



# Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds

Seema Pradhan, Nitesh Bandhiwal, Niraj Shah, Chandra Kant, Rashmi Gaur and Sabhyata Bhatia\*

National Institute of Plant Genome Research, New Delhi, India

## Edited by:

Ramin Yadegari, The University of Arizona, USA

## Reviewed by:

Rajandeeep Sekhon, Clemson University, USA

Dongfang Wang, Spelman College, USA

## \*Correspondence:

Sabhyata Bhatia, National Institute of Plant Genome Research, Aruna Asaf Ali Marg, Post Box No. 10531, New Delhi 110067, India  
e-mail: sabhyatabhatia@nipgr.ac.in

Understanding developmental processes, especially in non-model crop plants, is extremely important in order to unravel unique mechanisms regulating development. Chickpea (*C. arietinum* L.) seeds are especially valued for their high carbohydrate and protein content. Therefore, in order to elucidate the mechanisms underlying seed development in chickpea, deep sequencing of transcriptomes from four developmental stages was undertaken. In this study, next generation sequencing platform was utilized to sequence the transcriptome of four distinct stages of seed development in chickpea. About 1.3 million reads were generated which were assembled into 51,099 unigenes by merging the *de novo* and reference assemblies. Functional annotation of the unigenes was carried out using the Uniprot, COG and KEGG databases. RPKM based digital expression analysis revealed specific gene activities at different stages of development which was validated using Real time PCR analysis. More than 90% of the unigenes were found to be expressed in at least one of the four seed tissues. DEGseq was used to determine differentially expressing genes which revealed that only 6.75% of the unigenes were differentially expressed at various stages. Homology based comparison revealed 17.5% of the unigenes to be putatively seed specific. Transcription factors were predicted based on HMM profiles built using TF sequences from five legume plants and analyzed for their differential expression during progression of seed development. Expression analysis of genes involved in biosynthesis of important secondary metabolites suggested that chickpea seeds can serve as a good source of antioxidants. Since transcriptomes are a valuable source of molecular markers like simple sequence repeats (SSRs), about 12,000 SSRs were mined in chickpea seed transcriptome and few of them were validated. In conclusion, this study will serve as a valuable resource for improved chickpea breeding.

**Keywords:** Transcriptome, chickpea, seed, assembly, next generation sequencing, annotation, differential expression

## INTRODUCTION

Chickpea (*Cicer arietinum* L.) is one of the earliest annual pulse crops to be cultivated by man and consumed as a source of vegetable protein and the third most important food legume in the world. India is the leading producer of chickpea where the crop is grown on 8.3 million hectares of land yielding 7.7 million tons of chickpea crop. Of these, 0.33 million tons comprises the seeds (FAOSTAT, 2012<sup>1</sup>, <http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E>). Chickpea is a self-pollinating diploid ( $2n = 16$ ) annual crop with a genome size of 738.09 Mbp (Varshney et al., 2013). The seeds are an important source of protein for millions of people in the developing countries. In addition to having high protein content (20–23%), chickpea seeds are rich in carbohydrates (41.1–47.42%), fiber (6% crude fiber), minerals (phosphorus, calcium, magnesium, iron and zinc) and  $\beta$ -carotene (Friedman, 1996; Abbo et al., 2005; Iqbal et al., 2006). Apart from all the obvious dietary attributes, improving the yield and quality of chickpeas is economically desirable keeping in mind the

large population of developing countries that are in need of an affordable source of high protein nutrition. Seeds are essential for flowering plant reproduction because they protect, nourish and contain the developing embryo that represents the next sporophytic generation. Various studies on legume seed development show that the process is genetically programmed and correlated with changes in metabolite levels (Borisjuk et al., 2004; Weber et al., 2005).

In order to explore the possibilities of improvement in the nutritional quality and quantity of the seeds, it is necessary to have a clear understanding of the regulatory mechanisms involved in the process of seed development. Such an understanding may be best achieved through an exhaustive study of the transcriptome of the developing seed which may provide a blueprint of genes and pathways involved. Conventional methods for such studies have included generation of EST libraries by cloning and sequencing on the Sanger platform and have been used for analyzing the transcriptome of developing seeds in plants such as *Arabidopsis thaliana* (White et al., 2000). But the advent of high throughput NGS technology enables us to

<sup>1</sup> Available online at: <http://faostat.fao.org>

generate considerably larger amounts of data as compared to conventional transcript analysis methods. A host of such technologies from Illumina/Solexa, ABI/SOLiD and 454/Roche have provided unprecedented opportunities for high-throughput functional genomics research (Mardis, 2008; Schuster, 2008; Morozova et al., 2009). The efficiency of the Roche/454 Genome Sequencer FLX platform for improving the coverage of transcriptome as compared to conventional methods of EST sequencing has been well documented (Cheung et al., 2006). Also, many transcriptome sequencing projects have preferred to use the 454 platform as compared to other techniques (Vera et al., 2008; Salem et al., 2010) due to longer reads (~600–800 bp) which are thought to be more amenable to *de novo* assembly and annotation. Earlier studies related to in depth analysis of seed development have been undertaken using various high throughput techniques in plants like *Arabidopsis* (Girke et al., 2000; Le et al., 2010), wheat (Gregersen et al., 2005) rice (Xue et al., 2012) and oats (Gutierrez-Gonzalez et al., 2013) and also in important legumes such as soybean (Severin et al., 2010; Jones and Vodkin, 2013) and *Medicago* (Gallardo et al., 2007).

Although there have been studies reporting transcriptome analysis in chickpea (Hiremath et al., 2011; Garg et al., 2011a,b; Singh et al., 2013; Afonso-Grunz et al., 2014; Kudapa et al., 2014), none of these have focussed specifically on chickpea seed development. However, such studies have been undertaken in the related legume soybean where seed development was extensively studied with respect to different tissues (Severin et al., 2010) as well as in seeds taking developmental stages from few days post fertilization to maturity where major landmarks in seed development, including accumulation of nutrients, synthesis of storage proteins and desiccation were analyzed using RNA-seq (Jones and Vodkin, 2013).

Seed development in dicots generally proceeds through three stages. The first stage is that of morphological development where embryo formation occurs. In the second stage, i.e., cell expansion, food and storage reserves are produced while in the third stage, i.e., maturation drying, desiccation of seed occurs (Baud et al., 2002). Earlier studies on chickpea seed development (Setia and Malik, 1985) have shown that an exponential increase in fresh seed weight occurs between 10 and 40 days after anthesis (DAA) which was also observed in our chickpea samples (inset **Table 1**). Hence, we selected the following time points: 10 DAA, to correlate with early stages (cell division and morphogenesis); 20 DAA and 30 DAA, to correlate with middle stages (cell expansion and storage protein accumulation) and 40 DAA to correlate with maturation, thereby representing the

complete spectrum of seed development from early stages to maturity. Further, we compared the transcripts obtained from each stage with those from other tissues of chickpea to find out genes specific to chickpea seed development. Digital expression analysis of the transcripts from individual stages was carried out to provide information about developmental stage specific events. Identification of the transcription factors and metabolic pathways involved in seed development was done in order to gain an insight into the regulatory networks governing seed development.

## MATERIALS AND METHODS





### PLANT MATERIALS

Chickpea (*C. arietinum*) cultivar ICCV2 was grown in the field at NIPGR. The flowers were tagged on the day they opened completely. Seeds were collected at 10, 20, 30, and 40 DAA intervals, frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . At least three biological replicates were collected for each sample.

### RNA ISOLATION AND cDNA SYNTHESIS

Total RNA was isolated from tissue samples (pooled from three biological replicates) using the LiCl precipitation method as described by Choudhary et al. (2009). Briefly, 0.8 g of frozen seed tissue was ground to a fine powder in liquid nitrogen and transferred to an Eppendorf tube containing 500  $\mu\text{l}$  of extraction buffer (200 mM NaOAc pH-5.2, 1% SDS, 10 mM EDTA pH 8.0) and 500  $\mu\text{l}$  of phenol. This was centrifuged at 14,000g for 10 min at RT. The aqueous phase was separated and extracted twice with phenol:chloroform (1:1) followed by O/N precipitation at  $4^{\circ}\text{C}$  with 0.3 volume of 10 M LiCl for RNA precipitation. The RNA pellet was recovered by centrifugation at 10,000 rpm/9700g for 10 min at  $4^{\circ}\text{C}$ , and was washed twice with 2.5 M LiCl and once with 70% ethanol. The pellet was air-dried and dissolved in DEPC-treated ddH<sub>2</sub>O. mRNA was isolated from the total RNA using PolyAtract mRNA isolation system IV (Promega Corporation, Madison, WI, USA) according to the manufacturer's protocol. Both total RNA and mRNA samples were checked for quality on the Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA) and only samples having RIN (RNA Integrity Number) greater than 8.0 were selected for cDNA preparation. cDNA was synthesized from 1.5  $\mu\text{g}$  of mRNA using Promega's Universal Riboclone cDNA synthesis system (Promega Corporation, Madison, WI, USA) using random hexameric primers and following the manufacturer's protocol.

**Table 1 | Summary of sequenced reads obtained from the four stages of developing seed tissues.**

Tissue type/ Stage of seed development	Total no. of reads	Percentage of HQ filtered reads	Maximum read length	Average read length
 10 DAA	187,778	81.19	586	341.1
 20 DAA	514,370	81.68	734	344.8
 30 DAA	214,273	80.69	571	338.4
 40 DAA	464,048	82.69	649	345.4
COMBINED	1,380,469	81.81	734	342.4

#### 454 (GSFLX TITANIUM) SEQUENCING

Approximately 600ng of double-stranded cDNA was nebulized and fragments ranging in size from 300 to 800 bp were selected. Specific adapters were ligated to the fragmented cDNA and denatured to generate single-stranded cDNA followed by emulsion PCR amplification for sequencing. Four cDNA libraries were generated, one each from the mRNA isolated from 10, 20, 30, and 40 DAA seed tissue. The quality of libraries was assessed using High Sensitivity DNA kit on Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA). All the four cDNA libraries were sequenced in one complete run using the GS FLX Titanium series sequencing reagents and sequencer. Each of the cDNA libraries was tagged with unique RL MID (multiplex identifier) adaptors and mixed before emulsion PCR and sequencing.

