Characterisation of Photosynthesis related genes via T-DNA Mutant Screening

Christos Livitsanos

MSc Thesis Report

GEN80436

Name: Christos Livitsanos

Student number: 1145207

Master: Plant Biotechnology

Specialization: Functional Plant Genomics

Supervisor: Phuong Nguyen

Examiner: René Boesten

Date: 01/09/2022 – 28/02/2023

Chair Group: Laboratory of Genetics

Contents

Abstract

In the face of an ongoing climate crisis and a rising planet population breeding for improved crop quality and quantity is essential. Improving photosynthetic efficiency is a possible way to improve crop yield, but the trait is both complex and understudied. Identifying the genes that underly the trait is the first major step towards breeding for it. Several Genome-Wide Association studies had identified 3 major Quantitative Trait Loci (QTLs) related to the photosynthetic efficiency indicators ΦPSII and ΦNO on chromosomes 1, 2 and 3 for the the *A. thaliana* Dutch population. Subsequent analyses showed numerous candidate genes contained within the QTLs. In this project, 19 of these candidate genes were tested with the use of 23 T-DNA insert lines. The aim of the project was to identify the most promising candidate genes based on the phenotypic variation between the mutant and the natural accession plants. The mutant lines were genotyped for the insert and subsequently phenotyped via High Throughput Phenotyping. Six of the genes exhibited altered phenotypes with the most promising being AT2G18030 and AT3G45760. Gene AT3G45760 in particular showed signs of pleiotropy, since its disruption affected traits, seemingly unrelated to photosynthesis, like fertility and seed size. These results further consolidate the findings of previous projects that had identified AT2G18030 and AT3G45760 as promising candidates. Further research such as gene expression analysis and Quantitative Complementation will be essential to gain further insights into the roles of these candidate genes in photosynthetic efficiency.

Introduction

Photosynthesis and its importance

Oxygenic photosynthesis is the complex process, which allows plants, cyanobacteria and certain algae to convert light energy into organic molecules via the use of CO_2 and water. At the same time, the reaction releases molecular O_2 back into the atmosphere (Stirbet et al., 2020). This procedure is expressed in the global equation of photosynthesis:

$$6CO_2 + 12H_2O + \text{light energy} \rightarrow C_6H_{12}O_6 + 6O_2 + 6H_2O$$

Photosynthesis is a membrane-based process, at least at the first energy storage phases. In photosynthetic eukaryotes, all main phases of photosynthesis take place in the thylakoid membranes of the chloroplast (Figure 1). There, two different light reactions occur simultaneously in Photosystems I and II (PSI & PSII), where light absorbed by the photosystem's pigment-protein complexes is efficiently converted into chemical energy(Blankenship, 2021).

The procedure can be summarised as follows: Light energy is captured by the reaction centres of photosystem II (PSII) and used to oxidise water and reduce plastoquinone (PQ) to plastoquinol. The reducing equivalents of plastoquinol (two electrons and two protons) move through the electron transport chain via the cytochrome b6f complex until they reach photosystem I (PSI) where they will be used to reduce ferredoxin, a small protein acting as an electron carrier between the light-dependent and the light-independent reactions of photosynthesis. In turn, reduced ferredoxin reduces Nicotinamide Adenine Dinucleotide Phosphate ($NADP^+$) to NADPH via the enzyme ferredoxin NADP oxidoreductase (FNR). The oxidation of water by PSII and the oxidation of plastoquinol by the cytochrome b 6/f complex causes protons (H+) to be pumped from the stroma into the thylakoid lumen , generating a proton gradient that drives the phosphorylation of ADP to form ATP via ATP Synthase (van Bezouw et al., 2019).

Consequently, NADPH and ATP are used to assimilate CO_2 from the atmosphere via the Calvin-Benson cycle. Finally, the fixed CO_2 is used for the production of the organic molecules essential for the plant, such as glucose, sugars and starch (Blankenship, 2021). The whole procedure is depicted in Figure 1.

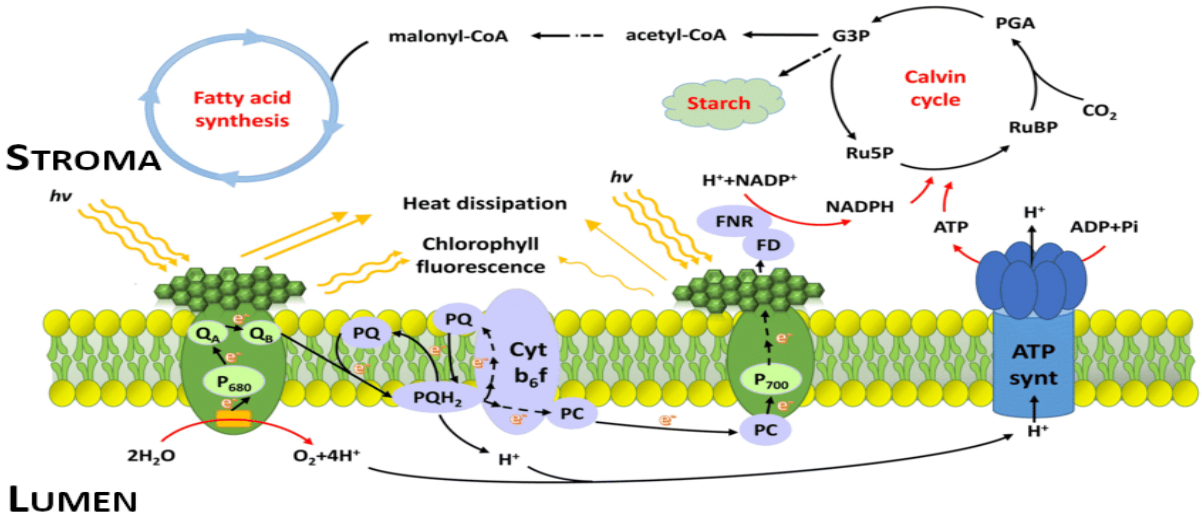


Figure 1 Light energy is used by photosystem II (PSII) to oxidise water and reduce plastoquinone(PQ) to plastoquinol. The reducing equivalents of plastoquinol move through the electron transport chain via the cytochrome b6f complex and are used by photosystem I (PSI) to reduce ferredoxin. In turn, reduced ferredoxin reduces $NADP^+$ to NADPH via the enzyme ferredoxin NADP oxidoreductase (FNR). The oxidation of water by PSII and the oxidation of plastoquinol by the cytochrome b $_6$/f complex

causes the release of protons (H^+) into the thylakoid lumen , generating a proton potential difference between the lumen and the stroma. Protons passing down this proton potential difference through the ATPase drive the phosphorylation of ADP to form ATP (adapted from Environmental Sciences Europe)

The importance of photosynthesis is paramount. Photosynthetic organisms not only form the basis of most food webs, but also produce the vast majority of the planet's oxygen (Britannica, 2022). As far as crop production is concerned, with the human population on the rise and projected to reach 10 billion by 2050 (Baillie & Zhang, 2018), there is an urgent need for improvement of crop quantity and quality. Improving photosynthesis could, potentially, lead to improved crop yield (Evans, 2013). However, it is a complex trait and controlled by a multitude of genes, many of which are still unidentified. Therefore, breeding for photosynthetic efficiency is a challenging task.

Photosynthetic Efficiency and High-Throughput Phenotyping

The photosynthetic efficiency is not perfect. Only a fraction of the photons absorbed by Photosystem II is used to fuel photochemistry. This fraction of photons is expressed as the quantum yield of Photosystem II (ΦPSII). The rest of the light energy absorbed by the plant is dissipated as heat. This loss of energy as heat can be induced by the plant, in a process known as Non-Photochemical Quenching (NPQ), a mechanism used by the plant as protection against excess sunlight. The fraction of photons dissipated through NPQ are expressed as the quantum yield of Non-Photochemical Quenching (ΦNPQ). Additionally, part of the light energy is constitutively lost as fluorescence or heat. This fraction of absorbed photons that are dissipated through fluorescence or other non-photochemical processes is expressed as the quantum yield of non-photochemical energy dissipation (ΦNO). Together, ΦPSII, ΦNPQ and ΦNO represent the total light energy absorbed by Photosystem II so that the sum of the three traits always equals to 1 (Müller et al., 2001)(Kramer et al., 2004)

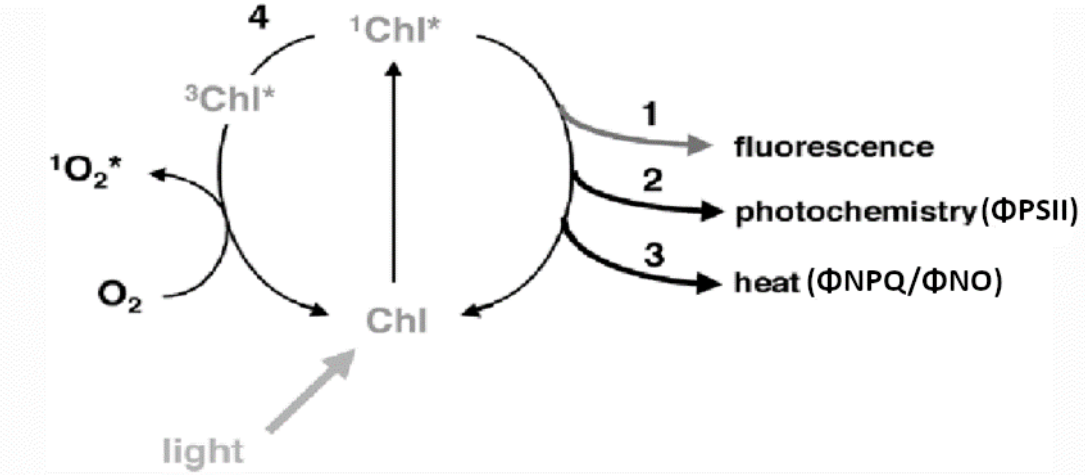


Figure 2 When Chlorophyll absorbs light it is converted to its excited state. It can return to its lower energy state via three ways: 1) by emitting photons as fluorescence, 2) by fuelling photochemistry, 3) by being dissipated as heat, either in an induced way or passively. Excited chlorophyll can also 4) produce triplet excited ^{3}Chl via intersystem crossing and by doing so, create Reactive Oxygen Species (adapted from* Müller et al., 2001*)*

Photochemistry, heat dissipation and fluorescence competitively contribute to the de-excitation of chlorophyll (Figure 2). Thus, by measuring the fluorescence emitted during the process the rest of the parameters can be measured, in a process known as chlorophyll fluorescence imaging.

In Wageningen University, these measurements can very efficiently be performed by the Phenovator II, an automated, high-throughput phenotyping platform. The system is capable of screening up to

1440 *A. thaliana* plants multiple times per day. It measures the following chlorophyll fluorescence parameters:

- Fm: maximal fluorescence, represents the maximum fluorescence emitted when all photosystem II (PSII) reaction centres are closed, and provides a reference point for measuring other fluorescence parameters
- Fo: basal fluorescence, represents the fluorescence emitted by open PSII reaction centres in the dark, and reflects the degree of PSII photoinactivation or damage
- Fv: variable fluorescence, represents the difference between maximal fluorescence (Fm) and basal fluorescence (Fo), and indicates the amount of light energy that can be absorbed and used for photochemistry by photosynthetic organisms
- Fp: effective quantum yield of photosystem II, represents the fraction of light energy that is absorbed and used for photochemistry by PSII, and reflects the efficiency of photosynthetic electron transport.
- Fmp: maximal fluorescence in the light, represents the maximum fluorescence emitted when all PSII reaction centres are open under high light conditions, and is used to estimate the rate of electron transport and photosynthetic efficiency.

These, then can be used to calculate ΦPSII, ΦNPQ and ΦNO. High Throughput Phenotyping creates vast amounts of Phenotyping data, which can then be used in Quantitative Genetics.

Genome Wide Association Studies on Photosynthesis

An efficient way of utilising the High Throughput Phenotyping data is in a Genome-Wide Association Study (GWAS). The purpose of the GWAS is to identify genetic variants or markers that are associated with a particular trait or phenotype of interest. It is a powerful tool for studying the genetic basis of complex quantitative traits and discovering the Quantitative Trait Loci (QTLs) underlying those traits (van Bezouw et al., 2019).

In GWAS, the genomes of individuals with a particular phenotypic trait are compared to those without the trait, looking for genetic markers that are significantly more common in one group compared to the other. These markers are usually Single Nucleotide Polymorphisms (SNPs). Hence, GWAS is a correlation between phenotypic and genetic variation in order to identify the genetic locus responsible for this variation (Brachi et al., 2011).

While GWAS is sufficient in locating QTLs, several tests, such as Linkage Disequilibrium analysis and Haplotype analysis must follow in order to determine the QTL interval and thus the most relevant candidate genes.

GWAS was performed to study photosynthesis in A. thaliana by two of my predecessors, who were led to the discovery of several QTLs and candidate genes, which will be further studied in this project.

Former student Jacky To used data from 4 different Genome-Wide Association Studies performed on the *A. thaliana* Dutch population (DartMap) and the Global population (HapMap). Two of them were performed on the DartMap population by Nguyen et al. for the traits ΦPSII and ΦNO under fluctuating light. One was performed by Boesten et al. for ΦPSII under stable light. Finally, one was performed on the HapMap population by Prinzenberg et al. (2019) These analyses yielded a QTL associated with the traits ΦPSII and ΦNO in chromosome 3. Subsequent tests, such as Linkage Disequilibrium and Haplotype analysis yielded a total of 14 candidate genes as the most likely to be causal. For 10 of these, T-DNA insert lines with the insert within the gene or the gene's promotor were available (Table 1).

Former student Laura Bos Calderó performed various GWAS analyses on data obtained from High Throughput Phenotyping of the *A. thaliana* Dutch population. The analyses resulted in three notable QTLs for the traits ΦPSII and ΦNO, only one of which was considered significant enough for additional studying (Figure 3). This QTL, situated on chromosome 2 positions 7818000-7853000, contains 14 genes ranging from AT2G17970 to AT2G18070. However, only 9 out of the 14 genes contained or were within close distance of SNPs in high LD with the SNP of interest. Haplotype analysis further limited the number to 8 genes (Table 1).

In addition to these genes, an extra candidate gene was added in by Phuong Nguyen. AT1G61860 was the SNP of interest of a past GWAS analysis and was included as an additional try.

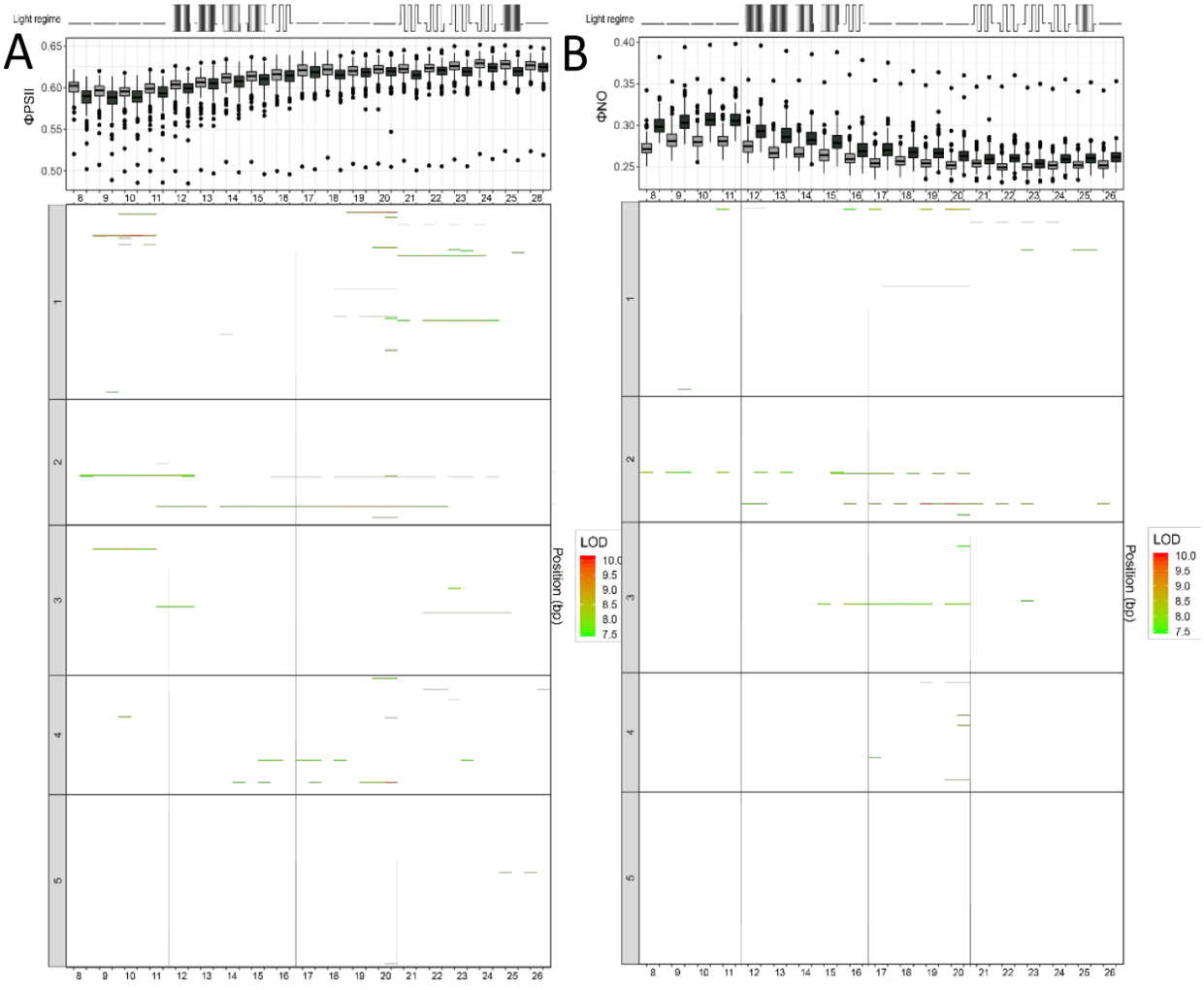


Figure 3 Example of a plot depicting the output of two GWAS analyses overtime. The traits studied are ΦPSII (figure A) and ΦNO (figure B). The X axis displays time in Days After Sowing. The Y axis depicts the 5 Arabidopsis chromosomes. The bars represent QTLs and the significance of the SNP-trait association is color-coded as the LOD score. Longer bars means that the SNP was significant with the trait for a longer period of time and so represent QTLs more likely to be affecting the trait. (Made by Laura Bos Calderó)

Table 1 The 19 candidate genes tested and the proteins they code for. All the information on the table comes from The Arabidopsis Information Resource – TAIR (Berardini et al., 2015), unless specified otherwise

Gene Name	Protein Coded	Protein Function
AT1G61860	Protein kinase superfamily protein	Transfer of a phosphate group from ATP/GTP to its protein substrate (protein phosphorylation) (Cheng et al., 2011)
AT2G17970	2-oxoglutarate/Fe (II)-dependent (2OG) superfamily protein	protein modification, nucleic acid modification and repair, secondary metabolite synthesis (Jia et al., 2017)
AT2G18010	Small Auxin Up-regulated RNA - like protein 10 (SAUR10)	Auxin-responsive growth and development. Could be induced by other hormones as well (Stortenbeker & Bemer, 2019)
AT2G18030	Methionine Sulfoxide Reductase A5 (MSRA5)	MetO reduction to Methionine, protection from oxidative stress (Rey & Tarrago, 2018)
AT2G18040	Peptidylprolyl cis/trans Isomerase (PIN1AT)	Conversion of phosphoserine-proline bonds between cis and trans forms (Landrieu et al., 2000)
AT2G18042 (pseudogene)	Non-functional ECA1 gametogenesis family protein	Facilitation of reaction between male/female gametes (Sprunck et al., 2014)
AT2G18050	Linker Histone protein (HIS1-3)	Structural organisation of DNA (Fyodorov et al., 2018) Regulation of gene expression (Laybourn & Kadonaga, 1991)
AT2G18060	Arabidopsis NAC-domain Containing protein 37 (ANAC037)	Plant growth, plant development, abiotic stress response, pathogen immunity (Yuan et al., 2019)
AT2G18070	Hypothetical protein	Unknown
AT3G45450	Nucleoside triphosphate hydrolase superfamily protein	ATP hydrolysis
AT3G45620	Predicted Transducin/WD40 repeat-like protein	Scaffolding protein, Involved in germination, growth, stress responses (Gachomo et al., 2014)
AT3G45630	RNA binding (RRM/RBD/RNP motifs) family protein	Binding to single-strand RNA (Lunde et al., 2007)
AT3G45650	Nitrate Excretion Transporter 1 protein (NAXT1)	Passive nitrate transport, nitrate excretion from roots (Segonzac et al., 2007)
AT3G45660	NAXT NPF superfamily protein	Transmembrane nitrate transport
AT3G45670	Protein kinase superfamily protein	Transfer of γ-phosphate groups from ATP/GTP to its protein substrates (Cheng et al., 2011)
AT3G45740	Haloacid dehydrogenase (HAD) superfamily of proteins	Nucleophilic substitution of carbon-halogene bonds, various hydrolytic enzyme activities (Caparrós-Martín et al., 2013)
AT3G45760	Nucleotidyltransferase Protein 5 (NAT5)	Transfer of nucleotides to the 3' end of RNA
AT3G45770	Mitochondrial enoyl-ACP reductase	Fatty acid biosynthesis
AT3G45780	Phototropin1 (PHOT1)	Blue light photoreceptor kinase for photo-induced movements like stomatal opening and root phototropism (UniProt Consortium, 2023)

T-DNA mutants and their use on gene function studies

The discovered candidate genes need to be tested in order to identify which ones -if any- are causal genes. One of the downstream approaches to identifying causal genes is through the use of mutant plants. This approach includes the genotyping and phenotyping of mutant lines, and has been used successfully for identifying genetic loci and genes essential for photosynthesis and plant adaptation to light changes (Cruz et al., 2016). For *Arabidopsis*, T-DNA insertion lines are commercially available for large number of genes, which facilitates the workflow.

T-DNA stands for transfer-DNA. The T-DNA fragment is inserted into the target organism's genome via a process where *Agrobacterium tumefaciens* infects the plant and transfers part of its DNA through a tumour-inducing plasmid (Tzfira & Li, 2004). The T-DNA is integrated randomly in the host's genome, thus, it is a highly efficient gene disrupting tool (Krysan et al., 1999). This, in combination with the fact that there is a wide range of protocols available for Arabidopsis T-DNA transformation, makes possible the creation of large mutant libraries for studying the function of virtually any gene (Clough & Bent, 1999) (O'Malley et al., 2015). Ideally, a T-DNA insertion should result in the full knockout of the disrupted gene but this is not always the case. The effectiveness of the insert depends on a variety of factors. A key factor is the insert's location on the gene. According to Wang (2008), if the insert lies in the protein coding region there is a 86% chance it knocks out the gene. However, this percentage drops to 41% if the T-DNA has been integrated in a region before the start codon. Usually, insertions before the start codon create a knockdown instead (53% of the time). Insertions after the stop codon, may also not create a knockout and 17% of the time, they do not affect gene expression at all. Thus, even if the insert lines are confirmed as homozygous for the insert by the producing company, they need to be genotyped.

Hence, the aim of this project is to identify the most promising candidate genes based on the phenotypic variation between the mutants and their wild type genetic background. Nienteen different genes will be tested across 3 QTLs: a single gene on the QTL of chromosome 1, 8 genes on the QTL of chromosome 2 and 10 genes on the QTL of chromosome 3. Mutants of possible causal genes are expected to exhibit statistically significant alterations in their photosynthetic efficiency compared to the wild type plants. The altered phenotype is dependent on the type of mutant and the effect of the gene on photosynthesis. For instance, a knockout mutant for a gene positively affecting photosynthetic efficiency will likely have reduced ΦPSII values compared to the wildtype. On the contrary, an overexpressing mutant of the same gene is expected to be more efficient in its use of light energy, at least to the extent that the overexpressing gene affects the process.

Materials and Methods

Plant material

Twenty-three T-DNA lines were used. The lines were provided by Phuong Nguyen and obtained from the Nottingham Arabidopsis Stock Centre (NASC, Table 2). In addition, lines Columbia-0 and Columbia-3 were used as wildtype controls.

Table 2 The 23 T-DNA insert lines used. e Short ID is arbitrarily assigned in order to easily identify the lines. Each line's genetic background, insert position and zygosity status are provided

Gene	***Insertion***	***Line ID***	***Short ID***	***Background***	***Zygosity***
AT1G61860	Exon	SALK_130390C	7	Col-0	Confirmed
AT2G17970	Exon	SALK_111811C	8	Col-0	Confirmed
AT2G18010	Falls on this gene and AT2G18020	SALK_065537	4	Col-0	Segregating
AT2G18030	Intron	SALK_201557C	18	Col-0	Confirmed
AT2G18030	Exon	SAIL_47_C10	21	Col-3	Segregating
AT2G18040	3' UTR	SAIL_15_A04	20	Col-3	Segregating
AT2G18042	Exon	SALK_064492C	13	Col-0	Confirmed
AT2G18050	5' UTR	SALK_025209C	12	Col-0	Confirmed
AT2G18060	3' UTR	SALK_022534C	14	Col-0	Confirmed
AT2G18070	Exon	SALK_017125C	17	Col-0	Confirmed
AT3G45450	Exon	SAILseq_828_A03.1	23	Col-0	Homozygous
AT3G45620	Exon	SALK_026980C	16	Col-0	Confirmed
AT3G45630	Exon	SALK_061949C	15	Col-0	Confirmed
AT3G45650	Exon	SALK_091226C	11	Col-0	Confirmed
AT3G45660	Promotor	SALK_030923	3	Col-0	Segregating
AT3G45670	Exon	SALK_201949C	19	Col-0	Confirmed
AT3G45740	Exon	SALK_118254	5	Col-0	Segregating
AT3G45740	Intron	SALK_118255C	10	Col-0	Confirmed
AT3G45760	Exon	SALK_018808	2	Col-0	Segregating
AT3G45760	1st variant Exon, 2nd variant 3'UTR	SALK_109748C	9	Col-0	Confirmed
AT3G45770	Intron	SALK_003308	1	Col-0	Segregating
AT3G45770	Exon	SALK_130583	6	Col-0	Segregating
AT3G45780	Exon	SAIL_1232_C01	22	Col-0	Homozygous

Seed preparation and sowing

A few seeds (~20) from each line were placed on small size petri dishes, on filter paper covered with 1 ml demineralized water. Care was taken for the seeds to be spread uniformly throughout the filter paper surface area in order to avoid trouble with picking up the small seedlings after germination. Then, the petri dishes were stored in a labelled container which was placed in the cold chamber (4^{o}C), where the seeds were left in the dark to stratify for 48 hours. Once the stratification of the seeds was complete, they were moved to a growth chamber to germinate. The growth chamber conditions were: 24^{o}C and a 16h/8h light/darkness cycle (light from 7:00 to 23:00). After around 48 hours of germination the seeds were moved to the greenhouse where they were implanted on rockwool blocks (1 seed per block). Hyponex was used as a growth medium. The seeds were planted as follows:

- From segregating lines 12 seeds/line
- From lines confirmed as homozygous 6 seeds/line
- From lines homozygous but not confirmed 10 seeds/line

Each block was numbered with the Short ID (Table 2) plus an individual seed number. Each line was labelled with GMO red labels.

