Genome-Wide Identification and Characterization of WRKY Gene Family in Camelina Sativa

Yanan Song  
Shanxi Agricultural University

Hongli Cui  
Shanxi Agricultural University

Ying Shi  
Shanxi Agricultural University

Jinai Xue  
Shanxi Agricultural University

Chunli Ji  
Shanxi Agricultural University

Chunhui Zhang  
Shanxi Agricultural University

Runzhi Li (rli2001@126.com)

Research article

Keywords: Camelina sativa, Genome-wide, WRKYs, Phylogenetic analysis, Expression profile, Salt stress

DOI: https://doi.org/10.21203/rs.3.rs-36113/v1

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

**Background:** WRKY gene family is one of the largest transcription factor families and WRKY proteins (WRKYs) have the complex biological functions to regulate plant metabolic processes. Although the WRKY genes were identified in many species and the functions were verified, there were no reports of *Camelina sativa* WRKY genes.

**Results:** In this investigation, a total of 202 CsWRKY genes were identified and encoded 242 CsWRKYs. The CsWRKYs were further classified into three major groups according to their structure and phylogeny. The comprehensive analysis showed the characteristic sequences of CsWRKYs were conserved in the evolutionary process. In addition, the 137 segmental duplication events were the major force to expand the CsWRKY members in evolution. Compared with other reported plant species, CsWRKYs family as the largest WRKY gene family had maximum members. Furthermore, expression profiling indicated that different CsWRKY members exhibited differently in shoots and roots, and some CsWRKY genes were also up-regulated to varying degrees under salt stress in shoots.

**Conclusions:** In this research, a detailed overview of CsWRKY family genes and expression patterns offered precious information for understanding the potential evolutionary process and the biological functions of CsWRKY genes, which was useful for the further characteristic research of CsWRKY genes and the development of high-quality *Camelina sativa* varieties.

**Background**

*Camelina sativa* (*C. sativa*), a dicotyledonous plant, belongs to the *Cruciferae* family. As a "low-input and environment-friendly" alternative industrial oil crop, *C. sativa* has been planted in large areas [1] and has begun trial planting and mass production in China. Compared with other oil-bearing crops, *C. sativa* has several prominent traits, such as some degree of salt tolerance, high drought resistance, low temperature resistance, short life cycle (80-100 days), highly resistant to common pests and diseases of many cruciferous crops [2-5]. There are high-level oil (36%–47%) and protein (30%) in its seed, the oil has the unsaturated fatty acids (90%), omega-3 fatty acids (40%), essential fatty acids and a variety of natural active ingredients, which can improve human immunity and prevent various diseases such as cancer [6, 7]. Thus, *C. sativa* may provide sustainable raw material for food, aviation fuel, biofuel production and other high-value industrial production [8]. Studies have shown that the expression of microRNA167A in *C. sativa* seeds reduced α-linolenic acid content in the leaves and increased the seed size [9]. The expression of diacylglycerol acyltransferase 1 and glycerol-3-phosphate dehydrogenase limited seed and oil yields of transgenic *C. sativa* [10]. However, there are few reports on the molecular mechanism of high tolerance of *C. sativa*.

During the plant growth and development, the plant is stressed by various stresses including salt and drought [4], resulting in the decline in the quality and yield of crop. Extracellular signals (salt, drought, etc.) are amplified into the cell nucleus by signal transduction pathway, then interact with defense-related
proteins to activate or inhibit the expression of downstream defense-related genes, thereby forming a chain of stress responses. Moreover, transcription factors (bHLH [11], MYB [12], bZIP [13], NAC [14], WRKY [15], et al.) play a significant role in stress responses, which combines the cis-elements to regulate the gene expression with other regulatory factors [16]. Among them, WRKY transcription factor is a large transcription factor gene family with complex biological functions, and specific to various species from single-celled algae to higher plants. For instance, in Arabidopsis thaliana (A. thaliana), overexpression of AtWRKY50 promoted the production of sinapic derivatives [17]. AtWRKY46, 54, and 70 have important effect on activating the expression of brassinosteroid-mediated gene and restraining the drought gene response [18]. In Taxus chinensis, TcWRKY8 and TcWRKY47 regulated the expression of taxol-biosynthesis-related genes [19]. TaWRKY33 significantly enhanced the wheat drought tolerance [20]. PtrWRKY18 and PtrWRKY35 increased the resistance to Melampsora in Populus [21].

All known WRKYS contain at least one WRKY conserved domain. WRKY domain (around 60 amino acid residues) comprises a highly conserved N-terminal (NT) amino acid sequence (WRKYGQK) and an atypical zinc finger structure (C-X4–5C-X22–23H-X1-H or C-X7C-X23H-X1-C) at the C-terminus [16]. The conserved amino acid sequence WRKYGQK only has the single base mutation, such as WRKYGKK, WRKYGEK, and some WRKY sequences are replaced by WRRY, WSKY, WKRY, WVKY or WKKY sequences sometimes [22, 23]. The WRKY domain specifically binds to W-box (T)(T)TGAC(C/T), the core sequence TGAC decides the binding ability and the play of relevant function [24]. Depending on the number of WRKY domain and the characteristic of zinc finger structure, WRKYS are generally classified into three main groups (1, 2, and 3). There are two WRKY domains in group 1 and zinc finger structure type is C–X4–C–X22–23–H–X1–H, while group 2 and 3 has one WRKY domain, and the type of zinc finger structure is C–X4–C–X22–23–H–X1–H and C–X4–5–C–X23–24–H–X1–H respectively [25]. Group 2 can be further divided into 2a, 2b, 2c, 2d and 2e. In addition, some special R-protein WRKYS were found in certain species, which may lead to signal diversity and even shorten the speed of signal transmission with other components of signaling pathways. There are three, one and one R-protein WRKYS in A. thaliana (AtWRKY16, AtWRKY19 and AtWRKY52), Glycine Max (GmWRKY1) [26] and pineapple (AcWRKY23) [27] respectively.