#### ASSEMBLY AND ANNOTATION OF 454 READS

The sequence data was obtained in Fastq format and various quality control steps were performed using the NGS QC tool kit (Patel and Jain, 2012). The sequence data obtained was deposited in the SRA database under the accession number SRX125162 (Run numbers SRR445088, SRR445089, SRR445090, and SRR445091 for 10 DAA, 20 DAA, 30 DAA, and 40 DAA seed tissue respectively). After passing through the quality controls, *de novo* assembly of filtered reads was done with Newbler v 2.3.5 software from Roche using default parameters. For the reference assembly, filtered reads were mapped onto the chickpea reference genome (Varshney et al., 2013) using gsMapper and assembled with Newbler v 2.3.5. Both assemblies were merged and clustered using TGICL (Pertea et al., 2003). To deduce their putative function, chickpea non redundant transcript data set was subjected to BLASTX analysis against the non-redundant protein database of UniProt Swiss-Prot (<http://www.uniprot.org/downloads>) and the hits with an  $E$ -value  $\leq 1E-05$  were considered to be significant. The GOSlim terms for molecular function, biological process, and cellular component categories were assigned to the corresponding chickpea transcripts using in house Perl scripts. Distribution of transcripts into various biological pathways in KEGG was done through KAAS (KEGG Automatic Annotation Server; <http://www.genome.jp/tools/kaas/>). The transcripts were aligned to the sequences in the COG database (<http://www.ncbi.nlm.nih.gov/COG>) and categorized under different functions accordingly.

#### DETERMINATION OF TRANSCRIPT ABUNDANCE AND DIFFERENTIAL GENE EXPRESSION

Sequence reads from the four tissues, i.e., 10, 20, 30, and 40 DAA seed, were mapped onto the non-redundant chickpea transcripts using GS Reference Mapper (gsMapper, <http://www.454.com/products/analysis-software/>). Number of reads mapped were normalized and measured in terms of RPKM (reads per kilobase per million) in order to determine the transcript abundance (Mortazavi et al., 2008). Unigene expression levels were calculated as  $RPKM(A) = (10,000,000 \times C \times 1,000)/(N \times L)$ , where  $A$  refers to the expression of gene  $A$ ,  $C$  to number of reads that uniquely aligned to gene  $A$ ,  $N$  to total number of reads that uniquely aligned to all genes, and  $L$  to the length of gene  $A$ . The genes with normalized RPKM  $\geq 1$  were considered to be expressed at a particular stage. The DEGseq package

was used to determine differentially expressed genes at each stage (<http://www.bioconductor.org/packages/release/bioc/html/DEGseq.html>).

#### GO TERM ENRICHMENT AND BIOLOGICAL PATHWAY ANALYSIS

GO enrichment and pathway analysis was performed as described by Singh et al. (2013). Briefly, BLAST search was used to identify the best *Arabidopsis* hit corresponding to each chickpea transcript. GO enrichment of various sets of genes was performed using BiNGO tool (Maere et al., 2005). Transcript diversity and abundance was scrutinized with MapMan (Thimm et al., 2004) using the *M. truncatula* peptide database downloaded from MapMan (<http://mapman.gabipd.org/web/guest/mapmanstore>) as reference. Accordingly, chickpea seed transcripts were first compared (Blastx,  $evalue \leq 1E-05$ ) against the *M. truncatula* peptide database downloaded from the ftp site of Phytozome v 9.1 (<ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Mtruncatula/annotation/>).

#### REAL TIME QUANTIFICATION

Gene specific primers were designed using PRIMER EXPRESS version 3.0 (PE Applied Biosystems, USA) with default parameters (Supplemental Table S2). First strand cDNA was synthesized by reverse transcription from 3  $\mu$ g of total RNA in 20  $\mu$ l of reaction volume using AccuScript High Fidelity 1st strand cDNA synthesis Kit (Agilent technologies, USA). 5X dilutions of all cDNA samples were used for Real time PCR analysis with 200 nM of each primer mixed with SYBR Green PCR master mix (Applied Biosystem, Life Technologies, USA) as per manufacturer's instructions. The reaction was carried out in 96-well optical reaction plates (Applied Biosystems, USA), using ABI Prism 7000 Sequence Detection System and software (PE Applied Biosystems, USA). To normalize the variance among samples, Elongation factor 1 $\alpha$  (Forward: 5'-TCCACCACTTGGTCGTTTTG-3', Reverse: 5'-CTTAATGACACCGACAGCAACAG-3') (Garg et al., 2010) was used as endogenous control. Relative expression values were calculated after normalizing against the maximum expression value. The values presented are the mean of the three biological replicates, each with three technical replicates. The error bars indicate standard error. Correlation between  $\log_2$  normalized values of qRT PCR and RPKM was plotted on MS Office Excel.

#### GENOMIC DNA ISOLATION AND TOUCHDOWN PCR

DNA was isolated from fresh, young leaf tissue of chickpea (ICCV2) using the CTAB method (Doyle and Doyle, 1987). The quality and final concentration was estimated by agarose gel electrophoresis using known concentration of uncut  $\lambda$  DNA as a standard. PCR amplification of genomic DNA was carried out in a 20  $\mu$ l reaction volume in an Eppendorf Mastercycler ProS containing 30–50 ng of genomic DNA, PCR buffer (20 mM Tris-HCl, 50 mM KCl), 0.75  $\mu$ M of each primer, 0.125 mM of each dNTP, 1.5 mM MgCl<sub>2</sub> and 0.5 U of Taq DNA polymerase (Life Technology, India). The following touchdown amplification profile was used: (1) initial denaturation 94°C 3 min, (2) 18 cycles of 94°C 50 s, 60°C 50 s, decreasing annealing temperature 0.5°C/cycle, 72°C 50 s, (3) 20 cycles of 94°C 50 s, 50°C 50 s, 72°C 50 s, and (4) final extension 72°C 7 min. The amplification products were separated on 2% agarose gel (Lonza,

Seakem), stained with ethidium bromide and analyzed using the gel documentation system Gel-Doc IT (UVP).

## RESULTS

### SEQUENCING AND ASSEMBLY OF CHICKPEA SEED TRANSCRIPTOME

cDNA libraries were prepared using the mRNA isolated from four stages of seed development i.e., 10 DAA, 20 DAA, 30 DAA and 40 DAA. Sequencing was carried out on the Roche 454 GS FLX Titanium platform and a total of about 1.38 million reads of Q20 quality were obtained (Table 1). The NGS QC ToolKit (Patel and Jain, 2012) was used for removing low quality reads, trimming adapter/primer sequences, trimming sequences containing homopolymers more than 7 bp in length and removing sequences less than 100 bp in length. The 1,129,347 high quality filtered reads obtained had an average length of 342 bp and covered a total of 332,667,201 bases. Amongst the four tissue types, the highest number of reads was recorded for 20 DAA seed tissue (420,329). All the high quality reads were *de novo* assembled using the gsAssembler software (Newbler v.2.3.5, from Roche, <http://www.454.com/products/analysis-software/>) and 86,905 unigenes were obtained. Reference based assembly was also done using the chickpea genome sequence (Varshney et al., 2013) as a reference which generated a total of 93,911 unigenes. Both the *de novo* and reference assemblies were merged and clustered using TGICL (Pertea et al., 2003) to generate the final assembly containing 51,099 unigenes which was used for all further analyses (Table 2). The unigenes were used for BLASTX comparison with the peptide sequences of *C. arietinum* (Varshney et al., 2013) and showed 69.7% similarity. Similar comparison with the Non-redundant (Nr) protein database of NCBI (<ftp://ftp.ncbi.nlm.nih.gov>) showed that about 36,024 unigenes (70.5%) found a match while 15,075 unigenes (29.5%) found no hits in the database and hence can be assumed to be putative novel genes. Also, BLASTX comparison with protein sequences of *Arabidopsis*, *Oryza*, *Glycine*, *Lotus japonicus* and *Medicago* downloaded from Phytozome v 9.1 (<http://www.phytozome.net/>) showed that the chickpea seed transcriptome showed highest similarity with *G. max* (64.7%) and *M. truncatula* (60.6%) as compared to *L. japonicus* (59.8%), *A. thaliana* (57.6%) and *O. sativa* (53.7%).

### FUNCTIONAL ANNOTATION

The assembly was annotated using Uniprot-SwissProt, COG and KEGG databases. The standalone version of NCBI Blast (version 2.2.25+) was downloaded from NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/>) and functional annotation of the transcripts was done by searching against the high quality non-redundant protein sequences in the UniProt Swiss-Prot database (<http://www.uniprot.org>) using BLASTX with a threshold *E*-value of  $\leq 1E-05$ . A total of 21,414 (41.9%) of the 51,099 unigenes could find a match in the Uniprot database, which were then annotated using in house Perl scripts. Accordingly, unigenes were assigned Gene Ontology (GO) terms and were grouped into the three main GO categories; Biological processes, Molecular function and Cellular components (Figure 1A). Majority of the unigenes were found to have a role in various Molecular functions followed by those related to Biological processes and Cellular components. Analysis of their further distribution into various sub categories revealed that the majority of the unigenes

under Biological processes were involved in cellular (25.6%) and metabolic processes (21.05%). About 29.7% and 21.06% of unigenes under Molecular functions category were involved in catalytic activity and binding respectively while amongst the Cellular components, cell (37.74%) and membrane (22.8%) accounted for most of the unigenes (Figure 1A).

Another 9170 unigenes were annotated using the COG (Cluster of Orthologous Groups) database. Annotation according to these clusters classified majority of the unigenes (2033) under the “General function prediction only” category followed by “Post-translational modification, protein turnover, chaperones” (923) and “Translation, ribosomal structure and biogenesis” (721) (Figure 1B). Further, only 9.04% of the annotated unigenes could be distributed into 215 biological pathways enlisted in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Figure 1C). Genes related to “Ribosome” were found to be most abundant in number (120) followed by those for “Spliceosome” (103) and “Biosynthesis of amino acids” (98).

