





## Article

# Gene Mapping of a Yellow-to-Lethal Mutation Based on Bulk-Segregant Analysis-Seq in Soybean

Yaqi Wang <sup>1,2</sup> , Fangguo Chang <sup>3,4</sup>, G M Al Amin <sup>5</sup> , Shuguang Li <sup>1</sup>, Mengmeng Fu <sup>1</sup> , Xiwen Yu <sup>1</sup>, Zhixin Zhao <sup>1</sup>, Haifeng Xu <sup>1,\*</sup> and Tuanjie Zhao <sup>2,\*</sup> 

<sup>1</sup> Huai'an Key Laboratory for Agricultural Biotechnology, Key Laboratory of Germplasm Innovation in Lower Reaches of the Huaihe River, Ministry of Agriculture and Rural Affairs, Huaiyin Institute of Agricultural Sciences of Xuhuai Region in Jiangsu, Huai'an 223001, China; yqwang\_01@126.com (Y.W.); dawn0524@126.com (S.L.)

<sup>2</sup> Soybean Research Institute, National Center for Soybean Improvement (Nanjing), Key Laboratory for Biology and Genetic Improvement of Soybean (General), Ministry of Agriculture and Rural Affairs, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China

<sup>3</sup> Agronomy College, Gansu Agricultural University, Lanzhou 730070, China

<sup>4</sup> State Key Laboratory of Aridland Crop Science, Gansu Agricultural University, Lanzhou 730070, China

<sup>5</sup> Department of Botany, Jagannath University, Dhaka 1100, Bangladesh; alamin@bot.jnu.ac.bd

\* Correspondence: hanksxf@163.com (H.X.); tjzhao@njau.edu.cn (T.Z.)

**Abstract:** Plant photosynthesis is mainly dependent on leaf color, and this has an impact on yield. Mutants lacking in chlorophyll have been analyzed to gain insight into the genetic processes involved in photosynthesis, chloroplast development, and chlorophyll metabolism. A yellow-to-lethal mutant, *yt1*, was selected from the M<sub>6</sub> generation of the <sup>60</sup>Co γ ray irradiation-treated soybean cultivar Nannong 1138-2. The mutant exhibited reduced chlorophyll content, with the thylakoid structure disrupted. Segregation of the cross between Williams 82 (W82) and *yt1* indicated that a recessive allele controlled yellow-to-lethal traits. The bulked-segregant analysis (BSA)-Seq method performed preliminary mapping, followed by simple sequence repeat (SSR) marker validation and further mapping. The candidate gene was mapped to a 418 Kb region containing 53 genes. High-throughput sequencing and first-generation sequencing results showed a two bp deletion in the second exon of *Glyma.08g106500*, leading to a frameshift mutation in *yt1*. As a promising candidate gene, *Glyma.08g106500* encoded a chloroplast-localized pentatricopeptide repeat (PPR) domain-containing protein involved in the assembly of chloroplast proteins. These results will contribute to cloning the mutant *yt1* gene and provide insight into the regulatory processes controlling photosynthesis and chloroplast development and growth in soybean.

**Keywords:** yellow-to-lethal mutant; BSA-Seq; gene mapping; simple sequence repeats (SSR)



**Citation:** Wang, Y.; Chang, F.; Al Amin, G.M.; Li, S.; Fu, M.; Yu, X.; Zhao, Z.; Xu, H.; Zhao, T. Gene Mapping of a Yellow-to-Lethal Mutation Based on Bulk-Segregant Analysis-Seq in Soybean. *Agronomy* **2024**, *14*, 185. <https://doi.org/10.3390/agronomy14010185>

Academic Editor: HongWei Cai

Received: 9 December 2023

Revised: 9 January 2024

Accepted: 10 January 2024

Published: 15 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Leaves play a crucial role in photosynthesis and significantly contribute to the overall productivity of green plants. Leaves are responsible for exchanging gases, allowing plants to take in carbon dioxide and release oxygen into the atmosphere. Without leaves, photosynthesis would not occur, and the survival of green plants would be compromised. The insufficiency of chlorophyll in plants leads to a decline in their photosynthetic efficacy, causing a decrease in overall production. In recent years, extensive research on the genome has led to the identification of chlorophyll-deficient mutants, which have proven to be crucial in identifying genes related to chlorophyll metabolism. These mutants have provided valuable insights into the molecular mechanisms involved in chlorophyll synthesis and degradation. By studying these genes, the genetic variations help to develop strategies to enhance chlorophyll production in plants, ultimately improving their photosynthetic

efficiency and overall productivity [1–7]. Several investigations have demonstrated that specific mutants lacking chlorophyll exhibit a gain in photosynthetic efficiency rather than a decrease [8,9]. These mutants are crucial to elucidating the metabolic pathways underlying chlorophyll photosynthesis.

Chlorophyll-deficient mutants can be divided into nonlethal and lethal groups according to their viability in the field. In recent years, more and more chlorophyll-deficient mutants have been found in soybean, including over ten nuclear gene-lethal mutants [10]. Mostly, the leaves of the lethal mutant are bright yellow and die quickly in the field. The Y<sub>11</sub> mutant is the earliest reported lethal mutation [11]. The segregating lines exhibited incomplete dominance: 1 Y<sub>11</sub>Y<sub>11</sub> (normal green); 2 Y<sub>11</sub>y<sub>11</sub> (light green); 1 y<sub>11</sub>y<sub>11</sub> (yellow, lethal). Further research has shown that a single base mutation in the *Glyma13g30560* gene, which codes for the Mg-chelatase subunit (ChlI1a), causes a phenotype that ranges from yellow to lethal [12]. However, a mutation in ChlI1b, a paralog of ChlI1a, led to a nonlethal mutant CD-5. Noble et al. [13] reported a soybean mutant called lethal yellow (LY), which can survive until the pod-setting stage if given an appropriate carbon source. The researchers believe that the difference in the content of chlorophyll levels in LY is due to the rapid degradation of chlorophyll rather than an inability to synthesize it. Reed et al. [14] revealed a recessive allele controlling leaf color by identifying a lethal mutant from the cultivar “BSR 101” in soybean. Sandhu et al. [15] found a yellow chlorophyll-deficient mutant in soybeans that was lethal because it could not make grana in the chloroplast and died three weeks after sprouting.

BSA-Seq is an efficient way to find causal loci in plants that combines the BSA method with a high-throughput sequencing technique. It has shown significant benefits in QTL/gene mapping [16,17]. Populations used in BSA include F<sub>2</sub>, F<sub>2,3</sub>, BC<sub>1</sub>, recombinant inbred lines (RILs), doubled haploids (DHs), nested association mapping (NAM) [18], multi-parent advanced generation inter-cross (MAGIC) [19], and natural populations [20]. Data collected from next-generation sequencing at all DNA, RNA, and protein levels are used for genetic polymorphism [21]. DNA bulks of individuals with significant phenotypic differences for the target trait are used for whole genome resequencing. After the analysis of the ΔSNP index between the data of the two gene bulks, the initial mapping intervals can be calculated. Then, the traditional molecular markers, SSR and InDel markers, validate the BSA-Seq results and narrow down the causal region. This method has been widely used to mine candidate genes for target traits in *Arabidopsis* [16], rice [22–24], cotton [25], maize [26], wheat [27], sunflower [28], soybean [29], and peanut [30].

This study identified a yellow-to-lethal mutant (*yt1*) from Nannong 1138-2 (NN1138-2) treated with <sup>60</sup>Co γ ray irradiation. Then, it used the BSA-Seq method combined with SSR markers to map the target gene to a 418 Kb interval on chromosome 8. Sequencing results and functional annotation showed *Glyma.08g106500* was predicted to be the candidate gene. This study lays the groundwork for cloning genes related to chlorophyll mechanisms or chloroplast development pathways to enhance photosynthetic efficiency in soybean varieties.

## 2. Materials and Methods

### 2.1. Plant Materials and Genetic Analysis

Soybean accessions NN1138-2 underwent irradiation with a 150 Gy dose of <sup>60</sup>Co γ ray irradiation in 2011. The M<sub>1</sub> plants were harvested and pooled, and the M<sub>2</sub> generation began to be harvested individually. In 2016, a yellow-to-lethal mutant (*yt1*) was identified from the segregating lines of the M<sub>6</sub> generation. Through consecutive selfing and selection, we achieved residual heterozygous lines with a relatively pure background except for the causal locus.

