9-2022

Polyploidy events shaped the expansion of transcription factors in Cucurbitaceae and exploitation of genes for tendril development

Yu Zhang

Yingchao Zhang

Bing Li

Xiao Tan

Changping Zhu

See next page for additional authors

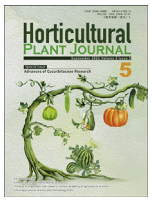
Authors

Yu Zhang, Yingchao Zhang, Bing Li, Xiao Tan, Changping Zhu, Tong Wu, Shuyan Feng, Qihang Yang, Shaoqin Shen, Tong Yu, Zhuo Liu, and Xiaoming Song

September 2022. Horticultural Plant Journal, 8 (5): 562–574

Horticultural Plant Journal

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



Polyploidy events shaped the expansion of transcription factors in Cucurbitaceae and exploitation of genes for tendril development

Yu Zhang [a,1], Yingchao Zhang [a,1], Bing Li [b,1], Xiao Tan [a], Changping Zhu [a], Tong Wu [a], Shuyan Feng [a], Qihang Yang [a], Shaoqin Shen [a], Tong Yu [a], Zhuo Liu [a], and Xiaoming Song [a,c,d,*]

[a] College of Life Sciences/Center for Genomics and Bio-computing, North China University of Science and Technology, Tangshan, Hebei 063210, China
[b] Institute of Cash Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, Hebei 050051, China
[c] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[d] Food Science and Technology Department, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

Received 28 January 2022; Received in revised form 12 April 2022; Accepted 24 June 2022

Available online 21 July 2022

A B S T R A C T

Cucurbitaceae is one of the most important plant families distributed worldwide. Transcription factors (TFs) regulate plant growth at the transcription level. Here, we performed a systematic analysis of 42 641 TFs from 63 families in 14 Cucurbitaceae and 10 non-cucurbit species. Whole-genome duplication (WGD) was the dominant event type in almost all Cucurbitaceae plants. The TF families were divided into 1 210 orthogroups (OGs), of which, 112 were unique to Cucurbitaceae. Although the loss of several gene families was detected in Cucurbitaceae, the gene families expanded in five species that experienced a WGD event comparing with grape. Our findings revealed that the recent WGD events that had occurred in Cucurbitaceae played important roles in the expansion of most TF families. The functional enrichment analysis of the genes that significantly expanded or contracted uncovered five gene families, *AUX/IAA*, *NAC*, *NBS*, *HB*, and *NF-YB*. Finally, we conducted a comprehensive analysis of the *TCP* gene family and identified 16 tendril-related (*TEN*) genes in 11 Cucurbitaceae species. Interestingly, the characteristic sequence changed from CNNFYFP to CNNFYLP in the *TEN* gene (*Bhi06M000087*) of *Benincasa hispida*. Furthermore, we identified a new characteristic sequence, YNN, which could be used for *TEN* gene exploitation in Cucurbitaceae. In conclusion, this study will serve as a reference for studying the relationship between gene family evolution and genome duplication. Moreover, it will provide rich genetic resources for functional Cucurbitaceae studies in the future.

Keywords: Cucurbitaceae; Transcription factors (TFs); Whole-genome duplication (WGD); Expansion and contraction; *TCP* gene family; Tendril-related genes (*TEN*)

1. Introduction

Cucurbitaceae is the fourth largest economic plant family in the world and comprises 115 genera containing nearly 1 000 species (Schaefer et al., 2009). Most of these plants are annual or perennial herbaceous or woody vines and are widely distributed in tropical and subtropical regions (Zhang et al., 2018; Lu et al., 2020). Cucurbitaceae plants are known for their important

[1] The authors contributed equally to the work.
* Corresponding author.
E-mail address: songxm@ncst.edu.cn
Peer review under responsibility of Chinese Society of Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS)

https://doi.org/10.1016/j.hpj.2022.07.004

economic value and are an important edible plant (Gao et al., 2020; Qi et al., 2020; Yang et al., 2021). This family comprises many vegetables and fruits, including *Cucurbita maxima* (pumpkin), *Cucurbita pepo* (squash), *Citrullus lanatus* (watermelon), *Cucumis sativus* (cucumber), *Cucurbita moschata* (gourd), and *Cucumis melo* (melon) (Chomicki et al., 2020; Zhou et al., 2020; Liu et al., 2021b; Yang et al., 2021). More than 30 species with local commercial value are widely planted in their native areas. The fruits and seeds can also be consumed and have rich medicinal value (Zhang et al., 2018; Qiao et al., 2021; Mukherjee et al., 2022). These plants have many pharmacological effects, such as anti-cancer, anti-oxidation, antithrombotic, and antibacterial effects (Dinan et al., 1997; Wu et al., 2002; Tsai et al., 2010; Xie et al., 2010; Karrar et al., 2019). Currently, the genomes of at least 14 Cucurbitaceae plants have been released (Table S1). With the development of sequencing technology and bioinformatics, the power of comparative genomics is becoming increasingly evident. Key genes involved in the domestication and improvement of Cucurbitaceae have been gradually found, thereby providing new impetus for studying Cucurbitaceae genomics.

Almost all major biological processes are regulated by transcription factors (TFs) (Ahmad et al., 2021; Li et al., 2021). TFs play important roles in plant evolution, growth, pathogen defense, and responses to biological and abiotic stress (Smith and Bryant, 1975; Rushton et al., 2010; Xie et al., 2019b). TFs regulate gene expression by binding to the *cis*-regulatory elements of target genes through the DNA-binding domain (DBD) (Velthuijs et al., 2021). Proteins encoded by similar TF families have a common DNA binding domain, and >60 TF families have been identified in plants. There are many TF families in almost all eukaryotes, but the size of these TF families varies greatly among species. Plant TF families are usually larger than those in animals, and higher plant TF families are larger than those in lower plants (Moharana and Venancio, 2020). The large TF family size in higher plants is usually related to polyploidization events (Lehti-Shiu et al., 2017).

Polyploidization is widely recognized as an important force in angiosperm evolution (Soltis et al., 2014). Polyploidization, such as whole-genome duplication (WGD) or whole-genome triplication (WGT), in angiosperms dating back to the last century has been integral to angiosperm biology (Renny-Byfield and Wendel, 2014). All existing eudicots have experienced at least one polyploid event, called a hexaploid event; it is also known as a γ polyploid or WGT event (Aköz and Nordborg, 2019). Recently, more plant genome sequences have been released and more WGD and WGT events have been detected after γ polyploidization. For example, *Arabidopsis* experienced two additional lineage-specific WGD events, called α and β polyploidization events (Blanc et al., 2000). Cucurbitaceae plants share additional WGD events, except an ancient WGT event (Wang et al., 2018); the *Cucurbita* genus shares another WGD event (Sun et al., 2017; Barrera-Redondo et al., 2019). Furthermore, chayote (*Sechium edule*) of Cucurbitaceae experienced a species-specific WGD event (Fu et al., 2021). Polyploidization has been widely considered a necessary condition for successful plant domestication (Salman-Minkov et al., 2016). In addition to the aforementioned large-scale duplication events that have promoted the expansion of gene families, small-scale repetitive events, such as tandem duplication (TD) and proximal duplication (PD), also play important roles in the expansion of gene families (Wang et al., 2011; Conant et al., 2014; Qiao et al., 2018). Gene loss is considered a common result of WGD/T events in plants (Zou and Yang, 2019). However, copies of some TF families are often more easily retained than those of other families (Liang and Schnable, 2018).

Cucurbitaceae species climb using tendrils, which are specialized organs that arise from leaf axils (Liu et al., 2021a). Tendrils provide access to advantageous or higher positions to capture more light for more effective photosynthesis (Kiss, 2006; Valladares et al., 2011). In production, Cucurbitaceae crops are mostly cultivated by climbing from the ground or by artificially tying vines; accordingly, tendrils have become redundant organs that consume nutrients. The removal of tendrils increases labor costs and leaves plants susceptible to pathogen exposure from these wounds. Therefore, tendril-free breeding is an important direction that will meet the needs of horticultural cultivation. Cucumber and melon tendrils are branchless, while pumpkin and watermelon have strip tendrils with 2–4 branches (Wang et al., 2015; Sousa-Baena et al., 2018). *Teosinte branched 1* (*tb1*), *CYCLOIDEA* (*CYC*), and *PROLIFERATING CELL FACTORS 1* and *2* (*PCF1* and *PCF2*) are TFs formed by the *TCP* gene family, which is a plant-specific gene family involved in a series of developmental processes (Martin-Trillo and Cubas, 2010; Yang et al., 2020; Pei et al., 2021a). Gene duplication and diversification have resulted in the development of two clades (classes I and II) with different *TCP* domains. Class I (known as *PCF* or TCP-P class) includes *PCF1* and *PCF2* and is constituted by a class of closely related proteins (Kosugi and Ohashi, 2002; Navaud et al., 2007). Class II (known as TCP-C class) includes *tb1* and *CYC* and can be further classified into two clades, *CINCINNATA* (*CIN*) and *CYC/TB1* (Navaud et al., 2007; Martin-Trillo and Cubas, 2010). The genes of the *CIN* clade mainly participate in lateral organ development, while *CYC/TB1* of the *TCP* genes plays a central role in the development of axillary meristems that give rise to either lateral shoots or flowers (Crawford et al., 2004; Howarth and Donoghue, 2006; Efroni et al., 2008; Martin-Trillo and Cubas, 2010). A rare single nucleotide polymorphism (SNP) was identified in the *TEN* gene, which is essential for tendril development in cucumber (Wang et al., 2015). The *TEN* gene encodes a *TCP* TF conserved within Cucurbitaceae and is expressed specifically in tendrils (Wang et al., 2015). Additionally, the *TEN* gene binds to the enhancers of target genes through the C-terminus and preferentially binds to lysine 56 and 122 of histone H3 domains as a histone acetyltransferase (HAT) (Yang et al., 2020).

