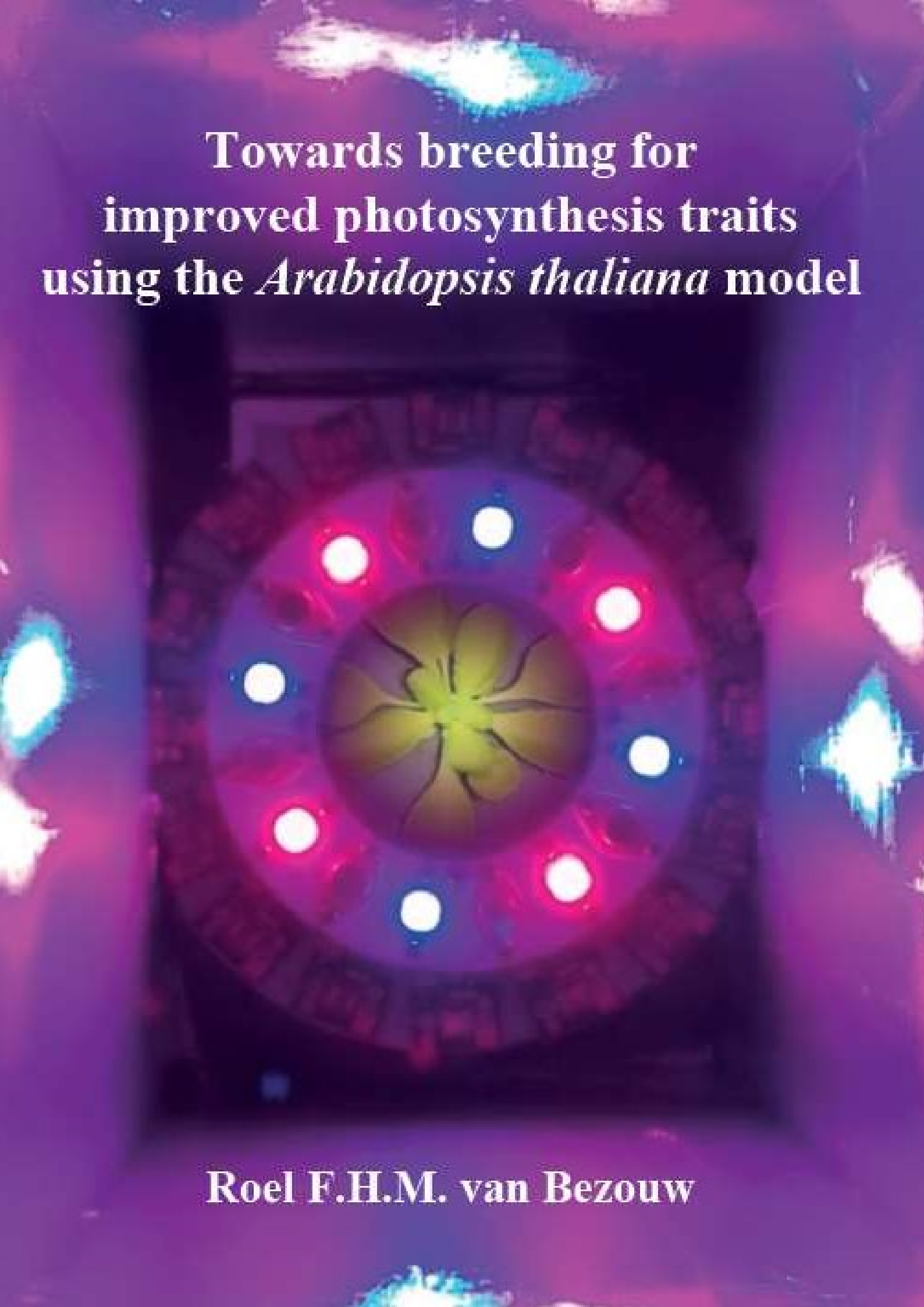
Towards breeding for improved photosynthesis traits using the *Arabidopsis thaliana* model



Roel F.H.M. van Bezouw

Propositions

1. The use of chromosome substitution line populations as efficient quantitative trait mapping tools is inseparable from reverse breeding strategies.
   (this thesis)

2. Copy number variation in SQUALENE EPOXIDASE-LIKE genes demonstrates that there are plenty of opportunities to further improve photosynthesis in plants.
   (this thesis)

3. In quantitative genetic studies, high diversity is more often than not a liability, rather than an advantage.

4. Advancing knowledge is most efficiently performed by invalidating scientific theories and hypotheses.

5. Increasing investment of interdisciplinary research should not come at the expense of specialists.

6. The value of developing soft skills is severely underrated in the PhDtrajectory.

7. The severe imbalance between the number of people trained to install solar panels compared to those trained to approve their installation is a sign of intellectual impoverishment.

Propositions accompanying the thesis entitled:

Towards breeding for improved photosynthesis traits using the Arabidopsis thaliana model

Roel F.H.M. van Bezouw

21 June 2022

Towards breeding for improved photosynthesis traits using the *Arabidopsis thaliana* model

Roel Franciscus Herman Maria van Bezouw

Thesis committee

Promotors
Prof. Dr Mark G.M. Aarts
Personal Chair at the Laboratory of Genetics
Wageningen University & Research

Dr Joost J.B. Keurentjes
Associate professor at the Laboratory of Genetics
Wageningen University & Research

Co-promotor
Dr Jeremy Harbinson
Associate professor at the Laboratory of Biophysics
Wageningen University & Research

Other members
Prof. Dr Fred A. van Eeuwijk, Wageningen University & Research), the Netherlands
Prof. Dr Tracy Lawson, University of Essex, United Kingdom
Dr Agnieszka Doroszuk – van der Wurff, Rijk Zwaan Breeding B.V., Fijnaart, the Netherlands
Dr Ute Armbruster, Max Planck Institute of Molecular Plant Physiology, Golm, Germany

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences (EPS)

Towards breeding for improved photosynthesis traits using the *Arabidopsis thaliana* model

Roel Franciscus Herman Maria van Bezouw

Thesis
submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 21 June 2022
at 4 p.m. in the Omnia Auditorium.

Roel F.H.M. van Bezouw

Towards breeding for improved photosynthesis traits using the *Arabidopsis thaliana* model, 247 pages.

PhD thesis, Wageningen University, the Netherlands (2022)
With references, with summary in English, Chinese & Ukrainian

DOI: 10.18174/571274

ISBN: 978-94-6447-264-6

Table of contents

Chapter 1. General introduction

Optimal photosynthesis; an evolutionary paradox?

Photosynthesis is a complex process, yet one of the most important on earth. The production of oxygen by primitive photoautotrophic organisms enabled the development of terrestrial aerobic life forms, including humans. In turn, these organisms themselves serve as a food source and form the basis of all webs of life. In plants, photosynthesis developed when cyanobacteria engaged in a complex form of endosymbiosis, were stripped from most of their functional genome to maximize their efficiency and eventually forming chloroplast organelles. In these organelles, complex proteins harvest light energy and use this to free electrons from water (in the light-dependent reactions). These are subsequently used as a resource to assemble carbohydrates from carbon dioxide (in the light-independent reactions) (**Figure 1**). The produced carbohydrates eventually form the quintessential building blocks of life itself.

Photosynthesis is in development for over approximately 3.2 billion years (Blankenship, 2010), of which the latest 430 million years in terrestrial plants. Thus, the process of photosynthesis should have had more than enough time to be optimized in the course of evolution as a result of natural selection. This assumption, together with the practical difficulties to robustly assess photosynthesis in plants (Murchie *et al*., 2018), led to this process being ignored for purposes of selection and improvement in plant breeding schemes (Flood *et al.*, 2011). However, with the advance of better phenotyping tools, the sharp increase in knowledge on the physiology and kinetic balances that are involved in photosynthesis has resulted in numerous studies that have highlighted inefficiencies in photosynthesis (Farquhar *et al.*, 1979; Pettersson & Ryde-Pettersson, 1988; Laisk *et al.*, 1989; Poolman *et al.*, 2000, Zhu *et al.*, 2007; Kubis & Bar-Even 2019). Most notably, modelling the quantity of incoming irradiance that is processed into photochemistry resulted in estimates of as little as 3.2% of light irradiance to be useful for plant productivity under optimal conditions (Long *et al.*, 2006, Zhu *et al.*, 2010) – a figure which is approximately three times below the theoretical maximum (Zhu *et al.*, 2010). This indicates that photosynthesis did not reach optimal functional efficiency.

Optimal photosynthesis in the context of species fitness and survival in natural environments, which are often hostial and limited for resources, may constitute a different meaning compared to optimal photosynthesis for purposes of production in the context of agricultural controlled environments. Capturing the efficiency of photosynthesis traits in measurable parameters will thus not always provide an obvious picture on why perceived suboptimal components in photosynthesis may exist. For example, ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) is the enzyme which assimilates CO_2 in the Calvin-Benson cycle to produce carbohydrates for plant production . However, there is a high probability that RuBisCO fails to capture a molecule of CO_2 and assimilates a molecule of O_2 instead, a process which is called photorespiration. This results in a net loss of energy in the Calvin cycle with no sugars produced (**Figure 1**). This property of RuBisCO has previously been characterized as “imperfect” (Erb & Zarzycki, 2018) and “inefficient” (Andersson, 2008), which led to scientists dreaming of perfecting the enzyme through genetic engineering (Yokota *et al*., 1999; Carmo *et al*., 2015). However, the photorespiratory pathway is found to contribute to abiotic stress tolerance by removing critical intermediates in the Calvin cycle, and may thus consequently contribute to plant fitness in natural environments (Wingler *et al.*, 2000;

Voss *et al.*, 2013; Timm *et al.*, 2019). Thus, photosynthesis must be considered in the context of particular growth environments before embarking on optimization of this process.

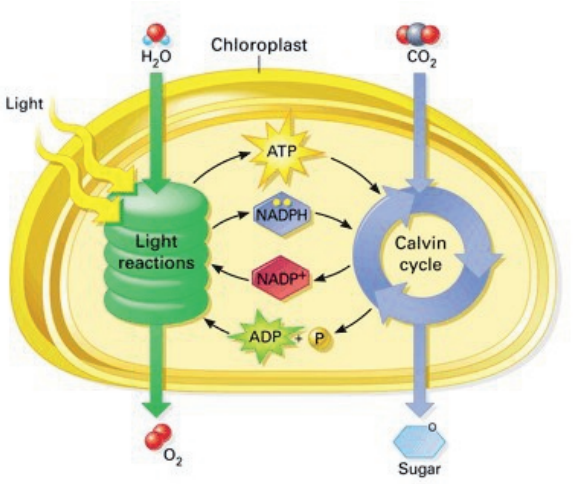


Figure 1. *A simplified overview of photosynthesis in the chloroplast; the light dependent reactions and Calvin Benson cycle (light independent reactions). In the light reactions, complex molecular structures in the form of light harvesting antennae (LHA), photoreaction centres and proton pumps are required to separate electrons from water (H_2O). These electrons in turn are used to charge (reduce) adenosine diphosphate (ADP) and nicotinamide adenine di-nucleotide phosphate ($NADP^+$) energy carrier molecules, which results into their energy depleted (oxidized) counterparts adenosine triphosphate (ATP) and NADPH, respectively. In turn, energy from the ATP and NADPH pools is used to drive the light-independent Calvin Benson cycle. The Calvin Benson cycle fixes carbon dioxide (CO_2), which results in the production of essential carbohydrates to be used as building blocks for all biological processes in plants. ATP and NADPH used and n lose their electron (oxidation), return to ADP and $NADP^+$, and are subsequently recharged via light energy captured by the photo systems in the light-dependent reactions completing a cycle. This cycle is repeated itself for as long as light is harvested in the LHA. Picture source: https://socratic.org/questions/why-is-the-calvin-cycle-referred-fo-as-a-dark-reaction*

Maximizing photosynthesis in plants for sustainable and improved crop productivity.

By significant improvement of agricultural practices in the past decades, large crop yield gains have been realized within a short time span (Kromdijk & Long, 2016), particularly during the "Green Revolution" (Zhu *et al*., 2010; Ray *et al*., 2013). However, as each percentage point of further yield advancement requires significantly more effort to realize than the previous, contemporary annual yield increments per unit land area are currently stalling in all major staple food crops (Ray *et al*., 2013). Key traits benefitting yield (the harvest index) and light use efficiency (the light interception coefficiency of leaves) are already being optimized (Zhu *et al*., 2010, Ort *et al.*, 2015; Kromdijk & Long, 2016; Evans & Lawson, 2020). This leaves improvement of photosynthesis efficiency – expressed as the rate by which carbon dioxide is assimilated into sugars per unit light incoming irradiance – as a major trait to hold significant potential to further staple crop productivity (Zhu *et al*., 2010). Many inefficiencies related to the core genes

photosynthesis have previously been highlighted and materialized as potential targets for transgenic improvement (Long *et al*., 2015; Zhu *et al*., 2020). Examples of these include overexpression of *SEDUHEPTULOSE-BISPHOSPHATASE* (*SBPase*), to accelerate the regeneration of ribulose 1,5-bisphotase – a rate-limiting step in the Calvin-Benson cycle – and improvement of carbon assimilation rates in wheat (Driever *et al.*, 2017); the introduction of highly efficient microbial photo-respiratory pathways in tobacco (South *et al.*, 2020); and increased production of RuBisCO in rice (Yoon *et al.*, 2020). In each of these examples, (grain) yield in plants receiving the transgenes increased 20 to 40% compared to wild types in greenhouse or field conditions – depending on the species or trait under investigation. These examples confirm that optimization of photosynthesis of crops under agriculture conditions is possible, and that there is ample room to boost crop productivity that way.

However, despite an increasing number of valuable examples being published in the past five years, genetic engineering of common photosynthesis pathways may not always translate into the same enhancement of productivity across species. Kromdijk *et al*. (2016) were able to increase biomass productivity in tobacco plants by up to 20% due to improving acceleration and deceleration speed of photo-protective mechanisms, so that plants are better able to optimize photosynthesis to fluctuating light conditions. However, Garcia-Molina & Leister (2020) demonstrated that improving the same photo-protective properties in *Arabidopsis thaliana*, unexpectedly, reduces plant growth by around 20% under growth conditions of fluctuating light conditions similar as those used by Kromdijk *et al*., (2016), thus achieving opposite effects. In addition, overexpression of *SBPase* in rice does not improve plant growth under non-stressed conditions (Feng *et al.*, 2007a,b), contrasting results obtained by Driever *et al.* (2017) in wheat. The pinnacle of an example of tempered ambitions in genetic engineering of photosynthesis for a better crop yield is the C4-Rice project – a major project involving the transplantation of the C4-photosynthetic pathway into rice to overcome rate-limitations associated with the native C3-pathway (Mitchell & Sheehy, 2006, Ermakova *et al*., 2020, 2021). The ambition set out in the project was to introduce a C4-rice cultivar to the market in 2030. However, aiming to transfer the whole pathway proved to be challenging, and eventually led to the initial goals being compromised (Araus *et al*., 2021).

Thus, although transgenic optimized "super-photosynthesizing crops" may play a role in future agriculture, their intended application for some crops may still be problematic. However, and regardless whether the experiments revolving around modifying plant photosynthesis proved to stimulate biomass accumulation, they established that there is still room to optimize photosynthesis for agricultural purposes. As with nearly all traits currently under optimization selection in plant breeding programmes, existing natural variation present a powerful source to further optimize photosynthesis efficiency (Flood *et al*., 2011; Zhu *et al*., 2020). The potential of studying genetic variation for photosynthesis, although previously deemed irrelevant (Long *et al*., 2015; Ort *et al*., 2015), now increasingly receives attention as a resource to consider for further improving the genetic yield potential in crops (Lawson *et al*., 2012; van Bezouw *et al*., 2019; Araus *et al.*, 2021), even among initial sceptics (Zhu *et al*., 2020). In the next section, I will discuss how genetic variation can be exploited for improving traits – including those associated with photosynthesis.

Utilizing natural variation to study complex traits.

Genetic mapping stands at the core of modern genetics (Bazakos *et al*., 2017), a method which is used to study natural variation for traits in many organisms. The ever-decreasing cost of genetic analyses and whole genome sequencing (WGS) technology has popularized this methodology in plant science and breeding. The genetic mapping of quantitative trait loci (QTLs) is now frequently used to explore alleles linked to trait performance in plant breeding and genomic variations that result in our understanding of the evolution of traits. The use of molecular markers and sequencing has resulted in various applications in the plant breeding industry to improve selection efficiency. By adopting such technology, plant breeders can now select for desired genetic material in early growth stages and unwanted traits can be easily selected against without the need for confirmation through direct, expensive phenotyping.

The success of genetic mapping is dependent on the number of genes and their interactions, and causal allelic variation therein. Together, these factors constitute the genetic architecture of a trait in populations of organisms (McKay, 2001; Hansen, 2006). The genetic complexity of a trait is then determined by the quantity of genes involved in the process under study, with an emphasis on those containing genetic variation. Examples of traits controlled by low numbers of genes (and thus not so complex) may include some pigment proteins (e.g., visible as petal colours) and race-specific disease resistances mediated by specialized resistance genes. Such traits may be controlled by one or few genes and are thus easily selected for (**Figure 2**). By contrast, traits such as tolerance to abiotic stresses of plants (e.g. salinity), fruit yield and others related to plant development are highly quantitative in nature and often controlled by genetic variation in numerous – often interacting – genes and pathways. Genes encoding proteins that affect photosynthesis are likely to be the most numerous (**Figure 2**), as more than 10% of the nuclear genome encodes chloroplast located proteins alone (Fristedt *et al*., 2017).

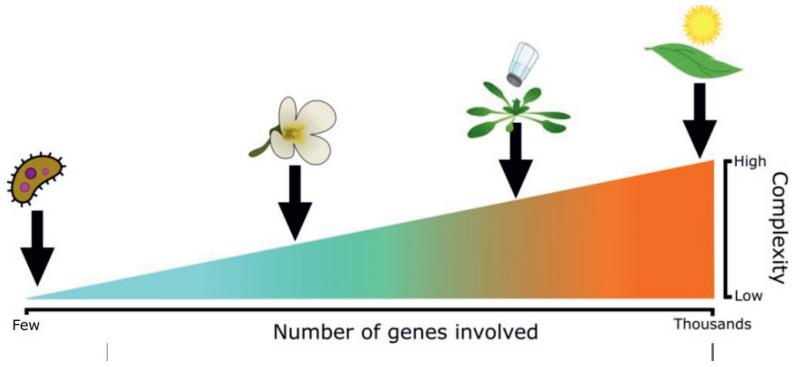


Figure 2. *A scale of genetic complexity for some plants traits. While hundreds of disease resistance genes are known, often only very few resistance genes are necessary to reliably cope with each disease strain (Hammond-Kosack & Jones, 1997). Many disease resistances are therefore considered to be traits that are explained by simple genetics. In Arabidopsis, 306 genes are known to affect flowering time in Arabidopsis thaliana (Bouché et al., 2016). Flowering time can be regarded as a moderately complex trait. In salt tolerance also several hundreds of genes are known to contribute, but the trait is poorly understood (Munns & Tester, 2008) and the total figure of involved genes will likely end up being higher. The most important process in plants is photosynthesis, which regulated by thousands of genes that encode for*

proteins locating to the chloroplasts alone (Fristedt et al., 2017), and possibly a large number of genes that co-regulate downstream further processes. The more genes are involved in a trait, the higher the genetic complexity will be.

High trait complexity undermines the ability to detect genetic loci underlying trait variation in genetic mapping studies. This is particularly the case if genetic variation is not related to genomic variation (e.g. in epigenetics) and if interactions between genetic loci (epistasis) are highly prevalent. A compounding difficulty in the detection of relevant genotypic variation for photosynthesis is the presence of genetic variation in the chloroplast genome (El-Lithy *et al.*, 2005; Flood *et al.*, 2016a; Flood *et al.*, 2020), which do not inherit in a Mendelian fashion. Depending on the genetic layout of mapping populations, these factors may contribute to some extend to the masking of genetic signals leading to relevant genetic variation for photosynthesis.

What the study of diversity panels can and cannot accomplish to advance our understanding of highly complex traits like photosynthesis.

Many genetic mapping population types require multiple generations of inbreeding or back-crossing before a sufficient level of homozygosity is reached for accurate mapping (Bazakos *et al.*, 2017). Due to the availability of genetic diversity panels and increasingly accessible genomic sequence data, genome wide association studies (GWAS) have become the plant geneticist' workhorse to study complex traits (Atwell *et al*., 2010; Bazakos et al., 2017). Collections of accessions are often readily available in seed centres and genetic variation represent generations of natural recombination events, exchange of genetic material and random mutations for millions of years (Atwell *et al.*, 2010; Korte & Farlow, 2013; Durvusalu *et al*., 2017). These properties allow genetic mapping at unparalleled genetic mapping accuracy (Kim *et al.*, 2007), while in cultivated species GWAS are generally less accurate and diverse due to genetic lineages having diversified more recently (but see maize, Tian *et al*., 2011; Yann *et al*., 2013). Phenotyping and genotyping at high density in diversity panels followed by GWAS allows to accurately pinpoint the genetic components for traits up to the level of causal polymorphisms in densely genotyped material. Exploiting diversity panels to identify genetic components contributing to traits in this way is less time and resource consuming than inbreeding, backcrossing and fine-mapping of QTLs using mapping populations from traditional crosses (Bazakos *et al*., 2017). These factors led to GWAS surpassing the use of traditional mapping populations in plant science.

However, the diversity present in such panels comes with several serious drawbacks that impair their use to detect novel traits and genetic variation to improve complex traits. These drawbacks include population genetic relationships related to genetic architecture (Korte & Farlow, 2013), epistasis introducing false additive associations and vice versa (Zan *et al.* 2018), the presence of multi-alleles at a single locus (e.g. Forsberg *et al*., 2015; Tang *et al*., 2018; Hazzouri *et al*., 2018, but see van Bezouw *et al*., 2019b), and chloroplastic variation being present while unaccounted for in nuclear genomic data (Flood *et*

al., 2016b, 2020). Lastly, due to the requirement of setting a minor allele frequency threshold to account for false positives, rare variants that are of interest in breeding may remain undetectable (Korte & Farlow, 2013; Bernardo 2016). Thus, while exploration of diversity panels is a highly useful method to obtain an impression of available genetic or phenotypic diversity for a trait within a plant species, the identification of elite alleles may be challenging. GWAS thus only have limited applications for use in crop species to identify novel genes and variation of use in plant breeding programmes – a view that is more and more acknowledged among plant breeders and geneticists (Bernardo, 2016; Liang *et al*., 2021).

The advantages of reducing genetic complexity; usingbiparental mapping populationsfor in-depthexploration of natural variation in photosynthesis traits.

Success in studying and discovering new and relevant traits in photosynthesis ultimately leans on three parameters: i) the genetic complexity of photosynthesis, ii) the ability to phenotype the trait of interest; and iii) the genetic complexity of the population at use. The number of genes and their complexity are factors that are difficult to control, i.e. it is impossible to reduce the number of genes that constitute photosynthesis in a given genotype. Phenotyping of photosynthesis traits is similarly unlikely to be made easier, as there are many different traits that cannot be ignored (Murchie *et al*., 2018). However, being able to more specifically define and phenotype certain traits may reduce the number of genes to be relevant to affect a biological process of interest, and thus improve genetic mapping accuracy (Crowell *et al*., 2016; van Bezouw *et al*., 2019). The factor that reduces genetic complexity most is the choice of the genetic mapping population itself. Mapping populations derived from only two – or few – parental lines have a limited genetic complexity compared to a diversity panel with long separate lineages. The result is that genetic variants – including elite ones – are always present at a rate of 50% in a segregating population and many of the previously discussed disadvantages of diversity panels also do not apply or impact results at a lesser degree in bi-parental mapping populations compared to diversity panels.

Bi-parental mapping populations have a high detection power for quantitative trait loci (Keurentjes *et al.*, 2007, Kooke *et al.*, 2012, Bazakos *et al.*, 2017). This makes such populations ideal to map traits with that are likely contributed by genetic loci that are small in effect size such as those related to photosynthesis. The reduction of genetic complexity further increases the likelihood of detecting a genetic factor affecting a trait, as often only few genetic loci will significantly impact traits between two different genotypes at a time. Despite these apparent advantages, only few reports describe the use of bi-parental mapping populations to study photosynthesis traits (Jung & Niyogi, 2009; Oakley *et al*., 2018; de Oliveira Silva *et al*., 2018; Honda *et al*., 2021). Oliveira Silva *et al.*, (2018) and Honda *et al*., (2021) found natural differences in carbon assimilation rates in rice and tomato to be as big as those reported for transgenic alteration of key photosynthesis genes in tomato (Nunes-Nesi *et al*., 2005), tobacco (Simkin *et al*., 2015) and wheat (Driever *et al*., 2017). The findings in these genetic studies thus establish the power of natural variation to improve photosynthesis traits and as an alternative to transgenic approaches. Unfortunately, in none of these studies genetic variation beyond the establishment and discovery of the QTLs is further analyzed, and thus the underlying causal genes unfortunately remain unknown.

A good model plant species to study natural variation is *Arabidopsis thaliana* (Koornneef *et al*., 2004), which has a fast breeding cycle and is a good model species to explore novel genes and traits that may be of interest in the context of plant breeding. In addition, many genetic mapping populations from contrasting accessions have been developed. These include populations F2, introgression lines, doubled haploids and recombinant inbred line populations, but also more advanced Arabidopsis Multiple Parental Recombinant Inbred Line (AMPRIL) and Multiparent Advanced Generation Inter-Cross (MAGIC)-populations (Wijnen & Keurentjes, 2014; Bazakos *et al*., 2017). All these come with specific advantages and disadvantages with regards to development, genetic variation and mapping accuracy. However, the high mapping power in all such populations may be highly advantageous in photosynthesis research.

While biparental crosses A novel biparental mapping population type is that composed of Chromosome Substitution Lines (CSLs). CSL-populations are a special type of segregating mapping population due to the absence of crossovers events in their chromosomes. This a property that opens new angles and opportunities for the study of complex traits due to the highly simplified chromosomal architecture (Wijnen, 2019). However, CSLs have so far seen little reporting in plant genetic studies. In the next section, I will further elaborate on the development, potential advantages and use of populations consisting of CSLs in quantitative genetics.

Chromosome substitution line populations as a novel bi-parental mapping population type.

The use of chromosome substitution lines (CSLs) is an established method to perform genetic mapping in animal model species to dissect quantitative traits at high efficiency with high mapping power (Nadeau *et al*., 2000; Belknap, 2003; Buchner & Nadeau, 2015). This is particularly true for the analysis epistatic interactions (Spiezio *et al.*, 2012), which are particularly hard to detect in conventionally segregating populations due to segregation distortion and low statistical power (Carlbog & Haley, 2004; Mackay 2014; Lachowiec *et al*., 2015). The development and use of single chromosome substitution populations in *A. thaliana* dates back as early as 2002 (Koumproglou *et al*. 2002), although the first published report dates back from 1967 (Laurie *et al*.). For inbreeding plant species, like *A. thaliana*, a wide variety of genetic mapping population types are already available that are more accurate and generally have a more cost-efficient developmental trajectory albeit at the cost of subjecting larger populations to be phenotyped in experiments (Keurentjes *et al*., 2007; Kooke *et al*., 2012; Wijnen & Keurentjes, 2014; Bazakos *et al.*, 2017). The use of populations of CSLs in plants thus comes with competing opportunity costs set off against segregating populations, particularly related to their development. Thus far, the opinion of plant geneticists is not positive towards potential advantages that CSL-populations offer over other population types that are generally easier to generate, as can be concluded from the small number applications reported in genetic studies in crops (see Unrau *et al*. 1956, Wu *et al.*, 2010, De León *et al*., 2011, Saha *et al.*, 2014).

The efficiency of substituting whole chromosomes in background genotypes is greatly enhanced by recently developed transgenic approaches that suppress meiotic crossovers in plants using RNAi silencing technology (Wijnker *et al.*, 2012, 2014; Calvo Baltanas *et al.*, 2020). Such technology allows a

more efficient development of genetic mapping populations consisting of lines each with multiple substitutions (Wijnker *et al*., 2012; Wijnen, 2019). A CSL-population completed to the maximum number of possible lines in this way offers several potential advantages in genetic mapping studies. These includes the genetic mapping of heterosis and enhanced detection power epistatic interactions that are harder to study in naturally segregating populations (Wijnen & Keurentjes, 2014; Wijnen, 2019). A preliminary trait analysis in a CSL-population developed from Col-0 and Ler-0 showed that phenotypic variation in CSLs appeared to be higher (Wijnen, 2019). In this analysis, more additive and epistatic genetic effects were detected compared to a conventionally genetic mapping population composed from the same parents despite the reduced genetic diversity resulting from a lack of crossovers (Wijnen, 2019). This suggests that, also in plants, CSLs may offer increased genetic mapping power over naturally segregating alternatives (Spiezio *et al*., 2012; Buchner & Nadeau, 2015). Next to potential opportunities, there are also disadvantages associated with the use of CSLs. The development of complete and balanced populations in species with larger chromosome numbers may be particularly laborious as the total number of possible combinations doubles with each additional chromosome in the haploid chromosome count (Wijnker *et al.*, 2012). Genetic mapping in CSLs will also result in QTLs that are mapped at the chromosome level, rather than specific genomic regions. This is, however, not a disadvantage if trait variation can be discovered this way that would otherwise be invisible, for example if trait variation is explained by epistatic interactions. The motivation for developing and using populations of CSLs in quantitative trait analysis will eventually depend on the value of being able to use small populations in experiments, the willingness to invest in the development of such a population, the total number of chromosomes in the target species and the interest in the analysis of (higher order) inter-chromosomal epistatic interactions. The use and advantages of populations of CSLs are in need of tested and proven examples before they are unanimously adopted into the plant geneticist's toolbox.

Towards implementation of natural variation in photosynthesis in plant breeding programmes

Chlorophyll fluorescence imaging greatly reduced the time and resource cost of mass phenotyping for photosynthesis traits (Baker, 2008; Murchie & Lawson, 2013; Rungrat *et al*., 2016), which can potentially support their deciphering and use in breeding programs. Due to the ease of screening genetic variation in diversity panels, the existence of variation that can be used to improve photosynthesis traits can now be considered as an established fact. This idea is supported by the large number of published studies showing large trait variation in relevant crop and model species (reviewed in van Bezouw *et al*., 2019a). Studying natural variation in photosynthesis for purposes of improving it imperatively demands a basis for a subsequent improvement of productivity, abiotic stress tolerance and/or quality in plants. However, to do so will require more knowledge on the causal genes controlling photosynthesis QTLs. In many genetic studies, those are often neither resolved up to a sufficient level to determine their causality in photosynthesis and/or empirically shown to eventually enhance productivity. Only few only few such

examples existing (Jung & Niyogi, 2009; Athanasiou *et al.*, 2010; Wang *et al*., 2016; van Rooijen *et al*., 2017; Honda *et al*., 2021).

My main goal is to exploit the low genetic complexity and high genetic mapping power of biparental mapping population to study novel traits and genes that may be used to improve photosynthesis and subsequent downstream processes. To do so, I will use three different biparental mapping population types (**Figure 3**), each differing in development time and genomic complexity, and grow these under three different light environments. RIL and F2 populations are standard genetic mapping population types that have been highly successful in studying quantitative traits (Keurentjes *et al*., 2011; Bazakos *et al*., 2017; Jiang *et al*., 2021), while genetic mapping in populations consisting of CSLs have thus far received little attention – particularly in those that contain multiple chromosome substitutions. To further explore the possibilities, opportunities and strategies that CSL-populations have to offer, I will explore their use in more depth using a modelling approach to help familiarize the concept and explore their potential.

In all experimental work in this thesis the tried and tested model species *A. thaliana* is used. This is an ideal species for the purposes set out in this chapter as large numbers of plants can efficiently be phenotyped for photosynthesis in high-throughput phenotyping facilities (Flood *et al*., 2016); extensive natural variation exists within the species (Atwell *et al*., 2010; Durvusalu *et al*., 2017); biparental mapping population of many types are available (Wijnen & Keurentjes, 2014); and it is the first species in which populations of chromosome substitution lines exist with multiple substitutions per line (Wijnker *et al*., 2012; Wijnen 2019). However, insights obtained from this research will likely be applicable to commercially relevant species as well.

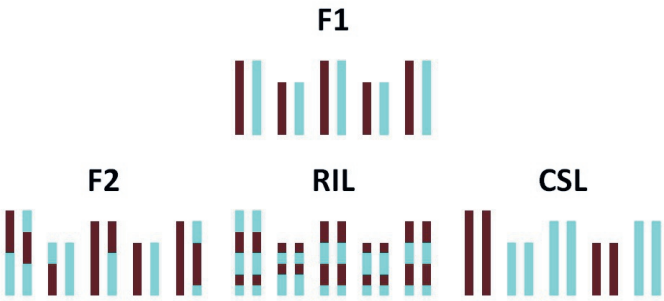


Figure 3. *The genetic layout of different types of biparental mapping populations in Arabidopsis thaliana. From an F1 hybrid, different types of mapping populations can be developed upon selfing. An F2 population is the most basic population and is directly derived from the progeny of an F1 hybrid (used in* ***Chapter 4****). Six to eight rounds of selfing F2's result in recombinant inbred lines (RILs; used in* ***Chapter 3****), which have an increased recombination frequency compared to an F2 population offering high mapping accuracy. Following transgenic procedures, populations of CSLs (chromosome substitution lines, used in* ***Chapter 5****) can be developed in a few generations. Such populations are developed with suppression of recombination events and are characterized by a absence of cross-over events.*

Thesis layout

This thesis starts off with a review, in **Chapter 2**, which provides the latest insight in nearly a decade of development, research, and utilization of genetic variation in photosynthesis since the latest review on this topic by Flood *et al.* (2011). I summarize developments contributing to our understanding of the phenomic and genomic complexity of photosynthesis, with a focus on innovation in high-throughput phenotyping methodology and I identify areas of study that have still not been covered sufficiently.

Chapter 3 describes the use of the more than thirty-years-old recombinant inbred line (RIL) population developed from the tried and tested lab accessions Ler-0 and Col-0 to study photosynthesis under various stable light environments. Natural variation in copy number variation of *SQUALENE EPOXDIASE-like* genes is found to causally impact photosynthesis. The outcomes provide novel insights in the evolution of photosynthesis and highlight the value of biparental mapping to discover new genes contributing to complex traits.

The stable conditions in climate-controlled growth chambers typically do not represent those encountered by plants in field conditions, where shifts in light intensity occur (van Rooijen *et al*., 2015). In **Chapter 4**, the genetic components underlying phenotypic variation between two *A. thaliana* accessions differing in their high light acclimation response are mapped in an F2 population. The QTLs in this study proved to have a high impact on the operating efficiency of photosynthesis. I tested their potential effects in impacting biomass accumulation under appropriate light conditions and set up a biomass essay for genotype x light environment to measure the effect of these QTLs on growth.

However, a single, a sudden shift does not represent natural daily light fluctuations, which range in the order of milliseconds to hours (Kaiser *et al.*, 2018). In **Chapter 5**, I aim to resolve the genetic architecture of losses in productivity and changes in plant morphology under fluctuating light conditions compared to stable light conditions. To do so, I complete and use a population of chromosome substitution lines (CSLs), first developed by former colleague Ramon Botet Vaca to assess whether it is possible to dissect genetic components related to productivity under influence of fluctuating light.

The development and use of CSL-populations include several challenges and strategic considerations. In **Chapter 6**, quantitative aspects of the development, and mapping power of additive and epistatic genetic effects of such populations are explored. Instead of performing experiments, permutation models are used to explore a variety of scenarios to support decision-making in the use of CSLs.

The significance of these findings, and a final synthesis, are discussed and placed in the context of the ongoing research of improving photosynthesis and the genetic mapping of complex traits, with an emphasis on chromosome substitution libraries in the general discussion (**Chapter 7)**.

Chapter 2. Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency

Roel F.H.M. van Bezouw[1], Joost J.B. Keurentjes[1], Jeremy Harbinson[2], Mark G.M. Aarts[1]

SUMMARY

In recent years developments in plant phenomic approaches and facilities have gradually caught up with genomic approaches. An opportunity lies ahead to dissect complex, quantitative traits when both genotype and phenotype can be assessed at a high level of detail. This is especially true for the study of natural variation in photosynthetic efficiency, for which forward genetics studies have yielded only a little progress in our understanding of the genetic layout of the trait. High-throughput phenotyping, primarily from chlorophyll fluorescence imaging, should help to dissect the genetics of photosynthesis at the different levels of both plant physiology and development. Specific emphasis should be directed towards understanding the acclimation of the photo-synthetic machinery in fluctuating environments, which may be crucial for the identification of genetic variation for relevant traits in food crops. Facilities should preferably be designed to accommodate phenotyping of photo-synthesis-related traits in such environments. The use of forward genetics to study the genetic architecture of photosynthesis is likely to lead to the discovery of novel traits and/or genes that may be targeted in breeding or bio-engineering approaches to improve crop photosynthetic efficiency. In the near future, big data approaches will play a pivotal role in data processing and streamlining the phenotype-to-gene identification pipeline.

---

[1]Laboratory of Genetics, Wageningen University, [2]Laboratory of Biophysics, Wageningen University.

This work was published in Plant Journal issue 97(1):112-133, 2019: https://doi.org/10.1111/tpj.14190

Introduction

In the past two decades annual increases in yield of major staple food crops such as rice, wheat, soybean and maize have stagnated globally despite agronomic and genetic improvements (Wei *et al*., 2015). It is expected that by the middle of this century 60–110% greater yield output is required from crops for food, feed and bio-fuel in order to keep up with the increasing demand of the growing human population worldwide (Kromdijk and Long, 2016; Tilman *et al*., 2011). To keep pace with this development a new Green Revolution will be required to double the realised average annual yield increment from 2000 until 2018 of approximately 1.2%, up to a minimum of 2.4% per year from 2018 until 2050 (Ray *et al*., 2013). The forecast negative effects of global warming on the yield of staple crops will further complicate the achievement of this goal, given that the negative effects largely offset the potential gains from increased atmospheric carbon dioxide (Schauberger *et al*., 2017; Zhao *et al*., 2017). The relationship between irradiance and crop yield is summarised by the yield model of Monteith (Long *et al*., 2006) in which the harvestable yield component is the product of irradiance, the interception of irradiance, the efficiency of conversion of intercepted irradiance into biomass (ε_c) and the harvest index. In the past, crop breeding programmes aimed to increase the harvest index and the light capture efficiency of the canopy, and did so successfully. As a result these traits are now reaching their theoretical potential, making it unlikely that there can be further significant improvement (Long *et al*., 2006; Zhu *et al*., 2010). The remaining component of yield, ε_c, is strongly influenced by the efficiency of photosynthesis, and ε_c is still far below its the theoretical maximum (Zhu *et al*., 2010). This means that there is still room for improving this trait by crop breeding. A modest 50% improvement in ε_c – which would still leave it far below its theoretical maximum – would already be enough to bridge the gap between human demand and production of plant resources. Given its importance to ε_c, photosynthesis is considered as a primary target for improvement in crop species (Long *et al*., 2006; Ort *et al*., 2015; Long *et al*., 2015).

The physiological basis of photosynthesis, and the genes coding for the proteome of photosynthesis, have been relatively well characterised during the last decades (e.g. Farquhar *et al*., 1980). In the past 15 years, attempts to improve plant photosynthetic efficiency have focused on identifying the efficiency bottlenecks in photosynthesis and then bio-engineering photosynthetic pathways to overcome them (Long *et al*., 2015). This approach has led to suggestions, amongst others, to modify specific subunits of photosystem II (PSII) to optimise light-harvesting capacity (Long *et al*., 2006; Ort *et al*., 2011; Walker *et al*., 2018), improve the catalytic properties of Rubisco (Carmo-Silva *et al*., 2014) and transplant the C4 photosynthesis machinery into C3 plants (Crovshoff and Hibberd, 2012; Schüler *et al*., 2016). Application of this strategy has been successful. For example, increasing the speed with which the photosynthetic machinery in tobacco adapts to fluctuating light resulted in greatly improved average photosynthetic efficiency throughout the photoperiod (Kromdijk *et al*., 2016). Improving the rate of ribulose-1,5-bisphosphate (RuBP) regeneration in the Calvin cycle in wheat by over-expressing sedoheptulose-1,7-bisphosphatase (SBPase) resulted in increased rates of CO_2 fixation (Driever *et al*., 2017). These improvements in photosynthesis translate to increases in plant biomass: in the case of tobacco, with improved responses to fluctuating light (Kromdijk *et al*., 2016), plant biomass was increased by 14–20% under field conditions, while for wheat, with improved RuBP regeneration, a 30–40% higher grain yield was obtained under greenhouse conditions (Driever *et al*., 2017). From these examples it should be clear that improving photosynthetic efficiency is both a promising and feasible route for increasing crop yields.

Natural variation is an underexploited genetic resource with which to improve plant photosynthesis and the genetic yield potential of crops (Flood *et al*., 2011; Lawson *et al*., 2012). The failure to better use natural variation in photosynthesis as a route to yield improvement stems partly from the misconception that photosynthesis has been fully optimised by millions of years of selection for this vital trait (Flood *et al*., 2011). Clearly this assumption of perfection is misguided – if nothing else, the improvements in photosynthesis outlined above show this. In addition, highly heritable variation in photosynthesis-related traits has been found for model and crop species such as wheat (Driever *et al*., 2014), Arabidopsis (*Arabidopsis thaliana*), soybean (*Glycine max*), sorghum (*Sorghum bicolor*), maize

(*Zea mays*) and rice (*Oryza sativa*) (see **Table 1**). The identification of functional allelic variations of important genes and traits related to photosynthesis using a forward genetics analysis of diversity panels is therefore a valuable approach, for three main reasons. First, the study of natural variation offers key insights into regulatory processes of photosynthesis, for example under fluctuating light conditions (Lawson *et al*., 2012; Murchie *et al*., 2018). Furthermore, genetic variation in the core genes of photosynthesis has been acknowledged to be an important resource for the bio-engineering of improved photosynthesis (Prins *et al*., 2016; Reeves *et al*., 2018). Lastly, when beneficial alleles are identified they can be exploited in current crop breeding programmes. The exploitation of natural variation in photosynthesis is likely to be particularly active in jurisdictions where the use of genetically modified crops is prohibited, or where the resources to generate genetically modified crops are not available. Note, however, that genes identified as being important in determining photosynthetic traits based on studies of natural variation of these traits could be valuable not only in conventional plant breeding approaches to improving photosynthesis but also in genetic modification approaches.

Despite the potential for increasing photosynthetic efficiency, major bottlenecks in the phenotypic and genotypic evaluation of photosynthesis-related traits exist (Flood *et al*., 2011; Murchie *et al*., 2018). In the last 10 years, however, both plant genomics and phenomics have matured to the point where, together, they can be used in forward genetics analyses of photosynthetic traits. In the case of photosynthesis, chlorophyll fluorescence techniques allow the measurement of a range of photosynthetic traits (Baker, 2008; Maxwell and Johnson, 2000; Murchie and Lawson, 2013) and fluorescence imaging techniques allow these traits to be measured quickly for large numbers of plants, enabling the high-throughput phenotyping of photosynthesis (**Box 1**, **Figure 1**). At the same time, whole-genome sequencing has become cheaper (Barabaschi *et al*., 2016; Jiao and Schneeberger, 2017), leading to the development of high-resolution mapping populations in the form of genotypically detailed diversity panels suitable for genome-wide association studies (GWAS) (**Box 2**, **Figure 2**). This convergence of phenomics and genomics is expected to lead to a better understanding of the genetics and molecular mechanisms of complex traits such as photosynthesis. Nonetheless, large-scale and detailed phenotyping of photosynthetic traits is relatively new and may be conceptually challenging due to the multidimensional nature of the traits. The possibilities for high-throughput phenotyping of many plant traits have been reviewed in detail in other works (e.g. Awlia *et al*., 2016; Rungrat *et al*., 2016; Großkinsky *et al*., 2017). If measurement of photosynthesis-related traits is discussed it is usually limited to the context of abiotic stress responses (Rungrat *et al*., 2016, Großkinsky *et al*., 2017). We believe that the time has come to examine the plethora of genomic and phenomic tools that have been developed and how these can be applied to successfully screen for natural variation in photosynthesis beyond its relatively simple use as a measure of plant stress.

This review aims to bring together developments in genomics and phenomics from the past 10 years, and evaluate their progress within the context of exploring natural variation in photosynthesis. We first aim to provide an overview of how insights into photosynthesis and genetic mapping studies have contributed to a deeper understanding of the genetic architecture of photosynthesis. As such, we have reviewed studies that aimed to unravel the genetics of photosynthesis by using phenotyping methods that are directly linked to photosynthesis-related traits and omit studies that assessed indirect traits, such leaf architecture. Next, phenotyping of photosynthesis-related traits will be critically reviewed in the context of mapping trait variation. In addition, we will argue for the need for phenomics to overcome the challenges that lie ahead of successful mapping. Finally, high-throughput phenotyping of photosynthesis in relation to statistics, data handling and gene identification will be discussed.

Table 1. *An overview of genome-wide association analyses found in the literature that target photosynthesis-related traits. Nature of data: brief description of the methodology of obtaining the data. Stress: Stresses subjected to plants as given in the corresponding literature. Growth conditions: Growth conditions of the experiment. Rep: describes the number of replicates of the entire experiment to obtain the complete dataset. Traits: the type of traits that were analysed in the study. Study: Roman numerals I, II and III correspond to repeated use of data collection obtained from the same experiment as described in the earliest published papers' methods of each pair.*

Species	Nature of data (photosynthesis-related only)	Stress	Data	Growth conditions	Rep	Traits (photosynthesis-related only)	Study
Soybean	Single measurements of chlorophyll content by extracts and spectral measurements	–	1	Climate chamber	1	Chlorophyll content and chlorophyll reflectance traits	Hao *et al*. (2012)
Maize	Single measurements of data using handheld pulse amplitude modulation (PAM) and soil plant analysis development (SPAD) index devices	Cold	1	Climate chamber and field grown	1	Chlorophyll content (SPAD) and chlorophyll reflectance traits	Strigens *et al*. (2013)
Rice	Single measurements of chlorophyll content by both leaf extracts and spectral measurements in mature plants	–	1	Field grown	2	Chlorophyll content (SPAD and leaf extracts)	Wang *et al*. (2015)b
Arabidopsis	Photosystem II efficiency data taken 1 h after a stepwise increase from low to high light	High light	1	Climate chamberb	1	Photosystem II (PSII) efficiency (PAM derived)	van Rooijen *et al*. (2015)II
Soybean	Single measurement of chlorophyll reflectance index of soybean canopies using multispectral devices	–	1	Field grown	2	Non-photochemical quenching (NPQ) (derived by broadview multispectral devices)	Herritt *et al*. (2016)a
Soybean	Measurements derived from handheld PAM-devices and assessment of chlorophyll content in seedling plants at multiple levels of cold stress	Cold	1	Climate chamber	1	Chlorophyll content (SPAD) and chlorophyll reflectance traits	Fiedler *et al*. (2016)
Soybean	Single measurement of chlorophyll content by both extracts and spectral measurements	–	1	Field-grown	2	Chlorophyll content (multispectral derived and leaf extracts)	Dhanapal *et al*. (2016)a
Rice	Measurements of NPQ in excised leaf discs by handheld PAM devices	–	1	Field grown	2	NPQ (PAM derived)	Wang *et al*. (2017)b
Sorghum	Pre-cold, cold and post-cold measurements of chlorophyll fluorescence traits and gas-exchange measurements	Cold	3a	Climate chamber	1	Chlorophyll reflectance and gas-exchange traits	Ortiz *et al*. (2017)
Arabidopsis	PSII efficiency measurements, under control (2 days) and high light (4 days) conditions	High light	18	Climate chamberb	1	PSII efficiency (PAM derived)	van Rooijen *et al*. (2017)II
Rice	Three measurements of chlorophyll content by SPAD meter in developing seedlings	–	3	Climate chamber	1	Chlorophyll content (SPAD)	Lin *et al*. (2017)

[a] *Before, during and after cold measurements.*
[b] *See Flood et al. (2016b) for a description.*

Despite the potential for increasing photosynthetic efficiency, major bottlenecks in the phenotypic and genotypic evaluation of photosynthesis-related traits exist (Flood *et al*., 2011; Murchie *et al*., 2018). In the last 10 years, however, both plant genomics and phenomics have matured to the point where, together, they can be used in forward genetics analyses of photosynthetic traits. In the case of photosynthesis, chlorophyll fluorescence techniques allow the measurement of a range of photosynthetic traits (Baker, 2008; Maxwell and Johnson, 2000; Murchie and Lawson, 2013) and fluorescence imaging techniques allow these traits to be measured quickly for large numbers of plants, enabling the high-throughput phenotyping of photosynthesis (**Box 1**, **Figure 1**). At the same time, whole-genome sequencing has become cheaper (Barabaschi *et al*., 2016; Jiao and Schneeberger, 2017), leading to the development of high-resolution mapping populations in the form of genotypically detailed diversity panels suitable for genome-wide association studies (GWAS) (**Box 2**, **Figure 2**). This convergence of phenomics and genomics is expected to lead to a better understanding of the genetics and molecular mechanisms of complex traits such as photosynthesis. Nonetheless, large-scale and detailed phenotyping of photosynthetic traits is relatively new and may be conceptually challenging due to the multidimensional nature of the traits. The possibilities for high-throughput phenotyping of many plant traits have been reviewed in detail in other works (e.g. Awlia *et al*., 2016; Rungrat *et al*., 2016; Großkinsky *et al*., 2017). If measurement of photosynthesis-related traits is discussed it is usually limited to the context of abiotic stress responses (Rungrat *et al*., 2016, Großkinsky *et al*., 2017). We believe that the time has come to examine the plethora of genomic and phenomic tools that have been developed and how these can be applied to successfully screen for natural variation in photosynthesis beyond its relatively simple use as a measure of plant stress.

This review aims to bring together developments in genomics and phenomics from the past 10 years, and evaluate their progress within the context of exploring natural variation in photosynthesis. We first aim to provide an overview of how insights into photosynthesis and genetic mapping studies have contributed to a deeper understanding of the genetic architecture of photosynthesis. As such, we have reviewed studies that aimed to unravel the genetics of photosynthesis by using phenotyping methods that are directly linked to photosynthesis-related traits and omit studies that assessed indirect traits, such leaf architecture. Next, phenotyping of photosynthesis-related traits will be critically reviewed in the context of mapping trait variation. In addition, we will argue for the need for phenomics to overcome the challenges that lie ahead of successful mapping. Finally, high-throughput phenotyping of photosynthesis in relation to statistics, data handling and gene identification will be discussed.

BOX 1. Genetic mapping populations for trait discovery

Quantitative trait locus (QTL) mapping is traditionally performed by crossing two distinct genotypes with contrasting phenotypes (parents) to generate an F_1 generation and then subsequently subjecting the F_2 or other next-generation progeny to phenotypic screening. By extensive genotyping of the F_2 generation using genomic markers [often based on single nucleotide polymorphisms (SNPs) or insertion–deletion (InDel) polymorphisms in the genomes of two parental lines], causal loci responsible for the observed differences in phenotype can be identified in QTL analysis, and will thus provide insights into the genetic architecture of the studied trait. Each genotypic marker is assessed by applying analysis of variance (ANOVA), and when the significance reaches a pre-formulated logarithm of the odds (LOD) threshold the genomic region is claimed to be a QTL for a phenotype.

Major disadvantages of mapping in an F_2 population are that there are no replicates for each genotype and that heterozygosity is prevalent in such a population, causing further segregation in subsequent generations (**Figure 1a**) – both of which will increase genotypic variance that is difficult to properly account for in QTL mapping. In addition, the limited number of recombination events per chromosome generated by a single round of self-fertilisation means that genetic linkage between distant regions is strong and large populations will need to be screened to achieve a sufficient mapping resolution to be able to pinpoint genomic regions of interest.

In order to overcome most of the disadvantages described above, so-called 'immortal' inbreeding and backcrossing populations have been developed in plant species that allow inbreeding (Wijnen and Keurentjes, 2014). The most common ones are recombinant inbred line (RIL) and near isogenic line (NIL) populations (Bazakos *et al*., 2017; Keurentjes *et al*., 2007). Recombinant inbred line populations are acquired by selecting several F_2 plants and repeatedly self-pollinating these plants for six to eight generations (**Figure 1b**). This will result in near homozygosity of all alleles and a large number of recombination events which will increase mapping resolution. Near isogenic line populations are generated through repeated backcrossing of F_2 plants to a recurrent parental line (**Figure 1c**). Molecular markers are then used to identify genomic regions of interest which will be selected for until the genetic background is isogenic to that of the parent. Near isogenic line populations effectively allow the study of genomic regions in isolation and individual lines can be easily backcrossed again to the parent of interest to fine map causal loci. The genetic layout of each of these solutions introduces different types of population structure and experimental pre-conditions that should be taken into account to properly identify genes related to traits (Keurentjes *et al*., 2007).

The ever decreasing costs of full genome sequencing (Barabaschi *et al*., 2016) have led to the development of diversity panels suitable for GWAS in many crop and model species (Bazakos *et al*., 2017; Ingvarsson and Street, 2011; Korte and Farlow, 2013). These studies exploit naturally occurring variants resulting from bi-allelic SNPs that have been recombining between genotypes through the course of history of plant species (**Figure 1d**), as identified through whole genome sequencing. In this way, thousands to millions of SNPs can be identified in diversity panels, to full saturation of the genome. The process of continuous recombination out-crossing events typically lasts thousands to millions of years in such panels. Co-segregating small genomic regions (also called linkage or haplotype blocks) in wild species often span less than 1 kilobase pair (kbp). In crop species that have only recently been domesticated, the size is generally less than 100 kbp. Compared with the linkage blocks of a few million base pairs that are typically found in bi-parental mapping populations such as F_2, RILs and NILs, this is a huge improvement. The result is that once a genomic region has been identified through QTL mapping there will be only a few relevant candidate genes. This strongly accelerates gene identification and overcomes the requirement for additional fine mapping experiments, especially for highly polygenic traits such as photosynthesis; this has led to GWAS being widely adopted by forward plant geneticists (Atwell *et al*., 2010; Ingvarsson and Street, 2011; Tian *et al*., 2011).

Potential disadvantages of GWAS, compared with conventional bi-parental mapping studies, generally include the low statistical power of finding QTLs. One of the reasons for this is that many independent marker–trait association tests need to be made – which increases the threshold to control for false positives. Additionally, in an ideal RIL population all alleles are present in a 1:1 ratio, while allelic distributions in GWAS populations may be heavily skewed towards the major allele (**Figure 1d**), further compromising QTL detection power (Korte and Farlow, 2013). Furthermore, the two-dimensional nature of bi-allelic SNPs may be an oversimplification of reality since in natural populations many different functional alleles exist and introduce extra variance into the analysis (Forsberg *et al*., 2015; Ingvarsson and Street, 2011; Korte and Farlow, 2013). On top of these potential disadvantages, epitasis and epigenetics

may further introduce variance that is difficult to account for in statistical models. Nevertheless, GWAS in inbreeding plant species can be considered highly potent since replications can be used, which opens the possibility of repeating the same experiment for confirmation or to account for genotype–environment interactions (Brachi *et al*., 2011).

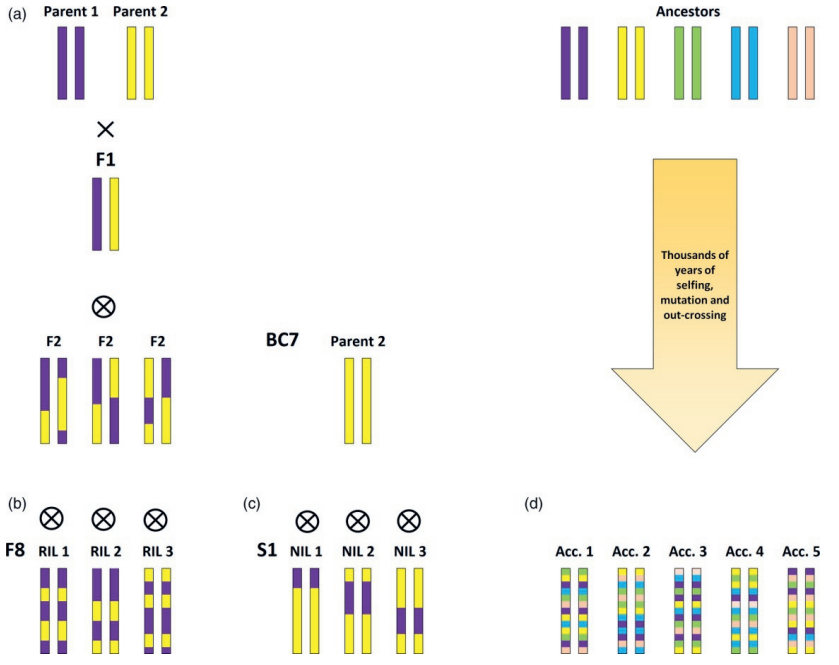


Figure 1. *Schematic overview of the generation of different genetic mapping populations in inbreeding plant species.(a) F_2 populations derived from a self-fertilised F_1 require only two generations to produce. (b) Further self-fertilizing of F_2 lines for more than eight generations through single-seed descent will result in fully homozygous recombinant inbred lines (RILs).(c) Parental backcrossing of the F_2 while selecting for specific chromosomal segments will result in homozygous near isogenic lines (NILs) after more than seven generations and a single round of self-fertilizing. (d) Diversity panels are collections of independent, self-fertilised accessions (Acc.) that result from thousands of years of self- and cross-fertilization and mutating between ancestral genotypes that can no longer be retraced*

BOX2. Evolution of photosynthesis phenotyping tools and their potential for phenomics

The genetic mapping of photosynthesis-related traits requires robust and efficient phenotyping protocols that can be used to reliably assess these traits between different plant genotypes. A range of methods have been developed in order to estimate the performance of photosynthesis-related traits in plant leaves. The most important ones are listed here.

Measurement of leaf gas exchange

Infrared gas analysis technology allows direct measurements of carbon dioxide uptake and release and water vapour release by plants (Long *et al*., 1996). Infrared gas-analyser systems intended for the measurement of photosynthetic gas exchange are available as highly integrated systems, including, for example, measurement systems for leaf temperature, irradiance and chlorophyll fluorescence, and with extensive on-board data processing. Infrared gas analysis systems are accurate, but also expensive, and the measurement protocols are generally time-consuming (Stinziano *et al*., 2017). They require an operator, making them unsuitable for robotic use. Recently a method for the faster measurement of *A*/*Ci* curves (net CO_2 assimilation rate, *A*, versus calculated substomatal CO_2 concentration, *Ci*) has been proposed, allowing a complete *A*/*Ci* curve to be measured in about 5 min (Stinziano *et al*., 2017) (this corresponds to about 100 assessments per day). While this is a welcome advance, it needs to be seen in the context of the potential need to measure 1000–2000 plants for a genome-wide association analysis, and to do this quickly enough to avoid problems due to the variation in assimilation that can occur over even a day.

Assessment of chlorophyll content and other leaf reflectance-based methods

The reflectance spectrum of a leaf and changes in this reflectance spectrum can be a useful source of biochemical and physiological information. In some cases reflectance-based measurements have been developed for high-throughput phenotyping applications (e.g. measurement of the content of leaf chlorophyll and other foliar pigments). In other cases optical methods, which have been developed for more intensive laboratory applications, have not yet been used for high-throughput applications despite being, in principle, suitable for this.

Measurement of chlorophyll content is often performed using laborious destructive protocols, which greatly reduce their suitability for the high-throughput assessment required in genetic mapping studies. Another problem arises from the variation of chlorophyll content over time (Flood *et al*., 2016b; Hardwick and Baker, 1974; Lin *et al*., 2017). The foliar content of chlorophyll and other photosynthetic (and non-photosynthetic) pigments can be measured via leaf reflectance measurements (Gitelson *et al*., 2001, 2003, 2006) or by means of chlorophyll fluorescence (Cerovic *et al*., 2002; Gitelson *et al*., 1999). These essentially optical methods can be readily implemented using narrow-band (i.e. via optical filters; e.g. Flood *et al*., 2016b) or (hyper)spectral imaging (Yendrek *et al*., 2017), making them suitable for high-throughput phenotyping.

Leaf reflectance or absorbance measurements have also been used to estimate physiological properties connected with photosynthesis. The photochemical reflectance index (Gamon *et al*., 1997), which is correlated with the de-epoxidation state of the xanthophyll-cycle pigment pool and thus the qE component of non-photochemical quenching (NPQ) (Gamon *et al*., 1997; Herritt *et al*., 2016; Ruban *et al*., 1993), has been developed for field- and even satellite-based remote sensing applications (Drolet *et al*., 2005). The physiological traits Vc_{max} and J_{max}, which are commonly obtained from an analysis of *A*/*Ci* curves, have been estimated from near-infrared reflectance data (Silva-Perez *et al*., 2018) using an approach that would be highly suited to high-throughput applications. Other light-induced absorbance change techniques that have been developed for intensive measurements of photosynthesis (the electrochromic shift and absorbance changes induced by near-infrared light; Baker *et al*., 2007) could be used in imaging-based, high-throughput applications – if this is technically feasible. So far, however, there are no examples of these other potentially useful techniques being applied in fully automated, high-

throughput applications, although some have been refined for use in hand-held field phenotyping instrumentation (Kuhlgert *et al*., 2016).

Chlorophyll fluorescence-based techniques

Chlorophyll fluorescence has been used extensively to measure photosynthetic processes in chlorophyll-containing samples, such as isolated photosynthetic pigment-binding complexes, isolated thylakoids, chloroplasts and leaves (reviewed in depth by Baker, 2008; Harbinson, 2018). Chlorophyll fluorescence is the reverse of the light absorption process that forms excited states of chlorophyll *a* in the first place, and while fluorescence occurs from all chlorophyll, it is from PSII that most fluorescence is emitted and the yield of PSII fluorescence is strongly influenced by its physiological state (**Figure 2**). The most important traits derived from chlorophyll fluorescence are photosynthetic efficiency (F_q'/F_m'), which is a measure of photosynthetic efficiency in light-adapted leaves, and the maximum photosynthetic efficiency (F_v/F_m) measured in dark-adapted leaves under conditions where the yield of photochemistry is maximal and no rapidly relaxing NPQ is present. Modulated chlorophyll fluorescence techniques are the most commonly used methods for measuring chlorophyll fluorescence *in folio*, and with these techniques a wide range of physiologically useful parameters can be derived (Baker, 2008; Harbinson, 2018; Murchie and Harbinson, 2014). Chlorophyll fluorescence-based techniques are continuously being developed in order to improve their efficiency. For example, a single assessment of NPQ in plants requires a long dark-adapted phase lasting for 10–20 min (Baker, 2008; Murchie and Lawson, 2013). To overcome this requirement, Tietz *et al*. (2017) developed a novel protocol termed $NPQ_{(T)}$, which drastically improved the speed at which each measurement takes place. The relative ease with which chlorophyll fluorescence methods can be applied using imaging techniques means that they are widely used in the high-throughput phenotyping of photosynthesis.

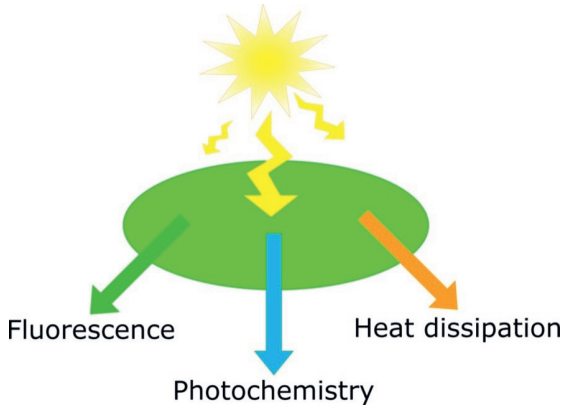


Figure 2. *The fate of quanta absorbed by photosystem II.The absorption of photons by photosystem II results in the formation of excited, energy-rich chlorophyll a molecules. These can be used to drive photochemistry, they may lose their energy as heat or they may lose the energy by re-emitting it as a photon (fluorescence). Photochemistry in photosystem II drives the oxidation of water and the reduction of plastoquinone and is the start of the linear electron transport chain. The linear electron transport chain drives carbon assimilation. Heat dissipation in the form of non-photochemical quenching is a regulated process that is controlled to dissipate excited states of chlorophyll a that are in excess of the needs of photochemistry. Non-photochemical quenching is a protective process, as by dissipating the excited states*

as heat it reduces the potential for producing damaging reactive oxygen species in photosystem II. Whenever the rate of one of these parameters changes, the rates of the other two are adjusted proportionately. This makes chlorophyll fluorescence a sensitive and thereby extremely useful tool for assessing photosynthetic performance in plants.

Basic Principles of the Green Machine

Photosynthesis is a diverse physiological process that couples the energy of absorbed quanta to metabolism, permitting otherwise endergonic reactions to proceed. Oxygenic photosynthesis, which occurs in cyanobacteria, algae and plants, is the dominant type of photosynthesis and sustains most life in the biosphere. In short, photosynthesis comprises the following steps: (i) light absorption by photosynthetically coupled pigments (primarily chlorophylls) in PSI and PSII; (ii) the use of this absorbed energy to drive chemical reactions that result in the formation of energy-rich metabolically useful compounds (reduced ferredoxin or NADPH, and ATP); (iii) and metabolic activity driven by NADPH, ferredoxin or ATP (see **Figure 3** for an overview). The most conspicuous of the metabolic processes coupled to photosynthesis is CO_2 fixation, which is driven by the light-independent photosynthetic carbon reduction cycle (also called the Calvin cycle), which eventually results in the formation of sugars (initially as sugar phosphates) and starch. Next to underpinning most life in the biosphere, photosynthesis is the fundamental process for agricultural productivity. Photosynthetic efficiency, i.e. the efficiency with which light is used to drive the photosynthetic processes, is an essential measure of photosynthetic performance (Baker, 2008; Maxwell and Johnson, 2000). It effectively dictates the amount of light energy that is required in photochemistry to fix a unit of carbon dioxide. It is this process of carbon dioxide fixation that contributes most to biomass formation. The regulation of photosynthetic light harvesting and electron transport is dominated by the need to moderate the production of reactive oxygen species (ROS) as a by-product of photosynthetic chemistry. Both superoxide and singlet dioxygen (molecular oxygen) can be formed directly by the photosynthetic light-harvesting and electron transport systems. To this end, various processes, such as NPQ, the production of anti-oxidants and the regulation of electron transport, are present to moderate the formation of ROS and detoxify them if they are formed. These protective mechanisms are activated in response to stressful light conditions, either due to too much incoming light and/or as a result of physiological stress, to reduce the rate of formation of ROS by photosynthesis.

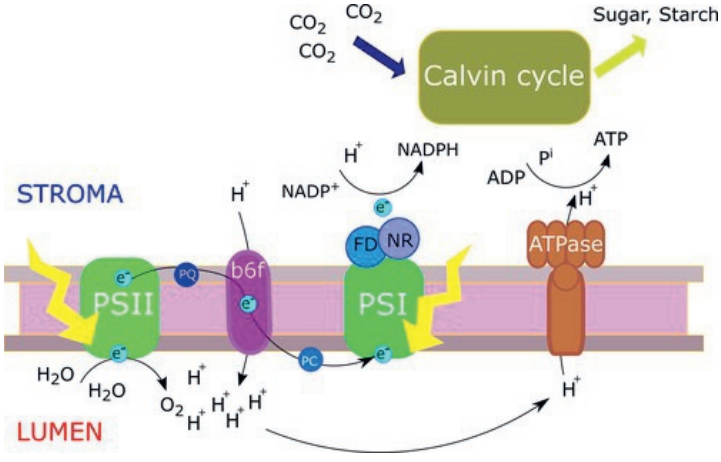


Figure 3. *A summary of photosynthesis. Light energy is used by photosystem II (PSII) to oxidise water and reduce plastoquinone (PQ) to plastoquinol. The reducing equivalents on plastoquinol move through the electron transport chain via the cytochrome b_6/f complex and plastocyanin, and are used by photosystem I*

(PSI) to reduce ferredoxin, a process that is also light-driven. Reduced ferredoxin in turn reduces $NADP^+$ to NADPH via the enzyme ferredoxin NADP oxidoreductase (NR). The oxidation of water by PSII and the oxidation of plastoquinol by the cytochrome b_6/f complex liberates protons (H^+) into the thylakoid lumen, generating a proton potential difference between the thylakoid lumen and the stroma. Protons passing down this proton potential difference through the ATPase drive the phosphorylation of ADP to form ATP.

Non-photochemical quenching in PSII is an important photoprotection mechanism. The absorption of quanta by PSII results in the formation of excited states of chlorophyll *a* in PSII. These are used to drive photochemistry in PSII and photosynthetic electron transport that result in the formation of NADPH and ATP. If the rate of formation of excited states of chlorophyll *a* in PSII exceeds the metabolic demands of photosynthetic metabolism, and thus photochemistry in PSII, the activation of NPQ ensures that the excess of excited states is dissipated as heat (Murchie and Harbinson, 2014). The activation of NPQ depends on decreases in the pH of the thylakoid lumen which will occur if the rate of proton deposition into the lumen, as a result of photosynthetic electron transport, exceeds the use of these protons by ATP synthesis (**Figure 3**). An increase in proton concentration leads to a fall in lumen pH, protonation of photosystem subunit S (PsbS) and the activation of violaxanthin epoxidase, which leads to the epoxidation of violaxanthin to antheraxanthin and eventually zeaxanthin (Latowski, *et al.,* 2011). Both the protonation of PsbS and the formation of antheraxanthin and zeaxanthin contribute to the formation of NPQ in PSII. Essentially then, NPQ thermally dissipates excited states of chlorophyll *a*, reducing the concentration of these excited states in PSII and reducing the risk of photodamage due to ROS formation. When light conditions become favourable again (i.e. lower), or when the plant adapts to the new light environment (i.e. photosynthetic capacity increases), lumen pH increases, deprotonating PsbS and activating a de-epoxidase that converts zeaxanthin and antheraxanthin back to violaxanthin. In this way, NPQ protects the integrity of the photosystems and reduces photodamage (Müller *et al*., 2001).

The Genomic Complexity of Photosynthesis

Photosynthesis is arguably the trait that above all others sets plants and algae apart from other, heterotrophic, eukaryotes, and their needs have largely driven the evolution of plants as terrestrial organisms over the past 450 million years. The importance of photosynthesis to the nature of plants can be seen by comparing parasitic plants with their more typical autotrophic relatives. The anatomy of parasitic plants is often reduced to only haustoria-like structures and reproductive organs (Kokla and Melnyk, 2018), and frequently there is the loss of plastids and photosynthetic genes (Hadariová *et al*., 2018; Vogel *et al*., 2018).

In the course of evolution, many genes originally residing on the chloroplast genome have moved to the nuclear genome, but their proteins are still mostly active in chloroplasts. A specific, conserved, sequence tag encoding the so-called ‘chloroplast transit peptide’ needs to be present in genes located on the nuclear genome in order for the translated protein to be properly transported into the chloroplasts (Bruce, 2000). Using the presence of this sequence tag for the transit peptide as a criterion for a photosynthetic gene, roughly 10–15% of plant genes are involved in photosynthesis (AGI 2000, Leister, 2003; Richy and Leister, 2004; van Rooijen *et al*., 2015). Given that the chloroplast genomes of embryophytes are estimated to contain only about a hundred of these genes, most of the genes coding for chloroplast-targeted proteins reside within the nuclear genome and can be genetically mapped – which amounts to roughly 3000 genes in the genome of Arabidopsis alone.

Although mutations in the chloroplast genomes can potentially have strong adverse effects on photosynthesis in plants they cannot be genetically mapped due to their strictly maternal inheritance in nearly all species and the absence of meiotic recombination. For example, a single nucleotide mutation in the chloroplast gene *Photosynthesis Subunit A* (*PsbA*) reduces the photosynthetic efficiency of Arabidopsis (El-Lithy *et al*., 2005). *PsbA* encodes for the D1 protein, a reaction centre protein of PSII, so it is involved in charge separation in PSII and the subsequent transfer of electrons to downstream electron acceptors (**Figure 3**). This process is blocked entirely by binding of the herbicide atrazine to the D1 protein (Barros

and Dyer, 1988; Kuhn and Böger, 1990). A point mutation causing a Ser264–Gly substitution in D1 leads to a reduction in the efficiency of PSII charge separation (Barros and Dyer, 1988; Kuhn and Böger, 1990). This mutation prevents atrazine from binding to the D1 protein, making the plant atrazine resistant, but at the cost of reduced photosynthetic efficiency and reduced growth. Despite the reduction in growth this is a viable phenotype, especially in atrazine-treated areas (El-Lithy *et al*., 2005; Flood *et al*., 2016a). This shows that strong artificial selection pressure is required in order to select for substantial negative chloroplast effects on photosynthetic efficiency. The contribution of other functional chloroplastic (or even mitochondrial) variation to phenotypic variance for photosynthesis in natural genetic populations or crop species accessions is currently unknown. However, such variation may in part be responsible for genetic variance for photosynthesis-related traits that cannot be accounted for in GWAS, also known as missing heritability (Brachi *et al*., 2011; Zuk *et al*., 2012). Analytical pipelines or statistical models that can take this variation into account will need to be developed. Identification of the phenotypic effect of chloroplast mutations could be performed by comparison of reciprocal F_2 or RIL mapping populations, in which the nuclear genome segregates in two different cytoplasmic backgrounds (El-Lithy *et al*., 2005). Alternatively, cybrids, cyto-swaps or cytolines may be used, in which a full nuclear genome is transferred to another cytoplasmic background, replacing the original nuclear genome, for example by recurrent backcrossing (Miclaus *et al*., 2016; Roux *et al*., 2016).

The conserved composition of photosynthetic complexes coupled with the importance of the process implies that not all nuclear-encoded chloroplast-targeted proteins are likely to have significant natural genetic variation (e.g. essential genes involved in maintaining the chloroplast membranes). For other genes associated with photosynthesis, but not coding for elements of the typically multicomponent photosynthetic complexes and enzymes, the situation may be different. van Rooijen *et al*. (2015) identified SNPs associated with variation in the recovery of photosynthetic efficiency after transition from low to high light. Associated with these SNPs were genes so far not linked to photosynthesis, although these were enriched for genes encoding chloroplast proteins. The large number of genes encoding chloroplast-targeted proteins found in the nuclear genome still leaves plenty of room for the discovery through forward genetics of novel nuclear genes involved in photosynthesis-related traits. Sequencing efforts in parasitic plants, which have lost a significant part of their photosynthetic capability, may provide further insight into their identification (Vogel *et al*., 2018). The fact that little research has been performed to identify such genes means that there is also still a lot to discover about the function of such novel genes in photosynthesis.

Insights from Genetic Mapping Studies into the Genetic Architecture of Plant Photosynthesis

From the above it is clear that photosynthesis is a complex quantitative trait that involves many genes. Quantitative trait locus mapping is an effective and widely used research tool to unravel quantitative traits, as it allows the unbiased discovery of the underlying functional variants of relevant genes. The diversity of genetic mapping studies seeking to dissect the genetic architecture of photosynthesis-related traits has increased markedly, which can be attributed to the awareness of the need to improve photosynthesis (Flood *et al*., 2011; Leister, 2003; Long *et al*., 2006; Zhu *et al*., 2010). Popular QTL-mapping population types such as RIL and NIL populations have been used to discover genetic variation and genes that contribute to variation in photosynthetic efficiency for a wide variety of species (e.g. Adachi *et al*., 2011, 2014; Gu *et al*., 2012; Yan *et al.,* 2015; Oakley *et al*., 2018; de Oliveira Silva *et al*., 2018). The design principles for these population types are described in Box 2. In line with what is expected of the genetic architecture of quantitative traits (Atwell *et al*., 2010), most studies find multiple QTLs for photosynthesis-related traits with a small effect size (Yan *et al*., 2015; de Oliveira Silva *et al*., 2018), which confirms the highly polygenic nature of the trait. Nevertheless, large-effect QTLs were found by Oakley *et al*. (2018) who assessed photosynthetic performance in a large RIL population of Arabidopsis exposed to cold stress. They concluded that the recovery of photosynthesis to cold was controlled by only a few loci, which indicates that some specific photosynthesis-related traits are influenced by functional genetic variation in a limited number of genes. Adachi *et al*. (2011) characterised several regions associated with improved leaf

photosynthesis by phenotypic analysis of a backcrossed NIL population composed of a commercial rice variety (Koshihikari) containing introgressions from a cultivar with high photosynthesis (Habataki). Subsequent introgressions of two of these loci into the Koshihikari genome significantly improved both total accumulated biomass and the rate of photosynthesis of the recipient genotype (Adachi *et al*., 2014). As such, Adachi *et al*. (2011, 2014) showed that improving photosynthetic efficiency through conventional breeding is a feasible strategy for further improving rice yield.

To summarise, photosynthesis-related traits are generally controlled by many QTLs with a small effect size and this reflects the results from genome-wide analyses of photosynthesis-related genes. Large-effect QTLs can nevertheless be detected for specific photosynthesis-related traits, which signifies the relevance for forward genetic screening. A random genomic region of a few million base pairs in size – which is a common size for introgressions in NIL populations and of QTL confidence regions identified in RIL populations – can easily encompass hundreds of genes, of which only one may be involved in the specific trait. To localise the causal gene(s) in such a gene pool would require a significant effort through follow-up fine mapping experiments (Adachi *et al*., 2017). Given the lengthy breeding process needed to improve the photosynthesis of crop species, based on newly identified genes, and the few decades left to meet the increased global demands of plant production (Kromdijk and Long, 2016), a more efficient gene identification pipeline is desirable. An advantage of GWAS over conventional bi-parental mapping population studies is the greatly improved genomic resolution with which causal genetic loci can be pinpointed (Box 1). As such, GWAS have the potential to result in much smaller lists of candidate genes that may control the traits of interest, and at a high density of SNP markers many associated loci may be identified. This would potentially allow faster identification and application of newly discovered genetic variation.

There have so far been only a limited number of GWAS aimed at elucidating photosynthesis-related traits, and an overview of all published studies we could identify is given in **Table 1**. Where generally few highly significant QTLs can be found in bi-parental mapping studies, most GWAS report numerous trait–marker associations. Many of these reach just above the LOD thresholds set to declare a region of significant interest, even after strong statistical corrections for population structure and/or experimental variance are applied (e.g. Hao *et al*., 2012). Most make only a small contribution to the total phenotypic variation. These findings illustrate the classic dichotomy between GWAS and bi-parental mapping studies; the former are more accurate at pinpointing causal variants and the latter are more powerful at detecting phenotypic differences caused by such variants (Bazakos *et al*., 2017). Nevertheless, QTLs in GWAS often identify SNPs that locate in or near genes that have a predicted role in photosynthesis-related traits following Gene Ontology enrichment analyses, which sets the precedent for a closer investigation of those genetic loci (Dhanapal *et al*., 2016; Herritt *et al*., 2016; van Rooijen *et al*., 2017; Wang *et al*., 2017). Functional characterisation of these genes may greatly advance our knowledge of the genetic regulation of photosynthesis and potentially lead to the discovery of new traits and genes that could be targets in plant breeding programmes (**Box 3**). Such characterisation will require additional physiological analysis of near-isogenic genotypes, such as mutants and wild types, including detailed, low-throughput leaf or plant gas exchange measurements, comparative transcriptomics, promoter studies, etc., as is common in gene function analysis.

BOX3. Highlights of two studies on functional genetic variation of plant photosynthetic efficiency

These studies have been selected to demonstrate the potential of using genetic mapping approaches to unravel genetic variation in photosynthetic traits up to the DNA sequence level. Causal genetic variation in selected candidate genes and their effects were confirmed in follow-up reverse genetics, and as such these studies fulfil the aim of characterising the impact of genetic variation on the phenotype.

Example 1: Natural variation in OsPsbS-1 affects NPQ in rice

Improving the rate at which NPQ is induced and relaxes will ensure that the level of photoprotection matches the level of incoming sunlight, which fluctuates drastically in field-grown crops (Murchie, 2017; Ruban, 2017). This trait is increasingly being recognised as a priority target for genetic improvement to boost crop productivity (Long *et al*., 2015; Taylor and Long, 2017). Violaxanthin de-epoxidase, zeaxanthin epoxidase and PsbS are the key regulators of the early, energy-dependent phase of NPQ (also known as the qE phase). Kromdijk *et al*. (2016) overexpressed these genes in tobacco and observed a higher accumulation of biomass in tobacco plants growing under field conditions compared with wild-type control plants, demonstrating the potential to improve NPQ to greatly benefit crop yield (Long *et al*., 2015).

Wang *et al*. (2017) explored natural variation of NPQ in 529 field-grown rice accessions and made use of a handheld pulse amplitude modulation fluorometer to assess the rate of NPQ in excised leaf parts. Two major marker–trait associations were discovered, one of which located to *OsPsbS-1* and explained more than 40% of the total genetic variance for NPQ in the rice diversity panel. Major allelic variation – which included a 2674-bp InDel located near the promoter region of *OsPsbS-1* – explained significant differences in the rate of NPQ, while no functional genetic variation was found that affected the strongly conserved protein sequence. Subsequent analysis confirmed that allelic regulation of *OsPsbS-1* expression was strongly correlated with NPQ. The ability to detect significant functional variation in *OsPsbS-1* by means of genetic mapping can be considered as a proof of concept of the idea that candidate genes belonging to important photosynthetic processes can be retrieved using genetic mapping approaches. Consistent with the hypothesis put forward by Müller *et al.* (2001) that the xanthophyll cycle and PsbS alone cannot account for total NPQ that takes place in plants during qE, Wang *et al*. (2017) discovered more than 30 additional loci that affect NPQ, some of which explained medium to high proportions of the phenotypic variance. This suggests that the regulation of NPQ in rice can be improved beyond the known core genes. Several of the marker–trait associations were confirmed using a bi-parental F_2 mapping population, and thus Wang *et al*. (2017) also demonstrated the power of applying both association panels and bi-parental mapping populations to discover and validate functional allelic variation in genetically complex traits such as photosynthesis.

Example 2: The role of YELLOW SEEDLING 1 in the high-light acclimation response of Arabidopsis

Plant acclimation to changes in irradiance intensity induces various responses at the genetic, molecular and physiological levels (Athanasiou *et al*., 2010; Bailey *et al*., 2004; Kouřil *et al*., 2013; van Rooijen *et al*., 2018). van Rooijen *et al*. (2015) presented the first association mapping study in which high-throughput phenotyping of photosynthesis-related traits was employed to elucidate these responses in Arabidopsis. They performed a GWAS of 344 accessions of Arabidopsis to study light acclimation after a stepwise increase in growth light intensity, after which photosynthetic efficiency (F_q'/F_m' or φPSII) was tracked multiple times a day for 4 days. Short- and long-term acclimation responses were identified, leading to the discovery of marker–trait associations that were present in different phases of light acclimation or throughout the experiment, effectively dissecting long-term light acclimation into different phases with the potential to derive different molecular mechanisms that account for them. By employing quantitative complementation methods with knockout mutants (Turner, 2014), allelic variations of several candidate genes were investigated to determine the light acclimation response in Arabidopsis. Allelic variation in

YELLOW SEEDLING 1 (*YS1*) made a significant contribution to long-term acclimation to high light. YS1, a pentatricopeptide repeat protein, locates to the chloroplasts and is involved in the modulation of RNA and chloroplast development (Kindgren *et al*., 2012; Zhou *et al*., 2009). It may be a key player in the development of chloroplasts in young leaves in response to high light conditions (van Rooijen *et al*., 2017). This finding may contribute to the little-understood transcriptional regulation of photosynthesis (Imam *et al*., 2014; Yu *et al*., 2014) and it demonstrates the power of using chlorophyll fluorescence in a high-throughput phenotyping facility to dissect a complex photosynthetic trait.