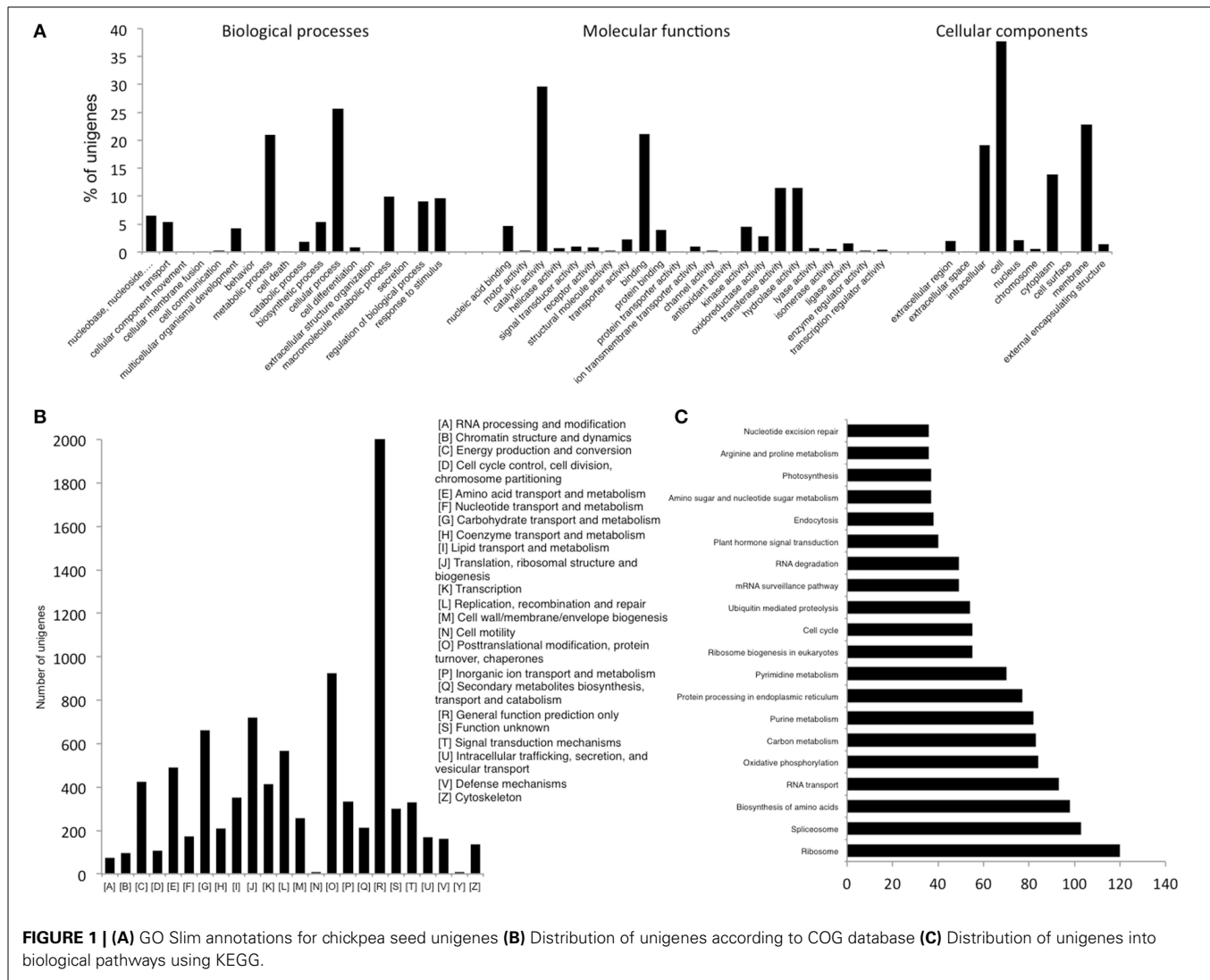
### TRANSCRIPTIONAL FLUX DURING PROGRESSION OF CHICKPEA SEED DEVELOPMENT

In order to elucidate the dynamics of seed development, we performed a number of analyses to determine the pattern of expression of unigenes across the four stages. The assembled set of 51,099 unigenes was used as the reference onto which raw reads from each stage were mapped to generate a putative expression profile for the transcripts. Unigenes with RPKM value  $\geq 1$  in at least one of the seed tissues were considered to be transcriptionally expressed. We found 93.3% of the 51,099 unigenes to be expressing in at least one of the developmental stages and of these, 65.8% expressed at early stages of seed development (10 and 20 DAA) while 43.15% expressed at later stages (30 and 40 DAA). The unigenes were sorted into clusters using k Means/Median clustering function and Figure of Merit (FOM) calculation (Yeung et al., 2001) embedded in MeV v 4.9 to determine the optimal number of clusters to classify gene expression profiles, which resulted in 10 representative clusters. We selected five major clusters (K1–5) which displayed distinct patterns of expression during the seed development process (Supplemental Figure S1). Clusters K1 and K5 contained unigenes that showed a gradual increase and a gradual decrease toward later stages of development respectively. Seed storage proteins constituted majority of cluster K1 while in cluster K5 most of the metabolic enzymes and transcription regulators were seen to decrease as seed development progressed. Clusters K2, K3, and K4 contained unigenes that are highly expressed at 10, 20, and 30 DAA of seed development respectively. A number of heat shock proteins and Zn finger containing transcription factors were present in 10 DAA seeds while both 20 and 30 DAA seed tissue had abundance of various metabolic enzymes and a variety of transcription regulators (Supplemental Table S1). In order to validate the RNA-seq digital expression data, 20 unigenes were randomly selected, primers were designed (Supplemental Table S2) and qRT PCR was performed whereby we obtained a high correlation between the log normalized RPKM and RQ values ( $r^2 = 0.69$ ) (Supplemental Figure S2).

Further, DEGseq (Wang et al., 2010) was used to determine the differentially expressed unigenes across the four stages of seed development where only those unigenes which showed four-fold

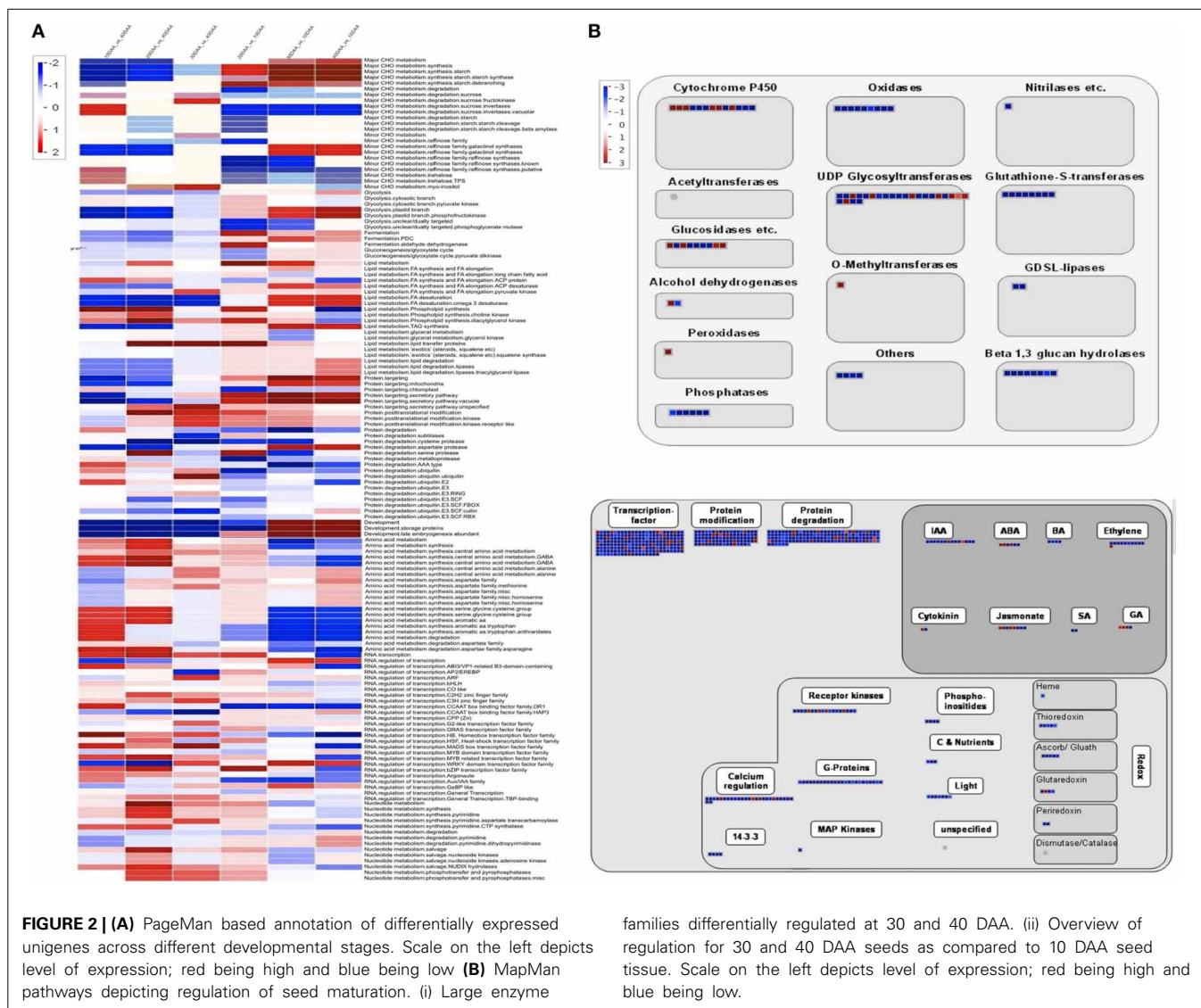
**Table 2 | Assembly statistics.**

Type of assembly	N50 value	N50 Index value	Total no. of transcripts contigs+ singletons)	Avg. size of contig (bp)	Length of largest contig (bp)	No. of contigs 100–500 bp in size	No. of contig >500 bp in size
De novo	441	22,089	86,905	437.6	6,483	72,490	14,415
Reference	541	22,065	93,911	415.96	12,845	64,743	29,168
De novo + reference (TGICL)	918	11,814	51,099	726.7	21,371	21,704	29,395



change in expression ( $p$ -value  $\geq 0.001$ , MARS method) were considered to be differentially expressed. Using this criterion, only 6.75% of the total set of unigenes was seen to be differentially expressed. In order to provide a more comprehensive idea of the pattern of gene expression, we generated two sets of data. The first set was derived using 10 DAA sample as control while in the second set, 40 DAA sample was used as control and the data was then subjected to a Bin-wise Wilcoxon test in PageMan (Figure 2A). Unigenes related to carbohydrate

metabolism (major and minor CHO metabolism) were more highly expressed at later stages as compared to 10 DAA and included genes for starch synthases, debranching, starch cleavage, galactinol synthase etc. but genes for degradation of sucrose, starch and metabolism of raffinose and trehalose had higher expression in 10 DAA seed tissue as compared to 40 DAA. The processes of glycolysis, fermentation, gluconeogenesis, lipid metabolism, protein targeting and degradation and storage protein accumulation were observed to be generally upregulated in



later stages compared to 10 DAA seed tissue. On the other hand, activities like amino acid metabolism, RNA: regulation of transcription, nucleotide metabolism were seen to be upregulated in earlier stages as compared to 40 DAA seed tissue. Although most DNA related processes were upregulated in 10 and 20 DAA seeds, some related to DNA synthesis (histones) were abundantly represented in 40 DAA seeds.

