SCIENTIFIC REPORTS

OPEN

Structural and Functional Characterization of the VQ Protein Family and VQ Protein Variants from Soybean

Yuan Zhou[1], Yan Yang[1], Xinjian Zhou[1], Yingjun Chi[1], Baofang Fan[2] & Zhixiang Chen[1,2]

Proteins containing the FxxxVQxhTG or VQ motif interact with WRKY transcription factors. Although VQ proteins have been reported in several plants, knowledge about their structures, functions and evolution is still very limited. Here, we report structural and functional analysis of the VQ protein family from soybean. Like Arabidopsis homologues, soybean VQ proteins bind only Group I and IIc WRKY proteins and a substantial number of their genes are responsive to stress-associated phytohormones. Overexpression of some soybean VQ genes in Arabidopsis had strong effects on plant growth, development, disease resistance and heat tolerance. Phylogenetic analysis, sequence alignment and site-directed mutagenesis revealed that the region immediately upstream of the FxxxVQxhTG motif also affects binding to WRKY proteins. Consistent with a larger WRKY-binding VQ domain, soybean VQ22 protein from cultivated soybean contains a 4-amino acid deletion in the region preceding its VQ motif that completely abolishes its binding to WRKY proteins. By contrast, the 4-amino acid deletion is absent in the VQ22 protein from wild soybean species (*Glycine soja*). Overexpression of wild soybean *VQ22* in cultivated soybean inhibited growth, particularly after cold treatment. Thus, the mutation of soybean VQ22 is associated with advantageous phenotypes and may have been positively selected during evolution.

As sessile organisms, plants are constantly exposed to a wide range of environmental conditions, favorable or unfavorable, and rely on complex molecular, biochemical and cellular processes to adapt or survive. Differential gene transcription through concerted action of transcription factors is central to the adaptive processes of plants. While many families of plant transcription factors have homologs in other organisms, some are plant-specific and, therefore, may act as regulators of biological processes unique to plants. WRKY proteins are a large super-family of sequence-specific DNA-binding transcription factors found almost exclusively in plants[1]. WRKY proteins contain the highly conserved WRKY domain, which contain the almost invariant WRKYGQK sequence at the N-terminus followed by a C2H2 or C2HC zinc-finger motif[1]. According to the number and sequence of the conserved WRKY zinc-finger motifs, WRKY proteins were initially classified into three groups[1]. The first group contains two C2H2 zinc-finger motifs, whereas the second group contains one C2H2 zinc-finger motif, and the third group contains one C2HC zinc-finger motif. More recent analyses have shown that Group II WRKY proteins can be further divided into five subgroups (IIa, IIb, IIc, IId, and IIe)[1,2]. Since their initial identification more than 20 years ago, plant WRKY transcription factors have been subjected to extensive analysis for their biological functions. These studies have established plant WRKY transcription factors as critical regulators of plant responses to a wide spectrum of biotic and abiotic stresses[3–11]. Some plant WRKY proteins are also involved in the regulation of plant growth and development such as trichome[12] and seed development[13], germination[14] and senescence[15,16]. Therefore, plant WRKY transcription factors are important regulators of diverse biological processes in plants.

Like many other regulatory proteins, WRKY transcription factors rarely act alone but rather often function as complexes with their interacting partners for regulation of important properties such as DNA-binding and transcription-regulating activities, stabilities and subcellular localization[17]. Studies over the last decade have

[1]Department of Horticulture, Zijingang Campus, 866 Yuhangtang Road, Zhejiang University, Hangzhou, 310058, China. [2]Department of Botany and Plant Pathology, 915 W. State Street, Purdue University, West Lafayette, IN 47907, USA. Correspondence and requests for materials should be addressed to Z.C. (email: zhixiang@purdue.edu)

identified a substantial number of WRKY-interacting proteins with roles in signaling, transcription, chromatin remodeling and other cellular processes[17]. Among these WRKY-interacting partners are proteins containing a conserved FxxxVQxLTG or VQ motif[18]. VQ proteins are encoded by a fairly large family with 25, 34 and 39 members in *Physcomitrella patens*, Arabidopsis and rice, respectively[19,20]. Using yeast two-hybrid assays, we have previously shown that Arabidopsis VQ proteins specifically interact with the C-terminal WRKY domain of Group I WRKY proteins and the single WRKY domain of Group IIc WRKY proteins[18]. A number of VQ proteins have also been subjected to molecular and genetic analysis for their biological functions. We have previously discovered that two VQ proteins from *Arabidopsis thaliana*, SIB1(AtVQ23) and SIB2 (AtVQ16), act as co-activators of AtWRKY33 in plant defense against necrotrophic pathogens by specifically recognizing the C-terminal WRKY domain and stimulating the DNA-binding activity of WRKY33[10]. Intriguingly, a number of other Arabidopsis VQ proteins including AtVQ21/MKS1, AtVQ22, AtVQ12 and AtVQ29 function as negative regulators of plant defense against necrotrophic pathogens[21–24]. The opposite roles of different VQ proteins in plant defense suggest the complex nature of plant defense involving interacting pathways that require tight regulation and modulation not only for effective defense against different types of microbial pathogens but also for balancing defense with growth. Other VQ proteins regulate plant responses to abiotic stress. AtVQ15 is a Calmodulin-binding protein and acts as a negative regulator in response to osmotic stress[25]. AtVQ9 is a AtWRKY8-interacting protein but its mutation has an opposite effect that that of AtWRKY8 on salt tolerance, suggesting that the VQ protein may function as a transcriptional repressor that antagonizes AtWRKY8 during plant response to salt stress[26]. A number of plant VQ proteins also regulate plant growth and development. Arabidopsis AtVQ14/IKU1 and its interacting WRKY10/MINI3 are both important regulators of endosperm development and seed size[27]. AtVQ29 negatively regulates seedling photomorphogenesis[28] whereas AtVQ8 may regulate chloroplast development[18]. The molecular and genetic analysis of these Arabidopsis VQ genes indicates their important roles in diverse biological processes in plants.

Although significant progresses have been made in the studies of VQ proteins, our understanding of this important family of transcription regulator is still very limited. Gene encoding VQ proteins have been identified in crop plants including rice[20], Chinese cabbage[29] and grapevine[30] in which the interactions with VviWRKYs were predicted[31], but there have been no molecular and genetic studies that rigorously interrogate the biological functions of the VQ genes in crop plants. Even in Arabidopsis, there have been only a few studies that provide in-depth analysis of the structures and functions of its VQ genes and gene products. As a result, there remain important gaps in our knowledge about the structure-function relationship and the evolution of plant VQ proteins. In the present study, we have analyzed the entire family of 74 structurally divergent VQ proteins from soybean. We demonstrate that soybean VQ proteins bind only Group I and Group IIc WRKY proteins but they differ greatly in binding specificity. Using phylogenetic analysis, sequence alignment and site-directed mutagenesis, we further revealed that the region immediately upstream of the short FxxxVQxhTG motifs also affects binding to WRKY proteins, thereby expanding the functional domain of VQ proteins for binding to WRKY proteins. Through genome-wide analysis, we also identified a VQ variant in cultivated soybean with a 4-amino acid deletion in the region preceding its FxxxVQxhTG motif that completely abolishes its activity to bind WRKY proteins. The 4-amino acid deletion apparently occurred recently as it is absent in the VQ protein from wild soybean species (*G. soja*). Molecular genetic analysis suggests that the loss-of-function mutation of the soybean VQ protein may have been positively selected during evolution because it is associated with advantageous phenotypes in growth.

Results

Identification and structural analysis of soybean VQ protein family.
Identification of the soybean VQ gene family with 74 members through searching the Phytozome Database has recently been reported[32]. However, in Table 1 of the reported study, the names and Phytozome Identifiers (IDs) of a majority of the 74 soybean VQ proteins were mismatched with their corresponding VQ motif sequences (Supplemented Table 1). As a result, it is unclear whether one can match individually the reported structural and expression properties with the 74 soybean VQ genes and their protein products. Using the same database, we also identified 74 genes encoding proteins containing the FxxxVQxxhTG motif in soybean (Fig. 1). As found in Arabidopsis, a majority of soybean VQ genes are also intronless and encode relatively small proteins with fewer than 300 aa residues. Thus, the VQ gene family from soybean is substantially large from those of other plants such as Arabidopsis (34 members) and rice (39 members). This is likely due to the fact that soybean is a paleopolyploid that has undergone at least two rounds of large-scale duplication at approximately 14 and 42 million years ago[33]. Indeed, alignment of amino acid residues and phylogenetic tree construction revealed that a majority of the soybean VQ proteins have one or more close homologs (Fig. 2). Among these close VQ homologs, some (e.g. GmVQ22/GmVQ23 and GmVQ39/GmVQ40/GmVQ41) are tandem repeats. GmVQ36 and GmVQ37 on chromosome IIX are closely related to GmVQ65 and GmVQ64 on chromosome XV, respectively and the synonymous distances between the two gene pairs are very similar (Fig. 2). Likewise, GmVQ16, GmVQ17, GmVQ18 and GmVQ19 on chromosome V are close homologs of GmVQ32, GmVQ33, GmVQ34 and GmVQ35 on chromosome IIX, respectively and the distances between the four gene pairs are also very similar. These observations suggest that a large number of close VQ homologs in soybean resulted from duplication of chromosome regions or even whole chromosomes (polyploids).

All 74 soybean VQ proteins contain the conserved FxxxVQxxhTG motif. Among the 74 VQ protein, 56 contain the FxxxVQxxLTG motif, 16 contain FxxxVQxxFTG motif and 2 contain the FxxxVQxxVTG motif (Figs 1 and 2). All these slightly varied VQ motifs are also found in VQ proteins from Arabidopsis[18], indicating that they are conserved across plant species. A notable discovery from the sequence alignment of the VQ motifs is a 4-amino acid deletion in GmVQ22 in a region immediately upstream of the conserved FxxxVQxxFTG motif (Figs 1 and 2). As will be described later, deletion of the 4 amino acid residues in GmVQ22 affects its binding to WRKY proteins, suggesting that the functional VQ domains for binding of WRKY proteins may be larger than

the highly conserved FxxxVQxxFTG motif. The phylogenetic analysis of the 74 soybean VQ proteins is consistent with this possibility. As shown in Fig. 2, the phylogenetic tree consists of a number of clades, which provide further information about the structural relationship among the VQ proteins. First, we observed that the soybean VQ proteins with FxxxVQxxLTG motif are classified into one branch of closely related clades, whereas those VQ proteins with the FxxxVQxxFTG or FxxxVQxxVTG are grouped into the other branch of clades (Fig. 2). This observation supports that the FxxxVQxxhTG motif is the most important determinant of the phylogenetic relationship of the VQ protein family. Second, within both branches of the phylogenetic tree, the amino acid sequences corresponding to those deleted in GmVQ22 can be divided into several types that are largely clade-specific, suggesting that subfamilies of VQ proteins are determined largely by the sequences immediately upstream of the conserved FxxxVQxxhTG motif (Fig. 2). Third, in the second branch of the phylogenetic tree, 14 VQ proteins with the FxxxVQxxFTG motif contain a TTL/VL upstream submit, whereas two VQ proteins containing the FxxxVQxxVTG motif have a TTFI upstream submotif (Fig. 2). Thus, the loss of an aromatic amino acid in the FxxxVQxxVTG motif of GmVQ6 and GmVQ60 appears to be associated with the complementary addition of an aromatic acid residue in the upstream submotif (Fig. 2).

Interacting WRKY partners of GmVQ proteins. Arabidopsis VQ proteins interact specifically with the C-terminal WRKY domain of Group I WRKY proteins and the single WRKY domain of Group IIc WRKY proteins[18]. Using yeast two-hybrid (Y2H) assays, we examined the partnership between soybean WRKY and VQ proteins. We selected 25 soybean VQ proteins from different clades of the phylogenetic tree and fused their corresponding genes with the activation domain (AD) of the Gal4 transcription factor in the Y2H prey vector pAD-Gal4 (Fig. 3). We also fused the DNA fragments for the C-terminal WRKY domain of 9 soybean Group I WRKY proteins and the single WRKY domain of 11 Group IIc WRKY proteins with the DNA-binding domain (BD) of the Gal4 transcription factor in the Y2H bait vector pBD-Gal4 (Fig. 3). As negative controls, the DNA fragments for the WRKY domains of three Group IIb, IIe and III WRKY proteins were also cloned into pBD-Gal4 (Fig. 3). The fused pAD and pBD recombinant vectors were then co-transformed into yeast cells and tested for *LacZ* reporter gene expression through assays of β-galactosidase activity. As shown in Fig. 3, 21 of the 25 soybean VQ proteins tested interacted with some of the Group I and Group IIc WRKY proteins, albeit with varying intensities based on the β-galactosidase activity. By contrast, none of the 25 VQ proteins interacted with the WRKY domains of Group IIb, IIe or III WRKY proteins (Fig. 3). These results indicated that like Arabidopsis homologs, soybean VQ proteins also appear to interact only with the C-terminal WRKY domain of Group I WRKY proteins and the single WRKY domain of the Group IIc WRKY proteins. Among the 25 tested soybean VQ proteins, some including GmVQ7, GmVQ16, GmVQ47 and GmVQ62 and GmVQ66 interacted with almost all 20 tested Group I and Group IIc WRKY proteins (Fig. 3). A majority of the other tested soybean VQ proteins, however, interacted only with a selected number of Group I and Group IIc WRKY proteins (Fig. 3). A few VQ proteins including GmVQ8, GmVQ20 and GmVQ28, on the other hand, interacted with only a very few Group I or Group IIc WRKY proteins (Fig. 3). Four tested VQ proteins (GmVQ9, GmVQ13, GmVQ59 and GmVQ74) failed to interact with any of the 20 tested Group I or Group IIc WRKY proteins (Fig. 3). Thus, soybean VQ proteins display a substantial diversity in their binding specificity to WRKY proteins.

Expression analysis of soybean VQ genes. In Arabidopsis, a large percentage of WRKY and VQ genes are responsive to biotic and abiotic stresses, supporting that the predominant roles of the two gene families in plant stress responses[4,18]. Induction of plant stress responses is associated with signaling of a number of phytohormones including salicylic acid (SA), abscisic acid (ABA), jasmonic acid (JA) and ethylene (ET). To investigate possible roles of soybean VQ genes in plant stress responses, therefore, we examined their expression in response to SA, ABA, ET and JA using quantitative real-time PCR (qRT-PCR). First, we analyzed in both root and leaf tissues the SA-responsiveness of the 67 soybean VQ genes that were initially identified based on the Wm82.a1.v1 annotation version of the soybean genome. In roots, 58 out of the 67 VQ genes examined displayed >2-fold induction at least at one of the three time points after SA treatment (Fig. 4), indicating that a vast majority of soybean VQ genes are responsive to SA in roots. In leave, 40 out of the 67 analyzed GmVQ genes (~60%) displayed >2-fold induction at least at one of the three time points after SA treatment (Fig. 4). We also analyzed about 30 soybean VQ genes selected from all clades of the phylogenetic tree for responses to ABA, ET and JA and observed that approximately 20 of them were induced by >2-fold by at least one of the three phytohormones (Fig. 4). Interestingly, a number of soybean VQ genes including *GmVQ9*, *GmVQ16*, *GmVQ43*, *GmVQ47*, *GmVQ54*, *GmVQ63* and *GmVQ64* were induced at high levels by all four phytohormones. The high responsiveness of these VQ genes to distinct types of stress-related phytohormones suggest their broad roles in plant stress responses.