Some of the seeds failed to germinate. As a result, a new seed batch was prepared from scratch, the only difference being that a larger container box was used, where all the seeds could be near the bottom of the container (this could perhaps aid stratification/germination) and no more than 2 petri dishes were stacked on top of each other. The box was put in the cold chamber for stratification for 48 hours and then in the growth chamber for germination for another 48 hours. This new batch was used three times to replace seeds that did not germinate in the greenhouse.

Genotyping for insert

In order to genotype the T-DNA mutants, the zygosity of the plants had to be determined. For this, DNA is extracted from plant tissue and the insert area is PCR amplified by using insert-specific primers (O'Malley & Ecker, 2010). These primers are the Left and Right Genomic Primers (LP,RP) and the T-DNA border primer (BP). Thus, when the PCR product is run in a gel, the user gets one band (LP to RP) if there is no insert. If the plant is homozygous for the insert, the user also gets one band (RP to Insertion site plus 110 bp from BP to LP). If the plant is heterozygous for the insert, both bands will appear on the gel (SIGnAL : Salk Institute Genomic Analysis Laboratory), (Figure 3).

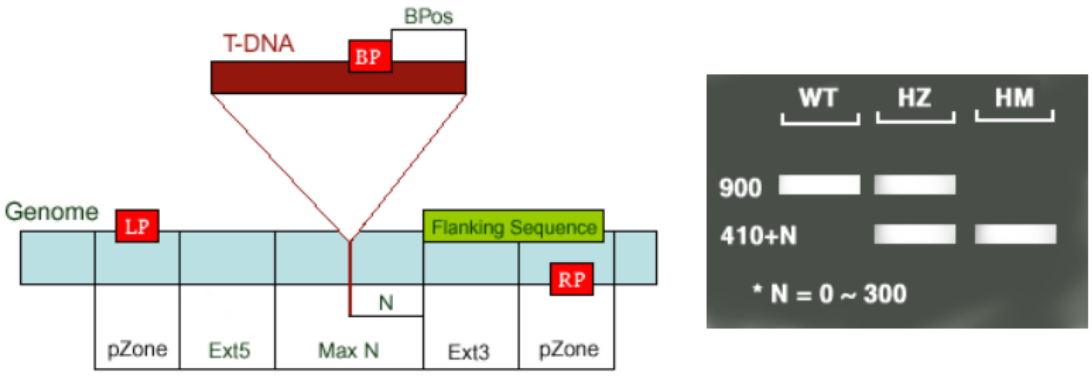


Figure 4 Schematic of the T-DNA insert where: N - Difference of the actual insertion site and the flanking sequence position, usually 0 - 300 bases, MaxN - Maximum difference of the actual insertion site and the sequence, default 300 bps, pZone - Regions used to pick up primers, default 100 bps, Ext5, Ext3 - Regions between the MaxN to pZone, reserved not for picking up primers, LP, RP - Left, Right genomic primer, BP - T-DNA border primer LB - the left T-DNA border primer and BPos - The distance from BP to the insertion site. On the right, the bands appearing after gel electrophoresis for Wild Type, Heterozygous mutant and Homozygous mutant (http://signal.salk.edu/tdnaprimers.2.html)

Leaf material, as well as flower buds were collected during the first 2 weeks after sowing and DNA was extracted using a version of the CTAB protocol, where each cleaning step was only performed once (Supplementary Text 1). This was intended to save time.

Following extraction, DNA concentration and quality was checked with the Nanodrop, looking for a single absorbance peak at 260 nm and a 260/280 absorbance ratio of 1.8-2.0.

After that, primers for the PCR were designed. The primers were automatically designed by the SALK T-DNA Primer design tool (Http://signal.salk.edu/tdnaprimers.2.html). However, primers were also manually designed for three out of the twenty-three lines (Short IDs 2,13 and 18). These primers were designed using the NCBI Primer-BLAST tool (Ye et al., 2012). First, the coding sequence for the disrupted gene of each line was recovered using the SALK Genome browser tool (Http://signal.salk.edu/atg1001/3.0/gebrowser.php). Then, the exact location for each insert was found and the primers were designed around it, in such a way that the LP-RP product is around 1100-1200 bp long. The gene sequence was inserted in primer-BLAST with the following parameters altered: PCR Min product size was set to 1100 and Max size was set to the total length of the template sequence. Number of primers to return was set to 20, in order to have multiple primer pairs to check for suitability. The organism was set to *A. thaliana*. All other parameters were left on default. Several primer pairs were selected for each gene and tested for suitability. First, each primer pair was blasted against "Refseq representative genomes" to find if the pair amplified another target of similar size. Then, each pair was tested for problematic structures such as hairpins, self-dimers and heterodimers. This was accomplished with IDT's OligoAnalyzer Tool (Https://www.idtdna.com/calc/analyzer). The full list of primers used can be found in Supplementary Tables 3 & 4.

Subsequently, the PCR was performed. A standard PCR protocol was used. First, all reagents were thawed on ice. Then, a reagent Master Mix was prepared, in a 1.5 ml tube and according to the number of reactions performed per PCR (number of individuals per line + wild type control + pipetting error). In some cases, 2 wild type controls were used. Additionally, for lines 2, 7, 12 and 17, a separate PCR was used as a positive control: all individuals were amplified with primers from line with Short ID 5,

which had given strong, clear bands in electrophoresis (Supplementary figures 7, 12, &17). Reagents were added in the Master Mix in the following order: MQ water, buffer (5X Green), dNTPs, primers, Taq polymerase. Next, the Master Mix was briefly vortexed and centrifuged in order for the tube contents to settle. Afterwards, 9 µl from the mix were added in thin walled 0.2 ml PCR tubes and 1 µl DNA template was added in each tube, except for the negative control tubes (10 µl reaction).

The following program was used:

Step	Temp	Time	# of cycles
Initial Denaturation	95°C	3 min	
Denaturation	95°C	30 sec	30-35
Primer Annealing	Tm-5°C	45 sec	
Extension	72°C	1 min per kb	
Final Extension	72°C	10 min	

Several variations of the program were run and are included in the Appendix (Supplementary Table 1).

Three different primers were used for each line: the primers targeting left and right of the target region (LP and RP), and a border primer situated in the insert DNA (BP). Two types of PCR were run: a) a single reaction per line containing all three primers at once and b) two separate reactions per line, where the first reaction contained the LP-RP primer pair and the second contained the BP-RP primer pair. The different reaction mixes used can be found in the Appendix (Supplementary Table 2).

After amplification PCR products were run in a 1% agarose gel with 0.005% ethidium bromide. A 1kb DNA ladder was also loaded for band size measurement. As described in the introduction, if there is no insert a band will appear with size equal to the max product size for each primer pair. If the plant is homozygous for the insertion a single band will appear as well, but smaller than the full product size. If the plant is heterozygous 2 bands will appear. Also, the gel should show no substantial band shearing or contamination, indicating good quality DNA.

Seed harvesting and resowing

The following step was to harvest seeds from Arabidopsis plants confirmed as homozygous for the insertions. The time of harvest varies and depends on several factors affecting plant development, such as light, temperature, nutrition and genetic makeup. Generally, when these conditions are optimal, the Columbia genotype seeds can be harvested within 8 weeks from sowing (Arabidopsis Biological Resource Center). The mutants however may grow on different rates, depending on which genetic loci have been disrupted by the inserts. In any case, seeds were harvested after siliques had browned completely. The plants were not watered after siliques started to turn brown. Carefully, the whole plant was cut at its base and gently tapped on a piece of paper, where the seeds were collected. Finally, the seeds were sieved through two separate sieves to remove any other plant tissue debris left on them (Arabidopsis Biological Resource Center)(Mulligan & Russell), and placed in small plastic bags, labelled with the line Short ID and the individual number. Homozygous seeds were then planted in rockwool blocks (several plants per genotype in a randomized block design) in the Phenovator II B8 room for High Throughput Phenotyping.

Gene expression analysis

To confirm the presence of the insert in the homozygous mutants and check the expression levels of the disrupted genes (True Knockouts or Knockdowns) an RT-qPCR Reaction would have to be performed. Leaf material was obtained from each plant. The leaves were grinded as described in the PCR protocol, and then RNA was extracted and purified using the ZymoResearch Direct-zol RNA Miniprep Kit (Supplementary Text 2). Afterwards, RNA was subjected to DNAse treatment and precipitation (Supplementary Text 3).

The RNA quality was checked in the Nanodrop. However, it was not good enough to proceed to Reverse Transcription and time constraints did not allow for a repeat of the process. Hence, the gene expression analysis was stopped at this point.

High-Throughput Phenotyping

Plants were grown in the Phenovator II B8 room, arranged in a randomised block design. From the 1st to the 12th Day After Sowing, the plants were exposed to constant light with an intensity of 300 umol $s^{-1}m^{-2}$ and a 12 hour light/darkness cycle. From the 13th to the 16th Day After Sowing the plants were exposed to fluctuating light, alternating between 100 umol $s^{-1}m^{-2}$ and 900 umol $s^{-1}m^{-2}$ every 15 minutes. From the 17th to the 19th Day After Sowing plants were again exposed to constant light of 300 umol $s^{-1}m^{-2}$. Finally, for the 20th and 21st Day After Sowing a fluctuating light was applied, alternating between 100 umol $s^{-1}m^{-2}$ and 900 umol $s^{-1}m^{-2}$ every 60 minutes.

On the 6th Day After Sowing, the Phenotyping started. The Phenovator measured Fp and Fmp twice per day and Fm, Fo and Fv once during the night. These were used to calculate the photosynthetic efficiency parameters Fv/Fm, ΦPSII, NPQ, ΦNPQ and ΦNO (Table 3). Additionally, plant size was measured once per day and once per night. The results were extracted as .csv files in Excel format.

Table 3 Formulas for the calculation of the photosynthetic efficiency parameters

ΦPSII	**(Fmp-Fp)/Fmp**
NPQ	**(Fm/Fmp) - 1**
ΦNPQ	**(Fp/Fmp)-(Fp/Fm)**
ΦNO	**Fp/Fm**

Subsequently, data processing was performed to make the data more usable and calculate additional parameters. The data in the files was sorted in the same order and the Fv/Fm datasheet was doubled to match the number of ΦPSII measurements. Some rows of data in the ΦPSII datasheet were missing, because of the Phenovator's failure to record data for some trays on the 25th of November, and these were filled with "NA". The ΦPSII and Fv/Fm datasheets were merged and Non-Photochemical Quenching (NPQ), the ratio of light energy that is dissipated via NPQ (ΦNPQ) and the ratio of light energy that is dissipated via unregulated processes (ΦNO) were calculated as described in Table 3. The data was then converted from a vertical to a horizontal format and the measurements file was merged with the conversion key file. False data of missing plants were replaced with "NA". Finally, the genotyping results were added in.

The genotyping data was not added earlier due to the fact that I was unable to genotype the majority of the plants before the start of phenotyping. There was no option to delay the start of Phenotyping

or change the start date and, as a result, most plants were added to the randomised block design without information on their genotype. This posed a challenge, since many of the phenotyped plants could very likely be non-mutants and, thus, useless for the experiment. At the same time, this meant a smaller number of mutated plants would be phenotyped, potentially affecting the statistical significance of the results. Despite these problems, the project continued.

Student's t-test and plotting data

The Student's t-test was performed, to compare the phenotyping data of the homozygous mutants of each line to those of its background. All lines were tested against Columbia-0, except for lines 20 and 21, which were tested against Columbia-3. A p-value was calculated to determine if the observed phenotypical difference between the mutants and their background was statistically significant. Differences with a p-value less than 0.05 were considered to be significant.

Promising lines were selected for further analysis and the photosynthetic efficiency trait values were plotted over time in boxplots. Each boxplot displayed the measurements for a trait of the homozygous individuals of mutant line and those of the wildtype, with the light treatment depicted above the boxplots as differently coloured bars. This design allowed for a comparison of the trait performance between the mutant and the wildtype, as well as an assessment of the potential effect of the light treatment on the trait. The graphs were designed in RStudio, using the ggplot2 and tidyr packages (Supplementary Text 4).

Results

Genotyping for insert

In order to determine the genotype of the T-DNA lines, the DNA of each individual of each line was extracted. Afterwards, it was PCR amplified using a combination of 3 primers: the Left Primer (LP) and Right Primer (RP), which targeted genomic DNA sequences flanking the insert and the Border Primer (BP), which targeted a sequence on the insert. The amplified products were then analysed by gel electrophoresis.

Initially, a single PCR reaction was performed per line, containing all three primers. This was intended to save time. However, in most cases, the reaction for the wild type allele was the only one that took place. In others, both reactions worked but exhibited faint bands and/or artifacts. Lines 1 through 6, as well as lines 20 and 21, all characterised by NASC as segregating, were initially tested using this method.

After the method proved unsuccessful, lines were instead tested using the two separate PCR reactions, one for the wild type allele and one for the mutant allele. The results were significantly improved on the second attempt, although some minor inconveniences persisted. In some cases, amplification of the control, which contained only wild type plant DNA, was unsuccessful, even though the rest of the individuals' DNA was amplified as expected. Thus, the reaction had to be repeated. This was most likely due to a poorer quality of the wild type DNA samples, as not all individuals were extracted on the same day. Subsequently, two negative controls, containing only wild type DNA, were included for each line, in order to avoid retesting the line, in case of a failed amplification.

On other occasions (Lines 2,7,12,17), the majority or all of the samples failed to produce a result during the PCR reaction. To determine if the issue was with the quality of the DNA, the reaction was repeated including positive controls. As positive controls, all individuals of the line were amplified with the wild type primers designed for line 5, the suitability of which, had been confirmed (Figure 5). Line 2, in particular, posed a significant challenge. Initial attempts to amplify failed completely, as none of the individuals produced any bands during amplification attempts. To overcome this, the line was amplified with the primers from line 5, revealing that the DNA quality was adequate. Next, a gradient PCR was conducted to identify the optimal temperature for the primers but unfortunately, the reaction failed across the whole temperature range. It was concluded that the issue, most likely, lay with the primers used and was resolved with the manual design of new primers.

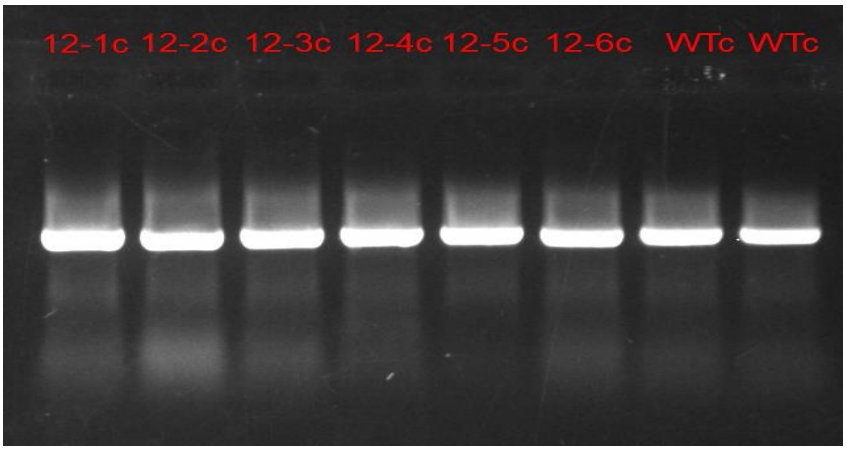


Figure 5 The individuals of Line 12 tested with wild type primers designed for Line 5. There are clear bands indicating that the extracted DNA is of good enough quality to be used in PCR

On several instances, lines 13 and 18 presented difficulties with amplification. Despite the individuals being homozygous, as expected, the negative controls failed to show any results (Supplementary figures 13 &18). Based on the facts that a) the rest of the lines characterised as “Homozygous Confirmed” were largely as expected and b) the negative control, containing only wild type DNA, had failed to amplify on previous occasions it was decided to treat these lines as homozygous for the insert.

Finally, in the case of segregating Lines 3, 5 and 20, no homozygous mutants were retrieved. Instead, all individuals were wild types (Supplementary figures 3, 5 & 20).

In total, the zygosity of the individual lines was largely in line with expectations. Lines 7-19, characterised by NASC as “Confirmed Homozygous” exhibited almost complete homozygosity with only a couple of exceptions (Lines 10 and 12). On the contrary, Lines 1, 2, 4, 6, 20 and 21, characterised by NASC as “Segregating” demonstrated a mix of homozygous mutants, heterozygous and wild type plants. A good example showcasing this is Line 4 (Figure 6). The majority of individuals in this line exhibit clear bands in both rows, indicating that they contain both the wild type allele of the gene, as well as the disrupted copy. Thus, these individuals are heterozygous for the mutation. Similarly, individuals 4-2 and 4-12 only have the disrupted copy of the gene, making them homozygous mutants. On the other hand, individuals 4-6, 4-8 and 4-9 are wild type plants, lacking any T-DNA integration. Finally, for plant 4-10, the reaction failed. Finally, Line 22 characterised as “Homozygous”, exhibited only homozygous individuals.

A few results deviated from expectations. These were “Segregating” Lines 3, 5 and 20, which only had wild type plants, “Homozygous” Line 23, were all individuals were heterozygotes and “Homozygous Confirmed” Lines 10 and 12, which were, in fact, segregating.

Complete results of genotyping can be found in Table 4.

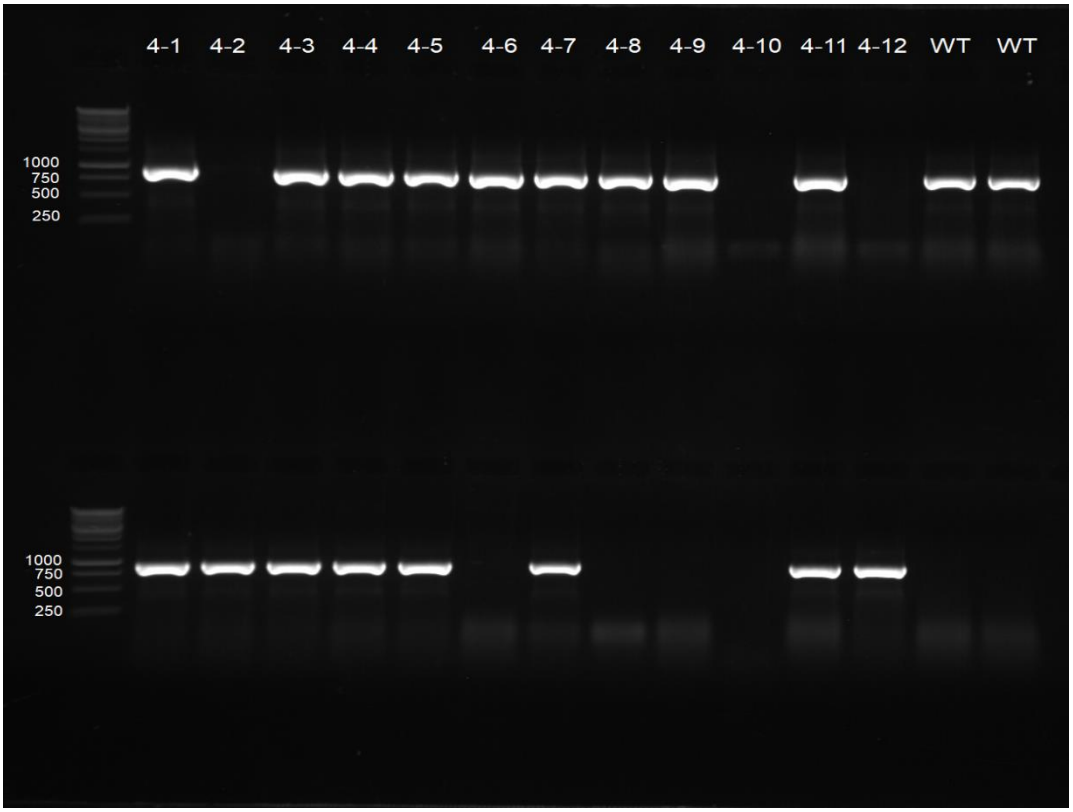


Figure 6 Gel results for Line 4. The upper gel represents the Left primer-Right Primer reaction, genotyping for the wild type allele. The lower gel shows the Border Primer-Right Primer reaction, genotyping for the mutant allele. DNA ladder is shown in

the first lane with size (in bp) indication for several bands. The following lanes are for tested individuals of line 4, indicated by number 4-1 to 4-12. The last two lanes are for wild type samples.

Table 4 Homozygosity screening for T-DNA lines. First Row: Individual number. First column: Line Short ID, second column: Gene ID, last column: Conclusion for the line's zygosity. HMZ: Homozygous mutants. HTZ: Heterozygous for the insertion. WT: Wild Type plants. NR: No Reaction, the PCR failed or was inconclusive. NP: No Plant, the individual's seed failed to germinate

		1	2	3	4	5	6	7	8	9	10	11	12	CONCLUSION
7	AT1G61860	HMZ	HMZ	HMZ	HTZ	HMZ	HMZ							Segregating
8	AT2G17970	HMZ	HMZ	NR	HMZ	HMZ	HMZ							Homozygous
4	AT2G18010	HTZ	HMZ	HTZ	HTZ	HTZ	WT	HTZ	WT	WT	NR	HTZ	HMZ	Segregating
18	AT2G18030	NP	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
21	AT2G18030	WT	HTZ	HMZ	HMZ	HMZ	WT	HMZ	HMZ	WT	HMZ	HMZ	HMZ	Segregating
20	AT2G18040	WT	WT	WT	WT	WT	WT	WT	NP	WT	WT	WT	WT	Wild Type
13	AT2G18042	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
12	AT2G18050	HTZ	HMZ	WT	HTZ	HTZ	HTZ							Segregating
14	AT2G18060	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
17	AT2G18070	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
23	AT3G45450	HTZ	NP	HTZ	HTZ	HTZ	HTZ	HTZ	HTZ	HTZ	HTZ			Segregating
16	AT3G45620	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
15	AT3G45630	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
11	AT3G45650	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
3	AT3G45660	WT	WT	NP	WT	WT	WT	WT	NP	WT	WT	NP	WT	Wild Type
19	AT3G45670	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
5	AT3G45740	WT	WT	WT	WT	WT	WT	WT	NP	WT	WT	WT	WT	Wild Type
10	AT3G45740	HMZ	HMZ	HTZ	HTZ	HMZ	WT							Segregating
2	AT3G45760	HTZ	HMZ	WT	WT	WT	WT	HTZ	HMZ	HTZ	HTZ	WT	WT	Segregating
9	AT3G45760	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ							Homozygous
1	AT3G45770	HTZ	NR	NR	NP	WT	WT	HMZ	HTZ	NR	NR	HMZ	HMZ	Segregating
6	AT3G45770	WT	HMZ	HTZ	HMZ	HMZ	HMZ	HTZ	HMZ	HMZ	HMZ	HMZ	NR	Segregating
22	AT3G45780	HMZ	HMZ	HMZ	HMZ	HMZ	HMZ	NP	HMZ	HMZ	HMZ			Homozygous

High Throughput Phenotyping

The photosynthetic phenotypes of the T-DNA insert lines were screened by the High Throughput Phenotyping platform Phenovator II, in order to evaluate the effect of the insertion on the function of the candidate genes. The phenotypic variation between each insert line and its background line was determined by Student's t-test, for which a p-value of 0.05 or less was considered statistically significant. This means that the probability of the observed difference between the line and the background being due to chance is less than 5%, and therefore, it can be concluded that the deviation is statistically significant and not random.

The t-test for the comparison between each line and the background is conducted per timepoint. Each timepoint represents a measurement of a parameter by the Phenovator. The platform measured 5 parameters (ΦPSII, ΦNO, NPQ, ΦNPQ and Size-ΦPSII) twice per day for a total of 16 days, which amounts to 160 measurements. The platform also measured 2 parameters (Fv/Fm and Size-Fv/Fm) once per day but these values were doubled to match those of the rest of the parameters, amounting to 64 timepoints. So, in total 224 measurements from each line were compared to the 224 corresponding measurements of the background. Table 5 shows the 32 measurements taken for the trait ΦPSII for 6 of the Lines.

It is important to note that the difference between each line and the wild type was considered significant only depending on the trait being examined. For example, while plant size differences can be an indicator of deviation from the background, it is not that important in this particular context, as it can be attributed to multiple reasons beyond photosynthetic efficiency. Therefore, more emphasis was placed on the lines that exhibited significant p-values for the photosynthetic traits, especially for ΦPSII and ΦNO, which were the traits studied in the Genome-Wide Association Studies.

T-DNA lines for candidate genes of QTL on Chr 3

The first line of interest was Line 9, an insert line for the gene AT3G45760, which codes for a nucleotidyltransferase involved in 3' RNA end processing. The gene was pinpointed by a previous student who studied the QTL on chromosome 3, as the most likely to be the causal gene underlying the QTL.

The line exhibited significant deviation from its background for 170 out of the 224 timepoints of measurement. It showed alternate phenotypes across the board, deviating from the wildtype in all traits studied.

The Fv/Fm parameter was higher for this insert line than the background. It was also observed that the significant difference between the line and the background only appears under constant light and seems to disappear when the fluctuating light is applied (Figure 7). This was also the case for the ΦPSII parameter (Figure 8).

Additionally, both NPQ and ΦNPQ of Line 9 were higher than the background, regardless of the light treatment. However, for both lines, there appears to be a notable difference between the morning and afternoon values of these traits, during the first Days After Sowing (Figures 9 & 10).

On the contrary, ΦNO, is lower in the insert line than the wild type. Here, too, a difference between morning and afternoon values is observed, early in the treatment. Also, the differences seem to exist regardless of the light (Figure 11).

Finally, plant size was a clear difference exhibited only by Line 9. The plants were much smaller than their wild type counterparts (Figure 12). The line also exhibited other non-photosynthesis related alterations, such as slow growth and development, partial sterility and larger seed size.

Table 5 Results of a t-test comparison between each of the significant lines and its background, per timepoint, for the trait ΦPSII. Highlighted boxes are p-values <0.05. Highlight is red if the ΦPSII of the mutant is lower than that of the wild type and green if the ΦPSII of the mutant is higher than that of the wildtype.

