Innovative Metabolomics Insights for Better Health

Phytohormone Metabolomics Assay Final Report

Metware Biotechnology Inc.

www.metwarebio.com

Contents

MWHW-23-xxx-a Phytohormone Metabolomics Assay Final Report

1 Abstract

Phytohormones are chemical messengers that coordinate the cellular functions of multicellular organisms. Plant hormones play important roles in regulating various aspects of plant growth and development as well as in biotic and abiotic stress responses. Qualitative and quantitative analysis of plant hormones can help us understand the molecular mechanism behind these processes. MetwareBio's Phytohormone Metabolomics Assay is a quantitative detection method, which include auxin (Auxin), cytokinin (CKs), abscisic acid (ABA) , jasmonates (Jas), salicylic acid (SA), gibberellins (Gas), ethylene (ETH), strigolactones.

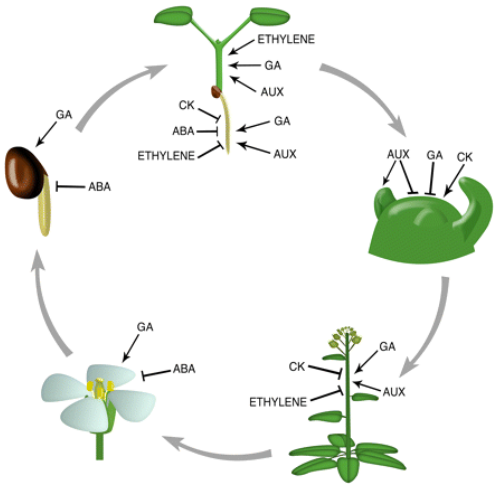

Fig 1: The effect of plant hormones on the growth and development of plants

2 The experimental process

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) can detect and quantify compounds with high polarity and poor thermal stability, and accurately quantify them. The overall process is as follows:

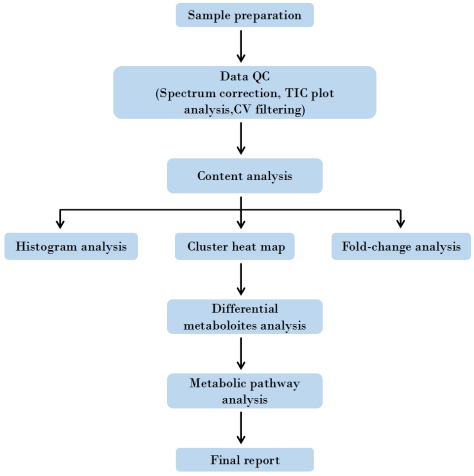

Fig 2: Flow chart of metabolomics analysis

Compounds to be detected:

Table 1: List of compounds in the panel

Number	Compounds	Index
1	Abscisic acid	ABA
2	ABA-glucosyl ester	ABA-GE
3	Abscisic aldehyde	ABA-ald
4	Indole-3-carboxylic Acid	ICA
5	Indole-3-carboxaldehyde	ICAld
6	3-Indolebutyric acid	IBA
7	Indole-3-acetic acid	IAA
8	1-O-indol-3-ylacetylglucose	IAA-Glc
9	Indoleacetyl glutamic acid	IAA-Glu
10	3-Indoleacetonitrile	IAN
11	Indole-3-acetyl glycine	IAA-Gly
12	2-oxindole-3-acetic acid	OxIAA
13	Indole-3-acetyl-L-aspartic acid	IAA-Asp
14	N-(3-Indolylacetyl)-L-leucine	IAA-Leu
15	N-(3-Indolylacetyl)-L-valine	IAA-Val
16	Indole-3-acetyl-L-tryptophan	IAA-Trp
17	N-(3-Indolylacetyl)-L-alanine	IAA-Ala
18	N-(3-Indolylacetyl)-L-phenylalanine	IAA-Phe
19	Indole-3-acetyl-Lglutamic acid dimethyl ester	IAA-Glu-diMe
20	Indole-3-acetyl-L-leucine methyl ester	IAA-Leu-Me
21	Indole-3-acetyl-L-valine methyl ester	IAA-Val-Me
22	Indole-3-acetyl-L-phenylalanne methyle ester	IAA-Phe-Me
23	3-Indoleacetamide	IAM
24	Tryptamine	TRA
25	Indole-3-lactic acid	ILA
26	3-Indoleacrylic acid	IA
27	L-tryptophan	TRP
28	Methyl indole-3-acetate	MEIAA

