A common whole-genome paleotetraploidization in Cucurbitales

Jiaqi Wang, Min Yuan, Yishan Feng, Yan Zhang, Shoutong Bao, Yanan Hao, Yue Ding, Xintong Gao, Zijian Yu, Qiang Xu, Junxin Zhao, Qianwen Zhu, Ping Wang, Chunyang Wu, Jianyu Wang, Yuxian Li, Chuanyuan Xu and Jinpeng Wang

1 Department of Bioinformatics, School of Life Sciences, North China University of Science and Technology, Tangshan 063000, China
2 State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Science, Beijing 100093, China
3 University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

Cucurbitales are an important order of flowering plants known for encompassing edible plants of economic and medicinal value and numerous ornamental plants of horticultural value. By reanalyzing the genomes of two representative families (Cucurbitaceae and Begoniaceae) in Cucurbitales, we found that the previously identified Cucurbitaceae common paleotetraploidization that occurred shortly after the core-eudicot-common hexaploidization event is shared by Cucurbitales, including Begoniaceae. We built a multigenome alignment framework for Cucurbitales by identifying orthologs and paralogs and systematically redateing key evolutionary events in Cucurbitales. Notably, characterizing the gene retention levels and genomic fractionation patterns between subgenomes generated from different polyploidizations in Cucurbitales suggested the autopolyploid nature of the Begoniaceae common tetraploidization and the allopolyploid nature of the Cucurbitales common tetraploidization and the Cucurbita-specific tetraploidization. Moreover, we constructed the ancestral Cucurbitales karyotype comprising 17 proto-chromosomes, confirming that the most recent common ancestor of Cucurbitaceae contained 15 proto-chromosomes and rejecting the previous hypothesis for an ancestral Cucurbitaceae karyotype with 12 proto-chromosomes. In addition, we found that the polyploidization and tandem duplication events promoted the expansion of gene families involved in the cucurbitacin biosynthesis pathway; however, gene loss and chromosomal rearrangements likely limited the expansion of these gene families.

Introduction

Among the eudicots, Cucurbitales is an important order of flowering plants with unisexual flowers, with more than 2,680 species belonging to 128 genera (Matthews and Endress, 2004; Schaefer et al., 2009; Filipowicz and Renner, 2010). Cucurbitales plants are mostly annual or perennial...
vines and herbs, and the remainder are thorny shrubs or trees (Schaef er et al., 2009). Cucurbitales includes many important edible and medicinal plants, such as cucumber (Cucumis sativa) \((2n = 2x = 14)\), melon (Cucumis melo) \((2n = 2x = 24)\), and pumpkin (Cucurbita moschata) \((2n = 2x = 40)\) (Xie et al., 2019), as well as many ornamental plants, such as Begonia masoniana \((2n = 2x = 30)\) and Begonia loranthoides \((2n = 2x = 38)\) (Li et al., 2022). Because of their high economic and horticultural value, genome sequences of 24 species have been obtained in the two largest families, Cucurbitaceae and Begoniaceae, in Cucurbitales. Cucurbitaceae is the second largest fruit and vegetable family among angiosperms after Solanaceae, containing more than 115 genera and 960 species (Wang et al., 2018a; Guo et al., 2020; Ma et al., 2022). Begoniaceae has more than 2 genera and 1,601 species (http://www.theplantlist.org/) and is known for its great diversity in leaf shapes, colors, patterns, and textures (Dewitte et al., 2011). Based on its rich species diversity, Begoniaceae is also expected to be a model species in the field of floral research. Deciphering the genomes of Cucurbitaceae and Begoniaceae provides important data materials for studying the evolutionary trajectory of Cucurbitales plants.

Recursive polyploidization or whole-genome duplication (WGD) is an important driver of plant evolution and is widely present in the evolution of angiosperms (Sol tis et al., 2015; Soltis and Soltis, 2016). Polyploidization results in the generation of a large number of duplicated genes, leading to genome instability (Wang et al., 2009), which is characterized by extensive chromosomal rearrangements and massive gene loss (Paterson et al., 2004; Wang et al., 2005, 2017, 2018a; Murat et al., 2017; Zhuang et al., 2019). Similar to most plant clades in eudicots, Cucurbitales have experienced the core eudicot common hexaploidization (ECH or gamma) event (The Angiosperm Phy logeny Group, 2016). The ECH event was detected in the Arabidopsis thaliana genome and subsequently better characterized in the Vitis vinifera genome, which revealed the triplication of seven ancient eudicot chromosomes at \(\sim 115–130\) million years ago (MYA) (Bowers et al., 2003; Jaillon et al., 2007). Previously, by analyzing the molecular distance of homologous genes generated by polyploidy in Cucurbitaceae, it was found that Cucurbitaceae experienced a paleotetraploidization event after the ECH; this WGD event occurred at approximately 90–102 MYA (Wang et al., 2018a). The tetraploidization event of Cucurbitales occurred at a time close to the ECH event. Considering the relatively close phylogenetic position of Cucurbitaceae and Begoniaceae (The Angiosperm Phy logeny Group, 2016), we speculate that Begoniaceae may also share this polyploidization event. However, recent studies on the estimates of the age of molecular differentiation based on DNA sequence data suggest that, apart from the ECH, Begoniaceae experienced only one duplication event at \(\sim 35 \pm 8\) MYA (Li et al., 2022). Therefore, whether Cucurbitaceae and Begoniaceae experienced two polyploidization events independently or shared one polyploidization event is a question worth investigating.

Chromosomal rearrangements are extensive in eukaryote genomes and involve chromosome inversions, translocations, and fusions (Zhao et al., 2021). Extensive rearrangements can promote genome remodeling and provide opportunities for genetic innovation (Pont et al., 2019). Among all patterns of chromosomal rearrangements, chromosome fusion is the main reason for the general reduction in chromosomes following recursive polyploidization (Pont et al., 2019). Fusion involves two patterns, chromosome end-to-end joining (EEJ) and nested chromosome fusion (NCF) (Schubert and Lysak, 2011; Wang et al., 2015b, 2016a). The specific mechanisms of ancestral karyotype evolution and the impact of chromosomal rearrangements on the structure and function of extant genomes are still very important questions (Schubert, 2007). Previous studies have shown that the ancestral eudicot karyotype (AKE) has seven chromosomes (Jaillon et al., 2007). However, there is still no uniform conclusion for the construction of the ancestral Cucurbitaceae karyotype (ACK-II), which is considered to have 12 or 15 proto-chromosomes (Wu et al., 2017; Xie et al., 2019), which has motivated us to conduct this study. Moreover, the ancestral Cucurbitales karyotype (ACK-I) has still not been studied. The evolutionary trajectory of karyotypes is an answer to the question of how the ancestors of Cucurbitales originated and evolved. Thus, it is crucial to reconstruct the ancestral karyotypes of Cucurbitales and Cucurbitaceae.

Here, by performing our previously developed comparative genomic pipeline (Wang et al., 2018a), we compared all available genomes with a chromosome-level assembly from the Cucurbitales and the well-characterized eudicot reference genome V. vinifera (Figure 1A). We aimed to determine the times of the Cucurbitales common tetraploidization (CCT) and other polyploidization events in Cucurbitales; to construct a multigene alignment framework for Cucurbitales through identified orthologs and paralogs produced by polyploidization and species divergence; to construct the ACK-I and clarify the ACK-II; to explore gene retention and loss and the situation of biased genomic fractionation in Cucurbitaceae and Begoniaceae; and to explore the gene expansion patterns of families in the cucurbitacin synthesis pathway.

Results
Cucurbitaceae common tetraploidization event shared by Cucurbitales
To infer the placement of previously identified Cucurbitaceae common tetraploidization event in Cucurbitales, we performed a comparative analysis of the 14 genomes in six genera from two families Cucurbitaceae and Begoniaceae by using V. vinifera as outgroup (Figure 1A and Supplemental Table S1). Through genome collinearity comparisons, we identified the homologous genes and blocks within and between the studied species (Supplemental Table S2). We
found that the Begoniaceae genomes contain better intragenomic collinearity than Cucurbitaceae genomes, and the plants from these two families maintain excellent intergenomic collinearity. Notably, for blocks with four or more collinear gene pairs, the similar numbers of identified collinear gene pairs were found in the plants of genus *Begonia* and *Cucurbita*, which are including 19,892–34,215 collinear gene pairs located in 2,226–2,647 blocks. Then, we characterized the synonymous nucleotide substitution rate ($K_s$) between collinear gene pairs, and estimated that the peaks of $K_s$ distributions were related to the polyploidization and speciation events (Figure 2A and Supplemental Table S3). Clearly, we found that the three $K_s$ peaks of *B. loranthoides* at 0.425, 1.034, and 1.174, and the two younger peaks are likely corresponding to the two rounds of WGDs occurred in *Begonia* after the ECH event, rather than the one WGD event reported previously (Li et al., 2022).

