Plant Communications
Research article

CellPress
Partner Journal

Differences in pseudogene evolution contributed to the contrasting flavors of turnip and Chiifu, two *Brassica rapa* subspecies

Xin Yin[1,2,3,6], Danni Yang[1,2,3,5,6], Youjie Zhao[4,6], Xingyu Yang[1,2,3,5,6], Zhili Zhou[1,2,3,6], Xudong Sun[1,2,3], Xiangxiang Kong[1,2,3], Xiong Li[1,2,3], Guangyan Wang[1,2,3], Yuanwen Duan[1,2,3], Yunqiang Yang[1,2,3,*] and Yongping Yang[1,2,3,*]

[1]Plant Germplasm and Genomics Center, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[2]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, China

[3]Institute of Tibetan Plateau Research at Kunming, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[4]College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming, Yunnan, China

[5]University of Chinese Academy of Sciences, Beijing 100049, China

[6]These authors contributed equally to this article.

***Correspondence: Yunqiang Yang (yangyunqiang@mail.kib.ac.cn), Yongping Yang (yangyp@mail.kib.ac.cn)**

https://doi.org/10.1016/j.xplc.2022.100427

ABSTRACT

Pseudogenes are important resources for investigation of genome evolution and genomic diversity because they are nonfunctional but have regulatory effects that influence plant adaptation and diversification. However, few systematic comparative analyses of pseudogenes in closely related species have been conducted. Here, we present a turnip (*Brassica rapa* ssp. *rapa*) genome sequence and characterize pseudogenes among diploid *Brassica* species/subspecies. The results revealed that the number of pseudogenes was greatest in *Brassica oleracea* (CC genome), followed by *B. rapa* (AA genome) and then *Brassica nigra* (BB genome), implying that pseudogene differences emerged after species differentiation. In *Brassica* AA genomes, pseudogenes were distributed asymmetrically on chromosomes because of numerous chromosomal insertions/rearrangements, which contributed to the diversity among subspecies. Pseudogene differences among subspecies were reflected in the flavor-related glucosinolate (GSL) pathway. Specifically, turnip had the highest content of pungent substances, probably because of expansion of the methylthioalkylmalate synthase-encoding gene family in turnips; these genes were converted into pseudogenes in *B. rapa* ssp. *pekinensis* (Chiifu). RNA interference-based silencing of the gene encoding 2-oxoglutarate-dependent dioxygenase 2, which is also associated with flavor and anticancer substances in the GSL pathway, resulted in increased abundance of anticancer compounds and decreased pungency of turnip and Chiifu. These findings revealed that pseudogene differences between turnip and Chiifu influenced the evolution of flavor-associated GSL metabolism-related genes, ultimately resulting in the different flavors of turnip and Chiifu.

Key words: turnip genome, comparative genomics, pseudogene evolution, GSL biosynthesis, flavor

Yin X., Yang D., Zhao Y., Yang X., Zhou Z., Sun X., Kong X., Li X., Wang G., Duan Y., Yang Y., and Yang Y. (2023). Differences in pseudogene evolution contributed to the contrasting flavors of turnip and Chiifu, two *Brassica rapa* subspecies. Plant Comm. **4**, 100427.

INTRODUCTION

The "triangle of U model" represents the genetic relationships among *Brassica* species (Nagaharu, 1935), including three diploid species, *Brassica rapa* (AA, 2n = 20) (Cheng et al., 2016; Zhang et al., 2018), *Brassica nigra* (BB, 2n = 16), and *Brassica oleracea* (CC, 2n = 18) (Liu et al., 2014). Hybridizations involving these species resulted in three amphidiploid species: *Brassica juncea* (AABB, 2n = 36) (Yang et al., 2016), *Brassica napus*

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

(AACC, 2n = 38) (Chalhoub, 2014), and *Brassica carinata* (BBCC, 2n = 34) (Johnston et al., 2005; Song et al., 2021). These analyses of *Brassica* reference genome sequences structurally characterized *Brassica* ancestral genes and revealed candidate genes controlling target traits. Because *Brassica* is closely related to *Arabidopsis thaliana*, both underwent paleopolyploidization events (γ, ~300 million years ago [mya]; β, ~112–235 mya; and α, ~20–100 mya) (Bowers et al., 2003). Brassicaceae lineage-specific polyploidization events (whole-genome triplication [WGT], ~12.4–22.5 mya) (Beilstein et al., 2010; Wang et al., 2011b) occurred after their divergence from the *Arabidopsis* lineage. These events were followed by a diploidization that involved substantial genomic recombination, pseudogenization, and eventual gene loss (Bowers et al., 2003; Lysak et al., 2005; Town et al., 2006; Mun et al., 2009; Wang et al., 2011b; Cheng et al., 2013; Chalhoub, 2014; Guo et al., 2021; Yang et al., 2022). Gene loss is an important process in the two-step evolutionary model of *Brassica* diploid plants (least fractionated subgenome [LF], medium fractionated subgenome [MF1], and most fractionated subgenome [MF2]) (Wang et al., 2011b; Cheng et al., 2012). Ancient polyploidization events generated *Brassica* vegetable and oilseed crops with various shapes and tastes that developed under natural conditions or through human activities (Wang et al., 2011b; Graham and May, 2011). *Brassica* vegetable crops include *B. rapa* (Chinese cabbage, pak choi, and turnip) and *B. oleracea* (broccoli, cabbage, and cauliflower), and oilseed crops include *B. napus*, *B. juncea*, and *B. carinata*. The same *Brassica* species may exhibit morphological diversity, and different *Brassica* species can form morphologically similar organs.

Pseudogenes are genetic elements related to functional genes, but they are nonfunctional because of disabling mutations that have occurred during long-term evolution (Xie et al., 2019). Pseudogenes may include in-frame stop codons, frameshifts, and truncated gene sequences (Zhang et al., 2003). Although pseudogenes are nonfunctional, they can affect plant development and adaptation (Gujas et al., 2012; Wu et al., 2017; Xie et al., 2019; Xu et al., 2019). They have also been associated with intellectual disabilities in humans (Green et al., 2017) and incipient balancing selection in bacteria (Will et al., 2010). Polyploidization events have produced thousands of pseudogenes in plant genomes (Wolfe, 2001; Xie et al., 2019). For example, analyses of *B. napus* and *B. oleracea* indicated that some genes in *Brassica* ancestors were lost or underwent pseudogenization, mainly affecting flowering time (Schiessl et al., 2014). Pseudogenes related to production of anticancer phytochemicals and morphological variations represent the consequences of genome duplications and genetic divergence via polyploidization, resulting in biochemical and morphological changes in *B. oleracea* (Liu et al., 2014). *Brassica* pseudogenes contributed to genome evolution after ancient polyploidization events, but there have been relatively few genome-wide, multi-species analyses of their rates of evolution and surrounding chromatin environment.

Glucosinolates (GSLs) and the products of their hydrolysis, especially isothiocyanates, are important secondary metabolites related to the typical flavors (bitterness and pungency) of Brassicaceae plants (Bell et al., 2018). Degradation of glucoraphanin generates products (e.g., sulforaphane) that reportedly have anticancer activities (Fimognari et al., 2002; Tortorella et al., 2015). GSLs and their hydrolysates that affect plant flavors vary greatly between species. Aliphatic GSLs (e.g., sinigrin, gluconapin, and progoitrin) and indolic GSLs (e.g., glucobrassicin and neoglucobrassicin) are responsible for the bitter taste of broccoli and some cauliflower varieties (Engel et al., 2002; Jones et al., 2006; Stotz et al., 2011). Isothiocyanate compounds derived from sinigrin, gluconapin, gluconasturtiin, glucoputranjivin, glucosinalbin, glucobrassicanapin, and glucoraphasatin are associated with the pungency of several Brassicaceae crops, including cabbage, broccoli, kale, wasabi, caper, maca, and radish (Bell et al., 2018). In other species, such as papaya, glucotropaeolin production leads to increased pungency (Bell et al., 2018). Because pungency is an undesirable trait, decreasing the abundance of these compounds may be critical for satisfying consumer taste preferences (Drewnowski and Gomez-Carneros, 2000; Suzuki et al., 2006; Bell et al., 2018). However, the effects of ancient polyploidization events on the flavor-related traits of *Brassica* crops have not been resolved at the genomic level.

Turnip (*Brassica rapa* ssp. *rapa*; AA, 2n = 20), an important *B. rapa* crop, is one of the oldest known taproot vegetables (Liang et al., 2006; Zhang et al., 2014). It was initially cultivated in Europe in 2500–2000 BC, but it subsequently spread to other parts of the world (Song et al., 1990; Wu et al., 2019). The turnip taproot is often used in French and Japanese cooking (Sasaki and Takahashi, 2002). In China, turnip has traditionally been cultivated on the Qinghai–Tibet Plateau as an edible crop used as animal feed; it is also valued for its pharmaceutical properties (Zheng et al., 2018). Thus, turnips were domesticated and cultivated by our ancestors (Ignatov et al., 2008; Cheng et al., 2016; Qi et al., 2017). Many studies have confirmed that turnips are a good source of vitamin C, dietary fiber, folate, niacin, and calcium (Parveen et al., 2015; Ma et al., 2016). However, the turnip taproot has a pungent taste, possibly because of the considerable abundance of aliphatic GSLs, which are the most abundant GSLs in turnip (Bell et al., 2018). Accordingly, the taste of turnip varies substantially from that of other *B. rapa* crops with the AA genome, and this has influenced the acceptance of turnips by consumers. A draft genome assembly at the scaffold level and a genome assembly using PacBio and chromosome conformation capture (Hi-C) technologies at the chromosome level have recently been constructed on the basis of an analysis of a European turnip (ECD04) (Park et al., 2021; Yang et al., 2022). Large-scale resequencing of *B. rapa* and a pan-genome revealed the diverse morphotypes and structural variations that arose during intraspecific diversification of *B. rapa*, providing researchers with useful genomic information on Chinese and European turnips (Cheng et al., 2016; Cai et al., 2021). However, studies of differences in the evolution of flavors between turnip and other *B. rapa* crops have been lacking. For the benefit of people living in need of turnips on the Tibetan plateau or other environmentally hostile areas, the effect of the evolution of the GSL biosynthesis pathway on the pungency of turnips should be elucidated.

In this study, we compared pseudogene evolution among *Brassica* diploid species. We analyzed a chromosome-level turnip genome sequence that was obtained using Illumina and PacBio data and was assembled according to information generated by Hi-C technologies. The results of this investigation highlight

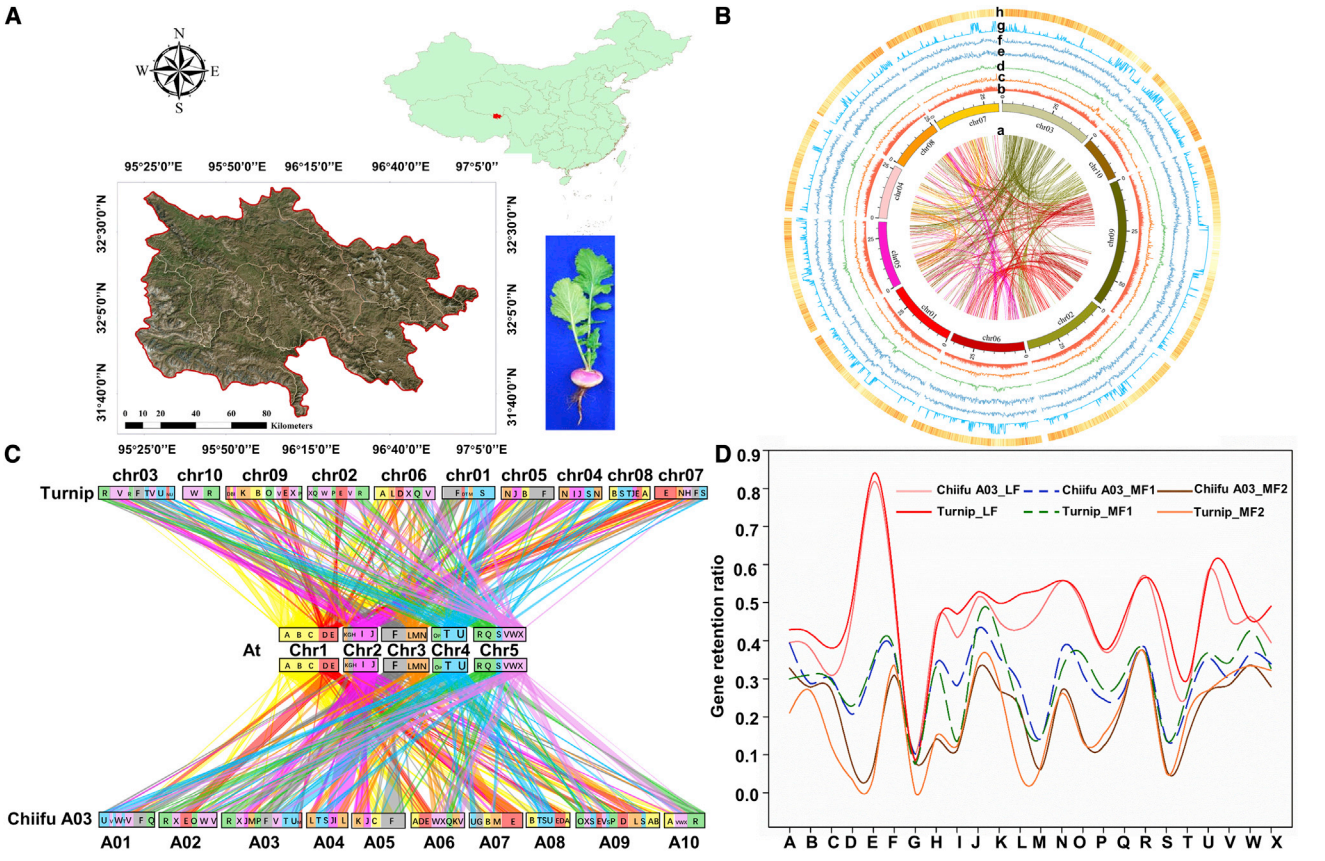


Figure 1. Genome sequencing of turnip.

(A) Turnip sample collection sites in Tibet.

(B) Overview of the turnip draft genome assembly. chr1–chr10, circular representation of the pseudomolecules. a, gene duplications; b, total number of repetitive elements; c, copia elements; d, gypsy elements; e, terminal inverted repeat elements; f, large retrotransposon derivative elements; g, GC density; h, gene density. The presented data are for a 100-kb window.

(C) Structure and segmental collinearity of the genomes of turnip, Chiifu A03, and *A. thaliana*. Syntenic blocks are labeled according to the *A. thaliana* genome (A–X).

(D) Gene retention ratio of the three subgenomes (LF, MF1, and MF2) of turnip and Chiifu A03 on the basis of a comparison with *A. thaliana* (A–X blocks). The x axis presents the physical position of each *A. thaliana* block (A–X). The y axis presents the percentage of the retained orthologous genes corresponding to a gene in the *A. thaliana* A–X blocks.

the differences in the pseudogenes of *Brassica* species. The objective of this study was to clarify the dynamics of *Brassica* pseudogene evolution and the effects of GSL metabolism on turnip pungency. The data presented here provide useful insights into turnip genome evolution and will serve as an important resource for breeding programs interested in optimizing the content of beneficial GSLs in turnips and other *Brassica* crops.

RESULTS

Genome sequencing, assembly, and annotation

For the turnip genome sequencing analysis, a single plant was collected in Nangqen, Qinghai province, China (96°29′24″, 32°12′36″) (Figure 1A). Genomic DNA was extracted and sequenced using PacBio and Illumina sequencing strategies. We obtained 44.93 Gb of PacBio reads, which corresponded to about 110× coverage of the 446.09-Mb genome; the genome size was estimated on the basis of k-mer statistics (Supplemental Tables 1 and 2). The assembled genome was 409.69 Mb, which included 2437 contigs with an N50 value of 1.21 Mb (Table 1), making it similar in size to the assembled Chiifu A03 genome (403.20 Mb with a contig N50 value of 4.29 Mb) (Sun et al., 2022). To evaluate the genome assembly quality, Illumina short reads were mapped to the assembly, which resulted in a mapping efficiency of 96.69% (Supplemental Table 3). The genome integrity, determined on the basis of benchmarking universal single-copy orthologs (BUSCO), was 97.20% (Supplemental Table 4).

To anchor the scaffolds to chromosomes, 26.79 Gb of clean reads (65.50×) were obtained via Hi-C library sequencing (Supplemental Table 5). Read pairs were mapped to the draft assembly using BWA (version 0.7.10-r789) (Li and Durbin, 2009). We determined that 85.28% of the reads were correctly mapped to the genome (Supplemental Table 6), including 22.01% uniquely mapped read pairs. LACHESIS (Burton et al., 2013) was used to group, sort, and orient all contigs, and 1501 scaffolds were successfully anchored to 10 pseudochromosomes (chr01–chr10) (Figure 1B; Supplemental Table 7). The scaffold N50 value for the final

Primary genome assembly	
Sequenced genome size (Mb)	409.69
Number of contigs	2437
Contig N50 (bp)	1 214 203
Contig N90 (bp)	54 084
Maximum contig size (bp)	8 713 068
Total size (bp)	409 691 509
Chromosome-level genome assembly	
Number of chromosomes/scaffolds	10/1501
Scaffold N50	37 215 573
Scaffold N90	26 283 954
GC content (%)	35.14
Maximum scaffold size	57 405 788
Total size	355 929 172

Table 1. Summary of the turnip genome assembly and annotation.

assembly was 37.22 Mb. Thus, the chromosome-level genome assembly for turnip comprised pseudochromosomes ranging in length from 23.02 to 57.40 Mb (Supplemental Figure 1; Supplemental Table 7). The number of pseudochromosomes was consistent with the previously reported number of chromosomes in the *Brassica* AA genome (i.e., n = 10 and 2n = 20) (Zhang et al., 2018).

The *de novo* prediction of repetitive sequences in the turnip genome indicated that repetitive sequences represented 42.60% of the assembled genome; this proportion was lower than that of Chiifu A03 (50.31%) (Sun et al., 2022) and ECD04 (46.9%) (Yang et al., 2022) (Supplemental Table 8). The long terminal repeat (LTR) retrotransposon was the most common repetitive sequence, accounting for 17.13% of the genome, and this proportion was between that of Chiifu A03 (19.62%) and ECD04 (20.77%) (Supplemental Table 8; Sun et al., 2022; Yang et al., 2022). Protein-coding genes were predicted via *ab initio* gene predictions, homology-based predictions, and RNA sequencing (RNA-seq) and then integrated using EVidenceModeler (version 1.1.1) (Haas et al., 2008). In total, 56 832 genes were obtained. In addition, 98.57% of the genes were annotated by screening databases (e.g., NR) (Supplemental Table 9), and microRNA target genes were predicted (Supplemental Table 10).

We next assessed the quality of the turnip genome. Specifically, the LTR assembly index score for the turnip genome was 13.8, indicative of good assembly continuity. Genome annotation completeness was estimated to be 95.30% on the basis of the BUSCO assessment. Illumina paired-end reads for TUA, TUE, A03, and ECD04 and reads derived from 48A resequencing data for a turnip population (Yang et al., 2019) were mapped to turnip chromosomes (Supplemental Figure 2; Supplemental Table 11). Approximately 96.47%–97.42% of these reads were mapped to the turnip chromosomes, indicating that the turnip genome assembly contained almost all of the information provided by the Illumina reads and 48A resequencing data. Similarly, 96.34%–98.56% of the Illumina paired-end reads and reads obtained from48A resequencing data were mapped to the TUA, TUE, A03, and ECD04 chromosomes (Supplemental Figure 3; Supplemental Tables 12, 13, 14, and 15). These results confirmed that the turnip genome used in this study was assembled and annotated appropriately.

The whole-genome triplication (WGT) in *Brassica* species was followed by extensive gene loss and frequent reshuffling of triplicated genomic blocks (Liu et al., 2014). Triplicated regions, which were determined on the basis of homologous gene pairing between *A. thaliana* and turnip, as well as Chiifu A03, were constructed; they were related to the 24 ancestral crucifer karyotype blocks (A–X) in *A. thaliana* (Schranz et al., 2006). Most of the regions in the turnip genome shared a conserved syntenic block with the *A. thaliana* genome (Figure 1C). The WGT-derived triplicated blocks in the turnip and Chiifu A03 genomes were partitioned into the LF, MF1, and MF2 subgenomes (Figure 1D; Wang et al., 2011b). These syntenic blocks occupied most of the genome assemblies of *A. thaliana* (25 054 genes, 90.59% of 27 655 genes), turnip (26 973 genes, 47.46%), and Chiifu A03 (26 738 genes, 55.96%), providing the foundation for comparative analyses (Supplemental Table 16).

Comparative genomics analysis

To analyze turnip evolution, 3593 single-copy orthologs from 12 species, including *Brassica* species with an AA genome (Chiifu v3.5, Chiifu A03, *B. rapa* L. ssp. *chinensis* [Bras], TUA, TUE, ECD04, *B. rapa* Z1 [Z1], and turnip) (Cai et al., 2021; Istace et al., 2021; Li et al., 2021; Sun et al., 2022; Yang et al., 2022; Zhang et al., 2022), a BB genome (*B. nigra*) (Perumal et al., 2020), and a CC genome (*B. oleracea*) (Cai et al., 2020), as well as *Raphanus sativus* (Kitashiba et al., 2014), were used for protein sequence alignments, with the *A. thaliana* genome (Cheng et al., 2017) serving as an outgroup (Figure 2A and Supplemental Figure 4). Turnips and other *B. rapa* subspecies, which were derived from a common ancestral *Brassica* species (AA genome), were clustered together on a branch. The divergence time between *B. nigra* and *B. oleracea* was approximately 8.42 mya, whereas the divergence time between *B. oleracea* and *B. rapa* subspecies was about 2.29 mya. The timing of this evolutionary process was consistent with the

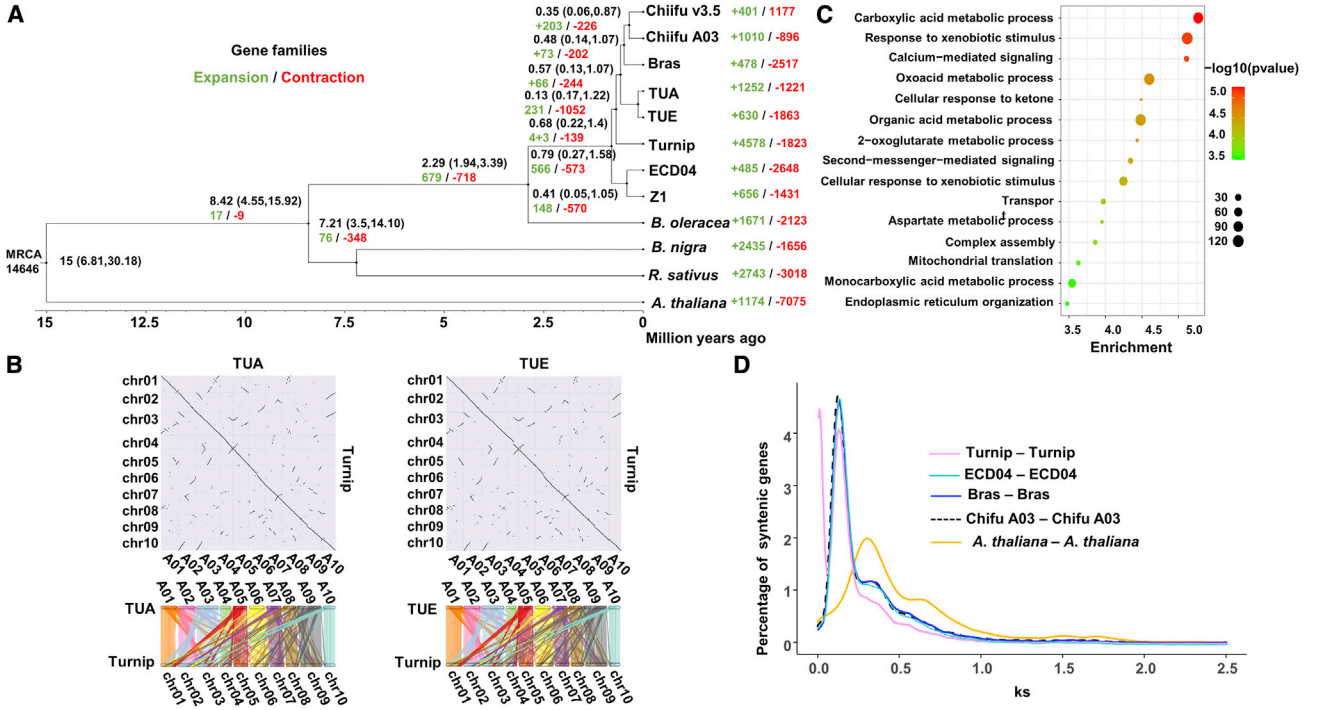


Figure 2. Evolution of the turnip genome.

