

SCIENTIFIC REPORTS



OPEN

Pivotal role of bZIPs in amylose biosynthesis by genome survey and transcriptome analysis in wheat (*Triticum aestivum* L.) mutants

Pankaj Kumar^{1,2}, Ankita Mishra^{1,2}, Himanshu Sharma¹, Dixit Sharma³, Mohammed Saba Rahim^{1,4}, Monica Sharma¹, Afsana Parveen^{1,2}, Prateek Jain^{1,2}, Shailender Kumar Verma³, Vikas Rishi¹ & Joy Roy¹

Starch makes up 70% of the wheat grain, and is an important source of calories for humans, however, the overconsumption of wheat starch may contribute to nutrition-associated health problems. The challenge is to develop resistant starch including high amylose wheat varieties with health benefits. Adapting advance genomic approaches in EMS-induced mutant lines differing in amylose content, basic leucine zipper (bZIP) regulatory factors that may play role in controlling amylose biosynthesis were identified in wheat. bZIP transcription factors are key regulators of starch biosynthesis genes in rice and maize, but their role in regulating these genes in wheat is poorly understood. A genome-wide survey identified 370 wheat bZIPs, clustered in 11 groups, showing variations in amino acids composition and predicted physicochemical properties. Three approaches namely, whole transcriptome sequencing, qRT-PCR, and correlation analysis in contrasting high and low amylose mutants and their parent line identified 24 candidate bZIP (positive and negative regulators), suggesting bZIPs role in high amylose biosynthesis. bZIPs positive role in high amylose biosynthesis is not known. *In silico* interactome studies of candidate wheat bZIP homologs in *Arabidopsis* and rice identified their putative functional role. The identified bZIPs are involved in stress-related pathways, flower and seed development, and starch biosynthesis. An in-depth analysis of molecular mechanism of novel candidate bZIPs may help in raising and improving high amylose wheat varieties.

Cereal grains largely contain starch^{1,2}, and are important sources of calories for humans. Starch is highly digestible, about 99% is digested in human gut and converted into glucose, raising the glycemic index. Its overconsumption causes nutrition-associated health problems³⁻⁵. There is a surge to develop cereal crops with resistant starch and food grains rich in dietary fibre^{6,7}. Resistant starch is categorized as 'good dietary fibre'³. Due to its slow digestion in gut, high amylose starch is categorized as resistant or healthy starch that results in low glycemic index. Amylose is further converted into short chain fatty acids (SCF) by bacteria in large intestine^{4,5}. SCFs are known prebiotics with proven health benefits^{4,5}. Functional genomics approaches are used to identify regulatory factors controlling amylose biosynthesis that may be manipulated to increase its content in grains⁸⁻¹⁰. In this study, genome-wide analysis of contrasting mutant lines for amylose content led to the identification of candidate bZIPs that correlate with the expression of two key genes of amylose biosynthesis pathway i.e., Granule-bound starch synthase I (GBSSI) and starch branching enzyme II (SBEII).

Briefly, starch is a semi-crystalline structure composed of two fractions, amylose and amylopectin^{11,12}. Amylose is largely linear chain of glucose moiety whereas amylopectin is highly branched. Amylose is a non-digestible or slow-digestible fraction of starch and thus considered as resistant starch. In cereal, amylose

¹National Agri-Food Biotechnology Institute (NABI), Sector-81, SAS Nagar, Mohali, 140306, Punjab, India.

²Department of Biotechnology, Panjab University, Chandigarh, 160014, India. ³Centre for Computational Biology and Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Kangra, 176206, Himachal Pradesh, India. ⁴Department of Plant Sciences, School of Basic and Applied Sciences, Central University of Punjab, Bathinda, 151001, India. Correspondence and requests for materials should be addressed to J.R. (email: joykroy@nabi.res.in)

⁵Department of Plant Sciences, School of Basic and Applied Sciences, Central University of Punjab, Bathinda, 151001, India. Correspondence and requests for materials should be addressed to J.R. (email: joykroy@nabi.res.in)

makes ~25% of total starch. In this study, two contrasting mutant lines differing in amylose content were used for genome-wide analysis. High amylose mutant line ‘TAC 75’ with 65% amylose content and low amylose mutant line ‘TAC 6’ with 7% amylose content along with the parent line ‘C 306’ with ~26% amylose content were investigated. The key enzymes involved in starch biosynthesis are ADP-glucose pyrophosphorylase, starch synthases (soluble and granule-bound), and starch branching and debranching enzymes^{11–13}. GBSSI is largely responsible for amylose biosynthesis^{14,15}, whereas SBEII is responsible for amylopectin biosynthesis^{16,17}. The differential expression of GBSSI and SBEII have been correlated with high amylose biosynthesis^{18,19}. Unlike in other plants, their regulation by bZIPs is largely unknown in wheat. bZIP transcription factor family members are well-known for their roles in growth and development^{20,21}. bZIPs are also involved in the regulation of starch biosynthesis in the endosperm that determine starch quality and quantity^{22–25}. OsbZIP58 in rice²³, ZmbZIP91 in maize²⁶ and bZIP58 in wheat¹⁸ are reported to regulate starch biosynthesis. However, their role in amylose biosynthesis in wheat is not well-understood.

Genome-wide analysis identified many bZIP family members (genes) in various plant species like maize²⁷, cucumber²⁸, rice²⁹ and *Arabidopsis*³⁰. For example, there are 94 bZIPs identified in *Oryza sativa subsp. indica*, 140 in *O. sativa subsp. japonica*, 216 in *Zea mays*, 127 in *Arabidopsis thaliana*, and 187 in wheat³¹. Due to its large genome size and lack of complete genome information, the structural and functional characterization of bZIPs is lacking in wheat. Publically available genome sequence of wheat can be used for the genome-wide analysis of bZIPs and their roles in high amylose biosynthesis. The sequence based structured domain information of bZIPs can be used to identify their putative functional role by *in silico* analysis using the published validated functional roles, phylogenetic group information, and interactome analysis in the other plant species such as *Arabidopsis*.

In the present study, genome-wide analysis and phylogenetic analysis of bZIPs were undertaken in wheat. The 284 Gb transcriptome sequence data was generated from the two contrasting mutant lines, ‘TAC 75’ and ‘TAC 6’ and their parent variety, ‘C 306’. Transcriptome analysis, qRT-PCR data, and correlation analysis identified candidate wheat bZIPs (TabZIPs) regulating high amylose biosynthesis. The putative functional role of the candidate TabZIPs in amylose biosynthesis were predicted by using phylogenetic group information and protein interacting networks of *Arabidopsis* databases.

Results

Identification and characterization of wheat bZIPs. *Genome survey of bZIP transcription factors in wheat.* Using whole genome sequence databases, wheat bZIPs TFs were identified through sequence similarity match with maize, rice, barley and *Arabidopsis*. A total of 370 wheat bZIPs (TabZIPs) were identified by a Hidden Markov Model (HMM) profile ‘PF00170’ search against the whole wheat proteome ensembl database by HMMER3.0 and BLASTp search using plant bZIP sequences. Subsequently, after validating the integrity of the bZIP domain using NCBI-CDD and InterProScan, a total of 370 wheat bZIP proteins encoded by 238 bZIP genes were identified. Each wheat bZIP protein was assigned a unique identifier from TabZIP1 to TabZIP370. The gene isoforms and their proteins were assigned the same gene/protein number with decimal point. The information regarding TabZIP transcription factors is listed in Supplementary Table S1.

Phylogenetic analysis of TabZIPs. The sequence homology relationship of 370 TabZIP proteins with that of *Arabidopsis*, maize, rice, and barley was analysed by multiple sequence alignment. The analysis revealed a low level of variation (Fig. 1, Supplementary Fig. S1). The phylogenetic analysis grouped the 370 TabZIP proteins in 11 clades (Fig. 2), and named as groups A to I and S to U, in accordance with those reported in *Arabidopsis*³⁰. Two groups, A and D, were the largest, each with 83 TabZIPs followed by group C (48 TabZIPs), group I (44 TabZIPs), group G (32 TabZIPs), group S (19 TabZIPs), group B, E and F (each with 14 TabZIPs), group H (9 TabZIPs), while group U is the smallest with only 5 TabZIPs.

TabZIP physicochemical properties and conserved motifs analysis. The 370 predicted TabZIPs varied in their amino acid composition ranging from 129 (TabZIP48.1 and TabZIP94) to 920 (TabZIP106.1) residues with a molecular mass from 14 kDa (TabZIP94) to 103 kDa (TabZIP106.1). Their theoretical isoelectric point (pI) varied between 4.65 (TabZIP198.1 and TabZIP219.1) to 11.38 (TabZIP133). The grand average of hydropathicity (GRAVY) of each amino acid residue in the 370 TabZIP proteins was very low, which indicates better interaction between TabZIP proteins and water molecules. The information regarding the physicochemical properties of TabZIPs is provided in Supplementary Table S2. In TabZIPs, ten motifs were identified by the MEME database that gave insight into their function and divergence (Table 1, Fig. 3, Supplementary Fig. S2, and Supplementary Table S3). Motif 1 was present in all the TabZIP proteins, while motif 7 was present in all the groups except group D. Group D comprised all motifs except motif 7 and 8. Motif 8 was shared by groups A, B, C, E, G and I. The results showed that the TabZIPs share similar sequences and clustered in the same group.

Analysis of cis-regulating elements of starch pathway genes and prediction of DNA binding domain (DBD) of TabZIPs. To understand the mechanism of transcriptional regulation of wheat starch amylose and amylopectin biosynthesis pathway genes, their promoter regions were analysed for the cis-regulating element using the available online wheat genome sequences datasets. It has been previously reported that in plants bZIPs prefer to bind ACGT core sequence like G-box (CACGTG), C-box (GACGTC), and A-box (TACGTA) motifs. Promoter analysis showed the presence of A and G boxes which are the putative sites for bZIP DNA binding domains. To analyse the cis-regulating region, up to 1.0 kb sequence upstream to open reading frames of starch pathways genes (GBSSI, GBSSII, SSI, SSII, SSIII, SSIV, SBEI, SBEIIa and SBEIIb) were identified. Five (GBSSI), seven (GBSSII), four (SSI), four (SSII), six (SSIII), one (SSIV), thirteen (SBEI), six (SBEIIa), and one (SBEIIb) ACGT core sequence were found in the up-stream sequences of these genes. Three A boxes were identified in SSII and one A box was identified in GBSSII and SSIV (Supplementary Fig. S3).