Further, the homologous gene dotplots were produced by our custom software (Wang et al., 2017, 2019b) and used to identify the orthologous and paralogous regions (Figure 1, B–G). The orthologous regions and paralogous regions can be identified according to the peaks of $K_s$ distributions related to the specific speciation and polyploidization event (see “Materials and methods” for details). We observed an orthology ratio of 1:4 between the *V. vinifera* and *B. loranthoides* genomes (Figure 1B and Supplemental Figures S1 and S2).
and detected a genomic region of *B. loranthoides*, matched to one paralogous region generated from the most recent WGD and two paralogous regions generated from the older WGD event (Figure 1C and Supplemental Figures S3 and S4). These results indicate that the *B. loranthoides* experienced two successive polyploidization events after the ECH. Then, we obtained an orthology ratio of 1:2 between the *V. vinifera* and *C. melo* genomes and an orthology ratio of 1:2 between the *C. melo* and *B. loranthoides* genomes (Figure 1, D and E and Supplemental Figures S5–S8). These results indicate that the Cucurbitaceae common tetraploidization also shared by Begoniaceae, which can be called the CCT event. If the CCT is shared only by Cucurbitaceae, the orthology ratio between *C. melo* and *B. loranthoides* should be 2:4, rather than 1:2. Similarly, we confirmed that the *C. moschata* experienced two successive polyploidization events after the ECH (Figure 1, F and G and Supplemental Figures S9–S12). Based on this analysis, we obtained an expected topology of homologous gene tree related to the WGDs and divergence of *V. vinifera* and three representative species of Cucurbitales, and showed the placements of the key polyploidization events on the species tree (Figure 1, A, H, and I).

**Evolutionary rates and dates of key events in Cucurbitales**

By analyzing the Ks distributions of collinear genes within and between the genomes, we characterized the divergence of evolutionary rates between Cucurbitales plants (Figure 2A and Supplemental Table S3). Within genomes, the Ks of paralogous gene pairs generated from the polyploidization events with very close time often show an overlapping distribution and with abnormally long tails. This makes it difficult to identify the Ks peaks corresponding to the polyploidization events. Here, we employed a comprehensive statistical approach to identify the major normal distributions within the observed Ks distributions, as proposed in previous studies (Wang et al., 2017, 2019b, 2020a). We found that the ECH-related Ks peaks from the different genomes varied and were present at \( t/C24 = 1.02 \) for *V. vinifera*, \( t/C24 = 1.174 \) for *B. loranthoides*, and \( t/C24 = 1.358 \) for *C. melo* (Supplemental Table S3). This result indicates that in Cucurbitales, the *B. loranthoides* has a slower evolutionary rate than *C. melo*, and the *C. melo* has been proved to be the slowest plant in Cucurbitaceae (Wang et al., 2018a). Therefore, we conclude that the evolutionary rate of the genomes of Begoniaceae is slower than that of Cucurbitaceae.

Differences in the rate of evolution can lead to the temporal distortion of evolutionary events. To date the key evolutionary events in Cucurbitales, we used an algorithm similar to the one proposed previously (Wang et al., 2015b, 2016a, 2016b, 2017; Liu et al., 2017; Sun et al., 2017b) to...
correct the $K_s$ distributions related to the polyploidizations and speciation. After three rounds of correction, we aligned the $K_s$ peaks of the ECH, CCT, and Cucurbita-specific tetraploidization (CST) events in each genome to the 1.02, 0.824, and 0.22, respectively (Figure 2B and Supplemental Table S4). Based on the occurrence time of the ECH event at $\sim$115–130 MYA (Paterson et al., 2004; Jaillon et al., 2007; Jiao et al., 2012; Vekemans et al., 2012), we infer that the CCT event occurred at $\sim$93–105 MYA and that the divergence of Begoniaceae from Cucurbitaceae (with a $K_s$ peak of $\sim$0.798) occurred shortly after the CCT event at $\sim$90–101 MYA. In addition, we inferred that the Begoniaceae common and Cucurbita-specific polyploidizations (Begoniaceae common tetraploidization [BCT] and CST) occurred at $\sim$38–43 and $\sim$25–28 MYA, and the divergence of $C.$ moschata and cushaw pumpkin (Cucurbita argyrosperma) occurred at $\sim$24–27 MYA, suggesting that the CST occurred closer to the time of $C.$ moschata and C. argyrosperma divergence. Divergence of C. melo from C. moschata, watermelon (Citrullus lanatus), and C. sativus occurred at $\sim$27–31, $\sim$22–25, and $\sim$12–14 MYA, respectively (Figures 1A and 2B).

Multigenome alignment framework and online resource for Cucurbitales
Clarifying the origin and collinearity of diverse homologous genes is critical to investigating the evolution of essential genes (Soltis et al., 2015). Based on intergenomic and intragenomic comparisons, we identified the orthologous genomic regions produced by species divergence among the 14 genomes from Cucurbitales and V. vinifera (Supplemental Tables S5–S21). We found 12,307 orthologous gene pairs produced by the divergence of Begoniaceae and Cucurbitaceae when comparing the B. loranthoides and $C.$ moschata genomes and 20,066 orthologous gene pairs produced by the divergence of two Cucurbita genomes ($C.$ moschata and C. argyrosperma) (Supplemental Table S21). These results indicated that there is better collinearity between the Cucurbitaceae genomes. Likewise, we identified the paralogous genes produced by ECH, CCT, and CST events within the 15 genomes (Supplemental Table S20). Despite Cucurbitales having experienced at least one more WGD event than V. vinifera, many Cucurbitales plants have fewer ECH-related genes; for example, 1,603 ECH-related paralogous genes were identified in the wax gourd (Benincasa hispida) genome, which is barely half that of V. vinifera. We found obvious differences in the number of CCT-related paralogous genes in Cucurbitaceae, which ranged from 1,910 to 10,435, and small differences in that number in Begoniaceae, ranging from 2,968 to 6,208. We identified 4,151–10,435 paralogous genes in Cucurbita, which is twice that in Begoniaceae, probably due to the late occurrence of the specific polyploidization event in Cucurbita resulting in less genome fractionation.

To explore the homologous genes among species at the Cucurbitales scale, we constructed a hierarchical and event-related multigenome alignment for Cucurbitales by identifying paralogous and orthologous genes, with the V. vinifera genome as the reference (Figure 3 and Supplemental Tables S22 and S23). We filled in the first column with all the gene IDs of V. vinifera and then added the gene IDs of the other species column by column based on genomic homology, creating a multigenome collinear gene table of 135 = $(1 + 8 \times 4 + 6 \times 2) \times 3$ columns. Without taking gene loss into account, each of the three V. vinifera paralogous genes produced by the ECH has four orthologous genes in Begoniaceae and Cucurbita and two orthologous genes in Benincasa and Cucumis. This multigenome alignment associated the paralogs within each genome and out-paralogs between genomes with each polyploidization event and associated the orthologs between genomes with their ancestral populations. Furthermore, to facilitate the utilization of this comparative genome resource, we built an online platform to store the collinear gene table and provided a user-friendly search function (http://www.cgrpoee.top/Cucurbitales/index.html).

Genomic fractionation and likely auto- and allopolyploidization events in Cucurbitales
To investigate the fractionation patterns of BCT-produced subgenomes in Begoniaceae and CST-produced subgenomes in Cucurbita, we estimated that the gene retention and loss levels of the subgenomes in four Begoniaceae genomes (B. loranthoides, Begonia dartervaderiana, B. masoniana, and Begonia peltatifolia) with V. vinifera as reference genome and the C. moschata subgenomes with C. melo as reference genome (Supplemental Tables S24–S27). All these polyploid-produced subgenome alignments of Begoniaceae and Cucurbita are stored in our bioinformatics research platform (http://www.cgrpoee.top/Cucurbitales/Polyploidy_subgenome.html). Compared with the reference genomes, we found that a large number of genes from Cucurbitales genomes are absent from the collinear locations. Genes from each chromosome of the reference showed a varying retention rate $\sim$10%–45% in five Cucurbitales genomes, and the subgenomes of B. loranthoides produced by BCT showed a highly similar retention pattern (Figure 4, A and B). Based on these results, we calculated the gene loss differences between subgenomes of considered species relative to each chromosome from reference genome and then took the average of these values (Supplemental Tables S24 and S25). We found that the BCT-produced subgenomes in four species from Begoniaceae have more balanced gene loss than the CST-produced subgenomes in C. moschata from Cucurbitaceae. For example, we obtained the values of 0.024 for the B. masoniana BCT-produced subgenomes and 0.038 for the C. moschata CST-produced subgenomes (Figure 4, A and B and Supplemental Tables S24 and S25). Further, we performed a statistical analysis of the lengths of chromosome segments with continuous loss of genes in B. loranthoides (Figure 4C and Supplemental Figure S13). Comparing with the V. vinifera genome, the deletions of 15 or fewer consecutive genes accounted for 48.87% of all lost genes in B.
However, using the *C. melo* as reference, the percentage of genes lost reached 55.30%. This result shows that the length of chromosomal regions of genes lost continuously accumulates with the increase of time, which may have changed the fractionation pattern of subgenomes produced by WGDs during the process of species evolution.