(A) Phylogenetic tree, with the number of gene family expansion and contraction events indicated by green and red numbers, respectively, below each species name. The estimated divergence times (million years ago) are indicated at each node (95% credibility intervals).

(B) Dot plot for the segmental collinearity between the turnip and TUA genomes and between the turnip and TUE genomes. The TUA and TUE chromosomes are indicated by different colors, and the orthologous chromosomal segments in turnip are indicated by the same color. Conserved collinear blocks of gene models are presented for the 10 turnip chromosomes and the TUA and TUE genomes.

(C) GO enrichment analysis of the most significantly expanded gene families in turnip. The enrichment factor indicates −Log10 (*P* value). The 15 most significantly enriched pathways are shown.

(D) Distribution of the synonymous substitution rate (Ks).

findings of an earlier study (Guo et al., 2021). The divergence times between turnip and ECD04 and between turnip and other *B. rapa* subspecies were approximately 0.79 and 0.68 mya, respectively. LTR retrotransposons were actively inserted into the turnip, ECD04, and Chiifu A03 genomes approximately 1.34, 1.88, and 1.65 mya, respectively (i.e., before the divergence between turnip and ECD04 and between turnip and *B. rapa* subspecies). A comparative analysis of the timing of LTR retrotransposon insertion into the genomes revealed that turnip and *B. rapa* subspecies had similar evolutionary histories (Supplemental Figure 5).

The MCScanX package (Wang et al., 2012) was used to analyze the collinearity between turnip and Chiifu A03, ECD04, TUA, and TUE (Figure 2B and Supplemental Figure 6). A total of 643 and 631 large syntenic blocks were detected between the turnip genome and the TUA and TUE genomes, respectively. Moreover, 335.62 Mb (97.00%) and 336.31 Mb (97.20%) of the sequences on the 10 turnip chromosomes were revealed to be collinear, covering almost all of the TUA and TUE chromosomes, respectively (Figure 2B). To verify the accuracy and continuity of the assembled sequences, sequence collinearity between the turnip genome and the Chiifu A03, ECD04, TUA, and TUE genomes was determined using the nucmer program of the MUMmer package (v4.0rc1) (Marcais et al., 2018), after which NGenomeSyn was used to detect highly similar sequence segments (Supplemental Figure 7). This analysis indicated that turnip is closely related to TUA, TUE, ECD04, and Chiifu A03. However, there were numerous chromosomal insertions/rearrangements that differentiated turnip from TUA, TUE, ECD04, and Chiifu A03. This is in accordance with previous studies that detected many genomic insertions/rearrangements after the WGT event in *Brassica* (Liu et al., 2014). An earlier analysis of retained or lost genes after the WGT in *Brassica* revealed over-retention of genes involved in metabolic pathways (Liu et al., 2014; Lou et al., 2012). To further resolve the diversity between turnip and *B. rapa* subspecies, we detected 4578 and 1010 gene families that had expanded in the turnip and Chiifu A03 genomes, respectively (Figure 2A). Gene Ontology (GO) enrichment analysis indicated that the expanded gene families in turnip were mainly related to metabolic processes and responses to environmental stimuli (Figure 2C), whereas those in Chiifu A03 were mainly associated with S-glycoside catabolism (Supplemental Figure 8). This implies that there may be differences in the survival strategies of turnip growing on the Tibetan plateau and domesticated Chiifu A03 (Cheng et al., 2016; Zhao et al., 2005). On the basis of duplicated gene pairs, we calculated the age distribution of the synonymous substitution rate (Ks) (Figure 2D). The results indicated that, other than the common polyploidization events among *Brassica* species, there were no additional species-specific WGD events in turnip, consistent

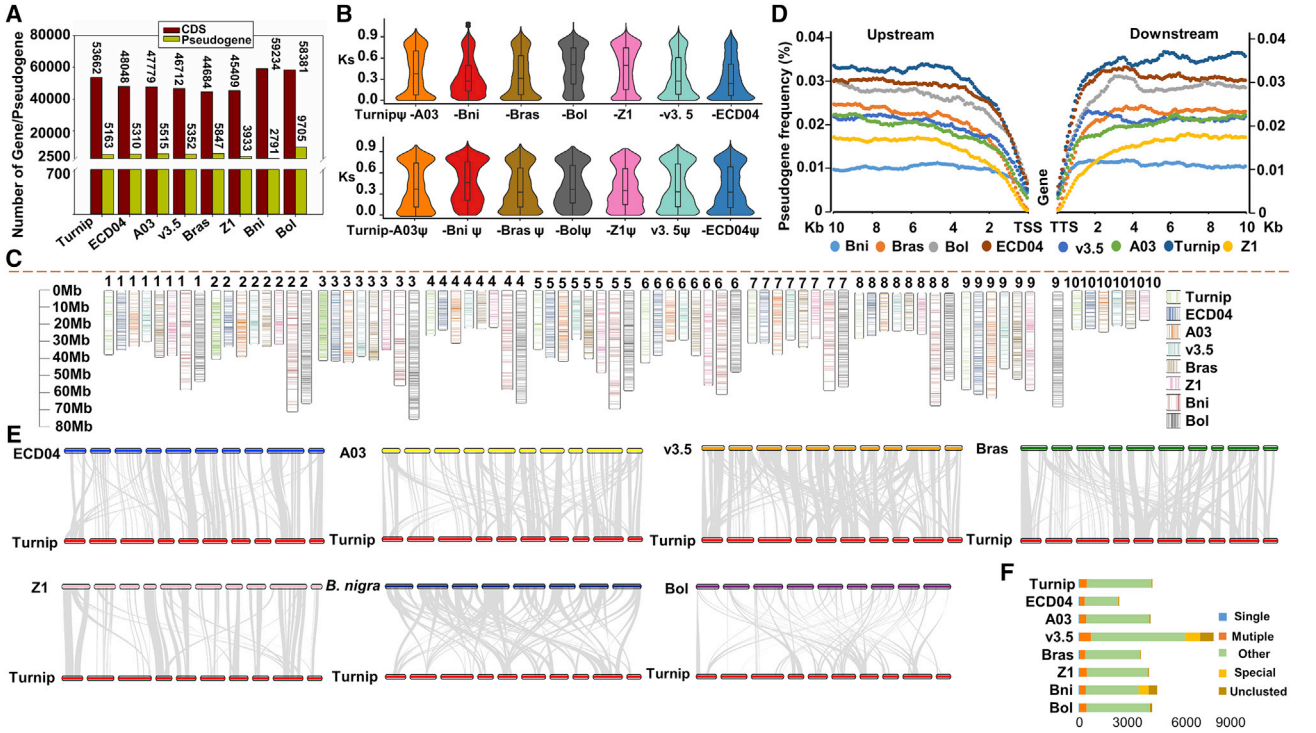


Figure 3. Identification and comparison of pseudogenes in diploid *Brassica* species.

(A) Number of genes/pseudogenes in eight diploid *Brassica* genomes, including the AA genomes of turnip, ECD04, Chiifu A03, Chiifu v3.5, *B. rapa* L. ssp. *chinensis* (Bras), and *B. rapa* Z1, as well as the BB genome of *B. nigra* (Bni) and the CC genome of *B. oleracea* (Bol).

(B) Comparison of pseudogene evolution rates (Ks) in eight diploid *Brassica* species/subspecies. Turnipψ–others indicates the pseudogenes of turnip and functional genes of other species/subspecies. Turnip–othersψ indicates the pseudogenes of other species/subspecies and functional genes of turnip.

(C) Comparison of pseudogene distribution between the turnip genome and the genomes of other diploid *Brassica* species/subspecies. Numbers represent the corresponding chromosome numbers.

(D) Distribution of pseudogenes 10 kb upstream and downstream of coding sequences (CDSs) in eight diploid *Brassica* species/subspecies. The presence of a pseudogene downstream of the transcription termination site or upstream of the transcription start site of each gene is indicated by 1, whereas the absence of a pseudogene is indicated by 0, with a total window size of 100 kb. The data presented on the y axis were calculated by dividing the number of genes with an upstream or downstream pseudogene (1) by the total number of genes.

(E) Comparative analysis of pseudogenes between turnip and other diploid *Brassica* species/subspecies. Syntenic blocks were determined according to alignment of the turnip chromosomes with the chromosomes of other diploid *Brassica* species/subspecies.

(F) Comparison of the number of unique and shared pseudogene families between turnip and other diploid *Brassica* species/subspecies.

with the results of the phylogenetic analysis of the *Brassica* AA genome. This may also explain the similarity in the genomic structures of turnip and *B. rapa* subspecies.

Comparative analysis of pseudogenes in diploid *Brassica* genomes

To systematically identify candidate pseudogenes in the genomes of eight diploid *Brassica* species, including the AA genomes of turnip, ECD04, Chiifu A03, Chiifu v3.5, Bras, and Z1, the BB genome of *B. nigra* (Bni), and the CC genome of *B. oleracea* (Bol), we used prediction software and performed homology searches with stringent filters to minimize noise and enhance positive signals (Figure 3A and Supplemental Figure 9A). Among the examined species, Bol (CC genome) and Bni (BB genome) had the most and fewest pseudogenes, respectively. The AA genomes included 3933–5847 pseudogenes. Turnip, ECD04, Chiifu A03, Chiifu v3.5, and Bras had similar numbers of pseudogenes (Figure 3A). We subsequently determined the pseudogene evolution rate by estimating the Ks values for the pseudogenes and their functional paralogs (Figure 3B). The Chiifu A03 pseudogenes evolved faster than the turnip pseudogenes, possibly because of the high artificial selection pressure to which Chiifu A03 was subjected.

Anchoring of the *Brassica* pseudogenes to each chromosome revealed an asymmetrical distribution, with the greatest variation in the distribution of pseudogenes on chr02 of turnip and on A02 of the other species. Considering the rearrangement results for chromosome A02, we hypothesized that genomic rearrangements may be responsible for the differences in pseudogene distribution (Figure 3C and Supplemental Figure 9B). We analyzed the distribution of pseudogenes 10 kb upstream and downstream of coding sequences (CDSs) in eight *Brassica* diploid species. Turnip had the most pseudogenes in the upstream/downstream regions, with similar pseudogene distribution trends detected in the AA genomes of Bras, Chiifu v3.5, and Chiifu A03. By contrast, the BB genome (Bni) had the fewest pseudogenes in the upstream/downstream regions. This result provides further evidence of the asymmetrical distribution of pseudogenes in eight *Brassica* diploid species (Figure 3D). In

addition, we identified syntenic blocks of pseudogenes in *Brassica*. More specifically, the synteny between turnip and other species/subspecies (i.e., the percentage of syntenic regions) was as follows: 38.40% (Chiifu A03), 43.10% (Bras), 40.20% (Chiifu v3.5), 41.00% (ECD04), 29.20% (Z1), 21.20% (Bni), and 19.90% (Bol) (Figure 3E and Supplemental Figure 9C). The lowest synteny between turnip and Bol may be related to the relatively extensive chromosomal rearrangements and asymmetrical gene loss in duplicated genomic blocks that occurred in the Bol genome (Liu et al., 2014). To further characterize the pseudogenes specific to turnip, syntenic orthologs of pseudogenes in *Brassica* were identified (Figure 3F).

We then performed a GO enrichment analysis of the pseudogenes in eight diploid *Brassica* species/subspecies by annotating their closest functional paralogs (Supplemental Figure 10). The turnip pseudogenes were mainly annotated with the GO terms "multicellular organismal process," "anatomical structure development," and "macromolecule biosynthetic process," probably because some nonfunctional or dispensable genes became pseudogenes. The Chiifu A03 pseudogenes are related to "cellular localization," "signal transduction," and "signaling," suggesting that Chiifu A03 had more pseudogenes related to stress responses. These findings may reflect functional differences in pseudogenes between the two subspecies, which may be the result of selection pressure during domestication. We assessed the functions associated with the Pfam domains encoded by the pseudogenes on the basis of the annotations of their functional paralogs (Supplemental Figure 11). A relatively small proportion of the pseudogenes were related to core genes, such as transcription factor genes (0.20%–1.41%) and kinase genes (0.86%–2.59%), but most of the pseudogenes were functionally unknown and unclassified. Accordingly, there were relatively few regulatory genes among the pseudogenes.

Effect of pseudogenes on the GSL biosynthesis pathway

GSLs and the products of their hydrolysis are determinants of the unique taste of *Brassica* crops (Bell et al., 2018). Compared with other *Brassica* species, turnips contain more aliphatic GSLs (Yang et al., 2020). On the basis of the pseudogene functional annotations, we analyzed the genes involved in the aliphatic GSL metabolic pathway (Figure 4A and Supplemental Figure 12); these genes are the predominant GSL-related genes in the *Brassica* AA genome. Genes encoding methylthioalkylmalate (MAM) synthase were identified in turnip, ECD04 (Yang et al., 2022), Chiifu A03 (Sun et al., 2022), Bras (Li et al., 2021), and 10 other representative *Brassica* species/subspecies (Cai et al., 2021), including TUA, TUE, Chiifu v3.5, BRO, Z1, CCA, CCB, MIZ, PCA, and TCA (Figure 4B, Supplemental Figure 13; Supplemental Table 17). Turnip had the most *MAM* functional genes, which were distributed mainly on chr02 but also on chr03 and chr04. In the other species/subspecies, the *MAM* genes were distributed on chromosomes A02, A03, and A04 (Supplemental Figure 14). Syntenic relationships were detected between the turnip *MAM* genes on chr03 and chr04 and the *MAM* genes on chromosomes A03 and A04 in the other species/subspecies. However, all five *MAM* genes on chr02 in turnip were functional, whereas the syntenic regions on chromosome A02 in the other species/subspecies contained 1–3 *MAM* pseudogenes. There were two main pseudogene types, those on blocks homologous to a turnip *MAM* gene (*Gene0495830*), which had a premature termination codon (Supplemental Figure 15), and those on blocks homologous to turnip *MAM* genes (*Gene0228790* or *Gene0464890*), which became pseudogenes. Because of the decrease in *MAM* gene family size in the other species/subspecies as a result of the development of pseudogenes, we hypothesized that *MAM* genes (*Gene0495830*, *Gene0228790*, and *Gene0464890*) may be critical for explaining the differences in GSL synthesis between turnip and the other species/subspecies.

To test this hypothesis, we first compared the aliphatic GSL content in *Brassica* species with the AA genome (turnip, Chiifu, Bras, and Z1) and Bol (CC genome) (Figure 4C). The aliphatic GSL content varied among organs (leaves/taproots) and developmental stages (10, 20, and 30 days). Compared with their abundance in other *Brassica* species/subspecies, several aliphatic GSLs were more abundant in turnip taproots (gluconapin, progoitrin, glucobrassicanapin, glucoraphanin, glucoerucin, and glucoberteroin) or turnip leaves (gluconapin and glucobrassicanapin). These compounds are the main sources of the pungency of *Brassica* plants (Bell et al., 2018; Kusznierewicz et al., 2013; Depree et al., 1998). We also obtained hairy roots from Chiifu and turnip plants that had been transformed using *Agrobacterium rhizogenes* for overexpression or silencing (via RNA interference [RNAi]) of the *MAM* genes (*Gene0228790*, *Gene0464890*, and *Gene0495830*) (Figure 4D and Supplemental Figure 16). The glucobrassicanapin and gluconapoleiferin contents determined by liquid chromatography-mass spectrometry (MS) were significantly higher in Chiifu hairy roots overexpressing *Gene0228790*, *Gene0464890*, and *Gene0495830* than in the control. As expected, the RNAi-mediated silencing of *Gene0228790*, *Gene0464890*, and *Gene0495830* in turnip hairy roots decreased the gluconapin, progoitrin, and glucobrassicanapin content. These results imply that *MAM* is important for development of the pungent flavor of turnips.

Effect of 2-oxoglutarate-dependent dioxygenase (AOP2)-encoding gene on the GSL biosynthesis pathway

Degradation of glucoraphanin produces sulforaphane, one of the best anticancer compounds identified to date (Fimognari, 2002; Tortorella et al., 2015). Other genes that influence the production of specific anticancer- and flavor-related aliphatic GSLs are *AOP* genes, which encode enzymes that convert sulforaphane-related GSLs into GSLs that lack anticancer properties (Liu et al., 2014). We identified *AOP2* genes in turnip, Bol (CC genome), Bni (BB genome), and nine other representative *Brassica* AA genome species/subspecies (ECD04, Chiifu A03, Bras, Chiifu v3.5, Z1, CXA, CXB, PCA, and TCA) (Cai et al., 2021; Figure 5A). Most species/subspecies had three *AOP2* genes, but among the three Bol *AOP2* genes, one was functional, whereas the other two were pseudogenes. This may help to explain the high glucoraphanin levels in Bol (inactive *AOP2* genes) but not in Chiifu (three active *AOP2* genes) (Wang et al., 2011a; Liu et al., 2014). Two *AOP2* genes in Z1 had syntenic relationships with two *AOP2* genes in turnip, but there

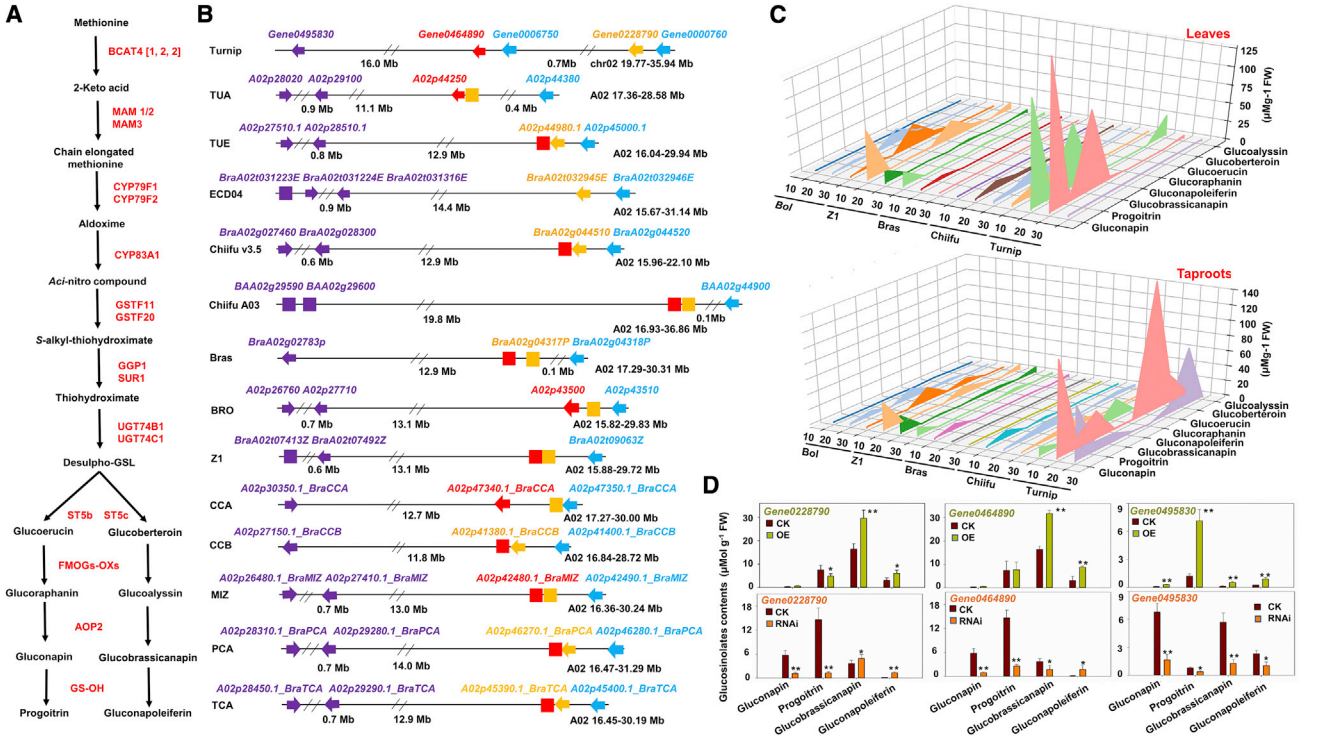


Figure 4. Comparison of pseudogenes and determination of flavor-related metabolite content in the aliphatic GSL metabolic pathways of diploid *Brassica* species.

(A) Overview of the aliphatic GSL biosynthesis pathway.

(B) Analysis of synteny among *MAM* genes in 14 diploid *Brassica* species/subspecies. Turnip chromosome chr02 with *MAM* genes and the collinear chromosome A02 in other *Brassica* species/subspecies are presented. Identical colors represent the same homologous region. Arrows indicate the gene orientation on the chromosome. Boxes represent pseudogenes into which those on blocks homologous to turnip *MAM* (*Gene0495830*, *Gene0228790*, or *Gene0464890*) were converted in other species/subspecies. Specifically, *MAM* (*Gene0495830*) in turnip was converted into pseudogenes because of codon termination in other species/subspecies. Syntenic regions in both genomes, with one turnip genome containing a functional gene (*Gene0228790* or *Gene0464890*) and the other containing a homologous sequence with clear markers indicative of a pseudogene, are presented.

(C) Differences in aliphatic GSL content in the leaves and taproots of *Brassica* species/subspecies 10, 20, and 30 days after germination (n = 4) as determined by HPLC-MS/MS. AA genome: turnip, Chiifu, Bras, and *B. rapa* Z1; CC genome: Bol.

(D) Pungency-related compound content in Chiifu hairy roots overexpressing *MAM* genes (*Gene0228790*, *Gene0464890*, and *Gene0495830*) (top graphs) and in turnip hairy roots in which these three genes were silenced via RNAi (bottom graphs). The compound content was determined by HPLC-MS/MS. Error bars indicate the standard deviation (n = 4). Asterisks indicate significant differences between the transgenic turnip/Chiifu hairy roots and the control (Student's *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$).

was no synteny between Z1 *AOP2* genes and one of the turnip *AOP2* genes (*Gene0486840*) (Figure 5B).

