

# In Silico Identification and Characterization of B12D Family Proteins in Viridiplantae

Zainab M Almutairi 

Department of Biology, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-kharj, Saudi Arabia.

Evolutionary Bioinformatics  
Volume 18: 1–12  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11769343221106795



**ABSTRACT:** B12D family proteins are transmembrane proteins that contain the B12D domain involved in membrane trafficking. Plants comprise several members of the B12D family, but these members' numbers and specific functions are not determined. This study aims to identify and characterize the members of B12D protein family in plants. Phytozome database was retrieved for B12D proteins from 14 species. The total 66 B12D proteins were analyzed in silico for gene structure, motifs, gene expression, duplication events, and phylogenetics. In general, B12D proteins are between 86 and 98 aa in length, have 2 or 3 exons, and comprise a single transmembrane helix. Motif prediction and multiple sequence alignment show strong conservation among B12D proteins of 11 flowering plants species. Despite that, the phylogenetic tree revealed a distinct cluster of 16 B12D proteins that have high conservation across flowering plants. Motif prediction revealed 41 aa motif conserved in 58 of the analyzed B12D proteins similar to the bZIP motif, confirming that in the predicted biological process and molecular function, B12D proteins are DNA-binding proteins. *Cis*-regulatory elements screening in putative *B12D* promoters found various responsive elements for light, abscisic acid, methyl jasmonate, cytokinin, drought, and heat. Despite that, there is specific elements for cold stress, cell cycle, circadian, auxin, salicylic acid, and gibberellic acid in the promoter of a few *B12D* genes indicating for functional diversification for B12D family members. The digital expression shows that *B12D* genes of *Glycine max* have similar expression patterns consistent with their clustering in the phylogenetic tree. However, the expression of *B12D* genes of *Hordeum vulgure* appears inconsistent with their clustering in the tree. Despite the strong conservation of the B12D proteins of Viridiplantae, gene association analysis, promoter analysis, and digital expression indicate different roles for the members of the B12D family during plant developmental stages.

**KEYWORDS:** B12D family, plants, gene expression, *cis*-elements, phylogenetics

**RECEIVED:** January 25, 2022. **ACCEPTED:** May 12, 2022.

**TYPE:** Original Research

**FUNDING:** The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported under Project No. 2021/01/17976, from the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University.

**DECLARATION OF CONFLICTING INTERESTS:** The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Zainab M Almutairi, Department of Biology, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, P.O. Box. 83, Al-kharj 11942, Saudi Arabia. Email: z.almutairi@psau.edu.sa

## Introduction

The coordination of plant growth and reproduction processes is complicated and involves many regulatory proteins. Previous studies in plant transcripts during seed germination have identified various proteins expressed in aleurone and embryo during differentiation called B12D proteins. The first member of B12Ds was identified by screening the transcriptome of barley aleurone and embryo.<sup>1</sup> Thereafter, other members of B12D proteins have been identified and found to be expressed in all plant tissues at different developmental stages.<sup>2,3</sup> Moreover, a member of B12D proteins is involved in leaf senescence in plants.<sup>4</sup>

B12D family proteins are small transmembrane proteins containing the B12D domain. Six B12D proteins have been identified in the rice genome.<sup>5</sup> On the other hand, at least 8 or 9 B12D proteins were suggested to be expressed in various barley tissues.<sup>2</sup> However, the number of B12D proteins appears to be different across plant species. One of the B12D proteins is previously known to be the subunit NDUFA4 of the mammalian electron transport chain. NDUFA4, also called NADH-ubiquinone reductase complex I subunit MLRQ,<sup>6</sup> is conserved in insects, fungi, and higher metazoans and is involved in ATP synthesis.<sup>7</sup> Subcellular localization reveals that B12D protein from *Arabidopsis* #AT3G48140 is localized in the mitochondrion,<sup>8</sup> plasma membrane,<sup>9</sup> and peroxisome.<sup>10</sup> However, B12D

proteins appear to have a single transmembrane helix and are embedded in the inner mitochondrial membrane.<sup>5,6</sup>

B12D proteins are expressed in various plant tissues, such as starchy endosperm, pericarp, immature and mature embryos, aleurone, seedling shoots, flowers during heading and early ripening spikelets.<sup>2,5,11</sup> *B12D* genes appear to be preferentially regulated by plant hormones like gibberellic acid, abscisic acid, and ethylene.<sup>1,2,12</sup> Additionally, *B12D* genes are regulated in response to different abiotic and biotic stress.<sup>5,12</sup> Rice *B12D* gene #Os07g41340 is induced by flooding, salt, heat, and cold stress during germination.<sup>5,13</sup> Likewise, Os07g41340 is involved in the rice defense response to biotic stress induced by blast disease caused by *Magnaporthe oryzae*.<sup>14</sup> Moreover, rice *B12D* gene #Os07g17330 is regulated in response to submergence in 5 rice genotypes.<sup>15</sup>

Identifying the *B12D* gene family in Viridiplantae is essential to characterize this family member and understand their role during plant development. The present study aims to characterize the members of the *B12D* gene family in Viridiplantae by identification and in silico characterization of gene structure, functional motifs, phylogenetics, screening of *cis*-regulatory elements in the putative promoters, and digital expression analysis. The result of this study is expected to help uncover the role of *B12D* gene family members during plant differentiation and maturation.



## Experimental Procedures

### Database sequence retrieval

The Phytozome database (<https://phytozome-next.jgi.doe.gov/><sup>16</sup>) was retrieved for B12D proteins across Viridiplantae. Fourteen species were selected to identify the B12D protein family in their genomes. The selection of species was based on their classification which the selected species present the main divisions of Viridiplantae. The 14 species are including 3 gymnosperms as follows: *Botryococcus braunii* v2.1 (chlorophyte), *Marchantia polymorpha* v3.1 (embryophyte), *Selaginella moellendorffii* v1.0 (tracheophyte), and 11 angiosperms include 1 Amborellales, *Amborella trichopoda* v1.0, 5 dicots (*Aquilegia coerulea* v3.1, *Helianthus annuus* r1.2, *Glycine max* Wm82.a2.v1, *Populus trichocarpa* v3.0, *Arabidopsis thaliana* TAIR10), and 5 monocots (*Musa acuminata* v1, *Hordeum vulgare* r1, *Oryza sativa* v7.0, *Setaria italica* v2.2, *Zea mays* RefGen\_V4). For each protein found in the selected 14 species, the CDS (coding sequence), genomic, and amino acid sequences were downloaded from the Phytozome database. To ensure that all retrieved proteins are members of the B12D family, we identified the B12D domain (Pfam #PF06522) using Pfam (<http://pfam.sanger.ac.uk>) and my hits scan tool ([http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan)). Chromosome location for each gene was detected using the CDS sequence for each protein as a query in the NCBI tBLASTn against the Whole Genome Shotgun Contigs database. The genome size for each species was retrieved from EnsemblePlants database (<https://plants.ensembl.org/index.html><sup>17</sup>) to investigate the relationship between genome size and the number of B12D members in each plant genome.

