Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The DNA binding landscape of four seed-specific bZIP transcription factors in bread wheat: use of Bind-n-seq *in vitro* analysis

Koushik Shah
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India. https://orcid.org/0000-0002-5631-8013
Raminder Kaur
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India.
Nishtha Sharma
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India.
Prateek Jain
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India.
Dalwinder Singh
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India.
Shrikant Mantri
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India.
Vikas Rishi (vikasrishi@nabi.res.in)
National Agri-Food Biotechnology Institute, Sector-81 (Knowledge City), S.A.S. Nagar, Mohali-140306, Punjab, India.

Research Article

Keywords: Basic leucine zippers(bZIPs), DNA-Protein interactions, Bind-n-seq, Seed development, Regulatory G-box, hexaploidy/bread wheat

Posted Date: June 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1787800/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Sequence-specific transcription factor (TF)-DNA binding is a vital step in biology since it directly regulates genetic information inside the organism. However, the binding sites for many TFs are unknown in plants and animals.Using high-throughput sequencing technologies,studies have shown that TFs, including basic leucine zipper (bZIP), bind to a myriad of sequences, many of which are biologically relevant. bZIPs are eukaryote-specific dimeric TFs, and in plants, they are reported to bind to 8–12 nts sequences (with ACGT core) in the promoter region of many genes. In sedentary plants, they are the conduit for converting environmental cues into gene activity. A key question is what determines bZIP DNA binding specificity.This study adopts the Bind-n-seq methodology, which uses a 108 nts oligonucleotides library with an 18 nts random sequence that either incorporates ACGT tetranucleotides core or a completely random sequence to generate *in vitro* DNA-bZIP interaction maps of four seed-specific bread wheat bZIPs. Subsequently, *de novo* binding motifs representing the complete binding landscape for each bZIP were determined. Binding analysis shows a high degree of overlap for bZIP binding to G-box and C-box but with distinct flanking bases, suggesting a discrete role of flanking regions in ordaining TF-DNA binding specificity. From this study binding motifs obtained for Hy5 and Embp-1 show similarity with previously reported sequences, suggesting the protocol's robustness.

Introduction

A comprehensive understanding of specificity of transcription and gene regulation continues to unfold (Biggin 2001, Todeschini, Georges et al. 2014). DNA-protein interaction is a central event of gene regulatory networks, which act as a connecting link between the genome and proteome. TFs act as regulatory switches known to modulate gene expression in response to various environmental and developmental cues (Lai, Stigliani et al. 2019). TFs exert their effect by sequence-specific interactions with cis-elements of genes, either activating or repressing these target genes.Therefore, identifying the putative binding sites of TFs helps us to decipher gene regulatory networks within the cellular system under different physiological conditions.

As a sessile organism, the plant has developed and further enhanced various regulatory mechanisms to survive under adverse conditions such as drought, salinity, and high temperature. Earlier studies suggested that gene regulation may play more imperative roles in plants than in animals, since a larger number of TF-coding genes (6–10%) are listed in plant genome databases compared to 2% in human (Riechmann, Heard et al. 2000). bZIPs are a pivotal dimeric class of transcription factors unique to eukaryotes that bind to the major groove of DNA in a sequence-specific manner. bZIP domain is 60–80 amino acids long bipartite alpha-helical structure with conserved N-terminal DNA binding half with invariant asparagine and a conserved N-x7-R/K motif rich in basic amino acids. In comparison, the other C-terminal amphipathic half is variable in length with the occurrence of leucine every seven amino acids. The basic region is responsible for sequence-specific DNA binding, while the leucine zipper is an amphipathic coiled-coil structure responsible for its dimerization propensity (Hurst 1994, Llorca, Potschin et al. 2014). Previous studies in plants, revealed that bZIP binds to DNA containing ACGT core elements

as in, G-box (CACGTG), C-box (GACGTC), and A-box (TACGTA) to regulate downstream target genes (Izawa, Foster et al. 1993, Foster, Izawa et al. 1994, Niu, Renshaw-Gegg et al. 1999).

In plants, bZIP's involvement in various biological processes such as seed development and maturation, pathogen defense, hormonal signaling, floral initiation, etc. is well documented. bZIP act as a key component that response to various abiotic and environmental stress conditions (Jakoby, Weisshaar et al. 2002). In recent years, through genome-wide analysis, bZIPs genes are identified in several organisms, such as in *Saccharomyces cerevisiae* (17), *Arabidopsis thaliana* (67), *Oryza sativa* (89), *Sorghum bicolor* (92), *Brachypodium distachyon* (96), maize (125), tomato (69), *Musa acuminata* (121), *Drosophila melanogaster* (27), *Homo sapiens* (56), and *Triticum aestivum* (~300) (Nijhawan, Jain et al. 2008, Wei, Chen et al. 2012, Li, Fu et al. 2015, Li, Gao et al. 2015, Liu and Chu 2015, Pourabed, Ghane Golmohamadi et al. 2015, Wang, Cheng et al. 2015).

Sequencing of 14.5 GB of wheat genome and transcriptome atlas of different development stages of the plant has opened new opportunities to analyze familiar and novel bZIP TFs (International Wheat Genome Sequencing, investigators et al. 2018, Ramirez-Gonzalez, Borrill et al. 2018). In human and model plant Arabidopsis, majority of the bZIPs are functionally annotated. However, in complex genome like hexaploidy wheat, assigning functions to TFs is challenging, since bZIP-DNA binding information is scarce. For example, wheat is predicted to have ~300 bZIP TFs, but only a few of these TFs e.g., Embp-1, and SPA have known binding motifs that have been experimentally validated (Guiltinan and Miller 1994, Albani, Hammond-Kosack et al. 1997, De Jong 2013). Genome-wide *in vivo* studies such as Chromatin immunoprecipitation (ChIP-seq and ChIP-chip) identified the binding sites of many transcription factors in plants and animals (Johnson, Mortazavi et al. 2007, Chen, Bhadauria et al. 2018). These techniques are limited by the availability of specific antibodies and the number of TFs that can be processed in one time (Geertz and Maerkl 2010). Due to difficulty in obtaining ample binding spectra of TFs using *in vivo* data, other *in vitro* binding techniques such as protein-binding microarrays (PBM)(Berger and Bulyk 2009), SELEX (Jolma, Kivioja et al. 2010, Smaczniak, Angenent et al. 2017), MITOMI (Rockel, Geertz et al. 2012), DAP-seq (O'Malley, Huang et al. 2016) and Bind-n-seq (Zykovich, Korf et al. 2009) are engaged to measure transcription factor binding affinities to a broad array of possible binding sites in various organisms.

This study adapted Bind-n-seq for *in vitro* analysis of seed-specific bZIPs DNA binding in the bread wheat genome that takes advantage of next generation short-read sequencing. The technique involves incubating proteins with random target DNA's sequences that carry distinct barcodes for downstream identification of bound sequences. The bound oligonucleotides are separated from unbound fraction, eluted and sequenced using next generation sequencing. The binding motifs are extracted from the sequences. Unlike SELEX, multiple rounds of binding and amplification are not required. In addition, many binding reactions can be assayed in parallel with barcoded oligonucleotides. This report uses an 18-bp binding region with and without ACGT core element for binding. We show the efficacy of Bind-n-seq by analyzing the DNA-binding activities of two well-characterized bZIPs HY5 (Long Hypocotyl 5) from *Arabidopsis* and Embp-1 from wheat. Using the protocol, we could obtain 5–10 thousand reads per

sample while simultaneously analyzing 45 samples and identifying the canonical binding site for each protein with as few as 1000 reads. From the sequencing runs G and C boxes were found to be enriched suggesting plant bZIPs preferential binding to these motifs. To define low and high-affinity binding sites, this study involves quantitative analysis of DNA binding sites of four seed-specific bZIPs i.e., Embp-1, ABI5, ABFB, and TabZIP-1.