Table 1: List of compounds in the panel

Number	Compounds	Index
29	3-Indolepropionic acid	IPA
30	Indole	Indole
31	N6-isopentenyladenine	IP
32	trans-Zeatin	tZ
33	cis-Zeatin	cZ
34	Dihydrozeatin	DZ
35	Dihydrozeatin ribonucleoside	DHZR
36	cis-Zeatin riboside	cZR
37	cis-Zeatin-9-glucoside	cZ9G
38	N6-Isopentenyl-adenine-9-glucoside	iP9G
39	N6-Isopentenyl-adenine-7-glucoside	iP7G
40	trans-Zeatin-O-glucoside	tZOG
41	Dihydrozeatin-O-glucoside riboside	DHZROG
42	cis-Zeatin-O-glucoside riboside	cZROG
43	4-[[(9-beta-D-Glucopyranosyl-9H-purin-6-yl)amino]methyl]phenol	pT9G
44	2-Chloro-trans-zeatin	2CltZ
45	para-Topolin	pT
46	meta-Topolin	mT
47	meta-Topolin riboside	mTR
48	ortho-Topolin	oT
49	6-Benzyladenine	BAP
50	6-Benzyladenosine	BAPR
51	Kinetin	K
52	Kinetin riboside	KR
53	para-Topolin riboside	pTR
54	ortho-Topolin riboside	oTR
55	meta-Topolin-9-glucoside	mT9G
56	ortho-Topolin-9-glucoside	oT9G
57	N6-Benzyladenine -9-glucoside	BAP9G
58	N6-Benzyladenine-7-glucoside	BAP7G
59	Kinetin-9-glucoside	K9G
60	2-Methylthio-N6-isopentenyladenine	2MeSiP
61	2-methylthio-cis-zeatin	2MeScZ
62	2-Methylthio-cis-zeatin riboside	2MeScZR
63	2-Methylthio-N6-isopentenyladenosine	2MeSiPR
64	N6-isopentenyladenosine	IPR
65	trans-Zeatin riboside	tZR
66	Dihydrozeatin-7-glucoside	DHZ7G
67	cis-Zeatin riboside monophosphate	cZRMP
68	N-6-iso-pentenyladenosine-5'-monophosphate	iPRMP
69	trans-Zeatin-9-glucoside	tZ9G
70	9-Ribosyl-trans-zeatin 5'-monophosphate	tZRMP
71	1-Aminocyclopropanecarboxylic acid	ACC
72	Gibberellin A1	GA1
73	Gibberellin A3	GA3
74	Gibberellin A4	GA4
75	Gibberellin A7	GA7

Table 1: List of compounds in the panel

Number	Compounds	Index
76	Gibberellin A9	GA9
77	Gibberellin A15	GA15
78	Gibberellin A19	GA19
79	Gibberellin A20	GA20
80	Gibberellin A24	GA24
81	Gibberellin A53	GA53
82	Gibberellin A44	GA44
83	Gibberellin A8	GA8
84	Gibberellin A5	GA5
85	Gibberellin A6	GA6
86	Gibberellin A29	GA29
87	Gibberellin A34	GA34
88	Gibberellin A51	GA51
89	Gibberellin A12 aldehyde	GA12-ald
90	Jasmonic Acid	JA
91	Methyl jasmonate	MEJA
92	Dihydrojasmonic acid	H2JA
93	Jasmonoyl-L-isoleucine	JA-ILE
94	cis(+)-12-Oxophytodienoic acid	OPDA
95	N-[(-)-Jasmonoyl]-(l)-phenalanine	JA-Phe
96	N-[(-)-Jasmonoyl]-(L)-valine	JA-Val
97	3-oxo-2-(2-(Z)-Pentenyl) cyclopentane-1-butyric acid	OPC-4
98	3-oxo-2-(2-(Z)-Pentenyl)cyclopentane-1-hexanoic acid	OPC-6
99	12-Hydroxyjasmonic acid	12-OH-JA
100	Jasmonate-1-aminocyclopropane-1-carboxylic acid	JA-ACC
101	Salicylic Acid	SA
102	Salicylic acid 2-O-β-Glucoside	SAG
103	L-Phenylalanine	Phe
104	trans-Cinnamic acid	t-CA
105	2-Coumarate	2-Coumarate
106	2-Methoxycarbonylphenyl beta-D-glucopyranoside	MeSAG
107	5-Deoxystrigol	5DS
108	（±）Strigol	ST

Original file path: Final report/data/component.xlsx

2.1 Sample information

This project has 24 samples divided into 4 groups. Sample information is shown in the following table:

Table 2: Sample information table

Species	Tissues	MW_ID	Sample_ID
_	_	A1	A1
_	_	A2	A2
_	_	A3	A3
_	_	A4	A4
_	_	A5	A5
_	_	A6	A6
_	_	B1	B1
_	_	B2	B2
_	_	B3	B3
_	_	B4	B4
_	_	B5	B5
_	_	B6	B6
_	_	C1	C1
_	_	C2	C2
_	_	C3	C3
_	_	C4	C4
_	_	C5	C5
_	_	C6	C6
_	_	D1	D1
_	_	D2	D2
_	_	D3	D3
_	_	D4	D4
_	_	D5	D5
_	_	D6	D6

Original file path: Final report/0.data/sample_info.xlsx

2.2 Reagents and instruments

Table 3: Instrument information

Instrument	Model	Manufacturer
LC-MS/MS	6500+	SCIEX
Centrifuge	5424R	Eppendorf
Electronic balance	AS 60/220.R2	RADWAG
Ball mill instrument	MM400	Retsch
centrifugal concentrator	CentriVap	LABCONCO
Vortex	Cubic B	Four E'S
Ultrasonic cleaner	CD-F15	Olenyer

Table 4: Information of standards and reagents

Reagent	level	brand
Methanol	Chromatographically pure	Merck
Acetonitrile	Chromatographically pure	Merck
Acetic acid	Chromatographically pure	Sigma-Aldrich
Formic acid	Chromatographically pure	Sigma-Aldrich
Standard	Chromatographically pure	BioBioPha/Sigma-Aldrich

2.3 Sample extraction process

Fresh plant sample was harvested and stored at -80°C. Samples were freeze-dried, ground into powder (30 Hz, 1 min). 15 mg of plant sample was weighed into a 2 mL plastic microtube and frozen in liquid nitrogen, dissolved in 1 mL methanol/water/formic acid (15:4:1, V/V/V). 10 μL internal standard mixed solution (100 ng/mL) was added into the extract as internal standards (IS) for the quantication. The mixture was vortexed for 10 minutes, followed by centrifugation for 5 min (12000 r/min, and 4°C). The supernatant was transferred to clean plastic microtubes and dried by evaporation. The samples were reconstituted in 100 μL 80% methanol (V/V), and filtered through a membrane filter (0.22 μm, Anpel) before LC-MS/MS analysis.[2, 4, 8]

2.4 Chromatography-mass spectrometry acquisition conditions

The sample extracts were analyzed using an UPLC-ESI-MS/MS system (UPLC, ExionLC™ AD, https://sciex.com/; MS, Applied Biosystems 6500 Triple Quadrupole, https://sciex.com/).