Using a sliding window along chromosomes, we further compared gene retention levels of BCT- and CST-produced paralogous regions in four Begoniaceae genomes and *C. moschata*. If the windows with retention rate difference \( \leq 0.05 \) are not considered, in *B. loranthoides*, we found that the percentages of dominant windows to preserve more genes of the two BCT-related subgenomes were 33.93% and 28.90%, with little difference in the retention levels (Figure 4, A and B and Supplemental Table S26). However, in *C. moschata*, the difference in the retention levels between the CST-related subgenomes was obvious, with a nearly two-fold difference of 53.54% and 27.16% (Supplemental Figure S14 and Supplemental Table S27). These results suggest that the gene retention of Begoniaceae BCT-produced subgenomes was balanced, while the CST-produced subgenomes in *C. moschata* was obviously imbalanced.

To quantify the balance of subgenome fractionation, we used the previously developed mathematical measurement algorithm P-index (Wang et al., 2019a). The value of P-indices must be falling in the range of \([0–1]\). Small values of P-indices indicate the similar fractionation patterns (retained gene content, density, and orders) shared by considered subgenomes comparing with the selected reference genome. Here, by using the *V. vinifera* genome as a reference, we
estimated that the P-index values of BCT subgenomes in four Begoniaceae genomes are 0.230, 0.274, 0.207, and 0.276, respectively. The CST subgenomes in *C. moschata* have the P-indices of 0.348 with *C. melo* as a reference. According to the previous study, the P-index /C240.3 could be used as a threshold to classify auto- and allopolyploidies. Because the known and previously inferred alloploidies always have larger P-index > 0.3, including that the *Brassica napus*, *Zea mays*, *Gossypium hirsutum*, and *Brassica oleracea* (Schnable et al., 2011; Chalhoub et al., 2014; Li et al., 2014; Wang et al., 2015a; Renny-Byfield et al., 2017) Those inferred autopolyploidy of *Glycine max*, *Populus trichocarpa*, and *Actinidia chinensis* (Murat et al., 2017; Wang et al., 2017, 2018b) have P-indices < 0.3. Together with prior knowledge and our analysis results, it suggests that the BCT to be a likely autopolyploidy event, while the CST to be a likely allopolyploidy event.

**Ancestral karyotypes of Cucurbitales and chromosome evolutionary trajectory**

The reconstruction of ancestral chromosomes can help reveal the evolutionary process of species that have experienced recursive polyploidizations and the possible effects of chromosome rearrangement. It provides an important reference resource for inferring the evolution of modern genomes. By comparing the 14 Cucurbitales genomes with the *V. vinifera* genome, we reconstructed the ACK-I before CCT event, which contains 17 proto-chromosomes (H1–H17) when the minor fragment of chromosome inversions and deletions are ignored (Figure 5, A and B and Supplemental Figures S15 and S17). To reconstruct the ancestral karyotype of Cucurbitales, we defined the 7 pre-ECH chromosomes of the AEK as G1–G7 and the 21 post-ECH chromosomes as A1–A7, B1–B7, and C1–C7, which correspond to the 19 chromosomes of *V. vinifera* (Figure 5, A and B). Then, we inferred that the 21 chromosomes of the tripled AEK went through two EEJ fusions, two nested chromosome (NCF) fusions, and three nonreciprocal translocations to form ACK-I. For example, we found that the orthologous chromosomes Bh14 of *B. hispida* and B1o1 of *B. loranthoides* shared the orthologous regions Vv4 (C2) and Vv18 (B4) in *V. vinifera*, which were preserved the full ancestral structure of Cucurbitales; we can define this region (C2 + B4) as H15 (Figure 5A). The orthologous chromosomes
Bhi10 and Blo14 shared the orthologous region Vv6 (A6), which were preserved the full ancestral chromosome A6 structure of AEK; we can define this region as H3 (Figure 5A). The orthologous regions of chromosome H3 in Bhi2 and Bhi8 can be explained by chromosome translocation (Figure 5A). Similarly, we inferred the other 15 ancestral chromosomes (H1–H2, H4–H14, and H16–H17) in the ACK-I (Node A) (Figure 5B). After the CCT, the chromosomes of ACK-I doubled, and then the chromosomes with 2 EEJ fusions formed 32 chromosomes in the common ancestor of Cucurbitaceae and Begoniaceae (Node B).

Surprisingly, although Begoniaceae experienced additional WGD after the CCT event, we found that the chromosome numbers of four species in this family ranged from 15 to 19, which are far less than the number of chromosomes in ancestor of Node B. This indicates that the Begoniaceae genomes have affected by large-scale chromosomal rearrangements after the divergence of Cucurbitaceae and Begoniaceae. By comparing the Begoniaceae genomes with the *V. vinifera* genome, we deduced the ancestral Begoniaceae karyotype (ABK) pre-BCT (Node C) and post-BCT (Node D) (Figure 5B and Supplemental Figures S18–S21). After the CCT event, through 12 EEJ fusions and 4 NCF fusions, the 32 chromosomes of the ACK-I were reduced to the 16 chromosomes of ABK (Node C). After the BCT event, the ancestral Begoniaceae genome doubled, and through nine EEJ fusions and two NCF fusions, the ABK reduced to 21 chromosomes (Node D). Among four Begoniaceae genomes, we found that the *B. loranthoides* has the fewest specific chromosomal rearrangements, involving three EEJ chromosome fusions and five nonreciprocal translocations (Supplemental Figures S18–S21).

Further, to investigate previous studies reported that the ACK-II has 15, or 12 chromosomes (Wu et al., 2017; Zhao et al., 2021), we compared the Cucurbitaceae genomes with the *V. vinifera* genome to reanalyze the most recent common ancestor of Cucurbitaceae. We confirmed that the ACK-II (Node E) has 15 chromosomes, which were reduced from the 32 chromosomes in ancestor of Node B through 12 EEJ fusions, 5 NCF fusions, and 2 nonreciprocal translocations (Figure 5B and Supplemental Figures S22–S31). Then, we checked that the chromosome fissions in identified previous study (Wu et al., 2017), and found that the many regions involving in fission events were independent ancestral chromosomes of Cucurbitaceae rather than the regions derived from species-specific fissions (Supplemental Figure S32). For example, the chromosome Cme11 of *C. melo* evolved from an ancestral chromosome of ACK-II in previous study. However, we found that the orthologous regions of Cme11 are separated in Cmo2, Cmo20, Cmo8, and Cme17 of *C. moschata*, and the orthologous regions of Cme11 are also separated in different chromosomes in bottle gourd (*Lagenaria siceraria*) and *C. lanatus*. Further, we detected that the location of the breakpoint of Cme11 (27.5 Mb) is similar when comparing *C. melo* with the genomes of *C. moschata*, *L. siceraria*, and *C. lanatus*, indicating that the Cme11 was fused by at least two ancient chromosomes of ACK-II. We also inferred that the most recent common ancestor of *Benincasa* and *Cucumis* (Node G) contains 15 chromosomes involving 7 nonreciprocal translocations from ACK-II. We further confirmed that the ancestral *Cucumis* karyotype (ACK-IV) contains 12 chromosomes, clearly confirming that the melon maintained the ancestral
and 14 Cucurbitales plants (Figure 6B and Supplemental Table S28). We found that the number of family members in the genomes of Begoniaceae after the BCT event, which made the number of modern chromosomes close to that of the ABK pre-BCT, while the more chromosome translocations occurred in Cucurbitaceae plants. We found that the genomes of Begoniaceae retain the chromosome structure of the AEK more completely than that of the Cucurbitaceae genomes (Figure 5, B and Supplemental Figures S18–S31). In addition, Cucurbita experienced additional WGD event after the CCT, which was resulted in the most chromosomal rearrangements among Cucurbitaceae; these rearrangements involved one chromosome fusion and four nonreciprocal translocations before the CST events and eight chromosome fusions, four nonreciprocal translocations, and one reciprocal translocation after the CST event (Figure 5, B and Supplemental Figures S28–S31).

Expansion and diversity evolution of the gene families involved in the cucurbitacin synthesis pathway

Cucurbitacins constitute a group of diverse triterpenoid substances that are well known for their bitterness, and they have been mainly reported in Cucurbitaceae plants (Chen et al., 2005). According to the chemical formula, the cucurbitacins can be divided into 14 types (A–N), of which the most famous is cucurbitacin B, which has been proved to have anticancer effects (Andeweg and Bruyn, 1959; Kaushik et al., 2015). The studies reported that the eight gene families of cytochrome P450 (P450), squalene epoxidase (SQE), squalene synthase (SQS), oxidosqualene cyclase (OSC), isopentenyl diphosphate isomerase, farnesyl diphosphate synthase (FPS), geranyl pyrophosphate synthase (GPS), and acetyltransferase (ACT) proteins play a crucial role in the regulation of cucurbitacin synthesis (CS), especially P450, ACT, and OSC have a decisive role in determining the final class of cucurbitacin synthesized (Osbourn, 2010; Hamberger and Bak, 2013; Thimmappa et al., 2014; Figure 6A).

