 |
| Kaempferol-3-O-glucuronide | 6.55 – 25,000.00 | 0.9974 | 2.29 | 0.69 |
| Kaempferol-3-O-rutinoside | 6.55 – 1,600.00 | 0.9996 | 2.40 | 0.72 |
| Kaempferol-3-O-sophoroside | 20.48 – 5,000.00 | 0.9997 | 8.40 | 2.52 |
| Kaempferol-7-O-glucoside | 3.28 – 800.00 | 0.9988 | 0.78 | 0.24 |
| Kaempferol-7-O-Neohesperidoside | 3.28 – 2,000.00 | 0.9946 | 1.27 | 0.38 |
| Keracyanin | 32.77 – 20,000.00 | 0.9992 | 7.38 | 2.21 |
| Luteolin | 13.11 – 8,000.00 | 0.9983 | 4.24 | 1.27 |
| Luteolin-7,3'-Di-O-glucoside | 3.28 – 128.00 | 0.9996 | 1.84 | 0.55 |
| Luteolin-7-O-glucoside | 1.31 – 2,000.00 | 0.9927 | 0.67 | 0.20 |
| Luteolin-7-O-glucuronide | 3.28 – 2,000.00 | 0.9923 | 1.62 | 0.49 |
| Matairesinol | 131.07 – 32,000.00 | 0.9939 | 40.96 | 12.29 |
| Mellein | 13.11 – 3,200.00 | 0.9996 | 3.96 | 1.19 |
| Methyl-jasmonate | 0.33 – 12.80 | 0.9878 | 0.12 | 0.04 |
| Miquelianin | 16.38 – 25,000.00 | 0.9985 | 9.36 | 2.81 |
| Mitragynine | 6.55 – 25,000.00 | 0.9977 | 4.12 | 1.24 |
| Morin | 32.77 – 20,000.00 | 0.9963 | 12.90 | 3.87 |

*(Continued)*

TABLE 3 Continued

| Metabolite | Range (fmol) | $R^2$ | LOQ (fmol) | LOD (fmol) |
|---|---|---|---|---|
| Myricetin | 327.68 – 1,250,000.00 | 0.9900 | 76.20 | 22.86 |
| Myricetin-3-O-Rhamnoside | 32.77 – 125,000.00 | 0.9925 | 9.13 | 2.74 |
| N,N-Dimethyltryptamine | 0.63 – 2,389.50 | 0.9925 | 0.19 | 0.06 |
| Naringenin | 6.55 – 4,000.00 | 0.9950 | 2.86 | 0.86 |
| Naringenin-7-O-glucoside | 1.31 – 800.00 | 0.9984 | 0.30 | 0.09 |
| Naringin | 1.31 – 2,000.00 | 0.9994 | 0.78 | 0.23 |
| Neobavaisoflavone | 1.31 – 320.00 | 0.9983 | 0.33 | 0.10 |
| Neodiosmin | 1.31 – 2,000.00 | 0.9999 | 0.46 | 0.14 |
| Orientin | 16.38 – 10,000.00 | 0.9866 | 6.94 | 2.08 |
| Paynantheine | 13.11 – 1,280.00 | 0.9915 | 6.43 | 1.93 |
| p-Coumaraldehyde | 131.07 – 12,800.00 | 0.9928 | 28.81 | 8.64 |
| p-Coumaryl-alcohol | 131.07 – 32,000.00 | 0.9915 | 37.24 | 11.17 |
| Phloretin | 3.28 – 2,000.00 | 0.9939 | 1.10 | 0.33 |
| Piceid | 32.77 – 20,000.00 | 0.9936 | 7.67 | 2.30 |
| Pinosylvin | 32.77 – 20,000.00 | 0.9995 | 14.50 | 4.35 |
| Prunetin | 6.55 – 4,000.00 | 0.9954 | 5.46 | 1.64 |
| Psilocybin | 131.07 – 80,000.00 | 0.9982 | 43.84 | 13.15 |
| Pterostilbene | 13.11 – 8,000.00 | 0.9922 | 3.34 | 1.00 |
| Quercetin | 32.77 – 8,000.00 | 0.9903 | 7.20 | 2.16 |
| Quercetin-3,4'-O-diglucoside | 3.28 – 800.00 | 0.9955 | 0.74 | 0.22 |
| Quercetin-3-O-galactoside | 16.38 – 10,000.00 | 0.9921 | 5.57 | 1.67 |
| Quercitrin | 16.38 – 10,000.00 | 0.9975 | 9.93 | 2.98 |
| Rauwolscine | 1.31 – 320.00 | 0.9954 | 0.32 | 0.09 |
| Reserpine | 3.28 – 5,000.00 | 0.9938 | 1.21 | 0.36 |
| Resveratrol | 65.54 – 40,000.00 | 0.9935 | 18.41 | 5.52 |
| Rhamnazin | 6.55 – 4,000.00 | 0.9988 | 2.32 | 0.70 |
| Rhamnetin | 13.11 – 8,000.00 | 0.9971 | 3.05 | 0.91 |
| Roridin L2 | 16.26 – 9,926.40 | 0.9937 | 7.49 | 2.25 |
| Rutin | 13.11 – 8,000.00 | 0.9912 | 3.23 | 0.97 |
| Sakuranetin | 3.28 – 2,000.00 | 0.9960 | 1.17 | 0.35 |
| Salicylic acid | 16.38 – 10,000.00 | 0.9939 | 4.45 | 1.34 |
| Schaftoside | 3.28 – 5,000.00 | 0.9950 | 1.70 | 0.51 |
| Scopoletin | 65.54 – 250,000.00 | 0.9953 | 29.93 | 8.98 |
| Scutellarein | 163.84 – 16,000.00 | 0.9984 | 77.10 | 23.13 |
| Seneciphylline | 6.55 – 4,000.00 | 0.9950 | 2.23 | 0.67 |
| Sinapaldehyde | 32.77 – 8,000.00 | 0.9975 | 18.94 | 5.68 |
| Sinapic acid | 131.07 – 32,000.00 | 0.9949 | 67.56 | 20.27 |

*(Continued)*

TABLE 3 Continued

| Metabolite | Range (fmol) | $R^2$ | LOQ (fmol) | LOD (fmol) |
|---|---|---|---|---|
| Sinapyl-alcohol | 1,310.00 – 2,000,000.00 | 0.9939 | 471.48 | 141.44 |
| Sophoricoside | 3.28 – 2,000.00 | 0.9958 | 1.20 | 0.36 |
| Spiraeoside | 3.28 – 5,000.00 | 0.9962 | 1.86 | 0.56 |
| Strychnine | 16.38 – 10,000.00 | 0.9974 | 4.38 | 1.31 |
| Swertiajaponin | 13.11 – 8,000.00 | 0.9978 | 4.52 | 1.36 |
| Swertisin | 6.55 – 1,600.00 | 0.9973 | 3.58 | 1.07 |
| Syringic acid | 65.54 – 40,000.00 | 0.9979 | 30.06 | 9.02 |
| Tamarixetin | 6.55 – 4,000.00 | 0.9936 | 1.46 | 0.44 |
| Theobromine | 6.55 – 4,000.00 | 0.9990 | 2.32 | 0.69 |
| Theophylline | 3.28 – 2,000.00 | 0.9992 | 0.84 | 0.25 |
| Tomatidine hydrochloride | 6.55 – 4,000.00 | 0.9906 | 2.13 | 0.64 |
| Tricetin | 32.77 – 1,250,000.00 | 0.9994 | 8.38 | 2.51 |
| Tricin | 1.31 – 2,000.00 | 0.9972 | 0.58 | 0.17 |
| Trigonelline | 1.31 – 2,000.00 | 0.9810 | 0.42 | 0.12 |
| Tryptamine | 16.38 – 4,000.00 | 0.9987 | 5.73 | 1.72 |
| t-Trimethoxyresveratrol | 16.38 – 4,000.00 | 0.9962 | 4.72 | 1.42 |
| Tyramine | 13.11 – 3,200.00 | 0.9997 | 5.24 | 1.57 |
| Vanillic acid | 327.68 – 32,000.00 | 0.9942 | 109.23 | 32.77 |
| Vanillin | 16.38 – 10,000.00 | 0.9943 | 4.89 | 1.47 |
| Vanillyl-alcohol | 32.77 – 8,000.00 | 0.9992 | 10.30 | 3.09 |
| Vicenin 3 | 3.28 – 5,000.00 | 0.9907 | 1.45 | 0.43 |
| Vicenin 2 | 6.55 – 1,600.00 | 0.9993 | 2.02 | 0.61 |
| Vincosamide | 1.31 – 800.00 | 0.9972 | 0.38 | 0.11 |
| Vitexin | 1.31 – 2,000.00 | 0.9948 | 0.60 | 0.18 |
| Xanthohumol | 3.28 – 800.00 | 0.9971 | 0.89 | 0.27 |
| Yohimbine | 1.31 – 320.00 | 0.9966 | 0.45 | 0.14 |

gibberellic acid, sinalpadehyde and sinapyl-alcohol were found to be the highest in WR2, while harmane and vincosamide, abundant in WR2, were not detected in CC (Figure 3).

To determine the consistency between untargeted and targeted results, an ANOVA-simultaneous comparative analysis (ASCA) was performed on the 74 polyphenols quantified in our samples, and it revealed that both techniques display similar results. Figure 4 are three examples of phytochemicals—caffeine, vicenin and 3-caffeoylquinic acid (3-CQA)—that were found to be significantly more abundant in CC using the untargeted (U) metabolomics. These results were validated by a quantitative targeted (T) approach, which confirmed that the three polyphenols were the highest in CC. These examples illustrate the consistency between the untargeted and targeted metabolomics in our pipeline. Figure 5 depicts the workflow of the metabolomic platform to identify and quantify polyphenols using liquid chromatography mass spectrometry.

## 4 Discussion

Plant secondary metabolites belong to a wide variety of chemical families and are present at very small concentrations, which makes it challenging to recover with a single extraction procedure and to quantify with a sensitive enough analytical technique. Moreover, most phytochemicals are poorly soluble in water, photosensitive and not thermostable, requiring careful

TABLE 4 Recovery efficiency (RE) of secondary metabolites and matrix effect (ME) from coffee leaf extracts.

| Family | Metabolite | RE | ME |
|---|---|---|---|
| | | (%) | (%) |
| alkaloid | 1,3,7-trimethyluric acid | 89.90 | -24.20 |
| alkaloid | caffeine | 76.10 | -29.22 |
| alkaloid | corynanthine | 56.00 | -24.06 |
| alkaloid | erucifoline | 61.10 | -6.21 |
| alkaloid | harmane | 80.30 | -23.95 |
| alkaloid | hordenine | 72.30 | -44.83 |
| alkaloid | mitragynine | 46.70 | 3.28 |
| alkaloid | seneciphylline | 61.40 | 26.53 |
| alkaloid | tomatidine | 35.80 | -3.46 |
| amino acid derivative | 5-hydroxytryptophan | 102.70 | -44.92 |
| amino acid derivative | tyramine | 74.80 | -44.75 |
| anthocyanidin | apigeninidin | 66.70 | 13.19 |
| anthocyanin glycoside | keracyanin | 23.30 | -7.01 |
| cinnamate ester | 3,4-di-O-caffeoylquinic acid | 194.30 | -182.44 |
| cinnamate ester | 4,5-dicaffeoylquinic acid | 57.20 | 122.65 |
| cinnamate ester | 4-caffeoylquinic acid | 82.20 | 75.55 |
| cinnamic acid derivative | 3,4-dimethoxycinnamic acid | 76.50 | -25.81 |
| coumarin derivative | 6-methylcoumarin | 78.10 | -24.91 |
| coumarin derivative | 7,8-dihydroxy-4-methylcoumarin | 79.50 | -19.74 |
| coumarin derivative | esculetin | 82.60 | -3.48 |
| coumarin derivative | mellein | 81.20 | -20.56 |
| coumarin derivative | scopoletin | 68.70 | -24.87 |
| dihydroxybenzoic acid | 2,5-dihydroxybenzoic acid | 77.10 | -15.47 |
| dihydroxybenzoic acid | 3,4-dihydroxybenzoic acid | 76.20 | 0.76 |
| dihydroxybenzoic acid | 3,5-dimethoxybenzoic acid | 80.50 | -24.28 |
| dimethoxybenzene | 3,5-dimethoxybenzaldehyde | 82.40 | -21.11 |
| diterpenoid | gibberellic acid | 66.70 | -35.04 |
| diterpenoid | ginkgolide A | 58.90 | -24.42 |
| flavanol | catechin | 81.60 | -23.50 |
| flavanol | epicatechin | 83.10 | 15.05 |
| flavanol | epigallocatechin | 73.60 | -85.83 |
| flavanone | eriodictyol | 73.10 | -29.82 |
| flavanone | hesperetin | 80.20 | -20.55 |
| flavanone | homoeriodictyol | 72.90 | -18.14 |
| flavanone | isosakuranetin | 80.10 | -20.79 |

*(Continued)*

10.3389/fpls.2022.1057645

TABLE 4 Continued

| Family | Metabolite | RE (%) | ME (%) |
|---|---|---|---|
| flavanone | naringenin | 74.30 | -22.78 |
| flavanone O-glycoside | naringenin-7-O-glucoside | 63.00 | -35.42 |
| flavanone-C-glycoside | swertiajaponin | 56.20 | 21.92 |
| flavanone-O-glycoside | naringin | 45.80 | -90.06 |
| flavanonol | dihydrokaempferol | 71.60 | -23.12 |
| flavanonol | dihydroquercetin | 69.40 | -24.51 |
| flavone | 3-deoxyrobinetin | 60.00 | 53.64 |
| flavone | apigenin | 78.90 | -28.58 |
| flavone | apigenin-7-glucuronide | 54.20 | -16.12 |
| flavone | baicalein | 55.20 | 7.23 |
| flavone | chrysin | 87.30 | -34.61 |
| flavone | chrysoeriol | 78.80 | -21.68 |
| flavone | diosmetin | 69.80 | -16.06 |
| flavone | eupatorin | 77.50 | -21.60 |
| flavone | flavopiridol | 72.20 | 80.23 |
| flavone | genkwanin | 76.90 | -22.52 |
| flavone | luteolin | 68.80 | -12.86 |
| flavone | myricetin | 51.00 | -18.64 |
| flavone | scutellarein | 50.40 | 51.01 |
| flavone | tectochrysin | 67.80 | -18.77 |
| flavone | tricetin | 53.80 | 36.58 |
| flavone C-glycoside | isoorientin | 53.50 | 4.16 |
| flavone C-glycoside | swertisin | 55.60 | -15.14 |
| flavone-C-glycoside | isoschaftoside | 66.20 | -22.12 |
| flavone-C-glycoside | isovitexin/vitexin | 58.80 | -16.56 |
| flavone-C-glycoside | orientin | 52.10 | 11.89 |
| flavone-C-glycoside | Vicenin 2 | 42.00 | -10.71 |
| flavone-O-glycoside | acacetin-7-O-rutinoside | 61.80 | -9.30 |
| flavone-O-glycoside | apigenin-7-O-glucoside | 56.90 | -25.08 |
| flavone-O-glycoside | baicalin | 62.00 | -20.13 |
| flavone-O-glycoside | luteolin-7,3'-di-O-glucoside | 64.50 | -32.82 |
| flavone-O-glycoside | luteolin-7-O-glucoside | 42.40 | -27.38 |
| flavone-O-glycoside | luteolin-7-O-glucuronide | 69.20 | -12.69 |
| flavone-O-glycoside | myricetin-3-O-rhamnoside | 44.60 | 47.77 |
| flavone-O-glycoside | neodiosmin | 43.60 | -88.65 |
| flavonol | fisetin | 72.80 | 17.56 |

*(Continued)*

TABLE 4 Continued

| Family | Metabolite | RE (%) | ME (%) |
|---|---|---|---|
| flavonol | gossypetin | 38.30 | 186.65 |
| flavonol | isorhamnetin | 57.70 | -13.77 |
| flavonol | kaempferide | 73.60 | -20.18 |
| flavonol | kaempferol | 69.30 | -21.47 |
| flavonol | morin | 54.00 | 28.56 |
| flavonol | quercetin | 67.10 | -17.24 |
| flavonol | quercitrin | 58.20 | -2.56 |
| flavonol | rhamnazin | 60.80 | -15.60 |
| flavonol | rhamnetin | 61.90 | -12.18 |
| flavonol | tamarixetin | 66.90 | -18.37 |
| flavonol-O-glycoside | afzelin | 68.60 | -26.06 |
| flavonol-O-glycoside | isoquercetrin | 68.40 | 16.38 |
| flavonol-O-glycoside | isorhamnetin-3-O-glucoside | 68.50 | 2.12 |
| flavonol-O-glycoside | kaempferitrin | 52.00 | 0.79 |
| flavonol-O-glycoside | kaempferol-3-O-glucoside | 60.20 | 28.18 |
| flavonol-O-glycoside | kaempferol-3-O-glucuronide | 71.20 | -18.36 |
| flavonol-O-glycoside | kaempferol-7-O-glucoside | 56.30 | -29.40 |
| flavonol-O-glycoside | kaempferol-7-O-neohesperidoside | 51.30 | 2.93 |
| flavonol-O-glycoside | quercetin-3,4'-O-diglucoside | 49.20 | -21.71 |
| flavonol-O-glycoside | quercetin-3-O-galactoside | 53.00 | -0.89 |
| flavonol-O-glycoside | rutin | 33.80 | -48.97 |
| flavonol-O-glycoside | spiraeoside | 55.40 | -20.78 |
| hydroxybenzoic acid | 4-hydroxybenzoic acid | 83.00 | -36.35 |
| hydroxybenzoic acid | gallic acid | 76.40 | -19.35 |
| hydroxybenzoic acid | salicylic acid | 51.40 | -74.19 |
| hydroxybenzoic acid | vanillic acid | 82.00 | -78.18 |
| hydroxycinnamic acid | caffeic acid | 65.30 | -17.55 |
| hydroxycinnamic acid | coumaric acid | 80.20 | -25.62 |
| hydroxycinnamic acid | ferulic acid | 75.40 | -18.89 |
| indole alkaloid | 5-hydroxydimethyltryptamine | 91.50 | -37.98 |
| indole alkaloid | 7-hydroxy mitragynine | 59.20 | -20.88 |
| indole alkaloid | anthranilic acid | 72.60 | 48.89 |
| indole alkaloid | N,N-dimethyltryptemine | 79.30 | -0.45 |
| indole alkaloid | psilocybin | 82.50 | 45.11 |
| indole alkaloid | rauwolscine | 57.20 | -22.77 |
| indole alkaloid | reserpine | 30.70 | 0.24 |

*(Continued)*

TABLE 4 Continued

| Family | Metabolite | RE (%) | ME (%) |
|---|---|---|---|
| indole alkaloid | theobromine | 76.50 | -17.88 |
| indole alkaloid | theophylline | 77.00 | -25.19 |
| indole alkaloid | tryptamine | 72.60 | -42.00 |
| isoflavone | biochanin A | 74.50 | -24.03 |
| isoflavone | calycosin | 70.90 | -21.78 |
| isoflavone | glycitein | 81.40 | -22.87 |
| isoflavone | ipriflavone | 66.40 | -24.26 |
| isoflavone | neobavaisoflavone | 65.30 | -21.12 |
| isoflavone | prunetin | 75.60 | -17.48 |
| isoflavone-O-glycoside | genistein-7-O-glucuronide | 69.30 | -19.96 |
| isoflavone-O-glycoside | glycitin | 53.50 | -12.45 |
| isoflavone-O-glycosylated | sophoricoside | 59.00 | -43.91 |
| lactone | brefeldin A | 68.30 | -33.62 |
| lignan | arctigenin | 63.50 | -21.16 |
| lignan | matairesinol | 64.90 | -24.19 |
| monocarboxilic acid | cinnamic acid | 80.50 | -21.71 |
| mycotoxin | neosolaniol | 87.90 | 36.21 |
| mycotoxin | roridin-L2 | 36.50 | -22.76 |
| oxylipin | jasmonic acid | 73.30 | -25.55 |
| oxylipin | methyl jasmonate | 95.30 | -30.97 |
| phenol | gingerol | 68.80 | -20.27 |
| phenolic acid | benzoic acid | 90.80 | -36.14 |
| phenolic acid | sinapic acid | 69.60 | -14.47 |
| phenolic acid | syringic acid | 80.10 | -55.23 |
| phenolic alcohol | caffeyl alcohol | 81.80 | -33.98 |
| phenolic alcohol | coniferyl alcohol | 77.60 | -20.35 |
| phenolic alcohol | sinapyl alcohol | 67.90 | -17.88 |
| phenolic alcohol | vanillyl alcohol | 83.70 | -22.66 |
| phenolic aldehyde | 3,4-dihydroxycbenzaldehyde | 93.40 | 41.47 |
| phenolic aldehyde | 4-hydroxybenzaldehyde | 78.50 | -70.69 |
| phenolic aldehyde | caffeyl aldehyde | 77.70 | -49.70 |
| phenolic aldehyde | coniferaldehyde | 77.40 | -20.36 |
| phenolic aldehyde | sinapaldehyde | 70.10 | -56.06 |
| phenolic aldehyde | vanillin | 79.60 | -24.83 |
| sesquiterpene | abscisic acid | 60.60 | -29.79 |
| sesquiterpene | artemisinin | 49.30 | -22.97 |

*(Continued)*

TABLE 4 Continued

| Family | Metabolite | RE (%) | ME (%) |
|---|---|---|---|
| sesquiterpene | heptelidic acid | 68.30 | -20.29 |
| sesquiterpene | α-cyperone | 70.30 | -19.83 |
| stilbenoid | 3-hydroxystilbene | 78.20 | -23.62 |
| stilbenoid | piceid | 50.90 | -7.63 |
| stilbenoid | pinosylvin | 77.80 | -18.70 |
| stilbenoid | pterostilbene | 66.80 | -33.96 |
| stilbenoid | resveratrol | 64.10 | -32.92 |
| stilbenoid | t-trimethoxyresveratrol | 58.60 | -19.81 |
| triterpenoid | enoxolone | 28.90 | -25.84 |

TABLE 5 Intra and inter-day accuracy percentages for secondary metabolites added at three different concentrations (0.25, 0.5 and 1 μM) to coffee leaf extracts.

| | | Intra-day assay | Inter-day assay |
|---|---|---|---|
| Metabolite | Concentration (μM) | n = 5 | n=15 |
| 1,3,7-trimethyluric acid | 0.25 | 11.3 | 16.0 |
| | 0.50 | 25.2 | 19.9 |
| | 1.00 | 16.3 | 15.9 |
| 3,4-dimethoxycinnamic acid | 0.25 | 3.0 | 2.1 |
| | 0.50 | 3.0 | 2.6 |
| | 1.00 | 1.0 | 0.6 |
| 3-Hydroxystilbene | 0.25 | 27.7 | 12.2 |
| | 0.50 | 6.1 | 6.8 |
| | 1.00 | 2.2 | 2.9 |
| 4-Hydroxybenzoic acid | 0.25 | 14.5 | 17.0 |
| | 0.50 | 15.5 | 18.5 |
| | 1.00 | 14.0 | 17.1 |
| 5-Hydroxytryptophan | 0.25 | 6.5 | 4.5 |
| | 0.50 | 46.1 | 28.7 |
| | 1.00 | 43.6 | 31.8 |
| Apigeninidin | 0.25 | 7.5 | 5.4 |
| | 0.50 | 9.5 | 9.2 |
| | 1.00 | 11.4 | 11.0 |
| Arctigenin | 0.25 | 7.1 | 7.7 |
| | 0.50 | 3.4 | 3.7 |
| | 1.00 | 2.6 | 3.4 |

*(Continued)*

TABLE 5 Continued

| | | Intra-day assay | Inter-day assay |
|---|---|---|---|
| Metabolite | Concentration (μM) | n = 5 | n=15 |
| Biochanin A | 0.25 | 6.1 | 8.5 |
| | 0.50 | 3.6 | 4.8 |
| | 1.00 | 1.0 | 1.9 |
| Coniferaldehyde | 0.25 | 7.9 | 7.0 |
| | 0.50 | 6.4 | 5.8 |
| | 1.00 | 3.0 | 3.5 |
| Coniferyl alcohol | 0.25 | 30.8 | 26.1 |
| | 0.50 | 28.3 | 23.1 |
| | 1.00 | 18.6 | 17.1 |
| Cyperone | 0.25 | 7.9 | 3.2 |
| | 0.50 | 1.8 | 2.3 |
| | 1.00 | 1.0 | 1.6 |
| Dihydrokaempferol | 0.25 | 5.9 | 5.3 |
| | 0.50 | 7.6 | 7.5 |
| | 1.00 | 6.2 | 5.7 |
| Epicatechin | 0.25 | 29.3 | 18.5 |
| | 0.50 | 31.7 | 23.6 |
| | 1.00 | 4.5 | 9.2 |
| Ferulic acid | 0.25 | 15.3 | 11.3 |
| | 0.50 | 17.1 | 14.6 |
| | 1.00 | 1.6 | 4.6 |
| Gibberellic acid | 0.25 | 5.5 | 8.9 |
| | 0.50 | 1.5 | 7.7 |
| | 1.00 | 4.3 | 8.2 |
| Genistein-7-O-glucuronide | 0.25 | 3.4 | 8.2 |
| | 0.50 | 1.8 | 8.2 |
| | 1.00 | 7.6 | 7.3 |
| Isorhamnetin-3-O-glucoside | 0.25 | 8.4 | 5.4 |
| | 0.50 | 4.6 | 5.3 |
| | 1.00 | 0.5 | 1.7 |
| Jasmonic acid | 0.25 | 6.5 | 6.7 |
| | 0.50 | 2.6 | 4.1 |
| | 1.00 | 2.4 | 3.0 |
| Luteolin | 0.25 | 1.1 | 4.9 |
| | 0.50 | 4.0 | 5.2 |
| | 1.00 | 5.0 | 3.8 |

*(Continued)*

10.3389/fpls.2022.1057645

TABLE 5 Continued

| | | Intra-day assay | Inter-day assay |
|---|---|---|---|
| Metabolite | Concentration (μM) | n = 5 | n=15 |
| Naringenin | 0.25 | 3.1 | 3.3 |
| | 0.50 | 0.1 | 1.6 |
| | 1.00 | 2.5 | 1.6 |
| Orientin | 0.25 | 8.1 | 15.1 |
| | 0.50 | 8.5 | 20.3 |
| | 1.00 | 14.0 | 17.0 |
| Scopoletin | 0.25 | 4.5 | 4.9 |
| | 0.50 | 6.7 | 7.5 |
| | 1.00 | 1.6 | 3.2 |
| Sinapic acid | 0.25 | 30.1 | 20.4 |
| | 0.50 | 28.7 | 24.3 |
| | 1.00 | 4.2 | 6.5 |
| Swertiajaponin | 0.25 | 0.8 | 25.7 |
| | 0.50 | 6.0 | 26.9 |
| | 1.00 | 41.4 | 40.1 |
| Theobromine | 0.25 | 35.4 | 17.9 |
| | 0.50 | 16.7 | 12.5 |
| | 1.00 | 5.2 | 9.1 |

Intra-day and inter-day accuracy percentages were determined by re-injecting the samples on the same day (n = 5) and on three different days (n = 15), respectively.

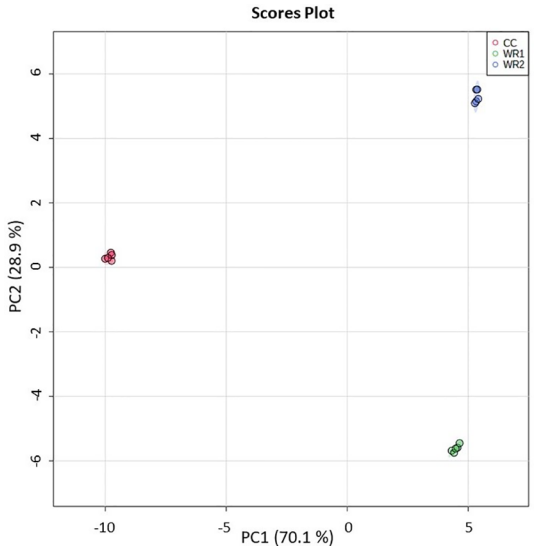



FIGURE 2
PCA analysis performed on targeted metabolomic analysis data from extracts of commercial coffee (CC) and two wild *Rubiaceae* species (WR1 and WR2). The extractions and analyses were performed with five pseudoreplicates per species. The shaded regions represent the 95% confidence intervals.

10.3389/fpls.2022.1057645

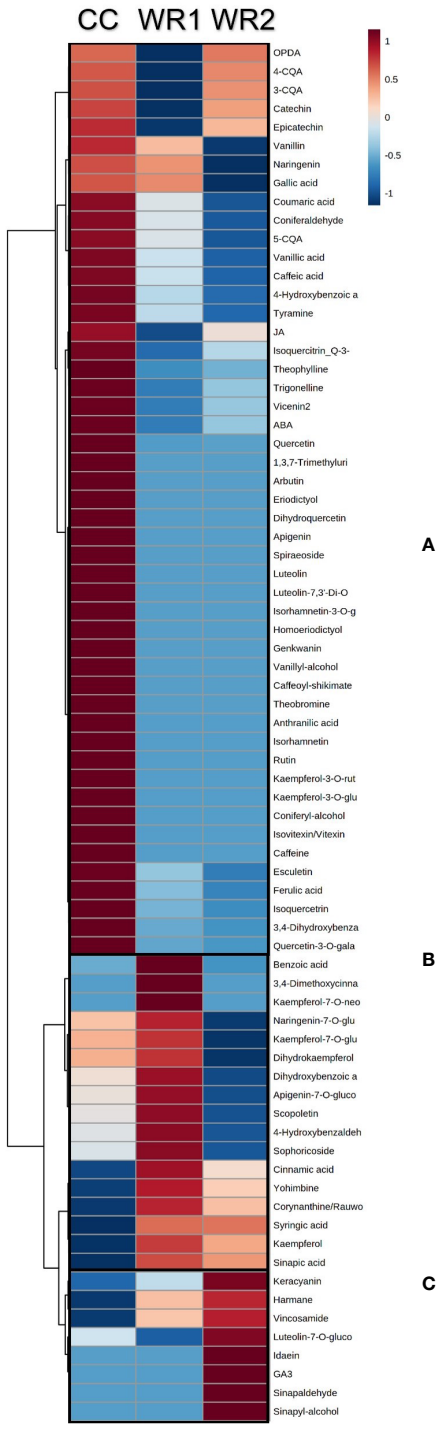



FIGURE 3
Heatmap analysis of the polyphenols from extracts of coffee (CC) and two wild *Rubiaceae* (WR1 and WR2) detected by targeted metabolomics. The metabolomics data were normalized by log transformation, mean-centered, and divided by the standard deviation of each variable. Ward's hierarchical clustering algorithm was used to group metabolites that have the same distribution pattern in the heat map. Color scale represents metabolite relative intensity, with the darkest red and blue symbolizing the highest and the lowest values, respectively. Boxes highlight metabolites that are higher in CC **(A)**, in WR1 **(B)**, and WR2 **(C)**.

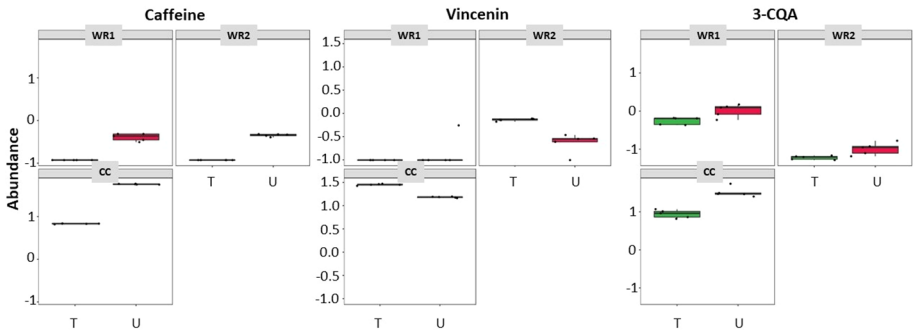



FIGURE 4
ASCA showing similarities of metabolites detected in coffee and wild *Rubiaceae* leaf extracts detected by targeted (T - green boxes) and untargeted (U – red boxes) metabolomics. 3-CQA is 3-caffeoylquinic acid.

procedures for their handling and storage. Choosing the right solvent composition and solid to liquid ratio is critical for isolating plant metabolites (Zhang et al., 2011). In general, the most common solvents are ethanol, methanol, chloroform, and water in different proportions for plant secondary metabolite extraction (Silva et al., 1998; Abubakar and Haque, 2020). Nonetheless, chloroform has a low polarity and is a carcinogen, therefore it is recommended to avoid it (Davidson et al., 2008). In the present study, methanol and water resulted in the most efficient mixture of solvents to recover polyphenols from coffee leaves, combined with sonication at 35-40°C. Indeed, ultrasonication has been widely used to reach higher yields of natural compounds (Annegowda et al., 2010; Masson et al., 2010; Hasan et al., 2017). For instance, ultrasonic-assisted extraction has been used in coffee leaves to improve the extraction of caffeine, trigonelline, rutin, chlorogenic acids, and mangiferin (Chen et al., 2020b). Furthermore, steroidal alkaloids have been successfully recovered from potato peel using ultrasound assisted extraction, obtaining at least 1.5 times more compounds in comparison with other extraction technique (Hossain et al., 2014). A comparative study in *Hibiscus* spp. concluded that using methanol and sonication resulted in better yields of phytosterols (Soares Melecchi et al., 2006). Sonication was also used to successfully extract quinones and flavonoids of six different species of *Dosera* sp. (Marczak et al., 2005). Other extraction techniques cited in literature involve microwave-assisted extraction, pressurized-liquid extraction and supercritical fluid extraction (Zhang et al., 2011; Abubakar and Haque, 2020). However, those are more time consuming and costly techniques (Roopashree and Naik, 2019).