Materials And Methods

Several binding reactions are executed simultaneously under varying salt and protein concentrations (Fig. 2). The key advantage of Bind-n-seq over other techniques is that the binding specificities and affinities for many proteins can be determined simultaneously.Two negative controls were performed; 1) Sequencing of the entire input library to assess the randomness of the input pool and 2) a no protein binding condition to detect potential apparatus-selected biases.

bZIP cloning

The coding region of the full-length DNA binding domain (DBD) of hexaploid bread wheat bZIPs Embp-1, ABI5, ABFB, and TabZIP-1 were amplified from wheat inflorescence cDNA (mature, developing seeds from different stages) and cloned into LIC-pT5 plasmid that used ligation-independent cloning (LIC) method. In addition, tissue-specific expression profiles of bZIPs were analyzed in seed developmental stages using plants ensemble (Borrill, Ramirez-Gonzalez et al. 2016). Further, the DNA-binding domain(DBD) and leucine zipper region of bZIPs were characterized as reported previously (Fassler, Landsman et al. 2002, Jakoby, Weisshaar et al. 2002, Vinson, Myakishev et al. 2002, Deppmann, Acharya et al. 2004). Also, Hy5 from *Arabidopsis*, involved in photomorphogenesis was cloned and used as a control for binding studies (Gangappa and Botto 2016).

bZIPs protein expression and purification

cDNA coding for DBDs of four bZIPs from LIC-pT5 were sub cloned in pETGEX-CT vector (Sharrocks 1994) as NcoI-SacI. Inserts were PCR amplified using primers carrying restriction sites for NcoI (forward) and SacI (reverse) (Supplementary Table S1).Restriction digestion of PCR products with NcoI and SacI was used to create bZIP-GST expression plasmids. The C-terminal GST-tag allows rapid one-step purification over the glutathione resin column. bZIPs DBDs were transformed and expressed in BL21(DE3) pLysE *E.coli* strain (Invitrogen) using heat shock method. 20ml of overnight culture with antibiotics (chloramphenicol (34mg/ml), ampicillin (100mg/ml) was used to inoculate 500 ml of Luria Broth (LB) media in a 2L flask supplemented with ampicillin (100mg/ml).The culture was incubated at 37°C with shaking (300 rpm) until the culture attained an OD_{600} = 0.4–0.6 (~ 3 hrs). The bZIP protein expression was induced by adding Isopropylthio-β-galactoside(IPTG) to a final concentration of 0.5mM.Bacterial cultures were grown for 3–4 hrs at 37°C or 16°C overnight.Subsequently,bacterial cells were harvested by centrifugation at 4000 rpm,4°C for 30 min. and the supernatant was discarded. Cell pellets were resuspended in 10 ml of lysis Buffer (50 mM Tris (pH 8.0), 100 mM KCl, 1mM DTT, 5 mM EDTA, and 0.2 mM PMSF) using several rounds of pipetting and vortexing. After repeating the freeze-

thaw cycles twice on ice, cell lysate was cleared by centrifugation at 25000 rpm for 30 min at 4°C. After lysis, clarified cell lysate was directly applied to the glutathione resin (GE Healthcare) column that was previously equilibrated with bZIP binding buffer. The column was washed two times with wash buffer (20mM Tris-HCl (pH 8.0), 100 mM KCl, 1mM DTT, 5mM EDTA, and 0.2 mM PMSF) to remove non-specific binding and impurities. GST-bZIP proteins were eluted from the column using 20 mM reduced glutathione (50 mM Tris-HCl, pH 8.0). Protein-containing fractions were run on 12% SDS-PAGE, and protein bands were visualized after Coomassie blue stain for four hours followed by de-staining with deionized water.Protein-containing fractions were pooled and dialyzed against 1L of dialysis buffer (20 mM Tris-HCl (pH 8.0), 100 mM KCl, 5 mM EDTA, 0.2 mM PMSF, and 1 mM DTT) for overnight with stirring at 4°C to remove reduced glutathione. Concentrations of bZIPs were measured by UV absorption spectrophotometer (Supplementary Fig. 1&2). Purified proteins were stored at – 20°C until further use.

Bind-n-seq library amplification and binding

To generate DNA libraries for Bind-n-seq,108-mer oligonucleotide pool with distinct barcodes were obtained from Eurofins and Sigma.The list of barcodes used here is given in the Supplementary File. Each library contains Illumina primer-binding sites, 2-nt CC leader sequence, 4-nt barcode for multiplexing and 18-nt binding region with and without ACGT core element (Foster, Izawa et al. 1994, Zykovich, Korf et al. 2009). Single-stranded (ss) DNA libraries were made double-stranded by asymmetric PCR using a reverse primer in a 25 μl reaction mixture containing 1 μM of each template,10 μM reverse primer, and a polymerase Go TaqPro Complete (Promega).PCR conditions used were: denaturation at 95°C for 2 min, annealing at 55°C for 1 min, extension at 72°C for 4 min (X 25 cycles) and finally 4°C for storage.To initiate the binding response, an extra 25 μl volume was added to attain a final concentration of 0.12mg/ml of herring sperm DNA (Sigma),100μM $MgCl_2$, 5mM DTT, 1% BSA, and indicated concentrations of KCl and purified binding protein (Supplementary Fig. 3). Binding reactions were incubated for 2 hrs at 22°C.

Pull down and enrichment of bZIP bound DNA

For enrichment, 50μl of glutathione resin was washed twice (by pelleting and resuspension) with 500μl of water,then equilibrated with bZIP binding buffer (20 mM Tris(pH8.0),150mM KCl,1mM $MgCl_2$, 5mM DTT,1mM EDTA). 50μl binding reaction was added to the 50μl of prepared resin, mixed, and incubated for 30 min at 23°C with gentle mixing for every 10 min. The reaction was then washed twice using the appropriate wash buffer (10mM-300mM salt) for 10-min incubations. Protein–DNA complexes were eluted by a 10-min incubation in 30 μl of elution buffer [50mM Tris (pH 8.0), 20mM reduced glutathione], pelleting the resin carefully transported the supernatant to a new tube.

Output DNA quantification and Library preparation

After binding,output samples were quantified using Real-time PCR (qRT-PCR) to measure the number of cycles required to equal concentrations for pooling and sequencing.Samples were analyzed on BioRad CFX manager(BioRad),10μl reaction containing 1μl of eluted TF-bound DNA, 0.5μM each of forward and reverse primer, SYBR Green PCR Master Kit (BioRad).PCR conditions used were: read 90°C for 10 min,40

cycles of 95°C for 15s,63°C for 1min, read every cycle, melting curve 40–95°C, hold for 3s and read every 0.7°C. A standard curve was generated using template DNA concentrations at 150, 15, 1.5, and 0.5 nM (Figure.3).

All TF-bound DNA were pooled in equimolar concentration. After quantification, sequencing run were performed on the Illumina MiSeq platform.Following manufacturer's instructions, the purified multiplexed PCR products (10-100ng) were end-repaired and ligated with adaptors using QIAseq 1-step amplicon library kit (Qiagen). The detailed procedure of adaptor ligation and sequencing is given in the supplementary section.

Sequence analysis and Bioinformatics

All raw FASTQ sequence files were trimmed using the Trimmomatic tool to remove adaptors.Then, trimmed sequencing reads were filtered and sorted using barcodes.The filters were only included: only A, C, G, T bases allowed,valid barcode to distinguish reads and constant ACGT core regions, and unique random regions in the random library.

After the library preparation and size selection steps, all the Bind-n-Seq reaction samples were mixed in equimolar ratios, and pair-end sequencing (2 × 75 bp) was performed using Illumina MiSeq sequencing platform according to the manufacturer’s instructions.First,the sequencing reads were obtained as fastq files (2�2.5M) and sorted according to their barcode tags. Next, the sequences were filtered using definite base calling (A, C, G, and T), valid barcode(4bp) and constant leader bases (2bp), and a unique 18bp random sequence. Finally, after quality control, the clean data of 108-mer reads were obtained and used for further analyses.

