Plant Science 248 (2016) 116–127

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Gene expression profiling during seed-filling process in peanut with emphasis on oil biosynthesis networks

Kapil Gupta[a], Galya Kayam[a], Adi Faigenboim-Doron[a], Josh Clevenger[b], Peggy Ozias-Akins[b], Ran Hovav[a,*]

[a] Department of Field Crops, Plant Sciences Institute, ARO, Bet-Dagan, Israel
[b] Department of Horticulture and Institute of Plant Breeding, Genetics & Genomics, The University of Georgia, Tifton, GA 31793, USA

ARTICLE INFO

Article history:
Received 23 February 2016
Received in revised form 21 April 2016
Accepted 26 April 2016
Available online 28 April 2016

Keywords:
Peanut
Oil biosynthesis pathway
RNA-Seq
Transcriptome
Seed filling

ABSTRACT

Pod-filling is an important stage of peanut (*Arachis hypogaea*) seed development. It is partially controlled by genetic factors, as cultivars considerably vary in pod-filling potential. Here, a study was done to detect changes in mRNA levels that accompany pod-filling processes. Four seed developmental stages were sampled from two peanut genotypes differing in their oil content and pod-filling potential. Transcriptome data were generated by RNA-Seq and explored with respect to genic and subgenomic patterns of expression. Very dynamic transcriptomic changes occurred during seed development in both genotypes. Yet, general higher expression rates of transcripts and an enrichment in processes involved "energy generation" and "primary metabolites" were observed in the genotype with the better pod-filling ("Hanoch"). A dataset of 584 oil-related genes was assembled and analyzed, resulting in several lipid metabolic processes highly expressed in Hanoch, including oil storage and FA synthesis/elongation. Homoeolog-specific gene expression analysis revealed that both subgenomes contribute to the oil genes expression. Yet, biases were observed in particular parts of the pathway with possible biological meaning, presumably explaining the genotypic variation in oil biosynthesis and pod-filling. This study provides baseline information and a resource that may be used to understand development and oil biosynthesis in the peanut seeds.

1. Introduction

Higher plants accumulate proteins, oils and carbohydrates in their seeds during a relatively prolonged phase, termed seed-filling (SF). SF is the period during which rapid metabolic and morphological (size, weight, and color) changes occur, encompassing cellular processes that include cell expansion and the early stage of desiccation [1–3]. In addition to their central role in the plant life cycle, seeds are major food sources. Thus, the quantity of oils, proteins, and carbohydrates accumulated during the SF period is an important yield component. In legumes (Fabaceae), seeds are developed in pods; simple fruit structures that develop from a carpel. At the beginning of pod development the shell wall or pericarp occupies the majority of the fruit volume [4]. This maternal, vascularized tissue serves as an intermediate source of nutrients that are transferred from the primary leaves source to the seed sink (embryo). Consequently, during the later stages of maturation, seeds accumulate nutrients and expand up to the point that in some legumes the seeds constitute the majority of total pod volume. This process is also known as pod-filling (PF) in legumes.

Peanut is a major source of protein and oil, ranked as the second-most important grain legume cultivated and the fourth largest edible oilseed crop in the world [5]. Peanut comprises a sustainable portion of the economy in many countries and plays an important role in global trade. As in other legume crops, PF is also an important factor in peanut influencing production both quantitatively and qualitatively. In the US peanut industry, for instance, PF (measured by "seed percentage" or the seed weight ratio from the total yield) is one of the components that is sampled and graded by the Federal-State Inspection Service to determine the overall quality and value. Inadequate PF can lead to general yield loss and low grades. In the "in-shell" peanut industries, inadequate PF also has another qualitative aspect. Incomplete PF of in-shell peanut causes severe burns during roasting, leading to sharp decreases in product price [6].

Studies have shown that external stress conditions during SF, particularly water deficit and high temperatures, are major fac-

Abbreviations: SF, seed-filling; PF, pod-filling; DE, differential expression; GO, gene ontology; RPKM, Reads per Kilobase per Million mapped reads; FA, fatty acids.

* Corresponding author.
E-mail address: ranh@agri.gov.il (R. Hovav).

http://dx.doi.org/10.1016/j.plantsci.2016.04.014
0168-9452/

tor decreasing seed size and PF in legumes [7–10]. Yet, PF in legumes also has a strong genetic component, whereas differences between genotypes are observed, as for example in soybean [8], *Lupinus* [4,11] and pea [12]. In peanuts, PF potential has a strong genetic component as well, as suggested by the significant difference between cultivars. For example, some cultivars suffered from incomplete PF when grown under insufficient irrigation in sandy soils [6]. Another example of a line with low PF potential is the "Peruvian-like" line (*A. hypogaea* ssp. *fastigiata* var. *peruviana*; PI338338). Although this accession constitutes a promising potential genetic source for breeding purposes, particularly for tolerance to pod wart [13], it was found to be characterized by incomplete PF, especially when grown under suboptimal conditions [14].