Untargeted metabolomics aims to capture the whole metabolome (Matsuda et al., 2009), while targeted analysis focuses on the use of commercially available standards to detect and measure the quantity of metabolites present in a biological sample (Shimizu et al., 2018). It is therefore essential

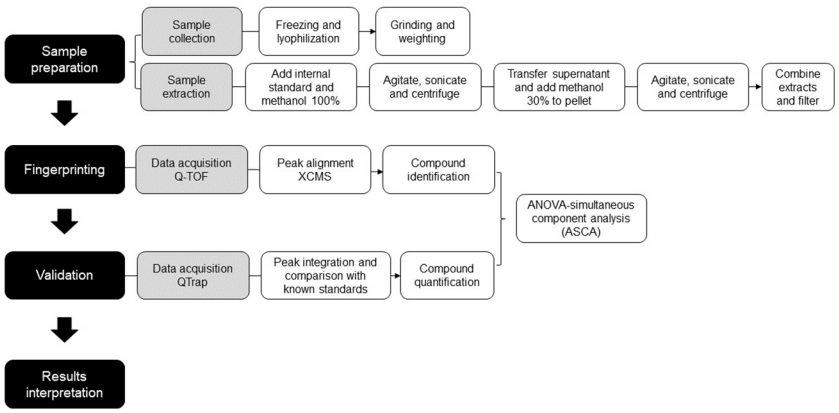



FIGURE 5
Workflow of the metabolomic platform to identify and quantify polyphenols using liquid chromatography mass spectrometry.

to have previous knowledge of the sample composition before performing targeted analysis. Most studies perform untargeted or targeted analysis, but usually do not combine both. However, the combination of both approaches, like the present study, creates a powerful tool to differentiate profiles and then detect/quantify the differences highlighted by metabolite fingerprinting, confirming the results. For instance, untargeted and targeted associated research on three *Coffea* sp. could differentiate the species metabolic profiles of leaf and fruit extracts. Additionally, five phytochemicals (caffeine, mangiferin and three caffeoylquinic acids) were identified, corroborating the identity of the differentiated metabolites between species and tissues (Montis et al., 2022). Also, Zhang et al. characterized cyclopeptides in 20 species of *Rubia* sp. (*Rubiaceae*) using LC-MS/MS (Zhang et al., 2018). Similarly to our study, the authors combined untargeted and targeted metabolomics using LC coupled to a triple TOF and a triple quadrupole, respectively, which provides reliability, precision, and sensitivity, with the additional advantage of requiring very small amounts of plant tissue. Nevertheless, our study reports an extraction that can be performed in less than two hours, and was optimized to recover a wide variety of compounds (over 40 families of phytochemicals). Indeed, 90% of the phytochemicals in our study had a recovery higher than 50%. Finally, 184 phytochemicals can be quantified in a sensitive and specific manner within a single LC-MS/MS run of 15min with polarity switch, which is particularly suitable for high-throughput analyses.

Our study compared the leaf phytochemical profiles from one commercial coffee and two wild *Rubiaceae* species. First, the untargeted metabolomics analysis identified 31 families of phytochemicals, of which flavone, flavone C-glycoside, flavonol, indole alkaloid, monolignol phenylpropanoid and triterpenoid were the most common families (Table 2). Several flavones and terpenes were not present in CC. Indeed, flavones 6,2'-dihydroxyflavone, afzelin and apigenin, the flavonoid tiliroside, flavonol nepetin-7-glucoside and the glycosylflavone robinin were not detected in CC leaves. Interestingly, flavonoids are the most abundant secondary metabolites in human diet (Alara et al., 2021): afzelin has been reported to have anti-inflammatory action (Kim et al., 2019a) while apigenin is an antioxidant and has anticancer properties (Shankar et al., 2017; Kim et al., 2019b). Flavonoids also play important roles in plant development and response to stress (Du Fall and Solomon, 2011; Nakabayashi and Saito, 2015). For example, robinin has been associated with plant drought resilience, as Chrysanthemum plants previously treated with this flavone had enhanced response to water stress and were able to maintain turgor pressure (Elansary et al., 2020). Gibberellic, sumaresinolic and madecassic acids, and soyasaponin were also absent in CC leaf extracts. Gibberellic acid is a plant hormone with multiple functions in growth regulation, flowering and stress response signaling (Bari and Jones, 2009; Schwechheimer and Willige, 2009; Iftikhar et al., 2019; Nagar et al., 2021). Madecassic acid, on the other hand, has been shown to have some medicinal properties, such as anti-inflammatory effects (Won et al., 2010), anti-colitis (Xu et al., 2017), and potential anti-cancer agent (Zhang et al., 2014; Valdeira et al., 2019). Equivalently, soyasaponins have been associated with health promoting properties, such as anti-inflammatory, anti-microbial, and cardiovascular protective activities (Guang et al., 2014; Lee et al., 2020; Wang et al., 2020).

Then, the targeted metabolomic analysis identified several compounds that were highly concentrated in CC or WR leaves (Figure 3; Supplemental Table 2). The most abundant compounds in CC leaves were chlorogenic acids (3,4 and 5-caffeoylquinic acids), caffeine, trigonelline, vicenin 2 and theobromine, which is consistent with previous reports (Campa et al., 2012; Funlayo et al., 2017; Chen, 2019; Cangeloni et al., 2022). Chlorogenic acids are well-known compounds for their antimicrobial activities (Sung and Lee, 2010; Su et al., 2014; Martínez et al., 2017). Indeed, they were shown to have deleterious effects on coffee microbial pathogens: a study showed that coffee plants supplied with silicon–a resistance inducer–had higher levels of chlorogenic acids and were therefore more resistant to *Hemilea vastatrix*, the causal agent of rust (Rodrigues et al., 2011). Caffeine content is one of the most important traits for coffee selection, either to bean processing for beverage consumption or for the pharmaceutical industry (Sawynok, 1995; Leroy et al., 2006; Patay et al., 2017; Carvalho et al., 2019). Having a resourceful method for caffeine detection and quantification, along with other desirable traits, would aid coffee breeders and studies on cultivar development. However, additional research is needed to correlate leaf and berry/bean composition in coffee to perform cultivar selection at earlier stages and speed up the breeding process. The content in kaempferol-7-O-neohesperidoside, kaempferol, yohimbine, corynanthine/rauwolscine, 3,4-dimethylcinnamic, cinnamic, benzoic and syringic acids were more abundant in WR1 leaves. A study in Litchi chinensis seeds revealed that kaempferol-7-O-neohesperidoside had a high cytotoxic activity against lung cancer cells (Xu et al., 2011). Similarly, kaempferol has also shown anti-cancer (Chen et al., 2020a; Kluska et al., 2021; Felice et al., 2022), anti-oxidant (Simunkova et al., 2021) and anti-malarial activities (Somsak et al., 2018), confirming more therapeutic uses of flavonoids. Alkaloids like yohimbine have promising clinical applications (Boğa et al., 2019; Saini et al., 2022), like anti-cancer activity (Jabir et al., 2022), and may be used as chemical markers for botanical selection (Osman et al., 2019). Several polyphenols were more abundant in WR2: harmane, vincosamide, keracyanin, idaein, gibberellic acid, luteolin-7-O-glycoside, synapaldehyde, and sinapyl-alcohol. Anthocyanins play important roles not only in plant reproduction, but also in response to abiotic and biotic stresses

(Liu et al., 2018). Additionally, keracyanin and idaein were proven to have potential anti-inflammatory and anti-cancer activities (Natarajan et al., 2016; Santamarina et al., 2021). Monolignols, key components for lignin biosynthesis, are crucial element for cell wall protection against stresses (Gallego-Giraldo et al., 2018; Xie et al., 2018). Synapaldehyde increases reactive oxygen species in plants and has anti-fungal activity (Millan et al., 2022). Also, this metabolite was more abundant in sugarcane resistant to the causal agent of ratoon stunting (Castro-Moretti et al., 2021), associating its role to plant defense. Overall, WR leaves had a higher diversity of phytochemicals in comparison to the CC. This may be due to the farmer selection of CC for fruit size, caffeine content, and yield at the detriment of other stress resistance and adaptation traits.

## 5 Conclusions

The present study reports: i) the optimization of an extraction procedure to recover 42 distinct families of phytochemicals from leaves, ii) the development of a robust and sensitive LC-MS/MS method to quantify 184 secondary metabolites, and iii) the complementarity between the untargeted and targeted metabolomics. This approach was applied to characterize the phytochemicals in three different species of *Rubiaceae*, including two wild species and one commercial coffee. The new targeted metabolomics approach was used to validate the identity of 74 compounds highlighted by the untargeted analysis. This work describes a sensitive and thorough pipeline (Figure 5) to detect, classify and quantify secondary metabolites in leaves of coffee and other *Rubiaceae*, and can be further applied to other plant organs and species.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

FC-M, J-CC, PC, JCS, and APA designed and conceptualized the project and method development; HC-G, EE-L, PC, and OG-F collected the leaves and prepared plant materials; FC-M and J-CC tested the different extraction methods, performed the LC-MS/MS method development, and method validation; FC-M. performed data processing and statistical analyses; FC-M, J-CC, and APA drafted the manuscript; APA, JCS and PC supervised, coordinated the project, and acquired funding for the project. All authors have read and agreed to the published version of the manuscript.

## Funding

This study was funded in part by National Science Foundation grant DEB-1638999 to JCS, PC and APA, and University of Costa Rica Vice rectory of Research grants B9662 and B7176 to PC. Samples were collected under the Institutional Biodiversity Commission (University of Costa Rica) resolutions VI07723-2018 and VI5152-2016.

## Acknowledgments

The authors would like to thank Pedro Juarez for helping with the collection and *Rubiaceae* identification in Costa Rica, Mariana C. Cia and Camila P. Carvalho for helping with plant collection in Brazil, and Mônica Veneziano Labate for access to the freeze dryer. All authors thank the BioAnalytical Facility (BAF) at the University of North Texas for providing the infrastructure and support to perform the metabolomic analyses.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1057645/full#supplementary-material

SUPPLEMENTARY FIGURE 1
PCA scores plot showing clustering of quality control samples for untargeted metabolomics in positive mode **(A)** and negative mode **(B)**. QC, quality control; CC, commercial coffee; WR1 and WR2, wild *Rubiaceae* species.

# References


Abubakar, A. R., and Haque, M. (2020). Preparation of medicinal plants: Basic extraction and fractionation procedures for experimental purposes. *J. Pharm. Bioallied Sci.* 12, 1. doi: 10.4103/JPBS.JPBS_175_19

Ahmad, R., and Salim, F. (2015). Oxindole alkaloids of uncaria (Rubiaceae, subfamily cinchonoideae): A review on its structure, properties, and bioactivities. *Stud. Nat. Prod. Chem.* 45, 485–525. doi: 10.1016/B978-0-444-63473-3.00012-5

Alara, O. R., Abdurahman, N. H., and Ukaegbu, C. I. (2021). Extraction of phenolic compounds: A review. *Curr. Res. Food Sci.* 4, 200–214. doi: 10.1016/J.CRFS.2021.03.011

Annegowda, H. V., Anwar, L. N., Mordi, M. N., Ramanathan, S., and Mansor, S. M. (2010). Influence of sonication on the phenolic content and antioxidant activity of terminalia catappa l. leaves. *Pharmacognosy Res.* 2, 368. doi: 10.4103/0974-8490.75457

Arias, C. L., Quach, T., Huynh, T., Nguyen, H., Moretti, A., Shi, Y., et al. (2022). Expression of AtWRI1 and AtDGAT1 during soybean embryo development influences oil and carbohydrate metabolism. *Plant Biotechnol. J.* 20, 1327–1345. doi: 10.1111/PBI.13810

Bari, R., and Jones, J. D. G. (2009). Role of plant hormones in plant defence responses. *Plant Mol. Biol.* 69, 473–488. doi: 10.1007/s11103-008-9435-0

Bataglion, G. A., Da Silva, F. M. A., Eberlin, M. N., and Koolen, H. H. F. (2015). Determination of the phenolic composition from Brazilian tropical fruits by UHPLC–MS/MS. *Food Chem.* 180, 280–287. doi: 10.1016/J.FOODCHEM.2015.02.059

Boğa, M., Bingül, M., Özkan, E. E., and Şahin, H. (2019). Chemical and biological perspectives of monoterpene indole alkaloids from rauwolfia species. *Stud. Nat. Prod. Chem.* 61, 251–299. doi: 10.1016/B978-0-444-64183-0.00007-5

Böttger, A., Vothknecht, U., Bolle, C., and Wolf, A. (2018). ""Plant secondary metabolites and their general function in plants,"," in *Lessons on caffeine, cannabis & co. learning materials in biosciences* (Cham: Springer), 3–17. doi: 10.1007/978-3-319-99546-5_1

Boudet, A. M. (2007). Evolution and current status of research in phenolic compounds. *Phytochemistry* 68, 2722–2735. doi: 10.1016/J.PHYTOCHEM.2007.06.012

Brar, D. S., and Khush, G. S. (2018). Wild relatives of rice: A valuable genetic resource for genomics and breeding research. In: Mondal, T., and Henry, R. (eds). *The Wild Oryza Genomes. Compendium of Plant Genomes* (Cham: Springer). doi: 10.1007/978-3-319-71997-9_1

Campa, C., Mondolot, L., Rakotondravao, A., Bidel, L. P. R., Gargadennec, A., Couturon, E., et al. (2012). A survey of mangiferin and hydroxycinnamic acid ester accumulation in coffee (Coffea) leaves: biological implications and uses. *Ann. Bot.* 110, 595–613. doi: 10.1093/aob/mcs119

Cangeloni, L., Bonechi, C., Leone, G., Consumi, M., Andreassi, M., Magnani, A., et al. (2022). Characterization of extracts of coffee leaves (Coffea arabica l.) by spectroscopic and chromatographic/spectrometric techniques. *Foods* 11, 2495. doi: 10.3390/FOODS11162495

Carvalho, H. F., da Silva, F. L., de Resende, M. D. V., and Bhering, L. L. (2019). Selection and genetic parameters for interpopulation hybrids between kouilou and robusta coffee. *Bragantia* 78, 52–59. doi: 10.1590/1678-4499.2018124

Castro-Moretti, F. R., Cocuron, J. C., Cia, M. C., Cataldi, T. R., Labate, C. A., Alonso, A. P., et al. (2021). Targeted metabolic profiles of the leaves and xylem sap of two sugarcane genotypes infected with the vascular bacterial pathogen leifsonia xyli subsp. xyli. *Metabolites* 11, 234. doi: 10.3390/metabo11040234

Castro-moretti, F. R., Gentzel, I. N., Mackey, D., and Alonso, A. P. (2020). Metabolomics as an emerging tool for the study of plant–pathogen interactions. *Metabolites* 10 (2), 52. doi: 10.3390/metabo10020052

Chandra Kala, S. (2015). Medicinal attributes of family rubiaceae. *Int. J. Pharm. Biol. Sci.* 5, 179–181. doi: 10.4103/0973-7847.95866

Chen, X. (2019). A review on coffee leaves: Phytochemicals, bioactivities and applications. *Crit. Rev. Food Sci. Nutr.* 59, 1008–1025. doi: 10.1080/10408398.2018.1546667

Chen, X., Ding, J., Ji, D., He, S., and Ma, H. (2020b). Optimization of ultrasonic-assisted extraction conditions for bioactive components from coffee leaves using the taguchi design and response surface methodology. *J. Food Sci.* 85, 1742–1751. doi: 10.1111/1750-3841.15111

Chen, S., Ma, J., Yang, L., Teng, M., Lai, Z. Q., Chen, X., et al. (2020a). Anti-glioblastoma activity of kaempferol *via* programmed cell death induction: involvement of autophagy and pyroptosis. *Front. Bioeng. Biotechnol.* 8. doi: 10.3389/FBIOE.2020.614419/BIBTEX

Chen, Q. C., Zhang, W. Y., Youn, U. J., Kim, H. J., Lee, I. S., Jung, H. J., et al. (2009). Iridoid glycosides from gardeniae fructus for treatment of ankle sprain. *Phytochemistry* 70, 779–784. doi: 10.1016/J.PHYTOCHEM.2009.03.008

Chiocchio, I., Mandrone, M., Tomasi, P., Marincich, L., and Poli, F. (2021). Plant secondary metabolites: An opportunity for circular economy. *Mol* 26, 495. doi: 10.3390/MOLECULES26020495

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 46, W486–W494. doi: 10.1093/nar/gky310

Cocuron, J.-C., Anderson, B., Boyd, A., and Alonso, A. P. (2014). Targeted metabolomics of physaria fendleri, an industrial crop producing hydroxy fatty acids. *Plant Cell Physiol.* 55, 620–633. doi: 10.1093/pcp/pcu011

Cocuron, J.-C., Casas, M. I., Yang, F., Grotewold, E., and Alonso, A. P. (2019). Beyond the wall: high-throughput quantification of plant soluble and cell-wall bound phenolics by liquid chromatography tandem mass spectrometry. *J. Chromatogr. A* 1589, 93–104. doi: 10.1016/J.CHROMA.2018.12.059

Cocuron, J. C., Tsogtbaatar, E., and Alonso, A. P. (2017). High-throughput quantification of the levels and labeling abundance of free amino acids by liquid chromatography tandem mass spectrometry. *J. Chromatogr. A* 1490, 148–155. doi: 10.1016/J.CHROMA.2017.02.028

Conserva, L. M., and Ferreira Júnior, J. C. (2012). Borreria and spermacoce species (Rubiaceae): A review of their ethnomedicinal properties, chemical constituents, and biological activities. *Pharmacogn. Rev.* 6, 46–55. doi: 10.4103/0973-7847.95866

Cragg, G. M., and Newman, D. J. (2013). Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* 1830, 3670–3695. doi: 10.1016/J.BBAGEN.2013.02.008

Davidson, I. W. F., Sumner, D. D., and Parker, J. C. (2008). Chloroform: A review of its metabolism, teratogenic, mutagenic, and carcinogenic potential. *Drug Chem. Toxicol.* 5 (1), 1–87. doi: 10.3109/01480548209017822

de Falco, B., and Lanzotti, V. (2018). NMR spectroscopy and mass spectrometry in metabolomics analysis of salvia. *Phytochem. Rev.* 17, 951–972. doi: 10.1007/S11101-018-9550-8

Du Fall, L. A., and Solomon, P. S. (2011). Role of cereal secondary metabolites involved in mediating the outcome of plant-pathogen interactions. *Metabolites* 1, 64–78. doi: 10.3390/metabo1010064

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. (2012). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* 91 9, 44–66. doi: 10.1007/S11306-012-0434-4

Elansary, H. O., Abdel-Hamid, A. M. E., Yessoufou, K., Al-Mana, F. A., El-Ansary, D. O., Mahmoud, E. A., et al. (2020). Physiological and molecular characterization of water-stressed chrysanthemum under robinin and chitosan treatment. *Acta Physiol. Plant* 42, 1–14. doi: 10.1007/S11738-020-3021-8/FIGURES/12

Felice, M. R., Maugeri, A., De Sarro, G., Navarra, M., and Barreca, D. (2022). Molecular pathways involved in the anti-cancer activity of flavonols: A focus on myricetin and kaempferol. *Int. J. Mol. Sci.* 23, 4411. doi: 10.3390/IJMS23084411

Feuillet, C., Langridge, P., and Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends Genet.* 24, 24–32. doi: 10.1016/J.TIG.2007.11.001

Funlayo, A. A., Adenuga, O. O., Mapayi, E. F., and Olaniyi, O. O. (2017). Coffee: Botany, distribution, diversity, chemical composition and its management. *J. Agric. Vet. Sci.* 10, 57–62. doi: 10.9790/2380-1007035762

Gallego-Giraldo, L., Posé, S., Pattathil, S., Peralta, A. G., Hahn, M. G., Ayre, B. G., et al. (2018). Elicitors and defense gene induction in plants with altered lignin compositions. *New Phytol.* 219, 1235–1251. doi: 10.1111/nph.15258

Guang, C., Chen, J., Sang, S., and Cheng, S. (2014). Biological functionality of soyasaponins and soyasapogenols. *J. Agric. Food Chem.* 62, 8247–8255. doi: 10.1021/JF503047A/ASSET/IMAGES/LARGE/JF-2014-03047A_0001.JPEG

Gulcin, I., Kaya, R., Goren, A. C., Akincioglu, H., Topal, M., Bingol, Z., et al. (2019). Anticholinergic, antidiabetic and antioxidant activities of cinnamon (cinnamomum verum) bark extracts: polyphenol contents analysis by LC-MS/MS. *Int. J. Food Prop.* 22, 1511–1526. doi: 10.1080/10942912.2019.1656232

Hasan, M. M., Bashir, T., and Bae, H. (2017). Use of ultrasonication technology for the increased production of plant secondary metabolites. *Molecules*. 22(7), 1046. doi: 10.3390/molecules22071046.

Hossain, M. B., Tiwari, B. K., Gangopadhyay, N., O'Donnell, C. P., Brunton, N. P., and Rai, D. K. (2014). Ultrasonic extraction of steroidal alkaloids from potato peel waste. *Ultrason. Sonochem.* 21, 1470–1476. doi: 10.1016/J.ULTSONCH.2014.01.023

Iftikhar, A., Ali, S., Yasmeen, T., Arif, M. S., Zubair, M., Rizwan, M., et al. (2019). Effect of gibberellic acid on growth, photosynthesis and antioxidant defense system of wheat under zinc oxide nanoparticle stress. *Environ. pollut.* 254, 113109. doi: 10.1016/J.ENVPOL.2019.113109

Jabir, N. R., Khan, M. S., Alafaleq, N. O., Naz, H., and Ahmed, B. A. (2022). Anticancer potential of yohimbine in drug-resistant oral cancer KB-ChR-8-5 cells. *Moleuclar Biol. Rep.* 49, 9565–9573. doi: 10.1007/s11033-022-07847-7

Jaini, R., Wang, P., Dudareva, N., Chapple, C., and Morgan, J. A. (2017). Targeted metabolomics of the phenylpropanoid pathway in arabidopsis thaliana using reversed phase liquid chromatography coupled with tandem mass spectrometry. *Phytochem. Anal.* 28, 267–276. doi: 10.1002/PCA.2672

Jain, C., Khatana, S., and Vijayvergia, R. (2019). Bioactivity of secondary metabolites of various plants: A review. *Artic. Int. J. Pharm. Sci. Res.* 10, 494–504. doi: 10.13040/IJPSR.0975-8232.10(2).494-04

Jez, J. M., Topp, C. N., Alvarez, S., and Naldrett, M. J. (2021). Mass spectrometry based untargeted metabolomics for plant systems biology. *Emerg. Top. Life Sci.* 5, 189–201. doi: 10.1042/ETLS20200271

Jorge, T. F., Mata, A. T., and António, C. (2016). Mass spectrometry as a quantitative tool in plant metabolomics. *Philos. Trans. R. Soc A Math. Phys. Eng. Sci.* 374 (2079), 20150370. doi: 10.1098/RSTA.2015.0370

Kessler, A., and Kalske, A. (2018). Plant secondary metabolite diversity and species interactions. *Annu. Rev. Ecol. Evol. Syst.* 49, 115–138. doi: 10.1146/annurev-ecolsys-110617-062406

Kim, M., Jung, J., Jeong, N. Y., and Chung, H. J. (2019b). The natural plant flavonoid apigenin is a strong antioxidant that effectively delays peripheral neurodegenerative processes. *Anat. Sci. Int.* 94, 285–294. doi: 10.1007/S12565-019-00486-2/FIGURES/6

Kim, J. H., Kim, M., Kim, J. M., Lee, M. K., Seo, S. J., and Park, K. Y. (2019a). Afzelin suppresses proinflammatory responses in particulate matter-exposed human keratinocytes. *Int. J. Mol. Med.* 43, 2516–2522. doi: 10.3892/IJMM.2019.4162/HTML

Kluska, M., Juszczak, M., Żuchowski, J., Stochmal, A., and Woźniak, K. (2021). Kaempferol and its glycoside derivatives as modulators of etoposide activity in HL-60 cells. *Int. J. Mol. Sci.* 22, 3520. doi: 10.3390/IJMS22073520

Korkina, L. G. (2007). Phenylpropanoids as naturally occurring antioxidants: From plant defense to human health. *Cell. Mol. Biol.* 53, 15–25. doi: 10.1170/T772

Kunert, O., Sreekanth, G., Babu, G. S., Rao, B. V. R. A., Radhakishan, M., Kumar, B. R., et al. (2009). Cycloartane triterpenes from dikamali, the gum resin of gardenia gummifera and gardenia lucida. *Chem. Biodivers.* 6, 1185–1192. doi: 10.1002/CBDV.200800339

Laus, G., and Teppner, H. (1996). The alkaloids of an uncaria rhynchophylla (Rubiaceae-coptosapelteae). *Phyton* (Horn, Austria)36, 185–196.

Lee, K. S., Woo, S. Y., Lee, M. J., Kim, H. Y., Ham, H., Lee, D. J., et al. (2020). Isoflavones and soyasaponins in the germ of Korean soybean [Glycine max (L.) merr.] cultivars and their compound-enhanced BMP-2-induced bone formation. *Appl. Biol. Chem.* 63, 1–8. doi: 10.1186/S13765-020-00508-Y/FIGURES/2

Leroy, T., Ribeyre, F., Bertrand, B., Charmetant, P., Dufour, M., Montagnon, C., et al. (2006). Genetics of coffee quality. *Braz. J. Plant Physiol.* 18, 229–242. doi: 10.1590/S1677-04202006000100016

Lin, Y., Xu, W., Huang, M., Xu, W., Li, H., Ye, M., et al. (2015). Qualitative and quantitative analysis of phenolic acids, flavonoids and iridoid glycosides in yinhua kanggan tablet by UPLC-QqQ-MS/MS. *Mol.* 20, 12209–12228. doi: 10.3390/MOLECULES200712209

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: A review. *Front. Chem.* 6. doi: 10.3389/FCHEM.2018.00052/BIBTEX

Liu, X., Zhou, L., Shi, X., and Xu, G. (2019). New advances in analytical methods for mass spectrometry-based large-scale metabolomics study. *TrAC Trends Anal. Chem.* 121, 115665. doi: 10.1016/J.TRAC.2019.115665

Ma, D., Li, Y., Zhang, J., Wang, C., Qin, H., Ding, H., et al. (2016). Accumulation of phenolic compounds and expression profiles of phenolic acid biosynthesis-related genes in developing grains of white, purple, and red wheat. *Front. Plant Sci.* 7. doi: 10.3389/FPLS.2016.00528/BIBTEX

Marchetti, L., Pellati, F., Graziosi, R., Brighenti, V., Pinetti, D., and Bertelli, D. (2019). Identification and determination of bioactive phenylpropanoid glycosides of aloysia polystachya (Griseb. et moldenke) by HPLC-MS. *J. Pharm. Biomed. Anal.* 166, 364–370. doi: 10.1016/J.JPBA.2019.01.033

Marczak, L., Kawiak, A., Łojkowska, E., and Stobiecki, M. (2005). Secondary metabolites in *in vitro* cultured plants of the genus drosera. *Phytochem. Anal.* 16, 143–149. doi: 10.1002/PCA.833

Martínez, G., Regente, M., Jacobi, S., Del Rio, M., Pinedo, M., and de la Canal, L. (2017). Chlorogenic acid is a fungicide active against phytopathogenic fungi. pestic. *Biochem. Physiol.* 140, 30–35. doi: 10.1016/J.PESTBP.2017.05.012

Martins, D., and Nunez, C. (2015). Secondary metabolites from rubiaceae species. *Molecules* 20, 13422–13495. doi: 10.3390/molecules200713422

Masson, P., Alves, A. C., Ebbels, T. M. D., Nicholson, J. K., and Want, E. J. (2010). Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS. *Anal. Chem.* 82, 7779–7786. doi: 10.1021/ac101722e

Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K., and Saito, K. (2009). MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* 57, 555–577. doi: 10.1111/J.1365-313X.2008.03705.X

McSorley, R., and Phillips, M. S. (1992). Origin, evolution, population genetics and resources for breeding of wild barley, hordeum spontaneum, in the fertile crescent. *Barley Genet. Biochem. Mol. Biol. Biotechnol.*, 19–43.

Millan, A. F. S., Gamir, J., Farran, I., Larraya, L., and Veramendi, J. (2022). Identification of new antifungal metabolites produced by the yeast metschnikowia pulcherrima involved in the biocontrol of postharvest plant pathogenic fungi. *Postharvest Biol. Technol.* 192, 111995. doi: 10.1016/J.POSTHARVBIO.2022.111995

Montis, A., Souard, F., Delporte, C., Stoffelen, P., Stévigny, C., and Van Antwerpen, P. (2022). Targeted and untargeted mass spectrometry-based metabolomics for chemical profiling of three coffee species. *Molecules* 27 (10), 3152. doi: 10.3390/MOLECULES27103152

Nagar, S., Singh, V. P., Arora, A., Dhakar, R., Singh, N., Singh, G. P., et al. (2021). Understanding the role of gibberellic acid and paclobutrazol in terminal heat stress tolerance in wheat. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.692252/BIBTEX

Nakabayashi, R., and Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. *Curr. Opin. Plant Biol.* 24, 10–16. doi: 10.1016/j.pbi.2015.01.003

Natarajan, T., Anandhi, M., Aiswarya, D., Ramkumar, R., Kumar, S., and Perumal, P. (2016). Idaein chloride induced p53 dependent apoptosis in cervical cancer cells through inhibition of viral oncoproteins. *Biochimie* 121, 13–20. doi: 10.1016/J.BIOCHI.2015.11.008

Ndagijimana, A., Wang, X., Pan, G., Zhang, F., Feng, H., and Olaleye, O. (2013). A review on indole alkaloids isolated from uncaria rhynchophylla and their pharmacological studies. *Fitoterapia* 86, 35–47. doi: 10.1016/J.FITOTE.2013.01.018

Orcic, D., Franciškovic, M., Bekvalac, K., Svircev, E., Beara, I., Lesjak, M., et al. (2014). Quantitative determination of plant phenolics in urtica dioica extracts by high-performance liquid chromatography coupled with tandem mass spectrometric detection. *Food Chem.* 143, 48–53. doi: 10.1016/J.FOODCHEM.2013.07.097

Osman, A. G., Haider, S., Chittiboyina, A. G., and Khan, I. A. (2019). Utility of alkaloids as chemical and biomarkers for quality, efficacy, and safety assessment of botanical ingredients. *Phytomedicine* 54, 347–356. doi: 10.1016/J.PHYMED.2018.03.064

Patay, E. B., Fritea, L., Antonescu, A., Antonescu, A., and Dobjanschi, L. (2017). Coffea arabica: A plant with rich content in caffeine. *The Question of Caffeine*. (London, United Kingdom: IntechOpen) doi: 10.5772/intechopen.68149

Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L. (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. biochim. *Biophys. Acta - Gene Regul. Mech.* 1829, 1236–1247. doi: 10.1016/J.BBAGRM.2013.09.006

Patti, G. J., Yanes, O., Siuzadak, G., and Siuzdak, G. (2012). Metabolomics: the apogee of the omics triology. *Nat. Rev. Mol. Cell Biol.* 13, 263–269. doi: 10.1038/nrm3314

Peleg, Z., Fahima, T., Abbo, S., Krugman, T., Nevo, E., Yakir, D., et al. (2005). Genetic diversity for drought resistance in wild emmer wheat and its ecogeographical associations. *Plant Cell Environ.* 28, 176–191. doi: 10.1111/J.1365-3040.2005.01259.X

Perez de Souza, L., Alseekh, S., Naake, T., and Fernie, A. (2019). Mass spectrometry-based untargeted plant metabolomics. *Curr. Protoc. Plant Biol.* 4, e20100. doi: 10.1002/CPPB.20100

Quatrin, A., Pauletto, R., Maurer, L. H., Minuzzi, N., Nichelle, S. M., Carvalho, J. F. C., et al. (2019). Characterization and quantification of tannins, flavonols, anthocyanins and matrix-bound polyphenols from jaboticaba fruit peel: A comparison between myrciaria trunciflora and m. jaboticaba. *J. Food Compos. Anal.* 78, 59–74. doi: 10.1016/J.JFCA.2019.01.018

Roberts, L. D., Souza, A. L., Gerszten, R. E., and Clish, C. B. (2012). Targeted metabolomics. *Curr. Protoc. Mol. Biol.* 98, 30.2.1–30.2.24. doi: 10.1002/0471142727.MB3002S98

Rodrigues, F. A., Carré-Missio, V., Jham, G. N., Berhow, M., and Schurt, D. A. (2011). Chlorogenic acid levels in leaves of coffee plants supplied with silicon and infected by hemileia vastatrix. *Trop. Plant Pathol.* 36, 404–408. doi: 10.1590/S1982-56762011000600010

Roopashree, K. M., and Naik, D. (2019). Advanced method of secondary metabolite extraction and quality analysis. *J. Pharmacogn. Phytochem.* 8, 1829–1842.