The three *AOP2* genes in turnip were expressed in taproots and leaves at different developmental stages, implying that the sulforaphane-related GSLs in turnip are converted by *AOP2* to GSLs that lack anticancer activities (Figure 5A, Supplemental Table 18, and Supplemental Figure 17). Thus, we attempted to inactivate all *AOP2* genes to decrease pungency and increase sulforaphane formation. We used RNAi technology to silence the expression of three *AOP2* genes in turnip (*Gene0405960*, *Gene0250680*, and *Gene0486840*) and Chiifu (*BraA02g028320*, *BraA09g001360*, and *BraA03g029140*) and then analyzed the pungency and accumulation of glucoraphanin in the resulting samples (Figure 5C and Supplemental Figure 18). As expected, inhibition of *AOP2* expression significantly decreased the content of pungency-related compounds (gluconapin and glucobrassicanapin) in turnip and enhanced glucoraphanin accumulation in turnip and Chiifu. These findings are relevant for future attempts to modulate turnip pungency (Supplemental Figure 19).

DISCUSSION

Pseudogenes are important for research on evolution and comparative genomics because they represent the molecular remnants of ancient genes that existed in the genome millions of years ago (Zou et al., 2009; Moghe et al., 2014; Xie et al., 2019; Xu et al., 2019). The only cross-species comparisons of pseudogenes in plants have focused on the evolution and expression signatures of pseudogenes in the *Arabidopsis* and rice genomes (Zou et al., 2009), the pseudogenization of duplicated genes in wild radish (*Raphanus raphanistrum*) and three other Brassicaceae species (Moghe et al., 2014), and the evolutionary origins of pseudogenes and their associations with regulatory sequences among seven angiosperms (Xie et al., 2019). The diversity in pseudogene evolution among *Brassica* species remains unclear. The comparison of *Brassica*

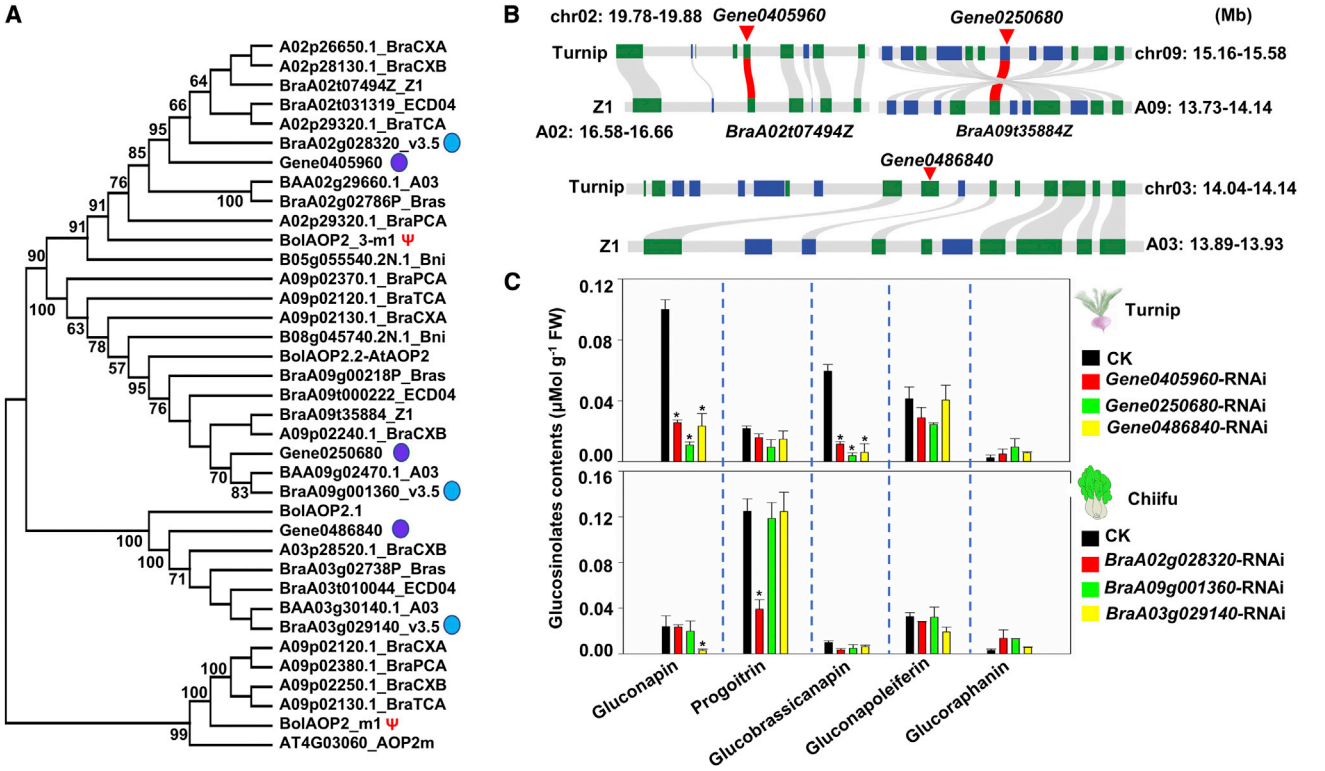


Figure 5. Effect of *AOP2* genes on the GSL biosynthesis pathway.

(A) Neighbor-joining trees of *AOP* gene families in genomes of 13 species, including the AA genome of CXA, CXB, Z1, ECD04, TCA, Chiifu v3.5, turnip, Chiifu A03, Bras, and PCA, the BB genome of Bni, and the CC genome of Bol, with *A. thaliana* serving as an outgroup, were constructed by aligning the CDSs with 1000 bootstrap replicates. Three *AOP2* genes were present in the turnip (purple circles; *Gene0405960*, *Gene0250680*, and *Gene0486840*) and Chiifu (blue circles; *BraA02g028320*, *BraA09g001360*, and *BraA03g029140*) genomes. Red asterisks and Ψ represent pseudogenes in Bol.

(B) Analysis of the synteny between the *AOP2* genes in *B. rapa* Z1 and turnip. Of the three *AOP2* genes in turnip, *B. rapa* Z1 lacked a collinear gene for *Gene0486840*.

(C) Pungency-related compound and glucoraphanin content in *AOP2*-RNAi turnip and Chiifu hairy roots. Error bars indicate the standard deviation (n = 4). Asterisks indicate significant differences between the transgenic turnip/Chiifu hairy roots and the control (Student's *t*-test; $*P < 0.05$, $**P < 0.01$).

pseudogenes revealed in this study suggests that the CC genome had more pseudogenes than the AA genome and fewer pseudogenes derived from a common ancestor, with relatively few syntenic blocks, indicating that the pseudogenes in the AA genome appeared after the divergence from Bol. The pseudogenes were asymmetrically distributed on the chromosomes among the subspecies. The Chiifu A03 pseudogenes evolved faster than the turnip pseudogenes. Hence, although Chiifu A03 and turnip have the AA genome, the genetic diversification of these two subspecies may be related to selection pressure imposed during domestication.

Pseudogenes usually result from gene duplications or retrotranspositions related to WGD events (Wolfe, 2001; Xie et al., 2019). The small difference in the number of pseudogenes among *Brassica* species with the AA genome is consistent with the lack of additional WGD and LTR events after divergence in these species. The asymmetrical distribution of pseudogenes on the chromosomes was due to numerous chromosomal insertions/rearrangements in the *Brassica* AA genomes (Lou et al., 2012; Liu et al., 2014). These pseudogene differences were revealed by gene functional annotations, which indicated that core genes, including those encoding transcription factors, were generally not converted to pseudogenes. Previous studies have determined that *Brassica* crops exhibit extreme morphological characteristics and diverse environmental adaptability because of artificial selection during domestication and breeding (Cheng t al., 2016; Qi et al., 2017). This phenomenon demonstrates that plant survival is the first priority and that the diversity in specific characteristics increased after domestication. Accordingly, the metabolic differences may have been the result of selection for agriculturally desirable traits by humans, especially the flavor-related characteristics of domesticated/semi-domesticated Chiifu and turnip.

The production of four aliphatic GSLs (gluconapin, progoitrin, glucobrassicanapin, and gluconapoleiferin) influences formation of the distinct flavors of *Brassica* crops (Bell et al., 2018). The variability in aliphatic GSL structures is due mainly to two major genetic loci (*MAM* and *AOP*) (Keurentjes et al., 2006; Wentzell et al., 2007; Liu et al., 2014). Specifically, *MAM* controls the variability in aliphatic GSL carbon chain length, whereas *AOP* is responsible for modification of side chain structure (Benderoth et al., 2009; Liu et al., 2014). Detailed analysis of five *Brassica* species in this study revealed considerable variation in the relative content of aliphatic GSLs among species and subspecies, consistent with the findings of earlier studies (Chen et al., 2008; Yang et al., 2020). Previous research confirmed

that there were significant differences in the GSL content of Chinese cabbage (Chiifu) germplasm, due largely to genetic changes because of selection pressure influenced by consumer preference (Kang et al., 2006). Thus, the high concentrations of pungency-related substances in turnip plants grown on the Tibetan plateau may be related to their minimal domestication by humans. In Bol, two nonfunctional *AOP2* genes are associated with decreased accumulation of these four metabolites, which, in turn, is related to an increase in anticancer GSL content (Liu et al., 2014); this may also be the result of flavor-related selection pressure during domestication. We determined that the Chiifu flavor preferred by humans is influenced mainly by nonfunctional *MAM* genes, whereas the pungency of turnip is mainly associated with expansion of the *MAM* gene family. This is in accordance with the results of GSL content surveys and explains why gluconapin and glucobrassicanapin are abundant in turnip (Padilla et al., 2007; Lee et al., 2013) but not in Chiifu. This inspired us to attempt to convert *AOP2* in turnip into a nonfunctional gene via RNAi. Doing so will enhance accumulation of anticancer substances and optimize turnip flavor. Therefore, agriculturally important *Brassica* crop traits may be improved by focusing on GSL pathway-related genes. The data generated in this study will help researchers and breeders develop crops with more desirable flavors and a greater abundance of anticancer compounds, satisfying worldwide consumer demands.

METHODS

Plant materials

Turnip seeds collected from Nangqen county, Qinghai province (N 32°12′11″, E 96°28′50″) were sown in a seedling raising plate. Seedlings were cultivated under controlled greenhouse conditions (12 h light [28°C]/12 h dark [25°C] cycle, 200 mmol photons m^{-2} s^{-1} light intensity, and 75%–80% relative humidity) and were watered appropriately. Genomic DNA was extracted from the leaves and used for subsequent genomic DNA sequencing analysis and construction of Hi-C libraries.

Illumina and PacBio sequencing

Genomic DNA was extracted from leaves according to a standard cetyltrimethylammonium bromide (CTAB) method. DNA quality was assessed using a 2100 Bioanalyzer (Agilent Technologies, USA). The Illumina sequencing library was constructed and then sequenced (150-bp paired-end mode) using the Illumina X Ten platform as described by the manufacturer. For PacBio sequencing, 10 μg genomic DNA was sheared, and then approximately 20-kb fractions were selected using the BluePippin Selection system (Sage Science, USA). The library was sequenced using the Pacific Biosciences Sequel platform.

Estimation of genome size

The turnip genome size was estimated on the basis of a k-mer frequency analysis of the Illumina short reads using the Jellyfish program (version 2.2.6) (http://www.genome.umd.edu/jellyfish.html) with a k-mer frequency of 21. The heterozygosity ratio was estimated using the GenomeScope online tool (http://qb.cshl.edu/genomescope/). Finally, genome size was calculated using the following formula: genome size = k-mer coverage/mean k-mer depth.

De novo assembly and genome refinement

The PacBio SMRT (Single Molecule Real Time) analysis package (https://www.pacb.com) was used for quality control screening of the raw reads with the following parameters: readScore, 0.75; minSubReadLength, 500, including removal of sequencing adapters and low-quality short reads. Errors in the PacBio long reads were corrected using the error correction module embedded in Canu (version 1.3) (Koren et al., 2017) with correctedErrorRate set to 0.045. *De novo* sequence assembly was performed using the default parameters of Canu to produce contigs. The clean PacBio reads were then aligned with the assembled contigs using Basic Local Alignment with Successive Refinement (version 1.3.1) (Chaisson and Tesler, 2012). The contigs were further corrected using Quiver from the SMRT Analysis package. Illumina paired-end reads from the same turnip species were aligned to the optimized contigs using BWA (version 0.7.10-r789) (Li, 2014). The assembled sequences were polished using Pilon (version 1.22) (Walker et al., 2014) with the following parameters: -mindepth 10 -changes -fix bases. The completeness of the genome assembly was evaluated using BUSCO (version 4.0.6) and the embryophyta_odb10 single-copy gene dataset (Simao et al., 2015).

Hi-C library construction

Leaves from individual turnip plants were harvested and immersed in a formaldehyde solution to crosslink and fix chromatin. The leaf cells were lysed, and HindIII endonuclease was used to digest the fixed chromatin. The DNA ends were marked with biotin-14-dCTP (2′-deoxycytidine 5′-triphosphate), and the blunt ends were ligated to each other using DNA ligase. Next, the nuclear complexes were reverse crosslinked during incubation with proteinase K at 65°C. The DNA was purified and sheared (100–500 bp) by sonication. The biotin-labeled fragments were enriched using streptavidin magnetic beads. Poly(A) tails were added to the fragment ends using the Klenow fragment (exo-) before adding the Illumina paired-end sequencing adapter in a ligation mixture. PCR amplification of fractions was performed using 12 cycles, and the PCR products were sequenced using the Illumina HiSeq platform (150-bp paired-end reads).

Pseudomolecule construction by Hi-C

Clean Hi-C reads were used with HiC-Pro (Servant et al., 2015) to map the Hi-C sequencing reads to the assembled contigs using the BWA-aln algorithm without any mismatches and with detection of valid contacts (Li and Durbin, 2009). The preassembled contigs split into 50-kb segments (on average) combined with uniquely matched Hi-C data were clustered, ordered, and directed onto the pseudochromosomes using LACHESIS software (Burton et al., 2013). Orientation errors with obvious discrete chromatin interaction patterns were manually adjusted to improve the chromosome-scale assembly quality. The final chromosome assemblies were divided into 100-kb bins with equal lengths. The interaction signals generated by the valid mapped read pairs between each bin were visualized in a heatmap.

Genome annotation

To annotate repetitive sequences, the assembled turnip genome was screened using LTR_FINDER (version 1.05) (Xu and Wang, 2007), MITE-Hunter (Han and Wessler, 2010), RepeatScout (version 1.0.5) (Price et al., 2005), and PILER-DF (version 2.4) (Edgar and Myers, 2005). All isolated sequences were then classified using PASTEClassifier (Hoede et al., 2014) and mapped using the Repbase database and RepeatMasker software (version 4.0.6) (Tarailo-Graovac and Chen, 2009). Using a substitution rate (r) of 7.3×10^{-9} substitutions per site per year (Exposito-Alonso et al., 2018), the insertion date (T) was calculated for each LTR retrotransposon (T = K/2r; K, genetic distance). Next, *ab initio*, homology-based, and RNA-seq-based prediction methods were combined to annotate gene models. The *ab initio* predictions were obtained using Genscan (Haas et al., 2008), Augustus (version 2.4) (Stanke and Morgenstern, 2005), GlimmerHMM (version 3.0.4) (Majoros et al., 2004), GeneID (version 1.4) (Blanco et al., 2007), and SNAP (version 2006-07-28) (Korf, 2004). GeMoMa software (version 1.3.1) was used to predict homologous species (mainly *A. thaliana*, *B. juncea*, *B. napus*, and Chiifu v3.0). Protein sequences were downloaded from the *Brassica* database (http://brassicadb.cn/#/). The RNA-seq reads were mapped to the genome assembly using HISAT and StringTie (Kim et al., 2015). TransDecoder (http://transdecoder.github.io) and GeneMarkS-T (Tang et al., 2015) were used to identify transcripts according to the mapping results. Finally, all

prediction results were integrated using EVidenceModeler (version 1.1.1) (Haas et al., 2008). Turnip genes were functionally annotated using the eggNOG, GO, KEGG_ko, and Pfam databases and the eggNOG online service (http://eggnog-mapper.embl.de/; Huerta-Cepas et al., 2019; Cantalapiedra et al., 2021). DIAMOND BLASTP (default alignment parameter) (Buchfink et al., 2021) was used to align turnip proteins to sequences in the NR and Swiss-Prot databases, with the best hit (-k 1) used for annotations.

Phylogenetic tree construction

Orthologous groups were identified using OrthoFinder (version 2.3.12) and all-versus-all BLASTP alignments ($E < 1e^{-5}$) with protein sequences encoded by the genomes of the following 12 species: turnip, *B. rapa* ssp. *pekinensis* (Chiifu-401-41, v3.5) (Zhang et al., 2022), *B. rapa* L. ssp. *pekinensis* cv. A03 (Chiifu A03) (Sun et al., 2022), European turnip ECD04 (Yang et al., 2022), Bras (Li et al., 2021), *B. rapa* ssp. *rapa* (TUA and TUE) (Cai et al., 2021), and *B. rapa* (Z1) (Chaisson and Tesler, 2012) with the AA genome; Bol (To1000, v2.0) (Cai et al., 2020) with the CC genome; Bni (Ni100_V2) (Perumal et al., 2020) with the BB genome; *Raphanus sativus* (Kitashiba et al., 2014; http://radish.kazusa.or.jp); and *A. thaliana* (*Athaliana*_447_Araport11) (Cheng et al., 2017) ($E < 1e^{-5}$, inflation factor = 1.5). Protein sequences encoded by single-copy genes were used to generate a multiple sequence alignment concatenated to a super alignment matrix. A maximum-likelihood phylogenetic tree was constructed according to the PROTCATJTT model in RAxML software (version 8.2.12) (Stamatakis, 2014). Species divergence times were estimated using MCMCtree in PAML (Yang, 2007) with an independent substitution rate (clock = 2) and GTR substitution model I. A Markov chain Monte Carlo analysis was run for 10 000 generations using a burn-in of 1000 iterations. Calibration points were applied according to the core Brassicaceae origin time of 21.3–29.8 mya (Guo et al., 2017). Homozygous gene pairs were identified for turnip, Chiifu A03, ECD04, and *A. thaliana*, and Ks values were calculated using WGDI (https://github.com/SunPengChuan/wgdi).

Determination of syntenic relationships between turnip and its relatives

Homologous genes were analyzed using MCScanX (Wang et al., 2012) with the following parameters: $E < 1e^{-10}$; Gap_penalty, −3. Syntenic blocks were defined as those with at least five syntenic genes. The sequence collinearity between turnip and other genomes was assessed using the nucmer program of the MUMmer package (v4.0rc1) (Marcais et al., 2018), and the syntenic relationships were visualized using NGenomeSyn (https://github.com/Hewm2008/NGenomeSyn). We assigned and partitioned multiple turnip or Chiifu A03 chromosomal segments that matched the same *A. thaliana* (*Athaliana*_447_Araport11) segment (24 ancestral crucifer blocks A–X) into the LF, MF1, and MF2 subgenomes (Schranz et al., 2006; Wang et al., 2011b).

Gene family expansion analysis

The expansion and contraction of gene families were determined using CAFE5 (default parameters) (https://github.com/hahnlab/CAFE5). Functional annotations were performed using eggNOG-mapper (http://eggnog-mapper.embl.de/). The GO annotation analysis was performed using TBtools_windows-x64_1_098685 (eggNOG-mapper Helper), and the results were visualized using online tools (http://www.bioinformatics.com.cn/).

Identification of pseudogenes in the *Brassica* diploid genomes

Pseudogenes were identified using two methods. Pseudogenes were first identified by examining the assembled genomes of turnip and seven other *Brassica* species (Chiifu v3.5, Chiifu A03, Z1, ECD04, Bol, Bni, and Bras) as described previously (Xie et al., 2019). In brief, the analysis consisted of five major steps. First, we identified intergenic regions (masked genic and transposon regions) with sequences similar to known proteins using Exonerate (https://github.com/nathanweeks/exonerate). The following steps focused on intergenic non-TE (transposable element) regions. We preliminarily screened the candidate pseudogene regions by comparing the genomic regions with known proteins; we accepted alignments with an E value of less than $1e^{-5}$, identity of 20% or greater, match length of 30 amino acids or more, and match length of 5% or greater of the query sequence. In chromosomal segments with multiple hits, the alignment hit with the best match was retained. Next, homologous segments were linked into contigs according to the distance between the hits on the chromosome (Gc) and the distance on the query protein (Gq); the distance was set to 50 bp. The candidate contigs were then realigned using a more accurate alignment program, tfasty, with the following parameters: -A -m 3 'q'. Accurate sequences and the positions of frameshifts and stop codons as well as insertions and deletions were determined in this step. In the final step, Exonerate was used to identify pseudogene–functional paralog pairs.

The second method was as follows. The CDSs and protein sequences were extracted from each genome using the gffread tool in Cufflinks (version 2.2.1). The annotated genes on the genome were masked to obtain the new genome mask_gene_genome.fa; the above pep was done after the query sequence using GenBlastA (version 1.0.4) (Gough and Chothia, 2002) to the new genome for homologous gene prediction using the following parameters: genblasta -P wublast -pg tblastn -q query.pep.fa -t mask_gene_genome.fa -p T -e 1e-5 -g T -f F -a 0.5 -d 100000 -r 10 -c 0.5 -s 0. Pseudogene prediction was performed using GeneWise v0.2 (Lees et al., 2012) to obtain the final results with the following parameters: -Identity 0.95 -cover 0.95.

Finally, the predicted pseudogenes were combined with the pseudogenes identified in the abovementioned search to obtain the final number of pseudogenes for each examined species. Numerical computations were performed at the Heifei Advanced Computing Center.

Pseudogene annotation and evolutionary analysis

Pseudogenes were annotated according to their functional paralogs in the non-redundant protein sequence (NR), Swiss-Prot, kyoto encyclopedia of genes and genomes (KEGG), gene ontology (GO), clusters of orthologous genes (COG), nucleotide sequence (NT), and Pfam databases. To evaluate the level of the selective constraint on the pseudogenes, we calculated the Ks and Ka values for each pseudogene and its closest functional paralog (Xie et al., 2019). First, collinear blocks in the genomes of turnip and the other species/subspecies were compared using MCScanX (Wang et al., 2012). We then extracted the pseudogene–functional paralog pairs, the pseudogenes in the other species/subspecies and the closest functional paralogs in turnip, and the turnip pseudogenes and the closest functional paralogs in the other species/subspecies. We subsequently extracted the paired nucleotide sequences separately and translated them into protein sequences for a comparison using multiple alignment using fast fourier transform (MAFFT, version 7.487) (Katoh and Standley, 2013). The protein sequence comparison results were converted to CDS comparison results using ParaAT. Finally, selection pressure was calculated using the KaKs Calculator (version 2.0) (Wang et al., 2010). A Fisher's test with KaKs < 3 was also performed.

Pseudogene Pfam domain analysis

We annotated all pseudogenes according to their functional paralogs in the Pfam database (Pfam-A.hmm) using HMMER 3.1b2 (February 2015) (http://hmmer.org/) with $\leq 1e^{-5}$ set as the threshold.

GSL extraction and analysis

GSL content was measured as described previously (Yang et al., 2020). In brief, 200 mg plant tissue was added to 80% (v/v) methanol solution containing 50 μL 1 mM sinalbin as an internal standard. The solution was mixed and then centrifuged. The collected supernatant was added to DEAE-Sephadex A-25 ion-exchange columns. The columns were washed with 80% methanol, double distilled H_2O (ddH_2O), and 20 mM [2(N-morpholino)ethanesulfonic acid] MES buffer (pH 5.2) before 30 μL sulfatase solution was applied. After overnight incubation at room temperature, the

eluted desulfo-GSLs were separated using a high-performance liquid chromatography system (HPLC, Agilent 1100) and a ultra-performance liquid chromatography/mass spectrometry/mass spectrometry (UPLC-MS/MS) system (LCMS-8040 system, Shimadzu) with a reverse-phase C18 column and a water–acetonitrile gradient. GSL content was calculated on the basis of the peak areas at 229 nm relative to the peak area of the internal standard using the recommended relative response factors reported in DIN EN ISO 9167. The results were calculated in terms of μmol/g fresh weight.