Functional analysis of soybean VQ genes in Arabidopsis. As an approach for functional analysis of the soybean *VQ* genes, we generated transgenic Arabidopsis plants that constitutively overexpressed soybean *VQ* genes. The coding sequences of soybean Gm*VQ* genes including those highly responsive to stress-related phytohormones were subcloned behind the *CaMV 35S* promoter in a binary plant transformation vector and transformed into *Arabidopsis* plants. qRT-PCR was used to identify F0 transformant plants with elevated levels of transcripts for the transformed VQ gene. Due to the large number of transgenic lines for a large number of soybean VQ genes, our analysis of the transgenic plants was carried out in transgenic F1 lines with increased VQ gene expression and focused primarily on alterations in growth, development and disease resistance. Generally speaking, overexpression of a majority of the soybean VQ genes had no effects on growth or development of transgenic plants. However, altered growth was observed in transgenic plants overexpressing some of the soybean VQ genes. For example, transgenic plants with elevated transcript levels for Gm*VQ19* or Gm*VQ27* had narrower leaves than those of untransformed control plants (Fig. 5A). Transgenic plants overexpressing Gm*VQ43 and*

```
GmVQ1:  ANNAARQQPQPQVYNISKNDFRDIVQQLTGSPSQDPPPRPPHNPPKPQSM      GmVQ38: STLHSPTTPNTTFVQADPSNFRAVVQKLTGASDDPSAPKLPLTLPSRLAA
GmVQ2:  NVGRQQQQPQPQVYNISKNEFRDIVQKLTGSPSQDPQSKPPQNNNSPKPQ      GmVQ39: GGRQQQQQPQPQVYNISKNDFRDIVQQLTGSPSQSQDHPQSKPPQNQNNS
GmVQ3:  APVAPPAPTPIRVYKVDAINFRDLVQQLTGAPEFKPDQEHHQLFQSVAVA      GmVQ40: RQQQQQPQPQPQVYNISKNDFRDIVQQLTGSPSQSQDHPQSKPPQNQNNS
GmVQ4:  PKIRIIHIFAPEIIKTDVENFRELVQKLTGRPSGENLKCCNKKKETRDHH      GmVQ41: GGRQQQQQPQPQVYNISKNDFRDIVQQLTGSPSQSQDHPQSKPPQNQNNS
GmVQ5:  KQYALPSESQPQVYNISKNDFRDMVQKLTGSPGHTPQPKPVATSRLHRLR      GmVQ42: QFNTLIKVLRPKVYITDSSSFKKLVQELTGNGSSNNTLSPPPLEPNMVEN
GmVQ6:  RKSRASKRSQTTFITADPANFRQMVQQVTGVRFGGAGQPIAPAAGVVKPE      GmVQ43: RRSRASKRTPTTLLNANPTNFRALVQQFTGCPRTTMSTLGVHKGPITLNF
GmVQ7:  HPVKVVYISNPMKIKTSASEFRALVQELTGQDAESPPDPTRFHGLIHPDS      GmVQ44: PKIRIIHIFAPEIIKTDVENFRELVQKLTGKPSGENLKYFCNNKKNKAIA
GmVQ8:  HSPVIVHLKSPKVIHVRPEEFMSLVQQLTRNPVSAAFAAPKMDKSSTVVE      GmVQ45: HHPVIIYTHSPKVIHTQPKDFMSLVQKLTGLSRSGEEKEEGGNQSPQQLS
GmVQ9:  TVKITYISNPVLVRACDASEFRSVVQQLTGKDTNKKVETPRKEYNSTLMQ      GmVQ46: INPIIIYTESPKIIHTKAKDFMALVQRLTGRSSSTNDNLSTASLPQEGSE
GmVQ10: VQAQQNLQHQPPVYNINKNDFRDVVQKLTGSPAHDRISTPPPIQPPKPPS      GmVQ47: AAVAPPAPTPIRVYKVDAINFRDLVQQLTGAPEFKPADQEQHQLFPSVAP
GmVQ11: KRSRASRRAPTTVLTTDTNNFRSMVQEFTGIPAPPFSPSFSRRIPLRPNP      GmVQ48: KRSRASRRAPPTVLTTDDTTNFRAMVQEFTGIPAPPFTSSSSSFPRTRLDL
GmVQ12: KNVKVTYISSPMKVKTSASNFRALVQELTGQASNVAEMFVEADYYYGVHH      GmVQ49: LRSDSTNPYPTTFVQADTSTFKQVVQMLTGSSETTKPQLQQQQQQQQQDL
GmVQ13: REPVIIYSVSPKVIHVTPGDFMDVVQRLTGASSGEYPPAGRGDVSPAARL      GmVQ50: IPRSDSNPYPTTFVQADTSTFKQVVQMLTGSSDTTKQASQDPPPSSRNFN
GmVQ14: RQPIIIYTVSPKVIHTTPSDFMNLVQRLTGSSSSSVVVPNNNNTHVDPF       GmVQ51: RRSRASRRTPTTLLNTDTTNFRAMVQQFTGGPSAPFASNPSLAPQVLPNL
GmVQ15: APVAPMPPTPIKVYKVDAINFRDVVQQLTGAPEHESQQQHLQIKIARAES      GmVQ52: RRSRASRRTPTTLLNTDTTNFRAMVQQFTGGPSAPFASNPSLAPQVLPNL
GmVQ16: RRSRTSKSTPITLLKANTSNFRALVQQFTGCPTTTMSLAIHKGPITLNF       GmVQ53: TRSESGNPYPTTFVQADTNSFKQVVQMLTGSTQTAKQASASASPKPNSEP
GmVQ17: TSGPKIVHIETRYVETDAINFRDLVQQFTGKNSSTTNWVGNGASFSSVAC      GmVQ54: KDIKVTYISSPVKVKTSASNFRALVQELTGQYSNVAETSMPMEEENGHS
GmVQ18: KRSRASRRAPTTVLTTDTTNFRAMVQEFTGIPAPPFTSSSFPRTRLDLFA      GmVQ55: MEAPKIVQIETRYVQTDAANFRDVVQSLTGKNSSTDWIGQNAESSYSSKR
GmVQ19: RKPVIIYTVSPKVLHVTVSDFMNVVQRLTGPSSGAEPNLRAGDVSPAARL      GmVQ56: TRSEPANSYPTTFVQADTSFKQVVQMLTGSSETAKQAAAAAAAASSSSS
GmVQ20: YGAASDCKPLTTFVQTNSDAFREVVQRLTGPSEASAAKEGDAKKVATMKR      GmVQ57: IPRSDSNPYPTTFVQADTSTFKQVVQMLTGSSDTTKQASQDPQPPQPPSS
GmVQ21: QVPNNQVQHQPPVYNINKNDFRDVVQKLTGSPAHDRISTPPPIQPPKPPS      GmVQ58: RRSRASRRTPTTLLNTDTSNFRAMVQQFTGAPSAPDMFGSPVPVNPNGLM
GmVQ22: KRSRSSRRAPTT----DTNNFRSMVQEFTGIPAPPFSSSFSRRIPLCPNP      GmVQ59: RQPVIIYTVSPKVIHTTPSDFMSLVQRLTGSSSSSSSSSTNKVSTEPLNG
GmVQ23: NESRASRRAPTTVLTTDTNNFRSMVQEFTGIPAPPFSSSFSRRIPLRPNP      GmVQ60: KRTRASRRAPTTVLTTDTSNFRAMVQEFTGIPAPPFSASSSYSRRLDLLT
GmVQ24: RQPIIIYTVSPKVIHTTPSDFMNLVQRLTGSSSSSSAEVVMSNNNNTTHV      GmVQ61: TRSEPANPYPTTFVQADTSSFKQVVQMLTGSSETAKQAAAGGSSKPSNSI
GmVQ25: REPVIIYSVSPKVIHVTPGDFMDVVQRLTGASIGEEPPSGRGDVSPAARL      GmVQ62: RRSRASKKTPTTLLNANTTNFRALVQQFTGCHSTTMPTLGVQKGPITLNF
GmVQ26: RSDSTTNPYPTTFVQADTSTFKHVVQMLTGSSETTNPQQQLQDPVQPQSS      GmVQ63: KEIKVTYISSPVKVKTSASNFRALVQELTGQYSNVAETSIPMEEDNGHSE
GmVQ27: STLHSPTTPNTTFVQANPSNFRAVVQKLTGASDDPSAHKLPLTLPTRLAA      GmVQ64: RAPVKVVIINTQYVETDATSFKSVVQKLTGKDSDILDAKVQRKRHDHKGG
GmVQ28: CHPVIIYTHSPKVIHTQPKDFMSLVQKLTGLSRSDEKEDEDGGNPTPQQP      GmVQ65: PKIRIIHVYAPEIIKTDAANFKELVQRLTGKPKEEEGRGKSKSKTALTKD
GmVQ29: TRSESANPYPTTFVQADTTSFKQVVQMLTGSTQTAKQASASASEPAKPHT      GmVQ66: HHVAPMPPTPVKVYKVDAINFREVVQQLTGAPKHKPQQLLQINIARAESS
GmVQ30: GAAASDCKPLTTFVQTNSDAFREVVQRLTGPSEASAAKEGDATKVATVKR      GmVQ67: KRSRASRRAPTTVLTTDTTNFRAMVQEFTGIPAPPFTSSSSSFPRTRLD
GmVQ31: RPPVIIYTHPPKVIHTHARNFMELVQKLTGLYRTDPEDNGGGEEVLKPPT      GmVQ68: QFNTLIKILRPKVYITDSSSFKKLVQELTGNGSPTTLSPPPLEPNMDENF
GmVQ32: RRSRASKSTPITLLKANTSNFRALVQQFTGCPTTTAMSLAIHKGPVTLNF      GmVQ69: RKSRASKRSQTTFITADPANFRQMVQQVTGVRFGGAPIAPAAVVKPEPQR
GmVQ33: TSGPKIVHIETRYVETDAINFRDLVQQLTGKNSSTTNWDVGNGALFSSLP      GmVQ70: HPVKVVYISNPMKIKTSASEFRALVQELTGQDAESPPDPSRFHGHHPDSS
GmVQ34: KRSRASRRAPTTVLTTDTTNFRAMVQEFTGIPAQPFTSSSFPRTRTQQHH      GmVQ71: HSPVIVHLKSPKVIHVRPEEFMSLVQQLTGNPVSAAAVAANSPKMEKSST
GmVQ35: RKPVIIYAVSPKVLHVPAGDFMNVVQRLTGPSSGDVSPAARLASIERTSP      GmVQ72: TVKITYISSPVLVRAYDASEFRSVVQQLTGNDSNSNKKVERPHKGYSSTL
GmVQ36: PKIRIIHVYTPEIIKTDAANFRELVQRLTGKPKEEGTGGSKSETALTKEP      GmVQ73: RRSRASRRTPTTLLNTDTTNFRAMVQQFTGGPSAPYASNPSLAPHVLPNL
GmVQ37: RAPVKIVIINTQYVETDATSFKSVVQKLTGKDSDNLDAKVQRKRHDHKSG      GmVQ74: INPIIIYTESPKVIHTKAKDFMALVQRLTGRSSTNDNLSTTASLPQEGSE
```

Figure 1. VQ domain sequence of soybean VQ proteins. The highly conserved residues in the FxxxVQxhTG motif are shown in red.

GmVQ62 flowered substantially earlier than control plants (Fig. 5B). Overexpression of GmVQ37 had no significant effect on growth and flowering time but reduced seed setting, particularly on the first 10 or so siliques (Fig. 5C). The reduced seed setting appeared to be associated with shorter filaments of the stamen in transgenic plant flowers, which would reduce pollination.

A number of Arabidopsis VQ proteins have been show to play roles in plant resistance to necrotrophic pathogens[21–24]. Therefore, we examined effects of overexpression of soybean VQ genes on response of transgenic Arabidopsis lines to *B. cinerea*. Using drop-inoculation assays of detached leaves, we observed that the lesion sizes of transgenic lines overexpressing *GmVQ35* and *GmVQ47* were substantially larger than those of control wild-type plants (Fig. 6A), suggesting that the two soybean VQ proteins act as negative regulators of plant defense against the necrotrophic pathogen. Transgenic Arabidopsis plants overexpressing *GmVQ47* also displayed increased sensitivity to heat (Fig. 6B). Thus, expression of a number of soybean VQ genes had various effects on plant growth, development and responses to both biotic and abiotic stresses.

Identification of VQ22 allelic variants in soybean. Annotated genome sequence from 'Williams 82', a cultivated soybean line, revealed a 12 base pair (bp) deletion in *GmVQ22*, resulting in removal of 4-amino acid residues (V146, L147, T148, and T149) in the region immediately upstream of the FxxxVQxxFTG motif (VQ22ΔV146_T149) (Figs 1 and 2). Sequencing of the genes from 'Williams 82' confirmed the 12-bp deletion in *GmVQ22*. Interestingly, sequencing of the same gene from a wild soybean (*G. soja*) line revealed the absence of the 12-bp deletion, suggesting possible allelic polymorphism at the *VQ22* locus in the *Glycine* genus. To study the polymorphism further, we developed two PCR approaches for rapid detection of the polymorphism. We designed a pair of PCR primers for detection of the PCR-amplified fragment length polymorphism (AFLP) between GsVQ22 and GmVQ22ΔV146_T149 (Fig. 7A). We also used the method of PCR Amplification of Specific Alleles (PASA) with PCR primers that specifically amplify either the Gs*VQ22* or Gm*VQ22ΔV146_T149* allele (Fig. 7B)[34]. Both PCR procedures detected the polymorphism reliably and faithfully and were used to investigate their occurrence in a large collection of wild (*G. soja*), semi wild (*G. gracilis*), and cultivated (*G. max*) soybean lines. As shown in Table 1, survey of 26 wild soybean lines revealed that they all contain the GsVQ22 allele without the 12-bp deletion. By contrast, cultivated soybean lines or cultivars all contain the *GmVQ22ΔV146_T149* allele (Table 1). Interestingly, semi wild soybean populations contain both the *VQ22* and *VQ22ΔV146_T149* alleles (Table 1). Among the 20 semi wild soybean lines tested, 16 contain the *VQ22* allele, whereas the remaining 8 lines contain the *VQ22ΔV146_T149* allele (Table 1).

