RESEARCH

Comparative transcriptome analysis of vegetable soybean grain discloses genes essential for grain quality

Bin Wang[1,2]*, Yuanpeng Bu[1,2], Guwen Zhang[1,2], Na Liu[1,2], Zhijuan Feng[1,2] and Yaming Gong[1,2]*

Abstract

Background Vegetable soybean is an important vegetable crop in world. Seed size and soluble sugar content are considered crucial indicators of quality in vegetable soybean, and there is a lack of clarity on the molecular basis of grain quality in vegetable soybean.

Results In this context, we performed a comprehensive comparative transcriptome analysis of seeds between a high-sucrose content and large-grain variety (Zhenong 6, ZN6) and a low-sucrose content and small-grain variety (Williams 82, W82) at three developmental stages, i.e. stage R5 (Beginning Seed), stage R6 (Full Seed), and stage R7 (Beginning Maturity). The transcriptome analysis showed that 17,107 and 13,571 differentially expressed genes (DEGs) were identified in ZN6 at R6 (vs. R5) and R7 (vs. R6), respectively, whereas 16,203 and 16,032 were detected in W82. Gene expression pattern and DEGs functional enrichment proposed genotype-specific biological processes during seed development. The genes participating in soluble sugar biosynthesis such as *FKGP* were overexpressed in ZN6, whereas those responsible for lipid and protein metabolism such as *ALDH3* were more enhanced in W82, exhibiting different dry material accumulation between two genotypes. Furthermore, hormone-associated transcriptional factors involved in seed size regulation such as *BEH4* were overrepresented in ZN6, exhibiting different seed size regulation processes between two genotypes.

Conclusions Herein, we not only discovered the differential expression of genes encoding metabolic enzymes involved in seed composition, but also identified a type of hormone-associated transcriptional factors overexpressed in ZN6, which may regulate seed size and soluble content. This study provides new insights into the underlying causes of differences in the soybean metabolites and appearance, and suggests that genetic data can be used to improve its appearance and textural quality.

Keywords Vegetable soybean, Grain quality, Soluble sugar, Large grain, Transcriptome analysis

---

*Correspondence:
Bin Wang
wb5682561@163.com
Yaming Gong
gongym07@126.com
[1]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, 198, Shiqiao Rd, Hangzhou 310021, Zhejiang, China
[2]Key Laboratory of Vegetable Legumes Germplasm Enhancement and Molecular Breeding in Southern China of Ministry of Agriculture and Rural Affairs, 198, Shiqiao Rd, Hangzhou 310021, Zhejiang, China

Background

Vegetable soybean (*Glycine max* L.), also called "edamame", is harvested at full-seed development stage [1, 2]. Vegetable soybean has been extensively used for more than 2000 years in East Asian countries [3]. In recent years, acceptance of vegetable soybean-based food products has been reported in North and South America, Europe, and Africa [4]. Vegetable soybean has become a popular food all over the world because of their taste and nutritional value.

For vegetable soybean, grain quality has three characteristics, i.e. eating quality, appearance quality, and nutritional quality [5, 6]. The seeds of soybean are rich in protein, oil, and carbohydrates; as much as 47% of the carbohydrates are soluble sugars [7, 8]. Soluble sugar content is an important determinant of eating quality for vegetable soybean [9]. Carbohydrates in soybean determine the taste and quality of soyfoods [10]. Soybean soluble sugars can also be used as a functional factor or food additive to produce various health foods [10–13]. Vegetable soybeans are high sucrose content cultivars. Soluble sugar content in seeds increased during early developmental stage and reached a peak at full seed stage (R6), thereafter decreasing [14, 15]. Studies have shown that abundant genetic variations in total soluble sugar content of soybean germplasm provide abundant germplasm resources for breeding [16]. Soluble sugar content is a complex quality trait controlled by multiple genes and influenced by distinct environmental factors [17]. Until now, few QTLs associated with a soluble sugar content of soybean seeds were characterized. Metabolome and transcriptome analyses have been successfully employed to identify key gene regulatory networks of soluble sugar in soybean seeds [18]. Compared with grain soybean, the free amino acids, carbohydrates, sterols and flavonoids were increased in vegetable soybean. Combined with transcriptome analysis, multiple DEGs were identified, including starch and sucrose metabolism and sucrose transport pathway genes [19]. However, the specific regulatory mechanisms are still unclear.

Seed size is an important aspect of appearance quality for vegetable soybean [20]. It is generally believed that 100-grain dry weight of vegetable soybean is more than 20–30 g. Seed size is a complex trait controlled by many quantitative trait loci (QTL). Only a small subset of genes has been functionally characterized to be involved in soybean seed size regulation. *Seed thickness 1* (*ST1*), encoding a UDP-D-glucuronate 4-epimerase, affected seed thickness by influencing the pectin biosynthesis [21]. *Seed Thickness 05* (*GmST05*), encoding a transcriptional factor, positively controlled soybean seed thickness and size *via* regulating *GmSWEET10a* expression level [22]. A *phosphatase 2 C-1* (*PP2C-1*) allele from wild soybean ZYD7 contributed to the increase in seed size by modulating the brassinosteroid signaling pathway [23]. Furthermore, it has been found that there was an association between seed size and seed composition. *ST1*, *GmST05*, and *PP2C-1* regulated seed weight, simultaneously affecting oil content in soybean.

Recent studies suggested that vegetable soybeans were domesticated independently and clustered on a separate branch distinct from grain and wild soybean [24]. Vegetable soybeans exhibit large seeds and high soluble sugar content, because of domestication and breeding for improving soybean quality. Intrigued by these findings, we wondered that what factors simultaneously affected seed quality (seed size and soluble sugar content) in vegetable soybean. In this study, we characterized the grain quality of two genotypes, Zhenong6 (ZN6) and Williams82 (W82). ZN6, an outstanding commercial vegetable soybean cultivar in China, has large grain and high soluble sugar content. In contrast, the grain soybean cultivar, W82, exhibited precisely the opposite phenotype. Comparative transcriptome analysis was then employed for the seeds of the two genotypes to clarify two major issues: (i) what are the differences in the transcriptional regulation associated with different genotypes; (ii) what is the genetic basis for the different phenotypes of seed composition and size between the two genotypes. On the basis of bioinformatics analysis, it is expected that results of this study could elucidate the molecular mechanism underlying grain quality of vegetable soybean and provide a fundamental basis for genetic improvement of high-quality vegetable soybean.