These studies are a first probe into identifying and characterising alleles of genes affecting photosynthetic processes. The extent to which these or other alleles may improve crop yield is unknown and may be tested in follow-up experiments.

Challenges of Phenotyping Photosynthesis-Related Traits

Diversity panel and bi-parental genetic mapping populations have been assessed for a large variety of different phenotypes related to photosynthesis. In the majority of these studies, however, single measurements are assumed to give a representative summary of the trait performance of plants. A particular problem with photosynthesis is that it quickly adapts to changes in the physical environment of the plant (e.g. to changes in light intensity, temperature and water or nutrient availability). Photosynthesis also responds strongly to stress, it often shows a diurnal rhythm and it changes according to the developmental stage of the plant. Choosing the exact timing for phenotypic evaluation of environmental responses to photosynthesis can be very challenging, as shown by the response curve of photosynthetic efficiency after a stepwise increase in light irradiance generated by high-throughput phenotyping in **Figure 4**. The outcome of the statistical test drastically changes over time, and even within a single day. Thus, if only a single measurement of the plants was made its timing would determine the conclusion about whether future investigations would be worthwhile or not.

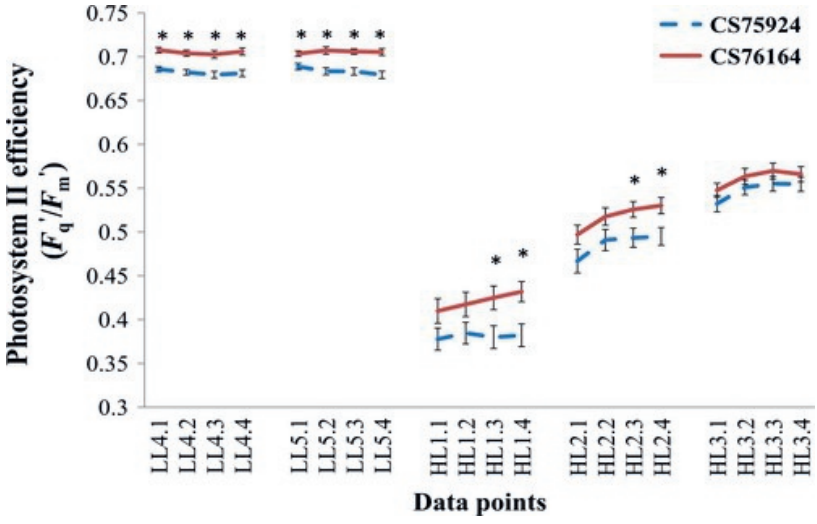


Figure 4. *High-throughput analysis of the response curve of natural Arabidopsis accessions Tanzania 1 (CS75924) and Landsberg erecta 0 (CS76164) to a stepwise increase in photosynthesis, starting from the second day of observations. Plants were grown as described by van Rooijen et al. (2015). Twelve replicates of each genotype have been assessed. Error bars depict ±1 standard error. Asterisks indicate*

significant ($P < 0.05$) differences over time and have been determined by applying Student's t-test. If no asterisk is indicated at a time point there was no significant difference between the photosynthetic efficiency of the two accessions. Each set of four data points, connected by a line, represents measurements for a single day [low light (LL), 100 µmol $m^{-2}sec^{-1}$ or high light (HL), 500 µmol $m^{-2}sec^{-1}$]. Data points should be read as follows: measurement of LL1.1 is taken at 09:00, LL1.2 at 11:30, LL1.3 at 14:30 and LL1.4 at 16:00, and the same for HL.

Even when applying stable light regimes that are used to reduce undesired phenotypic variation, both short- and long-term genetic variation in photosynthetic performance can be detected in the appropriate phenotypic setting (Flood *et al*., 2016b). This fluidity of photosynthesis creates particular challenges for its measurement. Accurate phenotyping of photosynthesis usually requires multiple measurements over the course of the day and over the life of the plant, or at least over a period long enough to encompass the changes in the process that are relevant to the photosynthesis-related trait of interest. In this section, various aspects of photosynthesis are discussed which require future attention in genetic mapping studies, but which impose serious challenges to phenotyping strategies. Photosynthetic responses due to environmental change are among those where most genetic variation is expected and novel traits are to be discovered (Lawson *et al.,* 2012), but which are simultaneously the most difficult to properly track (Murchie *et al*., 2018; Rungrat *et al*., 2016).

Fluctuating light

Light is the most variable resource throughout a plant's life. Plants growing under natural conditions experience constant fluctuations in the availability of light. Apart from the diurnal sun cycle, fluctuations are caused by the waving of leaves in the upper canopy, causing shadows on lower leaves (sunflecks), and drifting clouds blocking and exposing sunlight (cloudflecks) (Kaiser *et al*., 2018). Long-term light fluctuations include changes in weekly weather conditions and even seasonal variation, which affect plant development and photosynthetic efficiency. Recent insights into mismatches between incoming light and the ability to put that energy into use in carbon assimilation as a result of light fluctuations showed that these can cause considerable growth penalties and reduction in photosynthetic carbon assimilation in photoautotrophic organisms (Zhu *et al*., 2004; Graham *et al*., 2017; Vialet-Chabrand *et al*., 2017; Taylor and Long, 2017; Morales *et al*., 2018; Slattery *et al*., 2018). This has been further demonstrated in Arabidopsis plants that lack functional genes involved in photoprotection and light adaptation, such as in the PsbS *npq4-1* (Li *et al*., 2002) and *glucose 6-phosphate*/*phosphate translocator 2* (*gpt2*) (Athanasiou *et al*., 2010) mutants. Improvement of regulatory processes and adaptation mechanisms that allow plants to improve their physiological responses to fast and slow changes in light availability is highlighted as a promising path for improvement of crop photosynthetic efficiency (Long *et al*., 2015; Kromdijk *et al*., 2016; Taylor and Long, 2017; Morales *et al*., 2018).

Plants have developed various types of adjustment mechanisms to cope with incoming light, of which the short-term photoprotective mechanism of NPQ is one of the best investigated. In contrast to most traits related to light acclimation, rapid protocols are available for evaluating NPQ (Baker, 2008) and these have been successfully applied to discover significant allelic variation in *PsbS* in field-grown rice (see Box 3). Beyond PsbS-driven NPQ, additional mid- and long-term adaptations include movement of chloroplasts (Bailey *et al*., 2004; Wada *et al*., 2003) and the redistribution of light-harvesting complexes and reaction centres at the molecular level (Kouřil *et al*., 2013), as well as long-term adaptations in leaf architecture (Tardieu *et al*., 2017; Vialet-Chabrand *et al*., 2017). The constant switching of photosynthetic activity in response to changes in irradiance in both long and short terms requires a complex network of signalling mechanisms and regulatory gene networks in order to drive physiological adaptation; as sessile organisms, plants have to adapt physiologically as they are largely unable to evade increased irradiance or to relocate themselves to brighter spots under conditions of light shortage. Substantial heritability was found by van Rooijen *et al*. (2015) for several photosynthesis-related traits present during acclimation to a stepwise change in light intensity for 12 natural Arabidopsis accessions. Measuring rapid plant responses,

long-term acclimation and the relationships between these traits over time will impose challenges for high-throughput phenotyping and interpretation of short- and long-term light adaptation in populations of a size that allows bi-parental and/or association mapping.

Crop canopy architecture

In crop canopies, light availability is unevenly spread between top leaves and shade leaves and this affects the rate of photosynthesis between different canopy levels (Wilson and Cooper, 1969; Stewart *et al*., 2003; Zhu *et al*., 2012; Murchie, 2017). Strong discrepancies between incoming irradiance at the top canopy and bottom leaves are particularly prevalent in densely grown, vertically structured crop species such as maize, soybean, wheat and other cereals. Much of the scientific attention in crop modelling is on sun-exposed leaves, but shade leaves are a significant factor that may account for over 50% of carbon assimilation in crop canopies (Zhu *et al*., 2004, 2012). An uneven spread of incoming irradiance leads to heterogeneous exposure of plant leaves across the crop canopy and introduces patches of activated and de-activated photoprotective mechanisms (Murchie, 2017), which leads to de-optimisation of photosynthetic efficiency, carbon assimilation and eventually crop biomass accumulation. While optimising canopy architecture together with leaf-level functioning in the canopy is a clear route to improving photosynthetic efficiency of the canopy, and the canopy is the agricultural unit of production, the difficulties of fully assessing canopy architecture and within-canopy functioning are technologically burdensome. Not only must the canopy be mapped in space and time, but processes within the canopy (both physical and physiological) also need to be tracked in detail, which at the moment is not feasible on a large scale over time. Decreasing leaf chlorophyll content to allow more radiation into the canopy is a possible route to improving canopy photosynthetic efficiency (Ort *et al*., 2011, 2015). This should result in minor reductions in overall photosynthetic efficiency while significantly saving on nitrogen input (Slattery *et al*., 2018; Walker *et al*., 2018). Therefore, this trait has received considerable attention in GWAS (Dhanapal *et al*., 2016; Hao *et al*., 2012; Lin *et al*., 2017; Wang *et al*., 2015).

Photosynthesis affected by abiotic stresses

The importance of the effect of the environment on the photosynthetic machinery is reflected by the number of genetic studies that aim to investigate plant–environment interactions (Fiedler *et al*., 2016; Herritt *et al*., 2016; Oakley *et al*., 2018; Strigens *et al*., 2013). The response of the photosynthetic machinery to cold is the most popular area of study, which may be related to the conspicuous effect that low temperatures have on crop growth, especially in cool-temperate regions. Cold stress slows down the activity of photosynthesis – as it does all metabolic and diffusion-dependent processes – but low temperatures slow down plant growth more than photosynthesis, which results in photosynthesis becoming restricted by sink activity. Any restriction of photosynthesis will increase the extent of either photodamage to PSII (also known as photoinhibition; see Long *et al*., 1994; Takahashi and Murata, 2008; Nishiyama and Murata, 2014) or the extent of slowly reversible down-regulation of PSII (Demmig-Adams and Adams, 2006; Murchie and Harbinson, 2014). F_v/F_m represents the maximum quantum yield capacity of the photosynthetic machinery in plants, and can be determined on the basis of chlorophyll fluorescence. In the case of photodamage, low temperatures will simultaneously inhibit the synthesis of the D1 protein. Photodamage to PSII and the slowly reversible down-regulation of PSII will both manifest themselves as a slowly reversible loss of the F_v/F_m chlorophyll fluorescence parameter and are thus difficult to simply distinguish from each other (Murchie and Harbinson, 2014). This slow reversibility of the loss of F_v/F_m results in a loss of photosynthetic efficiency under more physiologically permissive conditions, and thus in lost productivity (e.g. Stirling *et al*., 1991). Photoinhibition and slowly reversible down-regulation of PSII are universally induced by many abiotic stressors, including, but not limited to, salinity, drought and heat stress (Nishiyama and Murata, 2014), even though there is variation in the exact physiological causes (e.g. ion toxicity, which interferes with protein cofactors, is specific to salinity stress). As photosynthesis is the

main driving force behind plant productivity, the discovery of traits that maintain photosynthetic functioning under environmental pressure is highly desirable (Rungrat *et al*., 2016). F_v/F_m (see Box 2) has been widely adopted in plant science to assess stress in plant leaves, primarily because unstressed leaves always tend to reach a value of around 0.83 for this trait in all plant species (Björkman and Demmig, 1987) but also because it is easily assessed using handheld devices (Murchie and Lawson, 2013). Generally, the use of the dark-adapted F_v/F_m to measure stress means that photosynthesis is being used as a probe of the plant's physiological state – the loss of photochemical efficiency may be a secondary effect to a more significant stress acting elsewhere. The loss of photosynthetic function in plants due to abiotic stresses is best described as a curvilinear transitory processes rather than a stepwise physiological transition. Insights into these plant response curves for photosynthesis may provide valuable information on how the photosynthetic machinery is preserved under environmental constraints – this is important for the recovery of photosynthetic efficiency after the stress has been alleviated. Response curves can only be constructed by repeated direct measurements of photosynthesis-related traits and can subsequently be used to map specific phases of stress induction, recovery and adaptation.

Photosynthesis affected by biotic stresses

Next to abiotic stresses, crop plants are being attacked continuously by different types of organisms that may have a severe impact on the integrity of plant tissue. Imaging techniques are being adopted as a convenient method to detect early infections of plant leaves (Scholes and Rolfe, 2017; Araus and Cairns, 2014; Singh *et al*., 2016; Zhao *et al*., 2016). However, as with abiotic stresses, biotic stresses also affect plant photosynthesis (Cheng *et al.,* 2016) and this may affect overall plant metabolism (e.g. root infections affect the nitrogen metabolism required for supporting photosynthesis-related proteins) (Berger *et al*., 2007) and expression of photosynthesis genes (Bilgin *et al*., 2010). Elucidating genetic variation that controls the physiological responses of the photosynthetic apparatus under biotic stress has not been studied in any of the QTL works reported in this review and might be a topic of future investigations. The difficulty in the phenotyping of such traits resides in the spatial tracking of the impact of microbes at their sites of infection and high-resolution phenotyping would be required to do this.

Many aspects of crop photosynthesis described here may become more important in the near future, for example because heat and drought stress are thought to affect crop production (Schauberger *et al*., 2017; Zhao *et al*., 2017) or because the traits linked to certain phenotypes are highly potent to further improve the genetic yield potential of crops (Lawson *et al*., 2012; Long *et al*., 2015; Ort *et al*., 2015). Single measurements – while they can be very important (Box 2) – may still only be able to capture a limited spectrum of traits that are involved in adapting and maintaining photosynthesis under growing conditions as described here. Thus, more physiologically complete phenotyping technologies have to be adopted to truly analyse traits and genes that may contribute to improve photosynthesis.

Towards High-Throughput Phenotyping to Study Natural Variation in Photosynthesis

Phenomics is often characterised as the collection of multidimensional phenotypic data covering physiological traits, from the cellular level up to the whole organism (Houle *et al*., 2010). This includes taking into account phenotypic plasticity induced by different environments and the developmental progression of the plant (Furbank and Tester, 2011; Tardieu *et al*., 2017). In theory, the phenome of plants comprises a near endless realm of spatial and temporal phenotypes that result from the interaction of a range of plant traits with each other and the environment. In practice, plant phenomics for genetic studies is often resolved as breaking down universal, whole-plant phenotypes into separate ones – in space or time, or both – following the logic that each 'sub' trait is being controlled by a smaller number of genes (**Figure 5**). Functional variation within such traits, which is only controlled by a subset of genes, is expected to show stronger trait associations in genetic analysis than when more general phenotypes are recorded, thus improving phenotypic resolution (Tian *et al*., 2011; Yang *et al*., 2014; Crowell *et al*., 2016; Zhang *et al*., 2017; Prado *et al*., 2018). This was conceptually verified by Crowell *et al*. (2016) who

dissected the trait of panicle size into 42 distinct and targeted traits in rice. Strong marker–trait associations were found linked to distinct genes in genomic regions that were otherwise identified as weaker single loci contributing to yield, in addition to new ones that could not be detected by using a universal major phenotype. This implies that an important bottleneck in GWAS, namely its low detection power, can be overcome by high-resolution phenotyping.

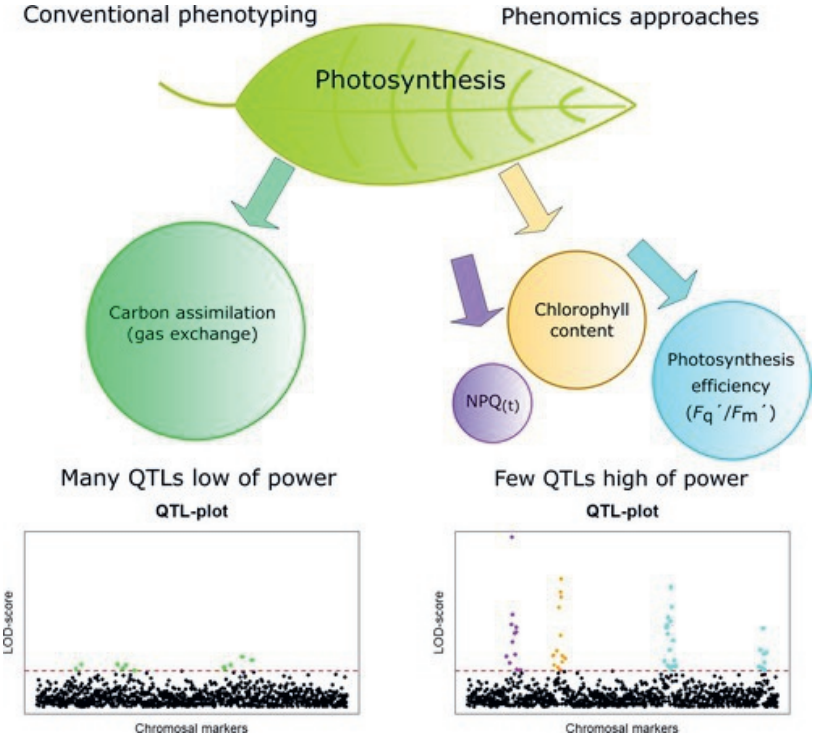


Figure 5. *Converging genomics and phenomics. In genetic mapping studies, a heritable, quantitative trait of interest is usually determined as a one-time assessment – here represented as an assessment of carbon assimilation in a plant leaf. Such traits are highly polygenic since different processes controlled by many more genes are responsible for their phenotypic outcomes. This results in many, barely significant (low power) quantitative trait loci (QTLs). Phenomics approaches allow the physiological and/or temporal phenotypic dissection of the trait, effectively splitting it into many traits. These traits are individually controlled by a subset of genes and/or pathways, which makes functional variation in one or a few genes belonging to this subset much more pronounced and thereby more detectable.*

The development of imaging technology has greatly expanded in the past decade as a result of the adoption of chlorophyll fluorescence and hyperspectral techniques as sophisticated plant phenotyping analysis tools (Box 2). The introduction of handheld fluorescence devices, such as the Multispec (Kuhlgert *et al*., 2016) and tools to measure the soil plant analysis development (SPAD) index, have reduced the time needed for assessing photosynthesis-related traits and opened the possibility for genetic mapping studies of photosynthesis even in the field. A thorough analysis of traits such as photosynthetic efficiency is still challenging, as is obvious from the limited number of phenotyping moments that can be accomplished in the time between sowing and harvesting, especially in field experiments (see **Table 1**). Particularly problematic is the incorporation of both short-term physiological responses and longer-term acclimation of photosynthesis in phenotyping assays that are highly sensitive to environmental fluctuations, that are

controlled by different genetic pathways and that are acting at different time points. These features strengthen the demand for sophisticated phenotyping facilities (Flood *et al*., 2011; Kaiser *et al*., 2018; Murchie *et al*., 2018; Rungrat *et al*., 2016).

Understanding the problem of this phenotypic bottleneck led to the development of phenotyping facilities capable of high-throughput measurement of photosynthesis-related traits and growth parameters for a large number of plants (e.g. Cruz *et al*., 2016; Jansen *et al*., 2009). Facilities that also have the capacity to conduct GWAS or bi-parental mapping populations are still rare, but are emerging (e.g. Cabrera-Bosquet *et al*., 2016; Flood *et al*., 2016b; Tschiersch *et al*., 2017). A key reason for this is the financial burden of investing in photosynthesis phenotyping facilities that are able to evaluate phenotypes for hundreds to thousands of plants – corresponding to the size of diversity panels with replicates, suitable for GWAS for example – several times per day within an interval of less than an hour (Fahlgren *et al*., 2015). Considerations concerning the experimental design of such facilities have to be taken into account, and will depend on traits of interest that are to be evaluated. We have already highlighted some of the photosynthetic traits that can be evaluated, each requiring different demands with respect to measuring tools and protocols to be used by the phenotyping facility. High-throughput data acquisition can be defined along two dimensions, temporal and spatial. High throughput on the temporal scale is achieved by repeating measurements over time. This requires a quick, but reliable, phenotyping protocol. Spatial high throughput is achieved by the ability to measure many plants within a short time, usually achieved by robotic instruments. The relevance of either axis will heavily depend on the trait under investigation, while the feasibility of doing so depends on the plant type, its developmental stage and its growing requirements.

Design approaches can be grouped into three different phenotyping principles and are generally formulated through trade-offs between crop size, high-throughput level of the method and the degree to which natural (field) growth conditions are met: moving the plant to the phenotyping sensors; moving the phenotyping sensors to the plant in a closed, controlled environment; and phenotyping in an open field or natural environment. The strengths and caveats of these design principles are critical to effective phenotyping.

Moving the plant to the phenotyping sensors

The most important staple crops are rice, maize, wheat and soybean. Given their relative importance it is unsurprising that such species are of major interest in GWAS that aim to identify the genes underpinning photosynthetic efficiency (**Table 1**). The complex architecture of these plants, especially when compared with Arabidopsis rosettes, implies that three-dimensional data need to be obtained for each plant from multiple phenotyping sensors, and this is difficult to do with a moving phenotyping device. One solution is to have a separate compartment with phenotyping equipment and have plants move to this by a conveyer belt system. Tschiersch *et al*. (2017) adapted a semi-controlled high-throughput phenotyping facility at IPK Gatersleben by incorporating chlorophyll fluorescence imaging systems into their existing phenotyping arsenal (**Figure 6a**, **b**). This is the first facility known to be able to three-dimensionally evaluate chlorophyll fluorescence-derived traits, such as photosynthetic efficiency, for the sizeable populations that are required for genetic analyses. ‘Moving-the-plant’ designs are particularly interesting for the implementation of tracking photosynthesis per leaf (Nagelmüller *et al*., 2016; Viaud *et al*., 2017), which would require sophisticated hardware for three-dimensional imaging. Another advantage of a static phenotyping chamber is that it can be equipped with several diverse imaging devices. It can therefore potentially be used to evaluate a larger number of traits per assessment for enhanced trait integration in downstream analyses (Cabrera-Bosquet *et al*., 2016; Tschiersch *et al*., 2017). Alternatively, modelling can be used based on parameters that are highly linked to key traits relevant to photosynthesis (Cabrera-Bosquet *et al*., 2016). It should be noted that moving the plants may add extra variance, or noise, to the measurements, because plants will respond to movement and be exposed to differential light irradiance and angles on their way to the phenotyping sensors. The need to move plants to a single phenotyping station

also limits the possibility of high-throughput measurements of the short-term photosynthetic responses of plants to the environment (i.e. over many minutes or hours).

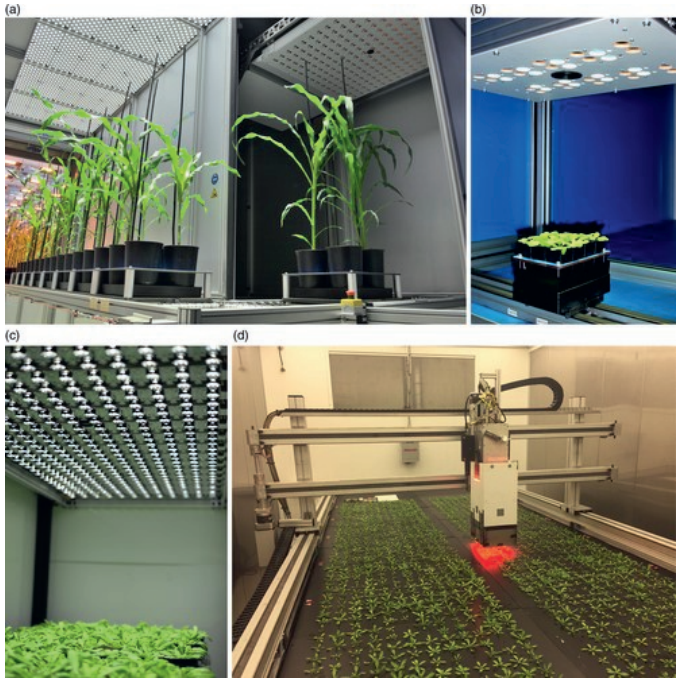


Figure 6. *A variety of high-throughput phenotyping facilities for photosynthesis following different experimental approaches that may be suitable for performing genetic mapping. (a) A high-throughput phenotyping facility located at IPK Gatersleben with phenotyping sensors to measure photosynthesis-related traits from all sides of plants the size of regular crop species (Tschiersch et al., 2017). (b) A close-up of the phenotyping module for small plants at the IPK phenotyping station. (c) The Dynamic Environmental Phenoyping Imaging system developed by Cruz et al. (2016) at Michigan State University. (d) the Phenovator, developed at Wageningen University and Research (Flood et al., 2016b).*

Photographs are used at courtesy of Henning Tschiersch, Thomas Altmann and Astrid Junker from IPK Gatersleben (a, b), Jeffry Cruz, David Hall and David M. Kramer from Michigan State University (c) and Tom Theeuwen from Wageningen University and Research (d).

Moving the phenotyping sensors to the plant in a closed environment

Arabidopsis is an excellent model species with respect to genetic analysis of photosynthesis, because of its compact size and flat growth architecture. These properties enable efficient phenotypic platform designs that have their measuring systems positioned above the plants (Cruz *et al*., 2016; **Figure 6c**; Flood *et al*., 2016b; **Figure 6d**). Ideally, these are systems that can assess each replicate simultaneously due to the implementation of a high density of fluorescence imaging systems (Cruz *et al*., 2016). A huge advantage of such systems is that growth light can be easily manipulated without disturbing the environment of the plants, as is the case when plants are moved to the phenotyping sensors. A high density of photosynthetic phenotypic measurements over consecutive time points allows tracking of short- and long-term photosynthetic responses under conditions of acclimation to a change in light conditions, as well as the short- and long-term physiological acclimation responses of photosynthesis that result from abiotic stresses (Rungrat *et al*., 2016) and light fluctuation (Vialet-Chabrand *et al*., 2017). Programmed, pre-determined

fluctuating light scenarios are of specific interest because they allow controlled experimental reproducibility, while simultaneously providing realistic phenotypic responses that are more likely to be observed in the field compared with stable light regimes (Vialet-Chabrand *et al*., 2017). This would also allow the study of abiotic stresses such as drought under more realistic growth conditions (Rungrat *et al*., 2016). The feasibility of this approach has been demonstrated by Athanasiou *et al*. (2010) and Kromdijk *et al*. (2016), who studied differences in light acclimation regulation of control plants compared with genetically transformed plants. This yielded no physiological response when plants were grown under stable conditions, but strong differences were found in seed yield (Athanasiou *et al*., 2010) and biomass accumulation (Kromdijk *et al*., 2016) under realistic growth conditions with naturally fluctuating light. These reports are promising for translating results from climate-controlled environments to field conditions.

Phenomic analysis of photosynthesis in field-grown plants

The ultimate challenge of plant phenomics is to develop high-throughput phenotyping methods to assess plants under field conditions (Araus and Cairns, 2014; Großkinsky *et al*., 2017; Murchie *et al*., 2018). This deserves more attention, as the correlation between results from experiments in climate-controlled conditions and those in the field is often low, which hampers the translation of results from controlled-environment experiments to applications in agronomic conditions (Araus and Cairns, 2014). Nevertheless, phenotyping in the field will always require that many uncontrollable and irreproducible variables are taken into account, adding to the environmental variance in genetic studies which will considerably affect the power to detect QTLs for photosynthesis. This has been demonstrated by Wang *et al*. (2017), who found that NPQ was related to genetic variation in OsPsbS-1 in only one of the experimental replications, not both, which they related to different weather conditions between years. As such, further validation of QTLs – either by using experimental repetitions or testing different environments to determine the sensitivity of the QTL to environmental fluctuations (Tardieu *et al*., 2017) – will be paramount for identifying key regulators of photosynthesis or using the QTL information to develop markers for marker-assisted breeding, especially when testing in field conditions. Up until now there have been no automated high-throughput phenotyping systems that can thoroughly assess plant photosynthesis in the field while accounting for in-field light fluctuations. Overcoming such problems may probably rely on field modelling of photosynthetic performance (Murchie *et al*., 2018). Scaling down on phenotypic precision enables measuring full crop canopy structures through broad view spectral and thermal imaging, which can significantly improve high-throughput evaluation of a variety of photosynthetic traits (**Box 1**). Such in-field imaging techniques have been successfully employed by Herritt *et al*. (2016) to measure the epoxidation state of xanthophyll pigments as an indirect measure of NPQ. High-throughput phenotyping of photosynthesis using such imaging devices should be relatively easy to include in phenotyping systems like those that support cameras which move above the crop canopy (Kirchgessner *et al*., 2016; Sun *et al*., 2018), and may provide a solution for measuring specific traits for field-grown crops.

Genetic Analysis of Photosynthesis Phenomics Data

Phenomics data for GWAS

Photosynthesis phenotyping facilities, including those that can be employed in field studies, are rapidly developing to increase their trait measuring capabilities, thanks to diversifying and improved measurement protocols (**Box 2**). Essential for dealing with phenomics datasets will be the adoption of a streamlined processing pipeline for data handling, standardisation and storage (Großkinsky *et al*., 2015). Similar to phenomics data for other traits, photosynthesis data obtained from high-throughput phenomics facilities will introduce new challenges associated with amalgamating gigabytes or even terabytes of imaging data and their derived measurements into a comprehensible whole with a meaningful biological interpretation

(Singh *et al*., 2016; Tardieu *et al*., 2017). Latent factors which might affect plant features without being explicitly measured can be detected by applying multivariate modelling (Dormann *et al*., 2012) or training unsupervised machine learning approaches that consider the influence of hidden or unmeasurable factors (Singh *et al*., 2016). These factors can then be corrected for or integrated in biological models. Furthermore, machine learning approaches and data modelling may significantly aid the use of phenomics data in helping to identify patterns related to specific traits (Cabrera-Bosquet *et al*., 2016; Murchie *et al*., 2018; Singh *et al*., 2016) – hereby referring to traits associated with development over time, but also traits that may be spatially linked, for example within the same leaf. This is especially useful in discerning true outliers from biologically interesting anomalies that may be lost in simpler models (Xu *et al*., 2015). To reduce data dimensions, alternative methods can also be explored, such as multitrait GWAS. Here, multivariate analyses are being employed to simplify the dimensions of phenotypic assessments which can improve the strength of QTLs – especially when different traits are assessed that share a common underlying genetic regulation (Thoen *et al*., 2016). However, care must be taken not to oversimplify data obtained from phenomics, as this could potentially hide important nuances in the data. The QTL leading to the discovery of the effect of natural genetic variation in *YS1* on acclimation of Arabidopsis to high light would not have been detected if such methods alone were employed (van Rooijen *et al*., 2017).

From candidate gene to improving crop productivity

The ultimate goal of improving photosynthetic efficiency in plants will eventually reside in improvement of harvestable plant products (Kromdijk and Long, 2016). Improving photosynthesis, though it might result in increased total biomass accumulation, does not necessarily directly translate into greater agricultural yield. For example, Adachi *et al*. (2014) introgressed QTLs supporting carbon assimilation into an existing rice cultivar; this resulted in an increase in total biomass but not grain yield. While transgenic approaches have been successful in improving plant yield through improved photosynthesis (e.g. Driever *et al*., 2017; Kromdijk *et al*., 2016), and similar results might be achieved based on spontaneous mutants, the natural allelic variation for many photosynthesis-related traits may be more subtle, often with many small-effect QTLs contributing to the trait (van Rooijen *et al*., 2017). This will complicate the identification of causal genes and limit the direct use of identified alleles in breeding programmes, although it will increase our knowledge about which genes are relevant for improving photosynthesis (**Box 3**) – which can then potentially be used in transgenic approaches for crop improvement and inspire the search for additional allelic variants with strong affects. When many small-effect QTLs are involved, a genomic selection approach may be adopted to improve photosynthesis, which may be more effective than breeding for a few selected loci (Hamblin *et al*., 2011). As indicated above, phenomics facilities are generally equipped with phenotyping sensors that measure growth and development traits beyond photosynthesis-related traits, allowing the setting up of a complete, physiologically comprehensible network analysis of how all traits impact upon, or are beneficial to, crop yield. The further use of low-throughput, but highly accurate, measuring tools, such as gas-exchange measurement devices on contrasting genotypes identified through high-throughput imaging analysis, will be needed to properly assess the impact of photosynthesis on crop physiology.

Epistasis in highly polygenic traits

Due to the existence of complex regulatory networks that affect photosynthesis (Imam *et al*., 2014; Yu *et al*., 2014) and the large number of genes that are involved, QTL mapping models that solely assume additivity are likely not sufficient to derive a complete comprehension of the genetic architecture of photosynthesis. Epistasis, meaning interaction between genes resulting in non-additivity of QTLs, has often been regarded as a rather elusive and difficult phenomenon to study, and is therefore often disregarded by plant geneticists and breeders. However, epistasis is expected to be common in highly polygenic traits (Mackay, 2014). The occurrence of epistasis is a probable explanation for the failure to

detect QTLs in bi-parental mapping approaches and GWAS, despite the presence of high heritabilities (Brachi *et al*., 2011; Lachowiec *et al*., 2015; Zuk *et al*., 2012). Early algorithms that were able to detect epistasis could only make use of small datasets and were too cumbersome to analyse the many gene-by-gene interactions to be found in diversity panels. New efficient ones are being developed that will make such analyses accessible to research groups without sophisticated high-performance computing infrastructure (Ning *et al*., 2018; Tsai *et al*., 2017; Zhu and Fang, 2018), which should lead to a more complete comprehension of highly complex traits such as photosynthesis.

Conclusion

The need for understanding and genetic improvement of plant photosynthesis is widely recognised as a vital development to combat looming future yield gap deficiencies (Flood *et al*., 2011; Lawson *et al*., 2012; Long *et al*., 2015; Zhu *et al*., 2010). This recognition, together with the development of chlorophyll fluorescence technologies that allow large numbers of plants to be evaluated within a short timeframe, has led to an increasing number of forward genetic studies aimed at detecting the genetic loci that contribute towards improving photosynthetic efficiency of crop plants. The introduction of GWAS has accelerated the speed with which candidate genes can be identified in photosynthesis-related traits when compared with bi-parental mapping population studies, but combining these population types to elucidate functional genetic variation and traits is likely to be the most rewarding approach (Bazakos *et al*., 2017).

In this review, we conclude that the current implementation of high-throughput phenotyping will become necessary to elucidate the genetic architecture of photosynthesis-related traits in mapping populations – especially with regard to the responses of photosynthesis to abiotic stresses, plant development and light acclimation. The successful identification of target QTLs and underlying genes using conventional methods greatly depends on the specific trait of interest, but has been well-demonstrated (**Box 3**). The past 5 years have seen a marked improvement of and increase in the application of high-throughput phenotyping tools, many of which can be incorporated with, employed or adapted to include sensors that measure photosynthesis. Exploring traits that contribute towards dynamic acclimation under conditions of fluctuating light may become a tractable option within a couple of years, given that more time-efficient tools are gradually being developed. The final, and very challenging, step would be to obtain field data at a desired level of phenotypic precision comparable to what can be achieved in climate-controlled growth facilities. The integration of both controlled and field studies will be necessary to validate gene candidacy and evaluate its application potential. We expect that the developments described here will enable further discoveries of photosynthetic trait in both science and industry to meet the ever growing demand for food for human consumption in the next three decades and beyond.

Acknowledgements

RFHMvB is supported by a research grant provided by Rijk Zwaan Breeding BV. JH and MGMA gratefully acknowledge the support of the research programme of BioSolar Cells, co-financed by the Dutch Ministry of Economic Affairs. Furthermore, we would like to acknowledge Tom P. M. J. Theeuwen (Wageningen University and Research), Henning Tschiersch, Thomas Altmann and Astrid Junker (Leibniz Institute of Plant Genetics and Crop Plant Research, IPK Gatersleben) and Jeffry A. Cruz, David Hall and David M. Kramer (Michigan State University) for use of their images in this work.

Conflict of interest

The authors are not aware of any conflict of interest arising from drafting this manuscript.

Chapter 3. Copy number variation in *SQUALENE EPOXIDASE-LIKE* genes determines photosynthesis efficiency in *Arabidopsis thaliana*

Roel van Bezouw[*1], Thu Phuong Nguyen*[*1], Rosanne van Velthoven[1], Elena Vincenzi[1], Hui Zhang[1], Renze Jacobi[1], Yanrong Gao[1], Bastiaan de Snoo[3], Francisca Reyes Marquez[1], Livia Vazquez Catalá[1], Muhammad Iqbal[1], Stan Jansen[1], Luc Buvelot[1], Cris Wijnen[1], Jeremy Harbinson[2], Joost Keurentjes[1†], Mark G.M. Aarts[1†]*

SUMMARY

The study of natural genetic variation is increasingly receiving attention and indicated as a potential target to improve crop species. In this work, an *Arabidopsis thaliana* recombinant inbred line population was grown to study photosynthesis efficiency under different environmental conditions. We provide evidence of copy number variation of a previously undescribed sub-family of a *SQUALENE EPOXIDASEs (SQEs)* to significantly impact photosynthesis efficiency and underlie the sole quantitative trait locus identified in this study. We additionally show that transgenic overexpression of *SQEs* can improve photosynthesis beyond values expressed by the parent demonstrating higher phenotypic values. The exact molecular pathways and processes in which these genes are involved remain unknown. In all, this study shows how bi-parental mapping populations and the availability of *de novo* genome sequences are a powerful combination to identify novel traits and genes involved in photosynthesis.

---

[1]Laboratory of Genetics, Wageningen University & Research, [2]Laboratory of Biophysics, Wageningen University & Research, [3]Rijk Zwaan Breeding B.V.
[*]*Leading author*, [†]*Senior author*

1. INTRODUCTION

Photosynthesis is the primary driver of plant production and, in fact, of most life in the biosphere. While all plants use basically the same photosynthetic mechanism, there is considerable variation between species and genotypes at the quantitative level on photosynthetic properties, such as the light-saturated rate of photosynthesis. This fine-tuning of photosynthesis arises most conspicuously from the need to optimize the overall efficiency of photosynthesis in terms of not only light-use efficiency, a complex parameter in the natural environment where fluctuating light is the norm, but also in terms of resource use efficiency (especially fixed carbon, nitrogen and other mineral nutrients) associated with the maintenance and development of the photosynthetic apparatus; water-use efficiency; and mechanisms that protect the photosynthetic machinery from damage, especially during periods where the intercepted quantum flux exceeds the metabolic demands of photosynthesis. As a result of this fine-tuning, photosynthesis in most species does not operate even at the biological limits defined by species with exceptional photosynthetic properties (Zhu *et al.*, 2010).

Selection against these limitations in photosynthesis efficiency had until recently not been considered as a target for plant breeding, leaving photosynthesis as one of the few major traits not to be optimized in crop species (Flood *et al.*, 2011; Lawson *et al.*, 2012). Thus, improving plant photosynthesis efficiency is regarded as one of the most promising ways to further improve the genetic yield potential of staple crops beyond rates that are currently realized (Zhu *et al.*, 2010; Ray *et al.*, 2013). Through modeling approaches limitations and options for improvement in key components of the photosynthesis apparatus have been identified. Many of these options are targets for genetic improvement (Zhu *et al.* 2015, 2020), including improving acceleration and relaxation of photo-protection mechanisms under fluctuating light conditions (Kromdijk *et al.*, 2016), enhancement of the regeneration rate of sedoheptulose-1,7-biphosphatase in the Calvin-Benson cycle (Driever *et al.*, 2017) and increasing the chance of photosynthetic recapture of CO_2 released during the photosynthetic carbon oxidation cycle (or photorespiration) (South *et al.* 2019). In each of these studies, biomass or yield improvements were reported to span the range of 15-40%, verifying that the potential exists for not only improving the properties of photosynthesis but using these improvements as means to increase crop yields.

So far the proof-of-principle experiments that have improved photosynthesis and yield relied on genetic modification. Making use of natural genetic variation is, however, an alternative route to seek for improvement of photosynthesis (Driever *et al.*, 2014; van Rooijen *et al.*, 2015; Faralli & Lawson, 2019). Better understanding the genetic basis for naturally occurring variation in photosynthesis will also open up new targets for improving photosynthesis by genetic modification or gene editing. The prospects of using natural variation in photosynthesis to improve crop yields has become increasingly feasible following the development of new phenotyping facilities in the last decade (Rungrat *et al.*, 2016; van Bezouw *et al.*, 2019). Such facilities are capable of rapidly measuring large numbers of plants and are thus suitable for the analysis of genetic mapping populations. As indicated by van Bezouw *et al.* (2019), studying natural variation in plants can lead to 1) the identification of novel genes previously not thought to affect

photosynthesis, 2) the identification of new potential targets for purposes of genetic engineering and 3) the identification of beneficial alleles that are readily available to be bred into existing cultivars. Genetic variation in photosynthesis traits is therefore a potential source for improving photosynthesis (Faralli & Lawson, 2019), which is underlined by studies that previously found significant variation for photosynthetic traits in several crop species (reviewed in Flood *et al.*, 2011, van Bezouw *et al.*, 2019). However, few examples of natural variation in photosynthesis traits have been explored up to the level of variation in genes (Jung & Niyogi 2009, Wang *et al.*, 2016; van Rooijen *et al.*, 2017; Basu *et al.*, 2018), and even fewer describe a functional connection between natural variation in a photosynthesis mechanism and plant productivity (Adachi *et al.*, 2017, 2019; Honda *et al.*, 2021). Thus, the hypothesis that natural variation in photosynthesis can be used to improve crops – in yield or other aspects – requires further examination in order to be attractive for more commercial applications.

Arabidopsis thaliana (hereafter: Arabidopsis) is a suitable model species to study this hypothesis. It is small enough to be conveniently grown in the large numbers needed for genetic mapping in high throughput phenotyping that use chlorophyll fluorescence imaging to evaluate photosynthesis traits (Rungrat *et al.*, 2016, Flood *et al.*, 2016, Oakley *et al.*, 2018). Furthermore, considerable variation exists among Arabidopsis genotypes for a variety of photosynthesis traits (Athanasiou *et al.*, 2010; van Rooijen *et al.*, 2015; Rungrat *et al.*, 2019; Flood *et al.*, 2020; Zhang *et al.*, 2021), which indicates that it is a suitable model species to explore the nature and impact of genetic variation on photosynthesis. The short reproductive cycle allows the development of a plethora of different genetic mapping population types with different parental lines (Wijnen & Keurentjes, 2014), which can be used to detect quantitative trait loci (QTLs) and the causal variants that underlie these. A particularly effective genetic mapping type is one composed of recombinant inbred lines (Bazakos *et al.*, 2017). Populations of recombinant inbred lines are derived from single seed descent and selfed for approximately eight generations to develop essentially homozygous, immortal lines that each carry a randomly determined, but unique mixture of the original parental genotypes. This subsequently leads to a resulting population with high genetic mapping power and a decent resolution (Keurentjes *et al.*, 2007). Because of the immortality of such lines in a selfing species, quantitative trait loci can be evaluated in replicate experiments covering a variation of environmental conditions to test their validity and robustness (El Soda *et al.*, 2014; Bazakos *et al.*, 2017).

In this study, we took advantage of the high mapping power, experimental replicability and genetic simplicity of a bi-parental recombinant inbred lines population to study photosynthesis efficiency under several environmental conditions. The population developed from the Ler-0 and Col-0 accessions previously developed by Lister & Dean (1993) was chosen because of the availability of *de novo* reference genomic sequences for both parental lines (Berardini *et al.*, 2015, Zapata *et al.*, 2016) and the availability of a high density linkage map (Singer *et al.*, 2006). The population was grown under a total of three treatments, control, nitrogen deficiency and low light conditions. Fine-mapping of the single photosynthesis QTL detected in the mapping population resulted in the discovery of a novel role in photosynthesis for a previously uncharacterized family of *SQUALENE EPOXIDASE*s.

2. MATERIALS AND METHODS

2.1 Seed material

We made use of a recombinant inbred line population developed from accessions Ler-0 and Col-0 (ABRC stock: CS1899, described in Lister & Dean, 1993). 97 lines that were previously genotyped for 676 genetic markers by Singer *et al.* (2006) were used in the mapping experiments. Segregating recombinant F2:3 families and introgression lines to fine-map and characterize the QTL were developed from a cross between double haploid introgression lines C5L-A10 x CSL32 and L5C-D12 x CSL1 (these lines are described in Wijnen *et al*., 2018). In each of the experiments, material with the same harvest date and from the same climate chamber or greenhouse compartment was used to avoid possible confounding effects introduced by parental growth conditions.

Up to three – or two when insertion lines were previously characterized – T-DNA insertion lines per target gene were ordered from the Eurasian Arabidopsis Stock Centre (http://arabidopsis.info/), potentially disrupting the function of the following genes; At5g24120 (SALK_101921, SALK_141383), At5g24130 (SALK_137191), At5g24140 (SALK_040805, SALK_012094), At5g24150 (SALK_024504, GABI_151G06), At5g24160 (SALK_083343, SALK_112366, SALK_118625) and At5g24165 (SALK_053722, SALK_069263). For At5g24155 no potential T-DNA insertion lines were available. Gene and T-DNA specific primers (**Supplementary table 1**) were used to identify homozygous T-DNA insertion plants among segregating progeny of these lines. Verified, but previously uncharacterized, insertion lines with T-DNAs locating outside the protein coding region were tested for transcription of the target gene. Only those lines in which the gene was not transcribed were considered to be a functional knock-out mutation of the target gene.

2.2 Plant phenotyping and experimental design

Plant phenotyping was performed in a robotic high-throughput platform based on chlorophyll fluorescence imaging as described by Flood *et al.* (2016) (hereafter called the Phenovator). The RIL-population was monitored in three different environmental conditions over two independent experiments, I & II. In experiment I, plants were grown under a growth irradiance of 100 µmol m^{-2} s^{-1} using normal N-supply ("100 µmol"). In experiment II, plants were grown a growth irradiance of 200 µmol m^{-2} s^{-1}, with two treatments of nutrient supply – control ("200 µmol") and nitrogen deficiency ("200 µmol -N", but fed with a solution of only 10% of the N-input compared to control conditions, see Erol, 2019, Dissertation, see https://library.wur.nl/WebQuery/wda/2253094 for details). Each RIL was grown in six replicates in experiment I and four replicates per condition for experiment II. For fine-mapping, T-DNA-insertion lines and transgenic analysis purposes, plants were grown under conditions of 100-120 µmol m^{-2} s^{-1} of light. Plants were phenotyped for the quantum yield of photosystem II electron transport in the light (ΦPSII) by dividing ambient light adapted chlorophyll fluorescence (*Fq'*) over the maximum chlorophyll fluorescence (*Fm'*) (Baker, 2008), for either four or five times per day depending on the experiment. Near infrared

imaging was used to obtain the projected leaf area (PLA) as arbitrary pixel counts of rosettes. In all experiments, and unless specified otherwise, plants with a PLA value below a 200 pixel value after ~20 days of growth were discarded from datasets as these mostly represent stunted or weakly developing individuals with aberrant phenotypes.

2.3 Statistical analysis and genetic mapping procedures

Prior to genetic mapping, phenotypic Best Linear Unbiased Estimates (BLUES) per genotype and heritability estimates were calculated for the RIL population using experimental block, table positions and camera position as random factors following Flood *et al.* (2016), with the R-package ade4 (Dray *et al.,* 2020). The quantitative trait locus (QTL) mapping procedure of r/QTLwas used to map QTLs in all experiments (Broman *et al.*, 2003; Arends *et al.*, 2010), based on the genetic markers and resulting linkage map previously developed by Singer *et al.* (2006). For comparisons of genotypes, analysis of variance involving a linear mixed model was used, using genotype as a fixed variable and the aforementioned blocking effects as random factors. All RNA expression data was log-transformed prior analysis. The package ggplot2 was used for visualization of the data (Wickham, 2016).

2.4 Genomic sequence analysis and development of genetic markers

The fine-mapped regions on chromosomes 5 of the Ler-0 *de novo* assembly (Genbank ref: CM004363.1, Zapata *et al.* 2016) and Col-0 (Genbank ref: CP002688.1, Cheng *et al.* 2016) were used to align the genomic regions and highlight genetic variation for the region of interest. A dotplot was made using the NCBI Blastn tool to visualize larger genomic polymorphisms in the region (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq). The NCBI multiple sequence alignment tool was used to map sequence polymorphisms and single nucleotide polymorphisms (SNPs). The SNP markers were used to develop Kompetitive Allele Specific PCR (KASP) probes used for analysis of recombinations and fine-mapping. The markers used in this study are listed in **Supplementary table 2**.

2.5 RNA expression analysis

For RNA expression analysis, 21-days old full rosettes were harvested and snap frozen in liquid nitrogen. The rosettes were ground to powder, and subsequently RNA was extracted using the Direct-Zol RNA isolation kit as provided by Zymo Research (Irvine, USA). A total of 1 µg of total RNA was added to initiate cDNA synthesis using the SensiFAST™ cDNA Synthesis Kit (Bioline, London, UK). The SYBR Green assay with SensiFAST™ SYBR no ROX kit was used for the quantification of transcripts in RT-PCR following the manufacturer's protocol. Primers were developed using the NCBI Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi), and in such a way that the 3' endings overlapped exon-exon junctions. Based on experience drawn from previous studies on RNA expression related to photosynthesis (van Rooijen *et al.* 2017), UBQ7 (At2g35635), CB5E (At5g53560) and OTUB1

(AT1G75780) were used as housekeeping genes for reference, All qRT-PCR gene primers used in this study are listed in **Supplementary table 3**. RNA expression values were calculated using the $2^{-\Delta\Delta Ct}$ method and $\log^{10}$ transformed prior analysis.

2.6 Cloning of vectors and selection of transformants

For each of the identified candidate genes, a transformation vector was developed that contains either the Col-0 or Ler-0 allele for the following genes: At5g24120 (*SIGMA FACTOR 5*, *SIG5*), At5g24150 (*SQUALENE EPOXIDASE 5*, *SQE5*), At5g24155 (*SQE7*), At5g24160 (*SQE6*) and At5g24165 (*PUTATIVE PLASTID PROTEIN, PPP*). The promoter region was assumed to cover at least 2000 base pairs (bp) upstream of the predicted ATG-start codon, or until the predicted transcribed part of the upstream gene. The terminator region was considered to cover at least 800 bp downstream from the predicted stop codon. Primers were designed to universally amplify both Col-0 and Ler-0 gene fragments (**Supplementary table 4**), except for *SQE5* for which two different forward primers had to be designed. Verifi high fidelity DNA polymerase (PCR Biosystems ltd, London, UK) was used to amplify each gene fragment. Restriction enzymes *Pst*I (pos: 2670) and *PsiI* (pos: 4573) were used to remove the *ccdB* gene from the pKGW_RedSeed vector and create a backbone (Ali *et al.*, 2012). Gene fragments and vector backbone were assembled following a modified version of the Gibson Assembly manufacturer's protocol (#E2611, New England Biolabs, Ipswich, MA, USA).

For overexpression (OX) of the Ler-0 allele of *SQUALENE EPOXIDASE 6*, RNA was isolated from a Ler-0 plant and cDNA was synthesized as previously described. The full-length cDNA was amplified using primers OX160_F and OX160_R (**Supplementary table 4**). The resulting product was subsequently cloned into the pFAST-R02 vector (Shimada *et al.* 2010). Note that the reverse primer had to be modified at the nucleotide level in order to have the fragment properly ligate to the cDNA in the correct orientation. This did not affect the amino acid sequence of the translated product.

An artificial micro-RNA (amiRNA) was designed using the Web microRNA designer (WMD3) tool as described in Schwab *et al.* (2006). The amiRNA sequence used, "TGTTGGTAAGGTAGAACACCG", was predicted to target At5g24160 at 100% identity, and At5g24155 and At5g24150 at 95% and 85% identity, respectively (**Supplementary figure 1**). The original protocol used the primers A and B to amplify the amiRNA precursor fragment, however we used nested primers (amiF and amiR, **Supplementary table 4**) to facilitate the cloning in the pENTR/D-TOPO vector (Thermo Fisher Waltham, Massachusetts, United States). Once cloned and sequenced, an LR reaction was performed to transfer the amiRNA precursor fragment into the pFAST-R02 binary vector, as described by the manufacturer (Gateway cloning protocol, https://www.thermofisher.com/nl/en/home/life-science/cloning/gateway-cloning/gateway-technology.html, Thermo Fisher). All the cloning steps were performed using standard molecular biology techniques as described in Sambrook *et al*. (1989).

All vectors were verified by sequencing and transformed into the *Agrobacterium tumefaciens* strain GV3101. *Arabidopsis thaliana* was transformed following an adapted protocol for floral dipping (Clough & Bent, 1998). Transgenic T1 seeds were selected for their dsRED marker phenotype under a UV

fluorescence microscope. After harvesting the T1 plants, T2 seeds were subsequently selected based on the increased fluorescence signal following Shimada *et al*. (2010). Copy number quantification in allelic transgenes was analysed following qPCR as previously described, but DNA sampled from T1 seedlings was used as template. Primers were designed on the dsRED gene to detect insertion copy number. Relative quantification of transgene copies was performed against two single copy reference genes. Primers used for copy number quantification are listed in **Supplementary table 5**. For RNAi and OX-lines, respective overexpression and silencing functioning were verified in T1 plants, from two leaves per independent transformant, after which the most promising lines were subsequently phenotyped for photosynthesis.

3. RESULTS

3.1 Genotypic differences in ΦPSII between Col-0 and Ler-0 are controlled by a single QTL on chromosome 5

Col-0 and Ler-0 exhibit a significantly different level of photosystem II efficiency (ΦPSII) at conditions of a growth irradiance of 100 µmol m^{-2} s^{-1} (**Figure 1a**), which makes their RIL-population suitable to screen for this trait in the Phenovator (for details; see Flood *et al*., 2016). Projected leaf area was similar between these lines (**Figure 1b**). The RIL-population was grown for twenty five days in two experiments and three conditions, 100 µmol m^{-2} s^{-1}, 200 µmol m^{-2} s^{-1} and –N 200 µmol m^{-2} s^{-1}, the latter which includes a growing solution that has only 10% of the nitrogen input compared to the other treatments. ΦPSII and projected leaf area were being monitored after approximately twenty days of growth for up to five days when rosettes matured. A single QTL for ΦPSII (*ΦPSII_c5*) on chromosome 5 is found in the Ler x Col population and found in all three conditions (**Figure 1c, supplementary figure 2**). Epistatic analysis for ΦPSII in all three treatments yields no detection of major additional genetic factors to significantly explain the observed phenotypic variation (**Supplementary figure 2, supplementary table 6**). Furthermore, the *ΦPSII_*c*5* locus is unaffected by any of the environmental conditions as determined by two-way ANOVA, using the single peak marker (m536) and treatment as the genotypic and environmental factors, respectively (**Supplementary table 7**). The QTL for PLA on chromosome 2 locates to the pleiotropic *erecta* locus, which is known to affecting plant physiology, development and growth (van Zanten *et al.*, 2009), and was not investigated further (**Figure 1d**).

3.2 Fine-mapping of the quantitative trait locus on chromosome 5 for ΦPSII

ΦPSII_c5 explains 25-50% of the genotypic contribution to phenotypic variation and is the only significant genetic factor explaining differences in ΦPSII between the studied genotypes under the tested environments (**Figure 1**). Two rounds of fine-mapping experiments were applied to determine the causal genetic variant underlying this quantitative trait locus. Introgression line C5L-A10 and CSL32 (described in Wijnen *et al.* (2018)) were crossed to start an F2:3 fine-mapping scheme (**Supplementary figure 3**). This leaves a heterozygous region of approximately ~10 Mbp with the Col-0 allele in an otherwise isogenic background of the Ler-0 genotype (**Supplementary figure 3**). The resulting F2 progeny of this population was grown in a greenhouse and screened for recombinants within the confidence interval between positions 7909425 and 8842971 on chromosome 5 (relative to CP002688.1, Cheng *et al.* 2016) of the QTL-region as defined by the 100 µmol m^{-2} s^{-1} RIL-mapping experiment (**Figure 2a**, **table 3**). Six genetic markers were used to identify recombinants for this region in ~1800 F2 lines (Two most important markers shown in **Figure 2a**, **supplementary table 2**). A total of 113 F2 lines were found to contain a recombination event within this selected region. For each of these F2 recombinants, eight of their progeny were grown in the Phenovator. In this first fine-mapping experiment, plants were phenotyped for both PLA and ΦPSII since the onset of germination. ΦPSII phenotypes were then corrected for developmental stage

by recalibrating the growth curves (based on PLA) starting at day 1 when 30+ pixels were visible for each plant. All individual plants were genotyped for a total of 17 genotypic markers after conclusion of the experiment (**Supplementary table 2**).

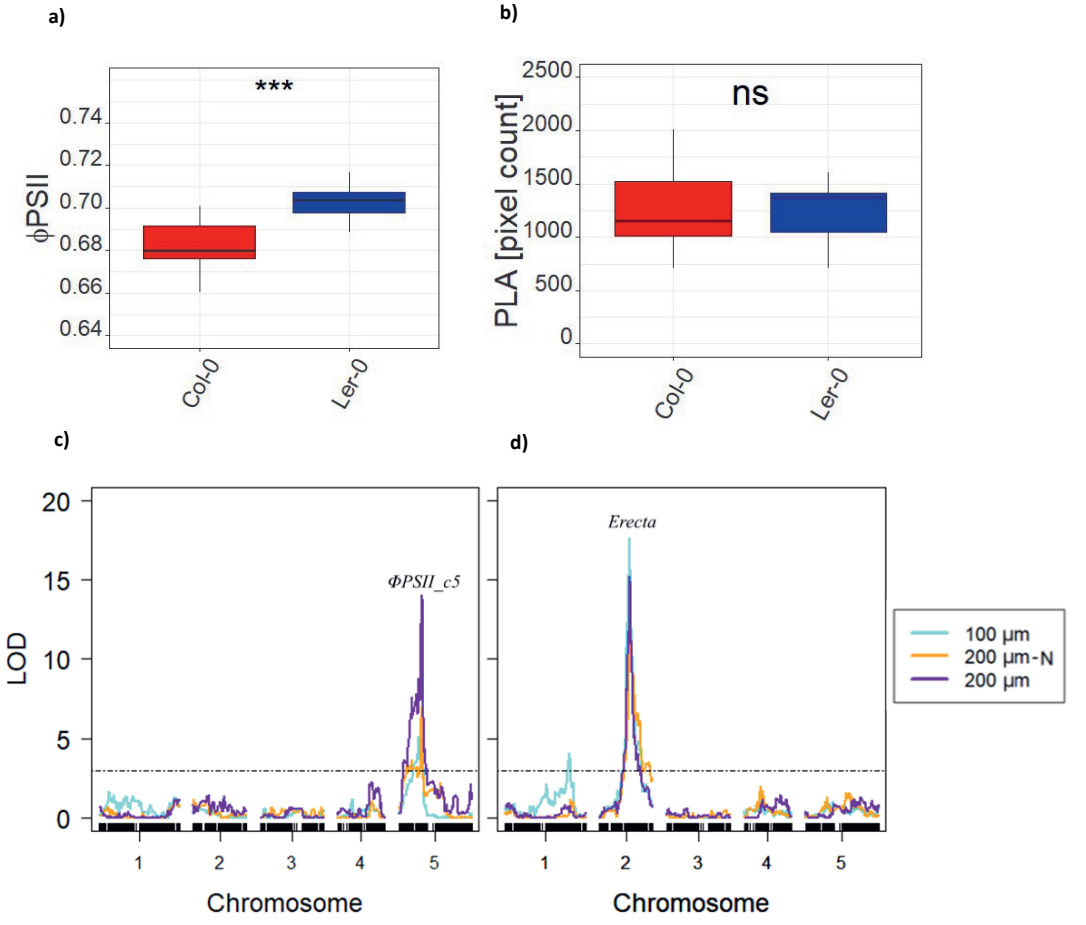


Figure 1. *Projected leaf area (PLA) and photosystem II efficiency (ΦPSII) in the Col-0 and Ler-0 genotypes, and genetic mapping for these two traits in their recombinant inbred line (RIL) progeny. Boxplots for* ***a****) ΦPSII and* ***b****) PLA in the Col-0 and Ler-0 parental lines after 24 days of growth under 100 µmol m-2 s-1 of irradiance. n = 16 for Col-0 and n = 9 for Ler-0. Genetic mapping of* ***c)*** *ΦPSII and of* ***d)*** *PLA under three different environmental conditions, 100 µmol m-2 s-1 irradiance (blue), 200 µmol m-2 s-1 irradiance (purple) and 200 µmol m-2 s-1 irradiance with a nutrient solution with only 10% nitrogen of the concentration of Nitrogen compared to control (orange). Genetic mapping was conducted in 97 lines of the Ler-0 x Col-0 populations using the Best Linear Unbiased Estimates of the genotypes. Genetic mapping was performed on PLA and ΦPSII at 09:00 om the 21st day after sowing. Logarithm of the odds (LOD) is expressed as the -log(p) value. The dotted line represents the LOD-threshold of significance (2.94) for assigning a quantitative trait locus based on an experiment-wide permutation analysis as described in Broman et al. (2003).*

Table 1. *Population statistics of ΦPSII data from three different treatments . 'Exp' =the independent experiments. 'Treatment': 100 µm: plants grown at an irradiance of 100 µmol m^{-2} s^{-1};200 µmol -N: plants grown at 200 µmol m^{-2} s^{-1} and grown in nutrient solution with only 10% nitrogen content compared to control (which introduces shortage); and 200 µm: plants grown at an irradiance of 200 µm m^{-2} s^{-1} in normal growing solution. 'Min': the minimum value of ΦPSII found for any of the lines, 'Max': the maximum value of ΦPSII found for any of the lines (both 'Max' and 'Min' are the Best Linear Unbiased Estimate); 'StdDev': the standard deviation; 'CoV: the Coefficient of variation; H^2: is the broad sense heritability. H^2 was calculated using a linear mixed model and using lme4 to separate the environmental variance from the genotypic variances. Data is presented for Φ_{PSII} at 09:00 on the 21st day after sowing.*

Exp	Treatment	Min	Average ΦPSII	Max	StdDev	CoV	H^2
I	100 µm	0.610	0.680	0.696	0.011	0.016	0.153
II	200 µm -N	0.695	0.723	0.740	0.008	0.011	0.093
II	200 µm	0.720	0.734	0.751	0.008	0.010	0.137

Table 2. *Characteristics of the ΦPSII_c5 locus in the three different treatments. 'Exp': the independent experiments; 'Treatments': '100 µm': plants grown at an irradiance of 100 µmol m^{-2} s^{-1}; '200 µm -N': plants grown at an irradiance of 200 µm m^{-2} s^{-1} and grown in nutrient solution with only 10% nitrogen content compared to control (which introduces shortage); and '200 µm' plants grown at an irradiance of 200 µm m^{-2} s^{-1} in regular growing solution. The 'Position', ' Marker' and 'Chr' (chromosome number) are the coordinates of the QTL, 'LOD' is the logarithm of the odds as –log(p). 'PVE (%)' is the percentage of genotypic variation explained by the peak marker. 'Effect' is the effect size of the quantitative trait locus in expressed as a Φ_{PSII} value. 'Allele' is parental origin of the alleles with higher phenotype. QTLs are based on measurements of Φ_{PSII} made at 09:00 h on the 21st day after sowing.*

Exp	Treatment	Chr	Position (cM)	Marker	LOD	PVE (%)	Effect	Allele
I	100 µm	5	42.962	m563	6.218	24.9	0.0106	Ler-0
II	200 µm -N	5	42.962	m563	6.296	25.2	0.0078	Ler-0
II	200 µm	5	42.962	m563	12.842	44.6	0.0103	Ler-0

The first round of fine-mapping experiment reduced the confidence interval of *ΦPSII_c5* to only 103.7 kbp (**Figure 2a, table 3,** see s**upplementary figure 4** for evaluation of each specific time point). The locus causes a ~2% phenotypic difference between the Ler-0 and Col-0 alleles over the duration of the experiment (**Figure 2b**). The fine-mapped genomic region still contained 23 genes (**Table 3**). To further delimit *ΦPSII_c5*, eight recombinant F2 lines were selected with crossovers between markers at positions 8131781 and 8220636 (**Figure 2a**). The experimental set-up was kept the same, except that for the selected recombinant families a total of eighty replicate progeny were grown in order to maximize statistical power for identification of recombination events that further decrease the region. These lines were genotyped with increased marker density compared to the first screen (**Supplementary table 2**). LOD-scores were similar as for the previous fine-mapping study (**Figure 2c**) and the region was reduced to only 28.9 kbp (**Table 2**).

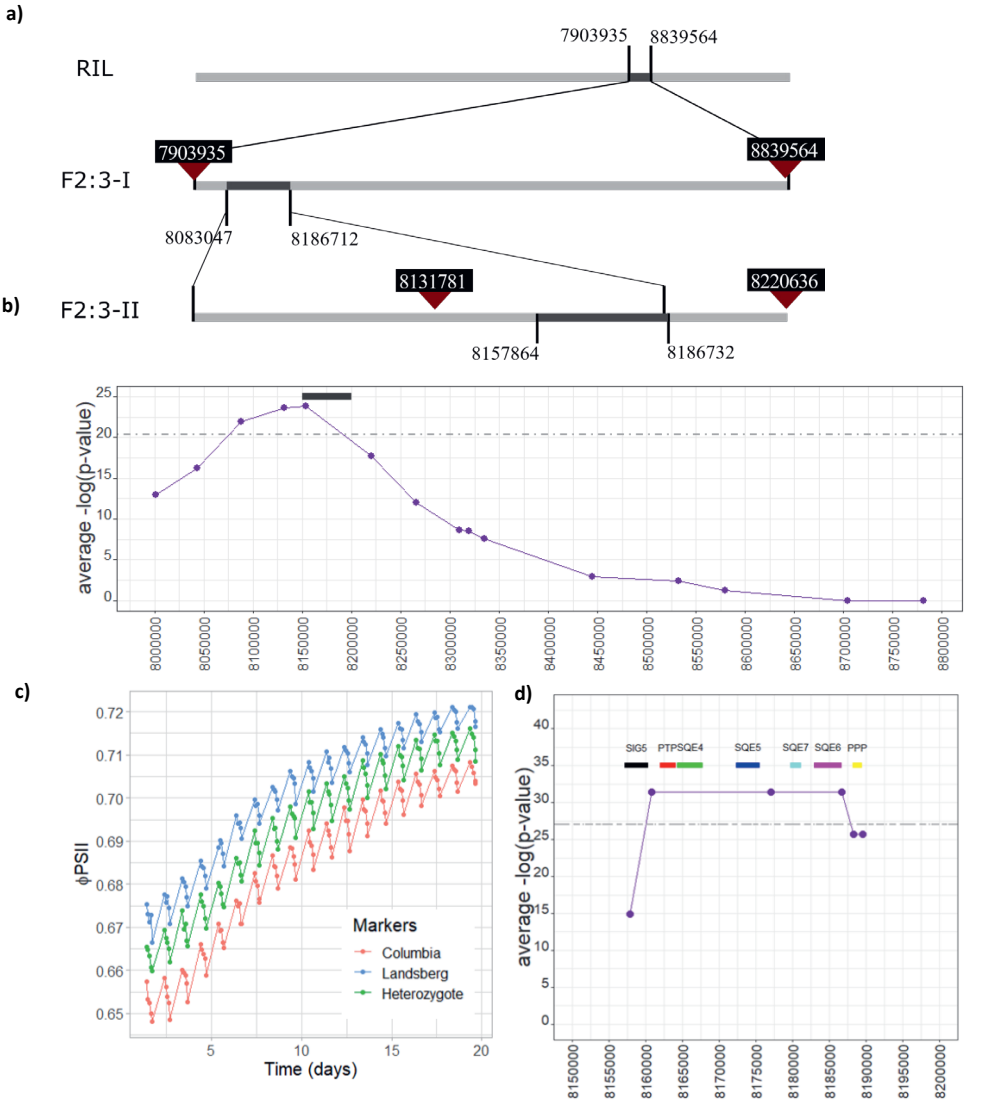


Figure 2. *Fine-mapping of the ΦPSII_c5 locus.* ***a)*** *schematic overview overview of the F2:3 family mapping gains (See* ***supplementary file 3*** *for elaboration of the method). Dark grey indicates the defined finemapped regions with their nucleotide positions indicated by the confidence intervals of the experiments. The black boxes indicate selection markers for recombinants. Image is not scaled.* ***b)*** *Outcomes of the first F2:3 mapping experiment (F2:3–I). Average LOD-value is given for 100 evaluations of the mapping population for ΦPSII over a twenty-day growing period. The x-axis gives the nucleotide positions on chromosome 5.* ***c)*** *Phenotypic plot with daily phenotypes for the first fine-mapping experiment. Each data point is comprised of all finemapping lines scoring Columbia / Col-0, Landsberg (Ler-0) or Heterozygote scores for the top three markers.* ***d)*** *As in* ***b****, but the second round of fine-mapping (F2:3–II) using a select number of eight F2:3 families of which eighty segregating replicates were grown to further delineate the QTL. Only the outcomes of the six peak markers are shown, with no common markers with* ***a****. Dashed grey lines (****b****,* ***d****) represent significance thresholds calculated following the permutation threshold procedure outlined in R/qtl (Broman et al. 2003). The black bar in* ***b*** *covers the full length of the x-axis of figure* ***d*** *as a reference. The gene identities are as follows: SIG5: At5g24120; PTP: At5g24130; SQE4: At5g24140; SQE5: At5g24150; SQE7: At5g24155; SQE6: At5g24160 and PPP: At5g24165.*

Table 3. *Delimitation of the ΦPSII_c5 QTL in the RIL experiment (at an irradiance of 100 µmol $m^{-2}s^{-1}$), the F2:3-I and F2:3-II fine mapping experiments. Start, end and size of the confidence intervals here defined as the genomic region with the inter-marker lines, with nucleotide positions following those from Genbank ref: CP002688.1 (Chr 5 of TAIR10). Reduction is defined by the size of the new interval relative to that determined in the previous mapping experiment. 'No. genes': total number genes in the region. -log(p) range gives the range of observed LOD-values over the duration of the respective the experiments, and represents the values generated from 20-100 datapoints. Although the duration of each experiment differed, in each experiment plants were phenotyped after at least 20 days of growth.*

Experiment	Start	End	Size (bp)	Reduction	No. genes	-Log(p) range
RIL	7909425	8842971	933546	-	238	6.08 - 11.20
F2:3-I	8083047	8186712	103665	11.1%	23	11.25 - 47.09
F2:3-II	8157864	8186732	28868	27.8%	7	12.59 - 46.11

3.3 Development and phenotypic analysis of reciprocal introgression lines

After fine-mapping, reciprocal introgression lines were developed from specific F2:3 family lines (**Figure 3a**). IL-A4 and F8-C3 are introgression lines for Col-0 and Ler-0 alleles in a Ler-0 and Col-0 background, respectively (**Figure 3a**). The lines were phenotyped for ΦPSII under similar conditions as the fine-mapping experiments and subjected to a two-way ANOVA to detect epistasis, including the genotype of the background and introgression as genotypic variables. In this experiment, a significant effect for the genotype of the genomic introgression was found ($F_{1,144.15}$= *53.3603, p < 0.001*), but not for genetic background ($F_{1,144.15}$= *1.1686, p = 0.28197*) (**Figure 3b**). However, a significant interaction effect between the genotype of the locus and the genetic background effect of the introgression lines was found ($F_{1,144.15}$= *5.2815, p = 0.02337*). This suggests that the Ler-0 allele exhibits a 30% larger increase of the phenotype in a Ler-0 genomic background than the Col-0 background (**Figure 3b**).

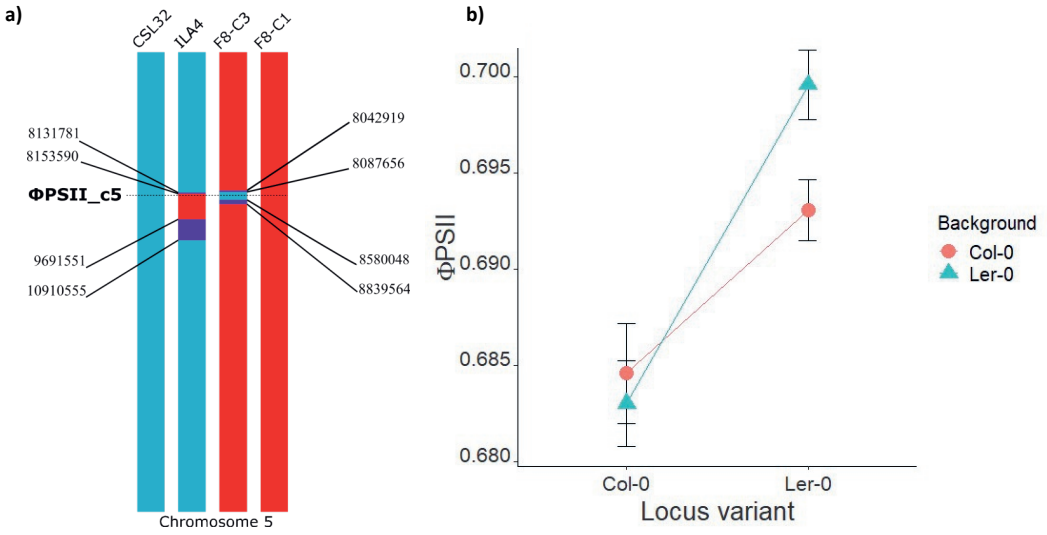


Figure 3. *Structure and phenotypic analysis of reciprocal introgression lines carrying variants of the ΦPSII_c5 locus.* ***a)*** *Genetic structure of reciprocal introgression lines (shown for chromosome 5 only). Blue, highlights Ler-0 genomic sequence; red, highlights Col-0 genomic sequence; and purple, indicates a region of uncertainty inbetween two recombinant markers (given as positions – see* ***supplementary table 2****). CSL32 is effectively Ler-0, but with a Col-0 cytoplasm. IL-A4 carries a Col-0 genomic introgression of at least 1.37 Mbp in a CSL32 background. F8-C1 is effectively Col-0, but having gone through a phase of haploidization. F8-C3 carries an introgression of at least 0.44 Mbp in an F8-C1 background. The position of marker c5_8177051, one of the molecular markers associated with enhanced photosynthesis during the fine-mapping experiments, is indicated in the figure.* ***b)*** *Phenotypic analysis of the reciprocal introgression lines. The locus variant indicates the genotype of the interaction, while background indicates the remainder of the genome. The ΦPSII average is calculated from a time series of measurements containing 24 data-points and is supported by 30+ biological replicates per line.*

3.4 Allelic variation of genes in the fine-mapped region

Based on the maximum LOD-scores from the second F2:3 mapping experiment, we defined a QTL region of 28.9 kbp in size, specifically between the start codon of At5g24120:(C5_pos_8159744) and the start codon of At5g24165:(C5_pos_8188622), to be our region of interest to contain causal variation and candidate genes (**Figure 2d, supplementary table 8**). This fine-mapped region encompasses five complete genes (At5g24130-At5g24160), and the promoter regions of two other genes (At5g24120, At5g24165) that locate partially within the QTL region (**Figure 2d**). In order to identify genomic variations between Ler-0 and Col-0, *de novo* genomic sequence data (Ler-0*;* Genbank ref: CM004363.1, Col-0; Genbank ref: CP002688.1) was analyzed and used to generate a dot plot to visualize larger genomic variations in the QTL-region. The region of interest contains various deletions, duplications, inversions and local re-arrangements, the majority which relate to *SQUALENE EPOXIDASE 4* (*SQE4*, At5g24140)*, SQE5* (At5g24150), *SQE7* (At5g24155) and *SQE6* (At5g24160) genes (**Figure 4a**). All these genes belong to a tandemly repeated family of sub-class II *SQUALENE EPOXIDASE* genes (Laranjeira *et al.*, 2015). Genetic variation between Col-0 and Ler-0 for the other genes (i.e. *SIG5*; At5g24120, *PTP;* At5g24130 and *PPP*; At5g24165) consists primarily of nucleotide changes resulting in synonymous amino acid

substitutions and polymorphisms in promoter regions, suggesting that only gene expression difference may underlie genetic variation for those genes.

The *SQE* genes encode proteins annotated as *SQUALENE EPOXIDASE*s (Rasbery *et al.*, 2007; Laranjeira *et al*., 2015). *SQE1*, *SQE2* and *SQE3* can functionally complement the yeast *Saccharomyces cerevisiae SQUALENE EPOXIDASE*-deficient *erg1* mutant, but *SQE4*, *SQE5*, *SQE7* and *SQE6* cannot and are thus referred to as *SQE-like* genes (Rasbery *et al*., 2007; Laranjeira *et al*., 2015). To detail specific and potentially impactful mutations within the *SQE-like* genes, the following described polymorphisms are determined relative to the Col-0 genome. The Ler-0 allele for *SQE4* is truncated due to missing an essential first exon (**Figure 4b),** which assumes this allele to be non-functional. For *SQE5*, the region of the 3' UTR termination region in Ler-0 is markedly different to that of the Col-0 allelic version (**Supplementary figure 5**). The cDNA sequences of both alleles of *SQE5* are relatively similar, but polymorphisms still cause a total of five amino acid substitutions in the resulting protein (200 F->L, 304 F->L, 379 E->Q, 426 C->S, 500 R->S, see **Figure 5**). *SQE7* (at5g24155) in Col-0 is previously described as a truncated member of the sub-class II *SQUALENE EPOXIDASE* family by Laranjeira *et al.* (2015), missing five out of the eight exons (**Figure 4b, Figure 5**). The genomic sequence of Ler-0 covering the *SQE7* gene suggests the presence of an intact version of *SQE7* in this genotype (**Figure 4b**). To confirm this, we sequenced the Ler-0 *SQE7* cDNA, aligned the sequence to the *de novo* genome assembly and recoded the cDNA sequence to amino acids proving this suggestion (**Figure 5)**. For the predicted amino-acid sequence of *SQE6*, only two amino acid substitutions are present between the Col-0 and Ler-0 alleles (3F->S, 72H->N, see **Figure 5**). In addition, at a distance of 316 bp upstream from the translation start side of the gene, the Col-0 promoter region of *SQE6* contains a larger 160 bp deletion and several smaller ones (**Supplementary figure 6**).

3.5 T-DNA-insertion mutant and RNA expression analysis of candidate genes

For four out of seven genes under consideration, homozygous T-DNA insertion knockout-lines were identified, *SIG5* (2x), *PTP* (1x), *SQE4* (2x) and *SQE6* (2x) (**Supplementary table 9**), but none of these showed deviating ΦPSII phenotypes compared to the Col-0 version of the locus under 100 m^{-2} s^{-1} (**Figure 6a**). To characterize the impact of genetic variation in the promoter region, 21-day old rosettes of CSL32 (Ler-0 allele) and IL-A4 (Col-0 allele) grown during the T-DNA line experiment were sampled for transcriptional analysis. Sampling took place at two hours and nine hours after the start of the photoperiod to anticipate the possibility of genetic variation impacting known diurnal expression patterns expressed by *SIG5* (Noordalley *et al.*, 2013). *PTP* and *SQE4* are not expressed in the rosettes (**Figure 6b, c**). For the other genes, there were no significant differences between the Col-0 and Ler-0 alleles, with the exception of *SQE6* for which the Col-0 allele had a significantly lower expression level than the Ler-0 allele (**Figure 6b, c**).