Seed maturation represents an interesting stage which may help understand the regulatory processes leading to quiescence. According to Setia and Malik (1985), increase in dry weight of chickpea seeds is more dramatic after 21 DAF (Days after Fertilization, synonymous to Days after Anthesis in case of chickpea which is a self-pollinating species) which continues upto 42 DAF. MapMan (Thimm et al., 2004) was used to analyse the large enzyme families represented in maturing chickpea seeds (30 and 40 DAA) (Figure 2B). We found that while most metabolic enzymes are downregulated in maturing seeds as compared to early stages, a few members of enzymes such as cytochrome P450, UDP-glycosyltransferases, peroxidases, alcohol dehydrogenases,

O- methyltransferases are expressed preferentially in 30 and 40 DAA seed tissue. Almost all regulatory processes slow down as the seeds reach quiescence except for those related to hormones like ABA, GA, jasmonate, cytokinin and few members of IAA (Figure 2B).

**IDENTIFICATION OF TRANSCRIPTS WITH PREFERENTIAL EXPRESSION IN SEEDS**

We employed two approaches for identifying unigenes that are specific to chickpea seeds. In the first approach, unigenes obtained from the chickpea seeds were compared to the transcripts of both Desi (ICC4958) and Kabuli (ICCV2) varieties of chickpea available in the Chickpea Transcriptome Database (CTDB) (Garg et al., 2011b) which contain transcripts obtained from tissues of shoot, root, mature leaves, flowers and young pods but not from seeds. A comparison between the transcript data from seed and from the whole plant (as present in the CTDB, <http://59.163.192.90:8080/ctdb/cgi-bin/downloadSeq.pl>) revealed that of the 51,099 seed unigenes, 42,160 (82.5%) were common with

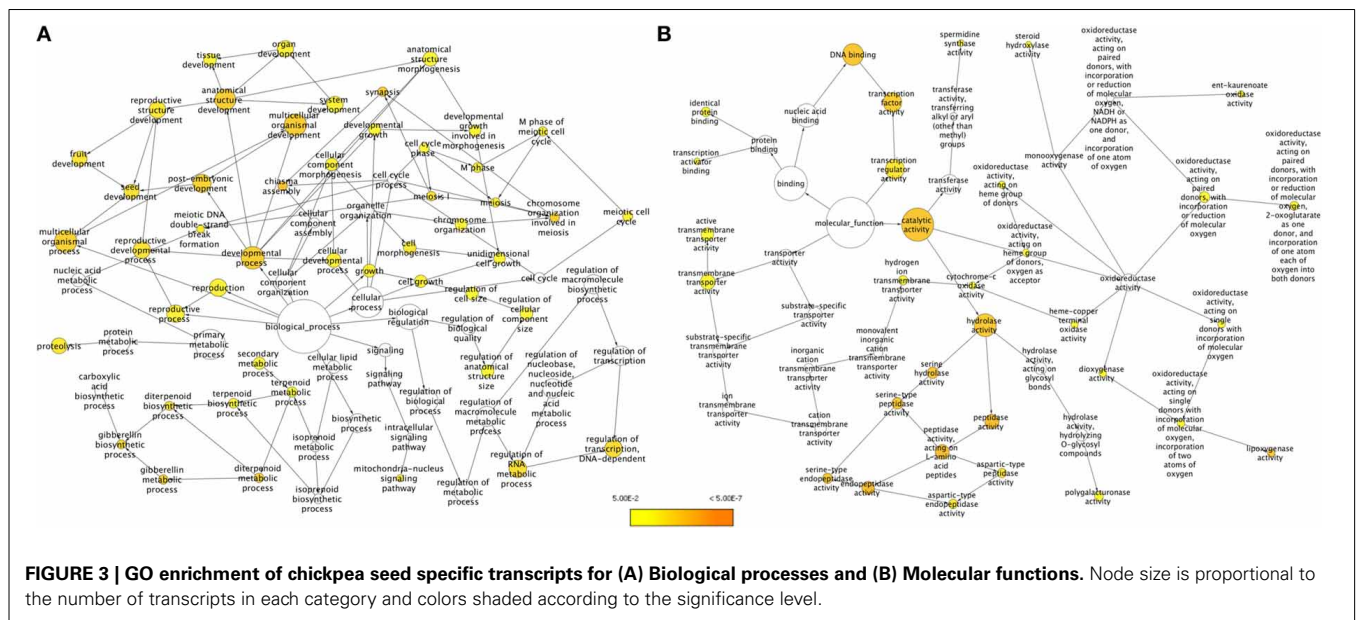
the whole plant while 8,939 (17.5%) of seed unigenes did not find a match in the CTDB. These were therefore assumed to be putatively specific to seed tissue. GO enrichment of these chickpea seed specific unigenes was carried out using BiNGO based on homology with the *Arabidopsis* protein sequences (Figure 3). Under the Biological processes category, a clear enrichment of genes related to development processes especially, genes for seed development, embryonic and post embryonic development, fruit development, reproductive structure development and chiasma formation were seen to be highly represented (Figure 3A). We also found an over representation of categories like lipid metabolic process, terpenoid metabolic process, proteolysis, regulation of transcription, regulation of RNA metabolic process and gibberellin metabolic process. Categories like hydrolase activity, transcription factor activity and lipoxygenase activity were found to be overrepresented under Molecular functions (Figure 3B).

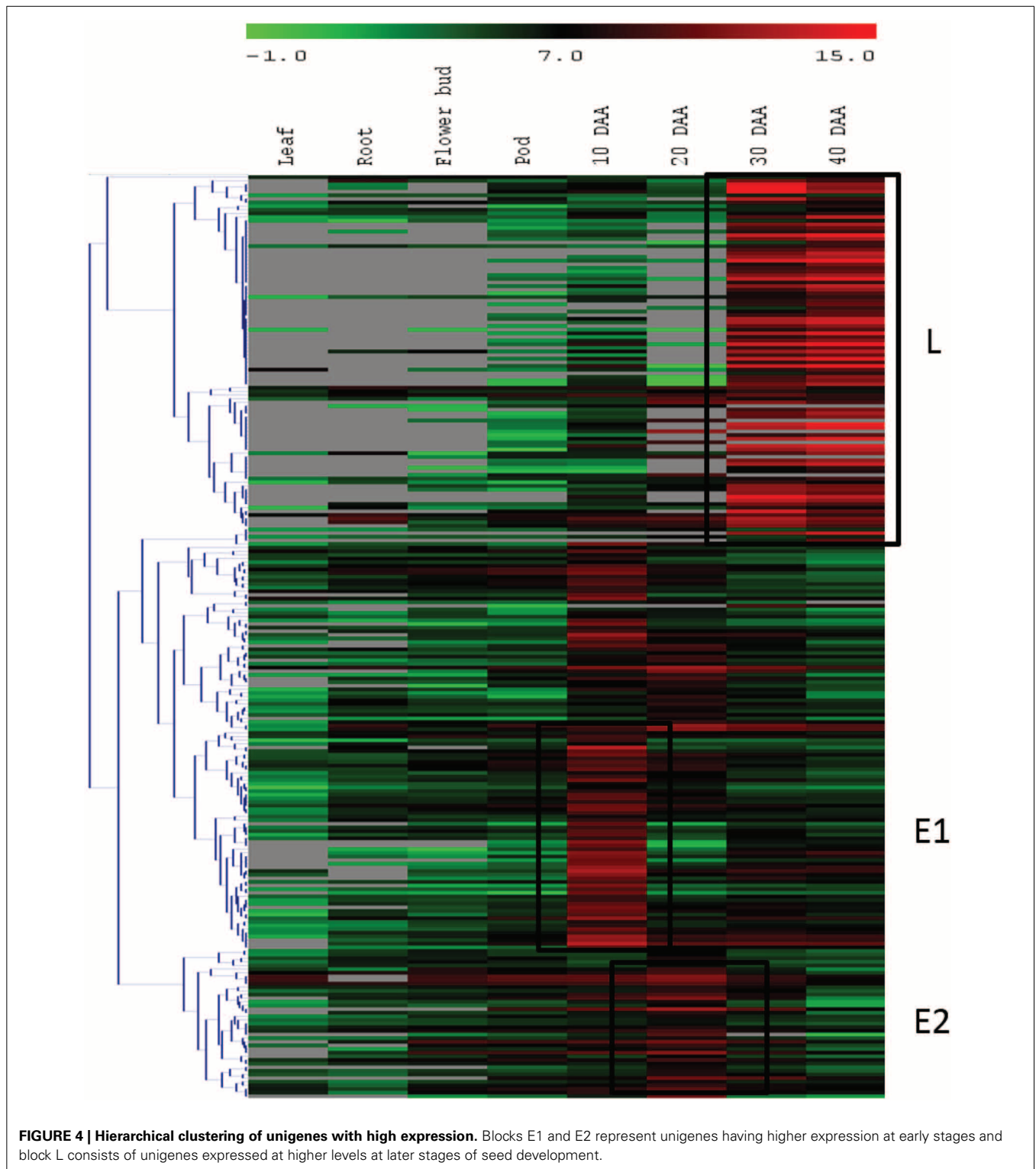
For the second approach, we mapped the transcript sequence reads generated from the four stages of developing seed (generated in this study) as well as the reads from the transcriptomes of leaf, root, flower bud and pod of chickpea reported by Garg et al. (2011b) onto the assembled unigenes from chickpea seed tissue. Log<sub>2</sub> normalized RPKM values were calculated and subjected to Z-score analysis as described by Severin et al. (2010). The heat map was generated taking 1000 most highly expressing unigenes (calculated by taking sum of raw reads mapped onto each unigene) as well as the Z-scores above 2 in each seed tissue. Hierarchical clustering of these unigenes showed distinct blocks where unigenes with preferential expression in seed tissue could be found (Figure 4). We further analyzed these seed specific genes and found that almost all of the 10 DAA specific unigenes (Block E1; Figure 4) encoded heat shock proteins and chaperones. The 20 DAA seed had preferential expression of genes related to the proteasome machinery along with a variety of metabolic enzymes (Block E2; Figure 4). In later stages i.e., 30 and 40 DAA,

while a sizeable portion of the highly expressing unigenes was constituted by storage proteins, a number of metabolic enzymes such as alcohol dehydrogenases, seed lipoxygenases and subtilisin proteases were also detected (Block I; Figure 4) (Supplemental Table S3). Five of the putative seed specific unigenes were analyzed using qRT PCR (Figure 5) in order to validate this approach.