In recent years, lots of research have focused on the WRKY gene families of many species. Since the first WRKY gene (SPF1) was studied in Ipomoea batatas [28], the WRKY genes of many species were identified (showed in Additional file 1: Table S1). Many WRKY genes had been performed the functional verification in response to salt stress. For example, the up-regulated expression of Glycine max WRKY92, 144 and 165 was connected with salinity stress [29]. VvWRKY30 from grape was confirmed to confer tolerance to salinity stress [30]. GmWRKY45 had a positive effect on the response to phosphorus and salinity stress, and a relation with fertility [31]. At present, there are no reports of the C. sativa WRKY proteins (CsWRKYS). In this study, a total of 242 CsWRKYS were identified and divided into three major groups. There were further comprehensive analyses of the CsWRKYS, including the structural characteristics, physicochemical property, construction of phylogenetic tree, identification of conserved motif, chromosomal locations and synteny analysis. Several specific WRKY members were investigated
in different tissues in the response to salt stress (SS) by the expression profiling of CsWRKYs. This research offered precious resources for understanding the biological roles of CsWRKYs.

**Results**

**Identification of CsWRKYs**

A total of 243 candidate WRKYs were initially obtained. All putative WRKYs were further identified for the presence of complete WRKY-specific domains. One predicted WRKY protein (XP_019082498.1) was removed due to incomplete WRKY domain. The remaining 242 WRKYs were renamed CsWRKY1-CsWRKY242 respectively (Additional file 2: Table S2). Their characteristics were showed including the group and subgroup, gene locus ID, start and end position in the chromosomes, SL, MW and PI. The annotation of 242 WRKYs were scanned using available *C. sativa* genome database. All 202 CsWRKY genes coded 242 CsWRKYs. The formation of 70 CsWRKYs was by alternative splicing of 30 CsWRKY genes, the rest of CsWRKY genes only had one coding protein (Additional file 3: Table S3). The length of protein sequences ranged from 136 (CsWRKY136) to 1699 amino acids (CsWRKY47), with an average of 371 amino acids. The MW ranged from 15.98 kDa (CsWRKY136) to 190.70 kDa (PvWRKY34), with an average of 41.33 kDa. The PI ranged from 4.13 (CsWRKY204) to 10.47 (CsWRKY107), with an average of 7.38 (Additional file 2: Table S2).

**Multiple sequence alignment and comparative phylogenetic analysis**

According to the highly conserved domain (about 60 amino acids) of each WRKY sequence, 242 CsWRKYs were aligned using Bioedit software. Some CsWRKYs selected randomly were showed in Figure 1 because of excessive sequences (Multiple sequence alignment of 242 CsWRKYs were displayed in Additional file 6: Figure S1).

The CsWRKYs included the highly conserved heptapeptide sequence (WRKYGQK) and the zinc-finger motif. A total number of 11 CsWRKYs in the group 2c only were different due to a change in one amino acid of WRKYGQK sequence, CsWRKY202 contained a WRKYGXK sequence, the others had a WRKYGKK sequence (CsWRKY127, CsWRKY128, CsWRKY135, CsWRKY223, CsWRKY228, CsWRKY229, CsWRKY230, CsWRKY232, CsWRKY233 and CsWRKY235).

The comparative phylogenetic tree was built using MEGA 7.0 based on the highly conserved WRKY domains (approximately 60 amino acids) of 242 CsWRKYs and 75 AtWRKYs. The classification of WRKYs was confirmed according to the specific conserved sequences of WRKYs and number of their WRKY zinc-finger motifs [32], which was consistent with some previous findings of *A. thaliana* [32] and *Common bean* [33]. According to the constructed phylogenetic tree of CsWRKYs and AtWRKYs, the WRKYs were classified into three main groups (group 1, 2 and 3) (Figure 2), 69 members belonged to group 1 (56 CsWRKYs), 198 members belonged to group 2 (149 CsWRKYs) and 50 members (37 CsWRKYs) to group 3. Group 2 was subdivided into five subgroups (2a, 2b, 2c, 2d and 2e), which had 16
(12 CsWRKYs), 35 (26 CsWRKYs), 82 (63 CsWRKYs), 34 (24 CsWRKYs) and 31 members (24 CsWRKYs) respectively, most of the CsWRKYs were present in subgroup 2c.

The CsWRKYs were put into the group 1 owing to their two WRKY conserved domains, which were located in the NT and C-terminal (CT) of these sequences respectively. In 242 CsWRKYs, 56 CsWRKYs of the group 1 had two WRKY conserved domains and the zinc-finger domains (C–X4–C–X22–23–H–X1–H). The remaining 186 CsWRKYs contained one WRKY conserved domain. 149 CsWRKYs of the group 2 had the zinc-finger domain of C–X4–5–C–X23–24–H–X1–H. 50 CsWRKYs of the group 3 had the zinc-finger domain of C–X7–C–X23–H–X1–C. There were exceptions, a number of CsWRKYs lacked some amino acids (CsWRKY100, CsWRKY101, CsWRKY111 CsWRKY186, CsWRKY201, CsWRKY208, CsWRKY226, CsWRKY240, CsWRKY241, CsWRKY242), which might be the reason of the incomplete sequencing database.