UPLC Conditions were as follows:

Column: Waters ACQUITY UPLC HSS T3 C18 (100 mm×2.1 mm i.d.,1.8 μm); solvent system: water (0.04% acetic acid):acetonitrile (0.04% acetic acid); gradient program: 95:5 V/V at 1min, 5:95 V/V at 8 min, hold for 1 min, 95:5 V/V at 9 min; hold for 3min; flow rate: 0.35 mL/min; temperature: 40°C; injection volume: 2 μL.

The mass spectrum conditions were as follows:

Triple quadrupole (QQQ) scans were acquired on a triple quadrupole mass spectrometer, 6500+ LC-MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in both positive and negative ion mode and controlled by Analyst 1.6.3 software (Sciex). The ESI source operation parameters were as follows: ion source, ESI+/-; source temperature 550 °C; ion spray voltage (IS) 5500 V (Positive) -4500 V (Negative); curtain gas (CUR) was set at 35 psi, respectively. Phytohormones were analyzed using scheduled multiple reaction monitoring (MRM). Data acquisitions were performed using Analyst 1.6.3 software (Sciex). Multi-quant 3.0.3 software (Sciex) was used to quantify all metabolites. Mass spectrometer parameters including the declustering potentials (DP) and collision energies (CE) for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period.[3, 5, 7]

2.5 Qualitative and quantitative principles of metabolites

Metabolites were quantified by multiple reaction monitoring (MRM) using triple quadrupole mass spectrometry. In MRM mode, the first quadrupole screened the precursor ions for the target substance and excluded ions of other molecular weights. After ionization induced by the impact chamber, the precursor ions were fragmented, and a characteristic fragment ion was selected through the third quadrupole to exclude the interference of non-target ions. After obtaining the metabolite spectrum data from different samples, the peak area was calculated on the mass spectrum peaks of all substances and analyzed by standard curves.

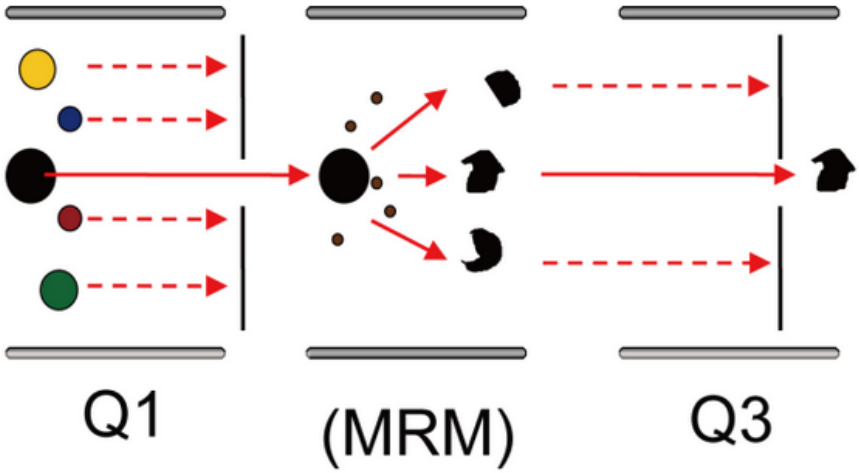

Fig 3:
Schematic diagram of multiple reaction monitoring mode by mass spectrometry

3 Data evaluation

3.1 Data pre-processing

Analyst 1.6.3 was used to process mass spectrum data. The following figure shows the total ions current (TIC) and MRM metabolite detection multi-peak diagram (XIC) of the mixed QC samples. The X-axis shows the Retention time (RT) from metabolite detection, and the Y-axis shows the ion flow intensity from ion detection (intensity unit: CPS, count per second).