Analysis of the upstream submotif of VQ proteins in interaction with WRKY proteins. Even though the four amino acid residues deleted in VQ22ΔV146_T149 are located outside the FxxxVQxxhTG motif, they are immediately upstream of the highly conserved motif and, therefore, could affect binding to WRKY proteins. To test the possibility, we fused both the GsVQ22 and GmVQ22ΔV146_T149 genes with the AD of
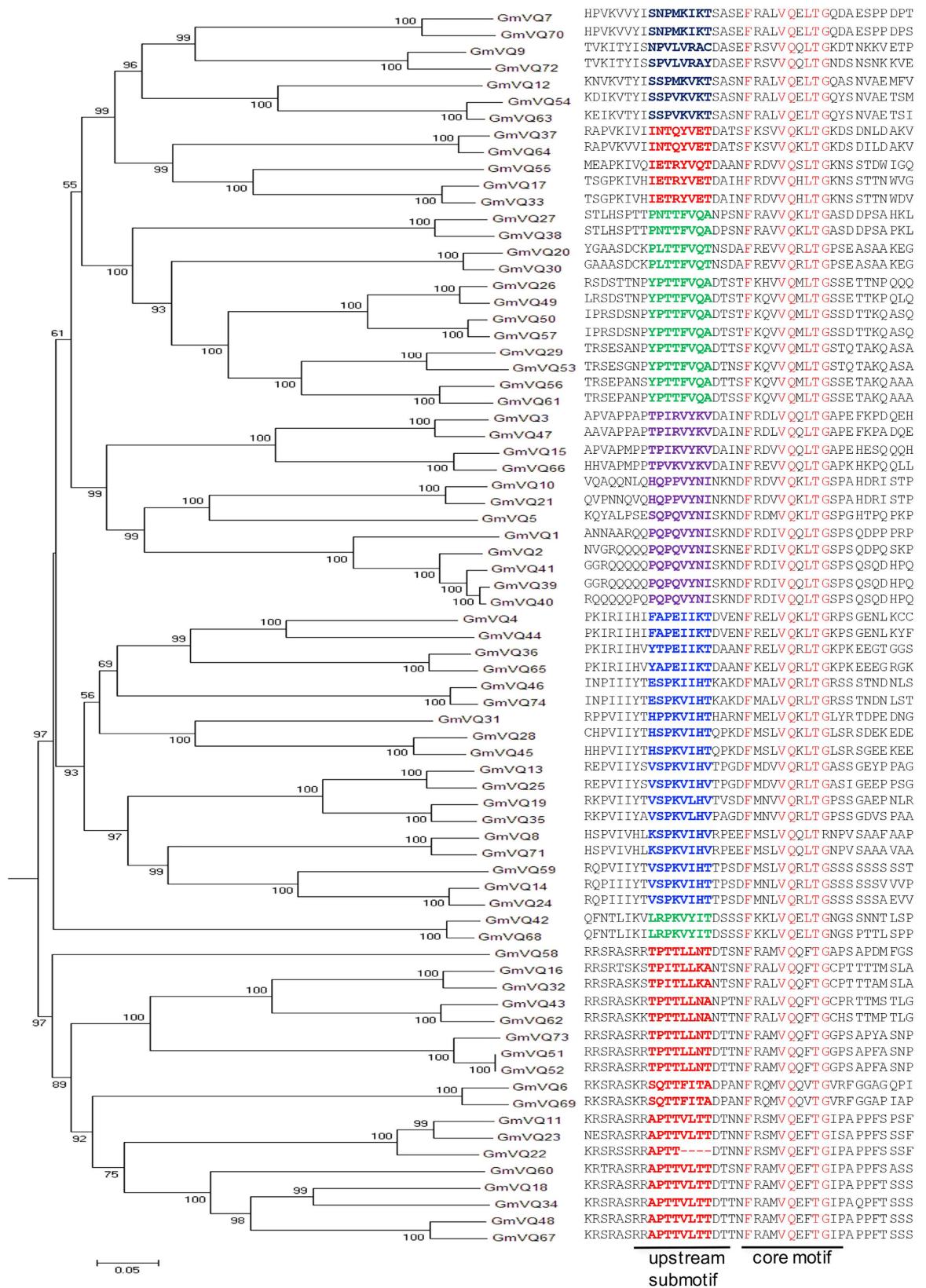
Figure 2. Phylogenetic analysis of soybean VQ proteins. The phylogenetic tree was inferred using the the neighbor-joining method from 74 soybean VQ proteins. Bootstrap values from 1,000 replicates were used to assess the robustness of the tree. VQ domain sequences are also shown for each VQ protein. The highly conserved residues in the FxxxVQxhTG motif are in red. Amino acid residues of the upstream submotifs are shown in different colors for different clades.
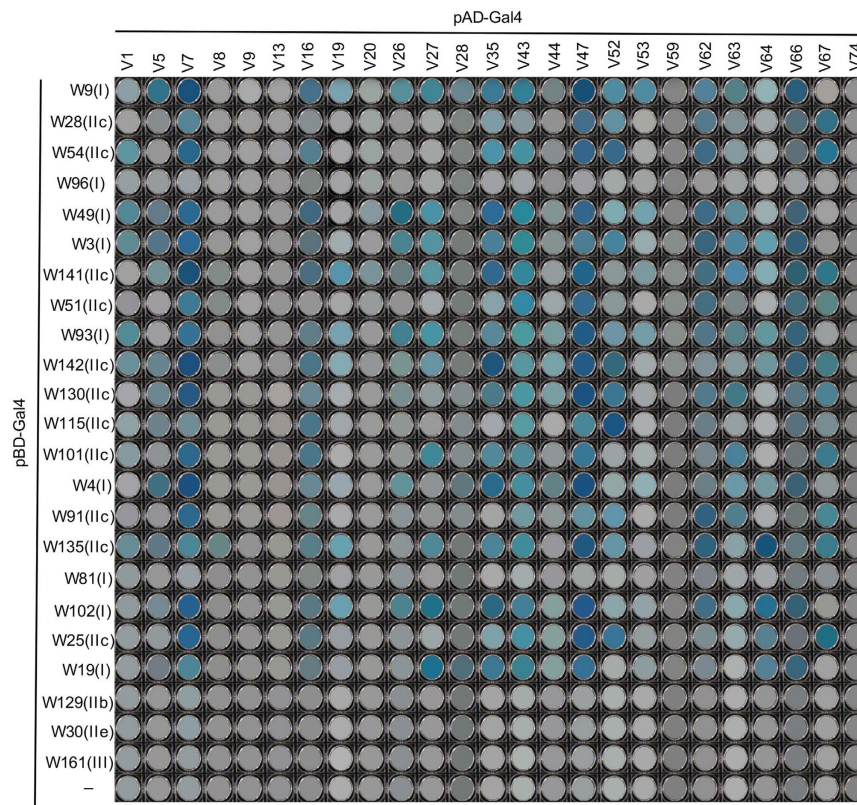
Figure 3. Interaction of GmVQ proteins with GmWRKY proteins in yeast cells. The Gal4 DNA BD-WRKY (W) domain fusion bait vectors were co-transformed with the activation domain (AD)-VQ (V) fusion prey vectors into yeast cells and the transformant cells were assayed for *LacZ* reporter gene expression. The empty pBD prey vector was used as negative control. The specific subfamilies to which the tested WRKY proteins belong to are indicated in parentheses.

the Y2H prey vector pAD-Gal4 and cotransformed into yeast cells with the pBD-Gal4-WRKY fusion constructs from 23 soybean WRKY genes. As shown in Fig. 8, GsVQ22 interacts with 18 of the 20 tested Group I and Group IIc WRKY proteins, albeit with varying intensities based on the β-galactosidase activity. On the other hand, no interaction was detected between GsVQ22 and the three Group IIb, IIe, or III WRKY proteins (Fig. 8). Thus, like other VQ proteins, GsVQ22 interacts only with Group I and Group IIc WRKY proteins. By contrast, when the pAD-GmVQ22ΔV146_T149 fusion vector was cotransformed into yeast cells with the 23 pBD-Gal4-WRKY bait vectors, none of them displayed detectable β-galactosidase activity (Fig. 8), indicating that deletion of the four amino acid residues in GmVQ22ΔV146_T149 completely abolishes its binding to WRKY proteins.

The complete loss of binding to WRKY proteins by GmVQ22ΔV146_T149 suggests a critical role of the upstream submotif of VQ proteins in their recognition of the WRKY domains of Group I and IIc WRKY proteins. However, the loss of WRKY-binding activity in GmVQ22ΔV146_T149 could also be due to the effects of the deletion of four amino acid residues on the folding and conformation of the protein. Indeed, phylogenetic analysis and sequence alignment have revealed that the amino acid residues in the upstream submotif are not highly conserved but largely clade-specific (Fig. 2). These observations suggest that the upstream sequence is not the core motif essential for WRKY-binding but could be required for clade-specific properties of VQ proteins. To distinguish between the two possibilities, we selected six GmVQ proteins (GmVQ7, GmVQ16, GmVQ27, GmVQ35, GmVQ44 and GmVQ47) from different clades in the phylogenetic tree with strong binding to Group I and IIc WRKY proteins and changed a single amino acid residue in the upstream clade-specific submotif using overlapping PCR (Fig. 9A). The mutated genes were fused with the AD of the pAD-Gal4 prey vector and cotransformed into the yeast cells with pBD-Gal4-WRKY vectors for protein interaction assays. As shown in Fig. 9B, change of a single amino acid residue in the upstream submotif of the VQ proteins did not abolish their binding to WRKY proteins. However, their binding specificity to a number of tested WRKY proteins was altered by these single amino acid substitutions. For example, the L76I substitution of GmVQ16 increased its binding to a number of Group I and IIc WRKY proteins (Fig. 9). By contrast, the Y46L substitution in the upstream motif of GmVQ47 resulted reduced binding to a number Group I and IIc WRKY proteins (Fig. 9). Interestingly, the F18I substitution in the upstream motif of GmVq18 increased binding to four Group I and IIc WRKY proteins but reduced binding to two other WRKY proteins. Single amino acid substitutions in the other three GmVQ proteins also alters their binding to WRKY proteins, although the number of affected WRKY proteins was smaller (Fig. 9).