Each WRKY domain had one highly conserved intron structure except the N-terminal conserved WRKY domain of the WRKYs in the group 1, and the position also was rather conservative (Figure 1). There were two kinds of introns, the constant amino acid PR intron was between the WRKY sequence and the zinc-finger domain in some groups (group 1-CT, 2c, 2d, 2e and 3). The other intron VQR was located in the interior of zinc-finger structure (C–X4–5–C–X5–VQR–X18–19–H–X1–H) in the group 2a and 2b [32]. The conserved intron made the identification of the correct WRKY domain easier.

Some specific WRKYs contained R proteins and the WRKY domains that belonged to the group R, which were one of the most significant features of the WRKY family genes in flowering plants [25]. For example, *A. thaliana* had three R protein-WRKY genes (AtWRKY16, AtWRKY19 and AtWRKY52) [26]. In the phylogenetic tree, AtWRKY19 and CsWRKY47 were gathered to the group 1, the other four sequences (AtWRKY16, AtWRKY52, CsWRKY186 and CsWRKY220) were clustered into the group 2d. The classified standard of R protein-WRKY genes have been characterized by Rinerson et al [26]. According to the further protein architecture analysis of CsWRKY47, CsWRKY186 and CsWRKY220, CsWRKY47 contained the typical domain of PAH-WRKY(1-NT)-WRKY(1-CT)-NB-ARC and the protein kinase domain in the C-terminal end of the sequence, which belonged to the group RW3 and was the only member. CsWRKY186 had the typical domain of NB-ARC-LRR-WRKY. CsWRKY220 had the TIR-NB-ARC-LRR-WRKY domain, which indicated that CsWRKY186 and CsWRKY220 were peculiar and not belong to any group. The three CsWRKYs formed possibly in order to adapt to especial change during the growth.

**Motifs analysis and chromosomal mapping**

A total of 10 motifs in all CsWRKYs were analyzed using the MEME, the motifs identified ranged from 15 to 50 residues in width (Figure 3) and shown with colored boxes according to the scale (Additional file 7: Figure S2). The CsWRKYs were clustered into groups according to the similar quantity and structure of motifs. The information of each motif was displayed in Figure 3. Motif 1, 2 and 3 were found to code the WRKY conserved domain. All CsWRKYs had more than two motifs, and some CsWRKYs helded 7 kinds of motif at most. The motifs were diverse among different groups. For example, the group 1 had 9 motifs (motif 1, 2, 3, 4, 5, 6, 8, 9 and 10), the number of motif 1, 2 and 3 was two, and motif 6 and 9 were unique
to these members of group 1, which may be significant to the specific functional WRKYs. Motif 10 only appeared both group 1 and 2c. Motif 4 and 8 were present in group 1, 2b and 2c. Motif 7 was present in group 2d, 2e and 3. Motif 5 was absent in group 2c and 2e merely. The number of motif 5 was two in all members of the group 2a. As a whole, the motif analysis of CsWRKYs showed that every group possibly had peculiar conservation, corresponding to the grouping of the phylogenetic tree.

All identified CsWRKY genes were distributed across all *C. sativa* chromosomes using MapChart (Chr1–Chr20, Figure 4). Some genes had alternative splicing, so some protein name represented one gene name (in the Additional file 7: Figure S2), the rest of WRKY genes were represented by corresponding protein labels. The map revealed the position of CsWRKY179 was not mapped to any chromosome because it’s present on scaffold region. The remaining distribution of 201 CsWRKY genes on the chromosomes was uneven. Some chromosomes had the same number of CsWRKY genes. For example, the minimal number of CsWRKY genes only was three in several chromosomes (chromosome 1, 15 and 19). Chromosome 3 and 17, chromosome 12 and 20, chromosome 16 and 7 had 6, 10, 15 genes respectively. Some chromosomes (chromosome 2, 8, 9 and 13) had 11 CsWRKY genes. A maximum number of 20 CsWRKY genes on chromosome 11 accounted for 8.3% of all CsWRKY genes. WRKY genes had variant distributions in different species, such as *Chickpea* [34], *Kiwifruit* [24] and *Vitis vinifera* [35].

**Gene duplication, synteny analysis and selection pressure analyses**

Gene duplication played a vital driving force in gene evolution including tandem and segmental duplication. The two main evolutionary patterns played a prominent role in the generation of new genes and the evolutionary formation of new functions, which was also the pivotal reason to expand the member of gene family in plants [22, 29, 36-38]. In order to analyze the expansion patterns of CsWRKY genes, tandem duplication event (the distance of two or more genes within 200kb in equal chromosome) was confirmed on the basis of the method by Guo et al [35]. The results indicated that the gene duplication of the CsWRKYs contained no tandem duplication. Meanwhile, the gene duplication of CsWRKYs among *C. sativa* chromosomes was identified using Blastp and MCScanX. The analysis showed that there was no tandem duplication, which was consistent with above result. There were 137 segmental duplication events (Figure 5, Additional file 4: Table S4). These revealed that certain CsWRKYs appeared with the segmental duplication events to a great extent. It has been reported that there were 17 and 55 segmental duplication events in *pineapple* and *kiwifruit* chromosomes [24, 27]. The above result indicated that segmental duplication was an uppermost mode of the evolution of CsWRKY genes.