Figure 1

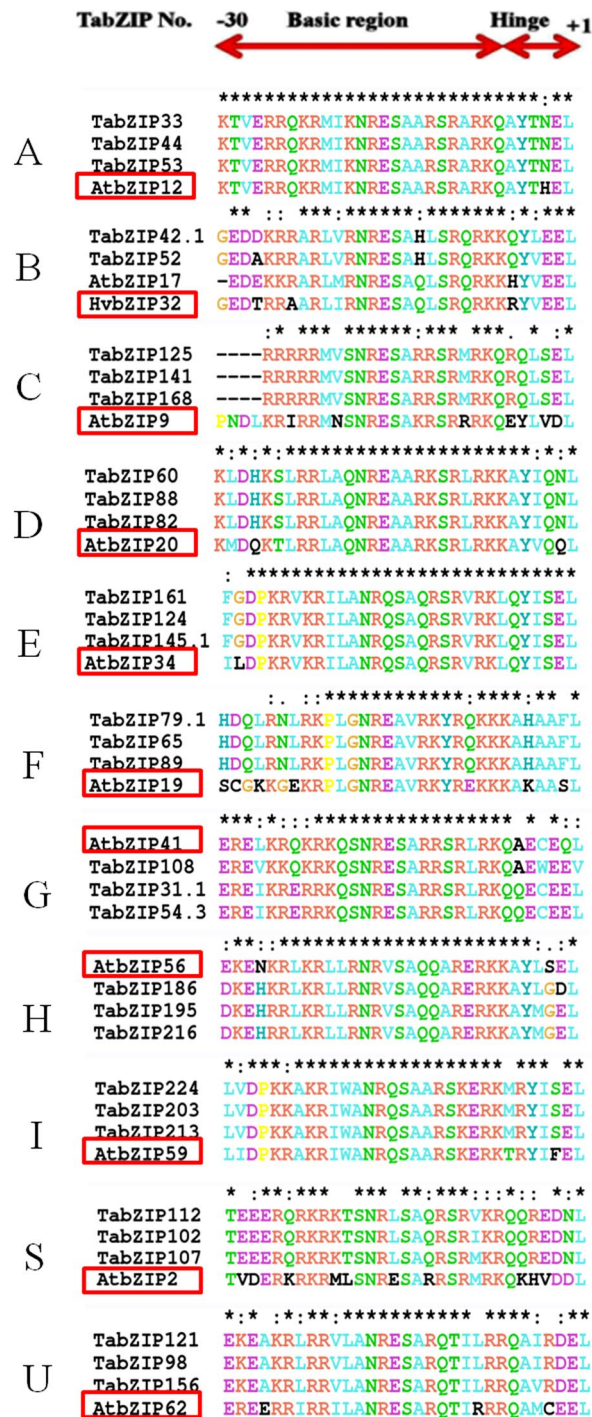


Figure 1. Multiple sequence alignment (MSA) of basic and hinge regions of representative wheat bZIP (TabZIP) proteins. The alignment of the amino acid code was generated using multiple sequence alignment with one homolog of *Arabidopsis* (AtbZIP) or barley (HvbZIP) along with TabZIP. The reference homologs are displayed in red boxes, the asterisk (*) and colon (:) above the alignment represent the conserved and variable region, respectively. The detail of MSA is provided in Supplementary Fig. S1.

Identification of candidate TabZIPs regulating amylose biosynthesis in wheat. Three approaches, namely genome-wide transcriptome sequencing, candidate genes based qRT-PCR, and a statistical correlation analysis of GBSSI and SBEII in the contrasting mutant lines for amylose content, were applied for the identification of candidate TabZIPs for high amylose biosynthesis.

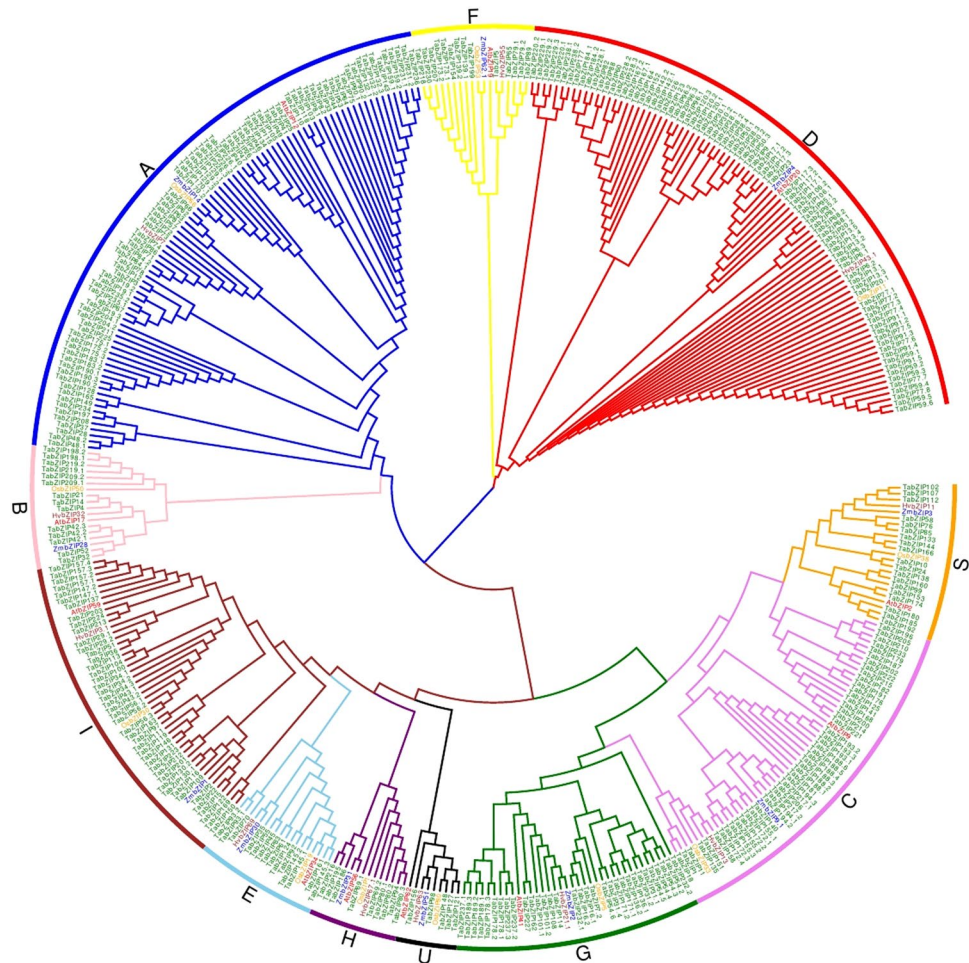


Figure 2. Phylogenetic tree of 370 wheat bZIP (TabZIP) proteins and representative bZIPs of *Arabidopsis* (AtbZIPs), maize (ZmbZIP), barley (HvbZIP), and rice (OsbZIP). The phylogeny was determined by the neighbour joining method using amino acid sequences in MEGA 6.0, at 1000 bootstrap. The tree was divided into 11 clades (A–I and S,U), represented by different colors. Details of acronyms used in the tree are given in Supplementary Table S1.

Sequence characterization and functional analysis of GBSSI, SBEIIa, and SBEIIb in mutants. The sequences of GBSSI, SBEIIa, and SBEIIb, which were retrieved from transcriptome sequence data, identified both synonymous and non-synonymous mutations in their coding regions (Supplementary Fig. S4). The mutations were also detected in their homoeologous loci. Three, one, and two non-synonymous mutation frequency were identified in the catalytic domain of GBSSI-2BL, GBSSI-7AS, and GBSSI-7DS, respectively. In SBEII isoforms (SBEIIa and SBEIIb), mutation frequency was high and the sequencing errors cannot be ruled out. Both SBEII isoforms were located at three homoeologous chromosomes i.e. 2AL, 2BL, and 2DL. The frequency of mutations in SBEIIa and SBEIIb were relatively higher than GBSSI (Supplementary Fig. S4).

The comparative gene expression analysis of GBSSI, SBEIIa and SBEIIb starch metabolic genes at three development stages (21, 28 and 35 days after anthesis, DAA) revealed variation in their expression levels. GBSSI showed high expression during seed development in the high amylose mutant line compared to the low amylose mutant line (Supplementary Fig. S5). During seed development, SBEII isoforms (SBEIIa and SBEIIb) showed very low expression levels in the high amylose mutant line in comparison to that of low amylose mutant line. The higher expression level of GBSSI and lower expression of both isoforms of SBEIIa and SBEIIb in the high amylose mutant line were correlated with increased amylose content during seed development (Supplementary Fig. S5).

Transcriptome sequencing of mutant lines. FPKMs (Fragments Per Kilobase of transcript per Million mapped reads) value is a normalization method for gene expression study. FPKMs of the 370 TabZIP genes were retrieved from the transcriptomic sequence data. The sequence data was generated on two biological replicates of the developing seeds (28 days after anthesis) belonging to the two mutant lines, ‘TAC 75’ with 65% amylose (high amylose mutant line) and ‘TAC 6’ with 7% amylose (low amylose mutant line), and the parent wheat variety, ‘C 306’ with 26% amylose. The genes having FPKM values of at least 0.02 were considered to be expressed and were used for differential gene expression analysis (Supplementary Table S4A, S4B). Considering an FPKM value of 0.02 as the cutoff point, 81 out of 370 (~22%) TabZIPs did not show expression in any three genotypes i.e. 289 (78%)

Motif number	Multiple consensus sequences	Number of TabZIP proteins with motif
1	[KR]RQ[RK]R[ML][ALI][SQK]NRE[SA]A[RA][RK]SR[LE]RK[QK]AY	365
2	DVFH[LV][LM][ST]G[MA]WA[TS]PAER[CF]F[LF]W[LM]GGFRPSE[LV]LK[IVL]L[IA][GP]	80
3	F[YVL][RQ]QADNLR[QL]QTL[HQ]QM[RH]RILTTTQAAR[CA][FL][LV][SVA][IL][GS][DE]Y[FY][RS]RLRALSSLV[AL][AS]RP	72
4	[MA]FD[MV]EYARW[LV][DE][ED][DH][NG][KR][RH][MIL][AN]ELR[GA][AG][LV][QN]AH[LA][AG]DS[DEN]L[GR]AIV[ED]EC[ML]	82
5	QL[ED]PLTEQQ[LM][MV]GI[CY][NG]LQ[QH]SS[EQ]QAE[ED]AL[SA]QG[LM][QE][QA]L[HQ]QSL[AS][DE]T[VL]A[AS]GTL[NA][DS]G	76
6	[VI]Q[QN]LE[TS]SR[IVL][RK]L[QA]Q[LMI]EQELQRRAR[QS]QGI[FL][LI][GS]G[GS]G[AD][GQ]GD[MSL]S[SP][GA]A	82
7	ELER[EKQ]VSLXRAENXXLXK[RQE]LX	268
8	GX[PT][LF][GS]SM[NT][MLV][DE]E[FL][LW]RNIWX[AV]EE	93
9	[NPD][VG][AP]NY[MT][GA][QI]MA[IL]A[LM][GE]KL[AG][STN]LE[SN]	63
10	[AQ][GE][RQ]P[PQ]TL[EN]IFPSWPM[PH]HPQQ[LP]H[SP]	43

Table 1. Details of conserved motifs identified in wheat bZIPs (TabZIP) proteins using MEME database. A set of 10 motifs were identified in TabZIPs by their consensus sequence analysis. The motif number was assigned following Bailey *et al.* (2009). The detail is provided in Supplementary Table S3.

