



Genome-Wide Identification and Expression, Protein–Protein Interaction and Evolutionary Analysis of the Seed Plant-Specific *BIG GRAIN* and *BIG GRAIN LIKE* Gene Family

OPEN ACCESS

Edited by:

Elena M. Kramer,
Harvard University, United States

Reviewed by:

Michael R. McKain,
University of Alabama, United States
Shihua Cheng,
China National Rice Research Institute
(CAAS), China

***Correspondence:**

Ashverya Laxmi
ashverya_laxmi@nipgr.ac.in

†Present address:

Bhuwaneshwar S. Mishra,
Department of Plant Molecular
Biology, University of Delhi,
New Delhi, India

‡These authors have contributed
equally to this work as first authors.

§These authors have contributed
equally to this work as second
authors.

Specialty section:

This article was submitted to
Plant Evolution and Development,
a section of the journal
Frontiers in Plant Science

Received: 06 June 2017

Accepted: 05 October 2017

Published: 25 October 2017

Citation:

Mishra BS, Jamsheer K M, Singh D,
Sharma M and Laxmi A (2017)
Genome-Wide Identification
and Expression, Protein–Protein
Interaction and Evolutionary Analysis
of the Seed Plant-Specific *BIG GRAIN*
and *BIG GRAIN LIKE* Gene Family.
Front. Plant Sci. 8:1812.
doi: 10.3389/fpls.2017.01812

**Bhuwaneshwar S. Mishra^{†‡}, Muhammed Jamsheer K[‡], Dhriti Singh[§], Manvi Sharma[§]
and Ashverya Laxmi^{*}**

National Institute of Plant Genome Research, New Delhi, India

BIG GRAIN1 (*BG1*) is an auxin-regulated gene which functions in auxin pathway and positively regulates biomass, grain size and yield in rice. However, the evolutionary origin and divergence of these genes are still unknown. In this study, we found that *BG* genes are probably originated in seed plants. We also identified that seed plants evolved a class of *BIG GRAIN LIKE* (*BGL*) genes which share conserved middle and C-terminal motifs with *BG*. The *BG* genes were present in all monocot and eudicot species analyzed; however, the *BGL* genes were absent in few monocot lineages. Both *BG* and *BGL* were found to be serine-rich proteins; however, differences in expansion and rates of retention after whole genome duplication events were observed. Promoters of *BG* and *BGL* genes were found to be enriched with auxin-responsive elements and the *Arabidopsis thaliana* *BG* and *BGL* genes were found to be auxin-regulated. The auxin-induced expression of *AthBG2* was found to be dependent on the conserved ARF17/19 module. Protein-protein interaction analysis identified that *AthBG2* interact with regulators of splicing, transcription and chromatin remodeling. Taken together, this study provides interesting insights about *BG* and *BGL* genes and incentivizes future work in this gene family which has the potential to be used for crop manipulation.

Keywords: *BIG GRAIN*, *BIG GRAIN LIKE*, Phylogenetic analysis, gene family, auxin, ARF7/19

INTRODUCTION

In seed crops like rice, the grain size is an important agronomic trait which determines yield. Studies identified many important determinants of grain size in rice which has the potential to be used for crop manipulation (Zuo and Li, 2014). The *BIG GRAIN 1* (*OsaBG1*) has been recently identified as a positive regulator of grain size in rice. Interestingly, the constitutive expression of *OsaBG1* in *Arabidopsis thaliana* also resulted in an increase in seed size (Liu et al., 2015).

Apart from increase in seed size, enhanced expression of *BG1* in rice and *A. thaliana* resulted in an increase in biomass suggesting that this gene is a general promoter of growth (Liu et al., 2015). *OsaBG1* is a primary auxin response gene and increased expression resulted in enhanced

auxin sensitivity suggesting that this gene is involved in auxin signaling. *OsaBG1* is predominantly localized in the plasma membrane and overexpression lines showed enhanced auxin levels and basipetal transport suggesting that this gene might be involved in the regulation of auxin transport. These results suggest that *OsaBG1* is involved in enhancing biomass and grain size through regulating auxin pathway in rice (Liu et al., 2015). Different studies have already implicated auxin signaling in regulating biomass and seed size in both monocots and eudicots. The loss of *AUXIN RESPONSIVE FACTOR 2* of *A. thaliana* results in an increase in seed size due to enhanced cell proliferation while ectopic expression of *AUXIN RESPONSIVE FACTOR 19 (ARF19)* from *Jatropha curcas* promotes seed size and yield (Schruff et al., 2005; Sun et al., 2017). Decreased synthesis of auxin in *Pisum sativum* resulted in small and wrinkled seeds with reduced starch content (McAdam et al., 2017). In rice, the *NARROW LEAF1*, an inhibitor of polar auxin transport was found to be a suitable candidate gene for crop engineering for better yield (Qi et al., 2008; Fujita et al., 2013; Takai et al., 2013; Zhang et al., 2014).

In field trials, the *OsaBG1* overexpressing transgenic lines showed a significant increase in yield suggesting that *BG* genes are ideal candidates for crop improvement (Liu et al., 2015). The *BG* homologs were found to be present in eudicots. However, except the single study in rice, so far, no attempt has been made to study the *BG* genes in eudicots or any other monocot species. The enhanced biomass and bigger seed phenotype of *OsaBG1*-expressing *A. thaliana* lines suggest that the *BG*-dependent pathway of growth regulation might be conserved in all angiosperms (Liu et al., 2015). Many important seed crops such as legumes belong to eudicotyledonae and the potential of *BG* genes in crop improvement makes them ideal candidates to be studied in eudicots. Although *BG* genes were identified from selected monocots and eudicots, the origin and evolution of these genes in plants are yet to be identified.

In this study, we made use of the availability of a large number of sequenced genomes in the plant lineage to trace the evolutionary origin and evolution of *BG* genes. The comprehensive phylogenetic analysis suggest that *BG* genes are originated in seed plants. We also identified a group of *BIG GRAIN LIKE (BGL)* genes, which show sequence similarity with *BG* in some motifs and seems to be originated along with *BG* genes. Detailed phylogenetic analysis identified difference in the rate of retention and expansion among *BG* and *BGL* genes. Expression analysis of *A. thaliana* *BG* and *BGL* genes in different developmental stages was done and the spatiotemporal expression dynamics of these genes were identified. Further, the promoter analysis and auxin-dependent expression analysis of *A. thaliana* *BG* and *BGL* genes suggest the conservation of auxin-dependent transcriptional control of *BG* and *BGL* genes. Finally, we identified that *AthBG2*, a member of *A. thaliana* *BG* family; interacts with conserved regulators of splicing, transcription and chromatin remodeling. This comprehensive analysis of *BG* and *BGL* genes in plants can serve as a springboard for future work in this agronomically important class of genes.

MATERIALS AND METHODS

Identification of BG and BGL Proteins

The *BG* and *BGL* proteins from different sequenced plant genomes were identified by BLASTP, iterated BLAST and profile-HMM based searches in various databases. The remaining *BG* proteins from *Oryza sativa* were identified from MSU Rice Genome Annotation Project Database using *OsaBG1* as the query with default parameters (BLASTP, *e*-value: 1e-5) (Ouyang et al., 2007). The *BG* and *BGL* proteins from *Amborella trichopoda* were identified from Amborella Genome Database using *OsaBG1* as the query (BLASTP, *e*-value: 1e-5) (Albert et al., 2013). The *BG* and *BGL* proteins from the remaining species were identified using *AtrBG1* or *AtrBGL1* as the query in BLASTP. The proteins from *Picea abies* were identified from ConGenIE (BLASTP, *e*-value: 1e-3) (Sundell et al., 2015). The *BG* and *BGL* proteins from *Nelumbo nucifera* and *Phoenix dactylifera* were identified from NCBI RefSeq database (BLASTP, *e*-value: 10) (Pruitt et al., 2007). The proteins from *Lotus japonicus* were identified from Lotus Base (BLASTP, *e*-value: 1e-5) (Mun et al., 2016). The proteins from the remaining species were identified from Phytozome v12 (BLASTP, *e*-value: -1) (Goodstein et al., 2012). The retrieved sequences were manually curated to remove repeats and partial sequences. These sequences were aligned and used for profile-HMM based search in hmmsearch available in HMMER web server v2.15.0 with default parameters in the respective reference proteomes (significance *e*-value: 0.01 for sequence and 0.03 for hits). Further, three rounds of iterated BLAST search was performed using the aligned sequences in different reference proteomes using jackhmmer available in HMMER web server v2.15.0 with default parameters (significance *e*-value: 0.01 for sequence and 0.03 for hits) (Finn et al., 2015). Finally, to identify the true *BG/BGL* hits, the hits obtained from profile-HMM and iterative searches were subjected to BLAST search in *A. thaliana* and/or *O. sativa* proteomes. The final set of verified protein sequences used for phylogenetic analysis is given in Supplementary Table S1. The corresponding CDS sequences were retrieved from the respective databases using the protein identifiers. The final set of CDS sequences used for phylogenetic analysis is given in Supplementary Table S2.

Conserved Motif Analysis

Presence of characterized domains in *BG* and *BGL* proteins was analyzed by batch CD search tool against CDD v3.16 using the default parameters (*e*-value: 0.01) (Marchler-Bauer et al., 2015). The conserved motifs of *BG* and *BGL* proteins were identified from analyzing the aligned sequences of *BG* and *BGL* proteins from eudicots and monocots individually in Jalview (Waterhouse et al., 2009). The sequences of conserved motifs were exported from Jalview and used for HMM logo construction in Skylign (Wheeler et al., 2014). The fold recognition, topology and disorder prediction analysis was done using Phyre v2.0 with default parameters (Kelley et al., 2015).

Phylogenetic Analysis

The phylogenetic trees were reconstructed based on Bayesian (MrBayes v3) and maximum likelihood (PhyML) estimation methods in TOPALI v2.5 (Milne et al., 2009). The protein and CDS sequences were aligned by ClustalX v2.1 (Larkin et al., 2007). The aligned file was used to identify the best-suited model of amino acid and DNA substitution in TOPALI v2.5 (Milne et al., 2009). JTT + I + G model was used for reconstruction of all protein-based phylogenetic trees (Jones et al., 1992). GTR + I + G model was used for the reconstruction of all DNA-based phylogenetic trees (Tavaré, 1986). For Bayesian phylogeny, two Metropolis-Coupled Markov Chain Monte Carlo runs with 16 chains were performed with the given settings [sample frequency: 10; burn in (%): 25] for 100,000 generations unless specified. The protein-based phylogenetic tree of eudicots was run for 500,000 generations to attain convergence. The protein-based phylogenetic tree of eudicots was run for 300,000 generations to attain convergence. The CDS-based phylogenetic tree of eudicots and monocots was run for 1,000,000 generations to attain convergence. To verify the convergence of the Bayesian phylogenetic reconstruction, potential scale reduction factor was confirmed to reach 1.0 as runs converged in MrBayes v3 (Ronquist et al., 2012). The maximum likelihood trees were constructed with 1000 bootstrap replicates. The species phylogenetic tree was obtained from NCBI Taxonomy Common Tree tool. The phylogenetic trees were edited and visualized with midpoint rooting in FigTree v1.4.3 (Rambout, 2016). The alignment files used for phylogenetic analysis and tree output obtained are deposited in Figshare¹.

Ka/Ks Ratio Calculation

The PAL2NAL program was used for Ka, Ks and Ka/Ks calculations (Suyama et al., 2006). The Ka/Ks ratios of paralogous and orthologous genes identified from the phylogenetic reconstruction were calculated.

