ELSEVIER

GENOMICS
Data

Data in Brief

RNA-Seq analysis of laser-capture microdissected cells of the developing central starchy endosperm of maize

CrossMark

Dhiraj Thakare [a], Ruolin Yang [a], Joshua G. Steffen [b,1], Junpeng Zhan [a], Dongfang Wang [a,2], Richard M. Clark [b], Xiangfeng Wang [a], Ramin Yadegari [a,*]

[a] *School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA*
[b] *Department of Biology, University of Utah, Salt Lake City, UT 84112, USA*

ARTICLE INFO

ABSTRACT

Endosperm is a product of double fertilization, and provides nutrients and signals to the embryo during seed development in flowering plants. Early stages of endosperm development are critical for the development of its storage capacity through synthesis and accumulation of starch and storage proteins. Here we report on the isolation and sequencing of mRNAs from the central portion of the starchy endosperm of *Zea mays* (maize) B73 at 6 days after pollination. We detected a high level of correlation among the four biological replicates of RNAs isolated using laser-capture microdissection of the cell type. Because the assayed developmental stage precedes the synthesis and accumulation of the major storage proteins and starch in the endosperm, our dataset likely include mRNAs for genes that are involved in control and establishment of these storage programs. The mRNA-Seq data has been deposited in Gene Expression Omnibus (accession number GSE58504).

Specifications	
Organism/cell line/tissue	*Zea mays* (maize) reference genotype B73, central portion of the starchy endosperm
Sex	NA
Sequencer or array type	Illumina Genome Analyzer IIx
Data format	Normalized reads mapped to B73 RefGen_v2
Experimental factors	Biological replicates of a single endosperm cell type
Experimental features	A unique replicated dataset of laser-capture microdissected RNA from central starchy endosperm at 6 days after pollination
Consent	NA
Sample source location	Tucson, AZ

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58504

* Corresponding author.
E-mail address: yadegari@email.arizona.edu (R. Yadegari).
[1] Present address: Department of Natural Sciences, Colby-Sawyer College, New London, NH 03257, U.S.A.
[2] Present address: Department of Biology, Spelman College, Atlanta, GA 30314, U.S.A.