Although the genome data of many Cucurbitaceae species have been released, the evolutionary history of the main gene families in Cucurbitaceae remains unknown. Therefore, in this study, we collected the existing well-annotated Cucurbitaceae genome datasets and other representative model species and analyzed the contraction and expansion of major TF families in Cucurbitaceae. We also explored the effects of WGD/T events on the evolution of major gene families. Finally, a systematic comparative analysis of tendril-related genes in Cucurbitaceae was carried out, which provides rich genetic resources for future investigations and breeding applications.

2. Materials and methods

2.1. Genomic data

Genome annotation datasets of 24 plants were obtained from public databases (Table S1). The sequences of *Benincasa hispida*, *Lagenaria siceraria*, *C. lanatus*, *C. maxima*, *C. pepo*, *C. moschata*, *Cucurbita argyrosperma*, and *Trichosanthes anguina* were downloaded from CuGenDB (Zheng et al., 2019). Gene sequences of *C. sativus*, *Momordica charantia*, *S. edule*, and *C. melo* were downloaded from the NCBI database. Gene sequences of *Cucumis hystrix* and *Luffa cylindrica* were obtained from figshare (https://figshare.com) and CNGBdb (https://db.cngb.org), respectively. The gene sequences of 10 non-cucurbit plants were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov). TBtools software was used to eliminate redundancy caused by alternative splicing variants and incomplete gene predictions (Chen et al., 2020). Specifically, we used Gtf/Gff3 Sequences Extract tool of TBtools to extract coding sequences (CDSs) and remove alternative splicing variants (Chen et al., 2020). Then, we obtained the protein sequences using the batch translate CDS to Protein tool in TBtools. Some pseudogenes do not encode protein products due to reading frame losses, insertions, and senseless mutations. Therefore, we deleted these incomplete genes with >1 stop codon using Perl script.

2.2. Identification and classification of TFs

First, pfam_scan.pl script and the pfam database (http://pfam.sanger.ac.uk/) were used to predict the domains of all protein sequences in each species (Mistry et al., 2021). Then, the TFs in each species were extracted from the Pfam results using Perl script. The TF families extracted in this study are shown in Table S2. To ensure high accuracy, we used an e-value of $<1e^{-4}$ to filter the Pfam results according to previous studies (Song et al., 2014; Pei et al., 2021b).

Here, we classified the TFs by integrating PlantTFDB and previous research on legumes (Jin et al., 2017; Moharana and Venancio, 2020). In PlantTFDB, super families with the same Pfam number were regarded as multiple families. In this study, we classified TFs with the same Pfam ID into one gene family (i.e., *AP2/ERF*, *MYB*, *MADS*, and *HB*) for a more accurate comparative analysis according to a previous report (Moharana and Venancio, 2020).

2.3. Phylogenetic tree construction and orthogroup (OG) detection

A phylogenetic tree of 14 Cucurbitaceae and 10 representative non-cucurbit species was constructed using orthologous genes in OrthoFinder v2.3.12 (Emms and Kelly, 2019). Specifically, DIAMOND software was used for all-vs-all sequence alignment (Kelly and Maini, 2013). Then, the Markov Cluster Algorithm (MCL) was used to cluster the results according to the comparison, and the direct orthologous group was obtained. Finally, the STAG algorithm was used to infer a rootless species tree (Emms and Kelly, 2018).

We used SpeciesTree_root.txt obtained by OrthoFinder as the inference tree and r8s (v1.81) to extract the ultrametric tree (https://sourceforge.net/projects/r8s/). We used the cafetutorial_prep_r8s.py script of CAFÉ in GitHub to generate an input file for r8s (De Bie et al., 2006). Then, r8s was used to estimate the divergence time using the penalized likelihood (PL) method. By setting multiple smoothing values to carry out calculation tests, we selected the optimal value. We used the divergence time of rice and *Arabidopsis thaliana* as a reference for estimating the divergence time of other species. The divergence time of rice and *A. thaliana* was 152 Mya, which was obtained from the TimeTree database (Kumar et al., 2017).

2.4. Duplication types and identification of TFs

We identified the gene duplication type of TF OGs using the DupGen_finder program (Qiao et al., 2019). The main duplication types included WGD, TD, PD, transposed duplication (TRD), and dispersed duplication (DSD). Because multiple duplication patterns can simultaneously occur in the same group of genes, the identification order was WGD > TD > PD > TRD > DSD based on previous reports (Wang et al., 2011, 2012; Qiao et al., 2018).

2.5. Estimation of gene family expansion and contraction

CAFÉ v5 was used to evaluate the expansion and contraction of TF families using an ultrametric tree and OGs (Han et al., 2013). First, Perl script was used to extract the protein sequences of all TFs, which were grouped by OrthoFinder to generate OGs. Second, the OGs and ultrametric tree generated by r8s were used as the input files in CAFÉ to estimate contraction and expansion. The cafetutorial_report_analysis.py script in CAFÉ was used to count the number of expansions and contractions. Finally, Evoview v3 software was used to visualize the results (Subramanian et al., 2019).

2.6. Enrichment analysis of gene family contraction and expansion

We conduct gene family enrichment analysis based on the contraction and expansion results. First, the statistical file of species Pfam results was used as the background. The contraction and expansion results of each family in each species were enriched and analyzed. The scipy package in Python was used to conduct the enrichment analysis (Virtanen et al., 2020). The *P*-values were corrected using the Bonferroni method in R (Song et al., 2021). A fold-change >2 and corrected *P*-value (Q-value) <0.05 were used to define significant enrichment terms. An online website (http://www.ehbio.com/test/venn) was used to visualize the results.

2.7. Multiple alignment and phylogenetic analysis of gene families

We extracted the protein sequences of gene families whose number changed significantly in each species using TBtools (Chen et al., 2020). Mafft v7.475 was used for sequence alignment (Katoh and Standley, 2013). FastTree v2.1 was used to construct a phylogenetic tree of the TF families (Price et al., 2009). EVOVIEW v3 and iTOL v6.5 were used to edit the tree and add related information (Subramanian et al., 2019).

Multiple alignments were illustrated using Jalview v2 (Waterhouse et al., 2009).

3. Results and discussion

3.1. Systematic identification and classification of TFs

Based on the specificity of each TF family domain, we identified and classified all TFs of 24 plants (Table S1). A total of 42 641 TFs were identified in 14 Cucurbitaceae and 10 non-cucurbit species (Table S3), which were further divided into 63 families (Table S4). Furthermore, we extracted 26 850 TFs from 14 Cucurbitaceae genomes. We compared the TFs to *A. thaliana* using plantTFDB, which showed good consistency with the corresponding gene families. The TF numbers in *C. maxima* and *Selaginella moellendorffii* were 2 658 and 875, which had the highest and lowest number of TFs, respectively (Fig. 1, a). Our results showed that the number of TFs was usually higher in Cucurbitaceae than

