Genome-wide identification and evolution of HECT genes in wheat

Xianwen Meng, Ting Yang, Jing Liu, Mingde Zhao and Jiuli Wang

The College of Ecological Environmental and Resources, Qinghai Provincial Key Laboratory of High Value Utilization of Characteristic Economic Plants, Qinghai Nationalities University, Xining, China

ABSTRACT

Background. As an important class of E3 ubiquitin ligases in the ubiquitin proteasome pathway, proteins containing homologous E6-AP carboxyl terminus (HECT) domains are crucial for growth, development, metabolism, and abiotic and biotic stress responses in plants. However, little is known about *HECT* genes in wheat (*Triticum aestivum* L.), one of the most important global crops.

Methods. Using a genome-wide analysis of high-quality wheat genome sequences, we identified 25 *HECT* genes classified into six groups based on the phylogenetic relationship among wheat, rice, and *Arabidopsis thaliana*.

Results. The predicted *HECT* genes were distributed evenly in 17 of 21 chromosomes of the three wheat subgenomes. Twenty-one of these genes were hypothesized to be segmental duplication genes, indicating that segmental duplication was significantly associated with the expansion of the wheat *HECT* gene family. The Ka/Ks ratios of the segmental duplication of these genes were less than 1, suggesting purifying selection within the gene family. The expression profile analysis revealed that the 25 wheat *HECT* genes were differentially expressed in 15 tissues, and genes in Group II, IV, and VI (*UPL8*, *UPL6*, *UPL3*) were highly expressed in roots, stems, and spikes. This study contributes to further the functional analysis of the *HECT* gene family in wheat.

INTRODUCTION

Ubiquitination is a post-translational modification that involves the covalent attachment of ubiquitin to a protein substrate. Ubiquitination is important for cellular homeostatic maintenance and plays essential roles in plant growth, development, and the regulation of abiotic and biotic stresses (*Downes et al., 2003*; *El Refy et al., 2003*; *Liu & Stone, 2011*; *Miao & Zentgraf, 2010*; *Moon, Parry & Estelle, 2004*; *Rotin & Kumar, 2009*; *Stone, 2014*; *Wang & Deng, 2011*). At the end of a three-enzyme cascade (E1 ubiquitin activating enzyme; E2 ubiquitin conjugating enzyme; E3 ubiquitin ligase), E3 recruits substrates and promotes or directly catalyzes ubiquitin transfer onto its targets (*Huibregtse et al., 1995*; *Scheffner, Nuber & Huibregtse, 1995*; *Wang et al., 2006*; *Wang & Pickart, 2005*). E3 generally determines the specificity of the ubiquitination reaction with different substrate recognition domains (*Huibregtse et al., 1995*; *Kim et al., 2007*; *Scheffner, Nuber & Huibregtse, 1995*). According to structural similarities and ubiquitination domains, plant E3 proteins can be classified as three main types (*Chen & Hellmann, 2013*; *Craig et al., 2009*; *Duplan & Rivas, 2014*;

Corresponding authors
Mingde Zhao, zmd1226@163.com
Jiuli Wang, wang_jiul@163.com

Guzman, 2014; *Huibregtse et al., 1995*; *Mach, 2008*; *Maspero et al., 2013*; *Qin et al., 2008*; *Scheffner, Nuber & Huibregtse, 1995*; *Schwechheimer & Calderon Villalobos, 2004*; *Stone, 2014*; *Wang et al., 2006*; *Wang & Pickart, 2005*; *Yee & Goring, 2009*): HECT (Homologous to E6-AP C-Terminal), RING (Really Interesting New Gene), and U-box.

The HECT-type ubiquitin ligase is an important class of E3s defined by the presence of a C-terminal catalytic HECT domain. The general features of HECT domains are an N-terminal lobe that contains the E2-binding site and a smaller C-terminal lobe that includes the active-site Cys residue that receives ubiquitin from E2 and links itself with the ubiquitin molecule (*Downes et al., 2003*; *Huibregtse et al., 1995*). Classification of HECT E3 proteins into different subfamilies is based on the N-terminal domains (*Downes et al., 2003*; *Grau-Bove, Sebe-Pedros & Ruiz-Trillo, 2013*; *Marin, 2010*; *Marin, 2013*) responsible for recognizing and binding protein substrates (*Kamadurai et al., 2013*; *Kim et al., 2011*; *Maspero et al., 2011*; *Maspero et al., 2013*; *Rotin & Kumar, 2009*), while the conserved C-terminal HECT domain catalyzes the transfer of ubiquitin to various substrates. Substrate proteins usually possess recognition motifs that can directly bind to the N-terminal domains, while the special HECT domains are essential to the prediction and evolution of the *HECT* genes in plants; however, comprehensive research on these genes is limited.

The HECT-type E3 ubiquitin ligases comprise a small class of E3s, and seven genes (*UPL1-UPL7*) have been identified in *Arabidopsis thaliana* (*Downes et al., 2003*). *UPL3* is involved in trichome development (*Downes et al., 2003*; *Patra, Pattanaik & Yuan, 2013*), genome endoreduplication (*El Refy et al., 2003*) and seed size (*Miller et al., 2019*). *UPL5* is involved in leaf senescence (*Miao & Zentgraf, 2010*), and *UPL1*, *UPL3*, and *UPL5* in plant immunity (*Furniss et al., 2018*). These seven *A. thaliana HECT* genes can be classified into five subfamilies or six groups according to the phylogenetic relationships provided in previous studies (*Marin, 2013*; *Meng et al., 2015*). However, little research has been conducted on the *HECT* genes in wheat, which is one of the most important crops produced worldwide (*Choulet et al., 2014*; *International Wheat Genome Sequencing Consortium, 2014*; *International Wheat Genome Sequencing Consortium et al., 2018*). In this research, we conducted a comprehensive genome-wide analysis of the wheat *HECT* genes to identify *HECT* genes conserved in wheat, rice, and *A. thaliana*. Gene exon-intron structure, conserved motif, domain structure, chromosomal distribution, duplication event, and expression profile were also analyzed in detail. Our research data will provide useful information for further functional investigation of the *HECT* gene family in allohexaploid wheat and their evolution in polyploid plants.

MATERIALS & METHODS

Sequence retrieval and identification of the HECT gene family in wheat

To identify the *HECT* genes in wheat, the protein sequences of all HECT genes in *A. thaliana* and rice were retrieved from the Phytozome v13 database (*Goodstein et al., 2012*) with the Ensembl Plants (*Howe et al., 2020*) as a complementary sequence database. These protein sequences were then used as queries to conducted local BlastP and tBlastN (*Camacho et al., 2009*) searches using default parameters (E-value $<10^{-5}$) against wheat reference sequence

database in the Chinese Spring reference IWGSC RefSeq v1.0 (International Wheat Genome Sequencing et al. 2018) from the Ensembl Plants database. The hmmsearch program in the HMMER software (version 3.3) (*Potter et al., 2018*) was also used to the identification of *HECT* genes with the HMM profile of the HECT domain (PF00632) in the Pfam 32.0 database (*El-Gebali et al., 2019*), using the default parameters (E-value $< 10^{-5}$). Then, the combined candidate *HECT* genes were used as queries to conduct BlastP and tBlastN searches of the wheat genome again to obtain more potential gene candidates with the default parameters (E-value $< 10^{-5}$). The obtained protein sequences were further verified using the InterProScan program (*Jones et al., 2014*) to confirm the presence of the HECT domain. Finally, each *HECT* gene was revised manually for conserved domain architectures using the Pfam (*El-Gebali et al., 2019*), PROSITE (*Sigrist et al., 2013*), and SMART (*Letunic & Bork, 2018*) databases. Proteins without a typical HECT domain or fewer than 300 amino acids were removed from the final sequence dataset.

Sequence alignment and phylogenetic analysis

Multiple sequence alignments of the wheat HECT protein sequences were performed by using MUSCLE (*Edgar, 2004*) with its default parameters, and MAFFT (L-INS-i strategy) (*Rozewicki et al., 2019*). The phylogenetic tree was constructed and visualized using MEGAX software (*Kumar et al., 2018*) based on the full-length HECT protein sequences through a neighbor-joining algorithm with 1,000 bootstrap repetitions. The maximum likelihood (ML) methods implemented in PhyML3.1 (*Guindon et al., 2010*) were also used to construct trees of full-length HECT protein sequences with 1,000 bootstrap repetitions.

Sequence analysis

The structures of *HECT* genes and the number of exons and introns were determined using the Gene Structure Display Server (*Hu et al., 2015*) by aligning the coding sequences with their corresponding genomic DNA sequences. The conserved motifs encoded by *HECT* genes were identified using MEME (Multiple EM for Motif Elicitation) (*Bailey et al., 2015*). The conserved domains of the HECT protein sequences within the phylogenetic trees were visualized and annotated using EvolView (*Subramanian et al., 2019*).