Due to the inability of the mutant to generate seeds, we used the plants of a residual heterozygous line as the male parent to cross with Williams 82 (W82) to construct F<sub>2</sub> populations. The green leaf individuals of the segregating F<sub>2</sub> population were harvested to generate F<sub>2,3</sub> and F<sub>3,4</sub> populations. For genetic analysis, the segregation ratio of alleles was

evaluated based on the chi-square ( $\chi^2$ ) test [31]. Due to the easy discrimination of leaf color, the phenotype of the individuals was assessed by visual inspection. The plants grew in the field with rows of 2 m in length and a space of 0.5 m between rows.

## 2.2. Chlorophyll Quantification and Transmission Electron Microscopy (TEM) Analysis

The chlorophyll content of the wild-type and mutant plants was directly measured in the field using the SPAD-502 Plus portable chlorophyll meter when the mutant leaves exhibited yellow. A total of ten plants were sampled for each material. The leaf tissue sections (5 mm  $\times$  3 mm) were collected from 2-week-old seedlings for transmission electron microscopy analysis. First, we used a pre-fixative solution on the sections with 2.5% (*w/v*) glutaraldehyde in 0.1 M sodium phosphate buffer (PBS, pH = 7.4) for 8 h at 4 °C. Following air extraction to aid in sinking, the sections underwent three washes using the same buffer. The samples were post-fixed in a 1% (*w/v*) osmic acid solution in 0.1 M PBS buffer (pH = 7.4) for 5 h. Afterward, samples were rinsed three times with the same buffer and dehydrated using a series of increasing concentrations of acetone (30–50–70–80–90–95–100–100%). After being permeated in a graded series of acetone embedding agents, the samples were embedded and polymerized in Spurr's resin medium at 60 °C for 48 h. Before being stained with a 1% (*w/v*) solution of uranyl acetate and a 1% (*w/v*) solution of Reynolds' lead citrate dye, the specimens were dehydrated with a graded series of acetone and cut into ultrathin slices [32]. An HT7700 Hitachi transmission electron microscope (Hitachi High Technology Corporation, Tokyo, Japan) was used for observation.

## 2.3. BSA-Seq

Resequencing was performed using the BSA technique. We used mutant phenotypic data from the F<sub>2.3</sub> generation of the W82  $\times$  *ytl* cross to make the DNA pool for sequencing. The DNA pool comprised the same amount of DNA from 30 wild-type and mutant individuals. They were selected at F<sub>2.3</sub> from a single F<sub>2</sub> plant. After liquid nitrogen freezing, the genomic DNA for bulk was isolated from the soybean leaves using hexadecyltrimethylammonium bromide (CTAB) method [33]. The DNA samples were sent to Biomarker Technologies Co., Ltd. (Qingdao, China) to construct a paired-end sequencing library with an average sequencing depth of 30 $\times$ .

The experiment was conducted following the typical methodology given by Illumina HiSeq® 2500 (Illumina Inc., San Diego, CA, USA) [34]. DNA was then modified with A at the 3' end and linked with the sequencing adapter. It was fragmented using ultrasonic disruption. Afterward, agarose gel electrophoresis was employed to isolate DNA fragments with a size range of around 200–300 bp. These fragments were then subjected to PCR amplification to generate the sequencing library. The library that successfully underwent quality assessment was sequenced using the Illumina platform.

## 2.4. Quality Control and Analysis of Sequencing Data

The generated sequencing data have paired-end 150 base (PE150) reads, and the corresponding Q20 and Q30 are calculated using the sequencing Phred value (QPred) for each base error rate. The percentage of samples with sequencing greater than Q20 and Q30 is calculated to qualify the sequencing data of the sample. Paired-end sequencing reads were mapped to the soybean Williams 82 reference genome (Wm82.a2.v1, [http://phytozome-next.jgi.doe.gov/info/Gmax\\_Wm82\\_a2\\_v1](http://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1) (accessed on 4 August 2014)) using BWA software (Version: 0.7.17-r1188) with the default parameters [35]. Duplicated reads were filtered with the Picard package (picard.sourceforge.net, Version:1.87). The Genome Analysis Toolkit (GATK, Version 4.2.5.0) was used to identify single nucleotide polymorphisms (SNPs) and insertion and deletion (InDels) variations in both gene bulks with default parameters: ‘–filter-expression “QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < –12.5 || ReadPosRankSum < –8.0”’ [36]. The ANNOVAR package (Version: 2017-07-17) was used for the functional annotation of variants. The seven algorithms, including deep learning (DL), K values, ED<sup>4</sup>,  $\Delta$ SNP index, SmoothG,

SmoothLOD, and Redit in DeepBSA software version 1.4 [37], were used to calculate candidate intervals associated with leaf color. Finally, the overlapped regions based on the above multi-methods were the most likely considered candidate intervals for leaf color.

### 2.5. Gene Mapping, Candidate Gene Prediction, and Clone

SSR markers within the primary mapping interval were used to screen for polymorphic markers between two bulks [38,39]. Then, the polymorphic markers were used to detect the genotype of recessive individuals in the F<sub>2:3</sub> and F<sub>3:4</sub> populations from the W82 × *yt1* cross. Based on the genotyping results, the target gene was mapped to a candidate interval flanked by SSR markers selected from SoyBase (<http://www.soybase.org/> (accessed on 27 February 2004)) and Song et al. [40]. They were synthesized by General Biology Co., Ltd. (Hefei, China). The DNA was extracted from soybean leaves using a One Step Plant Direct PCR kit (Senweis Co., Ltd., Huai'an, China). The SSR-PCR process was performed as reported before [32]. The PCR reactions took 10 µL of DNA (10–100 ng) as the template, 2 × Taq Master Mix (Sangon Biotech Co., Ltd., Shanghai, China), and 0.4 µM (10 µM) each of the forward and reverse primers. The PCR thermal cycle was programmed for 30 s at 95 °C, 30 s at 55 °C, 30 s at 72 °C, and a final cycle of 5 min at 72 °C before cooling to 4 °C. PCR products were separated with a capillary electrophoresis system (Agilent ZAG DNA Analyzer, Agilent Technologies, Inc. Santa Clara, CA, USA).

The genes within the mapping interval were downloaded from the SoyBase (<https://www.soybase.org/> (accessed on 4 August 2014)) website and annotated on the Phytozome (<https://phytozome-next.jgi.doe.gov/> (accessed on 4 August 2014)) website. According to the functional annotation, the candidate gene conferring chlorophyll content and chloroplast development were predicted. The candidate genes were cloned as follows: 100–1000 ng genomic DNA and 25 µL 2 × Phanta Max Master Mix (Vazyme, P515, Nanjing, China) were added to a 50 µL mixture with 2 µL (10 µM) each of the forward and reverse primers, which were programmed for 30 s at 95 °C, 30 s at 58 °C, 2.5 min at 72 °C, and a final cycle of 10 min at 72 °C before cooling to 4 °C in PCR thermal cycler. The samples used for amplifying the candidate genes came from residual heterozygous lines with contrasting patterns. PCR products of candidate genes were separated with 1% agarose gels stained with GoldView<sup>TM</sup> (Beijing Solarbio Science & Technology Co., Ltd. Beijing, China) and sequenced via Sanger (3730XL) DNA sequencing on an Applied Biosystems Automated 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