TIMEPOINT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
LINE 2	0.214697	0.030165	0.039897	0.025848	0.031147	0.009125	0.007491	0.002874	0.010438	0.015625	0.047317	0.027183	0.037619	0.072471	0.140975	0.23361
LINE 6	0.056398	0.017592	0.062285	0.074943	0.070478	0.103015	0.102476	0.017874	0.007583	0.045629	0.033048	0.022654	0.064198	0.036545	0.043936	0.039735
LINE 7	0.313324	0.558277	0.552123	0.833912	0.411908	0.434838	0.148278	0.07456	0.038032	0.056764	0.067183	0.033235	0.050021	0.030129	0.06162	0.050414
LINE 9	2.2E-06	7.03E-05	9.14E-05	0.069224	0.290453	0.325499	0.136515	0.923801	0.131644	0.11606	0.02774	0.015032	0.001317	0.006065	0.020772	0.07609
LINE 16	0.016021	0.007698	0.054967	0.009681	0.360524	0.01736	0.996351	0.013989	0.065988	0.010435	0.148086	0.027589	0.269301	0.023293	0.10315	0.015763
LINE 18	0.276668	0.116587	0.035616	0.215084	0.116941	0.157678	0.068953	0.055302	0.019048	0.051535	0.023063	0.036969	0.052251	0.034119	0.011945	0.034434
TIMEPOINT	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
LINE 2	0.241714	0.161599	0.321688	0.20643	0.275854	0.202711	0.205289	0.240579	0.286879	0.129733	0.185086	0.19601	0.291421	0.418887	0.490023	0.635205
LINE 6	0.034452	0.025516	0.112239	0.120248	0.111919	0.088707	0.098052	0.149503	0.126466	0.0179	0.028485	0.001675	0.013864	0.001292	0.012774	0.004559
LINE 7	0.032859	0.046843	0.151592	0.040302	0.05283	0.03202	0.019881	0.037241	0.030358	0.011077	0.01382	0.003771	0.034628	0.051716	0.06089	0.035001
LINE 9	0.141605	0.344924	0.189403	0.78079	0.434114	0.307613	0.105999	0.048712	0.020668	0.020443	0.024641	0.548843	0.664429	0.588152	0.520723	0.656158
LINE 16	0.095686	0.012557	0.234588	0.014108	0.057371	0.100019	0.12206	0.066593	0.25724	0.001958	0.044889	0.000262	0.032329	0.001214	0.015399	0.004721
LINE 18	0.016891	0.017849	0.043337	0.016987	0.019109	0.018575	0.022263	0.028097	0.033252	0.041274	0.042766	0.018189	0.02655	0.010496	0.040023	0.065035

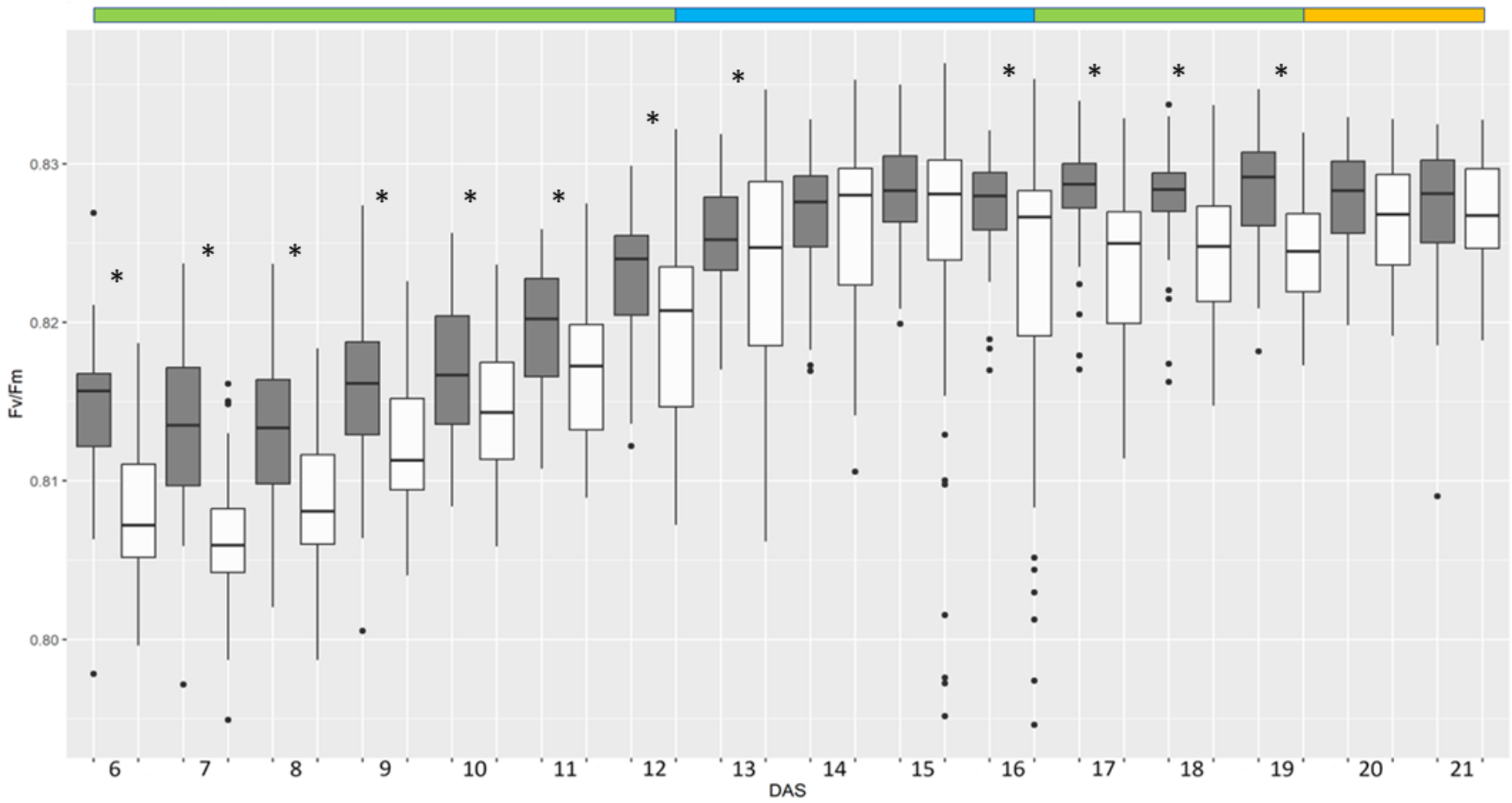


Figure 7 Fv/Fm comparison between Line 9 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 9 is demonstrating higher Fv/Fm than its background, though only under constant light

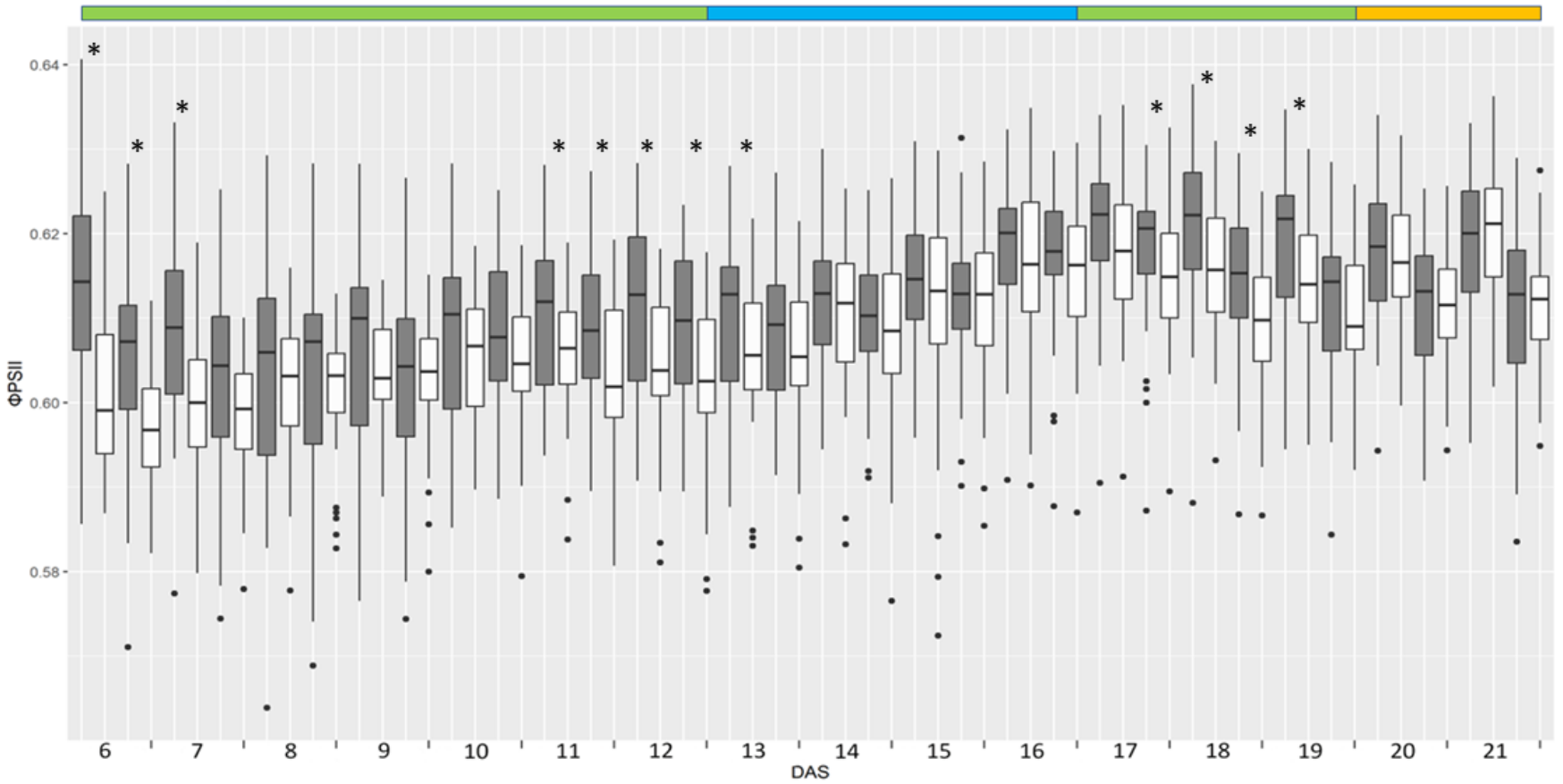


Figure 8 ΦPSII comparison between Line 9 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 9 exhibits higher ΦPSII than its background under constant light

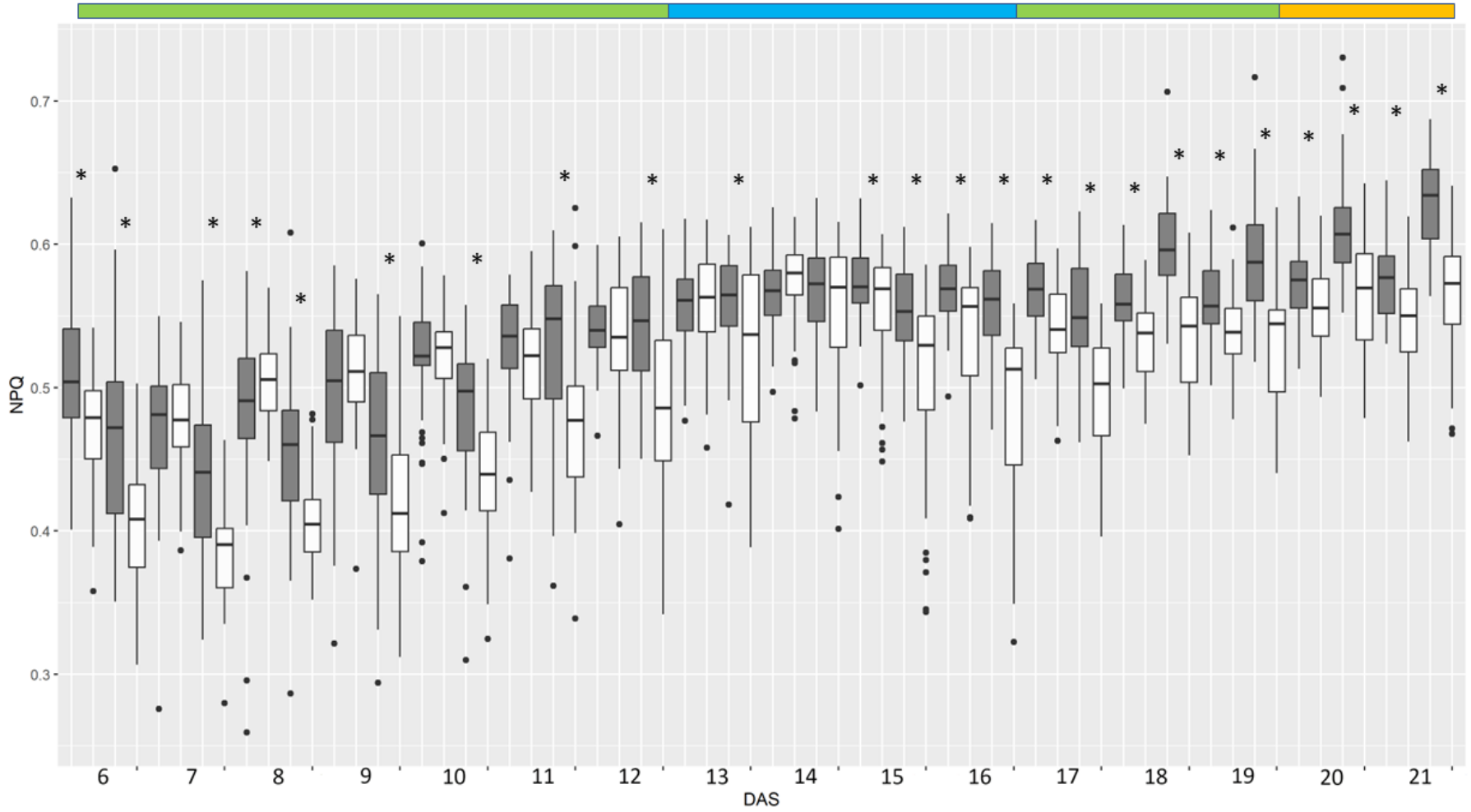


Figure 9 NPQ comparison between Line 9 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. The Non-Photochemical Quenching of Line 9 seems to be higher than the wildtype regardless of light treatment. From 6 to 10 Days After Sowing, there is an important difference between morning and afternoon values of the same day

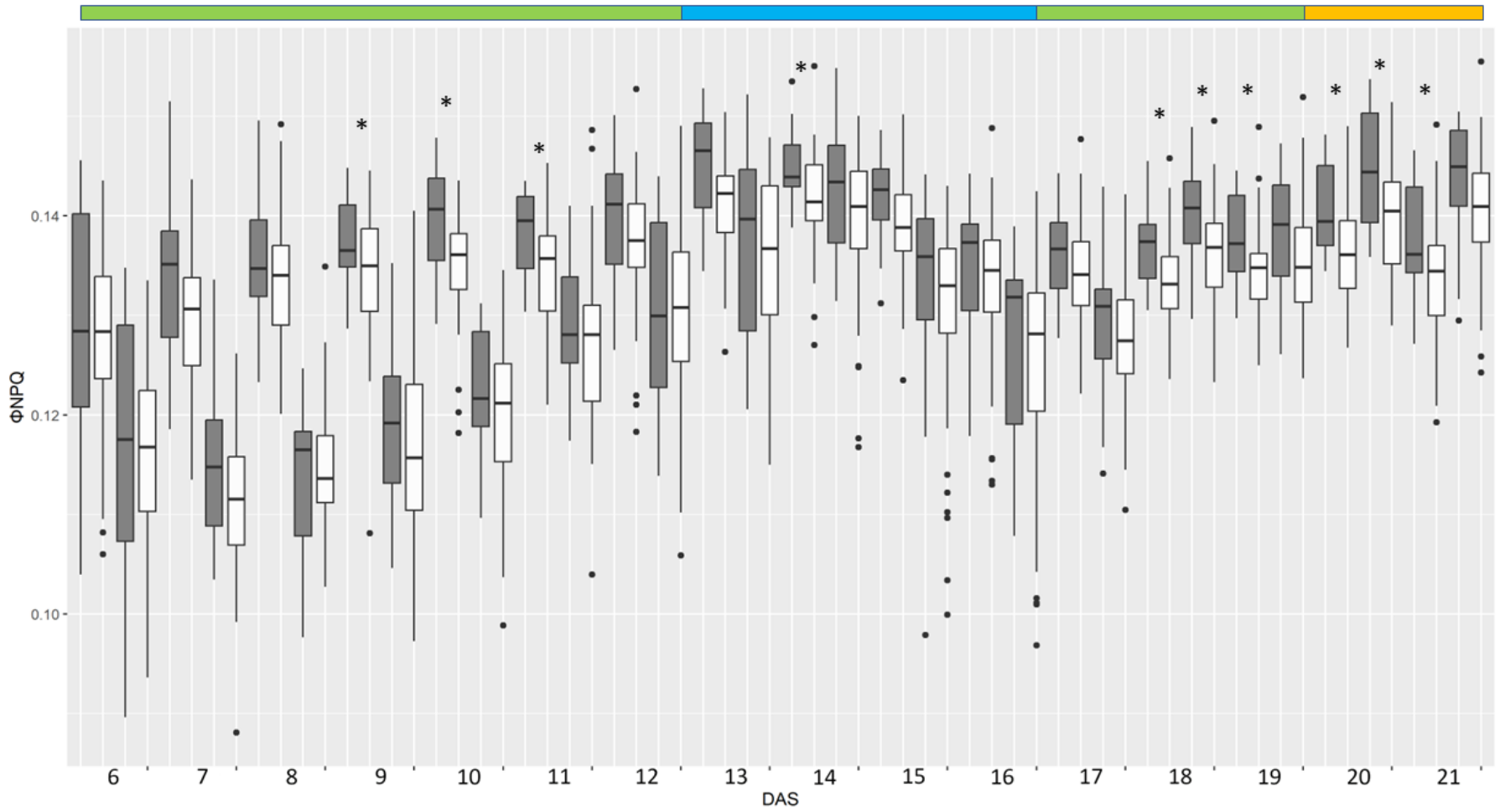


Figure 10 ΦNPQ comparison between Line 9 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 9 exhibits higher ΦNPQ than its background. From 6 to 10 Days After Sowing there seems to be significant differences between morning and afternoon values of the same day

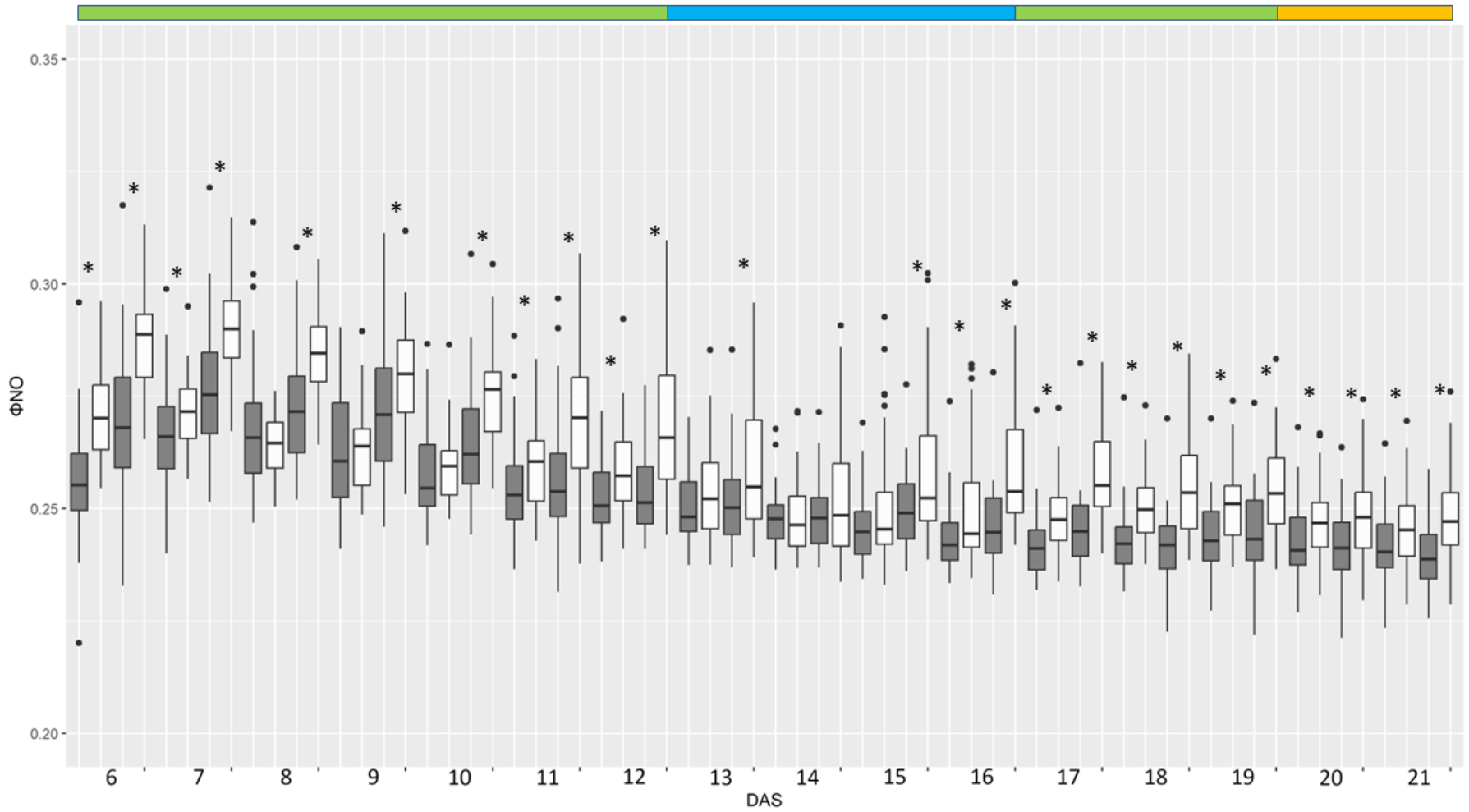


Figure 11 ΦNO comparison between Line 9 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 9 exhibits lower ΦNO values than its background, regardless of light treatment. From 6 to 10 Days After Sowing there seems to be significant differences between morning and afternoon values of the same day

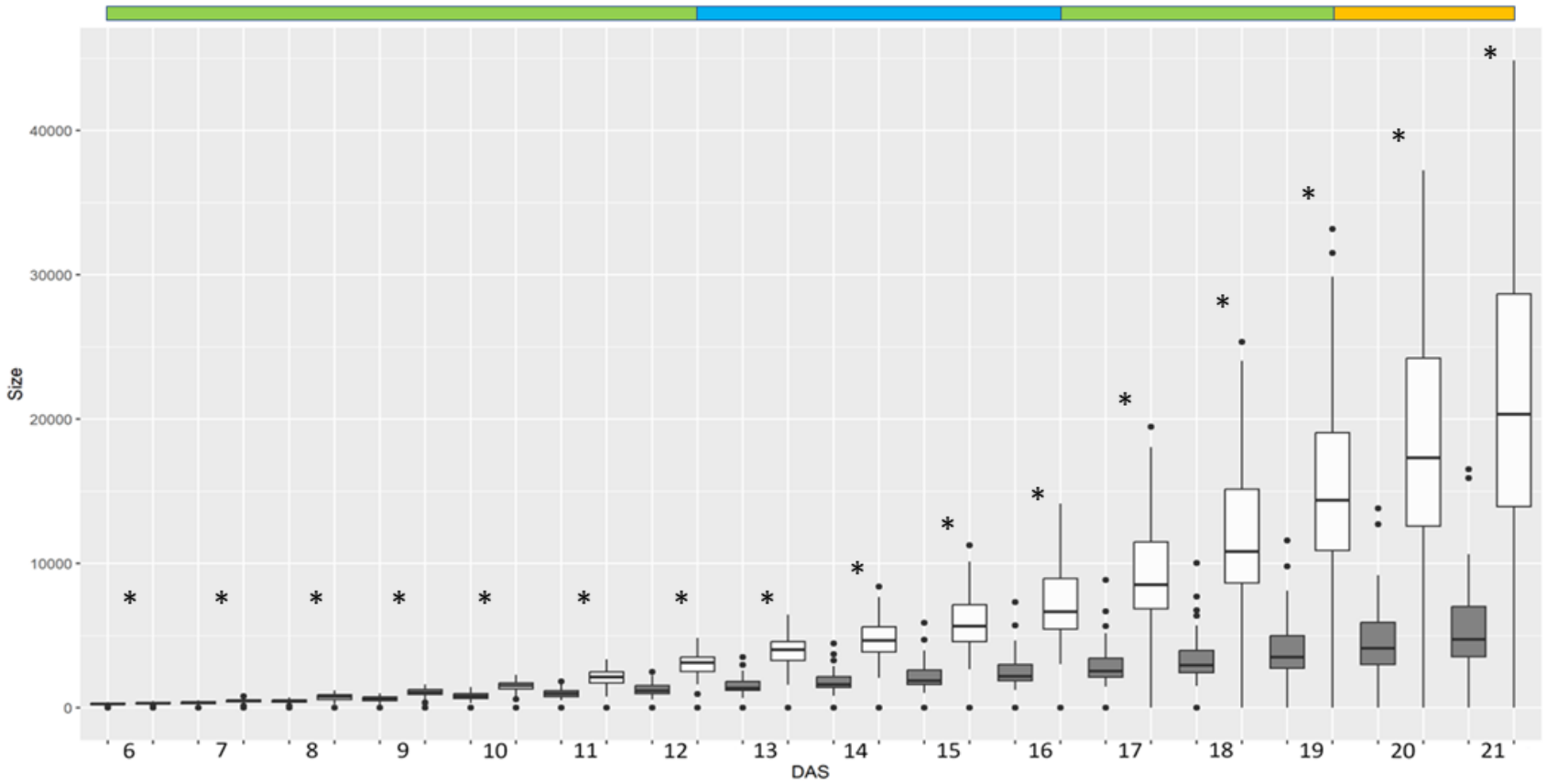


Figure 12 Size comparison between Line 9 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 9 is significantly smaller than its background

Line 2, is a second insert line for AT3G45760, and exhibited significant differences from the wild type for 24 out of the 224 measurements, and for traits ΦPSII and ΦNPQ. Both these parameters were heightened in Line 2, compared to the background. Similarly to Line 9, differences in ΦPSII only appeared under constant light (Figure 13) and ΦNPQ showed large differences between morning and afternoon for the first Days After Sowing (Figure 14). However, the rest of the parameters showed no deviation for this line.