Chromosomal location and duplication

To map all *HECT* genes to the wheat chromosomes, information of *HECT* gene chromosomal location was obtained from Ensembl Plants. Gene duplication events of the wheat *HECT* genes were inferred based on their location among the three wheat subgenomes (A, B, and D). Firstly, all-in-all BlastP of the wheat genome was performed to analyze sequence similarity among the three subgenomes (A, B, and D). Secondly, MCScanX (Multiple Collinearity Scan toolkit) (*Wang et al., 2012*) was then used with default parameters to detect possible gene duplication blocks. Finally, Chromosomal location and syntenic relationships were illustrated using Circos-0.67 (*Krzywinski et al., 2009*). Synonymous (Ks) and nonsynonymous substitution (Ka) rates were calculated with TBtools (*Chen et al., 2018*), as previously described (*Meng et al., 2015*). For each gene pair, the approximate divergence time (T, million years ago, Mya) of the duplication events for each paralogous gene pair was estimated using the mean Ks values from T = Ks/2 λ, in

which the mean synonymous substitution rate (λ) for wheat is 6×10^{-9} (*Wolfe, Li & Sharp, 1987*; *Wolfe, Sharp & Li, 1989*).

Expression analyses

RNA-Seq data (Accession number "choulet_URGI" and "PRJNA497810") were downloaded from expVIP (*Borrill et al., 2019*; *Borrill, Ramirez-Gonzalez & Uauy, 2016*; *Choulet et al., 2014*; *Ramírez-González et al., 2018*) (Table S1–S4) and used to explore the expression patterns of *HECT* genes in wheat. These transcript data were obtained from five organs (root, stem, leaf, spike, and grain) at three developmental stages and flag leaves harvested at 3, 7, 10, 13, 15, 17, 19, 21, 23, and 26 dpa (day post-anthesis). The expression data were log2 based TPM (transcripts per million mapped reads) values, and the heatmap of the expression patterns of wheat *HECT* genes was drawn using the R heatmap.2 function.

RESULTS

Identification of the HECT gene family in wheat

To identify *HECT* genes in wheat, the HMM HECT domain profile PF00632 (*El-Gebali et al., 2019*) and the HECT protein sequences from *A. thaliana* (*Downes et al., 2003*) and rice (*Meng et al., 2015*) were used to search against the wheat protein sequences in Ensemble Plants (*Howe et al., 2020*; International Wheat Genome Sequencing et al. 2018), then the potential *HECT* genes were confirmed by InterProScan (*Jones et al., 2014*) with Pfam (*El-Gebali et al., 2019*), Prosite (*Sigrist et al., 2013*) and SMART (*Letunic & Bork, 2018*) databases that helped to characterize the candidates by the existence of the complete HECT domain. Ultimately, we identified 25 putative *HECT* genes in the latest wheat genome (Table 1).

Phylogenetic analysis of HECT genes in wheat

To understand the evolutionary relationship of the wheat *HECT* genes, phylogenetic trees were constructed based on the alignment of the full-length protein sequences and HECT domain sequences of 25 wheat, 7 rice, and 7 *A. thaliana* HECT proteins (Fig. 1 and Fig. S1). According to the classification criteria used for *A. thaliana* and rice in previous studies (*Downes et al., 2003*; *Grau-Bove, Sebe-Pedros & Ruiz-Trillo, 2013*; *Marin, 2013*; *Meng et al., 2015*), the wheat *HECT* genes were categorized into seven groups (Group I, II, III, IV, V, VI and VII), which contained 0, 3, 5, 5, 3, 5 and 4 *HECT* genes, respectively. Genes in Group III, IV and VI were the most abundant and comprised 60% of the identified genes, while genes in Group I was absent in wheat. Nevertheless, in *A. thaliana*, Group I included two HECT genes, Group II did not include any *HECT* genes, and other Groups included one HECT gene, respectively. These seven groups can be further classified into five subfamilies that correspond to those described in a previous study (*Marin, 2013*).

Gene exon-intron structure and conserved motif and domain architecture of the wheat HECT genes

To investigate the structural characteristics of wheat *HECT* genes, the exon-intron structures of the wheat *HECT* genomic sequences, conserved motifs, and the domain

Table 1 Putative *HECT* genes identified in the wheat genome.

Gene symbol	Gene locus	Chromosome	Gene start (bp)	Gene end (bp)	Gene % GC content	length (AA)
TaHECT01	TraesCS1A02G059500	1A	40892389	40898032	47.13	1335
TaHECT02	TraesCS1A02G106100	1A	103311704	103326325	42.01	1010
TaHECT03	TraesCS1A02G288600	1A	485557978	485566303	43.98	1520
TaHECT04	TraesCS1B02G123400	1B	149601252	149613588	39.18	1027
TaHECT05	TraesCS1B02G298000	1B	518258672	518266604	42.38	1521
TaHECT06	TraesCS1D02G108900	1D	101929905	101943199	39.16	1030
TaHECT07	TraesCS1D02G287600	1D	386119575	386127952	44.18	1522
TaHECT08	TraesCS2A02G064700	2A	29095129	29105808	42.87	1862
TaHECT09	TraesCS2B02G076900	2B	42143087	42153426	43.11	1835
TaHECT10	TraesCS3B02G194900	3B	215607542	215618805	42.3	695
TaHECT11	TraesCS4A02G285800	4A	591699693	591709939	41.24	1036
TaHECT12	TraesCS4B02G027200	4B	20589014	20599094	41.07	1170
TaHECT13	TraesCS4D02G025000	4D	10931453	10942048	41.32	1122
TaHECT14	TraesCS5A02G121600	5A	263498258	263513761	42.56	3632
TaHECT15	TraesCS5A02G262600	5A	475656292	475662357	45.58	832
TaHECT16	TraesCS5B02G112800	5B	177183624	177199346	42.35	3628
TaHECT17	TraesCS5B02G261000	5B	445453118	445459346	44.89	858
TaHECT18	TraesCS5D02G118000	5D	154524415	154539554	42.35	3631
TaHECT19	TraesCS5D02G270200	5D	373832243	373838247	44.01	817
TaHECT20	TraesCS6A02G003300	6A	1494577	1504714	46.25	1839
TaHECT21	TraesCS6B02G000300	6B	163680	173488	45.87	1851
TaHECT22	TraesCS6D02G005600	6D	2399592	2409297	45.55	1841
TaHECT23	TraesCS7A02G244000	7A	220125320	220130919	46.62	862
TaHECT24	TraesCS7B02G313300	7B	559527374	559559604	45.86	897
TaHECT25	TraesCS7B02G496200	7B	747214104	747221564	42.45	796

