

New Insights Into Invertase Gene Family and Functional Characterization of Critical Cell Wall Invertase *TaCWINV40* For Male Fertility in Wheat (*Triticum Aestivum* L.)

Jiali Ye

Northwest A&F University: Northwest Agriculture and Forestry University

Yaning Bu

Northwest A&F University: Northwest Agriculture and Forestry University

Mengting He

Northwest A&F University: Northwest Agriculture and Forestry University

Yongfeng Wu

Northwest A&F University: Northwest Agriculture and Forestry University

Xuetong Yang

Northwest A&F University: Northwest Agriculture and Forestry University

Lingli Zhang

Northwest A&F University: Northwest Agriculture and Forestry University

Xiyue Song (✉ songxiyue@nwafu.edu.cn)

Northwest Agriculture and Forestry University <https://orcid.org/0000-0001-8158-3380>

Research Article

Keywords: gene family, invertase, thermo-sensitive male sterility, wheat

Posted Date: December 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1109325/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Invertase (INV, ec3.2.1.26) irreversibly hydrolyzes sucrose into fructose and glucose, and it is regulated by the environment to affect pollen fertility in some plant species. However, there has been a lack of systematic identification of INV gene family in wheat. In order to reveal the potential influence on the male fertility, a total of 130 wheat INVs that unevenly distributed on 21 chromosomes were systematically identified and analyzed in this study. According to physical and chemical properties, subcellular location, and phylogenetic tree, they were divided into two acidic INV (AINV) subtypes: cell wall group (TaCWINV1-68), vacuole group (TaVINV1-42), and two neutral/alkaline INV (A/NINV) subtypes: cytoplasmic α group (TaA/NINV1-11) and cytoplasmic β group (TaA/NINV12-20). The amplification of A/NINVs is mainly attributed to the polyploidization of wheat, and the multiple duplication events experienced in AINVs revealed their non-dose sensitivity characteristic. The wheat RNA-seq data revealed the tissue specificity of A/NINVs and AINVs, and six spike-specific CWINVs showed significant differential expression between the fertile and sterile anthers of thermo-sensitive male-sterile wheat KTM3315A. *TaCWINV40* localized in cell wall was effectively silenced in the fertile KTM3315A, and the malformed pollen grains and non-germinating pollen tubes shed light on its indispensability in the development of wheat anthers. This study will spur the interest on manipulating the novel genetic characteristics of *TaCWINVs* for the construction and improvement of wheat male sterile materials.

Introduction

Sucrose is the final product of photosynthesis, which is transported from the source tissue to the non-photosynthetic tissue (sink tissue) through the phloem (Koch 2004). In non-photosynthetic tissues, sucrose and its hexose products (glucose and fructose) play important roles in primary metabolism and special metabolism (Caretto et al. 2015). They are not only raw materials for many metabolic pathways, but also provide energy and carbon skeletons for the production of organic substances such as amino acids, nucleotides and structural carbohydrates, and can be used as signal molecules to coordinate source-sink relationships and resource utilization (Ruan 2012). Two main enzymes are responsible for introducing sucrose into plant metabolism: sucrose synthase (Susy, ec2.4.1.13) and invertase. Susy is glycosyltransferase that reversibly converts sucrose into UDP- glucose and fructose in the presence of UDP. In contrast, INV is hydrolase that irreversibly hydrolyzes sucrose into glucose and fructose.

INV, also called β -fructofuranosidase, means that the reaction catalyzed by the enzyme is the hydrolysis of the terminal non-reducing β -fructofuranoside residues in β -fructofuranoside (Romero-Gomez et al. 2000). According to subcellular location, INVs can be divided into cell wall invertase (CWINV), vacuolar invertase (VINV) and cytoplasmic invertase (CINV) (Sturm 1999). CWINVs and VINVs have similar conserved amino acid residues, belong to glycoside hydrolase family 32 (GH32), and contain two conserved sequences KNWINDPNGP and MWECDX (Gallagher et al. 2004; Den Ende et al. 2009). CINVs belong to glycoside hydrolase family 100 (GH100) and have almost no homology with CWINVs and VINVs (Lee and Sturm 1996). The different evolutionary origins of CWINVs, VINVs and CINVs are reflected in their biochemical and molecular properties. Insoluble CWINVs are glycosylated proteins bound to the

cell wall with molecular weights of 28-64kDa and optimal pH values of 3.5-5.0 (Verhaest et al. 2006). VINVs are soluble enzymes with molecular weights of about 70kDa and optimal pH values of 5.0-5.5 (Kim et al. 2011). These two are summarized as acid INVs (AINVs). Soluble CINVs are low-level expressed non-glycosylated polypeptides with molecular weights of 54-65kda, usually located in the cytoplasm, mitochondria, plastids and nucleus (Vargas and Salerno 2010). For their optimum pH values are 6.8-9.0, they are recognized as neutral/alkaline INVs (A/NINVs).

A/NINVs were found in all organs at different developmental stages of plants, especially in developing tissues. The presence of NINV in rice and *Arabidopsis thaliana* is essential for root cell development and reproductive development (Qi et al. 2007; Jia et al. 2008; Martin et al. 2013). The vegetative growth rate of *Arabidopsis thaliana* plants with two neutral invertase (NINV) subtypes deletion (*cin1/cin2* mutant) were seriously reduced (Barratt et al. 2009), because insufficient glucose production hinders cellulose biosynthesis and leads to the reduction of anisotropic growth and abnormal arrangement of cellulose (Barnes and Anderson 2018). In addition, A/NINVs were found involved in response to environmental stresses in wheat (Vargas et al. 2007). VINVs located in the vacuole usually perform some functions by adjusting the osmotic pressure, such as participating in the sweet taste regulation of tomatoes and potatoes), and response to drought and freezing stress (Kim et al. 2000; Andersen et al. 2002; Qian et al. 2018). Furthermore, the decrease of GhVIN expression in cottonseed coat seems to disturb the expression of some key genes involved in trehalose and auxin metabolism and signal transduction, revealing important roles of VINVs in flower organ development and male and female fertilization (Wang and Ruan 2016a, b).

CWINVs have been proved to play a crucial part in the tolerance of biotic stresses of pathogenic bacteria and other abiotic stresses (Webster et al. 2012; Albacete et al. 2015; Chang et al. 2017; Xu et al. 2017; Nishanth et al. 2018; Su et al. 2018; Chen et al. 2019). However, as a result of mainly expressed in anthers and grains, CWINVs directly affect plant male organ development and crop yield by participating in response to stresses (Goetz et al. 2017). OLIVER et al. found that low temperature induced the suppression of rice anther-specific CWINV gene OSINV4, which led to the obstruction of starch accumulation in pollen (Oliver et al. 2005). In wheat, the down-regulated expression of the anther-specific CWINV gene *lvr1* triggered by water stress was considered to be a sign of pollen development failure (Koonjul et al. 2004). Based on growing evidence, the function researches of CWINVs have moved from theoretical basis to practical applications such as obtaining male sterile plants and increasing plant seed setting rate. In tobacco and *Arabidopsis thaliana*, anther-specific RNA interference was used to inhibit the activity of CWINVs, effectively avoiding the supply of carbohydrates in pollen, and obtaining plants with reduced pollen germination ability and seed setting rate (Hirsche et al. 2009; Goetz et al. 2017). In *Arabidopsis*, rice and maize, the overexpression of CWINVs can promote the carbon distribution in the early stage of grain filling to improve grain yield and quality (Lammens et al. 2008; Wang et al. 2008; Coleman et al. 2009).

Wheat is the most important food crop worldwide, and with the continuous growth of world population and the decrease of cultivated land area, increasing wheat yield is of great significance to global food

demand security. The utilization of heterosis based on the development of male sterility is a widely recognized strategy to obtain high-yield and high-quality wheat varieties. Three subtypes of INV have been proved to have many important functions, such as regulating plant growth and development, responding to abiotic and biotic stresses, affecting fertility and yield. However, the members of the INV gene family in wheat have not been systematically identified and analyzed, and the fertility-related INVs that can be used in the development of male sterile wheat need to be discovered urgently. Therefore, we used the whole reference genome of wheat to identify the INV family comprehensively, and analyzed their physical and chemical properties, phylogenetic evolution, gene structures and duplication events, expression profiles, as well as screened out the candidate INVs related to male fertility by virus induced gene silencing (VIGS) verification.

Materials And Methods

Plant materials and treatments

KTM3315A was bred by the College of Agronomy of Northwest A&F University (34°29'N, 108°08'E), which is a two-line hybrid breeding material with thermo-sensitive male sterility characteristics (Ye et al. 2017). When the wheat reaches the stage of pollen development, the wheat was transplanted into flower pots and grown in two incubators with different growth conditions. Under sterile conditions (defined as AS) in the day and night 14h/10h, the temperature of the incubator is 17°C/15°C, the light is 20000xL/0xL, and the humidity is 50%/40%; under fertile conditions (defined as AF) the temperature of the incubator is 22°C/20°C in the 14h/10h day and night, the light is 2000xL/0xL, and the humidity is 50%/40%. When the wheat began to produce anthers, stain with acetic magenta dye to observe the anther stage. The fertile and sterile anthers in late uninucleate, binucleate, and trinucleate stages (Respectively referred to as Luns, Buns and Tns) were collected in different cryopreservation tubes and they were quickly frozen in liquid nitrogen and stored at -80 °C in the refrigerator.

The type of tobacco used in this study is *Nicotiana benthamiana*, The day and night time in the growth incubator is 16h/8h, the temperature is 25°C/23°C, the light intensity is 15000xL/0xL, and the humidity is 60%/50%.

