

Differentially expressed genes and phenotypic variation in florets among the isogenic Ogura-CMS, DGMS and inbred lines of broccoli

Zhansheng Li (✉ lizhansheng@caas.cn)

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers <https://orcid.org/0000-0003-0132-2741>

Yumei Liu

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Zhiyuan Fang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Limei Yang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Mu Zhuang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Yangyong Zhang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Honghao Lv

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Yong Wang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Fengqing Han

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Jialei Ji

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Research article

Keywords: Broccoli, Cytoplasmic male sterility, Genic male sterility, Gene expression, Florets

Posted Date: May 26th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-26810/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Male sterility contributes strongly to hybrid seed purity and production in *Brassica* crops. To detect plant phenotypes and the gene expression patterns involved in bioprocess of an Ogura cytoplasmic male sterility line (Ogura-CMS), a dominant genic male sterility line (DGMS) and their maintainer line, we analyzed the transcriptomes of broccoli florets among the isogenic Ogura-CMS (T54C), DGMS (T54M) and inbred lines (T54S).

Results: There were respectively 505, 585 and 469 upregulated genes, and 1109, 1073 and 543 downregulated genes in the comparison groups of T54C and T54M, T54C and T54S, and T54M and T54S. The head weight and head width showed stronger performance in the Ogura-CMS than the DGMS line or maintainer line. The Ogura-CMS line showed poorer performance in seed yield and seed germination than the DGMS line or maintainer line. The DGMS line had longer maturation and flowering periods than the Ogura-CMS and maintainer lines. The plant hormone genes, auxin (*TIR1*, *GH3* and *SAUR*) and salicylic acid (SA) (*NPR1*), and adenosine triphosphate-binding cassette (ABC) transporters-related genes (*ABCB1* and *ATM*) were highly up-regulated in T54C compared with T54M or T54S. The brassinosteroid upregulated gene *CYCD3* related to the function of late-flowering and delaying senescence in plants, was detected in T54M over two seasons.

Conclusions: Among the isogenic Ogura-CMS, DGMS and inbred lines of broccoli, the Ogura-CMS line showed strong performance in head yield than the DGMS line or the inbred line. The DGMS line had longer days to flowering than the Ogura-CMS and inbred lines. However, the inbred line presented a higher seed yield and seed germination rate than the DGMS or Ogura-CMS line. This study found that some potential plant hormone genes, auxin (*TIR1*, *GH3* and *SAUR*) and SA (*NPR1*), and ABC transporters related genes (*ABCB1* and *ATM*) might play a key role in regulation of the developmental trait in the Ogura-CMS line. The BR- upregulated gene *CYCD3* may function in late-flowering and delaying senescence of broccoli consistent with the investigations conducted over two seasons. Our findings provided a possible explanation for physiological and developmental differences of broccoli among the Ogura-CMS, the DGMS and the inbred lines.

Background

Broccoli (*Brassica oleracea* L. var. *italica*) is a well-known vegetable grown worldwide that belongs to the cruciferous vegetable family, which includes cabbage, kale, cauliflower, Brussels sprouts, Chinese cabbage, rutabaga, and turnips. Broccoli contains many nutrients and few calories, exhibiting anticancer, weight loss promoting, and cardiovascular and cerebrovascular disease-preventing properties that are beneficial to human health [1–5]. In 2017, the area in China under cultivation for broccoli has exceeded 70 000 ha based on data from the Ministry of Agriculture [6].

Currently, commercial broccoli crops primarily consist of hybrid cultivars benefiting from heterosis, and the hybrid cultivars are mostly male sterility lines, especially Ogura cytoplasmic male sterility (Ogura-CMS) lines. There are usually two types of hybrid cultivars: in the first, both parents in the cross are from inbred lines, and at least one of the parents has self-incompatibility; in the second, one of the parents is transferred from a male sterility line. The male sterility lines currently used in broccoli breeding were derived from a line with dominant genic male sterility (DGMS), which was discovered in cabbage, and from a line Ogura-CMS, which was discovered in radish [7–10]. However, there are still some disadvantages regarding bud development and seed production in male sterility lines, especially in Ogura-CMS plants; these disadvantages include bud abortion and low seed yield [7, 11, 12]. On the other hand, hybrid cultivars from male sterility lines have enhanced seed purity and hybrid rates, and DGMS lines usually have higher seed yields than Ogura-CMS line from the same inbred line [11, 13]. The Ogura-CMS lines of broccoli and other *Brassica* plants used today have been improved by protoplast fusion [11], and Ogura-CMS is stable and easily transferred between species; therefore, these lines are important sources of CMS in *Brassica* plants [14–16].

Transcriptome analyses based on RNA sequencing (RNA-Seq) technology are important methods with which to understand the differentially expressed genes (DEGs) involved in various biological processes, and these analyses have been widely performed on *Brassica* plants [12, 13, 17, 18]. For example, small RNA-Seq has revealed DEGs in the early development of broccoli pollen [19], and transcriptome analyses have been used to examine four transcription factors in normal and abortive buds of an Ogura-CMS line and its maintainer line [12]. In addition, the roles of pigment mechanisms in post-harvest broccoli yellowing [20], gene expression patterns associated with sulforaphane metabolism in broccoli florets [21], glucosinolate metabolism in seeds and sprouts [22, 23], and yield heterosis in curds of broccoli have also been elucidated by RNA-Seq technology [24]. To the best of our knowledge, it would be the first report of comparative transcriptome analyses of isogenic Ogura-CMS, DGMS and inbred lines in broccoli florets.

Our group first discovered a plant with DGMS, from the 79–399 spring cabbage in 1979 [7]. To obtain a new male sterile line of broccoli for cultivar breeding, we used multiple broccoli inbred lines with cabbage DGMS material and developed several good DGMS broccoli lines with stable agronomic performance. In the same way, Ogura-CMS broccoli lines derived from CMSR3629 (Ogura-CMS) introduced by the Asgrow Seed Co. (USA), have also been developed. Among these materials, the inbred line T54 (T54S) has selfed more than 25 generations. T54S was used to develop an isogenic Ogura-CMS line (T54C) and an isogenic DGMS line (T54M). In fact, there are some big differences of flowering time, seed yield, and other agronomic traits among isogenic Ogura-CMS, DGMS and inbred lines, which needs further research in transcription level for revealing the molecular mechanism. Therefore, these materials are ideal for studying the relationships among molecular mechanisms and important agronomic traits in the Ogura-CMS and DGMS lines and their maintainer line.

Results

Morphological analysis

There were no significant differences in head characteristics among the Ogura CMS line T54C, the DGMS line T54M and their maintainer line T54S (the inbred line), but there were some other differences in maturity days, flowering time, production and seed characters, which are described in detail in Table 1. Our analysis of the agronomic traits of broccoli at harvest revealed that the Ogura-CMS broccoli line showed stronger performance in head weight and head width than the DGMS line or the maintainer line (Table 1). The DGMS line had longer production (days to maturation and flowering) than the Ogura-CMS and inbred lines. In contrast, the inbred line presented a higher seed yield and seed germination rate than the DGMS and Ogura-CMS lines. The Ogura-CMS line showed the lowest seed yield and germination rate.

Transcriptome analysis

Six cDNA libraries from broccoli florets obtained at harvest times from the T54C, T54M and T54S lines were subjected to Illumina sequencing with two biological replicates for each sample. After filtering of invalid reads, 159,424,568 clean reads and 47,827,370,400 clean bases (47.8 Gb) were obtained (Table 2). The Q20 and Q30 percentages were 98.12–98.41% and 95.06–95.62%, respectively, with GC percentages of 45.08–46.80% (Table 2).

Genomic Characteristics and-Nucleotide Polymorphism (SNP) Distribution

As shown in Table 2, the average amounts of clean bases data for T54C, T54M and T54S were 8.77 Gb, 7.91 Gb and 7.25 Gb, respectively. The average GC percentages of T54C, T54M and T54S were 46.77%, 45.82% and 46.45%, respectively. There were no obvious differences in numbers of mapped reads. In addition, the alternative splicing pattern distributions of the samples were shown in Fig. 1. The T54C vs T54M comparison presented more alternative splicing of five patterns than the T54C vs T54S and T54M vs T54S comparisons, and the T54C vs T54S and T54M vs T54S comparisons showed similar alternative splicing patterns with regard to alternative 3' splice sites (A3SSs), alternative 5' splice sites (A5SSs), mutual exclusive exons (MEXs), mutual intron retention (IR) and exon skipping (ES).

As shown in Table 3, we found that eight regions of the genome contained SNP. The T54S line showed a higher percentage of SNP distribution in exonic regions (54.39%) than the T54C (52.35%) and T54M (52.79%) lines but a lower percentage in intergenic regions (15.85%) than the T54C (16.96%) and T54M (17.25%) lines. The remaining regions of SNP distribution showed no clear differences.

Gene optimization and correlation test

A total of 5754 optimized regions involving 3872 unigenes were predicted using the gene structure optimization method (Fig. 2; Table S1). Approximately 13 mitochondrial unigenes (chrMT) were optimized (Table S1).

To study the reliability the results and differences among the samples, correlation tests were performed for each sample comparison based on the \log_2 (FPKM) values of gene expression, as shown in Table 4. All the coefficients of correlation for the pairs of samples ranged from 0.913 to 0.973, with the Pearson's R^2 test values ranging from 0.867 to 0.973 ($R^2 > 0.8$). The coefficients of correlation among replicate T54C, T54M and T54S samples were 0.965 ($R^2 = 0.943$), 0.960 ($R^2 = 0.925$) and 0.966 ($R^2 = 0.919$), respectively. The coefficients of correlation between T54C and T54M, T54C and T54S, and T54M and T54S ranged from 0.867 to 0.918, 0.867 to 0.905, and 0.894 to 0.973, respectively. Therefore, the inbred line and the DGMS line are more closely related than the inbred line and the Ogura-CMS line.