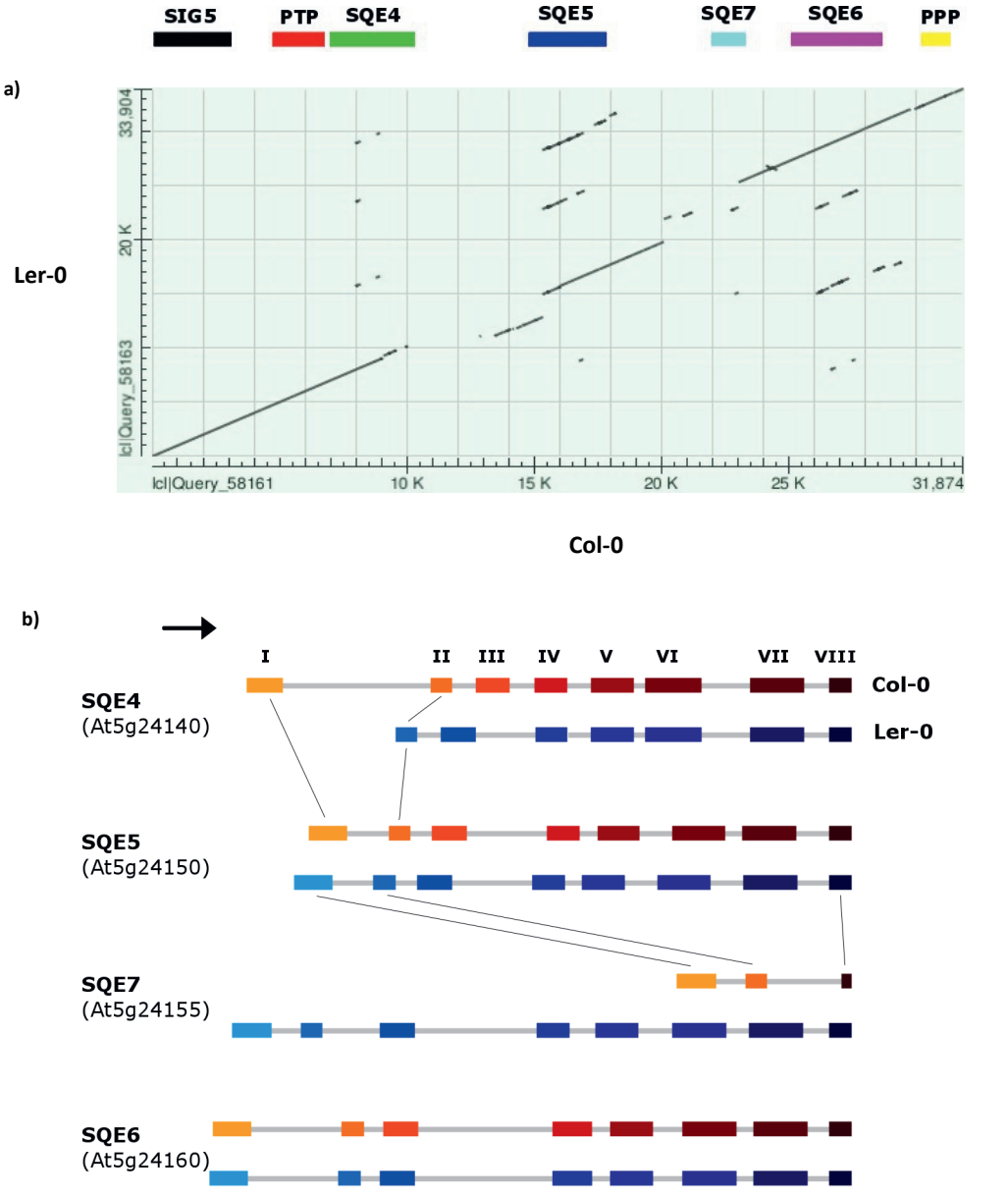


Figure 4. *DNA sequence variation in genes underlying the ΦPSII_c5 QTL.* ***a)*** *Dotplot alignment of the 8159744-8188622 region on chromosome 5, which is the identified region after fine-mapping, using Col-0 as a reference against Ler-0 (positions follow TAIR10 annotation).* ***b)*** *Layout of the introns (grey lines) and exons (coloured blocks) of SQE-like genes in Col-0 (red) and Ler-0 (blue). The black arrow indicates the direction of transcription. Roman numbers I-VIII and the colour gradient indicate the eight exons required for functional SQE-like genes. The sizes of introns and exons are scaled to match their relative sizes. The black lines are merely visual aids to correctly identify the exon ID among these highly similar genes.*

Figure 5. *Alignment of the SQE5, SQE7 and SQE6 amino acid sequences in SEAVIEW following the MUSCLE algorithm (Edgar, 2004; Guoy et al., 2010). The amino acid sequence for SQE7 was derived from cDNA sequencing of the full transcript PCR, followed by alignment of this sequence to the Col-0 chromosome 5 de novo sequence (NCBI Reference sequence: CM004363.1).*

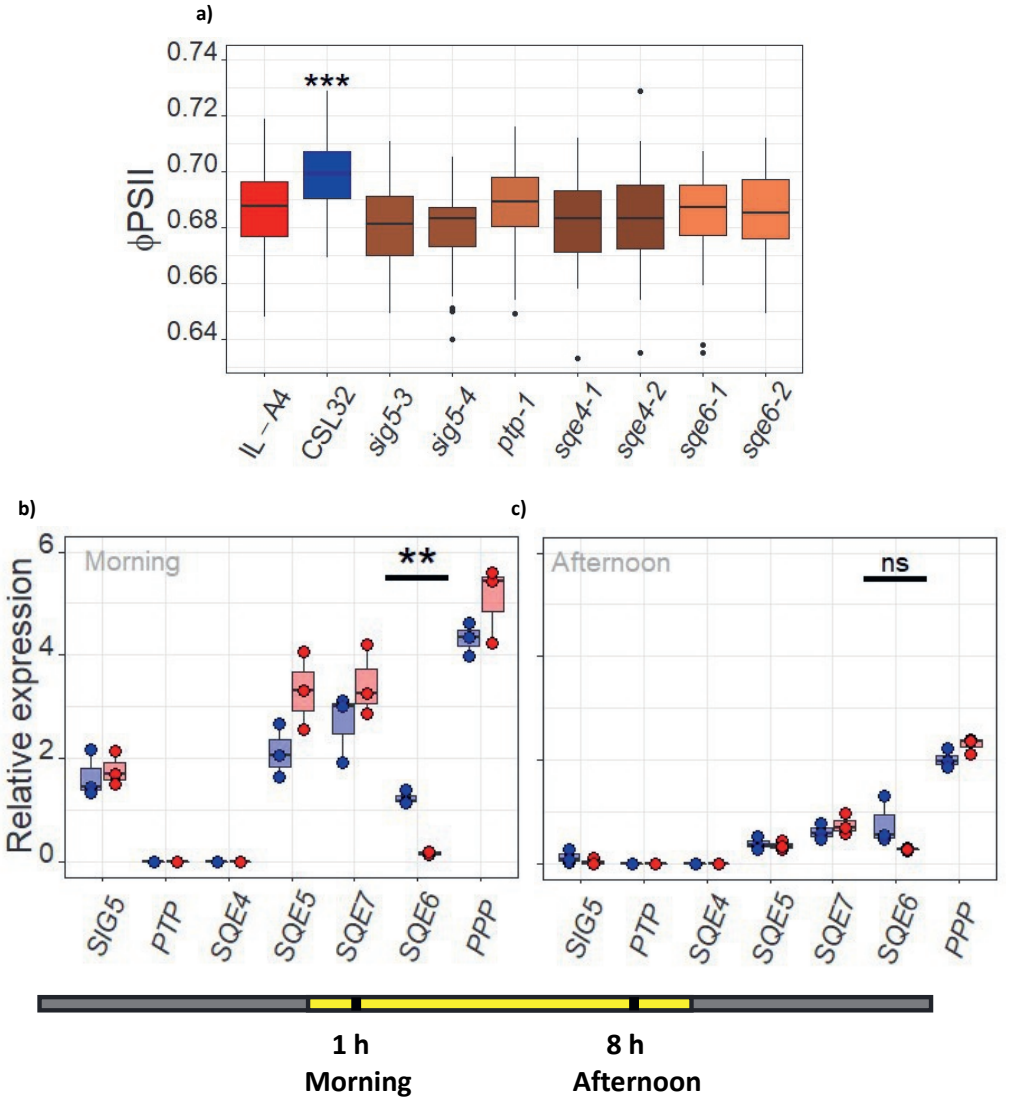


Figure 6. *Functional analysis of candidate genes underlying ΦPSII_c5 under 100 µmol m^{-2} s^{-1} of irradiance.* ***a)*** *Evaluation of ΦPSII in ILA-4 (red, Col-0 allele for ΦPSII_c5), CSL32 (blue, Ler-0 allele for ΦPSII_c5) and seven independent T-DNA knockout lines (brown, in Col-0 background) on day 25 of growth after sowing, results shown for 9:00 h (1 hour after onset of light). N = 48-60 per genotype.* ***b)*** *Transcription analysis of seven candidate genes between the CSL32 and IL-A4 lines. Samples were taken at 09:00 h in the morning (one hour after the onset of the photoperiod). For each gene, the results of three pools consisting of three independent replicates per genotype are shown.* ***c)*** *as* ***b****, but samples were taken in the afternoon at 16:00 h (eight hours after the onset of the photoperiod).Unless indicated with symbols, no significant differences were found between the control and T-DNA insertion lines or between allele specific expression patterns. Significant differences are indicated at ** p = < 0.01, *** = p < 0.001, ns = not significant. The black and yellow bar indicates the photoperiod and the time since the onset of lighting for which the samples have been taken.*

3.6 Transgenic complementation of IL-A4 and copy number manipulation of SQE-like genes

The T-DNA mutant and RNA expression analyses did not result in evidence leading to identification of a causal gene underlying *ΦPSII_c5*. The *PTP* and *SQE4* genes did not show a phenotype upon disruption of the gene and are also not expressed in the shoots (**Figure 6b, c,** Klepikova *et al.*, 2016). As *SQE4* is assumed to be non-functional in Ler-0 (**Figure 4b**), we expected a deviant phenotype from the T-DNA mutant in Col-0 background compared to the Col-0 functional alleles for the gene to be causal but this we did not find (**Figure 6a**). Based on these considerations we omitted these two genes from further consideration. In addition, no gene expression differences were found for *SIG5* and *PPP*. These two genes were also considered to be unlikely causal genes as only their promoters locate in the QTL region and thus we expected a difference in expression (**Figure 2d**). The high amino acid sequence similarity between *SQE5*, *SQE7* and *SQE6* (**Figure 5**) and the high allelic diversity between the parental lines (**Figure 4b; supplementary figure 5-6**) implies that copy number variation of these genes drives this locus. To further explore the causality of genetic variation, the different alleles belonging to *SIG5, SQE5*, *SQE7*, *SQE6* and *PPP* were transformed into IL-A4 as complementing the phenotype back to CSL32 indicates a causal relationship be. To explore a role for copy number variation in the *SQE-like* genes, a transgenic overexpression line in the IL-A4 background was generated for the *SQE6* cDNA from the Ler-0 allele, which should result in an effective increase of the number of active copies of *SQE-like* genes in Col-0 background. An amiRNA gene (*sqe576*) silencing vector was developed for post-transcriptional silencing of *SQE5*, *SQE6* and *SQE7*, to reduce activity of *SQEs*. The transgene was transformed into the CSL32 background to assess the impact of such a reduction on its phenotype.

Unfortunately, for allelic transgene of *SIG5* for the Col-0 allele (*SIG5[C]*), no transformants were obtained after repeated floral dipping experiments and it was thus excluded. For all other transgenic lines, T2 seeds were sown in the Phenovator (**Figure 6a**) and phenotyped for Φ_{PSII}. For each independent T1 line, four homozygous T2 seeds, characterized by higher intensity of seed fluorescence (Shimada *et al*., 2010), were chosen for 24-25 independent transformants per allele, leading to a total of 96-100 plants per allelic complementation transformant. We chose this approach as it would have been too tedious to validate all these transformants while at the same time ensuring all lines to have reached T2 stage to reach homozygosity in phenotyped progeny (**Figure 6a**). Analysis of allelic transgenic insertions revealed that a considerable – but variable – number of copies are present in the transgenic lines (**Supplementary figure 7**). Furthermore, two independent transformant lines of the gene silencing *sqe576* and overexpression *35S::SQE6[L]* transformants – each of which had the functionality of the transgenic alleles verified in the T1 – were grown in 32 homozygous replicates.

Introgression line IL-A4 transformed with additional gene copies of *SIG5* or *PPP* did not recover the ΦPSII phenotype of CSL32, which verifies that these genes are not causal to *ΦPSII_c5* (**Figure 7b**). However, transgenic complementation with *SQE5[C]*, *SQE6[L]* and *SQE7[L]* did (**Figure 7b**), while their allelic counterparts *SQE5[L]*, *SQE6[C]* and *SQE7[C]* do not. This proves a causal role for these *Squalene-like* genes in photosynthesis and verifies that allelic variation in these genes contributes to phenotypic difference between Col-0 and Ler-0. These results further underscore the deleterious genetic variation

earlier described in these genes (**Figure 4b**, **5**, **Supplementary file 5, 6**), and the causality of these genes to explain *ΦPSII_c5*. Gene expression analysis in T2 lines harvested during this experiment verified the intended functionality of the *35S::SQE6[L]* and *sqe576* lines in the T2 (**Figure 7c-f**). Gene transcription of *sqe576* lines shows that SQE5 is unaffected by activity of the amiRNA, but transcription of *SQE7* and *SQE6* is reduced to ~50% and ~30% compared to levels observed in the WT background CSL32. *35S::SQE6[L]* showed 70-96x higher transcription values of *SQE6* compared to the WT background IL-A4. In line with these observations, *sqe576*-lines show a ΦPSII phenotype that is at an equal level to the IL-A4 line (**Figure 7g**) and both independent *35S::SQE6[L]* overexpression transgenes in the IL-A4 background showed elevated levels of Φ_{PSII}, even exceeding that of CSL32 (**Figure 7h)**. Together, these results prove that natural variation and transgenic alteration of the effective dosage of *SQE-like* genes has a profound effect on Φ_{PSII} in Arabidopsis.

4. DISCUSSION

In this work we applied a multi-environmental QTL-mapping approach to study the genetic architecture of photosynthesis under different environmental conditions in a Ler-0 x Col-0 RIL-population. Only a single quantitative trait locus – *ΦPSII_c5* – was found to underly variation of Φ_{PSII} in all the treatments we used. Photosynthesis takes place in the chloroplast, in which only be accessed by proteins that can enter membranes by carrying a specific chloroplast transit peptide (ctp). Two out of seven genes underlying this QTL, namely *SIGMA 5* (*SIG5*) and *PUTATIVE PLASTID PROTEIN* (*PPP*), contain such a sequence, but our experiments did not indicate their involvement in the QTL. Instead, we discovered an unexpected causal relationship between photosynthesis and copy number variation of *SQUALENE EPOXIDASE-like* gene members which do not encode a ctp (Laranjeira *et al.*, 2015). Nonetheless, our work highlights a role for leaf-expressed *SQE-like* genes in photosynthesis.

4.1 The Ler-0 x Col-0 RIL population expresses a simple genetic architecture of Φ_{PSII} for plants grown under stable light conditions.

In order to carefully characterize the nature of *ΦPSII_c5*, we analyzed its interaction with the environment, genetic background, interaction with productivity, and plant development. Doing so may support the evaluation of specific hypotheses about the potential involvement of candidate genes underlying QTLs (El-Soda *et al.*, 2014; 2015; Lachowiec *et al.*, 2015; Knoch *et al.*, 2020). Using a Ler-0 x Col-0 population, we tested three different environments but found no evidence for GxE effects (**Figure 1b, Supplementary table 6**). Furthermore, we found no QTL for PLA to co-locate with *ΦPSII_c5* (**Figure 1c**). Instead a QTL for PLA located on chromosome 2, this QTL is driven by the well-known *Erecta* locus of which the variant in Ler-0 is known to cause extreme pleiotropic effects in many traits (van Zanten *et al.*, 2009) and may have masked a potential growth benefit associated with increased photosynthesis in *ΦPSII_c5*. We did not, however, find interactions with this *Erecta* locus in our genome-wide epistatic analysis which

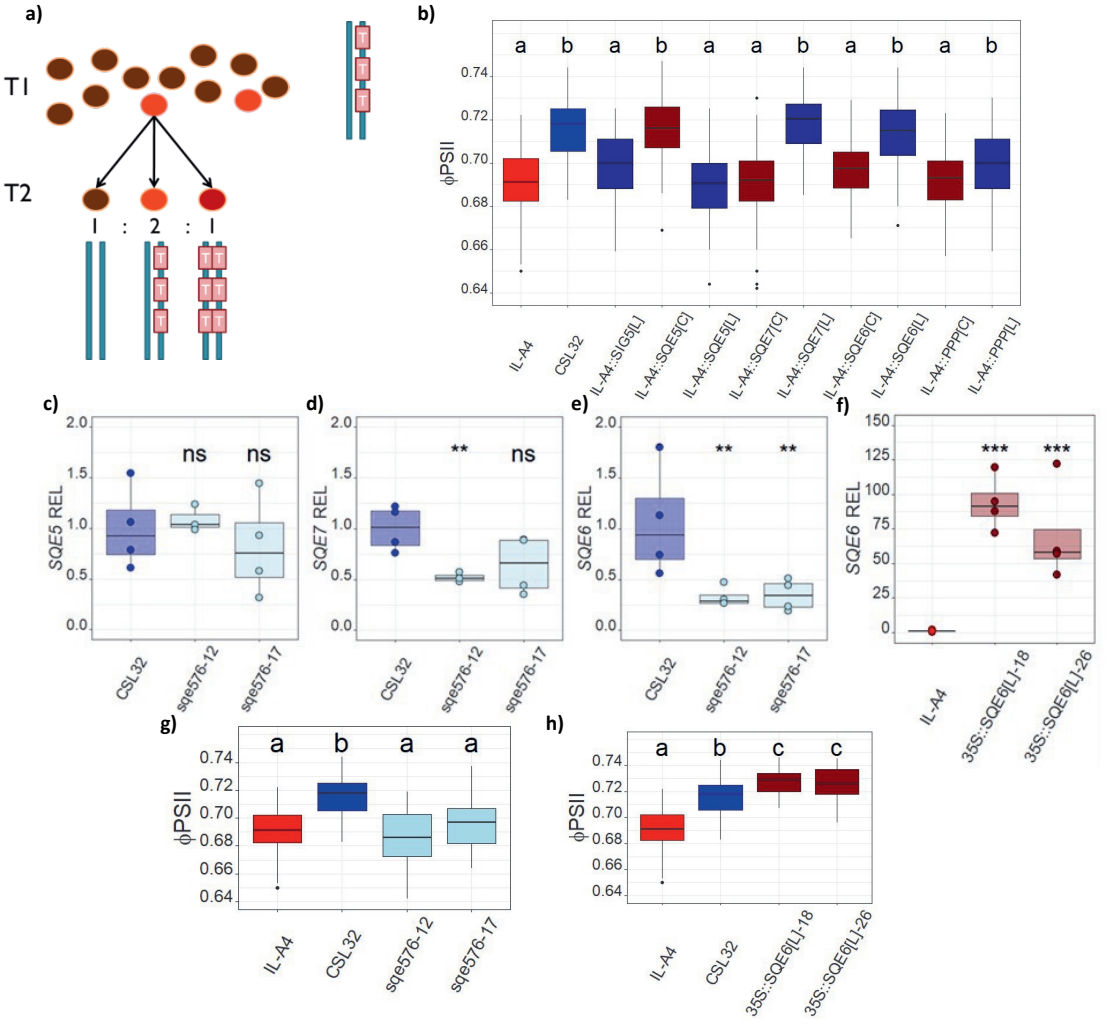


Figure 7. *Evaluation and validation of transgenic T2 lines under 100 umol m^{-2} s^{-1} of light irradiance.* ***a)*** *Selection strategy to obtain the number of active transgenic insertions for growing T2 seeds. The most intensely red fluorescing seeds (dark red) were picked from the T2 segregating offspring and assumed to be homozygous for the inserts (following Shimada et al., 2010).* ***b)*** *Φ_{PSII} values for transgenic allelic complementation lines measured at 09:00 h on day 21 of growth after sowing. IL-A4 and CSL32 are the control lines, representing respectively Col-0 (red) and Ler-0 (blue) alleles for the genes underlying the QTL (n= 92 for WT lines, n=96-100 from 24-25 independent transformants per construct). Dark red and blue indicate the parental origin of the alleles. Abundance of the SQE5* ***(c)****, SQE7* ***(d)****, and SQE6* ***(e)*** *transcripts in two independent sqe576 SQE-like silencing RNAi-lines.* ***f)*** *Transcript abundance of SQE6 in two independent SQE6 overexpression lines. Relative expression levels (REL) in* ***c-f*** *are normalized against the expression of the same gene in the WT control line (CSL32, resp IL-A4) (n=3-4 individual biological replicates).* ***g)*** *Evaluation of Φ_{PSII} in two independent sqe576 (n=32) against the indicated WT lines (n=96).*. ***h)*** *as in* ***g****, but for two independent, homozygous overexpression lines (n=32) against the indicated WT lines (n=96). For* ***c-f****, ns = no significant differences, * = p <0.05, ** = p <0.01, *** p <0.001 in student's t-test against the reference genotypes for log-normalized expression data. For* ***b****,* ***g****,* ***h****, letters indicate equal groups following Tukey post-hoc test after ANOVA.*

suggests that it does not in any way affect Φ_{PSII} (**Supplementary table 5**). Previously identified genetic variation between the chloroplast genomes of Col-0 and Ler-0 also does not appear to have an effect on *ΦPSII_c5* (Flood *et al.* 2020), which confirms our observation that the RILs, which are developed in the Ler-0 cytoplasmic background do not map differently than the fine-mapping population and introgression lines which are in the Col-0 background. The heritability estimates (<0.16) for each of the conditions indicate that the genotypic contribution to variation in ΦPSII is very low, (**Table 1**). These values are surprisingly similar to other estimates of heritability obtained using diversity panels screened for photosynthetic traits (such as ΦPSII) with the same phenotyping system (van Rooijen *et al.*, 2015; Prinzenberg *et al.*, 2020). Intensive phenotyping during our first round of fine-mapping allowed us to conclude that there is no developmental effect on ΦPSII (**Figure 2c**). Taking all observations together, we conclude that *ΦPSII_c5* is a low effect size QTL that is unresponsive to environmental conditions, is independent of plant development, and is not dependent on genetic variants in the nuclear or cytoplasmic genomes (but see **figure 3b**).

4.2 Advantages of a bi-parental mapping population and use of transgenic complementation for the discovery of novel traits and genetic variation for photosynthesis

The role of *ΦPSII_c5* as the sole genetic factor significantly contributing to genetic variation for Φ_{PSII} in the RIL is surprising given the plethora of QTLs typically found for this trait in Arabidopsis in both diversity panels (van Rooijen *et al.*, 2015; Prinzenberg *et al.*, 2020) and RIL-populations (Oakley *et al.*, 2018; Flood, 2016), and also because photosynthesis is generally acknowledged to be a highly complex trait (Flood *et al.*, 2011; van Bezouw *et al.*, 2019). Two rounds of fine-mapping in populations derived from cross with a near isogenic line allowed us to reliably narrow down the identity of *ΦPSII_c5* to just seven possible gene candidates (**Figure 2b, d**). By analysis of *de novo* genomic sequences of only two genotypes, Col-0 and Ler-0, we were able to examine and predict the impact of genetic variation on functionality and causality of these genes (**Figure 4b, 5, supplementary figure 5-6**), even though T-DNA knock-out lines and RNA expression analysis provide no conclusive evidence for which gene or genes exhibits variation that explains the *ΦPSII_c5* (**Figure 6a-c**). Although our RNA expression and T-DNA knockout line experiments reasonably excluded *SIG5, Polypyrimidine tractbinding-like protein (PTP), SQE4* and *PPP* as potential causal genes, definite proof for causality of the genetic polymorphisms in *SQE5*, *SQE7* and *SQE6* to contribute to variation in photosynthesis required a transgenic complementation experiment.

Transgenic complementation is a powerful method to prove the involvement of specific genetic variation. This is particularly useful for genes for which T-DNA-lines give a phenotype due to the presence of potential null alleles already present in the background genome (Bentsink *et al.*, 2006; Huang *et al.*, 2012; Tang *et al.*, 2018). However, plant transformation often results in the integration of multiple copies of the transgene in the recipient genome of independent transformants. This may affect the stability and functioning of these inserts and thus the validity of resulting phenotypes of the transformants (Buck *et*

al., 2009; Glowacka *et al*., 2018). Furthermore, high dosages of transgenic inserted natural alleles may not accurately reflect natural genetic variation as an increase in the number of copies of a gene may have a different impact than can be expected from natural copy numbers. In our complementation lines we found varying, but high numbers, of transgene copies, but no significant differences in the transgene copy number for between groups of independent transformants (S**upplementary figure 7).** Thus, we could reliably identify functional allelic variation that links to *ΦPSII_c5*, regardless of the exact number of copies in each independent line and without the need to consider alternative hypotheses to have caused the phenotypes.

In our experiments we transformed introgression line ILA4 – which contains an introgression including the Col-0 version of the *ΦPSII_c5* locus (**Figure 3a**) thus exhibiting lower ΦPSII compared to the Ler-0 version (**Figure 3b**) – with whole allele, overexpression or gene silencing transgenes to identify which gene would elevate ΦPSII to the level of CSL32 (**Figure 3b**) – carrying the Ler-0 version of *ΦPSII_c5*. We showed that alleles of *SQE5[C]*, *SQE7[L]* and *SQE6[L]*-alleles successfully complemented IL-A4 to the CSL32 phenotype (**Figure 7b**), while the alternative alleles of these genes (i.e. *SQE5[L]*, *SQE7[C]*. *SQE6[C]*) and the other tested genes failed to do so (**Figure 7b**). Our extensive genomic analyses provided insight as to why. The Ler-0 version of *SQE5* (*SQE5[L]*) may have reduced functionality, possibly due to low translation efficiency, as the 3'-UTR region is very different from *SQE5[C]* (**Supplementary figure 5**), which showed higher ΦPSII. The 3'-UTR region is known to affect the translation efficiency of a protein from RNA (Mayr, 2019). This gives a plausible explanation as no expression differences were found between for alleles of this gene (**Figure 6b, c**). *SQE7[C]* is truncated (**Figure 4b**, **Figure 5**), losing 5 out of 8 exons compared to Ler-0, and lacks most of its predicted catalytic domains as described in (Laranjeira *et al.*, 2015). The removal of these catalytic domains directly explains the failure of this allele to complement ILA4. *SQE6[C]* is expressed at a significantly lower level han *SQE6[L]* (**Figure 6b, c**), most likely as a result of a deletion in the promoter region (**Supplementary figure 6**). As we harvested full rosettes for RNA extraction we cannot conclude whether the expression of *SQE6[C]* is generally lower or restricted only in certain tissues. The latter is expected due to the high number of additionally introduced copies of these genes among our transgenes following floral dip (**Supplementary figure 7**), which should compensate for reduced transcriptional activity in the natural allele compared to CSL32 (**Figure 6b, c**). However, promoter reporter lines would need to be developed in order to verify this hypothesis. The overexpression and amiRNA-lines support our findings that the dose of *SQE-likes* influences ΦPSII (**Figure 7g, h**), with overexpression lines even expressing a superior phenotype to that of CSL32. All our outcomes suggest that the range of copy numbers where an increase in *SQE-like* results in an increase in the photosynthesis phenotype operates between two limits (**Figure 8**), a base level defined by the WT and *sqe576*-lines, and a saturation level defined by the overexpression line. However, the exact impact of this gene family on biochemical pathways requires further investigation to biologically explain this relationship.

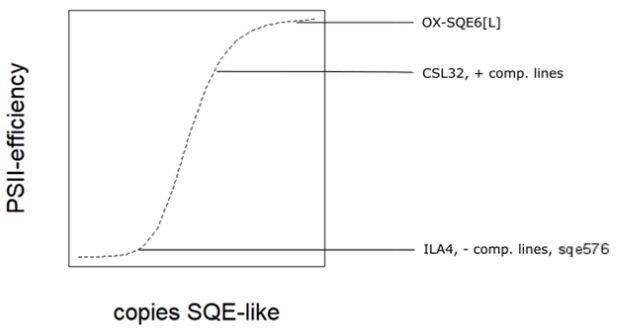


Figure 8. *Hypothesized relationship between activity of SQE-like genes and photosynthesis. An upper and lower horizontal asymptote are visible, within which SQE-like genes have a dose-response effect on photosynthesis. '+ comp. lines' denote the SQE-like genes for which the alleles complemented the CSL32 with a Col-0 introgression of the QTL (IL-A4). The '– comp.' lines denote those SQE-like genes for which complementation of the IL-A4 negative introgression line failed. 'sqe576' is the rnai line that suppresses the expression of SQE5, 6, and 7 in CSL32, while 'OX-SQE6[L]' is the overexpression line in the IL-A4 background line.*

4.3 How do SQE-like genes impact photosynthesis?

Sterol biosynthesis is a primary plant process that is initiated in two pathways, the chloroplastic methylerythritol phosphate pathway and the cytosolic mevalonate pathway. Farnesyl pyrophosphate is the final product of both pathways, and this in turn is converted to squalene by *SQUALENE SYNTHASE*. *SQUALENE EPOXIDASEs (SQE)* then catalyze the epoxidation of squalene to form (S)-2,3-epoxysqualene, a precursor from which a broad diversity of sterols is produced in downstream metabolism (Kalra *et al.* 2015; Valitova *et al.*, 2016).

SQEs exist in all living organisms and are essential because they catalyze a vital step in the formation of sterols, which are vital for cell functioning, signaling, development and growth (Piironen, 2000; Sonawane *et al.*, 2015; Valitova *et al.*, 2016). While species generally encode only a single copy of *SQE* in the genome, in the plant kingdom many species have extra copies – including those targeted to the chloroplasts (Rasbery *et al.*, 2007; Laranjeira *et al.*, 2015; Liu *et al.*, 2019; Manzoor *et al.*, 2021). *SQE-like* genes form a sub-clade of *SQUALENE EPOXIDASE*s, a gene sub-family which is exclusive to Brassicaceae (Laranjeira *et al.*, 2015). A major difference between *SQE4-7* (*SQE-likes*) and *SQE1-3* (true *SQEs*) genes is the presence of a flavin adenine di-nucleotide (FAD) dependent oxido-reductase domain, comparable to a nicotinamide adenine dinucleotide NAD(P)-binding Rossman-like domain, suggesting functionally different roles (Laranjeira *et al.*, 2015). Both clades contain FAD/NAD(P) binding domains that extend to most of the posterior region of the predicted genes. Unlike true *SQE*s , *SQE-likes* are unable to complement *SQE* deficient *erg1* mutant yeast (Rasbery *et al.*, 2007), a functional diversification between these two groups is likely to have occurred. A relationship between *SQE* or *SQE-like* genes and photosynthesis has so far not been shown. A knock-out mutant for *SQE1* in *Arabidopsis thaliana*, *dry2/sqe1-5*, had impaired growth, accumulated squalene, and had impaired stomatal closure (Rasbery *et al.*, 2007; Posé *et al.*, 2009), but apart from squalene accumulation *sqe3* mutants do not show these

phenotypes (Laranjeira *et al.,* 2015). Micro-array data of the *dry2/sqe1-5* mutant reveal that photosynthesis genes are down regulated, while stress-related genes are upregulated. These results imply that disruption(s) of *SQEs* directly or indirectly affects plant photosynthesis, but that functional redundancy exists among the members of this gene family. In this work we have established a dosage dependent relationship between *SQE-like* genes and photosynthesis, but the functional relationship has not yet been established. However, the mutants we developed in this work provide excellent starting material with which to further study these relationships.

4.4 Genetic variation in the SQE-like tandem reveals insight in Brassicaceae evolution

In the Brassicaceae multiple *SQE-like* genes are present in an interspecific conserved syntenic block signifying a copy number expansion event (Laranjeira *et al*., 2015). *SQE4* may be the first copy of *SQE-likes*. It displays root-restricted expression in Arabidopsis, similar to *SQE1-3* (Rasbery *et al*., 2007; Laranjeira *et al.*, 2015; Klepikova *et al.*, 2016) and the first diversification occurred in a species with only one orthologue of *SQE4* (Laranjeira *et al*., 2015). Phylogenetic clustering by Laranjeira *et al*., (2015) supports the idea that *SQE5* was the next formed *SQE-like*, followed by *SQE7* and then *SQE6*. The ~30kbp genomic region after finemapping in our experiment exactly matches a previously described conserved syntenic region in species within the Brassicaceae as identified by Laranjeira *et al.* (2015). Their alignment of this region suggests that related species all contain the truncated version of *SQE7* rather than the full-length gene. A more thorough analysis of the polymorphisms we describe here is required before we can be certain of nature and evolution of these genes within Arabidopsis genotypes and other species within the Brassicaceae clade.

4.5 SQUALENE EPOXIDASE-like genes as a target for improving crop photosynthesis

Established core proteins in the photosynthesis machinery, such as RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE-OXYGENASE, SEDUHEPTULOSE BISPHOSPHATASE and PHOTOSYSTEM SUBUNIT S, are popular targets for transgenic modification to enhance photosynthesis and subsequent productivity in a wide variety of (crop) species (Kromdijk *et al.*, 2016; Driever *et al.*, 2017; South *et al.*, 2019; Yoona *et al.*, 2020). Although the exact functionality of *SQE-like* genes is unknown, we found that in overexpression lines ΦPSII was higher than found with natural Ler-0 (CSL32) alleles (**Figure 7c, d**). However, ΦPSII is a relative measure of the light-use efficiency of photosystem II which is at its maximum in the dark-adapted state when it is normally referred to as Fv/Fm, which in plants reaches a maximum of around 0.83 (Baker, 2008). It is also linearly related to gross carbon dioxide fixation, a major factor in plant productivity. Thus changes in ΦPSII are expected to be paralleled by changes in assimilation. Although it is difficult to point towards the metabolic step that is catalysed by the *SQE-like* gene product, the potential beneficial effects on photosynthesis may not be restricted to *Arabidopsis* – or even the Brassicaceae. The use of exogenic photosynthesis genes has previously led to improved photorespiration properties in Tobacco and biomass growth (South *et al.* 2019), which is an attractive

application of the discoveries we report here. Improving photosynthesis traits may not always result in improved above-ground growth (Garcia-Molina & Leister, 2020), and normally growth is only judged above ground. The relationship between photosynthesis and plant growth is complex and variable, depending on factors such as leaf area, leaf area per unit leaf dry weight, and the leaf fraction of plant biomass, none of which are necessarily constant, thus Arabidopsis may be a poor model to study the link between photosynthesis and productivity. Introduction of *SQE-like* genes driven by a universally active promoter – such as the 35S – in other species such as crops or algae might yield potential benefits in terms of increased photosynthesis, potentially useful metabolites, and maybe even biomass production. However, before engaging in *SQE-like* mediated crop improvement, the potential molecular pathways in which *SQE-like* genes are involved in remains to be elucidated.

5. CONCLUSION

In this study, we used a recombinant inbred line population to explore natural variation in photosynthesis efficiency between Ler-0 and Col-0. We mapped this to an uncharacterized gene family of *SQE-like* genes and provided evidence of how natural genetic variation can affect photosynthesis. So far, the biological role of this gene family is unknown. Overexpression of these genes revealed that ΦPSII – and perhaps other photosynthesis parameters – can be increased to above that found in the high performance parent. *SQE-like* genes are not available in most of the clades in plant kingdom, which may suggest that other species may be enhanced by introduction of such genes. To conclude, we have shown that natural variation for photosynthesis is a powerful resource for improving photosynthesis that may reveal unexpected options for improving photosynthesis efficiency in crop species.

6. ACKNOWLEDGEMENTS

We would like to thank Gerrit Stunnenberg, David Brink, Taede Stoker for caring for our plants. Furthermore, we would like to thank Suzanne Vlinderveld, Tom Theeuwen and René Boesten for having participated in sowing the mapping and fine-mapping experiments. We would also like to thank Maarten Koornneef, Tom Theeuwen and René Boesten for discussions and helpful suggestions for interpretation and analysis of this work. In addition, we acknowledge Raul Wijfjes for his help on making the *de novo* genomes accessible for use and analysis. We would also like to thank Maarten Wassenaar for making sure the Phenovator repeatedly lived to phenotype another day. And last, the first author would like to thank René Boesten specifically for inspiring the title of this chapter. This research was generously funded by a grant from Rijk Zwaan B.V.

SUPPLEMENTARY FILES

Supplementary table 1. Primers used to confirm TDNA insertions. LP = left border primer, RP = right border primer.

Locus	Primer sequences	
Salk_LBb1.3	attttgccgatttcggaac	
SAIL_LB1	gccttttcagaaatggataaatagccttgcttcc	
	LP	**RP**
SALK_101921	ccattctctagtgtcagccac	gtttgagatgggaagacctcc
SALK_141383	tctcatacccgcttgacaaag	gttcagctgcaagatctccac
SALK_133729	tctcaagctcggatgctttac	aagctagctcacttgtgctgc
SALK_137191	ggatggtcttctcagcaacag	ggatggtcttctcagcaacag
SALK_130439	ggatggtcttctcagcaacag	aaattccactagccccagttg
SALK_012094	tgatgaagttcgttgtgttgc	tgtggcgatatagttccttcg
SALK_012094	tctagcggatgatttttggtg	tttcctctcaacaaatgtcgg
SALK_040805	tggttgttattcgaaccttcg	gtaattgaagcagccttgtcg
SALK_024504	ttttgcggaactgaaacaaac	gtcagtagcaccgtcttctcg
GK-151G06	actcgacgcccgtactatgt	agtggcgtcacagtttcaca
SALK_083343	cccttaattgccaattaaaagg	ctgaggatcaaaaagacggtg
SALK_118625	tgtgcaagagttccataactcc	tgtgtttagaacttttccccg
SALK_053722	cgacgtaaaatatcgtttcgc	ttctcatcttcagcttcctatgg
SALK_069263	tcatagcgttagacgccgtac	cgatttctttgtcggagacag

Supplementary table 2. The markers and their positions as used for finemapping in this study. In bold the markers defining the QTL for the first (I) fine-mapping experiment. The underlined markers mark the location of the quantitative trait locus according to the second (II) fine-mapping experiment.

Position	Polymorphism		Selection I	Fine-mapping I	Selection II	Fine-mapping II
	Col-0	Ler-0				
6776258	T	G	X			
7903935	A	G	X	X		X
8000694	T	C	X	X		
8042919	T	C		X		
8087656	T	C		X		
8131781	G	T		X	X	
8153590	A	G		X	X	
8157027	G	T			X	X
8160764	T	C			X	X
8177051	G	A			X	X
8186643	T	G			X	X
8188246	C	T			X	X
8189521	A	T			X	X
8220636	T	A		X	X	X
8265957	T	G		X		
8309822	C	A		X		
8319308	G	C	X	X		
8335443	A	G		X		
8445073	T	A		X		
8532284	T	C		X		
8580048	A	G		X		
8638564	C	T		X		
8704447	G	A		X		
8781505	T	C		X		
8839564	T	A	X	X		X
9691551	A	G	X			

Supplementary table 3. qRT-PCR Primers used in the gene expression experiment.

qRT PCR	Gene	Primer name	Primer sequence
Candidate	At5g24120	At120_F	ctcctagagtcctgtcgcc
	At5g24120	At120_R	cgctctcgacgatgtgctc
	At5g24130	At130_F	gtggattgaagcttcttgggg
	At5g24130	At130_R	caacgttcagcaatcctccc
	At5g24140	At140_F	gcagctgcgtttgttggaatc
	At5g24140	At140_R	gcctcagaattcttgcctggg
	At5g24150	At150_F	gtatgacagaacctccgcattg
	At5g24150	At150_R	caggcgaagaaacaacagcc
	At5g24155	At155_F	cgtcctcctggctgcataac
	At5g24155	At155_R	gcatattctcttgctaaggacggg
	At5g24160	At160_F	gatacgcggctgcatttgc
	At5g24160	At160_R	gcctcaagctttttggtttggc
	At5g24165	At165_F	ctatctcatctggttctggaccg
	At5g24165	At165_R	ctatctcatctggttcttcg
Reference	AT2G35635	UBQ_7_F	ccctcctgttcaacaaaggctc
	AT2G35635	UBQ_7_R	acagagcctccctctatcgc
	At5g53560	CB5E_F	gtcttgttgtcctcaacagggaa
	At5g53560	CB5E_R	tccatcatgtcccttgcagtg
	AT1G75780	OTUB1_F	cacaagatcgtgctgaagtcg
	AT1G75780	OTUB1_R	gttgctcaaggaacaacgca

Supplementary table 4. Primers used for molecular cloning

Molecular cloning	Primer name	Primer sequence
pKGW_RedSeed::*SIG5*$^{(C/L)}$	At120_F1	ctcgggctattcttttgatttaaatttaagtggttgtgataatggctctgc
	At120_R1	atacggtcaatgacaacttcaaacttcac
	At120_F2	gaagttgtcattgaccgtatgatgg
	At120_R2	atatggtcgacctgcactttacgtggctttgtgcatgg
pKGW_RedSeed::SQE5$^{(C/L)}$	At150_C_F1	ctcgggctattcttttgatttatgtatagagagggacgagactatcact
	At150_L_F1	ctcgggctattcttttgatttaattcggttctggttcgggtca
	At150_R1	cgccttcatgctgacttggacagtgt
	At150_F2	gtccaagtcagcatgaaggcgagtag
	At150_R2	atatggtcgacctgcagaacttatcgttgtaggtatgagaaacgt
pKGW_RedSeed::SQE7$^{(C/L)}$	At155_F1	ctcgggctattcttttgatttattttatctgcataaccctccctaaagg
	At155_R1	ctcttgctaaggtttgtaaagtggc
	At155_F2	ctttacaaaccttagcaagagaatatgcgagag
	At155_R2	atatggtcgacctgcacacgaatctatgtgattgtgcatttgg
pKGW_RedSeed::SQE6$^{(C/L)}$	At160_F1	ctcgggctattcttttgatttatgcaatataaggatcccatgcttctaga
	At160_R1	acatacaagaataaagaaggcgaagaaacaac
	At160_F2	gccttctttattcttgtatgtcactcctt
	At160_R2	atatggtcgacctgcatcagcttcctatggattcttcaagtaattg
pKGW_RedSeed::PPP$^{(C/L)}$	GA165_F	ctcgggctattcttttgatttaagatagttgtggaaagaggagtgaca
	GA165_R	atatggtcgacctgcaacctgatcgtgtgagtcgga
pFASTR02::SQE6^{L}	OX160_F	cgtgcgtaaaagccat
	OX160_R	ttattagataacggttgcagcca*
pFASTR02::RNAi_*sqe576*	RNAI_1	cggtgttctagcgttaccaaca
	I	gaTGTTGGTAAGGTAGAACACCGtctctcttttgtattcc
	II	gaCGGTGTTCTACCTTACCAACAtcaaagagaatcaatga
	III	gaCGATGTTCTACCTAACCAACTtcacaggtcgtgatatg
	IV	gaAGTTGGTTAGGTAGAACATCGtctacatatatattcct
	A	ctgcaaggcgattaagttgggtaac
	B	gcggataacaatttcacacaggaaacag
	amiF	caccgtcgacggtatcgataagct
	amiR	catggcgatgccttaaataaaga

**The reverse primer for the SQE6 had to be modified at the nucleotide level compared to the genomic sequenceto enable proper annealing for TOPO-cloning purposes. These modifications do not impact the amino acid sequence*

Supplementary table 5. Primers used for copy number quantification of transgenic complementation lines

Primer name	Primer sequence
CB5E_FWD	tgatcatcctggaggcgatg
CB5E_REV	ttgcagtgtcgctgtgacca
PSD1-1_FWD	gcggtcctggctagcaatgatg
PSD1-1_REV	gtagccctgcgccttcgttatg
DS3_FWD	ggttcgcatggaaggaacggtc
DS3_REV	aagcaaatggcaaaggtccccc
DS4_FWD	ccaccgggtgtaacaagacgga
DS4_REV	ttggaggtagtgccgttgggaa

Supplementary table 6. Epistatic analysis of ΦPSII efficiency under three different conditions. Chr x Chr represents the combination of chromosomes in which the largest interaction effect was found. PSII_ll, PSII_n & PSII_c represent three treatments; Low light (100 µm m^{-2} s^{-1} of light), Nitrogen deficiency (Nitrogen at 10% of control), Control (200 µm m^{-2} s^{-1}). All numbers represent the –log(p) values of the genetic interactions between genetic loci. The treatment-wide average calculated LOD-threshold for interactions is 3.9, which none of the interactions meet.

Chr x Chr	PSII_ll	PSII_n	PSII_c
c1:c1	0.000	0.001	0.418
c1:c2	0.484	0.065	1.540
c1:c3	0.275	2.337	1.744
c1:c4	1.210	1.367	0.973
c1:c5	0.022	0.066	0.263
c2:c2	0.620	0.949	0.301
c2:c3	0.818	1.205	0.832
c2:c4	1.950	0.756	1.521
c2:c5	0.160	0.024	0.022
c3:c3	0.783	0.609	0.254
c3:c4	0.326	0.590	0.494
c3:c5	1.430	0.239	0.097
c4:c4	0.253	1.639	0.019
c4:c5	1.070	0.608	1.160
c5:c5	0.000	1.782	0.795

Supplementary table 7. Two-way ANOVA to explore the Genotype x Environment component of *ΦPSII_c5*. Factor, Treatment has three levels (100 µmol, 200 µmol –N and 200 µmol) and marker M536 has two levels (Col-0 and Ler-0). Marker M536 is linked to *ΦPSII_c5*.

Factor	SumSq	Df	F-value	Pr(>F)	
Treatment	0.04053	2	493.6	2.20E-16	***
M536	0.001438	1	35.025	6.85E-08	***
Treatment x M536	0.000002	2	0.0228	0.9775	NS
Residuals	0.003449	84			

Supplementary table 8. All genes in the finemapped region. The data was generated through the *plant ensemble* bioMart tool (consulted in March 2019).

Gene locus	Gene Start*	Gene End*	Abbrevation†	Gene name†	Gene ontology
AT5G24120	8157151	8160343	*SIG5*	RNA polymerase sigma factor *SIG5*, chloroplastic/mitochondrial [Source:UniProtKB/Swiss-Prot;Acc:Q9ZNX9]	cellular response to blue light plastid sigma factor activity cellular response to salt stress response to far red light response to red light embryo sac development positive regulation of transcription, DNA-templated regulation of transcription, DNA-templated DNA binding chloroplast plastid mitochondrion transcription, DNA-templated DNA-binding transcription factor activity sigma factor activity DNA-templated transcription, initiation protein binding chloroplast organization regulation of RNA biosynthetic process photosystem II assembly
AT5G24130	8161928	8164066	*PTP*	Poly-pyrimidine tractbinding like protein [Source:UniProtKB/TrEMBL;Acc:Q9FL61]	biological_process molecular_function membrane integral component of membrane
AT5G24140	8164319	8167776	*SQE4*	*SQUALENE EPOXIDASE* 4 [Source:TAIR;Acc:AT5G24140]	sterol biosynthetic process membrane integral component of membrane monooxygenase activity oxidoreductase activity oxidation-reduction process flavin adenine dinucleotide binding squalene monooxygenase activity

					endoplasmic reticulum
AT5G24150	8172300	8175525	*SQE5*	*SQUALENE EPOXIDASE* 5 [Source:UniProtKB/Swiss-Prot;Acc:O65404]	sterol biosynthetic process membrane integral component of membrane oxidoreductase activity oxidation-reduction process flavin adenine dinucleotide binding squalene monooxygenase activity endoplasmic reticulum
AT5G24155	8179703	8181107	*SQE7*	*SQUALENE EPOXIDASE* 7 [Source:UniProtKB/TrEMBL;Acc:Q6ID26]	sterol biosynthetic process oxidation-reduction process FAD binding squalene monooxygenase activity endoplasmic reticulum
AT5G24160	8182917	8186704	*SQE6*	*SQUALENE EPOXIDASE* 6 [Source:UniProtKB/Swiss-Prot;Acc:O65402]	sterol biosynthetic process membrane integral component of membrane monooxygenase activity flavin adenine dinucleotide binding oxidation-reduction process squalene monooxygenase activity
AT5G24165	8188166	8189470	*PPP*	Putative plastid protein[Source:UniProtKB/TrEMBL;Acc:Q8LDQ8]	biological_process

*positions relative to the TAIR10 reference genome assemble

†gene nomenclature sensu Laranjeira *et al.* (2015). If no gene abbreviation convention was established, they were formulated ad-hoc.

Supplementary table 9. List of T-DNA lines used in this experiment. Names are determined ad-hoc unless the respective line was previously characterized.

Locus name[a]	T-DNA lines	Name	location	Knockout?[b]
At5g24120	SALK_101921	*sig5-4*	Exon	Yes[c]
	SALK_141383	*sig5-3*	Exon	Yes[d]
At5g24130	SALK_137191	*ptp-1*	Exon	Yes
	SALK_130439	-	Exon	No
At5g24140	SALK_012094	*sqe4-1*	Exon	Yes
	SALK_040805	*sqe4-2*	Exon	Yes
At5g24150	SALK_024504	-	3'-UTR[†]	Expression still detectable[†]
	GK-151G06	-	Exon	No TDNA present
At5g24155	-	-	-	-
At5g24160	SALK_083343	*sqe6-1*	Intron	Yes[±]
	SALK_112366	-	Exon	No TDNA present[±]
	SALK_118625	*sqe6-2*	Exon	Yes[±]
At5g24165	SALK_053722	-	Promoter	No TDNA present
	SALK_069263	-	5'UTR	No TDNA present

[a]Following TAIR10.

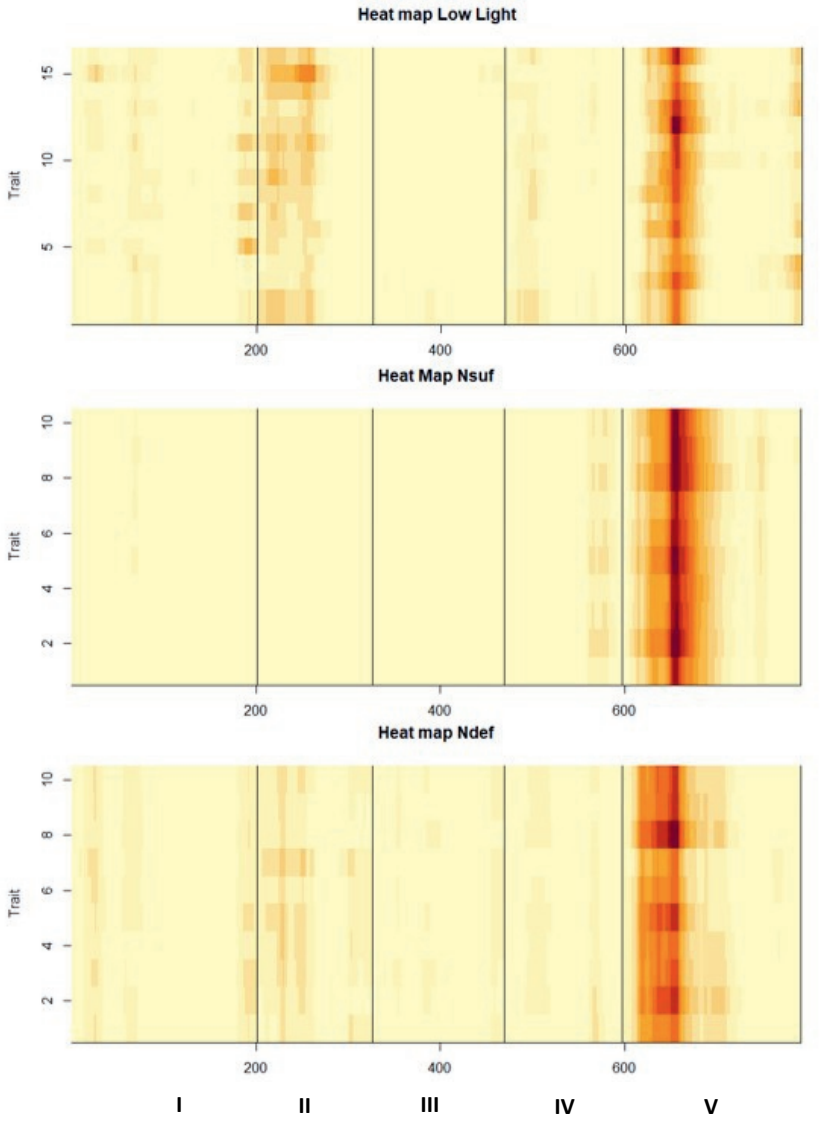
[b]TDNA's located between the translation start and end sites were assumed to be introduce a proper gene knockouts. TDNA insertions locating outside the coding region were experimentally verified to be or not a knock-out.

[c, d]experimentally verified by Zhao *et al.* (2017), Noordalley *et al.* (2013) respectively.

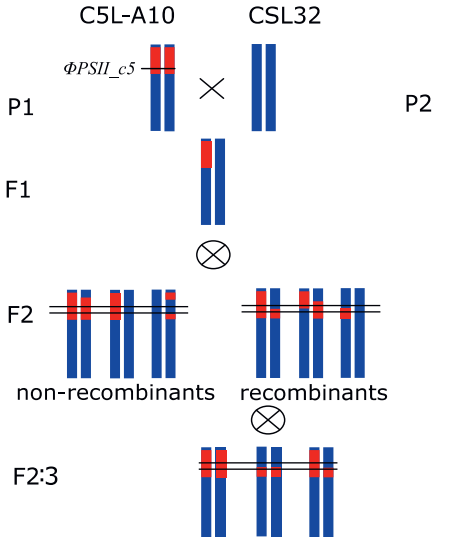
[†]Expression verified by qPCR. http://www.ruf.rice.edu/~bartel/projects/genesofinterest.html lists the location of the TDNA as 429 bp into the 5'UTR region, but this is incorrect. The 5'-UTR of At5g24150 is 130 bp in size (sensu TAIR10 https://www.*Arabidopsis*.org/servlets/TairObject?type=locus&name=AT5G24150, 12-03-2021), which means that the TDNA locates in the promoter region and functional copies can be produced.

[±]Consistent with http://www.ruf.rice.edu/~bartel/projects/genesofinterest.html

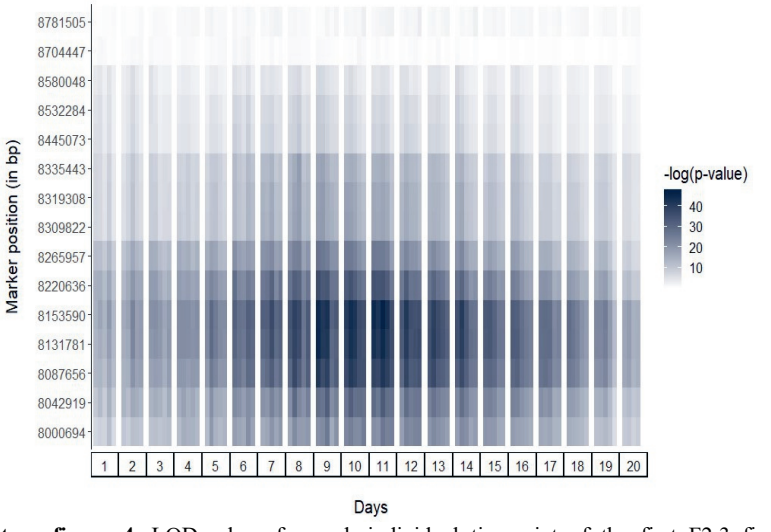
Supplementary figure 1. The artificial micro-RNA sequence and recognition site of the *sqe576* vector. The first ~100 bp (starting from ATG) of the genomic coding sequence of the *Ler-0* accession are given and aligned against the 21 bp long micro-RNA sequence (*sqe576*) for *SQE5*, *SQE7* and *SQE6*. *sqe576* binds to *SQE5*, *SQE7* and *SQE6* mrna at 3, 0 and 1 mismatches, respectively.



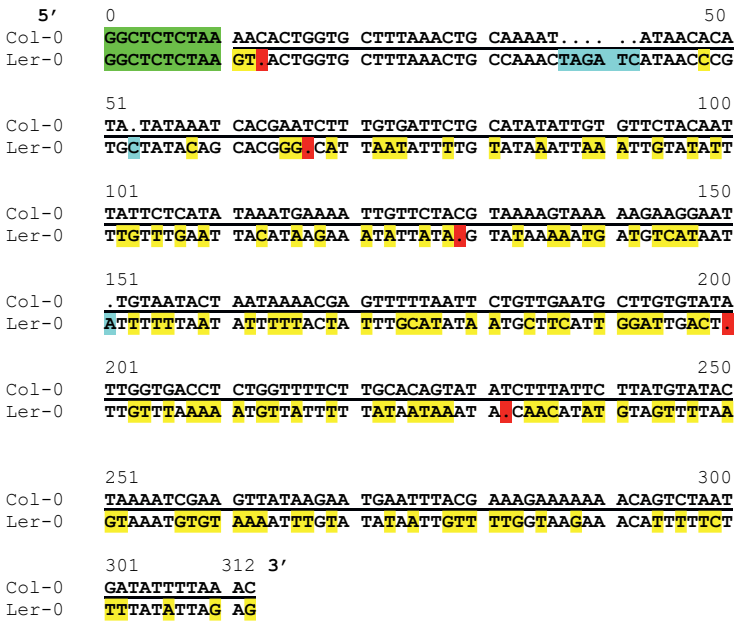
Supplementary figure 2. QTL mapping of time series for PSII efficiency of plants aged 19-21 days after sowing. Trait is given as the time point (5 points/day for the low light, 4 points/day for the others). During the Ndef (10% nitrogen supply, 200 µmol m^{-2} s^{-1} growth irradiance) and Nsuf experiments (200 µmol $^{-2\ s-1}$ growth irradiance) growth irradiance PSII efficiency was evaluated four times a day, while for the low light (200 µmol $^{-2\ s-1}$ growth irradiance) experiment this was evaluated five times a day. Marker position is given as a cumulative number from the first marker on chromosome 1 to the last marker on chromosome 5. Colour intensity indicates the LOD-value, with red values indicating higher LOD-scores.



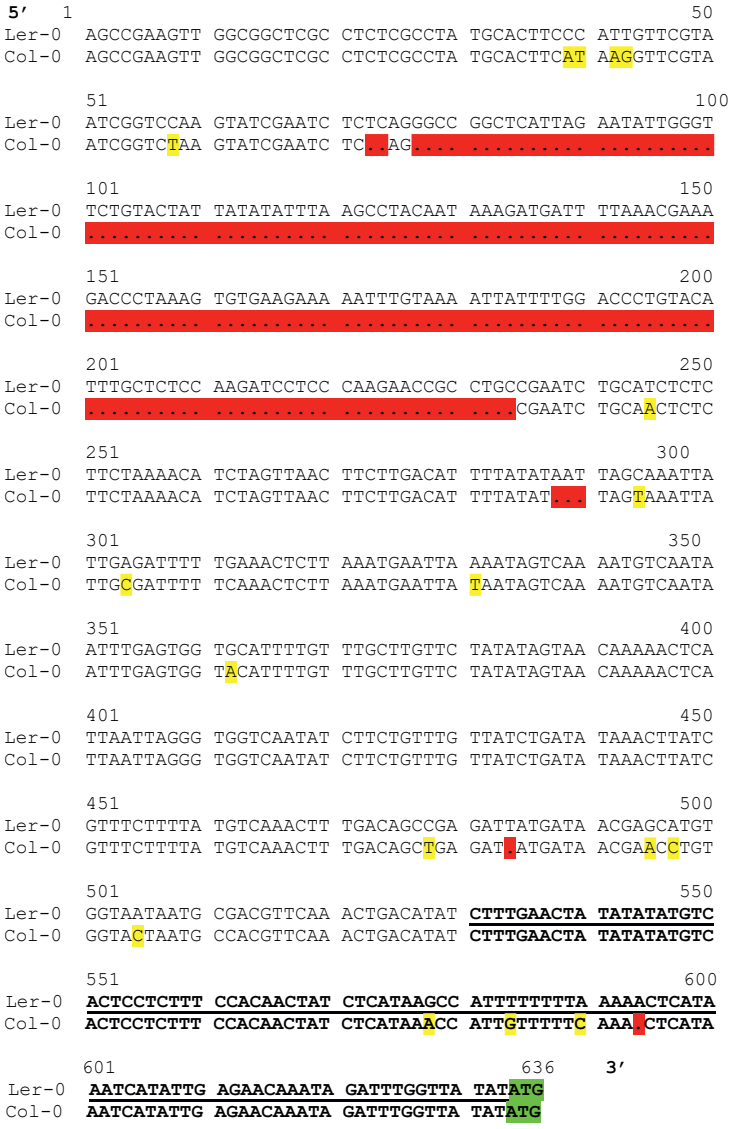
Supplementary figure 3. Schematic overview of the F2:3 family mapping approach targeting *ΦPSII_c5*. A cross between CSL32 - an isogenic Ler-0 line, but with a Col-0 cytoplasmic genome – and introgression line C5L-A10 – as CSL32 but with a Col-0 introgression spanning ~10 Mbp (Wijnen *et al*., 2018). The location of the *ΦPSII_c5* is annotated in the introgression line. Selfing of this F1 hybrid generates recombinant and non-recombinant lines in the F2, with recombinants defined as those lines that now exhibit crossovers between the borders (black solid lines) of the confidence of the *ΦPSII_c5* QTL. Selfing of these F2 lines generate F2:3 progeny that segregate for their identified cross-over sites, thus generating F2:3 families.



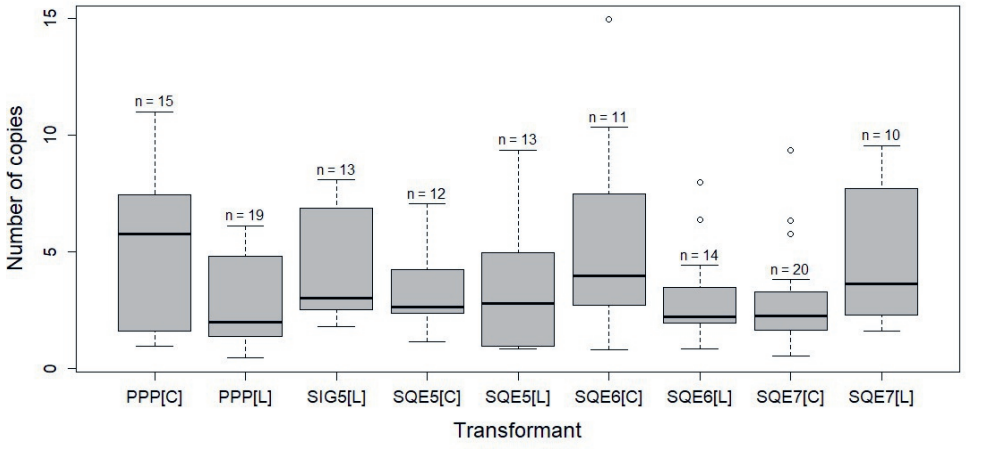
Supplementary figure 4. LOD-values for each individual timepoint of the first F2:3 fine-mapping experiment. The colour intensity indicates the LOD-values (as –log(p-value)). Days indicate the day of growth from which the average rosette size of each of the replicates reached 30 pixels, and thus represent the developmental stage of the plants.



Supplementary figure 5. Terminator region alignment of *SQUALENE EPOXIDASE 5* (*SQE5*, At5g24150) of Ler-0 and Col-0 alleles for the first 340 basepairs in beyond the stop codon. Green indicates the translation termination site of SQE5 (TAA) and several bp upstream into the coding region. Yellow indicates single nucleotide polymorphisms, red indicates deletions and cyan indicates insertions (all relative to Col-0). The bold, underscored sequence denotes the entire 3'-UTR region length of the Col-0 allele of the gene (NCBI Reference sequence; NM_122320.4; Tabata *et al*. 2000). The sequencing alignment was made in MULTALIN (Corpet, 1988).



Supplementary figure 6. Promoter alignment of *SQUALENE EPOXIDASE* 6 (At5g24160) of Ler-0 and Col-0 alleles for the first 633 base pairs in front of the ATG translation start site. Green indicates the translation start site, yellow indicates single nucleotide polymorphisms and red indicates deletions (all relative to Ler-0). The bold sequence denotes the 5'-UTR region of the gene (NCBI Reference sequence; NM_122321.5, Tabata *et al*., 2000). The reverse complement sequence is shown. Sequencing alignment was made in MULTALIN (Corpet, 1988).



Supplementary figure 7. Copy number estimation in sets of allelic transformants grouped per allelic construct. No significant differences were found among the number of copies detected in each of the lines, following ANOVA and Tukey post-hoc test. The number of independent lines per transformation vector tested is given above each boxplot. These particular lines were not phenotyped in the Phenovator, but independent lines originating from the same batches were and these results are thus representative for all lines tested.

Chapter 4. Genetic analysis of photosynthesis traits in response to high light irradiance in two *Arabidopsis thaliana* accessions

Roel F.H.M. van Bezouw[1], Roxanne van Rooijen[1], Stefan Meijs[1], Marieke van der Hoeven[1], Chi-Wei Pan[1], Yvet Boele[1], Tom Theeuwen[1], Bastiaan de Snoo[2], Jeremy Harbinson[3], Mark G.M. Aarts[1].

SUMMARY

The use of high-throughput phenotyping facilities has come increasingly more popular to assess and understand the genetic architecture of photosynthesis traits in plants. In this chapter, we used several genetic approaches to describe natural variation in the photosynthetic response to high light irradiance in an F2 population derived *Arabidopsis thaliana* accessions contrasting for this trait. Two novel quantitative trait loci (QTL) are discovered and described to have a profound impact on photosynthesis efficiency following a stepwise increase of irradiance. From the F2 progeny, further genotypic material was selected to genetically dissect one of the discovered QTL and assess the potential of these QTL to improve plant productivity under fluctuating light regimes. In this work, new genetic variation, phenotypes and the possibility of a gene with unknown function *TRANSCRIPTION INITIATION FACTOR II D (TFIID)* (At3g45210) as being involved in high light acclimation are described. The outcomes of this study highlight the potential that is present within plant species to improve photosynthesis efficiency, and consequently yield.

---

[1]Laboratory of Genetics, Wageningen University & Research, [2]Rijk Zwaan Breeding B.V., [3]Laboratory of Biophysics, Wageningen University & Research

1. INTRODUCTION

Light energy intercepted by plant leaves is the driver of plant photosynthesis, plant growth and, ultimately, for most of the life in the terrestrial biome. To function optimally, photosynthesis relies on the balanced operation of several distinct sub-processes; light-harvesting, electron and proton transport, photosynthetic metabolism, and the supply of carbon dioxide from the surrounding air. This balanced operation depends on regulation of the photosynthetic sub-processes. In the short-term (seconds - hours time range) this depends on the physiological regulation by changes in, for example, lumen pH, or the activation of enzymes by thioredoxin or stromal Mg^{2+} concentration or pH. In the longer term (several hours or longer) this regulation depends on changes in the composition or stoichiometry and organization of the photosynthetic machinery.

Disturbance to the operation of photosynthesis – and thus the activation of regulatory responses – occur in part due to changes in the physical environment of the leaf. Photosynthesis is known to be very sensitive to the environment, and the environment fluctuates over a large range of time scales. Fluctuations in irradiance, the energetic driving force of photosynthesis, have a particularly strong effect on photosynthesis. Partly this arises from the physiological role played by light in photosynthesis and partly from the fact that there are no buffering mechanisms for light, so fluctuations in light can be large, rapid and common (Kaiser *et al*., 2018, Matsubara 2018). When these changes in light intensity are prolonged, they result in alterations in the stoichiometry of the components of photosynthesis. The number of light-harvesting components decreases, while electron transport, ATP synthesizing capacity, and metabolic capacity (especially of rubisco) increases. Photoprotective mechanisms will also be upregulated to deal with the risks associated with the formation of reactive oxygen species in higher irradiance conditions.

Adaptation to changing incoming irradiance by either physiological or stoichiometric changes may be slow compared to the speed of fluctuations in irradiance (Kaiser *et al*., 2018). Light fluctuations that prevent a completely adjusted state of the photosynthesis apparatus impact leaf carbon assimilation and, consequently, productivity (Athanasiou *et al*., 2010; Vialet-Chabrand *et al*., 2017; Graham *et al*., 2017; Kaiser *et al*., 2020). Improving a plant's capacity to adapt to changing light conditions is therefore seen as a promising target with room for improvement for purposes of improving crop yield (Zhu *et al*. 2015). Within plant species, natural variation exists for the rate or extent of the response of photosynthetic sub-processes important for short and long-term adaptation to changes in irradiance, which means that such traits are also amenable to genetic mapping (Niyogi *et al*., 2009; van Rooijen *et al*. 2015, Wang *et al*., 2016, Rungrat *et al*., 2019). Consequently, this opens a route to improve these responses by exploiting this natural variation in crop breeding for improving adaptation under conditions of light fluctuation as a potential route of investigation to further advance crop production (Flood *et al*., 2011; Lawson *et al*., 2012, van Bezouw *et al*., 2019).

Genetic diversity for a trait can be explored in various ways. Diversity panels of model and crop species have become a popular tool to screen genetic variation in photosynthesis following genome wide association studies (van Bezouw *et al*., 2019). However, in such studies typically only few of the observed quantitative trait loci (QTL) for photosynthesis studies have been verified and are generally found to be weak in effect. Classic bi-parental mapping populations present a more powerful alternative for studying the genetics of quantitative traits (Bazakos *et al*. 2017), but these are not as efficient in mapping accuracy. However, for highly polygenic traits such as photosynthesis, the simplistic genetic layout of bi-parental mapping populations can be advantageous as the number of genetic loci contributing to the trait of interest and the number of genetic factors contributing to non-additive genetic portion of variation are much lower (Tiang *et al*., 2021). Furthermore, bi-parental mapping populations are a good basis for subsequent selection experiments to assess the impact of QTL on the trait of interest through the development of e.g. introgression lines.

Two *Arabidopsis thaliana* (hereafter; Arabidopsis) accessions, S96 and SLSP30, were previously identified in a genome wide association study by van Rooijen *et al*. (2017) to contrast in their photosynthetic response to long-term high light acclimation. The contrasting response of these two accessions is utilized in this study to further explore the high light acclimation response in Arabidopsis

using an F2-mapping approach. We discovered two high light responsive QTL, which combined account for a 25% improvement of photosynthesis efficiency in response to increased irradiance. The associated markers were used to evaluate selection for these QTL in fluctuating light experiments to assess the impact of these large effect size on productivity. Furthermore, one of the QTL found in this study was fine-mapped to the gene level, identifying *TRANSCRIPTION INITIATION FACTOR SUBUNIT D* (*TFIID*; at3g45210) with a function in the mitochondria underlying the genomic locus. In all, this study highlights the large impact of intra-specific natural variation that can be found for photosynthesis traits in Arabidopsis.

2. MATERIALS AND METHODS

2.1 Seed material

Arabidopsis accessions S96 (CS76153) and SLSP30 (CS76228) were crossed and the resulting F2 was used for genetic mapping. After phenotyping, 304 plants were genotyped using Illumina's Golden Gate Genotyping with Vera Code Technology for 384 markers, of which 117 markers were found to be polymorphic between the parents. These markers were subsequently used to generate a genetic map through Joinmap 4 (van Ooijen, 2006) (**Supplementary figure 1**) and were used for linkage analysis. F2 line progeny was allowed to set seed after the conclusion of this initial experiment. The progeny was then used for the development of fine-mapping populations and introgression lines as described in the results section.

2.2 Whole genome sequencing

Genomic DNA of the parental lines was isolated following the Maloof CTAB-protocol (see https://openwetware.org/wiki/Maloof_Lab:96well_CTAB). Genomes of S96 and SLSP30 were sequenced at 40x coverage by GenomeScan b.v.. The snpeff bioinformatics tool (Cingolani *et al*., 2013) was subsequently used to map polymorphisms between the S96 and S96 genomes, which were then used to develop KASP probes for genotyping and predicting the impact of genetic variation on candidate gene functioning. Primer sequences and their respective positions can be found in **supplementary table 1**. These KASP probes were subsequently used for genotyping in mapping experiments, recombinant selection and the development and selection of introgression lines

2.3 Growth conditions and high-throughput phenotyping using the Phenovator

For genetic mapping of photosynthesis efficiency and phenotyping of genetic material, plants were grown as described in van Rooijen *et al*. (2017). In short, plants were grown for 24 days under conditions of 100 µmol m^{-2} s^{-1} of light, then a stepwise increase of light irradiance to 450 µmol m^{-2} s^{-1} was introduced for four days. Unless otherwise noted, the Phenovator experiments were as described here. Four to six times per day, ΦPSII was measured as an indicator for photosynthesis performance using chlorophyll fluorescence techniques as described by Flood *et al*. (2016), Murchie and Lawson (2013) and in **Chapter 3**.

2.4 Mutant characterization and analysis

For all genes located between the flanking markers of the Q3 quantitative trait locus and found to be expressed in rosettes, up to three independent T-DNA insertion lines were ordered from the Nottingham Arabidopsis Seed Centre (Scholl *et al*. 2000), depending on availability. For At3g45140 *LYPOXYGENASE 2* (*LOX-2),* no TDNA insertion lines were available and thus we requested the *lox2-1* EMS-mutant from dr. Mats Andersson (Swedish University of Agricultural Sciences) and prof. Edward Farmers (University of Lausanne), which was kindly provided. A snp mutation introduces a W630* amino acid substitution which truncates the protein and causes a proper gene knockout, as previously described by Gauser *et al*. (2009). In addition, we ordered the RNAi-LOX2 (NASC stock: N3748) line, but contrary to what is reported by Bell *et al*. (2005), we were able to detect *LOX2* transcripts indicating a possible loss of function of the RNAi-transgene. All TDNA lines were tested for homozygosity of the insert and other mutants were genotypically verified using the primers as described in s**upplementary table 2**. All plants were grown and harvested simultaneously in the greenhouse prior to subsequent experimentation to avoid undesired variation resulting from seed age or unequal growth conditions.

2.5 RNA isolation and sequencing.

The B9-C11 and B9-E7 introgression lines (**Figure 2**), selected for genomewide homozygosity for the Q3 QTL from the progeny of an F2 line, were grown in the Phenovator, and grown as described previously. 48 replicates of each line were grown in a corner of the Phenovator, in a checkerboard experimental design to ease the collection of the samples. The table was subdivided in six blocks to ensure proper randomized sampling. For five time points, 09:00 low light (LL), 09:00 high light (day 1; HL1), 11:30 HL (day 1; HL2), 09:00 HL (day 2; HL3), 09:00 HL (day 5; HL4), six rosettes belonging to each genotype were harvested and snap-frozen in liquid nitrogen – for a total of thirty samples per genotype. Collecting all twelve (2x6) rosettes per time point this way took less than one minute. The rosettes were stored in a -80 freezer before subsequent isolation of RNA. RNA was isolated using protocols as described in C**hapter 3**. Five RNA-samples with the best RNA integreity scores per genotype per time point were chosen for sequencing and then sent to Novogene^{tm} (Cambridge), processed into cDNA, normalized and sequenced using an Illumina Novaseq 600. The raw reads were returned with the final data quality control report given in **supplementary table 3**. Raw sequences were then processed using a specialized in in-lab script using Kallisto (Bray *et al*., 2016) script and aligned against the Col-0 reference genome TAIR10.1 for purposes of gene identification and read countings. Subsequently, RNA-expression levels were analyzed in the Sleuth (Pimentel *et al*., 2017) and Deseq2 (Love *et al*., 2014) packages.

2.6 Fluctuating light regime experiments and phenotyping.

Inbred lines containing different variants of the QTLs in homozygous state (See **figure 1, supplementary figure 2**), and the introgression lines B9-C11 and B9-E7 (**Figure 2**), were grown in a climate chamber and illuminated by Fluence LED modules. Plants were grown for 5 days at 300 µmol m^{-2} s^{-1} before being exposed to a range of light regimes that differ in their fluctuation frequency. The "Constant" light (CL-)regime was set at 300 µmol m^{-2} s^{-1}, while fluctuating light regimes were set at three different frequencies, oscillating between stepwise changes in irradiance between low light, 100 µmol m^{-2} s^{-1}, and high light, 500 µmol m^{-2} s^{-1}; in the "24 h" treatment, changes in irradiance alternated between photoperiods over days, in "60 min" alternate and "5 min" the light irradiance changes changes every five minutes **(Figure 3)**. On the 28th day of growth, light intensity was set to control conditions (300 µmol m^{-2} s^{-1}) and plants were measured using an enclosed semi-automated chlorophyll fluorescence measuring facility. This facility, PSI PlantScreen^{tm} (see https://www.npec.nl/tool/the-robin-psi-plantscreentm-system/ for details), is capable of measuring only few plants each time compared to the Phenovator (**Chapter 3**; Flood *et al*., 2016). Furthermore, plants must be carried into the facility to be phenotyped. However, the main advantage of the PSI PlantScreen^{tm} is that it can be programmed to assess a much larger number of possible traits of interest and is thus more flexible in photosynthetis trait analysis. Measurement protocols for photosynthesis traits were developed that measure $\phi PSII$, $NPQt$, qEt (energy dependent, reversible photochemical quenching) and *qIt* (photoinhibtion), following equations as described in Baker (2008) and Tietz *et al*. (2017);

Equation 1) $\phi PSII = F_q / F_m'$
Equation 2) $NPQt = (4.88 / ((F_m' / F_o') - 1)) - 1$
Equation 3) $qEt = (F_m'' - F_m') / F_m'$
Equation 4) $qIt = (4.88 / ((F_m'' / F_o') - 1) - 1))$

To assess these parameters in the PSI PlantScreen^{tm}, all plants were acclimatized for 250 seconds at 300 µmol m^{-2} s^{-1}. At the end of this acclimation period F' and F_m' were measured, the latter with a saturating pulse of 5500 µmol m^{-2} s^{-1}. The recovery period of the saturating pulse was 20 seconds, after which plants were exposed to darkness for four seconds. Subsequently, plants were exposed to far-red light for 16 seconds. At the end of this far-red light acclimation, F_o' and F_m'' – the latter again under a saturating pulse of 5500 µmol m^{-2} s^{-1}. After phenotyping in the PlantScreen^{tm} system, the aboveground parts of the plants

were harvested and put in paper bags. Plants were oven-dried for 60 degrees for at least two days and measured using a closed chamber analytical balance with an accuracy of four digits.

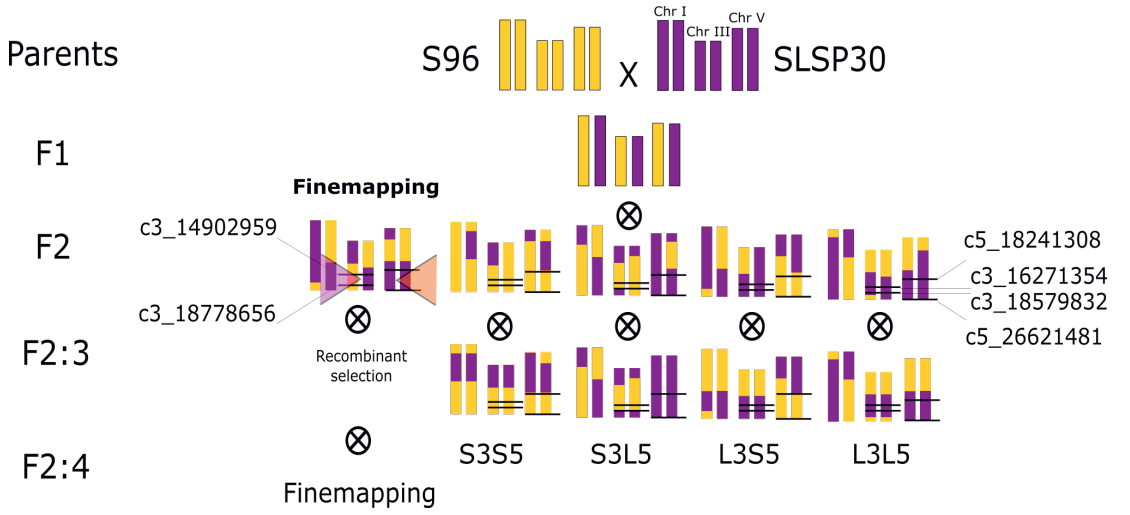


Figure 1. *Overview of crosses and generation of genetic material for the genetic analysis of high light acclimation of the Arabidopsis thaliana accessions S96 and SLSP30. Parents were crossed to generate the F2 population. From the F2 population, two experimental routes were followed, a fine-mapping approach and the generation of introgression families (IFs). In the fine-mapping approach, several F2 lines were chosen that only segregate for Q3 (the genomic locus indicated by the pink triangle) and are homozygous for Q5 (the genomic locus indicated by the orange triangle). Progeny of these specific F2 lines were grown to select for recombinants based on the molecular markers given on the left, giving the F2:3 generation. Inbreeding of the selected lines then form the F2:4 fine-mapping population for Q3. Further material was developed to study the cumulative effect of the Q3 and Q5 QTL on phenotypic performance and productivity. IFs are collections of lines that have been selected based on their homozygosity for either QTL. For two QTL, this can be subdivided into four classes, while they still segregate for their background (**Supplementary figure 2**). S3S5 represents F2 lines selected to be homozygous for S96 alleles for both the Q3 and Q5 alleles, S3L5 are homozygous S96 on Q3 and homozygous SLSP30 on Q5, and so on. The markers on the right are derived from polymorphic makers from the Illumina Golden Gate chip and its genetic map (**Supplementary figure 1**). The markers denoted on the left are KASP probes which have been developed exclusively for the fine-mapping study (**Supplementary table 1**). The markers give the chromosome number and nucleotide position in basepairs, following the TAIR10.1 reference genome sequence of Arabidopsis thaliana. Only chromosomes I, III and V are shown and the pictures represent concepts, not actual recombinants. The figures are not scaled.*

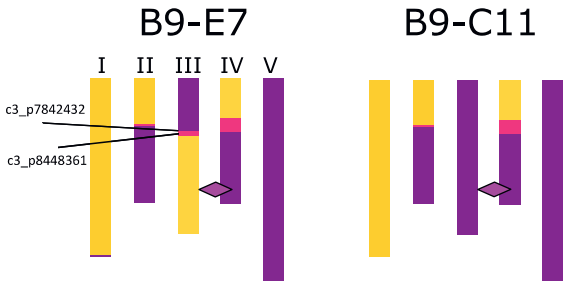


Figure 2. *The scaled genetic layout of the B9-E7 and B9-C11 introgression lines. Purple = marker scores SLSP30, yellow = marker scores S96, pink = ambiguous region. The purple diamond indicates the location of the Q3 QTL for high light acclimation (marker location c3_16271364).*

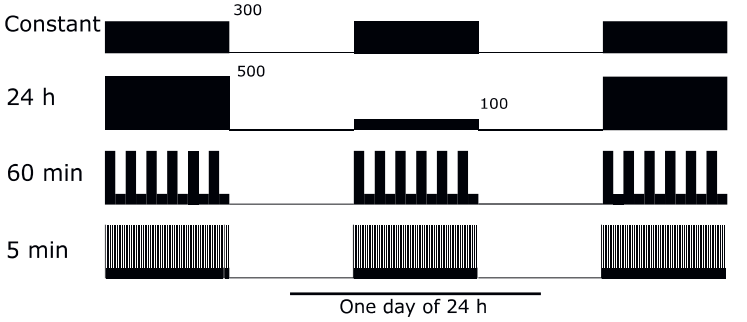


Figure 3. *Light regimes used during the fluctuating light experiments. The durations indicate how long, during each day, it takes for the next stepwise change in photosynthetically active radiation (in µmol m^{-2} s^{-1}). The "Constant" light treatment regime had plants grown stably at 300 µmol m^{-2} s^{-1}. All fluctuating light regimes have a maximum of 500 µmol m^{-2} s^{-1} and a minimum of 100 µmol m^{-2} s^{-1}. Each day of 24 hours contains a photoperiod of 12 hours. The figure shows three photoperiods and two dark periods (no light) per treatment for the "Constant", "24 h", "60 min" and 5 "min".*

2.7 Statistical analysis and QTL-mapping.

QTL-mapping was performed using the qtl package in R (Broman *et al*. 2003). Multiple QTL Mapping was used in order to increase the reliability of the observed results (Arends *et al*. 2010). For analysis of T-DNA lines and analysis of parental and introgression lines in the Phenovator, a linear mixed model was applied including block and camera block (See Flood *et al*. 2016). Camera block is an alternative blocking effect for camera position that results from a peculiar deviation within the camera head of the Phenovator (**Supplementary figure 3**). Analysis of data from a uniformity trial in the Phenovator suggested that camera block displays a strong, stepwise environmental gradient which is more influential than any of those listed individually in Flood *et al*. (2016). As such, the linear model for all line-to-line comparisons follows;

$$\Phi PSII = genotype + experimental.block + camera.block + \varepsilon$$

where genotype includes introgression lines or mutant lines, experimental.block follows from Flood *et al*. (2016) and camera.block is a factor as described previously.

3. RESULTS

3.1 Assessment of the photosynthetic response to high light acclimation in two Arabidopsis accessions and discovery of two large effect size quantitative trait loci

S96 and SLSP30 are two Arabidopsis accessions that respond differently to an increase in high light (**Figure 4a**). S96 always exhibits a 2% higher ΦPSII compared to its SLSP30 counterpart throughout the low light phase of the experiments. The relative difference increases five-fold, immediately after exposure to high irradiance (**Figure 4b**), but by the end of the first day of high light this difference increases to a 20% higher ΦPSII – a tenfold larger difference (**Figure 4c**). Meanwhile, ΦPSII of SLSP30 declines throughout the first day of high light, while ΦPSII in S96 does not. To investigate the underlying genetic components of this difference, 304 segregating F2 progeny plants of the F1 hybrid S96 between and SLSP30 were grown in the Phenovator and phenotyped for the low light (100 µmol m^{-2} s^{-1}) to high light (450 µmol m^{-2} s^{-1}) acclimation response (See van Rooijen *et al*. 2015, 2017). After phenotyping, the individual F2 lines were grown to maturity, harvested for seeds and post-experimental genotyping revealed a total of 117 markers to be polymorphic between the parental lines. A genetic map was drawn totaling 381.9 centimorgan (cM), with an average marker spacing of 3.26 centiMorgan, or 3.05 Mbp, (**Supplementary figure 1**). This map was then used to perform genetic mapping of quantitative trait loci (QTL) for this trait.

