

# Genome-wide characterization of WRKY gene family in *Helianthus annuus* L. and their expression profiles under biotic and abiotic stresses

**Chong Yang**

Guangzhou Sugarcane Industry Research Institute

**Juanjuan Li**

Zhejiang Sci-Tech University

**Faisal Islam**

Zhejiang University

**Luyang Hu**

Zhejiang University

**Jiansu Wang**

Zhejiang University

**Jian Wang**

Zhejiang University

**Ling Xu** (✉ [lxu@zstu.edu.cn](mailto:lxu@zstu.edu.cn))

Zhejiang Sci-Tech University

**Weijun Zhou**

Zhejiang University

---

## Research article

**Keywords:** *Helianthus annuus* L., WRKY, expression pattern, abiotic and biotic stress, evolution

**Posted Date:** March 17th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-17430/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** WRKY transcription factors play important roles in various physiological processes and stress responses in flowering plants. However, the information about WRKY genes in *Helianthus annuus* L. (common sunflower) is limited.

**Results:** Ninety WRKY (*HaWRKY*) genes were identified and renamed according to their locations on chromosomes. Further phylogenetic analyses classified them into four main groups including a species-specific WKKY group and *HaWRKY* genes within same group or subgroup generally showed similar exon-intron structures and motif compositions. The tandem and segmental duplication possibly contributed to the diversity and expansion of *HaWRKY* gene families. Synteny analyses of sunflower WRKY genes provided deep insight to the evolution of *HaWRKY* genes. Transcriptomic and qRT-PCR analyses of *HaWRKY* genes displayed distinct expression patterns in different plant tissues, as well as under various abiotic and biotic stresses.

**Conclusions:** Ninety WRKY (*HaWRKY*) genes were identified from *H. annuus* L. and classified into four groups. Structures of *HaWRKY* proteins and their evolutionary characteristics were also investigated. The characterization of *HaWRKY* genes and their expression profiles under biotic and abiotic stresses in this study provide a foundation for further functional analyses of these genes and will be beneficial to crop improvement.

## Background

The WRKY gene family is considered as one of the largest transcription factor (TF) family in higher plants [1], which basically contain an approximate 60-residue DNA-binding domain, named as WRKY domain, with a highly conserved heptapeptide motif WRKYGQK and a C<sub>2</sub>H<sub>2</sub>- or C<sub>2</sub>HC-type of zinc-finger motif included. Both the heptapeptide motif and zinc-finger motif are needed for binding of WRKY TFs to the cis-acting element W-box (C/T)TGAC(C/T) [2–3]. WRKY gene family can be classified into three main groups (I-III), based on the number of WRKY domains and the structure of their zinc-finger motifs [4]. The group I WRKY proteins consist of two WRKY domains, whereas groups II and III contain only one. The group II and III WRKY proteins are distinguished by the type of zinc-finger motif, with a C-X<sub>4</sub>-5-C-X<sub>22-23</sub>-H-X<sub>1</sub>-H type of motif in group II and a C-X<sub>7</sub>-C-X<sub>23</sub>-H-X<sub>1</sub>-C type in group III [5].

The first WRKY gene was cloned and identified from sweet potato, encoding a 549 amino acid protein called SPF1 (SWEET POTATO FACTOR1) [6]. Since then, a large number of WRKY genes have been discovered from different plants. Functional analyses show that WRKY genes are associated with various aspects of physiological processes, including seed dormancy and germination, root development, leaf senescence, modulation of flowering time, plant nutrient utilization etc. [7]. The knockout mutant of *AtWRKY2* resulted in hypersensitivity of *Arabidopsis* to ABA during seed germination and post-germination, early growth, suggesting that *AtWRKY2* mediates seed germination and post-germination development [8]. Overexpression of *OsWRKY31* in rice inhibited plant lateral root formation and

elongation, and also affected the transport process of auxin [9]. AtWRKY12, AtWRKY13 and AtWRKY71 are three main genes regulating Arabidopsis flowering time, with AtWRKY12 and AtWRKY13 working antagonistically under short daylight conditions [10], and AtWRKY71 accelerating flowering [11]. It has been also documented that 12 WRKY genes are involved in leaf senescence in Arabidopsis and rice, as their mutants inhibited or promoted leaf senescence to different extents [7].

In addition to plant growth and development, WRKY genes also participate in modulation of plant tolerance to abiotic and biotic stress. Qiu and Yu [12] reported that overexpression of OsWRKY45 in Arabidopsis significantly increased the expression level of PR genes and ABA/stress regulated genes, thus contributed to the enhancement of disease resistance and salt and drought tolerance of the plant. GmWRKY54 from soybean, which was confirmed in a DNA binding assay that could interact with the W-box, conferred salt and drought tolerance to transgenic Arabidopsis, possibly through the regulation of DREB2A and STZ/Zat10 [13]. In tobacco, overexpression of grape VvWRKY2 reduced the susceptibility to fungal pathogens like *Botrytis cinerea*, *Pythium* spp. and *Alternaria tenuis* [14]. The WRKY1 in tobacco could be phosphorylated by a salicylic acid-induced protein kinase (SIPK), resulting in enhanced DNA-binding activity to a W-box sequence from the tobacco chitinase gene CHN50, and subsequently formation of hypersensitive response-like cell death [15]. Thus, WRKY genes may be involved in mitigating the damage caused by stresses, through interacting with the cis-element W-box and activating downstream plant defense signaling [2].

Common sunflower (*Helianthus annuus* L.) is grown throughout the world as an industrial crop for edible oil and fruits. It is the fourth important oilseed crop which contributes to 12% of the edible oil produced globally. However, sunflower production has been threatened by different stresses, among which drought and salinity are two major abiotic constraints [16] and parasitic weed *Orobanche cumana* is a new emerged biotic issue worldwide [17]. WRKY transcription factors are involved in regulation of plant tolerance to both abiotic and biotic stresses. Thus, it is of great interest to characterize a WRKY gene family in sunflower and identify their functions under different stresses.

The WRKY gene family has been well studied in many plant species. However, the identification of WRKY gene family in *H. annuus* is limited. Although Giacomelli et al. [18] have identified a total number of 97 WRKY genes in the Asteraceae, only 26 of them belong to *H. annuus*, and this identification was all based on EST database. The publication of common sunflower reference genome provided an opportunity to reveal the organization, expression and evolutionary traits of common sunflower WRKY gene family at the genome-wide level. In the current study, we investigated WRKY gene family in common sunflower and identified 90 WRKY genes, which could be classified into four main groups. Analyses on exon-intron organization, motif composition, gene duplication, chromosome distribution, phylogenetic relationship and gene synteny were further conducted to systemically characterize these common sunflower WRKY genes. Additionally, the expression patterns of HaWRKY genes in different plant tissues and in responses to different abiotic and biotic stress were also recorded, to identify the implication of specific WRKY genes in different biological processes. The present findings provide a foundation for future research on functional characterization of WRKY genes in common sunflower.

## Results

### Identification of the WRKY genes

A total of 104 candidate WRKY genes were predicted from *H. annuus*, among which 14 were removed after domain check and the rest 90 were named as HaWRKY1-HaWRKY90 (Supplementary Table 2). The length of these genes ranged from 388 bp (HaWRKY5) to 8445 bp (HaWRKY49). HaWRKY51 was identified as the smallest protein (88 aa) while the largest one was HaWRKY52 (679 aa), with molecular weight (MW) 10.48 and 74.25 kDa, respectively. The isoelectric point (pI) of these proteins ranged from 4.81 (HaWRKY86) to 10.44 (HaWRKY53) (Supplementary Table 2).