Results

Significant grain size and seed composition differences between ZN6 and W82

ZN6 is a vegetable soybean variety widely grown in southern China. Compared to grain soybean variety W82, ZN6 has a larger grain size (Fig. 1A; Fig. S1). At the R6 stage, the size of ZN6 seeds is approximately 2 times larger than that of W82 seeds. In addition, we found that there are also significant differences between the two varieties in the seed composition (Fig. 1B and D). Among different seed development stages, the soluble sugar content showed a general trend of increasing first and then decreasing in ZN6 and W82 (Fig. 1B). The highest point appeared around R6, at which time the soluble sugar content of ZN6 was always about two-fold higher than that of W82. With the development of seeds, we found that the oil content and the protein content increased gradually. The protein and oil content of W82 was always higher than that of ZN6 at the R5-R7 stage of seed development (Fig. 1C-D; Fig. S2). These results confirmed genotype-dependent differences in soybean appearance quality and nutritional quality, as shown in previous studies. Genetic stability of soybean seed composition
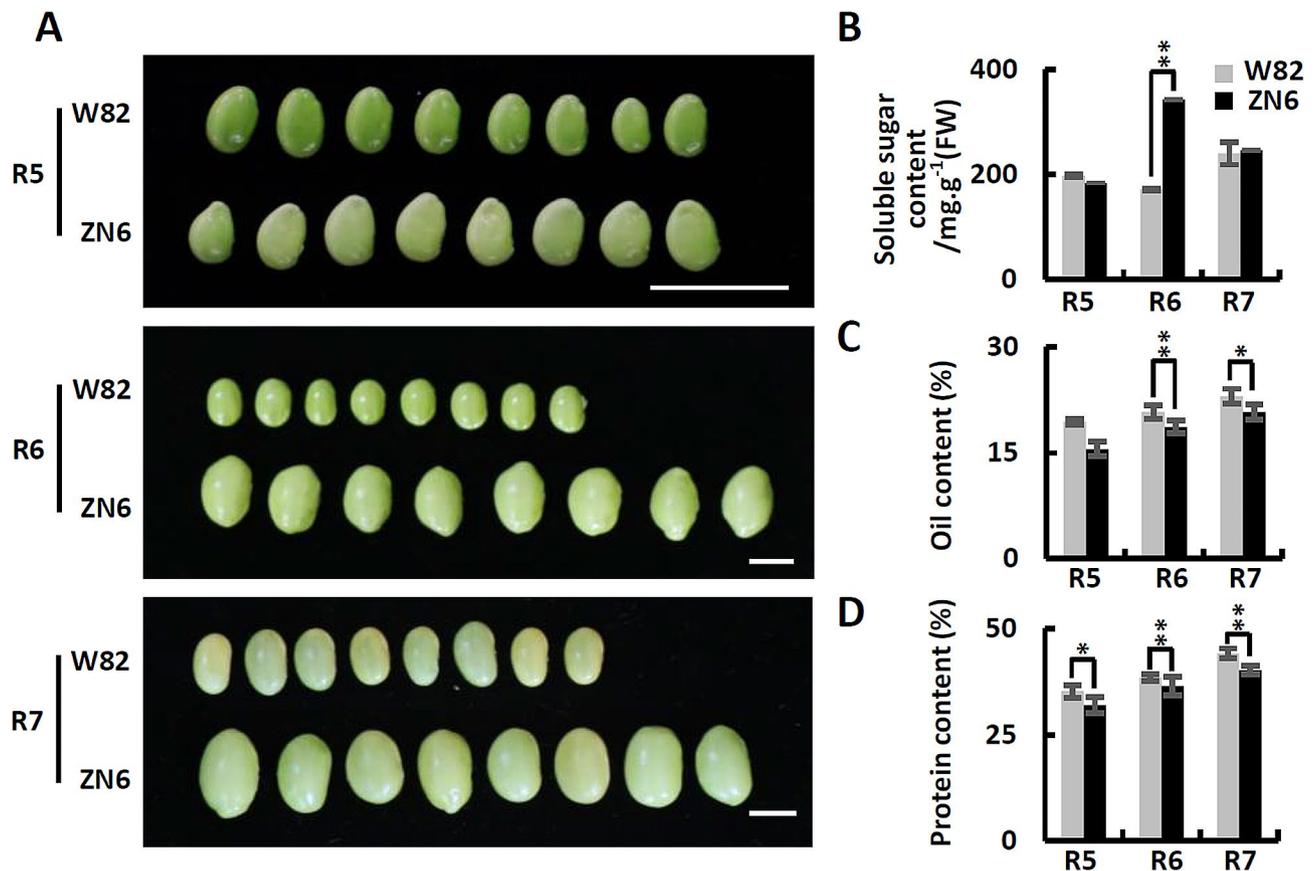
Fig. 1 Phenotype differences between high-sucrose content and large-grain variety (Zhenong 6, ZN6) and low-sucrose content and small-grain variety (Williams 82, W82) at three developmental stages. (**A**) Visual appearance of ZN6 and W82 seeds at stage R5 (Beginning Seed), stage R6 (Full Seed), and stage R7 (Beginning Maturity). Scale bars = 1 cm. (**B**) Dynamic changes of soluble sugar content in ZN6 and W82 seeds at stage R5, stage R6, and stage R7. (**C**) Dynamic changes of protein content in ZN6 and W82 seeds at stage R5, stage R6, and stage R7. (**D**) Dynamic changes of lipid content in ZN6 and W82 seeds at stage R5, stage R6, and stage R7. The data are presented as the mean ± SE of three replicates. Asterisks indicate the significant differences between the two groups (Student's t test, * $p < 0.05$; ** $p < 0.01$)

Table 1 Summary of transcriptome data by the Illumina platform

Sample	ZN6-R5	ZN6-R6	ZN6-R7	W82-R5	W82-R6	W82-R7	Sum/Ave
Raw reads	47,952,208	42,822,501	44,738,343	47,751,998	49,732,024	45,668,317	278,665,391
Raw bases	7,192,831,200	6,423,375,200	6,710,751,500	7,162,799,700	7,459,803,600	6,850,247,500	41,799,808,700
Q20 (%)	96.86	97.21	96.98	96.94	97.03	97.26	97.05
Q30 (%)	91.75	92.44	91.96	91.90	92.09	92.58	92.12
Clean reads	47,796,749	42,681,775	44,587,595	47,603,993	49,560,619	45,519,825	277,750,556
Clean bases	6,895,969,926	6,851,491,035	6,974,214,310	7,084,154,032	7,080,872,671	6,782,636,810	41,669,338,782
Q20 (%)	97.02	97.37	97.14	97.10	97.19	97.42	97.21
Q30 (%)	91.95	92.63	92.15	92.09	92.29	92.77	92.31
Mapped reads	46,032,705	41,073,613	42,827,007	46,093,739	47,568,359	43,938,133	267,533,556
Proportion (%)	96.41	96.84	96.34	96.99	96.81	96.91	96.72

and size implies differences in gene involvement between different varieties.

