TÜBİTAK

Turkish Journal of Biology

http://journals.tubitak.gov.tr/biology/

Research Article

Turk J Biol
(2014) 38: 296-306
© TÜBİTAK
doi:10.3906/biy-1308-57

Genome-wide identification and analysis of growth regulating factor genes in *Brachypodium distachyon*: in silico approaches

Ertuğrul FİLİZ[1,*], İbrahim KOÇ[2], Hüseyin TOMBULOĞLU[3]

[1]Department of Crop and Animal Production, Çilimli Vocational School, Düzce University, Çilimli, Düzce, Turkey
[2]Department of Molecular Biology and Genetics, Faculty of Science, Gebze Institute of Technology, Gebze, Kocaeli, Turkey
[3]Department of Biology, Faculty of Science and Arts, Fatih University, Büyükçekmece, İstanbul, Turkey

Received: 27.08.2013 • **Accepted:** 23.12.2013 • **Published Online:** 28.03.2014 • **Printed:** 28.04.2014

Abstract: Growth-regulating factor (*GRF*) genes may play important roles for regulating growth and development in different plant tissues and organs. Here we report the first genome-wide analysis of the *GRF* gene family in *Brachypodium*. We performed in silico comparative analysis of *GRF* genes, including their structure, duplication in the genome, conserved motifs, and phylogenetic relationship. At the end of the study, 10 Bd*GRF* genes were identified. The highest number of *GRF* genes was identified on chromosome 1 with 5 members, whereas the least number of genes (only 1 member) was found on chromosomes 2, 4, and 5. Of those, a single segmental duplication was observed in the *Brachypodium* genome. Average exon and intron numbers were determined as 3 and 4, respectively. Motif analysis showed that WRC and QLQ residues were consistent in all GRF protein sequences. Gene Ontology terms showed that 10 BdGRF proteins grouped in the same biological function, biological process, and cellular component groups. In addition, we compared the new BdGRF proteins with the other monocot and dicot GRF proteins sequences. Phylogenetic analysis revealed that GRF proteins of monocot and dicot species were clustered together in a joined tree; in particular, the monocot species (*Brachypodium*, maize, and rice) were grouped into the same cluster with high bootstrap values. We assume that the results of this study will provide molecular insights about GRF proteins in grass species.

Key words: *Brachypodium distachyon*, growth regulating factor, *GRF*, genome-wide analysis

1. Introduction

Transcriptional control of biological processes including development, differentiation, growth, and metabolism is related to specific cis-regulatory regions of genes. Additionally, transcription factor activities affect gene expression level (Zhang et al., 2008). In *Arabidopsis thaliana*, 1500 possible specific transcription factors were detected and approximately 45% of these are accepted as plant-specific transcription factors (Riechmann et al., 2000). These transcription factors were classified based on their DNA-binding domains (Yamasaki et al., 2008). Growth-regulating factor (*GRF*) genes are plant-specific transcription factors that are distributed in all genomes of seed plants (Kim et al., 2003). These genes may regulate growth and development of leaves and cotyledons (Kim and Kende, 2004). In general, GRF family proteins contain 2 conserved regions: the QLQ (Gln, Leu, Gln) and WRC (Trp, Arg, Cys) domains (van der Knaap et al., 2000; Kim et al., 2003; Zhang et al., 2008). The QLQ domain is similar to the N-terminal part of the yeast SWI2/SNF2 protein, which is located with the SWI/SNF chromatin-remodeling complex in yeast (Treich et al., 1995), that may play a role in protein–protein interactions (Kim et al., 2003; Choi et al., 2004). The WRC domain contains a functional nuclear localization signal and putative zinc finger motifs with 1 His and 3 Cys residues (van der Knaap et al., 2000). Recently, GRF-interacting factor (GIF) family proteins that interact with the QLQ domain of GRF proteins in *Arabidopsis* have been identified (Kim and Kende, 2004). *GIF* genes connect with some mice CREST-related transcription coactivators including calcium signaling mechanisms (Aizawa et al., 2004) and proteins of the GIF family have a conserved domain named SNH or SSXT (Kim and Kende, 2004).

GRF genes were found to comprise 9 and 12 members in *Arabidopsis* and rice, respectively (Kim et al., 2003; Choi et al., 2004). The GRF family proteins of *Arabidopsis* (AtGRF) and rice (OsGRF) contain the same characteristic regions of the QLQ (Gln, Leu, Gln) and WRC (Trp, Arg, Cys) domains. Many At*GRF* genes are expressed in growing and developing tissues, including

* Correspondence: ertugrulfiliz@gmail.com

shoot tips, flower buds, and roots. At*GRF1* through At*GRF6* genes were strongly expressed in roots, upper stems, and shoot tips; on the contrary, these genes were expressed at low levels in mature stems and leaves. In addition, At*GRF7* and At*GRF8* were mostly expressed in shoot tips and flowers. Overexpression of At*GRF1* and At*GRF2* correlated with larger leaves and cotyledons. On the other hand, triple insertional null mutants of At*GRF1*–At*GRF3* involved smaller leaves and cotyledons. These results suggest that AtGRF proteins affect the regulation of plant development and growth (Kim et al., 2003). Kim and Kende (2004) showed that At*GRF1* and At*GIF1* may act as a transcription activator and coactivator, respectively, and may be components of regulating the growth and shape of leaves and petals. The At*GRF5* gene was shown to regulate cell proliferation, especially during the development of leaf size and shape (Horiguchi et al., 2005). Recently, the *Arabidopsis* GRF7 protein was demonstrated to interact directly with the dehydration-responsive element-binding protein 2A (*DREB2A*) promoter and repress *DREB2A* activity (Kim et al., 2012).

