

RESEARCH ARTICLE

Metabolomics reveals significant variations in metabolites and correlations regarding the maturation of walnuts (*Juglans regia* L.)

Guodong Rao^{1,2,3}, Jinkai Sui² and Jianguo Zhang^{1,2,3,*}

ABSTRACT

The content of walnut metabolites is related to its nutritive value and physiological characteristics, however, comprehensive information concerning the metabolome of walnut kernels is limited. In this study we analyzed the metabolites of walnut kernels at five developmental stages from filling to ripening using GC-MS-based untargeted metabolomics; of a total 252 peaks identified, 85 metabolites were positively identified. Further statistical analysis revealed that these 85 metabolites covered different types of metabolism pathways. PCA scores revealed that the metabolic compositions of the embryo are different at each stage, while the metabolic composition of the endotesta could not be significantly separated into distinct groups. Additionally, 7225 metabolite-metabolite correlations were detected in walnut kernel by a Pearson correlation coefficient approach; during screening of the calculated correlations, 463 and 1047 were determined to be significant with $r^2 \geq 0.49$ and had a false discovery rate (FDR) ≤ 0.05 in endotesta and embryo, respectively. This work provides the first comprehensive metabolomic study of walnut kernels and reveals that most of the carbohydrate and protein-derived carbon was transferred into other compounds, such as fatty acids, during the maturation of walnuts, which may potentially provide the basis for further studies on walnut kernel metabolism.

KEY WORDS: Metabolome, Walnut kernel, Developmental stages, ANOVA, Metabolite-metabolite correlation

INTRODUCTION

Walnuts (*Juglans regia* L.) are classified as a strategic species and an indispensable food for human nutrition due to the high protein and oil content of its kernels (Martinez et al., 2010). Walnut trees, which are widely distributed around the world, are cultivated commercially throughout southern Europe, northern Africa, the USA, western South America, and eastern Asia and have been characterized as an oleaginous tree for biodiesel fuels (Moser, 2012). Walnuts rank second behind almonds in tree nut production, and China leads the world production of walnuts according to the Food and Agriculture Organization (FAO). In 2010, global walnut production was 1,500,000 tons, with China accounting for 33.33%

of the global walnut production (Mao and Hua, 2012). The walnut oil content of walnuts is 62–68%, thus making walnut oil the major product of the walnut; it is one of the most important specialty oils used for salad dressings and cooking due to the high content of monounsaturated and polyunsaturated fatty acids (Ros et al., 2004; Mao and Hua, 2012). Walnut protein is the main by-product of oil, which is considered to be another source of plant protein for use in human food products (Mao and Hua, 2014). Other contents, such as phenolic acids, tocopherols, flavonoids, juglone and phytosterols have also been researched (Martinez et al., 2010; Nour et al., 2013; Cosmulescu et al., 2014); it has been reported that the consumption of walnut kernel can provide protection against certain types of cancer due to its high concentration of natural antioxidants (Miraliakbari and Shahidi, 2008). Walnut nutrient composition has been investigated in several studies, the content of triterpenic, sterols, tocopherols, aliphatic alcohols, volatile compounds and carotenoids in the kernel oils from six walnut varieties were detected (Abdallah et al., 2015), and the level of tocopherols, fatty acids, total carotene and selenium of fifteen walnut varieties have also been studied (Ozrenk et al., 2012). Molecular weight distribution, gel electrophoresis and amino acid composition of walnut proteins and protein fractionations were analyzed by high performance liquid chromatography (HPLC) to get a comprehensive understanding of the walnut proteins (Mao et al., 2014); however, a comprehensive study of the dynamic metabolite changes during walnut kernel ripening has not yet been published. Metabolomic analysis involves detecting and quantifying metabolic changes with techniques such as mass spectrometry and NMR spectroscopy, and integrating the resulting data with multivariate statistical techniques such as principal component analysis (PCA) and orthogonal signal correction projection to latent structure discriminant analysis (OPLS-DA) (Zhang et al., 2011; Topfer et al., 2015). In this study, a GC-MS based metabolomic approach was utilized to investigate the metabolic composition and natural metabolite variations during walnut kernel development. The measured metabolites were integrated into a metabolism network, and metabolite-metabolite correlation analysis was conducted. Results from this study will provide new insights into the understanding of walnut kernel metabolism, and create a more comprehensive overview of changes in metabolism during walnut kernel development and maturation.

RESULTS

Metabolomic profiling of ripening kernels in walnuts

Kernels were sampled at five developmental stages from filling to ripening, with endotesta and embryos of kernels investigated separately. These developmental stages were identified as P1–P5 for endotesta, and R1–R5 for embryo. The metabolic profiling of kernels was investigated using an untargeted global metabolomics platform with GC-MS analysis. A total of 252 metabolites were detected, of which 85 named metabolites were confirmed using National Institute of Standards and Technology (NIST) and Wiley libraries.

¹State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, Republic of China.

²Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, Republic of China. ³Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, Republic of China.

*Author for correspondence (zhangjg@caf.ac.cn)

 J.Z., 0000-0002-8915-3155

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Hierarchical clustering analysis was performed to classify the walnut kernel metabolites, these 85 metabolites were classified into seven major metabolite groups which covered different metabolism pathways. The largest group contained 36.47% of the total number of identified metabolites: 31 organic acid metabolites. The second largest group had 20% of the total identified metabolites: 17 metabolites involved in carbohydrate metabolism. The third largest group (17.64%) consisted of 15 metabolites related to amino acid metabolism, followed by ten (11.76%) metabolites in amine metabolism, six (7.05%) in phosphate metabolism, and five (5.88%) in lipid metabolism. Principal component analysis (PCA) on all 85 metabolites was conducted to compare their metabolic composition throughout kernel development, with two principal components explaining 58.81% of the overall variance of the metabolite profiles, 42.66% and 16.15% for PC1 and PC2, respectively. The PCA scores revealed that metabolomes of the kernel and the endotesta are different from each other at all developmental stages (Fig. 1). The metabolic pathway of identified differentially abundant metabolites were then analyzed by the Kyoto Encyclopedia of Genes and Genomes (KEGG) which showed 63 identified metabolites that covered 14 pathways or metabolisms, including the citric acid cycle (TCA cycle), beta-alanine metabolism, D-arginine and D-ornithine metabolism, D-alanine metabolism, sphingolipid metabolism, propanoate metabolism, cysteine and methionine metabolism, pyruvate metabolism, arginine and proline metabolism, pyruvate metabolism, cysteine and methionine metabolism, pentose phosphate pathway, fructose and mannose metabolism, and glycolysis/gluconeogenesis (Fig. S2).