To identify genetic loci in the F2-population, a multiple QTL-mapping (MQM-)procedure was performed as described by Arends *et al*. (2010). Three genetic markers are identified as co-factors to improve mapping power and delineation of the QTL; c3_p16271354, c5_p22099562 and c5_p26203511 (**Supplementary figure 4**). No QTL are detected during the low light growth stage of the experiment, while two QTL emerge after a stepwise increase of irradiance on chromosome 3 (here designated as Q3) and chromosome 5 (Q5) (**Figure 5a, Table 1**). These QTL were found for two days before the strength of the associations decays from the third day of high light. The QTL on chromosome 5 potentially consisted of two co-locating independent QTL (**Figure 5b**). The MQM-mapping procedure is unable to make a further distinction and Q5 is thus considered to be one locus for the remainder of this work. For both QTL, alleles originating from S96 are associated with higher values of ΦPSII in the acclimatization response to high light and both QTL explain a maximum of 30% of the variance depending on the time point (**Table 1**). Q3 and Q5 each cause a relative effect size of, at maximum, 10% and 15%, respectively, with particularly high differences recorded at the end of day 1 and on day 2 of high light (**Figure 5c**). In addition, no evidence for epistatic interaction was found in the population (**Supplementary table 4**), suggesting that Q3 and Q5 act additively and do not interact.

Q3 and Q5 show a combined additive phenotypic effect of 25% for ΦPSII, which implies a significant impact on the photosynthesis capacity in lines carrying either both the stronger or the weaker alleles for these QTL. To further investigate their significance, a selection program was performed (**Figure 1**), to be able to assess the implication the effect of these photosynthesis QTL on plant growth. Two types of genetic material are developed from the initial F2 population (**Table 1**). Specific groups of lines were selected that are homozygous for the different alleles of Q3 and Q5 (**Figure 1**), but otherwise still segregate in the background (**Supplementary figure 2**). To minimize background effects, multiple of such lines are chosen per group, forming families of related lines. These families are defined as “introgression line families”, and are composed of progeny from five to seven F2 lines, selected for combinations of homozygous loci for Q3 and Q5 for the various alleles, but with variable backgrounds. Each introgression line family (IF) is designated as S3S5, L3S5, S3L5 and L3L5, with S as a designator for S96 alleles and L as a designator for SLSP30 alleles for Q3 and Q5 (**Figure 1**). Furthermore, two introgression lines (ILs), B9-C7 and B9-E7, were selected as different for the Q3 locus and carring SLSP30 alleles for Q5 (**Figure 2**). These lines were chosen to further the explore Q3 QTL in more detail as described further.

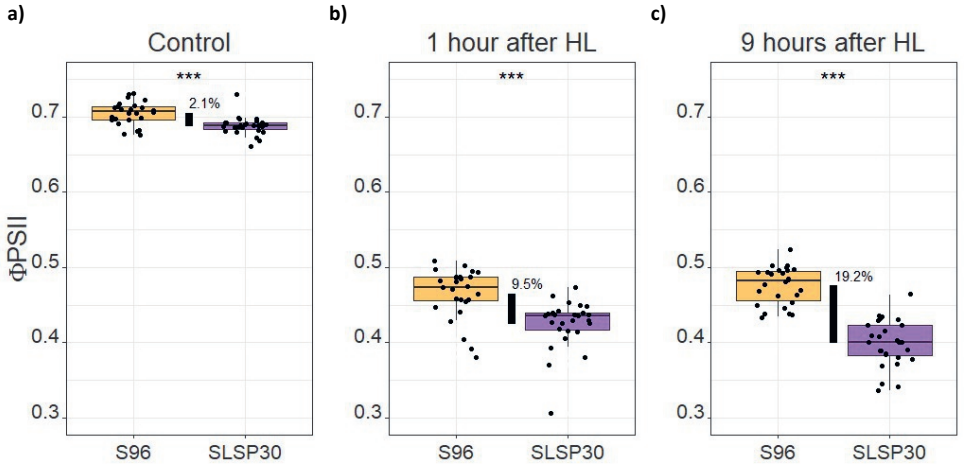


Figure 4. *ΦPSII performance of S96 and SLSP30 under conditions of increased irradiance. Evaluation of ΦPSII in **a**) 100 µmol m^{-2} s^{-1} on day 24 of growth, **b**) one hour after the onset of a stepwise increase in irradiance to 450 µmol m^{-2} s^{-1} on day 25 of growth, **c**) as in **b**, but now after 9 hours after the onset of the photoperiod. The black bars indicate the difference of the genotypic means between the parental lines, with the numbers indicating the percentage change between the accessions. n = 26 for S96, n = 27 for SLSP30. In 1**a-c**, black dots represent individual measurements. *** = p < 0.001, following a linear mixed model as described in the methods..*

Table 1. *Characteristics of the Q3 and Q5 quantitative trait loci associated with the photosynthetic high light acclimation response for three representative time points. "Treatment"; a select number of time points as described in **figure 5**. "QTL"; the name of the QTL. A two-QTL model involving both QTL is also included. "Position" = genetic position in centiMorgan (cM) (see also **Supplementary figure 1**), "Marker" = the peak marker associated with the QTL. The marker name is a combination of the chromosome and the marker position (in base pairs). "LOD"; Logarithm of the odds as -logp, "PVE"; percentage of the total genotypic variation that is explained by the marker. "Allele"; the allele contributing to higher values of ΦPSII.*

Treatment	QTL	Chromosome	Position (cM)	Marker (with position)	LOD (-log(p))	PVE (%)	Allele
LL2.1	Q3	3	45.9	c3_p16271354	0.93	1.40	S96
HL1.1	Q3	3	45.9	c3_p16271354	3.21	4.75	S96
HL2.1	Q3	3	45.9	c3_p16271354	8.65	12.27	S96
LL2.1	Q5	5	79.5	c5_p26203511	1.35	2.03	S96
HL1.1	Q5	5	79.5	c5_p26203511	5.65	8.20	S96
HL2.1	Q5	5	79.5	c5 p26203511	11.5	15.99	S96
LL2.1	Q3+Q5	-	-	-	2.22	3.31	-
HL1.1	Q3+Q5	-	-	-	8.90	12.60	-
HL2.1	Q3+Q5	-	-	-	21.05	27.30	-

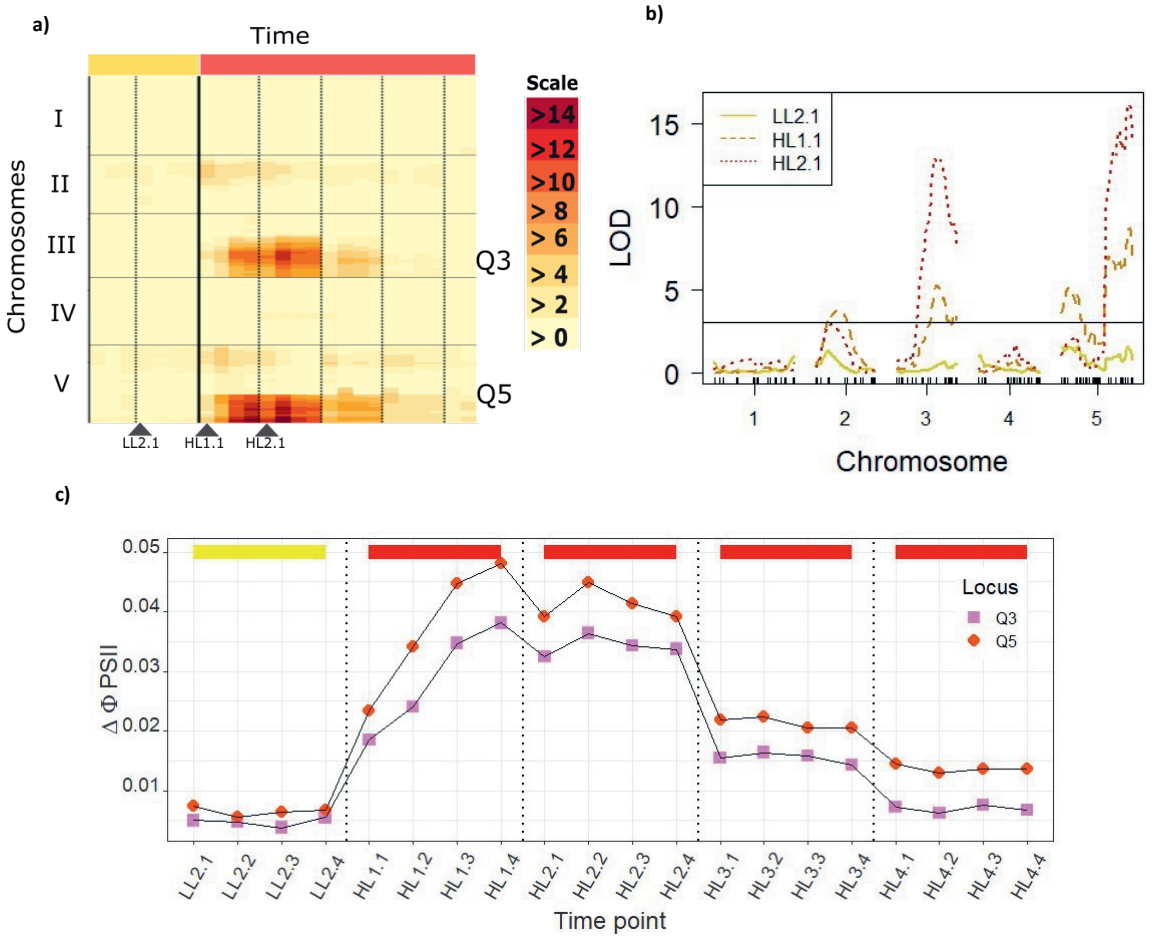


Figure 5. *Genetic mapping of the response of ΦPSII to a stepwise increase in high light in the S96 x SLSP30 F2-population.* ***a****) MQM mapping of high light acclimation QTL over time. The x-axis represents time points (4 per day) and the y-axis represents the genetic markers spread across the five chromosomes. Dotted vertical lines delineate the days, while the bold line indicates the switch of irradiance from 100 µmol m^{-2} s^{-1} (low light) to 450 µmol m^{-2} s^{-1} (high light). Scale gives the –log(p) value for each time point.* ***b****) QTL plot forΦPSII measured at three different time points as indicated with black triangles in* ***a****. The threshold was drawn at 2.9 as an average for all phenotypes after a permutation test of 450 permutations.* ***c****) Phenotypic difference in ΦPSII between the S96 and SLSP30 alleles for each time point for Q3 (homozygous for markers c3_p16271354:c3_p18579832) and Q5 (homozygous for markers c5_p18241308:c5_p26621481). Positive values indicate that S96 alleles are associated with higher photosynthesis, while negative values for SLSP30. The yellow and red bars in* ***a****,* ***c*** *represent the low light and high light phases. Time points represent indicated as LL = low light, HL = high light, with the first number being the days and the second number the specific time points within the days; 1 = 1.0 h after onset of photoperiod, 2 = 3.5 h, 3 = 5.0 h and 4 = 7.5 h.*

3.2 Analysis of allele-specific QTL effects on ϕPSII and growth in the introgression lines and families

To validate the effect of the QTL in the and introgression line families (IFs) (**Figure 1, supplementary figure 2**) and introgression lines (ILs) (**Figure 2**) lines were grown in the Phenovator in the same experimental conditions as the F2-population. S3S5 IFs, homozygous for S96 at both QTL, show an enhanced photosynthesis efficiency of maximally 20% higher than the L3L5 IFs, homozygous for SLSP30 at both QTL (**Figure 6a**), while this difference is only ~2% under low light conditions. It takes L3L5 a full day to reach the levels of photosynthesis efficiency as S3S5 on the first day. Tukey post-hoc tests conclude that the L3L5 IF exhibits lower ΦPSII compared to the other three introgression families for the majority of the high light acclimation period. This suggests an epistatic effect between Q3 and Q5 to take place (**Figure 6b**), although we found no evidence for such in the analysis of the whole F2-population (**Supplementary table 4**). S3L5 and L3S5 IFs take more intermediate positions in their ranking of ΦPSII values. Between the introgression lines, IL B9-E7 carrying the S96 allele for Q3 shows an approximate 10% higher ΦPSII compared to B9-C11 carrying the SLSP30 allele for Q3 (**Figure 6c**). This difference is consistent with the Q3 effect as identified in the F2-mapping (**Figure 5c**)

In the constant low light phase, the S96 and SLSP30 accessions exhibit a significant difference in growth after 24 days (**Figure 7a**). However, for the ILs and IFs, all genotypes reach similar sizes with no significant differences between B9-E7 and B9-C11 (**Figure 7b**) or between the IFs (**Figure 7c**). As such, these genotypic selections effectively eliminate the differences in growth observed between the parents in constant light conditions.

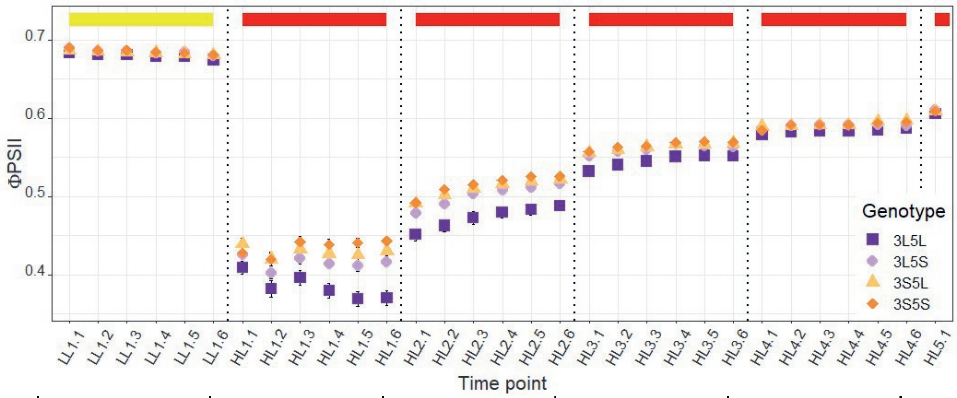
3.3 The effect of Q3 and Q5 on biomass and photosynthesis traits under fluctuating light conditions

During the high light acclimation experiments, only a single stepwise increase of irradiance introducing a photosynthesis response is used. However, in natural environments many changes in irradiance are typically observed and at various durations (Kaiser *et al*. 2018). Athanasiou *et al*. (2010) previously reported on the correlation between genotypic fitness under fluctuating light conditions and genotypic variation in the photosynthetic response to a stepwise change to high light. In line with these observations, we hypothesized that repeated changes of irradiance triggering the Q3 and Q5 QTL effects can potentially impact growth. To test this hypothesis, IFs and ILs were subjected to four different regimes of fluctuating light for 28 days and measured for productivity (as dry weight), but also photosynthesis efficiency and non-photochemical quenching (photoprotection, Tietz *et al*., 2017). The light regimes consist of constant light at 300 µmol m^{-2} s^{-1} ("Constant"), and three fluctuating light regimes involving stepwise changes to 100 µmol m^{-2} s^{-1} and 500 µmol m^{-2} s^{-1}, with intervals lasting full days (24 h), hours (60 min) to five minutes (5 min) (**Figure 3**).

Plants grown under the fluctuating light regimes show variable degrees of growth reductions compared to the constant light treatment. Most notably, the long-term 24h treatment leads to a ~40% reduction after 28 days of growth compared to plants grown in constant lights. 60 min fluctuations and 5 min fast fluctuations caused reductions of approximately 25% in dry weight (**Figure 8a**). Values for ϕPSII are lower (**Figure 8b**), while total non-photochemical quenching (NPQ_t) is higher (**Figure 8c**) for plants grown under any of the fluctuating light treatments.Photoinhibitory quenching (qI_t) is highest in the 24 h treatment (**Figure 8d**) and short term energy-dependent quenching qE_t is highest in plants grown in conditions of 60 min and 5 min treatments (**Figure 8e**). Between ILs B9-C11 and B9-E7, variation is found only found for ΦPSII, NPQ_t and qI_t in the 60 min growth conditions although generally higher values are reported the other conditions for B9-E7. Plant productivity remained the same in all conditions (**Figure 9a-e**). Most notably, qI_t is higher in B9-E7 than in B9-C11 (**Figure 9d**), particularly in the 60 min

treatment. Phenotypic variation among the IFs is larger than the ILs, due to selection for two QTLs. Among the IFs, S3S5 score lower values for dry weight (**Figure 10a**) and ΦPSII (**Figure 10b**), while values for the various NPQ_t components are consistently higher in these lines (**Figure 10c-d**). No genotypic variation is observed for qE_t among all conditions (**Figure 9e**, **10e**).

a)



b)

	LL						HL1						HL2						HL3						HL4						HL5
S3S5	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
S3L5	ab	ab	ab	ab	ab	ab	a	a	ab	ab	a	ab	ab	ab	a	a	ab	a	a	a	a	a	a	a	a	a	a	a	a	a	a
L3S5	ab	ab	ab	ab	ab	ab	ab	ab	b	b	b	b	b	b	a	a	b	a	a	a	a	a	a	ab	a	a	a	a	a	a	a
L3L5	b	b	b	b	b	b	b	b	c	c	c	c	c	c	b	b	c	b	b	b	b	b	b	b	a	a	a	a	a	a	a

c)

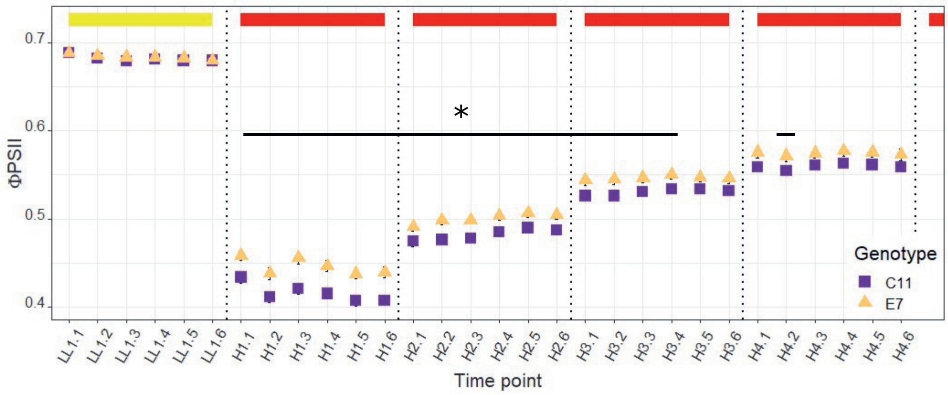


Figure 6. *Phenotypic confirmation of Introgression family (IF) and line (IL) analysis.* ***a)*** *Phenotypic evaluation of four F2:4 families each with a different selection of alleles for Q3 and Q5, that have either SLSP30 (L3L5) or S96 (S3S5) alleles – or a combination thereof (L3S5, S3L5). Dots represent the phenotypic mean of each ILF per time point, with n = 28 for each allelic group. Six or seven independent IFs were grown for each genotypic group in four or five replicates.* ***b)*** *Tukey post-hoc test outcomes per time point for the data depicted in* ***a.*** *Letters indicate groups of IFs with similar phenotypic means for ϕPSII.* ***c)*** *Phenotypic means for introgression lines B9-C11 and B9-E7, differing for only the Q3 QTL due to a large chromosomal introgression (**Figure 2**). The black bar indicates the range of time points in which ($p < 0.05$) = *. Time points represent low light (100 µmol m^{-2} s^{-1}) or high light (500 µmol m^{-2} s^{-1}) phases – also indicated with yellow and bars at the top of the graphs. First number indicating day, starting for plants grown for 24 days, second number specific time points of measurements (X.1 = 9:00, X.2 = 11:00, X.3 = 12:30 and X.4 = 14:00, X.5 = 15:30, X.6 = 17:00.*

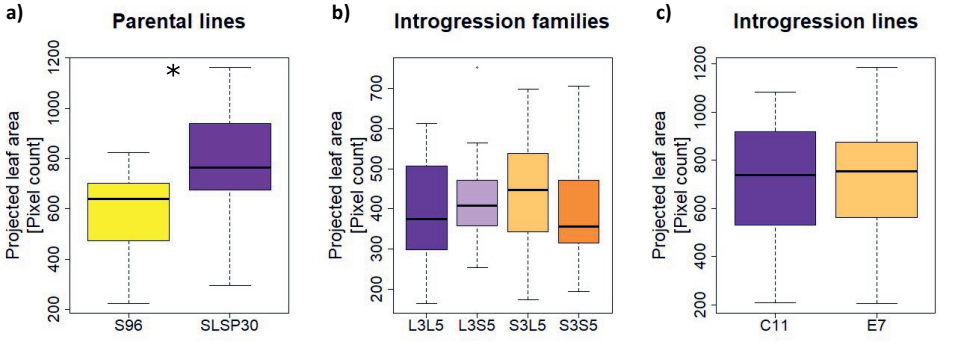


Figure 7. *Evaluation of projected leaf area of the **a**) parental accessions (S96 and SLSP30), **b**) introgression families (IFs)(L3L5, L3S5, S3L5, S3S5,) and **c**) introgression lines (ILs) (B9-C11 and B9-E7) after 24 days of growth at constant light (100 µmol m^{-2} s^{-1}). n = 21-28 for each independent IF and IL. A significant difference is only found between the parental lines ($p < 0.05$), following ANOVA on linear models as described in the methods.*

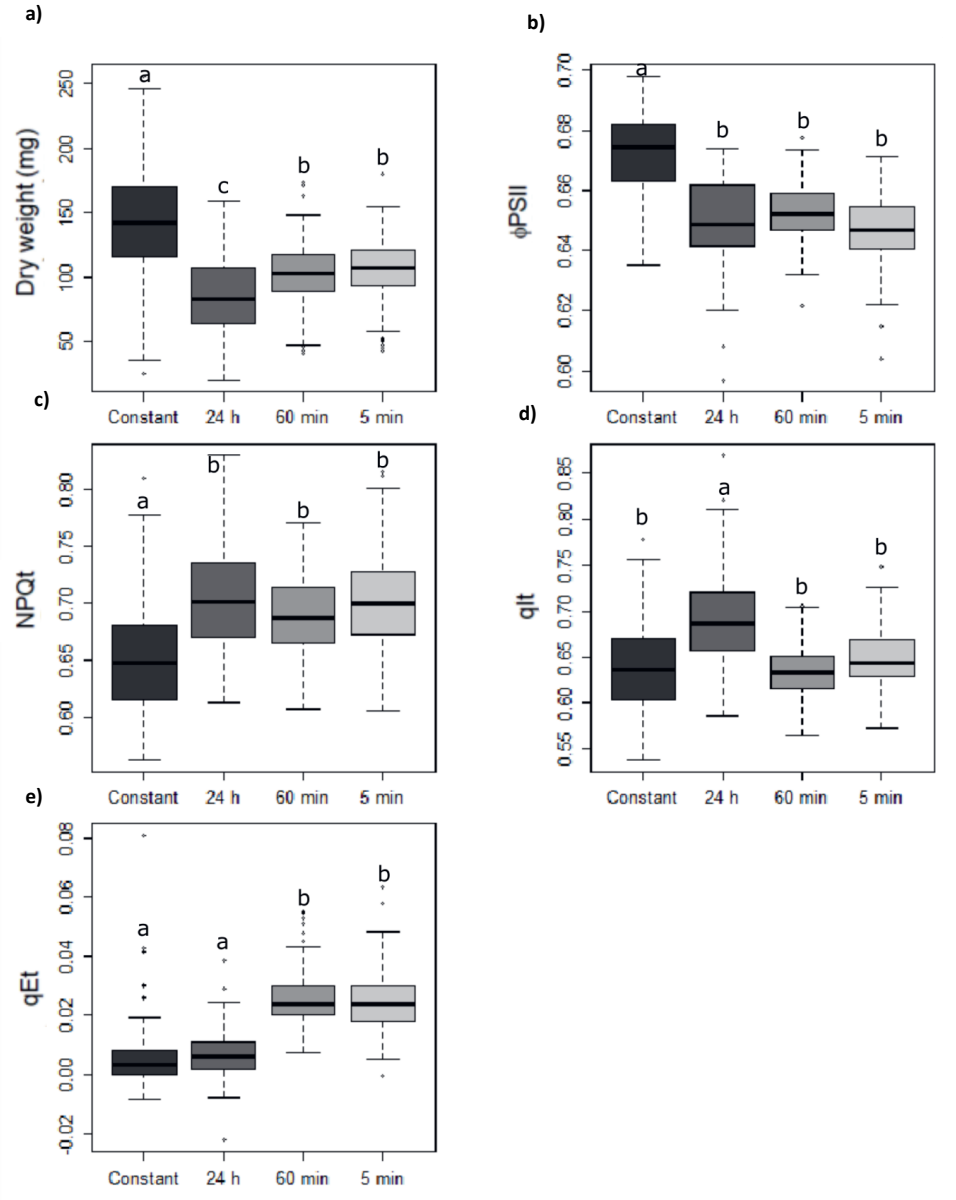


Figure 8. *Phenotypic response of plants grown under different light conditions. Constant light (grown at 300 µmol m^{-2} s^{-1}), 24 hour fluctuations (100/500 µm m-2 s-1 every other day), 60 minutes (100/500 µmol m^{-2} s^{-1}) and 5 minutes (100/500 µm m^{-2} s^{-1}). Total accumulated incoming irradiance among all treatments is equal. Traits are* ***a****) total rosette dry weight in milligrams,* ***b****) photosystem II efficiency (ϕPSII),* ***c****) theoretical non-photochemical quenching (NPQ_t),* ***d****) photo-inhibitory component of NPQ_t (qI_t).* ***e****) energy dependent component of NPQ_t (qE_t). Total plants evaluated per treatment; constant light (n=169), 24 hour (n=177), 60 minutes (n=217) and 5 minutes (n=220). Letters indicate groups with equal means after Tukey post-hoc test.*

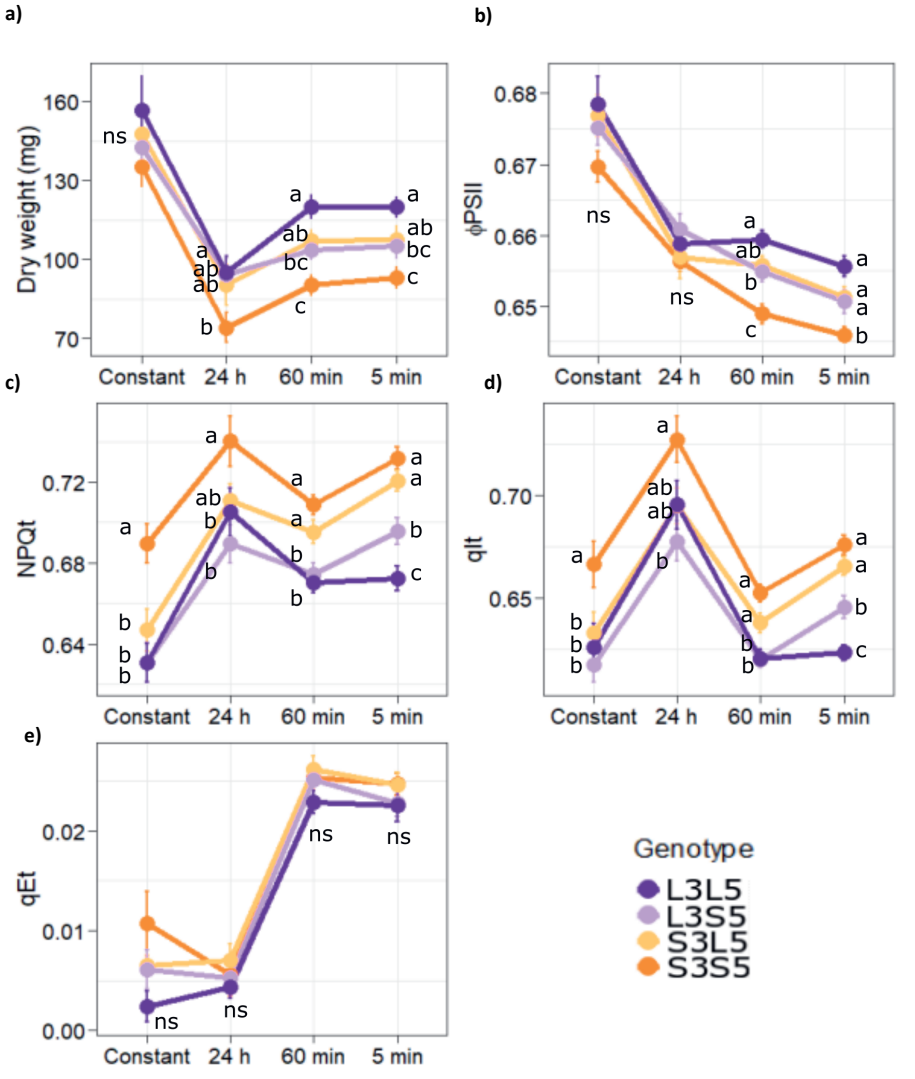


Figure 9. *Phenotypic response of introgression lines families under different light conditions. The dots represent the genotypic means with standard errors given. Treatments; constant light (grown at 300 µmol m-2 s-1), 24 hour fluctuations (100/500 µmol m-2 s-1 every other day), 60 minutes fluctuations (100/500 µmol m-2 s-1), 5 minutes fluctuations (100/500 µmol m-2 s-1). Traits are* ***a****) total rosette dry weight in milligrams,* ***b****) photosystem II efficiency (ϕPSII),* ***c****) non-photochemical quenching (NPQ_t),* ***d****) photo-inhibitory component of NPQ_t (qI_t).* ***e****) energy dependent component of NPQ_t (qE_t). L3L5 = selection of seven F3 lines homozygous SLSP30 for light acclimation quantitative trait loci Q3 and Q5, L3S5 = selection of seven F3 lines homozygous SLSP30 for Q3 and Q5, S3L5 = selections of seven F3 lines homozygous S96 for Q3 and SLSP30 for Q5, S3S5 = selections of seven F3 lines homozygous S96 for QTL Q3 and Q5. 18-39 replicates per genotype per treatment were evaluated. Letters indicate groups with equal means after Tukey post-hoc test. ns = no significant differences were detected in ANOVA prior post-hoc test, thus indicating that there is no genotypic difference in the treatment for the trait.*

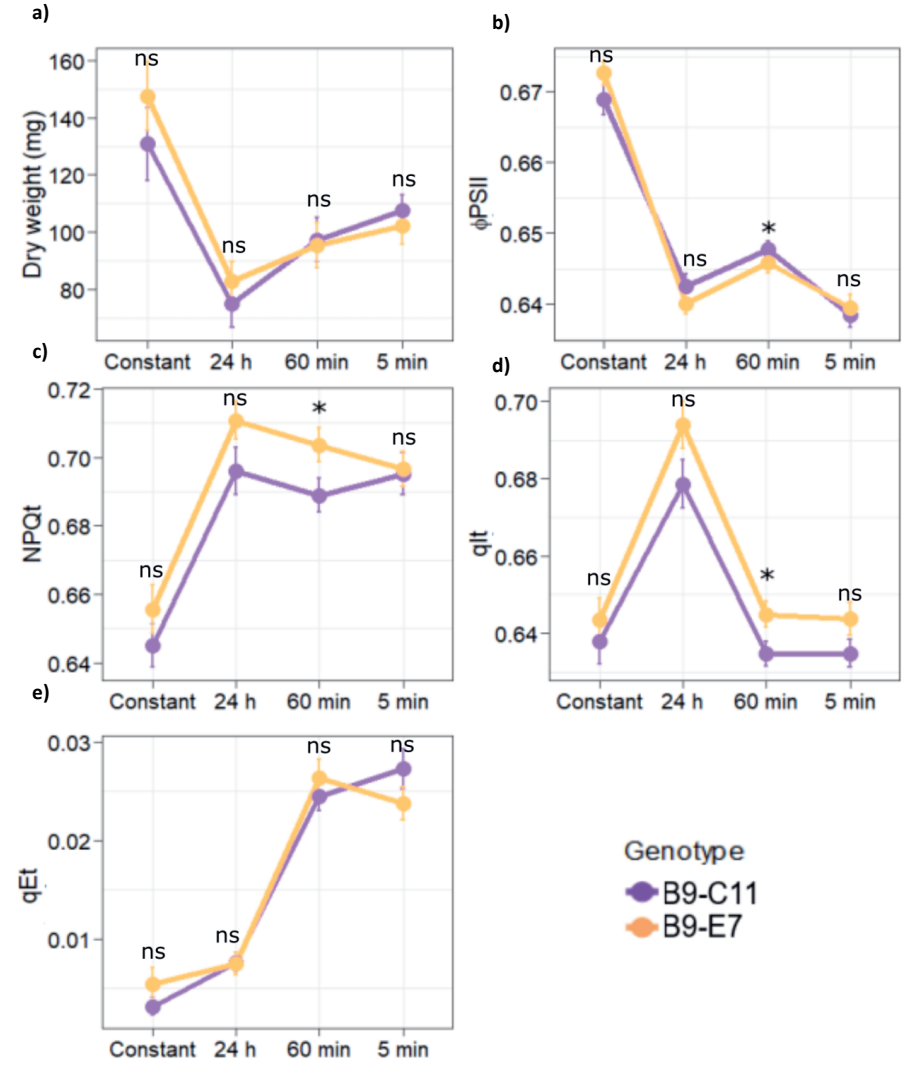


Figure 10. *Phenotypic response of introgression lines under different light conditions. The dots represent the genotypic means with standard errors given. Treatments; constant light (grown at 300 µmol m^{-2} s^{-1}), 24 hour fluctuations (100/500 µmol m^{-2} s^{-1} every other day), 60 minutes (100/500 µmol m^{-2} s^{-1}) and 5 minutes fluctuations (100/500 µmol m^{-2} s^{-1}). Total accumulated incoming irradiance among all treatments is equal. Traits are* ***a)*** *total rosette dry weight in milligrams,* ***b)*** *photosystem II efficiency (ϕPSII),* ***c)*** *non-photochemical quenching (NPQ_t),* ***d)*** *photo-inhibitory component of NPQ_t (qI_t).* ***e)*** *energy dependent component of NPQ_t (qE_t).. B9-C11 is recombinant inbred line that only differs for 12 Mbp compared B9-E7. Total n = 29-44 per genotype per treatment. ns = p > 0.05, * = p < 0.05, following Student's t-test.*

3.4 Fine-mapping of the Q3 quantitative trait locus for high light acclimiation

The fluctuating light experiments show that selection for the Q3 and Q5 QTL may be promising to elevate losses in productivity to fluctuating light conditions. However, knowledge of genes that are causal to photosynthesis phenotypes are required for further implementation in crop species (Theeuwen *et al*., 2022). Due to the small size of the F2 population and subsequent limited number of recombinants, both QTL could not be mapped to regions smaller than several million base pairs, thus requiring additional rounds of fine-mapping. Q3 could be mapped to a region of certainty of approximately ~3.6 Mbp, while the Q5 QTL could only be located to a region of uncertainty of nearly 9.0 Mbp (**Figure 5b**). Therefore, Q3 was chosen as a target for fine-mapping. For the development of the fine-mapping population, F2:3 progeny from three different, independent F2-genotypes heterozygous for Q3 and homozygous SLSP30 for Q5 were selected for the development of an F2:4 fine-mapping population (**Figure 1**). Q3 was defined as laying between positions 14902959 and 18778656 on chromosome 3 and a total of 2046 plants were genotyped using five markers across the region (**Figure 11a**). A total of 365 genotypes were identified as having crossovers between 15819786 and 17785729, thus effectively reducing the QTL to ~2 Mbp. 300 F2:3 lines were selected that show crossovers between the these markers in the region and the progeny of these lines (F2:4) were used to for finemapping. Selection criteria of F2:3 progeny to chose a balanced population included taking into account the occurrence and distribution of crossovers between each of the markers proportionally to the genomic distance covered between the markers on the physical map, if possible. Four plants of each F2:3 line, for a total of 1200 plants, were sown in the Phenovator, with conditions and phenotying similar as in the F2 mapping. All individuals were subsequently sampled for DNA isolation and genotyped for an additional 23 markers which were used to further delineate the region (See **supplementary table 1**).

All non-recombinant genotypes were discarded from the analysis prior to QTL-mapping in this F2:4 population. Fine-mapping confirms the presence of a strong QTL on chromosome 3 (**Figure 11b**), with an observed effect size of, at maximum, ΦPSII 0.033 (or ~10% difference between the alleles) for the last three time points on day one (**Figure 11c**). In this experiment, the confidence interval is reduced to at minimum 170000 and maximum 240000 basepairs, depending on the time point, but effectively between markers 163113585 and 16632585 (**Figure 11a**). In this region, which 37 genes are annotated (following TAIR10.0). In order to further decrease the size of the confidence interval of the QTL, a subset of six F2:3 families was chosen based on the presence of a recombination region (between positions 16481741-16632585). These plants were grown at a much higher replication number of sixty plants per F2:3 line each to maximize phenotypic contrasts of recombination. Plants were grown as previously described, although the light increase was set at 700 µmol m^{-2} s^{-1} instead of 450 µmol m^{-2} s^{-1}, to increase contrasts for the QTL and thus allow a sharper identification of the causal locus. In this experiment, the QTL reached a peak LOD-value of 6.11, but not before the third day of the increased irradiance treatment. The final round of fine-mapping using again extra markers with a further reduced marker spacing (**Supplementary table 1**) eventually lead to the conclusion that the causal locus locates inbetween markers c3_165481154 and c3_16632585 (**Figure 11c**), with a total of nine candidate genes (**Supplementary table 5**)

Remarkably, the phenotypic response for different alleles of the Q3 QTL to high light is apparently different in each of the fine-mapping experiments (**Figure 11d**), but S96 is consistently the allele associated with higher ΦPSII. In the F2 mapping, Q3 reached the largest ΔΦPSII on day one and two after an increase in high light irradiance reaching approximately 8-10% difference between alleles for S96 and of SLSP30. By contrast, such difference in ΔΦPSII could only be observed on the first day after high light acclimation during the first round of fine-mapping. On the second day, the difference is already smaller (**Figure 11d**). The final mapping experiment shows a different response as well, showing the largest absolute phenotypic difference in ΔΦPSII on the third day, although it should be noted that a much a higher irradiance level was used. This experiment is also the only one where ΔΦPSII remains larger than 0.01 after day 5 of increased irradiance.

3.5 RNA-seq analysis

To establish expression differences for the finemapped genes in the Q3 locus, B9-C11 and B9-E7 introgression lines (**Figure 2**) were grown as in van Rooijen *et al*. (2017) and harvested for RNA to conduct a transcriptome analysis of the candidate genes. A principal component analysis (PCA) was performed to validate the response of the full transcriptome to the introduction high irradiance (**Figure 12**). The HL2 samples are the most distance group, possibly indicating a circadian expression pattern affecting the transcriptome as previously observed in van Rooijen *et al*. (2018). The HL3 samples, taken on the second day of increased light irradiance, and the HL4 samples, taken on the fifth day of increased light irradiance, colocate in the PCA plot indicating a longer term acclimation to high light irradiance after the stepwise increase (HL1) (**Figure 12**). Throughout the experiment, the number of differentially expressed genes (DEGs) remaind the same, with the majority of DEGs associated with chromosome 3. This is consistent with the genetic layout of these genotypes, as only allelic variation is located on chromosome 3 in the introgression (**Figure 2**).

A total of 37 genes were considered to detect transcriptional differences, all which locate in the F2:4 - I fine-mapped QTL-region of chromosome 3 (**Figure 9c**). 18 of these are not expressed in the rosette tissues and were discarded (**Supplementary figure 5**). Considering the result of the finemapping (**Figure 9c**, **supplementary table 5**), nine genes are highlighted for differential expression. Furthermore, *LIPOXYGENASE 2* (*LOX2;* At3g45140) was included as this gene encodes a chloroplast transit peptide (Bell, 2001; Glauser *et al*., 2009) – a feature lacking in all nine candidate genes underlying the region identified by the F2:4 – II mapping experiment. Five candidate genes are differentially expressed ($p < 0.01$) in at least one time point during the treatment, although only at one time point, with the exception of *LOX2* and *TFIID* (At3g45210). Only expression for *LOX2* is significantly different in the response to high light.

Table 2. *Differentially expressed genes per time point and chromosome between the introgression lines B9-E7 and B9-C11,. "Day"; day of the experiment. "Photoperiod"; the time into the photoperiod when sampling of rosettes took place."C1-5" Chromosome one to five. For each chromosome, the total number of differentially expressed genes are given at p < 0.01.*

Time point	*Day*	*Irradiance*	*Photoperiod*	*C1*	*C2*	*C3*	*C4*	*C5*
LL1	1	100 µmol m^{-2} s^{-1}	+1h	15	13	136	5	15
HL1	2	500 µmol m^{-2} s^{-1}	+1h	30	16	129	19	18
HL2	2	500 µmol m^{-2} s^{-1}	+3.5h	46	23	145	22	34
HL3	3	500 µmol m^{-2} s^{-1}	+1h	13	11	120	2	9
HL4	6	500 µmol m^{-2} s^{-1}	+1h	18	8	129	6	18

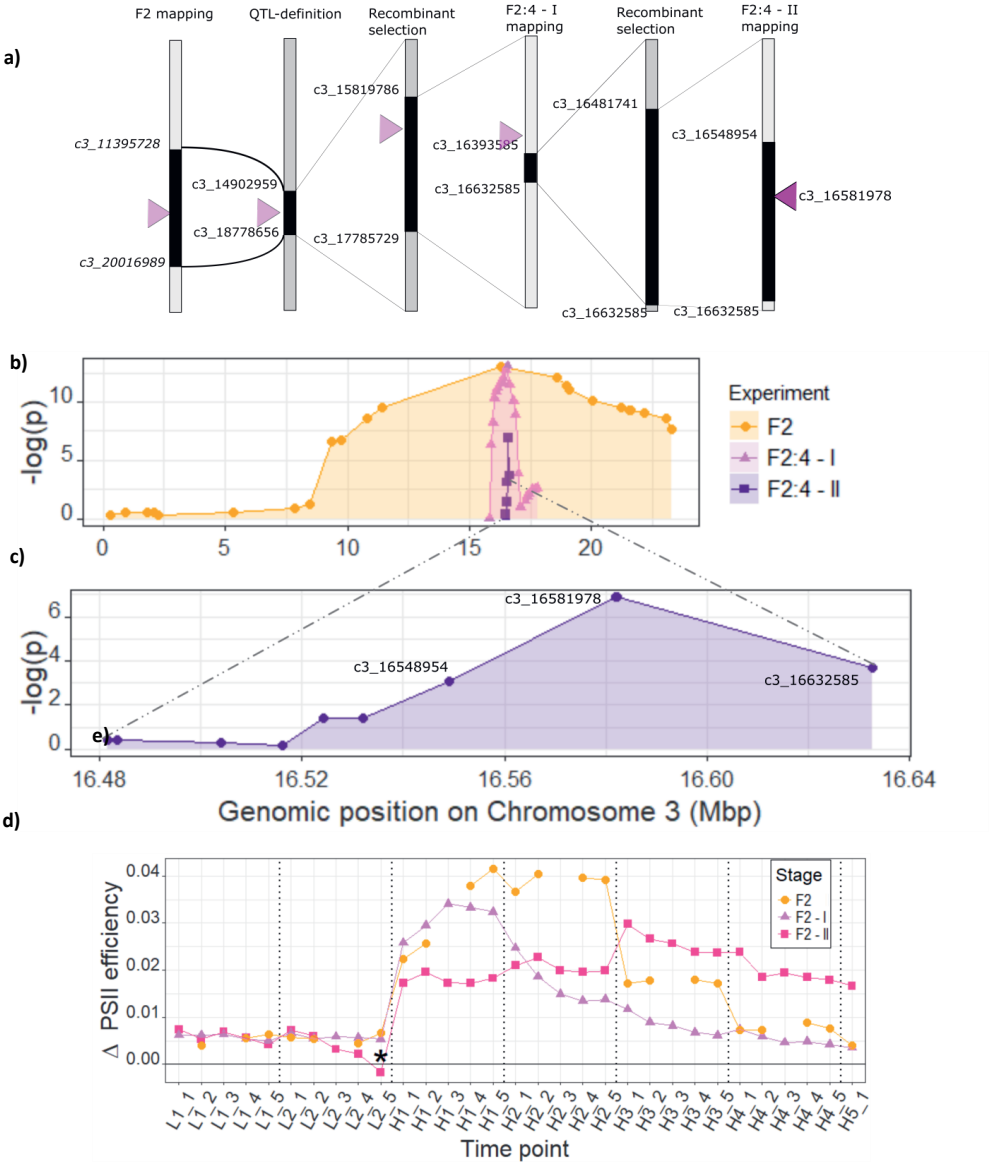


Figure 11. *Fine-mapping of the Q3 QTL on chromosome 3.* ***a)*** *Selection and mapping procedure on the Arabidopsis thaliana third chromosome. The grey bars represent the chromosomes or chromosomal regions, the lightgrey bars indicate the mapping experiments. The faint pink triangle points to the original QTL-location in the F2-mapping. The position of the final round of fine-mapping (F2:4-II) is indicated by an opaque pink triangle. Markers in italics are derived from the Illumina Golden Gate chip genotyping, while the other markers are KASP-probes as described in* ***supplementary table 1****.* ***b)*** *QTL plots of the maximum LOD-values during the high light phase for the different (fine-)mapping experiments, for F2, F2:4-I and F2:4-II experiments.* ***c)*** *The final delineation of the Q3 locus, including the peak and bordering markers for the QTL.* ***d)*** *The effect size differences of Q3 between the different finemapping experiments, positive values indicate superior S96 alleles. Genotypic means were calculated using at least 60 biological replicates per homozygous genotypic group in each experiment.*

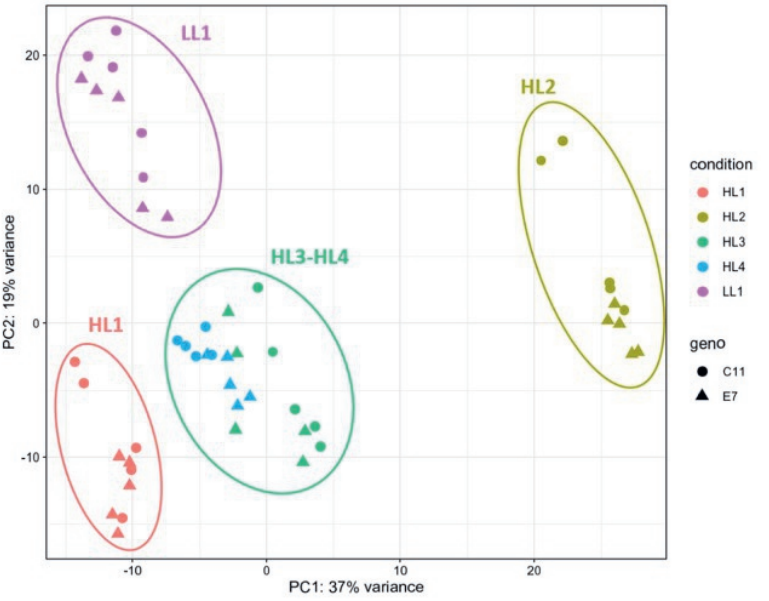


Figure 12. *Principal component analysis of differentially expressed genes between the B9-C11 and B9-E7 introgression lines, in response to a stepwise change of increased irradiance. Encircled are groups of samples belonging to the same treatment / time point. For time points, see* ***table 2***.

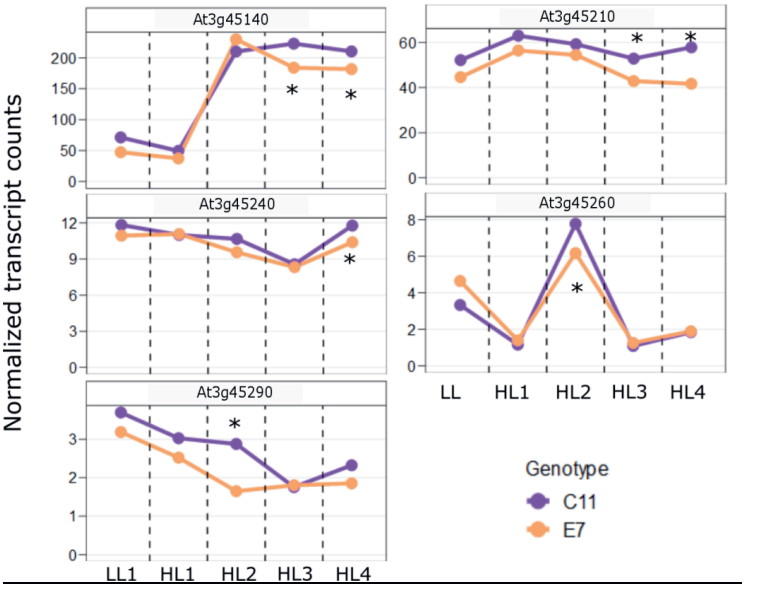


Figure 13. *Expression patterns of the five candidate genes differentially expressed belonging to the Q3 quantiative trait locus. Time points as described in* ***table 2***. **'s give time points with significant ($p < 0.01$) differential expression of the candidate genes between the B9-C11 and B9-E7 introgression lines.*

3.6 Mutant analysis and genetic variation in the candidate gene

A total of eleven knockout mutants were identified after genotyping (**Table 3**), with inserts disrupting genes in the fine-mapped Q3-region that are found to be (differentially) expressed following the RNA-sequencing experiment. In total, these mutants represent seven of the candidate genes listed following the fine-mapping experiments (**See supplementary table 5**), thus for some genes more than one mutant line could be verified. They were grown in Phenvoator to evaluate the impact of disruption these genes on ΦPSII. A knock out for *LYPOXYGENASE 2* (LOX2; At3g45140) was also included. Two mutant lines indeed show a deviating photosynthesis compared to the Col-0 control genotype in the initial screening, *lox2-1* and *tfiid-1* (**Figure 14a**) – a mutant for *TRANSCRIPTION INITIATION FACTOR SUBUNIT IID* (TFIID; At3g45210). Subsequent experimental repeats with higher number of replicates for *lox2-1* and *tfiid-1*, confirms a role for *TFIID* (**Figure 14b**), but not for *lox2-1* (**Figure 14c**). The phenotype of *tfiid-1* shows differences between compared to the Col-0 wild type at time points similar as those to show differences in the F2-mapping experiment, with an effect of ΦPSII present on day 1 and day 2 of high light irradiance (**Figure 11d**). *TRANSCRIPTION INITIATION FACTOR SUBUNIT IID* was further scrutinized for genetic variation in the sequence. Using snpeff (Cingolani *et al*., 2013), the 5”UTR, the exon and its 3”UTR were scrutinized for sequence polymorphisms. Several larger insertions are found to locate in the 3”UTR of the S96 allele, while the SLSP30 allele contains three minor insertions relative to S96.

Table 3. *Locations and short names of TDNA mutants used to characterize the candidate genes underlying Q3 that show detectable expression in the rosettes. “Location”; the gene component where the TDNA is located. For SALK_004101, no TDNA insertion could be found. “Homozygous”; a homozygous mutant line was found for the gene*

Mutant line	Type	Locus	Short	Location	Homozygous
LOX 2 EMS	EMS-mutant	At3g45140	*lox2-1*	SNP introducing Premature stopcodon	Y
SAIL 337 G05	T-DNA	At3g45190	*sit-1*	Exon	Y
SAIL_722_D01	T-DNA	At3g45190	*sit-2*	Intron	Y
SALK_144179	T-DNA	At3g45190	*sit-3*	Intron	Y
SALK 004101	T-DNA	At3g45210	*tfiid-2*	Exon	N
SALK_004102	T-DNA	At3g45210	*tfiid-1*	Exon	Y
SALK_096275	T-DNA	At3g45230	*apap-1*	Exon	Y
SALK 142938	T-DNA	At3g45240	*grik1-1*	intron	Y
SALK_204894	T-DNA	At3g45240	*grik1-2*	Intron	Y
SALK_206493	T-DNA	At3g45243	*scpl48-1*	Exon	Y
SALK_059260	T-DNA	At3g45260	*bib-1*	Intron	Y

Table 4. *Analysis of genetic variation in TRANSCRIPTION INITIATION FACTOR SUBUNIT IID (At3g45210). Position is given in base pairs, relative to Arabidopsis thaliana reference genome TAIR10.1. Sequence gives the part of the gene in which the SNP is located. R indicates which of the polymorphisms in S96 and SLSP30 are similar to the reference genome. 0 indicates SLSP30 has the reference allele, 1 indicates that S96 has the reference allele.*

Position	Sequence	SLSP30	S96	R
16556875	5”-UTR	C	T	0
16557409	5”-UTR	G	A	1
16557445	5”-UTR	AC	A	1
16557939	EXON1	G	A	1
16557954	EXON1	G	T	1
16558007	3”-UTR	T	C	0
16558064	3”-UTR	A	C	0
16558069	3”-UTR	A	T	0
16558127	3”-UTR	T	G	0
16558133	3”-UTR	A	G	0
16558142	3”-UTR	TA	T	0
16558164	3”-UTR	T	TGAGTGTTTTTTTGTTGA	0
16558168	3”-UTR	ACG	A	0
16558171	3”-UTR	G	GCTCATATCTACA	0

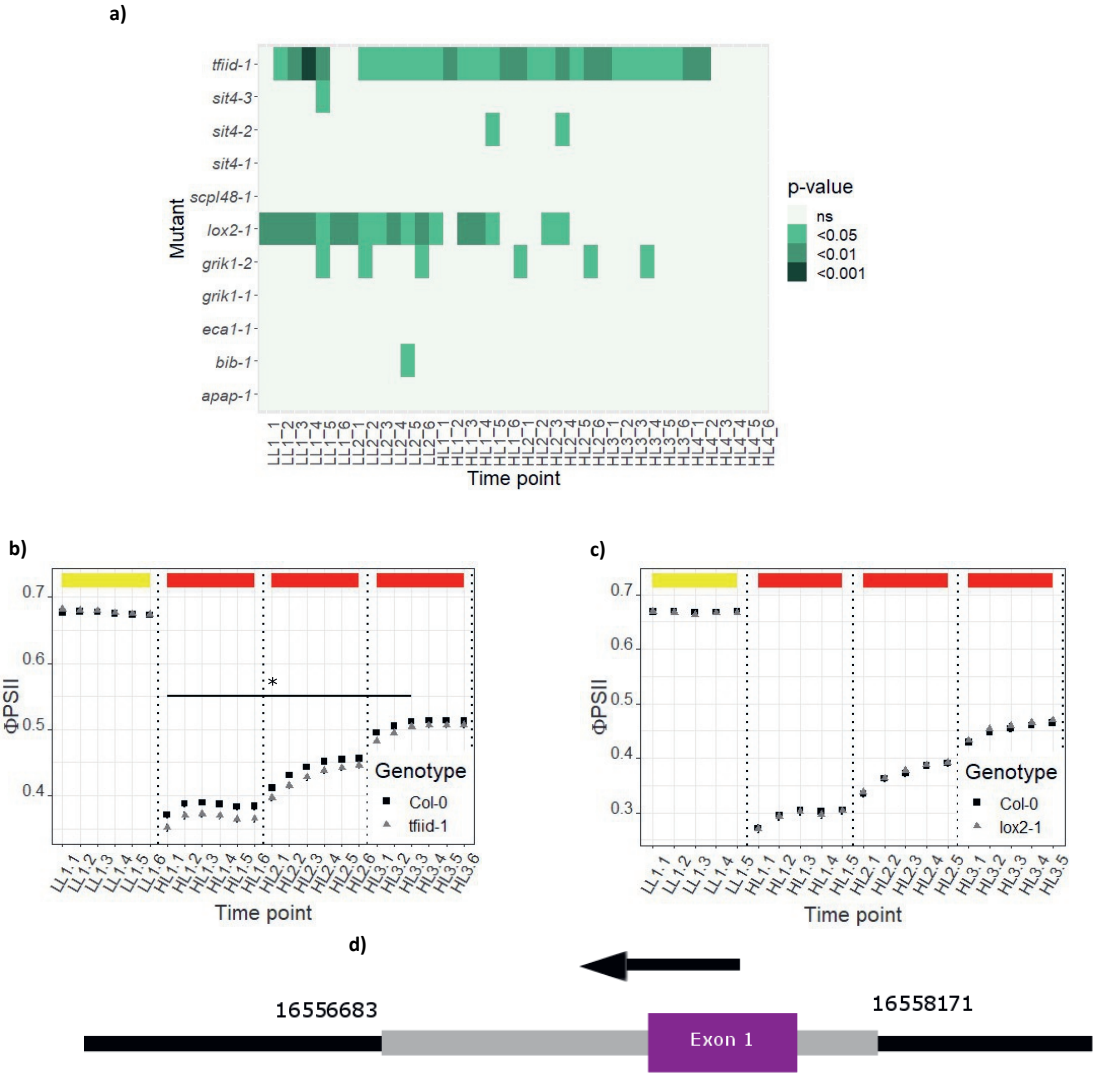


Figure 14. *Analysis of mutant lines underlying the Q3 QTL for high light acclimation.* ***a)*** *Time series analysis of T-DNA KO lines. p-values result from the analysis of a linear mixed models incorporating environmental factors of the T-DNA line against the Col-0 background line. On the x-axis, the time points are given as low light (LL) = 100 µmol m-2 s-1, high light (HL) = 450 µmol m-2 s-1. The first number after treatment indicates independent days and the specific time point of the measurement taken for ΦPSII; 1 = 1 h after start of the photoperiod, 2 = 2.5 h, 3 = 4 h, 4 = 5.5 h, 5 = 7 h, 6 = 8,5 h. n = 16-20 for all genotypes and the Col-0 control background line.* ***b)*** *Independent confirmation experiment for the tfiid-1 experiment, at 28 replicates for both genotypes (black bar indicates time points where $p < 0.05$.* ***c)*** *independent confirmation experiment for the lox2-1 mutant, at 28 replicates for both genotype.* ***d)*** *A gene model of At3g45210, TRANSCRIPTION FACTOR IID in its native, antisense orientation. In purple the exons, in grey the untranslated regions. The positions of the start and end of the transcribed region are indicated in the figure. The model is scaled to reflect to size of the different gene components.*

4. DISCUSSION

This study describes the identification of QTLs related to the photosynthetic acclimiation to high light. Two novel QTLs for high light acclimiation in *Arabidopsis thaliana* are described. Selection experiments were performed to show that selection for photosynthesis traits may reduce yield losses resulting from fluctuating light conditions in *Arabidopsis*. Furthermore, finemapping of the genes led to the identification *TRANSCRIPTION INITIATION FACTOR II D (TFIID)* (At3g45210), which is suggested as a candidate gene for the QTL on chromosome 3.

4.1 Genetic mapping of high light acclimation quantitative trait loci reveals two large effect size, high light acclimation QTLs

S96 and SLSP30 are two Arabidopsis genotypes that differ in photosynthesis efficiency in both low light and high light conditions (**Figure 4a-c**). Unlike in **Chapter 3**, where a single QTL determines differences in ΦPSII in a cross between Col-0 and Ler-0, no quantitative trait loci could be detected under stable growth light conditions in the F2 progeny of S96 and SLSP30 (**Figure 5a**). This indicates that copy number variation of *SQUALENE EPOXIDASEs* does not contribute to phenotypic differences between S96 and SLSP30 in the low light phase. Instead, the observed differences in photosynthesis efficiency between S96 and SLSP30 may be controlled by many smaller, undetectable loci (**Figure 5a**). The two QTL, Q3 and Q5, discovered after a stepwise increase to high light irradiance are found to be highly specific for the imposed treatment (**Figure 5a, b**). Q3 and Q5 express a similar phenotypic trajectory in the high light acclimation phase (**Figure 5c**), which may indicate that the underlying genes may be involved in a similar biological process. However, the total variance explained by both loci is relatively low with respect to the large effect size of the QTL (**Table 1**), which suggests that additional genetic factors may still play a significant role in the photosynthetic response to high light acclimation. These genetic components may be of little effect, however, and therefore masked by the large effect size of the two major QTL, Q3 and Q5.

The effect size of the two major QTL for high light acclimation of photosynthesis are individually much greater, at 10% for Q3 and 15% for Q5, than each of those described in a collection of 344 HapMap accessions recorded from a genome wide association studies (GWAS) by van Rooijen *et al*. (2017). This confirms the improved mapping power of biparental mapping populations over diversity panels (Tiang *et al*., 2021), which is relevant as mostly small effect size QTLs are in GWAS (van Bezouw *et al*., 2019). In each of the mapping F2-experiments a consistent and reliable response to high light is detected, characterized by high LOD-scores (**Figure 4a, c**). however, at the cost of a lower resolution for purposes of gene identification. The locations of the peak markers defining the Q3 and Q5 QTL do not overlap with the genomic positions of QTL as reported in van Rooijen *et al.* (2017). This may possibly indicate that between S96 and SLSP30 rare alleles for high light acclimiation are present for genes, which would otherwise remain undetectable in diversity panels due to low allele frequencies (Korte & Farlow, 2013; Bazakos *et al*., 2017).

4.2 Selection for the high light acclimation QTL Q3 and Q5 establishes a relationship between increased ϕPSII, decreased NPQ and increased productivity under fluctuating light conditions

Phenotypes for plant photosynthesis traits in genotypes grown in climate-controlled chambers under stable light regimes are rarely representative for those grown under fluctuating conditions that are often encountered in the open field (Mishra *et al*., 2012, Vialet-Chabrand *et al*., 2017, Slattery *et al*., 2019). Furthermore, genetic variation in photosynthetic adaptability had previously been hypothesized as being more promising to improve crop productivity than under constant conditions (Lawson et al., 2012). However, experiments aimed at exploring these genetic resources to test their potential are scarce (but see Athanasiou *et al*., 2010). This is particularly true for genes and traits associated with natural variation in photosynthesis that are derived from chlorophyll fluorescent imaging facilities (van Bezouw *et al*., 2019). To bridge this gap, the F2 population described here was utilized to select inbred material that allows the

assessment of the light acclimation QTL Q3 and Q5 in different combinations and to test their impact on plant productivity in different fluctuating light environments (**Figure 1, Supplementary figure 6**).

In the fluctuating light experiments, we found photosynthesis efficiency to have an inverse relationship with non-photochemical quenching, but a positive relationship with growth (**Figure 8-10**). Between introgressions lines B9-C11 and B9-E7 – differing for the Q3 QTL (**Figure 2**) – no significant differences are observed for growth (**Figure 9**). For the introgression families (IFs) (**Figure 9**) – which collectively still segregate in their background but with combinations of homozygous loci for the Q3 and Q5 QTL (**Supplementary figure 2**) – we did observe significant differences in productivity in the fluctuating light treatments. This is likely caused by the higher genetic variation in these lines due to the selection for two QTL rather than one. Counterintuitively, our data does suggest that maker-assisted selection for improved photosynthesis efficiency in response to high light in the Phenovator does not benefit productivity in Arabidopsis. Lines carrying a combination of weaker, SLSP30-alleles for both Q3 and Q5 (3L5L-lines) grow 10-20% larger in terms of rosette dry weight in the highly fluctuating light conditions compared to lines carrying S96 alleles for both QTL (S3S5-lines). In the Phenovator, IFs L3L5 exhibit at most a more than 20% lower ΦPSII following a stepwise increase in light intensity compared to the other IFs (**Figure 6a**), but L3L5 exhibits a slightly higher ΦPSII than IFs S3S5 in the fluctuating light experiment. This apparent reverse correlation is surprising as we expected that improved ΦPSII as observed in Phenovator would be correlated with increased growth. A possible explanation for these differences can be found in the fundamentally different nature of the genetic mapping and fluctuating light experiments.

In the genetic mapping experiments in the Phenovator, only a single stepwise change is introduced and plants were measured as the high light acclimation response resolved. In the fluctuating light experiments, plants were repeatedly exposed to changes in irradiance during the entire growth period after which a single measurement was taken in plants that were acclimatized to normalize the measurements. Plants are known to acclimatize and adapt their photosynthesis apparatus to fluctuating light conditions (Vialet-Chabrand *et al*., 2017; Schumann *et al*. 2017). This adaptation also increases the speed by which adaptations to a stepwise change to long-term high light irradiance are made (Schumann *et al*. 2017). The difference in the biological memory may explain these differences in the measurements of photosynthesis efficiency between the experiments. We obtained further insight in the photosynthetic acclimation in the analysis of total non-photochemical quenching (in this work measured as NPQ_t, Tietz *et al*., 2017), and its fast energy-dependent (qE_t) and slow photo-inhibitory (qI_t) components. While NPQ_t is the same for all three fluctuating light conditions (24 h, 60 min, 5), fluctuating light treatment spanning full days exhibited elevated photo-inhibitory components – rather than elevated qE_t as in the 60 min and 5 min treatments. This indicates that a different form of photosynthetic adaptation takes place in the 24 h treatment. These experiments at least showed that it may be possible to selet for photosynthesis QTLs leading to improved growth performance under photosynthetically challenging conditions. However, the exact roles of these QTLs in these observations requires a more indepth study of the molecular components.

4.3 A mismatch between nocturnal programming of photosynthetic acclimation and future incoming irradiance triggers photoinhibition and introduces larger biomass losses than rapid fluctuating light regimes

In the Phenovator, repeated measurements of ΦPSII indicate that the long term high light acclimation of photosystem II efficiency is characterized by elevated levels of ΦPSII overnight, as the difference in ΦPSII shows a near continuous relationship regardless of day or night time (as previously described in van Rooijen *et al*., 2017, 2018, see **Figure 5c, 6a, 6b**). This implies that in night time, biological processes are at play that enhance the photosynthetic capacity of plants to light to be more efficient during the next photoperiod. In their review, Morales & Kaiser (2020) describe studies aimed at understanding plant physiology under influence of fluctuating light to be classified in two broad categories; 1) those that seek to study the plants response to diurnal changes in irradiance and 2) those that use repeated stepwise

changes to trigger photoprotective mechanisms. Day to day fluctuations, as presented in the 24 h fluctuating light treatment, thus presents an underexplored scale of light fluctuation intervals. The long-term light fluctuations introduced here reduce biomass growth by more than 40% compared to stable conditions; higher than the treatments with higher frequencies of stepwise changes in irradiance that we exposed our material to (**Figure 9a**). The large scale and highly dynamic nature of the light regimes that we exposed our plants to prevent the evaluation of photosynthesis in all plants *in situ.* However, the acclimatory effects of longitudinal growth under conditions of fluctuating light are still detectable in *ex situ* measurements using PlantScreen^{tm}, showing elevated qI_t – but not qE_t – compared to the stable, control light conditions. This shows that under the 24 h treatment longer term biological changes are more important than in the faster fluctuating light treatments. Although it is impossible to speculate from the data in this work as deeper physiological or molecular work is lacking, long-term acclimiation responses as described here may hold promise to further improve field crop photosynthesis.

4.4 Selection for the high light acclimation QTL Q3 and Q5 establishes a relationship between increased ϕPSII, decreased NPQ and increased productivity under fluctuating light conditions in Arabidopsis

Transgenic improvement of photosynthesis has seen many recent examples (Driever *et al*., 2107; South *et al*., 2019; Yoon *et al*., 2020), but the earliest of such have involved transgenic alteration of total NPQ (Kromdijk *et al*., 2016). The main component of this alteration includes increased adaptation of the expression level of *PHOTOSYSTEM SUBUNIT S* (Jung & Niyogi, 2009) and enzymes involved in the regulation of the Xanthophyll cycle (Latowski *et al*., 2011), which in concert are important regulators of photoprotection in plants. Tobacco plants transformed in this way exhibit higher ΦPSII and reduced total NPQ, leading to a more efficient photosynthesis efficiency, which are associated with improved growth in field conditions (Kromdijk *et al*., 2016). In Arabidopsis, acceleration of these photoprotective properties introduces similar effects on ΦPSII and NPQ, but an unexpected reduction in growth was observed by Garcia-Molinia & Leister (2020). In this study, selection for the natural alleles of SLSP30 for both the Q3 and Q5 do correlate with increased ΦPSII (**Figure 9b**), decreased NPQ (**Figure 9c**) and increased productivity (**Figure 9a**) under fast fluctuating light conditions, while under stable light conditions no differences in growth are observed (**Figure 9a**). This combination of properties does match Kromdijk *et al.* (2016), demonstrating that also in *Arabidopsis thaliana* improved photosynthesis efficiency can be achieved through a modulation of NPQ – through the molecular mechanisms may be different what has been achieved using transgenic approaches .

It should be noted, though, that the introgression families have been selected for genetic markers that still represent larger genomic sections – particularly for QTL Q5 – which may impact productivity under fluctuating light conditions. However, careful selection of segregating lines ensured that mostly random backgrounds were established in each of the introgression families (**Supplementary figure 2**). This selection was effective to diminish the growth advantage observed between the inbred S96 and SLSP30 parental lines under stable light conditions (**Figure 7a, b, 9a**), while retaining the stronger differentiating effects on photosnythesis efficiency acclimation (**Figure 6a**). Furthermore, IFs 3S5L and 3L5S predominantly take intermediate phenotypes for traits under most conditions (**Figure 9**), which is in line with our expectations as S96 (S) always has the stronger alleles. To conclude, were are able to reduce selection bias for biomass accumulation that may have arisen from genetic yield components in the background of our lines that could alter our conclusions. Thus, our results will – at least predominantly –be attributed by selection on these QTL, although the genetic composition and exact molecular pathways remain elusive. This example shows that it is possible to exploit natural variation for photosynthesis traits to achieve similar outcomes as can be achieved in transgenic studies (Theeuwen *et al*., 2022).

4.5 The genetic dissection of one high light acclimation QTL points to a possible involvement for a mitochondrial TRANSCRIPTION INITIATION FACTOR II D (TFIID)

Finemapping and functional characterization of candidate genes underlying QTL is important for purposes of transferring traits between species (Theeuwen *et al.*, 2022). The high effect size of the light acclimation QTLs and their possible role in improved biomass under fluctuating light conditions – as can be perceived by fieldgrown crops (Kaiser *et al.*, 2018) – are particularly interesting targets. In this study, we several lines of evidence to point towards a role for genetic variation in a mitochondrial transcription factor to contribute to high light acclimation (**Figure 13, 14b, d**, **Table 4**). The involvement of this gene may not be evident by gene transcription alone, as the phenotypes of *tfiid-1* (**Figure 14b**) and the associated Q3 QTL (**Figure 14d**) are shown earlier in response to high light than the differences in transcripts between the S96 and SLSP30 alleles (**Figure 11d**). However, genetic variation in the *tfiid-1* appears to be more associated with the untranslated regions (Mayr, 2019) (**Table 4**). Genetic variation in the UTRs in *SQE5* (**Chapter 3**) and *XND-1* (Tang *et al*., 2018) was previously associated with phenotypic variation.

TFIIDs are essential components in RNA polymerase and are required in the recognition of promoters (Patel *et al*., 2019). At3g45210 is described as having an unknown function (**Supplementary table 5**), with the resulting protein locating to the mitochondria. One highlighted feature is the DUF458 domain, for which in Arabidopsis there are fifteen gene members of the gene (Fischer-Kilbienski *et al*., 2010). Umbach *et al.* (2005) previously associated expression of the gene with oxidative stress. Mitochondria are a less well studied in the context of photosynthesis, however, they act to function as maintaining biochemical balances in the cell during photosynthesis (Gardeström & Lernmark, 1995; Igamberdiev *et al*., 2006; Gardeström & Igamberdiev, 2016). A stepwise increase in irradiance introduces the formation of reactive oxygen species (ROS), particularly if the photosynthesis machinery is not adapted to process the excess of light (Cloudhury *et al*., 2017). Prolonged damaging can result in reduced photosynthetic activity (Long *et al*., 1994, 1996). Mitochondrial activity can scavenge the production of ROS, thus preventing damage in the cell and maintainting photosynthesis function. Our data aligns with the suggestion that *TFIID* may be involved in this process (**Figure 9d**, **12a, b**), and may potentially be involved in the transcription of ROS-scavenging genes in the mitochondria. However, conclusive evidence for the role of this particular *TFIID* gene must come from a cloning approach as described in **Chapter 3** and a deeper functional analysis of the gene.

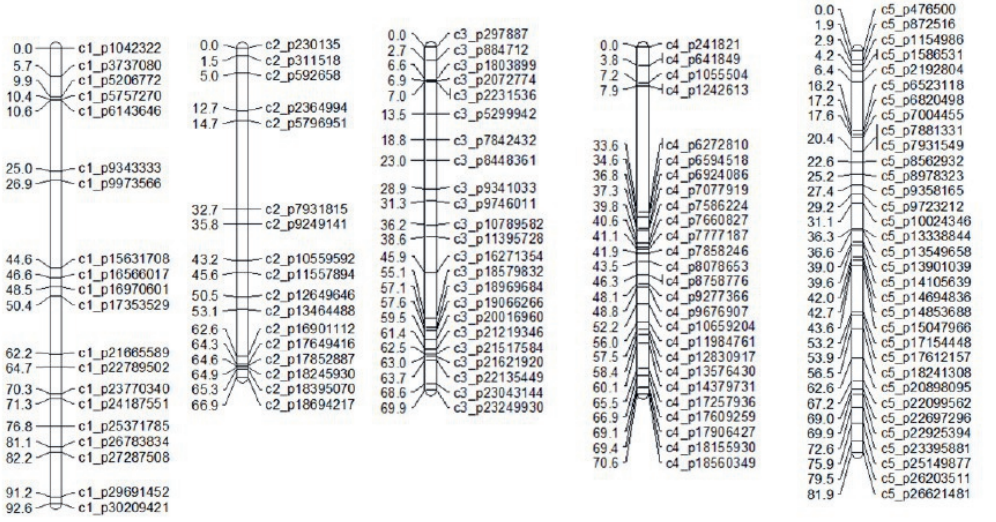
5. CONCLUSION

In this work we aimed to describe the genetic architecture of the high light acclimation response in two Arabidopsis accessions with contrasting properties for this trait. Two QTL were characterized and used as a basis for genetic dissection and assessment of productivity under fluctuating light conditions. We showed that genetic selection for the peak markers for these high light acclimiation QTL results in genotypic material that is more tolerant to losses in productivity under fluctuating light conditions. This study shows that photosynthetic responses to high light can be driven by strong genetic components. Genetic dissection and confirmation of the underlying candidate genes may reveal genes that can mediate the transfer of the here described traits to crop species.

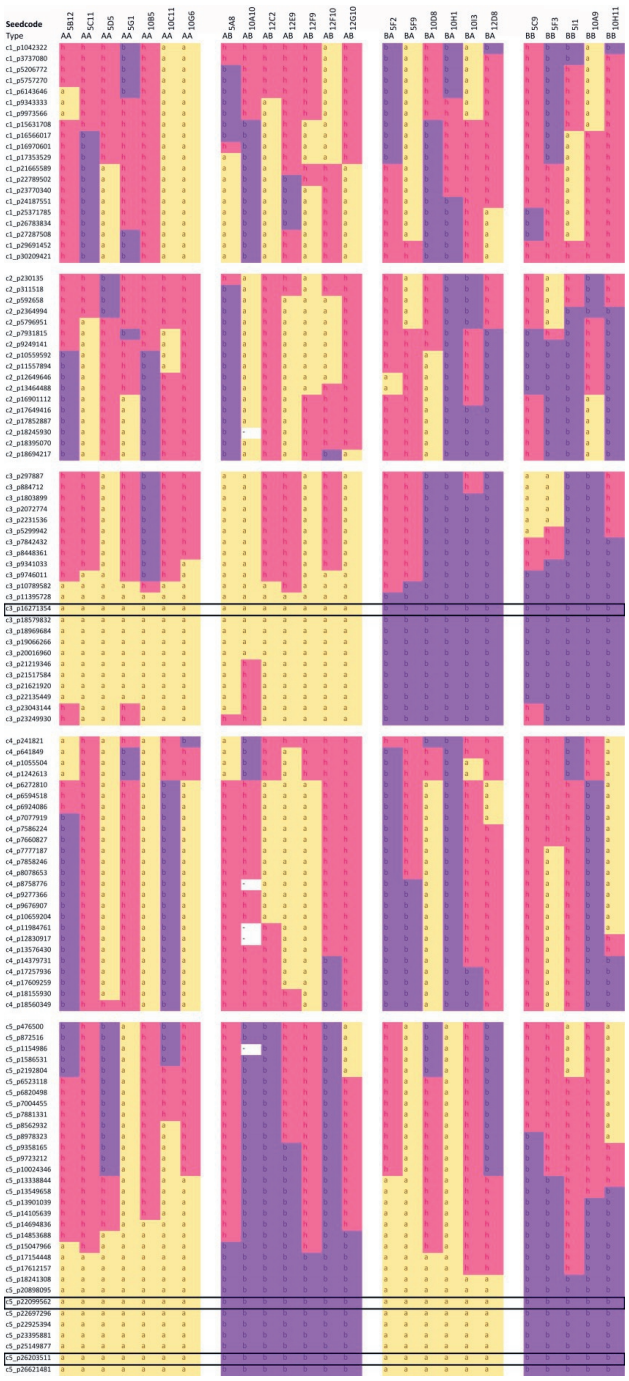
6. ACKNOWLEDGEMENTS

We would like to thank Gerrit Stunnenberg, David Brink, Taede Stoker for caring of our plants. Thim Harmsen, Yanrong Gao and Suzanne Vlinderveld for sowing the plants. We would also like to thank Louise Logie for discussions on operation of the PlantScreen^{tm}, and Mats Andersson and Edward Farmers for their kind donation of the *lox2-1* mutant. In addition, we acknowledge Raul Wijfjes for his help on the SLSP30 and S96 genomes. We also would like to thank Joost van den Heuvel for his help in processing of the transriptome. This research is funded by a generous grant from Rijk Zwaan B.V..

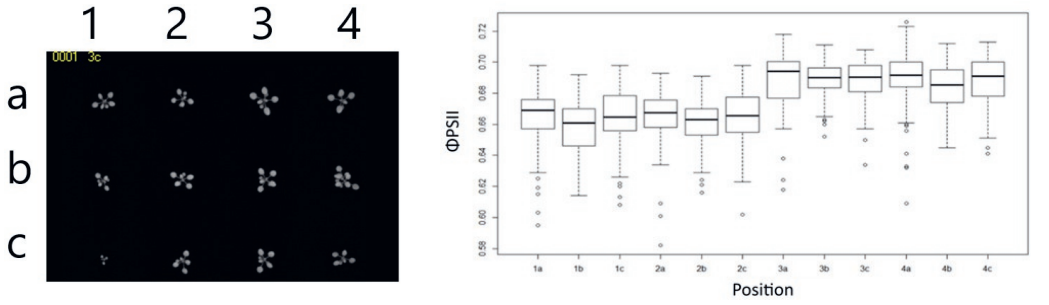
SUPPLEMENTARY FILES



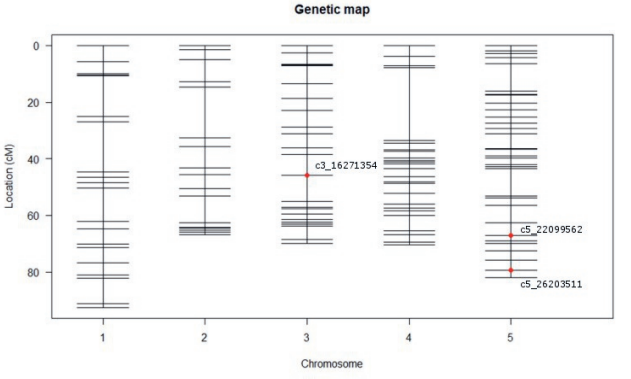
Supplementary figure 1. Genetic map of the F2 population (n = 304) developed from a cross between the Arabidopsis accessions S96 and SLSP30. Marker names indicate chromosome number and position in basepairs (following the TAIR10.1 Col-0 genome for reference). The genetic map was previously published in van Rooijen (2016).



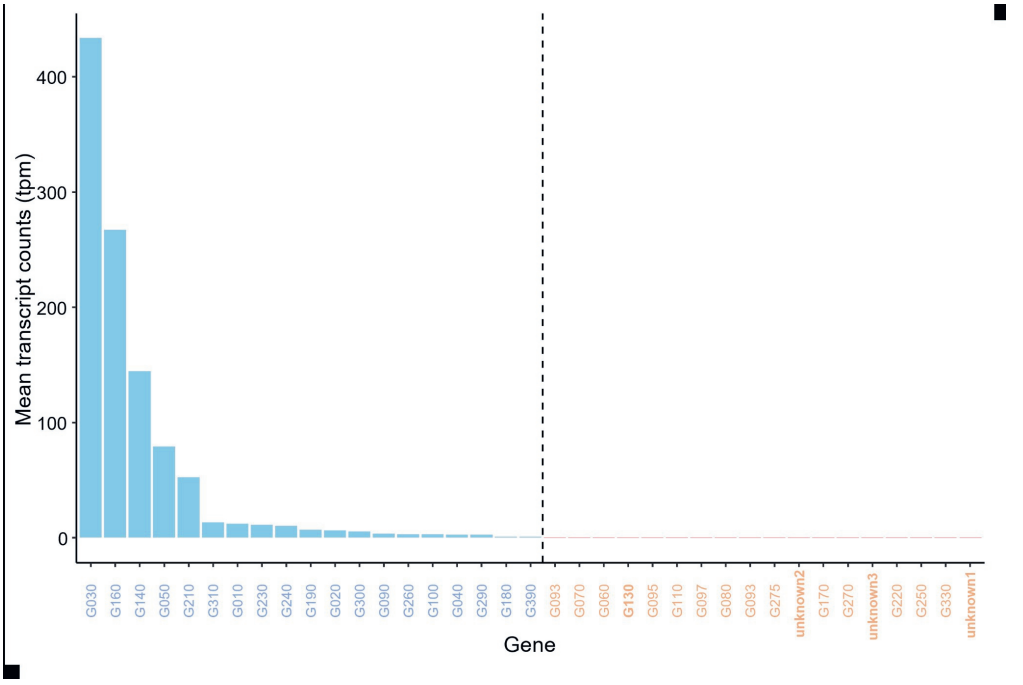
Supplementary figure 2. Genytopic profile of the introgression families (IFs) in the F2 generation. Markers are all derived from the Illumina Golden Gate chip. In black lines, the co-factors representing the Q3 (c3_1671354) and Q5 (c5_22099562, c5_26203511) QTL are highlighted. Yellow (A) = S96, purple (B) = SLSP30, pink = heterozygotes.



Supplementary figure 3. The cam.block parameter. a) Coordinate system within the camera head of the Phenovator (see Flood *et al*., 2016)**.** b) This data represents ΦPSII data at 09:00 of the 14th day of growth for 1440 replicates of Col-0 genotype grown in the Phenovator in a uniformity trial using the Col-0 genotype (Flood *et al*., 2016). Positions refers to those within the camera head (a), where numbers represent columns and letters represent rows. Phenotypes are constituted by ~100 replicates per individual camera head position.



