HOSTED BY

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

The Crop Journal

Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development

Yonglong Yu[a,1], Dong Zhu[a,1], Chaoying Ma[a,1], Hui Cao[a,1], Yaping Wang[a], Yanhao Xu[b], Wenying Zhang[b], Yueming Yan[a,b,*]

[a]College of Life Sciences, Capital Normal University, Beijing 100048, China
[b]Hubei Collaborative Innovation Center for Grain Industry/Yangtze University, Jingzhou 434025, China

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form
8 January 2016
Accepted 2 February 2016
Available online 13 February 2016

Keywords:
Wheat
Transcriptome microarray
Differentially expressed genes
Grain development

ABSTRACT

Wheat seed development is an important physiological process of seed maturation and directly affects wheat yield and quality. In this study, we performed dynamic transcriptome microarray analysis of an elite Chinese bread wheat cultivar (Jimai 20) during grain development using the GeneChip Wheat Genome Array. Grain morphology and scanning electron microscope observations showed that the period of 11–15 days post-anthesis (DPA) was a key stage for the synthesis and accumulation of seed starch. Genome-wide transcriptional profiling and significance analysis of microarrays revealed that the period from 11 to 15 DPA was more important than the 15–20 DPA stage for the synthesis and accumulation of nutritive reserves. Series test of cluster analysis of differential genes revealed five statistically significant gene expression profiles. Gene ontology annotation and enrichment analysis gave further information about differentially expressed genes, and MapMan analysis revealed expression changes within functional groups during seed development. Metabolic pathway network analysis showed that major and minor metabolic pathways regulate one another to ensure regular seed development and nutritive reserve accumulation. We performed gene co-expression network analysis to identify genes that play vital roles in seed development and identified several key genes involved in important metabolic pathways. The transcriptional expression of eight key genes involved in starch and protein synthesis and stress defense was further validated by qRT-PCR. Our results provide new insight into the molecular mechanisms of wheat seed development and the determinants of yield and quality.

1. Introduction

Wheat (*Triticum aestivum* L., 2*n* = 6*x* = 42, AABBDD), an allohexaploid species, is the principal food crop used for humans and livestock globally. Wheat is counted among the "big three" cereal crops and is unrivaled in its range of cultivation owing to its extensive agronomic adaptability, high yield potential, and nutritional profile (including

* *Corresponding author*. Tel./fax: +86 10 68902777.
E-mail address: yanym@cnu.edu.cn
[1] These authors contributed equally to this work.
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

http://dx.doi.org/10.1016/j.cj.2016.01.006

essential amino acids, minerals, vitamins, beneficial phytochemicals, and dietary fiber) [1,2].

Cereal seeds consist of two main tissue fractions: starchy endosperm/aleurone and embryo/scutellum. The starchy endosperm of mature wheat seeds is a major source of nutrition, as well as the primary site for the storage of starch and proteins important for grain yield and flour quality [3,4]. The aleurone is a single cell thick and has two biological roles: (1) digesting the starchy endosperm to release nutrients and amino acids to the germinating embryo and (2) protecting the endosperm from stress and pathogens [5]. In addition, the wheat seed aleurone layer is the most concentrated source of vitamins and minerals and is also rich in proteins and lipids [6].

Starch and protein are the principal storage reserves in the wheat seed. Wheat seed development includes five main phases: fertilization (0 days post-anthesis [DPA]), "coenocytic" endosperm (1–5 DPA), cellularization, and early grain-filling (6–13 DPA), maximum grain filling (14–24 DPA), and desiccation (25–38 DPA). In general, the grain-filling period is considered to comprise cellularization and the early and maximum grain-filling phases [7].