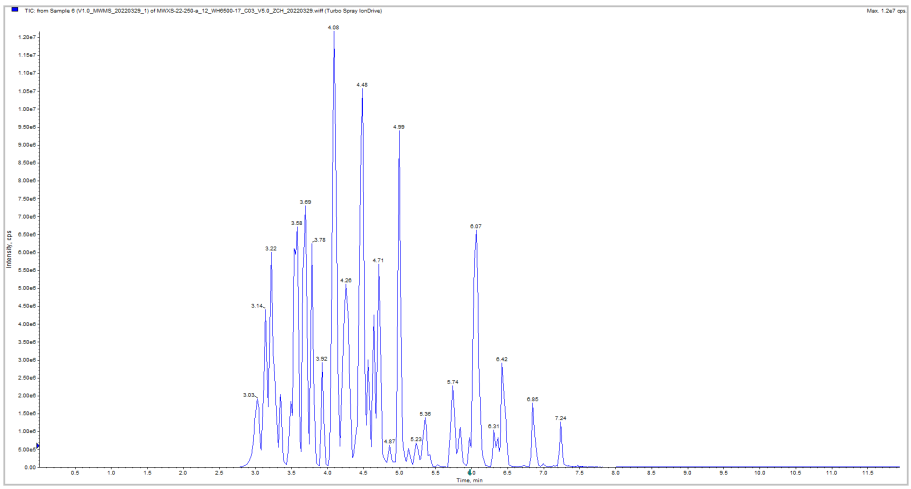

Fig 4: Total ion current diagram of mixed phase mass spectrum analysis

Original file path: Final report/0.data/QC/*QC_MS_TIC.png

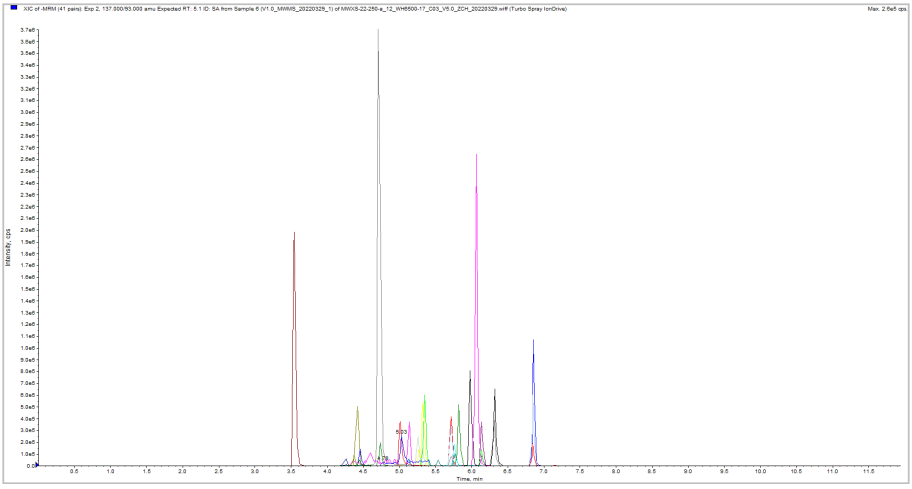

Fig 5: Extraction ion flow chromatogram

Original file path: Final report/0.data/QC/*MRM_detection_of_multimodal_maps*

The mass spectrometry data was analyzed using MultiQuant 3.0.3 software. The mass spectrum peaks detected in different samples were scored and corrected based on retention time and peak shape of the standard. The figure below shows the correction results of quantitative analysis of a substance randomly selected from different samples.

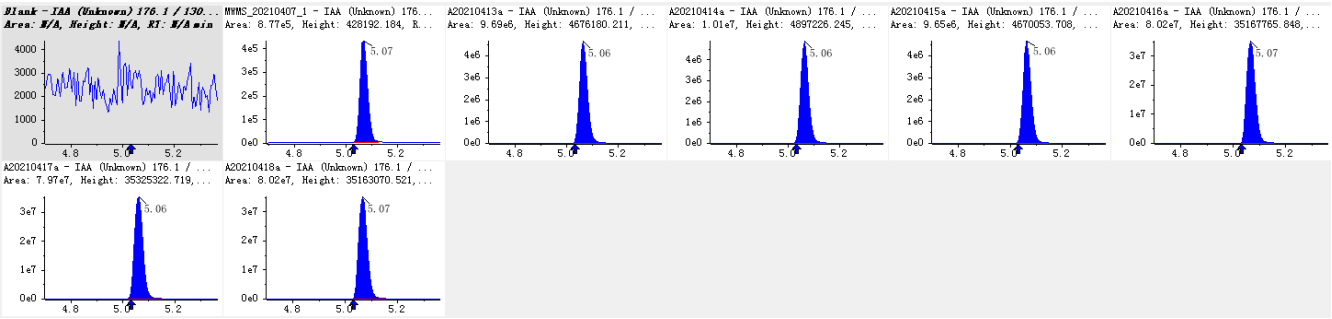

Fig 6: Scoring correction diagram for quantitative analysis of metabolites

Note: The figure shows the quantitative analysis integral correction results of randomly selected metabolites in different samples. The x-axis is the retention time (min) of metabolite detection, the y-axis is the ion flow intensity (CPS) of a certain metabolite ion detection, and the peak area represents the relative content of the substance in the sample.

Original file path: Final report/0.data/QC/*Integral_correction.png

3.2 Standard Solution Preparation

Standards were prepared with a series of concentrations from 0.01 ng/mL, 0.05 ng/mL, 0.1 ng/mL, 0.5 ng/mL, 1 ng/mL, 5 ng/mL, 10 ng/mL, 50 ng/mL, 100 ng/mL, 200 ng/mL, 500 ng/mL. For 7 compounds (IAA-Glc, TRP, H2JA, SAG, t-CA, 5DS, ST), each point of the serial dilution was 20-fold of the standard concentration. For Phe, each point of the serial dilution was 30-fold of the standard concentration. Each compound in the standard had obtained the chromatography peak intensity data and calibration curve. The equation of calibration curve are shown in the following table

Table 5: Equation of calibration curve

Index	Compounds	Class	RT
ABA	Abscisic acid	ABA	N/A
ABA-ald	Abscisic aldehyde	ABA	N/A
ABA-GE	ABA-glucosyl ester	ABA	N/A
IAA-Asp	Indole-3-acetyl-L-aspartic acid	Auxin	N/A
IPA	3-Indolepropionic acid	Auxin	N/A
ICAld	Indole-3-carboxaldehyde	Auxin	N/A
TRA	Tryptamine	Auxin	N/A
IAA-Glu	Indole-3-acetyl glutamic acid	Auxin	N/A
IAA-Ala	N-(3-Indolylacetyl)-L-alanine	Auxin	N/A
IAM	3-Indole acetamide	Auxin	N/A

Original file path/0.data/equation.xlsx

3.3 Quantification Results

Compound quantification was obtained by substituting the sample peak area ratio in the detected samples into the equation of calibration curve for calculation.