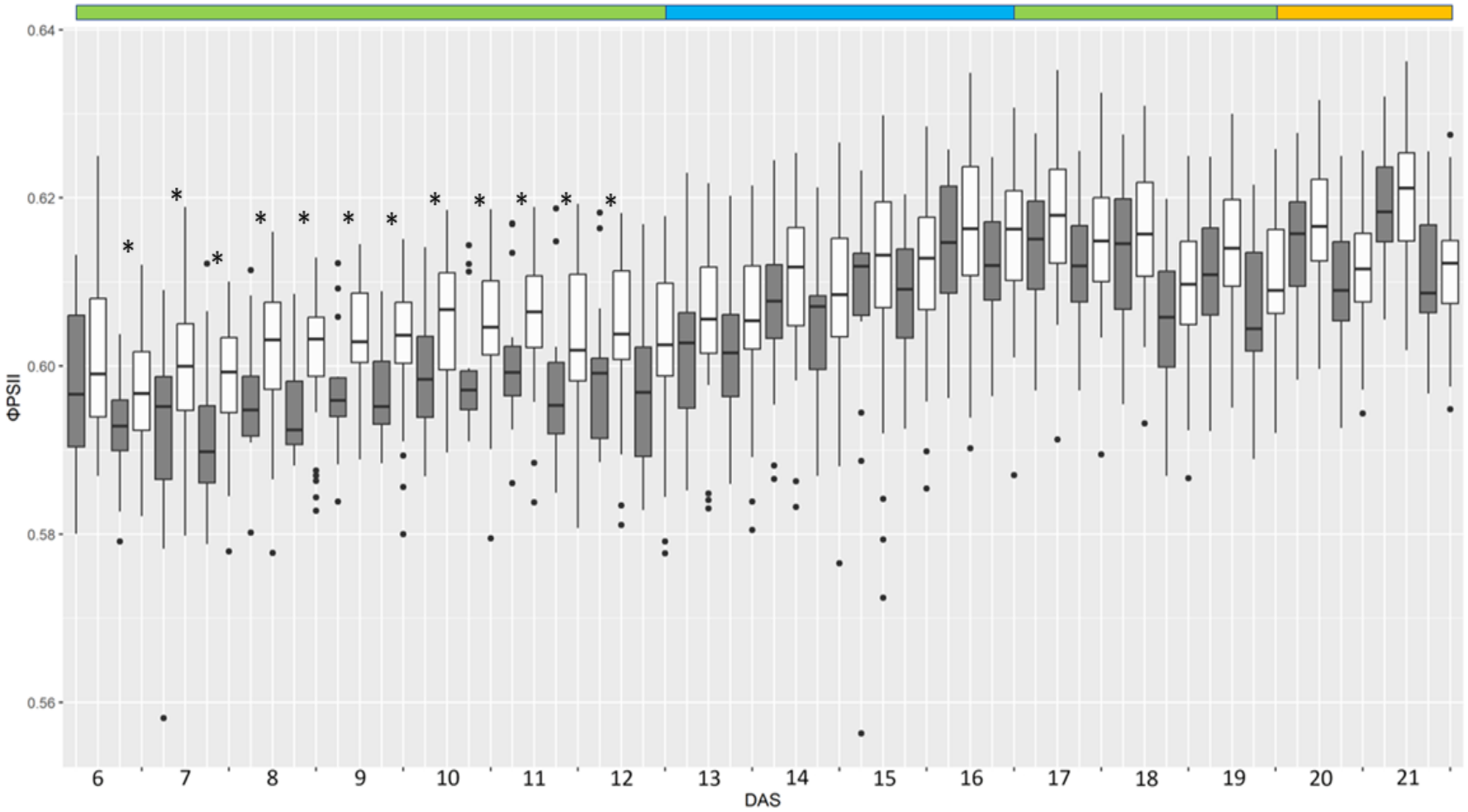


Figure 13 ΦPSII comparison between Line 2 (grey) and the wild type (white), per timepoint. Asterisks signify statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 2 exhibits higher ΦPSII than its background under constant light

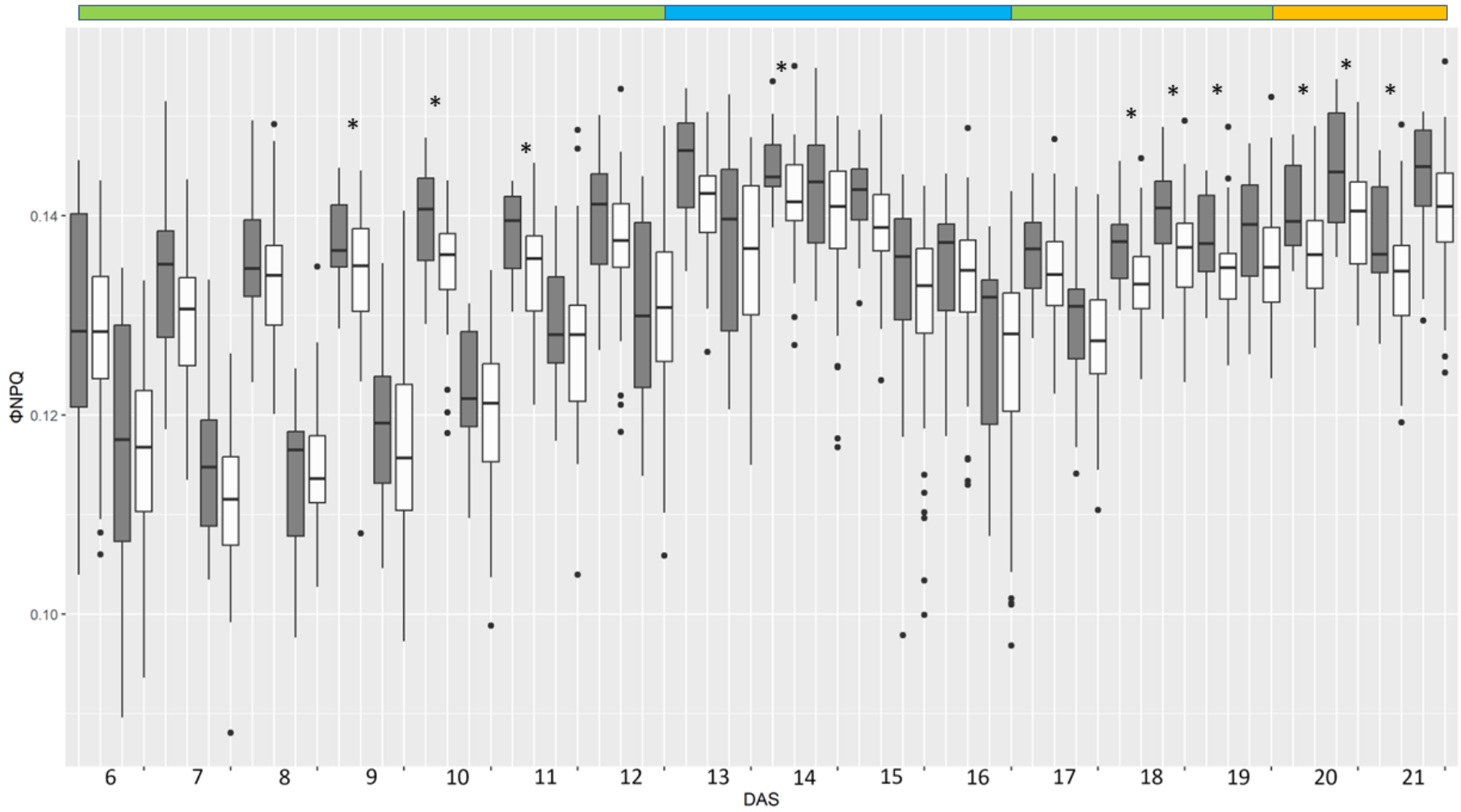


Figure 14 ΦNPQ comparison between Line 2 (grey) and the wild type (white), per timepoint. Asterisks signify statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes. Line 2 exhibits higher ΦNPQ than its background, regardless of light treatment. From 6 to 10 Days After Sowing there seems to be significant differences between morning and afternoon values of the same day

Lines 2 and 9 represent the only case where 2 insert lines for the same gene showed altered phenotypes. For the rest of the genes only 1 line was deviating from its background.

Another line of interest for the QTL in chromosome 3 was 16, an insert line for the AT3G45620 gene, which is predicted to code for a WD40 repeat-like scaffolding protein. Here, 93 out of the 224 timepoints showed differences between the mutants and the wild type plants for all the photosynthetic efficiency parameters.

Fv/Fm was lower in this line, and this decrease did not appear to be correlated with the light treatment applied (Supplementary Figure 24). The same observation was made for ΦPSII (Supplementary Figure 25). On the other hand, the parameters that measure energy dissipation NPQ, ΦNPQ and ΦNO were all lower in this line and showed differences between morning and afternoon values for the first 5 Days After Sowing. Values seem to be unaffected by light treatment (Supplementary Figures 26, 27 & 28).

The last notable line for chromosome 3 was Line 6, which differed from its background for 24 out of the 224 measurements and for traits ΦPSII and ΦNO. It is an insert line for AT3G45770, coding for a polyketide synthase involved in fatty acid biosynthesis. ΦPSII was the main difference, and was elevated compared to the background, regardless of light treatment (Supplementary Figure 29). ΦNO was lower, albeit only during 6 timepoints during the application of light fluctuating every 60 minutes (Supplementary Figure 30). Another insert line for that particular gene was Line 1, which, however, did not any alteration.

T-DNA lines for candidate genes of QTL on Chr 2

For the chromosome 2 QTL, phenotypes of 2 T-DNA lines for candidate gene AT2G18030 were examined. Only one of those, Line 18, showed an alternate phenotype for 35 out of the 224 measurements and mainly for ΦPSII. AT2G18030 codes for a methionine sulfoxide reductase and, therefore, acts as a protector of the plant from oxidative stress. The line had lesser ΦPSII values than the wildtype, regardless of light treatment (Supplementary Figure 31). Line 21, the second insert line for the same gene, did not show any statistically significant differences with its background.

T-DNA lines for candidate genes of QTL on Chr 1

Line 7 was another line showing promising results. It is an insert line for AT1G61860, which codes for a protein kinase involved in phosphorylation. It showed alternate phenotypes for 39 measurements, mainly regarding ΦPSII and ΦNO.

ΦPSII was lower in the insert line, irrespective of the light treatment. Also, the differences appeared relatively late, during the 10th Day After Sowing (Supplementary Figure 32). On the other hand, the mutants exhibited elevated ΦNO, with large differences between morning and afternoon until the 11th Day After Sowing (Supplementary Figure 33).

Discussion

Genotyping results

Regarding the genotyping results, it was observed that 11 out of the 13 lines characterised as "Confirmed Homozygous" actually turned out to be homozygous for the insertion. Lines characterised as "Segregating" were also in line with expectations, whereas the two "Homozygous" lines were, in fact, one homozygous for the mutant allele and one segregating.

In certain instances, the "Segregating" lines showed a complete absence of both homozygous and heterozygous plants. These were lines 3, 5 and 20, where all individuals were wild-type apart from those that did not germinate (Supplementary Figures 3, 5 & 20). This could be due to a negative impact on the homozygous plants' survival caused by major genomic rearrangement (Clark & Krysan, 2010) and suggests that the ungerminated plants in these lines might have been homozygous mutants. Alternatively, these lines are the result of homozygous lethal genes, knocked-out genes essential for the plant's survival (Lloyd & Meinke, 2012). However, in either scenario, one would expect to observe at least some heterozygous plants. Therefore, it is more likely that these lines are false positives that do not contain the T-DNA at the identified locus. This is further reinforced by the clear, strong bands obtained from the LP-RP reaction, indicating that the problem is not with the primers or the quality of the DNA (R. O'Malley et al., 2015).

It is also possible that the lack of homozygous and heterozygous plants observed in lines 3, 5, and 20 may be due to issues with the PCR genotyping protocol. One possibility is that the T-DNA insert was integrated in reverse orientation so that the correct primer combination for its amplification would be the Border Primer (BP) – Left Primer (LP) pair instead of the Border Primer (BP) – Right Primer (RP) pair used. However, according to the SIGnAL T-DNA Express Site (Http://signal.salk.edu/atg1001/3.0/gebrowser.php) only the insert for line 3 is in a reverse orientation.

Alternatively, it is possible that the PCR conditions used for amplifying the T-DNA insert were not optimal. For example, the annealing temperature or extension time may have been suboptimal, resulting in inefficient amplification of the insert.

The "Confirmed Homozygous" lines were largely as expected with two exceptions. One was Line 12 for which It seems that only one of the individuals is homozygous and the rest are heterozygous. However, the reaction for the wild type allele of this line was inconclusive since it only exhibited faint bands (Supplementary Figure 12). There is a chance that these bands are artifact amplicons of an unintended target. After all, based on the clear bands of the reaction for the mutant allele and the consistency of the rest of the "Confirmed Homozygous" lines, this premise could be accepted. However, for the sake of accuracy, only the one individual was considered homozygous. In general, it is hard to pinpoint the exact cause of faint or weak bands, as it could be attributed to numerous factors, such as high annealing temperature, low denaturation temperature, high/low denaturation time, short annealing time, short extension time and small number of cycles (Asif et al., 2021). The second exception was Line 10, which appeared to be segregating (Supplementary Figure 10).

For lines 13 and 18 the reaction for the mutant allele was successful but the reaction for the wild type was not (Supplementary Figures 13 & 18). After several attempts, it was decided for the wild type reactions to be repeated with manually designed primers. The primers were designed and ordered but

time constraints prevented their use. However, the reactions for the insertion worked. Additionally, problems had arisen before with the wildtype template DNA used and on several occasions, reactions had to be repeated because the wildtype sample did not initially amplify. This could be an issue with the quality of the wild type DNA. Presumably, not all individuals had the same DNA quality, as the extraction was performed in several batches. Alternatively, it is suggested that the absence of a single or a couple of bands from the gel could be attributed to incorrect handling of the master mix. One scenario involves the lack of an ingredient from the mix, which is highly unlikely as all ingredients were carefully checked during pipetting, especially in the second or third attempt. Another possibility is the non-homogenous mixing of the materials (Asif et al., 2021). This seems more likely, although the master mixes were both briefly vortexed and centrifuged prior to their use. Hence, the lines were considered homozygous for the insertion to continue with the experiment.

Finally, regarding the "Homozygous Unconfirmed" lines, 22 was predominantly homozygous for the insertion as expected (Supplementary Figure 22), whereas 23 was entirely heterozygous (Supplementary Figure 23).

Phenotyping

Phenotypes exhibited by mutant lines

Lines 9 and 2, T-DNA mutant lines of gene AT3G45760, consistently showed altered phenotypes. While it is generally expected that knockout mutants for candidate genes of a particular trait will exhibit a decrease or loss of function related to that trait, it is worth noting that there are also cases where a knockout can improve or enhance the phenotype (van Es et al., 2019). The latter was the case with Line 9, which showed some unexpected results. Firstly, it exhibited altered phenotypes for all 6 traits studied, including size. Additional differences were observed in the line such as a slower growth rate and partial sterility. This suggests that AT3G45760 has broad effects on cellular or developmental processes and thus, is a pleiotropic gene. Pleiotropy refers to the phenomenon where a single gene affects two or more, seemingly unrelated traits (Paaby & Rockman, 2013). It is often difficult to distinguish pleiotropy from physical gene linkage (Flint & Mackay, 2009). It is also challenging to discover whether pleiotropy is the result of a single locus with many different products, or of a single gene product that can be used in various different ways by the organism (He & Zhang, 2006). In this case, since the gene codes for a nucleotidyltransferase involved in 3' RNA processing, which is quite a generic function, the latter is more probable.

Secondly, line 9 displays augmented photosynthetic efficiency compared to the wildtype. One possible scenario to explain this, is the following: the Nottingham Arabidopsis Stock Centre (NASC) reports two variants of this insert line. In one, the insert lies in an exon, similar to Line 2. In the other variant, however, the insert is located in the 3' UTR of the same gene. Insertions after the stop codon, such as this, are known to be less effective in producing knockouts or knockdowns than insertions in the coding regions or before the start codon (Wang, 2008). However, they could still disrupt or interfere with the transcription of the upstream gene by separating it from downstream enhancer elements, causing a partial phenotype. Furthermore, the insert itself could contain enhancer elements and thus, when placed in the 3'UTR, cause the overexpression of the upstream gene (Wang, 2008). This would justify the increased efficiency displayed by mutants of Line 9. So, while not certain, it is more likely that Line 2 is a knockout line and Line 9 an overexpressing line of AT3G45760.

Alternatively, the random integration of T-DNA could provide another explanation for the differences between supposedly identical insert lines. In several lines, there are more than 1 inserts. The average insertion number reported is 2 inserts per line, based on Oxford Nanopore long read Sequencing and

this is likely to be an underestimate (Pucker et al., 2021). This could mean that a line could have 2 or more inserts. The additional inserts could be disrupting more genes, leading to the apparent different phenotypes between the lines. Even if the extra insert is in an intergenic region, it could affect the expression of the gene upstream/downstream of it by disrupting regulatory elements like promoters and enhancers (Wang, 2008). Indeed, according to the NASC website (Https://arabidopsis.info/) Lines 2 and 9 could contain additional insertions. Furthermore, TDNA integration can cause major genomic rearrangements, such as translocations and duplications in the target genome (Pucker et al., 2021). It can also cause deletions (De Schutter et al., 2007). Hence, it is possible that the insert on the same gene caused different types of such rearrangements between the lines, resulting in the phenotypical differences.

In any case, it seems that the differences between these two insert lines and the wild type, regarding usable energy were larger during the first Days After Sowing and gradually decrease over time. Then, they started deviating again. This pattern appears to correlate with the light treatment. It seems that the measurements start to get closer to each other with the first application of fluctuating light, alternating every 15 minutes. The values start deviating again after the reapplication of constant light. It is not clear if the second application of fluctuating light has the same effect on the phenotype, as it is only applied for the last 2 days. Nevertheless, while this is nothing conclusive, it could be an indicator of a correlation of the light treatment and the disrupted gene. This correlation seemed to only exist for the photochemistry parameters (Fv/Fm and ΦPSII) and not the rest of the metrics (Figures 7 & 8).

When looking at the rest of the lines that showed altered phenotypes, a common pattern emerges. Even though two lines are the knockouts of the same gene, one line exhibits an altered phenotype but the other does not. Specifically lines 18 and 21 (for gene AT2G18030), as well as 6 and 1 (for gene AT3G45770), all show this pattern. This is unexpected and suggests that there are other factors at play that are contributing to the observed phenotypical difference.

One possible factor could be a difference in knock-out efficiency, caused by variation in the location of the inserts within the knocked-out gene. As already mentioned, T-DNA insertions in the genome are random (Krysan et al., 1999). Therefore, two lines could have inserts in the same gene but in different locations, i.e. one could be in an intron and the other in a UTR. As a result, the knockout level of the two lines could be different. This is, in fact, the case for these 4 lines. The insert for Line 18 falls in an intron and for Line 21 in a Coding Sequence exon. Similarly, insert of Line 1 falls in an intron and that of Line 6 in an exon. However, in the former case it is the intron insertion that affects the phenotype and in the latter the exon one. Theoretically, both these types of insertions should produce full knockouts. The insert in the exon should disrupt the coding sequence. The insert in the intron should also knock out the gene, as the T-DNA is around 24.000 bp, hence, too large to be spliced out by Arabidopsis, where the largest introns are found to be 5.000 to 10.000 bp in length (Gielen et al., 1999)(Chang et al., 2017). However, splicing-out of the T-DNA along with the intron is not impossible, and has been reported to happen in a small degree (Verelst et al., 2007). In this case, the wild type transcript is produced, albeit “with decreased efficiency” (Wang, 2008). The splicing of the insert could explain the case of Lines 1 and 6, where the Line with the intron insertion exhibited wild type phenotype regarding photosynthetic efficiency.

Additionally, according to NASC (Https://arabidopsis.info/) Lines 1, 6 and 18 “may have additional insertions potentially segregating”. This could mean that similar to Lines 9 and 2, there could be additional inserts disrupting additional genes causing the difference in phenotype.

It is also plausible that the different phenotypes between supposedly identical lines could be the result of epistatic interactions. Epistasis is the phenomenon where the effect of a gene on a particular

phenotype is dependent on the presence or absence of a different gene. In other words, the effect of a gene is “masked” by another gene in the pathway (Vanderweele, 2010). This could lead to complex interactions between multiple genes, where the phenotype observed in a particular line depends on the genetic background and the presence or absence of other mutations. Thus, for epistasis to be the reason for the observed differences, there needs to be a difference in background. This could justify the differences between lines 18 and 21, where the former is based on a Col-0 background and the latter on a Col-3 background. However, the Col-0 and Col-3 accessions are genetically very similar but have shown polymorphisms (Https://arabidopsis.info/CollectionInfo?id=94, 2023). For the rest of the lines, which are all based on Col-0, random integration or differences in knockout efficiency seem more probable causes.

Finally, there is gene redundancy, where the function of the knocked out gene could be compensated by other genes in the same pathway. Normally, since both lines are knockouts of the same gene one would expect for the redundant gene to be compensated in both lines and thus, the same phenotype to appear. So, as is the case with epistasis, a difference in background is required for this possibility to be considered. Therefore, only lines 18 and 21 could potentially be affected by this.

Relation of disrupted genes to photosynthetic efficiency

A more in-depth look into the function of the genes that showed alternate phenotypes might provide an insight on the relation between each gene and photosynthetic efficiency.

One of the genes for which the mutant shows an altered phenotype is located on the first chromosome (Supplementary figures a & b). This is AT1G61860 which codes for a protein kinase involved in phosphorylation. Not a lot of information exists on the protein. The alternative name listed on The Arabidopsis Information Resource (TAIR) is PBS1-LIKE 41 (PBL41). It can be inferred from that that the protein could share some structural and functional similarities with the Arabidopsis PBS1 protein. PBS1 is a kinase involved in plant immunity. To defend against attacks from pathogens, plants have evolved complex defence systems, one of which is innate immunity. Plant innate immunity is initiated following pathogen recognition by transmembrane cell-surface Pattern Recognition Receptors (PRRs) and intracellular receptor proteins (Chisholm et al., 2006). PRRs recognize the molecular imprint of the pathogen in the form of Pathogen-Associated Molecular Patterns (PAMPs) and activate several of plant’s immune responses known as Pattern-Triggered Immunity (PTI) (Schwessinger & Zipfel, 2008). As a counterattack, several pathogens produce various effectors which cleave molecules involved in the activation of PTI, thus suppressing the immune response. However, certain plant proteins of the Receptor-Like Cytoplasmic Kinase (RCLK) group, one of which is PBS1, serve as alternative targets for the pathogen effectors. Therefore, the role of PBS1 and several PBS-like kinases is to serve as a decoy by mimicking actual effector targets, thus, protecting the plant’s immune responses (Zhang et al., 2010). While there is a connection between the innate immune system and photorespiration, which is closely linked to photosynthesis (Jiang et al., 2023), it is unclear why a knockout mutant of this gene exhibits altered ΦPSII and ΦNPQ values.

Three out of the five genes belong in the QTL in chromosome 3.

The first is AT3G45620 which codes for a Transducin/WD40 repeat-like protein. Transducin/WD40 repeat proteins mediate various protein-protein interactions acting as scaffolding proteins, as well as being involved in the assembly and regulation of multi-subunit protein complexes (Stirnimann et al., 2010). With such a generic role and an overabundance in eucaryotic proteomes, it comes as little surprise that WD40 repeats are integral to a very large number of biological processes, ranging from

cell division to environmental stress responses (Gachomo et al., 2014; Stirnimann et al., 2010). Despite their importance, WD40 domains are quite understudied, at least compared to other common regulators, such as kinases. In 2013, Kundu et al. demonstrated the interaction between receptor for activated C kinase 1 (RACK1), a WD40 scaffolding protein, with various proteins, 18% of which were photosynthesis-related. These included Rubisco small subunit 1A, Plastocyanin major isoform DRT112, Chlorophyll a-b binding protein 1CAB1 and Light-harvesting complex II chlorophyll a/b binding protein 1 (Kundu et al., 2013). This could explain the lowered photosynthetic efficiency (excluding ΦNO) values exhibited by Line 16 (Supplementary figures 24, 25, 26 and 27). It is worth noting that the protein coded by AT3G45620 is predicted and has not been experimentally discovered. However, the WD40 low sequence conservation, as well as their diversity of function makes their identification challenging (Stirnimann et al., 2010).

The second gene in this chromosome is AT3G45760, which codes for NTP5, a protein of the Nucleotidyltransferase family, catalysing the transfer of phosphorous-containing groups to substrates. Polymerases are a very known class of nucleotidyltransferases, transferring nucleotides to nucleic acid chains (Voet et al., 2016). NTP5 seems to have a polymerase function, acting on the formation of the 3' end of mature RNA (The Arabidopsis Information Resource (TAIR), 2022). Maturation of the 3' end of RNA is executed by endonucleolytic cleavage followed by polyadenylation, the addition of multiple adenosines forming the poly(A) tail (Shi, 2012). This modification enhances mRNA stability and facilitates mRNA transfer to the cytoplasm, as well as the initiation of translation. However, the specific role of RNA modifications in gene expression varies and can depend upon various factors like stage of development, cell type and environmental effects (Roundtree et al., 2017). It is, therefore, hard to pinpoint the exact connection between the disrupted gene and photosynthetic efficiency. Regardless, the gene's impact on multiple biological processes is supported by the observation that one of the insert lines exhibited altered phenotypes for numerous traits, not all of which were related to photosynthesis (Figures zxcvbn). The second insert line, Line 2, showed a lowered ΦPSII and ΦNPQ values (Figures a b)

The final gene in chromosome 3 is AT3G45770, which codes for an enoyl-ACP reductase, an enzyme involved in fatty acid biosynthesis. In plants, de novo fatty acid biosynthesis occurs in the plastids and mitochondria (Ohlrogge & Jaworski, 1997). There, two different biosynthetic systems use an acyl carrier protein (ACP)-dependent Fatty Acid Synthase (FAS) to catalyse the synthesis of the acids (Guan et al., 2017). Enoyl-ACP reductase is one of the core components of both the plastidial and mitochondrial FAS systems (Guan et al., 2020). Fatty acids themselves are an important component of the photosynthetic machinery. Photosystems I and II, as well as ATP synthase, all key parts of photosynthesis, are embedded in the thylakoid membranes of the chloroplasts. The lipid and fatty acid composition of the chloroplasts is unique, denoting a relationship between the thylakoid membrane structure and photosynthesis in photosynthetic organisms (Hernández & Cejudo, 2021). Indeed, plenty of research suggests that the degree of unsaturation of fatty acids in the chloroplast membrane can affect the fluidity and stability of the membranes and, thus, electron transport and the efficiency of light harvesting (Falcone et al., 2004). These changes in the level of unsaturation are often the response to environmental temperature stresses (Domínguez et al., 2010)(Routaboul et al., 2012). Furthermore, fatty acid biosynthesis and photosynthesis are reciprocally related, in that the energy created by photosynthesis is used in the synthesis of fatty acids and, oxidation of fatty acids can generate ATP that can be used in driving photosynthesis (Hernández & Cejudo, 2021). This complex interplay between fatty acids and the photosynthetic machinery could explain the lowered ΦPSII and heightened ΦNO phenotypes exhibited by line 6 (Supplementary figures a b). On the contrary, insert line 1, a knockout line for the same gene does not exhibit any observable difference from the background.

The alternate phenotypes observed in mutants of both AT3G45760 and AT3G45770 confirm the findings of former student Jacky To , who identified these two genes as the most likely to be causal genes underlying the chromosome 3 QTL. Specifically for AT3G45770, my predecessor discovered a correlation between gene expression and light treatment where, the gene's expression seems to increase after exposure to fluctuating light. In this project, and assuming that Line 6 is a knockout line, this would make the differences between mutant and wild type greater after exposure to light fluctuating every 15 minutes. This, however, does not seem to be entirely the case. While ΦPSII is indeed affected by both applications of fluctuating light, it also exhibits differences between wild type and mutant during constant light as well (Supplementary Figure 29). As for ΦNO, the differences there seem to appear only under light fluctuating every 60 minutes (Supplementary Figure 30). In any case, my predecessor determined AT3G45760 as more likely to be causal than AT3G45770, due to it containing two high LD SNPs causing severe consequence variants. This correlates with this study's findings, since AT3G45760 was the only gene for which, both insert lines had an alternate phenotype.