Expression and Promoter Analysis and Promoter: GUS Line Construction

The AtGenExpress expression data of *A. thaliana* BG and BGL genes in different developmental stages and auxin treatments were retrieved from Arabidopsis eFP browser (Schmid et al., 2005; Winter et al., 2007; Goda et al., 2008). The response of BG and BGL genes towards auxin treatments were calculated from this data. The values were plotted as heat maps using MultiExperiment Viewer (MeV, v4.8) (Saeed et al., 2006). For the qRT-PCR analysis of *AthBG2* in different developmental stages, the tissue and developmental stage-specific samples previously prepared in the lab were used (Jamsheer et al., 2015). The cDNA was prepared from 2 µg of total RNA using High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, United States). The 1:20 diluted cDNA samples were used for qRT-PCR reaction in 7500 Fast Real-Time PCR System using SYBR-Green chemistry as per the manufacturer's protocol (Applied Biosystems, United States). The qRT-PCR primers of

AthBG2 and the endogenous reference *UBQ10* were designed from the transcript region using Primer Express v3.0 (Applied Biosystems, United States). The primers used for qRT-PCR are given in Supplementary Table S4. The relative quantification was performed using $\Delta\Delta CT$ method (Livak and Schmittgen, 2001).

For the promoter analysis, 1kb upstream sequences from the TSS of *A. trichopoda*, *A. thaliana* and *O. sativa* BG and BGL genes were retrieved from Phytozome v12 (Goodstein et al., 2012) (Supplementary Table S3). The presence of different auxin responsive elements reported in the literature (TGTCNN, NNGACA, ACTTTA, CACGCAAT, TGACG, and KGTCCCAT) was analyzed and plotted according to the scale. For the expression analysis of *AthBG2* toward auxin treatment in *Col-0*, 5 days old seedling grown under standard growth conditions (16:8 h photoperiod, 60 µmol m⁻² s⁻¹ light and 22°C ± 2°C temperature) were subjected to 1 µM IAA treatment in 0.5X liquid MS medium for different time points. For gene expression comparison between WT and *arf7arf19* mutant, 7 days old seedlings grown under standard growth conditions were subjected to 1 µM IAA treatment in 0.5X liquid MS medium for different time points. The tissue was harvested at the end of time points in liquid nitrogen and stored in -80°C. RNA isolation and cDNA preparation from this tissue were performed as described previously (Mishra et al., 2009). The qRT-PCR experiment and calculation were performed as described above. The seeds of *arf7arf19* were obtained from ABRC (CS24625) (Okushima et al., 2005).

For the promoter: GUS line construction of *AthBG2*, 2 kb upstream region from the TSS was cloned in pENTR/D-TOPO using the primer given in Supplementary Table S4 and mobilized to pMDC164 (Curtis and Grossniklaus, 2003) vector using Gateway technology (Invitrogen, Carlsbad, CA, United States). The *pAthBG2::GUSA* construct was transformed in *Col-0* grown in standard growth conditions using floral dip method (Clough and Bent, 1998). The transformants were identified by screening the seeds in 0.5X MS medium with 15 µg/ml hygromycin A and homozygous lines were identified in the third generation. For GUS assay, the homozygous lines were grown under the standard growth conditions. The tissues were harvested and GUS assay was performed as described previously (Jefferson et al., 1987).

Yeast Two-Hybrid Assay

The *AthBG2* CDS cloned in pGBKT7 (BD vector) and transformed into Y2H gold strain was used as the bait to screen the normalized Mate & Plate Universal Arabidopsis Y2H cDNA library as per the manufacturer's protocol (Clontech, Mountain View, CA, United States). The primers used for cloning are given in Supplementary Table S4. Before the Y2H screening, the construct was tested for auto activation and toxicity respectively on SD/-Leu/X-α-Gal/AbA (SDO/X/A) and SD/-Leu (SDO) plates. After identifying that the construct lacked auto activation and toxicity properties, it was used for Y2H library screening. The colonies obtained from SD/-Leu/-Trp/X-α-Gal/AbA (DDO/X/A) after Y2H library screening were further screened on SD/-Ade/-His/-Leu/-Trp/X-α-Gal/AbA (QDO/X/A) medium. From the colonies survived on QDO/X/A medium, prey plasmids (pGADT7; AD vector) were isolated and

¹<https://figshare.com/s/54e4cf77da0709bbc1fb>

sequenced to identify the interacting protein. The interaction was further confirmed by co-transforming bait and recovered prey plasmids in Y2H gold strain and screening on QDO/X/A medium. A negative control experiment was conducted along with this experiment using bait vector and prey plasmids to identify the false interactors.

Bimolecular Fluorescent Complementation and Subcellular Localization

For the BiFC experiments, *AthBG2* and *Y14* CDS were cloned into pENTR/D-TOPO vector and mobilized respectively to pSAT4A-DEST-n(1-174)EYFP-N1 and pSAT5-DEST-c(175-END)EYFP-C1 (Tzfira et al., 2005) vectors using Gateway technology (Invitrogen, Carlsbad, CA, United States). Primers used for cloning are given in Supplementary Table S4. The clones were cotransformed in onion peel using PDS-1000 Helios Gene Gun (Biorad). The transformed onion peels were incubated at 22°C in darkness for 24 h and fluorescence was analyzed in TCS SP2 (AOBS) laser confocal scanning microscope after the incubation (Leica Microsystems). Negative control experiments were conducted along with this experiment.

For the subcellular localization of *AthBG2*, the CDS cloned into pENTR/D-TOPO vector was mobilized to pEG104 (Earley et al., 2006) vector using Gateway technology (Invitrogen, Carlsbad, CA, United States). The vector and YFP-CDS construct were transformed individually in onion peels and fluorescence was analyzed as described above.

RESULTS

Evolutionary Origin and Expansion of BG and BGL Genes

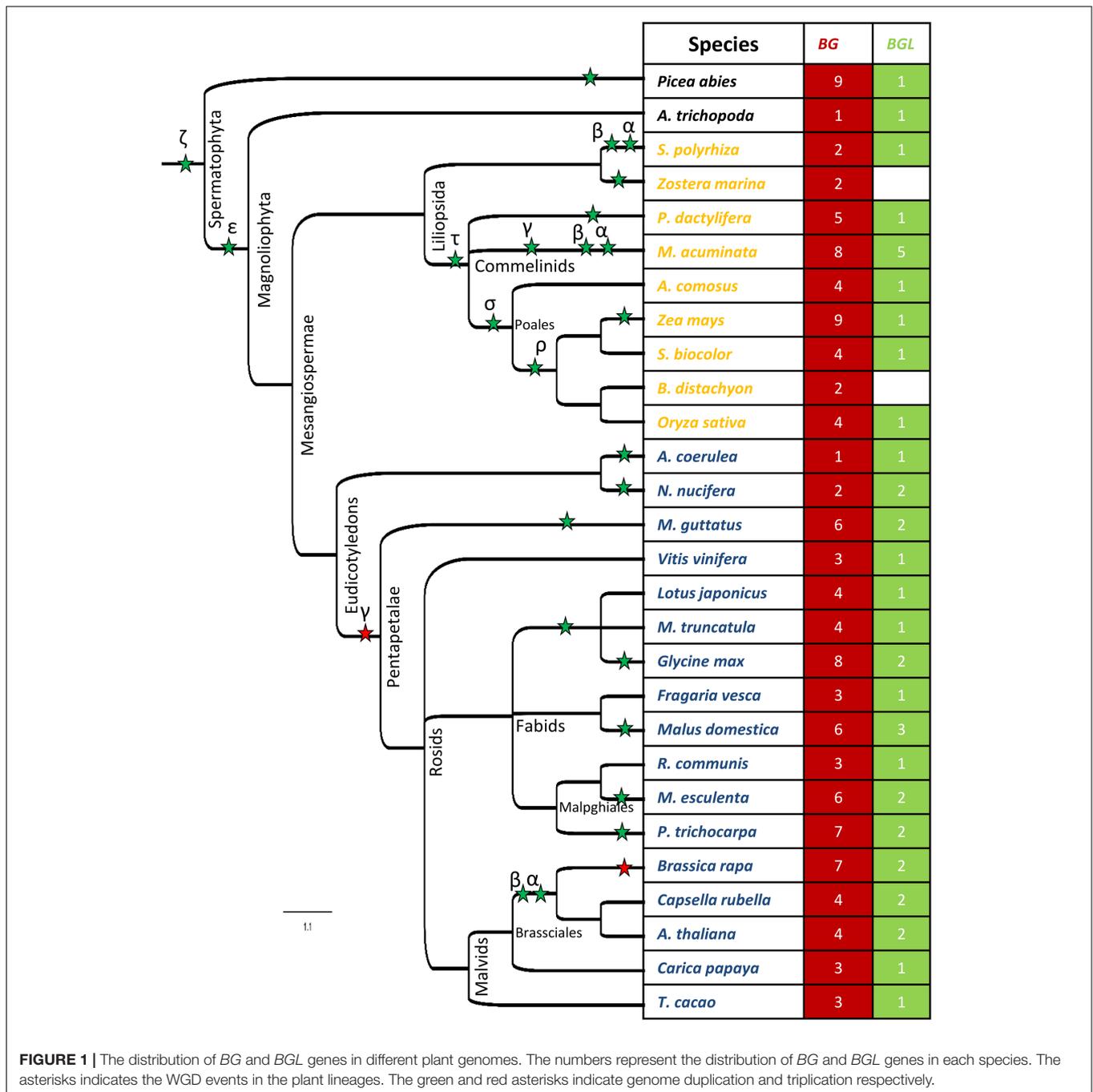
A combination of BLAST and HMM-based methods were employed to identify BG proteins from the selected sequenced viridiplantae members. Homology search in some of the sequenced chlorophyte genomes (*Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, *Volvox carteri*, and *Micromonas pusilla*) could not identify any homologs of *OsaBG1*. We also searched for the BG proteins in *Klebsormidium flaccidum*, which belongs to Charophyta and shows prototypes of signaling pathways required for terrestrial life (Hori et al., 2014). However, we couldn't retrieve any homologs of *OsaBG1* from this genome either. Further, iterative searches also failed to identify any hits of BG proteins in the algal genomes we analyzed, indicating that BG proteins are absent in algae.

We next analyzed the presence of BG1 homologs in embryophytes. The searches in Bryophyta (*Marchantia polymorpha*, *Physcomitrella patens* and *Sphagnum fallax*) and Pteridophyta (*Selaginella moellendorffii*) could not identify any hits of BG1. Interestingly, 9 hits of BG proteins were identified from the gymnosperm, *Picea abies* (Figure 1). Further, a hit of BG1 was identified from *A. trichopoda*, which is the only living species belonging to the sister lineage of angiosperms (Albert et al., 2013). In both *P. abies* and *A. trichopoda*, BLAST

search identified another protein which has similarity with the BG proteins in the middle and C-terminal regions (Figure 1, Supplementary Figure S1). We named these proteins as BIG GRAIN LIKE (BGL). Consistent with the sequence divergence observed among BG and BGL proteins, in the Bayesian phylogenetic interference of *A. trichopoda* and *P. abies* BG and BGL proteins, they recovered to distinct clades (Supplementary Figure S2). Relatively large number of BG proteins in *P. abies* indicates the possible contribution of whole genome duplications (WGD) in the expansion of BG genes in spermatophytes. In the plant lineage, there is evidence for an ancient WGD event (ζ) shared by seed plants and another WGD event (ϵ) common to angiosperms (Jiao et al., 2011). The absence of lineage-specific WGD in the *Amborella* lineage and the occurrence of Pinaceae-specific WGD event can be attributed to the increase in the number of BG genes in *P. abies* (Albert et al., 2013; Li Z. et al., 2015).