Concentration of the sample（ng/g）= c*V/1000/m

c（ng/mL）is the concentration obtained by substituting the sample peak area ratio into the equation of calibration curve.

V（μL）is the volume of solution used for reconstitution.

m（g）is the quality of the sample.

The metabolite ID, concentration and corresponding metabolite names of some metabolites detected in this experiment are shown in the following table:

Table 6: Statistical Table of metabolite quantity

Index	A1	A2
ABA	13.5173	11.4789
ABA-GE	58.6315	75.3755
ABA-ald	92.1498	79.8324
IAA-Val-Me	20.6642	25.7091
Indole	83.7167	65.6261
IPA	17.9413	12.4903
ILA	45.7197	50.3847
ICAld	61.7979	43.2415
ICA	38.0458	43.321
IBA	78.9966	60.8545

Original file path: Final report/0.data/*level.xlsx

3.4 Sample Quality Control Analysis

3.4.1 Total Ion Chromatogram Analysis

Using the mixed solution as the QC sample, one QC sample was inserted every 10 detection samples for analysis during the detection by the system. The stability of the device during the detection of the project can be assessed by analyzing the overlapped total ion flow chromatograms (TICs) obtained from the mass spectrometry detection and analysis of the same QC samples. The high stability of the testing device is a vital safeguard for the reproducibility and reliability of the data.

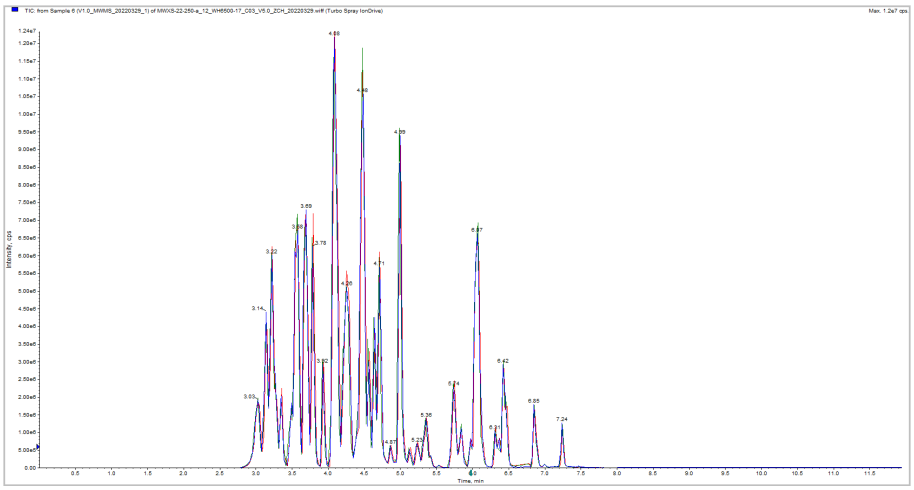

Fig 7: TIC overlap diagram detected by QC sample essence spectrum
Note: Superimposed spectrum from different QC samples. The results showed that the spectrum of total ion flow were highly consistent indicating that the signal stability was good when the same sample was detected at different times by mass spectrometry.

Original file path: Final report/0.data/picture/*QC_MS_tic_overlap*

3.4.2 CV value distribution of all samples

The Coefficient of Variation (CV) value is the ratio between the standard deviation of the original data and the mean of the original data, which can reflect the degree of data dispersion. The Empirical Cumulative Distribution Function (ECDF) can be used to analyze the frequency of CV of substances that is smaller than the reference value. The higher the proportion of substances with low CV value in QC samples is, the more stable the experimental data is. The proportion of substances with CV value less than 0.3 in QC samples was higher than 80% , indicating that the experimental data were relatively stable. The proportion of substances with CV value less than 0.2 in QC samples was higher than 80%, indicating that the experimental data were very stable.

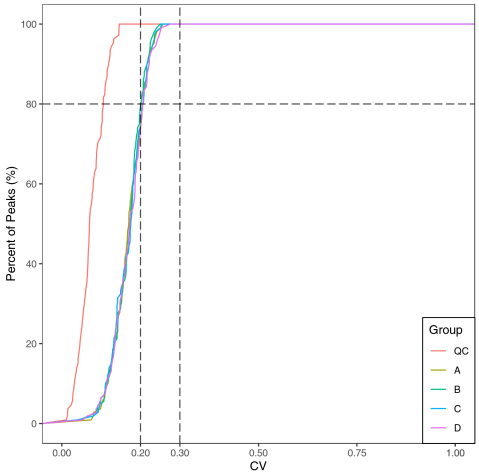

Fig 8: CV distribution of each group

Note: The X-axis represents the CV value, the Y-axis represents the proportion of metabolites with CV value less than a corresponding reference value. Different colors represent different sample groups. QC indicates quality control samples. The two dash lines on X-axis correspond to 0.2 and 0.3; the two dash lines on Y-axis correspond to 80% .