architecture of the wheat HECT proteins were compared based on their phylogenetic relationships. Each gene structure was revealed by aligning its coding sequences with the corresponding genomic sequences (*Chen et al., 2018*; *Hu et al., 2015*). Most of the wheat *HECT* genes contained abundant (more than ten) exons and only those in Group III had only three or four exons (Fig. 2A). Closely related *HECT* genes in the same phylogenetic group had similar exon-intron structures, and those with closer evolutionary relationships were more similar in their number and length of exons and introns. The conserved motifs of wheat HECT proteins in each group were analyzed using MEME software (*Bailey et al., 2015*). Fifteen conserved motifs (motif1-motif15) were predicted and these motifs were specific to each group (Fig. 2B). The composition of the structural motifs varied among the different HECT groups, while similar motifs were found in the same group. Additionally, the motifs encoding the HECT domain in the C-terminal regions of wheat HECT proteins were relatively conserved, suggesting that the functions of the HECT proteins were intergroup specific. The domain architecture of HECT proteins was obtained using the InterProScan program (*Jones et al., 2014*) with a three-database annotation (*El-Gebali et*
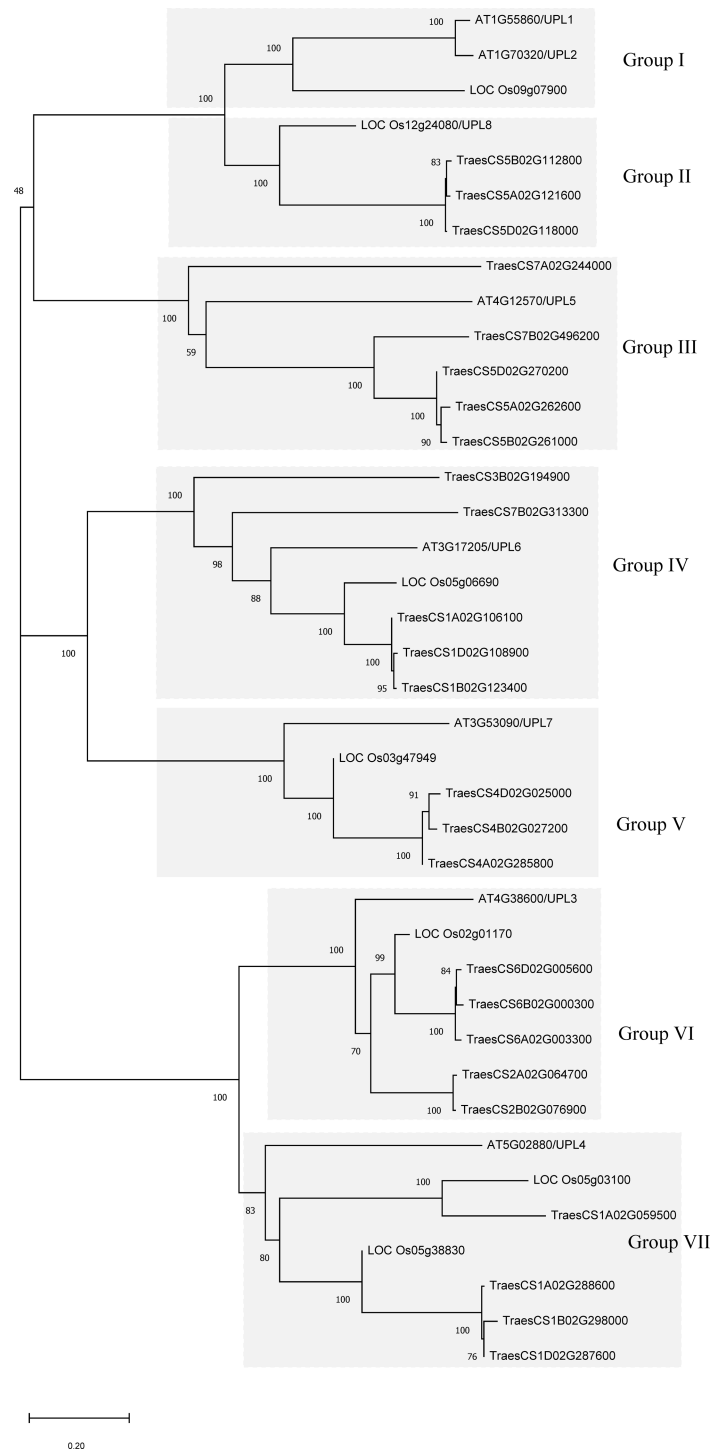
Figure 1 **Phylogenetic relationships of 39 *HECT* genes from wheat (25), rice (seven), and *Arabidopsis thaliana* (seven).** A neighbor-joining (NJ) unrooted tree is shown and the shaded areas indicate the main branches that correspond to the seven gene groups. MEGAX package was used to construct the NJ tree from full-length amino acid sequence alignments (File S1) of the three plant species, with 1000 bootstrap replicates. Numbers refer to bootstrap support in terms of percentage.
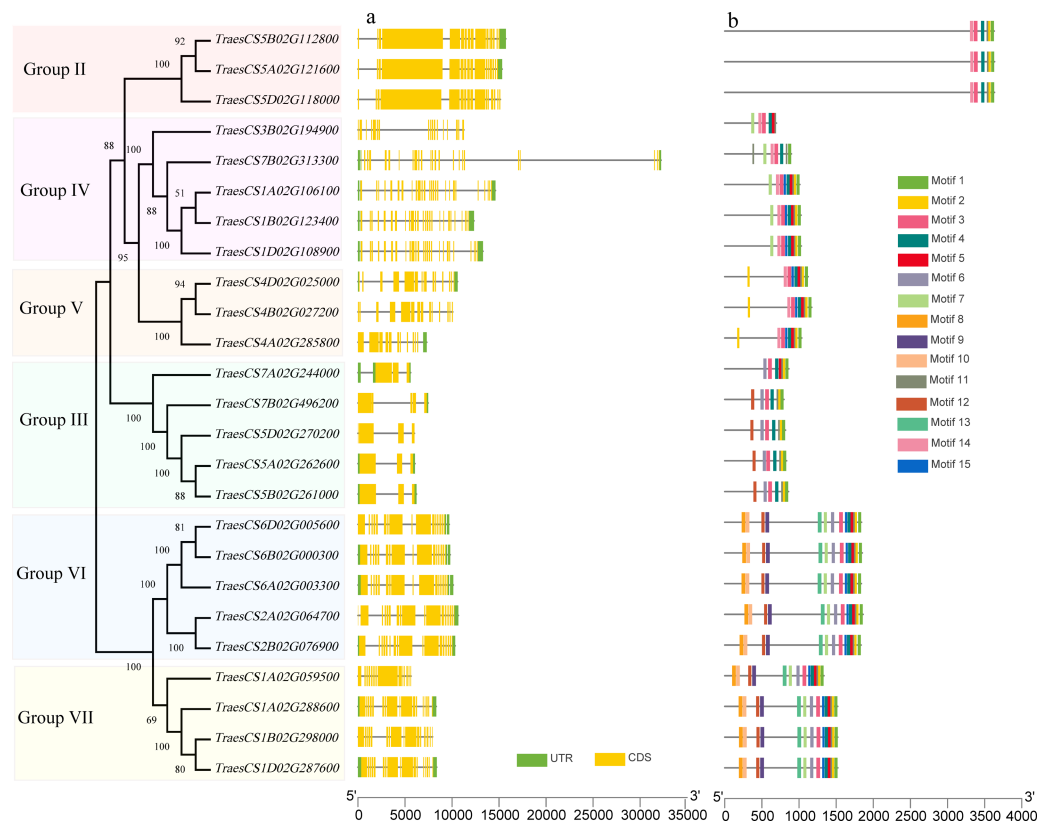
Figure 2 Gene structures and conserved motifs that encode 25 wheat HECT proteins based on phylogenetic relationships. The unrooted neighbor-joining tree was constructed using the alignment of full-length amino acid sequences (File S2) with the MEGAX package. The lengths of the exons and introns in each *HECT* gene are displayed proportionally. The green boxes, yellow boxes, and lines indicate UTRs, exons, and introns, respectively. Conserved motifs are showed in different colors.

al., 2019; *Letunic & Bork, 2018*; *Sigrist et al., 2013*). In addition to the HECT domain, other domains were found in the N-terminal regions of wheat HECT proteins (Fig. 3). The wheat *HECT* genes that were derived from the same group generally had similar exon-intron structures (Fig. 2A), motif compositions (Fig. 2B), and domain architectures (Fig. 3).

Chromosomal location and duplication of wheat HECT genes

To decide the chromosomal locations of the wheat *HECT* genes, the 25 putative wheat *HECT* genes were located in the 21 chromosomes of the wheat genome database available from Ensembl Plants (*Howe et al., 2020*; International Wheat Genome Sequencing et al. 2018). The wheat *HECT* genes were randomly distributed in 17 of 21 chromosomes; chromosome 2D, 3A, 3D, and 7D contained no *HECT* genes, chromosome 1A contained three *HECT* genes, chromosome 1B, 1D, 5A, 5B, 5D, and 7B each contained two *HECT* genes, and the other chromosomes each contained only one *HECT* gene (Fig. 4). The 25 wheat *HECT* genes were approximately evenly distributed among the A (9), B (10), and D (6) subgenomes, which was in accordance with the observation that most *HECT* genes have three homoeologous sequences located across three subgenomes. However, the *HECT*
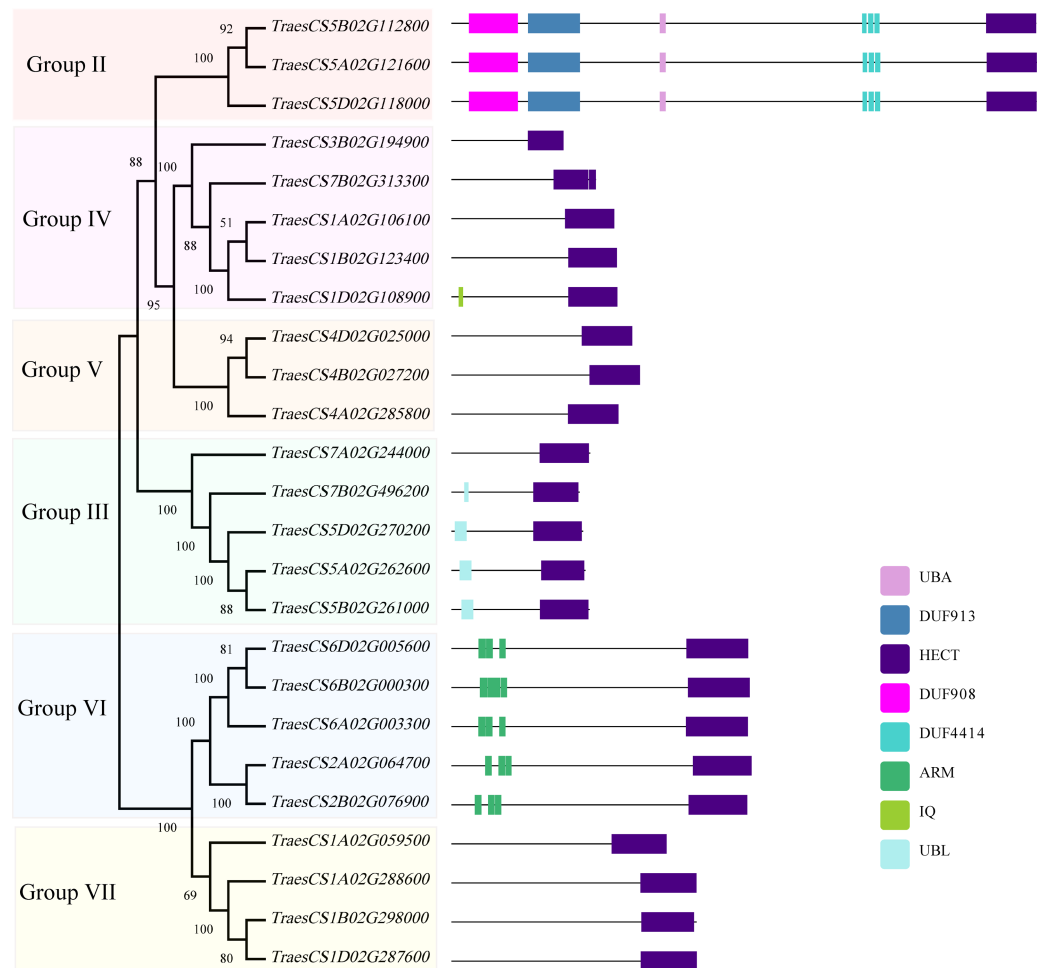
Figure 3 **Domain architecture of 25 wheat HECT proteins according to their phylogenetic relationships.** Each domain is represented by a colored box. UBA: Ubiquitin associated domain, DUF: Domain of Unknown Function, ARM: Armadillo repeats, IQ; IQ Short calmodulin-binding motif, UBL: Ubiquitin-like domain.