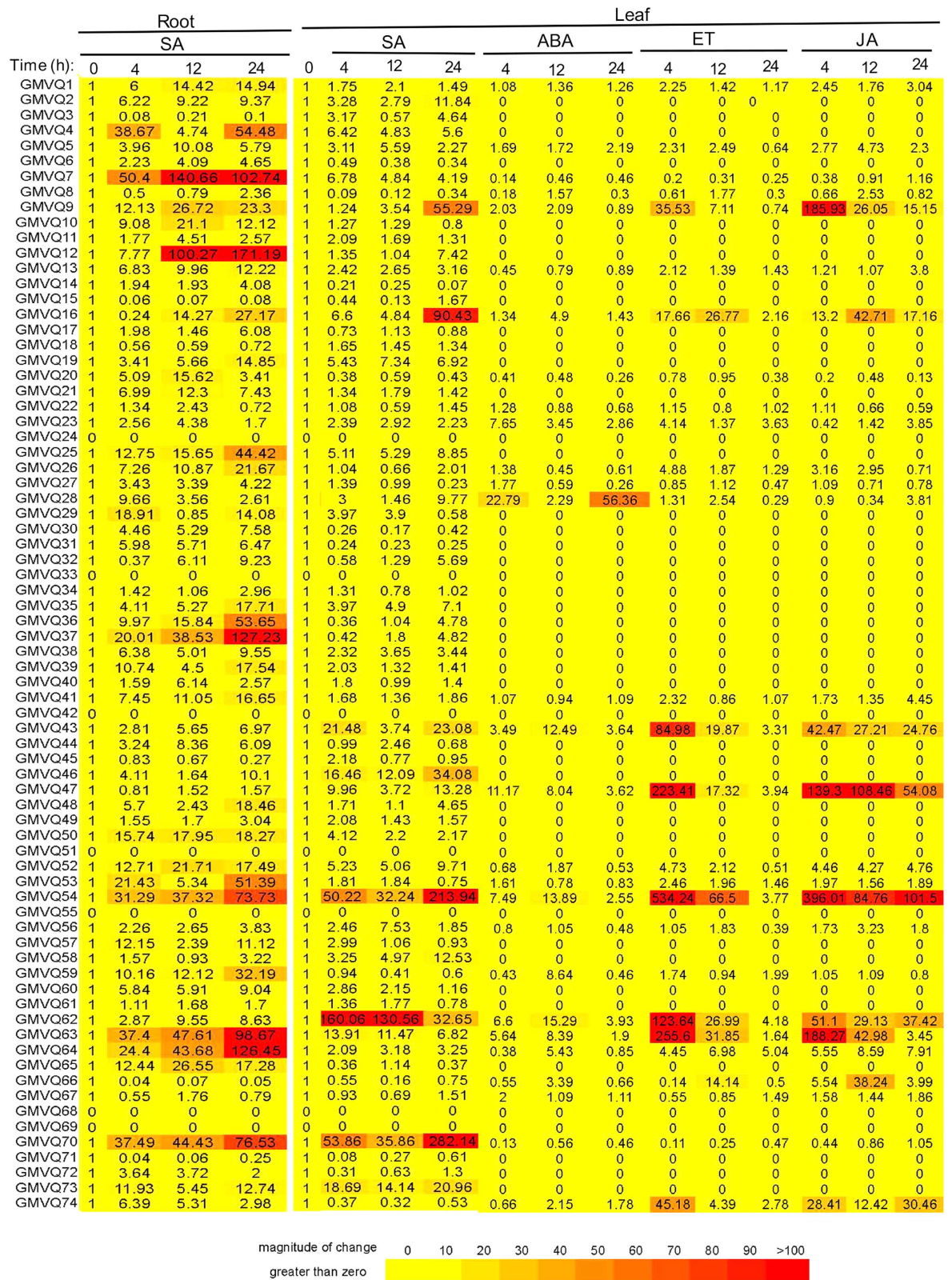|	Root				Leaf													
	SA				SA				ABA			ET			JA		
Time (h):	0	4	12	24	0	4	12	24	4	12	24	4	12	24	4	12	24
GMVQ1	1	6	14.42	14.94	1	1.75	2.1	1.49	1.08	1.36	1.26	2.25	1.42	1.17	2.45	1.76	3.04
GMVQ2	1	6.22	9.22	9.37	1	3.28	2.79	11.84	0	0	0	0	0	0	0	0	0
GMVQ3	1	0.08	0.21	0.1	1	3.17	0.57	4.64	0	0	0	0	0	0	0	0	0
GMVQ4	1	38.67	4.74	54.48	1	6.42	4.83	5.6	0	0	0	0	0	0	0	0	0
GMVQ5	1	3.96	10.08	5.79	1	3.11	5.59	2.27	1.69	1.72	2.19	2.31	2.49	0.64	2.77	4.73	2.3
GMVQ6	1	2.23	4.09	4.65	1	0.49	0.38	0.34	0	0	0	0	0	0	0	0	0
GMVQ7	1	50.4	140.66	102.74	1	6.78	4.84	4.19	0.14	0.46	0.46	0.2	0.31	0.25	0.38	0.91	1.16
GMVQ8	1	0.5	0.79	2.36	1	0.09	0.12	0.34	0.18	1.57	0.3	0.61	1.77	0.3	0.66	2.53	0.82
GMVQ9	1	12.13	26.72	23.3	1	1.24	3.54	55.29	2.03	2.09	0.89	35.53	7.11	0.74	185.93	26.05	15.15
GMVQ10	1	9.08	21.1	12.12	1	1.27	1.29	0.8	0	0	0	0	0	0	0	0	0
GMVQ11	1	1.77	4.51	2.57	1	2.09	1.69	1.31	0	0	0	0	0	0	0	0	0
GMVQ12	1	7.77	100.27	171.19	1	1.35	1.04	7.42	0	0	0	0	0	0	0	0	0
GMVQ13	1	6.83	9.96	12.22	1	2.42	2.65	3.16	0.45	0.79	0.89	2.12	1.39	1.43	1.21	1.07	3.8
GMVQ14	1	1.94	1.93	4.08	1	0.21	0.25	0.07	0	0	0	0	0	0	0	0	0
GMVQ15	1	0.06	0.07	0.08	1	0.44	0.13	1.67	0	0	0	0	0	0	0	0	0
GMVQ16	1	0.24	14.27	27.17	1	6.6	4.84	90.43	1.34	4.9	1.43	17.66	26.77	2.16	13.2	42.71	17.16
GMVQ17	1	1.98	1.46	6.08	1	0.73	1.13	0.88	0	0	0	0	0	0	0	0	0
GMVQ18	1	0.56	0.59	0.72	1	1.65	1.45	1.34	0	0	0	0	0	0	0	0	0
GMVQ19	1	3.41	5.66	14.85	1	5.43	7.34	6.92	0	0	0	0	0	0	0	0	0
GMVQ20	1	5.09	15.62	3.41	1	0.38	0.59	0.43	0.41	0.48	0.26	0.78	0.95	0.38	0.2	0.48	0.13
GMVQ21	1	6.99	12.3	7.43	1	1.34	1.79	1.42	0	0	0	0	0	0	0	0	0
GMVQ22	1	1.34	2.43	0.72	1	1.08	0.59	1.45	1.28	0.88	0.68	1.15	0.8	1.02	1.11	0.66	0.59
GMVQ23	1	2.56	4.38	1.7	1	2.39	2.92	2.23	7.65	3.45	2.86	4.14	1.37	3.63	0.42	1.42	3.85
GMVQ24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ25	1	12.75	15.65	44.42	1	5.11	5.29	8.85	0	0	0	0	0	0	0	0	0
GMVQ26	1	7.26	10.87	21.67	1	1.04	0.66	2.01	1.38	0.45	0.61	4.88	1.87	1.29	3.16	2.95	0.71
GMVQ27	1	3.43	3.39	4.22	1	1.39	0.99	0.23	1.77	0.59	0.26	0.85	1.12	0.47	1.09	0.71	0.78
GMVQ28	1	9.66	3.56	2.61	1	3	1.46	9.77	22.79	2.29	56.36	1.31	2.54	0.29	0.9	0.34	3.81
GMVQ29	1	18.91	0.85	14.08	1	3.97	3.9	0.58	0	0	0	0	0	0	0	0	0
GMVQ30	1	4.46	5.29	7.58	1	0.26	0.17	0.42	0	0	0	0	0	0	0	0	0
GMVQ31	1	5.98	5.71	6.47	1	0.24	0.23	0.25	0	0	0	0	0	0	0	0	0
GMVQ32	1	0.37	6.11	9.23	1	0.58	1.29	5.69	0	0	0	0	0	0	0	0	0
GMVQ33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ34	1	1.42	1.06	2.96	1	1.31	0.78	1.02	0	0	0	0	0	0	0	0	0
GMVQ35	1	4.11	5.27	17.71	1	3.97	4.9	7.1	0	0	0	0	0	0	0	0	0
GMVQ36	1	9.97	15.84	53.65	1	0.36	1.04	4.78	0	0	0	0	0	0	0	0	0
GMVQ37	1	20.01	38.53	127.23	1	0.42	1.8	4.82	0	0	0	0	0	0	0	0	0
GMVQ38	1	6.38	5.01	9.55	1	2.32	3.65	3.44	0	0	0	0	0	0	0	0	0
GMVQ39	1	10.74	4.5	17.54	1	2.03	1.32	1.41	0	0	0	0	0	0	0	0	0
GMVQ40	1	1.59	6.14	2.57	1	1.8	0.99	1.4	0	0	0	0	0	0	0	0	0
GMVQ41	1	7.45	11.05	16.65	1	1.68	1.36	1.86	1.07	0.94	1.09	2.32	0.86	1.07	1.73	1.35	4.45
GMVQ42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ43	1	2.81	5.65	6.97	1	21.48	3.74	23.08	3.49	12.49	3.64	84.98	19.87	3.31	42.47	27.21	24.76
GMVQ44	1	3.24	8.36	6.09	1	0.99	2.46	0.68	0	0	0	0	0	0	0	0	0
GMVQ45	1	0.83	0.67	0.27	1	2.18	0.77	0.95	0	0	0	0	0	0	0	0	0
GMVQ46	1	4.11	1.64	10.1	1	16.46	12.09	34.08	0	0	0	0	0	0	0	0	0
GMVQ47	1	0.81	1.52	1.57	1	9.96	3.72	13.28	11.17	8.04	3.62	223.41	17.32	3.94	139.3	108.46	54.08
GMVQ48	1	5.7	2.43	18.46	1	1.71	1.1	4.65	0	0	0	0	0	0	0	0	0
GMVQ49	1	1.55	1.7	3.04	1	2.08	1.43	1.57	0	0	0	0	0	0	0	0	0
GMVQ50	1	15.74	17.95	18.27	1	4.12	2.2	2.17	0	0	0	0	0	0	0	0	0
GMVQ51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ52	1	12.71	21.71	17.49	1	5.23	5.06	9.71	0.68	1.87	0.53	4.73	2.12	0.51	4.46	4.27	4.76
GMVQ53	1	21.43	5.34	51.39	1	1.81	1.84	0.75	1.61	0.78	0.83	2.46	1.96	1.46	1.97	1.56	1.89
GMVQ54	1	31.29	37.32	73.73	1	50.22	32.24	213.94	7.49	13.89	2.55	534.24	66.5	3.77	396.01	84.76	101.5
GMVQ55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ56	1	2.26	2.65	3.83	1	2.46	7.53	1.85	0.8	1.05	0.48	1.05	1.83	0.39	1.73	3.23	1.8
GMVQ57	1	12.15	2.39	11.12	1	2.99	1.06	0.93	0	0	0	0	0	0	0	0	0
GMVQ58	1	1.57	0.93	3.22	1	3.25	4.97	12.53	0	0	0	0	0	0	0	0	0
GMVQ59	1	10.16	12.12	32.19	1	0.94	0.41	0.6	0.43	8.64	0.46	1.74	0.94	1.99	1.05	1.09	0.8
GMVQ60	1	5.84	5.91	9.04	1	2.86	2.15	1.16	0	0	0	0	0	0	0	0	0
GMVQ61	1	1.11	1.68	1.7	1	1.36	1.77	0.78	0	0	0	0	0	0	0	0	0
GMVQ62	1	2.87	9.55	8.63	1	160.06	130.56	32.65	6.6	15.29	3.93	123.64	26.99	4.18	51.1	29.13	37.42
GMVQ63	1	37.4	47.61	98.67	1	13.91	11.47	6.82	5.64	8.39	1.9	255.6	31.85	1.64	188.27	42.98	3.45
GMVQ64	1	24.4	43.68	126.45	1	2.09	3.18	3.25	0.38	5.43	0.85	4.45	6.98	5.04	5.55	8.59	7.91
GMVQ65	1	12.44	26.55	17.28	1	0.36	1.14	0.37	0	0	0	0	0	0	0	0	0
GMVQ66	1	0.04	0.07	0.05	1	0.55	0.16	0.75	0.55	3.39	0.66	0.14	14.14	0.5	5.54	38.24	3.99
GMVQ67	1	0.55	1.76	0.79	1	0.93	0.69	1.51	2	1.09	1.11	0.55	0.85	1.49	1.58	1.44	1.86
GMVQ68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GMVQ70	1	37.49	44.43	76.53	1	53.86	35.86	282.14	0.13	0.56	0.46	0.11	0.25	0.47	0.44	0.86	1.05
GMVQ71	1	0.04	0.06	0.25	1	0.08	0.27	0.61	0	0	0	0	0	0	0	0	0
GMVQ72	1	3.64	3.72	2	1	0.31	0.63	1.3	0	0	0	0	0	0	0	0	0
GMVQ73	1	11.93	5.45	12.74	1	18.69	14.14	20.96	0	0	0	0	0	0	0	0	0
GMVQ74	1	6.39	5.31	2.98	1	0.37	0.32	0.53	0.66	2.15	1.78	45.18	4.39	2.78	28.41	12.42	30.46

magnitude of change greater than zero: 0 10 20 30 40 50 60 70 80 90 >100

Figure 4. Expression of soybean GmVQ genes in response to different hormones. Roots of 2-week old soybean seedlings (*G. max* cv 'Williams 82') were treated with 1 mM SA. The first trifoliolate leaves of about 2-week-old soybean seedlings were treated with 0.1 mM ABA, 0.5 mM ET and JA and 1 mM SA. Plant tissues were collected from three indeoednent biologcal repeats at indicated time of treatment with the phytohormones for RNA isolation and qRT-PCR analysis of the VQ gene transcripts using gene-specific primers. The numbers in the expression profile are from three technical repeats and are normalized induction fold over untreated control plants. Those VQ genes that were not analyzed for a specific phytohormone were given a number of zero.
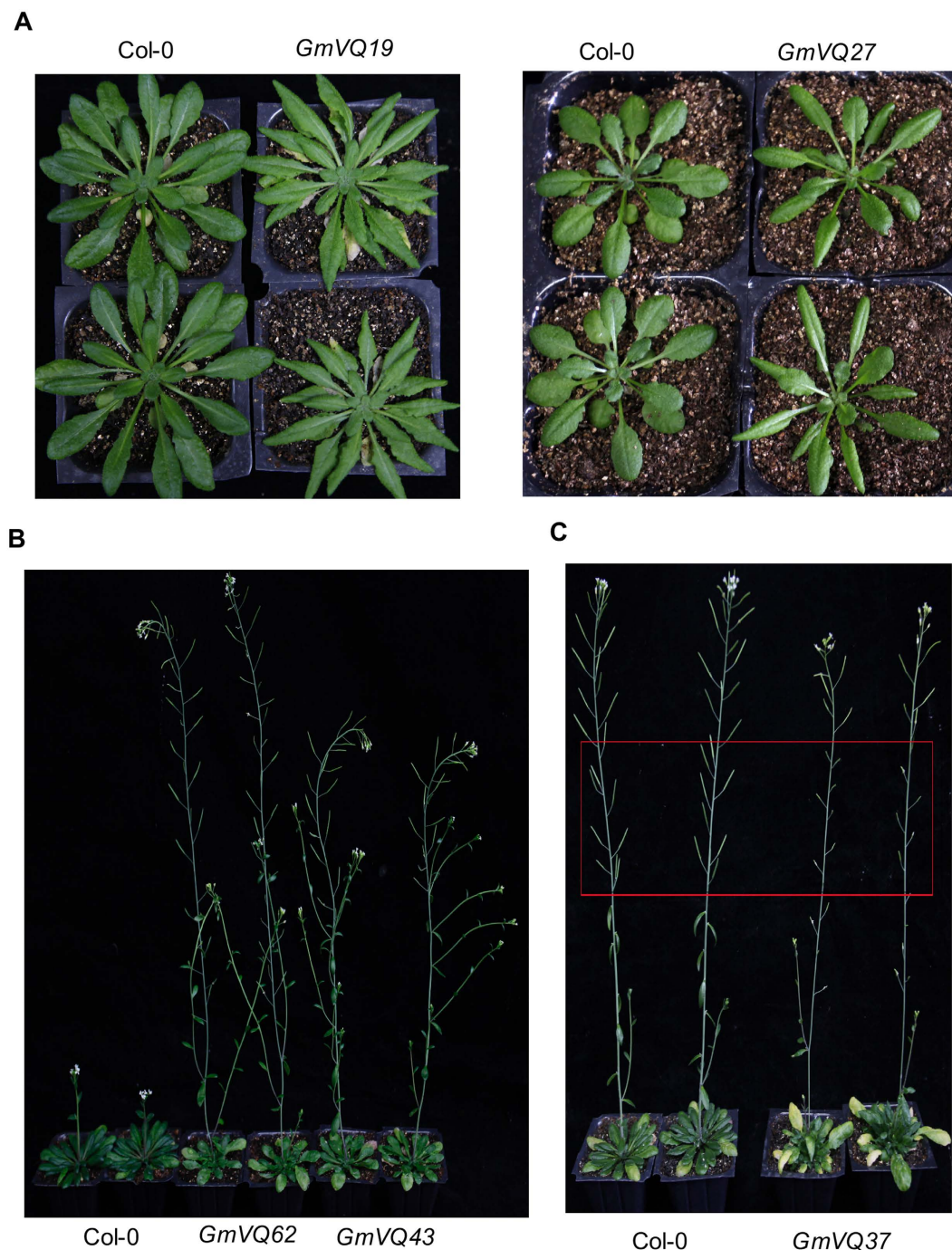
Figure 5. Growth and developmental phenotypes of transgenic VQ-overexpressing *Arabidopsis* plants.
(**A**) Altered leaf growth of transgenic plants overexpressing Gm*VQ19* and *GmVQ27*. The picture of Col-0 wild type and two lines of transgenic overexpression plants for each VQ gene was taken 7 weeks after germination. (**B**) Earlier flowering of transgenic plants overexpressing Gm*VQ62* and *GmVQ43*. The picture of Col-0 wild type and two lines of transgenic overexpression plants for *GmVQ* gene was taken 10 weeks after germination. (**C**) Reduced seed setting of transgenic plants overexpressing *GmVQ37*. The picture of Col-0 wild type and two lines of transgenic plants was taken 12 weeks after germination. The parts of the inflorescences of the transgenic plants with poor silique development and seed setting are indicated by a red box.