To dissect the evolutionary history of BG and BGL genes in angiosperms, we identified the BG and BGL genes from 17 eudicots and 9 monocots which represent important taxonomical positions (Figure 1). In eudicots, BG and BGL genes were found to be present in all species analyzed. Among the early diverging eudicots species analyzed, *Aquilegia coerulea* was found to have single BG and BGL genes each while their number was found to be increased to 2 in *Nelumbo nucifera*. Interestingly, both these genomes underwent lineage-specific WGD (Ming et al., 2013; Tiley et al., 2016). Further, the number of BG genes in majority of eudicot genomes without independent WGD (*Vitis vinifera*, *Fragaria vesca*, *Ricinus communis*, *Carica papaya* and *Theobroma cacao*) (Jaillon et al., 2007; Ming et al., 2008; Chan et al., 2010; Argout et al., 2011; Shulaev et al., 2011) events were found to be increased to 3 suggesting that the core eudicot genome triplication event contributed to this expansion (Jiao et al., 2012). However, in these species, the number of BGL genes remained 1 as in *A. trichopoda*. An expansion in the number of BG genes was observed in *Mimulus guttatus*, which can be attributed to the Lamiales-specific WGD event (Edger et al., 2016). Even in this species, only modest increase in the number of BGL genes was observed (Figure 1). These results indicate the difference in the rate of retention of BG and BGL genes after WGD events.

Legumes share an ancient papilionoid-specific WGD event that occurred ~ 54 Mya (Pfeil et al., 2005; Bertoli et al., 2009). Apart from this ancestral WGD, *Glycine max* underwent an independent WGD event ~ 13 Mya (Schmutz et al., 2010). Consistent with this, the legumes without independent WGD (*Lotus japonicus* and *Medicago truncatula*) (Sato et al., 2008; Young et al., 2011) possess 4 BG genes and 1 BGL gene each while *G. max* possesses 8 BG genes and 2 BGL genes (Figure 1). A similar increase in the number of BG and BGL genes was also observed in other species with independent WGD (*Malus domestica*, *Manihot esculenta*, *Populus trichocarpa* and *Brassica rapa*) (Tuskan et al., 2006; Velasco et al., 2010; Wang et al., 2011; Bredeson et al., 2016) events. The modest increase in the number of BG and BGL genes in *C. rubella* and *A. thaliana* could be due to the β and α duplications in crucifers (Bowers et al., 2003; Panchy et al., 2016).



In monocots, among the species analyzed, we could not identify BGL proteins from *Zostera marina* and *Brachypodium distachyon* even after repeated searches using various strategies (Figure 1). However, BGL proteins were found to be present in all other monocot species. The available evidence suggests that the monocot-specific τ duplication occurred after the divergence of Alismatales from other monocots (Jiao et al., 2014; Ming et al., 2015). The modest increase in the number of BG genes in species belonging to Alismatales (*Zostera marina* and *Spirodela polyrhiza*) can be attributed to the lineage-specific WGD event(s)

(Wang et al., 2014; Olsen et al., 2016). The σ WGD shared by Poales and ρ WGD common to grasses might have contributed to the expansion of BG genes in *Ananas comosus*, *Zea mays*, *Sorghum bicolor*, *Brachypodium Distachyon*, and *Oryza sativa* (Paterson et al., 2004; Tang et al., 2010; Ming et al., 2015). As observed in eudicots, in the monocots with independent WGD (*P. dactylifera*, *Musa acuminata* and *Z. mays*) (Schnable et al., 2009; D'Hont et al., 2012; Al-Mssallem et al., 2013) event(s), the number of BG genes are further increased. However, except in *M. acuminata*, not much increase in the number of BGL genes

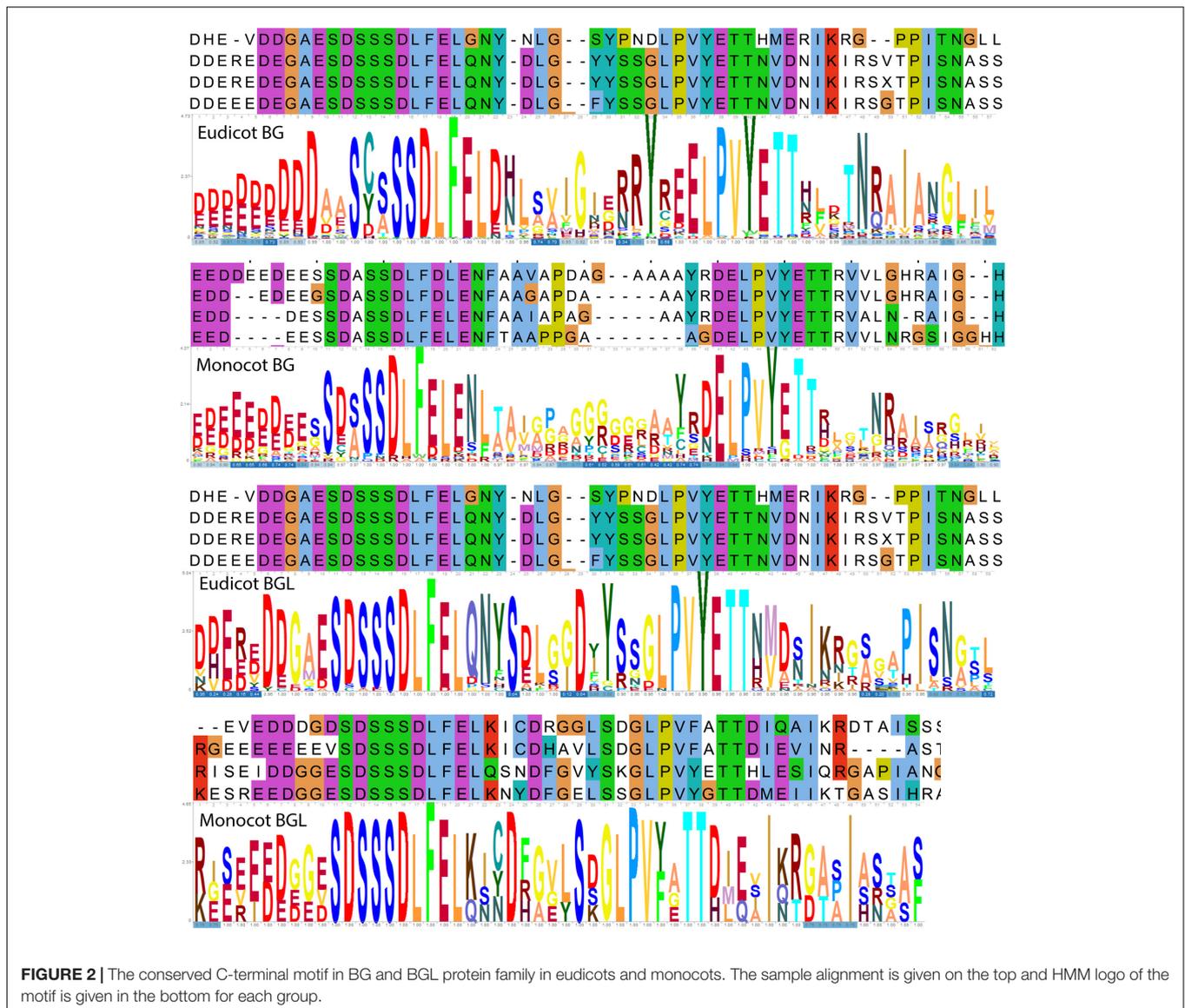


FIGURE 2 | The conserved C-terminal motif in BG and BGL protein family in eudicots and monocots. The sample alignment is given on the top and HMM logo of the motif is given in the bottom for each group.

was observed indicating the difference in the rate of retention in BG and BGL genes in monocots also.

Identification of the Conserved Motifs in BG and BGL

The preliminary analysis of representative BG proteins from eudicots and monocots suggest that this is a novel class of protein with no known protein domains (Liu et al., 2015). To find out whether the BG and BGL proteins acquired any known domains, we analyzed all 124 BG and 40 BGL proteins identified from different species for the presence of known protein domains. Almost in all species we analyzed, the size of BGL proteins was found to be smaller than BG (Supplementary Figure S3). Our analysis revealed that except two BG proteins from *Z. mays* and one protein from *S. bicolor*, BG and BGL proteins do not possess any known functional domains (Supplementary Figure S4). The middle region of SbiBG2 and ZmaBG3 show

partial similarity with the domain present in γ and τ subunits of DNA polymerase III (Jergic et al., 2007) (Supplementary Figures S4A,C,D). Similarly, the mid region of ZmaBG2 shows partial similarity with the domain present in the δ subunit of DNA polymerase III (Song et al., 2001) (Supplementary Figures S4B,D). Apart from these remote similarities, BG and BGL proteins did not show any significant similarities with any known protein domains suggesting that they belong to a novel class of proteins.

Since the domain search revealed that BG and BGL proteins do not possess any known domains, we analyzed the BG and BGL proteins from eudicots and monocots individually for the presence of conserved motifs shared across different species. As observed in the BLAST results, BG and BGL proteins in both eudicots and monocots show the presence of a conserved C-terminal region (Figure 2). This region was also found to be conserved among the *P. abies* BG and BGL proteins (Supplementary Figure S1). The N-terminal of this region was

found to be enriched with glutamic acid and aspartic acid repeats and showed a propensity for disorder in BG and BGL proteins we analyzed (Supplementary Figure S5). Further, many other amino acids were also found to be conserved in this C-terminal motif in the BG and BGL proteins in both eudicots and monocots (Figure 2). The topology prediction of the C-terminal motif of BG proteins from *A. trichopoda*, *O. sativa* and *A. thaliana* identified that this region has high propensity to form helices (Supplementary Figure S5A). However, analysis of fold similarity with known structures identified very low similarity suggesting that this C-terminal region is a novel domain (Supplementary Table S5). Topology prediction of the C-terminal motif of BGL proteins from *A. trichopoda* and *A. thaliana* predicted different topologies. The C-terminal motif of AtrBGL1 and AthBGL1 showed an N-terminal repeat of three small beta sheets followed by a C-terminal short helix while the shorter C-terminal motif of AthBGL2 predicted to have only the repeats of beta sheets (Supplementary Figure S5B). Analysis of fold similarity of the C-terminal motif of BGL2 also showed weak similarities with known structures (Supplementary Table S5).

Further, we identified four other conserved motifs present in BG proteins in both eudicots and monocots (Supplementary Figure S6). Similarly, conservation analysis identified two other conserved motifs in BGL proteins in both eudicots and monocots (Supplementary Figure S7). The motif III of BG proteins was found to have similarity with the N-terminus of the long motif II of BGL indicating that this is another region where the BG and BGL protein show similarity. The motif analysis identified that both BG and BGL proteins show many repeats of serine, glutamic acid, aspartic acid, arginine and lysine. As observed in the motif analysis, amino acid composition analysis of BG and BGL protein found the highest enrichment of serine followed by lysine, glutamine, arginine etc (Supplementary Figure S8).