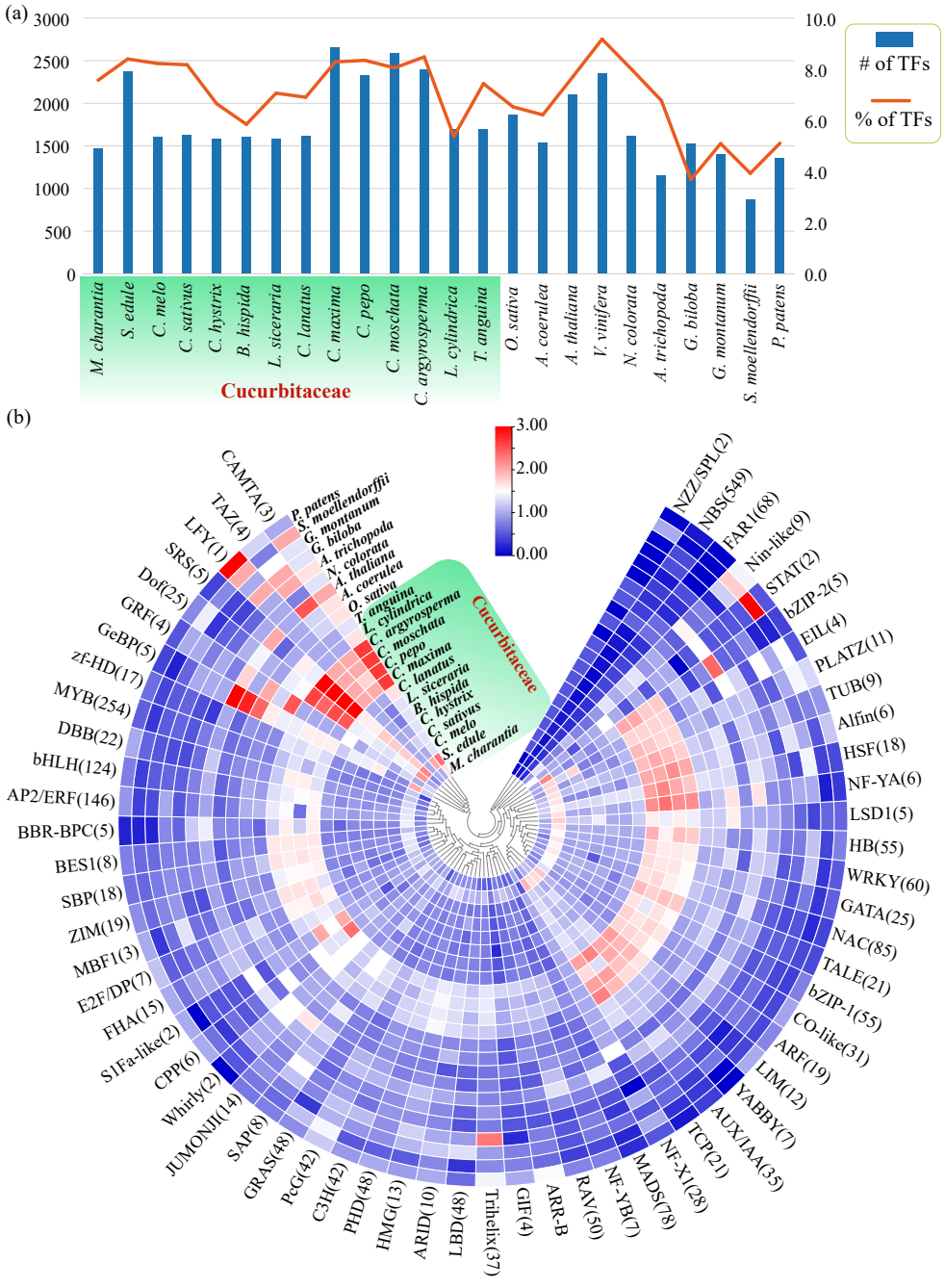


Fig. 1 Identification and comparison of transcription factor (TF) families in 14 Cucurbitaceae and 10 non-cucurbit species
(a) Absolute and relative numbers of the TFs in each species. The blue histogram shows the number of TFs in 24 species, and the orange line represents the percentage of TFs of each species in all genes. (b) Ratio of the number of 63 TF families of each species to *Vitis vinifera*. Values > 1 indicate that the TF families in the corresponding species were larger than *V. vinifera*. Numbers in brackets represent the number of related TF families in *V. vinifera*.

Table 1 The number of different duplication types of the transcription factors (TFs) in 14 Cucurbitaceae and 10 non-cucurbit species

Species	WGD	TD	PD	TRD	DSD	Singleton	Total TFs
Momordica charantia	441	83	18	418	375	136	1 471
Sechium edule	1 530	86	41	502	51	162	2 372
Cucumis melo	489	72	38	455	397	153	1 604
Cucumis sativus	577	68	30	460	363	130	1 628
Cucumis hystrix	514	81	47	430	369	145	1 586
Benincasa hispida	377	65	59	425	531	148	1 605
Lagenaria siceraria	554	84	61	450	296	142	1 587
Citrullus lanatus	572	78	37	444	350	137	1 618
Cucurbita maxima	2 005	50	19	364	35	185	2 658
Cucurbita pepo	1 665	46	8	394	38	175	2 326
Cucurbita moschata	1 935	53	32	359	25	189	2 593
Cucurbita argyrosperma	1 836	26	13	350	29	146	2 400
Luffa cylindrica	610	76	76	468	334	137	1 701
Trichosanthes anguina	679	85	46	445	300	146	1 701
*Oryza sativa**	644	144	100	488	315	172	1 863
*Aquilegia coerulea**	206	129	111	307	646	144	1 543
*Arabidopsis thaliana**	852	130	95	568	279	183	2 107
*Vitis vinifera**	598	404	269	500	350	235	2 356
*Nymphaea colorata**	508	215	62	390	275	163	1 613
*Amborella trichopoda**	35	85	46	77	800	110	1 153
*Ginkgo biloba**	6	204	137	14	1027	136	1 524
*Gnetum montanum**	0	117	43	0	1097	142	1 399
*Selaginella moellendorffii**	132	21	66	154	406	96	875
*Physcomitrella patens**	407	74	51	341	362	123	1 358

Note: The 10 non-cucurbit species are indicated with asterisks (*). WGD, whole-genome duplication; TD, tandem duplication; PD, proximal duplication; TRD, transposed duplication; DSD, dispersed duplication.

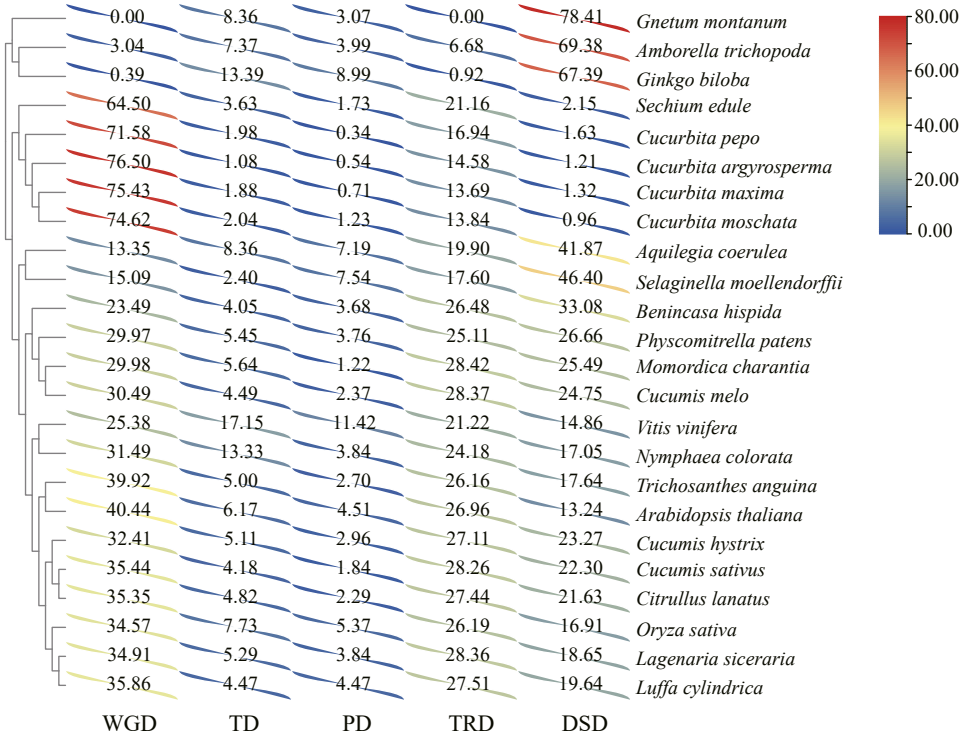


Fig. 2 The proportion of transcription factors of different duplication types to the total number of TFs in each species
WGD, whole-genome duplication; TD, tandem duplication; PD, proximal duplication; TRD, transposed duplication; DSD, dispersed duplication.

in non-cucurbit plants. This phenomenon indicated that some expansion events had occurred in Cucurbitaceae lineages.

To further our understanding of the different proportions of TFs in the genome, we compared the sizes of the TF families of different species. We found that the number of TFs was higher in *Amborella trichopoda* than in *S. moellendorffii* (Fig. 1, a). More importantly, > 70% (45/63) of the TF families in the basal angiosperm, *A. trichopoda*, was greater than the fern, *S. moellendorffii* (Table S4, Fig. 1, b). This result may be due to the expansion of TF lineages in seed plants by ancient WGD events (Moharana and

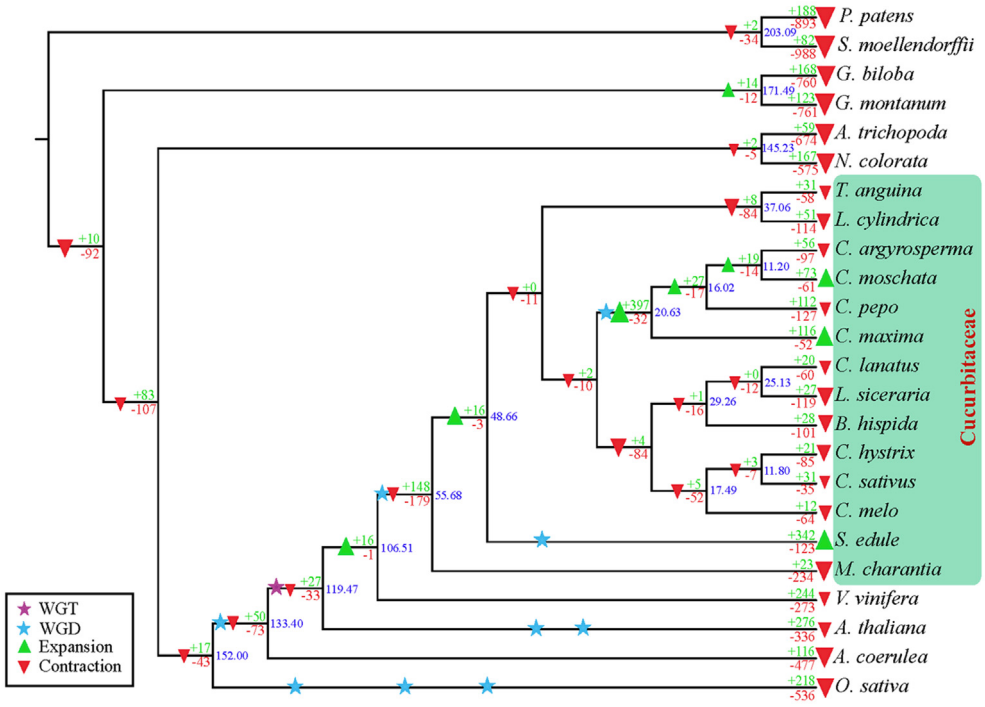


Fig. 3 Expansion and contraction analysis of the transcription factor families in 24 species

Green and red triangles refer to the nodes with more expansions and contractions, respectively. Blue numbers represent the divergence time between different species. Blue and purple stars represent whole-genome duplication (WGD) and whole-genome triplication (WGT) events, respectively.