In rice, 11 homologs of Os*GRF1* were identified and characterized. Totally, 12 OsGRF proteins contain 2 conserved regions: the QLQ (Gln, Leu, and Gln) and WRC (Trp, Arg, and Cys) domains. Studies showed that Os*GRF* genes were expressed especially in growing tissues. Gibberellic acid applications improved the expression of 7 Os*GRF* genes (Os*GRF1*, 2, 3, 7, 8, 10, and 12). Most Os*GRF* genes were expressed at the highest level in nodes and rapidly growing primary leaves, while Os*GRF* genes were expressed at very low levels in root tissues. Based on in situ localization analysis, Os*GRF1* mRNA was observed in the epidermis, vascular bundles of the intercalary meristem of the internode, and adventitious roots of the second highest node (Choi et al., 2004). In addition, 14 homologs of Zm*GRF* genes and 3 homologs of Zm*GIF* genes were identified and characterized in maize (*Zea mays* L.). In particular, overexpression of both Zm*GRF11*–Zm*GIF2* and Zm*GRF2*–Zm*GIF3* genes speeds the growth of the inflorescence stem when compared to wild-type *A. thaliana*, and these genes were suggested to be responsible for growth and development in maize (Zhang et al., 2008). Zm*GRF2*, Zm*GRF5*, Zm*GRF9*, and Zm*GRF13* were expressed at higher levels in immature leaves than old leaves, whereas Zm*GRF3*, Zm*GRF5*, Zm*GRF6*, Zm*GRF7*, Zm*GRF9*, Zm*GRF11*, and Zm*GRF13* were expressed notably in immature ears. Zm*GRF11* and Zm*GRF2* were also highly expressed in ears and shoots. Thus, the *GRF* and *GIF* gene families may play critical roles in the growth and development of these organs or tissues (Zhang et al., 2008).

Grasses have important potential in providing human and animal nutrition and they may be strategic candidates for renewable energy sources (Vain, 2011). *Brachypodium distachyon* (L.) Beauv., named "purple false broom", is a new model plant for grasses and herbaceous energy crops (Draper et al., 2001). The whole-genome sequence of *Brachypodium* was completed and it provides information for understanding grass genome evolution (International Brachypodium Initiative, 2010). Based on the genome sequencing data, the genus *Brachypodium* is more closely related to wheat, barley, and forage grasses than to rice (Opanowicz et al., 2008). Furthermore, *Brachypodium* is convenient for functional genomics research in grasses owing to its small genome size and physical stature, short lifecycle, and simple growth requirements (Ozdemir et al., 2008). In the present study, we aimed to investigate *Brachypodium GRF* genes at the genome-wide scale. *Brachypodium GRF* gene numbers and duplications, exon-intron structures, protein motif analysis, physicochemical properties, putative biological functions, and phylogeny were analyzed in detailed.

2. Materials and methods

2.1. Identification of the GRF family in *Brachypodium*

We used 9 *Arabidopsis* (Kim et al., 2003), 14 maize (Zhang et al., 2008), and 12 rice (Choi et al., 2004) GRF protein sequences collected from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/protein/) as query sequences. Subsequently, we performed a BLASTP search of the *Brachypodium distachyon* genome at the Joint Genome Institute (http://www.phytozome.net). The sequences were selected as predicted proteins if their E-value satisfied $E \leq e^{-10}$ and redundant sequences were removed. We obtained information on protein sequences, cDNA sequences, genomic sequences, intron distribution patterns and phases, and intron/exon boundaries. The Pfam (http://pfam.sanger.ac.uk) and SMART (http://smart.embl-heidelberg.de) proteomics servers were then used to verify the conserved domains of GRF proteins.

2.2. Motif and phylogenetic analysis of predicted GRF proteins in *Brachypodium*

All confirmed BdGRF protein sequences were aligned using Clustal W (Thompson et al. 1994) in BioEdit 7.1.3.0 (Hall, 1999). The conserved motif analysis was performed with MEME (Multiple Em for Motif Elicitation) software (Timothy et al., 2009). The following parameter settings were used: distribution of motifs, 0 or 1 per sequence; maximum number of motifs to find, 5; minimum width of motif, 6; maximum width of motif, 50. Phylogenetic analyses were conducted using MEGA version 5.1 (Tamura et al., 2011) by a neighbor-joining tree based on the multiple sequence alignment of all predicted GRF protein sequences including the following parameters: Poisson correction, pairwise deletion, and bootstrap analysis with 1000 replicates.

2.3. Chromosomal distribution, gene duplication, and structural analysis of GRF genes in *Brachypodium*

To identify gene duplications among all putative genes, the following parameters were adopted: the alignment of the coding nucleotide sequences covered 70% of the longest genes and the amino acid identity between the sequences was >70% (Yang et al., 2008). A structural analysis of *Brachypodium GRF* genes, including exon and intron numbers and locations as well as conserved domain locations, was performed and displayed using the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) (Guo et al., 2007). Open reading frames (ORFs) were determined by using an ORF finder online (http://www.ncbi.nlm.nih.gov/projects/gorf/). Physicochemical characteristics of GRF proteins were computed using the online ProtParam tool (http://www.expasy.org/tools/protparam.html), including the number of amino acids, molecular weight, and theoretical isoelectric point (pI).

2.4. Putative functional analysis of *Brachypodium* GRFs

Functional annotations of BdGRF proteins were surveyed based on the Gene Ontology (GO) term analysis tool of Gramene (http://www.gramene.org/) developed by the Gene Ontology Consortium (Ashburner et al., 2000). Accordingly, 10 BdGRF proteins were evaluated according to their molecular functions, biological processes, and cellular localizations.

3. Results

To identify the GRF coding genes in the *Brachypodium* genome, GRF proteins of *Arabidopsis* (9), maize (14), and rice (12) were used as query sequences. In total, 10 genes were identified as potential encoding GRF proteins. Subsequently, all predicted GRF proteins were surveyed to verify whether they contained QLQ and WRC motifs (Table 1), which are the main characteristic residual motifs in GRF proteins. It was confirmed that all predicted GRF proteins contained QLQ and WRC domains (Figure 1).