Identification of wheat INV gene family

A total of 55 INVs amino acid sequences of the three plant species *Arabidopsis* (17), rice (19), and *Brachypodium* (19) were extracted from the research results of Wang et al. (Wang et al. 2017). All wheat protein sequences were downloaded from Ensembl plants (release 47, http://plants.ensembl.org/Triticum_aestivum/Info/Index) (Kersey et al. 2018), and were build a library using the make blastdb command. The candidate wheat INVs (TaINVs) were obtained through the alignment of the above 55 INV and wheat protein library using the blast all -p parameter. the final wheat INVs were identified after confirming by CDD (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) (Marchlerbauer et al. 2011), pfam (<http://pfam.xfam.org/>) (El-Gebali et al. 2018) and Interproscan

(<https://www.ebi.ac.uk/interpro/search/sequence/>) (Mitchell et al. 2019). Finally, the intersection of the three software identification results was finally used as the wheat INV gene family. In order to more accurately determine the subcellular location of TaINVs, the comprehensive prediction results of Plant-mPLOC software (Chou and Shen 2010) and ProtComp 9.0 server (www.softberry.com) were finally adopted. The ProtParam tool (<https://web.expasy.org/protparam/>) on the ExPASy website (Gasteiger et al. 2003) was used to obtain the physical and chemical properties, and MapGene2Chrom web (http://mg2c.iask.in/mg2c_v2.0/) was used to draw the chromosome distribution location map of *TaINVs*.

Phylogenetic tree construction for wheat INVs

Taking into account the genetic relationship, the amino acid sequences of 19 rice INVs, 19 *Brachypodium* INVs and 130 wheat INVs were used to draw a phylogenetic tree. We performed clustalw alignment of a total of 168 amino acid sequences by MEGA (Kumar et al. 2016) with default setting parameters, and removed the gaps in the alignment result. With the above data, a phylogenetic tree was constructed using neighbor joining (NJ) method by passion model and Bootstrap replication 1000.

Duplication events analysis of TaINVs

To analyze genome-wide replication event, all TaINVs amino acid sequences were blast. Among them, each three TaINVs with a similarity of more than 95% and on different sub-genomes of the same chromosome were defined as homologous copies. The members involved in segmental duplications and tandem duplications in TaINVs were obtained by Mscscanx software (Wang et al. 2012). Three standards were recognized as necessary for the identification of tandem duplication: sequence similarity greater than 75%; the length of similar fragment greater than 75%; on the same chromosome. The results of multiple sequence alignments that comply with these three standards were recognized as tandem duplications of TaINVs, and KA/KS values of tandem duplications were obtained by KAKS_calculator 2.0 (Guo et al. 2014).

Analysis of gene structures and conserved domains in TaINVs

The gene structure information of *TaINVs* was extracted from the gff3 file of the wheat reference genome (IWGSC RefSeq v1.1) (Appels et al. 2018) downloaded from Ensembl plants (release 47, http://plants.ensembl.org/Triticum_aestivum/Info/Index) (Kersey et al. 2018). Using MEME software (Bailey et al. 2009), we obtained 15 motifs with a length of 6-50 for AINVs and A/NINVs in wheat, respectively. Motifs and gene structures of TaINVs together with its phylogenetic tree were displayed with TB-tools (Chen et al. 2018).

Expression analysis of TaINVs

To analyze the expression pattern of *TaINVs*, we obtained the express abundance data of *TaINVs* in nine different stages and tissues (roots, leaves at seeding, vegetative and reproductive stages; spikes at vegetative stage; spikes and grains at reproductive stage) in Wheat Expression Browser database (Clavijo

et al. 2017; Borrill et al. 2016). To have a more intuitive understanding for *TaINVs*, the TPM (Trans Per Million) expression data was selected to draw a horizontal and vertical clustering heatmap.

In this study, Total RNA was extracted by TriGene reagent (GeneStar) and reverse transcribed into cDNA by StarScript II First-strand cDNA Synthesis Kit-II (GeneStar). We used 2·RealStar Green Fast Mixture (GeneStar) reagent to prepare the qRT-PCR reaction system according to the manufacturer's instructions, and detected the amplification value of the *TaINVs* through the default program of the Applied Biosystems 7500 Real-Time PCR System. Housekeeping gene *actin* was used as an internal reference. Each sample had three technical replicates and $2^{-\Delta\Delta Ct}$ method was used to calculate the relative expression of *TaINVs*. All gene quantitative primers were shown in Table 6.

The verification of subcellular localization and cis-elements for TaCWINV40

The following primers were used to clone the CDS sequence of *TaCWINV40* with the stop codon removed from the cDNA of fertile anthers: F-CGAGCTCAAGCTTCGAAATGGGGATGGCGTCG, R-CGACTGCAGAATTCGAACTGCCCCAGACGCTTGTTTCATCG, in which the underlined part represented the homologous arms. The cloned *TaCWINV40* CDS fragment was integrated into *pCAMBIA1302* vector with a green fluorescence protein (GFP) to construct TaCWINV40: GFP by homologous recombination. Finally, 35S: GFP (negative control) and TaCWINV40: GFP transformed into *Agrobacterium tumefaciens* strain GV3101 were introduced into 4-week-old *N. benthamiana* tobaccos, and transgenic tobacco leaves showed fluorescent signals after 36h-40h. In addition, a cell wall maker with red fluorescence protein (RFP) signals and TaCWINV40: GFP fusion protein were co-injected into tobacco epidermal cells for co-expression. The fluorescent signal and bright field of transgenic leaves were observed and photographed by confocal microscope (Nikon A1, Kanagawa, Japan).

Moreover, through PLACE database (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>) (Higo et al. 1999), we obtained the cis-elements on *TaCWINV40* promoter, and focused on the analysis of elements that combine transcription factors and pollen specific regulators.

Function validation of TaCWINV40

VIGS (Virus-induced gene silencing) refers to infecting plants with viruses that carrying target gene fragments can induce plant endogenous gene silencing and cause phenotypic changes, which is a popular method in recent years to study the function of target genes based on phenotypic variation (Burchsmith et al. 2004). The Barley Stripe Mosaic Virus-VIGS (BSMV-VIGS) system is commonly used in wheat, and the BSMV vector consists of three components, BSMV- α , BSMV- β , and BSMV- γ . To explore the function of *TaCWINV40*. The following primers were used to clone the 200bp CDS sequence of *TaCWINV40*: F: TAGCTAGCTGATTAATTAAGCTGGGCTAACAGGGATG, R: TTGCTAGCTGAGCGGCCGCGAGAAGTCGCCGTAGTCGTAT, in which the underlined part represented the homologous arms. Using *NotI* and *PacI* digestion sites in cloned sequence and BSMV- γ vector, the 200bp

CDS of *TaCWINV40* was inserted into the BSMV- γ vector to generate BSMV: *TaCWINV40* by homologous recombination. BSMV: 0 was composed of BSMV- α , BSMV- β , and BSMV- γ . In the same way, the CDS fragment of *Phytoene desaturase (PDS)* gene was inserted into BSMV- γ and used as positive control (BSMV:*PDS*), because its silence would cause photobleaching characteristics on the leaves. Recombinant BSMV: *TaCWINV40* vector, BSMV: *PDS* vector and the negative control BSMV- γ vector were mixed with BSMV- α and BSMV- β in equal proportions, respectively. Then added appropriate buffer solution and applied to the flag leaves of wheat plants (Yang et al. 2021).

Fertility identification and phenotyping of infected plants

The leaves and anthers of the plants infected were collected to observe the phenotype and analyze the gene expression level. Anthers in Tns were used to identify male fertility by DAPI, I₂-KI staining, and pollen tube germination. Pollen tube germination was determined by germination medium (20g sucrose, 10g PEG₄₀₀₀, 4mg H₃PO₃, 4mg Ca(NO₃)₂, 1mg VB₁, dilute to 100ml with distilled), cultured at 37°C for 30 minutes. The pictures were obtained with the microscope Olympus SZX10 (Japan). In addition, scanning electron microscope (SEM) and transmission electron microscope (TEM) were used to observe the cell morphology in anthers. The calculation of the seed setting rates adopted the percentage of total number of grains in the total number of florets.

Statistical analyses

Student's t-test in the Statistical Product and Service Solutions (SPSS) software were used to perform all statistical analyses. The differences were examined using Student's t-test and the significance level was set at 0.05 ($P < 0.05$).

Results

Identification and annotation of INV gene family members in wheat

By aligning all the protein sequences of wheat with the INVs of *Arabidopsis* (17), rice (19) and *Brachypodium distachyon* (19), 130 wheat INVs were identified and strictly screened (Table S1, Table S2). Twenty of them were evaluated as A/NINVs and belong to the GH100 family, and mainly expressed in cytoplasm and chloroplast. According to the position order, these 20 A/NINVs were named as TaA/NINV1-TaA/NINV20. In addition, 68 INVs were predicted to be expressed in the cell wall and were named TaCWINV1-TaCWINV68, and 42 INVs were named TaVINV1-42 for their subcellular location was vacuole. By analyzing their physical and chemical properties, it was found that the number of amino acids of A/NINV was 505-653 aa, and the corresponding molecular weight was 56.31-72.8 kDa. 509-670 aa and 56.61-74.97 kDa for VINVs, respectively. However, the amino acid number and molecular weight of CWINVs are smaller than those of the former two, which are 332-657 aa and 37.88-74.94 kDa. The theoretical isoelectric points of these 130 TaINVs span a wide range, from 4.69 to 9.31 (Table S3). By

mapping their chromosome distribution, we found that the numbers of *TaINVs* were unequal in the three A, B and D sub-genomes (A:B:D=52:33:41), but it can be predicted that most genes have three homologous copies because of their uniform location distribution in the three sub-genomes (Fig S1).

Phylogenetic analysis of *TaINVs*

In order to understand the evolutionary relationship of the wheat INV gene family, 130 *TaINVs* with 19 rice *INVs* (*OsCINVs*) and 19 *Brachypodium distachyon* *INVs* (*BdINVs*) were used to construct a phylogenetic tree for cluster analysis. Compared with the previous studies on rice (Ji et al. 2005), *Brachypodium* (Wang et al. 2017), the classification results of *TaINVs* were similar. The 130 *TaINVs* were divided into two major categories: *AINVs* and *A/NINVs*. The *AINVs* group included *CWINs* and *VINs*, *A/NINVs* were divided into two branches including α (*TaA/NINV1-11*) and β (*TaA/NINV12-20*) (Fig. 1). In the *A/NINVs* group, almost every *OsINV* and *BdINV* corresponded to three homologous *TaINVs* from the wheat ABD sub-genome (wheat: 20, rice: 8, *Brachypodium*: 8). In the *AINVs* group, the situations are a little more complicated. Compared with rice and *Brachypodium distachyon*, the number of *INVs* in wheat was more than three times that of them (wheat: 110, rice: 11, *Brachypodium*: 11), and a higher level of increase. There were only two *VINVs* in rice and three *VINVs* in *Brachypodium*, however, the number of *VINVs* in wheat was as high as 42.

