2020

Comprehensive identification and analyses of the Hsf gene family in the whole-genome of three Apiaceae species

Qiaoying Pei

North China University of Science and Technology

Tong Wu

Qihang Yang

Ke Gong

See next page for additional authors

Authors

Qiaoying Pei, North China University of Science and Technology, Tong Wu, Qihang Yang, Ke Gong, Rong Zhou, Chunlin Cui, Ying Yu, Wei Zhao, Xi Kang, Rui Cao, and Xiaoming Song

September 2021. Horticultural Plant Journal, 7 (5): 457–468.

Horticultural Plant Journal

Available online at www.sciencedirect.com
The journal's homepage: http://www.keaipublishing.com/en/journals/horticultural-plant-journal

Comprehensive identification and analyses of the *Hsf* gene family in the whole-genome of three Apiaceae species

Qiaoying Pei[a], *Tong Yu*[a], *Tong Wu*[a], *Qihang Yang*[a], *Ke Gong*[a], *Rong Zhou*[b], *Chunlin Cui*[a], *Ying Yu*[a], *Wei Zhao*[a], *Xi Kang*[a], *Rui Cao*[a,*], *and Xiaoming Song*[a,c,d,*]

[a] *College of Life Sciences, North China University of Science and Technology, Tangshan, Hebei 063210, China*
[b] *Department of Food Science, Aarhus University, Aarhus 8200, Denmark*
[c] *School of Life Science and Technology and Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China*
[d] *Food Science and Technology Department, University of Nebraska-Lincoln, Lincoln, Nebraska 68526, USA*

Received 20 April 2020; Received in revised form 29 July 2020; Accepted 21 August 2020

Available online 25 August 2020

A B S T R A C T

Apiaceae is a major family from Apiales and includes many important vegetable and medicinal crops. Heat shock transcription factors (*Hsf*) play important roles in heat tolerance during plant development. Here, we conducted systematic analyses of the *Hsf* gene family in three Apiaceae species, including 17 *Apium graveolens* (celery), 32 *Coriandrum sativum* (coriander), and 14 *Daucus carota* (carrot). A total of 73 *Hsf* genes were identified in three representative species, including *Arabidopsis thaliana*, *Vitis vinifera*, and *Lactuca sativa*. Whole-genome duplication played important roles in the *Hsf* gene family's expansion within Apiaceae. Interestingly, we found that coriander had more *Hsf* genes than celery and carrot due to greater expansion and fewer losses. Twenty-seven branches of the phylogenetic tree underwent considerable positive selection in these Apiaceae species. We also explored the expression patterns of *Hsf* genes in three plant organs. Collectively, this study will serve as a rich gene resource for exploring the molecular mechanisms of heat tolerance. Additionally, this is the first study to report on the *Hsf* gene family in Apiaceae; thus, our research will provide guidance for future comparative and functional genomic studies on the *Hsf* gene family and others in Apiaceae.

Keywords: *Hsf* gene family; Gene duplication and loss; Expression pattern; Apiaceae

1. Introduction

Apiaceae is one of the largest aromatic flowering plant families. It contains approximately 450 genera and 3 700 species, which are widely distributed across the temperate zone (Lee et al., 2019). The typical representative plants of Apiaceae include celery, coriander, and carrot, among others. Celery, which is grown worldwide, is a popular medicinal herb (Maljaei et al., 2019; Yin et al., 2020). Coriander is a phytogenic herb that is widely distributed throughout North Africa, central Europe, and Asia (Sojic et al., 2019). Carrot is a biennial herbaceous species and belongs to the root vegetables (Que et al., 2019). Roots contain high quantities of alpha- and beta-carotene, and are a good source of vitamin K and vitamin B6, which may protect humans against certain diseases (Iorizzo et al., 2016; Liu et al., 2019c). Overall, these three plants are well known and are widely used as vegetables, spices, and medicine.

Plants suffer from several extreme environmental stresses, including high temperature, high salinity, drought, and water deficiency (Song et al., 2014; Yan et al., 2018; Castañares and Bouzo,

* Corresponding authors. Tel.: +86 315 8805607
E-mail addresses: caoruivv@163.com; songxm@ncst.edu.cn
Peer review under responsibility of Chinese Society for Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS)

https://doi.org/10.1016/j.hpj.2020.08.005

2019; Liu et al., 2019a; Qi et al., 2019). Among these stresses, heat is one of the most detrimental as it negatively affects crop yield (Kumar et al., 2018a). In response, plants need to relieve its adverse effects (Song et al., 2014; Wan et al., 2019). The ability of plants to adapt to stress is important for their growth and development (Guo et al., 2008; Zhou et al., 2019). Furthermore, adjustments to stress are important for controlling the expression of stress-responsive genes (Gomez-Pastor et al., 2018; Zhou et al., 2019). *Hsf* (Heat shock transcription factor) is a key transcription factor that responds to heat stress and plays an important role in heat resistance (Ohama et al., 2016; Wan et al., 2019).

Hsf shares a DNA-binding domain (DBD) at the N-terminus, an adjacent oligomerization domain (OD or HR-A/B), and nuclear localization signal (NLS) motifs (Scharf et al., 2012; Li et al., 2019). Additionally, *Hsf* has a C-terminal activation domain (CTAD), AHA motifs, and a nuclear export signal (NES) (Liu et al., 2018; Li et al., 2019). The OD sequence contains a heptad repeat pattern of hydrophobic amino acid residues. *Hsf* has been classified into three groups based on the number of amino acid residues between two heptad repeats, respectively named HsfA, HsfB, and HsfC (Nover et al., 2001; Liu et al., 2018). In groups A and C, there are 21 and 7 amino acid residue insertions in the HR-A/B regions, whereas there are no insertions in group B, which is similar to all non-plant *Hsf*s (Nover et al., 2001). Additionally, group A has AHA motifs, which have not been detected in groups B or C (Döring et al., 2000).

The first *Hsf* family gene was cloned and characterized in yeast (Sorger and Pelham, 1988). Subsequently, corresponding genes were cloned in *Drosophila melanogaster*, *Homo sapiens*, and tomato (Pirkkala et al., 2001; Guo et al., 2008). With the increase in available genomes, the *Hsf* gene family has been studied at the whole-genome level in many plants. A total of 21, 25, 24, 35, 36, 64, 32, 18, 29, 20, 19, 25, and 26 *Hsf* family genes have been identified in *Arabidopsis*, rice, tomato, *Brassica oleracea*, *B. rapa*, *B. napus*, cassava, *Prunus mume*, Tartary buckwheat, chickpea, grape, pepper, and soybean, respectively (Guo et al., 2008; Scharf et al., 2012; Chung et al., 2013; Song et al., 2014; Guo et al., 2015; Huang et al., 2015; Zafar et al., 2016; Zhu et al., 2017; Liu et al., 2018, 2019b; Lohani et al., 2019; Wan et al., 2019; Yu et al., 2019). Collectively, these studies have served as rich resources for comparative analyses on the *Hsf* gene family in plants. However, there has been no report on the heat shock transcription factor (*Hsf*) gene family in these three plants (celery, coriander and carrot) despite its important role in plant heat resistance. An investigation of the *Hsf* gene family in the whole-genome of celery, coriander, and carrot is valuable for plant heat resistance breeding.

In this study, we comprehensively described the *Hsf* gene family in celery, coriander, and carrot. A comparative analysis using lettuce, grape, and *Arabidopsis* as representative species was also conducted. The goals of this study were to (i) identify *Hsf* family genes in celery, coriander, and carrot; (ii) classify *Hsf* family genes using phylogenetic relationships; (iii) map *Hsf* family genes on corresponding chromosomes; (iv) identify orthologous and paralogous *Hsf* family genes; (v) detect duplications or losses of *Hsf* family genes; and (vi) analyze *Hsf* family gene expression patterns in three plant organs using RNA-seq. This study will serve as a useful resource for future studies on the biological functions of the *Hsf* gene family. Additionally, these findings enhance our understanding of the evolutionary history of the *Hsf* gene family in Apiaceae.

2. Materials and methods

2.1. Genome sequence collection and Hsf family gene identification

The gene sequences of coriander (*Coriandrum sativum*) and celery (*Apium graveolens*) were downloaded from the CGDB database (Song et al., 2020a). For *Arabidopsis*, they were retrieved from the *Arabidopsis* Information resource (TAIR10, http://www.arabidopsis.org/). The gene sequences of carrot (*Daucus carota*) (version 2), lettuce (*Lactuca sativa*) (version 5), and grape (*Vitis vinifera*) (Genoscope.12X) were downloaded from the Phytozome database (http://www.phytozome.net/) (Jaillon et al., 2007; Iorizzo et al., 2016; Reyes-Chin-Wo et al., 2017). *Arabidopsis* and grape were selected as they are typical model species and their *Hsf* family genes have been well characterized. Lettuce was selected as it is a close relative to Apiaceae. The Pfam database was used to identify *Hsf* family gene (PF00447) with an e-value $< 1e^{-4}$ (Punta et al., 2012). The SMART and CDD databases were used to perform domain validation and ensure accuracy (Marchler-Bauer et al., 2009; Letunic et al., 2012). Additionally, DBD and HR-A/B domains were detected using the SMART and MARCOIL databases, respectively (Letunic et al., 2012; Zimmermann et al., 2018).