Supplementary figure 4. Cofactor selection for multiple QTL mapping (MQM) of the F2 population. The red dots indicate the marker positions on the genetic map that were chosen as cofactors in the analysis.



Supplementary figure 5: Average transcription count per time point of candidate genes for Q3 in the F2:4 – I experiment. Gene names on the X-axis are abbreviations for genetic loci following TAIR10.1. All genes, At3g45**030**-At3g45**330** are presented, with red genes being expressed at an average of < 1 transcript per time point in the experimental conditions. Notable genes; G140 = *LOX2*, G210 = *TFIID*.

Supplementary table 1. The list of markers used in this study and their sequences. The marker names indicate chromosome and position (ref: TAIR10.0). Polymorphism: the first letter indicates the S96 (FAM-)allele and the second letter indicates the SLSP30 (HEX-)allele. Letters always follow the 5"-3" genome orientation. Screening; here an x indicates the use of the marker for screening of recombinants. Forward and reverse indicate the forward sequence in basepairs used for genetic screening of the segregating population. The order FAM and HEX sequences are in line with the indicated polymorphism.

Marker	Polymorphism	Screening-I	F2:4-I	Screening II	F2:4-II
c3_15819786	T-C	x	x		
c3_15885951	A-C		x		
c3_15960527	A-G		x		
c3_16017695	C-T		x		
c3_16100491	A-T		x		
c3_16170043	A-T		x		
c3_16230514	T-G		x		
c3_16299252	C-A	x	x		
c3_16393585	G-C		x	x	x
c3_16481741	C-A				x
c3_16483479	G-A				x
c3_16504019	T-C		x	x	x
c3_16516084	A-G				x
c3_16524164	A-T				x
c3_16531977	C-T				x
c3_16548954	C-G		x	x	x
c3_16581978	T-C				x
c3_ 16582104	A-C				x
c3_16632585	C-T		x	x	x
c3_16793672	T-C		x		
c3_16890226	A-G		x		
c3_16983646	A-C	x	x		
c3_17092735	C-A		x		
c3_17289352	A-G		x		
c3_17371446	A-G		x		
c3_17445561	C-G	x	x		
c3_17595550	G-A		x		
c3_17655256	C-T		x		
c3_17719568	T-C		x		
c3_17785729	T-A	x	x		

Supplementary table 2. Primrs for TDNA and the LOX2 mutants. The primers for the LOX2 RNAi and LOX 2 EMS mutants were used for RT-PCR and sequencing, respectively.

Locus	Primer sequences	
Salk LBb1.3	attttgccgatttcggaac	
SAIL LB1	gcctttcagaaatggataaatagccttgcttcc	
	LP	**RP**
SALK_059260	ATCCGATTGAACCCGAATAAC	CGATTACTGCTAGCCAAGGTG
SAIL_337_G05	ATGTTTTCCAGCAACTGGTTG	TCAGCAGCAAAAACAAGAATG
SALK_144179	TCTCCAATTTCTGTGAATCCG	ATGCGCAAAACAATTCATACC
SALK_004101	TGGACAATGCAGATGTCAAAC	TCCTCAACCACAAACAAAAGG
SALK_004102	TGGACAATGCAGATGTCAAAC	TCCTCAACCACAAACAAAAGG
SALK_142938	CTGTTTTGCAGAAGCTTGGAG	AACCCGTAGCCTCGATAAATG
SALK_096275	TATGCTGAAATTGGCTAACGG	AAGTCTCGTACTCGTCTCCGG
SALK_204894	TCATCCTCTGGTTTGGATCTG	TGGCTTTTTCCAGTACAATGC
SALK_206493	AGGGCATTTTTGAAAAAGAGG	TTCATATTGCATCCCAAAAGG
SAIL_722_D01	TGAAACACTGTGCACTATCGC	CAATCTAAGCAAGCCATCAGC
35S-LOX2 RNAi	GCCTCGTCCAAACCTCAGAA	AGGGGTCAGTCTCACAAGGT
LOX 2 EMS	GGATTATCATGATTTGCTTCTACC	TCAAATAGAAATACTATAAGGAACAC

Supplementary table 3. Quality control parameters for all samples. The five best samples were chosen out of a total of six samples, with the remainder discarded, per time point per genotype.

Sample	Raw reads	Raw data	Effective(%)	Error(%)	Q20(%)	Q30(%)	GC(%)
LL_E7_6	42059754	6.3	98.99	0.02	98.2	94.48	46.36
LL_E7_4	44219054	6.6	98.58	0.02	98.09	94.23	46.47
LL_E7_3	43691020	6.6	99.15	0.02	98.26	94.61	46.73
LL_E7_2	44828362	6.7	98.49	0.02	98.19	94.47	46.37
LL_E7_1	44125270	6.6	98.72	0.03	98.05	94.13	46.45
LL_C11_5	42387604	6.4	98.61	0.02	98.12	94.28	46.47
LL_C11_4	40579396	6.1	98.72	0.02	98.15	94.35	46.35
LL_C11_3	43113896	6.5	98.8	0.02	98.08	94.17	46.03
LL_C11_2	42885282	6.4	98.77	0.02	98.24	94.56	46.24
LL_C11_1	44080340	6.6	98.81	0.02	98.16	94.36	45.83
HL4_E7_6	42268486	6.3	98.24	0.02	98.27	94.61	46.19
HL4_E7_5	40587846	6.1	98.82	0.02	98.29	94.66	46.26
HL4_E7_3	42310260	6.3	99.05	0.02	98.21	94.51	46.25
HL4_E7_2	42822786	6.4	99.02	0.02	98.15	94.3	46.14
HL4_E7_1	48231858	7.2	99.08	0.02	98.16	94.34	46.08
HL4_C11_6	42175152	6.3	99.2	0.02	98.25	94.57	46.1
HL4_C11_5	42137288	6.3	99.12	0.02	98.25	94.53	46.18
HL4_C11_4	44500202	6.7	98.88	0.02	98.22	94.47	46.01
HL4_C11_2	43493030	6.5	99.22	0.02	98.25	94.56	46.21
HL4_C11_1	43068758	6.5	99.04	0.02	98.23	94.49	45.95
HL3_E7_6	53843618	8.1	98.81	0.02	98.11	94.22	46.09
HL3_E7_4	42200046	6.3	99.21	0.02	98.17	94.35	46.06
HL3_E7_3	42264226	6.3	99.12	0.02	98.31	94.73	46
HL3_E7_2	39689706	6	98.73	0.03	98.06	94.11	45.92
HL3_E7_1	41559800	6.2	98.72	0.02	98.1	94.22	46.16
HL3_C11_6	42600848	6.4	98.01	0.03	98	93.95	46.01
HL3_C11_5	48925052	7.3	98.41	0.02	98.25	94.61	46.06
HL3_C11_3	44419490	6.7	98.87	0.02	98.11	94.24	46.1
HL3_C11_2	46241834	6.9	98.62	0.02	98.11	94.25	46.19
HL3_C11_1	39884946	6	97.95	0.02	98.1	94.22	45.79
HL2_E7_5	44644130	6.7	98.25	0.02	98.31	94.74	45.99
HL2_E7_4	41350660	6.2	98.41	0.02	98.17	94.38	45.46
HL2_E7_3	45097084	6.8	98.52	0.02	98.35	94.85	45.94
HL2_E7_2	42548298	6.4	98.13	0.02	98.24	94.55	46.17
HL2_E7_1	42462392	6.4	98.51	0.02	98.34	94.78	45.96
HL2_C11_6	37470388	5.6	98.2	0.02	98.21	94.5	46.08
HL2_C11_5	40729658	6.1	97	0.02	98.13	94.26	45.99
HL2_C11_4	38081860	5.7	98.33	0.02	98.17	94.39	46.12
HL2_C11_2	41045900	6.2	98.14	0.02	98.22	94.54	45.86
HL2_C11_1	40672390	6.1	98.83	0.03	97.93	93.86	45.97
HL1_E7_6	43191822	6.5	98.32	0.02	98.23	94.52	45.65
HL1_E7_5	42819996	6.4	98.63	0.02	98.35	94.85	46.4
HL1_E7_4	49019728	7.4	99.11	0.02	98.22	94.46	46.56
HL1_E7_3	45212240	6.8	98.85	0.02	98.09	94.19	46.49
HL1_E7_2	43897390	6.6	98.92	0.02	98.17	94.41	45.98
HL1_C11_6	42273610	6.3	98.75	0.03	97.94	93.86	46.25
HL1_C11_5	41340120	6.2	99.06	0.03	98	93.97	46.28
HL1_C11_4	45392812	6.8	98.91	0.02	98.16	94.35	46.05
HL1_C11_3	37138128	5.6	98.32	0.02	98.11	94.27	46.21
HL1_C11_1	44493676	6.7	98.89	0.02	98.2	94.42	45.83

Supplementary table 4. Epistatic analysis for three representative time points in the initial S96 x SLSP30 F2-mapping for photosynthesis using the scantwo() function in R/qtl (Broman *et al*. 2003). The numbers give the highest –log(p) score of the interaction term between two markers located on the designated chromosomes (cX:cX). The threshold was calculated using the permutation threshold procedure in rQTL and was set at 3.9 for epistatic interactions. Time points are as described in **figure 4**.

Time point	c1:c1	c1:c2	c1:c3	c1:c4	c1:c5	c2:c2	c2:c3	c2:c4	c2:c5	c3:c3	c3:c4	c3:c5	c4:c4	c4:c5	c5:c5
LL1.1	1.327	1.300	3.995	1.645	2.832	1.697	2.207	2.403	2.054	2.719	1.173	3.831	0.524	1.856	0.951
HL1.1	1.897	2.208	1.647	1.623	1.762	0.887	2.036	1.521	1.187	1.413	2.885	2.967	2.058	1.446	1.450
HL2.1	0.908	2.224	2.493	1.661	1.414	1.565	1.275	0.919	2.020	1.696	2.926	2.095	1.703	2.412	0.906

Supplementary table 5. Table generated by the biomart tool in Plant ensemble (14-05-2020). "Locus name"; refers to the TAIR10.1 definition, "gene start" and "gene end" refer to the location of the open reading frame and is always defined in 5" to 3" orientation, regardless of the orientation of the gene itself. "Name" refers to the gene name – if available. "Description" refers to the protein name as given by UniProtKB and TrEMBL accessions. "GO terms" are those as assigned in TAIR10.1.

Locus name	Gene start	Gene end	Name	Descdription	GO terms
AT3G45190	16541615	16548367	SIT4	SIT4 phosphatase-associated family protein [Source:UniProtKB/ TrEMBL;Acc:F4J5I0]	regulation of phosphoprotein phosphatase activity protein phosphatase binding biological_process regulation of phosphoprotein phosphatase activity protein phosphatase regulator activity regulation of phosphoprotein phosphatase activity protein phosphatase binding
AT3G45200	16548828	16549730		Uncharacterized protein T14D3.140 [Source:UniProtKB/ TrEMBL;Acc:Q9M1T9]	cellular_component biological_process molecular_function
AT3G45210	16556684	16558169	TFIID	Transcription initiation factor II D AT3g45210/T14D3_150 [Source:UniProtKB/ TrEMBL;Acc:Q9M1T8]	
AT3G45220	16564881	16566330	Z4	Serpin-Z4 [Source:UniProtKB/ Swiss-Prot;Acc:Q9M1T7]	negative regulation of endopeptidase activity extracellular space peptidase inhibitor activity negative regulation of peptidase activity extracellular space serine-type endopeptidase inhibitor activity serine-type endopeptidase inhibitor activity
AT3G45230	16569051	16569860	APAP	Hydroxyproline-rich glycoprotein family protein [Source:UniProtKB/ TrEMBL;Acc:Q9M1T6]	plant-type cell wall biological_process molecular_function membrane integral component of membrane
AT3G45240	16570478	16573762	GRIK1	geminivirus Rep interacting kinase Serine/threonine-protein kinase GRIK1 [Source:UniProtKB/ Swiss-Prot;Acc:Q93V58]	response to virus viral process protein autophosphorylation transferase activity nucleotide binding ATP binding nucleus kinase activity kinase activity kinase activity phosphorylation protein kinase activity protein phosphorylation protein serine/threonine kinase activity protein serine/threonine kinase activity cytoplasm

					protein binding protein phosphorylation cytoplasm protein serine/threonine kinase activity intracellular signal transduction nucleotide binding ATP binding kinase activity phosphorylation protein kinase activity protein phosphorylation protein serine/threonine kinase activity response to virus viral process protein autophosphorylation transferase activity nucleotide binding ATP binding nucleus kinase activity kinase activity kinase activity phosphorylation protein kinase activity protein phosphorylation protein serine/threonine kinase activity protein serine/threonine kinase activity cytoplasm protein binding protein phosphorylation cytoplasm protein serine/threonine kinase activity intracellular signal transduction
AT3G45243	16575823	16576173	SCPL48	ECA1 gametogenesis related family protein [Source:UniProtKB/ TrEMBL;Acc:A8MSE2]	
AT3G45260	16596358	16598811	BIB	Protein indeterminate-domain 9 [Source:UniProtKB/ Swiss-Prot;Acc:Q944L3]	protein localization to nucleus regulation of meristem growth nucleic acid binding metal ion binding DNA binding nucleus nucleus nucleus DNA-binding transcription factor activity regulation of transcription, DNA-templated protein binding sequence-specific DNA binding DNA-binding transcription factor activity positive regulation of transcription, DNA-templated nucleus DNA-binding transcription factor activity protein localization to nucleus regulation of meristem growth nucleic acid binding metal ion binding DNA binding nucleus nucleus nucleus DNA-binding transcription factor activity regulation of transcription, DNA-templated protein binding sequence-specific DNA binding DNA-binding transcription factor activity

					positive regulation of transcription, DNA-templated nucleus DNA-binding transcription factor activity
AT3G45275	16605781	16606238	ECA1	ECA1 gametogenesis related family protein [Source:UniProtKB/ TrEMBL;Acc:A8MS30]	membrane integral component of membrane

Chapter 5. Genetic variation in the response to fluctuating light conditions in an *Arabidopsis thaliana* chromosome substitution line population.

Roel F.H.M. van Bezouw[1], Ramon Botet Vaca[1], Sietske Bijlsma[1], Bastiaan de Snoo[2], Jitpanu Jamyabok[1], Elias Kaiser[3], Jeremy Harbinson[4], Joost J.B. Keurentjes[1].

SUMMARY

In field conditions light availability to plants is not a constant factor, but dynamically changes on a variety of scales; ranging from an order of seconds to days and longer. Plants need to constantly adapt the the state of the photosynthesis machinery to avoid damage from excess light and overinvestment in the photosynthesis apparatus leading to wastes of resources that could be used in growth. Prolonged exposure to changes in light intensity eventually results in yield losses. In this study, we describe the development of a novel chromosome substitution line (CSL-)population from two *Arabidopsis thaliana* accessions differing in their productivity response to fluctuating light conditions, but that are similar in photosynthesis efficiency. This population was used to explore the role of genetic variation plant physiology and productivity in response to fluctuating light. Significant variation in the response to fluctuating light conditions is revealed among the CSL-lines. An F1 hybrid developed from these two parental accessions exhibits both increased tolerance to fluctuating light, which suggests a role for heterosis contributing to this trait. To conclude, this work describes a new approach to study genetic variation for the productivity of plants under fluctuating light regimes.

---

[1]Laboratory of Genetics, Wageningen University, [2]Rijk Zwaan Breeding B.V., [3]Horticulture and Product Physiology, Wageningen Univeristy, [4]Laboratory of Biophysics, Wageningen University

1. INTRODUCTION

The rate of photosynthesis in plants is crucial to their productivity, but it also represents a major investment by the plant in terms of its assets and resources. These include mineral nutrients that have to be won from the soil and then assimilated, and carbon skeletons acquired via photosynthesis. As a result the photosynthetic properties of plants, such as, for example, their irradiance response curves, are a balance between the costs and benefits - a process of optimization. The photosynthesis machinery, therefore, naturally adapts to growth irradiance in order to cost-effectively use those limiting resources, such as light energy, and mineral nutrients (especially nitrogen) for photosynthesis. However, optimal rates of photosynthesis may be difficult to maintain in field conditions where light availability changes rapidly and unpredictably at different timescale as a result of, e.g., leaf movement in the wind, moving clouds, the movement of the sun across the horizon and climatic events such as alternating periods of, e.g., cloudiness and sunshine (Kaiser *et al.*, 2018; Slattery *et al.*, 2018; Morales & Kaiser, 2020). Both increases and decreases of incoming irradiance will result in imbalances between the intrinsic capacity of the photosynthesis machinery (i.e. as can be achieved under steady-state conditions) and actual amount of light energy that can be processed (Ruban, 2017). Inefficient photosynthesis that results from such imbalances may limit effective carbon assimilation. The slow responses of plant photosynthesis to both increases and decreases in irradiance are believed to play a part in the lower growth under fluctuating irradiance conditions (Athanasiou *et al.*, 2010; Slattery *et al.*, 2018). Part of these losses can be explained by traits that are required for adaptation to changes in light not being maximized in crop species (Lawson *et al*., 2012, Long *et al.*, 2015). Such traits may thus be amenable to improvement as a means to improve crop productivity (Long *et al.*, 2015; Kromdijk *et al*., 2016)

Arabidopsis thaliana (hereafter; Arabidopsis) is a model species with a globally widespread distribution and shows particularly large genomic variation resulting from a long history of long distance migration, isolation and adaptation (Durvusalu *et al.*, 2017). Furthermore, the species is a self-pollinating, which means that it can be used to study genotype x environment interactions due to the possibility to grow replicated genotypes (EL-Soda *et al*., 2014). Arabidopsis genotypes derived from different localities show variation for traits related to the adaptation of the photosynthesis machinery (Athanasiou *et al*., 2010, Mishra *et al*., 2012, van Rooijen *et al.*, 2018; Zhang *et al*., 2020, **Chapter 4**). Such variation can be mapped in genetic approaches to discover genes and their alleles that may be of interest to improve corps (Theeuwen *et al*., 2022). Progress in photosynthesis phenotype imaging technology permits the analysis of large numbers of Arabidopsis genotypes, which makes this species suitable to perform genetic mapping of adaptive traits in photosynthesis (van Bezouw *et al*., 2019).

Genetic mapping studies involving Arabidopsis have so far not targeted the growth and physiological responses to dynamic light conditions. Instead, focus has so far been on chlorophyll fluorescence imaging phenotypes that directly measure photosynthetic activity at the molecular level and its response to changes in light conditions (Jung & Niyogi, 2009; van Rooijen *et al.*, 2015; Wang *et al.*, 2016; Rungrat *et al.*, 2019). Consequently, most of our knowledge related to losses of productivity in Arabidopsis as a result of fluctuating light conditions is mostly derived from experiments that involve only

a single genotype (Vialet-Chabrand *et al.*, 2017; Schneider *et al.*, 2019). Doing so leaves potential roles for genetic variation to adaptation unexplored. Kaiser *et al.* (2020) used a panel of more than thirty Arabidopsis accession to study the impact of fluctuating light on their growth. They conclude that genetic variation in the growth response to fluctuating light conditions is determined by limitations in the rate of photosynthesis in genotypes that are more productive under stable light conditions. As a result, productive genotypes are impacted by higher losses in productivity compared to genotypes growing a naturally smaller rosette, which implies that reduced functioning of photosynthesis as a result of adaptation is a biophysical growth-limiting factor in Arabidopsis. A genetic mapping strategy involving two contrasting parental lines in their physiological response to fluctuating light regimes may determine the role of genetic variation in these traits.

Stable and fluctuating light can be considered as a treatment similar as in other control-treatment studies involving (a)biotic stresses (Munns & Tester, 2008; Nguyen *et al.*, 2013; El-Soda *et al*., 2014; Thoen *et al*., 2017; van Bezouw *et al.*, 2019b). However, for no genetic mapping studies exist in which fluctuating light has been used as such. We aim to investigate the role of genetic variation as a cause for light regime-dependent variation in the productivity and physiology of Arabidopsis. We hypothesize that genetic variation exists for the traits in response to fluctuating light conditions and involve genotype x environment interactions. To do so, we made use of a novel genetic mapping resource in the form of a complete population of chromosome substitution lines (CSLs). A CSL-population is a bi-parental genetic mapping population type that segregates for full chromosomes only, thus without crossovers in the progeny (Wijnen & Keurentjes, 2014). In animals, the use of CSLs – although only in populations with single chromosome substitutions – has previously been associated with improved genetic mapping power and detection of epistasic interactions compared to conventionally segregating populations from the same parental genotypes (Spiezio *et al*., 2012; Bucher & Nadeau, 2015). The CSL-population as described and developed in this work is composed from the widely used Col-0, the core accession for Arabidopsis that shows high productivity, and MIB-22, a wild type accession that produces small, compact rosettes (**Figure 1**). These properties make this population ideal for the goals set out in this study.

2. MATERIALS AND METHODS

2.1 Development of a Col-0 x Mib-22 chromosome substitution panel

Development of chromosome substitution libraries is efficient when meiotic recombination can be suppressed to avoid the introduction of random cross-overs within the population (Wijnker *et al*., 2012). *dmc-4* and *dmc-5* are two independent transformants carrying an RNAi transgene in the Col-0 (CS76113) *Arabidopsis* accession that silences *DISRUPTED MEIOTIC cDNA1* (Wijnker *et al.*, 2014). Lines carrying these constructs have significantly reduced meiotic crossover rates in *Arabidopsis* (Wijnker *et al.*, 2012). Following the protocol described by Wijnker *et al*. (2014) (**See supplementary figure 1**), homozygous transformants carrying the RNAi construct were crossed to the MIB-22 genotype (CS76182) to create an achiasmatic F1 hybrid. Subsequently, this F1 hybrid was crossed to *gfp-tailswap* in order to generate haploid, non-recombinant offspring. Upon selfing, these haploid lines produce homozygous doubled haploid offspring. The last few CSLs with missing combinations were developed using a backcross strategy. A set of genomic markers described in Wijnen *et al*. (2018) was used for genotypic evaluation of the CSL-population. The inter-marker spacing averages ~3 Mbp/marker, which is sufficient to detect crossovers (Kooke *et al*., 2012).

2.2 Growth conditions and experimental design

For assessment of photosynthesis traits, the parental lines were grown in a high-throughput photosynthesis facility and under conditions as described by van Rooijen *et al*. (2015). This involves growth under 100 µmol m^{-2} s^{-1} and a stepwise change in irradiance on day 25 to 500 µmol m^{-2} s^{-1}, to measure the photosynthetic response to a change in light irradiance. For the genetic mapping experiment, all lines of the Col-0 x MIB-22 CSL-population were grown twice in the same climate chamber, in 10 hour photoperiods at 21/19 degrees day and night, and grown under TL-lighting tubes. The plants were watered twice, after two weeks and three weeks of growth. The plants were grown under different light regimes each time as depicted in **Figure 2**. Plants were grown under a stable light regime of 250 µmol m^{-2} s^{-1} of photosynthetically active irradiance and this served as the control experiment. In the treatment experiment, growth conditions were equal to control during the first two days of growth to allow for a proper establishment of the plants. On the 3rd day after sowing, the light regime was changed to an alternating rhythm of 20 minutes of high light irradiance (400 µmol m^{-2} s^{-1}) followed by stepwise changes with a duration of 20 minutes of low light irradiance (100 µmol m^{-2} s^{-1}) for a total of 23 stepwise changes every day (**Figure 2**), serving as the fluctuating light conditions. The light regimes were programmed in such a way that the total accumulated irradiance in each treatment was identical. Thus, only the daily distribution of light irradiance was different between the experiments as recommended by Morales & Kaiser (2020).

Plants were grown in a climate chamber in complete randomized block design, spread over two flood tables. On each of these tables, three equally sized blocks were established and plants were

randomized among these blocks. 24 replicates – 4 per block – were grown for all genotypes, except for genotypes CM01 and CM32, for which 32 replicates each were grown. Furthermore, the Col-0 and MIB-22 wild types and an F1 hybrid made from cross CM1xCM32 was also grown in 32 replicates. Blocks, columns (X) and rows (Y) were included as environmental random factors in our statistical models for purposes of analysis.

2.3 Phenotypic assessment of traits

The assessment of photosynthesis traits in the Col-0 and MIB-22 lines was performed described by Van Rooijen *et al*. (2015). The quantum efficiency of photosystem II (ΦPSII) and projected leaf area were measured for these lines under conditions of low light and a stepwise increase in high light following **Chapter 3**.

In the genetic mapping experiments, plants were imaged for the duration of growth using an automated camera system that allows the simultaneous measuring of all plants. Imaging took place once a day at 17:00 and were processed using custom written Python scripts to estimate the projected leaf area (PLA), in units of total number of pixels per plant. On day 28, two pools per genotype, consisting of two-three plants each, were harvested at around 13:00 to be used for the assessment of chlorophyll and anthocyanin content. A total of five pools were harvested this way from each CM01, CM32, their hybrid and the wild types, Col-0 and MIB-22. Chlorophyll and anthocyanin contents were estimated spectrophotometrically using ethanol extractions (Webb *et al*., 1992), after which they were measured using a spectrophotometer. These measurements yielded the content of chlorophyll *a* (chla), chlorophyll *b* (chlb) and their ratio (chlab). The remainder of the plants were put in bags, dried at 60 degrees for at least three days and weighed to obtain aboveground dry weight (DW) values. Thereafter, all visible leafs were counted per plant, giving the total leaf count per plant (L). Thus, in total six traits were directly measured. From the DW and PLA parameters the specific leaf area (SLA) was calculated. Furthermore, the relative growth rate (RGR) was calculated by log-transformation of the images derived area data over time, then fitting a linear model to obtain the generalized regression coefficient per replicate plant. This model was chosen as it would fit on all growth curves. As fluctuating light is the treatment factor in our study we defined “fluctuating light tolerance” (Tol) as the aboveground productivity (DW) in the fluctuating light experiment divided by the aboveground productivity in stable light conditions, similar to as is done in abiotic stress studies (van Bezouw *et al*., 2019b).

2.4 Statistical analysis

For the analysis of genotypic comparisons and calculations of Best Linear Unbiased Estimates (BLUEs) of genotypic means, we used a linear mixed model incorporating the experimental block, and x and y coordinates (as in Flood *et al*., 2016). Heritability values were calculated by deriving the fraction of genotypic variance by treating all factors as random variables. Since a (near) complete population of chromosome substitution lines offers a balanced experimental design for genetic analysis (Wijnen, 2019), we aimed to apply one-way or multi-way analysis of variance (ANOVA) for each combination of

chromosomes up to the third level of genetic interactions. This totals five additive, ten two-way epistatic and ten three-way epistatic interactions to be considered in a population developed from Arabidopsis. Including all additive and interactive terms in a linear model would imply a significant reduction in degrees of freedom, which would in turn severely limit statistical power to detect chromosomal effects. Thus, we opted for analyzing one-by-one each additive or interaction effect. All analyses where performed using the Anova() function in the cars-package (Fox & Weisberg, 2019). For the inclusion of random effects, we used lmer4 (Bates *et al*., 2015) and lmerTest (Kuznetsova *et al*., 2020) packages. These packages allow the implementation of linear mixed models in R and subsequent post-hoc analysis. For analysis of the full CSL-population, ANOVA was performed using standard analysis of variance on the BLUEs of genotypic means of base R (R Core Team, 2021). Correlation analysis was performed in base R and the correlation plot was made using the corrplot package (Wei & Simko, 2021). Figures were drawn in the ggplot2 package (Wickham, 2016) or using base R functions (R Core Team, 2021).

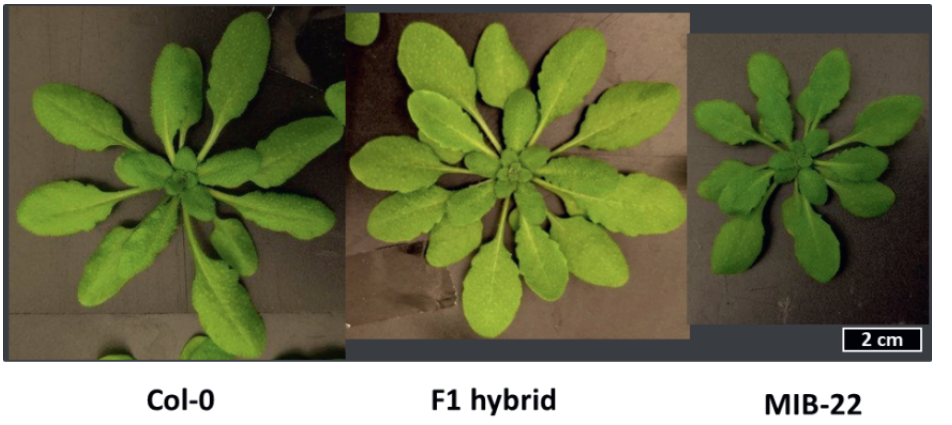


Figure 1. *Images of representative replicates belonging to the accessions Col-0, MIB-22 and the hybrid between the CSLs representing these lines at the nuclear genomic level. All images were taken after 28 days of growth under 250 um m-2 s-1 of growth irradiance, scaled and cropped for comparison purposes. The colouration of leaves may not be representative for the genotypes.*

3. RESULTS

3.1 The development of a chromosome substitution line population derived from Arabidopsis thaliana accessions Col-0 and MIB-22

By applying a reverse breeding scheme (Wijnker *et al*., 2014, **Supplementary figure 1**), we developed a chromosome substitution lines (CSLs) population consisting of the Col-0 and MIB-22. Lines containing any chromosomal combination of having both chromosome 4 and 5 donated from Col-0 – thus all possible lines carrying recalcitrant *dmc* RNAi-constructs conferring semi-sterility – were backcrossed to the appropriate CSLs to develop hybrids that then yielded construct-free lines in their progeny. After ~25 lines were developed, a backcross strategy was applied to retrieve all other lines from the population. In total, 32 lines – thus all possible combinations – were developed in this way. A final genotypic scoring of the population found that one CSL still contained a crossover *CM22:MCMCM*. In order to guarantee the absence of crossovers for purposes of population analysis, we decided not to grow the CM22 line as it contained a larger introgression. All lines were grown and genotyped once more in the greenhouse and seed from one genotyped plant was collected for each line. A phenotypic evaluation of silique elongation and seed set concluding that no lines containing Col-0 chromosomes 4 & 5 carry *dmc* constructs (Wijnker *et al*., 2012, 2014). After conclusion of the experiment, however, we found another minor introgression in CM19*::MCCMC* by using extra markers. The full set of genotypes used in this experiment and their genotypic evaluation can be found in **supplementary figure 2.**

3.2 Col-0 and MIB-22 exhibit similar rates of quantum yield of photosystem II under stable light conditions and in response to high light

Col-0 and MIB-22 genotypes are similar in their quantum yield of photosystem II (ΦPSII, a measure for photosynthetic activity Baker, 2008) under stable light (100 µmol m^{-2} $s^{-1,}$ **Figure 3a**) and in the short term response to high light (500 µmol m^{-2} $s^{-1,}$, **Figure 3b**). Under conditions of stable light, Col-0 grows a rosette nearly twice the size that of MIB-22, which confirms previous observations that Col-0 is an accession with much higher productivity (**Figure 1**). ΦPSII could not be measured under the experimental conditions that the CSL-population was grown in, particularly due to the short-term fluctuations that were imposed on the plants (**Figure 2**). Nonetheless, the data suggests that no significant variation for ΦPSII exists between these two parental lines for stable light conditions and the high light acclimation response (**Figure 3a, b**), thus concluding that those traits do not play a significant role in subsequent experimentats.

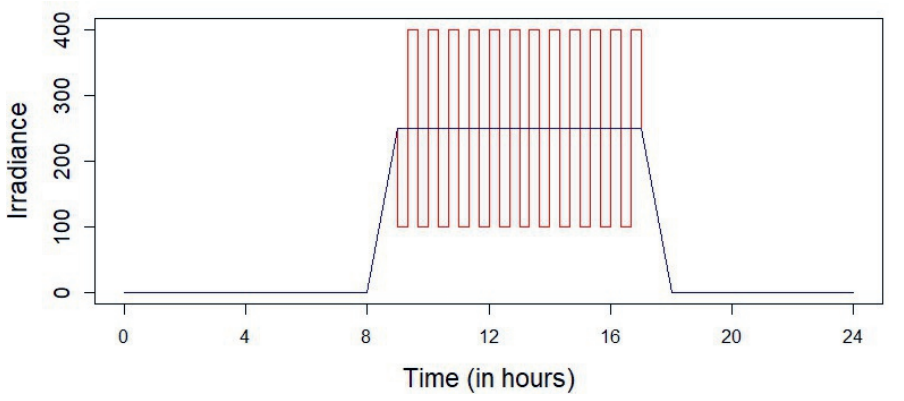


Figure 2. *The light regimes used during the experiments. The red line marks course of changes of the fluctuating light regime used in the experiments, at intervals of 20 minutes for a total of 23 changes in irradiance per day. Note that the increase and decrease at the days start and ending were identical. The total sum of incoming irradiance received by plants in each treatment is identical. Irradiance is measured as photosynthetically active radiation in µmol m^{-2} s^{-1}.*

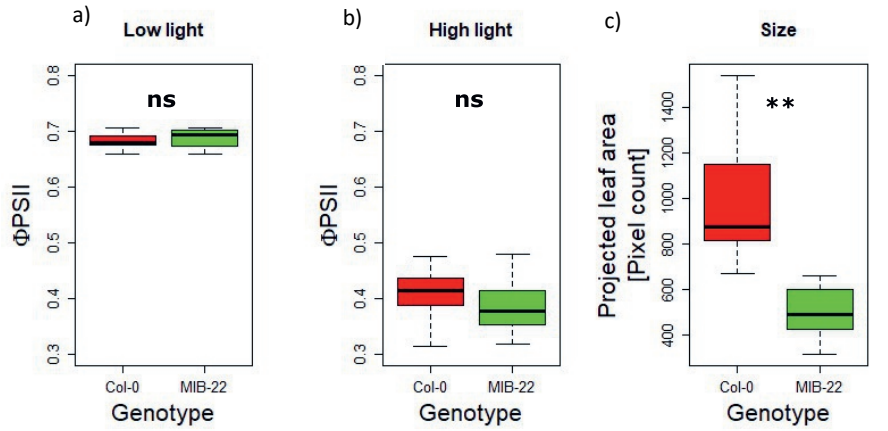


Figure 3. *Assessment of quantum yield of photosystem II (ΦPSII) and projected leaf area (PLA) in the wild type Col-0 and MIB-22 genotypes. a) ΦPSII of plants grown under low light conditions (100 µmol m^{-2} s^{-1}). Measurement was taken on the 24st day of growth, 1 hour after the onset of the photoperiod (See van Rooijen et al., 2015). b) ΦPSII of the same plants, but now taken on the 25th day of growth, 1 hour after the onset of the photoperiod with a stepwise increase of irradiance to 500 µmol m^{-2} s^{-1}. c) Projected leaf area of the genotypes on the 24st day of growth. ns = not significant, ** = significant at $p < 0.01$. n = 14 for Col-0 and 16 for MIB-22.*

3.3 Fluctuating light treatment negatively affects plant plant productivity and changes rosette architecture

The Col-0 x MIB-22 CSL-population, wild type parents and an F1 hybrid between CM1 and CM32 (**Supplementary figure 1**) were grown in two subsequent experiments. In both experiments, plants were grown under stable light conditions for three days after which either fluctuating or stable light conditions were continued for the remainder of the experiment (**Figure 2**). We assessed a total of eleven traits, including the assessment of projected leaf area (PLA) of three specific days of measurements of project leaf area, namely 6, 17 and 28 days after sowing to represent early, mid and late growth. Best linear unbiased estimates (BLUEs) were calculated for each genotype per condition prior analysis. Most of the traits evaluated in this experiment are negatively impacted by the fluctuating light treatment (**Table 1**). Three of these are unaffected by treatment, namely leaf count (LC), relative growth rate (RGR) and projected leaf area (PLA) measured 6 days after sowing. Rosette dry weight (DW) under fluctuating light conditions was reduced by 26.2% compared to stable light conditions, while rosette leaf area (pixel count at 27 DAS) only decreased by 14.7%. In turn, this discrepancy negatively affects the population-wide specific leaf area (SLA). Furthermore, chlorophyll *a* (chla) decreased significantly, while chlorophyll *b* content (chlb) increased significantly. This counteracting interplay together results in a triplication of the chlorophyll a/b ratio (chlab) under fluctuating light conditions. Leaf anthocyanin content was not detectable in the rosettes, and will thus not be discussed further.

Correlations between the best linear unbiased estimates (BLUES) for the measured traits between treatments revealed that traits indicating productivity (PLA, DW) under stable light condition are negatively correlated with fluctuating light tolerance (Tol) (**Figure 4**). Conversely, only the pixel to weight ratio of plants grown under fluctuating light conditions was correlated with tolerance of productivity. Chlab correlates negatively with chlb, and there are some minor correlations between chlorophyll content and leaf count (**Figure 4**). However, chlorophyll is generally a poor predictor for any of the other traits.

3.3 Analysis of (wild type) parental and F1 lines reveals a role for hybrid vigour for fluctuating light tolerance

Within the chromosome substitution lines (CSL-)population, the Col-0 and MIB-22 parental lines are essentially represented by two specific substitution lines, namely CM1 and CM32 (**Supplementary figure 1**). However, essential differences are that CM1 and CM32 have gone through the process of meiotic suppression and dihaploidzation (Wijnker *et al*., 2014). Furthermore, both these lines contain the cytoplasmic genome of the Col-0 genotype, thus CM32 differs from MIB-22 in that respect. These different genetic origins may impact photosynthetic performance as the chloroplastic genome in the cytoplasm is a source genetic variation (Flood *et al*., 2020). The Col-0 and CM1 genotypes show a roughly 30% reduction in dry weight when grown in fluctuating light conditions compared to stable light conditions (**Figure 5**). Conversely, dry weight in MIB-22 grown under fluctuating light conditions is reduced by <15% compared to those grown under stable light conditions, while CM32 is unaffected, thus implying a cytoplasmic contribution to the response to fluctuating light of this trait (**Figure 5a**). Leaf count is similarly

differentially affected by the fluctuating light treatment between CM32 and MIB-22 (**Figure 5b**). The $F1_{CM1xCM32}$ hybrid shows high values for all traits assessed (**Figure 5a-e**), expressing profound hybrid vigour. $F1_{CM1xCM32}$ approximates dry weight as in the Col-0/CM1 genotypes, but under fluctuating light conditions only a marginal decrease in biomass accumulation is observed compared to stable light conditions (**Figure 5a**), while having high leaf number (**Figure 5b**). Plant area size in early (6 DAS, **Figure 5c**), mid (17 DAS, **Figure 5d**) and later (27 DAS, **Figure 5e**) stages shows smaller reductions between the treatments meaning that specific leaf area changed, but this trait was unaffected in the MIB-22 genotypes (MIB-22 wild type and CM32). For chlorophyll content traits, no differences are found between these five genotypes (**Table 2**), although they are affected by treatment in a similar way as is observed in the whole populations (**Table 1**).

Table 1. *Descriptive statistics of trait values for the 31 genotypes and $F1_{CM1xCM32}$ of the Col-0 x MIB-22 chromosome substitution library. All values have been calculated from best linear unbiased estimates within each of the treatments. Trait values are calculated for stable (S) and fluctuating (F) treatments. Min = genotype with the lowest, Mean = the average and Max = highest trait value of the phenotype. Stdev is the standard deviation calculated among the best linear unbiased estimates. CoV = coefficient of variation (Stdev divided by the Mean), FvS t-test gives the p-value for the outcome of a t-test comparing the estimated means of the fluctuating and stable treatments. F/S ratio is the relative value of the trait measured in plants grown under fluctuating light conditions compared to stable light conditions. H^2 is the broad-sense heritability of traits. ns = not significant, * = p <0.05, ** = p < 0.01, *** =p < 0.001.*

Trait	Unit	Treatment	Min	Mean	Max	Stdev	CoV	FvS t-test	F/S -ratio	H^2
Dry weight	g	F	0.029	0.043	0.065	0.007	0.169	<0.001***	0.738	0.221
		S	0.036	0.058	0.075	0.010	0.171			0.254
Dry weight tolerance	-	F / S	0.554	0.748	1.037	0.122	0.163	-	-	-
Relative growth rate	pixels/day	F	1210	1670	2026	199	0.119	0.075 ns	0.971	0.040
		S	1296	1958	2588	278	0.142			0.090
Total leaf count	n	F	19.798	15.659	12.373	1.437	0.092	0.406ns	0.982	0.290
		S	18.649	15.941	13.703	1.261	0.079			0.336
Specific leaf area	-	F	0.957	1.110	1.388	0.095	0.086	<0.001***	0.865	-
		S	1.109	1.283	1.453	0.084	0.066			-
Pixel count (6 DAS)	n	F	83.78	106.9	138.84	15.4	0.144	0.314 ns	0.934	0.084
		S	87.23	114.4	156.56	17.6	0.154			0.128
Pixel count (17 DAS)	n	F	2321	3217.5	4304	433.3	0.135	<0.001***	0.795	0.114
		S	2820	4048.8	5557	593.7	0.147			0.175
Pixel count (27 DAS)	n	F	27913	38518.2	46737	4592.0	0.119	<0.001***	0.853	0.167
		S	29908	45150.9	59688	6396.0	0.142			0.298
chlab	mg/ml	F	3.32	8.214	17.29	4.030	0.491	<0.001***	3.090	0.103
		S	1.29	2.659	4.18	0.814	0.306			0.199
chla	mg/ml	F	0.24	0.320	0.43	0.037	0.116	<0.001***	1.234	0.068
		S	0.19	0.259	0.30	0.029	0.112			0.264
chlb	mg/ml	F	0.02	0.060	0.10	0.020	0.331	<0.001***	0.502	0.226
		S	0.07	0.119	0.17	0.028	0.235			0.156

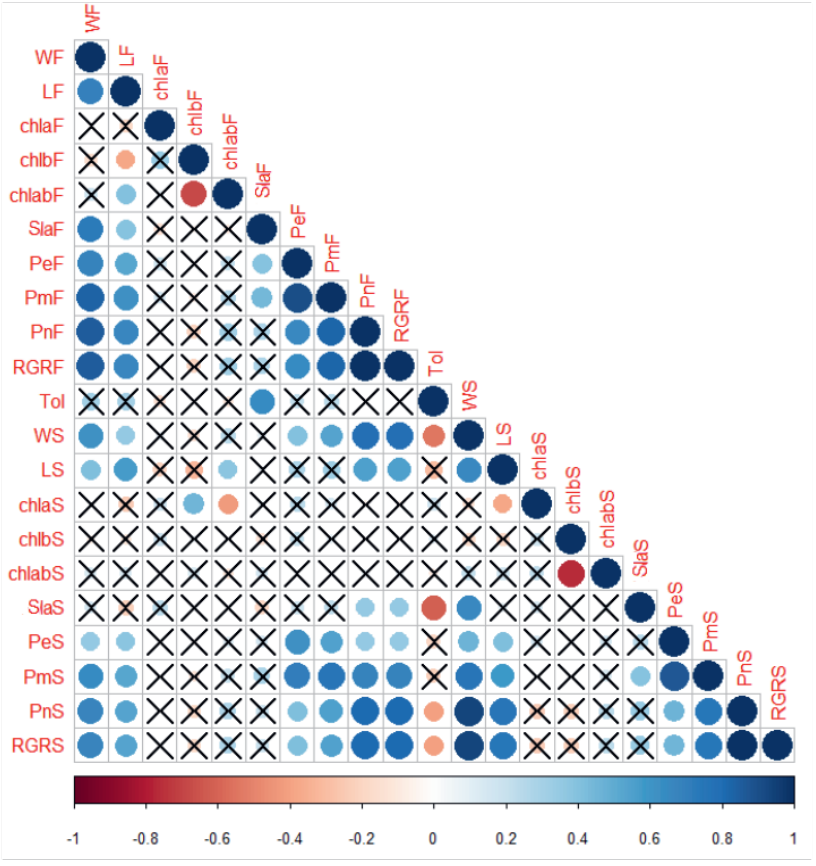


Figure 4. *Correlation of various traits measured in the stable light and fluctuating light experiments. W = Dry weight, L = Total leaf count, chla = chlorophyll a content, chlb = chlorophyll b content, chlorophyll ab is the chlorophyll a / b ratio, specific leaf area (Sla) = projected leaf area / dry weight, P = projected leaf area of which; Pe = early phase (6 DAS) rosette size, Pm = mid-phase (18 DAS), Pn = end phase (28), RGR = Relative growth rate, Tol = dry weight (fluctuating light) / dry weight (stable light) conditions. Capital S = Stable light conditions, F = fluctuating light conditions. X's denote correlative relationships to not be significant at the P <0.05 level (Pearson's method). Size and colour of circles indicate the strength and direction of the correlation relationship. Means of 30 substitution lines and the F1*$_{\text{CM1xCM32}}$ *were included in the data to calculate the correlation coefficients and their significance level. Chlorophyll data missed for* CM16 *and thus all phenotypes for this line are excluded from the analysis.*

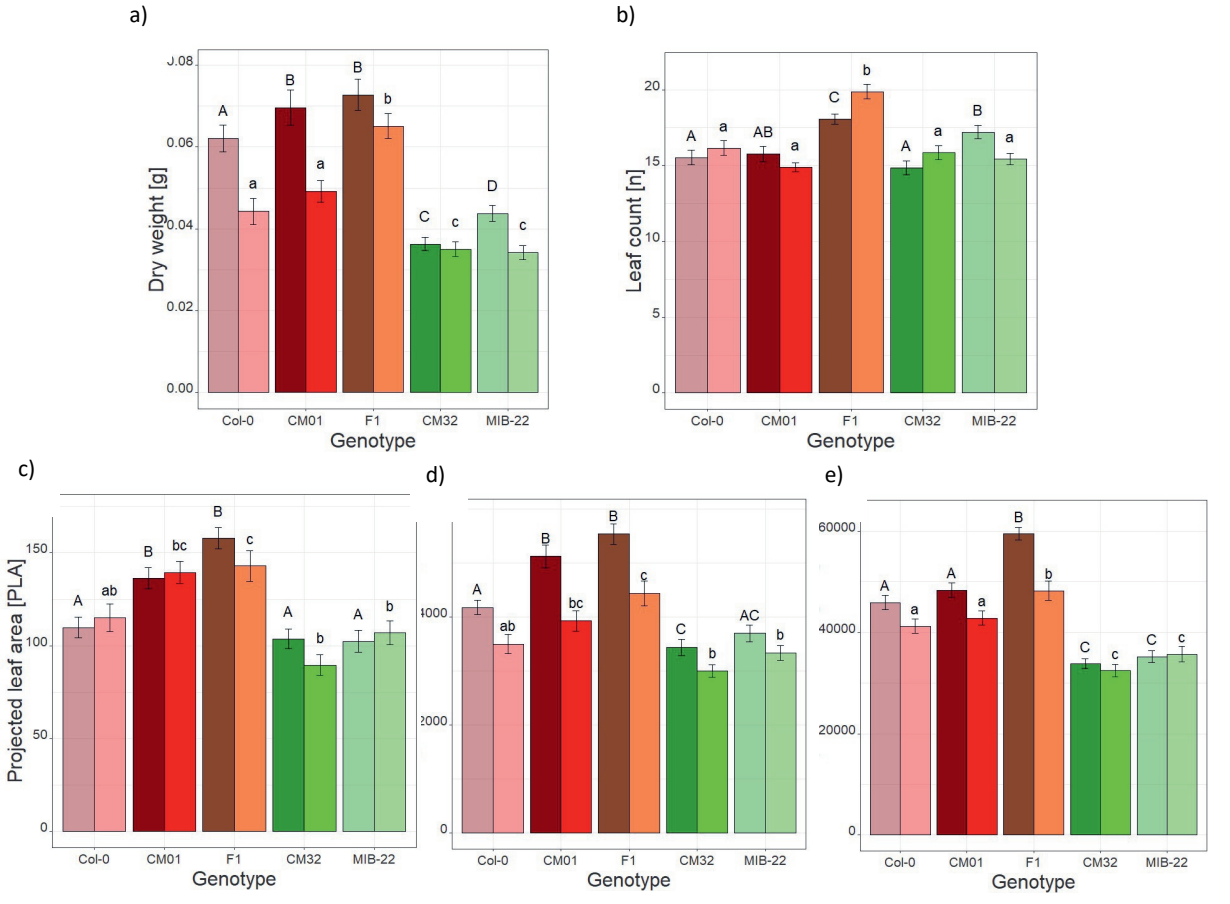


Figure 5. *Trait values for parental genotypes and their hybrid under stable and fluctuating light conditions. a) Dry weight, b) leaf count, c) projected leaf area at 6 days, d) 17, e) 27. The left, darker, bars indicate the trait value of plants grown for 28 days under stable light conditions, the right, brighter, bars indicate the trait value of plants grown under fluctuating light conditions. Capital letters indicate groups of genotypes sharing the same phenotypic mean following a post-hoc Tukey test for plants grown under stable light conditions, lower case letters are for those of plants grown under fluctuating light conditions. Bars show the means and standard error of at least 17 replicates per treatment per genotype.*

Table 2. *Chlorophyll content and chlorophyll a/b ratio in parental lines, reverse breeding parents and their hybrids. Means and standard errors are given for each genotype per treatment. Each replicate (N) is composed of up to three rosettes snap-frozen in liquid nitrogen. Letters indicate groups with equal means after Tukey post-hoc tests. FvS t.test gives the p-value for testing the genotypic means of the genotypes of stable light against the samples taken from the fluctuating light treatment. ns = not significant, * = p <0.05, ** = p < 0.01, *** =p < 0.001.*

Treatment	Genotype	N	chla	chlb	chla/b
Stable	Col-0	4	0.268+-0.022[a]	0.088+-0.023[a]	4.001+-1.379[a]
	CM01	5	0.295+-0.029[a]	0.163+-0.017[a]	1.927+-0.333[a]
	F1	5	0.255+-0.021[a]	0.095+-0.019[a]	3.596+-1.23[a]
	CM32	5	0.235+-0.019[a]	0.089+-0.013[a]	3.068+-0.737[a]
	MIB-22	5	0.231+-0.024[a]	0.129+-0.023[a]	2.229+-0.626[a]
Fluctuating	Col-0	5	0.327+-0.021[a]	0.048+-0.007[a]	7.629+-1.399[a]
	CM01	5	0.310+-0.037[a]	0.078+-0.013[a]	5.358+-2.177[a]
	F1	5	0.324+-0.024[a]	0.054+-0.014[a]	9.952+-4.494[a]
	CM32	4	0.286+-0.033[a]	0.056+-0.009[a]	5.545+-1.173[a]
	MIB-22	5	0.294+-0.008[a]	0.077+-0.016[a]	5.202+-1.731[a]
FvS t.test	-	-	0.0677[ns]	0.0136*	0.0054**

3.4 Plant productivity is characterized by large variation in the genotype x environment

From the previous, it should be clear that high genetic variation exists among the parental lines of the Col-0 x MIB-22 CSL-population and this is reflected in heritability estimates (**Table 1**). A CSL-population has a limited number of possible genotypes (Wijnen, 2019), but each can be grown in high replicate number thus permitting the analysis of genotypes against each other. By testing the between-environment effect of each individual line, we observe large variation for productivity tolerance to fluctuating light condition in the population (**Figure 6a**). The most intolerant lines lose almost half of their dry weight under fluctuating light conditions compared to growth under stable light conditions, while a group of six lines are unaffected by light regime. Further examination of the correlation between the genotypic BLUEs of the individual genotypes in stable and fluctuating light conditions reveals that genotypes with high dry weight under stable light conditions predict dry weight under fluctuating light conditions (**Figure 6b**)

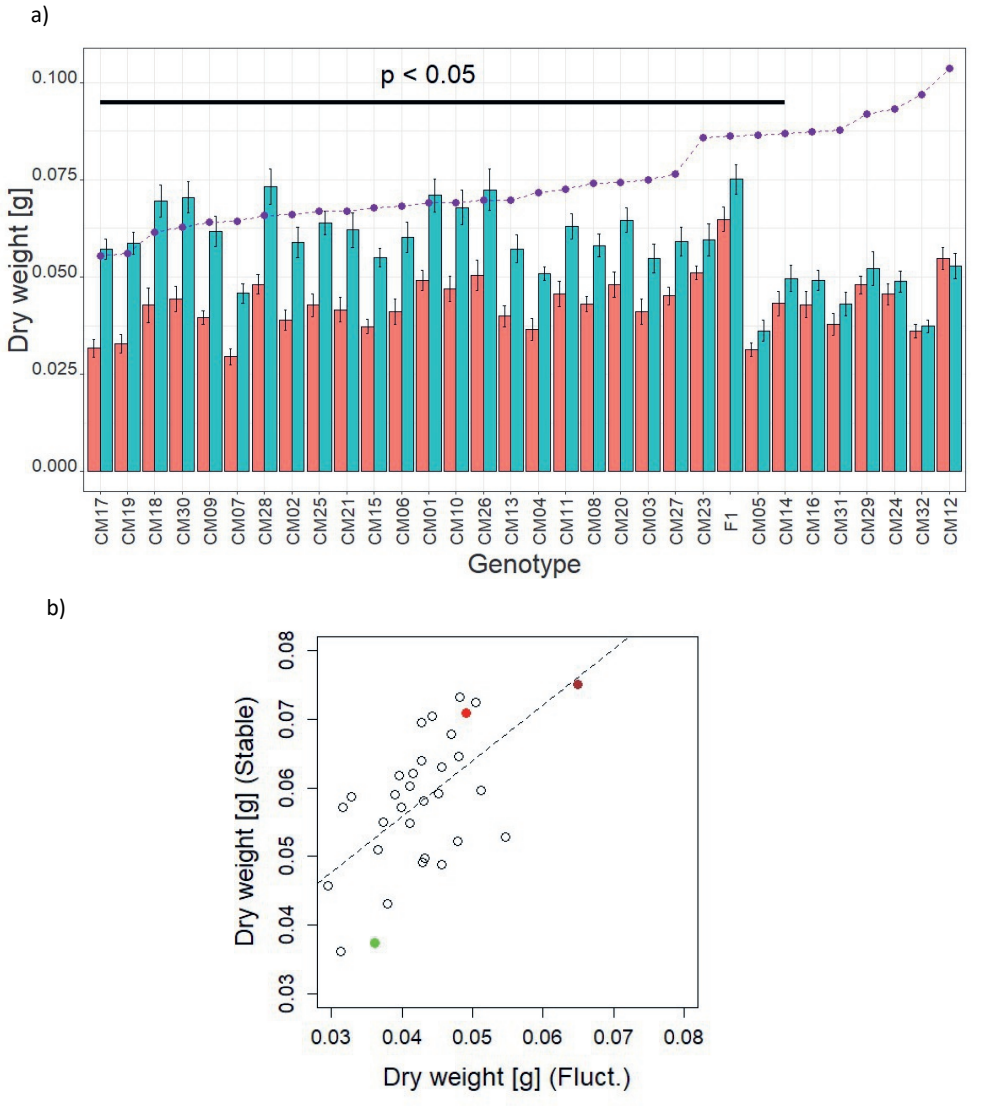


Figure 6. *Ranking of all Col-0 x MIB-22 substitution line genotypes and the* $F1_{CM1xCM32}$ *for dry weight.* ***a****) Bars depict the Best Linear Unbiased estimates of the genotypic means of rosette dry weight (DW) for fluctuating conditions (red) and stable light conditions (blue). The purple line depicts the tolerance of productivity, calculated as* $DW_{fluctuating}/DW_{stable}$ *and are ranked following this value. The horizontal black bar indicates a significant effect (p < 0.05) of treatment on dry weight for all genotypes graph, following analysis of variance. +-1 error bars are given.* ***b****) correlation between rosette dry weight under fluctuating and stable light conditions. All means have been calculated from more than 15 biological replicates per genotype per treatment.*

To unravel the genetic contribution to physiological traits under fluctuating light we analyzed the 31 chromosome substitution lines using classic analysis of variance, up to the third level of chromosomal interactions (**Figure 7a**). A total of 32 significant additive and interacting quantitative trait loci (QTLs) are found; four additive QTLs under fluctuating light and nine under stable light conditions. Tolerance of productivity to fluctuating light yielded another, but weak, QTL, thus we found a total of 14 additive effects. Of these, only a QTL for 27 DAS for projected leaf area on chromosome 3 and total leaf count on chromosome 5 are present in both light environments. The remainder of these associations are environment specific, suggesting a strong role for growth light regime to play a role. A total of six two-way and twelve three-way epistatic chromosomal interaction effects were found for various traits during the experiments, with the majority of these interactions being specific to a single environment. However, most of the genotypic effects we found are weak associations ($0.01 < p < 0.05$), which will probably be neglected after performing a post-hoc test correction.

QTLs for dry weight on chromosome 3 (**Figure 7b**) and leaf count on chromosome 5 (**Figure 7c**) were further examined due to the strength of the genotype-phenotype associations. A Col-0 chromosome 3 exhibits a 19.4% increase between for dry weight under stable light conditions. The phenotypic effect of this QTL is, however, suppressed under fluctuating light conditions (**Figure 7b**), with two-way ANOVA confirming the presence of a light regime-dependent effect for growth (interaction term; GxE; $F_{1,58} = 4.25$, $p = 0.043$). In line with these observations, chromosome 3 is also the main driver of genetic variation in the tolerance of productivity to fluctuating light conditions (**Figure 7d**). By contrast, and in accordance with our earlier analyses of the data (**table 1**), the leaf QTL on chromosome 5 is unaffected by environment of Col-0 and MIB-22 (**Figure 7c**). Two-way ANOVA) confirmed the absence of a genotype x environment effect for this QTL (interaction term; $F_{1,58} = 0.549$, $p = 0.462$).

3.5 Time series analysis of projected leaf area reveals temporal effects of fluctuating light on plant growth

High-throughput imaging of projected leaf area (PLA) of the CSL-population during the duration of the experiment allows the description of long-term trends in growth. Repeated tracking of PLA may lay bare trends that would otherwise go unnoticed in destructive harvests, which represent a single moment in time. Both the average best linear unbiased estimates (BLUE) of the genotypic means and heritability values per treatment for the PLA in the CSL-population increased during the experiment – particularly in the last week (**Figure 7a**). However, both the PLA and heritability are significantly higher in the population grown under stable light conditions, compared to those grown under fluctuating light conditions. Broad sense heritability values reach over 0.30 in the later stages of growth under stable light conditions. Under fluctuating light conditions the maximum heritability barely exceeds that of the population after 5 days of growth under stable light conditions. To identify the day on which the plants grown under stable light conditions show higher PLA values, we performed a student's t-test for each day of the experiment (**Figure 7b**). This revealed that from day nine after sowing, there is a significant difference of average PLA of the population between the stable and fluctuating light conditions. This difference gradually increases until the

end of the experiment, where the average PLA of plants grown under fluctuating light conditions is only 75-80% to that of the average PLA under stable light conditions (**Figure 7b**).

Based on visual inspection of the mean differences in PLA between the light treatments (**Figure 7a**), we hypothesized that the fluctuating light conditions introduces a growth delay for the whole population of approximately 1 day. To validate this assumption we re-calculated the student's t-test for each day analysis and the fraction of growth between the PLA from the fluctuating light condition, but this time compared the values for projected leaf area of fluctuating light one day ahead (PLA+1) of the values for stable light conditions (PLA) (**Figure 7b**). In doing so, we observe an opposite trend, with relative differences in PLA between the light treatments decreasing over time. This confirms that in our experiment plant growth under fluctuating light conditions lags one day behind the growth projection under stable light conditions.

Broad sense heritability values dynamically change between and within experiments (**Figure 7a**), which indicates that the contribution of genetic factors may vary of time. To analyze this, we applied analysis of variance on PLA on all time points as we did do for the other traits (**Table 1**). Under stable light conditions, we found that QTL on chromosome 3 for PLA (**Figure 8a**) to be detectable after 20 days of growth (**Figure 7c**). However, the effect of this QTL is lost in fluctuating light as chromosome 3 only reveals a marginal effect on the last day of that experiment. Two additional dynamic QTLs are present in the population, namely on chromosome 2 under stable light conditions, which is present from day 16, and chromosome 5 under fluctuating light conditions only appearing in the final days of phenotyping (**Figure 7c**).

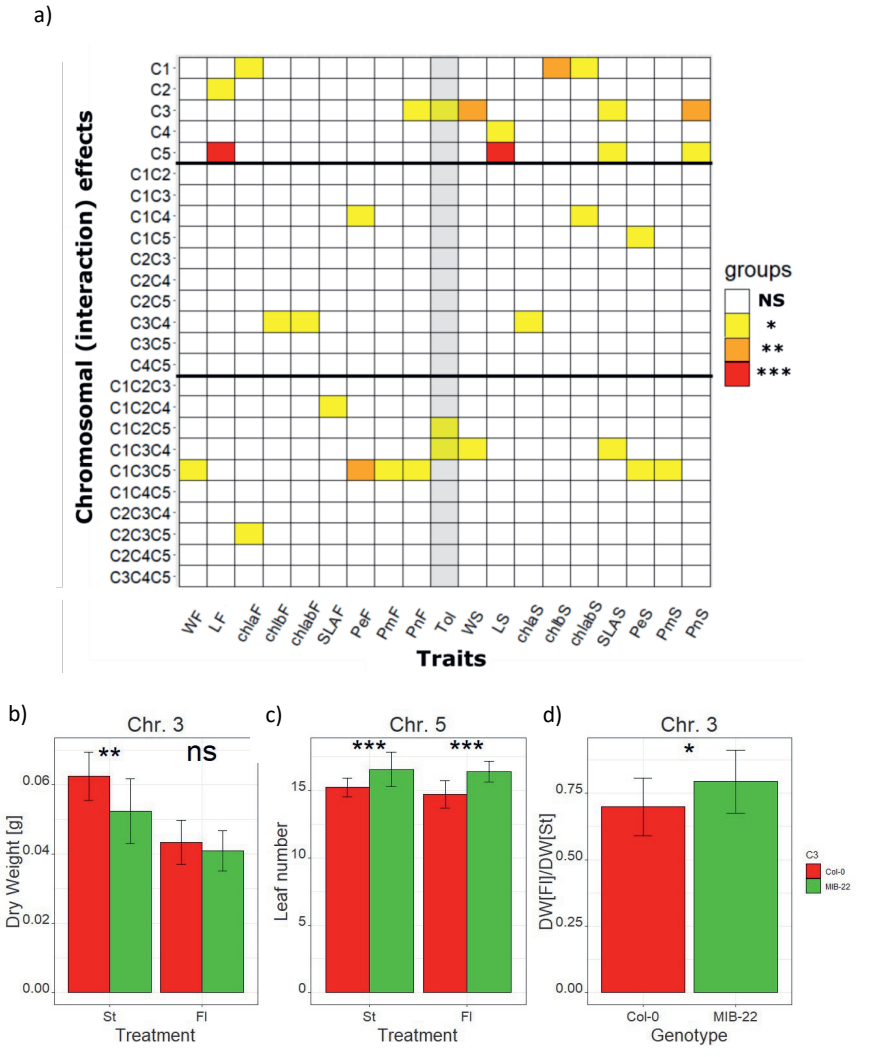


Figure 7. *Analysis of the Col-0 x MIB-22 chromosome substitution line population grown in fluctuating and stable light conditions.* ***a)*** *Analysis of variance matrix for traits. On the Y-axis each combination of chromosomes tested is presented and the testing outcome of the highest order of interaction with each analysis is given in colour groups, where NS (white) = p > 0.05, * (yellow) = p < 0.05, ** (orange) p < 0.01, *** (red) p < 0.001. W = Dry weight, L = number of leaves, chla = chlorophyll a, chlb = chlorophyll b, chlb = ratio chla/chlb, SLA = Specific leaf area, Pe = projected leaf area early (6 days after sowing, Pm = PLA medium (17 days after sowing), Pn = PLA end (27 days after sowing). Dry weight tolerance (Tol), here the dry weight under fluctuating light conditions divided by the dry weight under stable light conditions (WF/WS), is highlighted in grey. F denotes traits measured under fluctuating light conditions, S denotes traits measured under stable light conditions.* ***b)*** *Genotype x environment analysis for the dry weight (W) trait for chromosome 3,* ***c)*** *as in* ***b****, but for leaf count (L),* ***d)*** *fluctuating light tolerance (Tol). Treatments are St = stable, Fl = fluctuating light. Bars depict the means and standard error of at least 15 genotypic means per chromosome. Asterisks as in a, following from student's test per treatment.*

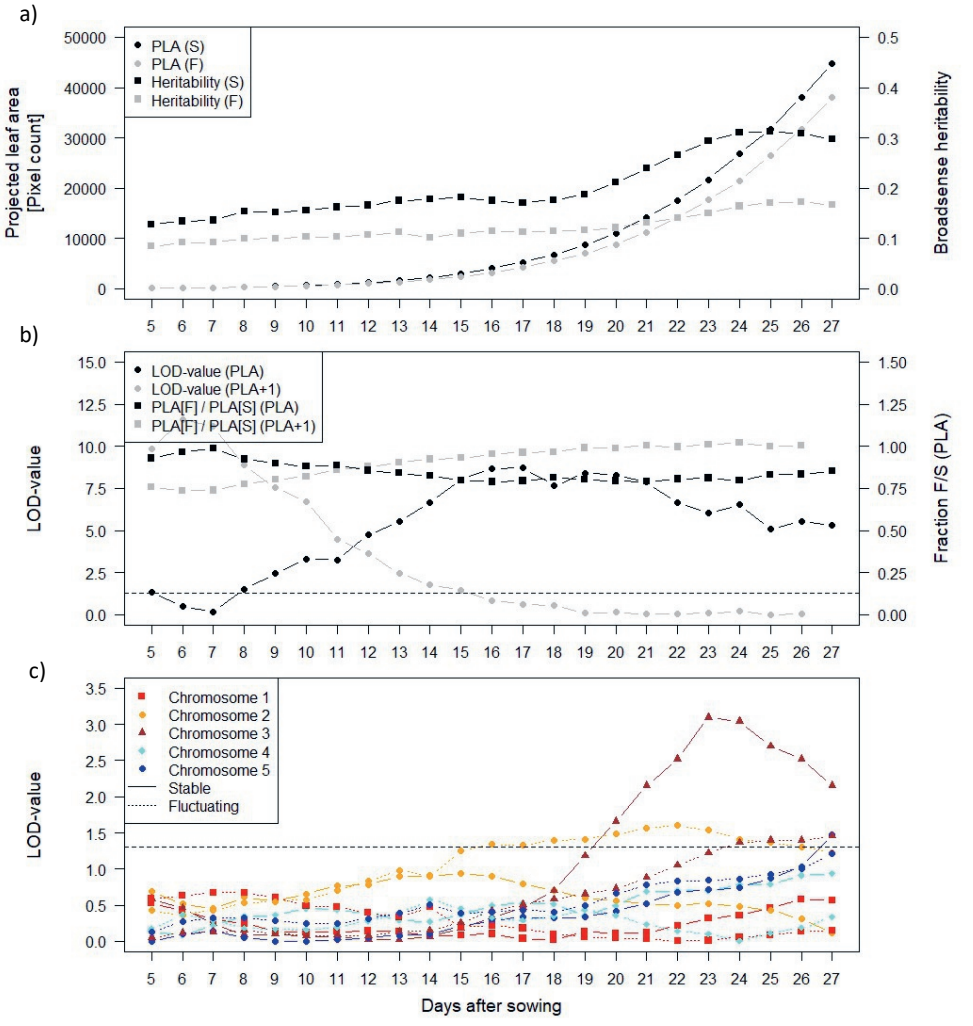


Figure 8. *Growth curve analysis of chromosome substitution lines grown under stable (S) and fluctuating light (F) conditions. Measurements per day are taken from images taken at 17:00. Plants were imaged from the 5th day after sowing in either condition.* ***a****) Average projected leaf area is taken as the mean number of pixels of the spatially corrected genotypic best unbiased linear (BLUES) estimate of each substitution line. Broadsense heritability is calculated as the percentage of variation explained by genotypic terms from 31 different chromosome substitution lines divided by the total variation.* ***b****) LOD-value represents the –log(p) transformed test statistics of testing the BLUES of the stable and fluctuating light conditions (thus n=31 per category), while Fraction F/S (PLA) represents the means of the BLUEs under fluctuating conditions divided by stable conditions. The +1 indicator indicates that these numbers have been generated from using the PLA values of the fluctuating light condition one day ahead that of the stable conditions.* ***c****) LOD-values of additive chromosomal effects, plotted per chromosome and treatment over time for both the stable and fluctuating light condition. In b, c the horizontal dashed line represent (LOD = 1.301) = (p = 0.05) thresholds.*

4. DISCUSSION

In this work, we describe the development of a population of chromosome substitution lines composed from two Arabidopsis accessions differ in productivity and rosette size (**Figure 1, 3c**), but are similar in relevant photosynthetic properties (**Figure 3a**, **b**). We subsequently used this population to study the genetic architecture of the response of various plant traits to fluctuating light, with a focus on plant productivity. Most, but not all, measured traits were subject to genotype x environment interactions (**Figure 7a-d**). To our knowledge, this is the first report of a genetic mapping study that focuses on physiological traits and productivity treating fluctuating light growing conditions as an abiotic stress treatment. In addition, this is one of the few reports on the use of a comprehensive chromosome substitution population. Here, we discuss the implications of our findings.

4.1 The physiological response under fluctuating light can be studied as an abiotic stress in genetic mapping studies in a low-tech setting

The impact of fluctuating light conditions can be regarded as a form of abiotic stress, categorically similar to, e.g, salinity (Munns & Tester, 2008) and drought (Tardieu *et al*., 2018). Fluctuation light challenges photosynthetic acclimation and results in compromised growth of plants under such conditions (Vialet-Chabrand *et al*., 2017; Schneider *et al*., 2019; Ding *et al*., 2019). As light fluctuations can be expected to always be encountered in the growth environment in field crops (Tardieu *et al*., 2017; Kaiser *et al*., 2018), dynamic light may be one of the most important yet overlooked abiotic stresses that challenge crop growth. Fluctuating light treatments applied in studies using Arabidopsis vary greatly in the frequency, nature and values of changes, varying from few daily step-wise changes (Athanasiou *et al*., 2010; van Rooijen *et al*., 2015) to high frequency schemes (Kaiser *et al*., 2020; Flood *et al*., 2020) and simulations of measured irradiance data from field conditions (Vialet-Chabrand *et al*., 2017). The frequency and variation in light intensity of the fluctuations we imposed in our study are lower than those applied in other studies investigating the effects of fluctuating irradiance on the Col-0 genotype of Arabidopsis (Vialet-Chabrand *et al.*, 2017; Schneider *et al.*, 2019). Reductions in dry weight that arose from growth under fluctuating irradiance compared to stable light conditions reached a population average of 26.2% in this study (**Table 1**), which is similar to the cited studies. However, line specific losses vary between 0 and 45%, which reveals revealing that a large degree of segregation exists for this trait in the CSL-population used in this study (**Figure 5**). The results in this study indicate that plant physiological and productivity responses to fluctuating light are amenable to genetic mapping. This study also shows that the use of extensive chlorophyll fluorescence imaging setups may not necessarily be a prerequisite to study relevant traits in fluctuating light conditions (van Bezouw *et al*., 2019). We showed that it is possible to study the plant's response to fluctuating light in a low-tech experimental setups leading to meaningful insights in plant growth, an approach which may be easy to extend to crop species as well.

4.2 Fluctuating light reduces genetic variation and the effects of a quantitative trait locus (QTL) for productivity

The majority of traits studied in this work showed low to moderate values of heritability (**Table 1**, **Figure 7a**), but under fluctuating light conditions heritability values were generally similar or lower for most traits than in stable light conditions. These observations are similar as those previously reported by Kaiser *et al*. (2020) – see heritability values (S**upplementary table 1) –** despite that they used a fluctuating light regime with shorter stepwise intervals and a larger difference between minimum and maximum irradiance compared to this study (**Figure 2**). Regardless of these differences, these observations imply that fluctuating light may reduce the trait variability in genotypes of Arabidopsis. Nonetheless, many (interacting) QTL were found to be specific for the growth conditions (**Figure 7a**), thus highlighting a role for genotype x environment interactions.

We discovered a strong QTL on chromosome 3 that positively affects dry weight (**Figure 6a, b, 7c**), but does not express under fluctuating light growth conditions thus creating a genotype x environment interaction. This QTL, however potent, is in this study only mapped to the chromosome level. Continuous and efficient fine mapping efforts to discover the causal gene – similar to as was previously demonstrated in **Chapter 3** and **4 –** may result in a novel trait that stimulates growth, but which is compromised by rate limitations of photosynthesis. This trait may be subsequently explored to identify bottlenecks that prevent plants from growing to their maximum extend in field conditions. Previously, activity of *GLUCOSE 6-PHOSPHATE/PHOSPHATE TRANSLOCATOR 2 (GPT2)* was found to negatively affect plant fitness under fluctuating light conditions compared to stable light conditions (Athanasiou *et al*. 2010). *GPT2,* located on chromosome 1, is thought to affect carbohydrate metabolism and knock-out mutants are associated with reduced starch synthesis in leaves (Dyson *et al*., 2015). The QTL on chromosome 3 may very well be driven by a gene involved in a similar process and may be amenable to transgenic exploitation to overcome the associated bottlenecks to improve plant growth under fluctuating light conditions.

4.3 Hybrid vigour contributes significantly to increased tolerance of plant productivity to fluctuating light

In this study, we used a hybrid line ($F1_{CM1xCM32}$) between the CSLs CM1 and CM32, that represent the Col-0 and MIB-22 wild type parental accessions respectively, and grew the progeny alongside the CSL and parental lines. $F1_{CM1xCM32}$ maintains high productivity associated with Col-0 genotypes under stable light conditions, while losing only 10.5% of growth under fluctuating light, a number comparable in MIB-22 (**Figure 2a**). These observations conflict with observations made by Kaiser *et al.* (2020), who conclude that Arabidopsis accessions with the greatest productivity under stable light conditions suffer only proportionately greater loss of productivity under fluctuating irradiance than did those genotypes with lower productivity under constant irradiance. Kaiser *et al*. (2020) imply that biophysical – rather than genetic – factors play a dominant role in explaining differences in yield losses under fluctuating light environments. The $F1_{CM1xCM32}$ demonstrate that high productivity can persist under fluctuating growth

light, suggesting an improved light use efficiency or photosynthetic activity of this hybrid under these adverse conditions. Although the exact genetic and molecular mechanisms remain to be elucidated, hybrid vigour is thought to result from diversifying alleles in the genome in which deleterious alleles are bi-directionally masked by beneficial alleles coming from either of the parents (Blum, 2013; Fujimoto *et al.*, 2018; Botet Vaca & Keurentjes, 2020). In the Col-0 x C24 hybrid system upregulation of genes encoding for proteins locating to the chloroplast are associated with 20% higher growth compared to the parental values (Fujimoto *et al*., 2012). In our study, $F1_{CM1xCM32}$ is nearly 40% bigger than Col-0 and 100% bigger than MIB-22 in dry weight and projected leaf area (**Fig 4a**, **d**). As both the quantum yield of photosystem II (ΦPSII), in stable and in response high light (**Figure 2a, b**), and chlorophyll content (**Table 2**) are equal between the parents, then other parameters associated with photosynthesis must explain hybrid vigour in $F1_{CM1xCM32}$. A particular trait that predicts hybrid vigour is the development of additional leaves, especially in the early stages of plant development (Fujimoto *et al.*, 2012; Saeki *et al.*, 2016; Ding *et al.*, 2021). Compared to the parental lines (CM1, CM32), $F1_{CM1xCM32}$ also grows a higher number of leaves (**Figure 3b**), regardless of the conditions tested. The Col-0 (CM1) x MIB-22 (CM32) hybrid may be an interesting system to study the role of hybrid vigour and growth under fluctuating light conditions in future studies.

4.4 Development of a novel Col-0 x MIB-22 population of chromosome substitution lines and its use in quantitative trait analysis

The first full and complete panel of chromosome substitution lines in Arabidopsis is the Col-0 x Ler-0 initiated by Wijnker *et al*. (2012) and completed and used for trait analysis by Wijnen (2019) and Wijnen *et al*. (2018). Developing genetic mapping populations including Ler-0 as a parent always bring segregation patterns related to the well studied and highly pleitropic *ERECTA*-locus (Van Zanten *et al.*, 2009). This locus significantly impacts a large number of physiological and developmental traits (Alonso-Blanco *et al.*, 1998; Ungerer *et al.*, 2002; van Zanten *et al.*, 2009, 2010), including rosette dry weight and projected leaf area. Genetic mapping of progeny from crosses involving Ler-0 thus usually involves the detection of this locus (Keurentjes *et al.*, 2007; Wijnen 2019). While this feature provides a good opportunity to validate genetic mapping methodology because of its clear phenotype and study the pleiotropic effects of the *ERECTA* gene in traits, the presence of this locus may epistatically mask the effects of other contributing loci due to it prominently contributing to genetic variation (Carlborg & Hailey, 2004; MacKay, 2014). The CSL-population developed from Col-0 and MIB-22 does not include this locus, making it valuable in quantitative trait analysis.

Complete panels of chromosome substitution panels are unique in that they offer an opportunity of analyzing a completely balanced experimental setup using analysis of variance (ANOVA) to detect QTL (Wijnen, 2019). Reportings on the use of CSLs for quantitative trait mapping in Arabidopsis are few (Koumproglou *et al*., 2002; Wijnen; 2019; Lardon *et al*., 2020). For each trait studied in this work we found several QTL – either additively or in interaction with other chromosomes (**Figure 7a**). However, most of these trait-chromosome associations are detected at only a low levels of confidence (**Figure 7a**), thus are unlikely to be considered after corrections for multiple pair-wise testing. This finding conflicts

with reports on the use of CSLs in animal model systems that describe increased genetic mapping power features over conventionally segregating populations (Spiezio *et al*., 2012; Buchner & Nadeau, 2015). A possible explanation for the low genetic mapping power of the population is that only a total of 32 lines is available in a population of *Arabidopsis thaliana*, thus each genotypic class in analysis is represented by only a limited number of lines. A framework for the statistical analysis of CSL-populations may still developed for best research practices. A major advantage of CSLs in this study is that the lines can be replicated in large numbers, thus increasing confidence of the line performance. In addition, the limited genomic architecture allows for a comprehensive analysis of chromosome-specific changes over time supported by high-throughput phenotyping (**Figure 8a-c**).

5. CONCLUSION

In this study, we developed a CSL-population developed from two parental lines that differ in their physiological responses to fluctuating light. We used this population to assess plant growth and physiological parameters under fluctuating light, considering fluctuating light for the first time as an abiotic stress in a control-treatment genetic mapping experiment. This experiment reveals that genotype x environment effects are prominent, particularly in relationship to a growth QTL that had its effects inhibited under fluctuating light conditions. By contrast, an F1 hybrid grown from the parental lines exhibited high performance regardless of the treatment. This experiment signifies the role of fluctuating light as impairing growth, but also highlights a role for heterosis to play a role in abiotic stress resistance. Furthermore, this study contributes to experience in the use of CSLs in plants, which is still limited up to date. To conclude, we delivered insights and approaches that can lead to new perspectives to study the role of fluctuating light in crop productivity and the use of CSLs in quantitative trait analysis.

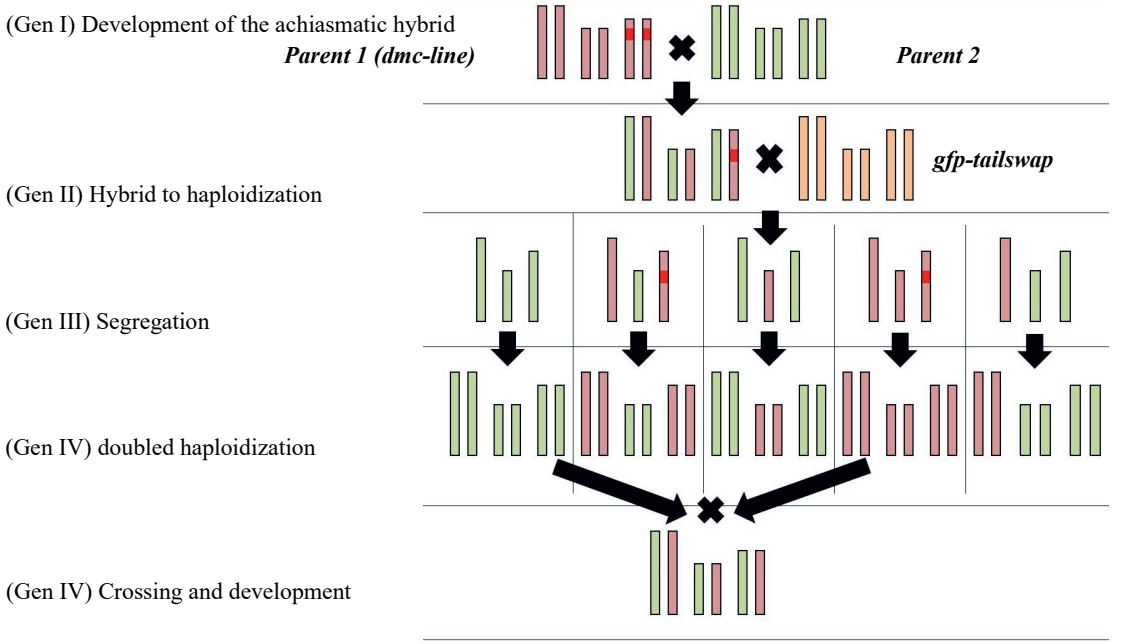
6. ACKNOWLEDGEMENTS

The authors would like to thank Gerrit Stunnenberg, Taede Stoker and David Brink for caring the plants and keeping an eye out. We also would like to thank Jan Aarts for developing the software that was used to extract meaningful data from the PLA images. Furthermore, we would like to thank Suzanne Vlinderveld and Gema Florés Andaluz for support with sowing and sampling activities. We also would like to acknowledge Rijk Zwaan bv for generously funding this research.