## 3. Results

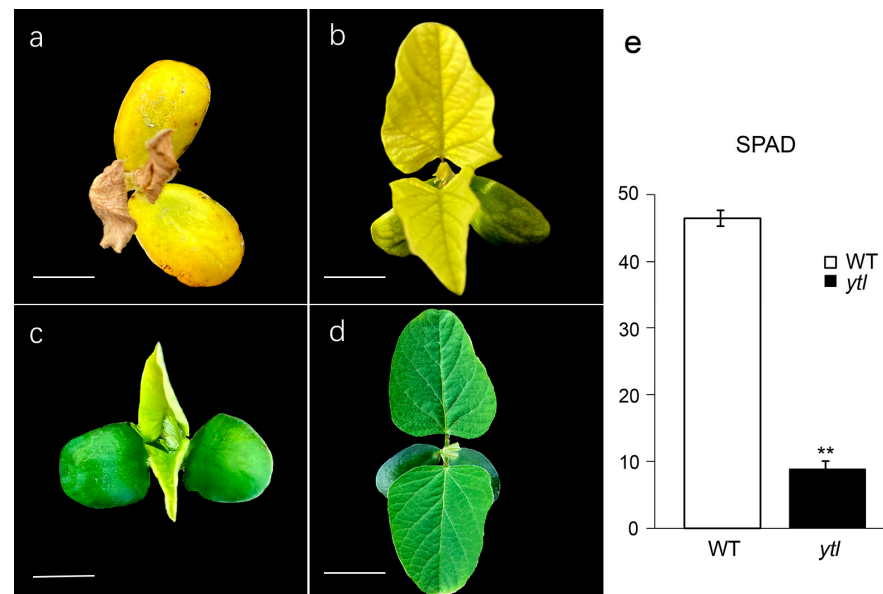
### 3.1. Characterization of the Mutant *yt1*

Under natural field conditions, the cotyledons of the wild type exhibited a green color (Figure 1d). In contrast, the *yt1*'s cotyledons were green and yellow (Figure 1a–c). However, all the true leaves of the *yt1* variant were consistently yellow. The seedlings exhibiting yellow cotyledons died rapidly (Figure 1a,b), whereas those displaying green cotyledons died gradually before reaching the first trifoliolate stage (Figure 1c). The chlorophyll content of both the wild type and mutant was measured in the field using portable chlorophyll meters during the true leaf stage. The mean chlorophyll content in wild-type leaves was 46.48, with a highest value of 47.6 and a lowest value of 44.9. The mean chlorophyll level in *yt1* leaves was 8.88, with a high of 11.1 and a minimum of 7.4. An analysis of variance revealed that the wild type had considerably greater relative chlorophyll content than the mutant plant (Figure 1e).

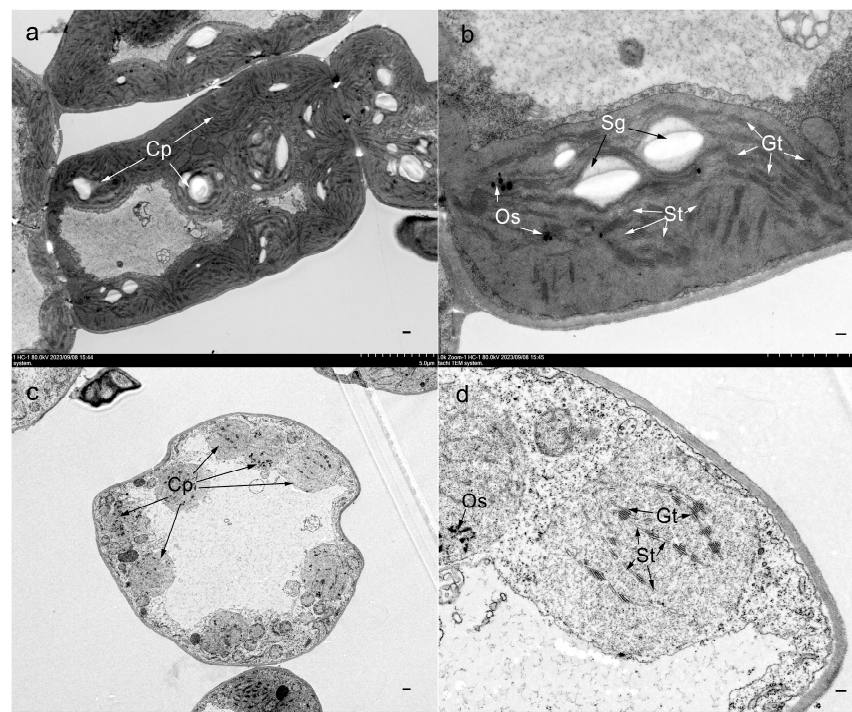
A transmission microscopy study revealed a higher abundance of chloroplasts with well-defined and intact membrane systems in the wild type than those with unclear and partial membrane systems in *yt1* (Figure 2a,c). The stroma lamellae in wild-type chloroplasts connect the grana structure. On the other hand, the stroma thylakoid was not easily visible in *yt1* chloroplasts, where the grana stack was smaller and more distorted (Figure 2b,d). Compared to the well-developed starch grains found in wild-type chloroplasts, the *yt1*



chloroplasts could not produce starch grains (Figure 2b,d). Therefore, abnormalities in the formation of chloroplasts and decreases in chlorophyll could account for the *ytl* phenotype.



**Figure 1.** Phenotypes of wild type and *ytl* mutant in the field. (a–c). Phenotypes of *ytl* mutant with different cotyledon color; (d) phenotypes of wild type; (e) chlorophyll content in the true leaves of wild type and *ytl* mutant. Bars: 1 cm. The significant test was carried out by the student's *t*-test (\*\*:  $p \leq 0.01$ ). Each value represents an average of ten replicates. Values are expressed as mean  $\pm$  SD.



**Figure 2.** Ultrastructure of chloroplast in wild type and *ytl* mutant. The thylakoid membrane organization in the chloroplasts of wild type (a,b) and *ytl* (c,d). Bars: 5  $\mu$ m (a,c); 2  $\mu$ m (b,d). Well-developed chloroplast in wild type and abnormal chloroplasts in *ytl* mutant. Each component was marked with arrows. Cp, chloroplast; Gt, grana thylakoid (stacked); St, stromal thylakoid (unstacked); Sg, starch granule. Os, osmiophilic granule.

### 3.2. Genetic Analysis

It was observed that a segregating line of M<sub>7</sub> derived from NN1138-2 treated with <sup>60</sup>Co γ ray showed a 3:1 segregation ratio between the wild type and mutant (45:14,  $\chi^2 = 0.01 < \chi^2_{0.05} = 3.84$ ,  $p = 0.94$ ). Then, the inheritance of phenotypes of F<sub>2</sub> populations derived from the W82 × *yt1* cross was investigated (Table 1). Of the 208 plants, 46 exhibited a yellow leaf color, while the remaining 162 showed green. The statistical analysis using the chi-square test indicated a value of 0.78, less than the critical value of 3.84 at a significance level of 0.05, which fit the expected ratio of 3:1. These data revealed that a single recessive nuclear allele governed the yellow-to-lethal phenotype according to Mendelian laws of inheritance.

**Table 1.** Chi-square tests of wild-type and mutant plants of F<sub>2</sub> population of W82 × *yt1* cross.

Population	Total No. of Plants	No. of Green Leaf Plants	No. of Yellow-to-Lethal Plants	Expected Ratio	$\chi^2$ (3:1)	<i>p</i>
F <sub>2</sub> (W82 × <i>yt1</i> )	208	162	46	3:1	0.78	0.38