We obtained data set of 0.2 � 2 million reads while simultaneously analyzing 45 samples and identifying the canonical binding site for each protein with 5,000-10000 reads.The fold enrichment of bound DNA sequences is proportional to the relative affinity of the protein. The motif analysis was performed with MEME package 4.5.0 in –DNA –comp –w 6 + 18 –n motifs parameters. The fold enrichment of a motif was calculated as the ratio of the reads containing the motif to the total motif obtained in a selected binding reaction. Reads were calculated for the occurrence of motifs for each binding reaction using the FIMO (Find Individual Motif Occurrence) tool of MEME software with a threshold P-value of 0.001. Results from FIMO were processed by filtering low complexity patterns and those seed sequences with an enrichment below 2.5-fold. Finally, we applied an R script to select the final list of sequences submitted to the 1) all the unique motifs, 2) shorter unique motifs that might be contained in longer ones, and 3) longer unique motifs. Sequence logos of binding motifs were generated using the WEBLOGOS web tool (Crooks et al.,2004).For a comparative study of Arabidopsis Hy5 binding motifs using Bind-n-seq, the transcription factor databases JASPER and CIS-BP were used.

Results

Bind-n-seq methodology and binding analysis for four bZIPs

Recently,Bind-n-Seq was applied to study DNA binding of few classes of TFs e.g., zinc fingers and RNA binding proteins(RBPs) in animal systems (Zykovich, Korf et al. 2009, Lambert, Robertson et al. 2015).In the current study,we adapted the technique to describe and analyze binding sites for TaABI5, TaABFB, TaEmbp-1, and Tabzip-1 that showed higher expression during seed development.The outline of Bind-n-seq used for bZIP binding is shown in Fig. 2.For binding,purified bZIP protein was mixed with randomized double-stranded(DS) target DNAs, which contained four bases barcode,18 bp binding region with and without ACGT core element, and flanking primer-binding sites.A randomized part of 14bp with ACGT core contains $2.6×10^7$ members (4^{14}), and a random library (4^{18}) has $6.8×10^9$ binding sites.Each binding reaction has an over-representation of each possible 18-mer,equivalent to 20 pmol or 384 ng of single-stranded 108-mer oligonucleotides.Double-stranded libraries were created by primer extension.After incubation, protein–DNA complexes were separated from unbound DNAs and TF bound DNA was quantified.The DNAs from various binding conditions were pooled in approximately equal concentrations, and massively parallel sequencing was performed.After sequencing, bar codes were used to sort the DNAs into their respective experimental conditions. For binding, each protein concentration ranging from 5nM to 500 nM and salt concentration from 10mM to 300mM were used(Shown in supplement data). The sequence-specific DNA binding specificities of bZIP proteins have been studied *in vitro* (supplementary tables 4–8). Our binding data showed that 50nM bZIP in presence of 150mM KCl is stringent enough for bZIP binding, as most reads were identified in this condition (Fig. 4).

Evaluation of binding specificity among Hy5 and wheat bZIPs

In total, 50,000 unique binding sites were identified in this study. *Arabidopsis* Hy5 was used as a control for binding reactions in parallel to other wheat bZIPs.In addition,binding landscape for seed-specific bZIPs i.e., Embp-1, TabZIP-1, TaABI-5, and TaABFB under various salt and protein concentrations conditions were obtained.Analysis of binding data shows that most of the bZIPs are preferentially bound to the ACGT core motif with distinct flanking bases, as shown in the supplementary file.

Comparing PWM models under varying salt conditions shows that Hy5 preferably binds to high-affinity GGCACGTG(G-box) sequences in 150mM and 300mM salt.Earlier studies revealed that in addition to G box, CG hybrid (GACGTG) and CA hybrid (GACGTA) are also bound well by Hy5 (Song, Yoo et al. 2008).The binding motifs were also analyzed under various protein concentrations (500 nM, 100 nM, 50 nM, and 5 nM). As shown in Fig. 5,results were similar to previously described Chip-chip and DAP-Seq cistrome databases for Hy5 (Zhang, He et al. 2011, O'Malley, Huang et al. 2016).

Embp-1 is bZIP TF and its homologs are found in many plants and earlier reported in abscisic acid response in plants. Embp-1 is an embryo-specific bZIP in wheat and binds with the 10bp palindromic ‘GCCACGTCGG’ (G-Box) sequence (Guiltinan, Marcotte et al. 1990, Guiltinan and Miller 1994, De Jong 2013). Embp-1 in 10 and 150mM salt conditions interacts explicitly with the ‘GCC/GACGTC/GGC’ sequence. In 300 mM salt, Embp-1 binds specifically to the G box. In 5 nM and 50 nM protein

concentrations,we found similar G and C box motifs, whereas, 500 nM protein condition shows bZIP binding to palindromic sequence ‘GCGACGTCGC’ (G-Box).

3. TaABI5 binding and conserved functions in orthologs

ABA INSENSITIVE 5 (ABI5) is involved in the abscisic acid (ABA) signaling pathway and gets activated in response to abiotic stresses.ABI5 is described as a regulator of seed germination and early seedling development in ABA and abiotic stresses (Finkelstein, Gampala et al. 2002, Skubacz, Daszkowska-Golec et al. 2016).However,the precise role of ABI5 in hexaploid bread wheat in ABA signaling and its function under stress remains elusive.Under different salt conditions, PWM models of TaABI5 show that it interacts with ACGTG (G-Box) with high specificity.TaABI5 binds specifically to the ‘GGGCACGTGCCC’ with ACGT core followed by G in all three salt conditions.TaABI5 interacts with ACGTC (C-Box) and ACGTG (G-Box) with high specificity in presence of different protein concentrations (5-500 nM). In 100 nM, 50nM, and 5 nM conditions, TaABI5 binds with ACGT core element ‘CGCGACGTCGCG’. In contrast, at 500 nM condition, the TaABI5 interacts with ACGTC (C-Box) flanked by different adjacent bases (supplementary table). TaABFB (ABA response element-binding factor) is seed-specific bZIP and a sequence homolog of ABI5. TaABFB is involved in various ABA-mediated signaling pathways (Choi, Hong et al. 2000). In our study, TaABFB binds to 10 bp palindromic sequence ‘GCGACGTCGC’ (C-Box) when binding reaction is performed in presence of different salt concentrations. In low protein concentrations, TaABFB binds to same C-box whereas at high protein concentration (500nM), TaABFB interacts CGGGACGTCGG (C-Box) but with different flanking region.

bZIPs binding in the absence of ACGT core element

Plant bZIPs are known to assemble around ACGT core DNA sequences (Izawa, Foster et al. 1993, Foster, Izawa et al. 1994). Here, we have explored the binding of four bZIPs to sequences that lack core ACGT element. Therefore, an additional library with a randomized region that may still carry ACGT by chance was used for binding under 150 mM salt and 50 nM protein concentration (Shown in supplement Table). We observed, that most of the bZIPs including Hy5 bind to the sequences enriched in the ACGT core element, and sequence specificity remain same for G-box and C-Box. The binding of bZIPs to the ACGT core suggest that bZIPs preferably bind to the ACGT core to regulate downstream genes. Using this randomized library, TaABI5 binding is enriched by ‘CGCGCGGGCGGG’ sequence that lack ACGT core, suggesting that at least TaABI5 preferentially binds to non-ACGT sequence(Shown in Fig. 4). Also, homolog of ABI5, TaABFB binds with high affinity to the ACGT core element sequence (GCC/TACGTCGG), suggesting a high degree of specificity that may be governed by the N-terminal DBD of proteins. Results show that all bZIPs but TaABI5 preferentially bind to ACGT core elements over the random sequence, and TaABI5 binds to GC-rich palindromic sequences in an enriched manner.