Differential genes expression analysis and DEG clustering

To explore the DEGs, the gene expression variations were analyzed in three comparisons T54C vs T54M, T54C vs T54S and T54M vs T54S (Fig. 3A, B). In the T54C line compared with the T54M line, 505 and 1109 genes were up- and downregulated, respectively. In the T54C line compared with the T54S line, 585 and 1073 genes were up- and downregulated, respectively. In the T54M line compared with the T54S line, 469 and 534 genes were up- and downregulated, respectively. The DEG cluster analysis (Fig. 4) revealed that the T54C line showed similar up- and downregulated genes in comparisons with the T54M and T54S lines, but the T54M vs T54S comparison presented very different numbers and types, as shown in Fig. 3B.

Analysis of the Ogura-CMS, DGMS and inbred lines

To better understand the function and biological processes associated with the DEGs, GO term ($p < 0.01$) and KEGG pathway ($p < 0.01$) analyses were performed for the DEGs between T54M and T54S, T54C and T54S, and T54C and T54M. The GO analysis classified the DEGs from these comparisons into 60 GO categories were separated into three main groups: the biological process (BP) group, the molecular function (MF) group and the cellular component (CC) group (Fig. 5). With regard to the T54M vs T54S comparison, GO analysis revealed that 332 and 331 upregulated genes corresponded to the intracellular and intracellular parts terms, respectively, both belonging to the CC group. However, other genes associated with terms in the CC group, including those associated with membrane (302), cell periphery (225), plasma membrane (182) and cellular region (104) terms, as well as 228 genes in the BP group were downregulated in the T54M line (Fig. 5A). With regard to the DEGs between T54C and T54S, GO analysis revealed that upregulated genes were associated with the response to stimulus (389), response to stress (271), or response to chemical (255) terms in the BP group, or with the nucleus (419) terms in the CC group. The organism metabolic process (270), extracellular region part (421), intrinsic component of membrane (300), and cell periphery (232) terms were associated with downregulated genes (Fig. 5B). In the T54C vs T54M comparison, more upregulated genes were associated with the cell (423) and cell part (423) terms in the CC group than with terms in the BP and MF groups, and the extracellular region (122) term in the CC group was associated with more downregulated genes than other terms (Fig. 5C).

The KEGG pathway results were comprehensively evaluated and filtered ($p < 0.01$), and there were eight terms strongly related to plant development. Specifically, photosynthesis-antenna proteins (7 upregulated genes), plant hormone signal transduction (41 upregulated genes), phenylpropanoid biosynthesis (39 downregulated genes), adenosine triphosphate (ATP)-binding cassette (ABC) transporters (4 up- and 5 downregulated genes), fatty acid elongation (5 upregulated genes), fatty acid metabolism (25 downregulated genes), flavonoid biosynthesis (15 downregulated genes) and polycyclic aromatic hydrocarbon degradation (8 upregulated genes) terms were distributed among different samples (Table 5). But there needs further validation of these genes in broccoli.

Identification of functional modules in PPI networks

The Protein-Protein Interactions Network (PPI-Net) was developed in 2011 [25, 26]. To better understand the interactions of the DEGs, PPI networks were constructed using Cytoscape software. As shown in Fig. 6, unlike the T54M vs T54 comparison, which showed a radiating PPI network, the T54C vs T54S comparison and T54C vs T54M comparison showed branching PPI network. These networks accurately reflected the different backgrounds of the Ogura-CMS, DGMS and inbred lines of broccoli. The results also suggested the existence of a close relationship between the DGMS and inbred lines of broccoli but a distant relationship between both of these lines and the Ogura-CMS line. In the PPI network for T54C and T54S, the most highly upregulated genes were *LOC106311788* (auxin-responsive protein IAA3-like) and *LOC106339384* (E3 ubiquitin protein ligase DRIP2), and the most highly downregulated genes were *LOC106343730* (auxin-responsive protein IAA7) and *LOC106331616* (auxin-responsive protein IAA7-like). In the PPI network for T54C and T54M, the most highly upregulated genes were *LOC106312313* (uncharacterized protein *At5g05190*), *LOC106327001* (auxin response factor 7) and *LOC106317959* (auxin response factor 7-like), and the most highly downregulated gene was *LOC106304816* (SKP1-like protein 14). In the PPI network for T54M and T54S, the most highly upregulated gene was *LOC106298512*, and the clusters of downregulated genes with higher degrees were *LOC106322852* (protein TRANSPARENT TESTA 12-like), *LOC106325277* (ribonuclease 1), *LOC106295628* (protein TRANSPARENT TESTA 12-like) and *LOC106340694* (protein ORGAN SIZE RELATED 1).

Special DEG selection and qRT-PCR validation

To better understand and explain the different agronomic traits of the broccoli lines T54S, T54M and T54C during developmental stages in the spring and autumn of 2016, fourteen DEGs were selected and their expression was validated by qRT-PCR. These genes were associated with photosynthesis-antenna proteins, plant hormone signal transduction, phenylpropanoid biosynthesis, ABC transporters,

fatty acid elongation, fatty acid metabolism, flavonoid biosynthesis and polycyclic aromatic hydrocarbon degradation. One additional gene associated with each term was targeted, and the transcriptome data and qRT-PCR results showed similar patterns for the fourteen DEGs (Figure S2).

Discussion

Male sterility lines, including Ogura-CMS and DGMS (especially Ogura-CMS), have been widely used in Brassica crops, such as broccoli, cabbage, cauliflower, Chinese cabbage, rape, and radish [7, 12, 27]. The Ogura-CMS line contributes significantly to hybrid seed production in *Brassica* crops; therefore, some studies focus on Ogura-CMS, although the mechanism of hybrid seed production has not been elucidated. Previous transcriptome and proteome analyses of the Ogura-CMS line have provided basic information about the DEGs and differential abundant proteins (DAPs) between Ogura-CMS lines and their maintainer lines [12, 13, 17, 18, 28]. In this study, some significant differences in agronomic traits associated with growth and reproductive growth and development were noted among the Ogura-CMS and DGMS lines and their maintainer line in a two-year investigation with statistical analysis. In contrast to previous studies, this study ought to elucidate the reasons for the stronger performance in head weight and head width in the Ogura-CMS line than in the DGMS line of broccoli and sought to determine why the DGMS line of broccoli had longer maturation and flowering periods than the Ogura-CMS and inbred lines. The Ogura-CMS line of broccoli showed poorer performance with regard to seed yield and germination rate than the maintainer line; this poorer performance was related to bud abortion during the bud development stage. According to transcriptome analysis and qRT-PCR verification, some DEGs and pathways might be associated with these significant differences in agronomic traits among the Ogura-CMS, DGMS and maintainer lines.

Plant hormone signal transduction and ABC transporters function in the Ogura-CMS line

Plant hormones play important roles in plant development in addition to increasing ATP supply. In this study, the plant hormone signal transduction and ABC transporters pathways (ko01075 and ko02010, respectively) were both significantly enriched for DEGs in the Ogura-CMS line compared to the inbred line and DGMS line, respectively. In the plant hormone signal transduction pathway, 26 of 548 genes were upregulated in the Ogura-CMS line and were associated with terms including auxin, cytokinin (CK), abscisic acid (ABA), ethylene (ET), jasmonic acid (JA) and salicylic acid (SA) metabolism terms; these genes are most likely related to reproductive growth. The promotion of these factors associated with plant growth might help to explain the growth trait advantages and reproductive development disadvantages in the Ogura-CMS line compared with its maintainer line.

With regard to auxin metabolism, the *TIR1*, *GH3* and *SAUR* genes were all upregulated, which could promote cell enlargement during plant growth promotion. The *SCF (TIR1) E3* ubiquitin ligase complex is involved in the auxin-mediated signaling pathway that regulates root and hypocotyl growth, lateral root formation, cell elongation, and gravitropism. In the PPI network for T54C and T54S, the expression of the genes *LOC106311788* (auxin-responsive protein IAA3-like) and *LOC106339384* (E3 ubiquitin protein ligase DRIP2) were elevated, which consistent with the previous results. The auxin-responsive *GH3* gene family is essential for plant growth and development in rice, and several genes of this family have been confirmed to have indoleacetic acid (IAA) amido synthetase biochemical activity [29, 30]. The *SAUR* gene family was initially defined as a set of auxin-inducible genes regulating development, especially in hypocotyls, and many *SAUR* genes act by promoting cell expansion [31, 32]. With regard to CK metabolism, only the *Arabidopsis* response regulators (*A-ARR*) gene was upregulated, which primarily regulates cell division in vivo in *Arabidopsis* [33]. With regard to ABA metabolism, the protein phosphatase 2C (*PP2C*) gene is present in plants and animals; plants express SAUR-immune *PP2C*, *D2* or *PP2C*. *D5* derivatives confer severe cell expansion defects and correspondingly cause constitutively low levels of PM H⁺-ATPase phosphorylation. It has been proven that PM H⁺-ATPase plays an important role in seed germination due to its role in supplying cell walls with the H⁺-ions necessary for cell elongation commencement in embryos, but the mechanism of PM H⁺-ATPase activation in germinating seeds has not been determined [34, 35]. With regard to ET metabolism, high expression of the *ERF1/2* gene is involved in fruit ripening, and overexpression of this gene has been found to be specifically induced by glucose during germination and early seedling development in *Arabidopsis* plants [36]. With regard to JA metabolism, *JAZ* and *MYC2* were both upregulated. The *JAZ* repressor family proteins are jasmonate coreceptors and transcriptional repressors that operate in the JA signaling pathway and are vital in regulating the development of plants. The *JAZ* gene family has been reported to exist in many plants [37, 38]; *MYC2*, a basic helix-loop-helix (bHLH)-type stress-related transcription factor involved in JA signaling, interacts with *ANAC019* to coregulate the expression of *NYE1* during JA-induced chlorophyll degradation in *Arabidopsis* [39]. In addition, JA promotes degreening via *MYC2/3/4*-mediated regulation of major chlorophyll catabolic genes, which has been reported in *Arabidopsis* [40, 41]. The *NPR1* gene associated with SA metabolism, was upregulated; this gene is a key regulator of the SA-mediated systemic acquired resistance (*SAR*) pathway and

plays an essential role in plant immunity [42]. In *Arabidopsis*, *NPR1* can interact with transcription factors to induce the expression of pathogenesis-related (*PR*) genes and strongly promote defense responses [43].