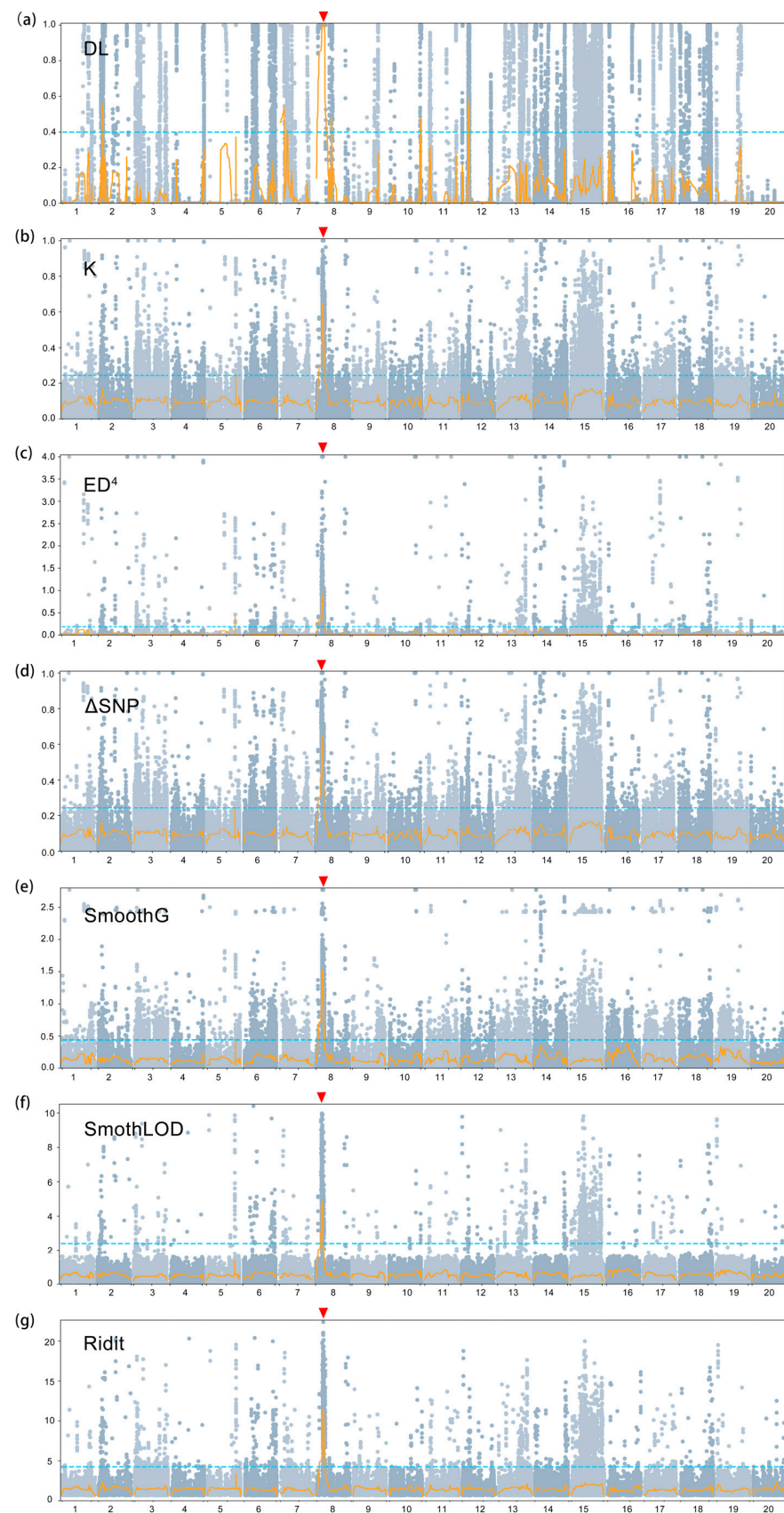
### 3.3. Analysis of BSA-Seq Quality

Thirty wild-type and thirty mutant plants from a segregating line of the F<sub>2:3</sub> generation were selected, and constructed green leaf and yellow leaf bulks. Through library construction and next-generation resequencing, we obtained approximately 209.26 M clean reads and 62.32 Gb clean data, which were evaluated as shown in Table S1. After filtering, the data volumes of green and yellow leaf bulk were 27.72 G and 34.60 G, respectively. The correct base recognition rates of 99% (Q20) and 99.9% (Q30) reached 98% and 91% of the total bases, with a GC content of about 35%. The base quality and the GC content meet the requirements for further analysis.

The reference genome size of soybean Williams 82 is 955,932,237 bp. Each mapping rate of wild-type bulk and mutant bulk was 89.58% and 88.49%. The sequencing quality of each chromosome is presented in Table S2. The coverage of each chromosome in the green leaf gene bulk was between 96.9% and 98.3%, and the sequencing depth was between 24.9% and 31.8%. The coverage of each chromosome in the yellow leaf bulk was between 96.9% and 98.4%, and the sequencing depth was between 31.9% and 38.5%. The sequencing data met the criteria of resequencing analysis and were suitable for SNP and InDel testing.

### 3.4. Preliminary Mapping Based on the BSA-Seq

Two bulks yielded 11,522,230 SNPs covered in the 20 soybean chromosomes. The seven algorithms included in DeepBSA software version 1.4 [37] were employed to compare the SNP frequency differences between the two bulks ( $\Delta$ SNP index) and estimate the mapping region related to the leaf color. As shown in Figure 3 and Table S3, one predominant signal located on chromosome 8 was detected by all the algorithms. Taking the intersection of the seven algorithms, the mapping interval was anchored finally to the region between 7,435,009 and 10,766,043 with a peak of 8,758,861~9,666,879.



**Figure 3.** Mapping intervals of yellow-to-lethal locus based on the different algorithms in DeepBSA. The  $x$ -axis represents the soybean genome chromosome, and the  $y$ -axis represents the association

value of the SNPs calculated by the algorithm. Each gray dot represents the position and association value of an SNP, the yellow line is a Tri-kernel smooth fitting for all dots, the blue line is the default threshold, which is three standard deviations above the genome-wide median, and the red triangle indicates the mapping interval identified by the algorithms. (a) DL algorithm; (b) K algorithm; (c) ED<sup>4</sup> algorithm; (d) ΔSNP index algorithm; (e) SmoothG algorithm; (f) SmoothLOD algorithm; (g) Redit algorithm.

### 3.5. Fine Mapping by Using SSR Markers

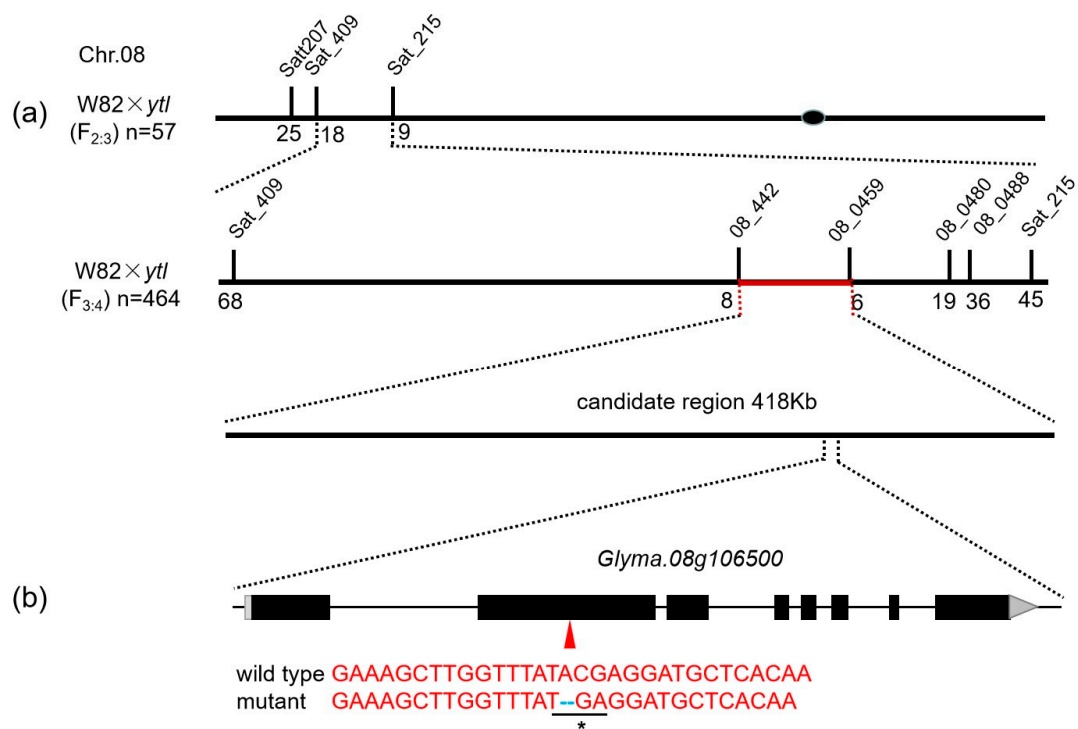
During the preliminary mapping stage on chromosome 8, we examined sixteen SSR markers from the published SSR genetic map on SoyBase (<http://soybase.org/> (accessed on 27 February 2004)) to identify polymorphic markers between the green leaf and *yt1* leaf bulks. Polymorphism was seen only in Satt207, Sat\_409, and Sat\_215 among the two bulks. The sequences of these polymorphic markers are provided in Table 2. Fifty-seven F<sub>2:3</sub> mutant individuals were genotyped using the three markers. Figure 4 showed 25, 18, and 9 recombinant individuals for SSR markers Satt207, Sat\_409, and Sat\_215, respectively. Based on the genotyping results, the target gene was located between Sat\_409 and Sat\_215. Among the twenty-four newly acquired SSR markers from Song et al. [40] located between Sat\_409 and Sat\_215, only four markers showed a variation between the wild-type and mutant bulks (Table 2). A total of 464 F<sub>3:4</sub> mutant individuals were genotyped using 4 markers. Among these, SSR markers BARCSOYSSR\_08\_0442, BARCSOYSSR\_08\_0459, BARCSOYSSR\_08\_0480, and BARCSOYSSR\_08\_0488 had 8, 6, 19, and 36 recombinant individuals, respectively. As a result, the *yt1* was narrowed down to a 418 kb region between BARCSOYSSR\_08\_0442 and BARCSOYSSR\_08\_459 markers containing 53 candidate genes.