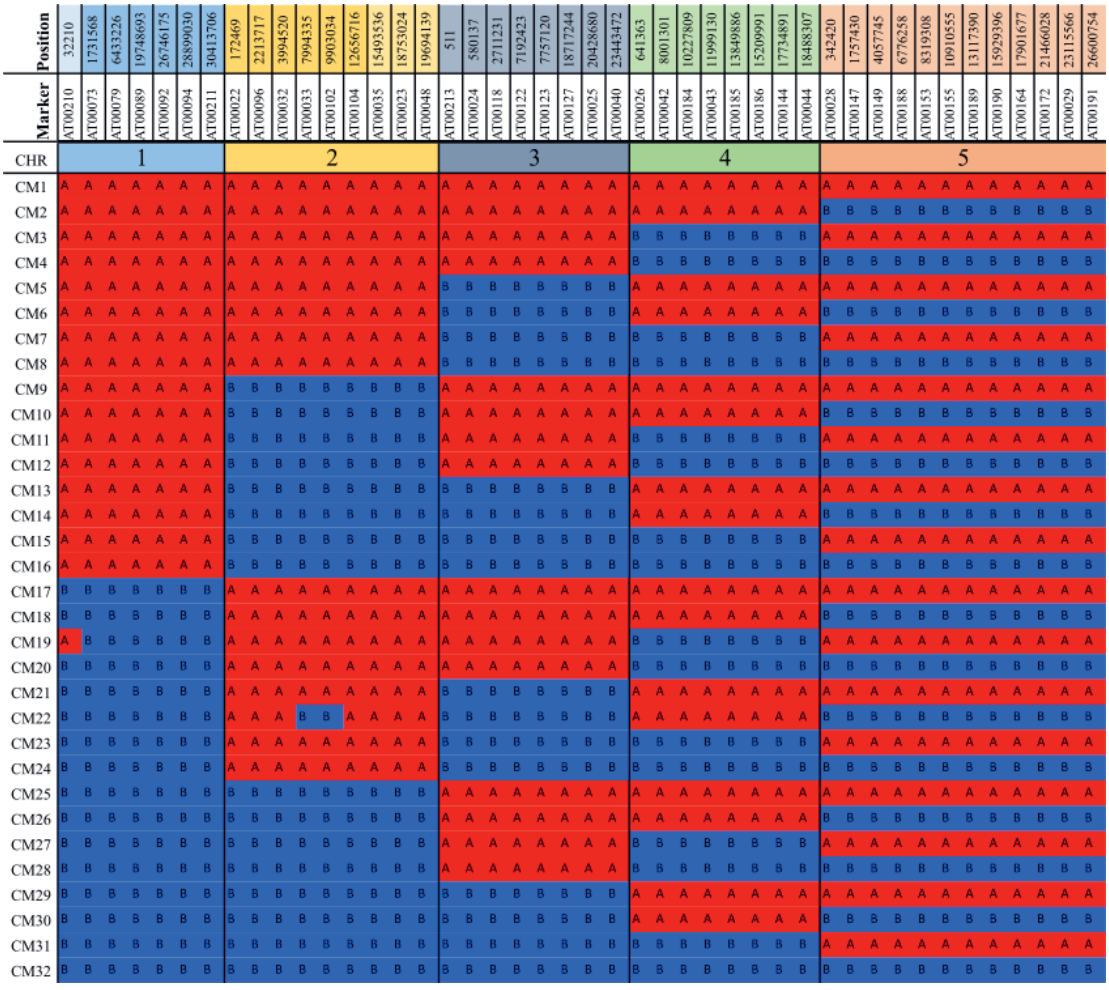
SUPPLEMENTARY FILES

Supplementary table 1. Heritability of various traits under conditions of fluctuating and constant light as measured in Kaiser et al. (2020). The reported values were obtained from data requested from the authors of the study. ΦPSII = quantum yield of photosystem II, NPQ = non-photochemical quenching, PLA = projected leaf area, Nleaf = leaf number.

	ΦPSII	NPQ	PLA	N_{leaf}
Fluctuating light	0.562	0.600	0.613	0.397
Constant light	0.243	0.246	0.405	0.381



Supplementary figure 1. Crossing scheme scheme of a chromosome substitution library Arabidopsis. Only chromosomes I, III and V are shown for simplicity. In generation I An achiasmatic genotype transformed with a dmc recombinant silencing construct (Wijnker et al., 2012) is crossed to a second parental line of interest. In generation II, such achiasmatic hybrids are crossed to gfp-tailswap (Ravi & Chan, 2010) to induce haploidization. The resulting offspring (generation III) consists of haploid lines segregating for whole chromosomes, but doubled haploidization in the next generation IV leads to homozygous offspring. If a large enough number of such lines are generated, all possible combinations (for Arabidopsis – 5 chromosome species – 2^5=32) can be retrieved. However, many lines will inevitably still contain the undesired dmcsilencer construct. V) To obtain lines free of dmc construct and those that may be harder to derive from random segregation, specific crossings can be performed to further develop missing lines, ensure absence of the dmc construct and the recreation of a transgene free hybrid.



Position	32210	1731568	6433226	19748693	26746175	28899030	30413706	172469	2213717	3994520	7994335	9903034	12656716	15493536	18753024	19694139	511	580137	2711231	7192423	7757120	18717244	20428680	23443472	641363	8001301	10227809	11999130	13849886	15209991	17734891	18488307	342420	1757430	4057745	6776258	8319308	10910555	13117390	15929396	17901677	21466028	23115566	26600754
Marker	AT00210	AT00073	AT00079	AT00089	AT00092	AT00094	AT00211	AT00022	AT00096	AT00032	AT00033	AT00102	AT00104	AT00035	AT00023	AT00048	AT00213	AT00024	AT00118	AT00122	AT00123	AT00127	AT00025	AT00040	AT00026	AT00042	AT00184	AT00043	AT00185	AT00186	AT00144	AT00044	AT00028	AT00147	AT00149	AT00188	AT00153	AT00155	AT00189	AT00190	AT00164	AT00172	AT00029	AT00191
CHR	1							2									3								4								5											
CM1	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM2	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM3	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM4	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM5	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM6	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM7	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM9	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM10	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM11	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM12	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM13	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM14	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM15	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM16	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM17	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM18	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM19	A	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM20	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM21	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM22	B	B	B	B	B	B	B	A	A	A	B	B	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM23	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM24	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM25	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM26	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM27	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM28	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
CM29	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
CM30	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	B	B
CM31	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	A	A	A
CM32	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B

Supplementary figure 2. Genotypic evaluation of the Col-0 x MIB-22 chromosome substitution library population. Each line is a chromosome substitution. A's represent markers that score the Col-0 allele, while B's represent markers that score the MIB-22 allele. Position gives the molecular position of the respective markers, relative to the Col-0 TAIR10 reference genome. The markers used to genotype the population are previously described in Wijnen et al. (2018).

Chapter 6. Development, use and analysis of chromosome substitution line populations in plants; a modelling approach

R.F.H.M. van Bezouw[1], T.G.Wijnker[1], J.J.B. Keurentjes[1]

SUMMARY

Chromosome substitution line (CSL-)populations are a novel type of genetic mapping tool popular in animals, but have only rarely been used in the study of the plants. The development of reverse breeding technology in the past decade, most notably the combination of transgenic suppression of meiotic recombination and doubled haploidization, has led to the opportunity of more rapidly developing CSLs. In this study, we explore the strategies involving the development and application of reverse breeding and CSLs using a permutation based modelling approach. The development of a CSL-population in which all possible CSLs are present may be an endeavour too costly to pursue. However, we do conclude that by only a marginal increase of efforts required to reconstruct an F1-hybrid in a reverse breeding program, powerful genetic mapping populations can be constructed. This knowledge will particularly aid breeding programmes involving multiple, simultaneous reverse breeding experiments. In all, the tools and insights developed in this study may aid those interested in engaging in reverse breeding – once biotechnological advancements will make this more feasible in crop species.

---

[1]Laboratory of Genetics, Wageningen University & Research

1. INTRODUCTION

Reverse breeding – first conceptualized by Dirks *et al*. (2009) – is a method that is aimed at retrieving homozygous progeny from an F1 hybrid via suppression of meiotic recombination and subsequent doubled haploidization. Doubled haploid progeny can then be used to recreate the F1 hybrid. This technology is expected to rapidly increase the rate by which new F1 hybrids can be developed in crop species. In addition, application of reverse breading may lead to the build-up of excess genotypic material that consists of lines with unique combinations of chromosomes originating from the parental lines (Wijnker *et al.*, 2012, 2014). Larger numbers of such lines are reminiscent of chromosome substitution line (CSL-)populations – sometimes referred to as chromosome substitution libraries (Wijnen, 2019) – that can be used as segregating populations in genetic mapping studies. Traditionally, CSL-populations have been developed in a select number of species by conventional backcrossing schemes, substituting one chromosome at a time in the recipient genotype (Nadeau *et al*., 2000; 2012; Koumproglou *et al*., 2002; Cowley *et al*., 2004; Singer *et al.,* 2004, Wu *et al*., 2010). Such populations essentially represent near isogenic lines (Kooke *et al*., 2012), but with the distinct feature that donor introgressions span full chromosomes. CSL-populations have been successfully applied to study various traits in a wide range of organisms, e.g., rat (Cowley *et al*., 2004), Drosophila (Wang *et al*., 2008), mouse (Nadeau *et al*., 2002; Krewson *et al*., 2004) and Arabidopsis (Koumproglou *et al*., 2002). In animals, CSL-populations are a standardized concept and considered to be a much more powerful genetic mapping tool than alternative, segregating genetic mapping population types, particularly with regards to the study of epistasis (Nadeau *et al*., 2002, 2012; Buchner & Nadeau, 2015; Spiezio *et al.*, 2012; Miller *et al*., 2020).

Plants are generally more tolerant to inbreeding than animals. Thus, a wide variety of immortal, inbred genetic mapping population types is available in plants that all have superior QTL-detection accuracy to the level of (groups of) molecular markers or smaller introgressions (Wijnen & Keurentjes, 2014; Bazakos *et al*., 2017). Although a CSL-population that consists of single substitution variants can be grown at a much lower volume of plants than immortal segregating population types, the disadvantages are perceived to outweigh the potential benefits as can be inferred from the very few reports on their use in quantitative trait analysis in plants (Koumproglou *et al*., 2002; Wu *et al*., 2010; Saha *et al*., 2021). Advances in doubled haploidization technology in crops (Dunwell *et al*., 2010, Ren *et al*., 2017; Kuppu *et al*., 2020; Fv *et al*., 2021), increased knowledge of meiotic crossovers (Wijnker & de Jong, 2008) and technology to control meiosis (Wijnker *et al*., 2012; 2014) has led to procedures of developing CSL-populations that have genotypes containing all possible combinations of multiple chromosome substitutions in *Arabidopsis thaliana* (Wijnen, 2019; **Chapter 5**). Beyond hybrid recreation, such a population introduces new angles and opportunities for the use of CSLs in breeding, including, but not limited, to the analysis of epistasis, efficient use in "omics" approaches and genetic mapping of heterosis (Wijnen, 2019). Meiotic suppression via virus induced gene silencing is a recent development, which improves flexibility of implementation and is transgene free, which may enable reverse breeding technology in crop species at a faster and more efficient rate than was previously possible (Calvo-Baltanás *et al*., 2020).

Despite such technological advances, the use of CSLs as a mapping population is still very limited – particularly because there is no to little experience with implementation of reverse breeding methodology. In species with larger genomes such as soy (*Glycine max*) (2n = 20) and tomato (*Lycopersicon esculentum*) (2n = 24), such populations may consist of large numbers of unique CSLs; a maximum of 1024 (2^{10}) and 4096 (2^{12}), respectively. Populations of this size exceed the number of lines that are required for a typical recombinant inbred line population of these species and thus their practical use is challenged. However, there may exist a degree of redundancy of how many CSLs may still make a reliable genetic mapping population, which would open more opportunities for their use. Obtaining such an incomplete CSL-population may also come with an imbalanced allelic distribution that cannot be controlled during the development of such lines, potentially affecting the reliability of a population to map traits and the interpretation of outcomes. This may impact the advantage of such a CSL-population, particularly the ability to map epistasis (Carlborg, 2004; Laurie *et al*., 2014; Lachowiec *et al*., 2015;

Mackay, 2014) – which is purported as an important advantage of CSL-populations (Wijnen & Keurentjes, 2014; Wijnen 2019). Such considerations may impact decision making in reverse breeding and initiation of the development of population of CSLs.

Technology supporting reverse breeding slowly gains traction and may increase efficiency of using CSLs-populations. However, the willingness to adopt novel breeding techniques always depend on strategic and economic considerations, and associated opportunity costs (Bernardo, 2016). To obtain more insight and quantify relevant considerations in reverse breeding, development of CSL-populations and their use in genetic mapping, we appled an *in silico* analysis to explore possibilities and pitfalls that are relevant for those engaged or wishing to engage, in applying reverse breeding, and the development and use of CSL-populations. In this work, we aim to answer research questions that involve the following topics; i) the impact of segregation distortion on reverse breeding and hybrid recreation, ii) the economic aspects related to the development of CSL-populations, iii) genetic mapping power of incomplete CSL-populations and iv) the influence of the genetic background and population architecture on the detection power of QTLs. To provide answers to these concerns, we developed several functions in R (R Core Team, 2021) to support them with data. In addition, we support our findings using an example of flowering time in a complete CSL-population that was previously developed (**Chapter 5**). In the discussion section we review the outcomes of our modelling study and translate these into strategic considerations that are useful when working on reverse breeding and development of CSL-populations.

2. METHODS

2.1 Assumptions on modeling development and production of (populations of) chromosome substitution lines

Despite technological advances, transgenic reduction of meiotic crossover suppression to develop chromosome substitution lines (CSLs) in plants are not flawless. Recombination events may still occur, though at a vastly reduced rate (Wijnker *et al*. 2012, Calvo-Baltanás *et al*., 2020). The ability to efficiently retrieve "clean" CSLs –i.e., CSLs in which no recombination events are present (**Figure 1**) – is essential to practice cost-effective reverse breeding and the development of CSL-populations. Regardless of the exact technology applied the feasibility of producing clean CSLs depends on two factors; the degree of efficiency of finding clean CSLs and the regeneration of doubled haploids. Efficiency of finding clean CSLs relates to the number of clean CSLs that are found relative to the residual CSLs that still contain crossovers. Such an efficiency applies to transgenic suppression of meiotic recombination (Wijnker *et al*. 2012, Calvo-Baltanás *et al*., 2020), but may also be deducted from extensive backcrossing strategies. The efficiency of obtaining doubled haploids relates to the number of non-aberrant plants that can be obtained among offspring and may be intra- and interspecific – if applied at all. CSLs that have one or few crossovers have also been considered to be useful in reverse breeding (Dirks *et al*., 2009) and quantitative trait analysis (Calvo-Baltanás *et al*., 2020). Thus, the efficiency of obtaining clean CSLs is dependent on many factors and in how far crossovers are tolerable following the objectives of any given projects. Accurate figures can't be easily predicted as too many variables have to be taken into account with only few cases being reported (Koumproglou *et al*., 2002; Wijnker *et al*., 2014; Guan *et al*., 2015; Calvo-Baltanás *et al*., 2019), and are thus not considered in our models. For modelling purposes, we assume that the development of or genetic mapping in CSL-populations are exclusively composed of transgene free, e.g., clean CSLs following a methodology as described in Calvo-Baltanás *et al*., (2020) at 100% efficiency – unless otherwise specified – of meiotic crossover reduction (**Figure 1**). We further assume that aneuploids or other aberrations following haploidization techniques are selected against before genotyping Wherever applicable in our work, predicted efficiencies for obtaining clean CSLs must be multiplied to obtain more realistic numbers, efficiency depending on the situation. Our models can also be adapted to fit any purpose including efficiencies with alternative underlying assumptions, would scenarios demand such.

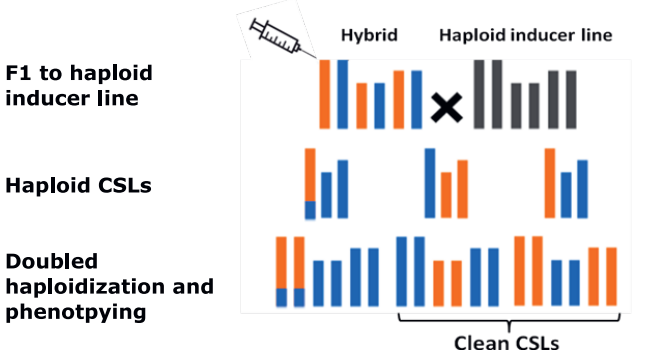


Figure 1. *Reverse breeding mediated by virus induced gene silencing (adapted from Calvo-Baltanás et al., 2020). Transgene free suppression of meiosis is introduced in F1 hybrids developed from unknown parents (blue/orange and cyan/magenta). The F1 is then crossed to a haploid inducer line (in grey). The haploid offspring then goes through a phase of spontaneous doubled haploidization, which should result into chromosome substitution lines. Some lines will experience incomplete meiotic suppression, and thus introduce CSLs with one or few cross-overs. Lines without cross-overs are considered clean CSLs.*

2.2 The development trajectory of CSL-populations follows the distribution of the mathematical coupon collector'sproblem

For research questions related to sampling for development of populations of CSLs we used equations underlying the mathematical problem known as the coupon collector's problem (Croucher, 2006). In short, this mathematical problems seeks to resolve the question of how many samples need to be drawn, at random and with replication, from an (assumed) infinite source of samples to collect a set including all unique samples, under the assumption of an equal probability of obtaining a sample. Following this distribution, diminishing returns are expected when increasing numbers of samples are being collected, as the chance of obtaining specific samples already in the collection increases when more unique samples had previously been found. In this work, we used the equations in Croucher (2006) to calculate sampling and developmentrates, and costs associated with the production of complete CSL-populations. However, a mathematical model is often unable to account for complicated deviations of likelihoods that are commonly found in genetics resulting, e.g. segregation distortion of genomic loci, and would thus require highly complicated mathematical solutions. To probe into questions related to these factors, we developed a series of permutation models as described below.

2.3 Permutation models for population development strategies and reverse breeding.

Statistical permutation models enable the answering of complex questions that are difficult, if not impossible, to resolve using mathematical equations, and are also easy to adapt for more complex and specific tasks. To answer questions that we were not able to resolve using mathematical equations given in Croucher (2006), we developed a series of permutation models in R (R Core team, 2021). Two functions form the basis of most of our models; SimLib (**Supplementary file 1**) and DrawIncCSL (**Supplementary file 2**). SimLib generates a complete CSL-population in a SimLib format, while DrawIncCSL gives random sets of unique genotypes from a SimLib simulated CSL-population vector. When generating random populations, we assume the parental lines to always be included in the selection. In doing so, we ignore potential genetic variation in cytoplasmic genomes – the mitochondria and chloroplasts – that might affect their performance, as such effects have previously been found to be negligible for most traits (Flood *et al*., 2020). The ReverseBreeding function (**Supplementary file 3**) is an adaptation to the DrawIncCSL function, and simulates a reverse breeding experiment in which infinite random of CSLs are drawn with replication from a SimLib CSL-population vector until a set of unique lines is sampled that can recreate the original F1. The function returns the total number of lines that were required to do so. Each run of the function can be interpreted as a unique, independent reverse breeding experiment, and so the output will vary accordingly. The ReverseBreeding function enables the inclusion of scenarios of segregation distortion, thus allowing for non-random passing over of specific or all chromosomes and at different rates.

2.4 Testing QTL detection power and analyzing population architecture in incomplete CSL-populations

To be able to test the detection power of (interacting) QTLs of incomplete populations, DrawIncCSL was extended to include code to perform analysis of variance (ANOVA) to test mapping power of (incomplete)CSL-populations (**Supplementary file 4, 5**) for additive (add.sampler) and interacting (int2.sampler) QTLs. In short, these models draw pre-determined numbers of random CSLs with their accompanying phenotypes, and then calculate the average -log10 linearized *p*-value, standard deviation of the $-\log_{10}(p)$ and total number of outcomes passing the threshold for significance following uncorrected ($p < 0.05$) and Bonferroni multiple testing thresholds ($p < 0.05/n_{\text{independent tests}}$) are scored for preset additive and interaction effects.The uncorrected and Bonferroni thresholdsare, respectively, the most liberal and conservative methodsto correct for multiple testing (van der Weele & Mathur, 2019) so outcomes of

alternative *post-hoc* tests are expected to fall within their range. In the case of epistasis, the number of random populations drawn in which no epistasis can be tested due to missing genotypic groups is also given. The function int3.sampler is an adaptation of the int2.sampler that can be used to score three-way interactions (**Supplementary file 6**), but is not used for analysis in this work. All these functions can be run for any number of permutations ("iter" term), thus it can be used to calculate confidence intervals for the reliability of incomplete populations in random subsets of any number of lines from the total possible number. These functions can become slow at high permutation numbers and larger chromosome counts. Finally, the CSL4 function was developed to analyze the population structure of incomplete CSLs and test the presence of "sets of four", i.e. the presence of sets of CSLs that enable the detection of a two-way epistatic analysis in an otherwise homogenous background (**Supplementary file 7**).

2.5 Flowering time experiment in the Col-0 x MIB-22 population of chromosome substitution lines

To validate outcomes obtained from our modelling approaches for incomplete populations of CSLs, we grew and phenotyped the Arabidopsis Col-0 x MIB-22 population (described in **Chapter 5**) for flowering time, a trait with a naturally high heritability in Arabidopsis (Brachi *et al*., 2010). Two of the lines, CM16 and CM22, do contain small introgressions of several million base-pairs in size on chromosome 1 and 2 respectively, but we assumed this genotype to be a clean CSL for purposes of this study. The population was grown in the greenhouse at once, confirmed by genotyping (see **Chapter 5**), then the seeds were harvested to ensure a uniform age of the seed batch across the population. Seeds were then stratified at four degrees for four days and grown in a greenhouse in September and October 2021, under stable temperatures with values of around 20/18 Day/Night. Plants were grown in three plastic trays at a rate of one replicate per genotype per block, except for the wild type parental lines which were grown for two replicates in each block, for a total of 36 (4x9) plants per block. Within each block plants were randomized completely. A fourth block was sown and served as a reserve block. After seven days of growth, several replicates that did not grow in the main three blocks were replaced by plants from the reserve block to make sure that our dataset would not contain any missing data. Flowering time was only scored for the three main blocks and was defined as the day, since sowing, of the onset of the opening of the first flowering bud to have visible petals. Flowering time was tracked until all plants had flowered and (interacting) QTLs for flowering time were detected following ANOVA on the genotypic means, using the add.sampler and 2int.sampler functions, to detect additive and epistatic QTLs, respectively.

3. RESULTS

3.1 How does segregation distortion affect the ability to recreate an F1 hybrid in reverse breeding?

Segregation distortion – the phenomenon in which specific alleles (or in this case chromosomes) pass over to progeny at a non-Mendelian rate – is not an uncommon phenomenon in crop species (Coulton *et al*., 2020), and thus may need to be taken into account for when planningfor reverse breeding experiments (Wijnker *et al*., 2012). Signatures for segregation distortion for whole chromosomesare previously reported and validated in an F2-mapping experiment for *Arabidopsis thaliana* in reverse breeding experiments performed by Wijnker *et al.* (2012). Dirks *et al*. (2009) provide mathematically derived numbers of CSLs to obtain from a reverse breeding experiment to be able to recreate the F1 hybrid from reverse breeding, however, inclusion of segregation distortion in mathematical equations may be challenging. Using the ReverseBreeding function, we simulated reverse breeding experiments to calculate the number of offspring lines required to obtain successful hybrid recreation in diploid species with 5, 7, 9 and 11 chromosomes (**Figure 2**). In addition, we tested the effect of several scenarios of segregation distortion, including distortions for single chromosomes and multiple chromosomes. We also assessed the impact of several arbitrarily defined "realistic" scenarios, as outlined in **supplementary table 1**, in which varying distortion numbers per chromosome are present with a maximum of 0.7/0.3 – thus the preferred chromosome variant has a 70% of passing over to the offspring whilethe non-preferred one has a 30% chance during meiosis –, but with standardized means and deviations of distortion for each chromosome. For Arabidopsis (five chromosomes), we included a scenario by using the numbers reported in Wijnker *et al*. (2012) (**supplementary table 1**).

Following these simulations, we predicted (nearly)the exact numbers of clean CSL lines required to recreate F1 hybrid in crop species of the indicated crop genome sizes at the 95% confidence level as were previously mathematically derived by Dirks *et al*. (2009) (**Figure 2**), which validates our approach.A single chromosome exhibiting a strong distortionfor one chromosome (0.9/0.1)almost doubles the required efforts to recreate a hybrid from crop species in all chromosome counts tested. By contrast, weak segregation distortion (0.6/0.5) on one or all chromosomes does only marginally increase the required number of lines to successfully perform reverse breeding compared to no segregation distortion. The arbitrary "realistic" scenarios were more difficult to place in context. While the average and standard deviation of each "realistic" scenario is equal (**Supplementary table 1**), the relative increase of CSLs required to recreate F1's at the 95% confidence threshold varies across chromosome count and the outcomes may thus be specific to the input distortion rates.In all, these simulations imply that the degree of segregation distortion on each chromosome may be crucial to predict the chance of finding a complementing pair of CSLs to recreate an F1 hybrid.

3.2 What are the production costs of developing a full population of CSLs, and how are these costs distributed during the developmental stages?

As with all genetic mapping population types, costs are associated with their development – be it growth for seed production, directed (back-)crossing or genotyping – of (intermediate) genotypes (Koumproglou *et al*., 2000; Kooke *et al*., 2012; Bazakos *et al*., 2017). For the development of CSL-populations, no numbers covering costs have previously been estimated that involve modern techniques supporting reverse breeding by combining crossover suppression (COS) and doubled haploidziation. Under the assumption that only clean CSLs are considered for the development of CSL-populations, the chance of obtaining new lines from a randomly drawn collection of clean CSLs follows a coupon collector's problem distribution (Croucher, 2006). Thus, we can mathematically infer how many clean CSLs need to be obtained in order to obtain a full population and thus determine the economic feasibility of doing. The total number of clean CSLs to be discovered in genotyping in order to complete a full population with lines containing all combinations of CSLs increases markedly, with each extra chromosome multiplying the total genotyping

effort by a factor ~2.4 (**Table 1**). As the total genotyping effort entirely depends on the achieved efficiency of meiotic suppression, the provided numbers require multiplication by that efficiency to account for genotyped lines that are not clean CSLs (**Figure 1**) – or otherwise passing a definition that can be project-specific. To provide some numbers from reported methodology we multiplied our outcomes with reported COS methodology in literature, including Wijnker *et al*. (2012) and Calvo-Baltanás *et al*. (2020) who report values of 94% and 22% efficiency, respectively. Several hypothetical scenarios with very low efficiency are also included for comparison. Our results indicate that at low COS efficiency and/or high chromosome count the required number of lines to genotype may easily run into hundreds of thousands – if not millions – to obtain enough clean CSLs to find all unique lines and develop a complete population. These numbers emphasize that highly efficient COS is required to develop complete populations for economically important crop species with larger genomes such as cotton (2n = 20), corn (2n = 24) and tomato (2n = 24).

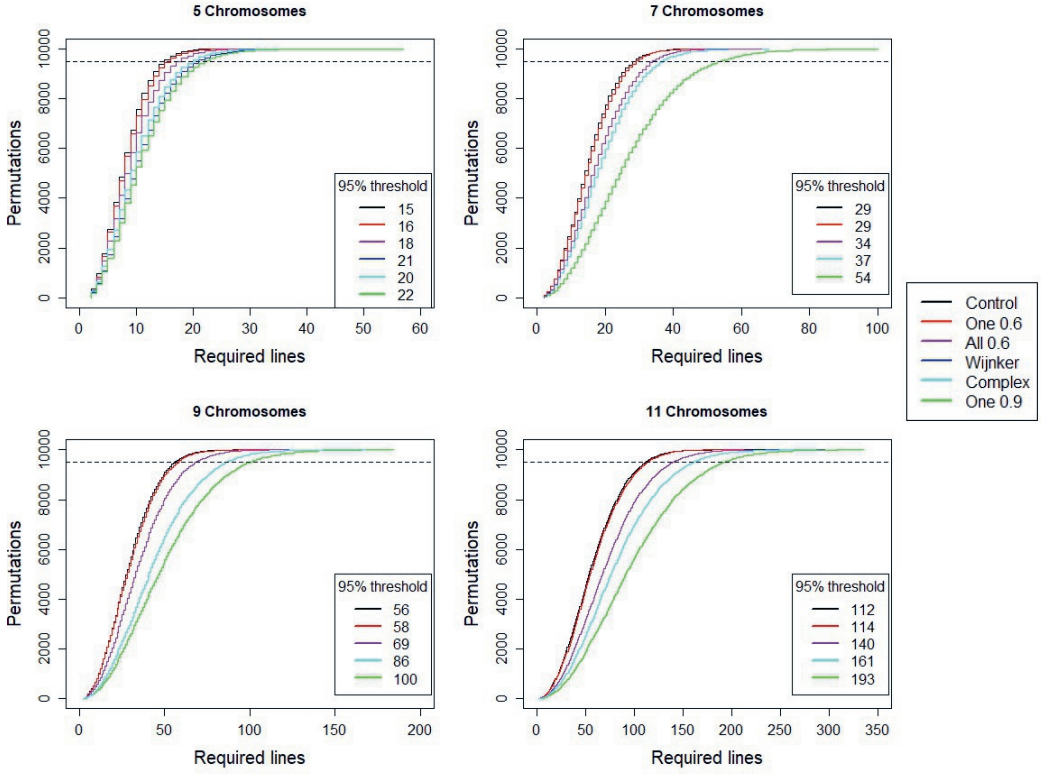


Figure 2. *Reverse breeding efficiency for populations of chromosome substitution lines (CSLs) originating from species with different unique chromosome counts. The curves are composed from 10.000 independent reverse breeding experiments, generated via permutation analysis using the ReverseBreeding function, for species with 5,7,9 or 11 chromosomes. The different colours indicate different scenarios in segregation distortion plays a role, thus where the probability of a chromosome passing over from the hybrid deviates from 50% (or 0.5/0.5). The scenarios are Control (Black): all chromosomes inherit at a rate of 0.5/0.5; One 0.6 (Red) one chromosome at a 0.6/0.4 rate, All 0.6 (Purple); all chromosomes have a 0.6/0.4 rate, Wijnker (Blue); Distortion rates as observed in progeny of the Col-0 x Ler-0 Arabidopsis thaliana hybrid by Wijnker et al. (2012), Complex (Cyan); arbitrary scenario's with equal average and standard deviation of distortion levels per chromosome (**Supplementary table 1**). Note that due to differences in chromosome*

number, the total rate of passing chromosomes of each species is different, but the mean and standard deviation are the same. One 0.9 (Green); One chromosome has a distortion rate of 0.9/0.1. The inserted legends give the number of clean CSL required for a 95% confidence of finding a complementing pair to reconstitute the F1 hybrid in each of the scenarios. In each of the plots, the 95% threshold line is indicated by the black dashed line.

Table 1. *Estimates of the total genotyping effort required to complete CSL-populations. For each unique chromosome count (x = 2n) we give the total number of unique CSLs that are expected (2^n) and the mean number of crossover suppressed doubled haploid lines that need to be discovered with the lower and upper bounds of the 95% confidence interval given at a 100% efficiency. The last five columns represent the means of several scenarios of the applied method for crossover reduction. Reported efficiencies for RNAi-mediated (Wijnker et al., 2012) and virus induced gene silencing (VIGS) methodology (Calvo-Baltanás et al., 2020) are given. In addition, three fictional methodologies M1, M2 and M3 are given with arbitrarily lower efficiencies of obtaining CSLs with no crossovers. In light grey shading, the combined chromosome count and method exceed an average of over 100.000 plants to be genotyped to obtain a full library. In dark grey, such combination require more than 1 million plants to be genotyped for a good chance of obtaining such a population.*

Chromosome count	Unique CSLs	Lower[a]	Mean[a]	Upper[a]	Wijnker[b]	Calvo-Baltanás[c]	M1[d]	M2[d]	M3[d]
			100%		*94%*	*22%*	*10%*	*2%*	*0.50%*
12	4096	44276	44628	44981	47477	202855	446280	2231400	8925600
11	2048	18609	18847	19085	20050	85668	188470	942350	3769400
10	1024	7529	7689	7849	8180	34950	76890	384450	1537800
9	512	3383	3490	3597	3713	15864	34900	174500	698000
8	256	1497	1568	1639	1668	7127	15680	78400	313600
7	128	649	695	742	739	3159	6950	34750	139000
6	64	273	304	334	323	1382	3040	15200	60800
5	32	110	130	149	138	591	1300	6500	26000

[a]*Calculated following equations in Croucher (2006),* [b] *Efficiency derived from Wijnkeret al. (2012),*[c] *Efficiency derived from Calvo-Baltanás et al. (2020),* [d]*Arbitrary efficiencies chosen for comparison*

Table 2. *Development of CSL-populations under fixed resource budgets. For each chromosome count (x = 2n) three scenarios are presented in which 100, 1000 or 5000 CSLs without meiotic crossovers can be developed. Statistics have been calculated using 1000 permutations from an adapted version of the ReverseBreeding function. For all populations we counted the average number of unique CSLs obtained mlines, the standard deviation of the average number of unique CSLs obtained (stdev), the percentage of the total number of unique lines are given (% ulines) and the fraction of completed – i.e. all unique lines possible are found – CSL-population (mean %).*

Chromosome count	100 non-crossover haploids				1000 non-crossover haploids				5000 non-crossover haploids			
	mlines	stdev	% ulines	mean %	mlines	stdev	% ulines	mean %	mlines	stdev	% ulines	mean %
12	98.8	1.09	2	0	887.6	22	22	0	2887.4	21.98	70	0
11	97.7	1.51	5	0	791.4	10.5	39	0	1869.4	10.7	91	0
10	95.4	2.05	9	0	637.8	9.98	62	0	1016.4	2.65	99	0
9	90.9	2.7	18	0	439.5	6.44	86	0	512	0.18	100	97.1
8	82.7	3.12	32	0	250.8	2.1	98	5	256	0	100	100
7	69.6	3.2	54	0	127.9	0.22	100	95	128	0	100	100
6	50.7	2.44	79	0	64	0	100	100	64	0	100	100
5	30.7	1.05	96	22.5	32	0	100	100	32	0	100	100

Following the distribution of the coupon's collector problem, there is an unequal chance of obtaining specific lines, which depends on the number of CSLs already found. The last several missing unique CSLs will be the most difficult to find and will require significantly more genotyping effort to obtain them. 62.5% of all possible lines in a species of 5 chromosomes (C5) (**Figure 3a**) and approximately 74.4% of all possible lines in C7 (**Figure 3b**), C9 (**Figure 3c**) and C11 (**Figure 3d**) species can be retrieved by only 25% of the genotyping efforts required to obtain the entire population. These numbers increase to 87.5% (C5) and ~93.5% (C7, C9, C11) of the total population by committing to 50% of these efforts. Thus, a relatively large proportion of CSL-populations can be obtained at reduced costs for genotyping.

Similarly, the possibility to build a population of CSLs may be further constrained by restrictions in resources to develop one, resulting from pre-established estimated and limited budgets. Thus, we also explored the development of CSLs under such constraints of genotyping by limiting the number of clean CSLs that we can obtain (**Table 2**). Again, these numbers must be multiplied against the efficiency of obtaining such lines to calculate realistic figures. We arbitrarily set the limitations on the possibility of obtaining a clean CSLs at three levels; 100, 1000 and 5000. These numbers of genotyped clean CSLs plants may already cover all – if not most – of the populations for species with up to 6, 9 and 11 chromosomes respectively (**Table 2**). At the level of COS efficiency reported for obtaining clean CSLs using virus induced gene silencing (22%, Calvo-Baltanás *et al*., 2020), this means that a population of a species with 7 chromosomes (e.g. rye, pea, barley, cucumber) may be within reach by genotyping a total of ~4.600 DH offspring. This amounts to a total genotyping effort of 48x 96-well plates for purposes of genotyping. If 1000 clean CSLs are obtained populations of species containing 8 and 9 chromosomes may already nearly reach completion. Production of 5000 clean CSLs may obtain 99% and 91% of the populations coming from crop species with 10 and 11 chromosomes (**Table 2**).

3.3 How effective are populations of CSLs in the detection of QTLs *at various levels of completeness?*

The development of incomplete CSLs may be much more feasible compared to aiming for obtaining a complete population for many crop species (**Figure 3a-d**). An incomplete CSLs may infer with the balanced genomic architecture that a complete CSL-population has to offer, but does retain the ability to study inter-chromosomal epistasis in a sophisticated way. Thus, we also aimed to test at which degree of incompleteness (interacting) QTLs for a given phenotype that are present in otherwise complete populations can still be detected at a confidence level of 95%. We tested several variables, including the chromosome count and the strength of QTLs and how these may impact performance. As epistasis is more difficult to detect in genetic mapping populations than additive effects (Ehrenreich, 2017; Zan *et al*., 2018), we modeled both additive and epistatic effects in full populations – with the epistatic effects being modeled as one combination of two chromosomes having a positive or negative effect on the phenotype. We performed these analyses in populations developed from species with chromosome counts 5 (C5), 7 (C7) and 9 (C9).

Random phenotypes were generated using the rnorm() function in base R (R core team, 2021) in complete CSLs for each of the chromosome count levels (C5, C7, C9). These random phenotypes were generated with a phenotypic average value of 5 and a standard deviation of 0.5 (10% of the mean). The base dataset was chosen so that the additive component for each chromosome explains less than 5% of phenotypic variation per chromosome (**Table 1**, **supplementary table 2**). The coefficient of variation in each these datasets is equal at 9-10% (**Table 3**). Each of the random values we assigned to a substitution line is assumed to be a best linear unbiased estimate of the genotypic means. Weak, medium and strong QTLs were introduced by increasing the phenotype of the appropriate genotypes by arbitrarily chosen values of 5%, 10% or 20% of the dataset means, respectively (**Table 3**), while the base dataset was used as a control without QTLs effects. No significant QTLs could be detected in the control datasets for all three levels of chromosome count (**supplementary table 2**), thus validating the dataset. We used the add.sampler and int2.sampler functions at 500 permutations per additional CSL per scenario of QTLs to

calculate the fraction of CSLs relative to a complete CSL-population that must be developed to obtain 95% confidence to detect the simulated (interacting) QTLs in the full populations.

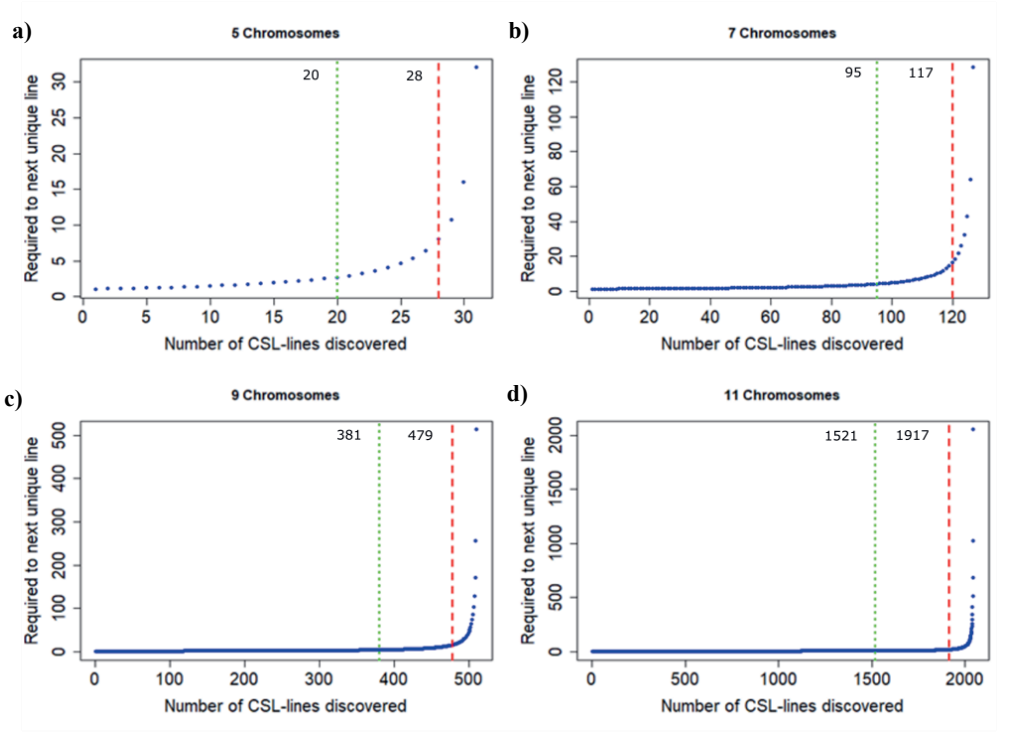


Figure 3. *Rate of discovering unique chromosome substitution lines (CSLs) under assumptions of the distribution of the coupon collector's problem (Croucher, 2006). The y-axis represents the average number of lines needed to genotype that would lead to the discovery of the next unique CSL-line, which depends on the number of CSLs already discovered on the x-axis. The figures are given for only CSLs in which no crossovers are found. In vertical lines, the mathematically derived number of obtained clean CSLs genotyped for purposes of developing a full population. The green dotted lines reflects the 25% effort threshold, while the red dashed line represents the 50% effort threshold. The cut-off numbers are given in number of CSL-lines discovered next to these lines.*

A much smaller proportion of possible CSLs of the C9-population is required to detect additive (**Figure 4a**) and epistatic QTLs (**Figure 4b**), compared to a C5-population. For strong additive effects, all populations populations require about 10-12 CSLs (or 13-19 at Bonferroni correction) for a 95% confidence to obtain the QTL, while for the 10% additive QTL this number increases to 41-42 (or 64-65 at Bonferroni correction) in the C7- and C9-population, while in a C5-population two-thirds of the entire possible population is required to reliably detect such an effect. Small effect size QTLs (here 5%) are not detectable in C5- and C7-populations. Populations of CSLs are hypothesized as methodology to detect epistatic interactions, but it is still more difficult than discovering additive effects in all populations (**Figure 4b**). From these figures, we conclude that a relataively small minimum number of lines is required to detect epistasis in species with larger chromosome counts implying high redundancy, while populations composed from species with few chromosomes must include most, if not all, of the possible lines to do so. Furthermore, introducing the Bonferroni correction effectively eliminates the option to detect small effect

size epistatic QTLs in C7 and C9. The detection of epistatic effects in the C5 population is not possible at all, unless no pair-wise comparison is taken into consideration and the epistatic effect size is large (here at 20% of the population mean).

3.4 Do background QTLs impact the ability to detect epistatic and additive effects in incomplete populations of CSLs?

Quantitative trait loci for traits in genetic mapping populations do, however, rarely come alone (Loudet *et al*., 2002; Ungerer *et al.*, 2002; El-Lithy *et al.* 2006; Keurentjes *et al.*, 2007, but see **Chapter 3**). The presence of multiple QTL may hinder the detection ability of individual QTLs due to introducing additional phenotypic noise, and in CSL-populations this will not be different. After considering the initial results across populations and scenarios (**Figure 4a, b**), we used the datasets of C7 carrying either an additive or epistatic QTL of medium (10%) effect size as a basis for introducing several additional combinations of varying numbers and strengths of epistatic QTLs to obtain a diverse set of testable scenarios (**Table 3, Figure 4c, d**). All background QTLs were located on chromosomes not already containing the initial target phenotype as in our original dataset. For analysis, we applied the same procedures as before, identifying the minimum population size of random CSLs required to identify the target 10% additive or epistatic effect QTL at a level of 95% confidence.

With increasing strength, background QTLs reduce the ability to reliably detect the 10% target QTL in the C7-population (**Figure 4c, d**), compared to no background effects (**Figure 4b, d**). Introducing a 20% effect size background QTL nearly doubles the required number of lines to reliably detect the 10% additive QTL (**Figure 4c**). Opposite effect background QTLs cause less severe effects in terms of reduced mapping power than introducing than introducing those in the same direction. For epistasis, introducing 10% and 20% background additive effects effectively eliminated the possibility to detect the target 10% epistatic effect in the C7 population under the Bonferroni threshold (**Figure 4d**). These examples show that for a full population of C7 the detection of epistatic effect interactions may still be challenging if background QTLs are present in the population.

3.5 What is the chance of obtaining combinations of CSLs that permit reliable analysis of epistatic interactions?

A potential application of CSLs is their use for analysis of inter-chromosomal epistatic interactions (Wijnen & Keurentjes, 2015; Bazakos *et al.*, 2017; Wijnen, 2019). Screening for epistasis proves to be difficult in normal segregating populations, but may be an important constituent of genetic variation (Laurie *et al*., 2014; Mackay, 2014). In CSLs, such an analysis can be performed population wide, but also for specific sets of lines that present enable such analysis. To optimally explore epistatic interactions the presence of specific lines that form good “sets of four” are to be present in the population, which strongly aids the interpretation of using a full population for assessment and validation purposes (**Figure 5a**). Thus, next to an evaluation of power to discover epistasis, the availability presence “sets of four” for all possible testable interactions may also benefit the use of CSL-populations for purposes of genetic mapping. We used the CSL4 function to permute the 95% probability threshold of obtaining a population containing “sets of four” for all possible two-way genetic interactions for populations with 5, 7 and 9 unique chromosomes at the 95% confidence level, using 1000 permutations. Similar to trends found in previous sections, smaller fractions of a complete CSL-population are required to obtain these sets of four for all interactions (**Figure 5b**), again concluding that there is more redundancy present for a population of CSLs at higher chromosome levels.

Table 3. *Summary statistics for each of the test datasets generated for the chromosome counts 5 (C5), 7 (C7) and 9 (C9), for purpose of testing the reliability and mapping power in incomplete populations. In brackets, the size of the datasets is given. The control datasets are base data sets without quantitative trait chromosomes (*QTL*s). Datasets were made by introducing one additive or interacting (epistatic) QTL, following the title descriptions. In some scenario's, background QTLs were added to test their influence on detection of the primary QTLs as indicated in the table. Add = Additive QTL, Epi = Two-way epistatic QTL. For background QTLs, a percentage denotes an additive QTL, a minus introduces a negative QTL and epi denotes the addition of a two-way epistatic QTL. Percentages denote the addition of the relative value to the average of the control. For each dataset the average, standard deviation and coefficient of variation are given.*

Datasets for C5 (n=32)

Primary QTLs	**Control**	**Add 5%**	**Add 10%**	**Add 20%**	**Epi 5%**	**Epi 10%**	**Epi 20%**	-
Average	4.991	5.116	5.241	5.491	5.054	5.116	5.241	-
Std Deviation	0.453	0.493	0.560	0.741	0.467	0.504	0.632	-
Coeff of Variation	0.091	0.096	0.107	0.135	0.092	0.099	0.121	-

Datasets for C7 (n=128)

Primary QTLs	**Control**	**Add 5%**	**Add 10%**	**Add 20%**	**Epi 5%**	**Epi 10%**	**Epi 20%**	-
Average	4.994	5.119	5.244	5.493	5.056	5.119	5.244	-
Std Deviation	0.495	0.508	0.550	0.697	0.508	0.543	0.662	-
Coeff of Variation	0.099	0.099	0.105	0.127	0.100	0.106	0.126	-

Datasets for C9 (n=512)

Primary QTLs	**Control**	**Add 5%**	**Add 10%**	**Add 20%**	**Epi 5%**	**Epi 10%**	**Epi 20%**	-
Average	5.038	5.164	5.416	5.920	5.101	5.227	5.479	-
Std Deviation	0.486	0.506	0.625	1.021	0.499	0.589	0.910	-
Coeff of Variation	0.096	0.098	0.115	0.172	0.098	0.113	0.166	-

Scenarios for C7_10 (additive)

Primary QTLs	**Add 10%**	**Add 10%**	**Add 10%**	**Add 10%**	**Add 10%**	**Add 10%**	**Add 10%**	**Add 10%**
Background QTLs	-	5%	10%	20%	-10%	-10% & 10%	epi 10%	epi -10%
Average	5.244	5.368	5.493	5.743	4.994	5.244	5.493	4.994
Std Deviation	0.550	0.568	0.612	0.757	0.603	0.661	0.606	0.603
Coeff of Variation	0.105	0.106	0.111	0.132	0.121	0.126	0.110	0.121

Scenarios for C7_10 (epistatic)

Primary QTLs	**Epi 10%**	**Epi 10%**	**Epi 10%**	**Epi 10%**	**Epi 10%**	**Epi 10%**	**Epi 10%**	**Epi 10%**
Background QTLs	-	5%	10%	20%	-10%	-10% & 10%	epi 10%	epi -10%
Average	5.119	5.244	5.368	5.618	4.869	5.119	5.368	4.869
Std Deviation	0.543	0.561	0.606	0.752	0.597	0.655	0.599	0.597
Coeff of Variation	0.106	0.107	0.113	0.134	0.123	0.128	0.112	0.123

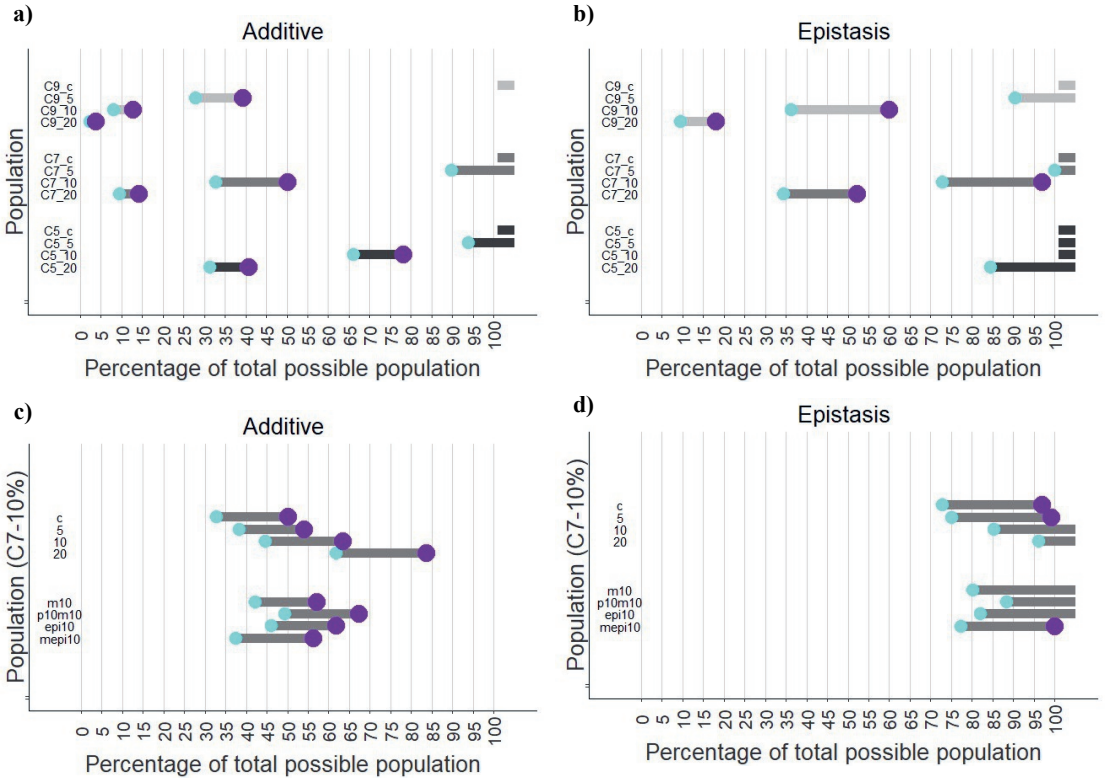


Figure 4. *Power analysis for incomplete CSL-populations for a set of scenarios. The x-axis represents a percentage of the total number of CSLs that can be produced from a the total number of possible lines n^2; so 5 chromosomes = 32, 7 = 128 and 9 = 456 total possible CSLs in a complete population. The introduced scenarios encompass simulated QTLs in each population, based on control phenotypes.* ***a)*** *additive effect scenarios,* ***b)*** *epistatic effect scenarios,* ***c)*** *additive scenario's for the C7_10 scenario population with introduced background effects of various sizes, types and quantitaties,* ***d)*** *As in* ***c****, but for the C7_10 epistatic scenario.C5 (in light grey) = population with 5 chromosomes, C7 (in grey) = population with 7 chromosomes, C9 (in dark grey)= population with 9 chromosomes. C = control dataset with no simulated QTLs, 5, 10, 20 = +5%, +10% or +20% simulated main effect (epistatic in "b") QTL added, m10 = -10% effect QTL added, p10m10 = -10% and +10% effect size QTLs added, epi10 = epistatic +10% effect added, mepi10 = -10% epsitatic effect added. In each graph, the small cyan point represents the minimum size that a random incomplete population can yield a confidence of 95% after of the uncorrected values p-values of the main additive or epistatic effect after 500 permutations at each level of incompleteness. Plot titles refer to the nature of the main effect (additive or epistatic). The larger, purple dot represents the 95% threshold after Bonferroni correction. The grey bars thus reflect these thresholds for v. Absence of dots indicate that the scenario fails to reliably detect QTLs. If no positive outcomes are found, a small bar is added at more than 100% of the population.*

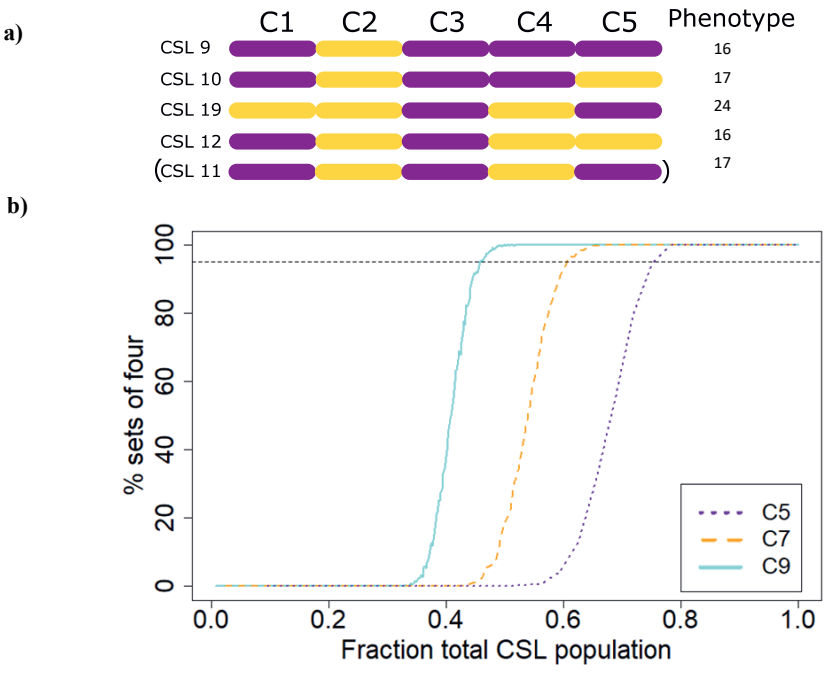


Figure 5. *Probability of obtaining "sets of four" CSLs for analysis of epistasis. a) An example subset of CSLs of a species with five chromosomes, with arbitrary phenotypes to underline the "sets of four" problem. Based on the first four CSLs, it is assumed that an epistatic interaction on chromosome 4 (C4) and C5 are present. However, because among these four lines the C1-C3 are not homogenous, difficulty might arise in interpreting the outcome due to background QTLs. In this case, the addition of CSL11 (between brackets) reveals that in fact there is no epistasis, but chromosome 1 is causal to the observed phenotypic effect. As such, incompleteness of a CSL-population may introduce a degree of uncertainty of correct interpretation of epistasis. b) The distribution of probability of obtaining sets of four in three different species following analysis of the CSL4 script (**Supplementary file 7**), for species with 5 (C5), 7 (C7) and 9 (C9) chromosomes at 1000 random permutations. On the y-axis, the probability of obtaining all possible "sets of four" in random population that represent a fraction (x-axis) of the total possible population that can be obtained for each chromosome number. For C5 a maximum of 10, for C7 a maximum of 21 and in C9 a maximum of 36 such sets can be found. The horizontal black lines represents the 95% threshold of confidence to obtain all these lines.*

3.6 Flowering time in a Col-0 x MIB-22 CSL-population; a case of testing the genetic mapping ability of incomplete CSL-populations.

The simulated data in the previous experiments follows a normal distribution and include a limited number of QTLs, both assumptions which in practice may not apply. Thus, we aimed to validate our outcomes using a simple quantitative trait with high heritability and scored flowering time in the Col-0 x MIB-22 CSL-population (**Chapter 5**) to provide an empirical phenotypic dataset. We found blocking to not affect flowering time ($F_{2,99}$ = 0.067, p = 0.935), and thus simply relied on the phenotypic means for each genotype. The broad-sense heritability is 0.91 (*genotypic variance/total variance*), average time of flowering is 29.67 days, the standard deviation of the genotypic means is 5.15 days, the coefficient of variation is 0.17 and the distribution of genotypic means skews slightly towards longer flowering times (**Figure 6a**). The parental lines exhibit a ten day difference in flowering time (**Figure 6b, supplementary table 5**) and a total of three additive QTLs are present as genetic factors explaining flowering time in the Col-0 x MIB-22 population on chromosomes C1, C2 and C4, with effect sizes ranging from 16.7% (relative to the Col-0 chromosome), -12.5% and 19.9% relative to the Col-0 alleles, respectively (**Figure 6b**). No epistatic interactions for flowering time are present in this population (data not shown), and thus we only focused on the additive QTLs by using this dataset as input for the add.sampler function to rediscover these in incomplete CSLs.

Would the Col-0 x MIB-22 population be incomplete, then the discovered QTLs can be discovered at a confidence level of 95% for a population size of 21 (0.66 of the total population), 31 (0.97) and 27 (0.84), random lines, respectively, at unprotected multiple comparison (**Figure 6c**). When implementing the Bonferroni correction threshold, only a complete CSL-population will consistently find the QTL on chromosome 1, at least 27 (0.84) random lines are required to detect the QTL for flowering time on chromosome 4 and the QTL on chromosome 2 is no longer detectable (**Figure 6d**). Furthermore, smaller incomplete populations reveal an increased probability to detect QTLs on chromosomes that are not present in the population with data from all lines (**Figure 6b**). This may probably be due to random pooling of outlier genotypes that are not present in the simulated datasets, which enhances the probably of detecting a false positive when the population is small as randomly selected extreme genotypes are more prominent.

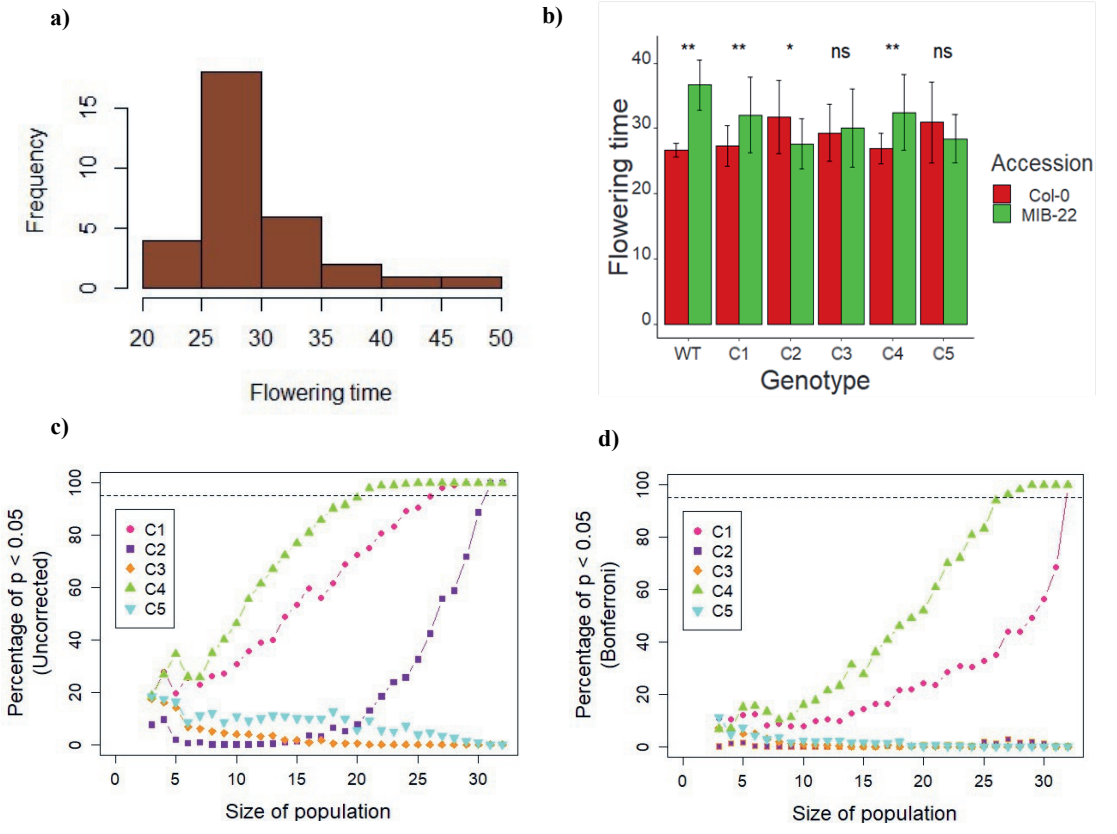


Figure 6. *Data and power analysis of the flowering time experiment of the Col-0 x MIB-22 CSL-population. a) Distribution of the genotypic mean flowering time for all 32 CSLs. b) Mean flowering time in the wild type parents (n=6 per WT parent) and the detection of additive QTLs on chromosome C1-C5 (n=16 per genotypic class), ** = p < 0.01, * = p < 0.05, ns = non-significant. c) Distribution of uncorrected p-values for additive effects for all five (C1-C5) individual chromosomes in incomplete CSL-populations, using the add.sampler() function for the flowering time phenotype. Each point represents the fraction of positives from 500 random permutations of the respective size of each incomplete population (x-axis). d) As in c, but for Bonferroni corrected values of p. In c) and d), the horizontal dotted line represents the 95% threshold of confidence.*

4. DISCUSSION

CSL (chromosome substitution line)-populations are a conspicuous genetic mapping population type with a unique genotypic architecture, characterized by absence of meiotic crossovers. By applying reverse breeding on an F1 hybrid, a finite number of unique clean chromosome substitution lines – i.e. without chromosomes carrying a crossover. However, little work is performed on development or use of CSL-populations in plants due to technological constraints and lack of standardization of analytical tools. As technology able to generate CSLs is improving (Calvo-Baltanás *et al*., 2020), we aimed to explore strategic considerations with regards to the economic development and use of CSLs for purposes of reverse breeding and genetic mapping. CSLs are an ideal model for analytic purposes as there is no reliance on the need to incorporate arbitrary and random recombination events that may affect the computational burden and/or the scope of generalizations for purposes of population modelling and development (e.g. Da Costa e Silva *et al*., 2007; Kooke *et al.*, 2012; Li *et al*., 2012; Garin *et al*., 2021). We primarily aimed to answer research questions related to the effect of segregation distortion on reverse breeding, the developmental costs of CSL-populations, redundancy in genetic mapping with CSLs and the impact of genetic population architecture. To accomplish and quantify these, we developed a series of R-scripts that can be applied to produce tailor made answers at need. To verify insights regarding genetic analysis of mapping in CSLs, we performed a flowering time experiment in the Col-0 x MIB-22 CSL-population.

4.1 Segregation distortion has a mild impact on the chance of recreating an F1 in reverse breeding.

The practice of reverse breeding itself is a relatively less laborious practice compared to developing an entire CSL-population. Still, numerical predictions with regards to the feasibility of reverse breeding derived from mathematical equations fail to account for random effects and unaccounted variation, which may lead to underestimated predictions. In the case of reverse breeding, as moderate segregation distortion is not uncommon in segregating genetic mapping populations (El-Lithy*et al*., 2006; Ashrafi *et al*., 2009; Ren *et al*., 2009; Wijnker *et al*. 2012), numbers provided by Dirks *et al*. (2009) that predict the chance of finding a complementing pair may be too optimistic. In this work, we tested this assertion by introducing various scenarios of segregation distortion in reverse breeding experiments (**Figure 2**). We conclude that a single chromosome exhibiting a high segregation distortion reduces chances of obtaining a complementary pair of CSLs forming an F1 pair more than moderate segregation at multiple chromosomes (**Figure 2**). Mild segregation distortion does not significantly impact the required effort to produce a complementary pair of CSLs that recreates the F1. Although we specifically and only tested the impact of segregation distortion in the context of reverse breeding, undoubtedly it will also come at play a relevant role in the development of complete CSL-population at similar rates of additional costs.

4.2 A high investment is required toe development complete CSL-populations

In line with the coupon's collector assumptions (Croucher, 2006), a considerable majority of the total number of possible combinations in a given complete CSL-population can be obtained by a quarter (25%) of the developmental efforts compared to one that consists of all possible unique lines (**Figure 3, Table 1, 2**). The remainder of these resources (75%) thus need to be spent on obtaining the last remaining lines, if the goal of the operation is to obtain a full population. These numbers imply that incomplete CSL-populations from four different hybrids can be developed for the efforts of completing one from a single hybrid, which makes the opportunity costs for the development of complete CSL-populations very high – particularly considering that incomplete populations maintain a relatively high mapping power compared to full populations (**Figure 4**). Nonetheless, if a complete CSL-populations is the goal, crossing of specific lines in near-complete populations may at some point be more resource efficient and deterministic than

continuing crossover suppression and doubled haploidization to find specific lines required to complete a population (Calvo-Baltanás *et al*., 2020). This is particularly relevant if segregation distortion of full chromosomes is observed, so that specific combinations can be targeted through specific crosses. In some plant systems, obtaining sufficient lines may become very challenging, particularly if the recovery efficiency for clean CSLs is low and the unique chromosome number in the species in question is high (**Table 1, 2**). Large quantities of unique CSLs may be discovered in the initial development phase, but the development of a complete CSL-population is in any way a very costly endeavour for crop species with higher chromosome numbers. Biotechnological advances need to improve development efficiency considerably, if complete CSL-populations are to compete with other genetic mapping population types that are much easier to develop in plants (Bazakos *et al*., 2017).

4.3 CSL-populations in species with low chromosome numbers may have low QTL-detection power, but high line redundancy increases reliability of incomplete CSL-populations made from species with higher chromosome counts.

An important aim of this study is to explore the potential of genetic mapping of chromosome substitution populations, particular those with higher chromosome numbers which are more challenging to develop. The adoption of new biotechnological tools have not yet led to the development of complete CSL-populations from species other than *Arabidopsis thaliana*(Arabidopsis). This species has only five chromosomes which greatly aids in developing such populations and testing proof of principle (**Table 1, 2; Figure 3**), but this advantage of having a small genome may also come at a downside in genetic mapping due to low observed QTL detection power (**Figure 4a**, **b**). The problem of limited number of genotypic means in such a segregating population can only slightly be accounted for by using higher replicates per genotype to increase the accuracy of the genotypic mean of individual CSLs (Keurentjes *et al*., 2007), but particular (combinations of) chromosomes may still produce outliers that increase statistical noise as shown in the flowering time experiment (**Figure 4a**). This may leave Arabidopsis as a sub-optimal model to study the potential of using of CSL-populations in analysis of inter-chromosomal epistasis. CSL-populations in species with higher chromosome counts are much more suitable in this regard (**Figure 4a, b**), but are much harder to obtain due to limitations in technology and longer growth cycles. In species with a higher unique chromosome counts, also a high degree of redundancy exists, i.e. minor fractions of all unique CSLs that can be produced are required to reliably detect QTLs (**Figure 4a**, **b**). This particularly true for the scenario's with simulated 10% and 20% additive QTL, as equal numbers of random CSLs are required to reliably detect QTLs for C7 an C9 populations in incomplete CSL-populations in the scenarios (**Figure 4a**). Thus, reliable genetic mapping in incomplete populations may be feasible to at least detect moderately strong effects in small, random collections of CSLs that may still be far away from a complete population.

Genetic mapping power in any genetic mapping population is dictated by the minimum number of lines that are appropriate to reliably detect a QTL (Keurentjes *et al*., 2007) and the balance of the allelic distribution per chromosome or marker. Accordingly, the presence of background QTLs reduces the feasibility of genetic mapping in CSL-populations due to observed reduction of detection power (**Figure 4c, d**), which is particularly true for epistatic QTLs (**Figure 4d**). Thus, the detection of medium effect size epistatic QTLs may still be challenging and requires larger numbers of CSLs from two parents, while for additive effects relatively incomplete populations will still suffice to reliably detect QTLs in the presence of background QTLs. The genetic mapping of traits in incomplete CSL-populations is still efficient for species with larger genome numbers.

To test the mapping power of QTLs in incomplete populations, we relied on a fixed phenotypic dataset per population type, rather than randomly introducing new datasets for every calculation. Although this puts the sampling strategy at risk of arbitrariness due to drawing conclusions from a single phenotypic dataset, random data invokes the risks of unforeseen outliers, and inhibits a comparative sensitivity analyses. The purpose of the modelling exercise was not to provide an exhaustive overview all possible outcomes that can be found in incomplete populations of CSLs, but rather to explore opportunities and

weaknesses of mapping in incomplete CSLs. Furthermore, we chose arbitrary values for the QTL effect sizes, the means and standard deviations in our simulated data. Discovery of a QTL always depends on the number of replicates for the relevant marker (or chromosome), the noise and effect size (Keurentjes *et al*., 2007). The use of data with different dimensions as reported here will always affect the outcomes of a modelling experiment, but we expect the generalizations to not deviate. The scripts in the **supplementary files** do allow the exploration of many different scenarios.

In our flowering time experiment we discovered a total of 3 QTLs controlling flowering time in the Col-0 x MIB-22 population. The observed confidence levels are much lower than those that are derived from, e.g., RIL and NIL populations for the same trait (Loudet*et al*., 2002; Ungerer*et al.*, 2002; Keurentjes *et al.*, 2007; Brachi*et al*., 2010), likely owing to the small number of lines in each genotypic class. MIB-22 exhibits a 40% longer observed flowering time in comparison to Col-0, and at high heritability of 0.91, which implies that high impact genetic variants do exist between these two accessions controlling the trait (**Figure 5b**). It is unknown how many genetic variants play a role explaining this difference between Col-0 and MIB-22 as flowering time in Arabidopsis thalianais controlled by many loci (Brachi *et al.*, 2010), but Col-0 and MIB-22 haven't previously been used for genetic mapping of the trait. The difference in flowering time between these two accessions is, however, much larger than between parents of many other reported bi-parental mapping populations (Loudet *et al*., 2002; El-Lithy *et al.* 2006; Keurentjes *et al.*, 2007; Wijnen, 2019). The lack of finding epistatic interactions in this population is in line with observations in **Figure 3a** and **b**, from which a CSL-population composed from a species with only five chromosomes has a limited ability to detect such. This may be further enhanced by the skewedness of the flowering time dataset compared to our simulated datasets in which several CSLs exhibit long flowering times (**Figure 5a**). Alternatively, it may also be that between Col-0 and MIB-22 no epistatic effects are present. In view of the considerations described here, our simulations may be too optimistic compared to real life data. However, all would depend on the standard deviation and the size of the tested and expected effects that are expected to be found in a given population. The models described in this study are excellent to extend towards such scenarios beyond those demonstrated in this work.

4.4 The Reverse Breeding & Mapping Strategy: Efficient use of CSLs in reverse breeding experiments.

Following analyses involving genetic mapping in this work (**Figure 4, 5**), the notion that CSL-populations exhibit particularly stronger genetic mapping power over alternatives – as is reported in animals (Buchner & Nadeau, 2015) – may not translate to plants. Importantly, as described in more detail in the previous section, we did not observe enhanced mapping power compared to alternative genetic mapping population types for additive or epistatic QTLs in CSLs phenotyped for flowering time (**Figure 5c, d)** or other traits as in (**Chapter 5**). We did, however, not perform a proper comparative research between a segregating population and a complete CSL-population composed from the same parents to test genetic mapping effectivity. Nonetheless, these findings may leave plant geneticists at risk with high opportunity costs to develop a CSL-population compared to alternatives that are developed at much lower costs and with higher precision (Bazakos *et al*., 2017). However, as incomplete CSL-populations are capable at relatively low numbers (**Figure 5**, combining reverse breeding and genetic mapping by introducing limited extra investments in the development of CSLs may prove to be a valuable breeding strategy.

Suppose a situation in which a breeder engages in reverse breeding on a hybrid of interest of a species containing 7 chromosomes. In such a species, a complementing pair of CSLs can be obtained after generating only 29-54 lines (**Figure 2**). Depending on segregation distortion present in the populations these will nearly all be unique (**Figure 3b**). Increased efforts to generate up to 80 CSLs may yield enough lines that can be used to detect additive QTLs and stronger epistatic effects (**Figure 3a-d**). The generation of these extra lines does not significantly increase the burden of obtaining lines, as these will still be far below the 25% effort of obtaining a complete population (**Figure 2b**). Such a population would also be sufficient to contain all "sets of four" for purposes of validating or detecting two-way epistatic interactions

(**Figure 4b**). The development of an incomplete genetic mapping population can thus be relatively cheap option and at a relatively low efficiency of generating clean CSLs, without losing much of the effectivity in genetic mapping that a complete CSL-population has. Most importantly, genetic mapping in this population can be performed in the doubled haploid stage, before a single hybrid has been (re-)created (**Figure 7**). Although less accurate than traditional segregating mapping populations, this is much more time-efficient than engaging in the development of new genetic mapping populations which takes several generations to complete (Kooke *et al.*, 2012; Bazakos *et al.*, 2017), while reverse breeding practice leans heavily on the capability to develop clean CSLs (Dirks *et al*., 2009). Genetic material suitable for mapping would thus already exist a hybrid is crossed and grown to evaluate its performance (**Figure 7**). This strategy can be employed in any crop species – provided that the generation of clean CSLs is possible and reasonably efficient (**Table 1, 2**). If multiple reverse breeding operations are run simultaneously, the best lines can then be selected for purposes of crossing between pools of CSLs which could potentially speed up the development of new hybrids even further.

5. CONCLUSIONS

The adoption of novel technologies should be scrutinized before use to carefully evaluate possible associated opportunities and weaknesses (Bernardo, 2016). A modelling approach is an appropriate and cost-effective form of identifying these, before putting such in practice to prevent high opportunity costs. In the case of reverse breeding and CSL-populations appropriate technology has advanced significantly, which prompted us to apply permutation and mathematical modelling to explore several points of attention for useful practical and strategic considerations. CSLs are particularly useful to perform such analyses because only a limited number of unique genotypes are available per cross. We conclude that segregation distortion is a factor that may slightly or moderately impact the chance of successful reverse breeding and the development of CSL-populations. Furthermore, we conclude that developing complete CSL-populations may be too expensive to compete with more conventional genetic mapping population types, but incomplete populations may be a reliable alternative. By translating strategic concerns into quantifiable outcomes we gained more insights in how to most effectively and efficiently plan experiments involving reverse breeding and the use of CSLs. Taking all our results together, we propose a strategy that involves genetic mapping in incomplete populations that result from producing a moderate excess of CSLs than is minimally required from reverse breeding experiments. Doing so may drastically improve the (re-)creation of new hybrids, particularly if multiple reverse breeding experiments being conducted simultaneously. In this work, we only considered the application of reverse breeding for and genetic mapping in CSL-populations. For additional possibilities in use CSLs we recommend the work of Wijnen (2019).

6. ACKNOWLEDGEMENTS

We would like to thank Evert Gutting, Kees van Dun and Geo Valikkakam James, for discussions in the early phases of this work. In addition, we would like to greatly thank Rijk Zwaan bv for a generous grant that supported this work.

7. COMPETING INTERESTS

Rijk Zwaan holds a patent for reverse breeding.

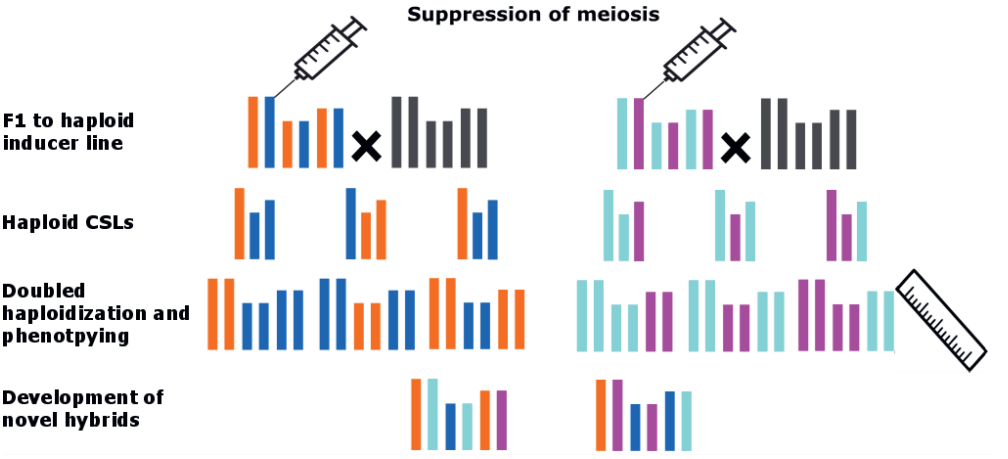


Figure 7. *The Reverse Breeding & Mapping strategy (adapted from Calvo-Baltanás et al., 2020). The figure depicts reverse breeding operations in two different hybrids (orange/blue and cyan/purple), as in figure 1. Immediate phenotyping and mapping in doubled haploid offspring should result in the detection of chromosome substitution lines that may be optimal for subsequent crossing and hybrid development. The perspective of using multiple, independent and simultaneous reverse breeding experiments is particularly interesting, as these may result into . In all, this strategy enables the production of novel hybrids in a time-span of just four generation, with information on the presence of QTLs.*

SUPPLEMENTARY FILES

Supplementary table 1. Complex **s**egregation distortion scenarios used as input for our Reverse Breeding modelling experiments using ReverseBreeding(). Each value represents the segregation distortion rate per chromosome for one of the alleles. Standard deviation and averages are given for the scenarios and are identical across scenarios. The values for Wijnker *et al*. (2012) are estimated from their figures, therefore this specific scenario thus only applies to Arabidopsis thaliana.

Population	1	2	3	4	5	6	7	8	9	10	11	Standard deviation	Average
C11	0.5	0.6	0.5	0.7	0.4	0.3	0.3	0.6	0.4	0.5	0.7	0.14142	0.5
C9	0.5	0.6	0.5	0.6	0.3	0.6	0.3	0.7	0.4	-	-	0.14142	0.5
C7	0.6	0.7	0.4	0.3	0.4	0.5	0.6	-	-	-	-	0.14142	0.5
C5	0.4	0.6	0.7	0.4	0.4	-	-	-	-	-	-	0.14142	0.5
Wijnker *et al*. (2012)	0.7	0.6	0.5	0.5	0.7	-	-	-	-	-	-	0.10000	0.6

Supplementary table 2. Test statistics of the control datasets for the randomly generated phenotypes of C5, C7 and C9 CSL-populations. All numbers are derived from a Type II sum of squares ANOVA model, which includes all chromosomes. Factor includes all terms as additive chromosomes, including the error term. % explained variance is calculated as the sum of squares of the individual chromosome divided by the total sum of squares.

C5	Factor	Sum Sq	df	F-value	*p*-value	% expl. var.
	C1	0.0084	1	0.038	0.848	0.132
	C2	0.1670	1	0.749	0.395	2.620
	C3	0.0233	1	0.105	0.749	0.366
	C4	0.1524	1	0.683	0.416	2.391
	C5	0.2254	1	1.011	0.324	3.536
	ε	5.7987	**26**			

C7	Factor	SumSq	df	F-value	*p*-value	% expl.var.
	C1	0.0011	1	0.004	0.948	0.0035
	C2	0.0002	1	0.001	0.978	0.0006
	C3	0.1356	1	0.533	0.467	0.4365
	C4	0.3569	1	1.404	0.238	1.1489
	C5	0.0490	1	0.193	0.662	0.1576
	C6	0.0021	1	0.008	0.928	0.0068
	C7	0.0132	1	0.052	0.820	0.0424
	ε	30.5036	**120**			

C9	Factor	SumSq	df	F-value	*p*-value	% expl.var.
	C1	0.0279	1	0.117	0.733	0.02316
	C2	0.0017	1	0.007	0.934	0.00137
	C3	0.0003	1	0.001	0.973	0.00022
	C4	0.0732	1	0.306	0.580	0.06072
	C5	0.0035	1	0.015	0.904	0.00289
	C6	0.3115	1	1.304	0.254	0.25848
	C7	0.0027	1	0.011	0.915	0.00228
	C8	0.0001	1	0.000	0.985	0.00007
	C9	0.1284	1	0.537	0.464	0.10656
	ε	119.9504	**502**			

Supplementary table 3. Flowering time data of the Col x MIB-22 CSL-population, including genetic layout (C1-5), the raw values for flowering time per block (b1-3), the average (mean) and the standard deviation (std) of the trait. Note that the Col-0 and MIB-22 genotypes have a total of six replicates each (i.e., two biological replicates in each block per genotype).