Generation of transgenic turnip and Chiifu hairy roots

The full-length CDSs of turnip *BrrMAM* genes (*Gene0495830*, *Gene0464890*, and *Gene0228790*) were cloned into the binary vector pRI101-AN to generate *35S::BrrMAM-GFP* constructs. For the RNAi constructs, the reverse complementary sequences of *BrrMAM* genes (*Gene0495830*, *Gene0464890*, *Gene0228790*) and *AOP2* genes (*Gene0405960*, *Gene0250680*, and *Gene0486840* in turnip; *BraA02g08320*, *BraA09g001360*, and *BraA03g029140* in Chiifu) were cloned into *pRI101-AN-FLAG* vectors. The resulting recombinant plasmids and the negative control vectors (*35S::GFP* and *35S::FLAG*) were inserted separately into *A. rhizogenes* strain LBA9402 cells.

Turnip and Chiifu hairy root cultures were established as described previously (Chung et al., 2016; Yin et al., 2020). In brief, a cotyledon infection method was used to insert the abovementioned genes into the turnip and Chiifu roots. The *35S::GFP* and *35S::FLAG* vectors were used as controls. The GSL content in the hairy roots was quantified using the UPLC-MS/MS system as described above. All primers are listed in Supplemental Table 19.

DATA AVAILABILITY

Raw Illumina and PacBio sequencing data and genome assembly data have been deposited in the Genome Sequence Archive in the China National Genomics Data Center (accession numbers CRA005412 and GWHBFXQ00000000).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences, Pan-Third Pole Environment Study for a Green Silk Road (Pan-TPE) (XDA2004010306); the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502); the National Natural Science Foundation of China (32070362, 41771123, and 32100315); the "Cross-Cooperative Team" of the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences; the Natural Science Foundation of Yunnan Province (202101AT070190); Digitalization, Development, and Application of Biotic Resource (202002AA100007); the Poverty Alleviation through Science and Technology Projects of the Chinese Academy of Sciences (KFJ-FP-201905); the Technology Transfer into Yunnan Project (202003AD150005); and the Postdoctoral Research Funding Projects of Yunnan Province (to X. Yin).

AUTHOR CONTRIBUTIONS

Conceptualization and Supervision, Yongping Yang and Yunqiang Yang; Methodology and Software, Y.Z., D.Y., X. Yang, and Z.Z.; Validation, X. Yin, X.S., X.K., and X.L.; Formal Analysis, X. Yin, G.W., Y.D., and Yunqiang Yang; Writing–Original Draft, X. Yin; Writing–Review & Editing, X. Yin and Yunqiang Yang; Funding Acquisition, Yongping Yang, Yunqiang Yang, and Y.D.

ACKNOWLEDGMENTS

No conflict of interest is declared.

Received: January 15, 2022
Revised: July 30, 2022
Accepted: August 29, 2022
Published: September 2, 2022

REFERENCES

Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R., and Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA. **107**:18724–18728. https://doi.org/10.1073/pnas.0909766107.

Bell, L., Oloyede, O.O., Lignou, S., Wagstaff, C., and Methven, L. (2018). Taste and flavor perceptions of glucosinolates, isothiocyanates, and related compounds. Mol. Nutr. Food Res. **62**:1700990. https://doi.org/10.1002/mnfr.201700990.

Benderoth, M., Pfalz, M., and Kroymann, J. (2009). Methylthioalkylmalate synthases: genetics, ecology and evolution. Phytochemistry Rev. **8**:255–268. https://doi.org/10.1007/s11101-008-9097-1.

Blanco, E., Parra, G., and Guigó, R. (2007). Using geneid to identify genes. Curr. Protoc. Bioinformatics **Chapter 4**, Unit 4.3. https://doi.org/10.1002/0471250953.bi0403s18.

Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422**:433–438. https://doi.org/10.1038/nature01521.

Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods **18**:366–368. https://doi.org/10.1038/s41592-021-01101-x.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. **31**:1119–1125. https://doi.org/10.1038/nbt.2727.

Cai, X., Wu, J., Liang, J., Lin, R., Zhang, K., Cheng, F., and Wang, X. (2020). Improved *Brassica oleracea* JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes. Theor. Appl. Genet. **133**:3187–3199. https://doi.org/10.1007/s00122-020-03664-3.

Cai, X., Chang, L., Zhang, T., Chen, H., Zhang, L., Lin, R., Liang, J., Wu, J., Freeling, M., and Wang, X. (2021). Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. Genome Biol. **22**:166. https://doi.org/10.1186/s13059-021-02383-2.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. **38**:5825–5829. https://doi.org/10.1093/molbev/msab293.

Chaisson, M.J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinf. **13**:238. https://doi.org/10.1186/1471-2105-13-238.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science **345**:950–953. https://doi.org/10.1126/science.1253435.

Chen, X., Zhu, Z., Gerendás, J., and Zimmermann, N. (2008). Glucosinolates in Chinese *Brassica campestris* vegetables: Chinese cabbage, purple cai-tai, choysum, pakehoi, and turnip. Hortscience **43**:571–574. https://doi.org/10.21273/hortsci.43.2.571.

Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J. **89**:789–804. https://doi.org/10.1111/tpj.13415.

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M.A., and Wang, X. (2013). Deciphering the diploid ancestral genome of the

Mesohexaploid *Brassica rapa*. Plant Cell **25**:1541–1554. https://doi.org/10.1105/tpc.113.110486.

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G., and Wang, X. (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PLoS One **7**:e36442.

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., Liu, B., Liang, J., Zhuang, M., Liu, Y., et al. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. Nat. Genet. **48**:1218–1224. https://doi.org/10.1038/ng.3634.

Chung, I.-M., Rekha, K., Rajakumar, G., and Thiruvengadam, M. (2016). Production of glucosinolates, phenolic compounds and associated gene expression profiles of hairy root cultures in turnip (*Brassica rapa* ssp *rapa*). Biotech **3**:6. https://doi.org/10.1007/s13205-016-0492-9.

Depree, J.A., M. Howard, T., and P. Savage, G. (1998). Flavour and pharmaceutical properties of the volatile sulphur compounds of Wasabi (*Wasabia japonica*). Food Res. Int. **31**:329–337. https://doi.org/10.1016/S0963-9969(98)00105-7.

Drewnowski, A., and Gomez-Carneros, C. (2000). Bitter taste, phytonutrients, and the consumer: a review. Am. J. Clin. Nutr. **72**:1424–1435. https://doi.org/10.1093/ajcn/72.6.1424.

Edgar, R.C., and Myers, E.W. (2005). PILER: identification and classification of genomic repeats. Bioinformatics **21**:i152–i158. https://doi.org/10.1093/bioinformatics/bti1003.

Engel, E., Baty, C., le Corre, D., Souchon, I., and Martin, N. (2002). Flavor-active compounds potentially implicated in cooked cauliflower acceptance. J. Agric. Food Chem. **50**:6459–6467. https://doi.org/10.1021/jf025579u.

Exposito-Alonso, M., Becker, C., Schuenemann, V.J., Reiter, E., Setzer, C., Slovak, R., Brachi, B., Hagmann, J., Grimm, D.G., Chen, J., et al. (2018). The rate and potential relevance of new mutations in a colonizing plant lineage. PLoS Genet. **14**:e1007155. https://doi.org/10.1371/journal.pgen.1007155.

Gough, J., and Chothia, C. (2002). SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. Nucleic Acids Res. **30**:268–272. https://doi.org/10.1093/nar/30.1.268.

Graham, N., and May, S. (2011). Genetics and Genomics of the Brassicaceae (New York: Springer).

Green, C., Willoughby, J., DDD Study, and Study, D.D.D. (2017). De novo SETD5 loss-of-function variant as a cause for intellectual disability in a 10-year old boy with an aberrant blind ending bronchus. Am. J. Med. Genet. **173**:3165–3171. https://doi.org/10.1002/ajmg.a.38461.

Gujas, B., Alonso-Blanco, C., and Hardtke, C.S. (2012). Natural *Arabidopsis* brx Loss-of-function alleles confer root adaptation to acidic soil. Curr. Biol. **22**:1962–1968. https://doi.org/10.1016/j.cub.2012.08.026.

Guo, N., Wang, S., Gao, L., Liu, Y., Wang, X., Lai, E., Duan, M., Wang, G., Li, J., Yang, M., et al. (2021). Genome sequencing sheds light on the contribution of structural variants to *Brassica oleracea* diversification. BMC Biol. **19**:93. https://doi.org/10.1186/s12915-021-01031-2.

Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X., Zhang, D., Ma, T., Hu, Q., Al-Shehbaz, I.A., et al. (2017). Plastome phylogeny and early diversification of Brassicaceae. BMC Genom. **18**:176. https://doi.org/10.1186/s12864-017-3555-3.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. **9**:R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Han, Y., and Wessler, S.R. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. **38**:e199. https://doi.org/10.1093/nar/gkq862.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., and Quesneville, H. (2014). PASTEC: an automatic transposable element classification tool. PLoS One **9**:e91929. https://doi.org/10.1371/journal.pone.0091929.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. **47**:D309–D314. https://doi.org/10.1093/nar/gky1085.

Ignatov, A.N., Artemyeva, A.M., and Hida, K. (2008). Origin and expansion of cultivated *Brassica rapa* in Eurasia: linguistic facts. 5th International Symposium on *Brassicas*/16th International Crucifer Genetics Workshop (*Brassica*). Acta Hortic. **867**:81–88. https://doi.org/10.17660/ActaHortic.2010.867.9.

Fimognari, C., Nüsse, M., Cesari, R., Iori, R., Cantelli-Forti, G., and Hrelia, P. (2002). Growth inhibition, cell-cycle arrest and apoptosis in human T-cell leukemia by the isothiocyanate sulforaphane. Carcinogenesis **23**:581–586. https://doi.org/10.1093/carcin/23.4.581.

Istace, B., Belser, C., Falentin, C., Labadie, K., Boideau, F., Deniot, G., Maillet, L., Cruaud, C., Bertrand, L., Chèvre, A.M., et al. (2021). Sequencing and chromosome-scale Assembly of plant genomes, *Brassica rapa* as a use case. Biology **10**:732. https://doi.org/10.3390/biology10080732.

Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J., Lopez, R., and Price, H.J. (2005). Evolution of genome size in Brassicaceae. Ann. Bot. **95**:229–235. https://doi.org/10.1093/aob/mci016.

Jones, R.B., Faragher, J.D., and Winkler, S. (2006). A review of the influence of postharvest treatments on quality and glucosinolate content in broccoli (*Brassica oleracea* var. *italica*) heads. Postharvest Biol. Technol. **41**:1–8. https://doi.org/10.1016/j.postharvbio.2006.03.003.

Kang, J.Y., Ibrahim, K.E., Juvik, J.A., Kim, D.H., and Kang, W.J. (2006). Genetic and environmental variation of glucosinolate content in Chinese cabbage. Hortscience **41**:1382–1385. https://doi.org/10.21273/hortsci.41.6.1382.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**:772–780. https://doi.org/10.1093/molbev/mst010.

Keurentjes, J.J.B., Fu, J., de Vos, C.H.R., Lommen, A., Hall, R.D., Bino, R.J., van der Plas, L.H.W., Jansen, R.C., Vreugdenhil, D., and Koornneef, M. (2006). The genetics of plant metabolism. Nat. Genet. **38**:842–849. https://doi.org/10.1038/ng1815.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357–360. https://doi.org/10.1038/nmeth.3317.

Kitashiba, H., Li, F., Hirakawa, H., Kawanabe, T., Zou, Z., Hasegawa, Y., Tonosaki, K., Shirasawa, S., Fukushima, A., Yokoi, S., et al. (2014). Draft sequences of the radish (*Raphanus sativus* L.) genome. DNA Res. **21**:481–490. https://doi.org/10.1093/dnares/dsu014.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722–736. https://doi.org/10.1101/gr.215087.116.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. **5**:59. https://doi.org/10.1186/1471-2105-5-59.

Kusznierewicz, B., Iori, R., Piekarska, A., Namieśnik, J., and Bartoszek, A. (2013). Convenient identification of desulfoglucosinolates on the basis of mass spectra obtained during liquid chromatography-diode array-electrospray ionisation mass spectrometry analysis: method verification for sprouts of different Brassicaceae species extracts. J. Chromatogr. A **1278**:108–115. https://doi.org/10.1016/j.chroma.2012.12.075.

Lee, J.G., Bonnema, G., Zhang, N., Kwak, J.H., de Vos, R.C.H., and Beekwilder, J. (2013). Evaluation of glucosinolate variation in a collection of turnip (*Brassica rapa*) germplasm by the analysis of intact and desulfo glucosinolates. J. Agric. Food Chem. **61**:3984–3993. https://doi.org/10.1021/jf400890p.

Lees, J., Yeats, C., Perkins, J., Sillitoe, I., Rentzsch, R., Dessailly, B.H., and Orengo, C. (2012). Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. Nucleic Acids Res. **40**:D465–D471. https://doi.org/10.1093/nar/gkr1181.

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics **30**:2843–2851. https://doi.org/10.1093/bioinformatics/btu356.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, P., Su, T., Zhao, X., Wang, W., Zhang, D., Yu, Y., Bayer, P.E., Edwards, D., Yu, S., and Zhang, F. (2021). Assembly of the non-heading pak choi genome and comparison with the genomes of heading Chinese cabbage and the oilseed yellow sarson. Plant Biotechnol. J. **19**:966–976. https://doi.org/10.1111/pbi.13522.

Liang, Y.S., Kim, H.K., Lefeber, A.W.M., Erkelens, C., Choi, Y.H., and Verpoorte, R. (2006). Identification of phenylpropanoids in methyl jasmonate treated *Brassica rapa* leaves using two-dimensional nuclear magnetic resonance spectroscopy. J. Chromatogr. A **1112**:148–155. https://doi.org/10.1016/j.chroma.2005.11.114.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A.P., Zhao, M., Ma, J., Yu, J., Huang, S., et al. (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. **5**:3930. https://doi.org/10.1038/ncomms4930.

Lou, P., Wu, J., Cheng, F., Cressman, L.G., Wang, X., and McClung, C.R. (2012). Preferential retention of circadian clock genes during diploidization following whole genome triplication in *Brassica rapa*. Plant Cell **24**:2415–2426. https://doi.org/10.1105/tpc.112.099499.

Lysak, M.A., Koch, M.A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. Genome Res. **15**:516–525. https://doi.org/10.1101/gr.3531105.

Ma, G., Wang, Y., and Xuan, Z. (2016). Analysis and comparison of nutritional compositions in Xinjiang turnip (*Brassica rapa* L.). Science & Technology of Food Industry **37**:360–364. https://doi.org/10.13386/j.issn1002-0306.2016.04.064.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**:2878–2879. https://doi.org/10.1093/bioinformatics/bth315.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. PLoS Comput. Biol. **14**:e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Moghe, G.D., Hufnagel, D.E., Tang, H., Xiao, Y., Dworkin, I., Town, C.D., Conner, J.K., and Shiu, S.-H. (2014). Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. Plant Cell **26**:1925–1937. https://doi.org/10.1105/tpc.114.124297.

Mun, J.H., Kwon, S.J., Yang, T.J., Seol, Y.J., Jin, M., Kim, J.A., Lim, M.H., Kim, J.S., Baek, S., Choi, B.S., et al. (2009). Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. Genome Biol. **10**:R111. https://doi.org/10.1186/gb-2009-10-10-r111.

Nagaharu, U. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn. J. Bot. **7**:389–452.

Padilla, G., Cartea, M.E., Velasco, P., de Haro, A., and Ordás, A. (2007). Variation of glucosinolates in vegetable crops of *Brassica rapa*. Phytochemistry **68**:536–545. https://doi.org/10.1016/j.phytochem.2006.11.017.

Park, S.G., Noh, E., Choi, S., Choi, B., Shin, I.G., Yoo, S.I., Lee, D.J., Ji, S., Kim, H.S., Hwang, Y.J., et al. (2021). Draft genome assembly and transcriptome dataset for European turnip (*Brassica rapa L.* ssp. *rapifera*), ECD4 carrying clubroot resistance. Front. Genet. **12**:651298. https://doi.org/10.3389/fgene.2021.651298.

Parveen, T., Hussain, A., and Someshwar Rao, M. (2015). Growth and accumulation of heavy metals in turnip (*Brassica rapa*) irrigated with different concentrations of treated municipal wastewater. Nord. Hydrol **46**:60–71. https://doi.org/10.2166/nh.2014.140.

Perumal, S., Koh, C.S., Jin, L., Buchwaldt, M., Higgins, E.E., Zheng, C., Sankoff, D., Robinson, S.J., Kagale, S., Navabi, Z.-K., et al. (2020). A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. Native Plants **6**:929–941. https://doi.org/10.1038/s41477-020-0735-y.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics **21**:i351–i358. https://doi.org/10.1093/bioinformatics/bti1018.

Qi, X., An, H., Ragsdale, A.P., Hall, T.E., Gutenkunst, R.N., Chris Pires, J., and Barker, M.S. (2017). Genomic inferences of domestication events are corroborated by written records in *Brassica rapa*. Mol. Ecol. **26**:3373–3388. https://doi.org/10.1111/mec.14131.

Sasaki, K., and Takahashi, T. (2002). A flavonoid from Brassica rapa flower as the UV-absorbing nectar guide. Phytochemistry **61**:339–343. https://doi.org/10.1016/s0031-9422(02)00237-6.

Schiessl, S., Samans, B., Huettel, B., Reinhard, R., and Snowdon, R.J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. Front. Plant Sci. **5**:3389.

Schranz, M.E., Lysak, M.A., and Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci. **11**:535–542.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. **16**:259. https://doi.org/10.1186/s13059-015-0831-x.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Song, K., Osborn, T.C., and Williams, P.H. (1990). *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPS).3. genome relationships in *Brassica* and related genera and the origin of *Brassica-oleracea* and *B-rapa* (SYN campestris). Theor. Appl. Genet. **79**:497–506. https://doi.org/10.1007/bf00226159.

Song, X., Wei, Y., Xiao, D., Gong, K., Sun, P., Ren, Y., Yuan, J., Wu, T., Yang, Q., Li, X., et al. (2021). *Brassica carinata* genome characterization clarifies U's triangle model of evolution and

polyploidy in *Brassica*. Plant Physiol. **186**:388–406. https://doi.org/10.1093/plphys/kiab048.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. **33**:W465–W467. https://doi.org/10.1093/nar/gki458.

Stotz, H.U., Sawada, Y., Shimada, Y., Hirai, M.Y., Sasaki, E., Krischke, M., Brown, P.D., Saito, K., and Kamiya, Y. (2011). Role of camalexin, indole glucosinolates, and side chain modification of glucosinolate-derived isothiocyanates in defense of *Arabidopsis* against *Sclerotinia sclerotiorum*. Plant J. **67**:81–93. https://doi.org/10.1111/j.1365-313X.2011.04578.x.

Sun, X., Li, X., Lu, Y., Wang, S., Zhang, X., Zhang, K., Su, X., Liu, M., Feng, D., Luo, S., et al. (2022). Construction of a high-density mutant population of Chinese cabbage facilitates the genetic dissection of agronomic traits. Mol. Plant **15**:913–924. https://doi.org/10.1016/j.molp.2022.02.006.

Suzuki, C., Ohnishi-Kameyama, M., Sasaki, K., Murata, T., and Yoshida, M. (2006). Behavior of glucosinolates in pickling cruciferous vegetables. J. Agric. Food Chem. **54**:9430–9436. https://doi.org/10.1021/jf061789l.

Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. **43**:e78. https://doi.org/10.1093/nar/gkv227.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Current Protocols in Bioinformatics **25**:4. https://doi.org/10.1002/0471250953.bi0410s25.

Tortorella, S.M., Royce, S.G., Licciardi, P.V., and Karagiannis, T.C. (2015). Dietary sulforaphane in cancer chemoprevention: the role of epigenetic regulation and HDAC inhibition. Antioxidants Redox Signal. **22**:1382–1424. https://doi.org/10.1089/ars.2014.6097.

Town, C.D., Cheung, F., Maiti, R., Crabtree, J., Haas, B.J., Wortman, J.R., Hine, E.E., Althoff, R., Arbogast, T.S., Tallon, L.J., et al. (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell **18**:1348–1359. https://doi.org/10.1105/tpc.106.041665.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9**:e112963. https://doi.org/10.1371/journal.pone.0112963.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Dev. Reprod. Biol. **8**:77–80. https://doi.org/10.1016/s1672-0229(10)60008-3.

Wang, H., Wu, J., Sun, S., Liu, B., Cheng, F., Sun, R., and Wang, X. (2011a). Glucosinolate biosynthetic genes in *Brassica rapa*. Gene **487**:135–142. https://doi.org/10.1016/j.gene.2011.07.021.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., et al. (2011b). The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. **43**:1035–1039. https://doi.org/10.1038/ng.919.

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**:e49. https://doi.org/10.1093/nar/gkr1293.

Wentzell, A.M., Rowe, H.C., Hansen, B.G., Ticconi, C., Halkier, B.A., and Kliebenstein, D.J. (2007). Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. PLoS Genet. **3**:1687–1701. https://doi.org/10.1371/journal.pgen.0030162.

Will, J.L., Kim, H.S., Clarke, J., Painter, J.C., Fay, J.C., and Gasch, A.P. (2010). Incipient balancing selection through adaptive loss of aquaporins in natural saccharomyces cerevisiae populations. PLoS Genet. **6**:e1000893. https://doi.org/10.1371/journal.pgen.1000893.

Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. **2**:333–341. https://doi.org/10.1038/35072009.

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H., et al. (2019). Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. Mol. Plant **12**:30–43. https://doi.org/10.1016/j.molp.2018.11.007.

Wu, W., Liu, X., Wang, M., Meyer, R.S., Luo, X., Ndjiondjop, M.N., Tan, L., Zhang, J., Wu, J., Cai, H., et al. (2017). A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. Native Plants **3**:17064. https://doi.org/10.1038/nplants.2017.64.

Xie, J., Li, Y., Liu, X., Zhao, Y., Li, B., Ingvarsson, P.K., and Zhang, D. (2019). Evolutionary origins of pseudogenes and their association with regulatory sequences in plants. Plant Cell **31**:563–578. https://doi.org/10.1105/tpc.18.00601.

Xu, Y.-C., Niu, X.-M., Li, X.-X., He, W., Chen, J.-F., Zou, Y.-P., Wu, Q., Zhang, Y.E., Busch, W., and Guo, Y.-L. (2019). Adaptation and phenotypic diversification in *Arabidopsis* through loss-of-function mutations in protein-coding genes. Plant Cell **31**:1012–1025. https://doi.org/10.1105/tpc.18.00791.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268. https://doi.org/10.1093/nar/gkm286.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat. Genet. **48**:1225–1232. https://doi.org/10.1038/ng.3657.

Yang, Y., Pu, Y., Yin, X., Du, J., Zhou, Z., Yang, D., Sun, X., Sun, H., and Yang, Y. (2019). A splice variant of *BrrWSD1* in turnip (*Brassica rapa* var. *rapa*) and its possible role in wax ester synthesis under drought stress. J. Agric. Food Chem. **67**:11077–11088. https://doi.org/10.1021/acs.jafc.9b04069.