Transcriptome profiles of the two genotypes during seed development

To identify the gene expression changes during grain development, we performed comparative transcriptomic analysis of W82 and ZN6 grains at R5, R6, and R7 stage. A total of 142,684,437 and 135,066,119 clean reads (47,603,993 for W82-R5, 49,560,619 for W82-R6, 45,519,825 for W82-R7, 47,796,749 for ZN6-R5, 42,681,775 for ZN6-R6, and 44,587,595 for ZN6-R7) were identified from the W82 and ZN6 transcriptome, respectively (Table 1; Fig. S3). An average of 96.72% of the reads were mapped to the soybean genome *Glycine max* Wm82.a4.v1 [25]. Using a cutoff of ≥2-fold change,

a total of 16,203 and 17,107 DEGs were detected in W82 and ZN6 at R6 stage, respectively, as compared with those at R5 stage, whereas 16,032 and 13,571 genes were differentially expressed at R7 respectively, as compared with those at R6 stage. W82 exhibited large transcriptional changes during later stages of seed development (Fig. S4). According to previous research, the developing stages (stages R5 to R7) refer to the most active phase of grain filling [26, 27]. For legume, sugar significantly preferentially accumulated in seeds before stage R6. After stage R6, most of the reduced sucrose may be converted into oil or protein during seed development [26, 27]. The transcriptional changes at the R6-R7 stages of seed development may partially be related to higher concentrations of total protein and oil in W82 seeds. These results further confirmed that grain quality was genotype-dependent for vegetable soybean.

The transcriptome validation for DEGs by qRT-PCR

To validate these expression patterns of genes from the RNA sequencing (RNA-Seq) results, we randomly selected 8 genes for qRT-PCR analysis. The results showed a high level of consistency between the qRT-PCR and RNA-sequencing data (Fig. 2). For each gene, the expression count values from RNA sequencing data exhibited similar expression profile at all the three stages comparing with the results of qRT-PCR (Fig. S5). The results indicated that the RNA-Seq data of the present study were accurate and reliable.
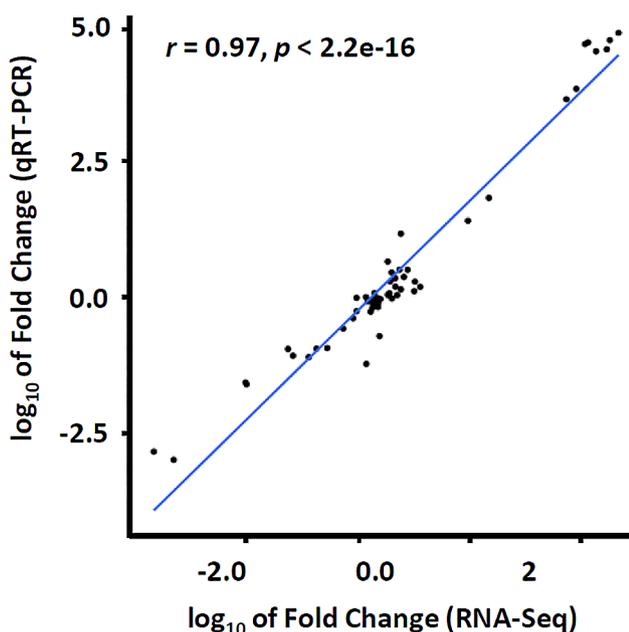
Fig. 2 Correlation between qRT-PCR data and RNA-sequencing for the selected genes. Each point represents a value of fold change of expression level at R6 or R7 comparing with that at R5 or R6. Fold change values were transformed by $\log_{10}$

Gene expression pattern analysis and functional enrichment of DEGs

Considering the differences in the genetic backgrounds of W82 and ZN6, we analyzed differences in gene expression pattern to find key genes that regulate grain quality [28, 29]. DEGs of each W82 and ZN6 at different time stages were clustered in eight profiles based on gene expression pattern (Fig. 3A). The 16 profiles were further divided into four classes: ZN6 up-regulated genes group (Profile 4, profile 6, and profile 7 from ZN6), ZN6 down-regulated genes group (Profile 0, profile 1, and profile 3 from ZN6), W82 up-regulated genes group (Profile 4, profile 6, and profile 7 from W82), and W82 down-regulated genes group (Profile 0, profile 1, and profile 3 from W82). To explore genes involved in grain quality in ZN6, we examined gene sets unique to W82 and ZN6 by performing clustering analysis with specific gene groups, respectively (Fig. 3B). A total of 1426 (Group 1 for W82) and 1593 (Group 2 for ZN6) genes were specifically upregulated during seed development (Fig. 3B; Table S2-S5). KEGG pathway enrichment analysis further revealed soluble sugar accumulation and seed size regulation ($p < 0.05$) were overrepresented simultaneously in ZN6 (Fig. 3C; Table S6). In contrast, genes involved in lipid and protein accumulation ($p < 0.05$) were significantly enriched in W82 (Fig. 3B; Fig. S6-S8; Table S6). These results suggested genotype-specific variation in genes and pathways regulating grain quality. The differentiation in gene expression patterns between the two genotypes was corresponding to their distinct seed composition accumulation and seed size regulation.