2.2. Phylogenetic analysis

Amino acid sequences of the *Hsf* family genes from six species were used to construct a phylogenetic tree using MEGA X software (Kumar et al., 2018b). First, multiple alignments were conducted using ClustalW (https://www.genome.jp/tools-bin/clustalw); then the phylogenetic tree was constructed using the neighbor-joining (NJ) method with 1 000 bootstrap replicates. Evolutionary distances were computed using the Poisson correction method.

2.3. Chromosomal localization, gene structure, and conserved motif analyses

The distributions of *Hsf* family genes on each chromosome in celery, coriander, and carrot were drawn according to the physical position of their general feature format (gff) files using Tbtools software (https://github.com/CJ-Chen/TBtools-Manual). The *Hsf* family gene structure in these species was performed using Gene Structure Displayer Server v2.0 (GSDS) (Hu et al., 2015). The gff files were submitted to GSDS website, which showed the position of exons, introns, and untranslated regions (UTR). The amino acid sequences of the *Hsf* family genes in celery, coriander, and carrot were analyzed using Multiple Expectation maximization for Motif Elicitation (MEME) to search for conserved motifs with the default parameters (Bailey et al., 2009).

2.4. Ortholog, paralog, and gene duplication analyses

The orthologous and paralogous of *Hsf* family genes in celery, coriander, and carrot were identified using OrthoMCL with an e-value of $1e^{-5}$ (Li et al., 2003). The relationships of the orthologous and paralogous genes in these three species were drawn using

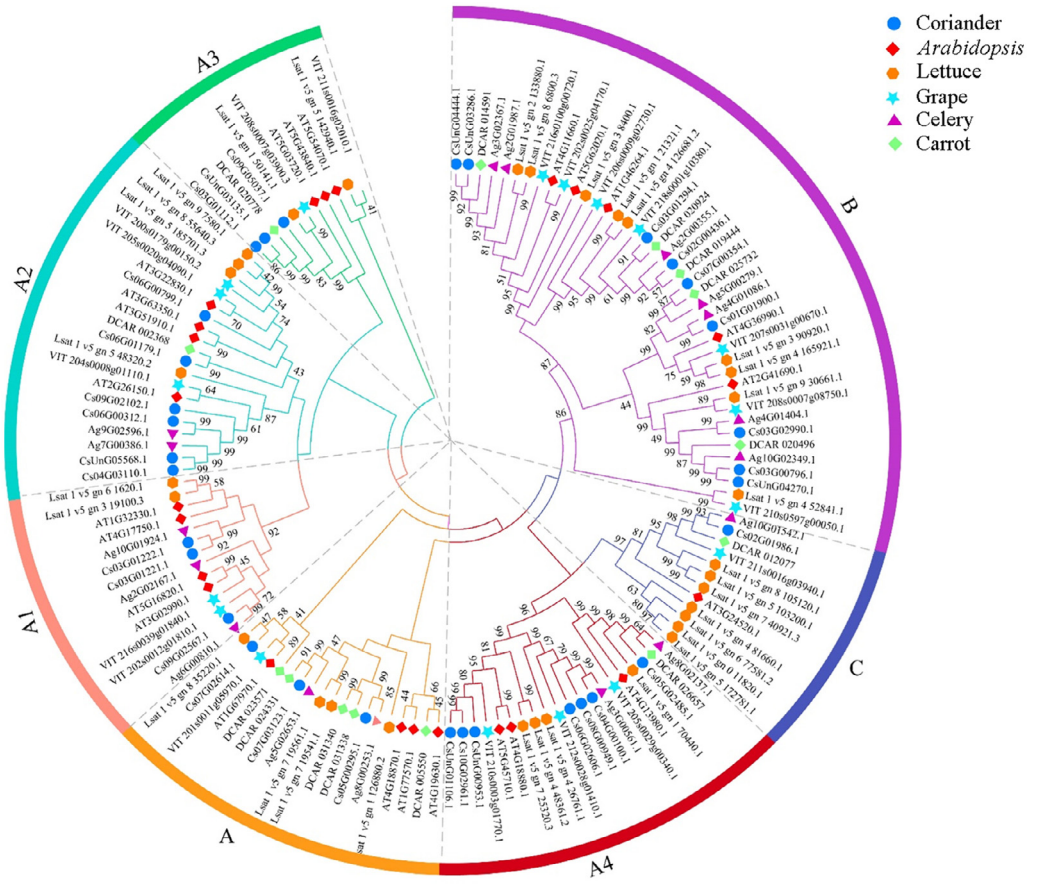

Fig. 1 Phylogenetic tree of the *Hsf* gene family in the three Apiaceae species (carrot, celery, and coriander) and lettuce, grape, and *Arabidopsis*

Phylogenetic tree topology was generated using MEGA X with the NJ method. Bootstrap values > 40% are shown.

Circos software (Krzywinski et al., 2009). The reconstructed *Hsf* family gene trees were compared with real species trees by lineage using Notung v2.9 software with the default parameters to conduct *Hsf* family gene duplication and loss estimation (Stolzer et al., 2012).

2.5. *Collinear analysis and duplication types of genes*

The collinear analysis of the *Hsf* family genes among celery, coriander, and carrot were conducted using MCScanX software (Wang et al., 2012). First, whole-genome protein sequences of these species were performed using the Blastp program with an e-value of 1×10^{-5}. Then, we searched the collinear blocks using MCScanX with the parameters: -k 50, -s 5, and -m 25. The duplication types of the genes were classified using a duplicate_gene_classifier program, which was incorporated in the MCScanX software. Finally, *Hsf* family genes located in the collinear blocks were extracted using Perl scripts.

2.6. *Ka/Ks calculation and divergence time estimation*

To estimate the divergence time of the genes, the sequences of orthologous *Hsf* family gene pairs in celery, carrot, and coriander were aligned using ClustalW. Then, the nonsynonymous rate (Ka), synonymous rate (Ks), and their ratio (Ka/Ks) between orthologous gene pairs were calculated based on their coding sequence alignments using the Nei-Gojobori method, which was implemented in the Ka/Ks_calculator program (Wang et al., 2010). Ks values of all of the orthologous gene pairs were used for divergence time estimation based on a neutral substitution rate of 5.2×10^{-9} substitutions per site per year (Song et al., 2020b).

2.7. *Selective pressure detection*

The likelihood ratio test of positive selection was used with maximum likelihood (ML) methods and codon substitution models. Alignment of the amino acid sequences of the *Hsf* family genes was conducted using ClustalW, and then translated into coding sequence (CDS) alignment. Based on previously reported methods to explore evolutionary traces of selective pressures (Mondragon-Palomino et al., 2002; Song et al., 2016, 2018), each branch of the phylogenetic tree was analyzed to infer the ratio of nonsynonymous to synonymous distances (ω) using CodeML implemented by PAML4.9 (Yang, 1997). A complete deletion method was employed for analyzing alignments with gaps. Sequences that contained > 40% of their length were eliminated. The varia-