genes were not randomly distributed among the different chromosomal groups of the three subgenomes. The chromosomal Group II, III, and VII contained two, one, and three sequences, respectively. The remaining 19 sequences were more evenly distributed across chromosomal Group I, IV, V, and VI, and ranged from three to seven genes per group. An interesting finding was that the location of the HECT genes on chromosome 4A was opposite with those of the homoeologous genes on chromosome 4B and chromosome 4D (Fig. 4).

Segmental and tandem duplication are considered two essential factors for gene family expansion in plants (*Cannon et al., 2004*; *Panchy, Lehti-Shiu & Shiu, 2016*; *Qiao et al., 2019*; *Zhu et al., 2014*). To examine duplication patterns of the wheat *HECT* genes, we identified tandem and segmental duplication events using MCscanX (*Wang et al., 2012*) employing default parameters with TBtools (*Chen et al., 2018*). No tandem duplicated HECT gene
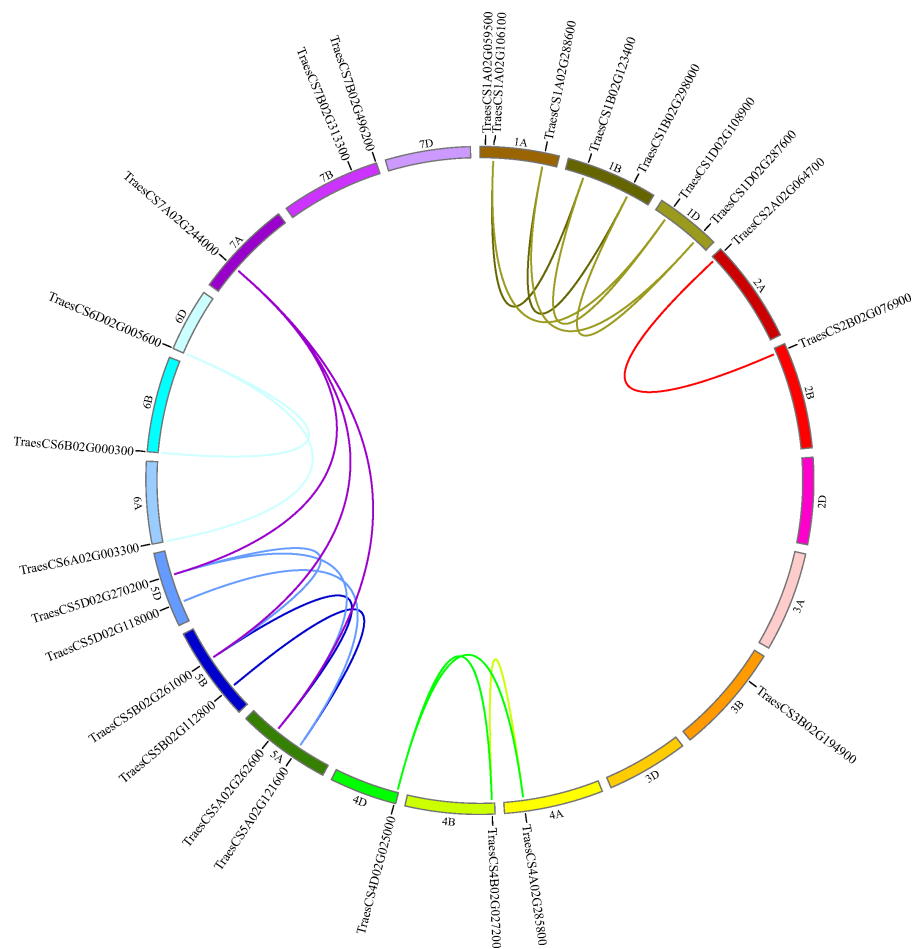
Figure 4 **Chromosome locations of *HECT* genes and segmentally duplicated gene pairs in the wheat genome.** Chromosomes are shown in different colors and in a circular form. The approximate distribution of each wheat *HECT* gene is marked on the circle with a short black line. Colored curves denote the details of syntenic regions between the wheat *HECT* genes. The purple curves represent the estimated time of duplication events that occurred 100–112 Mya, and the other curves represent the estimated time of duplication events 2–12 Mya.

pairs were identified in the 25 wheat *HECT* genes; however, 21 of the 25 wheat *HECT* genes were involved in segmental duplication. Twenty segmental duplicated HECT gene pairs were identified (Fig. 4 and Table 2), indicating that the segmental duplication events had contributed to *HECT* gene family expansion. To date the gene duplication time of these segmentally duplicated *HECT* genes, the Ks and Ka distances, as well as the Ka/Ks ratios were calculated. The Ka/Ks ratios for segmentally duplicated HECT gene pairs ranged from 0.07 to 0.44, with an average value of 0.20 (Table 2), implying that these segmentally duplicated *HECT* genes were under purifying selection, as indicated by the Ka/Ks ratios were less than 1. The divergence time of duplication events were inferred by Ks (Table 2). Results showed that within six existed phylogenetic groups, the two closest wheat *HECT*
Table 2 Estimates of the segmental duplication events in the wheat *HECT* gene pairs.

Group	Gene 1	Gene 2	Ka	Ks	Ka/Ks	Mya
II	*TraesCS5A02G121600*	*TraesCS5B02G112800*	0.01	0.04	0.16	2.93
	TraesCS5A02G121600	*TraesCS5D02G118000*	0.00	0.03	0.14	2.87
III	*TraesCS5A02G262600*	*TraesCS5B02G261000*	0.01	0.08	0.18	6.29
	TraesCS5A02G262600	*TraesCS5D02G270200*	0.01	0.07	0.18	6.09
	TraesCS5A02G262600	*TraesCS7A02G244000*	0.55	1.34	0.41	111.99
	TraesCS5B02G261000	*TraesCS5D02G270200*	0.01	0.06	0.23	4.79
	TraesCS5B02G261000	*TraesCS7A02G244000*	0.55	1.29	0.42	107.74
	TraesCS5D02G270200	*TraesCS7A02G244000*	0.55	1.24	0.44	103.15
IV	*TraesCS1A02G106100*	*TraesCS1B02G123400*	0.00	0.04	0.07	3.07
	TraesCS1A02G106100	*TraesCS1D02G108900*	0.00	0.03	0.15	2.44
	TraesCS1B02G123400	*TraesCS1D02G108900*	0.00	0.03	0.17	2.29
V	*TraesCS4A02G285800*	*TraesCS4B02G027200*	0.01	0.03	0.19	2.91
	TraesCS4A02G285800	*TraesCS4D02G025000*	0.02	0.05	0.30	4.37
	TraesCS4B02G027200	*TraesCS4D02G025000*	0.01	0.06	0.24	5.04
VI	*TraesCS2A02G064700*	*TraesCS2B02G076900*	0.00	0.03	0.15	2.66
	TraesCS6A02G003300	*TraesCS6D02G005600*	0.01	0.12	0.08	10.01
	TraesCS6B02G000300	*TraesCS6D02G005600*	0.01	0.13	0.07	11.03
VII	*TraesCS1A02G288600*	*TraesCS1B02G298000*	0.01	0.06	0.23	5.41
	TraesCS1A02G288600	*TraesCS1D02G287600*	0.01	0.05	0.11	4.25
	TraesCS1B02G298000	*TraesCS1D02G287600*	0.01	0.08	0.17	6.43

Notes.