One-way ANOVA analyses of the endotesta and embryo metabolites

A one-way analysis of variance (ANOVA) statistical assay was performed to calculate the variations within the detected metabolites, *P*-values had been corrected for multiple hypothesis testing. The results revealed that levels of 30 metabolites in the embryos were differentially expressed at different stages of development (Fig. 2A), and levels of 26 metabolites of the endotesta were differentially expressed at the different stages of development (Fig. 2B). The 30 significantly differentially abundant metabolites from embryos included nine amino acids, eight organic acids, three amines and carbohydrates, and one phosphate. The 26

significantly differentially abundant metabolites from the endotesta contained eight amino acids, six organic acids, five carbohydrates, and three amines. We then calculated the variations of the metabolites in stage 1, stage 3, and stage 5 to further analyze the metabolite differences in bigger time span stages. The results showed that levels of 39 metabolites differed significantly in the embryo (Fig. 2C), and levels of 29 metabolites of the endotesta were significantly different (Fig. 2D). These significantly upregulated metabolites are mainly composed of amino acids and organic acids. Comparison analysis of significantly variable metabolites between all five stage and three stage metabolites revealed that the metabolites in the tissue of embryo are more variable compared to the tissue of endotesta, and the majority of these differentially abundant metabolites participate in amino acid and organic acid metabolism.

Correlation analysis of endotesta and embryo metabolites in walnut

Person correlation coefficient analysis was used to analyze the metabolite-metabolite correlation among identified metabolites in endotesta and embryo (Carrari et al., 2006; Blekhman et al., 2014). This analysis allowed for the identification of metabolites that related to each other in the two tissues across walnut development and ripening. Metabolite-metabolite correlations between the tissue of endotesta and embryo showed unique profiles; in endotesta 1156 resulted in significant correlation coefficients ($P < 0.01$), out of these 1156 significant correlations, 790 were positive and 366 were negative (Fig. 3A). Notably amino acids dominated the significant metabolite correlations, followed by organic acids. In embryo, 1225 resulted in significant correlation coefficients ($P < 0.01$), out of these 1225 significant correlations, 715 were positive and 510 negative (Fig. 3B), amino acids and organic acids also dominated the significant metabolite correlations.

Further screening of the calculated correlations was performed, and 463 were determined to be significant with $r^2 \geq 0.49$ and had a false discovery rate (FDR) ≤ 0.05 in endotesta, among which 258 were positive correlations and 205 were negative correlations (Fig. S3). In the 258 positive correlations, organic acids dominated the most significant metabolite-metabolite correlations, followed by amino acids, carbohydrates, and amines. In embryos, there were 1047 significant correlations with $r^2 \geq 0.49$ and FDR ≤ 0.05 identified; among them, 911 were positive correlations while 136 were negative correlations. Organic acids and amino acids also played important roles in metabolite correlations.

DISCUSSION

Kernel maturity is an important developmental stage of plant reproduction, and the metabolites in kernels relate to kernel quality and nutritional values. Walnut is an agronomically important tree with high nutritional value due to the high contents of oils, proteins, and many other beneficial metabolites in the kernel. In this study, differential quantitative metabolite profiling was used to assess developmental changes of walnut kernels. Endotesta have the main function of protection of the embryo and also have a linkage between embryo and outer environment. The levels of metabolites in endotesta were relatively stable during walnut kernel maturation, PCA scores also revealed the metabolic compositions in this tissue could not be separated. Embryos are the energy source of kernels, metabolites in this tissue are synthesized for the germination as seed. The levels of most metabolic compositions in embryo were increased during the development of kernels. PCA scores also proved metabolites of the embryo are different at each stage of development.

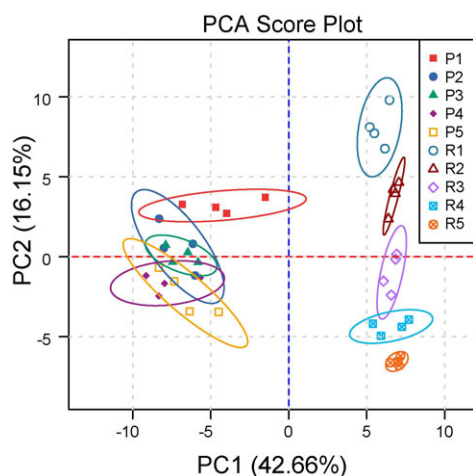


Fig. 1. Cluster analysis and PCA of metabolites in endotesta and embryos of walnuts. PCA score plot generated from all 85 metabolites of different samples.

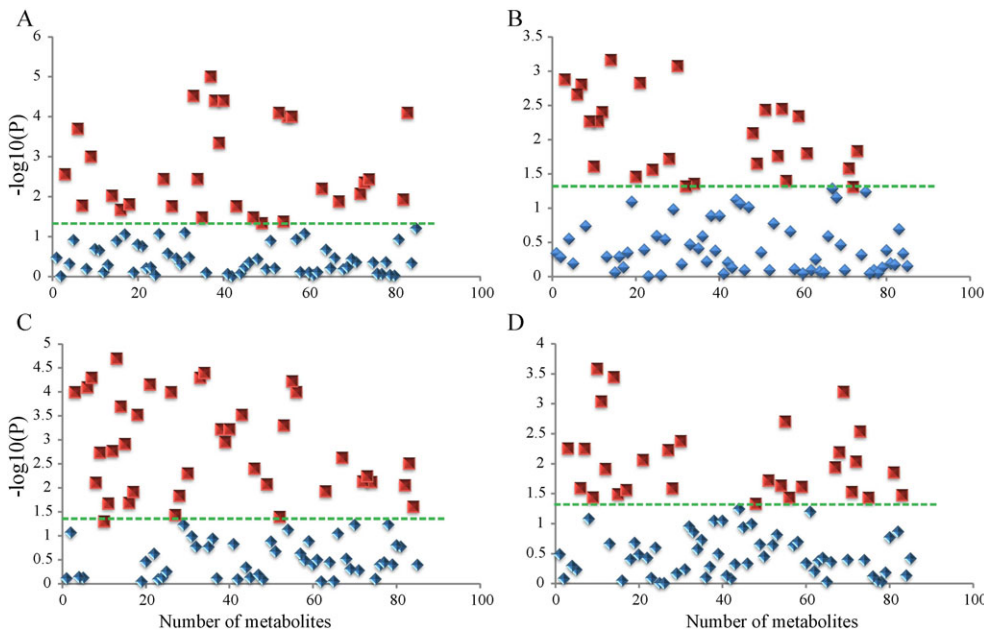


Fig. 2. One-way ANOVA analyses of endotesta and embryo metabolites in walnuts. Metabolites identified as statistically significant ($P < 0.05$; green dotted line) are shown as red squares while non-significant metabolites are shown as blue diamonds. (A) Metabolites of stage 1 to stage 5 in embryos. (B) Metabolites of stage 1 to stage 5 in endotesta. (C) Metabolites of stage 1, stage 3, and stage 5 in embryos. (D) Metabolites of stage 1, stage 3, and stage 5 in endotesta.