The duplication of wheat INV gene family

Gene duplication often occurred in the whole genome, the main reasons were genome-wide duplication, tandem duplication, and segmental duplication (Zhang 2003). Of the 130 *TaINVs*, 54 genes had one copy on each of the three homologous chromosomes in the three sub-genomes, 43 genes had one copy on each of the two homologous chromosomes in the three sub-genomes (Table S4). By analyzing the segmental duplications, we found that there were 82 pairs of segmental duplications involving with 101 *TaINVs* (Fig. 2). However, except for the six pairs of segmented duplications produced by *AINV* from different chromosomes, the rest was duplications between different ABD sub-genomes of the same chromosome. Moreover, 13 *TaINVs* involved with 22 pairs of tandem duplications were found in the wheat *INVs* gene family (Table S5), which all occurred in *TaAINVs* and *KA/KS* values less than one. Overall, polyploidization of wheat played an important role in the expansion of *TaINVs*, and the segmental and tandem duplication events partly caused the significantly higher number of *AINVs* in wheat.

The gene structures and motifs of *TaINVs*

Introns were characteristic of eukaryotes, which were subject to relatively little selective pressure, resulting in rapid changes in the size and order of genes structures (Lecharny et al. 2003; Rogozin et al. 2003). However, the positional correspondence between introns and exons was usually highly conserved among homologous genes, so they were used to classify paralogous genes into subfamily (Park et al. 2008). Obviously, different subtypes of *TaINVs* had different numbers of introns and exons (Fig. 3). In *TaAINVs*, the number of exons ranged from two to nine. Except for a few genes, most *TaCWINVs* contain 5-9 exons,

and most *TaVINVs* contain 3-4 exons. In *TaA/NINVs*, except for *TaA/NINV2* (3 exons) and *TaA/NINV18* (7 exons), there were four and six exons in α subgroup and β subgroup, respectively. On the contrary, *TaAINVs* of the same subtype contain relatively uniform intron and exon, for example, all *TaAINVs* contain a mini-exon, which encodes NDP. Furthermore, homologous copies, or paralogs formed by duplication events, had almost the same number and structure of introns and exons.

From the amino acid level, motif as a super secondary structure could facilitate the identification of functional differentiation within gene family. This study identified 15 conserved motifs in the *TaA/NINV* and *TaAINV* groups respectively, and they were completely different (Fig. 3). In *TaAINVs*, motif1 (β -fructosidase motif NDPN), motif6 (RDP) and motif9 (WECV/VD) were essential markers, while motif3 and motif15 were specific to *CWINV*, and motif 10 was unique to *VINV*. In *TaA/NINVs*, there were nine motifs (motif1-3, 5-9, 12) shared by 20 *A/N-INVs*, while two motifs (motif11, motif15) were specific in α subgroup and one motif (motif13) specific in the β subgroup.

Eight *AINVs* were specific expressed in wheat spikes during reproductive stage

Compared to identifying members with specific motifs, the time- and space-specificity characteristics of gene expression always provided straightforward information for the study of gene functions related to the expression position. In order to explore the expression pattern of *TaINVs* and screen important ones, the RNA-seq data from roots, leaves, spikes, and grains in the different growth stages were analyzed. The results showed that 124 *TaINVs* were detected in the above tissues (Fig. 4). The tissue specificity of these *TaINVs* was more obvious than the period specificity, as a result, the expression patterns of *TaINVs* in the same tissue between different stages were similar. The expression patterns of most *TaINVs* in vegetative organs and reproductive organs were just opposite. Therefore, the expression profile of 124 *TaINVs* could be roughly divided into two categories as Fig. 4, and 63 of them expressed preference during the vegetative period and 61 highly expressed during the reproductive period. Notably, eight wheat *AINVs* (*TaCWINV40*, *TaCWINV53*, *TaVVIN27*, *TaCWINV46*, *TaCWINV68*, *TaVVIN7*, *TaCWINV36*, and *TaCWINV2*) specifically expressed in spikes.

Six *TaINVs* differentially expressed in anthers of KTM3315A under different fertility conditions

To further investigate the *TaINVs* that may be related to wheat male fertility, we performed qRT-PCR on eight wheat spikes-specific *TaINVs* and four *TaINVs* highly expressed in wheat spikes. The sterile and fertile anthers of thermo-sensitive male sterile wheat KTM3315A at three stages (uninucleate, binucleate, trinucleate) were used as materials. The results showed that six of them (*TaCWINV2*, *TaCWINV3*, *TaCWINV4*, *TaCWINV41*, *TaVINV7*, and *TaVINV27*) had no significant difference in expression level at each stage of sterile and fertile anthers (Fig. 5). The other six *TaINVs* (*TaCWINV36*, *TaCWINV40*, *TaCWINV43*, *TaCWINV46*, *TaCWINV53*, and *TaCWINV68*) were not only similar in expression patterns, but

also had significantly up-regulated expression in the fertile anthers than that of sterile anthers at binucleate stage. Interestingly, three genes, *TaCWINV40*, *TaCWINV46*, and *TaCWINV53*, are orthologous genes of rice *OsCWINV2* (*LOCos04g33720*) (Fig. 1), which has been revealed to lead male abortion when suppressed by low temperature (Oliver et al. 2005).

Silencing of *TaCWINV40* induces a decrease in wheat fertility

Rice *OsCWINV2* was confirmed to be a cell wall invertase, which was anther-specific, mainly by affecting the hexose production and starch formation (Oliver et al. 2005). Here, *TaCWINV40* was used as a representative to study whether it had similar functions to *OsCWINV2*. By fusion expression with green fluorescent protein, the subcellular location of *TaCWINV40* was confirmed to be the cell wall (Fig. 6A). The analysis of the promoter region of *TaCWINV40* revealed the presence of four important cis-elements (Fig. 6B). POLLEN1LELAT52 and GTGANTG10 have been reported in the promoters of the tomato *lat52* gene and tobacco late pollen gene *g10*, respectively, and as regulatory elements responsible for their pollen specific activation (Bate and Twell 1998; Rogers et al. 2001). WRKY71OS and MYCCONSENSUSAT were recognition sites of transcription factor WRKY and MYC, which were revealed to be a regulator of cold-induced transcriptome and a transcriptional repressor of the gibberellin signaling pathway, respectively (Zhang et al. 2004; Chinnusamy et al. 2003). These four cis-elements were distributed at least 3 sites on the promoter of *TaCWINV40*, which illustrated that *TaCWINV40* may be regulated by them and then specifically expressed in pollen and regulated by cold and gibberellin pathway.

To investigate the effect of *TaCWINV40* on wheat fertility, VIGS technology was carried out using KTM3315A plants grown in fertile environments (>24°C). The infected plants showed abnormal leaves on about 14 days, and the white spots of positive control plants BSMV: *PDS* indicated that the barley virus successfully infected the plants and effectively silenced *PDS* gene (Fig. 7A). The qRT-PCR result showed that the expression of *TaCWINV40* in the anthers of BSMV: *TaCWINV40* was significantly lower than that of the negative control plants BSMV: 0 (Fig. 7G), which indicated that *TaCWINV40* had been silenced. Although the anthers of BSMV: *TaCWINV40* plants were still cracking and pollen grains were formed (Fig. 7B), the pollen microspores stained with I₂-KI and DAPI showed sterile characteristics, that is, transparent shrunken vacuoles and two round sperm nuclei. On the contrary, in BSMV: 0 plants, the microspores possessed two spindle-shaped sperm nuclei, and they were all dyed into solid regular circles by I₂-KI due to fulling of starch (Fig. 7C, D). In addition, the pollen tubes of BSMV: *TaCWINV40* germinated extremely low, while which of BSMV: 0 almost germinated (Fig. 7E). SEM and TEM observations of trinucleate microspores support the key to further understanding BSMV: *TaCWINV40*, from which we observed the sparse arrangement and abnormal secretion of ubisch body, as well as the adhesion of shrinking microspores on the anther wall (Fig. 7F). In mature plants, seed setting rate of BSMV: *TaCWINV40* was significantly lower than that of BSMV: 0 (Fig. 7H). once again confirmed the indispensable role of *TaCWINV40* in wheat anther development and fertility determination.

Discussion

The wheat INV gene family is a big family

In the last two decades, INV gene family in various plant species have been successively identified, and which is recognized as a small gene family with the small number of INVs in each species. For example, 17 INVs were identified in *Arabidopsis thaliana* (Ji et al. 2005), 19 INVs in rice (Ji et al. 2005), and 19, 21, 14, 24, 32, and 20 in *Brachypodium* (Wang et al. 2017), maize (Juarezcolunga et al. 2018), sugarcane (Wang et al. 2017), poplar (Bocock et al. 2008), *Glycine max* (Su et al. 2018), and cassava (Yao et al. 2015), respectively. Here, however, we identified 130 TaINVs, which is undoubtedly a big family compared with the number of other family members in wheat and the number of INV gene family members in other species.

TaAINVs are preferentially tandem duplicated

Angiosperms naturally tend to gene duplication and subsequent gene loss leads to faster structural evolution (Moore and Purugganan 2003), non-dose sensitive genes and genes at the top or end of metabolic pathways are preferentially tandem duplicated (Freeling 2009; Yu et al. 2015). Additionally, during the evolution of vascular plants, genes affecting stress response also have a higher probability of retention after tandem duplication (Hanada et al. 2008; Cannon et al. 2004). *AINVs* are non-dose sensitive genes regulated by the sugar-sensitive mechanism, and they are also genes that further affects sugar metabolism in response to stress through their expression (Qian et al. 2018), the occurrence of tandem duplication can help them evolve an increased copy number to cope with the complex metabolism of the huge wheat genome. More gene copy number means being ready at any time to translate more sucrose invertase, so that one sucrose molecule can be quickly turned into two hexose molecules, and to quickly adjust the cell osmotic pressure to respond to the environment stress. In this study, more than half of the 22 pairs of tandem duplications (14 pairs) came from chromosomes 2 and 6 in wheat, and each of these INVs had at least two homologous copies. For example, *TaCWINV7* (Chr: 2A) and *TaCWINV8* (Chr: 2A) form a pair of tandem duplication, and they each have two other homologous copies (*TaCWINV17* and *TaCWINV24*, *TaCWINV18* and *TaCWINV25*) on chromosomes 2B and 2D. In short, *TaCWINV7* may first be doubled because of tandem duplication, and then tripled because of wheat polyploidization, so the number of *TaAINV* increased from the original one to at least six. Therefore, the significant expansion of *TaAINVs* should be the result of deliberate breeding of high-resistant varieties and natural selection of wheat itself.

