Plant Bulletin, Vol 4, (1) 2025: 159-171 ISSN: 2959-3379 (Print), 2959-3387 (Online)

Research Article

Genome-wide identification and characterization of indeterminate domain (*IDD*) gene family in sunflower (*Helianthus annuus*)

Ayesha Khan[1,2], Muhammad Hammad Nadeem Tahir[2], Shahid Iqbal[1], Muhammad Nabeel[1,2], Fatima Javeria[1,2], Zulqurnain Khan[1,2*], Shoaib Ur Rehman[1,2*]

[1]*SINO-PAK Joint Research Laboratory, MNS-University of Agriculture-Multan, Pakistan*
[2]*Institute of Plant Breeding and Biotechnology, MNS-University of Agriculture-Multan, Pakistan*
Correspondence: shoaib.rehman@mnsuam.edu.pk; zulqurnain.khan@mnsuam.edu.pk

ABSTRACT

Indeterminate Domain (*IDD*) represents a large and evolutionary conserved gene family in plants playing critical roles in growth and seed development. Despite its agronomic significance as oilseed crop, a systematic analysis of *IDD* gene family is sunflower (*Helianthus annus*) has been lacking. In this work, 78 *IDD* genes were identified across six species, including 21 in *H. annus*. Phylogenetic analysis grouped *IDDs* genes into four distinct sub-groups, showing a closer evolutionary relationship between *H. annus*, *Theobroma cacao*, *Zea mays*, *Arabidopsis thaliana*, *Physcomitrella patens*, *Selaginella moellendorffii* and *Oryza sativa*. Gene structure analysis showed that most *HaIDD* genes have two intron and three exons. Conserved domain analysis identified ten different motifs related to abiotic stress tolerance. Sequence logo analysis further confirmed the high conservation of IDD family in *H. annus* and *A. thaliana*. The intracellular localization prediction showed that all HaIDD proteins were present inside nucleus. Upstream region analysis showed the existence of *cis*-acting element associated to abiotic stress responses, proposing a probable role for *HaIDDs* in stress tolerance sunflower breeding. The selection pressure was calculated by *Ka*/*Ks* ratio which indicated purifying selection because the value significantly less than one. Gene expression analysis demonstrated differential expression of *HaIDDs* across leaves, stamen bract, pollen, pistil, corolla, ligule, ovary and seeds. Especially, *HaIDD3* showed higher expression in leaves and seeds suggesting potential role in photosynthesis and seed development. Overall, these results indicate that *HaIDDs* are mainly involved governing plant developmental processes and abiotic stress responses in sunflower.

Keywords: *Helianthus annuus*, transcription factors, phylogenetic analysis, *semi qPCR*, Indeterminate Domain.

INTRODUCTION

Genome-wide characterization of transcription factors holds key importance in modern plant breeding. These transcription factors govern many biological processes in many plant species.

INDETERMINATE (*IDD*) is also an important transcription factor reported its role in growth and development in various crops. *IDD* genes contain C2H2 zinc finger domain which plays crucial functions in evolutionary process, growth, floral arrangements, under abiotic stress (salinity, drought, cold) and leaves, fiber and seed development (Ali *et al.* 2019; Kumar *et al.* 2019).

Sixteen *AtIDD* gene have been characterized in *A. thaliana* (Colasanti *et al.* 2006). Among these members, *AtIDD*-3, *AtIDD*-8, and *AtIDD*-10 have important function in the root development. *AtIDD*-3 is also link with gibberellic acid signaling pathway and on its binding with promoter, it plays role in the development of plant. *AtIDD*-3 interacts with DELLA and SCL3 proteins for the regulation of the genes in downstream region (Fan *et al.* 2017). *AtIDD*14, *AtIDD*-15, and *AtIDD*-16 are involved in the gene regulation which are related to the biosynthesis of auxin and the transport it in shoot Cui *et al.* (2013) gravitropic responses is also controlled by *AtIDD*-15. So, these play roles in the production of auxin.

Article History

Received: March 30, 2025, Accepted: May 06, 2025,
Published: June 30, 2025.

*AtIDD*8 associate with the *SnRK*1 and control the sugar metabolism during flowering, *AtIDD*11 is associate with the leaf arrangement (Seo *et al.* 2011, Jeong *et al.* 2015). Similarly, Zhang *et al.* 2020, characterized 15 *IDD* genes in *O. sativa*. *RIDI* (Rice Indeterminate 1) enact as master switch during transformation of vegetative to reproductive stage and act as activator for flowering (Matsubara *et al.* 2008). *OsIDD*2 function as a suppressor for the transcription of genes associate with biosynthesis of lignin and metabolism of sucrose (Huang *et al.* 2018). Whereas *OsIDD*3 act as an activator for cold stress responses by regulating expression of C-repeat binding factors (CBF1). It is reported that *OsIDD*14 link with shoot gravitropism and act as a regulator in signaling pathway of IAA and BRs. OsIDD10 play function in the roots for ammonium absorption. Three *IDD* genes have been identified in maize. *ZmIDI*, regulates the flowering during developmental stage and control the flowering from vegetative to flowering stages (Coneva *et al.* 2012). *ZmIDD*9 and *ZmIDDveg*9 associated with the endosperm formation and control the transcription of some storage proteins (Yi *et al.* 2015; Gontarek *et al.* 2016). Sixty-five IDD gene members have also been characterized in *Gossypium hirsutum* L. Spatiotemporal expression of *IDDs* in *G. hirsutum* suggest that these genes are high expressive in ovule and are up-regulated in response to drought, low-temperature, heat and NaCl stimuli (Ali et al., 2019).

Helianthus annuus (sunflower) is very important oilseed crop (oil contents 35 to 55%) and domesticated worldwide (Kaya *et al.* 2006). Like other crops, sunflower growth & development also faces biotic and abiotic stress factor challenges. The function of IDDs is well-characterized in *Arabidopsis*, rice, maize and cotton. However, investigation of *IDDs* in sunflower remained subtle. The current work displays the systematic analysis of *IDDs* in *H. annus*. A total of 21 *HaIDD* members were characterized to explore the exon-intron structures, scan-site analysis, conserved motif analysis, phylogenetic analysis, sequence logo, chromosome locations, *Ka/Ks* values, gene expression analysis in responses PEG conditions. This work will assist in understanding the basic functional mechanism of *HaIDDs* in sunflower.