Discussion

The binding of TF to DNA sequences is a significant step in the gene regulatory network in an organism. TF scan and binds to short DNA sequences present in adjacent DNA regions of their target genes, leading

to the tight regulation of genes by recruiting or impeding the general transcription machinery (Badis, Berger et al. 2009, Stormo and Zhao 2010). In plants, bZIPs are involved in various biological functions such as seed development and maturation, pathogen defence, environmental signals, and stress(biotic and abiotic) responses (Jakoby et al., 2002; Dröge-Laser et al., 2018). Seed development is a complex but key process, which is initiated as plant perceive various internal and external cues from the environment. This study suggests to catalog the DNA binding landscape of four seed-specific bZIPs (TaEmbp-1, TaABI5, TaABFB, and TabZIP-1) in hexaploid bread wheat, and by mapping it to genome may provide regulatory evidences related to seed development, abiotic stress, and other metabolic processes (Fig. 5). Genome-wide analysis in wheat has suggested ~300 putative bZIPs, but most of them are not functionally annotated and their binding sites are still to be decoded (Chen, Guo et al. 2015, Li, Gao et al. 2015). In recent studies, numerous *in vivo* (ChIP-based) and *in vitro* techniques such as PBM, SELEX and DAP-seq allowed the researchers to measure the binding specificity of TFs in plants under different experimental conditions. However, these methods are limited by the availability of antibodies, feature-length, and cyclic amplification of targets (Berger and Bulyk 2009, Zhu, Sun et al. 2012, Bartlett, O'Malley et al. 2017, Franco-Zorrilla and Solano 2017). Here, we used an efficient and high-throughput *in vitro* binding protocol to identify the binding sites of bZIPs that showed elevated expression during seed development. Earlier, the method was used successfully to study the binding of DBDs of zinc-finger proteins and RNA binding proteins (RBPs). One of the advantages of Bind-n-seq is that it allows the examination of a broad range of transcription factors from any source and system and subsequent profiling of the binding sites efficiently and cost-effectively (Zykovich, Korf et al. 2009, Lambert, Robertson et al. 2014, Lambert, Robertson et al. 2015). In bZIPs, the DNA binding specificity of the bZIP domain is governed by the N-terminal basic region, which interacts with the major groove of DNA. Further, the distinct base specificity of the binding motif depends on the presence of amino acids like conserved asparagine (N) and arginine (R) in the basic region (Ellenberger et al., 1992). The DBDs of seed-specific bZIPs have been annotated as reported earlier (Fassler, Landsman et al. 2002, Vinson, Myakishev et al. 2002). Asparagine interacts with the second and third base in consensus sequence to define specificity. Arginine preferably interacts with guanine in the major groove of DNA (Fujii et al., 2000).

Our study found that Bind-n-seq could recognize equitable *in vitro* binding motifs for Hy5 as reported earlier using ChIP-seq (Zhang et al., 2011) and novel binding motifs for wheat bZIP TFs (supplementary Table 4–8). However, we also obtained bound DNA motifs that display poly T and poly G/C low complexity traits in a few pull down binding reactions. This may be due to the lack of complete randomness of bases in the oligo library (Zykovich, Korf et al. 2009). Hy5, an *Arabidopsis* bZIP TF, is a well-studied master regulator that binds to the promoter of numerous genes, and regulates several biological processes during plant growth and development (Gangappa and Botto 2016). Therefore, it was used as a control for the wheat seed-specific bZIP TFs and their *in vitro* binding analysis. Using Bind-n-seq, comparative analysis of Hy5 bZIP frequency PWM model at 150 mM salt and 50 nM protein concentration, (considered as stringent) are in good agreement with those obtained from other *in vivo* and *in vitro* methods (ChIP-chip, DAP- seq) (Fig. 5), suggesting the validity and robustness of the technique.

Comparative analysis of different binding reactions showed that 50 nM, 100 nM protein, and 150 mM salt concentration provide efficient enrichment and a higher number of specific reads in the binding experiment. Furthermore, most bZIPs bind significantly to G box and C box sequences, suggesting that the bZIPs are involved in conserved and similar regulatory functions in hexaploid bread wheat. Gene ontology (GO) analysis also shows that bZIPs are involved in growth and development. Previous studies in *Arabidopsis* show that bZIP TFs bZIP10, bZIP25, bZIP53, bZIP72, and bZIP39 are involved in seed development (Lara et al., 2003; Alonso et al., 2009) and their binding site were characterized by *in vitro* methods like PBM, DAP-seq (Franco zorrilla et al., 2014; O'Malley et al., 2016). Furthermore, we observed slight differences in the binding of different bZIP TFs to DNA sequences. bZIP binding in suggests that binding specificity may be determined by factors other than individual TF binding site selection, such as flanking bases, methylation, and nearby region (**Refs.**) The binding of TaABI5 to G-Box is similar to *Arabidopsis* and rice and may predict their conserved evolutionary role in ABA response and abiotic stress during seed development.

Embp-1 is an embryo-specific bZIP TF in wheat involved and is involved in abscisic-acid mediated gene activation. It binds to 10 bp palindromic 'GCCACGTCGG'(G-Box) sequence (Guiltinan, Marcotte et al. 1990, Niu and Guiltinan 1994, Niu, Renshaw-Gegg et al. 1999). Embp-1 has higher sequence specificity for G-box compared to C-box. We analyzed Embp-1 binding in 10 and 150 mM salt condition where it interacts specifically with the 'GCC/GACGTC/GGC' DNA sequence. However,Embp-1 prefers to bind C-box over G-box in other salt and protein concentrations. This may be due to a change in sequences flanking the ACGT core element in the oligo library (De Jong et al., 2013; Gordan et al., 2013).

ABI5 is expressed during early seedling growth,and in the seed maturation phase and also has a vital role in ABA and abiotic stress-mediated responses.It is a core participant in the ABA signaling mechanism, and mutation in ABI5 results in the down regulation of stress-responsive genes (Finkelstein and Lynch, 2000). In rice, OsABI5 shows high homology with ABI5 and barley HvABI5 and can bind G-box (Zou et al., 2008). Bind-n-seq was able to determine comparable *de novo* binding motifs for Hy5 and ABI5 as described previously in other studies, as shown in Fig. 5 (Zhang, He et al. 2011, O'Malley, Huang et al. 2016)

ABI5 shows interacts with ACGTG (G-box) with high specificity. TaABI5 binds specifically to the 'GGGCACGTGCCC' binding site with ACGT core followed by G. In 50 nM protein concentration, TaABI5 prefers C-Box over G-Box.This study gives insights into the binding sites of four bZIP TFs that may have regulatory roles in seed development, abiotic stresses, and other biological processes in hexaploid bread wheat, an important food source worldwide. Surprisingly, our data observed few differences in binding specificity of four bZIPs and show functional redundancy in the binding sites. Previous studies show that specific binding by TFs is also determined by cooperativity, chromatin accessibility, and interaction with other factors, which may explains the differences between *in vitro* and *in vivo* studies (Gordân, Shen et al. 2013, Yin, Morgunova et al. 2017, Yella, Bhimsaria et al. 2018). HvABI5, a barley bZIP TF, is closely related to wheat TaABI5 and shows high sequence similarity to *Arabidopsis* AtABI5 and rice OsABI5. HvABI5 binds to ABA response elements promoters present in promoters of their target genes (Casaretto and Ho,

2003). Comparison of earlier DNA binding studies of AtABI5 and results of TaABI5 have shown similar sequence specificity for G-Box in core sequence with minor changes in flanking bases. The binding of TaABI5 to G-Box is similar to *Arabidopsis* and rice and may predict their conserved evolutionary role in ABA response and abiotic stress during seed development. Earlier studies in *Arabidopsis* and other plants suggests that G-box, compared to other sequences is prominently involved in various regulatory phenomena (Ezer, Shepherd et al. 2017). From our binding data, it was observed that plants bZIP TFs have higher sequence propensity for ACGTG (G-Box) and ACGTC (C-box) (supplementary Table 4–8). The results indicate that promoters regions enriched in G- and C-Box may be a prime targets of these bZIPs (Izawa, Foster et al. 1993, Foster, Izawa et al. 1994, Choi, Hong et al. 2000). TabZIP-1 binds with ACGTC (C- Box) and ACGTG (G-Box) in different salt conditions. For example, in 150mM salt condition, TabZIP-1 interacts with the 'GGCACGTCGC' sequence, while in 50 nm and 100 nM protein concentration, it also binds 'GCCACGTCCC/G' (C-Box) (Fig. 4). Izawa et al. 1993 showed that bZIP has a different binding affinity for G box, C Box, and CG box in core binding sites. bZIP TFs also regulate the expression of ABA-responsive genes, light, abiotic stress, and plant development genes by binding to promoters of genes enriched in G-box elements (Menkens et al., 1995). *Arabidopsis* Hy5 also binds to G-boxes to regulate the expression of genes involved in ABA, anthocyanin accumulation, and light signaling (Gangappa et al., 2016).

