3 Biotech (2020) 10:77
https://doi.org/10.1007/s13205-020-2052-6

ORIGINAL ARTICLE

Molecular characterization of *teosinte branched1* gene governing branching architecture in cultivated maize and wild relatives

Nitish Ranjan Prakash[1] · Rashmi Chhabra[1] · Rajkumar Uttamrao Zunjare[1] · Vignesh Muthusamy[1] · Firoz Hossain[1]

Received: 18 September 2019 / Accepted: 3 January 2020 / Published online: 29 January 2020

Abstract

We sequenced the entire *tb1* gene in six maize inbreds and its wild relatives (*parviglumis*, *mexicana*, *perennis* and *luxurians*) to characterize it at molecular level. *Hopscotch* and *Tourist* transposable elements were observed in the upstream of *tb1* in all maize inbreds, while they were absent in wild relatives. In maize, *tb1* consisted of 431–443 bp 5′UTR, 1101 bp coding sequence and 211–219 bp 3′UTR. In promoter region, mutations in the light response element in *mexicana* (~ 35 bp and ~ 55 bp upstream of TSS) and *perennis* (at ~ 35 bp upstream of TSS) were found. A 6 bp insertion at 420 bp downstream of the polyA signal site was present among teosinte accessions, while it was not observed in maize. A codominant marker flanking the 6 bp *InDel* was developed, and it differentiated the teosintes from maize. In *Tb1* protein, alanine (12.7–14.6%) was the most abundant amino acid with tryptophan as the rarest (0.5–0.9%). The molecular weight of *Tb1* protein was 38757.15 g/mol except 'Palomero Toluqueno' and HKI1128. R and TCP motifs in *Tb1* protein were highly conserved across maize, teosinte and orthologues, while TCP domain differed for *tb1* paralogue. *Tb1* possessed important role in light-, auxin-, stress-response and meristem identity maintenance. Presence of molecular signal suggested its localization in mitochondria, nucleus and nucleolus. *Parviglumis* and *mexicana* shared closer relationship with maize than *perennis* and *luxurians*. A highly conserved 59–60 amino acids long bHLH region was observed across genotypes. Information generated here assumes significance in evolution of *tb1* gene and breeding for enhancement of prolificacy in maize.

Keywords Evolution · Orthologue · Paralogue · Characterization · *Zea mays*

Introduction

Maize (*Zea mays* L.) assumes great significance as a model crop for studying crop domestication and evolution (Studer et al. 2017). Domestication is a series of selection events in wild relatives, and was practiced by ancient farmers to bring about necessary changes in plant morphology for its better utilization as a source of food, feed, fodder, clothing and medicinal values (Doebley et al. 2006). Specific gene(s) contributing to desirable morphological changes was the target of domestication, thereby leading to the increase in frequency of their desirable alleles in the population (Yao et al. 2019). Subsequent dispersal of crops along with hybridization, polyploidy and mutation added further diversity into the genome (Vann et al. 2015). It is, therefore, important to study the subsequent morphological changes and post-domestication diversification to effectively utilize the germplasm resources and wild relatives in the breeding programmes (Osterberg et al. 2017, Prakash et al. 2019).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-020-2052-6) contains supplementary material, which is available to authorized users.

✉ Firoz Hossain
fh_gpb@yahoo.com; fhossain@iari.res.in

Nitish Ranjan Prakash
nitishranjan240@gmail.com

Rashmi Chhabra
reshu0428@rediffmail.com

Rajkumar Uttamrao Zunjare
raj_gpb@yahoo.com; rajkumaruz@iari.res.in

Vignesh Muthusamy
pmvignesh@yahoo.co.in; vignesh@iari.res.in

[1] Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

Beadle (1939) elaborated that at least five major loci may be responsible for domestication of maize. Doebley (1992) mapped QTLs related to domestication of maize. Later, major domestication loci such as *teosinte branched1* (*tb1*) (Doebley et al. 1995), *teosinte glume architecture1* (*tga1*) (Dorweiler et al. 1993) and *grassy tiller1* (*gt1*) (Whipple et al. 2011) were mapped and cloned. Among these, *tb1* is a major locus controlling branching in maize and has been mapped on the short arm of chromosome 1 (Doebley 1992). *Tb1* is responsible for suppression of axillary bud outgrowth on main stem and development of female inflorescence in maize (Doebley 2004). In maize, *tb1* gene is expressed twice as that of teosinte, and over-expression of the suppressor protein causes reduction of branching in maize as compared to teosinte. This overexpression is due to presence of ~ 12 kb enhancer region (~ 58–69 kb upstream of *tb1* coding sequence) in maize (Doebley et al. 1997; Clark et al. 2006). Insertion of *Hopscotch* (~ 58–64 kb) and *Tourist* (~ 64–69 kb) transposable elements (TEs) within this ~ 12 kb region was observed in maize, while they were absent in teosinte (Zhou et al. 2011; Studer and Doebley 2012). Mutation in *Hopscotch* region may change the tissue specific expression of *tb1* gene resulting into differential branching architecture and inflorescence development. Studer and Doebley (2012) studied the presence of natural variation at *tb1* along with adjoining region in maize as well as teosinte, and found that several regions (genic and enhancer regions) are conserved. Mutation at different places in these regions had a direct effect on the traits associated with ear morphology and development.

Tb1 is a TCP (*Teosinte branched1* of maize; *Cycloidea* of snapdragon; *Proliferating cell nuclear antigen factor1* and *2* of rice) domain transcription factor and possesses role in growth of meristem, initiation of floral primordia, regulation of cell cycle and differentiation (Cubas et al. 1999; Leukens and Doebley 2001). Though several genes such as *enhancer of tb1.2* (*etb1.2*), *tassel replaces upper ears1* (*tru1*), *teosinte glume architecture1* (*tga1*), *barren stalk1* (*ba1*), *grassy tiller1* (*gt1*) and *barren stalk2* (*ba2*) regulate branching architecture in maize (Dorweiler et al. 1993; Gallavotti et al. 2004; Whipple et al. 2011; Yang et al. 2016; Dong et al. 2019; Yao et al. 2019), *tb1* acts as a master regulatory gene affecting the plant and inflorescence architectures (Studer et al. 2017). *Tb1* is a basic helix-loop-helix (bHLH) DNA-binding protein which consists of three conserved domains (Leukens and Doebley 2001). Furthermore, the homologue of *tb1* has also been reported in pearl millet (Remigereau et al. 2011), wheat (Dixon et al. 2018), *Arabidopsis* (Finlayson 2007) rice (Choi et al. 2012) and sorghum (Kebrom et al. 2006).

The recent advancement in molecular biology, sequencing technologies and ease of analyzing big data by bioinformatic tools has given tremendous opportunity for in silico characterization of target gene(s). So far, comprehensive characterization of *tb1* gene among elite maize inbreds in comparison with wild relatives has not been undertaken. Considering the importance of *tb1* in maize, the present study was undertaken to (i) characterize the *tb1* gene at molecular level in diverse maize inbreds and (ii) compare the sequence diversity with its wild relatives (teosinte accessions).

Materials and methods

Plant materials

Ten genotypes including five diverse maize inbreds (LM17, UMI1200, HKI1128, BML7 and CML425), one inbred (MGUSP101) developed from 'Sikkim Primitive' (prolific landrace of maize) and four teosinte accessions (*Zea mays* ssp. *parviglumis*, *Zea mays* ssp. *mexicana*, *Zea perennis* and *Zea luxurians*) were used in the present study. The details of genotypes have been presented in Table 1. Leaf samples were collected from the pot grown plants after 28 days of germination and DNA isolation was carried out as per standard protocol by CTAB-method.

Table 1 List of genotypes used for sequence analysis

S. no.	Genotype	Species	Institution
1.	Zm_LM17	*Zea mays* ssp. *mays*	PAU, Ludhiana, India
2.	Zm_UMI1200	*Zea mays* ssp. *mays*	TNAU, Coimbatore, India
3.	Zm_HKI1128	*Zea mays* ssp. *mays*	CCS-HAU, Uchani, India
4.	Zm_MGUSP101	*Zea mays* ssp. *mays*	IARI, New Delhi, India
5.	Zm_BML7	*Zea mays* ssp. *mays*	PJTSAU, Hyderabad, India
6.	Zm_CML425	*Zea mays* ssp. *mays*	CIMMYT, Mexico
7.	Zm_Luxurians	*Zea luxuriance*	CIMMYT, Mexico (EC889920)
8.	Zm_Perennis	*Zea perennis*	CIMMYT, Mexico (EC889921)
9.	Zm_Parviglumis	*Zea mays* ssp. *parviglumis*	CIMMYT, Mexico (EC889910)
10.	Zm_Mexicana	*Zea mays* ssp. *mexicana*	CIMMYT, Mexico (EC889911)

Amplification, cloning and sequencing of *tb1* gene

The ~ 3.5 kb region from B73 genome (Chr1:270,552,500…270,556,000 on B73 RefGen_v4) encompassing the *tb1* gene sequence along with promoter was retrieved from MaizeGDB (Portwood et al. 2018). The sequence was used to design 13 overlapping primer pairs spanning 2.8 kb region of the gene (promoter and coding region) using Primer3web_v4.1.0 (Untergasser et al. 2012) online tool (Table S1). Amplification of overlapping fragments was performed using polymerase chain reaction (PCR) for all ten genotypes. The amplified products in two replicates were sequenced through Macrogen Inc., South Korea. Sequence reads were then used to make consensus complete sequence (~ 2.8 kb approx.) of *tb1* gene for each genotype.