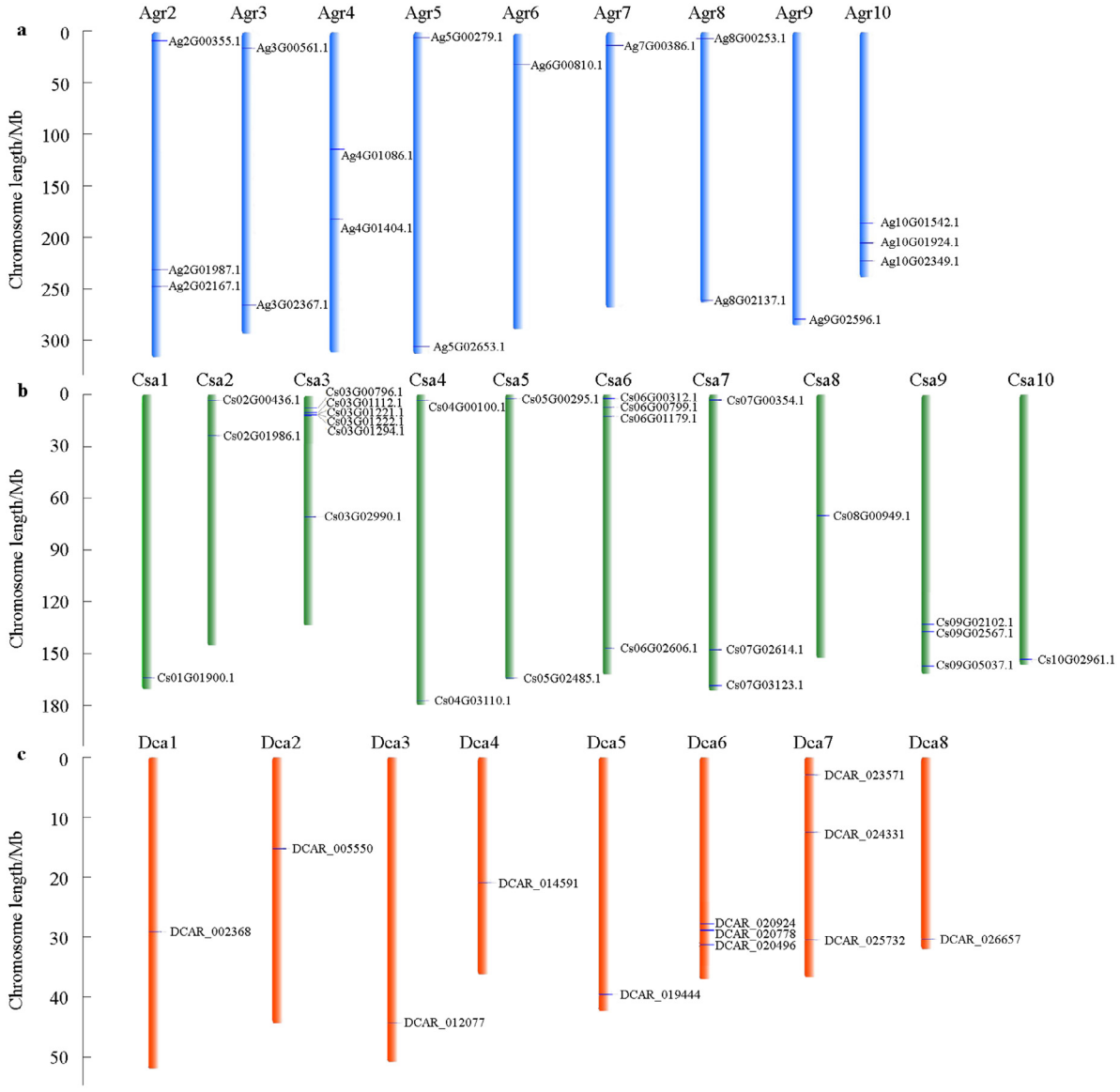

Fig. 2 Distribution of *Hsf* family genes on each chromosome in the three Apiaceae species
(a) celery; (b) coriander; (c) carrot.

tion among sites was calculated by employing a likelihood ratio test between the M0 and M1 and M7 and M8 models.

2.8. *Expression analysis*

To explore the *Hsf* family gene expression patterns, Illumina RNA-seq data with three replications were generated. These expression datasets contained three organs, including the root, petiole, and leaf in coriander and celery. The fragments per kilobase of exon model per million mapped (FPKM) values were used and $\log_2$-transformed. The expression patterns of the *Hsf* family genes in each organ were analyzed using Cluster v3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm). The hierarchical clustering heat maps were visualized using Tree View (https://sourceforge.net/projects/jtreeview/).

3. Results

3.1. *Phylogenetic and classification analyses*

In total, 17, 32, and 14 *Hsf* family genes were identified from the whole genomes of celery, coriander, and carrot (Table S1). Additionally, 24, 17, and 32 *Hsf* genes were identified in *Arabidopsis*, grape, and lettuce, respectively, which were used for the comparative analysis with the three Apiaceae species. To explore the evolutionary relationship and classification of the *Hsf* family genes, we constructed a phylogenetic tree using all of the amino acid sequences identified among these six species (Fig. 1). Results revealed that the *Hsf* family genes were divided into three groups, respectively named group A, B, and C. Group A further was divided into five subgroups according to their bootstrap values and phylogenetic relationships defined by *Arabidopsis* and grape.

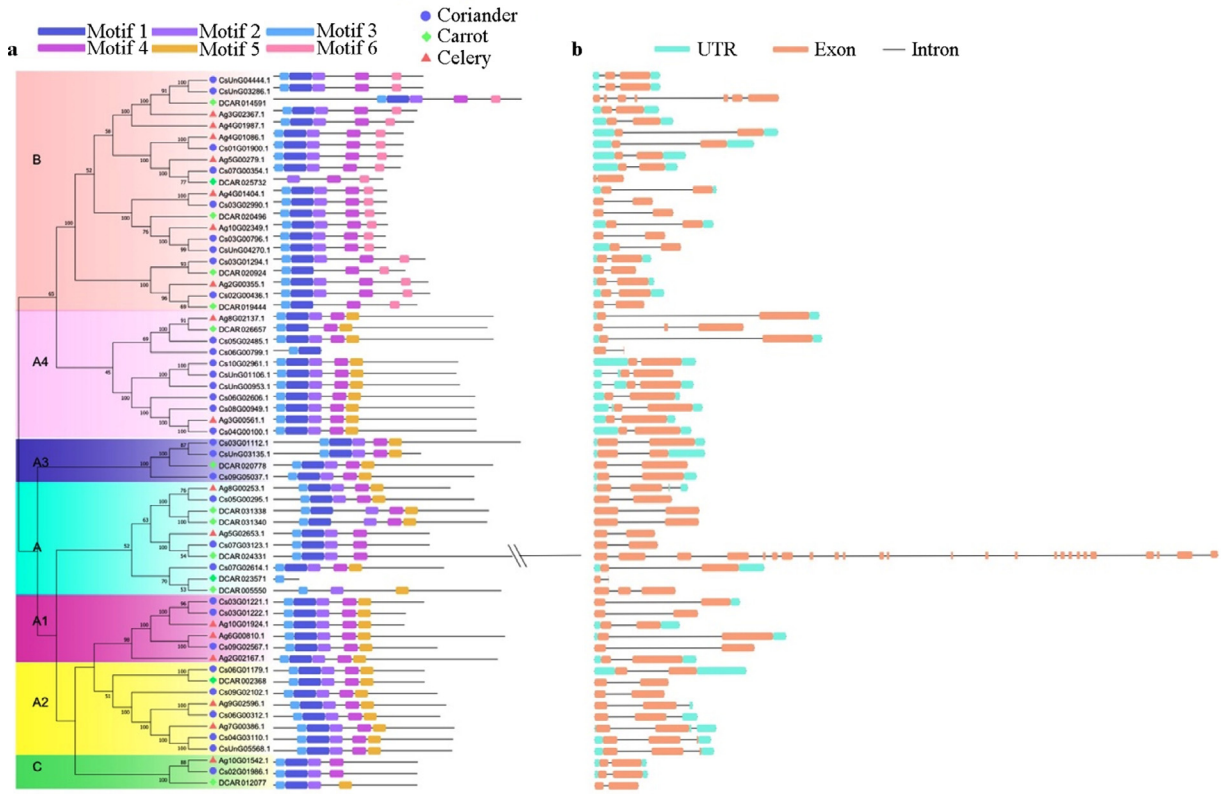

Fig. 3 The conserved motifs and gene structure of *Hsf* gene family in three Apiaceae species
(a) The conserved motifs of *Hsf* gene family; (b) The gene structure of *Hsf* gene family.

To gain further insights into the phylogenetic relationship among the *Hsf* family genes in the Apiaceae species, a phylogenetic tree was constructed only using the *Hsf* family genes from celery, coriander, and carrot (Fig. S1). The classifications of the phylogenetic tree showed that most *Hsf* genes (39, 61.9%) belonged to group A and its subgroups, whereas only three genes belonged to group C (Fig. S1). In the subgroups of group A, group A4 had the most genes (11), whereas group A3 only had four genes. Additionally, there were no carrot or celery genes in group A1 or A3, respectively.

3.2. *Chromosomal distribution of* Hsf *family genes in celery, coriander, and carrot*

To investigate the distribution of *Hsf* family genes on chromosomes, we performed an analysis of their locations on different chromosomes (Fig. 2). All of the 17 celery *Hsf* genes were mapped on nine corresponding chromosomes. Among the 32 coriander *Hsf* genes, 7 genes failed to be mapped on a chromosome, whereas 25 genes were mapped on 10 corresponding chromosomes. Among the 14 carrot *Hsf* genes, two genes, *DCAR_031338* and *DCAR_031340*, failed to map on a chromosome, whereas 12 genes were mapped on corresponding chromosomes. In celery, 17 genes were unevenly distributed on 9 chromosomes, except the Agr1 chromosome (Fig. 2, a). Chromosomes Agr2 and Agr10 had the most genes (three genes). In coriander, 25 genes were unevenly distributed on 10 chromosomes (Csa1–10) (Fig. 2, b). Chromosomes Csa1, Csa8 and Csa10 had only one gene, whereas chromosome Csa3 had the most genes (six genes). Moreover, five genes clustered at the end of Csa 3 may be due to gene duplication. In carrot, 12 *Hsf* genes were distributed on eight chromosomes (Dca1–8). Most genes were found on the Dca6 and Dca7 (three genes), while only one gene located on the each of other 6 chromosomes (Fig. 2, c).