TabZIPs showed expression in at least one genotype at a given time. Individual genotype data showed that 236 (63.8%) TabZIPs showed expression in the high amylose mutant, 241 (63.6%) in the low amylose mutant, and 226 (61.1%) in their parent variety. In pairwise comparison, 177 TabZIPs showed expression in all the three genotypes as shown in the Venn diagram (Fig. 4A). The graph also revealed that 12 TabZIPs were only expressed in the high amylose mutant and therefore unique to the mutant line. The twelve TabZIPs were TabZIP15, TabZIP50.1, TabZIP54.5, TabZIP56.3, TabZIP91.1, TabZIP91.2, TabZIP120.2, TabZIP128, TabZIP167.2, TabZIP173.2, TabZIP184.2, and TabZIP220.2. The FPKM values of these 12 TabZIPs were below 0.5, indicating low expression.

The pairwise differential gene expression analysis of TabZIPs was done among three genotypes (two mutant lines and one parent variety), using log 2-fold of mean FPKM data (two biological replicates). It involved three pairs, ‘TAC 75’ vs ‘TAC 6’, ‘TAC 75’ vs ‘C 306’, and ‘TAC 6’ vs ‘C 306’ (Fig. 4B). Out of the three, only two pairs (‘TAC 75’ vs ‘TAC 6’ and ‘TAC 75’ vs ‘C 306’) were used to identify candidate TabZIPs that may be involved in high amylose biosynthesis by comparing bZIPs expression in the above said genotypes. In comparison, the TabZIP that have FPKM value of zero in other genotype was not taken into consideration as reasons for absence cannot be determined. In this study a relatively stringent criterion of 5-fold differential expression was considered to identify putative candidate TabZIP for high amylose biosynthesis. A total of 147 TabZIPs showed at least 2-fold differential expression among the three pairs, ‘TAC 75’ vs ‘TAC 6’, ‘TAC 75’ vs ‘C 306’, and ‘TAC 6’ vs ‘C 306’ (Fig. 4B).

In the pair, ‘TAC 75’ vs ‘TAC 6’, a total of 89 (39 + 18 + 9 + 23) TabZIPs showed differential expression (at least 2-fold) in the high amylose mutant line (‘TAC 75’) in comparison to the low amylose mutant line (‘TAC 6’) (Fig. 4B, Supplementary Table S5). In this pair, 18 (out of 89) TabZIPs were unique to the pair ‘TAC 75’ vs ‘TAC 6’ (Fig. 4B). Out of 18 unique TabZIPs, two TabZIPs (TabZIP101.1 and TabZIP238.2) showed at least a 5-fold higher and three TabZIPs (TabZIP229.1, TabZIP229.3, and TabZIP238.1) showed at least a 5-fold lower in the high amylose mutant line in comparison to the low amylose mutant line. Out of 18 candidate TabZIPs, five TabZIPs (TabZIP101.1, TabZIP238.2, TabZIP229.1, TabZIP229.3, and TabZIP238.1) may be candidate genes, regulating high amylose biosynthesis.

In the pair, ‘TAC 75’ vs ‘C 306’, a total of 91 (9 + 8 + 39 + 35) TabZIPs showed at least 2-fold differential expression in the high amylose mutant line in comparison to the parent variety. In this pair, 8 TabZIPs were unique to the pair ‘TAC 75’ vs ‘C 306’ (Fig. 4B). Out of 8 unique TabZIP, only one TabZIP i.e. TabZIP145.3 showed at least 5-fold lower expression in the high amylose mutant line in comparison to the parent variety. Therefore, TabZIP145.3 may regulate high amylose biosynthesis.

Between the two pairs involving high amylose mutant line (‘TAC 75’ vs ‘TAC 6’ and ‘TAC 75’ vs ‘C 306’), 9 TabZIPs were common in both the pairs (Fig. 4B). Out of 9, 7 TabZIPs (TabZIP237.1, TabZIP110, TabZIP157.1, TabZIP188.5, TabZIP194.3, TabZIP117.1, and TabZIP137) showed >5-fold lower expression in the high amylose mutant line in both the pairs. These 7 TabZIPs may also regulate high amylose biosynthesis.

Among the three pairs (‘TAC 75’ vs ‘TAC 6’, ‘TAC 75’ vs ‘C 306’, and ‘TAC 6’ vs ‘C 306’), 39 TabZIPs were common (Fig. 4B). Out of 39, three TabZIPs (TabZIP117.2, TabZIP167.2, and TabZIP184.2) showed at least a 5-fold higher expression in the high amylose line than the low amylose mutant line (‘TAC 6’) and the parent variety (‘C 306’) in the two pairs (‘TAC 75’ vs ‘TAC 6’, and ‘TAC 75’ vs ‘C 306’). These three TabZIPs showed at least 5-fold lower expression in the low amylose mutant line (‘TAC 6’) compared to the parent variety ‘C 306’ (Supplementary Table S5). Similarly, three TabZIPs (TabZIP54.1, TabZIP59.2, and TabZIP77.1) showed at least a 5-fold lower expression in the high amylose line than the low amylose mutant line (‘TAC 6’) and the parent variety (‘C 306’) in the two paired genotypes (‘TAC 75’ vs ‘TAC 6’, and ‘TAC 75’ vs ‘C 306’). These three TabZIPs showed at least 5-fold higher expression in the low amylose mutant line (‘TAC 6’) than the parent variety (Supplementary Table S5). Therefore, these six TabZIPs (TabZIP117.2, TabZIP167.2, TabZIP184.2, TabZIP54.1, TabZIP59.2, and TabZIP77.1) are putative candidate genes for high amylose biosynthesis.

After three pair-wise comparisons, a total of 19 (5 + 1 + 7 + 6) candidate TabZIPs (5 positive regulators: TabZIP101.1, TabZIP117.2, TabZIP167.2, TabZIP184.2, and TabZIP238.2 and 14 negative regulators: TabZIP54.1,

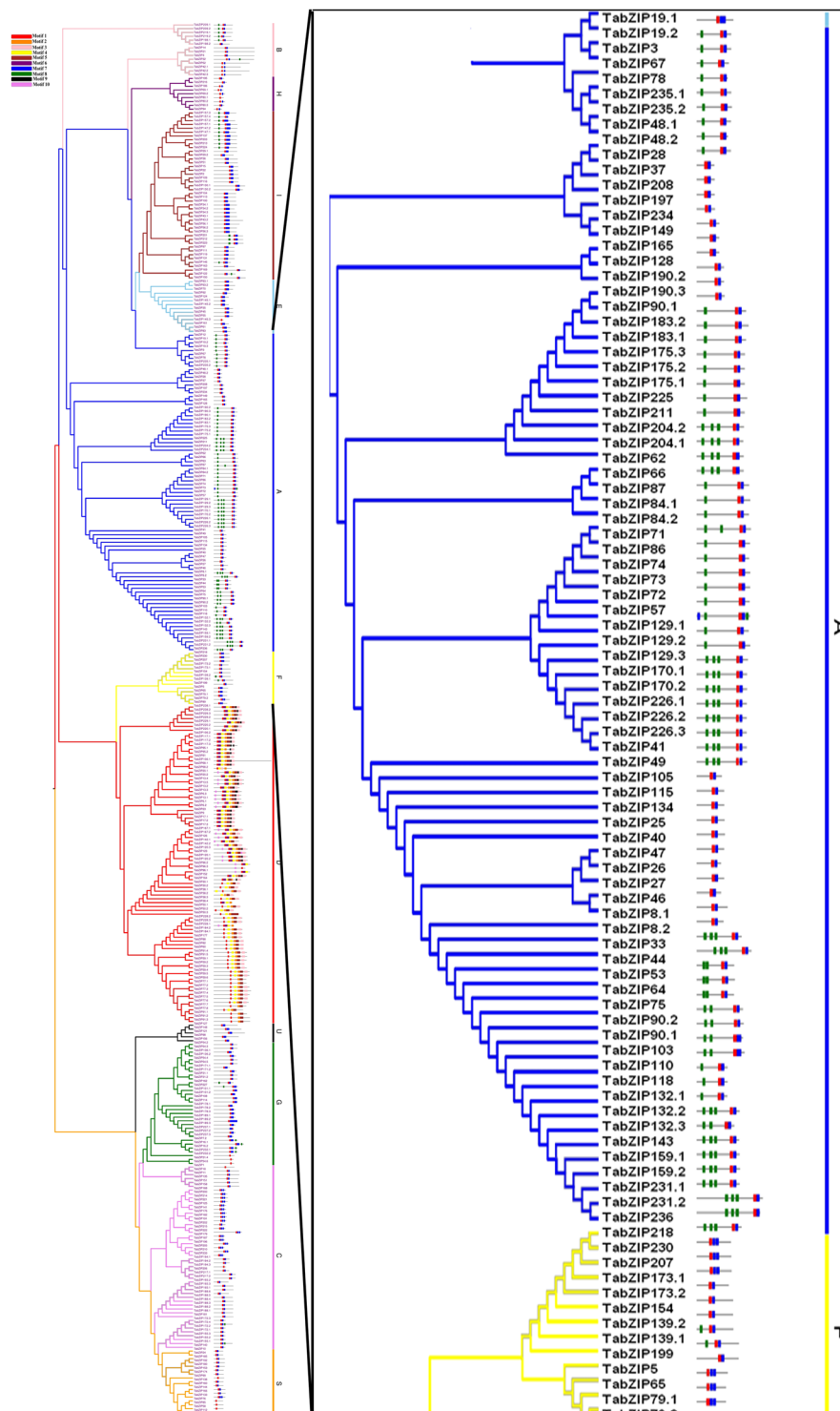


Figure 3. Identification and clustering of the conserved motifs of 370 wheat bZIP (TabZIP) proteins. Motif and phylogenetic analyses of TabZIPs were performed using the MEME database. Based on the phylogenetic relationship, TabZIPs were classified into 11 groups (A–I and S,U). The details of the motifs are given in Supplementary Fig. S2.