Yang, Y., Hu, Y., Yue, Y., Pu, Y., Yin, X., Duan, Y., Huang, A., Yang, Y., and Yang, Y. (2020). Expression profiles of glucosinolate biosynthetic genes in turnip (*Brassica rapa* var. *rapa*) at different developmental stages and effect of transformed flavin-containing monooxygenase genes on hairy root glucosinolate content. J. Sci. Food Agric. **100**:1064–1071. https://doi.org/10.1002/jsfa.10111.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591. https://doi.org/10.1093/molbev/msm088.

Yang, Z., Jiang, Y., Gong, J., Li, Q., Dun, B., Liu, D., Yin, F., Yuan, L., Zhou, X., Wang, H., et al. (2022). R gene triplication confers European fodder turnip with improved clubroot resistance. Plant Biotechnol. J. **20**:1502–1517. https://doi.org/10.1111/pbi.13827.

Yin, X., Yang, Y., Lv, Y., Li, Y., Yang, D., Yue, Y., and Yang, Y. (2020). BrrICE1.1 is associated with putrescine synthesis through regulation of the arginine decarboxylase gene in freezing tolerance of turnip (*Brassica rapa* var. *rapa*). BMC Plant Biol. **20**:504. https://doi.org/10.1186/s12870-020-02697-6.

Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F., Liang, J., Cai, C., Liu, Z., Liu, B., et al. (2018). Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation

capture technologies. Hortic. Res. **5**:50. https://doi.org/10.1038/s41438-018-0071-9.

Zhang, N., Zhao, J., Lens, F., de Visser, J., Menamo, T., Fang, W., Xiao, D., Bucher, J., Basnet, R.K., Lin, K., et al. (2014). Morphology, carbohydrate composition and vernalization response in a genetically diverse collection of Asian and European turnips (*Brassica rapa* subsp. *rapa*). PLoS One **9**:e114241. https://doi.org/10.1371/journal.pone.0114241.

Zhang, Z., Guo, J., Cai, X., Li, Y., Xi, X., Lin, R., Liang, J., Wang, X., and Wu, J. (2022). Improved reference genome annotation of *Brassica rapa* by pacific biosciences RNA sequencing. Front. Plant Sci. **13**:841618. https://doi.org/10.3389/fpls.2022.841618.

Zhang, Z., Harrison, P.M., Liu, Y., and Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. **13**:2541–2558. https://doi.org/10.1101/gr.1429003.

Zhao, J., Wang, X., Deng, B., Lou, P., Wu, J., Sun, R., Xu, Z., Vromans, J., Koornneef, M., and Bonnema, G. (2005). Genetic relationships within *Brassica rapa* as inferred from AFLP fingerprints. Theor. Appl. Genet. **110**:1301–1314. https://doi.org/10.1007/s00122-005-1967-y.

Zheng, Y., Luo, L., Liu, Y., Yang, Y., Wang, C., Kong, X., and Yang, Y. (2018). Effect of vernalization on tuberization and flowering in the Tibetan turnip is associated with changes in the expression of FLC homologues. Plant Divers. **40**:50–56. https://doi.org/10.1016/j.pld.2018.01.002.

Zou, C., Lehti-Shiu, M.D., Thibaud-Nissen, F., Prakash, T., Buell, C.R., and Shiu, S.-H. (2009). Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. Plant Physiol. **151**:3–15. https://doi.org/10.1104/pp.109.140632.

Plant Communications, Volume *4*

Supplemental information

Differences in pseudogene evolution contributed to the contrasting flavors of turnip and Chiifu, two *Brassica rapa* subspecies

Xin Yin, Danni Yang, Youjie Zhao, Xingyu Yang, Zhili Zhou, Xudong Sun, Xiangxiang Kong, Xiong Li, Guangyan Wang, Yuanwen Duan, Yunqiang Yang, and Yongping Yang

Differences in pseudogene evolution contributed to the contrasting flavors of turnip and Chiifu, two *Brassica rapa* subspecies

Xin Yin[1,2,3,6], Danni Yang[1,2,3,5,6], Youjie Zhao[4,6], Xingyu Yang[1,2,3,5,6], Zhili Zhou[1,2,3,6], Xudong Sun[1,2,3], Xiangxiang Kong[1,2,3], Xiong Li[1,2,3], Guangyan Wang[1,2,3], Yuanwen Duan[1,2,3], Yunqiang Yang[1,2,3*], Yongping Yang[1,2,3*]

[1]Plant Germplasm and Genomics Center, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[2]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Science, Kunming, 650204, China

[3]Institute of Tibetan Plateau Research at Kunming, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China

[4]College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming, Yunnan, China

[5]University of Chinese Academy of Sciences, Beijing, 100049, China

[6] These authors contributed equally to this work.

*Corresponding author:

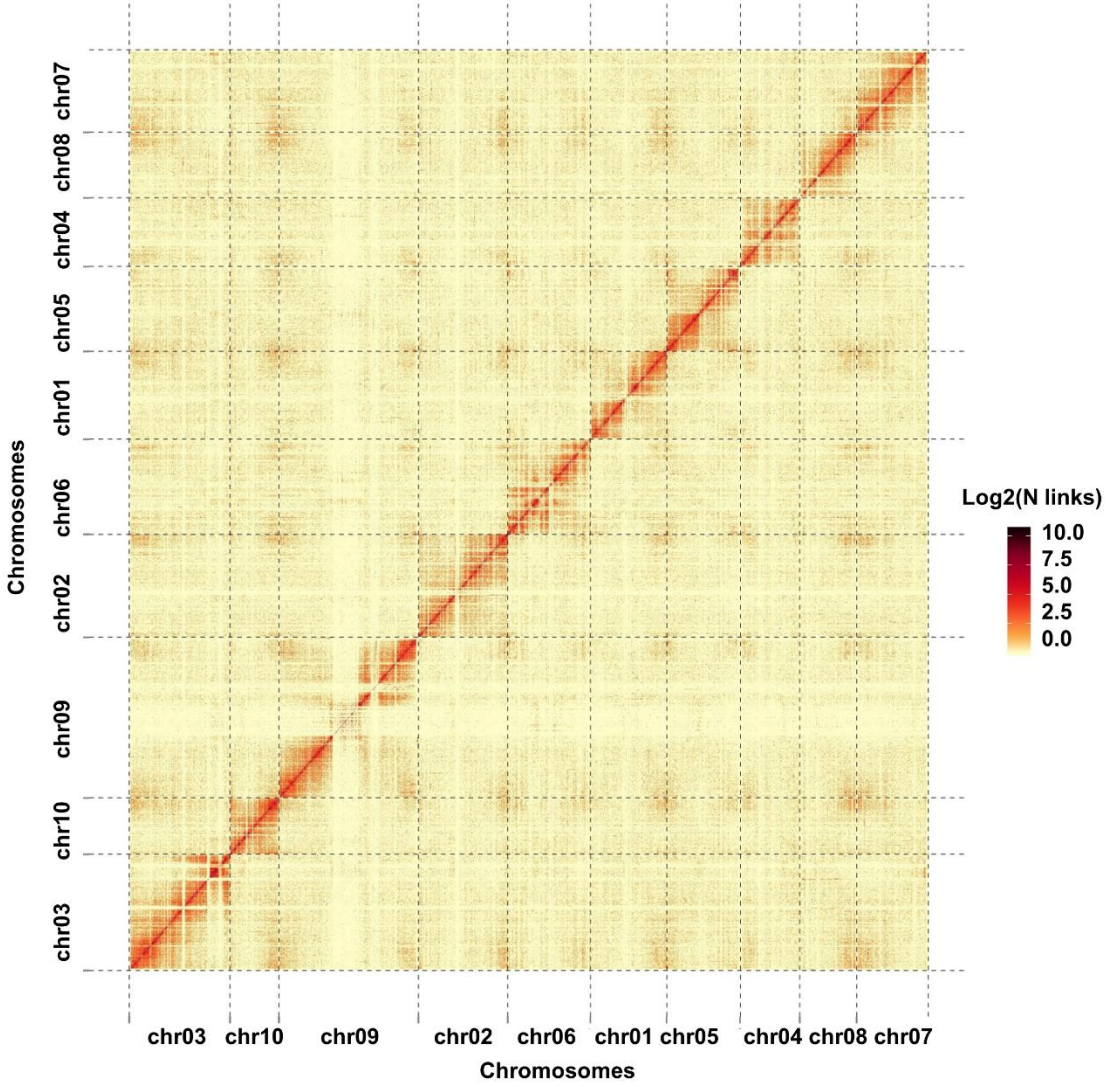
Yunqiang Yang

Email: yangyunqiang@mail.kib.ac.cn, Phone: 86-871-65223398

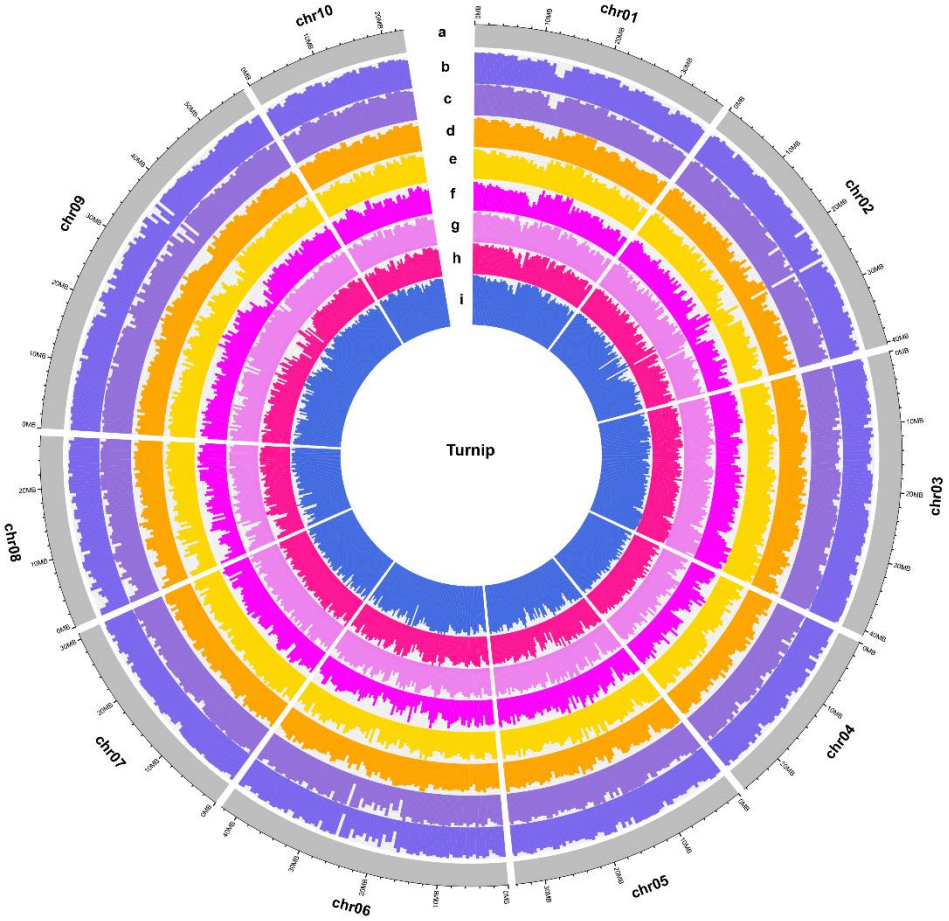
Yongping Yang

Email: yangyp@mail.kib.ac.cn, Phone: 86-871-65223398, ORCID: 0000-0002-0327-2664

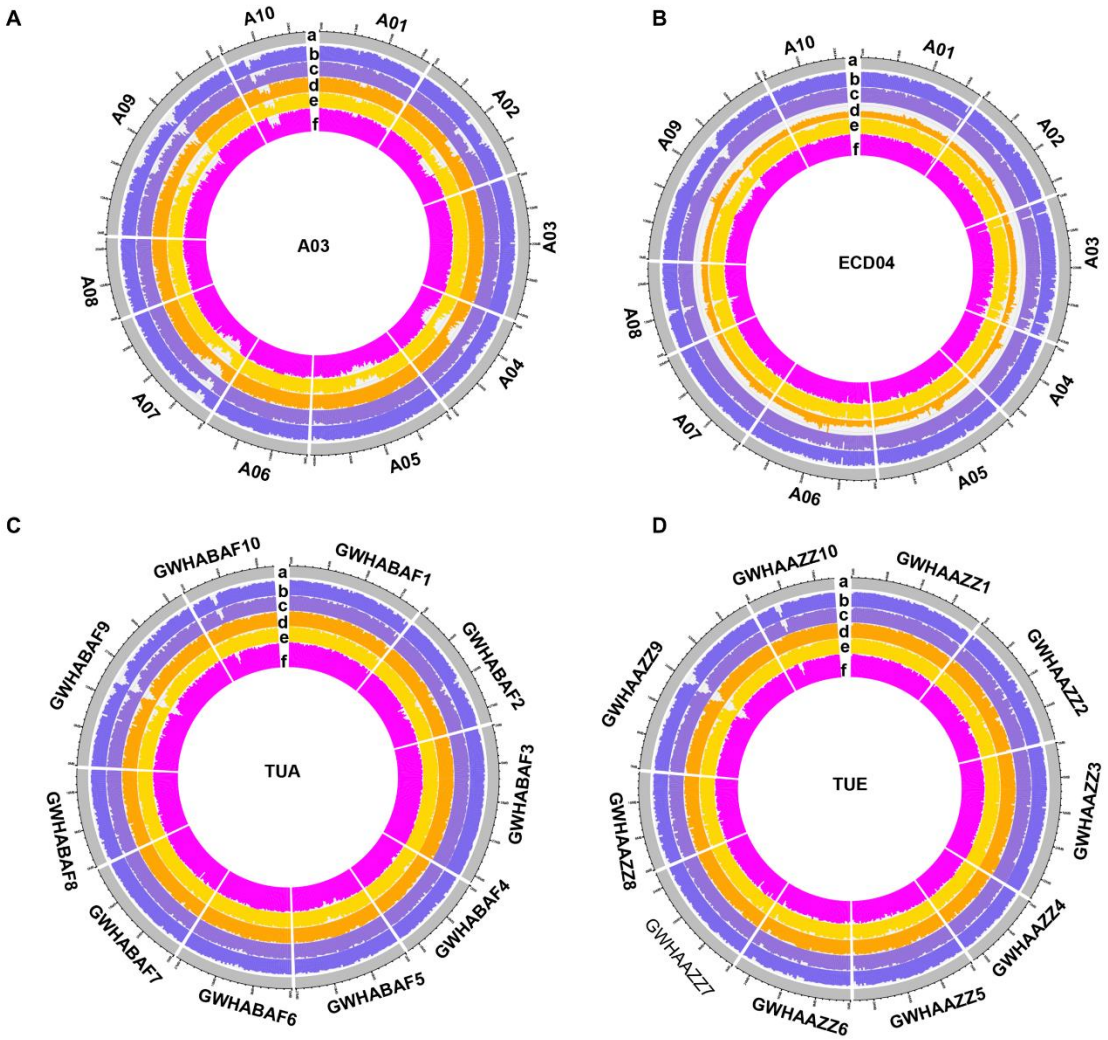
Supplemental Figures



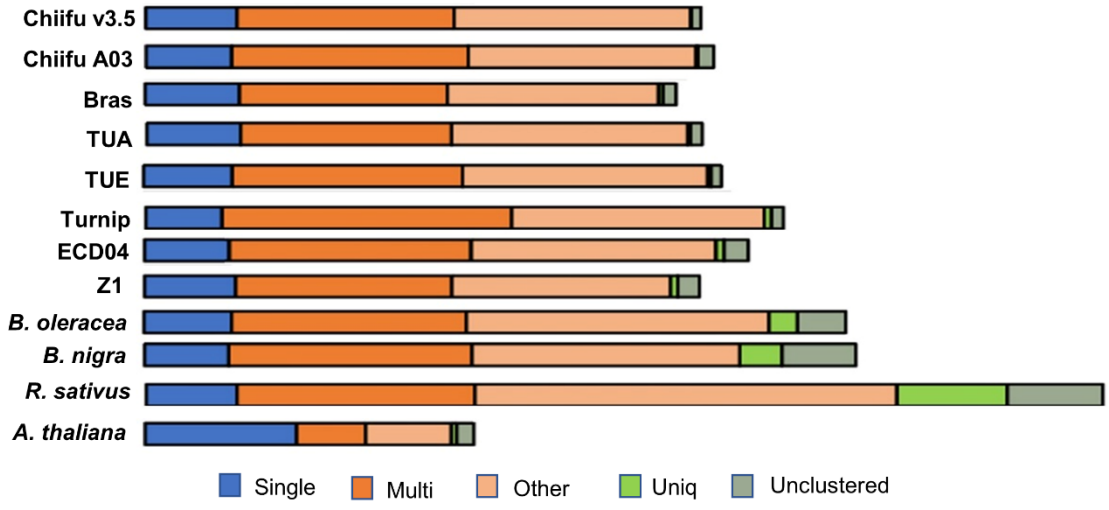
Supplemental Figure 1. Hi-C-assisted assembly of turnip pseudomolecules. The heat map presents the Hi-C chromosomal interactions (100 kb resolution). chr01–chr10, 10 turnip chromosomes. The x- and y-axes present the order of the positions of scaffolds on the corresponding pseudochromosomes.



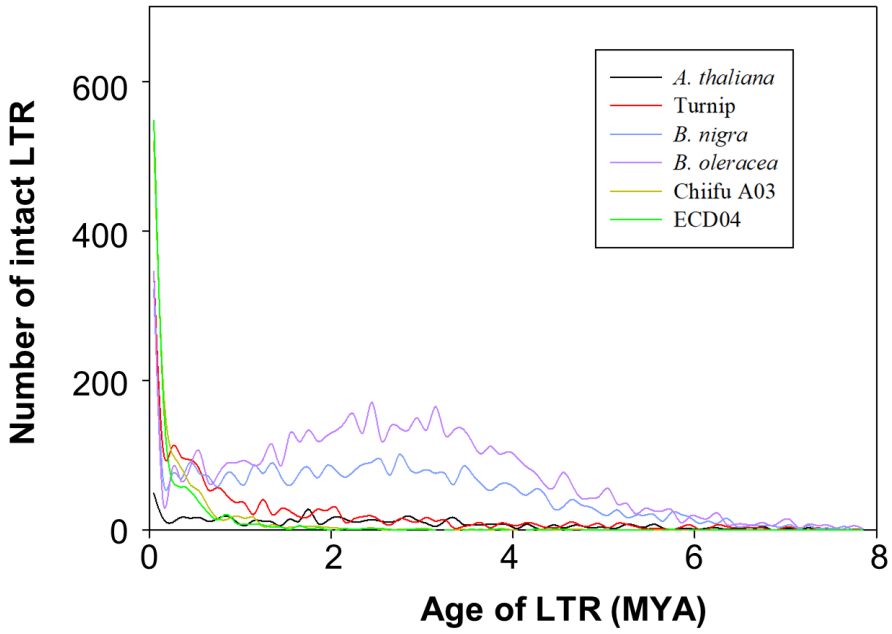
Supplemental Figure 2. Turnip, ECD04, TUA, and TUE Illumina paired-end reads (NGS) and 48A resequencing reads (48Areseq 1–3) from a turnip population were mapped to turnip chromosomes. a, chr1–chr10, circular representation of the pseudomolecules of turnip; b–f, turnip-DO1 NGS, turnip-DO2 NGS, ECD04 NGS, TUA NGS, and TUE NGS reads mapped to turnip chromosomes; g–i, 48A resequencing reads (48Areseq 1–3) from a turnip population mapped to turnip chromosomes.



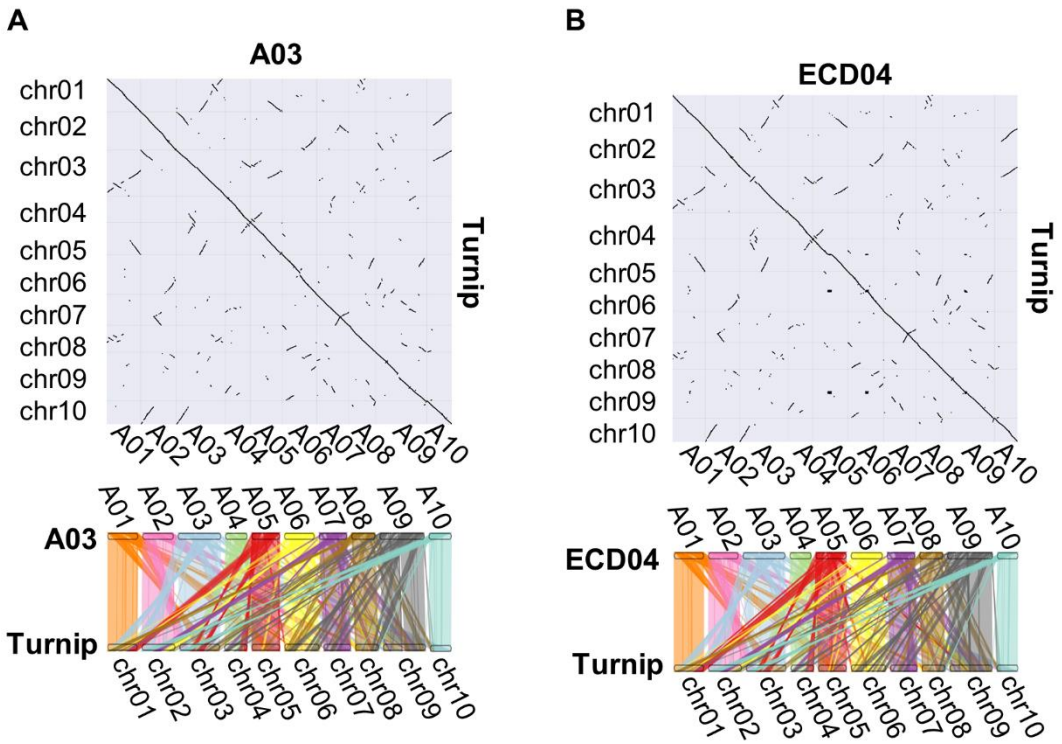
Supplemental Figure 3. Turnip Illumina paired-end reads (NGS) and 48A resequencing reads (48Areseq 1–3) from a turnip population were mapped to A03, ECD04, TUA, and TUE chromosomes (A–D). a, circular representation of the pseudomolecules of A03, ECD04, TUA, and TUE; b–c, turnip-DO1 NGS and turnip-DO2 NGS reads; d–f, 48A resequencing reads (48Areseq 1–3) from a turnip population mapped to turnip chromosomes.