Phylogenetic Analysis of BG and BGL Protein Family

In order to get more insight into the protein evolution, Bayesian phylogenetic reconstruction was used to decipher the evolutionary relationship of BG and BGL proteins in eudicots and monocots. The eudicot and monocot phylogenetic tree of BG and BGL protein was constructed individually along with the BG and BGL proteins from *A. trichopoda*. The eudicot BG and BGL proteins were recovered to separate clades in the phylogram (Figure 3). In the BGL clade, one protein each from *V. vinifera*, *C. rubella* and *A. thaliana* formed a clade with strong probability (posterior probability support value: 1) from other BGL proteins. We annotated them as BGL proteins because sequence alignment shows that it shares similarity with other BGL proteins in many regions. However, they also showed more sequence divergence compared to other BGL proteins (Supplementary Figure S9A). Further, in the individual phylogenetic tree with *A. trichopoda* BG and BGL proteins, these proteins were closely positioned with AtrBGL1 (Supplementary Figures S9B–D). The formation of separate clade from other BGL proteins can be due to the increased sequence divergence observed in these proteins and the consequent long-branch attraction (Kolaczowski et al.,

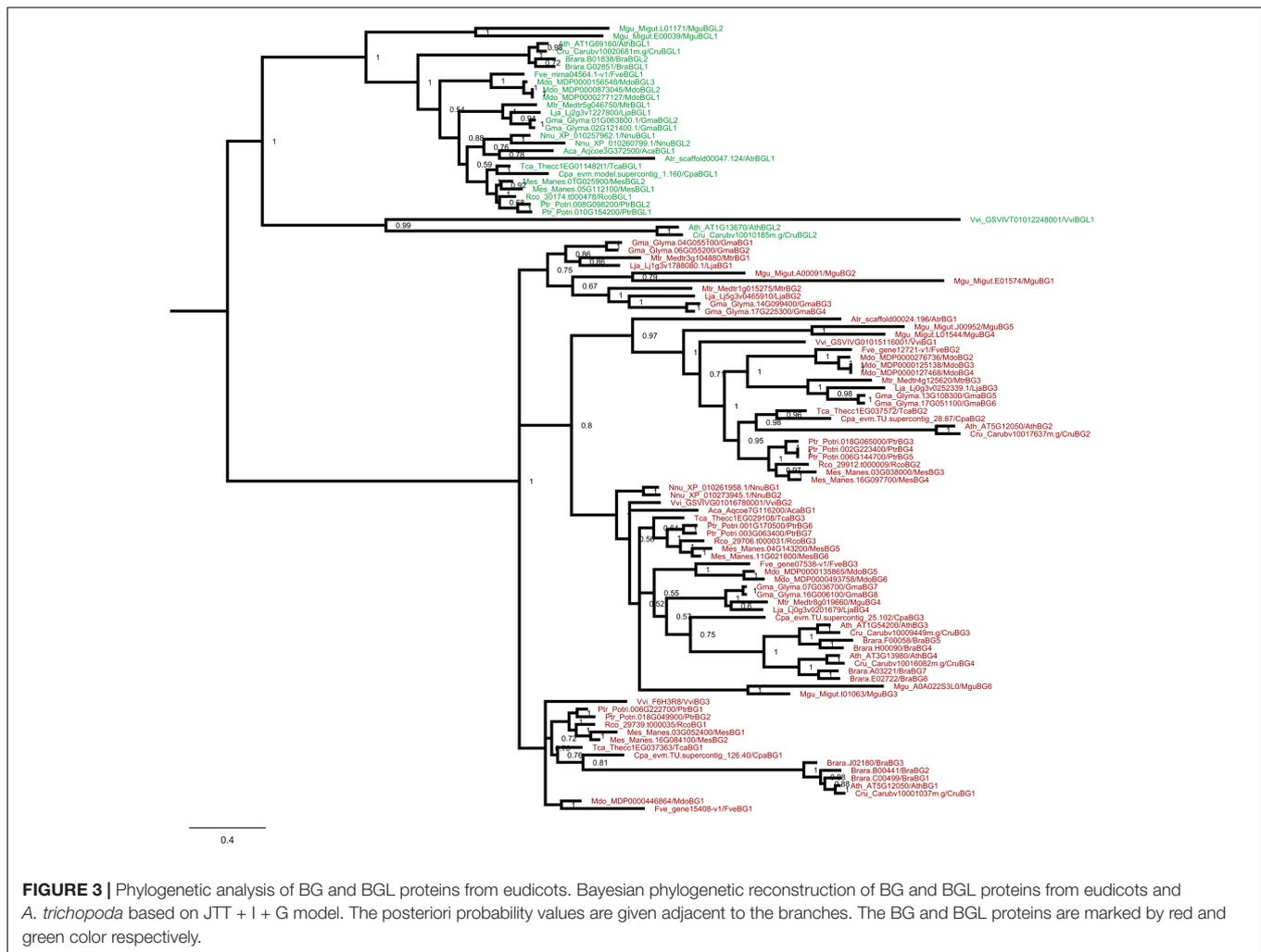
2009). Interestingly, in the BG proteins, we recovered four major sub-clades with strong probability (posterior probability support value: 0.8–1) suggesting sequence variation among the eudicot BG proteins. The BG proteins from eudicots without independent WGD were found to be positioned in most of these sub-clades suggesting the early origin of this sequence divergence. In the sub-clade recovered in our analysis, the positioning of BG proteins from different species fairly reflect the eudicot phylogeny (Zeng et al., 2017), suggesting the involvement of concerted evolution in protein divergence. The same pattern was also observed in the BGL sub-clades. The recovery of clades of BG and BGL proteins and different sub-clades in both groups is also observed in a maximum likelihood based estimation of the phylogenetic tree supporting the inferences obtained from the Bayesian phylogenetic reconstruction (Supplementary Figure S10). Further, we reconstructed the Bayesian phylogram of BG and BGL genes of eudicots using CDS sequences. As observed in the protein phylograms, the BG and BGL genes recovered to separate clades in the CDS-based phylogram (Supplementary Figure S11).

The recovery of BG and BGL proteins into separate clades was also observed in the Bayesian phylogram of monocots (Figure 4A). Further, BG proteins from monocots were also recovered to sub-clades with strong probabilities (posterior probability support value: 0.71–1). The maximum likelihood based estimation of the phylogenetic tree also supported these inferences (Supplementary Figure S12). However, in the Bayesian phylogenetic reconstruction using CDS sequences, the BGL genes from Poaceae were recovered to a clade closer to BG genes while other BGL genes formed another clade. Further, the solitary BG gene from *A. trichopoda* and one BG gene each from *Z. marina* and *S. polyrhiza* positioned a branch in between the two clades of BGL genes (Supplementary Figure S13).

Taken together, the phylogenetic analysis of BG and BGL proteins from eudicots and monocots indicates the possible role of concerted evolution in the protein divergence. Strikingly, the recovery of BG and BGL proteins into distinct clades in the phylogenetic analysis supporting the occurrence of distinct motifs specific to BG and BGL proteins. To further validate this observation, a phylogram was constructed with eudicot and monocot BG and BGL proteins (Figure 4B). Irrespective of the ancient monocot-eudicot divergence, the BG and BGL proteins recovered to separate clades supporting the hypothesis that the origin of both types of proteins predates the divergence of angiosperms.

Analysis of Selection Pressure on BG and BGL Genes

To identify the nature of evolutionary pressure acting on BG genes over time, the Ka/Ks ratio of 23 eudicot and 11 monocot putative paralogous gene pairs identified from the phylogenetic reconstruction was calculated. The Ka/Ks ratio of paralogous BG genes from eudicot was found to be ranging from 0.0108 to 0.7304 (Avg. 0.246) (Figure 5A). The Ka/Ks ratio of monocot BG genes was found to be ranging from 0.0601 to 0.3519 (Avg. 0.185) (Figure 5A). These results indicate strong purifying selection



acting on *BG* genes in both eudicots and monocots. Similarly, we calculated the K_a/K_s ratios of 6 eudicots and 1 monocot putative paralogous *BGL* gene pairs. The K_a/K_s ratios of *BGL* genes were found to be ranging from 0.238 to 0.434 (Avg. 0.326) in eudicots. The K_a/K_s ratio of *MacBGL1* and *MacBGL2* was estimated to be 0.435 (**Figure 5A**). These results suggest that both *BG* and *BGL* genes are under purifying selection in angiosperms.

To get better insight into the nature of selection in *BG* and *BGL* genes in closely related species, we calculated K_a/K_s ratio of putative orthologous genes inferred from the phylogenetic reconstruction (**Figure 5B**). The K_a/K_s ratio of 22 orthologous *BG* genes from eudicot was found to be ranging from 0.0917 to 0.3008 (Avg. 0.168). Similarly, the K_a/K_s ratio of 12 orthologous *BG* genes from monocots was found to be ranging from 0.0116 to 0.1851 (Avg. 0.116) confirming the purifying selection among *BG* genes. In 9 orthologous *BGL* eudicot genes analyzed, the K_a/K_s ratio varied from 0.0941 to 0.3856 (Avg. 0.177). Similarly, in the 3 orthologous *BGL* genes analyzed, the K_a/K_s ratio varied from 0.2065 to 0.4864 (Avg. 0.370) (**Figure 5B**). Taken together, these results confirm that both *BG* and *BGL* genes are under purifying selection in angiosperms.

Tissue and Developmental Stage-Specific Expression of *A. thaliana* *BG* and *BGL* Genes

For further analysis of *BG* and *BGL* gene family, we chose the model system *A. thaliana* which possess 4 *BG* and 2 *BGL* genes (**Figure 1**). The expression analysis of *A. thaliana* *BG* and *BGL* genes based on microarray data identified that these genes are expressed more in the seedling stages. Further, the expression of these genes was found to be more in the root, leaf, and flower and seed stages (**Figure 6A**). To validate the microarray expression data, we checked the expression of *AthBG2* in different tissues and developmental stages using qRT-PCR (**Figure 6B**). In microarray data, the expression of *AthBG2* was found to be high in the roots of different stages, leaf, flower etc. In qRT-PCR analysis, expression of *AthBG2* was found to be high in seedling of cotyledon and 2 leaves stage. The expression was found to be relatively high during seed germination, flower and root of the mature rosette. However, the expression was found to be relatively less in the rosette before and after bolting (**Figure 6B**). We further analyzed the promoter activity of *AthBG2* using

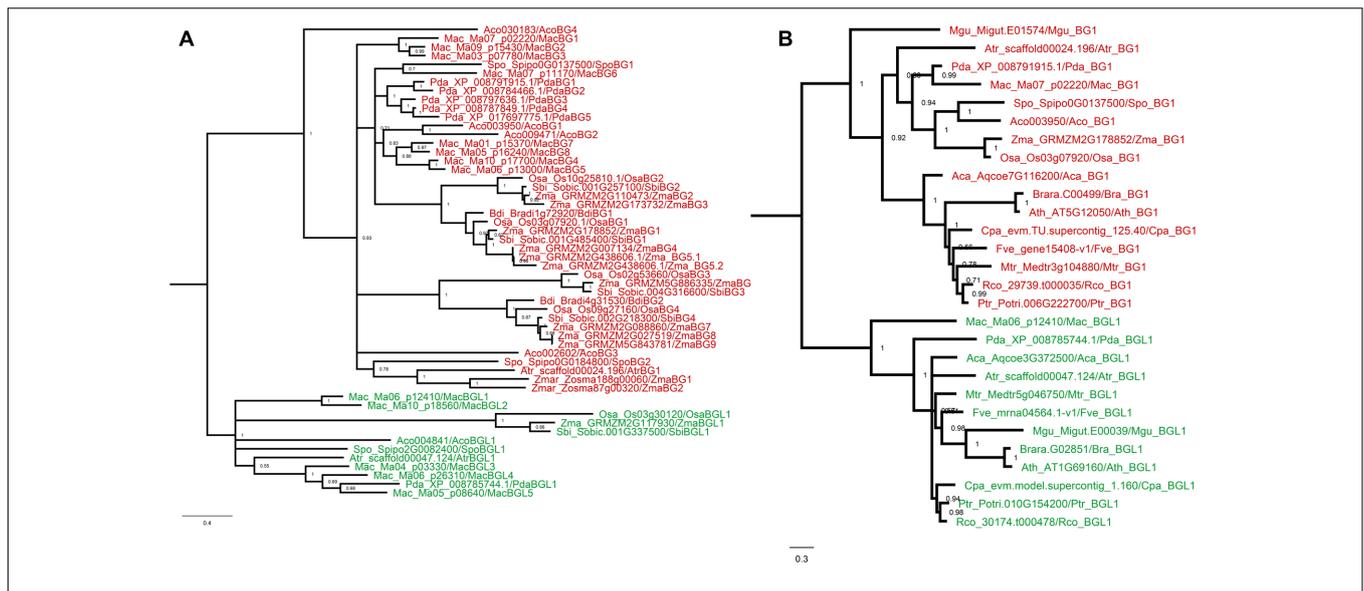


FIGURE 4 | Phylogenetic analysis of BG and BGL proteins from monocots and *A. trichopoda*. **(A)** Bayesian phylogenetic reconstruction of BG and BGL proteins from monocots and *A. trichopoda*. **(B)** Bayesian phylogenetic reconstruction of selected BG and BGL proteins from angiosperms. The phylograms were reconstructed on the basis of JTT + I + G model and the posterior probability values are given adjacent to the branches. The BG and BGL proteins are marked by red and green color respectively.