Their genome distributions and duplication analysis were studied and, among 10 Bd*GRF* genes, only a single segmental duplication was estimated between the 2 *GRF* genes (Bd*GRF3* and Bd*GRF6*) (Figure 2). In general, duplication events can cause gene expansion, especially in protein families. However, it seems that *GRF* gene duplications in *Brachypodium* did not cause gene expansion in the *GRF* gene family.

Bd*GRF* ORF lengths ranged from 645 bp (Bd*GRF9*) to 1347 bp (Bd*GRF1*), and molecular weights ranged from 22.55 kDa (BdGRF9) to 48.83 kDa (BdGRF10), and pI values ranged from 4.81 (BdGRF6) to 9.64 (BdGRF9) (Table 1). *GRF* genes were distributed in all chromosome of the *Brachypodium* genome (Figure 2). The largest number of *GRF* genes was detected on chromosome 1, including 5 genes; in contrast, only 1 gene was located on chromosomes 2, 4, and 5 (Figure 2; Table 1). Based on exon and intron structures, the average intron number was 2, while the average exon number of Bd*GRF* genes was 3 (Table 1; Figure 3). Eight genes had 2 or more introns, whereas only 2 genes had 1 intron. Motif distribution analysis was performed using the MEME web server and a total of 5 common motifs were observed (Figure 4; Table 2).

Motif I and motif II were especially distinctively observed in all predicted GRF proteins of *Brachypodium* and these motifs contained conserved GRF protein specific domains (QLQ and WRC). In addition, motif III and motif IV had the other GRF domains (FFD and TQL). The most similar motif types were determined in

Table 1. Bd*GRF* genes in *Brachypodium*, including their physiochemical, structural, and sequence properties.

Gene name	Sequence ID	Chr	Start	Stop	ORF length (bp)	Exon number	Intron number	Length (aa)	MW (kDa)	pI
BdGRF1	Bradi1g09900.1	1	7121044	7123408	1347	3	2	448	48.8354	6.36
BdGRF2	Bradi1g12650.3	1	9519556	9522488	1056	3	2	351	39.6293	9.34
BdGRF3	Bradi2g14320.1	2	12898050	12901905	780	3	2	259	27.7278	4.84
BdGRF4	Bradi4g16450.1	4	17177409	17180766	1074	4	3	393	41.9267	9.19
BdGRF5	Bradi5g20607.1	5	23439346	23443316	1182	4	3	393	43.1902	8.50
BdGRF6	Bradi1g28400.1	1	23635036	23639109	792	3	2	263	28.4674	4.81
BdGRF7	Bradi1g46427.1	1	44949148	44950328	966	2	1	321	34.9799	8.80
BdGRF8	Bradi1g50597.1	1	49186025	49188850	1035	3	2	344	37.3914	8.94
BdGRF9	Bradi3g51685.1	3	52726473	52727314	645	3	2	214	22.5554	9.64
BdGRF10	Bradi3g57267.1	3	56937080	56938992	1257	2	1	418	45.5962	7.78

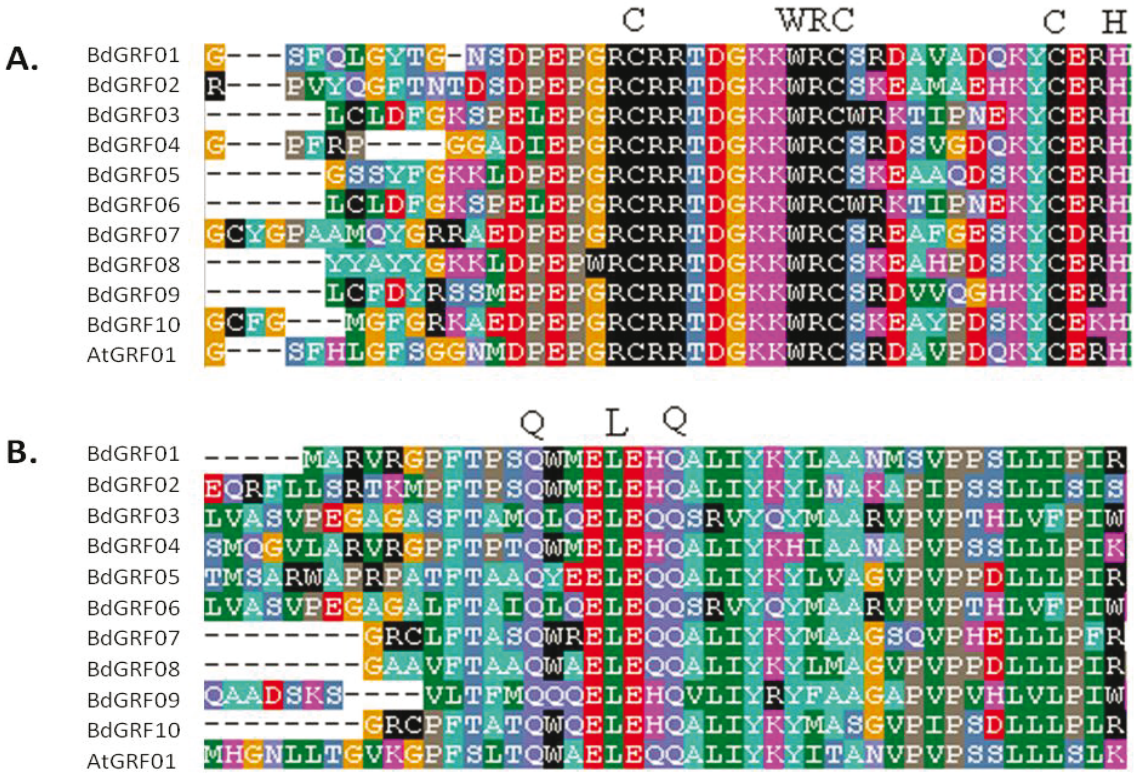


Figure 1. Comparison of the amino acid sequences of BdGRF proteins. A) The WRC domains of BdGRF and AtGRF1 proteins with the Cys3His zinc-finger motif. B) The QLQ domains of BdGRF and AtGRF1 proteins. AtGRF protein is selected to show the similarity with BdGRF protein sequences.