Saini, N., Grewal, A. S., Lather, V., and Gahlawat, S. K. (2022). Natural alkaloids targeting EGFR in non-small cell lung cancer: Molecular docking and ADMET predictions. *Chem. Biol. Interact.* 358, 109901. doi: 10.1016/J.CBI.2022.109901

Santamarina, A. B., Pisani, L. P., Baker, E. J., Marat, A. D., Valenzuela, C. A., Miles, E. A., et al. (2021). Anti-inflammatory effects of oleic acid and the anthocyanin keracyanin alone and in combination: effects on monocyte and macrophage responses and the NF-kappa b pathway. *Food Funct.* 12, 7909–7922. doi: 10.1039/D1FO01304A

Sawada, Y., and Yokota Hirai, M. (2013). ). integrated LC-MS/MS system for plant metabolomics. *Comput. Struct. Biotechnol. J.* 4, e201301011. doi: 10.5936/CSBJ.201301011

Sawynok, J. (1995). Pharmacological rationale for the clinical use of caffeine. *Drugs* 49, 37–50. doi: 10.2165/00003495-199549010-00004

Schwechheimer, C., and Willige, B. C. (2009). Shedding light on gibberellic acid signalling. *Curr. Opin. Plant Biol.* 12, 57–62. doi: 10.1016/J.PBI.2008.09.004

Shankar, E., Goel, A., Gupta, K., and Gupta, S. (2017). Plant flavone apigenin: An emerging anticancer agent. *Curr. Pharmacol. Rep.* 3, 423. doi: 10.1007/S40495-017-0113-2

Shimizu, T., Watanabe, M., Fernie, A. R., and Tohge, T. (2018). Targeted LC-MS analysis for plant secondary metabolites. *Methods Mol. Biol.* 1778, 171–181. doi: 10.1007/978-1-4939-7819-9_12/COVER

Silva, G. L., Lee, I.-S., and Kinghorn, A. D. (1998). "Special problems with the extraction of plants," in *Natural products isolation. methods in biotechnology*. Ed. R. J. Cannell (Totowa: Humana Press), 343–363. doi: 10.1007/978-1-59259-256-2_12

Simunkova, M., Barbierikova, Z., Jomova, K., Hudecova, L., Lauro, P., Alwasel, S. H., et al. (2021). Antioxidant vs. prooxidant properties of the flavonoid, kaempferol, in the presence of Cu(II) ions: A ROS-scavenging activity, fenton reaction and DNA damage study. *Int. J. Mol. Sci.* 22 (4), 1619. doi: 10.3390/IJMS22041619

Soares Melecchi, M. I., Péres, V. F., Dariva, C., Zini, C. A., Abad, F. C., Martinez, M. M., et al. (2006). Optimization of the sonication extraction method of hibiscus tiliaceus l. flowers. *Ultrason. Sonochem.* 13, 242–250. doi: 10.1016/J.ULTSONCH.2005.02.003

Somsak, V., Damkaew, A., and Onrak, P. (2018). Antimalarial activity of kaempferol and its combination with chloroquine in plasmodium berghei infection in mice. *J. Pathog.* 2018, 1–7. doi: 10.1155/2018/3912090

Souard, F., Delporte, C., Stoffelen, P., Thévenot, E. A., Noret, N., Dauvergne, B., et al. (2018). Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS. *Food Chem.* 245, 603–612. doi: 10.1016/J.FOODCHEM.2017.10.022

Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., and Gavrilenko, T. (2014). Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Bot. Rev.* 80, 283–383. doi: 10.1007/S12229-014-9146-Y/FIGURES/8

Su, Y., Ma, L., Wen, Y., Wang, H., and Zhang, S. (2014). Studies of the *in vitro* antibacterial activities of several polyphenols against clinical isolates of methicillin-resistant staphylococcus aureus. *Molecules* 19, 12630–12639. doi: 10.3390/molecules190812630

Sun, D., Dong, L., Guo, P., Shi, X., Gao, J., Ren, Y., et al. (2013). Simultaneous detection of flavonoids and phenolic acids in herba lysimachiae and herba desmodii styracifolii using liquid chromatography tandem mass spectrometry. *Food Chem.* 138, 139–147. doi: 10.1016/J.FOODCHEM.2012.09.096

Sung, W. S., and Lee, D. G. (2010). Antifungal action of chlorogenic acid against pathogenic fungi, mediated by membrane disruption. *Pure Appl. Chem.* 82, 219–226. doi: 10.1351/PAC-CON-09-01-08/MACHINEREADABLECITATION/RIS

Tautenhahn, R., Patti, G. J., Rinehart, D., and Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Anal. Chem.* 84, 5035–5039. doi: 10.1021/ac300698c

Valdeira, A. S. C., Darvishi, E., Woldemichael, G. M., Beutler, J. A., Gustafson, K. R., and Salvador, J. A. R. (2019). Madecassic acid derivatives as potential anticancer agents: Synthesis and cytotoxic evaluation. *J. Nat. Prod.* 82, 2094–2105. doi: 10.1021/ACS.JNATPROD.8B00864/ASSET/IMAGES/LARGE/NP-2018-00864R_0003.JPEG

Wang, F., Gong, S., Wang, T., Li, L., Luo, H., Wang, J., et al. (2020). Soyasaponin II protects against acute liver failure through diminishing YB-1 phosphorylation and Nlrp3-inflammasome priming in mice. *Theranostics* 10, 2714–2726. doi: 10.7150/THNO.40128

Wang, Y., Liu, H., Shen, L., Yao, L., Ma, Y., Yu, D., et al. (2015). Isolation and purification of six iridoid glycosides from gardenia jasminoides fruit by medium-pressure liquid chromatography combined with macroporous resin chromatography. *J. Sep. Sci.* 38, 4119–4126. doi: 10.1002/jssc.201500705

Won, J. H., Shin, J. S., Park, H. J., Jung, H. J., Koh, D. J., Jo, B. G., et al. (2010). Anti-inflammatory effects of madecassic acid *via* the suppression of NF-κB pathway in LPS-induced RAW 264.7 macrophage cells. *Planta Med.* 76, 251–257. doi: 10.1055/S-0029-1186142

Xie, S., Shi, Y., Wang, Y., Wu, C., Liu, W., Feng, F., et al. (2013). Systematic identification and quantification of tetracyclic monoterpenoid oxindole alkaloids in uncaria rhynchophylla and their fragmentations in q-TOF-MS spectra. *J. Pharm. Biomed. Anal.* 81–82, 56–64. doi: 10.1016/J.JPBA.2013.03.017

Xie, M., Zhang, J., Tschaplinski, T. J., Tuskan, G. A., Chen, J. G., and Muchero, W. (2018). Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01427

Xu, X., Wang, Y., Wei, Z., Wei, W., Zhao, P., Tong, B., et al. (2017). Madecassic acid, the contributor to the anti-colitis effect of madecassoside, enhances the shift of Th17 toward treg cells *via* the PPARγ/AMPK/ACC1 pathway. *Cell Death Dis.* 838, e2723–e2723. doi: 10.1038/cddis.2017.150

Xu, X., Xie, H., Hao, J., Jiang, Y., and Wei, X. (2011). Flavonoid glycosides from the seeds of litchi chinensis. *J. Agric. Food Chem.* 59, 1205–1209. doi: 10.1021/JF104387Y/ASSET/IMAGES/LARGE/JF-2010-04387Y_0001.JPEG

Yang, L., Peng, K., Zhao, S., Zhao, F., Chen, L., and Qiu, F. (2013). 2-methyl-l-erythritol glycosides from gardenia jasminoides. *Fitoterapia* 89, 126–130. doi: 10.1016/J.FITOTE.2013.05.018

Zhang, X., Bi, Q., Wu, X., Wang, Z., Miao, Y., and Tan, N. (2018). Systematic characterization and quantification of rubiaceae-type cyclopeptides in 20 rubia species by ultra performance liquid chromatography tandem mass spectrometry combined with chemometrics. *J. Chromatogr. A* 1581–1582, 43–54. doi: 10.1016/J.CHROMA.2018.10.049

Zhang, H. F., Yang, X. H., and Wang, Y. (2011). Microwave assisted extraction of secondary metabolites from plants: Current status and future directions. *Trends Food Sci. Technol.* 22, 672–688. doi: 10.1016/J.TIFS.2011.07.003

Zhang, H., Zhang, M., Tao, Y., Wang, G., and Xia, B. (2014). Madecassic acid inhibits the mouse colon cancer growth by inducing apoptosis and immunomodulation. *JBUON* 19, 372–376.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fpls.2022.1056082



OPEN ACCESS

EDITED BY
Ricardo Aroca,
Experimental Station of Zaidín (CSIC), Spain

REVIEWED BY
Esther Ngumbi,
University of Illinois at Urbana-Champaign, United States
Shaikhul Islam,
Bangladesh Agricultural Research Council, Bangladesh

*CORRESPONDENCE
Juliana Velasco de Castro Oliveira
juliana.velasco@lnbr.cnpem.br

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Symbiotic Interactions, a section of the journal Frontiers in Plant Science

RECEIVED 29 September 2022
ACCEPTED 13 December 2022
PUBLISHED 09 February 2023

CITATION
Almeida OAC, de Araujo NO, Mulato ATN, Persinoti GF, Sforça ML, Calderan-Rodrigues MJ and Oliveira JVdC (2023) Bacterial volatile organic compounds (VOCs) promote growth and induce metabolic changes in rice. *Front. Plant Sci.* 13:1056082. doi: 10.3389/fpls.2022.1056082



# Bacterial volatile organic compounds (VOCs) promote growth and induce metabolic changes in rice

Octávio Augusto Costa Almeida[1,2†], Natália Oliveira de Araujo[1,2†], Aline Tieppo Nogueira Mulato[1,2], Gabriela Felix Persinoti[1], Maurício Luís Sforça[3], Maria Juliana Calderan-Rodrigues[4] and Juliana Velasco de Castro Oliveira[1,2*]

[1]Brazilian Biorenewables National Laboratory (LNBR), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Brazil, [2]Graduate Program in Genetics and Molecular Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, Brazil, [3]Brazilian Biosciences National Laboratory, Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Brazil, [4]Group of Metabolic Regulation of Plant Growth, Max Planck Institute of Molecular Plant Physiology, Postdam, Germany

Plant growth-promoting bacteria (PGPB) represent an eco-friendly alternative to reduce the use of chemical products while increasing the productivity of economically important crops. The emission of small gaseous signaling molecules from PGPB named volatile organic compounds (VOCs) has emerged as a promising biotechnological tool to promote biomass accumulation in model plants (especially *Arabidopsis thaliana*) and a few crops, such as tomato, lettuce, and cucumber. Rice (*Oryza sativa*) is the most essential food crop for more than half of the world's population. However, the use of VOCs to improve this crop performance has not yet been investigated. Here, we evaluated the composition and effects of bacterial VOCs on the growth and metabolism of rice. First, we selected bacterial isolates (IAT P4F9 and E.1b) that increased rice dry shoot biomass by up to 83% in co-cultivation assays performed with different durations of time (7 and 12 days). Metabolic profiles of the plants co-cultivated with these isolates and controls (without bacteria and non-promoter bacteria—1003-S-C1) were investigated *via* $^{1}$H nuclear magnetic resonance. The analysis identified metabolites (e.g., amino acids, sugars, and others) with differential abundance between treatments that might play a role in metabolic pathways, such as protein synthesis, signaling, photosynthesis, energy metabolism, and nitrogen assimilation, involved in rice growth promotion. Interestingly, VOCs from IAT P4F9 displayed a more consistent promotion activity and were also able to increase rice dry shoot biomass *in vivo*. Molecular identification by sequencing the 16S rRNA gene of the isolates IAT P4F9 and E.1b showed a higher identity with *Serratia* and *Achromobacter* species, respectively. Lastly, volatilomes of these and two other non-promoter bacteria (1003-S-C1 and *Escherichia coli* DH5α) were

evaluated through headspace solid-phase microextraction coupled with gas chromatography–mass spectrometry. Compounds belonging to different chemical classes, such as benzenoids, ketones, alcohols, sulfide, alkanes, and pyrazines, were identified. One of these VOCs, nonan-2-one, was validated *in vitro* as a bioactive compound capable of promoting rice growth. Although further analyses are necessary to properly elucidate the molecular mechanisms, our results suggest that these two bacterial isolates are potential candidates as sources for bioproducts, contributing to a more sustainable agriculture.

KEYWORDS

**microbial volatile organic compounds, *Oryza sativa*, plant growth promotion, bioactive compounds, metabolomics**

## 1 Introduction

Microbial volatile organic compounds (VOCs), produced by organisms such as bacteria and fungi, are small signaling molecules (<C15) having low molecular masses (<300 Da), high vapor pressure, and low boiling point. Usually, these molecules are in the gas phase (at 25°C temperature/1 atm pressure), as their properties facilitate evaporation. VOCs are normally lipophilic compounds and can readily diffuse through water and gas-filled pores in soil and rhizosphere environments (Effmert et al., 2012; Schmidt et al., 2016; Schulz-Bohm et al., 2017). Therefore, differently from soluble metabolites, which are often involved in short-distance interactions, VOCs can reach and act even at long distances (Tyc et al., 2017; Westhoff et al., 2017; Schulz-Bohm et al., 2018).

Approximately 2,000 VOCs emitted from almost 1,000 bacterial and fungal species were already identified (Lemfack et al., 2018). However, the number of microorganisms investigated so far is still small since $10^{12}$ microbial species are expected to exist on Earth (Locey and Lennon, 2016). Bacterial VOCs belong to different chemical classes including alcohols, ketones, benzenoids, terpenoids, sulfur-containing compounds, alkenes, and others (Schulz and Dickschat, 2007; Peñuelas et al., 2014; Schmidt et al., 2016). The term "volatilome" describes all volatile compounds produced by an organism (Tilocca et al., 2020). Their composition depends on many factors, such as the nutrient medium, pH, aeration, and stage of culture growth. Moreover, some compounds are common in the volatilomes of a whole group of bacteria, but others are specific to particular strains (Schmidt et al., 2015; Hernández-Calderón et al., 2018; Guo et al., 2019).

VOCs are by-products of microbial metabolism, but still play important roles in intra- and inter-kingdom interactions. Bacterial VOCs can have antagonistic effects against other microorganisms (Raza et al., 2016a; Raza et al., 2016b; Freitas et al., 2022), promote their growth (Ryu et al., 2003; Song et al., 2019), and modulate virulence and resistance to antibiotics, biofilm formation, and motility (Effmert et al., 2012; Schulz-Bohm et al., 2017). These compounds can also benefit plants by serving as a direct nutrient source, modulating phytohormone pathways (such as auxin and cytokinin), increasing nutrient absorption, regulating enzyme activities, or even inducing systemic resistance against biotic and abiotic stressors (Bitas et al., 2013; Kanchiswamy et al., 2015; Fincheira and Quiroz, 2018). The effects of bacterial VOCs were mainly investigated in the model *Arabidopsis thaliana*, revealing an increase in total leaf surface area (Ryu et al., 2003), rosette leaf number, flowering time, biomass weight, seed set number (Xie et al., 2009), and root length (Maheshwari et al., 2021). For instance, Zhang et al. (2007) presented that the mechanism of action of VOCs from *Bacillus amyloliquefaciens* GB03 to promote the growth of *A. thaliana* involved the regulation of auxin homeostasis and cell wall expansion. Despite those beneficial actions of VOCs, some growth inhibition caused by volatiles from *Serratia* species has already been reported in *A. thaliana* (Plyuta et al., 2021). This might justify the commercial potentiality of identifying the effects of VOCs emitted by individual bacteria on plant growth promotion.

Few investigations revealed the VOCs' effectiveness in non-model species. Volatiles from *Pseudomonas pseudoalcaligenes* incremented biomass, germination, and drought tolerance in maize (*Zea mays* L.) by inducing changes in the defense system (Yasmin et al., 2021). Heenan-Daly et al. (2021) showed that VOCs from *Bacillus* and *Serratia* isolates controlled the expression of genes involved in photosynthetic activity, defense, and stress response in potato (*Solanum tuberosum*). Moreover, the compounds benzaldehyde, 1,2-benzisothiazol-3 (2 H)-one, and 1,3-butadiene, emitted by two *Bacillus* species, improved tobacco resistance against wilt disease using a dual approach; reducing *Ralstonia solanacearum* growth while

increasing the expression of plant genes involved in the salicylic acid pathway (Tahir et al., 2017a). Although several studies reinforce that VOCs can boost plant growth and health, their mechanisms of action are still poorly understood.

Rice (*Oryza sativa*) is a model monocotyledonous plant and has several specificities in relation to *A. thaliana* (a dicotyledonous) (Izawa and Shimamoto, 1996), such as in the shape of the leaves (Nelissen et al., 2016), branching of root (Pagès, 2016), and seed development (Sreenivasulu and Wobus, 2013). Furthermore, rice represents the main food crop for more than 50% of the world's population, besides playing an important role in animal feed (Parida et al., 2022). The world rice production in 2021/2022 is estimated at 514.07 million tons (United States Department of Agriculture, 2022). Although productivity per hectare has more than doubled since the 1960s, a further doubling will be necessary to feed the world's increasing population by 2050. However, rice production during the past decade has almost reached stagnation (Parida et al., 2022). VOC-producing bacteria are thus a promising eco-friendly alternative to increase its productivity (Bitas et al., 2013; Kanchiswamy et al., 2015; Fincheira and Quiroz, 2018; Brilli et al., 2019; Dias et al., 2021). To the best of our knowledge, there is no report about the growth promotion activity of bacterial VOCs on rice.

In this study, we have investigated the metabolome of rice treated with bacterial VOCs and identified two bacteria species causing several metabolic alterations (such as in the energy metabolism) and inducing rice growth. Furthermore, the compound nonan-2-one was detected in the volatilome and was shown to increase rice shoot biomass *in vitro*. VOCs produced by the strain IAT P4F9 could also promote biomass accumulation *in vivo*. This is the first report about the effects of bacterial VOCs on this important crop. Our findings are of particular significance for the agricultural application of VOCs to promote plant growth.

# 2 Materials and methods

## 2.1 Bacterial isolates and plant material

The bacterial strains from sugarcane and energy cane fields (rhizospheric soil and root), as well as from composting of filter cake, were previously isolated in our laboratory (Supplementary Table 1). We selected 14 isolates from different genera (formerly identified by sequencing of the 16S rRNA V3–V5 region), source, and location (Supplementary Table 1). These bacteria were stored in LB broth supplemented with cryoprotectant solution (25 g $L^{-1}$ gelatin, 50 g $L^{-1}$ lactose, 10 g $L^{-1}$ peptone, and 250 g $L^{-1}$ glycerol) at −80°C. Seeds of rice variety IRGA 424 RI (kindly donated by Corteva Agriscience, Mogi Mirim, Brazil), recommended to be cultivated in irrigated or dry land systems in the southern region of Brazil, were surface-sterilized by soaking in 1% sodium NaClO solution containing 50 µl of Tween-20 emulsifier for 30 min with manual stirring. Seeds were then rinsed eight times with sterile Milli-Q water.

## 2.2 Screening of rice growth-promoting activity

The sterilized rice seeds were placed on Petri dishes (150 × 15 mm) containing half-strength Murashige and Skoog salt (½ MS) medium (Murashige and Skoog, 1962) and 1% agar (pH 5.7) in a plant growth chamber (Fitotron® HGC Weiss Technik) under a 12-h photoperiod (21°C, irradiance of 150 µmol $m^{-2}$ $s^{-1}$/19°C, dark) and a relative humidity of 75%. One day before the co-cultivation experiments, bacterial isolates were grown on LB liquid media for 16 h. The culture was diluted with water to yield $10^8$ CFU $ml^{-1}$ based on optical density. The co-cultivation system consisted of a small Petri dish (49 × 12 mm) containing LB solid media inside a larger Petri dish (150 × 15 mm) filled with ½ MS solid medium (Supplementary Figure 1). Two germinated rice seedlings of 4-day-old with the prophyll emergence from the coleoptile (growth stage S3) (Counce et al., 2000) were transferred to the larger Petri dish and 20 µl of the diluted bacterial inoculum were spread onto the small dish. Rice seedlings cultivated on the same system without bacterial inoculum were used as the control, and co-culture plates inoculated with *Escherichia coli* DH5α were used as the negative control since it does not promote plant growth in several reports (Ryu et al., 2003; Ryu et al., 2005; Kanchiswamy et al., 2015; Tahir et al., 2017b; Tahir et al., 2017c; Rath et al., 2018). Plates were sealed with polyvinyl chloride (PVC) film, positioned to form a ~70° angle to support aerial root growth, and placed in the growth chamber for 7 days.

After this period, the dry weight of the roots and shoots was measured on an analytical balance. Morphological parameters of root architecture (total length, primary and secondary root lengths, surface area, and volume) were acquired by imaging using a scanner STD4800 (Regent Instruments Inc, Quebec, Canada), at 4,800 dpi, and the WinRHIZO Regular 2019a software. The experiment was performed in a complete randomized block design with four replicates (in total eight seedlings). Statistical differences for each evaluated parameter were calculated and compared by analysis of variance (ANOVA) and Tukey *post hoc* test ($p < 0.05$) using AgroEstat v.1.1.0.712 software (Barbosa and Maldonado, 2010). Outlier (one seedling) was removed from each treatment.

New co-cultivation assays were performed to evaluate the growth promotion effects in a longer period. For that, the bacterial isolates that promoted the highest increase in rice growth after 7 days of co-cultivation [IAT P4F9 and E.1b (Xilano01_11), hereafter referred to as the short name "E.1b"] were chosen. The experiments were conducted following the

same methodology, but the rice dry weight was evaluated after 12 days of co-cultivation.

## 2.3 Metabolome analysis

For metabolome analysis, plants were also co-cultivated with 1003-S-C1 (a bacterium that did not promote rice growth—negative control). The experiments were conducted following the same above-mentioned procedures and controls. The dry weight of shoots and roots from eight plants per treatment was evaluated after 12 days as previously described. Shoots dedicated to metabolomic analysis were harvested (15 per replicate), ground in liquid nitrogen with pestle and mortar, and stored at −80°C until metabolite extraction. The experiment was performed with four replicates.

Metabolites were extracted from 100 mg of frozen tissue with 600 µl of methanol (Sigma-Aldrich) and chloroform (Thermo Fisher Scientific) solution [2:1 (v/v) ratio]. After vortexing for 10 s, samples were sonicated for 5 min in a Branson Bransonic® (Emerson) ultrasonic bath, at room temperature. Samples were then kept on ice for 15 min and 300 µl of chloroform plus 300 µl of ice-cold Milli-Q water were added. Samples were homogenized again by vortexing for 10 s and centrifuged at 14,000 rpm for 20 min at 4°C (Eppendorf® 5430R). Afterward, 300 µl of the upper fraction of the three-phase solution was transferred to a new microtube and dried in a refrigerated concentrator (Refrigerated CentriVap Centrifugal Concentrator, Labconco) at 4°C for 48 h to allow complete methanol removal.

The remaining pellets were resuspended in 600 µl of $D_2O$-containing phosphate buffer (0.1 M, pH 7.4) and 0.5 mM of trimethylsilylpropionate ($TMSPd_4$). Samples were transferred to a 5-mm NMR tube for immediate acquisition using an Agilent DD2 500-MHz spectrometer (Agilent Technologies Inc., Santa Clara, CA, USA) equipped with a triple-resonance probe at 25°C. $^1$H-NMR spectra acquisition was performed with 256 scans collected with 32 K data points over a spectral width of 8,000 Hz. 2D NMR $^1$H-$^1$H-TOCSY spectra were acquired using a spectral width of 8,000 Hz and 128 increments with 56 transients of 2k complex points for each free induction decay. 2D NMR $^1$H-$^{13}$C-HSQC spectra were recorded with a spectral width of 8,000 Hz × 25,133 Hz and 128 increments with 60 transients of 2k complex points. In both 2D NMR spectra and $^1$H-NMR spectra, a 1.5-s relaxation delay was incorporated between scans with a continual water pre-saturation radiofrequency (RF) field to eliminate residual water signal.

The metabolites were processed and quantified using NMR Suite software version 8.1 (Chenomx Inc™, Edmonton, AB, Canada). The *Processor* module of this software was used to adjust the spectral phase and for baseline corrections. A 0.5-Hz line-broadening function was used to reduce signal noise and facilitate the fitting of the metabolite signals in spectral peaks. The water signal was suppressed, and the spectra were calibrated using the reference signal of the TMSP-$d_4$ as 0.5 mM. The spectra were individually transferred to the *Profiling* module of this software to determine the metabolomic profile of each sample. Metabolites were identified and their concentrations were measured and exported to Excel® (Microsoft Office™ *365*) and normalized by the fresh weight of the samples. When necessary, the 2D $^1$H-$^1$H-TOCSY/$^1$H-$^{13}$C-HSQC spectra were used to confirm the identity of some metabolites. Afterward, the metabolome data were analyzed using MetaboAnalyst 5.0 (Pang et al., 2021) following these procedures: data were normalized with $Log_{10}$ transformation, statistical analysis was performed by calculating significance with ANOVA and pairwise comparing mean differences with *t*-test ($p < 0.05$), corrected with false discovery rates (FDRs), principal component analysis (PCA), and hierarchical clustering (HCA) dendrogram based on Pearson distance measure and Ward clustering algorithm, and generation of a heatmap using the normalized data without standardization and with Pearson distance measure and Ward clustering algorithm. Outliers detected by the *RandomForest* function were removed from analyses. The module "Pathway analysis" at MetaboAnalyst 5.0 was used to identify which metabolic pathways have been affected. Data were normalized with $Log_{10}$ transformation, and no scaling was used. The parameter settings were as follows: visualization method as scatter plot (testing significant features), global test for the enrichment method, relative-betweenness centrality for the topology analysis, and the pathway library of *Oryza sativa* subsp. *japonica* (Japanese rice) (KEGG) as the reference metabolome. MetaboAnalyst 5.0 scores of a metabolic pathway may have been impacted from 0 to 1.0; thus, we considered those pathways with significance ($p < 0.05$) and an impact value higher than 0.1.

## 2.4 *In vivo* rice and bacteria co-cultivation

Bacteria and rice co-cultivation in a semi-open system were adapted from Park et al. (2015) (Supplementary Figure 2). Five-day-old rice seedlings (decontamination and germination as described in *Section 2.1*) were individually placed in a sterile plastic cup (pre-immersed in a 1% NaClO solution) filled with 70 ml of fine sterile vermiculite. About 10 holes were made at the bottom of the cups (protected with a piece of drainage blanket) to allow the bacterial VOCs to permeate through the substrate. The cups were then inserted into 115-ml glass bottles containing 5 ml of LB solid medium inoculated with 20 µl of $10^8$ CFU of rice growth promoter bacterial isolates IAT P4F9, E.1b, or 1003-S-C1 (negative control). Additional controls were co-cultivation systems with LB medium without bacterial inoculum and inoculated with the growth-promoting bacterium *B. amyloliquefaciens* GB03 (positive control) (Ryu et al., 2004;

Xie et al., 2009; Kwon et al., 2010; Delaplace et al., 2015). Rice was cultivated in a growth chamber under a 12-h photoperiod (as described in *Section 2.2*) for 15 days. As we used the rice cultivation system in dry soil, that period was established considering the rice growth limit stages (V3/V4) that precede the recommendation to flooding the field in the Brazilian context (SOSBAI, 2018). Bacterial inoculums were changed every 4 days of co-cultivation to maintain constant levels of VOCs within the system. During the experiment, irrigation was performed daily with 5 ml of sterile distilled water, and plants received 5 ml of Hoagland solution (Hoagland and Arnon, 1950) every 4 days. The dry shoot and root biomass of four plants per treatment were evaluated as described in *Section 2.2*.

## 2.5 Bacterial isolate identification

Improved identification of the best promoter bacteria (ITA P4F9 and E.1b) and the 1003-S-C1 (non-growth-promoting isolate) were performed according to Freitas et al. (2022). Briefly, the bacterial 16S rRNA gene was amplified using the combination of the primers 20F (5′ GAGTTTGAT CCTGGCTCAG 3′) and 1500R (5′ GTTACCTTGTTACG ACTT 3′) and sequenced using 20F, 1500R, 520F (5′ CAGCAGCCG CGGTAATAC 3′), 520R (5′ GTATTACC GCGGCTGCTG 3′), 920F (5′ AAACTCAAATGAATTG ACGG 3′), and 920R (5′ CCGTCAATTCATTTGAGTTT 3′) (Yukphan et al., 2004). The obtained contigs were compared with the EzTaxon database (Yoon et al., 2017). All 16S rRNA sequences were deposited in NCBI under accession numbers OP186019 [E.1b(Xilano01_11)], OP186020 (IAT P4F9), and OP186021 (1003-S-C1). In addition, phylogenetic analysis was performed. The 16S rRNA gene sequences of IAT P4F9, 1003-S-C1, and E.1b bacterial isolates were submitted to the online tool 16S-based ID available at the EzBioCloud server (Yoon et al., 2017) to identify 16S rRNA sequences of closely related taxa. These sequences were obtained from either the GenBank database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) or the EzBioCloud server. In addition, the *rpoB* gene of IAT P4F9 and 1003-S-C1 was amplified and sequenced using the primers CM7 (5' AACCAGTTCCGCGTTGGCCTGG 3') and CM31 (5' CCTGAACAACACGCTCGGA 3') (Mollet et al., 1997), respectively. The *rpoB* sequences were deposited in NCBI under accession numbers OP893533 (IAT P4F9) and OP893532 (1003-S-C1). Multiple sequence alignments of each marker were performed using Mafft v7.310 with parameters –maxiterate 1000 and –localpair (Katoh and Standley, 2013). Maximum likelihood phylogenetic trees were estimated using IQTree software v 2.0.3 using a supermatrix combining the 16S rRNA gene and *rpoB* alignments with 1000 Fastboostrap option. The evolutionary model TIM3+F+I+G4 was chosen according to BIC criteria for the IAT P4F9 and 1003-S-C1 strains. For the E.1b strain the 16S gene sequences were used and the model TN+F+I+G4 was chosen also based on BIC criteria. Phylogenetic trees were analyzed and visualized using the Itol web tool (Letunic and Bork, 2021).

## 2.6 Volatilome analysis

Identification of volatiles emitted from the bacteria identified in *Section 2.5*, as well as from the *E. coli* DH5α strain, was performed following Freitas et al. (2022). Succinctly, volatiles were extracted from bacterial cultures (five replicates per strain) in LB solid medium by the solid phase microextraction (SPME) method using a DVB/CAR/PDMS fiber (Supelco®, Sigma Aldrich). The fiber was conditioned at 250°C for 5 min in the hot injector and exposed in culture vials at 50°C for 30 min. After extraction, the SPME fiber was directly inserted into the front inlet of a gas chromatograph (Agilent 7890A, Agilent Technologies) connected to a mass spectrometer (Pegasus® HT TOFMS, LECO Corporation) and desorbed at 250°C for 5 min. The chromatography column was the DB5 (30 m × 0.25 mm × 0.25 µm) (Agilent Technologies) and the temperature of the oven was 30°C for 2 min, 30–300°C at a rate of 20°C $min^{-1}$, 300°C for 2 min, followed by a ramp of 20°C $min^{-1}$ until 330°C for 30 s. NIST v.17 (NIST/EPA/NIH Mass Spectral Library) and an in-house library were used to identify volatiles. The volatilomes were analyzed using MetaboAnalyst v5.0 (Pang et al., 2021) by following the same procedures described in *Section 2.3*, except for the statistical analysis, which was performed with ANOVA and comparing mean differences with Tukey's test ($p < 0.05$).

## 2.7 *In vitro* validation of VOCs' effects on rice

Some of the identified VOCs able to promote rice growth were validated using a co-cultivation system prepared as described in *Sections 2.1* and *2.2*. Instead of the bacterial inoculum, solutions with commercial synthetic compounds were applied onto a sterile qualitative filter paper placed in the smaller compartment. Six VOCs [(methyldisulfanyl)methane, (methyltrisulfanyl)methane, heptan-2-one, nonan-2-ol, nonan-2-one, and undecane-2-one] were selected based on the abundance, frequency, and chemical classes of compounds from the best promoter bacteria (ITA P4F9 and E.1b) volatilome profiles on. Compounds were tested at doses of 1.0, 0.1, and 0.01 mg diluted in ethanol, at a final volume of 20 µl. Two mixed solutions were also tested containing three VOCs each: the first one with (methyltrisulfanyl)methane (0.1 mg), nonan-2-one (0.1 mg), and undecane-2-one (0.1 mg), and the second one with (methyldisulfanyl)methane (1.0 mg), heptan-2-one (0.1 mg), and nonan-2-ol (0.1 mg). As a negative control, 20 µl of pure ethanol were used. After 12 days of co-cultivation in

the growth chamber under the same conditions described in *Section 2.2*, the shoot and root dry biomass of eight plants per treatment were measured by an analytical balance and evaluated as described in *Section 2.2*.

# 3 Results

## 3.1 Bacterial VOCs increased rice dry shoot biomass

Dry shoot biomass of rice co-cultivated for 7 days with the isolates IAT P4F9, E.1b, BNG P6F12, and 0277-S-C1 significantly increased by up to 39% (Tukey test, $p < 0.05$) when compared to plants grown in the absence of bacteria (control) and plants co-cultivated with *E. coli* DH5α (negative control) (Figure 1A). Contrastingly, rice seedlings co-cultivated with the isolate MTS P5D6 showed reduced dry shoot biomass by 40% (Figure 1A) and decreased root length, surface area, volume, and axial root length (Supplementary Figure 3).

Among the best isolates, IAT P4F9 and E.1b were selected to be evaluated in a 12-day co-cultivation assay. Plants co-cultivated with IAT P4F9 and E.1b had their dry shoot biomass significantly increased by 83% and 61%, respectively (Figure 1B). Curiously, plants co-cultivated with *E. coli* DH5α showed an increase in dry shoot (Figure 1B) and root biomass (Figure 1C) of 53% and 29%, respectively, and in the root morphological parameters (Supplementary Figure 4).

