Genome-wide identification and expression analysis of the VQ gene family in *Cucurbita pepo* L.

Xu Ke[1], Wang Ping[1]

[1] College of horticulture and plant protection, Inner Mongolia Agricultural University, Huhehaote, Inner Mongolia, China

Corresponding Author:
Wang Ping[1]
University of east, Huhehaote, Inner Mongolia, 01000, China
Email address: wangping@imau.edu.cn

Abstract

VQ protein is a plant specific protein, which plays an important role in plant growth and development and biological and abiotic stress response.This study aimed to systematically analyze for the first time the *VQ* of *C. pepo* and understand their expression patterns in response to different stimuli. Herein, 44 *VQ* genes were identified, which were divided into eight groups (I–VIII) based on phylogenetic analysis. Two genes (*CpVQ1* and *CpVQ2*) could not be located on the chromosome, whereas the remaining *CpVQ* genes were randomly distributed on the chromosomes, except for chromosomes 15 and 18. Noteworthy, the main event driving the expansion of the *VQ* gene family was chromosome fragment duplication. Based on qRT-PCR analysis, *VQ* genes are expressed in different tissues, and *VQ* genes are differentially regulated under a variety of abiotic stresses and powdery mildew stress. indicating that they play an important role in plant stress response and other aspects.This report presents the first systematic analysis of *VQ* genes from *C. pepo* and provides a solid foundation for further research of the specific functions of VQ proteins.

Keywords:VQ genes family;*Cucurbita pepo* L.;expression pattern;powdery mildew

Introduction

VQ protein is a highly conserved transcription factor that is widely expressed in biology, including several monocotyledonous and dicotyledonous plants(Jing and Lin, 2015; Song et al.,2016;Ding et al.,2019). This protein owns its name to its VQ motifs, which allows it to interact with WRKY transcription factors to regulate plant growth and development(Lai et al., 2011). For example, *VQ10* interacts with *WRKY25* and *WRKY33*, which when simultaneously overexpressed can impair plant growth(Cheng et al., 2012). In *Arabidopsis*, *AtWRKY2/34* and *AtVQ20* form a complex that negatively regulates the expression of MYB transcription factors to

40 regulate the development and function of pollen(Lei et al., 2017; 2018). The *ATVQ8* gene affects
41 leaf development, making the leaves appear yellow-green throughout the growth period, causing
42 plant malnutrition and affecting the normal growth of *Arabidopsis* plants. Plants overexpressing
43 *VQ17*, *VQ18*, and *VQ22* have slow growth and yellow leaves, and the leaf area becomes
44 smaller(Cheng at al., 2012). Moreover, *IKU1* encodes a protein containing a VQ motif that is
45 involved in the regulation of endosperm development, which in turn affects seed size(Wang et
46 al., 2010).
47   VQ plays an important protective role on plants, allowing them to fight pathogenic
48 bacteria and other biological and abiotic stresses(Jing et al.,2018). For example, *OsVQ13* is
49 highly expressed at 6 h after bacterial blight infection; thus, it may be involved in the resistance
50 pathway of *Oryza sativa* to bacterial blight(Kim et al., 2013a). Some *PbrVQ* genes are
51 differentially expressed under gibberellin and black spot stress(Cao et al.,2018).Furthermore,
52 transgenic *Arabidopsis* overexpressing *AtVQ15* is highly sensitive to salt stress during seed
53 germination and seedling growth, whereas mutants of this gene show strong salt tolerance(Perruc
54 at al., 2004). In response to harsh environments, plants have evolved several complex
55 mechanisms to deal with external stimuli(Fujita et al., 2006). This response is caused by the
56 interaction of transcription factors and cofactors to coordinate the transcription mechanisms of
57 the plant in response to the surrounding environment(Wray et al., 2003).For example,*AtVQ9* acts
58 antagonistically with *AtWRKY8* to mediate responses to salt stress(Hu et al., 2013).Studies have
59 shown that *AtVQ10* and *WRKY8* form a complex in the nucleus and positively regulate the
60 resistance of *Arabidopsis* to *Botrytis cinerea*(Chen et al., 2018). The banana fruit *MaWRKY26*
61 transcription factor physically interacts with the VQ motif-containing protein *MaVQ5*, which
62 leads to attenuated *MaWRKY26*-induced transactivation of jasmonic acid biosynthetic genes that
63 are associated with various stress responses(Ye et al.,2016).
64   *C. pepo* belongs to the *Cucurbita* genus of the Cucurbitaceae family. Because it contains
65 several nutrients, it has high edible and medicinal value, and is widely cultivated worldwide.
66 During the cultivation process, *C. pepo* is susceptible to various biological and abiotic stresses
67 that will affect the yield and quality of the fruit. Therefore, identification of resistance genes has
68 great significance for improving the yield and quality of *C. pepo* through molecular
69 breeding.Currently, 34, 39, 75, and 61 *VQ* genes have been identified in *Arabidopsis*, *O. sativa*,
70 *Glycine max* and Maize, respectively(Cheng at al., 2012; Kim at al., 2013a; Wang et al., 2019;
71 Song et al. ,2016). However, there is no detailed description of the *VQ* genes of *C. pepo*. In this
72 study, we investigated the *VQ* genes in the *C. pepo* genome, and analyzed their phylogeny, gene
73 structure, chromosomal location, and collinearity for a comprehensive analysis. In addition, we
74 also analyzed the expression levels of *CpVQs* under different abiotic stresses and powdery
75 mildew stress. This study is expected to provide basic information for the identification and
76 classification of *CpVQ* genes. Further experimental analysis can allow us to understand the
77 functions of *CpVQ*s involved in plant stress response.

78 **Materials & Methods**
79 **Identification of VQ genes in *C. pepo***

80   In present study, two methods were used to identify complete VQ members in *C. pepo*. Firstly,

81   The Hidden Markov Model (HMM) profiles of the VQ motif (PF05678) was download from

82   PFAM protein family database (http://pfam.xfam.org/)(Punta et al.,2004). Then the model as

83   a probe to perform a BLASTp against the *C. pepo* genome database using HMMER 3.0. The

84   cut-off E-value was set as 0.1. Secondly, The known VQ motif-containing members of *A.*

85   *thaliana* were obtained from The *Arabidopsis* Information Resource (TAIR) database

86   (http://arabidopsis.org) were used as queries to conduct a local BLASTp against the protein

87   database of *C. pepo*. The targeting genes with similarity of E-value less than $1e^{-20}$ were retained

88   for the further analysis.We set e as two values of 0.1 and $1e^{-20}$ for filtering. We set E-value to 0.1

89   for the initial filtering, and then set E-value to $1e^{-20}$ for in-depth filtering. After merging the two

90   results, the candidate *VQ* genes were evaluated using the online program

91   PFAM(http://pfam.xfam.org/search),Conserved Domains DataBase Search

92   (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and SMART tool (http://smart.embl-

93   heidelberg.de/) to confirm the presence of VQ motif. The biophysical properties of VQ members,

94   such as, peptide length, molecular weight (MW), isoelectric point (pI) were predicted using the

95   online ExPasy program (http://www.expasy.org/tools/)(Wilkins et al.,1999).

Multiple sequence alignment and Phylogenetic Analysis

97   In order to investigate the evolutionary relationships and classification of VQ members in *C.*

98   *pepo*, we firstly conducted a multiple sequence alignment based on the full-length amino acid

99   sequence of all *CpVQs* using the ClustalW tool with auto strategy parameters. According to the

100   alignment results, we constructed a phylogenetic tree using the neighbor-joining (NJ) method

101   with MEGAX software (Tamura et al.,2011).Branch support for the tree topology was

102   estimated by using a bootstrap analysis with 1000 replicates. The phylogenetic tree was

103   illustrated using the online ITOL (http://itol.embl.de/help.cgi) online tool.

Gene structure analysis and conserved motif detection

105   The online program Gene Structure Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/) was

106   used to draw a diagrammatic sketch of the intron-exon structure by comparing their coding

107   sequence (CDSs) with their corresponding genomic DNA sequences. The conserved motifs of

108   CpVQs were detected by the online tool MEME (http://meme.ebi.edu.au/), the parameters were

109   set as follows : zero or one occurrence per sequence; maximum number of motifs: 20; and other

110   optional parameters was set default(Bailey et al.,2015).

Chromosome Locations, Gene duplication and collinearity analysis

112   For the further investigation of evolution in *CpVQs*, we firstly constructed the   distribution map

113   of VQ members in *C. pepo*. The physical positions of all *VQ* genes were retrieved from the

114   GFF3 annotation file using a local script, the schematic diagram of chromosomal location was

115   visualized by TBTOOLS(Chen et al.,2020). For the identification of tandem duplication events

116   and segmental duplication events in CpVQ family, all *CpVQs* amino acid sequences were

117   aligned using BLASTp, with an e-value of $1e^{-10}$. As previous research described, two or more

118   genes located on the same chromosome were arranged in a 200 kb distance and shared more than

119   70% identity as analyzed with BLASTp can be defined as tandem duplication events. The

120 segmental duplication events were identified by using the Multiple collinear scanning toolkits
121 (MCScanX) with default parameters(Wang et al.,2012a). The Circos program was used to draw
122 collinearity maps to exhibit duplicated gene pairs between *CpVQ* genes as well as the synteny
123 blocks of the orthologous *VQ* genes between *C. pepo* and *Arabidopsis thaliana*, *Cucumis melo*,
124 *Oryza sativa*, *Zea mays*.

Calculating Ka and Ks

126 The Ka and Ks were calculated to assess the selection history and divergence time of gene
127 families. The values of synonymous (Ks) and nonsynonymous (Ka) substitutions of duplicated
128 *VQ* genes were estimated by using the KaKs_Calculator 2.0 with the NG method(Wang et
129 al.,2010a). The divergence time (T) was calculated using the formula $T = Ks/(2 \times 6.1 \times 10^{-9}) \times 10^{-6}$
130 million years ago (MYA)(Kim et al., 2013a; Kim et al., 2013b).

VQ gene expression analysis of *C. pepo*

132 The transcriptome expression data of *VQ* genes under Podosphaera xanthii stress was available
133 from our laboratory. The transcript abundance was represented by fragments per kilobase of
134 exon per million mapped reads (FPKM) values which were calculated based on RNA-Seq reads.
135 The heatmaps showing expression profiles were generated using log10-transformed FPKM
136 values.The results were presented as heatmaps using TBTOOLS software(Chen et al.,2020).

Plant materials and treatments

138 *C. pepo* (Self-delivered F2) was used in present study. Seeds were planted in a 1:1 (w/w) mixture
139 of soil and sand, cultured in an artificial climatic chamber kept at 30/22 ℃ with a 18/6 h
140 photoperiod (day/night). Seedlings that germinated after 8 weeks were subjected to different
141 stress conditions: 200 mM NaCl solution, 20% PEG6000 (drought), 4 ℃ (cold), immersing
142 (waterlogging) and cultured for a total of 24 h, leaves were collected after 0, 12 and 24 h.In order
143 to analyze the expression of *VQ* genes in different tissues, we collected plant roots, stems, and
144 leaves for RNA preparation. All of the samples were immediately frozen in liquid nitrogen and
145 stored at −80 ℃ for subsequent total RNA extraction. All samples were tested with three
146 technical replicates and three independent biological replicates.

RNA extraction and Quantitative real-time PCR (qRT-PCR)

148 Total RNA was extracted from *C.peop* using RNAsimple Total RNA Kit (TIANGEN
149 BIOTECH, Beijing) according to the manufacturer's instructions. First-strand cDNA synthesis
150 was accomplished using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix
151 (Transgen Biotech, Beijing). Quantitative Real-time PCR (qRT-PCR) was performed using TB
152 Green *Premix Ex Taq* II (TliRNaseH Plus) (RR420Q TaKaRa Biotechnology, Beijing, China) on
153 an FTC-3000P system (Applied Biosystems, CA, Canad) with the primers listed in Table S1.The
154 reaction procedure was completed under the following program:30 s of pre-denaturation at 95 ℃,
155 40 cycles of 5 s at 95 ℃, and 30 s at 60 ℃,and 4 °C to finish. All samples were tested with three
156 technical replicates and three independent biological replicates. The relative expression level was
157 calculated while using the $2^{-\Delta\Delta CT}$ method(Livak et al.,2001).The *CAC* genes were used as
158 internal control.(Obrero et al.,2011).