Finally, one gene belongs in the QTL in chromosome 2. This is AT2G18030, which codes for Methionine Sulfoxide Reductase A5 (MSRA5). Methionine Sulfoxide Reductases (MSRs) is a family of proteins the main role of which is the complete and direct protection of proteins from oxidative stress (Rey & Tarrago, 2018). More specifically, the oxidation of methionine leads to the creation of methionine sulfoxide (MetO) and methionine sulfone ($MetO_2$). MetO can be used as a readout for oxidative stress, as it can reach concentration levels in the class of low millimolar during severe oxidative stresses (Tarrago et al., 2015). The S-MetO and R-MetO diastereoisomers are formed, and are reduced by MSR types A and B respectively (Grimaud et al., 2001). In Arabidopsis the protein family consists of 14 members found in plastids, cytosol and the endoplasmic reticulum, where MSRA5 has been predicted to be localized (Rey & Tarrago, 2018). Most MSRs are expressed in all plant organs but are predominantly found in aerial green tissues where photosynthesis takes place. One could assume that MSRA5 will follow that pattern, although there is only evidence of MSRA5 expression in seeds in maize (Zhu et al., 2015)(Simonović & Anderson, 2007). In any case, both MSRAs and MSRBs seem to interact with exposure to light. It has been shown that the photosynthetic activity level, which controls the balance of plastidial redox, plays an important role in molecular pathways regulating the expression of the MSR genes. In A. thaliana, exposure to high light, as well as a combination of high light and low temperature results in the overexpression of at least one MSR gene (Romero et al., 2004)(Dos Santos et al., 2005). What's more, several photosynthesis-related proteins are potential MSR substrates, including RuBisCO and several ATPase subunits (Tarrago et al., 2012). All this can possibly explain the lowered ΦPSII phenotype exhibited by Line 18 (Supplementary Figure 18). The gene is (presumably) knocked out, thus, its overexpression cannot be induced by the oxidative stress brough about by excess light energy. Therefore, the protein is not produced or a faulty version of it is produced. In both cases, the photosynthesis-related substrates are not reduced by the MSRA5 protein, leading to reduced photosynthetic efficiency. Line 21 which was an insert line for the same gene, however, did not exhibit any altered phenotype. This line has a slightly different background (Col-3). This could mean that the Col-0 individuals are causing the apparent phenotypic differences, which is quite unlikely as the differences between Col-0 and Col-3 are minimal. It could also mean that Line 21 is a false positive and that individuals are not indeed homozygous mutants.

AT2G18030 was one of the candidate genes on the QTL in chromosome 2 identified by my predecessor Laura Bos Calderó. According to the GWAS performed, this particular QTL is significant during the first days of the experiment under constant light and light fluctuating every 15 minutes. After that, it loses significance and does not regain it with the application of light fluctuating every 15 minutes. In this project, however, while mutants of the gene indeed show significant deviation from the wild type from 6 to 15 Days After Sowing, there seems to be a restoration of significance with the reapplication of

fluctuating light (Supplementary figure 31). What's more, 3 severe consequence variants were predicted that could affect the gene and cause a difference in phenotypical differences within the population. One fell in the 3' UTR, one in the 5' UTR and one in an intron. All three variants could cause alterations in gene expression. However, gene expression analysis performed on the gene showed no differences between contrasting haplotypes, hence, my predecessor deemed this gene not likely to be causal.

Future research

To address the knowledge gaps and inconsistencies uncovered in this study, further research needs to be conducted. The gene expression analysis, performed on homozygous mutant plants, which would include:

a) The extraction of RNA from young plant tissue
b) the use of this RNA to synthesize total cDNA
c) the PCR amplification of that cDNA with specific primers for each studied gene, spanning the entirety of the coding sequence or a large portion of it. This would clarify if the insert lies indeed in an intron/exon or in a UTR.
d) for the inserts that do indeed fall in a UTR, a Reverse Transcription-quantitative PCR (RT-qPCR) will show the gene expression levels in the mutants and by comparison to a reference gene it can be concluded what type of disruption the insert causes: overexpression, knockdown or knockout.

Finally, genes that show altered phenotypes need to be cloned into vectors, reintroduced in the mutants and if the restoration of the wildtype takes place then can it be concluded that the gene is indeed causal. Whether this causal gene will indeed is suitable for breeding is another matter entirely.

Conclusion

The aim of this project was to screen a list of *A. thaliana* candidate genes that possibly affect photosynthetic efficiency under fluctuating light via the use of T-DNA knockout lines and determine which of these lines exhibit an altered phenotype.

In total, 19 candidate genes were tested contained in 3 QTLs in chromosomes 1, 2 and 3. The QTL of chromosome 2 was identified as candidate by two Genome-Wide Association Studies performed on the DartMap population by my predecessor Laura Bos Calderó for the traits ΦPSII and ΦNO. The QTL in chromosome 3 was identified by 4 different Genome-Wide Association Studies. Two of them performed on the DartMap population by Nguyen et al. for the traits ΦPSII and ΦNO under fluctuating light. One was performed by Boesten et al. for ΦPSII under stable light. Finally, one was performed on the HapMap population by, Prinzenberg et al. (2019)

Further evidence on the causality of gene AT3G45760 contained in the QTL on chromosome 3 was discovered by my predecessor Jacky To. This correlates with the findings of this study, where two different insert lines for AT3G45760 exhibited altered phenotypes compared to the wildtype. One line in particular showed altered phenotypes across the board for all traits studied. Other knockout lines showing alternate phenotypes were those for the genes AT1G61860, AT2G18030, AT3G45620, AT3G45770.

Additional research will need to be carried out, especially for AT3G45760. Gene expression analysis will be essential for determining if the altered phenotypes exhibited by the two lines are the result of a knockout, a knockdown or an overexpression of the gene. The cloning of the gene into a vector, the subsequent transformation of the mutant lines and the restoration of the wild type will conclusively prove its candidacy.

In conclusion, while this study represents a small step towards understanding the genetic basis of photosynthetic efficiency, its findings provide further support for previous research and highlight the need for continued efforts to identify and characterize candidate genes for photosynthetic efficiency. The altered phenotypes observed in the T-DNA knockout lines suggest that these genes are involved in the regulation of various aspects of the photosynthetic process. The identification of these candidate genes will provide a basis for future studies aimed at improving crop yield under fluctuating light conditions, which is of critical importance in the face of climate change and increasing food demands.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. Phuong Nguyen, for her guidance and support throughout this project. I am also deeply thankful to my examiner, René Boesten, for his feedback and for his help during the initial stages of the experiment. The Laboratory of Genetics staff members, including PhD candidates and technicians, have been instrumental in providing valuable advice and assistance. Finally, I would like to thank the students of the Laboratory of Genetics, for their support and for making this whole experience a bit more enjoyable.

References

Arabidopsis Biological Resource Center. (n.d.). *Handling Arabidopsis Plants and Seeds*.

Asif, S., Khan, M., Waqar Arshad, M., & Imran Shabbir, M. (2021). PCR Optimization for Beginners: A Step by Step Guide. *Research in Molecular Medicine*, *9*(2), 81–102. https://doi.org/10.32598/rmm.9.2.1189.1

Baillie, J., & Zhang, Y. P. (2018). Space for nature. *Science*, *361*(6407), 1051. https://doi.org/10.1126/science.aau1397

Blankenship, R. E. (2021). Molecular Mechanisms of Photosynthesis. In *Molecular Mechanisms of Photosynthesis* (3d ed.). https://doi.org/10.1002/9780470758472

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biology*, *12*(10). https://doi.org/10.1186/gb-2011-12-10-232

Britannica. (2022). Photosynthesis. In *Encyclopedia Britannica*. https://www.britannica.com/question/Why-is-photosynthesis-important

Caparrós-Martín, J. A., McCarthy-Suárez, I., & Culiáñez-Macià, F. A. (2013). HAD hydrolase function unveiled by substrate screening: Enzymatic characterization of Arabidopsis thaliana subclass I phosphosugar phosphatase AtSgpp. *Planta*, *237*(4), 943–954. https://doi.org/10.1007/s00425-012-1809-5

Chang, N., Sun, Q., Hu, J., An, C., & Gao, H. (2017). Large introns of 5 to 10 kilo base pairs can be spliced out in Arabidopsis. *Genes*, *8*(8). https://doi.org/10.3390/genes8080200

Cheng, H. C., Qi, R. Z., Paudel, H., & Zhu, H. J. (2011). Regulation and function of protein kinases and phosphatases. *Enzyme Research*, *2011*(1), 7–9. https://doi.org/10.4061/2011/794089

Chisholm, S. T., Coaker, G., Day, B., & Staskawicz, B. J. (2006). Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell*, *124*(4), 803–814. https://doi.org/10.1016/j.cell.2006.02.008

Clark, K. A., & Krysan, P. J. (2010). Chromosomal translocations are a common phenomenon in Arabidopsis thaliana T-DNA insertion lines. In *Plant Journal* (Vol. 64, Issue 6, pp. 990–1001). https://doi.org/10.1111/j.1365-313X.2010.04386.x

Clough, S. J., & Bent, A. F. (1999). *Floral dip : a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana*. *16*(October 1998), 735–743.

Cruz, J. A., Savage, L. J., Zegarac, R., Hall, C. C., Satoh-Cruz, M., Davis, G. A., Kovac, W. K., Chen, J., & Kramer, D. M. (2016). Dynamic Environmental Photosynthetic Imaging Reveals Emergent Phenotypes. *Cell Systems*, *2*(6), 365–377. https://doi.org/10.1016/j.cels.2016.06.001

De Schutter, K., Joubès, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., Van Der Schueren, E., Beeckman, T., Kushnir, S. T., Inzé, D., & De Veylder, L. (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. *Plant Cell*, *19*(1), 211–225. https://doi.org/10.1105/tpc.106.045047

Domínguez, T., Hernández, M. L., Pennycooke, J. C., Jiménez, P., Martínez-Rivas, J. M., Sanz, C., Stockinger, E. J., Sánchez-Serrano, J. J., & Sanmartín, M. (2010). Increasing ω-3 desaturase

expression in tomato results in altered aroma profile and enhanced resistance to cold stress. *Plant Physiology*, *153*(2), 655–665. https://doi.org/10.1104/pp.110.154815

Dos Santos, C. V., Cuiné, S., Rouhier, N., & Rey, P. (2005). The Arabidopsis plastidic methionine sulfoxide reductase B proteins. Sequence and activity characteristics, comparison of the expression with plastidic methionine sulfoxide reductase A, and induction by photooxidative stress. *Plant Physiology*, *138*(2), 909–922. https://doi.org/10.1104/pp.105.062430

Evans, J. R. (2013). Improving photosynthesis. *Plant Physiology*, *162*(4), 1780–1793. https://doi.org/10.1104/pp.113.219006

Falcone, D. L., Ogas, J. P., & Somerville, C. R. (2004). Regulation of membrane fatty acid composition by temperature in mutants of Arabidopsis with alterations in membrane lipid composition. *BMC Plant Biology*, *4*, 1–45. https://doi.org/10.1186/1471-2229-4-17

Flint, J., & Mackay, T. F. C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research*, *19*(5), 723–733. https://doi.org/10.1101/gr.086660.108

Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I., & Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology*, *19*(3), 192–206. https://doi.org/10.1038/nrm.2017.94

Gachomo, E. W., Jimenez-Lopez, J. C., Baptiste, L. J., & Kotchoni, S. O. (2014). GIGANTUS1 (GTS1), a member of Transducin/WD40 protein superfamily, controls seed germination, growth and biomass accumulation through ribosome-biogenesis protein interactions in Arabidopsis thaliana. *BMC Plant Biology*, *14*(1), 1–17. https://doi.org/10.1186/1471-2229-14-37

Gielen, J., Terryn, N., Villarroel, R., & Van Montagu, M. (1999). Complete nucleotide sequence of the T-DNA region of the plant tumour-inducing Agrobacterium tumefaciens Ti plasmid pTiC58. *Journal of Experimental Botany*, *50*(337), 1421–1422. https://doi.org/10.1093/jxb/50.337.1421

Grimaud, R., Ezraty, B., Mitchell, J. K., Lafitte, D., Briand, C., Derrick, P. J., & Barras, F. (2001). Repair of oxidized proteins: Identification of a new methionine sulfoxide reductase. *Journal of Biological Chemistry*, *276*(52), 48915–48920. https://doi.org/10.1074/jbc.M105509200

Guan, X., Okazaki, Y., Lithio, A., Li, L., Zhao, X., Jin, H., Nettleton, D., Saito, K., & Nikolau, B. J. (2017). Discovery and characterization of the 3-hydroxyacyl-ACP dehydratase component of the plant mitochondrial fatty acid synthase system. *Plant Physiology*, *173*(4), 2010–2028. https://doi.org/10.1104/pp.16.01732

Guan, X., Okazaki, Y., Zhang, R., Saito, K., & Nikolaua, B. J. (2020). Dual-localized enzymatic components constitute the fatty acid synthase systems in mitochondria and plastids. *Plant Physiology*, *183*(2), 517–529. https://doi.org/10.1104/pp.19.01564

He, X., & Zhang, J. (2006). Toward a molecular understanding of pleiotropy. *Genetics*, *173*(4), 1885–1891. https://doi.org/10.1534/genetics.106.060269

Hernández, M. L., & Cejudo, F. J. (2021). Chloroplast Lipids Metabolism and Function. A Redox Perspective. *Frontiers in Plant Science*, *12*(August), 2015–2022. https://doi.org/10.3389/fpls.2021.712022

Http://signal.salk.edu/atg1001/3.0/gebrowser.php. (n.d.). *SALK Genome browser*.

Http://signal.salk.edu/tdnaprimers.2.html. (n.d.-a). *http://signal.salk.edu/tdnaprimers.2.html*. http://signal.salk.edu/tdnaprimers.2.html

Http://signal.salk.edu/tdnaprimers.2.html. (n.d.-b). *SALK TDNA Primer design*.

Https://arabidopsis.info/. (n.d.). *https://arabidopsis.info/*.

Https://arabidopsis.info/CollectionInfo?id=94. (2023). *https://arabidopsis.info/CollectionInfo?id=94*. https://arabidopsis.info/CollectionInfo?id=94

Https://www.idtdna.com/calc/analyzer. (n.d.). *IDT OligoAnalyzer*.

Jiang, X., Walker, B. J., He, S. Y., & Hu, J. (2023). *The role of photorespiration in plant immunity*. *February*, 1–8. https://doi.org/10.3389/fpls.2023.1125945

Kramer, D. M., Johnson, G., Kiirats, O., & Edwards, G. E. (2004). New uorescence parameters for the determination of Q. *Biological Chemistry*, 209–218.

Krysan, P. J., Young, J. C., & Sussman, M. R. (1999). *T-DNA as an Insertional Mutagen in Arabidopsis*. *11*(December), 2283–2290.

Kundu, N., Dozier, U., Deslandes, L., Somssich, I. E., & Ullah, H. (2013). Arabidopsis scaffold protein RACK1A interacts with diverse environmental stress and photosynthesis related proteins. *Plant Signaling and Behavior*, *8*(5). https://doi.org/10.4161/psb.24012

Landrieu, I., De Veylder, L., Fruchart, J. S., Odaert, B., Casteels, P., Portetelle, D., Van Montagu, M., Inzé, D., & Lippens, G. (2000). The Arabidopsis thaliana PIN1At gene encodes a single-domain phosphorylation-dependent peptidyl prolyl cis/trans isomerase. *Journal of Biological Chemistry*, *275*(14), 10577–10581. https://doi.org/10.1074/jbc.275.14.10577

Lloyd, J., & Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiology*, *158*(3), 1115–1129. https://doi.org/10.1104/pp.111.192393

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. In *Nature Reviews Molecular Cell Biology* (Vol. 8, Issue 6, pp. 479–490). https://doi.org/10.1038/nrm2178

Müller, P., Li, X. P., & Niyogi, K. K. (2001). Non-photochemical quenching. A response to excess light energy. *Plant Physiology*, *125*(4), 1558–1566. https://doi.org/10.1104/pp.125.4.1558

Mulligan, B., & Russell, J. (n.d.). *Harvesting and Storage of Arabidopsis Seeds*. *i*, 2–3.

O'Malley, R., Barragan, C., & Ecker, J. (2015). *A User ' s Guide to the Arabidopsis T-DNA Insertional Mutant Collections*. 323–342.

O'Malley, R. C., & Ecker, J. R. (2010). Linking genotype to phenotype using the Arabidopsis unimutant collection. *Plant Journal*, *61*(6), 928–940. https://doi.org/10.1111/j.1365-313X.2010.04119.x

Ohlrogge, J. B., & Jaworski, J. G. (1997). Regulation of fatty acid synthesis. *Annual Review of Plant Biology*, *48*, 109–136. https://doi.org/10.1146/annurev.arplant.48.1.109

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in Genetics*, *29*(2), 66–73. https://doi.org/10.1016/j.tig.2012.10.010

Pucker, B., Kleinbölting, N., & Weisshaar, B. (2021). Large scale genomic rearrangements in selected Arabidopsis thaliana T-DNA lines are caused by T-DNA insertion mutagenesis. *BMC Genomics*, *22*(1), 1–21. https://doi.org/10.1186/s12864-021-07877-8

Rey, P., & Tarrago, L. (2018). Physiological roles of plant methionine sulfoxide reductases in redox homeostasis and signaling. *Antioxidants*, *7*(9). https://doi.org/10.3390/antiox7090114

Romero, H. M., Berlett, B. S., Jensen, P. J., Pell, E. J., & Tien, M. (2004). Investigations into the role of the plastidial peptide methionine sulfoxide reductase in response to oxidative stress in

Arabidopsis. *Plant Physiology*, *136*(3), 3784–3794. https://doi.org/10.1104/pp.104.046656

Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell*, *169*(7), 1187–1200. https://doi.org/10.1016/j.cell.2017.05.045

Routaboul, J. M., Skidmore, C., Wallis, J. G., & Browse, J. (2012). Arabidopsis mutants reveal that short-and long-term thermotolerance have different requirements for trienoic fatty acids. *Journal of Experimental Botany*, *63*(3), 1435–1443. https://doi.org/10.1093/jxb/err381

Schwessinger, B., & Zipfel, C. (2008). News from the frontline: recent insights into PAMP-triggered immunity in plants. *Current Opinion in Plant Biology*, *11*(4), 389–395. https://doi.org/10.1016/j.pbi.2008.06.001

Segonzac, C., Boyer, J. C., Ipotesi, E., Szponarski, W., Tillard, P., Touraine, B., Sommerer, N., Rossignol, M., & Gibrat, R. (2007). Nitrate efflux at the root plasma membrane: Identification of an Arabidopsis excretion transporter. *Plant Cell*, *19*(11), 3760–3777. https://doi.org/10.1105/tpc.106.048173

Shi, Y. (2012). Alternative polyadenylation: New insights from global analyses. *Rna*, *18*(12), 2105–2117. https://doi.org/10.1261/rna.035899.112

SIGnAL : Salk Institute Genomic Analysis Laboratory. (n.d.). *T-DNA Primer Design*. http://signal.salk.edu/tdnaprimers.2.html

Simonović, A. D., & Anderson, M. D. (2007). Analysis of methionine oxides and nitrogen-transporting amino acids in chilled and acclimated maize seedlings. *Amino Acids*, *33*(4), 607–613. https://doi.org/10.1007/s00726-007-0503-6

Sprunck, S., Hackenberg, T., Englhart, M., & Vogler, F. (2014). Same same but different: Sperm-activating EC1 and ECA1 gametogenesis- related family proteins. *Biochemical Society Transactions*, *42*(2), 401–407. https://doi.org/10.1042/BST20140039

Stirbet, A., Lazár, D., Guo, Y., & Govindjee, G. (2020). Photosynthesis: Basics, history and modelling. *Annals of Botany*, *126*(4), 511–537. https://doi.org/10.1093/aob/mcz171

Stirnimann, C. U., Petsalaki, E., Russell, R. B., & Müller, C. W. (2010). WD40 proteins propel cellular networks. *Trends in Biochemical Sciences*, *35*(10), 565–574. https://doi.org/10.1016/j.tibs.2010.04.003

Stortenbeker, N., & Bemer, M. (2019). The SAUR gene family: The plant's toolbox for adaptation of growth and development. *Journal of Experimental Botany*, *70*(1), 17–27. https://doi.org/10.1093/jxb/ery332

Tarrago, L., Kieffer-Jaquinod, S., Lamant, T., Marcellin, M., Garin, J., Rouhier, N., & Rey, P. (2012). Affinity chromatography: A valuable strategy to isolate substrates of methionine sulfoxide reductases? *Antioxidants and Redox Signaling*, *16*(1), 79–84. https://doi.org/10.1089/ars.2011.4153

Tarrago, L., Péterfi, Z., Lee, B. C., Michel, T., & Gladyshev, V. N. (2015). Monitoring methionine sulfoxide with stereospecific mechanism-based fluorescent sensors. *Nature Chemical Biology*, *11*(5), 332–338. https://doi.org/10.1038/nchembio.1787

The Arabidopsis Information Resource (TAIR). (2022). *The Arabidopsis Information Resource (TAIR)*. https://www.arabidopsis.org

Tzfira, T., & Li, J. (2004). *Agrobacterium T-DNA integration : molecules and models*. *20*(8). https://doi.org/10.1016/j.tig.2004.06.004

UniProt Consortium. (2023). *UniProt : the Universal Protein Knowledgebase in 2023*. 1–9.

van Bezouw, R. F. H. M., Keurentjes, J. J. B., Harbinson, J., & Aarts, M. G. M. (2019). Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *Plant Journal*, *97*(1), 112–133. https://doi.org/10.1111/tpj.14190

van Es, S. W., van der Auweraert, E. B., Silveira, S. R., Angenent, G. C., van Dijk, A. D. J., & Immink, R. G. H. (2019). Comprehensive phenotyping reveals interactions and functions of Arabidopsis thaliana TCP genes in yield determination. *Plant Journal*, *99*(2), 316–328. https://doi.org/10.1111/tpj.14326

Vanderweele, T. J. (2010). Epistatic interactions. *Statistical Applications in Genetics and Molecular Biology*, *9*(1). https://doi.org/10.2202/1544-6115.1517

Verelst, W., Saedler, H., & Münster, T. (2007). MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific arabidopsis promoters. *Plant Physiology*, *143*(1), 447–460. https://doi.org/10.1104/pp.106.089805

Voet, D., Voet, J. G., & Pratt, C. W. (2016). *Fundamentals of Biochemistry: Life at the Molecular Level, 5th Edition* (5th ed.). Wiley.

Wang, Y. H. (2008). How effective is T-DNA insertional mutagenesis in Arabidopsis? *J Biochem Tech*, *1*(1), 11–20.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Yuan, X., Wang, H., Cai, J., Li, D., & Song, F. (2019). NAC transcription factors in plant immunity. *Phytopathology Research*, *1*(1), 1–13. https://doi.org/10.1186/s42483-018-0008-0

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., Zou, Y., Gao, M., Zhang, X., Chen, S., Mengiste, T., Zhang, Y., & Zhou, J. M. (2010). Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a Pseudomonas syringae effector. *Cell Host and Microbe*, *7*(4), 290–301. https://doi.org/10.1016/j.chom.2010.03.007

Zhu, J., Ding, P., Li, Q., Gao, Y. K., Chen, F., & Xia, G. (2015). Molecular characterization and expression profile of methionine sulfoxide reductase gene family in maize (Zea mays) under abiotic stresses. *Gene*, *562*(2), 159–168. https://doi.org/10.1016/j.gene.2015.02.066

Appendix

Supplementary Table 1 Different PCR programs used during genotyping

PROGRAM 1			
Step	**Temp**	**Time**	**# of cycles**
Initial Denaturation	95°C	3 min	
Denaturation	95°C	30 sec	32
Primer Annealing	55°C	15 sec	
Extension	72°C	1:10 min	
Final Extension	72°C	10 min	
PROGRAM 2			
Step	**Temp**	**Time**	**# of cycles**
Initial Denaturation	95°C	3 min	
Denaturation	95°C	30 sec	32
Primer Annealing	55°C	30 sec	
Extension	72°C	1:10 min	
Final Extension	72°C	10 min	
PROGRAM 3			
Step	**Temp**	**Time**	**# of cycles**
Initial Denaturation	95°C	3 min	
Denaturation	95°C	30 sec	32
Primer Annealing	55°C	30 sec	
Extension	72°C	1:30 min	
Final Extension	72°C	10 min	

Supplementary Table 2 Different Master Mixes used for PCR in a single reaction and in two reactions during genotyping

PCR 1 X 3 primers mix	**10µl**
MQ H2O	6.3
Buffer 5XGreen	2
dNTPs	0.35
Primer LP	0.1
Primer RP	0.1
Primer BP	0.1
GoTaq	0.05
DNA Template	1
PCR 2 X 2 primers mix	**10µl**
MQ H2O	6.15
Buffer 5XGreen	2
dNTPs	0.4
Primer LP/Primer BP	0.2
Primer RP	0.2
GoTaq	0.05
DNA Template	1

Supplementary Table 3 Left Primer and Right Primer used for every line. Primer for line 2 was manually designed. The rest were designed the SALK T-DNA Primer design tool (Http://signal.salk.edu/tdnaprimers.2.html)

Short ID	Gene	Line ID	NASC ID	LP	RP
1	AT3G45770	SALK_003308	N503308	AAACCCACGCAACATATGAAG	TTCCCACTGTTCTGAAGGTTG
2	AT3G45760	SALK_018808		GCGAGTCATCAAAGGTGGTTC	GATGGTTTTGGACGGGTCATT
3	AT3G45660	SALK_030923	N530923	CTGTAACCATCCCAACGTAGC	TTGATGTTGAGCTTGTGATGC
4	At2g18010	SALK_065537	N565537	AGGGACTCTGGTTGGTAGGAC	TGGGTCTAACCATTCCTTGTG
5	AT3G45740	SALK_118254	N618254	TGGACCAGGTTTGAAAATGAG	GGGGATGTAGCGATTTCTTTC
6	AT3G45770	SALK_130583	N630583	AAACCGACAACACCTCTAAACC	GCTACTGATTTTAGGGCTGGG
7	At1g61860	SALK_130390C	N654894	TAAAAGAGAGAGACGCGTTGG	TCTTCCAAAGATCCGTTAGGC
8	At2g17970	SALK_111811C	N655143	GATCCCTAAACCCACGTGTTC	TGGAAATGGAGCTGATGTAGC
9	AT3G45760	SALK_109748C	N655825	CGTGGCTAGAAATTTTGCAAG	TCTGAAGCGAATTGGCATTAC
10	AT3G45740	SALK_118255C	N655868	ATTCACCCTTTCGGAAGCTAG	GATTTGGTTTGCCAAATGATG
11	AT3G45650	SALK_091226C	N657726	GTTGAACGCTGCATAGTCTCC	GCAGCTACGTTGTTAGGCTTG
12	At2g18050	SALK_025209C	N665594	CAAAGCCTCTCGGTAAATGTG	TATTCTTCTTGTCCTGCTGCC
13	At2g18042	SALK_064492C	N670847	TTCCCCATTTTGTTCTTAAAAAG	TTGTTGTTCATCAATCAAAAGC
14	At2g18060	SALK_022534C	N672318	CTCGTTTTAAGCGGATGTTTG	ATGACGGGAAATTGGAGAGAG
15	AT3G45630	SALK_061949C	N675478	CAGTGCCTTGAAACCAGAGAG	TGGTGACATAATTATTGCAGGTC
16	AT3G45620	SALK_026980C	N680453	CTTTCCTTGTGCAAACTCCTC	GTCTTACCAGACGCTCCTGTG
17	At2g18070	SALK_017125C	N682049	CTTTCAGATATTTTCGCGGTG	AAATTGTACTCCGGAAATCGC
18	At2g18030	SALK_201557C	N689098	TCATGCCTTAATTGTTAGCCG	TTCTTTGCTGGTAGAAGCCAG
19	AT3G45670	SALK_201949C	N689722	TTTCGTGCTAACGAAATCTGG	TGCAATCTTTGTCAAAAACCC
20	At2g18040	SAIL_15_A04	N800698	TTCTTGGCTCTCGTTAATCTCAC	TCCCATACAATGGAGTTCTGC
21	At2g18030	SAIL_47_C10	N802260	AAACTCTCAGTCGCAGAGGTG	TTGGCTCAATCCTCAAAATTG
22	AT3G45780	SAIL_1232_C01	N861611	ACATAGGATGCAGCAGAAACG	CAGTAGACTGGTGGGCTCTTG
23	AT3G45450	SAILseq_828_A03.1		ACATCACCACTAGCAATTCGC	CAAAGGTTCTCAGGGTTAGGG

Supplementary Table 4 Manually designed Left Primer and Right Primer pairs for Lines 13 and 18

Short ID	Gene	LP	RP
13	At2g18042	ACCAGAGCTTTGTTCGTCATC	TTCCTGGATGCATTGTAGAGG
18	At2g18030	TTCTTGTGTGTCTCTCTGCTCG	AGATGTCTCTCAAGACGGGC

Supplementary Table 5 T-DNA Border Primers (BP) used for genotyping based on whether the line was created from the SALK or SAIL insertion collections

BP SALK	LBb1.3	ATTTTGCCGATTTCGGAAC
BP SAIL	LB3	TAGCATCTGAATTTCATAACCAATCTCGATACAC

Supplementary Text 1 CTAB protocol used for DNA extraction

1. Label tubes and add 2 glass beads 3mm in diameter.
2. Collect the leaves in 1.5 ml screwed tubes, and store in -20^{o}C until grinding.
3. Grind material with liquid N2, frequency of 30Hz for 1.5 min. Repeat 1 or 2 times until the tissue has reached a fine powder state. Store at -80^{o}C until extraction.
4. Heat up water bath to 65^{o}C.
5. Add 500 µl of CTAB buffer, shake by hand for 1 min and tap the tube until all clumps are dissolved.
6. Incubate at 65^{o}C for 30 min, invert tubes 2 or 3 times during incubation. Afterwards, let them cool down to room temperature for 30 min.
7. Briefly centrifuge.
8. Add 500 µl of Chloroform:isoamylalcohol (24:1) (under hood), shake by hand vigorously for 80 sec. Centrifuge at max speed 13600 rpm for 5 min.
9. Transfer 350 µl of supernatant (gently pipet) into new Eppendorf tubes.
10. Add 350 µl of cold isopropanol in the supernatant. Gently invert tubes for several times.
11. Incubate on ice for 30 min.
12. Centrifuge at max speed 13600 rpm for 15 min.
13. Decant isopropanol and recover the DNA pellet on tissue paper to drain as much isopropanol.
14. Clean DNA pellet with 1 ml of 70% ethanol. Centrifuge at max speed for 13600 rpm for 5 min.
15. Decant ethanol and recover DNA pellet on tissue paper to drain as much ethanol.
16. Briefly centrifuge DNA pellet and drain the rest of ethanol by pipetting out.
17. Let DNA pellet dry completely when ethanol has evaporated.
18. Dissolve DNA in 100 µl of water (containing RNase A 10ug/ml) overnight at 4^{o}C.