### TRANSCRIPTION FACTORS IN CHICKPEA SEED DEVELOPMENT

Transcription factors play crucial roles in mediating transcriptional regulation. Therefore, we studied the expression dynamics of a number of transcripts which were identified to encode different transcription factors. Peptide sequences of different TFs from five different legumes (*Cajanus cajan*, *C. arietinum*, *G max*, *Lotus japonicus* and *M. truncatula*) were downloaded from Plant Transcription Factor Database (planttfdb.cbi.pku.edu.cn) and their HMM profiles were built using hmmbuild (HMMER 3.0, ftp://selab.janelia.org/pub/software/hmmer) and used for searching the chickpea seed transcriptome. We found 1,743 unigenes from the chickpea seed transcriptome to be transcription factors, which were further categorized into 54 transcription factor families of which the Zinc finger family was most abundant followed by B3, MYB, bHLH and MADS transcription factor families (Figure 6A). Based on the information obtained from previous reports (Verdier and Thompson, 2008; Agarwal et al., 2011), a number of transcription factors, which are known to be involved specifically in the process of seed development were analyzed for their pattern of digital expression in 10, 20, 30, and 40 DAA seed tissue (Figure 6B). We found that a majority of these transcription factors showed higher expression in early stages of seed development which decreased as the seed progressed to maturation. These included transcription factors belonging to AP2, ARE, HAP, bHLH, bZIP, NAC, and SBP families. However, a few members of MYB, bHLH, bZIP, C3H and WRKY had higher expression in 40 DAA seed tissue.





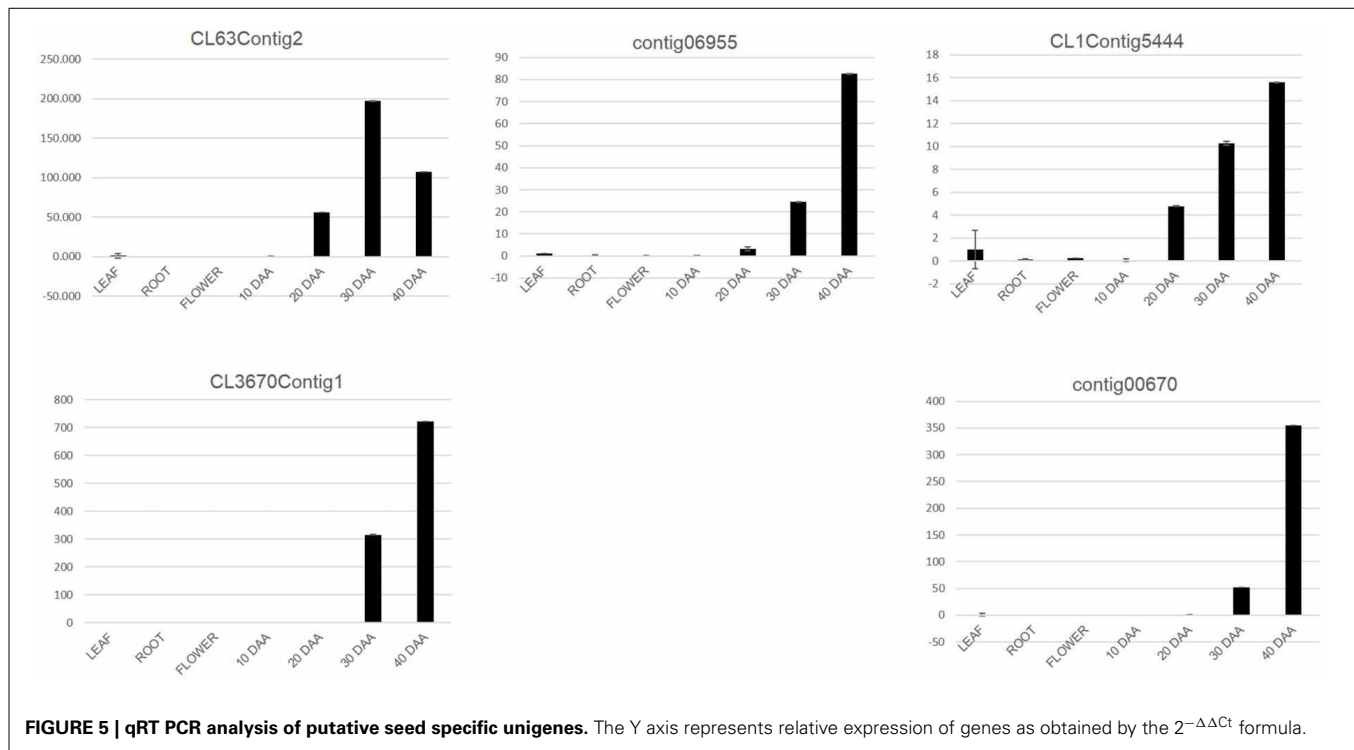
**FIGURE 4 | Hierarchical clustering of unigenes with high expression.** Blocks E1 and E2 represent unigenes having higher expression at early stages and block L consists of unigenes expressed at higher levels at later stages of seed development.

#### IMPORTANT SECONDARY METABOLITES IN CHICKPEA SEEDS

Flavonoids and isoflavonoids are one of the most popular groups of secondary metabolites found in plants. Many legume seeds have been reported to be rich sources of these secondary metabolites (Prati et al., 2007; Wang et al., 2008; Heiras-Palazuelos et al.,

2013). KEGG based annotation of the seed unigenes into various biological pathways revealed 12 enzymes involved in the flavonoid biosynthesis pathway and 5 in the isoflavonoid biosynthesis and these were represented by 29 unigenes in our transcriptome (Supplemental Figure S3).





**FIGURE 5 | qRT-PCR analysis of putative seed specific unigenes.** The Y axis represents relative expression of genes as obtained by the  $2^{-\Delta\Delta Ct}$  formula.

An RPKM based analysis taking into account leaf, roots, flower bud and pods along with the four stages of seed development showed that most of the 29 unigenes had higher expression in vegetative tissues as compared to seeds (Figure 7, Supplemental Table S4). However, distinct exceptions to this trend were the enzymes isoflavone-2-hydroxylase (4), which is known to be involved in biosynthesis of hydroxygenistein and flavonoid 3 hydroxylase 2, which is known to be involved in biosynthesis of dihydroquercetin and quercetin. Isoflavone hydroxylase (4) was seen to have especially high expression at 10 DAA and its expression gradually declined toward 40 DAA while its expression remained comparatively low in non-seed tissues. The enzyme Flavonoid 3 hydroxylase was seen to have a distinctly higher expression in 20 DAA seed tissue while remaining low in all other tissues (Figure 6). Geraniol 8-hydroxylase and Spermidine hydroxycinnamoyl transferase (1) also showed preferential expression in 10 and 30 DAA seed tissue while Chalcone-flavonone isomerase (2) and Trans-cinnamate 4-monooxygenase (2) showed higher expression in 20 DAA seed tissue as compared to other samples (Figure 6). Although differential expression analysis showed that most secondary metabolic activities are suppressed in later stages, some enzymes like Trans-cinnamate 4-monooxygenase (1), 2-hydroxyisoflavanone synthase, Spermidine hydroxycinnamoyl transferase (2) and Flavonol synthase showed higher expression in 30 and 40 DAA seed tissue as compared to 10 and 20 DAA seeds.