### In silico characterization and phylogeny

Functional protein association to predict the function of the B12D family proteins was predicted by STRING 11.0 ([string-db.org](http://string-db.org)) for the 6 B12D proteins from *O. sativa*. Transmembrane helix prediction was performed using MEMSAT-SVM.<sup>18</sup> The isoelectric point (PI) and molecular weight (MW) of B12D proteins were calculated by GeneScript (<https://www.genscript.com/tools/>). Motif prediction in protein sequences conducted by the MEME web server (<http://meme-suite.org/tools/meme><sup>19</sup>). Sequence alignment of conserved motifs was carried out using UniProt UGENE software.<sup>20</sup> Biological processes, molecular function, and cellular components were predicted by FFPred server<sup>21</sup> for *A. thaliana* and *O. sativa* B12D proteins. Exon/intron structures were generated by the Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/><sup>22</sup>) using the corresponding CDS and genomic sequences for each B12D protein. The phylogenetic tree for the 66 B12D proteins was constructed based on B12D domain sequences by excluding the N and C-terminals sequences. The tree is built using minimum evolution method<sup>23</sup>—with interior branch tests of 1000 replicates using MEGA X software.<sup>24</sup>

### Cis-regulatory analysis

In the current study, B12D proteins from 5 species, 3 monocots (*H. vulgare*, *S. italica*, and *O. sativa*) and 3 dicots (*G. max*, *P. trichocarpa*, and *A. thaliana*), were selected for regulatory *cis*-elements screening. The sequences of the putative promoters of the selected B12D genes were obtained using the genomic sequence for each protein available in the Phytozome database as a query in the NCBI tBLASTn against the Whole Genome Shotgun Contigs database. PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/><sup>25</sup>) was used to screen *cis*-elements in the putative promoter sequences 1500 bp upstream of the start codon.

### Digital expression of B12D genes

The digital expression analysis for B12D genes from *G. max* and *H. vulgare* was conducted using GENEVESTIGATOR v3 software (<https://genevestigator.com/><sup>26</sup>). The heatmap tree for gene expression in various plant tissues was obtained based on log<sub>2</sub> using the “anatomy” category and plant organs/tissues conditions. The heatmap tree for gene expression in various developmental stages was obtained based on log<sub>2</sub> using the “development” category and plant growth conditions.

### Investigating B12D genes duplication events

Gene duplication events were analyzed for paralogous clustered together in the phylogenetic tree and located in the same chromosome. The distance between the analyzed genes was not more than 50 kb.<sup>27</sup> The non-synonymous (K<sub>a</sub>) to synonymous (K<sub>s</sub>) substitutions ratio was calculated to investigate selection pressure for gene duplicates using the DnaSP v5.0 software (<http://www.ub.edu/dnasp/><sup>28</sup>). Alignment for the coding region of CDS sequences for each gene pair was used to calculate K<sub>a</sub>/K<sub>s</sub> ratio. A K<sub>a</sub>/K<sub>s</sub> ratio value >1 is considered a positive selection, and a value <1 is considered a negative selection.<sup>29</sup> The divergence time for each gene duplicate was estimated by dividing the K<sub>s</sub> value by the substitution rate, which is equal to 6.1 × 10<sup>-9</sup> per 10<sup>-6</sup> million years ago.<sup>30</sup>

## Results

### Characteristics of B12D proteins in Viridiplantae

The Phytozome database retrieval for B12D proteins in 14 species of Viridiplantae revealed a total of 66 B12D proteins in the selected species, as shown in Supplemental Table S1. Domain analysis showed that all retrieved proteins contained the B12D domain. The length of the retrieved B12D proteins ranges from 66 to 197 aa, of which 77% are between 86 and 98 aa. The isoelectric point values range from 5.71 to 11.37, of which 90% of analyzed proteins are between 9 and 10.8. The molecular weight ranges from 7.8 to 22.6 kDa, of which 80% of protein are between 9.4 and 11.2 kDa. The number of B12D proteins in each species was from 2 to 9 proteins as follows: 2

B12D proteins in each of *B. braunii* and *A. thaliana*, 3 B12D proteins in each of *S. moellendorffii* and *A. trichopoda*, 4 B12D proteins in each of *M. polymorpha*, *P. trichocarpa*, and *S. italica*, 5 B12D proteins in *M. acuminata*, 6 B12D proteins in each of *A. coerulea*, *H. annuus*, *G. max*, *O. sativa*, and *Z. mays*, and 9 B12D proteins in *H. vulgare*. The number of B12D proteins in lower Viridiplantae ranges from 2 to 4 B12D proteins in each species, while in higher Viridiplantae ranges from 4 to 6 B12D proteins in each species, with 2 exceptions: in *A. thaliana*, 2 B12D proteins and in *H. vulgare*, 9 B12D proteins. To investigate the relationship between the number of B12D members and the genome size in each species, the genome size data of the selected 14 species was retrieved and presented in Supplemental Table S1. In general, the number of B12D members increases from lower to higher Viridiplantae according to the increasing in genome size. However, the duplication events increase the number of B12D members in *M. polymorpha*, *A. coerulea*, and *O. sativa* despite the limited genome size in these species. Moreover, there are some exceptions as shown in *H. annuus* and *G. max* which both have an equal number of B12D members (6 proteins) despite the variation in the genome size (3500 and 1115 Mb, respectively).

The number of exons for each of the retrieved B12D proteins was between 1 and 4, of which 95% have 2 or 3 exons (Figure 1). Despite this similarity in exon numbers and lengths, the length of some introns appears significantly longer in 4 B12D proteins from 4 different species (HanXRQChr09g265771, A.trichopoda\_scaffold00055.169, Potri.012G74900, and LOC\_Os03g40440). Domain architecture appears conserved among most of the retrieved proteins, in which the B12D domain spans nearly from the first 10 aa at the N-terminus to the last 5 aa at the C-terminus of the proteins that range from 66 to 99 aa. However, longer proteins show a different domain architecture with the B12D domains located near the N-terminus (Supplemental Table S1). Transmembrane helix prediction showed that all retrieved B12D proteins comprise a single transmembrane helix except B12D protein #Potri.010G55300 of *P. trichocarpa*, which has 3 transmembrane helices (Supplemental Figure S1).

### Functional association and motifs

Functional protein association by STRING for the 6 B12D proteins from *O. sativa* showed the data only for 3 B12D proteins: LOC\_Os07g41340, LOC\_Os07g41350, and LOC\_Os06g13680 (Supplemental Figure S2). The other 3 B12D proteins of *O. sativa* (LOC\_Os03g40440, LOC\_Os07g17310, and LOC\_Os07g17330) have no data about their functional interaction in the STRING database. The 2 proteins, LOC\_Os07g41340 LOC\_Os07g41350, appear to interact and be associated with the same proteins, including 4 bidirectional sugar transporters (SWEET2A, SWEET3A, SWEET3B, and SWEET4), 2 deoxyribonucleoside kinases, 2 hypoxia-induced proteins, and a spiral 1-like protein. The *O. sativa* protein, LOC\_Os06g13680, interacted with 9 bidirectional sugar transporters

(SWEET1A, SWEET2A, SWEET3A, SWEET6A, SWEET1B, SWEET2B, SWEET3B, SWEET6B, and SWEET4) beside a berberine bridge enzyme, which is involved in the biosynthesis of peptidoglycan.

