Background

Cyclophilins (CYPs) are ubiquitous proteins found in organisms ranging from archaea and bacteria to plants and animals [1,2]. As they were originally identified as receptors for the immunosuppressive drug cyclosporine A (CsA), CYPs are classified in the immunophilin family of proteins possessing peptidyl prolyl cis/trans isomerase activity. Multiple CYPs have been found in genomes of various prokaryotes, but only a few have been studied in detail. The Escherichia coli genome encodes two CYPs, a cytosolic form (EcCYP-18) and its periplasmic counterpart (EcCYP-20) [3]. In the yeast Saccharomyces cerevisiae there are at least 8 different CYPs, Cpr1 to Cpr8 [4]. These proteins are not essential for growth, but are crucial for survival after heat stress [5]. The human genome encodes 16 unique CYPs, categorized into 7 major groups, namely human CYP A (hCYP-A), hCYP-B, hCYP-C, hCYP-D, hCYP-E, hCYP-40 and hCYP-NK [6]. The hCYP-A binds to CsA, and forms a ternary complex with calcineurin. The CsA-hCYP-A binding to calcineurin inhibits the phosphatase activity of calcineurin that results in a cascade of activities leading to the inactivation of T-cells [7].

Compared to human CYPs, very little is known about plant CYPs. The first plant CYPs were identified concurrently from tomato (Lycopersicon esculentum), maize (Zea mays), and oilseed rape (Brassica napus) [8]. Recently, with the availability of whole genome sequencing, the identification and characterization of plant CYPs has progressed substantially. However, compared to other organisms, the total number of plant CYPs in databases is still small, which suggests that many plant CYPs remain to be identified [9]. To date, Arabidopsis thaliana and rice (Oryza sativa) are the two plant species reported to have highest number of CYPs with 35 AtCYPs [10,11] and 28 OsCYPs [10,12], respectively. Among the identified AtCYPs, only 15 are characterized at the molecular level [11,13-21]. Their encoded proteins are found in the cytoplasm [17,19,20], endoplasmic reticulum (ER) [18,21], chloroplast [15,16], and nucleus [13]. An increase in the
expression of ROCI, an AtCYP, in response to light is associated with phytochromes and cryptochromes [19,22]. roc1 mutants display an early flowering phenotype [22], while gain-of-function mutations in ROCI reduce stem elongation and increase shoot branching [23]. In contrast, loss-of-function mutations in AtCYP40 reduce the number of juvenile leaves, with no change in inflorescence morphology or flowering time, and Arabidopsis plants with a defective AtCYP20-3 are found to be hypersensitive to oxidative stress conditions created by high light and high salt levels, and osmotic shock [24]. In addition to the AtCYPs having roles in various developmental processes, AtCYP59, a multi-domain CYP with a RNA recognition motif (RRM), regulates transcription and pre-mRNA processing by binding to the C-terminal domain of RNA polymerase II [13]. Collectively, these results show the roles of Arabidopsis CYPs in different cellular pathways, which necessitate further work to explore the function associated with each of the CYPs.

Compared to the Arabidopsis CYPs, little work has been done on the rice CYPs. Most of the studies on the latter show their roles in different types of stresses. OsCYP2 has been reported to have a role in different abiotic stress responses [25]. The expression of OsCYP2 is up-regulated towards salt stress, and its over-expression in rice enhances tolerance towards the salt stress. Similarly, overexpression in Arabidopsis and tobacco of the thylakoid-localized OsCYP20-2 increased tolerance towards osmotic stress, and to extremely high light conditions [26]. The expression levels of several other OsCYPs were increased by abiotic stresses such as desiccation and salt stress [10,12], indicating a critical role of OsCYPs during stress conditions.

Soybean (Glycine max [L.] Merr) is a legume plant belonging to the Papilionoideae family and is a rich source of protein, oil and plant natural products such as isoflavonoids. The soybean genome contains 56,044 protein coding loci located on 20 different chromosomes. Soybean has undergone two whole genome duplication events approximately 59 and 13 million years ago, as a result of which 75% of the genes have multiple copies [27]. Until now, not much was known about soybean CYPs except that a handful of CYP gene sequences had been deposited in the public databases. We present here a genome-wide identification of soybean CYPs, their phylogenetic analysis, chromosomal distribution, and structural and expression analysis. Our results indicate that soybean contains 62 CYPs, the largest family of CYP known to date in any organism. Further, the study describes a genome-wide segmental and tandem duplication during expansion of the GmCYP gene family.

