Data in Brief

Transcriptome profiling for discovery of genes involved in shoot apical meristem and flower development

Vikash K. Singh, Mukesh Jain *

Functional and Applied Genomics Laboratory, National Institute of Plant Genome Research, New Delhi 110067, India

ABSTRACT

Flower development is one of the major developmental processes that governs seed setting in angiosperms. However, little is known about the molecular mechanisms underlying flower development in legumes. Employing RNA-seq for various stages of flower development and few vegetative tissues in chickpea, we identified differentially expressed genes in flower tissues/stages in comparison to vegetative tissues, which are related to various biological processes and molecular functions during flower development. Here, we provide details of experimental methods, RNA-seq data (available at Gene Expression Omnibus database under GSE42679) and analysis pipeline published by Singh and colleagues in the Plant Biotechnology Journal (Singh et al., 2013), along with additional analysis for discovery of genes involved in shoot apical meristem (SAM) development. Our data provide a resource for exploring the complex molecular mechanisms underlying SAM and flower development and identification of gene targets for functional and applied genomics in legumes.

Specifications

Organism/cell line/tissue	*Cicer arietinum* L. genotype ICC 4958
Sequencer or array type	Illumina Genome Analyzer IIx
Data format	Raw data: FASTQ files, analyzed data: txt files
Experimental factors	Tissues/organs
Experimental features	RNA-seq dataset for gene expression profiling in shoot apical meristem and flower development in chickpea
Sample source location	New Delhi, India

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42679.

Materials and methods

Sample collection and RNA isolation

Chickpea (*Cicer arietinum* L. genotype ICC 4958) seeds were grown in culture room and field for collection of vegetative and reproductive tissues as described [1]. Three vegetative tissues [germinating seedling (GS), young leaves (YL) and shoot apical meristem (SAM)], and four stages each of flower bud (FB1–FB4) and flower (FL1–FL4) were harvested in at least three biological replicates (Fig. 1). Harvested tissues were snap frozen into liquid nitrogen and stored at −80 °C till RNA isolation. Total RNA was extracted from each tissue using TRI reagent (Sigma Life Science) according to manufacturer's instructions after grinding tissues in pre-chilled (in liquid nitrogen) mortar and pestle. Total RNA concentration and purity were determined using Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific) and microfludics-based RNA nano chip on 2100 Bioanalyzer (Agilent Technologies).

Generation of RNA-seq data

To understand the global expression profiles during flower development in chickpea, we performed RNA-seq analysis. Library for each sample was constructed using mRNA-Seq Sample Prep kit and single-end sequencing was performed following standard protocol of Illumina so as to generate 54 nucleotide long reads using Genome Analyzer IIx (Illumina Inc., San Diego, CA). High quality chastity-filtered reads were obtained in FASTQ file format after filtering the raw data through standard Illumina pipeline. Whole RNA-seq data were submitted to the Gene Expression Omnibus (series accession number GSE42679). An outline of strategy adopted for sample preparation and sequencing is provided in Fig. 2a.

Data processing and differential expression quantification

The strategy used for data processing and differential gene expression analysis is represented in Fig. 2b. The single-end chastity-filtered reads obtained in FASTQ format were further subjected to stringent sequence

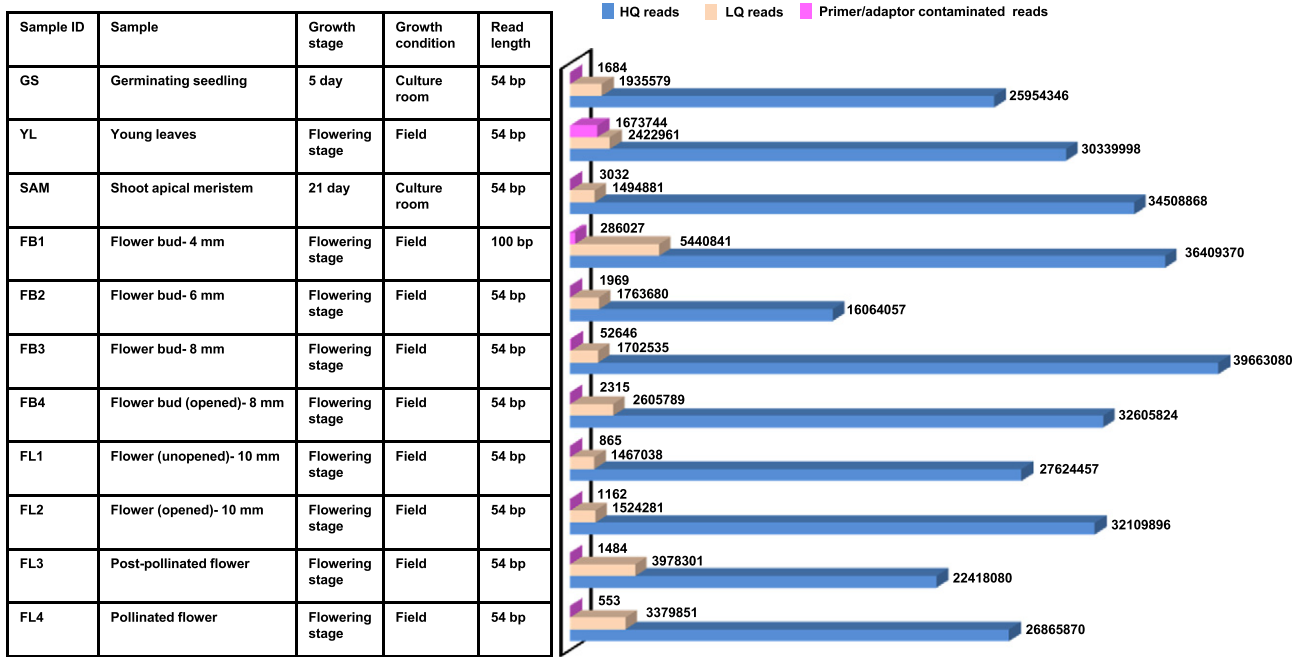| Sample ID	Sample	Growth stage	Growth condition	Read length
GS	Germinating seedling	5 day	Culture room	54 bp
YL	Young leaves	Flowering stage	Field	54 bp
SAM	Shoot apical meristem	21 day	Culture room	54 bp
FB1	Flower bud- 4 mm	Flowering stage	Field	100 bp
FB2	Flower bud- 6 mm	Flowering stage	Field	54 bp
FB3	Flower bud- 8 mm	Flowering stage	Field	54 bp
FB4	Flower bud (opened)- 8 mm	Flowering stage	Field	54 bp
FL1	Flower (unopened)- 10 mm	Flowering stage	Field	54 bp
FL2	Flower (opened)- 10 mm	Flowering stage	Field	54 bp
FL3	Post-pollinated flower	Flowering stage	Field	54 bp
FL4	Pollinated flower	Flowering stage	Field	54 bp