Results

Identification and sequence analysis of *VQ* genes in *C. pepo*

After manually removing redundant entries through screening(manually delete the genes without VQ motif) and validation of the search results, a total of 44 *VQ* genes were identified within the whole genome of *C. pepo*, which were named *CpVQ1–CpVQ44* based on their physical locations on the chromosomes. The characteristics concerning each gene, including gene number, length of coding sequence and amino acid sequence,MW and PI of the proteins, are summarized in **Table 1**. Subsequent sequence analysis of these 44 *CpVQ*s showed that the encoded CpVQ proteins ranged from 73 amino acids(aa) (*CpVQ34*) to 628 amino acids (*CpVQ14*) in length (average length: 219 amino acids). Similar to previous studies in *Arabidopsis* and *O. sativa*, most VQ proteins contained less than 300 amino acids. The calculated molecular weight and isoelectric points of these proteins varied from 8.34 kDa (*CpVQ34*) to 67.83 kDa (*CpVQ14*), and 4.68 (*CpVQ40*) to 10.67 (*CpVQ19*), respectively. The length of the coding sequences of this gene family was between 222 and 1,887 bp (average length: 660 bp).

Phylogenetic analysis of *CpVQs*

To detect the evolutionary relationships and classification of the VQ family in *C. pepo*, unrooted phylogenetic Neighbor-Joining trees were constructed with the 44 CpVQ proteins and the known VQ protein from *Arabidopsis*.Through the analysis of the phylogenetic and structural features of the VQ domains,and these proteins were divided into 8 clades(I–VIII) based on the nomenclature of the *Arabidopsis VQs*, with two proteins in I and VIII groups, three each in VI and VII, six members in III, eight in II, and group IV and V have the biggest amount of proteins with 10 *VQs* (**Fig. 1**). When compared with the other groups, the size of group IV and V were significantly larger. These results were not completely consistent with previous studies in *A. thaliana*, *O. sativa*, and *Zea mays*. In addition, a phylogenic tree was built using the 44 CpVQ protein sequences, which indicated that the groups IV and V in *C. pepo* had more VQ members than those in *Arabidopsis*. In addition, it is also possible that the expansion of the two subgroups may have resulted from gene duplications.

Gene structure and conserved motifs analysis

Gene structure can provide more information about the evolutionary relationship in a gene family; thus, the organization of exons and introns in *CpVQs* was investigated using the GSDS2.0 online tool. Structural analysis of *VQs* showed that half of the members of group V had introns, such as *CpVQ14* that had four introns, which may be related to their long sequences; genes in group VII have longer coding regions; whereas genes of group I have shorter coding regions than other groups. Moreover, 75% of *CpVQs* were found to be intronless genes, suggesting that many introns were lost during the long evolution. To further analyze the characteristics of *CpVQs*, the MEME online tool was used to predict the potential conserved motifs. A total of 20 motifs were identified, with lengths ranging between nine and 48 amino acids. Noteworthy, every *CpVQ* contained 1–6 conserved motifs (**Fig. 2**). Motif 1 contained the domain present in every VQ member; thus, representing an essential element in the VQ family. In addition, every group was found to have a clearly identifiable motif structure that distinguished it from the other groups. For example, Motif 2 and Motif 5 were prominent in

200 Group IV, Motif 17 was only observed in Group VII, and Motif 6 was only present in Group III.
201 It is worth mentioning that the gene structure and motif organization of *CpVQs* were similar
202 within the same group, but diverged between groups, which also indicated that the phylogenetic
203 classification was reliable.

Chromosome mapping and duplication events analysis

205 Two *VQ* genes (*CpVQ1* and *CpVQ2*) could not be mapped on any chromosome, but the
206 remaining *CpVQs* were randomly distributed on the chromosomes, except chomosomes 15 and
207 18 (**Fig. 3**). Chromosome 19 had the highest number of *CpVQs* (n = 7), followed by chromosome
208 10 (n = 6), chromosomes 1, 2, 6 , 8, and 14 (n = 3), chromosome 7 (n = 2), and the remaining
209 chromosomes contained one *VQ*. In addition, most *CpVQs* were found to be mainly located on
210 the two ends of the chromosomes.
211 To better understand the evolution mechanism of *VQs* in plants, the tandem duplication
212 and segmental duplication events of the CpVQ family were evaluated. Based on BLSTAp
213 results, no tandem duplication event was detected. A total of 21 pairs with 28 *CpVQs* were
214 identified in the whole genome (**Fig. 4**). Moreover, groups V and IV contained six and five
215 segmental duplication events, respectively. Hence, these results suggested that segmental
216 duplications were the primary driving force responsible for the expansion of the VQ family in *C.*
217 *pepo*, which is consistent with previous reports.

Synteny analysis of VQ genes

219 To further understand the expansion mechanisms of the CpVQ family, we constructed
220 comparative syntenic maps of *C. pepo* associated with four representative species, including two
221 dicots (*Arabidopsis* and *Cucumis melo*) and two monocots (*O. sativa* and *Z. mays*). Overall, the
222 *CpVQ* genes had the most homologous gene pairs with *C. melo* (n = 27), followed by
223 *Arabidopsis* (n = 16) . However, no homologous gene pairs were observed between *C. pepo* and
224 *O. sativa* or *Z. mays*(**Fig. 5**). The syntenic analysis results further suggested that the *VQ* genes
225 are highly conserved in dicotyledonous plants.
226 We calculated the Ka, Ks and Ka/Ks ratios of all para-homologous gene pairs (Table 2)
227 to explore the evolutionary constraints of repeated *CpVQs*. The Ka/Ks values of most gene pairs
228 were less than 1.0, which indicated that these gene pairs had undergone purification selection
229 pressure.In addition, 21 pairs of genes vaeried from 31 to 216 MYA.

Expression Pattern of the *CpVQ* Genes in Different Tissues

231 In order to explore the possible role of *CpVQ* gene in the growth and development of *C. pepo*,
232 we performed qRT-PCR expression analysis in three tissues of roots, stems and leaf. Expression
233 patterns varied among the randomly selected15 *CpVQ* genes (**Fig. 6**).Four *CpVQ* genes (*CaVQ1,*
234 *3,12* and *21*) were highly expressed in the root;two *CpVQ* genes (*CpVQ9* and *22*) were highly
235 expressed in leaf and root;four *CpVQ* genes (*CpVQ16, 26,34* and *40*) were highly expressed in
236 stem and root;two *CpVQ* genes (*CpVQ14* and *39*) were highly expressed in stem.

Expression pattern of the *CpVQs* under powdery mildew stress

238 We randomly selected 15 *CpVQ* genes from eight groups, and made sure their responses to the
239 powdery mildew-,drought-,cold-, salt-, and waterlogging-stress.Several studies have shown that

240   *VQs* play an important role in plant response to abiotic and biotic stress. Therefore, the
241   expression profiles of *CpVQs* under powdery mildew stress were evaluated next based on RNA-
242   seq data. Most of the *CpVQs* were found to be downregulated at early treatment timepoints, such
243   as the members in group I and IV(**Fig. 7**). In addition, some members were downregulated and
244   then upregulated. A total of 15 *CpVQs* from different groups were randomly selected for further
245   validation by qRT-PCR(**Fig. 8**). Accordingly, most of the tested *CpVQs* were confirmed to be
246   downregulated after powdery mildew infection, with the expression of these genes being
247   correlated with the RNA-seq data at different treatment timepoints. Among these genes,The
248   expression of two genes (*CpVQ1*, *CpVQ39*) significantly increased (more than 2-fold) at 12
249   h,whereas the expression of seven genes (*CpVQ3*, *CpVQ9*, *CpVQ12*, *CpVQ21*, *CpVQ22*,
250   *CpVQ36*, and *CpVQ40*) was significantly downregulated.In addition, three genes (*CpVQ26*,
251   *CpVQ33*, and *CpVQ34*) had a high expression value at 24 h after powdery mildew stress.

252   ***CpVQs* gene expression following abiotic stress by qRT-PCR**

253   To further investigate the roles of *CpVQs* in abiotic stress responses, the expression patterns of
254   *VQs* under different stress conditions, such as drought, cold, salt, and waterlogging were also
255   assessed.

256   Under drought treatment(**Fig. 9**), all *CpVQs* were upregulated at different treatment
257   timepoints. The expression of seven genes (*CpVQ1*, *CpVQ12*, *CpVQ14*, *CpVQ26*, *CpVQ33*,
258   *CpVQ34*, and *CpVQ40*) significantly increased (more than 1.5-fold) at 12 h, whereas *CpVQ21*
259   was significantly downregulated at 12h but it then increased at 24 h. In addition, six genes
260   (*CpVQ3*, *CpVQ8*, *CpVQ9*, *CpVQ16*, *CpVQ21*, *CpVQ22*, and *CpVQ39*) had a high expression
261   value at 24 h after drought stress.

262   During cold stress(**Fig. 10**), the expression of six *CpVQs* (CpVQ1, *CpVQ22*, *CpVQ33*,
263   *CpVQ39*, and *CpVQ40*) were significantly upregulated points (more than 2-fold) at 24 h. In
264   contrast, three genes (*CpVQ3*, *CpVQ9*, and *CpVQ21*) were downregulated at 12 h and their
265   levels were further decreased at 24 h. Five members (*CpVQ8*, *CpVQ14*, *CpVQ16*, *CpVQ26*, and
266   *CpVQ34*) were downregulated at early time points but their expression were significantly
267   increased and peaked at 24 h.

268   Upon salt stress(**Fig. 11**), seven genes (*CpVQ16*, *CpVQ22*, *CpVQ26*, *CpVQ33*, *CpVQ34*,
269   *CpVQ36*, and *CpVQ40*) shared a similar expression trend, rapidly rising at early time points but
270   with subsequently decreased expression. In contrast, *CpVQ8* and *CpVQ12* showed a trend of
271   downregulation from 0 to 12 h, but then their levels were rapidly increased. In addition, the
272   expression of *CpVQ21* showed a trend of downregulation from 0 to 24 h.

273   For the case of waterlogging treatment(**Fig. 12**),The three genes(*CpVQ3*, *CpVQ9*, and
274   *CpVQ21*) upregulated in the first 12h and then downregulated in the next 24h. In contrast,
275   *CpVQ8* and *CpVQ14* downregulated in the first 12h and then upregulated in the next 24h.
276   Interesting, the expression of nine *CpVQs* (*CpVQ1*, *CpVQ12*, *CpVQ16*, *CpVQ22*, *CpVQ26*,
277   *CpVQ33*, *CpVQ34*, *CpVQ39*, and *CpVQ40*) were upregulated from 0 to 24 h. Only *CpVQ36* was
278   downregulated after waterlogging treatment.

279   **Discussion**

280   VQ has been proven to be the main transcriptional regulator in plants.Currently,the VQ family
281   genes have been systematically analyzed in *Arabidopsis*(Cheng at al., 2012),rice(Kim at al.,
282   2013a),maize(Song et al. ,2016),tobacco(Liu et al.,2020),and cotton(Cheng et al.,2020),and
283   responding to biotic and abiotic stresses(Jing et al.,2018;Perruc at al., 2004).However, current
284   information on VQ characteristics in *C. pepo* is limited. Therefore, a comprehensive analysis of
285   VQ genes in *C. pepo* and their expression patterns under various non-biological and powdery
286   mildew treatments may pave the way for a better understanding on the mechanism of plant
287   growth and development. This will also help to select candidate genes and lay the foundation for
288   further in-depth study of the role of different regulatory networks in plant development and
289   stress-related processes.

Conservation of the VQ gene family of *C. pepo*

291   In higher eukaryotes, genes without introns are very common(Louhichi., 2011;Liu at al.,2021).
292   Herein, according to the gene structure, most *VQ* genes in *C. pepo* were also found to have no
293   introns. Only 11 *VQs* had introns, among which *CpVQ14* contained four introns. This result is
294   consistent with the lack of introns in 88.2% of *Arabidopsis*, 90% of Chinese Cabbage, and 92.3%
295   of tomato(Cheng at al., 2012; Zhang at al., 2015; Ding at al., 2019). Subsequently, a
296   phylogenetic tree was constructed based on the VQ proteins of *C. pepo* and *A. thaliana*.Through
297   the analysis of the phylogenetic and structural features of the VQ domains,and these proteins
298   were divided into 8 clades(I−VIII) based on the nomenclature of the *A. thaliana VQs*.This
299   analysis revealed that *VQs* with introns are located in different subfamilies, suggesting that these
300   introns appear relatively independently.Comparative these plants (*Cicer arietinum* and *Medicago*
301   *truncatula*, *A. thaliana* and lower plants, moss) indicate that most *VQ* genes have lost introns
302   during the long evolutionary period. Taken together, *VQs* identified in *C. pepo* L. and in other
303   species provide a certain reference value for the evolution of introns in plants.The average length
304   of the *CpVQ* is 219.3 amino acids, a majority of the *C. pepo VQ* genes are intronless so they
305   encode relatively small proteins with fewer than 300 amino acid residues, which is highly similar
306   to that reported for *Nicotiana tabacum*, *Arabidopsis*, tomato, *O. sativa*, and other plants(Liu at
307   al., 2020; Cheng at al., 2012; Ding at al., 2019; Kim at al., 2013a). Noteworthy, *CpVQ4*,
308   *CpVQ14*, *CpVQ23*, and *CpVQ32* were all found to be close to 400 amino acids.
309   The genetic structure of the *VQ* gene and the conserved Motif not only appear in higher
310   plants such as *Arabidopsis*, *O. sativa*, *Z. mays*, among others, but also in lower plants such as
311   moss, which also implies their ancient origins in the evolutionary history and their important role
312   in plant development(Jing at al., 2015; Song at al., 2016).Moreover, all 44 VQ proteins of
313   *C.pepo* were found to harbor conserved VQ domains and contained the same type of Motifs,
314   implying that in the same branch of the evolutionary tree, the closer the genes of VQ protein are,
315   the more similar the gene structure is.Motif 1 was identified as the core Motif that composes the
316   VQ domain, which was included in all pumpkin VQ proteins, thereby endowing the pumpkin
317   VQ protein with specific biological functions. These results were in agreement with those
318   reported for tomato and soybean species(Ding at al., 2019; Wang at al., 2019). In addition, VQ

319 proteins with similar Motif composition were also found to be located in the same sub-branch on
320 the evolutionary tree, as the Motifs composition between different branches was different.