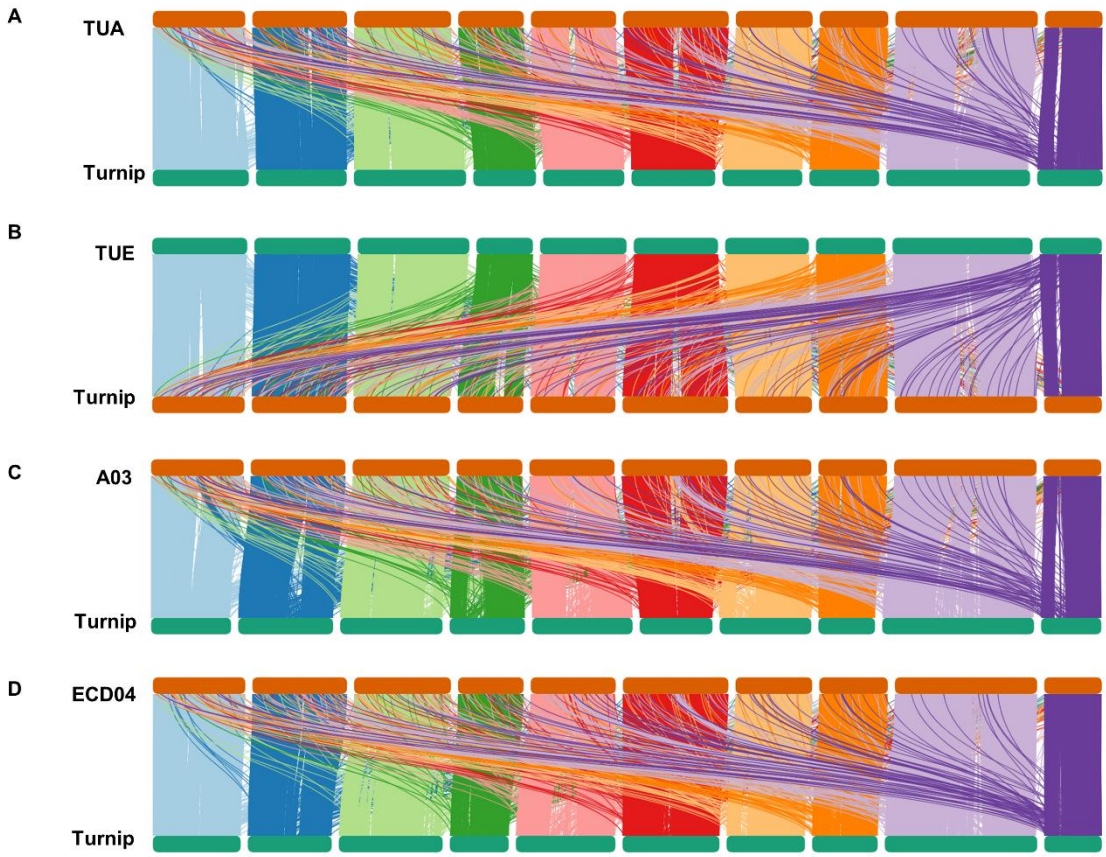
Supplemental Figure 4. Comparison of the number of gene families in the turnip genome and in the genomes of other diploid *Brassica* species and *A. thaliana*.



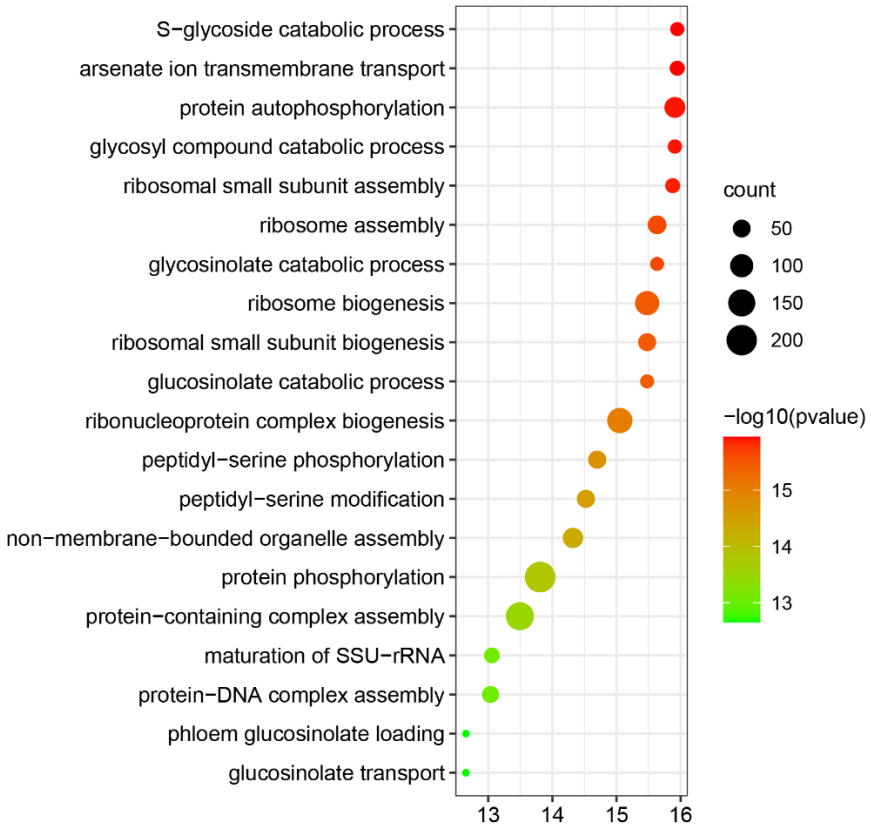
Supplemental Figure 5. Distribution of the LTR retrotransposon insertion times.



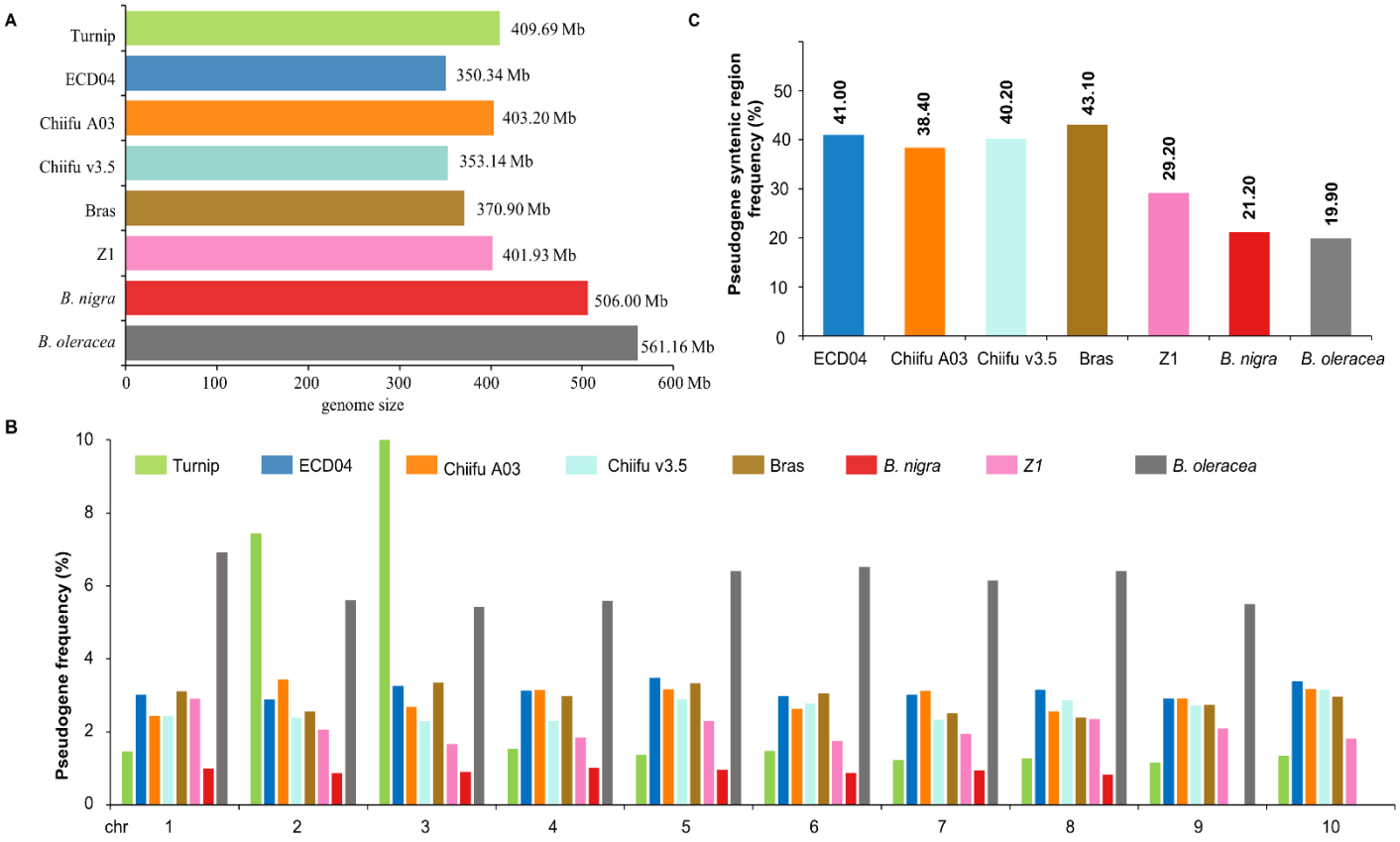
Supplemental Figure 6. Dot plot for the segmental collinearity between the turnip and Chiifu A03 genomes (A) and between the turnip and ECD04 genomes (B). Chiifu A03 and ECD04 chromosomes are presented in different colors, whereas orthologous chromosomal segments in turnip are presented in the same color. Conserved collinear blocks of gene models are presented for the 10 turnip chromosomes and the ECD04 and Chiifu A03 genomes.



Supplemental Figure 7. Sequence collinearity between the turnip genome and the TUA, TUE, Chiifu A03, and ECD04 genomes as determined by MUMmer.



Supplemental Figure 8. Results of the GO enrichment analysis of the most significantly expanded gene families in Chiifu A03. The enrichment factor, which indicates the degree of enrichment, was calculated as the ratio between the number of genes in the expanded families and all of the annotated genes in the pathway. The 15 most significantly enriched pathways are shown.

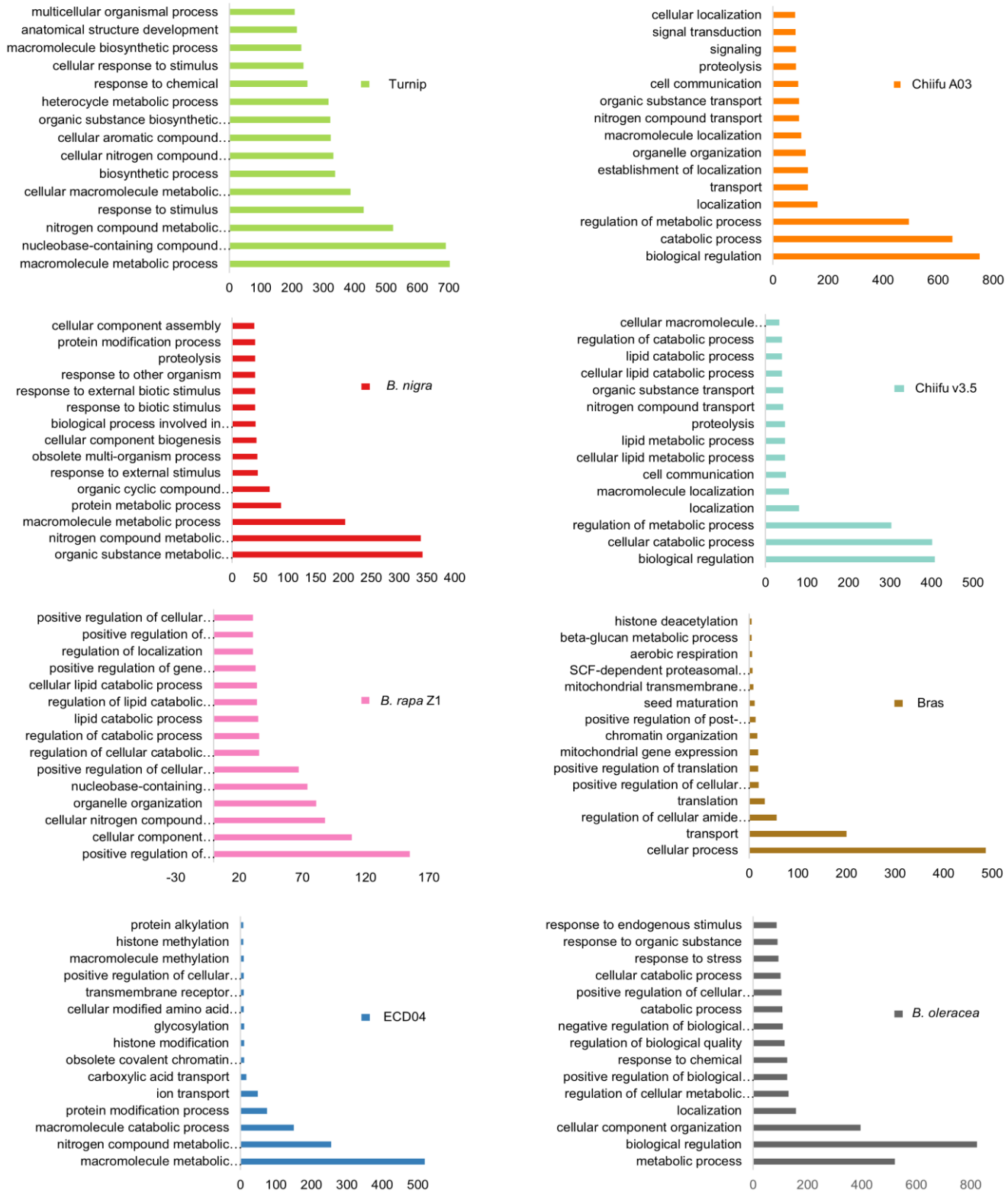


Supplemental Figure 9. Comparison of the pseudogenes in *Brassica* species/subspecies.

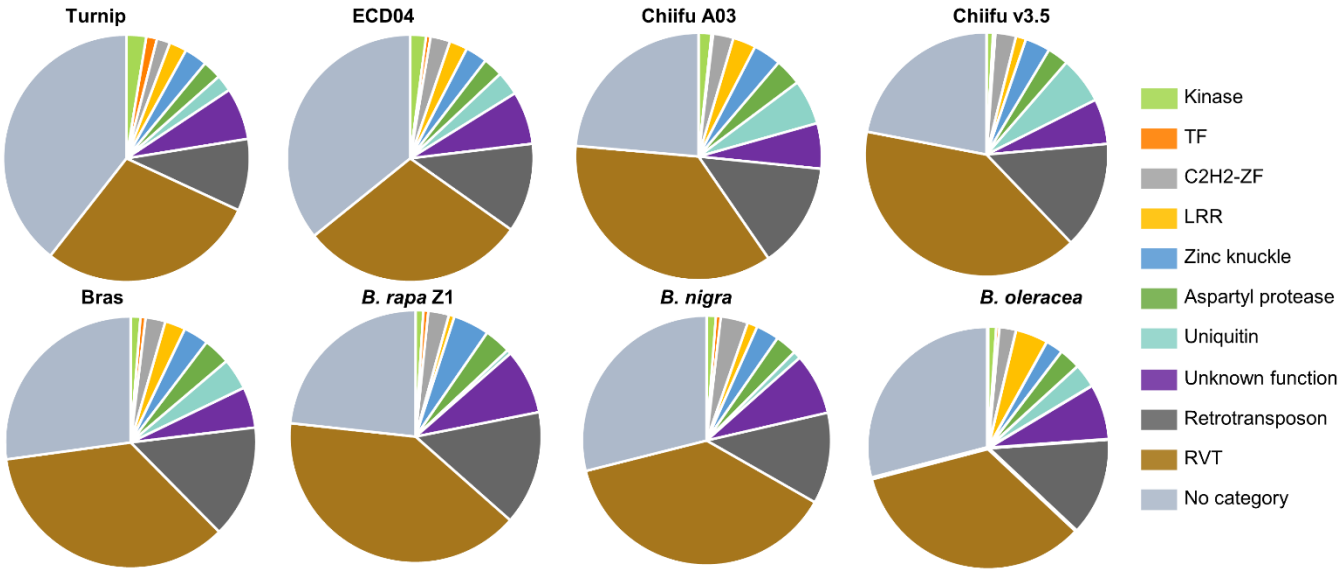
(A) Comparison of the genome sizes of *Brassica* species/subspecies, including turnip, ECD04, Chiifu A03, Chiifu v3.5, *B. rapa_chinensis* (Bras), *B. rapa* Z1, *B. nigra*, and *B. oleracea*

(B) Comparison of the pseudogene frequency (%) for each chromosome in *Brassica* species/subspecies. The y-axis presents the ratio of the pseudogene length to the chromosome length in *Brassica* species/subspecies. The x-axis presents the corresponding chromosome numbers.

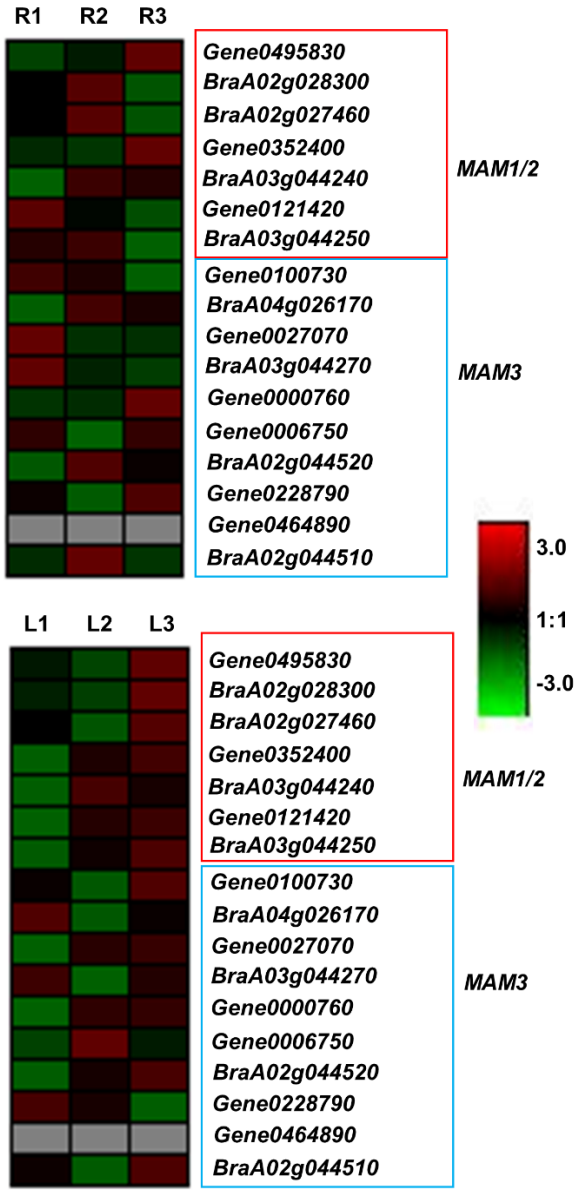
(C) Synteny (%) between turnip and other species/subspecies.



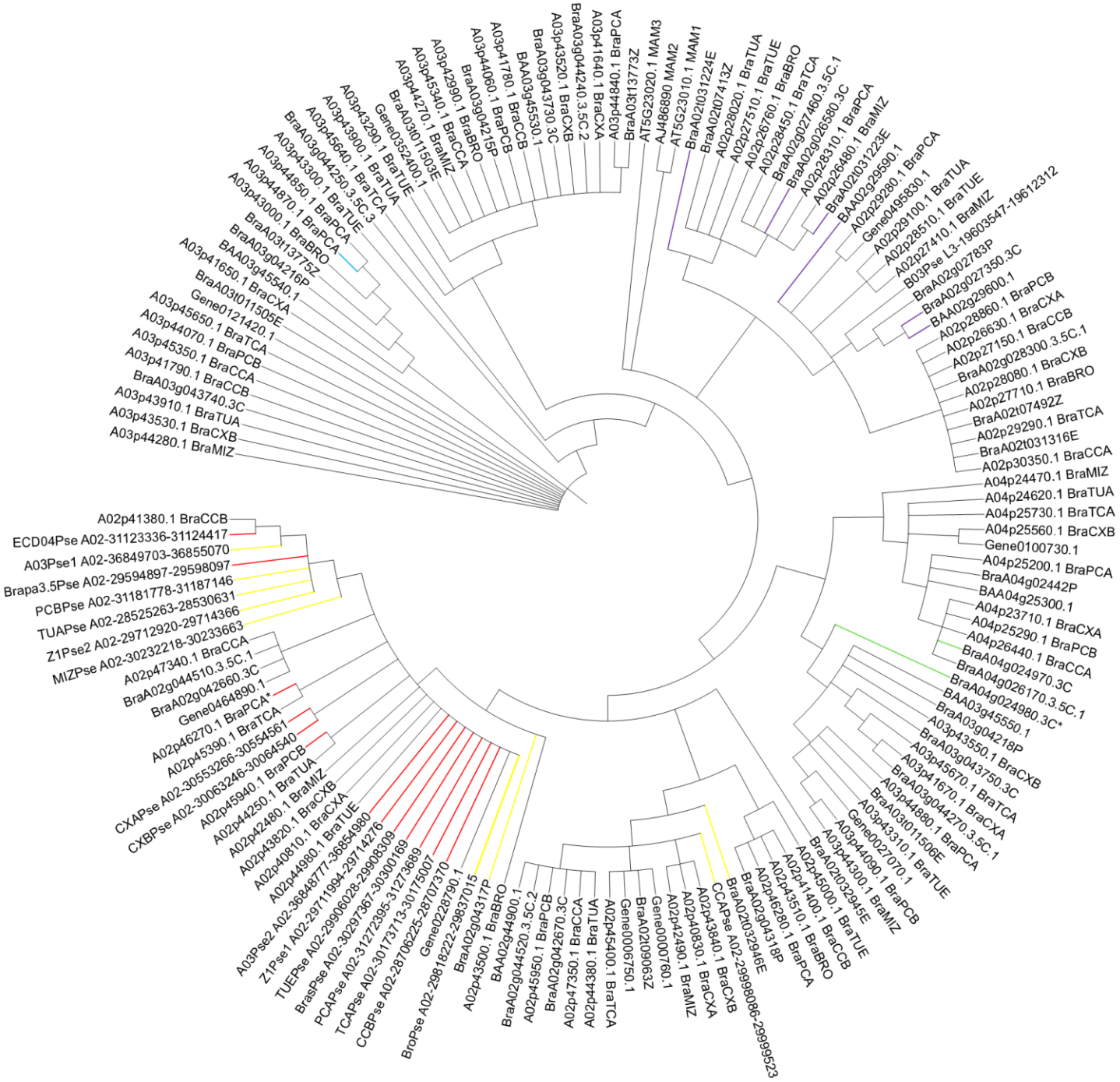
Supplemental Figure 10. Gene Ontology (GO)-based functional annotation of pseudogenes in *Brassica* species/subspecies. The GO enrichment analysis of the pseudogenes in *Brassica* species/subspecies was performed on the basis of the annotations of the closest functional paralogs.