## 3.2 VOCs from E.1b and IAT P4F9 altered the rice metabolism of amino acids, citric cycle intermediates, carbohydrates, and lipids

To evaluate the effects of bacterial VOCs on plant metabolism, the isolate 1003-S-C1 was included as a negative control since VOCs from *E. coli* DH5α promoted rice growth after 12 days (Figures 1B, C). Two samples, one belonging to 1003-S-C1 and the other to E.1b treatments, were excluded from further analyses as they were confirmed to be outliers. A total of 41 non-volatile metabolites were identified (Supplementary Table 2). Fold change values from statistical analysis (*t*-test, $p < 0.05$) revealed that the concentrations of 33 metabolites were

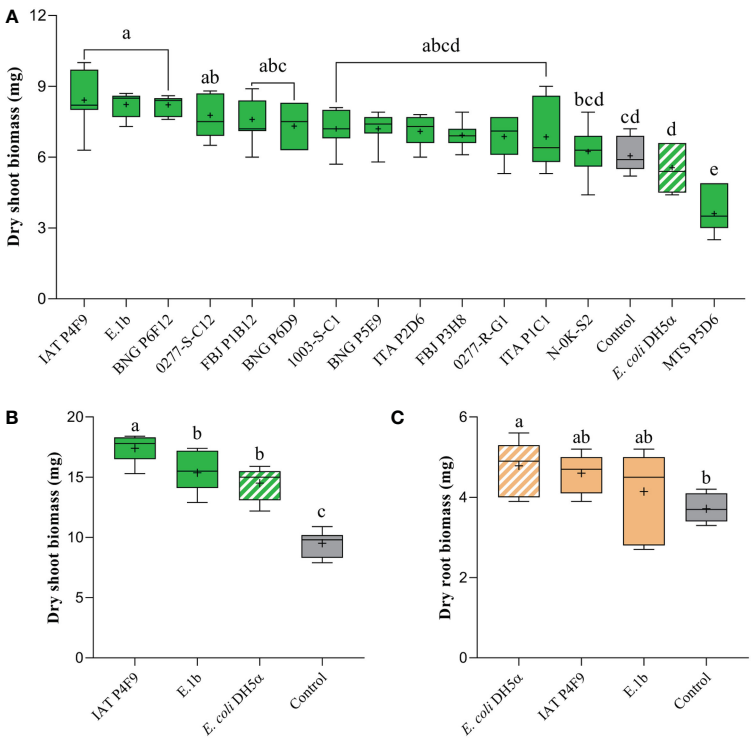



FIGURE 1
Effects of bacterial VOCs on **(A)** dry shoot biomass of rice plants, after 7 days of co-cultivation, and on **(B)** dry shoot and **(C)** root biomass, after 12 days of co-cultivation. Plants were co-cultivated with the bacterial isolates (full colored), the negative control *E. coli* DH5α (white striped), and grown without bacteria (control, gray). Significant differences (ANOVA followed by Tukey's test, $p < 0.05$) among treatments are indicated by letters ($n = 7$).

10.3389/fpls.2022.1056082

significantly different in at least one treatment comparison (Supplementary Table 3).

We performed HCA and PCA analyses to evaluate the general alterations on rice metabolomes due to VOCs (Supplementary Figure 5). On HCA, two main branches were formed separating the metabolome of plants co-cultivated with 1003-S-C1 from the others, and a sub-branch separated plants co-cultivated with IAT P4F9 and E.1b from control plants. PCA loadings (Supplementary Table 4) showed the metabolites that contributed most to the separation (e.g., 2-hydroxyisocaproate, 2-hydroxy-3-methylvalerate, lysine, 4-aminobutyrate, glutamine, asparagine, caprate, and tryptophan).

The heatmap of metabolic data (Figure 2) showed that several amino acids were upregulated in the plants co-cultivated with the two promoter strains, compared with one or both controls (Supplementary Table 3). Arginine, asparagine, leucine, and lysine increased to a similar extent in rice co-cultivated with IAT P4F9 and E.1b, while the amino glutamine, glycine, and valine were significantly more abundant in IAT P4F9 compared with E.1b-treated plants. The sugars sucrose and the hexoses derived from its breakdown, glucose and fructose, besides fucose, displayed contrasting behavior in plants co-cultivated with both promoter strains: reduced in E.1b while increased or unchanged in IAT P4F9, compared with one or both controls. Rice grown with IAT P4F9 presented a higher abundance of the tricarboxylic acid (TCA) cycle intermediate malate and the lipid-related caprate and glycerol, compared with E.1b treatment.

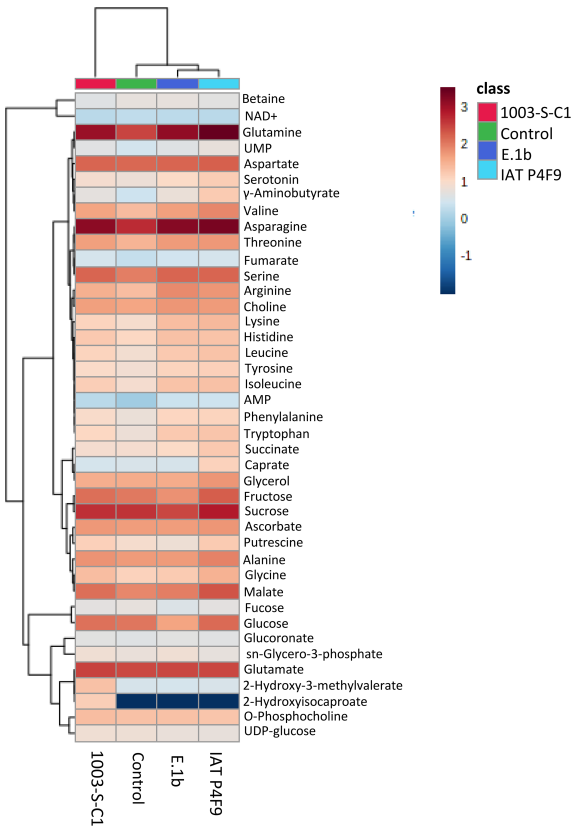



FIGURE 2
Heatmap of the metabolome of rice plants co-cultivated with the bacterial isolates (E.1b, IAT P4F9, and 1003-S-C1) and control plants (cultivated without bacteria). Columns represent each treatment and rows represent the different metabolites identified. The color code indicates the abundance of each metabolite (blue = low abundance; red = high abundance) based on the average of the $Log_{10}$-transformed values (concentrations determined based on the reference signal of 0.5 mM TMSP-$d_4$) ($n$ = 4). The darkest blue indicates that the metabolite was not identified in plants subjected to the respective treatment.

Thus, 16 and 15 metabolic pathways had impact values above the threshold in plants co-cultivated with the isolates IAT P4F9 and E.1b, respectively, compared with the control ($p < 0.05$) (Table 1). Curiously, by comparing their metabolomes with the metabolome of the negative control, 13 and 5 pathways were identified, all commonly detected in the comparisons with the control plants (except the "glycerophospholipid metabolism" pathway). In agreement with the pairwise comparison, co-cultivation with both promoter strains affected the plant pathways "Arginine and proline metabolism", "Arginine biosynthesis", "Glycine serine and threonine metabolism", "Glyoxylate and dicarboxylate metabolism", and "Starch and sucrose metabolism", compared with the control and the negative control. Interestingly, IAT P4F9, the strain that promoted the highest rice growth, also impacted "Alanine aspartate and glutamate metabolism", "Butanoate metabolism", "Citrate cycle (TCA cycle)", "Pyrimidine metabolism", "Pyruvate metabolism", "Tryptophan metabolism", and "Isoquinoline alkaloid biosynthesis".

## 3.3 *In vivo* validation of rice growth promotion mediated by bacterial VOCs

To evaluate if the growth promotion effects of VOCs produced by the isolates IAT P4F9 and E.1b were replicable *in vivo*, a new experiment was performed in a semi-open system where the plants were grown directly on the substrate. After 15 days, dry shoot biomass of plants co-cultivated with the isolate IAT P4F9 increased by 87% when compared to control plants, but no significant increase was observed in plants co-cultivated with isolates E.1b and 1003-S-C1 and with the strain *B. amyloliquefaciens* GB03 (Supplementary Figure 6).

## 3.4 Molecular identification of the best growth promoter isolates

The identification based on the 16S EzBiocloud database showed that the isolate IAT P4F9 presented 99.86% similarity

10.3389/fpls.2022.1056082

TABLE 1 Metabolic pathways affected in rice co-cultivated with IAT P4F9 and E.1b, compared to both controls.

| Pathway | | E.1b x CTL | | E.1b x 1003 | | IAT x CTL | | IAT x 1003 | |
|---|---|---|---|---|---|---|---|---|---|
| | | *p*-value | Impact value | *p*-value | Impact value | *p*-value | Impact value | *p*-value | Impact value |
| 1 | Alanine aspartate and glutamate metabolism | 4.07E–05 | 0.78 | | | 4.26E–06 | 0.78 | 0.001786 | 0.78 |
| 2 | Amino sugar and nucleotide sugar metabolism | 0.013033 | 0.14 | | | | | | |
| **3** | **Arginine and proline metabolism** | **3.56E–05** | **0.26** | **0.009642** | **0.26** | **0.00014547** | **0.26** | **0.008907** | **0.26** |
| **4** | **Arginine biosynthesis** | **2.40E–06** | **0.17** | **0.00023** | **0.17** | **2.47E–06** | **0.17** | **0.002125** | **0.17** |
| 5 | Butanoate metabolism | 9.33E–05 | 0.14 | | | 0.00026374 | 0.14 | 0.007963 | 0.14 |
| 6 | Citrate cycle (TCA cycle) | 0.037438 | 0.10 | | | 0.00065438 | 0.10 | 0.014002 | 0.10 |
| 7 | Glutathione metabolism | | | | | 0.00064237 | 0.12 | | |
| 8 | Glycerophospholipid metabolism | | | | | | | 0.007264 | 0.11 |
| **9** | **Glycine serine and threonine metabolism** | **9.51E–05** | **0.33** | **0.011322** | **0.33** | **3.08E–05** | **0.33** | **0.022301** | **0.33** |
| **10** | **Glyoxylate and dicarboxylate metabolism** | **9.85E–05** | **0.21** | **0.017527** | **0.21** | **1.52E–06** | **0.21** | **0.001638** | **0.21** |
| 11 | Isoquinoline alkaloid biosynthesis | 0.000411 | 0.50 | | | 9.18E–05 | 0.50 | 0.025039 | 0.50 |
| 12 | Nicotinate and nicotinamide metabolism | | | | | 0.015218 | 0.22 | | |
| 13 | Phenylalanine metabolism | 0.000754 | 0.47 | | | 0.0009957 | 0.47 | | |
| 14 | Pyrimidine metabolism | 0.000438 | 0.13 | | | 1.58E–06 | 0.13 | 0.000108 | 0.13 |
| 15 | Pyruvate metabolism | 0.018797 | 0.15 | | | 0.00016627 | 0.15 | 0.013085 | 0.15 |
| **16** | **Starch and sucrose metabolism** | **0.002559** | **0.16** | **0.022911** | **0.16** | **0.0035906** | **0.16** | **0.030246** | **0.16** |
| 17 | Tryptophan metabolism | 4.27E–05 | 0.14 | | | 0.0001443 | 0.14 | 0.017431 | 0.14 |
| 18 | Tyrosine metabolism | 4.09E–05 | 0.22 | | | 0.00022762 | 0.22 | | |

IAT, IAT P4F9; 1003, 1003-S-C1; CTL, Control (without bacteria).
Bold letters show commonly affected pathways.

with *Serratia marcescens* ATCC 13880(T), followed by *S. nematodiphila* DSM 21420(T) (99.72%), and *S. ureilytica* NiVa 51 (T) (98.39%). For E.1b, the highest similarity was with *Achromobacter insuavis* LMG 26845(T) (99.93%), followed by *A. ruhlandii* LMG 1866(T) (99.72%), and *A. aegrifaciens* LMG 26852(T) (99.72%). The best similarity hits to the bacterium isolate used as a negative control, 1003-S-C1, was with *Enterobacter cancerogenus* ATCC 33241(T) (99.93%), followed by *E. huaxiensis* 090008(T) (99.85%) and *E. bugandensis* EB-247 (T) (99.78%). The best hit strains for all bacteria are presented in Supplementary Table 5, as well as the similarity information. In addition, the phylogenetic tree showed that IAT P4F9 is closest to *S. bockelmannii* and 1003-S-C1 is closest to *E. asburiae* (Supplementary Figure 7). Lastly, E.1b grouped with *Achromobacter* strains.

## 3.5 Volatilome profiles are species-specific

Volatilome profiling of the isolates IAT P4F9, E.1b, and 1003-S-C1, as well of *E. coli* DH5α (since this strain unexpectedly induced the plant growth at 12 days) revealed a total of 55 VOCs (Supplementary Table 6) classified into 17 chemical classes. The largest numbers of volatiles were produced by E.1b and IAT P4F9 isolates (37 and 28 compounds, respectively), whereas in the volatilome of *E. coli* DH5α, only eight VOCs were identified (Supplementary Figure 8). From the total VOCs detected, only four (3-methylbutan-1-ol, methanethiol, 2-phenylethanol, and nonan-2-one) were commonly identified among all volatilomes (Supplementary Figure 8). The growth promoter isolates exclusively produced 21 and 9 VOCs, respectively, and five

VOCs were shared [heptan-2-one, (methyldisulfanyl)methane, undecane-2-one, toluene, and 1-(2-aminophenyl)ethanone]. The isolate 1003-S-C1 exclusively produced 3-methylbutyl acetate, 2-phenylacetaldehyde, hexadecanoic acid, and nonadecane.

The VOC profiles of the negative controls were more similar to each other than to the growth promoter isolates, and the E.1b volatilome was the most distinct among all (Supplementary Figure 9). PCA scores (Supplementary Table 7) established the VOCs that most explained the variation among isolates, such as phenylmethanol, (methyldisulfanyl)methane, (methyltrisulfanyl) methane, 2,4,6-trimethyl pyridine, anisole, methylsulfanylethane, phenol, butan-2-one, and undecan-2-one.

Although some VOCs were commonly produced by all bacteria, their abundance differed among them (Figure 3 and Supplementary Table 6). For instance, nonan-2-one was produced in higher concentrations by the growth promoter isolates, while 2-phenylethanol was more abundant in the volatilomes of the negative controls (1003-S-C1 and *E. coli* DH5α).

### 3.6 *In vitro* validation of VOCs as growth promoters

The compound nonan-2-one (0.1 mg) induced a significant increase of 26% (Tukey test, $p < 0.05$) in dry shoot weight (Supplementary Figure 10) when compared to plants grown without synthetic VOCs (control). However, the compound (methyltrisulfanyl)methane [also known as dimethyl trisulfide (DMTS)] in the highest concentration caused a decrease in dry root weight by 41% when compared to control plants (Supplementary Figure 10). None of the other doses, compounds, and mixtures showed significant changes in shoot and root biomass (data no shown).

## 4 Discussion

### 4.1 Bacterial isolate VOCs promote rice growth

Plant growth promotion mediated by bacterial VOCs has been studied since 2003, when two *Bacillus* strains were demonstrated to promote *Arabidopsis* growth when cultivated on divided Petri dishes where only airborne signals could be exchanged between bacteria and plants (Ryu et al., 2003). After that, several other studies were performed showing the role of these volatile molecules as growth inducers (Bailly et al., 2014; Bitas et al., 2015; Piechulla et al., 2017; Tahir et al., 2017b; Jiang et al., 2019). Nevertheless, just a few studies were conducted with species of economic interest, especially monocots. To the best of our knowledge, no study was performed on rice so far. Our results showed that rice growth can be promoted by bacterial VOCs, with potential implications for agricultural food production.

Four of our bacterial isolates were capable of increasing plant dry shoot biomass up to 39% after 7 days of co-cultivation, when compared to plants cultivated without bacterium (control) (Figure 1). These results are in line with previous studies on other monocots showing that fresh shoot biomass of sorghum (*Sorghum bicolor*) was increased by 67% in co-cultivation with *Arthrobacter agilis* UMCV2 (Castulo-Rubio et al., 2015). Also, an 81% increase in total dry biomass of *Brachypodium distachyon* (L.) Beauv. induced by VOCs emitted by *B. amyloliquefaciens* GB03 was reported by Delaplace et al. (2015). As the induction of these phenotypes could be influenced by the exposure period to VOCs (Xie et al., 2009; Zou et al., 2010; Velázquez-Becerra et al., 2011), we re-evaluated

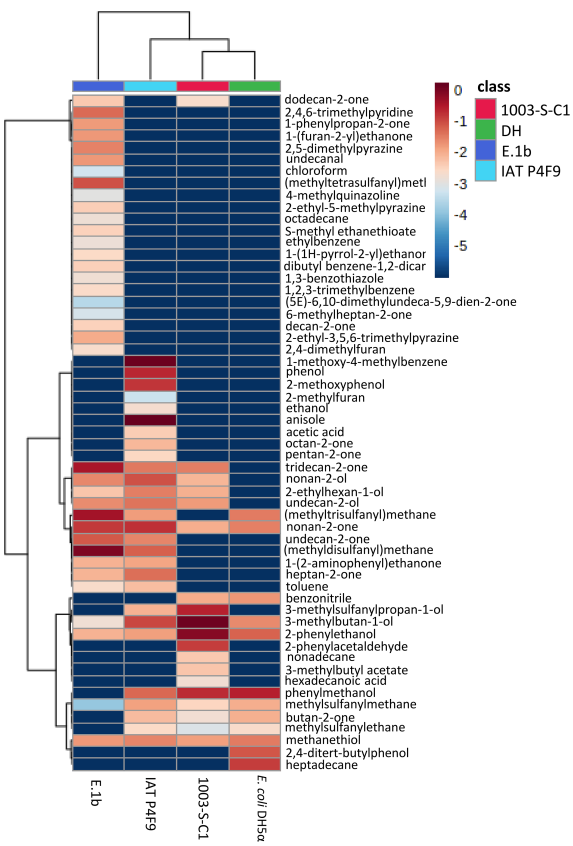



FIGURE 3
Heatmap of VOCs produced by the bacterial isolates (E.1b, IAT P4F9, and 1003-S-C1) and *E. coli* DH5α grown in LB medium. Columns represent each bacterium and rows represent the different VOCs detected. The color code indicates the abundance of each compound (blue = low abundance; red = high abundance), based on the average of the $Log_{10}$-transformed values (peak areas were normalized by cis-3-Hexenyl acetate) ($n = 5$). The darkest blue indicates that the VOCs were not identified in the volatilome of the respective bacterium.

the two isolates that promoted the highest rice growth, IAT P4F9 and E.1b, in a 12-day co-cultivation assay. Plants co-cultivated with both isolates had their dry shoot biomass augmented by 83% and 61%, respectively, in comparison to control plants, indicating that rice growth promotion increased over time. The first growth promotion effects (fresh shoot biomass) on *A. thaliana*, triggered by VOCs from *B. amyloliquefaciens* GB03, significantly started on the sixth day of co-cultivation and were observed until the 19th day, reaching an increase of nearly 260% (Zou et al., 2010). Therefore, it would be interesting to evaluate the long-term effect of these bacterial VOCs throughout the rice life cycle.

Curiously, the strain *E. coli* DH5α promoted rice plant growth after 12 days of co-cultivation. This strain was initially chosen as a negative control because it was not able to promote the growth of plants in previous studies (Ryu et al., 2003; Ryu et al., 2005; Kanchiswamy et al., 2015; Tahir et al., 2017b; Tahir et al., 2017c; Rath et al., 2018). However, Bailly et al. (2014) reported that VOCs emitted by *E. coli* DH5α induced an increase in biomass, secondary roots, and root hair length of *A. thaliana*. In the *Section 4.5*, we discuss which VOCs might be involved in rice growth promotion induced by this strain.

## 4.2 VOCs from IAT P4F9 and E.1b induce distinct metabolic changes in rice plants

By using $^{1}$H NMR, we characterized the metabolic profile of rice co-cultivated with the isolates and the control. The four metabolomes showed distinct metabolic profiles (Figure 2). Among the identified compounds, primary metabolites were the most abundant species, such as amino acids (44%), carbohydrates (15%), and organic acids (7%) (Supplementary Table 2). These compounds are the main intermediates of the plant's central metabolism, including cellular respiration, energy demand, storage, and cell division (Fernie and Pichersky, 2015).

Amino acids are considered building blocks for other metabolites and cellular components, such as nucleotides, nitrogen compounds, chlorophylls, and proteins (Hildebrandt et al., 2015). Interestingly, changes in the concentration of up to 16 amino acids were observed in growth-promoted treatments when compared with the controls. Kang et al. (2015) showed that the abundance of 17 amino acids increased in cucumber treated with *Enterobacter* sp. SE992 inoculated in soil in comparison to control plants. In growing plants, the high demand for protein translation is provided by upregulating amino acid biosynthesis (Hildebrandt et al., 2015). The higher abundance of arginine, asparagine, leucine, and lysine in both promoter treatments compared to controls may denote common pathways by which rice growth is enhanced. Besides the classic role in protein synthesis, these aspartate-derived or related amino acids may act on glycolysis, starch regulation, lipid and nucleotide metabolism, defense response, photorespiration, and nitrogen metabolism (reviewed by Yang et al., 2020).

The efficient utilization of nitrogen considerably impacts plant biomass and yield (Yousaf et al., 2021). Arginine, asparagine, and glutamine are considered nitrogen carriers that store or transport this nutrient through plant vascular bundle (Miflin and Habash, 2002; Lea et al., 2007; Hildebrandt et al., 2015). The high levels of arginine, asparagine, and glutamine can indicate more efficient nitrogen assimilation and storage mediated by the promoter bacterial isolates, at a higher magnitude in IAT P4F9. However, further experiments are necessary to validate this hypothesis. Moreover, the higher abundance of specific amino acids in rice plants could be not only a consequence of improved translation but also an indication that other pathways regulating plant growth, such as nitrogen metabolism, plant immunity, and hormone synthesis, can be stimulated by VOCs.

Carbohydrates are the products of photosynthesis, serving as plant energy source and storage, as well as precursors of structural components for the cell walls (Flowers, 1981; Buchanan et al., 2015). Interestingly, the abundance of fructose, fucose, glucose, and sucrose differed in each treatment. The higher levels of these sugars, especially sucrose, might suggest that photosynthesis was upregulated in plants co-cultivated with IAT P4F9, while the mechanism induced by E.1b seems to be less clear, as this isolate displayed an opposite effect on sugars. Such results may indicate that carbohydrate metabolism plays an important role in the molecular responses resulting from rice growth promotion induced by the bacterial VOCs.

The TCA intermediates malate, succinate, and fumarate were up to 1.52, 1.19, and 0.60-fold more abundant, respectively, in rice co-cultivated with IAT P4F9 compared to control and/or negative control plants. These organic acids are derived from the oxidation of carbohydrates produced by photosynthesis and participate in several metabolic pathways, such as respiration/energy metabolism, catabolism of amino acids, and even plant tolerance to stresses (Abadía et al., 2002; Zell et al., 2010; Hildebrandt et al., 2015). Interestingly, "citrate cycle" and "pyruvate metabolism" were also pointed out as affected pathways in plants co-cultivated with both promoter strains. Together, the levels of sucrose and TCA intermediates show that the process of transforming sugars into energy might be more active when rice was co-cultivated with IAT P4F9. Moreover, the higher amino acid contents in these plants might be an indicator of augmented translation, which could ultimately lead to higher growth and biomass.

Our study showed for the first time that microbial VOCs induce distinct metabolic changes in rice. Although the metabolic profiles cannot solely explain the molecular responses involved in growth promotion, they shed a light on possible pathways and targets for further investigations. Moreover, other approaches (e.g., transcriptomics, proteomics,

and mutant evaluation) are needed to elucidate the molecular and metabolic mechanisms involved in bacterial VOC-mediated plant growth.

## 4.3 VOCs from IAT P4F9 also promotes rice growth *in vivo*

Although the *in vitro* co-cultivation system is a practical method to screen bacteria as potential PGPB, the bacterial VOCs might be concentrated in the Petri dishes generating a response that would not occur in a natural environment. Thus, the isolates IAT P4F9 and E.1b were also evaluated in a semi-open system adapted from Park et al. (2015). Surprisingly, only VOCs emitted by IAT P4F9 promoted the growth of rice plants, which showed an 87% increase in shoot dry biomass. This was similar to the results from the *in vitro* experiment results (83% increase in shoot dry biomass in plants co-cultivated with IAT P4F9 compared to control plants). Although E.1b VOCs resulted in a 61% increase *in vitro*, this isolate did not affect plant growth *in vivo*. In the semi-open system used by Park et al. (2015), tobacco plants (*N. tabacum* cv. Xanthi-nc) co-cultivated with *P. fluorescens* SS101 had an increase of 150% in fresh biomass compared to control plants. Furthermore, an increase of 198% in dry biomass was observed in tomato plants co-cultivated with *Bacillus subtilis* SYST2 for 30 days *in vivo* (Tahir et al., 2017b). As for cucumber plants (*Cucumis sativa* L. cv backdadagi) cultivated in a greenhouse, their fresh shoot and root biomass increased by 57% and 30%, respectively, after 14 days in contact with VOCs emitted by *B. amyloliquefaciens* GB03 (Song et al., 2019). VOCs from *B. amyloliquefaciens* GB03 had no effect on rice plants in our experiment *in vivo*, strengthening the specificity of the interaction between plant and bacteria. Our results demonstrate for the first time that VOCs can promote growth of a monocot plant in semi-open systems. This reinforces their potential for field use, although it does not eliminate the need for further investigations on the long-term effects of bacterial volatiles on rice growing in paddy soil, approaching the ability of VOCs to diffuse into the water.

## 4.4 Bacterial isolates belong to different genera

Since we observed that the isolates IAT P4F9 and E.1b efficiently promoted rice growth, we investigated the taxonomic assignment of these strains by sequencing the full-length 16S rRNA and performed phylogenetic analysis using the 16S rRNA and *rpoB* nucleotide sequences for IAT P4F9 and the 16S rRNA sequence for the E.1b strain. According to EzBiocloud, the high identity of IAT P4F9 was with *S. marcescens* ATCC 13880(T) (98.86%). Phylogenetic analysis also supports the taxonomic assignment to *Serratia* genus, a γ-proteobacterium. Previous studies have reported antagonistic effects on *A. thaliana* growth caused by *Serratia* species (Kai et al., 2008; Blom et al., 2011a; Blom et al., 2011b; Heenan-Daly et al., 2021; Plyuta et al., 2021). Blom et al. (2011a) showed that *S. marcescens, S. entomophilia*, and *S. proteamaculans* were able to promote *A. thaliana* growth, but *S. plymuthica* could trigger positive or negative effects. In addition, Plyuta et al. (2021) showed that *S. proteamaculans* and *S. plymuthica* strongly inhibited *A. thaliana* growth.

The 16S of the isolate E.1b showed the highest identity with *Achromobacter insuavis* LMG 26845(T) (99.93%), and phylogenetics analysis reinforces its identification. Several studies report species of *Achromobacter* (β-Proteobacteria) acting as plant growth-promoting bacteria (Mayak et al., 2004; Jha and Kumar, 2009; Abdel-Rahman et al., 2017; Wang et al., 2020; Kapadia et al., 2022), but only a few have focused on VOC mechanisms. Recently, VOCs from *Achromobacter* sp. 5B1 were shown to increase both shoot and root biomass of *A. thaliana* by 66% and 55%, respectively (Jiménez-Vázquez et al., 2020), while no alteration was observed in lettuce (*Lactuca sativa*) co-cultivated with a strain from the same genus (Minerdi et al., 2011). Such opposite phenotypes can occur because VOC-mediated growth promotion is species-specific. The phenotype depends on the species or strain (bacterium and plant) that are interacting, the culture medium, growth condition, and other factors. For instance, increased dry biomass of *Mentha piperita* co-cultivated with *Pseudomonas fluorescens* WCS417r (Santoro et al., 2011) and tobacco (*Nicotiana tabacum* cv. Xanthi-nc) co-cultivated with *P. fluorescens* SS101 (Park et al., 2015) contrasts with reduced primary root formation in *A. thaliana* co-cultivated with *P. fluorescens* CHAO, probably due to hydrogen cyanide emission (Rudrappa et al., 2008). Thus, these results reinforce the importance of γ- and β-Proteobacteria for plant productivity and soil health, as previous studies have reported that they help in nutrient acquisition and provide plant protection (Lugtenberg and Kamilova, 2009; Malla et al., 2022).

## 4.5 Volatilome analysis revealed compounds known to promote plant growth

Bacterial volatilomes are diversified and dynamic, i.e., besides the differences in VOC profiles among bacteria, VOC production may change in quantity and quality due to different factors, such as media composition and growth period (Macías-Rubalcava et al., 2018; Heenan-Daly et al., 2021; Freitas et al., 2022). To characterize the volatilome produced by our bacterial isolates and further validate which VOCs are the bioactive compounds that induce growth, VOCs emitted by both promoter isolates IAT P4F9 and E.1b, and the isolate 1003-S-C1 (non-promoter, negative control) were analyzed by HS-

SPME/GC-MS. We also evaluated the *E. coli* DH5α volatilome to investigate why this strain promoted rice growth after 12 days of co-cultivation. All these bacteria belong to different genera, and in total, 58 VOCs were identified. These compounds can be classified according to the microbial volatile organic compounds database (mVOC, Lemfack et al., 2018) into several chemical classes, which are mostly benzenoids (24%), ketones (18%), sulfides (11%), alcohols (9%), furans (5%), pyrazines (5%), alkanes (4%), and others (24%). These are the most frequent chemical classes found in bacterial volatilomes (Wenke et al., 2012; Peñuelas et al., 2014).

Although it was not the focus of our study, we have also identified DMTS as a growth-inhibitor bio compound, at last in the highest concentration evaluated (Supplementary Figure 10). This compound is reported as capable of inhibiting the growth of several microorganisms (Ossowicki et al., 2017; Agisha et al., 2019; Guevara-Avendaño et al., 2019; Freitas et al., 2022) and inducing the expression of defense-related genes in apple fruit (Sun et al., 2022); however, it remains unclear if this could be due to a cytotoxic effect as our data suggested.

Here, we evaluated six compounds, individually and mixed in solutions, of which nonan-2-one (0.1 mg) increased dry shoot weight *in vitro* assays by up to 26% (Supplementary Figure 10). This compound was one of the five VOCs commonly identified in all the characterized volatilomes. Nonetheless, IAT P4F9 and E.1b produced nonan-2-one in higher concentrations compared with the negative controls. Fincheira et al. (2017) previously showed that nonan-2-one, identified in the volatilome of *Bacillus* sp. BCT9, was able to increase lettuce biomass by up to 48% at doses of 0.001 mg after 10 days of exposure to the volatile.

Although other compounds identified in the volatilome of our bacterial isolates were not evaluated or did not present significant effects on rice growth, they were validated in other plants. For instance, the application of 1.0 mg of DMDS promoted tobacco growth (Meldau et al., 2013), while 0.471 and 0.0471 mg of DMDS reduced the growth of *A. thaliana* by 20% and 30%, respectively (Kai et al., 2010). A VOC with a similar structure and belonging to the same chemical class (sulfide), DMTS, induced *A. thaliana* growth (Cordovez et al., 2018). Park et al. (2015) showed a significant increase in the fresh weight of *N. tabacum* when treated with butan-2-one. Camarena-Pozos et al. (2019) reported that 2-phenylethanol induced *A. thaliana* and *Agave tequilana* growth, while 3-methylbutan-1-ol induced *A. thaliana* and *A. salmiana* growth. Besides nonan-2-one, tridecan-2-one and undecane-2-one (both at $10^{-6}$ mg) have also elicited an increase in lettuce dry weight by 23% and 25%, respectively (Fincheira et al., 2017). Although a given compound can cause different effects on distinct plant species, we can hypothesize that one or some of these compounds might be responsible for the growth promotion of rice. Two of them (undecane-2-one and DMDS) were found exclusively on the volatilomes of the growth promoter isolates IAT P4F9 and E.1b. Intriguingly, 1003-S-C1 also shares some bioactive compounds that promote plant growth, such as 3-methylbutan-1-ol, 2-phenylethanol, and nonan-2-one, reinforcing that the growth-promoting phenotype is reached through a junction of several parameters, not only the presence or absence of a given compound.

Regarding growth promotion of rice co-cultivated with *E. coli* DH5α, some VOCs already validated as growth inducers were found in the volatilome of this strain. As already discussed in *Section 4.1*, Bailly et al. (2014) showed that *E. coli* DH5α can promote plant growth; the authors attributed these effects to the emission of VOC 1H-indole. This compound was not present in *E. coli* DH5α volatilome (Figure 3) because, after the manual peak-to-peak curing, it was considered peaks from single VOCs (i.e., co-elution metabolites were discarded) and VOCs present in at least three biological replicates and with an area at least twice as large in the samples as in the control (only LB media). However, as shown in Supplementary Figure 11, this compound was produced by this strain, but it co-eluted with others. To validate these data, we reanalyzed the *E. coli* DH5α volatilome using a quadrupole GC/MS and the 1H-indole eluted in a distinct GC peak (Supplementary Figure 12). In addition, as previously mentioned, other compounds that might have a growth-inducing effect were found in *E. coli* DH5α volatilome, such as nonan-2-one (Fincheira et al., 2017), butan-2-one (Park et al., 2015), 3-methyl-1-butanol, and 2-phenylethanol (Camarena-Pozos et al., 2019).

In conclusion, our study identified two bacterial isolates, belonging to *Achromobacter* and *Serratia* genera, which produced VOCs capable of inducing rice growth *in vitro*. Among these VOCs, nonan-2-one was validated *in vitro* as a growth inducer compound in this species. The bacterial VOCs also induced metabolic changes in these plants, especially in primary metabolites. In particular, rice energy metabolism seems to be upregulated by the isolate IAT P4F9. However, further studies are necessary to better understand the molecular changes mediated by bacterial VOCs on rice. Finally, this study also shows, for the first time, the *in vivo* growth promotion of a monocot, which is a potential sustainable tool for increasing rice productivity in the field.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/, accession numbers OP186019, OP186020, OP186021, OP893533 and OP893532.

## Author contributions

NA and OA carried out the experiments and performed the data analyses. OA and JO wrote the first draft of the manuscript.

AM and GP contributed to investigation. MS, GP and MC-R contributed to data analysis and wrote sections of the manuscript. JO contributed to conception, design of study, supervision, and funding acquisition. All authors revised the final manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2017/20521-6 and 2022/00474-1). NA (FAPESP 2019/08522-2) and OA (Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq, 141119/2019-5) are grateful for their fellowships.