Expansion mechanism of the *C. pepo* VQ family

322 Genome replication events play an important role in expanding the size of the genome and
323 diversifying gene functions, as replication events can result into genes with new
324 functions(Rensing., 2014). In higher plants, tandem repeat events and chromosome fragment
325 replication events are the main processes contributing for gene family expansion(Wang at al.,
326 2012).Previous studies have shown that diverse WGD events lead to the different sizes of plant
327 genomes(Adams et al.,2005).In this study, 44 *VQ* genes were identified in *C. pepo* L., which is a
328 slightly high number of genes as compared with other species, such as *Arabidopsis* with 34 *VQ*
329 members, or *O. sativa* and *Vitis vinifera* with 39 and 18 *VQs*, respectively(Cheng at al.,2012;Kim
330 at al.,2013a;Wang et al.,2015a). Nevertheless, the number of VQ members in *C. pepo* was lower
331 than initially predicted while considering its genome size of 261 Mb, as compared with the 125,
332 389, 486 Mb of *Arabidopsis*, *O. sativa*, and *V. vinifera*, respectively. Therefore, it can be
333 concluded that there is no necessary connection between the size of the genome and the number
334 of family members. Previous studies have shown that chromosome fragment replication is
335 considered to be the main mechanism of *VQ* gene expansion(Tang at al., 2008); thus, the
336 evolutionary process can explain the number of specific *VQ* genes in a species, not the size of the
337 genome. The main process for the expansion of the *C. pepo* VQ family was found to be
338 chromosome fragment duplication events, which is consistent with the results of previous
339 studies. Among the 44 *CpVQs*, a total of 28 members participated in 21 chromosome fragment
340 duplication events, accounting for 63.6%, and no tandem duplication event was identified. This
341 is consistent with the expansion of the VQ family in *Brassica napus*, which is driven by
342 chromosome fragment duplication, with tandem repeat events as the second driving force(Zou at
343 al., 2021). Gene duplication can produce gene function redundancy, and these repeated genes can
344 develop different gene expression patterns. We calculated the Ka, Ks and Ka/Ks ratios of all
345 para-homologous gene pairs to explore the evolutionary constraints of repeated CpVQs. The
346 Ka/Ks values of most gene pairs were less than 1.0, which indicated that these gene pairs had
347 undergone purification selection pressure.In the present study, some synlinear genes showed
348 different expression patterns, such as *CpVQ3* and *CpVQ33*, *CpVQ9* and *CpVQ21*, including
349 under drought treatment, suggesting that these homologous genes in *C. pepo* may have different
350 functions in regulation the normal growth and development of plants.

Expression patterns of VQ members of *C. pepo*

352 Previous studies have shown that members of the VQ family play an important role in the entire
353 plant development process and respond to various biotic and abiotic stresses. Molecular genetic
354 evidence suggests that the plant VQ protein may be an important regulator of disease resistance
355 and tolerance(Wang at al., 2015). In this study, we analyzed the expression level of *CpVQ* gene
356 in different tissues of *C. pepo* (**Fig.6**). The results showed that most genes were differentially
357 expressed in the tissues we analyzed. It indicates that the *CpVQ* gene may play an important role
358 in the growth and development of these organs or tissues. In addition, based on transcriptome

359   data, the expression *VQs* in *C. pepo* leaves under powdery mildew treatments at different
360   timepoints was evaluated(as observed in **Fig. 7**). Most VQ members showed a significant decline
361   in expression after 12 h of powdery mildew treatment, however, the expression patterns of the
362   third group is different from the other groups. In the third group, except *CpVQ20*, all were
363   significantly up-regulated upon powdery mildew infection, and 3 genes (*CpVQ18, 27* and *38*)
364   had higher expression at 24h, which may be because these genes responded late to powdery
365   mildew stress.Studies have shown that *SIB1(AtVQ23)* and *SIB2(AtVQ16)* function as activators
366   in plant defense against necrotrophic pathogens(Lai et al.,2011), so the *CpVQ* genes classified in
367   the third group may have the same function.In contrast, after 24 hours of powdery mildew
368   treatment, the expression of most members in other groups decreased significantly.Similar
369   results have been reported in grapevine that *VvVQ* genes are quickly responsive to powdery
370   mildew stresses(Wang et al.,2015).Therefore, the members of the VQ family in *C. pepo* may
371   play a key role in the powdery mildew signaling pathway.Moreover, analysis of *VQs* expression
372   under different abiotic stresses further demonstrated that the members of the VQ family are also
373   significantly induced under different stresses.In this study, six *CpVQ* genes (*CpVQ3, 8, 9, 16, 22*
374   and *39*) significantly upregulated during drought stress(as observed in **Fig. 9**), which results are
375   similar to the *OsVQ* genes that twenty-two *OsVQ* genes are upregulated under drought
376   stress(Kim et al.,2013a).Under cold stress(as observed in **Fig. 10**), five *CaVQ* genes (*CaVQ1, 22,*
377   *33, 39* and *40*) were significantly induced.Similar results have been reported in Chinese cabbage
378   that *BrVQ* genes are quickly responsive to cold stresses(Zhang et al.,2015).In addition, the *VQ*
379   genes are also sensitive to salt changes. Three *CpVQ* genes (*CaVQ*3, 9 and 14) were upregulated
380   salt treatment(as observed in **Fig. 11**).Similar changes occurred in Arabidopsis. *AtVQ9* and
381   *AtVQ15* were significantly expressed under salt stress.Nine *CaVQ* genes (*CaVQ1, 12, 16, 22, 26,*
382   *33, 34, 39* and *40*) were upregulated at waterlogging treatment(as observed in **Fig. 12**).In
383   summary, CpVQ members may be involved in regulating the response of plants to various
384   abiotic stresses and powdery mildew stress, and their response mechanisms may be complex and
385   diverse.
386   In summary, this study has systematically analyzed the evolutionary relationship,
387   conserved structure, and expression patterns of the members of the VQ family of *C. pepo* at the
388   whole genome level. The selection of candidate genes can provide reference for future
389   investigations.
390

Conclusions

392   In conclusion, this study provides the first comprehensive and systematic analysis of the 44 *VQ*
393   genes identified in *C. pepo* genome. All *CpVQ* genes can be divided into eight groups (I–VIII),
394   and were found to expand by chromosome fragment duplication events. RNA-seq analysis
395   further showed that half of the *VQs* have significantly different expression patterns at different
396   timepoints of powdery mildew infection.Therefore, the members of the VQ family in *C. pepo*
397   may play a key role in the powdery mildew signaling pathway. Genetic analysis data further

398 confirmed that the VQ family members respond to different abiotic stresses. Taken together,
399 these findings provide a theoretical basis for further research on the functions of *CpVQs*.
400

References

401
402 Adams et al.(2005)Adams KL, Wendel JF. Polyploidy and genome evolution in plants. Curr
403 Opin Plant Biol. 2005 Apr;8(2):135-41. doi: 10.1016/j.pbi.2005.01.001. PMID: 15752992.
404 Bailey et al.(2015)Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids
405 Res. 2015 Jul 1;43(W1):W39-49. doi: 10.1093/nar/gkv416. Epub 2015 May 7. PMID: 25953851;
406 PMCID: PMC4489269.
407 Chen et al.(2020)Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools:
408 An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant.
409 2020 Aug 3;13(8):1194-1202. doi: 10.1016/j.molp.2020.06.009. Epub 2020 Jun 23. PMID:
410 32585190.
411 Chen et al.(2018)Chen J, Wang H, Li Y, Pan J, Hu Y, Yu D. Arabidopsis VQ10 interacts with
412 WRKY8 to modulate basal defense against Botrytis cinerea. J Integr Plant Biol. 2018
413 Oct;60(10):956-969. doi: 10.1111/jipb.12664. Epub 2018 Jul 10. PMID: 29727045.
414 Cheng et al.(2012)Cheng Y, Zhou Y, Yang Y, Chi YJ, Zhou J, Chen JY, Wang F, Fan B, Shi
415 K, Zhou YH, Yu JQ, Chen Z. Structural and functional analysis of VQ motif-containing proteins
416 in Arabidopsis as interacting proteins of WRKY transcription factors. Plant Physiol. 2012
417 Jun;159(2):810-25. doi: 10.1104/pp.112.196816. Epub 2012 Apr 24. PMID: 22535423; PMCID:
418 PMC3375943.
419 Cheng et al.(2020)Chen P, Wei F, Cheng S, Ma L, Wang H, Zhang M, Mao G, Lu J, Hao P,
420 Ahmad A, Gu L, Ma Q, Wu A, Wei H, Yu S. A comprehensive analysis of cotton VQ gene
421 superfamily reveals their potential and extensive roles in regulating cotton abiotic stress. BMC
422 Genomics. 2020 Nov 16;21(1):795. doi: 10.1186/s12864-020-07171-z. PMID: 33198654;
423 PMCID: PMC7667805.
424 Chi et al.(2013)Chi Y, Yang Y, Zhou Y, Zhou J, Fan B, Yu JQ, Chen Z. Protein-protein
425 interactions in the regulation of WRKY transcription factors. Mol Plant. 2013 Mar;6(2):287-300.
426 doi: 10.1093/mp/sst026. Epub 2013 Mar 2. PMID: 23455420.
427 Caoet al.(2018)Cao Y, Meng D, Abdullah M, Jin Q, Lin Y, Cai Y. Genome Wide
428 Identification, Evolutionary, and Expression Analysis of VQ Genes from Two Pyrus Species.
429 Genes (Basel). 2018 Apr 23;9(4):224. doi: 10.3390/genes9040224. PMID: 29690608; PMCID:
430 PMC5924566.
431 Ding et al.(2019)Ding H, Yuan G, Mo S, Qian Y, Wu Y, Chen Q, Xu X, Wu X, Ge C.
432 Genome-wide analysis of the plant-specific VQ motif-containing proteins in tomato (Solanum
433 lycopersicum) and characterization of SlVQ6 in thermotolerance. Plant Physiol Biochem. 2019
434 Oct;143:29-39. doi: 10.1016/j.plaphy.2019.08.019. Epub 2019 Aug 27. PMID: 31479880.
435 Fujita et al.(2006)Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-
436 Shinozaki K, Shinozaki K. Crosstalk between abiotic and biotic stress responses: a current view

437   from the points of convergence in the stress signaling networks. Curr Opin Plant Biol. 2006

438   Aug;9(4):436-42. doi: 10.1016/j.pbi.2006.05.014. Epub 2006 Jun 8. PMID: 16759898.

439   Hu et al.(2013)Hu Y, Chen L, Wang H, Zhang L, Wang F, Yu D. Arabidopsis transcription

440   factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity

441   stress tolerance. Plant J. 2013 Jun;74(5):730-45. doi: 10.1111/tpj.12159. Epub 2013 Apr 1.

442   PMID: 23451802.

443   Jing et al.(2018)Jiang SY, Sevugan M, Ramachandran S. Valine-glutamine (VQ) motif coding

444   genes are ancient and non-plant-specific with comprehensive expression regulation by various

445   biotic and abiotic stresses. BMC Genomics. 2018 May 9;19(1):342. doi: 10.1186/s12864-018-

446   4733-7. PMID: 29743038; PMCID: PMC5941492.