Venancio, 2020). The *MADS* TF family is involved in regulating floral organ specificity (Zhao et al., 2021). It has been suggested that their diversity precedes angiosperm origin (Qiu and Köhler, 2022). Our study showed that the number of *MADS* family genes in *A. trichopoda* was more than two-times greater than that in *S. moellendorffii* (38/18) (Table S4), which was the result of the specific expansion of TF pedigrees. Compared to *S. moellendorffii*, *A. trichopoda* also had significantly expanded gene families with multiples of > 2, including *NF-YA*, *GeBP*, *NBS*, *BBR-BPC*, *bZIP-2*, *GATA*, *SAP*, and *TAZ* (Table S4). These results supported the occurrence of the specific expansion of TFs in early angiosperm diversification.

Aquilegia coerulea is a basic true dicotyledonous plant that is considered an ancient tetraploid. *A. coerulea* was key to the evolution of γ hexaploids shared by all core dicotyledons (Aköz and Nordborg, 2019; Moharana and Venancio, 2020). However, in *A. coerulea*, the number of some TF families was significantly higher than in *Vitis vinifera*, such as *GeBP* (*A. coerulea*, 14; *V. vinifera*, 5) and *JUMONJI* (*A. coerulea*, 23; *V. vinifera*, 14) (Fig. 1, b, Table S4). *V. vinifera* was used as a reference for comparative analysis with other core dicotyledons, as it has not experienced large-scale duplication after the γ hexaploid event (Jaillon et al., 2007; Moharana and Venancio, 2020). We found some significant species-specific expansions in Cucurbitaceae plants, such as *C. maxima* relative to *V. vinifera* (Fig. 1, b), which included *Dof* (62/25), *AUX/IAA* (78/35), *TCP* (43/21), *HB* (108/55), *CO-like* (55/31), *ARF* (33/19), *NAC* (141/85), *WRKY* (105/60), *bZIP-1* (92/55), and *bHLH* (192/124) (Table S4). Similarly, these families had also significantly expanded in *S. edule*, *C. pepo*, *C. moschata*, and *C. argyrosperma* when compared to *V. vinifera*. This result is supported by the findings of a previous report (Wan et al., 2013).

Shortly after the hexaploidization of the core dicotyledons, Cucurbitaceae experienced a cucurbit-common tetraploidization (CCT) event (Wang et al., 2018; Xie et al., 2019a). CCT events may have directly promoted the separation of Cucurbitaceae from other dicotyledons and may have even promoted the establishment of Cucurbitaceae (Wang et al., 2018). Two WGD events occurred in four *Cucurbita* species and *S. edule*, which were not shared with all other Cucurbitaceae (Sun et al., 2017; Fu et al., 2021). We found that the total number of TFs in *C. maxima*, *C. pepo*, *C. moschata*, *C. argyrosperma*, and *S. edule* was significantly higher than that in other Cucurbitaceae (Fig. 1, a). Moreover, 94% of the TF families expanded in the five aforementioned species (59/63), indicating that the polyploidization of Cucurbitaceae greatly increased the number of TF families. However, the number of *NBS* and *FAR1* gene families in Cucurbitaceae decreased significantly when compared to *V. vinifera* (Fig. 1, b). The number of *NBS* gene families in these five plants decreased significantly, especially in *C. pepo* (14) and *C. argyrosperma* (16). Previous studies showed that only a few tandem gene and fragment duplications had occurred in cucumber, which likely explains the small number of genes encoded by the *NBS* family (Wan et al., 2013). Too many disease-resistance genes could lead to a series of adverse traits, such as slow growth, low seed yield, and poor taste, according to previous reports (Zhai et al., 2011; Fei et al., 2013). Therefore, we speculated that the reduced number of *NBS* genes was an adaptive cost of Cucurbitaceae plants under

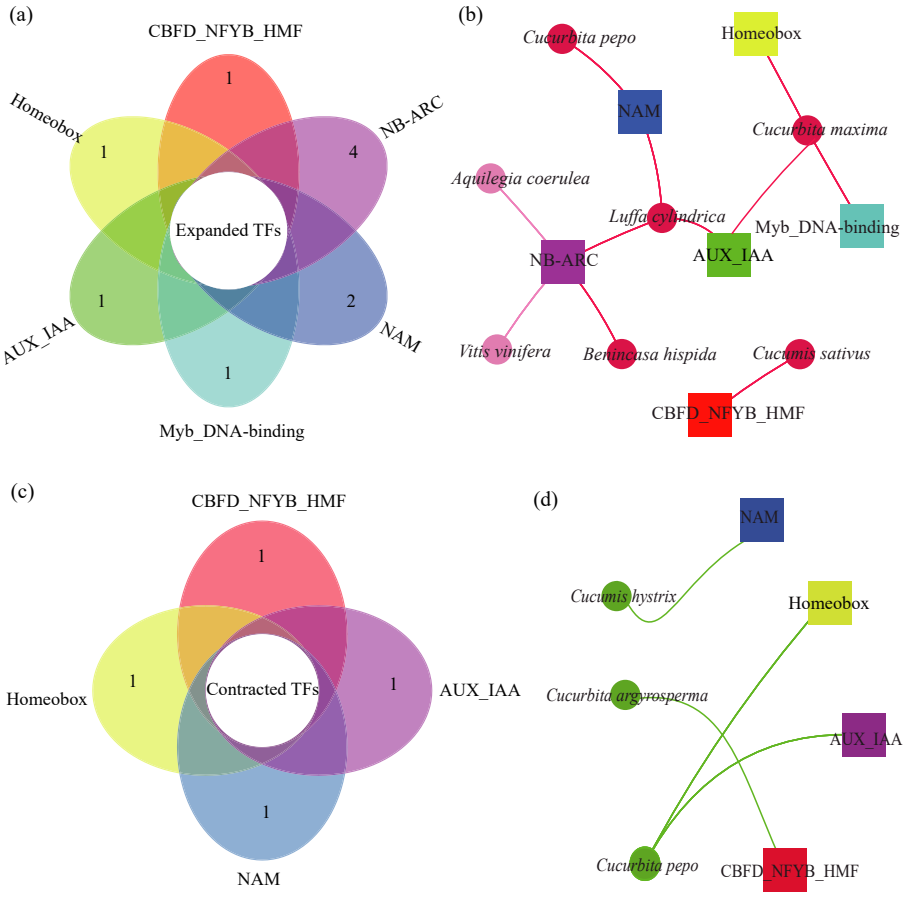


Fig. 4 Functional enrichment analysis of the genes in significantly expanded and contracted orthogroups (OGs)
(a) Petals with different colors represent different TF families with significant expansion, and numbers represent the number of species involved in this family. (b) Squares represent the distinct expansion of different TF families. Circles connected to the squares represent specific species involved in the family. Bright red represents Cucurbitaceae, and dark red represents non-cucurbit species. (c) Petals with different colors represent different TF families with significant contraction, and numbers represent the number of species involved in the family. (d) Squares represent different TF families with obvious contraction. Green circles connected to the squares represent specific species in the family.

strong artificial selection. We also learned that Cucurbitaceae use a specific defense mechanism against pests and diseases, such as lipoxygenase (*Lox*) genes, which resist various pathogens (Wan et al., 2013). Clearly, Cucurbitaceae are resistant to certain diseases, even with only a few *NBS* family genes.

3.2. *Contribution of duplication models to TF expansion*

Here, we studied the effects of a gene duplication model on TF family size in Cucurbitaceae. Each gene was assigned to a repeat type in the order of WGD > TD > PD > TRD > DSD. We found that > 90% of the TF families in Cucurbitaceae had at least one paralogous gene (Table 1 and Fig. 2). WGD was the main type in almost all Cucurbitaceae. A previous study reported that changes in the number of TFs were closely related to recent polyploidization events in Fabaceae (Moharana and Venancio, 2020). After continuous genome duplication events, most plants retained their genes and obtained many homologous TFs, which was consistent with our findings (Lehti-Shiu et al., 2017). *C. maxima* and *C. argyrosperma* had the largest number of WGD-type TFs, accounting for 76.50% and 75.43% of all TFs, respectively (Fig. 2). Similarly, the number of WGD-type TFs in *C. moschata*, *C. pepo*, and *S. edule* was significantly higher than that in other Cucurbitaceae, accounting for 60–70% of the total TFs (Fig. 2). Generally, WGD was the dominant repeat type in 13 Cucurbitaceae species. In one Cucurbitaceae species, *B. hispida*, the dominant type was DSD. Among the non-cucurbit species, WGD was the dominate type in *Oryza staiva*, *A. thaliana*, *V. vinifera*, *Nymphaea colorata*, and *Physcomitrella patens*. However, DSD was the dominate type in *A. coerulea*, *A. trichopoda*, *Ginkgo biloba*, *Gnetum montanum*, and *S. moellendorffii*. Our results clearly showed that most TF families in Cucurbitaceae had high repeatability and were greatly affected by polyploidization events.

3.3. *Phylogenetic analysis and contraction/expansion assessment*

To understand the diversity of plant TF families, the TF families of all plants were further divided into 1 210 OGs. Among all OGs, 60% (732/1 210) contained Cucurbitaceae and 112 were unique to Cucurbitaceae (Table S5).