3.3. Hsf *family gene structure and conserved motif analyses*

The gene structure and conserved motif analyses uncovered the conservative patterns of the *Hsf* family genes. Similar gene structures and conserved motifs were detected in the same group or subgroup. Here, the gene structure of 63 *Hsf* family genes from celery, coriander, and carrot are illustrated (Fig. 3). Almost all of the *Hsf* family genes had two exons, although there were some exceptions, such as the carrot gene, *DCAR_014591*, in group B, which had seven exons. Moreover, one carrot gene, *DCAR_024331*, in group A had 25 exons, whose length was much longer than other *Hsf* family genes. Thus, it was speculated that this phenomenon may be due to gene fusion or inaccurate predictions of some genes in the carrot genome.

Motif analyses were performed using MEME software, and six obvious motifs were selected (Fig. 3). For convenience, the motifs were respectively named motifs 1–6. Motif 3 was at the start of the genes, followed by motifs 1 and 2. In Apiaceae, most *Hsf* family genes had motifs 1–4. However, some motifs were lost in several *Hsf* genes, especially for carrot. For example, one carrot gene, *DCAR_025732*, lacked motifs 1 and 3, and two genes, *DCAR_020924* and *DCAR_019444*, lacked motif 2. Moreover, there

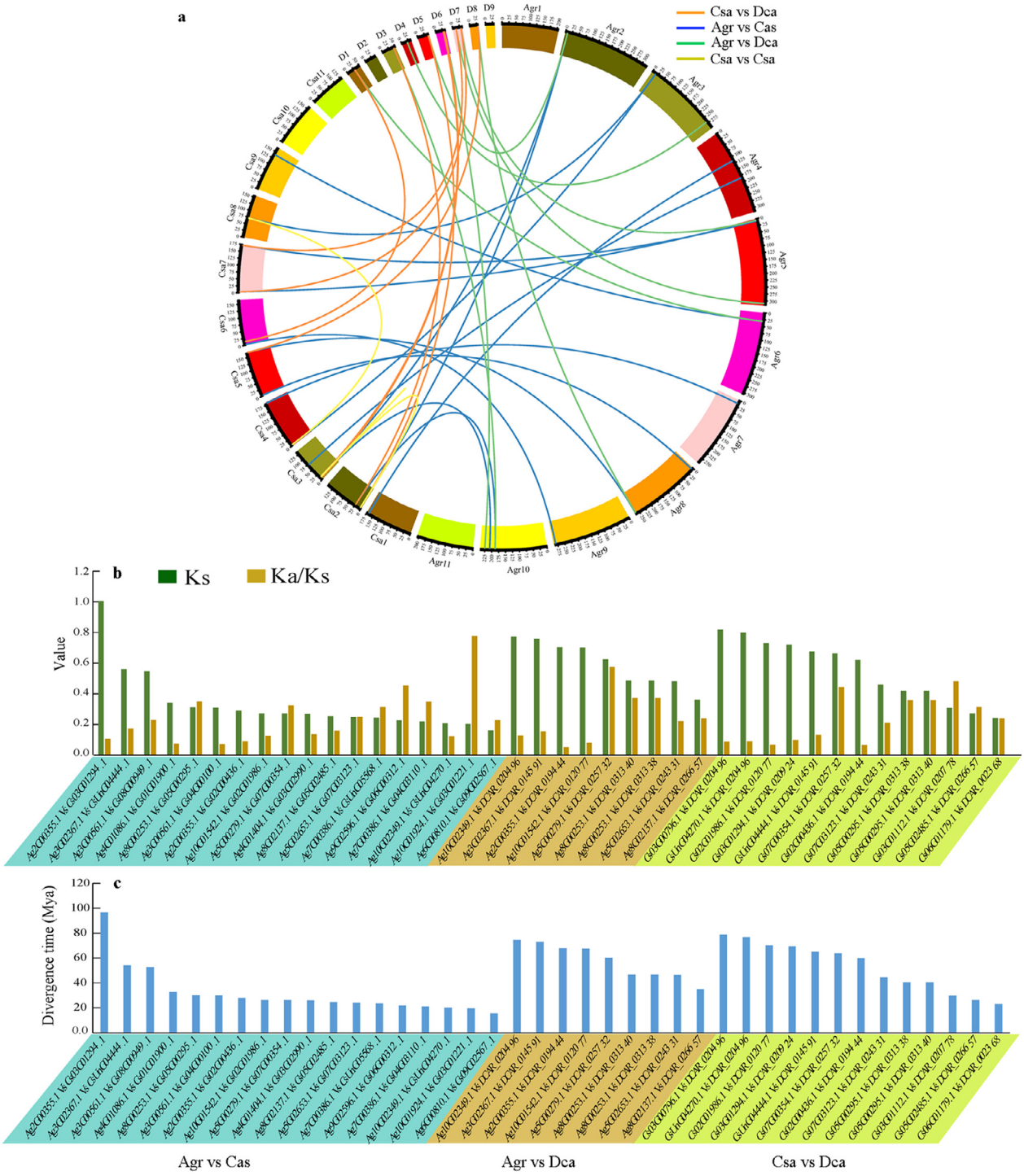

Fig. 4 Paralogous and orthologous analyses of the *Hsf* gene family

(a) Circle plot of paralogous and orthologous *Hsf* gene pairs among the three Apiaceae species. The number on the scale represents the physical location of each chromosome; (b) Ks and Ka/Ks values of orthologous *Hsf* gene pairs between any two of the three Apiaceae species; (c) Divergence time estimation of orthologous *Hsf* gene pairs between any two of the three Apiaceae species.

was only one motif (motif 3) in *DCAR_023571*, which indicated the carrot genome lost motifs throughout its evolution.

Interestingly, we found that all of the *Hsf* family genes in group B had motif 6, but this motif was not detected in other groups. Additionally, all of the genes in group B did not contain motif 5, which was detected in most genes of other groups. In group A4, most genes had motifs 1–5, except *Cs06G00799.1*, which lacked motifs 2, 4, and 5. Collectively, these results indicated that most motifs within the same group were highly similar and had more consistent position distributions within their gene structures.