447   Jing et al.(2015)Jing Y, Lin R. The VQ Motif-Containing Protein Family of Plant-Specific

448   Transcriptional Regulators. Plant Physiol. 2015 Sep;169(1):371-8. doi: 10.1104/pp.15.00788.

449   Epub 2015 Jul 28. PMID: 26220951; PMCID: PMC4577417.

450   Kim et al.(2013a)Kim DY, Kwon SI, Choi C, Lee H, Ahn I, Park SR, Bae SC, Lee SC, Hwang

451   DJ. Expression analysis of rice VQ genes in response to biotic and abiotic stresses. Gene. 2013

452   Oct 25;529(2):208-14. doi: 10.1016/j.gene.2013.08.023. Epub 2013 Aug 16. PMID: 23958655.

453   Kim et al.(2013b)Kim MY, Yang JK, Lee T, Lee SH. Divergence of flowering-related genes

454   in three legume species. Plant Genome. 2013b;6(3):841‒856. doi:

455   10.3835/plantgenome2013.03.0008.

456   Lai et al.(2011)Lai Z, Li Y, Wang F, Cheng Y, Fan B, Yu JQ, Chen Z. Arabidopsis sigma

457   factor binding proteins are activators of the WRKY33 transcription factor in plant defense. Plant

458   Cell. 2011 Oct;23(10):3824-41. doi: 10.1105/tpc.111.090571. Epub 2011 Oct 11. PMID:

459   21990940; PMCID: PMC3229152.

460   Lei et al.(2017)Lei R, Li X, Ma Z, Lv Y, Hu Y, Yu D. Arabidopsis WRKY2 and WRKY34

461   transcription factors interact with VQ20 protein to modulate pollen development and function.

462   Plant J. 2017 Sep;91(6):962-976. doi: 10.1111/tpj.13619. Epub 2017 Jul 20. PMID: 28635025.

463   Lei et al.(2018)Lei R, Ma Z, Yu D. WRKY2/34-VQ20 Modules in Arabidopsis thaliana

464   Negatively Regulate Expression of a Trio of Related MYB Transcription Factors During Pollen

465   Development. Front Plant Sci. 2018 Mar 19;9:331. doi: 10.3389/fpls.2018.00331. PMID:

466   29616054; PMCID: PMC5867338.

467   Livak et al.(2001)Livak KJ, Schmittgen TD. Analysis of relative gene expression data using

468   real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001 Dec;25(4):402-

469   8. doi: 10.1006/meth.2001.1262. PMID: 11846609.

470   Liu et al.(2020)Liu C, Liu H, Zhou C, Timko MP. Genome-Wide Identification of the VQ

471   Protein Gene Family of Tobacco (Nicotiana tabacum L.) and Analysis of Its Expression in

472   Response to Phytohormones and Abiotic and Biotic Stresses. Genes (Basel). 2020 Mar

473   7;11(3):284. doi: 10.3390/genes11030284. PMID: 32156048; PMCID: PMC7140788.

474   Liu et al.(2021)Liu H, Lyu HM, Zhu K, Van de Peer Y, Max Cheng ZM. The emergence and

475   evolution of intron-poor and intronless genes in intron-rich plant gene families. Plant J. 2021

476 Feb;105(4):1072-1082. doi: 10.1111/tpj.15088. Epub 2021 Feb 9. PMID: 33217085; PMCID:

477 PMC7116809.

478 Louhichi et al.(2011)Louhichi A, Fourati A, Rebaï A. IGD: a resource for intronless genes in

479 the human genome. Gene. 2011 Nov 15;488(1-2):35-40. doi: 10.1016/j.gene.2011.08.013. Epub

480 2011 Sep 2. PMID: 21914464.

481 Luo et al.(2005)Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A. MINISEED3

482 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE

483 gene, are regulators of seed size in Arabidopsis. Proc Natl Acad Sci U S A. 2005 Nov

484 29;102(48):17531-6. doi: 10.1073/pnas.0508418102. Epub 2005 Nov 17. PMID: 16293693;

485 PMCID: PMC1297679.

486 Obrero et al.(2011)Obrero A, Die JV, Román B, Gómez P, Nadal S, González-Verdejo CI.

487 Selection of reference genes for gene expression studies in zucchini (Cucurbita pepo) using

488 qPCR. J Agric Food Chem. 2011 May 25;59(10):5402-11. doi: 10.1021/jf200689r. Epub 2011

489 Apr 22. PMID: 21476515.

490 Perruc et al.(2004)Perruc E, Charpenteau M, Ramirez BC, Jauneau A, Galaud JP, Ranjeva R,

491 Ranty B. A novel calmodulin-binding protein functions as a negative regulator of osmotic stress

492 tolerance in Arabidopsis thaliana seedlings. Plant J. 2004 May;38(3):410-20. doi:

493 10.1111/j.1365-313X.2004.02062.x. PMID: 15086802.

494 Punta et al. (2004) Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N,

495 Forslund K, Ceric G, Clements J. The Pfam protein families database. Nucleic Acids Research.

496 2004;28(1):263–266. doi: 10.1093/nar/gkh121.

497 Rensing et al.(2014)Rensing SA. Gene duplication as a driver of plant morphogenetic

498 evolution. Curr Opin Plant Biol. 2014 Feb;17:43-8. doi: 10.1016/j.pbi.2013.11.002. Epub 2013

499 Nov 28. PMID: 24507493.

500 Song et al.(2016)Song W, Zhao H, Zhang X, Lei L, Lai J. Genome-Wide Identification of VQ

501 Motif-Containing Proteins and their Expression Profiles Under Abiotic Stresses in Maize. Front

502 Plant Sci. 2016 Jan 5;6:1177. doi: 10.3389/fpls.2015.01177. PMID: 26779214; PMCID:

503 PMC4700186.

504 Tamura et al.(2011)Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.

505 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary

506 distance, and maximum parsimony methods. Mol Biol Evol. 2011 Oct;28(10):2731-9. doi:

507 10.1093/molbev/msr121. Epub 2011 May 4. PMID: 21546353; PMCID: PMC3203626.

508 Tang et al.(2008)Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and

509 collinearity in plant genomes. Science. 2008 Apr 25;320(5875):486-8. doi:

510 10.1126/science.1153917. PMID: 18436778.

511 Wang et al.(2017)Wang J, Sun P, Li Y, Liu Y, Yu J, Ma X, Sun S, Yang N, Xia R, Lei T, Liu

512 X, Jiao B, Xing Y, Ge W, Wang L, Wang Z, Song X, Yuan M, Guo D, Zhang L, Zhang J, Jin D,

513 Chen W, Pan Y, Liu T, Jin L, Sun J, Yu J, Cheng R, Duan X, Shen S, Qin J, Zhang MC,

514 Paterson AH, Wang X. Hierarchically Aligning 10 Legume Genomes Establishes a Family-Level

515   Genomics Platform. Plant Physiol. 2017 May;174(1):284-300. doi: 10.1104/pp.16.01981. Epub
516   2017 Mar 21. PMID: 28325848; PMCID: PMC5411148.
517   Wang et al.(2010a)Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit
518   incorporating gamma-series methods and sliding window strategies. Genomics Proteomics
519   Bioinformatics. 2010 Mar;8(1):77-80. doi: 10.1016/S1672-0229(10)60008-3. PMID: 20451164;
520   PMCID: PMC5054116.
521   Wang et al.(2010)Wang A, Garcia D, Zhang H, Feng K, Chaudhury A, Berger F, Peacock
522   WJ, Dennis ES, Luo M. The VQ motif protein IKU1 regulates endosperm growth and seed size
523   in Arabidopsis. Plant J. 2010 Aug;63(4):670-9. doi: 10.1111/j.1365-313X.2010.04271.x. PMID:
524   20545893.
525   Wang et al.(2015)Wang H, Hu Y, Pan J, Yu D. Arabidopsis VQ motif-containing proteins
526   VQ12 and VQ29 negatively modulate basal defense against Botrytis cinerea. Sci Rep. 2015 Sep
527   23;5:14185. doi: 10.1038/srep14185. PMID: 26394921; PMCID: PMC4585807.
528   Wang et al.(2015a)Wang M, Vannozzi A, Wang G, Zhong Y, Corso M, Cavallini E, Cheng
529   ZM. A comprehensive survey of the grapevine VQ gene family and its transcriptional correlation
530   with WRKY proteins. Front Plant Sci. 2015 Jun 12;6:417. doi: 10.3389/fpls.2015.00417. PMID:
531   26124765; PMCID: PMC4464145.
532   Wang et al.(2019)Wang Y, Jiang Z, Li Z, Zhao Y, Tan W, Liu Z, Cui S, Yu X, Ma J, Wang
533   G, Li W. Genome-wide identification and expression analysis of the VQ gene family in soybean
534   (Glycine max). PeerJ. 2019 Aug 21;7:e7509. doi: 10.7717/peerj.7509. PMID: 31497394;
535   PMCID: PMC6708371.
536   Wang et al.(2012a)Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H,
537   Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary
538   analysis of gene synteny and collinearity. Nucleic Acids Res. 2012 Apr;40(7):e49. doi:
539   10.1093/nar/gkr1293. Epub 2012 Jan 4. PMID: 22217600; PMCID: PMC3326336.
540   Wang et al.(2012)Wang Y, Wang X, Paterson AH. Genome and gene duplications and gene
541   expression divergence: a view from plants. Ann N Y Acad Sci. 2012 May;1256:1-14. doi:
542   10.1111/j.1749-6632.2011.06384.x. Epub 2012 Jan 18. PMID: 22257007.
543   Wilkins et al.(1999)Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel
544   RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods
545   Mol Biol. 1999;112:531-52. doi: 10.1385/1-59259-584-7:531. PMID: 10027275.
546   Wray et al.(2003)Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV,
547   Romano LA. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 2003
548   Sep;20(9):1377-419. doi: 10.1093/molbev/msg140. Epub 2003 May 30. PMID: 12777501.
549   Ye et al.(2016)Ye YJ, Xiao YY, Han YC, Shan W, Fan ZQ, Xu QG, Kuang JF, Lu WJ,
550   Lakshmanan P, Chen JY. Banana fruit VQ motif-containing protein5 represses cold-responsive
551   transcription factor MaWRKY26 involved in the regulation of JA biosynthetic genes. Sci Rep.
552   2016 Mar 23;6:23632. doi: 10.1038/srep23632. PMID: 27004441; PMCID: PMC4804309.

553   Zhang et al.(2015)Zhang G, Wang F, Li J, Ding Q, Zhang Y, Li H, Zhang J, Gao J. Genome-
554   Wide Identification and Analysis of the VQ Motif-Containing Protein Family in Chinese
555   Cabbage (Brassica rapa L. ssp. Pekinensis). Int J Mol Sci. 2015 Dec 2;16(12):28683-704. doi:
556   10.3390/ijms161226127. PMID: 26633387; PMCID: PMC4691074.
557   Zou et al.(2021)Zou Z, Liu F, Huang S, Fernando WGD. Genome-Wide Identification and
558   Analysis of the Valine-Glutamine Motif-Containing Gene Family in Brassica napus and
559   Functional Characterization of BnMKS1 in Response to Leptosphaeria maculans.
560   Phytopathology. 2021 Feb;111(2):281-292. doi: 10.1094/PHYTO-04-20-0134-R. Epub 2021 Feb
561   8. PMID: 32804045.
562
563

Figure 1

Phylogenetic tree analysis of the *VQ* genes in *C.pepo and Arabidopsis thaliana*.

The clusters were designated as group I–VIII and indicated in a specific color.

Figure 2

Phylogenetic tree, conserved motifs and gene structure in *CpVQs*.

(A) Phylogenetic relationships (B) Conserved motifs of the *CpVQs*. Each motif is represented by a number in colored box. (C) Exon/intron structures of *CpVQ* genes.