Differential expression of genes involved in metabolic pathways contributed to seed composition variation in W82 and ZN6

To uncover the genes responsible for the different nature of seed composition accumulation between W82 and ZN6, gene expression pattern and DEGs functional enrichment were performed. The activity of the enzymes in the metabolite synthesis were suggested to play the major role in genotype difference of seed composition accumulation [30]. Between the two genotypes, nine DEGs involved in monosaccharides biosynthesis were upregulated in ZN6 (Fig. 4; Table S6). 43 DEGs belonged to three GO terms involved in lipid and protein metabolism, which were upregulated in W82 (Fig. 4; Table S6). Besides these genes, another two DEGs encoding serine/threonine-protein kinase SRK2, which have also been proposed to be related to sucrose accumulation in seeds [31], were also identified in ZN6 up-regulated genes group (Fig. 4; Table S6). The enhanced expressions of these genes in W82 and ZN6 may thus contribute to change in metabolic flow direction, leading to significant variation in seed components.
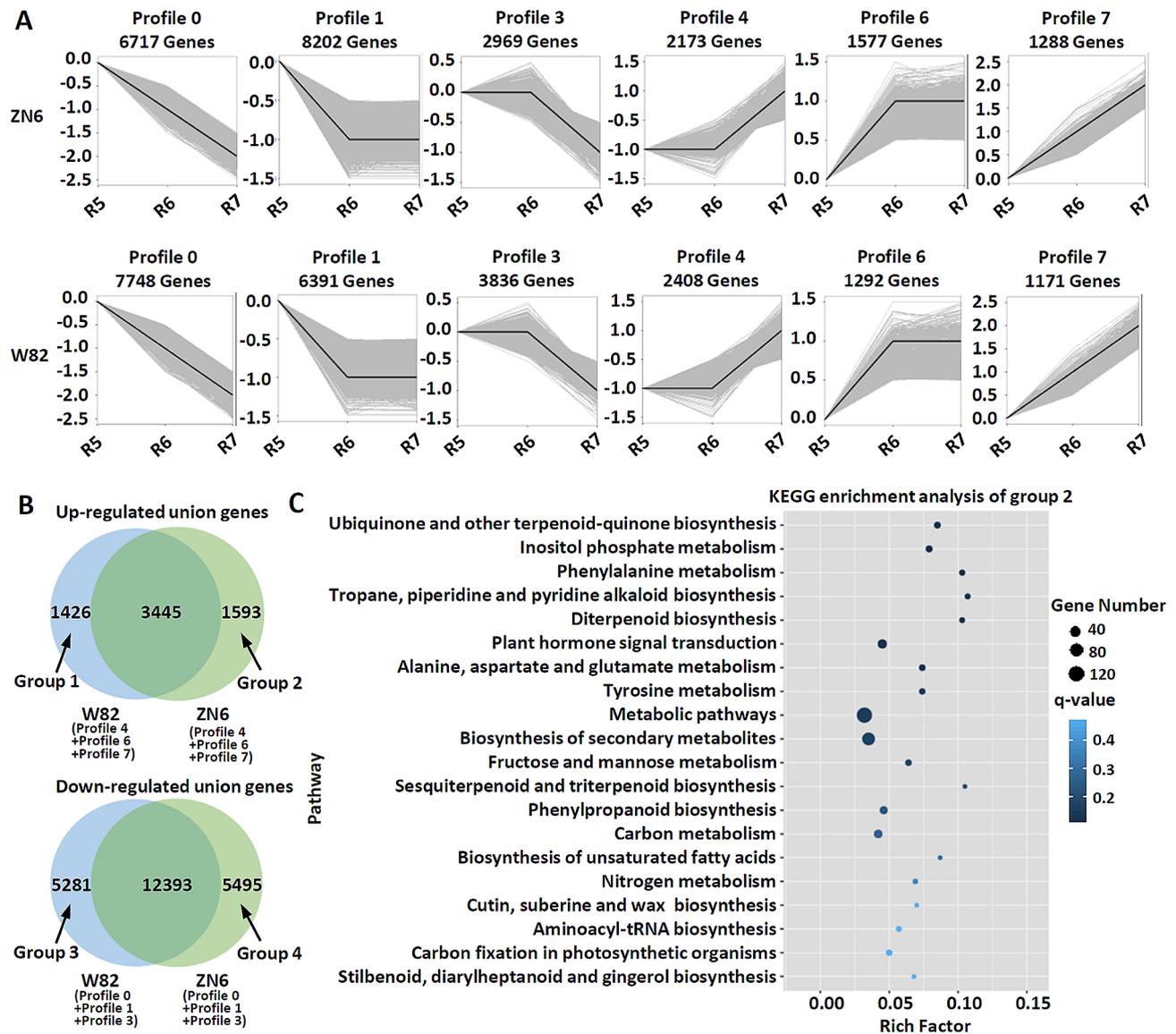
Fig. 3 Patterns of gene expressions and KEGG enrichment across three time points in ZN6 and W82 seeds. (**A**) Patterns of gene expressions across three time points in ZN6 and W82 inferred by STEM analysis. In each frame, the black line represented the expression tendency of all the genes, while the light gray lines represented the expression pattern of each gene. (**B**) Venn diagram analysis of genes groups shows gene sets unique to W82 and ZN6. The profiles were clustered into four classes: ZN6 up-regulated genes group (Profile 4, profile 6, and profile 7 from ZN6), ZN6 down-regulated genes group (Profile 0, profile 1, and profile 3 from ZN6), W82 up-regulated genes group (Profile 4, profile 6, and profile 7 from W82), and W82 down-regulated genes group (Profile 0, profile 1, and profile 3 from W82). (**C**) KEGG enrichment analysis of group 2

In addition, 11 genes involved in free amino acids metabolism, such as glutamate synthase and homogentisate 1,2-dioxygenase, were overexpressed in ZN6 (Fig. 4; Table S6). It is known that free amino acids are closely related to flavor and can be used as indicators of quality and freshness [32, 33]. The overexpression of the genes responsible for free amino acids metabolism was thought to result in the unique flavor of ZN6.