#### SSR MINING AND VALIDATION

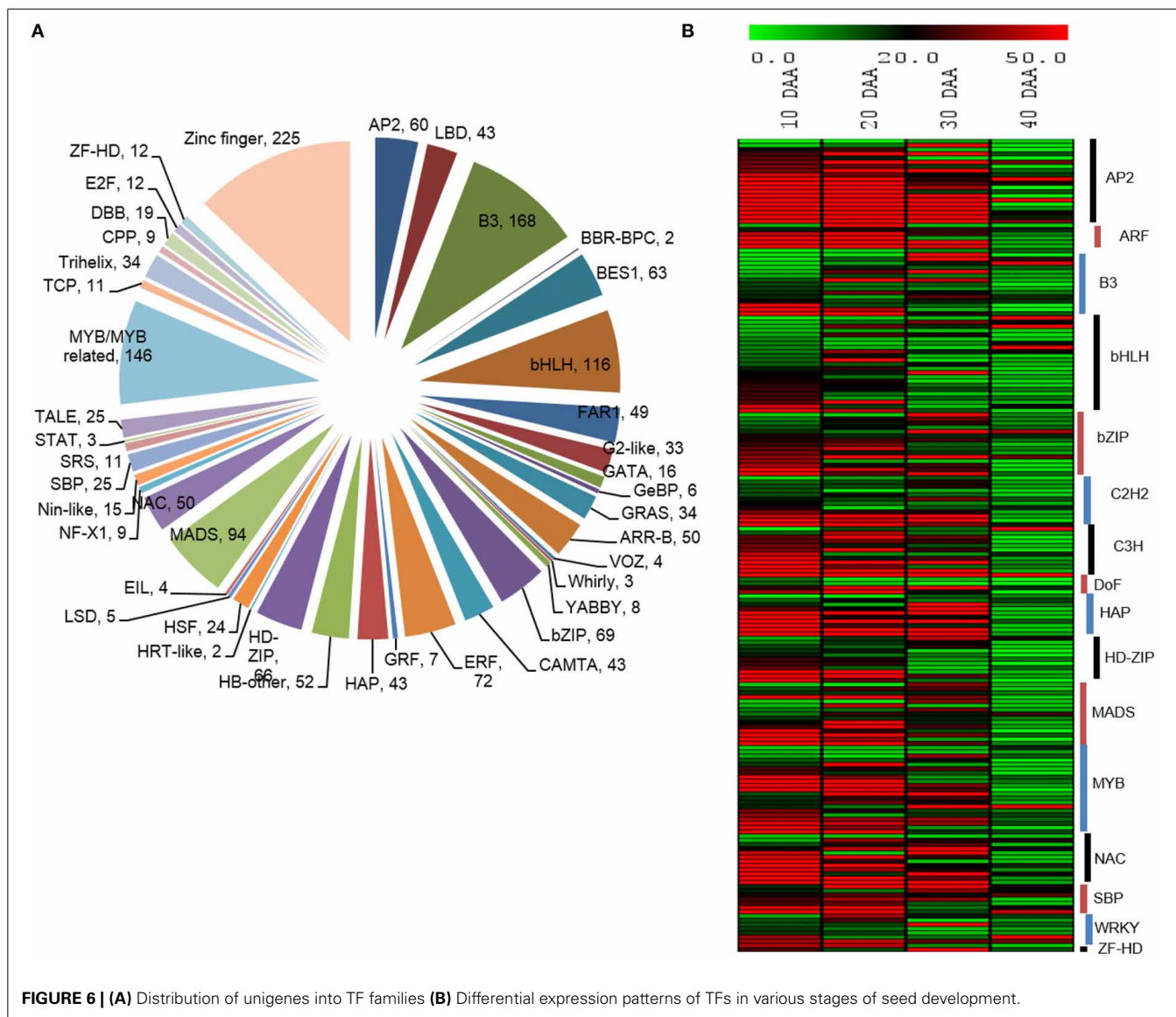
Transcriptome data is very widely used as a source to develop molecular markers especially the Simple Sequence Repeats (SSRs). Therefore, in order to identify SSRs (dinucleotide to

hexanucleotide repeats) in transcripts of developing seeds of chickpea, we mined the transcripts generated in this study using the MISA Perl script. Mononucleotide repeats were not considered as we trimmed sequence reads with homopolymers of more than seven bases as a part of quality control. A total of 12,606 SSRs could be identified in 9305 (18.2%) transcripts of chickpea (Table 3). The average frequency of SSRs was found to be one SSR per 2.94 kb of the chickpea seed transcriptome sequence. Furthermore, 2249 transcripts contained more than one SSR, and 841 SSRs were present in compound formation. The largest fraction of SSRs identified were trinucleotides (44.46%) followed by tetranucleotides (25.82%). Among the trinucleotide repeat containing SSRs identified, AAG/CTT accounted for majority (24%) of the repeats followed by AAT/ATT which accounted for 20% of trinucleotide repeats (Figure 8). Moreover, attempts were made to identify the novel SSRs out of the set here. These SSRs (including 200 bp upstream and downstream flanking sequences), were compared to the 48,328 SSR sequences reported from whole chickpea genome (Varshney et al., 2013). We found that only 1557 chickpea seed SSRs found matches with those from the chickpea genome. So, it can be assumed that the remaining 11,049 SSRs from chickpea seed are novel in chickpea. Forward and reverse primers were designed for 48 SSRs (Supplemental Table S2) and 43 of these (89.58%) were successfully amplified from genomic DNA of *C. arietinum* (ICCV2) (Supplemental Figure S4).

#### DISCUSSION

##### RNA SEQ IS A USEFUL METHOD FOR RESOURCE GENERATION

Analysis of transcriptomes is crucial as it not only helps in large scale identification of mRNA but also provides insights into the molecular basis of genes involved in developmental processes.



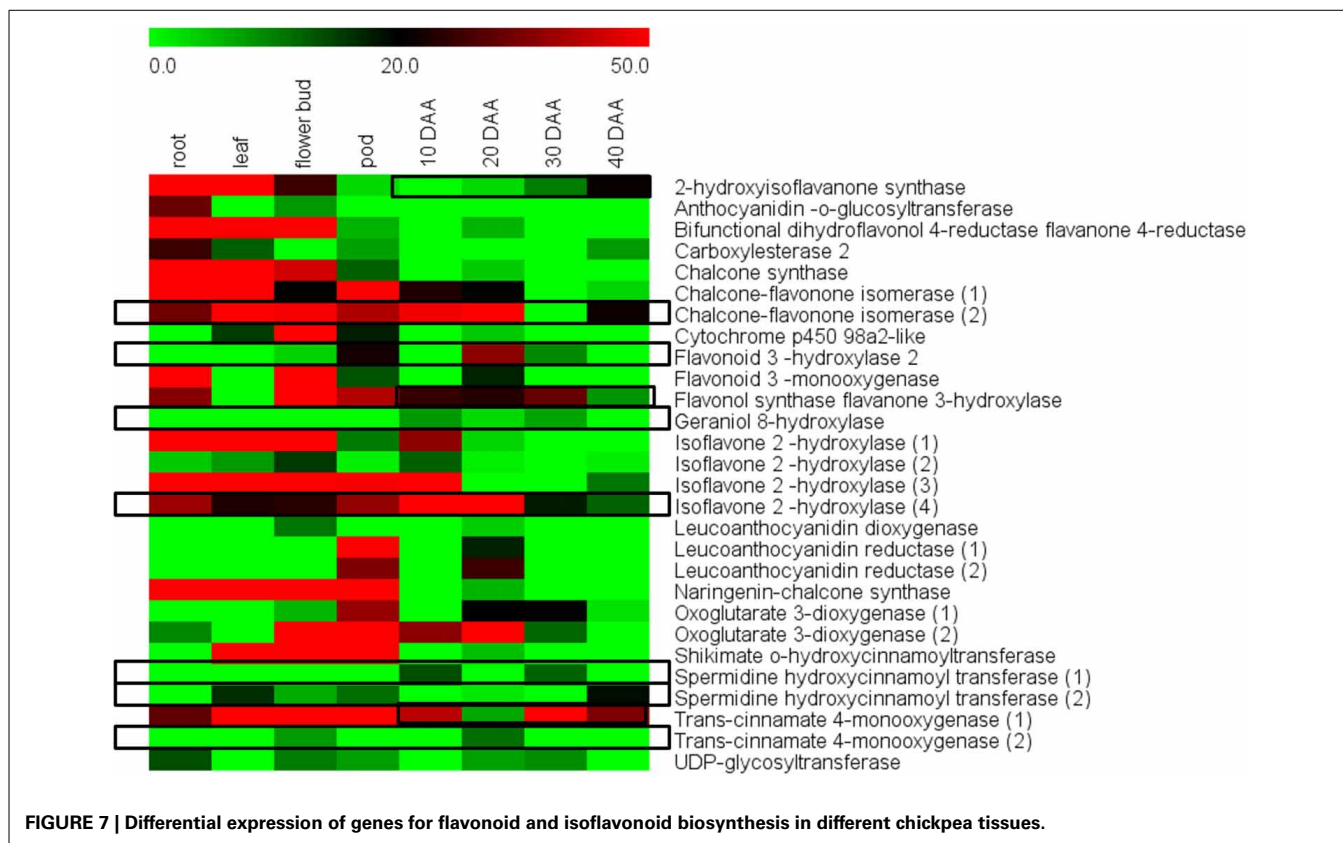
**FIGURE 6 | (A)** Distribution of unigenes into TF families **(B)** Differential expression patterns of TFs in various stages of seed development.

Seed development is an important process in plants and understanding the developmental dynamics, especially in plants of economic importance where seeds serve as a source of food and feed, may greatly help in developing crop improvement strategies. The ICCV2 cultivar of Kabuli chickpea was selected for analysis as it has gained distinction for being the first extra-short-duration (85–90 days) Kabuli variety with *Fusarium* wilt resistance and heat tolerance which can also escape terminal drought due to early maturation. In addition to this, the seeds of this cultivar are valued due to their higher protein content (23.4% as compared to 19.4% in Desi varieties) and better quality traits like biological value, true digestibility, net protein utilization etc., (ICRISAT, 1990).

Described by Margulies et al. (2005), the first NGS platform to integrate pyrosequencing using their PTP device was commercialized by Roche/454. Unlike platforms that produce shorter read lengths, the Roche/454 platform does not require the run time to be doubled for the sequencing of mate-pair templates. In

addition, the long reads produced by the 454 platform are especially useful for *de novo* assembly of non-model transcriptomes (Vera et al., 2008). We have used both the *de novo* as well as the reference based assemblies of the chickpea seed transcriptome to generate a more comprehensive and inclusive set of unigenes that provide an unbiased representation of the four distinct stages of chickpea seed development. The assembly statistics revealed that the merged assembly had better characteristics, such as higher N50 value, average contig length, total number of contigs, longest contig length etc. which serve as a yardstick to assess the quality of an assembly.