ABC transporters compose a transport system superfamily that is one of the largest and possibly one of the oldest gene families. This superfamily is essential for eukaryotes, especially for plants [44]. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20 and White). In this study, two genes *ABCB1* (*MDR1*) and *ATM*, both belonging to the ABCB gene family, were upregulated in the Ogura-CMS line compared with the DGMS line. ABCB1 is one of ABC genes are present in all kingdoms of life and are responsible for cellular homeostasis [45].

In summary, our findings might indicate that upregulation of the auxin-related (*TIR1*, *GH3* and *SAUR*), SA-related (*NPR1*) and ABC transporter (*ABCB1* and *ATM*) genes in the plant hormone signal transduction pathway (ko04075) in Ogura-CMS line compared with the DGMS line provides adequate and additional factors to promote the growth of Ogura-CMS broccoli, resulting in strong performance of head weight and head width. However, there needs further research of these genes in broccoli developmental regulation.

The plant hormone signal transduction and photosynthesis-antenna proteins responding to the DGMS line

In the plant hormone signal transduction pathway (ko04075), 15 of 548 total genes were upregulated in the DGMS T54M line compared with its maintainer line (T54S), two genes auxin genes (*Aux/IAA* and *SAUR*), one CK-related gene (*A-ARR*), one ABA-related gene (*PP2C*), one ET-related gene (*ERF1/2*), one brassinosteroid (BR) -related gene (*CYCD3*) and one JA-related gene (*JAZ*). Compared to the Ogura-CMS line, the DGMS line exhibited three differences in broccoli florets. With regard to auxin metabolism, two genes, *TIR1* and *GH3*, were absent; another gene, *Aux/IAA*, was present. This short-lived nuclear protein plays a crucial role in repressing the expression levels of genes activated by auxin response factors (ARFs) in *Arabidopsis* [46]. Biochemical and genetic evidence suggests that conserved domain II, found in most *Aux/IAA* proteins, is required for rapid degradation and for auxin-mediated acceleration of degradation [47]. The *CYCD3* gene associated with BR metabolism was upregulated in this study. Notably, *Arabidopsis* plants overexpressing *CYCD3-1* show extensive leaf curling, disorganized meristems, increased leaf numbers, late flowering and delayed senescence [48, 49]. One J-related gene (*JAZ*) was present in the DGMS line, while two JA-related genes (*JAZ* and *MYC2*) were present in the Ogura-CMS line; this finding could provide an additional explanation for the better seed yields in the DGMS line, as the absence of *MYC2* may have reduced degreening of florets (buds). No SA gene upregulation was found in the DGMS line. Therefore, we found that the BR gene *CYCD3* may function in late flowering and delayed senescence of broccoli during the growth and developmental stages consistent with the data obtained in the investigations conducted over two seasons.

In addition, we found that three upregulated genes (*Lhcb1*, *Lhcb3* and *Lhcb5*) were associated with the photosynthesis-antenna protein pathway (ko00196). The antenna proteins that exist in phycobilisomes in cyanobacteria and light-harvesting complex II (LHCII) protein complexes in green plants act as peripheral antenna systems, enabling more efficiency of light energy absorption. LHCII trimers are composed of *Lhcb1*, *Lhcb2*, and *Lhcb3* proteins in various trimeric configurations. *Lhcb4*, *Lhcb5*, and *Lhcb6* are monomeric proteins associated with photosystem II (PSII), while *Lhca* polypeptides make up the antenna of photosystem I (PSI). According to analyses of the functions and expressions of LHCII and PSII genes in previous studies, we hypothesize *Lhcb1*, *Lhcb3* and *Lhcb5* genes in the DGMS line may respond to the environment [50-53].

Conclusions

This study pointed out the Ogura-CMS broccoli line showed strong performance in head weight and head width than the DGMS line or the maintainer line. The DGMS line had longer production (days to maturation and flowering) than the Ogura-CMS and maintainer (inbred) lines. In contrast, the inbred line presented a higher seed yield and seed germination rate than the DGMS or Ogura-CMS line; the Ogura-CMS line showed the lowest seed yield and germination rate. According to RNA sequencing (RNA-Seq) and qRT-PCR verification, plant hormone genes, auxin (*TIR1*, *GH3* and *SAUR*) and SA (*NPR1*), and ABC transporters-related genes (*ABCB1* and *ATM*) might be important factors in the Ogura-CMS line. The BR- upregulated gene *CYCD3* may function in late-flowering and delaying senescence of broccoli consistent with the investigations conducted over two seasons. Our findings provided a possible explanation for physiological and developmental differences of broccoli among the Ogura-CMS, the DGMS and the inbred lines.

Methods

Plant materials and agronomic trait investigation

The broccoli (*Brassica oleracea* var. *italica*) maintainer T54S (the inbred line), Ogura-CMS T54C and the DGMS line T54M were all bred at the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences (IVF-CAAS). The original source of inbred line T54S was derived from the Sakata commercial cultivar (Japan). The broccoli line T54C is derived from the popular Ogura-CMS line (Sakata, Japan) [54, 55], and T54M was initially derived from the cabbage line 79-399-3 (our lab) [56, 57]. All the plant materials were cultivated and improved by our lab (IVF-CAAS). The T54S line produces an early-maturing, uniform, tight-headed, semidomed broccoli with large buds on a medium-framed plant. In the spring and autumn of 2016, three replicates were successively grown in the same experimental greenhouse at IVF-CAAS (Beijing, China). When plants harvest began, three types of bud samples were typically collected from T54S, T54M and T54C and labeled with two biological replicates (Fig. 7). The isolated buds were immediately frozen in liquid nitrogen and stored at -80°C for late use.

The agronomic traits of three lines were investigated during the harvest, flowering, seed ripening and seed germination stages. All the data were analyzed with SPSS 12.0 software (IBM Co., Ltd., New York, USA), and one-way ANOVA was carried out on the agronomic trait data at $p < 0.05$.

RNA extraction and Illumina sequencing

Total RNA was extracted using an RNA Pure Plant Kit (Tiangen Co., Ltd., Beijing, China) following the manufacturer's instructions. The integrity of the RNA was assessed with an RNA Nano 6000 Assay Kit on an Agilent Bioanalyzer 2100 platform (Agilent Technologies Inc., Santa Clara, CA, USA). High-quality RNA from each sample was used for cDNA library construction and RNA-Seq on an Illumina HiSeq™ 2500 platform (Berry Genomics Co., Beijing, China). To obtain clean high-quality reads, adapter sequences, low-quality reads (in which > 50% bases had Q-values ≤ 3) and unknown sequences ("N" reads > 10% unknown nucleotides) were removed from the raw reads. The Q20, Q30 and GC contents were then calculated based on the clean reads.

Quality control and transcriptome analysis

The clean reads from each sample were mapped to the reference genome (ftp://brassicadb.org/Brassica_oleracea/) with TopHat2 (Version 2.0.3.12) [58]. The gene expression levels were normalized using the fragments per kilobase of transcript per million mapped reads (FPKM) method [45]. The edgeR package (<http://www.rproject.org/>) was used to identify differentially expressed genes (DEGs) between two samples. Each gene was quantified by calculating the FPKM value with RSEM (V1.2.15). Thresholds of an FDR < 0.05 and a \log_2 |fold change| > 1 were set to select the DEGs. The DEGs were then analyzed for enrichment of Gene Ontology (GO; <http://www.geneontology.org/GO.downloads.shtml>) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/pathway.html>) pathways [21, 59]. GO terms or pathways with p values < 0.05 were defined as significantly enriched for the DEGs. Correlations among the expression levels of DEGs were analyzed with Pearson correlation tests. A protein-protein interaction (PPI) network was constructed with Cytoscape software with combined information from databases such as the Biomolecular Interaction Network Database (BIND) and an interactome database [60]. The reference genome was for *Brassica oleracea* (wild cabbage) (<https://www.ncbi.nlm.nih.gov/genome/10901>).