Fig. 1. Details of samples used for RNA-seq and summary of data after filtering low-quality (LQ) and primer/adaptor contaminated reads using NGS QC ToolKit. These data are taken from Singh et al. [5].

quality controls, using NGS QC Toolkit (v2.2.3) at default parameters [2]. This resulted into removal of 4–15% of low-quality sequences (Fig. 1). The high-quality reads were then aligned to the chickpea transcriptome (v1.0) available at Chickpea Transcriptome Database (v2.0; http://www.nipgr.res.in/ctdb.html) [3] with a tolerance of up to two mismatches using CLC Genomics Workbench (v4.7.2). The first round of alignment resulted into mapping of 66–88% of reads for individual samples. For the sequences that did not map to transcriptome, a second round of alignment was performed, in which we trimmed 14 bases from the 3′-end of unmapped reads after optimization of trimming one base recursively. This procedure enabled us to recover reads showing some misalignments caused by spurious or low-quality bases present at the read ends. After this step, total mapping percentage of reads increased to >90% for each sample. Read count obtained after mapping was normalized to 'reads per kilobase of exon model per million mapped reads (RPKM) for each sample. The genes having ≥1 RPKM at least in one sample were considered to be expressed. In total, 33,584 genes were found expressed in chickpea tissues analyzed, which represented 96.6% of the chickpea transcriptome.

Differential gene expression analysis was performed between tissues considering total reads mapped to chickpea transcripts using Bioconductor package DESeq (v1.5.24) [4]. DESeq uses a negative binomial distribution to model total read counts for biological and technical variation by a generalized Poisson distribution model. For differential gene expression analysis, data were normalized across the samples with size factors and variance. To estimate dispersion between samples without replicates, method = "blind" together with sharingMode = "fit-only" was used. Genes with P-value ≤ 0.05 and fold change ≥2 were regarded as differentially expressed. This represented a 100% linear fold change i.e. $\log_2 2 = 1$ or 100%. The expression of 1572 genes was significantly different in floral tissues as compared to vegetative tissues [5]. Among these, 1304 genes were up-regulated and 269 genes were down-regulated in floral tissues. These data have been integrated into the Chickpea Transcriptome Database (v2.0; http://www.nipgr.res.in/ctdb.html).

Differentially expressed transcripts in shoot apical meristem

Using similar strategy mentioned above, gene expression in SAM was compared with three combinations of other tissues to identify differentially expressed transcripts. In the first combination, differentially expressed genes in SAM were identified as compared to vegetative tissues (GS + YL) only, leading to identification of 882 up- and 1031 down-regulated genes (Fig. 3a). In the second combination, SAM was compared with vegetative tissues and first stage of flower bud development (FB1), which resulted into 1143 up- and 1008 down-regulated genes in SAM (Fig. 3a). Further, in the third combination, SAM was compared with all the tissues (GS + YL + FB1–4 + FL1–4) and 546 exclusively up- and 354 exclusively down-regulated genes in SAM were identified (Fig. 3a).

We performed gene ontology (GO) analysis using the BiNGO (v3.0.2) tool with P-value cut-off of ≤0.05 after applying Benjamini Hochberg correction to reveal functional categories enriched in the up-regulated genes in SAM. We found genes involved in various processes of SAM development, including G-protein coupled receptor protein signaling pathway, RNA biosynthetic process, regulation of RNA metabolic process, cell differentiation, stem cell differentiation and maintenance, leaf proximal distal/pattern formation, abscisic acid mediated signaling pathways, anatomical structure development, cell differentiation, responses to organic substances, responses to hormone stimulus, response to auxin stimulus and protein ubiquitination etc. enriched in SAM (Fig. 3b). The biological process GO terms found enriched in our study are in very good agreement with previous reports. For example, G-protein coupled receptor protein regulates cell proliferation, physiological responses of abscisic acid and gibberellins during SAM development [6–10]. Auxin signaling has been shown to promote leaf differentiation in SAM [11]. Furthermore, processes like abscisic acid mediated signaling pathways, anatomical structure development, responses to organic substances and responses to hormone stimulus have found in up-regulated genes in shoot apex of Arabidopsis by movement of WUSCHEL protein [12,13]. In addition, we found many