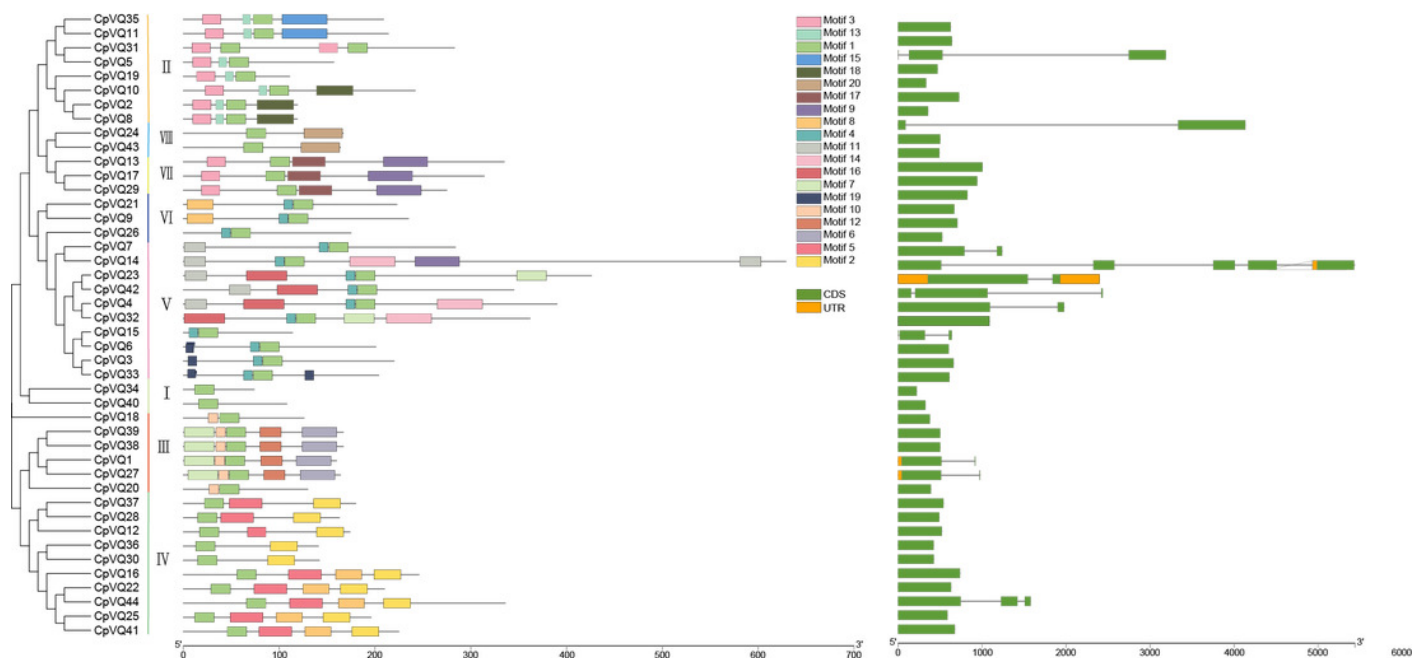



Figure 3

Chromosome location in *C.pepo.*

Chromosome numbers were indicated above each chromosome. The size of a chromosome was indicated by its relative length.

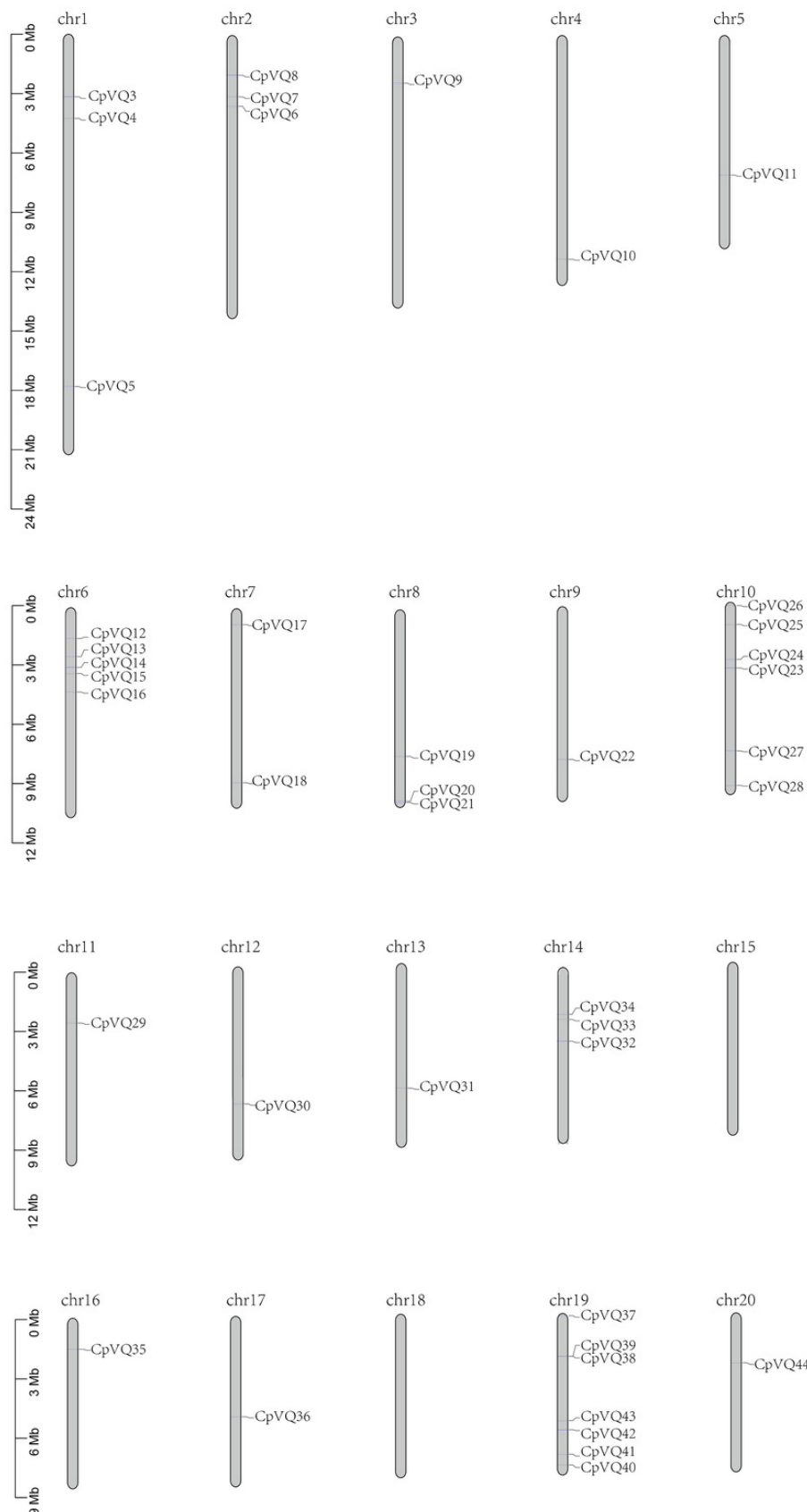



Figure 4

Syntenic analysis of *VQ* genes

Gray lines in the background indicate the collinear blocks within the *C.pepo* and oneself genomes, while the dark grey lines highlight the syntenic *VQ* gene pairs.

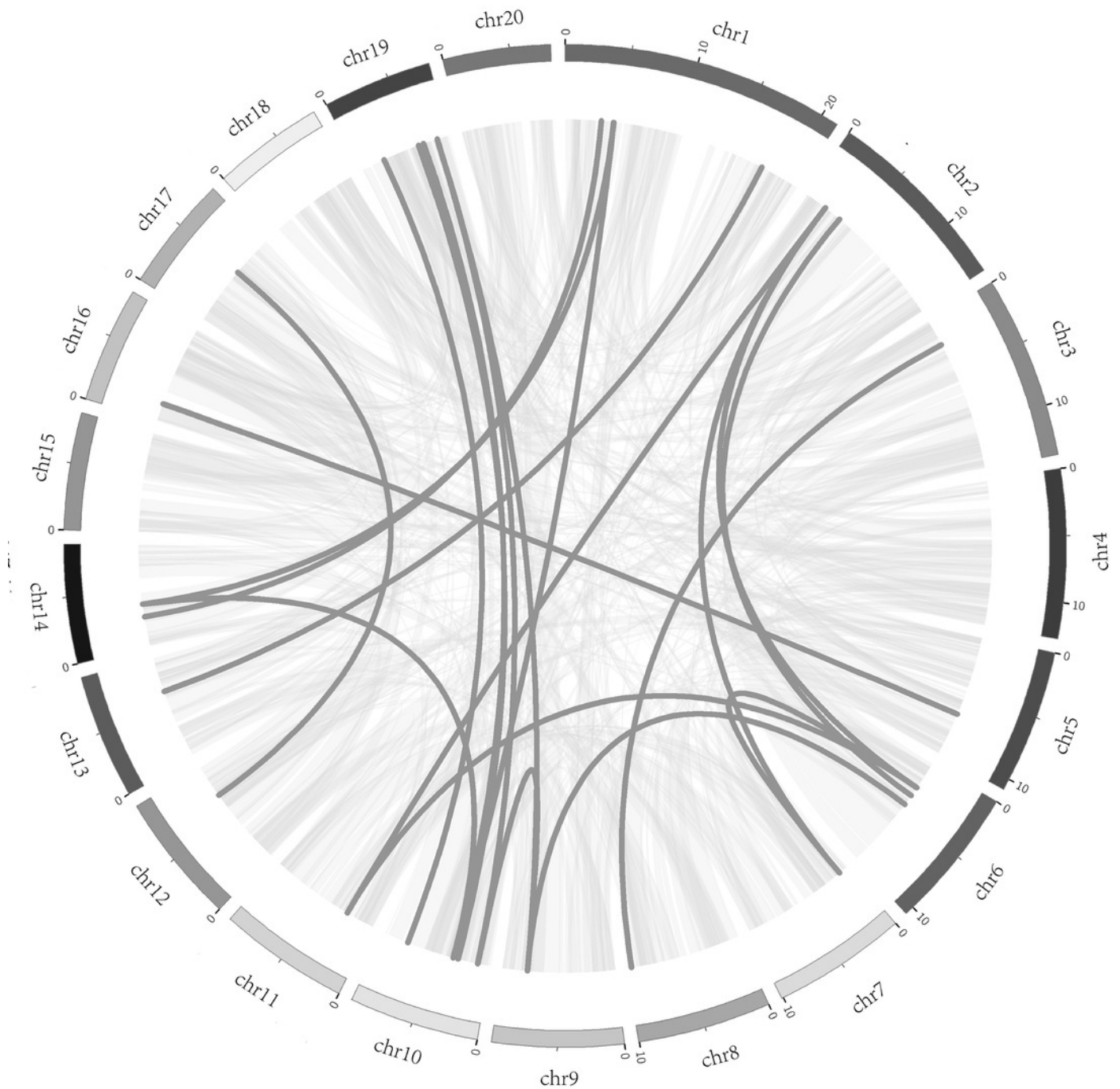



Figure 5

Synteny analysis of *VQ* genes between *C.pepo* and plant species.

Synteny analysis of the *VQ* genes between(A)*C.pepo* and *Arabidopsis thaliana*;(B)*C.pepo* and *Cucumis melo*;(C) *C.pepo* and *Oryza sativa*;(D) *C.pepo* and *Zea mays*.Gray lines in the background indicate the collinear blocks within the *C.pepo* and other plant genomes, while the black lines highlight the syntenic *VQ* gene pairs.

A

Arabidopsis thaliana

Chr1 Chr2 Chr3 Chr4 Chr5

Cucurbita peop

chr01 chr02 chr03 chr04 chr05 chr06 chr07 chr08 chr09 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20

B

Cucumis melo

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13

Cucurbita peop

chr01 chr02 chr03 chr04 chr05 chr06 chr07 chr08 chr09 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20

C

Oryza sativa

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12

Cucurbita peop

chr01 chr02 chr03 chr04 chr05 chr06 chr07 chr08 chr09 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20

D

Zea mays

chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10

Cucurbita peop

chr01 chr02 chr03 chr04 chr05 chr06 chr07 chr08 chr09 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20

Figure 6

Expression analysis of the *CpVQ* genes in different tissues of *C.pepo*

The surveyed tissues include root,stem,leaf.The $2^{-\Delta\Delta Ct}$ method was used to calculate the expression levels of target genes in different tissues.