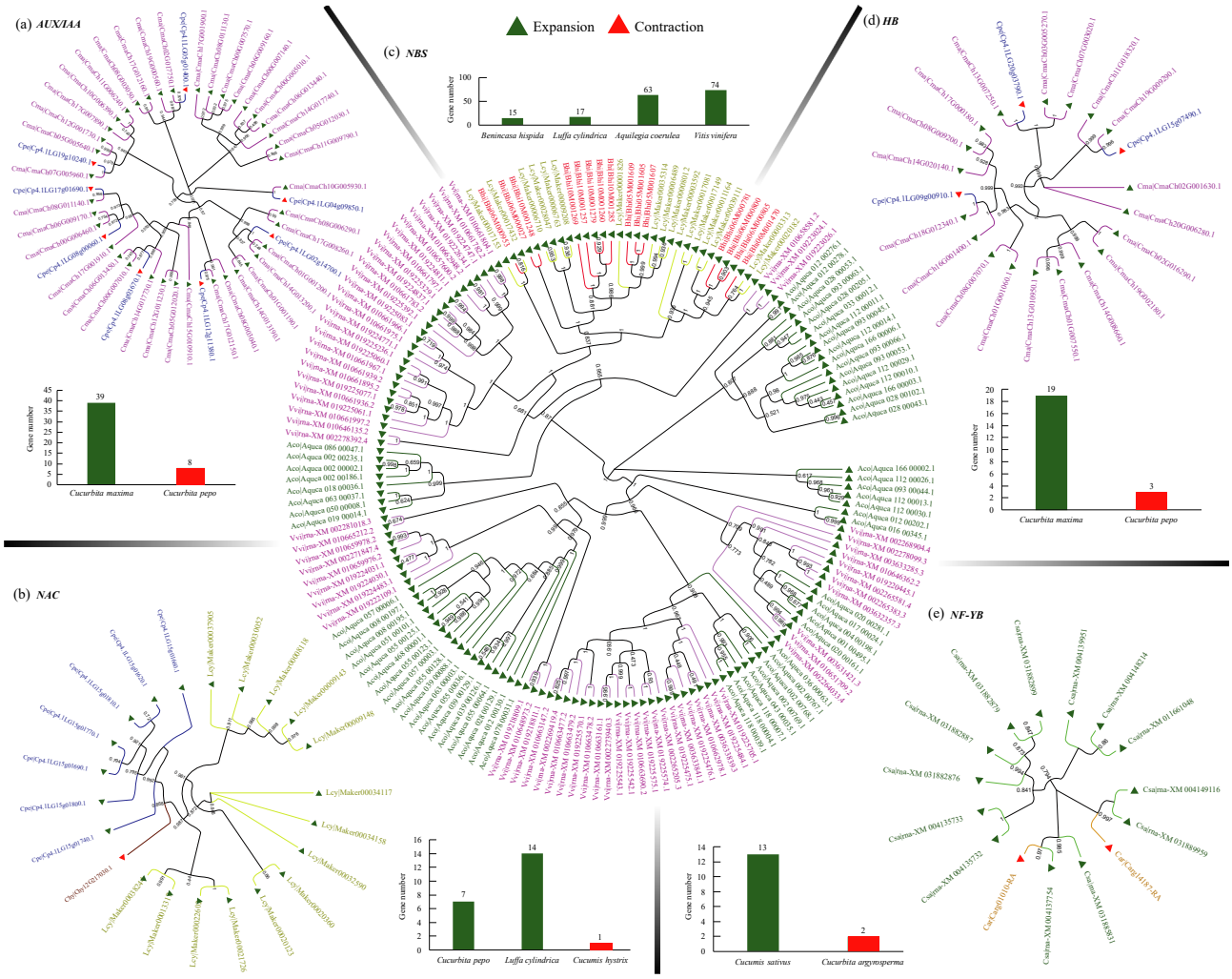


Fig. 5 Phylogenetic analysis of the enriched genes in significantly contracted and expanded OGs

Evolutionary tree constructed from the enriched *AUX/IAA* (a), *NAC* (*NAM*) (b), *NBS* (*NB-ARC*) (c), *Homeobox* (HB) (d), and *NF-YB* (*CBFD_NFYB_HMF*) (e) family genes. The colors of the genes represent different species. The enriched genes in significantly expanded and contracted OGs are marked with green and red triangles, respectively.

To better our understanding of the evolution of TF families in Cucurbitaceae, we analyzed the number of OG genes in each species. We calculated gene family expansions and contractions in different lineages of hypermetric trees based on gene-specific birth and mortality rates. We found that the TF families in Cucurbitaceae experienced recent WGD events and showed an overall trend of expansion, including *S. edule*, *C. maxima*, and *C. moschata* (Fig. 3). Moreover, this expansion trend was consistent with the aforementioned WGD-type TF expansions in these five species. Our results clearly revealed that WGD events promoted the expansion of most TF families. Overall, the expansion of TF families was strongly associated with two recent WGD events within Cucurbitaceae.

3.4. *Functional enrichment analysis of genes in significantly contracted or expanded gene families*

To further explore the function of genes in significantly expanded and contracted gene families, we performed a functional enrichment analysis and found that the number of expanded gene families was significantly greater than the number of contracted gene families in Cucurbitaceae and non-cucurbit plants. We identified 274 and 14 genes in the significantly expanded and contracted gene families, respectively (Table S6). A total of 137 genes had significantly expanded in Cucurbitaceae plants, of which, 52% (71/137) were from *C. maxima* and had originated during a recent WGD event (Table S6). Additionally, genes with significant expansion also existed in *B. hispida*, *C. pepo*, *C. sativus*, and *L. cylindrica*. Our results demonstrated that WGD events promoted the expansion of TF families. A total of 14 genes identified in the significantly contracted groups were from Cucurbitaceae, including *C. pepo*, *C. hystrix*, and *C. argyrosperma*. Although *C. pepo* and *C. argyrosperma* experienced a recent WGD event, some gene families exhibited significant contraction, which may have been caused by gene loss after a WGD event.

Moreover, we analyzed the TF family enrichment of the genes from significantly changed OGs. Our results showed that six TF families exhibited a significant expansion trend and four

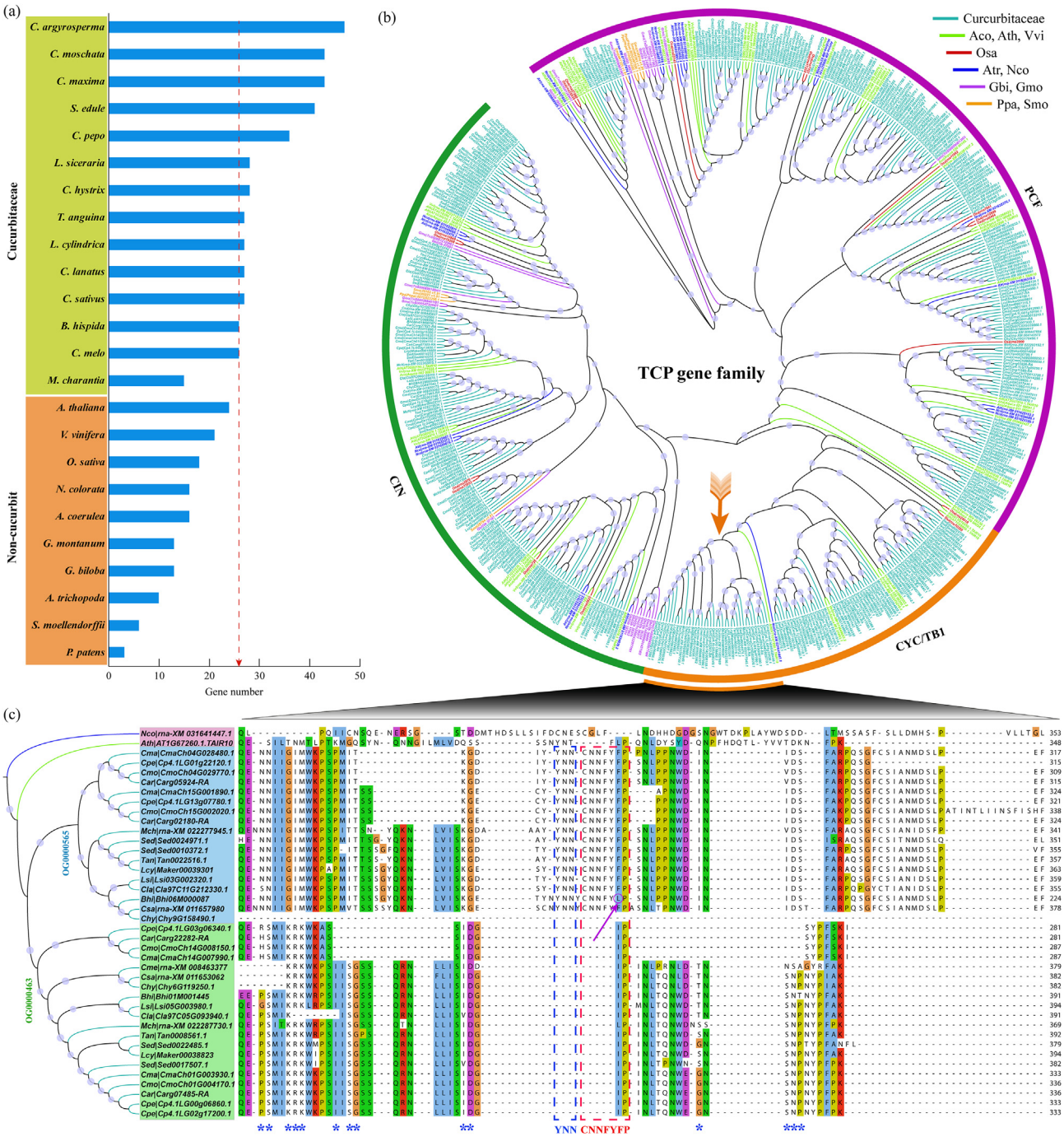


Fig. 6 TCP gene family analysis and *TEN* gene exploitation

(a) The number of *TCP* family genes in 14 Cucurbitaceae and 10 non-cucurbit species. (b) Maximum-likelihood trees of *TCP* family genes from 24 species using 1 000 bootstrap replicates. Bootstrap values >40% are indicated by a blue circle. Colored branches represent different species. Orange arrows indicate the branches of tendril-related *TEN* genes on the phylogenetic tree. (c) Multiple alignment of amino acid sequences of two groups (OG0000463 and OG0000565) and the *TEN* genes located in the OG0000565 group of each species. The red dashed box represents the conserved characteristic sequences within the tendril-related genes that were previously reported in Cucurbitaceae. The blue dashed boxes represent newly discovered conserved characteristic sequences in Cucurbitaceae. The purple arrow indicates the amino acid residue of the *TEN* gene in *B. hispida*, which changed from phenylalanine (F) to leucine (L). Blue asterisks indicate group-specific loci in the two groups.