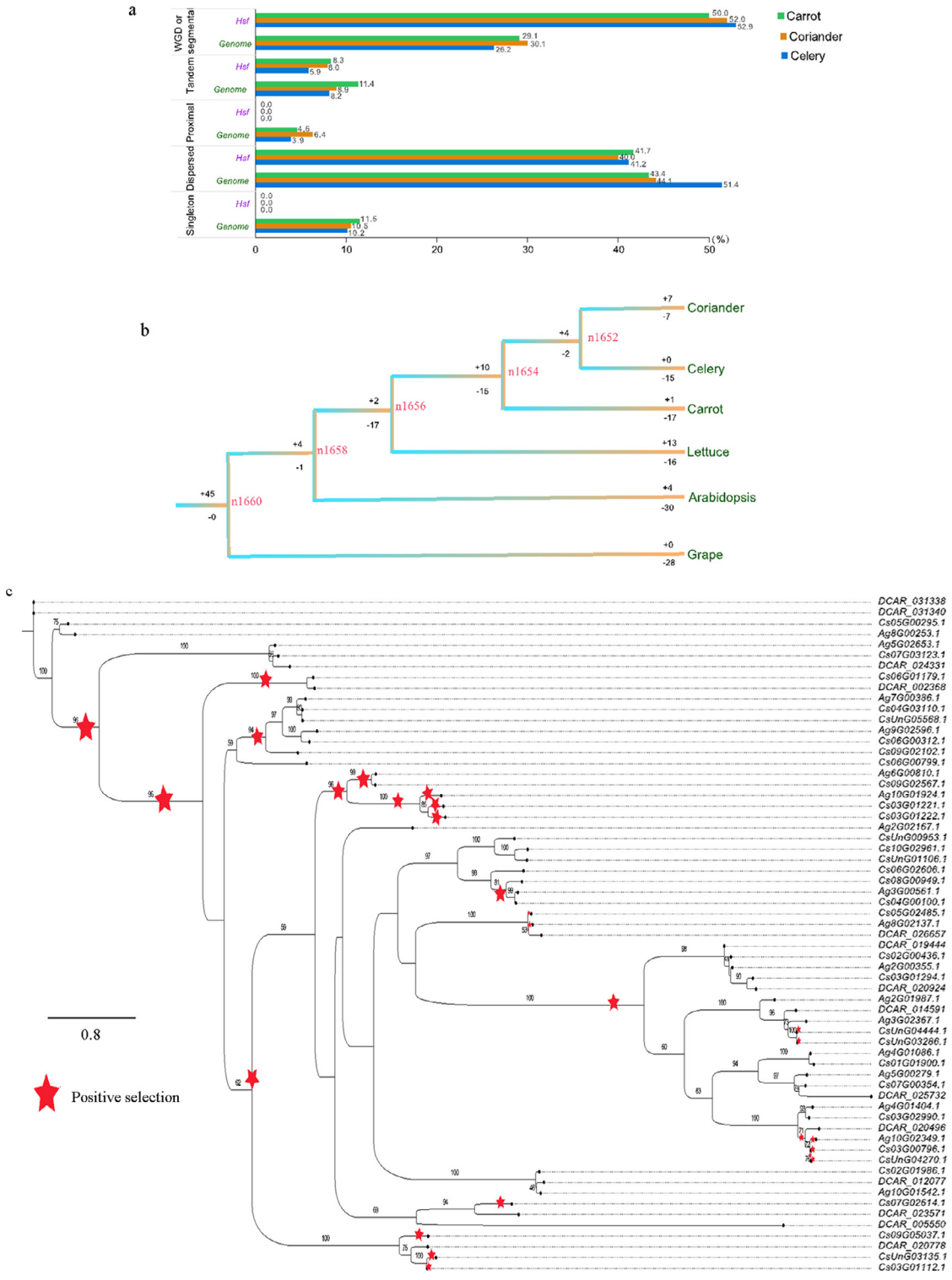

Fig. 5 Duplication, loss, and positive selection analyses of the *Hsf* gene family in three Apiaceae species

(a) Percentage of duplication types of the *Hsf* family genes and whole genome genes; (b) Duplication and loss analyses of the *Hsf* family genes, where "+" and "–" indicate duplication and loss, respectively, and the number after "+" and "–" represents the gene number; (c) Positive selection analysis of the *Hsf* family genes, where the red five-star represents positive selection branches. The maximum-likelihood (ML) phylogenetic tree was constructed using PhyML software. The numbers on each branch indicated the bootstrap values.

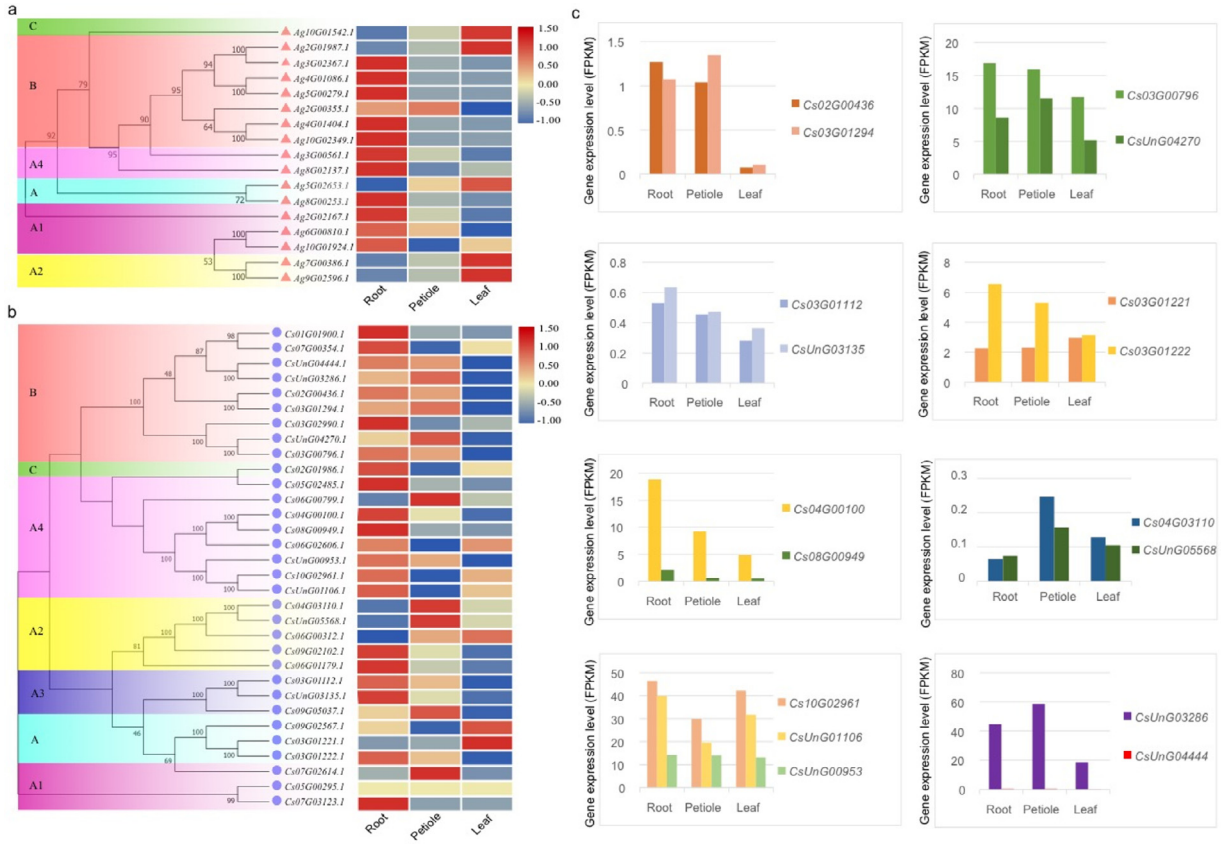

Fig. 6 Expression analysis of the *Hsf* family genes in three plant organs, including the root, petiole, and leaf

(a) Expression hierarchical clustering of the *Hsf* genes in celery based on RNA-seq data, where the expression values were calculated by reads per kilobase per million mapped reads (FPKM) and were $\log_2$-transformed; (b) Expression hierarchical clustering of the *Hsf* genes in coriander based on RNA-seq data; (c) Gene expression (FPKM) of the three organs was used for comparative analysis of the paralogous *Hsf* genes in coriander.

3.4. *Identification of orthologous and paralogous* Hsf *family genes*

To further explore the relationship of the *Hsf* family genes among celery, coriander, and carrot, orthologous and paralogous genes were identified (Fig. 4). There were 18 orthologous gene pairs between celery and coriander, 13 between coriander and carrot, and 10 between celery and carrot (Fig. 4, a; Table S2). The number of orthologous genes between celery and coriander was much greater than the other comparisons, indicating that the relationship between celery and coriander was closer than carrot and celery or coriander. Additionally, we identified 10 paralogous *Hsf* gene pairs in coriander, of which, only three were anchored on chromosomes (Figs. 4a and S2; Table S3). In carrot, there was only one paralogous gene pair, which was not anchored on a specific chromosome. Moreover, no paralogous gene pairs were detected in the celery genome.

Selected types of orthologous gene pairs were calculated according to nonsynonymous (Ka) and synonymous (Ks) substitutions (Fig. 4, b; Table 1). In total, we obtained the Ks, Ka, and Ka/Ks values, and selected the types of 41 orthologous gene pairs (Fig. 4, b; Table 1). Results revealed that most orthologous gene pairs (40/41) had Ka/Ks ratios < 1, indicating that purifying selection had acted upon these orthologous genes. One pair, *Ag6G00810.1* versus *DCAR_002368* could not be calculated Ks value, indicating that there might be more sequence divergence between these two genes.

Finally, we estimated the divergence time of orthologous *Hsf* gene pairs according to their synonymous substitution rates (Fig. 4, c; Table 1). Results revealed that the divergence time span was 15.48–96.63 million years for the orthologous gene pairs between celery and coriander, 34.70–74.34 million years for the orthologous gene pairs between celery and carrot, and 23.15–78.68 million years for the orthologous gene pairs between coriander and carrot.