INV gene family responds to temperature stress

Temperature stress is a kind of abiotic stress that TaINVs often responds to (Ruan et al. 2010). In tomatoes, CWINVs improve the fruit setting rate under long-term moderate heat stress by inhibiting the programmed cell death of non-reactive oxygen species, which may involve enhancing the introduction and catabolism of sucrose, the expression of HSP, and the reaction and biosynthesis of auxin (Liu et al. 2016). In order to protect tomato plants from low temperature stress, cold treatment will inhibit the

expression of *CWINV* inhibitors, so that *CWINVs* are strictly regulated by their inhibitors at the protein level, and the extracellular glucose and fructose content is maintained at an optimal level (Xu et al. 2017). The reproductive stage of flowering plants are very sensitive to temperature (Zinn et al. 2010), and temperature sensitive male sterile materials cascading amplify this sensitive. KTM3315A, the material of this study, shows sterility at normal autumn sowing temperature in Shaanxi, China, which is because it is so sensitive to temperature that 18°C is recognized as cold stress. Therefore, some spike-specific *TaINVs* down-regulated under sterile conditions, and the reason should be that they were inhibited just like rice *INVs* under cold stress (Oliver et al. 2005).

TaCWINV40 is pivotal to male fertility by affecting the function of tapetum

In this study, the cold response cis-element (MYCCONSENSUSAT, Fig. 6b) found in the promoter of *TaCWINV40* confirmed its response to low temperature, and the existence of two cis-elements required for pollen specific expression (GTGANTG10, POLLEN1LELAT52) predicted its expression site. In *Arabidopsis*, promoter analysis also confirmed the anther-specific expression pattern of sesame *Sicwinv1*, and the GUS staining was mainly detected in the anther tapetum (Zhou et al. 2019). The tapetum layer is the innermost layer of the microspore mother cell wall (epidermal layer, endothelial layer, middle layer and tapetum layer) after the anther morphogenesis of higher plants is completed, and its degradation at the right time is a necessary condition for the normal development of pollen (Mascarenhas 1989; Yi et al. 2016). The tapetum layer actively synthesizes proteins, fats, carbohydrates and other substances, and secretes them into the anther chamber to provide nutrients needed for the meiosis of the microspore mother cell and the development of microspores (Pacini et al. 1985). Our previous studies have shown that there is no difference in tapetum morphology between fertile and sterile anthers of KTM3315A at the uninucleate stage, but which in fertile anthers begin to degrade at the binucleate stage and almost completely absent at the trinucleate stage (Meng et al. 2016). Interestingly, in this study, we found that the expression level of *TaCWINV40* at the binucleate stage in fertile anthers was higher than that in sterile anthers, and the verification experiment showed that the silence of *CWINV40* led to the delayed degradation of tapetum and the thinning of ubisch body. Combining the above results, we speculated that *CWINV40* may be a key factor in promoting the degradation of the tapetum: when *CWINV40* was inhibited or expressed at low level, the tapetum was delayed degradation, which leading to insufficient energy supply of the pollen grains and ultimately male sterility. In fact, the provision of carbohydrates for plant male gametophytes is generally summarized as the mechanism of *INV* affecting fertility (Oliver et al. 2005).

The tapetum has also been reported to have another function, after the microspore mother cell completes meiosis, the tapetum secretes the β -1,3-glucanase in a timely manner to decompose the corpus callose that wraps the tetrad and release the microspores (Pacini et al. 1985). Under the control of the tapetum-specific promoter *Osg6B*, the introduced β -1,3-glucanase gene can prematurely dissolve the callosum layer of the pollen tetrad wall and cause male sterility (Tsuchiya et al. 1995). In this study, scanning electron microscope and transmission electron microscope showed an intuitive result, that is, after

TaCWINV40 was silenced, abnormal microspores adhered to the inner side of anther wall, and most of the microspores showed abnormal adhesion to each other, just like the phenotype that the tapetum fails to secrete β -1,3-glucanase. In addition, gibberellin was reported to induce β -1,3-glucanase genes up-regulation (Rinne et al. 2011), but a transcriptional repressor of the gibberellin signaling pathway (WRKY71OS) was found in the promoter region of *TaCWINV40*. Therefore, we speculated that *TaCWINV40* was an inducement of β -1,3-glucanase co-existing in the tapetum, and its down-regulation affected the synthesis of β -1,3-glucanase and caused abnormal adhesion of microspores. This finding is of great significance for guiding breeding practice of hybrid wheat.

Conclusions

Various life activities of plants are inseparable from the metabolism of sucrose. INV, as an important enzyme that hydrolyzes sucrose into hexose, participates in plant growth and development, response to stress, signal transduction and other important tasks. The publication of complete reference genome information of wheat provides great convenience and possibility for us to comprehensively identify members of the wheat INV gene family. In this study, we identified 130 wheat INVs and divided them into three categories: CWINV, VINV, and A/NINV according to their phylogeny. TaINVs with close genetic relationships show similarities in gene structures, motifs and expression patterns. Based on RNA-seq data and filtered by qRT-PCR, six *TaINVs* differentially expressed in wheat sterile and fertile anthers were identified. Inspired by the orthologous *LOCos04g33720*, we performed further functional verification of *TaCWINV40* which is subcellular located in the cell wall, and its silence can reduce wheat fertility and seed setting rate. Since it is expressed in the tapetum in other species, combined with the results of this study, we speculated that one of the mechanisms of *CWINV40* affecting pollen development is to regulate tapetum degradation, and the other was to induce the secretion of callose enzymes. This study is the first systematic identification of the wheat INV gene family, and the results of this study will help CWINVs become an ideal tool for the utilization of wheat heterosis in the near future.

Declarations

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jiali Ye, Yanning Bu, Mengting He, Yongfeng Wu, and Xuetong Yang. The first draft of the manuscript was written by Jiali Ye and Yanning Bu, all authors commented on previous versions of the manuscript. Xiyue song and Lingli Zhang critically revised the manuscript. All authors have read and approved the manuscript.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant number 32072060].

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

The authors declare no conflict of interest.

Ethical statement

There is no ethical statement to declare.

References

1. Albacete A, Canteronavarro E, Groskinsky DK, Arias CL, Balibrea ME, Bru R, Fragner L, Ghanem ME, Gonzalez MDLC, Hernandez JA (2015) Ectopic overexpression of the cell wall invertase gene CIN1 leads to dehydration avoidance in tomato. *Journal of Experimental Botany* 66 (3):863-878
2. Andersen MN, Asch F, Wu Y, Jensen CR, Naested H, Mogensen VO, Koch KE (2002) Soluble Invertase Expression Is an Early Target of Drought Stress during the Critical, Abortion-Sensitive Phase of Young Ovary Development in Maize. *Plant Physiology* 130 (2):591-604
3. Appels R, Eversole K, Stein N, Feuillet C, Keller B, Rogers J, Pozniak CJ, Choulet F, Distelfeld A, Poland J (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361 (6403)
4. Bailey TL, Boden M, Buske FA, Frith MC, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* 37:202-208
5. Barnes WJ, Anderson CT (2018) Cytosolic invertases contribute to cellulose biosynthesis and influence carbon partitioning in seedlings of *Arabidopsis thaliana*. *Plant Journal* 94 (6):956-974
6. Barratt DHP, Derbyshire P, Findlay K, Pike MJ, Wellner N, Lunn JE, Feil R, Simpson C, Maule A, Smith AM (2009) Normal growth of *Arabidopsis* requires cytosolic invertase but not sucrose synthase. *Proceedings of the National Academy of Sciences of the United States of America* 106 (31):13124-13129
7. Bate N, Twell D (1998) Functional architecture of a late pollen promoter: pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Molecular Biology* 37 (5):859-869
8. Bockock PN, Morse AM, Dervinis C, Davis JM (2008) Evolution and diversity of invertase genes in *Populus trichocarpa*. *Planta* 227 (3):565-576
9. Borrill P, Ramirez-Gonzalez R, Uauy C (2016) expVIP: a Customizable RNA-seq Data Analysis and Visualization Platform. *Plant Physiology* 170 (4):2172-2186
10. Burchsmith TM, Anderson JC, Martin GB, Dineshkumar SP (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant Journal* 39 (5):734-746

11. 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 (1):10-10
12. Caretto S, Linsalata V, Colella G, Mita G, Lattanzio V (2015) Carbon Fluxes between Primary Metabolism and Phenolic Pathway in Plant Tissues under Stress. *International Journal of Molecular Sciences* 16 (11):26378-26394
13. Chang Q, Liu J, Lin X, Hu S, Yang Y, Li D, Chen L, Huai B, Huang L, Voegelé RT (2017) A unique invertase is important for sugar absorption of an obligate biotrophic pathogen during infection. *New Phytologist* 215 (4):1548-1561
14. 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
15. Chen L, Liu X, Huang X, Luo W, Long Y, Greiner S, Rausch T, Zhao H (2019) Functional Characterization of a Drought-Responsive Invertase Inhibitor from Maize (*Zea mays* L.). *International Journal of Molecular Sciences* 20 (17):4081
16. Chinnusamy V, Ohta M, Kanrar S, Lee B-h, Hong X, Agarwal M, Zhu J-K (2003) ICE1: A regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Genes & development* 17:1043-1054
17. Chou K, Shen H (2010) Plant-mPLOC: A Top-Down Strategy to Augment the Power for Predicting Plant Protein Subcellular Localization. *PLOS ONE* 5 (6)
18. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman HD (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Research* 27 (5):885-896
19. Coleman HD, Yan J, Mansfield SD (2009) Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. *Proceedings of the National Academy of Sciences of the United States of America* 106 (31):13118-13123
20. Den Ende WV, Lammens W, Van Laere A, Schroeven L, Roy KL (2009) Donor and acceptor substrate selectivity among plant glycoside hydrolase family 32 enzymes. *FEBS Journal* 276 (20):5788-5798
21. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer EL L, Hirsh L, Paladin L, Piovesan D, Tosatto SC E, Finn RD (2018) The Pfam protein families database in 2019. *Nucleic Acids Research* 47 (D1):D427-D432
22. Freeling M (2009) Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem, Whole-Genome, Segmental, or by Transposition. *Annual Review of Plant Biology* 60 (1):433-453
23. Gallagher JA, Cairns AJ, Pollock CJ (2004) Cloning and characterization of a putative fructosyltransferase and two putative invertase genes from the temperate grass *Lolium temulentum* L. *Journal of Experimental Botany* 55 (397):557-569
24. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch AM (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* 31 (13):3784-3788