Different tissues

A VQ1

B VQ3

C VQ8

D VQ9

E VQ12

F VQ14

G VQ16

H VQ21

I VQ22

J VQ26

K VQ33

L VQ34

M VQ36

N VQ39

O VQ40

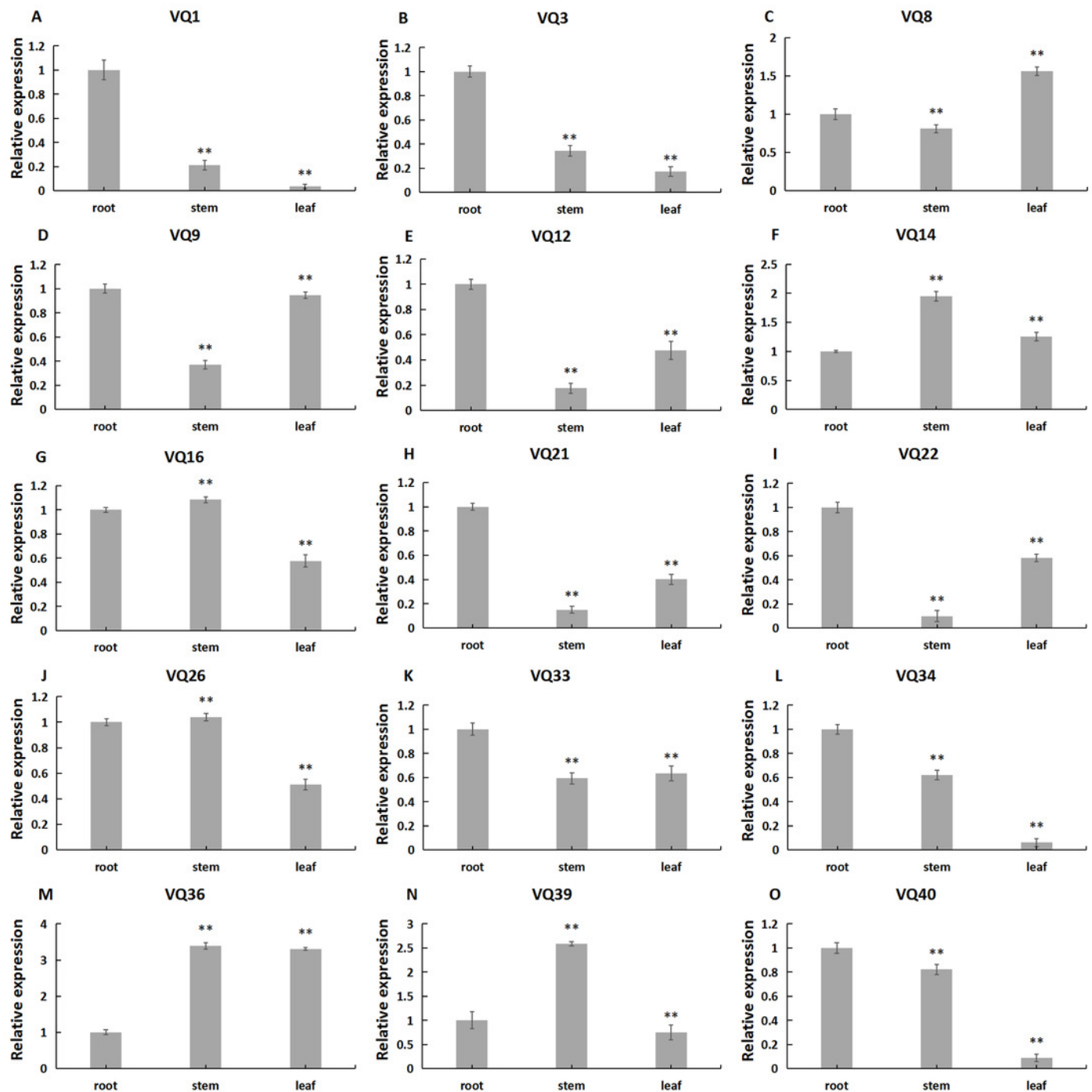



Figure 7

Expression profiles of *CpVQs* under powdery mildew stress based on the RNA-seq data.

The gene expression values are square-root transformed fragments per kilo-bases per million mapped reads (FPKM). Different colors in map represent gene transcript abundance values as shown in the color bar.

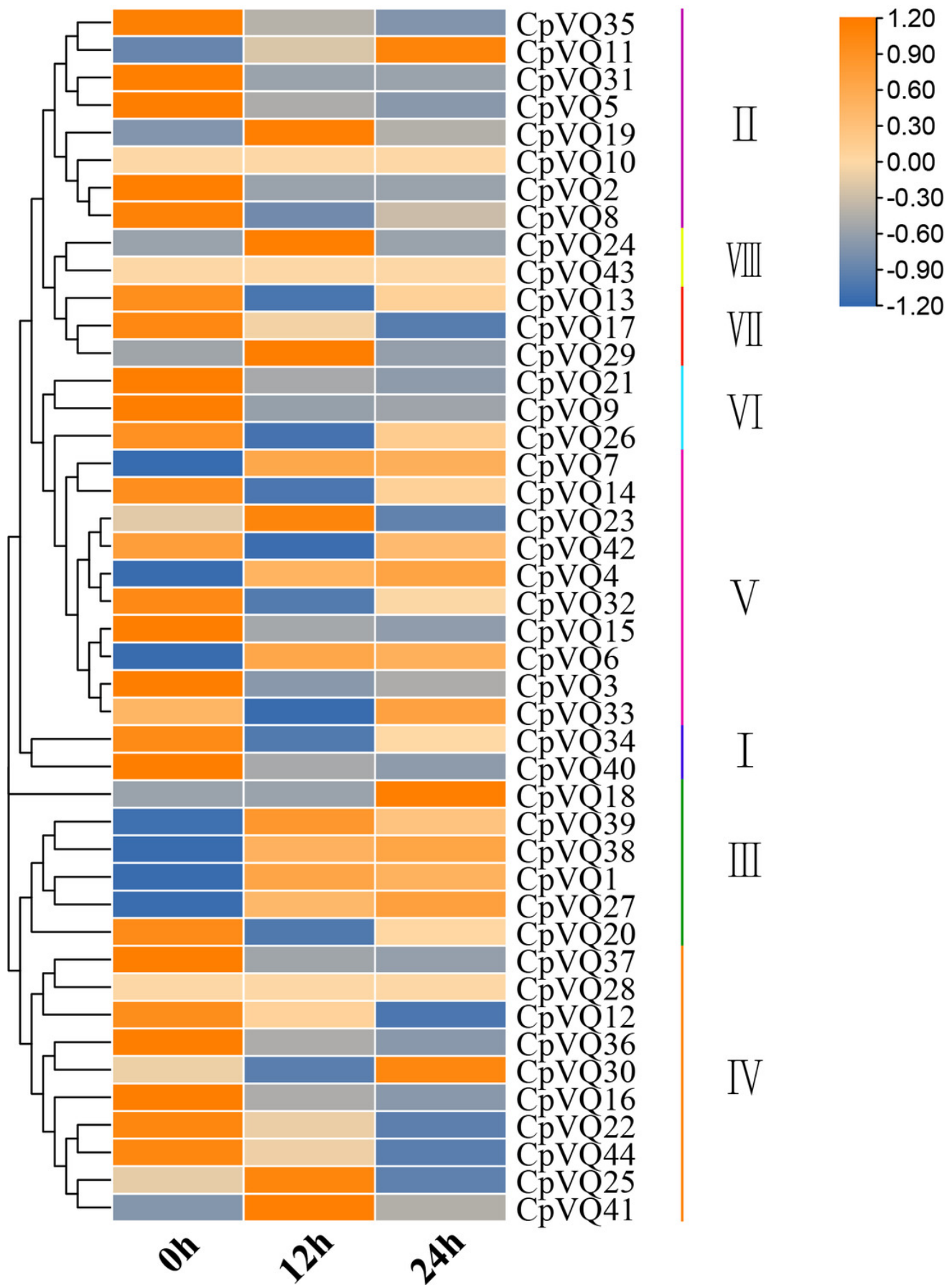



Figure 8

qRT-PCR validation of *VQ* genes in the response to powdery mildew treatment.

Stress treatments and time course are described in "Materials & Methods". (A–O) represent different genes that were evaluated by qRT-PCR. Asterisks indicate statistically significant differences between the stressed samples and counterpart controls ($*p < 0.05$, $**p < 0.01$).

Powdery mildew

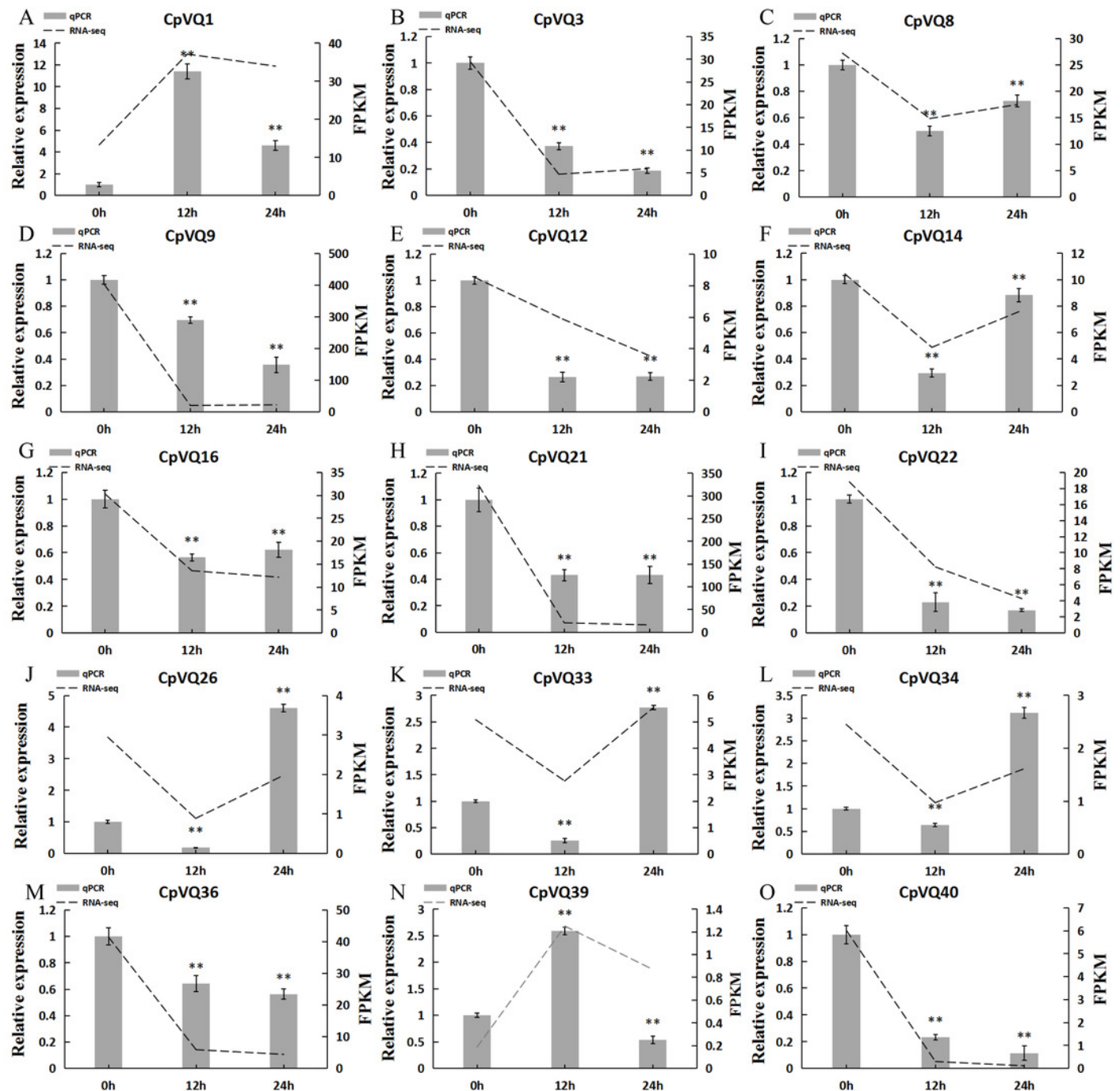



Figure 9

qRT-PCR validation of *VQ* genes in the response to drought treatment.

Stress treatments and time course are described in "Materials & Methods". (A–O) represent different genes that were evaluated by qRT-PCR. Asterisks indicate statistically significant differences between the stressed samples and counterpart controls ($*p < 0.05$, $**p < 0.01$).