Levels of most amino acids were stable, however, not for asparagine, proline, and leucine in endotesta, which decreased during kernel ripening (Fig. 4). Asparagine is a nitrogen transport compound found in many plants, it accounts for 50-70% of the nitrogen carried in translocatory channels serving fruit and seed of white lupin (*Lupinus albus* L.) (Atkins et al., 1975). In soybean, asparagine was proven to have a positive correlation with protein concentration in developing seeds, and the level of asparagine synthesis was high initially and progressively declined in the seed coat (Pandurangan et al., 2012). In the present study the level of asparagine was over tenfold higher in the early stage (stage 1), suggesting that the development of kernels requires a large amount of nitrogen which can be used for protein synthesis at the early stage. Many studies have shown that proline plays a regulatory role in response to biotic and abiotic stress by increasing its content. A

functional link between elevated proline and the ability of tolerance stress of seeds has been demonstrated in *Arabidopsis thaliana* (Hare et al., 2003). In sorghum seeds, proline also has the ability to increase the antioxidant compounds and oil antioxidant activity under normal and drought stress conditions in maize seed (*Zea mays* L.) (Ali et al., 2013). The level of proline is much higher at the early stage compared to other stages and progressively declines from stage 1 to stage 5, suggesting high levels of proline are required at the beginning of kernel development for preventing stress-induced damage. Observations have shown that leucine-derived carbon is incorporated into sugars and organic acids and is an alternate source of acetyl-CoA to sustain respiration and metabolic processes (Anderson et al., 1998). In the present study, leucine is present only at the early stage (stage 1), and barely detected at the later stages (stages 2-5), suggesting that leucine at stage 1 provides

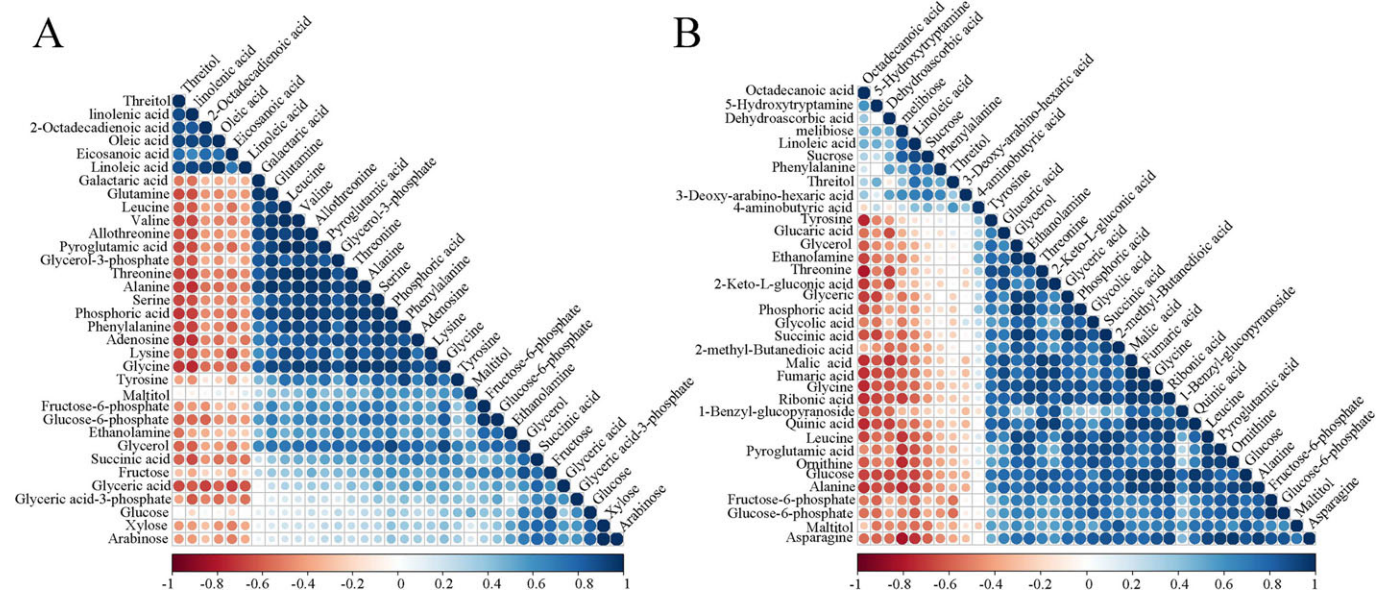


Fig. 3. Metabolite-metabolite correlation analysis. Positive correlations are shown in blue; negative correlations are shown in red. (A) Metabolites-metabolites correlation of endotesta. (B) Metabolites-metabolites correlation of embryos.