To determine the preference for binding these seed-specific bZIPs to other core sequences other than ACGT core, Bind-n-Seq was performed with random core library (N^{18}) under stringent salt (150 mM) and protein (50 nM) conditions. TaEmbp-1, TaABFB, TabZIP-1, and Hy5 bind to the ACGT core even when the random library is used, but surprisingly, TaABI5 binds with GC rich region instead of the ACGT core sequence (Fig. 4). Our results also suggest that two seed-specific bZIPs bind to ACGT core sequences with high affinity, whereas the other two recognize GC-rich sequences. This study, is an effort to advance the field of transcriptional gene regulation by experimentally obtaining the binding landscape of any TF. Understanding gene regulation is vital to essentially all areas of molecular biology, medicine, and crop improvement. In conclusion, this technique can be used in numerous studies to better predict the TF-DNA binding models, DNA-TF interaction landscape, and improving the understanding of the evolution of TFs.

Declarations

SUPPLEMENTARY DATA

Supplementary data are available online.

ACKNOWLEDGEMENT

We thank the Executive Director, National Agri-Food Biotechnology Institute (NABI), Mohali, for research facilities and research funding for the Department of Biotechnology (DBT), New Delhi. This manuscript has been released as a pre-print at research square(Shah et al., 2022). The work is supported by the core grant of the National Agri-Food Biotechnology Institute, Government of India (GOI).

Author Contributions

K.S. and V.R. designed the study. K.S. and V.R. designed the experiments, analyzed the data, and wrote the paper. K.S. conducted the experiments with help from P.J., N.S., and R.K. D.S. has helped in data analysis and statistics. NGS data were analyzed and reviewed by RK and DS. All authors have contributed with edits, seen, and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare that the research was conducted without any commercial or financial relationships construed as a potential conflict of interest.

None of declared.

References

1. Albani, D., et al. (1997). "The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes." Plant Cell **9**(2): 171–184.
2. Badis, G., et al. (2009). "Diversity and complexity in DNA recognition by transcription factors." Science **324**(5935): 1720–1723.
3. Bartlett, A., et al. (2017). "Mapping genome-wide transcription-factor binding sites using DAP-seq." Nat Protoc **12**(8): 1659–1672.
4. Berger, M. F. and M. L. Bulyk (2009). "Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors." Nat Protoc **4**(3): 393–411.
5. Biggin, M.D. (2001) To bind or not to bind. Nature genetics, 28, 303–304.
6. Choi, H., et al. (2000). "ABFs, a family of ABA-responsive element binding factors." J Biol Chem **275**(3): 1723–1730.
7. De Jong, A. T. (2013). "Effect of flanking bases on the DNA specificity of EmBP-1." Biochemistry **52**(5): 786–794.
8. Deppmann, C. D., et al. (2004). "Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs." Nucleic Acids Res **32**(11): 3435–3445.
9. Ezer, D., et al. (2017). "The G-Box Transcriptional Regulatory Code in Arabidopsis." Plant Physiol **175**(2): 628–640.
10. Fassler, J., et al. (2002). "B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners." Genome Res **12**(8): 1190–1200.
11. Finkelstein, R. R., et al. (2002). "Abscisic acid signalling in seeds and seedlings." Plant Cell **14 Suppl**: S15-45.
12. Foster, R., et al. (1994). "Plant bZIP proteins gather at ACGT elements." Faseb j **8**(2): 192–200.

13. Franco-Zorrilla, J. M. and R. Solano (2017). "Identification of plant transcription factor target sequences." Biochim Biophys Acta Gene Regul Mech **1860**(1): 21–30.
14. Gangappa, S. N. and J. F. Botto (2016). "The Multifaceted Roles of HY5 in Plant Growth and Development." Mol Plant **9**(10): 1353–1365.
15. Geertz, M. and S. J. Maerkl (2010). "Experimental strategies for studying transcription factor-DNA binding specificities." Brief Funct Genomics **9**(5–6): 362–373.
16. Gordân, R., et al. (2013). "Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape." Cell Rep **3**(4): 1093–1104.
17. Guiltinan, M. J., et al. (1990). "A plant leucine zipper protein that recognizes an abscisic acid response element." Science **250**(4978): 267–271.
18. Guiltinan, M. J. and L. Miller (1994). "Molecular characterization of the DNA-binding and dimerization domains of the bZIP transcription factor, EmBP-1." Plant Mol Biol **26**(4): 1041–1053.
19. Hurst, H. C. (1994). "Transcription factors. 1: bZIP proteins." Protein Profile **1**(2): 123–168.
20. International Wheat Genome Sequencing, C., et al. (2018). "Shifting the limits in wheat research and breeding using a fully annotated reference genome." Science **361**(6403).
21. Izawa, T., et al. (1993). "Plant bZIP protein DNA binding specificity." J Mol Biol **230**(4): 1131–1144.
22. Jakoby, M., et al. (2002). "bZIP transcription factors in Arabidopsis." Trends Plant Sci **7**(3): 106–111.
23. Johnson, D. S., et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." Science **316**(5830): 1497–1502.
24. Jolma, A., et al. (2010). "Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities." Genome Res **20**(6): 861–873.
25. Lai, X., et al. (2019). "Building Transcription Factor Binding Site Models to Understand Gene Regulation in Plants." Mol Plant **12**(6): 743–763.
26. Lambert, N., et al. (2014). "RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins." Mol Cell **54**(5): 887–900.
27. Lambert, N. J., et al. (2015). "RNA Bind-n-Seq: Measuring the Binding Affinity Landscape of RNA-Binding Proteins." Methods Enzymol **558**: 465–493.
28. Li, D., et al. (2015). "Genome-wide systematic characterization of the bZIP transcriptional factor family in tomato (Solanum lycopersicum L.)." BMC Genomics **16**: 771.
29. Li, X., et al. (2015). "Genome-wide identification and evolutionary analyses of bZIP transcription factors in wheat and its relatives and expression profiles of anther development related TabZIP genes." BMC Genomics **16**: 976.
30. Liu, X. and Z. Chu (2015). "Genome-wide evolutionary characterization and analysis of bZIP transcription factors and their expression profiles in response to multiple abiotic stresses in Brachypodium distachyon." BMC Genomics **16**: 227.
31. Llorca, C. M., et al. (2014). "bZIPs and WRKYs: two large transcription factor families executing two different functional strategies." Front Plant Sci **5**: 169.

32. Nijhawan, A., et al. (2008). "Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice." Plant Physiol **146**(2): 333–350.
33. Niu, X. and M. J. Guiltinan (1994). "DNA binding specificity of the wheat bZIP protein EmBP-1." Nucleic Acids Res **22**(23): 4969–4978.
34. Niu, X., et al. (1999). "Bipartite determinants of DNA-binding specificity of plant basic leucine zipper proteins." Plant Mol Biol **41**(1): 1–13.
35. O'Malley, R. C., et al. (2016). "Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape." Cell **165**(5): 1280–1292.
36. Pourabed, E., et al. (2015). "Basic leucine zipper family in barley: genome-wide characterization of members and expression analysis." Mol Biotechnol **57**(1): 12–26.
37. Ramirez-Gonzalez, R. H., et al. (2018). "The transcriptional landscape of polyploid wheat." Science **361**(6403).
38. Riechmann, J. L., et al. (2000). "Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes." Science **290**(5499): 2105–2110.
39. Rockel, S., et al. (2012). "MITOMI: a microfluidic platform for in vitro characterization of transcription factor-DNA interaction." Methods Mol Biol **786**: 97–114.
40. Sharrocks, A. D. (1994). "A T7 expression vector for producing N- and C-terminal fusion proteins with glutathione S-transferase." Gene **138**(1–2): 105–108.
41. Skubacz, A., et al. (2016). "The Role and Regulation of ABI5 (ABA-Insensitive 5) in Plant Development, Abiotic Stress Responses and Phytohormone Crosstalk." Frontiers in Plant Science **7**(1884).
42. Smaczniak, C., et al. (2017). "SELEX-Seq: A Method to Determine DNA Binding Specificities of Plant Transcription Factors." Methods Mol Biol **1629**: 67–82.
43. Song, Y. H., et al. (2008). "DNA-binding study identifies C-box and hybrid C/G-box or C/A-box motifs as high-affinity binding sites for STF1 and LONG HYPOCOTYL5 proteins." Plant Physiol **146**(4): 1862–1877.
44. Stormo, G. D. and Y. Zhao (2010). "Determining the specificity of protein-DNA interactions." Nat Rev Genet **11**(11): 751–760.
45. Todeschini, A. L., et al. (2014). "Transcription factors: specific DNA binding and specific gene regulation." Trends Genet **30**(6): 211–219.
46. Vinson, C., et al. (2002). "Classification of human B-ZIP proteins based on dimerization properties." Mol Cell Biol **22**(18): 6321–6335.
47. Wang, Z., et al. (2015). "Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes." BMC Genomics **16**: 1053.
48. Wei, K., et al. (2012). "Genome-wide analysis of bZIP-encoding genes in maize." DNA Res **19**(6): 463–476.