Ks, synonymous substitution rate; Ka, nonsynonymous substitution rate; Mya, million years ago.

genes were duplicated about 2–12 million years ago (Mya), while the other genes were duplicated about 100–112 Mya.

Expression profiles of wheat HECT genes

To discover the potential roles of these wheat *HECT* genes in growth and development, we used public RNA-seq data covering 15 tissues at different growth stages from expVIP (*Borrill et al., 2019*; *Borrill, Ramirez-Gonzalez & Uauy, 2016*; *Ramírez-González et al., 2018*). Based on the wheat RNA-seq data, the 25 wheat *HECT* genes were detected in all 15 tissues at the gene level (Fig. 5, Table S1, and Table S2). Moreover, the expression of these genes exhibits distinct expression and tissue-specific characteristics. Most *HECT* genes in Group II, IV, and VI were relatively highly expressed in the roots, stems and spikes, while those in the leaves were expressed at relatively low levels (Fig. 5). Interestingly, in wheat grain tissues, most wheat *HECT* gene expression in Group II, IV, and VI were high at 2 dpa and 30 dpa and low at 14 dpa. Moreover, genes within each group or in different groups had similar expression patterns in different tissues, such as the high expression of genes in Group II (*TraesCS5A02G121600, TraesCS5B02G112800, TraesCS5D02G118000*), Group IV (*TraesCS1A02G106100, TraesCS1B02G123400, TraesCS1D02G108900*), and Group VI (*TraesCS6A02G003300, TraesCS6B02G000300, TraesCS6D02G005600, TraesCS2A02G064700,* and *TraesCS2B02G076900*), except for *TraesCS3B02G194900* and *TraesCS7B02G313300* (Fig. 5). Furthermore, the genes in Group II, IV, and VI were
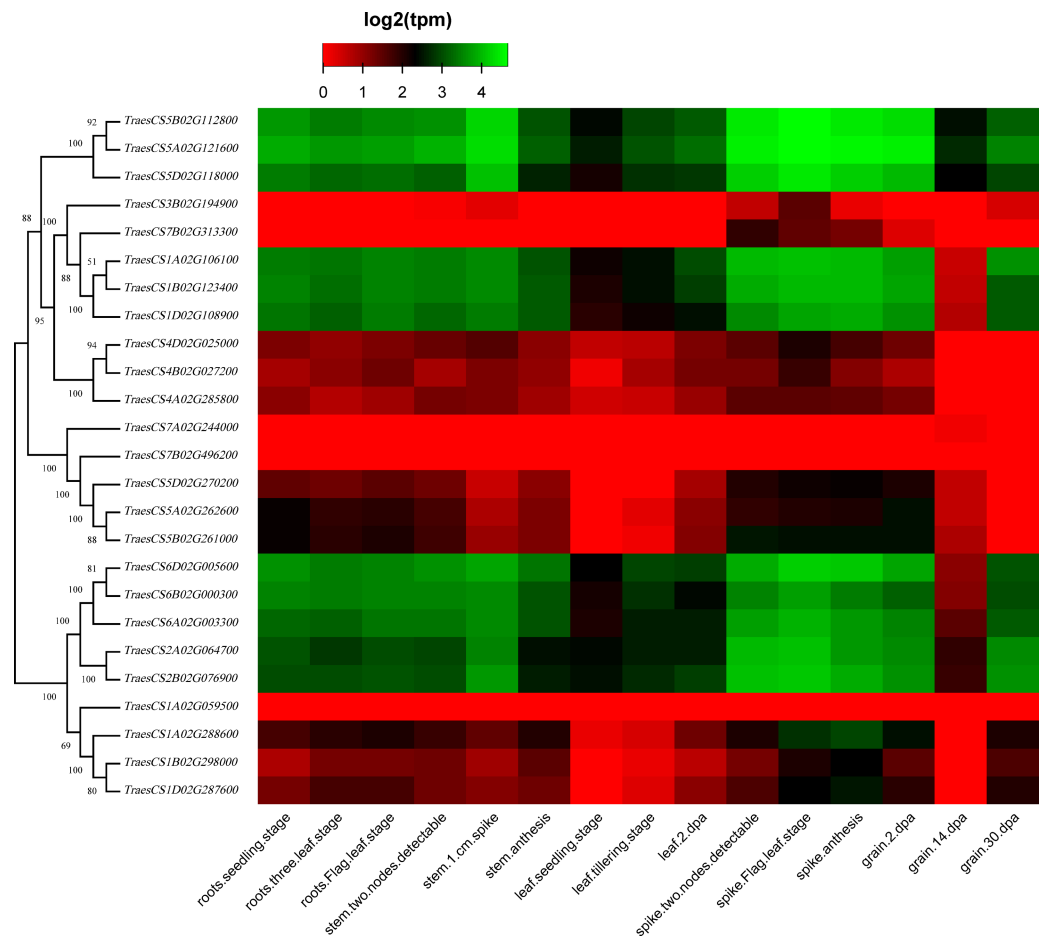
Figure 5 **Heatmap of wheat *HECT* gene expression patterns in 15 different tissues.** Transcriptional levels were obtained from expVIP. The RNA-seq relative expression data from 15 tissues was used to reconstruct the expression patterns of the wheat genes. The sources of the samples are provided on the *x*-axis.

relatively highly expressed in the spikes at different developmental stages and in stems at the one cm spike stage compared to those in other tissues (Fig. 5). According to RNA-seq data of the ten-time point expression time course of wheat senescence in the flag leaf, the expression level of most wheat HECT genes in Group II, IV, and VI gradually increased with the increase of dpa (Fig. 6, Table S3, and Table S4).

DISCUSSION

HECT genes play important roles in *A. thaliana* and diverse plant growth, developmental and physiological processes (*Downes et al., 2003*; (*El Refy et al., 2003*); *Furniss et al., 2018*; *Miao & Zentgraf, 2010*; *Miller et al., 2019*; *Patra, Pattanaik & Yuan, 2013*), including trichome development (*Downes et al., 2003*), genome endoreduplication (*El Refy et al., 2003*), seed size (*Miller et al., 2019*), leaf senescence (*Miao & Zentgraf, 2010*), and plant immunity (*Furniss et al., 2018*). However, this gene family has not been investigated in
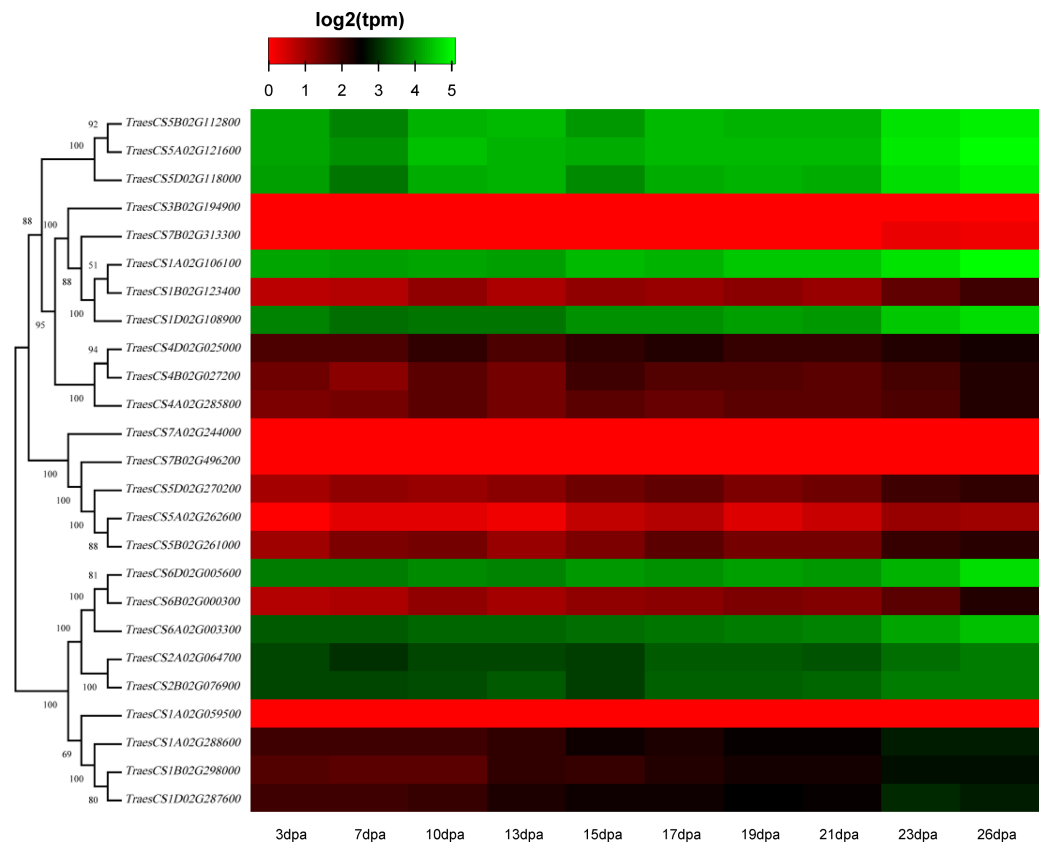
Figure 6 **Heatmap of wheat HECT gene expression patterns in wheat leaf senescence.** Transcriptional levels were obtained from expVIP. The RNA-seq relative expression data from flag leaves harvested at 3, 7, 10, 13, 15, 17, 19, 21, 23, and 26 dpa (day post-anthesis) was used to reconstruct the expression patterns of the wheat genes. The sources of the samples are provided on the *x*-axis.