MATERIAL AND METHODS

Sequence Retrieval and Sub-Cellular Localization

Proteome for *A. thaliana* was retrieved from TAIR data base (https://www.arabidopsis.org/index.jsp). Arabidopsis IDD protein sequences were employed as queries for IDD protein identification in rest of the species using Local BLASTp and HMMER search program. The reference proteome for *H. annus* (version 1.2) was retrieved from https://phytozome-next.jgi.doe.gov/ (Badouin et al., 2017). Protein sequences of *O. sativa* (version 7.0), *P. patens* (version 3.3), *T. cacao* (V-1.1) and *S. moellendorffii* (V-1.0) protein sequences were also retrieved from Phytozome web-database. Sequence search strategy having local BLASTp search with E-value cutoff of ≤ 1e-10 was used followed by HMMER search using PFAM domain model. SMART (http://smart.emblheidelberg.de/) was used to validate the predicted IDD protein sequences extracted through Local BLASTp (Letunic *et al.* 2015). The intra-cellular localization of *HaIDDs* were predicted by using web-database CELLO v.2.5 (http://cello.life.nctu.edu.tw/).

Phylogeny and Homology Assessment

IDD proteins from *H. annuus, T. cocao, S. moellendorffii*, *P. paten, A. thaliana* and *O. sativa* were aligned multiple times using ClustalW program followed by the construction of phylogenetic tree using Maximum Likely-Hood (ML) with 1000-bootstrap (Kumar et al., 2016).

Gene Structure, Protein Motif and *Cis*-Acting Element Analysis

Exon-Intron patterns were predicted using Gene Structure Display Server 2.0 (Guo et al., 2007). MEME tool was employed to perform protein motif as suggested by Bailey *et al.* (2006) while *cis*-acting elements were detected using The PlantCare database as suggested by Lescot et al., (2002).

Chromosomal Distribution, Synteny analysis, Sequence Logos and *Ka/Ks* values

Sunflower assemblies were used to generate chromosomal length file. Rates of non-synonymous and synonymous substitutions were computed using Tbtools (Chen et al., 2020) on the resulting groups of two pairs. *Ka*/*Ks* ratio was explored to observe the pattern of codon selection during evolution. Collinear pairs of *HaIDD* genes fabricated by MCscan were employed to generate collinearity map of *IDDs* using CIRCOS (Krzywinski et al., 2009). For sequence logos analysis, HaIDD and AtIDD protein sequences were aligned using ClustalW (Thompson et al., 2003) and graphical representation was generated using WebLogo (Crooks et al., 2004).

RNA SEQ Data and Gene Expression Analysis

High-throughput microarray data was downloaded from https://www.sunflowergenome.org/. *HaIDDs* expression in various tissues were observed and Tbtools was used to present the expression pattern in heatmap format. RNA was extracted from the seeds and seedling leaves of 20 days old seedling using GeneJET™ Plant RNA Purification Mini Kit. Extracted RNA was transformed into cDNA using Thermo Scientific RevertAid First-Strand cDNA synthesis kit. Absorbance of each extracted sample

was measured at 230 and 260 nm. Primers were designed by considering suitable region keeping in view the basic principles of designing primers to study gene expression. Semi q-PCR with *HaActin* as reference gene was performed. Thermal cycling conditions were; denaturation at 95°C for 180 sec (1 cycle), denaturation at 95°C for 30 sec, annealing at 60°C 30 sec (30 cycles), Final extension at 72°C 300 sec (1 cycle). Agarose gel of ~2.5% concentration was prepared, and PCR product was loaded to visualize the bands.

RESULTS

Identification of IDD Genes, Sub-Cellular Positioning and Scan Site Analysis

Seventy-eight *IDDs* were identified in the targeted six species, including dicots (*H. annus*, *T. cacao* and *A. thaliana*), monocots (*O. sativa*), mosses (*P. paten*) spike mosses (*S. moellendorffii*). Among these, 21 genes were identified in *H. annus*, 15 in *T. cacao*, 16 genes were identified in *A. thaliana*, 14 in *O. sativa*, 7 in *P. paten*, and 5 in *S. moellendorffii*. A higher number of *IDD* genes were detected in *H. annuus* as compared to *T. cacao*, Arabidopsis, rice and mosses highlighting expansion and effect on *HaIDD* genes in *H. annus*. Subcellular localization showed that the *IDD* gene family was highly expressed in nucleus (Table 1).

Scan-site analysis was performed to check the similarity and dissimilarity of *IDD* members among Arabidopsis and rice with sunflower. Most of the amino acids were polar and hydrophobic amino acids (Supplementary table 1).

The alignment between *IDD* members of Arabidopsis, rice and sunflower showed that there was similarity in the *IDD* members of these species (Supplementary Fig. 1). The similarity was from 90 to 280 amino acid region. There was very low similarity in the region of C-terminus and N-terminus.

Phylogenetic Analysis and Sequence Logos

Phylogenetic tree was constructed among 78 *IDDs*. For presentation purpose, the prefixes such as Pp, Sm, At, Th, Os and Ha were utilized to represent the *IDD* proteins from *P. patens*, *S. moellendorffii*, *A. thaliana*, *T. cacao*, *O. sativa* and *H. annuus*, respectively. Phylogenetic analysis grouped 78 IDDs into four sub-groups having 39 (sub-group-1), 13 (sub-group-2), 16 (sub-group-3) and 9 (sub-group-4) IDDs (Fig. 1A). Phylogenetic analysis demonstrated the strong relationship among sunflower, cacao, rice and Arabidopsis *IDDs*, as genes from these species were closely grouped with each other even in different clades and sub-groups. It also indicated that *T. cocao*, sunflower and *A. thaliana* were derived from same ancestors. However, one *IDD* gene (*OsIDD14*) did not fall in any clade showing its potential unique function in the evolution from aquatic to terrestrial individual. In green color clade, there was only 1 gene from fern and mosses which indicated that the genes had evolved after the parting of moss and fern.