### Phylogenetic analysis of WRKY family members

Phylogenetic analysis of WRKY family members of *H. annuus* and *A. thaliana* was conducted. The common sunflower WRKY domains were divided into four large groups (Fig. 1 and Fig. 2), corresponding to group I, II and III in Arabidopsis [2] and an extra WKKY group. Among 90 HaWRKY family members, group II accounts for the largest part with 48 HaWRKY proteins, followed by group I with 18 proteins and group III with 17 proteins (Fig. 1). There were 7 HaWRKY proteins in group WKKY, with C-X<sub>5</sub>-C-X<sub>23</sub>-H-X-H type zinc-finger motifs (Fig. 2), which is not found in Arabidopsis (Fig. 1). In addition, each group could be divided into several subgroups. Proteins with two WRKY domains were assigned as the N-terminal and the C-terminal WRKY domains according to their locations on protein. The proteins grouped either in N-terminal or C-terminal WRKY domains, usually followed by C<sub>2</sub>H<sub>2</sub>-type zinc-finger motifs (C-X<sub>4</sub>-C-X<sub>22-23</sub>-H-X-H), were classified as group I, with 16 identified as N-terminal WRKYs (I N) and 15 C-terminal (I C), and among them 13 members contained two WRKY domains (Fig. 2). Group II of HaWRKY family could be clustered into five subgroups, with four in IIa, 10 in IIb, 13 in IIc, 11 in IId and 10 in IIe (Fig. 1, Supplementary Table 2). 17 members of HaWRKYs in group III contained the C-X<sub>7</sub>-C-X<sub>23</sub>-H-X-C type zinc-finger motifs (Fig. 2, Supplementary Table 2), and were classified as subgroup IIIa. There were no HaWRKY proteins found in subgroup IIIb, not as that in Arabidopsis (Fig. 1).

### Gene structure and motif composition of WRKY family members

The distributions of exons and introns on HaWRKY genes were investigated via GSDS program, to gain further insight into the structure diversity of the WRKY family in sunflower. As shown in Fig. 3B, almost half number (43) of the HaWRKY genes had three exons, followed by 17 with two exons, 17 with four exons, 12 with five exons and one with seven exons. Genes within same groups generally shared similar structures, such as group IIIa, in which all HaWRKY genes possessed three exons and two introns. Most of the WRKY domains spanned an exon-exon junction, whereas HaWRKY genes with two WRKY domains in the group I at least had one complete domain within one exon, except HaWRKY1. Further analyses on introns indicated that HaWRKY genes only with phase-0 introns (between two consecutive codons) were clustered into group IIa and IIb, and only with phase-2 introns (between the second and third nucleotide of

a codon) into the group IId, Ile and IIIa. The phase-1 introns (between the first and second nucleotide of a codon) were widely distributed among these groups, except group IIa.

Motif structures on HaWRKY proteins were constructed via MEME program. As exhibited in Fig. 3C, HaWRKY family members within same groups usually have similar motifs with similar arrangements. Motifs 1 and 6 are WRKY domains, with motif 6 only limited in group I, whereas motif 1 distributed all over the groups. Motifs 12, 15, 17 and 18 are unique to group IId, as well as motifs 11, 13 and 19 to group Ile, and motif 20 to group IIIa. In addition, the clustered HaWRKY pairs, like HaWRKY10/19, HaWRKY20/76, HaWRKY25/26, HaWRKY37/90, HaWRKY46/80, HaWRKY71/72, have similar protein lengths and same motif distributions, indicating the conserved motif structures of HaWRKY proteins within same groups.

## Evolution of group III HaWRKY genes

The origin of group III WRKY genes appeared to have occurred prior to the divergence of monocots and dicots, while numerous duplications and diversifications happened after that event. In order to understand the evolution of common sunflower group III WRKY genes, a phylogenetic tree of group III WRKY proteins from two monocots (rice and maize) and three dicots (sunflower, Arabidopsis and grape) was constructed. All the group III WRKY family members were divided into 10 clades as shown in Fig. 4. WRKY proteins from closer species were clustered into same clades. Most proteins from dicots gathered in clade 1 and 3, whereas monocots in clade 2, 4, 5, 6, 8, 9 and 10. Clade 7 contained proteins from all 5 species, indicating these proteins might be orthologues from a single ancestral gene.

MEME analysis was also conducted to search the conserved motifs of group III WRKY proteins from five species. Proteins within same clades usually displayed similar motif structures, indicating potential functional similarities among WRKY proteins. Motifs 1 and 7 were WRKY domains. Interestingly, motif 1 was found in all clades, whereas motif 7 was unique to clades 9 and 10, two clades only containing rice proteins, implying that motif 1 might have common function among different species, while motif 7 might play specific roles in rice and contribute to the divergence of group III WRKY genes. Motifs 1, 10 and 18 were specific to dicots. In contrast, motifs 12 and 19 were only observed in monocots. These motifs might be also important to the divergence of WRKY genes.

## Chromosomal location and synteny of HaWRKY genes

HaWRKY genes are distributed unevenly on 17 chromosomes (Supplementary Table 2, Fig. 5A). Chromosomes Ha10 and Ha15 both have 13 HaWRKY genes as the largest groups, whereas there was no HaWRKY gene observed on chromosomes Ha2. No correlation between chromosome length and HaWRKY gene number could be determined.

Two or more genes located within 200 kb on same chromosome is defined as a tandem duplication event [19]. Five pairs of HaWRKY genes (HaWRKY8/9, HaWRKY53/54, HaWRKY65/66, HaWRKY66/67 and HaWRKY71/72) are tandem duplicated on sunflower chromosomes Ha3, Ha12, Ha14 and Ha15. In addition, four segmental duplication events with four HaWRKY gene pairs (HaWRKY15/82, HaWRKY25/65, HaWRKY28/55 and HaWRKY50/53) are also identified (Fig. 4A). These results indicated that tandem and segmental duplication possibly contributes to the diversity and expansion of HaWRKY gene families. The dN/dS ratio of these duplicated gene pairs were calculated to understand the evolutionary constraints. The synonymous substitution rates (dS) of all segmental and tandem duplicated HaWRKY gene pairs were higher than non-synonymous substitution rate (dN) as shown in Table 1, indicating that HaWRKY gene family probably went through strong purifying selection during evolution.

Table 1  
dN/dS analyses for the duplicated WRKY gene pairs of sunflower.

Duplicated gene 1	Duplicated gene 2	dN	dS	dN/dS	Purifying Selection	Duplicate type
HaWRKY8	HaWRKY9	0.3047	0.8041	0.3790	Yes	Tandem
HaWRKY53	HaWRKY54	1.0653	15.1286	0.0704	Yes	Tandem
HaWRKY65	HaWRKY66	0.8347	55.0897	0.0152	Yes	Tandem
HaWRKY66	HaWRKY67	0.1132	0.265	0.4273	Yes	Tandem
HaWRKY71	HaWRKY72	1.0141	1.2005	0.8447	Yes	Tandem
HaWRKY15	HaWRKY82	0.1847	1.2239	0.1509	Yes	Segmental
HaWRKY25	HaWRKY65	0.2458	0.715	0.3437	Yes	Segmental
HaWRKY28	HaWRKY55	0.7764	59.8252	0.0130	Yes	Segmental
HaWRKY50	HaWRKY53	0.1218	0.6674	0.1825	Yes	Segmental

Dual syntenies of common sunflower with Arabidopsis and rice were also conducted. A total of eight HaWRKY genes showed syntenic relationship with those in Arabidopsis, composing 9 orthologous pairs, whereas only one HaWRKY gene was collinear with one in rice (Fig. 5B and C). Similarly, in whole genome, more collinear gene pairs were observed between sunflower and Arabidopsis than rice, as sunflower is phylogenetically closer to Arabidopsis. HaWRKY16 was associated with two Arabidopsis genes and HaWRKY25 and HaWRKY 65 are syntenic with a same Arabidopsis gene. HaWRKY25 is also found to be syntenic with a rice gene, indicating that these orthologous pairs might occur before the divergence of monocots and dicots.

# Transcriptomic pattern of HaWRKY genes from different tissues

The transcriptome data of HaWRKY genes of different sunflower tissues were downloaded from Sunflower Genome Database. 20 of the 90 identified HaWRKY genes didn't expressed in all tissues (Fig. 6A), which might be pseudogenes, have special temporal and spatial expression patterns or express in other tissues. The expression patterns of HaWRKY genes in sunflower were organ-specific, as bract, corolla, ligule, ovary, seed and stamen, which are related to flower, were clustered into a big group, and leaves and stem were in another group (Fig. 6A). Most of the HaWRKY genes didn't express in pollen (Fig. 6A). In general, the expression levels of HaWRKY genes in bract, ligule, leaves and stem were higher than that in other tissues (Fig. 6B). HaWRKY17/22/79/81 displayed highest transcript abundances across all tissues except pollen and were clustered into a group, whereas the expression levels of HaWRKY23/31/37/40/68/84 were extremely low in all tested tissues (Fig. 6A). The expression patterns of some genes were tissue-specific, for example, HaWRKY73 was abruptly induced only in leaves, HaWRKY3 in stem, HaWRKY11 in style, etc.