25. Goetz M, Guivarch A, Hirsche J, Bauerfeind MA, Gonzalez M, Hyun TK, Eom SH, Chriqui D, Engelke T, Groskinsky DK (2017) Metabolic Control of Tobacco Pollination by Sugars and Invertases. *Plant Physiology* 173 (2):984-997
26. Guo J, Jain R, Yang P, Fan R, Kwok CK, Zheng J (2014) Reliable and Fast Estimation of Recombination Rates by Convergence Diagnosis and Parallel Markov Chain Monte Carlo. *IEEE/ACM Trans Comput Biol Bioinform* 11 (1):63-72
27. Hanada K, Zou C, Lehtishiu MD, Shinozaki K, Shiu S (2008) Importance of Lineage-Specific Expansion of Plant Tandem Duplicates in the Adaptive Response to Environmental Stimuli. *Plant Physiology* 148 (2):993-1003
28. Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* 27 (1):297-300
29. Hirsche J, Engelke T, Voller D, Gotz M, Roitsch T (2009) Interspecies compatibility of the anther specific cell wall invertase promoters from Arabidopsis and tobacco for generating male sterile plants. *Theor Appl Genet* 118 (2):235-245
30. Ji X, Den Ende WV, Van Laere A, Cheng S, Bennett J (2005) Structure, evolution, and expression of the two invertase gene families of rice. *Journal of Molecular Evolution* 60 (5):615-634
31. Jia L, Zhang B, Mao C, Li J, Wu Y, Wu P, Wu Z (2008) OsCYT-INV1 for alkaline/neutral invertase is involved in root cell development and reproductivity in rice (*Oryza sativa* L.). *Planta* 228 (1):51-59
32. Juarezcolunga S, Lopezgonzalez C, Moraleselias NC, Massangesanchez JA, Trachsel S, Tiessen A (2018) Genome-wide analysis of the invertase gene family from maize. *Plant Molecular Biology* 97 (4):385-406
33. Kersey P, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalhosilva D, Christensen MB, Davis P, Grabmueller C (2018) Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research* 46
34. Kim D, Lee G, Chang M, Park J, Chung Y, Lee S, Lee T (2011) Purification and Biochemical Characterization of Insoluble Acid Invertase (INAC-INV) from Pea Seedlings. *Journal of Agricultural and Food Chemistry* 59 (20):11228-11233
35. Kim J, Mahe A, Brangeon J, Prioul J (2000) A Maize Vacuolar Invertase, IVR2, Is Induced by Water Stress. Organ/Tissue Specificity and Diurnal Modulation of Expression. *Plant Physiology* 124 (1):71-84
36. Koch KE (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Current Opinion in Plant Biology* 7 (3):235-246
37. Koonjul P, Minhas JS, Nunes C, Sheoran IS, Saini HS (2004) Selective transcriptional down-regulation of anther invertases precedes the failure of pollen development in water-stressed wheat. *Journal of Experimental Botany* 56 (409):179-190
38. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33 (7):1870-1874

39. Lammens W, Roy KL, Van Laere A, Rabijns A, Den Ende WV (2008) Crystal Structures of Arabidopsis thaliana Cell-Wall Invertase Mutants in Complex with Sucrose. *Journal of Molecular Biology* 377 (2):378-385
40. Lecharny A, Boudet N, Gy I, Aubourg S, Kreis M (2003) Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure. *Journal of Structural and Functional Genomics* 3 (1):111-116
41. Lee HS, Sturm A (1996) Purification and characterization of neutral and alkaline invertase from carrot. *Plant Physiology* 112 (4):1513-1522
42. Liu Y, Offler CE, Ruan Y (2016) Cell Wall Invertase Promotes Fruit Set under Heat Stress by Suppressing ROS-Independent Cell Death. *Plant Physiology* 172 (1):163-180
43. Marchlerbauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Deweesescott C, Fong JH, Geer LY, Geer RC, Gonzales NR (2011) CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Research* 39:225-229
44. Martin M, Lechner L, Zabaleta E, Salerno GL (2013) A mitochondrial alkaline/neutral invertase isoform (A/N-InvC) functions in developmental energy-demanding processes in Arabidopsis. *Planta* 237 (3):813-822
45. Mascarenhas JP (1989) The Male Gametophyte of Flowering Plants. *The Plant Cell* 1 (7):657-664
46. Meng L, Liu Z, Zhang L, Hu G, Song X (2016) Cytological characterization of a thermo-sensitive cytoplasmic male-sterile wheat line having K-type cytoplasm of *Aegilops kotschyi*. *Breed Sci* 66 (5):752-761
47. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H, Elgebali S, Fraser M (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* 47
48. Moore RC, Purugganan MD (2003) The early stages of duplicate gene evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100 (26):15682-15687
49. Nishanth MJ, Sheshadri SA, Rathore SS, Srinidhi S, Simon B (2018) Expression analysis of Cell wall invertase under abiotic stress conditions influencing specialized metabolism in *Catharanthus roseus*. *Scientific Reports* 8 (1):15059
50. Oliver SN, Van Dongen JT, Alfred SC, Mamun EA, Zhao X, Saini HS, Fernandes SF, Blanchard C, Sutton BG, Geigenberger P (2005) Cold-induced repression of the rice anther-specific cell wall invertase gene OSINV4 is correlated with sucrose accumulation and pollen sterility. *Plant Cell and Environment* 28 (12):1534-1551
51. Pacini E, Franchi GG, Hesse M (1985) The tapetum: Its form, function, and possible phylogeny in Embryophyta. *Plant Systematics and Evolution* 149 (3):155-185
52. Park KCPC, Kwon SJKJ, Kim PHKH, Bureau TB, Kim NSKS (2008) Gene structure dynamics and divergence of the polygalacturonase gene family of plants and fungus. *Genome* 51 (1):30-40
53. Qi X, Wu Z, Li J, Mo X, Wu S, Chu J, Wu P (2007) AtCYT-INV1, a neutral invertase, is involved in osmotic stress-induced inhibition on lateral root growth in Arabidopsis. *Plant Molecular Biology* 64

(5):575-587

54. Qian W, Xiao B, Wang L, Hao X, Yue C, Cao H, Wang Y, Li N, Yu Y, Zeng J (2018) CsINV5 , a tea vacuolar invertase gene enhances cold tolerance in transgenic Arabidopsis. *BMC Plant Biology* 18 (1):228
55. Rinne PL, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjarvi J, van der Schoot C (2011) Chilling of dormant buds hyperinduces FLOWERING LOCUS T and recruits GA-inducible 1,3-beta-glucanases to reopen signal conduits and release dormancy in Populus. *Plant Cell* 23 (1):130-146
56. Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, Lonsdale DM, Twell D (2001) Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. *Plant Molecular Biology* 45 (5):577-585
57. Rogozin I, Wolf Y, Sorokin A, Mirkin B, Koonin E (2003) Remarkable Interkingdom Conservation of Intron Positions and Massive, Lineage-Specific Intron Loss and Gain in Eukaryotic Evolution. *Current biology : CB* 13:1512-1517
58. Romero-Gomez SJ, Augur C, Viniestra-Gonzalez G (2000) Invertase production by *Aspergillus niger* in submerged and solid-state fermentation. *Biotechnology Letters* 22 (15):1255-1258
59. Ruan Y (2012) Signaling Role of Sucrose Metabolism in Development. *Molecular Plant* 5 (4):763-765
60. Ruan Y, Jin Y, Yang Y, Li G, Boyer JS (2010) Sugar Input, Metabolism, and Signaling Mediated by Invertase: Roles in Development, Yield Potential, and Response to Drought and Heat. *Molecular Plant* 3 (6):942-955
61. Sturm A (1999) Invertases. Primary Structures, Functions, and Roles in Plant Development and Sucrose Partitioning. *Plant Physiology* 121 (1):1-8
62. Su T, Han M, Min J, Chen P, Mao Y, Huang Q, Tong Q, Liu Q, Fang Y (2018) Genome-Wide Survey of Invertase Encoding Genes and Functional Characterization of an Extracellular Fungal Pathogen-Responsive Invertase in *Glycine max*. *International Journal of Molecular Sciences* 19 (8):2395
63. Tsuchiya T, Toriyama K, Yoshikawa M, Ejiri S, Hinata K (1995) Tapetum-Specific Expression of the Gene for an Endo- β -1,3-glucanase Causes Male Sterility in Transgenic Tobacco. *Plant and Cell Physiology* 36 (3):487-494
64. Vargas WA, Pontis HG, Salerno GL (2007) Differential expression of alkaline and neutral invertases in response to environmental stresses: characterization of an alkaline isoform as a stress-response enzyme in wheat leaves. *Planta* 226 (6):1535-1545
65. Vargas WA, Salerno GL (2010) The Cinderella story of sucrose hydrolysis: Alkaline/neutral invertases, from cyanobacteria to unforeseen roles in plant cytosol and organelles. *Plant Science* 178 (1):1-8
66. Verhaest M, Lammens W, Roy KL, De Coninck B, De Ranter CJ, Van Laere A, Den Ende WV, Rabijns A (2006) X-ray diffraction structure of a cell-wall invertase from *Arabidopsis thaliana*. *Acta Crystallographica Section D-biological Crystallography* 62 (12):1555-1563
67. Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nature Genetics* 40 (11):1370-