Quantitative qRT-PCR analysis

Quantitative real-time (qRT-PCR) analyses were performed to validate the results from the DEGs with the three biological replicates. This is the results of the study, fourteen DEGs were chosen and tested; six of them were specifically upregulated, and eight of them were specifically downregulated in T54M vs T54S samples, T54C vs T54M samples and T54C vs T54S samples. Total RNA was extracted from the head buds of each sample at harvest time using an RNA Pure Plant Kit (Tiangen Co., Ltd., Beijing, China). First-strand cDNA was synthesized using a PrimeScript™ 1st Strand cDNA Synthesis Kit (Takara, Dalian, China). qRT-PCR was carried out according to the SYBR PrimeScript RT-PCR Kit manufacturer specifications (Takara, Dalian, China) on an ABI Prism® 7900HT Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Three biological replicates (with three technical replicates for each biological replicate) were analyzed for each gene, and the relative expression level was estimated by the $2^{-\Delta\Delta C_t}$ method [61, 62]. The *β -actin* gene was used as an internal control, and the specific primers were designed using Primer Premier 6 (Premier Co., Canada) (Table 6) [61].

Abbreviations

CORE: Collaborative Oat Research Enterprise

LOD: Logarithm of odds

Declarations

Acknowledgments

We thank the academic English editors (AJE) (available online: <https://www.aje.com/services/editing/>) for editing this manuscript.

Funding

This work was supported by grants from the National Key Research and Development Program of China (2017YFD0101805), the National Natural Science Foundation of China (31501761), the China Agriculture Research System (CARS-23-08A), the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-IVFCAAS), and the State Key Laboratory of Vegetable Germplasm Innovation (SKL-VGI). The funding organizations played no role in the design of study and collection, analysis, and interpretation of data, or the decision to submit the work for publication.

Contributions

ZL Conceived and designed experiments. YL Analysed the data. ZF, LY, MZ, YZ Conducted quantitative analysis. ZL Performed experiments. HL, YW, FH, JJ Carried out the investigation and data analysis. All authors have read and approved the manuscript.

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare there were no conflicts of interest.

Supplementary information

Supplementary Fig. S1: Verification of the selected DEGs was carried out by qRT-PCR.

References

1. Appendino G, Bardelli A: Broccoli, PTEN deletion and prostate cancer: where is the link? *Molecular cancer* 2010, 9.
2. Kensler TW, Fahey JW, Egner PA, Groopman JD, Hecht SS, Cornblatt BS, Ye L, Dinkova-Kostova A, Visvanathan K, Talalay P: Cancer chemoprevention with glucosinolate-rich broccoli sprouts. *Toxicol Pathol* 2006, 34(7):998-998.
3. Armah CN, Traka MH, Dainty JR, Doleman JF, Potter JF, Mithen RF: The effect of a high glucoraphanin broccoli diet on cardiovascular risk profile: a randomised controlled study. *P Nutr Soc* 2012, 71(Oce2):E62-E62.
4. Myzak MC, Dashwood RH: Chemoprotection by sulforaphane: Keep one eye beyond Keap1. *Cancer letters* 2006, 233(2):208-218.
5. Li ZS, Liu YM, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: Natural Sulforaphane From Broccoli Seeds Against Influenza A Virus Replication in MDCK Cells. *Natural product communications* 2019, 14(6).
6. Li ZS, Mei YJ, Liu YM, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: The evolution of genetic diversity of broccoli cultivars in China since 1980. *Sci Hortic-Amsterdam* 2019, 250:69-80.
7. Fang ZY, Sun PT, Liu YM, Yang LM, Wang XW, Hou AF, Bian CS: A male sterile line with dominant gene (Ms) in cabbage (*Brassica oleracea* var. *capitata*) and its utilization for hybrid seed production. *Euphytica* 1997, 97(3):265-268.
8. Kaminski P, Podwyszynska M, Starzycki M, Starzycka-Korbas E: Interspecific hybridisation of cytoplasmic male-sterile rapeseed with *Ogura* cytoplasm and *Brassica rapa* var. *pekinensis* as a method to obtain male-sterile Chinese cabbage inbred lines. *Euphytica* 2016, 208(3):519-534.

9. Kirti PB, Banga SS, Prakash S, Chopra VL: Transfer of Ogu Cytoplasmic Male-Sterility to Brassica-Juncea and Improvement of the Male-Sterile Line through Somatic-Cell Fusion. *Theor Appl Genet* 1995, 91(3):517-521.
10. Shu JS, Liu YM, Li ZS, Zhang LL, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: A generic SSR marker closely linked to a dominant genic male sterility gene (DGMS79-399-3) in broccoli (*Brassica oleracea* var. *italica*). *Mol Breeding* 2016, 36(7).
11. Chamola R, Balyan HS, Bhat SR: Transfer of cytoplasmic male sterility from alloplasmic Brassica juncea and B. napus to cauliflower (*B. oleracea* var. *botrytis*) through interspecific hybridization and embryo culture. *Indian J Genet PI Br* 2013, 73(2):203-210.
12. Shu JS, Zhang LL, Liu YM, Li ZS, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: Normal and Abortive Buds Transcriptomic Profiling of Broccoli ogu Cytoplasmic Male Sterile Line and Its Maintainer. *International journal of molecular sciences* 2018, 19(9).
13. Han FQ, Zhang XL, Yang LM, Zhuang M, Zhang YY, Li ZS, Fang ZY, Lv HH: iTRAQ-Based Proteomic Analysis of Ogura-CMS Cabbage and Its Maintainer Line. *International journal of molecular sciences* 2018, 19(10).
14. Wang QB, Zhang YY, Fang ZY, Liu YM, Yang LM, Zhuang M: Chloroplast and mitochondrial SSR help to distinguish allo-cytoplasmic male sterile types in cabbage (*Brassica oleracea* L. var. *capitata*). *Mol Breeding* 2012, 30(2):709-716.
15. Shu JS, Liu YM, Li ZS, Zhang LL, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: Organelle Simple Sequence Repeat Markers Help to Distinguish Carpeloid Stamen and Normal Cytoplasmic Male Sterile Sources in Broccoli. *PLoS one* 2015, 10(9).
16. Shu JS, Liu YM, Li ZS, Zhang LL, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: Detection of the Diversity of Cytoplasmic Male Sterility Sources in Broccoli (*Brassica Oleracea* var. *Italica*) Using Mitochondrial Markers. *Front Plant Sci* 2016, 7.
17. Wei XC, Zhang XH, Yao QJ, Yuan YX, Li XX, Wei F, Zhao YY, Zhang Q, Wang ZY, Jiang WS *et al*: The miRNAs and their regulatory networks responsible for pollen abortion in Ogura-CMS Chinese cabbage revealed by high-throughput sequencing of miRNAs, degradomes, and transcriptomes. *Front Plant Sci* 2015, 6.
18. Xing MM, Sun C, Li HL, Hu SL, Lei L, Kang JG: Integrated analysis of transcriptome and proteome changes related to the Ogura cytoplasmic male sterility in cabbage. *PLoS one* 2018, 13(3).
19. Li H, Wang Y, Wu M, Li LH, Jin C, Zhang QL, Chen CB, Song WQ, Wang CG: Small RNA Sequencing Reveals Differential miRNA Expression in the Early Development of Broccoli (*Brassica oleracea* var. *italica*) Pollen. *Front Plant Sci* 2017, 8.
20. Luo F, Cai JH, Kong XM, Zhou Q, Zhou X, Zhao YB, Ji SJ: Transcriptome profiling reveals the roles of pigment mechanisms in postharvest broccoli yellowing. *Hortic Res-England* 2019, 6.
21. Li ZS, Liu YM, Li LY, Fang ZY, Yang LM, Zhuang M, Zhang YY, Lv HH: Transcriptome reveals the gene expression patterns of sulforaphane metabolism in broccoli florets. *PLoS one* 2019, 14(3).
22. Gao J, Yu X, Ma F, Li J: RNA-Seq Analysis of Transcriptome and Glucosinolate Metabolism in Seeds and Sprouts of Broccoli (vol 9, e88804, 2014). *PLoS one* 2014, 9(4).
23. Ku KM, Becker TM, Juvik JA: Transcriptome and Metabolome Analyses of Glucosinolates in Two Broccoli Cultivars Following Jasmonate Treatment for the Induction of Glucosinolate Defense to *Trichoplusia ni* (Hubner). *International journal of molecular sciences* 2016, 17(7).
24. Li H, Yuan JY, Wu M, Han ZP, Li LH, Jiang HM, Jia YL, Han X, Liu M, Sun DL *et al*: Transcriptome and DNA methylome reveal insights into yield heterosis in the curds of broccoli (*Brassica oleracea* L var. *italica*). *Bmc Plant Biol* 2018, 18.
25. Lim D, Kim NK, Park HS, Lee SH, Cho YM, Oh SJ, Kim TH, Kim H: Identification of Candidate Genes related to Bovine Marbling using Protein-Protein Interaction Networks. *Int J Biol Sci* 2011, 7(7):992-1002.
26. Panga V, Raghunathan S: A cytokine protein-protein interaction network for identifying key molecules in rheumatoid arthritis. *PLoS one* 2018, 13(6).
27. Wu JY, Shen JR, Mao XZ, Liu KD, Wei LP, Liu PW, Yang GS: Isolation and analysis of differentially expressed genes in dominant genic male sterility (DGMS) Brassica napus L. using subtractive PCR and cDNA microarray. *Plant Sci* 2007, 172(2):204-211.
28. Wang S, Wang C, Zhang XX, Chen X, Liu JJ, Jia XF, Jia SQ: Transcriptome de novo assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in cabbage. *Plant Physiol Bioch* 2016, 105:224-232.
29. Du H, Wu N, Fu J, Wang SP, Li XH, Xiao JH, Xiong LZ: A GH3 family member, OsGH3-2, modulates auxin and abscisic acid levels and differentially affects drought and cold tolerance in rice. *J Exp Bot* 2012, 63(18):6467-6480.
30. Yang JH, Han SJ, Yoon EK, Lee WS: 'Evidence of an auxin signal pathway, microRNA167-ARF8-GH3, and its response to exogenous auxin in cultured rice cells'. *Nucleic acids research* 2006, 34(6):1892-1899.
31. Xu NF, Hagen G, Guilfoyle T: Multiple auxin response modules in the soybean SAUR 15A promoter. *Plant Sci* 1997, 126(2):193-201.