Genotype	C1	C2	C3	C4	C5	b1	b2	b3	mean	std
Col-0	c	c	c	c	c	26 27	25 27	28 27	26.67	1.03
cm01	C	C	C	C	C	28	25	27	26.67	1.53
cm02	C	C	C	C	M	30	27	26	27.67	2.08
cm03	C	C	C	M	C	29	27	28	28.00	1.00
cm04	C	C	C	M	M	32	36	34	34.00	2.00
cm05	C	C	M	C	C	26	25	24	25.00	1.00
cm06	C	C	M	C	M	26	28	28	27.33	1.15
cm07	C	C	M	M	C	29	30	29	29.33	0.58
cm08	C	C	M	M	M	33	35	33	33.67	1.15
cm09	C	M	C	C	C	26	25	23	24.67	1.53
cm10	C	M	C	C	M	26	25	25	25.33	0.58
cm11	C	M	C	M	C	27	27	26	26.67	0.58
cm12	C	M	C	M	M	27	27	26	26.67	0.58
cm13	C	M	M	C	C	24	25	23	24.00	1.00
cm14	C	M	M	C	M	22	23	22	22.33	0.58
cm15	C	M	M	M	C	27	26	27	26.67	0.58
cm16	C	M	M	M	M	31	26	30	29.00	2.65
cm17	M	C	C	C	C	29	31	31	30.33	1.15
cm18	M	C	C	C	M	30	29	30	29.67	0.58
cm19	M	C	C	M	C	41	34	36	37.00	3.61
cm20	M	C	C	M	M	40	40	41	40.33	0.58
cm21	M	C	M	C	C	27	29	30	28.67	1.53
cm22	M	C	M	C	M	29	30	28	29.00	1.00
cm23	M	C	M	M	C	34	36	35	35.00	1.00
cm24	M	C	M	M	M	46	46	45	45.67	0.58
cm25	M	M	C	C	C	27	24	27	26.00	1.73
cm26	M	M	C	C	M	29	32	31	30.67	1.53
cm27	M	M	C	M	C	29	26	26	27.00	1.73
cm28	M	M	C	M	M	29	28	27	28.00	1.00
cm29	M	M	M	C	C	29	25	26	26.67	2.08
cm30	M	M	M	C	M	27	25	28	26.67	1.53
cm31	M	M	M	M	C	33	32	34	33.00	1.00
cm32	M	M	M	M	M	36	41	39	38.67	2.52
MIB-22	m	m	m	m	m	41 34	31 37	39 37	36.5	3.56

```
#### SimLib
#
# n = the desired full population to be generated of size 2n = X

SimLib<- function(n) {
        l <- rep(list(0:1), n)
        Lines <- expand.grid(l)
        LineNames<- function(n) {
                Names = NULL
                for (i in 1:n) {
                Names <- append(Names, paste("C", i, sep = ""))
                        }
                return(Names)
        }
        names(Lines) <- LineNames(ncol(Lines))
        CSL <- c(1:nrow(Lines))
        Lines <- data.frame(CSL, Lines)
        return(Lines)
}

# Output generates a dataframe encompassing all possible lines from a cross between two genotypes of a species containing the desired number of chromosomes.
```

Supplementary file 1. *SimLib*; Generates a vector representing a simulation of an entire population of chromosome substitution lines of 2n = X.

```
#### DrawIncCSL
#
# Lines = a SimLib input file. Note that the file may include columns representing phenotypes
# y = the number of random lines to draw from the population. Note that p1 (all chromosomes 0) and p2 (all chromosomes 1) are always included (hence, to generate a population of size 16 of the 32 possible lines in Arabidopsis thaliana, one must use y = 14).

DrawIncCSL<- function(Lines, y = nrow(Lines)) {
                n <- ncol(Lines)
                Y <- ifelse(y > (nrow(Lines)-2), y <- (nrow(Lines)-2), round(y))
                Select <- Lines[c(1, sample(2:(nrow(Lines)-1), Y, replace = FALSE), nrow(Lines)), ]
                return(Select[order(Select$CSL),])
}
```

Supplementary file 2. *DrawIncCSL*; draw random populations of size y from a population in the *SimLib* format (Lines).

```
ReverseBreeding<- function(D = CSL, prb = NULL) {
        n <- 0
        count <- 0
        CSL <- D
        Collection <- CSL[-(1:nrow(CSL)),]
        while (n == 0) {
                Par <- CSL[c(sample(1:(nrow(CSL)), 1, prob = prb, replace = TRUE)), ]
                Collection <- rbind(Collection, Par)
                for (i in 1:nrow(Collection)) {
                        X <- sum(abs(Par[,2:ncol(CSL)] - Collection[i,2:ncol(CSL)]))
                        if (X == (ncol(CSL)-1)) {n <- 1}
        }
}
return(nrow(Collection))
}
```

Supplementary file 3. *ReverseBreeding*; Draws random CSLs (with replication) until a complementing pair is found that leads to the recreation of the F1 hybrid upon crossing (Dirks *et al*., 2009).

```
#### Add.sampler

#  Sample = the number of lines (with respective phenotypes) to draw from the population
#  database = a CSL-population with at least one phenotype (a SimLib file)
#  trait = Any trait column
#  nchr = the number of chromosomes in the population
#  iter = the number of random populations drawn

Add.sampler<- function(Sample = 1, database = Pop, trait = 7, nchr = 5, iter = 1000) {
P.values<<- list()
for (i in c(1:iter)) {
        p.outcome<<- c()
        Draw <<- DrawIncCSL(database, y = Sample)
        for (p in 2:(nchr+1)) {#the Chr columns
                AoV<<- AOV3(unlist(Draw[[trait]]) ~ as.factor(Draw[[p]]))
                p.outcome<<- c(p.outcome, -log(AoV["Pr(>F)"][[1]][[2]],10))
        }
        Val <- list(p.outcome)
P.values<<- c(P.values, Val)
        Sigs <- as.data.frame((do.call(cbind, P.values)))
}

Sigs <- as.data.frame(t(Sigs))

p95 <- c()
meanp<- c()
stdev<- c()
for (i in 1:ncol(Sigs)) {
        stdev<- c(stdev, sd(Sigs[[i]]))
        meanp<- c(meanp, mean(Sigs[[i]]))
        p95 <- c(p95, sum(Sigs[[i]]>-log(0.05,10))/nrow(Sigs)*100)
}
Results<- data.frame(t(Sigs))
Results$STDEV<- stdev
Results$P95 <- p95
Results$MEANP<- meanp
Results <- data.frame(t(Results))
return(Results[(nrow(Results)-2):(nrow(Results)),])
}

# STDEV = the standard deviation of the logarithm of the odds of the model outcomes
#
```

Supplementary file 4. *Add.sampler*; Sampling test statistics for tests of additive effects from incomplete populations of CSLs with phenotypes.

```
Int2.sampler <- function(Sample = 5, database = Pop, trait = 11, nchr = 5, iter = 10) {
P.values<<- list()
for (i in c(1:iter)) {
        p.outcome<- c()
        Draw <<- DrawIncCSL(database, y = Sample)
        for (p in 2:(nchr)) {     #the Chr columns
                for (q in p:(nchr)+1) {
                        if (q != p) {
                                summer <<- summary(aov(Draw[[trait]]~ as.factor(Draw[[p]]) * as.factor(Draw[[q]])))
                                if (length(summer[[1]][["Pr(>F)"]]) < 4) {
                                        p.outcome<- c(p.outcome, NA)
                                } else {
                                        p.outcome<- c(p.outcome, -log(summer[[1]][["Pr(>F)"]][[3]],10))
                }}
        }}
        Val <<- list(p.outcome)
P.values<<- c(P.values, Val)
        Sigs <- as.data.frame((do.call(cbind, P.values)))
}

Sigs <- as.data.frame(t(Sigs))

nna<- c()
p95 <- c()
p95_BF <- c()
meanp<- c()
stdev<- c()
for (i in 1:ncol(Sigs)) {
        nna<- c(nna, sum(is.na(Sigs[[i]])))
        stdev<- c(stdev, sd(Sigs[[i]], na.rm=TRUE))
        meanp<- c(meanp, mean(Sigs[[i]], na.rm=TRUE))
        Sigs[[i]][is.na(Sigs[[i]])] <- 0
        p95 <- c(p95, sum(Sigs[[i]]>-log(0.05,10))/nrow(Sigs)*100)
        p95_BF <- <- c(p95, sum(Sigs[[i]]>-log(0.05/ncol(Sigs),10))/nrow(Sigs)*100)

}

Results <- data.frame(t(Sigs))
Results$STDEV<- stdev
Results$P95 <- p95
Results$P95_BF<- p95_BF
Results$MEANP<- meanp
Results$NNA<- nna
Results <- data.frame(t(Results))
return(Results[(nrow(Results)-4):(nrow(Results)),])
}

# returns a file with the averaged LOD-values, the standard deviation, the number of ANOVA-outcomes being significant (p < 0.05, uncorrected), the number of outcomes being significant ((p < 0.05, Bonferroni).
```

Supplementary file 5. *Int2.sampler***.** Sampling test statistics for tests of two-way interactions from incomplete populations of CSLs with phenotypes.

```
Int3.sampler <- function(Sample = 1, database = Pop, nchr = 5, iter = 1000) {
P.values<<- list()
for (i in c(1:iter)) {
        p.outcome<- c()
        Draw <- DrawIncCSL(database, y = Sample)
        for (p in 2:(nchr+1)) { #the Chr columns
                for (q in p:(nchr+1)) {
                        for (r in q:(nchr+1)) {
                                if (p < q & q < r) {
                                summer <<- summary(aov(Draw[[trait]]~ as.factor(Draw[[p]])
*as.factor(Draw[[q]])as.factor(Draw[[r]])))
                                if (length(summer[[1]][["Pr(>F)"]]) < 4) {
                                        p.outcome<- c(p.outcome, NA)
                                } else {
                                        p.outcome<- c(p.outcome, -log(summer[[1]][["Pr(>F)"]][[3]],10))

                        }
                }}
        }
        Val <- list(p.outcome)
P.values<<- c(P.values, Val)
        Sigs <- as.data.frame((do.call(cbind, P.values)))
}
Sigs <- as.data.frame(t(Sigs))

nna<- c()
p95 <- c()
p95_BF <- c()
meanp<- c()
stdev<- c()
for (i in 1:ncol(Sigs)) {
        nna<- c(nna, sum(is.na(Sigs[[i]])))
        stdev<- c(stdev, sd(Sigs[[i]], na.rm=TRUE))
        meanp<- c(meanp, mean(Sigs[[i]], na.rm=TRUE))
        Sigs[[i]][is.na(Sigs[[i]])] <- 0
        p95 <- c(p95, sum(Sigs[[i]]>-log(0.05,10))/nrow(Sigs)*100)
        p95_BF <- <- c(p95, sum(Sigs[[i]]>-log(0.05/ncol(Sigs),10))/nrow(Sigs)*100)

}

Results <- data.frame(t(Sigs))
Results$STDEV<- stdev
Results$P95 <- p95
Results$P95_BF<- p95_BF
Results$MEANP<- meanp
Results$NNA<- nna
Results <- data.frame(t(Results))
return(Results[(nrow(Results)-4):(nrow(Results)),])
}

# returns a file with the averaged LOD-values, the standard deviation, the number of ANOVA-outcomes being significant (p
< 0.05, uncorrected), the number of outcomes being significant ((p < 0.05, Bonferroni).
```

Supplementary file 6. *Int3.sampler*. Sampling test statistics for tests of two-way interactions from incomplete populations of CSLs with phenotypes.

```
####        CSL4

#
# CHROMOSOMES;                  Indicate the number of chromosomes of the plant.A compete CSL will be generated sized 2^Y
# SAMPLEDRAW;                   Size of the sample (X) drawn from the entire population (Y)
# PERMUTATIONS;                 Number of sampledraws
# NINTERACTIONS;                Determine the interactions you wish to screen a population for #Note; Does not work yet!

####        Assumptions
#
#           No segregation distortion of chromosomes --> requires some model adjustments
#           Parents are always included (e.g., for CHROMOSOMES = 4, Line1 is 0 0 0 0 and Line16 is 1 1 1 1)

####        Output
#
#           Generates a list of the total number of "sets of four" present within a CSL-population for each permutation

CSL4 <- function(CHROMOSOMES, SAMPLEDRAW, PERMUTATIONS = 1000, NINTERACTIONS = 2) {

            nChr<- CHROMOSOMES
            nsubpop<- SAMPLEDRAW

            if (!require(plyr)) {
                        install.packages("plyr")
                        require(plyr)
            }
            SimLib<- function(n) {
                        l <- rep(list(0:1), n)
                        Lines <- expand.grid(l)
                        LineNames<- function(n) {
                                    Names = NULL
                                    for (i in 1:n) {
                                    Names <- append(Names, paste("C", i, sep = ""))
                                                }
                                    return(Names)
                        }
                        names(Lines) <- LineNames(ncol(Lines))
                        CSL <- c(1:nrow(Lines))
                        Lines <- data.frame(CSL, Lines)
                        return(Lines)
            }
            DrawIncCSL<- function(Lines, y = nrow(Lines)) {
                                    n <- ncol(Lines)
                                    Y <- ifelse(y > (nrow(Lines)-2), y <- (nrow(Lines)-2), round(y))
                                    Select <- Lines[c(1, sample(2:(nrow(Lines)-1), Y, replace = FALSE), nrow(Lines)), ]
                                    return(Select[order(Select$CSL),])
            }
            IntChr<- function(o) {
                        ChrNumber<- c(1:o)
                        Nint<- c()
                        for (i in ChrNumber) {
                                    for (j in i:length(ChrNumber))
                                                {if (i != j){
                                                            Nint<- append(Nint, list(c(i,j)))
                                                }
                                    }
                        }
                        return(Nint)
            }
            INT <<- IntChr(nChr)
            totCSLpop<<- SimLib(nChr)
            ResultINT<- c()
            for (i in 1:abs(PERMUTATIONS)) {
                        IntCount<<- 0
                        SLB <- DrawIncCSL(totCSLpop, nsubpop)
                        for (i in 1:length(INT)) {
                                    Haplo<- do.call(paste, SLB[-1][-INT[[i]]])
                                    df <- data.frame(Haplo)
                                    SLBdf<- data.frame(SLB[-1][INT[[i]]], df)
                                    DetectInt<- ddply(SLBdf, c("Haplo"), summarise, N = length(Haplo))
                                    if (any(DetectInt$N == 4)) {
                                                IntCount<<- IntCount + 1
                                    }
                        }
                        ResultINT<- append(ResultINT, IntCount)
            }
            return(ResultINT)
}
```

Supplementary figure 7. *CSL4*. This function to extract sets of four from incomplete populations of CSLs. The output can be used to calculate the probably of these sets for any number of randomly selected CSLs.

7. General discussion

1. Introduction

In **Chapter 2** (van Bezouw *et al*., 2019), I gave an update about the state-of-the-art of photosynthesis genetic research, and achievements and progress in high-throughput phenotyping of photosynthesis traits. In this work, I included many reports describing the genetic mapping of photosynthesis traits, thus marking the popularization and growing significance of studying natural variation, in crop and model species alike. In the next three chapters, I used biparental mapping population to identify novel genes and traits of relevance, and under three different light environments to demonstrate their potential in studying natural genetic variation in photosynthesis. In **Chapter 3**, I used a recombinant inbred line population from the lab accessions Col-0 and Ler-0 to study genetic factors that explain differences in photosynthesis under several environmental factors. Following up the QTL that I discovered here, I established a causal link between copy number variation between photosynthesis efficiency and an unexpected sub-family of *SQUALENE EPOXIDASES* genes using molecular cloning approaches, confirming the power and reliability of bi-parental mapping populations. In **Chapter 4**, I used a mixture of different strategies, including transcriptomics and genetic mapping to characterize the differential photosynthetic response to high light irradiance between the wild accessions S96 and SLSP30. I discovered QTLs showing a very large 25% difference in quantum yield of photosystem II (ΦPSII). To follow up, I performed growth assays under various fluctuating light conditions to see if they also affected biomass accumulation after repeated exposure to stepwise changes in irradiance. In **Chapter 5**, I aimed to describe variation in plant productivity to the response to fluctuating light in a newly developed population of chromosome substitution lines (CSLs), created from two parental lines that differ significantly in this response. A most interesting observation related to the hybrid between the parental lines (Col-0 x MIB-22). This hybrid shows increased growth associated with Col-0, but also high tolerance to fluctuating light conditions as associated with MIB-22. These findings rule out that biophysical processes limited productivity under such conditions. As populations of CSLs are laborious to develop, even in *Arabidopsis thaliana*, reports on CSLs used in quantitative trait analysis in plants are few. Thus, in **Chapter 6** I resorted to mathematical and statistical modelling to explore strategies and mapping power of CSLs, as an efficient way to highlight opportunities with regards to development of and mapping in CSLs. In this closing chapter of the thesis, I will elaborate my final thoughts and conclusions on the topics that I have addressed and scrutinized during my PhD. I will identify key strengths and weaknesses in methodology and approaches applied in this work, which will be of use in future considerations and interest in these topics.

2. Chromosome substitution lines as a genetic mapping tool in quantitative trait analysis.

2.1 QTL mapping in CSL-populations in *Arabidopsis thaliana* and Cucumber (*Cucumis sativus*)

Complete populations of chromosome substitution lines (CSLs), i.e. CSL-populations in which genotypes are present that contain all combinations of substituted chromosomes, are uniquely characterized by the

lack of crossovers across the genome. Wijnen (2019) discusses a variety of possibilities and opportunities that can be performed using populations of CSLs, but in this work an emphasis was put specifically on their use in the genetic mapping of complex traits. This is particularly fuelled by a shortage of effective methodology that allows for the proper detection of gene-gene interactions (epistasis) (Cordell, 2000; Mackay, 2014). The use of chromosome substitution lines is thought of as potent to unravel epistatic interations and low-effect genetic effects due to the balanced genetic architecture that can be found in a complete population (Wijnen & Keurentjes, 2014). Furthermore, promising reports involving the use of CSLs in rat and mouse exist, albeit using populations with only single chromosome substitutions, are found to be more powerful genetic mapping populations for both additive and epistatic genetic components compared to their segregating alternatives (Nadeau *et al*., 2000; Spiezio *et al*., 2012, Buchner & Nadeau, 2015). Building further on previous experiences by Wijnen (2019), and in line with the goal of exploring CSLs as a genetic mapping tool, I performed many experiments to evaluate the use of CSLs as a genetic mapping tool in plants.

My first experiments involved phenotyping of the Col-0 x Ler-0 CSL-population developed by Cris Wijnen (Wijnen *et al*., 2018), for a variety of photosynthesis traits. In line with my predecessor, I, too, performed experiments that do allow for a direct comparison of genetic mapping populations, in this case recombinant inbred lines (RILs). These populations were grown and phenotyped for photosystem II efficiency (ΦPSII) and projected leaf area (PLA) following methodology as described in **Chapter 3.** Between the RILs and CSLs, no major differences were found in terms of the number of QTLs found, QTL effect sizes, heritabilities and phenotypic variation within the populations (**Box 1**). The main contributors for both ΦPSII and PLA, *ΦPSII_c5* and *ERECTA* (**Chapter 3**), are present in all conditions and in both populations. The strength of the QTLs differed to some degree between the populations, but not significantly. The CSLs offered no evidence for epistasis to contribute to ΦPSII and PLA (**Box 1**).

The other CSL-population is composed of Arabidopsis accessions Col-0 and MIB-22, and was developed to near completion by Ramon Botet Vaca. This population has the advantage over the Col x Ler population of having no segregating *ERECTA* locus that may pleiotropically highly impact plant development (van Zanten *et al*., 2008; Keurentjes *et al*., 2007). I completed and used this population in **Chapter 5** to explore the genetic architecture of physiological adaptation to fluctuating light conditions (Vialet-Chabrand *et al*., 2017; Kaiser *et al*., 2020). Here, I did find more signs of epistasis playing a role in the genetic architecture of fluctuating light tolerance of productivity, although most of these interactions show only low degrees of confidence – even if multiple pair-wise comparison correction is not applied. The genetic simplicity of a CSL-population allows a more straightforward analysis of the genotypic data compared to traditionally segregating populations (Wijnen, 2019). This is particularly useful to describe longer term trends in context of the whole population and is helpful to characterize such (**Chapter 5**). Furthermore, I used the Col-0 x MIB-22 population to map flowering time (**Chapter 6**), but again no epistatic interactions were discovered that constitute this trait. A particularly noteworthy feature of the Col-0 x MIB-22 flowering time experiment is that only 96 plants had to be used during the entire experiment in order to map genetic components contributing to the trait. This figure is much smaller than is expected from segregating populations required to achieve similar outcomes (Brachi *et al*., 2010; Keurentjes *et al*., 2007), which makes the CSL-population cheap to use for genetic mapping purposes.

BOX 1: A comparison of trait trait variation and QTL detection power between CSLs and RILs

In all conditions as described in **Chapter 3,** I grew the CSL-population developed from the same parental accessions (Col-0 and Ler-0, Wijnen 2019) as the RIL. This allows a comparison of mapping power based on the genetic structures of these populations alone. Projected leaf area (PLA), which functions as a parameter for plant productivity, and the quantum yield of photosystem II (ΦPSII), as a parameter for photosynthetic activity of plants, were measured for all lines as described in **Chapter 3.** In these experiments the CSLs were grown alongside the RILs in 8-12 replicates per genotype, after which population statistics (**Table 1),** the detected chromosomal QTLs (**Table 2**) and the normalized, comparative effect sizes of the populations (**Table 3**) were recorded. As in the RILs (**Chapter 3**), no significant influence of epistatic interactions could be detected in the CSLs. Most of the population statistics are very similar between the populations (**Table 1-3**). The most striking difference, however, is that the RILs maps ΦPSII_c5 to a 30+ times higher resolution than the full length chromosomes of the CSL (**Figure 1**), although 50% more plants were grown in the RIL-population than in the CSL-population.

Table 1. *Population statistics in the RILs and CSLs for the PLA and ΦPSII traits. Mean = population average, StDev = standard deviation, CoV = coefficient of variation, Min = minimum, Max = maximum value, H^2 = broadsense heritability. Experiment and treatment correspond with those described in* ***Chapter 3.*** *All values are calculated from PLA and PSII measurements taken at one hour after the onset of light, 21 days after sowing. Treatments; LL = 100 µmol $m^{-2}s^{-1}$, C = 200 µmol $m^{-2}s^{-1}$, -N = 200 µmol $m^{-2}s^{-1}$ but with a solution containing only 10% of nitrogen content compared to the reference.*

Population	Trait	Exp	Treatment	Mean	StdDev	CoV	Min	Max	H^2
RILs	PLA	I	LL	657.0	157.6	0.240	387.7	1251.0	0.239
		II	C	637.6	160.8	0.252	349.1	1072.8	0.268
		II	-N	557.6	137.2	0.246	312.3	1003.4	0.252
	ΦPSII	I	LL	0.680	0.011	0.016	0.610	0.696	0.153
		II	N-	0.723	0.008	0.011	0.695	0.740	0.093
		II	C	0.734	0.008	0.010	0.720	0.751	0.137
CSLs	PLA	I	LL	700.2	171.1	0.244	428.9	1083.4	0.263
		II	C	469.3	116.4	0.248	256.3	706.5	0.301
		II	N-	398.6	112.2	0.281	224.2	709.5	0.342
	ΦPSII	I	LL	0.678	0.007	0.011	0.661	0.687	0.159
		II	N-	0.718	0.007	0.010	0.702	0.732	0.137
		II	C	0.727	0.008	0.011	0.712	0.739	0.083

Table 2. *Outcomes from analysis of variance for additive effects for each of the chromosomes (5) in Arabidopsis, following methodology of* ***Chapter 5****. The additive effect outcome from one-way ANOVA for each chromosome is shown in each column. Highly significant effects (p <0.01) are highlighted in dark grey, weakly significant effects (0.05 > p > 0.01) are given in light grey.*

Chr(s)	PLA_LL	PLA_C	PLA_-N	PSII_LL	PSII_C	PSII_-N
c1	0.78908	0.73321	0.92932	0.16253	0.3986	0.45914
c2	0.00001	0.00071	0.00099	0.22124	0.04311	0.09426
c3	0.10860	0.04827	0.00241	0.50056	0.04949	0.21049
c4	0.96516	0.03929	0.04610	0.09466	0.00358	0.16920
c5	0.07721	0.89932	0.74833	0.00060	0.01123	0.00047

Table 3. *Effect size analysis of RIL vs CSLs. For the populations and under each condition specifically, the mean value for the relevant allele is given. For PLA, values are given for Chromosome 2 (CSLs) or the corresponding peak marker in the RIL (M238). For PSII-efficiency, phenotypic values are given for Chromosome 5 (CSLs) or the corresponding peak marker (M563) in the RILs. Relative effect size gives the normalized effect size of the QTLs in either population.*

CSLs	PLA-C2			ΦPSII-efficiency-C5		
	LL	C	-N	LL	C	-N
Col	818.8	537.8	469.1	0.674	0.722	0.713
Ler	585.6	399.8	340.1	0.682	0.73	0.721
Abs(ΔCol-Ler)	233.1	138	129	0.00798	0.00769	0.00844
Relative effect size	**0.2847**	**0.2567**	**0.275**	**0.0118**	**0.0106**	**0.0118**

RILs	PLA-M238 (*Erecta*)			ΦPSII-efficiency-M563 (*ΦPSII c5*)		
	LL	C	-N	LL	C	-N
Col	816.3	795.8	680.2	0.673	0.728	0.718
Ler	574.9	556.2	494.5	0.684	0.738	0.726
Abs(ΔCol-Ler)	241.4	239.6	185.7	0.01068	0.01029	0.00779
Relative effect size	**0.2957**	**0.3011**	**0.273**	**0.0156**	**0.0139**	**0.0107**

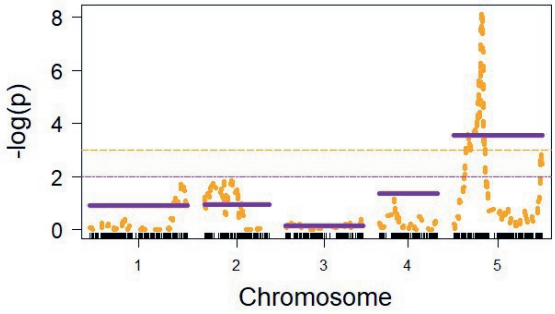


Figure 1. *Comparative Manhattan plot of the RILs (orange) and CSLs (purple). The result is shown for the All values are calculated from PLA and PSII measurements taken at one hour after the onset of light, 21 days after sowing for the 100 µmol $m^{-2}s^{-1}$ treatment. The dotted horizontal lines represent the LOD-threshold to claim QTLs for the CSL (orange, 2) and the RIL (purple, 3).*

Next to *Arabidopsis thaliana*, further experimentation of CSLs in a panel of single substitution lines in cucumber (*Cucumis sativus*) was performed by and with the collaborating partner in this project; Rijk Zwaan (**Box 2**). This cucumber CSL-population is derived from two contrasting parental lines, L and K, via advanced backcrossing schemes rather than by applying a transgenic method to suppress meiotic recombination (as described in Wijnker *et al*., 2012; Calvo-Baltanas *et al*., 2020). During the course of my PhD, these lines were grown in a high-throughput photosynthesis phenotyping system to obtain data assessing photosynthesis efficiency (Flood *et al*., 2016, **Chapter 3, 4**). In addition, a growth analysis was performed on plants grown in greenhouses, of which some of the results are shown in **Box 2**. The number of QTLs found in these cucumber CSLs is comparable to those typically found in segregating populations for similar traits that are scored in the same developmental stage (Wang *et al*., 2016; Liu *et al*., 2021). An obvious advantage of the CSLs in these experiments is the smaller number of lines that is required to perform genetic mapping in segregating populations – particularly when only small panels are grown.

Taken together, these experiments provide valuable experience for genetic mapping in CSLs with regards to their position in the genetic mapping population type taxonomy (Wijnen & Keurentjes, 2014; Bazakos *et al*., 2017, Tiang *et al.*, 2021). The total number of experiments performed allows for a more accurate generalization of outcomes, particularly following the use of CSLs generated from different parental origins and species. The total number of experiments described is substantial given the limited use of CSLs in scientific literature to this point (De Léon *et al*., 2011, Kuspira *et al*., 1956 (Wheat); Koumproglou *et al*., 2000; Lardon *et al*., 2020; Wijnen 2019 (*Arabidopsis thaliana*); Wu *et al*., 2010, Fuller *et al*., 2021 (Cotton). Having described the major insights gained from the experiments described in this work and considering the aforementioned studies, in the next section I will elaborate on general patterns and insights with regards to the use of chromosome substitution lines in quantitative trait analysis.

2.2 CSL-populations as an instrument in the quantitative geneticist' toolbox

In the previous section, I outlined several characteristics that can be considered potentially advantageous and disadvantageous while using CSLs as a mapping tool. In literature, three key advantages of the use of CSLs – although most has derived from use in animals applying only single substitution line populations. These include 1) the relatively small number of plants that are required to perform genetic mapping (Nadeau *et al*., 2000; Koumproglou *et al*., 2002), 2) elevated QTL detection mapping power (Spiezio *et al*., 2012; Buchner & Nadeau, 2015) and 3) improved ability of being able to map epistatic interactions (Nadeau *et al*., 2012; Wijnen, 2019).

In plants, genetic mapping in conventionally segregating mapping populations will easily require several times more total plants to achieve similar outcomes compared to CSLs (**Chapter 6**, Ungerer *et al*., 2002; Keurentjes *et al*., 2007; Brachi *et al*., 2010), which highlights the first key advantage as phenotyping can be very cost-efficient. The experiments in cucumber CSLs are noteworthy in this respect, as this advantage is particularly relevant in high-technological greenhouse crops as the total number of lines which are to be grown is smaller and each plant is costly to grow until their reproductive or phenotypically relevant phases. This factor may makes these lines highly efficient for purposes of genetic mapping, particularly in repeated experiments and cross-environmental evaluations. Long-term use of CSLs developed from two relevant genotypes may indeed prove to be cost-efficient in this way. However, for a complete population of CSLs developed from a plant species with high chromosome counts this advantage is compromised due to the large number of lines that can be made from such species (**Chapter 6**). In cucumber, a species with a low number of chromosomes (2n = 14), this would already require a total of 128 lines to develop a full population and thus increases developmental costs considerably (**Chapter 6**).

BOX 2: The use of chromosome substitution lines for quantitative trait analysis in cucumber

(Experimental design, performance and data contributed by Dr. Maaike Wubs)

Single chromosome substitution lines (**Figure 1**) have been developed by the vegetable breeding company Rijk Zwaan in cucumber (*Cucumis sativus*) prior the start of this project, following complex crossing schemes. These lines were phenotyped for a variety of traits, including growth parameters in a greenhouse (**Figure 2**), but also in the Phenovator (a high throughput phenotyping facility for photosynthesis traits, see Flood *et al*., 2016) for ϕPSII under 500 µm m-2 s-1 (**Figure 3**) and low light 100 µm m-2 s-1 (**Figure 4**). All phenotypes in these experiments were obtained from smaller plants of up to three weeks old.

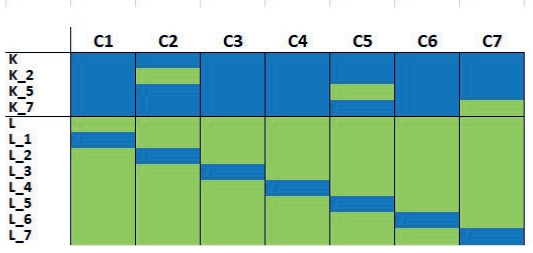


Figure 1. *The genotypic layout of the lines. Chromosomes belonging to line K (first parental line) belong in coloured blue, while chromosomes belonging to line L (second parental line) are coloured in green. C1-C7 represent the individual chromosomes of cucumber. Lines of K with L substitutions were not available for all chromosomes yet, although lines of L with K substitutions represent a completed panel of single chromosome substitution lines.*

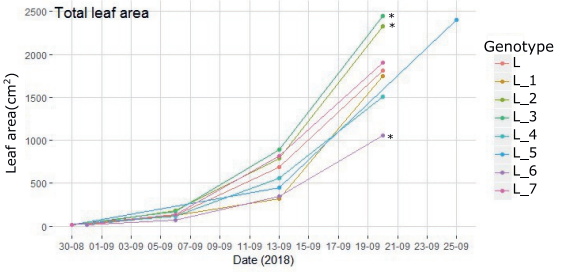


Figure 2. *Growth curve analysis for total leaf area in seedlings. Note that CSL20 was grown for a week more. Lines were analyzed by t-test and compared to the isogenic parental line*; * = $p < 0.05$. *32 plants were grown per genotype and eight replicates were harvested for each time point to calculate the total leaf area.*

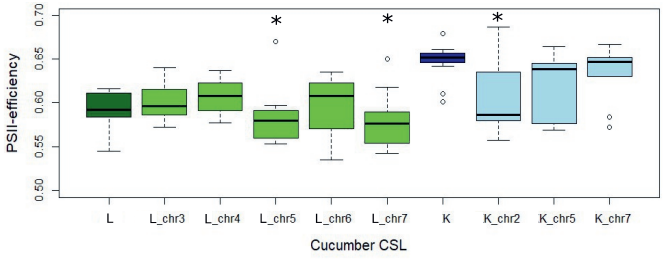


Figure 3. *ϕPSII values for cucumber seedlings grown after 12 days under 500 µm m-2 s-1 irradiance. All substitution lines are compared against the isogenic parent using student's t-test.* * = $p < 0.05$. *The reciprocal populations are coloured green or blue, while the parental lines are at a darker shade. n = 9–12 per genotype.*

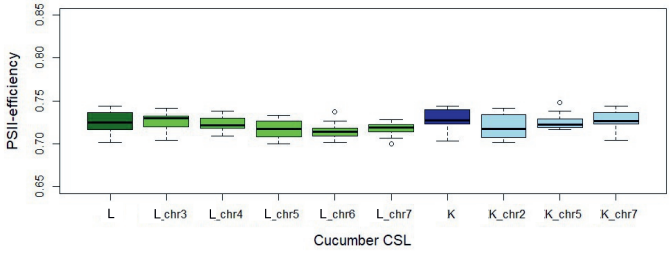


Figure 4. *ϕPSII value for cucumber seedlings one day after a stepwise reduction of growth irradiance to 100 µm m-2 s-1 . All substitution lines are compared against their parent using student's t-test. No significant differences were found. The reciprocal populations are coloured green or blue, while the parental lines are at a darker shade. n = 9–12 per genotype.*

However, CSL-populations also comes with several properties that are disadvantageous in genetic mapping compared to alternative populations. QTL can only be resolved to the chromosome level, which means that fine-mapping of those to identify genetic variation in candidate genes will require at least one extra experiment compared to QTL found in segregating populations. This may particularly be problematic in crops with long breeding cycles or that are expensive to grow per unit. To compensate such a burden, there must be significant benefits associated with genetic mapping in CSLs compared to alternative genetic mapping population types. If CSLs in plants indeed have improved QTL detection power levels, as is reported for animals (Bucher & Nadeau, 2015), then that advantage could potentially outweigh the disadvantage of resolution as discovering otherwise undetectable or unexplored genetic variation is highly valuable. However, in my experiments I conclude no significant advantages in the use of CSLs over RILs developed from the same parental lines (**Box 1**) or high power to detect additive effects and epistatic interactions (**Chapter 5, 6, Box 2**). Similarly, Lardon *et al.* (2020) also only found single QTL for shoot regeneration in the full Col-0 x Ler-0 CSL-population which indicates that such populations may not be particularly powerful to find additional loci. These outcomes conflict with conclusions drawn in Wijnen (2019), who report increased numbers of detection of additive and epistatic effects – although in that study only a limited number of traits were analyzed to come to that conclusion. In this thesis, more genotypic variation, species and a wider variety of traits are used, which should make purported advantages with regards to genetic mapping power apparent. I conclude that advantages of high mapping power in CSL-populations do not hold for plants, which means that the 2nd and 3rd key advantages as outlined at the start of this paragraph are not experienced. The lack of finding epistatic interactions may have to do with the fact that I simply did not look at variation in traits are constituted by epistatic interactions in the parental lines. However, this shapes a major secondary problem, as it is impossible to predict whether a trait is controlled by a major genetic component of driven by epistasis based on the phenotype of the hybrid or parental lines of interest. Some authors argue that epistasis may not, however, be an important contributor to genetic variation in many traits (Hill *et al.,* 2008), although this may result from a lack of robust methods to detect such in genetic mapping. The study of epistasis may then be aided more by the development of novel algorithms, genetic mapping strategies and awareness on epistasis in conventional mapping populations (e.g. Huang *et al*., 2012; Lachowiec *et al*., 2015; Ning *et al*., 2018), rather than the development of populations specifically designed to detect such interactions without prior knowledge of their presence.

2.3 Use and strategic niches of CSL-populations

In *Arabidopsis thaliana*, I did not find any signs of a prominent epistatic interaction or higher detection power of QTLs that were not also present in the RIL-population developed from the same parents (**Box 1**). One goal of this project was to engage in the functional characterization of genes, and I adapted and focussed on traditionally segregating genetic mapping population types like RILs (**Chapter 3**) and F2 (**Chapter 4**) populations. Although such populations require more plants to grow to detect QTLs, they much more efficient for purposes of mapping genes (**Box 1**). Another potential disadvantage of CSLs

compared to other genetic mapping population types that so far has received little attention, but came to light in this study, is the low chromosome numbers in Arabidopsis in full CSL-populations. Arabidopsis' five chromosomes are ideal for the development of a complete CSL-population due to the limited number of lines (32) that are required to develop a complete CSL-population. However, this characteristic may also compromise the genetic mapping power, as I demonstrated by statistical modelling in **Chapter 6** and experienced during the experiments performed in **Chapter 5**. This most likely results from the low number of replications in the specific genotypic classes, sixteen for single chromosome effects, eight for epistatic effects, etc.. A strategy of adding additional segregants to refine and empower the mapping – a strategy analyzed in detail by Keurentjes *et al*. (2007) – cannot be used here given the finite nature of such a population. Furthermore, increasing the number of line replicates will only support moderate improved mapping power (Keurentjes *et al*., 2007). By contrast, CSL-populations developed from species with larger chromosome counts will have a much higher genetic mapping power, even if not all possible lines have been developed (**Chapter 6**) and may thus be more ideal for genetic mapping purposes.

Popularization of the use of CSLs in rodents originates from a demand of an immortal mapping population that can efficiently map QTL, without the disadvantages that segregating populations have in terms of the number of subjects to phenotype, low statistical power in such populations, the influence of background QTL and the development time of proper fine-mapping populations for gene discovery (Nadeau *et al*., 2012, Buchner & Nadeau, 2015). In plant genetics the competition for mapping population types is much more severe (Wijnen & Keurentjes, 2014; Bazakos *et al*., 2017) and genetic loci are detected at high effectivity in plants due to the ability of replication of genotypes (Brachi *et al*. 2011; Bazakos *et al*., 2017; Liang *et al*., 2021). As the purported increased power of detecting QTLs does not hold against available alternatives, the developmental costs associated with CSLs balance unfavourably against the gains in genetic mapping strategies. Additionally, populations composed of chromosome segment substitution lines, also called near isogenic lines (Keurentjes *et al*., 2007; Kooke *et al*., 2012), that allow more accurate mapping, outcompete plant populations with full chromosome substitutions as first proposed by Koumproglou *et al*. (2000) as is reflected in the sparse use of CSLs in plant scientific literature.

From a strategic perspective, I conclude that the development of full and complete populations of CSLs for purposes of genetic mapping comes with opportunity costs that are too high to justify. The current technological state-of-art for development of such populations is not efficient enough, and there is a lack of consistent evidence for improved genetic mapping power in plants compared to mapping populations that are easier to develop. The focus of using CSLs in this study had been on quantitative trait mapping and I did not explore alternative genetic strategies that are enabled by their use. Many more ideas exist to effectively utilize populations that segregate for full chromosomes, such as the genetic mapping of heterosis, the assessment of breeding diallels, cost-effective, "-omics" experiments, the effect of maternal inheritance of particular chromosomes and so on. These approaches are all more difficult, if not impossible, to properly realize in segregating populations. As these ideas have been developed and thoroughly discussed by Wijnen (2019), I will thus not further elaborate those here.

An alternative strategy to effectively use CSLs in a genetic mapping population is one that can be deduced from the outcomes in my modelling study in **Chapter 6.** The conclusions of this study imply that

genetic mapping in CSLs and reverse breeding are most cost-effective if considered complementary. If future biotechnological developments would make reverse breeding an effective strategy, only a modest increase in efforts over the ability of exactly reproducing the hybrid of interest, incomplete CSL-populations can be developed that would permit the genetic mapping of additive and epistatic traits in populations with higher chromosome levels. The costs of such incomplete CSL-populations would only be a fraction of the costs that involves the development of a complete CSL-population, while maintaining sufficient mapping power.

To summarize and conclude, based on described studies in this work and elsewhere (Koumrpoglou *et al*., 2000; Wijnen, 2019; Lardon *et al*., 2020) and in context of studying quantitative traits, the development of complete CSL-populations for purposes of improved mapping power involves very high opportunity costs compared to other genetic mapping population types. An integrated and larger scale reverse breeding & mapping strategy may, however, serve as a basis for a highly effective breeding strategy by maximizing genetic mapping efficiency against reduced developmental costs. The feasibility of such a strategy would, however, still depend on future advanes in biotechnological solutions to suppress meiotic recombination.

3. Thinking outside the chloroplast: Studying natural genetic variants shines new light on functioning and evolution of photosynthesis

3.1 Bi-parental mapping populations broaden the scope to search for "photosynthesis" genes

The importance of chloroplast locating proteins to affect photosynthesis cannot be overstated as it is trivial that this process takes place in the chloroplast. Genetic mapping studies that involve the elucidation of the genetic architecture of photosynthesis traits generally tend to prioritize candidate genes that encode chloroplast transit peptides in the translated protein (ctp) (van Rooijen *et al*., 2015; 2017; Rungrat *et al*., 2019; Prinzenberg *et al*., 2020). Ctps are specialized oligopeptides that enable the translocation of a protein from the cytoplasm into the chloroplast (Bruce, 2000; Eseverri *et al*., 2020), thus signifying their use as a biological marker for "photosynthesis genes". In my genetic mapping studies I, too, found and initially hypothesized genes with this characteristic to underlie QTLs, including promising, well-studied genes such as *SIGMA-FACTOR 5* (At5g24120) (**Chapter 3**) and *LIPOXYGENASE 2* (At3g45140) (**Chapter 4**). However, none of such genes gave a profound photosynthesis. These outcomes indicate that factors influencing photosynthesis in plants go beyond the chloroplast.

Due to the very few – but strong – quantitative trait loci for ΦPSII found in the bi-parental mapping populations (**Chapter 3 & 4**), compared to the typically many – but weaker – QTLs found in genome wide association analysis (van Rooijen *et al*., 2015; 2017; Rungrat *et al*., 2019; Prinzenberg *et al*., 2020), I was able to more carefully scrutinize and fine-map QTLs of interest and consider all genes. Rather than relying on predictive tools such as gene ontological annotation (Ashburner, 2000; Huang *et al*., 2009; van Rooijen *et al*., 2015) or gene transcriptome assays (Klepikova *et al*., 2016), continuous fine mapping efforts until the regions were delineated by a number of genes small enough to conduct functional

genomics experiments is a more objective way of pointing gene targets. GWAS-studies often rely on knock-out mutants to research genes (O'Malley *et al*., 2015; Tang *et al*., 2018; van Rooijen *et al*., 2017), which results from the phenomenon that often many alleles are present underlying a single locus in highly diverse populations, and identifying the alleles inducing phenotypic changes then beomces a very laborious process (Forsberg *et al*., 2015; Tang *et al*., 2018). The absence of a TDNA line with a phenotype is a likely hypothesis which prevented Tang *et al*. (2018) from investigating the stronger of two QTLs for xylem pressure in a diverse set of Arabidopsis accessions in their study. In a suitable diversity panel of Arabidopsis, wherein a proper number of accessions carry alleles for the Q3 locus, could thus also lead to the discovery of *TFIID* due to the availability of a mutant line (**Chapter 4) –** assuming that this gene is indeed the causal gene. However**,** for the *SQE-likes* in **Chapter 3** I expect that being able to discern the causal genetic mechanism that underlies this QTL is much more complicated, if not impossible, within reasonable limits of resource commitment. In addition, in a diverse collection of Arabidopsis accessions, I expect the number and combinatios of variants for *SQE5*, *7* and *6* to be numerous given that genomic regions including copy number variants contain high genetic variation (Bilgin Sonay *et al.*, 2015; Hastings *et al*., 2019; Lye ZN, Purugganan, 2019), Arabidopsis is already particularly highly diverse species (Durvusalu *et al*., 2017) and there are no T-DNA mutant lines giving a phenotype. Although more laborious to delineate the genomic region of interest in finemapping, this study demonstrates that the identification power for causal genes that are not obvious to consider following from gene ontology or gene knockout mutants phenotypes is high in bi-parental mapping populations – particularly if high quality genome sequences are available. Thus, my work in **Chapter 3** & **4** confirms several of the advantages outlined in my introduction that bi-parental mapping populations offer to unravel genetic variation.

3.2 *SQUALENE EPOXIDASEs* and *TANSCRIPTION FACTOR SUBUNIT II D* reveal novel areas of plant biology that impact photosynthesis

In **Chapter 4** I hypothesize a possible role for *TRANSCRIPTION FACTOR SUBUNIT II D (TFIID)* gene to affect photosynthesis efficieny in response to long-term high light, underlying the Q3 QTL. This gene is located in the nuclear genome, but the protein translocates to the mitochondria. Although this might seem confusing as the trait of interest is in photosynthesis, thus expecting involvement in chloroplasts, mitochondria in plants are highly relevant organelles to maintain biochemical balances in the cell during photosynthesis (Gardeström & Lernmark, 1995; Igamberdiev *et al*., 2006; Gardeström & Igamberdiev, 2016). Under abiotic stresses – i.e., unfavourable conditions in which plant metabolic balances are disturbed which can lead to tissue damage – plant mitochondria are particularly relevant to maintain the cellular redox balance and scavenging or ROS (Möller, 2001; Pastore *et al*., 2007; Choudhury *et al*., 2017). A definitive proof of evidence would require a molecular cloning strategy as described in **Chapter 3**, as a mutant line with a phenotype and genetic variation alone at this point only imply – but yet not prove – the causality of the gene underlying the Q3 QTL.

In **Chapter 3**, the causality of copy number variation in a previously uncharacterized sub-family of *SQUALENE EPOXIDASE* (*SQE*) gene family underlying a QTL for ϕPSII is supported by solid evidence. This finding is probably the most surprising and fundamental discovery contributing towards the functioning of photosynthesis in my whole dissertation. In *Arabidopsis thaliana*, a total of seven *SQE* genes are present, but only the dispersed members *SQE1, SQE2* and *SQE3* – sub-family I; "true *SQEs*" – are known to epoxidize squalene as the name suggests (Rasbery *et al*., 2007; Laranjeira *et al*., 2015). *SQE4, SQE5*, *SQE7* and *SQE6* – sub-family II; "*SQE*-likes" –are located in a tandem-repeat do not epoxidize squalene (Rasbery *et al*., 2007) and have no further described function. However, the proteins locate outside of the chloroplast and as of yet I haven't been able to come up with a good working hypothesized role or mechanism for *SQE* in processes related to photosynthesis. Another enigmatic feature is that *SQE*-like genes are solely present in most Brassicacaea branch of plant evolution (Laranjeira *et al*., 2015), signifying the diversity of the photosynthesis machinery among plant species.

The products of both *TFIID* and *SQE-likes*, and the biological processes these genes are involved in, locate outside of the chloroplast and would therefore likely fall outside the scope of photosynthesis physiologists to consider for further investigations (Zoschke & Bock, 2018). These genes certainly do not belong to the, often cited, number of ~3000 genes identified in the *Arabidopsis thaliana* genome that have their translated protein product locate to the chloroplast (Reumann *et al*., 2005; Martin & Sabatar, 2010; Fristedt *et al*., 2017). The relationship between of *SQE*-likes and *TFIID* and ϕPSII may serve as a starting point to broaden the scope of potentially contributing factors to improve photosnythesis traits. More importantly, these discoveries may serve as a reminder that photosynthesis traits in plants can be found in many aspects of plant physiology or biochemistry as photosynthesis, not just inside the chloroplast, which should not be overlooked.

4. From high throughput image phenotyping to improved photosynthesis and increased genetic yield potential.

4.1 Screening and improving photosynthesis traits to predictor of yield

Studying natural variation in photosynthesis is primarily motivated by the prospect of eventually improving the performance of crops (Driever *et al*., 2014; Flood *et al*., 2011; Acevedo-Siaca *et al*., 2020; Faralli & Lawson, 2020; Long *et al*., 2020;). Here, performance is usually expressed in terms of (speed of) growth, although potential benefits in tissue quality and abiotic stress tolerance traits should be recognised as well. Chlorophyll fluorescence imaging, and similar technology such as hyperspectral- and thermal imaging, are proposed and identified as the go to methodology to measure *in vivo* photosynthesis repeatedly (Baker, 2008; Murchie & Lawson, 2013; Tschiersch *et al*., 2017) and high-throughput (Flood *et al*., 2011; Rungrat *et al*., 2016; Murchie *et al*., 2018; Faralli & Lawson, 2019) for large numbers of plants in genetic studies. The advantages of repeated and fast high-throughput chlorophyll imaging technology over traditional ways of measuring photosynthesis efficiency are highly valued for their speed, flexibility and robustness in plant genetic analysis (van Rooijen *et al*., 2015; van Bezouw *et al*., 2019; Rungrat *et al*., 2019). However, so far

no reports have established the connection between discovered natural genetic variation in photosynthesis traits derived from imaging technology and the added value of selection for those traits to increase crop performance.

In **Chapter 4**, I found that the combined effect of the Q3 and Q5 QTLs for high light acclimation in Arabidopsis accessions introduces a differential effect size of up to 25% higher ΦPSII when comparing lines carrying the opposite alleles. This combined effect size is an order of magnitude larger than any previously reported for a photosynthesis trait (van Rooijen *et al*., 2015, 2017; Prinzenberg *et al*., 2020; Flood *et al*., 2020) in the high-throughput photosynthesis phenotyping platform at the Wageningen University (Flood *et al*., 2016). This combined effect size meets photosynthetic gains as reported in transgenic studies (Driever *et al*., 2017; South *et al*., 2020), who report significant gains in biomass accumulation. The identification of Q3 and Q5 thus challenges the idea that photosynthesis is a polygenic trait with many low effect size QTLs (Ort *et al*., 2015; Flood*,* 2015; van Bezouw *et al*., 2019). Thus, these QTLs motivated me to explore a potential relationship between the larger effect size QTLs for ϕPSII and their potential to contribute to biomass accumulation through optimized photosynthesis. As these QTLs are active in the response to a change in high light, I expected that the cumulative effect of repeated exposure to high light perturbations during growth should affect biomass accumulation (as previously shown in Vialet-Chabrand *et al*., 2017; Schneider *et al*., 2019). I also expected that the faster photosynthesis-acclimatizing plants would also suffer less from impairments in productivity. However, the selected Arabidopsis lines with the combination of superior acclimation alleles grew smaller, rather than bigger, under these conditions. There are many factors as to why the outcomes of this experiment did not meet my expectations, including background segregation of alleles and acclimatory processes that we could not measure in the different growth conditions. However, a 25% difference of efficiency in a photosynthesis trait cannot be ignored or under-estimated and suggests that optimization of photosynthesis may not be as straightforward to improve as the high effect size of the Q3 and Q5 seems to suggest.

The quantum yield of photosystem II (ΦPSII) is one of the more intuitive parameters that arises from chlorophyll fluorescence imaging. It describes the fraction of incoming light irradiance that is being captured by photosystem II reaction centres and is subsequently used into photochemistry – and thus carbon assimilation (Baker, 2008). However, carbon assimilation is also highly dependent on the carboxylation rate of Rubisco – a key rate-limiting enzyme of the Calvin Benson cycle. Due to the physiological and chemical alterations, C4 species are able to carboxylize a much higher fraction of incoming CO_2. In turn, this results in a high correlation between ΦPSII and ΦCO2 (carbon assimilation efficiency) (Genty *et al*., 1989; Edwards & Baker, 1993). The relationship between ΦPSII and carbon assimilation in C3 species is more diffuse, particularly due to the role that photorespiration plays in diverting electron transport and if plants experience misbalances in resources (Baker, 2008; Murchie & Lawson, 2013). However, measures of carbon assimilation through gas-exchange analyzers – devices which enable direct quantification of ΦCO2 in plants – may also not necessarily lead to satisfying correlations with biomass accumulation either (Murchie *et al*., 2002; Janh *et al*., 2011; Driever *et al*., 2014). If the assessments of photosynthesis efficiency may not always show an obvious relationship with biomass accumulation, then the goal of enhancing yield through improved photosynthesis requires a more

systematic approach. Careful scrutiny of related traits may will then eventually point out key biological processes that may eventually point to such a relationship (van Bezouw *et al*., 2019; Zhu *et al*., 2020; Theeuwen *et al*., 2022).

4.2 Understanding functioning, opportunities and limitations of photosynthesis for improved crop yield

Flood (2015) and van Rooijen (2016) describe one of the first elaborate undertakings of using ϕPSII as a trait to assess larger populations for their photosynthetic activity and under different environments, but all performed in the same experimental setup (Flood *et al*., 2016). In Flood (2015), the general effect sizes of QTLs found in his study – which ran as low as a 0.25% difference between the strong and weak allele – made him wonder about reasons why to even dissect such QTLs – and even whether the whole exercise would be worth it. In **Chapter 3**, I dissected yet another such small effect size photosynthesis QTL (1.5%), which on itself appears as yet another one of those detected (Flood, 2015). By dissecting this QTL in *Arabidopsis thaliana*, we found that activity of *SQE-like* was the sole major component of differences in ϕPSII between the Col-0 and Ler-0 accessions. Findings like this may contribute to our understanding of which processes in a plant cell do contribute towards changes in observed ΦPSII – thus understanding what the trait may potentially imply from a plant physiological point of view. In the case of my work in *SQE-likes*, the signal that was picked up through assessment of ΦPSII is likely to represent a change in a rate-limiting catalytic process in a known or unknown metabolic pathway in *Arabidopsis*. If so, then observed changes in ΦPSII between genotypes should be treated with more caution, as the number of plant processes indirectly responsible for changes in this parameter is potentially big, but creatively, as this would imply that the applications of chlorophyll fluorescence imaging (CFI) may expand further. The exact meaning of differences in CFI parameters resulting from genotypic differences must be understood thoroughly to suggest biological mechanisms attributed to those changes.

In **Chapter 5**, I decided to eschew *in vivo* chlorophyll fluorescence imaging and try a different approach to understand growth and physiology of plants under influence of different photosynthetically demanding growth conditions. Particularly challenging conditions for the adaptation of photosynthesis is plant adaptation to fluctuating light, a topic that saw an increase in interest due to resulting potential yield losses (Vialet-Chabrand *et al*., 2017; Kaiser *et al*., 2018; Ding *et al*., 2019; Schneider *et al*., 2019; Morales & Kaiser, 2020). Transgenic alteration of a plant's ability to adapt to dynamic light results also results in alterations of growth (Athanasiou *et al*., 2010; Kromdijk *et al*., 2016; Garcia-Molina & Leister, 2020). Despite this, fluctuating light has not yet established itself as an abiotic stress, alike salinity and drought. Capitalizing on a recent discovery of natural variation in the growth response of various *Arabidopsis thaliana* lines to fluctuating light (Kaiser *et al*., 2020), I was able to describe a reduced effect of a growth-stimulating QTL on chromosome 3 using a population of CSLs (**Chapter 5**). An intensive fine mapping approach may yield a light regime sensitive regulator for plant growth, while a thorough phenotypic, molecular and/or biochemical analysis of the F1-hybrid may result in a deeper understanding in the role of photosynthesis in hybrid vigour and its tolerance to fluctuating light conditions.

The high-throughput phenotyping facility the Phenovator (Flood *et al*., 2016) is an excellent tool to conduct time series analyses of photosynthesis and projected leaf area, yet to assess the photosynthetic response to naturally fluctuating light conditions it is still too slow (Murchie *et al*., 2018). Furthermore, many other traits other than the efficiency of photosystem II can be efficiently phenotyped high-throughput, and many researchers have demonstrated being able to do so (Rungrat *et al*., 2019). A future lies ahead in which the focus of research will not rely on a single photosynthesis trait, but many, to accurately evaluate specific traits (van Bezouw *et al*., 2019).

By evaluating various components of non-photochemical quenching – a plant protective mechanism against excess light irradiance that operates at differen levels in time (Ruban, 2017) – I was able to obtain a deeper insight in the plant's photosynthetic response to differing fluctuating light regimes in **Chapter 4**. Even though the direct translation of observed ΦPSII and productivity fluctuating environments in this chapter may have been contradictory, this chapter does show that selection for lines carrying the large effect size, high light QTLs Q3 and Q5 may improve growth and reduced photo-protective properties. This relationship was previously established by Kromdijk *et al*. (2016) by modifying photoprotection, although it could not be reproduced in *Arabidopsis thaliana* García-Molina & Leister (2020). Although in **Chapter 4** the assessed genotypic material is not optimal due remaining segregation and the relationship between the QTL was inverse to what was expected, at least the experiments demonstrated that selecting for natural variants of photosynthesis traits can potentially result in improved biomass accumulation in adverse light conditions as observed in the field.

5. Towards breeding directed at improving photosynthesis and yield

In my general introduction, I briefly touched upon the current state of the C4-rice project, an effort to incorporate the more effective C4-photosynthesis molecular pathway into rice to benefit photosynthesis efficiency and, consequently, production (Evans, 2000; Hibberd *et al*., 2008; von Caemmerer *et al*., 2012). In the past 15 years this project garnered much insight and attention to the wider public, but the direct transfer of this pathway in rice turned out to be more complicated than previously anticipated (Ermakova *et al*., 2020; 2021). Around the same time, in the late 2000s, a group in Japan embarked on a line of research that involved the evaluation of the impact of genomic introgressions related to photosynthesis on biomass accumulation of rice cultivars (Adachi *et al*., 2010; 2011). In their latest study, a relationship between biomass accumulation and late stage maintenance of carbon assimilation was described (Honda *et al*., 2021). Up to date, this work serves as one of the better available examples to benefit productivity in a staple crop species by using natural variation in photosynthesis. When the exact contributing loci are resolved, a rice breeding program can potentially target this trait and lift off, well ahead of the supposed "final breeding cycle before starvation" as stipulated by Kromdijk & Long (2016), even if multiple loci need to be bred simultaneously into the desired commercial cultivar. Although this rice example should not de-emhpasize the significance of transgenes to improve photosynthesis, it does show that natural variation in photosynthesis traits can also be effective to achieve similar results.

To have the exploitation of natural variation in photosynthesis successfully integrated in plant breeding programs, working examples are in great need to show the feasibility as well as the development of easily selectable traits. Selection for alleles of *SQE-likes* (**Chapter 3**) and *TFIID* (**Chapter 4**) will at this point unlikely serve as promising genes to consider for future improvement of crops – unless the exact mechanisms, and thus the benefits, are being resolved (Theeuwen et al., 2022). The association of these genes with photosynthesis may ultimately broaden the scope of scientist to consider more potential targets for improvement. To be able to make quick genetic gains to improve photosynthesis, genomic prediction is a method which may drastically increase the speed by which these are being realized in highly polygenic traits like photosynthesis (Meeuwissen *et al*., 2001; Bernardo, 2016; Millet *et al*., 2019). Rather than using and elucidating photosynthesis traits to the molecular level or as a target in genomic prediction experiments, phenomic prediction may serve as another approach to make use of photosynthesis traits in breeding. Imaging phenotypes from chlorophyll imaging may serve as a basis to assess plant health and metabolic activity, thereby using photosynthesis traits as predictive markers for productivity in later stages Grzybowski *et al*. (2021). However novel, methodology to robustly and reliably phenotype plants at a larger scale is required to yield the benefits, which may be aided and validated by linking causal genotypic variation to phenotypes to understand the nature of measurements. Photosynthesis traits as described in **Chapter 4** (ΦPSII for light acclimation) and **Chapter 5** (productivity tolerance to fluctuating light conditions) may be two of such, particularly because of the large variation and moderate to high heritability values observed in these experiments, but there may be many more that still will require exploration.

If there would be any doubt in the existence and potential of utility of natural variation in photosynthesis traits, this must be have faded away. Having established the existence of natural variation for photosynthesis, the most important bottleneck that currently exists is the question of how to translate natural variation in photosynthesis into more resilient and better yielding crops. In this thesis I laid out several different genetic and phenotypic strategies to do so, which may serve to inspire those who wish to study natural variation in photosynthesis and find applications in plant breeding.

References

Acevedo-Siaca LG, Coe R, Wang Y, Kromdijk J, Quick WP, Long SP (2020). Variation in photosynthetic induction between rice accessions and its potential for improving productivity. *New Phytologist*. 227: 1097-1108.

Adachi S, Tsuru Y, Nito N, Murata K, Yamamoto T, Ebitani T, Ookawa T, Hirawasa T (2011). Identification and characterization of genomic regions on chromosomes 4 and 8 that control the rate of photosynthesis in rice leaves. *Journal of Experimental Botany.* 62(6). 1927– 1938.

Adachi S, Baptista LZ, Sueyoshi T, Murata K, Yamamoto T, Ebitani T, Ookawa T, Hirawasa T. (2014). Introgression of two chromosome regions for leaf photosynthesis from an *indica* rice into the genetic background of a *japonica* rice. *Journal of Experimental Botany* 65(8), 2049– 2056.

Adachi S, Yoshikawa K, Yamanouchi U, Tanabata T, Sun J, Ookawa T, Yamamoto T, Sage RF, Hirasawa T, Yonemaru J (2017). Fine Mapping of *Carbon Assimilation Rate 8*, a Quantitative Trait Locus for Flag Leaf Nitrogen Content, Stomatal Conductance and Photosynthesis in Rice. *Frontiers of Plant Sciences*. 8:60.

Adachi S, Yamamoto T, Nakae T, Yamashita M, Uchida M, Karimata R, Ichihara N, Soda K, Ochiai T, Ao R, Otsuka C, Nakano R, Takai T, Ikka T, Kondo K, Ueda T, Ookawa T, Hirasawa T (2019). Genetic architecture of leaf photosynthesis in rice revealed by different types of reciprocal mapping populations. *Journal of Experimental Botany*. 70(19):5131-5144.

AGI (Arabidopsis Genome Initiative) (2000). Analysis of the genome sequencing of the flowering plant *Arabidopsis thaliana*. *Nature*. 408:796– 815.

Ali MA, Shah KH, Bohlmann H (2012). pMAA-Red: a new pPZP-derived vector for fast visual screening of transgenic Arabidopsis plants at the seed stage. *BMC Biotechnology*. 12:37.

Alonso-Blanco C, El-Assal SE, Coupland G, Koornneef M (1998). Analysis of natural allelic variation at flowering time loci in the Landsberg erecta and Cape Verde Islands ecotypes of Arabidopsis thaliana. *Genetics*. 149(2):749-64.

Andersson I. (2008). Catalysis and regulation in Rubisco. *Journal of Experimental Botany*. 59(7):1555-1568.

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Sciences.*19(1). 52– 61.

Araus JL, Sanchez-Bragado R, Vicente R (2021). Improving crop yield and resilience through optimization of photosynthesis: panacea or pipe dream? Journal of Experimental Botany. 72(11):3936-3955.

Arends D, Prins P, Jansen RC, Broman KW (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinformatics*. 26(23):2990-2992.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*. 25(1):25-9.

Ashrafi H, Kinkade M, Foolad MR (2009). A new genetic linkage map of tomato based on a Solanum lycopersicum x S. pimpinellifolium RIL population displaying locations of candidate pathogen response genes. *Genome*. 52(11):935-56.

Athanasiou K, Dyson BC, Webster RE, Johnson GN (2010). Dynamic acclimation of photosynthesis increases plant fitness in changing environments. Plant Physiology. 152(1):366-373.

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature.* 465:627– 631.

Awlia M, Nigro A, Fajkus J, Schmoeckel SM, Negrão S, Santelia D, Trtílek M, Tester M, Julkowska MM, Panzarová K (2016). High-throughput non-destructive phenotyping of traits that contribute to salinity tolerance in *Arabidopsis thaliana*. *Frontiers in Plant Science.* 7:1414.

Bailey S, Horton P, Walters RG (2004). Acclimation of *Arabidopsis thaliana* to the light environment: the relationship between photosynthetic function and chloroplast composition. *Planta.* 218(5): 793– 802.

Baker NR (2008). Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annual Review of Plant Biology*. 59:89-113.

Baker NR, Harbinson J, Kramer DM (2007). Determining the limitations and regulation of photosynthetic energy transduction in leaves. *Plant, Cell & Environment.* 30(9):1107– 1125.

Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Valé G, Cattivelli, L (2016). Next generation breeding. *Plant Science.* 242:3-13.

Barros MDC, Dyer, TA (1988). Atrazine resistance in the grass *Poa Annua* is due to a single base change in the chloroplast gene for the D1 protein of photosystem II. *Theoretical and Applied Genetics.* 75(4):610–616.

Basu U, Bajaj D, Sharma A, Malik N, Daware A, Narnoliya L, Thakro V, Upadhyaya HD, Kumar R, Tripathi S, Bharadwaj C, Tyagi AK, Parida SK (2019). Genetic dissection of photosynthetic efficiency traits for enhancing seed yield in chickpea. *Plant, Cell and Environment*. 42(1):158-173.

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bazakos C, Hanemian M, Trontin C, Jiménez-Gómez, Loudet O (2017). New Strategies and Tools in Quantitative Genetics: How to Go from the Phenotype to the Genotype. *Annual Review of Plant Biology*. 68:435-455.

Belknap JK (2003). Chromosome substitution strains: some quantitative considerations for genome scans and fine mapping. *Mammalian Genome*. 14:723–732.

Bell E, Creelman RA, Mullet JE (1995). A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in Arabidopsis. *Proceedings of the Natational Acadamy of Sciences of the United States of America*. 92(19):8675-8679.

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015). The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*. 53(8):474-485.

Berger S, Sinha AK, Roitsch T (2007). Plant physiology meets phytopathology: plant primary metabolism and plant – pathogen interactions. *Journal of Experimental Botany* 58(15–16):4019– 4026.

Bernardo R (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetetics*. 129(12):2323-2332.

van Bezouw RFHM, Keurentjes JJB, Harbinson J, Aarts MGM (2019). Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *Plant Journal*. 97(1):112-133.

van Bezouw RFHM, Janssen EM, Ashrafuzzaman M, Ghahramanzadeh R, Kilian B, Graner A, Visser RGF, van der Linden CG (2019b). Shoot sodium exclusion in salt stressed barley (Hordeum vulgare L.) is determined by allele specific increased expression of HKT1;5. *Journal of Plant Physiology*. 241:153029

Bilgin DD, Zavala JA, Zhu J, Clough, SJ, Ort DR, DeLucia EH (2010). Biotic stress globally downregulates photosynthesis genes. *Plant, Cell & Environment.* 33(10), 1597– 1613.

Bilgin Sonay T, Carvalho T, Robinson MD, Greminger MP, Krützen M, Comas D, Highnam G, Mittelman D, Sharp A, Marques-Bonet T, Wagner A. (2015). Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. *Genome Research.* 25(11):1591-1159.

Björkman O, Demmig B (1987). Photon yield of O_2 evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. *Planta* 170(4):489– 504.

Blankenship RE (2010). Early evolution of photosynthesis. *Plant Physiology*. 154(2):434-438.

Blum A (2013). Heterosis, stress, and the environment: a possible road map towards the general improvement of crop yield. *Journal of Experimental Botany*. 64(16):4829-4837

Botet R, Keurentjes JJB (2020). The Role of Transcriptional Regulation in Hybrid Vigor. *Frontiers in Plant Sciences*. 11:410.

Bouché F, Lobet G, Tocquin P, Périlleux C (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. *Nucleic Acids Research*. 44(D1):D1167-1171.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordberg M, Bergelson J, Cuguen J, Roux F (2010). Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature. *PLoS Genetics.*6(5).

Brachi B, Morris GP, Borevitz JO (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology.* 12:232.

Bray NL, Pimentel H, Melsted P, Pachter L (2016). Near-optimal probabilistic RNA-seq quantification, *Nature Biotechnology*. 34**:**525–527.

Broman KW, Wu H, Sen Ś, Churchill GA (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*. 19:889-890.

Bruce BD (2000). Chloroplast transit peptides: structure, function and evolution. *Trends in Cell Biology.* 10(10):440– 447.

Buchner DA, Nadeau JH (2015). Contrasting genetic architectures in different mouse reference populations used for studying complex traits. *Genome Research*. 25(6):775-791.

de Buck S, Podevin N, Nolf J, Jacobs A, Depicker A. (2009). The T-DNA integration pattern in Arabidopsis transformants is highly determined by the transformed target cell. *Plant Journal*. 60(1):134-45.

Cabrera-Bosquet L, Fournier C, Brichet N, Welcker C, Suard B, Tardieu F (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytologist.* 212(1):269– 281.

von Caemmerer S, Quick WP, Furbank RT (2012). The development of C_4rice: current progress and future challenges. *Science*. 11(2):228-31

Calvo-Baltanás V, Wijnen CL, Yang C, Lukhovitskaya N, de Snoo CB, Hohenwarter L, Keurentjes JJB, de Jong H, Schnittger A, Wijnker E (2020). Meiotic crossover reduction by virus-induced gene silencing enables the efficient generation of chromosome substitution lines and reverse breeding in Arabidopsis thaliana. *The Plant Journal* 104(5):1437-1452.

Carmo-Silva E, Scales JC, Madgwick PJ, Parry MAJ (2014). Optimizing Rubisco and its regulation for greater resource efficiency. *Plant, Cell & Environment.* 38(9):1817– 1832.

Carlborg Ö, Haley CS (2004). Epistasis: too often neglected in complex trait studies? *Nature Review Genetics*. 5(8):618-25.

Cerovic, ZG, Ounis, A, Cartelat, A, Latouche, G, Goulas, Y, Meyer, S, Moya, Z (2002). The use of chlorophyll fluorescence excitation spectra for the non-destructive in situ assessment of UV-absorbing compounds in leaves. *Plant, Cell & Environment.* 25(12):1663– 1676.

Cheng DD, Zhang ZS, Sun XB, Zhao M., Sun GY, Chow WS (2016). Photoinhibition and photoinhibition-like damage to the photosynthetic apparatus in tobacco leaves induced by *Pseudomonas syringae* pv. *tabaci* under light and dark conditions. *BMC Plant Biology.*16:29.

Choudhury FK, Rivero RM, Blumwald E, Mittler R (2017). Reactive oxygen species, abiotic stress and stress combination. *Plant Journal*. 90(5):856-867.

Clough SJ, Bent AF (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*. 16(6):735-43.

Cordell HJ (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Moleculair Genetics*. 11(20):2463-2468.

Corpet F (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Resesearch*. 25;16(22):10881-10890.

Coulton A, Przewieslik-Allen AM, Burridge AJ, Shaw DS, Edwards KJ, Barker GLA (2020). Segregation distortion: Utilizing simulated genotyping data to evaluate statistical methods. *PLoS One*. 15(2):e0228951

Cowley Jr AW, Liang M, Roman RJ, Greene AS, Jacob HJ (2004). Consomic rat model systems for physiological genomics. *Acta Physiologica Scandinavica* 181(4):585-592.

Croucher JS (2006). Collecting coupons - a mathematical approach. *Australian Senior Mathematics Journal.*20(2):31-35.

Crovshoff S, Hibberd JM. (2012). Integrating C4 photosynthesis into C3 plants to increase yield potential. *Current Opinions in Biotechnology.* 23(2):209– 214.

Crowell S, Korniliev P, Falcão A, Ismail A, Gregorio G, Mezey J, McCouch S (2016). Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nature Communications.* 7:10527.

Cruz, JA, Savage, LJ, Zegarac, R., Hall, CC, Satoh-Cruz, M., Davis, GA, Kent Kovac, W, Chen, J, Kramer, DM (2016). Dynamic environmental photosynthetic imaging reveals phenotypes. *Cell Systems.* 2(6):365–377.

Da Costa e Silva, L., CD Cruz, MA Moreira, EG de Barros (2007). Simulation of population size and genome saturation level for genetic mapping of recombinant inbred lines (RILs). *Genetics and Molecular Biology*. 30, 1101–1108.

De León JLD, Escoppinichi R, Geraldo N, Börner A, Röder MS (2011). The Performance of Single Chromosome Substitution Lines of Bread Wheat Subjected to Salinity Stress. *Cereal Research Communications. 39*(3):317–324.

Demmig-Adams B, Adams WW (2006). Photoprotection in an ecological context: the remarkable complexity of thermal energy dissipation. *New Phytologist.* 172:11– 21.

Dhanapal AP, Ray JD, Singh SK, Hoyos-Vilegas V, Smith JR, Purcell, LC and Fritsch, FB (2016). .Genome-wide association mapping of soybean chlorophyll traits based on chanopy spectral reflectance and leaf extracts. *BMC Plant Biology.*16:174.

Ding Y, Zhang R, Zhu L, Wang M, Ma Y, Yuan D, Liu N, Hu H, Min L, Zhang X (2021). An enhanced photosynthesis and carbohydrate metabolic capability contributes to heterosis of the cotton (Gossypium hirsutum) hybrid 'Huaza Mian H318', as revealed by genome-wide gene expression analysis. *BMC Genomics*. 22(1):277.

Ding J, Zhao J, Pan T, Xi L, Zhang J, Zou Z (2019). Comparative Transcriptome Analysis of Gene Expression Patterns in Tomato Under Dynamic Light Conditions. *Genes (Basel)*. 10(9):662.

Dirks R, van Dun K, de Snoo CB, van den Berg M, Lelivelt CL, Voermans W, Woudenberg L, de Wit JP, Reinink K, Schut JW, van der Zeeuw E, Vogelaar A, Freymark G, Gutteling EW, Keppel MN, van Drongelen P, Kieny M, Ellul P, Touraev A, Ma H, de Jong H, Wijnker E (2009). Reverse breeding: a novel breeding approach based on engineered meiosis. *Plant Biotechnology Journal*. 7(9):837-845.

Dormann CF, Elith J, Bacher S, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schrörder B, Skidmore AK, Zurell D, Luatenback S (2012). Collinearity: a review of methods to deal with and a simulation study evaluating their performance. *Ecography.* 36(1):27-46.

Dray S, Dufour A (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4):1-20.

Drolet GG, Huemmrich, KF Hall, FG Middleton, EM Black, TA, Barr AG (2005). A MODIS-derived photochemical reflectance index to detect inter-annual variations in the photosynthetic light-use efficiency of a boreal deciduous forest. *Remote Sensing of Environment* 98(2–3):212– 224.

Driever SM, Lawson T, Andralojc PJ, Raines CA, Parry MA (2014). Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. *Journal of Experimental Botany*. 65(17):4959-4973.

Driever SM, Simkin AJ, Alotaibi S, Fisk SJ, Madgwick PJ, Sparks CA, Jones HD, Lawson T, Parry MAJ, Raines CA (2017). Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philosophical Transactions of the Royal Society of London B*. 372(1730):20160384

Dunwell JM (2010). Haploids in flowering plants: origins and exploitation. *Plant Biotechnology Journal*. 8(4):377-424.

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Picó FX, Alonso-Blanco C, Hancock AM (2017). African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings National Academy of Science of the United States of America*. 114(20):5213-5218.

Dyson BC, Allwood JW, Feil R, Xu Y, Miller M, Bowsher CG, Goodacre R, Lunn JE, Johnson GN (2015). Acclimation of metabolism to light in Arabidopsis thaliana: the glucose 6-phosphate/phosphate translocator GPT2 directs metabolic acclimation. Plant Cell & Environment. 38(7):1404-1417.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*.5:113.

Edwards GE, Baker NR (1993). Can CO2 assimilation in maize leaves be predicted accurately from chlorophyll fluorescence analysis? Photosynth Research. 37(2):89-10.

Ehrenreich IM (2017). Epistasis: Searching for Interacting Genetic Variants Using Crosses. *Genetics*. 206(2):531-535.

El-Lithy, ME, Rodrigues, GC, van Rensen, JJ, Snel, JF, Dassen, HJ, Koornneef, M, Jansen, MA, Aarts, MGM and Vreugdehil, D (2005) Altered photosynthetic performance of a natural Arabidopsis accession is associated with atrazine resistance. *Journal of Experimental Botany*, 56(416), 1625– 1634.

El-Lithy ME, Bentsink L, Hanhart CH, Ruys GJ, Rovito D, Broekhof JLM, van der Poel HJA, van Eijk MJT, Vreugdenhil D, Koornneef M (2006). New Arabidopsis thaliana Recombinant Inbred Line Populations Genotyped Using SNPWave and Their Use for Mapping Flowering-Time Quantitative Trait Loci. *Genetics*. 172(3):1867-1876.

El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM (2014). Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends in Plant Science*. 19(6):390-398

El-Soda M, Kruijer W, Malosetti M, Koornneef M, Aarts MGM (2015). Quantitative trait loci and candidate genes underlying genotype by environment interaction in the response of Arabidopsis thaliana to drought. *Plant Cell and Environment*. 38(3):585-599.

Erb TJ, Zarzycki J (2018). A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Current Opinions in Biotechnology*. 49:100-107.

Erol (2019). Genetic analysis of nitrogen use efficiency in Arabidopsis thaliana. *PhD thesis*. Wageningen University.

Ermakova M, Danila FR, Furbank RT, von Caemmerer S (2020). On the road to C_4 rice: advances and perspectives. *Plant Journal*. 101(4):940-950.

Ermakova M, Arrivault S, Giuliani R, Danila F, Alonso-Cantabrana H, Vlad D, Ishihara H, Feil R, Guenther M, Borghi GL, Covshoff S, Ludwig M, Cousins AB, Langdale JA, Kelly S, Lunn JE, Stitt M, Caemmerer S, Furbank RT (2021). Installation of C_4 photosynthetic pathway enzymes in rice using a single construct. *Plant Biotechnology Journal.* (3):575-588.

Eseverri Á, Baysal C, Medina V, Capell T, Christou P, Rubio LM, Caro E (2020). Transit Peptides From Photosynthesis-Related Proteins Mediate Import of a Marker Protein Into Different Plastid Types and Within Different Species. *Frontiers in Plant Science*. 11:560701.

Evans JR, von Caemmerer, S (2000). Would C4 rice produce more biomass than C3 rice? *Studies in Plant Science*. 7:53–71.

Evans JR, Lawson T (2020). From green to gold: agricultural revolution for food security. *Journal of Experimental Botany*. 71(7):2211-2215

Fahlgren, N, Gehan, MA and Baxter, Z (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curent Opinion in Plant Biology.* 24, 93– 99.

Farquhar, GD, von Caemmerer, S and Berry, JA (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. *Planta*, 149, 78– 90.

Faralli M, Lawson T (2020). Natural genetic variation in photosynthesis: an untapped resource to increase crop yield potential? *Plant Journal*. 101(3):518-528.

Feng LL, Han YJ, Liu G, An BG, Yang J, Yang GH, Li YS, Zhu YG. (2007). Overexpression of sedoheptulose-1,7-bisphosphatase enhances photosynthesis and growth under salt stress in transgenic rice plants. *Functional Plant Biology*. 34, 822–834.

Feng LL, Wang K, Li Y, Tan YP, Kong J, Li H, Li YS, Zhu YG (2007). Overexpression of SBPase enhances photosynthesis against high temperature stress in transgenic rice plants. *Plant Cell Reports*. 26:1635–1646.

Fiedler, K., Bekele, WA, Matschegewski, C, Snowdon, R, Wieckhorst, S, Zacharias, A, Uptmoor, R (2016). Cold tolerance during juvenile development in sorghm: a comparative analysis by genomewide association and linkage mapping. *Plant Breeding*, 135, 598– 606.

Fischer-Kilbienski I, Miao Y, Roitsch T, Zschiesche W, Humbeck K, Krupinska K (2010). Nuclear targeted AtS40 modulates senescence associated gene expression in Arabidopsis thaliana during natural development and in darkness. *Plant Molecular Biology*. 73(4-5):379-90.

Flood PF (2015). Natural genetic variation in *Arabidopsis thaliana* photosynthesis. *PhD thesis*. Wageningen University.

Flood PJ, Harbinson J, Aarts MGM (2011). Natural genetic variation in plant photosynthesis. *Trends in Plant Science*. (6):327-35.

Flood PJ, Kruijer W, Schnabel SK, van der Schoor R, Jalink H, Snel JF, Harbinson J, Aarts MGM (2016). Phenomics for photosynthesis, growth and reflectance in Arabidopsis thaliana reveals circadian and long-term fluctuations in heritability. *Plant Methods*. 15(12):14.

Flood, PJ, van Heerwaarden, J, Becker, F, de Snoo, CB, Harbinson, J, Aarts, MGM (2016a) Whole-genome hitchhiking on an organelle mutation. *Current Biology.* 26(10):1306– 1311.

Flood PJ, Theeuwen TPJM, Schneeberger K, Keizer P, Kruijer W, Severing E, Kouklas E, Hageman JA, Wijfjes R, Calvo-Baltanas V, Becker FFM, Schnabel SK, Willems LAJ, Ligterink W, van Arkel J, Mumm R, Gualberto JM, Savage L, Kramer DM, Keurentjes JJB, van Eeuwijk F, Koornneef M, Harbinson J, Aarts MGM, Wijnker E. Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nature Plants*. 2020 6(1):13-21.

Forsberg, SKG, Andreatta, ME, Huang, X.Y, Danku, J, Salt, DE Carlborg, Ö. (2015). The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. *PLoS Genetics.* 11(11), e1005648.

Fox J, Weisberg S (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA.

Fristedt R (2017). Chloroplast function revealed through analysis of GreenCut2 genes. *Journal of Experimental Botany*. 68(9):2111-2120.

Fujimoto R, Uezono K, Ishikura S, Osabe K, Peacock WJ, Dennis ES (2018). Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breeding Science*. 68(2):145-158.

Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES (2012). Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proceedings National Academy of Science of the United States of America*. 109(18):7109-14.

Fuller MG, Saha S, Stelly DM, Jenkins JN, Tseng TM (2021). Assessing the Weed-Suppressing Potential of Cotton Chromosome Substitution Lines Using the Stair-Step Assay. *Plants (Basel)*. 10(11):2450.

Furbank RT, Tester M (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Sciences.*16(12), 635– 644.

Gamon, JA, Serrano, L, Surfus, JS (1997) The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, 112(4), 492– 501.

Garcia-Molina A, Leister D (2020). Accelerated relaxation of photoprotection impairs biomass accumulation in Arabidopsis. *Nature Plants*. 6(1):9-12

Gardeström P, Lernmark U (1995). The contribution of mitochondria to energetic metabolism in photosynthetic cells. *Journal of Bioenergitcs & Biomembranes*. 27(4):415-21.

Gardeström P, Igamberdiev AU (2016). The origin of cytosolic ATP in photosynthetic cells. *Physiologia Plantarum*. 157(3):367-379.

Garin V, Wimmer V, Borchardt D, Malosetti M, van Eeuwijk F (2021). The influence of QTL allelic diversity on QTL detection in multi-parental populations: a simulation study in sugar beet. *BMC Genomic Data.* 22(4):1-12.

Glauser G, Dubugnon L, Mousavi SA, Rudaz S, Wolfender JL, Farmer EE (2009). Velocity estimates for signal propagation leading to systemic jasmonic acid accumulation in wounded Arabidopsis. *Journal of Biological Chemistry*. 284(50):34506-13.

Genty B, Briantais, JM, Baker, NR. (1989). The relationship between quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochimica &. Biophysica Acta.* 990:87–92.

Gitelson AA, Buschmann C, Lichtenthaler HK (1999). The chlorophyll fluorescence ratio F735/F700 as an accurate measure of the chlorophyll content in plants. *Remote Sensing of Environment.* 69(3), 296– 302.

Gitelson AA, Merzlyak MN, Chivkunova, OB (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry & Photobiology.* 74(1):38– 45.

Gitelson AA, Gritz Y, Merzlyak, MN (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology.* 160, 271– 282.

Gitelson, AA, Keydan, GP, Merzlyak, MN (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*. 33, L11402.

Głowacka K, Kromdijk J, Leonelli L, Niyogi KK, Clemente TE, Long SP. An evaluation of new and established methods to determine T-DNA copy number and homozygosity in transgenic plants (2016). *Plant Cell & Environment.* 39(4):908-17

Gouy M, Guindon S, Gascuel O (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biologand Evolution*. 27(2):221-4.

Graham PJ, Nguyen B, Burdyny T, Sinton D (2017). A penalty on photosynthetic growth in fluctuating light *Scientific. Reports.* 7(1), 12513.

Großkinsky DK, Svensgaard J, Christensen S, Roitsch T (2015) Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *Journal of Experimental Botany* 66(18), 5429– 5440.

Grzybowski M, Wijewardane NK, Atefi A, Ge Y, Schnable JC (2021). Hyperspectral reflectance-based phenotyping for quantitative genetics in crops: Progress and challenges. *Plant Communications*. 2(4):100209

Gu, J, Yin, X, Struik, PC, Stomph, GJ and Wang, H (2012). Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L) leaves under drought and well-watered field conditions. *Journal of Experimental Botany* 63(1), 455– 469.

Hadariová L, Vesteg M, Hampl V., Krajčovič, J (2018). Reproductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Current Genetics.* 64(2):365– 387.

Hamblin MT, Buckler ES, Jannink, JL (2011). Population genetics of genomics-based crop improvement methods. *Trends in Genetics* 27, 98– 106.

Hammond-Kosack KE, Jones JD (1997).. Plant Disease Resistance Genes. *Annual Reviews of Plant Physiology and Plant Molecular Biology*. 48:575-607.

Hansen TF, Alvarez-Castro JM, Carter AJ, Hermisson J, Wagner GP. Evolution of genetic architecture under directional selection. Evolution. 2006 Aug;60(8):1523-36.

Hao, D, Chao, M, Yin, Z and Yu, D (2012) Genome-wide association analysis detecting significant single nucleotide polymorphisms for chlorophyll and chlorophyll fluorescence parameters in soybean (Glycine max) landraces. *Euphytica*, 186, 919– 931.