Genes involved in hormone signal transduction acted as seed size control in ZN6

Besides soluble sugar and free amino acids biosynthesis, ZN6 also employed specific regulatory mechanism to improve grain quality. Among the up-regulated genes in ZN6, hormone-associated transcriptional factors genes were found to be overrepresented, most of which are involved in auxin, brassinosteroid, and cytokinin signal transduction (Fig. 4; Table S6). Multiple studies have demonstrated that auxin, brassinosteroid, and cytokinin regulate seed size in plants.
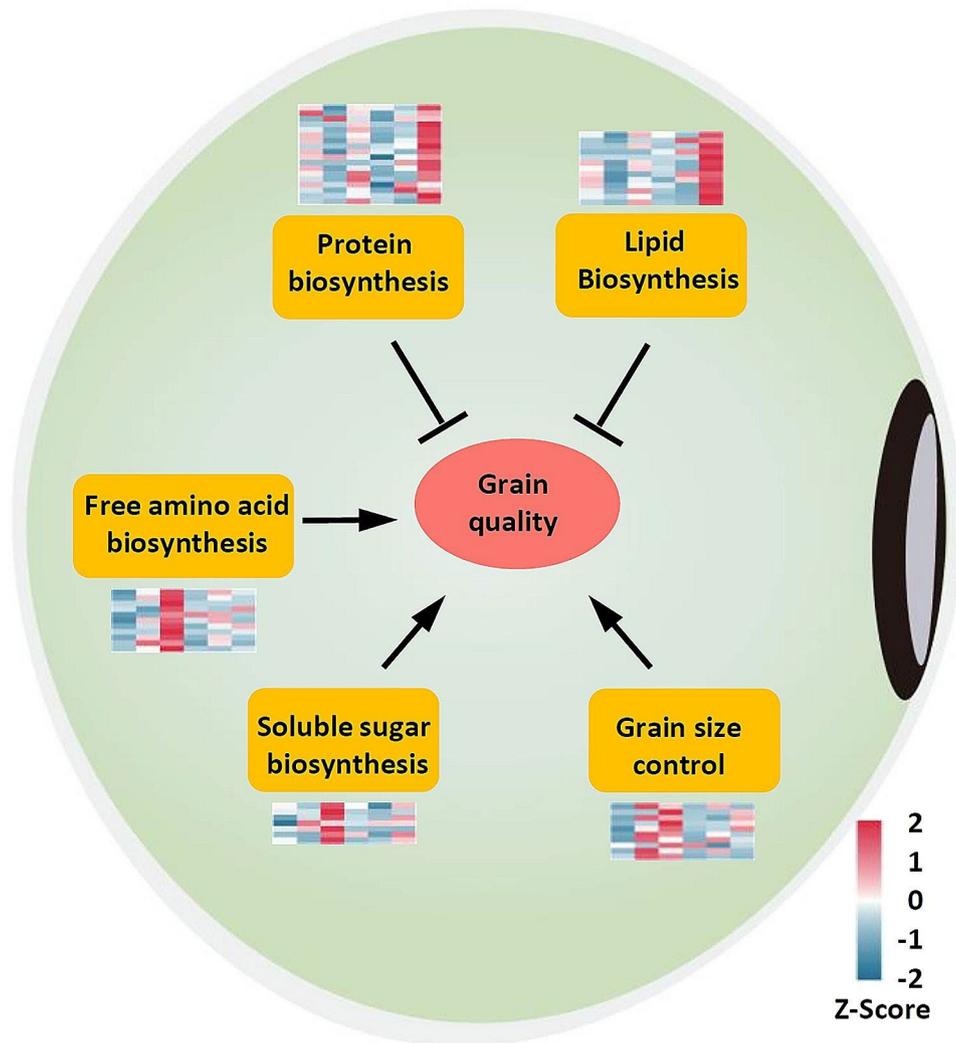
Fig. 4 Transcriptional changes of genes responsible for grain quality formation in seeds of the two soybean genotypes. The metabolites biosynthesis pathway and grain size control responsible for grain quality are represented in orange boxes. The expression pattern of the genes involved in metabolites biosynthesis pathway and grain size control over the three time points in both two genotypes are shown beside each box. Each square represents a comparative pair. The squares from left to right represent: ZN6-R5, ZN6-R6, ZN6-R7, W82-R5, W82-R6, and W82-R7. The red, blue squares indicate their expression levels. Arrows indicate promotion and blockhead arrows inhibition

To confirm that sugar metabolism and grain size control together contributed to grain quality traits in vegetable soybean, we analyzed expression of the selected quality-associated genes in different soybean varieties at stage R6 (Fig. S8A-C). The results showed that genes involved in sugar metabolism and grain size control were overexpressed in vegetable soybean varieties XNQY and ZNQF. In contrast, protein and lipid metabolism genes were upregulated in grain soybean varieties TL and HC6. These results suggested that there are conserved regulatory pathways of grain quality between vegetable soybean varieties.

Discussion

Vegetable soybean has become a popular food all over the world because of their taste and nutritional value. According to these previous studies [24], vegetable soybean was domesticated independently. A cluster analysis found that vegetable soybean is clearly clustered on a separate branch distinct from grain and wild soybean. Many useful quality traits have been bred into vegetable soybean cultivars, such as large seeds and high soluble sugar content. In contrast, for grain soybean, high protein content or oil content were selected. Seed size and soluble content are two important aspects of vegetable soybean quality. Harrigan et al. (2015) selected nine soybean varieties and performed metabolomics analysis of mature grains. They found that glucose and sucrose showed large

variation across varieties [34]. We characterized the grain quality of vegetable soybean cultivar ZN6. Compared to grain soybean variety W82, ZN6 has a larger grain size (Fig. 1A; Fig. S1). In addition, we found that ZN6 contained relatively high levels of soluble sugar (Fig. 1B). On the contrary, W82 contained relatively high levels of protein and oil (Fig. 1C-D; Fig. S2). This, combined with evolutionary analysis, allowed us to speculate that the grain quality in vegetable soybean is genotype-dependent.

Seed composition is a complex quality trait that is controlled by multiple genes and influenced by distinct environmental factors [17]. Until now, few QTLs and genes associated with soybean seed composition were reported. Previous studies also discovered some regulatory genes for soluble sugar accumulation in vegetable soybeans using metabolome and transcriptome [18, 35]. Compared with grain soybean, the free amino acids, carbohydrates, sterols, and flavonoids were increased in vegetable soybean. Combined with transcriptome analysis, multiple DEGs were identified, including starch and sucrose metabolism and sucrose transport pathway genes [19]. In this study, we found that the genes participating in soluble sugar biosynthesis were overexpressed in ZN6 (Fig. 4; Table S6), whereas those responsible for lipid and protein metabolism were more enhanced in W82, exhibiting different dry material accumulation between two genotypes.