32. Spartz AK, Ren H, Park MY, Grandt KN, Lee SH, Murphy AS, Sussman MR, Overvoorde PJ, Gray WM: SAUR Inhibition of PP2C-D Phosphatases Activates Plasma Membrane H⁺-ATPases to Promote Cell Expansion in Arabidopsis. *Plant Cell* 2014, 26(5):2129-2142.
33. D'Agostino IB, Deruere J, Kieber JJ: Characterization of the response of the arabidopsis response regulator gene family to cytokinin. *Plant physiology* 2000, 124(4):1706-1717.
34. Ren H, Park MY, Spartz AK, Wong JH, Gray WM: A subset of plasma membrane-localized PP2C.D phosphatases negatively regulate SAUR-mediated cell expansion in Arabidopsis. *Plos Genet* 2018, 14(6).
35. Obroucheva NV, Sin'kevich IA, Lityagina SV, Novikova GV: Activation of acid growth in germinating horse chestnut seeds. *Russ J Plant Physiol* 2013, 60(3):437-441.
36. Zhou XJ, Cooke P, Li L: Eukaryotic release factor 1-2 affects Arabidopsis responses to glucose and phytohormones during germination and early seedling development. *J Exp Bot* 2010, 61(2):357-367.
37. Wang YK, Qiao LY, Bai JF, Wang P, Duan WJ, Yuan SH, Yuan GL, Zhang FT, Zhang LP, Zhao CP: Genome-wide characterization of JASMONATE-ZIM DOMAIN transcription repressors in wheat (*Triticum aestivum* L.). *Bmc Genomics* 2017, 18.
38. Garrido-Bigotes A, Figueroa NE, Figueroa PM, Figueroa CR: Jasmonate signalling pathway in strawberry: Genome-wide identification, molecular characterization and expression of JAZs and MYCs during fruit development and ripening (vol 13, e0197118, 2018). *PLoS one* 2018, 13(10).
39. Hong GJ, Xue XY, Mao YB, Wang LJ, Chen XY: Arabidopsis MYC2 Interacts with DELLA Proteins in Regulating Sesquiterpene Synthase Gene Expression. *Plant Cell* 2012, 24(6):2635-2648.
40. Zhu XY, Chen JY, Xie ZK, Gao J, Ren GD, Gao S, Zhou X, Kuai BK: Jasmonic acid promotes degreening via MYC2/3/4-and ANAC019/055/072-mediated regulation of major chlorophyll catabolic genes. *Plant J* 2015, 84(3):597-610.
41. Dombrecht B, Xue GP, Sprague SJ, Kirkegaard JA, Ross JJ, Reid JB, Fitt GP, Sewelam N, Schenk PM, Manners JM *et al*: MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. *Plant Cell* 2007, 19(7):2225-2245.
42. Nic-Matos G, Narvaez M, Peraza-Echeverria S, Saenz L, Oropeza C: Molecular cloning of two novel NPR1 homologue genes in coconut palm and analysis of their expression in response to the plant defense hormone salicylic acid. *Genes Genom* 2017, 39(9):1007-1019.
43. Chen J, Mohan R, Zhang Y, Li M, Chen H, Palmer IA, Chang M, Qi G, Spoel SH, Mengiste T *et al*: NPR1 Promotes Its Own and Target Gene Expression in Plant Defense by Recruiting CDK8. *Plant physiology* 2019, 181(1):289-304.
44. Kushnir S, Babiychuk E, Storozhenko S, Davey MW, Papenbrock J, De Rycke R, Engler G, Stephan UW, Lange H, Kispal G *et al*: A mutation of the mitochondrial ABC transporter Sta1 leads to dwarfism and chlorosis in the Arabidopsis mutant starik. *Plant Cell* 2001, 13(1):89-100.
45. Hodges LM, Markova SM, Chinn LW, Gow JM, Kroetz DL, Klein TE, Altman RB: Very important pharmacogene summary: ABCB1 (MDR1, P-glycoprotein). *Pharmacogenet Genom* 2011, 21(3):152-161.
46. Muto H, Nagao I, Demura T, Fukuda H, Kinjo M, Yamamoto KT: Fluorescence cross-correlation analyses of the molecular interaction between an Aux/IAA protein, MSG2/IAA19, and protein-protein interaction domains of auxin response factors of Arabidopsis expressed in HeLa cells. *Plant Cell Physiol* 2006, 47(8):1095-1101.
47. Dreher KA, Brown J, Saw RE, Callis J: The Arabidopsis Aux/IAA protein family has diversified in degradation and auxin responsiveness. *Plant Cell* 2006, 18(3):699-714.
48. Dewitte W, Riou-Khamlichi C, Scofield S, Healy JMS, Jacquard A, Kilby NJ, Murray JAH: Altered cell cycle distribution, hyperplasia, and inhibited differentiation in arabidopsis caused by the D-type cyclin CYCD3. *Plant Cell* 2003, 15(1):79-92.
49. Menges M, Samland AK, Planchais S, Murray JAH: The D-type cyclin CYCD3;1 is limiting for the G1-to-S-phase transition in Arabidopsis. *Plant Cell* 2006, 18(4):893-906.
50. Pietrzykowska M, Suorsa M, Semchonok DA, Tikkanen M, Boekema EJ, Aro EM, Jansson S: The Light-Harvesting Chlorophyll a/b Binding Proteins Lhcb1 and Lhcb2 Play Complementary Roles during State Transitions in Arabidopsis. *Plant Cell* 2014, 26(9):3646-3660.
51. Zhang L, Melo TB, Li H, Naqvi KR, Yang CH: The inter-monomer interface of the major light-harvesting chlorophyll a/b complexes of photosystem II (LHCII) influences the chlorophyll triplet distribution. *J Plant Physiol* 2014, 171(5):42-48.
52. Kumar S, Fladung M: Gene stability in transgenic aspen (*Populus*). II. Molecular characterization of variable expression of transgene in wild and hybrid aspen. *Planta* 2001, 213(5):731-740.

53. Minagawa J, Han KC, Dohmae N, Takio K, Inoue Y: Molecular characterization and gene expression of lhcb5 gene encoding CP26 in the light-harvesting complex II of *Chlamydomonas reinhardtii*. *Plant molecular biology* 2001, 46(3):277-287.
54. Dong X, Kim WK, Lim YP, Kim YK, Hur Y: Ogura-CMS in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) causes delayed expression of many nuclear genes. *Plant Sci* 2013, 199-200:7-17.
55. Yu HL, Fang ZY, Liu YM, Yang LM, Zhuang M, Lv HH, Li ZS, Han FQ, Liu XP, Zhang YY: Development of a novel allele-specific Rfo marker and creation of Ogura CMS fertility-restored interspecific hybrids in *Brassica oleracea*. *Theor Appl Genet* 2016, 129(8):1625-1637.
56. Chen C, Zhuang M, Fang ZY, Wang QB, Zhang YY, Liu YM, Yang LM, Cheng F: A Co-Dominant Marker BoE332 Applied to Marker-Assisted Selection of Homozygous Male-Sterile Plants in Cabbage (*Brassica oleracea* var. *capitata* L.). *J Integr Agr* 2013, 12(4):596-602.
57. Han FQ, Zhang XL, Yuan KW, Fang ZY, Yang LM, Zhuang M, Zhang YY, Wang Y, Liu YM, Li ZS *et al*: A user-friendly KASP molecular marker developed for the DGMS-based breeding system in *Brassica oleracea* species. *Mol Breeding* 2019, 39(6).
58. Kim D, Perteu G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013, 14(4).
59. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research* 2008, 36(10):3420-3435.
60. Brysbaert G, Lorgouilloux K, Vranken WF, Lensink MF: RINspector: a Cytoscape app for centrality analyses and DynaMine flexibility prediction. *Bioinformatics* 2018, 34(2):294-296.
61. Li ZS, Liu YM, Fang ZY, Yang LM, Zhuang M, Zhang YY, Zhao W, Sun PT: Variation of Sulforaphane Levels in Broccoli (*Brassica Oleracea* Var. *Italica*) during Flower Development and the Role of Gene *Aop2*. *J Liq Chromatogr R T* 2014, 37(9):1199-1211.
62. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 2001, 25(4):402-408.