Supplementary Text 2 ZymoResearch Direct-zol™ RNA Miniprep protocol

- ✓ Perform all steps at room temperature and centrifugation at 10,000-16,000 x g for 30 seconds, unless specified.

1. Lyse cells in an appropriate volume of TRI Reagent® or similar and mix thoroughly
2. Add an equal volume ethanol (95-100%) to a sample lysed in TRI Reagent® or similar1 and mix thoroughly.
3. Transfer the mixture into a Zymo-Spin™ IICR Column2 in a Collection Tube and centrifuge3. Transfer the column into a new collection tube and discard the flow-through.

4. Add 400 µl Direct-zol™ RNA PreWash5 to the column and centrifuge. Discard the flow-through and repeat this step.
5. Add 700 µl RNA Wash Buffer5 to the column and centrifuge for 1 minute to ensure complete removal of the wash buffer. Transfer the column carefully into an RNase-free tube (not included).
6. To elute RNA, add 50 µl of DNase/RNase-Free Water directly to the column matrix and centrifuge.

Supplementary Text 3 RNA DNAse treatment and precipitation protocol

1. Set up the DNase digestion reaction as follows:

   RNA sample 25ul
   DNase buffer 3ul
   DNase RQ Promega 2ul

2. Incubate in at 37℃ for 45minutes
3. Add 70 µL of nuclease-free water, 50 µL of 7.5 M NH4Ac (ammonium acetate) and 400 µL of absolute 100% ethanol to 30 µL of RNA.
4. Incubate at -20℃ (fridge in molecular lab) for 1 hour or overnight.

RNA precipitation
Turn on the 4℃ centrifuge in advance.
1. Spin at 13000 rpm at 4℃ for 20-30 minutes and discard the supernatant.
2. Wash pellet with absolute 70% ethanol and spin for 15 min, discard the liquid. Repeat the step.

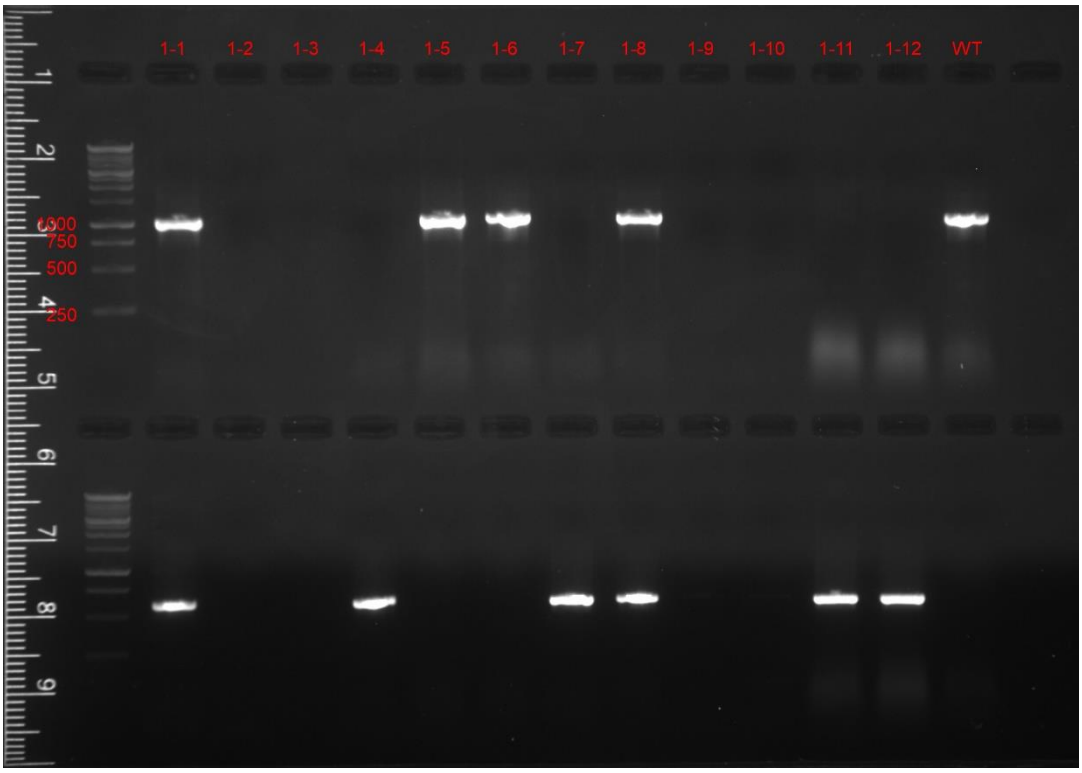
3. Air-dry in the fume hood.
4. Resuspend to make sure the final concentration is not less than 50 ng/ul.

Supplementary Text 4 R code used to create the boxplot graph comparing Fv/Fm of line9 and the wild type. All graphs used were created by modifying this code

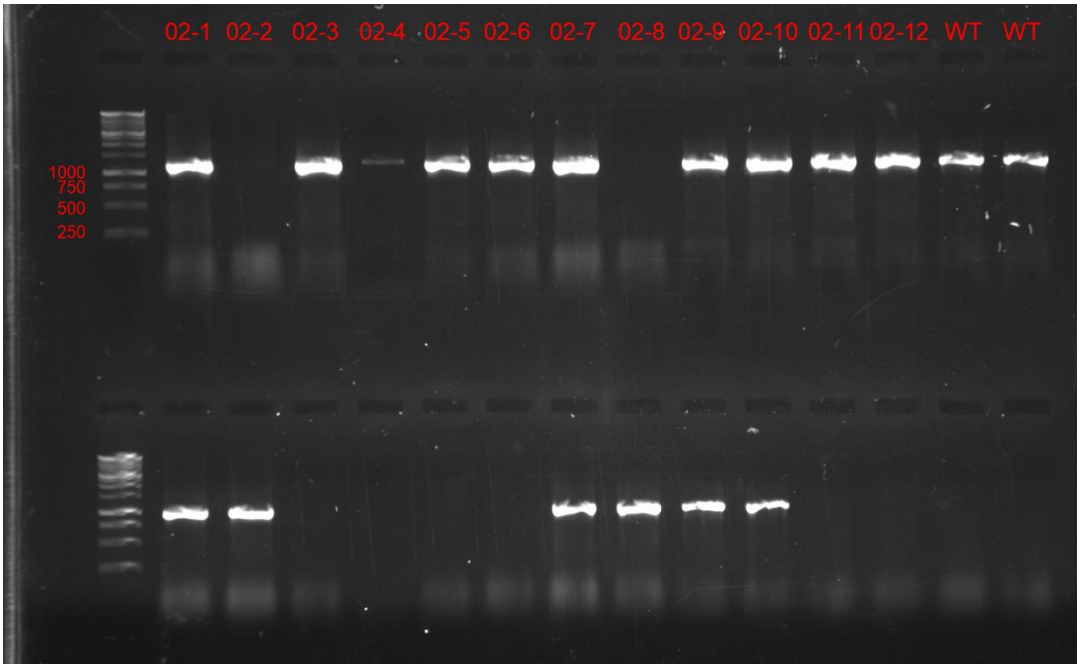
```
library(ggplot2)

library(tidyr)

data <- read.csv("LINE9COLFVFM.csv")


columns <- c("cFvFm1", "cFvFm1col", "cFvFm3", "cFvFm3col", "cFvFm5", "cFvFm5col", "cFvFm7",
"cFvFm7col", "cFvFm9", "cFvFm9col", "cFvFm11", "cFvFm11col", "cFvFm13", "cFvFm13col",
"cFvFm15", "cFvFm15col", "cFvFm17", "cFvFm17col", "cFvFm19", "cFvFm19col","cFvFm21",
"cFvFm21col", "cFvFm23", "cFvFm23col", "cFvFm25", "cFvFm25col", "cFvFm27", "cFvFm27col",
"cFvFm29", "cFvFm29col","cFvFm31", "cFvFm31col")


df <- data[, columns]


df_long <- gather(df, key = "variable", value = "value")
```

```
variable_names <- c("6_1", "6_2", "7_1", "7_2", "8_1", "8_2", "9_1", "9_2", "10_1", "10_2", "11_1",
"11_2", "12_1", "12_2", "13_1", "13_2", "14_1", "14_2", "15_1", "15_2", "16_1", "16_2", "17_1",
"17_2", "18_1", "18_2", "19_1", "19_2", "20_1", "20_2", "21_1", "21_2")


df_long$variable <- factor(df_long$variable, levels = columns)


levels(df_long$variable) <- variable_names


boxplot_colors <- rep(c("grey51", "grey99"), length(columns)/2)


ggplot(df_long, aes(x = variable, y = value, fill = variable)) +
  geom_boxplot(fill = boxplot_colors) +
  xlab("DAS") +
  ylab("Fv/Fm") +
  ggtitle("Line 9 - Columbia-0") +
  theme(legend.position = "none")


ggsave("9colfvfmcolor.png", width = 3840, height = 2160, units = "px")
```

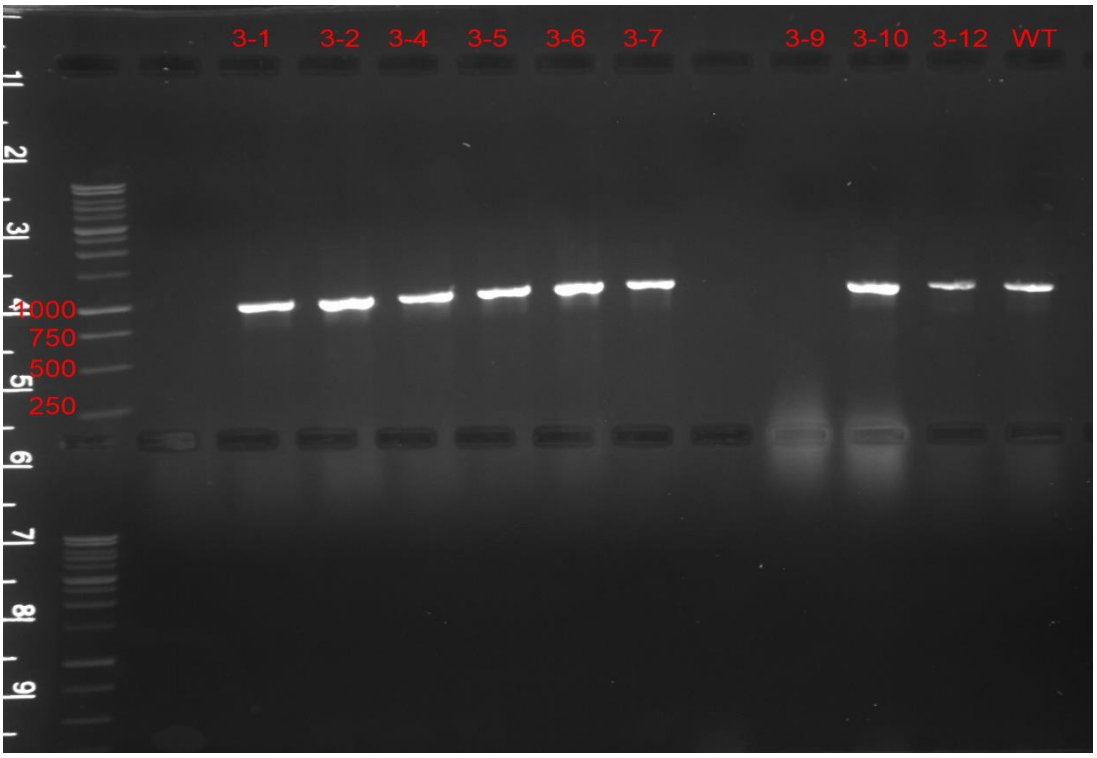
Supplementary Figure 1 Gel results for Line 1. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 1, indicated by number 1-1 to 1-12. Last lane: wild type control



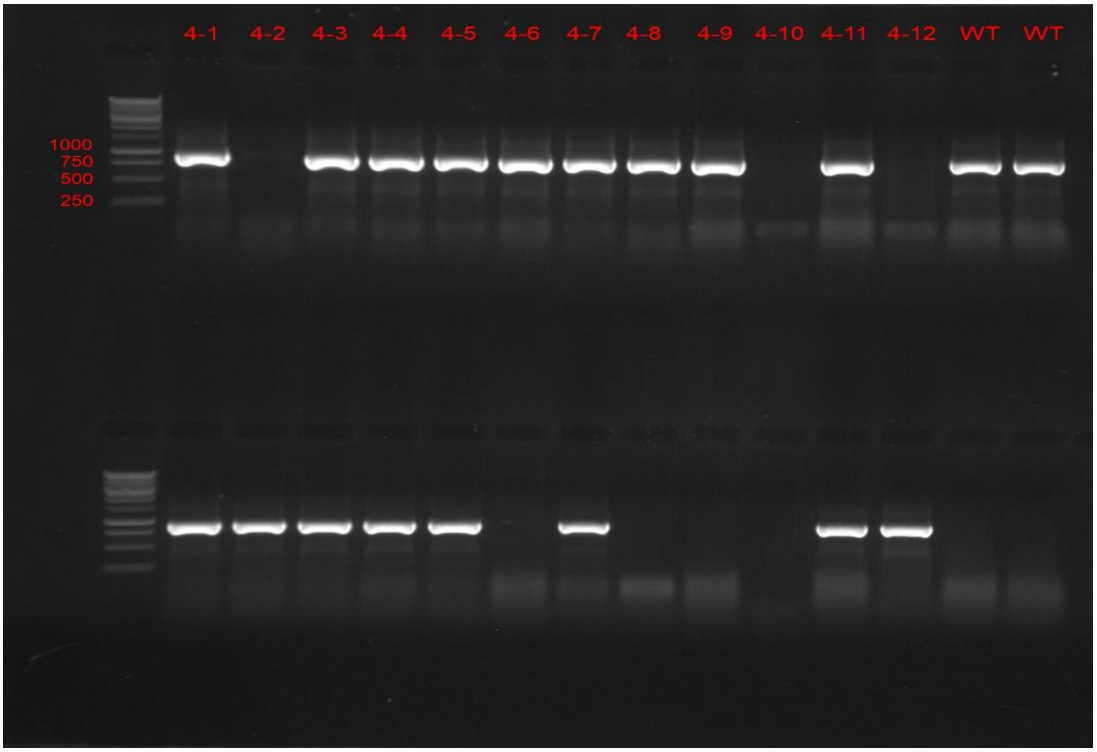
Supplementary Figure 2 Gel results for Line 2. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 2, indicated by number 2-1 to 2-12. Last two lanes: wild type controls



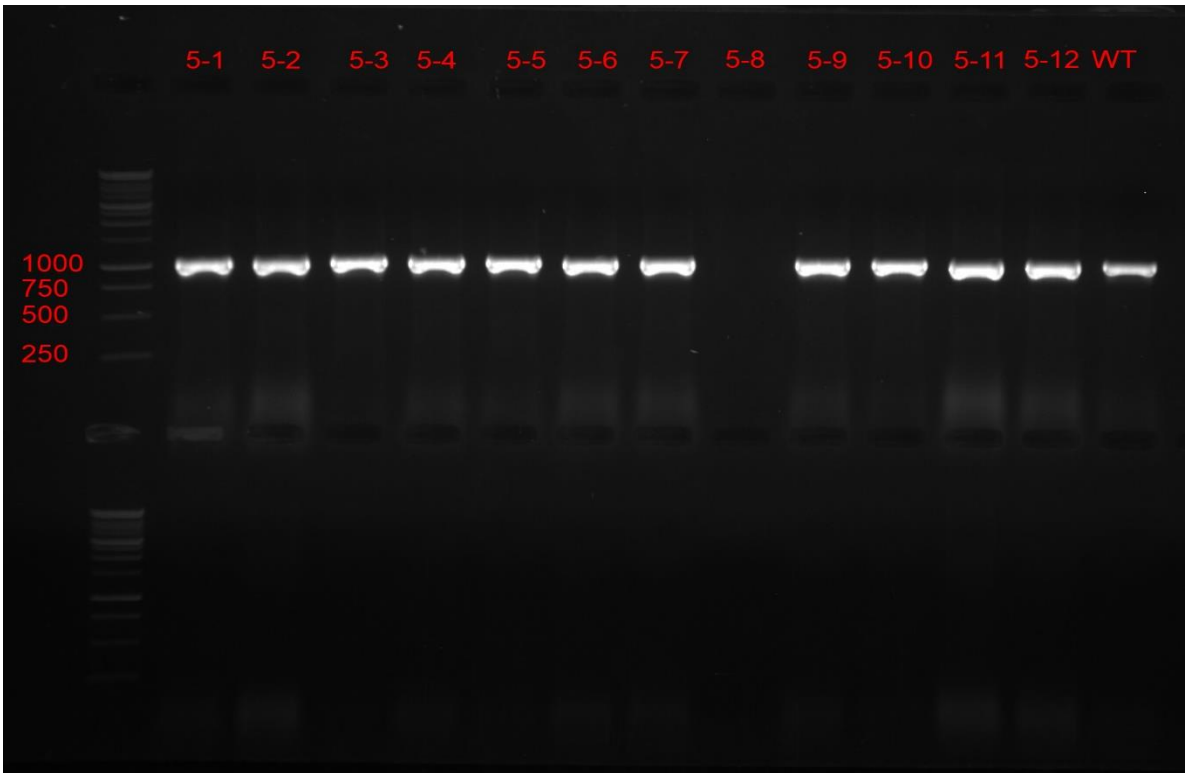
Supplementary Figure 3 Gel results for Line 3. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 3, indicated by number 3-1 to 3-12. Individuals 3-3 and 3-8 did not germinate. Last lane: wild type controls



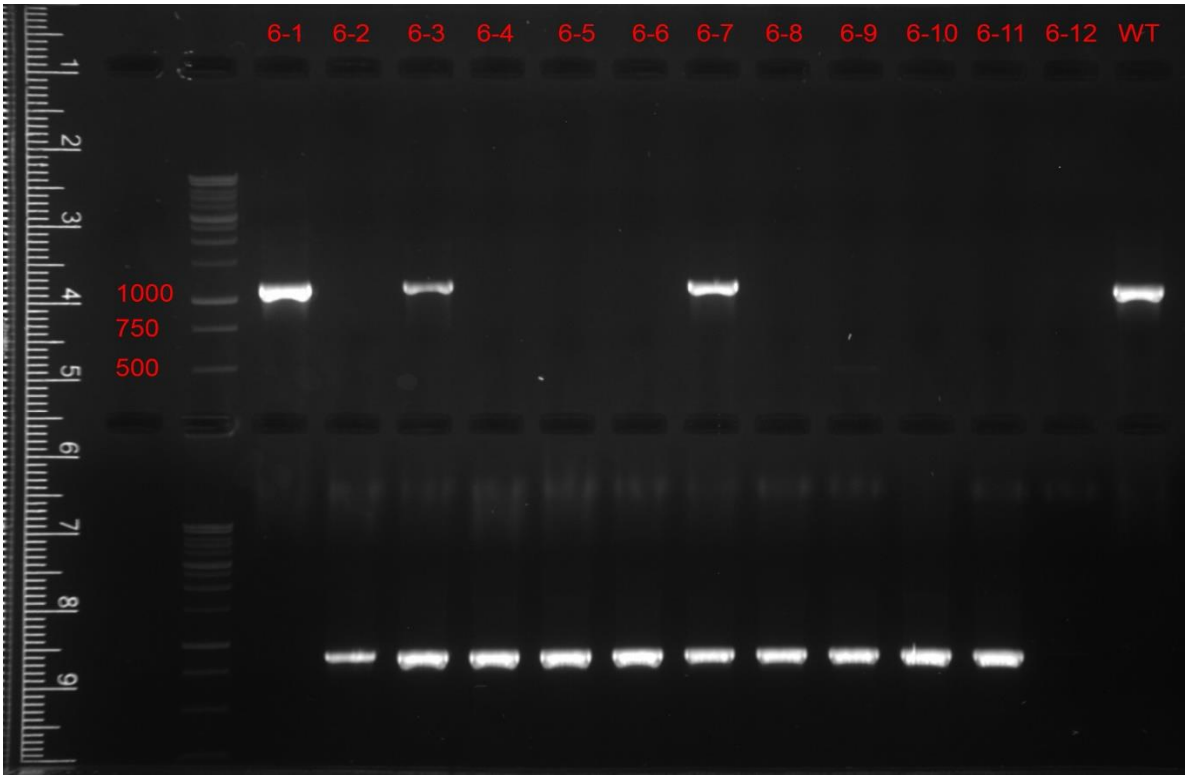
Supplementary Figure 4 Gel results for Line 4. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 4, indicated by number 4-1 to 4-12. Last two lanes: wild type controls



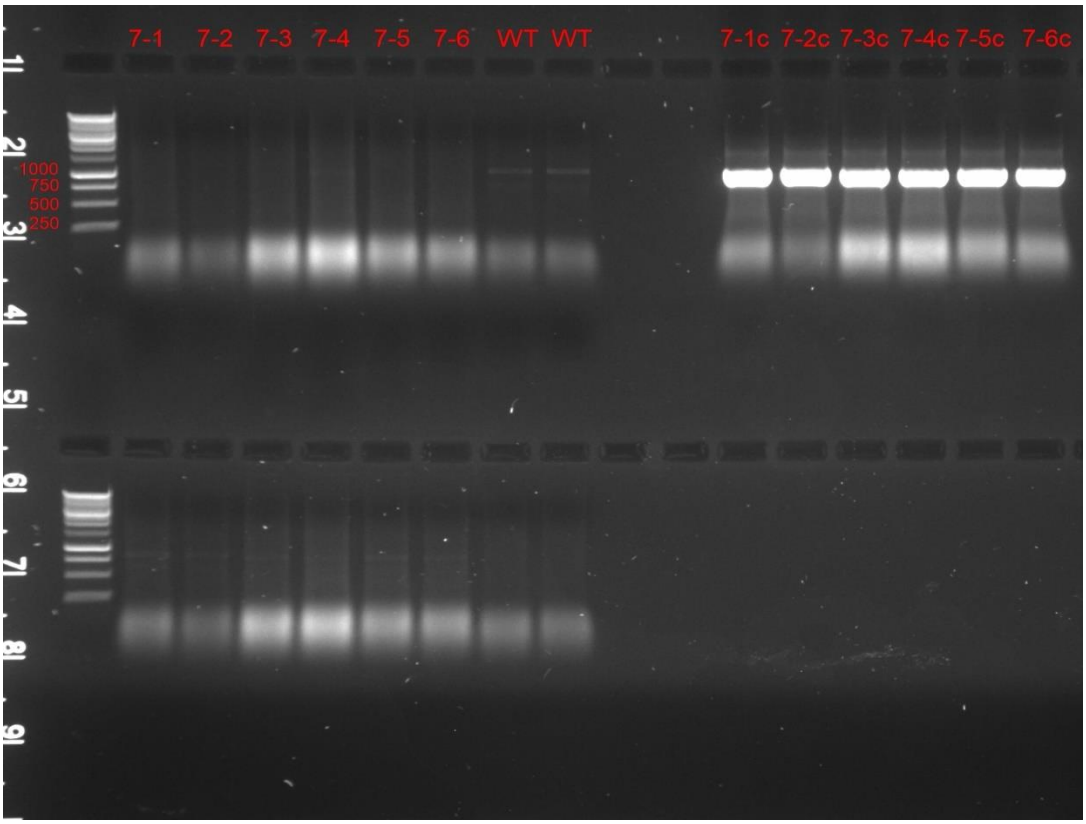
Supplementary Figure 5 Gel results for Line 5. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 5, indicated by number 5-1 to 5-12. Individual 5-8 did not germinate. Last lane: wild type control



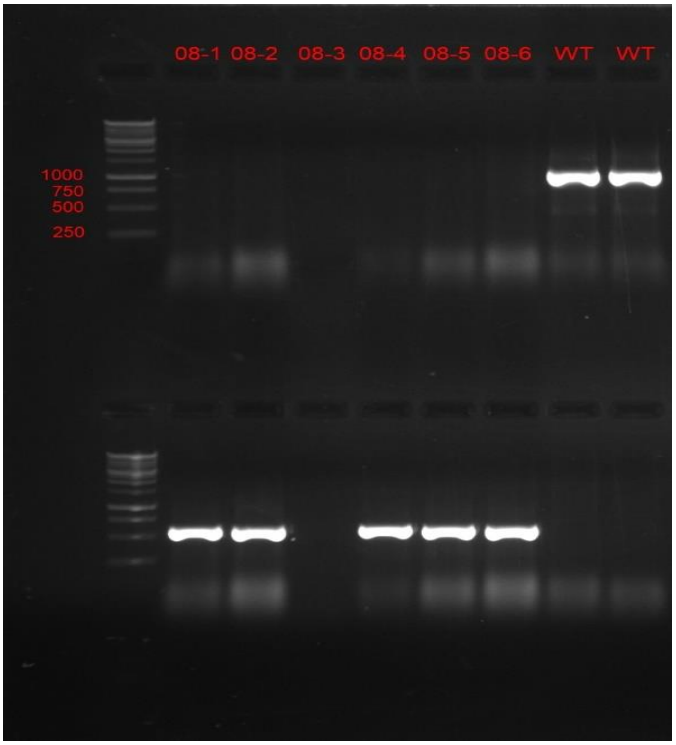
Supplementary Figure 6 Gel results for Line 6. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 6, indicated by number 6-1 to 6-12. Last lane: wild type control