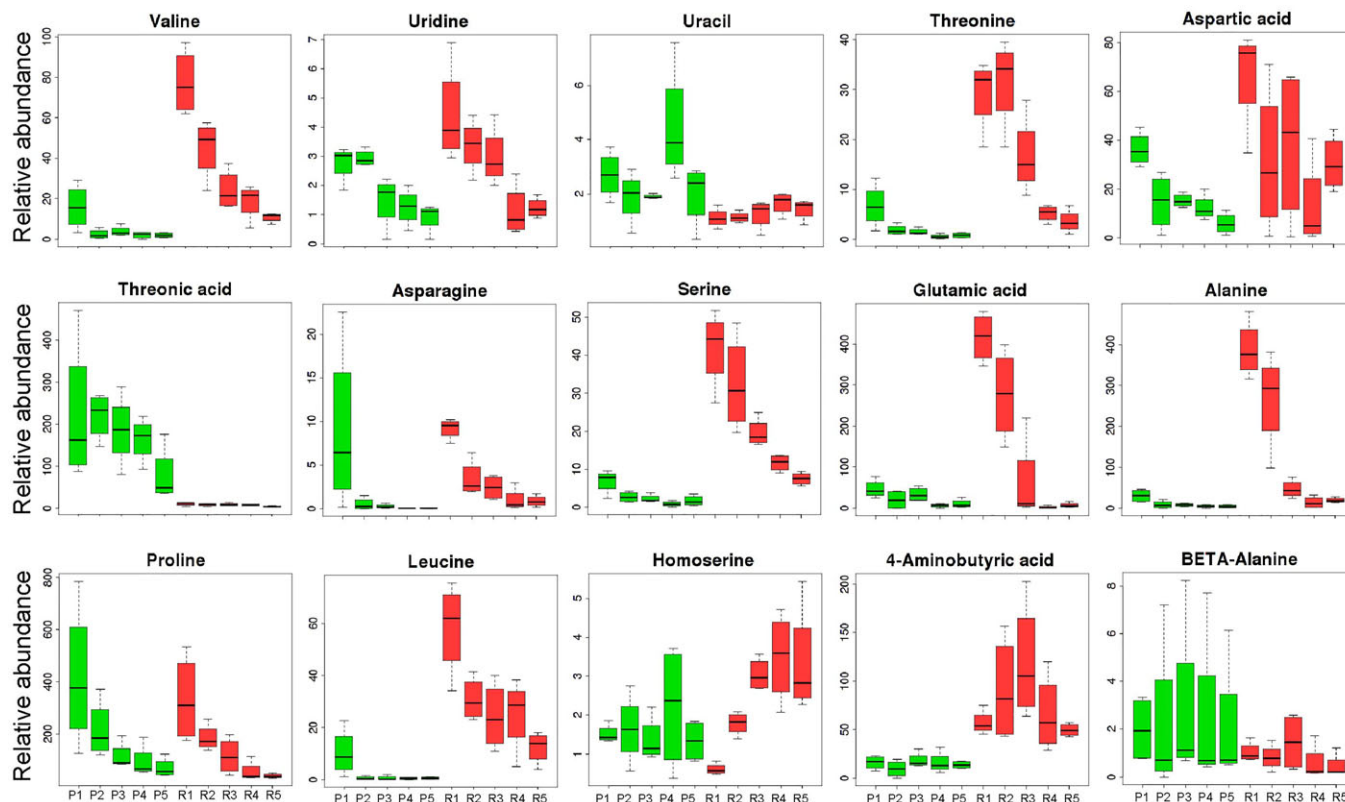


Fig. 4. Boxplot-visualizations of amino acid relative abundances in endotesta and embryo of walnut kernels. The relative abundance of endotesta amino acids at different time points (1-5) are show in green, and the relative abundance of embryo amino acids at different time points (P1-P5; R1-R5) are show in red.

leucine-derived carbon for the subsequent synthesis of other compounds (sugars or organic acids) at the following stages. Endotesta are the barrier between the embryo and outer environment, and proteins in this tissue are mainly comprised of structural proteins, regulatory proteins, and transport proteins. The levels of these proteins are stable during the kernel ripening, which complies with the stability of amino acid levels noted in comparison with the embryo. The embryo is the main part of seed, consisting of precursor tissues for leaves, stem and root. Proteins in the embryo mainly comprise of storage proteins which provide the energy for kernel germination. High levels of amino acids were detected at the early stage and the level of most amino acids decreased during the development of kernel in embryo until only low levels were detected at stage 5 (Fig. 5), suggesting that most of the protein-derived carbon was transferred into other compounds during kernel development. A targeted global metabolomics platform with GC-MS was used to verify the accurate content of amino acids of kernels, which showed the same variation of trends of amino acids contents detected by an untargeted method. A total of 20 amino acids with a decreasing pattern across five development stages were identified, the total content of amino acids decreased from 3282.65 $\mu\text{g/g}$ at T1 to 369.80 $\mu\text{g/g}$ at T5 (Table 1).

Fatty acids are the main component of plant seed oil and are usually unbranched compounds with an even number of carbons ranging from 12 to 22 and with 0 to 3 cis double bonds (Thelen and Ohlrogge, 2002). Six fatty acids including stearic acid, palmitic acid, α -linolenic acid, eicosanoic acid, 9,12-(Z, Z)-octadecadienoic acid, and linoleic acid were detected by the untargeted metabolomic during walnut kernel development. A targeted global metabolomics platform with GC-MS was used to verify the accurate content of fatty acids of kernels. A total of twelve fatty acids with an increasing

pattern across five development stages were identified. The total content of fatty acids were accumulated from 15.02 mg/g at T1 to 157.58 mg/g at T5, among which the unsaturated fatty acids were accumulated from 10.81 mg/g at T1 to 127.24 mg/g at T5 (Table 2).

The level of these six fatty acids increases during the kernel development (Fig. 5), indicating a continuous carbon flow into fatty acids from other compound such as amino acids mentioned above. Unsaturated fatty acid content is one of the important parameters of different vegetable oils. Linoleic acid and α -linolenic acid are the predominant unsaturated fatty acids found in most seed oils, which are required in the diet of higher animals, including humans, as they lack the desaturase that convert oleic acid to linoleic acid and α -linolenic (Gong et al., 2014). Fatty acids also have the function for plant tolerance biotic and abiotic stresses by affect the fluidity of cell membrane (Upchurch, 2008). Linoleic acid is the most abundant fatty acid in plant membranes, the increasing accumulation during the kernel ripening suggesting it may be involved in modifying membranes fluidity necessary for the development of walnut kernel. Free α -linolenic acid has been proved to possess antifungal activity in many plants (Göbel et al., 2002; Walters et al., 2004; Sánchez-Sampedro et al., 2007); it is also a precursor of jasmonic acid which plays a vital role in plant responses to biotic and abiotic stress (Wasternack, 2007; Wasternack and Hause, 2013). The accumulation of α -linolenic acid provided an indicator for the walnut fungal defense response during the kernel development. Three saturated fatty acids (stearic acid, palmitic acid, and eicosanoic acid) were detected to have increased levels during kernel ripening. The level of stearic acid and palmitic acid are much higher than eicosanoic acid, indicating stearic acid and palmitic acid are the main saturated fatty acids in walnut kernel. Interestingly, the levels of stearic acid and palmitic acid are different at the early stage