## Profiles of HaWRKY genes under abiotic and biotic stress

23 HaWRKY genes which were highly induced in different tissues of common sunflower (except in pollen) were selected to test the reactions of different WRKY genes to different abiotic stresses. Generally, HaWRKY genes were inhibited in sunflower leaves after treatment of PEG and NaCl with different concentrations, whereas HaWRKY29/30 at 150 Mm NaCl and HaWRKY48/89 at 20% PEG were significantly up-regulated and no any changes of HaWRKY55/57 were observed in response to abiotic stresses (Fig. 7A). Similarly, in roots as shown in Fig. 7B, expression levels of most HaWRKY genes decreased significantly in response to PEG, among them HaWRKY22/29/74 were significantly down-regulated only at high concentration of PEG. HaWRKY57/59/77/81/82/89 were significantly induced after treatment with PEG and HaWRKY30/35/38/48/55 were not sensitive to PEG in the roots. In contrast, Most of the HaWRKY genes were significantly up-regulated after treatment with NaCl in roots, as compared to the control. Among them, the transcript levels of 14 HaWRKY genes increased and 6 decreased, as the concentration of NaCl went up. HaWRKY13/52/79 were significantly depressed by treatment of NaCl in sunflower roots.

In order to understand the role of sunflower WRKY gene family against biotic stress, transcription levels of two contrasting common sunflower cultivars (TK0409, susceptible; JY207, resistant) under infection of root parasitic weed *Orobanche cumana* were tested. Most of the genes were inhibited in both cultivars under attack of *O. cumana*, as compared to their corresponding cultivars without attack, respectively (Fig. 8). Interestingly, HaWRKY 7/15/44/45/68/71/72/76/85 were induced after attack of *O. cumana* in resistant cultivar JY207, whereas still depressed in susceptible cultivar TK0409 (Fig. 8), suggesting these genes might partly contribute to the resistance of sunflower against *O. cumana*.

## Discussion

The WRKY transcription factor family is considered to be involved in diverse stress responses, developmental and physiological processes in plants. Systematical characterization of WRKY genes in several species has been studied, including Arabidopsis, rice, tomato, maize etc. In our study, 90 WRKY genes were identified from sunflower genome, assigned as HaWRKY1 to HaWRKY90 based on their locations on chromosomes.

Multiple protein sequence alignments revealed domain variations in common sunflower WRKY family. A WKKY group with 7 proteins was identified in our study, including HaWRKY 60 with WKKYGQK, HaWRKY33/50/53/54 with WKKYGEK, and HaWRKY51 with WKKYGKK (Fig. 2). Interestingly, this group has been also found in *Helianthus exilis*, *Helianthus petiolaris* and *Helianthus tuberosus* [18], indicating that the WKKY variation is common in the Asteraceae. In addition, WKKYGQK, WKKYGKK and WKKYGEK are also observed in different legumes, but with low frequencies [20]. However, there are no more reports about WKKY group in other plant species. Although the WRKYGQK is highly conserved in most WRKY domains, variation in the core sequence has been documented [20]. In our studies, mutations happened to R and Q sites, while the others were conserved. WRKYGKK and WRKYGEK are the most frequently occurring variants of the core sequence in most plant species [20]. As the WRKYGQK core sequence can interact with the W-box to activate downstream genes, the variations in this motif might influence the function of downstream target genes [13, 21]. Thus, further investigations on functions and binding specificities of these sunflower proteins with mutated WRKY motifs might provide deep insight into this transcription factor family.

A domain loss is common in the WRKY gene family in plants, which is recognized as a divergent force for expansion of this gene family [3, 22]. In the current study, 5 domain loss events were found in group I, suggesting a potential cause of the diversity of WRKY genes in this group. Tandem and segmental duplication events also played a pivotal role in the expansion of WRKY gene family [4]. Five pairs of tandem and four pairs of segmental duplicated genes were identified in the present study, with five pairs in group IIIa, two pairs in group IIId, one pair in group IIb and one pair in group IIc. This result indicated that tandem and segmental duplication events might contribute to the amplification of sunflower WRKY genes in these groups, as compared to those of Arabidopsis.

It has been reported that WRKY genes from group III play an important role in plant adaptation and evolution [23]. In order to explore how the WRKY group III gene family evolved, a phylogenetic tree of WRKY group III proteins from sunflower with two dicots (*Arabidopsis*, grape) and two monocots (rice, maize) was constructed, which divided the 17 group III HaWRKYs into three clades. WRKY proteins from closer species appeared to be clustered together. Both monocots and dicots proteins occurred in many clades, suggesting group III WRKY genes diversified before the monocot-eudicot split. In addition, clade 7 contained group III WRKY proteins from all 5 species, which tended to form monocot- and dicot-specific subclades, implying that group III WRKY genes evolved separately after the divergence of monocots and dicots.

It is well known that WRKY genes play essential roles in plant growth and development [7]. Li et al. [24] reported that AtWRKY13 functioned in stem development, as a weaker stem phenotype was observed and lignin-synthesis-related genes were repressed in Arabidopsis wrky13 mutants. In the current study, the orthologous of AtWRKY13 in sunflower, HaWRKY28 displayed high expression levels in stem, indicating that these genes might also act in stem development in sunflower. Overexpression of WRKY15 exhibited an increased leaf area of Arabidopsis, which implied that AtWRKY15 seemed to be involved in leaf growth [25]. HaWRKY79, which is the orthologous of AtWRKY15, were highly induced not just in leaves, but across all tissues, indicating that this gene might be constitutive in sunflower plant growth and development. In contrast, HaWRKY7 was specifically expressed in leaves, suggesting its role in sunflower leaf growth. Among all WRKY genes in sunflower, only HaWRKY30 displayed a high level of expression in pollen, with other WRKY genes extremely low expressed. Interestingly, two pollen-specific regulators in Arabidopsis, AtWRKY34 and AtWRKY2, have phylogenetically close relationship with this sunflower WRKY gene, indicating that HaWRKY30 might be associated with pollen developmental modulation.

In addition to their role in plant growth and development, WRKY TFs also play pivotal roles in various stress responses, providing an important basis for genetic improvement of crops. Drought and salinity, both of which can cause plant cellular dehydration [26], are two major constraints to sunflower production [27]. Responses of plants to drought and salinity usually result in accumulation of reactive oxygen species (ROS) and abscisic acid (ABA) [26, 28], which activate downstream WRKY genes [7].

Overexpression of a membrane-localized cysteine-rich receptor-like protein kinase, CRK5 in Arabidopsis, led to increase of ABA sensitivity and promotion of stomatal closure, and subsequent enhancement of plant drought tolerance. Knockout of AtWRKY18, AtWRKY40 and AtWRKY60 significantly increased the expression of CRK5, suggesting negative regulation of these three genes on CRK5 [29]. In our study, the relative expression levels of two orthologous of AtWRKY40, HaWRKY74 and HaWRKY81, were recorded in sunflower roots and leaves. Expression levels of both two genes decreased as the concentrations of PEG increased in sunflower leaves, while in sunflower roots, two genes were induced under low concentration of PEG and inhibited under high concentration. It has also been reported that AtWRKY46, AtWRKY54, and AtWRKY70 are implicated in promotion of BR-regulated plant growth and inhibition of drought response, as reduced BR-regulated growth and higher survival rates under drought stress was observed in their triple mutant [30]. HaWRKY9 and HaWRKY22, which were phylogenetically close to AtWRKY46, were both repressed under PEG treatments in sunflower roots and leaves. These results are suggesting that sunflower probably enhanced drought tolerance via down-regulating specific WRKY genes and subsequently activating downstream signal pathways. The increase of ABA level caused by drought usually induces high expression of AtWRKY57, which binds to W-box in the promoter region of the downstream response genes. HaWRKY57, the orthologous of AtWRKY57, displayed a high expression level under PEG treatment in sunflower roots. Interestingly, the increase of ROS level caused by salinity also activates AtWRKY57, and consistently, HaWRKY57 was highly expressed under treatment of 300 mM NaCl in sunflower roots. These results indicated that HaWRKY57 might share similar functions with AtWRKY57 in sunflower under drought and salinity. AtWRKY15 is another WRKY gene induced by ROS, but will make Arabidopsis more susceptible to osmotic stress and oxidative stress [25]. In our study,

HaWRKY79, the orthologous of AtWRKY15, was significantly suppressed in sunflower roots under treatment of NaCl, implying their similar roles in conferring salt tolerance.