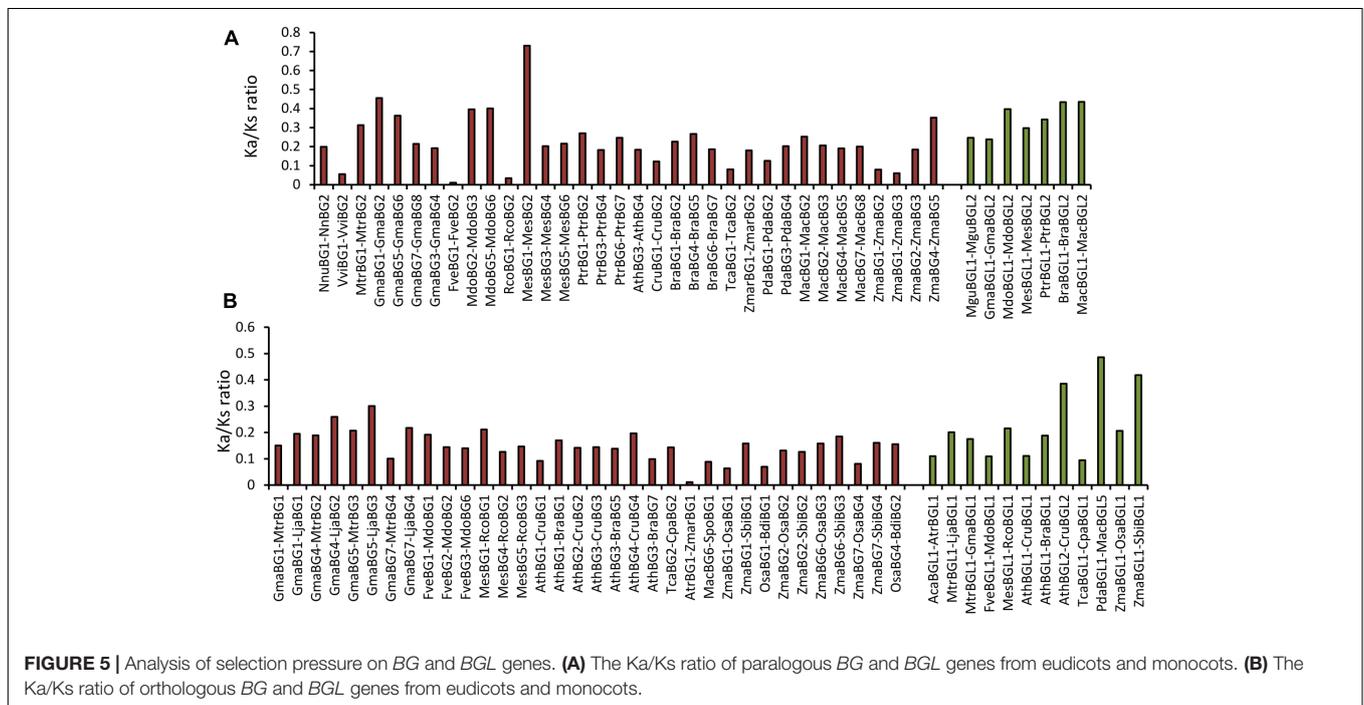
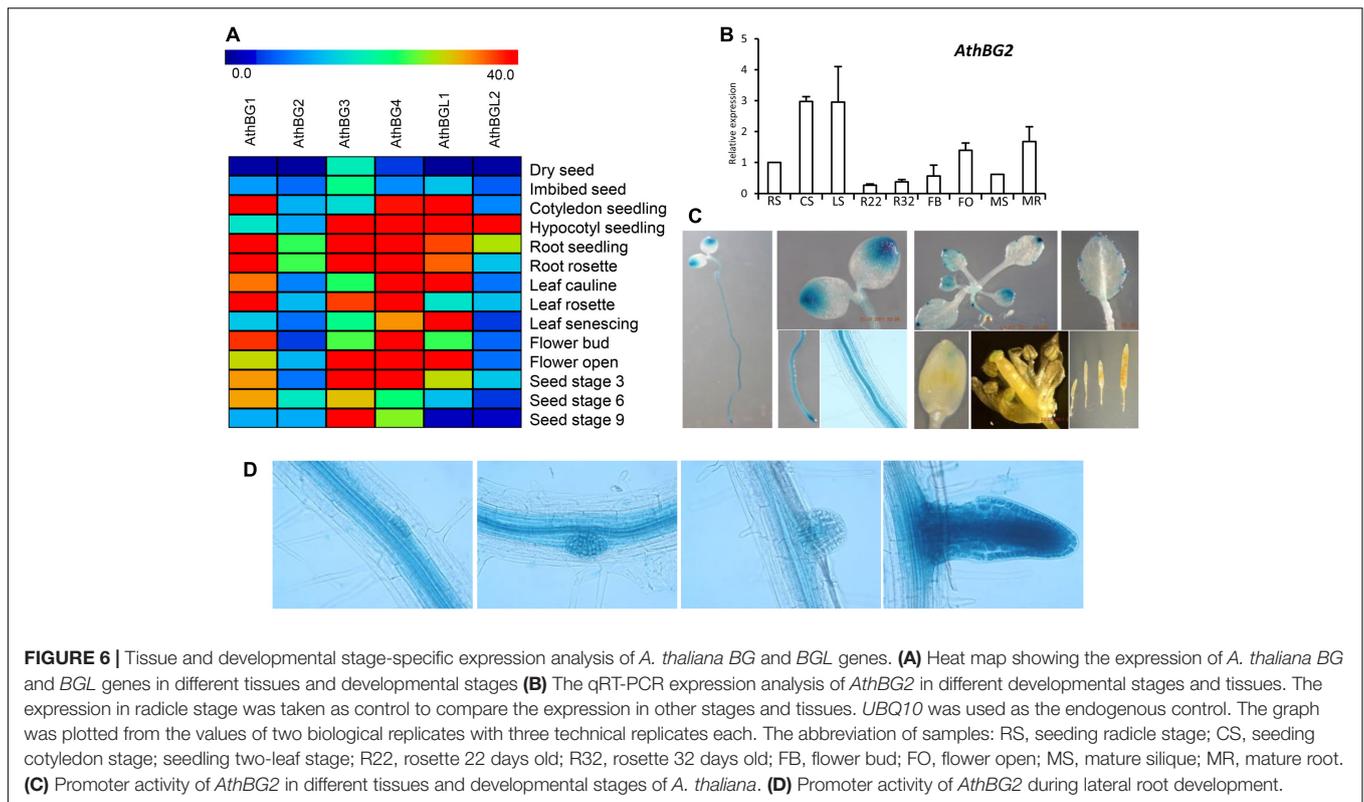


FIGURE 5 | Analysis of selection pressure on BG and BGL genes. **(A)** The Ka/Ks ratio of paralogous BG and BGL genes from eudicots and monocots. **(B)** The Ka/Ks ratio of orthologous BG and BGL genes from eudicots and monocots.

promoter: GUS assay (**Figure 6C**). In seedling stage, the *AthBG2* promoter was found to be active at cotyledon tip and stelar region of root and at the root tip. In mature leaves, the promoter activity was mainly restricted to serration tips. A small amount of promoter activity was observed in the stigma region. Further, promoter activity of *AthBG2* was observed during lateral root development in the young primordia and enhanced expression was observed in young lateral root (**Figure 6D**).

Taken together, the expression analysis of *A. thaliana* BG and BGL genes identified that these genes are expressed profusely in the seedling stages. Interestingly, the promoter: GUS assay of *AthBG2* identified that the promoter activity of this gene is overlapping with the activity of *DR5*: GUS auxin-responsive reporter, which show maximum activity in the regions of auxin accumulation such as root tip and cotyledon tip (Ulmasov et al., 1997). Further, the promoter activity of *AthBG2* overlaps

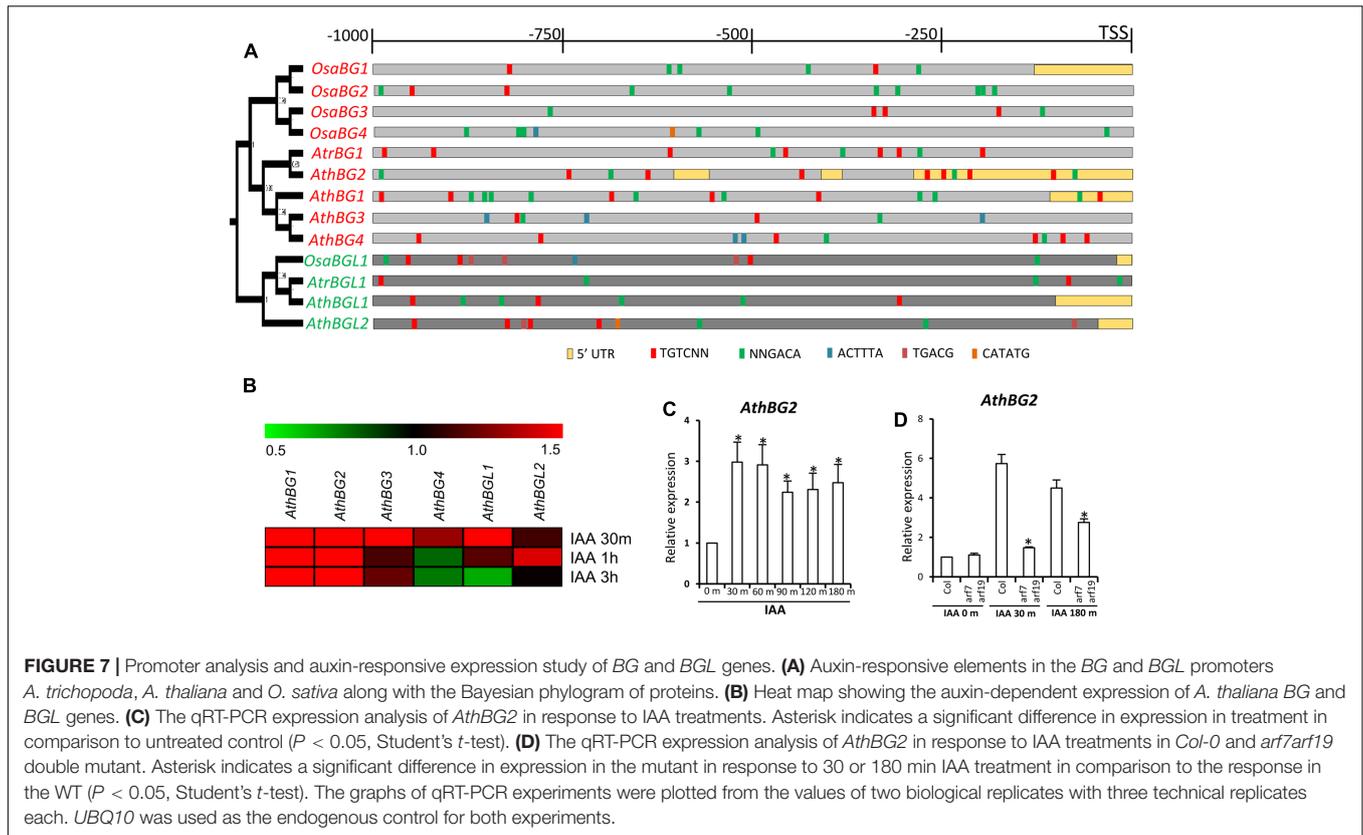


with the expression domain of genes such as *ARF7*, *ARF19*, *INDOLE-3-ACETIC ACID INDUCIBLE 2*, *LATERAL ORGAN BOUNDARIES-DOMAIN 16 (LBD16)*, *LBD29*, and *LBD33* which are involved in the regulation of auxin-dependent development (Okushima et al., 2005, 2007; Hay et al., 2006).