49. Yella, V. R., et al. (2018). "Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif." Nucleic Acids Res **46**(22): 11883–11897.
50. Yin, Y., et al. (2017). "Impact of cytosine methylation on DNA binding specificities of human transcription factors." Science **356**(6337).
51. Zhang, H., et al. (2011). "Genome-wide mapping of the HY5-mediated gene networks in Arabidopsis that involve both transcriptional and post-transcriptional regulation." Plant J **65**(3): 346–358.
52. Zhu, J. Y., et al. (2012). "Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq)." Methods Mol Biol **876**: 173–188.
53. Zykovich, A., et al. (2009). "Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing." Nucleic Acids Res **37**(22): e151.

Figures

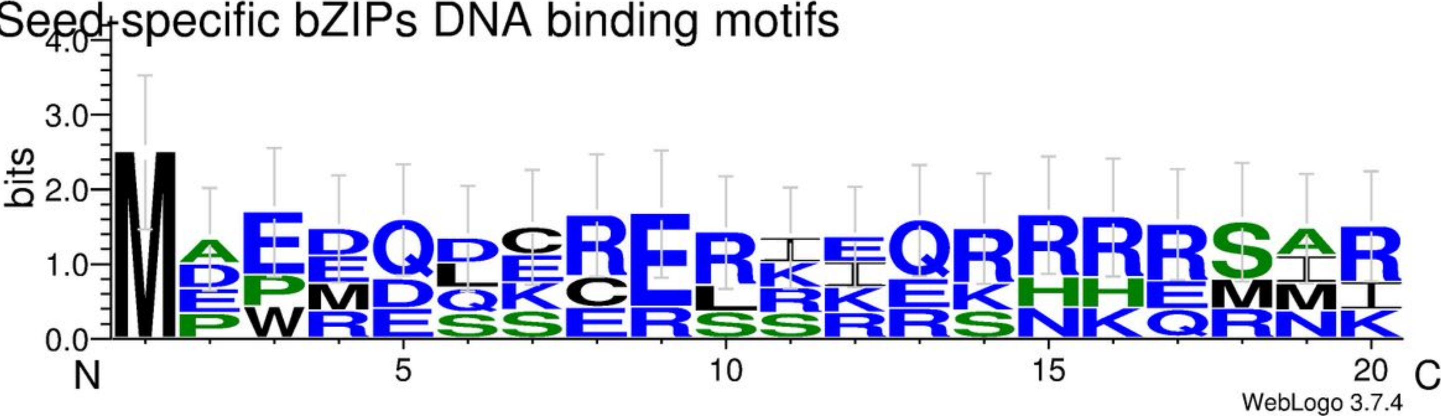

Figure 1

Visualization of multiple sequence alignment of DNA binding domains(DBDs) hexaploid wheat seed-specific bZIPs.The sequence logo shows the basic region (motif 1–20) of bZIPs. The bZIP domain contains a basic DNA-binding region and an adjacent leucine zipper structure. The basic region has an invariable N-X7-R/K motif, while the zipper region is composed of a heptapeptide repeat of Leucine (L). The total height of the letter piles at each position indicates the conservation of the sequence at that position (measured in bits). The height of a single letter in the letter piles represents the relative frequency of the corresponding amino acid at that position. For each logo, multiple sequence alignments of respective protein sequences were subjected to an online tool “ WebLogo” (Crooks 2004)

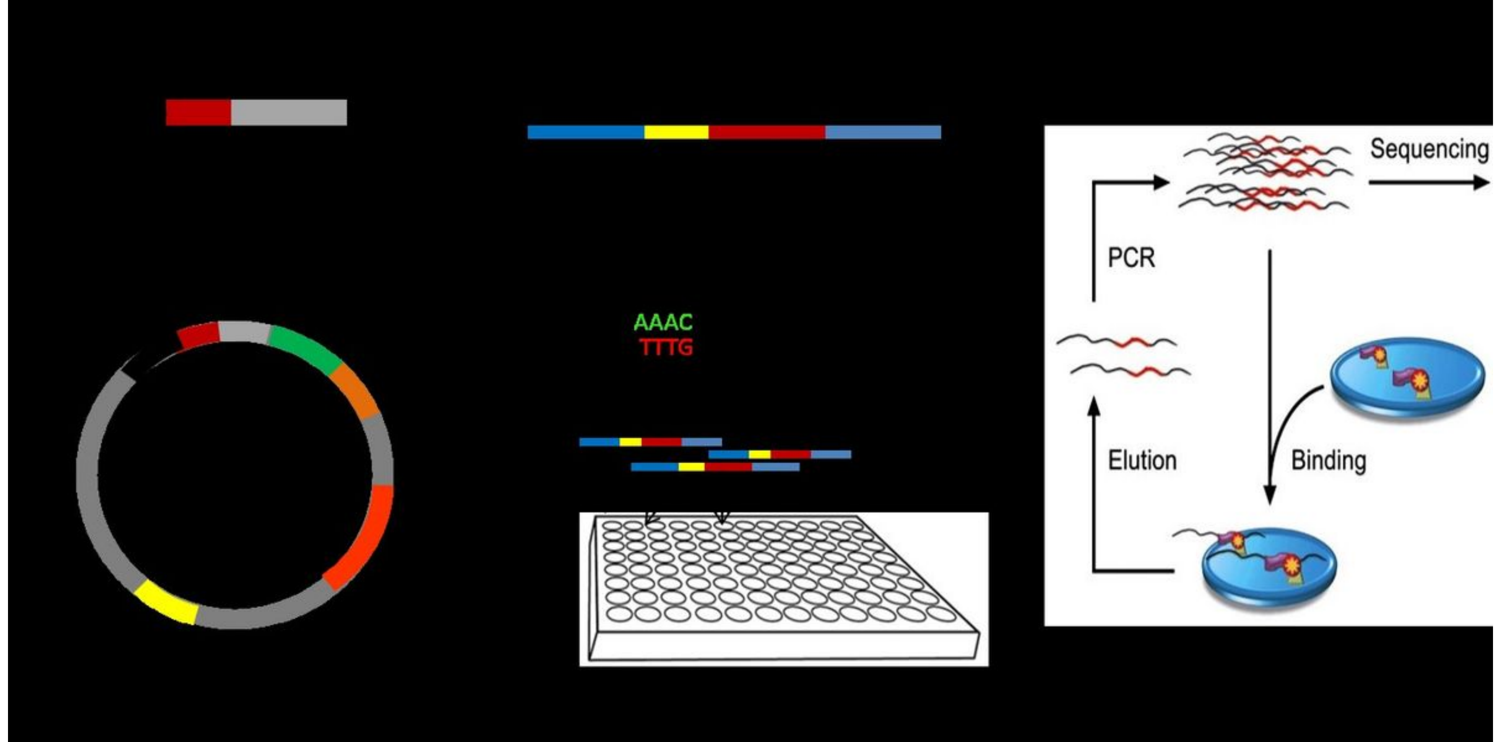

Figure 2

Outline of Bind-n-seq(BNS) experimental technique used for bZIP binding studies in hexaploid wheat (A) DNA-binding domains(DBDs) of bZIPs are cloned and expressed with GST-fusion tag in T7 promoter based expression vector (B) GST-tagged bZIP protein is incubated with a random pool of DNA oligonucleotides with same concentration at different concentrations of salt and proteins (C) After binding, bound target DNA is pulled down using affinity purification and associated DNA is amplified and sequenced. The counts of sequences in this library estimate the fraction of bound DNA molecules compared to input DNA, which was also sequenced.