The synergistic interaction between maternal and zygotic tissues determines the final seed size [36]. Recent advances have identified several signaling pathways that control seed size, including or involving the ubiquitin-proteasome pathway, G-protein signaling, mitogen-activated protein kinase (MAPK) signaling, phytohormone perception and homeostasis, and some transcriptional regulators. Brassinosteroids (BRs) and auxin are two major hormones involved in maternal control of seed size. The BR biosynthesis mutants and BR insensitive mutants produced short seeds, suggesting that BR promotes seed growth [34, 37–42]. Auxin homeostasis can influence seed size by regulating endosperm development. Evidence from Arabidopsis and rice suggested the roles of auxin signaling in seed size control [40, 43–50]. In this study, we found that hormone-associated transcriptional factors genes were found to be overrepresented in ZN6, most of which are involved in auxin, brassinosteroid, and cytokinin signal transduction (Fig. 4; Table S6). Previous studies have found that there is an association between seed size and seed composition. *ST1*, *GmST05*, and *PP2C-1* regulated seed weight, simultaneously affecting oil content in soybean [19, 21–23]. Thus, we speculate that hormone-associated transcriptional factors genes in ZN6 regulated seed size and affected the accumulation of soluble sugar.

Conclusion

In summary, there are two important conclusions from this work: (i) many genes related to soluble sugar, lipid, and protein metabolism were differentially expressed between W82 and ZN6 during seed development; (ii) hormone-associated transcriptional factors genes overexpressed in ZN6 regulated seed size and affected the accumulation of soluble sugar (Fig. 4). This study has greatly enhanced our knowledge of the genetic regulatory network of grain quality for vegetable soybean, providing important clues for molecular assisted screening and breeding of high-quality cultivars for vegetable soybean.

Materials and methods

Plant material and sampling

Two identified soybean genotypes, a high-sucrose content and large-grain variety (Zhenong 6, ZN6) and a low-sucrose content and small-grain variety (Williams 82, W82), were used for transcriptome analysis in this study. Soybean plants were cultivated at the farmland of Zhejiang Academy of Agricultural Sciences (29°27′N, 120°23′E) in the summers of 2020 and 2021. All plants were subjected to the same field management during the entire growth period. According to previous research [26, 27], about forty seed samples from three plants of two genotype (ZN6 and W82) were collected for RNA sequencing and grain composition measurement with three biological replicates at R5, R6, and R7 stages, respectively. Stage R5 (Beginning seed) is seed filling phase; total pod weight peaks at stage R6 (Full seed); Seeds and pods begin to lose green color at stage R7 (Beginning maturity).

Determination of soluble sugar content, oil content, and protein content

Three plants of each genotype as biological replicates were used for measurement of soluble sugar content, oil content, and protein content. Total soluble sugar content was measured using the Plant Soluble Sugar Content Assay Kit (Solarbio, China). The operation methods are strictly in accordance with the instructions. Soluble sugar content was expressed on a fresh weight basis as mg/g (FW). The oil content was quantified using the Soxhlet extraction method [51]. The protein content of seeds was determined by Kjeldahl analysis [52]. The total nitrogen content was converted to protein content by using a conversion factor (5.64 for soybean) [52].

Identification of differentially expressed genes

For RNA extraction, about forty seed samples from three plants of two genotype (ZN6 and W82) were collected with three biological replicates at R5, R6, and R7 stages, respectively. mRNA was enriched by hybridization to oligo(dT) beads. The cDNA library was constructed with

Wang *et al. BMC Plant Biology*   (2024) 24:491

Page 8 of 10

NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs, USA), and Sequencing was performed on a NovaSeq6000 platform (Illumina) by Gene Denovo Biotechnology Co. (China).

Reads from RNA sequencing were mapped to the soybean reference genome (*Glycine max* Wm82.a4.v1). Gene abundances were quantified using RSEM software. Differentially expressed genes (DEGs) were identified by DESeq2 (version 1.12.4). To agglomerate gene patterns along seeds development stages, Short Time-course Expression Miner (STEM) algorithm was used [53]. All the DEGs were clustered into eight expression profiles.

Functional annotation and KEGG classification

All expressed genes were compared against various databases for functional annotation, including Clusters of Orthologous Groups of proteins database (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), NCBI nonredundant protein database (Nr), and Swiss-Prot database, by BLASTX searches with an e-value cutoff of 1e-5 in Blast2GO. Only the annotation with the highest score was used for each protein sequence. For each KEGG pathway, the numbers of up- and down-regulated genes of each genotype were compared to the reference set by Fisher's exact test to find out the pathways enriched with up and down-regulated genes. KEGG enrichment analysis was also carried out for all the eight gene expression profiles.

Gene expression validation

Eight genes from RNA sequencing data were randomly selected for validation by quantitative real-time PCR (qRT-PCR). Seed samples (ZN6 and W82) were collected with three biological replicates at R5, R6, and R7 stage, respectively. Total RNA was extracted from three biological replicates using an RNA isolation kit (Thermo Fisher Scientific, USA). First strand cDNA was prepared using TransScript® One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen, China). Gene-specific primers for qPCR were designed using Primer3 Web site (Table S1). The soybean *GmActin* gene (*Glyma.18G290800*) was used as the internal reference gene (Table S1). RT-PCR was performed with ChamQ SYBR qPCR Master Mix (Vazyme, China) and reactions were run on CFX96 Touch Real-Time PCR Detection System (Bio-Rad, USA). Three technical replicates were performed. Linear regression analysis was performed to evaluate the correlation of fold change data between qRT-PCR and RNA sequencing using R package (version 3.1.3).

Statistical analysis

The RNA-seq analysis was described as above, and Student's t-tests and Tukey test were performed using Excel and SPSS, respectively.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05214-1.

Supplementary Material 1

Supplementary Material 2

Author contributions
B.W. and Y. G. conceived and designed the study; B.W. performed most of the experiments; G. Z., Z. F., Y. B., and N. L. contributed to data analyses. B.W. wrote the manuscript. Y. G. discussed and revised the manuscript.

Data availability
The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA010668).