To explore the expansion patterns of these families, we identified a total of 4,074 family members in the genomes of V. vinifera and 14 Cucurbitacaceae plants (Figure 6B and Supplemental Table S28). We found that the number of family members varied between two Cucurbitales families; for example, the number of P450 family members (mean: 202.3) in Cucurbitaceae was greater than that in Begoniaceae (mean: 152.0). Within the genomes, the number of family genes varied considerably; for example, the P450 family of L. siceraria has 249 members, substantially more than the GPS family only has 2 members. When comparing Cucurbitales plants with V. vinifera, we found that the genes in families SQS, ACT, and SQE show expansion pattern, which may be contributed by the CCT events. However, the genes of families OSC and P450 show reduced pattern, and the other three families have a similar number of genes. Then, combined with the identified homology information and the phylogenetic relationship of genes in families, we calculated the WGD expansion rate (WGD-ER), tandem duplication expansion rate (TD-ER), and the contraction rate (CR) caused by gene loss for each family (Figure 6C; Supplemental Figures S35–S37 and Supplemental Table S29) (see “Materials and methods” for details). We detected that the families with expansion patterns in Cucurbitales genomes often show that the duplication expansion rate is higher than the CR. For example, the SQS of B. darthvaderiana has a duplication expansion rate 100% (WGD-ER: 50% and TD-ER: 50%, respectively), which is higher than that the CR with 75%. In contrast, the families with contraction patterns often show that the duplication expansion rate is lower than the CR. For example, the OSC of B. darthvaderiana has a duplication expansion rate of 50% (TD-ER), which is lower than that the CR with 200%. Surprisingly, we also found that the OSC has a very active TD-ER to be 600% in V. vinifera. All these results suggest that the WGD and tandem duplication events provided the genetic basis for the expansion of the families in Cucurbitales, but the loss of duplicated genes limited the family expansion, and even appears contraction. Additionally, the families in Cucurbitales show contraction patterns most likely explained by the high frequency of gene loss in Cucurbitales and active tandem gene duplication in V. vinifera.

Further, we identified the gene clusters in family P450, a largest family in CS pathway, and found that the most clusters gradually “growing” during the diversification of Cucurbitales plants (Figure 6D and Supplemental Figures S38–S40). In C. moschata, we detected four largest clusters of P450 located on the Cmo04, Cmo09, and Cmo18 chromosomes, which are containing 14, 20, 11, and 15 genes, respectively (Figure 6D). Within the four representative genomes of L. siceraria, C. moschata, B. loranthoides, and V. vinifera, the sparse family genes and few clusters in P450 were distributed near the location of chromosomal rearrangements, and the clustered genes seem to prefer to nest in the middle regions of their ancestral chromosomes (Supplemental Figures S41–S44). Through intergenomic comparisons of these genomes, we found that the clusters of one species in another often show a decrease in the number of genes associated with chromosomal rearrangements (Figure 6E and Supplemental Figure S45). For example, the two clusters α1 and α2 on paralogous chromosomes 4 and 18 of V. vinifera have 14 and 34, respectively. In C. moschata, the two orthologous clusters β1 and β2 of α1 and α2 on chromosomes 1 and 9, which are related to an ancestral rearrangement at the base of Cucurbitales; while the two clusters β1 and β2 only have 10 and 27 genes, respectively (Figure 6E). These results suggested that the chromosomal rearrangements have an inhibitory effect on gene family expansion.
Figure 6 Evolution and diversification of gene families involved in the CS pathway in Cucurbitales. A, Cucurbitacin synthesis pathway. B, Identification of the gene families involved in the CS pathway in the studied genomes. C, Phylogenetic tree of the genes in FPS family of the Cucurbitales and V. vinifera genomes. Genes from different species are marked with different colors. The dots on the three layers represent the gene tandem duplication, paralogous, and orthologous relationships (marked with a, b, and c, respectively), and the same color represents the same group of relationships within and between genomes. D, P450 gene amplification model related to gene duplication events in C. moschata. P450 genes are displayed in blue patches. Curved lines (gray) link the collinear gene pairs generated at the period from the divergence between V. vinifera and Cucurbitales to the CCT event. Curved lines (red and plum) link gene pairs generated at the period from now to the CST and from the CST to the CCT event, respectively. E, The chromosomal distributions of P450 genes in the selected genomes. The lines between gene pairs on the same chromosome represent homologous genes generated at the period from the divergence between V. vinifera and CCT event. The lines between gene pairs on the different chromosome represent homologous genes generated by the divergence between the species. The β1, β2, γ1, and γ2 represent the gene clusters. F, The CDS–UTR structures of a set of orthologous FPS genes in selected genomes. The genes with different colors present the different species, and the shadow in one species shows that they come from the different subgenomes generated by the polyploidization event.
In addition to the change of gene copy numbers, the gene structure variations (GSVs) in these families present another prominent evolutionary feature. To clearly compare the GSVs in Cucurbitales, we selected five relatively small families (FPS, GPS, OSC, SQS, and SQE) and visualized the coding sequence (CDS) and untranslated region (UTR) of all genes on the phylogenetic trees (Supplemental Figures S46–S50). Among the five families, the gene structure in FPS is relatively conservative, because the genes in FPS can be clustered into two categories according to the number of CDS and phylogenetic relationship, while the genes in other families are clustered into three to five categories. For CDS–UTR structures, we found that the diversity of genes in Cucurbitaceae (each gene contains 1–6, or 9–33 CDS) was higher than Begoniaceae (each gene contains 1–3, or 10–12 CDS), and the Cucurbitaceae genes often contained 1–2 UTR (Supplemental Figures S46–S50). Next, we detected that the structures of genes in CCT-produced subgenomes of Begoniaceae with more conservative than their orthologous gene structures in Cucurbitaceae. For example, the paralogous gene pairs Bpe09g00111 and Bpe03g00747 in B. peltatifolia contain 12 and 12 CDS, respectively, but their orthologous genes Lsi02g01811 and Lsi02g01180 in L. siceraria contain 18 and 12 CDS, respectively, and have 0–2 UTR distributed in each gene (Figure 6F). These results implied that the chromosomal rearrangements might result in the biased genomic fractionation which likely promoted the diversity of gene structures.

**Discussion**

**Overlooked CCT in Begoniaceae**

Polyploidy is important in genome evolution and functional innovation and is common in flowering plants (Jiao et al., 2012; Soltis et al., 2015; Van de Peer et al., 2017; Landis et al., 2018); however, increasing genome complexity makes polyploidization events difficult to identify (Paterson et al., 2004; Soltis et al., 2015). For a recently sequenced genome, comparison to well-characterized genomes of related plants helps deconvolute the genome structure and evolutionary history (Schnable et al., 2011; Murat et al., 2014; Wang et al., 2016b). Here, we performed a comparative whole-genome analysis of Cucurbitaceae and the recently sequenced genomes of Begoniaceae, and confirmed that Begoniaceae underwent two consecutive polyploidizations after the ECH event, rather than only one polyploidization (Li et al., 2022). Our previous study showed that all Cucurbitaceae species share a tetraploidization event (Wang et al., 2018a). Based on homology comparison analysis, we found that the common tetraploidization of Cucurbitaceae was the same WGD event that was overlooked in Begoniaceae (Li et al., 2022), which we re-defined as the CCT.

In this study, we detected that the ECH and CCT events occurred at ~115–130 and ~93–105 MYA, respectively. Additional paralogs from more ancient polyploidization events are sometimes found within homologous regions formed in some well-characterized genome duplications, but they are more often obscured by subsequent reduplications and rediploidization (Jiao et al., 2014). Thus, it is likely that the paralogous homologous regions generated by the CCT mask the paralogous homologous regions generated by the ECH, and the two polyploid events are close in time, which eventually makes the two polyploidization events indistinguishable. In addition, we infer that Cucurbitaceae split from Begoniaceae at ~90–101 MYA. The corrected Ks distributions of the Cucurbitaceae–Begoniaceae orthologs and the Cucurbitales CCT paralogs overlap with one another. This implies that CCT events may have directly contributed to the split of Cucurbitaceae from Begoniaceae. Similarly, it is possible that the CST event led to the split of two branches of C. moschata and C. argyrosperma.