Drought

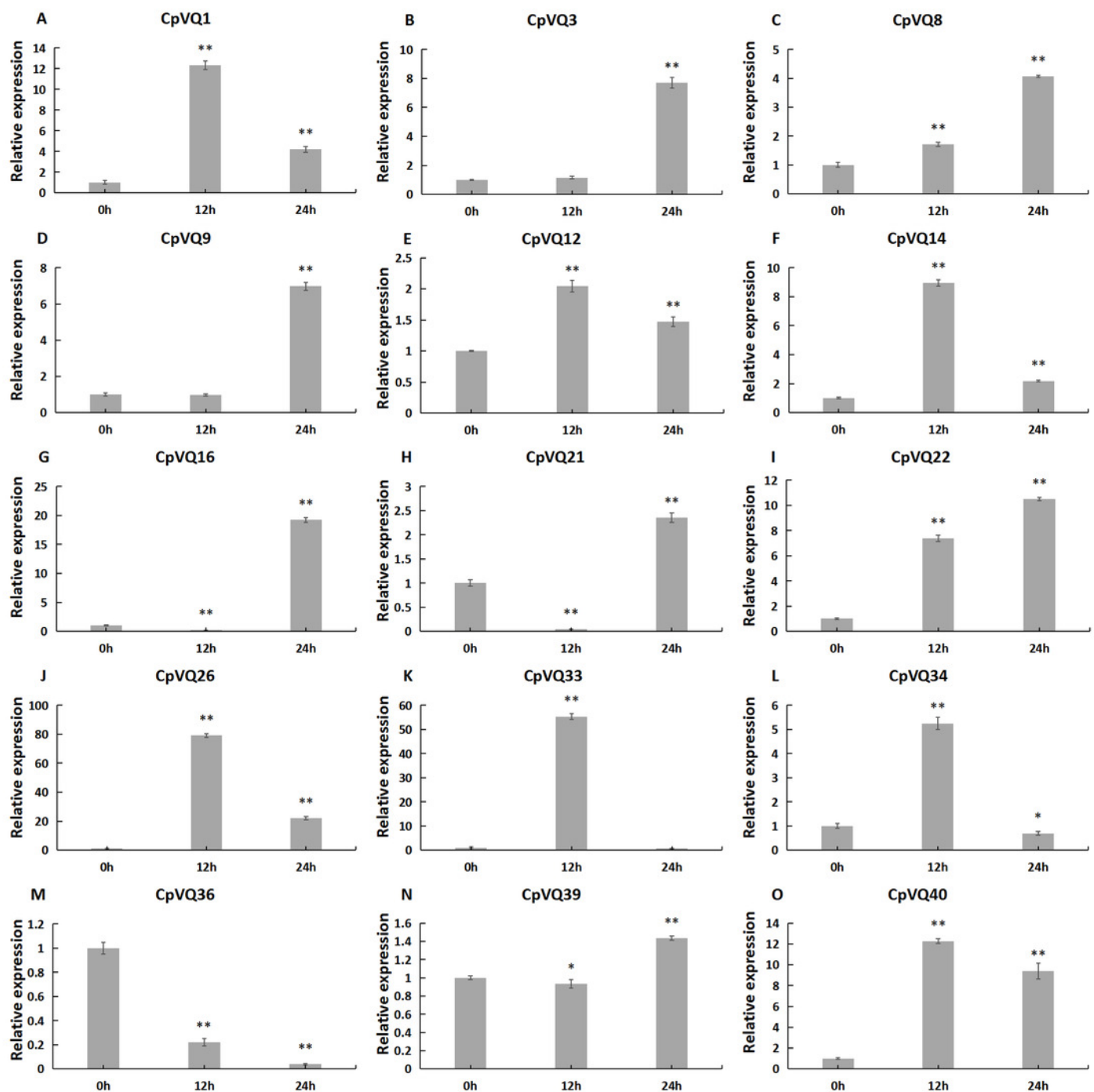



Figure 10

qRT-PCR validation of *VQ* genes in the response to cold treatment.

Stress treatments and time course are described in "Materials & Methods". (A–O) represent different genes that were evaluated by qRT-PCR. Asterisks indicate statistically significant differences between the stressed samples and counterpart controls ($*p < 0.05$, $**p < 0.01$).

Cold

A CpVQ1

B CpVQ3

C CpVQ8

D CpVQ9

E CpVQ12

F CpVQ14

G CpVQ16

H CpVQ21

I CpVQ22

J CpVQ26

K CpVQ33

L CpVQ34

M CpVQ36

N CpVQ39

O CpVQ40

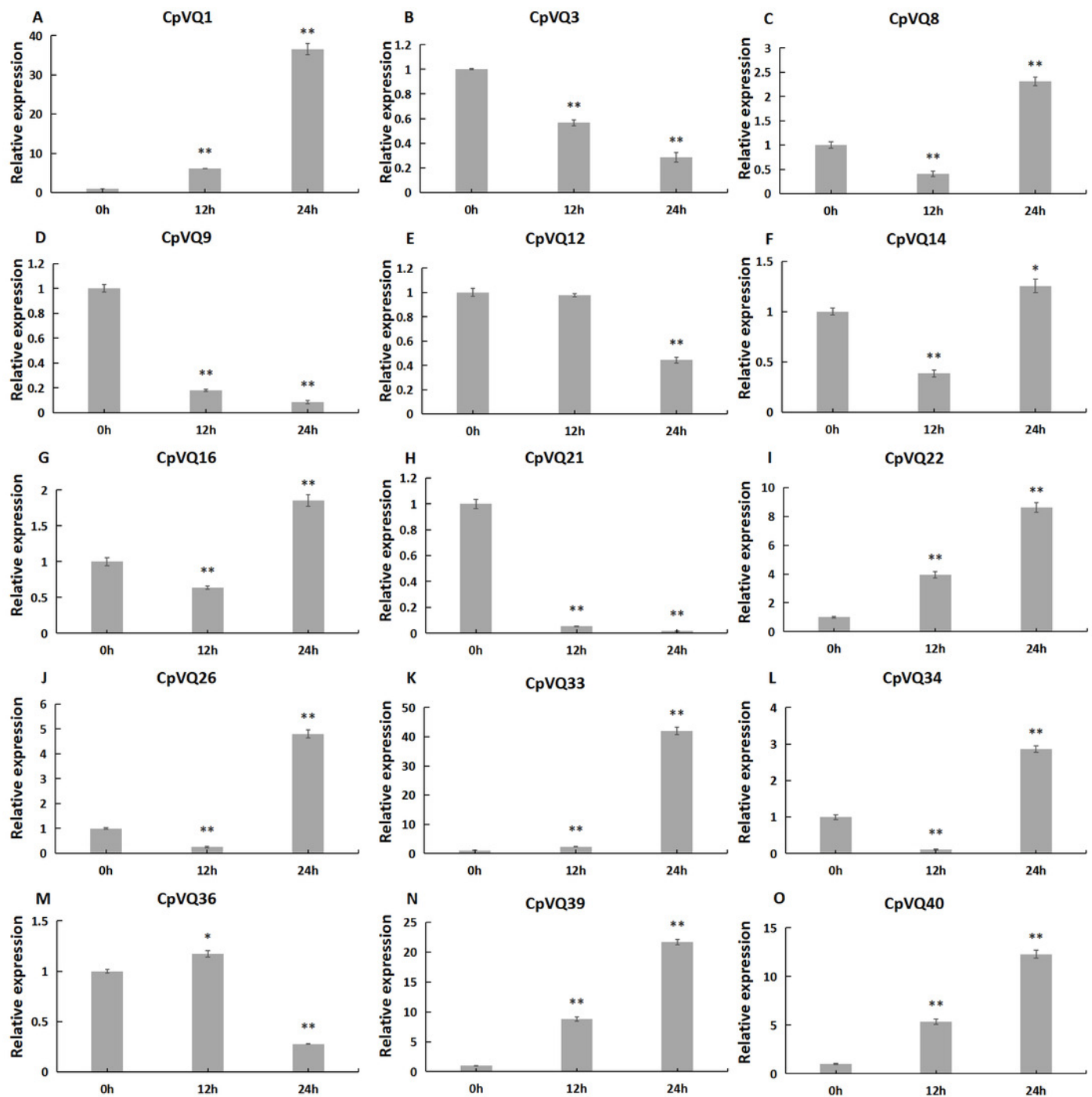



Figure 11

qRT-PCR validation of *VQ* genes in the response to salt treatment.

Stress treatments and time course are described in "Materials & Methods". (A–O) represent different genes that were evaluated by qRT-PCR. Asterisks indicate statistically significant differences between the stressed samples and counterpart controls ($*p < 0.05$, $**p < 0.01$).

Salt

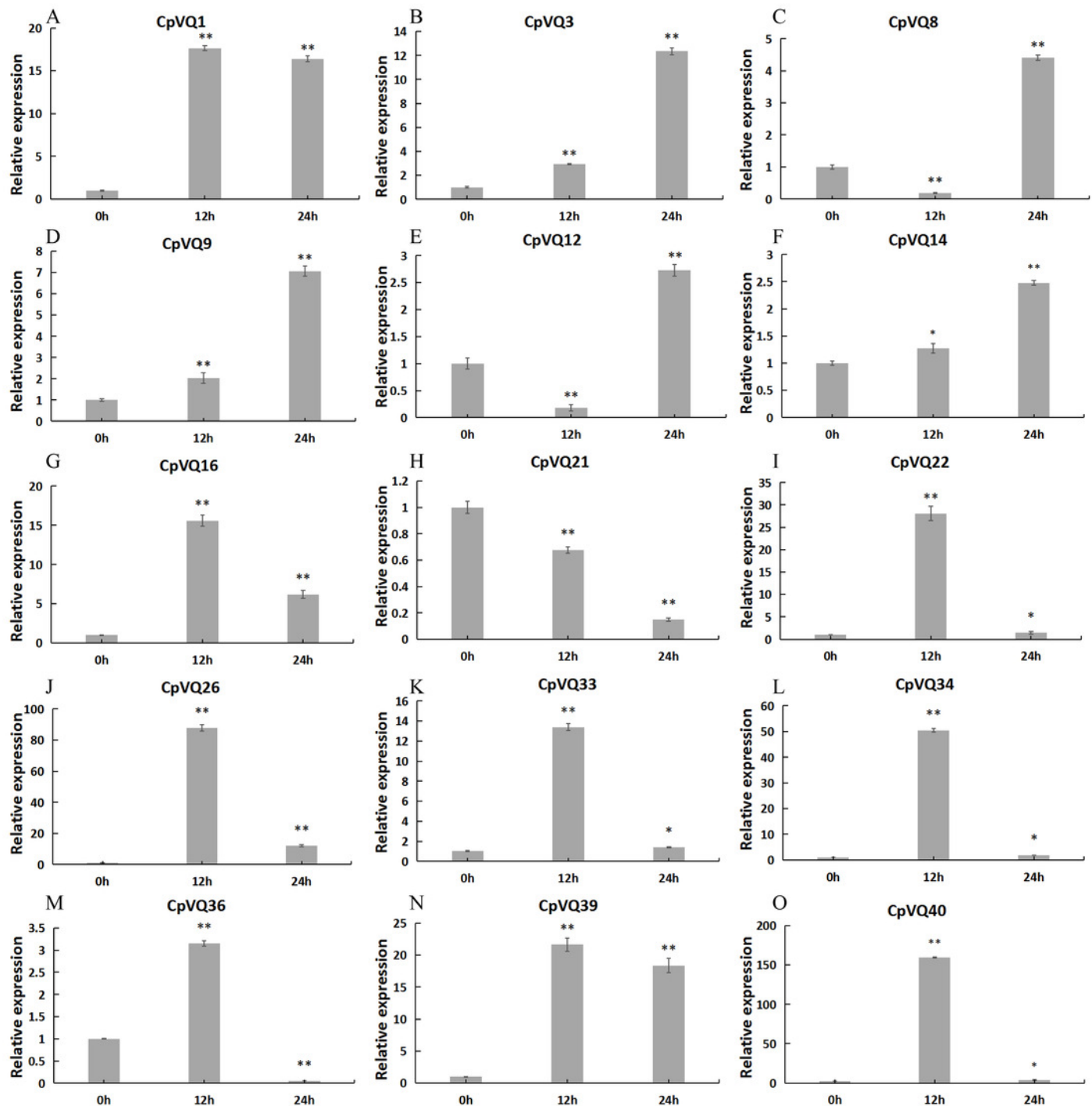
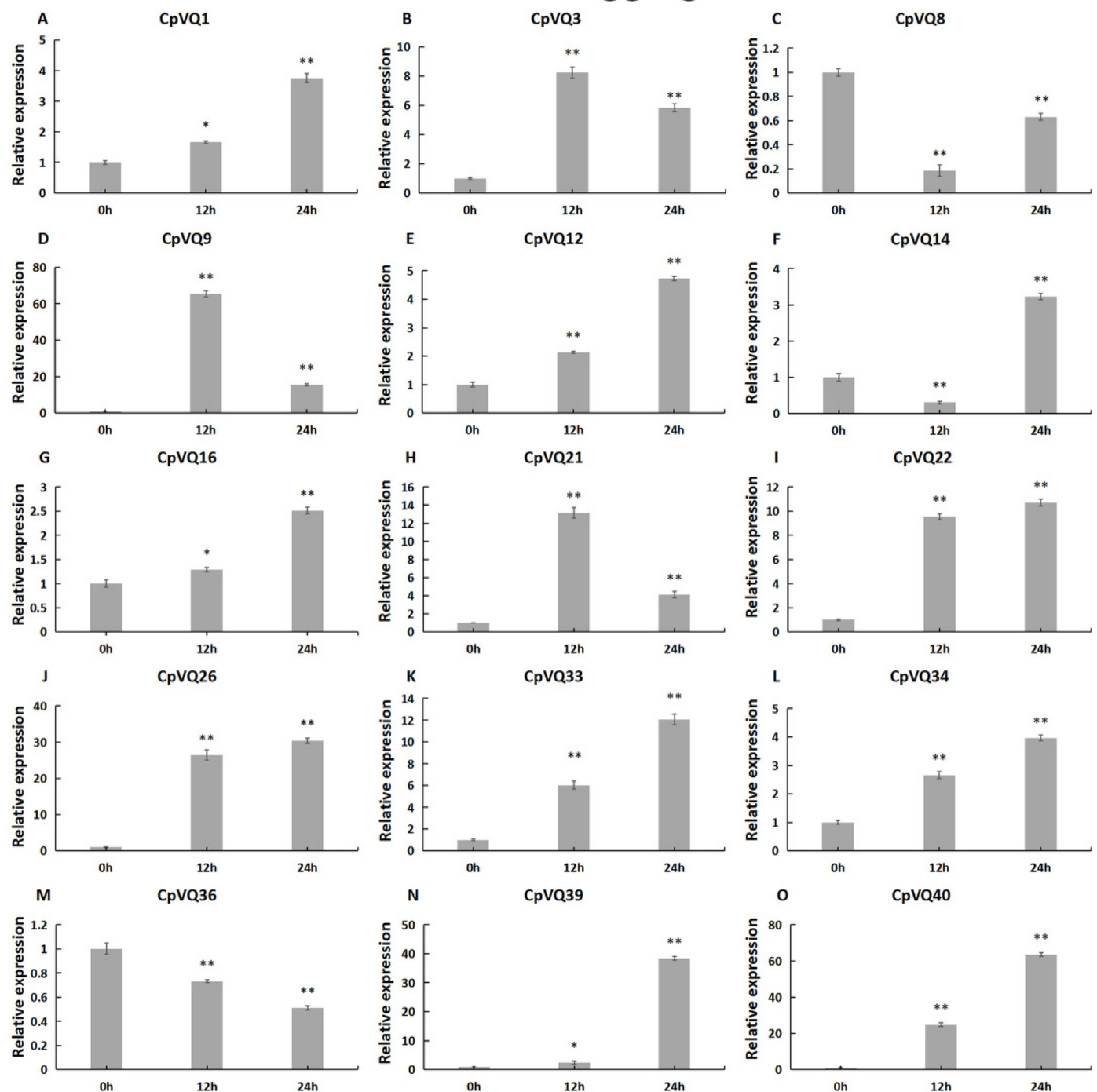