The comparative synteny maps of two different genomes (*C. sativa* and *A. thaliana*, *C. sativa* and *B. rapa*) were performed in order to explore the origin and evolution of CsWRKY genes (Figure 6, Additional file 5: Table S5). The corresponding orthologs between 173 CsWRKY genes and 65 AtWRKY genes were confirmed and had 173 gene pairs. The corresponding orthologs of 166 CsWRKY genes and 111 BrWRKY genes had 282 gene pairs. The multiple CsWRKY genes corresponded to a single AtWRKY gene. These syntenic relationship showed the expansion of CsWRKY genes possibly occurred after *A. thaliana* in evolution.
Expression profiling analysis

The transcript expression levels of 202 CsWRKY genes were displayed in Figure 7. The expression number of CsWRKY genes approximately reached 89.11% in R, followed by 88.61% (F), 81.68% (IF), 80.69% (S), 78.71% (OL), 77.23% (ESD), 76.73% (EMSD) and 75.25% (GS). The percent of CsWRKY genes expression in other tissues were 68.32% (C), 67.82% (LMSD) and 64.85% (YL) respectively. The lowest expression percent was 52.48% in LSD. The 70 CsWRKY genes (34.65%) expressed in twelve tissues including 28 CsWRKY genes in group 1, 41 CsWRKY genes in group 2 and 1 CsWRKY gene in group 3. Simultaneously, the 24 CsWRKY genes highly expressed in every tissue. Furthermore, CsWRKY36 and CsWRKY227 not expressed in every tissue. Some CsWRKY genes also had the tissue specific expression pattern (Figure 7). In group 1, there were two genes (CsWRKY25 and CsWRKY34) co-expressed in IF and LSD. CsWRKY174 in group 2b was co-expressed in GS and R. In group 2c, CsWRKY182 and CsWRKY202 specifically were expressed in F, CsWRKY204 only was expressed in ESD, CsWRKY73 was co-expressed in F and OL, CsWRKY203 was co-expressed in F and GS. CsWRKY118 in group 2e was co-expressed in GS and S. In group 3, there were three genes (CsWRKY111, CsWRKY208 and CsWRKY226) merely expressed in R, CsWRKY100 was expressed both in ESD and R.

The functions of most AtWRKY genes had been verified. It has been reported that AtWRKY25 and AtWRKY33 were sensitive to NaCl, and overexpression of either gene enhanced tolerance to NaCl [39], which belonged to group 1. According to structural features and phylogenetic analysis of CsWRKYs and AtWRKYs, we found that AtWRKY33 was the ortholog of CsWRKY8, CsWRKY9 and CsWRKY10, AtWRKY25 was the ortholog of CsWRKY48, CsWRKY49 and CsWRKY50. These CsWRKY genes belonged to group 1. To provide further clues to the function of CsWRKY genes under SS, the expression levels of six CsWRKY genes in roots and shoots under SS and normal conditions (NC) were displayed in Figure 8. These CsWRKY genes had higher levels in shoots than that in roots under same conditions. Compared with gene expression levels in NC, three CsWRKY genes (CsWRKY8, CsWRKY9 and CsWRKY10) under SS had a significant down-regulated expression in roots, three other CsWRKY genes (CsWRKY48, CsWRKY49 and CsWRKY50) under SS had no obvious change in roots. These CsWRKY genes had higher levels of expression to varying degrees under SS in shoots. In particular, CsWRKY50 were significantly up-regulated in shoots under SS. The expression analyses implied that these CsWRKY genes played an important part in the response to SS. These results meant that most genes were deliberately expressed in space and time, which were not useless genes. They may have functional differentiation in the process of evolution to cope with different environmental changes.

Discussion

Large amounts of WRKYs in Camelina sativa

WRKYs were widely present in various plants, and a large number of WRKYs were identified. In this study, a total of 202 CsWRKY genes were obtained and coded 242 CsWRKYs. The distribution of the CsWRKYs in groups was very similar to that of AtWRKYs. The comprehensive analysis showed the characteristic
sequences of CsWRKYs were conserved in the evolutionary process. Compared with other reported plant species, CsWRKYs family had the largest members and was the largest WRKY gene family (Additional file 1: Table S1). The number of CsWRKYs in the group 2 and 3 was threefold than AtWRKYs approximately, and the group 1 members increased largely compared with AtWRKYs, these might be the expanding result of the number of CsWRKYs in the process of evolution.