**Hypothesis of the nature of polyploidization events in Cucurbitales**

Polyploidy is indispensable in the evolutionary process of plants, with nearly 90% of species possibly being allopolyploid and only a small fraction being autopolyploid (Chen, 2010; Wang et al., 2019a). Autopolyploid was formed by duplication of genomes, whereas allopolyploid was generated from the crosses between different species with independent evolutionary histories, both of which are followed by chromosome doubling or fusion of unreduced gametes. It suggests that the WGD events can be divided into auto- and allopolyploidies, according to the unbiased and biased fractionation of genomes produced by the genome doubling (Garsmeur et al., 2014). Here, our findings on gene retention and loss among BCT-generated subgenomes suggest that there is little dominance among individual subgenomes in Begoniaceae, which raised the possibility that the BCT event was likely autotetraploidy in nature, as demonstrated in soybean and Salicaceae (Liu et al., 2017; Wang et al., 2017). In contrast, the CST-produced subgenomes of C. moschata showed a biased fractionation pattern, once again demonstrates that the genus Cucurbita has a likely common allotetraploid ancestor (Sun et al., 2017a). All of these Cucurbitales plants shared a likely allopolyploidization event has demonstrated by previous study (Wang et al., 2018a).

We observed that the Begoniaceae has a lower evolutionary rate than the Cucurbitaceae. The most likely explanation is that the Begoniaceae affected by an additional autotetraploidization BCT after the CCT event. As a sharp contrast, the Cucurbita evolved faster than Begoniaceae and almost all other Cucurbitaceae species (Figure 2A), probably because they experienced another round of allotetraploidization CST after the CCT event. Generally, allopolyploidization would increase the evolutionary rate of genomes due to homologous recombination and the presence of additional gene copies (Wang et al., 2018b). However, autotetraploidization results in multiple copies of chromosomes, which may prolong cell cycle time and thus generation time. Allopolyploids, doubled chromosomes, exhibiting diploid inheritance, but absent an evidence for prolonging the time of cell cycle and generation (Doyle and Egan, 2010). Plants
with slower rates of molecular evolution tend to be perennial, have long generation times, and produce taller plants (Gaut et al., 2011; Lanfear et al., 2013), which once had the characteristics of autoploidy history, such as Salicaceae and Rosaceae (Liu et al., 2017; Li et al., 2019b; Wang et al., 2019a). Begoniaceae species with slower rates of evolution tend to have these characteristics.

Although we used the method of genome fractionation bias commonly used in angiosperms to infer the nature of polyploidization events in Cucurbitales, and found some evolutionary characteristics of the genome consistent with the nature of the corresponding paleopolyploid, it is still insufficient as evidence to distinguish auto- and allopolyploidy. First, subgenomes produced by similar polyploidization events may result in different fractionation patterns in different lineages. For example, the allotetraploid Ethiopian cereal teff without clear signs of fractionation (vanBuren et al., 2020), but in the autotetraploid Teleost fish, it is surprising that the biased fractionation exists between subgenomes (Parey et al., 2022). Second, the neopolyploids differences in gene content between extant and progenitor species complicate simple interpretation of gene content patterns. One possible reason is that the biased subgenome of extant species mainly reflects inherit from the progenitors, and this difference can be detected before the polyploidization events (Liu and Wang, 2022). Finally, it should be noted that paleopolyploidization events often occurred a few, or even tens of MYA, which brings great challenges to assign collinear blocks to the subgenomes of polyploidy association (Emery et al., 2018). Although the current work utilizes the recently constructed pipeline of complex genome structure analysis (Wang et al., 2018a, 2019a), more effective tools are still need to be developed to clearly differentiate auto- and allopolyploidy.

Paleochromosomes and the chromosome evolutionary trajectory of Cucurbitales

Ancestral chromosome reconstruction is important for confirming the phylogenetic position of species and revealing the impact of different polyploidy events on genomic diversification (Murat et al., 2017). The karyotypes of Cucurbitales have complex chromosomal rearrangements (inversions, translocations, fusions, and fissions) resulting from recursive polyploidization events that are difficult to deduce clearly (Ren et al., 2019). Here, we reconstructed the 17 proto-chromosomes ancestral karyotype of Cucurbitales plants and investigated the karyotype evolutionary trajectory of representative Cucurbitales plants. The reconstruction of the ACK-I provides a reference for studying the origin and evolution of Cucurbitales plants, enabling us to better understand the evolutionary history of Cucurbitales plants at the chromosome level.

In addition, based on the AEK (Jaillon et al., 2007), we clarified the 15 proto-chromosomes of the ACK-II, and rejected the previous hypotheses of the ACK-II with 12 proto-chromosomes (Wu et al., 2017). We also confirmed that the ancestral Cucumis karyotype has 12 chromosomes, which is consistent with a recent report (Zhao et al., 2021). Using the same approach, we also deduced the ABK before and after the BCT event, that filling a gap in the chromosome karyotype evolution of Begoniaceae plants. Furthermore, we found that Cucurbitaceae exhibited more chromosomal rearrangements than Begoniaceae, resulting in greater diversity, and Begoniaceae genomes contained more complete ancestral chromosomal structures, which suggest they may become model plants for future genomic studies in the Cucurbitales. Notably, the main mode of chromosome fusions in Cucurbitales is the EEJ pattern, which is the same as that in most eudicots, such as V. vinifera and coffee (Jaillon et al., 2007; Li et al., 2019a).

Genetic evolution and diversity of important gene families

Cucurbitacin has excellent antitumor and antipest qualities, and its synthetic pathway involves many important proteases. Based on its functional importance, cucurbitacin has been studied in cucumber and muskmelon (Qing et al., 2022; Zhang et al., 2022). Cucurbitacins have been mainly reported in the Cucurbitaceae plants and can cause bitterness in many plants (Chen et al., 2005). We explored the expansion patterns of gene families involved in the Cucurbitacin synthesis pathway, and found the three expansion patterns of families including the increase, decrease, and basically stable maintenance of gene copy number. Our findings revealed that the WGD and tandem duplication events provided the genetic basis for the expansion of the families in Cucurbitales, but the gene loss and chromosomal rearrangements may be inhibiting the family expansion.

Gene structural diversity may promote protein functional diversity. The CDS contains exons that encode protein sequences, and the UTR is a noncoding region at either end of the mRNA that contains many different regulatory elements involved in variable splicing (Mayr, 2019; Navarro et al., 2021). For the gene structures of the families in Cucurbitales involved in the CS pathway, we found that the Cucurbitaceae shows more diversity than Begoniaceae. Although the chromosomal rearrangements limit the family expansion, we suggested that it is like a booster for the structural diversity of family genes. The reason for this speculation is that the biased subgenome fractionation of species may require more structural variations of genes to compensate for their adaptability to the environment. In this study, we demonstrated that the biased subfractionation exists in Cucurbitaceae and the balanced fractionation in Begoniaceae. The difference of subgenome fractionation may be positively related to the degree of chromosomal rearrangement, as we here detected that the Cucurbita genomes have more rearrangements than that the Begoniaceae genomes. Based on this analysis, although preliminary evidence supports our hypothesis, more in-depth research is needed in the future on how chromosomal rearrangement affects the genetic diversity and adaptability of gene families.
Materials and methods

Availability of data and materials
Genome sequences and annotation files were downloaded from their respective online public databases. Detailed data sources are provided in Supplemental Table S1.

Inference of gene collinearity
For gene collinearity analyses, we used BLASTP (Altschul, 2012) software to search the potential homologous gene pairs for each protein within and between genomes and set strict parameters (e-value < 1e−5 and score > 100). Then, the top 10 BLAST matches were selected as input for ColinearScan (Wang et al., 2006) to infer the collinear blocks. We set the maximum gap between the anchored genes to 50, and removed the gene copies more than 30 in a large gene family. This method has been commonly used in the comparative genome research of many angiosperms (Wang et al., 2017, 2018b, 2019a, 2019b).

Calculation and kernel function analysis of the Ks
The Ks of collinear gene pairs was calculated by the Nei–Gojobori approach (Nei and Gojobori, 1986) implementing the Bioperl statistical module (https://bioperl.org/). For the Ks density distribution curve of collinear gene pairs, we used the kernel smoothing density function ksdensity (the parameter set to 0.5) in MATLAB (Cichosz et al., 2020) to evaluate the probability density of each Ks list. Then, we used the Gaussian approximation function Gaussian in the fitting toolbox cftool to perform a Gaussian multipeak fitting of the curve (https://www.mathworks.com/help/curvefit/index.html), and analyzed the distribution of Ks for homologous genes within and between genomes. The parameter R-squared was used to evaluate the goodness of fit and set to at least 95%, the minimum number of positive-terminus distributions was used to represent the complex Ks distribution, and principle 1 was used to represent the corresponding evolutionary event.

Identification of orthology and paralogy
The gene collinearity dotplots are helpful for identifying the orthologous and paralogous chromosomal regions within and between genomes. We plotted the collinear gene pairs according to their genomic locations in dotplots, and used different color-coded dots to distinguish whether the anchor gene pairs are the best BLAST hit among the genomes. For collinear regions within genomes, if the median Ks of gene pairs is approximately equal to the Ks peak value of speciation, then this region is identified as an orthologous region. For a collinear region within genome, if the median Ks of gene pairs is approximately equal to the Ks peak value of polyploidization, then this region is identified as a paralogous region.