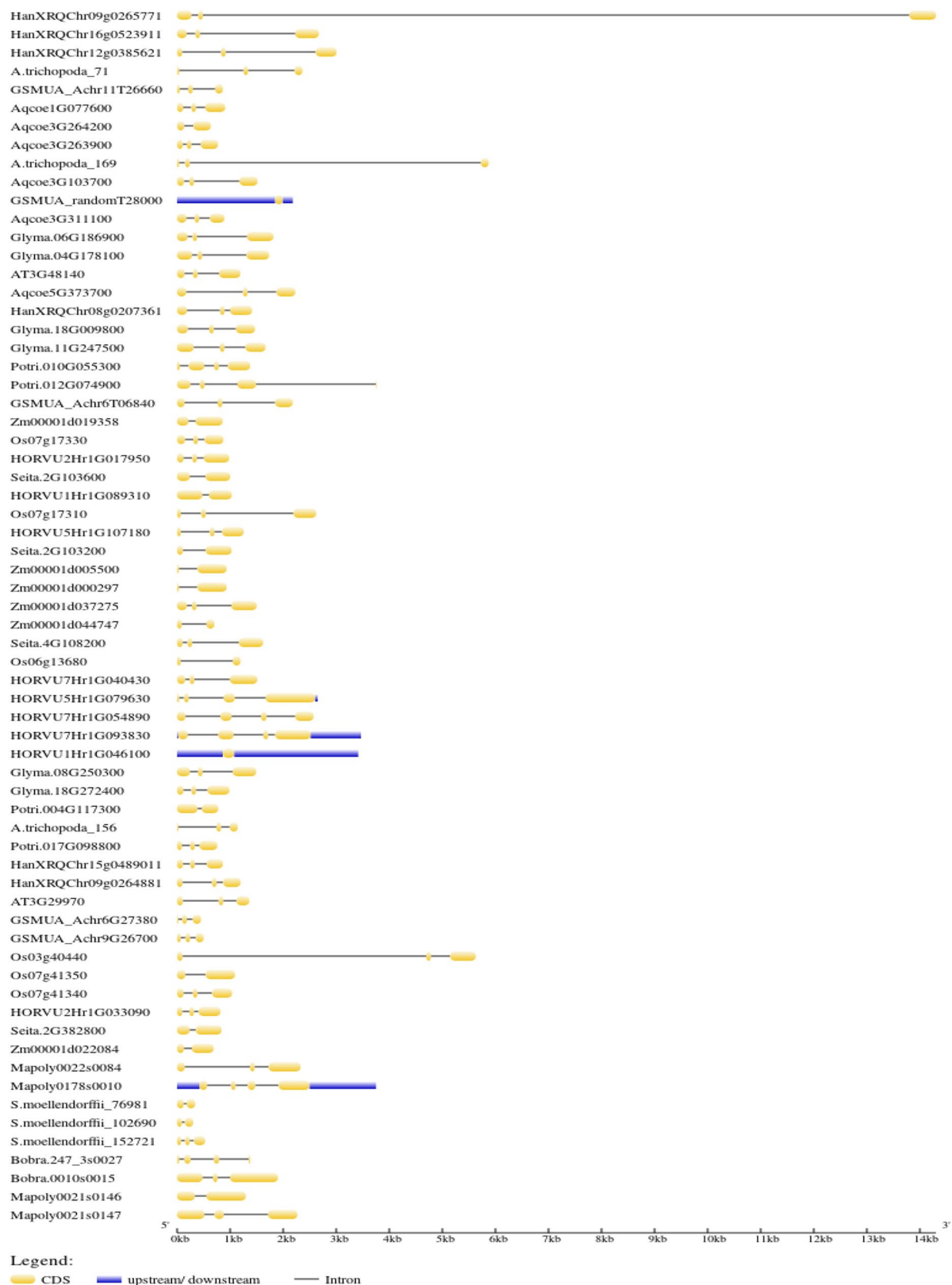
Motif prediction revealed 4 motifs conserved in most of the investigated B12D proteins (Figure 2). The first motif appeared as the basic leucine zipper (bZIP) motif with 41 aa length and was conserved in 58 of the analyzed B12D proteins. The second motif is 21 aa length and located in the N-terminal. The third motif is 15 aa in length and located in the C-terminal. The second and third motifs contain 2 proline residues and appear to be the PxxxP motif required for protein-protein interactions. The fourth motif (LRKFVR) is 6 aa in length and found in 65 of the analyzed B12D proteins close to the C-terminal (Figure 3). Some B12D proteins contain only 1 or 2 of the 4 motifs as shown in the 2 proteins of *B. braunii*; the proteins Mapoly0021s0146 and Mapoly0021s0147 of *M. polymorpha*; the proteins HORVU7Hr1G93830 and HORVU1Hr1G46100 of *H. vulgare*; and the protein Aqcoe3G264200 of *A. coerulea*.

Gene ontology analysis for *A. thaliana* and *O. sativa* proteins revealed a similar biological process and molecular function for the analyzed B12D proteins, which indicates the involvement of B12D proteins in transmembrane transport, and the establishment of localization in cells, and regulation of DNA-templated transcription. The molecular functions for the analyzed B12D proteins include catalytic activity, ion transmembrane transporter activity, nucleic acid binding, and cytokine activity. The cellular components for these proteins include the mitochondrial inner membrane, an integral component of the plasma membrane, endoplasmic reticulum, and extracellular vesicular exosome (Supplemental Tables S4-S9).

### Multiple sequence alignment and phylogenetics

Multiple sequence alignment of 66 B12D proteins shows that *M. polymorpha* B12D proteins Mapoly0021s0146 and Mapoly0021s0147 are less conserved among the aligned 66 proteins, followed by the 2 B12D proteins from *B. braunii* and the 3 B12D proteins from *S. moellendorffii*, which are all 3 species from the lower Viridiplantae. The B12D proteins from the other 11 flowering plant species appear to be more conserved, although the alignment shows 2 distinct groups. The first group includes 16 B12D proteins: 3 proteins from *O. sativa*, 2 proteins of each of *M. acuminata*, *H. annuus*, *G. max*, and *P. trichocarpa*, and a single B12D protein each of *A. trichopoda*, *A. thaliana*, *H. vulgare*, *S. italica*, and *Z. mays*. This group does not include any proteins of *A. coerulea*. The second group includes the other 41 B12D proteins from these 10 species besides the B12D proteins of *A. coerulea* (Supplemental Figure S4).