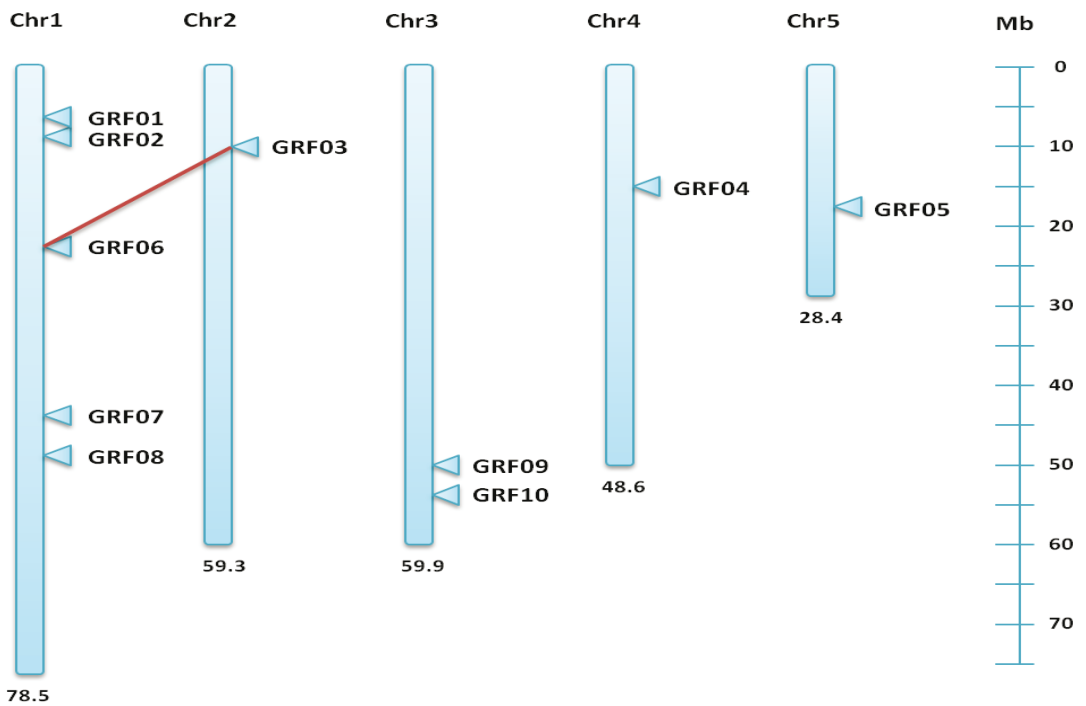


Figure 2. Genome distribution and duplication analysis of *Brachypodium GRF* genes on chromosomes 1 to 5. Red line shows segmental duplication between *GRF3* and *GRF6*. The ruler represents values in megabases (Mb).