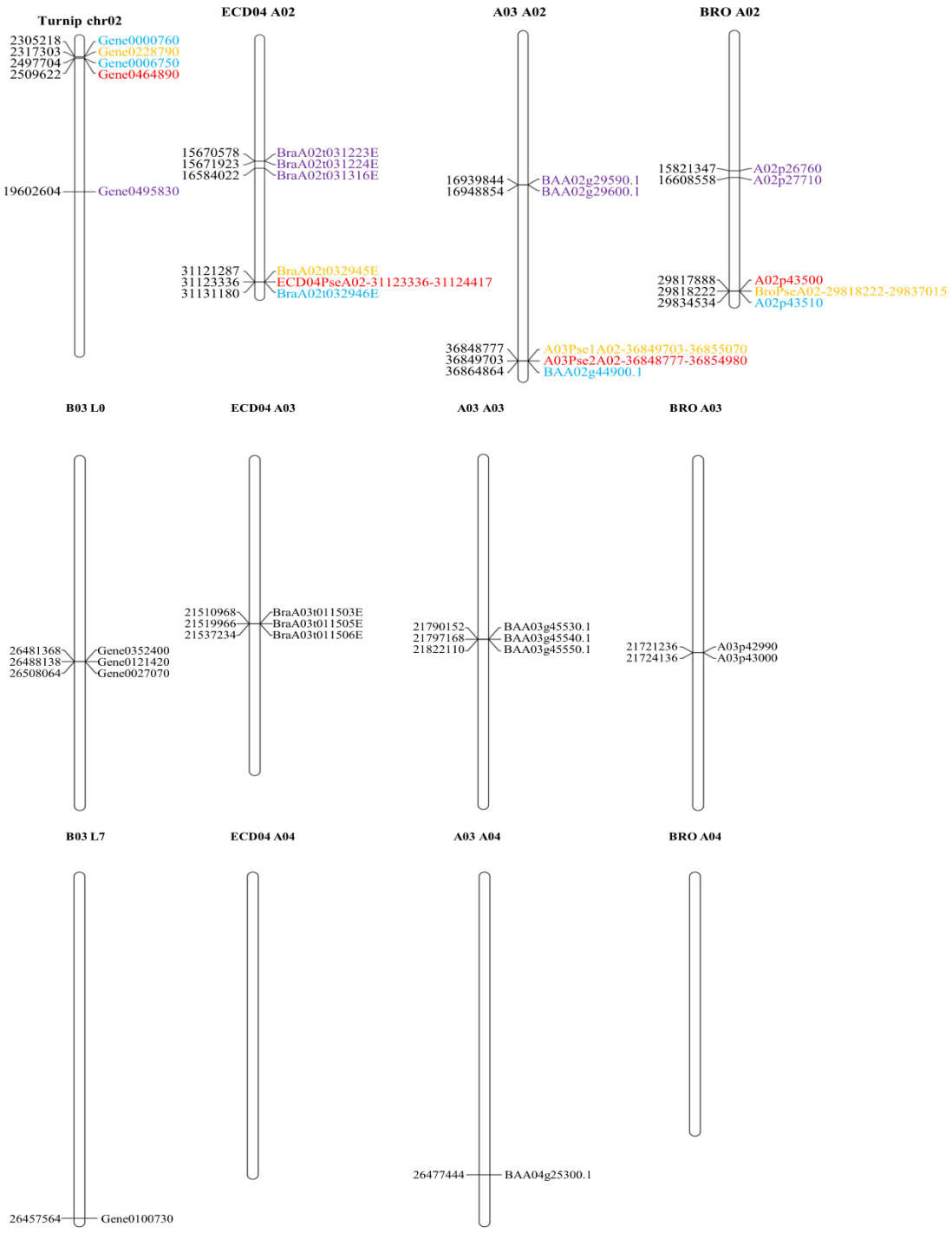
Supplemental Figure 11. Functional annotation of pseudogenes in *Brassica* species/subspecies according to Pfam domains.



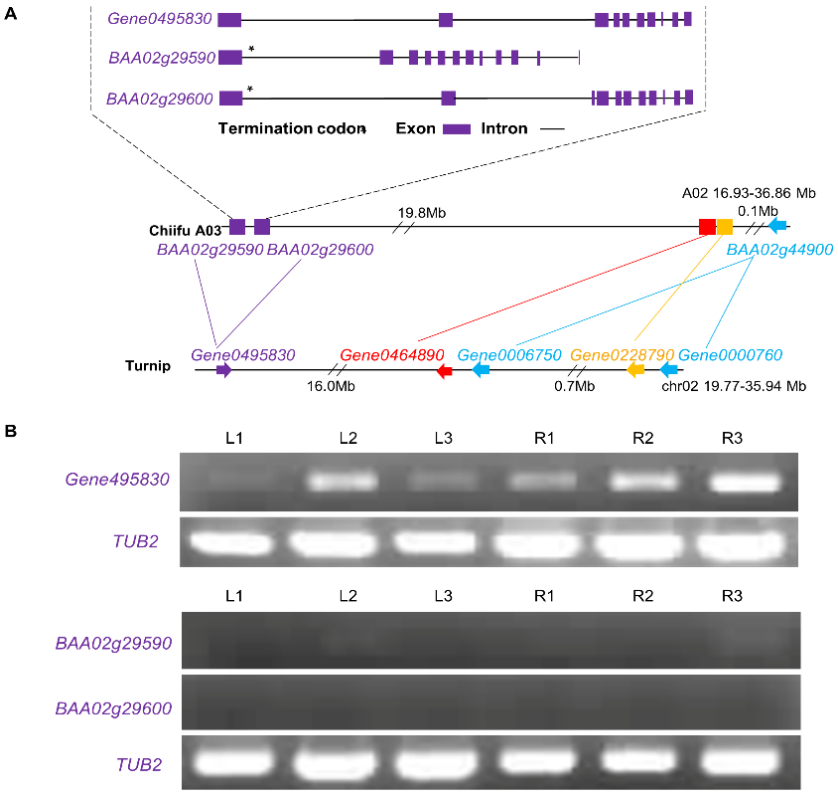
Supplemental Figure 12. Heat map of the quantitative real-time PCR (qRT-PCR) data for the turnip and Chiifu v3.5 *MAM* genes in different developmental stages. The qRT-PCR analysis was completed using three biological and technical replicates. R1, R2, and R3 represent the taproots collected at 10, 20, and 30 days after germination, respectively. L1, L2, and L3 represent the leaves collected at 10, 20, and 30 days after germination, respectively. Gene expression levels are presented in the colored bar. *TUB2* as reference (LOC103873913).



160 **Supplemental Figure 13. Phylogenetic tree comprising *MAM* genes and**
161 **pseudogenes.** Neighbor-joining trees consisting of *MAM* genes were constructed by
162 aligning coding sequences, with 1,000 bootstrap replicates. The *A. thaliana* genome
163 includes only *MAM1* and *MAM3*. Colored lines represent pseudogenes, with numbers
164 indicating the position of the pseudogene on the chromosome.
165



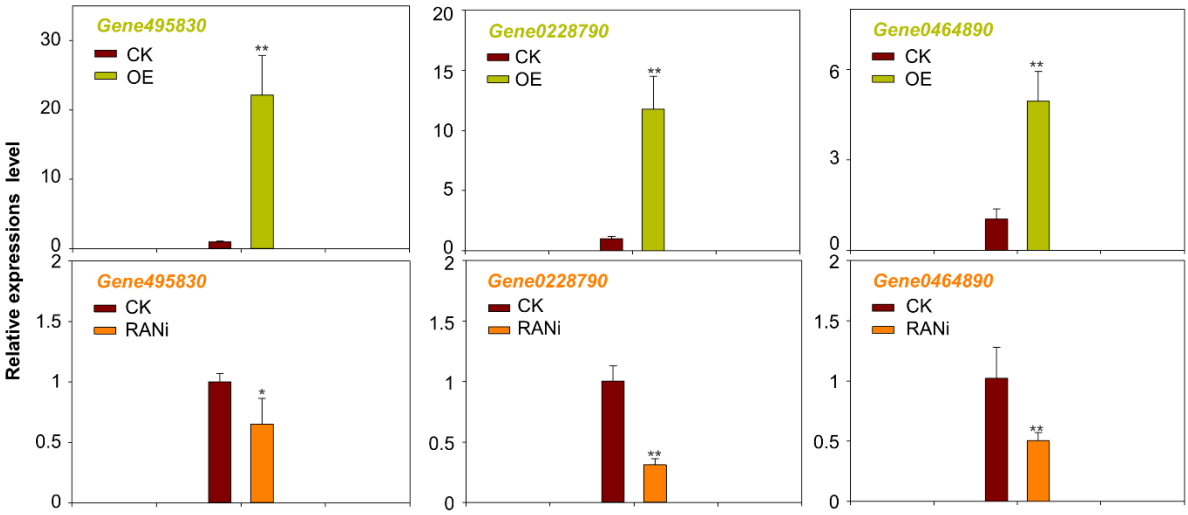
Supplemental Figure 14. Distribution of *MAM* genes on the turnip, ECD04, Chiifu A03, and BRO chromosomes. Turnip has the most *MAM* functional genes, which are distributed on chromosomes chr02, chr03, and chr04; the collinear chromosomes in other species/subspecies are A02, A03, and A04, respectively. Identical colors represent the same homologous regions on turnip chr02 and ECD04, Chiifu A03, and BRO A02 chromosomes.



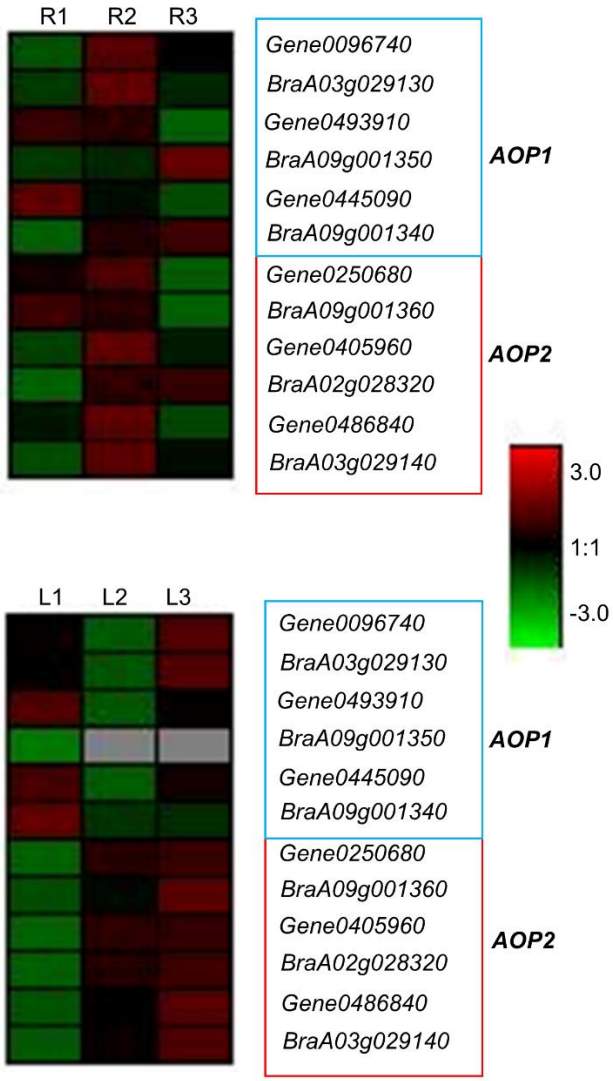
Supplemental Figure 15. Identification of pseudogenes.

(A) Analysis of the synteny among *MAM* genes in turnip and Chiifu A03. Turnip chromosome chr02 with *MAM* genes and the collinear chromosome A02 in Chiifu A03 are presented. Identical colors represent the same homologous region. Lines indicate genes with colinearity. Arrows indicate the gene orientation on the chromosome. Boxes represent pseudogenes in Chiifu A03. Specifically, dashed box represents *MAM* (*Gene0495830*) in turnip converted into pseudogenes (*BAA02g29590* and *BAA02g29600*) due to codon termination (asterisks) in Chiifu A03. Syntenic regions in both genomes, with one turnip genome containing two functional gene (*Gene0464890* and *Gene0228790*) and others containing homologous sequence in Chiifu A03 genome with clear markers indicative of a pseudogene, are presented.

(B) Validation of pseudogenes. The expression levels of pseudogenes *BAA02g29590* and *BAA02g29600* in Chiifu at different developmental stages (L1–L3 in leaves and R1–R3 in taproots) were compared with the expression levels of the functional gene *Gene0495830* in turnip via semi-quantitative PCR (RT-PCR). L1, L2, and L3 and R1, R2, and R3 represent the leaf samples and taproot samples collected at 10, 20, and 30 days after germination, respectively. *ACT* as reference.



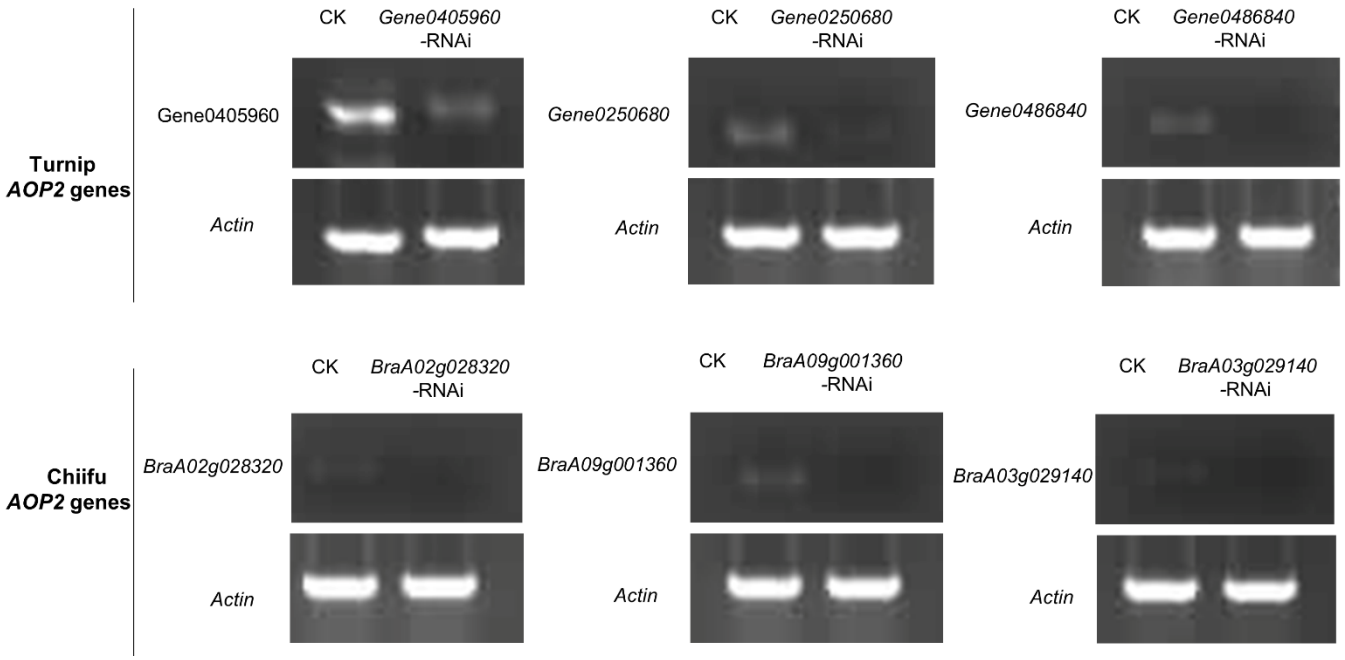
Supplemental Figure 16. Overexpression (OE) of three turnip *MAM* functional genes in Chiifu hairy roots (top) and silencing of these genes in turnip hairy roots via RNAi (bottom). Non-transgenic roots served as the control (CK). The analysis was completed using three biological and technical replicates. Error bars indicate the standard deviation. The *MAM* genes with significant differences in expression between the transgenic and CK samples are indicated by asterisks (*, $P < 0.05$; **, $P < 0.01$). *TUB2* as reference (LOC103873913).



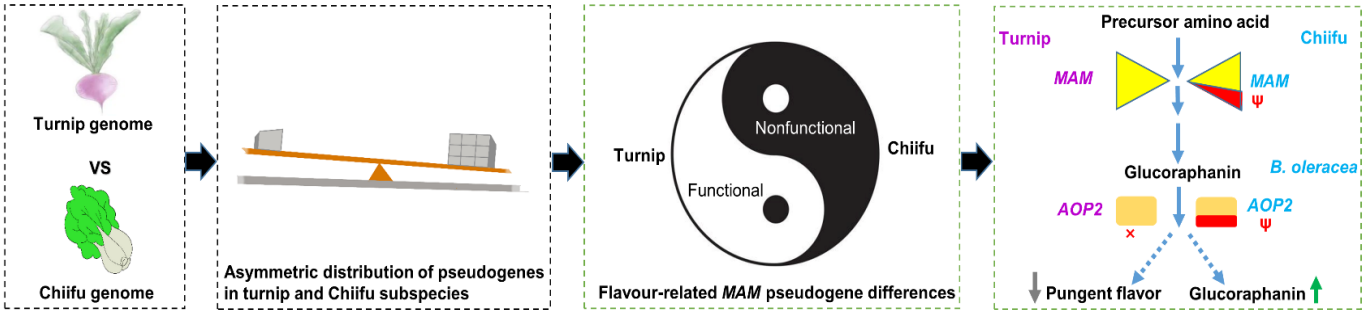
208

209 **Supplemental Figure 17. Heat map of the quantitative real-time PCR (qRT-PCR)**
210 **data for the turnip and Chiifu v3.5 *AOP* gene expression levels at different**
211 **developmental stages.** The qRT-PCR analysis was completed using three biological
212 and technical replicates. R1, R2, and R3 represent the taproot samples collected at 10,
213 20, and 30 days after germination, respectively. L1, L2, and L3 represent the leaf
214 samples collected at 10, 20, and 30 days after germination, respectively. Gene
215 expression levels are presented in the colored bar. *TUB2* as reference (LOC103873913).

216



Supplemental Figure 18. Results of the RT-PCR analysis of *AOP2* genes in turnip and Chiifu v3.5. *AOP2*-RNAi and control samples were analyzed. *Actin* as reference.



Supplemental Figure 19. Article flowchart. The pseudogenetic differences in turnip and Chiifu genomes resulted in the differential evolution of the flavor-related GSL metabolic pathway. Pseudogenes were distributed asymmetrically on the chromosomes in these two subspecies. The *MAM* gene family expanded in turnip but converted into pseudogenes in Chiifu. There are three functional *AOP2* genes both in turnip and *B. rapa*. Although *B. oleracea* also contains three *AOP2* genes, only one is functional; the other two genes are mutated. These differences explain the diversity in the flavors of turnip and Chiifu. In turnip, the accumulation of anticancer substances may be enhanced and the pungency may be decreased by the RNAi-based silencing of the flavor- and anticancer-associated *AOP2* gene in the GSL pathway. Pseudogenes are indicated by "ψ".

Supplemental tables

Supplemental Table 1. Estimation of the turnip genome size according to a k-mer analysis.

K-mer	K-mer number	K-mer depth	Genome size (Mb)
19	45,695,255,046	14	446.09

Supplemental Table 2. Sequencing data for the turnip genome assembly.

Sequencing libraries	Illumina reads	Pacbio reads	Hi-C reads
Insert size (bp)	270	20 000	270
Clean data (Gb)	59.4	44.93	26.79
Mean read length (bp)	-	8,595	-
Sequence coverage (×)	50	110	65.5

Supplemental Table 3. Reads mapped to the turnip genome assembly.

Library	Total reads	Mapped reads	Mapped (%)
270 bp	226,529,887	219,034,986	96.69

Supplemental Table 4. BUSCO results for the turnip genome assembly.

Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs
1,568 (97.20%)	1065 (66.00%)	503 (31.20%)	7 (0.40%)	39 (2.40%)

Supplemental Table 5. Hi-C sequencing data statistics.

Read Pairs Number	Base Number	GC Content（%）	% ≥Q30
89,707,160	26,796,366,884	40.46	92.39

Supplemental Table 6. Evaluation of the coverage on the basis of mapped clean reads.

Mapping Type	Number	Ratio（%）
Total Read Pairs	89,707,160	100
Mapped Reads	153,017,543	85.28
Unique Mapped Read Pairs	19,744,769	22.01

Supplemental Table 7. Chromosome length data statistics.

Group	Sequence Length (bp)
chr01	37215573
chr02	38061388
chr03	38969619
chr04	26283953
chr05	34184301
chr06	42502065
chr07	30802089
chr08	27475427
chr09	57405788
Chr10	23028969

Supplemental Table 8. Summary of the transposable elements in the turnip, ECD04, and Chiifu A03 genomes.

Species	class	Subclass	Copy Number	Length (bp)	% Genome
A03	classI	LTR/Copia	22050	20087298	5.56
		LTR/Gypsy	35435	37614783	10.41
		LTR/unkonwn	31179	13195197	3.65
	classII	CACTA	10046	5289144	1.46
		Mutator	51177	31478394	8.71
		PIF_Harbinger	9328	3677627	1.02
		Tc1_Mariner	3624	1009162	0.28
		hAT	13346	4372430	1.21
	classIII	helitron	127160	49668898	13.74
	others	-	48403	15461914	4.27
	total	-	351748	181854847	50.31
Turnip	classI	LTR/Copia	18379	14963893	4.21
		LTR/Gypsy	32017	30291604	8.51
		LTR/unkonwn	26760	15674789	4.41
	classII	CACTA	11923	6775596	1.90
		Mutator	22949	12410888	3.49
		PIF_Harbinger	11872	4237219	1.19
		Tc1_Mariner	5058	1308878	0.37
		hAT	13454	4530589	1.27
	classIII	helitron	110792	48125049	13.53
	others	-	31951	13248625	3.72
	total	-	285155	151567130	42.60
ECD04	classI	LTR/Copia	24172	21758690	6.27
		LTR/Gypsy	46055	35901048	10.35
		LTR/unkonwn	22666	14390627	4.15
	classII	CACTA	10731	5942413	1.71
		Mutator	32738	23415477	6.75
		PIF_Harbinger	9486	3689151	1.06
		Tc1_Mariner	3556	1109901	0.32
		hAT	15925	4825756	1.39
	classIII	helitron	99347	37286850	10.75
	others	-	32786	14347884	4.15
	total	-	297462	162667797	46.90

Supplemental Table 9. Functional annotations of turnip genes.

database	Annotated number	Percentage (%)
eggNOG	51069	89.86
GO	25322	44.56
KEGG_ko	23848	41.96
PFAM	34724	61.10
Swiss-Prot	38236	67.28
NR	56018	98.57
All_Annotated	56832	

Supplemental Table 10. miRNA data statistics.

Types	Number
miRNA	326
rRNA	2010
tRNA	1174

Supplemental Table 11. Statistics of the turnip Illumina paired-end reads (NGS) and turnip 48A resequencing reads mapped on the new turnip chromosomes.

NGS/resequencing reads	Data sources	Average sequencing depth	Mapping rate (%)	Coverage (%)
Turnip NGS-D01	This study	70.21	96.79	94.78
Turnip NGS-D02		54.82	96.50	94.61
ECD04 NGS	NCBI project: PRJNA672906	58.31	97.30	86.94
TUA NGS	GSA number: CRA003187 (Yang et al., 2019)	23.44	96.47	84.50
TUE NGS		76.42	96.63	86.41
B48Areseq-1		27.81	96.87	92.23
B48Areseq-2		21.44	97.41	90.27
B48Areseq-3		25.11	97.42	90.28

Supplemental Table 12. Statistics of the turnip Illumina paired-end reads (NGS) and turnip 48A resequencing reads mapped on A03 chromosomes.

NGS/resequencing reads	Accession number	Average sequencing depth	Mapping rate (%)	Coverage (%)
Turnip NGS-D01	This study	68.27	97.27	93.50
Turnip NGS-D02		53.33	97.06	93.08
B48Areseq-1	(Yang et	26.84	97.42	91.73
B48Areseq-2	al., 2019)	20.65	98.31	89.29
B48Areseq-3		23.52	98.56	89.19

Supplemental Table 13. Statistics of the turnip Illumina paired-end reads (NGS) and turnip 48A resequencing reads mapped on ECD04 chromosomes.

NGS/resequencing reads	Accession number	Average sequencing depth	Mapping rate (%)	Coverage (%)
Turnip NGS-D01	This study	88.17	97.00	93.41
Turnip NGS-D02		69.33	96.78	93.16
B48Areseq-1	(Yang et al., 2019)	34.83	97.24	92.76
B48Areseq-2		27.50	98.12	91.87
B48Areseq-3		32.59	98.42	91.90

Supplemental Table 14. Statistics of the turnip Illumina paired-end reads (NGS) and turnip 48A resequencing reads mapped on TUA chromosomes.