wheat. In this research, we conducted an extensive analysis of the wheat *HECT* genes, including their evolution, gene exon-intron structure, conserved motif, domain structure, chromosomal location, duplication event, and expression pattern.

We identified 25 *HECT* genes in the wheat genome, which is 3.6 times the number present in *A. thaliana* (*Downes et al., 2003*). However, a former study discovered 19 soybean *HECT* genes, which is 2.7 times the number found in *A. thaliana* (*Meng et al., 2015*). Our results showed six more *HECT* genes in wheat than the number previously estimated in the soybean genome. A possible explanation for this difference is that wheat is a hexaploid crop with 21 chromosomes containing three subgenomes (A, B, and D) (*International Wheat Genome Sequencing Consortium, 2014*; *International Wheat Genome Sequencing Consortium et al., 2018*), while soybean is a diploid crop with 20 chromosomes derived from an ancient tetraploid that may have had about two times more the number of *HECT* genes than other normal diploid species (*Schmutz et al., 2010*).

The phylogenetic analysis of the 25 wheat *HECT* genes classified them into subfamilies similar to those characterized by previous research (*Downes et al., 2003*; *Grau-Bove, Sebe-Pedros & Ruiz-Trillo, 2013*; *Marin, 2013*; *Meng et al., 2015*). The classification was

according to the corresponding HECT gene sequence homology. Based on the phylogenetic relationships among the *HECT* genes in wheat, rice, and *A. thaliana*, the wheat *HECT* genes were classified into seven groups. Compared with a former report in *A. thaliana* (*Marin, 2013*), subfamily IV *HECT* genes were absent in wheat. Wheat subfamily V (Group I, *UPL1/2* and Group II, *UPL8* in this study) contained three genes, subfamily VI (Group III, *UPL5*) contained five genes, subfamily III (Group IV, *UPL6*) contained five genes, Subfamily II (Group V, *UPL7*) contained three genes, and subfamily I (Group VI, *UPL3* and Group VII, *UPL4*) contained nine genes. The *HECT* gene Group I was not observed in the wheat genome. With the exception of Group I and II, other wheat Groups own HECT genes orthologous with *A. thaliana*. This is basically consistent with the results of a previous HECT gene investigation in plants (*Marin, 2013*), suggesting that *A. thaliana* HECT gene Group II (*UPL8* in this study) was lost, while the wheat HECT gene Group I (*UPL1/UPL2*) was not observed in our analysis. Gene members of each phylogenetic Group often possess identical gene exon-intron structures, conserved motifs, and domain architectures, indicating that they probably recognize, bind, and might interact with same or similar substrate protein.

Segmental duplication events, tandem duplication events, as well as transposition events are three main evolutionary mechanisms of duplication events that expand the members of gene family (*Cannon et al., 2004*; *Panchy, Lehti-Shiu & Shiu, 2016*; *Qiao et al., 2019*; *Zhu et al., 2014*). Segmental duplications frequently occur in higher plants, because they are diploidized polyploids that have maintained various duplicated chromosomal blocks in the existing genomes (*Cannon et al., 2004*; *Qiao et al., 2019*). In this present research, we discovered that 21 of the 25 wheat *HECT* genes were located in chromosomes across the three subgenomes (A, B, D), indicating that segmental duplication obviously contributed to expanding the wheat *HECT* gene family. A previous study has shown that the allohexaploid wheat subgenomes A, B, and D were originally derived from three diploid ($2x$; $2n = 14$) species and underwent three hybridization events (International Wheat Genome Sequencing 2014). The A and B subgenomes diverged from a common ancestor ∼7 million years ago and the first hybridization occurred ∼5.5 million years ago between A and B subgenomes, leading to the D subgenome through homoploid hybrid speciation. The second hybridization between the A and B subgenomes gave rise to the AABB genome <0.8 million years ago via polyploidization. Wheat originated <0.4 million years ago by allopolyploidization from a third hybridization. By estimating the approximate dates of the segmentally duplicated pairs of wheat *HECT* genes, we infer that the paralogous genes in wheat HECT groups originated from a relatively recent duplication event during the shaping of the three subgenomes (A, B, D) that occurred before the second hybridization event in wheat evolution history, except for *TraesCS7A02G244000* in Group III, which originated from a relatively ancient duplication event before the appearance of the common ancestor of the A and B subgenomes. Thus, segmental duplication events were the primary driving forces for *HECT* gene evolution during the speciation and evolution of allohexaploid wheat.

To better understand the roles of the *HECT* genes during the life cycle of wheat, we performed an expression analysis of public RNA-seq data (*Borrill, Ramirez-Gonzalez &*

Uauy, 2016; *Choulet et al., 2014*; *Ramírez-González et al., 2018*) in 15 tissues at different developmental stages. Analysis of the expression patterns of these wheat genes in 15 tissues showed that most wheat *HECT* genes in Group II, IV, and VI were relatively highly expressed in the roots, stems, and spikes. In particular, the genes in Groups II, IV, and VI were relatively highly expressed in the spikes at different developmental stages and in stems at the 1-cm spike stage compared to other tissues. Therefore, the expression of these genes may be closely related to wheat spike growth and development, suggesting that the *HECT* genes in highly expressed spikes may be involved in the regulation or degradation of proteins via ubiquitination during spike development stage. Previous studies have revealed that *A. thaliana AT4G38600/UPL3* plays a specific role during trichome development (*Downes et al., 2003*; *Patra, Pattanaik & Yuan, 2013*) and seed size (*Miller et al., 2019*) and that *AT4G12570/UPL5* is an important transcription factor that positively regulates leaf senescence by the ubiquitination and degradation of *AT4G23810/WRKY5 3* (*Miao & Zentgraf, 2010*). In our investigation, the wheat genes orthologous to *A. thaliana AT4G38600/UPL3* included five paralogous genes in Group VI. Except for grain at the 14 dpa stage, these five genes were all relatively highly expressed in wheat, particularly in spikes. A reasonable explanation is that the relatively low expression of *UPL3* at 14 dpa stage may be related to the size of wheat seeds and is an adaptive regulation mechanism during seed formation. This is consistent with a recent study of *UPL3* in *Brassica napus* (*Miller et al., 2019*). Miller et al. determined a mechanism in which the proteasomal degradation of LEC2, a transcription factor controlling seed maturation, is mediated by UPL3 and reduced *UPL3* expression would increase LEC2 protein levels and seed size. The wheat genes orthologous to *A. thaliana AT4G12570/UPL5* were five paralogous genes in Group III, which were expressed in different wheat tissues but showed distinct features. At different developmental stages, the expression levels of *TraesCS5D02G270200*, *TraesCS5A02G262600*, and *TraesCS5B02G261000* in roots, stems, and spikes were relatively unchanged, but gradually increased in leaves and decreased in grain. The three wheat genes *TraesCS1A02G106100*, *TraesCS1B02G123400*, and *TraesCS1D02G108900* are orthologous to *A. thaliana AT3G17205/UPL6* in Group IV, and the genes in Group II (*UPL8* absent in *A. thaliana*) showed similar expression patterns to those in Group VI (*UPL3*). RNA-seq data of wheat leaf senescence (*Borrill et al., 2019*) indicated that HECT genes in Group II, IV and VI (*UPL8*, *UPL6*, and *UPL3*) might also play crucial roles in plant leaf senescence. The differential expression of paralogous HECT genes in or among groups in wheat suggests that they might have the same or similar functions as their orthologous genes in *A. thaliana* and *Brassica napus*; nevertheless, they might have evolved functional differences.