Presence of *hopscotch* and *tourist* TEs

In all 10 genotypes including six maize inbreds (LM17, HKI1128, BML7, UMI1200, CML425 and MGUSP101) and four teosinte accessions (*parviglumis*, *perennis*, *luxurians* and *mexicana*), presence and absence of *Hopscotch and Tourist* TEs in the upstream of the *tb1*gene was confirmed by *InDel* markers as depicted by Studer et al. (2011). List of primers used are given in Table 2.

Sequence retrieval of other species and genotype

Sequences of genotypes of maize and other related species were obtained from various publicly available databases (Table 3). The 3.5 kb region of B73 containing *tb1* gene was used as query for nucleotide BLAST with an expectation value of $\leq 1e^{-5}$ at MaizeGDB (Portwood et al. 2018). The matching sequences from maize landrace 'Palamero

Table 2 List of primers used for identifying the presence and absence of Hopscotch and Tourist transposable element & 6 bp InDel in teosinte

Primer name	Primer sequence (5′ to 3′)	Type of primer	Reference
Tourist			
FM-F0372	ACCAGCAAGCAGCAAGAAAT	Forward	Studer et al. (2011)
IM-R0375	TTGAGTGTCGCCTAGACTGC	Reverse	
RM-R0377	CCTACTTTTTCATCTCCCGC	Internal reverse	
Hopscotch			
FH-F0378	CTGCGATGATGCAAGGAGTA	Forward	
IH-R0379	CTCAATGCATGCCGTTATTG	Reverse	
FH-R0381	CGTTGTCGACAGTCTCCTCA	Internal reverse	
6 bp InDel			
TST1F	TGCAAATCTGATTCGTTCCTT	Forward	Present study
TST1R	AGCCAGGATTATTATCACATAGAAAT	Reverse	

Table 3 List of sequences taken from databases for in silico comparison analysis

S. no.	Genotype	Species	Source	Position on reference genome
1.	Brachypodium	*Brachypodium distachyon*	Phytozome v12	Chr1:8277049..8281449 on Bd21-3_v1.1
2.	Coix	*Coix lacryma-jobi*	FJ487449.1	–
3.	Hordeum	*Hordeum vulgare*	Ensembl	Chr4:17,597,800..17,602,200 on IBSC_v2
4.	Oryza	*Oryza sativa*	IRGSC_Ensembl	Chr3:28427500..28431000 on IRGSP-1.0
5.	Pennisetum	*Pennisetum glaucum*	EF694124.2	–
6.	Sorghum	*Sorghum bicolor*	Ensembl	Chr1:9,465,750–9,471,250 on Sb_NCBIv3
7.	Setaria	*Setaria italica*	Ensembl	Chr9:7,677,500..7,681,500 on Si_v2.0
8.	Triticum_A	*Triticum aestivum*	IWGSC_Ensembl	Chr4A:582,837,800–582,842,200 on IWGSC
9.	Triticum_B	*Triticum aestivum*	IWGSC_Ensembl	Chr4B:30,360,322–30,364,227 on IWGSC
10.	Triticum_D	*Triticum aestivum*	IWGSC_Ensembl	Chr4D:18,462,728–18,466,728 on IWGSC
11.	Zm_B73	*Zea mays* ssp. *mays*	MaizeGDB	Chr1:270552500..270556000 on B73 RefGen_v4
12.	Zm_PalameroT*	*Zea mays* ssp. *mays*	MaizeGDB	AECO01001281.1
13.	Zm_ParB73	*Zea mays* ssp. *mays*	MaizeGDB	Chr5:122942600..122946600 on B73 RefGen_v4
14.	Zm_CML247	*Zea mays* ssp. *mays*	MaizeGDB	Chr1:277608178..277612178 on CML247
15.	Zm_Mo17	*Zea mays* ssp. *mays*	MaizeGDB	Chr1:270000530..270004530 on MO17

Toluqueno' native of Mexico, and inbreds such as CML247 and Mo17 were also taken. Due to non-availability of complete sequence in Genome Browser for 'Palamero Toluqueno', promoter region could not be retrieved. The sequence of most similar paralogue of *tb1* (86% similarity) located at chromosome 5 of B73 (maize inbred) was also considered for analysis. Since this gene is reported to be consisted of only one domain called TCP, we have directly used 3.5 kb region for similarity search in whole genome sequence of *Oryza sativa*, *Triticum aestivum* and *Hordeum vulgare* available at 'EnsemblPlants' database (Kersey et al. 2017), and 3.5 kb region from the best hit genomic sequences encompassing the *tb1* gene were downloaded. Similarly, sequences for *Brachypodium distachyon* were retrieved from whole genome sequence available at Phytozome_v12 (Goodstein et al. 2011). Sequences for *Coix lacryma-jobi* and *Pennisetum glaucum* were obtained from NCBI website by nucleotide BLAST (with *e*-value $\leq 1e^{-5}$) with best hit having complete sequence of homologous *tb1* gene with promoter (Altschul et al. 1990). Altogether 10 orthologues and one paralogue of *tb1* were used for in silico comparison with *tb1* of maize and it wild relatives. Besides, the *tb1* sequence of six maize inbreds and four wild relatives were undertaken in the present study (Table 1), *tb1* sequence of elite inbreds (B73, Mo17 and CML247) and one landrace (Palamero Toluqueno) retrieved from MaizeGDB (Portwood et al. 2018) were also taken for comparison. The details of these 15 sequences used for in silico comparison are given in Table 3.

Gene prediction and phylogenetic tree

All genomic sequences were then analyzed individually using online gene prediction software FGENESH (Solovyev et al. 2006), which predicted the number and position of exons, introns, splice sites, poly-A sites, transcription start site, mRNA sequence and protein sequence. Complete gene sequence, predicted mRNA and protein sequence of all genotype were then aligned using CLUSTALW and MUSCLE, respectively, in MEGA v7.0 (Kumar et al. 2016). The nucleotide sequences were then trimmed to have a consensus start and end point. Phylogenetic tree using mRNA and protein sequences separately was constructed by Neighbour-Joining method with 3000 bootstrap in MEGA v7.0.

Sequence analysis of *tb1* gene

Aligned nucleotide and protein sequences were manually analyzed for presence of conserved region, deletion, duplication, insertion and point mutation. Transition-Transversion bias, nucleotide composition and Tajima's Neutrality test were calculated using MEGA_7.0 and DnaSP (Rozas et al. 2017). The amino acid composition of predicted *Tb1* protein of all genotypes was calculated using PEPSTATS (Chojnacki et al. 2017). The chemical and physical parameters of protein such as molecular weight, theoretical isoelectric point (pI), total number of positively and negatively charged residue, extinction coefficient (M^{-1} cm^{-1}), instability index, aliphatic index and grand average of hydropathicity (GRAVY) were calculated using ProtParam tool at ExPASy (Gasteiger et al. 2005).

Structural analysis of *Tb1* protein

Secondary structure prediction of *Tb1* protein was carried out using SOPMA (Self-Optimized Prediction Method with Alignment) (Geourjon and Deleage 1995). Conserved domain in protein sequences were predicted using and NCBI-CDD (Marchler-Bauer et al. 2016). The folding state and disorder of secondary structure of predicted proteins was generated by FoldIndex server (Prilusky et al. 2005) and DISOPRED3 (Jones and Cozzetto 2014), respectively. Disulfide bond connectivity between cysteine residues was predicted using DiANNA1.1 (Ferre and Clote 2006).

Modeling of *Tb1* protein

Threading based modeling of protein for MGUSP101 was done by I-TASSER server (Yang et al. 2015). The best model among the top five was selected. Stereo-chemical property of predicted model was evaluated using RAMPAGE for Ramachandran plot (Lovell et al. 2003).

Functional characterization of *tb1* gene

Identification of promoter component was done using PlantCARE database (Lescot et al. 2002). Motif prediction was done using MOTIF Search and MEME tools (https://www.genome.jp/tools/motif/; Bailey et al. 2009). Prediction of function of motif (predicted by MEME) was done using GOMo (Buske et al. 2010) and other online resources. For functional prediction of protein sequences, we employed FFPred_3 at PSIPRED server (Cozzetto et al. 2016). Accessibility of solvent to *Tb1* protein was estimated by RaptorX ACC program (Wang et al. 2016). Sub-cellular localization of protein was predicted using TargetP_1.1 server (Emanuelsson et al. 2000). To determine the membrane helix and topology prediction, HMMTOP and MEMSAT-SVM programs from the PSIPRED server (Nugent and Jones 2009) were employed.

Results

Hopscotch and *tourist* TEs

InDel markers revealed that all six maize inbreds (LM17, HKI1128, BML7, UMI1200, CML425 and MGUSP101)

possessed insertion of both *Hopscotch* and *Tourist* TEs in the upstream of the gene. However, both TEs were absent in all four teosinte accessions (*parviglumis*, *perennis*, *luxurians* and *mexicana*). Gel picture depicting the presence of *Tourist* and *Hopscotch* transposable element is given in Fig. 1a, b.