The parasitic weed *Orobanche cumana* is a new emerged threat to sunflower production worldwide. Previous studies proposed that *O. cumana* deployed effectors in sunflower to suppress host defense responses and resistant sunflower cultivars recognized effectors with the help of R proteins to activate effector-triggered immunity. WRKY family has been found to be involved in the microbe-associated molecular pattern-triggered immunity, PAMP-triggered immunity or effector-triggered immunity [7]. Thus, it is worthy of studying WRKY genes in sunflower against *O. cumana*. According to our transcriptome data, HaWRKY7/15/44/45/68/71/72/76/85 were specifically induced in sunflower resistant cultivar under attack of *O. cumana*, whereas repressed in susceptible cultivar during same interactions, indicating their potential roles in conferring resistance to sunflower against *O. cumana*.

## Conclusion

We identified 90 WRKY genes from *Helianthus annuus* L. and they were classified into four main groups including a species-specific WKKY group. HaWRKY genes within same group or subgroup generally showed similar exon-intron structures and motif compositions. Synteny analyses of sunflower WRKY genes provided deep insight to the evolution of HaWRKY genes. The expressions of HaWRKY genes suggested that these genes could be involved in the regulation of sunflower growth and development, as well as various abiotic and biotic stresses. This study provides a foundation for further functional analyses of HaWRKY genes and will be beneficial to crop improvement.

## Methods

### Gene identification

The genome of *H. annuus* was downloaded from Sunflower Genome Database (<https://www.sunflowergenome.org/>) [31]. The protein sequences of the WRKY family of *A. thaliana* were obtained from Plant Transcription Factor Database (<http://planttfdb.cbi.pku.edu.cn/index.php>) [32], which were used to search the WRKY genes from *H. annuus* genome via BlastP and tBlastN (E-value  $\leq 1e-20$ ). Then Pfam database (<http://pfam.xfam.org/>) and SMART database (<http://smart.embl-heidelberg.de/>) were used for verification of the WRKY domains [33–34]. These potential sequences were further queried in the NCBI Conserved Domains Database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) and InterProScan Database (<http://www.ebi.ac.uk/interpro/search/sequence-search>) to validate the conserved domain [35–36]. The molecular weight (Mw) and isoelectric point (pI) of the full-length proteins were predicted using the pI/Mw tool ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)) in ExPASy.

### Phylogenetic analysis and gene structure

Multiple sequence alignment based on WRKY domain sequences were conducted by clustal W analysis with default parameters. A neighbor-joining (NJ) tree was constructed in MEGA 5.2 with the following criteria: Poisson model, pairwise deletion, and 1000 bootstrap replications. Further maximum likelihood (ML) analysis of WRKY gene family from sunflower and Arabidopsis was conducted, to confirm the reliability of the result. The intron-exon structures of sunflower WRKY genes were analyzed by comparing predicted coding sequences with their corresponding full-length sequences using the online tool Gene Structure Display Server (GSDS, <http://gsds.cbi.pku.edu.cn/>) [37]. The MEME online program (Multiple Expectation Maximization for Motif Elicitation) version 4.11.1 (<http://meme-suite.org/index.html>) was used to identify conserved motifs in the sunflower WRKY proteins [38].

## Chromosomal distribution and gene duplication

Multiple Collinearity Scan toolkit (MCScanX) was adopted to analyze the gene duplication events with default parameters [39]. PAL2NAL v14 was subsequently used to calculate dN and dS, with a dN/dS ratio of 1 indicative of neutral selection [40].

## Plant materials, growth conditions and treatments

Seeds of both *O. cumana* and two sunflower cultivars JY207 and TK0409 were provided by the Institute of Plant Protection, Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot, China. For abiotic stresses, common sunflower cultivar TK0409 was used. The seeds were germinated and grown in peat moss. At the five-leaf stage, seedlings with uniform size were selected for hydroponics in a half-strength Hoagland nutrient solution (pH = 6.0). All the plants were placed in a growth chamber with light intensity ranging from 250 to 350  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , temperature at 16–20 °C, and relative humidity at approximately 55–60%. The nutrient solution was renewed every five days. After one week of acclimatization, the solutions were supplemented with NaCl (0, 150, and 300 mM) and polyethylene glycol-6000 (PEG-6000, 0, 10% and 20% w/v), respectively. After another one week, roots and leaves of the sunflower seedlings were sampled for RNA isolation. For biotic stress, common sunflower cultivars TK0409 (susceptible) and JY207 (resistant), and root parasitic weed *Orobancha cumana* were applied. 200 mg of *O. cumana* seeds were homogeneously mixed with 0.5 kg of the peat and vermiculite (1:1, v/v) substrate. Sunflowers were grown in the substrate containing *O. cumana* seeds as mentioned above. All the plants were placed in a growth chamber with 20 °C at daytime and 14 °C at night, photoperiod for 14 hr, and an irradiance of 300  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Three weeks after inoculation, sunflower roots were collected for experiments. Each treatment was replicated three times.

## RNA extraction and gene expression analysis

TaKaRa MiniBEST Plant RNA Extraction Kit (Takara Bio, Kyoto, Japan) was used to extract total RNA from sunflower leaves and roots. 200 ng of total RNA was reverse transcribed by using TaKaRa PrimeScript™ RT reagent Kit with gDNA Eraser. The gene-specific primers of sunflower for qRT-PCR amplification were provided in Supplementary Table 1. SYBR Premix Ex Taq II (Tli RNaseH Plus, TaKaRa) in CFX96™ Real-Time PCR detection System (Bio-Rad, Hercules, CA, USA) was used to conduct the qRT-PCR experiments. The PCR conditions consisted of pre-denaturation at 95 °C for 30 s, 40 cycles of

denaturation at 95 °C for 5 s, and annealing and extension at 58 °C for 30 s. Each treatment had three biological replicates.

## Abbreviations

HaWRKY: *Helianthus annuus* WRKY; AtWRKY: *Arabidopsis thaliana* WRKY; PEG: Polyethylene Glycol; ROS: reactive oxygen species.

## Declarations

### Consent for publication

Not applicable.

### Availability of data and material

All data analyzed during this study are included in this article and its Supplementary files.

### Ethics approval and consent to participate

*Helianthus annuus* cv. TK0409 and JY207 are widely cultivated and *Orobancha cumana* is a common parasitic weed in Inner Mongolia, China. Both *Helianthus annuus* and *Orobancha cumana* are not listed in the appendices I, II and III of the Convention on the Trade in Endangered Species of Wild Fauna and Flora. Seeds of both *O. cumana* and two sunflower cultivars JY207 and TK0409 were collected in Bayannaer, Inner Mongolia and provided by the Institute of Plant Protection, Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot, China. Collection of plant materials complied with the institutional, national and international guidelines. No specific permits were required.

### Competing interests

The authors declare that they have no competing interests

### Funding

This work was supported by the National Natural Science Foundation of China (31701333, 31570434, 31871694), GDAS' Project of Science and Technology Development (2020GDASYL-20200103062), Zhejiang Provincial Natural Science Foundation (LGN18C130007), the Jiangsu Collaborative Innovation

Center for Modern Crop Production, China Scholarship Council (201808330058), and Inner Mongolia Science & Technology Plan (IMSTP). The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Author contributions

WJZ, LX and CY conceived and designed the study. CY, JJL, JSW, LYH and JW performed the experiments. CY wrote the paper. CY, FI, LX and WJZ reviewed and edited the manuscript. All authors read and approved the manuscript.

## Acknowledgements

The authors thank to the Institute of Plant Protection, Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot, China for providing plant materials.