A former research discovered that AT4G38600/UPL3 mediated UPS-dependent proteolysis of the two transcription factors AT5G41315/GL3 and AT1G63650/EGL3, which interact with the ARM domains of UPL3 and function as positive regulators during *A. thaliana* trichome development (*Patra, Pattanaik & Yuan, 2013*). The evolutionarily closely related Group VII (UPL4) and VI (UPL3) belong to the same subfamily I defined in previous studies (*Grau-Bove, Sebe-Pedros & Ruiz-Trillo, 2013*; *Marin, 2013*) and genes in Group VII contain no ARM domains (Fig. 3) and thus, are differentially expressed at relatively low levels (Fig. 5) compared to those genes in Group VI. More functional

explorations of these genes could improve our understanding of the roles of *HECT* genes in wheat and other plants during growth and development.

CONCLUSIONS

Herein, 25 identified wheat HECT genes were classified into six phylogenetic groups and distributed evenly in 17 of 21 chromosomes of the three subgenomes. Twenty-one hypothesized segmentally duplicated genes indicated that segmental duplication was significantly associated with the expansion of these HECT genes. The expression analysis revealed that most wheat HECT genes in Group II, IV, and VI (*UPL8*, *UPL6*, and *UPL3*) were highly expressed in roots, stems, and spikes at different developmental stages and gradually increased with the increase of dpa, suggesting that these genes may be involved in wheat growth, development and leaf senescence. This study provides useful information for further biological functional analysis of the HECT gene family in allohexaploid wheat.

Funding

This research was supported by the Natural Science Foundation of Qinghai Province (2018-ZJ-964Q) and the Scientific research Projects for High-level Talents of Qinghai Nationalities University (2017XJG02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Natural Science Foundation of Qinghai Province: 2018-ZJ-964Q.
Scientific Research Projects for High-level Talents of Qinghai Nationalities University: 2017XJG02.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xianwen Meng conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ting Yang and Jing Liu performed the experiments, prepared figures and/or tables, and approved the final draft.
- Mingde Zhao and Jiuli Wang conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The raw data is available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.10457#supplemental-information.

REFERENCES

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Research* **43**:W39–W49 DOI 10.1093/nar/gkv416.

Borrill P, Harrington SA, Simmonds J, Uauy C. 2019. Identification of transcription factors regulating senescence in wheat through gene regulatory network modelling. *Plant Physiology* **180**:1740–1755 DOI 10.1104/pp.19.00380.

Borrill P, Ramirez-Gonzalez R, Uauy C. 2016. expVIP: a Customizable RNA-seq data analysis and visualization platform. *Plant Physiology* **170**:2172–2186 DOI 10.1104/pp.15.01667.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421 DOI 10.1186/1471-2105-10-421.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. *BMC Plant Biology* **4**:10 DOI 10.1186/1471-2229-4-10.

Chen C, Xia R, Chen H, He Y. 2018. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv 289660* DOI 10.1101/289660.

Chen L, Hellmann H. 2013. Plant E3 ligases: flexible enzymes in a sessile world. *Molecular Plant* **6**:1388–1404 DOI 10.1093/mp/sst005.

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Berges H, Quesneville H, Wincker P, Feuillet C. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**:1249721 DOI 10.1126/science.1249721.

Craig A, Ewan R, Mesmar J, Gudipati V, Sadanandom A. 2009. E3 ubiquitin ligases and plant innate immunity. *Journal of Experimental Botany* **60**:1123–1132 DOI 10.1093/jxb/erp059.

Downes BP, Stupar RM, Gingerich DJ, Vierstra RD. 2003. The HECT ubiquitin-protein ligase (UPL) family in Arabidopsis: UPL3 has a specific role in trichome development. *The Plant Journal* **35**:729–742 DOI 10.1046/j.1365-313X.2003.01844.x.

Duplan V, Rivas S. 2014. E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. *Frontiers in Plant Science* **5**:42 DOI 10.3389/fpls.2014.00042.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797 DOI 10.1093/nar/gkh340.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* **47**:D427–D432 DOI 10.1093/nar/gky995.

El Refy A, Perazza D, Zekraoui L, Valay JG, Bechtold N, Brown S, Hulskamp M, Herzog M, Bonneville JM. 2003. The Arabidopsis KAKTUS gene encodes a HECT protein and controls the number of endoreduplication cycles. *Molecular Genetics and Genomics* **270**:403–414 DOI 10.1007/s00438-003-0932-1.

Furniss JJ, Grey H, Wang Z, Nomoto M, Jackson L, Tada Y, Spoel SH. 2018. Proteasome-associated HECT-type ubiquitin ligase activity is required for plant immunity. *PLOS Pathogens* **14**:e1007447 DOI 10.1371/journal.ppat.1007447.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**:D1178–D1186 DOI 10.1093/nar/gkr944.

Grau-Bove X, Sebe-Pedros A, Ruiz-Trillo I. 2013. A genomic survey of HECT ubiquitin ligases in eukaryotes reveals independent expansions of the HECT system in several lineages. *Genome Biology and Evolution* **5**:833–847 DOI 10.1093/gbe/evt052.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**:307–321 DOI 10.1093/sysbio/syq010.

Guzman P. 2014. ATLs and BTLs, plant-specific and general eukaryotic structurally-related E3 ubiquitin ligases. *Plant Science* **215–216**:69–75 DOI 10.1016/j.plantsci.2013.10.017.

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, Carbajo M, Chakiachvili M, Christensen M, Cummins C, Cuzick A, Davis P, Fexova S, Gall A, George N, Gil L, Gupta P, Hammond-Kosack KE, Haskell E, Hunt SE, Jaiswal P, Janacek SH, Kersey PJ, Langridge N, Maheswari U, Maurel T, McDowall MD, Moore B, Muffato M, Naamati G, Naithani S, Olson A, Papatheodorou I, Patricio M, Paulini M, Pedro H, Perry E, Preece J, Rosello M, Russell M, Sitnik V, Staines DM, Stein J, Tello-Ruiz MK, Trevanion SJ, Urban M, Wei S, Ware D, Williams G, Yates AD, Flicek P. 2020. Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Research* **48**:D689–D695 DOI 10.1093/nar/gkz890.

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**:1296–1297 DOI 10.1093/bioinformatics/btu817.

Huibregtse JM, Scheffner M, Beaudenon S, Howley PM. 1995. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proceedings of the National Academy of Sciences of the United States of America* **92**:2563–2567 DOI 10.1073/pnas.92.7.2563.

International Wheat Genome Sequencing Consortium. 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**:1251788 DOI 10.1126/science.1251788.

International Wheat Genome Sequencing Consortium, investigators IRp, Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Investigators Iw-gap, Pozniak CJ, Stein N, Choulet F, Distelfeld A, Eversole K, Poland J, Rogers J, Ronen G, Sharpe AG, Whole-genome S, assembly, Pozniak C, Ronen G, Stein N, Barad O, Baruch K, Choulet F, Keeble-Gagnere G, Mascher M, Sharpe AG, Ben-Zvi G, Josselin AA, Hi Cd-bs, Stein N, Mascher M, Himmelbach A, Whole-genome assembly quality c, analyses, Choulet F, Keeble-Gagnere G, Mascher M, Rogers J, Balfourier F, Gutierrez-Gonzalez J, Hayden M, Josselin AA, Koh C, Muehlbauer G, Pasam RK, Paux E, Pozniak CJ, Rigault P, Sharpe AG, Tibbits J, Tiwari V, Pseudomolecule a, Choulet F, Keeble-Gagnere G, Mascher M, Josselin AA, Rogers J, RefSeqgenome S, gene a, Spannagl M, Choulet F, Lang D, Gundlach H, Haberer G, Keeble-Gagnere G, Mayer KFX, Ormanbekova D, Paux E, Prade V, Simkova H, Wicker T, Automated a, Choulet F, Spannagl M, Swarbreck D, Rimbert H, Felder M, Guilhot N, Gundlach H, Haberer G, Kaithakottil G, Keilwagen J, Lang D, Leroy P, Lux T, Mayer KFX, Twardziok S, Venturini L, Manualgene C, Appels R, Rimbert H, Choulet F, Juhasz A, Keeble-Gagnere G, Subgenomecomparative A, Choulet F, Spannagl M, Lang D, Abrouk M, Haberer G, Keeble-Gagnere G, Mayer KFX, Wicker T, Transposable E, Choulet F, Wicker T, Gundlach H, Lang D, Spannagl M, Phylogenomic A, Lang D, Spannagl M, Appels R, Fischer I, Transcriptome a, data RN-s, Uauy C, Borrill P, Ramirez-Gonzalez RH, Appels R, Arnaud D, Chalabi S, Chalhoub B, Choulet F, Cory A, Datla R, Davey MW, Hayden M, Jacobs J, Lang D, Robinson SJ, Spannagl M, Steuernagel B, Tibbits J, Tiwari V, Ex Fvan, Wulff BBH, Whole-genome M, Pozniak CJ, Robinson SJ, Sharpe AG, Cory A, Histonemark A, Benhamed M, Paux E, Bendahmane A, Concia L, Latrasse D, tags BACcMI-BW-GP, Rogers J, Jacobs J, Alaux M, Appels R, Bartos J, Bellec A, Berges H, Dolezel J, Feuillet C, Frenkel Z, Gill B, Korol A, Letellier T, Olsen OA, Simkova H, Singh K, Valarik M, vander Vossen E, Vautrin S, Weining S, Chromosome LTCm, physical mapping quality c, Korol A, Frenkel Z, Fahima T, Glikson V, Raats D, Rogers J, mapping RH, Tiwari V, Gill B, Paux E, Poland J, Optical m, Dolezel J, Cihalikova J, Simkova H, Toegelova H, Vrana J, Recombination a, Sourdille P, Darrier B, Genefamily A, Appels R, Spannagl M, Lang D, Fischer I, Ormanbekova D, Prade V, family CBFg, Barabaschi D, Cattivelli L, Dehydrin gene F, Hernandez P, Galvez S, Budak H, family NLRg, Steuernagel B, Jones JDG, Witek K, Wulff BBH, Yu G, family PPRg, Small I , et al. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**:eaar7191 DOI 10.1126/science.aar7191.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240 DOI 10.1093/bioinformatics/btu031.