Sequence of *tb1* gene

The consensus nucleotide sequence of *teosinte branched1* gene of maize inbreds viz, LM17 (GenBank accession number—MN842297), HKI1128 (GenBank accession number—MN842299), BML7 (GenBank accession number—MN842301), UMI1200 (GenBank accession number—MN842298), CML425 (GenBank accession number—MN842306) and MGUSP101 (GenBank accession number—MN842300) were generated and subsequently submitted to have the accession number. Consensus sequences of teosinte accessions viz., *Zea luxurians* (GenBank accession number—MN842302), *Zea perennis* (GenBank accession number—MN842303), *Zea mays* ssp. *parviglumis* (GenBank accession number—MN842304) and *Zea mays* ssp. *mexicana* (GenBank accession number—MN842305) were also generated and made available in the public domain.

Gene structure

Sequence analysis of *tb1* gene and orthologues among 10 genotypes including maize and teosinte accessions, and 15 sequences retrieved from the public domain revealed presence of single exon with a range of 1029–1137 nucleotides. In all maize accessions, *tb1* exon was 1101 bp long. *Tb1* protein length varied among 25 gene sequences from 342 to 378 amino acid, while all maize inbreds were having 366 amino acid based protein. *Parviglumis* was having mRNA and protein of equal length to maize, while other teosinte accessions (*luxurians*, *perrenis* and *mexicana*) were having shorter mRNA (1089–1095 nucleotide) and protein (362–364 amino acids). The transcription start site (TSS) for *tb1* was found to be located 289–1205 bp upstream of coding sequence. In all maize genotypes, it was located at 431–438 bp upstream of coding sequence start site. In *parviglumis* and *perennis*, TSS was found to be located as in case of maize. Similarly poly-A site was located at 1136–1991 bp downstream of coding start site. In all maize and *parviglumis* accessions, poly-A site is 1320 bp upstream of coding sequence start site. The predicted exon length, promoter position, protein length, poly-A site and TSS for all genotypes has been presented in Table S2. A common consensus gene structure for all maize genotype is displayed in Fig. 1(c).

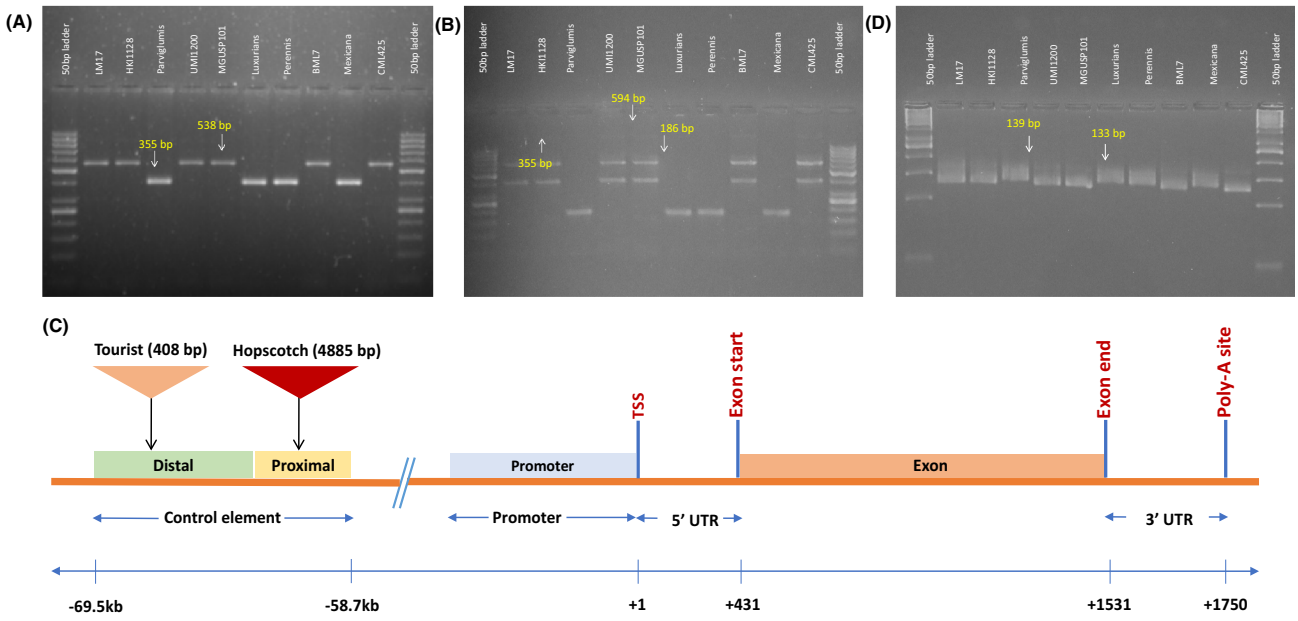


Fig. 1 Gel picture showing the presence and absence of **a** *Hopscotch* Region (with all three primers Forward, Reverse and Internal Reverse); F + R only amplified in Teosinte and gave ~ 355 bp product, F + IR only amplified in Maize and gave ~ 583 bp product. **b** *Tourist* transposable element (with Forward, Reverse and Internal reverse primer); F + R produced ~ 594 bp product in Maize and ~ 186 bp product in Teosinte; F + IR not amplified in teosinte but produced ~ 383 bp product in Maize; **c** Gene structure of *teosinte branched1* gene based on FGENESH for genotype MGUSP101 (representing *tb1* gene structure in all maize genotypes) and **d** Gel electrophoresis showing the presence of ~ 6 bp insertion in all teosinte accessions used in study. Product size amplified was ~ 133 bp in all maize and ~ 139 bp in all teosinte

Sequence variation

The multiple sequence alignment of 10 genotypes (maize inbreds and teosinte accessions) along with Mo17, B73 and CML247 revealed 6 bp insertion specific to all teosinte accessions at 420 bp downstream of predicted poly-A site. This has been confirmed by developing InDel-based markers from the flanking region (Table 1), the PCR amplicons for which were resolved on 4% metaphor agarose (Fig. 1d). A 3 bp deletion of 5′-CCG-3′ was also present in all the teosinte accessions at 17 bp downstream of predicted poly-A site. Similarly, a 6 bp deletion was present in all teosinte except *parviglumis* in the coding region. Analysis using DnaSP for 13 sequences (10 sequences + 3 retrieved from database) revealed total number of 194 polymorphic sites with 205 mutations with a nucleotide diversity (Pi) of 0.01386, while Tajima's *D* was found to be − 2.06018 (significant at 0.05). Since promoter sequence of 'Palamero Toluqueno' was unavailable in the public domain, coding sequence was used for identification of polymorphic sites. For all mRNA of 14 genotypes, MEGA (excluding the gaps and missing data ~ 1089 bp in final dataset) predicted overall transition/transversion bias (R) of 0.801. Tajima's neutrality test based on coding sequence using DnaSP revealed 17 polymorphic sites and 19 mutations with nucleotide diversity of 0.003340 and Tajima's test statistic (D) as − 1.62874 (not significant).

Phylogenetic tree

Phylogenetic tree construction using neighbor joining method at 3000 bootstrap for both protein and mRNA sequences revealed two major clusters, A and B (Fig. 2). Cluster-A consisted of sequences of temperate plants of genera *Triticum* (all three orthologue present in A, B and D genome of wheat), *Hordeum* and *Brachypodium*. Cluster-B had *Oryza* at single node and all other C_4-plants making a separate sub-cluster. Within the sub-cluster of cluster-B, the orthologue of *tb1* in maize clustered with *Sorghum* and *Coix*.

Physiochemical features

Primary structural analysis revealed alanine (12.7–14.6%) being the most abundant amino acid and tryptophan as the rarest (0.5–0.9%). In *tb1* paralogue of B73, alanine was 12%, while tryptophan was 1.4%. Across maize genotypes, alanine was 14.5% and tryptophan was 0.8%. Asparagine and glutamine were absent in all the protein sequences. A figure showing the types of amino acids for MGUSP101 is given in Figure S1. Average molecular weight of *Tb1* protein ranged from 36690.02 g/mol in *Hordeum vulgare* to 40404.69 g/mol in *Brachypodium distachyon*. In all maize genotypes, it was 38757.15 g/mol, except 'Palomero Toluqueno' and HKI1128. The predicted isoelectric point of *Tb1* protein ranged from 6.72 (for *tb1* paralogue in maize) to 8.74 (in *Coix*). Predicted isoelectric point was 8.3 in all maize and teosinte accessions. *Tb1* protein of all the genotypes had