**Table 2.** Polymorphic SSR markers for the genetic linkage map of the *yt1* mutant gene.

Primer Name	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
Satt207	GCGTTTTCTCATTTTGATTCTAAAC	GCGATTGTGATTGTAGTCCCTAAA
Sat_409	GCGGAGGTTTGTGCATTTCTAGGTCTTC	GCGACGCGTATGTACATAAAATATGCTGTT
Sat_215	GCGTAGCAACAAAGCAATCTACAG	GCGTCCCATTATTATCCACACTATGTAAT
BARCSOYSSR_08_0442	GAAACGGTTGGAGAATAGCG	TCCATGCCCCACTTAACAACA
BARCSOYSSR_08_0459	CAACTTTCGCATCGGTTACA	CGGAAACTGCTTATGGTTGC
BARCSOYSSR_08_0480	TCGTTGACTGCAATATGGTC	ACCGAATGAAGCGTAAAGGA
BARCSOYSSR_08_0488	TCTCATGGGAACCTGGAAAA	CGAAACAGCAACGAATCAAA
Glyma.08g106500-1	TTGGAGCTGGCAGAGTCTAATC	AATGAGCTGATGGCCACTGTA
Glyma.08g106500-2	GCCAGATGGGGACGTCATAAG	TCAATTGATGGAGCAACACCG
Glyma.08g106500-3	CAGGAAAACATCAGGAGGCAGA	GCTGCATTGCATCAAACGTGG

By searching the Phytozome database (<http://phytozome-next.jgi.doe.gov/pz/portal.html> (accessed on 4 August 2014)), the functional annotation of the 53 genes was downloaded and is shown in Table 3, only 11 genes of which had SNPs or InDels within their exon regions according to the BSA-Seq results: *Glyma.08g103200*, *Glyma.08g103300*, *Glyma.08g103400*, *Glyma.08g103700*, *Glyma.08g104000*, *Glyma.08g105400*, *Glyma.08g105600*, *Glyma.08g106500*, *Glyma.08g106700*, *Glyma.08g106900*, and *Glyma.08g107700*. Among these genes was a two bp deletion in the exon of *Glyma.08g106500*, leading to a frameshift with amino acid sequence changes. *Glyma.08g106500* encoded a PPR\_long domain-containing protein and was an ortholog of AT5G27270 in *Arabidopsis*, whose T-DNA mutant showed a seedling-lethal and albino phenotype with the chloroplasts broken down [41]. However, according to the *Arabidopsis* orthologs, the mutations in the other 10 genes could not lead to a yellow-to-lethal phenotype. Then, we sequenced the genomic DNA sequence of *Glyma.08g106500* between the wild type and mutant of a residual heterozygous line with three pairs of primers (Table 2), and a two bp deletion causing a frameshift and premature stop codon was observed in the second exon of mutant (Figure 4b). Then, the same deletion was detected in all 113 recombinant mutants from the segregating F<sub>3:4</sub> population. Therefore, *Glyma.08g106500* was predicted as the candidate gene for *yt1*.





**Figure 4.** Map-based cloning of *ytI* with SSR markers. (a) Gene mapping of *ytI* on chromosome 8 using  $F_{2:3}$  and  $F_{3:4}$  populations of  $W82 \times ytl$ . (b) Schematic diagram of *Glyma.08g106500* gene structure and variation in the wild type and mutant. There is a two bp deletion in the second exon of mutant *ytI*. The red triangle indicates the approximate position of the deletion in mutant *ytI*. Asterisk indicates the stop codon.

**Table 3.** Predicted genes and their putative functions within the mapped region.

Gene ID	Gene Start (bp)	Gene End (bp)	Putative Function
<i>Glyma.08G103100</i>	7,902,820	7,907,892	RING-type E3 ubiquitin transferase
<i>Glyma.08G103200</i>	7,918,306	7,932,242	Mannosyl-oligosaccharide glucosidase
<i>Glyma.08G103300</i>	7,934,866	7,944,355	Peptidase_M28 domain-containing protein
<i>Glyma.08G103400</i>	7,945,667	7,949,720	F-box domain-containing protein
<i>Glyma.08G103500</i>	7,947,843	7,948,181	Unknown
<i>Glyma.08G103600</i>	7,954,858	7,956,497	Signal peptidase complex subunit 1
<i>Glyma.08G103700</i>	7,959,250	7,962,513	Large subunit ribosomal protein L17e
<i>Glyma.08G103800</i>	7,966,298	7,969,958	Exostosin domain-containing protein
<i>Glyma.08G103900</i>	7,970,507	7,974,378	Caffeoyl-CoA O-methyltransferase
<i>Glyma.08G104000</i>	7,976,043	7,980,358	S-adenosyl-L-methionine-dependent methyltransferases
<i>Glyma.08G104100</i>	7,993,317	7,995,690	Flavonol biosynthesis
<i>Glyma.08G104200</i>	8,004,960	8,006,787	REF domain-containing protein
<i>Glyma.08G104300</i>	8,007,843	8,011,052	Aerobic respiration I (cytochrome c)
<i>Glyma.08G104400</i>	8,016,766	8,018,606	MYND-type domain-containing protein
<i>Glyma.08G104500</i>	8,041,175	8,044,478	DUF493
<i>Glyma.08G104600</i>	8,045,602	8,047,222	Polysacc_synt_4 domain-containing protein
<i>Glyma.08G104700</i>	8,055,039	8,059,859	Ribosome assembly protein RRB1
<i>Glyma.08G104800</i>	8,063,060	8,066,966	F-box domain-containing protein
<i>Glyma.08G104900</i>	8,067,568	8,069,734	RAMP4 in endoplasmic reticulum
<i>Glyma.08G105000</i>	8,079,687	8,087,576	<i>Arabidopsis</i> histidine kinase 2/3/4 (cytokinin receptor)
<i>Glyma.08G105100</i>	8,091,820	8,093,816	NTF2 domain-containing protein
<i>Glyma.08G105200</i>	8,094,716	8,102,735	TIG DNA binding
<i>Glyma.08G105300</i>	8,109,486	8,110,128	DUF3339
<i>Glyma.08G105400</i>	8,113,990	8,126,799	K-box domain-containing protein
<i>Glyma.08G105500</i>	8,128,245	8,135,906	K-box domain-containing protein

Table 3. Cont.