To explore the homologous domain sequences, Arabidopsis and sunflower (*H. annuus*) protein sequences were aligned to create sequence logos. The amino acid residues (AARs) pattern in *HaIDD* and *AtIDD* domain was substantially conserved between two species of plants. The findings also suggested that most IDD proteins have a similar metabolic activity and may target comparable components in downstream gene regulation. The conserved regions in Arabidopsis and sunflower also showed that these organisms also linked closely with each other (Fig. 1B). Table 2 catalogs the predicted *cis*-acting elements found in the promoter regions of *HaIDDs* indicating their function in stress responsiveness, light signaling, hormonal responses and plant development.

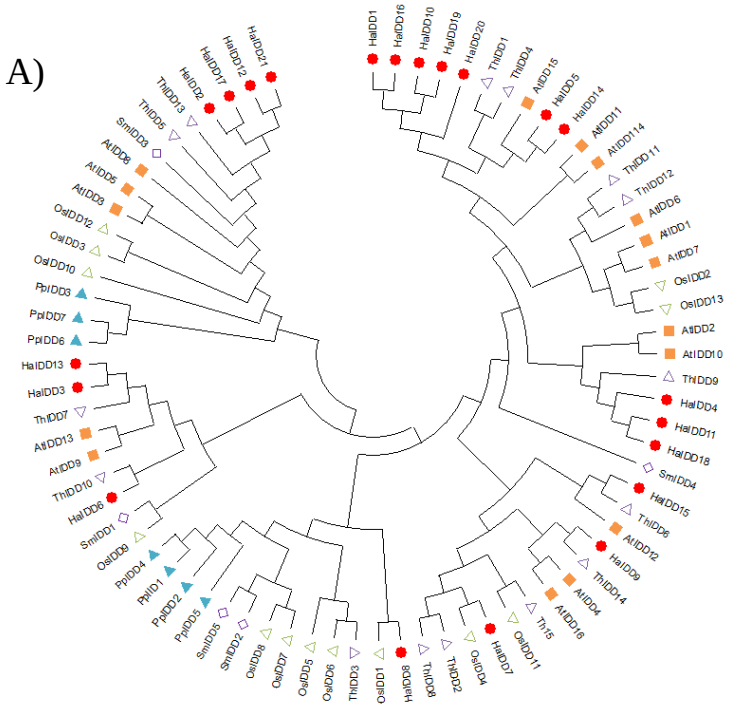
Gene structure, *cis*-acting Element and Motifs Analysis

Exon-intron structure, *cis*-acting and motifs analysis was conducted to explore variation and potential protein function among *HaIDD* members. *HaIDDs* showing similar exon-intron structure and distribution patterns were grouped in same sub-group. *HaIDDs* had three exons and two introns, but very few had four exons and three introns. Length of exons and introns was almost similar among *HaIDD* genes (Fig. 2). *HaIDD*6, *HaIDD*7 and *HaIDD*19 had four exons and three introns, while the other 18 genes have three exons. The genes had 4 exons that were closely related as they had same gene structure (Fig. 2).

The online MEME tool was exercised to search conserved motifs of HaIDD with the parameter settings: maximum number of motifs, twelve; maximum and minimum width of motifs, twenty and five. Motifs were manually corrected and analyzed.

Table 1. Sub-cellular localization values of IDD gene family in nucleus, plasma membrane, cytoplasm and mitochondria of sunflower**.**

Genes	Nucleus	Plasma membrane	Extracellular	Cytoplasmic	Mitochondria
*HaIDD*1	8.93	0.7	0	0	0
*HaIDD*2	8.44	1	0.02	0	0
*HaIDD*3	8.82	0.74	0	0	0
*HaIDD*4	8.7	0.73	0	0	0
*HaIDD*5	8.99	0.68	0	0	0
*HaIDD*6	8.32	0.99	0.1	0	0.01
*HaIDD*7	8.22	0.88	0.23	0	0.09
*HaIDD*8	8.53	1.03	0	0	0
*HaIDD*9	8.91	0.7	0	0	0
*HaIDD*10	8.85	0.7	0.01	0	0.01
*HaIDD*11	9.05	0.69	0	0	0
*HaIDD*12	8.88	0.69	0	0	0
*HaIDD*13	8.38	1.04	0.02	0	0
*HaIDD*14	8.41	I.06	0	0	0
*HaIDD*15	8.37	0.99	0.02	0	0
*HaIDD*16	8.26	1.12	0	0	0
*HaIDD*17	8.88	0.68	0	0	0
*HaIDD*18	8.98	0.69	0	0	0
*HaIDD*19	8.01	1.11	0	0	0.03
*HaIDD*20	8.84	0.65	0.02	0	0.03
*HaIDD*21	8.14	1.02	0	0	0.04



B) ***HaIDD*** **Sequence logo**

AtIDD **Sequence logo**

Figure 1. (A) Phylogeny of IDD members in sunflower and other species. (B) AtIDD and HaIDD sequence logos alignment. AARs shared by two Arabidopsis and sunflower are substantially conserved.

Table 2. Summary of *cis*-acting elements in the promoter regions of *HaIDDs*

Gene	LR	Hormone Responsive Elements				Stress Responsive Elements			Metabolic Regulators		
		MeJA-R	AR	GR	LAR	ABR	LTR	SAR	MYB-DI	MYB-FBR	*cis*-ZMR
HaIDD1	**Present**	1	-	1	-	1	1	1	1	-	-
HaIDD2	**Present**	-	-	-	-	1	-	1	-	-	-
HaIDD3	**3**	1	-	-	-	-	-	-	1	-	1
HaIDD4	**Present**	4	-	1	-	1	1	2	-	-	-
HaIDD5	**9**	2	-	1	-	-	-	2	-	-	1
HaIDD6	**18**	1	1		-	1	1	1	-	1	-
HaIDD7	**Majorly present**	2	1	1	-	1	1	-	-	-	-
HaIDD8	**10**	-	-	-	1	1	-	-	1		-
HaIDD9	**Present**	-	-	-	-	-	-	-	-	-	-
HaIDD10	**Present**	1	-	-	-	1	-	-	-	-	-
HaIDD11	**Present**	1	-	-	-	-	-	1	-	-	1
HaIDD12	**Present**	1	-	-	-	1	-	1	-	-	-
HaIDD13	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD14	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD15	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD16	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD17	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD18	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD19	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-

HaIDD20	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-
HaIDD21	**Majorly present**	Existent	Existent	-	-	Existent	Existent	Existent	-	-	-