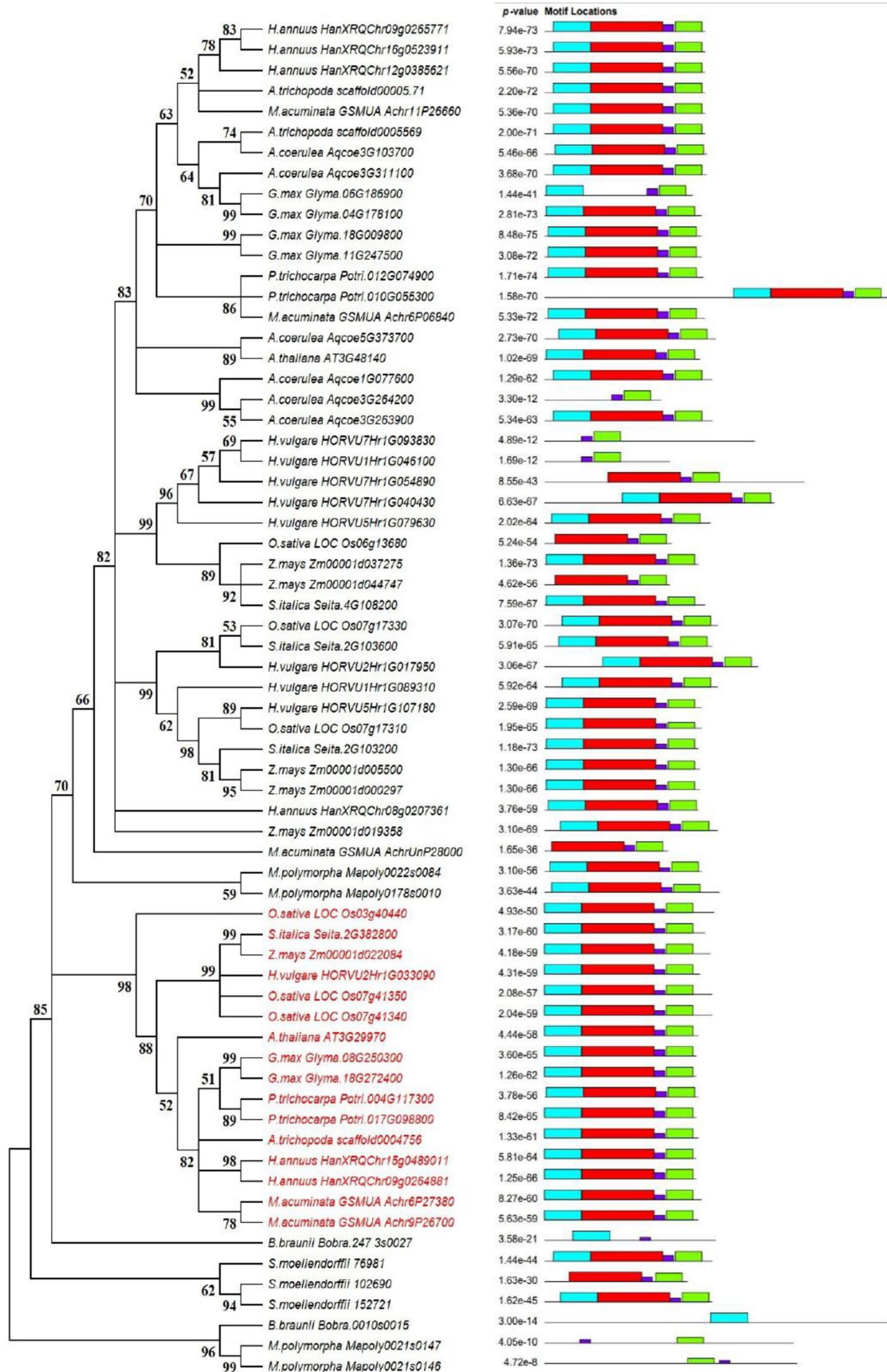
The phylogenetic tree shows the clustering of B12D proteins of the 11 flowering plant species. In contrast, the B12D proteins from the lower Viridiplantae branch from the main clusters include the B12D proteins from higher plants with 85% support except for the 2 *M. polymorpha* B12D proteins



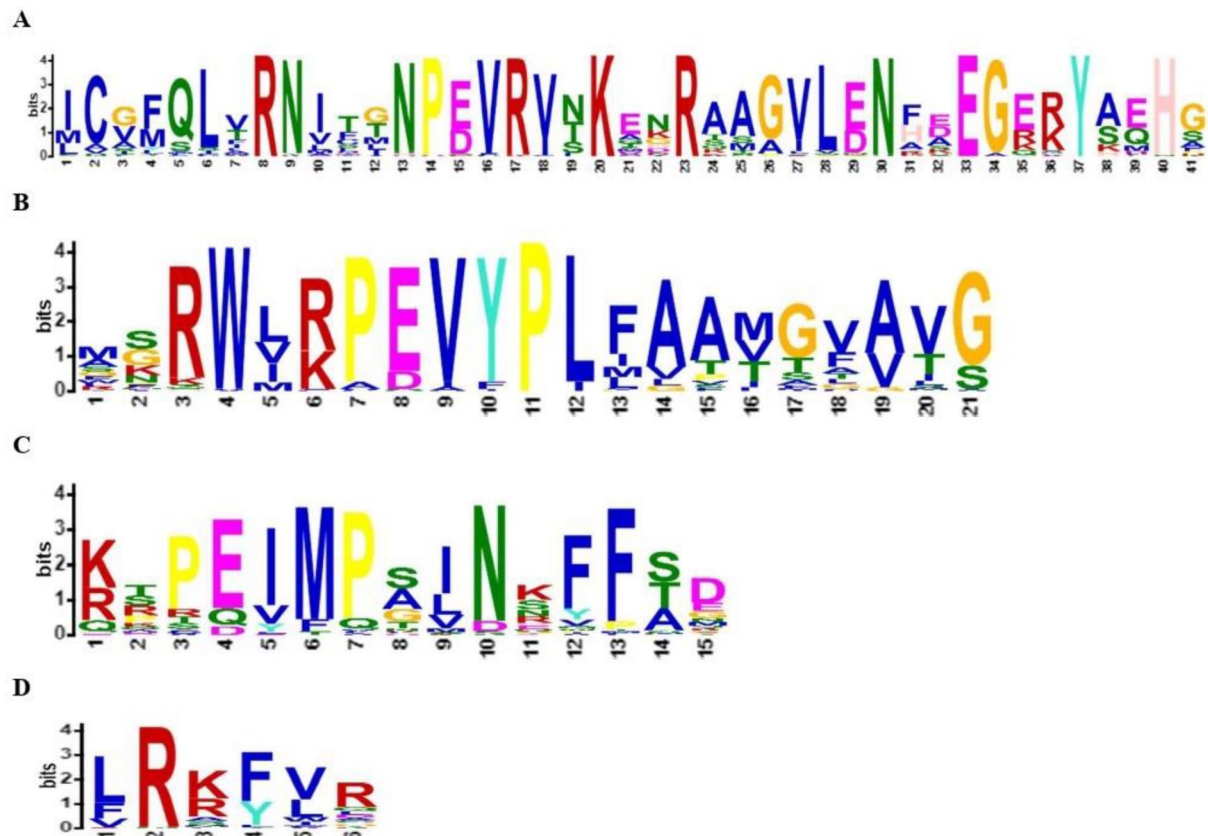
**Figure 1.** Exon-intron structure for 66 *B12D* genes from 14 Viridiplantae species visualized by Gene Structure Display Server. The CDS and genomic sequences for each gene were retrieved from the Phytozome database.

(Mapoly0178s0010 and Mapoly0022s0084) which clustered with flowering plants cluster (Figure 2). The tree shows 2 main clusters: the first cluster branches from a node supported by

98% and includes the 16 *B12D* proteins that aligned together in the first group of the multiple sequence alignment (shown in red in the tree). These 16 proteins separate into 2 subclusters



**Figure 2.** (A) Phylogenetic tree of 66 B12D proteins from 14 Viridiplantae species constructed by minimum evolution method with interior branch tests of 1000 replicates using MEGA X software. The tree was built based on B12D domain sequences and (B) conserved motifs in the 66 B12D proteins predicted by MEME server.