Figure 12

qRT-PCR validation of *VQ* genes in the response to waterlogging treatment.

Stress treatments and time course are described in "Materials & Methods". (A–O) represent different genes that were evaluated by qRT-PCR. Asterisks indicate statistically significant differences between the stressed samples and counterpart controls ($*p < 0.05$, $**p < 0.01$).

Waterlogging

List of all *VQ* genes identified in *C.pepo*

1 **Table 1**

2 **List of all VQ genes identified in *Cucurbita pepo L.***

Gene name	Gene locus	Chromosome location	Length (aa)	pI	Molecular weight (KDa)	Family group
CpVQ1	Cp4. 1LG00g04520	16693604–16694523	160	5. 44	17313. 57	III
CpVQ2	Cp4. 1LG00g15310	42412302–42412658	118	9. 83	13171. 89	II
CpVQ3	Cp4. 1LG01g01160	LG01:3157885–3158544	219	9. 03	22938. 39	V
CpVQ4	Cp4. 1LG01g09330	LG01:4248892–4250868	389	6. 8	41553. 79	V
CpVQ5	Cp4. 1LG01g21070	LG01:17809037–17809507	156	8. 13	17422. 62	II
CpVQ6	Cp4. 1LG02g02700	LG02:3577192–3577794	200	6. 91	21690. 73	V
CpVQ7	Cp4. 1LG02g03510	LG02:3102280–3103518	283	9. 16	30711. 29	V
CpVQ8	Cp4. 1LG02g05130	LG02:2005092–2009226	118	9. 83	13178. 19	II
CpVQ9	Cp4. 1LG03g04140	LG03:2346972–2347676	234	6. 53	25197. 91	VI
CpVQ10	Cp4. 1LG04g14070	LG04:11310784–11311509	241	5. 94	25762. 45	II
CpVQ11	Cp4. 1LG05g10550	LG05:7053658–7054299	213	6. 51	22984. 1	II
CpVQ12	Cp4. 1LG06g02740	LG06:1556739–1557260	173	10. 13	18950. 64	IV
CpVQ13	Cp4. 1LG06g04260	LG06:2472160–2473164	334	9. 91	35686. 43	VII
CpVQ14	Cp4. 1LG06g05110	LG06:3011929–3017926	628	5. 57	67834. 1	V
CpVQ15	Cp4. 1LG06g05610	LG06:3325430–3326068	113	9. 62	12485. 03	V
CpVQ16	Cp4. 1LG06g06820	LG06:4263922–4264659	245	8. 71	26210. 14	IV
CpVQ17	Cp4. 1LG07g01470	LG07:802433–803374	313	10. 08	33653. 95	VII
CpVQ18	Cp4. 1LG07g09800	LG07:8797116–8797493	125	6. 05	14220. 9	III
CpVQ19	Cp4. 1LG08g09380	LG08:7407091–7407423	110	10. 67	12069. 01	II
CpVQ20	Cp4. 1LG08g13390	LG08:9672254–9672643	129	4. 84	14612. 33	III
CpVQ21	Cp4. 1LG08g13490	LG08:9735143–9735811	222	5. 92	23933. 64	VI
CpVQ22	Cp4. 1LG09g08360	LG09:7719957–7720586	209	9. 64	22441. 4	IV
CpVQ23	Cp4. 1LG10g01130	LG10:3319719–3322117	425	8. 71	45497. 21	V
CpVQ24	Cp4. 1LG10g01930	LG10:2886496–2886996	166	9. 23	18288. 05	VIII
CpVQ25	Cp4. 1LG10g04880	LG10:1115416–1116003	195	9. 05	21186. 84	IV
CpVQ26	Cp4. 1LG10g06710	LG10:166830–167354	174	6. 28	18860. 78	VI
CpVQ27	Cp4. 1LG10g11090	LG10:7510307–7511283	164	5. 72	17629. 89	III
CpVQ28	Cp4. 1LG10g12290	LG10:9255696–9256184	162	10. 05	18042. 53	IV
CpVQ29	Cp4. 1LG11g04560	LG11:2540172–2540996	274	10. 43	30231. 7	VII

CpVQ30	Cp4. 1LG12g07320	LG12:6924419-6924844	141	5. 91	15294. 24	IV
CpVQ31	Cp4. 1LG13g04200	LG13:6294724-6297911	282	9. 99	31079. 58	II
CpVQ32	Cp4. 1LG14g01330	LG14:3722302-3723387	425	8. 71	45497. 21	V
CpVQ33	Cp4. 1LG14g03000	LG14:2622622-2623233	203	9. 68	21226. 55	V
CpVQ34	Cp4. 1LG14g03270	LG14:2365261-2365482	73	9. 15	8344. 47	I
CpVQ35	Cp4. 1LG16g00810	LG16:1567379-1568005	208	6. 3	22405. 61	II
CpVQ36	Cp4. 1LG17g08680	LG17:5075487-5075909	140	7. 84	14816. 74	IV
CpVQ37	Cp4. 1LG19g00160	LG19:110455-110994	179	10. 04	19803. 85	IV
CpVQ38	Cp4. 1LG19g02600	LG19:2163376-2163876	166	4. 93	17768. 9	III
CpVQ39	Cp4. 1LG19g02610	LG19:2160888-2161388	166	4. 93	17768. 9	III
CpVQ40	Cp4. 1LG19g06070	LG19:7655240-7655563	107	4. 68	12049. 37	I
CpVQ41	Cp4. 1LG19g07250	LG19:7111871-7112545	224	9. 62	24493. 82	IV
CpVQ42	Cp4. 1LG19g09780	LG19:5873440-5875876	344	6. 26	36896. 06	V
CpVQ43	Cp4. 1LG19g10580	LG19:5411418-5411909	163	9. 66	17828. 64	VIII
CpVQ44	Cp4. 1LG20g04410	LG20:2519217-2520794	335	8. 85	36902. 06	IV

Ka, Ks and Ka/Ks values calculated for homologous *VQ* gene pairs.

1 **Table2**

2 **Ka, Ks and Ka/Ks values calculated for homologous VQ gene pairs.**

Gene1	Gene2	Ka	Ks	Ka/Ks ratio	Differentiation time
CpVQ3	CpVQ33	0.143408874	0.41714233	0.343788832	34.19199426
CpVQ4	CpVQ23	0.56179148	1.733083058	0.324157274	142.0559883
CpVQ4	CpVQ32	0.114789648	0.572645983	0.200454821	46.93819533
CpVQ5	CpVQ31	0.244725293	0.735263756	0.332840142	60.26752102
CpVQ6	CpVQ15	1.018237411	1.77385189	0.574026172	145.3976959
CpVQ8	CpVQ13	0.074887932	0.430075478	0.174127416	35.25208839
CpVQ8	CpVQ17	0.302766804	1.921171796	0.157594862	157.473098
CpVQ8	CpVQ29	0.319999352	1.905949455	0.167894983	156.2253652
CpVQ9	CpVQ21	0.895462518	NaN	NaN	#VALUE!
CpVQ11	CpVQ35	0.083226176	0.468993438	0.177457016	38.4420851
CpVQ13	CpVQ17	0.304326774	1.957821306	0.155441548	160.4771563
CpVQ13	CpVQ29	0.301640689	2.424748395	0.12440082	198.7498684
CpVQ16	CpVQ22	0.351584842	2.560617395	0.137304715	209.8866717
CpVQ22	CpVQ25	0.193161647	1.846190406	0.104627153	151.3270824
CpVQ22	CpVQ41	0.206138134	1.465932276	0.140619139	120.1583833
CpVQ23	CpVQ32	0.614828236	2.638504545	0.233021481	216.2708644

CpVQ23	CpVQ42	0.202547017	0.536664673	0.377418203	43.98890764
CpVQ24	CpVQ43	0.142828118	0.572200927	0.249611826	46.90171536
CpVQ25	CpVQ41	0.111174117	0.440146678	0.252584246	36.07759653
CpVQ27	CpVQ39	0.182897122	0.455285224	0.401719872	37.31846099
CpVQ30	CpVQ36	0.207558725	0.385917729	0.537831536	31.6326007

List of primers used in qRT-PCR

List of primers used in qRT-PCR

Gene name	Forward primer (5'-3')	Reverse primer(5'-3')
CpVQ1	CACGGCGGATGCTGATCTGATG	ATAGAAAGAGCCAAGAAAGGGGTCATC
CpVQ3	ACCACCTCCTTCCGATTTACCTACC	CCAGCCGAACTTCCGATGATAGC
CpVQ8	TCGCCTCGCTGGAACAACTTAAC	TACTGCCGATGATGGTGGTGGAG
CpVQ9	AGACGACGACGACTTCTTCAATTCC	GGTGTGATTCTTCTTGGAGGCTCTAG
CpVQ12	CAAGCCGACCCAAACTCCTTCAG	ATGGTCTGTGGCGGGAGGTAAG
CpVQ14	GGCGGCGAAGAAGATTACGACTC	GGGAGGAGGTGGGTTTGGTAGG
CpVQ16	GCTCACTGGCTCATCGGAATCTTC	CTTGTGACGGCGGAGGAATAGAAC
CpVQ21	GGACGCAAGATCACCAAGAGGAAG	GGACCATGTGGCGGAAGTTAGC
CpVQ22	TCTGGGTTGGCGAATGCGAAATC	GTGTAGTGGTCGGTGATGGATGTAAG
CpVQ26	CTATCAGCAGCAGCAGCGATCAG	CTGGGCTTGGAAATGGAGTTGTTATTG
CpVQ33	CCGACGACCTTGCTCAATACAGAC	GATGAGGCGAAAGGCGGAGTTG
CpVQ34	AACGGATGCCAAGAGCTTCAAGAC	CTCTCGCAGCACTCTTTGGAACTC
CpVQ36	ATGGTTTGACAGGGATCGGAATTGG	GATCGCCTTCTCTTCCTCTTCACAAG
CpVQ39	CCGCCGCCCAGGACTATGATAG	TCCGATGAGTGGAGGAGGAAGAATC
CpVQ40	AGTCCGTCGTTCAGACCCTCAC	GTTCTCATCCACTCGCCGCTTC
CAC	GGACAAACAGAACCAACCATGA	GGTTTCCTTTCCGTCACTGTAGA