Original file path: Final report/1.Data_Assess/CV/*ECDF*

3.5 Sample quantification histogram

The results of sample content are grouped by statistics, and the statistical results are shown in the following table.

Table 7: Statistical results table

Index	Group	N	Mean
12-OH-JA	D	6	32.798
12-OH-JA	C	6	88.451
12-OH-JA	B	6	47.747
12-OH-JA	A	6	68.592
2-Coumarate	D	6	35.421
2-Coumarate	C	6	62.979
2-Coumarate	B	6	91.068
2-Coumarate	A	6	60.136
2CltZ	D	6	30.719
2CltZ	C	6	20.307

Original file path: Final report/1.Data_Assess/histogram/groups*.xlsx

The bar chart below shows the content difference of each substance in different groups.

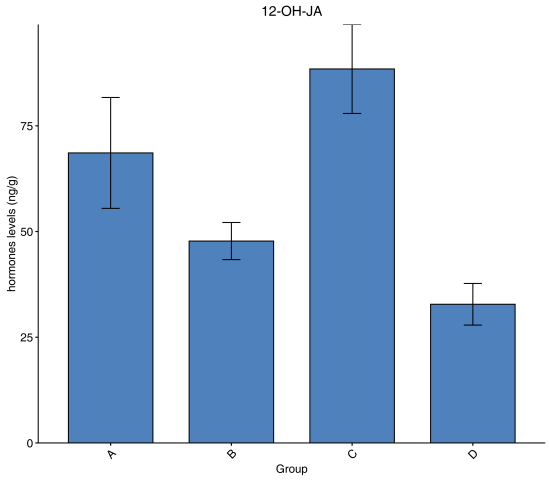

Fig 9: Sample content histogram

Note:The x-axis is the groups, the y-axis is the content, error bars are standard deviations.

Original file path: Final report/1.Data_Assess/histogram/histogram_compounds/*.png

3.6 Hierarchical Cluster Analysis

3.6.1 Principles of cluster analysis

Hierarchical Cluster Analysis (HCA) is a type of multivariate statistical analysis method. The samples are classified according to their features such that highest homogeneity is achieved between sample from the same group and highest heterogeneity is achieved between samples from different groups. In this report, the compound quantification data was normalized (Unit Variance Scaling, UV Scaling) and heatmaps were drawn by R software Pheatmap package. Hierarchical Cluster Analysis (HCA) was used to cluster the samples.

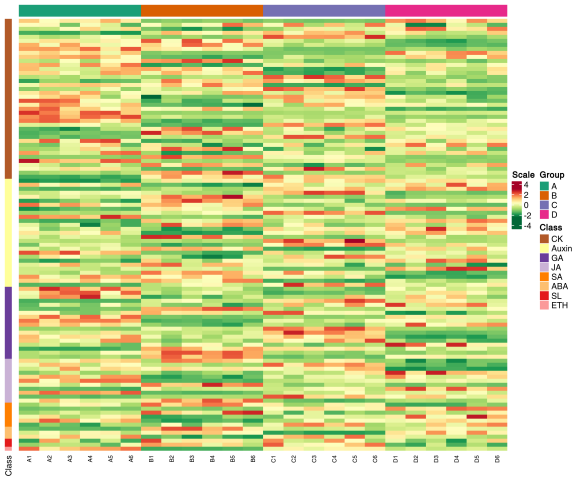

Fig 10: Sample clustering diagram

Note: X-axis indicates the sample name and the Y-axis are the metabolites. Group indicates sample groups. Z-Score indicates the relative quantification of each metabolite with red representing higher content and green representing lower content. Cluster analysis was performed on both metabolites (vertical cluster tree) and samples (horizontal cluster tree). “all_heatmap_class”: Heat map based on metabolite classification; “all_heatmap_no_cluster”: Showing only heatmap.

Original file path: Final report /1.Data_Assess/*all_heatmap*

4 Analysis results

4.1 Differential metabolite screening

Metabolomics data is characterized by “high-dimensional and massive”, requiring a large number of replicate samples to accurately uncover differential metabolites. Variance screening was performed based on the fold change value. Screening criteria:

For two sets of comparisons:

1.Metabolites with Fold Change ≥ 2 and Fold Change ≤ 0.5 were considered as significant and selected.

A partial result from the screening criteria is seen below:

Table 8: Screening results of differential metabolites

Index	Compounds	Type
ABA	Abscisic acid	down
ABA-ald	Abscisic aldehyde	up
IAA-Val-Me	Indole-3-acetyl-L-valine methyl ester	down
Indole	Indole	up
IPA	3-Indolepropionic acid	down
IAA-Ala	N-(3-Indolylacetyl)-L-alanine	up
IAA-Asp	Indole-3-acetyl-L-aspartic acid	down
IAA-Glu-diMe	Indole-3-acetyl-L-glutamic acid dimethyl ester	up
IAA-Gly	Indole-3-acetyl glycine	down
IAA-Phe-Me	Indole-3-acetyl-L-phenylalanne methyle ester	up

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/group-ID*_vs_group-ID*filter.xlsx.

4.1.1 Differential metabolite statistics

The number of different metabolites in each group is shown in the table below:

Table 9: Statistical table of differential metabolites

group name	All sig diff	down regulated	up regulated
A_vs_B	42	17	25

Statistical table of differential metabolites:Final report/2.Basic_analysis/Difference_analysis/sigMetabolitesCount.xl

4.2 Functional annotation and enrichment analysis of differential metabolites in KEGG database

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database that integrates compounds and genes into metabolic pathways. The KEGG database enabled researchers to study genes with their expression information and compounds with their abundances as a complete network.