In spite of the importance of the PF and SF process, the genetic components that control these processes have not been studied in detail in peanut. In particular, it is not known why some genotypes exhibit higher PF and SF potential. Also, since peanut seeds have a high oil content compared to other legume crops [15], oil accumulation may play an important role in SF related processes. Yet, relative to other oilseeds there are fewer studies of oil-related biosynthetic and regulatory pathways in peanut. Global transcriptome analysis of developing seeds can provide fundamental molecular understanding of SF processes in peanut seeds, as well as more specific processes such as oil metabolism. This has been illustrated recently for many oil-seed crops like soybean [16,17], rapeseed (*Brassica napus*) [18,19], oil palm [20], corn oil [21], Siberian apricot (*Prunus sibirica*) [22] and cotton [23]. Generally, these allow remarkable progress in our understanding of the molecular basis for energy flow during SF in oilseed plants. Global expression analyses were also reported for peanut by using de-novo assembly techniques of the transcriptome of the developing seeds [24,25]. The complexity of the peanut seed transcriptome was exemplified in these studies and possible metabolic pathways involved in the accumulation of oil during seed development were suggested, serving as an initial basis to understand the mechanisms of oil accumulation in this crop.

To extend this information over a broader developmental timeframe and genetic variation, here we present the global analysis of gene expression during the seed development program in peanut with an emphasis on lipid and oil related pathways. Utilizing gene annotation data from the recent UGA tetraploid transcript assembly (NCBI accession GDKN00000000.1), which was constructed using a guided assembly approach with newly sequenced peanut diploid genomes (www.peanutbase.org), we conducted transcriptome analysis of seed during development and SF. The study included a series of four, well defined developmental stages from two peanut genotypes that vary both in their oil content and SF potential. Because peanut is an allopolyploid ("AB-genome") derived from hybridization between two diploid genomes ("A", "B") [26] and due to the exclusivity of UGA tetraploid-transcriptome assembly, in which both subgenomes of tetraploid peanut are distinguished and separately represented, we were able to partition and separately characterize total duplicate gene expression for each gene pair into the individual contribution of each homoeologous (duplicated) gene copy (i.e., A and B). This has provided the possibility to explore transcriptome data with respect to genic and subgenomic patterns of expression, globally and with respect to oil pathways.

2. Materials and methods

2.1. Plant materials and tissue collection

Two peanut genotypes, Hanoch (*A. hypogaea* ssp. *hypogaea* var. *hypogaea*) and line 53 (*A. hypogaea* ssp. *fastigiata*), were used for the study. 'Hanoch' is a Virginia-type cultivar, has a spreading growth habit, is late maturing and has high "dead-end" ratio (aborted seeds in the distal part of the pod). Line 53 is the local name for PI338338, a Peruvian-type peanut with high tolerance to pod wart, relatively low oil content, and low PF potential [14]. Plants were grown under field conditions in a randomized block experimental design with three blocks as described [14]. The experiment was performed as part of an irrigation test experiment in a wheat-fallow field at the Israel Ministry of Agriculture Southern R & D Center, Negev, Israel. All 6 plots (2 genotypes × 3 blocks) were uprooted and harvested at 110 days post sowing and seeds from four specific stages of pod development [27] were harvested: R4-pods with tiny embryos, R5-beginning seed growth, R6-expanded but immature seeds and R7-expanded and mature seeds. The seeds were flash frozen in liquid nitrogen and stored at −80 °C till RNA extraction.

2.2. Oil content and fatty acid composition

Oil content was estimated in Hanoch and 53 during developmental stages R4 to R7. Oil content and composition was analyzed using 5 g of seeds that were collected from each developmental stage. Seeds were ground in an Abencor laboratory mill (Mc2 Ingenieria Y Sistemas, Seville, Spain) and dried for 24 h. The total oil content was then determined by a Soxhlet extraction using *n*-hexane. From each sample, 0.06 g oil was diluted with 0.3 ml chloromethane and 2 ml sodium methoxide, heated to 50 °C for 30 min and cooled to room temperature. 5 ml 2% acetic acid and 2 ml internal standard solution (C17-methylhepatadecanoate) were added. Following centrifugation (5000 rpm, 5 min) the upper phase was recovered using a glass syringe and then filtered through a 0.45 μm PTFE filter. The extract was dried at 37 °C under a slow stream of N_2 gas. The GC analyses were carried out on an Agilent 6890 GC–FID apparatus equipped with an SP™-2560 ('Supelco') (100 m × 0.25 mm I.D., 0.20 μm film thickness) bicyanopropyl polysiloxane fused-silica capillary column. The injector was kept at 250 °C and the detector at 260 °C. Helium at a constant pressure of 42.7 p.s.i. was used as a carrier gas with retention time locking. The column was maintained at 50 °C for 2 min and then programmed to 190 °C at 5 min and kept for 20 min at 190 °C. Identification of compounds was performed by comparing their relative retention indices with those of authentic samples.

2.3. Total RNA isolation, library preparation and high throughput sequencing

Samples of 400 mg were taken from each of the ground tissues that were used for oil extraction. These were used for RNA extraction using Hot Borate method as described by Brand and Hovav [28]. The total RNA was used for preparation of RNA-Seq libraries in three biological replicates using TruSeq RNA Sample Preparation Kit v2 (Illumina) following the Manufacturer's protocol. The integrity and quality of the total RNA, was checked on agarose gels. For construction of the RNA-Seq libraries, 4000 ng of the total RNA was used from each sample. mRNA was purified from total RNA using Agencourt Ampure-XP magnetic beads (Beckman Coulter). Following purification, mRNA was fragmented in the average size of 300 bp. The first-strand cDNA was generated using random hexamer-primed reverse transcription followed by second strand cDNA synthesis, end repaired and ligated to adaptor index, enriched by PCR for 15 cycles and purified. Libraries were validated by DNA Screen Tape D1000 using the Tapestation 2200 (Agilent Technologies). RNA-Seq libraries were sequenced by an Illumina HiSeq™ 2000.