exhibited a significant contraction trend (Fig. 4, a–d, Table S7). We found that the *NBS* gene family related to disease resistance exhibited an expansion trend in *B. hispida*, *L. cylindrica*, *V. vinifera*, and *A. coerulea*, and the *NAC* (*NAM*) TF family related to improved plant abiotic stress resistance significantly expanded in *C. pepo* and *L. cylindrica* (Fig. 4, b) (Duan et al., 2020; Ma et al., 2021). In *C. maxima*, the *Homeobox* (*HB*), *MYB*, and *AUX/IAA* families exhibited significant expansion, which are widely involved in the regulation of biological growth processes, such as meristem expression, cell cycle, and auxin mediated plant growth (Stracke et al., 2001; Yoon et al., 2020; Wei et al., 2021). In *C. sativus*, the *NF-YB* family also exhibited significant expansion; the genes in this family play important roles in promoting root formation, bud growth, and fruit ripening (Gan et al., 2013; Holland, 2013; Wu et al., 2014; Luo et al., 2018). Although these TF families had expanded in Cucurbitaceae, gene families that had contracted still existed. For example, the *AUX/IAA* and *HB* families of *C. pepo* exhibited significant contraction (Fig. 4, d).

Furthermore, we constructed phylogenetic trees of the *NBS*, *AUX/IAA*, *NAC*, *HB*, and *NF-YB* gene families with significantly contracted and expanded genes (Fig. 5). Generally, a WGD event expanded the scale of most TF families to improve resistance to various stressors and the growth regulation ability of Cucurbitaceae.

3.5. *Identification and phylogenetic analysis of TCP gene families in Cucurbitaceae*

The *TEN* gene is involved in cucumber tendril development, which is a *TCP* TF (Wang et al., 2015). In this study, we comprehensively analyzed the *TCP* gene family to explore the related genes that participate in Cucurbitaceae tendril development. We identified 581 *TCP* family genes from 24 species (Fig. 6, a, Table S4), among which, 441 were from 14 Cucurbitaceae and 140 were from 10 non-cucurbit species. The average number of *TCP* gene families in Cucurbitaceae (31.5) was greater than that in non-cucurbit species (14) (Fig. 6, a).

Among the examined non-cucurbit species, the greatest number of *TCP* genes was found in *A. thaliana* (24). However, almost all Cucurbitaceae (13/14, 92.86%) contained more genes than *A. thaliana* and other examined non-cucurbit species. In Cucurbitaceae, *M. charantia* was an exception and contained fewer *TCP* genes (15), which may have been the result of an incomplete genome or different developmental requirements. This phenomenon suggested that the *TCP* gene family plays an important role in most Cucurbitaceae species.

To further explore the evolution of the *TCP* gene family in Cucurbitaceae, we performed a phylogenetic analysis using *TCP* genes from 24 species (Fig. 6, b). The phylogenetic tree was clearly divided into two classes, I (*PCF*) and II. Class II was further classified into the *CIN* and *CYC/TB1* clades, similar to a previous report on *A. thaliana* (Martin-Trillo and Cubas, 2010). *TCP* genes were detected in moss (*P. patens*), fern (*S. moellendorffii*), and two gymnosperms (*G. biloba* and *G. montanum*) species in the *CIN* clade, but no gene was detected in the species of the *CYC/TB1* clade. Therefore, we speculated that the *CIN* clade may be more ancient than the *CYC/TB1* clade. *CYC/TB1* likely evolved later in angiosperms, as it was only detected in basal angiosperms (*N. colorata*) and other angiosperms in this study. Our findings corroborate the results of previous reports on the evolution of the *TCP* gene family in other species (Floyd and Bowman, 2007; Martin-Trillo and Cubas, 2010).

3.6. *Exploring candidate genes involved in Cucurbitaceae tendril development*

We found that the *TEN* gene, which is involved in cucumber tendril development, was located within the *CYC/TB1* clade of the phylogenetic tree (Fig. 6, b). We further explored the branch containing *TEN* genes and performed a multiple alignment analysis of these genes. The orthologous gene analysis showed that this branch further was divided into two gene families, OG0000463 and OG0000565 (Fig. 6, c). The *TEN* gene in cucumber was located in the OG0000565 group that contained 18 Cucurbitaceae genes. Through multiple sequence alignment, we explored whether these genes contained the characteristic sequence, CNNFYFP, which regulates tendril development genes reported in a previous study (Wang et al., 2015). Among these genes, *Chy9G158490.1* from *C. hystrix* was short (61 amino acids), which was likely due to the loss of segments during evolution or incomplete genome assembly; thus, the corresponding characteristic sequence, CNNFYFP, was not detected. In *C. melo*, no gene was detected in the OG0000565 group. Interestingly, in *B. hispida*, *Bhi06M000087* changed from CNNFYFP to CNNFYLP, that is, an amino acid changed from phenylalanine (F) to leucine (L), which may have mutated and resulted in the need for tendrils during natural or artificial selection.

The remaining 16 genes of the OG0000565 group contained characteristic sequences of tendril development genes, including two genes in *C. argyrosperma*, *C. maxima*, *C. moschata*, *C. pepo*, and *S. edule*, and one gene in *C. lanatus*, *C. sativus*, *L. cylindrica*, *L. siceraria*, *M. charantia*, and *T. anguina* (Fig. 6, c; Fig. S1). Interestingly, all five species with two genes had a WGD event within Cucurbitaceae, while the species with one gene had not undergone a WGD event within Cucurbitaceae. No tendril development-related genes containing the characteristic sequences were found in *B. hispida* and *C. hystrix*. We speculated that this finding may be due to incomplete genome assembly or other related genes performing similar functions.

In addition to the CNNFYFP characteristic sequence, we found a new characteristic sequence, YNN, which was located upstream of CNNFYFP (Fig. 6, c). This novel characteristic sequence was present in all genes of OG0000565, except *Chy9G158490.1* in *C. hystrix*. Therefore, this gene can be used to identify genes associated with Cucurbitaceae tendril development. In addition to these characteristic sequences, we found several group-specific loci in the genes belonging to the OG0000463 and OG0000565 groups (Fig. 6, c; Fig. S1). We speculated that these loci may have formed due to gradual site mutations of duplication genes during evolution.

In summary, we identified 16 TEN genes in 11 Cucurbitaceae species. These genes will provide abundant genetic resources for the construction of Cucurbitaceae tendril-related gene regulatory networks, evolutionary analyses, comparative genomics analyses, and functional studies of the maturation of genetic transformation systems of Cucurbitaceae plants (Tian et al., 2017, 2018; Zhang et al., 2020; Xin et al., 2022).

4. Conclusions

In this study, we compared and analyzed the TF families at the genome level of 14 Cucurbitaceae and 10 non-cucurbit plants. Our findings revealed that recent WGD events occurred in Cucurbitaceae, which had a profound effect on the expansion of most Cucurbitaceae TF families. Furthermore, we obtained a new characteristic sequence, YNN, which can be used for *TEN* gene exploitation in Cucurbitaceae crops. Finally, we identified 16 *TEN* genes in 11 Cucurbitaceae species, which will serve as rich genetic resources for functional studies and the marker-assisted breeding of Cucurbitaceae tendril-related genes in future studies.

Acknowledgments

This work was supported by the Natural Science Foundation of Hebei (Grant No. C2021209005), National Natural Science Foundation of China (Grant No. 32172583), the Natural Science Foundation for Distinguished Young Scholar of Hebei Province (Grant No. C2022209010), and the China Postdoctoral Science Foundation (Grant Nos. 2020M673188, 2021T140097).

Supplementary materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.hpj.2022.07.004.

REFERENCES

Ahmad, H., Rahman, M., Ahmar, S., Fiaz, S., Azeem, F., Shaheen, T., Ijaz, M., Bukhari, S., Khan, S., Mora-Poblete, F., 2021. Comparative genomic analysis of *MYB* transcription factors for cuticular wax biosynthesis and drought stress tolerance in *Helianthus annuus* L. Saudi J Biol Sci, 28: 5693–5703.

Aköz, G., Nordborg, M., 2019. The aquilegia genome reveals a hybrid origin of core eudicots. Genome Biol, 20: 1–12.

Barrera-Redondo, J., Ibarra-Laclette, E., Vázquez-Lobo, A., Gutiérrez-Guerrero, Y., de la Vega, G., Piñero, D., Montes-Hernández, S., Lira-Saade, R., Eguiarte, L., 2019. The genome of *Cucurbita argyrosperma* (silver-seed gourd) reveals faster rates of protein-coding gene and long noncoding RNA turnover and neofunctionalization within cucurbita. Mol Plant, 12: 506–520.