**Results and discussion**

The soybean genome contains 62 putative GmCYPs

To identify all the members of the CYP gene family in soybean, a BLASTN search of the soybean genome database G. max Wm82.a2.v1 (http://phytozome.jgi.doe.gov/pz/portal.html#search?show=BLAST&method=Org_Gmax) was performed using the nucleotide sequence of a previously identified soybean CYP cDNA (GenBank: AF456323) as a query. This search identified 11 unique CYP genes. Each of the 11 CYP genes was used separately as a query sequence in the BLAST search of soybean genome database. This process was repeated until no new CYP gene was found. A total of 62 soybean CYPs, located on 18 different chromosomes, were identified and named GmCYP1 to GmCYP62 (Table 1). Of the 62 GmCYPs, 52 encoded a protein with a single cyclophilin-like domain (CLD) which is responsible for the cis/trans isomerization of the peptidyl prolyl peptide bond. The remaining 10 GmCYPs contained the CLD and additional domains. As shown in Figure 1, GmCYP8, GmCYP9, GmCYP16, and GmCYP17 each contained two tetratricopeptide repeats (TPRs) at the C-terminus. The TPR motif is degenerate in nature and consists of a 34 amino acid repeat unit typically arranged in tandem arrays [28]. Such TPR motif containing proteins mediate protein-protein interactions and often help in the assembly of multi-protein complexes. AtCYP40 (AGI: At2g15790), the Arabidopsis ortholog of GmCYP8, GmCYP9, GmCYP16, and GmCYP17, contains 3 TPRs and is involved in microRNA-mediated gene regulation [29]. Loss-of-function mutation of AtCYP40 showed a precocious phase change with reduced number of juvenile leaves, but no alteration of flowering time [20]. Moreover, the conserved amino acids of the TPR domain of AtCYP40 are required for the interaction between AtCYP40 and cytoplasmic Hsp90 proteins. This interaction is essential for the function of AtCYP40 in planta [29] suggesting a critical role for the TPR domain in microRNA-mediated gene regulation. Here we speculate a possibly similar function for the TPR domain in GmCYP8, GmCYP9, GmCYP16 and/or GmCYP17.

GmCYP20 and GmCYP35 contain three tryptophan-aspartate (WD) repeats at the N-terminus (Figure 1). WD repeat-containing proteins are involved in a wide variety of cellular functions, providing binding sites for two or more proteins, or fostering transient interactions with other proteins [30,31]. The Arabidopsis CYP AtCYP71 (AGI:At3g44600), contains 2 WD repeats [11] and functions in chromatin remodelling [14]. The very high sequence identity of AtCYP71 with GmCYP20 (87%) and GmCYP35 (83%) suggests that these two GmCYPs may play similar roles in soybean.

The sequence analysis identified two soybean CYPs, GmCYP56 and GmCYP59, having an RNA recognition motif (RRM) and zinc knuckle (ZK) at the C-terminus.
along with CLD at the N-terminus end (Figure 1). RRM is a small RNA binding motif of 90 amino acids and is conserved in a wide variety of organisms [32]. AtCYP59 (AGI:At1g53720), the Arabidopsis ortholog of GmCYP56 (80%) and GmCYP59 (66%) (Additional file 1) [11], contains an RRM motif, and is a transcriptional regulator [13] that interacts with the conserved sequence of unprocessed mRNA, leading to the inhibition of the PPIase activity in vitro [33]. Based on the functional association of AtCYP59 in transcriptional regulation, we speculate that the multi-domain soybean CYPs GmCYP56 and GmCYP59 possibly play a role in regulation of transcript in soybean via their RRMs.

Lastly, GmCYP18 and GmCYP19 contain a U-box at the N-terminus end of the protein. The U-box domain is highly conserved in some ubiquitin ligases and predicted to be a part of the ubiquitination machinery. Mammalian CYC4 and Arabidopsis AtCYP65 are the U-box containing CYP where the CYP domain is predicted to have chaperone function [11,34].

Of the 62 GmCYPs, it was ascertained that 13 contain a chloroplast transit peptide, 13 contain a signal peptide, 5 contain a mitochondrial targeting peptide, 10 contain a nuclear localization signal, and the remaining 21 are cytosolic (Table 1). Unlike Arabidopsis and rice CYPs [10], none of the soybean CYPs are predicted to be localized to the ER or golgi or plasma membrane. Only one secretory GmCYP, GmCYP39, is predicted for