Correction of the evolutionary rates
The studied genomes evolved at different rates, making the timing of key evolutionary events difficult. Here, we employed an evolutionary rate correction algorithm commonly implemented in angiosperms, such as Fabaceae (Wang et al., 2017), Cucurbitaceae (Wang et al., 2018a), Durian (Wang et al., 2019b), to redate the key evolutionary events of Cucurbitales. The first step in the correction of evolutionary rates on the basis of ECH event, the Ks distribution peaks of ECH-produced duplicated genes in Cucurbitales aligned to have the same value with that of slowest evolving V. vinifera. We estimated the maximum-likelihood (ML) value μ through characterized the median Ks of duplicated gene pairs in each block associated with ECH event in studied genomes. Supposing that the Ks distribution of V. vinifera obeys a normal function $X_V \sim (\mu_V, \sigma_V^2)$, and the Ks distribution of specific Cucurbitales plant C obeys function $X_C \sim (\mu_C, \sigma_C^2)$. The relative difference of evolutionary rates between the genomes of Cucurbitales C and V. vinifera can be calculated with

$$r = (\mu_C - \mu_V)/\mu_V.$$  

We can define the Ks distribution of Cucurbitales plants C as

$$X_{C-correction} \sim (\mu_{C-correction}, \sigma_{C-correction}^2).$$

Then, the correction coefficient can be calculated as

$$\mu_{C-correction} = \frac{\mu_V \times \mu_C}{1 + r} \times \mu_C,$$

and, therefore, we get

$$X_{C-correction} \sim (\lambda_C \mu_C, \lambda_C^2 \sigma_C^2).$$

For the divergence of two Cucurbitales plants a and b, we can set the correction coefficient to be $\lambda_{ab} = \frac{\lambda_a + \lambda_b}{2}$, and supposing that the Ks distribution was $X_{ab} \sim (\mu_{ab}, \sigma_{ab}^2)$, then we calculated a corrected evolutionary rate

$$X_{ab-correction} \sim (\lambda_{ab} \mu_{ab}, \lambda_{ab}^2 \sigma_{ab}^2).$$

Specifically, for the divergence of V. vinifera and Cucurbitales plant a, we can set the correction coefficient to be $\lambda_{av} = \frac{\lambda_a + \lambda_v}{2}$, and supposing that the Ks distribution is $X_{av} \sim (\mu_{av}, \sigma_{av}^2)$, then, we calculated a corrected evolutionary rate

$$X_{av-correction} \sim (\lambda_{av} \mu_{av}, \lambda_{av}^2 \sigma_{av}^2).$$

After a correction of Ks of ECH-produced genes, the CCT-produced homologs in Cucurbitales still had very divergent Ks distribution. Therefore, similar to the first step correction, a further round of corrections was performed by aligning the Ks distributions peaks of CCT-produced homologs of Cucurbitales to have the same distribution means. After two rounds of correction, we performed a third round of correction for Cucurbita based on CST event. This correction
algorithm has been compiled into a calculation script with a detailed operation process noted and deposited in Github (https://github.com/wangjiaqi206/corrected-evolution-dating).

Construction of the event-related collinear gene table

The table was constructed using the *V. vinifera* genome as a reference, and all *V. vinifera* genes were placed in the first column. Each *V. vinifera* gene has two additional collinear genes due to the ECH event, and we specified two additional columns in the table to contain this information. For a gene in *V. vinifera*, if a corresponding orthologous gene appears in the expected position of Cucurbitales plants, the gene ID is filled into the cell of the corresponding column. When it is missing, usually owing to the absence or translocation of a gene in the genome, we place a dot in the cell of the corresponding column. Thus, for each species in Cucurbitales, a 135-column table was constructed with the gene IDs of these species filled sequentially according to homology (Supplemental Table S23). We also constructed a collinear gene table containing 33 columns of the three major Cucurbitales species, with *V. vinifera* as the reference (Supplemental Table S22).

Gene losses

Using the *V. vinifera* and *C. melo* genomes as references, we counted genes in the reference genome that did not have collinear orthologs in each Cucurbitales chromosomal region. The number of consecutive genes in the reference genome that had no Cucurbitales match was also counted. This "run length"—that is, the number of consecutive genes in the reference genome with no matching Cucurbitales gene—should follow a geometric distribution if gene loss/translocation occurred in a random manner. For a probability p that a gene was lost at random, the probability of loss of a consecutive run of n−1 genes can be estimated by the formula

\[ y = f(x = k|p) = (1 - p)^{(n-1)}p. \]

The fitting functions curve-fit and NumPy were run using Python scripts.

Calculation of the P-index

The polyploidy index (P-index) is a statistical approach for quantifying fluctuations in gene retention differences and provides a mathematical method to evaluate the similarity between homoeologous chromosomes in polyploids (Wang et al., 2019a). We used *V. vinifera* as a reference to calculate the P-index of Begoniaceae and *C. melo* as a reference to calculate the P-index of *C. moschata*, where the number of sliding windows was 50 and the parameter was 0.05−1.8 (Wang et al., 2019a). The core calculation formulas are as follows:

\[ P\text{-index} = \sum_{c=1}^{N_c} W_{abs} \left[ \frac{\sum_{i=1}^{N_c} \frac{A_i - B_i}{(A_i - B_i)^2} \times \delta_i}{N_c - \delta(N_c)} \right]. \]

More details can be found in previous article (Wang et al., 2019a).

Inference of ancestral karyotype and chromosome evolution

In the case of inference of the evolutionary trajectories of chromosomes, we used a "Telomere-centric genome repatterning model" proposed in previous study (Wang et al., 2015b). The main steps for tracking the chromosome rearrangements including: (1) generating syntenic dotplots and identifying orthologous genomic regions between the genomes of Cucurbitales and *V. vinifera*; (2) inferring the involved regions of the genomic rearrangements in compared genomes and revealing the connection and order of these orthologous regions; (3) reconstructing the ancestral state of these involved regions for the compared species based on orthologous regions in living species; and (4) comparing the chromosomes in each considered species with the predicted pattern of the most common ancestor of extant plants and identifying the genomic rearrangements during the evolution of Cucurbitales plants.

The dynamic evolutionary trajectory of chromosomes for the Cucurbitales plants was illustrated using the Adobe Animate (Adobe) software (Tepla and Klimova, 2015). The colors in the produced video match the colors assigned for the seven chromosomes in the AEK. The dynamic evolutionary trajectory of chromosomes in Cucurbitales has been deposited in the online platform (http://www.cgrpoee.top/Cucurbitales/index.html).

Identification and evolutionary analysis of the gene families

To identify the protein sequences of gene families involved in the CS pathway, the domains for eight families were downloaded from the PFAM protein dataset (http://pfam.xfam.org/) (Mistry et al., 2021; Supplemental Table S28). The hidden Markov model of HMMscan software (Chai et al., 2019) was employed to search the proteins of family genes from the genomes of Cucurbitales plants and *V. vinifera*, and setting a strict expectation (E-value < 1e−10).

For the phylogenetic analysis of gene families, we performed a multiple sequence alignment on the proteins of each family using the MAFFT program (Katoh et al., 2002; Katoh and Standley, 2013), and constructed the gene trees using the IQ-TREE software (Nguyen et al., 2015). Through the ModelFinder module in IQ-Tree, we obtained the best nucleotide substitution model for each family (Supplemental Table S29). Finally, the ML algorithm was employed to construct gene trees, and the bootstrap value was set to 1,000 to test the stability of the family gene trees. The gene tree of each family was displayed using an online tool Evolvview (https://www.evolgenius.info/evolvview/#/treeview) (Subramanian et al., 2019). Moreover, for the expansion rates of the whole-genome and tandem duplication (WGD-ER and TD-
ER) and CRs of the family in Cucurbitales, supposing that the family of V. vinifera has \( N = \frac{(\text{total number of genes}) - (\text{number of the newly gained genes from tandem duplication})}{\text{genes}} \), we designed the following algorithm:

1. WGD-ER = \( \frac{\text{total number of the newly gained genes from polyploidizations}}{N} \times 100\% \).
2. TD-ER = \( \frac{\text{number of the newly gained genes from tandem duplication}}{N} \times 100\% \).
3. CR = \( \frac{\text{number of V. vinifera genes lost in Cucurbitales}}{N} \times 100\% \).

In addition, identification of the gene clusters in families, we assigned the genes on each genome according to the physical positions. Then, we grouped all the genes into potential clusters, if the distance of a query gene to genes in potential clusters less than 2 Mb. To explore the relationship of chromosomal rearrangements and the expansion of gene cluster in families, we visualized the inferred paleochromosome rearrangements, homologous information, and the gene locations using TBtools (Chen et al., 2020). The gene structures of the families were analyzed through an online tool GSDS (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015).