Gene ID	Gene Start (bp)	Gene End (bp)	Putative Function
<i>Glyma.08G105600</i>	8,154,834	8,161,846	DNA primase large subunit
<i>Glyma.08G105700</i>	8,166,200	8,168,551	Rho GDP-dissociation inhibitor
<i>Glyma.08G105800</i>	8,171,582	8,177,075	MACPF domain-containing protein
<i>Glyma.08G105900</i>	8,182,656	8,185,890	U6 snRNA-associated Sm-like protein LSm7
<i>Glyma.08G106000</i>	8,189,588	8,192,602	Amidase domain-containing protein
<i>Glyma.08G106100</i>	8,194,006	8,194,604	Unknown
<i>Glyma.08G106200</i>	8,194,864	8,196,276	Amidase domain-containing protein
<i>Glyma.08G106300</i>	8,197,911	8,199,307	PORR domain-containing protein
<i>Glyma.08G106400</i>	8,199,636	8,206,598	kinesin family member 15
<i>Glyma.08G106500</i>	8,208,987	8,215,162	PPR_long domain-containing protein
<i>Glyma.08G106600</i>	8,215,384	8,215,533	Unknown
<i>Glyma.08G106700</i>	8,216,338	8,219,979	FK506-binding protein 4/5
<i>Glyma.08G106800</i>	8,221,055	8,223,572	Small subunit ribosomal protein S19
<i>Glyma.08G106900</i>	8,228,380	8,228,919	Unknown
<i>Glyma.08G107000</i>	8,229,694	8,231,371	mTERF domain-containing protein
<i>Glyma.08G107100</i>	8,232,784	8,238,471	C2H2-type domain-containing protein
<i>Glyma.08G107200</i>	8,242,346	8,245,459	Tic22 domain-containing protein
<i>Glyma.08G107300</i>	8,250,677	8,253,307	Xyloglucan:xyloglucosyl transferase
<i>Glyma.08G107400</i>	8,255,102	8,260,380	Unknown
<i>Glyma.08G107500</i>	8,266,733	8,268,742	Glycosyltransferase
<i>Glyma.08G107600</i>	8,276,759	8,278,195	Glycosyltransferase
<i>Glyma.08G107700</i>	8,284,689	8,288,765	Protein kinase domain-containing protein
<i>Glyma.08G107800</i>	8,296,196	8,307,328	Bifunctional aspartokinase/homoserine dehydrogenase 1
<i>Glyma.08G107900</i>	8,308,038	8,310,337	Histone H3
<i>Glyma.08G108000</i>	8,311,539	8,314,951	Selenium-binding protein 1
<i>Glyma.08G108100</i>	8,315,634	8,315,795	RALF domain-containing protein
<i>Glyma.08G108200</i>	8,317,340	8,318,268	RALF domain-containing protein
<i>Glyma.08G108300</i>	8,321,303	8,325,907	Microfibrillar-associated protein 1

#### 4. Discussion

The BSA-Seq strategy can be applied to multiple populations with genetic differences and combined with traditional gene mapping or association analysis methods. Several soybean genes controlling development and traits related to stress resistance or quality were mapped based on BSA-Seq in recent years. Song et al. [42] identified two candidate genes controlling soybean cotyledon color through a next-generation sequencing (NGS)-based BSA approach and marker-based classical gene mapping. Zhong et al. [43] finely mapped *RpsHC18* governing *Phytophthora sojae* resistance on soybean chromosome 3 based on QTL-seq, which was performed by the whole-genome resequencing (WGRS) of highly resistant and susceptible bulks from an F<sub>2.3</sub> population. Ochar et al. [34] mapped the candidate gene conferring crinkled leaf phenotype based on BSA-Seq technology and promised *Glyma.19G207100* would be a candidate gene. Zhang et al. [20] employed 1551 natural accessions with diverse worldwide origins to do BSA-Seq and identified 130 candidate genes underlying seed oil content in soybean. This research revealed that NGS-based BSA methods could be quickly and effectively used for mapping and mining agronomic traits and stress-resistance genes in soybean. DeepBSA is a deep learning-driven BSA method, and it is the first time deep learning has been applied to BSA for QTL detection or functional gene cloning [37]. DeepBSA outperforms all other algorithms regarding absolute deviation and the signal-to-noise ratio and performs well in various animal and plant databases. However, this method has yet to be reported in soybean. In the present study, green leaf and yellow leaf bulk from the F<sub>2.3</sub> population of a biparental population (W82 × *yt1*) was constructed for BSA-Seq. According to the results of ΔSNP performed by DeepBSA, we detected the most probable gene linked to a region with a physical distance of about 3.33 Mb on chromosome 8. Then, the SSR markers were used to validate the preliminary intervals

and narrow them down to a 418 Kb region between SSR markers BARCSOYSSR\_08\_0442 and BARCSOYSSR\_08\_459.

Many genes cause chlorophyll deficiency to lethality in plants, including nuclear and cytoplasmic genes, most of which are recessive nuclear genes involved in several pathways. Some genes participate in the biosynthesis and degradation of photosynthetic pigments pathway, and some genes are related to chloroplast development, viz., plastid-encoded RNA polymerase (PEP), posttranscriptional modification of plastid genes, nuclear-cytoplasmic signal transduction, and chloroplast protease. In *Arabidopsis*, the inactivation of ATAB2 leads to an albino phenotype, which strongly affects development and thylakoid membrane biogenesis because it damages proteins photosystem I and photosystem II [44]. The mutations in the G- or M-domain of atToc159 could cause a severe albino phenotype in *Arabidopsis*, and the atToc159 functions in the recognition and import of proteins into chloroplasts [45–47]. Moon et al. [48] reported that a T-DNA insertion in the *OsPDF1B* gene led to a seedling lethal phenotype in rice, which affected the development of chloroplasts and perhaps mitochondria. A single, nonsynonymous substitution in *Glyma13g30560* encoding the Mg-chelatase subunit (ChII1a) led to a chlorophyll-deficient and lethal phenotype in soybean [12]. A single base insertion in the *GmPsbP*, encoding an extrinsic photosystem II protein critical for oxygen evolution during photosynthesis, caused lethal yellow mutants [15].

The PPR proteins were characterized by a PPR repeat with highly degenerate units of 35 amino acids and participated in every step of chloroplast gene expression: transcription, RNA cleavage, RNA splicing, translation, and so on [49]. PPR mutants had various phenotypes attributed to little redundancy between family members, many related to leaf color mutants. A mutant of thylakoid assembly 8 (*tha8*) led to a phenotype of pale green and lethal seedling, and the gene encoded several PPR families [50]. As a chloroplast splicing factor, THA8 indirectly affected thylakoid energetics [50]. The LPA66-defect plant showed a high chlorophyll fluorescence phenotype and pale green leaf in *Arabidopsis*, and it is a chloroplast PPR protein editing *psbF* transcripts [51]. In this study, the mutant *ytl* showed a yellow-to-lethal phenotype with significantly decreased chlorophyll contents. TEM exhibited an abnormal chloroplast with a thylakoid structure mainly disrupted, indicating that the defects in chloroplast development caused the lethal phenotype. In the mapped region, a two bp deletion caused a frameshift mutation in *Glyma.08g106500* of *ytl*. The *Arabidopsis* ortholog of *Glyma.08g106500*, AT5G27270, encodes a chloroplast-localized P-type PPR protein with the function of the splicing of group II introns. The T-DNA insertion mutant lost the chloroplast development with defective pigment and seedling-lethal phenotype [41]. Thus, based on the results of gene mapping and the similar mutant phenotype in soybean and *Arabidopsis*, the PPR-encoded gene *Glyma.08g106500* was a promising candidate for mutant *ytl*. Further functional verification will be performed in the future.

## 5. Conclusions

In the present study, we identified a soybean yellow-to-lethal mutant, *ytl*, which showed a seedling-lethal phenotype and a significant decrease in chlorophyll content. The *ytl* showed developmental defects in the chloroplast membrane system and could not form starch granules. Genetic analysis showed that the phenotype of the mutant was controlled by a recessive allele, which was mapped to a 418 Kb interval flanked by SSR markers BARCSOYSSR\_08\_0442 and BARCSOYSSR\_08\_459 with 53 candidate genes. According to sequencing results, a two bp deletion in the second exon of *Glyma.08g106500* led to a frameshift mutation with a premature stop codon. *Glyma.08g106500* is a strong candidate for *ytl* because of the similar functions of its ortholog in *Arabidopsis*, whose T-DNA mutant showed a seedling-lethal and albino phenotype with the chloroplasts broken down. *Glyma.08g106500* encodes a PPR<sub>long</sub> domain-containing protein, which plays a vital role in the transcription and translation of proteins involved in chloroplast development. Further study of *Glyma.08g106500*, such as gene expression, identification

of interacting proteins, and the regulation of downstream genes, may provide a better understanding of the assembly of chloroplast developmental proteins.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy14010185/s1>. Table S1: Statistical analysis on sequencing data quality; Table S2: Matching of quality control data with reference genome; Table S3: Mapping regions by different algorithms based on DeepBSA.