3.5. Hsf *family gene evolutionary losses and duplications*

Gene duplication directly reflects gene family expansion. Here, we detected five types of gene duplication, including singleton, dispersed, proximal, tandem, and WGD/segmental. Among the five types, we found that WGD/segmental played a key role in *Hsf* gene family expansion of these three species (Fig. 5, a; Tables 2 and S4). The percentage of *Hsf* genes belonging to WGD/segmental was 52.9%, 52.0%, and 50.0% for celery, coriander, and carrot, respectively. There were no singleton or proximal duplication types detected in any of the three species. Additionally, >50% of the genes were detected in collinear blocks

Table 1 Ka/Ks calculations and divergence times of the orthologous *Hsf* gene pairs between celery (Agr), coriander (Csa), and carrot (Dca)

Species	Orthologous gene pairs	Ka	Ks	Ka/Ks	Purify	Divergence time (Mya)
Agr	*Ag2G00355.1 Vs Cs03G01294.1*	0.1072	1.0050	0.1067	Yes	96.63
vs	*Ag3G02367.1 Vs CsUnG04444.1*	0.0978	0.5611	0.1742	Yes	53.96
Csa	*Ag3G00561.1 Vs Cs08G00949.1*	0.1256	0.5472	0.2296	Yes	52.61
	Ag4G01086.1 Vs Cs01G01900.1	0.0253	0.3404	0.0745	Yes	32.73
	Ag8G00253.1 Vs Cs05G00295.1	0.1097	0.3129	0.3504	Yes	30.09
	Ag3G00561.1 Vs Cs04G00100.1	0.0222	0.3094	0.0716	Yes	29.75
	Ag2G00355.1 Vs Cs02G00436.1	0.0264	0.2905	0.0909	Yes	27.93
	Ag10G01542.1 Vs Cs02G01986.1	0.0344	0.2720	0.1264	Yes	26.15
	Ag5G00279.1 Vs Cs07G00354.1	0.0879	0.2718	0.3233	Yes	26.13
	Ag4G01404.1 Vs Cs03G02990.1	0.0369	0.2695	0.1369	Yes	25.92
	Ag8G02137.1 Vs Cs05G02485.1	0.0404	0.2535	0.1595	Yes	24.38
	Ag5G02653.1 Vs Cs07G03123.1	0.0623	0.2485	0.2507	Yes	23.89
	Ag7G00386.1 Vs CsUnG05568.1	0.0768	0.2433	0.3156	Yes	23.40
	Ag9G02596.1 Vs Cs06G00312.1	0.1028	0.2263	0.4543	Yes	21.76
	Ag7G00386.1 Vs Cs04G03110.1	0.0764	0.2189	0.3489	Yes	21.05
	Ag10G02349.1 Vs CsUnG04270.1	0.0254	0.2091	0.1213	Yes	20.11
	Ag10G01924.1 Vs Cs03G01221.1	0.1569	0.2022	0.7760	Yes	19.44
	Ag6G00810.1 Vs Cs09G02567.1	0.0368	0.1609	0.2287	Yes	15.48
Agr	*Ag10G02349.1 Vs DCAR_020496*	0.1000	0.7731	0.1293	Yes	74.34
vs	*Ag3G02367.1 Vs DCAR_014591*	0.1173	0.7602	0.1543	Yes	73.09
Dca	*Ag2G00355.1 Vs DCAR_019444*	0.0370	0.7049	0.0525	Yes	67.78
	Ag10G01542.1 Vs DCAR_012077	0.0560	0.7020	0.0798	Yes	67.50
	Ag5G00279.1 Vs DCAR_025732	0.3595	0.6247	0.5756	Yes	60.07
	Ag8G00253.1 Vs DCAR_031340	0.1813	0.4856	0.3733	Yes	46.69
	Ag8G00253.1 Vs DCAR_031338	0.1813	0.4856	0.3733	Yes	46.69
	Ag5G02653.1 Vs DCAR_024331	0.1069	0.4815	0.2219	Yes	46.29
	Ag8G02137.1 Vs DCAR_026657	0.0864	0.3609	0.2393	Yes	34.70
	Ag6G00810.1 Vs DCAR_002368	0.5834	NaN	NaN	NaN	NaN
Csa	*Cs03G00796.1 Vs DCAR_020496*	0.0724	0.8183	0.0884	Yes	78.68
vs	*CsUnG04270.1 Vs DCAR_020496*	0.0724	0.7985	0.0907	Yes	76.77
Dca	*Cs02G01986.1 Vs DCAR_012077*	0.0497	0.7304	0.0680	Yes	70.23
	Cs03G01294.1 Vs DCAR_020924	0.0702	0.7208	0.0974	Yes	69.31
	CsUnG04444.1 Vs DCAR_014591	0.0903	0.6743	0.1339	Yes	64.83
	Cs07G00354.1 Vs DCAR_025732	0.2944	0.6642	0.4432	Yes	63.86
	Cs02G00436.1 Vs DCAR_019444	0.0416	0.6219	0.0669	Yes	59.80
	Cs07G03123.1 Vs DCAR_024331	0.0977	0.4598	0.2125	Yes	44.21
	Cs05G00295.1 Vs DCAR_031338	0.1511	0.4202	0.3595	Yes	40.41
	Cs05G00295.1 Vs DCAR_031340	0.1511	0.4202	0.3595	Yes	40.41
	Cs03G01112.1 Vs DCAR_020778	0.1482	0.3071	0.4827	Yes	29.53
	Cs05G02485.1 Vs DCAR_026657	0.0860	0.2725	0.3157	Yes	26.20
	Cs06G01179.1 Vs DCAR_002368	0.0576	0.2407	0.2394	Yes	23.15

by detecting gene collinearity within a genome, which was considerably higher than the average level of the whole genome (Table S5).

We further performed gene loss and duplication analyses using Notung software. Results revealed the gene loss and duplication numbers of the *Hsf* genes in celery, coriander, carrot, lettuce, *Arabidopsis*, and grape based on the reconstructed phylogenetic tree (Figs. 5b and S3). In the lineages of the common ancestor of celery and coriander, 2 genes were lost and 4 were duplicated. In the lineages of the common ancestor of celery, coriander, and carrot, 15 genes were lost and 10 were duplicated. We found that the *Hsf* gene family underwent expansion in coriander when compared to celery and carrot. Additionally, there were more *Hsf* gene duplications in coriander (+7) than celery (+0) and carrot (+1), and fewer *Hsf* genes losses in coriander (–7) than celery (–15) and carrot (–17).

3.6. Positive selection analysis

Positive selection analyses are important for understanding gene functions during evolution. To investigate natural selection that acted upon celery, coriander, and carrot, a phylogenetic tree was constructed using PhyML and a positive selection analysis was conducted on all of the *Hsf* family genes from these three species (Fig. 5, c). In total, 27 significant nonsynonymous versus synonymous substitutions were detected, showing strong positive selection in most branches of the phylogenetic tree.

3.7. Expression analysis

To explore the expression patterns of the *Hsf* family genes in celery and coriander, we conducted an expression analysis of the *Hsf* family genes in three plant organs, including the root, petiole, and leaf. In celery, most genes (11, 64.71%) had higher expression levels in the root than the other organs (Fig. 6, a; Table S6). However, only one gene (*Ag2G00355*) had higher expression levels in the petiole than the root or leaf. In groups A2 and C, all of the three genes had higher expression levels in the leaf than the other organs. In groups A1 and A4, all of the five genes had higher expression levels in the root than the petiole or leaf. Phylogenetic tree of the celery *Hsf* genes was constructed to show their ex-

Table 2 Identification of duplicated gene types in the *Hsf* genes and the genome of celery, coriander, and carrot

Species	Singleton		Dispersed		Proximal		Tandem		WGD or segmental		
	Genome	Hsf	Genome	Hsf	Genome	Hsf	Genome	Hsf	Genome	Hsf	Percentage (%)
Celery	3 028	0	15 258	7	1 167	0	2 426	1	7 787	9	52.9
Coriander	3 577	0	14 963	10	2 161	0	3 032	2	10 200	13	52.0
Carrot	3 543	0	13 378	5	1 428	0	3 501	1	8 974	6	50.0

pression patterns in different subgroups. In group A2, two genes (*Ag7G00386* and *Ag9G02596*) had similar expression patterns in all the organs, but with higher expression levels in the leaf (Fig. 6, a).

Similarly, most genes (20, 62.50%) had higher expression levels in the root than the other organs in coriander (Fig. 6, b; Table S6). Among all of the coriander *Hsf* family genes, only one gene (*Cs05G00295*) was not detected in the root, petiole, or leaf, indicating that it may not play a role in the development of these three organs in coriander.

In coriander, expression levels were greatly similar between *Cs02G00436* and *Cs03G01294* in all the organs, and they belonged to paralogous genes (Fig. 6, c). We also detected the expression levels of other paralogous genes and found that all of the paralogous genes were distributed on the same branches in the phylogenetic tree. Furthermore, the expression patterns of most paralogous genes exhibited similar trends in the three organs.

4. Discussion

4.1. *Systematic and comprehensive analyses of the* Hsf *gene family*

Although the *Hsf* gene family has been reported in many plants, it had not been reported in Apiaceae until now. Celery, coriander, and carrot are important, globally grown vegetables and with their genomes completed, which provided us the opportunity to study the *Hsf* gene family in these three Apiaceae species (Iorizzo et al., 2016; Song et al., 2020b). Here, we identified 17, 32, and 14 *Hsf* genes in the celery, coriander, and carrot genomes, respectively. Furthermore, we identified 24, 17, and 32 *Hsf* genes in *Arabidopsis*, grape, and lettuce, respectively, for a comparative analysis with the three Apiaceae species.