Tables

Table 1 Agronomic trait investigation of the broccoli lines T54S, T54M and T54C during developmental stages in the spring and autumn of 2016

Trait Name	Autumn			Spring		
	T54S	T54M	T54C	T54S	T54M	T54C
Single head weight (kg) *	0.24 ± 0.04	0.23 ± 0.07	0.26 ± 0.04	0.22 ± 0.05	0.21 ± 0.08	0.24 ± 0.05
Head height (cm)	15.24 ± 1.16	15.26 ± 2.06	15.36 ± 2.18	15.11 ± 1.89	14.84 ± 1.78	15.26 ± 2.24
Head width (cm) *	15.37 ± 1.55	15.22 ± 2.08	16.04 ± 2.57	15.08 ± 1.87	14.39 ± 2.06	15.53 ± 3.15
Flower color	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow
Days to maturity (day) *	105 ± 3	109 ± 4	106 ± 4	126 ± 3	134 ± 4	128 ± 4
Days to flowering (day) *	126 ± 3	129 ± 3	126 ± 4	143 ± 2	147 ± 3	142 ± 2
Seeds yield per plant (g) *	4.13 ± 0.24	3.52 ± 0.58	2.41 ± 0.27	2.86 ± 0.29	2.21 ± 0.56	1.74 ± 0.36
Seeds germination rate (%) *	86.58 ± 3.77	83.04 ± 4.18	71.76 ± 2.47	85.03 ± 2.61	79.59 ± 2.92	71.46 ± 2.88

Note: T54C, T54M and T54S represent samples from the Ogura-CMS, DGMS and maintainer lines, respectively. All the data are shown as the mean ± standard deviation (n=3). "Days to maturity" represents the number of days from sowing to head maturity, and "days to flowering" represents the number of days from sowing to flowering. The asterisk (*) shows that a significant difference was identified among the T54S, T54M and T54C groups in the same season based on one-way ANOVA at $p < 0.05$

Table 2 Summary of the broccoli bud transcriptome sequencing data

Sample Name	Clean Reads	Clean Bases (Gb)	Clean N (%)	Clean GC (%)	Clean Q20 (%)	Clean Q30 (%)	Mapped Reads (Mapped/Clean)
T54C-1	30962081	9.29	0.00	46.74	98.34	95.47	72.70%
T54C-2	27469179	8.24	0.00	46.80	98.2	95.11	72.30%
T54M-1	28243036	8.47	0.00	46.56	98.34	95.48	73.00%
T54M-2	24472981	7.34	0.00	45.08	98.12	95.06	71.60%
T54S-1	25153030	7.55	0.00	46.27	98.33	95.43	72.60%
T54S-2	23124261	6.94	0.00	46.63	98.41	95.62	72.90%

Note: T54C, T54M and T54S represent the bud samples of the Ogura-CMS, DGMS and maintainer lines, respectively

Table 3 SNP distribution in genomic regions as determined by broccoli bud transcriptome sequencing data

Sample Name	Exonic	Splicing Site	5' UTR	3' UTR	Intronic	Upstream	Downstream	Intergenic
T54C-1	52.66%	0.40%	6.63%	9.82%	7.60%	2.32%	3.71%	16.86%
T54C-2	52.92%	0.40%	6.60%	9.61%	7.72%	2.25%	3.45%	17.05%
T54M-1	53.29%	0.40%	6.32%	10.10%	7.11%	2.24%	3.59%	16.95%
T54M-2	51.41%	0.35%	5.52%	10.89%	7.52%	2.44%	4.33%	17.55%
T54S-1	53.57%	0.40%	6.52%	10.18%	7.66%	2.12%	3.42%	16.13%
T54S-2	55.21%	0.41%	6.66%	9.98%	7.01%	2.06%	3.10%	15.56%

Note: T54C, T54M and T54S represent the bud samples of the Ogura-CMS, DGMS and maintainer lines, respectively

Table 4 Correlation test of gene expression levels derived from broccoli bud transcriptome sequencing data

Serial Number	Sample Name	Linear Equation	Pearson Test (R ²)
1	T54C-1 vs T54C-2_FPKM	$Y=0.965^*X + 0.025$	0.943
2	T54C-1 vs T54M-1_FPKM	$Y=0.939^*X - 0.053$	0.918
3	T54C-1 vs T54M-2_FPKM	$Y=0.915^*X + 0.046$	0.875
4	T54C-1 vs T54S-1_FPKM	$Y=0.913^*X + 0.177$	0.867
5	T54C-1 vs T54S-2_FPKM	$Y=0.937^*X + 0.150$	0.890
6	T54C-2 vs T54M-1_FPKM	$Y=0.945^*X - 0.073$	0.918
7	T54C-2 vs T54M-2_FPKM	$Y=0.917^*X + 0.028$	0.867
8	T54C-2 vs T54S-1_FPKM	$Y=0.918^*X + 0.158$	0.867
9	T54C-2 vs T54S-2_FPKM	$Y=0.946^*X + 0.130$	0.905
10	T54M-1 vs T54M-2_FPKM	$Y=0.960^*X + 0.099$	0.925
11	T54M-1 vs T54S-1_FPKM	$Y=0.947^*X + 0.231$	0.896
12	T54M-1 vs T54S-2_FPKM	$Y=0.973^*X + 0.205$	0.973
13	T54M-2 vs T54S-1_FPKM	$Y=0.947^*X + 0.141$	0.894
14	T54M-2 vs T54S-2_FPKM	$Y=0.967^*X + 0.013$	0.917
15	T54S-1 vs T54S-2_FPKM	$Y=0.966^*X - 0.013$	0.919

Note: T54C, T54M and T54S represent the bud samples of the Ogura-CMS, DGMS and maintainer lines, respectively. Y and X represent the values of \log_2 (FPKM) between two samples, and * indicates correlation is highly significant between two samples at $p < 0.05$

Table 5 KEGG pathways enriched for the DEGs

Term	Sample	ID	Input No	Background No	p-Value	Corrected p-Value	Up/Down
Photosynthesis -antenna proteins	T54M vs T54S	ko00196	7	35	0.00000	0.00004	up
Plant hormone signal transduction	T54M vs T54S	ko04075	15	548	0.00185	0.05459	up
	T54C vs T54S	ko04075	26	548	0.00000	0.00009	up
Phenylpropanoid biosynthesis	T54M vs T54S	ko00940	14	272	0.00025	0.02413	down
	T54C vs T54S	ko00940	25	272	0.00000	0.00019	down
ABC transporters	T54M vs T54S	ko02010	5	33	0.00036	0.02413	down
	T54C vs T54M	ko02010	4	33	0.00145	0.06517	up
Fatty acid elongation	T54C vs T54M	ko00062	5	41	0.00036	0.04808	up
Fatty acid metabolism	T54C vs T54M	ko01212	12	127	0.00036	0.03038	down
	T54C vs T54S	ko01212	13	127	0.00015	0.00754	down
Flavonoid biosynthesis	T54C vs T54M	ko00941	7	45	0.00045	0.03038	down
	T54C vs T54S	ko00941	8	45	0.00010	0.00680	down
Polycyclic aromatic hydrocarbon degradation	T54C vs T54S	ko00624	8	110	0.00072	0.01574	up

Table 6 Genes targeted by the specific primers for qRT-PCR analysis

Gene	Function Description	Chromosome	Forward Primer 5'-3'	Reverse Primer 5'-3'	Ta (°C)
<i>LOC106318096</i>	chlorophyll a-b binding protein 13, chloroplastic	C9	TCTCCACCAAACCAGCAAAG	TCAAGACCGTTGATGCGGAA	59
<i>LOC106313340</i>	auxin-responsive protein IAA4	C9	CTTGGGTTTTTCGGGGGAAGA	CCGTTTAAGCCTCTCACTGGT	60
<i>LOC106316002</i>	ethylene receptor 2 (<i>BO-ETR2</i>)	C1	TGGGTTTAAGCATTTTCATGGGA	GAAAGGGTGGGGACCGTAAG	60
<i>LOC106296937</i>	cinnamoyl-CoA reductase 2-like	C6	AACGGAGCCAAGTTCGTGAT	CGAAAAGAAGAGGGCATTTCAGC	60
<i>LOC106299021</i>	cytochrome P450 98A8-like	C6	TGGTGGGCATCCAACATACC	TTGTCGCTAACGAGCCACTT	60
<i>LOC106307625</i>	ABC transporter B family member 2	C1	GGAGGCGTCCAGTGATCC	TCCAACGAGTACTTGGCGAC	60
<i>LOC106305741</i>	ABC transporter B family member 19	C7	AGATCATCCGGGGAGGTGAA	TCATCAAGCATCAAATCATAAGCGT	60
<i>LOC106312287</i>	3-ketoacyl-CoA synthase 1 (KCS)	C8	GCCGTCTCTATCGGCAATGA	CACTCCACATATCCGGTCCA	60
<i>LOC106304303</i>	palmitoyl-protein thioesterase 1	C7	TTCGGGTTTTACCCGGATGG	TACGCAAAGGCTCCTTGGTT	60
<i>LOC106325611</i>	3-oxoacyl-[acyl-carrier-protein] synthase II, chloroplastic-like	C2	TGGCTGCCTCTTCCTGTTAC	CCGGAGTTAGTTGCTCGGTT	59
<i>LOC106328481</i>	chalcone synthase 3-like	C3	CCAAGCTCCTTGGTCTTCGT	ACGATGCTGGGAAGTATGTGT	59
<i>LOC106298325</i>	chalcone-flavonone isomerase-like	C6	TTGGCTACCTACCTCCTCCC	TCAAACGAACCCGTGACGAT	60
<i>LOC106306241</i>	cytochrome P450 81D11	C7	ACGAATCTGCGAAGGTGGAG	TGCCTGCATCTGCGGAATAA	61
<i>LOC106298990</i>	cytochrome P450 71B5	C6	CCGACCTAAGACGGTAGGGA	CCGGAGATCCGGTCAATGAG	60
<i>β-actin</i>	beta-actin		ATCTGGCATCACACTTTCTAC	ATCTCTTTGCTCATACGGTCT	55