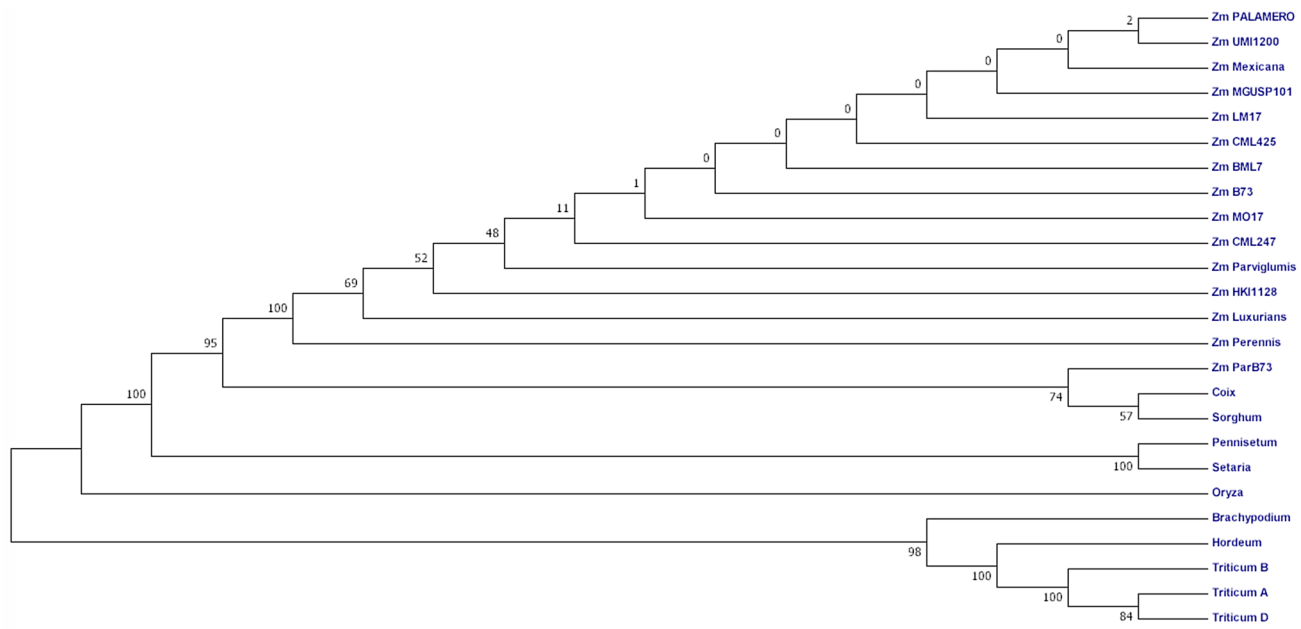


Fig. 2 Phylogenetic tree based on protein sequences in MEGA using Neighbour-joining method (The optimal tree with sum of branch length 1.08004339 has been shown)

instability index of > 40, ranging from 46.31 to 58.17. Across maize, it was 47.37 except 46.31 in 'Palomero Toluqueno'. Number of positively and negatively charged side chains in *Tb1* protein are 44 and 41 in all maize and teosinte, respectively, while varying in other genus. GRAVY (Grand Average of Hydropathy) ranged from −0.591 to −0.729 across genus for *Tb1* protein. Aliphatic index was 57.81 in all maize genotypes and it ranged from 57.54 to 59.25 in teosinte, and 55.19 to 61.64 in other genus. The molar extinction coefficient recorded at 280 nm measured in water, considering all pairs of cysteine form cystine ranged from 28670 to 30285 M^{-1} cm^{-1} in all genotypes. Molar extinction coefficient for *Tb1* paralogue in maize was abruptly high at 42900 M^{-1} cm^{-1} as compared to others, whereas in all maize and teosinte, it was 28920 M^{-1} cm^{-1}.

Secondary structure

Random coiling (41.76 to 49.28%) was the most abundant secondary structure followed by α-helix (30.42 to 41.23%), extended strand (8.72 to 14.81%) and β-turn (4.03 to 6.4%) (Figure S2). A figure depicting the predictions for secondary structure of individual amino acid in predicted protein of MGUSP101 has been given in Figure S2. Based on in silico analysis, we inferred that *Tb1* protein contains 4–8 unfolded regions across genera. Out of total 366 amino acids in maize genotypes, 133 (36.33%) remained unfolded. All maize and teosinte genotypes had four unfolded regions (Figure S3). Length of the largest unfolded region varied from 101 to 118 amino acids with all maize genotypes having 104 amino acids in the largest unfolded region. A graph showing disordered amino acids in MGUSP101 has been given in Figure S4. Disulfide bond formation predicted that there were 2–9 sites for disulfide bond formation, with only two in *Oryza sativa* and nine in all teosinte and maize accessions. In *Tb1* paralogue of maize, only eight disulfide sites were present. In all maize and *parviglumis*, sites for disulfide bonds formation were at 26, 29, 85, 112, 171, 203, 296, 307 and 348 amino acid positions. These resulted into four disulfide bonds between 26–348, 29–171, 85–296 and 112–203 amino acids.

Modeling of protein

Template based homology modeling is the best way to explore the 3-D structural model of any protein sequence, but for *Tb1* protein enough templates to cover maximum part of protein sequences were not found at SWISS-MODEL. Therefore, a threading based modeling was performed for MGUSP101 (Fig. 3a, b). Best model had Confidence (C)-score of −2.39, Template modeling (TM)-score of 0.43 ± 0.14 and RMSD of 12.3 ± 4.4Å. The consensus structure for MGU-SP-101 has been presented in Fig. 3a, b. Ramachandran plot analysis for stability of protein model revealed that 256 (70.3%) lied in favored region, 66 (18.1%) lied in allowed region and 42 (11.5%) lied in outlier region (Fig. 3c). A graph showing estimated local accuracy of the model by algorithm 'ResQ' has also been given in Fig. 3a, b.

Domain prediction

Domain prediction revealed that *Tb1* is a TCP domain protein, as reported earlier by Leukens and Doebley (2001). TCP domain was predicted to be 149 to 157 amino acids in length. All maize and teosinte genotypes were having 154 amino acids TCP domain. Result showing the TCP domain were shown in Figure S5. *Tb1* paralogue of maize had only 151 amino acids based long TCP domain. Motif prediction using MOTIF Search revealed two motifs, TCP (at position 104 to 162 amino acids) and R motif (at position 236 to 253 amino acids) in all maize and *parviglumis*. Motif identification at MEME_Suite_5.0.5 revealed best 12 motifs (range of motifs = 6 to 50 amino acids) (Fig. 4). Of these, motif at 105–154 amino acids and 228–268 amino acids were predicted to be TCP and R motifs. The TCP motif consisted of 59 amino acids and R-motif had 18 amino acids in all genotypes. Sequence alignment of R-motif revealed that it is completely conserved in *Tb1* of all maize and teosinte genotypes and its paralogue in maize. Sequence alignment of TCP motif (104 to 162 amino acids) displayed sequence as the most conserved in all maize and teosinte accessions. Major variation in this region was found for *Tb1* paralogue, even greater than other species. Sequence alignment of TCP and R motifs are given in Fig. 5.

Functional characterization

The *cis*-acting regulatory elements were identified for function using PlantCARE database. Only the elements having matrix score ≥ 5 on any of the strand were selected. Apart from prediction of promoter elements such as TATA-box, GC-box and CAAT-box, it has also predicted the elements responsible for regulating various pathways. *Tb1* is a transcription factor and it is having various regulatory roles such as abscisic acid responsiveness, anaerobic induction, light responsiveness, auxin responsiveness, meristem expression, MYBHv1 binding site, MeJA responsiveness, low temperature responsiveness, drought inducibility, defense and stress responsiveness. Mutation has been found in *mexicana* and *perennis* at a light responsive motif (5′AATCTCATCC3′ identified from *Pisum sativum*) at ~ 35 bp upstream from the TSS in promoter region. In addition, mutation in light response motif (5′TCCCTCA3′) at ~ 55 bp in the upstream from TSS was also observed in *mexicana*. It is possible that change in such motif may cause spatio-temporal expression of *tb1* resulting into high tillering. GOMo was unable to