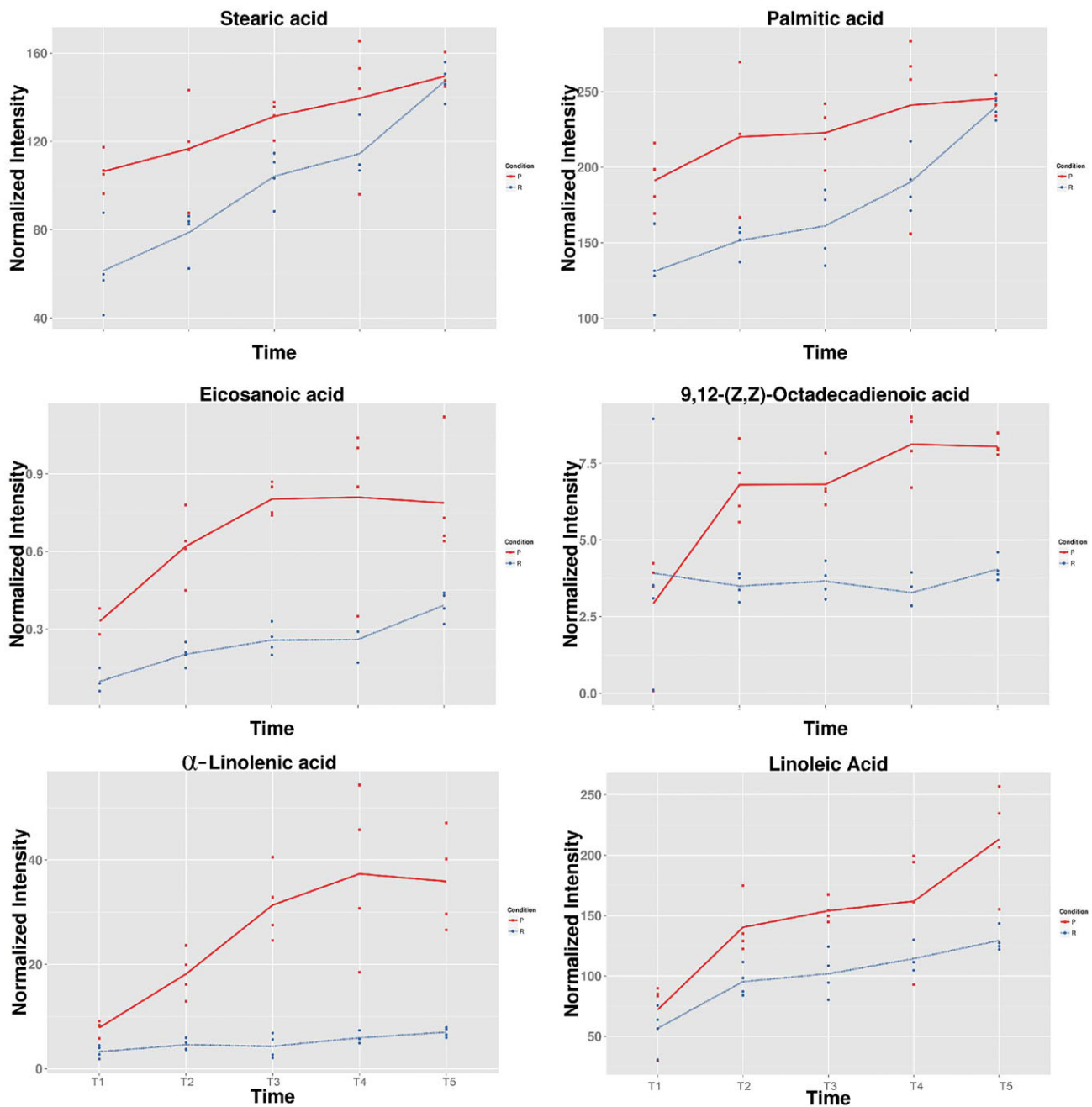


Fig. 5. Fatty acids normalized intensity in endotesta and embryo of walnut kernel. The normalized intensity of endotesta organic acids (P) at different time points (T1-T5) are show in red, and the normalized intensity of embryo amino acids (R) at different time points are show in blue.

of kernel development, while these two saturated fatty acids continuously increase to the same level until ripening, suggesting not only that the embryo has high levels of fatty acids but high levels of fatty acids were in endotesta.

The levels of carbohydrates were higher and relatively stable in endotesta, while they were more dynamic during the kernel development in embryo (Fig. 6). Glucose and cellobiose were predominant during the walnut kernel development; the difference between these two carbohydrates was that the level of glucose was continuously decreasing while cellobiose increased to a high and stable level during kernel ripening. The reverse content changing between these two carbohydrates supports the carbon flow into cellobiose from glucose. Glucose is an energy source that plays an important role in the respiration of several plant tissues. In endotesta

the levels of glucose were stable during development, suggesting that there is a balance between the glucose consumed and synthesized. Meanwhile, the glucose levels at the early stage were much higher than in the later stages in the embryos, suggesting that the glucose was mainly consumed as an energy source for respiration and metabolic synthesis. Cellobiose can be obtained by enzymatic or acidic hydrolysis of cellulose and cellulose rich materials, suggesting that the main carbohydrate of the ripe walnut kernel is cellulose. In higher plants, cell morphogenesis and anisotropic growth were mediated by a cellulose biosynthesis process which played a vital role to support the structural framework of the plant cell wall (Mendu et al., 2011). Cellulose also has been considered to have the key role to strengthen the secondary cell wall, which protects the embryo during seed development (Stork et al.,

Table 1. Amino acid content in developing seeds of walnut kernel ($\mu\text{g/g}$ FW)

Name	T1	T2	T3	T4	T5
α -Ala	356.00 \pm 79.691	132.90 \pm 40.229	52.00 \pm 2.691	5.00 \pm 0.566	4.70 \pm 0.301
Val	154.15 \pm 26.677	46.05 \pm 9.478	19.30 \pm 0.953	3.00 \pm 0.208	5.80 \pm 0.764
Leu	53.05 \pm 1.293	2.85 \pm 0.242	NA	NA	NA
Ile	50.75 \pm 8.523	10.00 \pm 3.306	NA	NA	NA
Pro	39.60 \pm 5.385	10.40 \pm 1.071	6.80 \pm 0.282	NA	NA
Gly	77.45 \pm 4.239	41.70 \pm 6.396	NA	NA	NA
Ser	132.80 \pm 19.940	85.30 \pm 8.788	49.70 \pm 4.472	17.55 \pm 1.640	12.20 \pm 1.356
Thr	346.70 \pm 12.497	333.05 \pm 13.135	71.200 \pm 7.038	14.80 \pm 1.798	NA
β -Ala	35.30 \pm 7.745	41.60 \pm 2.262	56.6 \pm 5.546	55.00 \pm 3.815	25.70 \pm 2.643
Met	17.65 \pm 5.567	7.65 \pm 0.739	6.60 \pm 0.893	3.20 \pm 0.541	NA
Pyr	568.80 \pm 36.938	206.10 \pm 20.285	208.60 \pm 15.939	133.00 \pm 12.091	101.40 \pm 10.230
GABA	63.75 \pm 6.594	73.15 \pm 4.707	38.70 \pm 2.027	73.90 \pm 2.352	22.40 \pm 7.333
Glu	794.80 \pm 24.530	609.25 \pm 11.205	258.80 \pm 11.097	110.10 \pm 10.005	41.90 \pm 3.122
Phe	33.25 \pm 2.258	13.05 \pm 1.524	13.60 \pm 1.692	11.70 \pm 1.312	6.80 \pm 1.104
Asp	261.60 \pm 17.684	41.27 \pm 2.915	15.60 \pm 3.534	14.90 \pm 3.466	12.3 \pm 1.181
Orn	25.40 \pm 2.979	11.07 \pm 1.931	6.90 \pm 0.906	NA	NA
Lys	49.95 \pm 1.194	79.20 \pm 2.705	27.70 \pm 1.187	14.00 \pm 0.275	6.40 \pm 0.305
His	163.85 \pm 10.703	103.30 \pm 13.625	95.50 \pm 9.647	91.4 \pm 1.608	91.20 \pm 0.718
Tyr	57.80 \pm 5.014	63.25 \pm 5.063	14.40 \pm 1.808	5.50 \pm 0.949	NA
Trp	NA	NA	NA	NA	39.00 \pm 11.931
Total	3282.65	1911.14	942.00	553.05	369.80

Data represented as mean \pm s.d.

FW, fresh weight; NA, not applicable.