Figures

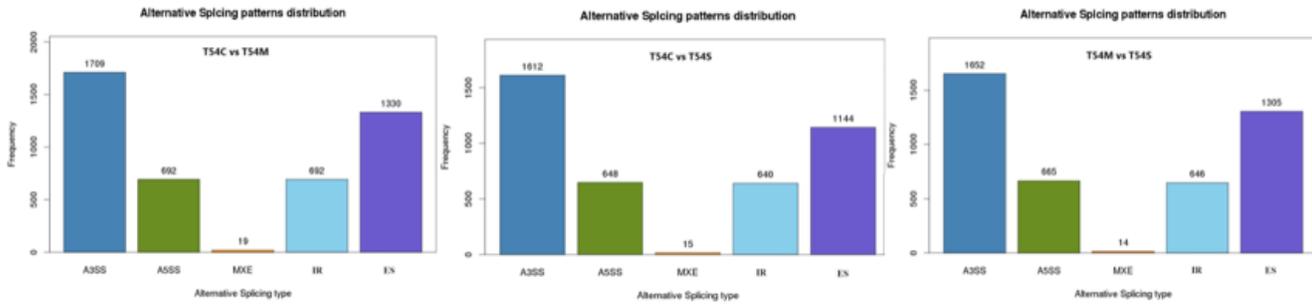


Figure 1

Alternative splicing pattern distributions of samples for A3SSs, A5SSs, MXEs, mutual IR and ES

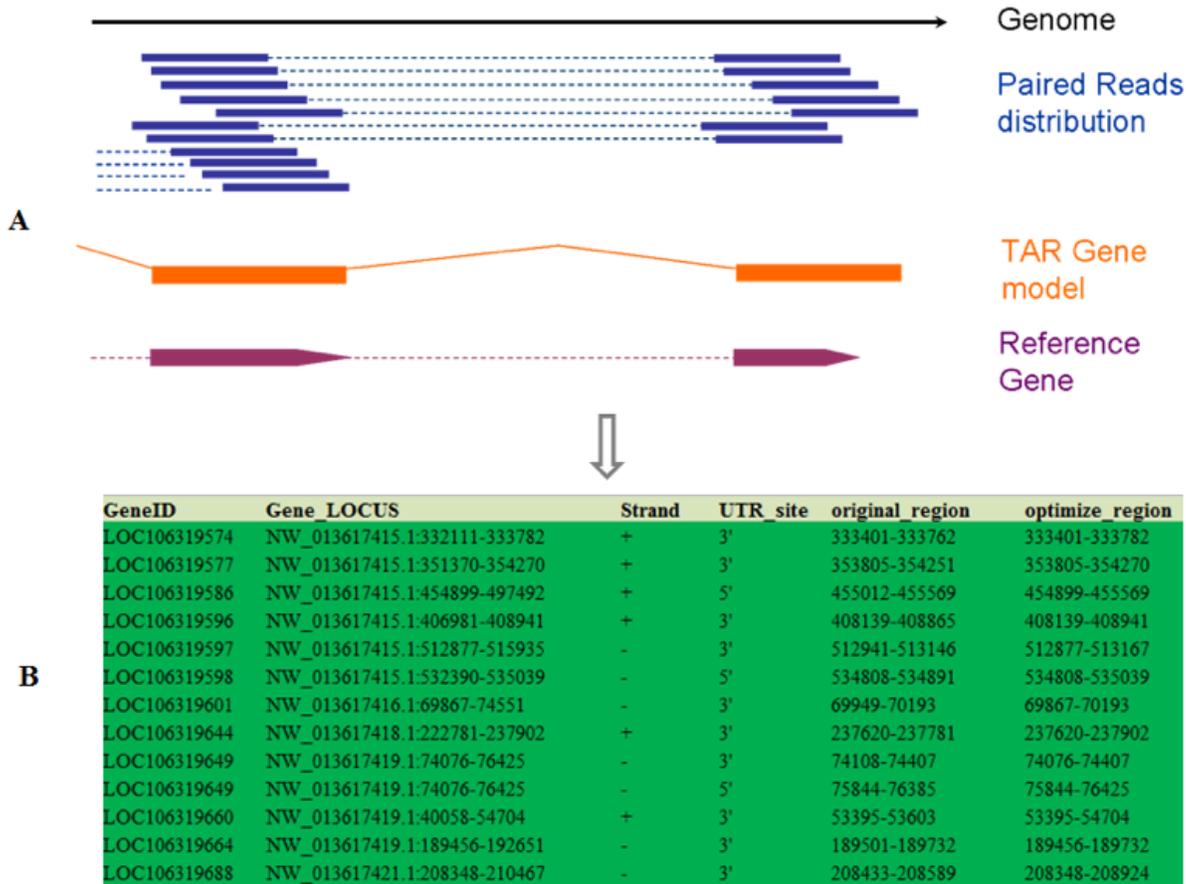


Figure 2

Summary of the process of gene optimization in the genome. The continuous and overlapping reads mapped to the genome were connected to form transcript active regions (TARs) marked in orange; the cutoff for coverage of the unique reads was > 2. Then, the different TARs were connected by paired-end reads (marked with orange lines) to form a potential gene model. Finally, the gene model was compared with reference gene regions marked purple; which enabled the 5' and 3' ends of the genes to be extended (A). The table in green shows some of the information on the optimized genes (B)

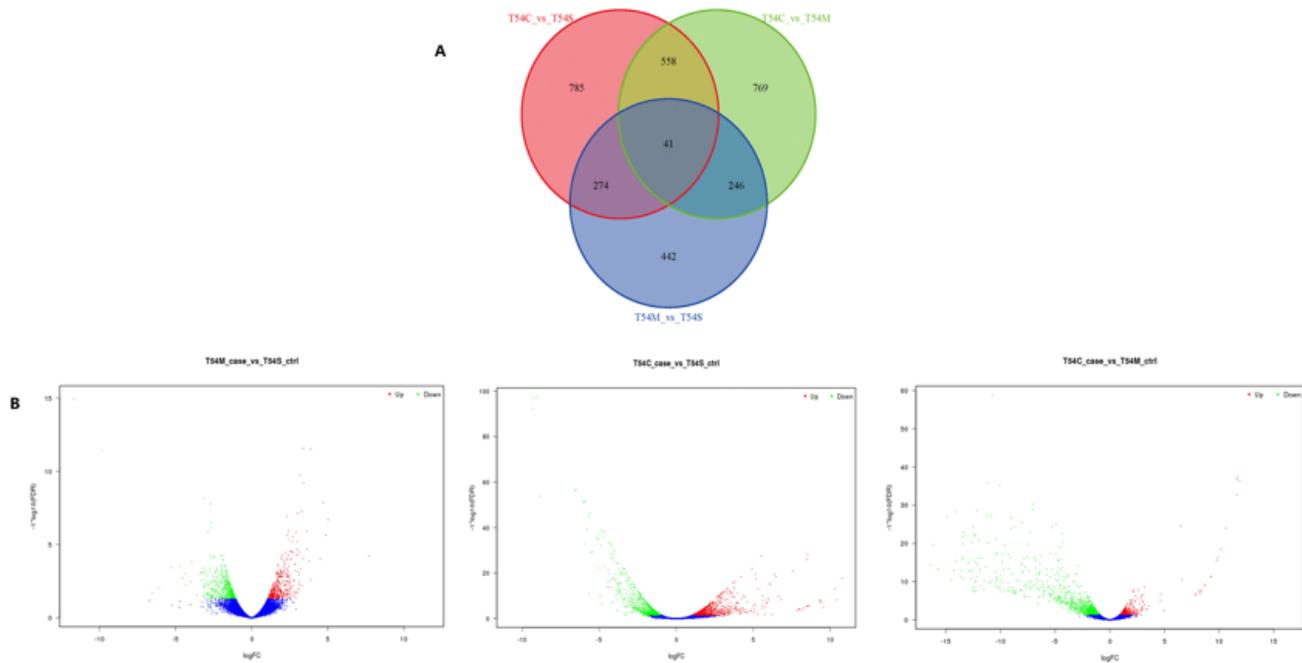


Figure 3

Summary of the process of gene optimization in the genome. The continuous and overlapping reads mapped to the genome were connected to form transcript active regions (TARs) marked in orange; the cutoff for coverage of the unique reads was > 2 . Then, the different TARs were connected by paired-end reads (marked with orange lines) to form a potential gene model. Finally, the gene model was compared with reference gene regions marked purple; which enabled the 5' and 3' ends of the genes to be extended (A). The table in green shows some of the information on the optimized genes (B)