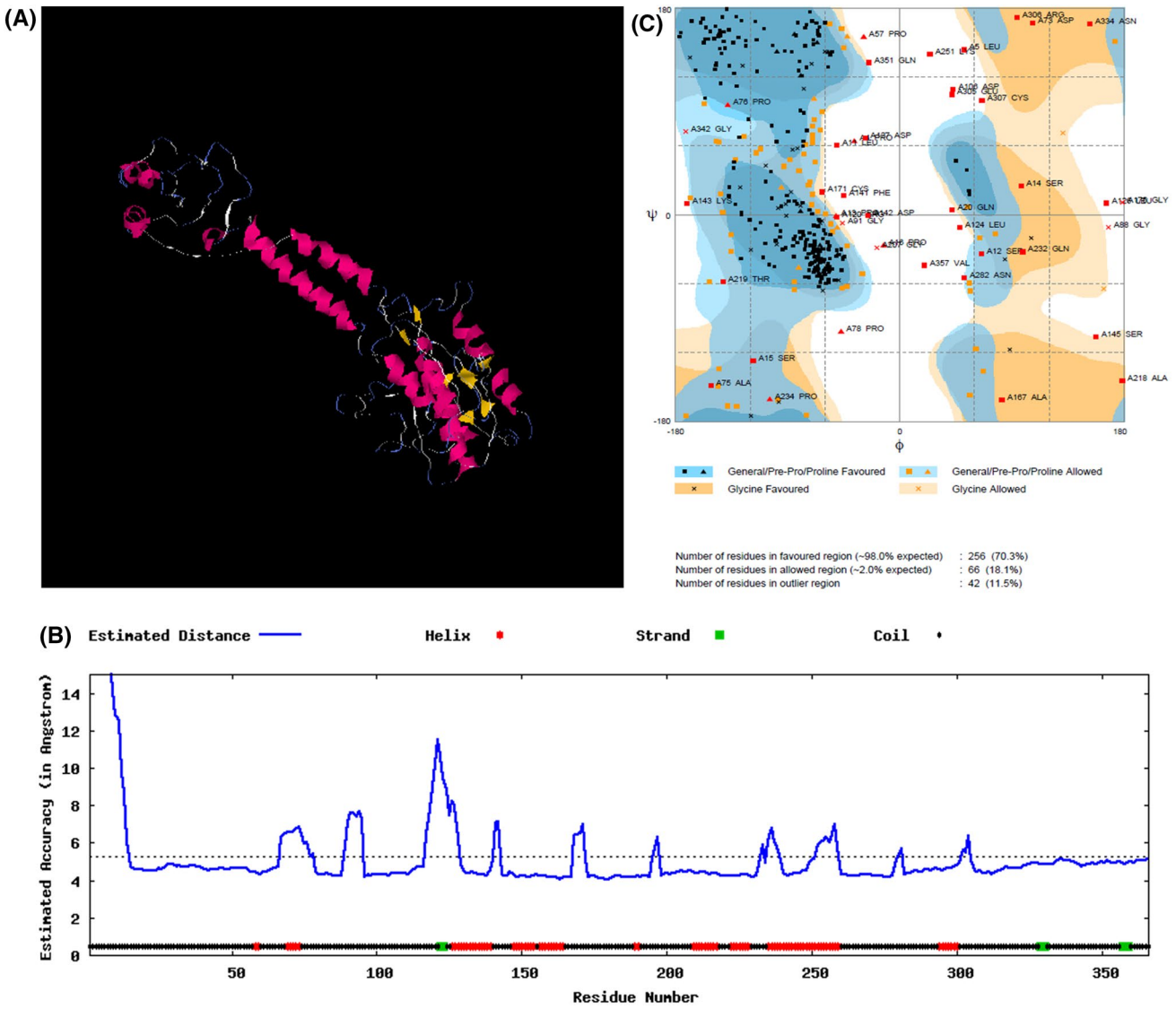
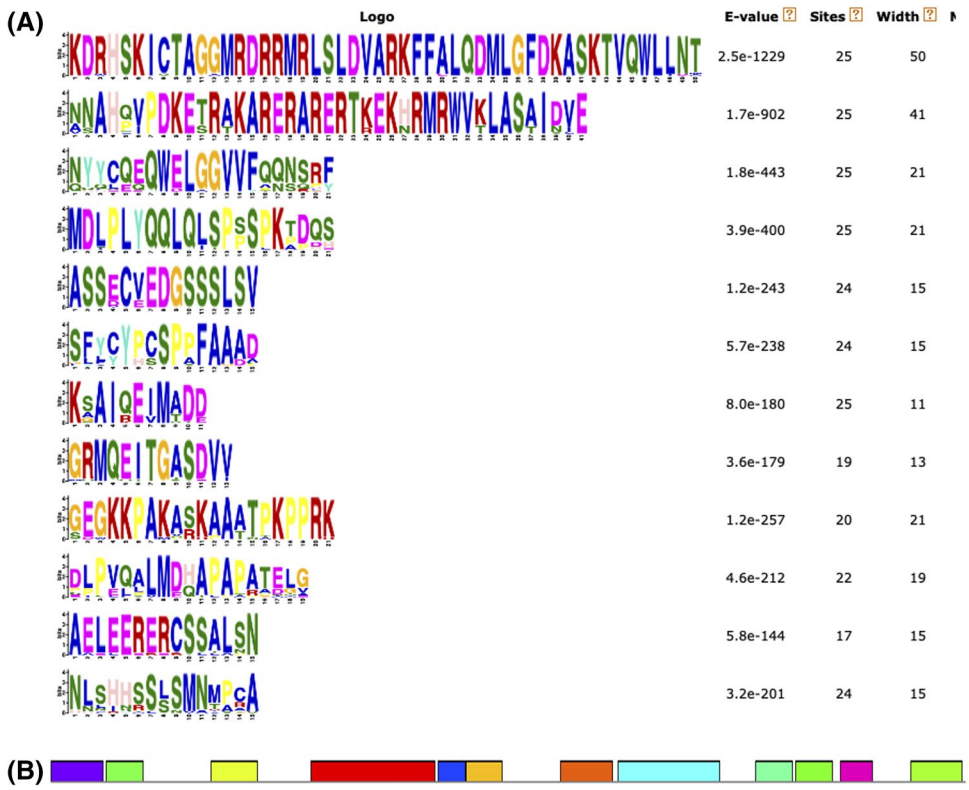


Fig. 3 **a** Thread-based model of *Tb1* protein by I-TASSER server for maize; **b** estimated local accuracy of model; **c** Ramachandran plot analysis for best model predicted by I-TASSER for *Tb1* protein of maize using RAMPAGE server

predict the function of motif predicted by MEME. The biological, molecular and cellular function prediction (> 0.9 probability) using FFPred server showed that *Tb1* protein may be involved in regulation of gene expression, transcription, metabolic process, nitrogen metabolism and nucleic acid binding transcription factor activity. TargetP_1.1 results (0.113 for cTP; 0.257 for mTP; 0.027 for SP; 0.689 for others) and FFPred cellular component has predicted mitochondrion, nucleolus and nuclear region as expressed region. Besides, it has also been predicted to be expressed in other sub-cellular components. Solvent accessibility prediction using RaptorX ACC program for 366 amino acids of *Tb1* protein of MGUSP101 revealed 235 amino acids (64.2%) as exposed, 92 amino acids (25.14%) as buried and 39 amino acids (10.66%) as intermediate based on relative solvent accessibility (exposed: 40 to 100%; intermediate: 10 to 40%; buried: 0 to 10% of RSA). Trans-membrane topology prediction using MEMSAT-SVM for predicted protein of MGUSP101 predicted only one helix at 124 to 139 amino acids.

Discussion

Branching and inflorescence development in cereals is a complex process. Various genes associated with cell division and proliferation, meristem growth and differentiation, meristem identity maintenance, auxin -transport, -production and -response factors, stress responsiveness and genes associated with floral organ development are



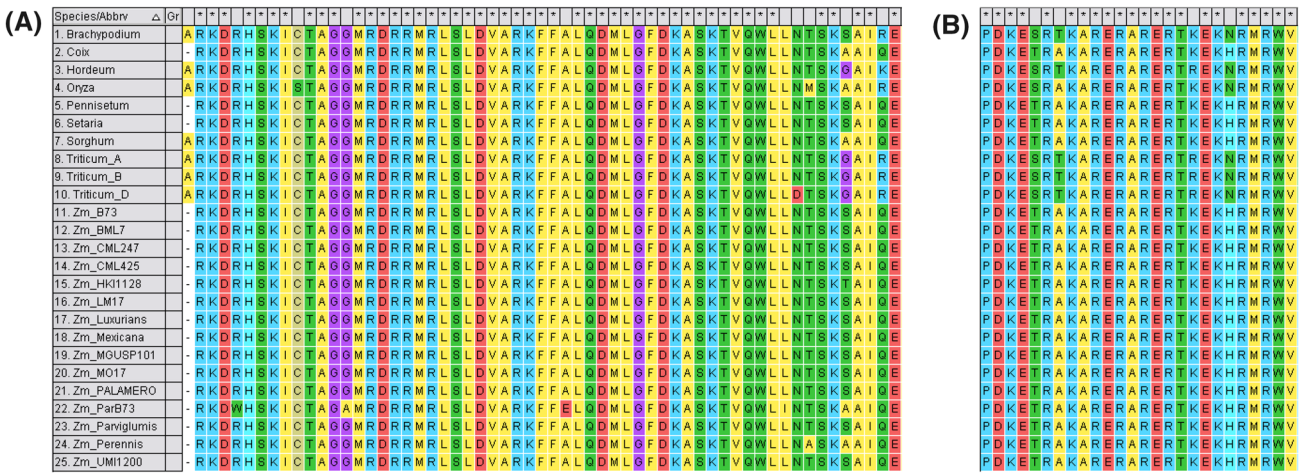


Fig. 5 Sequence alignment of **a** TCP and **b** R motif in MEGA 7 using Clustal_W

involved (Doust 2007; Barazesh and McSteen 2008; Rameau et al. 2015). In maize, genes such as *tb1*, *tga1*, *gt1*, *tru1*, *ba1*, *ba2* and *ramosa1* are involved in the process of domestication (Dong et al. 2019). Of these, *tb1* plays the pivotal role in modifying the plant architectures (Studer et al. 2011). Considering its paramount importance in regulating the axillary bud formation pathway in maize, we characterized the *tb1* gene and the encoded protein at molecular level and compared among cultivated maize, its wild relatives, and orthologues present in other cereal crops.

Molecular characterization

Hopscotch and *Tourist* TEs were present in the upstream of *tb1* gene of maize inbreds, while it was absent in the teosinte accessions. Zhou et al. (2011) first reported this while experimenting with the *tb1* gene in maize and teosinte. These TEs

acts as enhancer for transcription of *tb1*, and it expresses two folds in maize. Due to lack of these TEs, less expression of *tb1* is observed among teosinte accessions (Studer et al. 2011). Since *Tb1* protein acts as suppressor for genes involved in the axillary bud formation pathway, higher prolificacy (more ears per plant) is observed in teosinte, while apical dominance resulting in one ear per plant is a common phenomenon in cultivated maize (Vann et al. 2015). Interestingly, MGUSP101 is an inbred line developed from a prolific landrace 'Sikkim Primitive', which bears up to 7–9 ears per plant in its area of adaptation (Sikkim province of India). MGUSP101 bears up to five ears per plant in Delhi conditions as compared to maize inbreds that generally bear one ear per plant (Prakash et al. 2019). The analyses suggested that MGUSP101 did not differ from maize inbreds in relation to insertion of *Hopscotch* and *Tourist*, thereby suggesting the role of other mechanisms in conferring prolificacy in 'Sikkim Primitive'.