68. Wang L, Ruan Y (2016a) Critical Roles of Vacuolar Invertase in Floral Organ Development and Male and Female Fertilities Are Revealed through Characterization of GhVIN1-RNAi Cotton Plants. *Plant Physiology* 171 (1):405-423
69. Wang L, Ruan YL (2016b) Shoot-root carbon allocation, sugar signalling and their coupling with nitrogen uptake and assimilation. *Functional Plant Biology* 43 (2):105-113. doi:10.1071/Fp15249
70. Wang L, Zheng Y, Ding S, Zhang Q, Chen Y, Zhang J (2017) Molecular cloning, structure, phylogeny and expression analysis of the invertase gene family in sugarcane. *BMC Plant Biology* 17 (1):109
71. 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 Res* 40 (7):e49
72. Webster H, Keeble G, Dell B, Fosunyarko J, Mukai Y, Moolhuijzen P, Bellgard M, Jia J, Kong X, Feuillet C (2012) Genome-level identification of cell wall invertase genes in wheat for the study of drought tolerance. *Functional Plant Biology* 39 (7):569-579
73. Xu X, Hu Q, Yang W, Jin Y (2017) The roles of cell wall invertase inhibitor in regulating chilling tolerance in tomato. *BMC Plant Biology* 17 (1):195
74. Yang X, Ye J, Niu F, Feng Y, Song X (2021) Identification and verification of genes related to pollen development and male sterility induced by high temperature in the thermo-sensitive genic male sterile wheat line. *Planta* 253(4):83
75. Yao Y, Geng M, Wu X, Liu J, Li R, Hu X, Guo J (2015) Genome-Wide Identification, Expression, and Activity Analysis of Alkaline/Neutral Invertase Gene Family from Cassava (*Manihot esculenta* Crantz). *Plant Molecular Biology Reporter* 33 (2):304-315
76. Ye J, Duan Y, Hu G, Geng X, Zhang G, Yan P, Liu Z, Zhang L, Song X (2017) Identification of candidate genes and biosynthesis pathways related to fertility conversion by wheat KTM3315A transcriptome profiling. *Frontiers in Plant Science* 8 (449)
77. Yi J, Moon S, Lee Y, Zhu L, Liang W, Zhang D, Jung K, An G (2016) Defective Tapetum Cell Death 1 (DTC1) Regulates ROS Levels by Binding to Metallothionein during Tapetum Degeneration. *Plant Physiology* 170 (3):1611-1623
78. Yu J, Ke T, Tehrim S, Sun F, Liao B, Hua W (2015) PTGBase: an integrated database to study tandem duplicated genes in plants. *Database : the journal of biological databases and curation* 2015
79. Zhang J (2003) Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 18 (6):292-298
80. Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TH, Shen QJ (2004) A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant Physiol* 134 (4):1500-1513
81. Zhou T, Hao G, Yang Y, Liu H, Yang M, Zhao Y (2019) Sicwinv1 , a Cell Wall Invertase from Sesame, Is Involved in Anther Development. *Journal of Plant Growth Regulation* 38 (4):1274-1286

A phylogenetic tree constructed jointly by the INV gene family of wheat, rice, and *Brachypodium dilatatum*. Yellow, green and orange backgrounds are used to distinguish VINs, CWINVs and A/NINVs marked by the outer circle in the phylogenetic tree. The blue and purple fonts represent the gene IDs of rice and *Brachypodium dilatatum*, respectively, and the black fonts represent the names of wheat INVs.

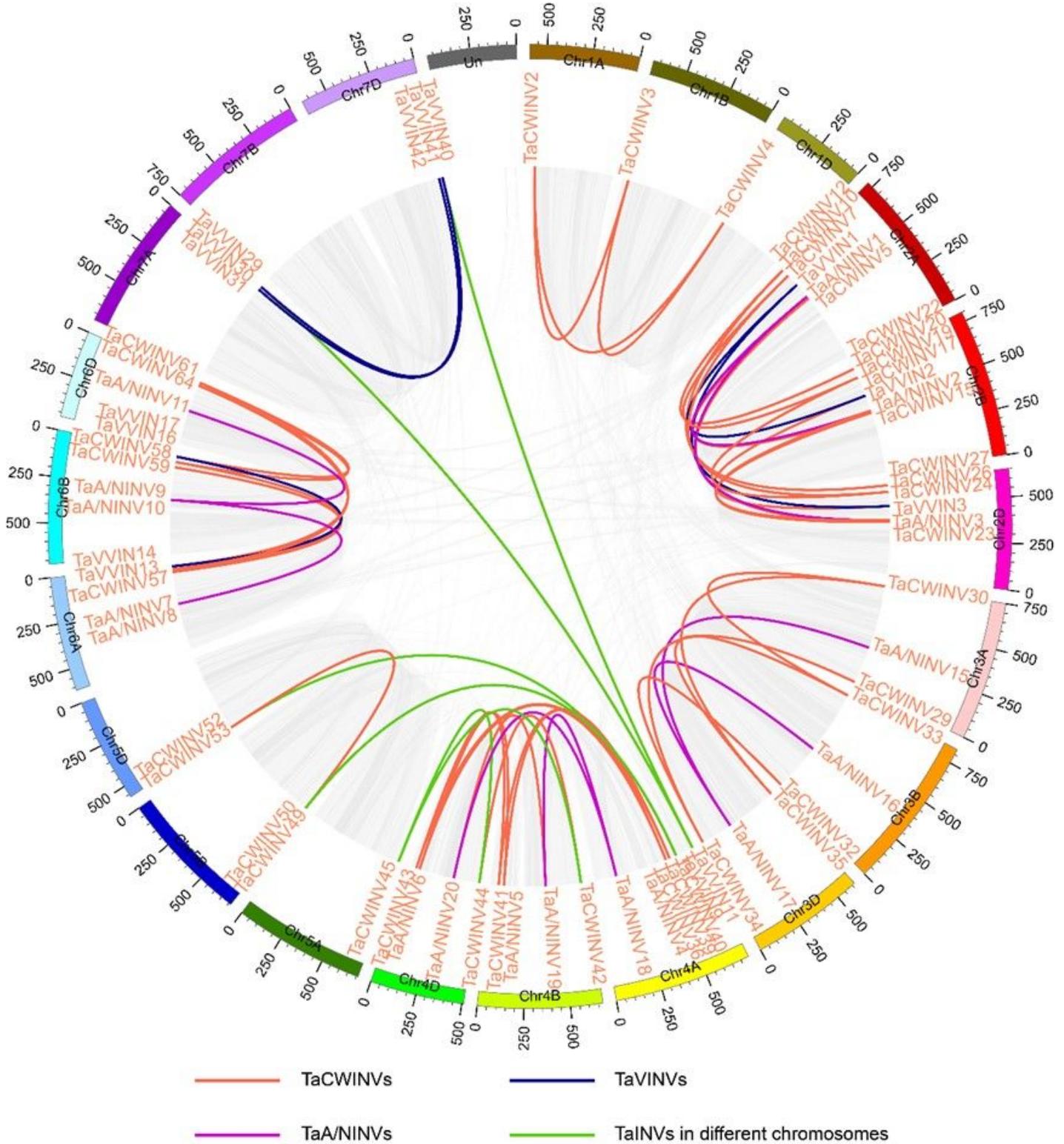


Figure 2

Segmental duplications of wheat INV gene family predicted by Mcscanx software. The outermost segment and the scale above it represent the 21 chromosomes of wheat and the positions on the chromosomes, and the un is used to place those members that have not been mapped to specific chromosomes. The orange, blue and purple lines respectively connect the two CWINVs, VINVs and A/NINVs that participate in the same segmental duplications, and these lines all connect INVs that on different sub-genomes of the same chromosome. The green lines represent segmental duplication between INVs on different chromosomes.