Accession numbers
Sequence data from this article can be found in their respective online data libraries. The accession number of genome sequence is provided in Supplemental Table S1. The detailed DNA sequences sources involving in studied families are provided in Supplemental Table S28.

Supplemental data
The following materials are available in the online version of this article.

**Supplemental Figure S1.** Intergenomic comparative analyses of the V. vinifera and B. loranthoides genomes.

**Supplemental Figure S2.** Intergenomic comparative analyses: syntenic dotplot between the V. vinifera and B. loranthoides genomes.

**Supplemental Figure S3.** Intragenomic comparison analyses of the B. loranthoides genome.

**Supplemental Figure S4.** Intragenomic comparative analyses: syntenic dotplot within the B. loranthoides genome.

**Supplemental Figure S5.** Intergenomic comparison analyses of the V. vinifera and C. melo genomes.

**Supplemental Figure S6.** Intergenomic comparative analyses: syntenic dotplot between the V. vinifera and C. melo genomes.

**Supplemental Figure S7.** Intergenomic comparison analyses of the C. melo and B. loranthoides genomes.

**Supplemental Figure S8.** Intergenomic comparative analyses: syntenic dotplot between the C. melo and B. loranthoides genomes.

**Supplemental Figure S9.** Intragenomic comparison analyses of the C. moschata genome.

**Supplemental Figure S10.** Intergenomic comparative analyses: syntenic dotplot within the C. moschata genome.

**Supplemental Figure S11.** Intergenomic comparison analyses of the C. melo and C. moschata genomes.

**Supplemental Figure S12.** Intergenomic comparative analyses: syntenic dotplot between the C. melo and C. moschata genomes.

**Supplemental Figure S13.** Gene loss rates of the B. loranthoides genome.

**Supplemental Figure S14.** Gene retention of C. moschata subgenomes along orthologous regions corresponding to C. melo chromosomes.

**Supplemental Figure S15.** Inference of the ancestral Cucurbitales karyotype (ACK-I) by comparing the genomes of C. melo (Cme), Cucumis hystrix (Chy), Cucumis sativus (Csa), and L. siceraria (Lsi) with the V. vinifera (Vvi) genome. ACK-II, ancestral Cucurbitaceae karyotype.

**Supplemental Figure S16.** Inference of the ancestral Cucurbitales karyotype (ACK-I) by comparing the genomes of C. moschata (Cmo), Cucurbita maxima (Cma), C. argyrosperma (Car), and Cucurbita pepo (Cpe) with the V. vinifera (Vvi) genome.

**Supplemental Figure S17.** Inference of the ancestral Cucurbitales karyotype (ACK-I) by comparing the genomes of C. lanatus (Cla), B. masoniana (Bma), B. peltatifolia (Bpe), and B. darthvaderiana (Bda) with the V. vinifera (Vvi) genome.

**Supplemental Figure S18.** Chromosome evolutionary trajectories of the B. loranthoides genome.

**Supplemental Figure S19.** Chromosome evolutionary trajectories of the B. masoniana genome.

**Supplemental Figure S20.** Chromosome evolutionary trajectories of the B. peltatifolia genome.

**Supplemental Figure S21.** Chromosome evolutionary trajectories of the B. darthvaderiana genome.

**Supplemental Figure S22.** Chromosome evolutionary trajectories of the B. hispida genome. ACK-II, ancestral Cucurbitaceae karyotype.

**Supplemental Figure S23.** Chromosome evolutionary trajectories of the L. siceraria genome.

**Supplemental Figure S24.** Chromosome evolutionary trajectories of the C. lanatus genome.

**Supplemental Figure S25.** Chromosome evolutionary trajectories of the C. melo genome.

**Supplemental Figure S26.** Chromosome evolutionary trajectories of the C. hystrix genome.

**Supplemental Figure S27.** Chromosome evolutionary trajectories of the C. sativus genome.

**Supplemental Figure S28.** Chromosome evolutionary trajectories of the C. maxima genome.

**Supplemental Figure S29.** Chromosome evolutionary trajectories of the C. moschata genome.

**Supplemental Figure S30.** Chromosome evolutionary trajectories of the C. pepo genome.
Paleotetraploidization events in Cucurbitales

Supplemental Figure S31. Chromosome evolutionary trajectories of the C. argyrosperma genome.

Supplemental Figure S32. Intergenomic comparison of C. hystrix (Chy), C. moschata (Cmo), C. lanatus (Cla), and L. siceraria (Lsi) genomes with the C. melo (Cme) genome.

Supplemental Figure S33. Intergenomic homologous structure comparison analyses between the C. melo and C. hystrix genomes.

Supplemental Figure S34. Intergenomic homologous structure comparison analyses between the C. melo and C. sativus genomes. ACK-IV, ancestral Cucumis karyotype.

Supplemental Figure S35. Inference of the expansion pattern of GPS (A) and FPS (B) gene families by comparing Cucurbitales with V. vinifera genes.

Supplemental Figure S36. Inference of the expansion pattern of SQS (A) and OSC (B) gene families by comparing Cucurbitales with V. vinifera genes.

Supplemental Figure S37. Inference of the expansion pattern of the SQE gene family by comparing Cucurbitales with V. vinifera genes.

Supplemental Figure S38. P450 gene amplification model related to gene duplication events in B. masoniana (A), B. loranthoides (B), B. darthvaderiana (C), and B. peltatifolia (D).

Supplemental Figure S39. P450 gene amplification model related to gene duplication events in C. argyrosperma (A), C. maxima (B), C. pepo (C), and V. vinifera (D).

Supplemental Figure S40. P450 gene amplification model related to gene duplication events in B. hispida (A), C. lanatus (B), L. siceraria (C), C. melo (D), C. hystrix (E), and C. sativus (F).

Supplemental Figure S41. Chromosomal distributions of P450 genes in L. siceraria.

Supplemental Figure S42. Chromosomal distributions of P450 genes in C. moschata.

Supplemental Figure S43. Chromosomal distributions of P450 genes in B. loranthoides.

Supplemental Figure S44. Chromosomal distributions of P450 genes in V. vinifera.

Supplemental Figure S45. Chromosomal distributions of P450 genes in V. vinifera and three representative Cucurbitales genomes.

Supplemental Figure S46. CDS–UTR structures of FPS family genes in the studied genomes.

Supplemental Figure S47. CDS–UTR of GPS family genes in the studied genomes.

Supplemental Figure S48. CDS–UTR structure of SQE family genes in the studied genomes.

Supplemental Figure S49. CDS–UTR structure of SQS family genes in the studied genomes.

Supplemental Figure S50. CDS–UTR structure of OSC family genes in the studied genomes.

Supplemental Table S1. Information on genomic data.

Supplemental Table S2. Number of homologous genes and blocks within a genome and between genomes.

Supplemental Table S3. Kernel function analysis of Ks distribution related to duplication events within each genome and between selected genomes (before evolutionary rate correction).

Supplemental Table S4. Kernel function analysis of Ks distribution related to duplication events within each genome and between selected genomes (after evolutionary rate correction).

Supplemental Table S5. Orthologous genomic regions between C. melo and C. moschata.

Supplemental Table S6. Orthologous genomic regions between V. vinifera and B. darthvaderiana.

Supplemental Table S7. Orthologous genomic regions between V. vinifera and B. hispida.

Supplemental Table S8. Orthologous genomic regions between V. vinifera and B. loranthoides.

Supplemental Table S9. Orthologous genomic regions between V. vinifera and C. argyrosperma.

Supplemental Table S10. Orthologous genomic regions between V. vinifera and C. hystrix.

Supplemental Table S11. Orthologous genomic regions between V. vinifera and C. lanatus.

Supplemental Table S12. Orthologous genomic regions between V. vinifera and C. maxima.

Supplemental Table S13. Orthologous genomic regions between V. vinifera and C. melo.

Supplemental Table S14. Orthologous genomic regions between V. vinifera and C. moschata.

Supplemental Table S15. Orthologous genomic regions between V. vinifera and C. pepo.

Supplemental Table S16. Orthologous genomic regions between V. vinifera and C. sativus.

Supplemental Table S17. Orthologous genomic regions between V. vinifera and C. argyrosperma.

Supplemental Table S18. Orthologous genomic regions between V. vinifera and C. moschata.

Supplemental Table S19. Orthologous genomic regions between V. vinifera and L. siceraria.

Supplemental Table S20. Number of duplicated genes within selected genomes related to WGD events.

Supplemental Table S21. Number of duplicated genes within selected genomes related to divergence between genomes.

Supplemental Table S22. Homologous alignments of B. loranthoides, C. moschata, and C. melo with V. vinifera as the reference.

Supplemental Table S23. Homologous alignments of Cucurbitales genomes with V. vinifera as the reference.

Supplemental Table S24. Gene loss and gene translocation rates with V. vinifera as the reference genome.

Supplemental Table S25. Gene loss and gene translocation rates with C. melo as the reference genome.