LR (Light Responsive); MeJA-R (Methyl Jasmonate Responsive); AR (Auxin Responsive); GR (Gibberellin Responsive); ABR (Abscisic Acid); LTR (Low Temperature Responsive); SAR (Salicylic Acid); MYB-DI (MYB-Binding site intricate in drought inducibility); MYB-FBR (MYB-Binding site intricate in flavonoid biosynthesis); *cis*-ZMR (Zein Metabolism Regulator)

The motif pattern was nearly identical in all members of the IDD gene family. *HaIDD*1 had motif 5 (IPNPNPNSGLPNKRKRNLPGT), motif 2 (PDPDAEVIALSPKTLMATNRFVCEICNKGFQ RDQNLQLHRRGHNLPWKLK), motif 3 (EEVKKKVYVCPEPSCVHHDPSRALGDLTGI KKHFSRKHGEK), motif 1 (KWKCEKCSKKYAVQSDWKAHSKICGTREY KCDCGTLFSRRDSFITHRAFC), motif 7 (DALAEENARLASKTQNEN), motif 4 (YAPPPSPHMSATALLQKAAQMGSTASNPS) and motif 6 (GBDDLTRDFLGVGGNERR). *HaIDD*2, *HaIDD*6, *HaIDD*7, *HaIDD*13, *HaIDD*19 had same motif pattern as *HaIDD*1. *HaIDD*3, *HaIDD*4, *HaIDD*5, *HaIDD*8, *HaIDD*14, *HaIDD*16, *HaIDD*17, *HaIDD*18, *HaIDD*20, *HaIDD*21 had the same motifs as *HaIDD*1 but another additional motif 8 (DGNGAKPRLPLWLDHNANQPH). *HaIDD*5, *HaIDD*8 and *HaIDD*14 had also motif 10 (MKGIFLDDNMSNLTSASNEASLSSSSNRNEI GTMYPPPQIQ) (Fig. 4, Table 3).

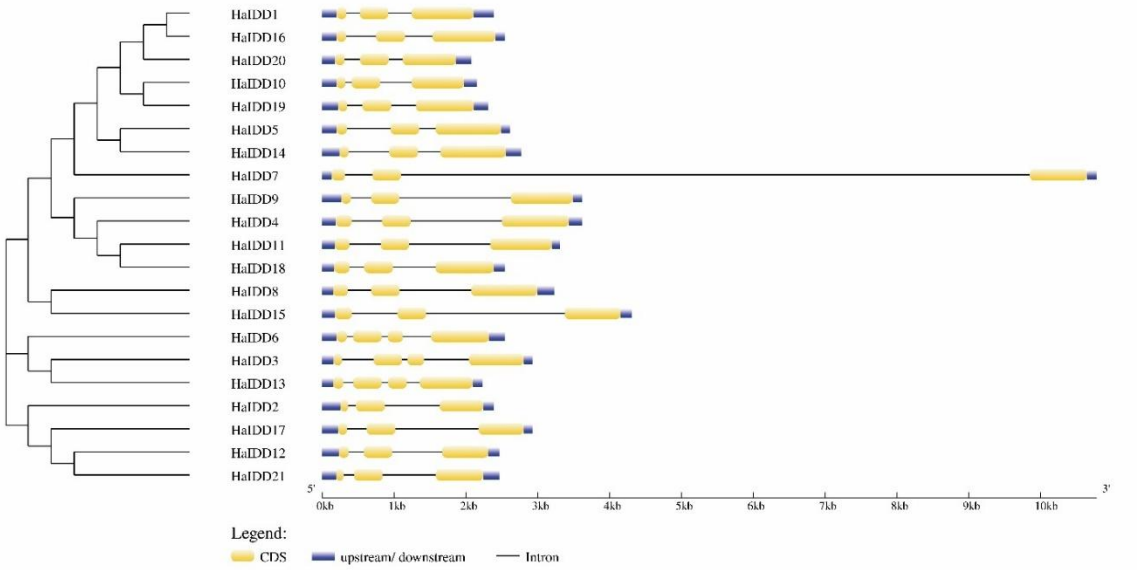


Figure 2. Gene structure of IDD gene family in sunflower. Green lines showed the Untranslated Regions (UTR) regions and yellow lines showed the intron regions. Scale bar is present at bottom.

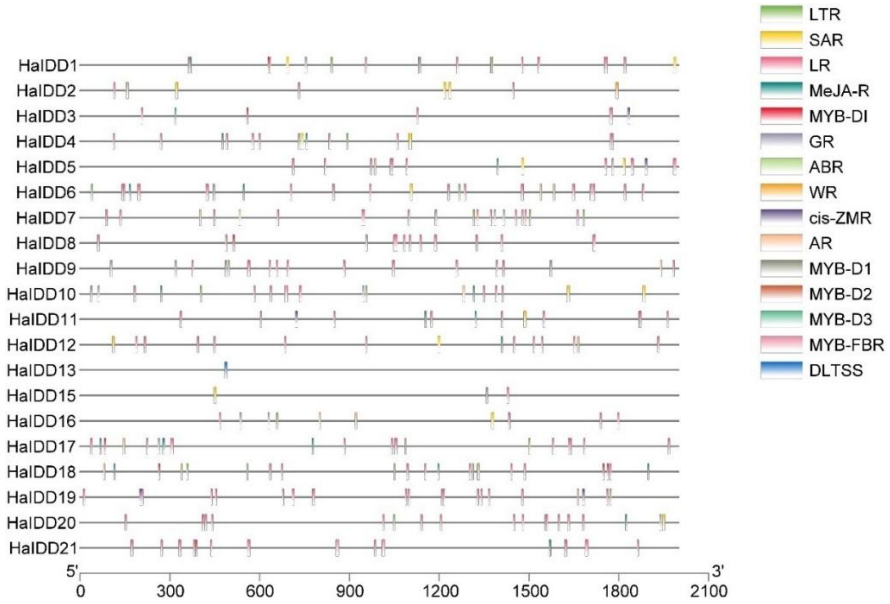


Figure 3. An expanded view of upstream region showing different coloured flags indicating various *cis*-acting elements