## References

- [1] Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 2014; 42: 1182-7.
- [2] Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 2000; 5: 199-206.
- [3] Brand LH, Fischer NM, Harter K, Kohlbacher O, Wanke D. Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and in vitro binding assays. *Nucleic Acids Res.* 2013; 41: 9764-78.
- [4] Zhang Y, Wang L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol Biol.* 2005; 5: 1.
- [5] Bakshi M, Oelmuller R. WRKY transcription factors: Jack of many trades in plants. *Plant Signaling Behav.* 2014; 9: e27700.
- [6] Ishiguro S., Nakamura K. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and  $\beta$ -amylase from sweet potato. *Mol General Genet.* 1994; 244: 563-71.
- [7] Chen F, Hu Y, Vannozzi A, Wu K, Cai H, Qin Y, Mullis A, Lin Z, Zhang L. The WRKY Transcription Factor Family in Model Plants and Crops. *Crit Rev Plant Sci.* 2018; 36: 311-35.

- [8] Jiang W, Yu D. Arabidopsis WRKY2 transcription factor mediates seed germination and postgermination arrest of development by abscisic acid. *BMC Plant Biol.* 2009; 9: 96.
- [9] Zhang J, Peng Y, Guo Z. Constitutive expression of pathogen-inducible OsWRKY31 enhances disease resistance and affects root growth and auxin response in transgenic rice plants. *Cell Res.* 2008; 18: 508-21.
- [10] Li W, Wang H, Yu D. Arabidopsis WRKY Transcription Factors WRKY12 and WRKY13 Oppositely Regulate Flowering under Short-Day Conditions. *Mol Plant.* 2016; 9: 1492-503.
- [11] Yu Y, Liu Z, Wang L, Kim SG, Seo PJ, Qiao M, Wang N, Li S, Cao X, Park C, Xiang F. WRKY71 accelerates flowering via the direct activation of FLOWERING LOCUS T and LEAFY in *Arabidopsis thaliana*. *Plant J.* 2016; 85: 96-106.
- [12] Qiu Y, Yu D. Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in *Arabidopsis*. *Environ Exp Bot.* 2009; 65: 35-47.
- [13] Zhou QY, Tian AG, Zou H, Xie ZM, Lei G, Huang J, Wang C, Wang H, Zhang J, Che S. Soybean WRKY-type transcription factor genes, *GmWRKY13*, *GmWRKY21*, and *GmWRKY54*, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants. *Plant Biotechnol J.* 2008; 6: 486-503.
- [14] Mzid R, Marchive C, Blancard D, Deluc L, Barrieu F, Corio-Costet MF, Drira N, Hamdi S, Lauvergeat V. Overexpression of VvWRKY2 in tobacco enhances broad resistance to necrotrophic fungal pathogens. *Physiol Plantarum.* 2007; 131: 434-47.
- [15] Menke FLH, Kang HG, Chen Z, Park JM, Kumar D, Klessig DF. Tobacco Transcription Factor WRKY1 Is Phosphorylated by the MAP Kinase SIPK and Mediates HR-Like Cell Death in Tobacco. *Mol Plant Microbe.* 2005; 18: 1027-34.
- [16] Liu X, Baird WV. Differential Expression of Genes Regulated in Response to Drought or Salinity Stress in Sunflower. *Crop Sci.* 2003; 43: 678-87.
- [17] Yang C, Xu L, Zhang N, Islam F, Song W, Hu L, D Liu, X Xie, Zhou W. iTRAQ-based proteomics of sunflower cultivars differing in resistance to parasitic weed *Orobancha cumana*. *Proteomics.* 2017; 17: 1700009.
- [18] Giacomelli JI, Ribichich KF, Dezar CA, Chan RL. Expression analyses indicate the involvement of sunflower WRKY transcription factors in stress responses, and phylogenetic reconstructions reveal the existence of a novel clade in the Asteraceae. *Plant Sci.* 2010; 178: 398-410.
- [19] Holub EB. The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat Rev Genet.* 2001; 2: 516-27.
- [20] Song H, Sun W, Yang G, Sun J. WRKY transcription factors in legumes. *BMC Plant Biol.* 2018; 18: 243.

- [21] van VMC, Pappaioannou D, Neeleman L, Bol JF, Linthorst HJ. A Novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. *Plant Physiol.* 2008; 146: 1983-95.
- [22] Wei KF, Chen J, Chen YF, Wu LJ, Xie DX. Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res.* 2012; 19: 153-64.
- [23] Wang Y, Feng L, Zhu Y, Li Y, Yan H, Xiang Y. Comparative genomic analysis of the WRKY III gene family in populus, grape, arabidopsis and rice. *Biol Direct.* 2015; 10: 48.
- [24] Li W, Tian Z, Yu D. WRKY13 acts in stem development in *Arabidopsis thaliana*. *Plant Sci.* 2015; 236: 205-213.
- [25] Vanderauwera S, Vandenbroucke K, Inze A, van de Cotte B, Muhlenbock P, De Rycke R, Naouar N, Van Gaeve T, Van Montagu MC, Van Breusegem F. AtWRKY15 perturbation abolishes the mitochondrial stress response that steers osmotic stress tolerance in *Arabidopsis*. *P Natl Acad Sci USA.* 2012; 109: 20113-8.
- [26] Bartels D, Sunkar R. Drought and Salt Tolerance in Plants. *Crit Rev Plant Sci.* 2005; 24: 23-58.
- [27] Kaya MD, Okçu G, Atak M, Çikılı Y, Kolsarıcı Ö. Seed treatments to overcome salt and drought stress during germination in sunflower :*Helianthus annuus* L., *Eur J Agron.* 2006; 24: 291-5.
- [28] Miller G, Suzuki N, Ciftci-Yilmaz S, Mittler R. Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant cell Environ.* 2010; 33: 453-67.
- [29] Lu K, Liang S, Wu Z, Bi C, Yu YT, Wang XF, Zhang D. Overexpression of an *Arabidopsis* cysteine-rich receptor-like protein kinase, CRK5, enhances abscisic acid sensitivity and confers drought tolerance. *J Exp Bot.* 2016; 67: 5009-27.
- [30] Chen J, Nolan TM, Ye H, Zhang M, Tong H, Xin P, Chu J, Chu C, Li Z, Yin Y. *Arabidopsis* WRKY46, WRKY54, and WRKY70 Transcription Factors Are Involved in Brassinosteroid-Regulated Plant Growth and Drought Responses. *Plant Cell.* 2017; 29: 1425-39.
- [31] Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, Legrand L, Gill N, Kane NC, Bowers JE, Hubner S, Bellec A, Bérard A, Bergès H, Blanchet N, Boniface M, Brunel D, Catrice O, Chaidir N, Claudel C, Donnadiou C, Faraut T, Fievet G, Helmstetter N, King M, Knapp SJ, Lai Z, Le Paslier M, Lippi Y, Lorenzon L, Mandel JR, Marage G, Marchand G, Marquand E, Bret-Mestries E, Morien E, Nambeesan S, Nguyen T, Pegot-Espagnet P, Pouilly N, Raftis F, Sallet E, Schiex T, Thomas J, Vandecasteele C, Varès D, Vear F, Vautrin S, Crespi M, Mangin B, Burke JM, Salse J, Muñoz S, Vincourt P, Rieseberg LH, Langlade NB. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature.* 2017; 546: 148-52.
- [32] Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 2017; 45: 1040-5.

- [33] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR. Pfam: the protein families database. *Nucleic Acids Res.* 2014; 42: 222-30.
- [34] Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.* 2012; 40: 302-5.
- [35] Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 2017; 45: 200-3.
- [36] Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014; 30: 1236-40.
- [37] Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015; 31: 1296-97.
- [38] Bailey T, Boden M, Buske FA, Frith M, Grant CE, Clementi L. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009; 37: 202-8.
- [39] Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012; 40: e49.
- [40] Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments, *Nucleic Acids Res.* 2006; 34: 609-12.