Auxin-Dependent Expression of *A. thaliana* BG and BGL Genes

The *OsaBG1* is a primary auxin-responsive gene (Liu et al., 2015). The promoter: GUS assay of *AthBG2* suggests that this gene might be involved in auxin signaling (Figures 6C,D). These results indicate that the auxin-dependent regulation of BG genes might be evolutionarily conserved. As the first step to test this hypothesis, we analyzed the presence of different auxin-responsive elements in the promoter and 5'UTR region of BG and BGL genes of *A. trichopoda*, *O. sativa* and *A. thaliana* (Figure 7A). The promoter and 5'UTR regions of both *AtrBG1* and *AtrBGL1* were found to be enriched with various auxin-responsive elements suggesting that the auxin-responsive expression of BG and BGL genes predates the angiosperm divergence. In rice, similar to *OsaBG1*, the promoter of other BG genes were also found to be enriched with auxin-responsive elements. The same pattern is also observed in the promoters of BG and BGL genes from *A. thaliana* (Figure 7A). As evident in the promoter analysis, the auxin-dependent expression analysis of *A. thaliana* BG and BGL genes using the available microarray data identified that these genes are responsive to auxin in various degrees (Figure 7B). The expression of *AthBG1* and *AthBG2* was found

to be induced in response to 30 min of auxin treatment and increased expression was retained in treatments of 1 and 3 h. The expression of *AthBG3* was induced in 30 min and in 1- and 3-h time points, expression was found to be almost similar to control levels. The expression of *AthBG4* was slightly induced in 30 min treatment and showed a small decrease in expression in the later time points. Among the BGL genes, *AthBGL1* expression was induced in response to 30 min treatment and repressed in response to 3 h treatment. In contrast, the expression of *AthBGL2* was induced in response to 1-h treatment (Figure 7B). These results indicate that the auxin-dependent regulation of expression of BG genes is conserved in monocots and eudicots. Further, the expression of BGL genes was also found to be responsive to auxin. In order to validate the microarray-based expression data, qRT-PCR expression analysis of *AthBG2* was done at various time points of auxin treatments (Figure 7C). As observed in the microarray data, the expression of *AthBG2* was found to be induced in 30 min and the increase in expression was found to be retained in all other time points. In rice, the increased expression of *OsaBG1* results in promotion of growth and increased grain size possibly through regulating auxin signaling pathway (Liu et al., 2015). In response to auxin, ARF7-ARF19 module redundantly activates cell cycle and positively regulates lateral and adventitious root formation and leaf cell expansion (Wilmoth et al., 2005; Okushima et al., 2007; Ito et al., 2016). Similar to *OsaBG1*, expression of *AthBG2* was found to be constitutively induced in response to auxin treatment at different time points (Figure 7C). Since *OsaBG1* is involved in the promotion of growth, we analyzed whether the auxin-dependent



expression of *AthBG2* is dependent on ARF7-ARF19 module using the *arf7 arf19* mutant (Okushima et al., 2005) (Figure 7D). In the qRT-PCR analysis, the expression of *AthBG2* was found to be similar in the mutant and WT suggesting that ARF7 and ARF19 are not involved in regulating the transcription of *AthBG2* in normal conditions. As observed in previous experiment, the expression of *AthBG2* was found to be induced in WT in response to auxin; however, this induction was found to be significantly abolished in the *arf7 arf19* mutant in both time points studied suggesting the role of ARF7-ARF19 module in regulating the auxin-dependent expression of *AthBG2* (Figure 7D).

Protein–Protein Interaction and Subcellular Localization Analysis of AthBG2

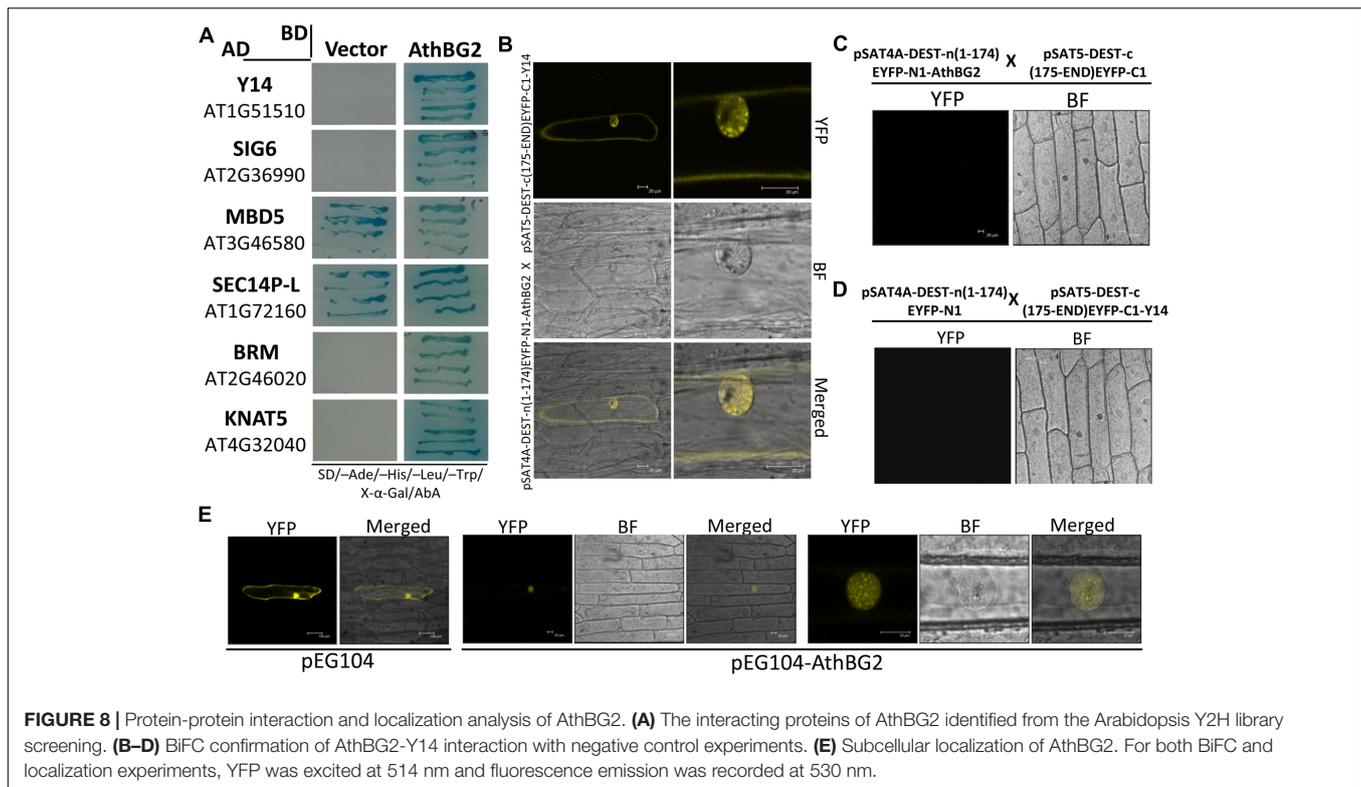
In order to get clues about the molecular function of AthBG2, we screened Arabidopsis Y2H library using AthBG2 as bait. We obtained 68 colonies after screening on QDO/X/A medium. Sequencing of these colonies identified 6 proteins in repeats (Figure 8A). Further screening to identify positive interaction confirmed 4 genuine interacting proteins of AthBG2. The 4 interacting proteins of AthBG2 were found to have conserved molecular functions. The Y14 is a conserved eukaryotic protein which is a part of exon junction complex (EJC) and involved in RNA splicing (Park and Muench, 2007; Koroleva et al., 2009; Mufarrije et al., 2011; Cilano et al., 2016). SIGMA FACTOR 6 (SIG6) is a chloroplast sigma factor of RNA polymerase involved

in chloroplast gene expression (Schweer et al., 2009; Chi et al., 2010). BRAHMA (BRM) is an SWI/SNF chromatin remodeling ATPase which is involved in the regulation of gene expression in response to many extrinsic and intrinsic cues (Han et al., 2015). KNOTTED1-LIKE HOMEBOX GENE 5 (KNAT5) is a homeobox transcription factor expressed majorly in the distal end of the epidermal cells in the elongation zone of primary root (Truernit et al., 2006).

In order to validate the Y2H result, we analyzed the interaction of AthBG2 with Y14 using BiFC assay. The AthBG2-Y14 interaction was found in both cytoplasm and nucleus confirming the Y2H result (Figure 8B). The removal of AthBG2 or Y14 abolished fluorescent signal confirming that this interaction is specific (Figures 8C,D). In the nucleus, the AthBG2-Y14 interaction was found to form speckle-like bodies. Y14 is reported to be localized predominantly in the nucleus (Park and Muench, 2007). Further, in the nucleus, Y14 was found in nucleolus and speckles (Koroleva et al., 2009). The subcellular localization analysis found that AthBG2 is predominantly localized in the nucleus, both in diffused and speckle-like form (Figure 8E).

DISCUSSION

The *BG* genes are recently identified as positive regulators of growth and grain size in rice (Liu et al., 2015). In this study, through homology-based searches in 39 viridiplantae genome, we suggest that *BG* genes are probably originated in seed



plants. *BG* genes are absent in *S. moellendorffii*, which belongs to Lycopodiophyta. Since there is no sequenced monophylyte genome available, at this point we cannot rule out whether the origin of *BG* genes predates the dawn of seed plants. Interestingly, a class of *BGL* genes were also found to be originated at the same time. The *BG* and *BGL* proteins share a conserved C-terminal motif and a middle motif. Apart from these shared motifs, *BG* and *BGL* proteins possess many unique conserved regions which resulted in their recovery to separate clades in the phylogenetic reconstruction. In the CDS-based phylogenetic reconstruction of *BG* and *BGL* genes from monocots, the recovery to specific *BG* and *BGL* clades was not that distinct, which could be due to the accumulation of synonymous mutations in the coding region. However, even in monocots, the phylogenetic reconstruction based on protein sequences using both Bayesian and maximum likelihood methods recovered distinct clades of *BG* and *BGL* proteins confirming the early divergence of *BG* and *BGL* proteins.

Unlike the animal and fungi genomes, plant genomes are characterized by many ancestral and recent WGD events. As a consequence of this, on an average, approximately 65% of genes in plant genomes have duplicated copies (Panchy et al., 2016). The genome sequencing of many plants identified specific gain or loss of many genes and gene families which ensure the survival of the species in their local environment (Tuskan et al., 2006; Jaillon et al., 2007; Chan et al., 2010; Young et al., 2011; Hori et al., 2014). The functional analysis in rice identified that *OsaBG1* is a general promoter of growth and seed size (Liu et al., 2015). Our study predicts that origin of these genes coincide with the origin of seed habit (Linkies et al., 2010). In our analysis, we

found that ancient and recent WGD increased the number of *BG* genes in most of the species. Interestingly, although *BGL* genes are also originated along with *BG* genes in seed plants before the divergence of angiosperms into eudicots and monocots, in eudicot species without independent WGD, their number is not increased suggesting the selective loss of *BGL* genes after the ancient eudicot genome triplication. Characteristically, in species with recent WGD event(s), the numbers of *BGL* genes have been increased suggesting the contribution of recent WGD events in the expansion of *BGL* genes in such species. Consistent with this, analysis of retention rate of duplicated genes after WGD events in *A. thaliana* identified that the rate of retention of duplicated genes is high in recent WGD events (Panchy et al., 2016).