Harbinson, J (2018) Chlorophyll fluorescence as a tool for describing the operation and regulation of photosynthesis *in vivo*. In Light Harvesting in Photosynthesis. (R Croce, R Grondelle, H Amerongen and Z Stokkum, eds). Boca Raton, FL: CRC Press.

Hardwick K, Baker NR (1974). In vivo measurement of chlorophyll content of leaves. *New Phytologist* 72, 51– 54.

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*. 10(8):551-564.

Hazzouri KM, Khraiwesh B, Amiri KMA, Pauli D, Blake T, Shahid M, Mullath SK, Nelson D, Mansour AL, Salehi-Ashtiani K, Purugganan M, Masmoudi K (2018). Mapping of *HKT1;5* Gene in Barley Using GWAS Approach and Its Implication in Salt Tolerance Mechanism. F*rontiers of Plant Science.* 9:156.

Herritt, M, Dhanapal, AP, Fritschi, FB (2016). Identification of genomic loci associated with the photochemical reflectance index by genome-wide association study in soybean. *Plant Genome*, 9(2), 1– 12.

Hibberd JM, Sheehy JE, Langdale JA (2008). Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Current Opinions in Plant Biology.* 11(2):228-31

Hill WG, Goddard ME, Visscher PM (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics.* 4(2):e1000008.

Honda S, Ohkubo S, San N, Nakkasame A, Tomisawa K, Katsura K, Ookawa T, Nagano A, Adachi S (2021). Maintaining higher leaf photosynthesis after heading stage could promote biomass accumulation in rice. *Scientific Report*. 11:7579.

Houle, D, Govindaraju, DR, Omholt, S (2010) Phenomics: the next challenge. *Nature Reviews Genetics* 11(12), 855– 866.

Huang X, Effgen S, Meyer RC, Theres K, Koornneef M. (2012). Epistatic natural allelic variation reveals a function of AGAMOUS-*like*6 in axillary bud formation in Arabidopsis.*Plant Cell*. 24(6):2364-2379.

Huang da W, Sherman BT, Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research. 37(1):1-13.

Igamberdiev AU, Shen T, Gardeström P (2006). Function of mitochondria during the transition of barley protoplasts from low light to high light. *Planta*. 224(1):196-204.

Imam S, Noguera DR, Donohue TJ (2014). Global analysis of photosynthesis transcriptional regulatory networks. *PLoS Genetics* 10(12), 1– 21.

Ingvarsson PK, Street NR (2011). Association genetics of complex traits in plants. *New Phytologist* 189(4), 909– 922.

Jahn CE, Mckay JK, Mauleon R, Stephens J, McNally KL, Bush DR, Leung H, Leach JE (2011). Genetic variation in biomass traits among 20 diverse rice varieties. *Plant Physiology*. 155(1):157-68

Jansen M, Gilmer F, Biskup B, Nagel KA, Rascher U, Fischbach A, Briem S, Dreissen G, Tittmann S, Braun S, De Jaeger I, Metzlaff M, Schurr U, Scharr H, Walter A. (2009). Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Functional Plant Biology.* 36(11):902– 914.

Jiao WB, Schneeberger K (2017). The impact of third generation genomic technologies on plant genome assemblY *Current Opinion in Plant Biology.* 36, 64– 70.

Jung HS, Niyogi KK (2009). Quantitative genetic analysis of thermal dissipation in Arabidopsis. *Plant Physiology*. 150(2):977-986.

Kaiser E, Morales A, Harbinson, J (2018). Fluctuating light takes crop photosynthesis on a rollercoaster ride. *Plant Physiology.* 176(2), 977– 989.

Kaiser E, Walther D, Armbruster U. (2020). Growth under Fluctuating Light Reveals Large Trait Variation in a Panel of Arabidopsis Accessions. *Plants (Basel)*. 9(3):316.

Kalra, S., Kumar, S. & Singh, K (2015). Molecular analysis of *SQUALENE EPOXIDASE* gene from *Chlorophytum borivilianum* (Sant. and Fernand.). *Journal of Plant Biochemistry &Biotechnology.***24,** 417–424.

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007). Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. *Genetics*. 175(2):891-905.

Keurentjes JJB, Willems G, Van Eeuwijk F, Nordborg M, Koornneef M (2011). A comparison of population types used for QTL mapping in Arabidopsis thaliana. *Plant Genetic Resources,* 9(2):185-188.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M. Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat Genet. 2007 Sep;39(9):1151-5.

Kindgren, P, Kremnev, D, Blanco, NE, de Dios Barajas López, J, Fernández, AP, Tellgren-Roth, C, Kleine, T, Small, Z, Strand, Å. (2012) The plastid redox insensitive 2 mutant of Arabidopsis is impaired in PEP activity and high light-dependent plastid redox signalling to the nucleus. *Plant Journal.* 70(2), 279– 291.

Kirchgessner, N, Liebisch, F, Yu, K, Pfeifer, J, Friedli, M, Hund, A, Walter, A (2016) The ETH field phenotyping platform FIP: a cable-suspended multi-sensor system. *Functional Plant Biology.* 44(1), 154–168.

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant Journal*. 88(6):1058-1070.

Knoch D, Abbadi A, Grandke F, Meyer RC, Samans B, Werner CR, Snowdon RJ, Altmann T (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnology Journal*. 18(1):68-82.

Kokla A, Melnyk CW (2018). Developing a thief: haustoria formation in parasitic plants. *Developmental Biologoy.* 442(1):53– 59.

Kooke R, Wijnker E and Keurentjes JJB (2012). Backcross populations and near isogenic lines. *Methods in Molecular Biology.* 871:3–16.

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004). Naturally occurring genetic variation in Arabidopsis thaliana. *Annual Review of Plant Biology.* 55:141-72.

Korte A, Farlow A (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29.

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsy M (2002). STAIRS: a new genetic resource for functional genomic studies of Arabidopsis thaliana. *The Plant Journal.*31(3):355-364.

Kouřil, R, Wientjes, E, Bultema, JB, Croce, R, Boekema, EJ (2013). High-light vs. low-light: effect of light acclimation on photosystem II composition and organization in *Arabidopsis thaliana*. *Biochimica et Biophysica Acta.* 1827(3):411– 419.

Krewson TD, Supelak PJ, Hill AEl, Singer JB, Lander ES, Nadeau JH, Palmert MR (2004). Chromosomes 6 and 13 Harbor Genes that Regulate Pubertal Timing in Mouse Chromosome Substitution Strains, *Endocrinology.* 145(10): 4447–4451.

Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, Long SP (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*. 354(6314):857-861.

Kromdijk, J and Long, SP (2016) One crop breeding cycle from starvation? How engineering crop photosynthesis for rising CO_2 and temperature could be one important route to alleviation. *Proceedings of the Royal Society B.* 283(1826), 20152578.

Kubis A, Bar-Even A. (2019). Synthetic biology approaches for improving photosynthesis. *Journal of Experimental Botany*. 70(5):1425-1433.

Kuhlgert S, Austic G, Zegarac R, Osei-Bonsu I, Hoh D, Chilvers MI, Roth MG, Bi K, TerAvest D, Weebadde P, Kramer DM (2016) MultispeQ Beta: a tool for large-scale plant phenotyping connected to the open PhotosynQ network. *Royal Society Open Science*. 3(10):160592.

Kuhn, M, Böger, P (1990). Studies on the light-induced loss of the D1 protein in photosystem-II membrane fragments. *Photosynthesis Research.* 23(3):291– 296.

Kuppu S, Ron M, Marimuthu MPA, Li G, Huddleson A, Siddeek MH, Terry J, Buchner R, Shabek N, Comai L., Britt AB (2020). A variety of changes, including CRISPR/Cas9-mediated deletions, in CENH3 lead to haploid induction on outcrossing. *Plant Biotechnology Journal.* 18(10):2068-2080.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software.* 82(13):1–26

Laranjeira S, Amorim-Silva V, Esteban A, Arró M, Ferrer A, Tavares RM, Botella MA, Rosado A, Azevedo H (2014). Arabidopsis Squalene epoxidase 3 (SQE3) Complements SQE1 and Is Important for Embryo Development and Bulk Squalene epoxidase Activity. *Molecular Plant*. 8(7):1090-102.

Lachowiec J, Shen X, Queitsch C, Carlborg Ö (2015). A Genome-Wide Association Analysis Reveals Epistatic Cancellation of Additive Genetic Variance for Root Length in Arabidopsis thaliana. *PLoS Genetics*. 11(9):e1005541.

Lardon R, Wijnker E, Keurentjes J, Geelen D (2020). The genetic framework of shoot regeneration in Arabidopsis thaliana comprises master regulators and conditional fine-tuning factors. *Communications Biology*. 3(549).

Laurie C, Wang S, Carlini-Garcia, L.A. Zeng ZB (2014). Mapping epistatic quantitative trait loci. *BMC Genetics.*15:112

Latowski, D, Kuczyńska, P, Strzalka, K (2011). Xanthophyll cycle - a mechanism protecting plants against oxidative stress. *Redox Report.* 16(2):78– 90.

Lawson T, Kramer DM, Raines CA (2012). Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. *Current Opinion in Biotechnology*. 23(2):215-20.

Leister D (2003). Chloroplast research in the genomic age. *Trends in Genetics* 19(1), 47– 56.

Li XP, Muller-Moule P, Gilmore AM, Niyogi KK (2002). PsbS-independent enhancement of feedback de-excitation protects photosystem II photoinhibition. *Proceedings National Academy of Science of the United States of America.* .99(23):15222– 15227.

Li HH, Zhang LY, Wang JK (2012). Estimation of statistical power and false discovery rate of QTL mapping methods through computer simulation. *Chinese Science Bulletin*. 57:2701-2710.

Liang Y, Liu HJ, Yan J, Tian F (2021). Natural Variation in Crops: Realized Understanding, Continuing Promise. *Annual Review of Plant Biology*. 72:357-385.

Lin, PC, Tsai, YC, Hsu, SK, Ou, JH, Liao, CT, Tung, CW (2017). Identification of natural variants affecting chlorophyll dynamics during rice seedling development. *Plant Breeding*.137:355– 363.

Lister C, Dean C (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana. Plant Journal*. 4**:**745–750.

Liu Y, Zhou J, Hu T, Lu Y, Gao L, Tu L, Gao J, Huang L, Gao W (2020). Identification and functional characterization of Squalene epoxidase and oxidosqualenecyclases from Tripterygium wilfordii. Plant Cell Rep. 39(3):409-418.

Liu D, Dong S, Bo K, Miao H, Li C, Zhang Y, Zhang S, Gu X (2021). Identification of QTLs Controlling Salt Tolerance in Cucumber (*Cucumis sativus* L.) Seedlings. *Plants (Basel)*. 10(1):85.

Long SP, Humphries S, Falkowski PG (1994). Photoinhibition of photosynthesis in nature. *Annual Review of Plant Physiology and Molecular Biology.* 45, 633– 662.

Long SP, Farage PK, Garcia RL (1996). Measurement of leaf and canopy photosynthetic CO_2 exchange in the field. *Journal of Experimental Botany* 47(11):1629– 1642.

Long SP, Zhu XG, Naidu SL, Ort DR (2006). Can improvement of photosynthesis increase crop yield? *Plant, Cell & Environment* 29(3), 315– 330.

Long SP, Marshall-Colon A, Zhu XG (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell*. 161(1):56-66.

Loudet O, Chaillou S, Camilleri C, Daniel-Vedele F (2002). Bay-0 × Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis thaliana. *Theoretical and Applied Genetics*.104:1173–1184.

Love MI, Huber W, Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15:550.

Lv J, Yu K, Wei J, Gui H, Liu H, Liu C, Liang D, Wang Y, Zhou H, Carlin R, Rich R, Lu T, Que Q, Wang WC, Zhang X, Kelliher T (2020). Generation of paternal haploids in wheat by genome editing of the centromeric histone *CENH3*. *Nature Biotechnology.*38:1397-1401.

Lye ZN, Purugganan MD (2019). Copy Number Variation in Domestication. *Trends in Plant Science*, 24(4):352-365.

Mackay TF (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*. 35:303-39.

Mackay TF (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nature Review Genetics*.15(1):22-33.

Manzoor MM, Goyal P, Pandotra P, Dar MS, Dar MJ, Misra P, Gupta AP, Vishwakarma RA, Ahuja A, Dhar MK, Gupta S (2021). Transcriptome-wide identification of Squalene epoxidase genes from Glycyrrhiza glabra L.: expression analysis and heterologous expression of GgSQE1 suggest important role in terpenoid biosynthesis. *Protoplasma*. 258(5):991-1007.

Martín M, Sabater B (2010). Plastid ndh genes in plant evolution. *Plant Physiology & Biochemistry.* (2010). 48(8):636-645.

Maxwell K, Johnson GN (2000). Chlorophyll fluorescence – a practical guide. *Journal of Experimental Botany* 51(345):659– 668.

Mayr C (2019). What Are 3' UTRs Doing? *Cold Spring Harbour Perspectives on Biology*.11(10):a034728.

Mellenthin M, Ellersiek U, Börger A, Baier M. Expression of the Arabidopsis Sigma Factor SIG5 Is Photoreceptor and Photosynthesis Controlled (2014). *Plants (Basel)*. 3(3):359-91.

Meuwissen TH, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4):1819-29.

Miclaus, M, Balacescu, O, Has, Z, Balacescu, L, Has, V., Suteu, D, Neuenschwander, S, Keller, Z, Bruggmann, R (2016). Maize cytolines unmask key nuclear genes that are under the control of retrograde signaling pathways in plants. *Genome Biology and Evolution.*. 8(11), 3256– 3270.

Miller AK, Chen A, Bartlett J, Wang L, Williams SM, Buchner DA (2020). A Novel Mapping Strategy Utilizing Mouse Chromosome Substitution Strains Identifies Multiple Epistatic Interactions That Regulate Complex Traits. *Genes, Genomes, Genetics.*10(12):4553-4563.

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F, Tardieu F (2019). Genomic prediction of maize yield across European environmental conditions. *Nature Genetics.* 51(6):952-956.

Möller IM (2001). PLANT MITOCHONDRIA AND OXIDATIVE STRESS: Electron Transport, NADPH Turnover, and Metabolism of Reactive Oxygen Species. *Annual Review of Plant Physiology and Plant Molecular Biology*. 52:561-591.

Morales A, Kaiser E. Photosynthetic Acclimation to Fluctuating Irradiance in Plants. (2020). Frontiers in Plant Sciences. 11:268.

Morales, A, Kaiser, E, Yin, X, Harbinson, J, Molenaar, J, Driever, SM and Struik, PC (2018). Dynamic modelling of limitations on improving leaf CO_2 assimilation under fluctuating irradiance. *Plant, Cell & Environment.* 41(3), 589– 604.

Mishra Y, Jänkänpää HJ, Kiss AZ, Funk C, Schröder WP, Jansson S (2012). Arabidopsis plants grown in the field and climate chambers significantly differ in leaf morphology and photosystem components. *BMC Plant Biology*. 11;12:6.

Müller, P, Li, X, Niyogi, KK (2001). Non-photochemical quenching. A response to excess light energY *Plant Physiology.* 125(4), 1558– 1566.

Mitchell PL, Sheehy JE. Supercharging rice photosynthesis to increase yield (2006). New Phytologist. 171(4):688-93

Munns R, Tester M. (2008). Mechanisms of salinity tolerance. *Annual Review of Plant Biology*. 59:651-81.

Murchie, EH (2017). Safety conscious or living dangerously: what is the 'right' level of plant photoprotection for fitness and productivity. *Plant, Cell & Environment.* 40(8):1239– 1242.

Murchie EH, Harbinson J (2014). Non-photochemical fluorescence quenching across scales: From chloroplasts to plants to communities. In Non-photochemical quenching and energy dissipation in plants, algae and cyanobacteria (B Demmig-Adams, G Garab, WW Adams and U.Z Govindjee, eds). Dordrecht: Springer; pp. 553– 582.

Murchie EH, Hubbart S, Chen Y, Peng S, Horton P (2002). Acclimation of rice photosynthesis to irradiance under field conditions. *Plant Physiology* 130(4):1999-2010.

Murchie EH, Kefauver S, Araus JL, Muller O, Rascher U., Flood PJ, Lawson T (2018). Measuring the dynamic photosynthome. *Annals of Botany.* 122(2), 207– 220.

Murchie, EH, Lawson, T (2013) Chlorophyll fluorescence analysis: a guide to good pratice and understanding some new applications. *Journal of Experimental Botany* 64(13), 3983– 3998.

Nadeau J, Singer J, Matin A, Lander ES (2000).Analysing complex genetic traits with chromosome substitution strains. *Nature Genetics.*24, 221–225.

Nadeau JH, Forejt J, Takada T, Shiroishi T (2012). Chromosome substitution strains: gene discovery, functional analysis, and systems studies. *Mammalian Genome*.23(9-10):693-705.

Nagashima A, Hanaoka M, Shikanai T, Fujiwara M, Kanamaru K, Takahashi H, Tanaka K. The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the psbD blue light-responsive promoter (BLRP) in Arabidopsis thaliana (2004). *Plant and Cell Physiology*. 45(4):357-68

Nagelmüller S, Kirchgessner N, Yates, S Hiltpod, M Walter A (2016). Leaf Length Tracker: a novel approach to analyse leaf elongation close to the thermal limit of growth in the field. *Journal of Experimental Botany* 67(6):1897– 1906.

Nguyen VL, Dolstra O, Malosetti M, Kilian B, Graner A, Visser RG, van der Linden CG (2013). Association mapping of salt tolerance in barley (Hordeum vulgare L.). *Theoretical and Applied Genetics.* 126(9):2335-51.

Ning C, Wang, D Kang, H Mrode, R Zhou, L Xu, S Liu JF (2018). A rapid epistatic mixed-model association analysis by linear retransformations of genomic estimated values. *Bioinformatics*, 34(11), 1817– 1825.

Nishiyama Y, Murata N (2014). Revised scheme for the mechanism of photoinhibition and its application to enhance the abiotic stress tolerance of the photosynthetic machinerY *Applied Microbioogy and Biotechnology.* 98(21):8777– 8796.

Noordally ZB, Ishii K, Atkins KA, Wetherill SJ, Kusakina J, Walton EJ, Kato M, Azuma M, Tanaka K, Hanaoka M, Dodd AN (2013). Circadian control of chloroplast transcription by a nuclear-encoded timing signal. *Science*. 339(6125):1316-9.

Nunes-Nesi A, Carrari F, Lytovchenko A, Smith AM, Loureiro ME, Ratcliffe RG, Sweetlove LJ, Fernie AR (2005). Enhanced photosynthetic performance and growth as a consequence of decreasing mitochondrial malate dehydrogenase activity in transgenic tomato plants. *Plant Physiology*. 137(2):611-22.

O'Malley RC, Barragan CC, Ecker JR. A user's guide to the Arabidopsis T-DNA insertion mutant collections (2015). *Methods in Molecular Biology*. 1284:323-342.

Oakley CG, Savage L, Lotz S, Larson GR, Thomashow MF, Kramer DM, Schemske DW (2018). Genetic basis of photosynthetic responses to cold in two locally adapted populations of Arabidopsis thaliana. *Journal of Experimental Botany*. 69(3):699-709.

de Oliveira Silva FM, Lichtenstein G, Alseekh S, Rosado-Souza L, Conte M, Suguiyama VF, Lira BS, Fanourakis D, Usadel B, Bhering LL, DaMatta FM, Sulpice R, Araújo WL, Rossi M, de Setta N, Fernie AR, Carrari F, Nunes-Nesi A. (2018). The genetic architecture of photosynthesis and plant growth-related traits in tomato. *Plant, Cell & Environment.* 41(2):327– 341.

Ort DR, Zhu X , Melis A (2011), Optimizing antenna size to maximize photosynthetic efficiency *Plant Physiology.* 155(1), 79– 85.

Ort DR, Merchant SS, Alric J, Barkan A, Blankenship RE, Bock R, Croce R, Hanson MR, Hibberd JM, Long SP, Moore TA, Moroney J, Niyogi KK, Parry MA, Peralta-Yahya PP, Prince RC, Redding KE, Spalding MH, van Wijk KJ, Vermaas WF, von Caemmerer S, Weber AP, Yeates TO, Yuan JS, Zhu XG (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings National Academy of Science of the United States of America*. 112(28):8529-36.

Ortiz, D, Hu, J, Salas Fernandez, MG (2017). Genetic architecture of photosynthesis in sorghum bicolour under non-stress and cold stress conditions. *Journal of Experimental Botany* 68(16):4545– 4557.

Pastore D, Trono D, Laus MN, Di Fonzo N, Flagella Z (2007). Possible plant mitochondria involvement in cell adaptation to drought stress. A case study: durum wheat mitochondria. *Journal of Experimental Botany*. 58(2):195-210.

Patel AB, Greber BJ, Nogales E (2019). Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Current Opinion in Structural Biology.* 61:17-24.

Pettersson G, Ryde-Pettersson U (1988). A mathematical model of the Calvin photosynthesis cycle. *Eurpean Journal of Biochemistry*. 175(3):661-672.

Piironen V, Lindsay DG, Miettinen TA, Toivo J, Lampi AM(2000). Plant sterols: biosynthesis, biological function and their importance to human nutrition. *Journal of the Science of Food and Agriculture.* 80: 939–966.

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*. 14(7):687-690.

Posé D, Castanedo I, Borsani O, Nieto B, Rosado A, Taconnat L, Ferrer A, Dolan L, Valpuesta V, Botella MA (2009). Identification of the Arabidopsis dry2/sqe1-5 mutant reveals a central role for sterols in drought tolerance and regulation of reactive oxygen species. *Plant Journal*. 59(1):63-76.

Poolman MG, Fell DA, Thomas S (2000). Modelling photosynthesis and its control. *Journal of Experimental Botany*. 51:319-328

Prado, SA, Cabrera-Bosquet, L, Grau, A, Coupel-Ledru, A, Millet, EJ, Welcker, C. Tardieu, F (2018). Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. *Plant, Cell & Environment.* 41:314– 326.

Prins, A, Orr, DJ, Andralojc, PJ, Reynods, MP, Carmo-Silva, E and Parry, MAJ (2016). Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. *Journal of Experimental Botany* 67(6), 1827– 1838.

Prinzenberg AE, Campos-Dominguez L, Kruijer W, Harbinson J, Aarts MGM (2020). Natural variation of photosynthetic efficiency in Arabidopsis thaliana accessions under low temperature conditions. *Plant, Cell and Environment*. 43(8):2000-2013.

Paul MJ, Foyer CH (2001). Sink regulation of photosynthesis. *Journal of Experimental Botany*. 52(360):1383-400.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ray DK, Mueller ND, West PC, Foley JA (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One*. 8(6):e66428.

Rasbery JM, Shan H, LeClair RJ, Norman M, Matsuda SP, Bartel B (2007). Arabidopsis thaliana Squalene epoxidase 1 is essential for root and seed development. *Journal of Biological Chemistry*. 282(23):17002-13.

Ravi M, Chan SW (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature*. 2010 25:464(7288):615-8.

Ren Y, Zhang Z, Liu J, Staub JE, Han Y, Cheng Z, Li X, Lu J, Miao H, Kang H, Xie B, Gu X, Wang X, Du Y, Jin W, Huang S (2009). An integrated genetic and cytogenetic map of the cucumber genome. *PLoS One*. 4:4(6):e5795

Ren J, Wu P, Trampe B, Tian X, Lübberstedt T, Chen S (2017). Novel technologies in doubled haploid line development. *Plant Biotechnology Journal*. 15(11):1361-1370.

Reeves G, Singh P, Rossberg TA, Sogbohossou D, Schranz ME Hibbard JM (2018). Natural variation within a species for traits underpinning C4 photosynthesis. *Plant Physiology.* 177(2):504–512.

Reumann S., Inoue K., Keegstra, K. *(*2005*)* Evolution of the general protein import pathway of plastids. *Molecular & Membrane Biology.* 22:73–86.

Richy E, Leister D (2004). An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. *Gene* 31(329):11– 16.

Roux, F, Mary-Huard, T, Barillot, E, Wenes, E, Botran, L, Durand, S, Villoutreix, R, Martin-Magniette, ML, Camilleri, C and Budar, F (2016). Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proceedings National Academy of Science of the United States of America.* 113(13), 3687– 3692.

van Rooijen R, Aarts MGM, Harbinson J. (2015). Natural genetic variation for acclimation of photosynthetic light use efficiency to growth irradiance in Arabidopsis. *Plant Physiology*. 167(4):1412-29.

van Rooijen R, Kruijer W, Boesten R, van Eeuwijk FA, Harbinson J, Aarts MGM (2017). Natural variation of YELLOW SEEDLING 1 affects photosynthetic acclimation of Arabidopsis thaliana. *Nature Communications*. 8(1):1421.

van Rooijen R, Harbinson J. Aarts MGM (2018). Photosynthetic response to increased irradiance correlates to variation in transcriptional response of lipid-remodelling and heat-shock genes. *Plant Direct.* 2(7):1– 19.

Ruban AV, Young A, Horton, P (1993) Induction of nonphotochemical energy dissipation and absorbance changes in leaves (evidence for changes in the state of the light-harvesting system of photosystem II in vivo). *Plant Physiology.* 102(3):741– 750.

Ruban AV. (2017). Quantifying the efficiency of photoprotection. *Philosophical Transactions of the Royal Society of London B.* 372(1730), 1– 10.

Rungrat T, Awlia M, Brown T, Cheng R, Sirault X, Fajkus J, Trtilek M, Furbank B, Badger M, Tester M, Pogson BJ, Borevitz JO, Wilson P (2016). Using Phenomic Analysis of Photosynthetic Function for Abiotic Stress Response Gene Discovery. *Arabidopsis Book.* 14:e0185.

Rungrat T, Almonte AA, Cheng R, Gollan PJ, Stuart T, Aro EM, Borevitz JO, Pogson B, Wilson PB (2019). A Genome-Wide Association Study of Non-Photochemical Quenching in response to local seasonal climates in Arabidopsis thaliana. *Plant Direct*. 3(5):e00138.

Saeki N, Kawanabe T, Ying H, Shimizu M, Kojima M, Abe H, Okazaki K, Kaji M, Taylor JM, Sakakibara H, Peacock WJ, Dennis ES, Fujimoto R (2016). Molecular and cellular characteristics of hybrid vigour in a commercial hybrid of Chinese cabbage. *BMC Plant Biology*. 16:45.

Saha S, Stelly DM, Raska DA, Wu J, Jenkins JN, McCarty JC, Makamov A, Gotmare V, Abdurakhmonov IY, Campbell BT (2012). Chromosome Substitution Lines: Concept, Development and Utilization in the Genetic Improvement of Upland Cotton. In *Plant Breeding,* Abdurakhmonov IY, Intech Open.

Sambrook J, Fritsch E.F, Maniatis T (1989). Molecular Cloning: A Laboratory Manual. Cold Spring harbor laboratory Press.

Schauberger B, Archontoulis S, Arneth A, Balkovic J, Ciais P, Deryng D, Elliott J, Folberth C, Khabarov N, Müller C, Pugh TA, Rolinski S, Schaphoff S, Schmid E, Wang X, Schlenker W, Frieler K. (2017). Consistent negative response of US crops to high temperatures in observations and crop models. *Nature Communications.* 8:13931.

Schneider T, Bolger A, Zeier J, Preiskowski S, Benes V, Trenkamp S, Usadel B, Farré EM, Matsubara S. (2019). Fluctuating Light Interacts with Time of Day and Leaf Development Stage to Reprogram Gene Expression. *Plant Physiology*. 179(4):1632-1657.

Schüler ML, Mantegazza O, Weber APM (2016). Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. *Plant Journal.* 87(1), 51– 56.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006) Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* 18, 1121–1133.

Scholes JD, Rolfe SA (2009). Chlorophyll fluorescence imaging as tool for understanding the impact of fungal diseases on plant performance: a phenomics perspective. *Functional Plant Biolology.* 36(11):880– 892.

Scholl RL, May ST, Ware DH (2000). Seed and molecular resources for Arabidopsis . Plant Physiology. 124(4):1477-80.

Seguí-Simarro JM, Moreno JB, Fernández MG, Mir R (2021). Species with Haploid or Doubled Haploid Protocols. *Methods in Molecular Biology*. 2287:41-103.

Shimada TL, Shimada T, Hara-Nishimura I (2010). A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana. *Plant Journal*. 61(3):519-528.

Silva-Perez, V., Molero, G, Serbin, SP, Condon, AG, Reynolds, MP, Furbank, RT Evans, JR (2018). Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. *Journal of Experimental Botany.* 69(3):483– 496.

Simkin AJ, McAusland L, Headland LR, Lawson T, Raines CA (2015). Multigene manipulation of photosynthetic carbon assimilation increases CO2 fixation and biomass yield in tobacco. *Journal of Experimental Botany*. 66(13):4075-4090.

Singh A, Ganapathysubramanian B, Singh AK, Sarkar S (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Sciences.*21(2), 110– 124.

Singer JB, Hill AE, Burrage LC, Olszens KR, Song J, Justice M, O'Brien W, Conti DV, Witte JS, Lander ES, Nadeau JH (2004). Genetic Dissection of Complex Traits with Chromosome Substitution Strains of Mice. *Science*. 304(5669):445-448.

Singer T, Fan Y, Chang HS, Zhu T, Hazen SP, Briggs SP (2006). A high-resolution map of Arabidopsis recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genetics*. 2(9):e144.

Slattery RA, Walker BJ, Webe, APM, Ort DR (2018) The impacts of fluctuating light on crop performances. *Plant Physiology.* 176(2), 990– 1003.

Spiezio SH, Takada T, Shiroishi T, Nadeau JH (2012). Genetic divergence and the genetic architecture of complex traits in chromosome substitution strains of mice. *BMC Genetics*. 18:13-38.

Sonawane PD, Pollier J, Panda S, Szymanski J, Massalha H, Yona M, Unger T, Malitsky S, Arendt P, Pauwels L, Almekias-Siegl E, Rogachev I, Meir S, Cárdenas PD, Masri A, Petrikov M, Schaller H, Schaffer AA, Kamble A, Giri AP, Goossens A, Aharoni A (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nature Plants*. 22(3):16205.

South PF, Cavanagh AP, Liu HW, Ort DR (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*. 363(6422):eaat9077.

Stewart DW, Costa C, Dwyer L (2003). Canopy structure, light interception and photosynthesis in maize. *Agronmy Journal* 95(6):1465– 1474.

Stinziano, JR, Morgan, PB, Lynch, DJ, Saathoff, AJ, McDermitt DK, Hanson DT (2017). The rapid A-C_i response: photosynthesis in the phenomic era. *Plant, Cell & Environment* 40(8), 1256– 1262.

Stirling, CM, Nie, GY, Aguilera, C, Nugawela, A, Long, SP, Baker, NR (1991) Photosynthetic productivity of an immature maize crop: changes in quantum yield of CO_2 assimilation, conversion efficiency and thylakoid proteins. *Plant, Cell & Environment.* 14:947– 954.

Strigens, A, Freitag, NM, Gilbert, X, Grieder, C, Riedelsheimer, C, Schrag, TA, Messmer, R, Melchinger, AE (2013). Association mapping for chilling tolerance in elite flint and dent maize inbred lines evaluated in growth chamber and field experiments. *Plant, Cell & Environment.* 36:1871– 1887.

Sun, S, Li, C, Paterson, AH, Jiang, Y, Xu, R, Robertson, J, Snider, JL, Chee, PW (2018). In-field high throughput phenotyping and cotton plant growth analysis using LiDAR *Frontiers in Plant Science.* 9:16.

Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, Miyajima N, Sasamoto S, Kimura T, Hosouchi T, Kawashima K, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakayama S, Nakazaki N, Naruo K, Okumura S, Shinpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Sato S, de la Bastide M, Huang E, Spiegel L, Gnoj L, O'Shaughnessy A, Preston R, Habermann K, Murray J, Johnson D, Rohlfing T, Nelson J, Stoneking T, Pepin K, Spieth J, Sekhon M, Armstrong J, Becker M, Belter E, Cordum H, Cordes M, Courtney L, Courtney W, Dante M, Du H, Edwards J, Fryman J, Haakensen B, Lamar E, Latreille P, Leonard S, Meyer R, Mulvaney E, Ozersky P, Riley A, Strowmatt C, Wagner-McPherson C, Wollam A, Yoakum M, Bell M, Dedhia N, Parnell L, Shah R, Rodriguez M, See LH, Vil D, Baker J, Kirchoff K, Toth K, King L, Bahret A, Miller B, Marra M, Martienssen R, McCombie WR, Wilson RK, Murphy G, Bancroft I, Volckaert G, Wambutt R, Düsterhöft A, Stiekema W, Pohl T, Entian KD, Terryn N, Hartley N, Bent E, Johnson S, Langham SA, McCullagh B, Robben J, Grymonprez B, Zimmermann W, Ramsperger U, Wedler H, Balke K, Wedler E, Peters S, van Staveren M, Dirkse W, Mooijman P, Lankhorst RK, Weitzenegger T, Bothe G, Rose M, Hauf J, Berneiser S, Hempel S, Feldpausch M, Lamberth S, Villarroel R, Gielen J, Ardiles W, Bents O, Lemcke K, Kolesov G, Mayer K, Rudd S, Schoof H, Schueller C, Zaccaria P, Mewes HW, Bevan M, Fransz P; Kazusa DNA Research Institute; Cold Spring Harbor and Washington University in St Louis Sequencing Consortium; European Union (2000). Arabidopsis Genome Sequencing Consortium. Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature*. 408(6814):823-826.

Taagen E., Bogdanove AJ, Sorrells ME (2020). Counting on Crossovers: Controlled recombination for Plant Breeding. *Trends in Plant Sciences*. 25(4):455-465.

Tang N, Shahzad Z, Lonjon F, Loudet O, Vailleau F, Maurel C (2018). Natural variation at XND1 impacts root hydraulics and trade-off for stress responses in Arabidopsis. Nature *Communication*. 9(1):3884.

Takahashi S, Murata N (2008), How do environmental stresses accelerate photoinhibition? *Trends in Plant Sciences.*13(4):178– 182.

Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennet M (2017). Plant phenomics, from sensors to knowledge. *Current Biology.* 27(15):R770– R783.

Tardieu F, Simonneau T, Muller B (2018). The Physiological Basis of Drought Tolerance in Crop Plants: A Scenario-Dependent Probabilistic Approach. *Annual Review in Plant Biology*. 69:733-759.

Taylor SH, Long SP (2017). Slow induction of photosynthesis on shade to sun transitions in wheat may cost at least 21% of productivity *Philosophical Transactions of the Royal Society of London B.* 372(1730):20160543.

Theeuwen TPJM, Logie LL, Harbinson J, Aarts MGM (2022). Genetics as a key to improving crop photosynthesis. *Journal of Experimental Botany*. 2:erac076.

Thoen MP, Davila Olivas NH, Kloth KJ, Coolen S, Huang PP, Aarts MG, Bac-Molenaar JA, Bakker J, Bouwmeester HJ, Broekgaarden C, Bucher J, Busscher-Lange J, Cheng X, Fradin EF, Jongsma MA, Julkowska MM, Keurentjes JJ, Ligterink W, Pieterse CM, Ruyter-Spira C, Smant G, Testerink C, Usadel B, van Loon JJ, van Pelt JA, van Schaik CC, van Wees SC, Visser RG, Voorrips R, Vosman B, Vreugdenhil D, Warmerdam S, Wiegers GL, van Heerwaarden J, Kruijer W, van Eeuwijk FA, Dicke M (2016). Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *New Phytologist.* 213(3):1346– 1362.

Tian, F, Bradbury, PJ, Brown, PJ, Hung, H, Sun, Q., Flint-Garcia, S, Rocheford, TR, McMullen, MD, Holland, JB, Buckler, S (2011). Genome-wide association of study of leaf architecture in the maize nested association mapping population. *Nature Genetics.* 43(2):159– 162.

Tietz S, Hall CC, Cruz JA, Kramer DM (2017). NPQ(T): a chlorophyll fluorescence parameter for rapid estimations and imaging of non-photochemical quenching of excitons in photosystem-II-associated antenna complexes. *Plant, Cell & Environment.* 40(8):1243– 1255.

Timm S, Woitschach F, Heise C, Hagemann M, Bauwe H (2019). Faster Removal of 2-Phosphoglycolate through Photorespiration Improves Abiotic Stress Tolerance of Arabidopsis. *Plants (Basel)*. 8(12):563

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. *Proceedings National Academy of Science of the United States of America*. 108(50), 20260–20264.

Tsai SF, Tung CW, Tsai CA, Liao CT (2017). An exhaustive scan method for SNP main effects and SNP x SNP interactions over highly homozygous genomes. Journal of *Computational Biology.* 24(12):1254–1264.

Tschiersch H, Junker A, Meyer RC, Altmann T (2017). Establishment of integrated protocols for automated high throughput kinetic chlorophyll fluorescence analyses. *Plant Methods.* 13:54.

Turner TL (2014). Fine-mapping natural alleles: quantitative complementation to the rescue. *Molecular Ecology.* 23(10):2377– 2382.

Umbach AL, Fiorani F, Siedow JN (2005). Characterization of transformed Arabidopsis with altered alternative oxidase levels and analysis of effects on reactive oxygen species in tissue. *Plant Physiology*. 139(4):1806-1820.

Unrau J, Person C, Kuspira J. (1957). CHROMOSOME SUBSTITUTION IN HEXAPLOID WHEAT. *Canadian Journal of Botany*. 34(4):629-640.

Ungerer MC, Halldorsdottir SS, Modliszewski JL, Mackay TF, Purugganan MD (2002). Quantitative trait loci for inflorescence development in Arabidopsis thaliana. *Genetics*. 160(3):1133-51.

VanderWeele TJ, Mathur MB (2019). Some desirable properties of the bonferroni correction: is the bonferroni correction really so bad? *American Journal of Epidemiology*. 188(3):617-618.

Vialet-Chabrand S, Matthews JS, Simkin AJ, Raines CA, Lawson T (2017). Importance of fluctuations in light on plant photosynthetic acclimation. *Plant Physiology.* 173(4):2164– 2179.

Viaud G, Loudet O Cournède PH (2017). Leaf segmentation and tracking in *Arabidopsis thaliana* combined to an organ-scale plant model for genotypic differentiation. *Frontiers in Plant Science.* 7:2057.

Vogel A, Schwacke R, Denton AK, Usadel B, Hollmann J, Fischer K, Bolger A, Schmidt MH, Bolger ME, Gundlach H, Mayer KFX, Weiss-Schneeweiss H, Temsch EM, Krause K. (2018). Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nature Communications* 9:2515.

Wada M, Kagawa T, Sato Y (2003). Chloroplast movement. *Annual Review of Plant Biology.* 54, 455–468.

Walker, BJ, Dewry, DT, Slattery, RA, Van Loocke, A, Cho, YB, Ort, DR (2018) Chlorophyll can be reduced in crop canopies with little penalty to photosynthesis. *Plant Physiology.* 176(2):1215– 1232.

Wang HY, Fu Y, McPeek MS, Lu X, Nuzhdin S, Xu A, Lu J, Wu ML, Wu CI (2008). Complex genetic interactions underlying expression differences between Drosophila races: analysis of chromosome substitutions. *Proclamation of the National Acadamey of Science of the United States of America*. 105(17):6362-7.

Wang M, Liu S, Zhang S, Miao H, Tian G, Lu H, Liu P, Wang Y, Gu X (2016). QTL mapping of seedling traits in cucumber using recombinant inbred lines. *Plant Breeding* 135:124-129.

Wang Q, Xie W, Xing H, Yan J, Meng X, Li X, Fu X, Xu J, Lian X, Yu S, Xing Y, Wang G. (2015). Genetic architecture of natural variation in rice chlorophyll content revealed by a genome-wide association study. *Molecular Plant*. 8(6):946– 957.

Wang Q, Zhao H, Jiang J, Xu J, Xie W, Fu X, Liu C, He Y, Wang G (2017). Genetic Architecture of Natural Variation in Rice Nonphotochemical Quenching Capacity Revealed by Genome-Wide Association Study. *Frontiers in Plant Science*. 8:1773.

Webb D, Burnison B, Trimbee A, Prepas EE (1992). Comparison of Chlorophyll a Extractions with Ethanol and Dimethyl Sulfoxide/Acetone, and a Concern about Spectrophotometric Phaeopigment Correction. *Canadian Journal of Fisheries and Aquatic Sciences*. 49:2331-2336.

Wei T, Simko V (2021). *R package corrplot: Visualization of a Correlation Matrix*. (Version 0.92)

Wei X, Zhang Z, Shi P, Wang P, Chen Y, Song X, Fulu, T (2015). Is yield increase sufficient to achieve food security in China? *PLoS ONE*, 10(2):e0116430.

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4

Wijnen CL (2019). Effects of reduced crossover recombination on quantitative trait analysis. *PhD thesis*. Wageningen University. 176p.

Wijnen CL, Keurentjes JJ (2014). Genetic resources for quantitative trait analysis: novelty and efficiency in design from an Arabidopsis perspective. *Current Opinion Plant Biology*. 18:103-109.

Wijnker E, Deurhof L, van de Belt J, de Snoo CB, Blankenstijn H, Becker F, Ravi M, Chan SWL, van Dun K, Lelivelt CLC, de Jong H, Dirks R, Keurentjes JJB (2014). Hybrid recreation by reverse breeding in Arabidopsis thaliana. *Nature Protocols.*9:761-772.

Wijnker E, van Dun KPM, de Snoo CB, Lelivelt CLC, Keurentjes, JJB, Naharudin NS, Ravi M, Chan SWL, de Jong H, Dirks R (2012). Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. *Nature Genetics.*44:467-470.

Wilson D, Cooper J (1969) Apparent photosynthesis and leaf characters in relation to leaf position and age, among contrasting *Lolium* genotypes. *New Phytologist.* 68:645– 655.

Wingler A, Lea PJ, Quick WP, Leegood RC (2000). Photorespiration: metabolic pathways and their role in stress protection. *Philosophical Transactions of the Royal Society London B*. 355(1402):1517-29.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson, GV, Provart, NJ (2007). An 'electronic fluorescent pictograph' browserfor exploring and analyzing large-scale biological data sets. *PLoSOne*2:e718.

Wu J, McCarty JC, Jenkins JN (2010). Cotton chromosome substitution lines crossed with cultivars: genetic model evaluation and seed trait analyses. *Theoretical and Applied Genetics*. 120(7):1473-83.

Valitova JN, Sulkarnayeva AG, Minibayeva FV (2016). Plant Sterols: Diversity, Biosynthesis, and Physiological Functions. *Biochemistry (Mosc)*. 81(8):819-34.

Voss I, Sunil B, Scheibe R, Raghavendra AS (2013). Emerging concept for the role of photorespiration as an important part of abiotic stress response. *Plant Biology*. 15(4):713-722

Xu L, Cruz JA, Savage LJ, Kramer DM, Chen J (2015). Plant photosynthesis phenomics data quality control. *Bioinformatics*, 31(11):1796– 1804.

Yan, XY, Qu, CM, Li, JN, Chen, L, Liu, LZ (2015). QTL analysis of leaf photosynthesis rate and related physiological traits in Brassica. *Journal of Integrated Agriclture*, 14(7):1261– 1268.

Yokota A., Uemura K., Kobayashi H. (1999). Super-Rubisco for Improvement of Photosynthetic Performances of Plants. In: Altman A., Ziv M., Izhar S. (eds) Plant Biotechnology and In Vitro Biology in the 21st Century. Current Plant Science and Biotechnology in Agriculture, vol 36. Springer, Dordrecht.

Yoon DK, Ishiyama K, Suganami M, Tazoe Y, Watanabe M, Imaruoka S, Ogura M, Ishida H, Suzuki Y, Obara M, Mae T, Makino A (2020). Transgenic rice overproducing Rubisco exhibits increased yields with improved nitrogen-use efficiency in an experimental paddy field. *Nature Food* 1**:**134–139.

Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H, Xie W, Lian X, Wang G, Luo Q, Zhang Q, Liu Q, Xiong L. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation rice. *Nature communications.* 5:5087.

Yao L, Zhang Y, Liu C, Liu Y, Wang Y, Liang D, Liu J, Sahoo G, Kelliher T (2018). *OsMATL* mutation induces haploid seed formation in indica rice. *Nature Plants.* 4:530-533.

Yendrek, CR, Tomaz, T, Montes, CM, Cao, Y, Morse, AM, Brown, PJ, McIntyre, LM, Leakey, ADB Ainsworth, EA (2017). High-throughput phenotyping of maize leaf physiology and biochemical traits using hyperspectral reflectance. *Plant Physiology.* 173(1), 614– 626.

Yu, X, Zheng, G, Shan, L, Meng, G, Vingron, M, Liu Q, Zhu, XG (2014) Reconstruction of gene regulatory networks related to photosynthesis in *Arabidopsis*. *Frontiers in Plant Science.* 5:273.

Zapata L, Ding J, Willing EM, Hartwig B, Bezdan D, Jiao WB, Patel V, Velikkakam James G, Koornneef M, Ossowski S, Schneeberger K (2016). Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. *Proceedings National Academy of Science of the United States of America*. 113(28):E4052-4060.

Zan Y, Forsberg SKG, Carlborg Ö (2018). On the Relationship Between High-Order Linkage Disequilibrium and Epistasis. *Genes, Genomes, Genetics*. 8(8):2817-2824.

van Zanten M, Snoek LB, Proveniers MC, Peeters AJ (2009). The many functions of ERECTA. *Trends in Plant Science*. 14(4):214-218.

Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Huang M, Yao Y, Bassu S, Ciais P, Durand JL, Elliott J, Ewert F, Janssens IA, Li T, Lin E, Liu Q, Martre P, Müller C, Peng S, Peñuelas J, Ruane AC, Wallach D, Wang T, Wu D, Liu Z, Zhu Y, Zhu Z, Asseng S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings National Academy of Science of the United States of America*.114(35):9326–9331.

Zhao P, Cui R, Xu P, Wu J, Mao JL, Chen Y, Zhou CZ, Yu LH, Xiang CB (2017). ATHB17 enhances stress tolerance by coordinating photosynthesis associated nuclear gene and ATSIG5 expression in response to abiotic stress. *Scientific Reports*. 7:45492.

Zhao YR, Li X, Yu KQ, Cheng F, He Y (2016). Hyperspectral imaging for determining pigment contents in cucumber leaves in response to angular leaf spot disease. *Scientific Reports* 6:22790.

Zhang X, Huang C, Wu D, Qiao F, Li W, Duan L, Wang K, Xiao Y, Chen G, Liu Q, Xiong L, Yang W, Yan J. (2017). High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiology.* 173(3):1554–1564.

Zhang M, Zhao J, Li W, Wen S, Huang H, Dong J, Liu B, Zhang G, Wang HB, Shen Y, Jin HL (2020). Increased photosystem II translation efficiency as an important photoprotective mechanism in an Arabidopsis thaliana ecotype (Tibet-0) adapted to high light environments. *Environmental and Experimental Botany*. 183:104350.

Zhou W, Cheng Y, Yap A, Chateigner-Boutin AL, Delannoy E, Hammani K, Small Z, Huang J (2009). The Arabidopsis gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant Journal.* 58(1):82– 96.

Zoschke R, Bock R (2018). Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation. *Plant Cell*. 30(4):745-770.

Zhu S, Fang G (2018). MatrixEpistasis: ultrafast, exhaustive epistasis scan for quantitative traits with covariate adjustmenT *Bioinformatics*, 34(15), 2341– 2348.

Zhu, XG, Ort, DR, Whitmarsh, J, Long, SP (2004) The slow reversibility of photosystem II thermal energy dissipation on transfer from high to low light may cause large losses in carbon gain by crop canopies: a theoretical analysis. *Journal of Experimental Botany* 55(400):1167–1175.

Zhu XG, Ort DR, Parry MAJ, von Caemmerer S (2020). A wish list for synthetic biology in photosynthesis research. *Journal of Experimental Botany*. 71(7):2219–2225.

Zhu XG, Song Q, Ort DR (2012). Elements of a dynamic systems model of canopy photosynthesis. *Current Opinion in Plant Biology.* 15(3):237–244.

Zhu XG, de Sturler E, Long SP (2007). Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant Physiology*. 145(2):513–26.

Zhu XG, Long SP, Ort DR (2010). Improving Photosynthetic Efficiency for Greater Yield. *Annual Review of Plant Biology.* 61(1):235–261.

Zuk, O, Hechter, E, Sunyaev, SR, Lander, ES (2012). The mystery of missing heritability: Gene interactions create phantom heritability. *Proceedings National Academy of Science of the United State sof America*. 109(4):1193–1198.

Acknowledgements

Completing a PhD is a tough journey that is generally thought to be a lonely journey, but you get to know many fantastic people along the road – sometimes from unexpected corners!

Jeremy, Joost and **Mark**, thanks for putting your trust in my ability to complete a PhD-thesis. You taught me valuable lessons in science which I will take at heart in my future career. **Jeremy**, I highly appreciated your easy going on just about every topic, be it politics, science or even the simplest things in life. I particularly enjoyed your non-genetic background, which kept me sharp on explaining genetics at a slower pace. **Joost**, discussions with you on genetics and quantitative analysis are seldomly short-lived, but highly enjoyable. Your easy take on how to approach science and kindness are most valuable. **Mark,** I enjoyed the many discussions, and the drinks and dinners with the group. You kept me motivated to pursue daring methods in molecular biology**,** a topic you are passionate about.

Maaike, **Magdalena**, **Evert** thanks a lot for supporting the project and the useful discussions we had during the project meetings. You greatly contributed to my understanding of the wider context of relevance of the research**. Bastiaan**, thank you a lot for performing the genotyping work in most of the chapters. Without you, the chance of developing the various finemapping populations and other genotypic material would be close to zero, well, below zero!

Jitpanu, René B and **Tom** and, I had a blast of spending time with you in the past five years. You were the salt of the earth, and some of the finest colleague-friends that I ever met. We were always debating everything important, including genetics, lab policies, the unequivocal blessings of the monarchy, the importance of diversity on the beverage chart, indigenous lifeforms in the People's Republic of Noord-Brabant, phenotyping equipment that never breaks down, 1001 ways to let your Arabidopsis plants die (particularly the important ones!) etc, etc, etc. With you, I lived through many good, bad and damn right stupid moments – some which I, (un)fortunately, will never forget! Together, we supervised nearly an entire generation of future plant science students of which I lost the count track. It must have been over 80! I am sure you will be able to develop your dream careers in the future - in frogs & teaching (**René B**), phenomics & photosynthesis (**Tom**) and shrimp & genetics (**Jitpanu**, 55555) respectively. In addition, **René B**, an important, but friendly reminder; I did find functional copy number variation and you did not ;-) .

Phuong, our paths were set to cross on multiple occasions and I was glad to shake hands on a number of projects with you. Thanks for everything, the life contemplations, discussions on scientific politics and of course your tremendous support and insight during the cooperative experimental trajectories.

Cris and **Ramon**, thanks for putting your effort in developing the CSL-populations and your nice discussions about their use throughout the project. **Erik**, your enthusiasm for plant science always flows in an upbeat motion and even resulted in the development of one of the experimental chapters. You are very good at spreading that enthusiasm around, and hope you will have passed that into your personal F1's. **Emilie** and **Vanesa,** you were the final members of the group of Joost and often contributed to – sometimes much needed – realism into the discussions we had.

Thanks to **Mark**, **Louise**, **Francesco**, **Valeria**, **Tanía, Robert, Tom**, **René**, **Frank**, **Phuong**, **Corrie**, **Grmay, Hedayat**, **Majid**, **Fatemeh, Maarten K** and the many, many others (the total number of faces passing by in the group of Mark must have totalled close to a hundred in the time I was involved, including all students. Suffice to say, impossible to remember all!) for the many discussions on my research. But also the research of others, which definitely served as starting points for further inspiration to move forward. Particularly your input, **Maarten K**, drawing from a full career of working in Arabidopsis was very valuable to carry on with, and I always loved your stories – particularly those about some of your favourite accessions.

Hedayat, **Majid**, **Fatemeh**, you came from Iran as visting researchers and I learned a lot from you, including the Persian cuisine. Fatemeh, I hope you will enjoy your time in the Netherlands, and I will surely visit you more.

Much more inspiring than acquiring new skills yourself is to successfully pass your knowledge onto the next generation of scientists and particularly when some excel beyond your own expectations and capabilities. I got to supervise and know twenty-five of you, from all parts of this colourful world; **Casper, Iqbal**, **Livia, Laura, Rosanne**, **Alan, Imme, Jeroen, Afke, Yanrong, Chih-Wei, Silvia, Thimo, Yvet, Hans, Elena, Marieke, Sietske, Stan, Stefan, Rik, Luc, Thirza, Rik, Luc** and **Sandra.** The group dinners made me got to know you and your motivations a little better. Without your motivation, I could have never figured out how to move forward all of the work that you have done, and you even managed to solve problems using unique ideas. Honourable mentions are reserved for **Rosanne** and **Stefan**, whose work led to hundreds and hundreds lines of sophisticated code representing complex genetic analyses which even I still have trouble figuring out by myself. Furthermore, I wish **Yanrong**, **Alan** and **Elena** the best of luck in the completion of PhD projects on their own!

My project extended technically into the group of Horticulture and Product Physiology, in which I met many nice people. Foremost **Francisca** (now in Genetics), with whom I could always have a good chatter. Your cloning work and ideas where highly valuable contributors to the development of Chapter 3. Each question that I asked about molecular cloning always came with additional answers on questions that I hadn't even thought of asking. And thanks **Julian**, for your pleasant occasional passing-by's. **Maarten W**, you proved to be indispensable to keep Phenovator running and to come up with all the lousy jokes in the process of doing so. **Elias**, you are an amazing scientist, covering an interesting topic that I was allowed to borrow from occasionally during my project. Thank you, **Leo**, for giving the opportunity to do several seminars in your group – it was nice to present and discuss my genetics work several times to a non-genetics environment.

Joost vdH, Corrie, Frank, José, Lidia, **Jordi** and **Patrick** and the other people at the labs, your help always proved to be useful, but particularly when I always had to learn yet another lab technique. **Corrie**, in my absence you will find that others also sometimes mess up the seedlab ;-) .

In addition, to those in the chair group that have directly contributing to my work, I would like thank the members of the **Genetics** for the times and stories at the lab. There were many people, and particularly Corona made it hard to keep track of all who passed by and arrived, making it hard to remember all. I would like to thank specifically **Ben** for his outstanding views on science, politics and sense of humour. Furthermore, thanks **Fons** and **Duur** for the valuable conversations about PhD politics and life trajectories, particularly during the final stages of completing my dissertation. **Sijmen** and **Kim** for organizing the GSS's. **Bwalya**, **Helena**, **Lennart**, **Mariska**, **Mathijs**, **Philippe** and **Kim**, you made for good occasional talks about (PhD-)life – particularly during corona when Radix wasn't so populated anymore. **Nyati**, this thesis shows that it is possible to complete one such, I wish you well in your fortunes to finish yours. I particularly enjoyed the vrijmibo's (**Alanna**, **Alex**, **Ben**, **Jenny**, **René B**, **Tom**, **Chipspanu**, **Helena, Mariska** and other regular visitors), lab outings and Christmas specials a lot. Thanks to **Bas**, for pulling the group through the pandemic in pleasant and caring way.

Thanks **Jaïr**, **René K** and **Johan** for putting your trust in having me as your new addition to the genomic breeding team at Rijk Zwaan Breeding B.V.. Your confidence clearly increased my motivation to complete this thesis, although it somehow made the final months feel to pass by at a very slow pace. I am looking forward to the years to come!

Harimurti, **Micha**, **Mariam**, **Kyrylo** and **Wei**, though often living at larger distance, you prove to be valuable and elevated both complaining and dreaming to higher levels. I would like to particularly thank **Kyrylo** and **Wei** for translating my summary in their mother tongues.

Bob, **Marius**, **Sebas**, **Nena** and **Shireen**, although you often were at the receiving end of my complaining, we always found times to come together. Now COVID passed on, I hope we will be able to visit each other more regularly again.

Particular thanks to **Elly,** for teaching me the required pipetting skills during my MSc thesis that eventually lead me to be able to succesfully clone ten genes, as demonstrated in this work. We also wrote a paper together on salt tolerance. And **John**, you believed in my capabilities to be able to complete an MSc in Plant Breeding – even without the proper educational background. The support of both of you proved to be invaluable to prepare me for and pull me through my chosen career path. I hope your new home in Bennekom will suit the both of you well.

Special thanks also the people of Unifarm; **Taede**, **Gerrit**, **David** and **Yannick,** for caring the plants and performing. And **Rinie**, apologies that we always demanded the technically impossible though it always seemed to work out! Thanks also for the people managing and maintaining my plants – at times very many – in Serre, **Bertus**, **Bert** and **Casper**, who where always here to help out growing my Arabidopsis plants.

My scientific journey started off in ecology. I am grateful for **Irma**, **Frank** and **Matty**, who are credited as being to the first to spark a lasting enthusiasm for science, quantitative analysis and...springtails!

Thanks **Frans** and **Matty**, for aiding in the identification of springtails that I wasn't able to during those nights in which I was able to grab my microscopes and could clear my mind from my PhD project. Thanks also, **Anne**, **Jan** and, particularly, **Koen**, for accompanying me during all of those field trips that carried me throughout The Netherlands. We made some valuable contributions to the Dutch Fauna already, adding three new species of springtails last year alone! Hopefully the planned Caribbean trip this year with you, **Koen,** will bring us closer to one new for science as well.

Incredibly, the first time I met you, **Padraíc**, was neither at a photosynthesis conference nor at a plant breeding seminar nor at a science outing, but rather wandering through the cold and deep mud in harsh, rainy weather, searching for the little collies. Perhaps one day you will find that Wijnkeriella species :-) . You have lots of great ideas for the future, and I am keen on hearing more of those in the future.

Mom and **dad**, you always supported me, regardless of the unpredictable twists I took in the past couple of years. Your patience was definitely worth it, even though I sometimes raised doubts about my chosen path myself. **Gido** and **Stein**, I finally joined you in the doctoral ranks. Together with **Sandra, René L, Karolien**, **Kees**, **Lucas** and **Floor**, the family events always replenished my senses and optimism during this journey.

Herman† and **Ria,** you both raised me as a little child, and always made sure to love the beauty of life in all of its forms. Even though you are no longer able to attend my defence, I will proudly carry the banner into my doctorate.

Suzanne, my dear love. Eight years ago, I made you a promise, during what must have been the among the coldest winters of your life and I consider it now fulfilled. At times, we have been standing back to back, facing harsh challenges in life – at times multiple at once. However, we never failed to prevail and the difficulties can only strengthen the special connection I enjoy with you ♥.

Summary

Photosynthesis is one of the last traits in which maximization is still feasible for purposes of improving crop productivity. The past decade has shown a sharp increase in the number of studies that prove and establish the significance of natural variation in photosynthesis traits, and this thus opens up the potential to use this resource in plant breeding programs. **Chapter 2** describes this trend in which photosynthesis phenotyping technology is finally starting to keep up with the biotechnological advances that lead to a revolution in genetic and genomic analyses in the past twenty years. Photosynthesis traits are, however, highly complex traits that often involve many genes that are low in effect and difficult to detect in genetic mapping studies. To counteract these, bi-parental mapping populations offer several important advantages including high detection power of quantitative trait loci (QTL), ease of generating genotypes to demonstrate the impact of QTL and a low genomic complexity which helps identifying causal variants. This thesis aims to demonstrate these advantages in studying natural genetic variation in photosynthesis traits in *Arabidopsis thaliana* using three different bi-parental mapping types in three different light environments. **Chapter 3** describes the genetic mapping of photosynthesis efficiency in the Ler-0 x Col-0 recombinant inbred line population, in different stable light conditions. Fine mapping of the single QTL resulted in the discovery of copy number variation in a previously uncharacterized gene family of *SQUALENE EPOXIDASE*s. This chapter demonstrates how converging genomics and phenomics can lead l to the discovery of novel photosynthesis traits, and how they can be used in transgenic improvement. In **Chapter 4** I describe the identification of QTL in an F2-population for photosynthesis efficiency in response to high light. These QTL, when combined, show a profound effect on photosynthesis efficiency. The large effect size prompted me to select lines carrying these QTL and study the response in productivity to a variety of fluctuating light regimes. To further anticipate on the use of fluctuating light regimes in photosynthesis research, I designed a genetic mapping approach, described in **Chapter 5**, that directly studied genetic variation for losses in productivity resulting from fluctuating light. The chromosome substitution line (CSL-)population used in this study showed large variation in their productivity tolerance to fluctuating light. Most notably, the F1 hybrid grown in this study expressed the high productivity of one parent, while maintaining the tolerance to fluctuating light of the other parent. Based in these results, I conclude that tolerance to fluctuating light may not only be a result of biophysical limitations, but also that by selection for this trait it is possible to genetically improve it. As the use of CSL-populations with multiple substitutions presents a novel concept, with only few literature reports available, in **Chapter 6** I modelled various aspects associated with reverse breeding, the development of CSL-populations and their use in quantitative trait analysis. The models developed and conclusions drawn in this chapter can be used to optimize strategies for the development of CSL-populations, also beyond the model species *A. thaliana*. In all, this thesis demonstrates the effectiveness and potential of biparental mapping populations to elucidate genetic variation for photosynthesis traits in *A. thaliana*, and provides a guideline for strategic use and development of CSL-populations in quantitative trait analysis.

最大化光合作用的效率是为数不多仍能提高作物产量的可行途径之一.在过去十年中，快速新增的大量研究证实了自然变异对光合性状的重要性，并**开**辟了其在植物育**种**中的应用潜力

第二章介绍了光合作用表型分型技术终于**开**始跟上引领了过去二十年遗传和基因组分析革命的生物技术的发展**步**伐。然而，光合性状是高度**复**杂的性状，通常**涉**及许多遗传定位研究中难以检测到的微效基因。双亲定位(bi-parental mapping）拥有克服这些影响的几个重要优势，包括对数量性状位点 (QTL) 的高检测能力、易于生成基因型用以证明 QTL的效力以及有助于鉴定**关**键变异位点的低基因组复杂性。本论文旨在利用三个不同的双亲定位群体，研究三**种**不同光环境下光合作用性状的遗传变异，以展示该技术在拟南芥光合作用自然遗传变异研究中的优势。

第三章描述了Ler-0 x Col-0 重组自交系群体在不同稳定光照条件下光合效率的遗传定位。通过单个数量性状位点（QTL）的精细定位，我们发现了之前未被报道的角鲨烯环氧酶（SQUALENE EPOXIDASEs）基因家族中的拷贝数变异。本章展示了基因组学和表型组学的结合如何助力新型光合作用性状的发掘，以及介绍了如何将它们用于转基因改良。

第四章描述了杂交二代（F2）群体中与强光下光合效率变化相**关**QTL的定位。这些QTL的组合对拟南芥光合作用效率有显著影响。因此，我筛选了含有这些高表型解释率位点的株系并研究了它们在不同波动光谱下生产效率的变化。

为进一**步**探索光合作用研究中波动光照机制的使用，我设计了第五章中描述的遗传定位方法，该方法研究了因光照波动导致的光合作用效率降低的遗传变异。使用的染色体代换系 (CSL-) 群体在波动光的光合作用耐受性方面具有很大的差异.最值得注意的是，本研究中生长的杂交一代（F1）表现出一个亲本的高光合效率，并同时保持了另一个亲本对波动光的耐受性。这些结果表明了对波动光的耐受性可能不仅是生物物理限制的结果，而且可能通过对这一性状的选择在遗传上进行改良。

由于只有很少文献报道使用了多替换染色体代换系（CSL）群体定位这一新概念，所以在第六章中，我从染色体代换系群体（CSL **种**群）的构建及其数量性状分析的角度建立了反向育**种**的**关**联模型。本章**开**发的模型和得出的结论可用于优化模式物**种**拟南芥即其他作物的染色体代换系群体（CSL **种**群）的构建策略。

综上，本论文证明了双亲定位群体在阐明拟南芥光合作用性状遗传变异方面的有效性和潜力，并为数量性状分析中染色体代换系(CSL)群体的战略性应用和发展提供了指导。

(Thanks to Wei Xiong for the translation)

Фотосинтез є однією з останніх ознак, максимізація якої все ще можлива з метою підвищення продуктивності сільськогосподарських культур. Останнє десятиліття продемонструвало різке збільшення кількості досліджень, які доводять і визнають значущість природної варіації ознак фотосинтезу, щовідкриває можливостідля використання цього ресурсу в програмах селекції рослин. **Розділ 2** описує тенденцію, за якою технологія фенотипування фотосинтезу нарешті наздоганяє біотехнологічні досягнення, які призвели до революції в генетичному та геномному аналізі за останні двадцять років. Однак ознаки фотосинтезу є дуже комплексними ознаками, які часто задіютьчисленні гени з низькоюефективністю і важкістю виявлення в дослідженнях з генетичного картування. Двобатьківськіпопуляції для картування допомагають оминути ці обмеження танадають кілька важливих переваг, включаючи високу здатність виявлення локусів кількісних ознак (ЛКО), легкість створення генотипів для демонстрації впливу ЛКО та низьку геномну комплексність, яка допомагає ідентифікувати причинні варіанти. Ця дисертація має на меті демонстрацію цих переваг для вивчення природної генетичної варіації ознак фотосинтезу *Arabidopsis thaliana* на прикладі використання трьох окремих двобатьківських популяцій для картування та вивчення генетичної варіації ознак фотосинтезу в трьох різних світлових середовищах. **Розділ 3** описує генетичне картування ефективності фотосинтезу в різних стабільних умовах освітлення в популяції рекомбінантних інбредних ліній Ler-0 x Col-0. Точне картування єдиного ЛКО призвело до виявлення варіації кількості копій у раніше неохарактеризованій родині генів, що кодують сквален епоксидази. Цей розділ демонструє якпоєднання геноміки та феноміки може привести до відкриття нових ознак фотосинтезу та як їх можна використати для трансгенного покращення. У **розділі 4** я описую ідентифікацію ЛКО ефективності фотосинтезу присильному освітленні у поколінні F2. Ці ЛКО в сукупності демонструють глибокий вплив на ефективність фотосинтезу. Висока величина цього ефекту спонукала мене вибрати лінії, що містять ці ЛКО, і дослідити зміни у продуктивності при різноманітних коливаннях світлових режимів. Для пришвидшення подальшого використання мінливих світлових режимів у дослідженнях фотосинтезу, я розробив підхід до генетичного картографування, описаний у **розділі 5**, для безпосередньогодослідження генетичних варіаційпов'язаних зі зменшенням продуктивності внаслідок світлових коливань. Популяція лінійз заміщенимихромосомами (ЛЗХ), використана в цьому дослідженні, мала велику варіацію в стійкості їхпродуктивності до світлових коливань. Зокрема використаний у цьому дослідженні гібрид F1 продемонстрував високу продуктивність однієї батьківської лінії, зберігаючи при цьому стійкість до світлових коливань від іншої батьківської лінії. У відповідності з цими результатами я роблю висновки, що толерантність до світлових коливань може бути не тільки результатом біофізичних обмежень, але й те, що шляхом відбору за цією ознакою можна генетично покращити її. Оскільки використання популяцій ЛЗХ з кількома заміщеннями представляє собою нову концепцію, що згадується лише у декількох літературних звітах, у **розділі 6** я змоделював різні аспекти, пов'язані зі зворотною селекцією, заснуванням популяцій ЛЗХ та їх використанням у аналізі кількісних ознак. Розроблені моделі та висновки, зроблені в цьому розділі, можуть бути використані для оптимізації стратегій створення популяцій ЛЗХ за межами модельного організму*A. thaliana*. Загалом, ця дисертація демонструє ефективність і потенціал використання двобатьківських популяцій для картування тавизначення генетичної варіації ознак фотосинтезу *A. thaliana*, а також надає рекомендації щодо стратегій використання та створення популяцій ЛЗХдля кількісного аналізу ознак.

(*Thanks Kyrylo Schenstnyi for the translation*)

Publications

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. doi: 10.1111/tpj.14190.

van Bezouw, R.F.H.M., Janssen, E.M., Ashrafuzzaman, M., Ghahramanzadeh, R., Kilian, B., Graner, A., Visser, R.G.F., van der Linden, C.G. (2019). Shoot sodium exclusion in salt stressed barley (Hordeum vulgare L.) is determined by allele specific increased expression of HKT1;5. *Journal of Plant Physiology*. 241:153029

van Bezouw, R.F.H.M, T.G. Wijnker, J.J.B. Keurentjes (in prep). Development, use and analysis of CSL-populations in plants – a modeling approach

van Bezouw, R. F. H. M (2012)**.** Groene ruggengraat als Europese verplichting. – Wolters Kluwer, 2012 - 2 - Milieu en Recht 40[2]: 95-97.

Berg M.P., , van Langevelde**,** F**,** Wynhoff I, **van Bezouw, R.F.H.M.** (2015). De springstaarten *Folsomides parvulus* en *F. angularis* nieuw voor de Nederlandse fauna. Nederlandse Faunistische Mededelingen 45:57-66

van Bezouw, R.F.H.M., van Langevelde**,** F**,** Wynhoff I, Berg, M.P. (2016). Springstaarten als indicatorsoort voor bodemkwaliteit in blauwgraslanden. Entomologische Berichten 76(2): 69-79

Berg, M.P., van Langevelde**,** F**,** Wynhoff I, **van Bezouw, R.F.H.M.** (2018). De stofluisspringstaart *Mackenziella psocoides* nieuw voor de Nederlandse fauna (Hexapoda: Collembola: Mackenziellidae). Nederlandse Faunistische Mededelingen 51: 23-32.

Berg, M.P., J. van Leeuwen & **R.F.H.M. van Bezouw** (2020). Het ruw dikstaartje *Paychyotoma crassicauda*, een springstaart nieuw voor de fauna van Nederland (Collembola: Isotomidae). Nederlandse Faunistische Mededelingen 54: 43-52.

van Bezouw, R.F.H.M., McCulloch J., Janssens F., Berg M.P. **(**in prep**)** Identification of black *Sminthurinus* species in Western Europe, with a redescription of *S. lawrencei* Gisin, 1963.

Van Veen, K. & **Van Bezouw, R.F.H.M.** (submitted to Nederlandse Faunistische Mededelingen). Vondsten van *Sminthurides bifidus* Mills, 1934 (gevorkte waterspringer) en *Sminthurinus domesticus* Gisin, 1963 *(*kaskogelspringer*)* in de Hortus botanicus in Leiden, nieuw voor Nederland.

Sminthurinus lawrencei in natural habitat, 27-II-2021, NL (Photo: Jan van Leeuwen)

Biography

Roel Franciscus Herman Maria van Bezouw was born on December 16th 1989 in Leiden, on the west coast of The Netherlands. He grew up here with his parents and his two brothers. From a young age he developed an interest in nature, science and history. After finishing havo in 2007 he went to study Forest- and Nature conservation at the hogeschool in Delft and graduated as a Bachelor of Applied sciences four years later. He studied chick-feeding in night jars for internship with ir. Henk Sierdsema at SOVON Vogelonderzoek Nederland and the legality of habitat connectivity for Natura 2000 at Natuurmonumenten with Toine Cooijmans. During the final year of his bachelor study he also got the opportunity to take a few courses at the Wageningen University to enjoy the taste of Academia, and continued to pursue an MSc degree in Ecology. Here, he completed an MSc thesis focussing on patterns of recovery of springtails (Collembola) under supervision of Prof. dr. Matty P. Berg, Prof. dr. Frank van Langevelde and dr. Irma Wynhoff, at de Vlinderstichting. In 2015 Roel received the prestigious Willem Barentsz Award for his work on springtails. Working on Collembola became a lasting passion. For his minor thesis he continued to work with Frank on vitellogenin modelling in honey bee brood as a cause for the decline of honey bee hive vitality.

Roel's personal life was tumultuos during his time studying in Wageningen. He lived away from his friends and family with little time to travel and visit them and his girlfriend fell sick to a long-term illness. He decided to drop the idea of doing something with nature in the academic world and set his sights on a career in plant breeding instead – a world he only just got in touch with at the Wageningen campus. After less than half a year of courses, he came to do a thesis with dr. C. Gerard van der Linden & Elly Janssen on natural variation in the potassium transporter *HKT1;5*, and QTL mapping of salt tolerance and ion content in barley. He published part of the results obtained during this thesis in Plant Breeding he published as first author in *Journal of Plant Physiology*. Keen on working in the plant breeding industry, he did do an internship at Hazera Seeds B.V. under supervision of dr. Rik van Wijk, in which he developed a hybrid performance prediction model. By the end of 2016 Roel eventually graduated at the Wageningen University for his degree in Forest- and Nature conservation, and early 2017 for his second master in Plant Sciences.

After finishing his second master he started looking for starter positions at various breeding companies, but kept his eyes open for PhD positions. He ended up only applying for industrially funded PhD position at the Laboratory of Genetics with dr. ing. Joost Keurentjes, Prof. dr. ir. Mark Aarts and dr. Jeremy Harbinson.

His time in Radix was dynamic to say the least. However, he developed a wide variety of skills, from genetic modelling and quantitative analysis to molecular cloning and plant transformation. He very much enjoyed supervising students on their projects. Last but not least, Roel's passion for the plant breeding field only grew further. When his contract ended, he engaged in a brief postdoctoral research to finish the work that was left behind by his premier collaborative partner Nguyen thu-Phuong.

In March 2022, Roel submitted his dissertation. Meanwhile, he continues in the plant breeding industry as Genomic Breeder Open Field Crops at Rijk Zwaan Breeding B.V. Here, he is still getting the challenge he needs to grow as a person, but his life has slowed down to a more enjoyable pace.

C: I always make fun of it and stuff, but isn't everything we dream in life a way to be loved a little more?

J: Yeah. I don't know. Sometimes I dream about being a good father and a good husband. And sometimes that feels really close. But then other times, it seems silly, like it would ruin my whole life. And it's not just a fear of commitment, or that I'm incapable of caring or loving, because... I can. It's just that, if I'm totally honest with myself, I think I'd rather die knowing that I was really good at something, that I had excelled in some way, you know, than that I'd just been in a nice, caring relationship.

C: Yeah, but, I had worked for this older man, and once he told me that he had spent all of his life thinking about his career and his work. And he was 52, and it suddenly struck him that he had never really given anything of himself. His life was for no one and nothing. He was almost crying saying that.

You know, I believe if there's any kind of God, it wouldn't be in any of us. Not you, or me. But just, this little space in between.

If there's any kind of magic in this world, it must be in the attempt of understanding someone, sharing something. I know, it is almost impossible to succeed. But who cares, really? The answer must be in the attempt.

- **Before Sunrise (1995)**

Education Statement of Graduate School Experimental Plant Sciences

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: **Roel F.H.M. van Bezouw**
Date: **21 June 2022**
Group: **Laboratory of Genetics**
University: **Wageningen University & Research**

1) Start-Up Phase	*date*	*cp*
► **First presentation of your project**		
Systematic analysis of epistatic interactions in photosynthesis	Feb 5, 2017	1.5
► **Writing or rewriting a project proposal**		
► **Writing a review or book chapter**		
Bezouw, R.F.H.M. et al. (2019) Converging genomics and phenomics to study national variation in photosynthesis efficiency. *The Plant Journal* doi: 10.1111/tpj.14190	Jan 2019	6.0
► **MSc courses**		
Subtotal Start-Up Phase		7.5

2) Scientific Exposure	*date*	*cp*
► **EPS PhD student days**		
EPS *Get2Gether*, Soest	Feb 15-16, 2018	0.6
EPS *Get2Gether*, Soest	Feb 11-12, 2019	0.6
► **EPS theme symposia**		
EPS theme 3 "Metabolism and Adaptation", Wageningen	Mar 13, 2018	0.3
EPS theme 3 "Metabolism and Adaptation", Wageningen	Nov 5, 2021	0.3
EPS theme 4 symposium & Willie Commelin Scholten day "Interactions between Plants and Biotic Agents", Amsterdam	Jan 24, 2018	0.3
► **Lunteren Days and other national platforms**		
Annual Meeting "Experimental Plant Sciences", Lunteren	Apr 10-11, 2017	0.6
Annual Meeting "Experimental Plant Sciences", Lunteren	Apr 9-10, 2018	0.6
Annual Meeting "Experimental Plant Sciences", Lunteren	Apr 8-9, 2019	0.6
Annual Meeting "Experimental Plant Sciences", online	Apr 12-13, 2021	0.5
Conference of the Netherlands Society for Evolutionary Biology, online	Apr 19-20, 2021	0.4
► **Seminars (series), workshops and symposia**		
Boosting Crop Productivity, KeyGene morning seminar	Jun 13, 2017	0.2
Veena Roa, The influence of Haldane on evolutionary biology in India	Aug 15, 2017	0.1
Richard Stouthamer, Nailing the borers; Biocontrol of invasive ambrosia beetles (Euwallaca fornicatus species complex) in Southern California	Sep 19, 2017	0.1
Richard Lenski, Adaptation and genome evolution in a long-term experiment	Sep 31, 2017	0.1
Gert Kema, The hidden sex life of a plant pathogen unveils a new model for disease epidemiology	Oct 31, 2017	0.1
Big data in Agrogenomics, KeyGene afternoon seminar	Nov 30, 2017	0.2
Gerard Jagers, Evolution; Process or Pattern	Sep 18, 2018	0.1
Bob Schmitz, Epigenomics studies of natural and induced alleles in plants	Jun 6, 2018	0.1

	Francois Tardieu, Phenomics in the era of sensor networks, big data and genomic prediction: where are we going?	Jan 15, 2019	0.1
	Andrew Simkin, Metabolic engineering to enhance photosynthesis and increase crop yield	Mar 22, 2019	0.1
	Maheshi Dassanavake, Multi-Ion salt stress adaptation explored using extremophyte genomics	May 20, 2019	0.1
	Mark Tester, Genetic characterization of salinity tolerance traits to increase salinity tolerance in crops - and the delivery of research outcomes	Sep 12, 2019	0.1
	Ivan Baxter, Mind the GxE = P, elemental content in plants	Dec 9, 2019	0.1
	Jacob Weiner, Applying Evolutionary Theory to Improve Plant Production	Feb 25, 2020	0.1
	Dr. Julien Dutheil, The evolution of stochastic gene expression.	Sep 22, 2020	0.1
	Mirjam Lemmens-Pott, PSG regulations concerning the Nagoya Protocol	Sep 29, 2020	0.1
	Antoine Languillaume, Caring about Code	Jun 9, 2020	0.1
	Two-part webinar on PanGenomes; Part I & Part II; KeyGene afternoon seminars	Feb 17 & Mar 18, 2021	0.2
	Daniel Croll, Drivers and brakes of pathogen emergence , WEES seminar	Feb 18, 2021	0.1
	"Artificial intelligence in plant science and breeding", afternoon seminar organized by Plant-RX Initiative	Feb 24, 2021	0.2
	Plantae Presents - Sheng-Yang He and Jonathan Jacobs	17 Mar, 2021	0.1
	Plantae Presents - Sophia Stone and Sjon Hartman	15 Apr, 2021	0.1
	Public lectures by Tracy Lawson & Amelia Henry	2 Dec, 2021	0.2
	Apomixis: the breakthrough breeding technology for the 2020s; KeyGene afternoon seminar	Feb 17, 2022	0.2
	Symposium: Genotype to Phenotype Modelling of Plant Adaption, Wageningen	Nov 16, 2017	0.3
	Symposium: Photosynthesis conference for crop yield and photosynthesis, Wageningen	Jun 20, 2018	0.3
	Symposium: Breeding Data: Statistical Advances in Modern Plant Breeding, Wageningen	Oct 16, 2018	0.3
	Symposium: Land Plant Evolution & Improving Photosynthesis and Crops, Wageningen	Jun 20, 2019	0.3
	Symposium: Optimising photosynthesis for societal needs with Professor Donald Ort, Wageningen	Mar 10, 2020	0.3
►	**Seminar plus**		
►	**International symposia and congresses**		
	5th Plant Genomics and Gene Editing Congress, Rotterdam	May 14-15, 2018	0.6
	Society of Evolutionary Biology meeting 2021, online	Jun 29-Jul 1, 2021	0.9
	Digicrops conference 2022, online	Mar 28-30, 2022	0.6
►	**Presentations**		
	Annual Meeting "Experimental Plant Sciences" 2017 (Poster)	Apr 11, 2017	1.0
	5th Plant Genomics and Gene Editing Congress (Poster)	May 14-15, 2018	1.0
	Annual Meeting "Experimental Plant Sciences" 2021 (Talk)	Apr 13, 2021	1.0
	Society of Evolutionary Biology meeting 2021 (Poster & Pitch)	Jun 29-Jul 1, 2021	1.0
►	**IAB interview**		
►	**Excursions**		
	Rijk Zwaan online company visit organized by EPS PhD council	Jun 16, 2021	0.2
	Subtotal Scientific Exposure		15.6

3) In-Depth Studies	*date*	*cp*
► **Advanced scientific courses & workshops**		
PE&RC & SENSE: R and Big Data, Wageningen	Oct 5-6, 2017	0.6
PE&RC & SENSE: Bayesian statistics, Wageningen	Oct 25-26, 2017	0.6
WUR Campus & Facilities: Basic course High Performance Computing Cluster, Wageningen	Oct 31, 2017	0.2
WUR Campus & Facilities: Basic couse Linux, Wageningen	May 19, 2018	0.2
STARGATE: An introduction to metabolomics, online	Dec 6-8, 2021	0.8

	Physalia-courses: Genomic-wide prediction of complex traits in humans, plants and animals, online	Feb 7-11, 2022	1.5
►	**Journal club**		
	Botanical Genetics Literature discussions	2017-2018	1.5
	Scientific peer review of a manuscript in *Plant, Cell & Environment*	Aug, 2021	0.3
►	**Individual research training**		
	Subtotal In-Depth Studies		5.7

4) Personal Development		*date*	*cp*
►	**General skill training courses**		
	EPS: PhD introduction course, Wageningen	Sep 26, 2017	0.3
	WGS: Competence Assessment, Wageningen	Jun 13, 2018	0.3
	WGS: Bridging across cultural differences, Wageningen	Nov 30 & Dec 12, 2018	0.7
	WGS: Critical thinking and argumentation, Wageningen	Nov 28, 2019	0.3
	SEB international: Presenting at an online conference, online	May 06, 2021	0.1
	WGS: Career perspectives, Wageningen	Jun 7-Jul 5, 2021	1.6
►	**Organisation of meetings, PhD courses or outreach activities**		
	Helping with Fascination of Plants Day 2019	May 16, 2019	0.3
►	**Membership of EPS PhD Council**		
	Subtotal Personal Development		3.6

TOTAL NUMBER OF CREDIT POINTS*	**32.4**
Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.	
** A credit represents a normative study load of 28 hours of study.*	

The work as described in this thesis was performed at the Laboratory of Genetics at the Wageningen University and financially supported by a research grant from Rijk Zwaan Breeding B.V..

The printing costs for this thesis were covered by the Laboratory of Genetics

Cover design; Roel van Bezouw & Suzanne Vlinderveld, with additional help from Proefschriftmaken.nl

Printed by: Proefschriftmaken.nl