**Author Contributions:** Conceptualization, Y.W. and T.Z.; data curation, Y.W. and F.C.; investigation, S.L., M.F. and H.X.; resources, G.M.A.A.; writing—original draft preparation, Y.W.; writing—review and editing, F.C. and G.M.A.A.; visualization, X.Y. and Z.Z.; supervision, T.Z. and H.X.; funding acquisition, Y.W., T.Z. and H.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (32201729), the Scientific Research Fund of Startup and Development for Introduced High-level Talents, Huai'an Academy of Agricultural Sciences, China (0112023014B), the Research and Development Fund Project of Huai'an Academy of Agricultural Sciences, China (HNY202221), the Core Technology Development for Breeding Program of Jiangsu Province (JBGS [2021]057), and the Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP) Program.

**Data Availability Statement:** The raw data of BSA-Seq can be downloaded from NCBI with the BioProject ID PRJNA1061757. The sequences for the candidate gene in wild type and mutant (PP098297 and PP098298) can be searched at NCBI when the paper is published.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Sharrock, R.A.; Parks, B.M.; Koornneef, M.; Quail, P.H. Molecular analysis of the phytochrome deficiency in an *aurea* mutant of tomato. *Mol. Gen. Genet. (MGG)* **1988**, *213*, 9–14. [\[CrossRef\]](#)
- Kato, K.K.; Palmer, R.G. Duplicate chlorophyll-deficient loci in soybean. *Genome* **2004**, *47*, 190–198. [\[CrossRef\]](#) [\[PubMed\]](#)
- Palmer, R.G.; Nelson, R.L.; Bernard, R.L.; Stelly, D.M. Genetics and linkage of three chlorophyll-deficient mutants in soybean: *y19*, *y22*, and *y23*. *J. Hered.* **1990**, *81*, 404–406. [\[CrossRef\]](#)
- Palmer, R.G.; Burzlaff, J.D.; Shoemaker, R.C. Genetic analyses of two independent chlorophyll-deficient mutants identified among the progeny of a single chimeric foliage soybean plant. *J. Hered.* **2000**, *91*, 297–303. [\[CrossRef\]](#) [\[PubMed\]](#)
- Davis, S.J.; Kurepa, J.; Vierstra, R.D. The *Arabidopsis thaliana* HY1 locus, required for phytochrome-chromophore biosynthesis, encodes a protein related to heme oxygenases. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6541–6546. [\[CrossRef\]](#)
- Ki-Hong, J.; Junghe, H.; Choong-Hwan, R.; Youngju, C.; Yong-Yoon, C.; Akio, M.; Hirochika, H.; An, G. Characterization of a rice chlorophyll-deficient mutant using the T-DNA gene-trap system. *Plant Cell Physiol.* **2003**, *44*, 463–472. [\[CrossRef\]](#)
- Sakuraba, Y.; Rahman, M.L.; Cho, S.H.; Kim, Y.S.; Koh, H.J.; Yoo, S.C.; Paek, N.C. The rice *faded green leaf* locus encodes protochlorophyllide oxidoreductase and is essential for chlorophyll synthesis under high light conditions. *Plant J.* **2013**, *74*, 122–133. [\[CrossRef\]](#)
- Gu, J.; Zhou, Z.; Li, Z.; Chen, Y.; Wang, Z.; Zhang, H. Rice (*Oryza sativa* L.) with reduced chlorophyll content exhibit higher photosynthetic rate and efficiency, improved canopy light distribution, and greater yields than normally pigmented plants. *Field Crops Res.* **2017**, *200*, 58–70. [\[CrossRef\]](#)
- Liu, M.; Wang, Y.; Nie, Z.; Gai, J.; Bhat, J.A.; Kong, J.J.; Zhao, T. Double mutation of two homologous genes *yl1* and *yl2* results in a leaf yellowing phenotype in soybean [*Glycine max* (L.) Merr]. *Plant Mol. Biol.* **2020**, *103*, 527–543. [\[CrossRef\]](#)
- Sandhu, D.; Coleman, Z.; Atkinson, T.; Krishan, M.R.; Mendu, V. Genetics and physiology of the nuclearly inherited yellow foliar mutants in soybean. *Front. Plant Sci.* **2018**, *9*, 471. [\[CrossRef\]](#)
- Weber, C.R.; Weiss, M.G. Chlorophyll mutant in soybeans provides teaching aid. *J. Hered.* **1959**, *50*, 53–54. [\[CrossRef\]](#)
- Campbell, B.W.; Mani, D.; Curtin, S.J.; Slattery, R.A.; Stupar, R.M. Identical substitutions in magnesium chelatase paralogs result in chlorophyll-deficient soybean mutants. *G3-Genes Genom. Genet.* **2015**, *5*, 123–131. [\[CrossRef\]](#) [\[PubMed\]](#)
- Noble, R.D.; Czarnota, C.D.; Cappy, J.J. Morphological and physiological characteristics of an achlorophyllous mutant soybean variety sustained to maturation via grafting. *Am. J. Bot.* **1977**, *64*, 1042–1045. [\[CrossRef\]](#)
- Reed, S.; Atkinson, T.; Gorecki, C.; Espinosa, K.; Przybylski, S.; Goggi, A.S.; Palmer, R.G.; Sandhu, D. Candidate gene identification for a lethal chlorophyll-deficient mutant in soybean. *Agronomy* **2014**, *4*, 462–469. [\[CrossRef\]](#)
- Sandhu, D.; Atkinson, T.; Noll, A.; Johnson, C.; Espinosa, K.; Boelter, J.; Abel, S.; Balpreet, K.; Dhatt, B.K.; Barta, T.; et al. Soybean proteins GmTic110 and GmPsbP are crucial for chloroplast development and function. *Plant Sci.* **2016**, *252*, 76–87. [\[CrossRef\]](#) [\[PubMed\]](#)



