Theoretical and Applied Genetics (2024) 137:105
https://doi.org/10.1007/s00122-024-04608-x

ORIGINAL ARTICLE

QTL mapping and candidate gene analysis reveal two major loci regulating green leaf color in non-heading Chinese cabbage

Aimei Bai[1] · Tianzi Zhao[1] · Yan Li[1] · Feixue Zhang[1,2] · Haibin Wang[1] · Sayyed Hamad Ahmad Shah[1] · Li Gong[1] · Tongkun Liu[1] · Yuhui Wang[1] · Xilin Hou[1] · Ying Li[1]

Received: 19 July 2023 / Accepted: 23 March 2024

Abstract

Key message **Two major-effect QTL *GlcA07.1* and *GlcA09.1* for green leaf color were fine mapped into 170.25 kb and 191.41 kb intervals on chromosomes A07 and A09, respectively, and were validated by transcriptome analysis.**

Abstract Non-heading Chinese cabbage (NHCC) is a leafy vegetable with a wide range of green colors. Understanding the genetic mechanism behind broad spectrum of green may facilitate the breeding of high-quality NHCC. Here, we used F_2 and $F_{7:8}$ recombination inbred line (RIL) population from a cross between Wutacai (dark-green) and Erqing (lime-green) to undertake the genetic analysis and quantitative trait locus (QTL) mapping in NHCC. The genetic investigation of the F_2 population revealed that the variation of green leaf color was controlled by two recessive genes. Six pigments associated with green leaf color, including total chlorophyll, chlorophyll a, chlorophyll b, total carotenoids, lutein, and carotene were quantified and applied for QTL mapping in the RIL population. A total of 7 QTL were detected across the whole genome. Among them, two major-effect QTL were mapped on chromosomes A07 (*GlcA07.1*) and A09 (*GlcA09.1*) corresponding to two QTL identified in the F_2 population. The QTL *GlcA07.1* and *GlcA09.1* were further fine mapped into 170.25 kb and 191.41 kb genomic regions, respectively. By comparing gene expression level and gene annotation, *BraC07g023810* and *BraC07g023970* were proposed as the best candidates for *GlcA07.1*, while *BraC09g052220* and *BraC09g052270* were suggested for *GlcA09.1*. Two InDel molecular markers (*GlcA07.1-BcGUN4* and *GlcA09.1-BcSG1*) associated with *BraC07gA023810* and *BraC09g052220* were developed and could effectively identify leaf color in natural NHCC accessions, suggesting their potential for marker-assisted leaf color selection in NHCC breeding.

Introduction

The plant leaf color serves dual purposes, both aesthetically and functionally, often signifying crucial growth stages. In nature, plant leaves display a wide spectrum of green hues, ranging from light to dark shades, dictated by varying concentrations of a number of pigments, among which chlorophyll and carotenoids are the two principal contributors (Tang et al. 2020). Generally, chlorophyll primarily contributes to authentic green color, while total carotenoids intermix the yellow undertones within the green spectrum (Li et al. 2018). The precise shade of green color in leaves is intricately determined by the complex interplay and ratio of these pigment concentration. A number of factors, such as constitutive genetics, environmental conditions (i.e., light, temperature), and senescence would significantly affect the accumulation of chlorophyll and total carotenoids in plants by modulating pigment biosynthesis localization, transportation, and degradation processes (Eckhardt et al. 2004; Sun and Li 2020).

The chloroplast stands as the primary site for pigment biosynthesis, intricately associated with production of chlorophyll and total carotenoids. The differentiation of chloroplasts from plastids involves three key stages:

Communicated by Isobel AP Parkin.

✉ Yuhui Wang
yuhui_wang@njau.edu.cn

✉ Ying Li
yingli@njau.edu.cn

[1] State Key Laboratory of Crop Genetics and Germplasm Enhancement and Utilization, Engineering Research Center of Germplasm Enhancement and Utilization of Horticultural Crops, Ministry of Education of the P. R. China, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, Jiangsu Province, China

[2] Huzhou Academy of Agricultural Sciences, Huzhou 313000, Zhejiang Province, China

Published online: 15 April 2024

plastid replication and plastid DNA synthesis, the transfer of nuclear-encoded plastid RNA polymerase (NEP) to plastid genes, and the expression of plastid and nuclear genes under the plastid-encoded RNA polymerase (PEP) (Pogson and Albrecht 2011). These complex processes engage a multitude of genes governing various aspects of chloroplast development and numerous involved genes have been identified, such as genes for chloroplast ribosome biogenesis (Sun et al. 2019; Zhou et al. 2021), chloroplast RNA editing and splicing (Zhou et al. 2021; Zhao et al. 2022a, b), and the NEP/PEP pathway (Lee et al. 2019). Malfunction in these genes often result in chloroplast dysfunction, causing leaf or plant discoloration. For example, mutation in *RH13*, which encodes a DEAD-box family RNA helicases in *Arabidopsis*, inhibits chlorophyll accumulation by disrupting development of chloroplast structure (Luo et al. 2023). Knockout of *OsCRS* and *OsPPR647*, which encodes a peptidyl-tRNA hydrolase protein and a pentatricopeptide repeat-containing protein, respectively, leads to albino leaf phenotype by interfering splicing of chloroplast genes in rice (Zhang et al. 2020a, b; Zhao et al. 2021a, b). Overexpression of *OsFSD3* reduces chlorophyll level and results in yellow-green leaves by downregulating the expression of PEP-dependent genes in rice (Lee et al. 2019). Loss of function of a tetratricopeptide repeat-containing protein (SLOWGREEN 1, SG1) delays greening and causes leaf albinism (Hu et al. 2014). Another pivotal example is *GUN1*, the *GENOMES UNCOUPLED* (*GUN*) gene, which encodes a PPR-containing protein in plastids. GUN1 plays a central role in the retrograde communication between chloroplasts and nucleus, and *gun1* mutants display pale green cotyledons and leaves (Llamas et al. 2017; Tadini et al. 2020).

Chlorophyll, is the main pigment imparting characteristic green appearance to plants, actively participates in photosynthesis by harvesting light and transferring electrons (Nelson and Yocum. 2006; Yang et al. 2015). Chlorophyll content is the net result of the balance between biosynthesis and biodegradation processes (Schlicke et al. 2014). Its synthesis commences from 5-aminolevulinic acid (ALA), which derives from Glu-1-semialdehyde (GSA) through the action of GSA-2,1-aminomutase (GSA-AM) (Hotta et al. 1997; Kumar and Söll 2000). GSA-AM, encoded by the nuclear *HEMA*, when silenced, leads to varying degrees of chlorophyll deficiency (Kumar and Söll 2000). Another critical step involves the conversion of protoporphyrin IX (Proto) to Mg-protoporphyrin (Mg-Proto), catalyzed by Mg chelatase (MgCh) consisting of three subunits: CHLH, CHLD, and CHLI. Additionally, *GUN4* activates MgCh that catalyzes the insertion of an Mg^{2+} ion into Proto (Masuda 2008; Adams et al. 2020; Ji et al. 2021). For example, silencing of the *CHLH* genes induces the yellow or white leaves in tobacco (Hiriart et al. 2002), while mutations in *CHLD* and *CHLI* result in yellow-green leaves in rice (Zhang et al. 2006). Downregulation of *OsGUN4* contributes to yellow-green leaves and chlorophyll deficiency in rice, as OsGUN4 can activate the CHLH subunit of MgCh (Zhou et al. 2012; Li et. 2021). In addition, GUN4 is able to interact with BALANCE of CHLOROPHYLL METABOLISM (BCM) to stimulate the MgCh activity during the early leaf development, while BCM shows a role in preventing chlorophyll degradation (Wang et al. 2020).

Non-heading Chinese cabbage (*Brassica rapa* ssp. *chinensis*, NHCC) is a widely marketable leafy vegetable cultivated on 1.3 million hectares in China (FAO 2023). Leaf color stands as a crucial determinant of the edible organ's value, meeting diverse consumer preferences and significantly influencing market value (Amagai et al. 2022). Moreover, leaf color is also related to biomass production by affecting efficiency of photosynthesis (Su et al. 2023), abiotic/biotic resistance, and nutrient substance content such as chlorophy ll, total carotenoids, and anthocyanidin, making it the prime target trait for NHCC breeding (Dai et al. 2018; Shen et al. 2018; Priyanka et al. 2023). Research in Chinese cabbage focusing on EMS-induced mutants and spontaneous mutations resulting in leaf color variations, particularly from pale to yellow-green, has identified several single-inherited genes involved in chlorophyll biosynthesis, notably *BrCHLH*, *BrCAO*, *DVR*, *BraA07001774*, and *BraA09004189* (Fu et al. 2019; Zhang et al. 2020a, b; Zhao et al. 2021a, b, 2022a). Conversely, the natural variance in leaf color within *Brassica rapa* is primarily regulated by quantitative trait loci (QTL) (Choi et al. 2017; Huang et al. 2017). Choi et al. (2017) report 13 QTL for leaf color across 6 linkage groups (LGs) (LG1, LG3, LG4, LG7, LG8, and LG9) from DH and RIL populations of Chinese cabbage (*B. rapa.* ssp. *pekinensis*). Huang et al. (2017) identify 8 QTL in 5 chromosomes (A03, A05, A07, A09, and A10) associated with core leaf color in NHCC. These findings suggest that leaf color is significantly influenced by environmental factors as well as genotype-by-environment effects (G × E). However, these QTL have yet to be fine-mapped and cloned to identify the underlying candidate genes. Furthermore, the lack of available molecular markers has hindered the effective utilization of these QTL in breeding through marker-assisted selection. Thus, delving into the genetic inheritance and identifying of underlying candidate genes responsible for natural variations in NHCC leaf color will be crucial for germplasm innovation in NHCC breeding.

In this study, two distinct NHCC varieties exhibiting with dark-green and lime-green leaf color were selected to construct an F_2 comprising 618 individuals and an $F_{7:8}$ RIL populations with 127 lines for genetic analysis and QTL mapping. Two major-effect QTL located on chromosomes A07 and A09 were identified and further fine-mapped with genomic regions spanning 170.25 kb and 191.41 kb, respectively. The resulting molecular markers tightly linked to

the identified QTL (*GlcA07.1* and *GlcA09.1*) were evaluated for their efficiency in the natural population of NHCC germplasm.

Method and material

Plant material and experimental design

Two non-heading Chinese cabbage (NHCC) inbred lines, namely, Wutacai (female parent) and Erqing (male parent) were utilized to develop an F_2 population comprising 618 individuals and an $F_{7:8}$ RIL population with127 lines. Wutacai, a prevalent landrace in Changzhou, Jiangsu Province, had dark-green leaves, while Erqing, a common landrace in Yangzhou, Jiangsu Province, displayed lime-green leaves. Both lines were advanced inbred lines without leaf wax exhibiting glossy leaf color. The natural population, comprising 73 NHCC germplasm accessions, primarily collected from different provinces of China (Table S11) and displaying a wide range of leaf color, was applied for further gene validation.

The F_2, RIL, and natural populations were all planted at the Baima Research Station (BRS, 31°35′N and 119°09′E) of Nanjing Agricultural University in Jiangsu Province for leaf color evaluation. Furthermore, apart from the BRS location, the leaf color assessments for RIL population were conducted at two additional sites: Jurong Experimental Station (JES, 31°956′N and 119°16′E) of Nanjing Agricultural University in Jiangsu Province and Huzhou Academy of Agricultural Sciences Experimental Station (HES, 30°90′N and 120°08′E) in Zhejiang Province.

The experiments were conducted using a randomized complete block design (RCBD) consisting of three replications with four plants per replication per line for RIL and natural populations. All seeds were direct-sowed in September from 2020 to 2022. Leaf color was visually rated at 90 days post-germination for the F_2 population categorized into green and lime-green groups. The individuals with lime-green leaves or lighter color with a certain degree of yellow tones were classified into lime-green category and the remaining individuals were grouped into green category.

Pigment measurement

Mature leaves at the rosette stage from the parental lines, F_1 hybrid, RIL, and the NHCC germplasm were sampled for pigment measurement, with each sample weighing approximately 0.1 g and undergoing three biological replicates. The six pigments associated with green leaf color, including chlorophyll a, chlorophyll b, total chlorophyll, carotene, lutein, and total carotenoids, were quantified following method described by Xu et al. (2013) with minor modification. Photosynthetic pigments were extracted with ethanol/acetone solution (ethanol: acetone = 1: 1) in the dark at room temperature for 24 h. Supernatant was measured at 470 nm, 474 nm, 485 nm, 642 nm, 649 nm, and 665 nm by multi-detection microplate reader (Cytation3, BioTek, USA). The concentration (mg/L) of chlorophyll a, chlorophyll b, total chlorophyll, carotene, total carotenoids, and lutein were calculated as follows:

$$C_{\text{chlorophyll}-a}(\text{mg/L}) = 9.99A_{665} - 0.0867A_{642}$$

$$C_{\text{chlorophyll}-b}(\text{mg/L}) = 17.7A_{642} - 3.04A_{665}$$

$$C_{\text{Totalchlorophyll}}(\text{mg/L}) = 27.9A_{649}$$

$$C_{\text{carotene}}(\text{mg/L}) = 12.6A_{485} - 6.00A_{470} - 0.0298[a] + 0.336[b]$$

$$C_{\text{carotenoids}}(\text{mg/L}) = 4.92A_{474} - 0.0255[a] - 0.225[b]$$

$$C_{\text{lutein}}(\text{mg/L}) = 10.2A_{470} - 11.5A_{485} - 0.0036[a] - 0.625[b]$$

a represented for the content of chlorophyll a; *b* represented for content of chlorophyll b. The final content (mg/100 g FW) was calculated using the formula: $C * V * 100/M$, whereas C represented for the concentration of pigment, V was the total volume of extract and M was the fresh weight of leaves for pigment extraction.

Statistical analysis of phenotypic data

Analysis of variance (ANOVA) was performed for each pigment in RIL population to estimate the variance of genetic and environmental effects. Pigments for the leaf color trait were fitted in a linear mixed model (LMM) containing fixed factors (genotypes) and random factors (locations, replications within locations, and the interaction of the genotypes and locations) using the *R/lme4* package (Bates et al. 2015). Best linear unbiased predictors (BLUPs) were extracted from the mixed model with all factors treated as random effects and were further applied for preliminary QTL mapping. Spearman's rank order correlation among the pigments traits was estimated based on the BLUPs value using the *R/corrplot* package (Murdoch and Chow 1996) and the *R/tidyverse* package (Wickham et al. 2019). Broad-sense heritability (H^2) was calculated from the variance components extracted from the mixed model (Wang et al. 2022).

DNA preparation, Illumina resequencing and marker development

Genomic DNA (gDNA) was extracted from leaves of the parental lines and RIL population using Genomic DNA

Extraction Kit (TIANGEN, Beijing, China). The quality and quantity of the resulting genomic DNA were examined by agarose gel electrophoresis, Nano Photometer® (IMPLEN, CA, USA) and Qubit® 3.0 Fluorometer (Life Technologies, CA, USA). The 0.5 µg DNA per sample was used for libraries construction and sequencing. The sequencing libraries were generated using the Annoroad® Universal DNA Fragmentase Kit V2.0 (AN200101-L) and the Annoroad® Universal DNA Library Prep Kit V2.0 (AN200101-L), and the libraries were sequenced on the Novaseq 6000 S4 platform (Illumina, USA) using the NovaSeq 6000 S4 Reagent Kit V1.5. The 150 bp paired ended reads were generated. The construction of library and sequencing were performed at Annoroad Gene Tech. (Beijing) Co., Ltd (https://www.annoroad.com/).

The 'NHCC001' reference genome sequences [*Brassica rapa* ssp. *chinensis* genome var. 'NHCC001', version 1.0] were available in the Non-heading Chinese Cabbage and Watercress Database (http://tbir.njau.edu.cn/NhCCDbHubs/index.jsp) (Li et al. 2020, 2022). Sequencing reads were aligned against 'NHCC001' using the Burrows-Wheeler Aligner (BWA, v0.7.17-r1198-dirty) -SAMtools v1.3.1 pipeline (Li et al. 2009; Li and Durbin 2010; Li 2011). Variant calling and genotyping followed the Genome Analysis Toolkit (GATK) pipeline (McKenna et al. 2010). The SNP/InDel markers with low quality were filtered according to the hard filter criteria (SNP: QUAL < 40.0; QD < 2.0; MQ < 40.0; FS > 60.0; SOR > 3.0; MQRankSum < − 12.5; ReadPosRankSum < − 8.0; Ave Depth < 8. InDel: QUAL < 40; QD < 2.0; FS > 200.0; SOR > 10.0; MQRankSum < − 12.5; ReadPosRankSum < − 8.0; Ave Depth < 8). Markers that followed an expected segregation pattern for advanced RIL population were selected to construct a linkage group. Non-heading Chinese cabbage was diploid with 10 chromosomes ($2n = 10$) and at least 20 independent genomic regions were expected. Thus, a threshold of 0.05/20 ~ 0.0025 with a Bonferroni correction was adopted to achieve a genome-wide error rate at $\alpha = 0.05$ for segregation distortion evaluation. The SNP markers meeting the following criteria were kept: 1) missing rate ≤ 10%; 2) the heterozygous individuals ≤ 7; 3) locus similarity < 99%.

Genetic linkage map construction and QTL analysis

The function *mstmap* in the *R/ASMap* package (Taylor and Butler 2017) was used to adjust the marker positions and reorder the markers based on maximum likelihood (ML). The function *est.map* in the *R/qtl* package was used to estimate the genetic distance using Kosambi mapping function (Lander et al. 1987).

QTL analysis was performed using the packages *R/qtl* (Broman et al. 2003) and *R/qtl2* (Broman et al. 2019). Simple marker analysis (SMA) and composite interval mapping (CIM) analysis were adopted for preliminary analysis. To reduce the possibility of ghost QTL detection (false discoveries) by SMA and CIM, we employed the multiple-QTL mapping (MQM) analysis method (Arends et al. 2010) for further QTL mapping. The function *Mqmsetcofactors* in the *R/qtl* package was used to set the cofactor by the significant QTL detected from the preliminary QTL analysis results. The LOD threshold for declaring significant QTL was determined using 1000 permutations at 0.05 level with the function *mqmpermutation*. The function *fitqtl* was used to estimate the additive effect and the percentage variance explained by the peak marker for each QTL. The 1.5 LOD drop intervals were applied to calculate the support intervals of significant QTL.

Genotyping

To further narrow down the candidate intervals, more InDel markers were developed based on the re-sequencing data from Wutacai and Erqing and were applied to genotype additional F_2 population to identify new recombinants. The corresponding primers were designed using the SnapGene Viewer software (v5.1.6, https://www.snapgene.com/) (Table S1). Products of PCR were used for genotyping by using 9% polyacrylamide gel electrophoresis and silver staining.

Candidate gene annotation analysis

SNP/InDel markers were annotated using the SnpEff (v5.0c) to predict their effect on genetic variants of the genes (Cingolani et al. 2012). Candidate genes were annotated using BLASTp, and the gene descriptions were obtained from TAIR database (https://www.arabidopsis.org/), Non-heading Chinese Cabbage and Watercress Database (http://tbir.njau.edu.cn/NhCCDbHubs/index.jsp) (Li et al. 2022), InterPro (https://www.ebi.ac.uk/interpro/) (Mistry et al. 2021), and HMMER (http://www.hmmer.org/) (Finn et al. 2011).

Comparative transcriptome analysis and quantitative real-time RT-PCR

For transcriptomic assays, total RNA was extracted from the mature leaf tissues of the parental lines and F_1 hybrid according to the manufacturer's instruction of the HiPure Plant RNA Mini Kit (Magen, Shanghai, China). The purity of the RNA sample was detected by the NanoPhotometer ® (IMPLEN, CA, USA) of which concentration was assessed by the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The MGIEasy RNA library Preparation Kit was used to construct library. DNA nanoball (DNB) was formed using the rolling circle amplification (RCA) technology. DNBSEQ-T7 (BGI,

Shenzhen, Guangdong, China) was used for sequencing and 150 bp paired ended sequencing reads were obtained. The cDNA library construction and sequencing were performed at Annoroad Gene Tech. (Beijing) Co., Ltd (https://www.annoroad.com/). STAR v2.7.9a (Dobin et al. 2013) was used to index the 'NHCC001' genome and align the reads to the reference genome. RSEM v1.3.1 (Li and Dewey 2011) was used to calculate the fragments per kilobase of exon per million reads mapped (FPKM). Principal component analysis (PCA) was applied for quality control using the *R/factoextra* package (https://cloud.r-project.org/package=factoextra/). The *R/DESeq2* package (Love et al. 2014) was adopted for analysis of differential expressed gene (DEG) analysis. Heatmap was generated by the *R/pheatmap* package (Hu 2021).

For qRT-PCR assays, total RNA was extracted using RNA Simple Total RNA Kit (TIANGEN, Beijing, China), and reverse transcribed into cDNA by Hiscript III RT SuperMix for qPCR (+gDNA wiper) (Vazyme, Nanjing, China). The Hieff qPCR SYBR Green Master Mix (No Rox) (Yeasen, Shanghai, China) was used for qRT-PCR with the CFX96 (Bio-rad, USA). The qPCR procedure (95 °C 5 min, 95 °C 10 s, 60 °C 30 s, 40 cycles) with *BcActin* (GeneID: BraC10g029590) as the internal reference gene followed Long et al. (2022) was applied. The $2^{-\Delta\Delta Ct}$ method (Pfaffl 2001) was used to determine the relative gene expression level. Each sample was run with three biological and three technical replicates.

Results

Variations of green leaf color in NHCC

The non-heading Chinese cabbage (NHCC) leaves from the parental lines, Wutacai and Erqing, exhibited typical dark-green and lime-green color, respectively, while their F_1 hybrid showed an intermediate color throughout their lifetime in all three environments (Fig. 1A). To quantify the green leaf color, we analyzed the pigments associated with chlorophyll (chlorophyll a, chlorophyll b,

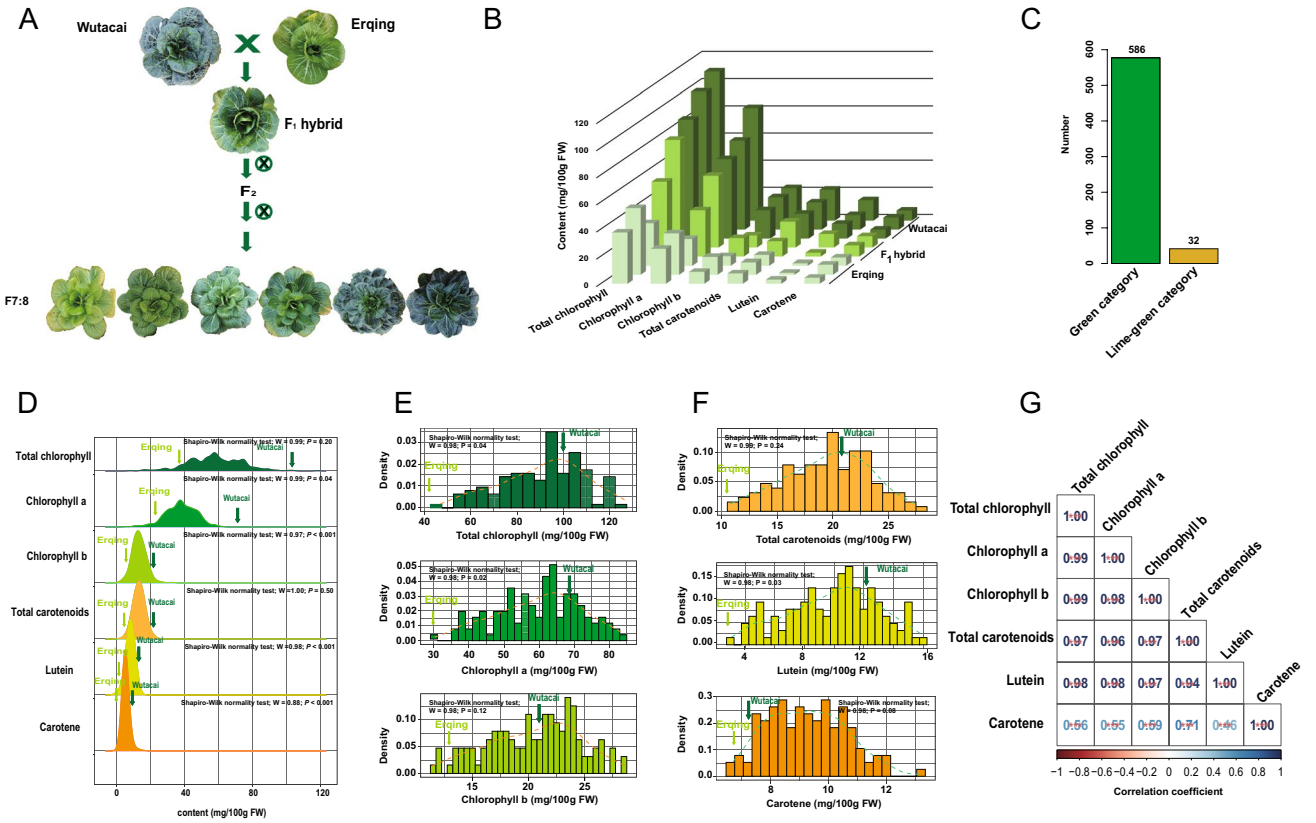

Fig. 1 The performance of green leaf color in the parental lines and RIL population. **A** The diversity of green leaf color in the parental lines, F_1 hybrid and representative lines in the RIL population. **B** The comparison of six pigments (chlorophyll a, chlorophyll b, total chlorophyll, carotene, lutein, and total carotenoids) between Wutacai, Erqing, and F_1 hybrid in three locations (***$P<0.001$). **C** The individual count for green category and lime-green category in F_2 population. **D** The distribution plots of pigments content of the F_2 population. **E** The distribution plots of BLUPs value of chlorophyll content (chlorophyll a, chlorophyll b, and total chlorophyll). **F** The distribution plots of BLUPs value of carotenoids content (lutein, carotene, and total carotenoids). **G** The correlation analysis for BLUP value of six pigments (***$P<0.001$)

and total chlorophyll) and carotenoids (carotene, lutein, and total carotenoids) in the parental lines and their F_1 hybrid collected from the three locations (Baima, Jurong, and Huzhou) (Fig. 1B; Table S2). Notably, the average content of total chlorophyll and total carotenoids in Wutacai (102.68 and 21.19 mg/100 g FW) were approximately 2–3 times higher than those in Erqing (39.22 and 8.15 mg/100 g FW), while F_1 plants (67.39 and 14.32 mg/100 g FW) showed intermediate level across all environments, suggesting a co-dominant or multi-allelic nature of the green leaf color regulation in Wutacai (Fig. 1B; Table S2). Particularly, the ratio of total carotenoids to total chlorophyll was consistently around 20% among the parental lines and F_1 hybrid, indicating that the lighter green hue in Erqing leaves resulted from an overall reduction in total pigments, rather than alterations in individual pigment components, such as increased carotenoids, contributors of orange hues, or decreased chlorophyll, contributors of green hues (Table S2).

We subsequently planted 618 F_2 individuals at Baima Research Station (BRS), revealing a spectrum of leaf colors ranging from yellowish/lime-green to dark-green. Upon visual identification, 32 individuals exhibited leaf color with yellowish tones, while 586 individuals displayed varying degrees of green, aligning with the expected segregation ratio of 15:1 for two recessive genes ($P = 0.27$; Chi-square test) (Fig. 1C). Subsequently, 358 F_2 individuals were randomly selected for pigment content measurement. The means, ranges, standard deviations (SD), skewness, and kurtosis are provided in Table S2.

The F_2 population demonstrated transgressive segregation leaning towards Erqing in terms of pigments content. Specifically, among these individuals, 40 exhibited lower total chlorophyll content, and 25 revealed lower total carotenoids content, compared to those present in Erqing (i.e., 39.22 mg/100 g Fresh weight (FW) of total chlorophyll and 8.15 mg/100 g FW of total carotenoids, respectively) (Fig. 1D; Table S2). Thus, we proposed that the lime-green with yellowish tones may depend on two major recessive genes, while the range of green to dark-green hues might be regulated by multiple loci.

Additionally, we constructed an $F_{7:8}$ RIL population comprising 127 lines which were planted across three locations for phenotypic data collection. In general, all pigment contents followed a normal distribution (Figs. 1E, F and S1). The average correlations (Spearman's rank r_s) among the six pigments showed positive associations ranging from 0.46 to 0.99 ($P < 0.05$), implying a potential shared regulatory network governing these pigments in this population (Figs. 1G and S3). Broad-sense heritability of six pigments ranged between 0.67 and 0.88 (Table 1). ANOVA for the RIL population clearly showed that genotypic effect was the main factor affecting pigment content ($P < 0.001$). In addition, ANOVA indicated that the environmental effect on the phenotypic variance of chlorophyll b and carotene was significant and no interaction effect of genotype-by-environment was detected significantly for all pigments (Table 1).

Genotyping and genetic map construction

The parental lines, Wutacai and Erqing, along with the RIL population, underwent re-sequencing via Novaseq 6000, yielding a total of 0.23 billion reads (~34 Gb data) for parental lines at an average 30× genome coverage. For the 127 lines in RIL population, sequencing generated 5.69 billion reads (~854 Gb data) with an average 10× genome coverage. After proper quality control, approximately 0.22 billion reads for the parental lines and 5.61 billion reads for RIL population retained. Out of 8,329,104 SNP markers were initially identified from the RIL population, 6,053,960 SNP markers passed the quality control (Fig. S4; Table S3).

Further filtration certified 883,635 SNP markers adhering to the correct genotyping patterns ($aa \times bb$) of parent lines (Fig. S4). The mis-genotyped SNP markers were eliminated by employing the Mendelian segregation ratio test with a significance threshold at $P < 0.0025$ with Bonferroni correction to maintain a genome-wide error rate of $\alpha = 0.05$, resulting in retained 454,477 SNP markers. Subsequently,

Table 1 Analysis of variance, variance component and heritability for pigments associated with green leaf color

Variance	df	Total chlorophyll	Chlorophyll a	Chlorophyll b	Total carotenoids	Lutein	Carotene
Genotype (G)	126	296.26*	124.70*	15.70 *	11.60***	8.35**	1.03***
Environment(E)	2	21.69	0	8.26***	0	0.71	1.31**
G *E	362	51.53	32.03	6.98	3.64	1.62	0.72
Block(L:Rep)	8	27.22	8.15	0.49	1.07	0.36	0.4
Residual	1120[a]	210.75	91.04	23.22	10.59	5.06	2.33
Heritability		0.88	0.86	0.76	0.83	0.88	0.67

*$P < 0.05$, **$P < 0.01$, *** $P < 0.001$

[a]The degree of residual for pigments (total chlorophyll, chlorophyll a, chlorophyll b, total carotenoids, lutein, carotene) associated with green leaf color is1007, 1120, 1015, 1011, 1008 and 1013, respectively.

markers with a missing data rate higher than 10% in the RIL population were filtered out (Fig. S4). Following this, 1,865 markers characterized by a threshold of heterozygous individuals (≤ 7) and lower similarity loci (< 99%) within the RIL population, designed as bin-markers and were utilized to construct the genetic map construction (Fig. S4; Tables S4 and S5). These markers collectively spanned 10 linkage groups, representing 10 chromosomes, constructing a genetic map spanning 4222.1 cM with and an average interval of 2.3 cM (Table S5).

QTL mapping for green leaf color in Wutacai × Erqing $F_{7:8}$ RIL population

The content of six pigments collected from three locations and their BLUPs values were subjected to QTL analysis using MQM method. The detected QTL for each leaf color pigment are illustrated in Figs. 2 and S5 and detailed information of each QTL was outlined in Table S6, encompassing peak marker, peak LOD value, peak location, 1.5 LOD interval, additive effect, phenotypic variance explained (R^2) and their physical locations in NHCC draft reference genome. Overall, a total of seven QTL were identified across chromosomes A01, A06, A07, A09, and A10, individually accounting for 3.00–21.91% of the phenotypic variance. Among them, *GlcA07.1* and *GlcA09.1* were recognized as two major-effect QTL that were consistently detected for the six pigments, explaining 4.45–21.91% and 9.30–18.32% of the phenotypic variance across different locations, respectively (Figs. 2 and S5; Table S6). *GlcA07.1* and *GlcA09.1* showed positive additive effect, signifying a contribution to pigment accumulation from the Wutacai allele (Fig. S6). An epistatic interaction between *GlcA07.1* and *GlcA09.1* was also detected for five pigments (except for carotene), elucidating 3.00–4.71% of the phenotypic variance (Fig. S7; Table S6). Conversely, minor-effect QTL such as *GlcA01.1*, *GlcA06.1*, and *GlcA10.1*, were specific to the pigment of carotene under different environments.

Refinement of *GlcA07.1* and *GlcA09.1*

To pinpoint the candidate genes underlying the two major-effect QTL, *GlcA07.1* and *GlcA09.1*, we employed additional polymorphic markers within the candidate regions to genotype the F_2 population to obtain new recombinants. By utilizing the mapped reads from the parental lines, we strategically developed polymorphic InDel markers (> 2 bp difference) distributed evenly across the candidate regions of the *GlcA07.1* and *GlcA09.1* loci (Table S7).

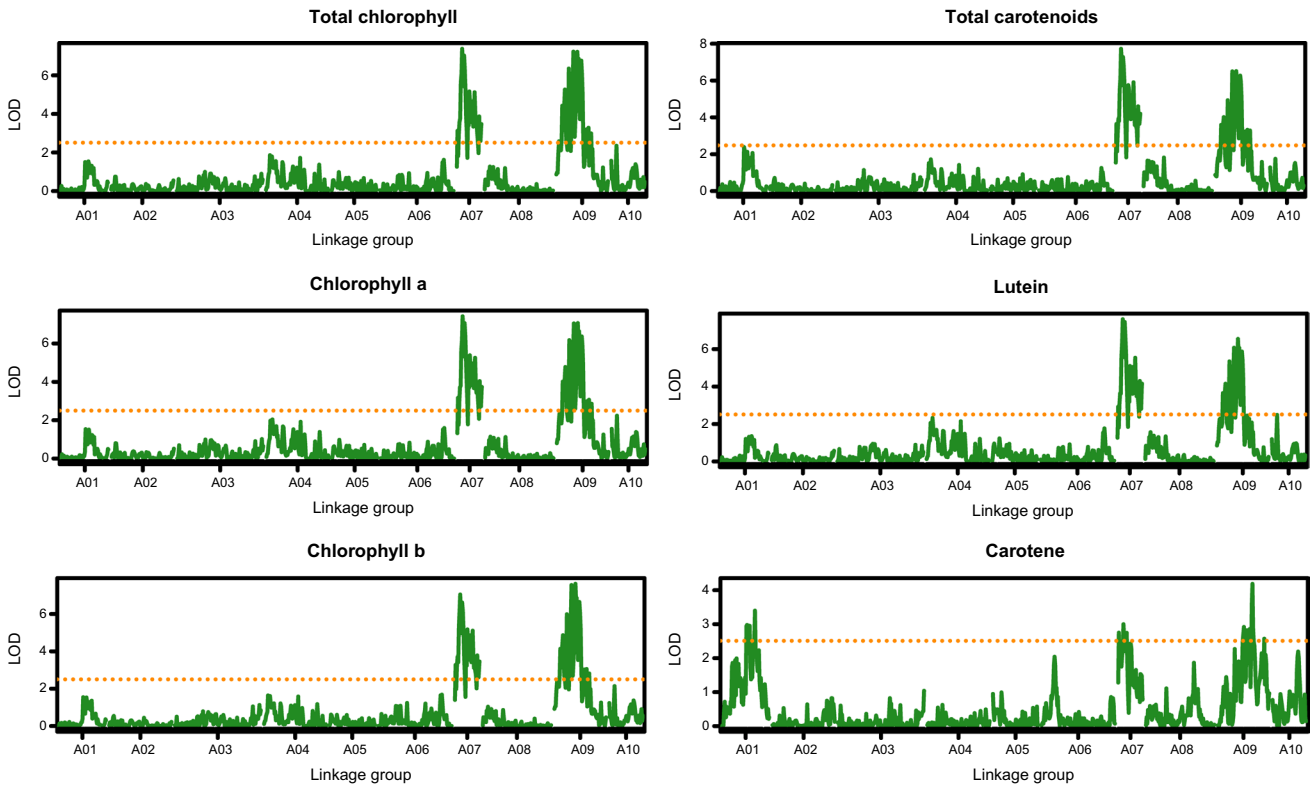

Fig. 2 Whole genome QTL for chlorophyll a, chlorophyll b, total chlorophyll, lutein, carotene, and total carotenoids detected with BLUPs value using MQM in the RIL population

The QTL *GlcA07.1* was identified in an approximately 4.1 Mb genomic region (17.04–21.11 Mb) flanking by SNP markers A07_17044251 and A07_21142380 on chromosome A07 (Fig. 3A). We genotyped a total of 530 F_2 individuals (part of the initial 618 F_2 population) with 14 InDel markers, and identified 49 recombinants grouped into 8 haplotypes (Fig. 3A, B). To assess the genetic effect contributed by *GlcA09.1* among these recombinants, we also genotyped them with the flanking InDel markers A09_53025154 and A09_53028979 to determine their haplotypes at *GlcA09.1* locus (Fig. 3C). The content of the six pigments of these recombinants were measured at the rosette stage (Fig. 3D). After carefully examination, the *GlcA07.1* was delimited into a ~ 170.25 kb genomic region (19.29–19.46 Mb) defined by InDel markers A07_19294439 and A07_19464687 (Fig. 3E). This region harbored 37 annotated candidate genes, among which 13 genes had non-synonymous mutations between the parental lines (Table S9).

The other major-effect QTL *GlcA09.1* was preliminarily mapped to ~ 7.5 Mb genomic region (51.23–58.77 Mb) on chromosome A09, delimited by SNP markers A09_51233595 and A09_58769541 (Fig. 4A). Genotyping 568 F_2 individuals (part of the initial 618 F_2 population) with 20 InDel markers led to the identification of 51 recombinants (Fig. 4B). Concomitantly, we employed InDel markers A07_19294439 and A07_19391005 to genotype the *GlcA07.1* locus among these 51 recombinants (Fig. 4C). After evaluating the content of six pigments from the recombinants and the effect attributed to *GlcA07.1*, we were able to narrow down *GlcA09.1* into ~ 191.41 kb genomic region (52.96–53.15 Mb) delineated by the InDel markers A09_52955063 and A09_53146472 (Fig. 4D, E). The NHCC genome annotation demonstrated 32 candidate genes in this genomic region, with 13 of which contained non-synonymous mutations between the parental lines (Table S10).

Candidate gene analysis

To refine the list of candidates, we performed RNA-seq on mature leaves of the parental lines and F_1 hybrid. The transcriptomic data yielded a total of 58.63 Gb data and 0.39 billion raw reads, with average 90% of raw reads were successfully mapped to the reference genome, indicating the high-quality of the transcriptome dataset (Table S8; Fig. S8). Among 69 annotated candidate genes located in both the *GlcA07.1* and *GlcA09.1* genomic regions, 26 genes were differentially expressed between the parental lines (Fig. 5A, B; Tables S9; and S10). In particular, 14 of 26 genes contained non-synonymous variations between parental lines, which were re-annotated based on the alignment with the *Arabidopsis thaliana* (https://www.arabidopsis.org/) and the Non-heading Chinese Cabbage and Watercress Database (Tables S9 and S10). According to the annotation, *BraC07g023970* and *BraC09g052220* emerged as two promising candidates associated with chlorophyll biosynthesis or chloroplast development (Tables S9 and S10).

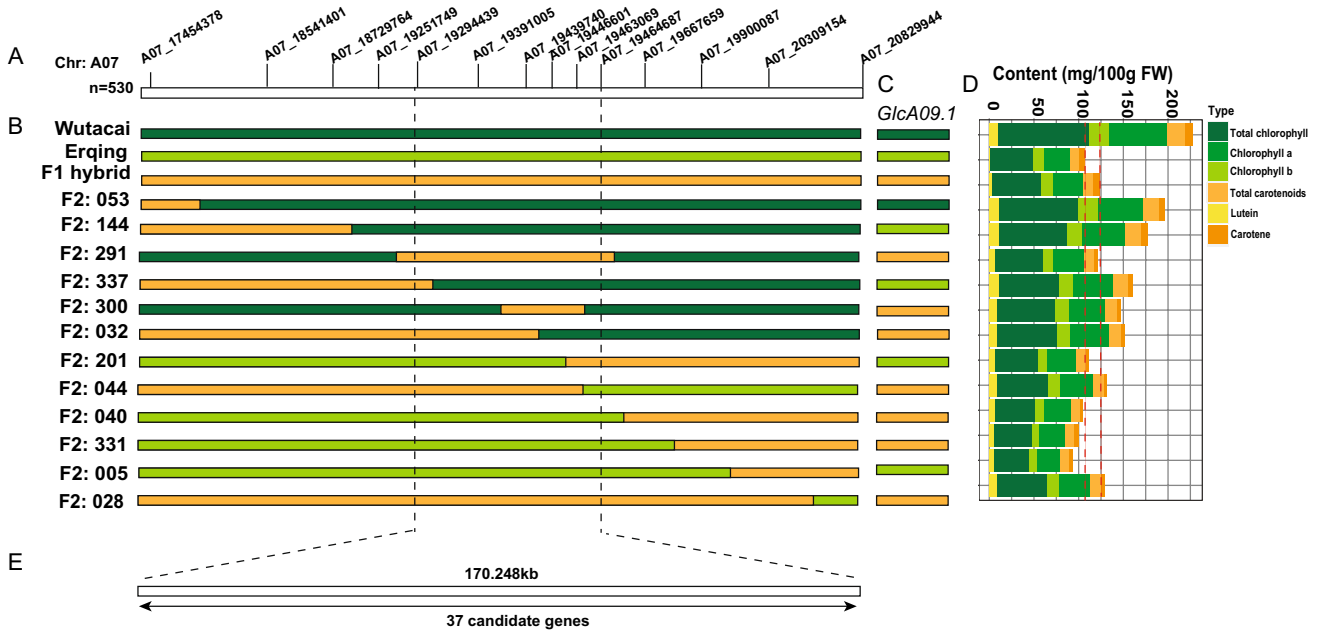

Fig. 3 Refinement of *GlcA07.1* on chromosome A07. **A** The primary physical mapping of employing 530 F_2 individuals. **B** The haplotypes of F_2 individuals with chromosomes segment substitution on target region. The dark-green indicated homozygous segment from Wutacai, lime-green indicated homozygous segment from Erqing and orange indicated the heterozygous region. **C** The genotype of *GlcA09.1*. **D** The phenotypes of recombinants. **E** The finally target genomic region of *GlcA07.1*

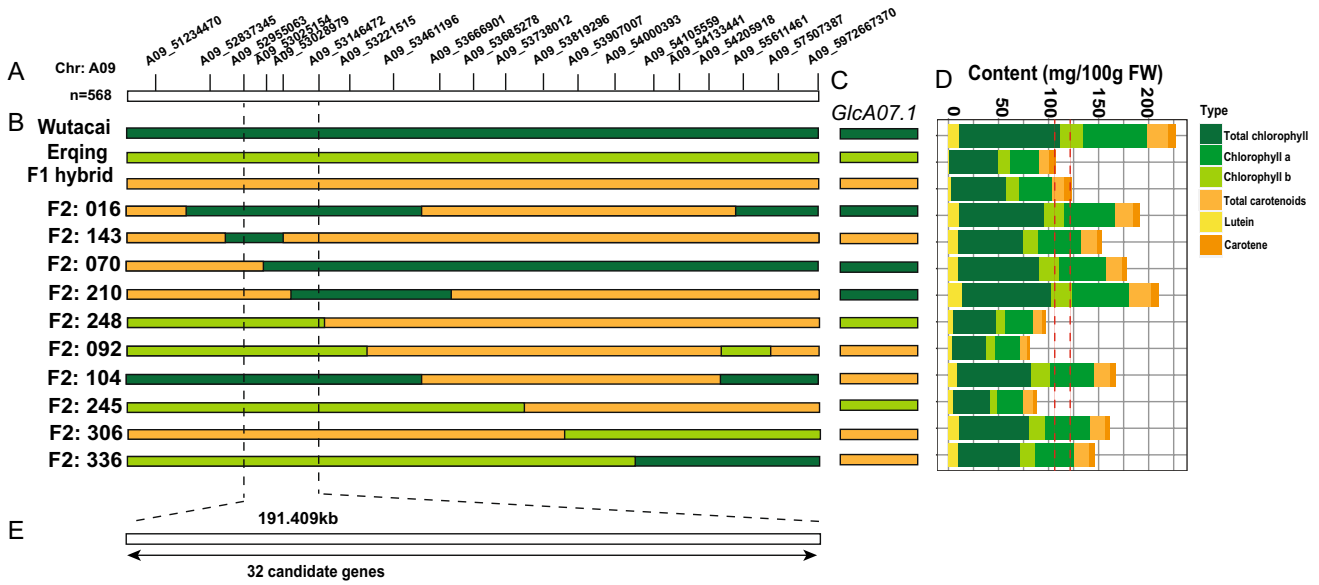

Fig. 4 Refinement of *GlcA09.1* on chromosome A09. **A** The primary physical mapping of employing 568 F_2 individuals. **B** The haplotypes of F_2 individuals with chromosomes segment substitution on target region. The dark-green indicated homozygous segment from Wutacai, lime-green indicated homozygous segment from Erqing and orange indicated the heterozygous region. **C** The genotype of *GlcA07.1*. **D** The phenotypes of recombinants. **E** The finally target genomic region of *GlcA09.1* (color figure online)

BraC07g023970 encodes a rhodanese/cell cycle control phosphatase superfamily protein (BcPPR) and is known to play a role in pigment metabolic process and photosynthesis (Fristedt 2017). We identified a single nucleotide change (A–G) in its tenth exon resulted in an amino acid substitution (I537V) in Erqing (Table S9). Another candidate, *BraC09g052220*, is a homolog of the *SLOW GREEN 1* (*SG1*) gene, which contributes to chloroplast development (Hu et al. 2014), denoted as *BcSG1* accordingly. We identified a 12-bp insertion in the exon, which resulted in a truncated protein with fewer amino acid residues in Erqing (Table S10).

Moreover, we also noticed that there were two genes (*BraC07g023810* and *BraC09g052270*) without sequence variations in exon region between the parental lines underlying QTL *GlcA07.1* and *GlcA09.1*, respectively, but carrying variations in the upstream region gave them the potential to be candidate genes (Tables S9 and S10). *BraC07g023810* is an orthologue of *AT3G59400* (*GENOME UCOUNP 4*, *GUN4*), which is known for its involvement in chlorophyll synthesis and coupling certain nuclear genes to chloroplast (Larkin et al. 2003), designated as *BcGUN4* here after. *BraC09g052270* (*BcGIP1*) encodes a GUN1-interacting protein (GIP1) and is involved in retrograde signaling pathway (Huang et al. 2021).

Quantitative real-time PCR (qRT-PCR) was employed to validate their relative expression levels in Wutacai, Erqing, Dark-green pool, and Lime-green pool. Consistent with RNA-seq findings, *BcGUN4* and *BcPPR* exhibited higher expression levels in Wutacai and Dark-green pool compared to that of in Erqing and Lime-green pool (Fig. 5C, D). Conversely, the expression levels of *BcGIP1* and *BcSG1* were lower in Wutacai and Dark-green pool compared to that of in Erqing and Lime-green pool (Fig. 5E, F). These findings implied that the candidate gene from *GlcA07.1* might play positive role in pigment accumulation and the candidate gene from *GlcA09.1* can be a negative regulator.

The development of molecular markers for marker-assisted selection

To assess the potential applicability of *BcGUN4* and *BcSG1* in leaf color breeding across various populations, we designed two polymorphic InDel markers (*GlcA07.1-BcGUN4* and *GlcA09.1-BcSG1*) targeting the promoter region of *BcGUN4* and the exon region of *BcSG1*, respectively. These markers were then utilized for genotyping 73 NHCC germplasm (Table S11). Our analysis revealed that combined selections for both genes significantly affected the content of five pigments (excluding lutein), resulting in observable leaf color difference (Fig. 6). Conversely, while selection for single gene individually exhibited only minimal impact on pigment levels (Fig. S9). These findings emphasized that concurrent selections for both *BcGUN4* and *BcSG1*, facilitated by markers *GlcA07.1-BcGUN4* and

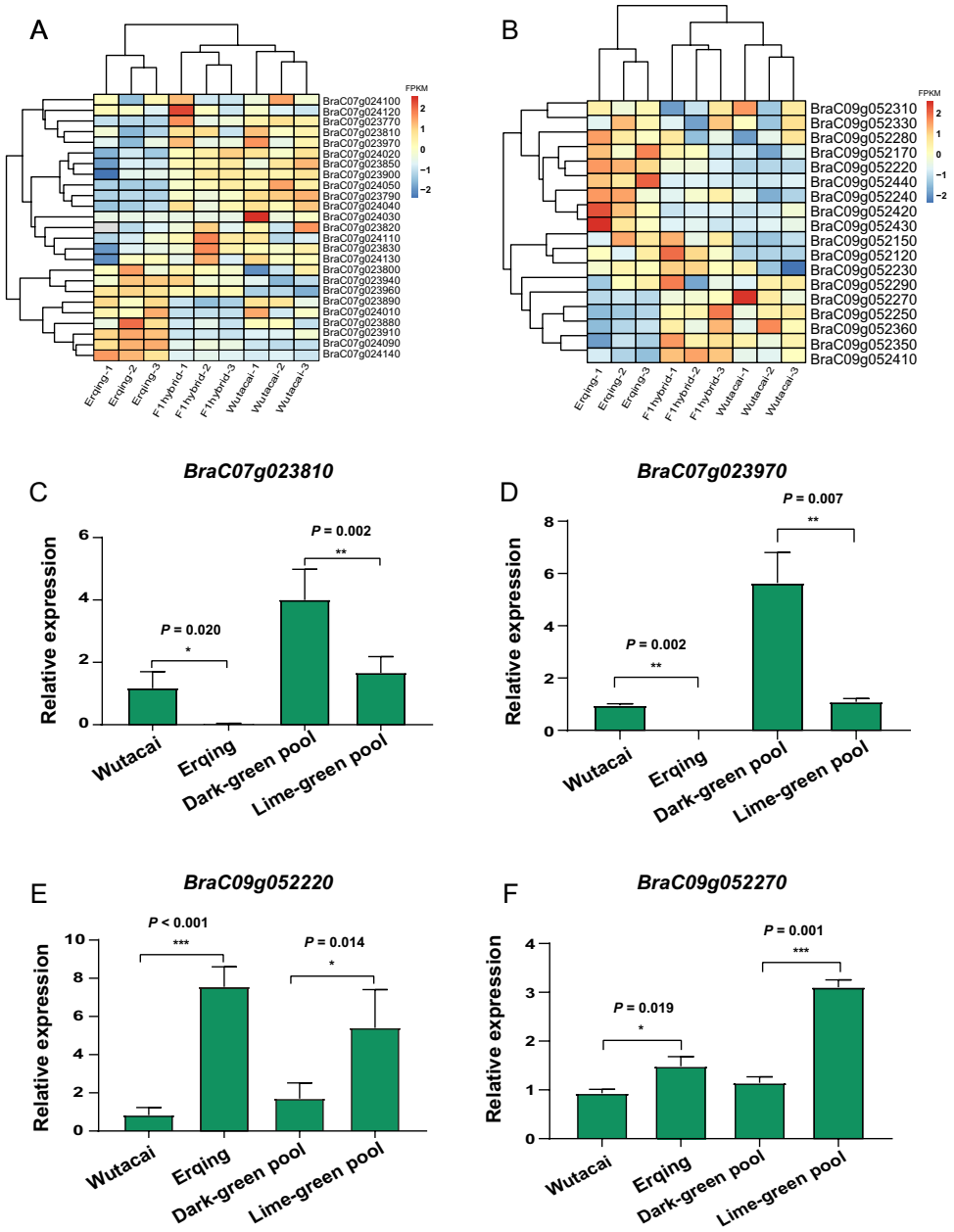

Fig. 5 The comparison for expression level of candidate genes **A** The expression heatmap of candidate genes from *GlcA07.1* between Wutacai, Erqing, and F_1 hybrid. **B** The expression heatmap of candidate genes from *GlcA09.1* between Wutacai, Erqing, and F_1 hybrid. **C** The qRT-PCR analysis of *BraC07g023810* between Wutacai and Erqing, Dark-green pool and Lime-green pool. **D** The qRT-PCR analysis of *BraC07g023970* between Wutacai and Erqing, Dark-green pool and Lime-green pool. **E** The qRT-PCR analysis of *BraC09g052220* between Wutacai and Erqing, Dark-green pool and Lime-green pool. **F** The qRT-PCR analysis of *BraC09g052270* between Wutacai and Erqing, Dark-green pool and Lime-green pool. Values represented the average of three replicates, and the error bar indicated the SD (standard deviation). *, **, and *** indicated the significance at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively (color figure online)

GlcA09.1-BcSG1, significantly affect leaf color variation within NHCC germplasm. This underscores their potential as key targets in leaf color breeding programs for non-heading Chinese cabbage.

Discussion

Green leaf color is an important agronomic trait of Brassica vegetables. It is not only related to the quality and marketability of organ product but also can indicate the photosynthetic ability and yield. It is essential for NHCC to clarify the genetic mechanism of green leaf color, which will provide valuable information for breeding. In our study, a total of seven QTL were detected while applying six pigments including chlorophyll a, chlorophyll b, total chlorophyll, carotene, lutein, and total carotenoids (Table S6). *GlcA07.1* and *GlcA09.1* on chromosomes A07 and A09, respectively, were the two major-effect QTL corresponding to all six pigments in displaying green leaf color. We further narrowed down the intervals of *GlcA07.1* and *GlcA09.1* into approximately 170.25 kb and 191.41 kb genomic regions, which harbored 37 and 32 candidate genes, respectively (Figs. 3 and 4). By examining their gene polymorphisms and differential expression levels between the parental lines, four promising candidate genes with two

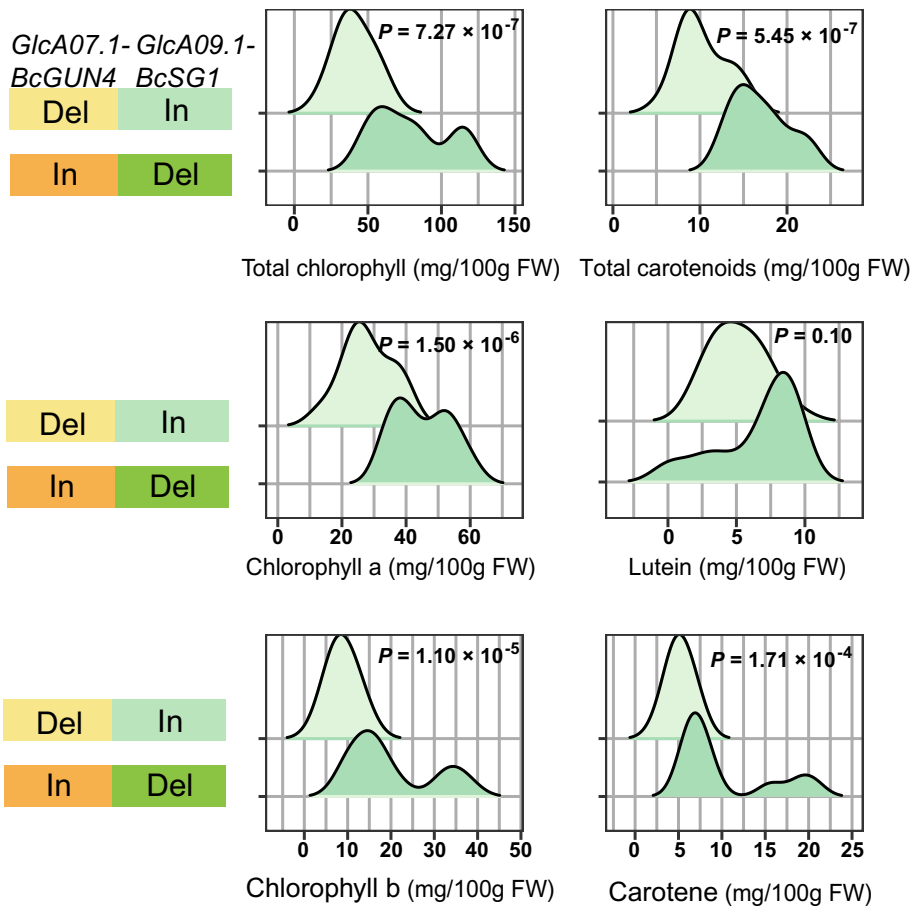

Fig. 6 The molecular markers (*GlcA07.1-BcGUN4* and *GlcA09.1-BcSG1*) is associated with pigments content in non-heading Chinese cabbage

genes underlying *GlcA07.1* (*BraC07g023810*/*BcGUN4* and *BraC07g023970*/*BcPPR*), and another two genes for *GlcA09.1* (*BraC09g052220*/*BcSG1* and *BraC09g052270*/*BcGIP1*), which were associated with chlorophyll biosynthesis and chloroplast development, respectively, were identified (Tables S9 and S10).

The common QTL for leaf color were detected in *Brassica rapa*

The investigation of natural variation in leaf color for *Brassica rapa* has revealed 31 QTL in total (Table S12). The QTL on chromosomes A01, A07, and A09 were consistently reported in two subspecies, *Brassica rapa* L. ssp. *pekinensis* and ssp. *chinensis* (Feng et al. 2012; Choi et al. 2017; Huang et al. 2017; Zhang et al. 2020a, b). In our investigation of genetic inheritance for green leaf color, we utilized an F_2 population derived from the cross between Wutacai and Erqing. Our analysis indicated that lime-green leaf color was determined by two recessive genes, fitting a 15:1 segregation ratio ($P = 0.27$ in Chi-square test) (Fig. 1C). This observation was further affirmed through quantifying pigment composition of green leaf color in the RIL population derived from the same parental lines, uncovering two major-effect QTL (*GlcA07.1* and *GlcA09.1*) detected across all 6 pigments and environments (Fig. 2 and S5). Intriguingly, upon comparing their physical locations to previous studies, *GlcA07.1* (A07: 19.29–19.46 Mb) and *GlcA09.1* (A09: 52.96–53.15 Mb) were found not to co-localize with previously reported QTL on chromosomes A07 and A09 reported by Huang et al. (2017) and Zhang et al. (2020a, b) (Figs. 3 and 4; Table S12). Additionally, the minor-effect QTL *GlcA07.2* (A07: 3.60–11.06 Mb), detected for the pigment lutein, partially overlapped with the QTL previously identified for core leaf color in Chinese cabbage (Tables S6 and S12).

In our evaluation of leaf color among 73 NHCC accessions, we employed two InDel markers for QTL *GlcA07.1* and *GlcA09.1,* resided in the upstream promoter region of *BcGUN4* and the CDS of *BcSG1*, respectively, to conduct the association analysis with six pigments. Our analysis unveiled that the combined selection for both genes significantly affected the content of five pigments (excluding lutein), consequently resulting in dark to light green leaf color difference (Fig. 6). This result implies the existence of potential common regulatory loci mediating

green leaf color in a substantial portion of the NHCC germplasm.

Candidate gene analysis for controlling green leaf color

Upon examining chlorophyll and carotenoid-related pigment contents, a consistent ratio of approximately 20% between total carotenoids and total chlorophyll was observed across the parental lines and F_1 hybrid. This consistency suggested that the lighter green appearance in Erqing leaves resulted from an overall reduction in total pigments, rather than specific alterations in individual pigment component (Table S2; Fig. 1B). This may implied that the regulatory genes governing natural green leaf color resided at an early stage within the pigment formation pathway, positioned prior to the divergence point for either chlorophyll or carotenoid synthesis branches. Furthermore, both biparental populations or natural accessions exhibited the presence of two recessive alleles in regulating the natural variation in green leaf color (Figs. 1C and 6). By examining the candidate region of *GlcA07.1* and *GlcA09.1*, four promising candidates, *BraC07g23810 (BcGUN4)*, *BraC07g023970* (*BcPPR*), *BraC09g052220* (*BcSG1*), and *BraC09g052270* (*BcGIP*), contributed to reverse effect from parental allele in regulating chloroplast-related pathway, thereby affecting both chlorophyll and carotenoid synthesis (Fig. 5; Tables S9 and S10).

The promising candidate gene in QTL *GlcA07.1*, *BraC07g023970*, is orthologous to *AT3G59780* and located in chloroplast involved in pigment metabolic process, but the gene function is still unclear. The specific mechanism is required for further analysis.

The gene *BcGUN4* encodes a tetrapyrrole-binding protein, sharing homology with Arabidopsis *AtGUN4* (82% identity) and Chinese cabbage *BrGUN4* (97.3% identity) (Table S9). Functionally, the GUN4 plays a pivotal role in the chlorophyll biosynthesis pathway by activating the Mg-chelatase enzyme (MgChl), responsible for converting protoporphyrinIX (ProtoIX) to Mg-protoporphyrinIX (Mg-ProtoIX) (Daloso et al. 2014). It functions by composing a complex with CHLH, CHLI, and CHLD, enhancing the MgCh complex activity (Jensen et al. 1996; Willows 2003; Adhikari et al. 2011; Gao et al. 2020). In this study, *BcGUN4* showed higher expression level in Wutacai and Dark-green pool compared to Erqing and Lime-green pool, implying its role in green color accumulation in NHCC (Fig. 5C). A similar complementary study in 'wucai' with yellow core leaves showed significantly down-regulated *GUN4* expression (*Brassica campestris* L.) (Xie et al. 2019).

BraC09g052220 (BcSG1), a candidate underlying QTL *GlcA09.1*, encodes a tetratricopeptide repeat (TPR) protein, a regulator of thylakoid membranes crucial for chloroplast differentiation and expression of chloroplast-related genes during early stage of *Arabidopsis* chloroplast development (Hu et al. 2014). Our previous Transmission Electron Microscope (TEM) examination of mature leaves from Wutacai and Erqing revealed a greater number of grana thylakoids in Wutacai under normal and high light intensity conditions (Liu et al. 2020). In this study, we found that the relative expression level of *SG1* in Wutacai is lower than Erqing, suggesting a potential cause for chloroplast grana failure (Fig. 5E; Table S10). Moreover, both *gun1* and *gun4* mutations in Arabidopsis are able to rescued the *sg1* phenotypes (Chan et al. 2016), yet the interplay between GUN4 and SG1 remains poorly understood, necessitating further exploration for a detailed mechanistic understanding.

Within the QTL *GlcA09.1*, another gene *BraC09g052270* encodes a GUN1-interacting protein (GIP1) designated as *BcGIP1*, functioning within the retrograde signaling pathway. GUN1, a nuclear-encoded plastid protein, plays an essential role in various retrograde signaling processes and acts as a pentatricopeptide repeat protein involved in RNA metabolism (Susek et al 1993; Barkan and Small 2014; Shimizu et al. 2019). Loss of GUN1 can exacerbates the phenotype of *Arabidopsis thaliana* mutant that deficient in chloroplast proteostasis (Tadini et al. 2016; Llamas et al. 2017; Pesaresi and Kim 2019; Lee et al. 2023). GIP1, localized in cytosol and chloroplast, could interacted with GUN1 by different approaches (Huang et al. 2021). GIP1 is found to be abundant in chloroplast with the presence of GUN1 under norflurazon treatment (Huang et al. 2021). Thus, GIP1 potentially influences chloroplast protein functions through its interaction with GUN1, likely participating in pigment regulation. Despite the absence of polymorphism in the coding sequences between the parental lines, its similar expression pattern to *BcSG1* in the parental lines suggests its potential involvement in the regulation of green leaf color mediated by *BcSG1* (Fig. 5D).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-024-04608-x.

Acknowledgements The authors thank Xin Yu and Liping Tang provided valuable advice and revised the manuscript. This work was supported by the Bioinformatics Center of Nanjing Agricultural University.

Author Contribution statement AMB performed majority of the reported work. Ying.L designed and supervised this experiment. HBW, FXZ, TZZ, and LG participated in filed trial, phenotype measurement, and InDel genotyping. Yan.L participated in phenotypic data collection. YHW participated in QTL mapping and data analysis. SHAS, TKL, and XLH participated in manuscript revision. AMB, YHW, and Ying.L wrote the manuscript with input from other coauthors. All authors reviewed and approved this submission.

Funding This work was supported by Jiangsu Seed Industry Revitalization Project (JBGS(2021)015), National Vegetable Industry Technology System (CARS-23-A-16) and the National Natural Science Foundation of China (32172562 and 31872106), the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Data availability The raw data has been deposited in Non-heading Chinese Cabbage and Watercress Database (NHCCP10001) (http://tbir.njau.edu.cn/NhCCDbHubs/DownloadDetail_detail.action?download_fileType=submitdata).

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval These experiments complied with the ethical standards in China.

References

Adams NBP, Bisson C, Brindley AA, Farmer DA, Davison PA, Reid JD, Hunter CN (2020) The active site of magnesium chelatase. Nat Plants 6:1491–1502. https://doi.org/10.1038/s41477-020-00806-9

Adhikari ND, Froehlich JE, Strand DD, Buck SM, Kramer DM, Larkin RM (2011) GUN4-porphyrin complexes bind the ChlH/GUN5 subunit of Mg-Chelatase and promote chlorophyll biosynthesis in *Arabidopsis*. Plant Cell 23:1449–1467. https://doi.org/10.1105/tpc.110.082503

Amagai Y, Lu N, Hayashi E, Takagaki M, Kikuchi M, Ibaraki Y, Kozai T (2022) External green light as a new tool to change colors and nutritional components of inner leaves of head cabbages. Food Meas 16:269–280. https://doi.org/10.1007/s11694-021-01150-y

Arends D, Prins P, Jansen RC, Broman KW (2010) *R/qtl*: high-throughput multiple QTL mapping. Bioinformatic 26:2990–2992. https://doi.org/10.1093/bioinformatics/btq565

Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol 65:415–442. https://doi.org/10.1093/plphys/kiad251

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01

Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA (2019) *R/qtl2*: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics 211:495–502. https://doi.org/10.1534/genetics.118.301595

Broman KW, Wu H, Sen Ś, Churchill GA (2003) *R/qtl*: QTL mapping in experimental crosses. Bioinformatics 19:889–890. https://doi.org/10.1093/bioinformatics/btg112

Chan KX, Phua SY, Crisp P, McQuinn R, Pogson BJ (2016) Learning the languages of the chloroplast: retrograde signaling and beyond. Annu Rev Plant Biol 67:25–53. https://doi.org/10.1146/annurev-arplant-043015-111854

Choi SR, Yu XN, Dhandapani V, Li XN, Wang Z, Lee SY, Oh SH, Pang WX, Ramchiary N, Hong CP, Park SY, Piao Z, Kim HR, Lim YP (2017) Integrated analysis of leaf morphological and color traits in different populations of Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Theor Appl Genet 130:1617–1634. https://doi.org/10.1007/s00122-017-2914-4

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Xiangyi L, Ruden DM (2014) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; *iso-2*; *iso-3*. Fly 6(2):80–92. https://doi.org/10.4161/fly.19695

Daloso DM, Antunes WC, Santana TA, Pinheiro DP, Ribas RF, Martins GS, Loureiro ME (2014) Arabidopsis *gun4* mutant have greater light energy transfer efficiency in photosystem II despite low chlorophyll content. Theor Exp Plant Physiol 26:177–187. https://doi.org/10.1007/s40626-014-0025-z

Dai X, He C, Zhou L, Liang M, Fu X, Qin P, Yang Y, Chen L (2018) Identification of a specific molecular marker for the rice blast-resistant gene *Pigm* and molecular breeding of thermo-sensitive genic male sterile leaf-color marker lines. Mol Breed 38:72. https://doi.org/10.1007/s11032-018-0821-2

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Eckhardt U, Grimm B, Hörtensteiner S (2004) Recent advances in chlorophyll biosynthesis and breakdown in higher plants. Plant Mol Biol 56:1–14. https://doi.org/10.1007/s11103-004-2331-3

Feng H, Li Y, Liu ZY, Liu J (2012) Mapping of *or*, a gene conferring orange color on the inner leaf of the Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). Mol Breeding 29:235–244. https://doi.org/10.1007/s11032-010-9542-x

Finn RD, Clements J, Eddy SR (2011) MMER web server: interactive sequence similarity searching. Nucleic Acids Res 392:W29–W37. https://doi.org/10.1093/nar/gkr367

Fu W, Ye X, Ren J, Li QQ, Du JT, Hou AL, Mei FB, Feng H, Liu ZY (2019) Fine mapping of *lcm1*, a gene conferring chlorophyll-deficient golden leaf in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Mol Breed 39:25. https://doi.org/10.1007/s11032-019-0945-z

Fristedt R (2017) Chloroplast function revealed through analysis of *GreenCut2* genes. J Exp Bot 68:2111–2120. https://doi.org/10.1093/jxb/erx082

Gao YS, Wang YL, Wang X, Liu L (2020) Hexameric structure of the ATPase motor subunit of magnesium chelatase in chlorophyll biosynthesis. Protein Sci 29:1040–1046. https://doi.org/10.1002/pro.3816

Hiriart JB, Lehto K, Tyystjärvi E, Junttila T, Aro EM (2002) Suppression of a key gene involved in chlorophyll biosynthesis by means of virus-inducing gene silencing. Plant Mol Biol 50:213–224

Hotta Y, Tanaka T, Takaoka H, Takeuchi Y, Konnai M (1997) New physiological effects of 5-aminolevulinic acid in plants: The increase of photosynthesis, chlorophyll content, and plant growth. Biosci Biotech Bioch 61:2025–2028. https://doi.org/10.1271/bbb.61.2025

Hu K (2021) Become Competent in generating RNA-Seq heat maps in one day for novices without prior R experience. Methods Mol Biol 2239:269–303. https://doi.org/10.1007/978-1-0716-1084-8_17

Hu Z, Xu F, Guan LP, Qian PP, Liu YQ, Zhang HF, Huang Y, Hou SW (2014) The tetratricopeptide repeat-containing protein *slow green1* is required for chloroplast development in *Arabidopsis*. J Exp Bot 65:1111–1123. https://doi.org/10.1093/jxb/ert463

Huang XQ, Wang LJ, Kong MJ, Huang N, Liu XY, Liang HY, Zhang JX, Lu S (2021) *At3g53630* encodes a GUN1-interacting protein under norflurazon treatment. Protoplasma 258:371–378. https://doi.org/10.1007/s00709-020-01578-x

Huang L, Yang YF, Zhang F, Cao JS (2017) A genome-wide SNP-based genetic map and QTL mapping for agronomic traits in Chinese cabbage. Sci Rep 7:46305. https://doi.org/10.1038/srep46305

Jensen PE, Gibson LCD, Henningsen KW, Hunter CN (1996) Expression of the *chlI*, *chlD*, and *chlH* genes from the *Cyanobacterium synechocystis* PCC6803 in *Escherichia coli* and demonstration that the three cognate proteins are required for magnesium-protoporphyrin chelatase activity. J Biol Chem 271:16662–16667. https://doi.org/10.1074/jbc.271.28.16662

Kumar AM, Soll D (2000) Antisense *HEMA1* RNA expression inhibits heme and chlorophyll biosynthesis in Arabidopsis. Plant Physiol 122:49–55. https://doi.org/10.1104/pp.122.1.49

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental

and natural populations. Genomics 1:174–181. https://doi.org/10.1016/0888-7543(87)90010-3
Larkin RM, Alonso JM, Ecker JR, Chory J (2003) GUN4, a regulator of chlorophyll synthesis and intracellular signaling. Science 299:902–906. https://doi.org/10.1126/science.1079978
Lee KP, Li M, Li M, Liu K, Medina-Puche L, Qi S, Cui C, Lozano-Duran R, Kim C (2023) Hierarchical regulatory module GENOMES UNCOUPLED1-GOLDEN2-LIKE1/2-WRKY18/40 modulates salicylic acid signaling. Plant Physiol 192(4):3120–3133. https://doi.org/10.1093/plphys/kiad251
Lee S, Joung YH, Kim JK, Choi YD, Jang G (2019) An isoform of the plastid RNA polymerase-associated protein FSD3 negatively regulates chloroplast development. BMC Plant Biol 19:524. https://doi.org/10.1186/s12870-019-2128-9
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323. https://doi.org/10.1186/1471-2105-12-323
Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698
Li RQ, Jiang M, Huang JZ, Møller IM, Shu QY (2021) Mutations of the *genomes uncoupled 4* gene cause ROS accumulation and repress expression of peroxidase genes in rice. Front Plant Sci 12:682453. https://doi.org/10.3389/fpls.2021.682453
Li X, He Y, Yang J, Jia YH, Zeng HL (2018) Gene mapping and transcriptome profiling of a practical photo-thermo-sensitive rice male sterile line with seedling-specific green-revertible albino leaf. Plant Sci 266:37–45. https://doi.org/10.1016/j.plantsci.2017.10.010
Li Y, Liu GF, Ma LM, Liu TK, Zhang CW, Xiao D, Zheng HK, Chen F, Hou XL (2020) A chromosome-level reference genome of non-heading Chinese cabbage [*Brassica campestris* (syn. *Brassica rapa*) ssp. *chinensis*]. Hortic Res 7:212. https://doi.org/10.1038/s41438-020-00449-z
Liu TK, Duan WK, Chen ZW, Yuan JP, Xiao D, Hou XL, Li Y (2020) Enhanced photosynthetic activity in pak choi hybrids is associated with increased grana thylakoids in chloroplasts. Plant J 103:2211–2224. https://doi.org/10.1111/tpj.14893
Li ZD, Li Y, Liu TK, Zhang CW, Xiao D, Hou XL (2022) Non-Heading Chinese cabbage database: an open-access platform for the genomics of *Brassica campestris* (syn. *Brassica rapa*) ssp. *chinensis*. Plants 11:1005. https://doi.org/10.3390/plants11081005
Llamas E, Pulido P, Rodriguez-Concepcion M (2017) Interference with plastome gene expression and Clp protease activity in Arabidopsis triggers a chloroplast unfolded protein response to restore protein homeostasis. PLoS Genet 13:e1007022. https://doi.org/10.1371/journal.pgen.1007022
Long Y, Zhao TZ, Xu LL, Zhang W, Huang FY, Wang JJ, Hou XL, Li Y (2022) *BcBRC1a* is a negative regulator for tillering in non-heading Chinese cabbage. Veg Res 2:11. https://doi.org/10.1016/j.plantsci.2021.110934
Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
Luo P, Shi C, Zhou Yi, Zhou J, Zhang X, Wang Y, Yang Y, Peng X, Xie T, Tang X (2023) The nuclear-localized RNA helicase 13 is essential for chloroplast development in *Arabidopsis thaliana*. J Exp Bot 74(17):5057–5071. https://doi.org/10.1093/jxb/erad225
Masuda T (2008) Recent overview of the Mg branch of the tetrapyrrole biosynthesis leading to chlorophylls. Photosynth Res 96:121–143. https://doi.org/10.1007/s11120-008-9291-4
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, Depristo MA (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110
Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45
Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A (2021) Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913
Murdoch DJ, Chow ED (1996) A graphical display of large correlation matrices. Am Stat 50:178–180. https://doi.org/10.2307/2684435
Nelson N, Yocum CF (2006) Structure and function of photosystems I and II. Annu Rev Plant Biol 57:521–565. https://doi.org/10.1146/annurev.arplant.57.032905.105350
Pesaresi P, Kim C (2019) Current understanding of *GUN1*: a key mediator involved in biogenic retrograde signaling. Plant Cell Rep 38:819–823. https://doi.org/10.1007/s00299-019-02383-4
Pogson BJ, Albrecht V (2011) Genetic dissection of chloroplast biogenesis and development: an overview. Plant Physiol 4:1545–1551. https://doi.org/10.1104/pp.110.170365
Priyanka S, Harsita N, Anamika B, Anurag B, Purabi B (2023) Nitrogen management by small farmers with the use of leaf color chart: a review. J Plant Nutr 46:1836–1844. https://doi.org/10.1080/01904167.2022.2144370
Schlicke H, Hartwig AS, Firtzlaff V, Richter AS, Glasser C, Maier K, Finkemeier I, Grimm B (2014) Induced deactivation of genes encoding chlorophyll biosynthesis enzymes disentangles tetrapyrrole-mediated retrograde signaling. Mol Plant 7:1211–1227. https://doi.org/10.1093/mp/ssu034
Shen JZ, Zou CW, Zhang XZ, Zhou L, Wang YH, Fang WP, Zhu XJ (2018) Metabolic analyses reveal different mechanisms of leaf color change in two purple-leaf tea plant (*Camellia sinensis* L.) cultivars. Hortic Res 5:7. https://doi.org/10.1038/s41438-017-0010-1
Shimizu T, Kacprzak SM, Mochizuki N, Masuda T (2019) The retrograde signaling protein GUN1 regulates tetrapyrrole biosynthesis. Proc Natl Acad Sci USA 116:24900–24906. https://doi.org/10.1073/pnas.1911251116
Su X, Yue X, Kong M, Xie Z, Yan J, Ma W, Wang Y, Zhao J, Zhang X, Liu M (2023) Leaf color classification and expression analysis of photosynthesis-related genes in inbred lines of Chinese cabbage displaying minor variations in dark-green leaves. Plants 12:2124. https://doi.org/10.3390/plants12112124
Susek RE, Ausubel FM, Chory J (1993) Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development. Cell 74:787–799. https://doi.org/10.1016/0092-8674(93)90459-4
Sun T, Li L (2020) Toward the 'golden' era: the status in uncovering the regulatory control of carotenoid accumulation in plants. Plant Sci 290:110331. https://doi.org/10.1016/j.plantsci.2019.110331
Sun Y, Tian Y, Cheng S, Wang Y, Hao Y, Zhu J, Zhu X, Zhang Y, Yu M, Lei J, Bao X, Wu H, Wang Y, Wan J (2019) WSL6 encoding an Era-type GTP-binding protein is essential for chloroplast development in rice. Plant Mol Biol 100:635–645. https://doi.org/10.1007/s11103-019-00885-z
Tadini L, Pesaresi P, Kleine T, Rossi F, Guljamow A, Sommer F, Mühlhaus T, Schroda M, Masiero S, Pribil M, Rothbart M, Hedtke B,

Grimm B, Leister D (2016) GUN1 controls accumulation of the plastid ribosomal protein s1 at the protein level and interacts with proteins involved in plastid protein homeostasis. Plant Physiol 170:1817–1830. https://doi.org/10.1104/pp.15.02033

Tadini L, Jeran N, Pesaresi P (2020) GUN1 and plastid RNA metabolism: learning from genetics. Cells 9:2307. https://doi.org/10.3390/cells9102307

Tang Y, Fang Z, Liu M, Zhao D, Tao J (2020) Color characteristics, pigment accumulation and biosynthetic analyses of leaf color variation in herbaceous peony (*Paeonia lactiflora* Pall.). 3 Biotech 10:76. https://doi.org/10.1007/s13205-020-2063-3

Taylor J, Butler D (2017) R Package *ASMap*: efficient genetic linkage map construction and diagnosis. J Stat Softw 6:1–29. https://doi.org/10.18637/jss.v079.i06

Wang P, Richter AS, Kleeberg JRW, Geimer S, Grimm B (2020) Post-translational coordination of chlorophyll biosynthesis and breakdown by BCMs maintains chlorophyll homeostasis during leaf development. Nat Commun 11:1254. https://doi.org/10.1038/s41467-020-14992-9

Wang Z, Dhakal S, Cerit M, Wang S, Rauf Y, Yu S, Maulana F, Huang W, Anderson JD, Ma XF, Rudd JC, Ibrahim AM, Xue Q, Hays DB, Bernardo A, Amand PS, Bai G, Baker J, Baker S, Liu S (2022) QTL mapping of yield components and kernel traits in wheat cultivars TAM 112 and duster. Front Plant Sci 13:1057701. https://doi.org/10.3389/fpls.2022.1057701

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the *tidyverse*. J Open Source Softw 43:1686. https://doi.org/10.21105/joss.01686

Willows RD (2003) Biosynthesis of chlorophylls from protoporphyrin IX. Nat Prod Rep 20:327–341. https://doi.org/10.1039/B110549N

Xie SL, Nie LB, Zheng YS, Wang J, Zhao MR, Zhu SD, Hou LF, Chen GH, Wang CG, Yuan LY (2019) Comparative proteomic analysis reveals that chlorophyll metabolism contributes to leaf color changes in Wucai (*Brassica campestris* L.) responding to cold acclimation. J Proteome Res 18:2478–2492. https://doi.org/10.1021/acs.jproteome.9b00016

Xu MY, Wei YH, Zeng JY (2013) Comparative analysis on the content of total carotenoids in the leaves of three general plants. J Guanxi Norm Uni Nat 30:18–19 (**in Chinese**)

Yang YX, Chen XX, Xu B, Li YX, Ma YH, Wang GD (2015) Phenotype and transcriptome analysis reveals chloroplast development and pigment biosynthesis together influenced the leaf color formation in mutants of *Anthurium andraeanum* 'Sonate.' Front Plant Sci 6:139. https://doi.org/10.3389/fpls.2015

Zhang H, Li J, Yoo JH, Yoo SC, Cho SH, Koh HJ, Seo HS, Paek- NC (2006) Rice Chlorina-1 and Chlorina-9 encode ChlD and ChlI subunits of Mg-chelatase, a key enzyme for chlorophyll synthesis and chloroplast development. Plant Mol Biol 62:325–337. https://doi.org/10.1007/s11103-006-9024-z

Zhang K, Mu Y, Li WJ, Shan XF, Wang N, Feng H (2020) Identification of two recessive etiolation genes (*py1, py2*) in pakchoi (*Brassica rapa* L. ssp. chinensis). BMC Plant Biol 20:68. https://doi.org/10.1186/s12870-020-2271-3

Zhang Q, Wang YL, Shen L, Ren D, Hu J, Zhu L, Zhang GH, Guo LB, Zeng DL, Qian Q (2020b) *OsCRS2* encoding a peptidyl-tRNA hydrolase protein is essential for chloroplast development in rice. Plant Growth Regul 92:535–545. https://doi.org/10.1007/s10725-020-00655-8

Zhao Y, Huang SG, Zhang MD, Zhang Y, Feng H (2021) Mapping of a pale green mutant gene and its functional verification by allelic mutations in Chinese cabbage (*Brassica rapa* L. ssp *pekinensis*). Front Plant Sci 12:699308. https://doi.org/10.3389/fpls.202

Zhao Y, Huang S, Wang N, Zhang Y, Ren J, Zhao Y, Feng H (2022) Identification of a biomass unaffected pale green mutant gene in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). Sci Rep 12:7731. https://doi.org/10.1038/s41598-022-11825-1

Zhao Y, Xu W, Zhang Y, Sun S, Wang L, Zhong S, Zhao X, Liu B (2021b) PPR647 protein is required for chloroplast RNA editing, splicing and chloroplast development in maize. Int J Mol Sci 22:11162. https://doi.org/10.3390/ijms222011162

Zhao Y, Gao R, Zhao Z, Hu S, Han R, Jeyaraj A, Arkorful E, Li X, Chen X (2022) Genome-wide identification of RNA editing sites in chloroplast transcripts and multiple organellar RNA editing factors in tea plant (*Camellia sinensis* L.): insights into the albinism mechanism of tea leaves. Gene 848:146898. https://doi.org/10.1016/j.gene.2022.146898

Zhou K, Zhang C, Xia J, Yun P, Wang Y, Ma T, Li Z (2021) Albino seedling lethality 4; chloroplast 30S ribosomal protein S1 is required for chloroplast ribosome biogenesis and early chloroplast development in rice. Rice 14:47. https://doi.org/10.1186/s12284-021-00491-y

Zhou SX, Sawicki A, Willows RD, Luo MZ (2012) C-terminal residues of *oryza sativa* GUN4 are required for the activation of the ChlH subunit of magnesium chelatase in chlorophyll synthesis. FEBS Lett 586:205–210. https://doi.org/10.1016/j.febslet.2011.12.026