Tb1 is a small sized (362 to 366 amino acids) TCP-domain transcription factor across maize and wild relatives, and therefore, it becomes easy to reach into chromatin unit and activate the genes (Maeshima et al. 2015). Orthologues of *tb1* identified in *Oryza* (*fine culm1*), *Triticum* (*TaTB1*), *Sorghum* (*SbTB1*), *Hordeum* (*INT-C*), *Pennisetum* (*PgTB1*) and *Arabidopsis* (*BRC1*) have the same TCP domain protein of comparable length (342–378 amino acids). Gene specific transcription factors are generally small sized proteins having nucleotide-binding domain, and are responsible for recruiting the general transcription factors for transcription. As maize is closely related to its progenitor *Balsas* teosinte (*Zea mays* ssp. *parviglumis*) (Doebley 1992), the coding sequence length (1101 bp), protein length (366 amino acids) and poly-A site are same, while they differ from other teosinte species (*luxurians, perennis* and *mexicana*). Promoter site, promoter component, redundancy, neofunctionalization, non-functionalization, sub-functionalization and distance from coding sequence may differ in different species for homologous genes (De Bodt et al. 2006). Similar trend was also observed in the present study while comparing the *tb1* in maize with other genera of grass family.

In our present study, *tb1* gene is found to be without intron, and it appears to be in consonance with the other predicted TCP domain proteins reported in strawberry (Wei et al. 2016), cotton (Ma et al. 2014; Zheng et al. 2018), wheat (Zhao et al. 2018), cassava (Lei et al. 2017) and *Arabidopsis* (Li 2015). Speculation is that intron may be gained during the process of evolution or may be lost (Chai et al. 2017). Tajima's *D* value is calculated to know if the genic region is departing from neutrality and have any selection pressure. A significant negative value was obtained for *tb1*, which indicated that the gene was under selection pressure and population is not evolving neutrally. While Tajima's *D* test was negative but insignificant for coding region, thereby implying that variation in coding sequences was by genetic drift of mutant alleles which suggest that selection pressure was imposed during domestication by the ancient farming community. A novel 6 bp insertion at 420 bp downstream of the polyA signal site was unique among four teosinte accessions (*parviglumis*, *perennis*, *luxurians* and *mexicana*) and was not observed in cultivated maize. The InDel markers (surrounding the 6 bp polymorphism) developed here can be used to introgress *tb1* allele of teosinte into elite maize inbreds for enhancing the prolificacy in maize. However, possession of 6 bp InDel as observed here needs further validation among other accessions of *parviglumis*, *perennis*, *luxurians* and *mexicana* besides other teosintes (*diploperennis, nicaraguensis* and *huehuetenangensis*).

Genetic relationships

The phylogenetic relationship with protein sequences confirmed the usual evolutionary relationship between members of grass family. All C_4-plants were clustered together thereby depicting their close relationship, while all temperate grasses (wheat, barley and *Brachypodium*) were grouped together. Since *tb1* gene was an extreme target for selection during domestication, all maize genotypes clustered together along with teosinte accessions. Among four teosintes, *parviglumis* and *mexicana* were more close to maize than *luxurians* and *perennis*. It also confirmed *parviglumis* as the progenitor of maize (Beadle 1939). The homologue of *Tb1* present in wheat genome A (342 amino acids), B (344 amino acids) and D (342 amino acids) differed only for 1–2 amino acids, as observed in maize and different teosinte accessions (362–366 amino acids). This suggests that this gene is still under evolution but at a very slow extent (Zheng et al. 2018; Chai et al. 2017). TCP family of transcription factor conditioned by *tb1* has evolved through duplication, genomic rearrangement, functional and structural divergence, therefore, are having diverse spatiotemporal functions, and it may be possible to obtain new function as paralogue. TCP proteins in maize are thought to be evolved by segmental duplication and later neo-functionalization. The close relationship between paralogue of *tb1* in maize inbred (B73) with *Sorghum* and *Coix* is possibly due to the fact that *tb1* paralogue has accumulated enough variations over a period of time that it now possesses much similarity with the *Sorghum* and *Coix* (Chai et al. 2017).

Physiochemical features of protein

Based on molecular weight, *Tb1* (36690 to 40404.62 g/mol) can be regarded as small protein. Most of the transcription factors are small proteins which can easily penetrate through the chromatin (Maeshima et al. 2015). Since the isoelectric point (pI) is near to 8.0 in most of the genotypes, it is

likely to precipitate in basic buffers. While in two species (*Hordeum* and *Brachypodium*) and *tb1*-paralogue, the pI is less than 7.0, indicating their precipitation in acidic buffers (Table 4). This is also evident from the phylogenetic tree, where *Hordeum* and *Brachypodium* were in different cluster compared to other maize and teosinte accessions. This information can be used in purification of *Tb1* protein. Size of protein and its composition were generally more conserved in genotypes than their isoelectric points (Nandi et al. 2005). Isoelectric point in maize and teosinte were 8.3, indicating highly conserved functions. All the *Tb1* proteins are predicted to be unstable based on Instability index (Ii > 40), this shows their highly fragile nature in vivo with an estimated half-life, $T_{1/2} < 5$ h (Rogers et al. 1986). Aliphatic Index (Ai) indicates about the thermostability of protein based on the relative volume occupied by the aliphatic side chains of alanine, valine, isoleucine and leucine. More aliphatic index indicates more thermostability of protein (Ikai 1980). Aliphatic index ranging from 55.19 to 61.64 represents that *Tb1* protein can sustain wide range of temperatures (Arora et al. 2009; Gupta et al. 2012). Grand Average of Hydropathy (GRAVY) value of range -0.591 to -0.729 also indicates that *Tb1* protein can have better interaction with water and thus hydrophilic in nature (Arora et al. 2009).

A highly conserved 59 to 60 amino acids long bHLH region was observed across genotypes (Cubas et al. 1999). Two domains, viz., TCP and R in *tb1* were observed in all species in the study (Kosugi and Ohashi 1997). A TCP-protein with R-domain is called as TCP-C type (Cubas 2004). TCP domain is well characterized in maize (Chai et al. 2017), wheat (Zhao et al. 2018), cotton (Ma et al. 2014; Zheng et al. 2018), cassava (Lei et al. 2017) and strawberry (Wei et al. 2016). Multiple Em for Motif Elicitation (MEME) identified 12 putative motifs in the *Tb1* protein sequences of all genotypes. Similarly MEME has also identified multiple putative motifs with unknown function while studying genome-wide TCP genes of rice (Danisman 2016), Arabidopsis (Li et al. 2015), wheat (Zhao et al. 2018), cotton (Ma et al. 2014; Zheng et al. 2018), maize (Chai et al. 2017), strawberry (Wei et al. 2016) and cassava (Lei et al. 2017). R-motif was highly conserved across maize, teosinte, orthologues and paralogue. Though TCP domain was highly conserved in all teosinte, maize and orthologues, it differed for *Tb1* paralogue. This may be true, because paralogue of *Tb1* in maize genome may have very different functions and different genes as their target sites.

Table 4 Physicochemical properties of predicted protein by ProtParam

S. no.	Genotype/accession	MW	pI	EC	II	Ai	GRAVY	R^-	R^+
1.	Zm_LM17	38,757.15	8.30	28,920	47.37	57.81	−0.611	41	44
2.	Zm_UMI1200	38,757.15	8.30	28920	47.37	57.81	−0.611	41	44
3.	Zm_HKI1128	38,757.15	8.30	28920	47.37	57.81	−0.611	41	44
4.	Zm_MGUSP101	38,757.15	8.30	28920	47.37	57.81	−0.611	41	44
5.	Zm_BML7	38,757.15	8.30	28920	47.37	57.81	−0.611	41	44
6.	Zm_CML425	38,757.15	8.30	28920	47.37	57.81	−0.611	41	44
7.	Zm_Luxurians	38,626.06	8.30	28920	48.29	58.54	−0.605	41	44
8.	Zm_Perennis	38,483.95	8.30	28920	47.93	59.25	−0.591	41	44
9.	Zm_Parviglumis	38,783.19	8.30	28920	48.42	57.54	−0.620	41	44
10.	Zm_Mexicana	38,643.05	8.30	28920	46.89	58.13	−0.612	41	44
11.	Brachypodium	40,404.69	6.90	30285	49.06	56.69	−0.718	45	44
12.	Coix	39,766.24	8.74	30285	52.63	59.57	−0.636	42	47
13.	Hordeum	37,077.32	6.84	28670	54.76	60.95	−0.681	41	40
14.	Oryza	40,384.56	7.08	28,545	58.17	55.19	−0.729	42	41
15.	Pennisetum	38,544.03	8.44	28,795	48.52	59.19	−0.652	41	44
16.	Sorghum	39,760.49	8.70	28,920	49.62	60.57	−0.633	41	46
17.	Setaria	38,374.94	8.39	28,795	52.54	61.64	−0.599	40	43
18.	Triticum_A	36,766.12	7.74	30,160	51.01	60.99	−0.636	39	40
19.	Triticum_B	37,078.44	8.27	30,160	53.48	59.77	−0.671	40	42
20.	Triticum_D	36,690.02	8.27	30,160	50.47	61.26	−0.649	39	41
21.	Zm_B73	38,757.15	8.30	28,920	47.37	57.81	−0.611	41	44
22.	Zm_PalameroT	38,761.14	8.30	28,920	47.37	57.81	−0.609	41	44
23.	Zm_ParB73	38,894.31	6.72	42,900	50.52	58.02	−0.635	42	41
24.	Zm_CML247	38,757.15	8.30	28,920	47.37	57.81	−0.611	41	44
25.	Zm_Mo17	38,757.15	8.30	28,920	47.37	57.81	−0.611	41	44