2010). Cellulose is also protects the seed from damage in the environment by seed mucilage that surrounds the seed (Arsovski et al., 2010). The dynamic upregulation of cellulose both in endotesta and embryo of walnut kernel suggests that cellulose plays an important role in carbohydrate metabolism during kernel development. Another two carbohydrates, gulose and fructose, have the same dynamic changing model as glucose; both of them were consumed at the early stage from a relative high level in embryos, indicating that these two carbohydrates are essential for the development of embryos at the early stage either by providing energy sources or as a chemical structure of enzymes.

MATERIALS AND METHODS

Plant materials

Four representative walnut (*Juglans regia* L.) cultivar kernels at five developmental stages were obtained from the Chinese Academy of Forestry, and the endotesta and embryos of kernels were collected separately. The five developmental stages span from fruit filling to ripening which were

harvested at 15 (stage 1), 30 (stage 2), 45 (stage 3), 60 (stage 4), and 80 (stage 5) days after full bloom (DAF). Embryos were aquiform at early stages (stage 1 and stage 2) and hardened as time went on (stage 3 to stage 5). All samples were frozen in liquid nitrogen immediately and stored at -80°C until further processing.

This study was carried out in strict accordance with the recommendations in the guide for observational and field studies (<http://bio.biologists.org/content/manuscript-prep>). All necessary permits were obtained for the described field studies. The sampling of all individuals of *J. regia* L. were approved by Shougong Zhang, director of Chinese Academy of Forestry.

Metabolite extraction, derivatization for GC-MS

Samples were ground under liquid nitrogen to obtain a fine powder, and 100 mg of lyophilized powder per sample was weighed for metabolite extraction. Cold extraction was employed, and the extraction protocol was followed according to previous studies with slight modifications (Weckwerth et al., 2004). 1 ml of cold solvent (maintained at -20°C), comprised of methanol:chloroform:water in a ratio of 5:2:1 together with 50 μl of internal

Table 2. Fatty acid contents in developing seeds of walnut kernel (mg/g FW)

	T1	T2	T3	T4	T5
C16:0	2.58 \pm 0.241	6.54 \pm 0.439	9.71 \pm 1.133	12.69 \pm 2.333	15.61 \pm 2.476
C18:0	0.58 \pm 0.054	1.82 \pm 0.142	4.94 \pm 0.777	7.33 \pm 1.273	9.60 \pm 2.197
C18:1 Δ^{9c}	0.74 \pm 0.366	4.22 \pm 1.057	13.61 \pm 2.357	28.84 \pm 3.185	27.93 \pm 3.933
C18:2 $\Delta^{9c, 12c}$	6.45 \pm 1.230	22.92 \pm 2.983	40.68 \pm 8.054	53.36 \pm 9.951	67.89 \pm 8.548
C18:3 $\Delta^{9c, 12c, 15c}$	2.02 \pm 0.186	6.14 \pm 0.970	10.56 \pm 2.635	17.59 \pm 7.638	23.34 \pm 1.953
C14:0	0.08 \pm 0.002	0.08 \pm 0.003	0.33 \pm 0.002	0.21 \pm 0.007	0.34 \pm 0.005
C16:1 Δ^{9c}	0.21 \pm 0.005	0.24 \pm 0.009	1.01 \pm 0.005	1.02 \pm 0.011	1.04 \pm 0.015
C17:1 Δ^{9c}	0.27 \pm 0.021	0.27 \pm 0.002	1.35 \pm 0.075	0.80 \pm 0.005	1.32 \pm 0.006
C18:3 $\Delta^{6c, 9c, 12c}$	0.56 \pm 0.005	0.56 \pm 0.008	2.76 \pm 0.025	2.81 \pm 0.031	2.77 \pm 0.046
C20:0	0.33 \pm 0.005	0.35 \pm 0.002	1.65 \pm 0.029	0.99 \pm 0.030	1.66 \pm 0.018
C20:1 Δ^{11c}	0.57 \pm 0.006	0.64 \pm 0.020	2.85 \pm 0.019	2.91 \pm 0.044	2.96 \pm 0.053
C22:0	0.63 \pm 0.009	0.63 \pm 0.000	3.13 \pm 0.001	1.88 \pm 0.003	3.13 \pm 0.027
Saturated FAs	4.21	9.41	19.76	23.10	30.33
Unsaturated FAs	10.81	34.98	72.83	107.34	127.24
Total FAs	15.02	44.39	92.59	130.43	157.58

Data represented as mean \pm s.d.

FW, fresh weight; FA, fatty acid.

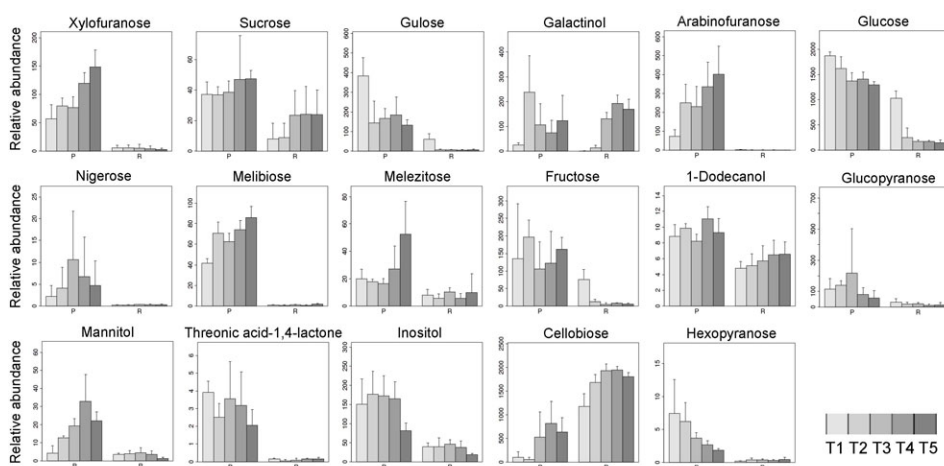


Fig. 6. Relative abundance of carbohydrates in endotesta and embryo of walnut kernels. Bar graph shows the normalized abundance of carbohydrates in endotesta (P) and embryo (R) at different development time points (T1-T5). Data represented as mean \pm s.d.

standard (Ribitol stock concentration, 0.2 mg/ml), was added to the ground material and vortexed for 10 s. The mixtures were then kept on ice for 25 min with intermittent vortexing or shaking for 10 s. The homogenate was then centrifuged at $13,000\times g$, at 4°C for 10 min and the supernatant was transferred to a new tube with 200 μl chloroform and 500 μl of deionized water and vortexed again. The mixture was then centrifuged at $13,000\times g$, at 4°C for 5 min. The solution was separated into the upper and lower phases; the upper phase was aqueous and the lower phase was organic. The aqueous phase was then dried under vacuum and derivatized for GC-MS analysis.