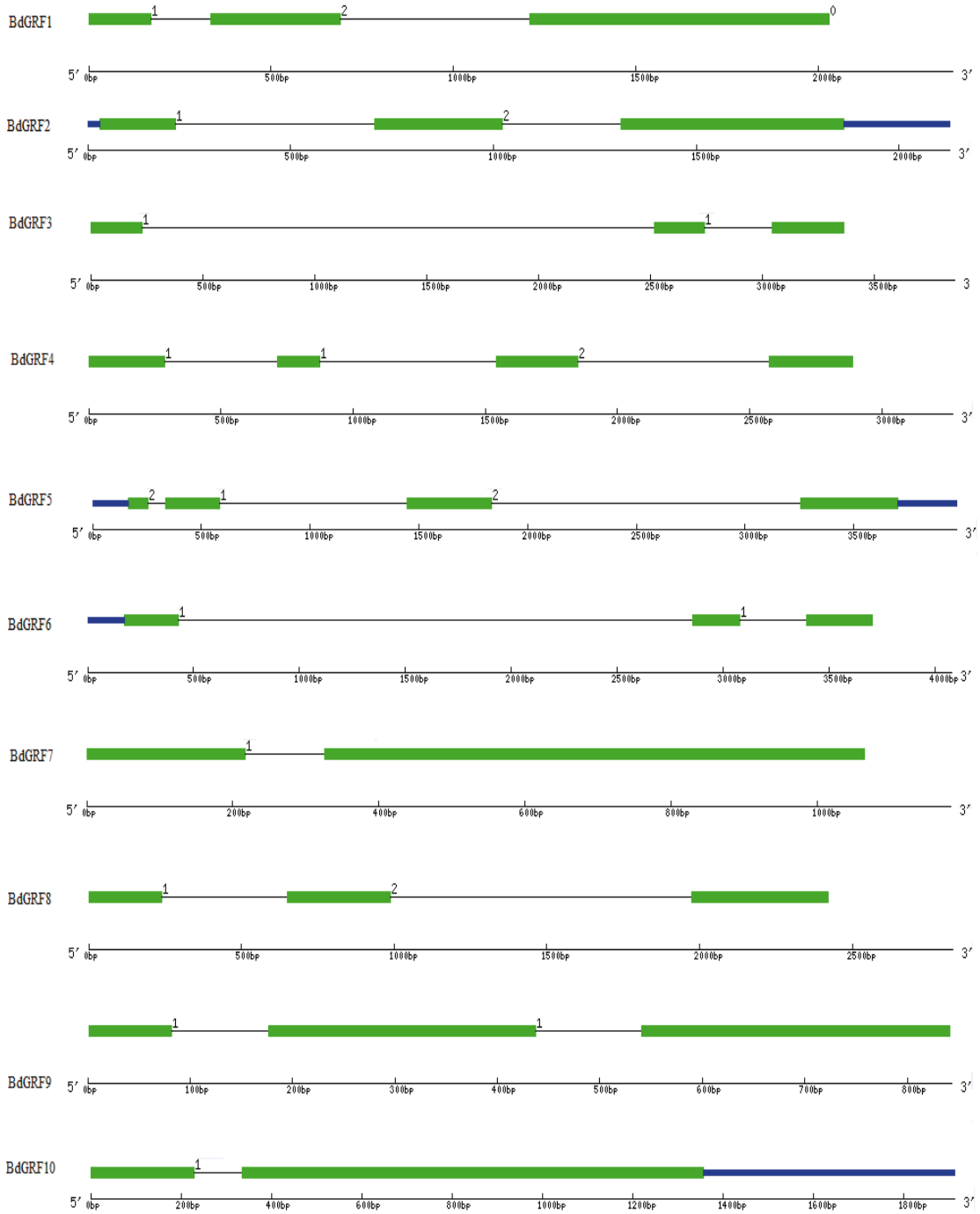


Figure 3. Gene structure of *Brachypodium GRF* genes. Exons and introns are depicted by filled green boxes and single lines, respectively. Intron phases 0, 1, and 2 are indicated by numbers 0, 1, and 2. Untranslated regions are displayed by thick blue lines at both 5' and 3' ends of each gene.

BdGRF5, BdGRF7, BdGRF8, and BdGRF10 with motif I, motif II, motif III, and motif IV, whereas the least common motif type was identified in BdGRF2, including motif I and motif II. Motif III (in BdGRF5, BdGRF7, BdGRF8, and BdGRF10), motif IV (BdGRF1, BdGRF5, BdGRF7, BdGRF8, and BdGRF10), and motif V (BdGRF3, BdGRF6, and BdGRF9) were observed 5, 6, and 3 times, respectively.

In order to analyze the phylogenetic organization of the GRF proteins in *Brachypodium*, 10 GRF proteins were used with MEGA 5.1 based on the neighbor-joining

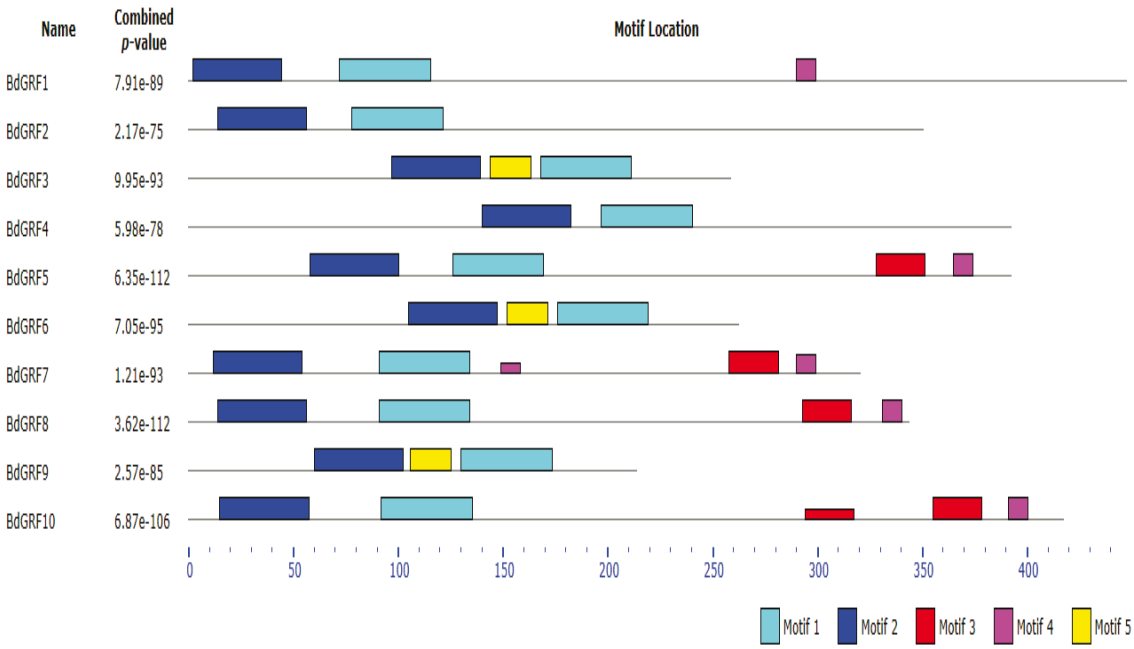


Figure 4. The most conserved protein motifs in BdGRFs (motif I, motif II, motif III, motif IV, and motif V, respectively). Each motif was represented in boxes with different colors: motif I, cyan; motif II, blue; motif III, red; motif IV, purple; and motif V, yellow.

Table 2. The most conserved protein motifs in GRF protein sequences of *Brachypodium*. Bolded residues show WRC, QLQ, FFD, and TQL domains, respectively.

Motif number	Width Sequence	Protein sequences
1	43	PEPGRCRRTDGKK**WRC**WREAIPDHKYCERHMHRGRNRSRKPVE
2	42	RVRCPFTAMQW**QE**L**E**H**Q**ALIYKYMAAGVPVPTHLLIPIWKSF
3	23	NVKQENKTLRP**FFD**EWPKERDNW
4	9	**TQL**SISIPM
5	19	EGAQNYPTLMGLATLCLDF

method (Figure 5). Accordingly, *Brachypodium* GRF proteins could be classified into 2 main groups (I and II). The main group I consisted of 6 GRF proteins (BdGRF1, BdGRF2, BdGRF5, BdGRF7, BdGRF8, and BdGRF10), whereas the main group II had 4 members (BdGRF3, BdGRF4, BdGRF6, and BdGRF9). The highest bootstrap value (100%) was observed in the main group II between BdGRF3 and BdGRF6.