## Acknowledgments

This research used facilities of the Brazilian Biorenewables National Laboratory (LNBR), part of the Brazilian Center for Research in Energy and Materials (CNPEM), a private nonprofit organization under the supervision of the Brazilian Ministry for Science, Technology, and Innovation (MCTI). The Metabolomics open access facilities staff is acknowledged for the assistance during the experiments. The authors would like to thank to Douglas Paixão for Sanger sequencing. The authors would also like to thank NMR facility (LNBio – CNPEM) for their support with the NMR sample preparation, spectrum data processing, and analyses, especially Dr. Marcos Alborghetti.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1056082/full#supplementary-material

## References

Abadía, J., López-Millán, A. F., Rombolà, A., and Abadía, A. (2002). Organic acids and fe deficiency: A review. *Plant Soil* 241 (1), 75–86. doi: 10.1023/A:1016093317898

Abdel-Rahman, H. M., Salem, A. A., Moustafa, M. M. A., and El-Garhy, H. A. S. (2017). A novice *Achromobacter* sp. EMCC1936 strain acts as a plant-growth-promoting agent. *Acta Physiol. Plant* 39, 1–15. doi: 10.1007/s11738-017-2360-6

Agisha, V. N., Kumar, A., Eapen, S. J., Sheoran, N., and Suseelabhai, R. (2019). Broad-spectrum antimicrobial activity of volatile organic compounds from endophytic *Pseudomonas putida* BP25 against diverse plant pathogens. *Biocontrol Sci. Technol.* 29, 1069–1089. doi: 10.1080/09583157.2019.1657067

Asari, S., Matzén, S., Petersen, M. A., Bejai, S., and Meijer, J. (2016). Multiple effects of *Bacillus amyloliquefaciens* volatile compounds: Plant growth promotion and growth inhibition of phytopathogens. *FEMS Microbiol. Ecol.* 92, 1–11. doi: 10.1093/femsec/fiw070

Bailly, A., Groenhagen, U., Schulz, S., Geisler, M., Eberl, L., and Weisskopf, L. (2014). The inter-kingdom volatile signal indole promotes root development by interfering with auxin signalling. *Plant J.* 80, 758–771. doi: 10.1111/TPJ.12666

Barbosa, J. C., and Maldonado, W.Jr. (2010). *Experimentação agronômica & AgroEstat – sistema para análises estatísticas de ensaios agronômicos* (Jaboticabal: Universidade Estadual Paulista). Júlio de Mesquita Filho.

Bhattacharyya, D., Garladinne, M., and Lee, Y. H. (2015). Volatile indole produced by rhizobacterium *Proteus vulgaris* JBLS202 stimulates growth of *Arabidopsis thaliana* through auxin, cytokinin, and brassinosteroid pathways. *J. Plant Growth Regul.* 34, 158–168. doi: 10.1007/s00344-014-9453-x

Bitas, V., Kim, H. S., Bennett, J. W., and Kang, S. (2013). Sniffing on microbes: Diverse roles of microbial volatile organic compounds in plant health. *Mol. Plant-Microbe Interact.* 26, 835–843. doi: 10.1094/MPMI-10-12-0249-CR

Bitas, V., Mccartney, N., Li, N., Demers, J., Kim, J., Kim, H., et al. (2015). *Fusarium oxysporum* volatiles enhance plant growth *via* affecting auxin transport and signaling. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.01248

Blom, D., Fabbri, C., Connor, E. C., Schiestl, F. P., Klauser, D. R., Boller, T., et al. (2011a). Production of plant growth modulating volatiles is widespread among rhizosphere bacteria and strongly. *Environ. Microbiol.* 13, 3047–3058. doi: 10.1111/j.1462-2920.2011.02582.x

Blom, D., Fabbri, C., Eberl, L., and Weisskopf, L. (2011b). Volatile-mediated killing of *Arabidopsis thaliana* by bacteria is mainly due to hydrogen cyanide. *Appl. Environ. Microbiol.* 77, 1000. doi: 10.1128/AEM.01968-10

Brilli, F., Loreto, F., and Baccelli, I. (2019). Exploiting plant volatile organic compounds (VOCS) in agriculture to improve sustainable defense strategies and productivity of crops. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00264

B. B. Buchanan, W. Gruissem and R. L. Jones (Eds.) (2015). *Biochemistry and molecular biology of plants. 2nd edition* (Hoboken, NJ, USA: Wiley-Blackwell).

Camarena-Pozos, D. A., Flores-Núñez, V. M., López, M. G., López-Bucio, J., and Partida-Martínez, L. P. (2019). Smells from the desert: microbial volatiles that affect plant growth and development of native and non-native plant species. *Plant Cell Environ.* 42, 1368–1380. doi: 10.1111/PCE.13476

Castulo-Rubio, D. Y., Alejandre-Ramírez, N. A., Orozco-Mosqueda, M.d. C., Santoyo, G., Macías-Rodríguez, L. I., and Valencia-Cantero, E. (2015). Volatile organic compounds produced by the rhizobacterium *Arthrobacter agilis* UMCV2 modulate *Sorghum bicolor* (Strategy II plant) morphogenesis and *SbFRO1* transcription *in vitro*. *J. Plant Growth Regul.* 34, 611–623. doi: 10.1007/s00344-015-9495-8

Cheng, X., Cordovez, V., Etalo, D. W., van der Voort, M., and Raaijmakers, J. M. (2016). Role of the GacS sensor kinase in the regulation of volatile production by plant growth-promoting *Pseudomonas fluorescens* SBW25. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01706

Cordovez, V., Schop, S., Hordijk, K., de Boulois, H. D., Coppens, F., Hanssen, I., et al. (2018). Priming of plant growth promotion by volatiles of root-associated microbacterium spp. *Appl. Environ. Microbiol.* 84, e01865–18. doi: 10.1128/AEM.01865-18

Counce, P., Keisling, T., and Mitchell, A. (2000). A uniform, objective, and adaptive system for expressing rice development. *Crop Sci. - Crop Sci.* 40, 436–443. doi: 10.2135/cropsci2000.402436x

Delaplace, P., Delory, B. M., Baudson, C., De Cazenave, M. M., Spaepen, S., Varin, S., et al. (2015). Influence of rhizobacterial volatiles on the root system architecture and the production and allocation of biomass in the model grass *Brachypodium distachyon* (L.) p. beauv. *BMC Plant Biol.* 15, 1–15. doi: 10.1186/s12870-015-0585-3

Dias, B. H. S., Jung, S.-H., Oliveira, J. V., de, C., and Ryu, C.-M. (2021). C4 bacterial volatiles improve plant health. *Pathogens* 10, 682. doi: 10.3390/pathogens10060682

Effmert, U., Kalderás, J., Warnke, R., and Piechulla, B. (2012). Volatile mediated interactions between bacteria and fungi in the soil. *J. Chem. Ecol.* 38, 665–703. doi: 10.1007/s10886-012-0135-5

Fernie, A. R., and Pichersky, E. (2015). Focus issue on metabolism: Metabolites, metabolites everywhere. *Plant Physiol.* 169, 1421–1423. doi: 10.1104/PP.15.01499

Fincheira, P., Parra, L., Mutis, A., Parada, M., and Quiroz, A. (2017). Volatiles emitted by *Bacillus* sp. BCT9 act as growth modulating agents on *Lactuca sativa* seedlings. *Microbiol. Res.* 203, 47–56. doi: 10.1016/j.micres.2017.06.007

Fincheira, P., and Quiroz, A. (2018). Microbial volatiles as plant growth inducers. *Microbiol. Res.* 208, 63–75. doi: 10.1016/j.micres.2018.01.002

Fincheira, P., and Quiroz, A. (2019). Physiological response of *Lactuca sativa* exposed to 2-nonanone emitted by *Bacillus* sp. BCT9. *Microbiol. Res.* 219, 49–55. doi: 10.1016/j.micres.2018.11.002

Flowers, H. M. (1981). Chemistry and biochemistry of d- and l-fucose. *Adv. Carbohydr. Chem. Biochem.* 39, 279–345. doi: 10.1016/S0065-2318(08)60208-5

Freitas, C. S. A., Maciel, L. F., Corrêa dos Santos, R. A., Costa, O. M. M. M., Maia, F. C. B., Rabelo, R. S., et al. (2022). Bacterial volatile organic compounds induce adverse ultrastructural changes and DNA damage to the sugarcane pathogenic fungus *Thielaviopsis ethacetica*. *Environ. Microbiol* 24, 1430–1453. doi: 10.1111/1462-2920.15876

Guevara-Avendaño, E., Bejarano-Bolívar, A. A., Kiel-Martínez, A.-L., Ramírez-Vázquez, M., Méndez-Bravo, A., von Wobeser, E. A., et al. (2019). Avocado rhizobacteria emit volatile organic compounds with antifungal activity against *Fusarium solani*, fusarium sp. associated with kuroshio shot hole borer, and *Colletotrichum gloeosporioides*. *Microbiol. Res.* 219, 74–83. doi: 10.1016/j.micres.2018.11.009

Guo, Y., Ghirardo, A., Weber, B., Schnitzler, J. P., Philipp Benz, J., and Rosenkranz, M. (2019). *Trichoderma* species differ in their volatile profiles and in antagonism toward ectomycorrhiza *Laccaria bicolor*. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.00891

Gutiérrez-Luna, F. M., López-Bucio, J., Altamirano-Hernández, J., Valencia-Cantero, E., de la Cruz, H. R., and Macías-Rodríguez, L. (2010). Plant growth-promoting rhizobacteria modulate root-system architecture in *Arabidopsis thaliana* through volatile organic compound emission. *Symbiosis* 51, 75–83. doi: 10.1007/s13199-010-0066-2

Heenan-Daly, D., Coughlan, S., Dillane, E., and Doyle Prestwich, B. (2021). Volatile compounds from *Bacillus*, *Serratia*, and *Pseudomonas* promote growth and alter the transcriptional landscape of *Solanum tuberosum* in a passively ventilated growth system. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.628437

Hernández-Calderón, E., Aviles-Garcia, M. E., Castulo-Rubio, D. Y., Macías-Rodríguez, L., Ramírez, V. M., Santoyo, G., et al. (2018). Volatile compounds from beneficial or pathogenic bacteria differentially regulate root exudation, transcription of iron transporters, and defense signaling pathways in *Sorghum bicolor*. *Plant Mol. Biol.* 96, 291–304. doi: 10.1007/s11103-017-0694-5

Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L., and Braun, H. P. (2015). Amino acid catabolism in plants. *Mol. Plant* 8, 1563–1579. doi: 10.1016/j.molp.2015.09.005

Hoagland, D. R., and Arnon, D. I. (1950). The water-culture method for growing plants without soil. *Calif Agric* (Berkeley). 347, 1–32.

Izawa, T., and Shimamoto, K. (1996). Becoming a model plant: The importance of rice to plant science. *Trends Plant Sci.* 1, 95–99. doi: 10.1016/S1360-1385(96)80041-0

Jha, P., and Kumar, A. (2009). Characterization of novel plant growth promoting endophytic bacterium *Achromobacter xylosoxidans* from wheat plant. *Microb. Ecol.* 58, 179–188. doi: 10.1007/S00248-009-9485-0

Jiang, C. H., Xie, Y. S., Zhu, K., Wang, N., Li, Z. J., Yu, G. J., et al. (2019). Volatile organic compounds emitted by bacillus sp. JC03 promote plant growth through the action of auxin and strigolactone. *Plant Growth Regul.* 87, 317–328. doi: 10.1007/s10725-018-00473-z

Jiménez-Vázquez, K. R., García-Cárdenas, E., Barrera-Ortiz, S., Ortiz-Castro, R., Ruiz-Herrera, L. F., Ramos-Acosta, B. P., et al. (2020). The plant beneficial rhizobacterium achromobacter sp. 5B1 influences root development through auxin signaling and redistribution. *Plant J.* 103, 1639–1654. doi: 10.1111/TPJ.14853

Kai, M., Crespo, E., Cristescu, S. M., Harren, F. J. M., Francke, W., and Piechulla, B. (2010). *Serratia odorifera*: Analysis of volatile emission and biological impact of volatile compounds on *Arabidopsis thaliana*. *Appl. Microbiol. Biotechnol.* 88 (4), 965–976. doi: 10.1007/S00253-010-2810-1

Kai, M., Vespermann, A., and Piechulla, B. (2008). The growth of fungi and *Arabidopsis thaliana* is influenced by bacterial volatiles. *Plant Signal. Behav.* 3, 482–484. doi: 10.4161/psb.3.7.5681

Kanchiswamy, C. N., Malnoy, M., and Maffei, M. E. (2015). Chemical diversity of microbial volatiles and their potential for plant growth and productivity. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00151

Kang, S. M., Radhakrishnan, R., Lee, S. M., Park, Y. G., Kim, A. Y., Seo, C. W., et al. (2015). *Enterobacter* sp. SE992-induced regulation of amino acids, sugars, and hormones in cucumber plants improves salt tolerance. *Acta Physiol. Plant* 37, 1–10. doi: 10.1007/s11738-015-1895-7

Kapadia, C., Patel, N., Rana, A., Vaidya, H., Alfarraj, S., Ansari, M. J., et al. (2022). Evaluation of plant growth-promoting and salinity ameliorating potential of halophilic bacteria isolated from saline soil. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.946217

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/MOLBEV/MST010

Kwon, Y. S., Ryu, C.-M., Lee, S., Park, H. B., Han, K. S., Lee, J. H., et al. (2010). Proteome analysis of *Arabidopsis* seedlings exposed to bacterial volatiles. *Planta* 232, 1355–1370. doi: 10.1007/s00425-010-1259-x

Lea, P. J., Sodek, L., Parry, M. A. J., Shewry, P. R., and Halford, N. G. (2007). Asparagine in plants. *Ann. Appl. Biol.* 150, 1–26. doi: 10.1111/j.1744-7348.2006.00104.x

Lemfack, M. C., Gohlke, B. O., Toguem, S. M. T., Preissner, S., Piechulla, B., and Preissner, R. (2018). MVOC 2.0: A database of microbial volatiles. *Nucleic Acids Res.* 46, D1261–D1265. doi: 10.1093/nar/gkx1016

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/NAR/GKAB301

Lo Cantore, P., Giorgio, A., and Iacobellis, N. S. (2015). Bioactivity of volatile organic compounds produced by *Pseudomonas tolaasii*. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.01082

Locey, K. J., and Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci.* 113, 5970–5975. doi: 10.1073/pnas.1521291113

Lugtenberg, B., and Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. *Annu. Rev. Microbiol.* 63, 541–556. doi: 10.1146/ANNUREV.MICRO.62.081307.162918

Macías-Rubalcava, M. L., Sánchez-Fernández, R. E., Roque-Flores, G., Lappe-Oliveras, P., and Medina-Romero, Y. M. (2018). Volatile organic compounds from *Hypoxylon anthochroum* endophytic strains as postharvest mycofumigation alternative for cherry tomatoes. *Food Microbiol.* 76, 363–373. doi: 10.1016/j.fm.2018.06.014

Maheshwari, A., Mmbaga, M., Bhusal, B., and Ondzighi-Assoume, C. (2021). Effect of volatile compounds produced by selected bacterial endophytes in promoting plant growth. *HortScience* 56, 1175–1182. doi: 10.21273/HORTSCI15668-21

Malla, M. A., Dubey, A., Kumar, A., and Yadav, S. (2022). Metagenomic analysis displays the potential predictive biodegradation pathways of the persistent pesticides in agricultural soil with a long record of pesticide usage. *Microbiol. Res.* 261, 127081. doi: 10.1016/J.MICRES.2022.127081

Mayak, S., Tirosh, T., and Glick, B. R. (2004). Plant growth-promoting bacteria that confer resistance to water stress in tomatoes and peppers. *Plant Sci.* 166, 525–530. doi: 10.1016/j.plantsci.2003.10.025

Meldau, D. G., Meldau, S., Hoang, L. H., Underberg, S., Wünsche, H., and Baldwin, I. T. (2013). Dimethyl disulfide produced by the naturally associated bacterium *Bacillus* sp. B55 promotes *Nicotiana attenuata* growth by enhancing sulfur nutrition. *Plant Cell* 25, 2731–2747. doi: 10.1105/tpc.113.114744

Miflin, B. J., and Habash, D. Z. (2002). The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.* 53, 979–987. doi: 10.1093/jexbot/53.370.979

Minerdi, D., Bossi, S., Maffei, M. E., Gullino, M. L., and Garibaldi, A. (2011). *Fusarium oxysporum* and its bacterial consortium promote lettuce growth and expansin A5 gene expression through microbial volatile organic compound (MVOC) emission. *FEMS Microbiol. Ecol.* 76, 342–351. doi: 10.1111/j.1574-6941.2011.01051.x

Mollet, C., Drancourt, M., and Raoult, D. (1997). *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol. Microbiol.* 26, 1005–1011. doi: 10.1046/J.1365-2958.1997.6382009.X

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nelissen, H., Gonzalez, N., and Inzé, D. (2016). Leaf growth in dicots and monocots: so different yet so alike. *Curr. Opin. Plant Biol.* 33, 72–76. doi: 10.1016/j.pbi.2016.06.009

Ossowicki, A., Jafra, S., and Garbeva, P. (2017). The antimicrobial volatile power of the rhizospheric isolate *Pseudomonas donghuensis* P482. *PloS One* 12, e0174362. doi: 10.1371/journal.pone.0174362

Pagès, L. (2016). Branching patterns of root systems: comparison of monocotyledonous and dicotyledonous species. *Ann. Bot.* 118, 1337. doi: 10.1093/aob/mcw185

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., et al. (2021). MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 49, W388–W396. doi: 10.1093/nar/gkab382

Parida, A. K., Sekhar, S., Panda, B. B., Sahu, G., and Shaw, B. P. (2022). Effect of panicle morphology on grain filling and rice yield: Genetic control and molecular regulation. *Front. Genet.* 13. doi: 10.3389/fgene.2022.876198

Park, Y., Dutta, S., Ann, M., Raaijmakers, J. M., and Park, K. (2015). Promotion of plant growth by *Pseudomonas fluorescens* strain SS101 *via* novel volatile organic compounds. *Biochem. Biophys. Res. Commun.* 461, 361–365. doi: 10.1016/j.bbrc.2015.04.039

Peñuelas, J., Asensio, D., Tholl, D., Wenke, K., Rosenkranz, M., Piechulla, B., et al. (2014). Biogenic volatile emissions from the soil. *Plant Cell Environ.* 37, 1866–1891. doi: 10.1111/pce.12340

Piechulla, B., Lemfack, M. C., and Kai, M. (2017). Effects of discrete bioactive microbial volatiles on plants and fungi. *Plant Cell Environ.* 40, 2042–2067. doi: 10.1111/PCE.13011

Plyuta, V. A., Chernikova, A. S., Sidorova, D. E., Kupriyanova, E. V., Koksharova, O. A., Chernin, L. S., et al. (2021). Modulation of *Arabidopsis thaliana* growth by volatile substances emitted by *Pseudomonas* and *Serratia strains*. *World J. Microbiol. Biotechnol.* 37, 1–9. doi: 10.1007/s11274-021-03047-w

Rath, M., Mitchell, T. R., and Gold, S. E. (2018). Volatiles produced by *Bacillus mojavensis* RRC101 act as plant growth modulators and are strongly culture-dependent. *Microbiol. Res.* 208, 76–84. doi: 10.1016/j.micres.2017.12.014

Raza, W., Ling, N., Liu, D., Wei, Z., Huang, Q., and Shen, Q. (2016a). Volatile organic compounds produced by *Pseudomonas fluorescens* WR-1 restrict the growth and virulence traits of *Ralstonia solanacearum*. *Microbiol. Res.* 192, 103–113. doi: 10.1016/j.micres.2016.05.014

Raza, W., Ling, N., Yang, L., Huang, Q., and Shen, Q. (2016b). Response of tomato wilt pathogen *Ralstonia solanacearum* to the volatile organic compounds produced by a biocontrol strain *Bacillus amyloliquefaciens* SQR-9. *Sci. Rep.* 6, 24856. doi: 10.1038/srep24856

Rudrappa, T., Splaine, R. E., Biedrzycki, M. L., and Bais, H. P. (2008). Cyanogenic pseudomonads influence multitrophic interactions in the rhizosphere. *PloS One* 3, 2073. doi: 10.1371/journal.pone.0002073

Ryu, C. M., Farag, M. A., Hu, C. H., Reddy, M. S., Kloepper, J. W., and Paré, P. W. (2004). Bacterial volatiles induce systemic resistance in *Arabidopsis*. *Plant Physiol.* 134, 1017–1026. doi: 10.1104/pp.103.026583

Ryu, C. M., Farag, M. A., Hu, C. H., Reddy, M. S., Wei, H. X., Paré, P. W., et al. (2003). Bacterial volatiles promote growth in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4927–4932. doi: 10.1073/pnas.0730845100

Ryu, C.-M., Farag, M. A., Paré, P. W., and Kloepper, J. W. (2005). Invisible signals from the underground: bacterial volatiles elicit plant growth promotion and induce systemic resistance. *Plant Pathol. J.* 21, 7–12. doi: 10.5423/PPJ.2005.21.1.007

Sánchez-López, Á.M., Baslam, M., De Diego, N., Muñoz, F. J., Bahaji, A., Almagro, G., et al. (2016). Volatile compounds emitted by diverse phytopathogenic microorganisms promote plant growth and flowering through cytokinin action. *Plant Cell Environ.* 39, 2592–2608. doi: 10.1111/pce.12759

Santoro, M. V., Zygadlo, J., Giordano, W., and Banchio, E. (2011). Volatile organic compounds from rhizobacteria increase biosynthesis of essential oils and growth parameters in peppermint (*Mentha piperita*). *Plant Physiol. Biochem.* 49, 1177–1182. doi: 10.1016/j.plaphy.2011.07.016

Schmidt, R., Cordovez, V., de Boer, W., Raaijmakers, J., and Garbeva, P. (2015). Volatile affairs in microbial interactions. *ISME J.* 9, 2329–2335. doi: 10.1038/ismej.2015.42

Schmidt, R., Etalo, D. W., de Jager, V., Gerards, S., Zweers, H., de Boer, W., et al. (2016). Microbial small talk: Volatiles in fungal–bacterial interactions. *Front. Microbiol.* 6. doi: 10.3389/fmicb.2015.01495

Schulz-Bohm, K., Gerards, S., Hundscheid, M., Melenhorst, J., de Boer, W., and Garbeva, P. (2018). Calling from distance: attraction of soil bacteria by plant root volatiles. *ISME J.* 12, 1252–1262. doi: 10.1038/s41396-017-0035-3

Schulz-Bohm, K., Martín-Sánchez, L., and Garbeva, P. (2017). Microbial volatiles: small molecules with an important role in intra- and inter-kingdom interactions. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.02484

Schulz, S., and Dickschat, J. S. (2007). Bacterial volatiles: the smell of small organisms. *Nat. Prod Rep.* 24, 814–842. doi: 10.1039/b507392h

Song, G. C., Riu, M., and Ryu, C.-M. (2019). Beyond the two compartments Petri-dish: optimising growth promotion and induced resistance in cucumber exposed to gaseous bacterial volatiles in a miniature greenhouse system. *Plant Methods* 15, 1–11. doi: 10.1186/S13007-019-0395-Y

SOSBAI (2018). *Arroz irrigado: recomendações técnicas da pesquisa para o sul do brasil*. Available at: https://www.sosbai.com.br/ (Accessed November 5, 2022).

Sreenivasulu, N., and Wobus, U. (2013). Seed-development programs: A systems biology–based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* 64, 189–217. doi: 10.1146/annurev-arplant-050312-120215

Sun, M., Duan, Y., Liu, J. P., Fu, J., and Huang, Y. (2022). Efficacy of dimethyl trisulfide on the suppression of ring rot disease caused by *Botryosphaeria dothidea* and induction of defense-related genes on apple fruits. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.796167

Tahir, H. A. S., Gu, Q., Wu, H., Niu, Y., Huo, R., and Gao, X. (2017a). *Bacillus* volatiles adversely affect the physiology and ultra-structure of *Ralstonia solanacearum* and induce systemic resistance in tobacco against bacterial wilt. *Sci. Rep.* 7, 1–15. doi: 10.1038/srep40481

Tahir, H. A. S., Gu, Q., Wu, H., Raza, W., Hanif, A., Wu, L., et al. (2017b). Plant growth promotion by volatile organic compounds produced by *Bacillus subtilis* SYST2. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00171

Tahir, H. A. S., Gu, Q., Wu, H., Raza, W., Safdar, A., Huang, Z., et al. (2017c). Effect of volatile compounds produced by *Ralstonia solanacearum* on plant growth promoting and systemic resistance inducing potential of *Bacillus volatiles*. *BMC Plant Biol.* 17, 1–16. doi: 10.1186/s12870-017-1083-6

Tilocca, B., Cao, A., and Migheli, Q. (2020). Scent of a killer: Microbial volatilome and its role in the biological control of plant pathogens. *Front. Microbiol.* 11. doi: 10.3389/FMICB.2020.00041/BIBTEX

Tyc,, Song, C., Dickschat, J. S., Vos, M., and Garbeva, P. (2017). The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends Microbiol.* 25, 280–292. doi: 10.1016/j.tim.2016.12.002

United States Department of Agriculture (2022) *USDA ERS - rice yearbook*. Available at: https://www.ers.usda.gov/data-products/rice-yearbook/ (Accessed July 27, 2022).

Velázquez-Becerra, C., Macías-Rodríguez, L. I., López-Bucio, J., Altamirano-Hernández, J., Flores-Cortez, I., and Valencia-Cantero, E. (2011). A volatile organic compound analysis from *Arthrobacter agilis* identifies dimethylhexadecylamine, an amino-containing lipid modulating bacterial growth and *Medicago sativa* morphogenesis *in vitro*. *Plant Soil* 339, 329–340. doi: 10.1007/s11104-010-0583-z

Wang, K. T., Li, Y. P., Wu, Y. C., Qiu, Z. Q., Ding, Z. X., Wang, X. J., et al. (2020). Improved grain yield and lowered arsenic accumulation in rice plants by inoculation with arsenite-oxidizing *Achromobacter xylosoxidans* GD03. *Ecotoxicol. Environ. Saf.* 206, 111229. doi: 10.1016/j.ecoenv.2020.111229

Wenke, K., Weise, T., Warnke, R., Valverde, C., Wanke, D., Kai, M., et al. (2012). "Bacterial volatiles mediating information between bacteria and plants," in *Biocommunication of plants*. Eds. G. Witzany and F. Baluška (Berlin, Heidelberg: Springer), 327–347. doi: 10.1007/978-3-642-23524-5_17

Westhoff, S., van Wezel, G. P., and Rozen, D. E. (2017). Distance-dependent danger responses in bacteria. *Curr. Opin. Microbiol.* 36, 95–101. doi: 10.1016/j.mib.2017.02.002

Xie, X., Zhang, H., and Pare, P. (2009). Plant signaling & behavior sustained growth promotion in *Arabidopsis* with long-term exposure to the beneficial soil bacterium *Bacillus subtilis* (GB03). *Plant Signal. Behav.* 4, 948–953. doi: 10.4161/psb.4.10.9709

Yang, Q., Zhao, D., and Liu, Q. (2020). Connections between amino acid metabolisms in plants: Lysine as an example. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.00928/BIBTEX

Yasmin, H., Rashid, U., Hassan, M. N., Nosheen, A., Naz, R., Ilyas, N., et al. (2021). Volatile organic compounds produced by *Pseudomonas pseudoalcaligenes* alleviated drought stress by modulating defense system in maize (*Zea mays* l.). *Physiol. Plant* 172, 896–911. doi: 10.1111/ppl.13304

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Yousaf, M., Bashir, S., Raza, H., Shah, A. N., Iqbal, J., Arif, M., et al. (2021). Role of nitrogen and magnesium for growth, yield and nutritional quality of radish. *Saudi J. Biol. Sci.* 28, 3021–3030. doi: 10.1016/J.SJBS.2021.02.043

Yukphan, P., Potacharoen, W., Tanasupawat, S., Tanticharoen, M., and Yamada, Y. (2004). *Asaia krungthepensis* sp. nov., an acetic acid bacterium in the alpha-proteobacteria. *Int. J. Syst. Evol. Microbiol.* 54, 313–316. doi: 10.1099/ijs.0.02734-0

Zell, M. B., Fahnenstich, H., Maier, A., Saigo, M., Voznesenskaya, E. V., Edwards, G. E., et al. (2010). Analysis of *Arabidopsis* with highly reduced levels of malate and fumarate sheds light on the role of these organic acids as storage carbon molecules. *Plant Physiol.* 152, 1251–1262. doi: 10.1104/PP.109.151795

Zhang, H., Kim, M. S., Krishnamachari, V., Payton, P., Sun, Y., Grimson, M., et al. (2007). Rhizobacterial volatile emissions regulate auxin homeostasis and cell expansion in *Arabidopsis*. *Planta* 226, 839–851. doi: 10.1007/s00425-007-0530-2

Zou, C., Li, Z., and Yu, D. (2010). *Bacillus megaterium* strain XTBG34 promotes plant growth by producing 2-pentylfuran. *J. Microbiol.* 48, 460–466. doi: 10.1007/s12275-010-0068-z

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 August 2023
DOI 10.3389/fpls.2023.1114398



OPEN ACCESS

EDITED BY
Marta Sousa Silva,
University of Lisbon, Portugal

REVIEWED BY
Antonio Giovino,
Council for Agricultural and Economics Research (CREA), Italy
José S. Câmara,
Universidade da Madeira, Portugal

*CORRESPONDENCE
Guillermo Federico Padilla-González
federico.padillagonzalez@wur.nl

†PRESENT ADDRESS
Guillermo Federico Padilla-Gonza´ lez,
Wageningen Food Safety Research,
Wageningen University &Research,
Wageningen, Netherlands

RECEIVED 02 December 2022
ACCEPTED 19 July 2023
PUBLISHED 10 August 2023

CITATION
Padilla-González GF, Rosselli A, Sadgrove NJ, Cui M and Simmonds MSJ (2023) Mining the chemical diversity of the hemp seed (*Cannabis sativa* L.) metabolome: discovery of a new molecular family widely distributed across hemp.
*Front. Plant Sci.* 14:1114398.
doi: 10.3389/fpls.2023.1114398



# Mining the chemical diversity of the hemp seed (*Cannabis sativa* L.) metabolome: discovery of a new molecular family widely distributed across hemp

Guillermo Federico Padilla-González[1*†], Abigail Rosselli[1], Nicholas J. Sadgrove[1,2], Max Cui[1] and Monique S.J. Simmonds[1]

[1]Enhanced Partnerships Department Royal Botanic Gardens, Kew, London, United Kingdom, [2]Department of Botany and Plant Biotechnology, University of Johannesburg (Auckland Park Campus), Auckland Park, Johannesburg, South Africa

Hemp (*Cannabis sativa* L.) is a widely researched industrial crop with a variety of applications in the pharmaceutical, nutraceutical, food, cosmetic, textile, and materials industries. Although many of these applications are related to its chemical composition, the chemical diversity of the hemp metabolome has not been explored in detail and new metabolites with unknown properties are likely to be discovered. In the current study, we explored the chemical diversity of the hemp seed metabolome through an untargeted metabolomic study of 52 germplasm accessions to 1) identify new metabolites and 2) link the presence of biologically important molecules to specific accessions on which to focus on in future studies. Multivariate analysis of mass spectral data demonstrated large variability of the polar chemistry profile between accessions. Five main groups were annotated based on their similar metabolic fingerprints. The investigation also led to the discovery of a new compound and four structural analogues, belonging to a previously unknown chemical class in hemp seeds: cinnamic acid glycosyl sulphates. Although variability in the fatty acid profiles was not as marked as the polar components, some accessions had a higher yield of fatty acids, and variation in the ratio of linoleic acid to α-linolenic acid was also observed, with some varieties closer to 3:1 (reported as optimal for human nutrition). We found that that cinnamic acid amides and lignanamides, the main chemical classes of bioactive metabolites in hemp seed, were more concentrated in the Spanish accession Kongo Hanf (CAN58) and the French accession CAN37, while the Italian cultivar Eletta Campana (CAN48) demonstrated the greatest yield of fatty acids. Our results indicate that the high variability of bioactive and novel metabolites across the studied hemp seed accessions may influence claims associated with their commercialization and inform breeding programs in cultivar development.

KEYWORDS

**Cannabis, cinnamic acid amines, chemical diversity, metabolomics, HEMP, fatty acids**

# 1 Introduction

*Cannabis sativa* L., specifically the subspecies, varieties and cultivars known as hemp, are considered one of the most versatile herbaceous plants in agriculture, because of its commercial value to the pharmaceutical, food, nutraceutical, cosmetics, textile, papermaking, and construction industries (Schluttenhofer and Yuan, 2017; Farinon et al., 2020). In the context of its nutraceutical use, the focus has been on its seeds, which are rich in polyunsaturated fatty acids, proteins, minerals, and specialized metabolites with noteworthy implications for human and animal nutrition (Callaway, 2004; Russo and Reggiani, 2015; Galasso et al., 2016; Farinon et al., 2020). In this regard, hemp seeds are processed to make edible oil, cake flour and protein powder (Farinon et al., 2020; Leonard et al., 2021b).

Considering that some varieties of *C. sativa* express psychoactive cannabinoids that are occasionally detected in *Cannabis*-based products, a distinction between hemp and the recreational variety has been made (Schluttenhofer and Yuan, 2017). The European Industrial Hemp Association (EIHA) defines hemp as "the plant *Cannabis sativa* L., or any part of it, with a delta-9 tetrahydrocannabinol (THC) concentration up to 0.3% on a dry weight basis" to differentiate it from psychoactive *C. sativa* in which the THC concentration exceeds 0.3% and can be up to 20% (European Industrial Hemp Association, https://eiha.org/). Subsequently, several countries mainly from Europe, North America and Asia legalized the cultivation of low THC hemp cultivars (Leonard et al., 2021b). By 2018 more than 50,000 hectares were allocated to the cultivation of this crop in Europe, one of the main producers of industrial hemp. Within Europe, France is currently the largest producer, followed by Italy and the Netherlands (European Industrial Hemp Association, https://eiha.org/). Over the years, varieties of hemp were developed to select for specific traits and seeds from many of these varieties are currently stored in international germplasm collections and botanical gardens. There are currently 75 varieties registered in the EU Catalogue, all of them with a THC content below 0.3% and a diverse profile of other metabolites (European Industrial Hemp Association, https://eiha.org/).