Supplemental Table S26. Gene retention in homoeologous plant regions with reference to V. vinifera genomes.

Supplemental Table S27. Gene retention in homoeologous plant regions with reference to C. melo genomes.

Supplemental Table S28. Gene IDs of gene families within the cucurbitacin biosynthesis pathway in selected genomes.
Supplemental Table S29. Nucleotide substitution models for phylogenetic trees within different gene families.

Funding
This work was supported by the National Natural Science Foundation of China (32170236 and 31501333 to J.P.W.); the Natural Science Foundation of Hebei Province (C2020209064 to J.P.W. and C2020209020 to M.Y.).

Conflict of interest statement. The authors declare no competing interests.

References
Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (2012) Basic local alignment search tool (BLAST). J Mol Biol 215: 403–410
Andeweg JM, Bruyn JJE (1959) Breeding of non-bitter cucumbers.

Chai G, Yu M, Jiang L, Duan Y, Huang J (2005) The worldwide holoparasitic

Filipowicz N, Renner SS (2020) A Matlab

analyses of big biological data. Mol Plant

(2020) TBtools: An integrative toolkit developed for interactive

netics. Early allopolyploid evolution in the post-Neolithic

Brassica

2020) Unravelling angiosperm genome evolution by phylogenetic analysis of chromo-

somal duplication events. Nature 422: 433–438

Chai G, Yu M, Jiang L, Duan Y, Huang J (2019) HMMCAS: A web tool for the identification and domain annotations of CAS proteins. IEEE/ACM Trans Comput Biol Bioinform 16: 1313–1315

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345: 950–953

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBTools: An integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13: 1194–1202

Chen JC, Chiu MH, Nie RL, Cordell GA, Qiu SX (2005) Cucurbetinics and cucurbitane glycosides: structures and biological activities. Nat Prod Rep 22: 386–399

Chen ZJ (2010) Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci 15: 57–71

Cichosz SL, Jensen MH, Larsen TK, Hejlesen O (2020) A Matlab tool for organizing and analyzing NHANES data. Stud Health Technol Inform 270: 1179–1180

Dewitte A, Twyford AD, Thomas DC, Kidner CA, Huylenbroeck JV (2011) The origin of diversity in begonia: genome dynamism, population processes and phylogenetic patterns. The dynamical processes of biodiversity—case studies of evolution and spatial dis-

tribution. InTech

Doyle JJ, Egan AN (2010) Dating the origins of polyploidy events. New Phytol 186: 73–85

Emery M, Willis MMS, Hao Y, Barry K, Oakgrove K, Peng Y, Schmutz J, Lyons E, Pires JC, Edger PP, et al. (2018) Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. PLoS Genet 14: e1007267

Filipowicz N, Renner SS (2010) The worldwide holoparasitic Apodantheaceae confidently placed in the Cucurbitales by nuclear

innovations. Mol Plant 13: 1117–1133

Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Trans R Soc Lond B Biol Sci 368: 20120426

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31: 1296–1297

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagranda A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–467

Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ, et al. (2012) A genome triplication associated with early diversification of the core eudicots. Genome Biol 13: R3

Jiao Y, Li J, Tang H, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell 26: 2792–2802

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3896–3905

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772–780

Kaushik U, Aeri V, Mir SR (2015) Cucurbitacin—an insight into evolution of chemical diversity. Front Plant Sci 6: 129

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS (2018) Impact of whole-genome duplication events on diversification rates in angiosperms. Am J Bot 105: 348–363

Lanfear R, Ho SY, Jonathan Davies T, Moles AT, Aarslen L, Swenson NG, Warman L, Zanne AE, Allen AP (2013) Taller plants have lower rates of molecular evolution. Nat Commun 4: 1879

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al. (2014) Genome sequence of the cultivated cotton Gossypium arboreum. Nat Genet 46: 567–572

Li J, Yuan J, Zhao Y, Meng F, Liu C, Zhang Z, Guo H, Xie Y, Hou Y, Li X, et al. (2019a) Reconstruction of evolutionary trajectories of coffee chromosomes.

Li L, Chen X, Fang D, Dong S, Guo X, Li N, Campos-Dominguez L, Wang W, Liu Y, Lang X, et al. (2022) Genomes shed light on the evolution of Begonia, a mega-diverse genus. New Phytol 234: 295–310

Li Q, Qiao X, Yin H, Zhou Y, Dong H, Qi K, Li L, Zhang S (2019b) Unbiased subgenome evolution following a recent whole-genome duplication in pear (Pyrus bretschneideri Rehd.). Hortic Res 6: 34

Liu C, Wang YG (2022) Does one subgenome become dominant in the formation and evolution of a polyploid? Ann Bot 15: mcac024

Liu Y, Wang J, Ge W, Wang Z, Li Y, Yang N, Sun S, Zhang L, Wang X (2017) Two highly similar poplar paleo-subgenomes suggest an autotetraploid ancestor of Salicaceae plants. Front Plant Sci 8: 571

Ma L, Wang Q, Zheng Y, Guo J, Yuan S, Fu A, Bai C, Zhao X, Zhang S, Wen C, et al. (2022) Cucurbitaceae genome evolution, gene function and molecular breeding. Hortic Res 6: uhab057

Matthews ML, Endress PK (2004) Comparative floral structure and systematics in Cucurbitaceae (Corynocarpaceae, Coriariaceae, Tetrameraceae, Datiscaeae, Begonieae, Cucurbitaceae, Anisophylleaceae). Bot J Linn Soc 145: 129–185

Mayr C (2019) What are 3’ UTRs doing? Cold Spring Harb Perspect Biol 11: a034728

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELM, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. (2021) Pfam: The protein families database in 2021. Nucleic Acids Res 49: D412–D419

Murat F, Armero A, Pont C, Klopp C, Salse J (2017) Reconstructing the genome of the most recent common ancestor of flowering plants. Nat Genet 49: 490–496

Murat F, Zhang R, Guizard S, Flores R, Armero A, Pont C, Steinbach D, Quevenelle H, Cooke R, Salse J (2014) Shared
subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K proteomes. Genome Biol Evol 6: 12–33

Navarro E, Mallen A, Hueso M (2021) Dynamic variations of 3’UTR length reprogram the mRNA regulatory landscape. Biomedicines 9: 1560

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 418–426

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268–274

Osbourn A (2010) Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol 154: 531–535

Parey E, Louis A, Montfort J, Guigue Y, Crollius HR, Berthelot C (2017) Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. Nat Commun 11: 884

Vevelmans D, Proost S, Vanneste K, Coenen H, Viaene T, Ruels P, Maere S, Van de Peer Y, Geuten K (2012) Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. Mol Biol Evol 29: 3793–3806

Wang J, Qin J, Sun P, Ma X, Yu J, Li Y, Sun S, Lei T, Meng F, Wei C, et al. (2019a) Polyploidy index and its implications for the evolution of polyploids. Front Genet 10: 807

Wang J, Sun P, Li Y, Liu Y, Yang N, Yu J, Ma X, Sun S, Xia R, Liu X, et al. (2018a) An overlooked polyploidization in Cucurbitaceae. Mol Biol Evol 35: 16–26

Wang J, Sun P, Li Y, Liu Y, Yu J, Ma X, Sun S, Yang N, Xia R, Lei T, et al. (2017) Hierarchically aligning 10 legume genomes establishes a family-level genomics platform. Plant Physiol 174: 284–300

Wang J, Yu J, Li Y, Wei C, Guo H, Liu Y, Zhang J, Li X, Wang X (2020a) Sequential polyploidization shaped the carrot genome. BMC Plant Biol 20: 52

Wang J, Yu J, Sun P, Li C, Song X, Lei T, Li Y, Yuan J, Sun S, Ding H, et al. (2020b) Paleo-polyploidization in lycocenomes. Genomics Proteomics Bioinformatics 18: 333–340

Wang J, Yu J, Sun P, Li Y, Xia R, Liu Y, Ma X, Yu J, Yang N, Lei T, et al. (2016a) Comparative genomic analysis of rice and pineapple contributes to understand the chromosome number reduction and genomic changes in grasses. Front Genet 7: 174

Paleotetraploidization events in Cucurbitales
provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus. Plant J 92: 963–975

Xie D, Xu Y, Wang J, Liu W, Zhou Q, Luo S, Huang W, He X, Li Q, Peng Q, et al. (2019) The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun 10: 5158

Zhang H, Zhao B, Wei H, Zeng H, Sheng D, Zhang Y (2022) Cucurbitacin B controls M2 macrophage polarization to suppresses metastasis via targeting JAK-2/STAT3 signalling pathway in colorectal cancer. J Ethnopharmacol 287: 114915

Zhao Q, Meng Y, Wang P, Qin X, Cheng C, Zhou J, Yu X, Li J, Lou Q, Jahn M, et al. (2021) Reconstruction of ancestral karyotype illuminates chromosome evolution in the genus Cucumis. Plant J 107: 1243–1259

Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, et al. (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 51: 865–876