Seed development is a very important stage in the cereal crop seed life cycle. The nutritive reserves of mature wheat seeds provide not only human food and livestock feed but also the energy for seed germination. Initial primary studies on seed development focused mainly on seed physiology and biochemistry [8], providing us with a basic understanding of the seed development process. In recent years, proteomics has been used to study the biochemical mechanisms of plant seed development, including barley [9], *Cunninghamia lanceolata* [10], *Medicago truncatula* [11,12], *Lotus japonicus* [13] and wheat [14]. However, the number of proteins identified by the proteomics approach is limited and does not permit a complete genome-wide comparison. Dry mature seeds of crops contain a vast number of mRNA species, which were first identified in cotton [15]. Since the 1990s, stored RNA in the mature dry seeds of plant species has been shown to be universal [16–18], and gene expression patterns can be detected in stored seed mRNA. Affymetrix arrays can provide a comprehensive, real-time picture of changes at the whole-transcriptome level during seed development, and this tool has been used to investigate the biological processes of wheat grain development [6,19,20]. Although transcriptome analysis during wheat grain development by RNA-Seq has been performed [21], studies of the comprehensive dynamic transcriptional characterization of grain filling stages are still limited.

Modern allohexaploid wheat has a huge and complex genome (up to 17,000 Mb) composed of the ancestral A, B, and D genomes. Recently, there has been much progress in wheat genome sequencing [22]. The A^u genome in *Triticum urartu* and D^t genome in *Aegilops tauschii* are the progenitors of the A and B genomes, respectively, of hexaploid wheat. Studies of the A^u and D^t genomes have recently been completed [23,24] and will facilitate further proteomics and transcriptomics research on wheat seed development. In the present study, we used an elite Chinese bread wheat cultivar (Jimai 20) with high yield, wide adaptability, and superior quality [25] to perform a dynamic transcriptome microarray analysis during grain-filling stages using the GeneChip Wheat Genome Array (Affymetrix, Santa Clara, CA, USA). We identified key differentially expressed genes involved in grain development. Our results shed new light on the molecular mechanisms of the accumulation of nutritive reserves in wheat, as well as the determinants of yield and quality.

2. Materials and methods

2.1. Plant material and field experiment

The wheat cultivar Jimai 20 was planted at the experimental station of Chinese Agricultural University, Beijing, in the 2013 to 2014 growing season. Experiments were performed in three biological replicates (each plot with 50 m^2). Cultivation and management followed local field production conditions. Grain samples were harvested at 11, 15, and 20 days post-anthesis (DPA). The collected samples were immediately placed in liquid nitrogen and stored at −80 °C until use.

2.2. Grain ultrastructure observation by scanning electron microscope (SEM)

Grain samples harvested at 11, 15, and 20 DPA were placed in fixative (5 mL 38% formalin, 5 mL glacial acetic, 90 mL 70% ethyl alcohol) for more than 12 h. Sequentially, the samples were dehydrated in 70% ethanol solution (20 min), 80% ethanol solution (20 min), 90% ethanol solution (overnight), and 100% ethanol solution (20 min). The samples were then treated stepwise for 20 min each in mixtures of ethanol and isoamyl acetate with ratios 3:1, 1:1, and 1:3 before soaking in isoamyl acetate. Finally, critical-point drying was performed for SEM observation. Observation of grain endosperm ultrastructures was performed with a SEM S-4800 FESEM instrument (Hitachi, Japan).

2.3. RNA isolation, microarray hybridization, and data treatments

Total RNAs isolation and microarray hybridization were performed as described previously [2]. All microarray data from the three biological replicates obtained in this study were deposited in the NCBI GEO database and are accessible under GEO Series accession number GSE75561 (http://www.ncbi.nlm.nih.gov/geo/query/acc. cgi?acc=GSE75561).

2.4. Filter of multi-class differential genes

As the random variance model (RVM) *F*-test can raise degrees of freedom effectively in the cases of small samples, it was applied to filter the differentially expressed genes for the control and experiment group. The differentially expressed genes were selected based on *P*-value threshold after significance analysis and false discovery rate (FDR) analysis [26–28].

2.5. Series test of cluster (STC) and gene ontology (GO) annotation analysis of differential genes

According to the random variance model (RVM) corrective ANOVA, the series test of cluster (STC) was performed by selecting differential expression genes. In accordance with the