NGS/resequencing reads	Accession number	Average sequencing depth	Mapping rate (%)	Coverage (%)
Turnip NGS-D01	This study	64.99	97.25	94.77
Turnip NGS-D02		50.72	97.03	94.60
B48Areseq-1	(Yang et al., 2019)	25.72	97.41	94.40
B48Areseq-2		19.42	98.30	93.63
B48Areseq-3		22.21	98.56	93.11

Supplemental Table 15. Statistics of the turnip Illumina paired-end reads (NGS) and turnip 48A resequencing reads mapped on TUE chromosomes.

NGS/resequencing reads	Accession number	Average sequencing depth	Mapping rate (%)	Coverage (%)
Turnip NGS-D01	This study	67.49	96.76	94.64
Turnip NGS-D02		52.68	96.54	94.47
B48Areseq-1	(Yang et al., 2019)	26.55	96.34	93.95
B48Areseq-2		20.18	97.92	93.61
B48Areseq-3		23.14	98.19	93.10

Supplemental Table 16. Statistical analysis of the retained genes in three subgenomic blocks in Chiifu A03 and turnip.

A. thaliana	No. of A03 genes							No. of Turnip genes					
Block	No. of genes	LF	LF-R	MF1	MF1-R	MF2	MF2-R	LF	LF-R	MF1	MF1-R	MF2	MF2-R
A	1902	751	0.39	748	0.39	623	0.33	815	0.43	570	0.30	400	0.21
B	1539	582	0.38	442	0.29	428	0.28	643	0.42	477	0.31	412	0.27
C	1081	334	0.31	325	0.30	290	0.27	412	0.38	319	0.30	132	0.12
D	586	278	0.47	121	0.21	43	0.07	298	0.51	134	0.23	20	0.03
E	1610	1315	0.82	542	0.34	114	0.07	1351	0.84	562	0.35	67	0.04
F	2691	1365	0.51	985	0.37	833	0.31	1435	0.53	1012	0.38	903	0.34
G	186	19	0.10	19	0.10	15	0.08	16	0.09	14	0.08	0	0.00
H	557	251	0.45	190	0.34	79	0.14	239	0.43	185	0.33	84	0.15
I	811	331	0.41	228	0.28	91	0.11	381	0.47	109	0.13	100	0.12
J	1797	925	0.51	777	0.43	589	0.33	950	0.53	820	0.46	627	0.35
K	255	115	0.45	91	0.36	69	0.27	127	0.50	104	0.41	73	0.29
L	439	189	0.43	124	0.28	98	0.22	229	0.52	94	0.21	22	0.05
M	666	319	0.48	95	0.14	41	0.06	356	0.53	97	0.15	47	0.07
N	1240	690	0.56	468	0.38	335	0.27	692	0.56	402	0.32	324	0.26
O	518	254	0.49	162	0.31	74	0.14	261	0.50	188	0.36	70	0.14
P	357	133	0.37	88	0.25	44	0.12	135	0.38	97	0.27	55	0.15
Q	602	267	0.44	164	0.27	149	0.25	280	0.47	189	0.31	156	0.26
R	2060	1175	0.57	783	0.38	752	0.37	1166	0.57	761	0.37	745	0.36
S	731	268	0.37	103	0.14	45	0.06	331	0.45	101	0.14	38	0.05
T	415	108	0.26	97	0.23	60	0.14	123	0.30	109	0.26	84	0.20
U	2477	1445	0.58	877	0.35	657	0.27	1421	0.57	912	0.37	666	0.27
V	620	284	0.46	186	0.30	176	0.28	357	0.58	212	0.34	193	0.31
W	1180	547	0.46	432	0.37	397	0.34	536	0.45	503	0.43	394	0.33
X	734	290	0.40	249	0.34	205	0.28	360	0.49	240	0.33	236	0.32
Total	25054	12235		8296		6207		12914		8211		5848	

Note: The LF (least fractionated), MF1 (medium fractionated), and MF2 (most fractionated) subgenomes of Chiifu A03 and turnip were obtained on the basis of the homology with *A. thaliana* blocks. The retention rates were calculated as the proportion of the genes in each subgenomic block that had a corresponding gene in *A. thaliana* blocks. R, retention rate.

623 **Supplemental Table 17. *MAM* genes and pseudogenes in *Brassica* genomes**

Gene ID	Species	Chr	Start	End
AT5G23010.1	*A. thaliana*	Chr5	7702848	7707172
AJ486890	*A. thaliana*	Chr5		
AT5G23020.1	*A. thaliana*	Chr5	7718118	7721866
BAA02g29590.1	*Brassica rapa* Chiifu A03	A02	16939844	16941344
BAA02g29600.1	*Brassica rapa* Chiifu A03	A02	16948854	16949309
BAA02g44900.1	*Brassica rapa* Chiifu A03	A02	36864864	36869292
BAA03g45530.1	*Brassica rapa* Chiifu A03	A03	21790151	21792578
BAA03g45540.1	*Brassica rapa* Chiifu A03	A03	21797168	21802282
BAA03g45550.1	*Brassica rapa* Chiifu A03	A03	21822110	21824526
BAA04g25300.1	*Brassica rapa* Chiifu A03	A04	26477444	26480189
Gene0352400.1	Turnip B03	chr03	26481369	26483548
Gene0121420.1	Turnip B03	chr02	2497704	2501665
Gene0027070.1	Turnip B03	chr03	26508063	26510950
Gene0000760.1	Turnip B03	chr02	2305218	2309166
Gene0228790.1	Turnip B03	chr02	2317303	2318748
Gene0006750.1	Turnip B03	chr02	2497704	2501665
Gene0464890.1	Turnip B03	chr02	2509622	2512810
Gene0495830.1	Turnip B03	chr02	19602605	19612312
Gene0100730.1	Turnip B03	chr04	26457563	26460316
BraA02g027460.3.5C.1	*Brassica rapa* Chiifu v3.5	A02	15960475	15974715
BraA02g028300.3.5C.1	*Brassica rapa* Chiifu v3.5	A02	16660024	16669763
BraA02g044510.3.5C.1	*Brassica rapa* Chiifu v3.5	A02	29594897	29598097
BraA02g044520.3.5C.2	*Brassica rapa* Chiifu v3.5	A02	29607024	29611499
BraA03g044240.3.5C.2	*Brassica rapa* Chiifu v3.5	A03	22082460	22085246
BraA03g044250.3.5C.3	*Brassica rapa* Chiifu v3.5	A03	22088206	22094550
BraA03g044270.3.5C.1	*Brassica rapa* Chiifu v3.5	A03	22114597	22117013
BraA04g026170.3.5C.1	*Brassica rapa* Chiifu v3.5	A04	18013584	18016329
BraA02g026580.3C	*Brassica rapa* Chiifu v3.0	A02	15960475	15960927
BraA02g027350.3C	*Brassica rapa* Chiifu v3.0	A02	16669308	16669763
BraA02g042660.3C	*Brassica rapa* Chiifu v3.0	A02	29594897	29598097
BraA02g042670.3C	*Brassica rapa* Chiifu v3.0	A02	29607154	29611410
BraA03g043730.3C	*Brassica rapa* Chiifu v3.0	A03	22082638	22085062
BraA03g043740.3C	*Brassica rapa* Chiifu v3.0	A03	22089652	22094767
BraA03g043750.3C	*Brassica rapa* Chiifu v3.0	A03	22114597	22117013
BraA04g024970.3C	*Brassica rapa* Chiifu v3.0	A04	18013584	18015367
BraA04g024980.3C	*Brassica rapa* Chiifu v3.0	A04	18015386	18016329
BraA02g02783P	*Brassica rapa* chinensis	A02	17294254	17304661
BraA02g04317P	*Brassica rapa* chinensis	A02	30297471	30300018
BraA02g04318P	*Brassica rapa* chinensis	A02	30306808	30312436
BraA03g04215P	*Brassica rapa* chinensis	A03	21567279	21569915
BraA03g04216P	*Brassica rapa* chinensis	A03	21572747	21579487

BraA03g04218P	*Brassica rapa* chinensis	A03	21598090	21600506
BraA04g02442P	*Brassica rapa* chinensis	A04	18181429	18184173
A02p26760.1_BraBRO	*Brassica rapa* BraBRO	A02	15821348	15826813
A02p27710.1_BraBRO	*Brassica rapa* BraBRO	A02	16608558	16618326
A02p43500.1_BraBRO	*Brassica rapa* BraBRO	A02	29817889	29819947
A02p43510.1_BraBRO	*Brassica rapa* BraBRO	A02	29834535	29839175
A03p42990.1_BraBRO	*Brassica rapa* BraBRO	A03	21721237	21723414
A03p43000.1_BraBRO	*Brassica rapa* BraBRO	A03	21724136	21732233
A02p30350.1_BraCCA	*Brassica rapa* BraCCA	A02	17275036	17284774
A02p47340.1_BraCCA	*Brassica rapa* BraCCA	A02	29985801	29989001
A02p47350.1_BraCCA	*Brassica rapa* BraCCA	A02	29997052	30001319
A03p45340.1_BraCCA	*Brassica rapa* BraCCA	A03	21954602	21956781
A03p45350.1_BraCCA	*Brassica rapa* BraCCA	A03	21961371	21966485
A04p26440.1_BraCCA	*Brassica rapa* BraCCA	A04	17458545	17461290
A02p27150.1_BraCCB	*Brassica rapa* BraCCB	A02	16847321	16857387
A02p41380.1_BraCCB	*Brassica rapa* BraCCB	A02	28707728	28710330
A02p41400.1_BraCCB	*Brassica rapa* BraCCB	A02	28717453	28722350
A03p41780.1_BraCCB	*Brassica rapa* BraCCB	A03	21576924	21579457
A03p41790.1_BraCCB	*Brassica rapa* BraCCB	A03	21582421	21589194
A02p26630.1_BraCXA	*Brassica rapa* BraCXA	A02	17062135	17071869
A02p40810.1_BraCXA	*Brassica rapa* BraCXA	A02	30553370	30555638
A02p40830.1_BraCXA	*Brassica rapa* BraCXA	A02	30566276	30570419
A03p41640.1_BraCXA	*Brassica rapa* BraCXA	A03	21664986	21667166
A03p41650.1_BraCXA	*Brassica rapa* BraCXA	A03	21671755	21676870
A03p41670.1_BraCXA	*Brassica rapa* BraCXA	A03	21695528	21699269
A04p23710.1_BraCXA	*Brassica rapa* BraCXA	A04	18315795	18318540
A02p28080.1_BraCXB	*Brassica rapa* BraCXB	A02	16976779	16986521
A02p43820.1_BraCXB	*Brassica rapa* BraCXB	A02	30063350	30065376
A02p43840.1_BraCXB	*Brassica rapa* BraCXB	A02	30076254	30080397
A03p43520.1_BraCXB	*Brassica rapa* BraCXB	A03	21584467	21587185
A03p43530.1_BraCXB	*Brassica rapa* BraCXB	A03	21591499	21596614
A03p43550.1_BraCXB	*Brassica rapa* BraCXB	A03	21615269	21617685
A04p25560.1_BraCXB	*Brassica rapa* BraCXB	A04	19089372	19092125
BraA03t011503E	Turnip ECD04	A03	21510967	21514685
BraA03t011505E	Turnip ECD04	A03	21519966	21525728
BraA03t011506E	Turnip ECD04	A03	21537234	21548661
BraA02t031223E	Turnip ECD04	A02	15670577	15671029
BraA02t031224E	Turnip ECD04	A02	15671924	15676050
BraA02t031316E	Turnip ECD04	A02	16584022	16594081
BraA02t032945E	Turnip ECD04	A02	31121287	31124919
BraA02t032946E	Turnip ECD04	A02	31131180	31136203
A02p26480.1_BraMIZ	*Brassica rapa* BraMIZ	A02	16362670	16366821
A02p27410.1_BraMIZ	*Brassica rapa* BraMIZ	A02	17153545	17163010
A02p42480.1_BraMIZ	*Brassica rapa* BraMIZ	A02	30231396	30232788

A02p42490.1_BraMIZ	*Brassica rapa* BraMIZ	A02	30241605	30245752
A03p44270.1_BraMIZ	*Brassica rapa* BraMIZ	A03	22804269	22806447
A03p44280.1_BraMIZ	*Brassica rapa* BraMIZ	A03	22811037	22816151
A03p44300.1_BraMIZ	*Brassica rapa* BraMIZ	A03	22834805	22837587
A04p24470.1_BraMIZ	*Brassica rapa* BraMIZ	A04	18680083	18682849
A02p28310.1_BraPCA	*Brassica rapa* BraPCA	A02	16472482	16486716
A02p29280.1_BraPCA	*Brassica rapa* BraPCA	A02	17220043	17229747
A02p46270.1_BraPCA	*Brassica rapa* BraPCA	A02	31272988	31275046
A02p46280.1_BraPCA	*Brassica rapa* BraPCA	A02	31282613	31287255
A03p44840.1_BraPCA	*Brassica rapa* BraPCA	A03	22000930	22005178
A03p44850.1_BraPCA	*Brassica rapa* BraPCA	A03	22006059	22010189
A03p44870.1_BraPCA	*Brassica rapa* BraPCA	A03	22017369	22018235
A03p44880.1_BraPCA	*Brassica rapa* BraPCA	A03	22030064	22032846
A04p25200.1_BraPCA	*Brassica rapa* BraPCA	A04	17231544	17234288
A02p28860.1_BraPCB	*Brassica rapa* BraPCB	A02	17379174	17388890
A02p45940.1_BraPCB	*Brassica rapa* BraPCB	A02	31180956	31182349
A02p45950.1_BraPCB	*Brassica rapa* BraPCB	A02	31196946	31201358
A03p44060.1_BraPCB	*Brassica rapa* BraPCB	A03	21740960	21743139
A03p44070.1_BraPCB	*Brassica rapa* BraPCB	A03	21747729	21752842
A03p44090.1_BraPCB	*Brassica rapa* BraPCB	A03	21771499	21774281
A04p25290.1_BraPCB	*Brassica rapa* BraPCB	A04	17060073	17062817
A02p28450.1_BraTCA	*Brassica rapa* BraTCA	A02	16458631	16472817
A02p29290.1_BraTCA	*Brassica rapa* BraTCA	A02	17182970	17192709
A02p45390.1_BraTCA	*Brassica rapa* BraTCA	A02	30174306	30182525
A02p45400.1_BraTCA	*Brassica rapa* BraTCA	A02	30185826	30189771
A03p45640.1_BraTCA	*Brassica rapa* BraTCA	A03	22258279	22260297
A03p45650.1_BraTCA	*Brassica rapa* BraTCA	A03	22265046	22270159
A03p45670.1_BraTCA	*Brassica rapa* BraTCA	A03	22288814	22291886
A04p25730.1_BraTCA	*Brassica rapa* BraTCA	A04	18645836	18648694
A02p28020.1_BraTUA	*Brassica rapa* BraTUA	A02	16413113	16414673
A02p29100.1_BraTUA	*Brassica rapa* BraTUA	A02	17355496	17364891
A02p44250.1_BraTUA	*Brassica rapa* BraTUA	A02	28524441	28525834
A02p44380.1_BraTUA	*Brassica rapa* BraTUA	A02	28583045	28587450
A03p43900.1_BraTUA	*Brassica rapa* BraTUA	A03	21650016	21652036
A03p43910.1_BraTUA	*Brassica rapa* BraTUA	A03	21655250	21661978
A04p24620.1_BraTUA	*Brassica rapa* BraTUA	A04	16803042	16807920
A02p27510.1_BraTUE	*Brassica rapa* BraTUE	A02	16044120	16049605
A02p28510.1_BraTUE	*Brassica rapa* BraTUE	A02	16897813	16907210
A02p44980.1_BraTUE	*Brassica rapa* BraTUE	A02	29906621	29909939
A02p45000.1_BraTUE	*Brassica rapa* BraTUE	A02	29918379	29923559
A03p43290.1_BraTUE	*Brassica rapa* BraTUE	A03	21400608	21402780
A03p43300.1_BraTUE	*Brassica rapa* BraTUE	A03	21408678	21414281
A03p43310.1_BraTUE	*Brassica rapa* BraTUE	A03	21430639	21433428
BraA02t07413Z	*Brassica rapa* BrapaZ1	A02	15881597	15884186

BraA02t07492Z	*Brassica rapa* BrapaZ1	A02	16577334	16587104
BraA02t09063Z	*Brassica rapa* BrapaZ1	A02	29723188	29727132
BraA03t13773Z	*Brassica rapa* BrapaZ1	A03	21573766	21576003
BraA03t13775Z	*Brassica rapa* BrapaZ1	A03	21579649	21584763
A03Pse1_A02-36849703-36855070	*Brassica rapa* A03	A02	36849703	36855070
A03Pse2_A02-36848777-36854980	*Brassica rapa* A03	A02	36848777	36854980
B03Pse_chr02-19603547-19612312	Turnip B03	chr02	19603547	19612312
Brapa3.5Pse_A02-29594897-29598097	*Brassica rapa* Chiifu v3.5	A02	29594897	29598097
BrasPse_A02-30297367-30300169	*Brassica rapa* chinensis	A02	30297367	30300169
Z1Pse1_A02-29711994-29714276	*Brassica rapa* BrapaZ1	A02	29711994	29714276
Z1Pse2_A02-29712920-29714366	*Brassica rapa* BrapaZ1	A02	29712920	29714366
CCAPse_A02-29998086-29999523	*Brassica rapa* BraCCA	A02	29998086	29999523
CCBPse_A02-28706225-28707370	*Brassica rapa* BraCCB	A02	28706225	28707370
CXAPse_A02-30553266-30554561	*Brassica rapa* BraCXA	A02	30553266	30554561
CXBPse_A02-30063246-30064540	*Brassica rapa* BraCXB	A02	30063246	30064540
ECD04Pse_A02-31123336-31124417	Turnip ECD04	A02	31123336	31124417
MIZPse_A02-30232218-30233663	*Brassica rapa* BraMIZ	A02	30232218	30233663
PCAPse_A02-31272395-31273689	*Brassica rapa* BraPCA	A02	31272395	31273689
PCBPse_A02-31181778-31187146	*Brassica rapa* BraPCA	A02	31181778	31187146
TCAPse_A02-30173713-30175007	*Brassica rapa* BraTCA	A02	30173713	30175007
TUAPse_A02-28525263-28530631	*Brassica rapa* BraTUA	A02	28525263	28530631
TUEPse_A02-29906028-29908309	*Brassica rapa* BraTUE	A02	29906028	29908309

624

625

626 **Supplemental Table 18. *AOP* genes in turnip**

	A. thaliana	Turnip
AOP1	AT4G03070.1	Gene0096740.1
		Gene0493910.1
		Gene0445090.1
AOP2	AT4G03060.1	Gene0250680.1
		Gene0405960.1
		Gene0486840.1

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641 **Supplemental Table 19. Primers used in this study**

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
The primers for real-time quantitative PCR (qRT-PCR) analysis.		
Gene0495830	CTATCCGCACAAGAAGCCAAC	GACGGGCTGGAGGTCAATAC
Gene0352400	CCTTTCACTCTCCCGGTCTG	ACGGAGCGTCGTGTCAAATA
Gene0121420	ATGATCTCCATCTCCCCCACT	TAGCACCAGTCCCAGCCTTA
Gene0100730	CTTCTGGCATCTCCCGCAAT	GTACTTCGGCCATCGTTCCA
Gene0027070	ATGTCCCACAATGATCCCCAC	CGATTCAGGCGTAGAGAGGG
Gene0000760	CCCTTACTCCACCGCAGAAG	CACGGTTTTGGCGATGGTTT
Gene0006750	GGAGCATTTGTGATGGGTGG	CAAATGCCTGAATTTGAGATTT
Gene0228790	AAGTGGAAATGCACCACTTGA	TTACCGCCTTACCATCTTGCT
Gene0464890	TAAGGCGGCTTGGGAATCAG	ATGTTGATCCCCACCGTGTC
Gene0096740	GACAAAGTGTCAGTGGAGC	CATGAGGACGAGGTGGTGTATA
Gene0493910	CAAAGTGGGATGAAGTGAAG	CTCTGAGAACTTGTGCATCGTC
Gene0445090	AGTGGGATAAAGTGAAGGCTG	GCGTCGAGAACTTGTGCATCAT
Gene0250680	GAGGAGTGATGTCCGTAAAGC	CATCCAATTCTGCTAACTTCT
Gene0405960	GTCCGTAAAGCTCTTGAAGAC	CATTGACTTGAGGTTCTCATCA
Gene0486840	ACAAGAGTACCAGCGAAAGG	CGCCAGCACCAACATCCGCACC
BraA03g029130	GGGACAAAGTGAAGACTGATG	ACACCACCTCGTCCTCATGAAG
BraA09g001350	CTCTCTGAATCTCTTGAGCTCC	TGAGACGATGCACAAATTCGC
BraA09g001340	ATAAAGTGAAGGCTGATGTCC	GATGATGCACAAGTTCTCGACG
BraA09g001360	CGTAAAGCTCTTGAAGACTAC	GCTGATGCTGATGATATTGCT
BraA02g028320	TAAAGCAATCGGTTTTGGAAG	CCCGCTGCAACATTAGTATCA
BraA02g029140	AAGATGATTAGCGAGCCGG	CGCGTCGATTCCCATCTCTAG
BraA03g044240	CCTTTCACTCTCCCGGTCTG	ACGGAGCGTCGTGTCAAATA
BraA03g044250	ATGATCTCCATCTCCCCCACT	TAGCACCAGTCCCAGCCTTA
BraA03g044270	CTTCTGGCATCTCCCGCAAT	GTACTTCGGCCATCGTTCCA
BraA02g044520	CCCTTACTCCACCGCAGAAG	CACGGTTTTGGCGATGGTTT
BraA02g044510	AAAGTGGAAATGCACCACTTGA	TTACCGCCTTACCATCTTGCT

BraA02g027460	TCTCCAGGTGGAGCCCTTAC	CGCATATTACCGGGACGTATC
BraA02g028300	GCCGCAACTAGTAGTATTGAC	GAGTTGCCTAGCAATCTCTAC
BraA04g026170	ACATTCACATGAAATATAAG	ATGCCCACCGTGGTTGCACC
TUB2	AGGCGTGTGAGTGAGCAGTT	CATCTCGTCCATTCCTTCACCTGT

The primers for RT-PCR analysis.

BAA02g29590	TACCGCCAACACAATCTCCG	GGCCCCAATGTCTTCTGGTG
BAA02g29600	ACCTATCCGCACAAGAAGCC	CAACGGTCTTGGCAATGCTT
Gene0495830	TACGTCCCGGTAATATGCGTC	GCTCCACGACATTTCAAAGC
Gene0405960	GCTGATGCCAATGCTAATAC	TGAATGGTTTGAAGACTCGT
Gene0250680	ACTGGTGATTGTGCTAATGT	TAAGCGTGAAGAGTAGAACG
Gene0486840	AACCCTAAAACCAGGAAGTG	CCCAAAGCTCTCAAATACCA
BraA02g028320	AGTATCAGCGAAACAATCCA	ACCTCCAAACCTTCAATCTC
BraA09g001360	AACACCTCAAGTCAACGAAT	CCTCCAAACCTTCAATCTCA
BraA03g029140	CGTCTACGACTGATGAAGTA	CAGCGATAACAACGAAAGTG
Actin	CCTGGTCAGCTTAACTCCGAC	CATCTCGTCCATTCCTTCACCTGT

642