To examine the phylogenetic relationship between monocot and dicot *GRF* genes, *Brachypodium*, rice, maize, and *Arabidopsis* GRF sequences were retrieved from genome databases. Ten BdGRF, 12 OsGRF, 14 ZmGRF, and 9 AtGRF protein sequences were used for comparative phylogenetic analysis (Figure 6). A total of 45 full-length protein sequences of GRF from monocot and dicot plant species were divided into 2 main groups, including subgroups named A, B, C, D, E, F, and G. In the comparative phylogenetic tree, subgroups A and F consisted of monocot species, whereas subgroups D and E had only dicot species. Notably, monocot (*Brachypodium*, maize, and rice) and dicot (*Arabidopsis*) GRF proteins were clustered together in subgroups B, C, and G. The highest number of BdGRFs were observed in subgroup C with 3 members (BdGRF1, 2, and 4), while the lowest was in subgroup F, including only 1 member (BdGRF9). BdGRF1 protein showed maximum similarity with maize at a 94% bootstrap value in subgroup C. Interestingly, a *Brachypodium* internal clade was observed between BdGRF3 and BdGRF6 with 57%

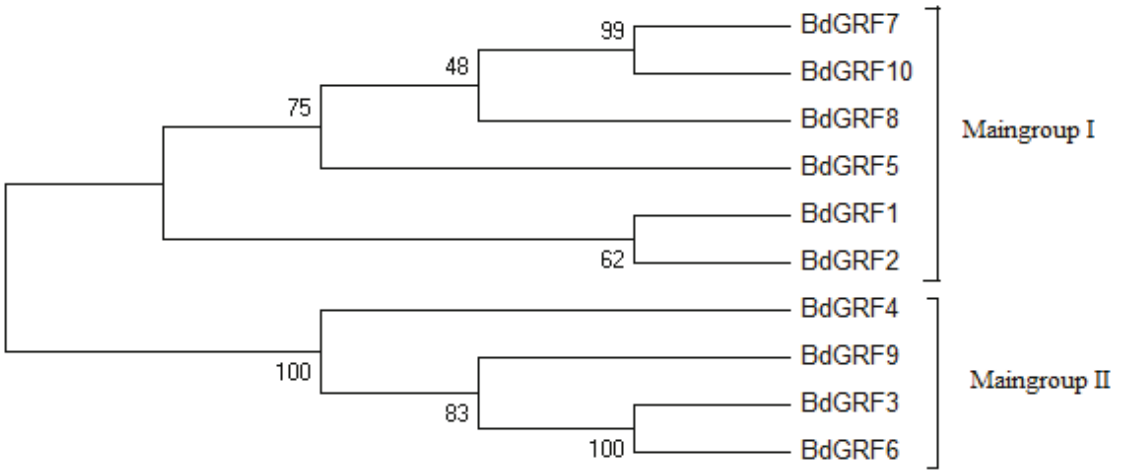


Figure 5. Phylogenetic tree of *Brachypodium* GRF proteins**.** The phylogenetic tree was constructed with MEGA 5.1 using the neighbor-joining method.

bootstrap value in subgroup G. AtGRF3–4 (in subgroup E) and AtGRF7–8 (in subgroup D) were separated from other GRFs. It was also observed that the outer bootstrap values were generally lower than the internal values. The overall phylogenetic analysis of *Brachypodium* GRF proteins clearly revealed a complicated phylogenetic relationship with other monocot and dicot plant species.

4. Discussion

In this study, we used *GRF* gene and protein sequences of *Arabidopsis*, rice, and maize as queries to find Bd*GRF* genes. Finally, 10 nonredundant *GRF* genes were identified and characterized in the *Brachypodium* genome. In the last decade, the *GRF* gene family has been identified and described in some plant species in detail. According to previous studies, 9 At*GRF* genes (Kim et al., 2003), 12 Os*GRF* genes (Choi et al., 2004), and 14 Zm*GRF* genes (Zhang et al., 2008) were identified in *Arabidopsis*, rice, and maize, respectively. When comparing the *Brachypodium* genome size with other grass genomes, *Brachypodium* has a much smaller genome size (0.3 Mb) than rice (0.4 Gb) or maize (2.5 Gb) (Vain, 2011). We found similar gene numbers in *Brachypodium* with 10 genes and it may support the idea that these *GRF* genes were conserved in monocot and dicot plant species. In rice, all OsGRF proteins include the highly conserved QLQ, WRC, and TQL domains in the N-terminal region (Choi et al., 2004). In maize, QLQ, WRC, TQL, and FFD domains were identified (Zhang et al., 2008). In our study, we identified 4 domains in BdGRF protein sequences containing WRC, QLQ, TQL, and FFD (Table 2; Figure 2), consistent with earlier studies.

It is widely accepted that the intron/exon structure contributes to the understanding of evolutionary relationships (Hu and Liu, 2011). Additionally, exon/intron gain/loss was substantial for structural divergence and functional differentiation (Xu et al., 2012). In *GRF* genes of *Arabidopsis*, 7 genes contain 3 introns while 2 genes have 2 introns. In rice, *GRF* genes contained between 2 and 4 introns. The intron number and exon-intron organization of *GRF* genes in *Arabidopsis* and rice are not well conserved, and these data showed that gene duplications in the *GRF* family have not been occurred recently (Choi et al., 2004). Our analysis showed similar findings to the previous studies, such that 2, 3, and 6 members had 1 intron, 2 introns, and 3 introns, respectively (Table 1). It was suggested that *GRF* genes in *Brachypodium* may have a similar history as in other monocot and dicot plants and that these *GRF* genes were not well conserved in the *Brachypodium* genome. Phylogenetic analysis revealed that *Brachypodium* GRF proteins were more closely clustered with maize and rice than in *Arabidopsis*, including high bootstrap values of 78% and 94% in maize and rice, respectively, in subgroup C. In contrast, the highest bootstrap value was found only between BdGRF3 and BdGRF6 (57%). As indicated in Figure 5, there were no higher bootstrap values in the joined tree (Figure 6). This could be explained by the fact that some Bd*GRF* genes were more similar to the *GRF* genes of other monocot species; this result may be related to *GRF* gene structures, which could be affected by some genomic forces, including insertion, deletion, and transposon activities. Furthermore, GRF proteins were clustered with *Arabidopsis* in subgroups B, C, and G (Figure 6). In a phylogenetic tree, it was shown that *Arabidopsis* and rice GRF proteins were clustered together within 3 subfamilies as A, B, and C (Choi et al., 2004). Our findings are consistent with these results. It can be proposed that the orthology of *GRF* genes may cause clustering of monocot and dicot subgroups in the joined tree and may reflect the functional conservation of plant *GRFs*. Although