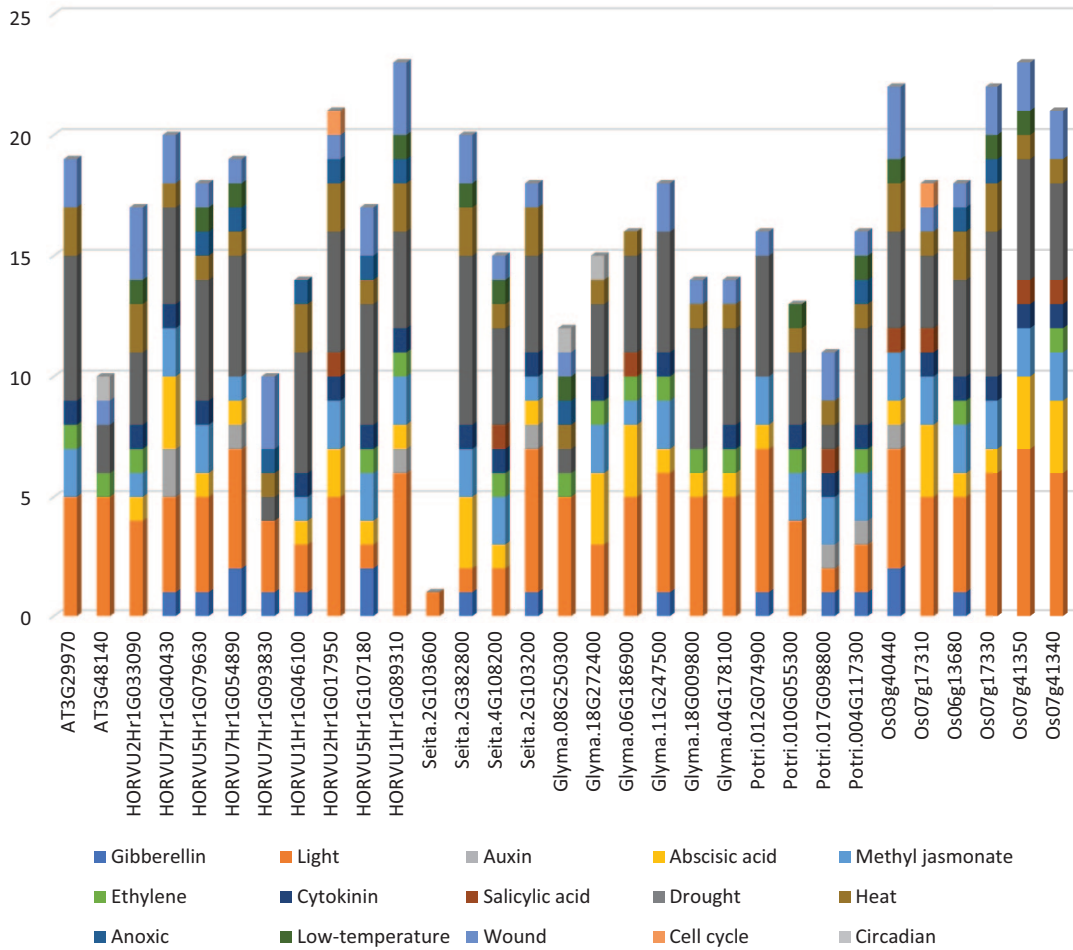
**Figure 3.** Sequence logo for the conserved motifs in the 66 B12D proteins from 14 Viridiplantae species predicted by MEME server.

with 88% support: 1 for monocot proteins and the other for dicot proteins except the 2 proteins (Achr6P27380 and Achr9P26700) of the monocot: *M. acuminata* and the scaffold00047.156 protein of the Amborellales; *A. trichopoda* which clustered with dicots. However, B12D protein #AT3G29970 from *A. thaliana* separated from the dicots subcluster with 52% support. These 16 proteins include 2 B12D proteins of each of *M. acuminata*, *H. annuus*, *P. trichocarpa*, and *G. max*; single protein of each of *A. trichopoda*, *A. thaliana*, *H. vulgare*, *Z. mays*, and *S. italica*; and 3 B12D proteins of *O. sativa*. However, B12D protein #Os03g40440 from *O. sativa* forms a separate branch from this cluster with 98% support value. Notably, these cluster of 16 B12D proteins does not include the B12D protein of *A. coerulea* which is the only exception of 11 flowering plants that is not represented in this cluster.

The second cluster is separated into multiple subclusters from a node supported with 82%. However, B12D protein from *M. acuminata* separates from this subcluster from a node supported with 66%. The first and second subclusters each branches from a node supported by 99%. These 2 subclusters include 18 B12D proteins from monocots which each includes at least 1 B12D protein from the 4 species: *H. vulgare*, *Z. mays*, *S. italica*, and *O. sativa*. The third subcluster branches from a node supported with 83% and includes 13 B12D proteins from the 5 dicot species beside *A. trichopoda* and *M. acuminata*.

#### Cis-elements in the putative promoter of B12D genes

The screening of *cis*-elements in the promoters for selected B12D genes reveals that the 1500bp putative promoter for B12D genes includes binding sites for various stress-responsive factors and plant hormones. These include the regulatory elements for gibberellic acid (GARE-motif, CARE, P-box, and TATC-box), the regulatory elements for abscisic acid (ABRE, ABRE2, ABRE4a, and ABRE4), the regulatory elements for auxin (AuxRR-core and TGA-element), the regulatory elements for methyl jasmonate (CGTCA-motif, JERE, and TGACG-motif), the regulatory element for ethylene (ERE), the regulatory element for cytokinin (as-1), the regulatory element for salicylic acid (TCA-element) the regulatory elements for wound (WRE3, WUN-motif, and W box); the regulatory elements for drought (DRE1, DRE core, MYB, MBS, and MYC), the regulatory element for anaerobic induction (ARE), the regulatory element for low-temperature (LTR), the regulatory elements for light (chs-CMA1a, box S, TCCC-motif, TCA-motif, Sp1, MRE, LS7, L-box, GT1-motif, GATA-motif, G-box, I-box, Box II, Box 4, ATC-motif, ATCT-motif, AT1-motif, AE-box, ACE, A-box, 3-AF3 binding site), the regulatory element for anoxic (GC-motif), the regulatory element for heat (STRE), the regulatory element for heat shock (CCAAT-box), the RY-element for seed-specific regulation (the CAT-box, which is specific for meristem, cell cycle, re2f-1, and



**Figure 4.** Cis-elements in the putative promoter of 31 *B12D* genes from 6 Viridiplantae species; *H. vulgare*, *S. italica*, *O. sativa*, *G. max*, *P. trichocarpa*, and *A. thaliana*; predicted by PlantCARE database.