Beyond the well-known cannabinoids, *C. sativa* contains a diverse metabolome of bioactive metabolites implicated in human health and nutrition. Hemp seeds are low in cannabinoids, and rich in phenylpropionamides (cinnamic acid amides and lignanamides) and unsaturated fatty acids, attracting nutraceutical and commercial interest (Crescente et al., 2018; Leonard et al., 2021a; Leonard et al., 2021b). The principal nutraceutical value of hemp seed oil is in its fatty acid composition, dominated by >90% of polyunsaturated fatty acids (Schluttenhofer and Yuan, 2017). It contains two dietary essential fatty acids, linoleic acid and α-linolenic acid in the ratio of 2.5–3:1. This is allegedly ideal for human nutrition and cardiovascular health (Leizer et al., 2000; Simopoulos, 2008; Galasso et al., 2016). However, the presence of high quantities of phenylpropionamides in hemp seeds have also been linked to some of their biological properties (Chen et al., 2012; Moccia et al., 2020; Leonard et al., 2021a). Hemp seeds accumulate the highest structural diversity of lignanamides among other lignanamides-producing taxa, with more than 80 different compounds (Leonard et al., 2021a).

Cinnamic acid amides and lignanamides display potent anti-inflammatory, antioxidant and anti-cancer activities in both *in-vitro* and *in-vivo* studies (Chen et al., 2012; Farinon et al., 2020; Moccia et al., 2020; Leonard et al., 2021a). Previous studies have promoted these hemp polyphenols as protective agents against human chronic diseases (Leonard et al., 2021a). For example, phenylpropionamides such as cannabisin B and N-*trans*-caffeoyltyramine are significantly stronger antioxidants than the standard soy isoflavones (Chen et al., 2012). Subsequent studies demonstrated that cannabisin B also enacts antiproliferative activity by inducing autophagic cell death in liver hepatocarcinoma (HepG2) cells (Chen et al., 2013). Other hemp lignanamides, such as cannabisin F, have demonstrated a potential neuro-protective effect by reducing mRNA levels of pro-inflammatory mediators, intracellular reactive oxygen species (ROS) and tumor necrosis factor α (TNF-α) in lipopolysaccharide-stimulated BV2 microglia cells (Wang et al., 2019).

Thus, the health and nutritional properties of hemp seeds are dependent upon their chemical composition that differs mainly according to the variety used. With such a high number of hemp varieties currently available in international germplasm collections, a comprehensive chemical assessment of these accessions is especially relevant, specifically to discover new metabolites and link the presence of biologically important molecules to certain accessions, ultimately to exploit the hemp germplasm by designing commercial cultivars with specific attributes. However, to date the phytochemical diversity of the hemp seed metabolome has not been explored in detail. In the present study, we focused on characterizing the chemical diversity of the hemp seed metabolome across cultivars/accessions using an untargeted metabolomics study of 52 hemp seed accessions.

# 2 Material and methods

## 2.1 Plant material

A collection of 51 hemp seed accessions from fourteen countries were obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany (http://gbis.ipk-gatersleben.de/) (Table 1). These accessions, maintained *ex-situ* by the IPK, were provided as small quantities of seeds (< 1 g). An additional sample of hemp seeds (1 kg) (S1, Table 1) was acquired in a local supermarket for further isolation of targeted metabolites (see section 2.5), bringing the total number of samples to 52.

The 51 externally sourced samples derive from sixteen recognised cultivars, including five historic cultivars whose seeds have a "deleted" status (Table 1) in the EU database of registered plant varieties (https://ec.europa.eu/info/index_en). Samples that were attained in replicate include: three accessions of the cultivar "Fibrimon" (plus two related but deleted strains, Fibrimon 21 and 56), four of "Cinepa" and three of "Kompolti" (Table 1).

TABLE 1 Hemp seed accessions and cultivars included in the present study.

| Code | Scientific name | Cultivar or Accession name | Country of origin | EU catalogue status |
|---|---|---|---|---|
| CAN16 | *Cannabis sativa* L. | – | Slovakia | – |
| CAN17 | *Cannabis sativa* L. | – | Hungary | – |
| CAN18 | *Cannabis sativa* L. subsp. *sativa* | – | Germany | – |
| CAN19 | *Cannabis sativa* L. | – | Italy | – |
| CAN20 | *Cannabis sativa* L. | – | North Korea | – |
| CAN21 | *Cannabis sativa* L. subsp. *sativa* | – | Romania | – |
| CAN22 | *Cannabis sativa* L. subsp. *spontanea* Serebr. | – | Georgia | – |
| CAN23 | *Cannabis sativa* L. subsp. *sativa* | – | North Korea | – |
| CAN24 | *Cannabis sativa* L. | – | Italy | – |
| CAN26 | *Cannabis sativa* L. | – | Turkey | – |
| CAN27 | *Cannabis sativa* L. | – | unknown | – |
| CAN28 | *Cannabis sativa* L. | Forose | unknown | – |
| CAN29 | *Cannabis sativa* L. subsp. *sativa* | – | Romania | – |
| CAN30 | *Cannabis sativa* L. | Bernburger | Germany | – |
| CAN31 | *Cannabis sativa* L. subsp. *sativa* | Cunepa | Romania | – |
| CAN32 | *Cannabis sativa* L. subsp. *sativa* | Conopla (ukr.) | Romania | – |
| CAN33 | *Cannabis sativa* L. subs. *sativa* | Cinepa | Romania | – |
| CAN34 | *Cannabis sativa* L. subs. *sativa* | Cinepa | Romania | – |
| CAN35 | *Cannabis sativa* L. subs. *sativa* | Cinepa | Romania | – |
| CAN36 | *Cannabis sativa* L. subsp. *sativa* | – | unknown | – |
| CAN37 | *Cannabis sativa* L. subsp. *sativa* | – | France | – |
| CAN38 | *Cannabis sativa* L. subsp. *sativa* | Cinepa | Romania | – |
| CAN39 | *Cannabis sativa* L. | – | China | – |
| CAN40 | *Cannabis sativa* L. | – | Italy | – |
| CAN42 | *Cannabis sativa* L. subsp. *sativa* | – | Croatia | – |
| CAN43 | *Cannabis sativa* L. subsp. *sativa* | Hohenthurmer Gleichzeitig Reifender | Germany | – |
| CAN44 | *Cannabis sativa* L. subsp. *sativa* | Fibrimon | Germany | – |
| CAN45 | *Cannabis sativa* L. subsp. *sativa* | Krasnodarskaya | Russia* | – |
| CAN46 | *Cannabis sativa* L. | Fibrimon 56 | France* | Deleted |
| CAN47 | *Cannabis sativa* L. subsp. *sativa* | – | Turkey | – |
| CAN48 | *Cannabis sativa* L. | Eletta campana | Italy | Registered |
| CAN50 | *Cannabis sativa* L. subsp. *sativa* | Fibrimon | France* | – |
| CAN51 | *Cannabis sativa* L. subsp. *sativa* | – | Argentina | – |
| CAN52 | *Cannabis sativa* L. subsp. *sativa* | Fibridia | unknown | – |
| CAN53 | *Cannabis sativa* L. subsp. *sativa* | Fibrimon | France* | – |
| CAN54 | *Cannabis sativa* L. subsp. *sativa* | Fibrimon 21 | France* | Deleted |
| CAN55 | *Cannabis sativa* L. subsp. *sativa* | Havellander | Russia* | – |
| CAN56 | *Cannabis sativa* L. subsp. *sativa* | Kompolti | Hungary* | Registered |
| CAN58 | *Cannabis sativa* L. subsp. *sativa* | Kongo Hanf | Spain | – |

*(Continued)*

TABLE 1 Continued

| Code | Scientific name | Cultivar or Accession name | Country of origin | EU catalogue status |
|---|---|---|---|---|
| CAN59 | *Cannabis sativa* L. subsp. *sativa* | Turken P 494 | unknown | – |
| CAN60 | *Cannabis sativa* L. subsp. *sativa* | Schurig | Russia* | – |
| CAN61 | *Cannabis sativa* L. subsp. *sativa* | Carmagnola in Selezione | Italy | Registered |
| CAN62 | *Cannabis sativa* L. subsp. *sativa* | Bredemann P | Germany* | – |
| CAN63 | *Cannabis sativa* L. subsp. *sativa* | Ramo | Germany | – |
| CAN64 | *Cannabis sativa* L. subsp. *sativa* | – | Turkey | – |
| CAN65 | *Cannabis sativa* L. subsp. *sativa* | Juso 14 | Germany | Deleted |
| CAN66 | *Cannabis sativa* L. subsp. *sativa* | Lovrin 110 | Germany | Registered |
| CAN67 | *Cannabis sativa* L. subsp. *sativa* | Fasamo | Germany | Deleted |
| CAN68 | *Cannabis sativa* L. subsp. *sativa* | Kompolti | Germany | Registered |
| CAN69 | *Cannabis sativa* L. subsp. *sativa* | Futura | Germany | Deleted |
| CAN70 | *Cannabis sativa* L. subsp. *sativa* | Kompolti | Germany | Registered |
| S1 | *Cannabis sativa* L. | – | unknown | – |

* Country of origin not reported in the IPK database.

## 2.2 Extraction and LC-MS analyses

The extraction and metabolic profiling by LC-MS was based on the protocol reported by De Vos et al. (2007). Seeds from each of the hemp accessions were ground in liquid nitrogen with a pestle and mortar. Fifty milligrams of the resultant seed powder were then transferred to Eppendorf tubes and 2 mL of 80% MeOH (HPLC grade) were added as extraction solvent. Tubes were then placed in an ultrasonic bath for 15 min at 25°C using a frequency of 40 kHz. After extraction, samples were centrifuged at 19,975 × g for 10 min at room temperature, and the supernatant was transferred to an LC-MS vial. Samples were randomly analyzed by arbitrarily selecting vials to minimize statistical bias. Furthermore, 30 µL of each of the 52 samples were pooled into a LC-MS vial labelled as "Quality Control". This sample was analyzed multiple times along the chromatographic sequence to check for reproducibility (De Vos et al., 2007).

The metabolic profiles of the hemp seed accessions were recorded on a Vanquish UHPLC system coupled to a 100 Hz photodiode array detector (PDA) and an Orbitrap Fusion Tri-hybrid high-resolution tandem mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). Chromatographic separation of hemp seed extracts (5 µL) was performed on a Luna C18 column (150 mm × 3 mm i.d., 3 µm, Phenomenex, Torrance, CA, USA) using a linear mobile phase gradient of 0:90:10 to 90:0:10 [MeOH (A): water (C): acetonitrile +1% formic acid (D)] over 20 min at a flow rate of 400 µL $min^{-1}$. Ultraviolet data were recorded between 210 nm and 550 nm.

Mass spectrometric detection was performed in the positive and negative ionization modes using the full scan and data dependent $MS^2$ and $MS^3$ acquisition modes. Total ion current (TIC) chromatograms were obtained over the range of 125 – 1800 *m/z* using a spray voltage of +3.5 and −2.5 kV for the positive and negative ionization modes, respectively. Four different scan events were recorded for each ionization mode as follows: (1) full scan, (2) $MS^2$ of the most intense ion in *full scan*, (3) $MS^3$ of the most intense ion in $MS^2$, and (4) $MS^3$ of the second most intense ion in $MS^2$. Additional parameters for the MS included the following: full scan resolution set to 60,000 (full-width at half-maximum, FWHM), capillary temperature set to 350°C, ion transfer tube temperature set to 325°C, RF lens set to 50%, automatic gain control target set to 4.0 × 105 (*full scan*) or 1.0 × 104 ($MS^2$ and $MS^3$), intensity threshold set to 1.0 × $10^4$, collision-induced dissociation energy set to 35 eV, activation *Q* set to 0.25, and isolation window set to 4 *m/z*. Nitrogen was used as the drying, nebulizer and fragmentation gas.

## 2.3 LC-MS data processing and multivariate analysis

The LC-MS raw data was sliced into two sets according to the ionization mode (positive and negative) and transformed to mzXML format using the MSConvert package from the software ProteoWizard 3.0.9798 (Proteowizard Software Foundation, Palo Alto, CA, USA). Each data set was then processed by MZmine 2.53 (Pluskal et al., 2010) following the protocol previously described (Padilla-González et al., 2020a; Padilla-González et al., 2020b). After MZmine pre-processing, the results were exported as.csv and.mgf files which were then uploaded to the GNPS platform (https://gnps.ucsd.edu) for Feature Based Molecular Networking (FBMN) analysis (Wang et al., 2016; Nothias et al., 2020). The.csv file containing quantitative information related to ion abundances in each hemp seed sample was also submitted to multivariate statistical analysis by principal component analysis (PCA), hierarchical clustering analysis (HCA) and heatmaps in the software R 3.0.3 (R Foundation for Statistical Computing, Austria). Prior to the

multivariate analyses the data matrix was centered and scaled by unit variance to avoid statistical bias influenced by differences between high and low abundant metabolites. LC-MS files, R scripts and detailed parameters used in MZmine (available as an MZmine batch file) are freely available in the MassIVE repository (study identifier MSV000090725).

## 2.4 Feature-based molecular networking and annotation tools

Feature-based molecular networking (FBMN) was performed following the workflow by (Nothias et al., 2020) on the GNPS platform (https://gnps.ucsd.edu). For this analysis, the data were first filtered by removing all MS/MS fragment ions within a window of +/−17 Da of the precursor ion value and by choosing only the top 6 fragment ions in the +/− 50 Da window throughout the spectrum. The precursor and fragment ion mass tolerance were set to 0.05 Da and 0.99 Da respectively, in accordance with the MS settings used. A cosine score above 0.65 and more than six matched peaks were considered for creating the molecular network. Furthermore, edges between two nodes were kept in the network only if each of the nodes appeared in each other's respective top 10 most similar nodes. Experimental $MS^2$ spectra were then searched against the GNPS spectral libraries (Wang et al., 2016) using the same filters as for the input data. To enhance chemical structural information within the molecular network and provide a more comprehensive overview of the hemp seed chemical space, information from *in silico* structure annotations from the GNPS library search were incorporated into the network using the GNPS MolNetEnhancer and MS2LDA workflows (Wandy et al., 2018; Ernst et al., 2019). MolNetEnhancer and MS2LDA were performed using the default values reported in GNPS, except for bin width in MS2LDA, which was set at 0.1. Chemical class annotations were performed using the ClassyFire chemical ontology (Feunang et al., 2016). Final molecular networks including quantitative data, substructure information and chemical classifications were then visualized using the software Cytoscape (Ono et al., 2014).

Lastly, to confirm and expand the spectral library annotation made by molecular networking and its *in-silico* match against the GNPS spectral libraries, accurate mass values, $MS^2$ and $MS^3$ data of all the annotated metabolites were manually inspected and compared with information available in the literature to achieve different confidence values following the Metabolomics Standards Initiative (Sumner et al., 2007).

## 2.5 Isolation of new metabolites

One kilogram of seeds was acquired in a local supermarket and submitted to an MS-guided isolation protocol to target the isolation of a potentially new metabolite detected by LC-MS, following the same steps as done previously (Padilla-González et al., 2017). Briefly, the seeds were grinded into a fine power and extracted with 80% water in methanol for 24 h, with two repeat extractions of the same material, to exhaust the polar metabolome from the seed flour. Extracts were combined and then evaporated on a centrifugal evaporator (Genevac EZ-2, SP Industries, Warminster, PA, USA) to remove the organic solvent fraction and then lyophilized to remove water and yield a dry crude extract residue. Ten grams of the extract were submitted to Sephadex LH-20 column chromatography (120 g, 400 × 40 mm i.d.) employing mixtures of 200 mL of water–methanol (100:0, 98:2, 96:4, 94:6, and 50:50) that were combined to afford ten fractions (Fr1–Fr10) after LC-UV-MS monitoring for similarities. Fr2 was further purified by semipreparative HPLC (Waters e2695, Waters, Milford, MA, USA) using a Luna C18 column (250 × 10 mm i.d., 10 µm, Phenomenex) and a mobile phase of acetonitrile and water both acidified with 0.1% of formic acid. Chromatographic separation was performed using a linear gradient of 5 – 20% acetonitrile over 30 min, to yield 4 mg of ferulic acid 4-*O*-glucosyl-6'-*O*-sulphate. The structure of the isolated metabolite was determined using uni- and bidimensional NMR experiments (Bruker ARX 400, Billerica, MA, USA, MeOD) and by high-resolution MS (Orbitrap Fusion, Thermo Scientific) and low-resolution tandem mass spectrometry (Ion Trap Velos Pro, Thermo Scientific).

## 2.6 Extraction and derivatization of fatty acids

Powdered seeds (100 mg) from each hemp accession were extracted with 2 mL of *n*-hexane in an ultrasonic bath for 15 min at 25°C using a frequency of 40 kHz. After extraction, samples were evaporated under vacuum to obtain an oily residue corresponding to hemp seed fixed oil. Five microliters of the oil were added to a 2 mL glass vial containing 500 µL of boron trifluoride in methanol (20%). The vial was filled with nitrogen, tightly sealed and gently mixed. The sample vial was subsequently heated at 100 °C for 15 min in a heating block. After cooling to room temperature, 500 µL of water and 750 µL of pentane were added to the vial and vigorously shaken. The organic phase (upper) was then transferred to another vial and evaporated under nitrogen. This partition was repeated for a second time and the organic phases were mixed and evaporated under nitrogen. The concentrated sample containing fatty acids methyl esters was re-dissolved in 200 µL of *n*-hexane and placed in a glass vial with insert for GC-MS analysis.

## 2.7 GC-MS analyses of fatty acids methyl esters

GC-MS analyses of fatty acids methyl esters were performed in an Agilent 7890A GC chromatograph coupled to a single quadrupole MS analyser (Agilent 5975C). Chromatographic separation was performed on a ZB-WAX column (30 m x 0.25 mm i.d. x 0.25 µm, Phenomenex), using helium as carried gas at a flow rate of 1 mL per minute. The temperature program was configured from 70 – 250 °C at a linear increment of 3 °C per minute. The injection temperature was set to 220 °C and the source temperature to 180 °C. 70 eV was used as ionization energy and the mass spectrometer was configured to record data from 38 to 650 *m/*

*z*. Compounds were identified by comparing their retention indices (calculated against a series of *n*-alkanes analysed under the same experimental conditions) and by comparing their mass spectra with the NIST library.

The percentage composition of the fatty acid methyl esters detected in each hemp cultivar/accession was semi-quantitatively obtained by manual integration of GC-MS data. This information was then centered and scaled by unit variance and submitted to statistical analyses in the software R.

# 3 Results

## 3.1 Metabolic profiling by LC-MS/MS

### 3.1.1 Exploratory methods and mining of the hemp seed metabolome

Metabolic profiling by LC-MS and concomitant data processing by MZmine 2.53 of 52 hemp seed accessions detected 437 and 242 mass features in the positive and negative ionization modes, respectively. A principal component analysis (PCA) of the positive mode dataset explained 37.1% of the total variance in the first two components and grouped the seed extracts according to the similarities of their metabolic fingerprints (Figures 1A, B). This analysis revealed that while most accessions have similar metabolic profiles, the Italian cultivar Eletta Campana (CAN48), the Spanish accession Kongo Hanf (CAN58) and the French accession CAN37 (Table 1) are clear chemical outliers (Figure 1A).

A hierarchical clustering analysis with bootstrap resampling (HCAbp) grouped the 52 hemp seed accessions into five main clusters with support values >80% (clusters 1-5, Figure 1C). This analysis further confirmed that accessions CAN37 and CAN58 form an independent group chemically different from the remaining hemp seed accessions (cluster 1, Figure 1C), while the Italian cultivar Eletta Campana (CAN48) is closely related in chemical composition to the Croatian accession (CAN42) and the German cultivar Kompolti (CAN68) (Figure 1C). Similar results were obtained in the PCA and HCAbp of the negative mode dataset with accessions CAN37 and CAN58 and cultivars CAN48 and CAN68 as clear outliers based on their metabolic fingerprints (Figure S1). It is noteworthy that the chemical differences between the samples in groups 1 – 5 are visibly distinguishable by interrogation of the base peak chromatograms. To exemplify this, a representative sample from each group is provided in Figure 1D.

Close inspection of the five HCA-based groups reveals that in some cases a correlation of chemical composition with the hemp seed cultivar/accession occurs. For example, the five Fibrimon cultivars included in the present study clustered in group 5, while

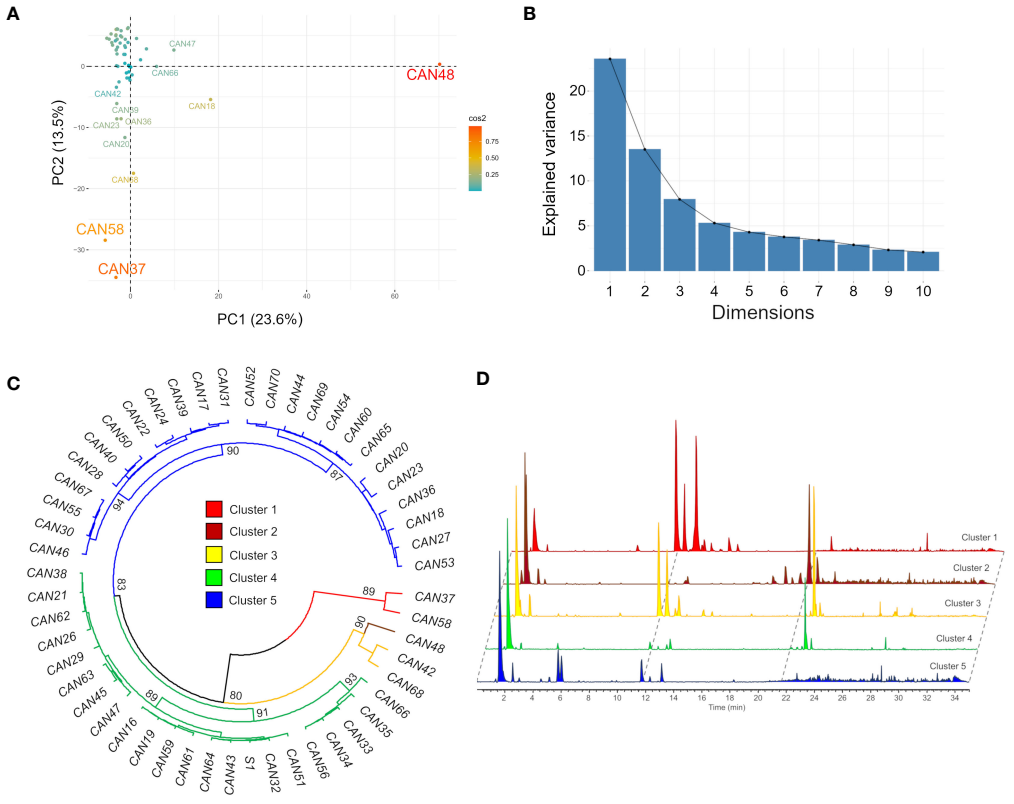



FIGURE 1
Clustering of the fifty-two hemp seed accessions based on the similarities of their metabolic fingerprints detected by LC-MS in positive ion mode. **(A)** PCA scores plot showing accessions CAN48, CAN58 and CAN37 as outliers; **(B)** Scree plot and the percentage of variance explained by the first 10 components; **(C)** HCAbp showing the metabolic clustering of the 52 hemp seed accessions in five main clusters highlighted in different colours and **(D)** Overlaid LC-MS chromatograms of one representative sample from each of the five groups.

all Cinepa accessions clustered in group 4. This pattern, however, was not consistent across all hemp cultivars, since the three investigated Kompolti samples were grouped in different clusters (Figure 1C). The correlation between chemical compositions and country of origin was also not evident. The eleven German accessions included in the present study were distributed across three different clusters, as were the four Italian accessions (Figure 1C).

A feature-based molecular networking analysis (FBMN) with *in silico* annotation tools of the positive mode LC-MS dataset revealed a clustering tendency by chemical class, where cinnamic acid amides (including lignanamides) and lipid-like molecules clustered the higher number of nodes (Figure 2A). Cinnamic acid amides and lipid-like molecules were consistently the most structurally diverse chemical classes in hemp seeds. Other classes detected include nucleosides and analogues, sugars and amino acids with derivatives (Figure 2A). The FBMN analysis, along with database searches and manual inspection of MS data, facilitated the identification of the majority of previously reported cinnamic acid amides in hemp (Table 2) although some nodes from the same chemical class were not identified. This analysis also revealed the accumulation patterns of cinnamic acid amides in the different groups of hemp seed accessions maintained *ex-situ* in the IPK germplasm collection (Figure 2B). As observed in Figure 2C, while

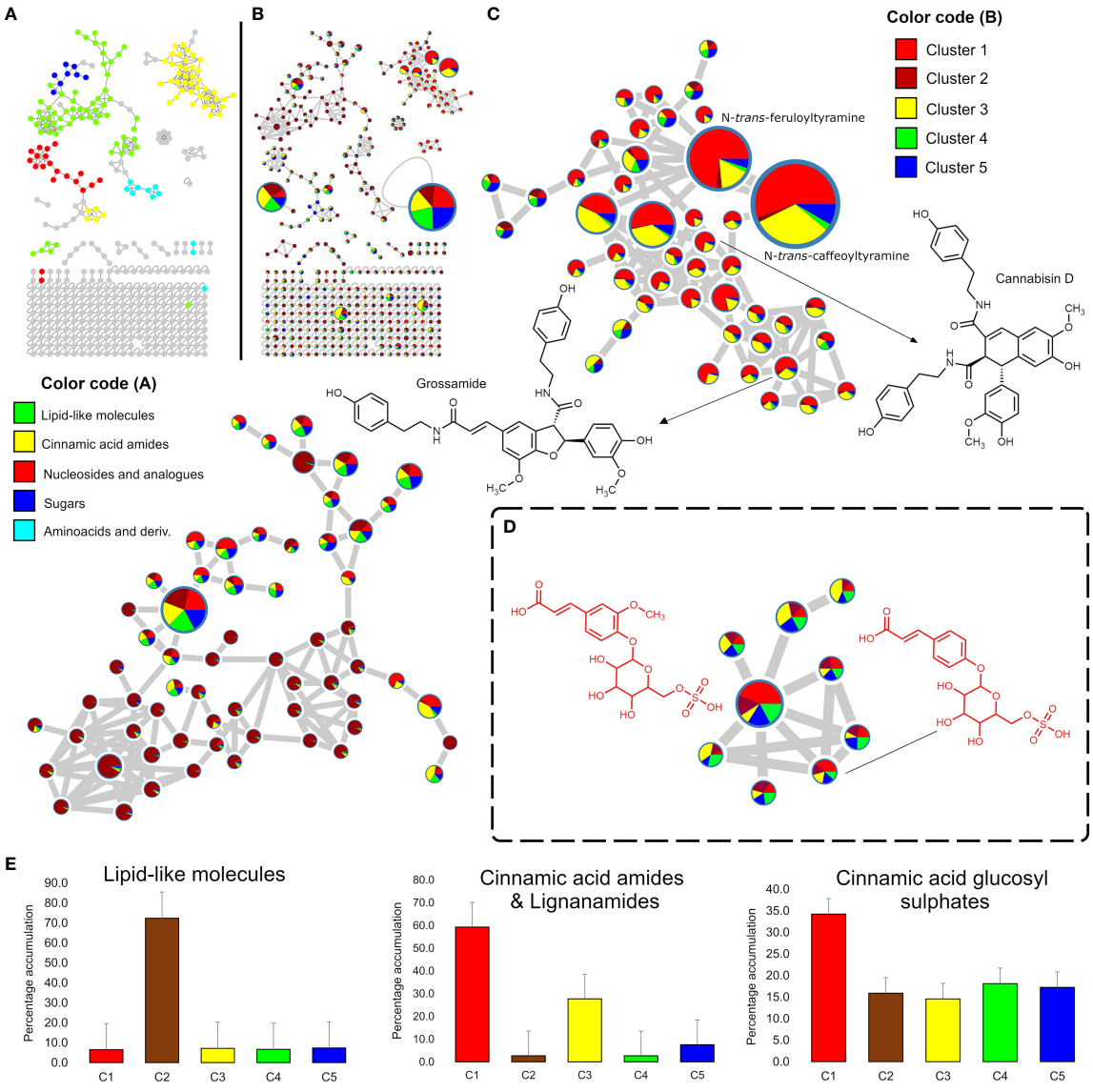



FIGURE 2
Mining of the hemp seed metabolome by Feature Based Molecular Networking (FBMN) of LC-MS data. **(A)** FBMN showing the chemical class assigned to each network through the MolNetEnhancer analysis of the positive ion mode LC-MS data; **(B)** FBMN showing pie charts representing the proportional accumulation of each node (representing *m/z* values) in the five groups identified by HCAbp (Figure 1C). The node size is proportional to the ion intensity. **(C)** Amplified nodes of cinnamic acid amides and lignanamides and lipid-like molecules detected in the positive ion mode dataset. **(D)** Amplified nodes of a previously unreported chemical class in hemp "Cinnamic acid glycosyl sulphates" detected only in the negative ion mode. **(E)** Bar charts summarizing the percentage accumulation of lipid-like molecules, cinnamic acid amides and lignanamides and cinnamic acid glycosyl sulphates in the five HCAbp clusters.