**Structure and classification analysis**

The phylogeny and motifs of 242 CsWRKYs were analyzed (Additional file 7: Figure S2). The identification of motifs also might provide support force of evidence to reveal the duplicate evolutionary process of the WRKY family genes and explore the mechanism of functional conservation. All CsWRKYs were classified to these groups (group 1, 2a, 2b, 2c, 2d, 2e and 3), and each group or subgroup had largely semblable structure and the specific domain combination. The group-specific motifs might due to functional differentiation of CsWRKYs. And CsWRKYs had the similar characters in the average amino acids length, MW and PI compared to other species, such as chickpea [34] and common bean [33]. The CsWRKYs in group 1 contained two WRKY domains and were the ancient members during the WRKY evolution. The WRKYs in group 2a and 2b came from an algal single WRKY domain or the other Group 1-derived lineage [26, 34]. The members of group 2c had been evolved from the lack of N-terminal domain of CsWRKYs in group 1, and the number of CsWRKYs reached the highest in all groups. The same results had been observed in some species, such as *Salvia miltiorrhiza*, *Panicum miliaceum* L and common bean [16, 33, 40]. The increasing largely quantity might due to adopt to various changes of the whole evolution. There were eleven variant CsWRKYs including WRKYGXK and WRKYGKK in group 2c, the different variants (WRKYGKK, WRKYGEK, WKKYEDK, et al) also were found in many species [16, 35, 40, 41], these variants might give WRKYs multiple biological functions [33]. In addition, R-type CsWRKY 47 belonged to RW3, CsWRKY 186 and CsWRKY 220 were new R-type proteins in *C. sativa*. From present research *Arabidopsis* WRKY coded a NBS-LRR-WRKY protein, the protein worked as chimeric protein and the WRKY domain had the DNA-binding activity [42]. The interaction of AtWRKY52/RRS1 with R protein RPS4 protected plant from the infection of fungal and bacterial pathogens [38]. These indicated that various R-type CsWRKYs possibly gave *C. sativa* variant disease resistance.

**Duplication events plays an important role in evolution of CsWRKY**

The expansion reason and complex mechanism of gene families was mainly caused by gene duplication [34]. CsWRKY genes had no tandem duplication event and 137 segmental duplication events, which revealed that segmental duplication events were the major force to expand the CsWRKY members. There were similar results in *Chickpea*, *Pineapple* and *Kiwifruit* [24, 27, 34]. Thus, it’s speculated that segmental duplication events were the primary cause to reach the largest number for CsWRKY genes. The genetic evolutionary process was reestablished probably by comparing all gene sequences within the same genome or among the different genomes [43]. The comparative synteny maps (*C. sativa* and *A. thaliana*, *C. sativa* and *B. rapa*) were performed. According to the synteny map of *C. sativa* and *A. thaliana*, 173 CsWRKY genes have 65 corresponding *Arabidopsis* orthologs, the multiple CsWRKY genes corresponded
to a single AtWRKY gene. According to the syntenic map of *C. sativa* and *B. rapa*, 166 CsWRKY genes have 111 corresponding *B. rapa* orthologs. Thus, the expansion of CsWRKY genes probably occurred after *A. thaliana* and *B. rapa*. These CsWRKY genes probably formed new gene functions to adapt to various changes so that the member was expanded. The majority Ka/Ks ratios of the above syntenic WRKY gene pairs were less than 1, which demonstrated that CsWRKY genes had experienced strongly the evolution of purify selection. Purify selection (Ka/Ks ratio < 1) usually caused the elimination of deleterious genes in evolution, which indicated the key conserved sequences of WRKY genes was beneficial to the plant survival and growth [8, 34].

**Vital regulation role of CsWRKYs**

AtWRKY25 and AtWRKY33 responded to SS had three orthologs in *C. sativa* respectively, which was owing to expansion of CsWRKY members in evolution. To further investigate the response of CsWRKYs to SS, six CsWRKY genes in roots and shoots were monitored, which were induced to varying degrees under SS (Figure 8). The results revealed that CsWRKY genes may perform complex function and play an important role in response to SS. Previous studies have indicated that the high expression of WRKYs had the vital influence by transcription activation of downstream target genes in the plant growth and development [44]. The tissue-specific expression of WRKY genes might have great effect during the plant growth and development by regulating the transcriptional process [15]. Therefore, the comprehensive analyses and stress response of CsWRKYs would lay a solid foundation for further study of the WRKY gene function in transcriptional regulation.

**Conclusion**

WRKYs were widely present in various plants, the members of CsWRKYs identified expanded largely compared with AtWRKYs. Comprehensive analysis revealed that expansion of CsWRKY members had multiple reasons, segmental duplication events were the major force, which increased the awareness of CsWRKY genes during the functional conservation and evolution. Expression patterns establishes a foundation for the future series of exploration in functional characteristics of CsWRKY genes and mechanism of *C. sativa* in respond to salt stress.

**Methods**

**Sequence identification**

The complete genome, proteome and CDS sequence files of *C. sativa* were downloaded from webpage of NCBI ([ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF_000633955.1_Cs](ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF_000633955.1_Cs)). The WRKY domain HMM (Hidden Markov Model) profile numbered PF03106 was extracted from the Pfam protein family database ([http://pfam.xfam.org/](http://pfam.xfam.org/)) [33]. The candidate WRKY protein sequences were discovered by the comprehensive research of utilization of HMMER (E-value cut-off < 1E-5) and BLAST analyses (75 AtWRKYs as queries) in *C. sativa* whole genome protein database [45]. These CsWRKYs sequences were
identified by checking the complete WRKY conserved domain with SMART (http://smart.embl-heidelberg.de/) and InterPro (http://www.ebi.ac.uk/interpro/), the redundant sequences were manually removed. The confirmed CsWRKYs used the ExPasy online tool website (http://web.expasy.org/protparam/) to calculate physicochemical properties including the protein sequence length (SL), molecular weight (MW) and isoelectric point (PI).