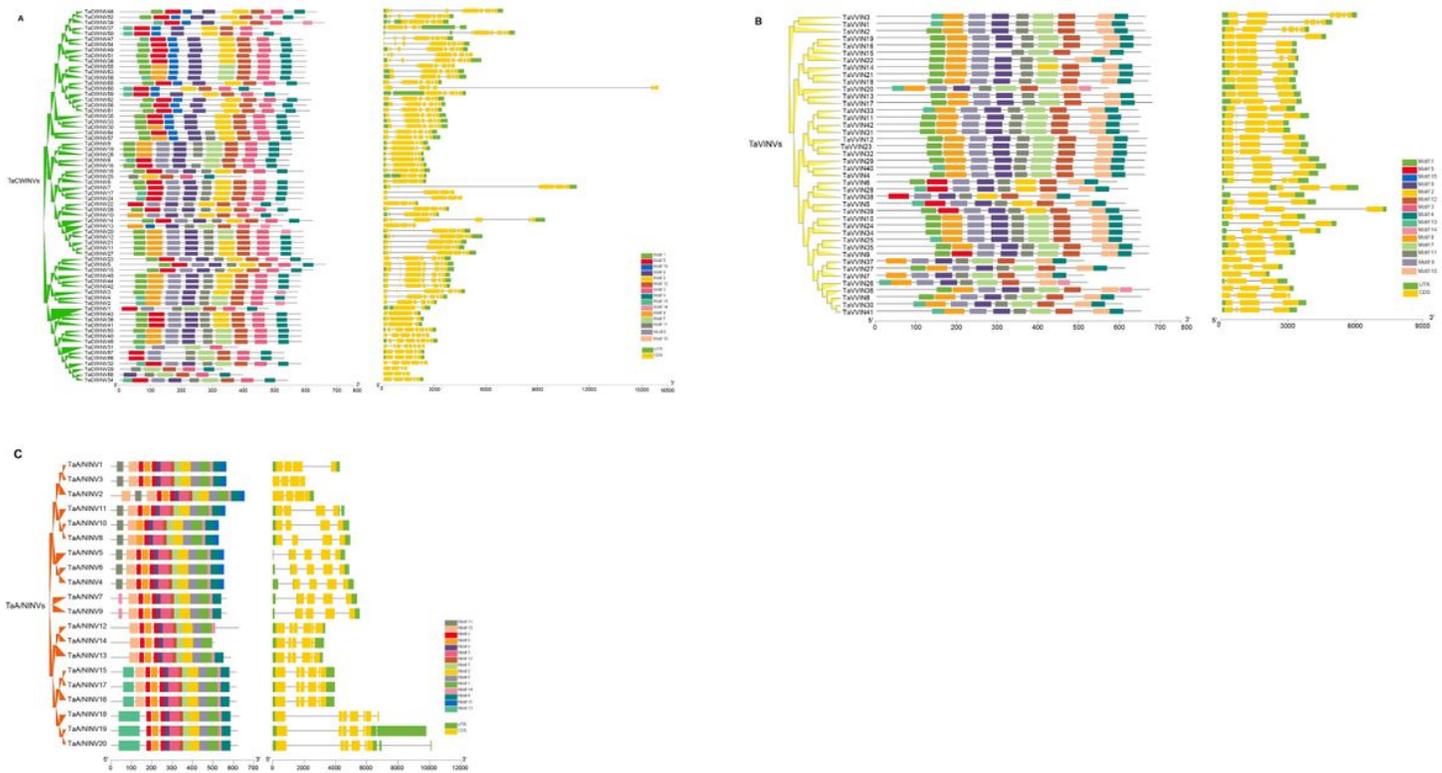


Figure 3

The gene structure and protein motifs of 130 wheat invertases. (A-C) The gene structure and protein motifs of CWINVs, VINVs, A/NINVs respectively. The rectangles with different colors represent different motifs, as shown in the legend on the right. The two rulers at the bottom of each figure are used to measure the position of motif or exon in amino acid sequence and gene sequence. The phylogenetic tree on the far-left clusters members with similar structures together.

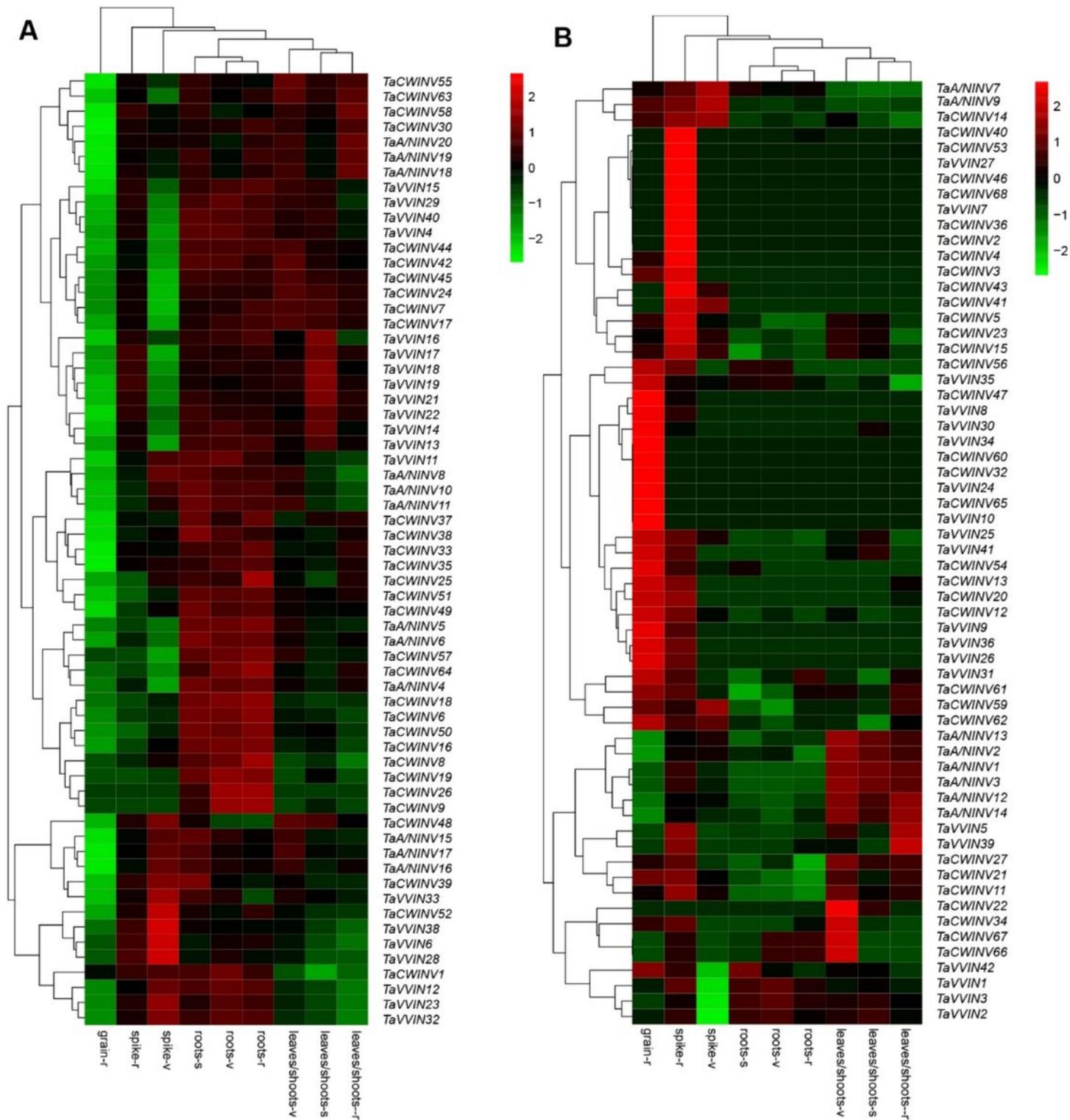


Figure 4

Heatmap of expression profiles of TaINVs in 9 different development stages and tissues. (A) The expression heatmap of TaINVs with relatively high expression in the vegetative tissues. (B) The expression heatmap of TaINVs with relatively high expression in the reproductive tissues. Horizontal and vertical clustering are used to observe TaINVs with similar expression patterns. The annotations on the

left represent gene names, the annotations at the bottom represent different tissues and periods, -s, -v, -r represent seedling stage, vegetative stage and reproductive stage, respectively.

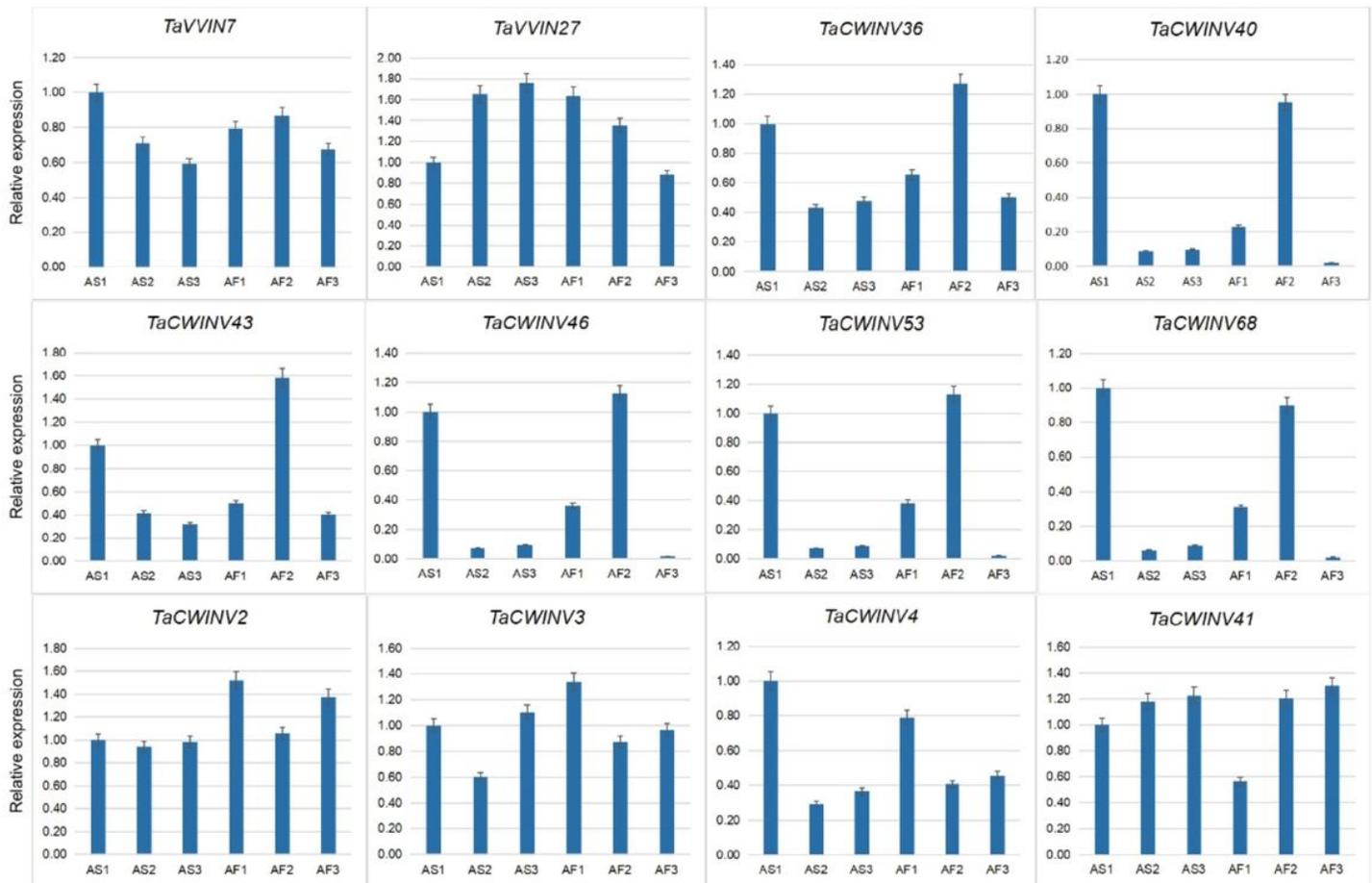


Figure 5

QRT-PCR verification of the differential expression of 12 TaINVs in sterile and fertile anthers of KTM3315A. The AS1-3 and AF1-3 on the abscissa represent the uninucleate, binucleate and trinucleate stages of the sterile and fertile anthers of KTM3315A, respectively. The ordinate represents the relative expression level of each gene, with the expression level in uninucleate stage of sterile anthers as the control. Error bars represent the standard deviation of three replicates.