TabZIP59.2, TabZIP77.1, TabZIP110, TabZIP117.1, TabZIP137, TabZIP145.3, TabZIP157.1, TabZIP188.5, TabZIP194.3, TabZIP229.1, TabZIP229.3, TabZIP237.1, and TabZIP238.1) were identified for their suggested role in high amylose biosynthesis.

In summary, differential expression (at least 5-fold) analysis of the transcriptome data of the two contrasting mutants and parent varieties identified 19 candidate TabZIPs for high amylose biosynthesis. These TabZIPs may

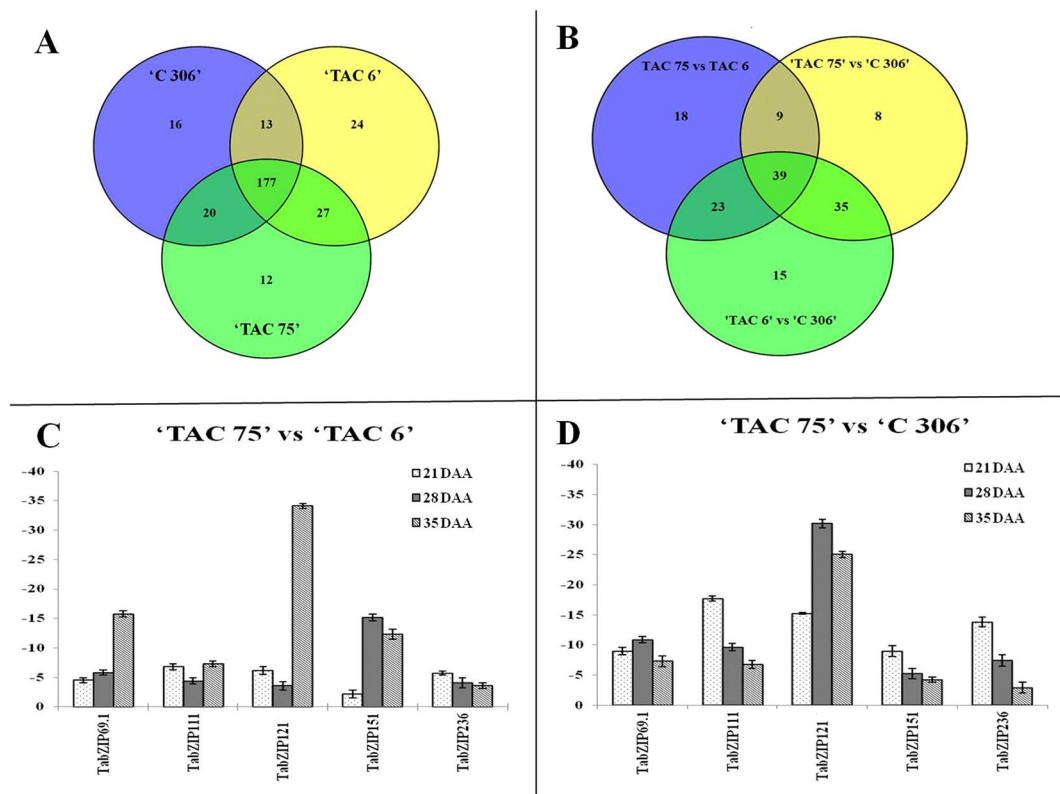


Figure 4. Venn diagrams indicating the number of wheat bZIPs (TabZIPs) showing expression in and among three individual lines (A) and bar charts showing differential expression of candidate TabZIPs (B). (A) Venn diagrams showing the number of bZIPs expressed (FPKM, fragment per kilo per million, ≥ 0.02) in two mutant lines, 'TAC 75' (amylose content = 65%) and 'TAC 6' (amylose content = 7%) and the parent variety, 'C 306' (amylose content = 26%). (B) Comparative analysis of the differentially expressed bZIPs (≥ 2 -fold FPKM) in the three pairs, 'TAC 75' vs 'TAC 6', 'TAC 75' vs 'C 306', and 'TAC 6' vs 'C 306'. The overlapped regions show the common TabZIPs. The FPKM values were determined from the NGS sequencing data (mean of two biological replicates). (C,D) Differential expression level (qRT-PCR) of 5 candidate TabZIPs is represented in 'TAC 75' vs 'TAC 6' (C) and 'TAC 75' vs 'C 306' (D). The expression was analysed at three stages of seed development (21, 28 and 35 DAA). 'TAC 75', 'TAC 6' and 'C 306' represent a high amylose mutant line, a low amylose mutant line, and their parent line, respectively. All the data is represented as mean \pm SD from three technical replicates.

play a pivotal role in the high amylose biosynthesis regulation. The stringent criterion used in this study led to the identification of only few candidate TabZIPs.

qRT-PCR-based candidate gene expression analysis in mutant lines. The detailed information on primers designed for qRT-PCR is provided in Supplementary Table S6. Pairwise differential expression analysis of quantitative expression data of the randomly selected 52 (out of 370) TabZIPs was done at three seed developmental stages (21, 28, and 35 DAA) and is given in Supplementary Table S7. The pairwise differential gene expression analysis was done among three genotypes (two mutant lines, 'TAC 75' & 'TAC 6' and one parent variety, 'C 306'), using log 2-fold of mean expression data. It involved three pairs, 'TAC 75' vs 'TAC 6', 'TAC 75' vs 'C 306', and 'TAC 6' vs 'C 306' (Supplementary Table S7). Out of the three pairs, only two pairs ('TAC 75' vs 'TAC 6' and 'TAC 75' vs 'C 306') are of interest and were analysed further to identify candidate TabZIPs for high amylose biosynthesis by comparing expression of TabZIPs in the high amylose mutant line ('TAC 75') in comparison to the low amylose line ('TAC 6') and the parent variety ('C 306'). Differential expression analysis of the 52 TabZIPs revealed largely negative expression in the high amylose mutant line in comparison to the low amylose mutant line ('TAC 75' vs 'TAC 6') and the parent variety ('TAC 75' vs 'C 306') (Supplementary Table S7). The majority of the TabZIPs showed a similar pattern during three different seed development stages i.e. 21, 28 and 35 DAA. Differential gene expression analysis revealed five bZIPs (TabZIP69.1, TabZIP111, TabZIP121, TabZIP151, and TabZIP236) that showed low expression in the high amylose line in comparison to the low amylose mutant line ('TAC 75' vs 'TAC 6') as well as to the parent variety ('TAC 75' vs 'C 306') (Fig. 4C,D) at the three seed development stages. These five bZIPs may be involved in negative regulation of high amylose biosynthesis. However, two bZIPs, TabZIP50.2 and TabZIP54, showed at least 4-fold positive expression in the high amylose line in comparison to the low amylose mutant line and the parent variety during late stage of seed development i.e. 28 and 35 DAA, the stages at which amylose biosynthesis is very high¹⁸. Further analysis of differential gene expression data between the high amylose ('TAC 75') and low amylose ('TAC 6') mutant line showed development stage specific expression, for example,

at the 21 DAA seed development stage, 46 TabZIPs showed >2-fold negative expression and 2 TabZIPs showed >2-fold positive expression in the high amylose mutant line. At 28 DAA seed development stage, 46 TabZIPs showed >2-fold negative expression and 5 TabZIPs showed >2-fold positive expression in the high amylose line. At 35 DAA seed development stage, 26 TabZIPs showed >2-fold positive expression and 47 TabZIPs showed >2-fold negative expression. Three biological replications, each with three technical replicates, by and large provided similar results. Therefore, the above analysis identified five bZIPs, TabZIP69.1, TabZIP111, TabZIP121, TabZIP151, and TabZIP236 that showed lower expression in the high amylose line in comparison to the low amylose mutant line and the parent variety (Fig. 4C,D) during the three seed development stages and may be the candidate genes for high amylose biosynthesis. These five TabZIPs are distinct from the 19 TabZIPs identified in the transcriptome sequence data. Therefore, a total 24 candidate TabZIPs regulating high amylose biosynthesis are identified in the two transcriptome studies.

Correlation analysis of TabZIPs with key enzymes for amylose biosynthesis. Two enzymes, 'GBSSI' and 'SBEIIb', are key genes mainly responsible for amylose and amylopectin biosynthesis, respectively. The over-expression of GBSSI or down-expression of SBEIIb is functionally validated to high amylose biosynthesis and vice-versa for amylopectin biosynthesis³². Therefore, it is important to analyze statistical correlation of expression data of TabZIPs with that of GBSSI and SBEIIb to identify genes responsible for the regulation of the starch biosynthesis pathway. The pairwise statistical correlation analysis of the normalized expression data of the 52 TabZIPs with that of GBSSI and SBEIIb identified 31 TabZIPs showing positive correlation with GBSSI, while 34 of 52 TabZIPs showed positive correlation with SBEIIb (Fig. 5). Among them, 14 TabZIPs showed positive correlation with both enzymes, and therefore, possibly regulate the expression of both GBSSI and SBEIIb. However, no negative correlation was observed for SBEIIb. The correlation analysis identified three TabZIPs (TabZIP 151, TabZIP121, TabZIP69.1) showing moderate negative to moderate positive correlation with GBSSI and SBEIIb, respectively (Fig. 5). The three are candidate genes for high amylose biosynthesis. These three TabZIPs are also present in the identified 24 candidate TabZIPs. In summary, three approaches (transcriptome studies, qRT-PCR analysis, and correlation data) identified 24 candidate TabZIPs for high amylose biosynthesis.

In silico prediction of functional role of candidate TabZIPs regulating high amylose biosynthesis. The functional role of the 24 candidate TabZIPs were predicted by two methods: 1) engaging phylogenetic group information and 2) protein interacting network analysis using *Arabidopsis* databases.