Increased gene dosage provides a selective advantage for the organisms and therefore polyploids generally show hybrid vigor and increased resilience (del Pozo and Ramirez-Parra, 2015; Panchy et al., 2016). Since *BG1* gene in rice is a general growth promoter, the increased retention of these genes in comparison to *BGL* genes could be because of the selective advantage it can provide for the organisms. Along the same line, the reduced retention of *BGL* genes in angiosperms might have occurred because these genes might be acting as a competitive inhibitor of *BG* functions in a manner similar to microProteins (Eguen et al., 2015). This preliminary hypothesis can be tested in model species like *A. thaliana* which will provide interesting insights about the molecular connection between *BG* and *BGL* genes.

In our intra- and inter-species analysis, both *BG* and *BGL* genes were found to be under purifying selection. Consistent with this, although both these protein families do

not have any known domains, they possess many novel motifs which are conserved across the plant kingdom. The gene duplication relaxes the selection pressure which results in the concomitant neo-functionalization or sub-functionalization and hypo-functionalization or non-functionalization in one or both duplicated genes. Eventually, the survivors of this selection would undergo strong purifying selection. In *BG* genes, the paralogous gene pairs from species without recent WGD (*V. vinifera*, *T. cacao*, *R. communis*, and *F. vesca*) events were found to be under more strong purifying selection than the species with recent lineage-specific or independent WGD (*A. thaliana*, *C. rubella*, *B. rapa*, *M. esculenta*, *M. domestica*, *M. truncatula*, *G. max*, *Z. mays*, and *M. acuminata*). Further, difference in the Ka/Ks values observed among paralogous gene pairs from different species can be due to the difference in the timing of WGD event in particular lineage and slow nucleotide substitution rate observed in species such as *P. dactylifera* and *N. nucifera* (Wilson et al., 1990; Ming et al., 2013).

The diverse expression pattern of *A. thaliana* *BG* and *BGL* genes in different developmental stages indicates their general role in regulating plant growth. The *OsaBG1* is expressed abundantly in the vascular tissue of culm and young panicles (Liu et al., 2015). Our expression analysis identified that *A. thaliana* *BG* and *BGL* genes show high expression in seedling tissues and flowers suggesting common expression domains in eudicots and monocots. Further, the promoter: GUS assay identified that similar to *OsaBG1*, *AthBG2* is expressed profusely in the vascular tissues. Interestingly, the expression of *A. thaliana* *BG* and *BGL* genes was high in roots while the expression of *OsaBG1* was found relatively low in roots indicating differences in the expression domain in eudicots and monocots. The divergence in the promoter regions in monocots and eudicots might have contributed to this divergence in expression domain.

Among the phytohormones, auxin acts as the major regulator of plant development and *OsaBG1* is found to enhance growth through promoting auxin signaling (Ljung, 2013; Liu et al., 2015). Our expression study identified that *A. thaliana* *BG* and *BGL* genes are responsive to auxin in varying degrees suggesting that the auxin-dependent transcriptional regulation is a conserved mechanism by which the transcription of *BG* and *BGL* genes are controlled. Promoter analysis identified significant enrichment of auxin-responsive elements in the promoters of *AtrBG1* and *AtrBGL1* suggesting that the auxin-dependent transcriptional regulation predates the angiosperm divergence. The promoter: GUS assay of *AthBG2* identified that the promoter activity of this gene is high wherever localized auxin maxima is present. Further, the auxin-dependent expression of *AthBG2* was found to be mediated through the ARF7-ARF19 module. These results indicate auxin as a major regulator of transcription of *BG* and *BGL* genes. The root-specific expression and the role of ARF7-ARF19 in the regulation of auxin-dependent expression of *AthBG2* gene suggest the possible involvement of this gene in primary and lateral root development. Further, ARF19 from *J. curcas* is already implicated in the regulation of seed size and yield suggesting the interconnection between ARF19 pathway and seed development (Sun et al., 2017).

The localization analysis identified the predominant localization of *AthBG2* in the nucleus. *OsaBG1* is predominantly localized in the plasma membrane (Liu et al., 2015). These results suggest that *BG* genes from monocots and eudicots may have undergone divergence in the localization pattern which might be contributing to differences in the molecular functions. Protein-protein interaction analysis identified that *AthBG2* interact with nuclear proteins with conserved molecular functions. *AthBG2* interacts with Y14, which is the component of EJC, in the cytoplasm and nuclear speckles suggesting the role of *AthBG2* in mRNA splicing. Further, both *BG* and *BGL* proteins were found to be serine-rich with multiple conserved serine repeats. This is a characteristic feature of proteins involved in splicing (Richardson et al., 2011). The interaction with SIG6 and KNAT5 suggest the possible role of *AthBG2* in regulating transcription. The other interacting protein we identified from the Y2H library screening is BRM; which is an SWI/SNF chromatin remodeling ATPase reported to be essential for the maintenance of root stem cell niche through positively regulating auxin level and the transcription of specific *PIN* and *PLETHORA* genes (Yang et al., 2015). Further, BRM is involved in all aspects of plant growth from the embryonic to reproductive development and stress mitigation (Tang et al., 2008; Han et al., 2012; Li C. et al., 2015; Peirats-Llobet et al., 2016; Xu et al., 2016). The interaction of *AthBG2* with BRM suggests its possible involvement in auxin-regulated processes like embryonic, root, shoot and reproductive development (Ljung, 2013). This comprehensive analysis of *BG* and *BGL* proteins reveals insights into their origin and expansion in plants. Further, the expression, localization and protein-protein interaction analyses highlight the possible differences between eudicot and monocot *BG* proteins and encourage more functional analysis in this gene family.

AUTHOR CONTRIBUTIONS

AL, BM, and MJ conceived and designed the study. MJ performed phylogenetic and bioinformatic analysis. BM performed cloning, promoter: GUS analysis and BiFC. MJ and BM performed the Y2H analysis. BM, DS, and MS performed qRT-PCR analysis. MJ, DS, and MS performed the Ka/Ks estimation. AL, BM, and MJ analyzed the data. MJ wrote the manuscript. All authors reviewed the manuscript

ACKNOWLEDGMENTS

MJ acknowledges National Institute of Plant Genome Research for the Post-Doctoral Fellowship. DS and MS acknowledge the University Grant Commission, Government of India for the Senior Research Fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01812/full#supplementary-material>

REFERENCES

- Albert, V. A., Barbazuk, W. B., dePamphilis, C. W., Der, J. P., Leebens-Mack, J., Ma, H., et al. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* 342, 1241089–1241089. doi: 10.1126/science.1241089
- Al-Mssallem, I. S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., et al. (2013). Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.* 4, 520–524. doi: 10.1038/ncomms3274
- Argout, X., Salse, J., Aury, J.-M., Gultinan, M. J., Droc, G., Gouzy, J., et al. (2011). The genome of *Theobroma cacao*. *Nat. Genet.* 43, 101–108. doi: 10.1038/ng.736
- Bertioli, D. J., Moretzsohn, M. C., Madsen, L. H., Sandal, N., Leal-Bertioli, S. C., Guimaraes, P. M., et al. (2009). An analysis of synteny of *Arachis* with Lotus and Medicago sheds new light on the structure, stability and evolution of legume genomes. *BMC Genomics* 10:45. doi: 10.1186/1471-2164-10-45
- Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438. doi: 10.1038/nature01521
- Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., et al. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34, 562–570. doi: 10.1038/nbt.3535
- Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., et al. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nat. Biotechnol.* 28, 951–956. doi: 10.1038/nbt.1674
- Chi, W., Mao, J., Li, Q., Ji, D., Zou, M., Lu, C., et al. (2010). Interaction of the pentatricopeptide-repeat protein DELAYED GREENING 1 with sigma factor SIG6 in the regulation of chloroplast gene expression in Arabidopsis cotyledons. *Plant J.* 64, 14–25. doi: 10.1111/j.1365-313X.2010.04304.x
- Cilano, K., Mazanek, Z., Khan, M., Metcalfe, S., and Zhang, X.-N. (2016). A new mutation, hap1-2, reveals a C terminal domain function in AtMago protein and its biological effects in male gametophyte development in *Arabidopsis thaliana*. *PLOS ONE* 11:e0148200. doi: 10.1371/journal.pone.0148200
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743.
- Curtis, M. D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* 133, 462–469. doi: 10.1104/pp.103.027979
- del Pozo, J. C., and Ramirez-Parra, E. (2015). Whole genome duplications in plants: an overview from *Arabidopsis*. *J. Exp. Bot.* 66, 6991–7003. doi: 10.1093/jxb/erv432
- D'Hont, A., Denoeud, F., Aury, J.-M., Baurens, F.-C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241
- Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., et al. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* 45, 616–629. doi: 10.1111/j.1365-313X.2005.02617.x
- Edger, P. P., Smith, R. D., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., et al. (2016). *Subgenome Dominance in an Interspecific Hybrid, Synthetic Allopolyploid, and a 140 Year Old Naturally Established Neo-Allopolyploid Monkeyflower*. Available at: <http://www.biorxiv.org/content/early/2016/12/16/094797> [accessed August 9, 2017].
- Eguen, T., Straub, D., Graeff, M., and Wenkel, S. (2015). MicroProteins: small size – big impact. *Trends Plant Sci.* 20, 477–482. doi: 10.1016/j.tplants.2015.05.011
- Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER web server: 2015 update. *Nucleic Acids Res.* 43, W30–W38. doi: 10.1093/nar/gkv397
- Fujita, D., Trijatmiko, K. R., Tagle, A. G., Sapasap, M. V., Koide, Y., Sasaki, K., et al. (2013). NAL1 allele from a rice landrace greatly increases yield in modern indica cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20431–20436. doi: 10.1073/pnas.1310790110
- Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., et al. (2008). The AtGenExpress hormone- and chemical-treatment data set: experimental design, data evaluation, model data analysis, and data access. *Plant J.* 0, 80414150319983. doi: 10.1111/j.0960-7412.2008.03510.x
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozone: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Han, S.-K., Sang, Y., Rodrigues, A., Wagner, D., Wu, M.-F., Rodriguez, P. L., et al. (2012). The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in *Arabidopsis*. *Plant Cell* 24, 4892–4906. doi: 10.1105/tpc.112.105114
- Han, S.-K., Wu, M.-F., Cui, S., and Wagner, D. (2015). Roles and activities of chromatin remodeling ATPases in plants. *Plant J.* 83, 62–77. doi: 10.1111/tpj.12877
- Hay, A., Barkoulas, M., and Tsiantis, M. (2006). ASYMMETRIC LEAVES1 and auxin activities converge to repress BREVIPEDICELLUS expression and promote leaf development in *Arabidopsis*. *Development* 133, 3955–3961. doi: 10.1242/dev.02545
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5, 2914–2927. doi: 10.1038/ncomms4978
- Ito, J., Fukaki, H., Onoda, M., Li, L., Li, C., Tasaka, M., et al. (2016). Auxin-dependent compositional change in mediator in ARF7- and ARF19-mediated transcription. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6562–6567. doi: 10.1073/pnas.1600739113
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148
- Jamsheer, K. M., Mannully, C. T., Gopan, N., and Laxmi, A. (2015). Comprehensive evolutionary and expression analysis of FCS-like zinc finger gene family yields insights into their origin, expansion and divergence. *PLOS ONE* 10:e0134328. doi: 10.1371/journal.pone.0134328
- Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* 6, 3901–3907.
- Jergic, S., Ozawa, K., Williams, N. K., Su, X.-C., Scott, D. D., Hamdan, S. M., et al. (2007). The unstructured C-terminus of the subunit of *Escherichia coli* DNA polymerase III holoenzyme is the site of interaction with the subunit. *Nucleic Acids Res.* 35, 2813–2824. doi: 10.1093/nar/gkm079
- Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J. E., McKain, M. R., McNeal, J., et al. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* 13:R3. doi: 10.1186/gb-2012-13-1-r3
- Jiao, Y., Li, J., Tang, H., and Paterson, A. H. (2014). Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* 26, 2792–2802. doi: 10.1105/tpc.114.127597
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Pyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Kolaczowski, B., Thornton, J. W., Philippe, H., Raucourt, G., de, and Philippe, H. (2009). Long-branch attraction bias and inconsistency in bayesian phylogenetics. *PLOS ONE* 4:e7891. doi: 10.1371/journal.pone.0007891
- Koroleva, O. A., Calder, G., Pendle, A. F., Kim, S. H., Lewandowska, D., Simpson, C. G., et al. (2009). Dynamic behavior of *Arabidopsis* eIF4A-III, putative core protein of exon junction complex: fast relocation to nucleolus and splicing speckles under hypoxia. *Plant Cell* 21, 1592–1606. doi: 10.1105/tpc.108.060434
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Li, C., Chen, C., Gao, L., Yang, S., Nguyen, V., Shi, X., et al. (2015). The *Arabidopsis* SWI2/SNF2 chromatin remodeler BRAHMA regulates polycomb function during vegetative development and directly activates the flowering repressor gene SVP. *PLOS Genet.* 11:e1004944. doi: 10.1371/journal.pgen.1004944
- Li, Z., Baniaga, A. E., Sessa, E. B., Scascitelli, M., Graham, S. W., Rieseberg, L. H., et al. (2015). Early genome duplications in conifers and other seed plants. *Sci. Adv.* 10:e1501084. doi: 10.1126/sciadv.1501084
- Linkies, A., Graeber, K., Knight, C., and Leubner-Metzger, G. (2010). The evolution of seeds. *New Phytol.* 186, 817–831. doi: 10.1111/j.1469-8137.2010.03249.x