TABLE 2 Compounds annotated in a collection of 52 hemp seed accessions.

| # | Rt | Class | Annotation | Molecular formula | Positive ionization | Positive $MS^2$ ions (intensity) | Negative ionization | Negative $MS^2$ ions (intensity) | Error (ppm) | Reference | MSI[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.65 | Sugars | Sucrose | $C_{12}H_{22}O_{11}$ | $[M+NH_4]^+$ 360.15067 | 325 (100), 343 (46), 289 (17), 307 (11), 259 (7), 198 (7), 145 (5) | $[M - H]^-$ 341.10840 | 179 (100), 161 (21), 143 (16), 113 (14), 101 (7), 131 (6) | 0.04 | Current study | 1 |
| 2 | 2.50 | Organic acids | Citric acid | $C_6H_8O_7$ | – | – | $[M - H]^-$ 191.01965 | 111 (100), 172 (24) | 2.46 | Current study | 1 |
| 3 | 4.02 | Phenolic acids | Protocatechuic acid hexoside | $C_{13}H_{16}O_9$ | – | – | $[M - H]^-$ 315.07214 | 152 (100), 271 (21), 165 (13), 253 (12), 108 (12), 225 (9), 243 (7), 299 (5) | 1.69 | Kew database | 2 |
| 4 | 5.19 | Phenylpropionamides | Coumaroyl-amino butanol glucopyranoside | $C_{19}H_{27}O_8N$ | $[M + H]^+$ 398.18103 | 147 (100), 309 (65), 291 (14), 236 (14) | – | – | 1.17 | (Moccia et al., 2020) | 2 |
| 5 | 6.84 | Glucosyl sulphates | Ferulic acid 4-*O*-glucosyl-6'-*O*-sulphate (Cannabigail) | $C_{16}H_{20}O_{12}S$ | – | – | $[M - H]^-$ 435.06061 | 241 (100), 223 (5), 193 (5) | 0.78 | Current study | 1 |
| 6 | 5.88 | Glucosyl sulphates | Caffeic acid glucosyl sulphate | $C_{15}H_{18}O_{12}S$ | – | – | $[M - H]^-$ 421.04361 | 241 (100), 179 (10) | 1.10 | Current study | 3 |
| 7 | 6.23 | Glucosyl sulphates | Cinnamic acid glucosyl sulphate (Cuinnabis) | $C_{15}H_{18}O_{11}S$ | – | – | $[M - H]^-$ 405.04938 | 241 (100), 325 (5), 223 (4), 139 (4), 163 (3) | 0.81 | Current study | 3 |
| 8 | 7.85 | Glucosyl sulphates | Dimethoxy cinnamic acid glucosyl sulphate | $C_{17}H_{22}O_{13}S$ | – | – | $[M - H]^-$ 465.07037 | 241 (100), 223 (5) | 0.18 | Current study | 3 |
| 9 | 8.44 | Glucosyl sulphates | Ferulic acid glucosyl sulphate isomer | $C_{16}H_{20}O_{12}S$ | – | – | $[M - H]^-$ 435.05972 | 241 (100), 193 (5) | 1.26 | Current study | 3 |
| 10 | 10.31 | Cinnamic acid amides | N-*trans*-feruloyltyramine glycoside | $C_{24}H_{29}NO_9$ | $[M + H]^+$ 476.19138 | 314 (100), 177 (8), 145 (2) | $[M - H]^-$ 474.17648 | 312 (100) | 1.42 | (Leonard et al., 2021b) | 2 |
| 11 | 11.69 | Cinnamic acid amides | N-*trans*-caffeoyltyramine | $C_{17}H_{17}NO_4$ | $[M + H]^+$ 300.12317 | 163 (100), 121 (10), 138 (9) | $[M - H]^-$ 298.10846 | 135 (100), 178 (46), 161 (38), 256 (12), 107 (6) | 1.38 | (Leonard et al., 2021b) | 2 |
| 12 | 13.13 | Cinnamic acid amides | N-*trans*-feruloyltyramine | $C_{18}H_{19}NO_4$ | $[M + H]^+$ 314.13882 | 177 (100), 145 (15) | $[M - H]^-$ 312.12427 | 297 (100), 178 (92), 135 (44), 148 (18), 270 (15), 176 (12), 191 (8), 312 (6), 253 (5), | 1.32 | (Leonard et al., 2021b) | 2 |
| 13 | 12.90 | Cinnamic acid amides | N-*trans*-coumaroyltyramine | $C_{17}H_{17}NO_3$ | $[M + H]^+$ 284.12814 | 147 (100) | $[M - H]^-$ 282.11328 | 119 (100), 145 (80), 162 (50), 134 (21), 240 (12) | 1.85 | (Leonard et al., 2021b) | 2 |
| 14 | 14.89 | Cinnamic acid amides | tri-*p*-coumaroylspermidine | $C_{34}H_{37}N_3O_6$ | $[M + H]^+$ 584.27612 | 420 (100), 438 (80) | $[M - H]^-$ 582.26062 | 462 (100), 342 (12), 436 (9) | 0.09 | (Nigro et al., 2020) | 2 |
| 15 | 8.95 | Cinnamic acid amides | N-*trans*-caffeoyloctopamine | $C_{17}H_{17}NO_5$ | $[M + H]^+$ 316.11804 | 298 (100), 255 (10), 168 (8) | $[M - H]^-$ 314.10349 | 296 (100), 202 (10) | 1.44 | (Leonard et al., 2021b) | 2 |

*(Continued)*





10.3389/fpls.2023.1114398

TABLE 2 Continued

| # | Rt | Class | Annotation | Molecular formula | Positive ionization | Positive $MS^2$ ions (intensity) | Negative ionization | Negative $MS^2$ ions (intensity) | Error (ppm) | Reference | MSI[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 10.11 | Lignanamides | N- caffeoyltyramine/N-caffeoyloctopamine dimer | $C_{34}H_{32}N_2O_9$ | $[M + H]^+$ 613.21814 | 595 (100), 460 (12), 476 (4) | $[M - H]^-$ 611.20239 | 593 (100), 430 (26), 501 (10), 320 (10), 448 (7) | 0.75 | (Nigro et al., 2020) | 3 |
| 17 | 12.31 | Lignanamides | Cannabisin B | $C_{34}H_{32}N_2O_8$ | $[M + H]^+$ 597.22308 | 460 (100), 432 (4) | $[M - H]^-$ 595.20789 | – | 1.03 | (Leonard et al., 2021b) | 2 |
| 18 | 12.69 | Lignanamides | Cannabisin B isomer | $C_{34}H_{32}N_2O_8$ | $[M + H]^+$ 597.22290 | 460 (100), 432 (20), 350 (9) | $[M - H]^-$ 595.20746 | 485 (100), 432 (15) 322 (15), 456 (4) | 1.33 | (Leonard et al., 2021b) | 3 |
| 19 | 12.91 | Lignanamides | Cannabisin B isomer | $C_{34}H_{32}N_2O_8$ | $[M + H]^+$ 597.22369 | 460 (100), 297 (20), 434 (7), 418 (5), 279 (4) | $[M - H]^-$ 595.20728 | 485 (100), 432 (92) 322 (28), 458 (14), 416 (14), 269 (12) | 0.008 | (Leonard et al., 2021b) | 3 |
| 20 | 13.95 | Lignanamides | 3'3-demethylgrossamide | $C_{34}H_{32}N_2O_8$ | $[M + H]^+$ 597.22357 | 434 (100), 297 (40), 460 (32), 323(13) 271 (8), 279 (6) | $[M - H]^-$ 595.20837 | 432 (100), 458 (19), 269 (12), 338 (5) | 0.21 | (Leonard et al., 2021b) | 2 |
| 21 | 14.82 | Lignanamides | Cannabisin M | $C_{34}H_{32}N_2O_8$ | $[M + H]^+$ 597.22345 | 434 (100), 297 (78), 300 (44), 460 (39), 187 (8) | $[M - H]^-$ 595.20721 | 298 (100), 432 (5) | 0.41 | (Leonard et al., 2021b) | 2 |
| 22 | 14.96 | Lignanamides | Cannabisin Q | $C_{34}H_{32}N_2O_8$ | $[M + H]^+$ 597.22327 | 434 (100), 297 (95), 460 (55), 300 (30) 279 (4), 187 (4) | $[M - H]^-$ 595.20715 | 298 (100), 432 (5) | 0.71 | (Leonard et al., 2021b) | 2 |
| 23 | 12.24 | Lignanamides | Cannabisin A | $C_{34}H_{30}N_2O_8$ | $[M + H]^+$ 595.20770 | 458 (100), 350 (2) | $[M - H]^-$ 593.19159 | 456 (100), 430 (38), 293 (4) | 1.35 | (Nigro et al., 2020) | 2 |
| 24 | 12.04 | Lignanamides | Rac-sativamide A | $C_{33}H_{30}N_2O_8$ | $[M + H]^+$ 583.20758 | 565 (100), 446 (87), 418 (33), 255 (8) | $[M - H]^-$ 581.19086 | 400 (100), 513 (8), 535 (7), 418 (6), 443 (5) | 0.76 | (Leonard et al., 2021b) | 2 |
| 25 | 12.69 | Lignanamides | Cannabisin H | $C_{28}H_{31}NO_8$ | $[M + H]^+$ 510.21204 | 492 (100), 312 (42), 462 (35), 355 (37), 177 (8), 201 (7), 474 (6), 338 (5) | $[M - H]^-$ 508.19772 | 312 (100), 460 (68), 490 (8), 366 (6) | 1.48 | (Leonard et al., 2021b) | 2 |
| 26 | 14.43 | Lignanamides | Grossamide K | $C_{28}H_{29}NO_7$ | $[M + H]^+$ 492.20178 | 462 (100), 325 (37), 337 (14), 355 (8) | $[M - H]^-$ 490.18640 | 472 (100), 460 (13) | 0.90 | (Leonard et al., 2021b) | 2 |
| 27 | 13.04 | Lignanamides | Cannabisin C | $C_{35}H_{34}N_2O_8$ | $[M + H]^+$ 611.23846 | 474 (100), 446 (4) | $[M - H]^-$ 609.22308 | 446 (100), 283 (20), 499 (8) | 1.45 | (Leonard et al., 2021b) | 2 |
| 28 | 13.53 | Lignanamides | Cannabisin C isomer | $C_{35}H_{34}N_2O_8$ | $[M + H]^+$ 611.23859 | 474 (100), 311 (32), 448 (12), 432 (6) | $[M - H]^-$ 609.22260 | 446 (100), 283 (12), 485 (5), 322 (5) | 1.23 | (Leonard et al., 2021b) | 3 |
| 29 | 14.56 | Lignanamides | Cannabisin N or isomer | $C_{35}H_{34}N_2O_8$ | $[M + H]^+$ 611.24005 | 448 (100), 311 (34), 474 (29), 337 (19), 285 (8), 293 (5), 355 (4) | $[M - H]^-$ 609.22333 | 446 (100), 283 (15), 472 (5) | 1.45 | (Leonard et al., 2021b) | 3 |
| 30 | 15.48 | Lignanamides | Cannabisin N or isomer | $C_{35}H_{34}N_2O_8$ | $[M + H]^+$ 611.23877 | 448 (100), 311 (60), 474 (42), 314 (32), 187 (10), 298 (7) | $[M - H]^-$ 609.22229 | 312 (100), 417 (76), 296 (69), 446 (32), 501 (16), 581 (16), 591 (10) | 0.94 | (Leonard et al., 2021b) | 3 |
| 31 | 15.75 | Lignanamides | Cannabisin N or isomer | $C_{35}H_{34}N_2O_8$ | $[M + H]^+$ 611.23822 | 311 (100), 448 (78), 474 (56), 324 (6), 293 (6) | $[M - H]^-$ 609.22235 | 312 (100), 296 (62), 446 (38), 417 (12), 473 (10) | 1.84 | (Leonard et al., 2021b) | 3 |

*(Continued)*





10.3389/fpls.2023.1114398

TABLE 2 Continued

| # | Rt | Class | Annotation | Molecular formula | Positive ionization | Positive $MS^2$ ions (intensity) | Negative ionization | Negative $MS^2$ ions (intensity) | Error (ppm) | Reference | MSI[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 32 | 13.71 | Lignanamides | Cannabisin D | $C_{36}H_{36}N_2O_8$ | $[M + H]^+$ 625.25446 | 488 (100), 460 (6) | $[M - H]^-$ 623.23877 | 460 (100), 283 (7), 446 (7) | 0.85 | (Leonard et al., 2021b) | 2 |
| 33 | 14.21 | Lignanamides | Cannabisin G | $C_{36}H_{36}N_2O_8$ | $[M + H]^+$ 625.25409 | 488 (100), 460 (6) | $[M - H]^-$ 623.23907 | 460 (100), 446 (20), 336 (12), 499 (9) | 1.44 | (Leonard et al., 2021b) | 2 |
| 34 | 15.46 | Lignanamides | Cannabisin F | $C_{36}H_{36}N_2O_8$ | $[M + H]^+$ 625.25397 | 462 (100), 325 (41), 488 (10), 351 (9), 307 (5), 299 (5) | $[M - H]^-$ 623.23853 | 460 (100), 297 (20), 486 (12) | 1.64 | (Leonard et al., 2021b) | 2 |
| 35 | 16.08 | Lignanamides | Grossamide | $C_{36}H_{36}N_2O_8$ | $[M + H]^+$ 625.25433 | 488 (100), 462 (48), 351 (13), 325 (7) | $[M - H]^-$ 623.23871 | 592 (100), 433 (37), 551 (28), 471 (24), 460 (22), 503 (12), 515 (8) | 1.06 | (Leonard et al., 2021b) | 2 |
| 36 | 14.35 | Lignanamides | Cannabisin E | $C_{36}H_{38}N_2O_9$ | $[M + H]^+$ 643.26483 | 462 (100), 625 (19), 338 (7), 488 (5) | $[M - H]^-$ 641.24988 | 489(100), 623 (82), 431 (12), 312 (11), 460 (9), 328 (9), 517 (6) | 1.12 | (Leonard et al., 2021b) | 2 |
| 37 | 14.64 | Lignanamides | *cis*-Cannabisin E | $C_{36}H_{38}N_2O_9$ | $[M + H]^+$ 643.26495 | 462 (100), 625 (24), 338 (8), 478 (8), 450 (5) | $[M - H]^-$ 641.24908 | 489 (100), 623 (75), 431 (11), 312 (11), 460 (10), 517 (6), | 0.94 | (Leonard et al., 2021b) | 2 |
| 38 | 16.25 | Lignanamides | Cannabisin O | $C_{54}H_{53}N_3O_{12}$ | $[M + H]^+$ 936.36957 | 799 (100), 771 (5) | $[M - H]^-$ 934.35425 | 902 (100), 771 (80), 765 (54), 739 (38), 797 (25), 826 (10), 641 (8) | 1.23 | (Leonard et al., 2021b) | 2 |
| 39 | 21.82 | Cannabinoids | Cannabidiol or Tetrahydrocannabinol isomer | $C_{21}H_{30}O_2$ | $[M + H]^+$ 315.23193 | not fragmented | – | – | 1.51 | (Citti et al., 2019) | 3 |
| 40 | 23.35 | Cannabinoids | Tetrahydrocannabinol | $C_{21}H_{30}O_2$ | $[M + H]^+$ 315.23187 | 297 (28), 259 (100), 235 (40), 233 (42), 221 (32), 207 (22), 193 (82), 181 (32), 135 (32) | – | – | 1.70 | (Citti et al., 2019) | 3 |
| 41 | 22.16 | Cannabinoids | Tetrahydrocannabinolic acid | $C_{22}H_{30}O_4$ | $[M + H]^+$ 359.22186 | $[MS^2]$ 341 (100), $[MS^3]$ 299 (10), 285 (48), 261 (72), 247 (8), 233 (30), 219 (100), 207 (12), 135 (4) | – | – | 1.04 | (Citti et al., 2019) | 3 |
| 42 | 20.13 | Lipids | Unknown fatty acid | $C_{18}H_{32}O_4$ | $[M + H]^+$ 313.23755 | 295 (100), 277 (54), 227 (16), 209 (8), 259 (4) | – | – | 1.07 | – | 4 |
| 43 | 21.78 | Lipids | Unknown fatty acid | $C_{18}H_{30}O_3$ | $[M + H]^+$ 295.22678 | 277 (100), 257 (7), 237 (3), 221 (3) | – | – | 1.81 | – | 4 |
| 44 | 22.35 | Lipids | Unknown fatty acid | $C_{18}H_{32}O_3$ | $[M + H]^+$ 297.24258 | 279 (100), 261 (6), 243 (3), 163 (3) | – | – | 1.29 | – | 4 |

[a]Confidence level of identification according to the Metabolomics standards initiative (1: Rt, HRMS and $MS^2$ match with a pure standard; 2: HRMS and $MS^2$ match with literature information; 3: tentative identity suggested by interpretation of $MS^2$ spectra and literature information; unknown compounds).





10.3389/fpls.2023.1114398

most of the nodes representing cinnamic acid amides are accumulated in all hemp seed accessions, the Spanish accession Kongo Hanf (CAN58) and the French accession CAN37 in cluster 1 seem to accumulate the highest levels of this chemical class with close to 60% of the total ion intensities, while intermediate levels (*ca.* 30%) are found in the accessions in cluster 3 (CAN42 and CAN68) (Figure 2E). The remaining hemp seed accessions collectively account for only *ca.* 13% of the total ion counts for this chemical class (Figure 2E). Regarding the lipid-like molecules, the Italian cultivar Eletta Campana (CAN48) showed the highest accumulation of this chemical class with 72% of the total ion counts, while the remaining 51 accessions together represent only 28% of the total ion counts (Figure 2E).

FBMN of the negative ionization mode revealed similar patterns to the positive mode dataset with cinnamic acid amides clustering the higher number of nodes (Figure S2). However, several unidentified compounds were detected in this mode grouping into a previously unreported "molecular family" in hemp seeds (Figure 2D). Manual dereplication of the nodes belonging to this chemical class and further isolation and structure elucidation of targeted metabolites (see section 3.1.3) allowed the identification of this molecular family as "cinnamic acid glycosyl sulphates". This unusual chemical class includes several structurally related metabolites, characterized by the presence of a cinnamic acid moiety (or a related derivative) *para*-*O*-linked to a glycosyl sulphate unit (Figure 2D). Interestingly, all hemp seed accessions showed consistently similar amounts of glycosyl sulphates with 15% to 20% of the total ion counts, except for the accessions in cluster 1 (CAN37 and CAN58) which showed a percentage accumulation of *ca.* 34% (Figure 2E).

### 3.1.2 Metabolic differences among hemp seed accessions

Although the current study provides a snapshot of the metabolic composition of different hemp seed accessions obtained from plants grown in different conditions and geographic locations, we aimed to compare their metabolic profiles to link the presence of biologically important molecules to certain accessions in order to inform future studies. Therefore, to have a deeper understanding of semiquantitative differences among the annotated metabolites in each hemp seed accession, a heatmap analysis based on manually extracted peak intensities was performed (Figure 3). The annotation of these compounds was performed by HRMS and $MS^2$ spectral matching with literature information and with spectra available in the GNPS database and in our *in*-house library of $MS^2$ data. After *in*-silico spectral matching, manual inspection of annotated peaks was performed to confirm and expand the identifications. This combined analysis allowed the annotation of 44 metabolites belonging to five main chemical classes: Cinnamic acid glycosyl sulphates, cinnamic acid amides and their oxidative coupling products, lignanamides, cannabinoids and lipids (Table 2). These compounds were annotated with different levels of confidence according to the metabolomics standards initiative (Sumner et al., 2007).

#### 3.1.2.1 Cinnamic acid glycosyl sulphates

Detailed analyses of the metabolites assigned as cinnamic acid glycosyl sulphates allowed the annotation of four new compounds (5, 6, 8 and 9) and one known metabolite (compound 7). The assigned structures of compounds 6 – 9 are based on the absolute chemical structure of compound 5, ferulic acid 4-*O*-glucosyl-6'-*O*-sulphate (cannabigail), which was purified and assigned by 1D and 2D NMR interpretation in MeOD, supported by HRMS and $MS^2$ data (see section 3.2. for details about its structural elucidation). All the compounds in this chemical class (compounds 5 – 9, Table 2) showed similarities in their $MS^2$ spectra, characterized by the presence of a base peak ion at 241 *m/z* $[M–H]^-$ (Table 2), suggesting close similarities in their chemical structures. However, detailed analyses of MS data revealed that the phenylpropanoid moiety in

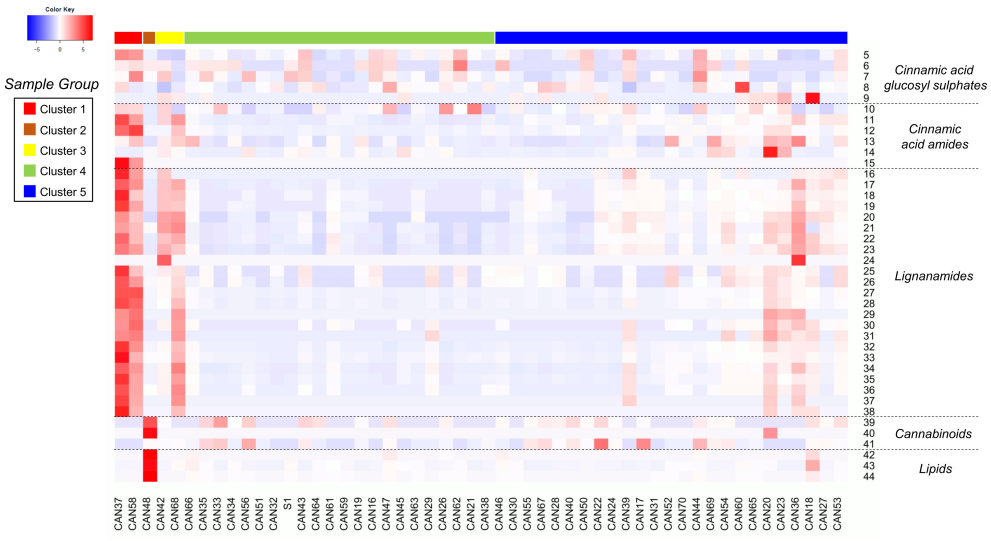



FIGURE 3
Heatmap showing the relative accumulation of the annotated metabolites in the 52 hemp seed accessions analysed by LC-MS. For compounds identities refer to Table 2.

these compounds differed, being ferulic acid in compounds 5 and 9, caffeic acid in 6, cinnamic acid in 7 (cuinnabis) and dimethoxycinnamic acid in 8. Analysis of accumulation patterns of this chemical class in the 52 hemp seed accessions showed that the accumulation of cinnamic acid glycosyl sulphates is rather homogeneous across all hemp seed accessions, although the Spanish Kongo Hanf (CAN58) and the French accession CAN37 (cluster 1) accumulate an overall higher proportion of these metabolites. Interestingly, an inverse tendency in the accumulation of compound 5 and its isomer, compound 9, was observed across accessions CAN37 and CAN18, while compounds 8 is preferentially accumulated in the Schurig cultivar (CAN60, Figure 3).

#### 3.1.2.2 Cinnamic acid amides and lignanamides

A total of six cinnamic acid amides and 23 lignanamides were identified in all hemp seed accessions based on the HRMS and $MS^2$ spectral match with information from the literature and online databases. N-*trans*-caffeoyltyramine (compound 11) and N-*trans*-feruloyltyramine (compound 12) represent the compounds accumulated in the highest proportion in hemp seeds out of all the detected polar metabolites. These two compounds are accumulated in the highest proportions in the French accession CAN37 and the Spanish accession Kongo Hanf (CAN58), although CAN37 accumulates higher amounts of N-*trans*-caffeoyltyramine relative to N-*trans*-feruloyltyramine, while the opposite is true for CAN58. Overall, CAN37 and CAN58 show the highest proportion of cinnamic acid amides and lignanamides among all hemp seed accessions (Figure 3), while an Argentinian accession (CAN51) and a Turkish accession (CAN26) both showed the lowest proportions for these two chemical classes, respectively (Figure 3). The Italian cultivar Eletta Campana (CAN48) and most of the accessions in cluster 4 accumulate the lowest proportion of cinnamic acid amides, while intermediate levels were found in accessions in cluster 3 (CAN42 and CAN68) followed by accessions in cluster 5 (Figure 3).

Similar to the monomeric cinnamic acid amides, lignanamides are accumulated in higher proportions in accessions CAN37 and CAN58, followed by the Croatian and German accessions CAN42 and CAN68, respectively, in cluster 3 and those in cluster 5 (Figure 3). Lignanamides constitute the most diverse group of metabolites in hemp seeds, among which cannabisin B (compound 17) and cannabisin C (compound 27) represent the two most abundant metabolites of this chemical class. Compounds in this chemical class are generally found as isomeric molecules apparently distinguishable by the intensity of their diagnostic $MS^2$ ions (Table 2), although no study has been performed to validate this hypothesis. Interestingly, although accessions CAN42 and CAN68 are closely related based on their overall chemical compositions (Figure 1C), the lignanamides profile of these two accessions showed salient differences (Figure 3). While both accessions accumulate similar proportions of compounds 16 – 23, accession CAN68 has a much higher proportion of cannabisins D, E, F, G, N and O, relative to CAN42 (compounds 25 – 38, Figure 3). Interestingly, while most cinnamic acid amides and lignanamides seem to be preferentially accumulated in accessions in cluster 1 (CAN37 and CAN58), compounds 13 (N-*trans*-coumaroyl tyramine) and 14 (tricoumaroyl spermidine) were accumulated mostly in accessions CAN36 and CAN20, respectively, both in cluster 5.

#### 3.1.2.3 Cannabinoids and lipid-like molecules

Considering the previous reports and current regulations on the presence of cannabinoids in hemp, we performed target mass searches for the main cannabinoids previously reported in hemp seed oil and organic extracts (Citti et al., 2019; Jang et al., 2020). Trace amounts (< 0.3%) of only three cannabinoids were detected in 39 of the 52 hemp seed accessions: cannabidiol (CBD), tetrahydrocannabinol (THC) and tetrahydrocannabinolic acid (THCA). THC and CBD were detected in comparatively higher proportions in the Italian cultivar Eletta Campana (CAN48, Figure 3), while THCA was mainly detected in the accessions belonging to cluster 5 (especially CAN17 and CAN22), as well as in a Kompolti cultivar (CAN56, Figure 3).

Similarly, three unidentified lipid-like molecules (compounds 42 – 44, Figure 3) were also detected mainly in the Italian Eletta Campana (CAN48), although trace amounts were also found in other accessions. Manual inspection of MS data indicated that these metabolites likely correspond to oxidated fatty acids. However, their structural characterization is yet to be determined.

## 3.2 Structural characterization of new cinnamic acid glycosyl sulphates

Considering that the analysis by FBMN and our dereplication approach suggested the presence of several potentially new metabolites belonging to a previously unreported molecular family in hemp (cinnamic acid glycosyl sulphates), we performed a target isolation of the most abundant metabolite from this chemical class, detected at 6.84 min with a mass of 435.06061 *m/z* $[M–H]^-$ (Table 2). The process started with a dried 20% methanolic extract of hemp seeds, which was submitted to the classic isolation processes using Sephadex LH-20 column chromatography followed by semipreparative HPLC (see Materials and Methods section). Approximately 4 mg of ferulic acid 4-*O*-glucosyl-6'-*O*-sulphate (compound 5) was afforded. Furthermore, the tentative structures of four additional cinnamic acid glycosyl sulphates (compounds 6 – 9; Figure 4) in the crude extracts are tentatively reported based on similarity to 5 according to HRMS and $MS^2$ data. All the compounds in this chemical class (compounds 5 – 9, Figure 4) showed a characteristic base peak ion in their $MS^2$ spectra at 241 *m/z* $[M–H]^-$, representing a deprotonated glucopyranosyl-sulphate unit formed after the neutral loss of the phenylpropanoid moiety. A subsequent neutral loss of a water molecule generated the ion at 223 *m/z* (Figure 4). In addition to these diagnostic ions, minor peaks representing a deprotonated ferulic acid (193 *m/z*), caffeic acid (179 *m/z*), cinnamic acid (163 *m/z*) and dimethoxycinnamic acid unit (223 *m/z*), were observed in the $MS^2$ spectra of compounds 5, 6, 7 and 8, respectively. The suggested fragmentation mechanism leading to the formation of these diagnostic ions is provided in Figure 4.

Compounds 5 and 9 were assigned as two isomers of the previously undescribed ferulic acid glucosyl sulphate metabolite,

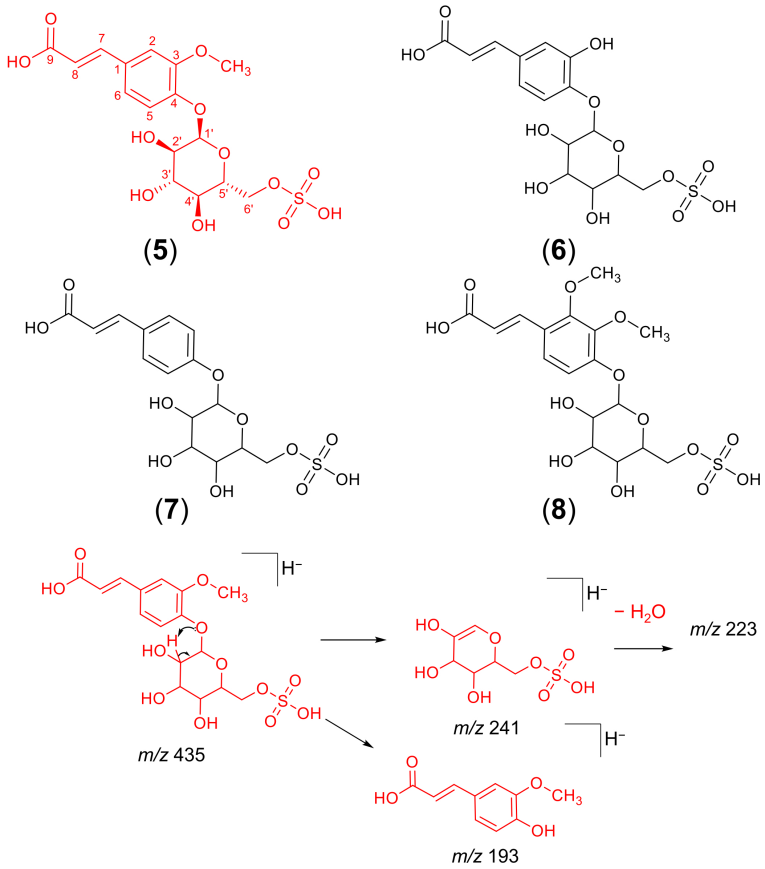



FIGURE 4
Chemical structures of cinnamic acid glycosyl sulphates detected in hemp seed and fragmentation mechanism of the isolated ferulic acid glucosyl sulphate. The chemical structure of ferulic acid-4-*O*-glucosyl-6'-*O*-sulphate (compound 5, in red) was confirmed by NMR experiments, while the chemical structures of the remaining compounds (6 – 9) were suggested by interpretation of HRMS and $MS^2$ data alone and, therefore, await spectroscopic confirmation.

based on chromatographic and spectrometric data (HRMS and $MS^2$). Both compounds showed a deprotonated molecule $[M–H]^-$ at *m/z* 435.0606 (Table 2), consistent with the molecular formula of $C_{16}H_{20}SO_{12}$ (calculated for $C_{16}H_{19}SO_{12}$, 435.0603). Their online UV spectra displayed absorbance maxima at 289 and 315 nm, characteristic of phenylpropanoid derivatives.

The 1H NMR spectrum of compound 5 included a proton spin system consistent with a six membered meta-para-disubstituted aromatic ring (Table 3), and two adjacent exocyclic olefinic monoprotonated carbons with a coupling constant (15.8 Hz) consistent with the trans configuration. The 13C spectrum is consistent with that of a substituted free ferulic acid moiety. Positions 3 and 4 of the ferulic acid moiety were assigned by HMBC couplings with other protons on the ring, and by an online NMR shifts simulator (nmrdb.org). The position of the methoxy singlet was established by a HMBC coupling to the carbon at position 3, and an NOE coupling to the proton doublet (1.2 Hz) at position 2. The attachment of the glucosyl moiety was established by a HMBC coupling from the anomeric proton to the fully substituted carbon at position 4.

The glucosyl moiety was assigned using COESY couplings, coupling constants and HSQC. The position of attachment of the sulphate was confirmed by the re-shielding of the carbon at position 6', which is normally measured at 61 – 62 ppm in similar glycosyl-arene structures (Schuster et al., 1986).

## 3.3 METABOLIC profiling by GC-MS

### 3.3.1 Fatty acid profiles of different hemp seed accessions

Metabolic profiling by GC-MS of the derivatized oil extracted from 52 hemp seed accessions demonstrated homogeneous chemistry with a predominance of linoleic acid (Table 4). Six fatty acids were detected as the major oil components in all accessions: palmitic acid, oleic acid, linoleic acid, γ-linolenic acid, α-linolenic acid and stearidonic acid. Integration of chromatographic peak areas showed that linoleic acid (53.65% ± 2.18), α-linolenic acid (20.18% ± 3.30) and oleic acid (13.62% ± 1.86) are the three main fatty acids, representing more than 87% of the total oil components (Table 4). The ratio of linoleic acid and α-linolenic acid among all hemp seed accessions was found to be 2.66:1, with accessions CAN19, CAN32, CAN43, CAN44, CAN50, CAN53 and CAN54 showing the closest values to 3:1, reported as optimal for human nutrition (Table 4).

TABLE 3 NMR data of ferulic acid 4-O-glucosyl-6′-O-sulphate (compound 5) in MeOD.

| Feruloyl | | | | Glucose | | |
|---|---|---|---|---|---|---|
| No. | $^{13}C$ | 1H | HMBC, H→C | No. | $^{13}C$ | 1H |
| 1 | 130.73 | – | – | 1'* | 102.34 | 4.94, d (3.5 Hz) |
| 2 | 112.37 | 7.22, d (1.2 Hz) | 1, 3 | 2' | 74.8 | 3.52, dd (3.5, 8.4 Hz) |
| 3 | 150.98 | – | – | 3' | 71.27 | 3.42, dd (7.6, 8.4 Hz) |
| $OCH_3$ | 56.8 | 3.82, 3H s | 3 | 4' | 77.58 | 3.48, dd (7.5, 7.6 Hz) |
| 4 | 149.8 | – | – | 5' | 71.27 | 3.68, dd (5.9, 7.5 Hz) |
| 5 | 116.3 | 6.69, d (7.9 Hz) | 6 | 6' | 68.15 | 4.13, dd (5.9, 11.1 Hz), |
| 6 | 130.95 | 7.03, dd (1.2, 7.9 Hz) | 1 | | | 4.36, d (11.1 Hz) |
| 7 | 145.13 | 7.56, d (15.8 Hz) | 2, 8, 9 | | | |
| 8 | 119.3 | 6.40, d (15.8 Hz) | 1 | | | |
| 9 | 171.2 | – | – | | | |

*HMBC coupling observed from the proton on the anomeric 1' to position 4 of the feruloyl moiety.