Blanc, G., Barakat, A., Guyot, R., Cooke, R., Delseny, M., 2000. Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell, 12: 1093–1101.

Chen, C., Chen, H., Zhang, Y., Thomas, H., Frank, M., He, Y., Xia, R., 2020. Tbtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant, 13: 1194–1202.

Chomicki, G., Schaefer, H., Renner, S., 2020. Origin and domestication of Cucurbitaceae crops: insights from phylogenies, genomics and archaeology. New Phytol, 226: 1240–1255.

Conant, G., Birchler, J., Pires, J., 2014. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Curr Opin Plant Biol, 19: 91–98.

Crawford, B., Nath, U., Carpenter, R., Coen, E., 2004. Cincinnata controls both cell differentiation and growth in petal lobes and leaves of *Antirrhinum*. Plant Physiol, 135: 244–253.

De Bie, T., Cristianini, N., Demuth, J., Hahn, M., 2006. CAFE: a computational tool for the study of gene family evolution. Bioinformatics, 22: 1269–1271.

Dinan, L., Whiting, P., Sarker, S., Kasai, R., Yamasaki, K., 1997. Cucurbitane-type compounds from *Hemsleya carnosiflora* antagonize ecdysteroid action in the *Drosophila melanogaster* BII cell line. Cell Mol Life Sci, 53: 271–274.

Duan, A., Yang, X., Feng, K., Liu, J., Xu, Z., Xiong, A., 2020. Genome-wide analysis of NAC transcription factors and their response to abiotic stress in celery (*Apium graveolens* L.). Comput Biol Chem, 84: 107186.

Efroni, I., Blum, E., Goldshmidt, A., Eshed, Y., 2008. A protracted and dynamic maturation schedule underlies *Arabidopsis* leaf development. Plant Cell, 20: 2293–2306.

Emms, D., Kelly, S., 2018. STAG: species tree inference from all genes. BioRxiv. https://doi.org/10.1101/267914.

Emms, D., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol, 20: 1–14.

Fei, Q., Xia, R., Meyers, B.C., 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. Plant Cell, 25: 2400–2415.

Floyd, Sandra K., Bowman, John L., 2007. The ancestral developmental tool kit of land plants. Int J Plant Sci, 168: 1–35.

Fu, A., Wang, Q., Mu, J., Ma, L., Wen, C., Zhao, X., Gao, L., Li, J., Shi, K., Wang, Y., 2021. Combined genomic, transcriptomic, and metabolomic analyses provide insights into chayote (*Sechium edule*) evolution and fruit development. Horticult Res, 8: 1–15.

Gan, D., Zhuang, D., Ding, F., Yu, Z., Zhao, Y., 2013. Identification and expression analysis of primary auxin-responsive *AUX/IAA* gene family in cucumber (*Cucumis sativus*). J Genet, 92: 513–521.

Gao, Z., Zhang, H., Cao, C., Han, J., Li, H., Ren, Z., 2020. QTL mapping for cucumber fruit size and shape with populations from long and round fruited inbred lines. Hortic Plant J, 6: 132–144.

Han, M., Thomas, G., Lugo-Martinez, J., Hahn, M., 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol, 30: 1987–1997.

Holland, P., 2013. Evolution of homeobox genes. Wiley Interdiscipl Rev Dev Biol, 2: 31–45.

Howarth, D., Donoghue, M.J., 2006. Phylogenetic analysis of the "ECE" (CYC/TB1) clade reveals duplications predating the core eudicots. Proc Natl Acad Sci USA, 103: 9101–9106.

Jaillon, O., Aury, J., 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., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pe, M., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.F., Weissenbach, J., Quetier, F., Wincker, P., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449: 463–467.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., Gao, G., 2017. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res, 45: D1040–D1045.

Karrar, E., Sheth, S., Navicha, W.B., Wei, W., Hassanin, H., Abdalla, M., Wang, X., 2019. A potential new source: nutritional and antioxidant properties of edible oils from *Cucurbit* seeds and their impact on human health. J Food Biochem, 43: e12733.

Katoh, K., Standley, D.M., 2013. Mafft multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol, 30: 772–780.

Kelly, S., Maini, P.K., 2013. DendroBLAST: approximate phylogenetic trees in the absence of multiple sequence alignments. PLoS ONE, 8: e58537.

Kiss, J.Z., 2006. Up, down, and all around: how plants sense and respond to environmental stimuli. Proc Natl Acad Sci USA, 103: 829–830.

Kosugi, S., Ohashi, Y., 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. Plant J, 30: 337–348.

Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. Timetree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol, 34: 1812–1819.

Lehti-Shiu, M.D., Panchy, N., Wang, P., Uygun, S., Shiu, S.-H., 2017. Diversity, expansion, and evolutionary novelty of plant DNA-binding transcription factor families. Biochim Biophys Acta (BBA) Gene Regul Mech, 1860: 3–20.

Li, X., Bao, T., Osire, T., Qiao, Z., Liu, J., Zhang, X., Xu, M., Yang, T., Rao, Z., 2021. *MarR*-type transcription factor *RosR* regulates glutamate metabolism network and promotes accumulation of L-glutamate in *Corynebacterium glutamicum* G01. Bioresour Technol, 342: 125945.

Liang, Z., Schnable, J.C., 2018. Functional divergence between subgenomes and gene pairs after whole genome duplications. Mol Plant, 11: 388–397.

Liu, X., Chen, J., Zhang, X., 2021a. Genetic regulation of shoot architecture in cucumber. Horticult Res, 8: 143.

Liu, X., Dong, S., Miao, H., Bo, K., Li, C., Yang, Y., Gu, X., Zhang, S., 2021b. Genome-wide analysis of expansins and their role in fruit spine development in cucumber (*Cucumis sativus* L.). Hortic Plant J. https://doi.org/10.1016/j.hpj.2021.11.004.

Lu, J., Nawaz, M., Wei, N., Cheng, F., Bie, Z., 2020. Suboptimal temperature acclimation enhances chilling tolerance by improving photosynthetic adaptability and osmoregulation ability in watermelon. Hortic Plant J, 6: 49–60.

Luo, J., Zhou, J., Zhang, J., 2018. *AUX/IAA* gene family in plants: molecular structure, regulation, and function. Int J Mol Sci, 19: 259.

Ma, J., Wang, L., Dai, J., Wang, Y., Lin, D., 2021. The *NAC*-type transcription factor *CaNAC46* regulates the salt and drought tolerance of transgenic *Arabidopsis thaliana*. BMC Plant Biol, 21: 1–11.

Martin-Trillo, M., Cubas, P., 2010. *TCP* genes: a family snapshot ten years later. Trends Plant Sci, 15: 31–39.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G., Sonnhammer, E., Tosatto, S., Paladin, L., Raj, S., Richardson, L., Finn, R., Bateman, A., 2021. Pfam: the protein families database in 2021. Nucleic Acids Res, 49: D412–D419.

Moharana, K., Venancio, T., 2020. Polyploidization events shaped the transcription factor repertoires in legumes (Fabaceae). Plant J, 103: 726–741.

Mukherjee, P., Singha, S., Kar, A., Chanda, J., Banerjee, S., Dasgupta, B., Haldar, P., Sharma, N., 2022. Therapeutic importance of Cucurbitaceae: a medicinally important family. J Ethnopharmacol, 282: 114599.

Navaud, O., Dabos, P., Carnus, E., Tremousaygue, D., Herve, C., 2007. *TCP* transcription factors predate the emergence of land plants. J Mol Evol, 65: 23–33.

Pei, Q., Li, N., Bai, Y., Wu, T., Yang, Q., Yu, T., Wang, Z., Liu, Z., Li, Q., Lin, H., Song, X., 2021a. Comparative analysis of the *TCP* gene family in celery, coriander and carrot (family Apiaceae). Veget Res, 1: 5.

Pei, Q., Yu, T., Wu, T., Yang, Q., Gong, K., Zhou, R., Cui, C., Yu, Y., Zhao, W., Kang, X., Cao, R., Song, X., 2021b. Comprehensive identification and analyses of the *Hsf* gene family in the whole-genome of three Apiaceae species. Hortic Plant J, 7: 457–468.

Price, M., Dehal, P., Arkin, A., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol, 26: 1641–1650.

Qi, X., Zhu, Y., Li, S., Zhou, H., Xu, X., Xu, Q., Chen, X., 2020. Identification of genes related to mesocarp development in cucumber. Hortic Plant J, 6: 293–300.

Qiao, A., Fang, X., Liu, S., Liu, H., Gao, P., Luan, F., 2021. QTL-seq identifies major quantitative trait loci of stigma color in melon. Hortic Plant J, 7: 318–326.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A., 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. Genome Biol, 20: 38.

Qiao, X., Yin, H., Li, L., Wang, R., Wu, J., Wu, J., Zhang, S., 2018. Different modes of gene duplication show divergent evolutionary patterns and contribute differently to the expansion of gene families involved in important fruit traits in pear (*Pyrus bretschneideri*). Front Plant Sci, 9: 161.

Qiu, Y., Köhler, C., 2022. Endosperm evolution by duplicated and neofunctionalized type I *MADS*-box transcription factors. Mol Biol Evol, 39: msab355.

Renny-Byfield, S., Wendel, J., 2014. Doubling down on genomes: polyploidy and crop plants. Am J Bot, 101: 1711–1725.

Rushton, P., Somssich, I., Ringler, P., Shen, Q., 2010. *WRKY* transcription factors. Trends Plant Sci, 15: 247–258.