4.2.1 Functional annotation of differential metabolites

Metabolites are annotated using the KEGG database, and only metabolic pathways containing differential metabolites are shown. Detailed results are found in the attached results. A portion of the results is shown below:

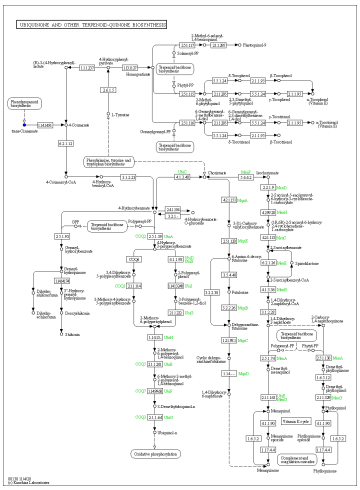

Fig 11: KEGG pathway of metabolites

Note: Red circles indicate that the metabolite content was significantly up-regulated in the experimental group; the blue circles indicate that the metabolite content was detected but did not change significantly; Green circles indicate that the metabolite content was significantly down-regulated in the experimental group. The orange circles indicate a mixture of both up-regulated and down-regulated metabolites.This allows searching for metabolites that may contribute to the phenotypic differences.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/Graph/ko*.

Statistical analysis of KEGG database annotation of screened metabolites with significant differences. Some of the results are as follows:

Table 10: KEGG annotations for differential metabolites

Index	Compounds	Type	cpd_ID
ABA	Abscisic acid	down	C06082
ABA-ald	Abscisic aldehyde	up	C13455
IAA-Val-Me	Indole-3-acetyl-L-valine methyl ester	down	-
Indole	Indole	up	C00463
IPA	3-Indolepropionic acid	down	-
IAA-Ala	N-(3-Indolylacetyl)-L-alanine	up	-
IAA-Asp	Indole-3-acetyl-L-aspartic acid	down	-
IAA-Glu-diMe	Indole-3-acetyl-L-glutamic acid dimethyl ester	up	-
IAA-Gly	Indole-3-acetyl glycine	down	-
IAA-Phe-Me	Indole-3-acetyl-L-phenylalanne methyle ester	up	-

Table 11: Enrichment Statistics of KEGG annotations for differential metabolites

ko_ID	Sig_compound	compound	Sig_compound_all	compound_all
ko00906	2	4	18	40
ko01100	12	24	18	40
ko01110	14	34	18	40
ko04075	4	9	18	40
ko00380	2	5	18	40
ko00400	2	3	18	40
ko00402	1	1	18	40
ko00999	2	5	18	40
ko00908	2	8	18	40
ko00270	1	1	18	40

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_filter_kegg.xlsx.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_KEGG.xlsx.

4.2.2 KEGG classification of differential metabolites

The significant differential metabolites were classified based on pathway annotation . The results are as follows:

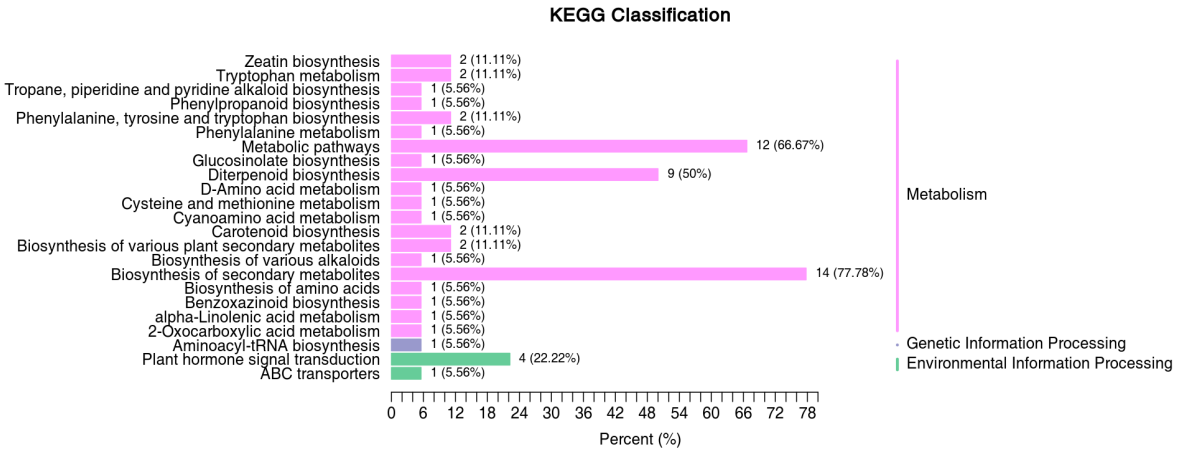

Fig 12: KEGG classification of differential metabolites

Note: the Y-axis shows the name of the KEGG pathway. The number of metabolites and the proportion of the total metabolites are shown next to the bar plot.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID_KEGG_barplot.*.