Functional analysis of soybean VQ22 variants. Deletion of four amino acid residues in the upstream submotif and abolishment of WRKY-binding activity of GmVQ22 suggest that this is likely a loss-of-function mutation that might be advantageous and positively selected for during domestication. To examine this, we tried to express the wild soybean GsVQ22 in cultivated soybean to determine the effects on the transgenic soybean plants. For this purpose, we placed the *GsVQ22* coding region under the control of the CaMV *35S* promoter in the binary plant transformation vector pFGC5941. Using cotyledonary node method, we transformed the *GsVQ22*
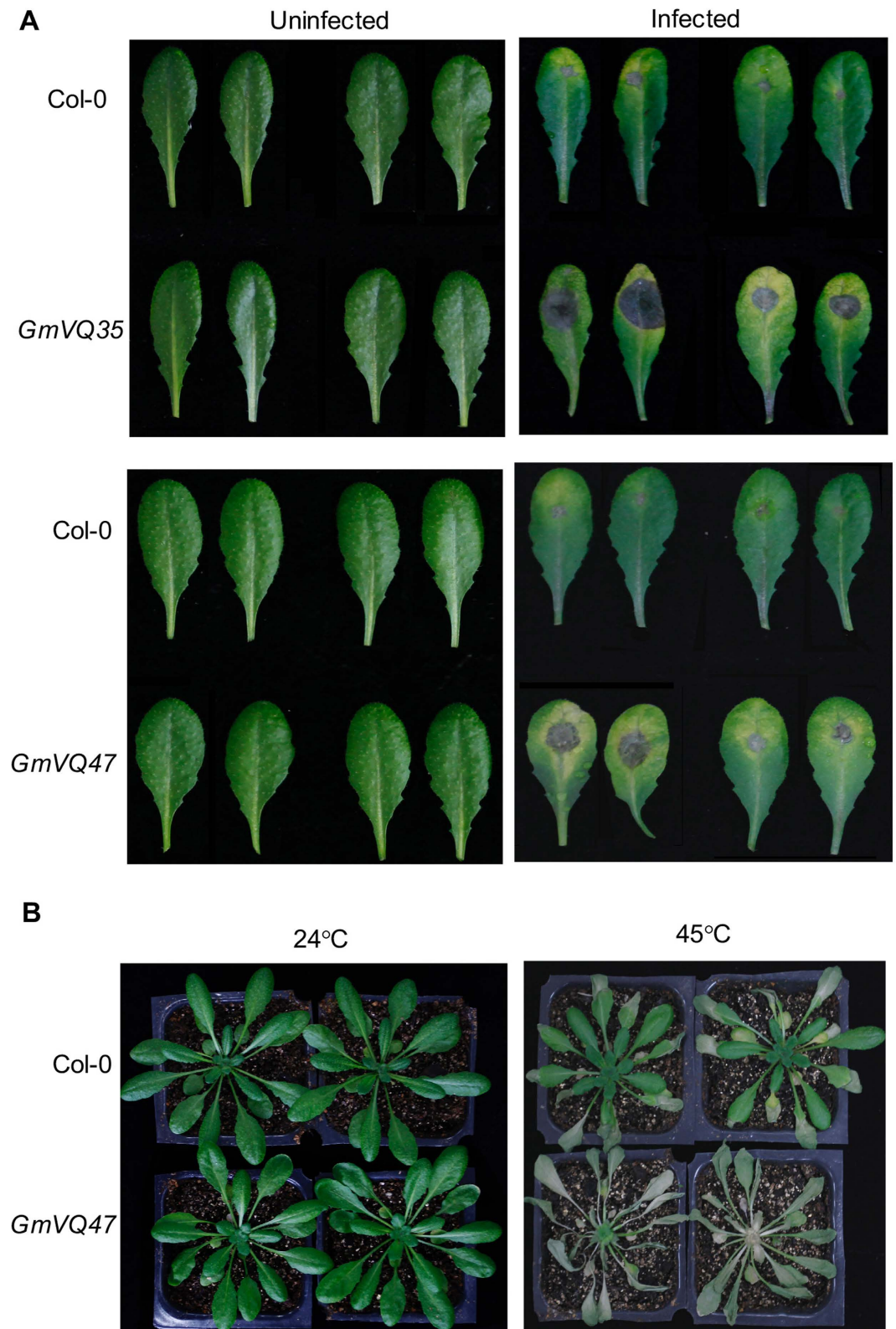
Figure 6. Altered disease resistance and heat tolerance of transgenic Gm*VQ*-overexpressing *Arabidopsis* plants. (**A**) Enhanced susceptibility to *B. cinerea*. Fully expanded leaves of Col-0 wild type and two independent lines of transgenic plants expressing Gm*VQ35* or *GmVQ47* were drop-inoculated with *Botrytis* and the picture was taken at the 4th day post inoculation (dpi). Both uninoculated and inoculated leaves are shown (**B**). Enhanced heat sensitivity. Col-0 wild type and transgenic plants expressing *GmVQ47* were placed in growth chamber 45 °C for 10 hours. The plants were then moved back to a growth room at 25 °C and pictures of representative plants were taken after 4-day recovery.

Figure 7. PCR-based detection of *VQ22* polymorphism in soybean. (**A**) AFLP-based detection of VQ22 and VQ22ΔV146_T149 alleles. Two primers flanking the 12-nt deletion in the VQ22ΔV146_T149 alleles (pVQ22-F and pVQ22-R1) were designed (shown in the diagram in upper panel) and used to detect AFLP in *VQ22* alleles from cultivated (*G. max*) and wild (*G. soja*) soybean. PCR products were separated on a 12% polyacrylamide gel (lower panel) (**B**). PASA-based detection of VQ22 and VQ22ΔV146_T149 alleles. Two pairs of PCR primers that specifically amplify either the *VQ22* (pGmVV22-F and pVQ22-R2) or *VQ22ΔV146_T149* (pGsVQ22-F and pVQ22-R2) allele (shown in the diagram in upper panel) were designed and used to detect the two alleles from cultivated (*G. max*) and wild (*G. soja*) soybean. PCR products were separated on a 0.8% agarose gel (lower panel).

expression construct into the soybean cultivar Heinong (HN) 37 and generated eight independent transgenic lines. However, six of the eight transgenic plants grew very poorly and didn't set seeds. The two remaining lines grew and set seeds largely normally and homozygous F3 progeny from the two independent lines were obtained. At early stages, homozygous F3 progeny from the transgenic *GsVQ22* plants were largely normal when compared with nontransgenic control plants (Supplemental Fig. 1). However, at older stages, the transgenic *GsVQ22* plants displayed a slight but significant reduction in growth. At 4-week age, the average heights of the transgenic plants were about 10–15% less than those of untransformed control plants (Fig. 10). qRT-PCR analysis showed that the *VQ22* transcript levels in the two independent transgenic lines were elevated by 10–20-fold when compared to those in untransformed HN37 plants (Fig. 11). To determine possible roles in stress tolerance, we also examined their responses to cold, heat and drought stresses but failed to observe significantly altered phenotypes. Interestingly, we consistently observed that reduced growth of the transgenic plants was enhanced after cold treatment. In these experiments, we first subjected 12-days old transgenic and control plants to 48-hour cold treatment at 3 °C. After the cold treatment, all the plants were moved back to a growth room at 24 °C for recovery of growth. This cold treatment significantly reduced the growth of nontransgenic control plants as indicated from the ~20% reduction in plant heights after 2-week recovery at 24 °C when compared to those of unstressed plants (Fig. 10). Interestingly, the cold treatment reduced the growth of transgenic *GsVQ22* plants to a larger extent as these plants displayed almost no increase in their height after 2 week recovery at 24 °C after the cold treatment (Fig. 10). As a result, the heights of the cold-treated transgenic *GsVQ22* plants were only about 60% of cold-treated nontransgenic plants after recovery for 2 weeks at 24 °C (Fig. 10). Thus, *GsVQ22* had an inhibitory effect on soybean growth and this inhibitory effect was strengthened by cold treatment.

To understand how the inhibitory effect of GsVQ22 on soybean growth was temperature-sensitive, we analyzed the effects of cold treatment on expression of *VQ22* in both transgenic and non-transgenic control plants. As described earlier, prior to cold treatment, transgenic *GsVQ22* plants contained the levels of *VQ22* transcripts 10–20 times higher than those in nontransgenic plants (Fig. 11). Interestingly, after 48-cold cold treatment, transgenic *GsVQ22* plants had an ~10-fold increase in *VQ22* transcripts, while nontransgenic plants displayed no such increase (Fig. 1). As a result, the *VQ22* transcript levels in the transgenic *GsVQ22* plants were ~200 times higher than those in nontransgenic plants after the cold treatment (Fig. 11). Thus, there was a correlation between the degrees of growth inhibition and the *VQ22* transcript levels. Without cold treatment, the transcript levels of *VQ22* were elevated by 10–20-fold in the transgenic *GsVQ22* plants (Fig. 11), leading to slight but significant inhibition

Line name	species	VQ22 allele[a]	Line name	species	VQ22 allele[a]
PI81762	*G. soja*	VQ22	PI81763	*G. gracilis*	VQ22
PI101404B	*G. soja*	VQ22	PI81771	*G. gracilis*	VQ22
PI342618B	*G. soja*	VQ22	PI81772	*G. gracilis*	VQ22
PI342620A	*G. soja*	VQ22	PI86046	*G. gracilis*	VQ22
PI342622A	*G. soja*	VQ22	PI135590	*G. gracilis*	VQ22
PI406684	*G. soja*	VQ22	PI232987	*G. gracilis*	VQ22
PI407288	*G. soja*	VQ22	PI232989	*G. gracilis*	VQ22
PI407289	*G. soja*	VQ22	PI232992	*G. gracilis*	VQ22
PI407296	*G. soja*	VQ22	PI417138	*G. gracilis*	VQ22
PI407297	*G. soja*	VQ22	PI417139	*G. gracilis*	VQ22
PI407298	*G. soja*	VQ22	PI291309C	*G. gracilis*	VQ22
PI407299	*G. soja*	VQ22	PI326580	*G. gracilis*	VQ22
PI424004A	*G. soja*	VQ22	PI253651C	*G. gracilis*	VQ22ΔV146_T149
PI424004B	*G. soja*	VQ22	PI291275	*G. gracilis*	VQ22ΔV146_T149
PI440913A	*G. soja*	VQ22	PI416762	*G. gracilis*	VQ22ΔV146_T149
PI464890A	*G. soja*	VQ22	PI88788	*G. gracilis*	VQ22ΔV146_T149
PI464891B	*G. soja*	VQ22	Peking	*G. gracilis*	VQ22ΔV146_T149
PI464891C	*G. soja*	VQ22	PI209332	*G. gracilis*	VQ22ΔV146_T149
PI468916	*G. soja*	VQ22	PI90763	*G. gracilis*	VQ22ΔV146_T149
PI479744	*G. soja*	VQ22	PI89772	*G. gracilis*	VQ22ΔV146_T149
PI479746B	*G. soja*	VQ22	Williams 82	*G. max*	VQ22ΔV146_T149
PI479748	*G. soja*	VQ22	Heinong 37	*G. max*	VQ22ΔV146_T149
PI479750	*G. soja*	VQ22	Jack	*G. max*	VQ22ΔV146_T149
PI483460B	*G. soja*	VQ22	F002-4-4	*G. max*	VQ22ΔV146_T149
PI507581	*G. soja*	VQ22	Zhongdou27	*G. max*	VQ22ΔV146_T149
PI522182B	*G. soja*	VQ22	ZDD5957	*G. max*	VQ22ΔV146_T149
PI65388	*G. gracilis*	VQ22	A8901	*G. max*	VQ22ΔV146_T149
PI79593	*G. gracilis*	VQ22	Wupiqingren	*G. max*	VQ22ΔV146_T149
PI79648	*G. gracilis*	VQ22	L21778	*G. max*	VQ22ΔV146_T149
PI79727	*G. gracilis*	VQ22	L6326	*G. max*	VQ22ΔV146_T149

Table 1. Genotyping of the *VQ22* alleles in a wide range of soybean lines. [a]The alleles of the VQ22 locus were determined by both AFLP- and PASA-based procedures. The *VQ22* allele encodes a normal VQ22 protein with an intact VQ motif, whereas the *VQ22ΔV146_T149* allele encodes a VQ22 protein with the deletion of four amino acid residues in the upstream submotif of the VQ22 protein.
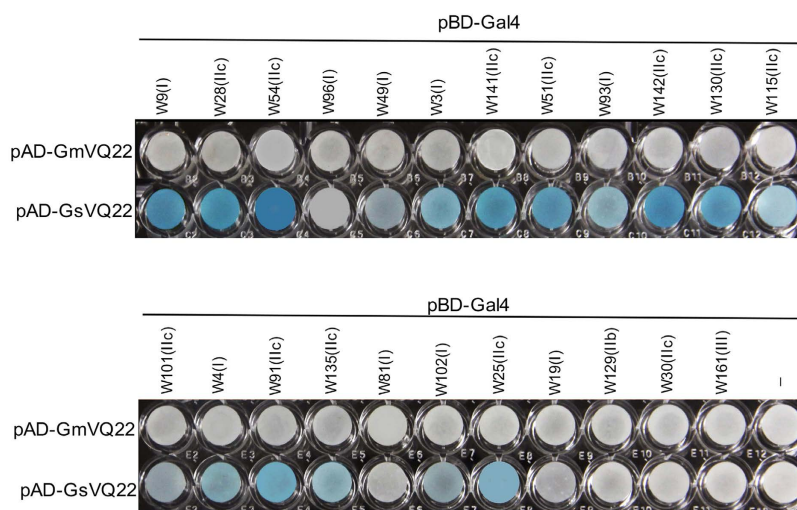
Figure 8. Comparison of GsVQ22 and GmVQ22 for interaction with WRKY protein in yeast cells. The Gal4 DNA BD-WRKY domain fusion bait vectors were contransformed with the activation domain (AD)-VQ fusion prey vectors into yeast cells and the transformant cells were assayed for *LacZ* reporter gene expression. The empty pBD prey vector was used as negative control. The specific subfamilies to which the tested WRKY proteins belong to are indicated in parentheses.