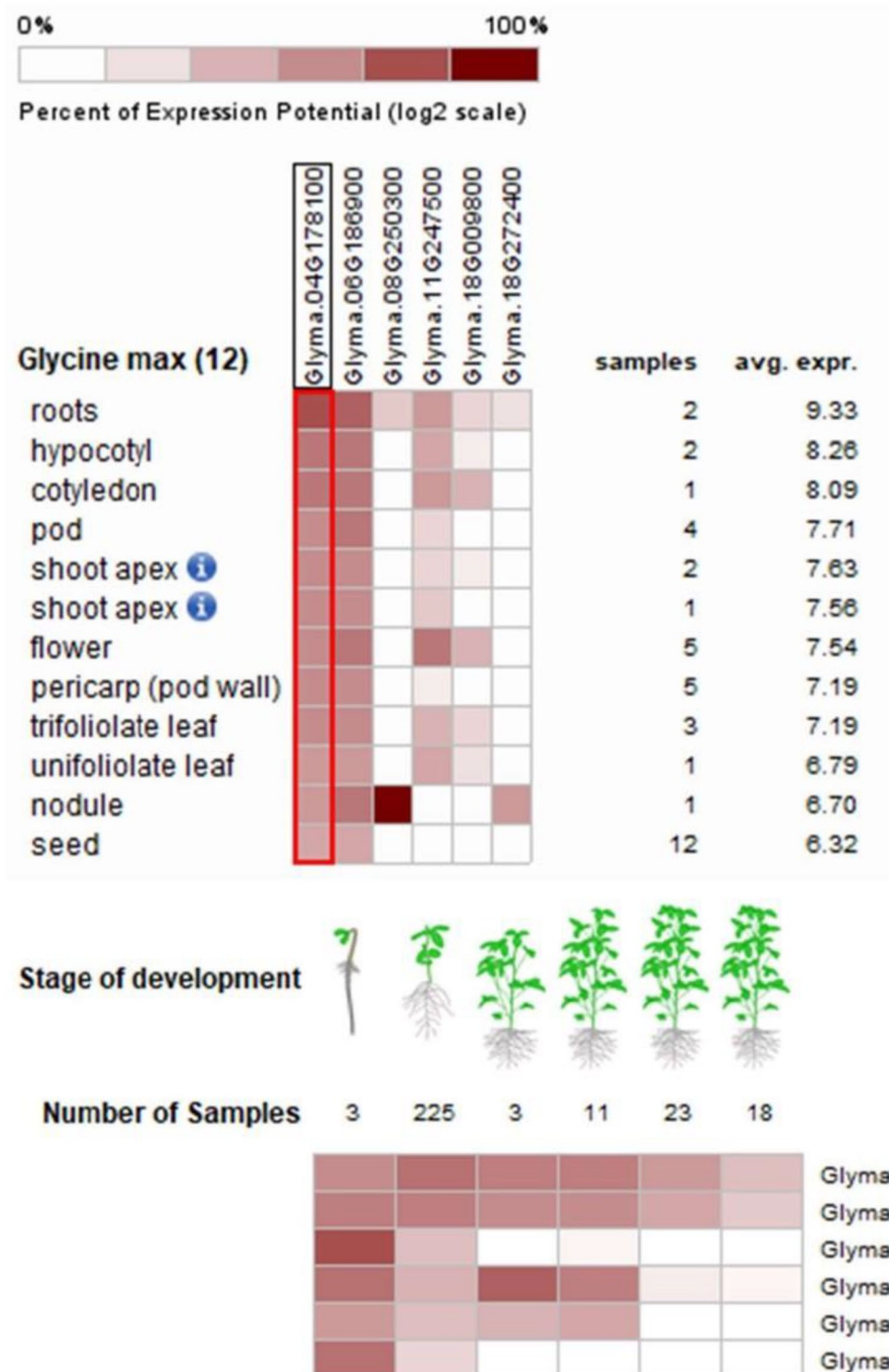
circadian). Other regulatory elements such as CAAT-box, TATA-box, AP-1, AAGAA-motif, AC-I, ACTCATCCT sequence, unnamed\_1, unnamed\_2, unnamed\_4, unnamed\_6, unnamed\_16, CTAG-motif, and F-box that are not assigned to any function in the PlantCARE database were found in some screened promoters (Supplemental Table S2). Some regulatory elements such as ABRE, ARE, AT~TATA-box, CAAT-box, CGTCA-motif, G-Box, MYB, MYC, STRE, TATA-box, as-1, and unnamed\_4 are found in most analyzed promoters. Moreover, the elements related to light response and drought appear to be extensively present in all analyzed promoters (Figure 4). However, other elements such as re2f-1, circadian, chs-CMA1a, Unnamed\_16, O<sub>2</sub>-site, regulatory elements for low temperature, auxin, salicylic acid, and gibberellic acid were found only in a few promoters (Supplemental Table S2).

#### Digital expression of *B12D* genes

Digital expression was analyzed for 6 *B12D* genes from *G. max* and 9 *B12D* genes from *H. vulgare*. Two genes of *G. max*, Glyma.06G186900 and Glyma.04G178100, were highly expressed in all plant tissues at all developmental stages. One

gene (Glyma.11G247500) was expressed in moderate levels in all tissues except nodule and seed and at a moderate to high level at different developmental stages. One gene (Glyma.18G009800) was expressed at a low level in cotyledon, flower, root, shoot, and leaf and at a high level in the early and middle developmental stages. Two genes (Glyma.18G272400 and Glyma.08G250300) were expressed only on nodules and roots and only during the early developmental stages (Figure 5).

Different expression patterns are seen with *H. vulgare* genes, in which the gene HORVU7Hr1G40430 is highly expressed in all plant tissues at all developmental stages. In contrast, HORVU1Hr1G89310 and HORVU2Hr1G17950 are expressed from moderate to high at all developmental stages in all plant tissues except shoot apex and spike. Three genes, HORVU7Hr1G93830, HORVU2Hr1G33090, and HORVU7Hr1G54890, show low expression at early developmental stages and high expression at late stages, but different expression patterns at the different tissues. Two other genes, HORVU1Hr1G46100 and HORVU5Hr1G79630, show low expression at early developmental stages and moderate expression at late stages and are expressed highly at lodicule, rachis, and palea while showing no expression to low expression in other tissues (Figure 6).



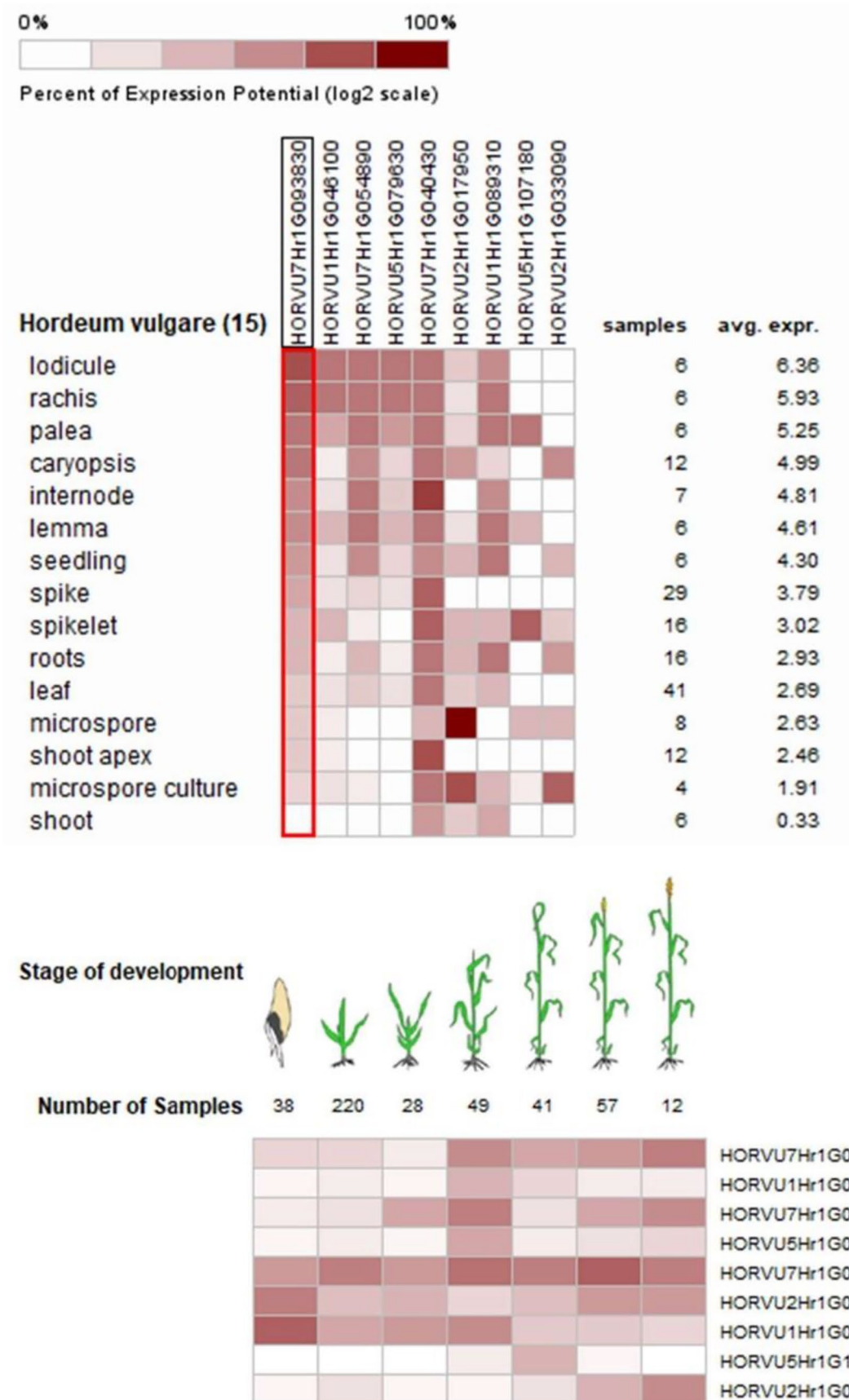
**Figure 5.** Digital expression analysis of 6 *B12D* genes of *G. max* conducted by Geneinvestigator v3 software. Gene accession numbers are illustrated in the diagrams: (A) expression patterns in various plant tissues and (B) expression patterns in the developmental stages; cotyledon, trifoliolate, flowering, pod fill, seeding, maturation.