Functional predictions using phylogenetic group information. The functional role of the 24 candidate TabZIPs was determined using the functional information of *Arabidopsis* bZIP groups³⁰. Their phylogenetic group information (Supplementary Table S1, Supplementary Fig. S2) revealed that the majority of the TabZIPs (10 TabZIPs) belonged to Group D (Table 2). Other TabZIPs belonged to Group C (3 TabZIPs), Group I (3 TabZIP), Groups A (2 TabZIPs), Group G (2 TabZIPs), and one TabZIP each in Groups E, H, and U. However, one TabZIP is not assigned to any group. The 10 TabZIPs of Group D were TabZIP59.2, TabZIP77.1, TabZIP117.1, TabZIP117.2, TabZIP167.2, TabZIP184.2, TabZIP229.1, TabZIP229.3, and TabZIP238.2. The majority of Group D members are involved in plant development and defense³⁰. Group I TabZIPs were TabZIP111, TabZIP137, and TabZIP157.1. Group I members are involved in vascular development^{33,34}. The Group C TabZIPs were TabZIP151, TabZIP188.5, and TabZIP194.3, and this group shows homology to Opaque 2, which regulates starch and carbohydrate biosynthesis in maize²³. Group A TabZIPs were TabZIP110 and TabZIP236 and its member are regulators of ABA-mediated signaling pathways and abiotic stress responsive genes. Group G TabZIPs were TabZIP101.1 and TabZIP237.1. Members of this group are reported to be involved in regulation of light-mediated cell elongation. Member of Group E (TabZIP145.3) is reported to be involved in pollen development³⁵. Member of Group H (TabZIP69.1) is reported to be involved in systemic acquired resistance³⁶. Member of Group U (TabZIP121) are involved in cellular transport and lipid metabolism³⁷.

Putative functional prediction using protein-protein interaction network analysis. Due to availability of limited genomic information and protein databases in wheat, the functional role of possible 24 candidate bZIPs regulating high amylose biosynthesis were predicted *in silico*, using their *Arabidopsis* homologs. The study of gene family interaction (protein-protein interaction) is important for prediction of function in biological processes. In this the study, 13 potential protein interaction networks (N1 to N13) were identified for 24 TabZIPs (Table 2, Fig. 6). The master regulators of the 13 networks were GBF1, AREB3, BZ02H3, BZIP9, PAN, TGA10, BZIP34, BZIP16, TGA6, AT1G06070, VIP1, AT1G19490, and TGA9. The four master regulators, TGA6, TGA9, TGA10, and PAN belonged to TGACG (TGA) motif-binding bZIP sub-family. These master regulators are regulating 9 (out of 24) candidate TabZIPs for high amylose biosynthesis. The previous studies reported their high degree of functional redundancy, mainly in plant disease resistance and stress responses. bZIPs like PAN, TGA9 and TGA10 are also involved in flower development. The master regulator, AREB3 regulates 4 TabZIPs (TabZIP110, TabZIP117.1, TabZIP117.2, and TabZIP236) and belonged to ABA-response elements (ABREs) binding proteins (AREB). AREBs are involved in ABA-responsive abiotic stress, mainly drought and high salinity stresses, and are involved in ABA-dependent signal transduction pathway. The functional roles of other master regulators are given in Table 2.

Discussion

The genome-wide analysis of wheat genome sequence data identified many TabZIPs, which are more than that reported in other major plant species, for example 75 in *Arabidopsis*³⁰, 89 in rice²⁹, 92 in sorghum³⁸, 170 in maize²⁷, 121 in banana³⁹, 77 in cassava⁴⁰, and 96 in *Brachypodium*⁴¹. The previous report identified 182 bZIP proteins in wheat³¹. In this study, out of 370, 184 new TabZIPs were identified (Supplementary Tables S1 and S2) and were

Figure 5

TabZIPs	GBSS I	SBE II
TabZIP50.2	0.84	-0.04
TabZIP216	0.69	-0.10
TabZIP203	0.69	-0.06
TabZIP188.2	0.62	0.36
TabZIP97	0.62	0.36
TabZIP224	0.61	-0.29
TabZIP158	0.60	-0.14
TabZIP219.1	0.59	-0.24
TabZIP198.1	0.57	0.18
TabZIP121	0.57	-0.13
TabZIP209.1	0.56	0.07
TabZIP193.1	0.55	-0.25
TabZIP208	0.54	0.16
TabZIP120.3	0.52	-0.23
TabZIP157.1	0.52	-0.21
TabZIP159.1	0.52	-0.16
TabZIP156.2	0.50	-0.11
TabZIP160	0.47	-0.13
TabZIP51	0.40	0.58
TabZIP212	0.40	0.00
TabZIP219	0.37	-0.13
TabZIP54	0.34	0.08
TabZIP228.3	0.29	0.43
TabZIP56.3	0.27	0.13
TabZIP8.2	0.26	-0.21
TabZIP123.1	0.22	0.00
TabZIP175.1	0.13	0.48
TabZIP137	0.11	0.55
TabZIP52	0.06	0.53
TabZIP120.2	0.04	0.39
TabZIP167.2	0.04	0.06
TabZIP113	0.00	0.03
TabZIP38.2	-0.02	-0.18
TabZIP148	-0.04	0.59
TabZIP38.1	-0.07	0.55
TabZIP143	-0.08	0.56
TabZIP36	-0.10	0.57
TabZIP13.3	-0.13	0.57
TabZIP29.1	-0.15	0.66
TabZIP59.4	-0.18	0.17
TabZIP152	-0.25	0.51
TabZIP54.1	-0.28	0.32
TabZIP167.1	-0.28	0.71
TabZIP147.1	-0.33	0.72
TabZIP136.1	-0.34	0.43
TabZIP236	-0.34	0.23
TabZIP91.2	-0.35	0.52
TabZIP98	-0.37	0.56
TabZIP151	-0.48	0.43
TabZIP121.1	-0.48	0.44
TabZIP69.1	-0.51	0.49
TabZIP111	-0.52	0.11

Figure 5. Heatmap of Pearson's correlation coefficient (r) between the expression data of 52 TabZIPs and two starch biosynthesis genes, granule bound starch synthase I (GBSSI), and starch branching enzyme II (SBEIIb). The expression data is the normalized delta C_T (ΔC_T) values of the genes measured on a quantitative real-time PCR (qRT-PCR). Positive and negative correlations are highlighted in green and red colour, respectively.

confirmed by motif and domain analysis. The presence of a large number of TabZIPs in wheat is expected due to its large genome size. bZIP numbers may further increase when the complete reference genome will be available.

Multiple sequence alignment and phylogenetic analysis clustered the wheat bZIP into 11 groups (Figs 1–3 and Supplementary Fig. S1), which is largely in agreement with the previous phylogenetic classification of plant bZIPs. Earlier, major plant bZIPs were grouped in 10–11 groups, for example, 10 groups are reported in *Arabidopsis*³⁰ and cassava⁴⁰, and 11 groups in maize²⁷, rice²⁹, and banana³⁹. Earlier bZIPs of wheat and its two wild relatives and

Wheat bZIPs (TabZIPs)	TabZIP group	Arabidopsis bZIPs (TabZIP homologs)	Protein-protein interaction (PPI)		Predicted function/role using <i>in silico</i> database	PPI network (Refer to Fig. 6)	References of predicted Functional role
			Master regulator	Interacting partners			
TabZIP110	A	bZIP12	AREB3	SNRK2.1, SNRK2.10, DPBF2, NRK2.3	Regulates the expression of stress (abscisic acid) related genes.	N2	Garcia <i>et al.</i> 2008
TabZIP236		BZIP66					
TabZIP151	C	bZIP63	BZ02H3	BZIP1, BZIP53, RR18	Regulates seed storage protein expression	N6	Alonso <i>et al.</i> 2009
TabZIP188.5		bZIP9	BZIP9	BZIP53, GBF6, BZIP25, BZIP44, ATB2	Regulation of sugar responsive genes	N5	Kang <i>et al.</i> 2010
TabZIP194.3							
TabZIP117.2	D	BZIP12	AREB3	SNRK2.1, SNRK2.10, DPBF2, NRK2.3	Regulates the expression of stress (abscisic acid) related genes.	N2	Garcia <i>et al.</i> 2008
TabZIP117.1		bZIP29					
TabZIP229.1		BZIP20	PAN	ROXY1, BOP2, NPR1, NPR3, NPR4, LFY	Regulates petal development	N4	Maier <i>et al.</i> 2009
TabZIP229.3							
TabZIP238.1							
TabZIP238.2							
TabZIP167.2		bZIP65	TGA10	NPR1, NPR4, NPR3, ROXY1, ROXY2, BOP2	Promotes anther development	N3	Murmu <i>et al.</i> 2010
TabZIP184.2							
TabZIP59.2		bZIP21	TGA9	NPR3, NPR1, ROXY1, ROXY2	Promotes anther development	N1	Murmu <i>et al.</i> 2010
TabZIP77.1							
TabZIP145.3	E	BZIP61	bZIP34	BZIP43, VIP1, GBF4, ROXY2		N11	
TabZIP101.1	G	bZIP16	BZIP16	GBF1, BZIP68, GBF2, GBF3	Represses the hormone responsive expression of genes in seed development	N13	Schindler <i>et al.</i> 1992 and Shen <i>et al.</i> 2008
TabZIP237.1							
TabZIP69.1	H	BZIP45	TGA6	NPR1, NPR4, NPR3, ROXY1	Regulates gene expression of mature fruit abscission	N9	Liu <i>et al.</i> 2005
TabZIP137	I	bZIP69	AT1G06070	BZIP75, GBF4, BZIP4, FD, BZIP27	Floral pathway development	N7	Abe, Mitsutomo, <i>et al.</i> 2005
TabZIP157.1							
TabZIP111	I	BZIP51	VIP1	MYBR1, AGB, BZIP52	Regulates the osmosensory signals of stress (abscisic acid) related genes	N8	Tsugama <i>et al.</i> 2012
TabZIP121	U	ND	AT1G19490	BZIP61, BZIP34, BZIP75, BZIP4, BZIP27	Induction of mesoderm and endoderm at earlier embryogenesis	N10	Shen <i>et al.</i> 2007
TabZIP54.1	ND	bZIP16	GBF1	HYH, HY5, MYC2, CKA2	Represses the hormone responsive expression genes in seed development	N12	Schindler <i>et al.</i> 1992

Table 2. Detail of 24 candidate TabZIPs including their protein-protein interaction networks and putative predicted functions using *Arabidopsis* databases. ND = not determined.

other plant species were grouped into 14 groups³¹. In this study, among the eleven phylogenetic groups identified, Group A and D are large cluster groups, each containing 83 TabZIPs. It is reported that bZIPs in Group A play an important role in abscisic acid (ABA) signaling and abiotic stresses. The abiotic stresses and ABA help to induce transcriptional and post-translational regulation^{42,43}. Several bZIPs were identified and shown to be ABA-responsive and improve multiple abiotic stress tolerance^{42–44}. For example, OsbZIP23, a bZIP from rice, plays a major role in ABA dependent drought and salinity tolerance⁴³. Group D members are involved in plant development and defence³⁰. For example, TGA (TGACG sequence binding protein) family in *Arabidopsis* interacts with the non-expressor of PR1 (NPR1), which is a key component in the salicylic acid defense signaling pathway^{45–49}. The tobacco TGA1 and *Arabidopsis* TGA2 proteins are responsive to salicylic acid and bind to xenobiotic responsive promoters^{47–49}.