This study systematically compared and analyzed the *Hsf* gene family. We constructed a phylogenetic tree of celery, coriander, carrot, *Arabidopsis*, lettuce, and grape. We found that conserved motifs of genes within the same group were highly similar and had more consistent position distributions in gene structure. There were no paralogous gene pairs in celery, and only one pair in carrot, which was far less than coriander (10). Based on the collinearity analysis, we found that WGD/segmental played a very important role in the *Hsf* gene family's expansion.

As the first study on the *Hsf* gene family in Apiaceae, our research will serve as an abundant data resource for future comparative and functional genomic studies on the *Hsf* gene family. We comprehensively identified and analyzed the gene structure, conserved motifs, chromosomal distribution, gene duplication and losses, orthologous and paralogous genes, and gene collinearity of *Hsf* genes in three Apiaceae species, and further conducted the expression pattern analysis of celery and coriander. A total of 136 *Hsf* genes were identified in these species, of which, 63 genes were identified in the Apiaceae species.

4.2. *Rich resource for future interaction studies on* Hsf *genes and Hsps in Apiaceae*

Hsf is specifically recognized as heat shock elements (HSEs) of the heat shock protein (Hsp) promoter (Nguyen et al., 2016). The N-terminal of *Hsf* was a highly conserved DBD, whose hydrophobic core forms a helix-turn-helix conformation (Nover et al., 2001). The accumulation of Hsps under the control of *Hsf* genes played a key role in plant heat stress responses (Kotak et al., 2007). Plant Hsps are grouped into five families based on their approximate molecular weights, including Hsp100, Hsp90, Hsp70, Hsp60, and small Hsp (Aevermann and Waters, 2008; Waters et al., 2008). Here, we identified 63 *Hsf* family genes in three Apiaceae species, celery, coriander, and carrot. The findings of this study provide rich resources for future interaction studies on *Hsf* family genes and Hsps in Apiaceae species.

Until now, only a few gene family studies have been characterized in Apiaceae species. Thus, this is the first comprehensive and systematic analysis of the *Hsf* gene family in three Apiaceae species. Identification of these transcription factors will assist in clarifying the heat tolerance ability and regulatory molecular genetic basis for Apiaceae genetic improvement, as well as provides a functional gene resource for transgenic research and gene editing. Moreover, this study will serve as a useful resource for future studies on the evolution and function of *Hsf* genes in Apiaceae. It may also facilitate our understanding of the effects of duplications and losses during Apiaceae evolution or other plants.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31801856), Hebei Province Higher Education Youth Talents Program (Grant No. BJ2018016), the China Postdoctoral Science Foundation (Grant No. 2020M673188), and the innovation and entrepreneurship training program for college students of North China University of Science and Technology (Grant No. X2019256).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.hpj.2020.08.005.

REFERENCES

Aevermann, B.D., Waters, E.R., 2008. A comparative genomic analysis of the small heat shock proteins in *Caenorhabditis elegans* and *briggsae*. Genetica, 133: 307–319.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. Meme suite: tools for motif discovery and searching. Nucleic Acids Res, 37: W202–W208.

Castañares, J.L., Bouzo, C.A., 2019. Effect of exogenous melatonin on seed germination and seedling growth in melon *(Cucumis melo L.)* under salt stress. Hortic Plant J, 5: 79–87.

Chung, E., Kim, K.M., Lee, J.H., 2013. Genome-wide analysis and molecular characterization of heat shock transcription factor family in *Glycine max*. J Genet Genom, 40: 127–135.

Döring, P., Treuter, E., Kistner, C., Lyck, R., Chen, A., Nover, L., 2000. The role of aha motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. Plant Cell, 12: 265–278.

Gomez-Pastor, R., Burchfiel, E.T., Thiele, D.J., 2018. Regulation of heat shock transcription factors and their roles in physiology and disease. Nat Rev Mol Cell Biol, 19: 4–19.

Guo, J.K., Wu, J., Ji, Q., Wang, C., Luo, L., Yuan, Y., Wang, Y.H., Wang, J., 2008. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. J Genet Genom, 35: 105–118.

Guo, M., Lu, J.P., Zhai, Y.F., Chai, W.G., Gong, Z.H., Lu, M.H., 2015. Genome-wide analysis, expression profile of heat shock factor gene family (CaHsfs) and characterisation of CaHsfA2 in pepper (*Capsicum annuum* L.). BMC Plant Biol, 15: 151.

Hu, B., Jin, J.P., Guo, A.Y., Zhang, H., Luo, J.C., Gao, G., 2015. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics, 31: 1296–1297.

Huang, X.Y., Tao, P., Li, B.Y., Wang, W.H., Yue, Z.C., Lei, J.L., Zhong, X.M., 2015. Genome-wide identification, classification, and analysis of heat shock transcription factor family in Chinese cabbage (*Brassica rapa pekinensis*). Genet Mol Res, 14: 2189–2204.

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P., Yildiz, M., Macko-Podgorni, A., Moranska, E., Grzebelus, E., Grzebelus, D., Ashrafi, H., Zheng, Z., Cheng, S., Spooner, D., Van Deynze, A., Simon, P., 2016. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat Genet, 48: 657–666.

Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyere, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Gaspero, D., Dumas, G., Felice, V., Paillard, N., Juman, S., Moroldo, I., Scalabrin, M., Canaguier, S., Le Clainche, A., Malacrida, I., Durand, G., Pesole, E., Laucou, G., Chatelet, V., Merdinoglu, P., Delledonne, D., Pezzotti, M., Lecharny, M., Scarpelli, A., Artiguenave, C., Pe, F., Valle, M.E., Morgante, G., Caboche, M., Adam-Blondon, M., Weissenbach, A.F., Quetier, J., Wincker, F., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449: 463–467.

Kotak, S., Larkindale, J., Lee, U., von Koskull-Doring, P., Vierling, E., Scharf, K.D., 2007. Complexity of the heat stress response in plants. Curr Opin Plant Biol, 10: 310–316.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S., Marra, M., 2009. Circos: an information aesthetic for comparative genomics. Genome Res, 19: 1639–1645.

Kumar, R.R., Goswami, S., Singh, K., Dubey, K., Rai, G.K., Singh, B., Singh, S., Grover, M., Mishra, D., Kumar, S., Bakshi, S., Rai, A., Pathak, H., Chinnusamy, V., Praveen, S., 2018a. Characterization of novel heat-responsive transcription factor (TaHSFA6e) gene involved in regulation of heat shock proteins (HSPs) - a key member of heat stress-tolerance network of wheat. J Biotechnol, 279: 1–12.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018b. Mega X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol, 35: 1547–1549.

Lee, H.O., Joh, H.J., Kim, K., Lee, S.C., Kim, N.H., Park, J.Y., Park, H.S., Park, M.S., Kim, S., Kwak, M., Kim, K.Y., Lee, W.K., Yang, T.J., 2019. Dynamic chloroplast genome rearrangement and DNA barcoding for three Apiaceae species known as the medicinal herb "bang-poong". Int J Mol Sci, 20: 2196.

Letunic, I., Doerks, T., Bork, P., 2012. Smart 7: recent updates to the protein domain annotation resource. Nucleic Acids Res, 40: D302–D305.

Li, L., Stoeckert, C.J., Roos, D.S., 2003. Orthomcl: identification of ortholog groups for eukaryotic genomes. Genome Res, 13: 2178–2189.

Li, W., Wan, X.L., Yu, J.Y., Wang, K.L., Zhang, J., 2019. Genome-wide identification, classification, and expression analysis of the *Hsf* gene family in carnation (*Dianthus caryophyllus*). Int J Mol Sci, 20: 5233.

Liu, B.B., Hu, J.J., Zhang, J., 2019a. Evolutionary divergence of duplicated *Hsf* genes in *Populus*. Cells, 8: 438.

Liu, G.T., Chai, F.M., Wang, Y., Jiang, J.Z., Duan, W., Wang, Y.T., Wang, F.F., Li, S.H., Wang, L.J., 2018. Genome-wide identification and classification of Hsf family in grape, and their transcriptional analysis under heat acclimation and heat stress. Hortic Plant J, 4: 133–143.

Liu, M.Y., Huang, Q., Sun, W.J., Ma, Z.T., Huang, L., Wu, Q., Tang, Z.Z., Bu, T.L., Li, C.L., Chen, H., 2019b. Genome-wide investigation of the heat shock transcription factor (*hsf*) gene family in tartary buckwheat (*Fagopyrum tataricum*). BMC Genom, 20: 871.