- Liu, L., Tong, H., Xiao, Y., Che, R., Xu, F., Hu, B., et al. (2015). Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11102–11107. doi: 10.1073/pnas.1512748112
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Ljung, K. (2013). *Auxin Metabolism and Homeostasis During Plant Development*. Available at: <http://dev.biologists.org/content/140/5/943> [accessed May 16, 2017].
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221
- McAdam, E. L., Meitzel, T., Quittenden, L. J., Davidson, S. E., Dalmais, M., Bendahmane, A. I., et al. (2017). Evidence that auxin is required for normal seed size and starch synthesis in pea. *New Phytol.* 216, 193–204. doi: 10.1111/nph.14690
- Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D. F., et al. (2009). TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25, 126–127. doi: 10.1093/bioinformatics/btn575
- Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., et al. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452, 991–996. doi: 10.1038/nature06856
- Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L.-T., et al. (2013). Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* 14:R41. doi: 10.1186/gb-2013-14-5-r41
- Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47, 1435–1442. doi: 10.1038/ng.3435
- Mishra, B. S., Singh, M., Aggrawal, P., and Laxmi, A. (2009). Glucose and auxin signaling interaction in controlling *Arabidopsis thaliana* seedlings root growth and development. *PLOS ONE* 4:e4502. doi: 10.1371/journal.pone.0004502
- Mufarrege, E. F., Gonzalez, D. H., and Curi, G. C. (2011). Functional interconnections of *Arabidopsis* exon junction complex proteins and genes at multiple steps of gene expression. *J. Exp. Bot.* 62, 5025–5036. doi: 10.1093/jxb/err202
- Mun, T., Bachmann, A., Gupta, V., Stougaard, J., and Andersen, S. U. (2016). Lotus base: an integrated information portal for the model legume *Lotus japonicus*. *Sci. Rep.* 6:39447. doi: 10.1038/srep39447
- Okushima, Y., Fukaki, H., Onoda, M., Theologis, A., and Tasaka, M. (2007). ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in *Arabidopsis*. *Plant Cell* 19, 118–130. doi: 10.1105/tpc.106.047761
- Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., et al. (2005). Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17, 444–463. doi: 10.1105/tpc.104.028316
- Olsen, J. L., Rouzé, P., Verhelst, B., Lin, Y.-C., Bayer, T., Collen, J., et al. (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* 530, 331–335. doi: 10.1038/nature16548
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2007). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 35, D883–D887. doi: 10.1093/nar/gkl976
- Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Park, N., and Muench, D. G. (2007). Biochemical and cellular characterization of the plant ortholog of PYM, a protein that interacts with the exon junction complex core proteins Mago and Y14. *Planta* 225, 625–639. doi: 10.1007/s00425-006-0385-y
- Paterson, A. H., Bowers, J. E., and Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9903–9908. doi: 10.1073/pnas.0307901101
- Peirats-Llobet, M., Han, S.-K., Gonzalez-Guzman, M., Jeong, C. W., Rodriguez, L., Belda-Palazon, B., et al. (2016). A direct link between abscisic acid sensing and the chromatin-remodeling ATPase BRAHMA via core ABA signaling pathway components. *Mol. Plant* 9, 136–147. doi: 10.1016/j.molp.2015.10.003
- Pfeil, B., Schlueter, J., Shoemaker, R., and Doyle, J. (2005). Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. *Syst. Biol.* 54, 441–454. doi: 10.1080/10635150590945359
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35, D61–D65. doi: 10.1093/nar/gkl842
- Qi, J., Qian, Q., Bu, Q., Li, S., Chen, Q., Sun, J., et al. (2008). Mutation of the rice narrow leaf1 gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiol.* 147, 1947–1959. doi: 10.1104/pp.108.118778
- Rambout, A. (2016). *FigTree*. Available at: <http://tree.bio.ed.ac.uk/software/figtree/> [accessed May 8, 2017].
- Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H., et al. (2011). Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLOS ONE* 6:e24542. doi: 10.1371/journal.pone.0024542
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., et al. (2006). TM4 microarray software suite. *Methods Enzymol.* 411, 134–193. doi: 10.1016/S0076-6879(06)11009-5
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15, 227–239. doi: 10.1093/dnares/dsn008
- Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501–506. doi: 10.1038/ng1543
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534
- Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2005). The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133, 251–261. doi: 10.1242/dev.02194
- Schweer, J., Geimer, S., Meurer, J., and Link, G. (2009). *Arabidopsis* mutants carrying chimeric sigma factor genes reveal regulatory determinants for plastid gene expression. *Plant Cell Physiol.* 50, 1382–1386. doi: 10.1093/pcp/pcp069
- Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43, 109–116. doi: 10.1038/ng.740
- Song, M.-S., Pham, P. T., Olson, M., Carter, J. R., Franden, M. A., Schaeper, R. M., et al. (2001). The delta and delta subunits of the DNA polymerase III holoenzyme are essential for initiation complex formation and processive elongation. *J. Biol. Chem.* 276, 35165–35175. doi: 10.1074/jbc.M100389200
- Sun, Y., Wang, C., Wang, N., Jiang, X., Mao, H., Zhu, C., et al. (2017). Manipulation of auxin response factor 19 affects seed size in the woody perennial *Jatropha curcas*. *Sci. Rep.* 7:40844. doi: 10.1038/srep40844
- Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y.-C., Sjödin, A., et al. (2015). The plant genome integrative explorer resource: PlantGenIE.org. *New Phytol.* 208, 1149–1156. doi: 10.1111/nph.13557
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Takai, T., Adachi, S., Taguchi-Shiobara, F., Sanoh-Arai, Y., Iwasawa, N., Yoshinaga, S., et al. (2013). A natural variant of NAL1, selected in high-yield rice breeding programs, pleiotropically increases photosynthesis rate. *Sci. Rep.* 3, 2149. doi: 10.1038/srep02149
- Tang, H., Bowers, J. E., Wang, X., and Paterson, A. H. (2010). Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl. Acad. Sci. U.S.A.* 107, 472–477. doi: 10.1073/pnas.0908007107
- Tang, X., Hou, A., Babu, M., Nguyen, V., Hurtado, L., Lu, Q., et al. (2008). The *Arabidopsis* BRAHMA chromatin-remodeling ATPase is involved in repression

- of seed maturation genes in leaves. *Plant Physiol.* 147, 1143–1157. doi: 10.1104/pp.108.121996
- Tavaré, S. (1986). *Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences*. Available at: <http://www.citeulike.org/user/aprasad/article/4801403> [accessed August 17, 2017].
- Tiley, G. P., Ané, C., and Burleigh, J. G. (2016). Evaluating and characterizing ancient whole-genome duplications in plants with gene count data. *Genome Biol. Evol.* 8, 1023–1037. doi: 10.1093/gbe/evw058
- Truernit, E., Siemerling, K. R., Hodge, S., Grbic, V., and Haseloff, J. (2006). A map of KNAT gene expression in the Arabidopsis root. *Plant Mol. Biol.* 60, 1–20. doi: 10.1007/s11103-005-1673-9
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691
- Tzfira, T., Tian, G.-W., Lacroix, B., Vyas, S., Li, J., Leitner-Dagan, Y., et al. (2005). pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Mol. Biol.* 57, 503–516. doi: 10.1007/s11103-005-0340-5
- Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* 9, 1963–1971. doi: 10.1105/tpc.9.11.1963
- Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., et al. (2010). The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat. Genet.* 42, 833–839. doi: 10.1038/ng.654
- Wang, W., Haberer, G., Gundlach, H., Gläßer, C., Nussbaumer, T., Luo, M. C., et al. (2014). The *Spirodela polyrhiza* genome reveals insights into its neotenus reduction fast growth and aquatic lifestyle. *Nat. Commun.* 5:3311. doi: 10.1038/ncomms4311
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wheeler, T. J., Clements, J., and Finn, R. D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* 15:7. doi: 10.1186/1471-2105-15-7
- Wilmoth, J. C., Wang, S., Tiwari, S. B., Joshi, A. D., Hagen, G., Guilfoyle, T. J., et al. (2005). NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. *Plant J.* 43, 118–130. doi: 10.1111/j.1365-313X.2005.02432.x
- Wilson, M. A., Gaut, B., and Clegg, M. T. (1990). Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Mol. Biol. Evol.* 7, 303–314.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N. J. (2007). An “electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* 2:e718. doi: 10.1371/journal.pone.0000718
- Xu, Y., Guo, C., Zhou, B., Li, C., Wang, H., Zheng, B., et al. (2016). Regulation of vegetative phase change by SWI2/SNF2 chromatin remodeling ATPase BRAHMA. *Plant Physiol.* 172, 2416–2428. doi: 10.1104/pp.16.01588
- Yang, S., Li, C., Zhao, L., Gao, S., Lu, J., Zhao, M., et al. (2015). The Arabidopsis SWI2/SNF2 chromatin remodeling ATPase BRAHMA targets directly to PINs and Is required for root stem cell niche maintenance. *Plant Cell* 27, 1670–1680. doi: 10.1105/tpc.15.00091
- Young, N. D., DeBellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625
- Zeng, L., Zhang, N., Zhang, Q., Endress, P. K., Huang, J., and Ma, H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytol.* 214, 1338–1354. doi: 10.1111/nph.14503
- Zhang, G.-H., Li, S.-Y., Wang, L., Ye, W.-J., Zeng, D.-L., Rao, Y.-C., et al. (2014). LSCHL4 from japonica cultivar, which is allelic to NAL1, increases yield of indica super rice 93-11. *Mol. Plant* 7, 1350–1364. doi: 10.1093/mp/ssu055
- Zuo, J., and Li, J. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* 48, 99–118. doi: 10.1146/annurev-genet-120213-092138

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mishra, Jamsheer K, Singh, Sharma and Laxmi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.