In silico analysis for the presence of HARDY an Arabidopsis drought tolerance DNA binding transcription factor product in chromosome 6 of *Sorghum bicolor* genome

Arun K. Shanker, M. Maheswari, S.K.Yadav and B.Ventateswarlu

Department of Crop Sciences, Central Research Institute for Dryland Agriculture (CRIDA)

Santosh Nagar, Hyderabad - 500 059, India

Water insufficiency, due to quickly escalating world population and the additional increase in water use for various other non agricultural purposes has put immense pressure on sustainable world crop production that uses a high percentage of quality water available in the planet. Water use efficiency (WUE), if increased in terms of biomass produced at the expense of transpiration measured, will have a positive effect on crop yield, under water limiting conditions. Expression of the Arabidopsis HARDY (*hrd*) DNA binding transcription factor (555 bp present on chromosome 2) has been shown to increase WUE in rice by Karaba et al 2007 (PNAS, 104:15270–15275). The *Sorghum bicolor* genome project was initiated through Community Sequencing Program (CSP) by a consortium led by Paterson et al has been completed and the version 1 of the genome is released publically

(http://www.phytozome.net/sorghum). Detailed annotation of the genome is under progress as a collaborative effort worldwide. The presence of at least 60 percent similarity in nucleotide, mRNA or protein product of Arabidopsis *hrd* gene in sorghum offers excellent possibility of *cis/trans* genetic improvement of *Sorghum bicolor* for higher yields in the dry and water scarce rainfed areas of the world. In this study we took up detail analysis of the complete sorghum genome for the similarity/presence of either DNA, mRNA or protein product of the above mentioned Arabidopsis *hrd* DNA binding transcription factor. We conducted i) BLASTN- nucleotide query to sorghum nucleotide database, ii)TBLASTX – translated nucleotide query sequence into protein sequences in all 6 reading frames compared to sorghum nucleotide database translated on all six reading frames and iii) TBLASTN- protein query to translated 6 frame sorghum nucleotide database. Our results indicate that with a target of 3304

sequences and 738540932 total letters, there is a poor similarity of only 24 nucleotides in chromosome 1 with *hrd* gene. On the other hand, mRNA translated nucleotide query sequence produced matches with 11 identified regions in chromosome 1, 4, 6, 3, 10, 5, 2, 9, 7 and 8 among which 1, 4 and 6 has significant E values of less than 1.6e-18. Chromosome 6 showed a sequence match of 61.5 percent positive between 61 and 255 mRNA residues of the query region. Further confirmation was obtained by TBLASTN which showed that chromosome 6 of the sorghum genome has a region between 54948120 and 54948668 which has 80 amino acid similarities out of the 185 residues Arabidopsis *hrd* DNA binding transcription factor. A homology model was built (figure 1 Above) with the positive match protein sequence and 20 flanking upstream and downstream sequences. SIM alignment was done and all templates with sequence identities above 25% were selected and model was generated from exPDB database scan by ProModII method. Missing side chains and deleted loops were added and energy minimization was done after adding hydrogen atoms. The model was verified using Anolea, Gromos and Verify3D. Scanning the motif for possible activation sites revealed that there was a protein kinase C phosphorylation site between 15th and 20th residue. The study indicates the possibility of the presence of a DNA binding transcription factor in chromosome 6 of *Sorghum bicolor* with 60 percent similarity to that of Arabidopsis *hrd* DNA binding transcription factor.