Liu, X., Liu, J.N., Bi, J.F., Cao, F., Ding, Y.Y., Peng, J., 2019c. Effects of high pressure homogenization on physical stability and carotenoid degradation kinetics of carrot beverage during storage. J Food Eng, 263: 63–69.

Lohani, N., Golicz, A.A., Singh, M.B., Bhalla, P.L., 2019. Genome-wide analysis of the Hsf gene family in *Brassica oleracea* and a comparative analysis of the Hsf gene family in *B. oleracea*, *B. rapa* and *B. napus*. Funct Integr Genom, 19: 515–531.

Maljaei, M., Moosavian, S., Mirmosayyeb, O., Rouhani, M., Namjoo, I., Bahreini, A., 2019. Effect of celery extract on thyroid function; is herbal therapy safe in obesity? Int J Prev Med, 10: 55.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., Deweese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Bryant, S.H., 2009. CDD: specific functional annotation with the conserved domain database. Nucleic Acids Res, 37: D205–D210.

Mondragon-Palomino, M., Meyers, B., Michelmore, R., Gaut, B., 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. Genome Res, 12: 1305–1315.

Nguyen, A., Gotelli, N., Cahan, S., 2016. The evolution of heat shock protein sequences, *cis*-regulatory elements, and expression profiles in the eusocial Hymenoptera. BMC Evol Biol, 16: 15.

Nover, L., Bharti, K., Doring, P., Mishra, S., Ganguli, A., Scharf, K., 2001. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones, 6: 177–189.

Ohama, N., Kusakabe, K., Mizoi, J., Zhao, H., Kidokoro, S., Koizumi, S., Takahashi, F., Ishida, T., Yanagisawa, S., Shinozaki, K., Yamaguchi-Shinozaki, K., 2016. The transcriptional cascade in the heat stress response of *Arabidopsis* is strictly regulated at the level of transcription factor expression. Plant Cell, 28: 181–201.

Pirkkala, L., Nykanen, P., Sistonen, L., 2001. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. FASEB J, 15: 1118–1131.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res, 40: D290–D301.

Qi, J.G., Sun, S.M., Yang, L., Li, M.J., Ma, F.W., Zou, Y.J., 2019. Potassium uptake and transport in apple roots under drought stress. Hortic Plant J, 5: 10–16.

Que, F., Hou, X.L., Wang, G.L., Xu, Z.S., Tan, G.F., Li, T., Wang, Y.H., Khadr, A., Xiong, A.S., 2019. Advances in research on the carrot,

an important root vegetable in the Apiaceae family. Hortic Res, 6: 69.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D.O., Truco, M.J., Xia, R., Zhu, S., Xu, C., Xu, H., Xu, X., Cox, K., Korf, I., Meyers, B.C., Michelmore, R.W., 2017. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nat Commun, 8: 14953.

Scharf, K., Berberich, T., Ebersberger, I., Nover, L., 2012. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. Biochim Biophys Acta, 1819: 104–119.

Sojic, B., Pavlic, B., Ikonic, P., Tomovic, V., Ikonic, B., Zekovic, Z., Kocic-Tanackov, S., Jokanovic, M., Skaljac, S., Ivic, M., 2019. Coriander essential oil as natural food additive improves quality and safety of cooked pork sausages with different nitrite levels. Meat Sci, 157, 107879.

Song, X.M., Liu, G.F., Duan, W.K., Liu, T.K., Huang, Z.N., Ren, J., Li, Y., Hou, X.L., 2014. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. Mol Genet Genom, 289: 541–551.

Song, X.M., Ma, X., Li, C.J., Hu, J.J., Yang, Q.H., Wang, T., Wang, L., Wang, J.P., Guo, D., Ge, W.N., Wang, Z.Y., Li, M.M., Wang, Q.M., Ren, T.Z., Feng, S.Y., Wang, L.X., Zhang, W.M., Wang, X.Y., 2018. Comprehensive analyses of the *BES1* gene family in *Brassica napus* and examination of their evolutionary pattern in representative species. BMC Genom, 19: 346.

Song, X.M., Nie, F.L., Chen, W., Ma, X., Gong, K., Yang, Q.H., Wang, J.P., Li, N., Sun, P.C., Pei, Q.Y., Yu, T., Hu, J.J., Li, X.Y., Wu, T., Feng, S.Y., Li, X.Q., Wang, X.Y., 2020a. Coriander genomics database: a genomic, transcriptomic, and metabolic database for coriander. Hortic Res, 7: 55.

Song, X.M., Wang, J.P., Li, N., Yu, J.Q., Meng, F.B., Wei, C.D., Liu, C., Chen, W., Nie, F.L., Zhang, Z.K., Gong, K., Li, X.Y., Hu, J.J., Yang, Q.H., Li, Y.X., Li, C.J., Feng, S.Y., Guo, H., Yuan, J.Q., Pei, Q.Y., Yu, T., Kang, X., Zhao, W., Lei, T.Y., Sun, P.C., Wang, L., Ge, W.N., Guo, D., Duan, X.Q., Shen, S.Q., Cui, C.L., Yu, Y., Xie, Y.Q., Zhang, J., Hou, Y., Wang, J.Y., Wang, J.Y., Li, X.Q., Paterson, A.H., Wang, X.Y., 2020b. Deciphering the high quality genome sequence of coriander that causes controversial feelings. Plant Biotechnol J, 18: 1444–1456.

Song, X.M., Wang, J.P., Ma, X., Li, Y.X., Lei, T.Y., Wang, L., Ge, W.N., Guo, D., Wang, Z.Y., Li, C.J., Zhao, J.J., Wang, X.Y., 2016. Origination, expansion, evolutionary trajectory, and expression bias of *AP2/ERF* superfamily in *Brassica napus*. Front Plant Sci, 7: 1186.

Sorger, P., Pelham, H., 1988. Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. Cell, 54: 855–864.

Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., Durand, D., 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics, 28: i409–i415.

Wan, X.L., Yang, J., Guo, C., Bao, M.Z., Zhang, J.W., 2019. Genome-wide identification and classification of the *Hsf* and *sHsp* gene families in *Prunus mume*, and transcriptional analysis under heat stress. Peer J, 7: e7312.

Wang, D.P., Zhang, Y.B., Zhang, Z., Zhu, J., Yu, J., 2010. Kaks_calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinform, 8: 77–80.

Wang, Y.P., Tang, H.B., Debarry, J.D., Tan, X., Li, J.P., Wang, X.Y., Lee, T.H., Jin, H.Z., Marler, B., Guo, H., Kissinger, J.C., Paterson, A.H., 2012. Mcscanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res, 40: e49.

Waters, E., Aevermann, B., Sanders-Reed, Z., 2008. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. Cell Stress Chaperones, 13: 127–142.

Yan, Y.Y., Wang, S.S., Wei, M., Gong, B., Shi, Q.H., 2018. Effect of different rootstocks on the salt stress tolerance in watermelon seedlings. Hortic Plant J, 4: 239–249.

Yang, Z.H., 1997. Paml: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci, 13: 555–556.

Yin, L., Liu, J.X.£¬, Chen, L.Z.£¬, Shen, D., Duan, A.Q., Feng, K., Xu, Z.S., Xiong, A.S., 2020. Cloning and expression profiles analysis of *AgHAT4* gene in Apium graveolens. Acta Horticulturae Sinica, 47: 143–152.

Yu, X.Y., Yao, Y., Hong, Y.H., Hou, P.Y., Li, C.X., Xia, Z.Q., Geng, M.T., Chen, Y.H., 2019. Differential expression of the Hsf family in cassava under biotic and abiotic stresses. Genome, 62: 563–569.

Zafar, S., Hussain, M., Raza, M., Ahmed, H., Ahmad, I., Sadia, B., Atif, R., 2016. Genome wide analysis of heat shock transcription factor (Hsf) family in chickpea and its comparison with *Arabidopsis*. Plant Omics, 9: 136–141.

Zhou, M., Zheng, S.G., Liu, R., Lu, J., Lu, L., Zhang, C.H., Liu, Z.H., Luo, C.P., Zhang, L., Yant, L., Wu, Y., 2019. Genome-wide identification, phylogenetic and expression analysis of the heat shock transcription factor family in bread wheat (*Triticum aestivum* L.). BMC Genom, 20: 505.

Zhu, X.Y., Huang, C.Q., Zhang, L., Liu, H.F., Yu, J.H., Hu, Z.Y., Hua, W., 2017. Systematic analysis of *Hsf* family genes in the *Brassica napus* genome reveals novel responses to heat, drought and high CO_2 stresses. Front Plant Sci, 8: 1174.

Zimmermann, L., Stephens, A., Nam, S., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A., Alva, V., 2018. A completely reimplemented mpi bioinformatics toolkit with a new hhpred server at its core. J Mol Biol, 430: 2237–2243.