Derivatization steps were carried out as follows. The dried organic phase was dissolved in 20 μl of methoxyamine hydrochloride (M. HCl) solution (40 mg/ml in pyridine) and was kept at room temperature for 2 h. Subsequently, 80 μl of N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was added to the M. HCl mixture and incubated at 37°C for 30 min with shaking. The derivatized samples were centrifuged at $13,000\times g$ for 5 min and then transferred to autosampler vials for GC-MS analysis. A mixed n-alkane standard solution C8–C20 and C21–C40 (Sigma Aldrich, St Louis, MO, USA) was used for the determination of retention indices (RI). The sample set also included a quality control (QC) with an aliquot (40 μl) of all prepared sample extracts mixed.

Untargeted metabolomic analysis

The derivatized samples were analyzed using a global unbiased mass spectrometry-based platform with GC-MS, and data normalization was performed according to a previous study (Zhang et al., 2011). The samples were randomized and the data acquisition was done in one batch to eliminate system errors. GC-MS was carried out on an Agilent 7890A/5975C GC-MS and auto-sampler unit. An HP-5MS (Agilent J&W Scientific, Folsom, CA) column with 0.25 μm thickness, 250 μm diameter, and 30 m length was used to separate derivatized metabolites. 1 μl of sample was injected in split mode in a 1:20 split ratio by the auto-sampler. The injection temperature was set at 280°C and the column oven temperature was 80°C , with helium as the carrier gas. Mass spectrometry settings were as follows: ion source temperature was 250°C , interface temperature was 280°C , and solvent cut time was 5 min. Temperature program: 5 min hold at 40°C , followed by $10^{\circ}\text{C}/\text{min}$ ramp to reach a final temperature of 300°C held for 5 min. The scan range was 35–750 m/z.

Raw GC/MS data were converted into CDF format (NetCDF) files by Agilent GC/MS 5975 Data Analysis software and subsequently processed by the XCMS (www.bioconductor.org) using XCMS's default settings with the following changes: xcmsSet (fwhm=3, snthresh=3, mzdiff=0.5, step=0.1, steps=2, max=300), group (bw=2, minfrac=0.3, max=300). The signal integration area of each metabolite was normalized to the internal standard (ribitol) for each sample. Identification of metabolites using the automated mass spectral deconvolution and identification system (AMIDS) was searched against commercial available databases such as National Institute of Standards and Technology (NIST) and Wiley libraries. Metabolites were confirmed by comparison of mass spectra to the spectra library using a cut-off value of 70%, and by matching the experimental retention time index (RI) of

each metabolite with the mass spectral and RI collection of the Golm Metabolome Database (GMD). Principle component analysis (PCA) was performed with R software (www.r-project.org). Heat map packages available in R were used to draw heat maps, and Mev (MultiExperiment Viewer) 4.8 software was used to perform a one-way ANOVA with standard Bonferroni correction. Identified metabolites were mapped onto general biochemical pathways according to annotation in KEGG. Metabolic network maps were constructed by incorporating the identified and annotated metabolites using Cytoscape 3.2.0 software (<http://www.cytoscape.org/>).

Targeted metabolomic analysis of fatty acid and amino acid

The extraction of fatty acids was conducted with a chloroform-methanol (1:2) mixture, according to the method of previous study (Bligh and Dyer, 1959; Thurnhofer and Vetter, 2005). Fatty acids were transesterified with 2 ml 1% methanol with an internal standard (D-25 tridecanoic acid) were added in sulfuric acid for 2 h at 80°C , then 2 ml μl H_2O and 2 ml hexane were added and vortex mixed. The derivatised organic layer was injected in an Agilent 7890A/5975C GC-MS in the selected ion monitoring (SIM) mode with a DB-WAX (Agilent J&W Scientific; 30 m \times 0.25 mm ID \times 0.25 μm) column. The initial column temperature was 50°C , after held for 3 min the column temperature was increased to 220°C by 10°C per minute, and held for 20 min. The MS ion source temperature was 230°C and the Quadrupole temperature was 150°C . Supelco 37 component FAME mix (Sigma-Aldrich) were used as reference standards to identify the fatty acids. The quantity of each methyl fatty esters was calculated from calibration curves of standards. For analysis of soluble amino acid content, 50 mg of samples were ground to powder and extracted in 1 ml methanol:chloroform: water in a ratio of 5:2:1 together with Norvaline (50 μl 50 $\mu\text{g}/\text{ml}$) as an internal standard and selected ion monitoring (SIM) mode was performed instead of full scan mode (Amira et al., 2005). The quantity of each amino acid was calculated from calibration curves of standards.

Acknowledgements

We thank Suzhou BioNovoGene (<http://www.bionovogene.com>) for help with data analysis, Tao Xu for technical support and LetPub (www.letpub.com) for linguistic assistance during the preparation of this manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Planning: G.R., J.Z. Sampling: J.S. Sample preparation, sample analysis and data analysis: G.R., J.S. Manuscript preparation and editing: G.R., J.Z.

Funding

This work was supported by grants from the National Natural Science Foundation of China [31400569], the Fundamental Research Funds for the Central Non-profit Research Institution of CAF [CAFYBB2014QB028], Collaborative Innovation Plan of Jiangsu Higher Education [2013-2015], the Fundamental Research Funds for the Central Non-profit Research Institution of CAF [RIF2013-11].

Supplementary information

Supplementary information available online at
<http://bio.biologists.org/lookup/doi/10.1242/bio.017863.supplemental>