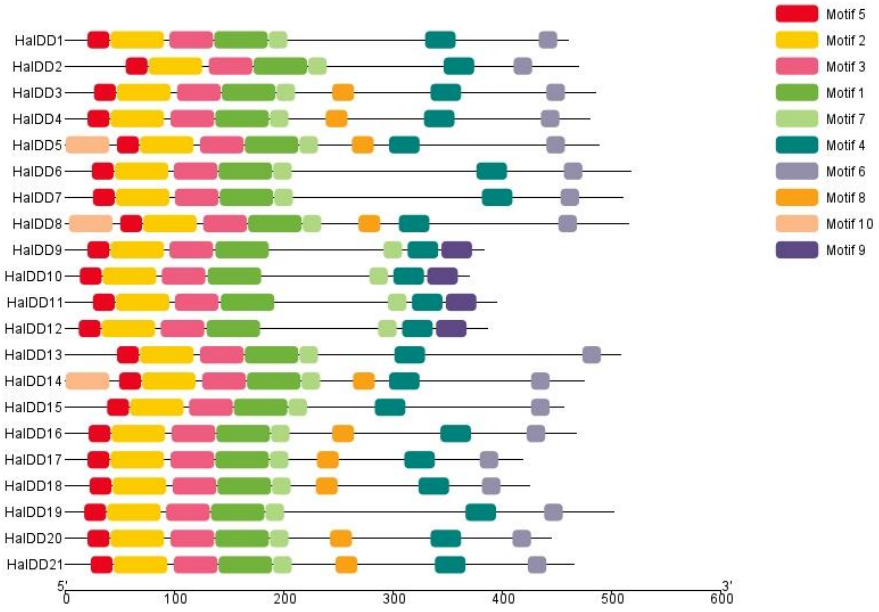


Figure 4. The online MEME software (http://memesuite.org) was exercised to find-out conserved motifs. Different colored boxes indicated the various themes.

Table 3. Sequences of conserved motifs and their length.

Motif No.	Sequence	Length
1	KWKCEKCSKKYAVQSDWKAHSKICGTREYKCDCGTLFSRRDSFITHRAFC	50
2	PDPDAEVIALSPKTLMATNRFVCEICNKGFQRDQNLQLHRRGHNLPWKLK	50

3	EEVKKKVYVCPEPSCVHHDPSRALGDLTGIKKHFSRKHGEK	41
4	YAPPPSPHMSATALLQKAAQMGSTASNPS	29
5	IPNPNPNSGLPNKRKRNLPGT	21
6	GBDDLTRDFLGVGGNERR	18
7	DALAEENARLASKTQNEN	18
8	DGNGAKPRLPLWLDHNANQPH	21
9	TILEITCHSCKQRFZASRNNGTAVTTDEA	29
10	MKGIFLDDNMSNLTSASNEASLSSSSNRNEIGTMYPPPQIQ	41

*HaIDD*5, *HaIDD*8 and *HaIDD*14 had also motif (MKGIFLDDNMSNLTSASNEASLSSSSNRNEIGTMYPPPQIQ). All the members of *HaIDD* family had same motif pattern which showed that this gene family was conserved. It had also been showed in the gene structure and *cis*-acting element analysis. These motifs play specific functions.

Chromosomal distribution, gene duplication and synteny analysis

Chromosomal distribution pattern of *HaIDDs* on respective chromosomes were determined. All *HaIDDs* were scattered on different chromosomes. Maximum genes (three on each) were located on Chromosome 7 and 9. Out of 17 chromosomes, 9 chromosomes contained only one *HaIDD* gene; three chromosomes contained only two genes. Rest of the chromosomes contained no genes (Figure 5a). Synteny analysis was executed to study the locus relationship between gene pairs. Segmental duplication event was detected between HaIDD1 and HaIDD16, HaIDD3 and HaIDD13, HaIDD5 and HaIDD14, HaIDD8 and HaIDD15, HaIDD12 and HaIDD21, HaIDD10 and HaIDD11, HaIDD 11 and HaIDD18 (Figure 5a, 5b). Non-synonymous (*Ka*) and synonymous (*Ks*) divergence levels for the duplicated genes were also investigated. It was observed that seven duplicated gene pairs demonstrated *Ka*/*Ks* value less than 0.5 advocating that the *HaIDDs* experienced strong purifying selection pressure with very less functional variance (Table 4).

Table 4. Calculation of Non-synonymous (*Ka*) and synonymous (*Ks*) divergence levels of sunflower IDD genes.

Sequence-1	Sequence -2	*Ka*	*Ks*	*Ka*/*Ks*	pS
HaIDD13	HaIDD3	0.099901	0.632026	0.158064	0.427090532
HaIDD21	HaIDD12	0.090478	0.594976	0.15207	0.410738255
HaIDD15	HaIDD8	0.316965	1.609818	0.196895	0.662323944
HaIDD18	HaIDD11	0.078317	0.488855	0.160206	0.359172414
HaIDD14	HaIDD5	0.036296	0.391414	0.092729	0.304949441
HaIDD19	HaIDD10	0.100008	0.474684	0.210683	0.351717617
HaIDD16	HaIDD11	0.384312	NaN	NaN	0.777173913

High Sequence Divergence Value (pS>=0.75). pS (probability of being synonymous)

