Zmcyp genes identification and characterisation

Methods:

- Gene structure, chromosomal locations and promoter analysis
- Conserved motifs and phylogenetic analysis of ZmCYPs
- Gene expansion and synteney
- Gene expression

Methods:

Identification and Charechterization of ZmCYP genes:

Identification of all ZmCYP genes in the Zea mays genome was done by performing nucleotide Blast search using a reported ZmCYP sequence as a query (NCBI accession number: M55021) against maize genome database (http://www. phytozome. net). Each of the newly identified ZmCYP's was again used as a query sequence until no new ZmCYP sequence found. All non redundant sequences were again checked for the presence of CLD domain using Pfam data base. Various physicochemical charecheterstics like M. Wt, PI and signal peptides examined using appropriate tools of SIB bioinformatics portal (http://www. expasy. org). Cellular/sub cellular targeting sites predicted using WolfP sort, Predator, Target P1 servers. Physical location of a gene, domain information obtained from Maize Genome Database and Phytozome database.

Conserved motifs and phylogenetic analysis of ZmCYP genes:

The conserved motif analysis of ZmCYP proteins was performed using the Multiple Expectation Maximization for Motif Elicitation (MEME) program. ZmCYP protein sequences were used for determining their phylogenetic

relationships. Multiple sequence alignment was done using clustal W with gap opening penalty10 and gap extension penalty 0. 1. Then the aligned sequences subjected to MEGA 6. 0 software for unrooted phylogenetic tree construction using Neighbor joining statistical method with the parameters like pairwise gap deletion and 1000 bootstrap replications.

Cis elements:

To identify cis elements we isolated upstream sequences from +1 to -1000 bp of all 47 ZmCyp's from Phytozome database and subsequently they subjected for analyzing for the presence of stress responsive elements using Plant Care database. Intron, exon structures were determined using Gene structure display server2. 0.

Results and Discussion:

To identify all the members of zmCYPs we used ZmCyp ( ) gene as a query sequence for blastn search against zeamays genome database (Phytozome). We identify 11 non redundant sequences against this search. We subsequently used all this 11 non redundant sequences for blastn against maize genome until unless no new cyp sequences found. As a result of theses we have obtained a total of 47 non redundant sequences. Naming was based on ascending order of their physical location on a chromosome. Further, we also reconfirmed all the 47 putative Zmcyp peptide sequences for the presence of CLD domain, a unique domain of this class of the family responsible for peptidyl prolyl cis trans isomerization activity by executing in pfam data base. To date this is the second largest cyp family identified after soybean. Further, all these 47 deduced polypeptides of Zmcyps were

analyzed for their amino acid length, M. Wt, PI. Interestingly amino acid length varied from 7. 6KDa to 131. 8KDa, similarly M. Wts of ZmCYP peptides varied from 65 to 1206. Gene locaton, transcript size, no of alternative transcripts, peptide length, M. Wt and PI were summarized in Table 1.

Results and Discussion:

Identification and Charechterization of ZmCYP genes:

A total of 47 putative non redundant ZmCYPs identified by performing BLASTN search against maize genome database. A known cDNA ZmCYP sequence (NCBI accession number: M55021) used as a query sequence for performing a BLASTN search. As a result, 10 sequences identified in the first round of search, which were subsequently used as a query sequences for new rounds of BLASTN search. These BLASTN search performed until no new sequences found. Further, all the 47 zmcyps cross checked for the presence of CLD (Cyclophylin like domain) domain with the help of Pfarm database. This is the second largest cyclophilin family identified in plants after soybean (62 CYPs). Deduced polypeptides of all the 47 zmcyps subjected for various Expasy tools for prediction of their various biochemical properties like number of amino acids, Molecular weight, Isoelctric point (pI) and GRAVY values. Peptide lengths and pI values varied from 65 amino acids to 1206 amino acids and 4. 49 to 11. 61 respectively. Similar results were also observed in rice ( ), Soybean ( ), Arabidopsis ( ). Most of the Zmcyps showed -ve GRAVY values indicating that, they were prominently involved in many cellular functions as corroborated with other reports ( ). Of all 47 zmcyp peptides 36 contain CLD domain and the remaining 9 contains additional

domains along with CLD domain. multidomains peptides which contain both domains like TPR (CYP32, 33), RNA recognition motif and Zinc knuckle (CYP31), MSP domain (CYP25). Out of 47 zmcyps 23 has chloroplast targeting signal peptides, 8 have nuclear localization signals, 3 have KDEL sequences, 2 have mitochondrial targeting signals, 1 contain golgi complex retion signals and the remaining 9 were cytosolic.

Gene structure, Chromosomal locations and promoter analysis:

Distribution of Zmcyps occurred in all the 10 chromosomes. However, percentage of distribution is uneven. For example 1 st chromosome contain the highest percentage of distribution (20. 93) with 9 zmcyps followed by 18. 6% on 5 th chomosome, 13. 9% on 2 nd , 6 th and 7 th chromosomes, 9. 3% on 8 th and 9 th chromosomes, 6. 9% on 3 rd and 4 th chromosome and lowest percentage (2. 3%) on 10 chromosomes. Naming has given based on ascending order of their physical location on chromosomes as ZmCYP1 to ZmCYP47. It was also observed that very less number of cyps located on centromeric regions and most of them were found in chromosomal ends which indicate that possible role of segmental duplications in the course of evolution. Further, the expansion of zmcyp family was discussed clearly in gene expansion and synteny sub headings. The schematic diagram of zmcyps distribution on each chromosome is depicted in Fig.

Gene structure analysis revealed that, varying lengths of zmcyps has observed. For example zmcyp5 has contained least no of base pairs (412) and zmcy28 has contained highest number of base pairs (32, 725). These vast variations are due to varying number and lengths of introns. For

instance zmcyp 7, 22 and 26 contain highest number (12) of introns in their ORF regions whereas, zmcyp 1, 6, 8 and 9 contains no introns in ORF. 5’ and 3’ UTRs are considered to be as important regulatory elements, out of 47 zmcyps only zmcyp14, 26 and 33 contains one intron in 5’ UTR region and the remaining were intron less 5’ UTRs. However, none of the zmcyps contained introns in the 3’UTR regions. These variations are may be due to gain or loss of exons/introns or exonization/pseudo exonization or Insertion/deletions (Xu et al, 2011) in course of evolution. This gene structural divergence may be one of the reasons for structural and functional diversity of different Zmcyp peptides.

Promoter analysis:

5’ upstream positions (500-1000bp) from translational start site considered as promoter sequences. Various types of cis regulatory elements are present in this region. In the present study 47 zmcyp promoter sequences retrived from maize genome database and subjected for analysis by PlantCare data base for finding out cis elements. These cis elements are considered as switches for transcription regulation.