| Gene name | Predicted transcript size (bp) | Predicted protein size (AA) | Predicted subcellular location | Domain information |
|-----------|-------------------------------|----------------------------|-------------------------------|--------------------|
| GmCYP1    | 973                           | 172                        | Cytosol                       | SD                 |
| GmCYP2    | 1595                          | 337                        | Secretory                     | SD                 |
| GmCYP3    | 1224                          | 172                        | Cytosol                       | SD                 |
| GmCYP4    | 1301                          | 373                        | Secretory                     | SD                 |
| GmCYP5    | 854                           | 172                        | Cytosol                       | SD                 |
| GmCYP6    | 1065                          | 204                        | Secretory                     | SD                 |
| GmCYP7    | 775                           | 172                        | Cytosol                       | SD                 |
| GmCYP8    | 2543                          | 616                        | Cytosol                       | MD                 |
| GmCYP9    | 354                           | 117                        | Cytosol                       | SD                 |
| GmCYP10   | 2595                          | 668                        | Nucleus                       | SD                 |
| GmCYP11   | 393                           | 130                        | Cytosol                       | SD                 |
| GmCYP12   | 1982                          | 493                        | Cytosol                       | SD                 |
| GmCYP13   | 1072                          | 175                        | Cytosol                       | SD                 |
| GmCYP14   | 582                           | 114                        | Cytosol                       | SD                 |
| GmCYP15   | 1611                          | 360                        | Cytosol                       | MD                 |
| GmCYP16   | 1233                          | 232                        | Secretory                     | SD                 |
| GmCYP17   | 1241                          | 360                        | Cytosol                       | MD                 |
| GmCYP18   | 1264                          | 236                        | Secretory                     | SD                 |
| GmCYP19   | 1380                          | 253                        | Chloroplast                   | SD                 |
| GmCYP20   | 1238                          | 236                        | Secretory                     | SD                 |
| GmCYP21   | 1062                          | 175                        | Cytosol                       | SD                 |
| GmCYP22   | 1087                          | 165                        | Cytosol                       | SD                 |
| GmCYP23   | 711                           | 236                        | Chloroplast                   | SD                 |
| GmCYP24   | 3138                          | 850                        | Nucleus                       | SD                 |
| GmCYP25   | 793                           | 164                        | Cytosol                       | SD                 |
| GmCYP26   | 2766                          | 167                        | Secretory                     | SD                 |
| GmCYP27   | 1253                          | 260                        | Chloroplast                   | SD                 |
| GmCYP28   | 1085                          | 226                        | Secretary                     | SD                 |
| GmCYP29   | 1200                          | 221                        | Cytosol                       | SD                 |
| GmCYP30   | 2532                          | 843                        | Nucleus                       | SD                 |
| GmCYP31   | 1745                          | 361                        | Cytosol                       | MD                 |
| GmCYP32   | 1836                          | 387                        | Chloroplast                   | SD                 |
| GmCYP33   | 1770                          | 361                        | Cytosol                       | MD                 |
| GmCYP34   | 1751                          | 439                        | Chloroplast                   | SD                 |
| GmCYP35   | 2576                          | 597                        | Nucleus                       | MD                 |
| GmCYP36   | 2532                          | 439                        | Mitochondria                  | SD                 |
| GmCYP37   | 1324                          | 225                        | Mitochondria                  | SD                 |
| GmCYP38   | 2292                          | 597                        | Nucleus                       | MD                 |
| GmCYP39   | 2922                          | 227                        | Chloroplast                   | SD                 |
| GmCYP40   | 2554                          | 616                        | Nucleus                       | MD                 |
| GmCYP41   | 1988                          | 445                        | Chloroplast                   | SD                 |
| GmCYP42   | 2374                          | 350                        | Chloroplast                   | SD                 |
| GmCYP43   | 1869                          | 263                        | Chloroplast                   | SD                 |
| GmCYP44   | 2403                          | 640                        | Nucleus                       | MD                 |
| GmCYP45   | 1822                          | 326                        | Secretory                     | SD                 |
| GmCYP46   | 1943                          | 230                        | Mitochondria                  | SD                 |
| GmCYP47   | 1645                          | 337                        | PM/Mitochondria#              | SD                 |
| GmCYP48   | 1489                          | 292                        | Mitochondria                  | SD                 |
localization in the mitochondrial inner membrane or plasma membrane. A search for ER retention signal did not locate KDEL or HDEL in any GmCYP. We also searched for CYP genes in the DFCI soybean gene index that contains 1,354,268 ESTs representing 73,178 TC sequences (http://compbio.dfci.harvard.edu/tgi/). Screening this database confirmed that 15 of the 62 GmCYP genes we identified were represented with 99-100% identity, and 100% coverage, implying that at least 25% of the GmCYPs are transcribed in various soybean tissues during normal growth and development, or in response to stress. Additionally, 33 GmCYPs displayed greater than 95% sequence identity with TC sequences in the soybean EST database, but with less than 100% query coverage. The lower sequence identities could be due simply to cultivar-specific sequence differences between the two databases, with the whole genome sequence originating solely from the cultivar Williams82 [27], and the DFCI soybean gene index comprising EST data from cDNA libraries of several different soybean cultivars, the number of transcribed GmCYPs in soybean can be expected to be more than 15. A list of all the soybean CYP gene family members and their detailed information is provided in Additional file 1.

Chromosomal distribution and phylogenetic analysis of soybean CYP genes

To determine the genome organization and distribution of GmCYPs on different chromosomes in soybean, a chromosome map was constructed. The results showed that the 62 GmCYPs are located on 18 different chromosomes. As depicted in Figure 2, the gene density per chromosome is uneven. Chromosome 11 and 19 contain the most, and show a relatively dense occurrence of CYP genes (6 each), whereas only one CYP (GmCYP54) is present on chromosome 14. No CYPs were found on chromosome 8 or 16. Most GmCYPs were localized towards the chromosome ends, and only GmCYP52, GmCYP49 and GmCYP54 were found near centromeres (Figure 2), suggesting the possibility of inter-chromosomal rearrangements, after genome duplication, between different soybean chromosomes.

Figure 1 Schematic representation of multi-domain GmCYPs. CLD, cyclophilin-like-domain; TPR, tetratrico peptide repeat; WD, tryptophan-aspartate repeat; RRM, RNA recognition motif; ZK, zink knuckle, U-box, U-box domain.

To explore the evolutionary relationship among soybean CYPs, a phylogenetic analysis was performed using their predicted amino acid sequences (Figure 3). As observed for many other genes in soybean, most of the predicted GmCYPs clustered together in pairs, reflecting the ancient genome duplication event [27,35]. Such events result in two copies of each gene which undergo shuffling and rearrangement, creating the potential for new diversity. There are four possible fates of duplicated genes [36]. First, one copy of the gene may be deleted during the course of evolution, resulting in loss of functional redundancy. Second, both copies of the genes may be retained and share the ancestral function, but gradually develop partially different functions (sub-functionalization). Third, one copy of the gene may acquire new function(s) during the course of evolution (neo-functionalization). Finally, there may be an intermediate stage between sub- and neo-functionalization. Which of these outcomes occur depends on the role of the specific gene in plant growth and development. Only those genes that are associated with critical functions for normal plant growth and development are retained, while others may be lost. The large number of CYPs present in the soybean genome thus likely reflects a combination of duplication and the important role of GmCYPs in soybean during normal growth and development, as well as in response to environmental stimuli.

Of the 62 GmCYPs, 54 are clustered in pairs (27 pairs) in the phylogenetic tree. The remaining 8 GmCYPs branched-off from the terminal branch of another pair of GmCYPs. This analysis further revealed that the multi-domain GmCYPs cluster together.

We also attempted to correlate the clustering of GmCYPs in the phylogenetic tree with their predicted subcellular localization. Interestingly, the GmCYPs predicted to be targeted to the same subcellular compartment grouped together as a separate clade. For example, GmCYPs with chloroplast transit peptide (GmCYP10, GmCYP23, GmCYP14, GmCYP28 and GmCYP12) formed a distinct clade on the tree. Another 4 chloroplast-localizing GmCYPs (GmCYP48, GmCYP52, GmCYP53, and GmCYP60) also formed a distinct clade, but in a different location on the tree. Similarly, the GmCYPs...
with nuclear localization signal (GmCYP27, GmCYP36, and GmCYP54, GmCYP43, GmCYP46) also formed separate clades on the tree (Figure 3). A similar pattern of gene clustering was observed for the GmCYPs predicted to localize in mitochondria or that were secretory.

By comparing the positions of GmCYPs on the chromosome map (Figure 2) and in the phylogenetic tree (Figure 3), an interesting grouping pattern was observed. If the GmCYPs were localized together on a chromosome, their paralogs were also found together on a different chromosome. For example, GmCYP4, GmCYP39, and GmCYP46 are clustered at the sub-telomere region of chromosome 4, and are most similar to GmCYP3, GmCYP38, and GmCYP43, respectively, which are clustered together in the sub-telomere region of chromosome 6 (Figure 2). Similarly, GmCYP36, GmCYP13, and GmCYP7 (chromosome 3) paired with GmCYP27, GmCYP15, and GmCYP11, respectively, from chromosome 19, and GmCYP18 and GmCYP32 (chromosome 11) paired with GmCYP19 and GmCYP31, respectively (chromosome 1), whereas GmCYP1 and GmCYP41 (chromosome 11) paired with GmCYP2 and GmCYP40 (chromosome 12). Moreover, GmCYP34 and GmCYP26, from chromosome 11, paired up with GmCYP24 and GmCYP25, respectively, from chromosome 18. These findings provide strong evidence for segmental duplication of chromosomal regions containing the GmCYPs, such as has been shown to play a vital role in the evolutionary generation of members of other gene families [37,38].

### Gene structures of GmCYPs

Analysis of the exon-intron structure of the GmCYP genes showed several variations (Figure 4). Six GmCYP genes (GmCYP1-GmCYP4, GmCYP6 and GmCYP7) contained no intron in their open reading frame (ORF). The number of introns varied from 1 to 13 in the ORFs of other GmCYP genes. The GmCYP5, GmCYP47, GmCYP50, GmCYP55 and GmCYP58 contained a single intron in their ORF while the largest numbers of introns were found in GmCYP56. The size of intron also varied considerably between different GmCYP gene family members with their size ranging from 39 bp (GmCYP5) to 9359 bp (GmCYP6) in the primary transcripts. Several other genes such as GmCYP22, GmCYP30, GmCYP34, GmCYP39-GmCYP42, GmCYP45, GmCYP49, GmCYP51- GmCYP53, GmCYP55- GmCYP57 and GmCYP59 contained introns larger than 4.0 kb in their ORFs. It has been suggested that the genome size may be correlated with intron size and that some elements of genome size evolution occurs within the gene [39]. However, in Gossypium spp., intron and genome size evolution are not coupled [40]. In the regions of low recombination, longer introns are selectively advantageous as they improve recombination and possibly counterbalance the mutational bias towards deletion [41]. A large-scale comparative analysis of intron positions among different kingdoms (animal, plant and fungus) identified a large number of positions that are likely to be ancestral [42]. Analysis of intron
sizes and positions in paralogs among *GmCYP* family did not show any specific pattern. However, in the majority of cases, the exon-intron numbers were similar in the genes that clustered together in the phylogenetic tree (Figure 3), for example, *GmCYP25* and *GmCYP26* or *GmCYP47* or *GmCYP20* and *GmCYP35*. The 5′ and 3′ untranslated regions (UTR) that border protein-coding sequences are important structural and regulatory elements of eukaryotic genes [43] and also contain large numbers of introns [44]. Out of 62 *GmCYPs*, 12 contained a single intron in the 5′UTR region while remaining *GmCYPs* consisted of intronless 5′UTR. The 3′UTR of two *GmCYPs*, *GmCYP16* and *GmCYP50*, were interrupted by a single intron whereas *GmCYP17* and *GmCYP44* contained 2 and 5 introns, respectively. The number of exon and intron in each *GmCYP* gene is shown in Additional file 2.

Expression analysis of *GmCYP* genes

To determine expression patterns of *GmCYP* genes, we used publicly-available genome-wide transcript profiling data of soybean tissues as a resource (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29163). The dataset contains RNAseq reads from soybean seeds across several stages of seed development (globular, heart, cotyledon, early-maturation, dry), and reproductive (floral buds) and vegetative (leaves, roots, stems, seedlings) tissues. As shown in Figure 5, most of the *GmCYP* genes showed distinct tissue-specific expression pattern. Out of the 62 *GmCYP* genes, 26 were expressed in the vegetative tissues whereas 34 were expressed in floral buds and different stages of seed development. Two *GmCYPs*, *GmCYP5* and *GmCYP6*, contained no sequence read in any of the soybean tissues included in the study. In addition, there were no EST or TC sequences in the DFCI gene index database with a perfect match to *GmCYP5* and

---

**Figure 3 Phylogenetic relationships of GmCYP proteins.** A Neighbor-joining tree was generated by MEGAS1.1 software [51] using putative amino acid sequence of 62 GmCYPs, and the tree was annotated using Interactive Tree of Life [52]. The numbers next to the branch shows the 1000 bootstrap replicates expressed in percentage. The solid line represents the real branch length and dotted lines added later for better visualization. The multi-domain CYPs are underlined and the predicted subcellular locations of GmCYPs are shown by colors as indicated.
Figure 4 (See legend on next page.)
GmCYP6 (Additional file 1). These evidences indicated that GmCYP5 and GmCYP6 are pseudogenes or expressed under special conditions or at specific developmental stages. The gene expression data revealed that the majority of GmCYPs (41%) were expressed in leaf tissue with the highest transcript accumulation level. Furthermore, it is interesting to note that the GmCYPs predicted to localize in the chloroplast were expressed in leaf tissues, suggesting their possible role in photosynthesis. Expression of several GmCYP genes in seed tissues during development indicates an important role of these genes in seed development.

Conclusions
Taken together, we have performed a comprehensive sequence analysis of soybean CYP genes (GmCYPs), and provided detailed information on them. Specifically, our results show that the soybean genome contains 62 CYP genes, the largest CYP gene family identified in any organism to date. The presence of predicted motifs, subcellular localization and their sequence homology with other identified CYPs from other organisms provided insight into their putative function. Results of the present study indicate a genome-wide segmental and tandem duplication during expansion of the GmCYP gene family.

Methods
Database search for CYP genes in soybean
To identify all the CYPs present in the soybean genome, the nucleotide sequence of the GmCYP1 (GenBank: AF456323) was used for a BLASTN [45] query against the new soybean genome database (Wm82.a2.v1) (http://phytozome.jgi.doe.gov/pz/portal.html) [46]. The newly identified sequences were subsequently used as queries to find other less similar GmCYPs. The chromosomal locations for all GmCYPs were obtained from the soybean genome database to draw the chromosomal map. The molecular weight for each GmCYP was calculated using ProtParam software [47] (http://web.expasy.org/protparam/), TargetP1 [48] (http://www.cbs.dtu.dk/services/TargetP/) and WoLF-PSORT [49] (http://wolfsporst.org/) were used to identify putative sub-cellular localization of the predicted protein sequences, and domain information was obtained from the soybean genome database [27]. To identify the transcribed GmCYPs in soybean, the coding sequence of each GmCYP was used as a query to BLAST against the soybean gene index (http://combio.dfci.harvard.edu/tgi/). The Tentative Contig (TC) sequences in the soybean gene index database were aligned with the corresponding GmCYP sequences to identify the percentage identity and coverage. Similarly, to find the GmCYP orthologs in Arabidopsis, the amino acid sequences of GmCYPs were used as queries to BLAST against the Arabidopsis protein database (http://www.arabidopsis.org/) [50].

Multiple sequence alignment and phylogenetic analyses
To investigate the phylogenetic relationships among GmCYP proteins, and their molecular evolution, a phylogenetic tree was generated. Multiple sequence alignment of the deduced amino acid sequences of all GmCYP proteins were aligned by Clustal X and the alignment was imported into MEGA5.1 to create a phylogenetic tree [51]. Neighbour-joining method was used with 1000 bootstrap replicates. The tree was exported into the Interactive Tree Of Life (http://itol.embl.de) for annotation and manipulation [52].

Expression analysis of soybean CYP genes
To determine the expression patterns of CYP genes in soybean tissues, the publically available transcriptome data (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29163) was used as a main source. The illumina sequencing of transcripts from ten different soybean tissues were downloaded from the NCBI database (http://www.ncbi.nlm.nih.gov/) with accession numbers SRX062325-SRX062334. After normalization of the dataset, the value of each gene was centered by subtracting the mean normalized value for each gene and scaled by dividing the centered value by the standard deviation of the gene following Eisen et al. [53]. The heatmap for GmCYP genes was generated in R using the heatmap.2 function from the gplots CRAN library (http://CRAN.R-project.org/package=gplots).

Availability of supporting data
Phylogenetic data (tree and data used to generate them) have been deposited in TreeBASE repository and is available under the URL http://purl.org/phylo/treebase/phylows/study/TB2:S16455.
Figure 5 Expression analyses of soybean CYP genes. The transcriptome data of soybean across different tissues and developmental stages were obtained from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29163) for heatmap generation. The color scale below the heat map indicates expression values, green indicating low transcript abundance and red indicating high levels of transcript abundance. Clustering of GmCYPs in (I) vegetative tissue or (II) reproductive or seed tissue is shown.
Additional files

Additional file 1: Soybean cyclophilin gene family.

Additional file 2: Number of exon/introns and splice variants in GmCYP genes.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

HRM designed and performed experiments, analysed data and wrote draft manuscript. PC conducted expression analysis of RNAseq data, and SD conceived and designed experiments, analysed data and wrote the final draft of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge Dr. Katherine Dobinson for the critical review of the manuscript and helpful comments. Ling Chen for technical assistance, and Alex Molnar for graphic designs. This research was funded by the Natural Sciences and Engineering Research Council of Canada’s Discovery Grant and Agriculture and Agri-Food Canada’s Crop Genomics Initiative grant to SD.

Received: 17 June 2014 Accepted: 9 October 2014

Published online: 29 October 2014

References

1. Galat A. Variations of sequences and amino acid compositions of proteins that sustain their biological functions: an analysis of the cyclophilin family of proteins. Arch Biochem Biophys 1999, 371(2):149–162.

2. Maruyama T, Suzuki R, Furutani M: Archeal peptidyl prolyl cis-trans isomerases (PPIases) update 2004. Front Biosci 2004, 9:1680–1720.