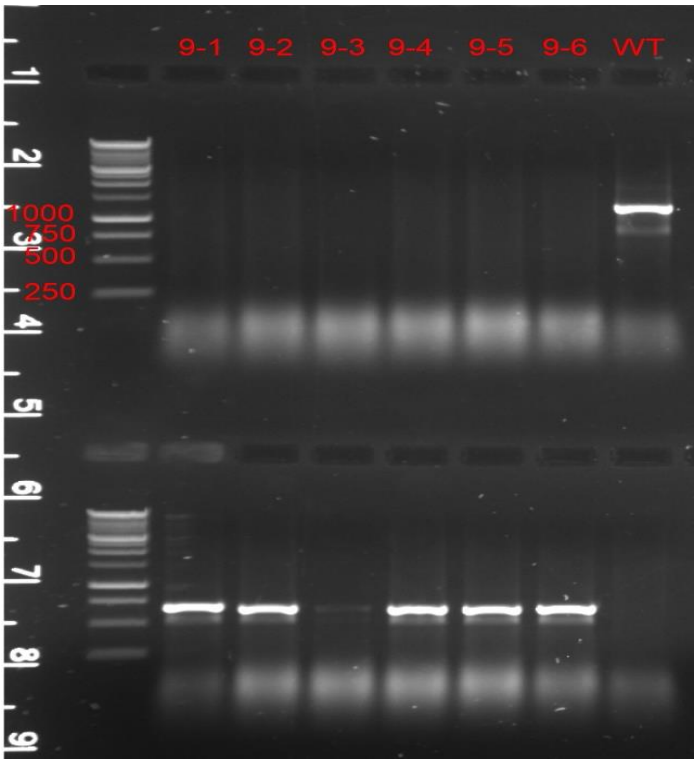
Supplementary Figure 7 Gel results for Line 7. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 7, indicated by number 7-1 to 7-6.. Last lanes: wild type controls. 7-1c to 7-6c: individuals of line 7 amplified with LP-RP primers from line 5, to check for DNA quality



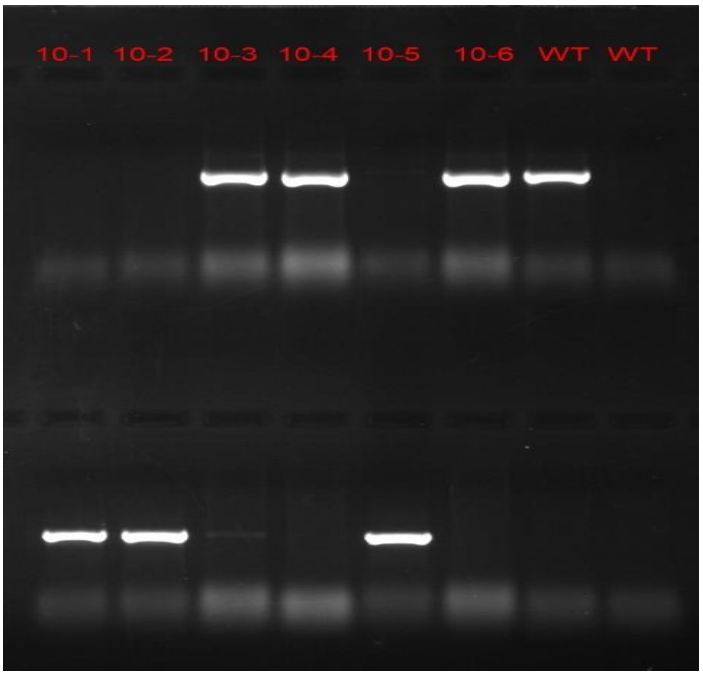
Supplementary Figure 8 Gel results for Line 8. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 8, indicated by number 8-1 to 8-6. Last lanes: wild type controls



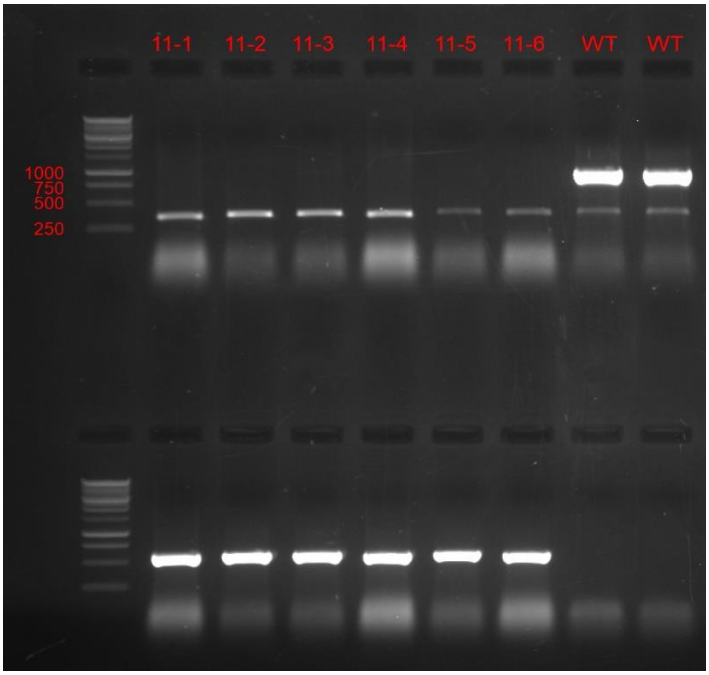
Supplementary Figure 9 Gel results for Line 9. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 9, indicated by number 9-1 to 9-6. Last lane: wild type control



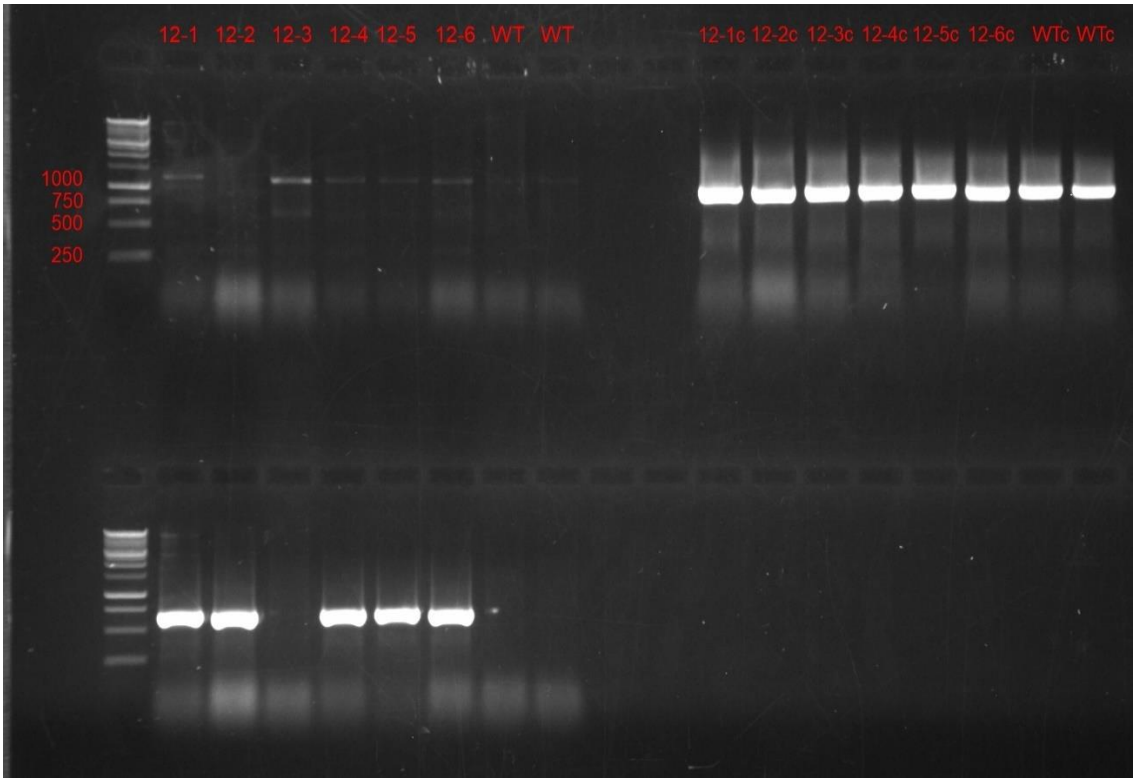
Supplementary Figure 10 Gel results for Line 10. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 10, indicated by number 10-1 to 10-6. Last lanes: wild type controls. Example where one of the wild type control did not amplify, despite them both being under the same conditions



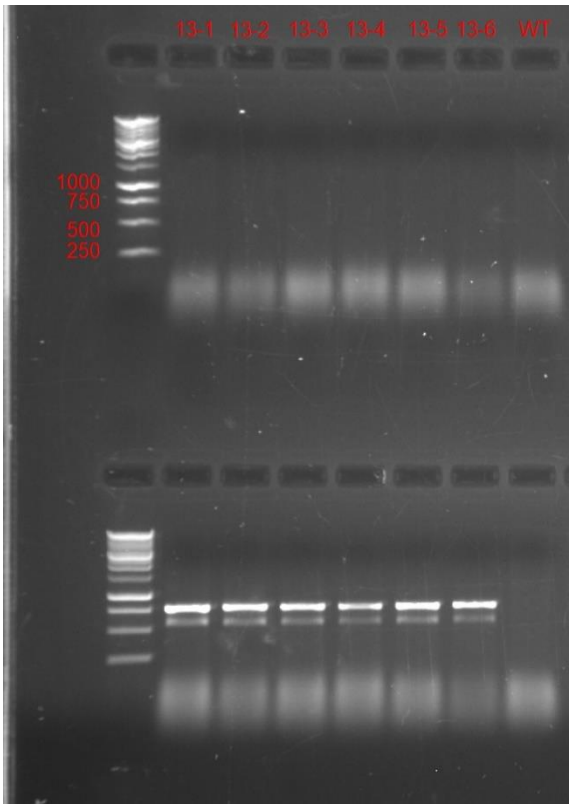
Supplementary Figure 11 Gel results for Line 11. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 11, indicated by number 11-1 to 11-6. Last lanes: wild type controls



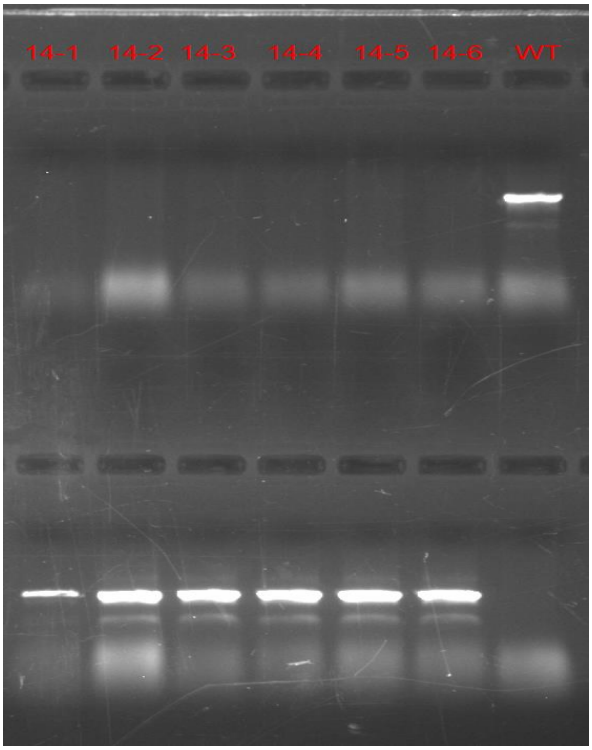
Supplementary Figure 12 Gel results for Line 12. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 12, indicated by number 12-1 to 12-6. Last lanes: wild type controls. 12-1c to 12-6c: individuals of line 12 amplified with LP-RP primers from line 5, to check for DNA quality. Quality of the wild type was checked as well



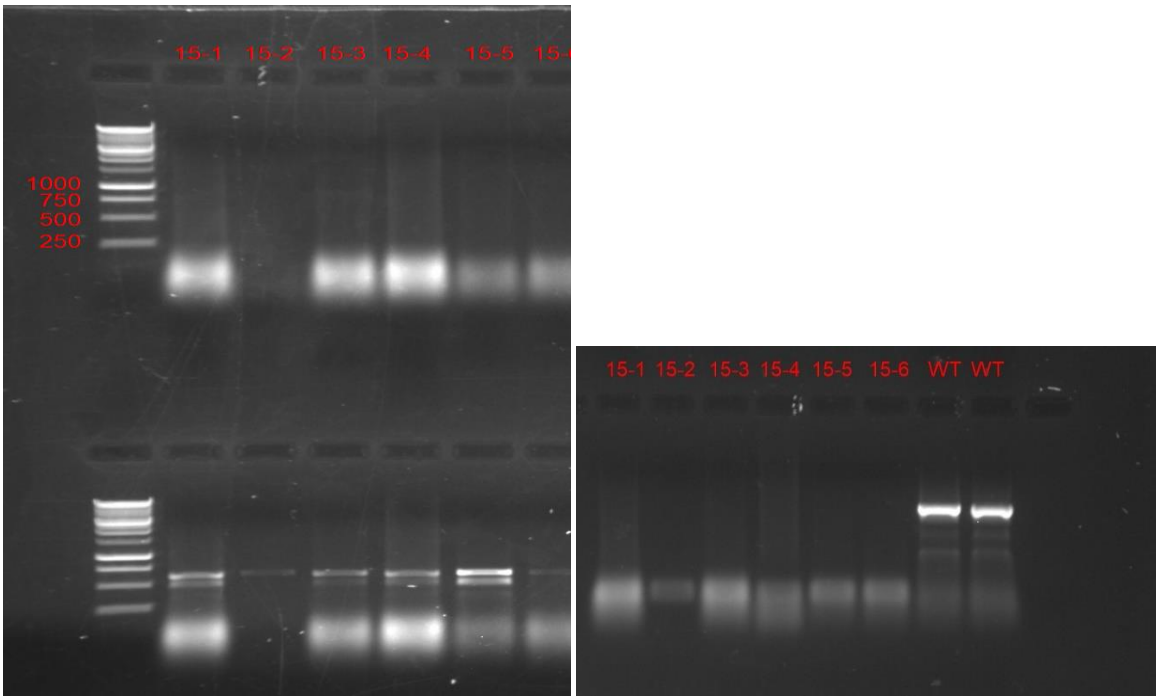
Supplementary Figure 13 Gel results for Line 13. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 13, indicated by number 13-1 to 13-6. Last lane: wild type control. One of the two cases where wild type DNA could not be amplified



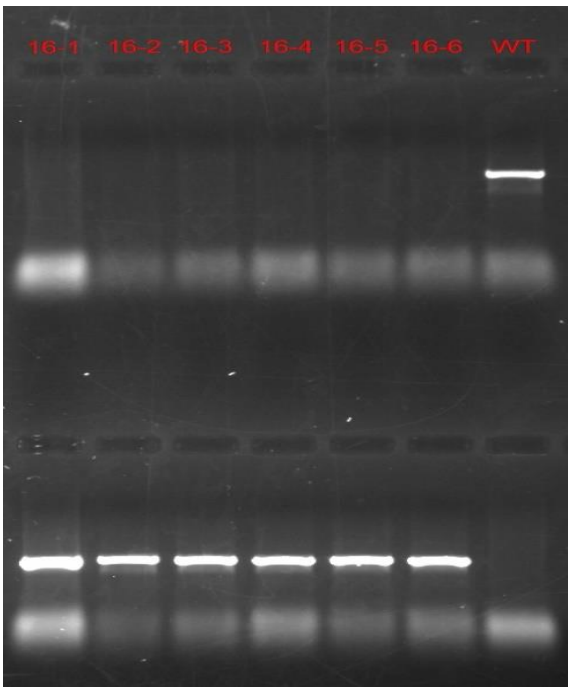
Supplementary Figure 14 Gel results for Line 14. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 14, indicated by number 14-1 to 14-6. Last lane: wild type control



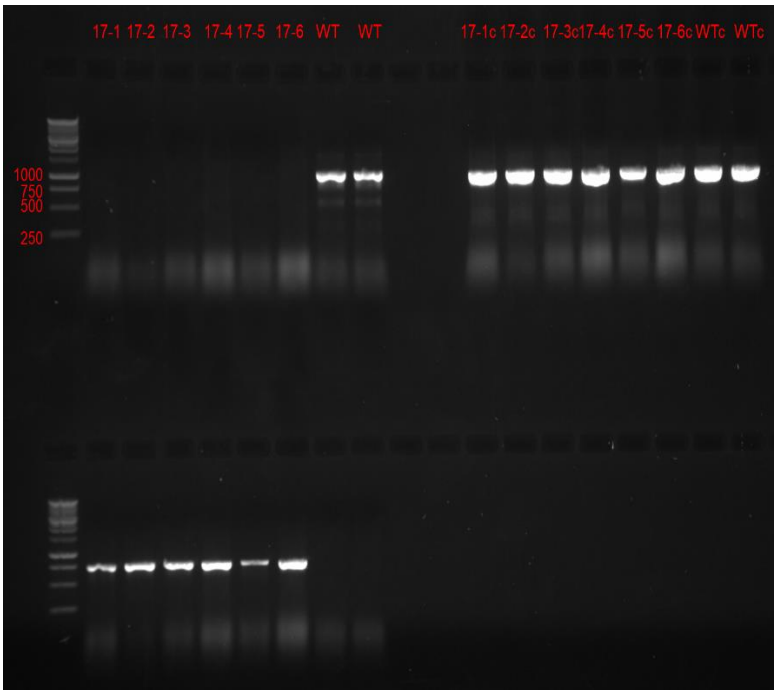
Supplementary Figure 15 Left picture: Gel results for Line 15. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 15, indicated by number 15-1 to 15-6. Right picture: the wild type reaction repeated due to initial failure to amplify the wild type controls



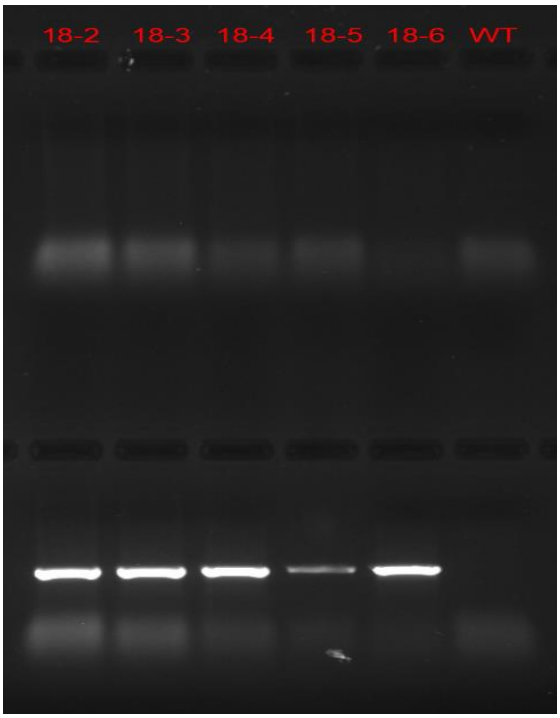
Supplementary Figure 16 Gel results for Line 16. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 16, indicated by number 16-1 to 16-6. Last lane: wild type control



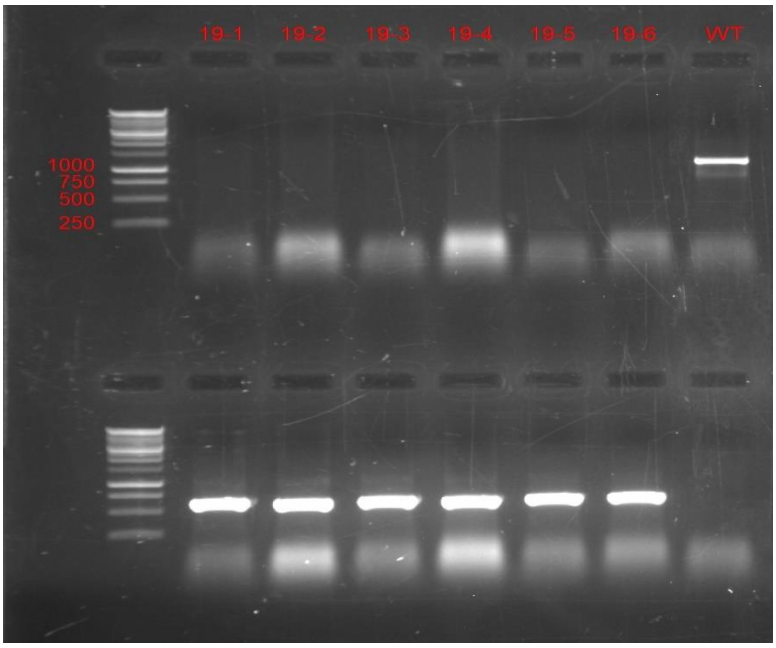
Supplementary Figure 17 Gel results for Line 17. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 17, indicated by number 17-1 to 17-6. Last lanes: wild type controls. 17-1c to 17-6c: individuals of line 17 amplified with LP-RP primers from line 5, to check for DNA quality. Quality of the wild type was checked as well



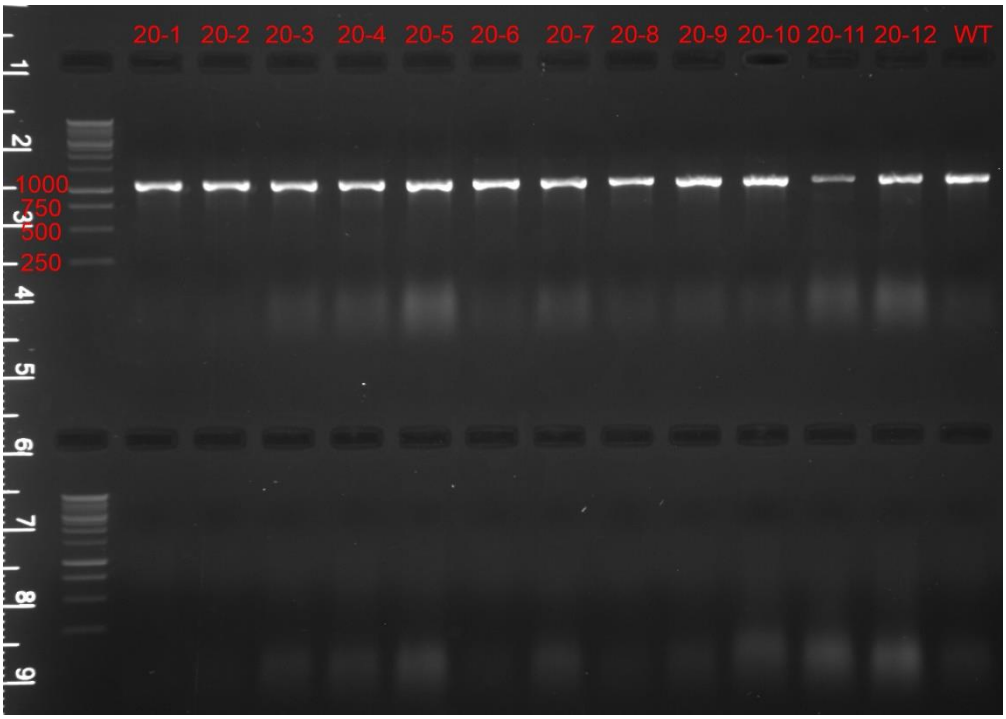
Supplementary Figure 18 Gel results for Line 18. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 18, indicated by number 18-2 to 18-6. Individual 18-1 did not germinate. Last lane: wild type control. The second case where wild type DNA could not be amplified



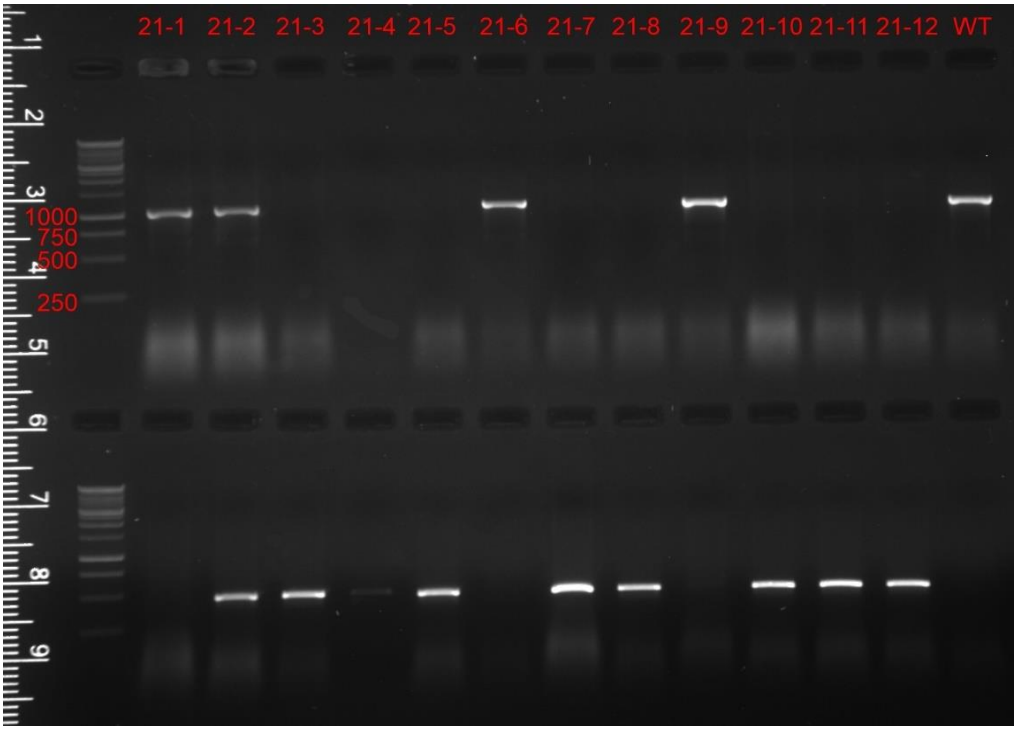
Supplementary Figure 19 Gel results for Line 19. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 19, indicated by number 19-1 to 19-6. Last lane: wild type control