TABLE 4 Percentage composition of the main fatty acids found in a collection of 52 hemp seed accessions.

| Accession | Palmitic acid | Oleic acid | Linoleic acid | γ-Linolenic acid | α-Linolenic acid | Stearidonic acid | LA: αLA ratio |
|---|---|---|---|---|---|---|---|
| CAN16 | 7.17 | 12.89 | 52.38 | 5.89 | 18.87 | 2.80 | 2.78 |
| CAN17 | 7.55 | 11.20 | 53.11 | 4.54 | 21.78 | 1.81 | 2.44 |
| CAN18 | 9.05 | 16.45 | 50.19 | 2.96 | 20.01 | 1.33 | 2.51 |
| CAN19 | 7.49 | 13.54 | 54.27 | 5.27 | 17.57 | 1.86 | 3.09 |
| CAN20 | 7.36 | 12.12 | 53.80 | 3.37 | 21.55 | 1.80 | 2.50 |
| CAN21 | 7.29 | 13.72 | 53.24 | 2.62 | 21.76 | 1.37 | 2.45 |
| CAN22 | 7.76 | 12.37 | 51.87 | 3.07 | 23.16 | 1.77 | 2.24 |
| CAN23 | 6.11 | 14.20 | 51.69 | 0.00 | 28.00 | 0.00 | 1.85 |
| CAN24 | 8.33 | 14.38 | 52.91 | 0.00 | 23.39 | 0.99 | 2.26 |
| CAN26 | 8.70 | 12.60 | 54.88 | 0.00 | 22.67 | 1.15 | 2.42 |
| CAN27 | 8.23 | 12.59 | 51.53 | 4.03 | 21.26 | 2.38 | 2.42 |
| CAN28 | 7.70 | 12.55 | 51.22 | 4.04 | 22.83 | 1.66 | 2.24 |
| CAN29 | 8.35 | 11.93 | 54.67 | 2.70 | 20.91 | 1.43 | 2.61 |
| CAN30 | 7.83 | 17.28 | 55.14 | 4.90 | 13.46 | 1.39 | 4.10 |
| CAN31 | 7.14 | 11.43 | 52.78 | 5.52 | 20.38 | 2.76 | 2.59 |
| CAN32 | 7.99 | 13.44 | 54.79 | 4.07 | 18.23 | 1.48 | 3.01 |
| CAN33 | 8.03 | 11.16 | 53.35 | 4.13 | 21.35 | 1.97 | 2.50 |
| CAN34 | 6.59 | 12.14 | 58.79 | 4.44 | 16.45 | 1.59 | 3.57 |
| CAN35 | 6.70 | 13.98 | 59.29 | 0.00 | 18.19 | 1.84 | 3.26 |
| CAN36 | 7.30 | 9.46 | 52.02 | 3.05 | 26.28 | 1.91 | 1.98 |
| CAN37 | 6.91 | 12.63 | 50.47 | 0.00 | 28.19 | 1.80 | 1.79 |
| CAN38 | 7.76 | 14.89 | 54.34 | 4.56 | 16.62 | 1.82 | 3.27 |
| CAN39 | 6.80 | 12.01 | 55.82 | 0.00 | 23.21 | 2.16 | 2.40 |

*(Continued)*

10.3389/fpls.2023.1114398

TABLE 4 Continued

| Accession | Palmitic acid | Oleic acid | Linoleic acid | γ-Linolenic acid | α-Linolenic acid | Stearidonic acid | LA: αLA ratio |
|---|---|---|---|---|---|---|---|
| CAN40 | 8.60 | 11.98 | 48.84 | 3.93 | 24.48 | 2.16 | 2.00 |
| CAN42 | 7.87 | 13.89 | 59.15 | 0.00 | 17.81 | 1.28 | 3.32 |
| CAN43 | 6.67 | 12.77 | 54.59 | 5.36 | 18.51 | 2.10 | 2.95 |
| CAN44 | 7.77 | 15.45 | 53.34 | 4.08 | 17.89 | 1.47 | 2.98 |
| CAN45 | 8.07 | 15.14 | 52.80 | 3.43 | 19.33 | 1.23 | 2.73 |
| CAN46 | 7.42 | 14.37 | 54.44 | 4.71 | 17.30 | 1.76 | 3.15 |
| CAN47 | 8.29 | 15.24 | 51.05 | 2.52 | 21.86 | 1.04 | 2.34 |
| CAN48 | 8.46 | 17.22 | 54.84 | 2.68 | 15.70 | 1.09 | 3.49 |
| CAN50 | 7.29 | 14.13 | 53.24 | 5.79 | 17.82 | 1.72 | 2.99 |
| CAN51 | 7.38 | 13.90 | 56.52 | 0.00 | 20.74 | 1.46 | 2.72 |
| CAN52 | 7.56 | 15.73 | 52.62 | 3.20 | 19.61 | 1.28 | 2.68 |
| CAN53 | 7.71 | 12.40 | 53.43 | 6.44 | 17.83 | 2.19 | 3.00 |
| CAN54 | 7.84 | 14.01 | 52.65 | 5.48 | 18.02 | 2.00 | 2.92 |
| CAN55 | 7.88 | 15.48 | 54.27 | 5.00 | 15.88 | 1.49 | 3.42 |
| CAN56 | 6.71 | 15.01 | 55.04 | 0.00 | 21.92 | 1.32 | 2.51 |
| CAN58 | 7.34 | 12.48 | 51.89 | 0.00 | 27.17 | 1.11 | 1.91 |
| CAN59 | 7.80 | 11.06 | 54.49 | 0.00 | 25.12 | 1.53 | 2.17 |
| CAN60 | 7.41 | 12.56 | 52.74 | 5.88 | 18.98 | 2.43 | 2.78 |
| CAN61 | 8.18 | 19.67 | 53.29 | 3.14 | 14.97 | 0.75 | 3.56 |
| CAN62 | 7.64 | 12.65 | 55.40 | 6.48 | 15.80 | 2.03 | 3.51 |
| CAN63 | 8.35 | 14.80 | 53.31 | 3.22 | 19.19 | 1.12 | 2.78 |
| CAN64 | 9.00 | 14.63 | 51.62 | 3.05 | 20.07 | 1.64 | 2.57 |
| CAN65 | 7.90 | 13.88 | 55.35 | 5.22 | 15.89 | 1.77 | 3.48 |
| CAN66 | 7.87 | 12.24 | 52.74 | 3.85 | 21.61 | 1.68 | 2.44 |
| CAN67 | 8.06 | 10.87 | 50.71 | 5.93 | 21.86 | 2.58 | 2.32 |
| CAN68 | 7.48 | 16.30 | 51.70 | 3.00 | 20.27 | 1.25 | 2.55 |
| CAN69 | 7.75 | 13.97 | 54.08 | 3.48 | 18.97 | 1.75 | 2.85 |
| CAN70 | 7.29 | 13.55 | 54.86 | 2.68 | 20.39 | 1.24 | 2.69 |
| S1 | 8.70 | 13.80 | 58.08 | 0.00 | 18.19 | 1.22 | 3.19 |
| Mean | 7.70 ± 0.62 | 13.62 ± 1.86 | 53.65 ± 2.18 | 3.23 ± 2.06 | 20.18 ± 3.30 | 1.63 ± 0.51 | 2.66 |

LA, Linoleic acid; αLA, α-Linolenic acid.

Chromatographic peak area comparisons revealed that the Italian cultivar Eletta Campana (CAN48) represents the accession with the highest accumulation of fatty acids, followed by samples of the cultivar "Fibrida" (CAN52) and "Forose" (CAN28). However, these cultivars showed interesting differences in the accumulation of specific metabolites. CAN48 showed the highest concentration of palmitic acid, oleic acid and linoleic acid, while CAN28 showed the highest concentration of γ-linolenic acid, α-linolenic acid and stearidonic acid. A boxplot analysis based on the fatty acids composition further confirmed that CAN48 and CAN28 are clear outliers in the accumulation of specific lipophiles and the mean values for the six major fatty acids (Figure 5A). Further analysis of the same dataset by PCA revealed that the fatty acid composition of the 52 hemp seed accessions is rather homogeneous with only two PCs accounting for more than 92% of the total variance (Figures 5B, C). The chemical structures of the main fatty acids detected in hemp seeds are reported in Figures 5D.

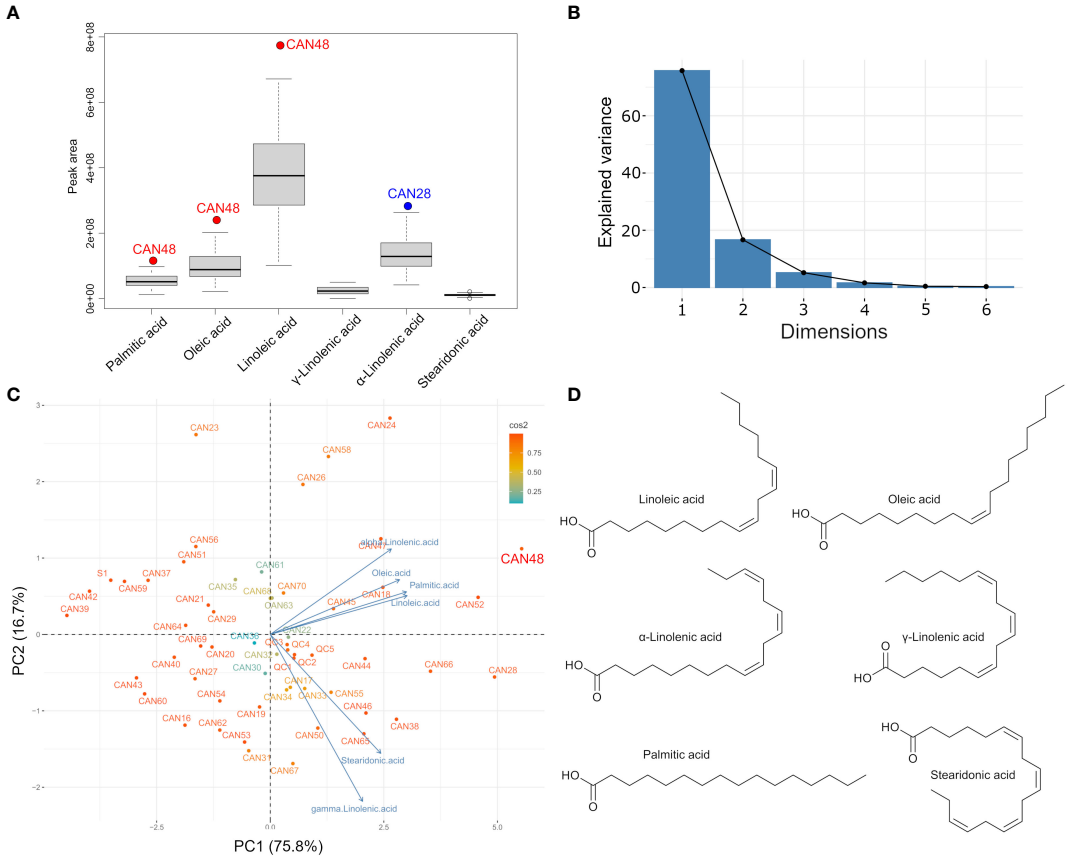



FIGURE 5
Fatty acid profile and clustering of the 52 hemp seed accessions based on their fatty acid compositions. **(A)** Boxplot showing the mean accumulation major fatty acids detected in all hemp seed accessions; **(B)** Scree plot and percentage of explained variance by the first 6 components; **(C)** PCA biplot showing the fatty acids-based clustering of all hemp seed accessions and **(D)** Chemical structures of the main fatty acids detected in hemp seeds.

# 4 Discussion

Hemp is an industrial crop that has been widely investigated due to its diversity of applications in the pharmaceutical, nutraceutical, food, cosmetic, textile, and material industries (Schluttenhofer and Yuan, 2017; Farinon et al., 2020). However, the chemical diversity of the hemp seed metabolome has not been explored in detail and new metabolites with unknown properties are likely to be discovered. In the current study, following an untargeted metabolomics approach by LC-MS and GC-MS, we discovered a previously unreported molecular family in hemp seeds "cinnamic acid glycosyl sulphates" and linked the presence of biologically important molecules to certain accessions. Chemical knowledge of these cultivars and accessions could be used as a base in future to optimize commercial cultivars in the context of nutritional and other functional uses.

Metabolic fingerprinting by LC-MS and data mining tools revealed that the 52 hemp seed accessions can be categorized into five main groups based on the similarities of their metabolic fingerprints. Although it is expected that when growing the same seeds under homogeneous conditions there will be some differences in the metabolic grouping, we believe that we can only have an accurate representation of the true metabolic phenotype of an organism by studying the chemical composition of plants/organs obtained from individuals collected in their natural environment. Germplasm and herbarium collections provide valuable samples for this. In the current study, among all the studied accessions and cultivars, the Spanish accession Kongo Hanf (CAN58) and the French accession CAN37 showed a remarkable structural diversity of bioactive and novel metabolites, some of them accumulated in high amounts. These two varieties showed the highest accumulation of phenylpropionamides (including cinnamic acid amides and their oxidative coupling products lignanamides), as well as new cinnamic acid glycosyl sulphates.

Phenylpropionamides (cinnamic acid amides and lignanamides) represents the main chemical class of bioactive metabolites in hemp seeds (Moccia et al., 2020; Leonard et al., 2021a; Leonard et al., 2021b). These compounds display potent antioxidant, anti-inflammatory, cytotoxic and acetylcholinesterase inhibitory activities, both *in vitro* and *in vivo* studies (Yan et al., 2015; Farinon et al., 2020; Moccia et al., 2020; Leonard et al., 2021a). For instance, previous studies have recognized the strong potential of hemp seed phenylpropionamides as protective agents against human chronic diseases (Leonard et al., 2021a), although results from clinical trials or cohort studies are not available to support

this. Our results indicate that N-*trans*-caffeoyltyramine (compound 11) and N-*trans*-feruloyltyramine (compound 12) constitute the main phenylpropionamides accumulated in hemp seeds, especially in the Spanish accession Kongo Hanf (CAN58) and the French accession CAN37. These two compounds have demonstrated strong antioxidant activities (Chen et al., 2012; Moccia et al., 2020) and *in vivo* anti-neuroinflammatory action (Maiolo et al., 2018; Zhou et al., 2018). Furthermore, the lignanamides fraction of hemp seeds are also highly active. Like compounds 11 and 12, lignanamides (compounds 16 – 38) are also accumulated in high amounts in the accession CAN58 and CAN37.

Recent studies by Chen et al. (2012) observed stronger DPPH scavenging activity by cannabisin B (compound 17) isolated from hemp seed than common soy (*Glycine max* (L.) Merr.) isoflavones and secoisolariciresinol diglucoside from flax seed (*Linum usitatissimum* L.). Cannabisins M, A, C, D, and 3,3'-demethylgrossamide (compounds 21, 23, 27, 32 and 20, respectively) from hemp seed also displayed potent antioxidant activities *in vitro* (Yan et al., 2015). These compounds are especially accumulated in accessions CAN37, CAN58, CAN68 and CAN42.

The anti-inflammatory properties that these and other hemp seed lignanamides display have also been widely recognized. For example, grossamide, (compound 35) accumulated in higher amounts in accession CAN37 (Figure 3), showed a significant reduction in the production of the proinflammatory cytokines interleukin-6 (IL-6) and tumor necrosis factor-a (TNF-a) at a concentration of 10 – 20 μM in a lipopolysaccharide (LPS) induced model (Luo et al., 2017). These cytokines have been linked to the stimulation of inflammatory and auto-immune processes in many diseases such as rheumatoid arthritis, multiple sclerosis, psoriasis as well as different types of cancer (Bradley, 2008; Blauvelt, 2017; Shang et al., 2017). Like grossamide, cannabisin F (compound 34) has also been reported to reduce the production of IL-6 and TNF-a in LPS-induced models, although through a different mechanism (Wang et al., 2019).

Regarding the cytotoxic effects of hemp seed lignanamides, previous studies have reported that it varies in accordance with their chemical structures. For example, cannabisin B has been shown to induce autophagic cell death on HepG2, a human liver cancer cell line (Chen et al., 2013), while cannabisin F is especially active against SH-SY5Y, a human neuroblastoma cell line (Li et al., 2019). Nevertheless, other hemp seed lignanamides, such as grossamide, cannabisin D and cannabisin G, have displayed weak cytotoxic activity against human prostate cancer LNCaP cells (Ma et al., 2002). Considering that CAN37, CAN58, CAN68 and CAN42 have a higher concentration of bioactive cinnamic acid amides and lignanamides, these accessions could be commercially prioritized as potential sources of antioxidant, anti-inflammatory and cytotoxic metabolites or in selective breeding programs to improve the concentration of specific metabolites in other commercialized cultivars. Interestingly, none of the five historic cultivars with deleted status in the EU database (Table 1) showed a particularly high concentration of bioactive metabolites compared to the newer accessions.

In addition to the 29 phenylpropionamide peaks annotated in the current study (Table 2), our results suggest that other lignanamides might also be present. Manual inspection of the FBMN results indicate that there are at least eleven additional nodes representing potentially new phenylpropionamides, some of them showing *m/z* values previously unreported in hemp seeds. For example, two nodes at *m/z* 600.270 and *m/z* 614.286 $[M+H]^+$ appear to be structurally related to tricoumaroyl spermidine (compound 14). Manual interpretation of MS data suggests that the node at *m/z* 600.270 $[M+H]^+$ likely represents a hydroxylated analogue, while the node at *m/z* 614.286 $[M+H]^+$ is likely its methoxylated version. Several isomers of *m/z* 597.33 and *m/z* 611.23 $[M+H]^+$ were also observed. The isolation and complete structural characterization of these potentially new metabolites is still necessary. Recent studies have proved that despite being exhaustively commercialized, hemp seed continues to be a promising source of new lignanamides. For example, Yan et al. (2015) identified the first trimeric feruloyltyramine in hempseeds, of which its coupling resulted in the formation of an 8-*O*-4 linkage. Considering that hemp hulls are the main site of accumulation of lignanamides, future agricultural waste valorisation programs are highly encouraged. Hemp hulls are a by-product of the hemp seed oil industry (Pojić et al., 2014; Mattila et al., 2018), which represents an opportunity. This is further supported by the fact that the total lignanamides in hemp hull can be up to four times higher than in whole hemp seed (3.36 mg/g vs 0.77 mg/g, respectively) (Mattila et al., 2018).

Aside from phenylpropionamides, hemp seed oil is also a commodity held in high regard for its nutraceutical properties due to a perfectly balanced content of omega 3 and omega 6 polyunsaturated fatty acids (Crescente et al., 2018; Citti et al., 2019). Mean concentration values for the six main fatty acids detected in the current study: palmitic acid oleic acid, linoleic acid, γ-linolenic acid, α-linolenic acid and stearidonic acid (Table 4), are within ranges comparable to previous studies (Galasso et al., 2016; Pavlovic et al., 2019). The Italian cultivar Eletta Campana (CAN48) showed the highest accumulation of fatty acids among the investigated hemp seed accessions. However, CAN32, CAN50 and CAN53 were the accessions with the ratio of linoleic acid: α-linolenic acid closest to 3:1, reported as optimal for human nutrition (Leizer et al., 2000; Simopoulos, 2008; Galasso et al., 2016). These results agree and further expand the studies of Galasso et al. (2016). Similar to previous studies, the cultivar with the lowest proportion of α-linolenic acid was Eletta Campana (CAN48) (Galasso et al., 2016). However, while the mean concentration values for the six fatty acids are also within comparable ranges to previous reports (Galasso et al., 2016; Pavlovic et al., 2019), there are some discrepancies in the concentration of specific fatty acids in certain accessions. For example, we found a higher proportion of oleic acid in CAN48 than in CAN39 and the higher percentage of α-linolenic acid was found in CAN40 instead of CAN24 as reported by Galasso et al. (2016). These discrepancies can be due to different extraction methods and the fact that Galasso et al. (2016) studied the oil composition of seeds coming from individuals grown under the same conditions while our study directly analysed the seeds stored in the IPK genebank.

In addition to the main fatty acids previously mentioned, other fatty acids have also been reported in minor concentrations in hemp

seed oil (Galasso et al., 2016; Pavlovic et al., 2019). These metabolites were not reported in the current paper due to identification and quantification discrepancies in these low concentrated metabolites. However, considering that the biological properties of hemp seed oil are related to the presence of the major fatty acids, especially the ratio between linoleic acid and α-linolenic acid, our results are biologically relevant. Complete reports of minor metabolites in hemp seed oil can be found in other studies (Galasso et al., 2016; Pavlovic et al., 2019).

A recent study reported the cannabinoid profile of ten commercially available hemp seed oils (Citti et al., 2019). Besides tetrahydrocannabinol and cannabidiol, another 30 cannabinoids were also identified (Citti et al., 2019). In the current study, we found traces of only three cannabinoids using a non-selective LC-MS method. Our results demonstrated that Eletta Campana (CAN48) was the cultivar with the highest concentration of cannabinoids, although their concentration was around 200 times lower than that of the most abundant metabolite (N-*trans*-caffeoyltyramine). Previous studies on the same accession (CAN48) obtained from the same genebank (IPK) reported an absolute quantification value of THC of 0.08%, which is well below the 0.3% threshold set by international regulations (Citti et al., 2019). Based on the low THC content of this cultivar, we might assume that all the hemp seed accessions included in the current study meet the international regulatory standards of < 0.3% THC content.

Recent studies have proved that despite being heavily studied, hemp seed continues to be a promising source of new metabolites. In the current study we discovered a previously unknown molecular family in hemp seeds "cinnamic acid glycosyl sulphates" widely distributed across all hemp seed accessions in similar concentrations. The identity of one novel compound from this chemical class was confirmed by isolation and interpretation of spectroscopic data, and four structural analogues, three of them new, is also suggested. The presence of these compounds in all hemp accessions suggest that cinnamic acid glycosyl sulphates might have important implications as quality control markers for the authentication of commercial hemp seed products extracted with highly aqueous solvents. Considering that these highly polar extracts, commonly used in the manufacture of cosmetics, are usually devoid of characteristic hemp seed phenylpropionamides due to insolubility issues, the need of more suitable quality control markers is justified. Further studies are still needed to identify the biological properties of these new metabolites.

In conclusion, our results confirm that some hemp accessions present in the IPK germplasm collection appear to contain more interesting profiles of metabolites than some of the cultivated hemp cultivars, as first suggested by Galasso et al. (2016). We suggest that accessions CAN37 and CAN58, should be scientifically prioritized for the discovery of new metabolites as well as in selective breeding programs for the development of new cultivars with high contents of bioactive phenylpropionamides. However, despite the functional and nutraceutical properties of this crop, the seeds of hemp are not totally free of antinutritional compounds, such as phytic acid (Russo, 2013; Galasso et al., 2016). According to previous reports, including ten of the accessions investigated here, CAN48 presents the highest levels of phytate, although other cultivars not included here like Futura75 and Felina32 have even higher values. The presence of high phytate contents in some cultivars limit their use in human nutrition and other monogastric animals since a high level of phytic acid may lead to mineral deficiencies of macro-and microelements, protein digestibility and poor organoleptic properties (Russo, 2013; Russo and Reggiani, 2015; Galasso et al., 2016). Therefore, an improvement for this trait might be necessary if the high yielding oil cultivar CAN48 is prioritized. According to previous studies (Galasso et al., 2016), the Italian accession CAN40 might be a good candidate to reduce the content of phytate through hybridization and selective breeding. Other accessions such as CAN32, or the Fibrimon cultivars CAN50 and CAN53, could also be prioritized if a 3:1 ratio of linoleic acid/α-linolenic acid is desired.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp, MSV000090725.

## Author contributions

GP-G conceived and designed the study, performed the LC-MS and GC-MS data analysis, guided the isolation methodology and wrote the first draft of the manuscript. AR prepared the samples for LC-MS and GC-MS acquisition and assisted in the isolation of metabolites. NS elucidated the structure of the new metabolite and assisted in the writing and proofreading of the manuscript. MC assisted in the extraction and isolation of the new metabolite and MS provided financial assistance for LC-MS and GC-MS analyses and assisted in the writing and proofreading of the manuscript. All authors contributed to the article and approved the submitted version.

## Acknowledgments

The authors would like to thank the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, for kindly proving the plant material. Li Hsun-Wen and Lin Che-Hsien (China Medical University, Taichung, Republic of China) are also acknowledged for their technical assistance preparing the LC-MS data for the heatmap analysis and to Rui Fang and Christopher Wallis (Royal Botanic Gardens, Kew) for the enlightening discussions.

## Conflict of interest

The authors declare that we have received funding from companies to assess the quality of commercialized hemp seed extracts including "The Procter & Gamble Company". However, no company contributed scientifically or financially to this research or was involved in the decision to publish it.

The authors declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1114398/full#supplementary-material

## References


Blauvelt, A. (2017). IL-6 differs from TNF-α: unpredicted clinical effects caused by IL-6 blockade in psoriasis. *J. Invest. Dermatol.* 137, 541–542. doi: 10.1016/j.jid.2016.11.022

Bradley, J. (2008). TNF-mediated inflammatory disease. *J. Pathol.* 214, 149–160. doi: 10.1002/path.2287

Callaway, J. C. (2004). Hempseed as a nutritional resource: An overview. *Euphytica* 140, 65–72. doi: 10.1007/s10681-004-4811-6

Chen, T., Hao, J., He, J., Zhang, J., Li, Y., Liu, R., et al. (2013). Cannabisin B induces autophagic cell death by inhibiting the AKT/mTOR pathway and S phase cell cycle arrest in HepG2 cells. *Food Chem.* 138, 1034–1041. doi: 10.1016/j.foodchem.2012.11.102

Chen, T., He, J., Zhang, J., Li, X., Zhang, H., Hao, J., et al. (2012). The isolation and identification of two compounds with predominant radical scavenging activity in hempseed (seed of *Cannabis sativa* L.). *Food Chem.* 134, 1030–1037. doi: 10.1016/j.foodchem.2012.03.009

Citti, C., Linciano, P., Panseri, S., Vezzalini, F., Forni, F., Vandelli, M. A., et al. (2019). Cannabinoid profiling of hemp seed oil by liquid chromatography coupled to high-resolution mass spectrometry. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00120

Crescente, G., Piccolella, S., Esposito, A., Scognamiglio, M., Fiorentino, A., and Pacifico, S. (2018). Chemical composition and nutraceutical properties of hempseed: an ancient food with actual functional value. *Phytochem. Rev.* 17, 733–749. doi: 10.1007/s11101-018-9556-2

De Vos, R. C., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., and Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2, 778–791. doi: 10.1038/nprot.2007.95

Ernst, M., Kang, K., Caraballo-Rodriguez, A. M., Nothias, L.-F., Wandy, J., Chen, C., et al. (2019). MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites* 9, 144. doi: 10.3390/metabo9070144

Farinon, B., Molinari, R., Costantini, L., and Merendino, N. (2020). The seed of industrial hemp (*Cannabis sativa* L.): nutritional quality and potential functionality for human health and nutrition. *Nutrients* 12, 1935. doi: 10.3390/nu12071935

Feunang, Y. D., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* 8, 61. doi: 10.1186/s13321-016-0174-y

Galasso, I., Russo, R., Mapelli, S., Ponzoni, E., Brambilla, I. M., Battelli, G., et al. (2016). Variability in seed traits in a collection of *Cannabis sativa* L. genotypes. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00688

Jang, E., Kim, H., Jang, S., Lee, J., Baeck, S., In, S., et al. (2020). Concentrations of THC, CBD, and CBN in commercial hemp seeds and hempseed oil sold in Korea. *Forensic Sci. Int.* 306, 110064. doi: 10.1016/j.forsciint.2019.110064

Leizer, C., Ribnicky, D., Poulev, A., Dushenkov, S., and Raskin, I. (2000). The composition of hemp seed oil and its potential as an important source of nutrition. *J. Nutraceuticals Funct. Med. Foods* 2, 35–53. doi: 10.1300/J133v02n04_04

Leonard, W., Zhang, P., Ying, D., and Fang, Z. (2021a). Lignanamides: sources, biosynthesis and potential health benefits – a minireview. *Crit. Rev. Food Sci. Nutr.* 61, 1404–1414. doi: 10.1080/10408398.2020.1759025

Leonard, W., Zhang, P., Ying, D., Xiong, Y., and Fang, Z. (2021b). Extrusion improves the phenolic profile and biological activities of hempseed (*Cannabis sativa* L.) hull. *Food Chem.* 346, 128606. doi: 10.1016/j.foodchem.2020.128606

Li, C.-X., Song, X.-Y., Zhao, W.-Y., Yao, G.-D., Lin, B., Huang, X.-X., et al. (2019). Characterization of enantiomeric lignanamides from *Solanum nigrum* L. and their neuroprotective effects against MPP+-induced SH-SY5Y cells injury. *Phytochemistry* 161, 163–171. doi: 10.1016/j.phytochem.2019.01.001

Luo, Q., Yan, X., Bobrovskaya, L., Ji, M., Yuan, H., Lou, H., et al. (2017). Anti-neuroinflammatory effects of grossamide from hemp seed *via* suppression of TLR-4-mediated NF-κB signaling pathways in lipopolysaccharide-stimulated BV2 microglia cells. *Mol. Cell. Biochem.* 428, 129–137. doi: 10.1007/s11010-016-2923-7

Ma, C.-Y., Liu, W. K., and Che, C.-T. (2002). Lignanamides and nonalkaloidal components of *Hyoscyamus Niger* seeds. *J. Nat. Prod.* 65, 206–209. doi: 10.1021/np010073b

Maiolo, S. A., Fan, P., and Bobrovskaya, L. (2018). Bioactive constituents from cinnamon, hemp seed and *Polygonum cuspidatum* protect against H2O2 but not rotenone toxicity in a cellular model of Parkinson's disease. *J. Tradit. Complement. Med.* 8, 420–427. doi: 10.1016/j.jtcme.2017.11.001

Mattila, P. H., Pihlava, J.-M., Hellström, J., Nurmi, M., Eurola, M., Mäkinen, S., et al. (2018). Contents of phytochemicals and antinutritional factors in commercial protein-rich plant products. *Food Qual. Saf.* 2, 213–219 doi: 10.1093/fqsafe/fyy021

Moccia, S., Siano, F., Russo, G. L., Volpe, M. G., La Cara, F., Pacifico, S., et al. (2020). Antiproliferative and antioxidant effect of polar hemp extracts (*Cannabis sativa* L., Fedora cv.) in human colorectal cell lines. *Int. J. Food Sci. Nutr.* 71, 410–423. doi: 10.1080/09637486.2019.1666804

Nigro, E., Crescente, G., Formato, M., Pecoraro, M. T., Mallardo, M., Piccolella, S., et al. (2020). Hempseed lignanamides rich-fraction: chemical investigation and cytotoxicity towards U-87 glioblastoma cells. *Molecules* 25, 1049. doi: 10.3390/molecules25051049

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nat. Methods* 17, 905–908. doi: 10.1038/s41592-020-0933-6

Ono, K., Demchak, B., and Ideker, T. (2014). Cytoscape tools for the web age: D3.js and Cytoscape.js exporters. *F1000Research* 3, 143. doi: 10.12688/f1000research.4510.2

Padilla-González, G. F., Amrehn, E., Frey, M., Gómez-Zeledón, J., Kaa, A., Da Costa, F. B., et al. (2020a). Metabolomic and gene expression studies reveal the diversity, distribution and spatial regulation of the specialized metabolism of yacón (*Smallanthus sonchifolius*, Asteraceae). *Int. J. Mol. Sci.* 21, 4555. doi: 10.3390/ijms21124555

Padilla-González, G. F., Diazgranados, M., Ccana-Ccapatinta, G., Casoti, R., and Da Costa, F. B. (2017). Caffeic acid derivatives and further compounds from *Espeletia barclayana* Cuatrec. (Asteraceae, Espeletiinae). *Biochem. Syst. Ecol.* 70, 291–293. doi: 10.1016/j.bse.2016.12.016

Padilla-González, G. F., Sadgrove, N. J., Ccana-Ccapatinta, G. V., Leuner, O., and Fernandez-Cusimamani, E. (2020b). Feature-based molecular networking to target the isolation of new caffeic acid esters from yacon (*Smallanthus sonchifolius*, asteraceae). *Metabolites* 10, 407. doi: 10.3390/metabo10100407

Pavlovic, R., Panseri, S., Giupponi, L., Leoni, V., Citti, C., Cattaneo, C., et al. (2019). Phytochemical and ecological analysis of two varieties of hemp (*Cannabis sativa* L.) grown in a mountain environment of Italian Alps. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01265

Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinf.* 11, 395. doi: 10.1186/1471-2105-11-395

Pojić, M., Mišan, A., Sakač, M., Dapčević Hadnađev, T., Šarić, B., Milovanović, I., et al. (2014). Characterization of byproducts originating from hemp oil processing. *J. Agric. Food Chem.* 62, 12436–12442. doi: 10.1021/jf5044426

Russo, R. (2013). Variability in antinutritional compounds in hempseed meal of Italian and French varieties. *Plant* 1, 25. doi: 10.11648/j.plant.20130102.13

Russo, R., and Reggiani, R. (2015). Evaluation of protein concentration, amino acid profile and antinutritional compounds in hempseed meal from dioecious and monoecious varieties. *Am. J. Plant Sci.* 06, 14–22. doi: 10.4236/ajps.2015.61003

Schluttenhofer, C., and Yuan, L. (2017). Challenges towards revitalizing hemp: A multifaceted crop. *Trends Plant Sci.* 22, 917–929. doi: 10.1016/j.tplants.2017.08.004

Schuster, B., Winter, M., and Herrmann, K. (1986). 4-O-β-D-glucosides of hydroxybenzoic and hydroxycinnamic acids — their synthesis and determination in

berry fruit and vegetable. *Z. für Naturforsch. C* 41, 511–520. doi: 10.1515/znc-1986-5-603

Shang, G.-S., Liu, L., and Qin, Y.-W. (2017). IL-6 and TNF-α promote metastasis of lung cancer by inducing epithelial-mesenchymal transition. *Oncol. Lett.* 13, 4657–4660. doi: 10.3892/ol.2017.6048

Simopoulos, A. P. (2008). The importance of the Omega-6/Omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Exp. Biol. Med.* 233, 674–688. doi: 10.3181/0711-MR-311

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221. doi: 10.1007/s11306-007-0082-2

Wandy, J., Zhu, Y., van der Hooft, J. J. J., Daly, R., Barrett, M. P., and Rogers, S. (2018). Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics* 34, 317–318. doi: 10.1093/bioinformatics/btx582

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34, 828–837. doi: 10.1038/nbt.3597

Wang, S., Luo, Q., and Fan, P. (2019). Cannabisin F from hemp (*Cannabis sativa*) seed suppresses lipopolysaccharide-induced inflammatory responses in BV2 microglia as SIRT1 modulator. *Int. J. Mol. Sci.* 20, 507. doi: 10.3390/ijms20030507

Yan, X., Tang, J., dos Santos Passos, C., Nurisso, A., Simões-Pires, C. A., Ji, M., et al. (2015). Characterization of lignanamides from hemp (*Cannabis sativa* L.) seed and their antioxidant and acetylcholinesterase inhibitory activities. *J. Agric. Food Chem.* 63, 10611–10619. doi: 10.1021/acs.jafc.5b05282

Zhou, Y., Wang, S., Ji, J., Lou, H., and Fan, P. (2018). Hemp (*Cannabis sativa* L.) seed phenylpropionamides composition and effects on memory dysfunction and biomarkers of neuroinflammation induced by lipopolysaccharide in mice. *ACS Omega* 3, 15988–15995. doi: 10.1021/acsomega.8b02250

# Frontiers in
# Plant Science

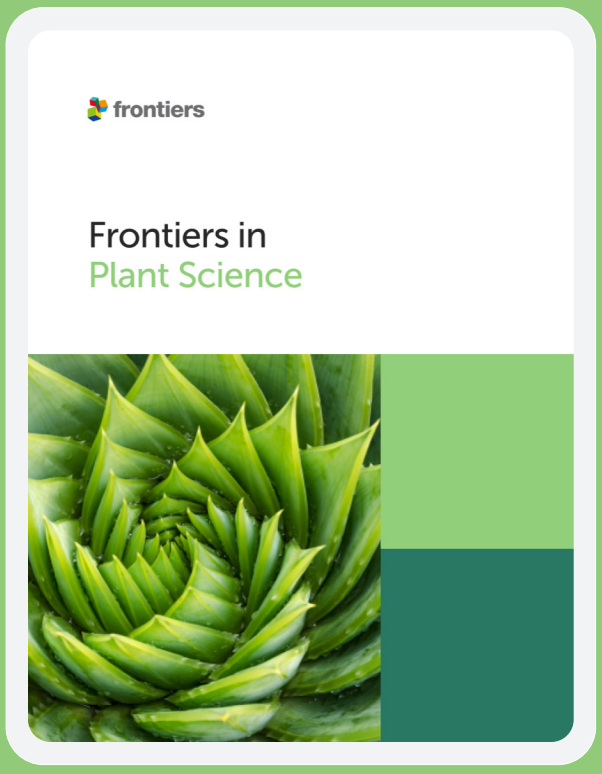
**Cultivates the science of plant biology and its applications**

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact