

# Frequent Gain- and Loss-of-Function Mutations of the BjMYB113 Gene Accounted for Leaf Color Variation in Brassica Juncea

Guanghai An

Huazhong Agricultural University: Huazhong Agriculture University <https://orcid.org/0000-0002-7919-6194>

Jiongjiong Chen (✉ [jjchen@mail.hzau.edu.cn](mailto:jjchen@mail.hzau.edu.cn))

Huazhong Agricultural University: Huazhong Agriculture University <https://orcid.org/0000-0002-3263-0002>

---

## Research article

**Keywords:** anthocyanins, MYB transcription factor, BSR-seq, map-based cloning, Brassica juncea

**DOI:** <https://doi.org/10.21203/rs.3.rs-154002/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Mustard (*Brassica juncea*) is an important vegetable, and some cultivars have purple leaves and accumulate more anthocyanins than the green. The genetic and evolution of purple trait in mustard has not been well studied.

**Result:** In this study, free-hand sections and metabolomics showed that the purple leaves of mustard accumulated more anthocyanins than green ones. The trait of purple leaves in mustard, *Mustard Purple Leaves* (*MPL*), was genetically mapped and a *MYB113-like* homolog was identified as the candidate gene. We identified three alleles of the *MYB113-like* gene, *BjMYB113a* from a purple cultivar, *BjMYB113b* and *BjMYB113c* from green cultivars. A total of 45 single nucleotide polymorphisms (SNPs) and 8 InDels were found between the promoter sequences of the purple allele *BjMYB113a* and the green allele *BjMYB113b*. On the other hand, the only sequence variation between the purple allele *BjMYB113a* and the green allele *BjMYB113c* is an insertion of 1,033-bp fragment in the 3' region of *BjMYB113c*. Transgenic assay and promoter activity studies showed that the polymorphism in the promoter region was responsible for the up-regulation of the purple allele *BjMYB113a* and high accumulation of anthocyanin in the purple cultivar. The up-regulation of *BjMYB113a* increased the expression of genes in the anthocyanin biosynthesis pathway including *BjCHS*, *BjF3H*, *BjF3'H*, *BjDFR*, *BjANS* and *BjUGFT*, and consequently led to high accumulation of anthocyanin. However, the up-regulation of *BjMYB113* was compromised by the insertion of 1,033-bp in 3' region of the allele *BjMYB113c*.

**Conclusion:** Our results contribute to a better understanding of the genetics and evolution of the *BjMYB113* gene controlling purple leaves and provide useful information for further breeding programs of mustard.

## Background

Mustard (*Brassica juncea*) is an important economic vegetable in the world (Liu et al. 2013; Xie et al. 2014). This cultivated species contains several different horticultural types, which exhibit extreme morphologic polymorphisms, with leaves, stems, roots, or seeds as harvest organs. For leafy mustard cultivars, some have green leaves while others have purple leaves. The purple color is caused by the accumulation of anthocyanins in leaf epidermal cells (Heng et al. 2020).

Anthocyanins are a large group of water-soluble natural pigments and widely distribute in higher plants. They are largely responsible for the red, purple, and blue colors in flowers, fruits, leaves, seeds and other plant organs (Kong et al. 2003; Stuurman et al. 2004). In plants, anthocyanins play important roles in attracting pollinators and seed distributors (Stuurman et al. 2004). Anthocyanins can protect plants from UV damage, and they have strong capacity to scavenge reactive oxygen species (ROS), which is critical for biotic and abiotic tolerance (Albert et al. 2009; Cao et al. 2017; Gould 2004; Neill and Gould 2003; Olsen et al. 2009; Zhang et al. 2013). This group of metabolites also promote human health with obvious functions against obesity and cancers (Tsuda et al. 2003; Wang and Stoner 2008; Zhang et al. 2014a).

Previous studies have shown that transcriptional regulations of anthocyanin biosynthesis in higher plants are mainly controlled by the MBW ternary complex containing MYB, bHLH and WD repeat transcription factors (Koes et al. 2005; Ramsay and Glover 2005). The MBW ternary complex transcriptionally regulates the genes in the anthocyanin biosynthetic pathway, including *phenylalanine-ammonia lyase (PAL)*, *chalcone synthase (CHS)*, *chalcone isomerase (CHI)*, *flavanone 3-hydroxylase (F3H)*, *dihydroflavonol 4-reductase (DFR)*, and *anthocyanidin synthase (ANS)* (Holton and Cornish 1995; Springob et al. 2003). Upregulation of components of the MBW complex has been shown to be responsible for high anthocyanin accumulation in many plant species such as peach (Zhou et al. 2015), apple (Espley et al. 2007), pear (Li et al. 2012), grape (Walker et al. 2007), blood orange (Butelli et al. 2012), petunia (Quattrocchio et al. 1999), *Arabidopsis* (Baudry et al. 2004), rice (Sweeney et al. 2006), lettuce (Su et al. 2020), bokchoy (Zhang et al. 2014b), kale (Yan et al. 2019), cauliflower (Chiu et al. 2010), barrel medic (Li et al. 2016), etc. The genetics controlling purple mustard has been reported recently, and a MYB-encoding gene was suggested to control for the color polymorphism of purple and green between different genotypes (Heng et al. 2020). However, the function of the *MYB* gene in mustard has not been verified, and the evolution and mechanism of activation of this gene remain unknown.

Compared with wild progenitors, cultivated horticultural crops showed rich color polymorphisms. The gain-of-function mutations on the anthocyanin pathway have mainly occurred in the encoding genes of members of the MBW complex (see above). For example, the *BoMYB2* gene in *Brassica oleracea* was activated independently for at least three times causing purple color in kale, cauliflower, and cabbage. The activation events in *BoMYB2* included an insertion of a CACTA transposon, an insertion of a harbinger transposon and point mutations in its promoter region, respectively (Chiu et al. 2010; Yan et al. 2019). Four genes encoding transcription factors bHLH, R2R3-MYB, R3-MYB and WD40, respectively, showed sequence and functional polymorphisms in cultivated lettuce, and mutations on three of them promoted accumulation of anthocyanins on lettuce leaves (Su et al. 2020).

In this study, we dissected the genetics controlling purple mustard using Bulk Segregant Analysis and RNA-seq (BSR-seq). The gene was genetically fine-mapped and functionally confirmed through agrobacteria-mediated transformation. The mutation events were analyzed in detail, and the evolution of the causal gene was proposed. Our study contributes to a better understanding of the genetics and evolution of the gene controlling purple leaves and provides useful information for further breeding programs of mustard.

## Results

### Characterization of anthocyanin in mustard cultivars

Mustard has both green and purple cultivars (Fig. 1a, b, c). We compared the total anthocyanin contents of one purple mustard cultivar (pl102) and two green cultivars (rt104 and gre101). As expected, the total anthocyanin content in the purple cultivar was considerably higher than those in the two green cultivars (Fig. 2a). Free-hand sections of leaf tissues of the purple cultivar showed that the anthocyanins

accumulated mainly in the epidermal cells and in less extent in the mesophyll cells near the epidermises. In contrast, no obvious anthocyanins were found in green cultivars (Fig. 1d).

We further used Liquid Chromatography-Electrospray Ionization-Mass Spectrometry (LC-ESI-MS) to analyze the types of anthocyanins and their concentrations in the purple and green cultivars. A total of 18 anthocyanins were identified (Table S1). Of them, six had significantly higher concentration in the purple cultivar than in green one ( $p < 0.05$ ), including Delphinidin 3,5-O-diglucoside, Delphinidin 3-O-(6"-O-malonyl)-beta-D-glucoside, Cyanidin 3-O-(6"-O-malonyl)-beta-D-glucoside, Pelargonidin 3-O-glucoside, Peonidin 3,5-O-diglucoside, and Peonidin 3-O-glucoside, all emitted dark color (Fig. 2b). We conclude that the purple color in mustard leaves was caused by high accumulation of anthocyanins.

## Map-based cloning of the gene controlling color variation in mustard

A previous study suggested that an insertion of 1,268-bp fragment in the first intron of a *MYB* gene was responsible for its loss-of-function and green leaves (Heng et al. 2020). However, how the *MYB* gene was activated in purple cultivars was unclear. Furthermore, some green cultivars such as rt104 and gre101 do not have the 1,268-bp insertion, and therefore should have distinct mechanisms for loss-of-function. To answer these questions, purple cultivar pl102 was crossed with green cultivars rt104 and gre101, respectively. Both  $F_1$  hybrids were purple, and they were selfed to generate two  $F_2$  segregating populations.

Of the 137 individuals in the  $F_2$  segregating population derived from the cross pl102  $\times$  rt104, 102 individuals had purple leaves and 35 individuals had green leaves, with an expected ratio of 3:1 ( $\chi^2 = 0.0219 < \chi^2(0.05,1) = 3.84, P > 0.05$ , Table S2), suggesting a single gene controlling the color variation, which was referred to as *Mustard Purple Leaves (MPL)*. Then, we used BSR-seq to genetically map the gene underlying the color variation. The *MPL* gene was mapped to chromosome J15 (B05) (Fig. 3a). Screening a total of 1,353 individuals from the  $F_2$  population ultimately mapped the *MPL* gene between markers AGH260 and AGH263 on J15 (B05), in an interval of approximately 170 kb region (Fig. 3b, c). Only 12 open reading frames (ORFs) were predicted in this interval. The *MYB* gene, which is an ortholog of *MYB113* in *Arabidopsis* and was previously shown to control color variation in mustard, was also located in the candidate interval and remained a reliable candidate gene (*BjMYB113*) (Fig. 3d). As mentioned above, the 1,268-bp insertion, which was suggested to inactivate the candidate gene in a previous study, was absent in the green parent (Heng et al. 2020). To verify the *BjMYB113* as the candidate gene, we first investigated its sequence variation between the two parents, including its 2,842 bp upstream sequences. A total of 45 SNPs and 8 InDels were found between the two alleles from the purple parent pl102 and the green parent rt104 (Fig. 4).

Interestingly, such sequence variation was not found between the purple parent pl102 and another green cultivar gre101. The *BjMYB113* gene from green cultivar gre101 is almost identical to that from the purple cultivar, with only one SNP at -2,269 of promoter region (Fig. 4), and the *BjMYB113* gene from either

cultivar does not have the 1,268-bp insertion in intron 1. BSR-seq analysis of the pl102 × gre101 segregating population also showed that the gene controlling leaf color was located in the same region as *BjMYB113* on chromosome J15 (Figure S1) (see above). We then sequenced its flanking region and discovered an insertion of 1,033-bp fragment at 2,873 bp downstream its stop codon. Therefore, there are at least four alleles of the *BjMYB113* gene, including *BjMYB113a* (purple allele), *BjMYB113b* (green allele) and *BjMYB113c* (green allele with the 1,033-bp insertion in its 3' downstream), as well as an allele identified previously with an insertion of 1,268-bp fragment in the first intron (Fig. 4).

### **Overexpressing the *BjMYB113a*, *BjMYB113b*, or *BjMYB113c* alleles promoted anthocyanin accumulation in *Arabidopsis***

The coding sequence of the three alleles *BjMYB113a*, *BjMYB113b* and *BjMYB113c* are highly conserved in different mustard cultivars (Figure S2). The coding sequence of *BjMYB113a* and *BjMYB113c* are identical, while *BjMYB113a* and *BjMYB113b* vary by three SNPs, with only one of them leading to amino acid change (M120K) (Figure S2). To verify the function of these three *BjMYB113* alleles, their coding sequence were driven by cauliflower mosaic virus (CaMV) 35S promoter and transferred into *Arabidopsis thaliana* Col-0. All *Arabidopsis* lines overexpressing *BjMYB113* alleles had purple leaves and stems in comparison with green leaves and stems in wild type (Fig. 5c, d, e, f). As expected, the color change was due to high accumulation of anthocyanins and high expression of *BjMYB113* (Fig. 5a, b). Above results indicated that the proteins encoded by the three *BjMYB113* alleles were functionally identical. Therefore, the color variations of mustard cultivars were most likely caused by expression difference of the *BjMYB113* alleles.

### **Promoter region but not coding region is responsible for the activation of the *BjMYB113a* gene**

We then investigated the expression level of the *BjMYB113* gene in different mustard cultivars. A quantitative real-time PCR analysis demonstrated that the expression level of *BjMYB113a* (from purple cultivar) was significantly higher than *BjMYB113b* and *BjMYB113c* (from green cultivars) (Fig. 6a). We hypothesize that the expression difference between alleles *BjMYB113a* and -b was caused by their promoter regions, and the expression difference between alleles *BjMYB113a* and *BjMYB113c* was caused by the 1,033-bp insertion in downstream region of the latter.

To test above hypothesis, promoter activity of the three *BjMYB113* alleles was analyzed. *GUS* gene driven by the promoter of *BjMYB113a*, *BjMYB113b*, and *BjMYB113c*, respectively, was transformed into *Arabidopsis* Col-0. Six transgenic lines were obtained for each allele, and the GUS activities of these transgenic lines were measured. The GUS activity driven by promoter of *BjMYB113a* or *BjMYB113c* was significantly higher than that driven by the promoter of *BjMYB113b*, but no activity difference was observed between the promoters of *BjMYB113a* and *BjMYB113c* (Fig. 6b). Therefore, the polymorphisms in the promoter region accounts for the expression difference between *BjMYB113a* and *BjMYB113b*, and the low expression of *BjMYB113c* was likely caused by the 1,033-bp insertion in its downstream rather than the point mutation in its promoter region.

Forty-five SNPs and eight InDels were found between the promoter regions of *BjMYB113a* and *BjMYB113b*. The motif analysis was carried out for these two promoters. The activation of the *BjMYB113a* in purple cultivar pl102 was likely due to the point mutations at -1,317 (C to G), -1,576 (A to G), -1,588 (G to T), -2,593 (C to T). These four SNPs produced new transcription factor binding sites, including a MBSI motif (AAAAAAC(G/C)GTTA), a bHLH motif (CANNTG), and two TATA-box motifs (ATTATA).

### **The 1,033-bp insertion suppresses the expression of *BjMYB113c***

To verify the effects of the 1,033-bp insertion on the expression of *BjMYB113*, the *BjMYB113c* allele, including its native promoter and 3' region was transformed to *Arabidopsis* Col-0. No color change was observed in positive transgenic lines (Fig. 5i). In contrast, transgenic line of a construct containing *BjMYB113a* gene with its native promoter and 3' downstream sequences did change the color from green to purple (Fig. 5g). Note the main sequence difference between above two constructs was the 1,033-bp insertion. We conclude that high expression of *BjMYB113a* was due to mutations in its promoter region, and that the low expression of *BjMYB113c* was caused by the suppression effects of the 1,033-bp fragment inserted in its downstream region.

### **High expression of the *BjMYB113a* activated the anthocyanin biosynthetic genes**

As an important component in the MBW transcription factor complex, activated *BjMYB113* gene may up-regulate multiple genes in the anthocyanin biosynthesis pathway. qRT-PCR was carried out to investigate the expression difference of genes, including *BjPAL* (*BjuA036480*), *BjC4H* (*BjuB015902*), *BjCHS* (*BjuA041225*), *BjCHI* (*BjuA004576*), *BjF3H* (*BjuA035478*), *BjF3'H* (*BjuA047311*), *BjDFR* (*BjuA033678*), *BjANS* (*BjuB044852*), and *BjUGFT* (*BjuA047199*). The qRT-PCR results showed that the expression of *BjCHS*, *BjF3H*, *BjF3'H*, *BjDFR*, *BjANS*, and *BjUGFT* in purple cultivar pl102 were significantly higher than that in green cultivars rt104 and gre101 (Fig. 7).

### **The evolution of the *BjMYB113* gene**

To understand the evolution of the *BjMYB113* gene in mustard, approximately 3 kb promoter sequences of 58 mustard accessions including 5 purple and 53 green ones were PCR amplified and sequenced. All accessions were also genotyped for the 1,268-bp insertion in intron 1 and the 1,033-bp insertion in its downstream. A neighbor-joining tree was constructed using the full-length gene sequences, with *BoMYB113* from *Brassica oleracea* as an outgroup (Figure S3). Two major clades were found for these gene sequences from mustard, varying at least 50 nucleotides in promoter sequences between members from the two clades. The six sequences from Clade 2 are identical, all with the inactivated promoter as in green cultivar rt101 (*BjMYB113b*). The Clade 1 contains five distinct sequences including the sequence of *BjMYB113a*, which has the activated promoter and lacks the 1,033-bp insertion in its downstream. The second sequence (with three cultivars) is *BjMYB113c* with one nucleotide different from *BjMYB113a* in promoter sequence and has the 1,033-bp insertion in its downstream. The third sequence (two cultivars) varies only two nucleotides from *BjMYB113a* and also has the 1,033-bp insertion. The fourth sequence

(five cultivars) has the same promoter sequence as *BjMYB113a* and with the 1,033-bp insertion. The fifth sequence (one cultivar LY) varies a 38-bp insertion in the promoter sequence and has the 1,268-bp insertion in intron 1. Therefore, all promoter sequences from Clade 1 are most likely activated ones (see above). Surprisingly, of the 52 cultivars in Clade 1, 36 have green leaves but they do not have the large insertions in either intron 1 or in its 3' downstream, suggesting that there might be other mutation events leading to the inactivation of the *BjMYB113* gene or loss-of-function mutations in other genes of the anthocyanin biosynthesis pathway.

## Discussion

# MYB plays a central role in color variation in plant species

In this study, we showed that the *BjMYB113* gene was responsible for the color variation in mustard, and point mutations and small InDels in the promoter region accounted for its activation and accumulation of the anthocyanins. Color polymorphisms might be caused by any genes in the anthocyanin biosynthesis pathways. However, the most frequent causal genes for natural color variations were MYB-encoding genes. For example, the *RLL2* gene which encodes an R2R3-MYB transcription factor, regulates multiple genes from the anthocyanin biosynthesis pathways and promote the high-level accumulation of anthocyanins in lettuce leaves (Su et al. 2020). In *B. oleracea* species, the *BoMYB2* gene has been activated independently for at least three times (Chiu et al. 2010; Yan et al. 2019). The blood orange arose by insertion of a Copia-like retrotransposon to a MYB gene *Ruby* (Butelli et al. 2012). The *OsC1* mutations in the coding region varies degrees of apiculus coloration in rice (Saitoh et al. 2004).

While wild species in natural populations have green leaves, many horticultural cultivars have red/purple leaves, such as red leaf plum, purple lettuce and purple mustard. With the assumption that their wild ancestors had green leaves, the red or purple cultivars arose from mutations and artificially selected and maintained during domestication or breeding. We showed in this study, the *BjMYB113* gene was activated due to the accumulated point mutations and/or InDels. It will be interestingly to investigate which point mutation(s)/InDel(s) played the critical role in its activation.

In this study, we also found a mutation event that inactivated the *BjMYB113* gene, a large insertion in its downstream. BLASTN search of the inserted sequences showed that it had a high copy number in the genome of mustard. The inserted sequences do not show similarity to known transposable elements, and the sequences *per se* do not show a typical transposon structure such as target site duplication (TSD) or terminal inverted repeats (TIR). The mechanism for the inactivation of the *BjMYB113* gene by the insertion remains unknown. One possibility is that the insertion disrupts an enhancer motif of the gene. It is also possible that the insertion might have altered the chromatin structure and consequently down-regulates gene expression in its vicinity. Alternatively, the insertion of a high copy sequence may introduce methylation to this region to silence the gene.

Our sequence analysis suggested that at least 36 cultivars with green leaves have the activated promoter of the *BjMYB113a* allele from the purple cultivar, but they do not have the 1,033-bp insertion in its 3' downstream or 1,268-bp insertion in intron 1 that could silence the gene. We sequenced the entire *BjMYB113* gene from some cultivars, but no sequence variations were identified. Therefore, these green phenotypes were most likely caused by loss-of-function mutations of other genes in the anthocyanin biosynthesis pathway rather than the *BjMYB113* gene.

In a recent report, a 1,268-bp insertion in the first intron suppressed the *BjMYB113* gene, leading to a low anthocyanin accumulation (Heng et al. 2020). The only difference between the promoter sequences of the *BjMYB113* gene from that cultivar LY and the *BjMYB113a* allele is a 38-bp insertion, and it is still unclear whether the low expression of *BjMYB113* gene in the cultivar is due to the 38-bp insertion in its promoter or the insertion of the 1,268-bp in its intron 1.

## Different mutation events activating genes associated with anthocyanin biosynthesis

Many mutation events may activate genes associated with anthocyanin biosynthesis, including point mutations, insertions of transposable elements, gene duplications, and InDels in promoter region. Insertions of a CACTA-like transposon and a harbinger transposon activated *BoMYB2*, respectively, leading to anthocyanin accumulation in purple kale and purple cauliflower (Chiu et al. 2010; Yan et al. 2019). A non-LTR retrotransposon activated *CaAn2* in purple *Capsicum annuum* (Jung et al. 2019). A Copia-like retrotransposon activated *Ruby* leading to blood orange in *Citrus* (Butelli et al. 2012). Gene duplications and mutations in promoter region activated the *RLL2* gene leading to activated anthocyanin biosynthesis in lettuce (Su et al. 2020). The promoter variation of two *B* alleles determine distinct tissue specificities of anthocyanin production in maize (Radicella et al. 1992).

## BSA in combination with RNA-seq is an efficient method to map and clone causal genes

BSR-seq, an efficient and economic method, has been widely applied to identify genes that control important traits in many crops, including wheat (Li et al. 2020), cabbage (Li et al. 2019), watermelon (Sun et al. 2020), rice (Gao et al. 2019), cauliflower (Zhao et al. 2020), Chinese kale (Tang et al. 2017). In this study, the *MPL* gene was also cloned through the BSR-seq analysis of  $F_2$  segregating populations. Using RNA-seq may not detect as many SNPs as DNA-seq would. However, in most cases, the number of SNPs identified by RNA-seq is enough for genetic analysis. Large difference of allele frequency of a SNP between two extreme pools provides direct evidence for the linkage between the SNP and a causal gene. The RNA-seq of the two pools also provides information on differentially expressed genes (DEGs). The DEGs, if located in the candidate region, might facilitate the identification of candidate gene. The GO analysis of the DEGs between the two extreme pools may also suggest the pathways that are associated with the trait of interest. In this study, the causal gene for the color variation in mustard, *BjMYB113*, showed expression difference between the two pools. Furthermore, several genes in the anthocyanin biosynthesis pathway were up-regulated in the purple pool compared with the green pool of the  $F_2$

segregating population. All these information greatly helped us pinpointing the *BjMYB113* gene as the causal gene and understanding the pathway involved for the color variation in mustard.

## Conclusions

We genetically cloned a R2R3-MYB coding gene controlling purple leaves in mustard. *BjMYB113* was activated due to promoter variations, which leading to the high expression of genes in the anthocyanin biosynthesis pathway and the high accumulation of anthocyanin in the purple cultivar. However, a large insertion in its 3' region or in its first intron compromised the high expression in the active allele leading to green color. These results indicated that *BjMYB113* transcription factor, as the important member of the MBW ternary complex, has experienced both gain-of-function and loss-of-function mutations during artificial selection and domestication. Our results contribute to a better understanding of the genetics and evolution of the *BjMYB113* gene controlling purple leaves and provide useful information for further breeding programs of mustard, which will be of broad interest to biologists.

## Methods

### Plant materials and genetic segregating populations

Purple mustard cultivar pl102 and green cultivars rt104 and gre101 were chosen to study the genetics of purple color leaves. Purple cultivar pl102 was crossed with green cultivars rt104 and gre101, respectively, to develop two F<sub>2</sub> segregating populations. These F<sub>2</sub> segregating populations along with its parents were acquired from professor Hanhui Kuang's lab (College of Horticulture and Forestry Sciences, Huazhong Agricultural University) in this study. And the parents' seeds were originally acquired at National Center for Vegetable Improvement (Central China). The materials were germinated and planted in a greenhouse under 16/8 h photoperiod at 25 °C ± 2 °C. All plants were planted to the field on campus of Huazhong Agriculture University, Wuhan, China.

### Analysis of total anthocyanin contents

Method of total anthocyanin content analysis followed the guidelines and protocols described previously (Shin et al. 2007). For anthocyanin extraction, 0.1 g tissues of each sample were incubated in 300 µL of extraction buffer (methanol containing 1% HCl) overnight at 4 °C in the dark. After extraction, 200 µL of water and 200 µL of chloroform were added to each sample, and absorbances were read at 530 and 657 nm. The quantity of anthocyanin was determined by calculating absorbance at 530 nm (A<sub>530</sub>) – 0.33A<sub>657</sub>, and each sample was extracted and measured in three independent experiments.

### Flavonoid metabolite profiling

The relative quantities of flavonoid metabolites in *Brassica juncea* leaf samples were analyzed with a liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) system by MetWare (Wuhan, China). The liquid chromatography–electrospray ionization-tandem mass

spectrometry (LC-ESI-MS/MS) system was used for the relative quantification of anthocyanin metabolites in *Brassica juncea* leaves samples. The protocols were described detailedly in Methods S1.

## Bulked Segregant Analysis and RNAseq (BSRseq)

Bulked Segregant Analysis in combination with RNA-seq (BSR-seq) was used to map genes controlling purple leaves (Liu et al. 2012). A total of 20 purple individuals from a segregating population were mixed as the purple-pool, and 20 green individuals were mixed as the green-pool. Total RNAs were extracted from the two pools using RNAiso plus (Takara, Japan). RNA-seq was performed on Illumina HiSeq2500 platform (Novogene, China), and approximately 5 GB clean data were obtained for each pool. The data were mapped to *Brassica juncea* reference genome (Yang et al. 2016), using the Bowtie software (Langmead et al. 2009). SNP callings were performed using SAMtools (Li et al. 2009). Low-quality SNPs with map quality value < 30, reads depth < 10× or base quality value < 20 were excluded. The key parameter of  $\Delta(\text{SNP-index})$  was used to identify the target region for purple/green traits (Takagi et al. 2013). The  $\Delta(\text{SNP-index})$  was calculated by subtracting the SNP-index value of the green-pool from the SNP-index value of purple-pool. Cleaved amplified polymorphic sequence (CAPS) markers were designed in the candidate region, and were used to screen the population to fine map the casual gene. The primers used in map-based cloning were shown in Table S3. The RNA-seq data supporting the results of this study is available in the NCBI SRA (Sequence Read Archive, <http://www.ncbi.nlm.nih.gov/sra/>) repository under BioProject PRJNA672814.

### RNA extraction and quantitative Real-Time reverse transcription polymerase chain reaction (qRT-PCR) analysis.

Total RNAs were extracted from leaves using RNAiso plus (Takara, Japan). The cDNA was synthesized using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransScript, China) with Oligo-dT18 primer. qRT-PCR analysis followed the guidelines and protocols described previously (Udvardi et al. 2008). All reactions were performed using the SYBR Premix (5.0  $\mu\text{L}$  of 2× SYBR Premix Go Taq II, 0.5  $\mu\text{L}$  of primers, 1.0  $\mu\text{L}$  of cDNA, and 3.5  $\mu\text{L}$  of ddH<sub>2</sub>O). Melting curve analysis of qRT-PCR samples revealed that there was only one product for each gene primer reaction. The PCR products were sequenced to confirm the specific amplification. A house-keeping gene *BjEF-1- $\alpha$ <sup>37</sup>* was used as an internal standard in tissues. The relative expression levels were counted using the formula  $2^{-\Delta\Delta\text{Cq}}$  as described in Bio-Rad protocol, and statistical differences were calculated using student's test. Three biological replications and three technical replications were performed in qRT-PCR. The primers used in qRT-PCR analysis were shown in Table S3.

## Plasmid construction and plant transformation

The full-length *BjMYB113* cDNA was cloned into the pRI101-GFP vector with CaMV 35S promoter to construct the overexpression vector. The full-length *BjMYB113* gene sequence (including native promoter and 3' region sequence) was amplified and cloned into the pRI101-GFP to construct the complementary vector. Approximately 3 kb promoter region of three *BjMYB113* alleles were amplified and recombined

into the pCAMBIA1301-GUS vector for checking GUS activity. All vectors were constructed using homologous recombination. Positive plasmids were verified by sequencing and then transformed into *Agrobacterium tumefaciens* GV3101 using thermal stimulation method. The vectors were transformed into *Arabidopsis* using floral-dip method (Clough and Bent 1998). All primers used for vector construction were shown in Table S3.

## Promoter activity analysis

The quantitative GUS activity was measured using the Lu's methods with slight modification (Lu et al. 2016). GUS activity was detected in 1-month-old *Arabidopsis* leaf tissues (10 mg) from three independent transgenic lines and six individuals in each line. Total proteins were extracted using 300  $\mu$ L GUS extraction buffer (50 mM phosphate buffer, pH 7.0; 10 mM EDTA, pH 8.0; 0.1% Sodium Dodecyl Sulfate; 10 mM  $\beta$ -mercaptoethanol). BCA Protein Assay (Beyotime Biotechnology, China) was used to measure the protein concentrations. Extraction (100  $\mu$ L) was added to 900  $\mu$ L GUS extraction buffer containing 1 mM 4-methylumbelliferyl glucuronide (MUG, Sigma) and incubated at 37 °C. The 900  $\mu$ L stop solution (1 M Sodium Carbonate) immediately added into 100  $\mu$ L the above reaction mixture and 60 min later, respectively. Fluorescence of 4-methylumbelliferone (MU) was monitored using Tecan Infinite™ at 455 nm emission and 365 nm excitation. GUS activity was expressed as  $\mu$ moles 4-methylumbelliferone (MU)  $\text{min}^{-1} \text{mg}^{-1}$  protein.

## Sequences analysis and neighbor-joining tree

The promoter sequence of *BjMYB113* alleles was analyzed using PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

Sequence alignments were conducted using Muscle program and manually adjusted in GeneDoc (<http://www.nrbsc.org/gfx/genedoc/>). The neighbor-joining tree was constructed using MEGA 7.0 (Kumar et al. 2016) and bootstrap values were calculated using 1,000 times.

The gene sequences (*BjMYB113a*, *-b*, and *-c*) supporting the results of this study are available in the NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession number MW166171- MW166173 of BankIt2394807. The primers used in gene sequences analysis were shown in Table S3.

## Abbreviations

### ANS

Anthocyanidin synthase

### BSR-seq

Bulked Segregant Analysis and RNA-seq

### CaMV

Cauliflower mosaic virus

### CAPS

Cleaved amplified polymorphic sequence

**CHI**

Chalcone isomerase

**CHS**

Chalcone synthase

**DEGs**

Differentially expressed genes

**DFR**

Dihydroflavonol 4-reductase

**F3H**

Flavanone 3-hydroxylase

**GO**

Gene Ontology

**InDels**

Insertions and Deletions

**LC-ESI-MS**

Liquid Chromatography-Electrospray Ionization-Mass Spectrometry

**MPL**

Mustard Purple Leaves

**MU**

4-methylumbelliferone

**MUG**

4-methylumbelliferyl glucuronide

**ORFs**

Open reading frames

**PAL**

Phenylalanine-ammonia lyase

**qRT-PCR**

quantitative Real-Time reverse transcription polymerase chain reaction

**ROS**

reactive oxygen species

**SNPs**

single nucleotide polymorphisms

**TIR**

terminal inverted repeats

**TSD**

target site duplication

**Declarations****Ethics approval and consent to participate**

Not applicable.

## Consent for publication

Not applicable.

## Availability of data and materials

The data supporting the result of this study are within the paper and its additional files. All sequencing datasets are deposited in the National Centre for Biotechnology Information (NCBI) under the BioProject ID PRJNA672814 with the Sequence Read Archive (SRA) accession SRR12919405-SRR12919408. The gene sequences (*BjMYB113a*, *-b*, and *-c*) supporting the results of this study are available in the NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession number MW166171- MW166173 of BankIt2394807.

## Competing interests

The authors declare no conflict of interests.

## Funding

This work was supported by grants from the National Key Research and Development Program of China [2018YFD1000800], and the Fundamental Research Funds for the Central Universities [2662018PY070]. All funders were not involved in the design of the study, data analysis, and writing the manuscript.

## Author contributions

J.C. designed the project. G.A. performed the experiments. G.A. wrote the manuscript with the helps from J.C.

## Acknowledgements

We thank Dr. Changchun Yu (Huazhong Agricultural University) for the vector pRI101-GFP and pCAMBIA1301-GUS, Dr. Weiyi Zhang (Huazhong Agricultural University) for technique support of bioinformatics, Dr. Hanhui Kuang (Huazhong Agricultural University) for plant materials and helpful suggestions on this work.

## References

- Albert NW, Lewis DH, Zhang H, Irving LJ, Jameson PE, Davies KM (2009) Light-induced vegetative anthocyanin pigmentation in *Petunia*. *J Exp Bot* 60:2191-2202
- Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L (2004) *TT2*, *TT8*, and *TTG1* synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J* 39:366-380
- Butelli E et al. (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24:1242-1255
- Cao X et al. (2017) A putative *R3 MYB* repressor is the candidate gene underlying *atroroseum*, a locus for anthocyanin pigmentation in tomato fruit. *J Exp Bot* 68:5745-5758
- Chiu LW, Zhou X, Burke S, Wu X, Prior RL, Li L (2010) The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiol* 154:1470-1480
- Clough SJ, Bent AF (1998) Floral dip\_ a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735-743
- Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2007) Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J* 49:414-427
- Gao J et al. (2019) Mapping and identifying a candidate gene *Plr4*, a recessive gene regulating purple leaf in rice, by using bulked segregant and transcriptome analysis with next-generation sequencing. *Int J Mol Sci* 20:4335
- Gould KS (2004) Nature's swiss army knife: the diverse protective roles of anthocyanins in leaves. *J Biomed Biotechnol* 2004:314-320
- Heng S et al. (2020) Fine-mapping of the *BjPur* gene for purple leaf color in *Brassica juncea*. *Theor Appl Genet*:doi.org/10.1007/s00122-00020-03634-00129
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7:1071-1083
- Jung S, Venkatesh J, Kang MY, Kwon JK, Kang BC (2019) A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in *Capsicum annuum*. *Plant Sci* 287:110181
- Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci* 10:236-242
- Kong J-M, Chia L-S, Goh N-K, Chia T-F, Brouillard R (2003) Analysis and biological activities of anthocyanins. *Phytochemistry* 64:923-933

- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870-1874
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25
- Li H et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079
- Li L, Ban ZJ, Li XH, Wu MY, Wang AL, Jiang YQ, Jiang YH (2012) Differential expression of anthocyanin biosynthetic genes and transcription factor *PcMYB10* in pears (*Pyrus communis* L.). *PLoS One* 7:e46070
- Li P et al. (2016) Regulation of anthocyanin and proanthocyanidin biosynthesis by *Medicago truncatula* bHLH transcription factor MtTT8. *New Phytol* 210:905-921
- Li R et al. (2019) Conjunctive analyses of BSA-Seq and BSR-Seq to reveal the molecular pathway of leafy head formation in Chinese cabbage. *Plants (Basel)* 8:603
- Li Y et al. (2020) Identification of the vernalization gene *VRN-B1* responsible for heading date variation by QTL mapping using a RIL population in wheat. *BMC Plant Biol* 20:331
- Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One* 7:e36406
- Liu X, Lu Y, Yuan Y, Liu S, Guan C, Chen S, Liu Z (2013) *De novo* transcriptome of *Brassica juncea* seed coat and identification of genes for the biosynthesis of flavonoids. *PLoS One* 8:e71110
- Lu S, Zhang Y, Zheng X, Zhu K, Xu Q, Deng X (2016) Isolation and functional characterization of a lycopene beta-cyclase gene promoter from citrus. *Front Plant Sci* 7:1367
- Neill SO, Gould KS (2003) Anthocyanins in leaves: light attenuators or antioxidants? *Funct Plant Biol* 30:865-873
- Olsen KM et al. (2009) Temperature and nitrogen effects on regulators and products of the flavonoid pathway: experimental and kinetic model studies. *Plant Cell Environ* 32:286-299
- Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R (1999) Molecular analysis of the *anthocyanin2* gene of petunia and its role in the evolution of flower color. *Plant Cell* 11:1433-1444
- Radicella JP, Brown D, Tolar LA, Chandler VL (1992) Allelic diversity of the maize *B* regulatory gene: different leader and promoter sequences of two *B* alleles determine distinct tissue specificities of anthocyanin production. *Genes Dev* 6:2152-2164
- Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci* 10:63-70

- Saitoh K, Onishi K, Mikami I, Thidar K, Sano Y (2004) Allelic diversification at the *C (OsC1)* locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics* 168:997-1007
- Shin J, Park E, Choi G (2007) PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in *Arabidopsis*. *Plant J* 49:981-994
- Springob K, Nakajima J, Yamazaki M, Saito K (2003) Recent advances in the biosynthesis and accumulation of anthocyanins. *Nat Prod Rep* 20:288-303
- Stuurman J, Hoballah ME, Broger L, Moore J, Basten C, Kuhlemeier C (2004) Dissection of floral pollination syndromes in *Petunia*. *Genetics* 168:1585-1599
- Su W et al. (2020) Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol J* 18:479-490
- Sun L et al. (2020) Linkage mapping and comparative transcriptome analysis of firmness in watermelon (*Citrullus lanatus*). *Front Plant Sci* 11:831
- Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18:283-294
- Takagi H et al. (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 74:174-183
- Tang QW, Tian MY, An GH, Zhang WY, Chen JJ, Yan CH (2017) Rapid identification of the purple stem (*Ps*) gene of Chinese kale (*Brassica oleracea* var. *alboglabra*) in a segregation distortion population by bulked segregant analysis and RNA sequencing. *Molecular Breeding* 37:153
- Tsuda T, Horio F, Uchida K, Aoki H, Osawa T (2003) Dietary cyanidin 3-O-beta-D-glucoside-rich purple corn color prevents obesity and ameliorates hyperglycemia in mice. *J Nutr* 133:2125-2130
- Udvardi MK, Czechowski T, Scheible WR (2008) Eleven golden rules of quantitative RT-PCR. *Plant Cell* 20:1736-1737
- Walker AR, Lee E, Bogs J, McDavid DAJ, Thomas MR, Robinson SP (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant J* 49:772-785
- Wang LS, Stoner GD (2008) Anthocyanins and their role in cancer prevention. *Cancer Lett* 269:281-290
- Xie Q et al. (2014) Accumulation and molecular regulation of anthocyanin in purple tumorous stem mustard (*Brassica juncea* var. *tumida* Tsen et Lee). *J Agric Food Chem* 62:7813-7821
- Yan C et al. (2019) Independent activation of the *BoMYB2* gene leading to purple traits in *Brassica oleracea*. *Theor Appl Genet* 132:895-906

- Yang J et al. (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet* 48:1225-1232
- Zhang Y et al. (2013) Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. *Curr Biol* 23:1094-1100
- Zhang Y, Butelli E, Martin C (2014a) Engineering anthocyanin biosynthesis in plants. *Curr Opin Plant Biol* 19:81-90
- Zhang Y, Chen G, Dong T, Pan Y, Zhao Z, Tian S, Hu Z (2014b) Anthocyanin accumulation and transcriptional regulation of anthocyanin biosynthesis in purple bok choy (*Brassica rapa* var. *chinensis*). *J Agric Food Chem* 62:12366-12376
- Zhao Z, Sheng X, Yu H, Wang J, Shen Y, Gu H (2020) Identification of candidate genes involved in curd riceyness in cauliflower. *Int J Mol Sci* 21:1999
- Zhou H et al. (2015) Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J* 82:105-121