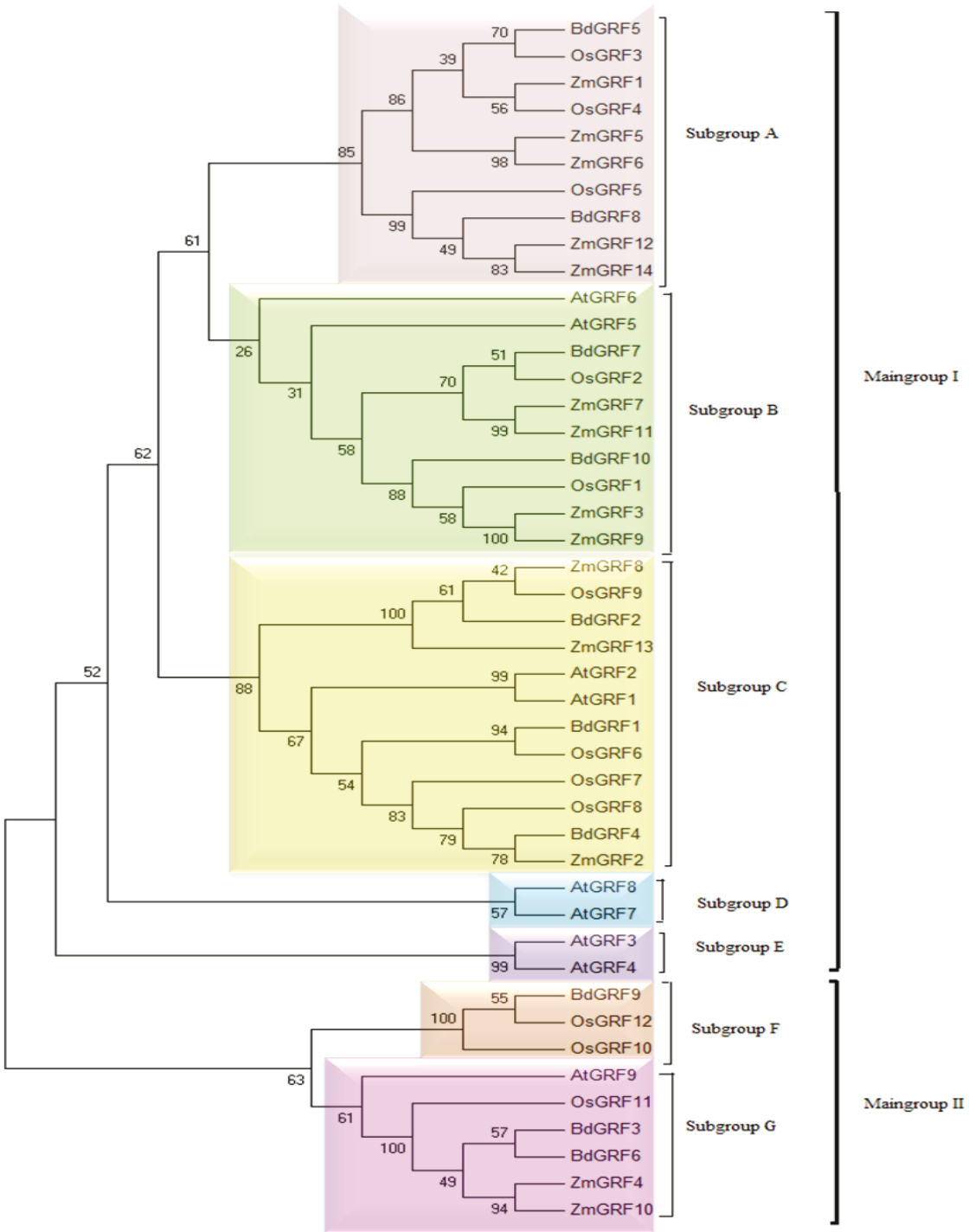


Figure 6. Phylogenetic tree of *Brachypodium*, *Arabidopsis*, maize, and rice GRF proteins. The phylogenetic tree was generated using the MEGA 5.1 program with the neighbor-joining method.

some gene families are more dynamic, others are more conserved and orthologous (Martinez, 2011). *GRF* genes could be conserved and they showed orthology in plants that generated mixed subgroups, including monocot and dicot plants, in the joined phylogenetic tree (Figure 6).

Gene duplication events affect gene family distribution in the genome (Cannon et al., 2004). Duplications in plant genomes include various scales containing tandem and segmental duplications (small-scale) or whole-genome duplications (large-scale) (Ramsey and

Table 3. Putative functions and cellular localizations of GRF proteins in *Brachypodium*.

Gene name	Sequence ID	GO*: Molecular function	GO: Biological process	GO: Cellular component
BdGRF1	Bradi1g09900.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF2	Bradi1g12650.3	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF3	Bradi2g14320.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF4	Bradi4g16450.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF5	Bradi5g20607.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF6	Bradi1g28400.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF7	Bradi1g46427.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF8	Bradi1g50597.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF9	Bradi3g51685.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus
BdGRF10	Bradi3g57267.1	ATP binding Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	Regulation of transcription, DNA-dependent	Nucleus

GO: Gene Ontology terms for BradiGRF proteins were derived from the data on the Gramene GO server (Ashburner et al., 2000; Jaiswal et al., 2002).

Schemske, 1998). Tandem duplication contains 2 or more genes located in the same chromosome; on the contrary, segmental duplications require gene duplications between different chromosomes (Liu et al., 2011). In *Arabidopsis*, approximately 25% of genes were produced by whole-genome duplication; in contrast, approximately 16% of genes were tandem duplicates (Rizzon et al., 2006). In our study, single gene duplication was identified as segmental between Bd*GRF3* and Bd*GRF6* (Figure 2). Genome distribution of *GRF* genes indicated that segmental duplication somewhat contributed to the expansion of *Brachypodium GRF* genes. In addition, the phylogenetic tree supports that these genes were clustered together with the highest bootstrap value (100%) in the main group II (Figure 5).