16. Schneeberger, K.; Ossowski, S.; Lanz, C.; Juul, T.; Petersen, A.H.; Nielsen, K.L.; Jrgensen, J.-E.; Weigel, D.; Andersen, S.U. SHOREmap: Simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* **2009**, *6*, 550–551. [\[CrossRef\]](#)
17. Baek, G.; Kim, C.W.; Kim, S. Development of a molecular marker tightly linked to the C locus conferring a white bulb color in onion (*Allium cepa* L.) using bulked segregant analysis and RNA-Seq. *Mol. Breed.* **2017**, *37*, 94. [\[CrossRef\]](#)
18. Yu, J.; Holland, J.B.; McMullen, M.D.; Buckler, E.S. Genetic design and statistical power of nested association mapping in maize. *Genetics* **2008**, *178*, 539–551. [\[CrossRef\]](#)
19. Kover, P.X.; Valdar, W.; Trakalo, J.; Scarcelli, N.; Ehrenreich, I.M.; Purugganan, M.D.; Durrant, C.; Mott, R. A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **2009**, *5*, e1000551. [\[CrossRef\]](#)
20. Zhang, S.; Abdelghany, A.M.; Azam, M.; Qi, J.; Li, J.; Feng, Y.; Liu, Y.; Feng, H.; Ma, C.; Gebregziabher, B.S.; et al. Mining candidate genes underlying seed oil content using BSA-seq in soybean. *Ind. Crops Prod.* **2023**, *194*, 116308. [\[CrossRef\]](#)
21. Zou, C.; Wang, P.; Xu, Y. Bulk sample analysis in genetics, genomics and crop improvement. *Plant Biotechnol. J.* **2016**, *14*, 1941–1955. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Abe, A.; Kosugi, S.; Yoshida, K.; Natsume, S.; Takagi, H.; Kanzaki, H.; Matsumura, H.; Yoshida, K.; Mitsuoka, C.; Tamiru, M.; et al. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **2012**, *30*, 174–178. [\[CrossRef\]](#)
23. Takagi, H.; Tamiru, M.; Abe, A.; Takagi, H.; Tamiru, M.; Abe, A.; Yoshida, K.; Uemura, A.; Yaegashi, H.; Obara, T.; et al. MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nat. Biotechnol.* **2015**, *33*, 445–449. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Frouin, J.; Filloux, D.; Taillebois, J.; Grenier, C.; Montes, F.; Lamotte, F.D.; Verdeil, J.; Courtois, B.; Ahmadi, N. Positional cloning of the rice male sterility gene *ms-IR36*, widely used in the inter-crossing phase of recurrent selection schemes. *Mol. Breed.* **2014**, *33*, 555–567. [\[CrossRef\]](#)
25. Rodriguez-Urbe, L.; Higbie, S.M.; Stewart, J.M.; Wilkins, T.; Lindemann, W.; Sengupta-Gopalan, C.; Zhang, J. Identification of salt responsive genes using comparative microarray analysis in Upland cotton (*Gossypium hirsutum* L.). *Plant Sci.* **2011**, *180*, 461–469. [\[CrossRef\]](#)
26. Klein, H.; Xiao, Y.; Conklin, P.A.; Govindarajulu, R.; Kelly, J.A.; Scanlon, M.J.; Whipple, C.J.; Bartlett, M. 2018. Bulk-segregant analysis coupled to whole genome sequencing (BSA-Seq) for rapid gene cloning in maize. *G3-Genes Genom. Genet.* **2018**, *8*, 3583–3592. [\[CrossRef\]](#)
27. Forrest, K.; Pujol, V.; Bulli, P.; Pumphrey, M.; Wellings, C.; Herrera-Foessel, S.; Huerta-Espino, J.; Singh, R.; Lagudah, E.; Hayden, M.; et al. Development of a SNP marker assay for the *Lr67* gene of wheat using a genotyping by sequencing approach. *Mol. Breed.* **2014**, *34*, 2109–2118. [\[CrossRef\]](#)
28. Livaja, M.; Wang, Y.; Wieckhorst, S.; Haseneyer, G.; Seidel, M.; Hahn, V.; Knapp, S.J.; Taudien, S.; Schön, C.-C.; Bauer, E. BSTA: A targeted approach combines bulked segregant analysis with next-generation sequencing and *de novo* transcriptome assembly for SNP discovery in sunflower. *BMC Genom.* **2013**, *14*, 628. [\[CrossRef\]](#)
29. Al Amin, G.M.; Kong, K.; Sharmin, R.A.; Kong, J.; Bhat, J.A.; Zhao, T. Characterization and Rapid Gene-Mapping of Leaf Lesion Mimic Phenotype of *spl-1* Mutant in Soybean (*Glycine max* (L.) Merr.). *Int. J. Mol. Sci.* **2019**, *20*, 2193. [\[CrossRef\]](#)
30. Galya, K.; Yael, B.; Adi, F.D.; Abhinandan, P.; Ilan, H.; Ran, H. Fine-mapping the branching habit trait in cultivated peanut by combining bulked segregant analysis and high-throughput sequencing. *Front. Plant Sci.* **2017**, *8*, 467. [\[CrossRef\]](#)
31. Pearson, K.X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **1900**, *50*, 157–175. [\[CrossRef\]](#)
32. Wang, Y.; Liu, M.; Ge, D.; Bhat, J.A.; Li, Y.; Kong, J.; Liu, K.; Zhao, T. Hydroperoxide lyase modulates defense response and confers lesion-mimic leaf phenotype in soybean (*Glycine max* (L.) merr.). *Plant J.* **2020**, *104*, 1315–1333. [\[CrossRef\]](#)
33. Porebski, S.; Bailey, L.G.; Baum, B.R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **1997**, *15*, 8–15. [\[CrossRef\]](#)
34. Ochar, K.; Bo-Hong, S.U.; Zhou, M.; Liu, Z.; Gao, H.; Flamlo, S.; Qiu, L. Identification of the genetic locus associated with the crinkled leaf phenotype in a soybean (*Glycine max* L.) mutant by BSA-Seq technology. *J. Integr. Agric.* **2022**, *21*, 3524–3539. [\[CrossRef\]](#)
35. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [\[CrossRef\]](#)
36. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [\[CrossRef\]](#)
37. Li, Z.; Chen, X.; Shi, S.; Zhang, H.; Wang, X.; Chen, H.; Li, W.; Li, L. DeepBSA: A deep-learning algorithm improves bulked segregant analysis for dissecting complex traits. *Mol. Plant* **2022**, *15*, 1418–1427. [\[CrossRef\]](#)
38. Wang, Y.; Chen, W.; Zhang, Y.; Liu, M.; Kong, J.; Yu, Z.; Jaffer, A.M.; Gai, J.; Zhao, T. Identification of two duplicated loci controlling a disease-like rugose leaf phenotype in soybean. *Crop Sci.* **2016**, *56*, 1611–1618. [\[CrossRef\]](#)
39. Garzon, L.N.; Blair, M.W. Development and mapping of SSR markers linked to resistance-gene homologue clusters in common bean. *Crop J.* **2014**, *2*, 183–194. [\[CrossRef\]](#)
40. Song, Q.; Jia, G.; Zhu, Y.; Grant, D.; Nelson, R.T.; Hwang, E.Y.; Hyten, D.L.; Cregan, P.B. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR\_1.0) in soybean. *Crop Sci.* **2010**, *50*, 1950–1960. [\[CrossRef\]](#)
41. Wang, X.; Zhao, L.; Man, Y.; Li, X.; Xiao, J. Pdm4, a pentatricopeptide repeat protein, affects chloroplast gene expression and chloroplast development in *Arabidopsis thaliana*. *Front. Plant Sci.* **2020**, *11*, e1198. [\[CrossRef\]](#)

42. Song, J.; Li, Z.; Liu, Z.; Guo, Y.; Qiu, L. Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. *Front. Plant Sci.* **2017**, *8*, e919. [[CrossRef](#)] [[PubMed](#)]
43. Zhong, C.; Sun, S.; Li, Y.; Duan, C.; Zhu, Z. Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel phytophthora resistance gene, *RpsHC18*, in soybean. *Theor. App. Genet.* **2018**, *131*, 525–538. [[CrossRef](#)]
44. Barneche, F.; Winter, V.; Crèvecoeur, M.; Rochaix, J.D. *ATAB2* is a novel factor in the signalling pathway of light-controlled synthesis of photosystem proteins. *Embo J.* **2014**, *25*, 5907–5918. [[CrossRef](#)] [[PubMed](#)]
45. Bauer, J.; Hiltbrunner, A.; Weibel, P.; Vidi, P.A.; Alvarez-Huerta, M.; Smith, M.D.; Schnell, D.J.; Kessler, F. Essential role of the G-domain in targeting of the protein import receptor atToc159 to the chloroplast outer membrane. *J. Cell Biol.* **2002**, *159*, 845–854. [[CrossRef](#)] [[PubMed](#)]
46. Lee, K.H.; Kim, S.J.; Lee, Y.J.; Jin, J.B.; Hwang, I. The M domain of atToc159 plays an essential role in the import of proteins into chloroplasts and chloroplast biogenesis. *J. Biol. Chem.* **2003**, *278*, 36794–36805. [[CrossRef](#)]
47. Agne, B.; Infanger, S.; Wang, F.; Hofstetter, V.; Rahim, G.; Martin, M.; Lee, D.W.; Hwang, I.; Schnell, D.; Kessler, F. A Toc159 import receptor mutant, defective in hydrolysis of GTP, supports preprotein import into chloroplasts. *J. Biol. Chem.* **2009**, *284*, 8670–8679. [[CrossRef](#)]
48. Moon, S.; Giglione, C.; Lee, D.Y.; An, S.; Jeong, D.H.; Meinel, T.; An, G. Rice peptide deformylase PDF1B is crucial for development of chloroplasts. *Plant Cell Physiol.* **2008**, *49*, 1536–1546. [[CrossRef](#)]
49. Barkan, A.; Small, I. Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 415–442. [[CrossRef](#)]
50. Khrouchtchova, A.; Monde, R.A.; Barkan, A. A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA* **2012**, *18*, 1197–1209. [[CrossRef](#)]
51. Cai, W.; Ji, D.; Peng, L.; Guo, J.; Ma, J.; Zou, M.; Lu, C.; Zhang, L. LPA66 is required for editing *psbF* chloroplast transcripts in *Arabidopsis*. *Plant Physiol.* **2009**, *150*, 1260–1271. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.