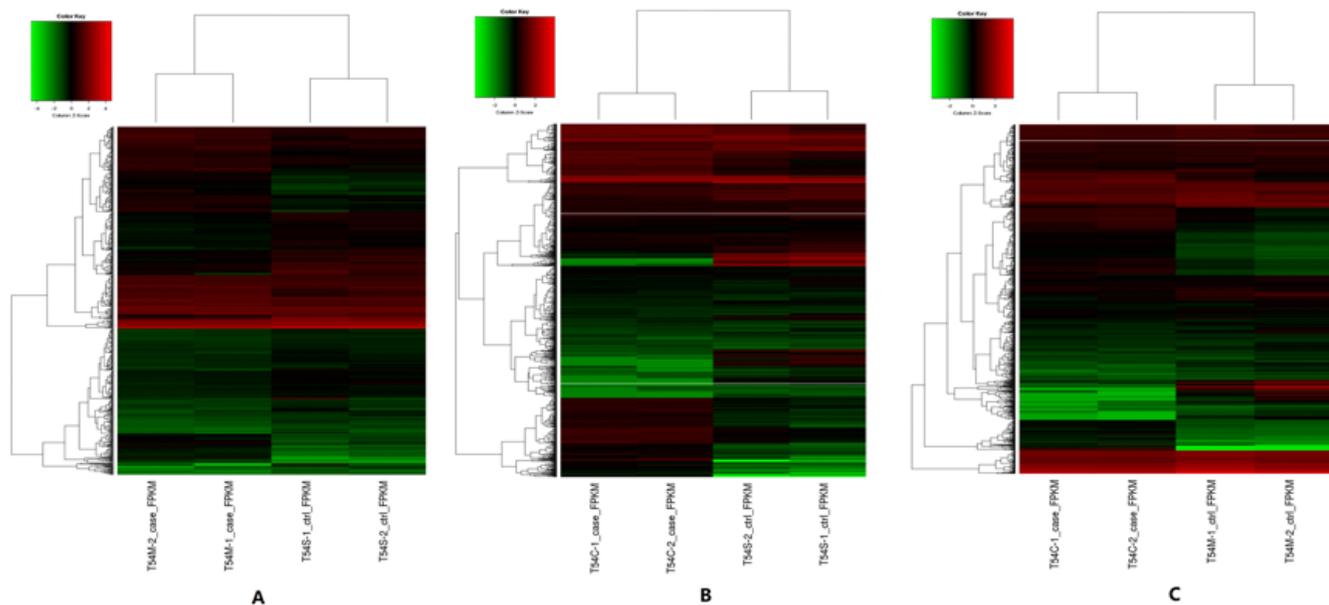


Figure 4

Cluster analysis of DEGs in the three comparisons of T54M vs T54S (A), T54C vs T54S (B), and T54C vs T54M (C). In the heatmap of DEGs, red indicates genes with high expression genes, and green indicates genes with low expression. The colors from red to green represent descending $\log_{10}(\text{FPKM} + 1)$ values

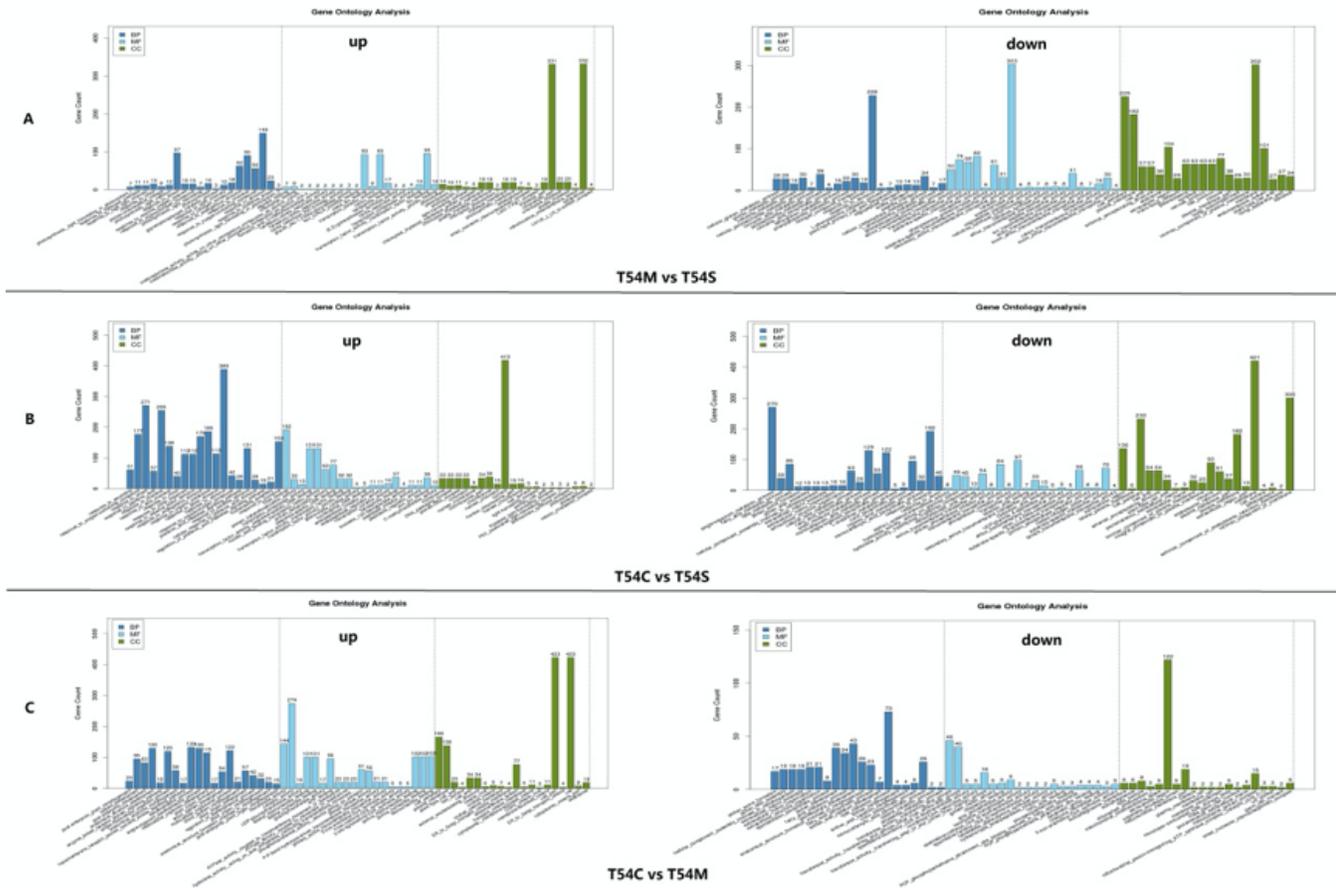


Figure 5

GO classifications of the DEGs in the broccoli bud transcriptome in three comparisons: T54M vs T54S (A), T54C vs T54S (B), and T54C vs T54M (C)

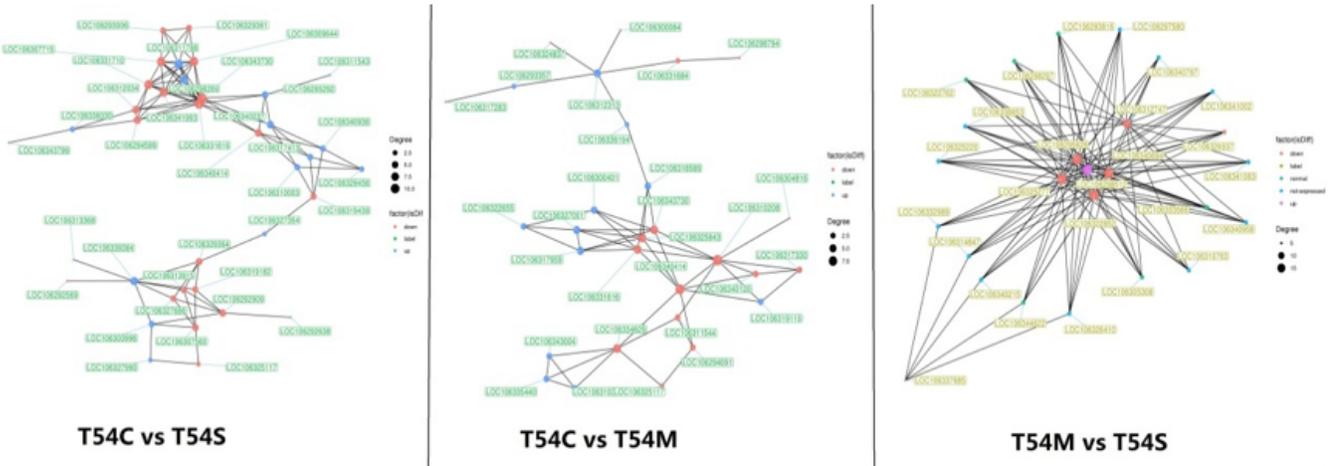


Figure 6

GO classifications of the DEGs in the broccoli bud transcriptome in three comparisons: T54M vs T54S (A), T54C vs T54S (B), and T54C vs T54M (C)

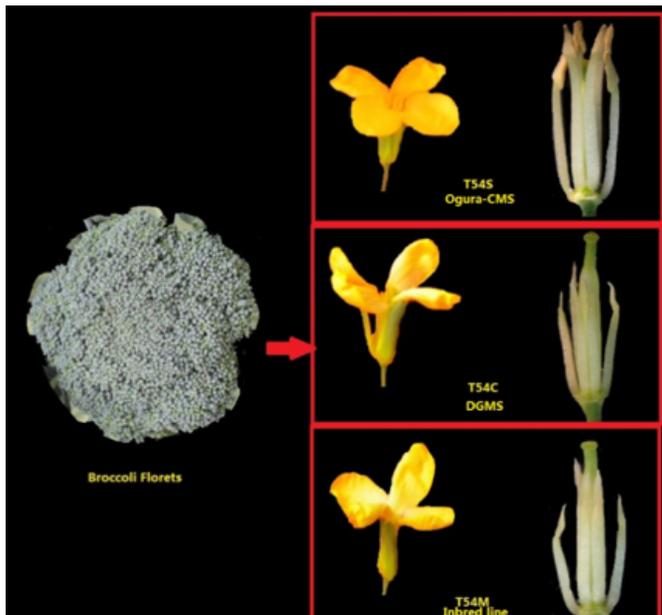


Figure 7

Head, flower, ovary and stamens characteristics among three broccoli lines derived from the Ogura-CMS line T54C, the DGMS line T54M and their maintainer line T54S

Supplementary Files

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

- [TableS1Informationoftheoptimizedgenes..xls](#)
- [FigureS1.doc](#)