Salman-Minkov, A., Sabath, N., Mayrose, I., 2016. Whole-genome duplication as a key factor in crop domestication. Nat Plants, 2: 1–4.

Schaefer, H., Heibl, C., Renner, S., 2009. Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. Proc Biol. Sci, 276: 843–851.

Smith, R., Bryant, R., 1975. Metal substitutions in carbonic anhydrase: a halide ion probe study. Biochem Biophys Res Commun, 66: 1281–1286.

Soltis, D., Visger, C., Soltis, P., 2014. The polyploidy revolution then… and now: Stebbins revisited. Am J Bot, 101: 1057–1078.

Song, X., Li, N., Guo, Y., Bai, Y., Wu, T., Yu, T., Feng, S., Zhang, Y., Wang, Z., Liu, Z., Lin, H., 2021. Comprehensive identification and characterization of simple sequence repeats based on the whole-genome sequences of 14 forest and fruit trees. For Res, 1: 7.

Song, X., Liu, T., Duan, W., Ma, Q., Ren, J., Wang, Z., Li, Y., Hou, X., 2014. Genome-wide analysis of the *GRAS* gene family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Genomics, 103: 135–146.

Sousa-Baena, M., Lohmann, L., Hernandes-Lopes, J., Sinha, N., 2018. The molecular control of tendril development in Angiosperms. New Phytol, 218: 944–958.

Stracke, R., Werber, M., Weisshaar, B., 2001. The *R2R3-MYB* gene family in *Arabidopsis thaliana*. Curr Opin Plant Biol, 4: 447–456.

Subramanian, B., Gao, S., Lercher, M., Hu, S., Chen, W., 2019. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. Nucleic Acids Res, 47: W270–W275.

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., Zhang, J., Zhang, H., Gong, G., Jia, Z., 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. Mol Plant, 10: 1293–1306.

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., Zhang, H., Ren, Y., Gong, G., Zong, M., Liu, F., Chen, Q., Xu, Y., 2018. Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. Plant Cell Rep, 37: 1353–1356.

Tian, S., Jiang, L., Gao, Q., Zhang, J., Zong, M., Zhang, H., Ren, Y., Guo, S., Gong, G., Liu, F., Xu, Y., 2017. Efficient CRISPR/Cas9-based gene knockout in watermelon. Plant Cell Rep, 36: 399–406.

Tsai, Y., Lin, C., Chen, B., 2010. Preparative chromatography of flavonoids and saponins in *Gynostemma pentaphyllum* and their antiproliferation effect on hepatoma cell. Phytomedicine, 18: 2–10.

Valladares, F., Gianoli, E., Saldana, A., 2011. Climbing plants in a temperate rainforest understorey: searching for high light or coping with deep shade? Ann Bot, 108: 231–239.

Velthuijs, N., Meldal, B., Geessinck, Q., Porras, P., Medvedeva, Y., Zubritskiy, A., Orchard, S., Logie, C., 2021. Integration of transcription coregulator complexes with sequence-specific

DNA-binding factor interactomes. Biochim Biophys Acta (BBA) Gene Regul Mech, 1864: 194749.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy, C., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. Nat Methods, 17: 261–272.
Wan, H., Yuan, W., Bo, K., Shen, J., Pang, X., Chen, J., 2013. Genome-wide analysis of *NBS*-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of *NBS*-encoding genes in Cucurbitaceae crops. BMC Genom, 14: 1–15.
Wang, J., Sun, P., Li, Y., Liu, Y., Yang, N., Yu, J., Ma, X., Sun, S., Xia, R., Liu, X., 2018. An overlooked paleotetraploidization in Cucurbitaceae. Mol Biol Evol, 35: 16–26.
Wang, S., Yang, X., Xu, M., Lin, X., Lin, T., Qi, J., Shao, G., Tian, N., Yang, Q., Zhang, Z., Huang, S., 2015. A rare SNP identified a *TCP* transcription factor essential for tendril development in cucumber. Mol Plant, 8: 1795–1808.
Wang, Y., Tang, H., Debarry, J., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J., Paterson, A., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res, 40: e49.
Wang, Y., Wang, X., Tang, H., Tan, X., Ficklin, S., Feltus, F., Paterson, A., 2011. Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. PLoS ONE, 6: e28150.
Waterhouse, A., Procter, J., Martin, D., Clamp, M., Barton, G., 2009. Jalview version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25: 1189–1191.
Wei, S., Chen, Y., Hou, J., Yang, Y., Yin, T., 2021. *AUX/IAA* and *ARF* gene families in salix suchowensis: identification, evolution, and dynamic transcriptome profiling during the plant growth process. Front Plant Sci, 12: 769.
Wu, J., Liu, S., Guan, X., Chen, L., He, Y., Wang, J., Lu, G., 2014. Genome-wide identification and transcriptional profiling analysis of auxin response-related gene families in cucumber. BMC Res Notes, 7: 1–13.
Wu, J., Wu, Y., Yang, B., 2002. Anticancer activity of *Hemsleya amabilis* extract. Life Sci, 71: 2161–2170.
Xie, D., Xu, Y., Wang, J., Liu, W., Zhou, Q., Luo, S., Huang, W., He, X., Li, Q., Peng, Q., Yang, X., Yuan, J., Yu, J., Wang, X., Lucas, W., Huang, S., Jiang, B., Zhang, Z., 2019a. The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun, 10: 5158.
Xie, Z., Liu, W., Huang, H., Slavin, M., Zhao, Y., Whent, M., Blackford, J., Lutterodt, H., Zhou, H., Chen, P., 2010. Chemical composition of five commercial *Gynostemma pentaphyllum* samples and their radical scavenging, antiproliferative, and anti-inflammatory properties. J Agric Food Chem, 58: 11243–11249.
Xie, Z., Nolan, T.M., Jiang, H., Yin, Y., 2019b. *AP2/ERF* transcription factor regulatory networks in hormone and abiotic stress responses in *Arabidopsis*. Front Plant Sci, 10: 228.
Xin, T., Tian, H., Ma, Y., Wang, S., Yang, L., Li, X., Zhang, M., Chen, C., Wang, H., Li, H., Xu, J., Huang, S., Yang, X., 2022. Targeted creating new mutants with compact plant architecture using CRISPR/Cas9 genome editing by an optimized genetic transformation procedure in Cucurbit plants. Hortic Res, 9: uhab086.
Yang, J., Zhang, J., Du, H., Zhao, H., Mao, A., Zhang, X., Jiang, L., Zhang, H., Wen, C., Xu, Y., 2021. Genetic relationship and pedigree of Chinese watermelon varieties based on diversity of perfect SNPs. Hortic Plant J, 8: 489–498.
Yang, X., Yan, J., Zhang, Z., Lin, T., Xin, T., Wang, B., Wang, S., Zhao, J., Zhang, Z., Lucas, W.J., Li, G., Huang, S., 2020. Regulation of plant architecture by a new histone acetyltransferase targeting gene bodies. Nat Plants, 6: 809–822.
Yoon, J., Cho, L., Yang, W., Pasriga, R., Wu, Y., Hong, W., Bureau, C., Wi, S., Zhang, T., Wang, R., 2020. Homeobox transcription factor *OsZHD2* promotes root meristem activity in rice by inducing ethylene biosynthesis. J Exp Bot, 71: 5348–5364.
Zhai, J., Jeong, D.H., De Paoli, E., Park, S., Rosen, B.D., Li, Y., Gonzalez, A.J., Yan, Z., Kitto, S.L., Grusak, M.A., Jackson, S.A., Stacey, G., Cook, D.R., Green, P.J., Sherrier, D.J., Meyers, B.C., 2011. MicroRNAs as master regulators of the plant *NB-LRR* defense gene family via the production of phased, trans-acting siRNAs. Genes Dev, 25: 2540–2553.
Zhang, M., Liu, Q., Yang, X., Xu, J., Liu, G., Yao, X., Ren, R., Xu, J., Lou, L., 2020. CRISPR/Cas9-mediated mutagenesis of *Clpsk1* in watermelon to confer resistance to *Fusarium oxysporum* f.Sp. *niveum*. Plant Cell Rep, 39: 589–595.
Zhang, X., Zhou, T., Yang, J., Sun, J., Ju, M., Zhao, Y., Zhao, G., 2018. Comparative analyses of chloroplast genomes of Cucurbitaceae species: lights into selective pressures and phylogenetic relationships. Molecules, 23: 2165.
Zhao, D., Chen, Z., Xu, L., Zhang, L., Zou, Q., 2021. Genome-wide analysis of the *MADS-box* gene family in maize: gene structure, evolution, and relationships. Genes-Basel, 12: 1956.
Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., Zhao, K., Blanca, J., Zhang, Z., Huang, S., Xu, Y., Weng, Y., Mazourek, M., U, K., Ando, K., McCreight, J., Schaffer, A., Burger, J., Tadmor, Y., Katzir, N., Tang, X., Liu, Y., Giovannoni, J., Ling, K., Wechter, W., Levi, A., Garcia-Mas, J., Grumet, R., Fei, Z., 2019. Cucurbit genomics database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. Nucleic Acids Res, 47: D1128–D1136.
Zhou, M., Guo, S., Tian, S., Zhang, J., Ren, Y., Gong, G., Li, C., Zhang, H., Xu, Y., 2020. Overexpression of the watermelon ethylene response factor *ClERF069* in transgenic tomato resulted in delayed fruit ripening. Hortic Plant J, 6: 247–256.
Zou, Z., Yang, J., 2019. Genome-wide comparison reveals divergence of cassava and rubber aquaporin family genes after the recent whole-genome duplication. BMC Genom, 20: 1–16.