**Multiple sequence alignment and the construction of comparative phylogenetic tree**

Multiple sequence alignment of the CsWRKY domain was performed using ClustalW with Bioedit software. Based on the alignment WRKY domains of CsWRKYs and AtWRKYs, the phylogenetic tree was constructed with MEGA 7.0 using Neighbor-Joining method and the parameters (Poisson model, pairwise deletion, and 1000 bootstrap replications). All identified CsWRKYs were divided into different groups according to classification of AtWRKYs sequences. The AtWRKYs sequences were obtained from the TAIR database (http://www.arabidopsis.org/).

**Motifs analysis**

The MEME (http://meme.nbcr.net/meme/intro.html) was used to analyze CsWRKYs and searched for 10 conserved motifs [46]. The interrelated parameters as follow: the repetitive time was any, the maximum motif number was 20 and the motif width was between 5 and 50 residues. The MEME results were displayed with the TBtool software [47].

**Chromosomal distribution**

The chromosomal position of all CsWRKY genes was determined from the genome annotation file. The physical position of CsWRKY genes from short-arm to long-arm telomeres on the chromosome was mapped using MapChart [48].

**Gene duplication and selection pressure analyses**

The gene duplication events within *C. sativa* genome were studied by using MCScanX (Multiple Collinearity Scan toolkit), the whole analytic process was utilized with default setting. The synteny relation between two different genomes (*C. sativa* and *A. thaliana*, *C. sativa* and *Brassica rapa* (*B. rapa*)) were identified and analyzed. The *B. rapa* data was downloaded from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html). The synonymous substitution rates (Ks) and non-synonymous substitution rates (Ka) of the identified CsWRKY gene pair were calculated using KaKs Calculator 2.0, the Ka/Ks was also calculated to determine there had the selective pressure act on protein-coding CsWRKY genes or not [49].

**Expression profiling of CsWRKY genes in different tissues**

The transcriptional data of twelve tissues of *C. sativa* was obtained by Kagale et al [7]. During twelve tissues of various developmental stages, the expression levels of CsWRKY genes were utilized to
analysis, including cotyledon (C), early-mid seed development (EMSD), early seed development (ESD), flower (F), germinating seed (GS), inflorescence (IF), late seed development (LSD), late-mid seed development (LMSD), mature leaf (OL), root (R), stem (S) and young leaf (YL). The hierarchical clustering and the heatmap-based expression profiles of CsWRKY genes were performed by HemI1.0. In order to further explore the function of CsWRKY genes, we researched whether some CsWRKY genes responds to SS. Salt stress was imposed by treating 21 days old seedlings with 192 mM NaCl, and the RNA-Seq data of *C. sativa* we used in response to SS came from Heydarian et al [4].

**Abbreviations**

WRKYs, WRKY proteins; NT, N-terminal; CT, C-terminal; SL, protein sequence length; MW, molecular weight; PI, isoelectric point; Ks, synonymous substitution rates; Ka, non-synonymous substitution rates; C, cotyledon; EMSD, early-mid seed development; ESD, early seed development; F, flower; GS, germinating seed; IF, inflorescence; LSD, late seed development; LMSD, late-mid seed development; OL, Mature leaf ; R, Root; S, stem; YL, young leaf

**Declarations**

**Authors’ contributions**

YNS and HLC conceived of the study, participated in all the bioinformatics analysis, including the sequence alignment and phylogeny analysis and drafted the manuscript; Ying Shi helped in bioinformatics analysis. JNX, CLJ and CHZ designed the study. RZL guided the writing of the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

This work was supported by grants from National Natural Science Foundation of China (31902394), Specialized Scientific Project of Shanxi Agricultural Valley Construction (SXNGJSDKYX201906), Applying Basic Research Planning Project of Shanxi Province (201801D221250), Key Research and Development (R&D) Planning Project of Shanxi Province (201803D31063), Key Research and Development (R&D) Planning Project of Jinzhong City (201903D01100013 and Y192012), Science and Technology Innovation Planning Project of Shanxi Agricultural University (2018YJ16), the State Ministry of Agriculture "948" Project (2014-Z39), Shanxi Province Key Projects of Coal-based Science and Technology (FT-2014-01), Research Project Supported by Shanxi Scholarship Council of China (2015-064), the Project of the Key Research and Development Program of Shanxi Province, China (Grant No. 201603D312005).