The functional information on Group B 14 TabZIPs is very limited. The members of this group have trans-membrane domain and specific domain at C-terminus which are important for ER stress response³⁹. Group C contains 48 TabZIPs and the members of this group include *Opaque2*, which plays an important role in modulating seed-specific gene expression³⁰. *Opaque2* regulates seed storage protein expression by interacting with the PBF protein. Groups E and F each contain 14 TabZIPs, but their functional information is not available. Group G comprises 32 TabZIPs and it is named so due to its members being G-box binding factors³⁰. The G-box binding factor genes from *Arabidopsis* are connected to the regulation of light-responsive promoters, and are involved in biotic and abiotic stresses^{39,50}. Group H contains nine TabZIPs. The name of this group refers to the HY5 gene which is a bZIP transcription factor in *Arabidopsis* that binds to a G-box, and regulates the stimulus-induced development³⁰. The control of HY5 activity by light is also well documented in dark-grown *Arabidopsis*^{30,40}.

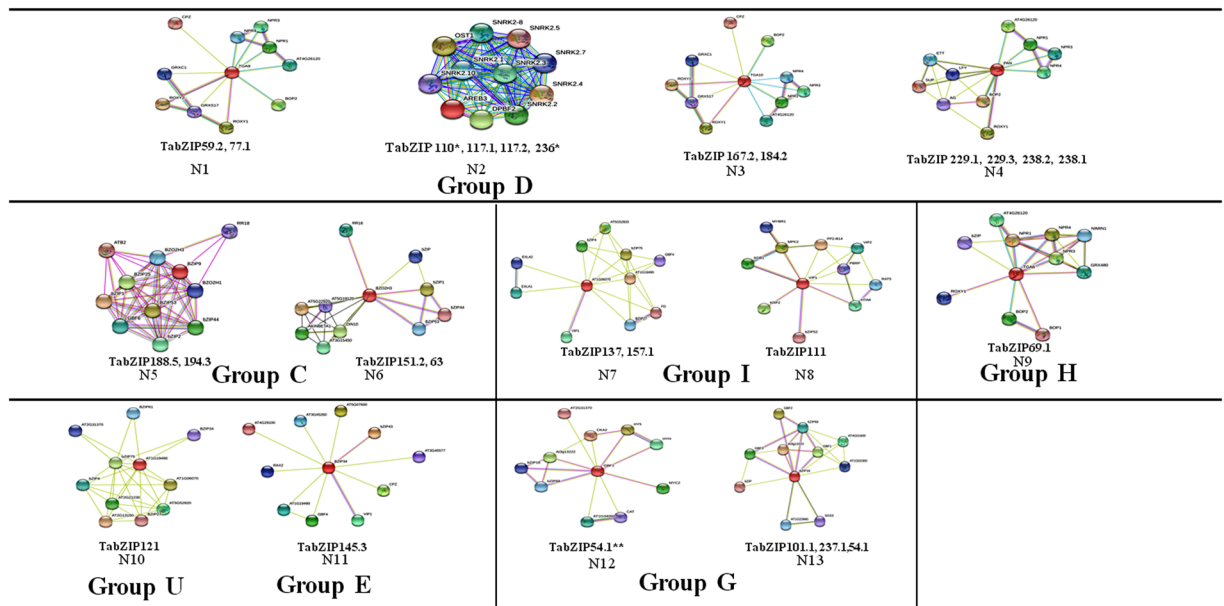


Figure 6. Protein-protein interaction networks (N1 to N13) identified for 24 candidate wheat bZIPs (TabZIPs) controlling high amylose biosynthesis in wheat using *Arabidopsis* databases. The 24 candidate TabZIPs were categorised into 8 wheat bZIP groups (A, C, D, H, I, U, E, G) and their identified 13 protein-protein interaction (PPI) networks (numbered as N1 to N13). Their interactions were analysed online using STRING database (<https://string-db.org/>). ‘*’ indicates two TabZIPs (TabZIP110 and TabZIP236) belonging to Group A identified PPI network N2 that is also identified by Group D members. ‘**’ indicates TabZIP54.1 whose wheat bZIP group was not determined.

Group I contains 44 TabZIPs. Various studies on Group I genes have been reported from different plant species (Rice RF2a and tomato VSF-1), which indicate their role in vascular development^{51,52}.

Group S in wheat is the third smallest group in wheat with 19 TabZIPs, whereas in *Arabidopsis* it is the largest³⁰. In *Arabidopsis* the *ATBZIP11/ATB2* gene is regulated by light, and play role in carbohydrate-consuming⁵³ (i.e. sink) and in the vascular system³⁰. The *ATB2* is involved in post-transcriptional repression by sucrose and carbohydrate balancing⁵³. While S Group, in monocot and dicot species get activated during stress⁵⁴. Group U in wheat is the smallest with five TabZIPs, which are very similar to the members of other plant species, such as maize and rice³⁰. Members of this group have hydrophobic isoleucine instead of conserved arginine 10 (Supplementary Fig. S1). This substitution in amino acid affects the DNA binding specificity of bZIP, which is documented in earlier studies⁵⁵. This classification was further supported by conserved motif analysis. Conserved motif analysis indicated that almost all the TabZIPs in wheat contained typical bZIP domains. Additionally, each subfamily had some common motifs and some subfamilies also contained the special motifs. These features in conserved bZIP motifs were also observed in maize, rice and *Arabidopsis*. Generally, most TabZIP genes in the same subfamilies showed similar gene structure and conserved motifs, which support their close evolutionary relationship and the classification of subfamilies (Table 1, Fig. 3, and Supplementary Table S3).

As wheat is among the most significant crops, the biological processes like seed development and maturation are vital for starch quality. The bZIP family has been reported to have role in the seed development processes of many plant species, however, the role of bZIPs in amylose or amylopectin biosynthesis regulation is undetermined in wheat. TabZIPs act as positive as well as negative regulators of genes, which actively take part in physiological processes mainly via DNA binding. In the current study, cis-regulatory sequences analysis of the starch biosynthesis pathway indicate the presence of high density of G and A boxes in GBSS I, and SBE II genes (Supplementary Fig. S3), indicating a high probability of bZIPs binding at their promoter regions. Therefore, some may have active role in regulation of amylose and amylopectin biosynthesis.

Before comparative gene expression analysis and next generation sequencing (NGS), the mutation lines were characterized by sequencing two key genes (GBSSI and SBEII) responsible for amylose and amylopectin biosynthesis, both synonymous and non-synonymous mutations were detected in their isoforms. The frequency of mutations in SBEIIa and SBE IIb were relatively higher than GBSSI (Supplementary Fig. S4). The comparative gene expression analysis of GBSSI, SBEIIa, and SBEIIb starch metabolic genes at three development stages (21, 28 and 35 DAA) in mutant lines validated high and low amylose mutants as reported earlier in wheat and other crops^{10,12,17,18,56}.

Differential gene expression analysis using whole transcriptome NGS data (two biological replicates) and qRT-PCR data using randomly selected 52 TabZIPs of the two contrasting mutants and parent varieties identified a total of 24 TabZIPs, which are potential candidate genes for high amylose biosynthesis. These TabZIPs may play role in high amylose biosynthesis by either positive or negative regulation. bZIP transcription factors are key regulators of starch biosynthesis genes in rice and maize, but their role in regulating these genes in wheat is poorly

understood. It is reported earlier in rice where bZIP transcription factor OsbZIP58²³ was regulator of six starch metabolic pathway genes including GBSSI and SBEII in starch biosynthesis²³. The identified candidate TabZIPs were categorized into 8 bZIP groups, namely A, C, D, E, G, H, I and U and one TabZIP (TabZIP54.1) was not assigned to any groups (Table 2). The putative functional role of the 24 candidate TabZIPs were predicted by using phylogenetic group information as well as protein interacting network analysis using *Arabidopsis* databases. Their putative functional role is described in the Results section. The putative functional role can be grouped into four categories- stress (abiotic and biotic), flower and seed development, starch biosynthesis, and seed storage protein regulations.

Interactome analysis identified 13 protein interaction networks (N1 to N13, Fig. 6, Table 2). The master regulators and partners of networks showed that many TabZIPs play important role in biological processes. The four master regulators, TGA6, TGA9, TGA10, and PAN identified in this study belonged to TGACG (TGA) motif-binding bZIP sub-family. These master regulators are represented by 9 (out of 24) candidate TabZIPs for high amylose biosynthesis. Group D TabZIPs (TabZIP59.2, 77.1, 167.2, 184.2) show homology with TGA9 and TGA10 (networks N1 and N4). TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are required for anther development⁵⁷. They also interact with other TGAs (TGA4) and PAN⁵⁸. The master regulator of other Group D members (TabZIP 229.1, TabZIP229.3, TabZIP238.1 and TabZIP238.2) in network N4 is PAN (PERIANTHIA). PAN is involved in the determination of floral organ number and also in a post-translational modification by GRXC7/ROXY1⁵⁹. It also binds with BOP1 and BOP2 which are involved in growth asymmetry, an important aspect of patterning in leaves and flowers⁶⁰.