Transcriptomes serve as a source of valuable information, especially, discovery of genes and markers (Garg et al., 2011a). Most of the conventional methods fall short in terms of coverage as well as handling of data. Out of the unigenes generated in this study, 18.2% were found to be containing SSRs with an average frequency of 1 SSR per 2.94 Kb of sequence, which is higher than that of the rest of the chickpea plant (Garg et al., 2011a). Further



**Table 3 | Numbers and distribution of SSRs in chickpea seed tissue.**

Total number of sequences examined	51,099
Total size of examined sequences (bp)	37,134,000
Total number of identified SSRs	12,606
Number of SSR containing sequences	9,305
Number of sequences containing more than 1 SSR	2,249
Number of SSRs present in compound formation	841

**DISTRIBUTION OF SSRs ACROSS DIFFERENT REPEAT TYPE CLASSES**

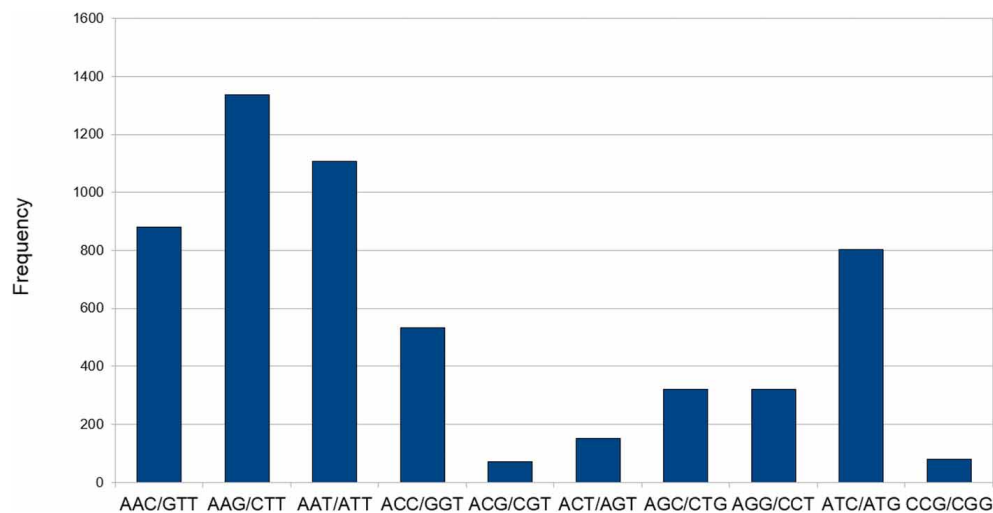
Unit size	Number of SSRs
2 (dinucleotides)	1,624
3 (trinucleotides)	5,604
4 (tetranucleotides)	3,255
5 (pentanucleotides)	1,075
6 (hexanucleotides)	1,048

analysis revealed that of the total set of mined SSRs, about 87% were novel and moreover about 90% were validated in genomic DNA of chickpea thereby suggesting that the SSRs generated in this study had considerable utility for marker-assisted selection in chickpea breeding programs.

**CHICKPEA SEED DEVELOPMENT IS A DYNAMIC PROCESS**

The overall process of chickpea seed development was seen to be similar to that of other dicot seeds whereby the cell division phase is followed by cell expansion and then by maturation.

Nevertheless, a more in depth analysis revealed a number of processes that may be more specific to seed development as compared to the rest of the plant. Gene Ontology based annotation revealed enrichment of categories like catalytic activity, binding, cellular and metabolic processes and cellular components like cell and membrane which clearly indicate that the process of seed development is genetically programmed and therefore may be correlated to increased metabolic and catalytic activities in the cell and membranes (Borisjuk et al., 2004). In contrast, the study of the transcriptomes from 5 other tissues of chickpea (Garg et al., 2011a) revealed the highest number of transcripts to have a function in the chloroplast and plasma membrane, as well as in processes like protein metabolism and response to various stimuli and stress. It was very clear that expression of most of the unigenes occurred at early stages suggesting a gradual decrease in transcriptome dynamics as the seed progressed toward desiccation and attaining quiescence. Moreover, the early stages were marked by increased metabolic activity involving high protein turnover (synthesis and degradation) whereas the later stages were characterized by protein storage and nutrient reservoir activities. It was also evident that the majority of the unigenes expressed across all stages of seed development with only 6.75% of unigenes having differential expression in different stages of seed development. This was in agreement with earlier studies of soybean (Severin et al., 2010) and *Arabidopsis* (Le et al., 2010) which had suggested that most genes involved in various seed functions are shared by all stages of seed development although there might be significant changes in their transcriptional activities. Nevertheless, each



**FIGURE 8 | Abundance of different types of trinucleotide repeat containing SSRs.**

developmental stage may have a very small gene set specific to that stage (Sreenivasulu and Wobus, 2013). Hierarchical clustering of the most highly expressing genes revealed a remarkably high expression of heat shock proteins in 10 DAA tissue suggesting the formation of nascent proteins since the general role of heat shock proteins is to act as molecular chaperones regulating the folding and accumulation of proteins as well as localization and degradation in all plant and animal species (Feder and Hofmann, 1999; Schulze-Lefert, 2004; Panaretou and Zhai, 2008; Hu et al., 2009; Gupta et al., 2010) especially the Hsp70, which functions as a chaperone for newly synthesized proteins to prevent their accumulation as aggregates by folding in a proper way during transfer to their final location (Sung et al., 2001; Su and Li, 2008). Similarly, at later stages of development (30 and 40 DAA), there was an abundance of genes related to nutrient reservoir activity, thus emphasizing the process of storage protein accumulation. But distinct expression of any particular category of transcripts was not seen in 20 DAA tissue although this stage displayed a high expression of most genes such as those belonging to categories like catalytic activity, metabolic process, carbohydrate metabolic process etc. This reiterates the hypothesis that the process of seed development is a time of high metabolic activity where early stages are associated with nascent protein generation as well as degradation while most of the storage protein accumulation occurs at later stages.

### GENES REGULATING SEED DEVELOPMENT

Comparison with the known chickpea transcripts from 5 tissues in the CTDB (Garg et al., 2011b) helped in identifying putative seed specific transcripts among which, presence of transcripts related to seed storage proteins and lipoxygenases along with enrichment of genes for seed, embryo, fruit development transcription activity, lipid metabolism and terpenoid metabolism not only further validated the seed transcriptome but also pointed to the importance of these gene groups in the process of seed development. Lipoxygenases are an important group of enzymes

which catalyze the oxidation of polyunsaturated fatty acids such as linoleic (18:2) and  $\alpha$ -linolenic (18:3) to produce unsaturated fatty acid hydroperoxides (Liavonchanka and Feussner, 2006) while alcohol dehydrogenases encode glycolytic enzymes. Preferential expression of these enzymes at later stages of seed development along with seed storage proteins add value to the nutritive quality of chickpea seeds.

Transcription factors are crucial regulatory proteins which mediate transcriptional regulation. They are known to be involved in almost every key process that occurs during plant development. Since seed development is such a dynamic process, it was imperative that we study the TFs involved. We found that majority of the TFs involved in seed developmental processes were highly expressed in early stages i.e. in 10 and 20 DAA seed tissue. Most of the well-known TFs known to be involved in seed development were also found in our transcriptome thereby reiterating the validity of our study. Almost all of the AP2 and HAP TFs were seen to be expressed at early stages while a few members of WRKY and C3H Zn finger containing TFs had higher expression at later stages. This reinforces the observation that gene activity in seeds decreases as they reach maturity (Sreenivasulu and Wobus, 2013). Such biased pattern of expression favoring early stages of chickpea seed development also suggests that most TFs are involved in processes like protein turnover, synthesis of storage compounds and general metabolism since these processes are abundant during early seed development. This also suggests that a very small proportion of these regulatory proteins are involved in the process of achieving dormancy since most gene activity ceases during later stages of chickpea seed development. Further studies related to this small group of TFs can lead to deciphering of many processes related to seed dormancy.

### CHICKPEA SEEDS CAN BE A SOURCE OF ANTIOXIDANTS

During this study, we recurrently encountered a number of genes related to secondary metabolic processes. The potent antioxidant properties of flavonoids like genistein and quercetin have

been well investigated (Borras et al., 2006; Boots et al., 2008; Zhang et al., 2010). Enzymes involved in biosynthesis of these flavonoids had distinctly higher expression in 10 and 20 DAA seed tissue suggesting that genistein and quercetin accumulate preferentially in chickpea seeds as compared to rest of the plant. We also observed that although most gene activity slows down as the seed nears quiescence, some enzymes related to flavonoid biosynthesis are upregulated in the later stages. This suggests that flavonoid biosynthesis may be closely associated with chickpea seed maturation. Hence, chickpea seeds can be considered as a good source of antioxidants along with proteins and carbohydrates.

In conclusion, studying molecular mechanisms underlying plant development are usually difficult but extremely important. Although molecular aspects of seed development have been investigated in model plants such as *Arabidopsis* and *Medicago*, studying the process in important crop legumes such as chickpea is necessary to unravel any unique processes regulating chickpea seed development. This deep transcriptome sequencing analysis revealed that the overall pattern of seed development was largely common in chickpea and other dicot seeds. However, preferential expression of certain genes/families may be responsible for the unique nutritive values of chickpea seeds. Hence this study would serve as a foundation for characterization of candidate genes which would not only provide novel insights into understanding seed development but also provide resources for improved chickpea breeding.

## ACKNOWLEDGMENTS

This work was funded by the Department of Biotechnology, Government of India, under the Next Generation Challenge Programme on Chickpea Genomics (grant number BT/PR12919/AGR/02/676/2009). SP acknowledges the award of research fellowship from the Department of Biotechnology. CK acknowledges the award of research fellowship from the Council of Scientific and Industrial Research. Authors acknowledge Kamlesh Sahu for his help.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpls.2014.00698/abstract>