3. Hayano T, Takahashi N, Kato S, Maki N, Suzuki M: Two distinct forms of peptidylprolyl-cis-trans-isomerase are expressed separately in periplasmic and cytoplasmic compartments of Escherichia coli cells. Biochemistry 1991, 30(12):3041–3048.

4. Arevalo-Rodriguez M, Wu X, Hanes SD, Heitman J: Prolyl isomerases in yeast. Front Biosci 2004, 9:2402–2446.

5. Sykes K, Gettring MJ, Sandbrook J: Proline isomerases function during heat shock. Proc Natl Acad Sci USA 1993, 90(12):5853–5857.

6. Galat A: Peptidylprolyl cis/trans isomerases (immunophilins): biological diversity–targets–functions. Curr Top Med Chem 2003, 3(12):1315–1347.

7. Liu J, Farmer JD Jr, Lane WS, Friedman J, Weiseman I, Schreiber SL: Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. Cell 1991, 66:830–835.

8. Gasser CS, Gunning DA, Budelier KA, Brown SM: Structure and expression of cytosolic cyclophilin/peptidyl-prolyl cis-trans isomerase of higher plants and production of active tomato cyclophilin in Escherichia coli. Proc Natl Acad Sci USA 1990, 87(24):5919–923.

9. Obiya SO, Moriyama EN: Mining the Arabidopsis and rice genomes for cyclophilin protein families. Int J Bioinform Res Appl 2009, 5(3):295–309.

10. Trivedi D, Yadav S, Vaid N, Tuteja N: Genome wide analysis of Cyclophilin gene family from rice and Arabidopsis and its comparison with yeast. Plant Signal Behav 2012, 7(12):1633–1666.

11. Romano PG, Horton P, Gray JE: The Arabidopsis cyclophilin gene family. Plant Physiol 2004, 134(4):1268–1282.

12. Ahn JC, Kim DW, You YN, Seok MS, Park JM, Hwang H, Kim BG, Luan S, Park HS, Cho HS: Classification of rinse (Oryza sativa L japonica nipoimbore) immunophilins (FKBPs, CYPs) and expression patterns under water stress. BMC Plant Biol 2010, 10:253.

13. Gutierrez M, Barta A, Lorkovic ZJ: AtCyp59 is a multidomain cyclophilin from Arabidopsis thaliana that interacts with SR proteins and the chromatin domain of the RNA polymerase II. RNA 2006, 12(4):631–643.

14. Li H, Luan S: The cyclophilin AtCYP1P1 interacts with CAF-1 and LHP1 and functions in multiple chromatin remodelling processes. Molecular Plant 2011, 4(4):748–758.

15. Schubert M, Peterson UA, Haas BJ, Funk C, Schroder WP, Kieselbach T: Proteome map of the chloroplast lumen of Arabidopsis thaliana. J Biol Chem 2002, 277(10):8354–8365.

16. Lippuner V, Chou IT, Scott SV, Ettinger WF, Theg SM, Gasser CS: Cloning and characterization of chloroplast and cytosolic forms of cyclophilin from Arabidopsis thaliana. J Biol Chem 1994, 269(11):7863–7868.

17. Hayman GT, Memyk JA: The nucleotide and deduced amino acid sequences of a peptidyl-prolyl cis-trans isomerase from Arabidopsis thaliana. Biochim Biophys Acta 1996, 1219(1):536–538.

18. Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld JU, Selchert K, Koncz C, Jurgens G: A conserved domain of the arabidopsis GNom protein mediates subunit interaction and cyclophilin S binding. Plant Cell 2000, 12(3):343–356.

19. Chou IT, Gasser CS: Characterization of the cyclophilin gene family of Arabidopsis thaliana and phylogenetic analysis of known cyclophilin proteins. Plant Mol Biol 1997, 35(6):873–892.

20. Berardini TZ, Boillman K, Sun H, Poethig RS: Regulation of vegetative phase change in Arabidopsis thaliana by cyclophilin 40. Science 2001, 291(5512):2405–2407.

21. Jackson K, Soll D: Mutations in a new Arabidopsis cyclophilin disrupt its interaction with protein phosphatase 2A. Mol Gen Genet 1999, 262(4–5):830–838.

22. Trupkin SA, Mora-Garcia S, Casal JJ: The cyclophilin ROC1 links phytochrome and brassinosteroid sensitivity. Plant J 2012, 71(5):712–723.

23. Ma X, Song L, Yang Y, Liu D: A gain-of-function mutation in the ROC1 gene alters plant architecture in Arabidopsis. New Physiol 2013, 197(3):751–762.