Kamadurai HB, Qiu Y, Deng A, Harrison JS, Macdonald C, Actis M, Rodrigues P, Miller DJ, Souphron J, Lewis SM, Kurinov I, Fujii N, Hammel M, Piper R, Kuhlman B, Schulman BA. 2013. Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *Elife* **2**:e00828 DOI 10.7554/eLife.00828.

Kim HT, Kim KP, Lledias F, Kisselev AF, Scaglione KM, Skowyra D, Gygi SP, Goldberg AL. 2007. Certain pairs of ubiquitin-conjugating enzymes (E2s) and ubiquitin-protein ligases (E3s) synthesize nondegradable forked ubiquitin chains containing all possible isopeptide linkages. *Journal of Biological Chemistry* **282**:17375–17386 DOI 10.1074/jbc.M609659200.

Kim HC, Steffen AM, Oldham ML, Chen J, Huibregtse JM. 2011. Structure and function of a HECT domain ubiquitin-binding site. *EMBO Report* **12**:334–341 DOI 10.1038/embor.2011.23.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* **19**:1639–1645 DOI 10.1101/gr.092759.109.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**:1547–1549 DOI 10.1093/molbev/msy096.

Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* **46**:D493–D496 DOI 10.1093/nar/gkx922.

Liu H, Stone SL. 2011. E3 ubiquitin ligases and abscisic acid signaling. *Plant Signaling & Behavior* **6**:344–348 DOI 10.4161/psb.6.3.13914.

Mach J. 2008. Ubiquitin ligation RINGs twice: redundant control of plant processes by E3 ubiquitin ligases. *The Plant Cell* **20**:1424 DOI 10.1105/tpc.108.200611.

Marin I. 2010. Animal HECT ubiquitin ligases: evolution and functional implications. *BMC Evolutionary Biology* **10**:56 DOI 10.1186/1471-2148-10-56.

Marin I. 2013. Evolution of plant HECT ubiquitin ligases. *PLOS ONE* **8**:e68536 DOI 10.1371/journal.pone.0068536.

Maspero E, Mari S, Valentini E, Musacchio A, Fish A, Pasqualato S, Polo S. 2011. Structure of the HECT:ubiquitin complex and its role in ubiquitin chain elongation. *EMBO Report* **12**:342–349 DOI 10.1038/embor.2011.21.

Maspero E, Valentini E, Mari S, Cecatiello V, Soffientini P, Pasqualato S, Polo S. 2013. Structure of a ubiquitin-loaded HECT ligase reveals the molecular basis for catalytic priming. *Nature Structural & Molecular Biology* **20**:696–701 DOI 10.1038/nsmb.2566.

Meng X, Wang C, Rahman SU, Wang Y, Wang A, Tao S. 2015. Genome-wide identification and evolution of HECT genes in soybean. *International Journal of Molecular Sciences* **16**:8517–8535 DOI 10.3390/ijms16048517.

Miao Y, Zentgraf U. 2010. A HECT E3 ubiquitin ligase negatively regulates Arabidopsis leaf senescence through degradation of the transcription factor WRKY53. *The Plant Journal* **63**:179–188 DOI 10.1111/j.1365-313X.2010.04233.x.

Miller C, Wells R, McKenzie N, Trick M, Ball J, Fatihi A, Dubreucq B, Chardot T, Lepiniec L, Bevan MW. 2019. Variation in Expression of the HECT E3 Ligase UPL3

Modulates LEC2 Levels, Seed Size, and Crop Yields in Brassica napus. *The Plant Cell* **31**:2370–2385 DOI 10.1105/tpc.18.00577.

Moon J, Parry G, Estelle M. 2004. The ubiquitin-proteasome pathway and plant development. *The Plant Cell* **16**:3181–3195 DOI 10.1105/tpc.104.161220.

Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. *Plant Physiology* **171**:2294–2316 DOI 10.1104/pp.16.00523.

Patra B, Pattanaik S, Yuan L. 2013. Ubiquitin protein ligase 3 mediates the proteasomal degradation of GLABROUS 3 and ENHANCER OF GLABROUS 3, regulators of trichome development and flavonoid biosynthesis in Arabidopsis. *The Plant Journal* **74**:435–447 DOI 10.1111/tpj.12132.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Research* **46**:W200–W204 DOI 10.1093/nar/gky448.

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology* **20**:38 DOI 10.1186/s13059-019-1650-2.

Qin F, Sakuma Y, Tran LS, Maruyama K, Kidokoro S, Fujita Y, Fujita M, Umezawa T, Sawano Y, Miyazono K, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K. 2008. Arabidopsis DREB2A-interacting proteins function as RING E3 ligases and negatively regulate plant drought stress-responsive gene expression. *The Plant Cell* **20**:1693–1707 DOI 10.1105/tpc.107.057380.

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, Ex Fvan, Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout CJ, Chalhoub B, Mayer KFX, Benhamed M, Latrasse D, Bendahmane A, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe AG, Paux E, Spannagl M, Bräutigam A, Uauy C. 2018. The transcriptional landscape of polyploid wheat. *Science* **361**:eaar6089 DOI 10.1126/science.aar6089.

Rotin D, Kumar S. 2009. Physiological functions of the HECT family of ubiquitin ligases. *Nature Reviews Molecular Cell Biology* **10**:398–409 DOI 10.1038/nrm2690.

Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research* **47**:W5–W10 DOI 10.1093/nar/gkz342.

Scheffner M, Nuber U, Huibregtse JM. 1995. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**:81–83 DOI 10.1038/373081a0.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* **463**:178–183 DOI 10.1038/nature08670.

Schwechheimer C, Calderon Villalobos LI. 2004. Cullin-containing E3 ubiquitin ligases in plant development. *Current Opinion in Plant Biology* **7**:677–686 DOI 10.1016/j.pbi.2004.09.009.

Sigrist CJ, Castro Ede, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2013. New and continuing developments at PROSITE. *Nucleic Acids Research* **41**:D344–D347 DOI 10.1093/nar/gks1067.

Stone SL. 2014. The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. *Frontiers in Plant Science* **5**:135 DOI 10.3389/fpls.2014.00135.

Subramanian B, Gao S, Lercher MJ, Hu S, Chen WH. 2019. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Research* **47**:W270–W275 DOI 10.1093/nar/gkz357.

Wang M, Cheng D, Peng J, Pickart CM. 2006. Molecular determinants of polyubiquitin linkage selection by an HECT ubiquitin ligase. *EMBO Journal* **25**:1710–1719 DOI 10.1038/sj.emboj.7601061.

Wang F, Deng XW. 2011. Plant ubiquitin-proteasome pathway and its role in gibberellin signaling. *Cell Research* **21**:1286–1294 DOI 10.1038/cr.2011.118.

Wang M, Pickart CM. 2005. Different HECT domain ubiquitin ligases employ distinct mechanisms of polyubiquitin chain synthesis. *EMBO Journal* **24**:4324–4333 DOI 10.1038/sj.emboj.7600895.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**:e49 DOI 10.1093/nar/gkr1293.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America* **84**:9054–9058 DOI 10.1073/pnas.84.24.9054.

Wolfe KH, Sharp PM, Li WH. 1989. Rates of synonymous substitution in plant nuclear genes. *Journal of Molecular Evolution* **29**:208–211 DOI 10.1007/BF02100204.

Yee D, Goring DR. 2009. The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. *Journal of Experimental Botany* **60**:1109–1121 DOI 10.1093/jxb/ern369.

Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, Hu Y. 2014. Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. *BMC Plant Biology* **14**:93 DOI 10.1186/1471-2229-14-93.