**Availability of data and materials**

All data generated or analysed during this study are included in this published article [and its supplementary information files].
Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Liang, C., et al., De novo assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. BMC Genomics, 2013. 14: p. 146.
2. Brock, J.R., et al., Phylogenetics of *Camelina Crantz*. (Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*). Mol Phylogenet Evol, 2018. 127: p. 834-842.
3. Abdullah, H.M., et al., Transcriptome profiling of *Camelina sativa* to identify genes involved in triacylglycerol biosynthesis and accumulation in the developing seeds. Biotechnol Biofuels, 2016. 9: p. 136.
4. Heydarian, Z., et al., Changes in gene expression in *Camelina sativa* roots and vegetative tissues in response to salinity stress. Scientific Reports, 2018. 8(1).
5. Yuan, L., et al., Characterisation of phospholipid: diacylglycerol acyltransferases (PDATs) from *Camelina sativa* and their roles in stress responses. Biol Open, 2017. 6(7): p. 1024-1034.
6. Yuan, L., et al., Characterisation of phospholipid: diacylglycerol acyltransferases (PDATs) from *Camelina sativa* and their roles in stress responses. Biology Open, 2017. 6(7): p. 1024-1034.
7. Kagale, S., et al., The developmental transcriptome atlas of the biofuel crop *Camelina sativa*. Plant J, 2016. 88(5): p. 879-894.
8. Biswas, S. and J.M. Akey, Genomic insights into positive selection. Trends Genet, 2006. 22(8): p. 437-46.
9. Na, G., et al., Enhancing microRNA167A expression in seed decreases the alpha-linolenic acid content and increases seed size in *Camelina sativa*. Plant J, 2019. 98(2): p. 346-358.
10. Abdullah, H.M., et al., Comparative transcriptome and metabolome analysis suggests bottlenecks that limit seed and oil yields in transgenic *Camelina sativa* expressing diacylglycerol acyltransferase 1 and glycerol-3-phosphate dehydrogenase. Biotechnol Biofuels, 2018. 11: p. 335.
11. Jin, C., et al., Overexpression of a bHLH1 transcription factor of pyrus ussuriensis confers enhanced cold tolerance and increases expression of stress-responsive genes. Front Plant Sci, 2016. 7: p. 441.
12. Guo, H., et al., Expression of the MYB transcription factor gene BpMYB46 affects abiotic stress tolerance and secondary cell wall deposition in *Betula platyphylla*. Plant Biotechnol J, 2017. 15(1): p.
107-121.
13. Ji, C., et al., Analysis of bZIP transcription factor family and their expressions under salt stress in *Chlamydomonas reinhardtii*. Int J Mol Sci, 2018. 19(9).
14. Yang, X., et al., Miscanthus NAC transcription factor MINAC12 positively mediates abiotic stress tolerance in transgenic *Arabidopsis*. Plant Sci, 2018. 277: p. 229-241.
15. Tang, J., et al., Genome-wide fractionation and identification of WRKY transcription factors in Chinese Cabbage (*Brassica rapa ssp. pekinensis*) reveals collinearity and their expression patterns under abiotic and biotic stresses. Plant Molecular Biology Reporter, 2013. 32(4): p. 781-795.
16. Li, C., et al., Molecular cloning and expression analysis of WRKY transcription factor genes in *Salvia miltiorrhiza*. BMC Genomics, 2015. 16: p. 200.
17. Hussain, R.M.F., et al., Overexpression of AtWRKY50 is correlated with enhanced production of sinapic derivatives in *Arabidopsis*. Metabolomics, 2018. 14(3): p. 25.
18. Chen, J., et al., *Arabidopsis* WRKY46, WRKY54, and WRKY70 transcription factors are involved in brassinosteroid-regulated plant growth and drought responses. Plant Cell, 2017. 29(6): p. 1425-1439.
19. Zhang, M., et al., Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in *Taxus chinensis*. Sci Rep, 2018. 8(1): p. 5197.
20. He, G.H., et al., Drought-responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat confer drought and/or heat resistance in *Arabidopsis*. BMC Plant Biol, 2016. 16(1): p. 116.
21. Jiang, Y., et al., The WRKY transcription factors PtrWRKY18 and PtrWRKY35 promote Melampsora resistance in *Populus*. Tree Physiol, 2017. 37(5): p. 665-675.
22. Zhao, J., et al., Evolutionary and expression analyses of the apple basic leucine zipper transcription factor family. Front Plant Sci, 2016. 7: p. 376.
23. Song, H., et al., Genome-wide identification and characterization of WRKY gene family in peanut. Front Plant Sci, 2016. 7: p. 534.
24. Jing, Z. and Z. Liu, Genome-wide identification of WRKY transcription factors in kiwifruit (*Actinidia spp.*) and analysis of WRKY expression in responses to biotic and abiotic stresses. Genes Genomics, 2018. 40(4): p. 429-446.
25. Rushton, P.J., et al., WRKY transcription factors. Trends Plant Sci, 2010. 15(5): p. 247-58.
26. Rinerson, C.I., et al., The evolution of WRKY transcription factors. BMC Plant Biol, 2015. 15: p. 66.
27. Xie, T., et al., Genome-wide investigation of WRKY gene family in pineapple: evolution and expression profiles during development and stress. BMC Genomics, 2018. 19(1): p. 490.
28. Characterization of a cDNA encoding
29. Song, H., et al., Global analysis of WRKY genes and their response to dehydration and salt stress in soybean. Front Plant Sci, 2016. 7: p. 9.
30. Zhu, D., et al., VvWRKY30, a grape WRKY transcription factor, plays a positive regulatory role under salinity stress. Plant Science, 2019. 280: p. 132-142.
31. Li, C., et al., GmWRKY45 enhances tolerance to phosphate starvation and salt stress, and changes fertility in transgenic Arabidopsis. Frontiers in Plant Science, 2020. 10.
32. Eulgem, T., et al., The WRKY superfamily of plant transcription factors. Trends Plant Sci, 2000. 5(5): p. 199-206.
33. Wu, J., et al., Genome-wide investigation of WRKY transcription factors involved in terminal drought stress response in common bean. Front Plant Sci, 2017. 8: p. 380.
34. Waqas, M., et al., Genome-wide identification and expression analyses of WRKY transcription factor family members from chickpea (Cicer arietinum L.) reveal their role in abiotic stress-responses. Genes Genomics, 2019. 41(4): p. 467-481.
35. Guo, C., et al., Evolution and expression analysis of the grape (Vitis vinifera L.) WRKY gene family. J Exp Bot, 2014. 65(6): p. 1513-28.
36. Zhu, Y., et al., Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. Bmc Plant Biology, 2014. 14.
37. Vatansever, R., et al., Genome-wide identification and expression analysis of sulfate transporter (SULTR) genes in potato (Solanum tuberosum L.). Planta, 2016. 244(6): p. 1167-1183.
38. Narusaka, M., et al., RRS1 and RPS4 provide a dual resistance-gene system against fungal and bacterial pathogens. Plant J, 2009. 60(2): p. 218-26.
39. Li, S., et al., Arabidopsis thaliana WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. Planta, 2011. 233(6): p. 1237-1252.
40. Yue, H., et al., Transcriptome-wide identification and expression profiles of the WRKY transcription factor family in Broomcorn millet (Panicum miliaceum L.). BMC Genomics, 2016. 17: p. 343.
41. Zou, Z., et al., Gene structures, evolution and transcriptional profiling of the WRKY gene family in castor bean (Ricinus communis L.). PLoS One, 2016. 11(2): p. e0148243.
42. Noutoshi, Y., et al., A single amino acid insertion in the WRKY domain of the Arabidopsis TIR-NBS-LRR-WRKY-type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. Plant J, 2005. 43(6): p. 873-88.
43. Koonin, E.V., Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet, 2005. 39: p. 309-38.
44. Xu, H., et al., WRKY transcription factor genes in wild rice Oryza nivara. DNA Res, 2016. 23(4): p. 311-23.
45. Wei, Y., et al., Genome-wide identification and expression analysis of the WRKY gene family in Cassava. Front Plant Sci, 2016. 7: p. 25.
46. Bailey, T.L., et al., MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res, 2009. 37(Web Server issue): p. W202-8.
47. Song, M. and X. Peng, Genome-wide identification and characterization of DIR genes in Medicago truncatula. Biochem Genet, 2019.
48. Voorrips, R.E., MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered, 2002. 93(1): p. 77-8.
49. Wang, Y., et al., MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res, 2012. 40(7): p. e49.