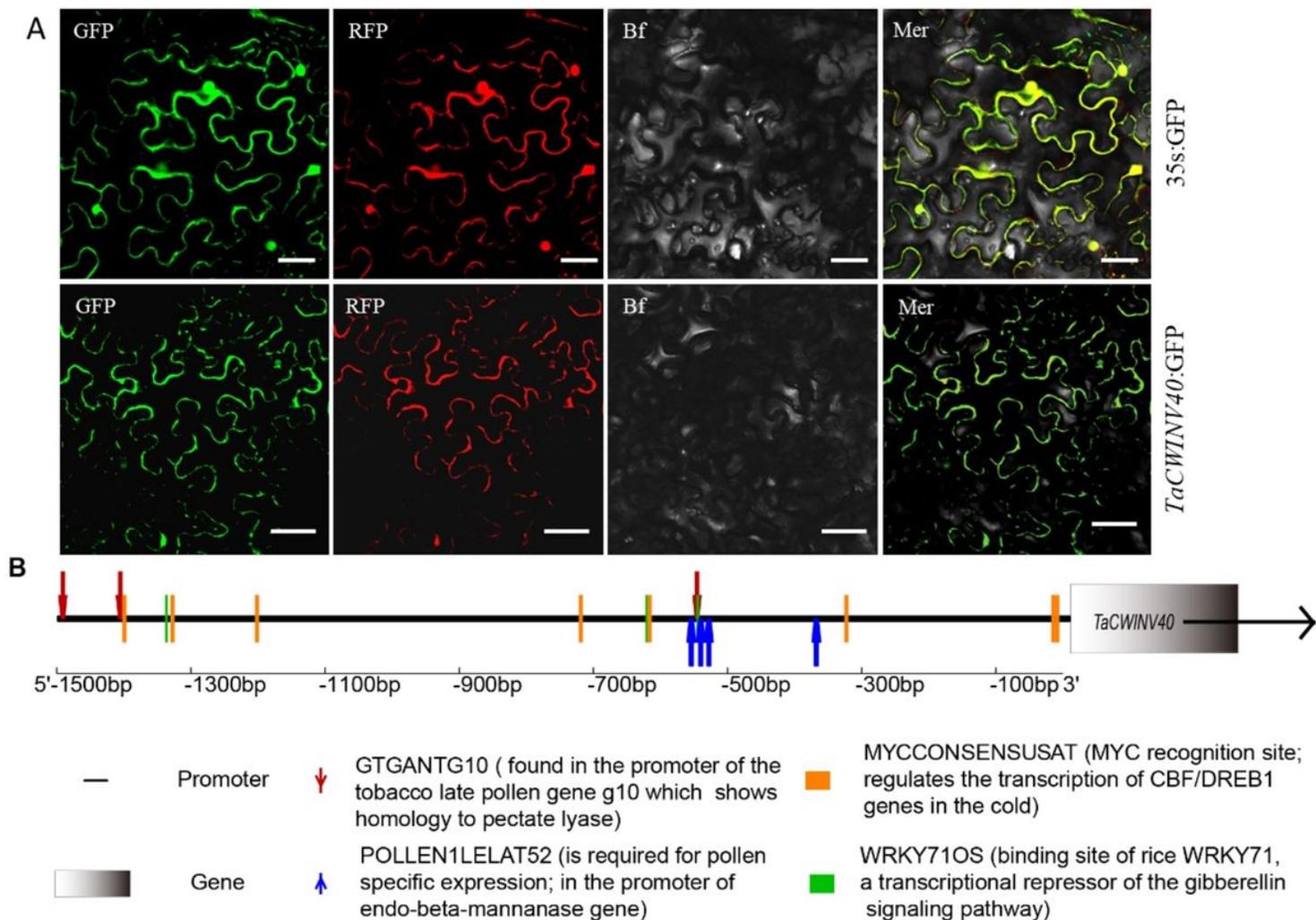


Figure 6

Protein subcellular localization and promoter cis-element of TaCWINV40. (A) Subcellular localization of TaCWINV40 in tobacco leaves. GFP: Green fluorescent protein. RFP: Red fluorescent protein. Bf: Bright field. Mer: Merged image. Scale bars=20µm (35s:GFP) 40µm (TaCWINV40:GFP). (B) Important cis-elements in the promoter of TaCWINV40.

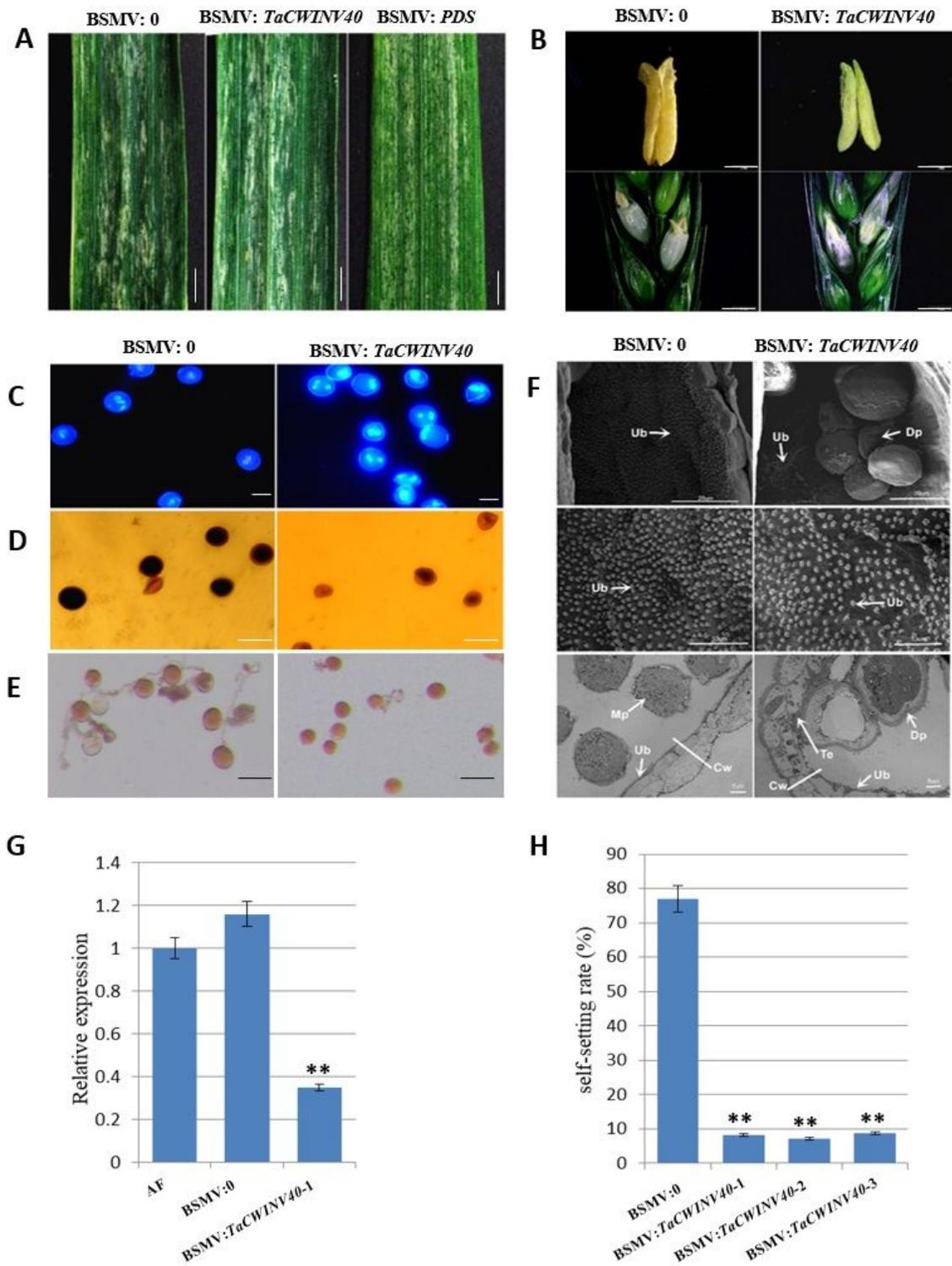


Figure 7

Phenotypic characteristics of BSMV-VIGS-mediated *TaCWINV40* silencing in KTM3315A plants. (A) The leaf phenotypes of plants BSMV: 0, BSMV: *PDS*, BSMV: *TaCWINV40* after treatment 14 days. Bar=50 μ m. (B) anthers and spikes phenotypes of BSMV: 0 and BSMV: *TaCWINV40*. Bar=50 μ m. (C) Pollen grains of BSMV: 0 and BSMV: *TaCWINV40* stained with DAPI. Bar=10 μ m. (D) Pollen grains of BSMV: 0 and BSMV: *TaCWINV40* stained with KI. Bar=50 μ m. (E) Pollen tube germination results of BSMV: 0 and BSMV:

TaCWINV40. Bar=50 μm . (F) Ultra-micro observation of Ubisch body and microspores of BSMV: 0 and BSMV: TaCWINV40. Cw, cell wall of anther; Dp, deformed pollen; Mp, mature pollen; Te, tectum; Ub, Ubisch body. (G) The relative expression level of TaCWINV40 in the anthers at the binucleate stage, AF stands for fertile anthers without virus infection, and the expression level of TaCWINV40 in AF was used as a control for the treatment plants BSMV: 0 and BSMV: TaCWINV40 to calculate the relative expression level. (H) Self-setting rate of BSMV: 0 and BSMV: TaCWINV40. Each data had three repetitions. Analysis of significance of differences was identified by Students't text (* $p < 0.05$, ** $p < 0.01$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.docx](#)
- [TableS1.xls](#)
- [TableS2.xls](#)
- [TableS3.xls](#)
- [TableS4.xls](#)
- [TableS5.xls](#)
- [TableS6.xls](#)