Experimental design, materials and methods

Sample collection and RNA analysis

Plants were grown under greenhouse conditions (16-h day) at the University of Arizona during January–June 2011, and self-pollinated to obtain 6-DAP kernels [1]. Using a protocol for laser-capture microdissection (LCM) of plant tissues [2,3] and further modified using methods that were subsequently published [4], we isolated the central portion of the starchy endosperm (Fig. 1) with a Leica LMD 6500 instrument (Leica Microsystems, Inc.). The captured region included the presumptive central starchy endosperm and a portion of the conducting zone cells [5–7]. Sections for replicates 1 through 3 (Reps. 1–3, June 2011) were obtained from nine separate kernels (three each) on a single ear while sections for the fourth replicate (Rep. 4) were obtained from a separate ear (on a separate plant, March 2011). RNA was extracted from microdissected sections using an ARCTURUS PicoPure RNA Isolation Kit (Applied Biosystems/Life Technologies, cat. no. KIT0204), and its size and integrity were evaluated using an RNA 6000 Pico Kit (Agilent Technologies, cat. no. 5067-1513) on an Agilent 2100 Bioanalyzer (Agilent Technologies) before and after DNase treatment (Fig. 1, TURBO DNase, Ambion/Life Technologies, cat. no. AM2238). About 50–60 ng of DNAase-treated RNA was used as template for cDNA synthesis in order to amplify the captured RNA (two rounds) using a T7 polymerase-
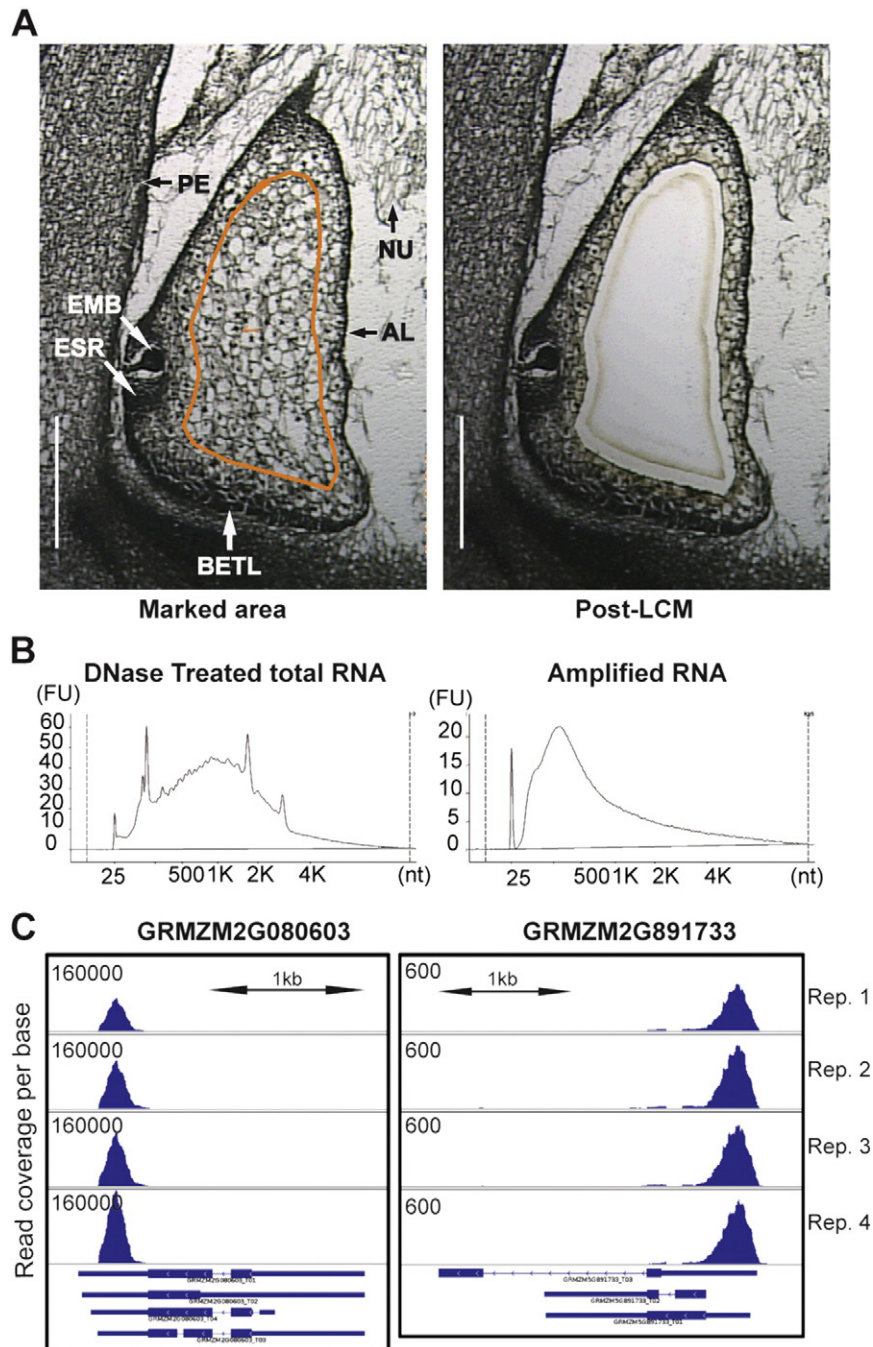
Fig. 1. Representative description of the biological material collected. (A) A representative section of a fixed 6-DAP kernel (Rep. 1) showing the central portion of endosperm marked for laser-capture microdissection. An estimated 3000–5000 cells were captured for each replicate. AL, aleurone; BETL, basal endosperm transfer layer; EMB, embryo; ESR, embryo-surrounding region; NU, nucellus; PE, pericarp. Scale bar: 400 μm. (B) Profiles of representative RNA (Rep. 1) obtained from LCM sections (left) and after the amplification of the isolated RNA (right). (C) Profiles of RNA-Seq reads for two representative genes in the maize genome showing reads mapped predominately to the 3′ regions of the genes.

based linear amplification system (Arcturus RiboAmp HS PLUS RNA Amplification Kit, Applied Biosystems/Life Technologies, cat. no. KIT0525) to produce a peak size of ~200–500-nt RNA fragments (Fig. 1).

Sequencing, mapping and normalization

Standard barcoded RNA-seq libraries were generated to facilitate sequencing of four samples using a single Illumina flowcell lane. Libraries were generated using protocols adapted from Illumina mRNA sample preparation protocol (cat. no. 1004894 Rev. A) described previously

[8]. Cluster generation and sequencing was carried out on the Illumina Cluster Station and Genome Analyzer IIx (GAIIx) instrument using Single Read Cluster Generation Kit (cat. no. 15003972) and v4 Sequencing Kit (cat. no. 15003925), respectively [8]. 82 cycles of imaging were carried out using a modification of the "GA2_76Cycle_SR_v7.xml" sequencing program. Methods for RNA-Seq data filtering and processing were essentially as those described previously [8].

Nearly 24.30 million reads were generated from the four replicates ranging from 4.33 million in Rep. 1 to 6.96 million in Rep. 2. Using Tophat [9], ~88.5–90.2% of the reads were mapped to the reference

Table 1
Mapping of reads aligned to the B73 reference genome.

Sample	Total read number	Filtered	Kept	Mapped	Unique-mapped	Multi-mapped (≤5 locations)	Multi-mapped (>5 locations)
Rep. 1	4,332,880	10,801	4,322,079	3,826,888	3,588,669	232,759	5460
Rep. 2	6,866,277	16,655	6,849,622	6,077,467	5,695,442	373,324	8701
Rep. 3	6,131,203	15,187	6,116,016	5,445,812	5,122,070	315,684	8058
Rep. 4	6,969,851	17,001	6,952,850	6,276,299	5,931,966	337,487	6846

B73 genome (release 5b.60) (Table 1). Using BEDTools [10], 31,130 of the 110,028 genes in the working gene set (WGS) were detected in at least 2 replicates with at least one raw read count. We defined these genes as expressed in our experiment. Inter-sample differences in library size were eliminated by total count (TC) normalization method. Read counts for all the genes in WGS were divided by the total counts of mapped reads (or library size) associated with their sample and multiplied by the mean total count across all the samples of the dataset to calculate the normalized read counts [11] and deposited in GEO.