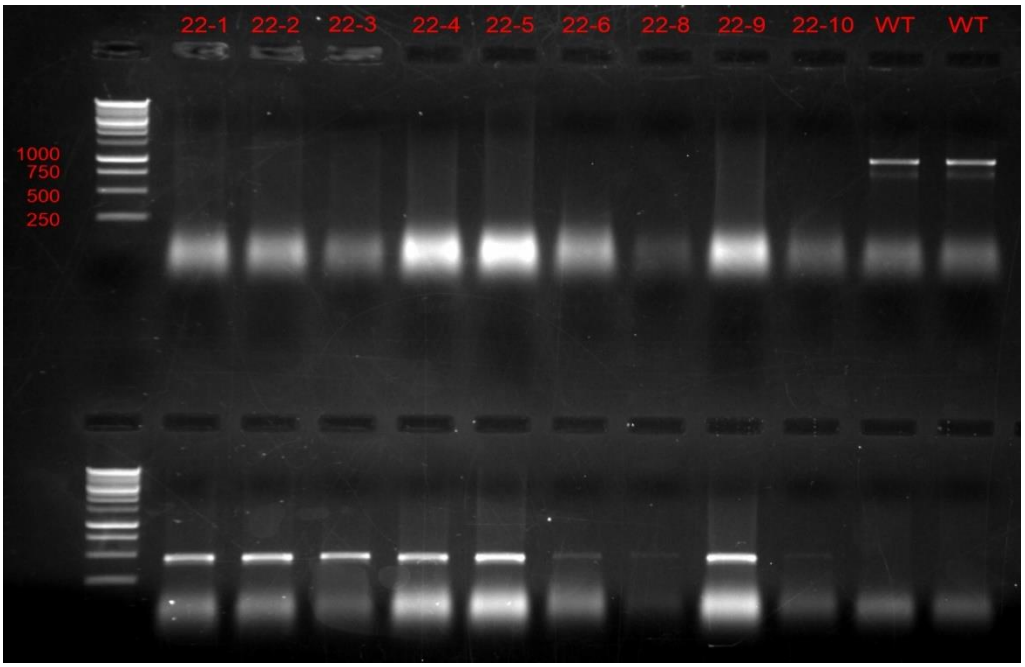
Supplementary Figure 20 Gel results for Line 20. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 20, indicated by number 20-1 to 20-12. Last lane: wild type control



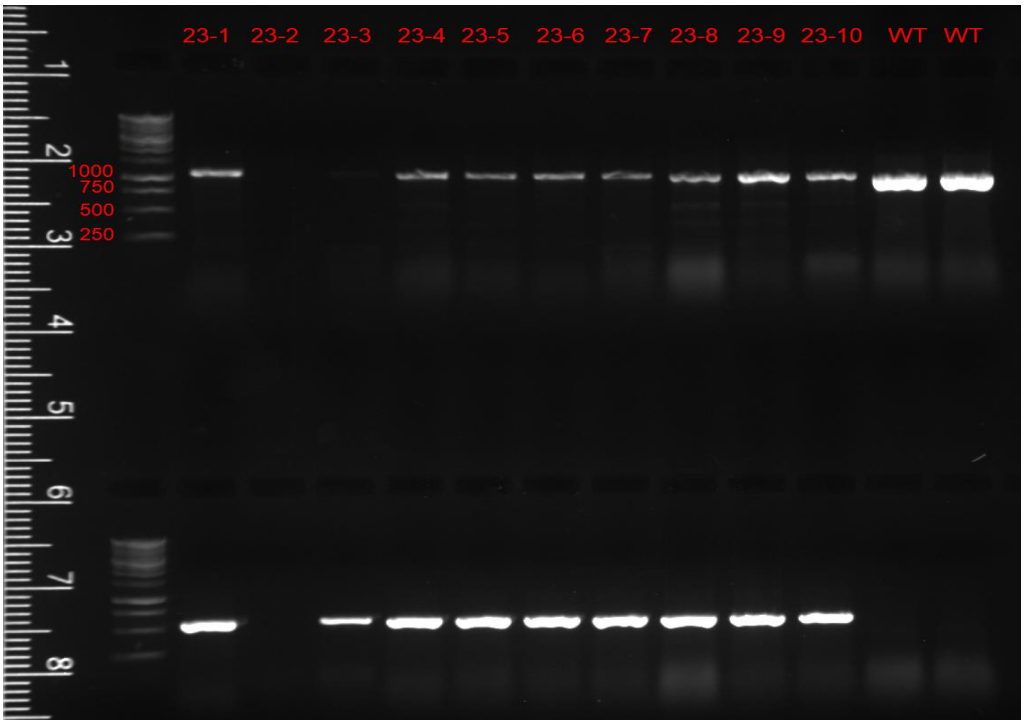
Supplementary Figure 21Gel results for Line 21. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 21, indicated by number 21-1 to 21-12. Last lane: wild type control



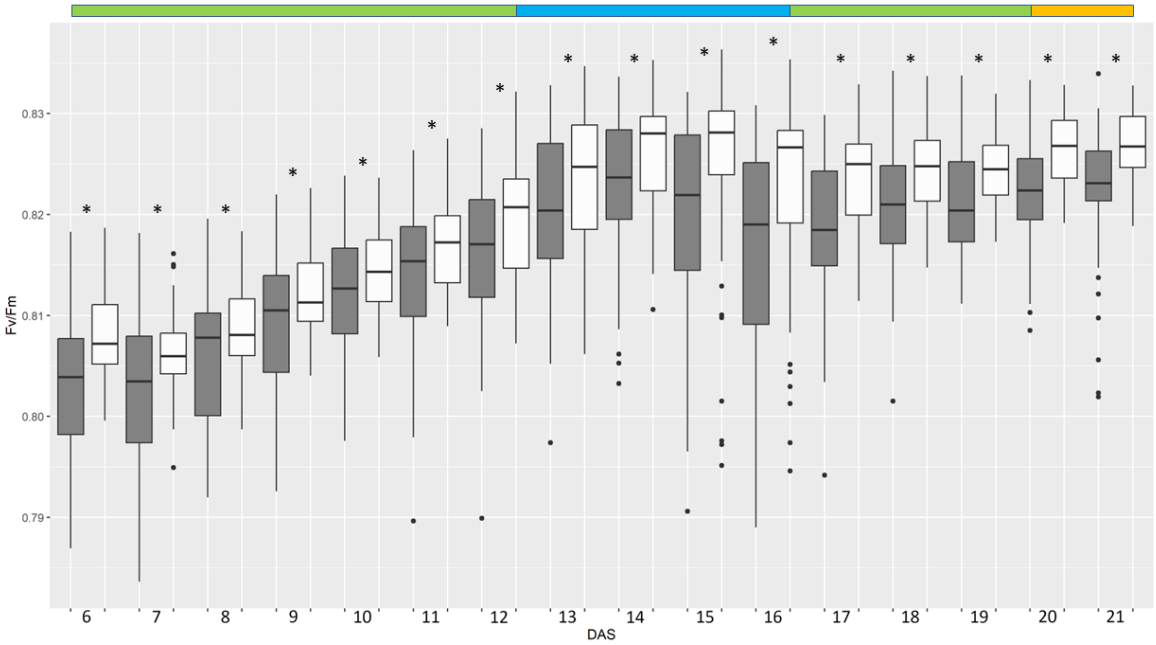
Supplementary Figure 22 Gel results for Line 22. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 22, indicated by number 22-1 to 22-10. Individual 22-7 did not germinate. Last lanes: wild type controls



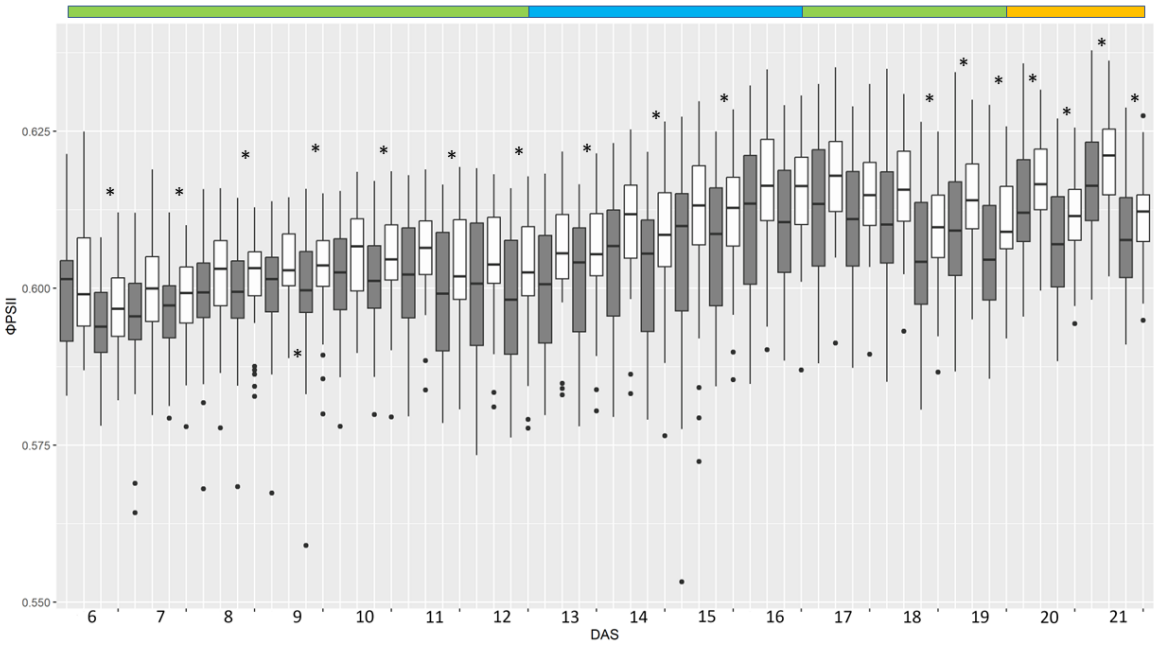
Supplementary Figure 23 Gel results for Line 23. Upper gel: Left primer-Right Primer reaction, genotyping for the wild type allele. Lower gel: Border Primer-Right Primer reaction, genotyping for the mutant allele. First lane: DNA ladder with size (in bp) indication for several bands. The following lanes: individuals of line 23, indicated by number 23-1 to 23-10. Individual 23-7 did not germinate. Last lanes: wild type controls



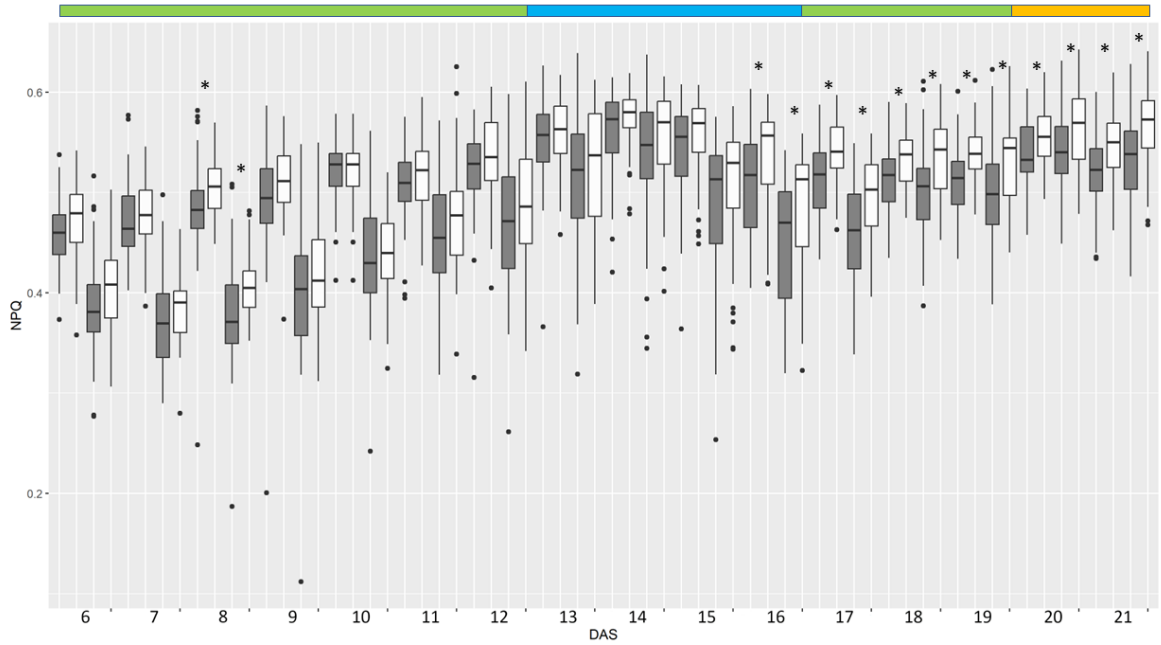
Supplementary Figure 24 Fv/Fm comparison between Line 16 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



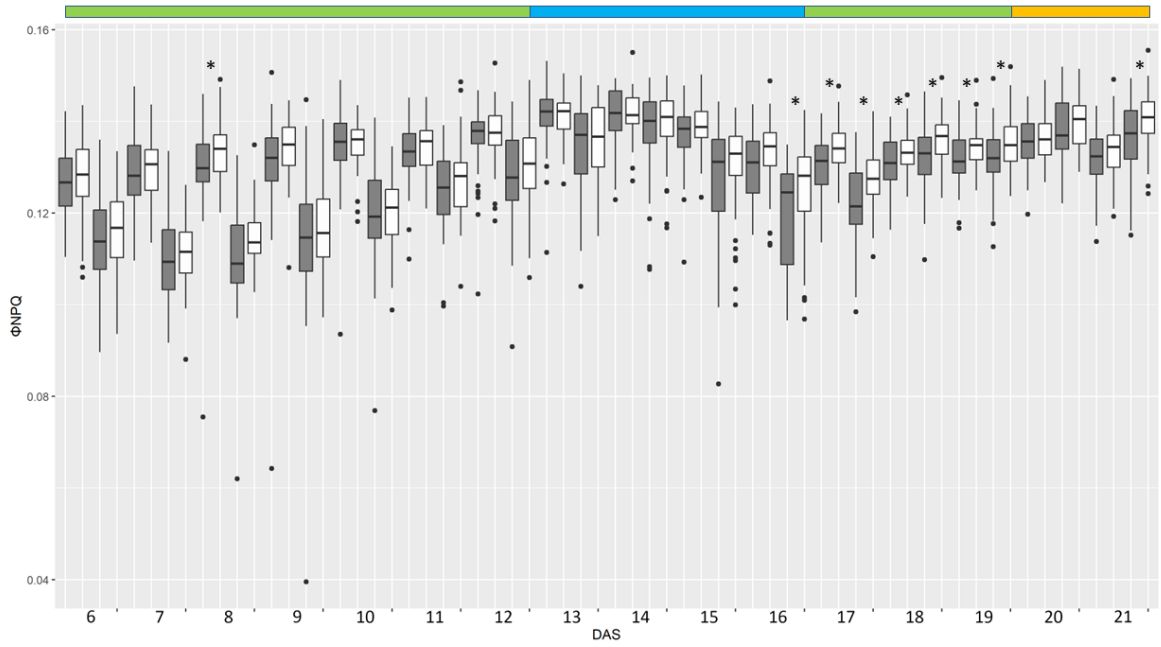
Supplementary Figure 25 ΦPSII comparison between Line 16 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



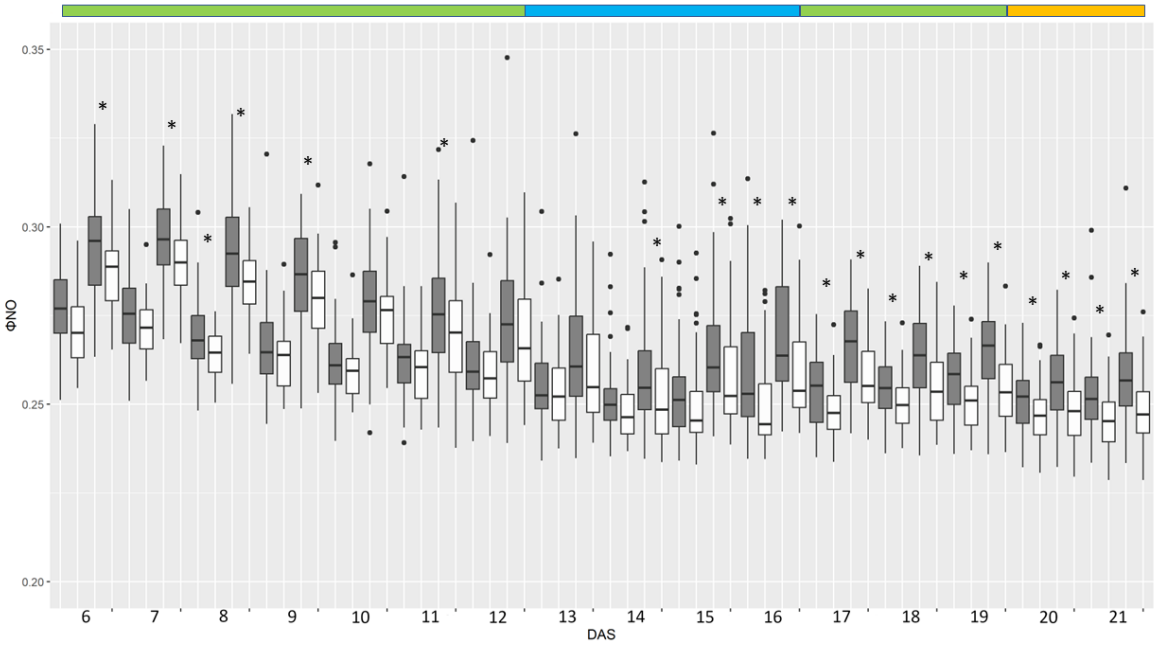
Supplementary Figure 26 NPQ comparison between Line 16 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



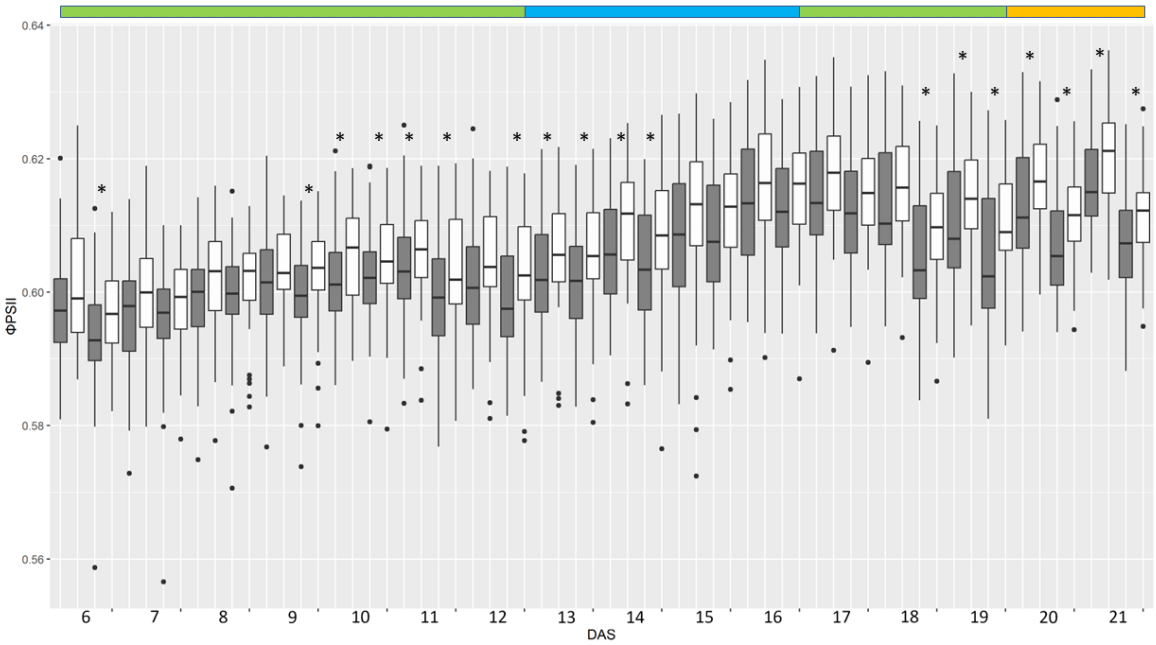
Supplementary Figure 27 ΦNPQ comparison between Line 16 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



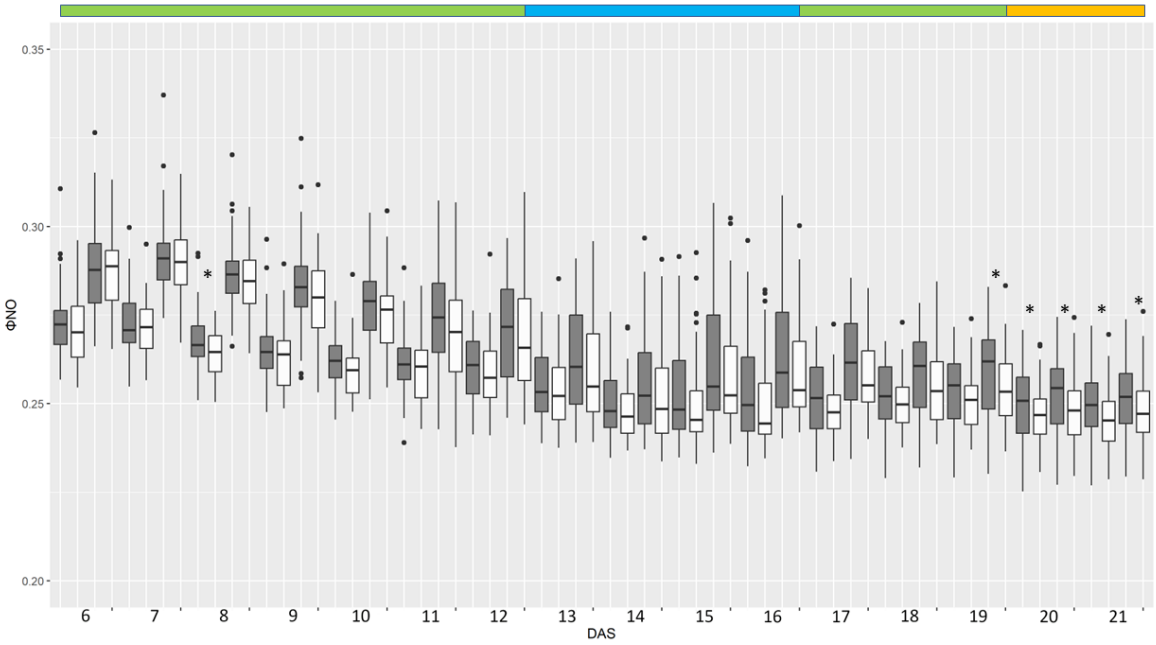
Supplementary Figure 28 ΦNO comparison between Line 16 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



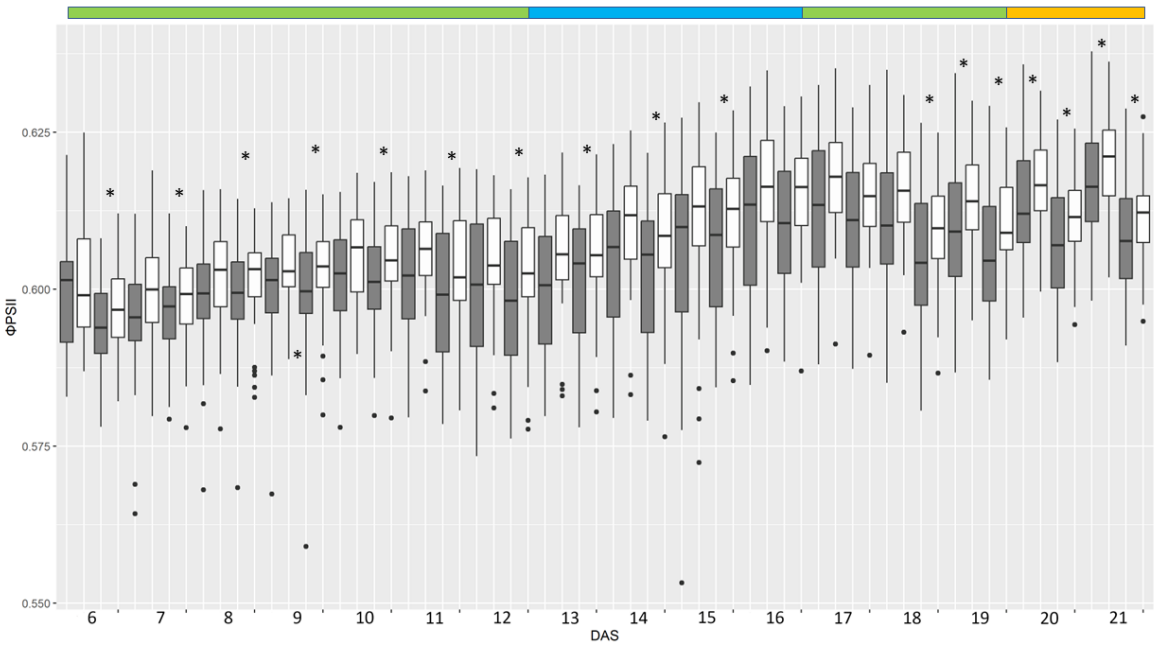
Supplementary Figure 29 ΦPSII comparison between Line 6 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



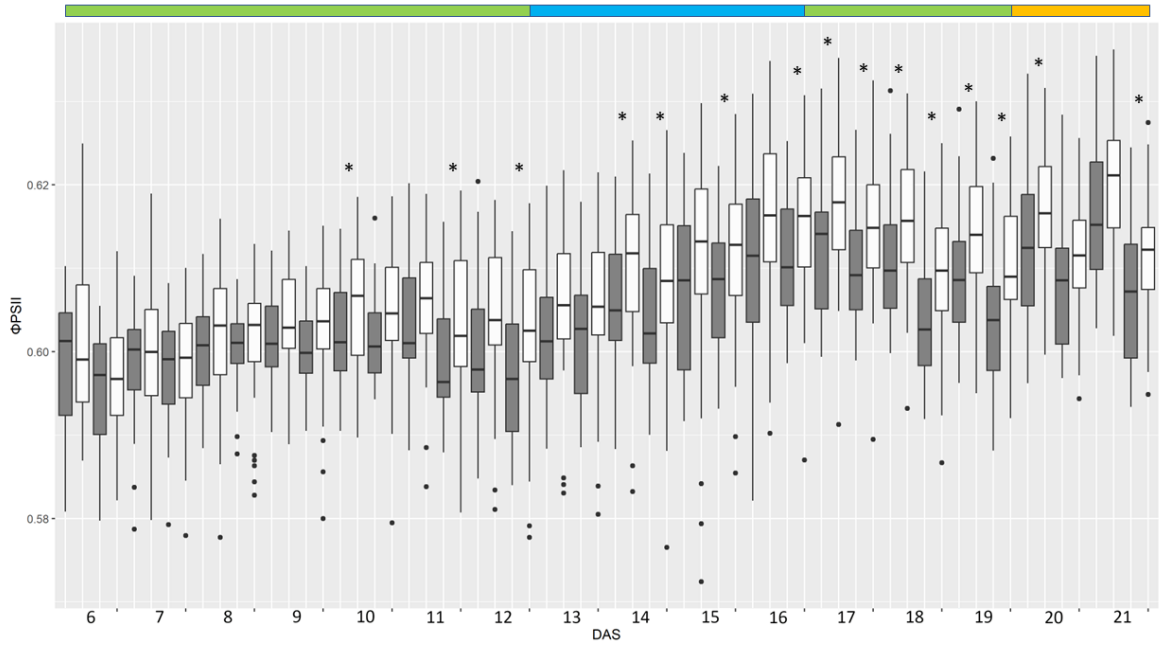
Supplementary Figure 30 ΦNO comparison between Line 6 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes.



Supplementary Figure 31 ΦPSII comparison between Line 18 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes



Supplementary Figure 32 ΦPSII comparison between Line 7 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes



Supplementary Figure 33 ΦNO comparison between Line 7 (grey) and the wild type (white), per timepoint. Asterisks indicate statistically significant deviation. The coloured bar demonstrates the light treatment. Green: constant light. Blue: light fluctuating every 15 minutes. Orange: Light fluctuating every 60 minutes