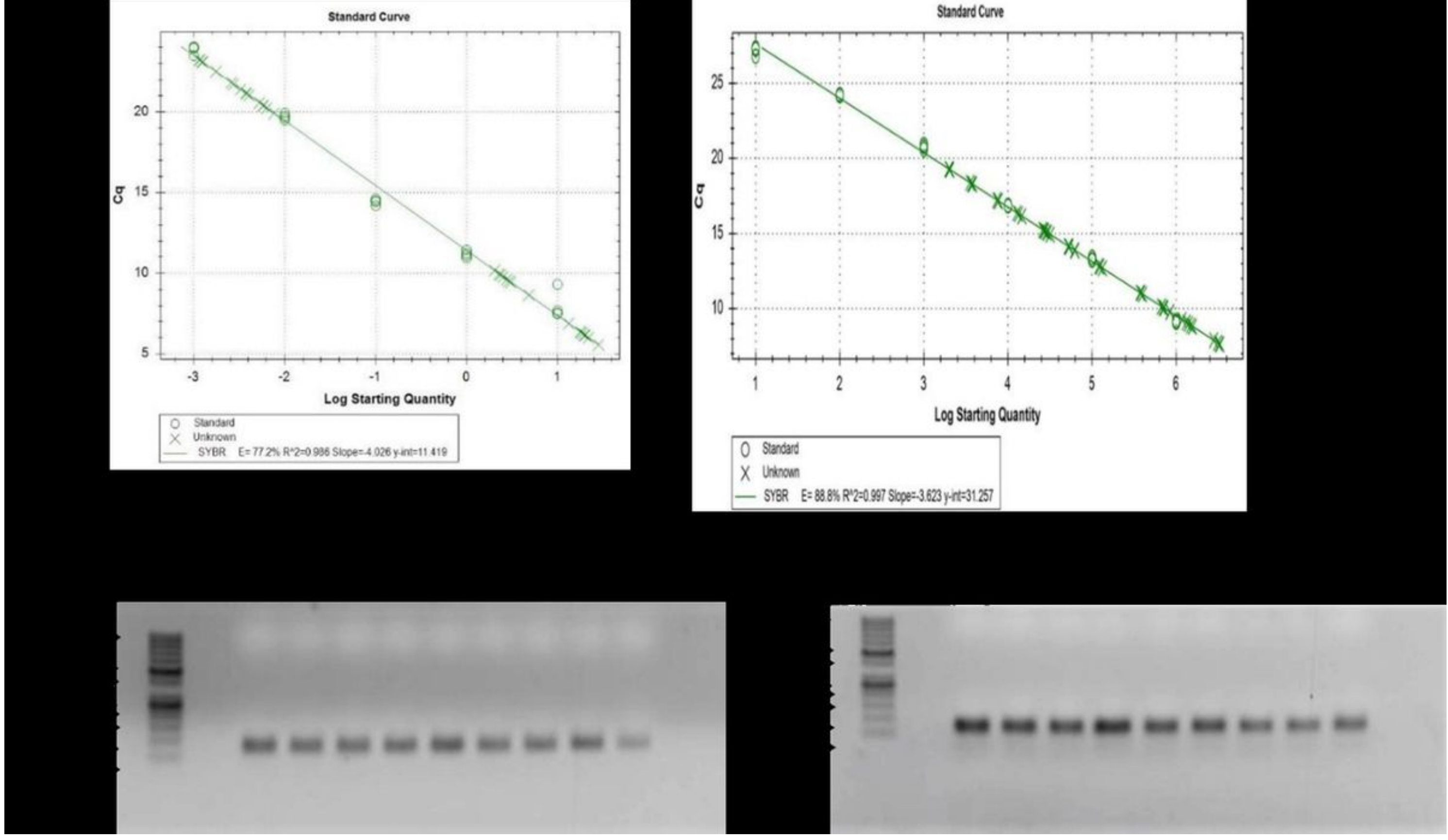

Figure 3

Absolute quantification of TF-bound output DNA from Different binding conditions using qRT-PCR and PCR enrichment for Multiplexing of samples(A) TF-bound DNA sequences after elution from each experiment were quantified using qRT-PCR to normalize the sample concentration for sequencing. A standard curve was generated using template DNA (A) Standard curve was generated for total output DNA quantification using different dilutions of control template DNA at 150,15,1.5 and 0.5 nM.(B)The output DNA concentration after the adaptor ligation step was measured using Kapa Library Quantification Kit. The dilutions of 1:2000, 1:4000, 1:8000, and 1:16000 of standard libraries generated a standard curve. (C) PCR amplification of Hy5 binding reactions for enrichment (D) PCR amplification of TabZIP-1 protein-bound DNA. 10µl of PCR reactions were checked on 2.5 % agarose gel

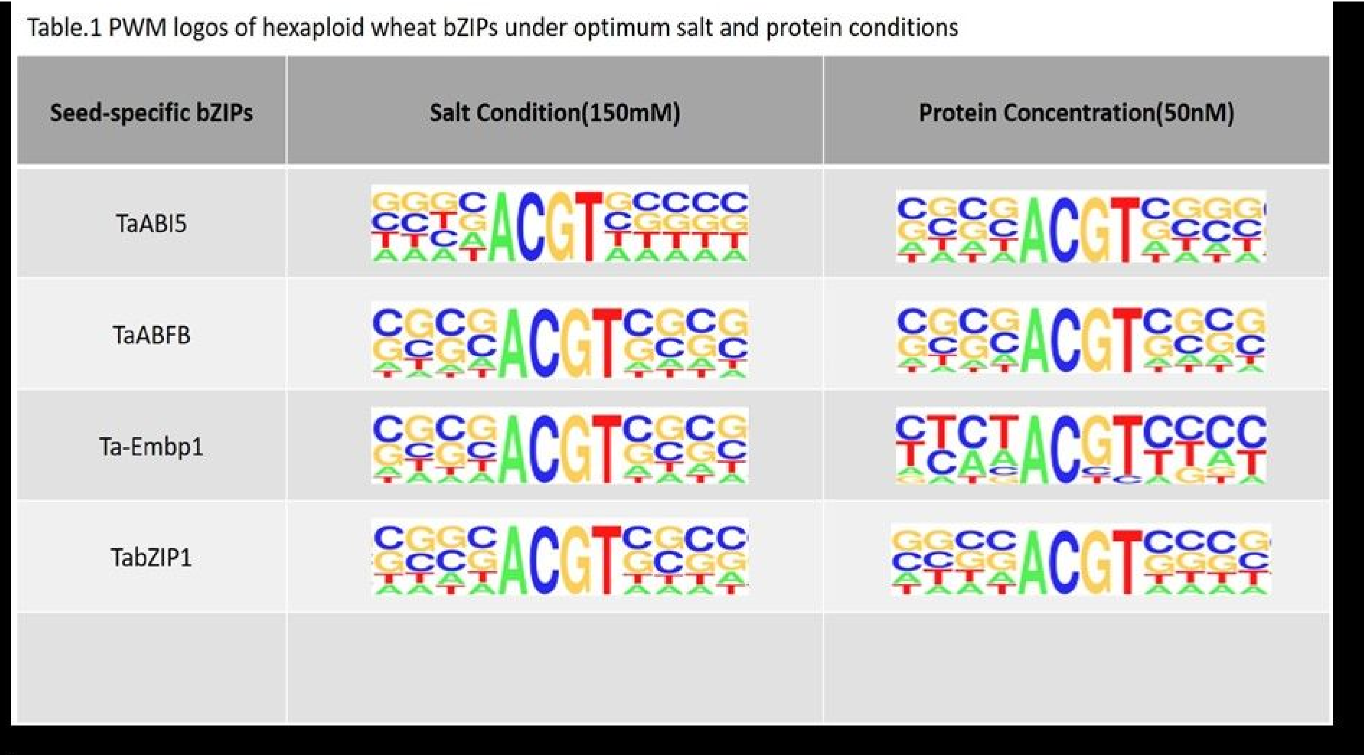

Table.1 PWM logos of hexaploid wheat bZIPs under optimum salt and protein conditions