Putative functional evaluations of BdGRF proteins were performed based on the information retrieved from the Gramene GO database (Table 3). The GO classification method improves our understanding of gene classifications in terms of their associated biological processes, cellular components, and molecular functions (Conesa et al., 2005). In this study, the functional classifications of BdGRF proteins were observed in the same GO groups. For instance, according to their molecular function predictions,

all BdGRF proteins were found to reside in the same groups as "ATP binding" (GO:0005524) and "hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides" (GO:0016818). Additionally, their cellular biological processes were proposed to belong to the "regulation of transcription, DNA-dependent" group (GO:0006355). Furthermore, the cellular components of all BdGRF proteins were determined in the same localization, as "nucleus" (GO:0005634) (Table 3). These data showed that BdGRF proteins could have similar biological roles and functions. Based on the detailed functional studies on *Arabidopsis*, GRF proteins were proposed to be related to development, regulations of leaf size and shape, and the transcriptional regulation of stress genes acting as activators or repressors (Horiguchi et al., 2005; Kim et al., 2012). By considering the homology and/or orthology modeling of *Arabidopsis* and *Brachypodium GRF* genes, particular functions of Bd*GRF* genes can be estimated. In our study, it is noteworthy that functional predictions of BdGRF proteins were based on the gene annotations. For future studies, these predictions need to be proven by functional evidence. We assume that the identification of Bd*GRF* genes and the represented data will be helpful for the further functional identifications of BdGRF proteins.

In conclusion, this research has contributed to the understanding of the *GRF* gene family of *Brachypodium*. In addition, identification and phylogenetic and comparative analyses of Bd*GRF* genes could be useful for the discovery of new *GRF* members in other plant species, especially in grasses.

References

Aizawa H, Hu SC, Bobb K, Balakrishnan K, Ince G, Gurevich I, Cowan M, Ghosh A (2004). Dendrite development regulated by CREST, a calcium-regulated transcriptional activator. Science 303: 197–202.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. (2000). Gene Ontology: tool for the unification of biology. Nat Genet 25: 25–29.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4: 53–62.

Choi D, Kim JH, Kende H (2004). Whole genome analysis of the Os*GRF* gene family encoding plant-specific putative transcription activators in rice (*Oryza sativa* L.). Plant Cell Physiol 45: 897–904.

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM (2001). *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol 127: 1539–1555.

Guo AY, Zhu QH, Chen X, Luo JC (2007). GSDS: a gene structure display server. Yi Chuan 29: 1023–1026.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acid S 41: 95–98.

Horiguchi G, Kim GT, Tsukaya H (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana. Plant J 43: 68–78.

Hu LF, Liu SQ (2011). Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers. Genet Mol Biol 34: 624–633.

International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463: 763–768.

Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, Pan X, Clark K, Teytelman L, Cartinhour S et al. (2002). Gramene: development and integration of trait and gene ontologies for rice. Compar Funct Genom 3: 132–136.

Kim JH, Choi D, Kende H (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in *Arabidopsis*. Plant J 36: 94–104.

Kim JH, Kende H (2004). A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in *Arabidopsis*. P Natl Acad Sci USA 101: 13374–13379.

Kim JS, Mizoi J, Kidokoro S, Maruyama K, Nakajima J, Nakashima K, Mitsuda N, Takiguchi Y, Ohme-Takagi M, Kondou Y et al. (2012). Arabidopsis GROWTH-REGULATING FACTOR7 functions as a transcriptional repressor of abscisic acid and osmotic stress–responsive genes, including *DREB2A*. Plant Cell 24: 3393–3405.

Liu Y, Jiang HY, Chen W, Qian Y, Ma Q, Cheng B, Zhu S (2011). Genome-wide analysis of the auxin response factor (*ARF*) gene family in maize (*Zea mays*). Plant Growth Regul 63: 225–234.

Martinez M (2011). Plant protein-coding gene families: emerging bioinformatics approaches. Trends Plant Sci 16: 1360–1385.

Opanowicz M, Vain P, Draper J, Parker D, Doonan JH (2008). *Brachypodium distachyon*: making hay with a wild grass. Trends Plant Sci 13: 172–177.

Ozdemir BS, Hernandez P, Filiz E, Budak H (2008). *Brachypodium* genomics. Int J Plant Genomics 2008: 536104.

Ramsey J, Schemske DW (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Evol Syst 29: 467–501.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290: 2105–2110.

Rizzon C, Ponger L, Gaut BS (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLoS Comput Biol 2: e115.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731–2739.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

Timothy L, 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: 202–208.

Treich I, Cairns BR, de los Santos T, Brewster E, Carlson M (1995). SNF11, a new component of the yeast SNF-SWI complex that interacts with a conserved region of SNF2. Mol Cell Biol 15: 4240–4248.

Vain P (2011) *Brachypodium* as a model system for grass research. J Cereal Sci 54: 1–7.

Van der Knaap E, Kim JH, Kende H (2000). A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. Plant Physiol 122: 695–704.

Xu G, Guo C, Shan H, Kong H (2012). Divergence of duplicate genes in exon–intron structure. P Natl Acad Sci USA109: 1187–1192.

Yamasaki K, Kigawa T, Inoue M, Watanabe S, Tateno M, Seki M, Shinozaki K, Yokoyama S (2008). Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. Plant Physiol Bioch 46: 394–401.

Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genomics 280: 187–198.

Zhang DF, Li B, Jia GQ, Zhang TF, Dai JR, Li JS, Wang SC (2008). Isolation and characterization of genes encoding GRF transcription factors and GIF transcriptional coactivators in Maize (*Zea mays* L.). Plant Sci 175: 809–817.