24. Dominguez-Solis JR, He Z, Lima A, Ting J, Buchanan BB, Luan S: A cyclophilin links redox and light signals to cytokine biosynthesis and stress responses in chloroplasts. Proc Natl Acad Sci USA 2008, 105(42):16386–16391.

25. Ruan SL, Ma HS, Wang SH, Fu YP, Xin Y, Liu ZW, Wang F, Tong JX, Wang SZ, Chen HZ: Proteomic identification of OsCYP2, a rice cyclophilin that confers salt tolerance in rice (Oryza sativa L) seedlings when overexpressed. BMC Plant Biol 2011, 11:34.

26. Kim SK, You YN, Park JC, Jeoung Y, Kim BG, Ahn JC, Cho HS: The rice thylakoid luminal cyclophilin OsCYP20-2 confers enhanced environmental stress tolerance in tobacco and Arabidopsis. Plant Cell Rep 2012, 31(4):417–426.

27. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umewa Z, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, et al: Genome sequence of the palaeopolyploid soybean. Nature 2010, 463(7278):178–185.

28. Bishat GL, Lasle M: The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays 1999, 21(11):932–939.

29. Earley KW, Poethig RS: Binding of the cyclophilin 40 ortholog SQUINT to Hsp90 protein is required for SQUINT function in Arabidopsis. J Biol Chem 2011, 286(44):38184–38189.

30. van Nocker S, Ludwig P: The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. BMC Genomics 2003, 4(1):50.

31. Strimminger CU, Petalas E, Russell BE, Muller CW: WD40 proteins propel cellular networks. Trends Biochem Sci 2010, 35(10):565–574.

32. Bandziulis RJ, Swanson MS, Dreyfuss G: RNA-binding proteins as developmental regulators. Genes Dev 1989, 3(4):431–437.

33. Bannikova O, Zywicki M, Marquez Y, Skrahina T, Kalyna M, Barta A: Identification of RNA targets for the nuclear multidomain cyclophilin atCyp59 and their effect on PPIase activity. Nucleic Acids Res 2013, 41(7):1783–1798.

34. Cyr DM, Hofhiedl F, Patterson C: Protein quality control: U-box-containing E3 ubiquitin ligases join the fold. Trends Biochem Sci 2002, 27(7):368–375.

35. Li X, Dhaubhadel S: Soybean 14-3-3 family genes: identification and molecular characterization. Planta 2011, 233(1):569–582.

36. Charon C, Arnaud C, Bertiainen Q, Therau V, Henry Y: Gene duplication within the Green Lineage: the case of TEL genes. J Exp Bot 2012, 63(14):5061–5077.

37. Cannon S, Mitra A, Baumgarten A, Young N, May G: The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol 2004, 4(1):10.
39. McLysaght A, Enright AJ, Skrabanek L, Wolfe KH: Estimation of synteny conservation and genome compaction between pufferfish (Fugu) and human. Yeast 2000, 17(1):22–36.
40. Wendel JF, Cronn RC, Alvarez I, Liu B, Small RL, Sennchina DS: Intron size and genome size in plants. Mol Biol Evol 2002, 19(12):2346–2352.
41. Carvalho AB, Clark AG: Genetic recombination: intron size and natural selection. Nature 1999, 401(6751):344–344.
42. Fedorov A, Merican AF, Gilbert W: Large-scale comparison of intron positions among animal, plant, and fungal genes. Proc Natl Acad Sci USA 2002, 99(25):16128–16133.
43. Wilkie GS, Dickson KS, Gray NK: Regulation of mRNA translation by 5′- and 3′-UTR-binding factors. Trends Biochem Sci 2003, 28(4):182–188.
44. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S: Structural and functional features of eukaryotic mRNA untranslated regions. Gene 2001, 276(1–2):73–81.
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215:403–410.
46. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazio J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS: Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 2012, 40(D1):D1178–D1186.
47. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel R, Bairoch A: Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook. Edited by Walker JM. Totowa, NJ: Humana Press; 2005:571–607.
48. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protocols 2007, 2(4):953–971.
49. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K: WoLF PSORT: protein localization predictor. Nucleic Acids Res 2007, 35(Web Server issue):WS85–WS87.
50. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E: The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 2008, 36(suppl 1):D1009–D1014.
51. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28(10):2731–2739.
52. Letunic I, Bork P: Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 2011, 39(Web Server issue):W475–W478.
53. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998, 95(25):14863–14868.

doi:10.1186/s12870-014-0282-7
Cite this article as: Mainali et al.: Genome-wide analysis of Cyclophilin gene family in soybean (Glycine max). BMC Plant Biology 2014 14:282.