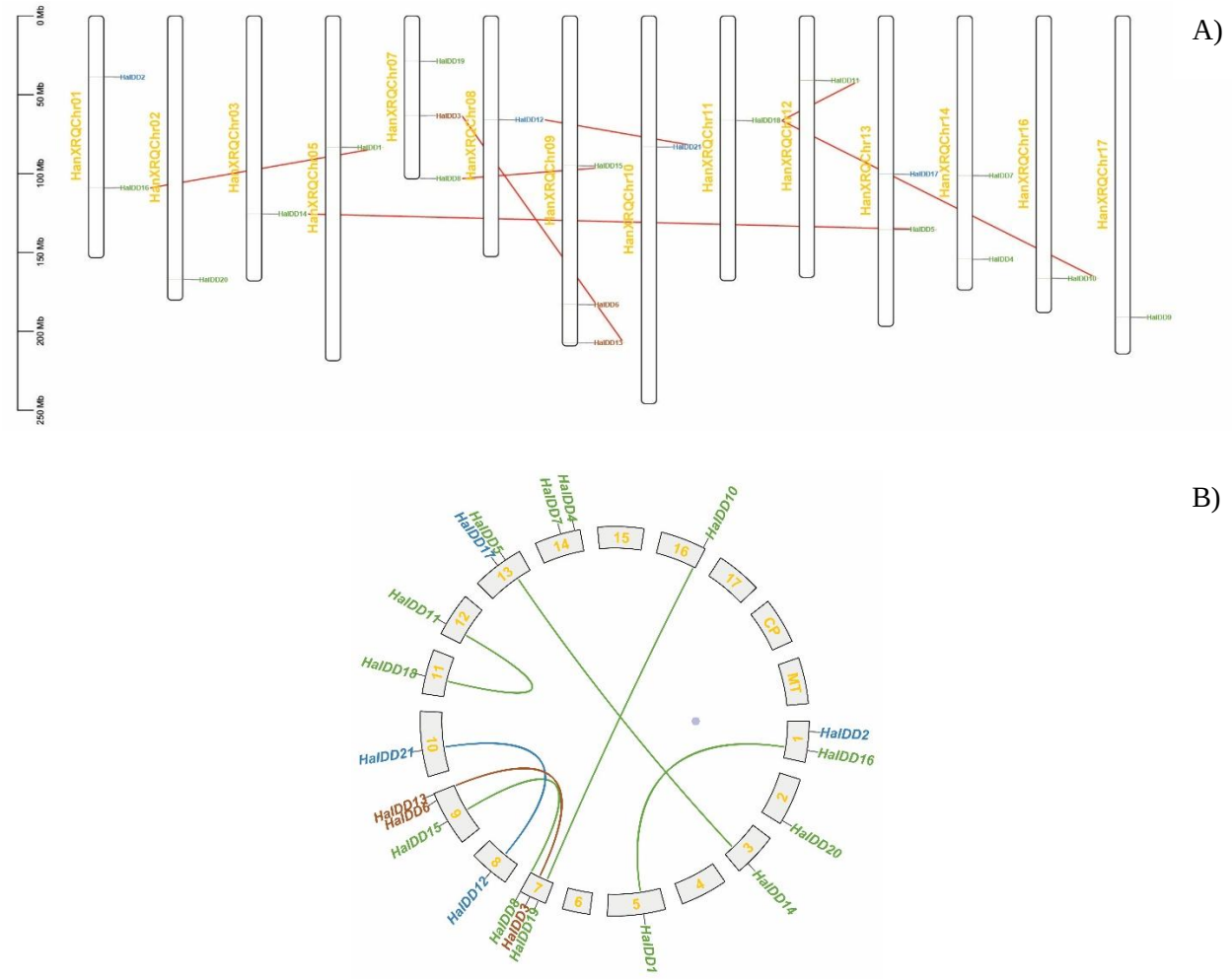


Figure 5. Genomic distribution and synteny analysis. Chromosomal location (a) and synteny analysis (b) of HaIDD genes.

Expression Pattern of *HaIDDs* in Different Tissues

The *IDD* gene family in plant is crucial for the developmental process and growth, such as root formation, starch and sugar metabolism, and transition of flower. Publicly available high-throughput microarray database was operated to analyze the pattern of *HaIDD* expression in different tissues of sunflower (https://www.sunflowergenome.org/). Expression of *HaIDDs* varies in different tissues and organs throughout the life cycle. The transcriptional level of *HaIDDs* was obtained from Sunflower Genome Database. Transcriptional level was detected in stamen, pollen, pistil, corolla, ligule, ovary, seed and bract. The expression pattern of *HaIDD*1, *HaIDD*5, *HaIDD*9, *HaIDD*10, *HaIDD*11, *HaIDD*13, *HaIDD*17, *HaIDD*18, *HaIDD*19 and *HaIDD*21 was undetectable. Expression level of other remaining gene members (*HaIDD*2, *HaIDD*3, *HaIDD*4, *HaIDD*6, *HaIDD*7, *HaIDD*8, *HaIDD*12, *HaIDD*14, *HaIDD*15, *HaIDD*16 and *HaIDD*20) was observed and presented in Fig. 6.

*HaIDD*2 had high level of expression in leaves and comparatively very low expression in seed. *HaIDD*3 had higher bract, seed ovary, pistil and leaves. *HaIDD*4 had highest expression in only seeds. *HaIDD*6 had highest expression in ovary, pistil, corolla, ligule and seed. While in bract and leaves the expression level is relatively low. In *HaIDD*7, there was down-regulation of the *IDDs*. *HaIDD*12 had high expression level in pistil, ligule, and leaves. *HaIDD*14 had highest expression level in ligules and high expression level in ovary,

leaves, seed and corolla. *HaIDD*15 had very low expression level while in *HaIDD*16 the expression level was very high in leaves, bract and seed. *HaIDD*20 had high expression level in leaves and bract. *HaIDD*6 and *HaIDD*14 showed almost similar expression pattern. *HaIDD*12 and *HaIDD*3 showed similar expression pattern. The high expression level of specific members in specific tissues showed that these genes play significant role in particular tissues during different stages. In *HaIDD*3 the expression level is highest in leaves. It means that this gene might be play role in photosynthesis and have effect on development of plant. This gene may play role in stress condition like heat, light. Semi q-PCR was conducted to check the expression of *HaIDD3* in seedling leaves and seeds. *HaIDD3* expression was compared with *HaActin.* As shows in the figure, it is evident that *HaIDD3* expression is relatively higher than *HaActin.*