Group I members are involved in vascular development^{30,51,52}. TabZIP111 shows homology with VIP1 in network N8 and TabZIP137 and TabZIP157.1 to *Arabidopsis* bZIP (AT1G06070) in network N7 in *Arabidopsis*. They play role in vascular development in tobacco and when over-expressed in *Arabidopsis*, caused growth retardation under a mannitol-stressed condition⁶¹. Group G members (TabZIP101.1 and 237.1) show homology with GBF and bZIP16 in network N13 (Table 2). GBF2 and AtbZIP16, G-box binding proteins, are involved in the regulation of light or hormone induced stresses^{62,63}. Group C members (TabZIP151, 188.5 and 194.3) show homology with BZ02H3 in network N6 and AtbZIP9 in network N5. BZ02H3 regulates seed storage protein expression²⁴. Group C TabZIPs also shows homology with *Opaque2*, which is more closely related to monocot species and regulate seed storage protein production by interacting with the PBF protein in *Arabidopsis*'s embryo. Rice OsbZIP58 and OsbZIP20 regulate starch and carbohydrate biosynthesis²³. The master regulator of Group A members (TabZIP110 and TabZIP236 in network N2) is AREB (ABA-response elements binding proteins) which is induced by ABA and osmotic stresses⁶⁴. AREB is activated by SnRKs and is involved in the PP2C-SnRK-AREB pathway and is an important component of the ABA signalling pathway. The Interactome analysis provides broad sight to understand the regulation and interacting partners' of these candidate TabZIPs and suggest their putative functions. These wheat bZIPs will be used for validation in molecular breeding programme on large germplasm set through e-QTL analysis and using functional genomics tools.

Conclusion

High amylose starch is considered to be a good dietary fibre, rich healthy starch, as it is not easily or slowly digested in our guts and is finally transformed into small chain fatty acids (SCFAs) (prebiotics) by bacteria in the large intestine. There is a global demand to develop cereal crops with high levels of resistant starch or dietary fibre rich food grains. Starch is composed of two fractions, amylose and amylopectin, which are synthesised by starch metabolic pathway genes. We identified 370 TabZIP genes from wheat and unravelled their basic classification and evolutionary relationships using evolutionary and conserved protein motif analyses. This will provide ample knowledge for functional characterization of bZIP genes. 24 candidate TabZIPs were identified for high amylose biosynthesis using whole transcriptome data of wheat contrasting mutants for amylose content. Their putative functional roles were determined using protein-protein network analysis. The bZIPs are being used in our lab in molecular breeding for the improvement of amylose content in wheat. All this information will lay a platform for future research on the functional characterization of potential TabZIPs and regulatory mechanism of high amylose biosynthesis in cereal crops. This study therefore is advancing our understanding of the molecular basis of genetic enhancements of amylose content in wheat.

Materials and Methods

Plant materials and transcriptome sequence data. Two contrasting mutant lines, 'TAC 75' (amylose content ~65%) and 'TAC 6' (amylose content ~7%) in M6 generation were used for the identification of putative bZIPs for high amylose biosynthesis. The lines were developed after the EMS treatment of the parent bread wheat (*Triticum aestivum* L.) variety 'C 306' (amylose content ~26%)⁵⁶. The 284 Gb transcriptome sequence data was generated from the two biological replicates of two contrasting mutant lines, 'TAC 75' and 'TAC 6' and their parent variety, 'C 306' (unpublished). The whole transcriptome data will be made available by requesting the corresponding author and it is available at NABI's intranet. For transcriptome sequencing, RNAs were extracted from developing seeds at 28 days after anthesis.

In silico identification, phylogenetic analysis, and physiochemical properties of bZIPs in bread wheat. The identification of the genome wide distribution of bZIPs was performed in two steps. In the first step, the hidden Markov model profiles⁶⁵ of the bZIP domain, viz. PF00170⁶⁶, were used as queries to search the bZIP proteins in the wheat proteome Ensembl database (<http://plants.ensembl.org/index.html>) using HMMER3.0⁶⁵. In the second step, a local BLASTp search was performed to identify the predicted wheat bZIPs by HMMER3.0 with already known bZIPs from *Arabidopsis*³⁰, maize²⁷, rice²⁹ and barley⁶⁷. These potential wheat bZIPs were further examined for the existence and integrity of the bZIP domain by using NCBI-CDD⁶⁸ and InterProScan⁶⁹. The bZIP protein sequences of wheat, rice, maize, barley, and *Arabidopsis* were aligned by using

ClustalW⁷⁰ with gap opening and gap extension penalties of 10 and 0.1, respectively. The neighbour-joining (NJ) method was used to infer the evolutionary history of all bZIP protein sequences. The associated taxa clustered together in the bootstrap test of 1000 replicas. The phylogenetic tree was constructed using MEGA software, version 6⁷¹. The visualization and annotation of the constructed phylogenetic tree was carried out by using EvolView⁷². The physicochemical properties of protein sequences such as molecular weight, isoelectric point, theoretical pI and GRAVY (Grand average of Hydropathicity) values of TabZIPs were calculated using ExPASy ProtParam⁷³. MEME (version 4.11.2) was used for the prediction of conserved motifs. The limits specified for minimum width, maximum width, and maximum numbers of motifs were 8, 50 and 10, respectively. The motifs were numbered according to their order displayed by MEME Suite.

In Silico cis-regulating elements map analysis of starch biosynthesis pathway genes. Cis-regulating elements of starch metabolic pathway genes were analysed to explore the DNA binding domains of bZIPs. The genomic sequences of starch metabolic pathway genes (GBSSI, GBSSII, SSI, SSII, SSIII, SSIV, SBEI, SBEIIa and SBEIIb) were retrieved from the International Genome Sequencing Consortium database (<https://www.wheatgenome.org/>) (IWGSC). They were processed through Regulatory Sequence Analysis Tools (RSAT: <http://rsat.ulb.ac.be/rsat/>)⁷⁴ to determine binding sites using their up to 1000 bp upstream sequences.

In Silico differential gene expression analysis. The transcriptome sequence data of the mutant lines and parent (unpublished) was used for the identification of wheat bZIPs (TabZIPs). The identified bZIPs were further confirmed based on gene ontology and Pfam domain analysis. FPKMs (Fragments per kilobase of transcript per million mapped reads) values of TabZIPs were used for gene expression study. FPKMs of the 370 TabZIP genes were retrieved from the transcriptomic sequence data on two biological replicates of the developing seeds (28 days after anthesis) belonging to the two mutant lines and the parent wheat variety. In this study the TabZIP genes having FPKM values of at least 0.02 were considered to be expressed and used for differential gene expression analysis.

Gene expression analysis by qRT-PCR during seed development. The primers of 52 TabZIPs along with two key genes (GBSSI and SBEII) of starch metabolic pathway that are largely responsible for amylose and amylopectin biosynthesis were designed using Primer Express Software Tool version 3.0 (Thermo Fisher Scientific, USA). The tagged spikes were harvested at three seed developmental stages i.e. 21, 28, and 35 DAA for RNA extraction by Trizol method and cDNA preparation. The relative expression levels of the target bZIPs were calculated by $\Delta\Delta C_t$ method (Schmittgen and Livak *et al.*, 2001). Wheat ADP-Ribosylation Factor, ARF (AB050957.1) was used as an internal control gene for normalization of gene expression data.

Statistical correlation analysis. Pearson's correlation analysis was performed between the normalized expression data (qRT-PCR) of 52 TabZIPs with that of SBEIIb and GBSSI (all values taken in this study are normalized Ct values with housekeeping ARF gene).

Interactome analysis. The putative function was determined for the candidate TabZIPs identified for high amylose biosynthesis through interactome analysis i.e. protein interacting network analysis. Wheat bZIPs homologs were determined in *Arabidopsis* and then their protein-protein interaction (PPI) networks were identified in *Arabidopsis thaliana* databases (<https://string-db.org/>) using default parameters.

References

- Pfister, B. & Samuel, C. Z. Formation of starch in plant cells. *Cellular and Molecular Life Sciences*. **73**(14), 2781–2807 (2016).
- Zeeman, S. C., Jens, K. & Smith, A. M. Starch: its metabolism, evolution, and biotechnological modification in plants. *Annual review of plant biology*. **61**, 209–234 (2010).
- Birt, D. F. *et al.* Resistant starch: promise for improving human health. *Advances in Nutrition: An International Review Journal*. **4**(6), 587–601 (2013).
- Topping, D. L. & Peter, M. Clifton. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiological reviews*. **81**(3), 1031–1064 (2001).
- Jenkins, D. J. A. *et al.* Glycemic index: overview of implications in health and disease. *The American journal of clinical nutrition*. **76**(1), 266S–273S (2002).
- Eskin, N. A. Michael, and Fereidoon Shahidi. *Biochemistry of foods*. Academic Press, (2012).
- Messer, E. Food Systems and Dietary Perspective: Are Genetically Modified Organisms the Best Way to Ensure Nutritionally Adequate Food? *Indiana Journal of Global Legal Studies*. 65–90 (2001).
- Fitzgerald, M. A., Susan, R. M. C. & Robert, D. Hall. Not just a grain of rice: the quest for quality. *Trends in plant science*. **14**(3), 133–139 (2009).
- Tang, G., Galili, G. & Zhuang, X. RNAi and microRNA: breakthrough technologies for the improvement of plant nutritional value and metabolic engineering. *Metabolomics*. **3**(3), 357–369 (2007).
- Tetlow, I. J. Understanding storage starch biosynthesis in plants: a means to quality improvement. *Botany*. **84**(8), 1167–1185 (2006).
- Tester, R. F., Karkalas, J. & Qi, X. Starch composition, fine structure and architecture. *Journal of Cereal Science*. **39**(2), 151–165 (2004).
- Jenkins, P. J. & Donald, A. M. The influence of amylose on starch granule structure. *International Journal of Biological Macromolecules*. **17**(6), 315–321 (1995).
- Jobling, S. Improving starch for food and industrial applications. *Current opinion in plant biology*. **7**(2), 210–218 (2004).
- Tatsuro, H. & Terao, T. A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa* L.). *Planta*. **220**(1), 9–16 (2004).
- Pandey, M. K. *et al.* Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (*Oryza sativa* L.). *Biotechnology advances*. **30**(6), 1697–1706 (2012).
- Smith, A. M. The biosynthesis of starch granules. *Biomacromolecules*. **2**(2), 335–341 (2001).
- Carciofi, M. *et al.* Concerted suppression of all starch branching enzyme genes in barley produces amylose-only starch granules. *BMC plant biology*. **12**(1), 223 (2012).