A

```
GmVQ7:              HPVKVVYISNPMKIKTSASEFRALVQELTGQDAESPPDPTRFHGLIHPDS
GmVQ7m(K42V):       HPVKVVYISNPMVIKTSASEFRALVQELTGQDAESPPDPTRFHGLIHPDS

GmVQ16:             RRSRTSKSTPITLLKANTSNFRALVQQFTGCPTTTTMSLAIHKGPITLNF
GmVQ16m(L76I):      RRSRTSKSTPITLIKANTSNFRALVQQFTGCPTTTTMSLAIHKGPITLNF

GmVQ27:             STLHSPTTPNTTFVQANPSNFRAVVQKLTGASDDPSAHKLPLTLPTRLAA
GmVQ27m(F18I):      STLHSPTTPNTTIVQANPSNFRAVVQKLTGASDDPSAHKLPLTLPTRLAA

Gm35:               RKPVIIYAVSPKVLHVPAGDFMNVVQRLTGPSSGDVSPAARLASIERTSP
GmVQ35m(K71L):      RKPVIIYAVSPLVLHVPAGDFMNVVQRLTGPSSGDVSPAARLASIERTSP

GmVQ44:             PKIRIIHIFAPEIIKTDVENFRELVQKLTGKPSGENLKYFCNNKKNKAIA
GmVQ44m(E53L):      PKIRIIHIFAPLIIKTDVENFRELVQKLTGKPSGENLKYFCNNKKNKAIA

GmVQ47:             AAVAPPAPTPIRVYKVDAINFRDLVQQLTGAPEFKPADQEQHQLFPSVAP
GmVQ47m(Y46L):      AAVAPPAPTPIRVLKVDAINFRDLVQQLTGAPEFKPADQEQHQLFPSVAP
```

B
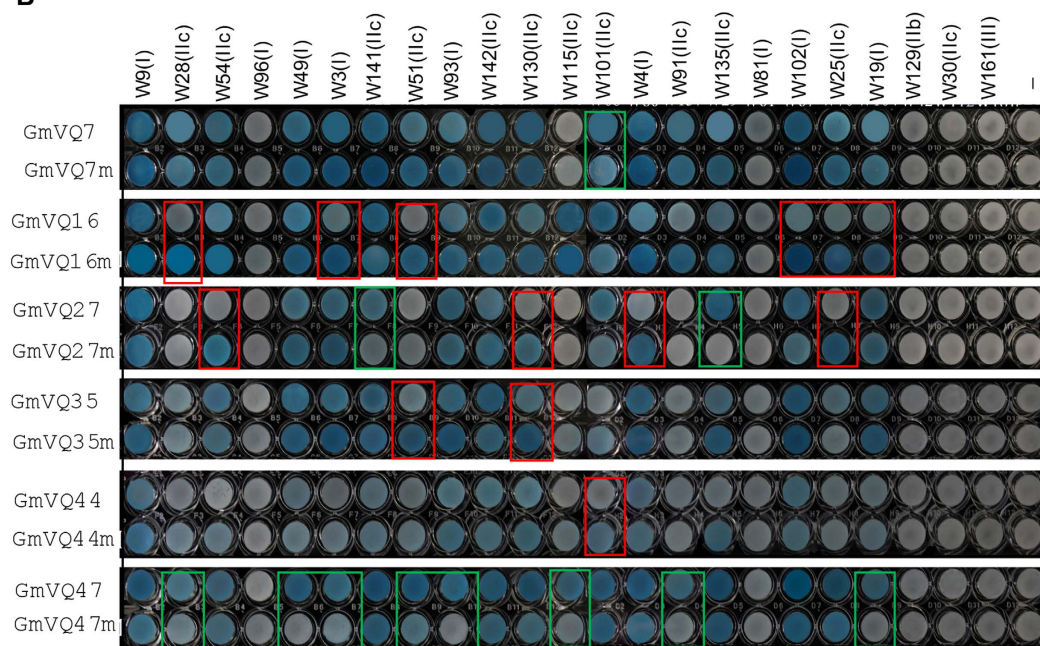
Figure 9. Effects of single amino acid substitutions in the upstream submotif of VQ proteins on binding to WRKY proteins. (**A**) Single amino acid substitutions in the upstream submotifs of six GmVQ proteins. Only the sequences of the extended VQ motifs are shown. The amino acid residues for the upstream submotifs and the FxxVQxxhTG core motifs are in red. The changed amino acid residues in mutated (m) VQ proteins in their upstream submotifs are in blue. (**B**) Interaction of wild-type GmVQ and mutated (m) GmVQ proteins with GmWRKY proteins in yeast cells. The Gal4 DNA BD-WRKY domain fusion bait vectors were co-transformed with the activation domain (AD)-VQ fusion prey vectors into yeast cells and the transformant cells were assayed for *LacZ* reporter gene expression. The empty pBD prey vector (−) was used as negative control. The specific subfamilies to which the tested WRKY proteins belong to are indicated in parentheses.

of growth (Fig. 10). With cold treatment, the transcript levels of *VQ22* in the transgenic *GsVQ22* plants were further elevated by ~10-fold (Fig. 11), resulting in even more inhibition of the growth of the transgenic plants (Fig. 10).

In the transgenic *GsVQ22* plants, the *GsVQ22* transgene is almost identical to the endogenous *GmVQ22* gene in DNA sequence except for the 12-bp deletion. Both the *GsVQ22* and *GmVQ22* genes are also highly homologous to *GmVQ11* and *GmVQ23* and, as a result, there could be cross-amplification during qRT-PCR analysis of the *VQ22* transcripts. To determine whether the large induction of *VQ22* transcripts upon cold treatment is due to increased expression of the endogenous *GmVQ22*, *GmVQ11* and *GmVQ23* genes or the *GsVQ22* transgene, we designed gene-specific primers for the four VQ genes and used them for qRT-PCR analysis of gene-specific transcripts. qRT-PCR analysis using gene-specific primers revealed no significant difference between nontransgenic plants and transgenic *GsVQ22* plants for the transcript levels of the endogenous *GmVQ11*, *GmVQ22* and *GmVQ23* genes (Fig. 11). Furthermore, cold treatment did not substantially affect the transcript levels of the
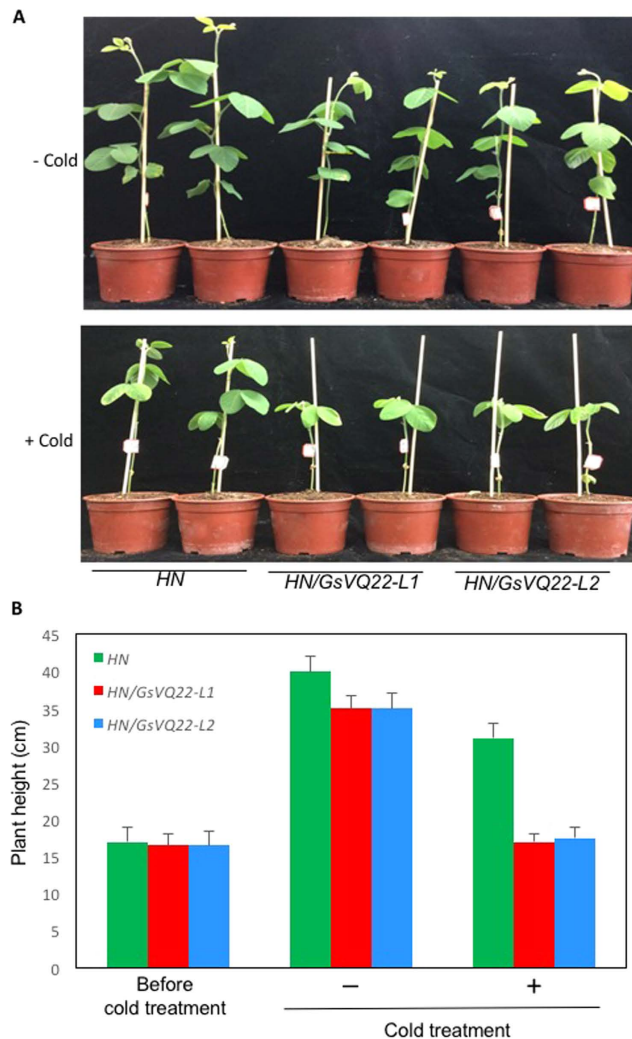
Figure 10. Growth of transgenic GsVQ22 soybean plants. (**A**) Growth of transgenic GsVQ22 soybean plants with or without cold treatment. Untransformed (HN) and two independent transgenic GsVQ22 (HN/GsVQ22) soybean plants were grown in a growth room at 24 °C with a 12/12 h hour light/dark photoperiod. Two weeks old seedlings were placed in a growth chamber at 3 °C for 48 h with a 12/12 h hour light/dark photoperiod. After the cold treatment, the plants were placed in a growth chmaber at 24 °C for recovery. Pictures were taken two weeks after cold treatment. Plants without cold treatment were also show as control (upper panel). (**B**) Heights of transgenic GsVQ22 soybean plants with or without cold treatment. Growth and cold treatment were as in A. Plant heights with (+) or without (−) cold treatment (at 3 °C for 48 h) were determined after two weeks at 24 °C following the cold treatment. Heights of plants before the cold treatment are also shown.

three endogenous VQ genes (Fig. 11). By contrast, when using *GsVQ22* transgene-specific primers, we observed >10-fold increase in its transcripts in the transgenic *GsVQ22* plants after cold treatment (Supplemental Fig. 2). Thus, the expression levels of the *GsVQ22* transgene in the two transgenic *GsVQ* lines were modest under normal condition but were substantially induced by cold temperature. The cold-induced elevation in expression of the *GsVQ22* transgene was associated with increased growth inhibition of the transgenic *GsVQ22* plants.

Discussion

Structures and functions of soybean VQ protein family.

Soybean VQ gene family contains 74 members, which is substantially larger than those from several other plants including Arabidopsis[18], rice[20], Chinese cabbage[29] and grapevine[30], where genome-wide identification of their VQ gene families have been reported. The large VQ gene family in soybean is likely related to the polyploid nature of soybean, which has undergone at least two rounds of large-scale duplication at approximately 14 and 42 million years ago[33]. Indeed, a majority of soybean VQ genes have one or more close homologs on different chromosomes (Fig. 2), suggesting that they resulted mostly from duplication of large chromosome regions or whole chromosomes. Despite the expansion of the gene family, all identified soybean VQ proteins contain the FxxxVQxxhTG motif flanked by divergent sequences as found with VQ proteins from other plant species (Figs 1 and 2). Y2H assays with 20 soybean VQ proteins from different subfamilies (or clades on the phylogenetic tree) showed that they all bind only to the C-terminal
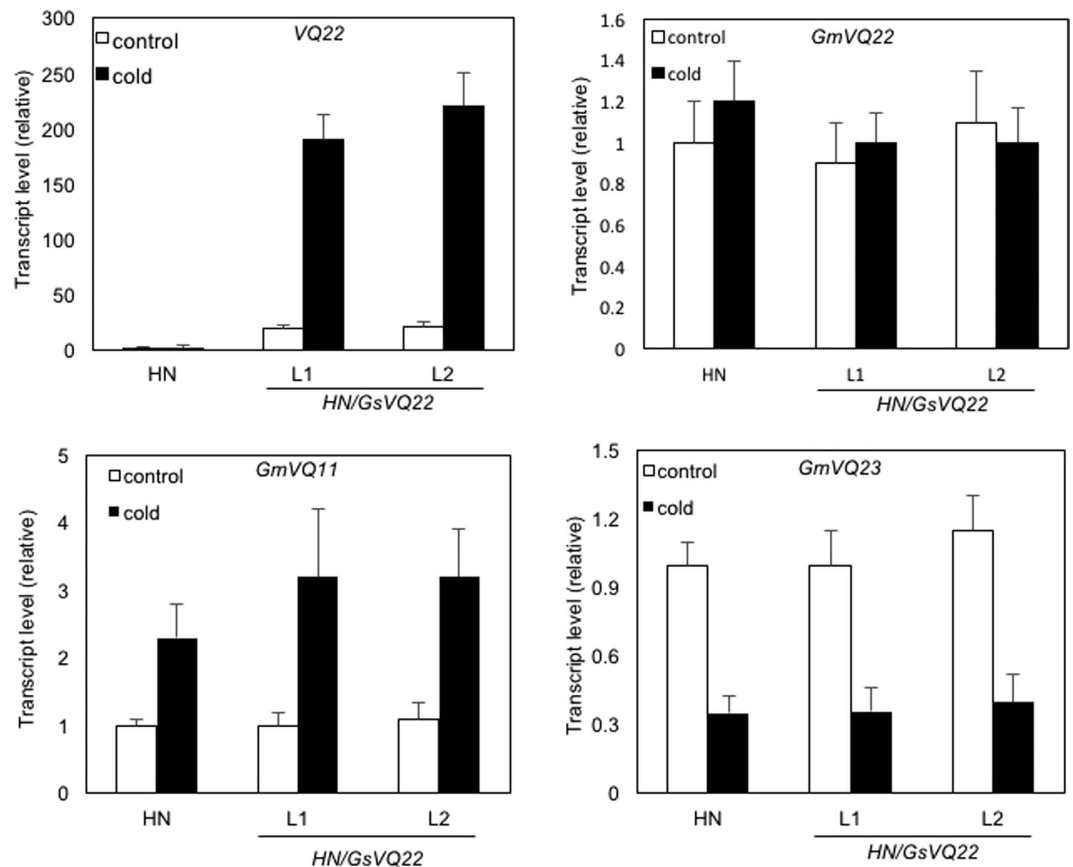
Figure 11. Expression of *GsVQ22*, *GmVQ22*, *GmVQ11* and *GmVQ23* in the transgenic *GsVQ22* soybean plants. Two weeks old nontransgenic (HN) and two independent lines (L1 and L2) of transgenic *GsVQ22* plants (HN/GsVQ22) were subjected to cold treatment (at 3 °C for 48 hours). Leaf samples were collected before and after cold treatment for RNA isolation and qRT-PCR analysis of transcript levels for *VQ22* (*GsVQ22+ GmVQ22*), *GmVQ11* and *GmVQ23* using gene-specific primers.

WRKY domains of Group I WRKY proteins and the single WRKY domain of Group IIc WRKY proteins (Fig. 3). Previously we have shown that the conserved V and Q residues in the FxxxVQxxhTG motif of Arabidopsis SIB1/VQ23 are required for its interaction with WRKY33, suggesting that the FxxxVQxxhTG motif is the core motif for WRKY binding[10]. The fact that all identified VQ proteins from distantly related plant species contain the highly conserved FxxxVQxxhTG motif, despite the great expansion of the gene family and the highly divergent flanking sequences, strongly suggests that binding of WRKY proteins remains the most important or even perhaps the central activity of plant VQ proteins.

Plant WRKY proteins have a predominant role in plant responses to biotic and abiotic stresses[3,5,35]. As their interacting proteins, plant VQ proteins are likely to be important regulators of plant disease resistance and stress tolerance as well. Molecular genetic analysis of a number of Arabidopsis VQ genes support the critical roles of VQ proteins in plant stress responses[10,18,19,24,26,28,36]. In this study, we have provided further evidence for roles of VQ proteins in defense responses of soybean plants. First, expression of a substantial percentage of soybean VQ genes was responsive to a number of phytohormones associated with plant defense and stress responses (Fig. 4). In particularly, a large number of soybean VQ genes were responsive to SA, an important defense signal in plants (Fig. 4). Previously, a large percentage of Arabidopsis WRKY and VQ genes are responsive to SA as well[4,18]. Survey of Arabidopsis WRKY and VQ gene promoters have further revealed strong enrichment of TTGAC W-box elements, suggesting extensive auto- and cross-regulation of Arabidopsis WRKY and VQ genes[4,18]. Survey of soybean VQ gene promoters likewise revealed strong enrichment of W boxes as well (data not shown). These findings indicate that auto- and cross-regulation is a conversed regulatory mechanism of plant WRKY and VQ genes, leading to formation of both positive or negative feedback loops and, consequently, rapid induction or repression of the genes under stress conditions. Second, overexpression of both soybean *GmVQ35* and *GmVQ47* caused enhanced susceptibility of transgenic Arabidopsis plants to *Botrytis* (Fig. 6A). Overexpression of *GmVQ47* also reduced heat tolerance of transgenic plants (Fig. 6B). Similarly, Arabidopsis contains not only positive but also negative VQ regulators of plant stress responses[10,18,19,24,26,28,33]. The opposite roles of VQ proteins in plant stress responses could be at least in part mediated by their opposite effects on interacting WRKY proteins. Thus, while Arabidopsis SIB1/AtVQ23 proteins enhance the DNA-binding activity of the interacting AtWRKY33 protein, other VQ proteins may inhibit the DNA-binding activity of their interacting WRKY proteins. Furthermore, some Arabidopsis VQ proteins act as transcriptional activators while others function as transcriptional repressors

in plant cells[19,28]. Therefore, a VQ protein could positively or negatively regulate the expression of the target genes of their interacting WRKY proteins by altering their DNA-binding and transcription-regulating activities. Coordinated action of both positive and negative VQ transcription regulators could potentially provide a mechanism for tight regulation and fine-tuning of genes associated with plant stress responses.