### Gene duplication analysis

*B12D* gene duplicates were analyzed for selection pressure, and only 4 gene pairs were found on the same chromosomal region (Supplemental Table S1).  $K_a/K_s$  values for the analyzed gene

duplicates from *O. sativa* and *M. polymorpha* were all less than 1, indicating that these 3 duplicates were evolved by negative selection, limiting the function of the genes after duplication. For *A. coerulea* duplicate,  $K_a$  and  $K_s$  values are 0, so the  $K_a/K_s$





**Figure 6.** Digital expression analysis of 9 *B12D* genes in *H. vulgare* conducted by Genevestigator v3 software. Gene accession numbers are illustrated in the diagrams: (A) expression patterns in various plant tissues and (B) expression patterns in the developmental stages; germination, tillering, booting, heading, flowering, spikelet, and ripening.

ratio cannot be calculated. The divergence time for the *O. sativa* gene duplicates is 49.4 and 52.9 million years ago, and for the *M. polymorpha* gene duplicate is 169.5 million years ago (Supplemental Table S3).

## Discussion

The B12D family has transmembrane proteins containing the B12D domain ranging in length from 80 to 98 aa found in plants, animals, and fungi.<sup>5,7</sup> Our results show that B12D proteins appear conserved among the Viridiplantae in protein length, amino acid sequence, domain architecture, exon number, transmembrane helices number, and motifs. However, the number of B12D members appears different in the selected 14 species. This variation in gene copy numbers among species occurs during speciation which contributes to the increasing in genome size. In plants, gene copy number divergence between species is found to be caused by evolutionary adaptation to environmental stress.<sup>31</sup> Similarly, the number of B12D proteins in vertebrate species is different which there are at least 2 B12D members in vertebrates.<sup>32</sup> However, there are 3 copies in human of NUDFA4, the MLRQ subunit of mitochondrial NADH-ubiquinone reductase complex I, a member of B12D family.<sup>33</sup>

Digital expression analysis shows different expression patterns for the paralogous genes from the same species in higher plants, as shown in *G. max* and *H. vulgare*. The phylogenetic tree and expression patterns of *G. max* B12D genes show that each 2 paralogous are clustered together in the tree and show similar expression patterns. However, expression patterns for *H. vulgare* B12D genes do not reflect their clustering in the phylogenetic tree. Each 3 paralogous *H. vulgare* B12D genes show similar expression patterns at the various developmental stages but show different expression patterns in the various tissues. These expression patterns of B12D family members indicate for the diversification in the function of B12D genes which are differentially expressed in the various developmental stages and the various organs of plant. This functional diversification in the highly conserved genes might occur as solution to diversifying a gene regulated by multiple enhancers with a high transcription activity.<sup>34</sup> The sequence conservation among B12D family member that has different roles is caused by the strong selection that results in divergence in the function faster than in the protein sequences as seen in histones and ribosomal RNA gene families.<sup>35,36</sup>

B12D proteins appear to be DNA binding proteins containing the bZIP motif, which is known to comprise 2 regions: a basic region that is rich in arginine (R), asparagine (N), and lysine (K), and a leucine region that is rich in leucine (L) and isoleucine (I).<sup>37</sup> Similarly, characterization of the wheat bZIP transcription factor family revealed that the conserved bZIP motif has 41 aa in length. It has a similar structure to the bZIP motif conserved across B12D proteins in the present study.<sup>38</sup> The involvement of B12D proteins in the regulation of transcription is confirmed by the predicted biological process of

B12D proteins. From these findings, B12D proteins appear to be membrane-bound transcription factors, and their target genes might be nuclear or mitochondrial as some B12D proteins are localized in the mitochondrial inner membrane.<sup>5,8,33</sup> Interestingly, some transcription factors are embedded in mitochondrial membranes in a dormant state and start to regulate nuclear genes in response to external or internal signals by translocation to the nucleus during plant growth and reproduction under environmental stress.<sup>39,40</sup> However, our knowledge about the target genes of B12D proteins is still very limited.

The 2 motifs on the N and C terminal contain the conserved PxxxP repeat involved in protein-protein interaction.<sup>41</sup> PxxxP motif is found in the mitochondrial ADP/ATP carrier proteins. It is known to allow close-packing of transmembrane helices alternating closing and opening of the carrier on the 2 sides of the inner mitochondrial membrane.<sup>42</sup>

The phylogenetic tree and multiple sequence alignment reveal independent divergence of ancient B12D proteins from the 3 gymnosperm species; *B. braunii*, *M. polymorpha*, and *S. moellendorffii*. The B12D proteins from flowering plants show clear divergence into 2 main clades each includes a member of *A. thaliana* B12D proteins; AT3G29970 or AT3G48140. Clade AT3G29970 includes 16 B12D proteins while clade AT3G48140 includes 41 B12D proteins. Notably, each of the 2 clades; AT3G29970 and AT3G48140 in the tree shows separate clustering for dicots and monocots B12D proteins except for B12D proteins of the monocot, *M. acuminata*, which are clustered with dicot species in AT3G48140 clade. Moreover, each of the 2 clades includes at least single B12D protein from the 11 flowering plant species except *A. coerulea* which all 5 B12D proteins clustered in AT3G48140 clade. Clustering of B12D proteins into 2 clades suggests an independent divergence of these clades from ancient orthologous B12D proteins from lower Viridiplantae. Moreover, the functional association of rice B12D proteins LOC\_Os07g41340 and LOC\_Os07g41350 that belong to AT3G29970 clade shows different protein associations from the LOC\_Os06g13680 protein, confirming that some members of the B12D family have a specific function.

Despite the high conservation of B12D proteins, only 4 pairs of B12D genes from 3 species were found to evolve by duplication. The duplicates from *O. sativa* and *M. polymorpha* resulted after experiencing negative selection pressure that silence one of the duplicates. This finding agrees with the study by He et al<sup>5</sup> shows that 1 of B12D duplicate genes (#LOC\_Os07g17330) is expressed in various rice tissues but its duplicate (#LOC\_Os07g17310) has very low expression, suggesting that some B12D duplicate genes have been silenced during plant evolution.<sup>5</sup>


Cis-regulatory elements responsive to abscisic acid, methyl jasmonate, anaerobic induction, cytokinin, light, drought, and heat were found in most analyzed B12D promoters. However,

the elements involved in the cell cycle, circadian, O<sub>2</sub>-site, auxin, regulatory elements for low temperature, and gibberellic acid are found only in the promoters of a small number of *B12D* genes. This diversity of regulatory elements in promoters of *B12D* genes indicates that some *B12D* genes are regulated by specific signals, suggesting that members of the *B12D* family appear to have different roles during plant growth and stress response. The wide range of stimuli that regulate *B12D* gene expression indicates rapid evolution in the promoter region in comparison with the change in amino acid sequences. This diversification in the promoter of *B12D* genes supports the suggestion that the functional divergence of *B12D* gene family might occur as an adaptation to environmental stress.<sup>31</sup>

## Conclusion

*B12D* family proteins are *B12D* domain-containing proteins include several transmembrane proteins. The number and specific function of each member of *B12D* proteins are not determined. In this study, we retrieved and characterized *B12D* proteins from 14 species of Viridiplantae. A high degree of conservation was observed among the analyzed *B12D* protein from higher plants. Despite this strong conservation, some members of the *B12D* proteins appear to have a different role during plant developmental stages as revealed from gene association analysis, promoter analysis, and digital expression. Furthermore, comprehensive identification of the *B12D* family members through functional proteomics, cellular localization, and protein-protein interactions is needed for cognizing the specific function of each member of the *B12D* family.