Seed-specific bZIPs	Salt Condition(150mM)	Protein Concentration(50nM)
TaABI5		
TaABFB		
Ta-Embp1		
TabZIP1		

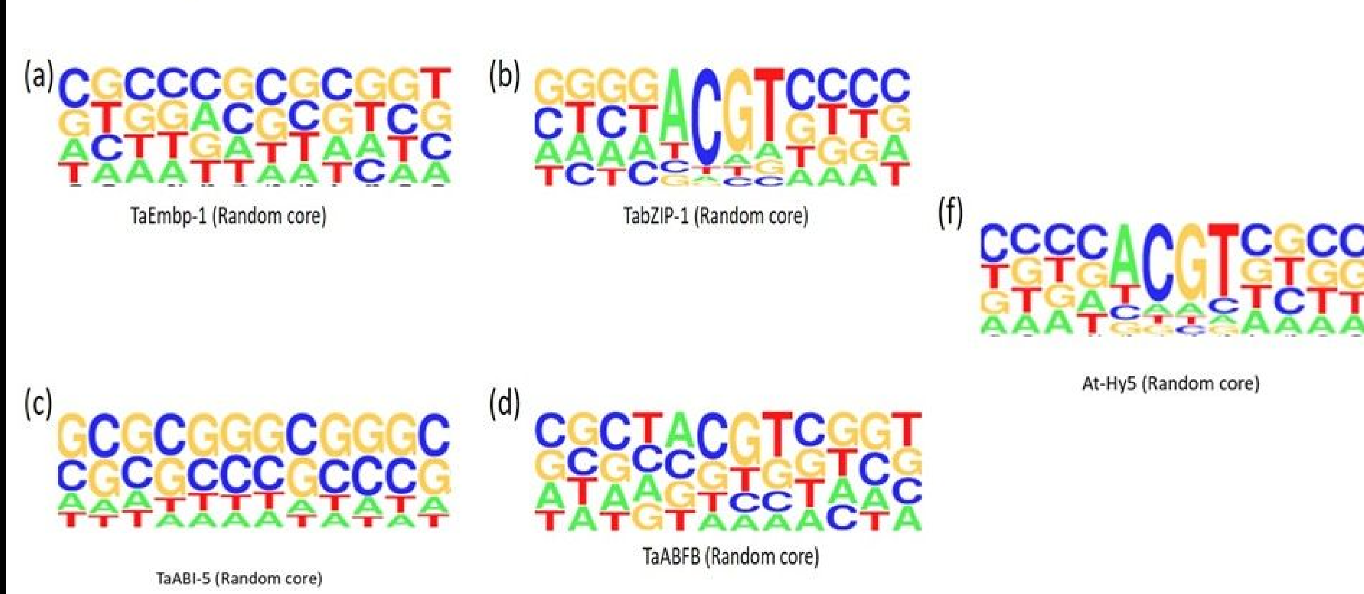

Figure 4

Characterization of seed-specific bZIPs binding sites under different salt and protein concentration. Sequence logos for binding sites from (A) PWM logos and sites for TaABI5, TaABFB, TaEmbp-1, and TabZIP-1 are shown under 150mM salt and 50nM protein concentrations where pull-down sequences are highly enriched in comparison to other conditions (B) PWM logos of bZIPs obtained using randomized region libraries (Without ACGT) used for binding under stringent 150mM salt and 50nM protein conditions

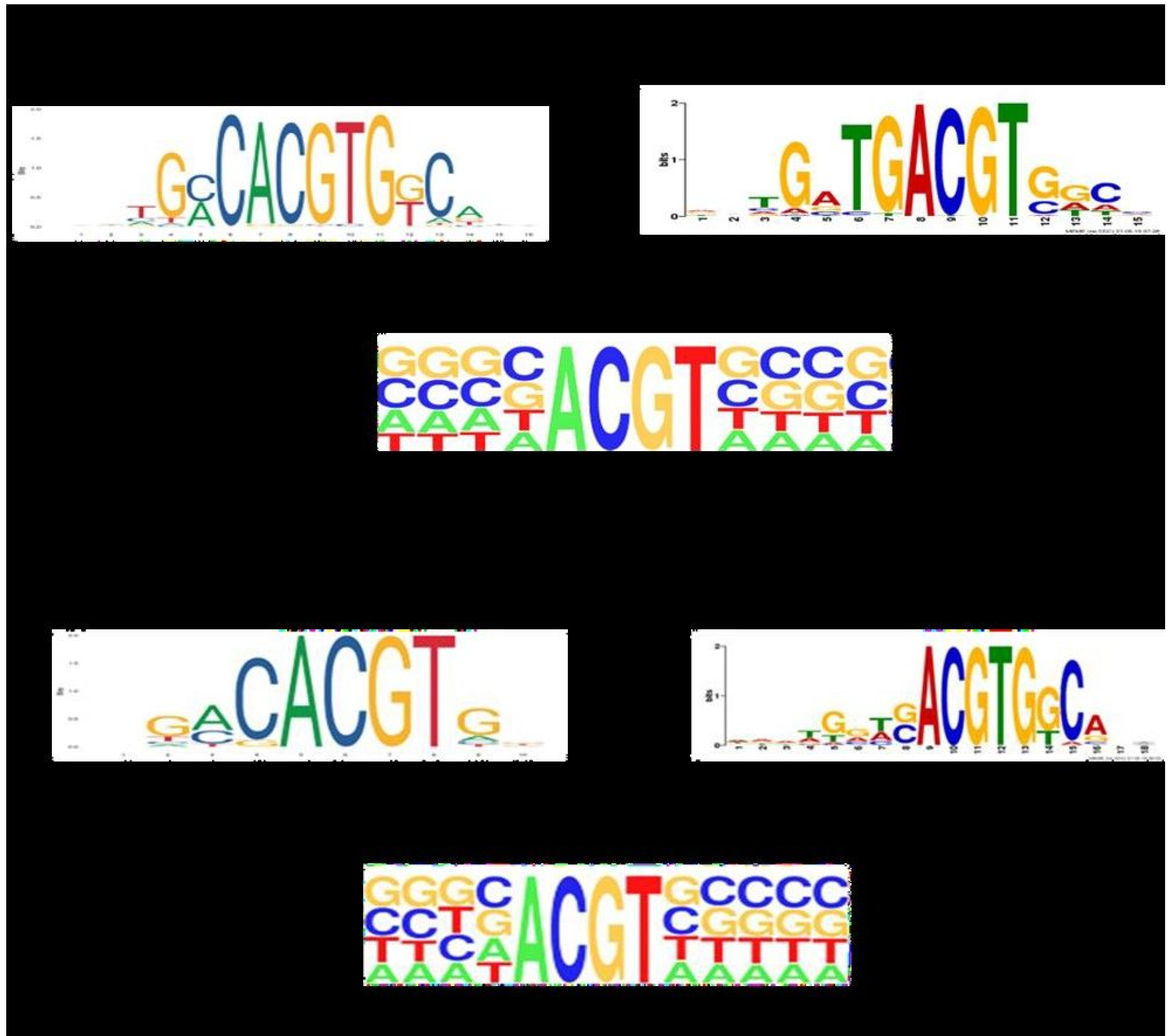

Figure 5

Figure shows binding specificity of Hy5 using the Bind-n-seq technique and comparison of TaABI5 binding to other Arabidopsis ortholog(A). Describes the comparison of At-Hy5 bZIP PWM models in different in vivo (ChIP-chip) and in vitro (DAP-Seq, Bind-n-Seq) DNA protein interaction studies in Arabidopsis. (B) PWM models of TaABI5 protein binding under conditions of salt 150 mM and protein 50 nM, earlier reported binding sites of ATABI5 in JASPER and cistrome databases.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.docx