Although WRKY proteins have been established to play a role in specific developmental processes such as trichome and seed development, a broad role of plant WRKY proteins in regulation of plant flowering has yet to emerge. Interestingly, a number of plant VQ proteins have been show to alter flowering time when overexpressed in Arabidopsis. Overexpression of Arabidopsis AtVQ29 substantially delayed Arabidopsis flowering[18]. By contrast, overexpression of soybean GmVQ43 and GmVQ62 promoted flowering when overexpressed in Arabidopsis (Fig. 5B). Interestingly, a number of soybean WRKY proteins were also able to alter flowering time in Arabidopsis plants[37,38]. Further analysis revealed that a number of Arabidopsis flowering time genes including a substantial number of W boxes in their elements that are recognized by the expressed soybean WRKY proteins[38]. Even though these altered flowering phenotypes were resulted from overexpression of transgenes, they nevertheless provide clues to perhaps an important role of plant WRKY and VQ proteins in the regulation of plant development, particularly under stress conditions.

Structural elements of VQ motifs affecting WRKY-binding specificity. Arabidopsis contains 34 VQ genes and 32 Group I and IIc WRKY genes[18]. In soybean, there are 74 VQ genes and 72 Group I and IIc WRKY genes[39]. Despite the similar numbers of VQ and Group I and IIc WRKY genes, the interaction patterns clearly indicated that they are not one-to-one partners. A single VQ protein often interacts with multiple WRKY proteins and vice versa (Fig. 3)[18]. Obviously, the specific biological functions of a VQ protein requires not only the conserved VQ motif for binding of WRKY proteins but also the flanking sequences, which are highly divergent among VQ proteins. The divergent flanking sequences of a VQ protein could determine or influence its subcellular localization, protein stability and interacting partners, thereby affecting how a specific WRKY/VQ complex regulates transcription of target genes and associated biological processes. In addition, the promoter sequence of a VQ gene would determine its spatial and temporal expression patterns and consequently determine which co-expressed WRKY proteins the VQ protein will be able to partner with.

Despite the facts that all identified VQ proteins contain the FxxxVQxxhTG motif and they all appear to interact only with Group I and IIc WRKY proteins, there is difference in WRKY-binding specificity among Arabidopsis and soybean VQ proteins (Fig. 3)[18]. While some VQ proteins are able to interact with a large number of Group I and IIc proteins, other VQ proteins display very high binding specificity. To understand the structural basis of the WRKY-binding specificity of VQ proteins, we have analyzed a region immediately upstream of the FxxxVQxxhTG motif in several soybean VQ proteins and provided new evidence for a role of the upstream submotif in determining WRKY-VQ binding specificity. First, we have previously shown that mutating the V and Q residues in the FxxxVQxxhTG motif completely abolish its binding to a WRKY protein, indicating that this short VQ motif is the core element in binding of WRKY proteins[10]. On the other hand, change of a single amino acid residue in the upstream submotif does not completely abolish WRKY protein binding (Fig. 9). However, deletion of four amino acid residues in the same region in GmVQ22 resulted in a complete loss of binding to WRKY proteins (Fig. 8). These observations suggest that unlike the FxxxVQxxhTG core motif essential for binding WRKY proteins, the upstream submotif is likely to be a modulating element of the core binding motif, affecting WRKY-VQ protein binding affinity and specificity. Consistent with this interpretation, we were able to alter their binding specificity to Group I and IIc WRKY proteins by changing a single amino acid residue in the upstream submotif of several soybean VQ proteins (Fig. 9). Alteration in binding specificity to group I and IIc WRKY proteins as a result of a single amino acid change was observed for all six soybean VQ proteins but were particularly strong and extensive with GmVQ27 and GmVQ47, in which an aromatic amino acid residue (F or Y) was changed to a non-aromatic residue (I or L) (Fig. 9). This would argue for a role of steric hindrance in determining the VQ-WRKY binding specificity. Alteration in binding specificity was also very strong with GmVQ16 when a leucine residue was changed to an isoleucine residue (GmVQ16L76I) (Fig. 9), which is surprising given the similarity between the two amino acid residues. To provide clues to the questions, we have analyzed the secondary structures of the VQ domains using programs for protein structure prediction. The FxxxVQxxhTG motif is predicted to be largely an α-helical structure, whereas the upstream submotif is predicted to display largely a β-sheet structure. The β-sheet secondary structure of the upstream submotifs of different VQ proteins also appear to be highly conserved despite their variation in amino acid residues. Interestingly, even though leucine and isoleucine both have branched and strongly hydrophobic four-carbon side chains, leucine is a strong α-helix forming residue while isoleucine strongly favors the β-conformation[40–42], which the upstream submotif of VQ protein probably adopts and may explain the overall enhancement of binding of GmVQ16L76I to a number of Group I and IIc WRKY proteins over wild-type GmVQ76.

Identification of the upstream submotif of the VQ proteins as a critical factor for the VQ-WRKY binding specificity could be very useful, particularly if the crystal structure of a VQ/WRKY protein complex becomes available. Like the VQ motif of VQ proteins, the C-terminal WRKY domain of Group I WRKY proteins and the single WRKY domain of Group IIc WRKY proteins consist of invariable WRKYGQK and zinc finger residues as well as variable residues, some of which are likely in direct contacts with the VQ motifs and determine their binding specificity to VQ proteins. Once the specific motifs from both WRKY and VQ proteins that determine their binding specificity are identified, it should be possible to elucidate the underlying structural basis with a combined approach of structural biology, site-directed mutagenesis and bioinformatics. This information could then be used to study how plant VQ proteins and their interacting WRKY proteins proliferate and evolve to form new interacting complexes with distinct biological functions. This knowledge could also be used to develop novel VQ proteins that target specific WRKY proteins for manipulation of their molecular properties and ultimately the associated biological processes.

Roles and evolution of soybean VQ22 variants.   Molecular genetic analysis of crop domestication has been a topic of active research not only for addressing the basic questions on the geographical origin and the events of domestication for a given crop species but also for understanding the specific molecular changes underlying important crop traits during domestication and subsequent selection, which can be highly useful for further crop improvement. In our study of soybean VQ gene family, we have discovered that soybean *VQ22* gene has apparently undergone a genetic change during soybean evolution. While wild soybean (*G. soja*) lines contain a VQ22 protein with a functional VQ motif capable of binding WRKY proteins, cultivated soybean (*G. max*) lines and cultivars have a VQ22 protein (VQ22ΔV146_T149) incapable of binding WRKY proteins due to deletion of four amino acid residues in the region immediately upstream of the FxxxVQxxhTG motif (Table 1). In different semi-wild soybean lines (*G. gracilis*), interestingly, both forms of VQ22 are found, with a majority of them containing the GsVQ22 form as in wild soybean lines (Table 1). Semi-wild soybean is an important intermediary type of soybean with both wild and domesticated soybean characteristics. Recent genetic and genomic analyses have revealed that semi-wild soybean is probably not an intermediate transition type of soybean domestication[43,44]. Instead, semi-wild soybean was likely resulted from the hybridization of wild and domesticated soybean[43,44]. Accordingly, deletion of the 4 amino acid residues in VQ22 likely occurred during soybean domestication or subsequent selection and both the *VQ22* and *VQ22ΔV146_T149* alleles occur in semi-wild soybean due to the hybridization between the *GsVQ22*-containing wild soybean and *GmVQ22ΔV146_T149*-containing cultivated soybean.

Deletion of the four amino acid residues in the upstream submotif of VQ22 abolishes its activity to bind WRKY protein (Fig. 8) and, therefore, is likely to be a loss-of-function mutation, although a possible new activity of VQ22ΔV146_T149 cannot be completed excluded. For functional analysis of the deletion mutation of *VQ22*, we have transformed the *GsVQ22* gene into a cultivated soybean cultivar and observed significant reduction in growth, particularly with increased expression after cold treatment, in the transgenic plants (Figs 10 and 11). *GsVQ22*-mediated inhibition of growth would be undesirable for maximizing plant biomass and seed yield of soybean and a loss-of-function mutation of *GsVQ22* would be advantageous and therefore could have been positively selected and fixed in domesticated soybean. The absence of the loss-of-function mutation of *GsVQ22* in wild soybean, despite its negative effect on plant growth, could be due to the association of the gene with other beneficial traits such as disease resistance and stress tolerance that are particularly important to wild soybean.

One intriguing discovery from the study is that expression of the *GsVQ22* transgene under control of the constitutive *CaMV 35S* promoter was elevated ~10-fold after cold treatment (Fig. 11), which led to further inhibition of growth of the transgenic plants (Fig. 10). Two independent transgenic lines displayed similar cold-responsiveness of the transgene induction by cold temperature, suggesting that it is probably not due to specific DNA elements at the insertion sites of the transgene construct in the transgenic plants. More likely, cold-induced expression of the *GsVQ22* transgene was a result of reduced silencing of the transgene expression at low temperatures. Due to its inhibitory effect on growth, overexpression of *GsVQ22* at high levels would have detrimental effects on transgenic plants and this might explain why 6 of the 8 transgenic lines generated from the study had poor growth and no seed setting. In the two surviving lines with modest levels of growth inhibition, the *GsVQ22* transcript levels were elevated by 10–20-fold (Fig. 11), which is not particularly high for the strong *CaMV 35S* promoter. The relatively low levels of the *GsVQ22* transgene expression in the two transgenic lines could be due to partial silencing of the transgene, which is common for transgenes driven by a strong promoter[45]. Upon cold treatment, silencing of the *GsVQ22* transgene was reduced, leading to increased expression of the transgene and enhanced growth inhibition of transgenic plants (Figs 10 and 11). Temperature-sensitive gene silencing has been reported in both animals and plants[46–51]. In addition, the general levels of DNA methylation are reduced in plants at low temperature and gene silencing is less efficient[52]. In addition, the Antirrhinum T*am3* transposon is excised at 15 °C, but not at 25 °C, and the transposition correlates with reduced methylation status of mobile element[53–55]. Further analysis will be necessary to determine whether increased expression of the transgenic *GsVQ22* gene in transgenic plants is resulted from reduced silencing at the transcriptional or post-transcriptional levels.

Methods

Plant materials and growth conditions.   Soybean (*Glycine max* cv 'Williams 82' and 'Heinong 37') and Arabidopsis were grown in a greenhouse or growth room at 24 °C with a photoperiod of 12 h.

Identification and phylogenetic analysis of soybean VQ proteins.   Published Arabidopsis VQ protein sequences were used in BLASTp searches for GmVQ proteins in the soybean genome (http://www.phytozome.net, Glycine max Wn82.a2.v1). All final data sets were downloaded in May. 2016. The Pfam database was employed to ascertain if the candidate proteins contained features typical of VQ proteins. The phylogenetic tree based on complete amino acid sequences of VQ proteins from soybean was inferred using the neighbor-joining method. Phylogenetic analyses were conducted in MEGA6. Bootstrap values from 1,000 replicates were used to assess the robustness of the tree.

qRT-PCR analysis of VQ gene expression.   Soybean tissue samples were lyophilized and stored at −80 °C until use. Total RNA was isolated from soybean tissues using the Trizol reagent according to the supplier's instruction. Extracted RNA was treated with DNase to remove contaminating DNA and reverse transcribed using the ReverTran Ace® qPCR RT kit (Toyobo) for reverse transcriptase-PCR. qRT-PCR was performed with an StepOnePlus™ Real-Time PCR System (ABI). PCRs were performed using the SYBR® Green qPCR Master Mixes (Takara) and gene-specific primers (Supplemental Table 2). The PCR conditions consisted of denaturation at 95 °C for 1 min, followed by 40 cycles of denaturation at 95 °C for 5s, 55 °C annealing for 30s and extension at 72 °C for 30s. Melt curve analysis was performed on the end products of PCR, to determine the specificity of

reactions. Relative quantification of gene expression was calculated according to the $\Delta\Delta$Ct method. The soybean actin gene (Gm18g52780) was used as internal control. The heatmap was visualized using Heatmapper Plus tool at the Bio-Array Resource for Plant Functional Genomics.

Analysis of protein-protein interactisn using yeast two-hybrid assays. Interactions between WRKY and VQ proteins were assayed using Gal4-based two-hybrid system in yeast (*Saccharomyces cerevisiae*). pAD-Gal-VQ and pBD-Gal-WRKY fusion constructs were generated from PCR-amplified coding sequences for GmVQ genes using the gene-specific primers (Supplemental Tables 3 and 4). Mutant VQ genes were generated with overlapping PCR using gene-specific primers (Supplemental Fig. 6). The prey and bait plasmids were transformed to yeast strain YRG-2. After two days culture, the transformants were picked from the selection plates and inoculated into snap-cap tubes containing 3 ml of selective medium. Cultures were collected when the absorbance A_{546} reached 0.8–1. Centrifugation at $2000 \times g$ for 5 min and the supernatant was discarded. Cell lysis was performed by two freeze-thaw cycles (3 min in liquid nitrogen, 3 min in a 37 °C water bath). Pellets were subsequently resuspended in 20 μl water, transferred to a transparent flat bottom 96-well microplate, mixed with 100 μL phosphate-buffered saline buffer, pH 7.4, containing 500 μg/mL X-gal, 0.5% (w/v) agarose, and 0.05% (v/v) β-mercaptoethanol and incubated at 28 °C[56].

Generating transgenic GmVQ-expressing Arabidopsis plants. For generating transgenic VQ overexpression lines, the full-length coding sequences for VQ genes were amplified using gene-specific primerers (Supplemental Table 5). The amplified fragments were digested using appropriate restriction enzymes and inserted into the plant transformation vector pFGC5941 containing the CaMV35S promoter. The fused plasmids were transformed into Col-0 wild-type plants using the Agrobacterium-mediated floral-dip procedure[57]. Transformants were identified for resistance to basta herbicide. Transgenic plants overexpressing the transformed GmVQ transgenes were identified using qRT-PCR.

Analysis of disease resistance and heat tolerance of Arabidopsis plants. For analyzing resistance of trangenic Arabidopsis plants to *Botrytis cinerea*, fully expanded rosette leaves from 5–6-week-old soil-grown plants were inoculated by adding 5 μl of a *Botrytis* spore suspension at a density of 5×10^5 spores ml^{-1} and kept at high humidity in Petri dishes. Photographs of representative leave were taken 4 days after inoculation[58]. For analying heat tolerance of transgenic arabidopsis plants, 5–6-week-old soil-grown plants were placed in growth chamber 45 °C for 10 hours. The plants were then moved back to a growth room at 25 °C and pictures of representative plants were taken after 4-day recovery.