Figures
Figure 1

Multiple sequence alignment of the WRKY domain from CsWRKYs. Black box indicated the conserved WRKY amino acid sequence and zinc-finger domain. Black line indicated the position of the conserved PR intron and VQR intron.
Figure 2

Comparative phylogenetic tree showed the domains relationship of CsWRKYs and AtWRKYs. The identified CsWRKYs (black hollow circle) and scattered AtWRKYs (black solid circle) were clustered to the different groups in phylogenetic tree. The various colors represent the different groups (or subgroups) of CsWRKY domains.
| Domain | E-value   | Sites | Width | Multilevel consensus sequence |
|--------|-----------|-------|-------|-------------------------------|
| 1      | $1.4e-6488$ | 292   | 27    | ![Sequence 1](image)           |
| 2      | $6.3e-2076$ | 283   | 15    | ![Sequence 2](image)           |
| 3      | $1.2e-2017$ | 243   | 21    | ![Sequence 3](image)           |
| 4      | $8.6e-1017$ | 111   | 21    | ![Sequence 4](image)           |
| 5      | $6.1e-629$   | 91    | 29    | ![Sequence 5](image)           |
| 6      | $1.1e-401$   | 12    | 50    | ![Sequence 6](image)           |
| 7      | $2.3e-380$   | 38    | 24    | ![Sequence 7](image)           |
| 8      | $1.6e-368$   | 22    | 41    | ![Sequence 8](image)           |
| 9      | $3.3e-364$   | 16    | 50    | ![Sequence 9](image)           |
| 10     | $2.4e-329$   | 30    | 29    | ![Sequence 10](image)          |
Figure 3

Details of 10 Motifs of CsWRKYs.

Figure 4

Chromosomal distribution of CsWRKY genes. The position of every CsWRKY gene can be determined using the left scale.
Figure 5

Synteny analysis of interchromosomal relationships of CsWRKY genes. All gene pairs and CsWRKY gene pairs in the C. sativa genome were indicated by gray lines and red lines respectively.
Figure 6

Syntenic analysis of CsWRKY genes between C. sativa and two plant species (A. thaliana and B. rapa). (A) C. sativa and A. thaliana (B) C. sativa and B. rapa. The collinear blocks between C. sativa and other species was showed gray lines. The syntenic WRKY gene pairs between C. sativa and A. thaliana were highlighted with blue. The syntenic WRKY gene pairs between C. sativa and B. rapa were highlighted with yellow. The chromosome number was indicated at the top of every chromosome.
Figure 7

Heat map representation and hierarchical clustering of the CsWRKY gene expression profiles in twelve tissues. (A) Group 1. (B) Group 2. (C) Group 3. The expression levels of CsWRKY genes were showed by different colors on the right scale.
Figure 8

Expression profiles of CsWRKY genes in roots and shoots under salt stress. SS, salt stress. NC, normal conditions. Data are means ± SE calculated from three biological replicates. The expression levels of CsWRKY genes in different tissues were compared to NC with one-way ANOVA at significance levels of ** P ≤ 0.01 and * P ≤ 0.05.

Supplementary Files

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

- TableS1.xlsx
- FigureS2.jpg
- TableS3.xlsx
- TableS2.xlsx
- TableS5.xlsx
- TableS4.xlsx
- FigureS1.jpg