Declarations

Ethics approval and consent to participate
Our research did not involve any human or animal subjects, material, or data. The plant material ZN6, W82, ZNQF, and XNQY used in this study comes from Zhejiang Academy of Agricultural Sciences. The plant material TL used in this study comes from Chinese Academy of Agricultural Sciences. The plant material HC6 used in this study comes from South China Agricultural University. The authors declare that experimental research and field studies on plants (either cultivated or wild), including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Lee JY, Popp MP, Wolfe EJ, Nayga RM Jr., Popp JS, Chen P, Seo HS. Information and order of information effects on consumers' acceptance and valuation for genetically modified edamame soybean. PLoS ONE. 2018;13(10):e0206300.
2. Djanta MKA, Agoyi EE, Agbahoungba S, Quenum FJ-B, Chadare FJ, Assogbadjo AE, Agbangla C, Sinsin B. Vegetable soybean, edamame: Research,

production, utilization and analysis of its adoption in Sub-saharan Africa. J Hortic Forestry. 2020;12(1):1–12.

3.   Shurtleff W, Aoyagi A. History of Edamame, Vegetable soybeans, and vegetable-type soybeans (1000 BCE to 2021): extensively annotated bibliography. and Sourcebook: Soyinfo Center; 2021.

4.   Dhakal K, Zhu Q, Zhang B, Li M, Li S. Analysis of shoot architecture traits in edamame reveals potential strategies to improve harvest efficiency. Front Plant Sci. 2021;12:614926.

5.   Zhang Q, Li Y, Chin KL, Qi Y. Vegetable soybean: seed composition and production research. Italian J Agron. 2017;12(3).

6.   Wang Z, Yu D, Morota G, Dhakal K, Singer W, Lord N, Huang H, Chen P, Mozzoni L, Li S. Genome-wide association analysis of sucrose and alanine contents in edamame beans. Front Plant Sci. 2023;13:1086007.

7.   Krober OA, Cartter JL. Quantitative interrelations of protein and nonprotein constituents of soybeans 1. Crop Sci. 1962;2(2):171–2.

8.   Hymowitz T, Collins F. Variability of sugar content in seed of *Glycine max* (L.) Merrill and *G. Soja* Sieb. And Zucc. 1. Agron J. 1974;66(2):239–40.

9.   Li Y-S, Du M, Zhang Q-Y, Wang G-H, Hashemi M, Liu X-B. Greater differences exist in seed protein, oil, total soluble sugar and sucrose content of vegetable soybean genotypes [*Glycine max* (L.) Merrill] in Northeast China. Aust J Crop Sci. 2012;6(12):1681–6.

10. Zeng A, Chen P, Shi A, Wang D, Zhang B, Orazaly M, Florez-Palacios L, Brye K, Song Q, Cregan P. Identification of quantitative trait loci for sucrose content in soybean seed. Crop Sci. 2014;54(2):554–64.

11. Crittenden RG, Playne MJ. Production, properties and applications of food-grade oligosaccharides. Trends Food Sci Technol. 1996;7(11):353–61.

12. Rivero-Urgëll M, Santamaría-Orleans A. Oligosaccharides: application in infant food. Early Hum Dev. 2001;65(Suppl):S43–52.

13. Zhan S, Ho SC. Meta-analysis of the effects of soy protein containing isofla-vones on the lipid profile. Am J Clin Nutr. 2005;81(2):397–408.

14. Saldivar X, Wang Y-J, Chen P, Hou A. Changes in chemical composition during soybean seed development. Food Chem. 2011;124(4):1369–75.

15. Yu X, Yuan F, Fu X, Zhu D. Profiling and relationship of water-soluble sugar and protein compositions in soybean seeds. Food Chem. 2016;196:776–82.

16. Hou AC, Li PAJ, Mozzoni D, Zhang L, Shi B. A. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Sci. 2009;49(3):903–12.

17. Shangguan L, Song C, Leng X, Kayesh E, Sun X, Fang J. Mining and comparison of the genes encoding the key enzymes involved in sugar biosynthesis in apple, grape, and sweet orange. Entia Horticulturae. 2014;165:311–8.

18. Lin H, Rao J, Shi J, Hu C, Cheng F, Wilson ZA, Zhang D, Quan S. Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. J Integr Plant Biol. 2014;56(9):826–36.

19. Chen Z, Zhong W, Zhou Y, Ji P, Wan Y, Shi S, Yang Z, Gong Y, Mu F, Chen S. Integrative analysis of metabolome and transcriptome reveals the improvements of seed quality in vegetable soybean (*Glycine max* (L.) Merr). Phytochemistry. 2022;200:113216.

20. Kumar R, Saini M, Taku M, Debbarma P, Mahto RK, Ramlal A, Sharma D, Rajendran A, Pandey R, Gaikwad K. Identification of quantitative trait loci (QTLs) and candidate genes for seed shape and 100-seed weight in soybean [*Glycine max* (L.) Merr]. Front Plant Sci. 2023;13:1074245.

21. Li J, Zhang Y, Ma R, Huang W, Hou J, Fang C, Wang L, Yuan Z, Sun Q, Dong X, et al. Identification of ST1 reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. Plant Biotechnol J. 2022;20(6):1110–21.

22. Duan Z, Zhang M, Zhang Z, Liang S, Fan L, Yang X, Yuan Y, Pan Y, Zhou G, Liu S, et al. Natural allelic variation of *GmST05* controlling seed size and quality in soybean. Plant Biotechnol J. 2022;20(9):1807–18.

23. Lu X, Xiong Q, Cheng T, Li Q-T, Liu X-L, Bi Y-D, Li W, Zhang W-K, Ma B, Lai Y-C, et al. A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. Mol Plant. 2017;10(5):670–84.

24. Liu N, Niu Y, Zhang G, Feng Z, Bo Y, Lian J, Wang B, Gong Y. Genome sequencing and population resequencing provide insights into the genetic basis of domestication and diversity of vegetable soybean. Hortic Res. 2022;9:uhab052.

25. Valliyodan B, Cannon SB, Bayer PE, Shu S, Brown AV, Ren L, Jenkins J, Chung CYL, Chan TF, Daum CG. Construction and comparison of three reference-quality genome assemblies for soybean. Plant J. 2019;100(5):1066–82.

26. Purcell LC, Salmeron M, Ashlock L. Soybean growth and development. Ark Soybean Prod Handb. 2014;197:1–8.

27. Pedersen P, Kumudini S, Board J, Conley S: Soybean growth and development: Iowa State University, University Extension Ames, IA; 2004.28.