## Figures

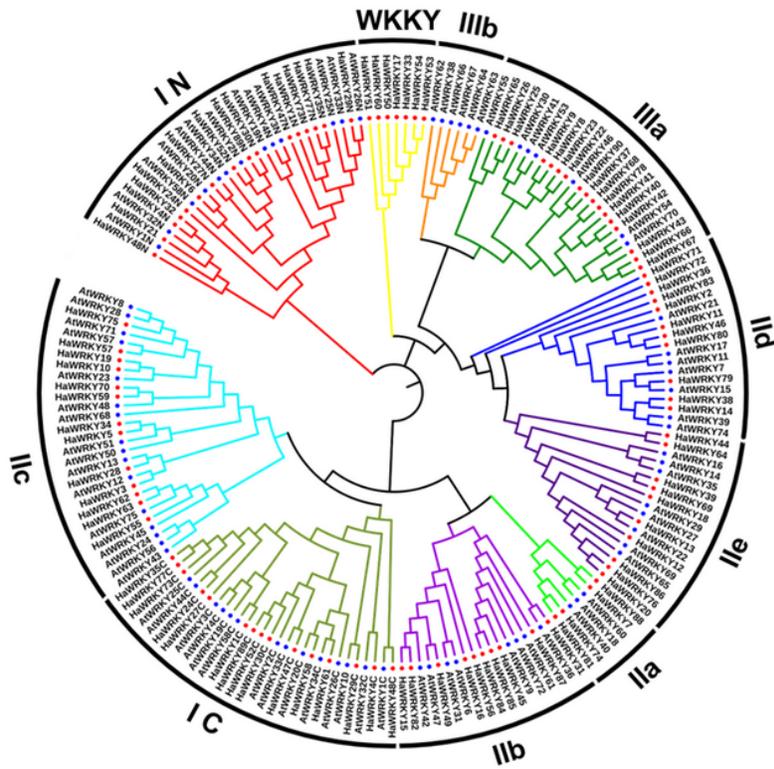


Figure 1

## Figure 1

Phylogenetic relationships of WRKY genes from common sunflower and Arabidopsis. The different-colored branches indicate different groups (or subgroups). The red solid circles and blue solid squares represent WRKY genes from Arabidopsis and common sunflower, respectively.

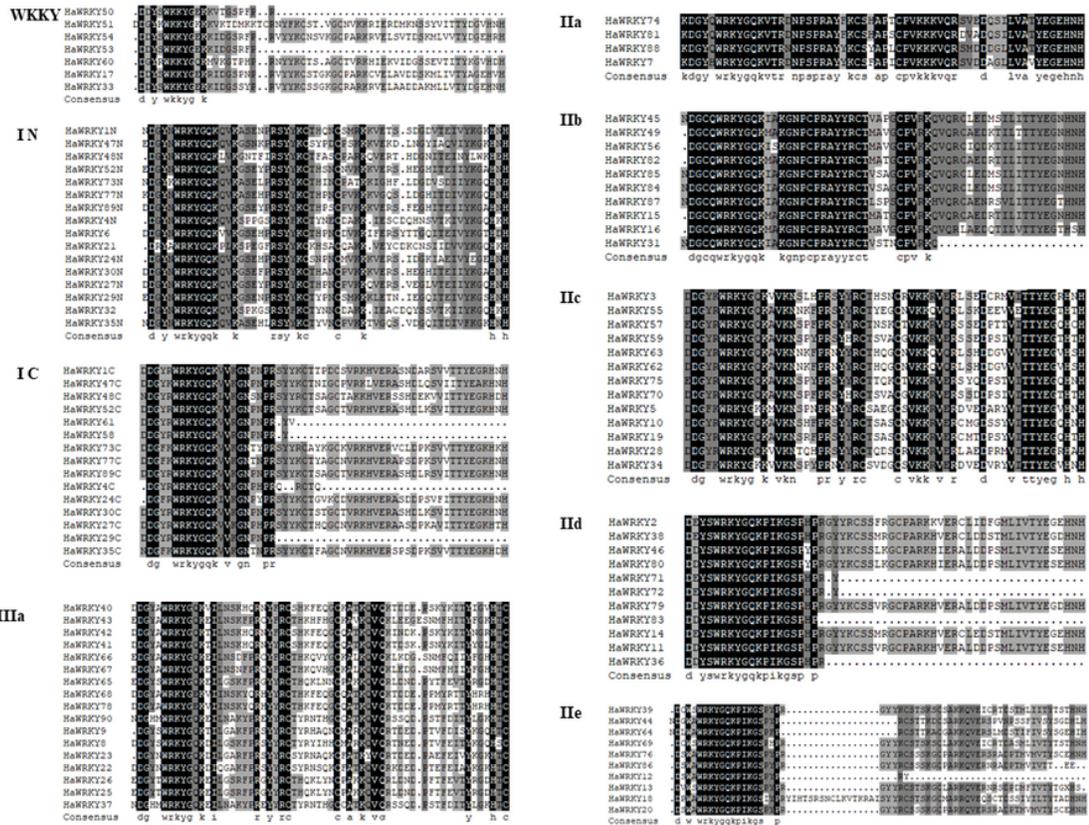


Figure 2

Figure 2

Alignment of multiple HaWRKY protein sequences. "N" and "C" indicate the N- and C-terminal of WRKY domains.

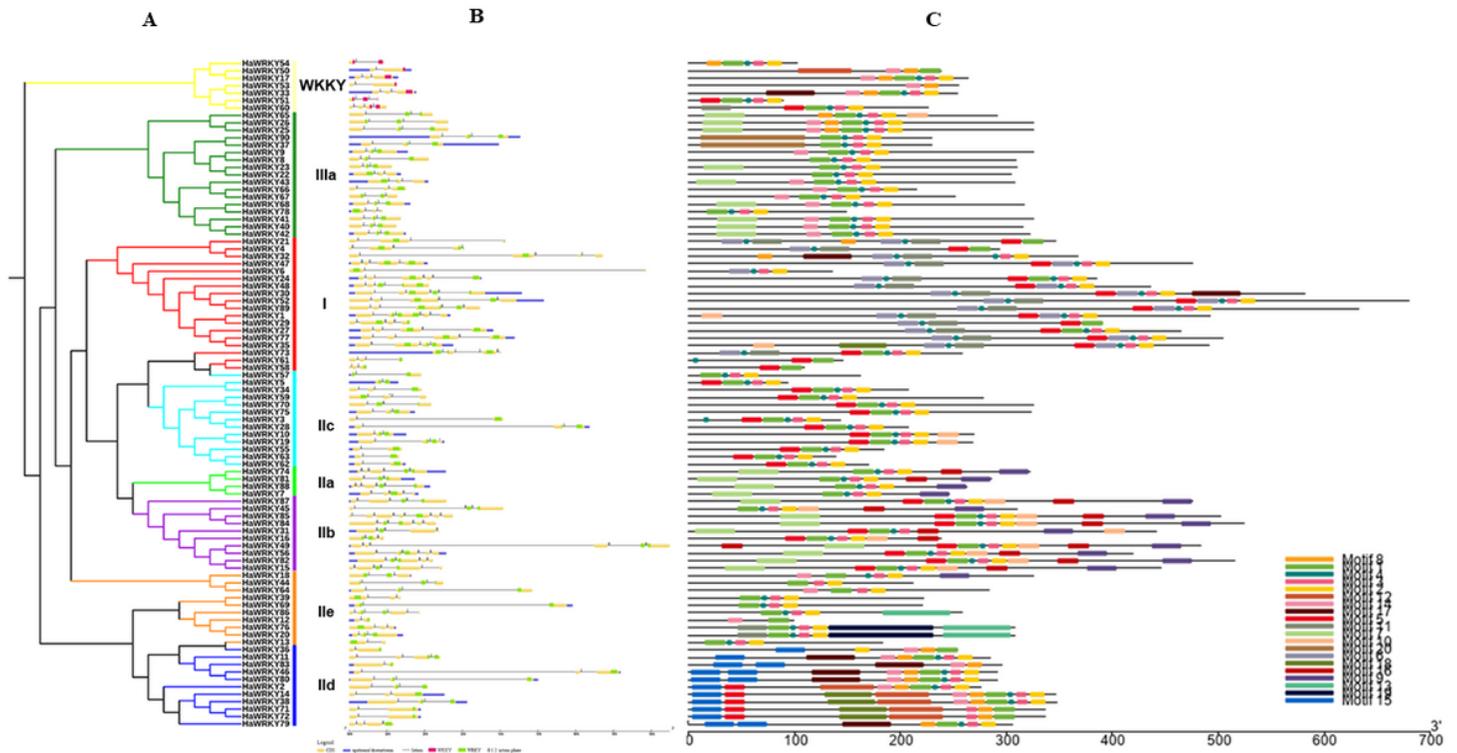
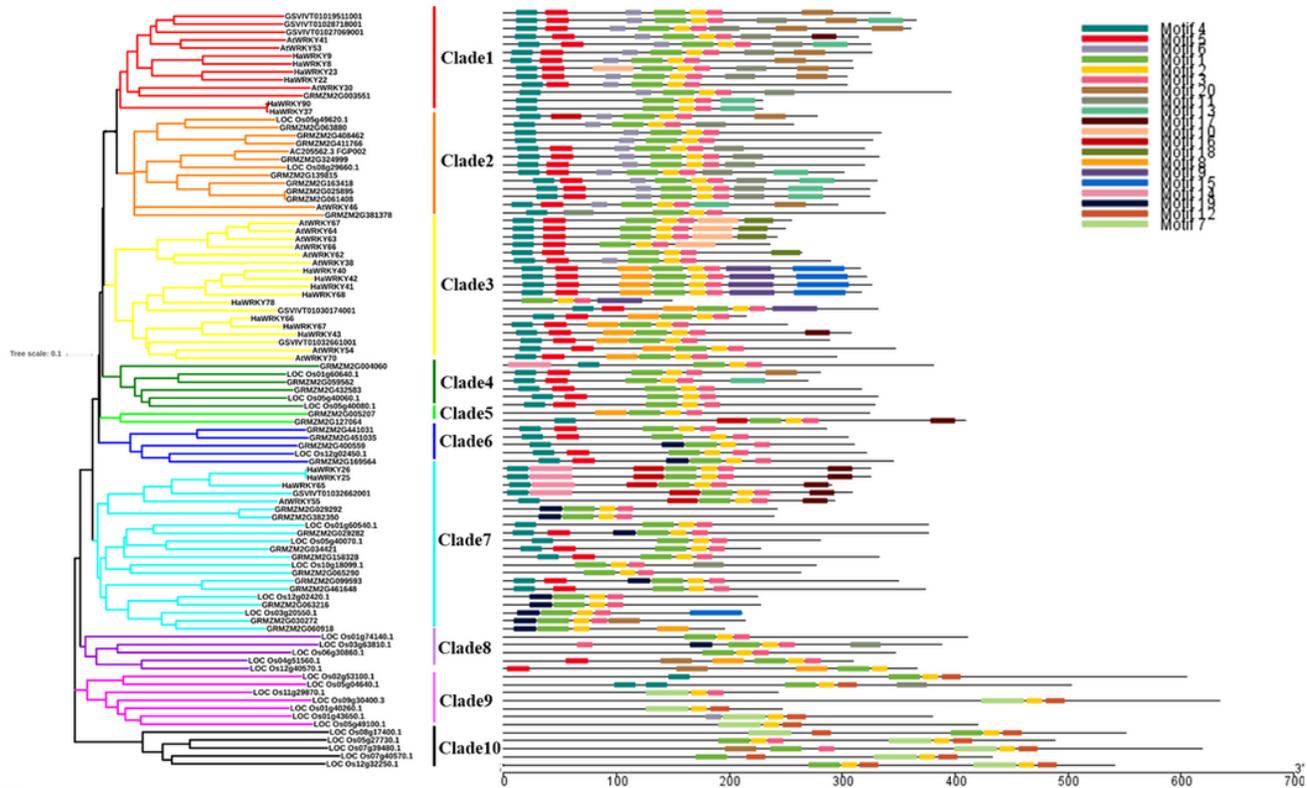