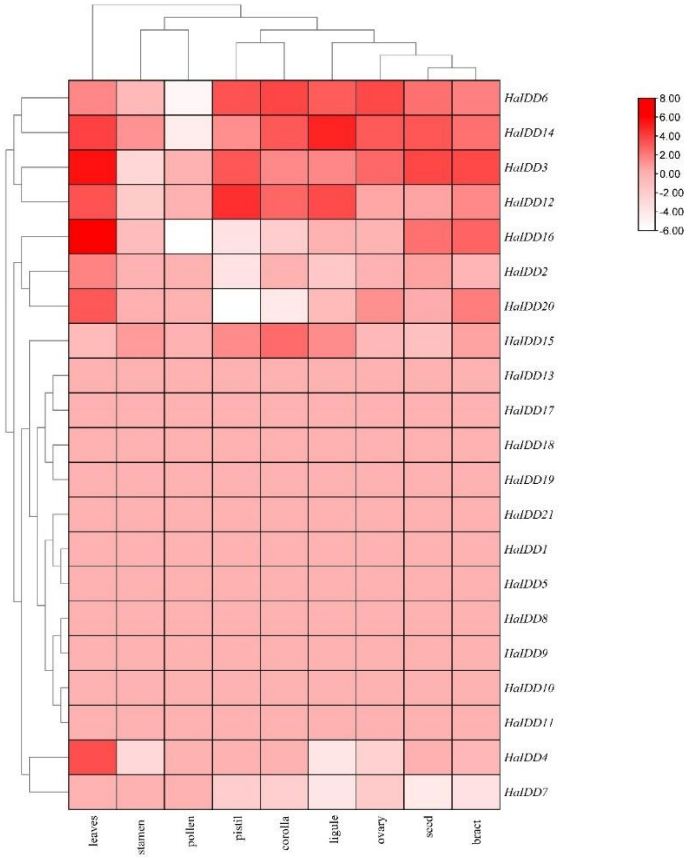


Figure 6. Expression patterns of *IDD* genes in various organs. Microarray data has been used to create a heat map. Pink indicated low expressiveness and red represented high expression on the right-hand color bar. Stamen, pollen, pistil, and corolla were among the tissues discovered.

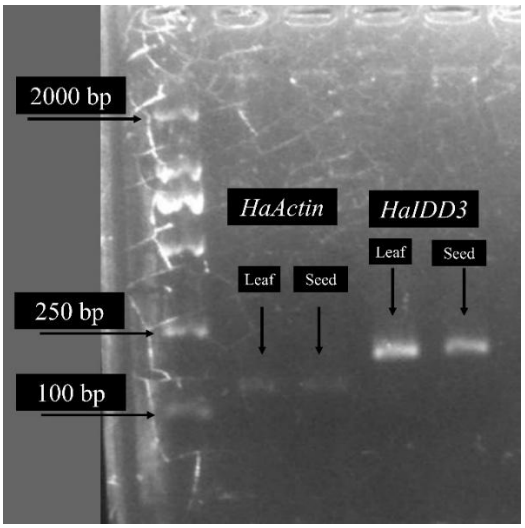


Figure 7. Semi qPCR results of HaIDD3 in comparison to *HaActin* in sunflower seedling leaves and seeds. Gel-electrophoresis based image showing strong bands for *HaIDD3* in sunflower leaf and seed correlating with heatmap results.

DISCUSSION

IDD gene encodes C2H2-type zinc finger transcription factors, which is among the biggest gene family in plants. *IDD* involved in plant growth & developmental processes. The *IDD* gene family has been identified i.e. apple, maize, rice and *Arabidopsis*. However, there has been no genome-wide discovery and study of IDD genes in sunflower. In this work, *IDD* genes in *H. annus*, *A thaliana*, *T. cacao*, moss (patens), *O. sativa* and *S. moellendorffii* (fern) were identified and analyzed in detail.

Absence of IDD gene in algal species and presence in mosses indicates that this gene family originated in early land plants (Wu et al., 2013 and Ali et al., 2019). In total, 78 *IDDs* were grouped into four sub-groups which demonstrated that most *HaIDDs* are in proximity of *TaIDD*, *OsIDD*, *AtIDD* genes predicted that these organisms may evolve from common ancestor. Furthermore, conserved amino acids residues pattern of IDDs in *H. annus* and *A. thaliana* during the evolution phase. Besides this, several different motifs were also identified which suggested that IDD proteins might perform differently in physiological pathways related to different co-factors. Some motifs are conserved in *HaIDD* members indicating the chances of gene evolution and duplication events. During the evolution of eukaryotes, loss or gain of intronic regions have been reported (Roy and Penny 2007). GSDS analysis showed that duplicated *HaIDDs* have similar exon-intron pattern whereas intronic region changes among *HaIDDs* highlighting that length of intron plays crucial roles in functional divergence of *HaIDD* genes. Moreover, *HaIDDs* have less number of introns (2 or 3) which suggest that *H. annus* is relatively new species.

Biophysical attributes of *HaIDD* genes predict that *HaIDDs* are present in nuclear. These results are also in accordance with Ali et al., 2019 and Fan et al., 2017). Furthermore, 21 IDD members of sunflower are unevenly distributed in the sunflower genome. The probable reason for this pattern of distribution would be gain or loss of genes during evolutionary history of *H. annus* (Park et al., 2020). In current work, 21 IDDs were identified in sunflower which were more in numbers as compared to model plants i.e. *Arabidopsis* and rice. The probable reason for higher number of *HaIDDs* would be that sunflower underwent gene duplication process. According to researchers, gene duplication played crucial role in crop evolution. Gene duplication consists of segmental, tandem and whole genome duplication (Xu et al., 2012). Most transcription factors experienced segmental duplication which resulted in gene family expansion functional divergence in crop plants (Qanmber et al., 2018, Yang et al., 2017). Gene duplication correlates with environmental and selection pressure (Ali et al., 2019). *Ka*/*Ks* ratio were used to explore the environmental and selection pressure on *HaIDDs*. *Ka*/*Ks* ratio were

less than 1 highlighting that *HaIDDs* experienced selection pressure. In evolutionary genetics, *Ka*/*Ks* ratio compares the rate of non-synonymous changeovers to the rate of synonymous changeovers. A pS >= 0.75 was applied so that high quality and conservative set of synonymous sites can be built. This approach was used to have more accurate Ks estimates.

HaIDDs had crucial roles in sunflower growth and development. In *Arabidopsis*, IDDs regulate GA signaling pathways (Long et al., 2015). In rice, *IDDs* regulate shoot response to gravity (Wu et al., 2013). In cotton, IDDs mainly expressed in post anthesis stage and upregulated in response to abiotic stress factors. In current study, *HaIDD* members showed variable expression. *HaIDD3*, *HaIDD6*, *HaIDD12*, *HaIDD14*, *HaIDD16* were more expressive in different tissues i.e. leaves pistil, ligule and seeds which clearly highlight its involvement in plant growth & development. *HaIDD*3 was also involved in light response or light stress condition as analysis showed that it has element of light response. Therefore, it can be deduced that it plays role in abiotic stress condition. In conclusion, our results showed that *HaIDDs* are crucial to sunflower vegetative and seed development and might prove an important member in governing other roles in sunflower.