Protein modeling

Secondary structure is more conserved than primary structure (protein sequences) as the most common amino acid substitution in protein are synonymous, without affecting the protein structure, hence this may be a great tool to study evolutionary relationship (Reehana et al. 2013). Correct prediction of secondary structure is required for prediction of 3-D structure, ligand binding site, domain and motif structure and functional prediction. SOPMA and secondary structure prediction using PSIPRED server revealed random coil as most common secondary structure in *Tb1* protein. Random coils are a kind of structure which predominantly depends upon the surrounding conditions of amino acids and least affected by the overall structure as it may attain any structure of its own to fit into the external condition of polypeptide. Since *Tb1* is a TCP domain bHLH transcription factor, we could find a significant proportion of α-helix, extended strand and β-turn (Kosugi and Ohashi 1997; Chai et al. 2017). The disordered region of 133 amino acids in maize may have a role in molecular recognition, binding activity and regulatory processes, and are, therefore, functionally important (Cozzetto and Jones 2013; Wright and Dyson 2015). Number of cysteine residue differed in different homologue of *Tb1* in different species. It resulted into different number of disulfide bond formation, and ultimately affects the protein structure and functions. Protein modeling is an important way to decide its three dimensional structure, interaction, homo-heteromerization, ligand binding site, domain structure and protein function (Yang et al. 2015; Moturu et al. 2018).

Functional characterization

Genes encoding transcription factor are known to have several pleiotropic effects as it regulates a series of genes responsible for a specific phenotype. Their expression is most commonly influenced by the environmental factors such as light, drought, diseases, stress, hormonal activity and various other external factors. They respond to such stimuli via a signaling protein binding to specific location in promoters. TCP proteins are known to have role in branching (Doebley et al. 1997), bud outgrowth (Takeda et al. 2003), working in strigolactones pathway to regulate axillary bud (Minakuchi et al. 2010), regulation of shoot apical meristem development (Hubbard et al. 2002), leaf and flower development (Aguilar-Martinez and Sinha 2013; Koyama et al. 2007), and expression of cell-cycle genes (Peng et al. 2015). PlantCARE database was searched for presence of functional element which identified several promoter and genic component depicting the potential role of *tb1* in response to several biotic (auxin, abscisic acid and meristem growth) and abiotic (anaerobic induction, light, drought and low temperature) factors. FFPred server predicted its role in regulation of gene expression, metabolic processes and nitrogen metabolism. Several authors have reported the various role of *tb1* in other crops, such as panicle and spike development in wheat (Zhao et al. 2018; Dixon et al. 2018), tiller development in wheat (Lewis et al. 2008), lateral branching in rice (Takeda et al. 2003), tillering in rice (Choi et al. 2012), lateral spikelet fertility in barley (Ramsay et al. 2011) and axillary bud outgrowth in sorghum (Kebrom et al. 2006). Light signaling may have an effect in developmental pathways in maize, as plant grown in shade develop only one ear, while in isolation generally develops more than one ear. Such kind of influence may be the outcome of shade avoidance syndrome (Kebrom and Brutnell 2007, Whipple et al. 2011). Similarly, Kebrom et al. (2006) reported that Phytochrome-B represses the *tb1* expression and negatively regulates axillary bud outgrowth in sorghum. Our prediction based on TargetP1.1 and FFPred depicts that *Tb1* protein localizes in nucleus and mitochondria. Relative solvent accessibility using RaptorX_ACC program revealed 64.2% of amino acid are exposed, and it correlates with the fact that *tb1* is actively engaged in molecular interaction and should bind to several sites in genome in response to various growth responses.

Conclusion

Tb1 gene encoding bHLH transcriptional factor having TCP domain is the key regulator of branching architecture of maize and teosinte. *Hopscotch* and *Tourist* TEs were absent in teosinte, while TEs were present in the upstream of *tb1*. A 6 bp deletion unique to maize was also identified, and a codominant marker differentiating the teosinte and maize was developed. *Parviglumis* and *mexicana* shared closer relationship with cultivated maize than *perennis* and *luxurians*. R domain is highly conserved in all genotypes, while TCP domain varied in paralogue of *Tb1*. *Tb1* protein besides regulating the axillary bud formation pathway, is also involved in various other regulations including abscisic acid-, light- and auxin- responsiveness. The information generated here holds great importance in the evolution of *tb1* gene and breeding for enhancement of prolificacy in maize.

Acknowledgements First author is thankful to Council of Scientific and Industrial Research for providing CSIR-Senior Research Fellowship during his Ph.D. programme. We thank ICAR-IARI, New Delhi for financial assistance. We sincerely acknowledge CIMMYT, Mexico for providing teosinte accessions.

Authors' contribution NRP: conduct of the experiment; FH: design and plan of experiment; RUZ and VM: in silico analysis; NRP and RC: manuscript drafting; FH: manuscript editing

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest in the publication.

References

Aguilar-Martinez JA, Sinha NR (2013) Analysis of the role of Arabidopsis class I TCP genes AtTCP7, AtTCP8, AtTCP22, and AtTCP23 in leaf development. Front Plant Sci 4:406

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arora N, Banerjee AK, Mutyala S, Murty US (2009) Comparative characterization of commercially important xylanase enzymes. Bioinformation 3(10):446

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37(2):W202–W208

Barazesh S, McSteen P (2008) Hormonal control of grass inflorescence development. Trends Plant Sci 13(12):656–662

Beadle GW (1939) Teosinte and the origin of maize. J Hered 30:245–247

Buske FA, Boden M, Bauer DC, Bailey TL (2010) Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics 26(7):860–866

Chai W, Jiang P, Huang G, Jiang H, Li X (2017) Identification and expression profiling analysis of TCP family genes involved in growth and development in maize. Physiol Mol Biol Plants 23(4):779–791

Choi MS, Woo MO, Koh EB, Lee J, Ham TH, Seo HS, Koh HJ (2012) *Teosinte Branched 1* modulates tillering in rice plants. Plant Cell Rep 31(1):57–65

Chojnacki S, Cowley A, Lee J, Foix A, Lopez R (2017) Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. Nucleic Acids Res 45:W550–W553

Clark RM, Wagler TN, Quijada P, Doebley JF (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescence architecture. Nat Genet 38(5):594

Cozzetto D, Jones DT (2013) The contribution of intrinsic disorder prediction to the elucidation of protein function. Curr Opin Struct Biol 23(3):467–472

Cozzetto D, Minneci F, Currant H, Jones DT (2016) FFPred 3: feature-based function prediction for all gene ontology domains. Sci Rep 6:31865

Cubas P (2004) Role of TCP genes in the evolution of morphological characters in angiosperms. In: Cronk QC et al (eds) Developmental genetics and plant evolution. CRC Press, Boca Raton, pp 262–281

Cubas P, Lauter N, Doebley J, Coen E (1999) The TCP domain: a motif found in proteins regulating plant growth and development. Plant J 18(2):215–222

Danisman S (2016) TCP transcription factors at the interface between environmental challenges and the plant's growth responses. Front Plant Sci 7:1930

De Bodt S, Theissen G, Van de Peer Y (2006) Promoter analysis of MADS-box genes in eudicots through phylogenetic footprinting. Mol Biol Evol 23(6):1293–1303

Dixon LE, Greenwood JR, Bencivenga S, Zhang P, Cockram J, Mellers G, Ramm K, Cavanagh C, Swain SM, Boden SA (2018) *TEOSINTE BRANCHED1* regulates inflorescence architecture and development in bread wheat (*Triticum aestivum*). Plant Cell 30(3):563–581

Doebley JF (1992) Mapping the genes that made maize. Trends Genet 8(9):302–307

Doebley JF (2004) The genetics of maize evolution. Annu Rev Genet 38:37–59

Doebley JF, Stec A, Gustus C (1995) *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics 141(1):333–346

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386(6624):485

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127(7):1309–1321

Dong Z, Alexander M, Chuck G (2019) Understanding grass domestication through maize mutants. Trends Genet 35(2):118–128

Dorweiler J, Stec A, Kermicle J, Doebley J (1993) *Teosinte glume architecture1*: a genetic locus controlling a key step in maize evolution. Science 262(5131):233–235

Doust A (2007) Architectural evolution and its implications for domestication in grasses. Ann Bot 100(5):941–950

Emanuelsson O, Nielsen H, Brunak S, Von Heijne G (2000) Predicting subcellular localization of proteins based on their *N*-terminal amino acid sequence. J Mol Biol 300(4):1005–1016

Ferrè F, Clote P (2006) DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. Nucleic Acids Res 34(2):W182–W185

Finlayson SA (2007) Arabidopsis *Teosinte Branched1-like 1* regulates axillary bud outgrowth and is homologous to monocot Teosinte Branched1. Plant Cell Physiol 48(5):667–677

Gallavotti A, Zhao Q, Kyozuka J, Meeley RB, Ritter MK, Doebley JF, Enrico Pe M, Schmidt RJ (2004) The role of *barren stalk1* in the architecture of maize. Nature 432(7017):630