## ORCID iD

Zainab M Almutairi  <https://orcid.org/0000-0002-7293-5767>

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES

- Aalen RB, Opsahl-Ferstad HG, Linnestad C, Olsen OA. Transcripts encoding an oleosin and a dormancy-related protein are present in both the aleurone layer and the embryo of developing barley (*Hordeum vulgare* L.) seeds. *Plant J*. 1994;5:385-396.
- Steinum TM, Berner HS, Stacy RAP, Salehian Z, Aalen RB. Differential regulation of the barley (*Hordeum vulgare*) transcripts B22E and B12D in mature aleurone layers. *Physiol Plant*. 1998;102:337-345.
- Zhu T, Budworth P, Han B, et al. Toward elucidating the global gene expression patterns of developing *Arabidopsis*: Parallel analysis of 8300 genes by a high-density oligonucleotide probe array. *Plant Physiol Biochem*. 2001;39:221-242.
- Huang YJ, To KY, Yap MN, Chiang WJ, Suen DF, Chen SC. Cloning and characterization of leaf senescence up-regulated genes in sweet potato. *Physiol Plant*. 2001;113:384-391.
- He D, Zhang H, Yang P. The mitochondrion-located protein OsB12D1 enhances flooding tolerance during seed germination and early seedling growth in rice. *Int J Mol Sci*. 2014;15:13461-13481.
- Walker JE, Arizmendi JM, Dupuis A, et al. Sequences of 20 subunits of NADH: ubiquinone oxidoreductase from bovine heart mitochondria. Application of a novel strategy for sequencing proteins using the polymerase chain reaction. *J Mol Biol*. 1992;226:1051-1072.
- Gabalón T, Rainey D, Huynen MA. Tracing the evolution of a large protein complex in the eukaryotes, NADH: ubiquinone oxidoreductase (complex I). *J Mol Biol*. 2005;348:857-870.
- Senkler J, Senkler M, Eubel H, et al. The mitochondrial complexome of *Arabidopsis thaliana*. *Plant J*. 2017;89:1079-1092.
- McLoughlin F, Arisz SA, Dekker HL, et al. Identification of novel candidate phosphatidic acid-binding proteins involved in the salt-stress response of *Arabidopsis thaliana* roots. *Biochem J*. 2013;450:573-581.
- Reumann S, Quan S, Aung K, et al. In-depth proteome analysis of *Arabidopsis* leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. *Plant Physiol*. 2009;150:125-143.
- Almutairi ZM. Molecular characterization and expression profiling of NADH-ubiquinone reductase complex 1 MLRQ subunit in pennisetum glaucum. *J Plant Biochem Biotechnol*. 2022;31:361-369.
- Aalen RB, Salehian Z, Steinum TM. Stability of barley aleurone transcripts: dependence on protein synthesis, influence of the starchy endosperm and destabilization by GA3. *Physiol Plant*. 2001;112:403-413.
- Yun KY, Park MR, Mohanty B, et al. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biol*. 2010;10:16.
- Bevitori R, Sircar S, de Mello RN, et al. Research article identification of co-expression gene networks controlling rice blast disease during an incompatible reaction. *Genet Mol Res*. 2020;19:3.
- Mohanty B. Promoter architecture and transcriptional regulation of genes upregulated in germination and coleoptile elongation of diverse rice genotypes tolerant to submergence. *Front Genet*. 2021;12:639654.
- Goodstein DM, Shu S, Howson R, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012;40:D1178-D1186.
- Bolser D, Staines DM, Pritchard E, et al. Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. *Methods Mol Biol*. 2016;1374:115-140.
- Nugent T, Jones DT. Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. *PLoS Comput Biol*. 2010;6:e1000714.
- Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings Int Conf Intell Syst Mol Biol*. 1994;2:28-36.
- Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28:1166-1167.
- Cozzetto D, Minnici F, Currant H, Jones DT. FFPred 3: feature-based function prediction for all gene ontology domains. *Sci Rep*. 2016;6:31865.
- Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31:1296-1297.
- Rzhetsky A, Nei M. A simple method for estimating and testing minimum-evolution trees. *Mol Biol Evol*. 1992;9:945-967.
- Stecher G, Tamura K, Kumar S. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol Biol Evol*. 2020;37:1237-1239.
- Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 2002;30:325-327.
- Hruz T, Laule O, Szabo G, et al. Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*? 2008;2008:420747.
- Zhang C, Zhang L, Wang D, et al. Evolutionary history of the glycoside hydrolase 3 (GH3) family based on the sequenced genomes of 48 plants and identification of jasmonic acid-related GH3 proteins in *Solanum tuberosum*. *Int J Mol Sci*. 2018;19:1850.
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299-3302.
- Doerks T, Copley RR, Schultz J, Ponting CP, Bork P. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res*. 2002;12:47-56.
- Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000;290:1151-1155.
- Suryawanshi V, Talke IN, Weber M, et al. Between-species differences in gene copy number are enriched among functions critical for adaptive evolution in *Arabidopsis halleri*. *BMC Genomics*. 2016;17:1034.
- De Grassi A, Lanave C, Saccone C. Genome duplication and gene-family evolution: the case of three OXPHOS gene families. *Gene*. 2008;421:1-6.
- De Grassi A, Caggese C, D'Elia D, Lanave C, Pesole G, Saccone C. Evolution of nuclearly encoded mitochondrial genes in Metazoa. *Gene*. 2005;354:181-188.
- Tunnaclyffe E, Corrigan AM, Chubb JR. Promoter-mediated diversification of transcriptional bursting dynamics following gene duplication. *Proc Natl Acad Sci USA*. 2018;115:8364-8369.
- Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*. 2007;175:477-485.
- Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. The human and mouse replication-dependent histone genes. *Genomics*. 2002;80:487-498.
- Miller M. The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Curr Protein Pept Sci*. 2009;10:244-269.

38. Li X, Feng B, Zhang F, et al. Bioinformatic analyses of subgroup-A members of the wheat bZIP transcription factor family and functional identification of TabZIP174 involved in drought stress response. *Front Plant Sci.* 2016;7:1643.
39. Ng S, Ivanova A, Duncan O, et al. A membrane-bound NAC transcription factor, ANAC017, mediates mitochondrial retrograde signaling in *Arabidopsis*. *Plant Cell.* 2013;25:3450-3471.
40. Liu Y, Li P, Fan L, Wu M. The nuclear transportation routes of membrane-bound transcription factors. *Cell Commun Signal.* 2018;16:12.
41. Teyra J, Huang H, Jain S, et al. Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities. *Structure.* 2017;25:1598-1610.e3.
42. Ruprecht JJ, King MS, Zögg T, et al. The molecular mechanism of transport by the mitochondrial ADP/ATP carrier. *Cell.* 2019;176:435-447.e15.