28. Zhou Q, Guo JJ, He CT, Shen C, Huang YY, Chen JX, Guo JH, Yuan JG, Yang ZY. Comparative transcriptome analysis between low- and high-cadmium-accumulating genotypes of *Pakchoi* (*Brassica chinensis* L.) in response to cadmium stress. Environ Sci Technol. 2016;50(12):6485–94.

29. Fang R, Chen X, Zhang X, Wang B. Comparative transcriptome analyses shed light on the regulation of harvest duration and shelf life in sweet corn. Food Res Int. 2023;163:112188.

30. Hayashi M, Crofts N, Oitome NF, Fujita N. Analyses of starch biosynthetic protein complexes and starch properties from developing mutant rice seeds with minimal starch synthase activities. BMC Plant Biol. 2018;18(1):59.

31. Zheng Z, Xu X, Crosley RA, Greenwalt SA, Sun Y, Blakeslee B, Wang L, Ni W, Sopko MS, Yao C. The protein kinase SnRK2. 6 mediates the regulation of sucrose metabolism and plant growth in Arabidopsis. Plant Physiol. 2010;153(1):99–113.

32. Zhang Q, Liu M, Ruan J. Integrated transcriptome and metabolic analyses reveals novel insights into free amino acid metabolism in Huangjinya tea cultivar. Front Plant Sci. 2017;8:291.

33. Zhang J, Zhang W, Xing L. Effects of ultrasound on the taste components from aqueous extract of unsmoked bacon. Food Chem. 2021;365:130411.

34. Harrigan GG, Skogerson K, MacIsaac S, Bickel A, Perez T, Li X. Application of 1H NMR profiling to assess seed metabolomic diversity. A case study on a soybean era population. J Agric Food Chem. 2015;63(18):4690–7.

35. Pan WJ, Xue HA, Huang SY, Yu JY, Ying ZH, Qu KX, Zhang ZX, Yin ZG, Qi HD, Yu GL, Zhang Y. Identification of candidate genes related to soluble sugar contents in soybean seeds using multiple genetic analyses. J Integr Agr. 2022;21(7):1886–902.

36. Li N, Xu R, Li Y. Molecular networks of seed size control in plants. Annu Rev Plant Biol. 2019;70:435–63.

37. Fang N, Xu R, Huang L, Zhang B, Duan P, Li N, Luo Y, Li Y. SMALL GRAIN 11 controls grain size, grain number and grain yield in rice. Rice. 2016;9:1–11.

38. Hong Z, Ueguchi-Tanaka M, Fujioka S, Takatsuto S, Yoshida S, Hasegawa Y, Ashikari M, Kitano H, Matsuoka M. The rice *brassinosteroid-deficient dwarf2* mutant, defective in the rice homolog of Arabidopsis *DIMINUTO/DWARF1*, is rescued by the endogenously accumulated alternative bioactive brassino-steroid, dolichosterone. Plant Cell. 2005;17(8):2243–54.

39. Jiang WB, Huang HY, Hu YW, Zhu SW, Wang ZY, Lin WH. Brassinosteroid regulates seed size and shape in Arabidopsis. Plant Physiol. 2013;162(4):1965–77.

40. Morinaka Y, Sakamoto T, Inukai Y, Agetsuma M, Kitano H, Ashikari M, Matsuoka M. Morphological alteration caused by brassinosteroid insensitivity increases the biomass and grain production of rice. Plant Physiol. 2006;141(3):924–31.

41. Tanabe S, Ashikari M, Fujioka S, Takatsuto S, Yoshida S, Yano M, Yoshimura A, Kitano H, Matsuoka M, Fujisawa Y. A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, *dwarf11*, with reduced seed length. Plant Cell. 2005;17(3):776–90.

42. Yuan H, Fan S, Huang J, Zhan S, Wang S, Gao P, Chen W, Tu B, Ma B, Wang Y. 08SG2/OsBAK1 regulates grain size and number, and functions differently in Indica and Japonica backgrounds in rice. Rice. 2017;10(1):1–12.

43. Aya K, Hobo T, Sato-Izawa K, Ueguchi-Tanaka M, Kitano H, Matsuoka M. A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. Plant Cell Physiol. 2014;55(5):897–912.

44. Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C. Auxin production in the endosperm drives seed coat development in Arabidopsis. Elife. 2016;5:e20542.

45. Hirano K, Yoshida H, Aya K, Kawamura M, Hayashi M, Hobo T, Sato-Izawa K, Kitano H, Ueguchi-Tanaka M, Matsuoka M. SMALL ORGAN SIZE 1 and SMALL ORGAN SIZE 2/DWARF AND LOW-TILLERING form a complex to integrate auxin and brassinosteroid signaling in rice. Mol Plant. 2017;10(4):590–604.

46. Hughes R, Spielman M, Schruff MC, Larson TR, Graham IA, Scott RJ. Yield assessment of integument-led seed growth following targeted repair of auxin response factor 2. Plant Biotechnol J. 2008;6(8):758–69.

47. Liu L, Tong H, Xiao Y, Che R, Xu F, Hu B, Liang C, Chu J, Li J, Chu C. Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice. Proc Natl Acad Sci USA. 2015;112(35):11102–7.

48. Noguero M, Le Signor C, Vernoud V, Bandyopadhyay K, Sanchez M, Fu C, Torres-Jerez I, Wen J, Mysore KS, Gallardo K. DASH transcription factor impacts *Medicago truncatula* seed size by its action on embryo morphogenesis and auxin homeostasis. Plant J. 2015;81(3):453–66.

49. Okushima Y, Mitina I, Quach HL, Theologis A. AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. Plant J. 2005;43(1):29–46.

50. Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ. The *AUXIN RESPONSE FACTOR 2* gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. Development. 2006;133(2):251–61.

51. Li Y, Du M, Zhang Q, Wang G, Jin J, Herbert S, Liu X. Planting date influences fresh pod yield and seed chemical compositions of vegetable soybean. HortScience. 2014;49(11):1376–80.

52. Beljkaš B, Matić J, Milovanović I, Jovanov P, Mišan A, Šarić L. Rapid method for determination of protein content in cereals and oilseeds: validation, measurement uncertainty and comparison with the Kjeldahl method. Accred Qual Assur. 2010;15:555–61.

53. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics. 2006;7(1):1–11.

Publisher's Note