Figure 3

**Figure 3**

Phylogenetic relationships, gene structures and motif compositions of WRKY genes in common sunflower. (A) Phylogenetic tree of WRKY genes. The different-colored branches indicate different groups (or subgroups). (B) Exon-intron structures of WRKY genes. Blue boxes indicate 5' and 3' UTRs. Yellow boxes indicate exons. Black lines indicate introns. Red boxes indicate WRKY domains. The numbers indicate the phases of introns. (C) Motif compositions of WRKY proteins. Different motifs are displayed with different colored boxes.



**Figure 4**

**Figure 4**

Phylogenetic relationships and motif compositions of group III WRKY proteins from five monocot and dicot plants. On the left side, proteins are clustered into 10 clades, marked with different colors. On the right side, Different motifs are displayed with different colored boxes.

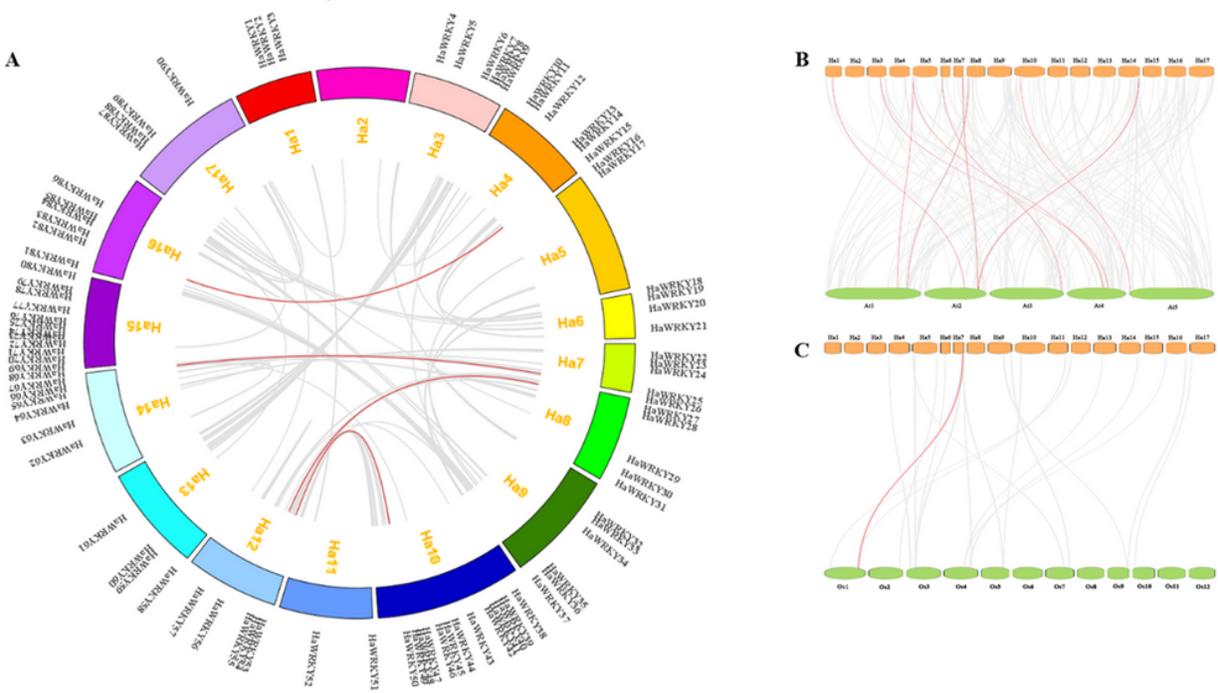


Figure 5

## Figure 5

Genome localization and synteny analyses of WRKY genes within common sunflower, and between common sunflower and two representative plant species. (A) Chromosomal distribution and interchromosomal relationships of common sunflower WRKY genes. Gray lines indicate all syntenic gene pairs in common sunflower genome and red lines indicate duplicated WRKY gene pairs. (B-C) Synteny analyses of common sunflower WRKY genes with Arabidopsis and rice, respectively. Gray lines indicate collinear gene pairs between common sunflower and other plant genomes and red lines indicate syntenic WRKY gene pairs.