PCR detection of the VQ22ΔV146_T149 deletion in soybean VQ22 gene. Soybean genomic DNA was extracted and used for PCR delection of the VQ22ΔV146_T149 deletion in soybean VQ22 gene in a wide range of wide, semi-wild and cultivated soybean lines using two procedures: PASA[34] and AFLP. Detection with the PASA method used primers specific for GsVQ22 (5′-TGCACCCACCACCGTTCT-3′ and 5′-AGGGGAATTTCCATTTGCAT-3′) and GmVQ22 (5′-GGCGTGCACACACCACAGA-3′ and 5′-AGGGGAATTTCCATTTGCAT-3′). Detection of AFLP between *GsVQ22* and *GmVQ22* was performed using the primers 5′-CACGACAACATTAATGATAATTCAAT-3′ and 5′-AGTTGTTGGTGTCGGTGGT-3′.

Soybean transformation of the *GsVQ22* gene. The intronless *GsVQ22* gene was PCR-amplified from the genomic DNA isolated from a wild soybean line with gene-specific primers (5-AGCCTCGAGATGGACTCTGGTAACAGTGGA-3 and 5-AGCTCTAGATCACACCCCACGCGAGATCATATT-3). The amplified fragment cloned into the plant transformation vector pFGC5941 containing the *CaMV 35S* promoter. The *GsVQ22* gene was transformed into the soybean cultivar Heinong 37 using Agrobacterium-mediated cotyledonary-node transformation method[59,60]. Transgenic soybean plants were verified by resistance to the Basta herbicide and PCR analysis of both the *Bar* and *GsVQ22* genes using gene-specific primers (*Bar*: 5′- GTACCGGCAGGCTGAAGTC -3 and 5-GCACCATCGTCAACCACTAC-3′; *GsVQ22*: 5′-TTTCATTTGGAGAGGACACG-3 and 5-GGTGAGAGATGGAACCGAAA-3′). Homozygous T_3 transgenic plants were used for qRT-PCR analysis of *VQ* expression and for analysis of growth phenotypes.

References

1. Rushton, P. J., Somssich, I. E., Ringler, P. & Shen, Q. J. WRKY transcription factors. *Trends in plant science* **15,** 247–258, doi: 10.1016/j.tplants.2010.02.006 (2010).
2. Zhang, Y. & Wang, L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol Biol* **5,** 1 (2005).
3. Chen, L. *et al.* The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et biophysica acta* **1819,** 120–128, doi: 10.1016/j.bbagrm.2011.09.002 (2012).
4. Dong, J., Chen, C. & Chen, Z. Expression profile of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* **51,** 21–37 (2003).
5. Eulgem, T. Dissecting the WRKY web of plant defense regulators. *PLoS Pathog* **2,** e126 (2006).
6. Eulgem, T. & Somssich, I. E. Networks of WRKY transcription factors in defense signaling. *Current opinion in plant biology* **10,** 366–371, doi: 10.1016/j.pbi.2007.04.020 (2007).
7. Ishihama, N. & Yoshioka, H. Post-translational regulation of WRKY transcription factors in plant immunity. *Current opinion in plant biology* **15,** 431–437, doi: 10.1016/j.pbi.2012.02.003 (2012).
8. Kim, K. C., Fan, B. & Chen, Z. Pathogen-Induced Arabidopsis WRKY7 Is a Transcriptional Repressor and Enhances Plant Susceptibility to Pseudomonas syringae. *Plant Physiol* **142,** 1180–1192 (2006).
9. Kim, K. C., Lai, Z., Fan, B. & Chen, Z. Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. *Plant Cell* **20,** 2357–2371, doi: 10.1105/tpc.107.055566 (2008).

10. Lai, Z. *et al.* Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *The Plant cell* **23,** 3824–3841, doi: 10.1105/tpc.111.090571 (2011).

11. Li, J., Brader, G., Kariola, T. & Palva, E. T. WRKY70 modulates the selection of signaling pathways in plant defense. *Plant J* **46,** 477–491 (2006).

12. Johnson, C. S., Kolevski, B. & Smyth, D. R. TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. *Plant Cell* **14,** 1359–1375 (2002).

13. Luo, M., Dennis, E. S., Berger, F., Peacock, W. J. & Chaudhury, A. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proc Natl Acad Sci USA* **102,** 17531–17536 (2005).

14. Jiang, W. & Yu, D. Arabidopsis WRKY2 transcription factor mediates seed germination and postgermination arrest of development by abscisic acid. *BMC Plant Biol* **9,** 96, doi: 10.1186/1471-2229-9-96 (2009).

15. Miao, Y., Laun, T., Zimmermann, P. & Zentgraf, U. Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. *Plant Mol Biol* **55,** 853–867 (2004).

16. Robatzek, S. & Somssich, I. E. Targets of AtWRKY6 regulation during plant senescence and pathogen defense. *Genes Dev* **16,** 1139–1149 (2002).

17. Chi, Y. *et al.* Protein-protein interactions in the regulation of WRKY transcription factors. *Molecular plant* **6,** 287–300, doi: 10.1093/mp/sst026 (2013).

18. Cheng, Y. *et al.* Structural and Functional Analysis of VQ Motif-Containing Proteins in Arabidopsis as Interacting Proteins of WRKY Transcription Factors. *Plant physiology* **159,** 810–825, doi: 10.1104/pp.112.196816 (2012).

19. Jing, Y. & Lin, R. The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators. *Plant Physiol* **169,** 371–378, doi: 10.1104/pp.15.00788 (2015).

20. Kim, D. Y. *et al.* Expression analysis of rice VQ genes in response to biotic and abiotic stresses. *Gene* **529,** 208–214, doi: 10.1016/j.gene.2013.08.023 (2013).

21. Andreasson, E. *et al.* The MAP kinase substrate MKS1 is a regulator of plant defense responses. *Embo J* **24,** 2579–2589 (2005).

22. Hu, P. *et al.* JAV1 controls jasmonate-regulated plant defense. *Mol Cell* **50,** 504–515, doi: 10.1016/j.molcel.2013.04.027 (2013).

23. Petersen, K. *et al.* Arabidopsis MKS1 is involved in basal immunity and requires an intact N-terminal domain for proper function. *PLoS One* **5,** e14364, doi: 10.1371/journal.pone.0014364 (2010).

24. Wang, H., Hu, Y., Pan, J. & Yu, D. Arabidopsis VQ motif-containing proteins VQ12 and VQ29 negatively modulate basal defense against Botrytis cinerea. *Sci Rep* **5,** 14185, doi: 10.1038/srep14185 (2015).

25. Perruc, E. *et al.* A novel calmodulin-binding protein functions as a negative regulator of osmotic stress tolerance in Arabidopsis thaliana seedlings. *Plant J* **38,** 410–420, doi: 10.1111/j.1365-313X.2004.02062.x (2004).

26. Hu, Y. *et al.* Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. *Plant J* **74,** 730–745, doi: 10.1111/tpj.12159 (2013).

27. Wang, A. *et al.* The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. *Plant J,* doi: 10.1111/j.1365-313X.2010.04271.x (2010).

28. Li, Y., Jing, Y., Li, J., Xu, G. & Lin, R. Arabidopsis VQ MOTIF-CONTAINING PROTEIN29 represses seedling deetiolation by interacting with PHYTOCHROME-INTERACTING FACTOR1. *Plant Physiol* **164,** 2068–2080, doi: 10.1104/pp.113.234492 (2014).

29. Zhang, G. *et al.* Genome-Wide Identification and Analysis of the VQ Motif-Containing Protein Family in Chinese Cabbage (Brassica rapa L. ssp. Pekinensis). *Int J Mol Sci* **16,** 28683–28704, doi: 10.3390/ijms161226127 (2015).

30. Wang, M. *et al.* A comprehensive survey of the grapevine VQ gene family and its transcriptional correlation with WRKY proteins. *Front Plant Sci* **6,** 417, doi: 10.3389/fpls.2015.00417 (2015).

31. Wang, M. *et al.* Genome and transcriptome analysis of the grapevine (Vitis vinifera L.) WRKY gene family. *Hortic Res* **1,** 14016, doi: 10.1038/hortres.2014.16 (2014).

32. Wang, X., Zhang, H., Sun, G., Jin, Y. & Qiu, L. Identification of active VQ motif-containing genes and the expression patterns under low nitrogen treatment in soybean. *Gene* **543,** 237–243, doi: 10.1016/j.gene.2014.04.012 (2014).

33. Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463,** 178–183, doi: 10.1038/nature08670 (2010).

34. Sommer, S. S., Groszbach, A. R. & Bottema, C. D. PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *Biotechniques* **12,** 82–87 (1992).

35. Eulgem, T. & Somssich, I. E. Networks of WRKY transcription factors in defense signaling. *Curr Opin Plant Biol* **10,** 366–371 (2007).

36. Pecher, P. *et al.* The Arabidopsis thaliana mitogen-activated protein kinases MPK3 and MPK6 target a subclass of 'VQ-motif'-containing proteins to regulate immune responses. *New Phytol* **203,** 592–606, doi: 10.1111/nph.12817 (2014).

37. Luo, X. *et al.* Ectopic expression of a WRKY homolog from Glycine soja alters flowering time in Arabidopsis. *PLoS One* **8,** e73295, doi: 10.1371/journal.pone.0073295 (2013).

38. Yang, Y. *et al.* Functional analysis of structurally related soybean GmWRKY58 and GmWRKY76 in plant growth and development. *J Exp Bot,* doi: 10.1093/jxb/erw252 (2016).

39. Rinerson, C. I., Rabara, R. C., Tripathi, P., Shen, Q. J. & Rushton, P. J. The evolution of WRKY transcription factors. *BMC Plant Biol* **15,** 66, doi: 10.1186/s12870-015-0456-y (2015).

40. Chou, P. Y. & Fasman, G. D. Prediction of protein conformation. *Biochemistry* **13,** 222–245 (1974).

41. Chou, P. Y. & Fasman, G. D. Empirical predictions of protein conformation. *Annu Rev Biochem* **47,** 251–276, doi: 10.1146/annurev.bi.47.070178.001343 (1978).

42. Fasman, G. D. Protein conformational prediction. *Trends Biochem Sci* **14,** 295–299 (1989).

43. Qiu, J. *et al.* Genome re-sequencing of semi-wild soybean reveals a complex Soja population structure and deep introgression. *PLoS One* **9,** e108479, doi: 10.1371/journal.pone.0108479 (2014).

44. Wang, K. J., Li, X. H. & Liu, Y. Fine-scale phylogenetic structure and major events in the history of the current wild soybean (Glycine soja) and taxonomic assignment of semi-wild type (Glycine gracilis Skvortz.) within the Chinese subgenus Soja. *J Hered* **103,** 13–27, doi: 10.1093/jhered/esr102 (2012).

45. Luo, Z. & Chen, Z. Improperly Terminated, Unpolyadenylated mRNA of Sense Transgenes Is Targeted by RDR6-Mediated RNA Silencing in Arabidopsis. *Plant Cell* **19,** 943–958 (2007).

46. Tuttle, J. R., Idris, A. M., Brown, J. K., Haigler, C. H. & Robertson, D. Geminivirus-mediated gene silencing from Cotton leaf crumple virus is enhanced by low temperature in cotton. *Plant Physiol* **148,** 41–50, doi: 10.1104/pp.108.123869 (2008).

47. Fu, D. Q. *et al.* Enhancement of virus-induced gene silencing in tomato by low temperature and low humidity. *Mol Cells* **21,** 153–160 (2006).

48. Russell, P., Walsh, E., Chen, W., Goldwich, A. & Tamm, E. R. The effect of temperature on gene silencing by siRNAs: implications for silencing in the anterior chamber of the eye. *Exp Eye Res* **82,** 1011–1016, doi: 10.1016/j.exer.2005.12.003 (2006).

49. Sos-Hegedus, A., Lovas, A., Kondrak, M., Kovacs, G. & Banfalvi, Z. Active RNA silencing at low temperature indicates distinct pathways for antisense-mediated gene-silencing in potato. *Plant Mol Biol* **59,** 595–602, doi: 10.1007/s11103-005-0354-z (2005).

50. Kameda, T., Ikegami, K., Liu, Y., Terada, K. & Sugiyama, T. A hypothermic-temperature-sensitive gene silencing by the mammalian RNAi. *Biochem Biophys Res Commun* **315,** 599–602, doi: 10.1016/j.bbrc.2004.01.097 (2004).

51. Fortier, E. & Belote, J. M. Temperature-dependent gene silencing by an expressed inverted repeat in Drosophila. *Genesis* **26,** 240–244 (2000).

52. Finnegan, E. J., Genger, R. K., Peacock, W. J. & Dennis, E. S. DNA Methylation in Plants. *Annu Rev Plant Physiol Plant Mol Biol* **49,** 223–247, doi: 10.1146/annurev.arplant.49.1.223 (1998).
53. Hashida, S. N. *et al.* The temperature-dependent change in methylation of the Antirrhinum transposon Tam3 is controlled by the activity of its transposase. *Plant Cell* **18,** 104–118, doi: 10.1105/tpc.105.037655 (2006).
54. Hashida, S. N., Kitamura, K., Mikami, T. & Kishima, Y. Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in Antirrhinum majus. *Plant Physiol* **132,** 1207–1216 (2003).
55. Kitamura, K., Hashida, S. N., Mikami, T. & Kishima, Y. Position effect of the excision frequency of the Antirrhinum transposon Tam3: implications for the degree of position-dependent methylation in the ends of the element. *Plant Mol Biol* **47,** 475–490 (2001).
56. Mockli, N. & Auerbach, D. Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system. *Biotechniques* **36,** 872–876 (2004).
57. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16,** 735–743 (1998).
58. Zheng, Z., Qamar, S. A., Chen, Z. & Mengiste, T. Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J* **48,** 592–605 (2006).
59. Song, Z. Y. *et al.* Screening Chinese soybean genotypes for Agrobacterium-mediated genetic transformation suitability. *J Zhejiang Univ Sci B* **14,** 289–298, doi: 10.1631/jzus.B1200278 (2013).
60. Paz, M. M., Martinez, J. C., Kalvig, A. B., Fonger, T. M. & Wang, K. Improved cotyledonary node method using an alternative explant derived from mature seed for efficient Agrobacterium-mediated soybean transformation. *Plant Cell Rep* **25,** 206–213, doi: 10.1007/s00299-005-0048-7 (2006).

Acknowledgements

Author Contributions

Z.C. conceived and designed the experiments. Y.Z., Y.Y., X.Z., Y.C. and B.F. performed the experiments. Y.Z. and Z.C. analyzed the data and wrote the manuscript. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhou, Y. *et al.* Structural and Functional Characterization of the VQ Protein Family and VQ Protein Variants from Soybean. *Sci. Rep.* **6**, 34663; doi: 10.1038/srep34663 (2016).