4.2.3 KEGG enrichment analysis of differential metabolites

KEGG pathway enrichment analysis was conducted based on the annotation results. We calculated the Rich Factor for each pathway, which is the ratio of the number of differenetial metabolites in the corresponding pathway to the total number of metabolites annotated in the same pathway. The greater the Rich Factor, the greater the degree of enrichment. P-value is the calculated using hypergeometric test as shown below:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

N represents the total number metabolites with KEGG annotation, n represents the number of differential metabolites in N, M represents the number of metabolites in a KEGG pathway in N, and m represents the number of differential metabolites in a KEGG pathway in M. The closer the p-value to 0, the more significant the enrichment. The size of the dots in the figure represents the number of significantly different metabolites enriched in the corresponding pathway. The results are shown below:

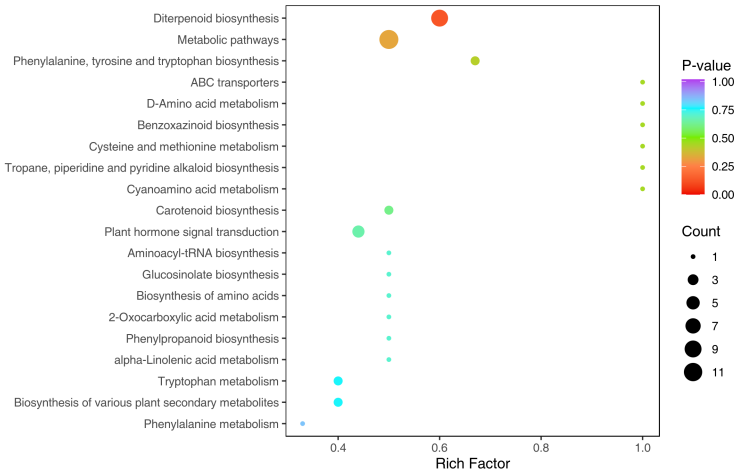

Fig 13: KEGG enrichment diagram of differential metabolites
Note: The X-axis represents the Rich Factor and the Y-axis represents the pathway. The color of points reflects the p-value. The darker the red, the more significant the enrichment. The size of the dot represents the number of enriched differential metabolites.

Final report/2.Basic_analysis/Difference_analysis/group-ID*_vs_group-ID*/enrichment/group-ID*_vs_group-ID*_KEGG_Enrichment.*.

5 References

1. Li Y, Zhou C, Yan X, et al. Simultaneous analysis of ten phytohormones in Sargassum horneri by high-performance liquid chromatography with electrospray ionization tandem mass spectrometry[J]. Journal

of Separation Science, 2016, 39(10): 1804-1813

2. Cai BD, Zhu JX, Gao Q, et al. Rapid and high-throughput determination of endogenous cytokinins in Oryza sativa by bare Fe3O4 nanoparticles-based magnetic solid-phase extraction[J]. Journal of Chromatography A, 2014, 1340:146-150.
3. Pan, X Q, Welti, et al. Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography-mass spectrometry[J]. NAT PROTOC, 2010.
4. Niu Q, Zong Y, Qian M, et al. Simultaneous quantitative determination of major plant hormones in pear flowers and fruit by UPLC/ESI-MS/MS[J]. Analytical Methods, 2014, 6(6): 1766.
5. Šimura J, Antoniadi I, Široká J, et al. Plant Hormonomics: Multiple Phytohormone Profiling by Targeted Metabolomics.[J]. Plant Physiol, 2018, 177(2):476-489.
6. Floková K, Tarkowská D, Miersch O, et al. UHPLC-MS/MS based target profiling of stress-induced phytohormones.[J]. Phytochemistry, 2014, 105:147-157.
7. Kunyan C, Yaying L, Xi Z , et al. Comparison of sample pretreatment methods for the determination of multiple phytohormones in plant samples by liquid chromatography-electrospray ionization-tandem mass spectrometry[J].Microchemical Journal, 2015, 25-31.
8. Hua-Ming X, Wen-Jing C, ian-Tian Y, et al. Spatio-temporal profiling of abscisic acid, indoleacetic acid and jasmonic acid in single rice seed during seed germination[J]. Analytica Chimica Acta, 2018, 119-127.

6 Appendix

6.1 Analytical methods

1.Hierarchical Cluster Analysis

The HCA (hierarchical cluster analysis) results of samples and metabolites were presented as heatmaps with dendrograms. HCA was carried out by R package pheatmap. For HCA, normalized signal intensities of metabolites (unit variance scaling) are visualized as a color spectrum.

2.Differential metabolites selected

Significantly regulated metabolites between groups were determined by absolute Log_2FC (fold change).

3.KEGG annotation and enrichment analysis

Identified metabolites were annotated using KEGG compound database (http://www.kegg.jp/kegg/compound/), annotated metabolites were then mapped to KEGG Pathway database (http://www.kegg.jp/kegg/pathway.html). Pathways with significantly regulated metabolites mapped to were then fed into MSEA (metabolite sets enrichment analysis), their significance was determined by hypergeometric test's p-values.

6.2 List of software and versions

Data processing methods were mainly adopted in the analysis process in two ways:

(1) unit variance scaling (UV)

Unit variance Scaling (UV) is also called Z-Score standardization, i.e., auto scaling. This method standardizes data according to mean and standard deviation of original data. The processed data conform to the standard normal distribution, that is, the mean value is 0 and the standard deviation is 1.

Calculation method: Divide the original data center by standard deviation.

The formula is as follows:

$$x' = \frac{x - \mu}{\sigma}$$

Where μ is the mean value and σ is the standard deviation.

(2) Centralization/zero-mean-centered (Ctr)

Calculation method: subtract the mean of the variables from the original data.

The formula is as follows:

$$x' = x - \mu$$