Data reproducibility

Similarity of expression profiles between the replicates was determined by a Pearson correlation coefficient (PCC) analysis [12]. The $\log_2$-transformed normalized read counts for the 31,130 expressed genes were used as input. The four replicates showed high correlation with each other, with PCC scores ranging from 0.95 to 0.96 (Fig. 2). The correlations were visualized using scatterplots (Fig. 2). Together, the evidence suggests that our LCM-generated RNA-Seq data is highly reproducible.

References

[1] G. Li, et al., Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. Proc. Natl. Acad. Sci. U. S. A. 111 (21) (2014) 7582–7587.
[2] N.M. Kerk, T. Ceserani, S.L. Tausta, I.M. Sussex, T.M. Nelson, Laser capture microdissection of cells from plant tissues. Plant Physiol. 132 (1) (2003) 27–35.
[3] M. Nakazono, F. Qiu, L.A. Borsuk, P.S. Schnable, Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15 (3) (2003) 583–596.
[4] M.F. Belmonte, et al., Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc. Natl. Acad. Sci. U. S. A. 110 (5) (2013) E435–E444.
[5] P.W. Becraft, Cell fate specification in the cereal endosperm. Semin. Cell Dev. Biol. 12 (5) (2001) 387–394.
[6] W.L. Charlton, et al., Endosperm development in *Zea mays*; implication of gametic imprinting and paternal excess in regulation of transfer layer development. Development 121 (9) (1995) 3089–3097.
[7] D.C. Cooper, Caryopsis development following matings between diploid and tetraploid strains of *Zea mays*. Am. J. Bot. 38 (9) (1951) 702–708.
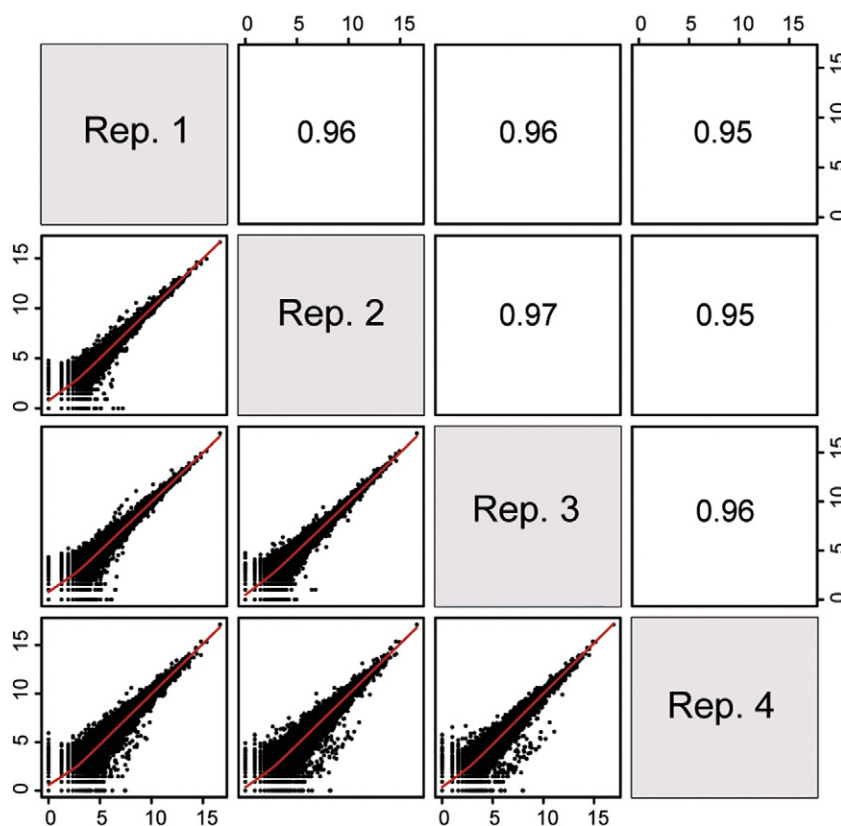

Fig. 2. Basic analysis of RNA-Seq reads. Scatterplots and Pearson correlation coefficient (PCC) analysis among the four replicates.

[8] X. Gan, et al., Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature 477 (7365) (2011) 419–423.

[9] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25 (9) (2009) 1105–1111.

[10] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26 (6) (2010) 841–842.

[11] M.A. Dillies, et al., A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. 14 (6) (2013) 671–683.

[12] K. Pearson, On Further Methods of Determining Correlation. Dulau and Company, 1907.