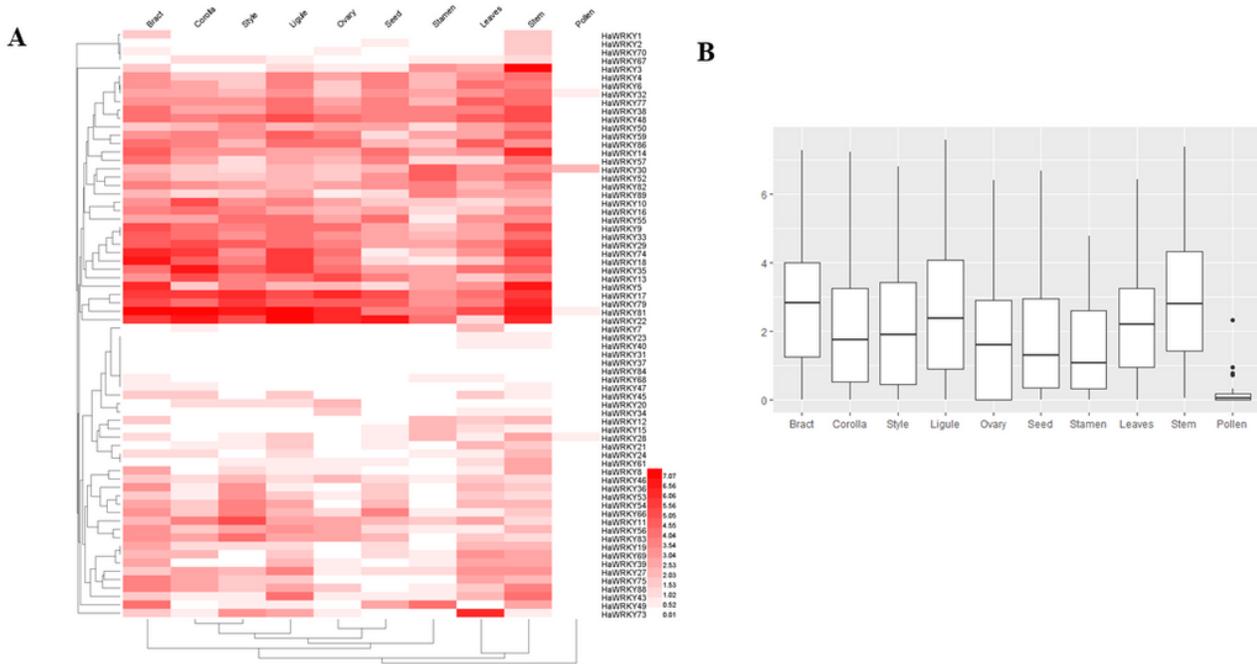
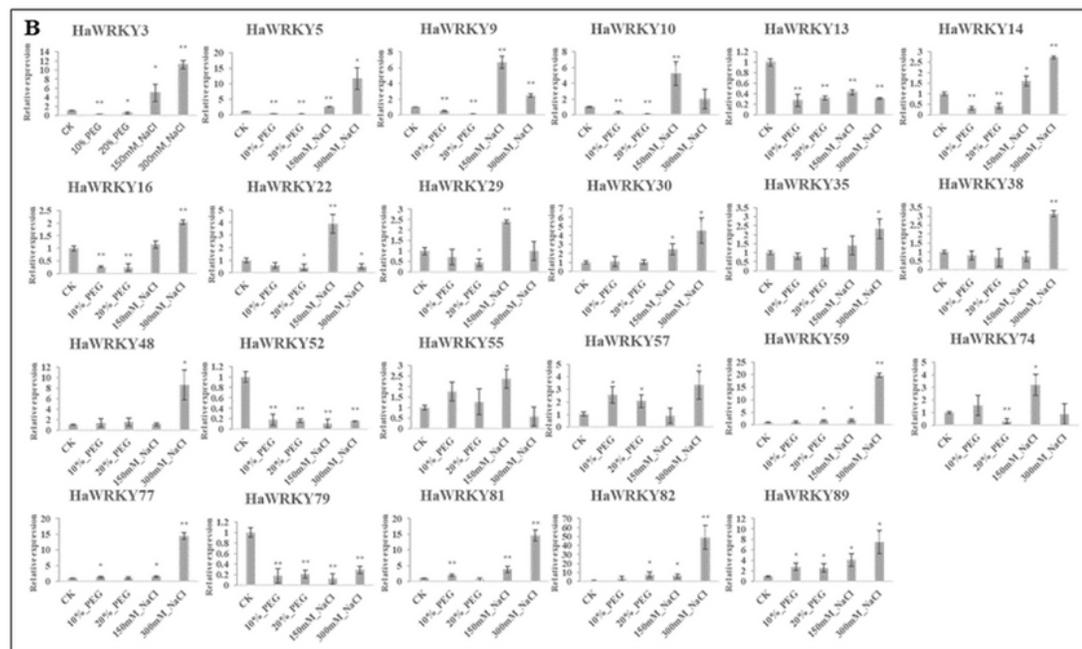
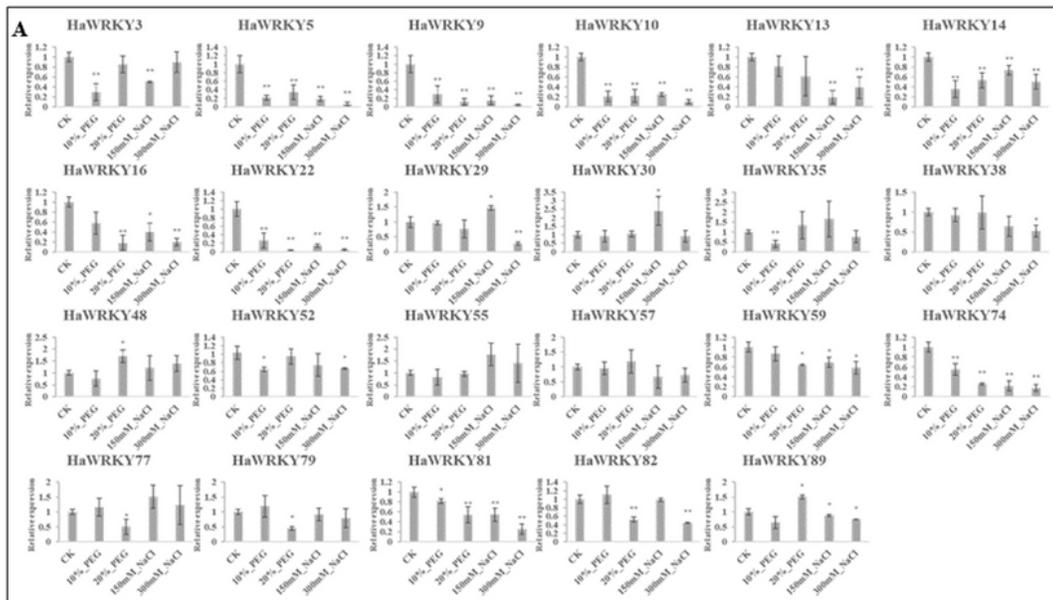


Figure 6

## Figure 6

Expression profile of WRKY genes in different tissues of common sunflower. (A) Hierarchical clustering of expression profile of WRKY genes from different tissues. Data were transformed with a  $\log_2(\text{FPKM}+1)$  transformation. (B) Boxplot of expression levels of WRKY genes in different tissues.



**Figure 7**

Expression profile of 23 selected HaWRKY genes in responses to treatments of PEG and NaCl. (A) Expression profile of WRKY genes in common sunflower leaves. (B) Expression profile of WRKY genes in common sunflower roots.

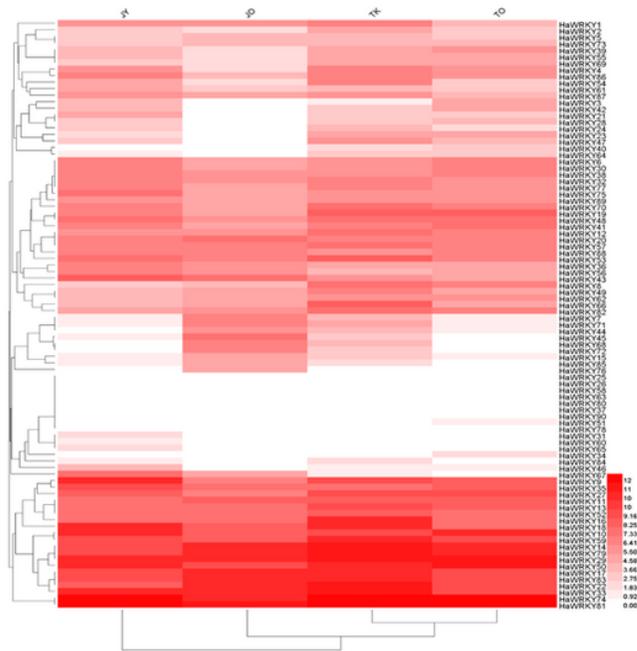


Figure 8

## Figure 8

Expression profile of HaWRKY genes in response to infection of parasitic weed *Orobancha cunama*. JY represents resistant sunflower cultivar JY207; JO represents JY207 with infection of *O. cunama*; TK represents susceptible sunflower cultivar TK0409; TO represents TK0409 with infection of *O. cunama*.

## Supplementary Files

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

- [SupplementaryTables.docx](#)