References

- Abdallah, I. B., Tlili, N., Martinez-Force, E., Rubio, A. G. P., Perez-Camino, M. C., Albouchi, A. and Boukhchina, S. (2015). Content of carotenoids, tocopherols, sterols, triterpenic and aliphatic alcohols, and volatile compounds in six walnuts (*Juglans regia* L.) varieties. *Food Chem.* **173**, 972-978.
- Ali, Q., Anwar, F., Ashraf, M., Saari, N. and Perveen, R. (2013). Ameliorating effects of exogenously applied proline on seed composition, seed oil quality and oil antioxidant activity of maize (*Zea mays* L.) under drought stress. *Int. J. Mol. Sci.* **14**, 818-835.
- Amira, G., Ifat, M., Tal, A., Hana, B., Shmuel, G. and Rachel, A. (2005). Soluble methionine enhances accumulation of a 15 kDa zein, a methionine-rich storage protein, in transgenic alfalfa but not in transgenic tobacco plants. *J. Exp. Bot.* **56**, 2443-2452.
- Anderson, M. D., Che, P., Song, J., Nikolau, B. J. and Wurtele, E. S. (1998). 3-Methylcrotonyl-coenzyme A carboxylase is a component of the mitochondrial leucine catabolic pathway in plants. *Plant Physiol.* **118**, 1127-1138.
- Arsovski, A. A., Haughn, G. W. and Western, T. L. (2010). Seed coat mucilage cells of *Arabidopsis thaliana* as a model for plant cell wall research. *Plant Signal. Behav.* **5**, 796-801.
- Atkins, C. A., Pate, J. S. and Sharkey, P. J. (1975). Asparagine metabolism—key to the nitrogen nutrition of developing legume seeds. *Plant Physiol.* **56**, 807-812.
- Blekhman, R., Perry, G. H., Shahbaz, S., Fiehn, O., Clark, A. G. and Gilad, Y. (2014). Comparative metabolomics in primates reveals the effects of diet and gene regulatory variation on metabolic divergence. *Sci. Rep.* **4**, 5809.
- Bligh, E. G. and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911-917.
- Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanon, M.-I., Nunes-Nesi, A., Nikiforova, V., Centero, D., Ratzka, A., Pauly, M. et al. (2006). Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.* **142**, 1380-1396.
- Cosmulescu, S., Trandafir, I. and Nour, V. (2014). Seasonal variation of the main individual phenolics and juglone in walnut (*Juglans regia*) leaves. *Pharm. Biol.* **52**, 575-580.
- Göbel, C., Feussner, I., Hamberg, M. and Rosahl, S. (2002). Oxylipin profiling in pathogen-infected potato leaves. *Biochim. Biophys. Acta* **1584**, 55-64.
- Gong, Y., Wan, X., Jiang, M., Hu, C., Hu, H. and Huang, F. (2014). Metabolic engineering of microorganisms to produce omega-3 very long-chain polyunsaturated fatty acids. *Prog. Lipid Res.* **56**, 19-35.
- Hare, P. D., Cress, W. A. and van Staden, J. (2003). A regulatory role for proline metabolism in stimulating *Arabidopsis thaliana* seed germination. *Plant Growth Regul.* **39**, 41-50.
- Mao, X. and Hua, Y. (2012). Composition, structure and functional properties of protein concentrates and isolates produced from walnut (*Juglans regia* L.). *Int. J. Mol. Sci.* **13**, 1561-1581.
- Mao, X.-Y. and Hua, Y.-F. (2014). Chemical composition, molecular weight distribution, secondary structure and effect of NaCl on functional properties of walnut (*Juglans regia* L.) protein isolates and concentrates. *J. Food Sci. Technol.* **51**, 1473-1482.
- Mao, X., Hua, Y. and Chen, G. (2014). Amino acid composition, molecular weight distribution and gel electrophoresis of walnut (*Juglans regia* L.) proteins and protein fractionations. *Int. J. Mol. Sci.* **15**, 2003-2014.
- Martinez, M. L., Labuckas, D. O., Lamarque, A. L. and Maestri, D. M. (2010). Walnut (*Juglans regia* L.): genetic resources, chemistry, by-products. *J. Sci. Food Agric.* **90**, 1959-1967.
- Mendu, V., Griffiths, J. S., Persson, S., Stork, J., Downie, A. B., Voiniciuc, C., Haughn, G. W. and DeBolt, S. (2011). Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. *Plant Physiol.* **157**, 441-453.
- Miraliakbari, H. and Shahidi, F. (2008). Oxidative stability of tree nut oils. *J. Agric. Food Chem.* **56**, 4751-4759.
- Moser, B. R. (2012). Preparation of fatty acid methyl esters from hazelnut, high-oleic peanut and walnut oils and evaluation as biodiesel. *Fuel* **92**, 231-238.
- Nour, V., Trandafir, I. and Cosmulescu, S. (2013). HPLC determination of phenolic acids, flavonoids and juglone in walnut leaves. *J. Chromatogr. Sci.* **51**, 883-890.
- Ozrenk, K., Javidipour, I., Yarılgac, T., Balta, F. and Gundogdu, M. (2012). Fatty acids, tocopherols, selenium and total carotene of pistachios (*P. vera* L.) from Diyarbakir (Southeastern Turkey) and walnuts (*J. regia* L.) from Erzincan (Eastern Turkey). *Food Sci. Technol. Int.* **18**, 55-62.
- Pandurangan, S., Pajak, A., Molnar, S. J., Cober, E. R., Dhaubhadel, S., Hernandez-Sebastian, C., Kaiser, W. M., Nelson, R. L., Huber, S. C. and Marsolais, F. (2012). Relationship between asparagine metabolism and protein concentration in soybean seed. *J. Exp. Bot.* **63**, 3173-3184.
- Ros, E., Núñez, I., Pérez-Heras, A., Serra, M., Gilabert, R., Casals, E. and Deulofeu, R. (2004). A walnut diet improves endothelial function in hypercholesterolemic subjects: a randomized crossover trial. *Circulation* **109**, 1609-1614.
- Sánchez-Sampedro, A., Kim, H. K., Choi, Y. H., Verpoorte, R. and Corchete, P. (2007). Metabolomic alterations in elicitor treated *Silybum marianum* suspension cultures monitored by nuclear magnetic resonance spectroscopy. *J. Biotechnol.* **130**, 133-142.
- Stork, J., Harris, D., Griffiths, J., Williams, B., Beisson, F., Li-Beisson, Y., Mendu, V., Haughn, G. and DeBolt, S. (2010). CELLULOSE SYNTHASE9 serves a nonredundant role in secondary cell wall synthesis in *Arabidopsis* epidermal testa cells. *Plant Physiol.* **153**, 580-589.
- Thelen, J. J. and Ohlrogge, J. B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. *Metab. Eng.* **4**, 12-21.
- Thurnhofer, S. and Vetter, W. (2005). A gas chromatography/electron ionization–mass spectrometry–selected ion monitoring method for determining the fatty acid pattern in food after formation of fatty acid methyl esters. *J. Agric. Food Chem.* **53**, 8896-8903.
- Topfer, N., Kleessen, S. and Nikoloski, Z. (2015). Integration of metabolomics data into metabolic networks. *Front. Plant Sci.* **6**, 49.
- Upchurch, R. G. (2008). Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnol. Lett.* **30**, 967-977.
- Walters, D., Raynor, L., Mitchell, A., Walker, R. and Walker, K. (2004). Antifungal activities of four fatty acids against plant pathogenic fungi. *Mycopathologia* **157**, 87-90.
- Wasternack, C. (2007). Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann. Bot.* **100**, 681-697.
- Wasternack, C. and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. *Ann. Bot.* **111**, 1021-1058.
- Weckwerth, W., Wenzel, K. and Fiehn, O. (2004). Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4**, 78-83.
- Zhang, J., Wang, X., Yu, O., Tang, J., Gu, X., Wan, X. and Fang, C. (2011). Metabolic profiling of strawberry (*Fragaria xananassa* Duch.) during fruit development and maturation. *J. Exp. Bot.* **62**, 1103-1118.
- Zhang, J., Zhang, Y., Du, Y., Chen, S. and Tang, H. (2011). Dynamic metabolomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res.* **10**, 1904-1914.