REFERENCES

Ali, F., Qanmber, G., Li, Y., Ma, S., Lu, L., Yang, Z., ... & Li, F. (2019). Genome-wide identification of Gossypium INDETERMINATE DOMAIN genes and their expression profiles in ovule development and abiotic stress responses. *Journal of Cotton Research*, *2*(1), 3.

Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., ... & Langlade, N. B. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, *546*(7656), 148-152.

Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, *34*(suppl_2), W369-W373.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, *13*(8), 1194-1202.

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research*, *14*(6), 1188-1190.

Colasanti, J., Tremblay, R., Wong, A. Y., Coneva, V., Kozaki, A., & Mable, B. K. (2006). The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC Genomics*, *7*(1), 158.

Coneva, V., Guevara, D., Rothstein, S. J., & Colasanti, J. (2012). Transcript and metabolite signature of maize source leaves suggests a link between transitory starch to sucrose balance and the autonomous floral transition. *Journal of Experimental Botany*, *63*(14), 5079-5092.

Cui, D., Zhao, J., Jing, Y., Fan, M., Liu, J., Wang, Z., ... & Hu, Y. (2013). The *Arabidopsis* IDD14, IDD15, and IDD16 cooperatively regulate lateral organ morphogenesis and gravitropism by promoting auxin biosynthesis and transport. *PLoS Genetics*, *9*(9), e1003759.

Fan, S., Zhang, D., Xing, L., Qi, S., Du, L., Wu, H., ... & Han, M. (2017). Phylogenetic analysis of IDD gene family and characterization of its expression in response to flower induction in Malus. *Molecular Genetics and Genomics*, *292*(4), 755-771.

Gontarek, B. C., Neelakandan, A. K., Wu, H., & Becraft, P. W. (2016). NKD transcription factors are central regulators of maize endosperm development. *The Plant Cell*, *28*(12), 2916-2936.

Guo, A. Y., Zhu, Q. H., Chen, X., & Luo, J. C. (2007). GSDS: a gene structure display server. *Hereditas*, *29*(8), 1023-1026.

Huang, P., Yoshida, H., Yano, K., Kinoshita, S., Kawai, K., Koketsu, E., ... & Ueguchi-Tanaka, M. (2018). OsIDD2, a zinc finger and INDETERMINATE DOMAIN protein, regulates secondary cell wall formation. *Journal of Integrative Plant Biology*, *60*(2), 130-143.

Jeong, E. Y., Seo, P. J., Woo, J. C., & Park, C. M. (2015). AKIN10 delays flowering by inactivating IDD8 transcription factor through protein phosphorylation in *Arabidopsis*. *BMC plant biology*, *15*(1), 110.

Kaya, M. D., Okçu, G., Atak, M., Çıkılı, Y., & Kolsarıcı, Ö. (2006). Seed treatments to

overcome salt and drought stress during germination in sunflower (*Helianthus annuus* L.). *European Journal of Agronomy*, *24*(4), 291-295.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639-1645.

Kumar, M., Le, D. T., Hwang, S., Seo, P. J., & Kim, H. U. (2019). Role of the INDETERMINATE DOMAIN genes in plants. *International Journal of Molecular Sciences*, *20*(9), 2286.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, *33*(7), 1870-1874.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., ... & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, *30*(1), 325-327.

Letunic, I., Doerks, T., & Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic acids research*, *43*(D1), D257-D260.

Long, Y., Smet, W., Cruz-Ramírez, A., Castelijns, B., De Jonge, W., Mähönen, A. P., ... & Blilou, I. (2015). *Arabidopsis* BIRD zinc finger proteins jointly stabilize tissue boundaries by confining the cell fate regulator SHORT-ROOT and contributing to fate specification. *The Plant Cell*, *27*(4), 1185-1199.

Matsubara, K., Yamanouchi, U., Wang, Z. X., Minobe, Y., Izawa, T., & Yano, M. (2008). Ehd2, a rice ortholog of the maize INDETERMINATE1 gene, promotes flowering by up-regulating Ehd1. *Plant Physiology*, *148*(3), 1425-1435.

Park, B., & Burke, J. M. (2020). Phylogeography and the evolutionary history of sunflower (*Helianthus annuus* L.): wild diversity and the dynamics of domestication. *Genes*, *11*(3), 266.

Qanmber, G., Yu, D., Li, J., Wang, L., Ma, S., Lu, L., ... & Li, F. (2018). Genome-wide identification and expression analysis of *Gossypium* RING-H2 finger E3 ligase genes revealed their roles in fiber development, and phytohormone and abiotic stress responses. *Journal of Cotton Research*, *1*(1), 1-17.

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, (1), 2-3.

Roy, S. W., & Penny, D. (2007). Patterns of intron loss and gain in plants: intron loss–dominated evolution and genome-wide comparison of *O. sativa* and *A. thaliana*. *Molecular Biology and Evolution*, *24*(1), 171-181

Seo, P. J., Ryu, J., Kang, S. K., & Park, C. M. (2011). Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in *Arabidopsis*. *The Plant Journal*, *65*(3), 418-429.

Wu, X., Tang, D., Li, M., Wang, K., & Cheng, Z. (2013). Loose Plant Architecture1, an INDETERMINATE DOMAIN protein involved in shoot gravitropism, regulates plant architecture in rice. *Plant Physiology*, *161*(1), 317-329.

Xu, G., Guo, C., Shan, H., & Kong, H. (2012). Divergence of duplicate genes in exon–intron structure. *Proceedings of the National Academy of Sciences*, *109*(4), 1187-1192.

Yang, Z., Gong, Q., Qin, W., Yang, Z., Cheng, Y., Lu, L., ... & Li, F. (2017). Genome-wide analysis of WOX genes in upland cotton and their expression pattern under different stresses. *BMC Plant Biology*, *17*(1), 113.

Yi, G., Neelakandan, A. K., Gontarek, B. C., Vollbrecht, E., & Becraft, P. W. (2015). The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation. *Plant Physiology*, *167*(2), 443-456.

Zhang, T., Tan, M., Geng, L., Li, J., Xiang, Y., Zhang, B., & Zhao, Y. (2020). New insight into comprehensive analysis of INDETERMINATE DOMAIN (IDD) gene family in rice. *Plant Physiology and Biochemistry*, *154*, 547-556.