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and analysis tools on the ExPASy Server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, Totowa, pp 571–607

Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics 11(6):681–684

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(D1):D1178–D1186

Gupta SK, Rai AK, Kanwar SS, Sharma TR (2012) Comparative analysis of zinc finger proteins involved in plant disease resistance. PLoS ONE 7(8):e42578

https://www.genome.jp/tools/motif/. Accessed 10 Apr 2019

Hubbard L, McSteen P, Doebley J, Hake S (2002) Expression patterns and mutant phenotype of *teosinte branched1* correlate with growth suppression in maize and teosinte. Genetics 162(4):1927–1935

Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88(6):1895–1898

Jones DT, Cozzetto D (2014) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31(6):857–863

Kebrom TH, Brutnell TP (2007) The molecular analysis of the shade avoidance syndrome in the grasses has begun. J Exp Bot 58(12):3079–3089

Kebrom TH, Burson BL, Finlayson SA (2006) Phytochrome B represses *Teosinte Branched1* expression and induces sorghum axillary bud outgrowth in response to light signals. Plant Physiol 140(3):1109–1117

Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, Kumar N, Liu Z, Maurel T, Moore B, McDowall MD, Maheswari U, Naamati G, Newman V, Ong CK, Bolser DM, De Silva N, Howe KL, Langridge N, Maslen G, Staines DM, Yates A (2017) Ensembl

genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res 46(D1):D802–D808
Kosugi S, Ohashi Y (1997) *PCF1* and *PCF2* specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. Plant Cell 9(9):1607–1619
Koyama T, Furutani M, Tasaka M, Ohme-Takagi M (2007) TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis. Plant Cell 19(2):473–484
Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874
Lei N, Yu X, Li S, Zeng C, Zou L, Liao W, Peng M (2017) Phylogeny and expression pattern analysis of TCP transcription factors in cassava seedlings exposed to cold and/or drought stress. Sci Rep 7(1):10016
Lescot M, Dhais P, Thijs G, Marchal K, Moreau Y, dePeer YV, Rouz P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30(1):325–327
Lewis JM, Mackintosh CA, Shin S, Gilding E, Kravchenko S, Baldridge G, Zeyen R, Muehlbauer GJ (2008) Overexpression of the maize *Teosinte Branched1* gene in wheat suppresses tiller development. Plant Cell Rep 27(7):1217–1225
Li S (2015) The *Arabidopsis thaliana* TCP transcription factors: a broadening horizon beyond development. Plant Signal Behav 10(7):e1044192
Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins Struct Funct Bioinf 50(3):437–450
Lukens L, Doebley JF (2001) Molecular evolution of the *teosinte branched1* gene among maize and related grasses. Mol Biol Evol 18(4):627–638
Ma J, Wang Q, Sun R, Xie F, Jones DC, Zhang B (2014) Genome-wide identification and expression analysis of TCP transcription factors in *Gossypium raimondii*. Sci Rep 4:6645
Maeshima K, Kaizu K, Tamura S, Nozaki T, Kokubo T, Takahashi K (2015) The physical size of transcription factors is key to transcriptional regulation in chromatin domains. J Phys Condens Matter 27(6):064116
Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45(D1):D200–D203
Minakuchi K, Kameoka H, Yasuno N, Umehara M, Luo L, Kobayashi K, Hanada A, Ueno K, Asami T, Yamaguchi S, Kyozuka J (2010) *FINE CULM1 (FC1)* works downstream of strigolactones to inhibit the outgrowth of axillary buds in rice. Plant Cell Physiol 51(7):1127–1135
Moturu TR, Thula S, Singh RK, Nodzyński T, Vařeková RS, Friml J, Simon S (2018) Molecular evolution and diversification of the SMXL gene family. J Exp Bot 69(9):2367–2378
Nandi S, Mehra N, Lynn AM, Bhattacharya A (2005) Comparison of theoretical proteomes: identification of COGs with conserved and variable pI within the multimodal pI distribution. BMC Genom 6(1):116
Nugent T, Jones DT (2009) Transmembrane protein topology prediction using support vector machines. BMC Bioinform 10(1):159
Osterberg JT, Xiang W, Olsen LI, Edenbrandt AK, Vedel SE, Christiansen A, Landes X, Andersen MM, Pagh P, Sandoe P, Nielsen J, Christiansen SB, Thorsen BJ, Kappel K, Gamberg C, Palmgren M (2017) Accelerating the domestication of new crops: feasibility and approaches. Trends Plant Sci 22(5):373–384
Peng Y, Chen L, Lu Y, Wu Y, Dumenil J, Zhu Z, Bevan MW, Li Y (2015) The ubiquitin receptors *DA1*, *DAR1*, and *DAR2* redundantly regulate endoreduplication by modulating the stability of TCP14/15 in Arabidopsis. Plant Cell 27(3):649–662
Portwood JL, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Sen TZ, Cho KT, Schott DA, Braun BL, Dietze M, Dunfee B, Elsik CG, Manchanda N, Coe E, Sachs M, Stinard P, Tolbert J, Zimmerman S, Andorf CM (2018) MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Res 47(D1):D1146–D1154
Prakash NR, Zunjare RU, Muthusamy V, Chand G, Kamboj MC, Bhat JS, Hossain F (2019) Genetic analysis of prolificacy in '*Sikkim Primitive*'—a prolific maize (*Zea mays* L.) landrace of North-Eastern Himalaya. Plant Breed 138:781–789. https://doi.org/10.1111/pbr.12736
Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL (2005) FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21(16):3435–3438
Rameau C, Bertheloot J, Leduc N, Andrieu B, Foucher F, Sakr S (2015) Multiple pathways regulate shoot branching. Front Plant Sci 5:741
Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WTB, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, Hayes PM, Lundqvist U, Franckowiak JD, Close TJ, Muehlbauer GJ, Waugh R (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED1*. Nat Genet 43(2):169
Reehana N, Ahamed AP, Ali DM, Suresh A, Kumar RA, Thajuddin N (2013) Structure based computational analysis and molecular phylogeny of C-Phycocyanin gene from the selected cynobacteria. Int J Agric Biol Eng 7:47–51
Remigereau MS, Lakis G, Rekima S, Leveugle M, Fontaine MC, Langin T, Sarr A, Robert T (2011) Cereal domestication and evolution of branching: evidence for soft selection in the *Tb1* orthologue of pearl millet (*Pennisetum glaucum* [L.] R. Br.). PLoS ONE 6(7):e22404
Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234(4774):364–368
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34(12):3299–3302
Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7(1):S10
Studer AJ, Doebley JF (2012) Evidence for a natural allelic series at the maize domestication locus *teosinte branched1*. Genetics 191(3):951–958
Studer A, Zhao Q, Ross-Ibarra J, Doebley JF (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet 43(11):1160
Studer AJ, Wang H, Doebley JF (2017) Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. Genetics 207(2):755–765
Takeda T, Suwa Y, Suzuki M, Kitano H, Ueguchi-Tanaka M, Ashikari M, Matsuoka M, Ueguchi C (2003) The *OsTB1* gene negatively regulates lateral branching in rice. Plant J 33(3):513–520
Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40(15):e115–e115
Vann L, Kono T, Pyhäjärvi T, Hufford MB, Ross-Ibarra J (2015) Natural variation in teosinte at the domestication locus *teosinte branched1* (*tb1*). Peer J 3:e900

Wang S, Li W, Liu S, Xu J (2016) RaptorX-property: a web server for protein structure property prediction. Nucleic Acids Res 44(W1):W430–W435

Wei W, Hu Y, Cui MY, Han YT, Gao K, Feng JY (2016) Identification and transcript analysis of the TCP transcription factors in the diploid woodland strawberry *Fragaria vesca*. Front Plant Sci 7:1937

Whipple CJ, Kebrom TH, Weber AL, Yang F, Hall D, Meeley R, Schimdt R, Doebley JF, Brutnell TP, Jackson DP (2011) *Grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. Proc Natl Acad Sci 108(33):E506–E512

Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16(1):18

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. Nat Methods 12(1):7

Yang CJ, Kursel LE, Studer AJ, Bartlett ME, Whipple CJ, Doebley JF (2016) A gene for genetic background in Zea mays: fine-mapping *enhancer of teosinte branched1. 2* to a YABBY class transcription factor. Genetics 204(4):1573–1585

Yao H, Skirpan A, Wardell B, Matthes MS, Best NB, McCubbin T, Durbak A, Smith T, Malcomber S, McSteen P (2019) The *barren stalk2* gene is required for axillary meristem development in maize. Mol Plant 12(3):374–389

Zhao J, Zhai Z, Li Y, Geng S, Song G, Guan J, Jia M, Wang F, Sun G, Feng N, Kong X, Chen L, Mao L, Li A (2018) Genome-wide identification and expression profiling of the TCP family genes in spike and grain development of wheat (*Triticum aestivum* L.). Front Plant Sci 9:1282

Zheng K, Ni Z, Qu Y, Cai Y, Yang Z, Sun G, Chen Q (2018) Genome-wide identification and expression analyses of TCP transcription factor genes in *Gossypium barbadense*. Sci Rep 8(1):14526

Zhou L, Zhang J, Yan J, Song R (2011) Two transposable element insertions are causative mutations for the major domestication gene *teosinte branched1* in modern maize. Cell Res 21(8):1267