18. Singh, A. *et al.* Expression patterns of genes involved in starch biosynthesis during seed development in bread wheat (*Triticum aestivum*). *Molecular breeding*. **35**(9), 184 (2015).
19. Yamakawa, H. *et al.* Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. *Plant physiology*. **144**, 258–277 (2007).
20. Landschulz, W. H., Peter, F. J. & Steven, L. M. K. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*. **240**(4860), 1759–1765 (1988).
21. Ellenberger, T. E. *et al.* The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein–DNA complex. *Cell* **71**, 1223–1237 (1992).
22. She, K. –C. *et al.* A novel factor FLOURY ENDOSPERM2 is involved in regulation of rice grain size and starch quality. *The Plant Cell*. **22**(10), 3280–3294 (2010).
23. Wang, J. –C. *et al.* OsbZIP58, a basic leucine zipper transcription factor, regulates starch biosynthesis in rice endosperm. *Journal of experimental botany*. **64**(11), 3453–3466 (2013).
24. Maddaloni, M. *et al.* The transcriptional activator Opaque-2 controls the expression of a cytosolic form of pyruvate orthophosphate dikinase-1 in maize endosperms. *Molecular and General Genetics*. **250**(5), 647–654 (1996).
25. Fu, F.-F. & Xue, H.-W. Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. *Plant Physiology*. **154**(2), 927–938 (2010).
26. Chen, J. *et al.* ZmbZIP91 regulates expression of starch synthesis-related genes by binding to ACTCAT elements in their promoters. *Journal of Experimental Botany* **67**(5), 1327–1338 (2016).
27. Wei, K. *et al.* Genome-wide analysis of bZIP-encoding genes in maize. *DNA research*. **19**(6), 463–476 (2012).
28. Baloglu, M. C. *et al.* Genome-wide analysis of the bZIP transcription factors in cucumber. *PLoS One* **9**(4), e96014 (2014).
29. Nijhawan, A. *et al.* Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant physiology* **146**(2), 333–350 (2008).
30. Jakoby, M. *et al.* bZIP transcription factors in Arabidopsis. *Trends in plant science* **7**(3), 106–111 (2002).
31. Li, X. *et al.* 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** (2015).
32. Szklarczyk, D. *et al.* STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**(D1), D447–52 (2015).
33. Liu, C., Wu, Y. & Wang, X. bZIP transcription factor OsbZIP52/RISBZ5: a potential negative regulator of cold and drought stress response in rice. *Planta*. **235**(6), 1157–1169 (2012).
34. Nakase, M., Aoki, N., Matsuda, T. & Adachi, T. Characterization of a novel rice bZIP protein which binds to the α -globulin promoter. *Plant molecular biology*. **33**(3), 513–522 (1997).
35. Gibalova, A. Rena kD, Matczuk K, Duplakova N, Chab D, Twell D, Honys D. AtbZIP34 is required for Arabidopsis pollen wall patterning and the control of several metabolic pathways in developing pollen. *Plant Mol Biol* **70**, 581–601 (2009).
36. Zhang, Y. *et al.* Knockout analysis of Arabidopsis transcription factors TGA2, TGA5, and TGA6 reveals their redundant and essential roles in systemic acquired resistance. *The Plant Cell*. **15**(11), 2647–2653 (2003).
37. Gibalová, A. *et al.* Characterization of pollen-expressed bZIP protein interactions and the role of ATbZIP18 in the male gametophyte. *Plant reproduction*. **30**(1), 1–17 (2017).
38. Wang, J. *et al.* Genome wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on Sorghum. *Journal of integrative plant biology*. **53**(3), 212–231 (2011).
39. Hu, W. *et al.* Genome-wide analyses of the bZIP family reveal their involvement in the development, ripening and abiotic stress response in banana. *Scientific reports*. **6** (2016).
40. Hu, W. *et al.* Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Scientific reports*. **6**, 22783 (2016).
41. Xiang, L. & Chu, Z. 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**(1), 227 (2015).
42. Johnson, R. R., Wagner, R. L., Verhey, S. D. & Walker-Simmons, M. K. The abscisic acid-responsive kinase PKABA1 interacts with a seed-specific abscisic acid response element-binding factor, TaABF, and phosphorylates TaABF peptide sequences. *Plant Physiology*. **130**(2), 837–846 (2002).
43. Xu, D. B. *et al.* ABI-like transcription factor gene TaABL1 from wheat improves multiple abiotic stress tolerances in transgenic plants. *Functional & integrative genomics*. **14**(4), 717–730 (2014).
44. Nakamura, S., Komatsuda, T. & Miura, H. Mapping diploid wheat homologues of *Arabidopsis* seed ABA signaling genes and QTLs for seed dormancy. *Theoretical and Applied Genetics*. **114**(7), 1129–1139 (2007).
45. Xiang, Y., Tang, N., Du, H., Ye, H. & Xiong, L. Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. *Plant physiology*. **148**(4), 1938–1952 (2008).
46. Alves, M. S. *et al.* Plant bZIP transcription factors responsive to pathogens: a review. *International journal of molecular sciences*. **14**(4), 7815–7828 (2013).
47. Singh, K. B., Foley, R. C. & Oñate-Sánchez, L. Transcription factors in plant defense and stress responses. *Current opinion in plant biology*. **5**(5), 430–436 (2002).
48. Després, C., DeLong, C., Glaze, S., Liu, E. & Fobert, P. R. The Arabidopsis NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. *The Plant Cell*. **12**(2), 279–290 (2000).
49. Zhou, J. M. *et al.* NPR1 differentially interacts with members of the TGA/OBF family of transcription factors that bind an element of the PR-1 gene required for induction by salicylic acid. *Molecular Plant-Microbe Interactions*. **13**(2), 191–202 (2000).
50. Zhang, L. *et al.* A novel wheat bZIP transcription factor, TabZIP60, confers multiple abiotic stress tolerances in transgenic Arabidopsis. *Physiologia plantarum*. **153**(4), 538–554 (2015).
51. Ringli, C. & Keller, B. Specific interaction of the tomato bZIP transcription factor VSF-1 with a non-palindromic DNA sequence that controls vascular gene expression. *Plant molecular biology*. **37**(6), 977–988 (1998).
52. Yin, Y., Zhu, Q., Dai, S., Lamb, C. & Beachy, R. N. RF2a, a bZIP transcriptional activator of the phloem specific rice tungro bacilliform virus promoter, functions in vascular development. *The EMBO Journal*. **16**(17), 5247–5259 (1997).
53. Rook, F. *et al.* Sucrose-specific signalling represses translation of the Arabidopsis ATB2 bZIP transcription factor gene. *The Plant Journal* **15**(2), 253–263 (1998).
54. Chen, H. *et al.* Basic leucine zipper transcription factor OsbZIP16 positively regulates drought resistance in rice. *Plant science*. **193**, 8–17 (2012).
55. Wei, K. A. I. F. A. *et al.* Genome-wide analysis of bZIP-encoding genes in maize. *DNA research* **19**(6), 463–476 (2012).
56. Mishra, A. *et al.* Development of EMS-induced mutation population for amylose and resistant starch variation in bread wheat (*Triticum aestivum*) and identification of candidate genes responsible for amylose variation. *BMC plant biology* **16**(1), 217 (2016).
57. Murmu, J. *et al.* Arabidopsis bZIP transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. *Plant Physiology*. pp-110 (2010).
58. Deppmann, C. D., Rebecca, S. & Elizabeth, J. T. Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. *Molecular biology and evolution*. **23**(8), 1480–1492 (2006).

59. Li, S. *et al.* Nuclear activity of ROXY1, a glutaredoxin interacting with TGA factors, is required for petal development in *Arabidopsis thaliana*. *The Plant Cell* **21**(2), 429–441 (2009).
60. Hepworth, S. R. *et al.* BLADE-ON-PETIOLE-dependent signaling controls leaf and floral patterning in *Arabidopsis*. *The Plant Cell* **17**(5), 1434–1448 (2005).
61. Tsugama, D., Shenkui, L. & Tetsuo, T. Analysis of functions of VIP1 and its close homologs in osmosensory responses of *Arabidopsis thaliana*. *PLoS One* **9**(8) (2014).
62. Menkens, A. E. & Anthony, R. C. Isolation and characterization of a fourth *Arabidopsis thaliana* G-box-binding factor, which has similarities to Fos oncoprotein. *Proceedings of the National Academy of Sciences*. **91**(7), 2522–2526 (1994).
63. Shaikhali, J. GIP1 protein is a novel cofactor that regulates DNA-binding affinity of redox-regulated members of bZIP transcription factors involved in the early stages of *Arabidopsis* development. *Protoplasma*. **252**(3), 867–883 (2015).
64. Kim, S. *et al.* ARIA, an *Arabidopsis* arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signaling component. *Plant physiology* **136**(3), 3639–3648 (2004).
65. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **367** (2011).
66. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*. **44**(D1), D279–D285 (2016).
67. Pourabed, E., Golmohamadi, F. G., Monfared, P. S., Razavi, S. M. & Shobbar, Z. S. Basic leucine zipper family in barley: genome-wide characterization of members and expression analysis. *Molecular biotechnology*. **57**, 12–26 (2015).
68. Marchler-Bauer, A. *et al.* NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–226 (2015).
69. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics*. **30**, 1236–1240 (2014).
70. Thompson J. D., Gibson T. & Higgins D. G. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, 2–3 (2002).
71. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*. **30**(12), 2725–2729 (2013).
72. Zhang, H., Gao, S., Lercher, M. J., Hu, S. & Chen, W. H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic acids research*. **40**(W1), W569–W572 (2012).
73. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExpASY server (pp. 571–607). Humana Press (2005).
74. Thomas-Chollier, M. *et al.* RSAT: regulatory sequence analysis tools. *Nucleic acids research*. **36**, suppl 2, W119–W127 (2008).

Acknowledgements

We would like to thank the Executive Director of the National Agri-Food Biotechnology Institute (NABI), Mohali, India for funds and support. The research work presented in the manuscript was funded by Department of Biotechnology (DBT), Ministry of Science & Technology, Government of India. PK is thankful to the ICMR, India, for the grant of Junior Research Fellowship (JRF) and Senior Research Fellowship (SRF) towards his Ph.D. The authors are thankful to Gregory Hoover of Clemson University for manuscript editing. The authors are also thankful to DeLCON for online library facility.

Author Contributions

J.R. conceived and designed the experiments. P.K. conducted experiment works, data analysis, and manuscript writing. A.M. had developed mutant lines. H.S., M.S.R., M.S., A.P., P.J. helped in experimental works. D.S. helped in the identification, phylogenetic and conserved motif analysis. V.R. and S.K.V. helped in data analysis and edited the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-35366-8>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018