## REFERENCES

- Abbo, S., Molina, C., Jungmann, R., Grusak, M. A., Berkovitch, Z., Reifen, R., et al. (2005). Quantitative trait loci governing carotenoid concentration and weight in seeds of chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 111, 185–195. doi: 10.1007/s00122-005-1930-y
- Afonso-Grunz, E., Molina, C., Hoffmeier, K., Rycak, L., Kudapa, H., Varshney, R. K., et al. (2014). Genome-based analysis of the transcriptome from mature chickpea root nodules. *Front. Plant Sci.* 5:325. doi: 10.3389/fpls.2014.00325
- Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33, 189–202. doi: 10.1002/bies.201000107
- Baud, S., Boutin, J.-P., Miquel, M., Lepiniec, L., and Rochat, C. (2002). An integrated overview of seed development in *Arabidopsis thaliana* ecotype WS. *Plant Physiol. Biochem.* 40, 151–160. doi: 10.1016/S0981-9428(01)01350-X
- Boots, A. W., Haenen, G. R. M. M., and Bast, A. (2008). Health effects of quercetin, From antioxidant to nutraceutical. *Eur. J. Pharmacol.* 585, 325–337. doi: 10.1016/j.ejphar.2008.03.008
- Borisjuk, L., Rolletschek, H., Radchuk, R., Weschke, W., Wobus, U., and Weber, H. (2004). Seed development and differentiation, a role for metabolic regulation. *Plant Biol.* 6, 375–386. doi: 10.1055/s-2004-817908
- Borras, C., Gambini, J., Gomez-Cabrera, M. C., Sastre, J., Pallardó, F. V., Mann, G. E., et al. (2006). Genistein, a soy isoflavone, up-regulates expression of antioxidant genes, involvement of estrogen receptors, ERK1/2, and NFκB. *FASEB J.* 20, 2136–2138. doi: 10.1096/fj.05-5522fj
- Cheung, F., Haas, B. J., Goldberg, S. M. D., May, G. D., Xiao, Y., and Town, C. D. (2006). Sequencing *Medicago truncatula* expressed sequenced tags using 454 life sciences technology. *BMC Genomics* 7:272. doi: 10.1186/1471-2164-7-272
- Choudhary, S., Sethy, N. K., Shokeen, B., and Bhatia, S. (2009). Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theor. Appl. Genet.* 118, 591–608. doi: 10.1007/s00122-008-0923-z
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15
- Feder, M. E., and Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and stress response, evolutionary and ecological physiology. *Annu. Rev. Physiol.* 61, 243–282. doi: 10.1146/annurev.physiol.61.1.243
- Friedman, M. (1996). Nutritional value of food proteins from different food sources. *J. Agric. Food Chem.* 44, 6–29. doi: 10.1021/jf9400167
- Gallardo, K., Firnhaber, C., Zuber, H., Hélicher, D., Belghazi, M., Henry, C., et al. (2007). A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds, Evidence for metabolic specialization of maternal and filial tissues. *Mol. Cell Proteomics* 6, 2165–2179. doi: 10.1074/mcp.M700171-MCP200
- Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., et al. (2011b). Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* 156, 1661–1678. doi: 10.1104/pp.111.178616
- Garg, R., Patel, R. K., Tyagi, A. K., and Jain, M. (2011a). *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res.* 18, 53–63. doi: 10.1093/dnares/dsq028
- Garg, R., Sahoo, A., Tyagi, A. K., and Jain, M. (2010). Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochem. Biophys. Res. Commun.* 396, 283–288. doi: 10.1016/j.bbrc.2010.04.079
- Girke, T., Todd, J., Ruuska, S., White, J., Benning, C., and Ohlrogge, J. (2000). Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* 124, 1570–1581. doi: 10.1104/pp.124.4.1570
- Gregersen, P. L., Rinch-pedersen, H. B., and Holm, P. B. (2005). A microarray based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic Res.* 14, 887–905. doi: 10.1007/s11248-005-1526-y
- Gupta, S. C., Sharma, A., Mishra, M., Mishra, R., and Chowdhuri, D. K. (2010). Heat shock proteins in toxicology, How close and how far? *Life Sci.* 86, 377–384. doi: 10.1016/j.lfs.2009.12.015
- Gutierrez-Gonzalez, J. J., Tu, Z. J., and Garvin, D. F. (2013). Analysis and annotation of the hexaploid oat seed transcriptome. *BMC Genomics* 14:471. doi: 10.1186/1471-2164-14-471
- Heiras-Palazuelos, M. J., Ochoa-Lugo, M. I., Gutierrez-Dorado, R., Lopez-Valenzuela, J. A., Mora-Rochin, S., Milan-Carrillo, J., et al. (2013). Technological properties, antioxidant activity and total phenolic and flavonoid content of pigmented chickpea (*Cicer arietinum* L.) cultivars. *Int. J. Food Sci. Nutr.* 64, 69–76. doi: 10.3109/09637486.2012.694854
- Hiremath, P. J., Farmer, A., Cannon, S. B., Woodward, J., Kudapa, H. B., Tuteja, R., et al. (2011). Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol. J.* 9, 922–931. doi: 10.1111/j.1467-7652.2011.00625.x
- Hu, W., Hu, G., and Han, B. (2009). Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. *Plant Sci.* 176, 583–590. doi: 10.1016/j.plantsci.2009.01.016
- ICRISAT. (1990). Chickpea Kabuli Variety ICCV2, Plant Material Description no.22. Available online at: [http://oar.icrisat.org/524/1/PMD\\_22.pdf](http://oar.icrisat.org/524/1/PMD_22.pdf)
- Iqbal, A., Khalil, I. A., Ateeq, N., and Sayyar Khan, M. (2006). Nutritional quality of important food legumes *Food Chem.* 97, 331–335. doi: 10.1016/j.foodchem.2005.05.011

- Jones, S. I., and Vodkin, L. O. (2013). Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS ONE* 8:e59270. doi: 10.1371/journal.pone.0059270
- Kudapa, K., Azam, S., Sharpe, A. G., Taran, B., Li, R., Deonovic, B., et al. (2014). Comprehensive transcriptome assembly of chickpea (*Cicer arietinum* L.) using Sanger and next generation sequencing platforms: development and applications. *PLoS ONE* 9:e86039. doi: 10.1371/journal.pone.0086039
- Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8063–8070. doi: 10.1073/pnas.1003530107
- Liavonchanka, A., and Feussner, I. (2006). Lipoxygenases, occurrence, functions and catalysis. *J. Plant Physiol.* 163, 348–357. doi: 10.1016/j.jplph.2005.11.006
- Maere, S., Heymans, K., and Kuiper, M. (2005). *BiNGO*, a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551
- Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402. doi: 10.1146/annurev.genom.9.081307.164359
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959
- Morozova, O., Hirst, M., and Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genomics Hum. Genet.* 10, 135–151. doi: 10.1146/annurev-genom-082908-145957
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Panaretou, B., and Zhai, C. (2008). The heat shock proteins, their roles as multi-component machines for protein folding. *Fungal Biol. Rev.* 22, 110–119. doi: 10.1016/j.fbr.2009.04.002
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit, A toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., et al. (2003). TIGR gene indices clustering tools (TGICL), a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651–652. doi: 10.1093/bioinformatics/btg034
- Prati, S., Baravelli, V., Fabbri, D., Schwarzingler, C., Brandolini, V., Maietti, A., et al. (2007). Composition and content of seed flavonoids in forage and grain legume crops. *J. Sep. Sci.* 30, 491–501. doi: 10.1002/jssc.200600383
- Salem, M., Rexroad, C. E., Wang, J., Thorgaard, G. H., and Yao, J. (2010). Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches. *BMC Genomics* 11:564. doi: 10.1186/1471-2164-11-564
- Schulze-Lefert, P. (2004). Plant immunity, the origami of receptor activation. *Curr. Biol.* 14, R22–R24. doi: 10.1016/j.cub.2003.12.017
- Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nat. Methods* 5, 16–18. doi: 10.1038/nmeth1156
- Setia, N., and Malik, C. P. (1985). Distribution of mineral nutrients in developing fruits of chickpea (*Cicer arietinum*). *Phyton (Austria)* 25, 117–121.
- Severin, A. J., Woody, J. L., Bolon, Y. T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq Atlas of Glycine max, a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Singh, V. K., Garg, R., and Jain, M. (2013). A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnol.* 11, 691–701. doi: 10.1111/pbi.12059
- Sreenivasulu, N., and Wobus, U. (2013). Seed-development programs, a systems biology based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* 64, 189–217. doi: 10.1146/annurev-arplant-050312-120215
- Su, P. H., and Li, H. M. (2008). *Arabidopsis* stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. *Plant Physiol.* 146, 1231–1241. doi: 10.1104/pp.107.114496
- Sung, D. Y., Kaplan, F., and Guy, C. L. (2001). Plant Hsp70 molecular chaperones, protein structure, gene family, expression and function. *Physiol. Plant.* 113, 443–451. doi: 10.1034/j.1399-3054.2001.1130402.x
- Thimm, O., Blaessing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN, a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-313X.2004.02016.x
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I., et al. (2008). Rapid transcriptome characterization for a non-model organism using 454 pyrosequencing. *Mol. Ecol.* 17, 1636–1647. doi: 10.1111/j.1365-294X.2008.03666.x
- Verdier, J., and Thompson, R. D. (2008). Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. *Plant Cell Physiol.* 49, 1263–1271. doi: 10.1093/pcp/pcn116
- Wang, L., Feng, Z., Wang, X., and Zhang, X. (2010). DEGseq, An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612
- Wang, M. L., Gillaspie, A. G., Morris, J. B., Pittman, R. N., Davis, J., and Pederson, G. A. (2008). Flavonoid content in different legume germplasm seeds quantified by HPLC. *Plant Genet. Resour.* 6, 62–69. doi: 10.1017/S1479262108923807
- Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201
- White, J. A., Todd, J., Newman, T., Focks, N., Girke, T., Martínez de Ilárduya, O., et al. (2000). A new set of *Arabidopsis* expressed sequence tags from developing seeds, The metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* 124, 1582–1594. doi: 10.1104/pp.124.4.1582
- Xue, L. J., Zhang, J. J., and Xue, H. W. (2012). Genome-Wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS ONE* 7:e31081. doi: 10.1371/journal.pone.0031081
- Yeung, K. Y., Haynor, D. R., and Ruzzo, W. L. (2001). Validating clustering for gene expression data. *Bioinformatics* 17, 309–318. doi: 10.1093/bioinformatics/17.4.309
- Zhang, Z. J., Cheang, L. C., Wang, M. W., and Lee, S. M. (2010). Quercetin exerts a neuroprotective effect through inhibition of the iNOS/NO system and pro-inflammation gene expression in PC12 cells and in zebrafish. *Int. J. Mol. Med.* 27, 195–203. doi: 10.3892/ijmm.2010.571

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 September 2014; accepted: 24 November 2014; published online: 16 December 2014.

Citation: Pradhan S, Bandhiwal N, Shah N, Kant C, Gaur R and Bhatia S (2014) Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Front. Plant Sci.* 5:698. doi: 10.3389/fpls.2014.00698

This article was submitted to *Plant Genetics and Genomics*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2014 Pradhan, Bandhiwal, Shah, Kant, Gaur and Bhatia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.